An integrated, multicriteria, spatial decision support system, incorporating environmental, social and public health perspectives, for use in geoenergy and geoenvironmental applications

Muhammad Irfan

Geoenvironmental Research Centre

School of Engineering

Cardiff University

Thesis submitted in candidature for the degree of Doctor of Philosophy at Cardiff University

December 2014

DECLARATION

This work has not previously been accepted in substance for any degree and is not concurrently submitted in candidature for any degree.

Signed ………………………. (Muhammad Irfan) Date ………08/12/2014……..

STATEMENT 1

This thesis is being submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy (PhD).

Signed ………………………. (Muhammad Irfan) Date ………08/12/2014……..

STATEMENT 2

This thesis is the result of my own independent work/investigation, except where otherwise stated. Other sources are acknowledged by explicit references.

Signed ………………………. (Muhammad Irfan) Date ………08/12/2014……..

STATEMENT 3

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed ………………………. (Muhammad Irfan) Date ………08/12/2014……..

ACKNOWLEDGEMENTS

Praise be to Almighty Allah for his endless blessings bestowed on me, for giving me the opportunity, wisdom, courage and strength to complete my PhD from the highly prestigious Cardiff University.

I would like to say special thanks to my Supervisors Prof Hywel Thomas, Dr Aleksandra Koj and Dr Majid Sedighi for their guidance, coaching, encouragement and support throughout the time that enabled me to carry out this research and to complete my PhD.

The financial support from Welsh European Funding Office (WEFO) through the GRC's Seren project is highly acknowledged. It would have been impossible for me to pursue PhD without this financial support.

I would like to acknowledge the Welsh Government, Local Government Data Unit Wales, NHS Wales, Election Centre of Plymouth University, BGS, CCW, DTI, DECC, DEFRA, EEA, Experian, ONS, OS and UKDA for giving me access to their highly valuable datasets required for this research.

I would like to say thanks to my colleagues in Seren project for their cooperation. Special thanks to my friends Manju, Nishant, Shauna, Ram, Jamie, Lee, Mojgan, Mat, Sanjana, Shakil, Ali, Adeel, Saad, Hassan, Salman, Faisal, Wasim, AJ, Renato, Alejandro, Claire, Dan, Vasilis, Siva, Jack, Eleni, Richard, Manhui, Amilia, Dariusz and Panos for their cheer.

Above all, I would like to thank my family for their love, support and reassurance which enabled me to complete my PhD. The prayers, affection, kindness and backing of my parents, my wife, my lovely little daughter, brother and sisters have been the source of strength and encouragement for me. Special thanks to my lovely wife for all her sacrifices during the hard times which let me focus on my PhD.

SUMMARY

A new Spatial Decision Support System (SDSS) has been designed and developed to address a wide spectrum of semi-structured spatial decision problems. These problems are related to site selection, site ranking and impact assessment. The proposed SDSS is conceptualised as a holistic, informed and impact-based multicriteria decision framework.

The system has been developed using the .NET C# programming language and open source geoinformatics technologies such as DotSpatial and SpatiaLite. A combination of existing Multi Criteria Decision Analysis (MCDA) and Artificial Intelligence (AI) techniques, with a few novel variations have been developed and incorporated into the SDSS. The site selection module utilises a theme-based Analytical Hierarchy Process (AHP) and Weighted Linear Combination (WLC). Two site ranking techniques have been introduced in this research. The first technique is based on the systematic neighbourhood comparison of sites with respect to key indicators. The second technique utilises multivariate ordering capability of the one-dimensional Self-Organizing Maps (SOM) to rank the sites. The site impact assessment module utilises a theme-based Rapid Impact Assessment Matrix (RIAM). A spatial variant of the General Regression Neural Networks (GRNN) with a genetic algorithm for optimisation has been developed for the prediction and regression analysis. A number of other spatial knowledge discovery and geovisual-analytics tools have been provided in the system to facilitate spatial decision making process.

An application of the SDSS has been presented to investigate the potential of Coalbed Methane (CBM) development in Wales, UK. Most potential sites have been identified by utilising the site selection and site ranking tools of the developed SDSS. An impact assessment has been carried out on the best sites by using Rapid Impact Assessment Matrix. Further analysis has uncovered the spatial variability expected in the potential impacts of the sites, considering key indicators. The application has demonstrated that the developed system can help the decision makers in providing a balanced regime of social, environmental, public health and economic aspects into the decision making process for engineering interventions..

The generic nature of the developed system has extended the concept of Spatial Decision Support System to address a range of spatial decision problems, thereby enhancing the effectiveness of the decision making process. The developed system can be considered as a useful modern governance tool, incorporating the key factors into decision making and providing optimal solutions for the critical questions related to energy security and economic future of the region.

LIST OF ACRONYMS

AADF	*Annual Average Daily Flow*
ADO	*Active Data Object*
AHP	*Analytical Hierarchy Process*
AI	*Artificial Intelligence*
ANN	*Artificial Neural Network*
ANP	*Analytical Network Process*
AONB	*Areas of Outstanding Natural Beauty*
BGS	*British Geological Survey*
BHPS	*British Household Panel Survey*
BLOB	*Binary Large Object*
BMU	*Best Matching Unit*
CBM	*Coalbed Methane*
CCS	*Carbon Capture and Storage*
CCW	*Countryside Council for Wales*
CDM	*Clean Development Mechanism*
CI	*Consistency index*
CP	*Component Planes*
CR	*Consistency Ratio*
CSM	*Criterion Sorting Mechanism*
DECC	*Department of Energy & Climate Change, UK*
DEFRA	*Department of Environment Food and Rural Affairs, UK*
DESYRE	*DEcision Support sYstem for the REqualification of contaminated sites*
DLUA	*Developed Land Use Areas*
DoT	*Department of Traffic, UK*
DSS	*Decision Support System*
DTI	*Department of Trade and Industry, UK*
ECBM	*Enhanced Coalbed Methane*
EEA	*European Environment Agency*
EF	*Ecological Footprint*
EIA	*Environmental Impact Assessment*

EPA	*Environmental Protection Agency*
EU	*European Union*
GA	*Genetic Algorithms*
GHG	*Green House Gas*
GIS	*Geographical Information System*
GRNN	*General Regression Neural Networks*
GSH	*Ground Source Heat*
GUI	*Graphical User Interface*
GWR	*Geographically Weighted Regression*
HMW	*Health Maps Wales*
IAIA	*International Association for Impact Assessment*
IDW	*Inverse Distance Weighting*
IEA	*International Energy Association*
LA	*Local Authority*
LNR	*Local Nature Reserves*
LSOA	*Lower Super Output Areas*
LULC	*Landuse Landcover*
MBR	*Minimum Bounding Rectangles*
MCDA	*Multicriteria Decision Analysis*
MDI	*Multi Document Interface*
MLP	*Multi-layer perceptron neural network*
MNR	*Marine Nature Reserves*
MSOA	*Medium Super Output Areas*
NAEI	*National Air Emissions Inventory, UK*
NIMBY	*Not in My Back Yard*
NNR	*National Nature Reserves*
NSW	*National Survey for Wales*
OGC	*Open Source Geospatial Consortium*
ONS	*Office of the National Statistics*
OS	*Ordinance Survey of Britain*
OWA	*Ordered Weighted Averaging*
PCP	*Parallel Coordinate Plots*
PM2	*Particulate Matter*

PNN	*Probabilistic neural network*
RDBMS	*Relational Database Management System*
RI	*Random Index*
RIAM	*Rapid Impact Assessment Matrix*
RMSE	*Root Mean Squared Error*
SAC	*Special Areas of Conservation*
SADA	*Spatial Analysis and Decision Assistance*
SAW	*Simple Additive Weighting*
SDLC	*Software Development Life Cycle*
SDSS	*Spatial Decision Support Systems*
SNR	*Signal to Noise Ratio*
SOM	*Self-Organizing Maps*
SPA	*Special Protection Areas*
SQL	*Structured Query Language*
SSSI	*Site of Special Scientific Interest*
TEV	*Total Economic Value*
TOPSIS	*Technique for Order of Preference by Similarity to Ideal Solution*
UCG	*Underground Coal Gasification*
USOA	*Upper Super Output Areas*
VOC	*Volatile Organic Compounds*
WHO	*World Health Organization*
WIMD	*Welsh Index of Multiple Deprivation*
WLC	*Weighted Linear Combination*
WHS	*Welsh Health Survey*

CONTENTS

1

INTRODUCTION

1.1 Introduction

Decision makers are often interested in achieving higher benefits with least risks involved through an effective decision making. A modern outlook of the decision making involves a number of socio-economic, environmental and public health risks and benefits to answer critical questions related to engineering interventions.

Many decision problems faced by decision makers involve spatial information (Rushton 2001). A rapid growth has been observed in the use of spatial information in various fields. Spatial decision problems combine a certain degree of both soft and hard information. Hard information is represented by quantitative and qualitative data, whereas soft information is comprised of a decision maker's preferences, priorities and judgements (Malczewski 1999). Spatial decision problems are often multicriteria in nature therefore it is hard to find a solution which clearly dominates other alternatives, in terms of the entire criteria (Malczewski 1999). Multicriteria decision making can include both multi-objective and multi-attribute scenarios which can increase problem complexity (Malczewski 1999).

Decision makers increasingly rely on Spatial Decision Support Systems (SDSS) to address multicriteria, semistructured spatial decision problems (Sugumaran and Degroote 2010). Multicriteria SDSS are being used in a number of applications especially in the renewable energy sector, e.g. (Simão et al. 2009; Perpiña et al. 2013; Silva et al. 2014; Wanderer and Herle 2014; Mekonnen and Gorsevski 2015) and in Geoenvironmental applications, e.g. (Wang et al. 2010; Agostini et al. 2012; Salim 2012; Cothren et al. 2013; Demesouka et al. 2013; Uyan et al. 2013; Comino et al. 2014; Zanuttigh et al. 2014)

An SDSS provides advanced analytical modelling techniques to address complex spatial decision problems, in addition to the core functions of a Geographical Information System (GIS) (Patel 2007). An SDSS is aimed at improving the effectiveness of the spatial decision making process (Malczewski 1999). The spatial decision making process has three important phases: a) Intelligence, b) Design and c) Choice (Malczewski 1999). Intelligence is the identification of the decision problem or an opportunity for improvement. Design is the identification or development of the possible solutions (alternates) by using adequate analytical techniques. Choice is the selection process of the best alternate through a systematic evaluation of the alternatives.

SDSSs are specialised in nature and are designed to address domain-specific spatial decision problems (Sugumaran and Degroote 2010). However, several spatial decision problems are repetitive in nature and common to many application areas. For example, decision problems related to site selection, site ranking and impact assessment are common.

Decision makers often face similar spatial decision problems in Geoenergy and Geoenvironmental applications. Geoenergy developments, e.g. unconventional gas, can not only provide energy security but also contribute to the local socio-economic developments. Some of these new technologies can at the same time help global climate change mitigation through carbon capture and storage (IEA 2014). The spatial decision problems related to these engineering interventions are multi-objective and multi-attribute in nature, based on their potential impacts on economy, environment, public health and society. There is a pressing need for an integrated and holistic approach to incorporate these impacts into the decision making process, through designing and developing a generic SDSS with adequate, reliable and robust modelling techniques.

Several, commercial and non-commercial GIS software such as IDRISI, ArcGIS, SAGA and ILWIS provide a number of modelling techniques and a mechanism to customise and bundle them together to serve as an SDSS (Sugumaran and Degroote 2010). To the author's knowledge, there is no existing generic SDSS that can be used to address site selection, site ranking and impact assessment under one system that can be applied in a variety of Geoenergy and Geoenvironmental applications. In this research, it is envisaged to design and develop a generic SDSS to tackle these common spatial decision problems. It is suggested that an integrated and scalable system can serve the purpose if it is provided with a range of adequate analytical techniques based on Multicriteria Decision Analysis (MCDA) and Artificial Intelligence (AI). Also, the system should be aided with a comprehensive geodatabase comprising key aspects from socio-economic, environmental and public health domains in an integrated manner. There are certain scientific challenges involved in achieving the aim stated above. Firstly, the identification of the most appropriate analytical modelling techniques to tackle these problems. Secondly, the identification of the key environmental, socio-economic, public health and techno-economic factors and indicators that should be incorporated in the multicriteria spatial decision analysis.

Subsequent to the above research elements, the aim and objectives of this research are stated in Section 1.2. A review of the Geoenergy and Geoenvironmental applications is presented in Section 1.3. Scope and limitations of the research are discussed in Section 1.4. An overview of the entire thesis is presented in Sections 1.5.

1.2 Aim and objectives

This research aims at designing and developing an impact-based, multicriteria Spatial Decision Support System to address a wide spectrum of spatial decision problems related to Geoenergy and Geoenvironmental applications.

In order to achieve this aim, the following objectives have been identified for the research:

- The design and development of a multicriteria Spatial Decision Support System (SDSS), thereby facilitating decision making process related to site selection, site ranking and impact assessment.
- The development of a set of adequate advanced analytical techniques and their integration into the SDSS to tackle semi-structured spatial decision problems.
- The exploration and identification of the key socio-economic, environmental, public health and techno-economic factors and indicators to be adopted in informed, impact-based decision making related to Geoenergy and Geoenvironmental applications.
- The exploration and development of a geodatabase containing the key socio-economic, techno-economic, environmental and public health data for the study area, i.e. Wales, UK.
- The investigation of the potential of CBM-ECBM development in Wales considering the socio-economic, environmental, public health and techno-economic aspects.

1.3 Geoenergy and Geoenvironmental applications

The phenomena of climate change and anthropogenic activities and their impacts on our environment, health and society are widely being acknowledged nowadays. The global economy is still largely dependent on the fossil fuel resources (IEA 2014). The policies responding to the climate change and reducing greenhouse gas emissions may impact the future economic development and sustainability. Therefore, energy security and climate change are together seen as the key drivers for future energy policy (IEA 2007). Natural gas is among the cleanest fossil fuel that can contribute to energy security and diversity during the transition period between fossil fuel dominated economy to the renewable energy based economy (Weijermars et al. 2011).

The term “Geoenergy” in this research has been used to represent a combination of unconventional ground source energy resources and CO_2 storage opportunity. Unconventional gas is an umbrella term used by the International Energy Association (IEA) (IEA 2013a) for the natural gas recovered from organic rich shale formations, Underground Coal Gasification (UCG), Coalbed Methane (CBM) and tight gas. Shale gas is composed of methane that is trapped in organic rich shale formations whereas tight gas is methane trapped in very low permeable rock formations such as sandstone (IEA 2013a).

Coalbed Methane (CBM) is the methane resource that is adsorbed in the coal within the coal seams (IEA 2013a). Also, gas recovery can be enhanced from the coalbed by injecting CO2 into the coal seam in superficial form (Mazzotti et al. 2009). This process is called Enhanced Coalbed Methane (ECBM) (White et al. 2005).

Figure 1.1 shows a worldwide potential resource of CBM, shale and tight gas by the end of 2012 (IEA 2013b). According to the information displayed in Figure 1.1, there are large reserves of remaining unconventional gas in different regions of the world including the

European Union (EU). Shale gas is the dominant resource in many regions, also there is a huge potential of CBM development in Russia, China, USA and Australia.

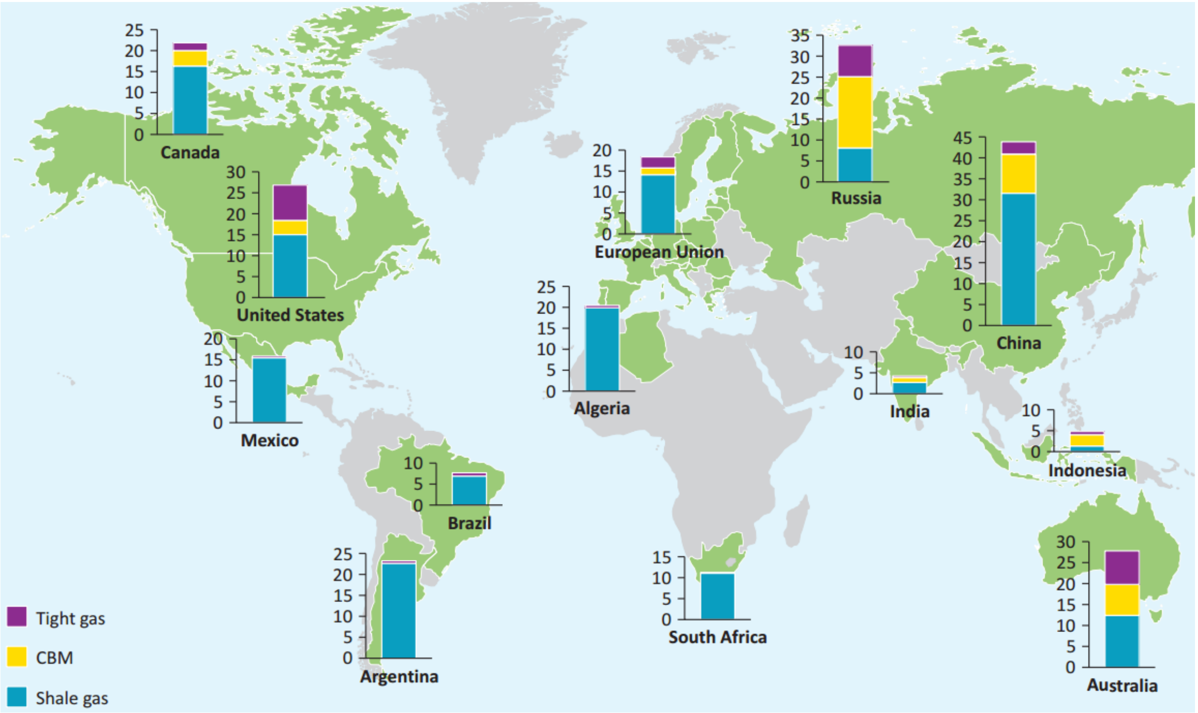


Figure 1.1 Remaining unconventional gas resources (Unit: trillion cubic meters) in selected regions by the end of year 2012 (IEA 2013b)

Similarly, UK has a large CBM resource that can be utilised to meet the energy demand (Jones et al. 2004). An estimate of this resource is reported to be as large as 2,900 billion cubic meters (DECC 2010). A considerable CBM potential has also been identified in the study area of this research, i.e. in North and South Wales coalfields in UK (Jones et al. 2004).

Complex spatial decision problems such as Site selection, ranking and impact assessment are often faced in other Geoenvironmental applications, for example waste disposal or nuclear power plant site (Malczewski 1999). The proposed SDSS is designed and developed to facilitate spatial decision making process in a large number of Geoenergy and Geoenvironmental applications.

1.4 Scope and limitations

The scope and limitation of the research are listed below:

- Although the developed SDSS can be used for a range of Geoenvironmental applications, Geoenergy applications remain the focus in this research.
- The developed system is independent of the geographical location and underlying data in the geodatabase. However, to demonstrate an application of the system, Wales UK has been selected as the study area in this research. Data has been collected and incorporated into the geodatabase to present an application of the developed system to explore CBM-ECBM opportunities in the Wales UK.
- The functionality of the SDSS developed in this research is explored for site selection, site raking, site impact assessment and spatial knowledge discovery.
- Selection of indicators and their relative weights in the site selection and ranking process entirely reside with the decision makers and not with the system.
- Different datasets collected for this research cover different timeframes and they have been acquired from various sources during 2011-12. Some of the datasets may have new versions available.
- Data has been acquired from multiple sources and exploration of their accuracy and scale is beyond the scope of this research.
- The survey data used and GIS modelling carried out to generate composite indicators such as “Social Acceptance” and “Social Capital” is of secondary importance in this research. The focus is therefore on how SDSS can utilise such information and assist with the spatial decision making process.
- Site characterisation is not included in the SDSS since it requires details about the site and its operations. Also, there are existing useful tools available for the purpose.

1.5 Thesis overview

A brief description of the thesis structure is given below:

Chapter 2 covers a selected literature review of the design and essential components of the Spatial Decision Support Systems (SDSS) especially in accordance with Geoenergy and Geoenvironmental spatial problems. It also covers the existing Multi Criteria Decision Analysis (MCDA) techniques used for the tackling spatial decision problems. This chapter also highlights the important environmental, public health and socio-economic risks and benefits linked with the considered engineering interventions. Key indicators representing these risks and benefits have also been identified.

Chapter 3 covers the design consideration and architecture of the system developed in this research. The modular approach taken to develop the system is explained. The system design is based on the general three-component architecture of the SDSS (Malczewski 1999). These components are i) Geodatabase, ii) Model base and iii) User Interface.

Chapter 4 and 5 explain the development of the system and its analytical modules. Chapter 4 covers those SDSS modules that are based on Artificial Intelligence techniques such as Artificial Neural Network (ANN). Chapter 5 presents those SDSS modules that are based on Multicriteria Decision Analysis (MCDA) such as Analytical Hierarchy Process (AHP). Some novel variations of the existing techniques have been introduced in the analytical modules.

Chapter 6 presents the verification performed for the different analytical modules of the SDSS. Existing tools such as Matlab and ArcGIS have been used to verify the code of the analytical modules and in some cases results are also validated by comparing the results with other published material.

In Chapter 7, the design and development of the Geodatabase has been explained. For the application of the SDSS, adequate spatial data from Wales (UK) has been incorporated in the geodatabase to cover environmental, socio-economic, public health and techno-economic aspects. Using these indicators, an application of the developed SDSS has been presented in Chapter 8. The application covers the site selection, ranking and impact assessment for potential Coalbed Methane (CBM) and Enhanced Coalbed Methane (ECBM) sites in Wales.

In Chapter 9, overall research work is summarised, conclusions are drawn by highlighting the contributions of this research and finally suggestion are made for improvement and further research.

References

Agostini, P.,Pizzol, L.,Critto, A.,D'Alessandro, M.,Zabeo, A. and Marcomini, A. 2012. Regional risk assessment for contaminated sites Part 3: Spatial decision support system. *Environment International* 48(0), pp. 121-132.

Comino, E.,Bottero, M.,Pomarico, S. and Rosso, M. 2014. Exploring the environmental value of ecosystem services for a river basin through a spatial multicriteria analysis. *Land Use Policy* 36(0), pp. 381-395.

Cothren, J.,Thoma, G.,DiLuzio, M. and Limp, F. 2013. *Integration of Water Resource Models with Fayetteville Shale Decision Support and Information System*.

DECC. 2010. *The unconventional hydrocarbon resources of Britain's onshore basins - shale gas*. UK: Department of Energy and Climate Change.

Demesouka, O. E.,Vavatsikos, A. P. and Anagnostopoulos, K. P. 2013. Suitability analysis for siting MSW landfills and its multicriteria spatial decision support system: Method, implementation and case study. *Waste Management* 33(5), pp. 1190-1206.

IEA. 2007. *Energy Security and Climate Policy - Assessing Interactions*. Paris: IEA.

IEA. 2013a. *Resources to Reserves 2013*. Paris: IEA.

IEA. 2013b. *World Energy Outlook 2013*. Paris: IEA.

IEA. 2014. *World Energy Outlook 2014*. Paris: IEA.

Jones, N. S.,Holloway, S.,Smith, N. J. P.,Browne, M. A. E.,Creedy, D. P.,Garner, K. and Durucan, S. 2004. *UK coal resources for new exploitation technologies*.

Malczewski, J. 1999. *GIS and Multicriteria Decision Analysis*. Wiley.

Mazzotti, M.,Pini, R. and Storti, G. 2009. Enhanced coalbed methane recovery. *The Journal of Supercritical Fluids* 47(3), pp. 619-627.

Mekonnen, A. D. and Gorsevski, P. V. 2015. A web-based participatory GIS (PGIS) for offshore wind farm suitability within Lake Erie, Ohio. *Renewable and Sustainable Energy Reviews* 41(0), pp. 162-177.

Patel, N. 2007. Spatial Decision Support System.*Modeling In Resource Management And Environment Through Geomatics*. pp. 195-205.

Perpiña, C.,Martínez-Llario, J. C. and Pérez-Navarro, Á. 2013. Multicriteria assessment in GIS environments for siting biomass plants. *Land Use Policy* 31(0), pp. 326-335.

Rushton, G. 2001. Spatial Decision Support Systems. In: Baltes, N.J.S.B. ed. *International Encyclopedia of the Social & Behavioral Sciences*. Oxford: Pergamon, pp. 14785-14788.

Salim, M. G. 2012. Selection of groundwater sites in Egypt, using geographic information systems, for desalination by solar energy in order to reduce greenhouse gases. *Journal of Advanced Research* 3(1), pp. 11-19.

Silva, S.,Alçada-Almeida, L. and Dias, L. C. 2014. Biogas plants site selection integrating Multicriteria Decision Aid methods and GIS techniques: A case study in a Portuguese region. *Biomass and Bioenergy* (0).

Simão, A.,Densham, P. J. and Haklay, M. 2009. Web-based GIS for collaborative planning and public participation: An application to the strategic planning of wind farm sites. *Journal of Environmental Management* 90(6), pp. 2027-2040.

Sugumaran, R. and Degroote, J. 2010. *Spatial decision support systems: Principles and practices*. CRC Press.

Uyan, M.,Cay, T. and Akcakaya, O. 2013. A Spatial Decision Support System design for land reallocation: A case study in Turkey. *Computers and Electronics in Agriculture* 98(0), pp. 8-16.

Wanderer, T. and Herle, S. 2014. Creating a spatial multi-criteria decision support system for energy related integrated environmental impact assessment. *Environmental Impact Assessment Review* (0).

Wang, J.,Chen, J.,Ju, W. and Li, M. 2010. IA-SDSS: A GIS-based land use decision support system with consideration of carbon sequestration. *Environmental Modelling & Software* 25(4), pp. 539-553.

Weijermars, R.,Drijkoningen, G.,Heimovaara, T. J.,Rudolph, E. S. J.,Weltje, G. J. and Wolf, K. H. A. A. 2011. Unconventional gas research initiative for clean energy transition in Europe. *Journal of Natural Gas Science and Engineering* 3(2), pp. 402-412.

White, C. M.,Smith, D. H.,Jones, K. L.,Goodman, A. L.,Jikich, S. A.,LaCount, R. B.,DuBose, S. B.,Ozdemir, E.,Morsi, B. I. and Schroeder, K. T. 2005. Sequestration of carbon dioxide in coal with enhanced coalbed methane recovery - A review. *Energy and Fuels* 19(3), pp. 659-724.

Zanuttigh, B.,Simcic, D.,Bagli, S.,Bozzeda, F.,Pietrantoni, L.,Zagonari, F.,Hoggart, S. and Nicholls, R. J. 2014. THESEUS decision support system for coastal risk management. *Coastal Engineering* 87(0), pp. 218-239.

2

LITERATURE REVIEW

2.1 Introduction

The main objective of this research is to design and develop a Spatial Decision Support System (SDSS) to support decision makers in confronting a spectrum of Geoenergy and Geoenvironmental spatial decision problems. SDSSs are used increasingly to facilitate the spatial decision making process in a variety of domains. Malczewski (2006) reviewed and classified over 300 articles on GIS-based Multicriteria Decision Analysis (GIS-MCDA). 11 % of these articles referred to a fully integrated SDSS, combining GIS and multicriteria analysis techniques in one system. This classification reveals that GIS-MCDA has been applied in a variety of fields, including environment, ecology, transportation, urban/regional planning, waste management, hydrology/water resources, agriculture, forestry, natural hazards and recreational/tourism (Malczewski 2006).

The literature review provided in this chapter covers different aspects of the SDSS designed and developed in this research. Section 2.2 presents the fundamentals of an SDSS, including its design components and characteristics. Section 2.3 covers the effective modelling techniques that are frequently used in a SDSS. In particular, MCDA and Artificial Intelligence (AI) techniques used for the modelling components of an SDSS, are covered in detail.

Section 2.4 provides a review of the Geoenergy applications considered in this research such as unconventional gas with possibility of carbon capture and storage. An application of the SDSS has been provided in Chapter 8 for the development of Coalbed Methane (CBM) and Enhanced Coalbed Methane (ECBM) in Wales, therefore these two resources are covered in detail in Section 2.4.

SDDSs are specialised in nature, they are designed and developed to cater for domain specific spatial decision problems. Whereas, Geographical Information Systems (GIS) are general purpose tools which can be applied in any field. Therefore, Section 2.5 covers a review of common spatial decision problems that can be faced in the Geoenergy applications. These problems are mostly related to site selection, site ranking, impact assessment and spatial knowledge discovery. Section 2.6 covers the most frequently used SDSS in Geoenvironmental spatial decision problems.

The last part of this chapter provides a comprehensive review of the significant aspects of Geoenergy applications. Aspects considered in this research are categorised into four domains: i) Socio-Economic, ii) Environmental, iii) Public Health and iv) Techno-Economic. The indicators, spatial and aspatial datasets associated with these domains are discussed. Using an effective design and development strategy, these key aspects will be integrated in the multicriteria SDSS.

2.2 Fundamentals of Spatial Decision Support System

This section covers the fundamentals of SDSS including its main components and characteristics. An SDSS is defined as (Malczewski 1999):

> "An interactive, computer-based system, designed to support a user or groups of users in achieving a higher effectiveness of decision making while solving a semistructured spatial decision problem".

"Structured" decisions are those that are well understood, repetitive and based on some relevant theory. These decision problems can normally be tackled by computer programs. On the other side of the decision spectrum, are those decision problems that are unstructured, not repetitive and are normally not based on a theory. As shown in Figure 2.1, semi-structured problems lie in the middle of the problem spectrum where decision makers input is required by the computers to solve complex, real life problems (Malczewski 1999).

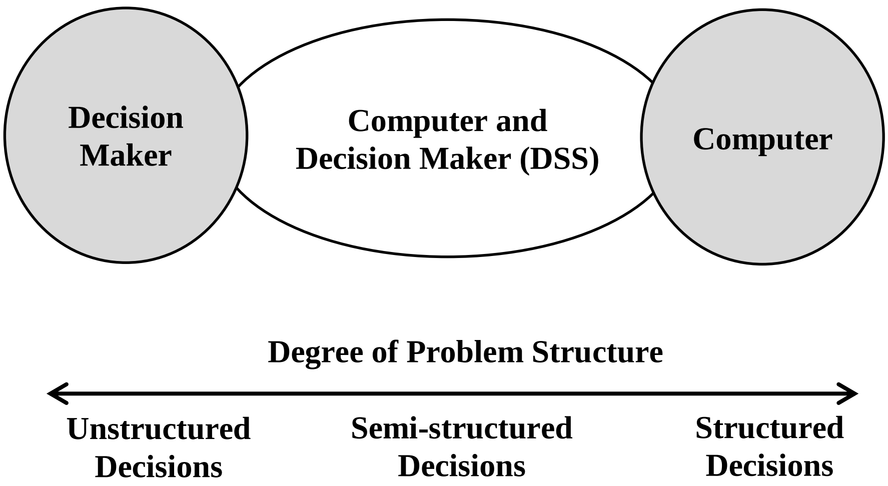


Figure 2.1 Degree of decision problem structure (Malczewski 1999)

The process of spatial decision making is explained as the tasks and features involved in the entire flow of the decision making process to solve the semi-structured or ill-structured spatial problems. Figure 2.2 shows the three-phase generalised decision making process

(Malczewski 1999). The three important phases of the decision making process are i) Intelligence, ii) Design and iii) Choice. Intelligence is the identification of the decision problem or an opportunity for improvement. Design is the identification or development of the possible solutions (alternates) based on adequate analytical models. The Choice is the selection process of the best alternate through a systematic evaluation of the alternatives (Malczewski 1999).

A computer system is considered as SDSS if it has some specific characteristics. Some of these characteristics are common to GIS including spatial data management and analysis tasks. While others are common to Decision Support Systems (DSS), such as facilitation of the decision making process to tackle semi-structured spatial problems. Sugumaran and Degroote (2010) identified the key characteristics of an SDSS as shown in Figure 2.2.

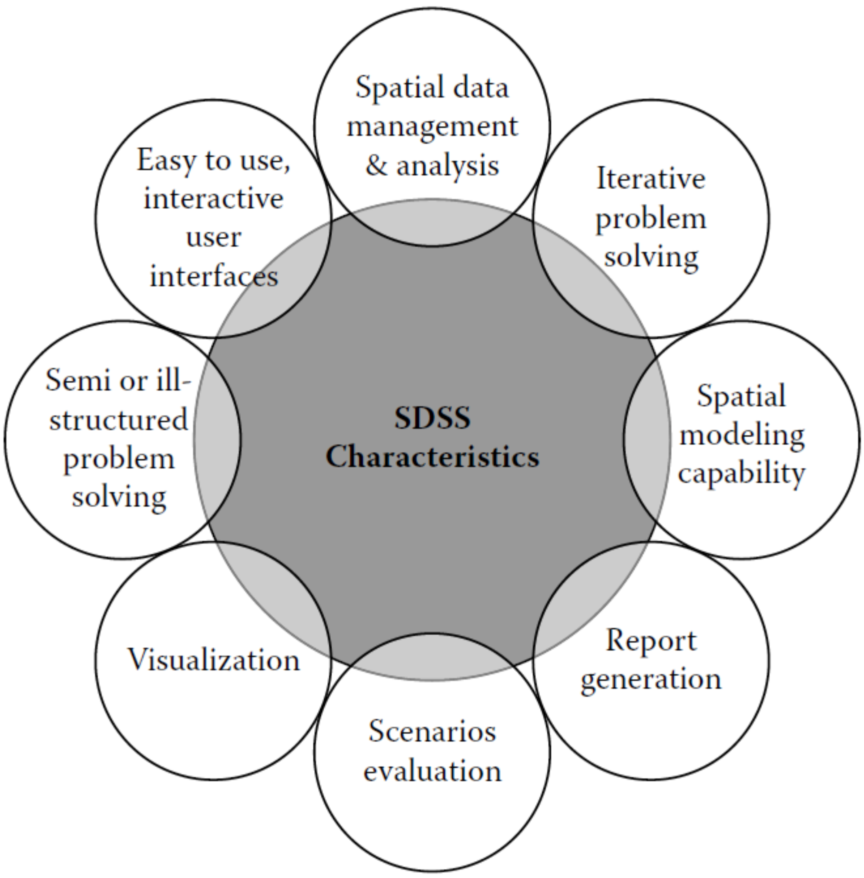


Figure 2.2 SDSS main characteristics (Sugumaran and Degroote 2010)

In order to manage large amount of spatial and non-spatial data, a Relational Database Management System (RDBMS) is usually required for an SDSS. Spatial modelling capability is the core of the SDSS and different analytical models can be used for spatial data manipulation to confront spatial decision problems. These models include numerical formulation, rule-based (if-then) models, Artificial Neural Networks (ANN), statistical models and Multicriteria Decision Analysis (MCDA) (Sugumaran and Degroote 2010). SDSS facilitates the decision making process by providing interactive graphical user interfaces, spatial modelling capability, scenario evaluation, iterative problem solving, visualisation and report generation.

Sugumaran and Degroote (2010) empahsised that SDSS are ought to address complex spatial decision problem of a specific role/domain. Because of this domain-speficic nature of the SDSS, there is no universal SDSS that can address all types of spatial decision problems.

Malczewski (1999) categorised the functions of SDSS in three broader groups: i) Database and management, ii) Model base and management and iii) Dialogue generation and management. Database management system deals with data storage, retrieval, manipulation, queries, indexing and topological relationships. Model base and management subsystem deals with functions related to the processing and analysis of information for problem solving. Model base can use different models including MCDA and ANN as described earlier. Dialogue generation provides the interface for the system to be used by the decision makers for input, output, report generation and visualisation (Malczewski 1999).

2.3 Modelling techniques used in SDSS

As discussed earlier many spatial problems are complex in nature and require appropriate analytical solutions. A number of analytical modelling techniques based on the nature of the spatial problems have been reported in literature. Some of the common techniques have been discussed in detail in the following sections, where they have been categorised into two main types: i) Spatial Multi Criteria Decision Analysis based and ii) Artificial Intelligence based.

2.3.1 Spatial Multi Criteria Decision Analysis based analytical techniques

Spatial Multicriteria Decision Analysis (MCDA) combines decision maker's preferences with geographical data to solve spatial decision problems such as site selection (Malczewski 1999). A number of techniques have been presented in the literature for MCDA including Analytical Hierarchy Process (AHP), Weighted Linear Combination (WLC) and Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS). A review of the selected publications is provided below to illustrate how these techniques have been used for spatial decision support.

AHP has been used frequently in GIS multicriteria analysis to confront spatial decision problems. AHP was first introduced by Saaty (1980), whereas Malczewski (1999) proposed its use in the GIS-based multi criteria decision analysis. AHP can be used in situations where a direct and established empirical relationship between dependent and independent variables is unknown. Also, when multiple options are available to choose from but there is no direct ranking available to help with the decision making process.

AHP is based on three principles: i) decomposition, ii) comparative judgment and iii) synthesis of priorities (Malczewski 1999). The decision problem is first decomposed into a hierarchical structure covering all the essential elements. At each level of the decision hierarchy, the components are compared and relative weights are assigned using pairwise comparison method. The priorities are then constructed at each level of the hierarchy with the help of the relative weights and scaled values of the components. These priorities are aggregated at each level, all the way up to the top level of the hierarchical tree to achieve the overall Goal.

In spatial MCDA, AHP is commonly used in combination with several other methods, including the Pairwise Comparison Method, Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS), Weighted Linear Combination (WLC) and Ordered Weighted Averaging (OWA). A review of the selected publications is given below to elaborate the integration of AHP with other techniques in spatial MCDA to facilitate spatial decision process in different domains.

TOPSIS is one of the most commonly used site ranking method in MCDA problems (Chen et al. 2011b; Jia et al. 2012). A combination of AHP and TOPSIS method has been suggested for municipal solid waste landfill site selection in a case study of Thrace region in Greece (Demesouka et al. 2013).

WLC is another commonly used multicriteria GIS analysis technique (Chen et al. 2001) and it is often used together with AHP. It is also known as Simple Additive Weighting (SAW).The relative weights are multiplied to the criterion maps and then summed up to produce the final decision map (Chen et al. 2001). For example, WLC is used in the SDSS for the site suitability of wind farms in northwest Ohio, USA (Gorsevski et al. 2013). Also, WLC has been applied in GIS based multicriteria analysis to find the optimum water harvesting ponds in Northern Jordan (Al-Adamat et al. 2010). Similarly, AHP, fuzzy membership functions and Simple Additive Weighting (SAW) has been used together to find the most suitable site for a tourist building in the rural landscape of Hervás in Spain (Jeong et al. 2013). Similarly, Rahman et al. (2012) presented spatial decision tool for the managed-aquifer recharge that uses a combination of AHP, WLC and OWA methods. A case study of this tool has been presented for the site selection of managed-aquifer recharge in the Algarve Region of Portugal (Rahman et al. 2012).

There are also certain known limitations to the AHP process. The weights assigned to the different levels in the decision hierarchy are subject to the decision maker's preferences (Nefeslioglu et al. 2013). Another problem associated with the AHP based site selection is that, it often results in a larger suitable area, especially in the case when raster based analysis is used. In such conditions, Site ranking can be useful as it provides a logical basis for the selection of the final sites out of the suitable area (Irfan et al. 2014).

2.3.2 Artificial Intelligence based analytical techniques

Artificial Intelligence (AI) techniques are becoming more popular to be used in the spatial decision support systems. This increase is due to the flexible soft computing paradigm of the AI techniques which helps facilitating the decision making in confronting complex spatial decision problems. It is evident from the literature review provided below that Artificial Neural Networks (ANN) and Genetic Algorithms (GA) based techniques have been used extensively in SDSS. Following sections describe the fundamentals of ANN and GA and how they are being utilised in the SDSS for spatial decision problem solving.

2.3.2.1 Artificial Neural Networks

Artificial Neural Network (ANN) mimics the learning and decision making process of the biological neurons to solve real life complex problems (Schalkoff 2011). The structure and working of ANN depend on its type. However, a general structure of ANN consists of layers of interconnected nodes (neurons). The input is given at the input layer. This information is then processed and progressed towards the neurons present at the next layers. The output is generated at the output layer of neurons.

Some of the commonly used ANN include: feed forward, back propagation, Self-Organizing Maps (SOM) and General Regression Neural Networks (GRNN). Implementation of ANN is not always straight forward and user input is required for setting up the right structure of ANN and its efficient training for a given problem (Nasseri et al. 2008). For this reason, only those types of ANN are considered in this research that i) require very little input from user, ii) are simple in structure, iii) require unsupervised or semi-supervised training and iv) their robustness and reliability have already been proven. Considering this criteria two types of ANN are selected for this research a) SOM and b) GRNN. SOM are unsupervised ANN and are useful for dimension reduction, clustering and pattern recognition from multidimensional data (Olawoyin et al. 2013). GRNN are powerful function approximates, capable of modelling linear and non-linear relationships in data and are very simple in their structure and working (Currit 2002). A review of selected publications is provided below to demonstrate the use of ANN in spatial decision making process.

Olawoyin et al. (2013) have used SOM to visualise trends in the quality of water, soil and sediment samples from four areas in the Niger Delta (Nigeria) for the effective decision making and remedial actions. SOM was able to identify the areas with high concentration of contaminants using the physical, eco-toxicological and chemical features in the samples from different geographical regions (Olawoyin et al. 2013). Similarly, SOM has been utilised to study the inter-relationships of multivariable soil data and to analyse the effects of soil physical properties on soil chemical/hydraulic processes (Merdun 2011).

Carlei and Nuccio (2014) presented a new approach to study the spatial agglomeration of the economic activities by recognising patterns in data. The approach used is based on the Self-Organizing Maps (SOM) which is a type of unsupervised ANN (Carlei and Nuccio 2014). A case study of clothing industry in Italy has been presented. The results show that the SOM based approach can be used to effectively identify spatial agglomeration in terms of industrial patterns (Carlei and Nuccio 2014).

Tayfur et al. (2014) investigated the use of GRNN in predicting the runoff in two small sub-catchments of Tiber River Basin in Italy. The rainfall and soil moisture information at different depths of soil have been used for prediction of runoff. The GRNN prediction was found to be satisfactory in relation to the actual runoff, with coefficient of determination R_2 equal to 0.87 (Tayfur et al. 2014).

Mostafa and Nataraajan (2009) used three types of neural networks to predict and classify per-capita Ecological Footprint (EF) of 140 nations. The neural networks used in the study include multi-layer perceptron neural network (MLP), probabilistic neural network (PNN) and GRNN. The results reveal that neural networks outperform traditional statistical methods used for the mentioned purpose (Mostafa and Nataraajan 2009).

Literature review presented above shows that SOM and GRNN have been effectively used for spatial decision support, particularly, for clustering, prediction, and knowledge discovery.

2.3.2.2 Genetic Algorithms

Although GRNN and SOM are simple in terms of their structure, input from user is still required to define their structure and other essential parameters. Genetic Algorithms (GA) since they were first introduced by Holland (1975), have been used extensively for machine learning and ANN optimisation. GA are structurally inspired by the natural process of evolution. Mutation and crossover phenomenon of the genes in natural evolutionary process is incorporated in the Genetic Algorithms (Holland 1975).

GA contains multiple generations and each generation has an even number of individuals. Each individual contains a number of genes which may contain a potential solution for a given problem. New individuals are created by combining different genes of the two selected individuals in a given population. The number of genes coming from each parent depends on the cross over rate provided by the user. Some of the genes in the offspring mutate and flip during this process and take a new shape which is different than the parent genes. This ensures the variety and diversity of genes in the individuals (Holland 1975). Natural process of evolution is adopted in the GA by creating new individuals from those having better solutions to the problem (Holland 1975).

Similarly, Nasseri et al. (2008) discussed the network structure and efficient training being the major obstacles in the effectiveness of ANN for rainfall forecasting. To overcome this obstacle, they have used GA with back-propagation training algorithm to find the appropriate ANN architecture for rainfall forecasting (Nasseri et al. 2008).

GA has also been used for the identification of appropriate parameters for ANN. Polat and Yildırım (2008) described an approach in which a GA has been used to identify the best parameters for a general regression neural network. This approach has been used for the pattern recognition in 2D and 3D images. Similarly GA has been used to generate optimal guidance training data set to train the GRNN used for a real-time dynamic optimal guidance scheme for a large missile defence space (Hossain et al. 2013). This literature review shows that GA can be used to identify the most suitable parameters for ANN, hence reducing inputs required from users.

2.4 Geoenergy

As mentioned earlier, the focus of the SDSS developed in this research remains on Geoenergy applications. The term "Geoenergy" used in this research refers to a number of technologies including Ground Source Heat (GSH), unconventional gas resources, and Carbon Capture and Storage (CCS) in deep geological formations. Unconventional gas is an umbrella term used for shale gas, Underground Coal Gasification (UCG), Coalbed Methane (CBM) and tight gas. Also, it covers the opportunity for Enhanced Coalbed Methane Recovery (ECBM) by injecting and storing CO_2 in unminable deep coal seams, thus mitigating greenhouse gas effects.

There is a considerable potential of CBM and ECBM in Wales, UK. Therefore, in this research CBM and ECBM have been considered in detail and have been used for an application of the SDSS. This application covers specific decision aspects of the site selection, site ranking and site impact assessment for CBM and ECBM development in Wales. Details and results of the application are provided in Chapter 8. A review of the CBM and ECBM technology is provided below.

2.4.1 Coalbed Methane and Enhanced Recovery

CBM is considered as a clean coal technology as the natural gas is known to be the cleanest and the most hydrogen-rich fossil fuel, whereas the coal itself is the most polluting fuel source (Economides and Wood 2009). Natural gas will remain the main source of energy during the transition period until the targets of renewable energy resources are entirely achieved (Weijermars et al. 2011). There is a huge potential globally to utilise CBM-ECBM as a Low-Carbon source of energy (IEA 2013). Countries, such as USA, Canada, Australia and China are already commercially exploring CBM (Moore 2012).

The coal seam is a net carbon sink which means that it can absorb CO_2 while enhancing the release of methane from its surface and cavities. This offers a great opportunity for not only enhancing the recovery of the coalbed methane but also to store CO_2 in deep and unminable coal seams. This process is called Enhanced Coalbed Methane (ECBM) (White et al. 2005) and it can be used to reduce the global greenhouse gas emissions in the atmosphere.

Figure 2.3 shows the process of CO_2 injection into a coal seam at the injection well, enhancing the flow of produced methane at the production well and storing CO_2 permanently in coal.

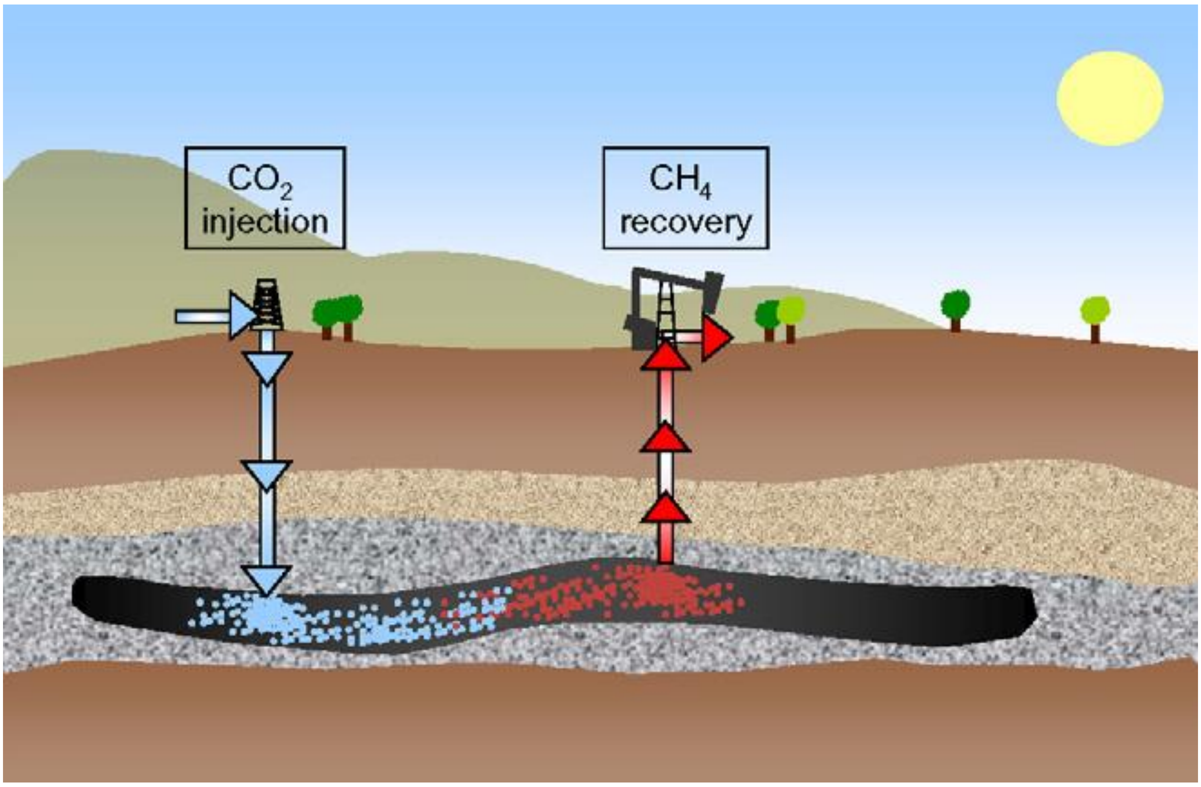


Figure 2.3 Enhanced coalbed methane recovery process (Mazzotti et al. 2009)

Jones et al. (2004) have emphasised that only unminable and ungasifiable coal seams should be targeted for CO_2 sequestration. Such areas in UK have already been identified in a report published by the Department of Trade and Industry (DTI) and the British Geological Survey (BGS) (Jones et al. 2004). GIS criteria used in this report for the identification of CBM potential zones is provided below (Jones et al. 2004):

- Coal seams greater than 0.4m in thickness and at depths between 200m and 1200m.
- Seam gas content >1m^3/tonne.
- 500 metres or more horizontal separation from underground coal workings.
- Vertical separation of 150m above and 40m below a previously worked seam.
- Vertical separation of >100m from major aquifers.
- Vertical separation of >100m from major unconformities.

A number of CBM potential zones have been identified in UK that meets above mentioned GIS criteria. Some of these potential CBM zones are situated in South and North coalfields in Wales, UK. According to this report, South Wales coalfield has the highest seam methane contents in the entire UK, i.e. within a range of 5.5 – 22 m^3/t (Jones et al. 2004). Also, the identified coal seams are at a depth between 200m and 1200m which mean that they are unlikely to be used for conventional mining or in-situ gasification of coal. This shows that Wales has a considerable potential of CBM-ECBM development. However, it is important to consider other important environmental, socio-economic, public health and techno-economic aspects into decision making process to ensure that such developments are useful for the society and safer for the environment and public health in Wales.

2.5 Considered spatial decision problems

The aim of the SDSS developed in this research is to assist the decision makers in solving a wider spectrum of spatial problems related to Geoenergy and Geoenvironmental applications, in particular those related to unconventional gas development. Spatial decision problems associated with Geoenergy and Geoenvironmental applications that are considered in this research are:

a. Site selection, suitability and ranking
b. Impact assessment
c. Spatial knowledge discovery: analysis of relationships between key environmental, public health and socio-economic indicators.

2.5.2 Site selection and ranking

Site selection, suitability and ranking problems are common to a wide range of fields, such as renewable energy, infrastructure development, health care, commercial, industrial, mining, unconventional energy, waste management, forestation and ecological services. Site selection is also one of the most common spatial problems faced in Geoenergy and Geoenvironmental applications. A review of multicriteria GIS analysis techniques used for the purpose of site selection and site ranking is provided below.

The NatCarb project has developed a GIS to facilitate potential sites identification for carbon capture and storage in US and Canada. The tool also provides online analysis features for the stakeholders (Carr et al. 2009).

An integrated GIS and analytical hierarchy fuzzy prediction method is used in Southern Qinshui basin in North China, to evaluate the CBM resources. The tool is used to identify the most suitable areas for CBM development in terms of the gas contents (Cai et al. 2011).

A GIS based site ranking using neighbourhood comparative analysis, TOPSIS and Criterion Sorting Mechanism (CSM) has been applied to rank the potential CBM sites in Wales (UK) (Irfan et al. 2014).

Damen et al. (2005) used multicriteria GIS analysis to explore an early opportunities for carbon capture and storage potential together with enhanced oil (EOL) and enhanced gas (ECBM) recovery at the global level. Potential areas where CO_2 can be captured and stored for ECBM and EOL, were identified at the global scale (Damen et al. 2005).

Chen et al. (2011a) developed a decision support system (DSS) to find the least cost pathway between carbon source and sink. This DSS utilises carbon source-sink models based on the transportation cost and complex terrain factors. The DSS has been utilised to identify the least cost pathways between the GreenGen, China's first near-zero-carbon-emission integrated gasification combined cycle (IGCC) power plant, and three neighbouring oilfields (Chen et al. 2011a).

The site selection, suitability and ranking problem is commonly faced in other renewable energy and Geoenvironmental applications. A review of such applications is provided below.

A hybrid multicriteria SDSS has been developed for the identification and prioritization of suitable regions for construction of solar power plants in Iran. This SDSS considers economic, environmental, technical, social and risk criteria in MCDA models to rank and prioritise Iranian cities for the solar projects (Vafaeipour et al. 2014).

Satkin et al. (2014) presented a multi criteria site selection model for wind-compressed air energy storage power plants in Iran (Satkin et al. 2014). Similarly, Weighted Linear Combination (WLC) in combination with fuzzy set theory has been applied for decisions on wind farm site selection in Northwest Ohio (Gorsevski et al. 2013).

A combination of Analytical Hierarchy Process (AHP) and TOPSIS method has been suggested for municipal solid waste landfill site selection in a case study of Thrace region in Greece (Demesouka et al. 2013). Similarly, site selection for temporary municipal waste storage, is carried out in in Sweden, using neighbourhood analysis performed on key demographic and metrological indicators (Ibrahim et al. 2013).

AHP, fuzzy membership functions and Simple Additive Weighting (SAW) has been used together in a study to find the most suitable site for a tourist building a in the rural landscape of Hervás in Spain (Jeong et al. 2013).

A spatial decision tool for managed-aquifer-recharge combining AHP, WLC and Ordered Weighted Averaging (OWA) has been presented in (Rahman et al. 2012). A case study of this tool has been presented for the site selection of managed-aquifer recharge in the Algarve Region of Portugal (Rahman et al. 2012).

Similarly, Site neighbourhood analysis has been adopted to identify potentially suitable sites for storm water harvesting in an urban area (Inamdar et al. 2013). In this case study a similar two-stage approach for site selection and ranking was adopted.

Literature review presented above illustrates how different spatial MCDA techniques have been used for site selection and site ranking covering a range of Geoenergy and Geoenvironmental aspects. Literature review shows that site selection and ranking are common problems and spatial multicriteria decision analysis techniques including AHP, WLC and TOPSIS have been intensively used to tackle them.

2.5.3 Impact assessment

Impact assessment is another aspect of the spatial decision making process. The Environmental Impact Assessment (EIA) is a process of assessing negative and positive impacts that a proposed project is going to have on society, environment and natural resource. The International Association for Impact Assessment (IAIA) defines the EIA as *(IAIA 2009)*:

> *"The process of identifying, predicting, evaluating and mitigating the biophysical, social and other relevant effects of development proposals prior to important decisions being taken and commitments made"*.

It is critical to identify and estimate the negative and positive impact of every significant aspect of a project on the status quo of the environment and society. The interest in impact assessment analysis has recently increased due to the global climate change concerns. Impact assessment is a qualitative and judgmental process and heavily involves decision makers input. Conducting EIA of an intervention (policy, plan, program or project) is a procedural task and the way it is carried out depends on the law of the land. In some cases it is also governed by regional and international laws (IAIA 2009).

A review of selected publications on the impact assessment of engineering interventions related to Geoenergy and Geoenvironmental applications is presented below. Although the impact assessment is designed as a qualitative process but there are methods available that can facilitate the impact assessment process in a semi-quantitative way, e.g. Rapid Impact

Assessment Matrix (RIAM) (Pastakia and Jensen 1998). RIAM is a commonly used method found in literature for the mentioned purpose. The Rapid Impact Assessment Matrix (RIAM) is a semi-quantitative way of executing the EIA in the form of a structured matrix containing the subjective judgements of the EIA assessors. The graphical form of RIAM can be useful in assessing the subjective and quantitative judgements with better clarity, as compare to other traditional methods of EIA which are more qualitative in nature (Pastakia and Jensen 1998). Using RIAM method, each aspect of the project is evaluated against the environmental components and is assigned to one of the four categories, i.e. (a) Physical/Chemical, (b) Biological/Ecological, (c) Social/Cultural and (d) Economics/Operational (Pastakia and Jensen 1998).

Koornneef et al. (2008) studied the institutional and procedural aspects of screening and scoping of environmental and strategic environmental impact assessment of CO_2 capture and storage in Netherlands. They identified three elements of the CCS for the impact assessment: i) power plant with CO_2 capture, ii) transportation of CO_2 and iii) the underground storage of the CO_2 (Koornneef et al. 2008).

Stamford and Azapagic (2014) have carried out a complete life cycle impact assessment of shale gas for electricity generation in UK. Study provides a comparison of shale gas with coal, conventional and liquefied gas, nuclear, wind and solar power for electricity generation. A number of environmental impact components were considered, including global warming, abiotic depletion, acidification, eutrophication, freshwater aquatic eco-toxicity, human toxicity, marine aquatic eco-toxicity, ozone layer depletion, photochemical ozone creation and terrestrial eco-toxicity potential (Stamford and Azapagic 2014). This study suggests that shale gas can be an environmentally sound option for electricity generation in UK if it is controlled by strict regulations.

RIAM has been applied to provide a systematic and quantitative evaluation of the socio-economic and environmental impacts of planned structural flood mitigation measures in metropolitan Manila, Philippines. The scale is determined for each perceived impact, and the results present both negative and positive impacts (Gilbuena Jr et al. 2013).

Mondal et al. (2010) have applied RIAM on four possible options of municipal solid waste management in Varanasi, India: i) open dumping, ii) sanitary landfill, iii) biomethanation and iv) gasification and incineration. RIAM results identified sanitary landfill as the best option under the existing circumstances in the study area (Mondal et al. 2010).

Similarly, RIAM has been applied to systematically evaluate and compare the four different types of potential biomass facilities in UK (Upham and Smith 2014). In this application, RIAM has been presented as a useful tool for non-specialist users for synthesising the results of different types of impact assessment (Upham and Smith 2014).

Literature presented above illustrates the vital component of an engineering intervention, i.e. the impact assessment. RIAM has been identified as an effective and quicker method to carry out impact assessment in a semi-quantitative manner. The assessed impacts of an engineering intervention at a particular location can also feedback into the decision on site selection for better public acceptance and a safer execution of the project. Also, RIAM can be a useful tool for non-specialist users to understand the potential impacts of a proposed engineering intervention.

2.5.4 Spatial knowledge discovery

Spatial knowledge discovery is about discovering hidden, unseen or unknown information from spatial information. Koperski and Han (1995) have defined the knowledge discovery from spatial data as:

> *"The extraction of interesting spatial patterns and features, general relationships that exist between spatial and non-spatial data, and other data characteristics not explicitly stored in spatial databases"*.

This section provides a review of the selected techniques that are being used for the spatial knowledge discovery. Techniques considered for this research are geovisual analytics, clustering, correlation and regression analysis.

Mennis and Guo (2009) have presented a review on the most common tasks related to spatial data-mining and knowledge discovery. The study reveals that the most common tasks in this field include spatial classification and prediction, spatial association rule mining, spatial cluster analysis and geovisualisation etc. The study also presents applications of genetic algorithm for optimisation, classification and interpolation (Mennis and Guo 2009).

Chae et al. (2014) developed a tool for spatial decision support environment that assists in the evacuation planning and disaster management and is aided with geo visual analytics of spatio-temporal data. The study also demonstrated an effective use of the tool in extracting public behaviour responses to social media before, during and after the occurrence of a natural calamity such as hurricane (Chae et al. 2014).

Henriques et al. (2012) presented a new tool called GeoSOM that facilitates the pattern finding in spatial data, based on the interaction between spatial and aspatial variables. The tool is capable of incorporating both spatial and non-spatial parameters in identifying clusters in a geographical dataset. According to the author, clustering is one of the most popular and important tasks in both spatial and non-spatial data analysis (Henriques et al. 2012).

Špatenková and Virrantaus (2013) studied the use of multiple spatio-temporal analysis methods to explore the causal relations in the building fire incident data from the city of Helsinki, Finland. The study uses both visual and computational methods for the purpose, such as Geographically Weighted Regression (GWR), bivariate analysis and Parallel Coordinate Plots (PCP) (Špatenková and Virrantaus 2013). PCP can be used for both spatial and spatio-temporal data visualisation to facilitate knowledge extraction and understanding (Edsall 2003).

Literature review presented above has identified some useful techniques that can be used for spatial knowledge discovery. These techniques include i) PCP, ii) SOM based clustering, iii) SOM based correlation analysis and iv) GRNN based regression analysis. Including these techniques in SDSS can also facilitate the spatial decision making process by providing an insight of any relationships that may exist between the key indicators.

2.6 SDSS in Geoenvironmental applications

This section provides a review of the existing non-commercial SDSS designed and developed to tackle the Geoenergy and Geoenvironmental spatial decision problems. Several, commercial and free GIS software such as IDRISI, ArcGIS, SAGA and ILWIS provide a number of modelling techniques and a mechanism to customise and bundle them together to serve as an SDSS (Sugumaran and Degroote 2010).

Commonly used decision support systems in the Geoenvironment field are: i) Spatial Analysis and Decision Assistance (SADA) (Stewart and Purucker 2011; SADA 2014) and ii) DEcision Support sYstem for the REqualification of contaminated sites (DESYRE) (Carlon et al. 2007). SADA provides a localised and site specific tools for human health and ecological risk assessment, cost-benefit analysis, area of concern analysis and geospatial visualisation (Stewart and Purucker 2011; SADA 2014).

DESYRE is a GIS-based Decision Support System (DSS) that is designed for the integrated management and remediation of contaminated sites (Carlon et al. 2007). DESYRE covers the mains aspects of a site remediation process such as the site characterization, risk assessment, analysis of social and economic benefits and constrains, remedial technology selection and residual risk analysis (Carlon et al. 2007).

Notably, both SADA and DESYRE, are site specific decision support tools and have limitations in tackling on regional and national level spatial decision problems, including site selection and ranking.

Other SDSSs have been developed for the non-local spatial decision analysis such as the facility siting, natural resource management and environmental monitoring and management. A review of the recently developed SDSS in the above mentioned domains is provided below:

Fayetteville shale gas SDSS has been developed to analyse and assess the impacts of water consumption for hydraulic fracturing (Cothren et al. 2013). The system is used by the regulatory agencies and producers, to study the potential impacts on the environmental flow components (EFCs) of the river.

Zuo et al. (2013) developed an SDSS to assist mineral planning practice in England and Wales. The main objective of this SDDS is to reduce CO_2 emissions of the supply chain, using alternate policies. This system utilises a spatial interaction model, a micro simulation model and the data on production and transportation of aggregates across the country. Using this system, new policy scenarios have been discussed for the cost-effectiveness and CO_2 reduction in the supply chain of the aggregates markets (Zuo et al. 2013).

Ruiz et al. (2012) have presented the design and construction of a multicriteria SDSS for the identification of sustainable industrial areas incorporating socio-economic, physical–environmental, infrastructures and urban development factors. The SDSS uses fuzzy logic and weighted score for the construction of the multicriteria decision model. This tool has been applied in Cantabria region, Spain for the identification of suitable areas for sustainable industrial areas (Ruiz et al. 2012).

Perpiña et al. (2013) presented a multicriteria GIS assessment process for the identification of suitable sites for the construction of biomass plants. A combination of Analytical Hierarchy Process (AHP), Weighted Linear Combination (WLC) and Ideal Point Method (IPM) is used to incorporate environmental, economic and social consideration in the site selection process. Sensitivity analysis has also been carried out to obtain the influential factors of the model (Perpiña et al. 2013).

Gorsevski et al. (2013) introduced a prototype SDSS to facilitate the group decision making for wind farms site suitability in Northwest Ohio. The framework integrates environmental and economic criteria in the analysis using fuzzy set theory, Borda count and WLC methods. The criterion maps created by participants are aggregated to produce a group solution using Borda count method. Sensitivity analysis has also been performed to check the sensitivity of the model against the weights assigned to different criterion (Gorsevski et al. 2013).

Zanuttigh et al. (2014) developed an SDSS for the management of coastal risks including assessment of erosion, flood risk, socio-economic and ecological vulnerability. This system allows the user to set up multiple scenarios by assigning different weights within the multi criteria risk analysis and then to compare different options (Zanuttigh et al. 2014).

Comino et al. (2014) have developed multicriteria SDSS for the assessment of environmental quality of the Pellice river basin in Italy. The model has been developed in IDRISI and has the capacity to identify the environmental quality of the study area in terms of “naturalness” and "pressures". An economic evaluation of the ecosystem services has been performed using the system. This evaluation compares the percentage of area covered under key landuse classes in comparison with the two environmental quality classes, i.e. "naturalness" and "pressures" (Comino et al. 2014).

2.7 SDSS Considerations for Geoenergy developments

Due to the increasing global awareness of greenhouse gas effects on the environment, society and public health, it is important for the decision makers to consider the entire life cycle assessment of unconventional gas technologies. As explained earlier in the Introduction to this chapter, there is a huge global potential for the exploitation of unconventional gas resources. A number of commercial and demonstration projects are at different stages of their life cycle. There are also a number of research publications highlighting the environmental, socio-economic, public health and techno-economic aspects of the unconventional gas developments.

It is important for the effective utilisation of these recourses that all the associated key aspects are considered for an informed risk based spatial decision making process. This can be helpful in raising the public acceptance of unconventional gas developments, also can ensure their sustainability. A comprehensive review is provided in the following sections to identify significant environmental, socio-economic, public health and techno-economic aspects that will be considered in SDSS design and development.

2.8 Socio-Economic aspects

This section covers the socio-economic aspects related to the unconventional gas developments. Systematic review of available literature revealed that there are more publications highlighting the socio-economic aspects of the shale gas development and CCS than other Geoenergy technologies. To enhance the understanding, studies from other relevant fields can also be used including mining, petrochemical, geoengineering and renewable energy sector.

Common socio-economic issues related to the unconventional gas development considered in this research include social acceptance, social capital and socio-economic uplifting through employment generation and business activities. Following sections cover review of these socio-economic aspects. Indicators that can be used to incorporate these important aspects in multicriteria spatial decision analysis are also presented.

2.8.1 Social acceptance

Social acceptance of Geoenergy applications is crucial for its development. A number of socio-economic and political factors can influence the level of social acceptance for the new technologies. A review of selected publications is provided below to illustrate the influence of social acceptance on the development of new technologies. Also, to identify key factors that should be incorporated in the spatial decision making process.

Social acceptance has affected the desired growth of new technologies. Wüstenhagen et al. (2007) conducted a research to show how social acceptance is delaying the renewable energy targets in many countries. A number of reasons have been identified in the literature that can affect the social acceptance for new technologies to provide alternate source of energy and help mitigate climate change effects. For example, Pidgeon et al. (2012) have presented the early findings of public response in UK, about geoengineering interventions to control climate change effects. One of the findings reveals public acceptance for such engineering interventions increases with the awareness about the consequences of climate change (Pidgeon et al. 2012).

Similarly, Wolsink (2000) studied the NIMBY (Not in My Back Yard) syndrome with respect to the public acceptance of wind energy. Wolsink (2000) identified major reasons behind the low public acceptance including noise pollution, annoyance, spoiled scenery, interference with natural areas and unreliability of the energy supply. This shows that if the sites are selected carefully, incorporating all these key aspects into spatial decision making, it can address the public concerns and help increasing the level of social acceptance for new technologies.

Shackley et al. (2009) have presented the outcomes of the ACCEPT project, researching acceptability of CO2 Capture and Storage in Europe. The project investigated stakeholders and their opinions about the role of CCS in Europe’s energy future. Participants selected from European countries, found to be moderately supportive of CCS deployment in their own country. Participants had shown higher trust and acceptability towards European level projects. Common concerns identified, were related to the environmental risks and the divergence of investments from renewable energy resource development towards CCS. This study shows that if environmental risks are controlled and incorporated into the spatial decision making process, a higher level of social acceptance can be achieved.

Shackley et al. (2006) have also presented a case study on the UCG project lead by the Coal Authority in Silverdale Colliery, Staffordshire, UK analysing the social and political issues underpinning the acceptance of UCG projects. The Silverdale project was withdrawn by the Coal Authority because of the public outcry and legal challenge. Concerns raised by the public were mostly related to noise pollution, visual impact, uncontrollable burning of coal in the seam, aquifer contamination, underground explosions and the experimental nature of the work (Shackley et al. 2006).

Ha-Duong et al. (2011) have presented the social aspects of the Total's Lacq plant, CO_2 integrated capture, transport and storage pilot project in south western France. The significance of social issues related to this particular project was really high as there is a large population that lives close to the storage site (a 4,500 meters deep depleted gas field). Also this project is the first of its kind to be carried out in France. The social conditions were favourable as the operator had already led the economic activities in this area for two generations and had proven that they could tackle higher risks. The project offered jobs and economic activities which contributed to the future of the area as previous operations due to depleted reserves went down (Ha-Duong et al. 2011). This study shows that there is a higher social acceptance if a project is being setup in an area where economic activities have been driven successfully by similar projects in past or by the same operator. Also, higher unemployment rate and slowing down economic activities in the area can result in higher social acceptance of unconventional gas developments, if it can offer new jobs and create business activity.

Huijts et al. (2007) have presented the results of a study on CCS acceptability conducted in Netherlands in 2003. Interviews were conducted with the stakeholders to obtain multi-facet viewpoints about CCS. The survey reveals that general public have slightly higher acceptability for storing CO_2 outside the urban areas, while opinion regarding storage near to the populated areas was found to be slightly negative (Huijts et al. 2007). This study shows that distance of unconventional technology sites from populated area plays an important role in shaping the level of the acceptance.

Bradbury et al. (2009) have presented a public outreach research study that was conducted in five different communities, living in the potential CCS areas in USA. Outcomes of the study shows that factors such as past experience with government, existing low socio-economic status, desire for compensation and perceived benefits to the community were of greater concern than the concern about the risks of the CCS technology itself (Bradbury et al. 2009). This shows that if a project has to offer economic benefits to a community in terms of job creation, business generation and royalties, the level of acceptance is usually higher.

Literature review presented above covers a number of social studies from different parts of the world, including UK. Key factors influencing the social acceptance of unconventional gas and renewable energy have been identified. This review also identifies certain characteristics of a community that can influence the level of social acceptance for considered technologies. The indicators identified for the purpose include awareness about climate change and greenhouse gases, proximity to populated places, the level of trust to the operator, existing socio-economic status of the communities, economic benefits and the existing exposure to similar economic activities in the areas including mining and natural gas. If these indicators can be measured using the relevant datasets, then a general level of social acceptance can be mapped across the study area. Areas with potentially higher social acceptance can be given more importance in the multicriteria spatial decision analysis for site selection of unconventional gas developments.

2.8.2 Social capital

Social capital can influence the level of acceptance, cooperation and involvement of a community in a positive or a negative way. According to World Bank’s research, Social Capital is a still evolving concept and instead of a narrow definition and it can be defined broadly as (Grootaert and Bastelaer 2002):

> “Institutions, relationships, attitudes and values that govern interactions among people and contribute to economic and social development”.

The UK government has adopted the OECD’s (Organisation for economic Co-operation and Development, Paris) definition of social capital as (Foxton and Jones 2011):

> "Networks together with shared norms, values and understandings that facilitate co-operation within or among groups".

Social capital is the bonding, bridging and linking between different groups such as geographical groups, professional groups, social groups and virtual groups (Foxton and Jones 2011). Social capital has an objective and a subjective part. The objective part refers to the observable and tangible networks, associations and institutions in a society. Whereas, the non-tangible part is the mutual trust, reciprocity and generally accepted attitudes, norms and behaviours (Grootaert and Bastelaer 2002).

Social capital is an important factor that can influence all stages of a project. Even if a technology is socially accepted, community support and involvement in the project throughout its lifecycle is crucial to ensure maximum socio-economic benefit of the project for them. A selected literature review is provided in this section to show how social capital can influence the acceptance and development of new technologies and what factors are important to be incorporated in the spatial decision making process.

Anderson et al. (2012) conducted a qualitative research and assessment of human and social capital in Otway community in Australia. Study aimed to determine characteristics of community where the acceptance of the CCS is achieved with relative ease. Also, to identify the best practices for public participation. Results of study reveal that Otway community accepted the technology, however at the later stages of the project execution, problems linked with disturbances in day-to-day farming activities were faced. The study reveals that one part of community had higher social capital and education level. The other side of community had a low social capital and their capacity to raise their voice was limited. These were mostly dairy farmers having low literacy rate and small social circle. Because of this low social capital they couldn’t raise their voice against the problems faced by them in their day-to-day farming activities. They reported that they were unprepared for such level of interference and couldn’t benefit from the project as a community (Anderson et al. 2012). This study shows that acceptance is not consistent throughout the project lifecycle. Also, that a community with lower level of human and social capital can be more vulnerable for exploitation by the proponents that could consequently have negative impacts on level of acceptance for future projects.

Literature review presented above shows that a higher level of social capital of communities can lead to better execution of projects and increase the potential of project benefits to the community. With higher level of social capital and effective communication strategies, communities can be involved at all stages of the project. Therefore, project related socio-economic or environmental challenges faced by the community can be raised and addressed in timely manner. This can also help in shaping a higher level of social acceptance for future projects.

There are a number of ways to measure social capital in communities. Foxton and Jones (2011) have devised a framework for the measurement of social capital in UK which will be used in this research as explained in Chapter 7. Indicators identified for the measurement of a general level of social capital in communities include civic participation, social participation, views about the local area, reciprocity and trust, crime rate and education level of the community (Foxton and Jones 2011).

2.8.3 Employment generation and socio-economic uplifting

This section present review of examples highlighting the role of the Geoenergy and renewable energy developments in changing local, regional and national economies. Also, it presents adequate indicators that can be incorporated in the multicriteria decision framework for unconventional gas development targeting local and regional growth.

Roddy and Younger (2010) presented the reasons behind the failure of some UCG and CCS projects in UK and Europe. This failure is mostly caused by public outcry on environmental and socio-economic concerns related to these technologies. For the success of the these new technologies, the agenda has to integrate wider local development initiatives aimed at creating new employment opportunities, business activities, improving quality of life and be beneficial to the environment (Roddy and Younger 2010). Similarly, the sustainable development agenda of the Clean Development Mechanism (CDM) stresses the importance of environmental and socio-economic benefit from such developments (Olsen and Fenhann 2008).

Perry (2012) presented a study analysing how a rural agro-based economy has gone through a rapid socio-economic transition caused by shale gas industry in rural north eastern Pennsylvania. Most of the community seemed to be in the favour of the gas exploration, as it has created employment and business opportunities for the locals. Initial critics of shale gas developments had shown positive attitudes towards it, based on its economic benefits to the local community. On the other hand some landlords, who were initially in the favour of shale gas developments, went against it after witnessing the environmental degradation on their lands (Perry 2012).

The literature review suggests that the Geoenergy developments can only be widely accepted and sustainable if they can contribute to the local economy by creating new jobs, infrastructure development and increasing local and regional business activities. Therefore, indictors that can reflect the current socio-economic status of the local communities should be included in the multicriteria decision framework for Geoenergy. These indicators include employment-unemployment rates, employment type by industry sectors, living standard, access to services and multiple deprivation. Those areas can be given higher importance where socio-economic conditions are relatively poor as compare to other areas. In this way Geoenergy developments, through job creation and business generation, can contribute to the uplifting of the local socio-economic conditions. Exploiting these indigenous energy resources can also help in reducing fuel poverty. Also, incorporation of the socio-economic parameters can increase the level of acceptance of such developments in future.

2.9 Environmental aspects

This section covers the known environmental hazards associated with the application of unconventional gas technologies. Environmental issues have to be incorporated into the spatial decision making for the site selection and impact assessment of the Geoenergy applications. This not only helps with an environmentally safe execution of the project but also creates favourable conditions for the social acceptance and future initiation of similar projects.

The environmental laws and regulations govern the safe execution of any critical project. However, the legal aspects are beyond the scope of this research. Therefore, environmental aspects incorporated in the SDSS are entirely based on the literature review. These environmental aspects are categorised as related to i) water quality, ii) air quality, iii) natural resource and landuse change. A review of selected publications is provided in the sections below to identify the key environmental factors and indicators associated with the Geoenergy developments.

2.9.1 Water quality

This section covers the environmental aspects of the Geoenergy developments related to the water quality. According to the literature review provided below, most common water related issues include presence of toxics in the produced water and the risk of aquifer and surface water contamination. There is also concern on the quantity of water used in the process that could lead to water shortages and would affect local ecosystems.

Pashin (2007) studied the hydrodynamics and environmental issues related to the coalbed methane reservoirs in the Black Warrior Basin (USA). It is reaffirmed that the biggest environmental concerns associated with CBM is the risk of contamination of shallow aquifers and surface water. Also, huge volumes of water required during the early stages of the project may result in the drawdown of water table, effecting domestic water supply (Pashin 2007).

Similarly, Orem et al. (2014) characterised the organic substances found in the produced and formation water generated from five CBM and two shale gas plays in USA. The quality of produced water is determined by the added chemicals (fracturing fluids) and the chemical composition of the coal or shale formations. A large quantity of different organic chemicals has been identified in the samples of produced and formation water. These chemical include polycyclic aromatic hydrocarbons, heterocyclic compounds, alkyl phenols, aromatic amines, alkyl aromatics, long-chain fatty acids, and aliphatic hydrocarbons in both CBM and shale gas regions. Whereas, some additional solvents, biocides and scale inhibitors are also found only in shale regions (Orem et al. 2014). This study shows that the produced water in both cases may contain hazardous chemicals and a proper treatment is required before disposal. Apart from strict regulation, the risk can also be reduced if the sites are placed strategically, i.e. away from water bodies and rivers and any major aquifers to avoid any accidental release.

Amount of water used in process of the exploitation of unconventional gas is another major challenge. Uliasz-Misiak et al. (2014) presented key environmental and legal issues related to the exploitation of unconventional gas (shale and tight gas) in Poland. Study reiterates that the amount of water required for the fracturing process may stress the water resources in the region. There is a limited possibility of geological storage of produced water therefore a huge amount of produced water will need to be treated and discharged. This can pose a high risk of contamination to the Polish rivers (Uliasz-Misiak et al. 2014). The study also emphasised that in order to reduce these environmental risks, the sites should be selected at a safe distance from environmental protection areas (Uliasz-Misiak et al. 2014).

Literature review presented above stresses the importance of strict regulatory regime for the effective utilisation of Geoenergy resources. However, at the same time, these environmental risks can be minimised by selecting the site location carefully. If sites are developed at a safer distance from environmental protected areas, important aquifers, water bodies and rivers, the risk of ground and surface water contamination can be reduced.

2.9.2 Air quality

This section covers a review on the environmental aspects of the unconventional gas developments related to the air quality. Existing studies show that some of the considered Geoenergy applications can potentially help to reduce the GHG emissions. Natural gas produced from these resources can be utilised as interim sustainable energy source, until the targets for the renewable energy based economy are achieved (Weijermars et al. 2011).

Some of the Geoenergy applications have positive impacts on the environment. For example, Jenner and Lamadrid (2013) examined the environmental impacts of coal, shale and conventional gas on air, water, land, also the overall effect on the quality of life in US. The results reveal that as compare to coal, shale gas is going to benefit local natural environment, it has smaller Green House Gas (GHG) footprint, better safety of workers, less water consumption and lower public health impacts.

Yu et al. (2007) studied the primary factors affecting the CBM and ECBM potential in China. The study suggests that the ECBM potential in China is so huge that it can store CO_2 produced in 50 years based on China's CO_2 emissions levels in year 2000 (Yu et al. 2007). Similarly, Imran et al. (2014) have presented UCG as a low carbon, environment friendly and economically feasible option for the utilization of deep and unminable coal resources (Imran et al. 2014). The authors suggest that utilizing UCG for power generation will reduce the emission of air contaminants as compared to the conventional coal fired power plants (Imran et al. 2014).

However, there are some issues associated with air quality which should be incorporated in the multicriteria decision analysis. These issues include the emissions from the site machinery and transportation involved in the unconventional gas development. Other issues are related the risk of fugitive methane and CO_2 release into the atmosphere caused by any natural disaster or structure failure. These issues are briefly discussed below.

For example, the impact of shale gas on ozone is about the same as that of the natural gas but it poses a higher risk of creation of photochemical oxidants (smog) as compare to coal or natural gas (Stamford and Azapagic 2014).

Kemball-Cook et al. (2010) estimated the production of shale gas and the associated effects on ground level Ozone, in Haynesville shale in USA. Photochemical modelling of the year 2012 showed an estimated increase in Ozone (8 hour) values of up to 5ppb the region from the developments in the Haynesville shale gas extraction (Kemball-Cook et al. 2010).

To conclude from the review presented above, it is important to incorporate the existing levels of air quality into multicriteria decision analysis. Areas with relatively higher emission rates or concentration levels of such chemicals should be avoided for site development. It is also suggested that a systematic monitoring of site should be conducted before and during the project execution. This will help to avoid any controversy related to the adverse effects of the project on the environment. Similar controversy has been discussed by Force and Graham (2011), cited in (Soeder et al. 2014). It was argued that the background concentration of Ozone and Particulate Matter (PM2.5) were already exceeding the national emission limits in the Marcellus Shale development region and there was little contribution from the shale gas developments. Therefore, the current level of emissions for the key substances should be included in the decision making. This also conforms to the environmental justice practices, i.e. to avoid environmental disease burden on population already living under stressed environmental conditions (Fisher et al. 2006; Maantay 2007).

2.9.3 Natural resources and landuse change aspects

This section covers a review of the selected publications identifying the impacts of unconventional gas developments on natural resources, ecosystem and landuse change.

Soeder et al. (2014) have discussed the risks of landscape deterioration associated with the shale gas related engineering interventions, i.e. pad construction, drilling, fracturing, access roads and pipelines. Study suggests that multiple horizontal wells drilled from the same pad can reduce the footprint on the landscape (Soeder et al. 2014). However, there are some associated risks including change of land use, slopes, hydrology and soil compaction (Soeder et al. 2014). The stress on ground water recharge and change in surface flow can affect the ecosystem and biodiversity of the area (Soeder et al. 2014).

Wei et al. (2011) have analysed the potential effects on soil caused by an accidental CO_2 leakage either i) during the transportation or ii) from the CCS reservoir through natural fractures, faults or abandoned wells. Injected CO_2 may also contain impurities that can be more toxic than the CO_2 itself (Wei et al. 2011).

Meng (2014) modelled the fracking pad sites with landscape variables in Marcellus shale region in USA. The results reveal that the gas fracking pad sites are not randomly placed in the landscape. Higher elevated areas and landcover class of "Wetland" is very likely to be intruded by the pads (Meng 2014). Similarly, while studying the environmental risks associated with unconventional gas development in Poland, Uliasz-Misiak et al. (2014) discussed the importance of incorporating environmental protection areas in the decisions to ensure an environmentally safe exploitation of the shale gas resource.

The selected review provided in this section identifies the potential impacts of the Geoenergy developments on natural resources, key landcover and landscape classes. It is important that the landcover and landscape variables should be incorporated into the multicriteria decision making process of site selection and impact assessment. This can help to avoid any critical landscape being selected for the site development. The key landscapes, landcover classes, scenic areas, biodiversity hotspots and key habitats should be protected from the development of Geoenergy resources. This also includes the traffic, noise and night time light pollution associated with the site construction and its operations.

2.10 Public Health aspects

This section covers a selected review on the impacts of Geoenergy developments on public health. As discussed in Section 2.9, these technologies have less harmful effects on the environment as compare to some of the conventional resources such as burning coal. Also, CCS and ECBM technologies are going to help reducing emissions of GHG in the atmosphere and CO_2 storage. However, some of the Geoenergy technologies are relatively new and therefore there is a lack of substantial epidemiological research on the population living in the proximity of such developments.

Limited information is available in the literature about environmental and public health related aspects of the considered technologies. For example, Bunch et al. (2014) analysed a comprehensive data of more than 4.6 million measurements of 105 Volatile Organic Compounds (VOC) from the air samples of Barnett shale regions in USA. The study aimed at analysing any potential acute and chronic health effects by comparing sample measurements against the national guidelines. The results reveal that the measurements of VOC associated with the shale gas were all below these guideline values. Annual average concentrations were also found to be below the levels of health concern. Results have concluded that extensive shale gas developments in Barnett region have not resulted in community-wide exposures and have not raised any health concern (Bunch et al. 2014). Similarly, Jenner and Lamadrid (2013) studied the environmental and public health risk associated with unconventional gas developments. Comparing various lifecycle assessments, shale gas is considered potentially safer for local natural environment, public health and workers safety when compared to the coal (Jenner and Lamadrid 2013).

Despite the absence of concrete evidence on public health impacts associated with Geoenergy developments, public health is an important factor to be considered in the spatial decision making. It is a proven fact that environment, socio-economic conditions and personal life styles can affect the public health. The engineering risks to environment and ecosystem can rightly be considered as risks to the public health. For example, air pollution is a major cause of environmental related deaths in many countries. According to the World Health Organization (WHO) estimates, 7 million premature deaths resulted from air pollution in 2012 (UNEP 2014).

Finkel and Hays (2013) also emphasised the need for quantified and evidence-based epidemiological research to assess short and long term exposure-related health effects on the population living in close proximity to the unconventional drilling sites. Finkel and Hays (2013) have also discussed the reports on several health complications attributed to natural gas operations, including respiratory, dermatological and gastrointestinal problems.

McKenzie et al. (2012) have estimated the health risks associated with the engineering interventions to the residents living within and those living outside 1/2 miles radius of the unconventional natural gas developments in Garfield county, Colorado, USA. Environmental Protection Agency (EPA) guidelines have been followed in the study to estimate chronic and sub-chronic non cancer hazard indices (McKenzie et al. 2012). The results reveal that the estimated health risks are higher for the residents living in close proximity to the engineering interventions, especially the health risk related to Benzene. The cumulative cancer risks were found to be 10 and 6 per million for the residents living within the proximity of 1/2 mile and those living more than 1/2 miles from the wells (McKenzie et al. 2012). This study suggests that appropriate distance from populated places should be a key consideration in the multicriteria decision analysis for Geoenergy developments.

Literature review presented above suggests that Geoenergy developments have less harmful effects on human health, compared to some of the other fuels such as coal. At the same time, these technologies are relatively new and evidence based information about the public health aspects is limited. It is therefore important to incorporate public health aspects into informed risk based multicriteria decision making process. If sites are selected at a distance from the populated places, and away from critical natural resources such as rivers, it is more likely that the public health effects can be minimised. Also, those areas where current public health status is relatively better can be given higher preference in the multicriteria decision analysis. Environmental justice also demands to avoid putting environmental disease burden on population, already living under stressed environmental conditions (Fisher et al. 2006; Maantay 2007). Indicators of public health status including the mortality rate, morbidity rate and hospital admission rate can be used for this purpose. Health surveys can also be used for the assessment of the spatial variation of general public health (Meng et al. 2010).

2.11 Techno-Economic aspect

This section covers the literature review of the Techno-Economic aspects of the Geoenergy resource development. Apart from having relatively less harmful effects on environment and public health, the economic benefits are one of the main driving forces behind the exploitation of Geoenergy resources. Technical and economic aspects considered in this research include the geological constraints, resource estimation, terrain parameters (slope, elevation, and aspect), distance from existing infrastructure (electricity and gas grid) and proximity of CO_2 producers. Cost-benefit, economic risk analysis and other site specific technical and economic parameters are beyond the scope of this research. Literature review presented below highlights the economic benefits of the considered resource, also identifies the key techno-economic aspects to be incorporated in the spatial decision making to ensure an effective and sustainable utilisation of the resource.

Yu et al. (2007) studied the primary factors affecting the CBM and ECBM potential in China. The study suggests that the CBM resources have the potential for providing gas for 218 years at a production rate equal to the China's gas production rate in year 2002.

Jones et al. (2004) carried out GIS based estimation of the potential CBM resource in UK. Different geotechnical parameters were used to estimate the potential resource, such as coal thickness, coal density, coal seam area and its average gas content (Jones et al. 2004). GIS data produced for this report has been acquired from British Geological Survey (BGS) for this research as explained in details in Chapter 7.

Zhou et al. (2013) have analysed geoengineering and economic aspects of CBM and ECBM applications in South Shizhuang CBM field in China. The parameters used in the economic analysis of CBM development include costs involved in gas collection points, gas stations, gas pipelines, road construction, well drilling, rig equipment, gas and water processing facility, fixed and variable costs involved with the day-to-day operations. For ECBM, additional indicators are involved in the analysis such as CO_2/N_2 capture, processing, compression and transportation to the site, separation cost after breakthrough, safety, monitoring and verification costs (Zhou et al. 2013).

van den Broek et al. (2010) have developed a GIS based linear optimization energy model to design a cost effective CO_2 storage infrastructure. This tool can support policy makers in the effective infrastructure development for CO_2 storage. Core features of the model include the identification of CO_2 source and sink cluster. Also, cost effective routing of CO_2

transportation pipelines van den Broek et al. (2010). Similarly, Chen et al. (2010) developed a GIS based multi-criteria decision model for a systematic analysis of CO_2 source sink matching, based on the least-cost pathway. The model considers the cost of CO_2 transportation based on landform and landuse between the source and sink (Chen et al. 2010).

Hsu et al. (2012) have suggested the utilisation of a general form of AHP called Analytical Network Process (ANP) approach to tackle the multicriteria decision problem to select the most suitable sites for the geologic storage of CO_2. The study also identifies useful techno-economic parameters to be incorporated in the ANP based site selection process such as reservoir area, reservoir thickness, porosity, depth, storage capacity, caprock permeability, caprock thickness and cost (Hsu et al. 2012).

Literature review presented in this section identifies the key techno-economic aspects considered for the unconventional gas development. Resource estimation is the main techno-economic aspect, followed by the site developmental cost based on the landuse and landform parameters. Also, the proximity to the existing infrastructure is important for development and operational cost including, roads, railways, existing gas and electricity network.

2.12 Conclusions

The focus of this research is the design and development of an integrated multicriteria SDSS to support decision makers in solving Geoenergy and Geoenvironmental spatial decision problems. A comprehensive literature review is provided in this chapter to cover different aspects of the proposed SDSS. Section 2.2 covers the structure, characteristics and design aspects of an SDSS. Section 2.3 identifies effective modelling techniques that have been used in an SDSS for spatial problem solving. The analytical modelling techniques discussed in this research are categorised into i) Spatial Multi Criteria Decision Analysis (S-MCDA) based techniques and ii) Artificial Intelligence (AI) based techniques. Section 2.4 covers the

Geoenergy applications considered in this research. Particularly, CBM and ECBM technologies have been discussed in detail due to the fact that there is a known CBM-ECBM potential in the study area, i.e. Wales in UK. An application of the SDSS for CBM-ECBM development in Wales is provided in Chapter 7. Section 2.5 covers the most commonly faced Geoenergy and Geoenvironmental spatial decision problems. These problems are categorised in three themes: a) site selection and ranking b) impact assessment c) spatial knowledge discovery. Section 2.8-2.11 covers the key socio-economic, environmental, public health and techno-economic aspects that should be incorporated in multicriteria spatial decision support system to ensure an effective, safer and sustainable utilisation of the considered Geoenergy resource.

For the design and development of the proposed SDSS, it is envisaged to utilise a range of S-MCDA and AI based analytical techniques. Artificial intelligence techniques, i.e. ANN can be complex in structure and difficult to be used by non-specialised decision makers. However, semi-supervised or un-supervised ANN can be useful for this purpose including SOM and GRNN. GA can also be utilised to find the most appropriate set of essential parameters to run ANN based analysis.

For site selection process, techniques such as AHP, WLC, Pairwise Comparison Method and Sensitivity Analysis are considered. For site ranking process, site neighbourhood analysis using TOPSIS and CSM can be utilised. Also the One-Dimensional SOM can be used to reduce dimensions, cluster and rank the sites based on key indicators. For site impact assessment, RIAM can be utilised as an effective and rapidly developed semi-quantitative approach. GRNN can also be utilised for the impact assessment and for regression analysis. PCP, GRNN and SOM based clean correlation finding techniques can be utilised for spatial knowledge discovery process.

Although SDSS are domain and problem specific, in this research an effort has been made to design and develop an integrated system to support the decision makers in confronting a set of most commonly faced spatial decision problem. These spatial decision problems are common to the Geoenergy developments and Geoenvironmental applications such as solid waste disposal and artificial aquifer recharge.

The literature review presented in this chapter, highlights the importance of an informed risk based multicriteria spatial decision support system to effectively utilise Geoenergy resources. It is emphasised that along with the techno-economic feasibility, sites should be selected where the environmental, socio-economic and public health conditions are also suitable and favourable:

i. Site development and operations should help in the socio-economic uplifting of the surrounding communities. This includes creation of new jobs, business activities, infrastructure development and an improvement in the existing facilities. Public acceptance and participation should be addressed to ensure maximum socio-economic benefits of the developments to the society.

ii. The surrounding areas of the proposed sites should not be under environmental stress already. The sites should be at a safe distance from sensitive environmental areas and key natural resources. The sensitive environmental areas should cover important environmental, natural resource and ecological aspects including landuse, landscape, surface and sub-surface hydrology, forestry, biodiversity and key habitats.

iii. Public health conditions in the surrounding areas should not be under stress already as the site developments and operations will result in some level of increase in pollution. This to ensure environmental justice in the surrounding areas.

iv. Sites should be feasible in terms of geotechnical and economic parameters to ensure a cost effective and sustainable regime for unconventional gas developments.

References

Al-Adamat, R.,Diabat, A. and Shatnawi, G. 2010. Combining GIS with multicriteria decision making for siting water harvesting ponds in Northern Jordan. *Journal of Arid Environments* 74(11), pp. 1471-1477.

Anderson, C.,Schirmer, J. and Abjorensen, N. 2012. Exploring CCS community acceptance and public participation from a human and social capital perspective. *Mitigation and Adaptation Strategies for Global Change* 17(6), pp. 687-706.

Bradbury, J.,Ray, I.,Peterson, T.,Wade, S.,Wong-Parodi, G. and Feldpausch, A. 2009. The Role of Social Factors in Shaping Public Perceptions of CCS: Results of Multi-State Focus Group Interviews in the U.S. *Energy Procedia* 1(1), pp. 4665-4672.

Bunch, A. G.,Perry, C. S.,Abraham, L.,Wikoff, D. S.,Tachovsky, J. A.,Hixon, J. G.,Urban, J. D.,Harris, M. A. and Haws, L. C. 2014. Evaluation of impact of shale gas operations in the Barnett Shale region on volatile organic compounds in air and potential human health risks. *Science of The Total Environment* 468–469(0), pp. 832-842.

Cai, Y.,Liu, D.,Yao, Y.,Li, J. and Qiu, Y. 2011. Geological controls on prediction of coalbed methane of No. 3 coal seam in Southern Qinshui Basin, North China. *International Journal of Coal Geology* 88(2–3), pp. 101-112.

Carlei, V. and Nuccio, M. 2014. Mapping industrial patterns in spatial agglomeration: A SOM approach to Italian industrial districts. *Pattern Recognition Letters* 40(0), pp. 1-10.

Carlon, C.,Critto, A.,Ramieri, E. and Marcomini, A. 2007. DESYRE: DEcision Support sYstem for the REhabilitation of contaminated megasites. *Integr Environ Assess Manag* 3(2), pp. 211-222.

Carr, T. R.,Iqbal, A.,Callaghan, N.,Dana Adkins, H.,Look, K.,Saving, S. and Nelson, K. 2009. A national look at carbon capture and storage—National carbon sequestration database and geographical information system (NatCarb). *Energy Procedia* 1(1), pp. 2841-2847.

Chae, J.,Thom, D.,Jang, Y.,Kim, S.,Ertl, T. and Ebert, D. S. 2014. Public behavior response analysis in disaster events utilizing visual analytics of microblog data. *Computers & Graphics* 38(0), pp. 51-60.

Chen, K.,Blong, R. and Jacobson, C. 2001. MCE-RISK: integrating multicriteria evaluation and GIS for risk decision-making in natural hazards. *Environmental Modelling & Software* 16(4), pp. 387-397.

Chen, W.,Huang, L.,Xiang, X.,Chen, J. and Sun, L. 2011a. GIS based CCS source-sink matching models and decision support system. *Energy Procedia* 4(0), pp. 5999-6006.

Chen, W.,Le Nindre, Y.-M.,Xu, R.,Allier, D.,Teng, F.,Domptail, K.,Xiang, X.,Guillon, L.,Chen, J.,Huang, L. and Zeng, R. 2010. CCS scenarios optimization by spatial multi-criteria analysis: Application to multiple source sink matching in Hebei province. *International Journal of Greenhouse Gas Control* 4(2), pp. 341-350.

Chen, Y.,Li, K. W. and Liu, S.-f. 2011b. An OWA-TOPSIS method for multiple criteria decision analysis. *Expert Systems with Applications* 38(5), pp. 5205-5211.

Comino, E.,Bottero, M.,Pomarico, S. and Rosso, M. 2014. Exploring the environmental value of ecosystem services for a river basin through a spatial multicriteria analysis. *Land Use Policy* 36(0), pp. 381-395.

Cothren, J.,Thoma, G.,DiLuzio, M. and Limp, F. 2013. *Integration of Water Resource Models with Fayetteville Shale Decision Support and Information System*.

Currit, N. 2002. Inductive regression: overcoming OLS limitations with the general regression neural network. *Computers, Environment and Urban Systems* 26(4), pp. 335-353.

Damen, K.,Faaij, A.,van Bergen, F.,Gale, J. and Lysen, E. 2005. Identification of early opportunities for CO2 sequestration—worldwide screening for CO2-EOR and CO2-ECBM projects. *Energy* 30(10), pp. 1931-1952.

Demesouka, O. E.,Vavatsikos, A. P. and Anagnostopoulos, K. P. 2013. Suitability analysis for siting MSW landfills and its multicriteria spatial decision support system: Method, implementation and case study. *Waste Management* 33(5), pp. 1190-1206.

Economides, M. J. and Wood, D. A. 2009. The state of natural gas. *Journal of Natural Gas Science and Engineering* 1(1–2), pp. 1-13.

Edsall, R. M. 2003. The parallel coordinate plot in action: design and use for geographic visualization. *Computational Statistics & Data Analysis* 43(4), pp. 605-619.

Finkel, M. L. and Hays, J. 2013. The implications of unconventional drilling for natural gas: a global public health concern. *Public Health* 127(10), pp. 889-893.

Fisher, J. B.,Kelly, M. and Romm, J. 2006. Scales of environmental justice: Combining GIS and spatial analysis for air toxics in West Oakland, California. *Health & Place* 12(4), pp. 701-714.

Force, C. A. T. and Graham, J. 2011. *Fine Particulate Matter (PM2. 5) and Ozone (O3) Air Quality in Western Pennsylvania in the 2000s*.

Foxton, F. and Jones, R. 2011. Social Capital Indicators Review. In: Statistics, O.f.N. ed. Crown copyright 2011.

Gilbuena Jr, R.,Kawamura, A.,Medina, R.,Amaguchi, H.,Nakagawa, N. and Bui, D. D. 2013. Environmental impact assessment of structural flood mitigation measures by a rapid impact assessment matrix (RIAM) technique: A case study in Metro Manila, Philippines. *Science of The Total Environment* 456–457(0), pp. 137-147.

Gorsevski, P. V.,Cathcart, S. C.,Mirzaei, G.,Jamali, M. M.,Ye, X. and Gomezdelcampo, E. 2013. A group-based spatial decision support system for wind farm site selection in Northwest Ohio. *Energy Policy* 55(0), pp. 374-385.

Grootaert, C. and Bastelaer, T. V. 2002. *Understanding and Measuring Social Capital
A Multidisciplinary Tool for Practitioners*. The World Bank.

Ha-Duong, M.,Gaultier, M. and deGuillebon, B. 2011. Social aspects of Total's Lacq CO2 capture, transport and storage pilot project. *Energy Procedia* 4(0), pp. 6263-6272.

Henriques, R.,Bacao, F. and Lobo, V. 2012. Exploratory geospatial data analysis using the GeoSOM suite. *Computers, Environment and Urban Systems* 36(3), pp. 218-232.

Holland, J. H. 1975. *Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence*. Oxford, England: U Michigan Press, pp. viii, 183.

Hossain, M. A.,Madkour, A. A. M.,Dahal, K. P. and Zhang, L. 2013. A real-time dynamic optimal guidance scheme using a general regression neural network. *Engineering Applications of Artificial Intelligence* 26(4), pp. 1230-1236.

Hsu, C.-W.,Chen, L.-T.,Hu, A. H. and Chang, Y.-M. 2012. Site selection for carbon dioxide geological storage using analytic network process. *Separation and Purification Technology* 94(0), pp. 146-153.

Huijts, N. M. A.,Midden, C. J. H. and Meijnders, A. L. 2007. Social acceptance of carbon dioxide storage. *Energy Policy* 35(5), pp. 2780-2789.

IAIA. 2009. *What is impact assessment?* IAIA International Headquarters, Fargo, USA: International Association for Impact Assessment.

Ibrahim, M. A.,Göransson, G.,Kaczala, F.,Hogland, W. and Marques, M. 2013. Characterization of municipal solid waste temporary storage sites: Risks posed to surrounding areas as a consequence of fire incidents. *Waste Management* 33(11), pp. 2296-2306.

IEA. 2013. World Energy Outlook 2013. IEA, pp. 116-117.

Imran, M.,Kumar, D.,Kumar, N.,Qayyum, A.,Saeed, A. and Bhatti, M. S. 2014. Environmental concerns of underground coal gasification. *Renewable and Sustainable Energy Reviews* 31(0), pp. 600-610.

Inamdar, P. M.,Cook, S.,Sharma, A. K.,Corby, N.,O'Connor, J. and Perera, B. J. C. 2013. A GIS based screening tool for locating and ranking of suitable stormwater harvesting sites in urban areas. *Journal of Environmental Management* 128(0), pp. 363-370.

Irfan, M.,Koj, A.,Sedighi, M. and Thomas, H. 2014. GIS based site ranking using neighbourhood analysis and comparison. *ThinkMind*, pp. 51-57.

Jenner, S. and Lamadrid, A. J. 2013. Shale gas vs. coal: Policy implications from environmental impact comparisons of shale gas, conventional gas, and coal on air, water, and land in the United States. *Energy Policy* 53(0), pp. 442-453.

Jeong, J. S.,García-Moruno, L. and Hernández-Blanco, J. 2013. A site planning approach for rural buildings into a landscape using a spatial multi-criteria decision analysis methodology. *Land Use Policy* 32(0), pp. 108-118.

Jia, J.,Fan, Y. and Guo, X. 2012. The low carbon development (LCD) levels' evaluation of the world's 47 countries (areas) by combining the FAHP with the TOPSIS method. *Expert Systems with Applications* 39(7), pp. 6628-6640.

Jones, N. S.,Holloway, S.,Smith, N. J. P.,Browne, M. A. E.,Creedy, D. P.,Garner, K. and Durucan, S. 2004. *UK coal resources for new exploitation technologies*.

Kemball-Cook, S.,Bar-Ilan, A.,Grant, J.,Parker, L.,Jung, J.,Santamaria, W.,Mathews, J. and Yarwood, G. 2010. Ozone impacts of natural gas development in the Haynesville Shale. *Environmental science & technology* 44(24), pp. 9357-9363.

Koornneef, J.,Faaij, A. and Turkenburg, W. 2008. The screening and scoping of Environmental Impact Assessment and Strategic Environmental Assessment of Carbon Capture and Storage in the Netherlands. *Environmental Impact Assessment Review* 28(6), pp. 392-414.

Koperski, K. and Han, J. eds. 1995. *Discovery of spatial association rules in geographic information databases*. Advances in spatial databases. Springer.

Maantay, J. 2007. Asthma and air pollution in the Bronx: Methodological and data considerations in using GIS for environmental justice and health research. *Health & Place* 13(1), pp. 32-56.

Malczewski, J. 1999. *GIS and Multicriteria Decision Analysis*. Wiley.

Malczewski, J. 2006. GIS-based multicriteria decision analysis: a survey of the literature. *International Journal of Geographical Information Science* 20(7), pp. 703-726.

Mazzotti, M.,Pini, R. and Storti, G. 2009. Enhanced coalbed methane recovery. *The Journal of Supercritical Fluids* 47(3), pp. 619-627.

McKenzie, L. M.,Witter, R. Z.,Newman, L. S. and Adgate, J. L. 2012. Human health risk assessment of air emissions from development of unconventional natural gas resources. *Science of the Total Environment* 424, pp. 79-87.

Meng, G.,Law, J. and Thompson, M. E. 2010. Small-scale health-related indicator acquisition using secondary data spatial interpolation. *Int J Health Geogr* 9, p. 50.

Meng, Q. 2014. Modeling and prediction of natural gas fracking pad landscapes in the Marcellus Shale region, USA. *Landscape and Urban Planning* 121(0), pp. 109-116.

Mennis, J. and Guo, D. 2009. Spatial data mining and geographic knowledge discovery—An introduction. *Computers, Environment and Urban Systems* 33(6), pp. 403-408.

Merdun, H. 2011. Self-organizing map artificial neural network application in multidimensional soil data analysis. *Neural Computing and Applications* 20(8), pp. 1295-1303.

Mondal, M. K.,Rashmi and Dasgupta, B. V. 2010. EIA of municipal solid waste disposal site in Varanasi using RIAM analysis. *Resources, Conservation and Recycling* 54(9), pp. 541-546.

Moore, T. A. 2012. Coalbed methane: A review. *International Journal of Coal Geology* 101(0), pp. 36-81.

Mostafa, M. M. and Nataraajan, R. 2009. A neuro-computational intelligence analysis of the ecological footprint of nations. *Computational Statistics & Data Analysis* 53(9), pp. 3516-3531.

Nasseri, M.,Asghari, K. and Abedini, M. J. 2008. Optimized scenario for rainfall forecasting using genetic algorithm coupled with artificial neural network. *Expert Systems with Applications* 35(3), pp. 1415-1421.

Nefeslioglu, H. A.,Sezer, E. A.,Gokceoglu, C. and Ayas, Z. 2013. A modified analytical hierarchy process (M-AHP) approach for decision support systems in natural hazard assessments. *Computers & Geosciences* 59(0), pp. 1-8.

Olawoyin, R.,Nieto, A.,Grayson, R. L.,Hardisty, F. and Oyewole, S. 2013. Application of artificial neural network (ANN)–self-organizing map (SOM) for the categorization of water, soil and sediment quality in petrochemical regions. *Expert Systems with Applications* 40(9), pp. 3634-3648.

Olsen, K. H. and Fenhann, J. 2008. Sustainable development benefits of clean development mechanism projects: A new methodology for sustainability assessment based on text analysis of the project design documents submitted for validation. *Energy Policy* 36(8), pp. 2819-2830.

Orem, W.,Tatu, C.,Varonka, M.,Lerch, H.,Bates, A.,Engle, M.,Crosby, L. and McIntosh, J. 2014. Organic substances in produced and formation water from unconventional natural gas extraction in coal and shale. *International Journal of Coal Geology* 126(0), pp. 20-31.

Pashin, J. C. 2007. Hydrodynamics of coalbed methane reservoirs in the Black Warrior Basin: Key to understanding reservoir performance and environmental issues. *Applied Geochemistry* 22(10), pp. 2257-2272.

Pastakia, C. M. R. and Jensen, A. 1998. The rapid impact assessment matrix (Riam) For eia. *Environmental Impact Assessment Review* 18(5), pp. 461-482.

Perpiña, C.,Martínez-Llario, J. C. and Pérez-Navarro, Á. 2013. Multicriteria assessment in GIS environments for siting biomass plants. *Land Use Policy* 31(0), pp. 326-335.

Perry, S. L. 2012. Development, Land Use, and Collective Trauma: The Marcellus Shale Gas Boom in Rural Pennsylvania. *Culture, Agriculture, Food and Environment* 34(1), pp. 81-92.

Pidgeon, N.,Corner, A.,Parkhill, K.,Spence, A.,Butler, C. and Poortinga, W. 2012. Exploring early public responses to geoengineering. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 370(1974), pp. 4176-4196.

Polat, Ö. and Yildırım, T. 2008. Genetic optimization of GRNN for pattern recognition without feature extraction. *Expert Systems with Applications* 34(4), pp. 2444-2448.

Rahman, M. A.,Rusteberg, B.,Gogu, R. C.,Lobo Ferreira, J. P. and Sauter, M. 2012. A new spatial multi-criteria decision support tool for site selection for implementation of managed aquifer recharge. *Journal of Environmental Management* 99(0), pp. 61-75.

Roddy, D. J. and Younger, P. L. 2010. Underground coal gasification with CCS: A pathway to decarbonising industry. *Energy and Environmental Science* 3(4), pp. 400-407.

Ruiz, M. C.,Romero, E.,Pérez, M. A. and Fernández, I. 2012. Development and application of a multi-criteria spatial decision support system for planning sustainable industrial areas in Northern Spain. *Automation in Construction* 22(0), pp. 320-333.

Saaty, T. L. 1980. *The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation*. New York: McGraw-Hill.

SADA. 2014. *Spatial Analysis and Decision Assistance (SADA)* [Online]. Available at: http://www.sadaproject.net/ [Accessed: Februrary 06].

Satkin, M.,Noorollahi, Y.,Abbaspour, M. and Yousefi, H. 2014. Multi criteria site selection model for wind-compressed air energy storage power plants in Iran. *Renewable and Sustainable Energy Reviews* 32(0), pp. 579-590.

Schalkoff, R. J. 2011. Intelligent Systems: Principles, Paradigms and Pragmatics. Jones & Bartlett Learning, pp. 493-495.

Shackley, S.,Mander, S. and Reiche, A. 2006. Public perceptions of underground coal gasification in the United Kingdom. *Energy Policy* 34(18), pp. 3423-3433.

Shackley, S.,Reiner, D.,Upham, P.,de Coninck, H.,Sigurthorsson, G. and Anderson, J. 2009. The acceptability of CO2 capture and storage (CCS) in Europe: An assessment of the key determining factors: Part 2. The social acceptability of CCS and the wider impacts and repercussions of its implementation. *International Journal of Greenhouse Gas Control* 3(3), pp. 344-356.

Soeder, D. J.,Sharma, S.,Pekney, N.,Hopkinson, L.,Dilmore, R.,Kutchko, B.,Stewart, B.,Carter, K.,Hakala, A. and Capo, R. 2014. An approach for assessing engineering risk from shale gas wells in the United States. *International Journal of Coal Geology* 126(0), pp. 4-19.

Špatenková, O. and Virrantaus, K. 2013. Discovering spatio-temporal relationships in the distribution of building fires. *Fire Safety Journal* 62, Part A(0), pp. 49-63.

Stamford, L. and Azapagic, A. 2014. Life cycle environmental impacts of UK shale gas. *Applied Energy* 134(0), pp. 506-518.

Stewart, R. N. and Purucker, S. T. 2011. An environmental decision support system for spatial assessment and selective remediation. *Environmental Modelling & Software* 26(6), pp. 751-760.

Sugumaran, R. and Degroote, J. 2010. *Spatial decision support systems: Principles and practices*. CRC Press.

Tayfur, G.,Zucco, G.,Brocca, L. and Moramarco, T. 2014. Coupling soil moisture and precipitation observations for predicting hourly runoff at small catchment scale. *Journal of Hydrology* 510(0), pp. 363-371.

Uliasz-Misiak, B.,Przybycin, A. and Winid, B. 2014. Shale and tight gas in Poland—legal and environmental issues. *Energy Policy* 65(0), pp. 68-77.

UNEP. 2014. *Emerging issues in our global environment*. Nairobi: UNEP.

Upham, P. and Smith, B. 2014. Using the Rapid Impact Assessment Matrix to synthesize biofuel and bioenergy impact assessment results: the example of medium scale bioenergy heat options. *Journal of Cleaner Production* 65(0), pp. 261-269.

Vafaeipour, M.,Hashemkhani Zolfani, S.,Morshed Varzandeh, M. H.,Derakhti, A. and Keshavarz Eshkalag, M. 2014. Assessment of regions priority for implementation of solar projects in Iran: New application of a hybrid multi-criteria decision making approach. *Energy Conversion and Management* 86(0), pp. 653-663.

van den Broek, M.,Brederode, E.,Ramírez, A.,Kramers, L.,van der Kuip, M.,Wildenborg, T.,Turkenburg, W. and Faaij, A. 2010. Designing a cost-effective CO2 storage infrastructure using a GIS based linear optimization energy model. *Environmental Modelling & Software* 25(12), pp. 1754-1768.

Wei, Y.,Maroto-Valer, M. and Steven, M. D. 2011. Environmental consequences of potential leaks of CO2 in soil. *Energy Procedia* 4(0), pp. 3224-3230.

Weijermars, R.,Drijkoningen, G.,Heimovaara, T. J.,Rudolph, E. S. J.,Weltje, G. J. and Wolf, K. H. A. A. 2011. Unconventional gas research initiative for clean energy transition in Europe. *Journal of Natural Gas Science and Engineering* 3(2), pp. 402-412.

White, C. M.,Smith, D. H.,Jones, K. L.,Goodman, A. L.,Jikich, S. A.,LaCount, R. B.,DuBose, S. B.,Ozdemir, E.,Morsi, B. I. and Schroeder, K. T. 2005. Sequestration of carbon dioxide in coal with enhanced coalbed methane recovery - A review. *Energy and Fuels* 19(3), pp. 659-724.

Wolsink, M. 2000. Wind power and the NIMBY-myth: institutional capacity and the limited significance of public support. *Renewable Energy* 21(1), pp. 49-64.

Wüstenhagen, R.,Wolsink, M. and Bürer, M. J. 2007. Social acceptance of renewable energy innovation: An introduction to the concept. *Energy Policy* 35(5), pp. 2683-2691.

Yu, H.,Zhou, G.,Fan, W. and Ye, J. 2007. Predicted CO2 enhanced coalbed methane recovery and CO2 sequestration in China. *International Journal of Coal Geology* 71(2–3), pp. 345-357.

Zanuttigh, B.,Simcic, D.,Bagli, S.,Bozzeda, F.,Pietrantoni, L.,Zagonari, F.,Hoggart, S. and Nicholls, R. J. 2014. THESEUS decision support system for coastal risk management. *Coastal Engineering* 87(0), pp. 218-239.

Zhou, F.,Hou, W.,Allinson, G.,Wu, J.,Wang, J. and Cinar, Y. 2013. A feasibility study of ECBM recovery and CO2 storage for a producing CBM field in Southeast Qinshui Basin, China. *International Journal of Greenhouse Gas Control* 19(0), pp. 26-40.

Zuo, C.,Birkin, M.,Clarke, G.,McEvoy, F. and Bloodworth, A. 2013. Modelling the transportation of primary aggregates in England and Wales: Exploring initiatives to reduce CO2 emissions. *Land Use Policy* 34(0), pp. 112-124.

3

SYSTEM DESIGN AND ARCHITECTURE

3.1 Introduction

This chapter presents the design and architecture of the SDSS developed in this research. The system design is based on the logical components of an SDSS as described in Section 2.1. The SDSS is developed following a modular approach where each component has a particular functionality embedded into it. This modular approach is useful to extend the capabilities of the system without any major structural changes to the overall system. The architecture of the system is designed to support the key functional aspects of the SDSS that will solve a range of problems described in Chapter 1.

Intended users of the system are suggested in Section 3.2. The overall design considerations are discussed in Section 3.3. The analytical modules and their overall functionalities are presented in Section 3.4. Ethical issues related to the user judgement on the selection of indicators and their relative weights are discussed in Section 3.5. Software development aspects, including the tools and technologies used, are explained in section 3.5. The overall architecture of the system is shown in section 3.6. A summary of this chapter is given in the last part of this chapter.

3.2 Intended users

The intended users of the SDSS developed in this research are the decision makers in local and national government institutions in particular those involved with planning, energy, socio-economic development, environment and public health related affairs. These decision makers may or may not be specialist of the GIS and IT. This has been considered in the overall system design and in the selection of the analytical modules developed to support decision making process as discussed in Section 3.4.

The developed system can also be used by the specialised users of GIS and IT, e.g. consultants, academicians and researchers. As explained in Section 3.5, the system has been developed using open source technologies so that the analytical modules individually or the overall system can be distributed to the target audience with minimum licencing issues.

3.3 Design considerations

The main system design consideration is to provide a reliable platform that can facilitate informed impact-based decision making while confronting semi-structured spatial decision problems related to Geoenergy and Geoenvironmental applications. Considering that the system has to be comprehensive, reliable, flexible and user friendly. The system should be able to provide the adequate functionalities to address a range of problems as discussed in Chapter 2. The system should be flexible enough to accommodate the decision maker's choices (soft information) in the analysis. The analytical modules used in the system should be verified and reliable. Finally, the system should be user friendly enabling decision makers with limited knowledge of GIS to use the system effectively.

Second important aspect of the design is to incorporate indicators from four different domains in the decision making context in an integrated manner. As discussed in Chapter 2, these four domains are: a) Environmental, b) Public Health, c) Socio-Economic and d) Techno-Economic. The key systems of environment, health and society are interlinked therefore they may affect each other.

Another important design consideration is to incorporate the complete process flow of the spatial decision problem solving in one platform. As discussed in Chapter 2, spatial decision making process consists of the three phases: a) Intelligence, b) Design and c) Choice. During the Intelligence phase a problem or an opportunity is conceptualised and required key datasets are identified. In Design phase, based on adequate analytical modules, possible solutions (alternates) are identified and developed. The Choice is the selection process of the most suitable solution based on the systematic evaluation of the alternates identified during the Design phase.

The system design is independent of the study area and the underlying spatial data in Geodatabase. Therefore, the system can be applied anywhere in the world, subject to the availability of the data. Any new indicators can be added to the geodatabase and existing data can be updated without affecting the system.

3.4 System components and modules

Based on the literature review provided in Chapter 2, the system is designed and developed around the three key components of the SDSS: a) Geodatabase, b) Model Base and c) User interface (Malczewski 1999). These main components are expanded to incorporate the entire set of functional requirements of the system. Figure 3.1 presents a schematic diagram of the three main components of the system.

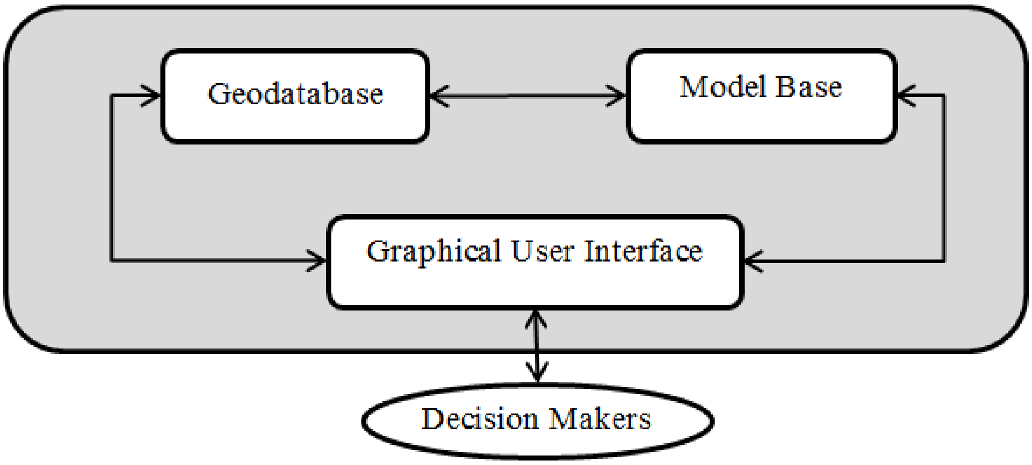


Figure 3.1 Components of an SDSS (Malczewski 1999)

Model Base is the central processing component and works as the brain of the system. The analytical modules of the system functions reside in the Model Base component where decision problems are solved by processing spatial and non-spatial data.

Spatial and non-spatial data is processed and stored in the Geodatabase component. The analytical modules utilise this data as required. The Geodatabase management component facilitates the storage, retrieval, manipulation, query and indexing of spatial and non-spatial data. Spatial data is in the form of GIS layers representing various topographic information of the study area whereas non-spatial information is the required statistical information linked with the study area.

Input and output functions reside in the user Graphical User Interface (GUI) part. User interacts with the SDSS using these interfaces or dialogues, in order to provide essential parameters for different types of analysis. The system generates results in the form of reports, maps and graphs. These components and their underlying functionalities are described in the following sections.

3.4.1 Geodatabase

Geodatabase is an essential component of the SDSS and it serves as the information backbone for the whole system. Spatial and non-spatial information is stored here and used by the system to solve spatial decision problems.

Designed system is generic but to demonstrate its capacity and validate developed components, spatial data from Wales (UK) has been used as case study. Therefore, spatial datasets related to Wales is acquired from available sources and incorporated into the geodatabase. The GIS layers are related to geology, elevation, hydrology, mineralogy, landuse-landcover (LULC), environment, public health and socio-economic aspects.

Key indicators are also acquired from various sources to be used within the decision making framework. These indicators cover environment, public health and socio-economic domains. Some indicators are directly added to the geodatabase as layers, while others are pre-processed using different GIS modelling techniques to form composite indicators representing complex socio-economic phenomena. GIS modelling, geodatabase design and development are explained in detail in Chapter 6.

The selection of indicators and other datasets incorporated into the geodatabase is based on the key environmental, socio-economic, public health and techno-economic aspects identified in the Literature Review. Sustainable development has been another key consideration in the selection of the indicators. Although sustainable development can be an independent theme but its underlying indicators falls within the broader geodatabase domains such as environment, social and economic. Lastly, the availability of the data (indicators) at a suitable scale defined the final shape of the geodatabase. However, the limitations related to the data availability do not affect the design and analytical capability of the system.

3.4.2 Model Base

Model Base serves as the brain of the SDSS assisting the decision making process. Model Base has the analytical modules to tackle the most commonly faced spatial decision problems related to Geoenergy and Geoenvironmental applications such as site selection, site ranking and impact assessment. It also provides tools for spatial knowledge discovery and geovisual analytics. Detailed discussion of these common spatial decision problems and tasks has been presented in Chapter 2.

Some of the analytical modules are based on artificial neural networks including Self-Organizing Maps (SOM) and General Regression Neural Networks (GRNN). Other analytical modules are based on different techniques, i.e. Analytical Hierarchy Process (AHP), Weighted Linear Combination (WLC), Rapid Impact Assessment Matrix (RIAM) and Parallel Coordinate Plots (PCP) etc.

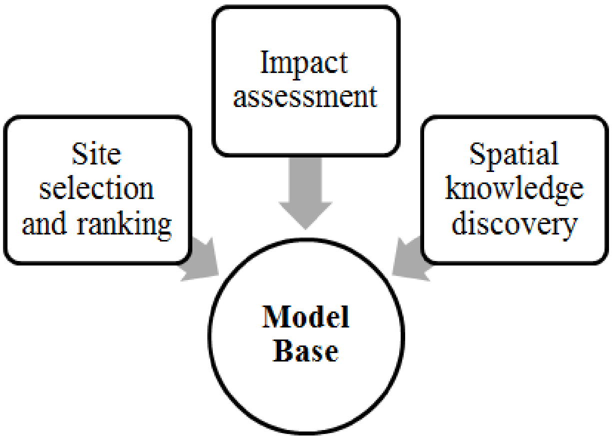


Figure 3.2 Main functionalities embedded within the Model Base

The functional aspects of the Model Base are divided into three main categories, as shown in Figure 3.3. These categories are explained below.

3.4.2.1 Site selection and ranking

Site selection and ranking subsystem has critical modules of the system to facilitate the site selection and ranking process. This is where the main emphasis of the SDSS application will be, i.e. to find the most suitable sites in terms of key socio-economic, environmental, public health and techno-economic indicators. There are three tools in this subsystem i.e. a) AHP based site selection tool b) SOM based site ranking tool c) Site ranking by neighbourhood analysis tool.

The AHP based site selection tool is the heart of the SDSS and it provides a hierarchical platform for decision problem solving. It utilises a combination of different techniques: a) Analytical Hierarchy Process, b) Weighted Linear Combination, c) Sensitivity Analysis, d) Commensuration tool and e) Pairwise Comparison Method. Additionally the site selection tool allows the user to view the analysis results on maps, in a number of different ways.

A number of techniques for site selection have been discussed in Section 2.5.2. AHP is found to be the most commonly used and well-established method for the purpose. It facilities the combination of hard and soft information into the decision making process. Also it is simpler to use even for the non-specialised end users. Some of the other techniques related to fuzzy classification, were not selected because it is not straightforward to select the most appropriate fuzzy membership function for a given dataset.

A commensuration tool is provided in the site selection module to scale the data into single currency. The Pairwise Comparison method provided in the tool helps the decision makers in checking the consistency of relative weights assigned to different indicators by the decision makers for a given analysis. The sensitivity analysis tool helps in checking the sensitivity of the analysis with respect to the relative weights. These features make it a comprehensive site selection tool to facilitate the decision makers.

The site selection tool is designed to be used in the first level of site selection process. It highlights the most potential areas in terms of the selected indicators (hard information) and user's preferred relative weights (soft information). At the next stage, Site ranking techniques are required to prioritise the most potential sites for further investigation.

As discussed in Section 2.5.2, at this stage of site ranking, it is envisaged to reduce user judgment and choice to the minimum. The potential sites should be ranked based on the key indicator's values at the site and in their surrounding neighbourhood.

The SOM based ranking tool uses the capabilities of Self Organizing Maps in order to reduce the number of dimensions, find natural clusters in the data and rank geographical areas in terms of the quality of indicators used in the analysis. SOM is preferred over some of the other clustering techniques because it is almost unsupervised and requires minimum input from the user. Also, one-dimensional SOM has the capability of ordering the data in ascending and descending order. Based on this capability, a novel site ranking mechanism has been introduced in the SDSS.

The site ranking by neighbourhood analysis tool is useful at the second stage of the site selection process where a number of potential candidate sites meet the basic criteria as set by the decision maker in the first stage of the site selection process. It ranks the candidate sites based on a systematic comparison of the surrounding areas of each site in accordance with key environmental, socio-economic and public-health indicators. The sites are ranked according to the status of key indicators in the given neighbourhood of sites being compared. Site ranking can be carried out using either a novel Criterion Sorting Mechanism (CSM) or Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS). TOPSIS has been selected as it is one of the most established ranking techniques as discussed in Section 2.5.2. TOPSIS can be used to cross compare the site ranks produced by CSM method.

3.4.2.2 Impact assessment

Impact assessment subsystem contains tools that can be used for the assessment of environmental and social impacts that engineering intervention may have at the potential site and its surrounding areas. This subsystem comprises three tools: a) Rapid Impact Assessment Matrix (RIAM) based impact assessment tool, b) General Regression Neural Networks (GRNN) based regression analysis and prediction tool and c) Traffic impact assessment tool.

The RIAM based tool is the heart of the impact assessment subsystem that quantifies the impact of an engineering intervention based on its positive and negative outcomes. RIAM is a semi-quantitative risk assessment mechanism that calculates the negative and positive impact in numerical format. RIAM has been selected as the impact assessment tool, as it is based on a mathematical background for the calculation of the impacts as discussed in Section 2.5.3. The categories of the impacts in terms of their importance, magnitude, permanence and reversibility are well structured as compared to some of the other similar techniques such as Life Cycle Assessment (LCA). Additionally, the components of RIAM are divided into four categories which are similar to the domains of the geodatabase applied in this research.

The GRNN based regression analysis and prediction tool is useful to predict the value of considered indicators at an unknown location, based on the similarity of indicators at the known locations. GRNN based tool can be used as an interpolator and also for impact assessment. This tool provides an option for the user to select the essential parameters for GRNN using either Holdout Method or Genetic Algorithm (GA).

As discussed in Section 2.3.2, GRNN has been preferred over other types of ANN for a number of reasons such as its simple structure and semi-supervised learning capability. Also unlike some of the other type of ANNs, GRNN does not work as a “black box”. Rather it predicts the values at an unknown location on the basis of its proximity to known location in terms of the selected independent variables. Additionally, because of its structure, it is easier to incorporate spatial parameters as one of the independent variable to support local variation in the regression analysis, for which a modified GRRN has been introduced in this research.

The traffic impact assessment tool can be used to estimate the percentage of increase in heavy traffic on the road network connecting the sites identified for Geoenergy. It uses existing traffic volume and emission data along the roads in the study area and calculates the percentage of the increase in traffic volume and emissions.

3.4.2.3 Spatial knowledge discovery

The spatial knowledge discovery component contains the tools that can facilitate the knowledge extraction from the data. These tools can be used for correlation and regression analysis of indicators and also to visually analyse the spatial variation of indicators across the geographical regions. These functions allow understanding the mutual relationships between key indicators from different domains. They can also determine how these indicators impact each other.

This information is also useful in the identification and selection of the most appropriate set of indicators for a given analysis. The selection of appropriate indicators is essential for the effectiveness of the decision making process. Not all the available indicators are useful for every analysis. In some cases, two or more of the independent indicators are strongly correlated with each other and hence only one of them should be used in the decision making context.

There are three tools in this category: a) SOM based clean correlation finding tool, b) GRNN based regression analysis tool and c) Parallel coordinate plotting tool. The SOM based clean correlation finding tool generates a matrix of clean correlation found among the indicators. Using the clean correlation matrix, number of indicators in the analysis can be reduced by selecting only those that are mutually independent and have strong correlations with the dependent indicator. The Parallel Coordinate Plots (PCP) tool can be used to select the most appropriate indicators (variables). PCP is also an effective exploratory analysis and data visualisation technique for the exploration of the structure of the data.

3.4.3 Graphical User interface

The third main component of the SDSS is the Graphical User Interface (GUI) which is used by the decision makers to interact with the system. Each module has a graphical user interface which contains necessary parameters to run that particular model. All the modules can be accessed from the main interface of the SDSS.

This main interface is the basic infrastructure of the SDSS which provides GIS functionalities and a hosting environment for the other analytical modules. This is the core of the system which links everything together. Decision making process starts here with loading the spatial data and then by applying one or more analytical models for decision support.

The main interface of the SDSS is designed in a way that can be extended easily by adding more functionality in form of additional modules. The modular design of the system ensures that minimum structural changes are required while adding more functionality to the existing system.

Analytical modules are provided using a menu bar control on this main interface under four different categories. First three categories cover analytical modules related to a) Site selection and ranking, b) Impact assessment and c) spatial knowledge discovery. These modules are explained in Section 3.4.2 as part of the Model Base. The fourth category contains the functions for geodatabase management including: a) Load thematic GIS layers, b) Select and load indicators as GIS layers and c) View metadata information regarding the available indicators in the Geodatabase. Figure 3.4 explains how different functionalities are combined together in groups within the SDSS main interface. This combination is based on the similarities of the functional behaviour of different analytical modules. The screen shots of GUIs are provided in Chapter 4, with the explanation of parameters required by each analytical module and its working.

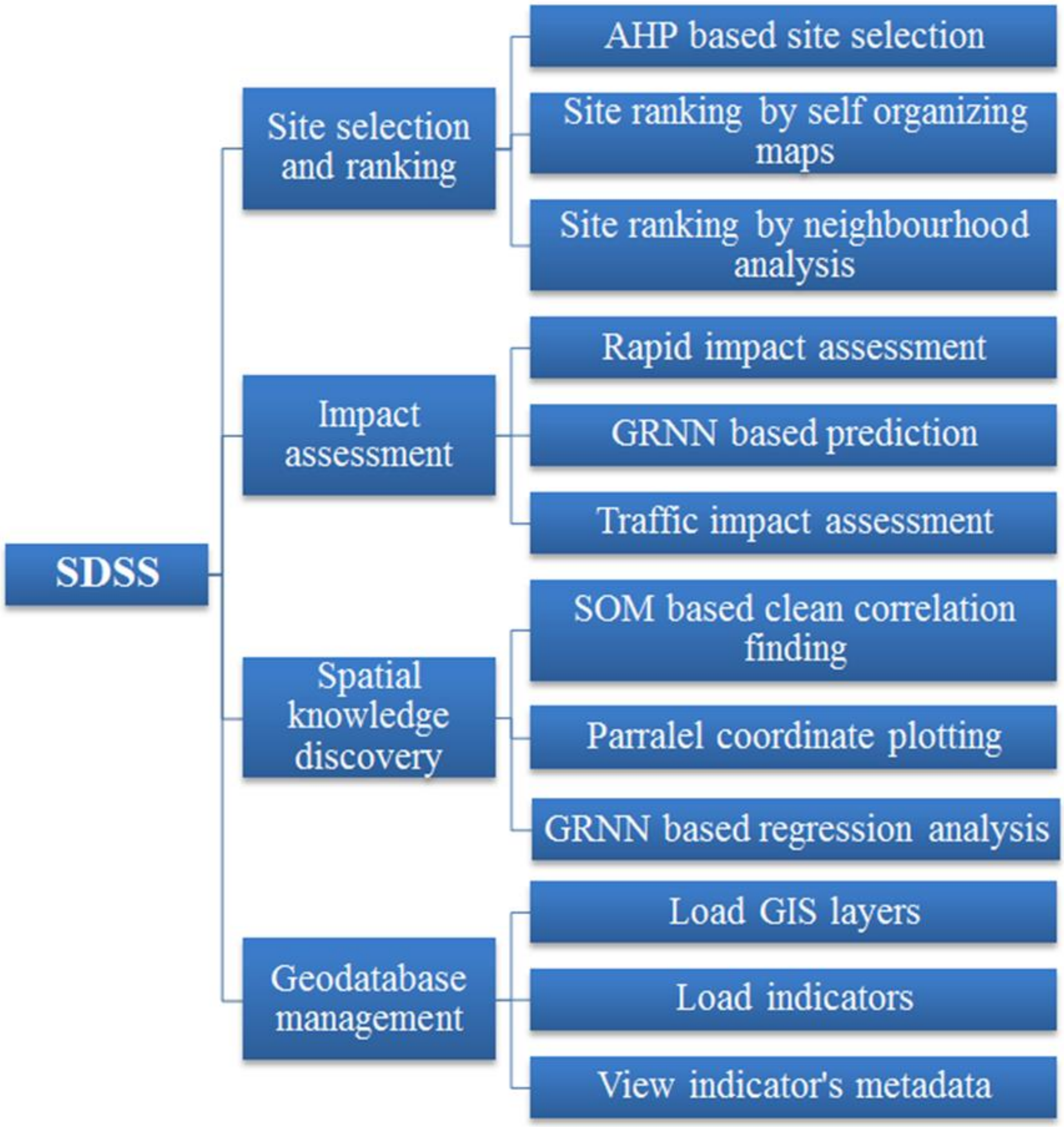


Figure 3.3 The SDSS User interfaces

3.5 Ethical issues with decision makers judgement and choice

As discussed above, the system is designed in a way that incorporates both data (hard information) and user judgements (soft information) into the decision making process. The ethical issues related to the soft information are of importance for fairness and justice. For example an operator would like to give more weight to the site economic parameters as compared to environmental and socio-economic parameters. This is a known limitation to multicriteria decision support systems as discussed in Section 2.3.1. Nonetheless, decision support systems are to support the decision makers and not to replace them.

Although, it is hard to place checks on the decision makers choice and judgements, however efforts have been made in the design and development of the SDSS to minimise the associated risks. This has been achieved by designing two novel site ranking techniques that rank the sites based on the status of selected key indicators at the sites and in their neighbourhood with minimum input required from the decision makers. Similarly, sensitivity analysis and Pairwise Comparison Method have been provided in the site selection tool to help the selection of an appropriate set of weights for the indicators used in a given analysis.

RIAM based site impact tool has been designed in a way that checks the spatial variability of the impacts using underlying spatial data, comparing the immediate neighbourhood of the sites with the entire study area. The quantitative nature of the site impact assessment methodology makes it easier to interpret the results and compare multiple sites. This reduces the risk associated with the qualitative judgements made by the decision makers about the impacts of potential sites.

3.6 Software development

The development of the system has been carried out using a modular approach where each module is conceptualised on the basis of its functional aspects. This section explains development lifecycle and the technological aspects of the system development.

3.6.1 Software development life cycle

Considering modular architecture of the SDSS, the Incremental Model of the Software Development Life Cycle (SDLC) has been adopted to develop the system. Each module is considered to be an increment to the overall system. Incremental approach that has been taken is useful in scenarios, when the broader system requirements are known but detailed requirements to each component are unclear.

The incremental model can deliver initially the fundamental structure of the system and then functionalities or modules can be added until the whole system is developed (Sommerville 2007). Figure 3.5 presents the flow chart of the incremental SDLC model adopted in the development of the system.

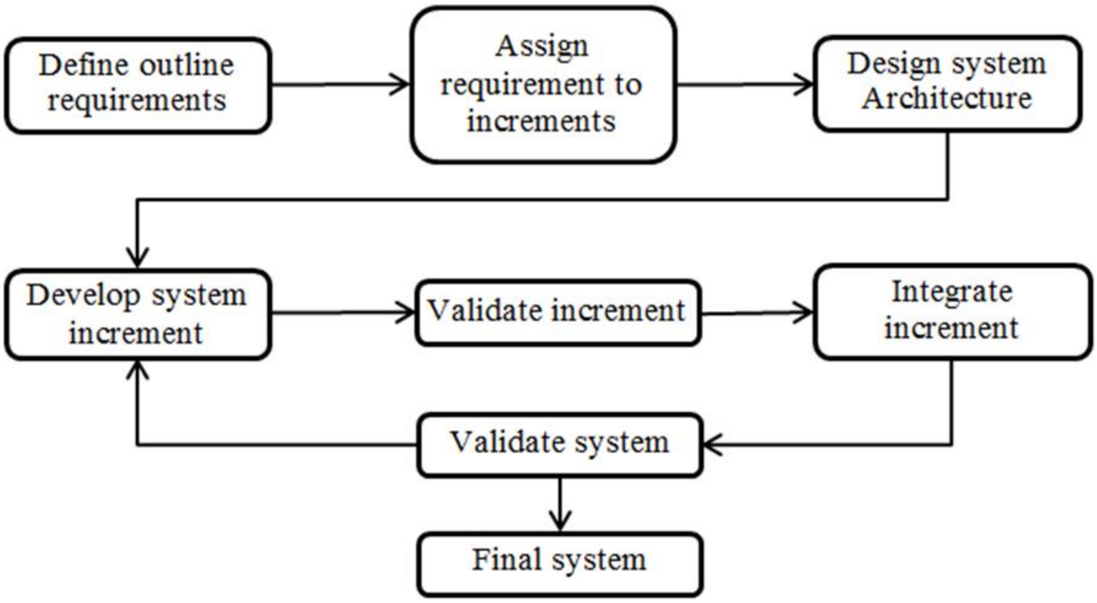


Figure 3.4 The Incremental Model of the SDLC (Sommerville 2007)

The process starts with an implementation of the basic requirements and then with each increment more functionality is added to the existing system until all the requirements are met (Sommerville 2007). This model was selected because of the conceived modular design of the system. In addition, a working system with the basic functionality can be used from the beginning. Any changes and new components can be incorporated into the system without any major changes to the existing structure. Therefore incremental model has been recognized as potentially suitable for research-oriented applications.

At first, the main interface of the system has been designed and developed, providing basic GIS functionalities such as visualisation, zoom-in, zoom-out, query, importing and exporting spatial data. Also, this main structure provides a platform for the analytical modules to be added in a sequential manner. Each analytical module has been conceived, designed and developed separately and added to the main SDSS as an increment to the overall SDSS functionality.

This section explains details of the tools and technologies considered for the development of different components of the SDDS. A number of options have been studied for the development of the Geodatabase and the System (Model Base and GUI). The emphasis has been on open source and non-commercial tools and technologies.

3.6.2 Technology selection for Geodatabase

Geodatabase is a relational database capable of storing, querying and indexing geographical data. Geodatabase is needed when a large amount of spatial data has to be stored and analysed. The traditional file structures for storing geographical data as layers becomes very difficult to manage and very time consuming to query. Considering its user-base and conformance to the Open Source Geospatial Consortium (OGC), three most widely used geodatabases are considered: a) PostGIS, b) SpatiaLite and c) mySQL-Spatial.

PostGIS is a spatial extension of PostgreSQL, which is a well-established, non-commercial and open source database. PostGIS has the largest user base and in terms of functionalities, reliability and robustness it is considered as the free alternative to propriety Oracle11g-Spatial database (PostGIS 2014).

SpatiaLite is an open source, light weight and single file geodatabase and provide full SQL engine. SpatiaLite is not based on the complex Client-Server architecture, therefore it is simple to implement (SpatiaLite 2014). In terms of its structure and functionalities, it is comparable to the ESRI's propriety File Geodatabase.

MySQL-Spatial is the spatial extension of famous light weight MySQL database but its spatial queries are limited to the Minimum Bounding Rectangles (MBR) around the features instead of the actual geometry of the features. Therefore mySQL-Spatial has its limitations in terms of the accuracy of spatial queries (Steiniger and Hunter 2013).

Both PostGIS and SpatiaLite are open source databases which can support handling, manipulation and querying of spatial data. Both are OGC compliant geodatabases and offer all the features required by the SDSS for spatial data storage, retrieval, management and query. PostGIS is based on the client-server architecture. In order to be accessed by the client applications, an installation of the database server is independently required (PostGIS 2014).

The SpatiaLite is however a light weight single file geodatabase where data can be packed together with the system for distribution. There is no limit for the size of the geodatabase file and it also comes with a user friendly GUI application for management. For these technical reasons, SpatiaLite was selected for the geodatabase development. Further details of the SpatiaLite geodatabase can be found on the SpatiaLite web page (SpatiaLite 2014).

3.6.3 Technology selection for system development

For the development of the SDSS, an existing framework or library was required to provide the basic GIS functionalities. These functions are to deal with the spatial data visualisation, creation, manipulation and queries. Literature review suggested that a number of open source and free libraries are available for the Java, C++ and .Net based developments (Steiniger and Hunter 2013). The .Net based spatial libraries were considered for the system development. The two most widely used .Net spatial libraries are SharpMap and DotSpatial.

SharpMap is .Net framework 4.0 based mapping library for embedding GIS functionality in web and desktop applications. SharpMap is released under Lesser General Public License (ShapMap 2014).

DotSpatial is an open source spatial library written for .NET framework 4.0. It supports developers incorporating the GIS functionalities in their .NET applications. DotSpatial has been used to develop an entire open source GIS platform called MapWindow6 (MapWindow 2014). It has been used for the development of several spatial information/decision support systems (Osna et al. ; Zanuttigh et al. ; Ames et al. 2012; Steenbeek et al. 2013; Steiniger and Hunter 2013; Horsburgh and Reeder 2014).

DotSpatial library was chosen for the development of the SDSS. It provides all the basic GIS functionalities such as spatial data visualisation, projection, query and manipulation. It also provide readymade controls that can be used in the development of graphical interface i.e. map, legend and layout controls. More information about DotSpatial can be found on the website (DotSpatial 2014).

The main interface of the SDSS utilises different readymade controls provided by DotSpatial library to support basic GIS functionality such as the Map, Legend, Status Bar and Toolbar Control etc. Map Control renders the GIS layers for visualisation. It is also connected with the Legend Control, where each layer has a legend item. Layers can be turned on and off from the Legend Control. Toolbar Control provides a number of GIS tools to zoom in and out, identify feature and to access the attribute table.

3.6.4 High level system architecture

The high level system architecture design is an abstract representation of the overall system where three key components of the SDSS, technological aspects of the design and the interaction between different system components are highlighted.

As described earlier, the application is developed in Microsoft C# .Net programming language. Microsoft Visual Studio 2012 is used as the Interactive Development Environment (IDE). Open source DotSpatial library is used to provide basic GIS functionality. Open source SpatiaLite geodatabase is used for the development of geodatabase. System.Data.SqLite is used as the Active Data Object for .Net (ADO.NET) to connect the .Net application with the SpatiaLite geodatabase.

The high level system architecture design is presented in Figure 3.7. Architecture diagram highlights the analytical modules in the Model Base component of the system. The geodatabase component included the key spatial and aspatial datasets. The interaction of the decision makers with the system by using the GUI component is also presented.

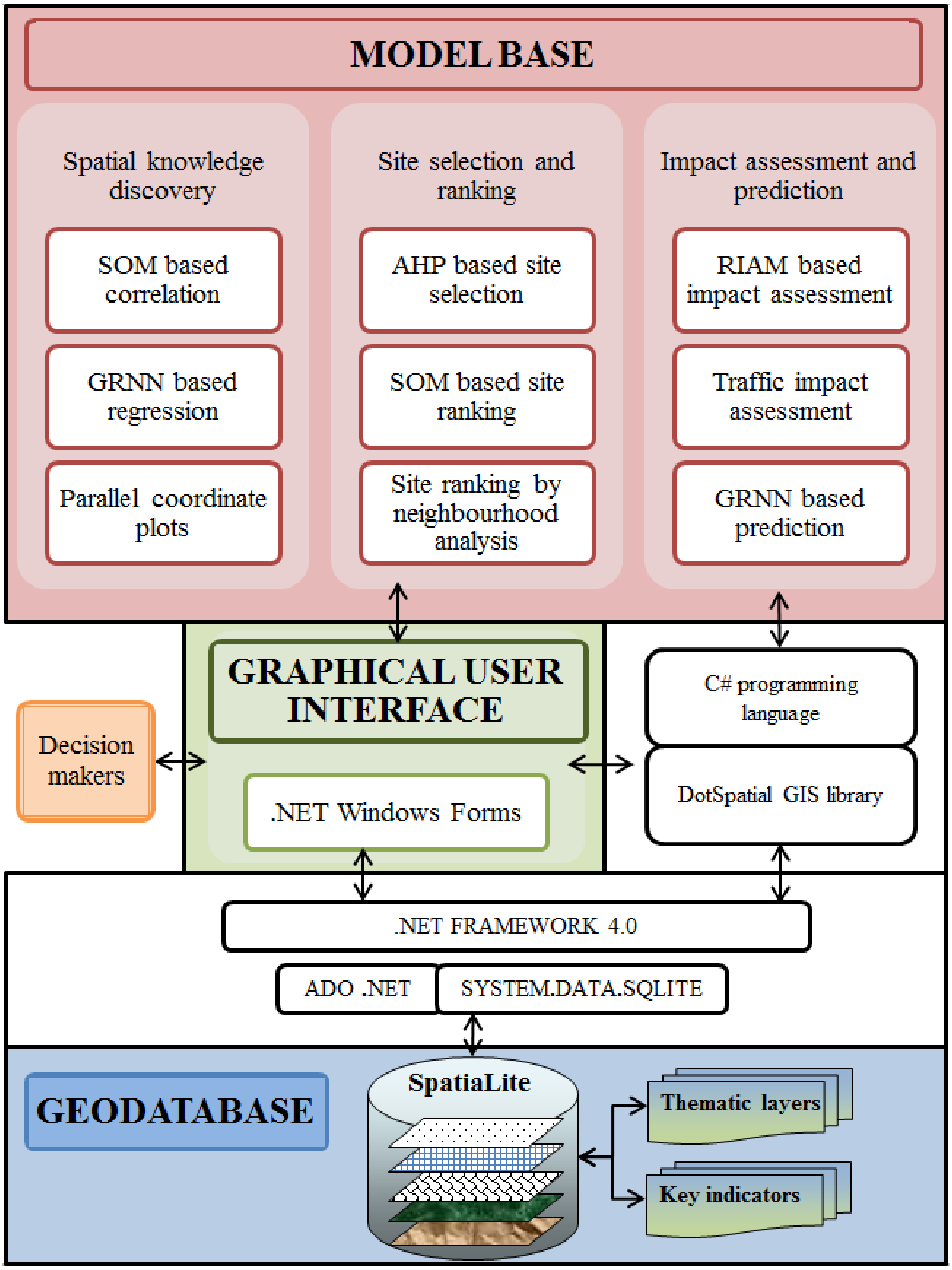


Figure 3.5 High level system architecture

3.7 Summary

This chapter presents the system architecture design for the spatial decision support system that has been developed. The SDSS is designed in a way to facilitate the decision makers in solving semi-structured spatial decision problems related to Geoenergy and Geoenvironmental applications. The three component based SDSS design presented by Malczewski has been consolidated. These components are a) Geodatabase, b) Model Base and c) Graphical user interface. The spatial and non-spatial data is stored in the geodatabase. The analytical models are part of the Model Base and the user interacts with the system using the GUIs.

The integrated environmental, public health, socio-economic and techno-economic domains based decision making has been considered in the design and development of different analytical modules of the system.

To develop this system, the Incremental Model of the SDLC that follows modular development approach, has been adopted. The technological aspects of the system development are also highlighted in this chapter. The open source GIS library DotSpatial is used for the development of the system using .NET C# programming language. The open source single-file and light-weight SpatiaLite geodatabase is used for storing and manipulation of the spatial and non-spatial data.

An integrated system architecture design diagram is shown, presenting the three main components of the system and their mutual connectivity, analytical modules, technological aspects and the interaction of decision maker with the system.

References

Ames, D. P.,Horsburgh, J. S.,Cao, Y.,Kadlec, J.,Whiteaker, T. and Valentine, D. 2012. HydroDesktop: Web services-based software for hydrologic data discovery, download, visualization, and analysis. *Environmental Modelling & Software* 37(0), pp. 146-156.

DotSpatial. 2014. *DotSpatial Library* [Online]. Available at: http://dotspatial.codeplex.com/ [Accessed: October 14].

Horsburgh, J. S. and Reeder, S. L. 2014. Data visualization and analysis within a Hydrologic Information System: Integrating with the R statistical computing environment. *Environmental Modelling & Software* 52(0), pp. 51-61.

Malczewski, J. 1999. GIS and Multicriteria Decision Analysis. Wiley, pp. 280-286.

MapWindow. 2014. *MapWindow 6* [Online]. Available at: http://www.mapwindow.org/ [Accessed: Februrary 06].

Osna, T.,Sezer, E. A. and Akgun, A. GeoFIS: An Integrated Tool for the Assessment of Landslide Susceptibility. *Computers & Geosciences* (0).

PostGIS. 2014. *PostGIS Geodatabase* [Online]. Available at: http://postgis.net/ [Accessed: February 06].

ShapMap. 2014. *SharpMap geospatial application framework* [Online]. Available at: http://dotspatial.codeplex.com/ [Accessed: October 14].

Sommerville, I. 2007. Software Engineering.8 ed. Addison Wesley, pp. 71-73.

SpatiaLite. 2014. *SpatiaLite Geodatabase* [Online]. Available at: https://www.gaia-gis.it/fossil/libspatialite/index [Accessed: February 06].

Steenbeek, J.,Coll, M.,Gurney, L.,Mélin, F.,Hoepffner, N.,Buszowski, J. and Christensen, V. 2013. Bridging the gap between ecosystem modeling tools and geographic information systems: Driving a food web model with external spatial–temporal data. *Ecological Modelling* 263(0), pp. 139-151.

Steiniger, S. and Hunter, A. J. S. 2013. The 2012 free and open source GIS software map – A guide to facilitate research, development, and adoption. *Computers, Environment and Urban Systems* 39(0), pp. 136-150.

Zanuttigh, B.,Simcic, D.,Bagli, S.,Bozzeda, F.,Pietrantoni, L.,Zagonari, F.,Hoggart, S. and Nicholls, R. J. THESEUS decision support system for coastal risk management. *Coastal Engineering* (0).

4

SDSS DEVELOPMENT

ARTIFICIAL INTELLIGENCE BASED ANALYTICAL MODULES

4.1 Introduction

This chapter covers the developmental aspects of the SDSS, in particular those analytical modules that utilises Artificial Intelligence (AI) techniques such Artificial Neural Network (ANN) and Genetic Algorithm (GA). The functionality of these modules is explained using flow charts, mathematical formulations, figures and tables. Graphical User Interfaces (GUIs) are explained and user interaction with the system is discussed.

As explained in Chapter 3, modular approach has been adopted to develop the system. In order to provide required functionality, adequate tools have been developed. Based on these different functionalities, analytical modules are grouped together in four sections: a) Site Selection and Ranking, b) Impact Assessment, c) Spatial Knowledge Discovery and d) Geodatabase Management. The analytical modules can be used separately or in conjunction with each other. A typical use case scenario of the SDSS can based on utilising each module, sequentially to facilitate the decision making process.

As explained in chapter 3, the system is developed using Microsoft .Net C# programming language and using an open source GIS library called DotSpatial (DotSpatial 2014). The system can be deployed on windows XP or Windows 7 operating systems (32 or 64 bit). Initially the main interface of the SDSS is developed, to provide a foundation for these analytical modules. This main interface is called SDSS Core which is explained in section 4.2.

Some of the analytical modules utilize the artificial intelligence based soft computing techniques including ANN and GA. Therefore the structure and working of these techniques are explained first in section 4.3 before their development and utilisation in different analytical modules is explained. The working of artificial intelligence techniques can be quite complex and it may require significant amount of assistance from the user in order to perform well. Therefore only certain type of ANN and GA techniques are incorporated in the SDSS that require minimum assistance from the user. For this purpose two types of ANN are selected, i.e. a) The Self Organizing Maps (SOM) and b) the General Regression Neural Networks c).

SOM is a semi-supervised ANN and only a small number of parameters are required from the user to perform. Although GRNN is supervised ANN but they also require a small number of parameters to perform and GA is incorporated to facilitate the decision makers in the selection of most appropriate parameters required to enable the GRNN to perform well.

4.2 SDSS Core

The main interface of the SDSS is composed of different .Net C# and DotSpatial controls. It provides the basic GIS functionality and hosting environment for the rest of the modules. It is the starting point for any analysis using the SDSS. The user loads data into the system using this interface to be utilised by other analytical modules as required. Figure 4.1 shows the main SDSS interface.

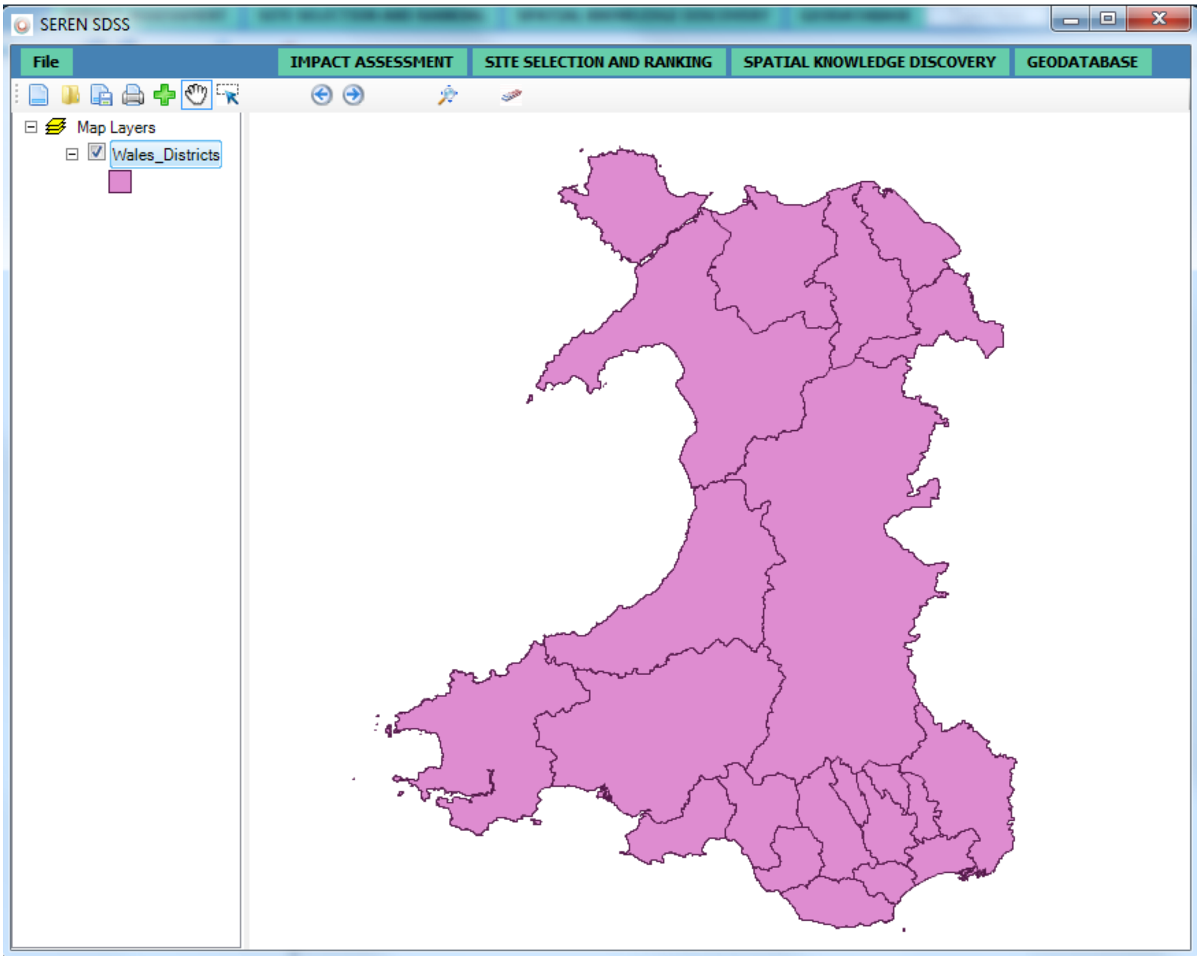


Figure 4.1 SDSS main interface

The map control is used for map visualisation and also for the user inputs on maps using mouse. The legend control has an entry for each GIS layer loaded in the map control to turn on/off or change the aesthetics of an individual GIS layer. The map control and legend controls are hooked together using an application manager control. The toolbar control on the top contains common tools for GIS operations such as zoom in, zoom out, feature selection and for adding more GIS layers for overlay operations.

The menu bar control hosts different modules (tools) of the SDSS where core functionality resides. As explained earlier, these tools are grouped together in four different sections based on the similarity of their functional aspects: a) Site Selection and Ranking, b) Impact Assessment, c) Spatial Knowledge Discovery and d) Geodatabase Management.

4.3 Artificial Neural Network approaches used in the SDSS

As mentioned earlier, some of the analytical modules utilize the artificial intelligence based soft computing techniques such as ANN and GA. Before different SDSS modules and their functionality is explained, it is important to explain the types of neural networks used in the SDSS. Two types of neural network techniques are used in the analytical modules of the SDSS i.e. a) The Self-Organizing Maps (SOM) and b) the General Regression Neural Network (GRNN).These two types of ANN are used in: i) The Self-Organizing Maps (SOM) based indicator selection tool, (ii) The SOM based site ranking tool and (iii) The General Regression Neural Networks (GRNN) based prediction and regression analysis tool.

4.3.1 Self-Organizing Maps (SOM)

Self-Organizing Maps (SOM) are unsupervised artificial neural networks. SOM are used extensively for exploratory data analysis. SOM are used to visualise high dimension data, to reduce the number of dimensions in the data and for clustering the data (Kohonen 2013). SOM preserve the natural structure of data using a neighbourhood function, while mapping it from high dimensional space (large number of variables in data) to a low dimensional space (usually 2 dimensional), i.e. the output map. It converts the non-linear statistical relationship in original data space (high dimensions), into simple geometrical relationship on the output space (low dimensions) (Kohonen 2001).

4.3.1.1 Structure and algorithm of SOM

Figure 4.2 illustrates a Self-Organizing Map. Output map consists of the model vectors. Model vectors are normally organized in either one dimensional or two dimensional output spaces. Model vectors contain a vector entity whose size is equal to the number of dimensions in the input data. The vector entity stores the values of each dimension (variable) for the model vector it is associated with. Initially, these values are randomly generated for each model vector in the output map.

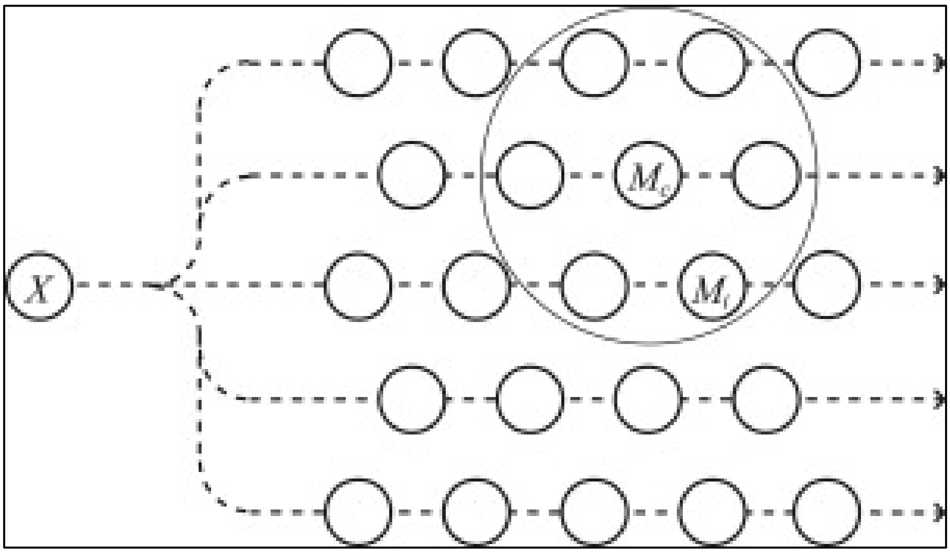


Figure 4.2 Illustration of a self-organizing map. An input data item X is broadcast to a set of models M_i, of which M_c matches best with X. All models that lie in the neighbourhood (larger circle) of Mc in the grid match better with X than with the rest. (Kohonen 2013)

Each input signal (input data vector) is represented by its corresponding model vector in the output map. Since the number of model vectors is normally less than the number of input data points, therefore one model vector can represent several input data points which are similar to each other. The number of input data points represented by each model vector can vary and depends on the natural clusters found in data. Input data points are randomly exposed to the neural networks and its Best Matching Unit (BMU) is identified from the model vectors in the output map. BMU is closest to the input data point in terms of all the dimensions as compare to the rest of the model vectors.

After the selection of the BMU, the neighbouring model vectors adjust their position in the output map in order to get closer to the best matching unit. The movement of each model vector is defined by a neighbouring function e.g. a Gaussian distribution function, time passed in the self-organizing process and its distance from the BMU in the output map.

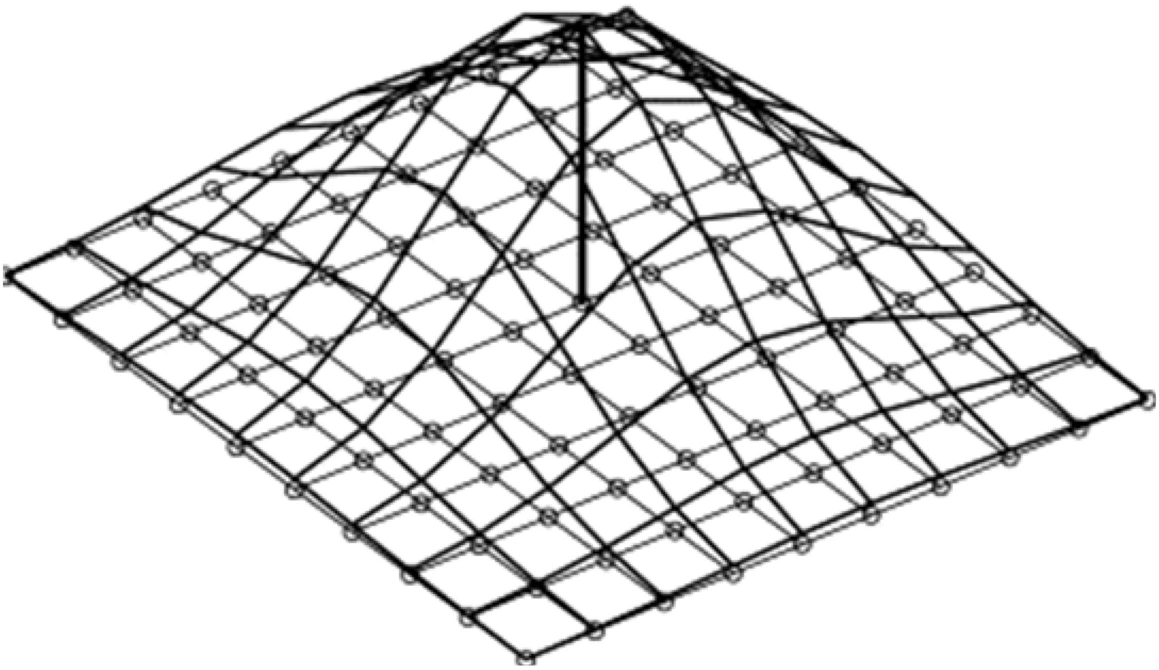


Figure 4.3 Movement of model vectors towards the BMU in self-organized output map (Kohonen and Somervuo 2002)

In Figure 4.3, the model vectors in output map are moving towards the best matching unit node. This movement is controlled by a neighbourhood function, time passed during the self-organization process and the distance of the model vector from the BMU. Immediate neighbouring model vectors move more, to get closer to the BMU as compare to the distant ones. After some time, the output map is self-organized and converged to represent the relationships in input data. Model vectors in the self-organized output map represent the mean values of clusters, naturally found in the data. After some time the output map converges to represent the relationships found in the input dataset (Kohonen 2001).

4.3.1.2 Training of SOM

The original SOM algorithm as defined in (Kohonen 2001), is used for the training of SOM. The process of training is explained in Figure 4.4. During the training process each data item is presented to the output map and the BMU is searched. After some time the output map converges to represent the relationships found in the input dataset (Kohonen 2001). The duration of training process is controlled by using a learning rate which has a very small value closer to zero. The movement of each model vector towards the BMU is compared to this learning rate. In the beginning of the training process these movements are quite large but after some time these movements get smaller and smaller and when it is smaller than the learning rate, the training process is stopped.

After convergence, some of the model vectors may not represent any of the input data vectors. It depends on the size of the output map, how the data is naturally clustered in the input domain and the training parameters of the neural network. If the output map has a very small number of neurons (model vectors) as compare to the input data points, then each model vector may become BMU for one or more input data vectors. If the output map size is larger, then only a few model vectors will become the BMU for input data vectors. Therefore, determining an appropriate size of the output map can be a challenging task. The Topographic and Quantization error terms can be used to select the most appropriate size of the output map for given dataset. The details of how to calculate these two error terms is given in Section 4.3.1.4. The mathematical aspects of the SOM training and convergence algorithm are given in details in Section 4.3.1.3 including the identification of BMU and movement of model vectors towards the BMU at a given time step.

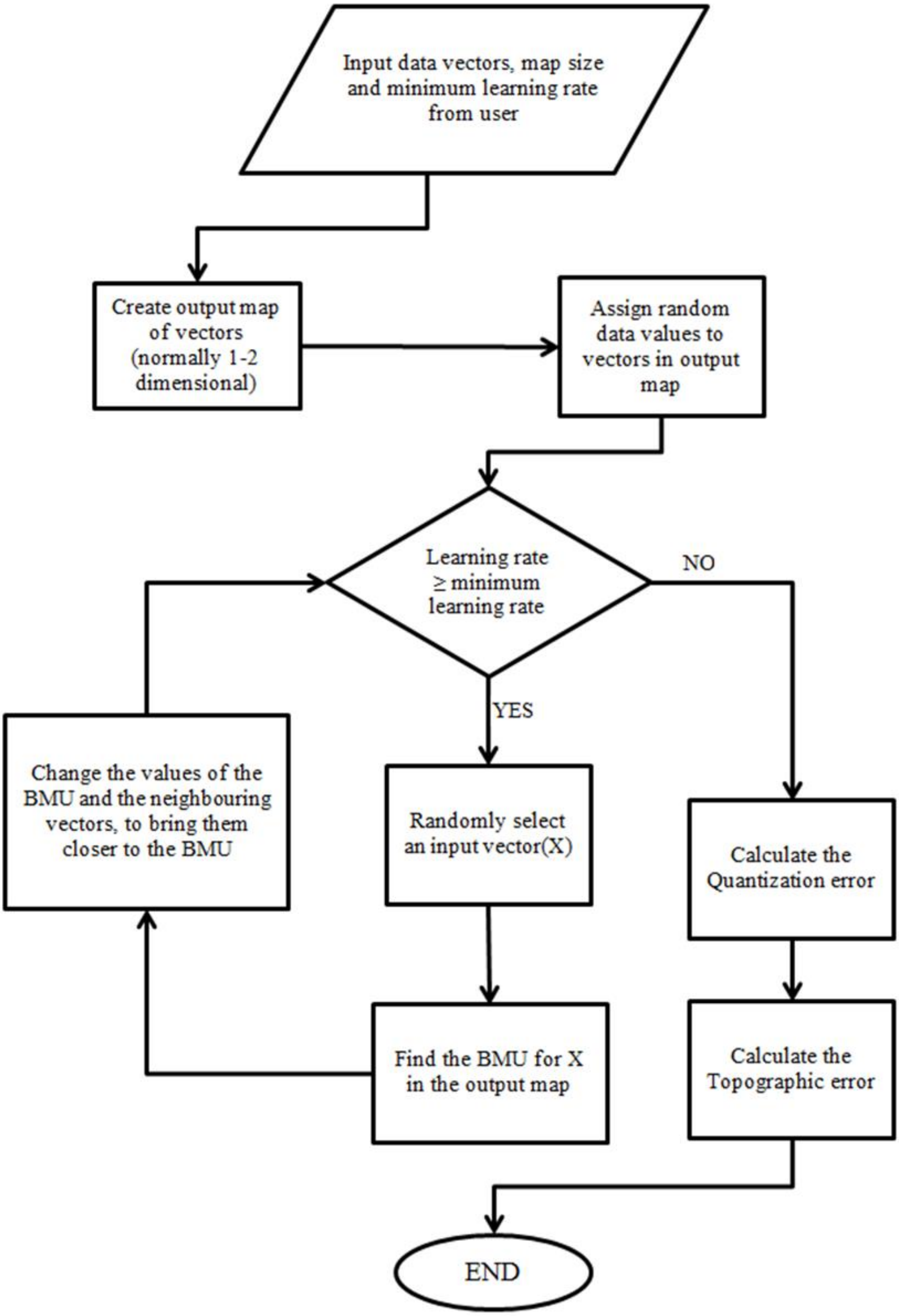


Figure 4.4 Flow chart showing the training process of SOM

4.3.1.3 Mathematical formulation

The BMU is identified by calculating the distance of the input vector with every model vector in the output space and selecting the one with shortest distance. BMU is identified by parameter C which is the index of the model vector that has the smallest Euclidean distance from the input data vector (Kohonen 2013).

$$C = \arg min_i \ \{||x - m_i||\} \tag{4.1}$$

where x is the input data vector and m_i is the model vector at ith index in the output map. The distance between the input data vector and each model vector is calculated and the closest model vector is selected as the BMU. Self-organization is an iterative process and every time a model vector is identified as MBU, the surrounding model vectors in the defined neighbourhood of the BMU are moved closer to the BMU (Kohonen 2001). This process continues until the convergence is achieved. The model vector at step interval (t+1) in the self-organizing output map is calculated using its value at step interval (t), its difference with the input vector and a neighbourhood function (Kohonen 2013).

$$m_i(t + 1) = m_i(t) + hc_i(t)\ [\mathrm{x}(t) - m_i(t)] \tag{4.2}$$

where x is the input data vector and m_i is the model vector at ith index in the output map. The modified value of m_i is also dependent on a neighbourhood, e.g. $hc_i(t)$. The neighbourhood function is some type of an exponential decay function that shrinks with the time. At the beginning, the neighbourhood is big and covers the entire map of output vectors but as the time passes, this is reduced to the immediate neighbourhood of the BMU. The neighbourhood function used here is the same as defined by (Kohonen 2001).

$$hc_i(t) = \alpha(t).\exp\left(\frac{||r_c - r_i||^2}{2\sigma^2(t)}\right) \tag{4.3}$$

The terms $\alpha(t)$ and $\sigma(t)$ are both monotonically-decreasing functions of time. The term $\alpha(t)$ is called the learning rate factor and has a value between 0-1. The term $\sigma(t)$ defines the kernel size and it decreases with time. The value of $hc_i(t)$ tends to become zero when time tends to become infinity. The terms rc and ri are the location vectors of nodes c and i in the output space, where c is the node location of the BMU for the current data input point. The larger is distance between the best matching unit and other model vectors, the smaller is the value for the term $hci(t)$ (Kohonen 2001).

4.3.1.4 Error estimation in SOM

The only input parameter required from the user to run SOM is an appropriate size of the output map (dimensions and number of nodes). That is why SOM is considered to be in the category of unsupervised neural networks. The size of the output map is very important for a good convergence. There are two methods to check the SOM convergence and its error estimation, i.e. a) Quantisation error and b) Topographic error (Kohonen 2001). The Quantization error is the average distance of all the input nodes from their respective BMU in the self-organized output map. Whereas, the Topographic error is calculated by identifying the first and second BMU for each input data vector and then by checking if these two are situated next to each other in the output map. If they are immediate neighbours then it is a perfect preservation and the Topographic error term would be zero and otherwise one. The overall Topographic Error of the self-organized is the mean Topographic error of each input data vector.

4.3.2 General Regression Neural Networks (GRNN)

General Regression Neural Network or GRNN is a type of Radial Basis Functions (RDF) and belong to the category of Probabilistic Neural Network (Specht 1991). GRNN is one pass neural networks and highly parallel in structure. GRNN do not require an iterative process to learn from the training dataset. Predictions can rather be made just after one pass of input data through the GRNN (Specht 1991). GRNN approach is used for the regression analysis and estimation or prediction of the quantitative (continuous) variables. GRNN is also known as the universal interpolators and can be used as an effective technique for data interpolation (Polat and Yildirim 2008a). GRNN are useful for the regression analysis of underlying relationship in the variables, both linear or nonlinear (Cigizoglu and Alp 2006).

4.3.2.1 Structure of GRNN

GRNN is very simple in its structure and has four layers of neurons, i.e. a) Input layer, b) Pattern layer, c) Summation layer, d) output layer. Figure 4.5 shows the general structure of the GRNN with these four layers which is originally suggested by Specht (1991).

The structure of GRNN is shown in Figure 4.5 which works as a feed forward network. It can approximate a function and estimate the value of a dependent variable from a set of independent variables.

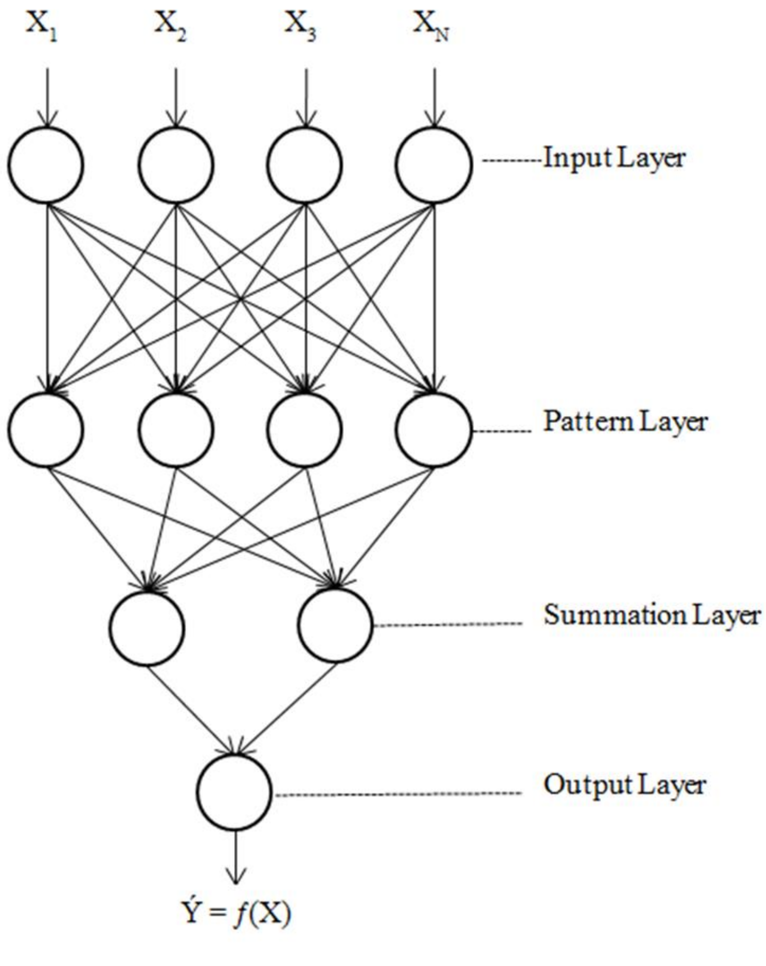


Figure 4.5 Structure of General Regression Neural Network (Specht 1991)

The Input layer contains as many neurons as there are variables in the input dataset. The input data points are presented to the Input layer which merely feeds into the second layer called the Pattern layer. Each input data point is stored in the Pattern layer. During the network learning process one data point is selected at a time and its difference (Euclidean distance) is calculated from other data points using 4.5. The summation layer computes the nominator and denominator terms for 4.4 by using the difference factor, independent variables (at known and unknown location) and dependent variable (at known location). The last layer is called the Output layer where the value of function $\acute{Y} = f(x)$ is computed using 4.4.

4.3.2.2 Mathematical formulation

The mathematical formulation to implement GRNN is straight forward and similar to probability distribution function. The output function of the GRNN can be given as (Specht 1991):

$$\acute{Y} = f(x) = \frac{\sum_{i=1}^{n} Y^i \ exp^{\left(-D_i{}^2/2\sigma^2\right)}}{\sum_{i=1}^{n} \ exp^{\left(-D_i{}^2/2\sigma^2\right)}} \tag{4.4}$$

where $\acute{Y}$ is the estimated value of the dependent variable at the unknown location, Y^i is the value of dependent variable at known locations and D_i is a scalar term that shows the difference between the prediction point and the training sample in terms of all the independent variables (dimensions) and it can be calculated as (Specht 1991):

$$D_i{}^2 = \left(X - X^i\right)^T \left(X - X^i\right) \tag{4.5}$$

The distance between the prediction point and a training sample defines the influence of that training sample in the calculation of the $f(x)$ (the dependent variable $\acute{Y}$). If this distance is small, the term $exp^{\left(-D_i{}^2/2\sigma^2\right)}$ increases and becomes equal to 1 if the difference is 0. A larger value of this term means the known value of dependent variable at this training sample will have more influence in the calculation of the dependant variable at the prediction point. If the distance is large, the value of the term $exp^{\left(-D_i{}^2/2\sigma^2\right)}$ decreases and becomes 0 for very large distances. Such sample points will have no contribution in the estimation of dependent variable at the predicted location. The predicted output always remains in between the maximum and minimum known values of the dependent variable (Polat and Yildirim 2008b).

4.3.2.3 Smoothing parameter sigma (σ)

The σ parameter can have single or multiple values for different variables (dimensions) in input dataset. If a single value is used, it is very important to normalise the input data. Normalisation ensures that single σ value can be used for all the dimensions (dependent variables). If normalization is not carried out and different dimensions are in different units, a single σ value will cover different distances in each dimension and the value of $D_i{}^2$ will not represent the actual difference between the training sample and the prediction point (Specht 1991). If dimensions have different influence in the estimation of dependent variable, then different σ parameters can be used. A smaller σ value will result in a localised regression analysis i.e. only the sample points that are very close to the prediction point in terms of their distances on different axis (domains) will contribute to the calculation of dependent variable. A larger σ value results in a more globalised regression where almost the entire set of data samples contributes to the calculation of the dependent variable. In such cases, results are very close to the mean value of the dependent variable in the entire set of sample points.

4.3.2.4 Holdout Method for training of GRNN

GRNN require supervised training and the selection of the most suitable value for the σ smoothing parameter is very important for the reliable results of GRNN (Leung et al. 2000). The Holdout Method is a useful and common method for the selection of σ (Specht 1991). In Holdout Method only one training sample is selected from the training set and the value of Ý is predicted at this sample point using rest of the samples (Specht 1991). The predicted value can be compared with the actual value and the difference can be used in the calculation of mean squared error (Specht 1991). RMSE is calculated by using the difference between actual and predicted value at each sample using:

$$RMSE = \frac{1}{n}\sqrt{\sum_{i=1}^{n}(\acute{Y}_i - Y_i)^2} \tag{4.6}$$

Each sample is taken out of the training set, one by one and RMSE value is re-adjusted. After the sample is placed back in the training set and the next sample is held out. Different values of σ are used to calculate the RMSE value and the one with least RMSE is selected for the actual prediction at the prediction point (Specht 1991).

Figure 4.6 explains the algorithm of Holdout method in a flow chart. The structure of GRNN is based on the number of variables in the data. Number of neurons at the input layer is equal to the number of variables in the data whereas the number of neurons in the patterns layer is equal to the total number of data sample point.

It is important to first normalize all the data if a single value for σ parameter is used for all the variables. Next, the user selects a random σ value and the Holdout Method is initiated. One sample points is selected at a time from the dataset and prediction (Ý) is made at this point using the rest of the sample points. After the prediction is made, the RMSE value is adjusted and the sample point is placed back in the dataset to be used in the estimation at other sample points.

When the prediction is done at all the sample points, one epoch (iteration) is completed. At this stage the RMSE values are compared with the desired RMSE value, as provided by the user. If the target RMSE has been achieved then the value of σ is selected and calculations are made at the prediction points. However, if the RMSE value is larger than the target RMSE value, the user assigns a new σ value to be tested with another epoch using Holdout method. This process can be repeated several times until the desired results are achieved (Specht 1991).

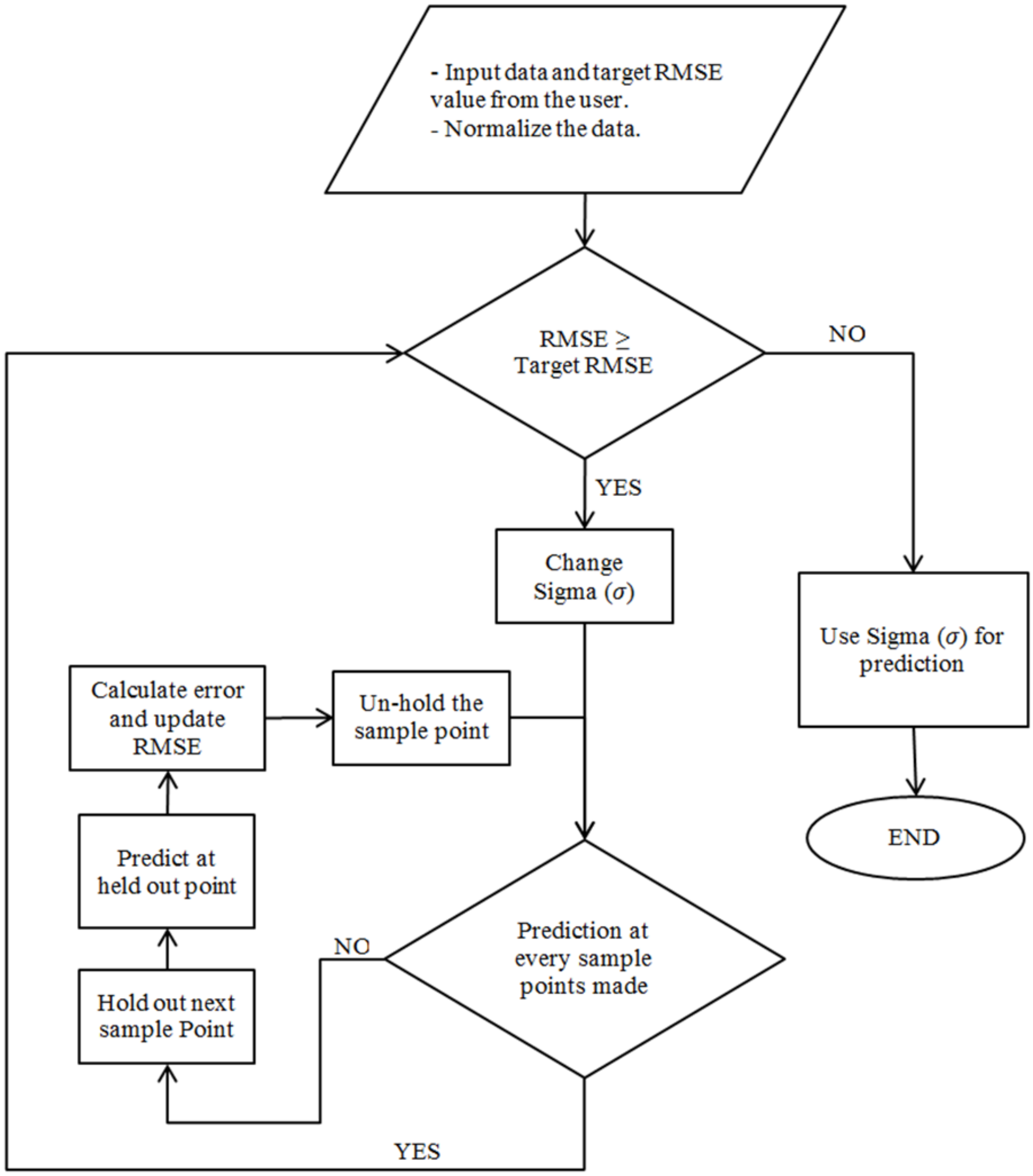


Figure 4.6 Flow chart of the Holdout Method for the selection of sigma (σ)

4.3.2.5 Genetic algorithm for the training of GRNN

The most essential part of the Holdout Method is the input of the σ parameter in each cycle from the user. The user has to decide whether an increase or a decrease in the σ value is going to help in reducing the RMSE value to the desired level. Normally, a few initial values provide a general trend of this effect, but user has to enter a new value of σ for every cycle.

Thus, the testing of a large number of σ parameters to find an optimum solution can be very time consuming and demanding.

If different σ parameters are to be used for different variables (domains) the selection of appropriate values can become even more tedious. In this case the user has to assess and provide an appropriate σ value for each dimension in every epoch to find an optimum set of σ parameters. It is possible that σ values are exhibiting different effects in different directions. The RMSE can be optimised by trying different combinations of values in the set of σ parameters. Genetic Algorithm (GA) can also be used for the selection of an appropriate set of σ parameters along with the Holdout Method as discussed in Section 2.3.2.

The structure and working of the GA is discussed in Chapter 2. Initially a population of individuals (sets of σ parameters) is randomly generated. The Holdout method is applied on the entire population and RMSE is calculated for each set of σ parameters. As the evolution process continues, only the fittest of the individuals (set of σ parameters) survive in the given population. The process stops after a fixed number of generations or if a desired RMSE value is achieved. The set σ parameters with least value of RMSE, is selected to be used for the calculation at the prediction points. The process of crossover and mutation is explained in Chapter 2. The holdout method used in this flow chart for the calculation of RMSE is the same as explained in Figure 4.6. Therefore it is shown as a predefined process.

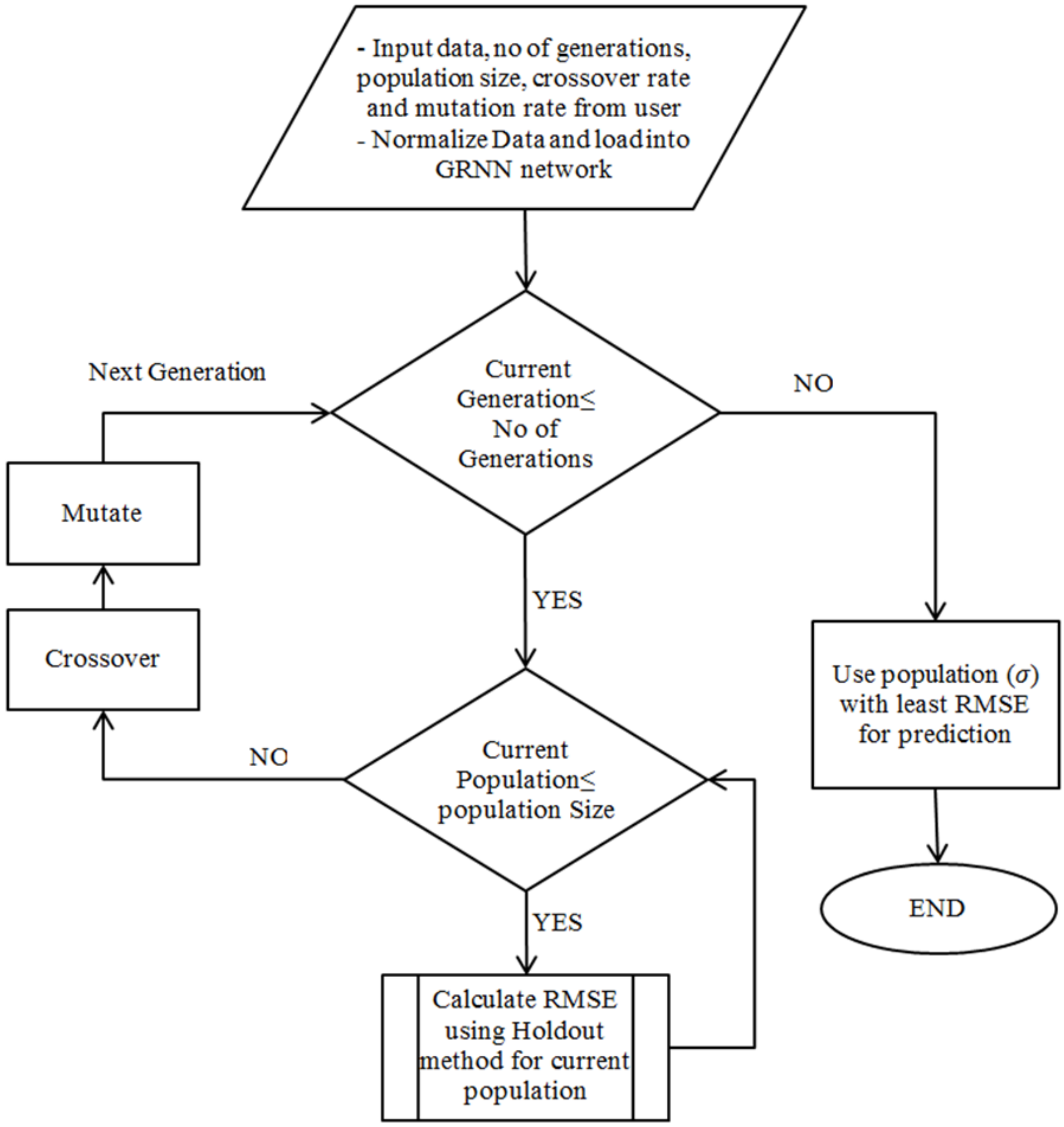


Figure 4.7 Flow chart of Genetic Algorithm for the tuning of GRNN sigma parameter (σ)

The user inputs are essential for running the algorithm for σ optimization. User provides number of generations, population size in each generation, crossover rate, mutation rate and the data itself. The algorithm requires that the population size should always be an even number to function properly.

4.4 Spatial Knowledge Discovery

Spatial Knowledge Discovery subsystem has tools to facilitate the decision makers in knowledge extraction from the data. As explained in Chapter 2, data driven approach is very useful in informed risk based decision making. The tools provided in this subsystem can be used for analysis related to correlation, regression and spatial-colocation of different indicators in a given geographical space. Also, the tools provided in this subsystem can facilitate the decision makers for the selection of most appropriate indicators. The selection of an appropriate set of variables is the key to the reliability and effectiveness of the analysis.

This section contains two tools that can be useful for the spatial knowledge discovery and for the selection of the most appropriate indicators: a) SOM based clean correlation tool and b) Parallel coordinates plotting tool. Along with these two tools, GRNN tool can also be utilised in finding the most useful set of variables that can be used in the analysis. For example, GRNN has been used by (Bowden et al. 2006) for the determination of the most appropriate variables to forecast chlorine in preventing the spread of waterborne diseases.

4.4.1 SOM based clean correlation tool

The working of self-organizing maps has been explained in Section 4.3.1. SOM has the capability of preserving the most important topological and metric relationships found in input dataset (Kohonen 1998). SOM preserves these relationships in the low dimensional output map in the form of geometrical connections where similar data points are clustered together.

As described earlier, SOM can be used for exploratory data analysis in a variety of different ways, e.g. reducing dimensions and clustering similar elements in the input data. Another possible use of SOM is to inquire about the structure of the data itself. For correlation hunting, the component planes are visualised together to see any visible similarity. A component plane is a sliced version of the output map showing one dimension of data at a time. If the number of component planes is high, then it is very difficult to trace any correlation that exists, by visual inspection only (Vesanto and Ahola 1999). In such scenarios where visual inspection is not appropriate due to the presence of a large number of dimensions, correlation can be calculated directly from the output plane (Corona and Lendasse 2005).

SOM based clean correlation tool can be used in the identification of correlation that may exists between different indicators in a given dataset. The independent variables having strong correlation with dependent variable can be selected and used in different type of analysis, e.g. regression or causal analysis.

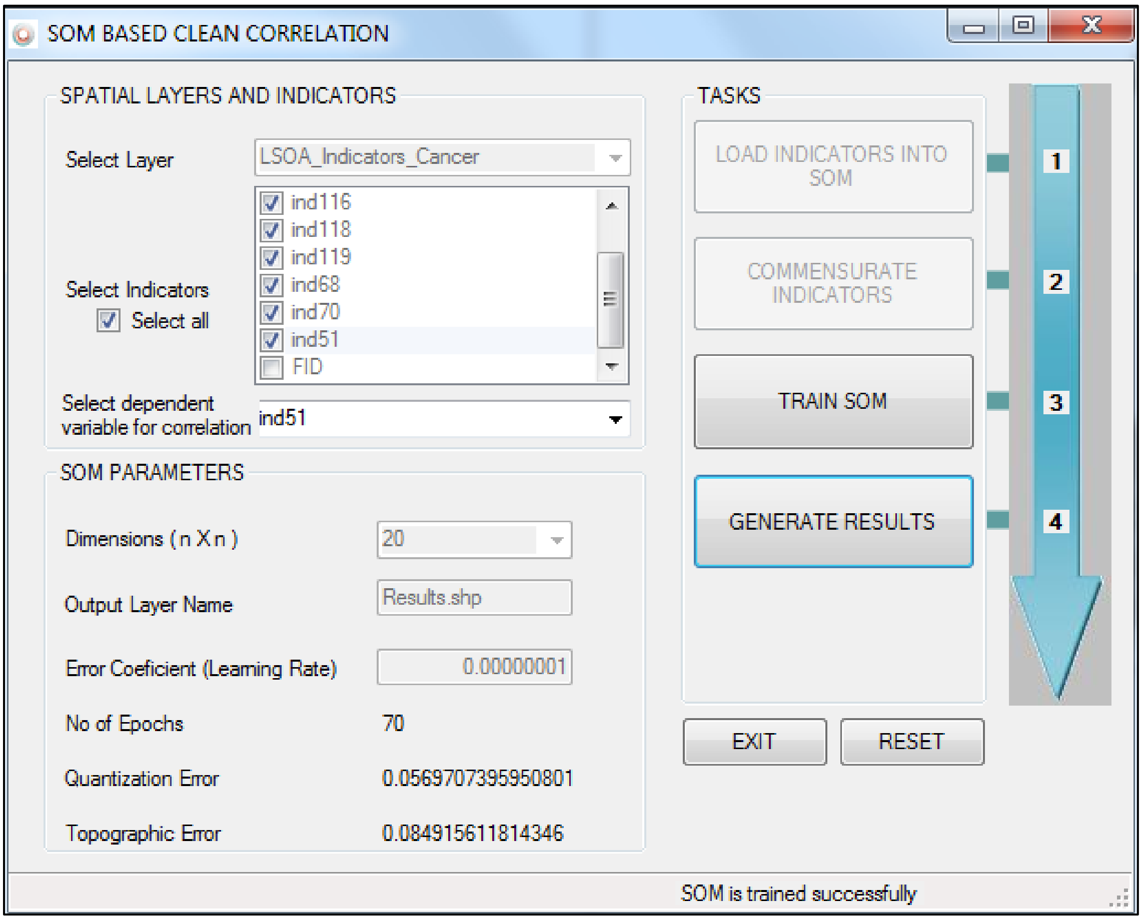


Figure 4.8 GUI of SOM based clean correlation tool

Figure 4.8 shows the GUI of the SOM based clean correlation tool. The user loads a GIS layer into the SDSS that already contains all the indicators for analysis. This layer is loaded into the tool using the GUI, from where the user can select the dependent and independent variables. The essential parameters of SOM are also provided by the user. These parameters have been explained in detail in Section 4.3.1. Once indicators are selected, they can be commensurate if they are in different units. The commensuration process brings all the indicators in the same scale between 0-1. Moreover during the commensuration process, the "Cost" or "Benefit" nature of the indicators is also described by the user. The commensuration process is described in detail in Section 5.4.1.1. After the commensuration, training process of the SOM is carried out as explained in Section 4.3.1.

The convergence of SOM can be checked with the help of topographic and quantization errors as explained in Section 4.3.1.4. After the convergence of SOM is achieved, the input data vectors become associated with their best matching unit and the correlation can be calculated directly from these BMUs instead of the entire data set. This also ensures that the presence of noise in the data has less effect on the correlation finding than as compare to the original data. This is because of the noise resistance capabilities of SOM as suggested in (Vesanto and Ahola 1999). The clean correlations can be calculated directly from the BMUs using (Corona and Lendasse 2005):

$$CC_{j\,k} = \frac{1}{\sigma_j \sigma_k} \sum_{l=1}^{M} (m_{lj} - \mu_j) * (m_{lk} - \mu_k) \tag{4.7}$$

where $CC_{j\,k}$ is the Clean Correlation between the variable j and k, σ_j and σ_k are the Standard deviations of j and k. Mean values of j and k are represented by μ_j and μ_k, l is the index number of model vector from (1 - M) where M is the total number of model vectors.

Each of the input vectors will be associated with one of the model vectors in the output map as its best matching unit. The results are generated in the form of a table showing correlations between dependent and each independent variable. Those variables having stronger correlation can be selected for further analysis. The same procedure can be used for the identification of any collinear independent variables in the input set.

The correlation is always between -1 and 1, where -1 states that the two variables are extremely correlated but in opposite directions, i.e. if one increases, the other decreases and vice versa; 1 means that the two variables are extremely positively correlated and 0 means no correlation exists between the two in given dataset.

4.5 Site Selection and Ranking

The site selection and ranking section has one of the key analytical modules developed in the SDSS. There are three different tools in this subsystem: a) AHP based site selection tool, b) SOM based site ranking tool and c) Site ranking by neighbourhood analysis tool.

4.5.1 SOM based site ranking tool

Self-organizing maps based site ranking tool is useful for multi criteria sorting or ranking of sites of geographical regions in terms of the appropriateness of their attributes (indicators). One-dimensional SOM have the proven capabilities of converging and ordering of the inputs when t approaches ∞. After a certain time the output maps is self-organized in ascending or descending order. After convergence has been achieved and the output map is in ascending or descending order, even with further processing, the SOM will retain this order (Kohonen 2001).

The Site ranking tool utilises this capability of one-dimensional SOM to cluster, order and rank the geographical features of sites based on their key attributes. This is a novel use of One-Dimensional SOM introduced in the current research. The indicators are first commensurate according to their “Cost” or “Benefit” nature and using an appropriate scaling method as described in Section 5.4.1.1. After commensurations all the indictors are scaled between 0-1 (having 0 as the worst value and 1 as the best values).

The one-dimensional SOM can order itself in both the directions depending on the input. Once ordered, the first and last neuron (model vector) of the output map can be verified to check whether it is in ascending or descending order. Once this order is known, an ordered rank is assigned to each BMU so that the BMU having higher values of key indicators gets the highest rank and the BMU with lower values of indicators get the lowest rank. Total number of ranks is equal to the size of the one-dimensional SOM. The GUI of SOM based site ranking tool is shown in Figure 4.9.

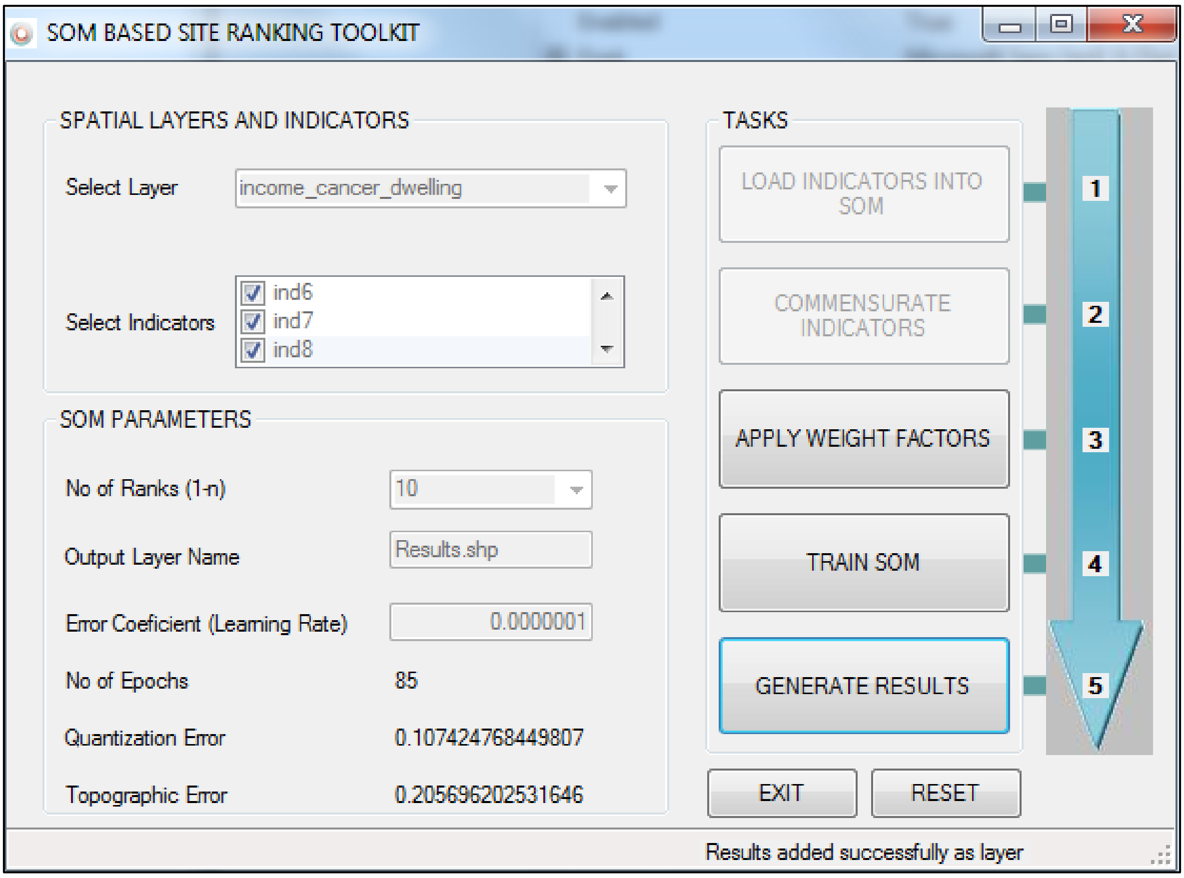


Figure 4.9 GUI of the SOM based site ranking tool

Training of SOM process is the same as explained in Section 4.3.1. Once the convergence has been achieved, each data point is represented by it corresponding BMU in the output map. In the first step, user loads data into the tool and assigns the essential parameters of the SOM. The tool then scales all the indicators between 0-1 during the commensuration step. At this stage user also identifies each indicator as “Cost” or “Benefit” in nature. This is necessary to bring all the data into single currency. For scaling, the tool provides Maximum Score Procedure and Score Range Procedure by using an appropriate equation (5.1 to 5.5).

In the next step, user assigns weights to emphasize the importance of different indicators over each other. User can assign weights directly or use Pairwise Comparison Method to find a suitable relative weighting scheme. Tool opens a separate window to facilitate the weighting process. If all the indicators are of equal importance then this step can be skipped. Next, the algorithm specified in Section 4.3.1 is used to train SOM.

The training is stopped when the user-given error coefficient (learning rate) is achieved. The smaller this number is, the longer it takes for SOM to train and the more chances are there for the SOM to reach its convergence and to order its neurons in the output space. The results are loaded in the map viewer window of the SDSS as a GIS layer.

The resultant layer contains all the attributes as they were in the input layer. An additional column "Clusters" is added to the layer with a number (1-n), where n is the total number of ranks, given by the user. It is the ranked order of each geographical feature, as assigned by the SOM ranking tool, based on the selected indicators. A colour ramp is used to symbolise the geographical features based on their rank order. The geographical features in a GIS layer are clustered and ranked as show in Figure 4.10. The commensuration process is important in this clustering and ranking process. As shown in Figure 4.10, the same geographical areas that are shown in darker colours in the left side map are shown in lighter colours in the right side map. This is because the left map is the result of a commensuration process where all the indicators were marked as "Benefit" in nature, whereas the map to the right is the result of a commensuration process where all the indicators were marked as "Cost" in nature. Clustering and ranking of the geographical features depends on the "Cost" or "Benefit" nature of the indicators as well.

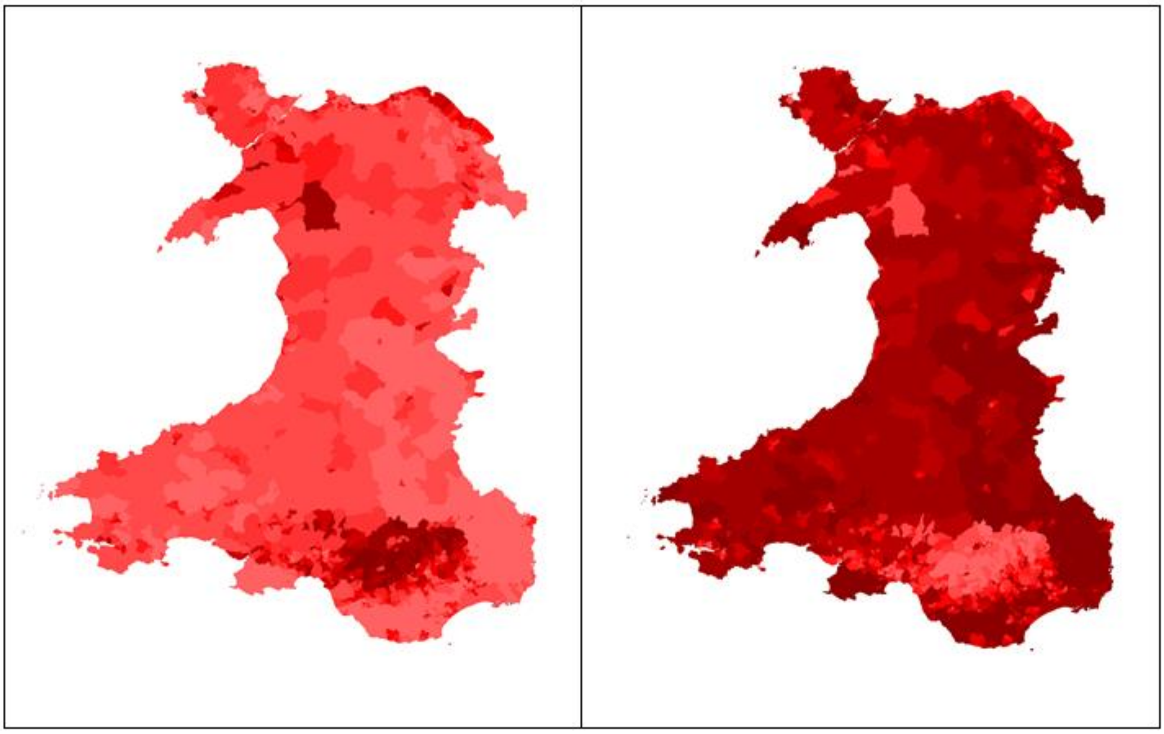


Figure 4.10 Clustering and ranking of features using SOM based site ranking tool

4.6 Impact Assessment

Impact assessment is another important group of functionalities included in SDSS. Once indicators are identified and the potential sites have been selected, it is important to carry out the impact assessment of engineering interventions. Impact assessment has been considered for different domains such as socio-economic, environmental health and road traffic impact etc. Impact assessment can be carried out separately or it can feed into the decisions on final site selection. The impact assessment section contains the tools related to prediction and impact assessment of sites that can be used to facilitate the decision making process. It contains three tools: i) GRNN based prediction tool, ii) RIAM based impact assessment tool and iii) Traffic impact assessment tool.

4.6.1 GRNN based prediction tool

Geographical General Regression (GRNN) based prediction tool utilises the capabilities of the GRNN for prediction and regression. It is a useful tool when the relationship between dependant and independent variables is unknown and complex. It supports both linear and nonlinear relationships.

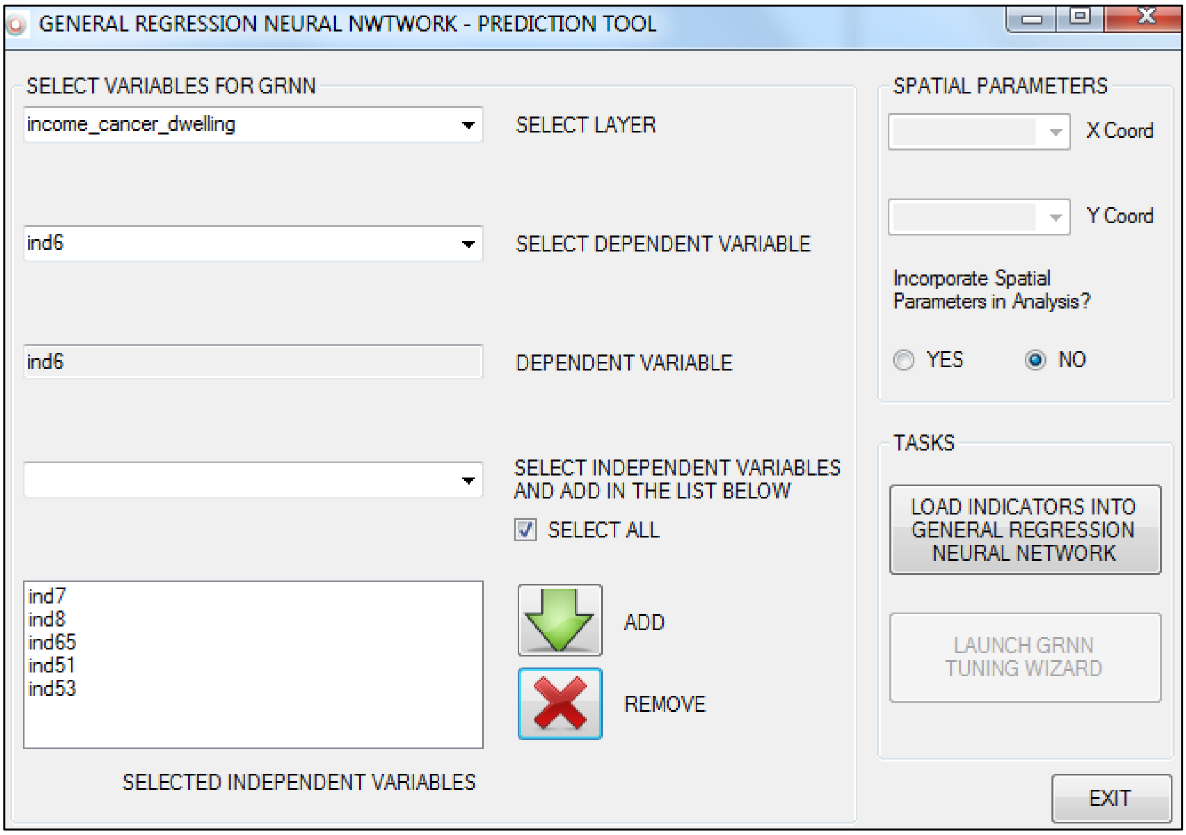


Figure 4.11 GUI of GRNN based prediction and regression analysis tool

Figure 4.11 shows the user interface of the GRNN based prediction tool. The user first selects the GIS layer containing the indicators. Any combination of indicators can be added as a GIS layer in the SDSS using the geodatabase management tools. Once layer is selected, user identifies the dependent and independent variables, and loads the data into the GRNN tool.

This tool provides a novel technique by using spatial data in the GRNN. This is achieved by considering the spatial distances between different geographical regions as one of the variable in the analysis. The spatial distance is used for the prediction of dependent variable in the same way as the difference between other dependent variables as explained in Section 4.3.2. The distance between geographical regions is calculated by calculating the Euclidean distance between their centroids.

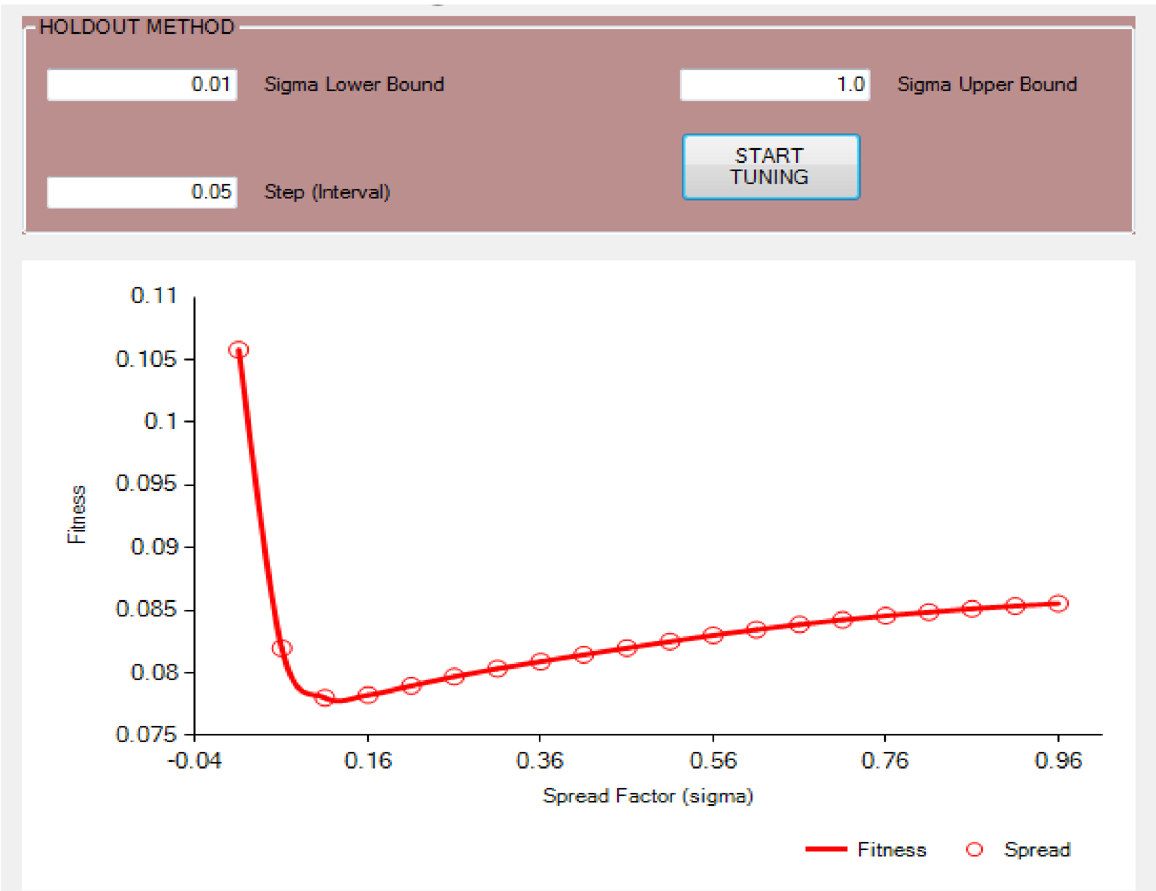


Figure 4.12 GUI of SIGMA optimization tool using Holdout Method

The user can select whether or not to use the spatial distance in the analysis. Once the data is incorporated in the neural network, a tuning wizard is launched helping the user to select best sigma (σ) parameters for the analysis. The tuning wizard utilises the Holdout Method for the calculation of the Root Mean Square Error (RMSE). User can give a range (upper and lower bound) for sigma parameters and a step (interval) to calculate the RMSE using Holdout method. The system plots the RMSE values against the corresponding sigma spread factors as shown in Figure 4.12. This helps in finding the best sigma value for the analysis.

The tool also allows the user to enter manually, different sigma values or use the genetic algorithm approach for the identification of a set of sigma values with least possible RMSE. Figure 4.13 depicts the GUI of the GRNN tuning wizard.

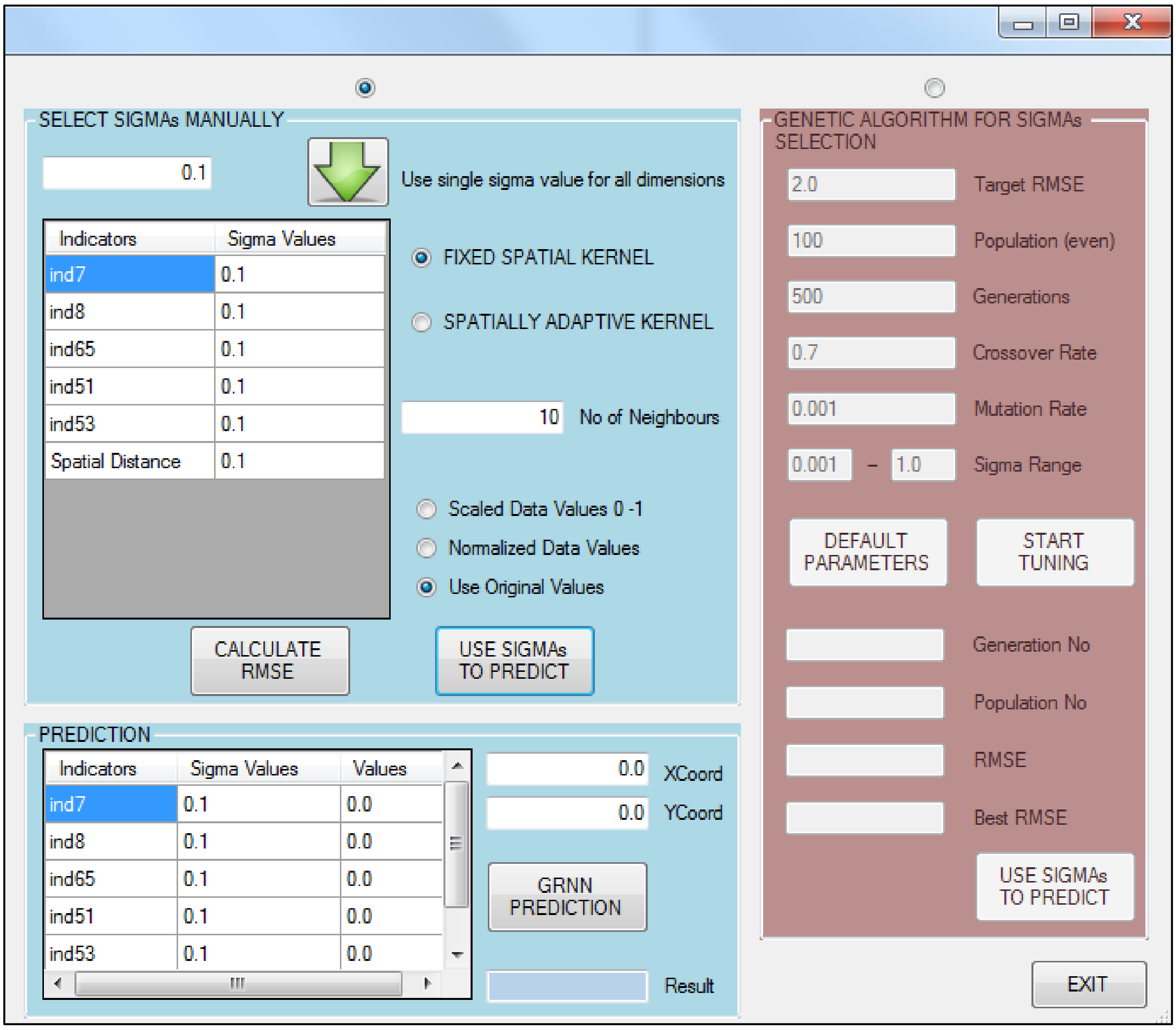


Figure 4.13 GUI for sigma selection and prediction

Either the actual, scaled or normalised data values can be used for the calculation of RMSE for a given set of σ. User can assign the same σ parameter for all the independent variables if the data is normalised or scaled. However, if the original data values of the independent variables are used for the estimation of the dependent variable, then it is important to assign the sigma values with care. This is important as some of the variables may have a different spread and range of data values as compare to the others and using a similar sigma value can affect the results.

Adopting spatial parameters in the regression analysis in GRNN is similar to the Geographically Weighted Regression (GWR) suggested by (Fotheringham et al. 2003). If spatial parameters are included in the analysis then the tool provides two different methods to identify a specific number of neighbouring geographical features to be used for the prediction analysis. These two methods are a) Fixed spatial kernel and b) Adaptive spatial kernel.

If an adaptive spatial kernel is selected, only a given number of neighbouring geographical features are selected for the analysis. This option should be used when the indicators have more localised regression effect. The system first calculates the distance of each geographical feature from the prediction point. Then only N closest neighbours are selected and used in the process. N is number of the neighbours to be incorporated and it is provided by the user. The sigma parameter is then used to find the influence of each feature to contribute in the prediction analysis. The closer features within the set of N nearest neighbours will have more influence than those distant apart.

If the indicators have more globally smooth regression effect, then fixed spatial kernel should be used (Fotheringham et al. 2003). If fixed spatial kernel is used then spatial distance is between all the geographical features and the prediction point is calculated. The smoothing parameter sigma used for the spatial dimension is then used for the identification of influence each feature will have in the prediction analysis. The influence region will depend on the sigma parameter and closer features will contribute more in the prediction than those distant apart. A very large value of sigma parameter used for the spatial parameter, may include the entire study area in the prediction analysis. A very small value of sigma parameter will result in the use of immediate neighbours in the prediction analysis.

Once a set of sigma parameters have been selected with acceptable RMSE value, the user can select to use them for the actual prediction at unknown location. If spatial parameter were not used in the analysis, only the independent variables need to be provided by the user at the unknown location, where prediction is to be made for the dependent variable. If however, spatial parameters were used, then user should provide the X and Y coordinated of the centroid of the geographical feature, for which the dependent variable is being predicted.

4.7 Conclusions

The development aspects of the SDSS have been discussed in this chapter in particular those analytical modules that utilise Artificial Intelligence (AI) techniques including Artificial Neural Network (ANN) and Genetic Algorithm (GA). Analytical modules of the system are divided in four sections namely a) Site Selection and Ranking, b) Impact Assessment, c) Spatial Knowledge Discovery and d) Geodatabase Management.

Some of the analytical modules utilise Artificial Intelligence (AI) techniques such as a) Self-Organizing Maps (SOM), b) General Regression Neural Networks (GRNN) and c) Genetic Algorithm (GA). The structure, mathematical background and algorithms of these AI techniques are first explained in the chapter. The development of analytical modules utilising such AI techniques is then explained. SOM (one-dimensional) is utilised in the site ranking tool and the clean correlation finding tool. GRNN has been utilised in the prediction and regression analysis tool. GA tool has been incorporated for the identification of essential parameters for the GRNN based analysis.

The Graphical User Interfaces (GUI) of the analytical modules is also described in the Chapter to understand the essential parameters required to carry out respective analysis. Some novel techniques have been introduced in the development of the analytical modules in this research, such as:

- The SOM based ranking module utilises the ordering capability of the one-dimensional SOM to cluster and rank geographical areas (or sites) in terms of the appropriateness of the key indicators used in the analysis.

- The GRNN based module incorporates a novel way of incorporating spatial dimension in the analysis. In this way the distance between geographical features is also considered as one of the dependant variables for the prediction of dependent variable. This can be useful in understanding the spatial colocation and local variations found in the regression analysis.

References

Bowden, G. J.,Nixon, J. B.,Dandy, G. C.,Maier, H. R. and Holmes, M. 2006. Forecasting chlorine residuals in a water distribution system using a general regression neural network. *Mathematical and Computer Modelling* 44(5–6), pp. 469-484.

Cigizoglu, H. K. and Alp, M. 2006. Generalized regression neural network in modelling river sediment yield. *Advances in Engineering Software* 37(2), pp. 63-68.

Corona, F. and Lendasse, A. 2005. Input Selection and Regression using the SOM.

DotSpatial. 2014. *DotSpatial Library* [Online]. Available at: http://dotspatial.codeplex.com/ [Accessed: October 14].

Fotheringham, A. S.,Brunsdon, C. and Charlton, M. 2003. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Wiley.

Kohonen, T. 1998. The self-organizing map. *Neurocomputing* 21(1-3), pp. 1-6.

Kohonen, T. 2001. *Self-Organizing Maps*. Third ed. Springer Berlin Heidelberg.

Kohonen, T. 2013. Essentials of the self-organizing map. *Neural Networks* 37(0), pp. 52-65.

Kohonen, T. and Somervuo, P. 2002. How to make large self-organizing maps for nonvectorial data. *Neural Networks* 15(8–9), pp. 945-952.

Leung, M. T.,Chen, A.-S. and Daouk, H. 2000. Forecasting exchange rates using general regression neural networks. *Computers & Operations Research* 27(11–12), pp. 1093-1110.

Polat, Ö. and Yildirim, T. 2008a. Genetic optimization of GRNN for pattern recognition without feature extraction. *Expert Systems with Applications* 34(4), pp. 2444-2448.

Polat, Ö. and Yildirim, T. 2008b. Hand geometry identification without feature extraction by general regression neural network. *Expert Systems with Applications* 34(2), pp. 845-849.

Specht, D. F. 1991. A general regression neural network. *Neural Networks, IEEE Transactions on* 2(6), pp. 568-576.

Vesanto, J. and Ahola, J. eds. 1999. *Hunting for correlations in data using the self-organizing map*. Proceeding of the International ICSC Congress on Computational Intelligence Methods and Applications (CIMA'99).

5

SDSS DEVELOPMENT

MCDA BASED ANALYTICAL MODULES

5.1 Introduction

This chapter covers the developmental aspects of the SDSS, in particular those analytical modules that utilise Multicriteria Decision Analysis (MCDA) techniques. The functionality of these modules is explained in flow charts, mathematical formulations, figures and tables. Graphical User Interfaces (GUIs) are explained and user interaction with the system is discussed. Each module is explained separately including its working and functional aspects, analytical background, essential inputs parameters and outputs.

As explained in Chapter 3, modular approach has been adopted to develop the system. In order to provide required functionality, adequate tools have been developed. Based on these different functionalities, analytical modules are grouped together in four sections: a) Site Selection and Ranking, b) Impact Assessment, c) Spatial Knowledge Discovery and d) Geodatabase Management. The analytical modules can be used separately or in conjunction with each other. A typical use case scenario of the SDSS can based on utilising each module, sequentially to facilitate the decision making process.

As explained in Chapter 4, the main interface of the SDSS is developed initially, to provide a foundation for these analytical modules. Some of the analytical modules utilise the artificial intelligence based soft computing techniques including Artificial Neural Network (ANN) and Genetic Algorithm (GA). These analytical modules are covered in detail in Chapter 4.

The Site Selection and Ranking section contains the most important tools of the SDSS to assist in the site selection and site ranking analysis and it is explained in detail in Section 5.4. The Impact Assessment section contains the tools for impact assessment and prediction analysis. These tools are useful for informed risk based decision making and it is explained in detail in Section 5.5. The Spatial Knowledge Discovery Subsystem contains the tools to assist decision makers in identifying the relationship between different indicators in the given geographical space. It can also be used for the selection of most appropriate indicators to be used in the analysis. It is explained in details in Section 5.2. The Geodatabase Management Subsystem contains the tools to load spatial data into the SDSS from geodatabase and it is explained in details in Section 5.3.

5.2 Spatial Knowledge Discovery

Spatial Knowledge Discovery subsystem has tools to facilitate the decision makers in knowledge extraction from the data. As explained in Chapter 2, data driven approach is very useful in informed risk based decision making. The tools provided in this subsystem can be used for analysis related correlation, regression and spatial-colocation of different indicators in a given geographical space. Also, the tools provided in this subsystem can facilitate the decision makers for the selection of most appropriate indicators. The selection of an appropriate set of variables is the key to the reliability and effectiveness of the analysis.

There are two analytical modules developed in the SDSS to facilitate the spatial knowledge discovery: a) SOM based clean correlation tool and b) Parallel coordinates plotting tool. The SOM based tool is explained in Chapter 4. Parallel coordinate plot tool is explained in section 5.2.1. The GRNN tool can also be utilised in finding the most useful set of variables that can be used in the analysis. For example, GRNN has been used by (Bowden et al. 2006) for the determination of the most appropriate variables to forecast chlorine in preventing the spread of waterborne diseases.

5.2.1 Parallel coordinates plot tool

Parallel Coordinates Plot (PCP) is another tool for exploratory data analysis that has been included in the tools for Spatial Knowledge Discovery section of the SDSS. Spatial data representing the real world (natural and man-made) features or phenomenon are called thematic maps. Traditional techniques for the visualisation of thematic maps e.g. Choropleth maps, can only represent one phenomenon (variable) on a map at a time.

PCP is a useful way of analysing two of more variables together. The PCP tool developed in the SDSS can be used for: a) visualising how different variables are correlated to each other, b) visualising how different variables are clustered together in a given geographical area and c) identifying the peculiar values of the variables different from normal patterns (Andrienko 2001). The PCP tool can also commensurate the variables if they are not in the same units of measurements. PCP tool has a simple GUI by which user can load a GIS layer from the main interface of the SDSS. Any number of attributes associated with this layer can be selected for plotting. Figure 5.1 shows the GUI of the PCP tool.

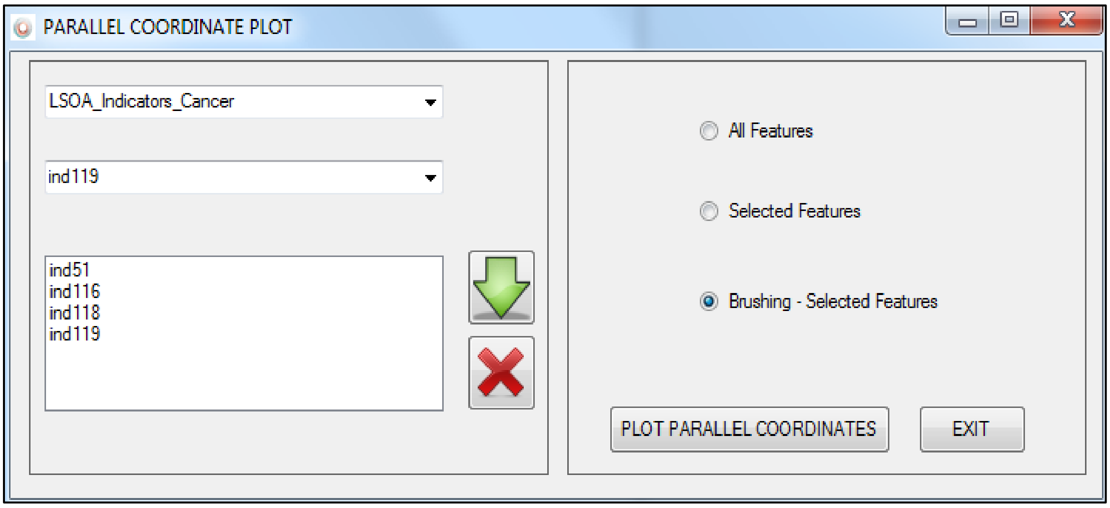


Figure 5.1 GUI of Parallel Coordinate Plot tool

PCP tool also provides function of brushing of selected features on the layer. Brushing is a technique used for highlighting a certain part of the data to make it more prominent than the rest of the data. This is helpful if a decision maker is interested in knowing the relationship of different variables in a given geographical region as compare to the entire area.

Figure 5.1 shows the interface for the selection of GIS layers, variables and plotting parameters. A GIS layer with necessary indicators (attributes) is first loaded into the SDSS. Some geographical areas are selected on the map. Then brushing option is selected and the variables are plotted on the parallel coordinates. The selected features on the map are highlighted in the plot to compare with rest of the features.

PCP tool uses the Microsoft chart control for plotting (Microsoft 2014) . User can change the type of chart to Line, Spline or Step-Line types (Microsoft 2014). Figure 5.2 shows the plot generated by the PCP tool. X-Axis represents the indicators selected for plotting and Y-Axis represents the scaled values (commensurate) of all the geographical features for each indicator.

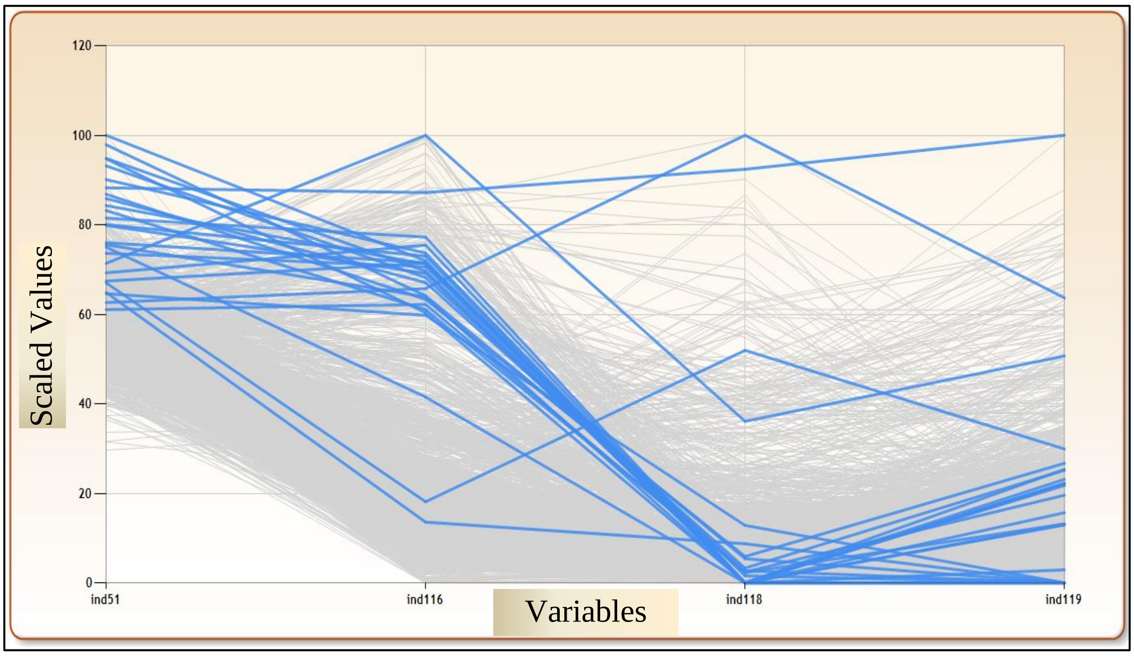


Figure 5.2 GUI of PCP with brushing of selected features

The selected features are shown in darker colour to distinguish them from the rest (Brushing technique). The plot depicts the general trend of data in the entire study areas and the relative position of selected features with respect to it.

5.3 Geodatabase Management

The geodatabase management subsystem provides a link between the backend spatial and non-spatial data and the frontend SDSS. Data available in Shapefiles can be directly loaded in the SDSS using the "Load Data" tool available in the tool-strip on the main interface. However, Shapefiles have their own limitations as discussed in Chapter 3. Therefore, an open source geodatabase called SpatiaLite (SpatiaLite 2014) has been incorporated in the SDSS for data storage, retrieval and query.

The geodatabase management subsystem offers different tools to interact with the spatial and non-spatial data stored in the SpatiaLite based geodatabase. These are a) View indicators tool, b) Load indicators tool and c) Load GIS layers tool. DotSpatial library has a plugin for connecting SpatiaLite to access spatial data (SpatiaLite 2014). This library has been extended to be used in the SDSS Geodatabase management tools. Functionalities of these three tools are explained in the following sections.

5.3.1 View indicators tool

There are various datasets that can be loaded from the geodatabase into SDSS. These are explained in details in Chapter 6 and used in the application of the SDSS in Chapter 7. These datasets can have some associated metadata information. It is important to understand this information in order to utilise these datasets appropriately. The view indicators tool included in the system can provide all the information about the datasets.

5.3.2 Load Indicators tool

The load indicators tool provides a user friendly graphical user interface (GUI) to select a combination of required indicators and load them into the SDSS. Indicators are divided into four different themes: i) Public Health Indicators, ii) Socio-Economic Indicators, iii) Environment Indicators and iv) Geo Technical Indicators. The Welsh Index of Multiple Deprivation (WIMD) covers more than one domain and therefore it is kept as a separate group. The GUI of load indicators tool is shown in Figure 5.3.

The user can load any combination of these indicators into a single or multiple layers for each domain. User can also select different scales, at which the indicators are available. These scales are a) Lower Super Output Areas (LSOA) level, b) Medium Super Output Areas (MSOA) level, c) Local Authority (LA) level or d) Fishnet (500×500m). The first three represent administrative or demographical divisions in Wales. Whereas, Fishnet is a grid of cells (500×500m) generated over the on onshore areas of Wales to combine all the datasets together for this research. All the datasets required for the site selection tools are combined together in this layer. This is explained in detail in Chapter 6.

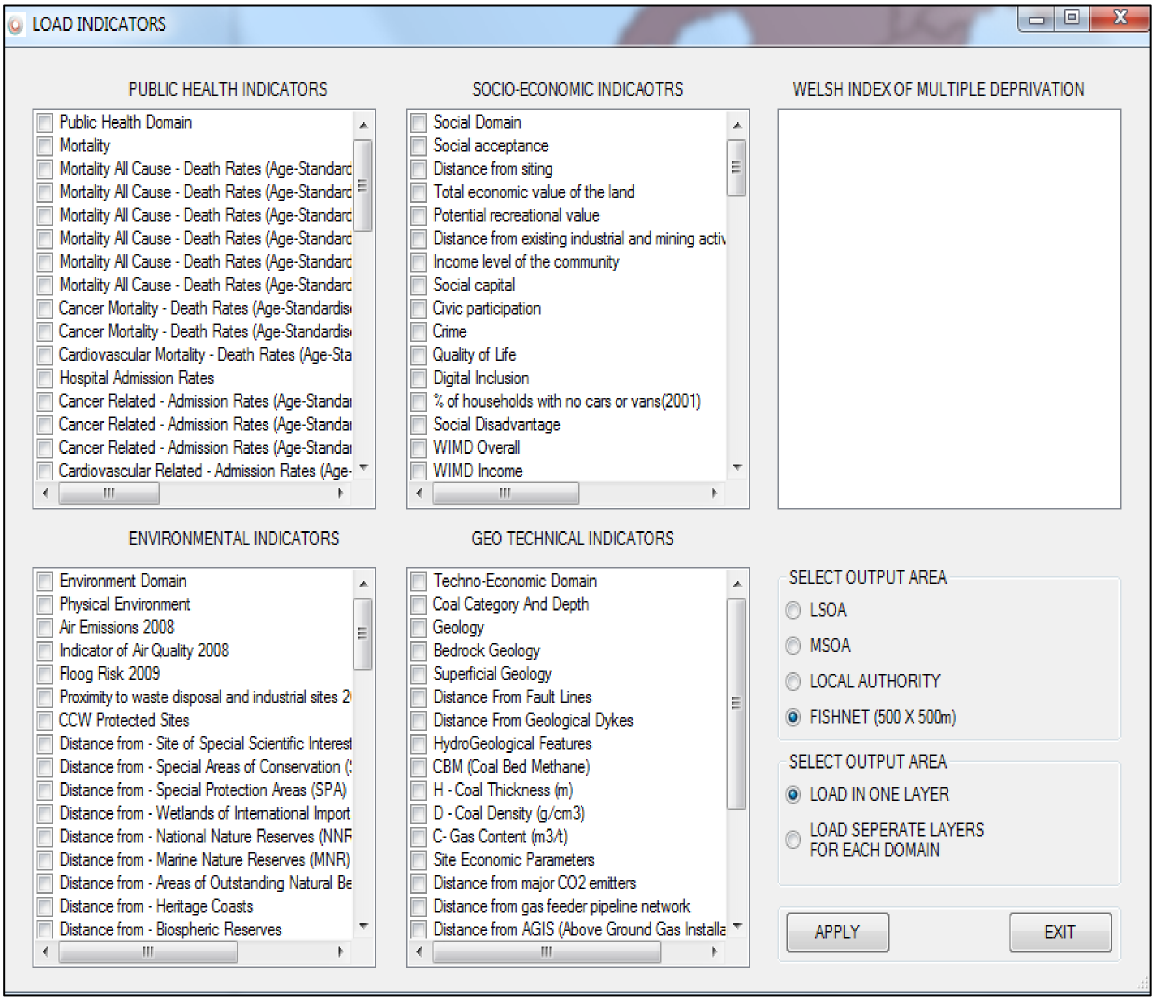


Figure 5.3 GUI of the Load Indicators tool

5.3.3 Load GIS layers tool

The final tool in the geodatabase management subsystem is to load the SpatiaLite GIS layers in the SDSS. Layers are stored in the SpatiaLite geodatabase as tables. Each row in the table represents a geographical feature, such as a point, line or polygon. The geometry is stored in the geometry column of the table as a Binary Large Object (BLOB). The attribute information linked to the geographical features is stored in other data columns within the same table. The details of how this combination of spatial and non-spatial information is stored within the tables can be found at (SpatiaLite 2014).

5.4 Site Selection and Ranking

The site selection and ranking section has one of the key analytical modules developed in the SDSS. There are three different tools in this subsystem: a) AHP based site selection tool, b) SOM based site ranking tool and c) Site ranking by neighbourhood analysis tool. The SOM based ranking tool has been discussed in detail in Chapter 4.

5.4.1 AHP based site selection tool

As discussed in Chapter 2, Analytical Hierarchy Process (AHP) can be used to facilitate the decision making process in a variety of application areas such as healthcare, environment, finance and government etc. AHP can be used in situations where a direct and established empirical relationship is unknown between dependent and independent variables. The concept of AHP is also used when multiple options are available to choose from but there is no direct ranking available to help the decision making process.

A generalized structure of the AHP tree is shown in Figure 5.4. The tree structure in this case consists of an overall goal to be achieved. This goal is dependent on multiple objectives and sub-objectives, which are finally dependent on the variables (independent variables) at the leaf level (lowest level) in the hierarchy tree. AHP allows the decision maker to assign different relative weights to objectives, sub-objectives and variables at different levels of the decision hierarchy as shown in Figure 5.4.

The relative weights highlight the importance of one input over the others in the analysis. Each parent node (in the decision tree hierarchy) is calculated from the values and relative weights of the child nodes it is dependent on. This structure of the decision tree can consist of several levels and depends on the complexity of the problem (Malczewski 1999).

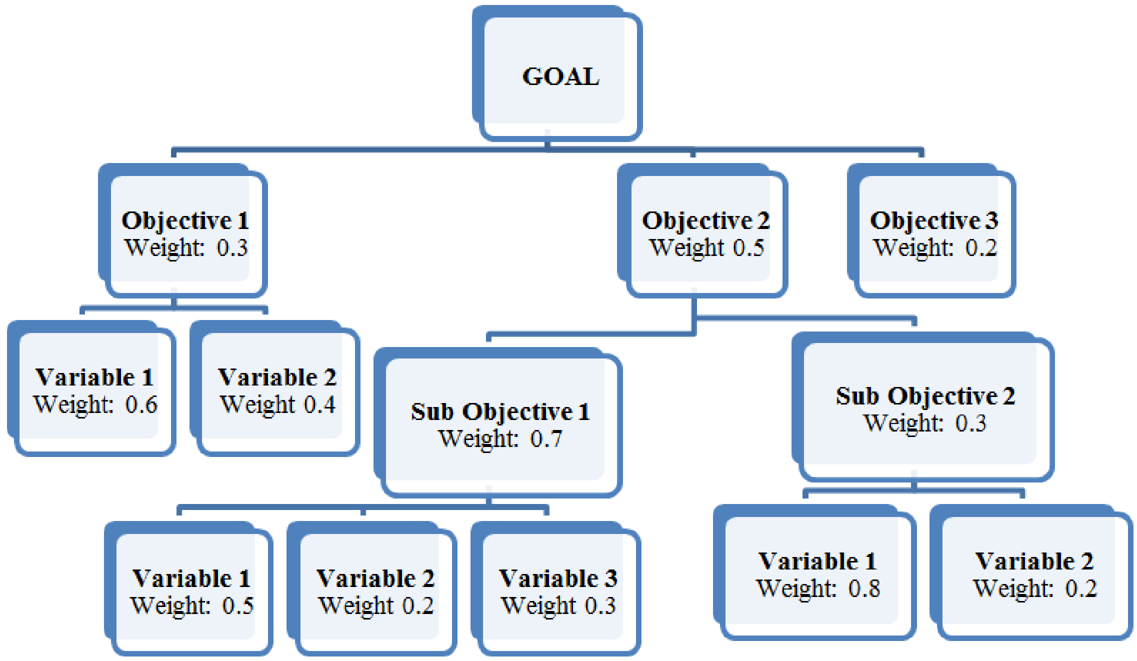


Figure 5.4 Analytical Hierarchy Process with Weighted Linear Combination

5.4.1.1 Commensuration of data

As discussed above the actual data is normally stored in the leaf level nodes which are placed at the bottom most level of the decision hierarchy tree. The indicators (variables) used in the AHP process can be in different units of measurement and it is important to first commensurate all of them between 0-1 where 0 is the least suitable and 1 is the most suitable.

The decision maker can use the commensurate tool to scale all the variables between 0 and 1, or can use the original units if all the data is in the same units. Scaling can be done by using either a) Maximum Score Procedure or b) Score Range Procedure. The commensurate tool shown in Figure 5.6 provides all these options to the user for scaling the data. The nature of a criterion can be either positive or negative. User defines whether a particular indicator is Benefit (the more, the better) or Cost (the less, the better) in nature. During the scaling process, an appropriate equation is used by the tool to cater for cost and benefit nature of individual criterion. If Maximum Score Procedure is selected by the user for scaling, Equations 5.1 and 5.2 are used for benefit and cost criterion respectively (Malczewski 1999).

$$\text{(Benefit)} \quad X'_{ij} = \frac{X_{ij}}{X_{j.max}} \tag{5.1}$$

$$\text{(Cost)} \quad X'_{ij} = 1 - \frac{X_{ij}}{X_{j.max}} \tag{5.2}$$

In decision environments, where both 'Benefit' and 'Cost' criteria exist together, above equations cannot be used together because of their different bases. In such cases 5.2 becomes (Malczewski 1999):

$$\text{(Cost)} \quad X'_{ij} = \frac{X_{j.min}}{X_{ij}} \tag{5.3}$$

The advantage of this method is that it's a proportional linear scaling of the raw data. The largest value of each variable always gets the score of 1 and smaller values get lower scores accordingly. If Score Range Procedure is selected for scaling, Equations 5.4 and 5.5 are used for benefit and cost criterion respectively. In this case the smallest value of each variable always gets the score of 0 whereas higher values get higher score accordingly (Malczewski 1999).

$$\text{(Benefit)} \quad X'_{ij} = \frac{X_{ij} - X_{j.min}}{X_{j.max} - X_{j.min}} \tag{5.4}$$

$$\text{(Cost)} \quad X'_{ij} = \frac{X_{j.max} - X_{ij}}{X_{j.max} - X_{j.min}} \tag{5.5}$$

Where X_{ij} is the value of the i^{th} location (potential site) for the j^{th} criterion. X'_{ij} is the scaled (standardized) value of X_{ij}. $X_{j.min}$ and $X_{j.max}$ are the minimum and maximum values of the j^{th} variable in the entire dataset (Massam 1988; Malczewski 1999).

5.4.1.2 Pairwise Comparison Method

The pairwise comparison method has been developed in the context of analytical hierarchy process as a comparative procedure to assign relative weights (Saaty 1980). In pairwise comparison method the decision maker has to assign a relative importance of a variable (between 1- 9) over every other peer variable (placed in the same tree level) for the calculation of the parent node (in the level above).

Decision makers normally use verbal judgments in order to emphasise the relative importance of variables with each other. Saaty (1980) also suggested a lookup table to convert these verbal judgments or preference into the equivalent numerical values. Table 5.1 provides the conversion scheme as suggested by Saaty (1980). 1 means the two variables are equally important and none has any priority over the other. The highest numerical value is 9 which mean that one variable is extremely important compare to the other.

Table 5.1 Preference lookup table of the Pairwise Comparison Method (Saaty 1980)

Numerical rating	Verbal judgement or preference
1	Equal importance
2	Equal to moderate importance
3	Moderate importance
4	Moderate to strong importance
5	Strong importance
6	Strong to very strong importance
7	Very strong importance
8	Very to extremely strong importance
9	Extreme importance

One of the advantages of using Pairwise Comparison Method is that the decision maker doesn't have to think about the relative weights of all the variables at the same time. Rather the process is carried out by comparing only two variables at a time Saaty (1980). Once all the relative preferences have been provided by the user, the system then produces relative weights (sum of which will be equal to 1) to be assigned to the indicators.

The relative weights are based on human judgement and can be inconsistent. There is an empirical way of checking this consistency. It mathematically checks the consistency ratio of the weights provided by the user (Saaty 1980). Following steps are taken to calculate the consistency ratio. The procedure used here for the calculation of consistency ratio is the same as explained by Saaty (1980) sited in (Malczewski 1999) where further details can be found.

Step 1- Create a pairwise comparison matrix by assigning relative importance of indicators over each other.

Step 2- Create the normalised pairwise comparison matrix by dividing each element in the matrix by the sum of its column in the pairwise comparison matrix. Then compute the average of each row by summing up all the elements in the normalised pairwise comparison matrix and then dividing it by the number of variables use.

Step 3- Calculate the weighted sum vector by multiplying the weight of first column in the original pairwise comparison matrix with the weight of the first criterion, second column with the weight of second criteria and so on.

Step 4- Determine the consistency vector by summing the values (created in step 3) over rows and then dividing it with the criterion weights.

Step 5- Calculate the value of lambda (λ) by taking the average value of the consistency vector determined in step 4.

Step 6- Calculate the value of Consistency index (CI) by using:

$$\mathrm{CI} = \frac{\lambda - \mathrm{n}}{n-1} \tag{5.6}$$

where n is total number of variables used in the pairwise comparison matrix.

Step 7- Finally, calculating the term Consistency Ratio (CR) which can be calculated by dividing the CI with Random Index (RI). A reference Random Index Table is also provided by (Saaty 1980). RI depends on the number of variables used in the pairwise comparison matrix.

Once consistency ratio is calculated for a given set of relative weights, then it can be determined whether or not the scheme is consistent. If the consistency ratio is less than 0.10, the ratio indicates an acceptable consistency. If it is more than 0.10 then it shows that the judgments are inconsistent and relative weights should be re adjusted (Malczewski 1999).

5.4.1.3 Weighted Linear Combination

Weighted Linear Combination (WLC) is often used with AHP in solving multicriteria spatial problems as discussed in Chapter 2. When is applied in conjunction with AHP, each parent node in the decision tree is calculated by multiplying the criterion maps (child nodes) by their relative weights and then summing up this product. This process starts from the lower most level of the decision tree where criterion maps are placed. The process continues upwards in the decision tree until the Goal (top node in the AHP decision tree) is processed (Moeinaddini et al. 2010). As described earlier, there is a weight associated with each of these nodes that shows its relative importance in the calculation of the parent node in the decision hierarchy tree. The score (weighted sum) for each node is calculated by using (Malczewski 1999):

$$A_i = \sum_j w_j x_{ij} \tag{5.7}$$

Where A_i is the suitability index for the ith location (Fishnet cell in this research) for a given evaluation, Wj is the relative importance weight of criterion j in the evaluation, x_{ij} is the value of the ith location for jth criterion. Sum of w_j is always equal to 1 for a given evaluation and assigned by the user (e.g. using pairwise comparison method).

5.4.1.4 AHP and WLC based spatial multicriteria analysis in geodatabase

An AHP and WLC based spatial multicriteria analysis routine has been developed within the geodatabase. This tool can be used to solve various different types of spatial multicriteria problems. There are several advantages of implementing the spatial multicriteria analysis routine within the geodatabase such as:

- Simple SQL queries can be used to multiply the variables (columns) of each feature (rows) with the corresponding weights assigned by the decision makers. The results are saved in the resultant columns.
- User doesn't have to manage several GIS layers (Raster and Vector) and the secondary files produced during the implementation process of AHP at the application level.
- User can save the weights in the database and compare the results with other combination of weights combinations from different decision makers.
- Individual objectives and sub-objectives can be seen as a layer in the system at any time as they are stored as separate columns.
- The overall processing time is much less than what it takes for processing WLC and AHP analysis in the file-based GIS layers. This is because each layer is stored as a table in the geodatabase instead of multiple files structure.
- If the weighting scheme and structure of the tree is unchanged, the user can retrieve the results at any level of the tree without reprocessing.
- Spatial filters and constraints can be applied along with the spatial queries to carry out the processing only in the desired regions. This can save substantial amount of processing time.

The details of the geodatabase developed, are covered in detail in Chapter 6.

5.4.1.5 Graphical User Interface

A user friendly interface is designed for the AHP based site selection tool. Decision maker can process any particular domain, i.e. Socio-Economic, Public Health, Environmental and Techno-Economic to explore the most suitable areas. Combined effect of all these domains can also be aggregated to find the most suitable sites in terms of multiple criteria associated with all four domains.

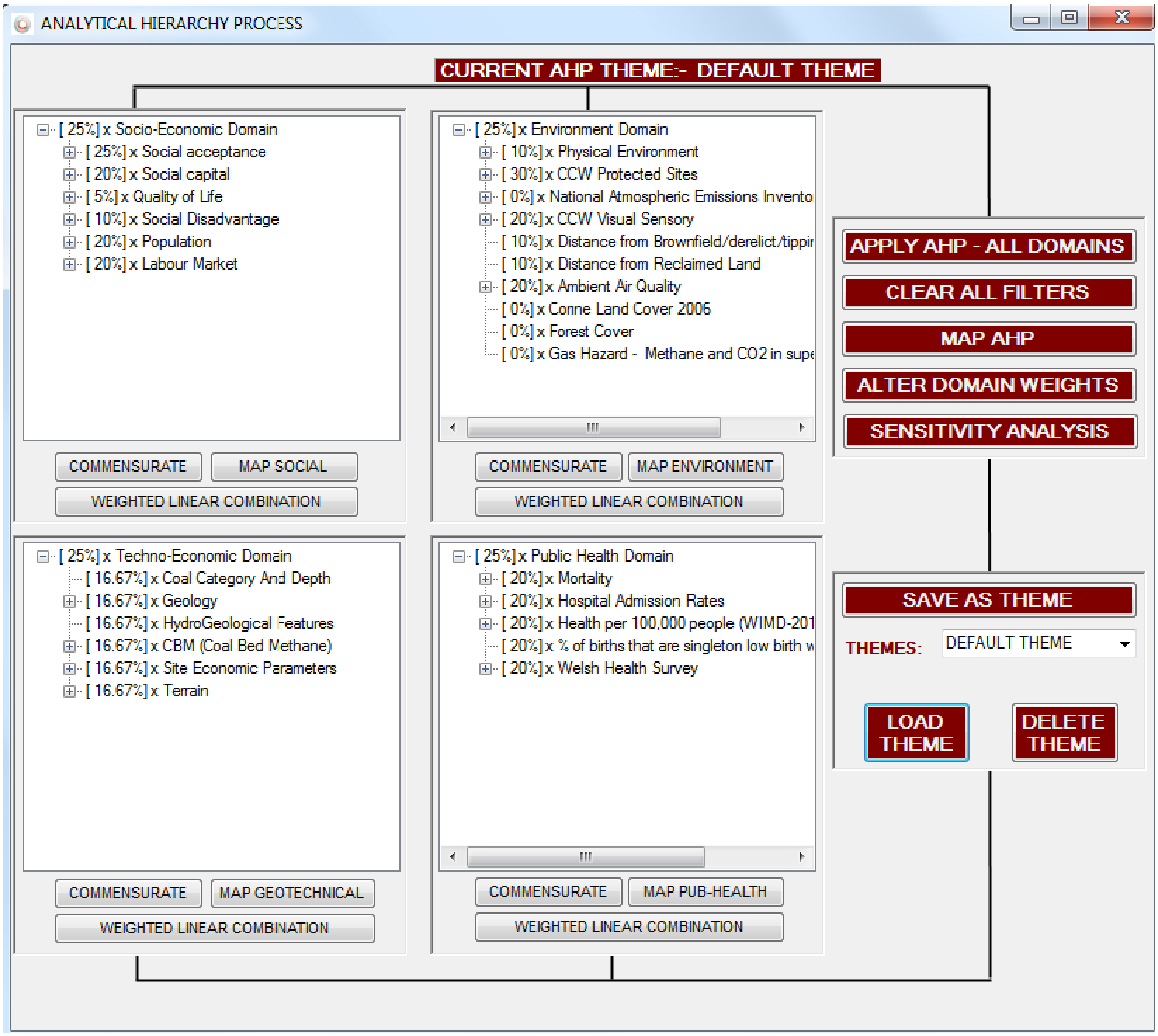


Figure 5.5 GUI of the Analytical Hierarchy Process based site selection tool

To commensurate (scale) the variables into the same currency, a commensurate tool is provided in the site selection tool as shown in Figure 5.6. User can either select the Score Range Process or the Maximum Score Procedure to scale the data. If the units of all the variables are same then the process can be applied without scaling.

The user should also select the Cost of Benefit nature of each variable. This information is necessary for the system to use an appropriate equation (5.1 to 5.5) to scale a variable.

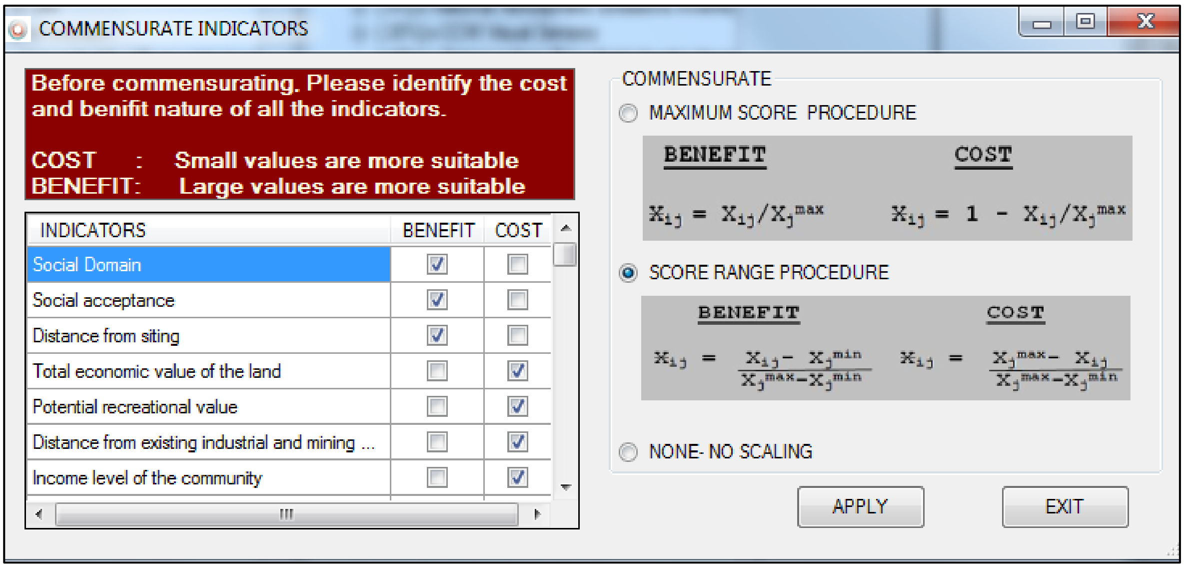


Figure 5.6 GUI for the commensuration of indicators

User can also change the relative weights of nodes at different levels of the AHP decision tree, i.e. objectives, sub objectives, indicators, sub-indicators using the weight assigning tool provide in the AHP based site selection tool. The GUI of the weight assigning tool is shown in Figure 5.7.

Two different weighting mechanisms are provided to user to choose from. User can either assign the weights directly (sum of which should be equal to one). Or the user can utilise the pairwise comparison method provided in the tool, to let the system calculate the relative weights of the different entities with precision. Pairwise comparison method is explained in Section 5.4.1.2. The user assigns a relative importance scheme to the indicators in use and the system calculates the relative weights from it and also suggests whether or not the weights can be used with confidence.

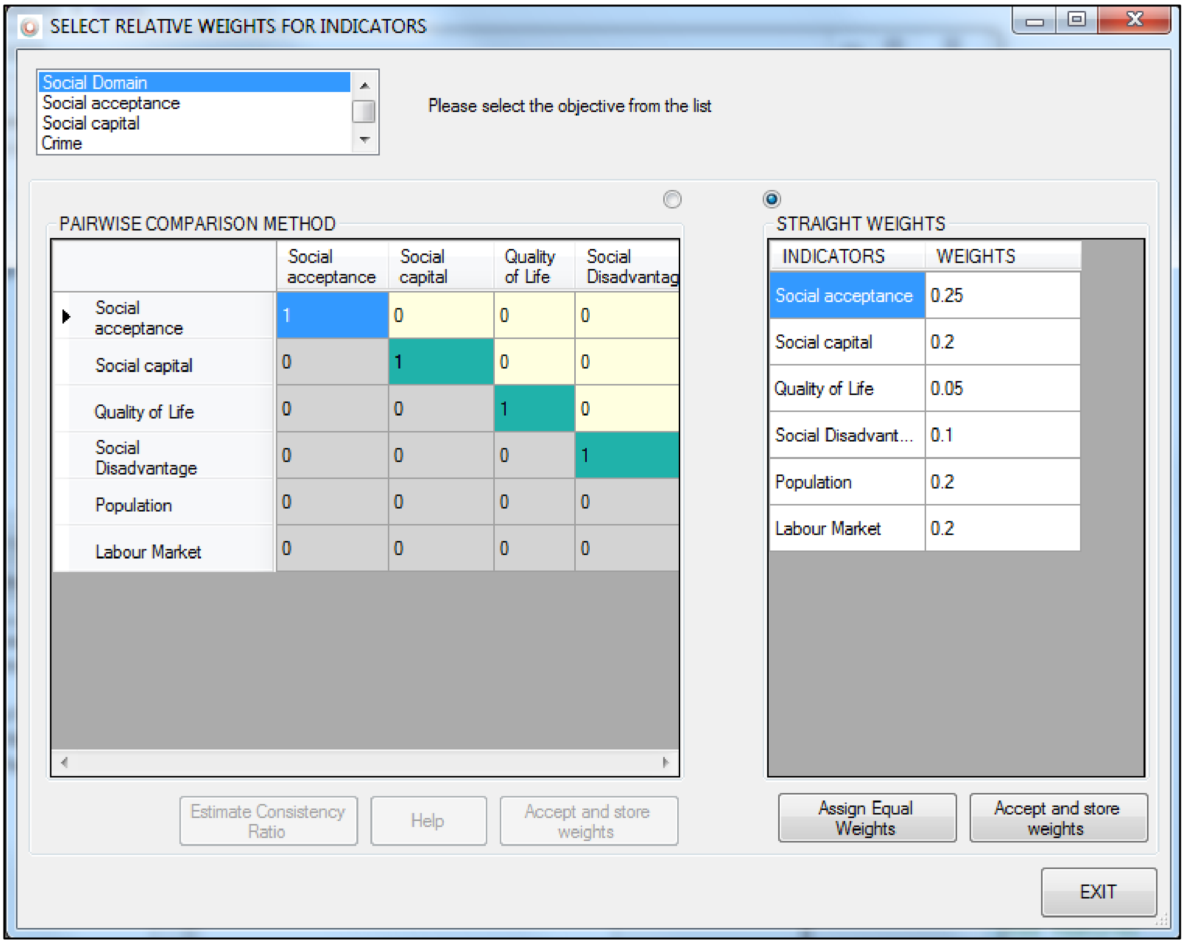


Figure 5.7 GUI for the selection of relative weights for indicators

The mechanism for the assigning of weight is different for continuous (quantitative) and discrete (qualitative) indicators. For quantitative (continuous) variables, the relative weight is assigned directly to the indicator (criterion map).

The qualitative (discrete) indicators are treated in a different way compared with the quantitative (continuous) indicators. An individual weight is assigned to each class of the qualitative indicator. For example, hydrogeology layer can have discrete values of hydrogeological features, e.g. "productive aquifer", "non-productive aquifer". Some of these hydrogeological classes will be preferred over others by the decision makers, hence a relative weights is assigned to each class separately. The sum of all the weights assigned to a qualitative variable is always equal to 1. The weights can be assigned using the qualitative weight assigning tool provided with the AHP based site selection tool as shown below.

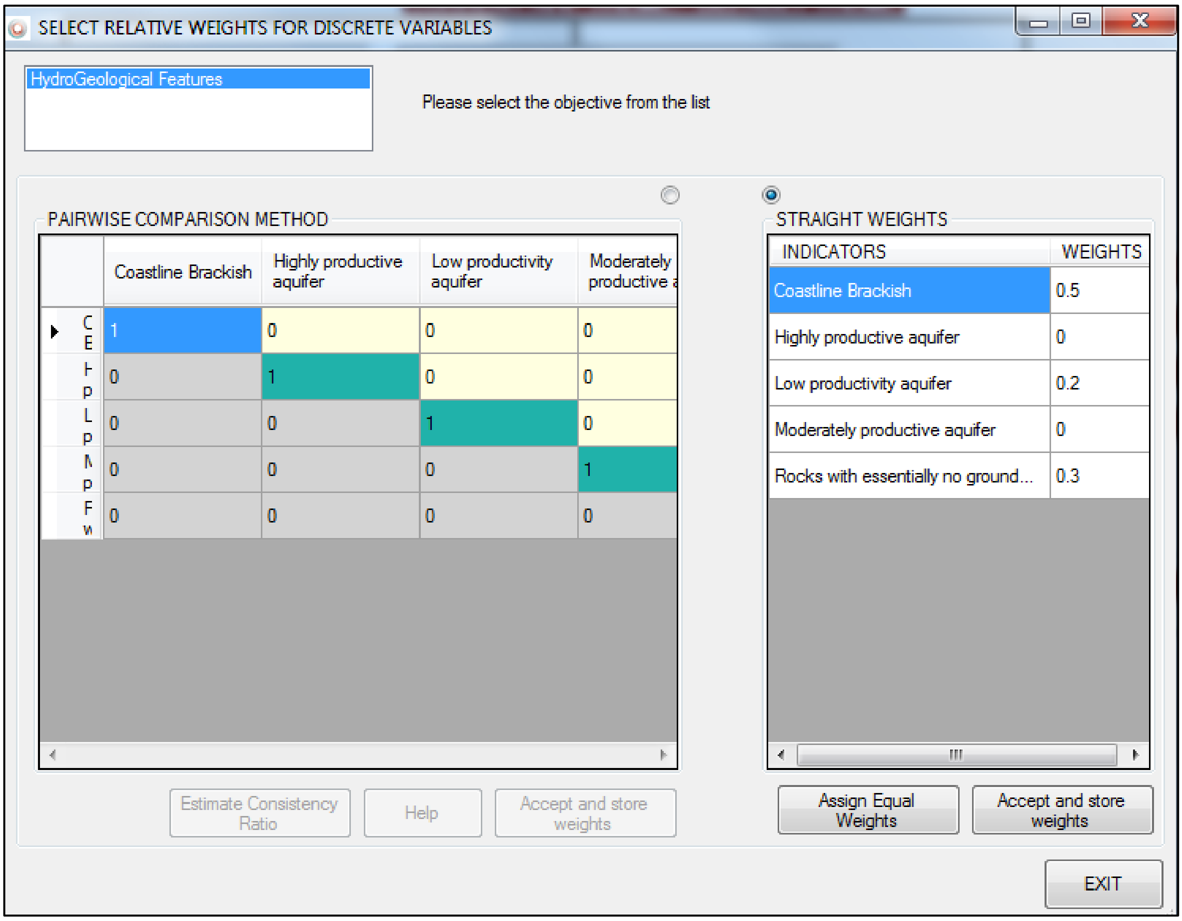


Figure 5.8 GUI for the selection of relative weights for different classes of discrete (qualitative) indicators

Once all the indicators are scaled and relative weights assigned, the user can apply the AHP at the Geodatabase level and the corresponding columns are updated with the results. These weights and results are saved in the database for future reference to allow the comparison of results. The user can see the results in the form of a map in the map viewer of the SDSS.

User can apply AHP to the entire region or filter out the regions not feasible for the site selection in order to reduce the computational load and in order to refine the results. This can be done by applying the constraints on any variable using the constraint tool shown below.

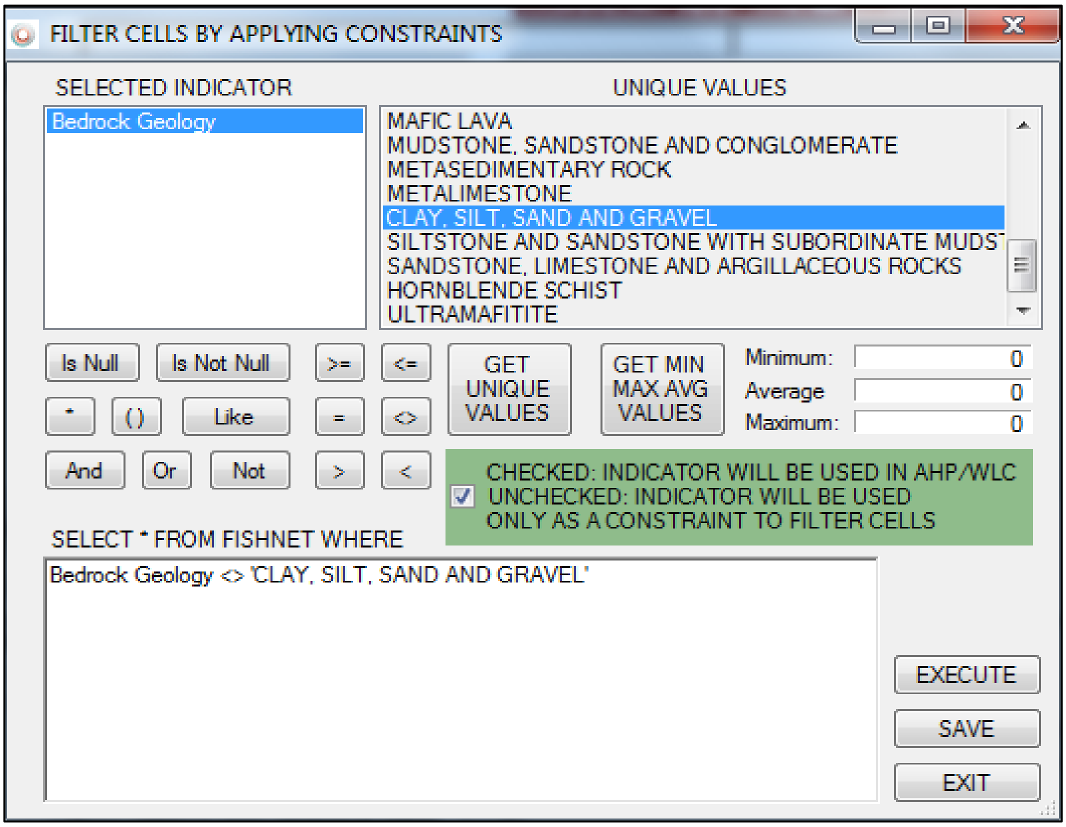


Figure 5.9 GUI for applying constraints and filters for the site selection process

The constraint tool provides a user friendly way of filtering out any regions, that the decision maker would not like to include in the AHP process. The user can also select any particular variable not to be used in the AHP process at all. The user can see the unique data values of the selected indicators. If the selected indicator is quantitative, useful information about the range of the data values is also shown, i.e. the minimum, maximum and average values. If the selected indicator is qualitative, unique data values (classes) are listed to be selected.

5.4.1.6 AHP analysis themes

Although there are many indicators provided in the SDSS as explained in Chapter 6, specific problem may need only a subset of the indicators and selected geographical region. Different decision maker would assign different weights to the participating indicators and use the filters and constraints as per their own requirements. The decision maker can save all this information in the form of a theme in the geodatabase. Using this capability a considerable amount of time and processing can be saved, if the same analysis has to be repeated with only a few changes in the essential parameters. User can load a particular AHP theme from the database, modify it or remove it permanently. A default theme is loaded with all the indicators to be used to start a new analysis. Themes can be managed from the main interface of the AHP based site selection tool as shown in Figure 5.5.

5.4.1.7 AHP results

The tool provides a useful way of visualising the results of AHP process. WLC can be applied at four different domains separately or as batch processing of all four together. The results of WLC can be viewed at any level of the AHP tree structure, i.e. overall goal, individual domains or indicators. As described in the Section 5.4.1.5, all the numerical calculations are carried out at the geodatabase level using SQL queries. The results of WLC are stored in the temporary columns at each level. Using this, the decision maker can explore the results at any level without running the process every time. The original values of the base indicators are stored and not changed during this process. If user changes any parameters, e.g. relative weights, the process can be run again to reflect the changes in results and AHP theme can be updated.

5.4.1.8 Visualisation of result

AHP resultant map can be customised in different ways using the map filter, benchmarking and symbology tool as shown in Figure 5.10. This tool provides three different options to customise the resultant maps which include i) Data categories and unique values, ii) Percentage filter and iii) Suitability benchmarking.

First option is to symbolise the map based on its underlying attribute data. User can select either an equal interval with a given number of classes or unique data values, to classify the map symbology. First option divides the attribute data into equal intervals and a colour ramp is used to show these intervals on the map. If unique values option is selected then a different colour is assigned to each unique data value on the map.

The Percentage filter option can be used to view only a percentage of the data on map, e.g. top 5 % of areas in terms of elevation or bottom 10 % of areas in terms of number of cancer patients. This is a useful tool for viewing the areas containing high or low data values for a given parameter. This tool is particularly useful in viewing a base indicator at the leaf level of the AHP tree rather than the composite indicators, sub-domains, domains or overall goal. For these higher levels, suitability benchmarking technique is useful.

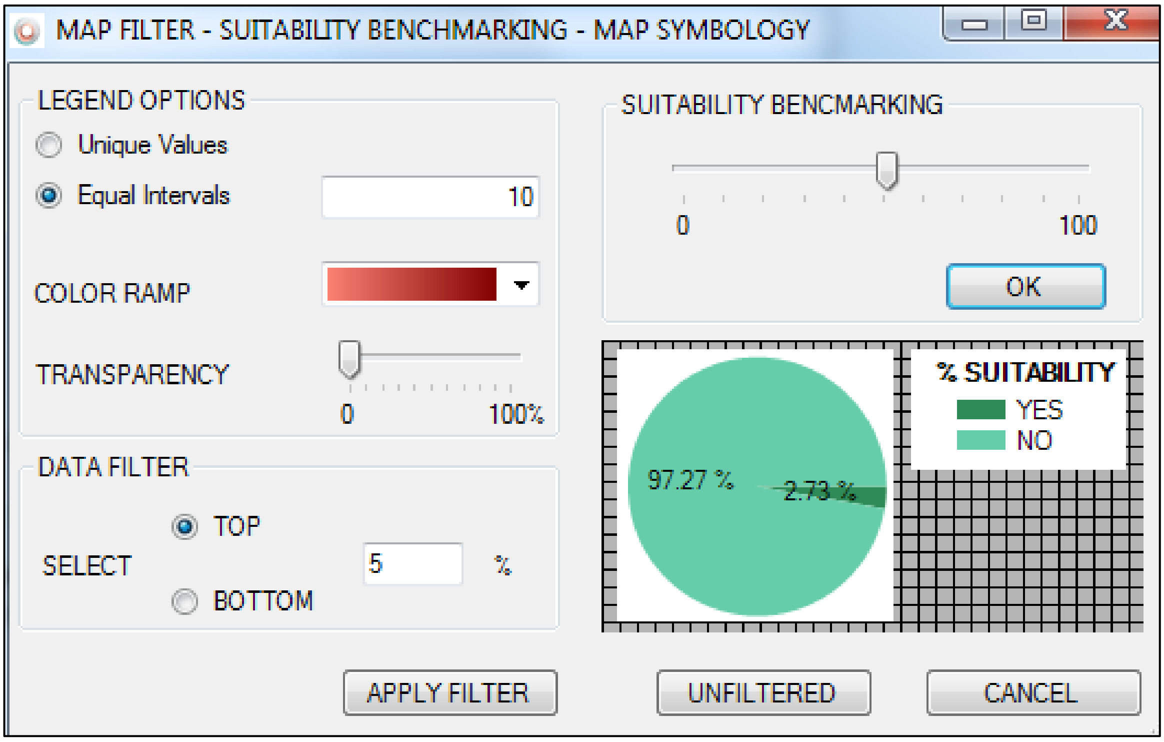


Figure 5.10 GUI for suitability benchmarking, data filter and legend options

Suitability benchmarking can be used to view only a certain number of features having values larger than the set benchmarking value. The indicator values are first commensurate (scaled) between 0-1 (virtually 0-100 for ease). System then reads only those cells (features) from the geodatabase that are above the benchmarking value and show the results on map. Figure 5.11 shows an example of the suitability benchmarking where elevation layer is scaled and the benchmarking is set to 50. The resultant cells are shown in 10 equal intervals in red colour ramp on the map. Pie chart shows that only 2.73% of the entire study area has the elevation that is more than the benchmark value.

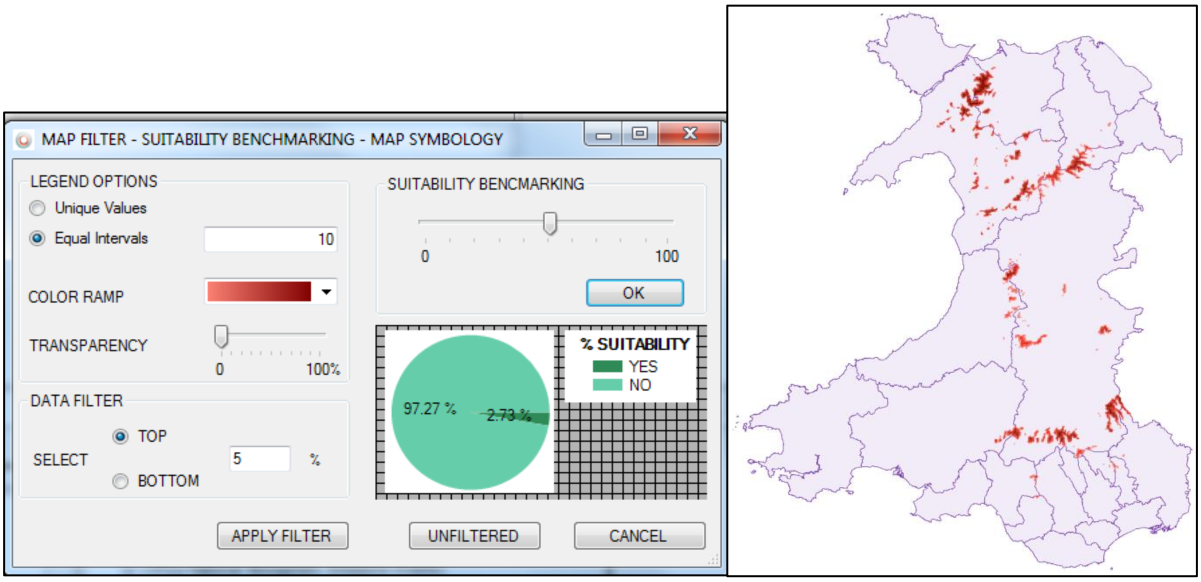


Figure 5.11 Suitability benchmarking example on elevation data

5.4.1.9 Sensitivity analysis

The sensitivity of the AHP based site suitability analysis is mostly associated with the user assigned relative weights. These criterion weights are essential for the analysis. The criterion weights are based on the decision maker's choice on the relative importance of different indicators. Therefore the decision results may be sensitive to these relative weights and hence it is a known drawback of AHP (Nefeslioglu et al. 2013; Feizizadeh et al. 2014). If the suitability of alternatives is very sensitive to this change in the relative weights, then the relative weights should be examined carefully (Malczewski 1999).

In order to assess the sensitivity of the relative weights in AHP hierarchy, a useful sensitivity analysis tool has been developed and incorporated in the AHP based site selection tool. The interface of sensitivity analysis is shown in Figure 5.12. Firstly, user selects a level in AHP at which the sensitivity analysis is to be performed. It can be performed at any level of the AHP tree, provided that the node has some child nodes at the level below in the decision tree. The user sets the essential parameters to run the tool, i.e. the sliding weight, direction of weight change (positive or negative or in both directions) and the benchmark value of the parent node to calculate the percentage change.

The tool calculates the area (Fishnet cells) on the map that fulfils the set benchmarking criteria. The system then increases or decreases the weights of the selected entity and slightly adjusts the weights for the other entities in the same level of the decision tree and having the same parent entity. This is to ensure that sum of all the relative weights is always equal to 1.

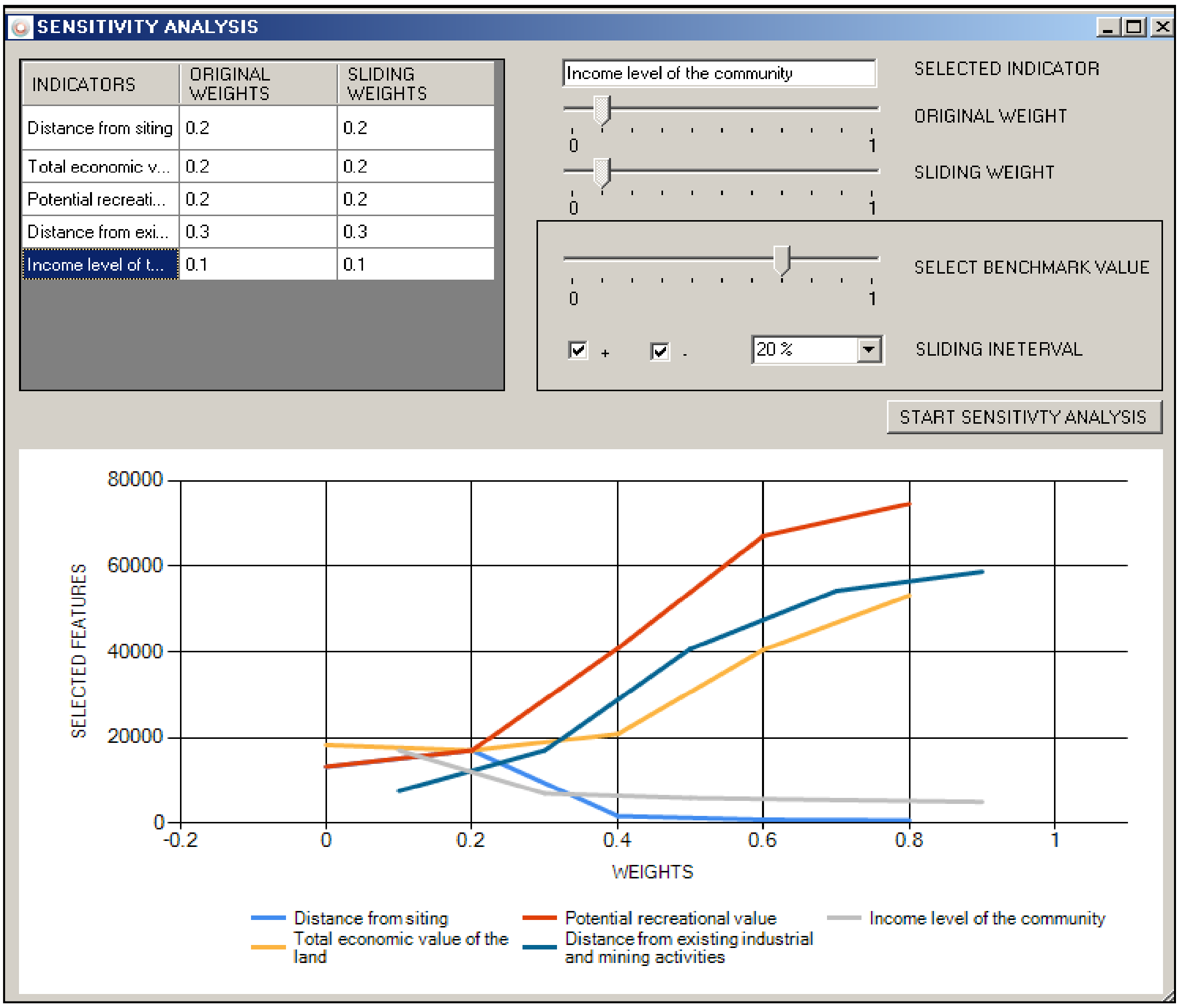


Figure 5.12 GUI for Sensitivity Analysis of the AHP based site selection tool

The percentage of the study area (Fishnet cells) that is above the set benchmark values is calculated repeatedly, every time the relative weights are modified. The process continues until the weight of the selected entity becomes 0 or 1, depending on the sliding direction (negative or positive). The process also stops if the change of weight cannot be adjusted in other nodes to make the sum equal to 1.

Visual inspection of the resultant graphs gives an understanding of the sensitivity of the decision process associated with the relative weights of the indicators. Any abrupt change in the graph with the modification of the relative weights, suggests the decision makers about the sensitivity of the relative weights being assigned to the indicators.

5.4.2 Site ranking by neighbourhood analysis tool

The tool provides a systematic mechanism to rank the potential sites based on the neighbourhood analysis and comparison. This tool can be used along with the AHP to refine the site selection process. AHP based site selection tool can produce a large area or multiple sites which have equal potential. At this stage the siting decision is mainly based on the choice and expert knowledge of the decision maker(s). To facilitate the decision making in the second phase, under the conditions described above, site ranking using neighbourhood analysis can be very useful. Using this tool, inputs required from the decision maker and the risks associated with personal judgement and choice can be minimized. The following steps are adopted in the tool to analyse the information and generate the results:

The tool uses two GIS layers as input from the user. One layer holds all potential sites whereas the other layer presents the key socio-economic and environmental health indicators in the neighbourhood of the potential sites. The tool then scales all the indicators between 0-1(where 0 is the minimum value and 1 is the maximum value of each indicator in the layer). During the commensuration process, the user also defines whether a particular indicator is Benefit (the more, the better) or Cost (the less, the better) in nature. The user can select either the original values or scaled values of the data to be used in the analysis. For scaling, the tool provides Maximum Score Procedure and Score Range Procedure by using an appropriate equation (5.1 to 5.5).

The tool then selects the neighbouring areas of each potential site based on criteria specified by the user. This is done by applying buffers around the candidate sites and selecting the intersecting regions in indicator layer. The tool calculates the minimum, maximum and average values of each indicator in the selected neighbourhood of each site. Either of these average, maximum or minimum values can be selected to rank the sites.

The final step is to assign ranks to the sites. User can choose the ranking either based on Criterion Sorting Mechanism (CSM) or the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) method. CSM is a novel method introduced here and it is based on the ordering (sorting) and ranking of geographical regions based on each attribute (Criterion) and then a cumulative rank is also generated based on all the attributes. Figure 5.13 presents the Geographical User Interface (GUI) of the tool. The user assigns the buffer radius in map units to define the neighbourhood of each potential site. The tool first generates the buffer polygons around each site and then the second layer containing indicators is intersected with these buffer polygons. The tool can rank the sites based on the average, maximum or minimum value of each indicator in the given surrounding regions of each site.

In order to compare the results of CSM technique, TOPSIS method is also incorporated within the tool. TOPSIS method is incorporated in the tool to compare the results of the CSM ranking method, since TOPSIS is one of the most commonly used ranking methods in MCDA problems as discussed in Chapter 2.

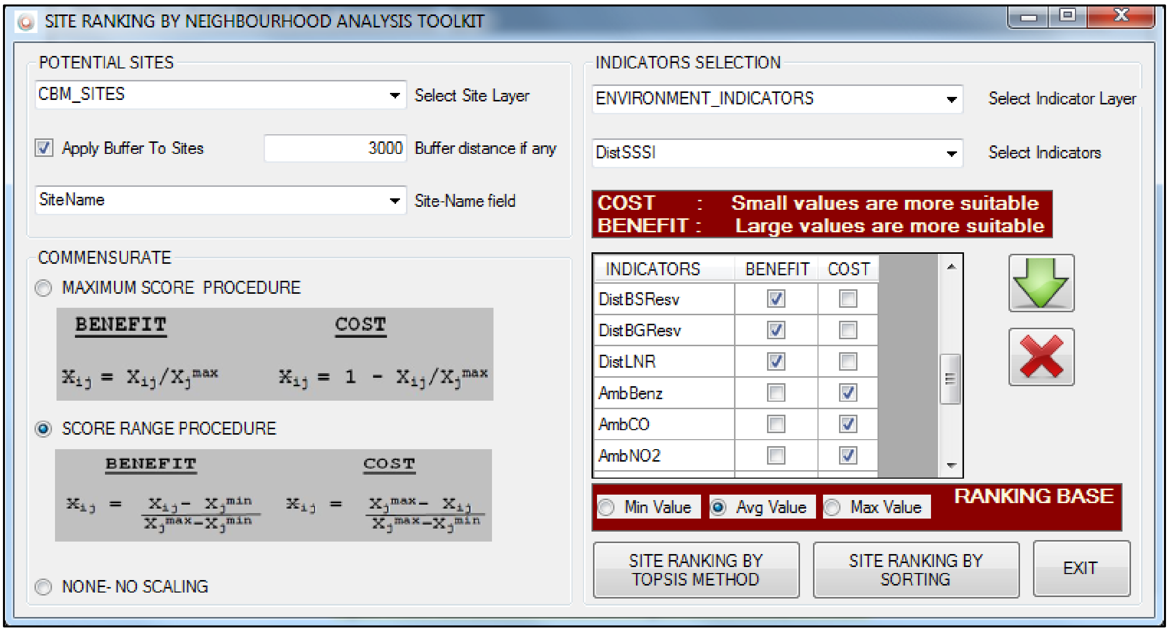


Figure 5.13 GUI of the site ranking by neighbourhood analysis tool

5.4.2.1 Mathematical formulation

Criterion Sorting Mechanism (CSM) is a novel method introduced here and it is based on the ordering and ranking of geographical regions based on each attribute (Criterion). A rank is assigned to the geographical region based on the average, minimum or maximum value of a given attribute (Criterion). Since all the attributes are scaled to the same currency, i.e. between 0-1, therefore it is possible to assign an overall rank.

The overall rank is based on the individual ranks assigned to each criterion. Each site may obtain different ranks for different indicators. A cumulative rank is then constructed by the tool using CSM. For this purpose, a rank sum is constructed for each site by summing up individual ranks of all the indicators. Sites are sorted in ascending order in terms of this rank sum. The site with lowest rank sum gets the overall Rank 1. Figure 5.14 shows the steps taken in the algorithm of CSM to rank sites.

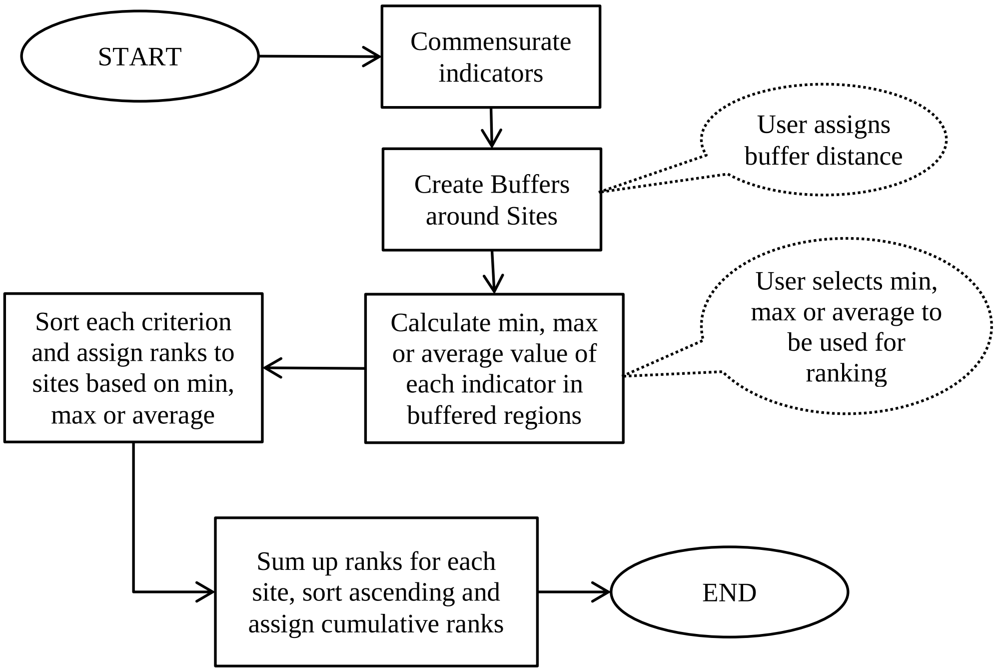


Figure 5.14 Flow chart of CSM based site ranking

In order to compare the results of CSM technique, TOPSIS method is also incorporated within the tool. TOPSIS is selected because it is a commonly used ranking method in MCDA problems (Chen et al. 2011) and (Jia et al. 2012). It ranks the sites based on their distances from the most ideal and the least ideal solution (Hwang and Yoon 1981). If TOPSIS method is used, the cumulative site ranks are constructed from the average, minimum or maximum values using the empirical formulation of TOPSIS method provided in (Hwang and Yoon 1981).

The results are generated in the form of a report containing charts and a table, which provide the ranks of each site with respect to the indicators. Also they present the cumulative rank produced by CSM and TOPSIS.

5.5 Impact Assessment

Impact assessment is another important group of functionalities included in SDSS. Once indicators are identified and the potential sites have been selected, it is important to carry out the impact assessment of engineering interventions. Impact assessment has been considered for different domains such as socio-economic, environmental health and road traffic impact etc. Impact assessment can also assist in the decisions on final site selection. The impact assessment section contains the tools related to prediction and impact assessment of sites that can be used to facilitate the decision making process. It contains three tools: i) GRNN based prediction tool, ii) RIAM based impact assessment tool and iii) Traffic impact assessment tool. GRNN based prediction tool has been explained in Chapter 4 whereas RIAM and Traffic impact assessment tools are explained below.

5.5.1 RIAM based site impact assessment tool

The Rapid Impact Assessment Matrix (RIAM) is a semi-quantitative way of executing the EIA in the form of a structured matrix containing the subjective judgements of the EIA assessors. The graphical form of RIAM can be useful in assessing the subjective and quantitative judgements with clarity and rapidly as compare to other traditional methods of EIA which are more qualitative in nature. RIAM can organize, analyse and present the results of EIA rapidly (Pastakia and Jensen 1998).

5.5.1.1 RIAM mathematical formulation

The RIAM bases site impact assessment tool utilises the same mathematical formulations of the original RIAM method as suggested in (Pastakia 1998), (Pastakia 1998). In RIAM method, each aspect of the project is evaluated against the environmental components. The environmental components are divided into four major categories, i.e. a) Physical/Chemical, b) Biological/Ecological, c) Social/Cultural and d) Economics/Operational. The Physical/Chemical (PC) components cover all the physical and chemical aspects of the environment, e.g. the impact of civil works on existing infrastructure and the long term effect of the project on ground water quality etc. The Biological/Ecological (BE) covers all the biological and ecological aspects of the environment, e.g. the effect of noise and pollution on biodiversity in the area and the fragmentation of key habitats etc. The Social/Cultural (SC) covers all the human aspects of the environment, e.g. culture, social acceptance, job creation and cohesion of the communities etc. The Economics/Operational (EO) covers all the financial and economic aspects of the project e.g. initial cost, profits, financing and future investments (Pastakia and Jensen 1998). A score is assigned to each component falling under any of the four categories, to construct an overall matrix. The individual score is calculated by evaluating each component against the following two criteria:

Group A. Criteria that are of importance to the condition, that individually can change the score obtained.

Group B. Criteria that are of value to the situation, but should not individually be capable of changing the score obtained.

Individual components that fall under Group A, are more important compared to those in Group B, therefore the scoring system is slightly different for both the groups. The score of individual components in Group A, is multiplied together to emphasize the weights of each component in the overall score. On the other hand the score of individual components in Group B is added together. This ensures that the collective importance of Group B is incorporated in the calculation of the overall score but without over influencing the overall score. The total score for A and B can be calculated as (Pastakia and Jensen 1998):

$$aT = (a1) \times (a2) \tag{5.8}$$

$$bT = (b1) + (b2) + (b3) \tag{5.9}$$

where $(a1)\ and\ (a2)$ are the individual scores for the components in Group A and $(b1)$, $(b2)$ and (b3) are the individual scores for the components in Group B. The overall assessment score (ES) is then calculated by the product of total score of Group A and Group B (Pastakia and Jensen 1998):

$$ES = (aT) \times (bT) \tag{5.10}$$

Group A and B have further categories in them, with specific scale values. Each project component is evaluated against these categories and then the given scale is used to construct the environmental assessment score (ES) of the component using Equations 5.8, 5.9 and 5.10.

Table 5.2 Assessment criteria – Adapted from (Pastakia 1998)

Criteria	Scale	Description
A1: Importance of condition	4	Important to national/international interests
	3	Important to regional/national interests
	2	Important to areas immediately outside the local condition
	1	Important only to the local condition
	0	No importance
A2: Magnitude of change/effect	3	Major positive benefit
	2	Significant improvement in status quo
	1	Improvement in status quo
	0	No change/status quo
	−1	Negative change to status quo
	−2	Significant negative dis-benefit or change
	−3	Major dis-benefit or change
B1: Permanence	1	No change/not applicable
	2	Temporary
	3	Permanent
B2: Reversibility	1	No change/not applicable
	2	Reversible
	3	Irreversible
B3: Cumulative	1	No change/not applicable
	2	Non-cumulative/single

	3	Cumulative/synergistic

Once the overall environmental assessments score (ES) has been calculated for individual project components, a band can also be assigned to it. A reference lookup table is provided by Pastakia (1998) to assign an impact assessment band. These band range values are given in Table 5.3. These bands are very useful in the comparison of different project components according to their sensitivity to environment. These bands can be counted for each project component or each candidate site. A site with more positive bands will have less environmental impact than those carrying more negative bands.

Table 5.3 Look up table for Environmental Scores and Range Bands – As in (Pastakia 1998)

Environmental	Range	Description of Range Bands
+72 to +108	+E	Major positive change/impacts
+36 to +71	+D	Significant positive change/impacts
+19 to +35	+C	Moderately positive change/impacts
+10 to +18	+B	Positive change/impacts
+1 to +9	+A	Slightly positive change/impacts
0	N	No change/status quo/not applicable
−1 to −9	−A	Slightly negative change/impacts
−10 to −18	−B	Negative change/impacts
−19 to −35	−C	Moderately negative change/impacts
−36 to −71	−D	Significant negative change/impacts
−72 to −108	−E	Major negative change/impacts

The process of RIAM based site impact assessment is explained in Figure 5.15. The main interface of the tool provides a matrix based results viewer where each added component can be viewed with its overall score (ES) and individual scores for A1, A2, B1, B2 and B3 criteria as described in Table 4.3. The tool also calculates the cumulative score and range band for each component. User can also save the overall RIAM scheme in the geodatabase as themes similar to the AHP based site selection tool.

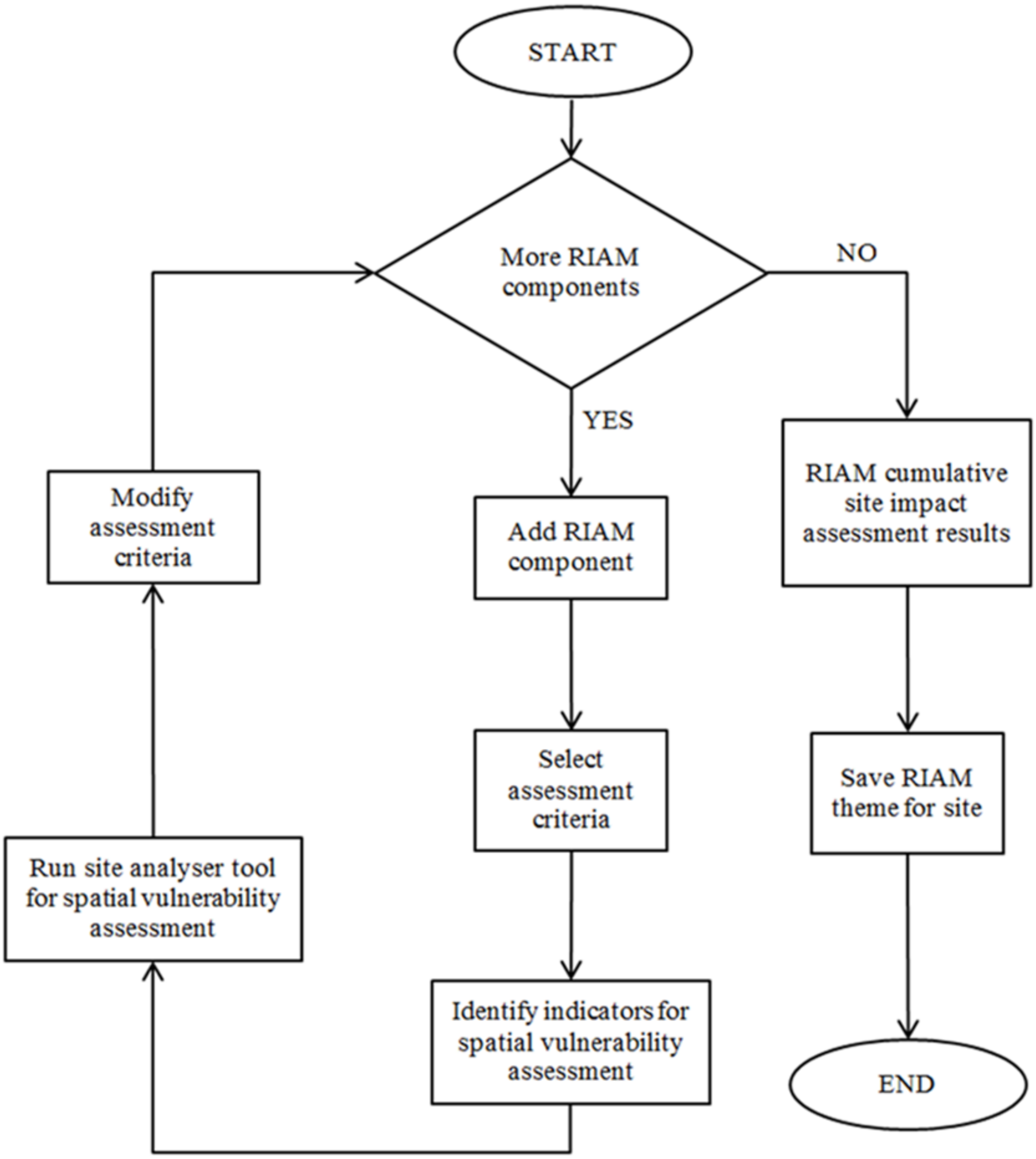


Figure 5.15 Flow chart of RIAM based site impact assessment

5.5.1.2 Graphical User Interface

A user friendly GUI has been developed for the RIAM based site impact assessment tool. User can also add, modify and delete any RIAM component under the four categories, i.e. PC, SC, BE and EO. User can assign the scale of different criteria under Group A and Group B by using the drop down menus. This ensures that only the valid values are entered by the user. User can save the composition of RIAM components as themes and they are stored in the geodatabase to be used in future. This is useful to compare impact assessment of multiple sites and to facilitate group decision making. Once all RIAM components are added to the impact assessment strategy, the theme can be saved in the geodatabase and can be accessed, modified or deleted in future. This theme based RIAM strategy can be applied on different potential sites and then the best site can be selected for development. The main interface of the RIAM tool is shown below.

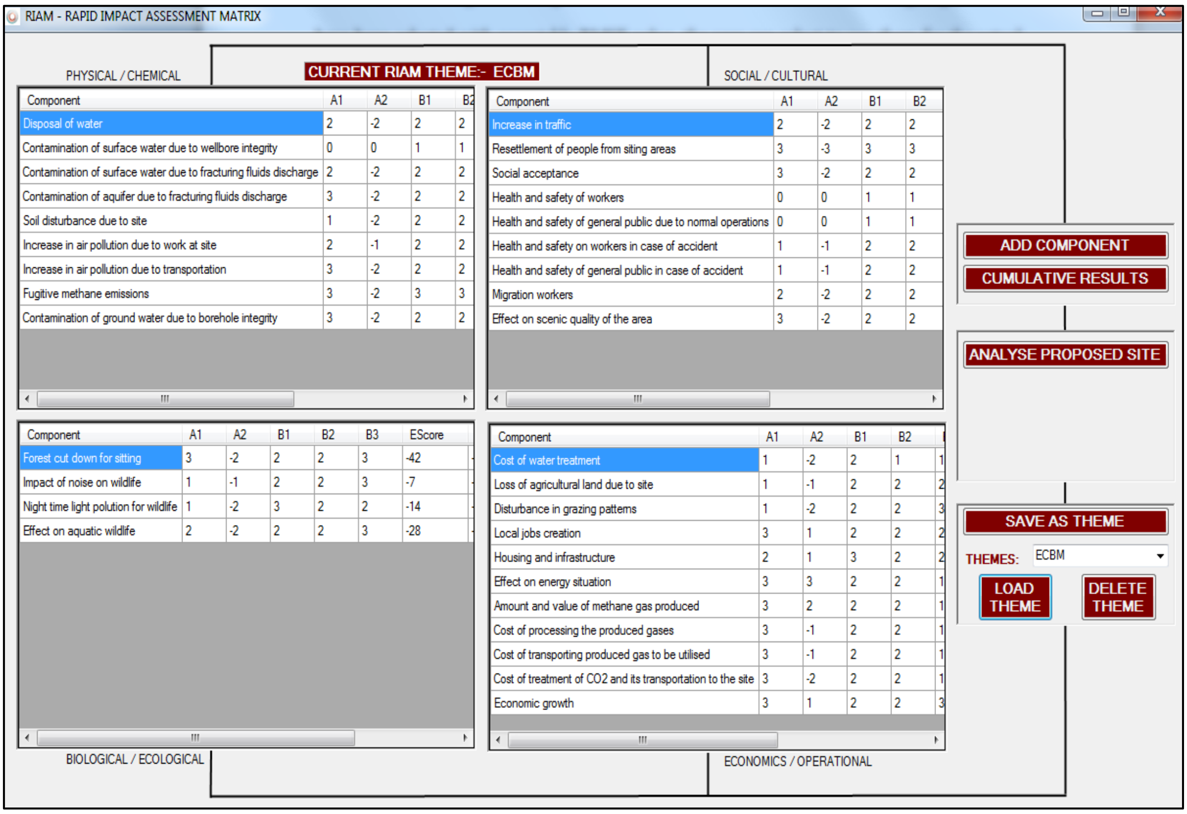


Figure 5.16 GUI of the RIAM based site impact assessment tool

The user enters each component in the system for the first time using the GUI shown in Figure 5.17 by providing the RIAM component name, its category and the score for A1, A2, B1, B2, B3 components. User can also identify and link any spatial data that can be influenced by the components. This information is used for spatial vulnerability assessment as explained in sections 5.5.1.1.

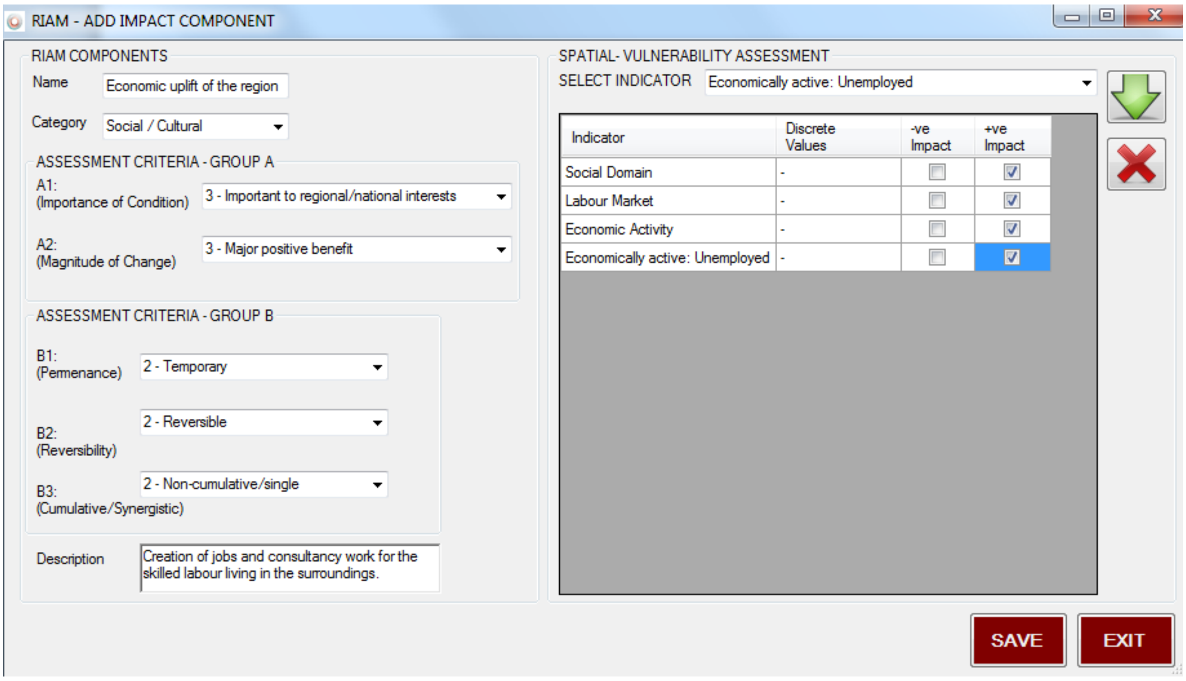


Figure 5.17 GUI for adding new RIAM component

5.5.1.3 Spatial vulnerability assessment

A novel addition has been developed in the RIAM based site impact assessment to assist the decision making process in identifying the spatial vulnerability of different RIAM components in the surrounding regions of a site. For this, user identifies key indicators and links it with appropriate RIAM components for a given site. User can select multiple indicators for a given RIAM component and suggests the system whether this indicator will have positive or negative impact by the given RIAM component.

All the spatial data linked with 500×500 Fishnet vector grid associated with the AHP based site selection tool is available in this tool as well. If the selected indicator is quantitative in nature, e.g. population, percentage of unemployment, then user can simply suggests whether it is going to be negatively or positively affected. However, if it is qualitative in nature then the system searches for discrete classes in the geodatabase and adds it to the GUI, where user needs to identify the negative or positive influence of the RIAM component on each class individually. As an example, if the user is adding a RIAM component that highlights the negative impact of a site development on the ground water quality. For this, the user links the discrete classes of the Hydrogeological features layer (e.g. 'Highly productive aquifer', 'Low productivity aquifer' and 'Coastline Brackish') to the appropriate RIAM components. Some of these classes will be negatively impacted by the site while others will not be affected. User can add or remove any related discrete classes from a given qualitative dataset to be linked with the RIAM component.

The tool provides an interface for analysing a site on the basis of such spatial vulnerability analysis. User provides the centroid coordinates of the given site and a buffer distance to select the affected neighbouring region around the site. The site analyser tool creates the buffer around the site and calculates the minimum, maximum and average values of the indicators in the buffered region and also in the entire study area.

This is helpful for the decision makers to analyse the proposed site according to its spatial vulnerability against the given indicators. For qualitative indicators, the site analyser tool work differently. It creates a buffer around the site and calculates the percentage of the each class in the given buffered region and also in the entire study area. On the basis of this spatial vulnerability assessment of the proposed site, it is easier for the decision maker to select the scale of a given RIAM component under different RIAM criteria. The overall results for the RIAM based impact assessment for a given site can also be viewed by clicking on the cumulative result button. The results are shown in Figure 5.18, using a table and graphs for all four domains, i.e. PC, SC, BE and EO. The visualisation of RIAM results using different graphs assists the decision makers in selecting the site which has least negative impact and more positive impact factors.

RIAM- CUMULATIVE RESULTS

	-E	-D	-C	-B	-A	NI/NC	+A	+B	+C	+D	+E
Physical / Chemical	0	3	3	2	0	1	0	0	0	0	0
Biological / Ecological	0	1	1	1	1	0	0	0	0	0	0
Social / Cultural	1	2	2	0	2	2	0	0	0	0	0
Economics / Operational	0	0	1	3	2	0	0	2	2	1	0

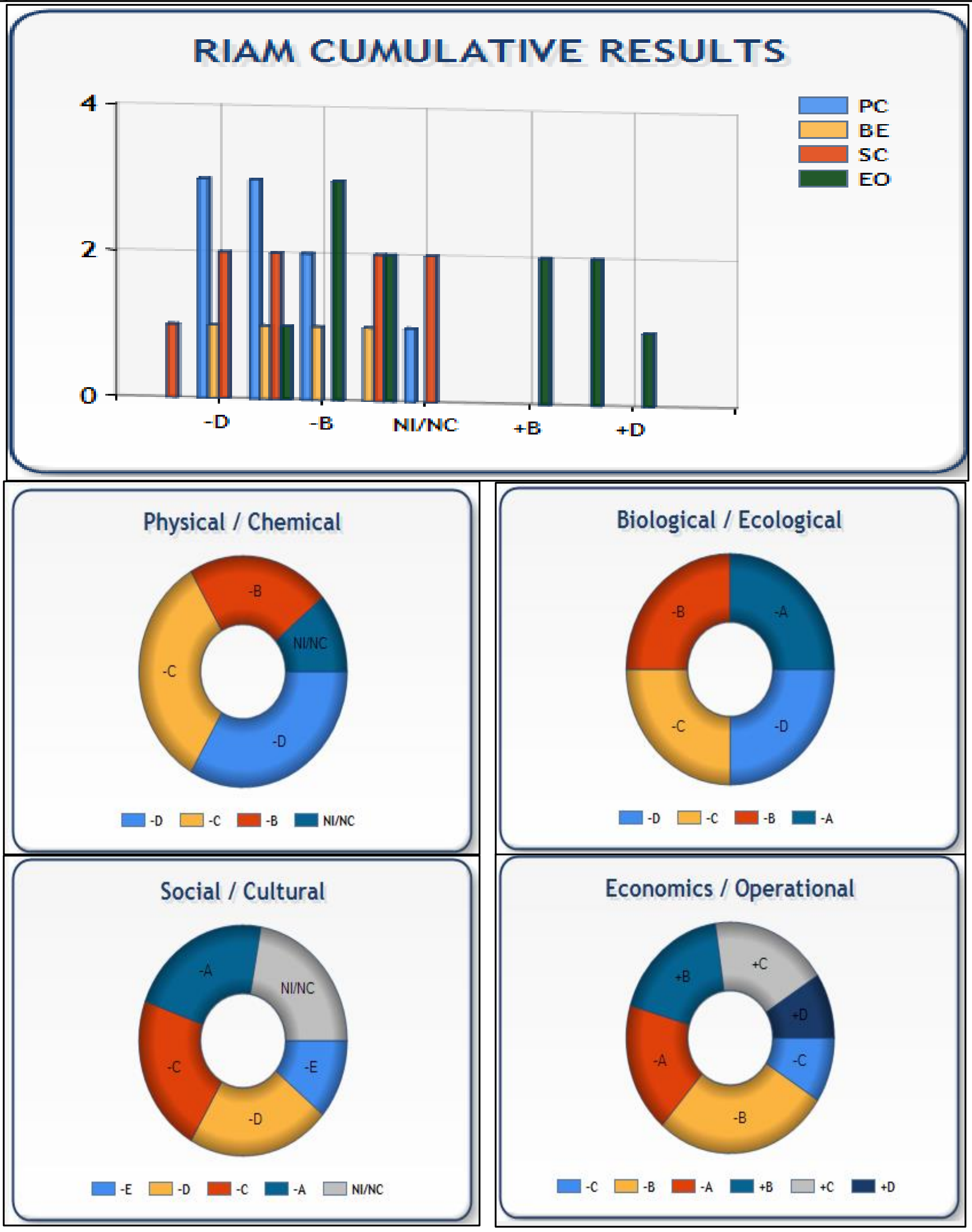


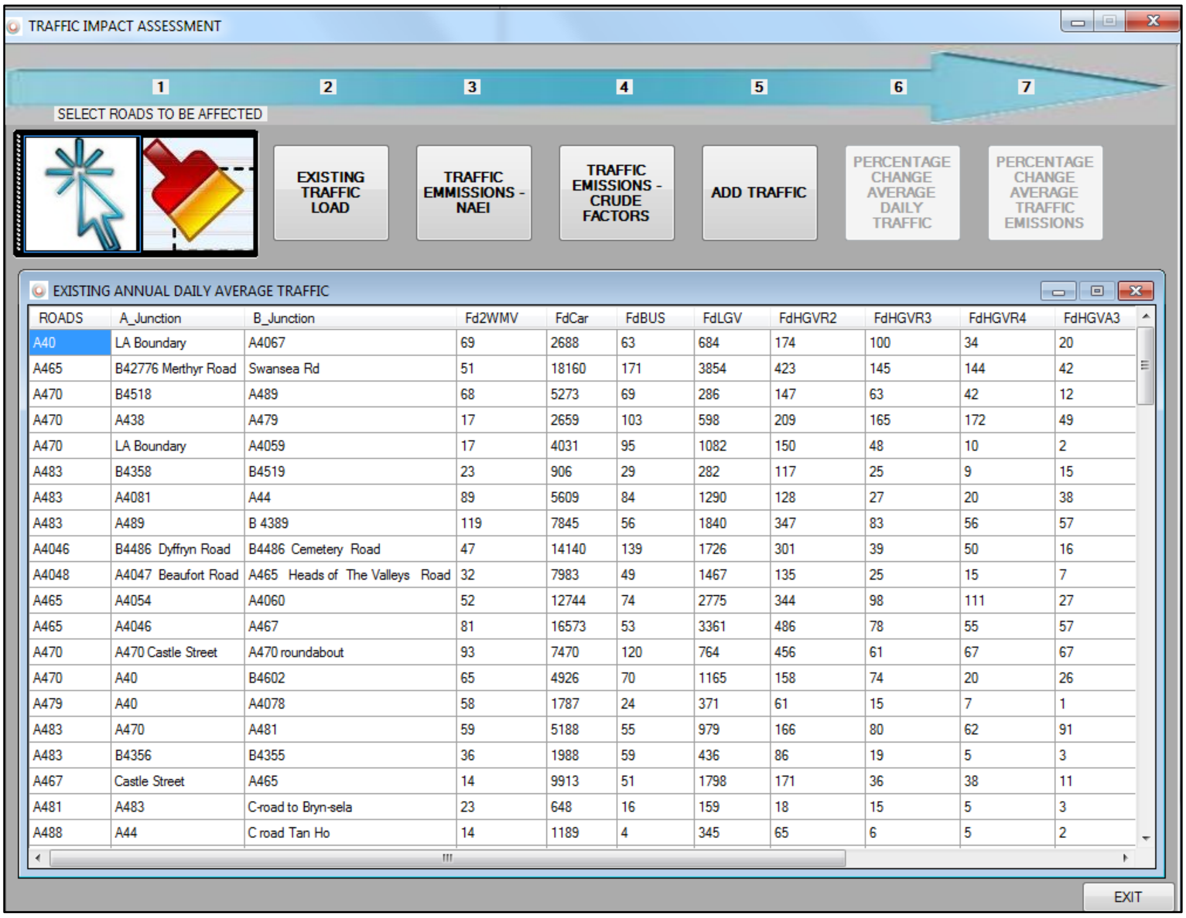
Figure 5.18 Cumulative results of RIAM based site impact assessment tool

5.5.2 The traffic impact assessment tool

As explained in Chapter 2, increase in road traffic, noise and pollution is a big concern linked to the engineering interventions of Geoenergy. Some of these applications result in an increase in the road traffic on routes linking the sites with existing network. This has created a lot of disruption for the local population and new projects have faced resilience because of this. It is very important for the future of Geoenergy applications that these impacts are assessed in detail and mitigating arrangements are made accordingly. Therefore traffic impact assessment tool is developed and has been made an essential part of the impact assessment subsystem of the SDSS.

The traffic impact assessment tool loads a traffic layer into the SDSS to be analysed for a given engineering intervention. The traffic layer contains the important road segments in the study area (Wales), existing traffic load and emissions. The data itself is explained in details in Chapter 6. The main interface of the tool is an MDI Container (parent window that can have child windows in it). It has multiple child windows that are used to show different information related to traffic, emissions and percentage of the traffic change etc. Figure 5.19 shows the main interface of the traffic impact assessment tool. User first selects the potentially affected road segments using the selection tool provided in the tool. User can view the existing load of traffic on these segments. User can also view the associated measured values of different emissions in the surrounding regions of these road segments.

The traffic information included in the geodatabase consists of Annual Average Daily Flow (AADF) from department of Traffic, UK (DoT 2014). It has AADF data calculated for different types of vehicles attached to the road segments. User can also view the air emissions along the selected segments of the roads.



ROADS	A_Junction	B_Junction	Fd2WMV	FdCar	FdBUS	FdLGV	FdHGVR2	FdHGVR3	FdHGVR4	FdHGVA3
A40	LA Boundary	A4067	69	2688	63	684	174	100	34	20
A465	B42776 Merthyr Road	Swansea Rd	51	18160	171	3854	423	145	144	42
A470	B4518	A489	68	5273	69	286	147	63	42	12
A470	A438	A479	17	2659	103	598	209	165	172	49
A470	LA Boundary	A4059	17	4031	95	1082	150	48	10	2
A483	B4358	B4519	23	906	29	282	117	25	9	15
A483	A4081	A44	89	5609	84	1290	128	27	20	38
A483	A489	B 4389	119	7845	56	1840	347	83	56	57
A4046	B4486 Dyffryn Road	B4486 Cemetery Road	47	14140	139	1726	301	39	50	16
A4048	A4047 Beaufort Road	A465 Heads of The Valleys Road	32	7983	49	1467	135	25	15	7
A465	A4054	A4060	52	12744	74	2775	344	98	111	27
A465	A4046	A467	81	16573	53	3361	486	78	55	57
A470	A470 Castle Street	A470 roundabout	93	7470	120	764	456	61	67	67
A470	A40	B4602	65	4926	70	1165	158	74	20	26
A479	A40	A4078	58	1787	24	371	61	15	7	1
A483	A470	A481	59	5188	55	979	166	80	62	91
A483	B4356	B4355	36	1988	59	436	86	19	5	3
A467	Castle Street	A465	14	9913	51	1798	171	36	38	11
A481	A483	C-road to Bryn-sela	23	648	16	159	18	15	5	3
A488	A44	C road Tan Ho	14	1189	4	345	65	6	5	2

Figure 5.19 GUI of the traffic impact assessment tool

The user can add the anticipated number of each type of vehicles in the form of AADF in the data to check the percentage change in traffic flow and air emissions. The percentage change in traffic and existing information on AADF and air emissions can help decision makers in planning the routes and selecting the sites with least impact. The road transport emission data has been collected from the National Air Emissions Inventory (NAEI 2014). NAEI uses COPERT model for the calculations of emissions considering all important parameters. The actual model is very complex and requires a lot of parameters for emission modelling such as the fleet age, type, cold start, hot exhaust, vehicle speed, tyres wear and tear, road abrasions and hot soak emissions etc. (NAEI 2014). For a general estimation of the emission, the emission factors provided by the NAEI (Table 5.4) for different types of vehicles have been used to calculate a rough estimate of the change in the emission rates with the increase in traffic on a given road segment.

Table 5.4 Average road transport emission factors for UK fleet in 2011(NAEI 2014)

Hot exhausts Including cold start	g/km NH_3	g/km Benzene	g/km CO	Kg/km CO_2	g/km VOC	g/km NOx	g/km NO_2	g/km PM10	g/km PM2.5	g/km SO^2
M/cycle	0.002	0.033	9.858	0.21191	0.855	0.215	0.002	0.014	0.013	3E-04
Petrol cars	0.036	0.007	2.513	0.21191	0.120	0.208	0.002	0.001	0.001	0.001
Diesel cars	0.001	4E-04	0.090	0.24721	0.015	0.611	6E-03	0.026	0.024	0.001
Busses	0.003	0.009	0.730	0.82475	0.151	6.452	0.016	0.081	0.077	0.004
Petrol LGVs	0.042	0.013	8.043	0.21191	0.307	0.663	0.008	0.001	0.001	0.001
Diesel LGVs	0.001	0.001	0.355	0.24721	0.055	0.898	0.006	0.055	0.052	0.001
Rigid HGVs	0.003	0.003	0.689	0.82475	0.101	3.603	0.014	0.058	0.055	0.004
Artic HGVs	0.003	0.002	0.422	0.98753	0.065	3.694	0.029	0.058	0.056	0.005

The average CO_2 emission factors used by the NAEI for modelling are consistent with the Greenhouse Gas Conversion Factors for Company Reporting Factors and the same are used here. Also Volatile Organic Compound (VOC) and Benzene emission factors include evaporative emissions. CO_2 emission factors for buses, cars and motorcycles are not given in the Greenhouse Gas Conversion Factors for Company Reporting Factors. For Buses, same values are used as that for the rigid HGVs. For cars and motorcycles, same values are used as that for LGVs.

The GRNN prediction tool (explained in Chapter 4) can also be used to make different types of predictions using this data, such as changes in the emission rates, congestions and noise pollution etc.

5.6 Conclusions

The development aspects of the SDSS have been discussed in this chapter in particular the MCDA based analytical modules. Analytical modules of the system are divided into four sections namely a) Site Selection and Ranking, b) Impact Assessment, c) Spatial Knowledge Discovery and d) Geodatabase Management.

Some of the analytical modules of the SDSS utilises Multi Criteria Decision Analysis (MCDA) techniques including A) Analytical Hierarchy Process, b) Weighted Linear Combination (WLC), Criterion Sorting Mechanism (CSM) and d) Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS). The structure and algorithm for these analytical modules are explained along with their mathematical formulation. Site selection tool utilises AHP and WLC techniques. Sensitivity analysis has also been incorporated in the tool for the decision support on the selection of relative weights for criterion maps. CSM and TOPSIS techniques have been incorporated in the site ranking by neighbourhood analysis and comparison tool. Rapid Impact Assessment Matrix (RIAM) has been utilised in the impact assessment tool.

The Graphical User Interfaces (GUI) of the analytical modules is also described in the Chapter to understand the essential parameters required to carry out respective analysis. Some novel techniques have been introduced in the development of the analytical modules in this research, such as:

- The AHP and WLC based site ranking tool facilitates a novel theme based structure where the decision maker can save their preferences (selected criterion and relative importance) as themes in the geodatabase. Existing themes can be loaded in the module to compare the results generated with other themes. Depending on the requirements, the AHP analysis can be applied on individual domains separately or all domains together. The tool provides a mechanism to apply constraints and filters so that the processing is reduced to the selected potential areas only instead of the entire study area.

- A novel approach has been introduced to link spatial dimension to the impact assessment components of the RIAM module. The tool facilitates the decision makers to associate any number of key indicators (available in the geodatabase) likely to be impacted (positive or negative) with the RIAM components. This can facilitate the spatial vulnerability assessment of a given site and its surrounding areas in terms of the RIAM components.
- A novel site ranking method (CSM) is introduced in the site ranking by neighbourhood analysis tool. It can calculate the site ranks based on individual indicators and a cumulative rank, by sorting the values (scaled) of indicators in the given neighbourhood.

References

Andrienko, G. A. N. 2001. Exploring spatial data with dominant attribute map and parallel coordinates. *Computers, Environment and Urban Systems* 25(1), pp. 5-15.

Bowden, G. J.,Nixon, J. B.,Dandy, G. C.,Maier, H. R. and Holmes, M. 2006. Forecasting chlorine residuals in a water distribution system using a general regression neural network. *Mathematical and Computer Modelling* 44(5–6), pp. 469-484.

Chen, Y.,Li, K. W. and Liu, S.-f. 2011. An OWA-TOPSIS method for multiple criteria decision analysis. *Expert Systems with Applications* 38(5), pp. 5205-5211.

DoT. 2014. *Department for Traffic* [Online]. Available at: http://www.dft.gov.uk/traffic-counts/index.php [Accessed: 04 March].

Feizizadeh, B.,Jankowski, P. and Blaschke, T. 2014. A GIS based spatially-explicit sensitivity and uncertainty analysis approach for multi-criteria decision analysis. *Computers & Geosciences* 64(0), pp. 81-95.

Hwang, C. L. and Yoon, K. 1981. *Multiple attribute decision making—Methods and applications*. Heidelberg: Springer-Verlag.

Jia, J.,Fan, Y. and Guo, X. 2012. The low carbon development (LCD) levels' evaluation of the world's 47 countries (areas) by combining the FAHP with the TOPSIS method. *Expert Systems with Applications* 39(7), pp. 6628-6640.

Malczewski, J. 1999. *GIS and Multicriteria Decision Analysis*. Wiley.

Massam, B. H. 1988. Multi-Criteria Decision Making (MCDM) techniques in planning. *Progress in Planning* 30, Part 1(0), p. 33.

Microsoft. 2014. *Microsoft .Net Chart control sample enviornment* [Online]. Available at: http://code.msdn.microsoft.com/Samples-Environments-for-b01e9c61 [Accessed: 14 October].

Moeinaddini, M.,Khorasani, N.,Danehkar, A.,Darvishsefat, A. A. and zienalyan, M. 2010. Siting MSW landfill using weighted linear combination and analytical hierarchy process (AHP) methodology in GIS environment (case study: Karaj). *Waste Management* 30(5), pp. 912-920.

NAEI. 2014. *National Atmospheric Emissions Inventory* [Online]. Available at: http://naei.defra.gov.uk/ [Accessed: 04 March].

Nefeslioglu, H. A.,Sezer, E. A.,Gokceoglu, C. and Ayas, Z. 2013. A modified analytical hierarchy process (M-AHP) approach for decision support systems in natural hazard assessments. *Computers & Geosciences* 59(0), pp. 1-8.

Pastakia, C. M. R. 1998. The rapid impact assessment matrix (RIAM)—A new tool for environmental impact assessment. In: Jensen, K. ed. *Environmental impact assessment using the rapid impact assessment matrix (RIAM)*. Olsen & Olsen, pp. 8-18.

Pastakia, C. M. R. and Jensen, A. 1998. The rapid impact assessment matrix (Riam) For eia. *Environmental Impact Assessment Review* 18(5), pp. 461-482.

Saaty, T. L. 1980. *The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation*. New York: McGraw-Hill.

SpatiaLite. 2014. *SpatiaLite Geodatabase* [Online]. Available at: https://www.gaia-gis.it/fossil/libspatialite/index [Accessed: February 06].

6

VERIFICATION OF ANALYTICAL MODULES

6.1 Introduction

This chapter considers the verification of different analytical modules developed for the SDSS. As described in Chapter 3, existing techniques including SOM, PCP, GRNN, TOPSIS and AHP have been used and developed in the various functionalities of the SDSS. Appropriate benchmarks and alternative software such as Matlab have been used to examine the accuracy of the code developed for analytical modules.

As discussed in Chapter 4, two types of ANN have been used in the SDSS, i.e. a) SOM and b) GRNN. The SOM has been used for finding correlations that may exist in the data and also for the site ranking purpose. GRNN has been used in the prediction tool along with the "Holdout" method and Genetic Algorithm. AHP, Pairwise Comparison Method, CSM and

TOPSIS have been utilised for site selection and ranking. RIAM has been utilised for the site impact assessment.

The following sections explain the verification of the analytical modules by comparing their results with the results of appropriate alternative codes. The behaviour of the two codes has been compared under similar conditions, using the same datasets. Verification of SOM based tools for site ranking and clean correlation is covered in Section 6.2 and 6.3. GRNN based prediction tool and the associated Holdout Method and Genetic algorithm based tuning is tested in Section 6.4. Verification of the AHP based site selection tool and the associated Pairwise Comparison Method for calculating relative weights is covered in Section 6.5. Verification of the site ranking by neighbourhood analysis tool and its associated CSM and TOPSIS ranking is covered in Section 6.6. Conclusions are presented in Section 6.7.

6.2 SOM based site ranking tool

In order to compare the results of the one dimensional SOM used for the site ranking purpose, a sample dataset has been prepared and tested against the GeoSOM toolbox. GEOSOM is a MATLAB based tool that utilises the SOMToolBox of the MATLAB (Bação et al. 2005). GeoSOM has been selected for the purpose of verification of the code for two reasons: a) It allows the processing of the one dimensional SOM and b) it can read the geographical data, e.g. a Shapefile. The GeoSOM toolbox can be used for the spatial data clustering and also for knowledge discovery within the dataset.

One dimensional SOM can cluster and order (ascending or descending) the input data on output lattice after convergence. At this stage the input features are represented by the position of their respective Best Matching Unit (BMU) in the ordered one dimensional output lattice. A novel approach has been adopted in this research to use this feature of the one dimensional SOM to assign relative ranks to the sites or geographical regions based on their attributes.

The GeoSOM toolbox does not rank the sites but it can be used to compare the results of the ordering of the dataset using one dimensional SOM to verify the SOM based site ranking tool. The ranking process is carried out after the one-dimensional SOM has reached convergence and the BMUs in the output map have been ordered in ascending or descending order. The site ranking tool then assigns ranks to the geographical regions according to the position of their BMU in the ordered one-dimensional SOM.

6.2.1 Data preparation

For the purpose of verifying the code of the SOM based site ranking tool, a geographical data has been selected first. A Shapefile containing Welsh Index of Multiple Deprivation (WIMD-2011) data for Wales is used for this purpose. WIMD is a reliable source of socio-economic indicator dataset developed and used by the Welsh Government to study multiple deprivation faced by the Welsh population at the Lower Super Output Areas (LSOA) level (WIMD 2011). This dataset is explained in detail in Chapter 7. The Shapefile contains seven individual WIMD indicators and an overall indicator reflecting the cumulative index of deprivation assigned to the 1896 LSOA regions in Wales.

6.2.2 Application

The data has been clustered, ordered and analysed using one-dimensional SOM in both the GeoSOM toolbox and the SOM based site ranking tool developed in this research. The structure of the output map is 1×20 in both cases. Since there is no unit linked with the WIMD indexes, therefore there is no need for commensuration process. An equal weight is assigned to the indicators in the dataset in the case of SOM based site ranking tool as there is no provision in the GeoSOM toolbox to assign different weights to different indicators.

Hexagonal output map has been selected as the crests and troughs of the one-dimensional hexagonal shaped output map can make the visualisation easier. The BMUs are represented by the bottom most and top most hexagon at each crest and trough.

6.2.3 Comparison of results

Ranks assigned to WIMD dataset by the SOM based site ranking tool are shown in Figure 6.1. Each LSOA is presented by its corresponding BMU and the rank assigned to it which is based on the position of its representative BMU in the output map.

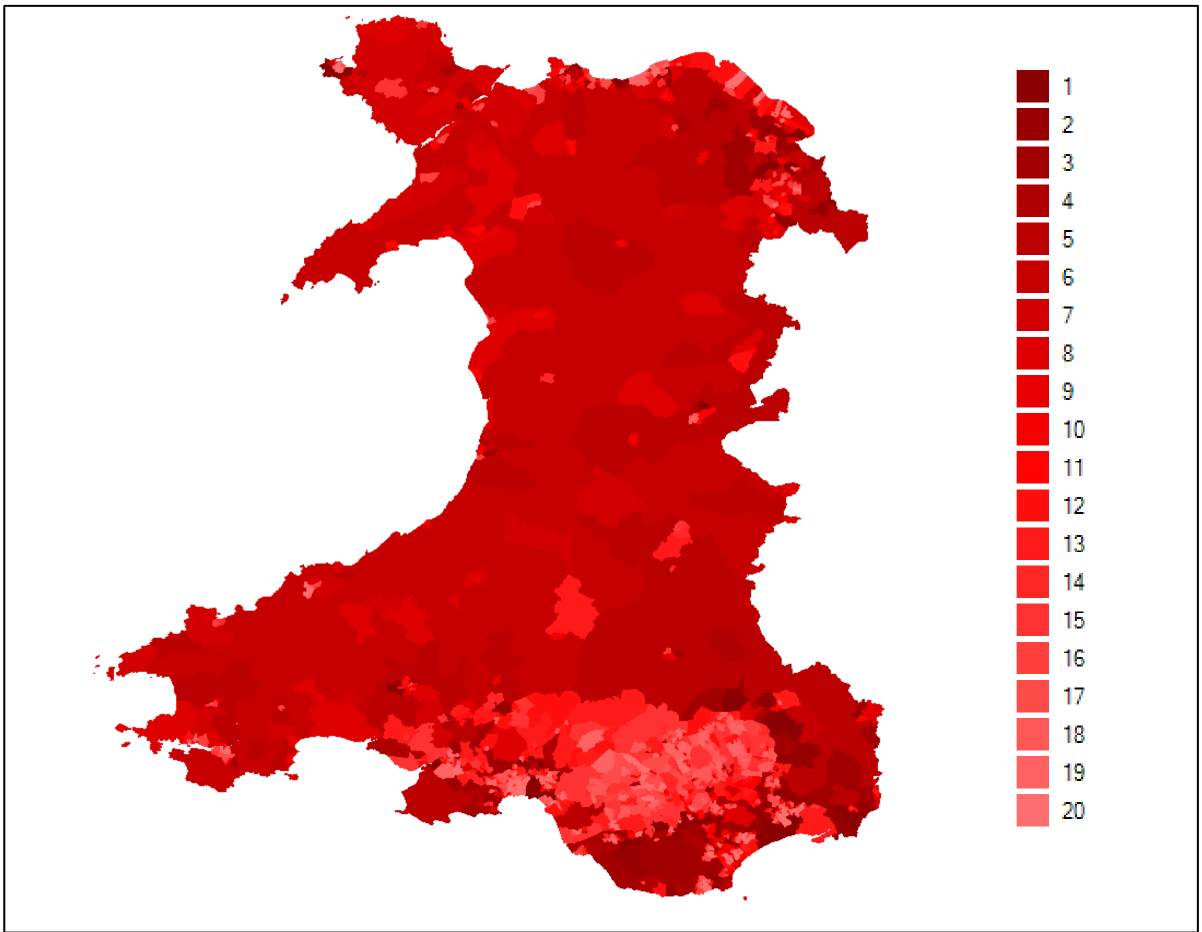
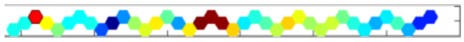


Figure 6.1 Ranks generated by SOM based site ranking tool

After the convergence, the results generated from the two codes are compared. For this purpose the ranks of randomly selected LSOAs generated by the SOM based site ranking tool is taken. To make it a representative sample, one LSOA is randomly selected from each of the twenty BMUs, since one BMU can represent multiple data points. These ranks are compared with position of the BMUs of the same LSOAs as generated by the GeoSOM toolbox after convergence.

Table 6.1 Comparison of results between GeoSOM and SOM based site ranking tool

LSOA	Rank of the selected LSOA: SOM based site ranking tool	Position of BMU representing the selected LSOA: GeoSOM toolbox	Ordering difference
W01000133	1		0
W01000008	2		0
W01000416	3		0
W01000013	4		1
W01000020	5		1
W01001407	6		1
W01000147	7		0
W01000808	9		0
W01001125	10		0
W01001313	12		0
W01000496	14		0
W01001230	15		0
W01001061	16		0
W01000957	17		1
W01001215	18		1
W01001387	19		0
W01001803	20		0

The hexagon highlighted by red colour represents the position of the BMU of randomly selected LSOAs as shown in the third column of the Table 6.1 for each of the 20 selected LSOAs. Comparison of the results is presented in Table 6.1 which shows the rank of the selected LSOAs generated by SOM based site ranking toolkit and the ordered position of the BMU representing the selected LSOAs as generated by the GeoSOM tool.

The comparison shown in Table 6.1 reflects that the two codes have produced similar order for the randomly selected LSOAs. This relative order is important for the ranking of the sites or geographical regions. It is noted that for some LSOAs the position of its represented BMU is different in the two cases. This difference is never exceeding 1 and the order is still retained, which means that the given LSOA is represented by the immediate neighbouring BMUs in the two codes. This behaviour of SOM is expected as the convergence is achieved slightly differently every time even using the same code. Some of the data nodes lying at the boundaries on clusters represented by a BMU can become part of the neighbouring clusters if the difference between the two BMUs is very small. The important thing is the order of the data points (LSOAs in this case), which is the same in both cases.

The main purpose of this comparison was the verification of the SOM code developed in this research with an existing reliable code. The comparison of the two results proves the reliability of the SOM based ranking tool. It is noted that there was no qualitative comparison made between the two codes, e.g. which code converges faster and takes less computational resources etc.

6.3 SOM based clean correlation tool

In order to verify the results of the SOM based clean correlation finding tool, MATLAB based GeoSOM toolbox was used again for the same reasons as described in Section 6.2.1. The individual Component Planes (CP) of the indicators can be exported as Shapefiles from the GeoSOM toolbox. CP is the cross section of the output map representing one variable at a time. The CPs can be used to find any correlation that may exist between different dependent and independent variables, directly from the BMUs (output map) rather than the entire dataset as explained in section 4.4.1.

6.3.1 Data preparation

A geographical data has been selected as shown in Table 6.2 as the benchmark dataset for verification. Selected indicators are combined together at the LSOA level in the form of a Shapefile. The Ind51, i.e. rate of cancer incidence per 100,000 of population, is selected as the dependent variable and the rest of the indicators are taken as the independent variables.

Table 6.2 Indicators selected for the verification of the SOM based clean correlation tool

Abbreviatio	Indicator	Year
Ind6	% of dwellings by council tax band; band A	31-Mar-2011
Ind7	% of dwellings by council tax band; band B	31-Mar-2011
Ind8	% of dwellings by council tax band; band C	31-Mar-2011
Ind9	% of dwellings by council tax band; band D	31-Mar-2011
Ind10	% of dwellings by council tax band; band E	31-Mar-2011
Ind11	% of dwellings by council tax band; band F	31-Mar-2011
Ind12	% of dwellings by council tax band; band G	31-Mar-2011
Ind13	% of dwellings by council tax band; band H	31-Mar-2011
Ind14	% of dwellings by council tax band; band I	31-Mar-2011
Ind51	Rate of cancer incidence per 100,000 population	2000-2009
Ind65	CACI - Mean income (average value)	2011
Ind97	Welsh Index of Multiple Deprivation -	2011

The key concept here is to find the correlation between the cancer incidence and the independent variable directly from the BMUs rather than the entire dataset. The rest of the indicators included in the dataset are the percentage of dwellings by council tax bands and the CACI's PayCheck gross household income data at the LSOA level (CACI 2012).

6.3.2 Application

The data has been used in a two-dimensional SOM in both GeoSOM toolbox and the SOM based site ranking tool. The output map size is kept the same in both cases (10×10 Matrix). Data is first normalised in both cases using 5.4. Once converged, data is exported from the GeoSOM toolbox as individual Component Planes for each indicator. These attribute data of all CPs is joined together using their unique feature IDs. The correlation coefficient of the dependent and independent variables is then calculated using Matlab. On the other hand, the SOM based clean correlation tool calculates the correlation as an automated process and the results are displayed as a matrix.

6.3.3 Comparison of results

The correlation between cancer incidents and the independent variables as calculated by the SOM based clean correlation tool and as manually calculated in the Matlab is given in Table 6.3. By calculating the correlation just at the BMUs level, dataset is reduced approximately eighteen times than the original dataset and still the system is able to calculate the correlation between dependent and independent variables.

In order to verify the property of the SOM to calculate clean correlation in the presence of any noise in the data, another test was performed by adding some noise in the data and repeating the process. Random noise was added to original dataset in each independent variable with a Signal to Noise Ratio (SNR) of 1%. The SOM based clean correlation tool give robust results even with random noise added to the dataset as shown in Table 6.3. the first two columns show the dependent and independent variables. Third column shows the correlation calculated from the original dataset using Matlab. Fourth column shows the clean correlation calculated by SOM based clean correlation finding tool from the converged output map, with and without noise. Last column shows the clean correlation calculated from the Components Planes produced by GeoSOM tool after the convergence.

Table 6.3 Clean correlation results comparison

Dependent variable	Independent variables	Matlab correlation from entire dataset	SOM based CC tool correlation from BMUs		GeoSOM correlation (without noise) from BMUs
			Without noise	With noise (1 % SNR)	
Ind51	Ind6	0.2243	0.2957	0.2566	0.2708
Ind51	Ind7	0.2280	0.3172	0.2729	0.3091
Ind51	Ind8	0.0402	0.0947	0.1239	0.0969
Ind51	Ind9	-0.1503	-0.1399	-0.1577	-0.2029
Ind51	Ind10	-0.2992	-0.4538	-0.4021	-0.4139
Ind51	Ind11	-0.2426	-0.4015	-0.3274	-0.3017
Ind51	Ind12	-0.2154	-0.3040	-0.2596	-0.2720
Ind51	Ind13	-0.1498	-0.2187	-0.2038	-0.1918
Ind51	Ind14	-0.0997	-0.1744	-0.1866	-0.1615
Ind51	Ind65	-0.3161	-0.4656	-0.4180	-0.4671
Ind51	Ind97	-0.4448	-0.6192	-0.6054	-0.5825

Three observations can be made from the results shown in Table 6.3. Firstly, the correlation between cancer incident data and the rest of the indicators conforms to the facts that are already known. A negative correlation exists between higher income groups and number of cancer incidents, a positive correlation exists between the multiple deprivation and the cancer incidents and similarly a positive correlation exists between the percentage of houses located in lower tax bands and the cancer incidents. These observations depicts that the number of cancer incidents increase in the areas where income and therefore living standard is comparatively lower. This can be used as the basis for “causal relationship” for further investigation in local areas considering other environmental and genetic causes. Secondly, the results also reveal that using SOM and calculating the clean correlation from the BMUs directly can yield similar results compared to the calculations made from the entire dataset. Furthermore the SOM based clean correlation can be robust even if the data has some random noise added to it, which makes it a useful analytical tool to solve real life problems.

Comparison of the results confirms that the SOM based clean correlation finding tool can be used for knowledge discovery from the dataset in the form of correlation/causality and also that the tool is resistant to noise in the dataset.

6.4 GRNN based prediction tool

The General Regression Neural Network (GRNN) based prediction tool can be used for prediction, regression or interpolation. A novel feature has also been added to the GRNN in this research to incorporate spatial parameter (distance) as an independent variable for the regression analysis as explained in Section 4.3.2. In order to verify the code of the General Regression Neural Networks Based prediction tool, its results can be compared with those calculated by using the NewGRNN tool in Matlab (Mathworks 2013). A GRNN can be created by using 6.1 in the Matlab and the network can be used for prediction or regression analysis.

$$net = newgrnn(P, T, spread) \tag{6.1}$$

where P is an $R \times Q$ matrix of Q input vectors and R independent variables, T is an $S \times Q$ matrix of Q target vectors and S dependent variables. The spread factor is the value of Sigma

(σ) parameter which is explained in Section 4.3.2. The default value of Sigma in NewGRNN function is 1.0. After the network is created using 6.1, the value of target vector can be predicted at the prediction point.

6.4.1 Data preparation

In order to compare the two codes, a prediction at unknown location can be made using the same data and same neural network parameters. For this purpose the Cancer data prepared earlier for the verification of the SOM based clean correlation tool has been used to train the GRNN network and for prediction. As described in Table 6.2, this data contains the rate of cancer incidence per 100,000 of population, percentage of dwellings under different council tax bands and the CACI's PayCheck gross household income data at the LSOA level (CACI 2012). Cancer incidence is taken as the dependent variable, whereas rest of the indicators are considered as the independent variables.

6.4.2 Application

In order to make a prediction and verify the results, one of the randomly selected sample points has been deliberated taken out of the dataset. Prediction has to be made at this point to verify the two codes by comparing the error terms. The Holdout Method used for the training of the neural network in GRNN based prediction tool, works on the same principal as explained in Section 4.3.2. The Sigma parameter defines the size of the neighbourhood sample points that are going to be used in the prediction. The closer points have more influence than those distant apart. Data has been scaled between 0-1 so that a single Sigma parameter can be used for the all indicators.

6.4.3 Comparison of results

In order to compare the results under similar conditions, the dataset was scaled using the Maximum Score Procedure using 5.1. Scaling all the data between 0 and 1 is also useful because a single Sigma parameter can be used for all dimensions as the Matlab based GRNN function does not support the use of a separate Sigma parameter for every dimension. There are eleven independent variables (indicators) and one dependent variable as described in Table 6.2. First the data is scaled and then one out of 1896 LSOAs was held out from the data so that the prediction can be made at this point and compared with the actual value. Different values for Sigma parameter were tested in both the codes, i.e. 0.1, 0.2, 0.5 and 1.0. The actual and predicted value of the Y parameter (dependent variable) at the prediction point, have been compared in the Table 6.4. As it can been seen from the comparison presented in Table 6.4, the two codes have produced very similar results with different values of Sigma parameter. The slight difference can be due to the fractional changes in rounding off, or the way the distance in calculated in the two codes.

Table 6.4 General Regression Neural Networks results verification

Sigma	Y	Ý-MATLAB GRNN	Ý-GRNN PREDICTION TOOL		
			ASPATIAL	SPATIAL WITH FIXED KERNEL	SPATIALLY ADAPTIVE KERNEL (20)
0.1	742.5	612.52	643.90876	644.86409	503.23761
0.2	742.5	614.6253	613.03788	615.81926	566.84211
0.5	742.5	600.1191	587.37432	586.83366	576.3426
1.0	742.5	580.1521	575.94501	575.51005	575.40803

The GRNN based prediction tool also offers a novel technique of incorporating spatial parameters in the calculation of the dependent variable. The results have been shown with and without the incorporation of spatial parameters. There are two different ways in which spatial parameters can be incorporated using the GRNN based prediction tool as explained in 4.6.1. The predicted values are given in Table 6.4 using spatial parameters, both with a fixed kernel size and spatially adaptive kernel size (with 20 neighbouring features).

The average value of the dependent variable (Rate of cancer incidence per 100,000 population from 2000-2009) is 570.5 in the entire study area. As the value of Sigma parameter is increased, the prediction becomes close to the average value. Conversely, smaller values of Sigma parameter, results in better prediction since the neighbouring points are given more weight in the calculation as compare to those at a distance. This is an expected trend of the GRNN as described in detail in Section 4.3.2. Very small Sigma parameters can also result in over-fitting which can be avoided by using Holdout Method for the entire dataset and finding the most suitable Sigma parameters with least RMSE values.

6.5 Holdout Method and GA based optimisation of GRNN

The Sigma parameter is the only parameter that a GRNN requires from the user and its value can impact the predictions made using the network. Therefore the GRNN based prediction tool also provides a mechanism to find a best set of Sigma parameters using either Holdout Method or Genetic Algorithm (GA) or a combination of both. It is useful when there are multiple independent variables and their relationship with dependent variable is complex. In order to verify the code of Holdout Method and Genetic Algorithm and to find a suitable set of Sigma parameters, Matlab was used. Holdout Method is an established approach to find the fitness of GRNN as explained in Section 4.3.2. Fitness of GRNN is assessed by calculating the RMSE for a given set of Sigma parameters, holding out one data point at a time and then predicting at the held out point.

6.5.1 Data preparation

In order to compare the two codes, a prediction at unknown location can be made considering the same data, same neural network parameters and using the Holdout Method and GA. For this purpose the Cancer data prepared earlier for the verification of the SOM based clean correlation tool has been used, as described in Table 6.2. Rate of cancer incidence (per 100,000 Population) for LSOAs is considered as the dependent variable and all other indicators are considered as the independent variables. The RMSE of the two codes can be compared while predicting the values of the dependent variable (cancer incidents) using different values of Sigma parameter.

6.5.2 Application

In order to compare the results of the Holdout Method for Sigma parameter selection in the GRNN based prediction tool, "NewGRNN" tool of the Matlab (Mathworks 2013) has been used as explained in Section 6.4. Holdout Method is implemented in a Matlab script that takes as input the data, an upper, lower bound and a step (interval) of the Sigma parameter. It holds out one sample point and use the rest to predict the value of the dependent variable at this point using the "newgrnn" function of Matlab and the given Sigma parameter. The Matlab based code for Holdout Method was run on the Cancer dataset with the following Sigma parameters:

- Upper Bound: 1.0
- Lower Bound: 0.01
- Step (interval): 005

RMSE is calculated from the predicted value and the actual value of the dependent variable. Once prediction has been made at all sample points (1896 in this case), the Sigma parameter is changed using the step parameter. This process is repeated until the upper bound of the Sigma parameter is reached. The RMSE values are plotted against the corresponding values of the Sigma parameter used. The plot suggests the optimum value of the Sigma parameter to be used for the prediction purpose, i.e. the one that results in the lowest RMSE value. The Holdout Method tool developed in the GRNN based prediction tool is also applied on the same dataset, Sigma parameter range and the step interval to calculate the RMSE value while predicting the dependent value at one data point at a time. Similarly the GA based Sigma selection tool is used in the GRNN based prediction tool to look for the best Sigma parameters under the similar conditions. The Sigma parameter suggested by the GA code is then compared with the one calculated using Holdout Method and “newgrnn” function of Matlab earlier to verify the applicability of the GA code.

6.5.3 Comparison of results

Comparison of the results of the Holdout Method implemented in the GRNN based prediction tool and using Matlab is presented in Figure 6.3. The plot shows the relation between different Sigma values and the resultant RMSE calculated by the two codes using Holdout Method. Same dataset is analysed using same Sigma range and step interval values in the two codes. The two codes have produced very similar patterns for RMSE plotted against associated Sigma values as it can be seen in Figure 6.3. In both cases, an optimum RMSE value is calculated when the Sigma parameter value is approximately 0.16.

The plot also depicts that the RMSE is high when very small and very large values of Sigma parameter is used.

For the combined application of the Genetic Algorithm and Holdout Method, the same dataset is used under similar conditions, i.e. the Sigma upper and lower bounds. Target RMSE is set at as low as 0.001, population size is set to be 20 and the total number of generations set to be 10. The cross over rate and mutation rate is set to the recommended levels of 0.7 and 0.001. Similar Sigma was used for all the variables. These are the essential parameters required for GA as explained in 4.3.2.The best Sigma found after 20 generation has a value of 0.163, which is very close to the optimum Sigma parameter value as calculated earlier using Holdout Method by using the GRNN based prediction tool and in Matlab.

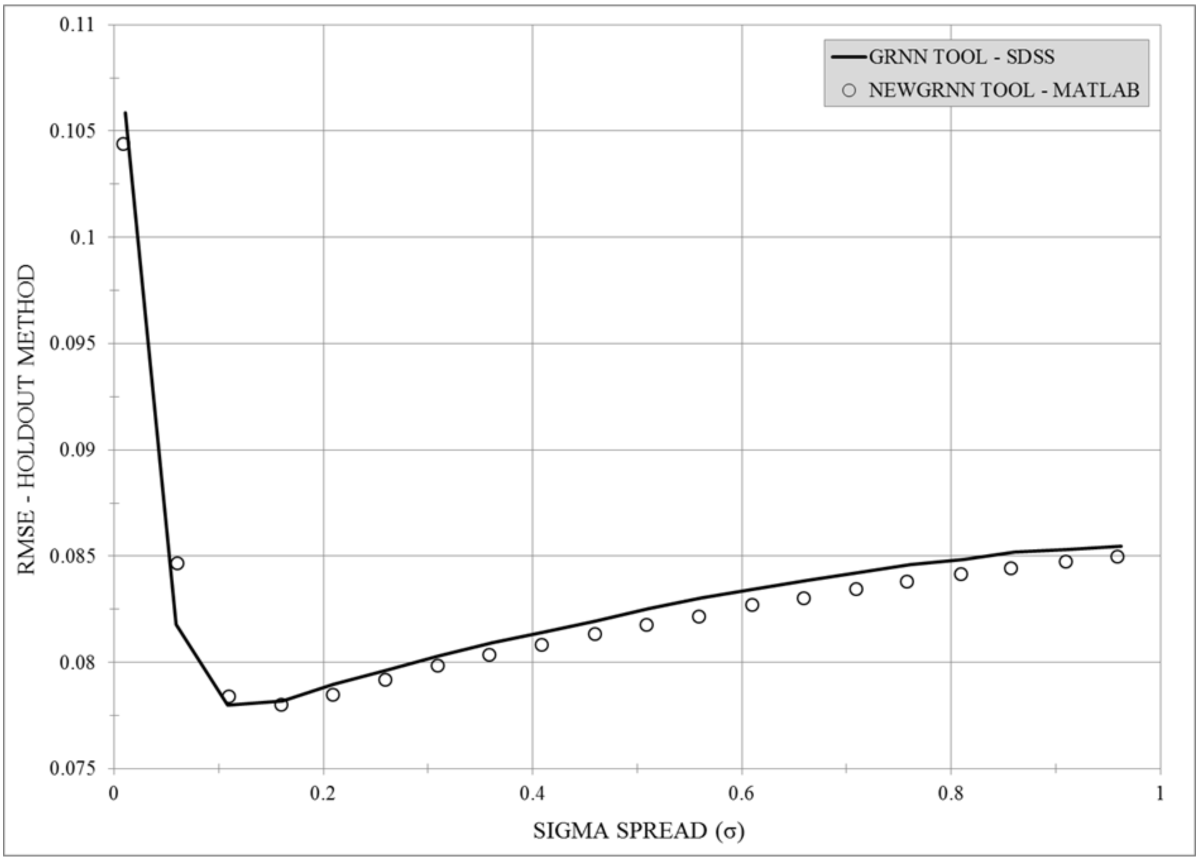


Figure 6.2 Sigma optimisation comparison - using Holdout Method implemented in GRNN based prediction tool and Matlab based NEWGRNN tool

The GA based Sigma optimisation tool can be used in both cases whether a single or a separate value is used for Sigma parameter for each variable (dimension) used in the analysis. However, it is more useful in the second case where a number of possible combinations of Sigma parameter values are to be tested to find a set with least RMSE. If a single spread factor is used, and data is scaled, then Holdout Method can be used with a given range and step interval for the Sigma parameter values and the value with least RMSE can be used in the prediction.

The comparison of the SOM based site ranking tool with appropriate Matlab based tools give satisfactory proof that the tool can be used with confidence. Also the associated tools

(Holdout Method and GA) can be useful for the selection of appropriate values for the Sigma parameters.

6.6 AHP based site selection tool

For the verification of AHP based site selection tool, a site suitability example is tested using the tool. The site suitability problem considered is similar to that provided by (Malczewski 1999). The problem involves evaluation of the three sites: Parcel A, B and C.

The overall goal of the AHP based analysis is to identify the most suitable parcel based on the desired criteria. The desired criteria are constituted by two objectives: (Objective-1) Economic and (Objective-2) Environment.

There are three indicators at the last level of the hierarchy: a) Price, b) Slope and c) Views. These indicators are used to achieve the above two objects which are used to achieve the overall goal. The Economic objective has only one indicator, i.e. the Price. The Environment objective is dependent on the other two indicators, i.e. Slope and Views. The values of these indicators for each parcel are given in Table 6.5.

Table 6.5 Data for site suitability problem – Adopted from (Malczewski 1999)

Parcel of Land	**Criterion**		
	Price ($)	**Slope (%)**	**Views(rank)**
A	96000	5	1
B	80000	8	3
C	110000	4	2
Relative Weights	0.667	0.250	0.083

Using the pairwise comparison method, relative weights have been assigned to the indicators and objectives. In the site suitability example used here for verification, the Economic objective is given two times more importance compared to the Environment objective. Price of parcel is the only attribute in the Economic objective, so the entire weight of this objective is carried forward to the Price attribute. In Environment objective, Slope attribute is given three times more importance than the Views attribute. Therefore the weight of the Environment objective will be accordingly divided into Slope and Views attributes.

6.6.1 Application

The same relative importance is assigned to the objectives and attributes in the Pairwise Comparison tool developed in the AHP based site selection tool that converts it into the equivalent percentage of weight for each node. At both levels of the AHP hierarchy, the consistency ratio calculated by the Pairwise comparison tool is zero, which shows that the calculated weights can be used with confidence.

The AHP hierarchy is shown in Figure 6.5 as processed in the AHP based site selection tool. The relative weights of the objectives and indicators are shown as percentage. This percentage is converted into the actual relative weight of each node during the calculations.

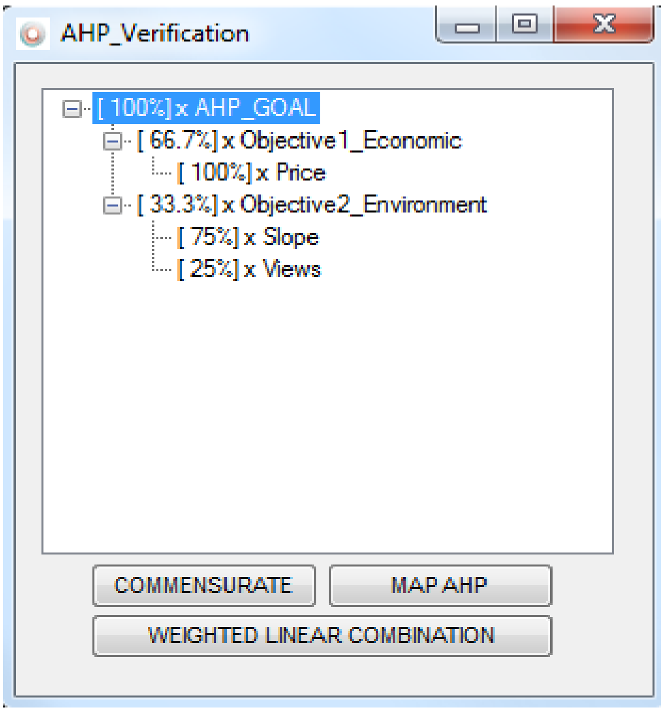


Figure 6.3 AHP hierarchical structure of the site suitability problem

The tool provides multiple ways of scaling the data as explained in Section 5.4.1.1. Maximum Score Procedure is used to scale the data in the site suitability problem in (Malczewski 1999). Therefore the same procedure is used to commensurate the data in order to compare the two results as shown in Figure 6.4. It is noted that all three attributes Slope, View and Price are "Cost" in nature, i.e. the less the better.

Weighted Linear Combination is applied in the tool to the AHP tree structure from leaf nodes to the root node. The final rating of the parcels is divided by the sum of all three ratings to assign a standardised rating to each parcel. Finally the standardised rating is converted into Ranks.

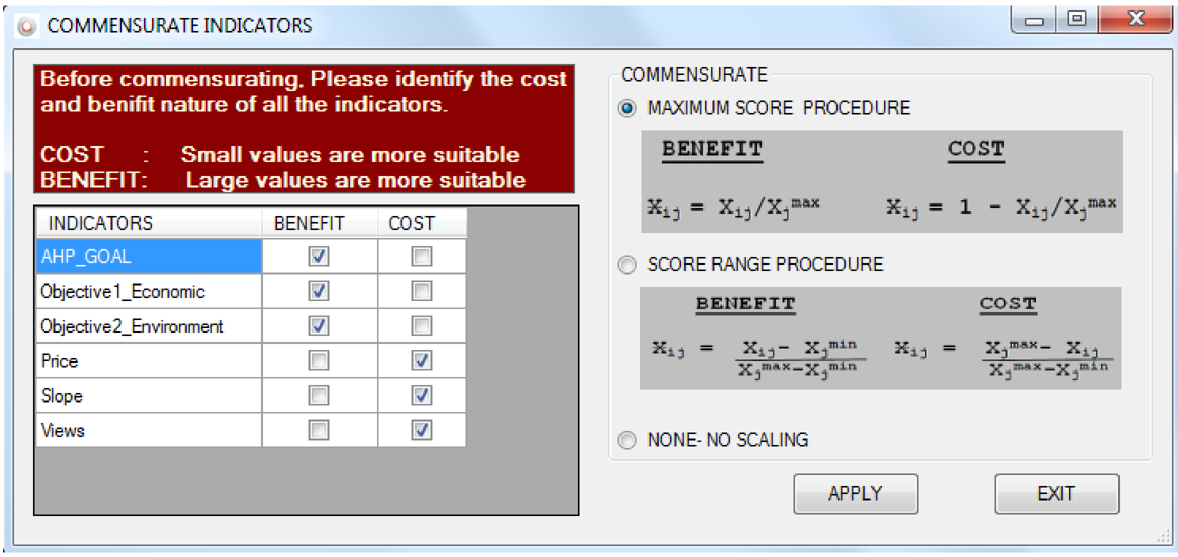


Figure 6.4 Commensuration process - AHP based site suitability tool

6.6.2 Comparison of results

The results generated by the tool and those given in (Malczewski 1999) are summarised in the Table 6.6 for comparison. The tool has produced exactly the same ranks for the three parcels as given in (Malczewski 1999).

The fractional difference in the standardised value is due to the fact that the weights are processed in slightly different ways as explained above. Also the rounding-off the decimal numbers at each calculation can result in fractional differences in iterations and is reflected in the final standardised values.

Table 6.6 Comparison of site suitability results

Parcel of Land	Rating Value	Standardised Value	Ranks
Site Suitability Results (Malczewski 1999)			
A	0.839	0.839/(0.839+0.82+0.777) = 0.344	1
B	0.82	0.82/(0.839+0.82+0.777) = 0.337	2
C	0.777	0.777/(0.839+0.82+0.777) = 0.319	3
Site Suitability Results – AHP based Site Selection Tool			
A	87.93	87.93/(87.93+84.14+81.81) = 0.346	1
B	84.14	84.14/(87.93+84.14+81.81) = 0.331	2
C	81.81	81.81/(87.93+84.14+81.81) = 0.322	3

Comparison of the result suggests that the code implemented in the AHP based site selection tool and its associated Pairwise comparison method and commensuration tools give expected results. Therefore the tool can be used for the site selection process with confidence.

6.7 Site ranking by neighbourhood analysis tool

As discussed in Section 5.4.2, a new site ranking method is introduced in the site ranking by neighbourhood analysis tool. The ranking method introduced is called the Criterion Sorting Mechanism (CSM). It is important to verify whether the CSM can rank the sites accurately: a) based on the individual indicators and b) based on the cumulative impact of all the indicators. For the verification of the first part, a testing environment is created where the ranks are already known for the individual indicators and the tool is run to find the ranks. For the second test, a known ranking method is to be used, to cross check the cumulative site ranks generated by CSM. For this purpose the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) method (Hwang and Yoon 1981) is adopted.

6.7.1 Data preparation

To verify the working of the tool, the same site suitability example is used as described in Section 6.5. A dummy spatial dataset (Polygon Shapefile) was created for the purpose. It contains three parcels A, B, C with three attributes namely Price, Slope and Views. The data values of the parcel for the three attributes are given in Table 6.5. A fishnet (Polygon Shapefile) of five hundred squared meters was generated over the parcels and attribute information was assigned to each cell. A separate site layer (Point Shapefile) was generated over the parcels which contains three sites, one in each parcel.

6.7.2 Application

The data is loaded in the tool in two separate GIS layers a) Fishnet containing indicator dataset and b) Site layer containing three sites, one in each parcel. Maximum score procedure is used to commensurate the data as it is the method used in the site suitability example (Malczewski 1999). All three indicators are "Cost" indicators, i.e. the less the better. A buffer of 3km is applied around the sites for the creation of the neighbourhood around them. Average value of the indicators is used for the ranking purpose using CSM and TOPSIS method.

As depicted in Figure 6.5, the 3km neighbourhood of each site is well within the parcel boundaries so the average value of each indicator is essentially the same as their original value. This approach is adopted as it enables the comparison of the ranks generated by the CSM technique with those already known. Since all three indicators are "Cost" in nature, therefore the Site with rank-1 should be site with minimum value of the individual indicator under consideration. On the other hand, the individual indicator's rank 3 should be assigned to the site having the highest value of the indicator. The values of each indicator in the parcels are shown as percentage of total in the map to the left in Figure 6.5. Sites with 3km radius neighbourhood and intersected Fishnet cells are shown in the map to the right in Figure 6.5.

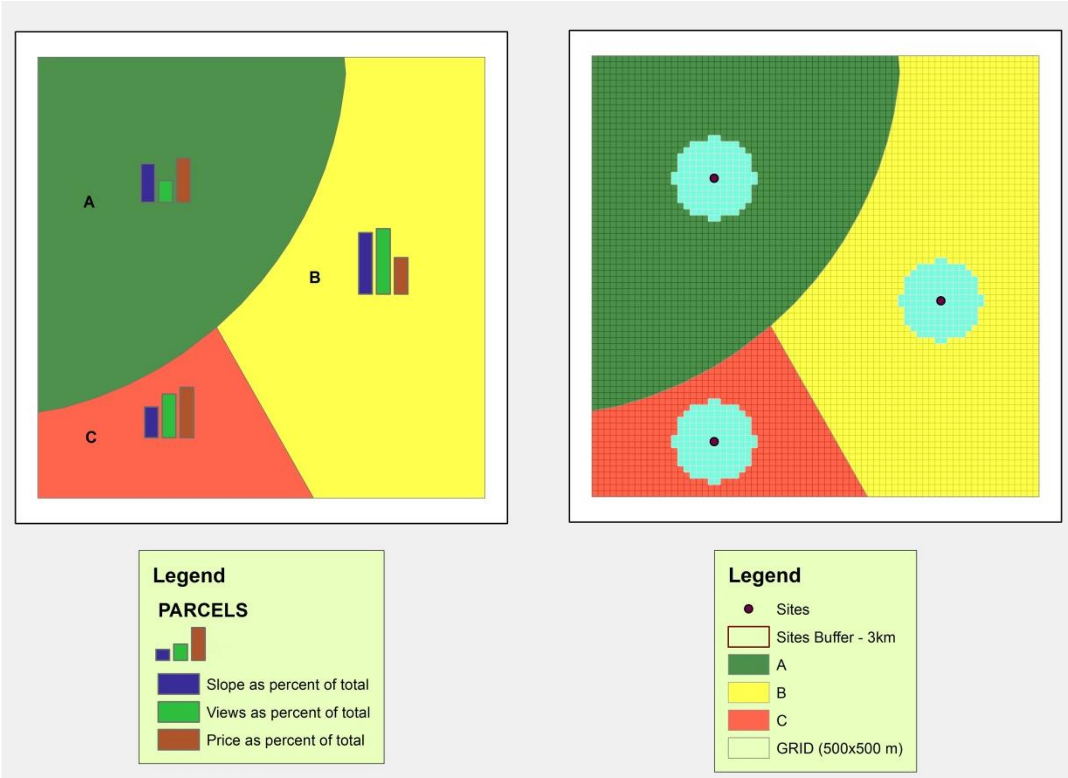


Figure 6.5 Site suitability based on neighbourhood analysis of key indicators

For the cumulative ranks, the results of CSM method cannot be compared to those generated by the AHP technique in Section 6.6 and given in (Malczewski 1999). This is because the weights assigned to the attributes and to the objectives at different levels of the AHP tree structure. Therefore cumulative results of CSM technique are compared with those generated by the TOPSIS method.

6.7.3 Comparison of results

Two types of ranks are generated for each site using the CSM method i) Site ranks based on each indicator and ii) site ranks based on all indicators. These ranks are shown in Figure 6.6. The individual ranks are compatible with the known ranks. Site A has the lowest value for "*Views*" indicator hence it is assigned rank 1. Site B has the lowest value for the "*Price*" indicator and therefore it is assigned rank 1 for this indicator. Site C has the lowest value for the "*Slope*" and therefore it is assigned rank 1 for this indicator.

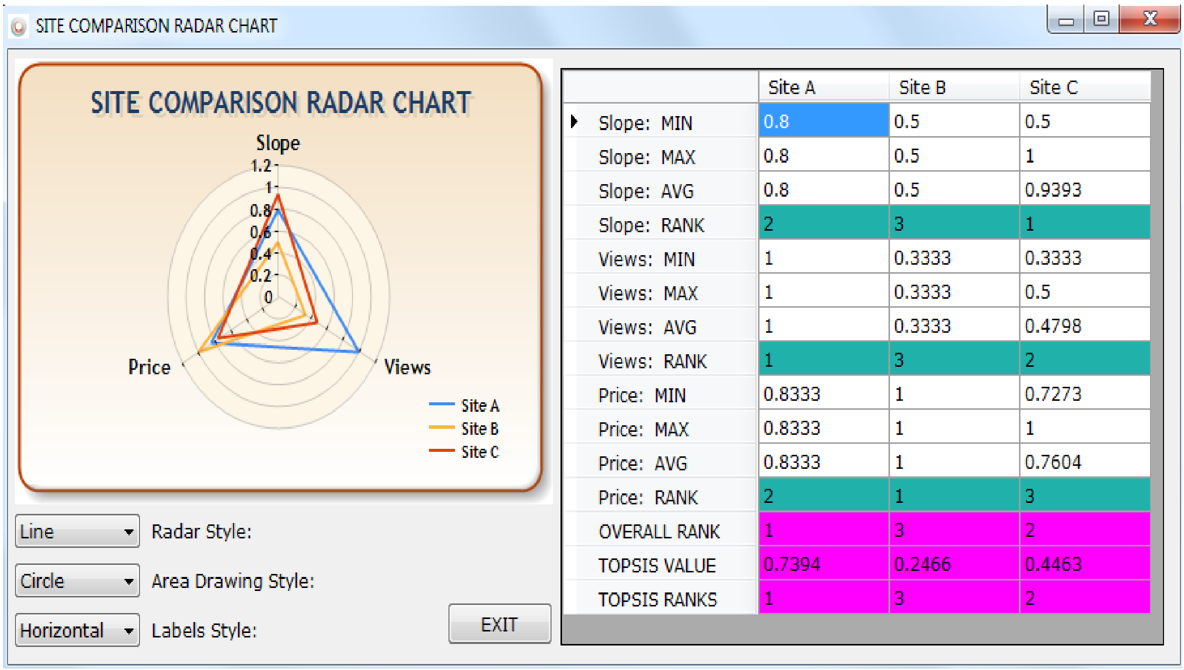


Figure 6.6 Site ranking results of CSM and TOPSIS methods

Looking at the cumulative results, both CSM and TOPSIS have assigned exactly the same ranks to the site. These results are not comparable to those produced by the AHP method because there are no relative weights involved in the analysis and the cumulative ranks are based on the ranks assigned against individual indicators. However, Site A has been assigned the cumulative rank 1 by all three methods: AHP, TOPSIS and CSM. If the weights are assigned to the indicators in CSM and TOPSIS method, the results would become comparable. The site ranking by neighbourhood analysis tool is developed to be used at the second level of the site selection process where decision maker is left with multiple equal potential sites after the first level of site selection process using techniques such as AHP.

6.8 Conclusions

This chapter covers the verification of different analytical modules developed in the SDSS. The modules verified are those based on AI techniques such as ANN or GA and also those based on MCDA techniques such as AHP and WLC.

To verify these tools, reliable software such as Matlab has been used to compare the results while solving the same problem using the same dataset and essential parameters. A systematic verification approach has been adopted in all the tests. First, the tool and its features to be verified have been explained along with the selection of the reliable alternate software that is to be used for the verification purpose. Secondly, a dataset is generated to be analysed by the two tools. The dataset is analysed using the two tools with the same parameters and problem structure. In the end, the results are compared and similarities and dissimilarities are highlighted and explained.

A critical comparison of the results has been made after each test to analyse any differences and possible reasons are discussed. It is however noted that all the results are closely matched and confirmed the accuracy of the code developed for different tools in the SDSS. The AHP and WLC based site selection tool provided results as expected and can be used for the first level of site selection, incorporating the key indicators with further confidence. The site ranking tools based on either Self-Organizing Maps or neighbourhood analysis, gave reliable ranking results and therefore can be used to further reduce the number of potential sites in the second level site selection process.

Based on the results of verifications carried out, the GRNN based regression and prediction tool has provided reliable results and the “Holdout” method and GA proved to be useful in the selection of the most appropriate Sigma parameter for the GRNN network. The SOM based correlation finding tool proved to be able to find correlation between dependent and independent variables from the Best Matching Units (BMU) of the SOM output map rather than analysing the entire dataset. It was also verified that the SOM based tool was able to find correlation between the dependent and independent variables even after adding artificial random noise to the dataset.

The verification process has shown the reliability of the underlying methods used and further developed in this research. The work has increased the confidence in the usability of the SDSS developed in this research, to facilitate the considered spatial decision problems as explained in Chapter 2.

References

Bação, F.,Lobo, V. and Painho, M. eds. 2005. *Geo-SOM and its integration with geographic information systems*. WSOM 05, 5th Workshop On Self-Organizing Maps, Paris.

CACI. 2012. *CACI-PayCheck household income* [Online]. Available at: http://www.caci.co.uk/location-planning/datadepot/datasets/paycheck [Accessed: 07 May].

Hwang, C. L. and Yoon, K. 1981. *Multiple attribute decision making—Methods and applications*. Heidelberg: Springer-Verlag.

Malczewski, J. 1999. *GIS and Multicriteria Decision Analysis*. Wiley.

Mathworks. 2013. *Generalized Regression Neural Networks* [Online]. Available at: http://www.mathworks.co.uk/help/nnet/ug/generalized-regression-neural-networks.html [Accessed: April 16].

WIMD. 2011. *Welsh Index of Multiple Deprivation* [Online]. Available at: http://wales.gov.uk/topics/statistics/theme/wimd/?lang=en [Accessed: 14 October].

7

GEODATABASE

7.1 Introduction

This chapter covers the design and development of Geodatabase that serves as an essential component of the spatial decision support system. The SDSS has been designed and developed independently of the study area, however to demonstrate its application; Wales (UK) has been selected as the study area. As discussed in Chapter 2, potential for the exploitation of unconventional gas resources has been identified in Wales. This natural resource can be exploited to meet the growing energy demands using indigenous resources. Jones et al. (2004) mentioned a considerable CBM and ECBM potential in North and South Wales coalfields.

As discussed in Chapter 2, there are certain environmental, socio-economic, public health and techno-economic aspects to be considered in the decision making context related to the exploitation of unconventional gas. Therefore, key datasets and indicators for the study area (Wales) have been identified and acquired from various sources for this research.

Section 7.2 highlights the four main domains within the geodatabase: i) Socio-Economic domain, ii) Environmental domain, iii) Public Health domain and iv) Techno-Economic domain.

As explained in Section 3.4, an OGC compliant SpatiaLite technology has been adopted for the development of geodatabase. SpatiaLite is a single file, light weight and open source database that can store, manipulate, index and query both spatial and aspatial data (SpatiaLite 2014).

The datasets acquired for Wales, are in different scales and units. In order to bring various datasets together a Fishnet (vector grid Shapefile) has been created over the study area. The Fishnet has been generated in ArcGIS on British National Grid reference system with a cell size of $500m^2$. There are a total of 86860 cells, covering the entire on-shore area of Wales. Each Fishnet cell is then populated with all the key indicators using different GIS analysis performed using ArcGIS10 software (ESRI 2014). The data used and type of the GIS analysis performed for each key indicator, are explained in detail in section 7.3 to section 7.6. Fishnet is stored as a GIS layer in the geodatabase with the original units of measurement for each indicator. These different units can be scaled between 0-1 by using the commensuration process as explained in Section 5.4.1.1.

7.2 Geodatabase Domains

As discussed in Chapter 2, the key indicators covering the four domains are essential for informed risk based spatial decision making process involved in the Geoenergy and Geoenvironmental problems. These domains are i) Socio-Economic domain, ii) Environmental domain, iii) Public Health domain and iv) Techno-Economic domain. The combination of all four domains in the geodatabase serves as the key data backbone of the SDSS. The analytical modules discussed in Chapter 4 and 5 utilise this information to facilitate informed risk based spatial decision making process.

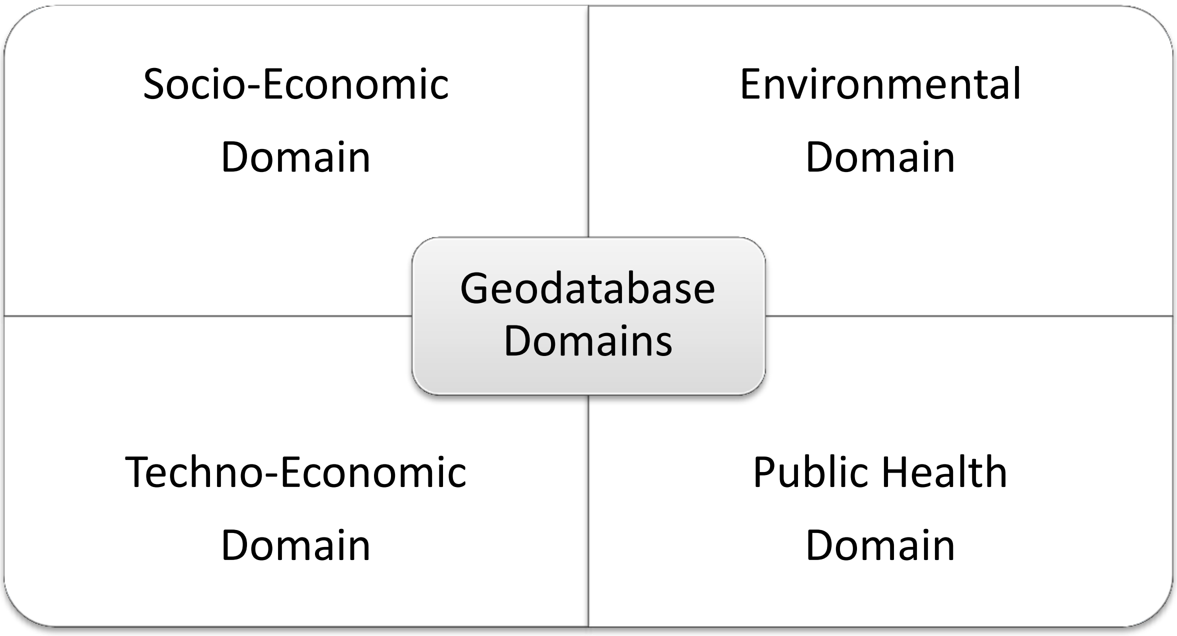


Figure 7.1 The four generalised domains of the datasets incorporated in the geodatabase

Some of the indicators are directly incorporated in the geodatabase while others are used in the GIS modelling to construct composite indicators such as Social Capital and Social Acceptance. GIS modelling is carried out in ArcGIS 10 software using different techniques as explained in sections below.

Socio-economic indicators reflect the current state of socio-economic condition of the communities. This domain covers unemployment, income level, deprivation, access to facilities and employment rates by industry type. These socio-economic indicators can be incorporated in the decision making process to give priority to those areas where the development of new technologies can help in socio-economic uplift of the area through job creation and business generation. Some of these indicators are used for the estimation of public acceptance of the engineering interventions in their vicinities.

Techno-economic dataset covers the resource estimation and other geotechnical aspects that are important for site selection related to a particular technology. It also covers the proximity of consumers (domestic and commercial), CO_2 emitters and the existing gas and electric infrastructure of the national grid to the potential Geoenergy resources in particular unconventional gas.

Environmental Domain covers the key indicators to reflect the state of environment, e.g. air quality, soil quality and ground water quality etc. These indicators not only show the current state but are also used as a measure of the fragility of certain areas where key environmental parameters are already under stress. Also the proximity of each Fishnet cell from the strategic environmental areas is calculated.

Public health indicators reflect the spatial variation in the state of public health in study area. These indicators cover disease, mortality and hospital admission rates caused by key illnesses. These indicators can be used to estimate a population's fragility or capability to face the environmental challenges.

7.3 Socio-Economic Domain

This section presents the indicators related to the socio-economic domain, incorporated into the geodatabase. Based on the literature review (Chapter 2), a number of key socio-economic indicators are identified as shown in Figure 7.2. These indicators can be incorporated into spatial decision making for the effective exploitation of unconventional ground source energy resources in Wales e.g. CBM, ECBM, UCG, CCS and Shale gas.

Some of the indicators are directly incorporated into the geodatabase while others required pre-processing and the application of appropriate GIS modelling techniques. In the latter case, composite indicators are developed such as social capital and public acceptance. Also, information extracted from relevant surveys, is interpolated across the study area using appropriate GIS modelling techniques.

In the former case, well established indicators are acquired and directly incorporated into the geodatabase under different domains. Welsh Index of Multiple Deprivation (WIMD) is one such dataset that is used directly under different headings. WIMD is an official measure of the multiple deprivations (lack of opportunities and resources) faced by the Welsh population (WIMD 2011b). WIMD is measured at the small area level, i.e. the Lower Super Output Areas (LSOA) with a mean population of about 1500 people. WIMD assigns a deprivation index to each LSOA with respect to eight different domains: income, employment, health, education, geographical access to services, community safety, physical environment, housing and an overall index combining all domains (WIMD 2011b). In 2011 there were 1896 LSOAs in Wales; therefore the index for each domain is always between 1 and 1896. The lower the index, the higher the deprivation, means that the respective LSOA is more deprived in the given domain as compare to other areas.

Figure 7.2 depicts the hierarchy (domain, indicators, sub-indicators) of the socio-economic domain. This hierarchy is similar to the one used in the AHP based site selection tool of the SDSS as explained in Section 5.4.1.

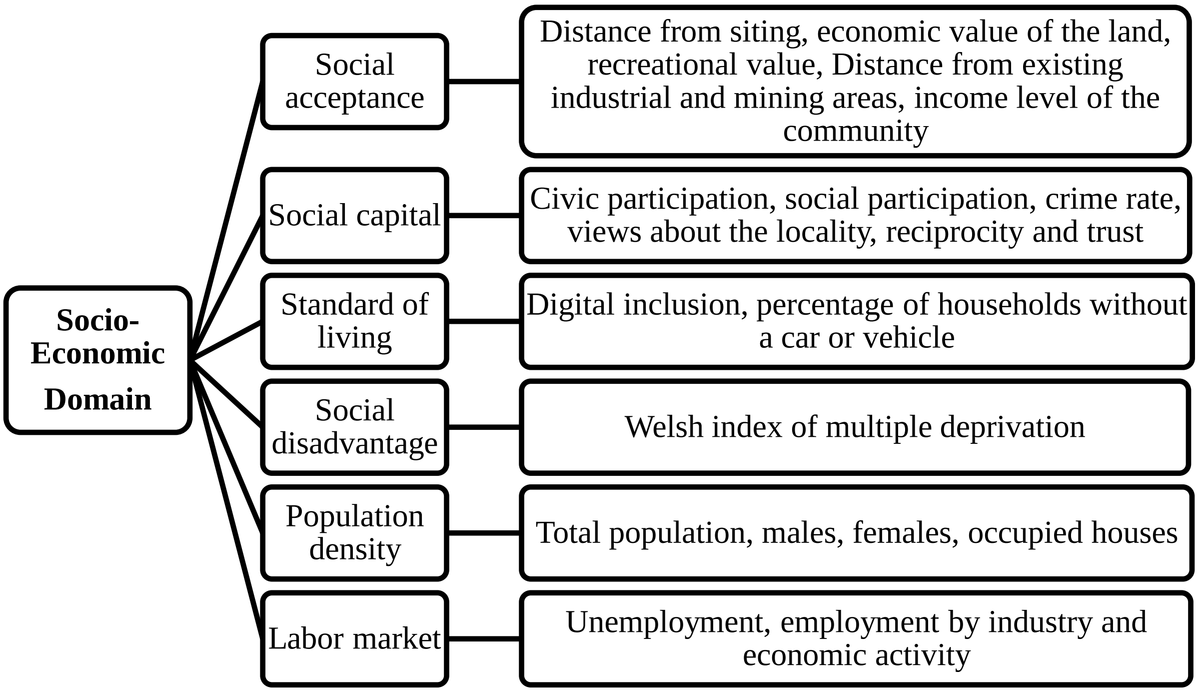


Figure 7.2 Key Socio-Economic indicators and datasets incorporated into the geodatabase

7.3.1 Social Acceptance indicators mapping

In current research the term “social acceptance” is used to represent an anticipated level of general acceptance of the engineering interventions related to the unconventional gas development in potential areas of the resource. This acceptance is generalised for all the stake holders including public, local councils, environmental health organisations and the policy makers. It is based on a number of key indicators that can influence the level of acceptance as discussed in Chapter 2. These indicators can influence the acceptance or opposition for the development of unconventional gas in a given geographical region. Following sections covers these indicators and details of any GIS analysis that has been carried out in this research.

7.3.1.1 Social Acceptance – Distance from siting

Distance of the residential areas and any strategic infrastructure from the proposed site can have a strong affect in shaping the general public acceptance. Communities living in immediate vicinity of the proposed sites may have more concerns for environment, health, safety and socio-economic impacts than those living at a distance. This is based on the NIMBY (not in my back yard) effect as explained by (van der Horst 2007).

For this purpose, Developed Land Use Areas (DLUA) polygon dataset is obtained from Ordinance Survey of Britain (OS). Then "Near" tool of the ArcGIS is used to calculate the distance of each Fishnet cell from its nearest DLUA polygon.

7.3.1.2 Social Acceptance – Total economic value of the land

The Total Economic Value (TEV) covers all the factors of market and non-market value of the area (van der Horst 2007). If the TEV around the proposed site is higher, then it may be more difficult to acquire social acceptance. It may also increase the cost of the project and make it financially less viable. TEV is a complex indicator and it is dependent on multiple factors (van der Horst 2007). Therefore for simplicity, the average household income is used as an indicator of TEV in current research.

For this analysis, the Experian average household income data is acquired at the Lower Super Output Area (LSOA) level (Experian 2012). Average household income for each LSOA polygon is assigned to its centroid by using "Feature-To-Point" tool in ArcGIS and then Fishnet cells are populated using the spatial join in ArcGIS.

7.3.1.3 Social Acceptance – Potential Recreational value

Recreational value of the land is another factor that influences the social acceptance of new engineering interventions in an area (van der Horst 2007). The more the recreational value of a potential sitting area, the less is the chance to get approval from the authorities and the acceptability of the public. Recreational value is inversely proportional to distance of scenic landscape to the populated places and the availability of alternate natural and manmade recreational places in the surrounding (van der Horst 2007).

In this research, the potential scenic and recreational value is calculated using the distance of each Fishnet cell from national parks, woodlands, recreational areas and leisure facilities and

Ramsar sites. These layers have been acquired from Ordnance Survey's Meridian-2 dataset (OS 2012), Corine-2006 land cover dataset of European Environment Agency (EEA 2006) and Countryside Council for Wales (CCW) datasets (CCW 2012). "Union" tool of the ArcGIS is used to combine all these potential recreational areas into one layer. Once these layers are combined, the "Near" tool is used to calculate the distance of each Fishnet cell centroid from the nearest feature in the combined recreational layer.

7.3.1.4 Social Acceptance – Distance from industrial, commercial and mining areas

Distance from exiting or historical industrial/mining areas can play an important role in the public acceptance of similar engineering interventions (van der Horst 2007). Generally the communities living closer to such activities have benefited from job creation and business generation for them and are more likely to accept similar projects in their surroundings. However, this is not always a positive indicator since the environmental health degradation in some industrial and mining areas and the availability of jobs in alternate sectors can also affect the social acceptance (van der Horst 2007). For this analysis the industrial and commercial areas are extracted from Corine-2006 land cover dataset of the European Environment Agency (EEA 2006) and mining areas map from Department of Energy & Climate Change (DECC 2012) are combined together. "Near" tool of ArcGIS is then used to calculate the distance of each Fishnet cell from the nearest feature in the two layers mentioned above.

7.3.1.5 Social Acceptance – Income level of the community

Income level of communities is another important indicator that can influences the acceptance for the engineering interventions conditions (Garrone and Groppi 2012). Generally the low and medium income communities, potentially show more acceptance towards the engineering interventions as they can create opportunities for them to uplift their socio-economic conditions (Garrone and Groppi 2012).

To assess the general trend of income level in the communities falling under each cell of the Fishnet, Welsh Index of Multiple Deprivation (WIMD) (income index) is used. WIMD income ranks for each LSOA polygon is assigned to its centroid by using “Feature-To-Point” tool in ArcGIS and then Fishnet cells are populated using the spatial join in ArcGIS.

7.3.2 Social capital indicators mapping

Social capital is another important indicator of a community that can influence the phenomenon of acceptance, cooperation and involvement of a community in a positive or a negative way. This section covers the GIS analysis and mapping involved in the development of social capital indicators.

The phenomenon of social capital is complex and qualitative in nature and covers a wide range of socio-economic and geographical characteristics. A general assessment of the social capital of a geographical region can be derived using relevant socio-economic and demographic characteristics. In current research a number of indicators have been used to measure the social capital in different geographical regions in Wales such as the crime rate, education, household income, percentage of minorities in the community, voting turnout and participation of the community in decision making at local and national level.

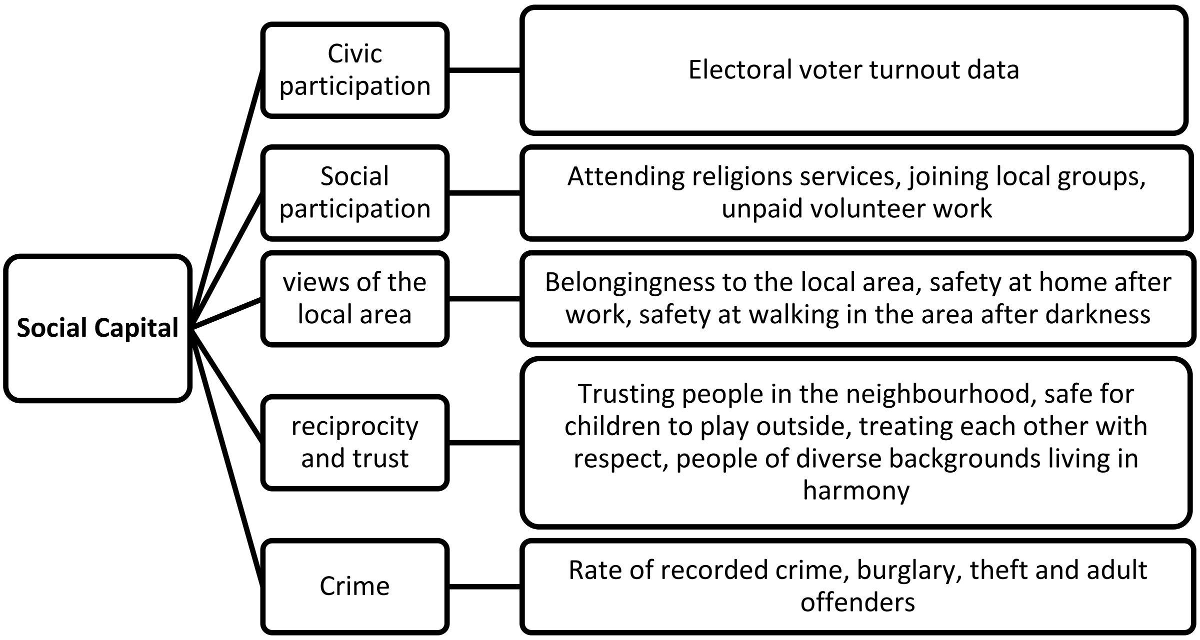


Figure 7.3 Social capital key indicators. Adapted from (Foxton and Jones 2011)

Figure 7.3 shows the factors identified by Foxton and Jones (2011) that can influence the social capital of a geographical region and their associated indicators (variables) as used in this research.

7.3.2.1 Surveys and datasets used for social capital mapping

Social capital is the bonding, bridging and linking relationships between different groups such as professional groups, social groups, virtual groups and communities grouped by geographical regions (Foxton and Jones 2011). In current study only geographical groups are studied and social capital is estimated via key indicators. Office of the National Statistics (ONS) has devised a framework for the measurement of social capital in UK (Foxton and Jones 2011). The key indicators suggested in this report for the social capital measurement are provided in Table A.1 (Appendix A).

In order to map the social capital, a number of indicators have been used along with two survey datasets (a) British Household Panel Survey (BHPS wave 18, 2009) and (b) National Survey for Wales (NSW 2012-13). These surveys contain questions about the neighbourhood, trust, reciprocity and volunteer work that can be used for mapping social capital. BHPS wave 18 is the most recent available data of the series and therefore it is used for the analysis purpose. The Special License Access includes the LSOA number with each survey record that can be used to interpolate and map the survey results at LSOA level. The selected questions from these two surveys that are used in this research are provided in Table A.2 in (Appendix A).

7.3.2.2 GIS analysis and mapping

Survey data can be used to map a particular phenomenon across a geographical region. Since surveys involve the opinion of respondents, therefore their locational information is normally not published along with the survey data. Survey results are normally summarised and disseminated at larger geographic levels like district, county or at national level. In order to map the results of the selected questions from the two surveys, a spatial interpolation technique, i.e. Inverse Distance Weighting (IDW) is used in this research. Similar spatial interpolation technique has been used to map the interrelationship of health and neighbourhood in (Meng et al. 2010). Spatial interpolation techniques like IDW and Krigging were used to obtain small area level variables using the Canadian Community Health Surveys (CCHS) in order to map health indicators at small-scale neighbourhood level (Meng et al. 2010).

Figure 7.4 presents a flowchart showing the general process carried out to map the relevant survey data at the LSOA level in Wales. Records with no answers are removed for the selected questions and average values (survey response) are calculated for each LSOA. Some questions contained only binary answers such as ‘Yes’ and ‘No’ and some contained a range of numbers showing a preference scale such as “best to worst”. An average value of the answer for each question is computed for each LSOA. Then using the spatial join technique these averaged values (of answers) are linked with the centroid of corresponding LSOA.

Inverse Distance Weighting (IDW) is used to spatially interpolate the averaged answer for each question. In this way a surface map is generated to represent the spatial variation in the response for each of the selected survey question. Finally, the Fishnet cells are populated for each map produced using the above defined process.

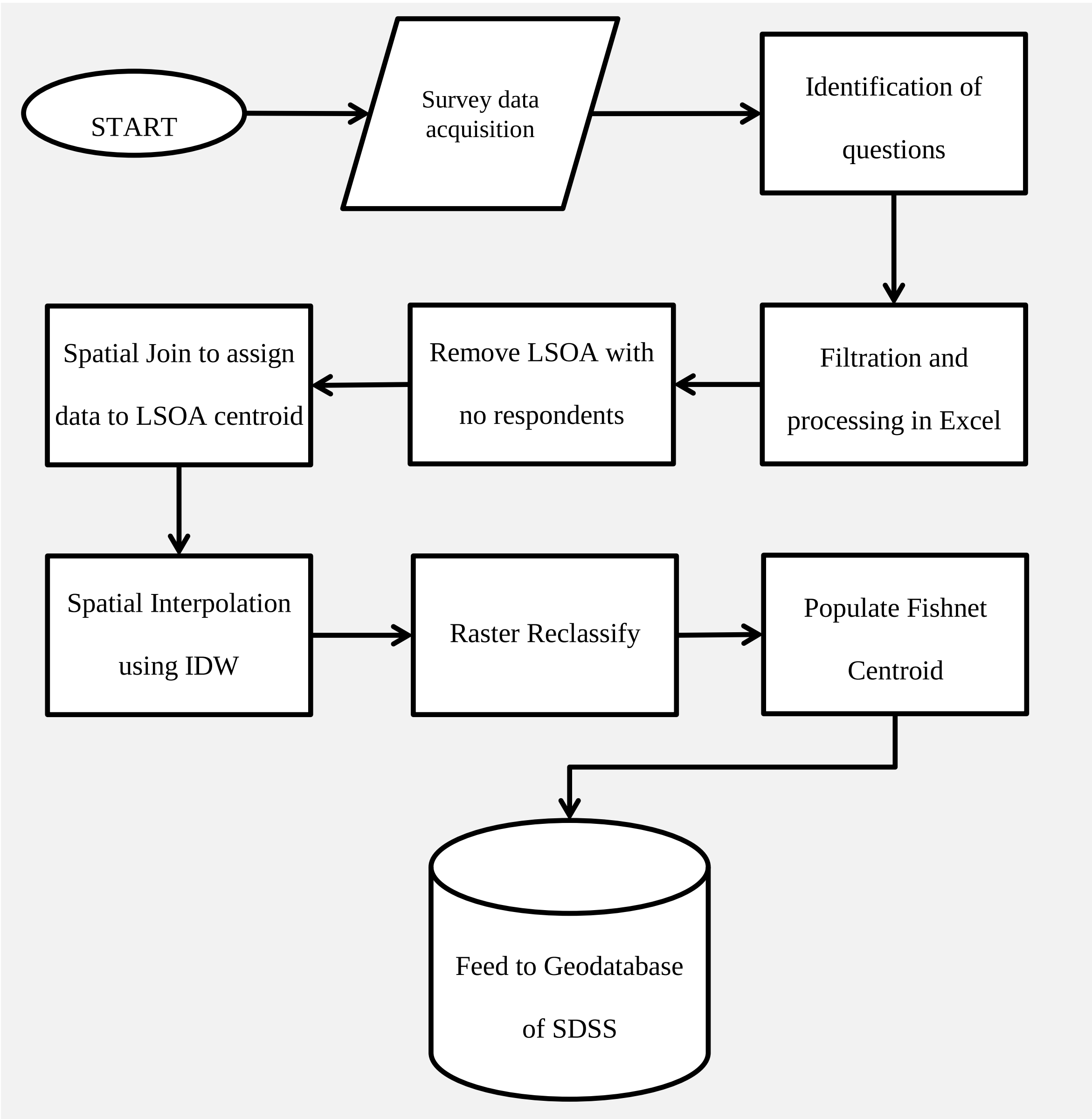


Figure 7.4 Processing and GIS analysis on survey data

7.3.2.3 Social Capital - civic participation

For analysing the civic participation across Wales, voting turnout dataset is used. Voting turnout is a key indicator of the civic participation since it reflects the involvement in local and national affairs, and perceptions of ability to influence them (Foxton and Jones 2011). For this purpose, voter turnout data for 2012 local elections is obtained at the electoral wards level from the Election Centre of Plymouth University (Plymouth-University 2013). Since local elections were not contested in all the wards in 2012, therefore voting turnout from 2011 of the Welsh Assembly election has been used for the missing wards.

7.3.2.4 Social Capital - Social participation

There are no direct questions related to the "social participation" in NSW data. For this reason BHPS survey Wave 18 is selected. A few questions are identified that could be used as sub-indicators for the mapping of "social participation". These questions are:

a. Attends religious services.
b. Attend local group/voluntary organisation.
c. Do unpaid voluntary work.

The process used for mapping the survey data has been described in Figure 7.4.

7.3.2.5 Social Capital - Attends religious services

BHPS wave 18 data is acquired with special license. There are 1417 respondents from Wales in the dataset. 1289 records out of 1417 had response to this question covering 597 LSOAs out of 1896. The result is scaled on 1 to 5 where 1 refers to "Very Frequently" and 5 to "Never".

7.3.2.6 Social Capital - Attends local group/voluntary organisation

BHPS wave 18 data is acquired with special license. There are 1417 respondents from Wales in the dataset. 1289 records out of 1417 had response to this question covering 597 LSOAs out of 1896. The result is scaled on 1 to 5 where 1 refers to "Very Frequently" and 5 to "Never".

7.3.2.7 Social Capital - Do unpaid volunteer work

BHPS wave 18 data is acquired with special license. There are 1417 respondents from Wales in the dataset. 1289 records out of 1417 had response to this question covering 598 LSOAs out of 1896. The result is scaled on 1 to 5 where 1 refers to “Very Frequently” and 5 to “Never”.

7.3.2.8 Social Capital - Views of the local area

The question about the views of the local area is taken from the NSW dataset. NSW has more records as compare to BHPS and is dedicatedly designed for Wales only. Hence it provides an opportunity for more detailed mapping. The following three questions are identified from the survey as sub-indicators of “views of the local area”:

- Belonging to the neighbourhood.
- Safety at home after work.
- Safety walking in the local area after dark.

7.3.2.9 Social Capital - Belonging to the neighbourhood

This question is selected from the NSW dataset. There are 14552 records in the dataset. Similar approach is used as described in previous section to map the survey results on this question across Wales. 14481 records out of 14552 had response to this question covering 1881 LSOAs out of 1896. The result is scaled on 1 to 5 where 1 refers to “Strongly Agree” and 5 to “Strongly Disagree”.

7.3.2.10 Social Capital - Safety at home after dark

This question is also selected from NSW dataset and the same procedure is applied to map it across Wales. 144537 records out of 14552 had response to this question covering 1881 LSOAs out of 1896. The result is scaled on 1 to 4 where 1 refers to “Very Safe” and 4 to “Very Unsafe”.

7.3.2.11 Social Capital - Safety walking in the area after dark

This question is also selected from National Survey of Wales and same procedure applied to map it across Wales. 144287 records out of 14552 had response to this question covering 1879 LSOAs out of 1896. The result is scaled on 1 to 4 where 1 refers to “Very Safe” and 4 to “Very Unsafe”.

7.3.2.12 Social Capital - Social network and social support

To map social network and social support the questions “Are people willing to help their neighbours in the area?” is selected from the NSW dataset. There are 14414 records out of 14552 with a valid response to this question covering 1880 LSOAs out of 1896. The result is scaled on 1 to 5 where 1 refers to “Strongly Agree” and 4 to “Strongly disagree”. The same procedure is used to interpolate the data as used in previous sections.

7.3.2.13 Social Capital - Reciprocity and trust

For the purpose of mapping this indicator of social capital, NSW (National Survey of Wales) 2012-13 data is acquired from Welsh Government with special license to have LSOA linkage. NSW has more records as compare to BHPS and is dedicatedly designed for Wales only. Hence it provides an opportunity for more detailed mapping. The following three questions are identified from the survey as sub-indicators of “view of the local area”:

a. Trusting people in the neighbourhood.
b. Safe for children to play outside.
c. People from different background get on well together.
d. People treating each other with respect and consideration.

7.3.2.14 Social Capital - Trusting people in the neighbourhood

This question is also selected from National Survey of Wales and this same procedure is applied to map it across Wales. 13974 records out of 14552 had response to this question covering 1878 LSOAs out of 1896. Records with 5 “just moved to the area” are also removed. The result is scaled on 1 to 4 where 1 refers to “Many people in the neighbourhood can be trusted” and 4 to “None of the people in the neighbourhood can be trusted”.

7.3.2.15 Social Capital - Safe for children to play outside

This question is also selected from NSW dataset and this same procedure is applied to map it across Wales. 14334 records out of 14552 had response to this question covering 1881

LSOAs out of 1896. The result is scaled on 1 to 5 where 1 refers to "Strongly agree" and 5 to "Strongly disagree".

7.3.2.16 Social Capital - People from different backgrounds get on well together

This question is also selected from National Survey of Wales and this same procedure is applied to map it across Wales. 13368 records out of 14552 had response to this question covering 1881 LSOAs out of 1876. The result is scaled on 1 to 5 where 1 refers to "Strongly agree" and 5 to "Strongly disagree".

7.3.2.17 Social Capital - People treating each other with respect and consideration

This question is also selected from National Survey of Wales and same procedure applied to map it across Wales. 14442 records out of 14552 had response to this question covering 1880 LSOAs out of 1876. The result is scaled on 1 to 5 where 1 refers to "Strongly agree" and 5 to "Strongly disagree".

7.3.2.18 Social Capital - results verification

Social capital is a qualitative term and it is not straight forward to measure it directly hence it should be treated in a latent context. It is important to verify the usefulness and effectiveness of the indicators and sub-indicators created above using primary, secondary data and GIS analysis. For this purpose a comprehensive literature review is carried out. There are some positive outcomes of the social capital on the overall social fabric. Higher social capital should theoretically contribute to the better standard of living and satisfaction level amongst the residents of a community. Different authors have analysed and suggested the various positive outcomes of social capital in different places. A study has been conducted in Netherlands to observe the effects of social capital on crime in the country. The results suggest that the social capital provides an informal network support for crime prevention. The communities in Netherlands with higher level of social capital have relatively lower crime rate accordingly (Akçomak and ter Weel 2012). Another study conducted in Japan also reveals similar results about the inverse relationship of social capital and crime (Takagi et al. 2012).

In order to investigate this causality-relationship in the current research, GIS analysis has been performed to verify if similar trends between crime and social capital exists in Wales. For this purpose the AHP based site selection tool of the SDSS is used. AHP tool has been explained in Section 5.4.1. First the Crime data is selected-out of the AHP so that it does not contribute in the mapping of social capital. The WLC is then applied on the AHP hierarchy across Wales and the Social Capital map is exported. Top and bottom 20 % cells (with respect to the social capital and crime data) are then exported as separate layers. There are 17372 cells in each of the four layers (20% of 86860). The “Intersection tool” in ArcGIS is then used to select only the following areas:

- Top 20% cells of social capital intersecting top 20% cells of crime.
- Bottom 20% cells of social capital intersecting bottom 20% cells of crime.
- Top 20% cells of social capital intersecting bottom 20% cells of crime.
- Bottom 20% cells of social capital intersecting top 20% cells of crime.

The results are shown in Table 7.1 where the two intersecting layers are given for each analysis. The number or intersected cells (out of 17372) are shown along with the percentage of cells being intersected. The main objective of the analysis is to check whether the best areas in terms of social capital coincide with low crime rate and the worst areas in terms of social capital coincide with the high crime rates. The results can be used to test the causality-relationship in Wales but also to verify the applicability of the social capital mapping method adopted in the SDSS.

Table 7.1 The results of social capital and crime rate causal analysis

Layer 1	Layer 2	Intersected cells	%
Social Capital : Top 20% cells	Crime Rate : Top 20% cells	4781	27.5 %
Social Capital : Top 20% cells	Crime Rate : Bottom 20% cells	1524	8.77 %
Social Capital : Bottom 20% cells capital	Crime Rate : Top 20% cells capital	1562	8.99 %
Social Capital : Bottom 20% cells capital	Crime Rate : Bottom 20% cells capital	7230	41.61 %

Top 20% of the crime rate cells are those cells where crime rate is low because this indicator is selected as the "COST" indicator in the AHP based site selection toolkit. Similarly the bottom 20 % of the crime rate cells is representing those areas where the crime rate is the highest.

On the other hand, the social capital is used as a "BENEFIT" in the AHP scheme. It means the top 20 % cells are representing those areas where the social capital is high. As it can be seen in Table 7.1, the results are quite promising even with using a crude method, i.e. interpolation for mapping the survey data with a small number of respondents. As shown in Table 7.1, areas with high crime rate coincide with the areas with very low social capital. Areas with low crime rate coincide with the areas having high social capital. The percentage of such areas is found to be very low where the results are opposite to this trend. These findings are very similar to those discussed by Akçomak and ter Weel (2012) and by Takagi et al. (2012) as explained earlier in this section.

Since the crime rate itself is a key indicator of the social capital, therefore it is understood that the method used here is robust and with the inclusion of crime rate, the results will be even more reliable.

Another technique is also used to verify the method of estimation for the social capital in Wales, by using the WIMD indicators. In Wales, the WIMD is an official mean of ranking of areas in terms of multi-faceted deprivation. The highest rank (1) suggests that the area is most deprived in the given domain. An overall index of multiple deprivation is also included that shows a cumulative effect of all the deprivations. These indexes are established at the LSOA level which is considered as “small area” in terms of population (roughly 1500 per LSOA). WIMD is constructed from eight different types of deprivation, i.e. income, housing, employment, access to services, education, health, community safety, physical environment and an overall index (WIMD 2011a).

In order to verify whether the derived social capital mapping is in line with the literature, a series of maps are created in ArcGIS showing whether high social capital areas coincide with the least deprived areas or not. To check this, “select by location” and “Intersection” tools in ArcGIS are used to produce the following maps:

- Top 20% cells of social capital intersecting with top 20% most deprived areas (WIMD-Overall Rank).
- Top 20% cells of social capital intersecting with top 20% most deprived areas (WIMD-Education Rank).
- Top 20% cells of social capital intersecting with top 20% most deprived areas (WIMD-Income Rank)

Table 7.2 presents the results of the analysis carried out to check the coexistence of high social capital values and low multiple deprivation. Results show that high social capital coexist with low multiple deprivation (WIMD overall index). Similarly, 63.81 % of area has coexistence of high social capital and low income deprivation. Similarly, over 45 % of the area has shown coexistence of high social capital and low education deprivation. This is according to the facts mentioned in Literature Review (Chapter 2). However the percentage of the top cells in terms of social capital coexisting with the top cells in terms of WIMD-overall index is low, i.e. 15 % which requires further investigation.

Table 7.2 Spatial Coexistence analysis results of social capital and WIMD indices

Layer 1	Layer 2	Intersected cells	%
Social Capital : Top 20% cells	Top 20 % WIMD (Overall index)	2630	15.14 %
Social Capital : Top 20% cells	Top 20 % WIMD (Education index)	7838	45.11 %
Social Capital : Top 20% cells	Top 20 % WIMD (Income Index)	11086	63.81 %

7.3.3 Social Capital - Crime

For mapping crime rates across Wales two datasets are used: i) Rate of recorded criminal damage per 100 people of the daytime population (2008-2010) and ii) the rate of adult offenders per 100 people of the daytime population (2008-2010). These indicators are acquired at the LSOA level and assigned to Fishnet cells.

7.3.4 Population density and occupied dwellings

It is important to have records of the occupied dwellings and population density in the surrounding areas of a proposed site. This is helpful in planning for the development of additional infrastructure that will be required, e.g. to overcome the load of migration workers etc. Also it is important for the contingency planning in the event of a hazard. For this purpose two datasets are acquired: a) Census 2011 Population Estimates (ONS 2011) and b) Code-Point ® Open (OS 2014).

Census 2011 Estimates, contain information about the expected number of occupied houses and number of male and female population for every postcode in the country. Whereas the Code-Point ® Open is Ordinance Survey (OS) data that contains precise location of 1.7 million postcodes in Britain (OS 2014). These two datasets are combined in ArcGIS using join attributes tool. Then spatial join tool is used to summarise this information into each Fishnet cell. This gives accurate information about expected number of male and female population and occupied dwellings in each cell.

7.3.5 Social disadvantage

Social disadvantage is presented as an umbrella term in this research that includes poverty, exclusion and deprivation. It is an important indicator as it can be used to identify the areas where multiple deprivations exists and new unconventional geo energy resources can be exploited to help reducing the deprivation in such areas. This can be done by supporting the local economy, creating new jobs and building new infrastructure. Welsh index of multiple deprivations is a reliable source of data developed and used by the Welsh Government (WIMD 2011a). It has eight domains covering different aspects of deprivations and a cumulative index reflecting the overall deprivation of an area at LSOA level. WIMD deprivation domains are income, employment, health, education, geographical access to services, community safety, physical environment and housing (WIMD 2011a).

7.3.6 Standard of living

Standard of living is used as an umbrella term in this research to present indicators that reflect the facilities, wealth and necessities available to individuals and communities. It can be used in different analysis in AHP based site selection toolkit. For the geodatabase, two sub-indicators have been acquired at the LSOA level: a) “Wales’s digital inclusion” and b) the “Percentage of households with a car or van”. The digital inclusion data of the Welsh Government is to assess the percentage of population in Wales digitally connected and to identify any relationships between digital engagement and socioeconomic and demographic characteristics (Welsh-Government 2012).

7.3.7 Labour market

The labour market data is an important part of the socio-economic domain. It can be used in the AHP based site selection toolkit in order to find the areas where a particular type of skilled labour is available, e.g. mine workers. the labour market data can also be used to identify the deprived areas where a large percentage of people are without jobs. These areas would be given priority over others if the resource potential exists for the applications of unconventional energy. The attraction of new jobs and economic activity will also create a positive environment for a general public perception and acceptance of engineering interventions. For the geodatabase two sub-indicators have been included, i.e. a) Economic activity and b) Employment by industry. These two datasets have been acquired from census 2011 datasets at the LSOA level.

7.4 Techno-Economic Domain

The techno-economic domain contains technical dataset for resource estimation and feasibility of an engineering intervention in terms of technical and economic parameters. This domain contains datasets which are used to locate potential sites for Geoenergy resources in Wales based on its feasibility in terms of geological, geographical, topographic and economic parameters. Figure 7.5 shows the hierarchy of data elements used in the techno-economic domain. The main elements consist of the unconventional gas resource estimation, geological parameters, terrain and site economic parameters.

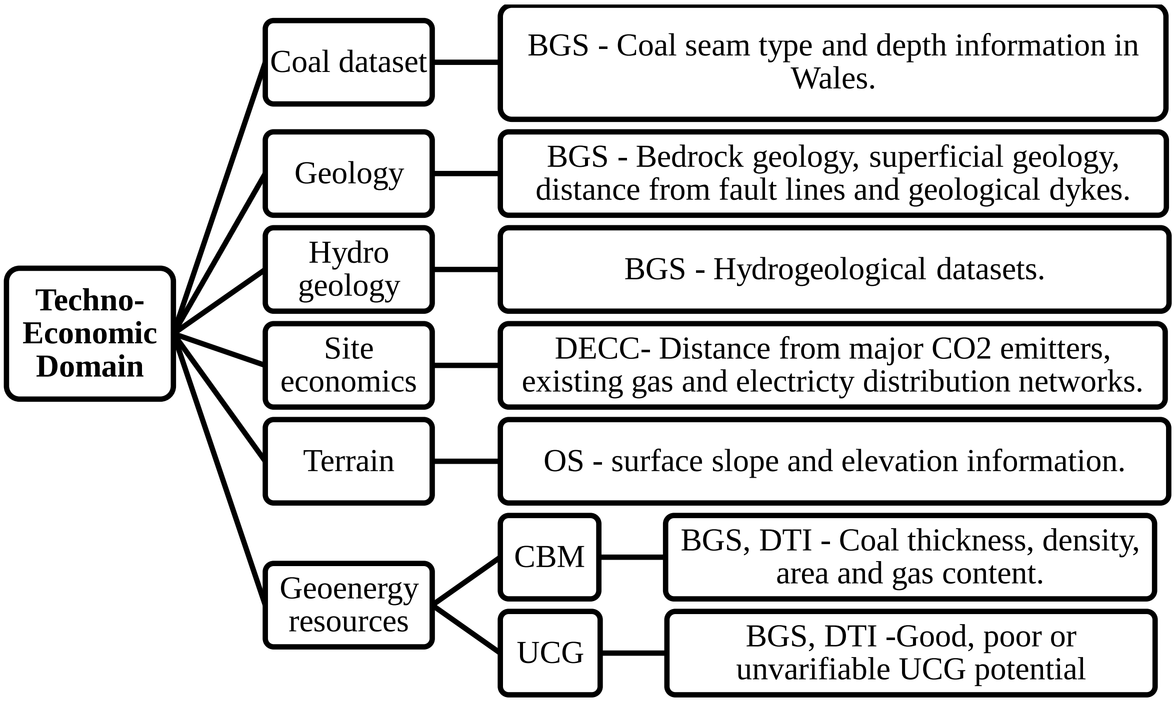


Figure 7.5 Key Techno-Economic indicators and datasets incorporated into the geodatabase

7.4.1 Geoenergy resources

For geodatabase, two geoenergy resources are considered, i.e. Coalbed Methane (CBM) and Underground Coal Gasification (UCG). These two datasets are used in the report " UK coal resources for new exploitation technologies" (Jones et al. 2004) and have been acquired from BGS with a special licence for academic use only. Although the decision analysis performed in SDSS developed in this research adopt a data centric approach but the system itself is independent of the data and the study area. Therefore, in future new energy resources can be incorporated easily, subject to the availability of the data. Shale gas can be one of the other important and potential Geoenergy resources in Wales, to be considered and included in the geodatabase in future.

7.4.1.1 Coalbed Methane (CBM)

The CBM data is comprehensive and contains all necessary parameters required for the CBM resource estimation. It is the same data produced and published in a report by the Department of Trade and Industry (DTI) and British Geological Survey (BGS) (Jones et al. 2004). CBM resource estimation ($10^6 m^3$) is carried out using equation 7.1 with the parameters such as coal thickness (m), density (g/cm^3), coal seam area (m^2) and its average gas content (m^3/tonne) (Jones et al. 2004). To calculate the resource density (10^6m^{3}ha^{-1}) from the resource estimation, equation 7.2 can be used.

$$CBM\ Resource = Area \times Clean\ coal\ thickness \times Average\ methane\ value \times Average\ coal\ density \quad (7.1)$$

$$Resource\ Density = (CBM\ resource\ /\ Area)\ /\ 100 \quad (7.2)$$

where clean coal thickness (meters) is total thickness of coal meeting the criteria minus 15% ash and dirt allowance, area in square meters, average methane value is in cubic meters per tonne and average coal density is in gram per cubic centimetres. There are some other important GIS criteria considered by the BGS and DTI in the preparation of this dataset which is discussed in Section 2.4.

The south and north coalfields in Wales are divided into sub categories based on the analogy of the parameters, e.g. average coal thickness, coal density and gas content. The South Wales Coalfield has the highest seam methane contents in the UK, reaching over 22 cubic meters per tonne (Jones et al. 2004). There is also a significant total thickness of coal meeting the criteria. These two properties make it an excellent resource for potential CBM applications. However, the extent of previous underground workings in coal seams could be a major hurdle in South Wales Coalfield. The areas in northwest of the coalfield still have large unmined areas that could offer potential sites for engineering interventions.

Table 7.3 CBM regions in south and north coalfields in Wales (Jones et al. 2004)

DTI CBM Area No.	Coal Thickness (m)		Coal Density (g/cm^3)	Gas Content (M^3/Tonne)
	Total coal	Clean coal		
		SOUTH WALES COALFIELD		
1	20.6	17.51	1.33	19-21 (Avg. 20)
2	15.6	13.26	1.33	21-24 (Avg.
3	13.9	11.82	1.33	16-19 (Avg.
4	23.8	20.23	1.33	7-10 (Avg. 8.5)
5	11.5	9.78	1.33	13-16 (Avg.
6	14.5	12.33	1.33	10-13 (Avg.
7	24	20.4	1.33	4-7 (Avg. 5.5)
8	14.31	12.16	1.33	(Avg. 12)
		NORTH WALES COALFIELD		
1	30.2	25.67	1.26	(Avg. 8)
2	4	3.4	1.26	(Avg. 7.1)
3	23	19.55	1.26	(Avg. 8)
4	21.8	18.53	1.26	(Avg. 8)
5	19.4	16.49	1.26	(Avg. 8)

Table 7.3 summarise these parameters for the eight distinct regions in South Wales Coalfields and five distinct regions in North Wales Coalfields as given in (Jones et al. 2004). The resource measured in this way does not imply that this amount of methane can be extracted right away or in near future. This depends on other factors that are favourable for the CBM extraction such as physical properties of coal e.g. coal permeability, technology, factors related to the planning permission.

The Fishnet cells in the geodatabase are populated with these key parameters for CBM resource estimation. The area parameter is however not required because Fishnet cells are equal in area, i.e. 500 m^2. The cells found on the boundaries are likely to give unreliable resource estimation as some of the cells may not fall entirely inside a region. Figure 7.6 shows the map of the CBM resource in Wales as calculated in the Fishnet cells using 7.1.

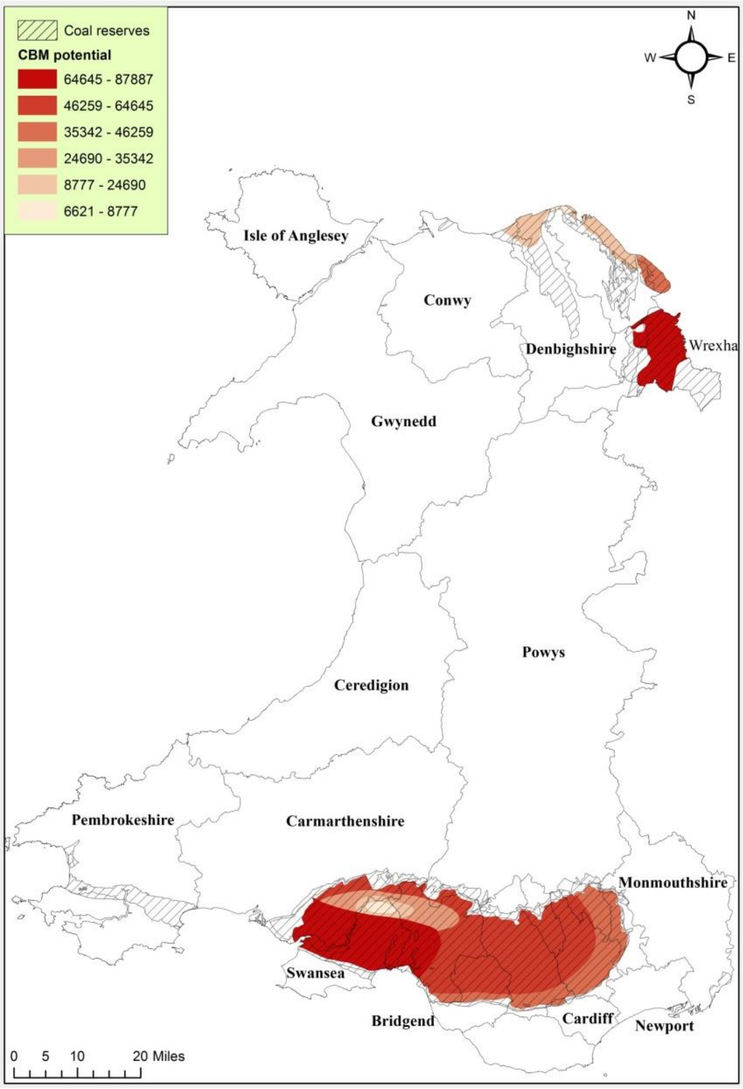


Figure 7.6 Coalbed methane resource potential areas in wales

7.4.1.2 Underground Coal Gasification (UCG)

The UCG dataset provides a qualitative resource potential only. It shows whether the area under consideration has a Good, Poor or Unverifiable UCG resource potential. Unlike CBM, the UCG dataset does not include parameters that can be used directly to roughly estimate the resource potential. However the AHP based site selection toolkit implemented in the SDSS, facilitates the use of qualitative dataset. Every discrete class in a qualitative dataset has assigned a weight separately, sum of which is equal to one for all the classes in that qualitative dataset. This is slightly different to the quantitative datasets used in the AHP process, where each dataset has assigned only one weight, sum of which is equal to one for all the sibling datasets at the same level of the tree and feeding into the same parent node of the tree. This way a combination of both quantitative and qualitative datasets can be used in the AHP base site selection toolkit. GIS modelling criteria adapted by BGS and DTI in the UCG resource estimation and mapping as reported in (Jones et al. 2004) is given as under:

- Seams of 2m thickness or greater.
- Seams at depths between 600 and 1200m from the surface.
- 500m or more horizontal and vertical separation from underground coal workings and current coal mining licences.
- More than 100m separation from major aquifers.
- More than 100m vertical separation from major overlying unconformities.

The qualitative UCG resource data is assigned to Fishnet cells as described above and the map is shown in Figure 7.7 with Good, Poor and Unverifiable UCG resource potential in Wales. Areas with “Good” UCG resources are those areas that meet all the criteria given above and high quality and extensive borehole data are available. Areas marked with “Poor” UCG prospects are those where the coal is present within the 600-1200m depth range but its thickness is not more than 2 meters (Jones et al. 2004).

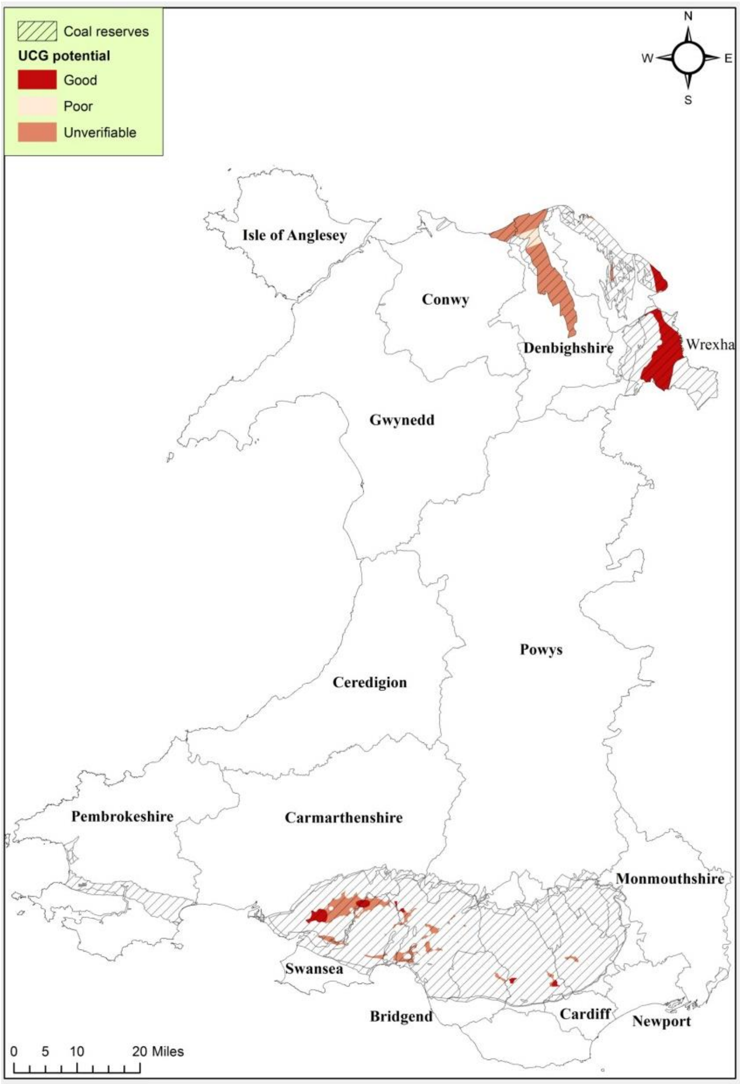


Figure 7.7 UCG resource potential areas in wales

7.4.2 Coal dataset

Coal category and depth dataset for Wales is acquired from the BGS and it is incorporated into the geodatabase as shown in Figure 7.8. This dataset is also qualitative in nature and it is processed using the same procedure as used for the qualitative dataset of UCG (Section 7.4.1.2). The discrete classes of the coal in this dataset are:

- Deep coal at more than 1200m.
- Extent of Coal Basin (Carboniferous Coal Measures - beneath sea).
- Inactive or unproductive coalfield.
- No coal identified at any level.

- Shallow coal with less than 50m overburden.
- Deep coal between 50m and 1200m.

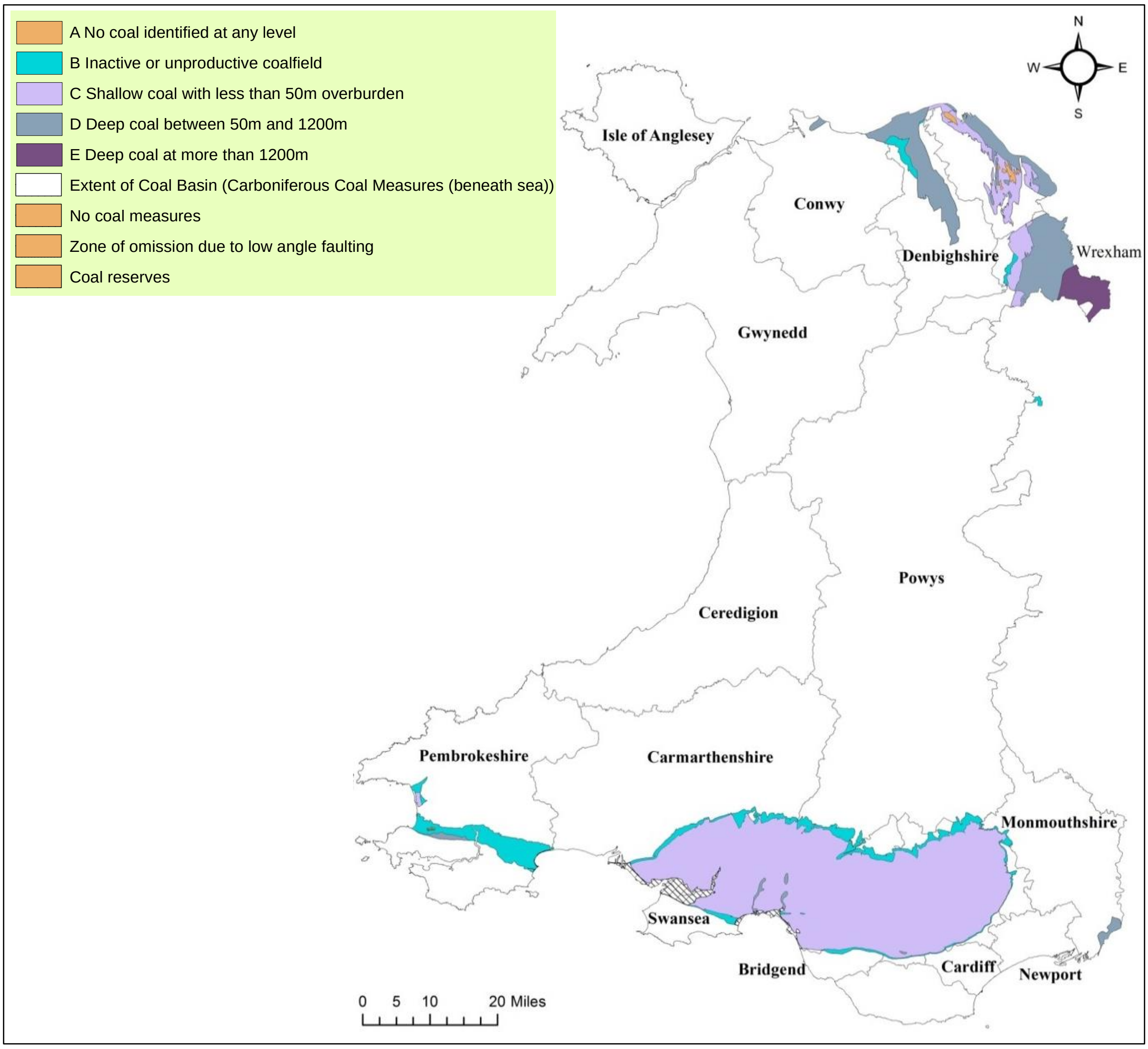


Figure 7.8 BGS - coal category and depth map

Coal dataset is important as the two unconventional gas resources incorporated in the geodatabase are CBM and UCG. This layer can be useful in applying the filters and constraints in AHP based site selection toolkit if the decision maker is interested only in those areas where coal exists at the shallower depths. This reduces the system resources and time required in carrying out the analysis.

7.4.3 Geological dataset

Geological information includes bedrock, superficial, dykes and faults. These are important parameters of the techno-economic domain for spatial decision making of site selection and impact assessment as discussed in Section 2.7. For this purpose DiGMapGB-625 (BGS 2013) bedrock and superficial geology dataset is acquired and incorporated in the geodatabase.

7.4.4 Hydrogeology dataset

The hydrogeological features are important for spatial decision making of site selection and for the impact assessment of potential contamination on important aquifers. For this purpose the hydrogeological data is acquired from the BGS (BGS 2012) as shown in Figure 7.9.

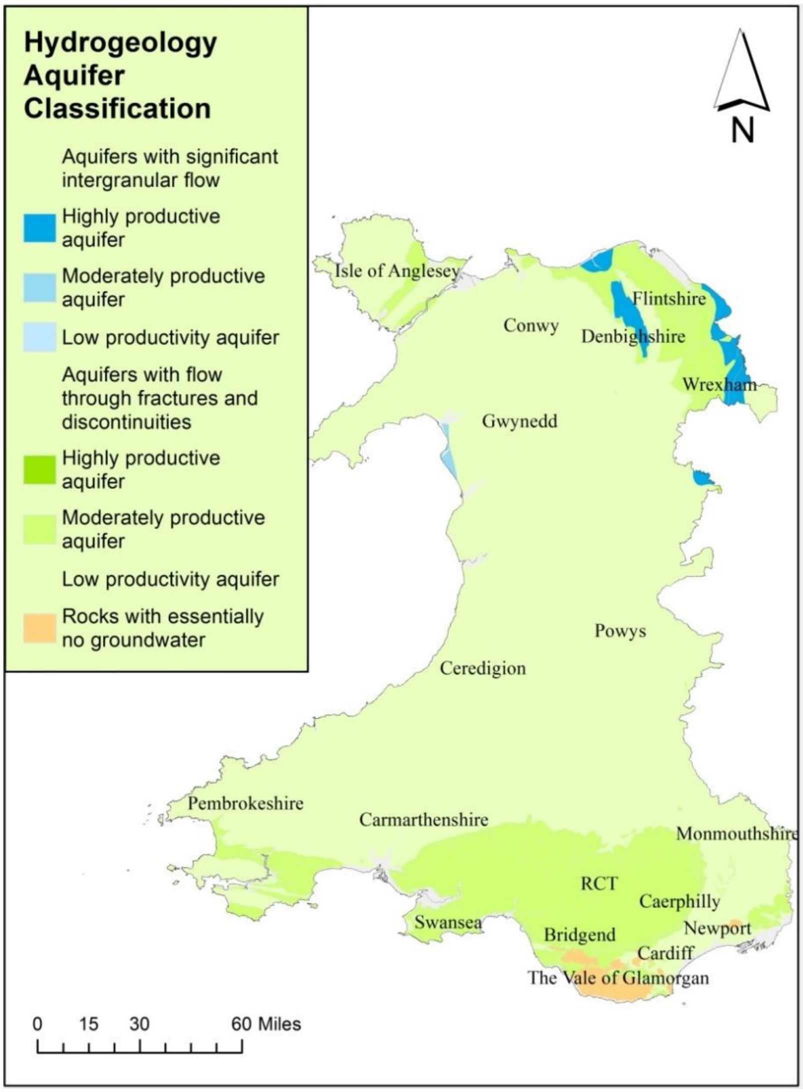


Figure 7.9 BGS - hydrogeology and aquifer potential map

7.4.5 Site economic parameters

Site economic parameters are the key consideration in the decision making process to ensure potential investments and sustainability of the unconventional gas development as discussed in Section 2.7. Therefore the following key datasets are identified and incorporated into geodatabase:

- Distance from major CO_2 emitters.
- Distance from existing gas feeder pipeline network.
- Distance from existing Above Ground Gas Installations (AGIS).
- Distance from High Transmission Lines (HTL).
- Distance from sub-stations.
- Distance from existing road and railway network.

Above datasets have been acquired from DECC (DECC 2012). These distances are calculated from the centroid of each Fishnet cell from the nearest feature of a given dataset one by one using the distance tool in ArcGIS. Existence of gas network is important to the economics of a CBM and UCG application so that the produced gas is easily injected into the grid. If electricity is produced from the site then the distance from the existing electric network is also important. Distance from CO_2 emitters has a dual role: i) CO_2 can be easily transported from the sources to the site for Enhanced Coal Bed Methane (ECBM) applications and ii) produced gas can be supplied to these industries as an energy source or to be used as stock. Distance from existing road and rail network is important throughout the lifespan of the sites. Another important factor that contributes to the site development cost is the land acquisition price for site development. For this purpose the TEV and distance from DLUA indicators can be used as explained in Section 7.3.

7.4.6 Terrain dataset

Terrain information such as Slope and Elevation are important factors to be considered in the decision making pertaining to site selection of the Geoenergy applications. Terrain parameters are essential considerations for the feasibility of the site development and can contribute to the project cost as well. Terrain parameters are also used in the environmental impact and risk assessment, e.g. the contamination of hydrology of the surrounding regions in case of a spillage. For this purpose, OS Terrain 50 data has been acquired and assigned to Fishnet cells.

7.5 Environmental Domain

As discussed in Section 2.6, Geoenergy applications may have significant effects on our environment. Therefore, it is important to include key environmental indicators into consideration while addressing spatial decision problems related to Geoenergy applications. The environmental domain contains these key indicators and the location of sensitive environmental areas in Wales as shown in Figure 7.10. These indicators reflect the current state of the environment across wales at various scales and can be useful in the informed risk based spatial decision making.

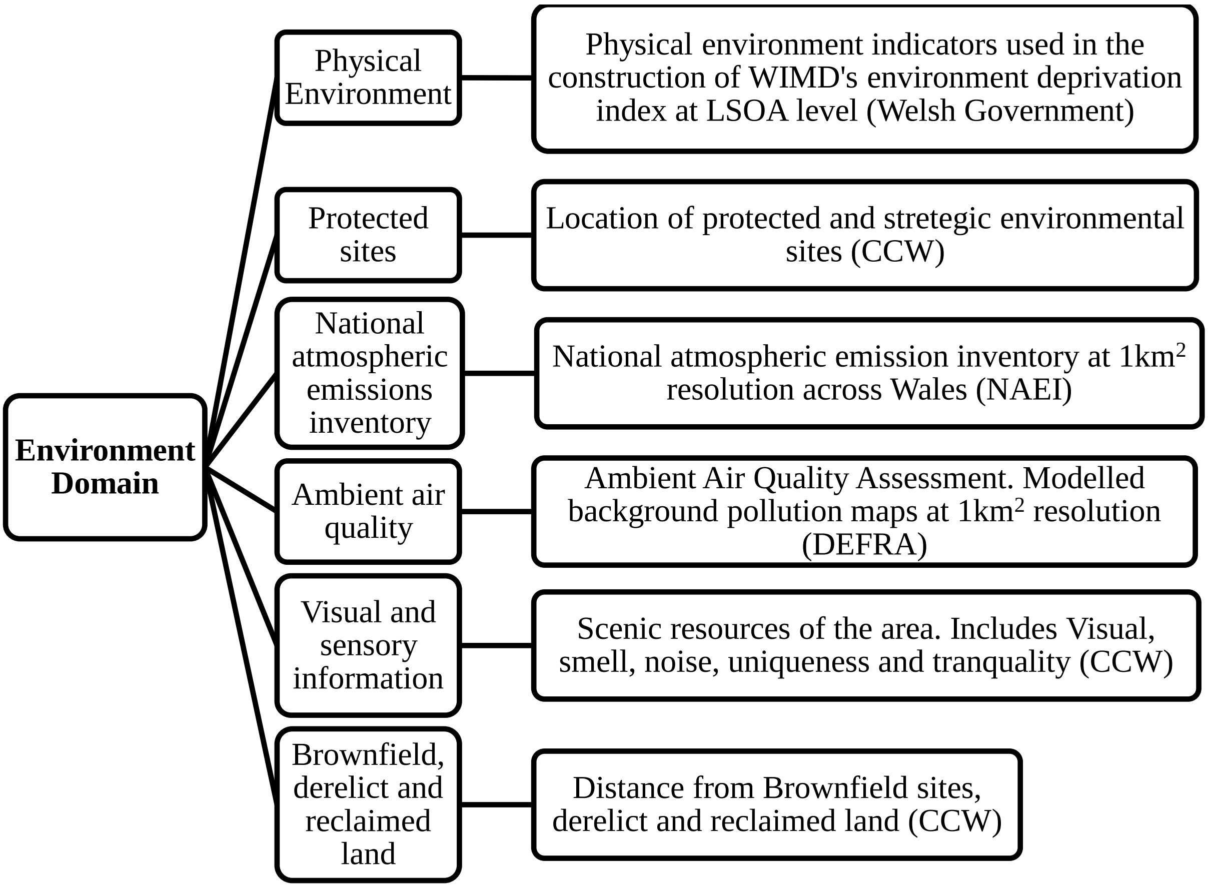


Figure 7.10 Key Environmental indicators and datasets incorporated into the geodatabase

7.5.1 Physical Environment

The physical environment covers important environmental factors that may impact the quality of life. The physical environment indicators used in the construction of the environmental deprivation ranks in WIMD-2011, are acquired at the LSOA level. These indicators are: a) Air quality, b) Air emissions, c) Flood risk and d) Proximity to waste disposal and industrial sites (WIMD 2011b).

7.5.2 Protected Sites

The exploitation of Geoenergy resources is subjected to licences and permission at a specific site level. This permission depends on meeting certain conditions. One important consideration is that the site is not going to negatively impact on the current state of the protected areas. Therefore it is important to include all the protected sites in Wales into the decision making context. For this purpose the protected sites data is acquired from the

Country Side Council for Wales (CCW 2012). The details of the protected sites covered in the CCW GIS dataset are provided in Table A.3 (Appendix A).

In order to incorporate these protected sites in the decision making context, the “Near” tool of ArcGIS is used to calculate the distance between the centroid of each Fishnet cell to the nearest edge of the protected site boundary.

7.5.3 National atmospheric emission inventory

In UK, the National Atmospheric Emissions Inventory (NAEI) collects emissions data from different sources, such as power stations, industrial processes, traffic, household heating and agriculture (NAEI 2014). This data is acquired from Department of Environment Food and Rural Affairs (DEFRA) NAEI portal in ASCII format and covers a number of pollutants such as: 13-Butadiene, Ammonia, Arsenic, Benzene, Benzo-a-Pyrene, Cadmium, CO_2 as Carbon, Chromium, Copper, Dioxins(PCDD/F), Hydrogen Chloride, Lead, Mercury, Nickel, Nitrogen Oxides as NO_2, Non Methane VOC, PM10 (Particulate Matter < 10um), Selenium and Sulphur Dioxide.

The NAEI datasets are developed by combining the information from point emission sources, road transport data and from distribution grid information. The cell size of each dataset is $1Km^2$. These emission datasets are acquired in the ASCII format and then processed in ArcGIS using “ASCII to Raster” tool and then populated in the Fishnet cells.

7.5.4 Ambient air quality

DEFRA UK is responsible for meeting the obligations of European Union Air Quality guidelines. For this purpose they use different modelling techniques applied over the monitored air quality and then produce modelled ambient air quality maps (DEFRA 2014). These modelled ambient air quality maps are acquired at $1Km^2$ resolution for a number of pollutants including Benzene, CO, PM2.5, PM10, SO, NOX, NO_2 and Ozone.

The modelled ambient air quality maps are mostly represented as Annual-Means, with units in µg m^3. However CO concentrations are represented in mg/m^3 (DEFRA 2014). Some areas near the coastlines have “no data” and these cells were removed first. Then Fishnet cells were populated using the spatial “Spatial Join” tool in ArcGIS.

This dataset can be used in two ways in the SDSS: i) it can be used in the AHP based site selection toolkit as a key environmental indicator by assigning less weights or completely filtering out an area already under environmental stress and ii) it can also be used along with the European Union’s National Emissions Ceiling Directive (European_Commission 2014). The Directive sets the upper limit of certain pollutants for each of the member states.

7.5.5 Visual and sensory information

Landscape, scenic beauty and aesthetics of an area are of interest to the decision makers and other stakeholders including communities. As discussed in Chapter 2, this factor is one of the reasons behind the negative public perception and acceptability for engineering interventions and site developments. It is therefore important to include this factor into the SDSS and make it an integral part of the decision making context of site selection, site ranking and impact assessment.

For this purpose, the visual and sensory section of the CCW’s LANDMAP dataset is acquired. The Visual and Sensory data maps the aspects that we perceive through our senses e.g. visual, hearing, smell and touch. These perceptions are based on the physical attributes of landform and land cover, visible patterns of distribution and their consistent relationships in particular areas. These perceptions are mostly limited to qualitative judgements of human beings. In LANDMAP dataset however, consistent definitions and terminologies are used along with the assessment methodologies. This is to ensure that the qualitative visual and sensory aspects of the landscape are mapped consistently across the Wales (CCW 2011).

The Intrinsic Evaluation Matrix of LANDMAP (Visual & Sensory) dataset is also incorporated in the geodatabase for aesthetic coverage across Wales. This dataset contains records of the ordinary and spectacular landscapes as well as information about the physical, ecological, visual, historical and cultural landscape of Wales (CCW 2011). The Intrinsic Evaluation Matrix covers scenic quality, integrity, character and rarity of the area and an "overall" index, which divides the Welsh landscape into Outstanding, High, Moderate and Low values. Some other aesthetic qualities are also included in the geodatabase including the level of human access in the area, night time light pollution and sense of place (CCW 2011).

The Land-Form and Land-Cover attributes are also incorporated in the geodatabase. These two are qualitative variables and can be useful for the site selection process and also for applying the filters and constraints. There are many categories, sub-categories and attributes linked to the visual and sensory data of LANDMAP. The selected parameters for this study have been incorporated into the geodatabase as shown in Figure 7.11.

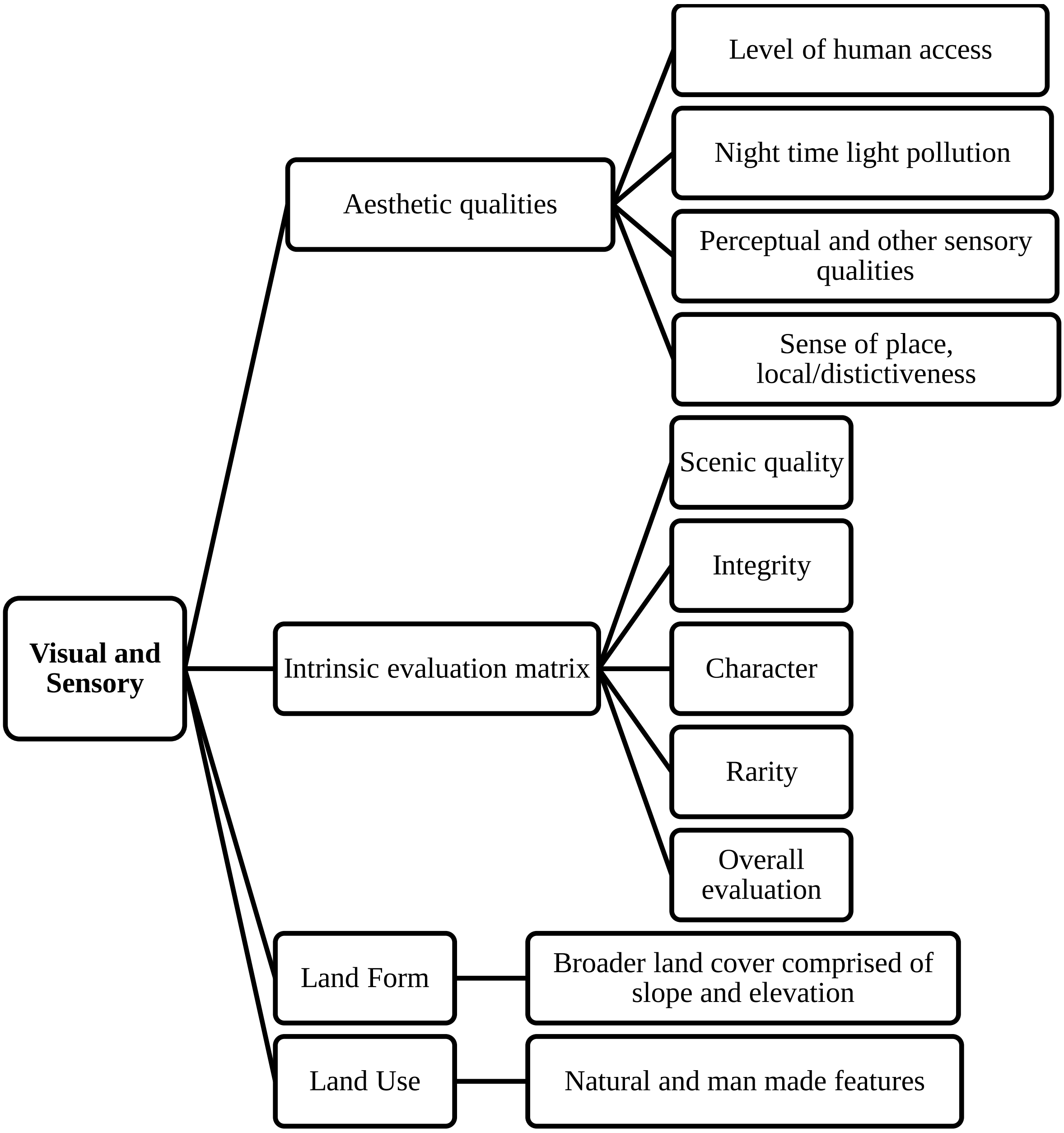


Figure 7.11 Selected visual and sensory attributes incorporate into the geodatabase

7.5.6 Brownfield, derelict and reclaimed land

Pauleit et al. (2005) referred to the UK's Planning policy Guidance Notes which favoured the regeneration and rehabilitation of the brownfields and derelict land. South Wales (UK) has been a centre for mining, querying and related industrial work for over 250 years, resulting in a large number of derelict and brownfield sites. Regeneration work has been carried out on the derelict and contaminated land to improve the environmental conditions (Thornton and Walsh 2001).

There is no direct source of information on the brownfield, derelict and regenerated land in Wales. In order to incorporate this information in the geodatabase, relevant classes of the landuse and land cover have been extracted from the CCW's LANDMAP data (CCW 2011). Visual Sensory (VS) section of the LANDMAP dataset is used for this purpose. VS30 attribute contains general comments of the surveyors about the land cover. All those areas are extracted with keywords such as derelict, brownfield, landfill and tipping. to extract these areas from the entire Welsh land cover.

Cultural Landscape (CLS) section of the LANDMAP dataset is used for extracting the reclaimed land CLS3 attribute contains the level-3 classification of cultural context. All those areas with "Reclaimed" keyword in CLS_3 are extracted. The "Select by Attribute" tool of the ArcGIS is used for this purpose and then each Fishnet cell is populated with the data.

7.6 Public Health Domain

Public health domain covers the public health aspects of the spatial decision context. Figure 7.12 shows the hierarchy of indicators and sub-indicators used in the public-health domain. These indicators have been selected as they represent the spatial variations in public health status in Wales. As discussed in Chapter 2, risks associated with public health can be minimized by carefully selecting the sites for Geoenergy applications. Area that is already under stress in terms of public health should be given less weights or filtered out during the site selection. Selected indicators cover mortality rate, hospital admission rate for major illnesses, health related indicators used in the construction of WIMD, percentage of low birth weights and results of the Welsh Health Survey (WSH) 2003-2009.

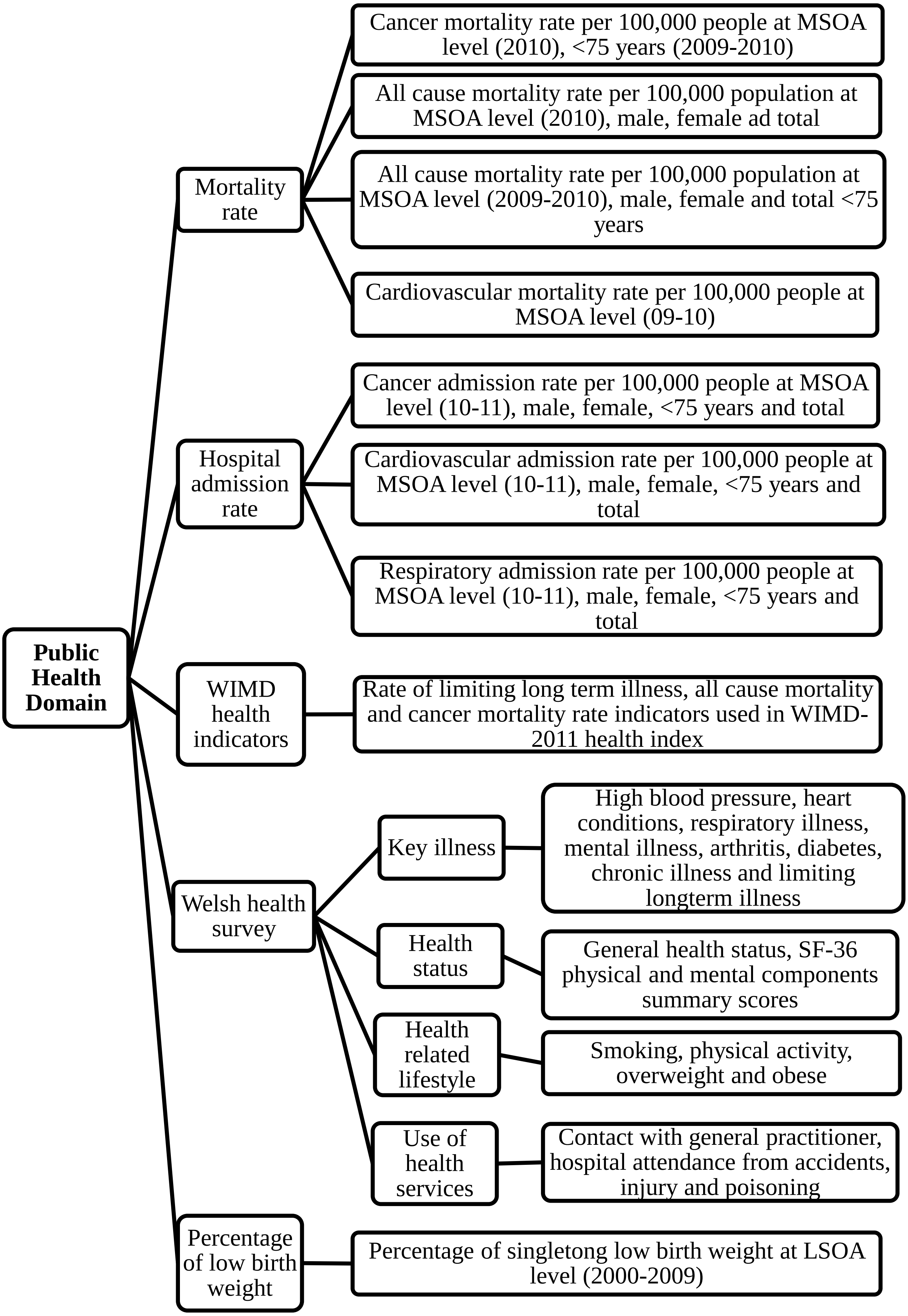


Figure 7.12 Key Public Health indicators and datasets incorporated into the geodatabase

Public health indicators can be used in the decision context of site selection and impact assessment in a way that helps decision makers for the advocacy of Geoenergy applications in particular the unconventional gas development in Wales with least impact on public health.

7.6.1 Mortality rate

Mortality rate data has been published by the NHS Wales Informatics Service (NWIS) and is obtained from the Health Maps Wales (HMW) data portal (NHSWales 2014). Spatial coverage of the data is focussed on the Welsh residents. The mortality data is published at Local Authority (LA), Upper Super Output Areas (USOA) and Medium Super Output Areas (MSOA). For geodatabase, only the data at MSOA has been incorporated as it represents the smaller areas as compare to LA and USOA. MSOAs have a mean population of about 7,000 people. Explanation of the sub-indicators used in this indicator is given below.

7.6.1.1 All-cause mortality rate

All-cause mortality data is the death rate per 100,000 of population in Wales at the MSOA level. The data is already age standardise (per 100,000 European standard population) as different age groups may have different death rates. All-cause mortality data is incorporated in the geodatabase for male, female and total population for all ages in 2010. All-cause mortality data for male, female and total population less than 75 years of age for 2009-2010 is also incorporated in the geodatabase.

7.6.1.2 Cancer mortality rate

Cancer is one of the main causes for disease related mortality and NHS records mortality caused by each type of cancer. The two datasets incorporated in geodatabase covers the death rates caused by all types of cancers: a) Cancer related mortality rate (per 100,000 European standard population) for all age groups and b) Cancer related mortality rate (per 100,000 European standard population) for less than 75 years of age.

7.6.1.3 Cardiovascular mortality rate

Cardiovascular diseases related mortality rate is also included in the geodatabase. The indicator used here is the 2 years range (2009-2010) of mortality rate cause by all types of cardiovascular diseases in Wales (per 100,000 European standard population) for all age groups.

7.6.2 Hospital admission rate

Hospital admission rate for key illnesses is another good indicator for assessing the public health status across the study area. For this purpose, hospital admission rates data is acquired from HMW for the key illnesses at the MSOA level such as cancer, cardiovascular and respiratory diseases. This data covers all the patients admitted in hospitals whether patient Class-1 (ordinary inpatient) or Class-2 (day case) (NHSWales 2014). The data is age standardised using European standard population and covers the two year range 2010-2011.

7.6.2.1 Cancer related admission rate

Spatial variations in cancer related hospital admission rates across Wales are included in the geodatabase. The datasets incorporated in geodatabase cover the Cancer related hospital admission rate (per 100,000 European standard population) in 2010-2011 for male, female and total population.

7.6.2.2 Cardiovascular related admission rate

Spatial variations in cardiovascular related hospital admission rates across Wales are included in the geodatabase. The datasets incorporated in geodatabase cover the cardiovascular related hospital admission rate (per 100,000 European standard population) in 2010-2011 for male, female and combined. Fourth data in this category covers only the coronary heart diseases related hospital admission rates (per 100,000 European standard population) in 2010-2011.

7.6.2.3 Respiratory related admission rate

Respiratory disease related emergency hospital admission rates (per 100,000 European standard population) in 2010-2011 for male, female and combined population in Wales are incorporated in the geodatabase.

7.6.3 WIMD health indicators

In order to provide the data at a finer level than MSOA, indicators used in the construction of WIMD health index are also incorporated in the geodatabase. WIMD is constructed at the LSOA level where mean population is about 1500 as compare to about 7000 people in MSOAs. Health indicators used in the construction of WIMD give finer detail for local area level analysis. Three indicators are incorporated in the geodatabase in this respect a) Rate of all cause deaths (per 100,000 European standard population) recorded between 2000-2009, b) Rate of cancer related deaths (per 100,000 European standard population) recorded between 2000-2009 and c) Rate of limiting long term illness (per 100,000 European standard population) in 2011.

7.6.4 Singleton low birth weight

The singleton low birth weight (less than 2500g) is also a WIMD (2011) indicator and is acquired at the LSOA level for a period of last ten years. It is evident from literature that low birth rate is linked to the mother's lifestyle and health. This indicator can be used for both socio-economic and public health domains. This data is acquired from the InfoBase Cymru which is a spatial data portal for of the Local Government Data Unit Wales (Infobase 2014).

7.6.5 Welsh Health Survey

The Welsh Health Survey is used by the Welsh Government to plan health services and policy decision making for promoting better health. WHS provides an overall picture about the nation's health and spatial variations across Wales. It covers different groups, such as children and older people. The survey uses a representative sample of Welsh population every year with around 15,000 adults and 3,000 children (WHS 2014). The survey results are summarised at the Upper Super Output Areas (USOA) level. The upper super output areas have a mean population of about 31,000.

Different headline indicators and question from the WHS for the period of 2003/04-2009 are selected and incorporated in the geodatabase as explained in the sections below.

7.6.5.1 Key illness

It covers the percentage of adults (age standardised) who reported that they are suffering and currently being treated for any key illness including high blood pressure, heart conditions, respiratory illness, mental illness, arthritis, diabetes, chronic illness and limiting long term illness. This percentage of each USOA in Wales for each key illness separately is incorporated in the geodatabase.

7.6.5.2 Health status

The survey responses for general health status are also acquired at the USOA level and incorporated into the geodatabase. This covers the summary scores of self-perceived SF-36 questionnaire. SF-36 is a well-used Short Form (SF) survey consisting of 36 questions about the physical and mental health of an individual.

7.6.5.3 Health related lifestyle

The percentage of adult respondents (age standardised) at the USA level, who reported health related lifestyles, e.g. smoking, overweight or obesity are incorporated in the geodatabase. This headline indicator also covers the percentage of adults who meet the guidelines for physical activities.

7.6.5.4 Use of health services

This headline covers the percentage of the adult respondents (age standardised) who have reported using the selected health services during the last week of the survey either (a) contact with the general practitioner and (b) hospital attendance from accidents, injury and poisoning. The data is also acquired at the USOA level and incorporated in the geodatabase.

7.7 Conclusions

This chapter covers the design and development of the geodatabase which is an essential component of the SDSS. The SDSS is designed and developed independently of the study area. However, for its applications, Wales has been considered as the study area in this research. Therefore the key indicators and essential spatial datasets have been identified for Wales and incorporated into the geodatabase. The data has been categorised into four domains: i) Socio-Economic, ii) Techno-Economic, iii) Environmental and iv) Public Health. The key indicators are acquired from various sources to facilitate informed risk based decision making related to Geoenergy and Geoenvironmental problems, in particular those related to unconventional gas development. Geodatabase serves as the data backbone of the SDSS and different analytical modules utilises this information while facilitating spatial decision making.

In order to provide a multicriteria spatial decision support environment, all important dimensions, indicators, sub-indicators have been considered, subject to the availability of the data. Some missing key indicators such as the soil quality, biodiversity and other ecological information could not be incorporated due to the cost or unavailability of the data.

The selection of key indicators and spatial dataset was based on the literature review presented in Chapter 2. For this purpose, the most common Geoenergy and Geoenvironmental spatial decision problems were reviewed first. Then, environmental, socio-economic, public health and techno-economic aspects of spatial decision problems were highlighted. Following that, appropriate datasets in Wales were identified and processed using different GIS modelling techniques and then incorporated into the geodatabase. Some of these datasets were acquired in the GIS format that could be stored directly in the geodatabase. While others were in the form of statistical or tabular information which required pre-processing and application of appropriate GIS modelling techniques before incorporating into the geodatabase. Verification of the GIS analysis results was carried out where necessary e.g. the GIS modelling and mapping of social capital across Wales has been verified as explained in section 7.3. In some cases, where direct information was not accessible, required data was derived through other datasets e.g. the information about brownfield and derelict land was extracted from different attributes (variables) of the CCW LANDMAP dataset as explained in Section 7.5.6.

The acquired datasets had different scale and units. To accommodate that, a Fishnet of $500m^2$ is created across the onshore area of Wales. This approach enables combining all the information in the same scale and the data can therefore be processed easily within the geodatabase using spatial and non-spatial queries. A commensuration tool is provided in the SDSS that can be used to scale the data and to bring it to the same currency between 0 and 1.

The multicriteria decision solving paradigm is a data-centric approach, the choice, quality and scale of the datasets can directly impact the quality of the decision outcomes. Therefore efforts have been made to acquire the most suitable and best available datasets. At the same time, the flexible design of the geodatabase makes it easier to add, replace or modify existing data incorporated in the geodatabase.

References

Akçomak, İ. S. and ter Weel, B. 2012. The impact of social capital on crime: Evidence from the Netherlands. *Regional Science and Urban Economics* 42(1–2), pp. 323-340.

BGS. 2012. *BGS Hydrogeology and aquifer potential* [Online]. BGS. Available at: http://www.bgs.ac.uk/products/hydrogeology/maps.html [Accessed: 07 May].

BGS. 2013. *BGS DiGMapGB-625* [Online]. Available at: http://www.bgs.ac.uk/products/digitalmaps/digmapgb_625.html [Accessed: 07 May].

CCW. 2011. *CCW LANDMAP – the landscape atlas* [Online]. Available at: http://www.ccw.gov.uk/landmap [Accessed: 14 October].

CCW. 2012. *Countryside Council for Wales GIS datasets* [Online]. Available at: http://www.ccw.gov.uk/landscape--wildlife/protecting-our-landscape/gis-download---welcome/gis-dataset-information.aspx [Accessed: 14 October].

DECC. 2012. *Department of Energy & Climate Change, Onshore GIS data* [Online]. Available at: https://www.gov.uk/oil-and-gas-onshore-maps-and-gis-shapefiles#onshore-gis-data [Accessed: 07 May].

DEFRA. 2014. *DEFRA Air Quality Modeling* [Online]. Available at: http://uk-air.defra.gov.uk/data/pcm-data [Accessed: 14 October].

EEA. 2006. *European Environment Agency, Corine Land Cover 2006 seamless vector data* [Online]. Available at: http://www.eea.europa.eu/data-and-maps/data/clc-2006-vector-data-version [Accessed: 07 May].

ESRI. 2014. *ESRI ArcGIS10 Mapping Platform* [Online]. Available at: https://www.arcgis.com/features/ [Accessed: 18 August].

European_Commission. 2014. *National Emission Ceilings Directive* [Online]. Available at: http://ec.europa.eu/environment/air/pollutants/ceilings.htm [Accessed: 05 Sept].

Experian. 2012. *Experian Mosaic June 2012* [Online]. Available at: http://www.experian.co.uk/marketing-services/products/mosaic-uk.html [Accessed: 07 May].

Foxton, F. and Jones, R. 2011. Social Capital Indicators Review. In: Statistics, O.f.N. ed. Crown copyright 2011.

Garrone, P. and Groppi, A. 2012. Siting locally-unwanted facilities: What can be learnt from the location of Italian power plants. *Energy Policy* 45(0), pp. 176-186.

Infobase. 2014. *Infobase Cymru* [Online]. Local Government Data Unit Wales. Available at: http://www.infobasecymru.net/IAS/ [Accessed: 14 October].

Jones, N. S.,Holloway, S.,Smith, N. J. P.,Browne, M. A. E.,Creedy, D. P.,Garner, K. and Durucan, S. 2004. *UK coal resources for new exploitation technologies*.

Meng, G.,Law, J. and Thompson, M. E. 2010. Small-scale health-related indicator acquisition using secondary data spatial interpolation. *Int J Health Geogr* 9, p. 50.

NAEI. 2014. *National Atmospheric Emissions Inventory* [Online]. Available at: http://naei.defra.gov.uk/ [Accessed: 04 March].

NHSWales. 2014. *Health Maps Wales* [Online]. NHS Wales Informatics Service (NWIS). Available at: https://www.healthmapswales.wales.nhs.uk/IAS/home/ [Accessed: 14 October].

ONS. 2011. *Cencus 2011* [Online]. Available at: http://www.ons.gov.uk/ons/index.html [Accessed: 30 Sept].

OS. 2012. *Ordnance Survey Meridian 2 Dataset* [Online]. Available at: http://www.ordnancesurvey.co.uk/business-and-government/products/meridian2.html [Accessed: 07 May].

OS. 2014. *Code-Point Open* [Online]. Available at: http://www.ordnancesurvey.co.uk/business-and-government/products/code-point-open.html [Accessed: 14 October].

Pauleit, S.,Ennos, R. and Golding, Y. 2005. Modeling the environmental impacts of urban land use and land cover change—a study in Merseyside, UK. *Landscape and Urban Planning* 71(2–4), pp. 295-310.

Plymouth-University. 2013. *The Elections Centre* [Online]. Available at: http://www1.plymouth.ac.uk/research/ceres/TEC/Pages/default.aspx [Accessed: 30 Sept].

SpatiaLite. 2014. *SpatiaLite Geodatabase* [Online]. Available at: https://www.gaia-gis.it/fossil/libspatialite/index [Accessed: February 06].

Takagi, D.,Ikeda, K. i. and Kawachi, I. 2012. Neighborhood social capital and crime victimization: Comparison of spatial regression analysis and hierarchical regression analysis. *Social Science & Medicine* 75(10), pp. 1895-1902.

Thornton, G. J. P. and Walsh, R. P. D. 2001. Heavy metals in the waters of the Nant-y-Fendrod: change in pollution levels and dynamics associated with the redevelopment of the Lower Swansea Valley, South Wales, UK. *Science of The Total Environment* 278(1–3), pp. 45-55.

van der Horst, D. 2007. NIMBY or not? Exploring the relevance of location and the politics of voiced opinions in renewable energy siting controversies. *Energy Policy* 35(5), pp. 2705-2714.

Welsh-Government. 2012. *Wales digital inclusion strategy* [Online]. Available at: http://wales.gov.uk/topics/people-and-communities/research/community/digincanalysis/?lang=en [Accessed: 07 May].

WHS. 2014. *Welsh Health Survey* [Online]. Welsh Government. Available at: http://wales.gov.uk/statistics-and-research/welsh-health-survey/?lang=en [Accessed: 14 October].

WIMD. 2011a. *Welsh Index of Multiple Deprivation* [Online]. Available at: http://wales.gov.uk/topics/statistics/theme/wimd/?lang=en [Accessed: 14 October].

WIMD. 2011b. *Welsh Index of Multiple Deprivation 2011 technical report.*

8

APPLICATION

8.1 INTRODUCTION

This chapter presents an application of the Spatial Decision Support System (SDSS). The developed software has been applied to facilitate spatial decision making for an effective utilisation of Coalbed Methane (CBM) and Enhanced Coalbed Methane recovery (ECBM) in Wales, UK. Both technologies have been explained in detail in Section 2.3. The spatial decision problems considered in this application include i) site selection, ii) site ranking and iii) impact assessment while considering key environmental, socio-economic, public health and techno-economic aspects.

As discussed in Chapter 2, unconventional gas developments may have impacts on economic, social, and environmental aspects. Also, effective and sustainable use of an unconventional gas resource depends on various techno-economic parameters. Therefore the main objective for this application is to ensure that the proposed sites have: a) minimum negative impact on the environment, b) positive impact on the socio-economic conditions of the communities living nearby, c) are located in areas where public health is not already under stress, and d) are economically and technically more viable than other potential areas. In order to achieve these objectives, the following analytical modules of the SDSS have been utilised: i) Analytical Hierarchy Process (AHP) based site selection tool, ii) Site ranking by neighbourhood analysis and comparison tool, iii) Self-Organising Maps (SOM) based site ranking tool, and iv) Rapid Impact Assessment Matrix (RIAM) based impact assessment tool.

The systematic approach developed in this case study starts with the application of the AHP based site selection tool to identify the potential areas with consideration of the four domains: i) Socio-Economic, ii) Environmental, iii) Public Health and iv) Techno-Economic, individually and combined, as discussed in Section 8.3. Equal weight is assigned to all four domains and using a sensitivity analysis carried out in section 8.4 the effect of uncertainty in the decision on weight will be discussed. A sensitivity analysis has not been applied to the individual weights assigned at the indicators levels, as this would have taken a substantial amount of processing and time. Also, for this case study, only selected indicators from the four domains have been used.

The selection of indicators and decision regarding weights assigned to them is subject to the decision maker’s choice and is based on their relative importance, quality and scale of data. This is a known limitation associated with AHP as discussed in Section 2.2. To overcome this, a pairwise comparison matrix can be used to find the weights that are consistent and can be used with confidence. The scope of this case study is limited to the demonstration of the applicability of the SDSS developed in this research. For real life application, more precise information can be used and group decision making can be considered. Weights can be obtained from a panel of experts representing environmental, socio-economic, public health and techno-economic side.

Section 8.6 and 8.7 covers the site ranking process using the SOM based site ranking module and site ranking by neighbourhood analysis modules respectively. Subsequently, the number of potential sites has been reduced. The ECBM scenario is discussed in section 8.8 where proximity of the potential CBM sites to the existing national grid (gas and electricity) and major CO_2 producers, is incorporated in the spatial decision making process.

Site impact assessment is covered in section 8.9 where the RIAM based module is applied to estimate the potential impact of the proposed sites on economic, social, and environmental aspects. The criteria used for the selection of sites remains the same for the entire study area and most of the identified sites are clustered together within similar geographical conditions. Therefore, the same RIAM matrix is applied on the selected sites and the results are discussed.

In order to identify the current state of the sites in terms of potentially harmful chemicals, a geochemical baseline is also provided in Section 8.8. The geochemical baseline gives the current concentration levels of the key substances around the proposed sites, including Arsenic Cadmium, Copper, Iron, Lead, Magnesium, Nickel, Potassium, Tin, Vanadium and Zinc.

8.2 Study area

The study area selected for the onshore CBM application is located in the North and South Wales coalfields in UK. According to a report published by the Department of Trade and Industry (DTI), there is a considerable potential for CBM recovery in these two coalfields (Jones et al. 2004).

Figure 8.1 shows an overview of the variable resource density ($10^6 m^3 ha^{-1}$) at different locations within the two coalfields. The CBM resource density has been calculated to populate the Fishnet cells using the same parameters as used in (Jones et al. 2004). Each Fishnet cell is of equal area, i.e. $500m^2$ therefore its respective CBM resource density calculation is the same for every cell covered by the given CBM region.

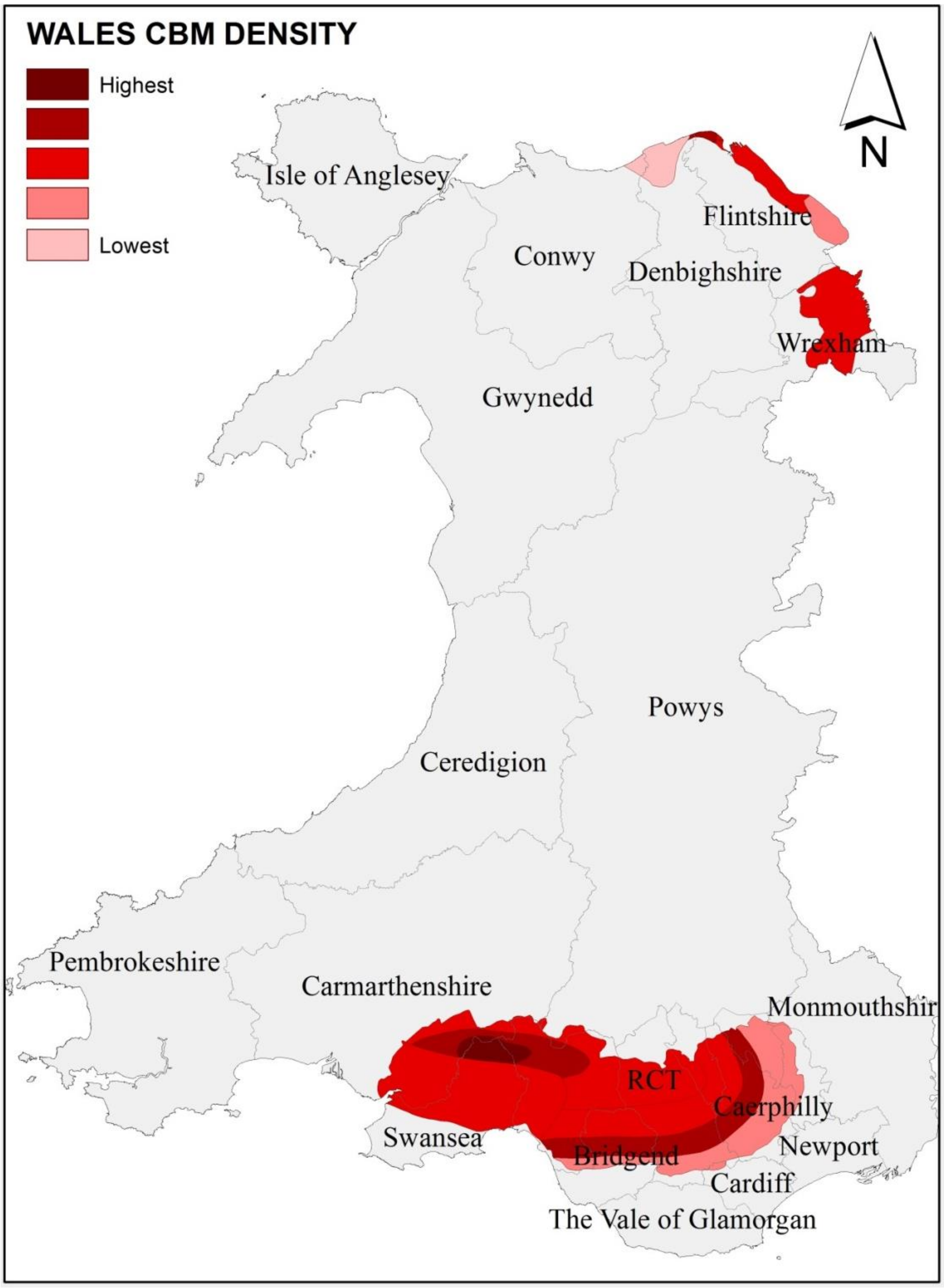


Figure 8.1 Onshore CBM resource density Map of Wales. Adapted from Jones et al. (2004)

8.3 CBM site selection

The AHP based site selection tool has been applied to identify the most suitable areas with CBM resource potential in the study area. The tool has the capability of applying and viewing the results of WLC analysis at any level of the AHP decision tree as explained in Section 5.4.1. Initially, the selection process is applied to the four domains separately i.e. a) Techno-Economic, b) Socio-Economic, c) Public Health and d) Environmental domain.

As explained in Section 8.1, the decision on the indicator selection and weights assigned to them, is taken by decision maker and it is based on their relative importance for a given analysis, quality and scale of indicators being used. For this case study, in the site selection process, only selected indicators from the four domains have been used.

Figure 8.2 depicts the performed systematic process of site selection, ranking and impact assessment. First, the AHP and WLC based site selection module is applied on the selected indicators from all four domains. Constraints are applied to limit the processing to the geographical areas covering South and North Wales coalfields. Selected indicators, constraints, weights and the results for each domain are discussed in Sections 8.3.1-8.3.4. Section 8.3.5 covers the combined effect of the all four domains in the AHP based site selection process.

Once potential sites or area for CBM application in Wales is identified, then site ranking is carried out to prioritise the sites. For this purpose, two site ranking tools developed in the SDSS are applied: i) SOM based site ranking tool and ii) Site ranking by neighbourhood analysis and comparison tool. This process ranks the Fishnet cells in terms of the potential for CBM development while considering key aspects as discussed in Section 8.5 and 8.6. Finally, the RIAM based site impact assessment tool is applied to generate an impact assessment report for the potential sites. This approach covers Physical/Chemical, Biological/Ecological, Social/Cultural and Economics/Operational aspects as explained in Section 8.9.

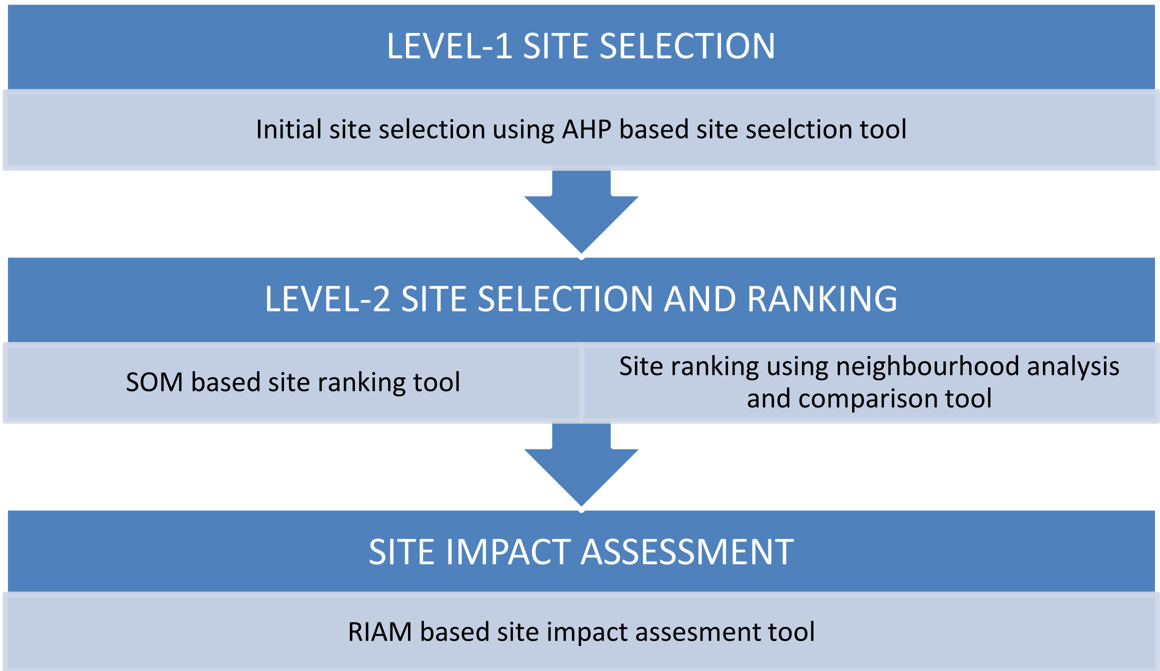


Figure 8.2 Utilising SDSS tools for CBM application in Wales

8.3.1 Level-1 site selection applying AHP based site selection tool to the Techno-Economic domain

At the first level of CBM site selection process, the AHP based site selection tool is applied. The tool facilitates implementation of the underlying WLC procedure at any level of the AHP decision tree. This feature is utilised to apply the Level-1 site selection process for each domain individually and then a combined process, considering all four domains, is also applied.

Headline indicators selected from Techno-Economic domain are: i) Geology, ii) Site economic parameters and iii) Terrain parameters. Table 8.1 shows all the indicators selected from the Techno-Economic domain to be used in the procedure. The “Cost” or “Benefit” nature of these indicators and the assigned weights are also given in the table. Details of the headline indicators, sub-indicators and their sources are provided in Geodatabase, Chapter 7.

CBM resource potential has been assigned 70% of the total weight, in order to give it more importance as compare to the other indicators in this domain. The rest of the weight is equally divided between geology, site economic parameters and terrain indicators. Weights of the composite indicators are further divided among their child indicators according to their relative importance.

Table 8.1 Indicators used in the Techno-Economic domain

Indicator	Weight	Nature (Cost/Benefit)
CBM resource	**70%**	**Benefit**
Geology	**10%**	**Benefit**
Distance from fault lines	50%	Benefit
Distance from geological dykes	50%	Benefit
Site economic parameters	**10%**	**Benefit**
Distance from major CO_2 emitters	50%	Cost
Distance from gas feeder pipeline network	50%	Cost
Terrain	**10%**	**Benefit**
Elevation	50%	Cost
Slope	50%	Cost

Filter: CBM Resource > 0 AND 'Highly productive aquifer' not present

As shown in above table, two filters are applied to restrict the AHP process only to those geographical areas where there is some CBM potential and where no highly productive aquifers are present. In this way, two advantages are achieved: i) processing time is reduced by excluding the Fishnet cells which are outside coalfields and ii) risk of contamination to highly productive aquifers is reduced to minimum. The same filter and constraint are applied during the processing of other three domains and all four domains combined.

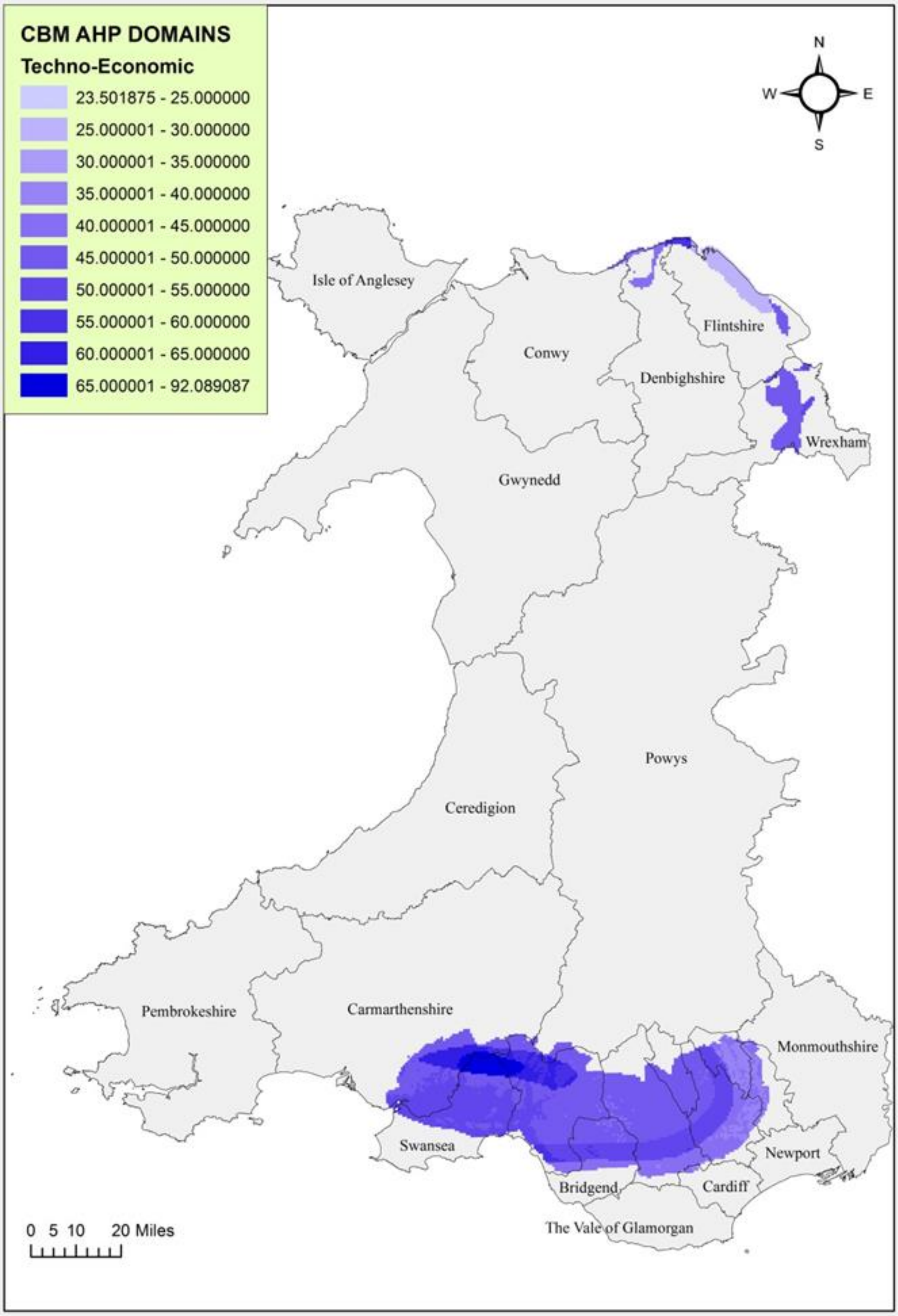


Figure 8.3 Level-1 CBM site selection – Techno-Economic domain

Map obtained in result of AHP based site selection tool application to the Techno-Economic domain is shown in Figure 8.3. This new maps is very similar to the map shown in Figure 8.1. This similarity is due to the fact that a high weight is assigned to the CBM resource density to emphasize the economic viability of the CBM site development. The slight variation found within a given CBM zone is caused by some of the other Techno-Economic factors considered in this application.

Fishnet cell values are arbitrary numbers (without unit) hence the cells with higher value are more suitable in terms of the indicators and their weights used in the analysis. Map reveals that the most suitable cells from the point of view of Techno-Economic domain are found in the northwest part of the South Wales coalfield. These areas are in the north of Swansea, northwest of Bridgend and in the southeast of Carmarthenshire counties.

8.3.2 Level-1 site selection applying AHP based site selection tool to the Environmental domain

As discussed in Section 8.1, one of the objectives of this application is to find those areas in Wales which are environmentally safer for the CBM development as compare to the other areas. This is achieved by designing a weighting scheme for the Environmental domain. This scheme promotes those Fishnet cells which are at a distance from critical environmental areas such as the protected sites. Furthermore, less weight is given to those areas where the existing emission rates of pollutants are comparatively higher than at other areas. Other, physical environment parameters are also considered for this purpose.

Headline indicators selected from Environmental domain are: i) Physical environment, ii) Protected sites, iii) Visual and sensory information, iv) Landform, v) Ambient air quality and vi) distance from brownfield, derelict and reclaimed land. Table 8.2 shows all the indicators, “Cost” or “Benefit” nature, also the assigned weights.

Table 8.2 Indicators used in the Environmental domain

Indicator	Weight	Nature
Physical environment	**10%**	**Benefit**
Air emissions 2008, Indicator of air quality 2008, Flood Risk 2009, Proximity to waste disposal and industrial sites	25% each	Cost
Protected sites	**30%**	**Benefit**
Distance from Site of Special Scientific Interest. Distance from Special Areas of Conservation. Distance from Special Protection Areas. Distance from Ramsar Sites. Distance from National Nature Reserves. Distance from Marine Nature Reserves. Distance from Areas of Outstanding Natural Beauty. Distance from Heritage Coasts. Distance from Biospheric Reserves. Distance from Biogenetic Reserves. Distance from Local Nature Reserves.	9.09% for each	Benefit
Visual and sensory information	**30%**	**Benefit**
Intrinsic evaluation matrix	***70%***	***Benefit***
Scenic quality, Integrity, Character, Rarity and Overall evaluation Discrete Classes Used: ‘Low’ and ‘Moderate’	20% for each class	Benefit
Land Form	***30%***	***Benefit***
Discrete Classes Used: ‘Lowland Valleys’ and ‘Flat lowland’	50% each	Benefit
Ambient air quality	**20%**	**Benefit**
Benzene, CO, NO2, NOX, PM2.5, PM10, SO2 and Ozone	12.5 % each	Cost
Distance from brownfield and derelict	**5%**	**Cost**
Distance from reclaimed land	**5%**	**Benefit**

Filter: CBM Resource > 0 AND 'Highly productive aquifer' not present

Details of the headline indicators, sub-indicators and their sources are provided in Geodatabase, Chapter 7. Some of the indicators from the AHP decision tree are qualitative (discrete) in nature such as the “Intrinsic evaluation matrix”. In such cases, the relative weight is assigned to the discrete classes directly. Only those discrete classes are shown in Table 8.2 which has assigned any weight. Rest of the classes have not been assigned any weight and they do not contribute to the process.

For protected areas, straight distance is calculated from the centroid of each Fishnet cell from the nearest edge of the protected area. “Benefit” nature is assigned to all these distance indicators to give more emphasis to the cells that are at a distance from the protected sites. Similarly, “Benefit” nature is assigned to the distance from regenerated land to avoid regenerated land being selected as potential sites. “Cost” nature is assigned to the distance from existing brownfield, derelict and tipping sites, since it is UK government’s policy to reuse such lands for useful purposes as explained in Section 7.5.6.

The ambient air quality modelling data, including Benzene, CO, NO_2, NOX, $PM_{2.5}$, PM_{10}, SO_2 and Ozone, is used as “Cost” indicators to avoid the polluted areas being highlighted in the potential zone. The Land Form data is a combination of slope and elevation and it reflects the shape and form of the land. In order to give more preference to low elevated and flat areas, these two classes are given 50% of the parent’s weight and rest of the classes are excluded from the process.

Visual and sensory information is used to assign low priority for the site selection in the areas where scenic beauty is evaluated to be high. The intrinsic evaluation matrix is assigned 70% of the weight within the visual and sensory information composite indicator. More weight is assigned to it in order to avoid scenic areas being highlighted as potential sites. The intrinsic evaluation matrix contains discrete variables, such as Scenic quality, Integrity, Character, Rarity and an Overall evaluation. For each of these qualitative variables, only two discrete classes have been included in the analysis. These classes are i) Low and ii) Moderate with a relative weight of 60% and 40% respectively. This ensures that the areas with extra ordinary natural landscape are supressed in the site selection process.

Similarly, Landform is a qualitative data and only two discrete classes have been assigned weight: i) Lowland valleys and ii) Flat lowland. The rest of the classes in Landform have not been included in the analysis hence not shown in the table.

The Score Range Procedure is used to scale all the indicators between 0-1 as explained in Section 5.4.1.1. The constraints used for the Environmental domain are the same as those used earlier in the Techno-Economic domain. Map obtained in result of AHP based site selection tool application to the Environmental domain is shown in Figure 8.4.

Fishnet cell values are arbitrary numbers (without unit) hence the cells with higher value are more suitable in terms of the indicators and their weights used in the analysis. The map reveals that according to the weighting scheme used and considering the Environmental domain, the Fishnet cells in the North Wales coalfields are randomly distributed and it is impossible to identify any clear trend. However, the distribution of Fishnet cell values in the South Wales coalfields reveals that the northeast part is relatively more suitable as compare to the rest of the South Wales coalfield.

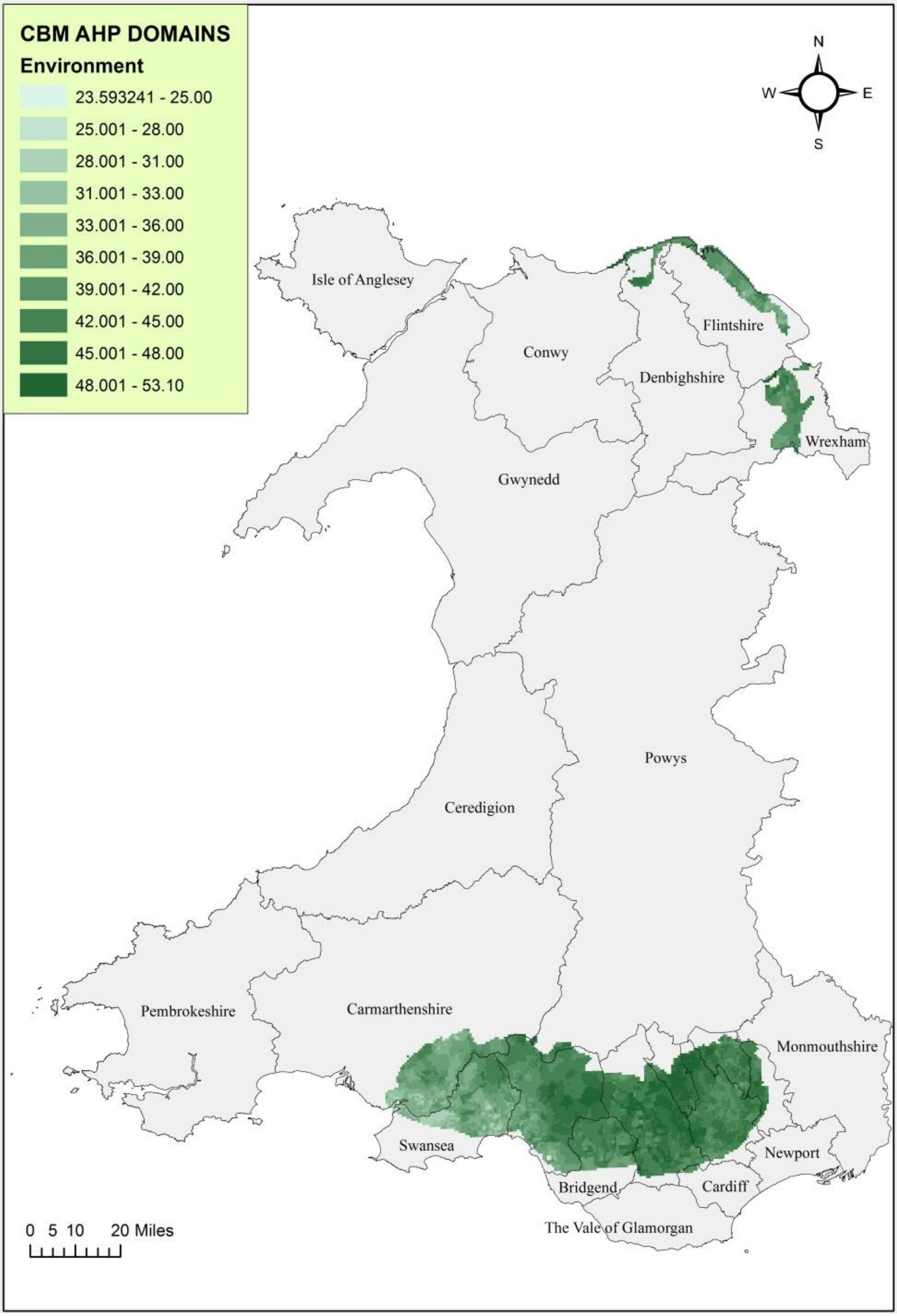


Figure 8.4 Level-1 CBM site selection – Environmental domain

8.3.3 Level-1 site selection applying AHP based site selection tool to the Public-Health domain

As discussed in Section 8.1, one of the objectives of this application is to incorporate Public Health indicators into the decision making and to highlight those areas for site selection which are not under stress in terms of Public Health domain. This is achieved by selecting appropriate indicators and designing a weighting scheme for the Public Health domain. This weighting scheme highlights the areas where general public health is better and supresses those areas where health conditions are worse as compare to the rest of the Wales. This approach refers to an Environmental Justice practice discussed in Section 2.8, which is suggesting that healthier communities living in the vicinity of the site are in better position to face environmental challenges caused by the development and running of the site. If such criterion is used for the site selection, it can increase the social acceptance of the unconventional gas developments.

Headline indicators selected from Public health domain are: i) Mortality, ii) Hospital admission rates and iii) Health related indicators used in WIMD-2011. Details of the headline indicators, sub-indicators and their sources are provided in Geodatabase, Chapter 7. Table 8.3 shows all the indicators selected from the Public Health domain to be used in the procedure. The "Cost" or "Benefit" nature of these indicators and the assigned weights are also given in the table. Constraints used for the Public Health domain are the same as in the Techno-Economic and Environmental domains.

Table 8.3 Indicators used in the Public-Health domain

Indicator	Weight	Nature
Mortality	**30%**	**Benefit**
All-cause mortality, Cancer mortality and	Equal % of	All Cost Indicators
Hospital admission rates	**20%**	**Benefit**
Cancer, Cardiovascular and Respiratory disease	Equal % of	All Cost Indicators
WIMD-2011 health per 100,000 people	**30%**	**Benefit**
Rate of limiting long term illness, all-cause mortality	Equal % of	All Cost Indicators
WIMD-2011 % of singleton low birth weight at	**20%**	**Cost**

Filter: CBM Resource > 0 AND 'Highly productive aquifer' not present

The Score Range Procedure is used to scale all the indicators between 0-1 as explained in Section 5.4.1.1. Map obtained in result of AHP based site selection tool application to the Public Health domain is shown in Figure 8.5. The darker red areas are those where public health conditions are already under stress as compare to the other areas which are shown in green. Green colour represents those cells where public health conditions are better and therefore they are considered as safer for unconventional gas development.

Map reveals that in terms of the Public Health domain and the given weighting scheme, areas in North Wales coalfield are better than South Wales coalfields. There is a variation within the South Wales coalfields, as the western side is more suitable comparing to the eastern side of the coalfield. Notably, this result is quite opposite to the Environmental domain results.

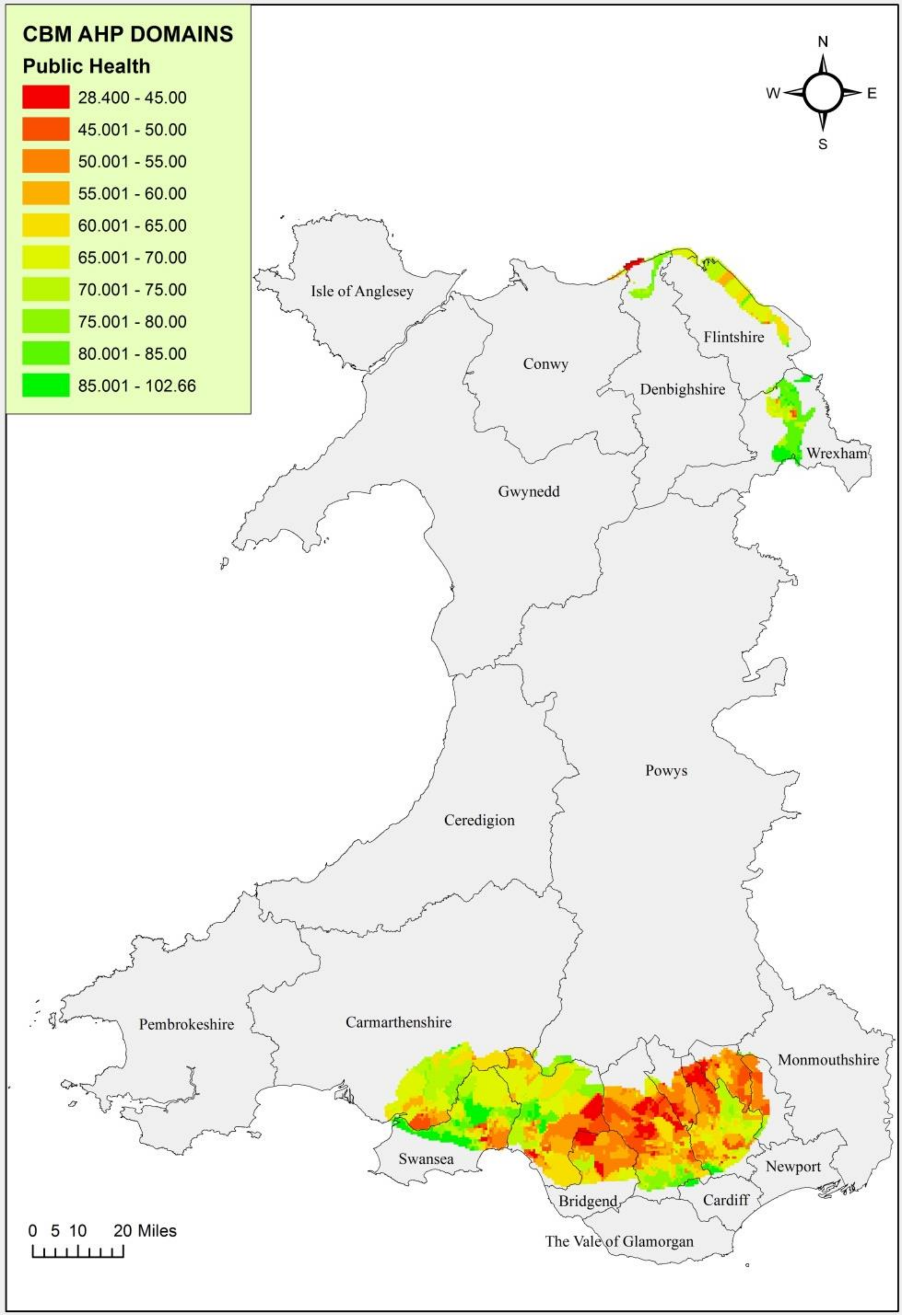


Figure 8.5 Level-1 CBM site selection – Public Health domain

8.3.4 Level-1 site selection applying AHP based site selection tool to the Socio-Economic domain

The Socio-Economic domain is one of the main components influencing decision making regarding the site selection for unconventional gas applications. Acceptance and sustainability of the developments are directly linked to impacts that these new technologies will have on the communities in the local and regional scale. The Socio-Economic domain contains indicators that reflect the key aspects, including anticipated public acceptance, social capital, multiple deprivation and condition of labour market.

Headline indicators selected from Socio-Economic domain are: i) Social acceptance, ii) Social capital, iii) Social disadvantage and quality of life, iv) Demographic structure and v) Labour market. Details of the headline indicators, sub-indicators and their sources are provided in Geodatabase, Chapter 7. Table 8.4 shows all the indicators selected from the Socio-Economic domain to be used in the procedure. The “Cost” or “Benefit” nature of these indicators and the assigned weights are also given in the table.

In this application, selection of the indicators and weighting scheme for the Socio-Economic domain are designed to highlight the areas in Wales where there is higher probability of social acceptance for unconventional gas technologies. Social acceptance is a complex qualitative phenomenon as explained in Section 2.6.1. Measures used to quantify social acceptance are discussed in details in Section 7.3.1.

Another objective of this process is to highlight those areas in Wales with potentially higher level of social capital. Social capital can help in the positive engagement of the communities in the utilisation of the new coal technologies in an effective and sustainable manner. Concept of social capital is explained in Section 2.6.2 and the GIS modelling for a general assessment of the social capital across Wales is provided in in Section 7.3.2.

Table 8.4 Indicators used in the Socio-Economic domain

Indicator	Weight	Nature
Social acceptance	**25%**	**Benefit**
Distance of developed land use areas from siting	20%	Benefit
Economic value of the land	20%	Cost
Recreational value	20%	Cost
Distance from existing industrial and mining areas	30%	Cost
Income level of the community	10%	Cost
Social capital	**20%**	**Benefit**
Civic participation	20%	Benefit
Social participation	15%	Benefit
Crime rate	5%	Cost
Views about the locality	30%	Benefit
Reciprocity and trust	30%	Benefit
Quality of life	**5%**	**Benefit**
Digital inclusion, % of households without a car or vehicle	Equal	Cost
Social disadvantage	**10%**	**Benefit**
Welsh index of multiple deprivation	Equal	Cost
population density	**20%**	**Benefit**
Total population	60%	Cost
Occupied houses	40%	Cost
Labour market	**20%**	**Benefit**
Employment by industry	**60%**	**Cost**
Mining and quarrying	70%	Benefit
Electricity, gas, steam and air-conditioning supply	15%	Benefit
Water supply, sewerage, waste management and remediation	15%	Benefit
Economic activity	**40%**	**Benefit**
Economically active: Unemployed	70%	Benefit
Economically inactive: Unemployed (Age 16-24)	30%	Benefit

Filter: CBM Resource > 0 AND 'Highly productive aquifer' not present

Finally, the Socio-Economic domain weighting configuration highlights those areas where unemployment rate is higher as compare to the other areas. Also, those areas are to be given preference where relevant job skillset is present. This is achieved by giving higher weights to mining and energy related sectors. This ensures that expected economic benefits of the unconventional gas developments are purposely focused in those areas where unemployment rate is comparatively higher and where a good percentage of relevant skillset is also present.

Geographical regions facing multiple deprivations are also highlighted in the site selection process. It is anticipated that natural resource development can be used to improve conditions of local communities through increased business activity, infrastructure development and subsequently job creation and fuel poverty alleviation. The weighting configuration also highlights the areas with sparse population as compare to densely populated areas in the potential CBM zones.

The Score Range Procedure is used to scale all the indicators between 0-1 as explained in Section 5.4.1.1. The constraints used for the Socio-Economic domain are the same as those used earlier for other domains.

Map obtained in result of AHP based site selection tool application to the Socio-Economic domain is shown in Figure 8.5. Red areas are showing less favourable areas in terms of the Socio-Economic domain and the given weighting scheme, whereas, green areas are more suitable. Map shows that more suitable cells have been identified in the North Wales coalfields, especially in the Flintshire County. In South Wales coalfield, more suitable cells are found in the middle and northern parts as compare to the rest of the coalfield.

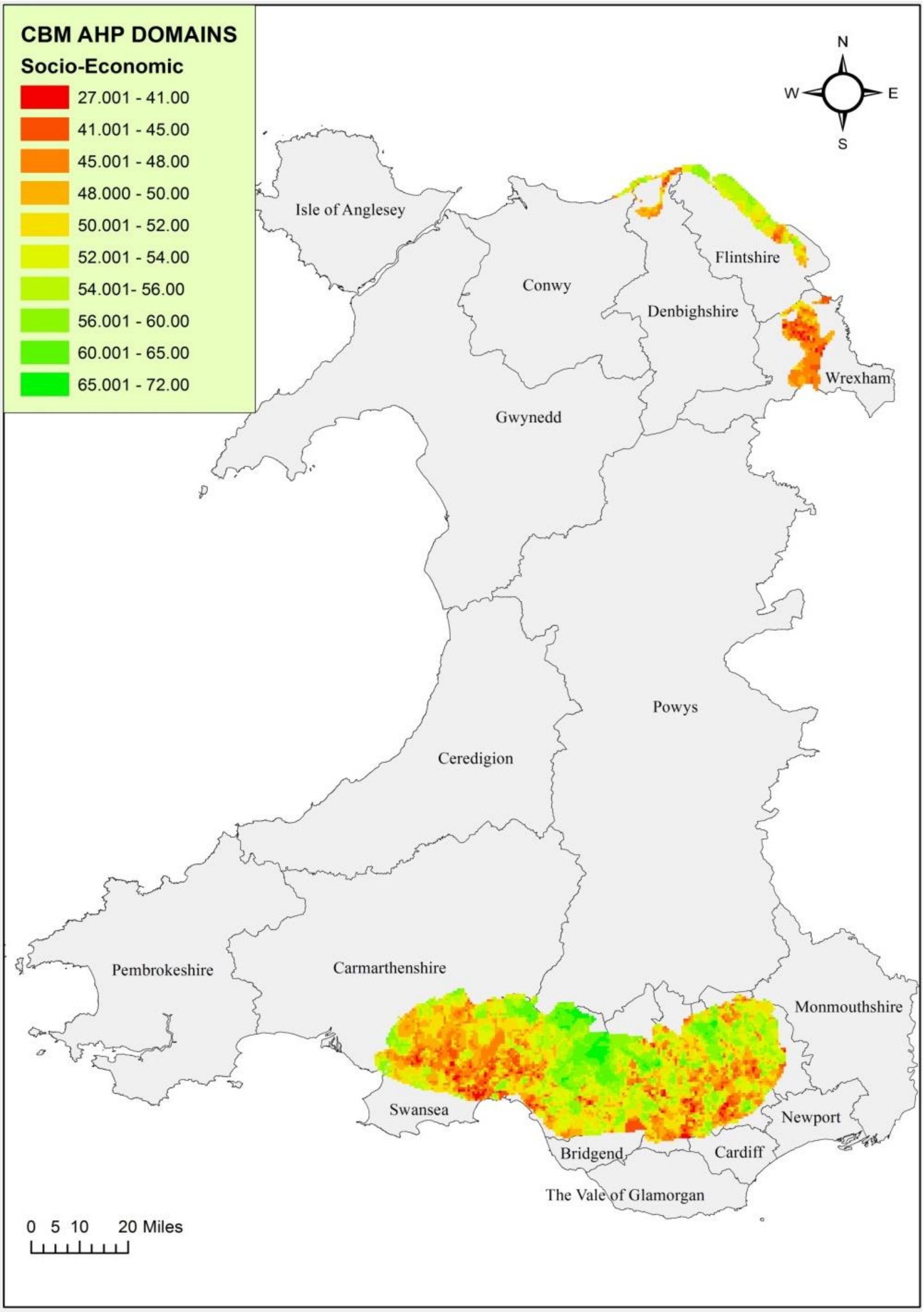


Figure 8.6 Level-1 CBM site selection – Socio-Economic domain

8.3.5 Level-1 site selection using AHP based site selection tool considering all four domains

A final step of the Level-1 site selection process using AHP tool considers all four domains. Equal weight (25%) is assigned to each domain in order to ensure that all four domains participate equally in the decision making process. In this way, the key aspects of the Environmental, Public Health and Socio-Economic domains are given equal importance to the Techno-Economic parameters. Hence, results of this analysis could be used for the transparent advocacy for the CBM development in Wales.

A sensitivity analysis has also been performed on the weighting scheme as discussed in Section 8.4. Number and location of the most suitable Fishnet cells can be sensitive to the assigned weights and sensitivity analysis is performed to assess this effect.

The geographical constraints and filter remains the same as used in other domains. Results of the weighting of all four domains together with their selected indicators are presented in Figure 8.7. Dark blue colour represents the most suitable cells whereas red colour represents the least suitable cells.

Resultant maps shows that northeast parts of the South Wales coalfield are more suitable for the CBM developments than rest of the CBM potential zones. The most suitable cells are found in the north of Neath Port Talbot and Swansea, also in the southeast of Carmarthenshire County. Fishnet cells in the North Wales coalfield, especially in the Flintshire region, are found to be the least suitable considering all four domains together. Although, according to the Socio-Economic domain results, this area was identified as the most suitable, but with the combined effect of all four domains, it has been superseded by other areas in the South Wales coalfield.

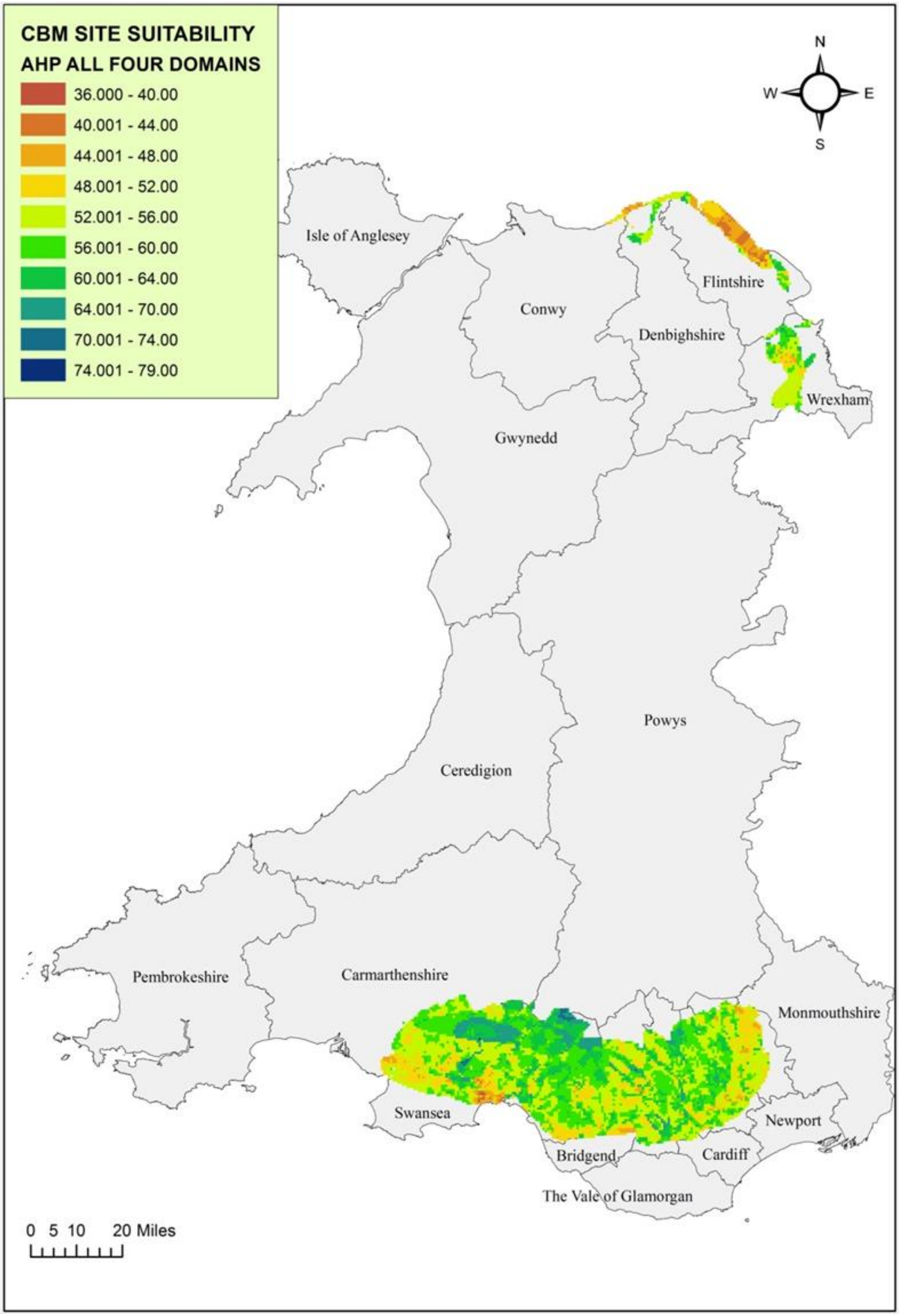


Figure 8.7 Level-1 CBM site selection - AHP all four domains

8.4 Sensitivity analysis

There are some uncertainties and risks associated with the Multi Criteria Spatial Decision Analysis (MCDA) techniques, including the AHP and WLC based site selection. The associated risks are mostly related to the weights assigned to different indicators in the hierarchical decision model (Feizizadeh et al. 2014). These risks are minimised by employing a sensitivity analysis to the site selection process as a function of weights assigned to the four main domains.

The sensitivity analysis feature is part of the AHP based site selection tool. The weight of the selected entity is increased or decreased by a fixed percentage. To adjust this difference in weight, weights of the peer entities at the same level of the hierarchy tree, are modified accordingly. The benchmarking value is set to calculate the number of Fishnet cells that are selected using the given weight scheme. In current analysis only the weights of the four domains, i.e. (a) Socio-Economic, (b) Environment, (c) Public-Health and (d) Techno-Economic are analysed for sensitivity.

This process can be applied at any level of the AHP decision hierarchy. Two different weight slabs are selected to check the sensitivity of each domain. At first, a weight slab of 20 is selected in the positive direction. The benchmarking value is set to 50. The results are shown in Figure 8.8. The weight of each domain is given at the X-Axis, whereas the number of selected features (Fishnet Cells) is given at the Y-Axis of the Graph. The number of features selected in each iteration is based on the user-defined benchmarking criteria. The weight of each domain is incremented one at a time and the rest of the weight difference is adjusted until there is no further increment possible.

The Public-Health domain is most sensitive to the weight increment. If its weight is increased from 25 to 45, the number of cells that are selected on the benchmark basis increases to almost three times. If the weight is further increased, the curve becomes steady and it remains almost the same when 65% or 85% weight is assigned to this domain. The Socio-Economic domain is second more sensitive in the analysis. The number of cells above the benchmark value is increased if the weight is increased. There is a steep increment in the number of selected cells when the weight is increased from 25% to 45% but afterward it takes a steady upward trend.

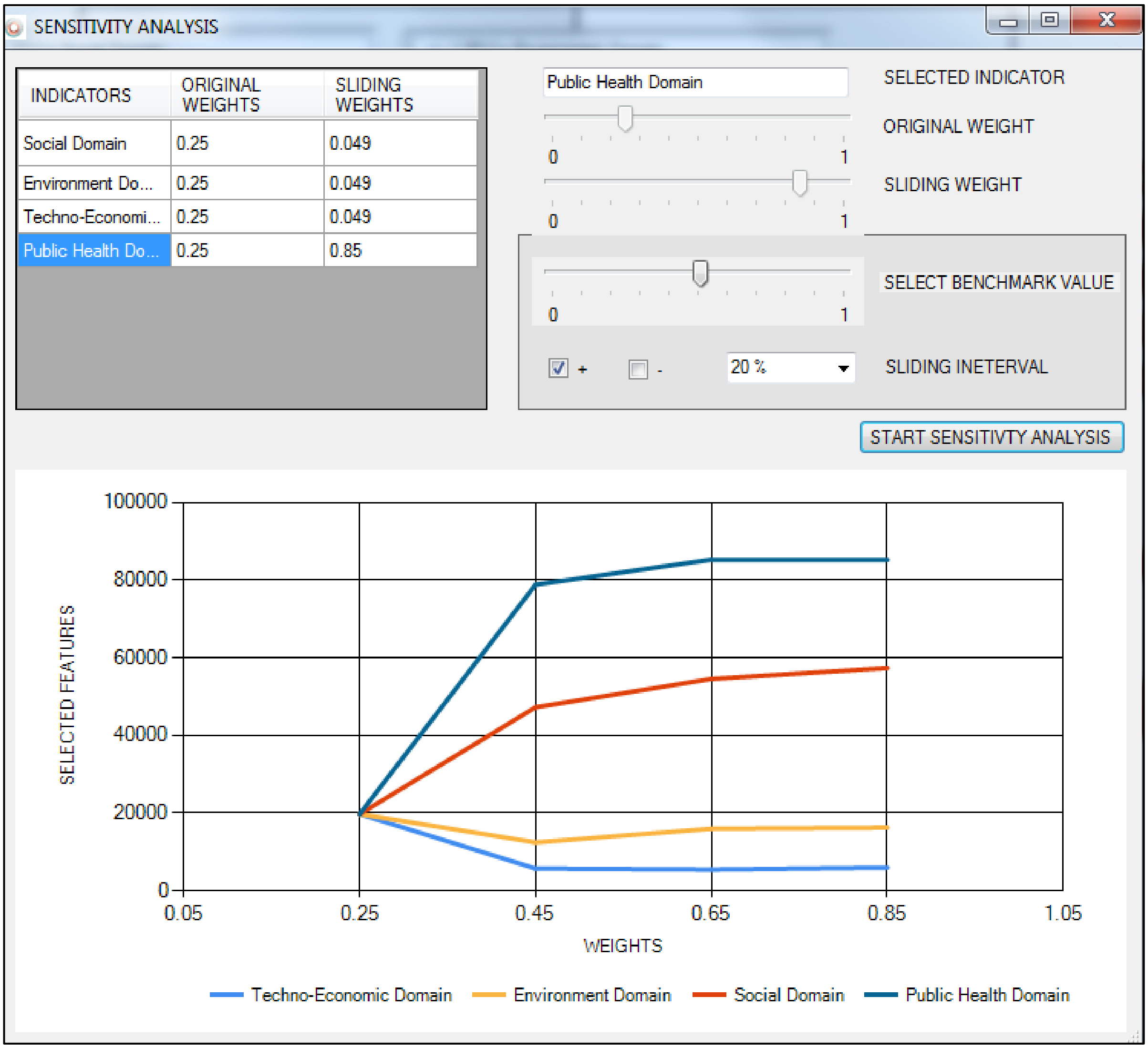


Figure 8.8 AHP sensitivity analysis - weight increment slab of 20%

The Environmental and the Techno-Economic domains are less sensitive to any change in their relative weights. It means that the number of selected cells is not changed to a great extent if their weight is changed, as compare to the other two domains. The number of selected cells is reduced if the weight is increased for Environmental and Techno-Economic domains, which is opposite to the trend shown by the other two domains discussed earlier. This trend is due to the fact that a very strict criterion has been used for Environmental domain so that the cells closer to the strategic environmental areas, e.g. protected sites, are less likely to be selected. Similarly in Techno-Economic domain, a large portion of the relative weight has been assigned to the CBM resource density. When the weights of these two domains are increased, the criteria become stricter, resulting in reduction of selected cells fulfilling the criteria.

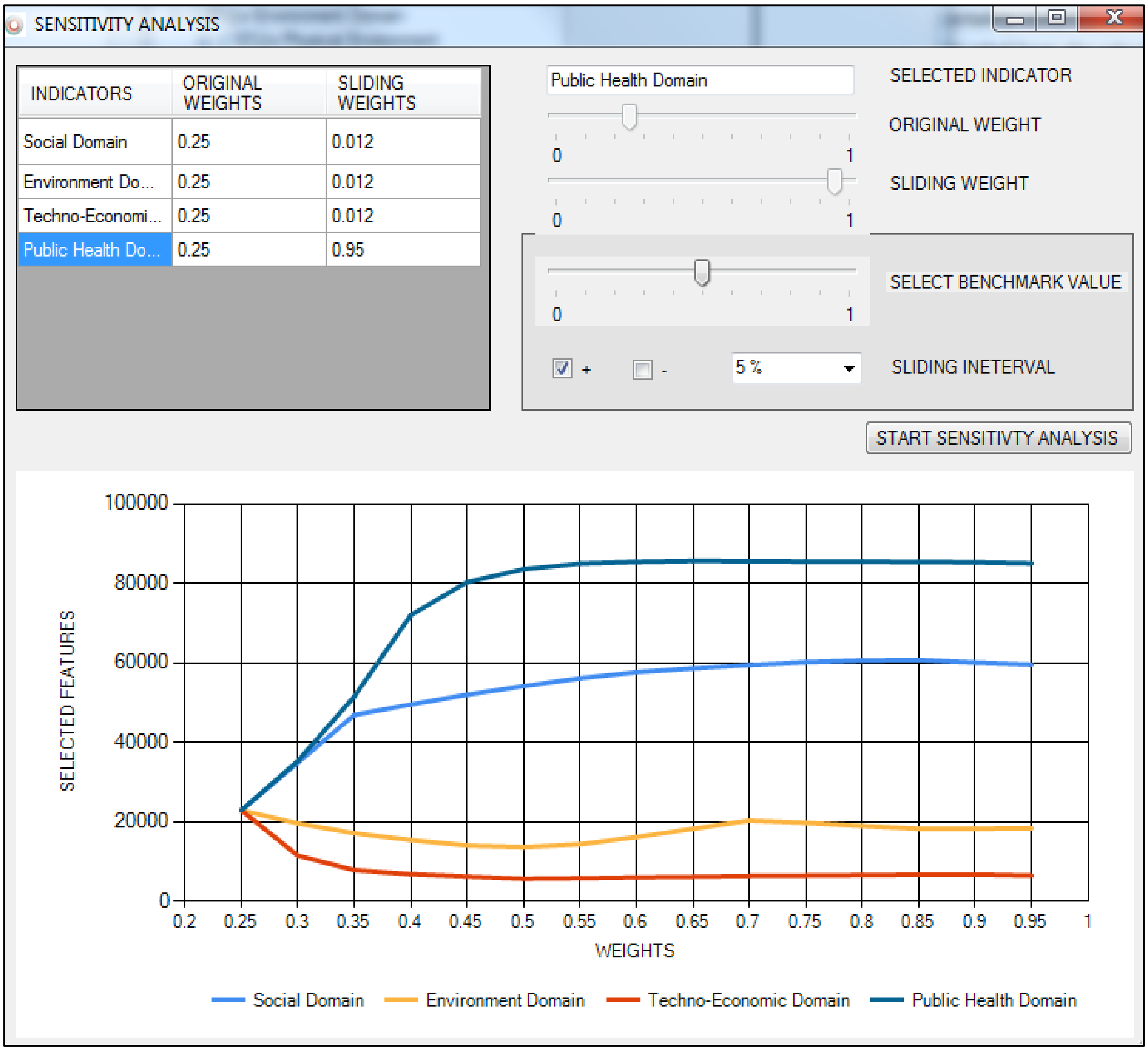


Figure 8.9 AHP sensitivity analysis - weight increment slab of 5%

During a second analysis, the sliding weight interval is reduced to 5 and the benchmark is kept at 50. The resulting graph is provided in Figure 8.9 and it shows a very similar trend to what is discussed. Therefore a change in step interval does not have a significant effect on the trend. Public-Health and Socio-Economic domains are more sensitive to increased weights. This justifies the selection of equal weights for all four dimensions in the analysis.

8.5 CBM site Ranking Using “SOM based site ranking tool

Following the results of Level-1 site selection process, a number of potential sites are identified for CBM development. At this stage, another spatial decision problem is faced by the decision makers, i.e. to rank and prioritise these potential sites (Irfan et al. 2014). Figure 8.7 shows a number of cells as highly suitable for CBM development according to the criteria used in AHP based site selection tool. In Level-2 site selection and ranking process, site ranking tools are applied to assigns relative ranks to the Fishnet cells according to the configuration of indicators in all four domains.

In order to facilitate the site ranking process, two site ranking tools are developed in the SDSS: i) SOM based site ranking tool and ii) GIS-based site ranking using neighbourhood analysis and comparison. Although both tools can be used for the site ranking purpose, the nature of ranking is different. The SOM based site ranking tool utilises the unsupervised artificial neural networks called Self-Organizing Maps (SOM). It finds natural clusters based on the site attributes within the data and then, ranks the sites in order. The second tool utilised the neighbourhood characteristics of each site in order to rank them. Ranking is carried out using CSM and TOPSIS methods as explained in Section 5.4.2.

In the first step of the Level-2 site selection process, all potential sites (Fishnet cells) are clustered and ranked using the SOM based ranking tool. The tool has been explained in detail in Section in 4.5.1. Five attributes are used for the clustering and ranking purpose at this stage, i.e. (a) The cumulative result of AHP process, (b) Socio-Economic Domain, (c) Environment Domain, (d) Public-Health Domain and (e) Techno-Economic Domain.

The individual domain scores are also incorporated in the clustering and ranking process ensuring that any areas dominant by one domain cannot supress other areas in the process. This facilitates the selection of those fishnet cells where all four domains and the cumulative AHP results are higher (relatively better) than others. All sites (Fishnet cells) in the CBM potential zone are divided into ten clusters according to their score for cumulative AHP and all four domains. Each cluster has been assigned a Rank (1-10) which is subsequently assigned to all the sites in a given cluster.

The initial total number of Fishnet cells across Wales is 86860 and then it is reduced to 8209 by applying constraint and filters in the AHP based site selection. After the processing of SOM based site ranking tool, number of cells in the top 3 clusters (Rank-1, Rank-2 & Rank-3) is reduced to 1739 cells. The Quantisation error is found to be 0.013 and the Topographic error is 0 which is an indication that the convergence is reasonable and the clustering and ranking generated from the SOM can be used with confidence as explained in Section 4.5.1. Fishnet cells having Rank-1, Rank-2 and Rank-3 are selected and exported as a GIS layer.

At this stage the number of potential sites is further reduced by application of additional filters to ensure that no site falls inside any critical environmental area or intersect with any infrastructure, including motorways, trunk roads, railways and airports. The GIS layers used for this purpose are provided in Table 8.5 along with any buffer distance if applied.

All those Fishnet cells that intersect with the features of these layers directly or with the buffers around them are filtered out. A buffer distance of 500 meters is applied to ensure a reasonable separation of the potential Fishnet cells from populated places, industrial and commercial areas, rivers and lakes.

Table 8.5 Filters applied on selected sites using GIS layers of infrastructure, stretegic environmental areas and human access

GIS Layer	Buffer Distance
Infrastructure	
Motorways, roads type A and B, rail road and airports	Intersecting cells
Strategic environmental areas	
All CCW protected sites, green urban areas, sports-leisure areas	Intersecting cells
Rivers, lakes and water bodies	500 meters buffer
Human Access	
Populated areas, industrial and commercial sites	500 meters buffer

After these filters are applied, 124 potential sites are left for further consideration. In results of this process, all sites in the North Wales coalfields have been filtered out and the remaining potential sites are situated in the South Wales coalfields.

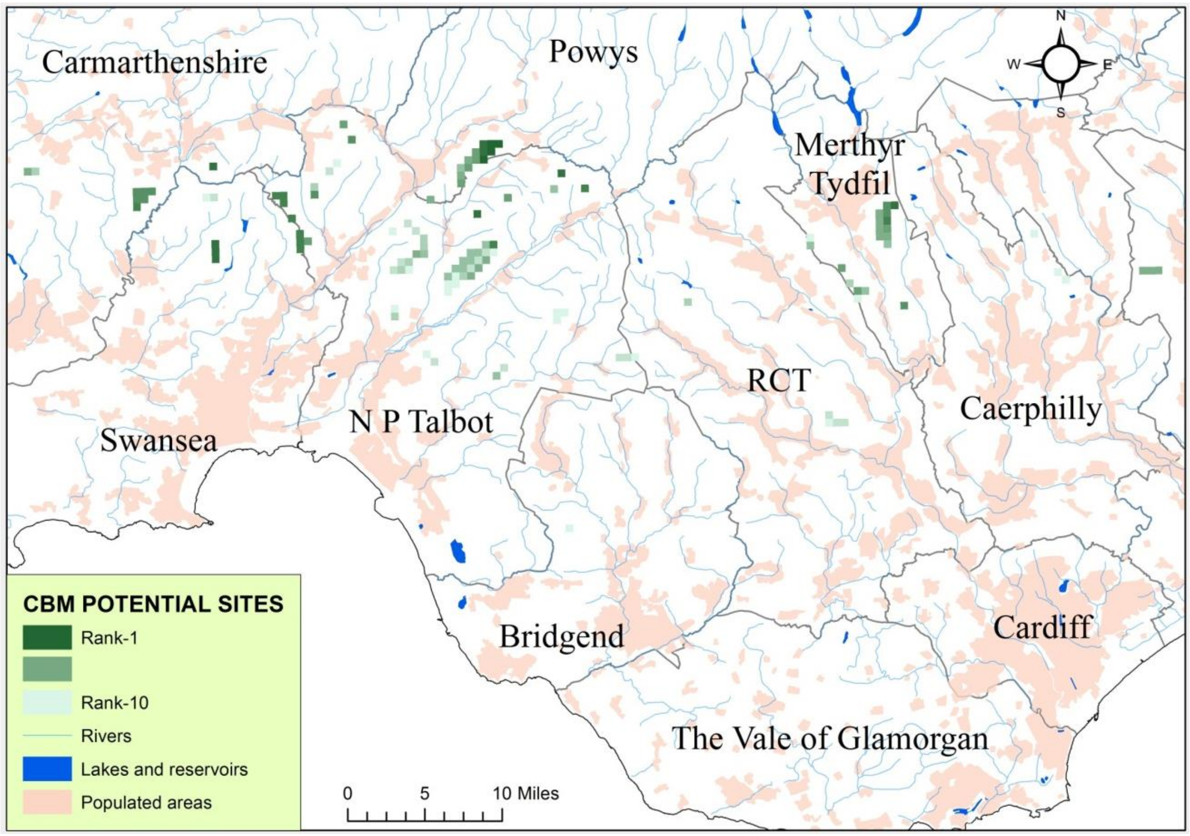


Figure 8.10 CBM potential sites after first level of site selection and filters

The resultant map is shown in Figure 8.10 along with populated areas, water bodies and rivers. For the visualisation purpose, remaining 124 potential sites (Fishnet cells) are clustered and ranked from 1-10 using the SOM-based site ranking tool. Best sites, i.e. the Rank-1 cells are shown in Dark Green colour on the map whereas the Rank-10 cells represent the least ranked sites in Light Green colour. Map shows that most of the remaining potential sites are located in the Northern part of the South Wales coalfields. These sites can be presented as the most suitable CBM sites with respect to the four domains together. Further filtration of sites at the second level of site selection process ensures that none of the proposed site: i) is located in any protected area, ii) lies near any river or water body, iii) is adjacent to any residential or commercial areas, and iv) intersects with any infrastructure. At the same time the proposed sites are located in the areas which are considered to be relatively better in terms of the Environmental, Public Health, Socio-Economic and Techno-Economic parameters.

8.6 CBM site Ranking Using "GIS based site ranking by neighbourhood analysis and comparison tool"

In the second step of the Level-2 site selection process, the GIS based site ranking using neighbourhood analysis and comparison tool is applied to further reduce the number of potential sites. The working of the tool is explained in Section in 5.4.2. This ensures that not only the selected sites but their neighbourhoods (surrounding areas) are also considered in the decision making regarding the engineering interventions with respect to key environmental, public health and socio-economic indicators (Irfan et al. 2014). For this purpose, neighbourhood analysis and comparison tool is applied to rank the remaining 124 sites.

Figure 8.11 shows the top 50 sites identified in a result of the neighbourhood analysis and comparison. The ranks have been assigned using the Criterion Sorting Mechanism (CSM) as discussed in Section 5.4.2.

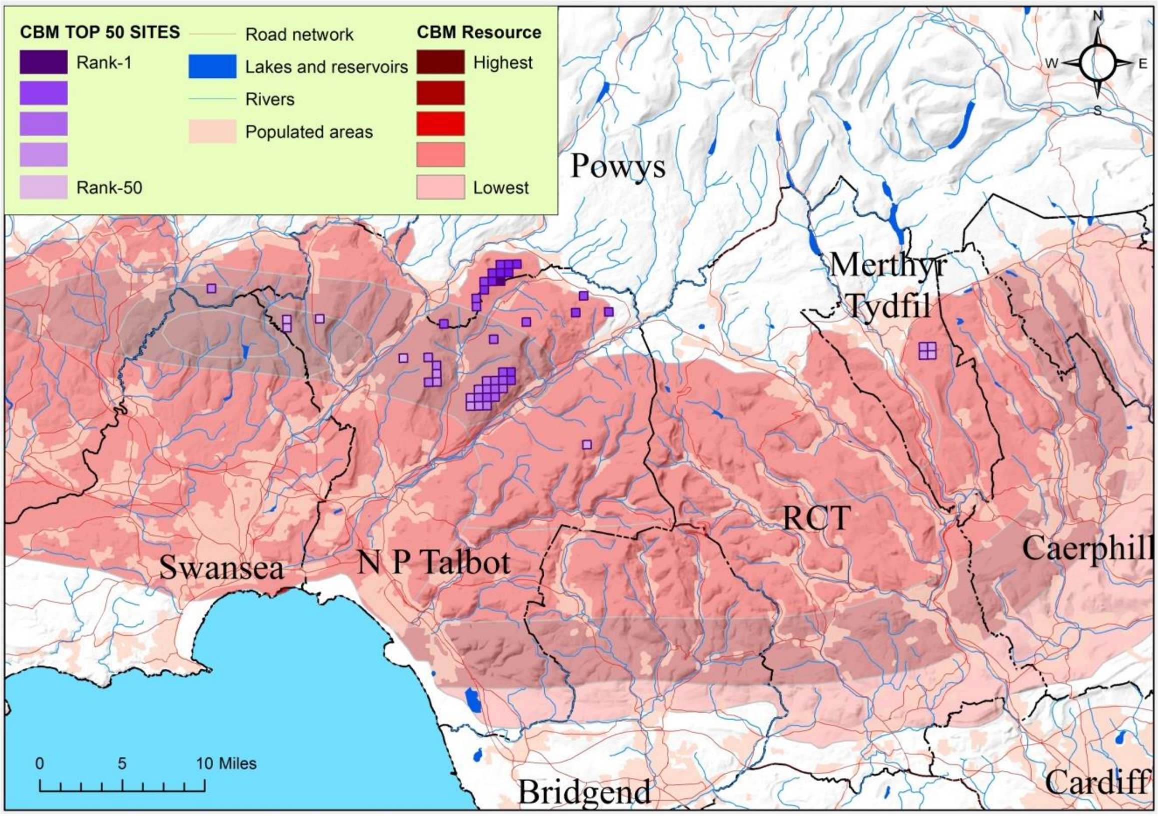


Figure 8.11 Top 50 potential CBM sites identified in the South Wales coalfields

The value of AHP and the values of all four domains are used in the neighbourhood analysis and comparison for ranking of the sites. This ensures the selection of sites that stand out from all sites, according to the of key indicator's values not only at the site-level but also at its selected neighbourhood. The top 50 sites for onshore CBM application in Wales are given in Table A.4 (Appendix A) along with their locations (X, Y), Ranks, AHP value and values for each of the four domains.

The location of these sites with respect to the CBM potential zones (Jones et al. 2004) is shown in Figure 8.11. It is noticeable that these sites are not situated in the best CBM zone in Wales. Most of them are situated in the second or third best zones in terms of CBM resource potential. It is due to the fact that environmental, socio-economic and public-health aspects were also considered together with the techno-economic parameters for the selection and ranking of the CBM sites.

8.7 Enhanced Coalbed Methane (ECBM) Scenario

As discussed in Literature Review chapter, in order to enhance the production of gas from CBM sites, CO_2 can be injected into the coal seams. This process is called Enhanced Coalbed Methane (ECBM) Recovery. This technology, not only increases the amount of gas released from coal surface but also offers an opportunity for long term storage of CO_2, helping to reduce greenhouse gases (White et al. 2005). In order to verify whether any of the earlier identified potential CBM sites can be used for ECBM, further GIS analysis is carried out using ArcGIS.

A GIS layer containing the major CO2 emitters in Wales is acquired from the Department of Energy & Climate Change UK (DECC 2012). This layer contains major CO_2 emitters from different sectors such as energy, cement, fertiliser and metal. To see if the potential CBM sites are suitable for ECBM, the locations of the potential CBM sites are analysed with the proximity of these emitter. Near tool of ArcGIS is used to calculate the distance of each CBM potential site from its nearest CO_2 emitter. The minimum distance was found to be 5.5 km and the maximum distance was 15.5 km. This proximity is a positive sign for the ECBM application as it reduces the cost of CO_2 transportation. For this reason, weight was assigned to the relevant parameter of the Techno-Economic domain, i.e. the "Proximity to major CO_2 emitters" as shown in Table 8.1.

The potential area for ECBM is encircled on the map presented in Figure 8.12. The nearest major CO_2 emitter is in close proximity (5.5 km) to an energy generation plant as shown with the arrows. The CO_2 emitted from the plant can be transported to the potential CBM sites through pipelines in a super critical form and then injected into the coal seams at a higher depth. This will enhance the production of the gas which then can be processed and transported back to the plant for electricity generation.

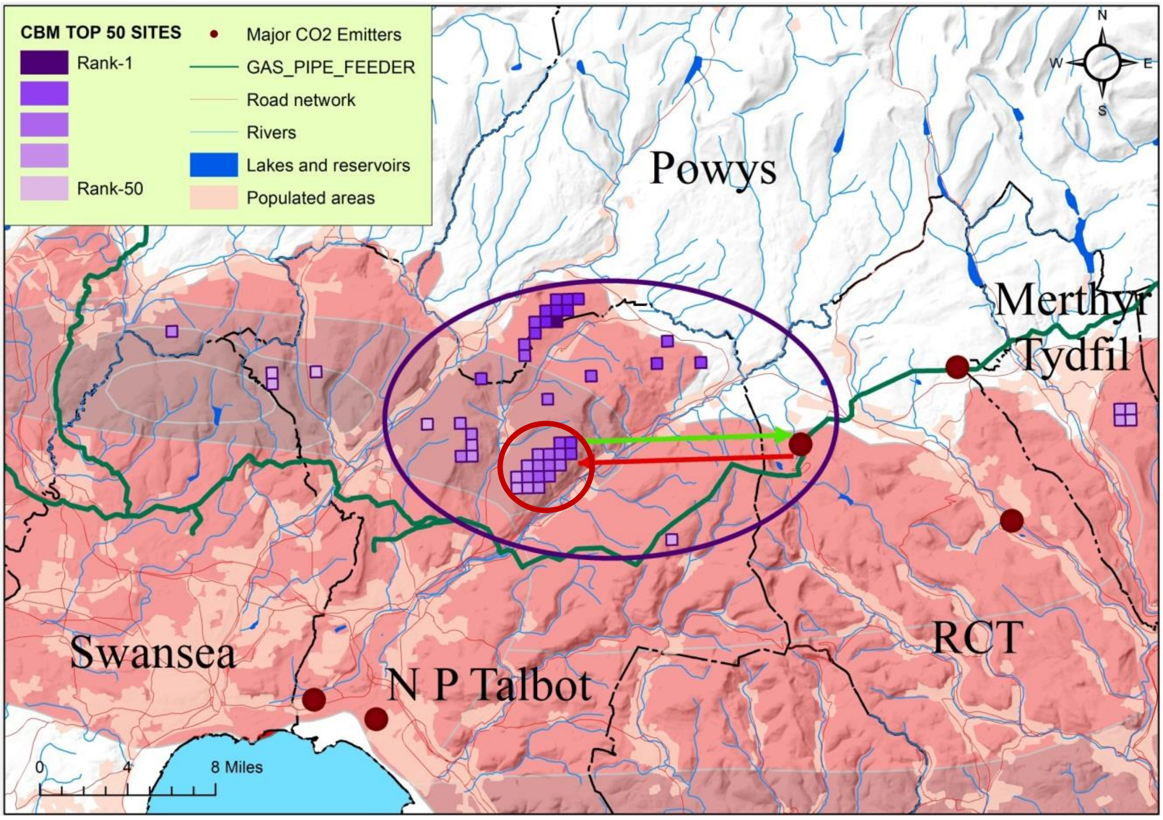


Figure 8.12 ECBM scenario – Major Co2 emitters and existing Gas-Feeder-Pipelines in the proximity of potential CBM sites

The location of existing gas feeder pipelines is also shown in the map. One of the main feeder pipelines runs through, in the close proximity of the CBM potential sites. Gas produced from the potential sites encircled in the map can be added to the existing network of the National Gas Grid and can be used for domestic and industrial purpose. This proximity between CBM sites and existing gas feeder pipelines reduces the cost of transportation of the produced gas from site to the consumers. For this reason, weight was assigned to the relevant parameter of the Techno-Economic domain, i.e. the "Proximity to existing gas feeder pipelines" as shown in Table 8.1.

As the map shows, there is a cluster (shown in red circle) of 17 potential CBM-ECBM sites (Fishnet cells of 500m^2) connected with each other. This cluster creates a big zone for the development of CBM-ECBM sites in South Wales Coalfields.

8.8 Geochemical baseline

As discussed in Literature review (Chapter 2), there have been incidents reported about the environmental damage caused by unconventional gas developments. These incidents are related to the escape of Methane into water table and presence of phenol and other hazardous chemicals in the vicinity of the sites. Therefore it is important to provide baseline information about presence and level of any Potentially Harmful Chemicals (PHC) around the potential CBM-ECBM sites proposed in this research.

In order to check the current status of PHC in sediments, geochemical baseline dataset is acquired from the BGS. It provides concentrations for Arsenic Cadmium, Copper, Iron, Lead, Magnesium, Nickel, Potassium, Tin, Vanadium and Zinc. The geochemical baseline data is processed using ArcGIS to interpolating each PHC point layer into a continuous raster map. The Inverse Distance Weighting (IDW) technique is used for the interpolation, with a cell size of $500m^2$ and same extent as that of Fishnet to align the two grids together. Using ArcGIS, value of each PHC is then extracted for the potential CBM/ECBM sites at their centroids.

In second step, datasets are acquired from the BGS containing existing values of Methane and CO_2 found in both bedrock and superficial geology from the natural sources. The data is acquired in vector format as polygons and directly assigned to the Top 50 CBM-ECBM sites using the intersection tool in ArcGIS. All these sites have known CO_2 and Methane hazard in the bedrock geology whereas some of the sites have this hazard from peat in the superficial geology. The values of geochemical baseline, CO_2 and Methane from natural sources are extracted for the top 50 CBM/ECBM sites and provided in Table A.8 (Appendix A).

8.9 Impact assessment

This section presents the final step in the SDSS application which is impact assessment of proposed CBM-ECBM sites. Impact assessment identifies negative and positive effects of an anthropogenic activity on environment, economy and public health. As discussed in Literature Review, it is important to carry out environmental and social impact assessment of certain developments. In some cases it is enforced by international, national or regional laws. In the UK, the Environment Agency is responsible for issuing policy guidelines for the environmental impact assessment. These guidelines are governed by the policies on air quality, biodiversity, landscape, adapting to climate change, noise and nuisance, waste management and water including flood risk (DEFRA 2013).

Rapid Impact Assessment Matrix (RIAM) based impact assessment tool of the SDSS is used for the impact assessment of proposed sites. RIAM based tool has been explained in detail in Section 5.5.1. In RIAM tool, each aspect of the project is evaluated against the environmental components and is assigned to one of the four categories, i.e. (a) Physical/Chemical, (b) Biological/Ecological, (c) Social/Cultural and (d) Economics/Operational. A novel theme-based RIAM tool has been developed in the SDSS as explained in Section 5.5.1. A new theme for CBM-ECBM site impact assessment was created and various RIAM components were added to this theme to analyse the impacts of proposed sites. RIAM components were created following results of the Literature Review that identified possible positive and negative impacts of CBM-ECBM sites. RIAM components used in this application are provided in Table A.5 (Appendix A).

As shown in Figure 8.13, most of the potential sites (Fishnet cells) are adjacent to each other forming clusters. Only a few sites are scattered and not connected to each other in clusters. Therefore, only one site (top ranked) has been selected from each geographical cluster for impact assessment. The selected sites are shown in Figure 8.13.

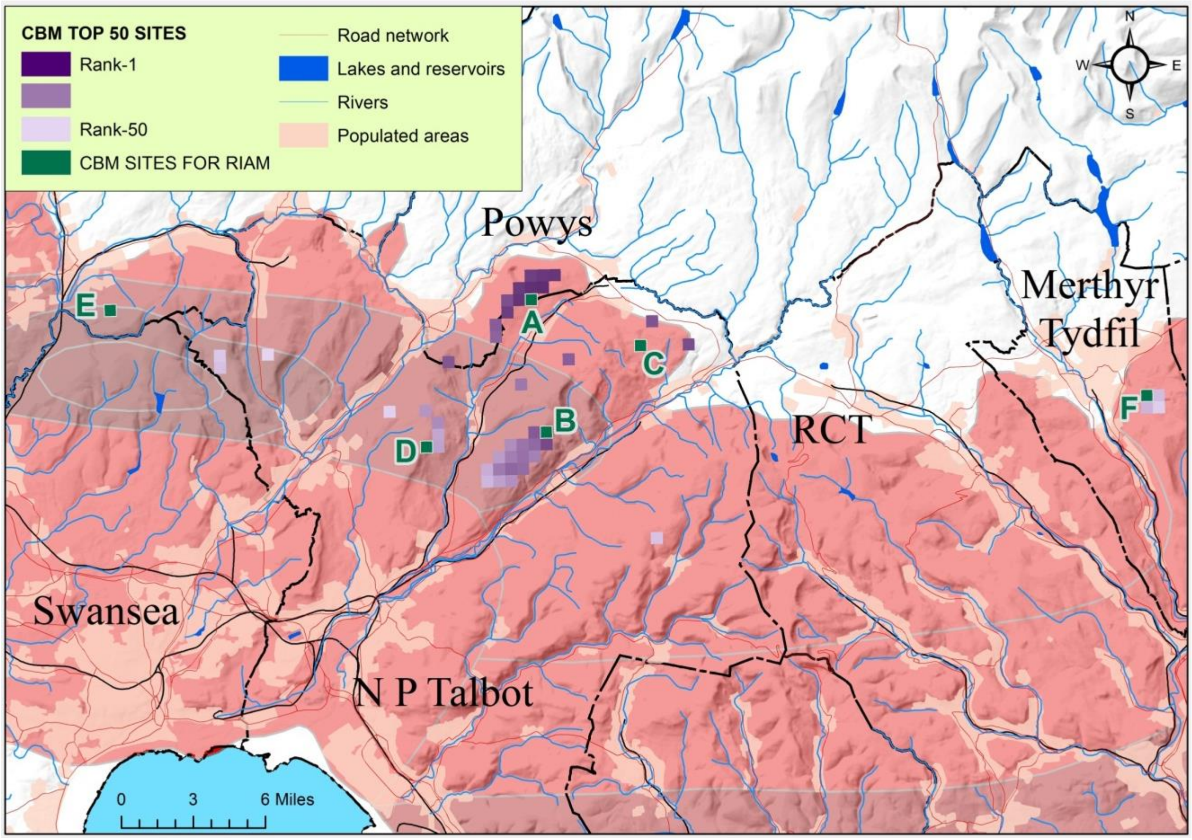


Figure 8.13 Selected CBM-ECBM sites for RIAM based impact assessment

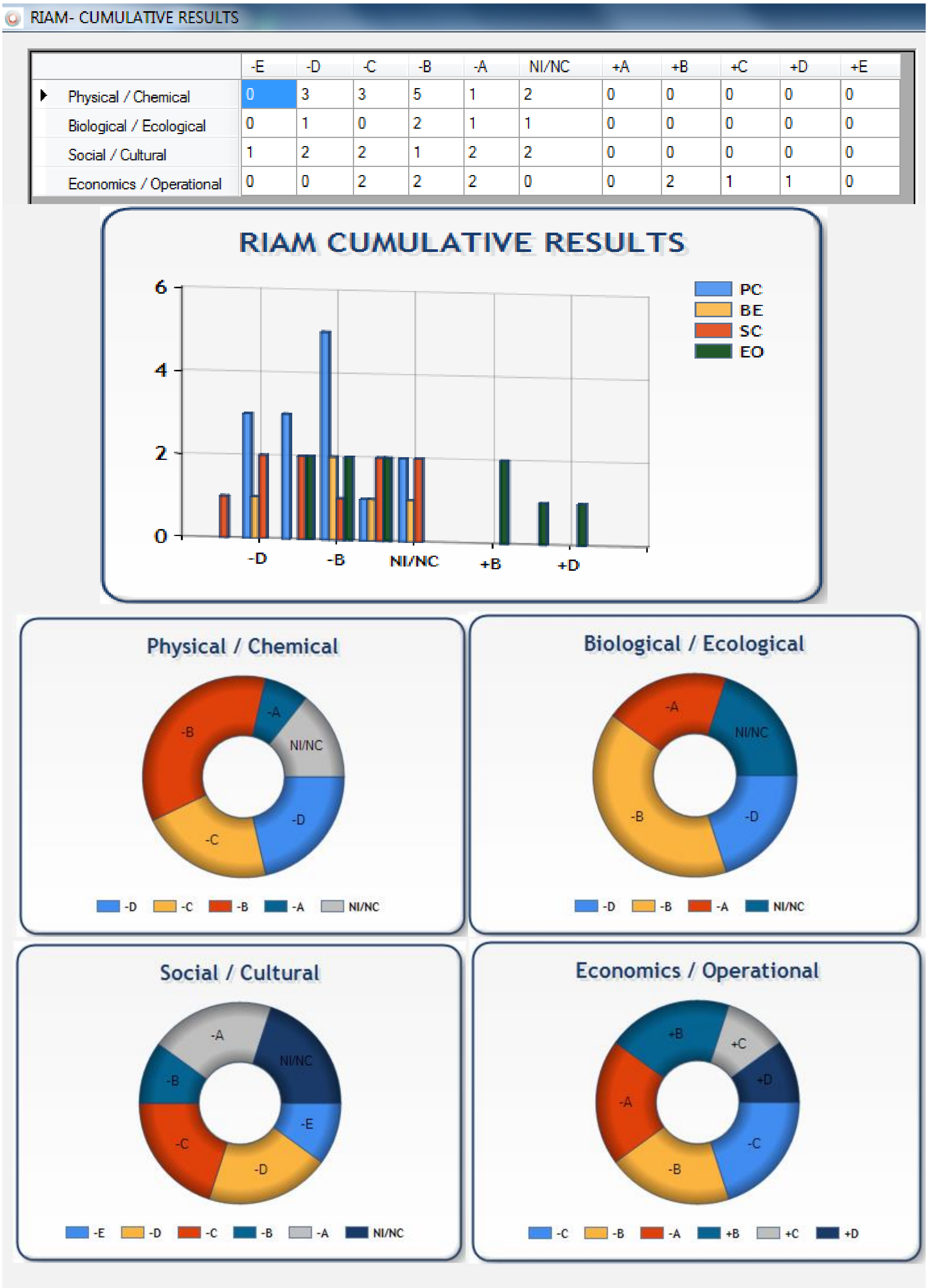
As described in Section 5.5.1, in addition to the conventional RIAM components and the scoring system, a new technique is developed, linking different spatial indicators with these components. RIAM components are going to have a negative or positive impact on these spatial indicators. This tool provides a functionality to analyse a given site against these indicators. There are two types of indicators, i.e. (a) continuous and (b) discrete. If a continuous indicator is linked with the RIAM component, its minimum, maximum and average value is calculated for the entire study area, in this case for Wales. Therefore, depending on the type of the spatial data (continuous or discrete) appropriate values are calculated in the neighbourhood of the given site by creating a buffer around it. Then the two values can be compared in the output table. This gives a better understanding of the possibility of an impact of the site on one or more key indicators in its neighbourhood. If the associated indicator is a discrete variable, then its different discrete classes are individually linked with the RIAM component by the decision maker. Each discrete class may be impacted negatively or positively by the engineering intervention at the given site.

When the site is analysed against the spatial components, the tool calculates the percentage of cells associated with the given discrete class in entire study area (Wales). The same is calculated for the site within a buffered region created around the site. Location of sites (X and Y coordinates) and buffer radius for neighbourhood is assigned to the Site Analyser tool. The results are generated in a tabular form and the status of each associated indicator can be compared between the entire study area and the buffered region. Five sites are selected from the top 50 sites. Each site represents a different cluster and it is then analysed against the indicators spatially linked to the RIAM components in the CBM-ECBM theme.

A buffer of 3km is selected around these sites to analyse the impact of RIAM components on their respective spatially linked indicators. The Site Analyser tool generates the results of the impact analysis of the RIAM components on their respective spatially linked indicators.

Not all of the RIAM components have indicators linked to them. The Comp. ID column shows the RIAM components unique ID which can be referred back to the Table A.6 (Appendix A). The indicator column shows the spatial indicators that are linked to the given RIAM component.

The tool calculates the minimum, maximum and average values of each continuous indicator in the entire study area and also in the buffered neighbourhood. The average value of each continuous indicator across the entire study area and within the given buffered neighbourhood is given for the selected sites (A-E) in Table A.7 (Appendix A). Similarly, for the discrete indicators, the percentage of Fishnet cells covered by a given discrete class of the indicator is given for the entire study area (Wales) and for the buffered regions around these sites. Values for the selected sites (A-E) are given in Table A.7 (Appendix A).



RIAM- CUMULATIVE RESULTS

	-E	-D	-C	-B	-A	NI/NC	+A	+B	+C	+D	+E
Physical / Chemical	0	3	3	5	1	2	0	0	0	0	0
Biological / Ecological	0	1	0	2	1	1	0	0	0	0	0
Social / Cultural	1	2	2	1	2	2	0	0	0	0	0
Economics / Operational	0	0	2	2	2	0	0	2	1	1	0

Figure 8.14 Cumulative results of RIAM based impact assessment tool for selected CBM-ECBM sites

Cumulative results generated by the RIAM based impact assessment tool are shows in Figure 8.14. Although, different RIAM components can be assigned to different sites and then analysed separately, for consistency, the same RIAM components have been used for all the sites analysed in this application.

The comparison of these values gives a better understanding of the current status of a given indicator in entire Wales and in the given neighbourhood of a site. This is helpful for the decision makers to select those sites where a RIAM component is supposed to have a negative impact on an indicators and the current status of the indicator is already worse than the national average. If a benchmark value is available for a given indicator then it can be used to compare with these values such as the upper limit of a particular emission type.

The results of each spatially linked RIAM component for every site are given in Table A.7 (Appendix A). A comparative analysis of the potential impacts of each site on spatial indicators, linked with RIAM components is provided below.

Both the indicators used for the air quality i) “Air emissions” and ii) “Air quality” are “Cost” in nature. These two indicators have also been used in the construction of the environment index of the WIMD-2011, as explained in Section 7.5. As it can be seen from the results given in Table A.7 (Appendix A), sites A, C and E are better than the Welsh average for “Air emissions” indicator, hence they are safer for site development. Values of “Air emissions” for Sites B, D and F are already higher than the Welsh average values, suggesting that any further deterioration in the situation is not desirable. For “Air quality” indicator, only site F has a better value than the Welsh average and rest of the sites are already worse than the Welsh average for this indicator.

Distance from geological dykes and fault lines are both “Benefit” indicators. The further the sites are from these geological features, the safer they are considered. Since it is a measured distance, therefore the comparison with Welsh average value is not suitable. However, all six sites are quite far from these two geological features. Among all the sites, Site F is the closest to any geological fault line however still at a distance of 1km away from it.

Distance from protected sites, i.e. Special Areas of Conservation (SAC) and Special Protection Areas (SPA) are both “Benefit” indicators and further the sites are from these two areas, more suitable and safer they are considered. These two indicators are linked with multiple RIAM components including the risk of habitat fragmentation, loss of habitat and impact of noise on the wildlife. As shown in the results given in Table A.7 (Appendix A), none of these sites is very close to such protected areas. The closest site is the Site C which is still at a distance of 2.5Km from the boundary of SAC.

During the filtration process all cells that are within a distance of 500 meters from any populated place were removed. Therefore, there is no risk associated with any site for vacation of properties or residents moving out of the area.

Comparing the positive impacts, especially in terms of job creation, there is a good percentage of local population associated with the mining, quarrying and construction works. This suggests that relevant skilled labour is already present in the proximity of sites. In terms of the WIMD income, employment and overall indexes, Sites C and F are situated in more deprived areas as compare to the other sites. Business generation and job creation at these sites will have a positive impact and it will help to address the problem of multiple deprivations.

Considering the economics of the site development and its running cost, Site B is the closest to one of the major CO_2 emitters in Wales. This provides an opportunity since the CO_2 can be transported to these sites with lower cost of transportation. At the same time, they could be potential consumers of the produced gas. This provides a favourable source-sink scenario for the ECBM development in Wales. Other sites are relatively far from any major CO2 emitters in Wales. Also, sites B is very close to the existing gas feeder pipelines. If the produced gas is processed at the site, then it can be directly injected into the national gas grid, hence cost of transportation will be reduced.

The RIAM components that are linked with the qualitative indicators are given in Table A.13 (Appendix A) showing the percentage of areas covered by a given discrete class of a discrete variable both in entire Wales and within the given buffered regions around the sites.

All discrete indicators linked to RIAM components are set to have negative impact on the given discrete class of the variable. As it can be seen, all six sites and their given neighbourhood regions have 0% area lying over highly productive aquifers and they all are situated in the moderately productive acquires zones. Neighbourhood of all six sites have negligibly low percentage of areas that have risk of CO_2 and methane hazards from the natural sources in superficial geology. Also, all the sites have 0% areas that are designated as permanent crop covers in the Corine land Cover dataset of 2006. Hence there is no risk of land cover change for the important class of croplands. All these factors are good for the site economics, local food security, environmental safety and consequently for the social acceptance of CBM-ECBM developments in Wales.

Sites B, C and D have quite high percentage of the areas covered by Conifer forest, which is the highest (42.3%) for site B. This implies that there is a risk of Conifer forest lost if the sites are established at these locations. Rest of the forest cover types, e.g. broad leave and mixed forests are less affected by any of the sites. Sites and E and F are having the least percentage of any type of forest in their surrounding regions and therefore the least negative impact is expected at these sites in terms of forest loss.

8.10 Conclusions

This chapter presents an application of the SDSS developed in this research. SDSS has been used to identify potential sites for CBM-ECBM development in Wales. The main aim of this case study was to demonstrate the value of the developed SDSS in facilitating the spatial decision making involved in Geoenergy applications. The objectives were set to identify CBM sites in Wales (UK) that should: a) have minimum negative impact on the environment, b) have positive impact on the socio-economic conditions of the area, c) are located in areas where public health conditions are comparatively better and d) are techno-economically more viable, compared to other areas. These objectives have been achieved and the results can be used to argue for the effective utilisation of the CBM resource in Wales (UK).

At the first level of the site selection process, AHP based site selection tool has been used to identify the most suitable sites for CBM development considering the key environmental, public health, socio-economic and techno-economic indicators. A sensitivity analysis has been carried out to analyse uncertainties linked with the domain weights.

At the second level of the site selection process, SOM based ranking tool has been applied to the most suitable sites to cluster and rank them according to the key indicators. Further spatial filters were applied to exclude any sites that may interfere with i) strategic environmental areas, ii) hydrological features, iii) infrastructure and iv) populated areas. As a result of the site ranking carried out based on neighbourhood analysis and comparison, the top 50 potential CBM sites in Wales were identified. This two-level site selection and ranking process has successfully narrowed down the CBM potential zones in Wales and helped to identify and prioritise the top 50 sites for further investigation. Also, the ECBM scenario is presented for the potential sites identified earlier. The proximity of the potential sites from the nearest CO_2 emitters and existing gas feeder pipeline network has been analysed. During the AHP based site selection process, weight was assigned to the indicators, representing the distance of each cell from the existing gas infrastructure and CO_2 emitters. This has been reflected in the results and a number of top sites with a potential for ECBM have been identified. This provides an opportunity for the long term storage of CO_2 at the selected locations in South Wales coalfields, enhancing the economic viability of the site at the same time.

Notably, the top CBM-ECBM sites identified in this application are not located in the best CBM zone in Wales, as identified in (Jones et al. 2004). However, these sites are still located in the top 3 best CBM zones in terms of CBM density. This shows that apart from techno-economic viability, key environmental, public health and socio-economic aspects have been given equal importance and influenced the outcome.

Finally, an impact assessment has been carried out on the selected top ranked CBM-ECBM sites. For this purpose, RIAM based site impact assessment tool has been applied on five sites selected from the fifty best sites. These five sites are selected from different clusters of sites, as the impact assessment is based on the key indicators in the neighbourhood of sites and those sites in very close vicinity are likely to exhibit similar results. The new technique of linking spatial data with RIAM components introduced in this research has also been applied in this case study. RIAM components that are identified in this research and their potential impact on different factors from Environmental, Socio-Economic and Public Health domains are linked together with the most relevant spatial data. This helped to analyse the spatial impact of each site on the key aspects. The results of impact assessment can also feed into the site selection process helping to identify the safest sites for development.

Also, a geochemical baseline is developed for the most suitable CBM-ECBM sites. This baseline provides the current level of PHCs found in the sediments around these sites. This can be helpful to monitor sites for any potential environmental degradation.

The details of top 50 sites in terms of AHP value, all four domain values, indicator score and geochemical baseline information has been provided in Appendix-A. Brief detail of the top 5 selected CBM-ECBM sites is provided in Table 8.6. As discussed earlier, the area of each site (Fishnet cell) is 500m^2. The centroid location of the 5 selected best sites is provided in Table 8.6 in the format of British National Grid.

Table 8.6 AHP domain values and location of selected CBM-ECBM sites

		AHP Domains				British National Grid	
Site Name	**AHP Value**	**Techno-Economic**	**Socio-Economic**	**Environment**	**Public Health**	**X**	**Y**
A	71.04	50.46	72.20	66.70	65.44	281750	210250
B	66.07	61.95	71.99	53.02	60.52	282250	204750
C	62.17	50.69	72.02	51.27	60.52	286250	208250
D	62.89	63.23	68.02	50.76	58.52	277250	204250
E	67.36	92.23	67.22	44.62	57.76	264250	210250
F	63.01	49.80	67.36	63.14	53.23	307250	205750

Results reveal that Site A is the overall best site with highest AHP score. Considering ECBM scenario, Site B is the closest to one of the major CO_2 emitters and in a close proximity to the existing gas feeder pipelines in Wales. Site B is connected to a cluster of few other sites that came up in the top 50 CBM sites in the entire Wales, considering multicriteria approach. This provides a considerable gas resource at one location that is also suitable for ECBM. The application results can be useful to argue the development of CBM-ECBM in Wales in a manner that is useful for the society and safer for environment and public health.

References

DECC. 2012. *Department of Energy & Climate Change, Onshore GIS data* [Online]. Available at: https://www.gov.uk/oil-and-gas-onshore-maps-and-gis-shapefiles#onshore-gis-data [Accessed: 07 May].

DEFRA. 2013. *Assessing environmental impact* [Online]. Dpartment for Environment, Food & Rural Affairs. Available at: https://www.gov.uk/assessing-environmental-impact-guidance [Accessed: 11 June].

Feizizadeh, B.,Jankowski, P. and Blaschke, T. 2014. A GIS based spatially-explicit sensitivity and uncertainty analysis approach for multi-criteria decision analysis. *Computers & Geosciences* 64(0), pp. 81-95.

Irfan, M.,Koj, A.,Sedighi, M. and Thomas, H. 2014. GIS based site ranking using neighbourhood analysis and comparison. *ThinkMind*, pp. 51-57.

Jones, N. S.,Holloway, S.,Smith, N. J. P.,Browne, M. A. E.,Creedy, D. P.,Garner, K. and Durucan, S. 2004. *UK coal resources for new exploitation technologies*.

White, C. M.,Smith, D. H.,Jones, K. L.,Goodman, A. L.,Jikich, S. A.,LaCount, R. B.,DuBose, S. B.,Ozdemir, E.,Morsi, B. I. and Schroeder, K. T. 2005. Sequestration of carbon dioxide in coal with enhanced coalbed methane recovery - A review. *Energy and Fuels* 19(3), pp. 659-724.

9

CONCLUSIONS

9.1 Introduction

In this chapter, the overall research work is summarised and conclusions are drawn. In addition, suggestions are made for further research.

This research aimed at designing and developing an impact-based, multicriteria Spatial Decision Support System to address a wide spectrum of spatial decision problems related to Geoenergy and Geoenvironmental applications.

In order to achieve this aim, the following objectives have been identified for the research:

- The design and development of a multicriteria Spatial Decision Support System (SDSS), thereby facilitating decision making process related to site selection, site ranking and impact assessment.
- The development of a set of adequate advanced analytical techniques and their integration

into the SDSS to tackle semi-structured spatial decision problems.

- The exploration and identification of the key socio-economic, environmental, public health and techno-economic factors and indicators to be adopted in informed, impact-based decision making related to Geoenergy and Geoenvironmental applications.
- The exploration and development of a geodatabase containing the key socio-economic, techno-economic, environmental and public health data for the study area, i.e. Wales, UK.
- The investigation of the potential of CBM-ECBM development in Wales considering the socio-economic, environmental, public health and techno-economic aspects.

These objectives have been achieved successfully. The design and development aspects are summarised in Section 9.2. The analytical modelling techniques developed and their variations introduced in this research are outlined in Section 9.3. The geodatabase developed for Wales is highlighted in Section 9.4. The application of the developed SDSS to investigate the potential of CBM-ECBM application in Wales is discussed in Section 9.5. Finally, the overall conclusions are drawn from the research work and suggestions are made for further research.

9.2 SDSS design and development

In this research, the concept of SDSS has been expanded and efforts have been made to design and develop a system that is capable of addressing the commonly faced spatial decision problems related to the applications of Geoenergy in an integrated manner. These problems are categorised as i) site selection and ranking, i) site impact assessment and iii) spatial knowledge discovery.

The system has been designed and developed considering the three key components of SDSS: a) Geodatabase, b) Model Base and c) User interface. SDSS has been developed following a modular approach where each component has a particular functionality embedded into it.

First a main skeleton of the software has developed using DotSpatial and Microsoft C# .Net programming language. This is the core of the system which links everything together. Then different analytical models have been developed and integrated into this main system. The modular design of the SDSS allows extending the capabilities of the system without any major structural changes to the overall system.

Overall functionalities of the system are bundled together into four distinct groups. The first three groups cover analytical modules related to a) Site selection and ranking, b) Impact assessment and c) spatial knowledge discovery. The fourth group contains the tasks related to the geodatabase management.

To conclude, an integrated yet scalable design of the SDSS helped achieving the design considerations. The modular approach adopted for the development, enables expansion of the system to incorporate new analytical modules. The user-friendly interfaces provide an easy interaction between the system and the decision makers.

9.3 Model-base development and verification

As discussed in Chapter 3, the Model Base is the brain of the system and it contains a set of appropriate analytical modelling techniques for the target spatial decision problems. A number of analytical modelling techniques have been developed and integrated into the SDSS for the considered spatial decision problems. These techniques are broadly categorised as i) Multi Criteria Decision Analysis (MCDA) and ii) Artificial Intelligence (AI) techniques.

For site selection and ranking, three tools have been developed and integrated into the SDSS: i) Analytical Hierarchy Process (AHP) based site selection tool, ii) Self-Organizing Maps (SOM) based site ranking tool and iii) Site ranking by neighbourhood analysis tool using Criterion Sorting Mechanism (CSM) and Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS).

For spatial knowledge discovery, three tools have been provided: i) Self-Organizing Maps based clean correlation finding tool, ii) Parallel Coordinate Plots (PCP) and iii) General Regression Neural Network (GRNN) based regression analysis tool.

For site impact assessment, two tools have been developed: i) Rapid Impact Assessment Matrix (RIAM) and ii) Traffic impact assessment tool. The GRNN based prediction tool can also be used for the purpose of site impact assessment and prediction.

Analytical modules developed in the SDSS, utilise a number of existing techniques. As discussed in Chapter 4 and 5, some new features have also been added to these existing techniques in this research. For example, the data ordering capability of one-dimensional SOM has been used for the purpose of site ranking. Similarly, a spatial parameter has been added to the GRNN tool to support Geographically Weighted Regression (GWR) analysis. Also, a mechanism has been developed to link spatial data with the RIAM components to assess and analyse spatial variations of the environmental and socio-economic impacts.

In order to verify the accuracy of the code developed in these analytical modules, reliable platforms such as ArcGIS, Matlab and GeoSOM have been used. First a dataset is prepared to verify the given SDSS module. Then using the same parameters, this dataset is analysed in the given SDSS module and using one of the above mentioned existing reliable tools for comparison. Finally, the results are compared and similarities and dissimilarities are discussed.

To conclude, the integration of reliable MCDA and AI techniques for the Model-base of the SDSS has helped addressing the considered spatial decision problems under one system. Incorporation of semi-supervised and un-supervised AI techniques makes it easier to use by the non-specialised decision makers. The additional spatial knowledge extraction and geovisual-analytics tools will support the decision makers in benefiting more from the underlying spatial information. Also, the verification process has increased the confidence in the utilisation of the SDSS and its different analytical modules.

9.4 Geodatabase development

Geodatabase serves as an essential component of the spatial decision support system. A geodatabase has been developed for the SDSS using SpatiaLite technology. As discussed in Chapter 2, there are certain environmental, socio-economic, public health and techno-economic aspects that need to be incorporated into the decision making context. Therefore, a number of indicators have been identified, acquired and incorporated into the geodatabase for the study area, i.e. Wales, UK. These key indicators are categorised into four domains: i) Socio-Economic, ii) Environmental, iii) Public Health and iv) Techno-Economic.

A number of indicators were directly used, while GIS modelling was carried out to develop some of the composite indicators such as “Social Acceptance” and “Social Capital”. A Fishnet of $500m^2$ was generated across Wales and all the indicators were populated into the Fishnet cells. This helped in overcoming the issues of scale and measuring units of different. Geodatabase contain key GIS layers for the study area including landuse, landform, geology, protected areas, air emissions, mortality, morbidity, multiple deprivation and unemployment.

To conclude, a rich variety of factors and indicators have been incorporated into the geodatabase developed for Wales, UK. The integration of different indicators from the socio-economic, environmental, public health and techno-economic domains into the multicriteria spatial decision analysis will increase the potential scope of application of the developed system. Critical issues of scale and resolution of the spatial data have been addressed by using a Fishnet, which combines all indicators together in one data structure.

9.5 Application

Although the SDSS has been designed and developed independent of the study area, for its application, Wales, UK has been selected as the study area. An application of the SDSS is presented in the research to investigate the potential of Coalbed Methane (CBM) and Enhanced Coalbed Methane (ECBM) development in Wales. The main consideration for this application has been to ensure that the proposed sites have minimum negative impact on the environment and public health and maximum socio-economic benefit to the communities living nearby.

At the first level of the site selection process, AHP based site selection tool has been used to identify the most suitable sites for CBM development considering the key Environmental, Public Health, Socio-Economic and Techno-Economic indicators. A sensitivity analysis has been carried out to analyse uncertainties linked with the domain weights.

At the second level of the site selection process, potential sites have been clustered and ranked based on their key indicators. For this purpose, SOM based site ranking and site ranking by neighbourhood analysis tools have been applied. Finally, the top 50 potential CBM sites in Wales have been identified after applying the spatial filters to exclude any cells interfering with key features such as protected areas and infrastructure.

A scenario of ECBM development has also been analysed with respect to the proximity of the potential CBM sites to major CO_2 emitters and gas feeder pipelines in Wales. Site impact assessment has also been performed on the selected sites to estimate the positive and negative impacts using RIAM based impact assessment tool. Also, appropriate indicators have been linked with different RIAM components to analyse the spatial variability in the impact of the selected sites.

To conclude, the application has demonstrated that the developed system can be used to facilitate the decision making process in Geoenergy and Geoenvironmental applications. The analytical modules developed for site selection, ranking and impact assessment have been applied successfully during the application process.

9.6 Overall conclusions

The overall contributions of the research work are presented below:

1. A new impact-based, multicriteria Spatial Decision Support System has been developed to address a wide spectrum of spatial decision problems in a holistic way.
2. The system utilises a combination of Multi Criteria Decision Analysis (MCDA) and Artificial Intelligence (AI) techniques for site selection, site ranking, impact assessment and spatial knowledge discovery. Verification of these analytical modules enhanced the confidence in the applicability of the developed system for spatial decision making.
3. A new site ranking technique has been introduced in this research that utilises the multivariate ordering capability of one-dimensional Self-Organizing Map (SOM). The ranks are assigned to the naturally existing clusters in the dataset after ordering. Therefore, user preference is minimised in the site ranking process which can be useful in the second level of site selection process to prioritise the potential sites.
4. The other site ranking technique developed in this research provides a new approach for comparison of the status of selected indicators in the neighbourhood of each site. The ranks are assigned to the sites based on the status of key indicators in their neighbourhood using a novel Criterion Sorting Mechanism (CSM).
5. A modified approach for the General Regression Neural Network (GRNN) has been introduced in this research to support Geographically Weighted Regression (GWR). The approach is based on introducing a spatial parameter and extending the original

GRNN algorithm to provide support for both fixed spatial kernel and spatially adaptive kernel for GWR.

6. The developed system has been applied in Wales, UK to investigate the potential of Coalbed Methane (CBM) and Enhanced Coalbed Methane (ECBM). To the author's knowledge, the work is the first of its kind, where socio-economic, environmental, public health and techno-economic aspects have been integrated together into a comprehensive decision making for site selection, site ranking and impact assessment for CBM-ECBM development.
7. A novel concept of incorporating "Social Acceptance" and "Social Capital" into the spatial decision making for engineering interventions has been introduced.
8. As a result of this research, an integrated Spatial Decision Support System has been developed to facilitate informed spatial decision making process. The generic nature of the developed system has extended the concept of SDSS to address a range of spatial decision problems, thereby enhancing the effectiveness of decision making process. It has been demonstrated that the developed system can be applied effectively to address multicriteria spatial decision problems faced in a number of Geoenergy and Geoenvironmental applications.
9. The developed system can be considered as a useful modern governance tool, incorporating the key factors into decision making and providing optimal solutions for the critical questions related to energy security and economic future of the region.

9.7 Suggestion for further research

In view of the current work, the following areas have been identified for further research:

1. A web-based SDSS can be useful for collaborative and group spatial decision making.
2. A public participatory GIS tool is recommended to be part of the web-based SDSS to incorporate local spatial knowledge into the decision making process. This tool can

also be used for reaching a consensus on the relative weights for key indicators.

3. Instead of pre-populating the geodatabase, a mechanism should be devised to acquire up-to-date information from the relevant spatial data infrastructures at the run-time.
4. Significant opportunities exist in the application of the SDSS for other Geoenergy developments such as shale gas and Underground Coal Gasification (UCG). Similarly, the system can be useful in Geoenvironmental applications such as solid waste management.
5. Further research is recommended to incorporate survey data used in the mapping of socio-economic and public-health indicators in a more concise manner. Spatial interpolation (IDW) has been used in this research to map the survey responses across the study area. Other computationally advanced techniques such as Spatial Micro Simulation can be used in future for increased reliability. Furthermore, the indicators mapped from the survey response data, should be verified in the field especially in the communities living around the most potential sites identified in this research in order to boost the confidence of the decision makers.
6. More accurate coal seam data can be added to the geodatabase to accurately calculate the gas resource confined in it. Three dimensional coal seams data can be useful in this regard.
7. Location and structural information for key infrastructure such as bridges can be incorporated in the multicriteria decision making. This will ensure the selected sites have least impact on critical structure both in day to day operations and in case of an engineering disaster.

- The soil quality data should also be incorporated in the decision making as it is an important indicator for the ecosystem, landuse and agricultural productivity.

Appendix A

INDICATORS AND OTHER DATA

A.1 INTRODUCTION

Appendix A contains the tabular information used in this research such as the indicators and survey datasets. It also presents detailed results of the application presented in Chapter 8.

A.2 UK Social Capital Measurement Framework

Table A.1 UK Social Capital Measurement Framework (Foxton and Jones 2011)

Definition	Examples of Indicators
Civic participation.	
Individual involvement in local and national affairs, and perceptions of ability to influence them.	• Perceptions of ability to influence events. • How well informed about local/national affairs. • Contact with public officials or political representatives. • Involvement with local action Groups. • Propensity to vote
Social networks and social support.	
Contact with, and support from, family and friends. These are seen as important sources of social capital. The number and types of exchanges between people within the network, and shared identities that	• Frequency of seeing/speaking to relatives/friends/neighbours. • Extent of virtual networks and frequency of contact. • Number of close friends/relatives who live nearby. • Exchange of help.

develop, can influence the amount of support an individual has, as well as giving access to other sources of help.	• Perceived control and satisfaction with life.
Social participation.	
Involvement in, and volunteering for, organised groups. Some indicators are measuring sources of social capital (e.g. those related to the personal contacts and interactions that are made by meeting people through clubs, churches, organisations, etc). Others are measuring outcomes of social capital. For instance, voluntary work is an important indicator of people's willingness to undertake activity that benefits others and the wider community.	• Number of cultural, leisure, social groups belonged to and frequency and intensity of involvement. • Volunteering, frequency and intensity of involvement. • Religious activity.
Reciprocity and trust.	
The amount of trust individuals have in others, those they know and do not know, as well was trust in formal institutions. Trust is seen as being closely linked to social capital, either as a direct part of it or as an outcome.	• Trust in other people who are like you. • Trust in other people who are not like you. • Confidence in institutions at different levels. • Doing favours and vice versa. • Perception of shared values.

Views of the local area.	
Individual perceptions of the area in which they live. This dimension is included as an aid for analysis and is not considered an aspect of social capital. Positive views of the local area are a good correlate for how happy, safe and secure people are within their environment.	• Views on physical environment. • Facilities in the area. • Enjoyment of living in the area. • Fear of crime.

A.3 Survey data used in ocial capital mapping

Table A.2 BHPS and NSW questions selected for analysis

Question	Answer Type	Total respondents	Total LSOAs represented	Potential indicators
British Household Panel Survey Wave 18				
Likes present neighbourhood	Yes 1 No 2	1402	672	Views of the local area
Attends religious services Classes modified as: 4 into 5 (Never) 5 into 4 (for weddings etc)	1 Very frequently 5 Never	1289	597	Social Participatio n

Attend local group/voluntary organisation	1 very frequent 5 Never	1289	597	Social Participation
Do unpaid voluntary work	1 very frequent 5 Never	1289	598	Social Participation

National Survey of Wales 2012-2013

Belonging to local area?	1 Strongly Agree 5 Strongly Disagree	14481	1881	Views of the local area
People like to help their neighbours?	1 Strongly Agree 5 Strongly Disagree	14414	1880	Social network & Social Support
Safety at home after dark?	1 Very Safe 4 Very Unsafe	14537	1881	Views of the local area
Safety walking in local area after dark?	1 Very Safe 4 Very Unsafe	14287	1879	Views of the local area
Trusting people in the	1 Many people in	13974	1878	Reciprocit

neighbourhood? (records with "5 just moved in the area" also removed)	the neighbourhood can be trusted 4 None of the people in the neighbourhood can be trusted			y and Trust
Safe for children to play outside?	1 Strongly Agree 5 Strongly Disagree	14334	1881	Reciprocity and Trust
People from different backgrounds get on well together (records with "6 too few people in the area" removed) (records with "7 all same background" removed)	1 Strongly Agree 5 Strongly Disagree	13368	1876	Reciprocity and Trust
People treating each other with respect and consideration	1 Strongly Agree 5 Strongly Disagree	14442	1880	Reciprocity and Trust

Table A.3 Countryside Council for Wales Protected areas and their significance (CCW 2012)

Protected sites	Description
Site of Special Scientific Interest (SSSI)	SSSIs include a wide range of habitats from small fens, bogs and riverside meadows to sand dunes, woodlands and vast tracks of uplands.
Special Areas of Conservation (SAC)	SACs are strictly protected sites designated under European law.
Special Protection Areas (SPA)	SPAs are strictly protected sites, also known as the Birds Directive, on the conservation of wild birds.
Wetlands of International Importance (Ramsar Sites)	Wetlands are hugely important areas which boast fantastic biological diversity and provide the water and means by which large numbers of plant and animal species survive.
National Nature Reserves (NNR)	The great nature reserves of Wales, stretching from the Great Orme in the north to the Mawddach Valley in the south.
Areas of Outstanding Natural Beauty (AONB)	An AONB is usually a mark of great landscape and scenic beauty. This means that an AONB is not necessarily an area of high nature conservation value, but in practice it often includes many areas which are.
Marine Nature Reserves (MNR)	The purpose of MNRs is to conserve marine life and geological features of special interest, whilst also providing opportunities for study of these marine systems.

Heritage Coasts	The "heritage coast" classification scheme was initiated in 1972 to protect coastline of special scenic and environmental value from undesirable development. The Countryside Council for Wales administers the coasts in Wales with some 42% of coast in Wales is protected under the scheme.
Biospheric Reserves	These sites are dedicated to understanding how human activity affects the biosphere. They are an international designation made by UNESCO – the United Nations Educational, Scientific and Cultural Organisation and sites are based on nominations made by more than 110 countries. Wales has one internationally recognised reserve that is dedicated to studying and understanding the way our way of life affects the land, air and water around us.
Biogenetic Reserves	These reserves highlight those areas of land and water that are of tremendous value for wildlife.
Local Nature Reserves (LNR)	Many local authorities in Wales - in urban as well as rural areas - have set up Local Nature Reserves (LNRs). There are a total of 62 throughout the country, all with natural features that are of special interest to their local area.

A.4 Top 50 CBM sites location and domain ranks

Table A.4 Top 50 CBM sites in wales after neighbourhood analysis and comparison

SITE RANKS	DOMAIN RANKS				LOCATION	
CSM	SOCIO-ECONOMIC	ENVIRONMENT	PUBLIC-HEALTH	TECHNO-ECONOMIC	X	Y
1	84	5	19	24	281750	210250
2	94	15	15	16	282250	210750
3	100	14	22	18	281750	210750
4	89	10	25	34	281250	210250
5	117	25	14	10	282750	211250
6	118	24	17	12	282250	211250
7	105	19	28	21	281250	210750
8	80	9	39	50	280750	209750
9	119	22	23	14	281750	211250
10	46	33	50	63	282250	204750
11	81	6	71	37	286250	208250
12	43	37	44	74	282250	204250
13	124	7	59	17	288250	208250
14	114	20	38	35	280750	210250

15	123	1	66	19	286750	209250
16	35	66	81	28	278250	207750
17	42	45	58	66	281750	204750
18	70	11	76	54	280250	209250
19	71	2	77	62	283250	207750
20	62	11	82	57	280250	208750
21	34	57	37	86	280750	203250
22	44	46	45	79	281750	204250
23	33	47	48	90	281250	203250
24	38	52	43	85	281250	203750
25	37	53	54	77	281250	204250
26	40	23	95	64	281250	206750
27	29	71	42	81	280250	203250
28	32	61	49	82	280750	203750
29	41	49	47	89	281750	203750
30	25	65	84	52	277250	204250
31	28	80	64	55	277750	205250
32	30	67	61	70	280750	204250
33	48	56	40	88	280750	202750

34	31	91	69	41	277250	205750
35	15	32	114	73	264250	210250
36	27	72	63	72	280250	203750
37	45	70	35	84	280250	202750
38	21	67	86	61	277750	204250
39	24	73	80	58	277750	204750
40	22	83	60	71	279750	203250
41	92	43	4	97	307250	205750
42	91	40	2	104	307250	205250
43	23	81	53	80	279750	202750
44	101	31	7	98	307750	205750
45	5	106	98	29	268750	207750
46	84	8	26	123	286750	200250
47	8	100	106	27	268750	208250
48	96	34	9	102	307750	205250
49	26	105	87	25	275750	205750
50	17	99	102	26	270750	208250

A.5 CBM-ECBM Rapid Impact Assessment Matrix

Table A.5 CBM-ECBM Rapid Impact Assessment Matrix

ID	COMPONENT	A1	A2	B1	B2	B3	E-SCORE	E-BAND
	PHYSICAL/CHEMICAL							
1	Disposal of water	2	-2	2	2	3	-28	-C
2	Contamination of surface water due to wellbore integrity	0	0	1	1	1	0	NI/NC
3	Contamination of surface water due to fracturing fluids discharge	2	-2	2	2	3	-28	-C
4	Contamination of aquifer due to fracturing fluids discharge	3	-2	2	2	3	-42	-D
5	Soil disturbance due to site	1	-2	2	2	2	-12	-B
6	Increase in air pollution due to work at site	2	-1	2	2	3	-14	-B
7	Increase in air pollution due to transportation	3	-2	2	2	1	-30	-C
8	Fugitive methane emissions	3	-2	3	3	3	-54	-D
9	Contamination of ground water due to borehole integrity	3	-2	2	2	2	-36	-D
10	Contamination of soil in the surrounding areas	1	-1	2	2	2	-6	-A
11	Lowered ground water table	2	-1	2	2	1	-10	-B
12	Minor tremors caused by the fracturing process	1	0	2	2	1	0	NI/NC
13	Methane migration in aquifers	1	-2	2	2	3	-14	-B
14	Infrastructure wear and tear	3	-1	2	2	1	-15	-B
	BIOLOGICAL / ECOLOGICAL							

15	Forest cut down for sitting	3	-2	2	2	3	-42	-D
16	Impact of noise on wildlife	1	-1	2	2	3	-7	-A
17	Night time light pollution for wildlife	1	-2	3	2	2	-14	-B
18	Effect on aquatic wildlife	2	0	1	1	1	0	NI/NC
19	Habitat fragmentation and loss	3	-1	2	3	1	-18	-B

SOCIAL / CULTURAL

20	Increase in traffic	2	-2	2	2	3	-28	-C
21	Resettlement of people from siting areas	3	0	1	1	1	0	NI/NC
22	Social acceptance	3	-2	2	2	3	-42	-D
23	Health and safety of general public due to normal operations	0	0	1	1	1	0	NI/NC
24	Health and safety on workers in case of accident	1	-1	2	2	2	-6	-A
25	Health and safety of general public in case of accident	1	-1	2	2	3	-7	-A
26	Migration workers	2	-2	2	2	3	-28	-C
27	Effect on scenic quality of the area	3	-2	2	2	3	-42	-D
28	Employment generation for surrounding communities	4	3	2	2	3	84	+E
29	Disturbance in grazing patterns	1	-2	2	2	3	-14	-B

ECONOMICS / OPERATIONAL

30	Cost of water treatment	1	-2	2	1	1	-8	-A
31	Loss of agricultural land due to site	1	-1	2	2	2	-6	-A

32	Disturbance in grazing patterns	1	-2	2	2	3	-14	-B
33	Local jobs creation	3	1	2	2	2	18	+B
34	Housing and infrastructure	2	1	3	2	2	14	+B
35	Effect on energy situation	3	3	2	2	1	45	+D
36	Amount and value of methane gas produced	3	2	2	2	1	30	+C
37	Cost of processing the produced gases	3	-1	2	2	1	-15	-B
38	Cost of transporting produced gas to be utilised	3	-1	2	2	1	-15	-B
39	Cost of treatment of CO2 and its transportation to the site	3	-2	2	2	1	-30	-C
40	Economic growth	3	1	2	2	3	21	+C

A.6 CBM-ECBM spatial variation of impacts – continuous variables

Table A.6 RIAM components impact analysis on linked indicators (Continuous variables)

COMP ID	Indicator	Impact	Wales Average	A	B	C	D	E	F
6	Air Emissions 2008	Negative	51.2	25.0	52.2	34.4	56.4	34.8	53.3
6	Indicator of Air Quality 2008	Negative	49.0	64.4	64.4	68.4	68.0	61.3	49.1
12	Distance From Fault Lines	Negative	1799.7	2123.0	1753.3	1736.1	2252.8	1193.7	990.1
12	Distance From	Negative	34870.8	61009.6	66050.1	61555.3	68329.5	69716.9	60935.2

	Geological Dykes								
19	Distance from - Special Areas of Conservation (SAC)	Negative	2884.4	6564.3	7523.6	2597.8	9592.6	5527.6	9116.7
19	Distance from - Special Protection Areas (SPA)	Negative	13953.0	27028.9	25360.0	30281.3	20470.5	14293.5	31453.5
21	Occupied Houses	Negative	14.6	25.1	14.3	19.9	8.7	27.5	90.5
23	Population	Negative	99.2	98.6	99.2	98.9	99.5	98.5	94.8
33	Mining and quarrying	Positive	0.3	3.9	2.2	4.4	1.4	0.5	0.3
33	Construction	Positive	9.4	10.1	11.0	10.6	9.5	10.2	9.7
33	Transport and storage	Positive	3.2	4.1	4.2	4.0	4.4	3.8	3.8
33	Accommodation and food service activities	Positive	6.7	4.5	4.7	4.5	4.0	3.6	4.6
33	Electricity, gas,	Positive	0.6	0.7	0.6	0.3	0.6	0.3	0.4

	steam and air conditioning supply								
33	Water supply; sewerage, waste management and remediation act	Positive	0.8	1.0	0.8	0.7	0.7	1.3	1.6
33	Professional, scientific and technical activities	Positive	4.7	3.0	2.9	2.9	3.5	3.8	2.2
38	Distance from major CO2 emitters	Positive	28466.1	12537.0	10591.0	7705.3	12460.5	15819.6	6793.8
38	Distance from gas feeder pipeline network	Positive	31440.6	9352.9	4587.4	5951.2	2895.2	3321.4	3937.0
40	WIMD Income	Positive	1281.9	821.8	643.7	613.5	1023.3	956.4	558.8
28	WIMD Employment	Positive	1290.7	494.4	528.1	398.0	814.7	840.6	320.4
35	WIMD Overall	Positive	1138.6	681.1	639.1	553.8	1077.7	840.1	440.8

A.7 CBM-ECBM spatial variation of impacts – discrete variables

Table A.7 RIAM components impact analysis on linked indicators (Discrete variables)

COMP ID	Indicator	Discrete Value	Wales %	A	B	C	D	E	F
4	HydroGeological Features	Highly productive aquifer	1.4	0.0	0.0	0.0	0.0	0.0	0.0
4	HydroGeological Features	Moderately productive aquifer	18.7	100.0	100.0	100.0	100.0	100.0	100.0
8	Gas Hazard - Methane and CO2 in superficial geology	Potential gas hazard from peat	0.6	2.2	1.5	1.5	1.5	1.5	0.0
10	Corine Land Cover 2006	Heterogeneous agricultural areas	1.3	11.7	5.1	10.9	8.8	0.0	5.1
10	Corine Land Cover 2006	Pastures	51.1	22.6	16.8	18.2	29.2	38.0	7.3
10	Corine Land Cover 2006	Permanent crops	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10	Corine Land Cover 2006	Scrub and/or herbaceous vegetation associations	20.0	22.6	15.3	10.2	25.5	43.1	47.4

13	Hydrogeological Features	Highly productive aquifer	1.4	0.0	0.0	0.0	0.0	0.0	0.0
13	Hydrogeological Features	Moderately productive aquifer	18.7	100.0	100.0	100.0	100.0	100.0	100.0
17	Night Time Light Pollution	Negligible	45.9	21.2	63.5	45.3	13.1	52.6	39.4
15	Forest Cover	Broadleaved	5.2	9.5	5.8	7.3	16.1	9.5	5.8
15	Forest Cover	Conifer	5.8	8.8	42.3	22.6	23.4	1.5	0.0
15	Forest Cover	Mixed mainly broadleaved	0.3	0.0	0.7	0.7	1.5	0.7	0.0
15	Forest Cover	Mixed mainly conifer	0.3	0.0	0.7	0.7	0.0	0.0	0.0
15	Forest Cover	Young trees	1.1	2.2	6.6	6.6	3.7	0.0	0.0
20	Scenic Quality	High	41.1	8.8	14.6	0.0	25.5	51.8	18.2
20	Scenic Quality	Outstanding	12.7	0.0	0.0	0.0	0.0	0.0	0.0
26	Sense of Place/Local Distinctiveness	Strong	44.1	1.5	0.0	2.2	0.0	58.4	35.8
31	Corine Land Cover 2006	Heterogeneous agricultural areas	1.3	11.7	5.1	10.9	8.8	0.0	5.1
31	Corine Land Cover	Pastures	51.1	22.6	16.8	18.2	29.2	38.0	7.3

	2006								
31	Corine Land Cover 2006	Permanent crops	0.0	0.0	0.0	0.0	0.0	0.0	0.0
32	Corine Land Cover 2006	Pastures	51.1	22.6	16.8	18.2	29.2	38.0	7.3

A.8 Geochemical baseline for top 50 CBM sites

Table A.8 PHC, CO2 and Methane hazard in top 50 CBM sites in wales

SITE RANKS	NATURALLY OCCURRING HAZARD		POTENTIALLY HARMFUL CHEMICALS											LOCATION	
CSM	**CO2/Methane in bedrock**	**CO2/Methane in superficial**	**Arsenic**	**Vanadium**	**Tin**	**Potassium**	**Magnesium**	**Nickel**	**Lead**	**Iron**	**Copper**	**Cadmium**	**Arsenic**	**X**	**Y**
1	Yes	No	338.19	120.49	6.82	20500.40	5081.78	79.34	67.31	50852.50	34.42	3.51	65.7138	281750	210250
2	Yes	Yes	341.17	117.07	7.02	20107.80	5088.08	69.78	74.63	50105.50	34.31	3.86	76.7296	282250	210750
3	Yes	No	333.37	117.86	6.97	20227.80	5063.13	69.33	72.93	50488.70	34.12	3.64	74.5177	281750	210750
4	Yes	No	325.92	120.68	6.99	20547.60	5062.78	72.74	67.91	50833.80	34.25	3.33	65.9936	281250	210250
5	Yes	Yes	356.09	113.89	7.19	19751.30	5152.70	65.22	81.05	49208.90	34.02	4.29	86.7814	282750	211250
6	Yes	Yes	344.77	115.12	7.08	19923.60	5101.84	65.75	79.07	49275.00	34.06	4.06	84.5875	282250	211250
7	Yes	No	317.41	119.69	6.82	20515.90	5050.22	67.20	69.09	51007.40	33.78	3.24	69.2932	281250	210750
8	Yes	No	325.39	121.83	7.43	20600.80	5055.15	71.21	67.78	51623.80	34.49	3.25	63.4769	280750	209750

9	Yes	No	332.32	116.62	6.86	20151.10	5054.34	65.87	75.72	49828.80	33.86	3.74	80.1357	281750	211250
10	Yes	Yes	294.00	121.35	6.83	21177.70	4575.70	68.96	70.00	51789.10	38.03	3.27	68.1415	282250	204750
11	Yes	No	337.33	119.07	6.60	20728.20	5220.41	69.21	67.67	47763.30	34.98	3.90	60.4982	286250	208250
12	Yes	Yes	281.54	122.44	7.00	21409.90	4669.44	68.61	65.30	50913.60	38.23	2.87	62.7084	282250	204250
13	Yes	No	352.75	116.53	5.87	20652.40	5627.60	79.49	60.78	45087.50	41.15	3.04	53.8105	288250	208250
14	Yes	No	308.42	122.90	7.09	20887.10	5080.86	67.88	65.72	50054.00	34.16	3.00	62.232	280750	210250
15	Yes	Yes	422.50	110.27	6.87	19428.90	5333.70	66.41	78.89	46918.50	34.02	5.76	68.7013	286750	209250
16	Yes	No	314.56	117.89	8.43	19999.20	4789.96	65.19	99.32	53738.00	36.00	3.39	68.2909	278250	207750
17	Yes	Yes	272.43	121.19	6.58	21396.60	4576.68	62.64	75.41	53702.30	35.79	3.06	77.23	281750	204750
18	Yes	No	331.83	122.86	8.16	20599.40	5042.95	69.01	70.13	53117.40	34.87	3.26	62.503	280250	209250
19	Yes	No	287.74	127.11	6.00	21364.10	4820.72	76.78	56.55	52489.60	37.67	2.93	52.1255	283250	207750
20	Yes	No	341.40	124.17	7.71	20636.80	5033.48	70.09	70.32	54370.90	34.77	3.34	60.5344	280250	208750

21	Yes	Yes	295.53	122.66	7.44	21686.70	4608.99	63.78	71.57	55202.30	38.99	3.47	71.1673	280750	203250
22	Yes	Yes	274.33	121.78	6.97	21458.50	4649.12	64.18	70.61	52507.90	36.93	2.97	70.1395	281750	204250
23	Yes	No	269.37	125.40	7.15	22161.80	4562.89	61.38	66.09	56509.60	38.01	3.06	66.1113	281250	203250
24	Yes	Yes	274.56	122.20	7.19	21609.00	4643.46	62.65	70.65	53586.80	37.47	3.05	70.5354	281250	203750
25	Yes	Yes	255.85	122.64	6.92	21763.80	4648.22	59.92	74.47	53357.30	36.60	2.80	76.0101	281250	204250
26	Yes	Yes	255.62	129.76	5.45	21180.60	4732.79	74.32	53.11	73619.30	34.25	2.15	50.3941	281250	206750
27	Yes	Yes	313.19	120.15	7.60	21236.80	4597.79	65.05	74.02	54643.30	39.43	3.76	75.5848	280250	203250
28	Yes	Yes	277.47	121.07	7.31	21444.80	4616.46	61.59	73.31	54161.70	37.84	3.18	75.3035	280750	203750
29	Yes	Yes	271.03	122.16	7.12	21538.50	4648.71	64.41	66.11	51966.60	37.04	2.88	64.6447	281750	203750
30	Yes	No	351.29	117.33	8.63	20261.00	4632.88	71.95	79.83	54005.30	40.67	4.35	77.8623	277250	204250
31	Yes	No	246.17	129.94	7.23	22863.30	5207.93	56.90	81.99	53683.60	38.90	2.89	89.7763	277750	205250
32	Yes	No	228.63	124.13	6.91	22177.70	4650.19	55.63	76.40	53390.50	36.87	2.46	79.615	280750	204250

33	Yes	No	311.94	125.95	7.54	22257.20	4569.49	65.15	71.58	57182.90	40.94	3.86	68.5539	280750	202750
34	Yes	No	286.12	118.17	8.54	20422.00	4733.48	63.88	86.36	53314.30	38.43	3.40	76.4172	277250	205750
35	Yes	No	223.47	116.59	6.76	19904.00	5440.34	64.23	47.96	50797.10	30.21	1.94	42.0532	264250	210250
36	Yes	No	277.71	118.01	7.42	20923.80	4503.54	59.74	73.84	55667.00	37.82	3.28	83.4652	280250	203750
37	Yes	No	347.51	122.25	7.91	21578.20	4643.60	69.46	76.66	54352.10	41.95	4.35	72.1452	280250	202750
38	Yes	No	423.90	117.37	9.20	20261.60	4569.89	78.48	83.97	55703.30	43.31	5.55	85.9593	277750	204250
39	Yes	No	311.76	125.40	7.90	21941.50	4983.89	65.10	83.47	54103.30	40.63	3.98	88.0445	277750	204750
40	Yes	No	332.23	118.77	7.53	20973.50	4564.17	66.60	73.38	54505.20	39.54	4.07	77.4171	279750	203250
41	Yes	No	260.38	128.84	9.07	20811.80	5408.84	78.35	59.67	53450.80	39.98	1.54	29.9327	307250	205750
42	Yes	No	337.54	124.22	7.08	18883.10	4863.43	89.18	48.93	58881.00	34.52	1.26	26.5541	307250	205250
43	Yes	No	358.09	120.74	7.90	21286.70	4654.40	70.82	75.11	53693.50	41.28	4.47	72.5882	279750	202750
44	Yes	No	210.51	142.28	12.21	23020.10	5706.71	86.02	58.63	50699.90	51.16	1.37	27.8354	307750	205750

45	Yes	No	374.91	113.98	6.84	20742.20	4895.30	80.38	62.50	58060.30	31.16	5.27	91.6803	268750	207750
46	Yes	Yes	188.12	118.13	6.01	22309.30	5164.84	50.37	49.96	46997.50	28.52	1.75	60.8946	286750	200250
47	Yes	Yes	305.27	120.21	7.43	20680.00	4947.52	68.81	62.74	54609.30	33.11	3.31	63.3811	268750	208250
48	Yes	No	208.85	142.42	11.09	22811.80	5719.65	85.51	54.18	50968.50	52.54	1.28	25.9893	307750	205250
49	Yes	No	384.45	106.87	11.22	18729.10	4592.81	89.39	95.98	48573.30	53.30	5.99	69.9964	275750	205750
50	Yes	No	276.33	123.85	6.54	20977.20	5053.31	65.64	55.20	53413.00	32.37	2.76	50.0492	270750	208250

References

CCW. 2012. *Countryside Council for Wales GIS datasets* [Online]. Available at: http://www.ccw.gov.uk/landscape--wildlife/protecting-our-landscape/gis-download---welcome/gis-dataset-information.aspx [Accessed: 14 October].

Foxton, F. and Jones, R. 2011. Social Capital Indicators Review. In: Statistics, O.f.N. ed. Crown copyright 2011.