Accepted Manuscript

Design and Development of a Generic Spatial Decision Support System, based on Artificial Intelligence and Multicriteria Decision Analysis

Muhammad Irfan , Aleksandra Koj , Majid Sedighi ,
Hywel Thomas

Please cite this article as: Muhammad Irfan , Aleksandra Koj , Majid Sedighi , Hywel Thomas , Design and Development of a Generic Spatial Decision Support System, based on Artificial Intelligence and Multicriteria Decision Analysis, *GeoResJ* (2017), doi: 10.1016/j.grj.2017.08.003

Title:

Design and Development of a Generic Spatial Decision Support System, based on Artificial Intelligence and Multicriteria Decision Analysis

Authors:

Muhammad Irfan [a] (First Author, Corresponding Author, Email: MuhammadI2@cf.ac.uk, Tel: 00442920876264)

Aleksandra Koj [a] (Email: KojA@cardiff.uk)

Majid Sedighi [a, b] (Email: majid.sedighi@manchester.ac.uk)

Hywel Thomas [a] (Email: ThomasHR@cardiff.ac.uk)

Affiliation:

[a] Geoenvironmental Research Centre, School of Engineering, Cardiff University, Queen's Building, The Parade, Cardiff, CF24 3AA, UK.

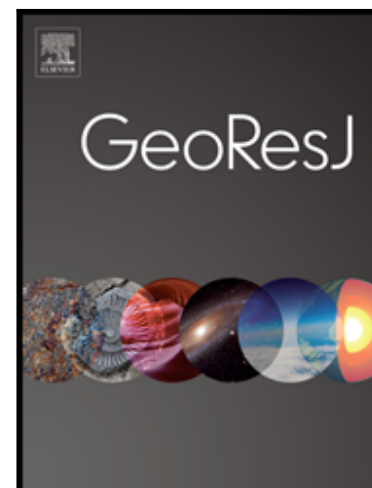[b] School of Mechanical, Aerospace and Civil Engineering, The University of Manchester, Manchester M13 9PL, UK.

Design and Development of a Generic Spatial Decision Support System, based on Artificial Intelligence and Multicriteria Decision Analysis

Abstract

A new integrated and generic Spatial Decision Support System (SDSS) is presented based on a combination of Artificial Intelligence and Multicriteria Decision Analysis techniques. The approach proposed is developed to address commonly faced spatial decision problems of site selection, site ranking, impact assessment and spatial knowledge discovery under one system. The site selection module utilises a theme-based Analytical Hierarchy Process. Two novel site ranking techniques are introduced. The first is based on a systematic neighbourhood comparison of sites with respect to key datasets (criterions). The second utilises multivariate ordering capability of one-dimensional Self-Organizing Maps. The site impact assessment module utilises a new spatially enabled Rapid Impact Assessment Matrix. A spatial variant of General Regression Neural Networks is developed for Geographically Weighted Regression (GWR) and prediction analysis. The developed system is proposed as a useful modern tool that facilitates quantitative and evidence based decision making in multicriteria decision environment. The intended users of the system are decision makers in government organisations, in particular those involved in planning and development when taking into account socio-economic, environmental and public health related issues.

1 Introduction

Decision makers increasingly rely on SDSS to address multicriteria, semi structured spatial decision problems. The concept of SDSS is mostly limited to domain specific applications [1]. However, certain spatial decision problems are common to many application areas. For example site selection, site ranking and site impact assessment problems are faced commonly in different environmental applications, public health risk assessment, land use planning, resource allocation, geoenergy initiatives and development of new facilities etc. These spatial decision problems have some common traits, i.e. they are multicriteria in nature and they 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 decision maker's preferences, priorities and judgements [2].

Although the above mentioned spatial decision problems are common to many application areas, it is hard to find a generic SDSS in literature that can readily be utilised. Sugumaran and Degroote [3] discussed the possibility of developing a generic SDSS that can be useful in many application areas. Some of the commercial and open source GIS software such as IDRISI, ArcGIS, SAGA and ILWIS provide a variety of modelling techniques and an underlying mechanism for software customisation to serve the purpose of a generic SDSS [3]. For example, ArcGIS model builder provides a mechanism to combine different geoprocessing components together. Despite these customisation features, it requires deeper understanding of the structure of different modules and/or relevant programming/scripting knowledge to

develop generic decision support tools from such existing software. Spatial Analysis and Decision Assistance (SADA) and Decision Support System for the Requalification of Contaminated Sites (DESYRE) are two freely available and frequently used decision support tools used for environmental and public health risk assessment. SADA provides a comprehensive decision support environment for site specific human health and ecological risk assessment [4]. DESYRE provides integrated management and remediation of contaminated sites, providing features for site characterization, socioeconomic constraints and risk assessment [5]. Both SADA and DESYRE provide site specific risk assessment features but lack in other commonly faced decision problems, e.g. site selection, site ranking or spatial knowledge discovery.

On the other hand, a number of other SDSS have been presented in the literature for domain specific applications related to site selection or site risk assessment. For example, Escalante et al. [6] presented an SDSS to evaluate crop residue energy potential to analyse the potential and geographic dispersion of biomass production. A set of biomass points was generated through the transport municipal network. Neighbourhood analysis was used to assign biomass potential to each study point. Fuzzy AHP and multi-criteria decision analysis has been used for the assessment of each biomass point for the selection of most suitable sites for anaerobic digestion plants [6]. 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 cities for the solar projects in Iran [7]. Zanuttigh et al. [8] 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 to compare different options subsequently [8]. Comino et al. [9] presented a 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 assess 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 environmental quality classes considered [9]. Gorsevski et al. [10] 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 Weighted Linear Combination (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 [10]. Fayetteville shale gas SDSS has been developed to analyse and assess the impacts of water consumption for hydraulic fracturing [11]. The system is used by the regulatory agencies and producers, to study the potential impacts on the environmental flow components of the river. Fayetteville

2

shale gas SDSS utilises the Soil and Water Assessment Tool (SWAT) as its underlying modelling unit to analyse changes in hydrological patterns in the study area as a result of Shale gas exploration [11]. Ruiz et al. [12] 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 [12].

The review presented above suggests that although a number of SDSS have been developed specifically for site selection or site risk assessment, but they do not offer a holistic decision support environment, i.e. they are limited to a specific decision problem or they are domain/application specific. It is hard to find a generic system that can tackle these frequently occurring spatial decision problems in one system, is not domain specific and is not limited to a given study area. Therefore there is a need for an integrated and holistic approach. This can be achieved by designing and developing a generic SDSS with an adequate model base to assist the decision makers tackling multicriteria decision problems in different domains. Furthermore, it is envisaged that using open source Geoinformatics technologies and a modular development approach can ensure the easy adoption and further extension of the capabilities of the system.

This paper presents the design, development and verification of such an integrated and generic SDSS based on a number of Artificial Intelligence (AI) and Multicriteria Decision Analysis (MCDA) techniques. The system can be applied in a variety of applications in environmental, socioeconomic, geotechnical and public health domains. Analytical modules can be used independently or in combination with each other as per the application requirement. The system also provides features for spatial knowledge discovery and geo-visual analytics to gain evidence based information for a given geographical region. The intended users of the system are decision makers in local and national government organisations, consultants and researchers.

2   SDSS architecture design

Architecture of the developed SDSS consists of three main components: I) Geodatabase, II) Model base and (c) Graphical User Interfaces (GUI) based on the SDSS architecture presented by Malczewski [2]. Geodatabase is used for spatial data management, Model Base provides analytical capability and GUI are utilised in decision making process by the user. The system design is independent of the study area and the underlying spatial data in Geodatabase. Therefore, the system can be applied independent of geographical location, subject to the availability of the data. Because of the modular design, any new analytical modules can be added to the model base without any architectural changes.

McIntosh et al. [13] identified key challenges and made recommendations in Environmental Decision Support Systems (EDSS) development and its successful adoption to help facilitate the achievement of desirable social and environmental outcomes. One of the main challenges exists in relation to ensuring EDSS longevity and financial

sustainability. A recommendation to overcome this challenge was to focus on EDSS development that is relatively easy and inexpensive to use and update. This can be achieved by employing open source software technology which enables easy model expansion and reusability to reduce development costs [13]. The .Net based open source spatial library DotSpatial has been used for the development of current SDSS in order to read, manipulate and visualise spatial data [14]. The analytical modules in the Model Base have been developed using Microsoft .NET C# programming language and DotSpatial library. The GUIs were developed using .Net Windows Forms. In order to cover the most commonly faced spatial decision problems, several analytical modules have been developed using MCDA and Artificial Neural Networks (ANN) techniques. As shown in Fig. 1, the analytical modules designed and implemented in the system are divided into three main categories in accordance to their functional similarity and include:

1. Site selection and ranking: The analytical modules developed are:

   a. Analytical Hierarchy Process (AHP) based site selection tool.

   b. Self-Organizing Maps (SOM) based site ranking tool.

   c. Site ranking by neighbourhood analysis tool.

2. Impact assessment and prediction: The analytical modules developed are:

   a. Rapid Impact Assessment Matrix (RIAM) based site impact assessment tool.

   b. General Regression Neural Network (GRNN) based regression and prediction tool.

3. Spatial knowledge discovery: The analytical modules developed are:

   a. SOM based correlation finding tool.
   b. Parallel Coordinate Plots (PCP) based geo-visual analytics tool.

Technical detail of these analytical modules is covered in Section 4 and verification is provided in Section 5. A schematic diagram of the system architecture design of the SDSS is shown in Fig. 1.
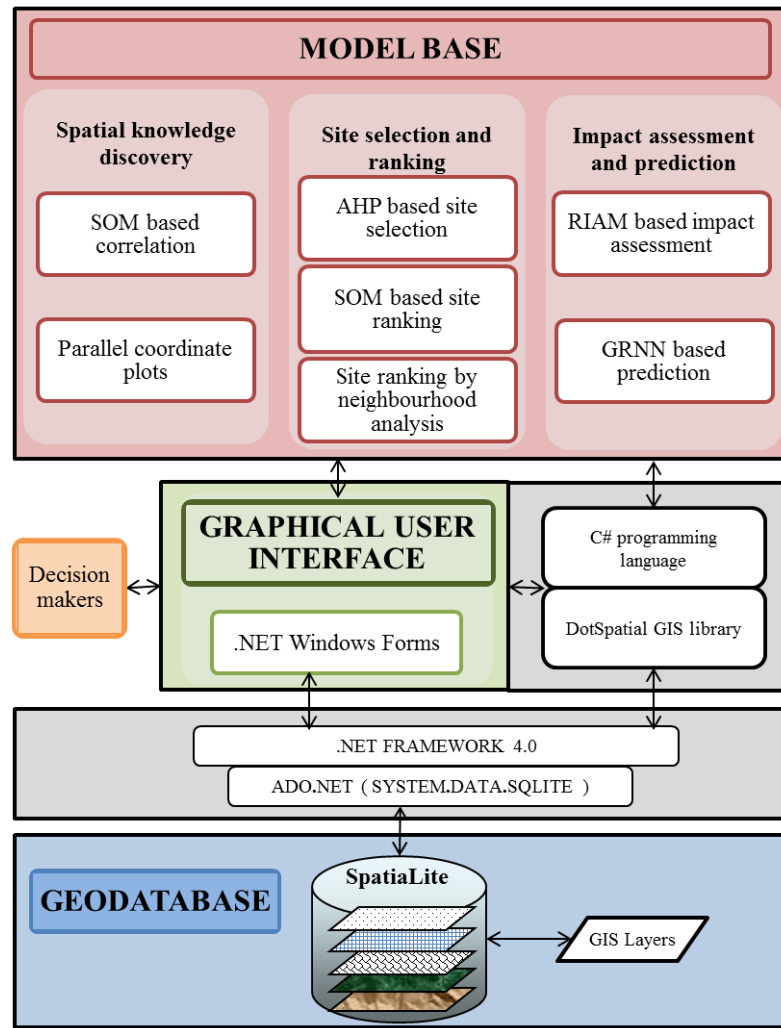
4

Fig. 1 - High level system architecture of developed SDSS

For geodatabase development, open source SpatiaLite technology has been used. SpatiaLite is an extension library of the popular SQLite database to support geometrical storage and geoprocessing operations [15]. SpatiaLite was selected because it is a lightweight single file-based geodatabase that can be distributed along with the software without any need to install database servers. To provide connectivity between the .Net based application and SpatiaLite geodatabase, the System.Data.SqLite has been used as the Active Data Object for .Net (ADO.NET) provider.

3   Geodatabase management

As mentioned earlier, SpatiaLite geodatabase has been used to store data. Functions have been provided through main GUI of the SDSS to load GIS layers from geodatabase. The developed SDSS is independent of the study area, but in order to verify the analytical modules and demonstrate the applicability of the system, the geodatabase has been populated with a number of GIS layers for Wales, UK. These demo applications are provided in Section 5 and cover the spatial decision problems of site selection, site ranking and impact assessment.

These GIS layers cover different aspects of four domains namely: I) Socio-Economic, II) Environmental, III) Public Health and IV) Techno-Economic. Some of the key GIS layers include physical environment, protected areas,

5

demography, index of multiple deprivation, mortality rates, hospital admission rates, social capital, labour market, topography, geology and hydrogeology. Techno-economic domain contains GIS layers to support applications related to facility siting, renewable energy and unconventional gas exploration etc. Details of data collection and GIS analysis performed in the development of these GIS layers is out of the scope of this paper, but can be found in [16].

These GIS layers differ to each other in format (raster, vector), scale and units. For meaningful analysis it is important to bring them into same scale and units. In order to overcome this issue, a Fishnet grid (a 2D mesh of squared cells) of 500×500m cell size has been created, covering the entire onshore area of Wales, UK. All GIS layers are then joined together with this Fishnet data structure using different geoprocessing functions in SpatiaLite. If the system is to be applied for smaller areas, e.g. county level, then a smaller cell sized Fishnet can be used to provide detailed mapping. Other tables are also created in the geodatabase to store the key information related to the parent-child relationship of different layers. This information is used in the analytical modules, e.g. in the AHP based site selection module. Analytical modules that require many input parameters from users, are also provided with functions to save these parameters in the geodatabase as a theme. For example the GIS layers used in site selection process and their relative weights can be saved and loaded as a theme for future analysis. This is a useful feature that can help in group decision making while comparing results of different combinations of parameters.

4   SDSS Development

The main interface of the system was developed first that provides basic features, e.g. add/remove GIS layers, visualise, zoom, pan and legend settings etc. The analytical modules were then developed and added to the Model Base of the system. As explained in Section 2, the analytical modules developed are divided into three categories: I) Site selection and ranking, II) Impact assessment and III) Spatial knowledge discovery. The detail of these analytical modules is given below.

4.1   Site selection and ranking

The site selection module is based on the AHP technique which combines hard and soft information together to identify potential sites (Fishnet cells). At this stage, another spatial decision problem is faced by the decision makers, i.e. to rank and prioritise these potential sites. The potential sites can be ranked based on the key criterion's values of each site and in its surrounding neighbourhood. This approach can also reduce the potential risks associated with personal judgment and choice of the decision maker [17]. To achieve this, two site ranking techniques have been developed as explained below.

4.1.1   AHP based site selection tool

6

The AHP based site selection tool was designed to be used as a first step in the site selection process as it delineates the potential areas from the entire study area and filters out any unsuitable areas. User selects the relevant socio-economic, environmental, public health and techno-economic GIS layers (criterions) and assigns relative weights to them for a given site selection problem. The AHP based site selection tool provides all the required functions for AHP analysis including, data commensuration, relative weight assignment, sensitivity analysis and storing the user preferences in the geodatabase as a theme. These features are discussed in details below.

If the relative importance of all criterions is known then the user can directly assign the relative weights to them. The sum of all the weights must be equal to 1. However, this is not practical when a large number of criterions are used in the analysis. In such cases, the tool allows to assign the relative weights using the Pairwise Comparison Method [2]. Using this approach, the relative importance of only two criterions is compared at a time. The tool then calculates the relative weights and checks its consistency using the same methodology as provided in [2].

In order to bring all criterions into same units, a data commensuration tool is also developed. It scales each criterion between 0-1 using either i) Maximum Score Procedure or ii) Score Range Procedure [2]. After scaling, the value for each location (Fishnet cell) is multiplied with its relative weight, a process called Weighted Linear Combination [2]:

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

(1)

where A_i is the suitability index for the ith location (Fishnet cell), w_j is the relative weight assigned to the *jth* criterion and x_{ij} is the value of the *jth* criterion at *ith* location. It is noted that the sum of w_j is always equal to 1 (or 100%, if used as percentage). In AHP, the criterions are arranged in a tree hierarchy, where the top most node represents the overall goal of the analysis. This goal is achieved by carrying out objectives and sub-objectives that form the AHP hierarchy tree. The AHP based site selection process is essentially the application of Weighted Linear Combination at each level of the criterion hierarchy, starting from the bottom most criterions and traversing all the way up to the parent node. The parent node in this case is the overall site selection process formed by the four domains and their criterions in a tree structure.

The main GUI of the AHP based site selection tool is shown in Fig. 2. Criterion selected by the user and their relative weights can be saved as theme in the database using the save them button and already stored themes can be loaded. AHP process can be applied at any level of the hierarchy and results can be visualised as a map. Tool also allows applying filters on criterion to restrict the processing to a certain geographical area. For example, a filter can be set on the geological layer to restrict the AHP based site selection process within the suitable geological formations only. This feature can also be used to filter out any critical environmental or strategic areas from the site selection process right at

the beginning. The tool provides features for data commensuration and weight assignment using this interface and associated popup dialogues.
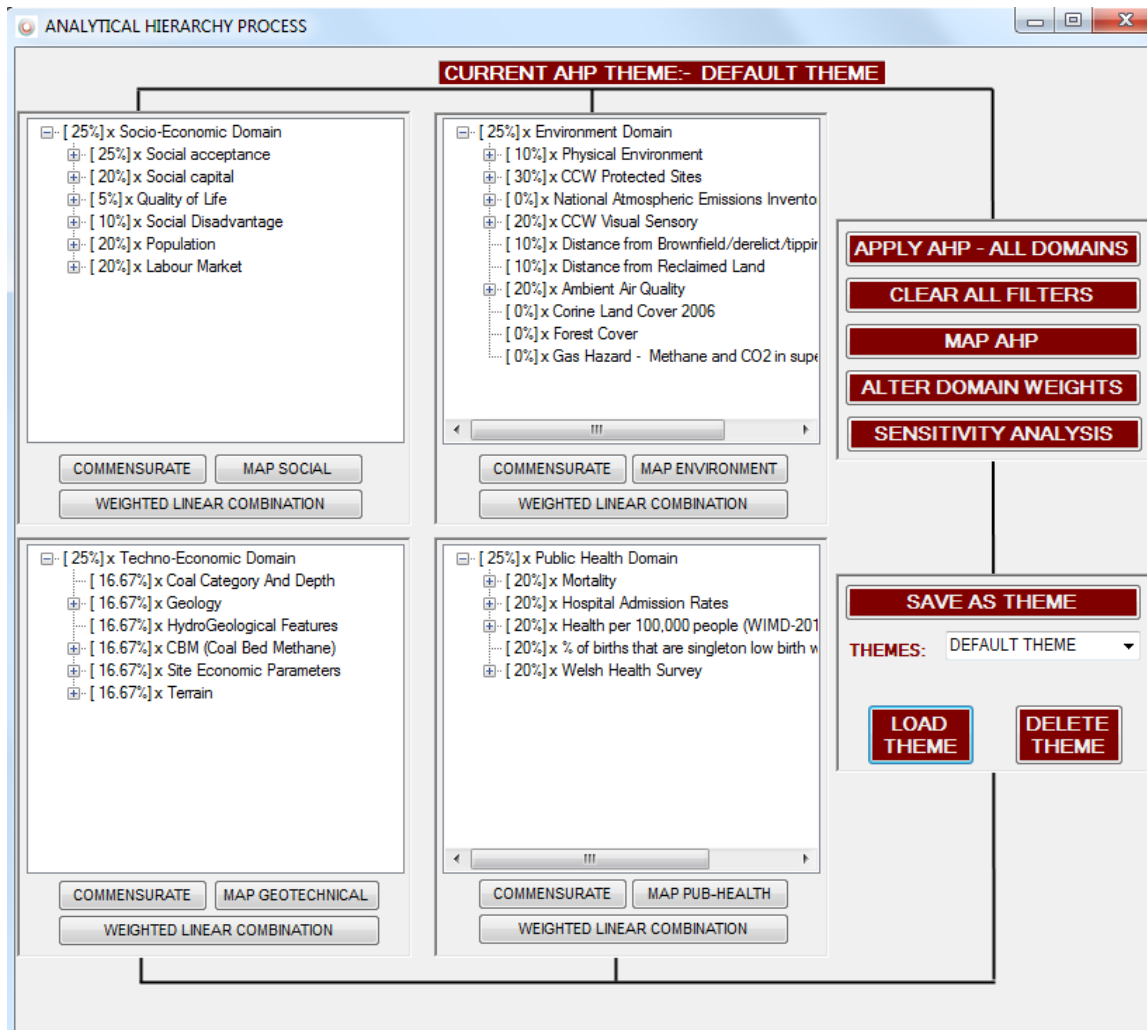

Fig. 2 GUI of the AHP based site selection tool loaded with selected GIS layers for Wales (UK)

AHP analysis result is sensitive to the relative weights assigned to criterions [18,19]. In order to assess this sensitivity of the relative weights, a sensitivity analysis tool has been developed. The GUI of the sensitivity analysis tool is shown in Fig. 3.

User can check the sensitivity of AHP results at any level of the AHP hierarchy with respects to the weights assigned to the criterions that contribute in the analysis for that level. For example if the user selects the top most level, its sensitivity is checked against the weights assigned to the four domains that construct it. User selects a criterion whose weight has to be changed. A new series (line) is added to the resultant graph for each selected criterion as shown in Fig. 3. In this particular example, the sensitivity analysis has been performed on the criterion "social acceptance" from Socio-Economic domain. This criterion is constructed by four sub-criterions as listed in Fig. 3.
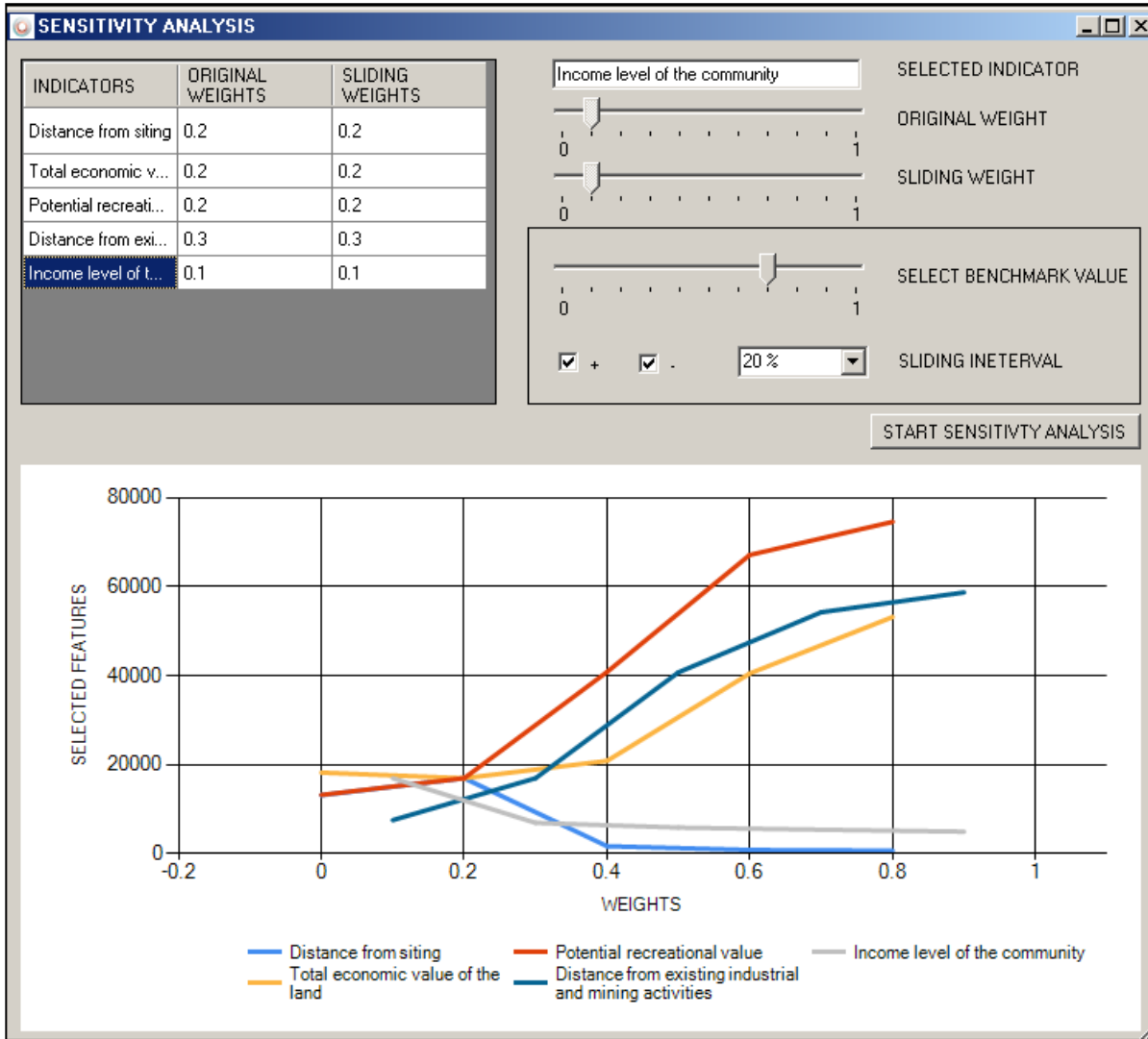
Fig. 3 GUI of the sensitivity analysis tool showing the effect of change in relative weights of different GIS layers on the number of selected features (fishnet cells) meeting the benchmark criteria

User also selects a benchmarking value between 0 and 1 using the benchmarking value slider on the sensitivity analysis tool. This value is used to count the number of sites (Fishnet Cells) that have values greater than the set benchmark using a given weighting scheme. User also selects a sliding interval which is used to increase or decrease the weight of the selected criterion. This difference of weight is equally distributed among other criterions to maintain the sum of weights equal to 1 (or 100%) all the time. Visual inspection of the resultant graphs provides an understanding of the sensitivity of AHP analysis with respect to the relative weights assigned to contributing criterions. This helps decision makers in choosing an appropriate set of relative weights for the criterions used in a given AHP analysis.

4.1.2 SOM based site ranking tool

A SOM is a type of ANN that uses a neighbourhood function to preserve the relationship in multidimensional input space into a low dimension output space called output map. The output map is usually a one or two-dimensional map.

SOM is used for visualizing high dimensional data as low-dimensional space. The training of SOM is unsupervised therefore it is very easy to use [20].

The one-dimensional SOM has the capability of clustering and ordering (sorting) multidimensional data in ascending or descending order [20,21]. Based on this capability, a novel site ranking tool has been developed in the SDSS. The SOM based site ranking tool groups the potential sites into clusters based on the values of criterion associated with the sites, e.g. socio-economic and environmental criterions. User selects these criterions on the basis of which site ranking is performed. Data (all criterions) is first scaled between 0 and 1 and then loaded into a one-dimensional SOM. User also provides the number of neurons (or number of clusters) in the output one-dimensional map. Once the unsupervised training is performed and the output map is converged to represent the multi-dimensional input space, an ordered rank is assigned to each cluster based on its position in the output map. Fig. 4 illustrates a one-dimensional SOM with 5 neurons in the output map, before and after self-organisation (convergence) over a multidimensional input data.
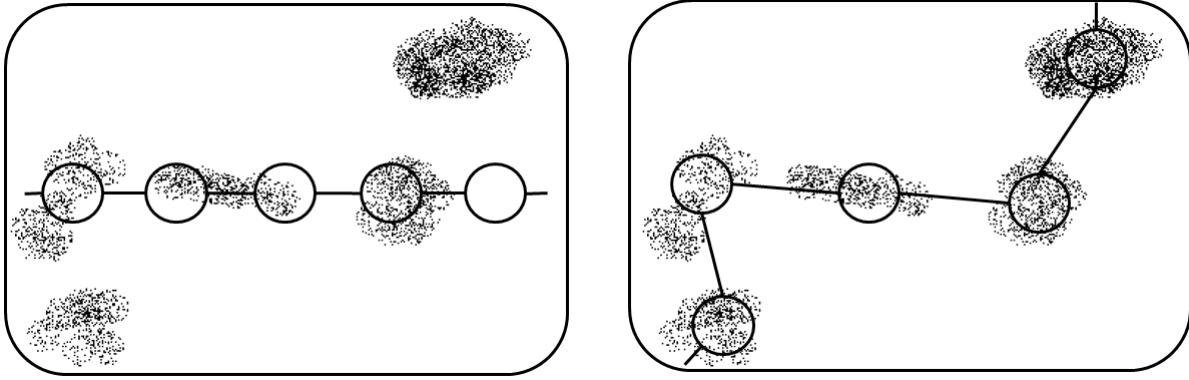

Fig. 4 One-dimensional SOM before and after convergence over the input data

Convergence of out map as shown in Fig. 4 is achieved as a result of an unsupervised competitive learning using a neighbourhood function. To start the learning process, the system first generates random criterion values for each neuron (models) in the SOM network. Number of criterions in the models is same as that of the input data. An input data item X is then exposed to the SOM network to obtain the Best Matching Unit (BMU). The BMU is the one with shortest Euclidean distance between the two in terms of their criterion values. This distance is calculated using [20]:

$$C = \arg min_i \{\|X - m_i\|\} \tag{2}$$

where C is the index of the BMU, X is the input data item and m_i is the model at ith index in the output map. Self-organization is an iterative process. Every time a model vector is identified as BMU, its neighbouring model vectors are moved closer to it in the output map [20]. The new location of a model vector in the output map at step interval (t+1) is calculated using its previous value at step interval (t), its difference from the input vector and a neighbourhood function, given as [20].

10

$$m_i(t + 1) = m_i(t) + hc_i(t)\left[x(t) - m_i(t)\right] \tag{3}$$

where x is the input data item introduced to the SOM network at time interval t, m_i is the model at ith index and $hc_i(t)$ is the neighbourhood function given as [20]:

$$hc_i(t) = \alpha(t).\exp\left(\frac{\|r_c - r_i\|^2}{2\sigma^2(t)}\right) \tag{4}$$

where, the terms $\alpha(t)$ and $\sigma(t)$ are both monotonically-decreasing functions of time. The term t is called the learning rate factor and has a value between 0 and 1. The term $\sigma(t)$ defines the kernel size and it decreases with time. The value of hc_i tends to become zero when time tends to become infinity. The term $r_c - r_i$ defines the distance between neighbouring model and the BMU [20].

The self-learning process continues and in each iteration the models in the output map adjust their position, to represent the multidimensional input data. This movement is significant in the beginning, but slows down with time. Only slight adjustments are made after a number of iterations. The learning process is stopped when these adjustments are not significant anymore. Each input data is represented by its corresponding BMU in the output map. A model in the output map can be a BMU for more than one input data items that are similar to each other in the multi-dimensional input space. Eventually each BMU represents a cluster of candidate sites that are similar to each other with respect to the criterions used for clustering and ranking purpose.

The number of neurons in SOM output map is important for a good convergence. In order to check this convergence, the tool calculates two error terms: a) Quantisation error and b) Topographic error using the same methodology defined by Kohonen [20]. The Quantization error is the average distance of all the input nodes from their respective BMU in the output map after convergence. The Topographic error is calculated by identifying the best two BMUs for each input data vector and then by checking if these two are also placed next to each other in the output map [20]. Based on these two error terms, user can increase or decrease the number of neurons in the output map to achieve better convergence. Once a satisfactory level of convergence is achieved, the tool applies ordered rank to each candidate site. This rank is based on the position of BMU of the site, in one-dimensional output map after convergence. The tool also assigns a gradient colour scheme to the resultant map showing higher ranks with darker colours for better visualisation.

4.1.3   Site ranking by neighbourhood analysis and comparison tool

This tool carries out a systematic comparison of criterions in the given neighbourhoods of potential sites for the purpose of site ranking. The candidate sites are ranked according to the status of key criterions in their neighbourhood. This neighbourhood is defined by the decision makers as a buffer radius which is calculated in map's distance units, e.g. meters. Decision maker also provides the criterions and the potential sites as two GIS layers. After scaling the criterions to same units, the tool calculates the minimum, the maximum and the average values for each criterion in the given

11

neighbourhood of each site. The final step is to assign ranks to the sites using one of these; the average, the maximum or the minimum values. The tool provides two options to rank the sites: I) Criterion Sorting Mechanism (CSM) [17] and II) Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) [22].

CSM is a novel site ranking method developed in this SDSS. In CSM, a separate rank is assigned to each site based on its value for every criterion used in the analysis. For example, a site with best status of an environmental criterion in its neighbourhood is assigned a Rank-1. Whereas a site with worst status of the same criterion gets a Rank-N, where N is total number of potential sites used in the analysis. Each site may obtain different ranks for different criterions. In the end a Rank-Sum is constructed for each site by summing up individual ranks for the criterions. Sites are sorted in ascending order in terms of this Rank-Sum. The site with the lowest Rank-Sum gets the overall Rank-1. Similarly the site with highest Rank-Sum gets the overall Rank-N, where Rank-1 site is the most suitable out of N potential sites [17].

In order to verify the results and increase the confidence in site ranking process, an existing site ranking method, TOPSIS is also implemented within this tool. TOPSIS is a commonly used site ranking method in MCDA problems. It ranks the sites based on their distances from the most ideal and the least ideal solution. The empirical formulations used in this tool are same as provided in [22].

4.2   Impact assessment

The impact assessment section contains tools that can be used for the assessment of environmental and social impacts of the potential site. Impact assessment is usually a qualitative procedure. However, two tools have been developed in the system that provides the capability to carry out semi-quantitative impact assessment. The two approaches adopted are: I) RIAM based impact assessment tool and II) GRNN based regression analysis tool. Further details of these methods are provided below.

4.2.1   RIAM based impact assessment tool

Impact assessment is a process of assessing environmental and socio-economic consequences of a program, a project or a development. Impact assessment is a qualitative process that is based on expert's judgement or community's sentiments. However, RIAM is a semi-quantitative method for carrying out impact assessment in the form of a structured matrix that contains the subjective judgements of the impact assessors [23]. The impact components are divided into four major categories, i.e. I) Physical/Chemical, II) Biological/Ecological, III) Social/Cultural and IV) Economics/Operational. The individual Environmental Assessment Score (ES) for each component is calculated by evaluating them against the following two criteria [23]:

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

12

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

The total score for Group A and B can be calculated using [23]:

$$aT = (a1) \times (a2) \tag{5}$$

$$bT = (b1) + (b2) + (b3) \tag{6}$$

$$ES = (aT) \times (bT) \tag{7}$$

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**. ES is the overall assessment score. The corresponding values for *a1, a2, b1, b2* and *b3* are provided by Pastakia and Jensen [23] in the form of a table.

A novel spatial variant of RIAM has been developed in this research to assist decision makers in checking the sensitivity of key environmental and socioeconomic parameters likely to be effected by RIAM impact components. For this purpose, decision maker can link the key environmental and socioeconomic criterions to the relevant RIAM components and assign a buffer distance to identify the neighbouring region around the site. For quantitative criterions, the tool calculates the minimum, maximum and average values of the linked criterions within the site's neighbourhood. This value is then compared to the minimum, maximum and average values of the same criterion in the entire study area (e.g. national average). However, if a qualitative criterion is linked to the RIAM component, then the tool calculates the percentage of the site's neighbourhood covered by the given discrete class of the qualitative criterion, e.g. the Conifer class of the forest layer.

The spatial variant of the RIAM technique can be useful in analysing how a proposed site can have negative or positive impact on the surrounding neighbourhood based on the existing status of the key criterions there. For example, if an area has higher air pollution than the national average, an addition of a coal fire power plant will then only deteriorate the air quality in the neighbourhood. This approach is evidence based as it records the existing status of key environmental, socio-economic or public health criterions in a quantified manner. Therefore, using this approach has advantage over the traditional qualitative judgement based impact assessment process.

4.2.2 GRNN based regression analysis tool

A novel GWR analysis tool has been developed in this research with a modified version of the GRNN. The GRNN tool facilitates the local variations in the regression analysis and helps decision makers in prediction and regression analysis.

GRNN is a one pass neural network, highly parallel in structure and belongs to the category of Probabilistic Neural Network [24]. It predicts the values at an unknown location on the basis of its proximity to a known location in terms of

the selected independent variables. A GRNN has been selected because it is useful when the relationship between dependent and independent variables is unknown and complex. In addition, due to its simple structure, it would be easier to incorporate the spatial parameters as one of the independent variables to support the local variation in the regression analysis. The output function of the GRNN is defined as [24]:

$$\acute{Y} = \frac{\sum_{i=1}^{n} Y^i \ exp^{(-D_i^2/2\sigma^2)}}{\sum_{i=1}^{n} \ exp^{(-D_i^2/2\sigma^2)}} \tag{8}$$

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 σ is the spread factor that defines the influence of neighbouring locations in the calculation of $\acute{Y}$. A small sigma value will result in a localised regression, whereas a very large value will result in average of the entire dataset. D_i is a scalar term that shows the Euclidean distance between the prediction point and the training sample in terms of all the independent variables (dimensions) and is defined as [24]:

$$D_i^2 = \left(X - X^i\right)^T \left(X - X^i\right) \tag{9}$$

where X and X^i represents the independent variables at known and unknown locations respectively.

The size of neighbourhood is important for the model to fit properly. Many iterations and comparison of standardised error can help in the selection of an appropriate model. However, if it is not clear what type and size of kernel is to be used, the GGRNN tool also provide a mechanism that uses spatial distance between geographical features as one of the independent variables for the prediction of the dependent variable. The neighbouring areas of the prediction location will influence more in the calculation of the dependent variable.

The training of GRNN is unsupervised and only requires the σ parameter from the user. GUI of the GGRNN tool is shown in Fig. 5. User provides can provide a sigma value for each independent variable. User can also use scaled, normalised or original values of the variables in the process. If scaled or normalised values are used, then a same σ value can be used for every independent variable. It is important to use an appropriate value of the σ parameter for a given analysis. Therefore, the tool also provides the "Holdout Method" [24] for the identification of most appropriate σ value. In "Holdout Method" a known location is held out of the dataset at a time and the value of dependent variable is calculated from using the rest of the data and a given σ value. User provides an upper and lower limit for the σ parameter and a step interval as shown in Fig. 5. The tool then applies "Holdout Method" and plots the Root Mean Square Error (RMSE) against the σ parameters within the user defined range.
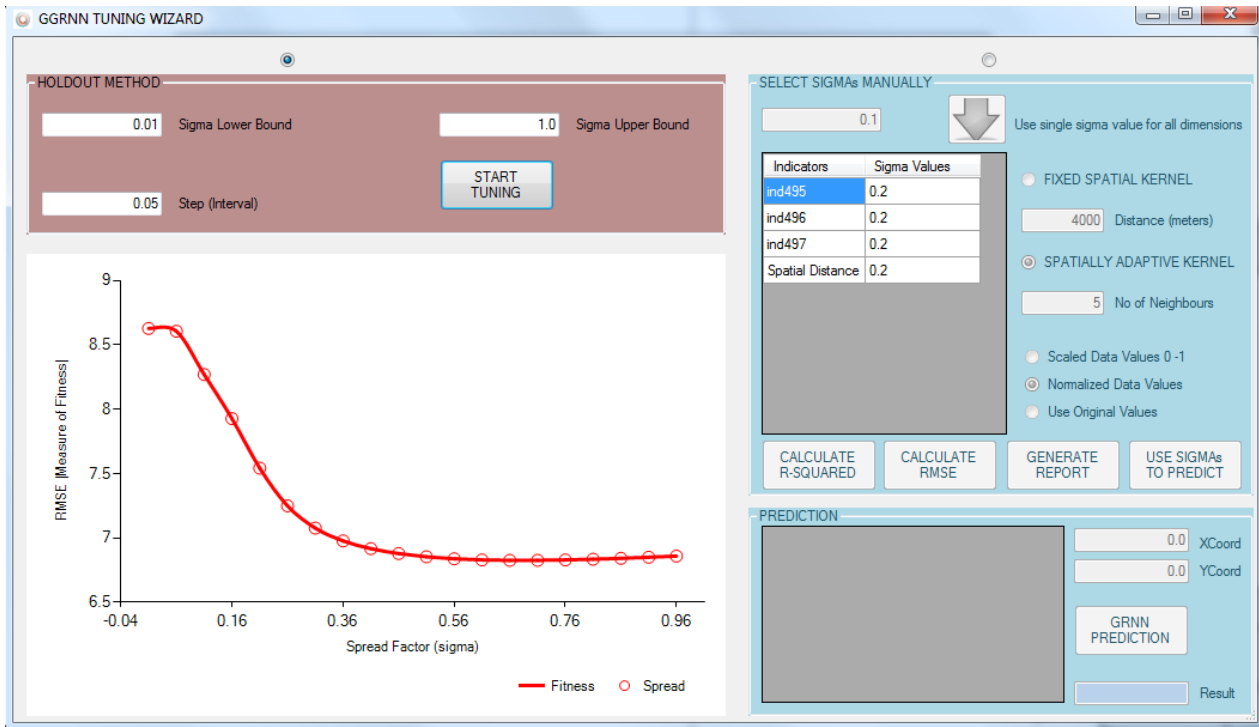
14

Fig. 5 GUI of the GRNN tool showing sigma parameter optimisation using Holdout Method with arbitrary data

GGRNN presented in this study, extends this basic GRNN calculation to allow semi-parametric or mixed GWR models presented by Fotheringham et al. [25]. This is achieved by incorporating the local and global independent variables and by analysing the local variation in the relationship between different parameters [26].

Influence of local and global variables are computed separately and then summed up together. The influence of global independent variables is calculated from the entire study area. The influence of local independent variables is calculated only within the given neighbourhood. In order to define the neighbourhood two different techniques are used: I) Fixed spatial kernel and II) spatially adaptive kernel as shown in Fig. 5. For fixed spatial kernel a spatial distance (e.g. 10km) is used to define the neighbourhood. For spatially adaptive kernel, a fixed number of neighbouring geographical features are included in the neighbourhood. Since the geometries of the geographical features, e.g. district boundaries are asymmetrical therefore resulting in a varying spatial kernel.

The GGRNN tool can be used to carry out Geographical Weighted Regression (GWR) analysis and also for the prediction of dependent variable at unknown location with the help of dependent and independent variables at the known location in the neighbourhood of prediction point.

4.3 Spatial knowledge discovery

The spatial knowledge discovery tools are provided to facilitate knowledge extraction from the data. These tools can be used to obtain an understanding of the relationship between environmental, socio-economic and public health parameter

15

in a given area. Tools developed in this section are: I) SOM based clean correlation finding tool and II) Parallel coordinate plotting tool.

4.3.1 SOM based clean correlation finding tool

SOM preserves the correlations found among input variables in the form of geometrical connections in low dimensional output map [20]. The SOM based clean correlation finding tool generates a matrix of clean correlation found among the criterions. Using the clean correlation matrix, the number of criterions in the analysis can be reduced by selecting only those that are mutually independent and have strong correlations with the dependent criterion.

After convergence, the correlation among input variables can be directly calculated from the BMUs in the output map using [27].

$$CC_{jk} = \frac{1}{\sigma_j \sigma_k} \sum_{l=1}^{M} (m_{lj} - \mu_j) * (m_{lk} - \mu_k) \tag{10}$$

where CC_{jk} 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. This also ensures that the presence of noise in the data has less effect on the correlation finding as compared to the original data. This is because of the noise resistance capabilities of SOM as suggested in [28].

4.3.2 Parallel Coordinate Plots based geo-visual analytics tool

The Parallel Coordinate Plots (PCP) is an effective exploratory analysis and data visualisation technique [29]. The PCP is an effective way of visualising two of more variables together. PCP can be used for: i) visualising how different variables are correlated to each other, ii) visualising how different variables are clustered together in a given geographical space and iii) identifying the peculiar values of the variables different from normal patterns [29].

The PCP tool in SDSS provides a function for "Brushing". Brushing is a technique used to highlight a certain part of the data to make it more prominent than the rest of the data [29]. It is useful if decision makers are interested in exploring the relationship between different variables in a given geographical region in comparison to the entire study area.
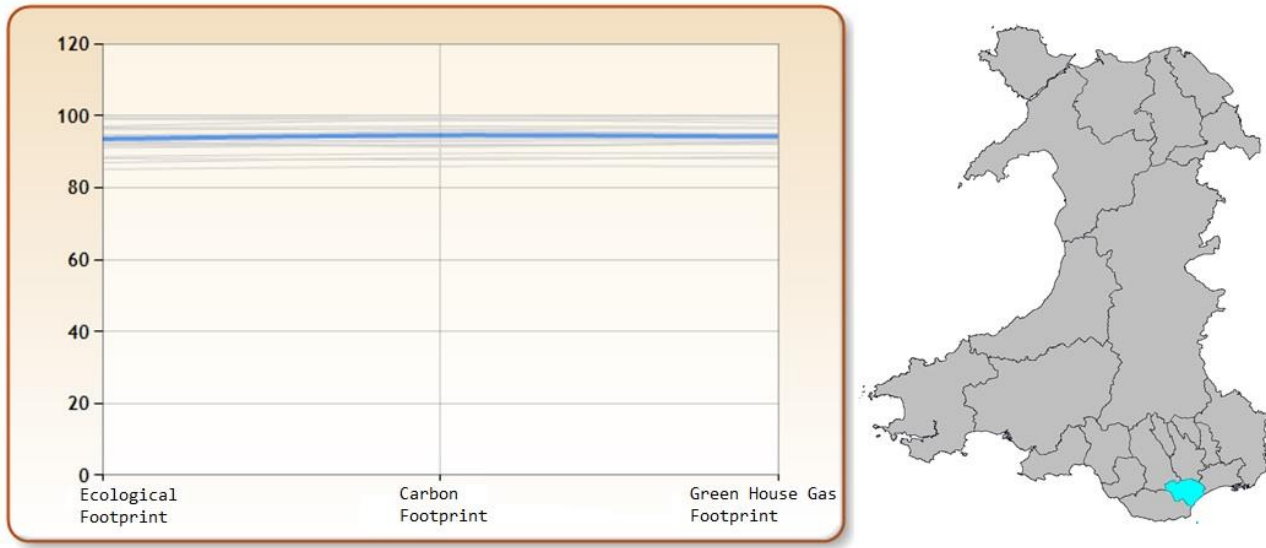
Fig. 6 PCP results with brushing of selected features on map. X-axis shows the variables and Y-axis shows the scaled values. Darker line shows the data for selected feature on map

An example of the PCP tool is shown Fig. 6 to demonstrate how multi-attribute data can be visualised. Attributes of selected features in the map are shown in darker colour in PCP to distinguish them from the rest. Plot shows the relative position of Cardiff Council, Wales, UK, among others with respect to Ecological, Carbon and Green House Gas footprints. Data has been scaled between 0 and 100 for all the variables.

5 SDSS application and verification

As described in section 4, some of the analytical modules developed in this SDSS are either new techniques or present a considerable developments or new variant based on the existing methods. These analytical modules can be used independently or in combination with each other as per the application requirement. Demo applications of these analytical modules are provided in this section and results are verified with the help of alternative reliable software, e.g. Matlab and ArcGIS. For this purpose the geodatabase is populated with a number of environmental, socio-economic and public health criterion maps for Wales, UK, details of which can be found in [17].

5.1 Application of the SOM based site ranking tool

An application of the SOM based site selection tool is provided here to rank the geographic units of Lower Super Output Areas (LSOA) in Wales, UK in terms of Welsh Index of Multiple Deprivation [30]. WIMD is the official measure of relative deprivation for small areas in Wales. Data consists of seven individual deprivation ranks and a cumulative rank showing the multiple deprivations of 1896 LSOA regions in Wales. Using the SOM based site ranking tool, data has been grouped together into 20 clusters and ranked from 1-20 based on WIMD. These clusters and associated ranks assigned by the SOM based site ranking tool are shown in Fig. 7.
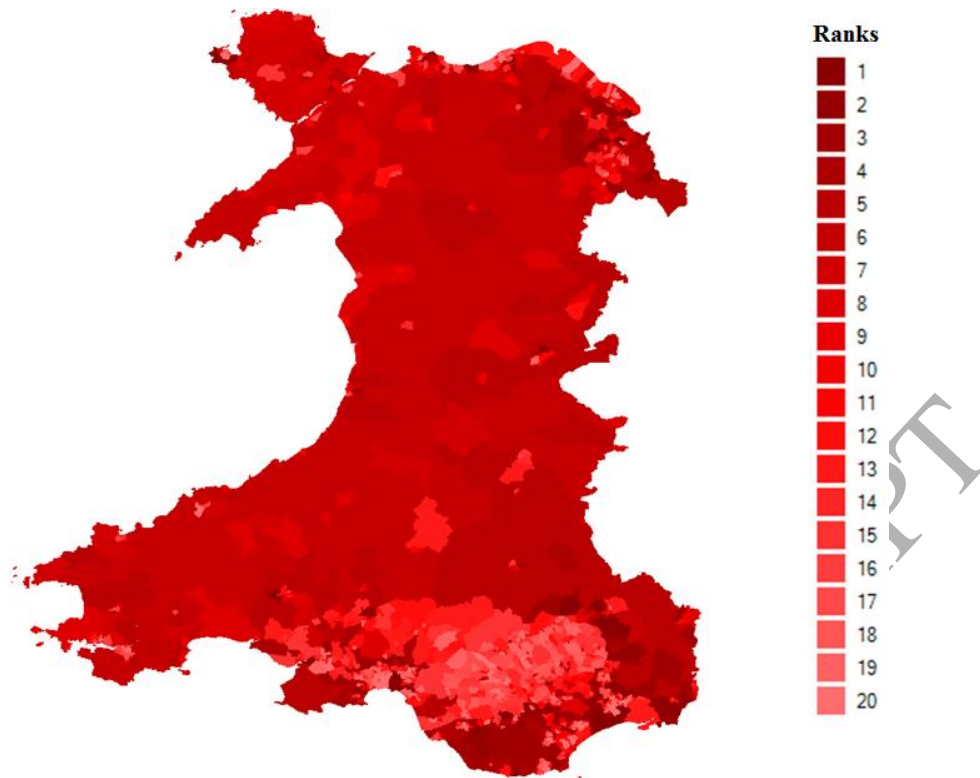
17

Fig. 7 Ranks generated by SOM based site ranking tool, LSOA regions in Wales are ranked from 1-20

In order to verify the site ranking carried out using SOM based tool, its results are compared with Matlab based GeoSOM toolbox [31]. GeoSOM has been used for verification for two reasons: I) GeoSOM allows the processing of one-dimensional SOM and II) GeoSOM can read the GIS data formats, e.g. a Shapefile. Although the GeoSOM toolbox does not explicitly provide the site ranking, it can be used to compare the ordering of clusters in the output map after convergence. A hexagonal one-dimensional output map has been selected in the GeoSOM toolbox where the BMUs are represented by the nodes at each crest and trough as shown in Fig. 8.

For comparison of ranks, one LSOA is randomly selected from each of the 20 clusters. The position of its BMU is then compared in the self-organized output maps generated by the two tools. In Fig. 8, the position of the BMUs designated by GeoSOM tool for each of the 20 LSOAs is encircled. The ordered position of the BMU is compared to the rank assigned by the SOM based site ranking tool. It is noted that for some LSOAs the position of its represented BMU is slightly different in the two cases but this difference never exceeds 1 and the order is still retained. This depicts that the given LSOA is represented by the immediate neighbouring BMUs in the two output maps. This behaviour of the SOM is expected, since the convergence of the neural network is achieved slightly differently every time even using the same.

18

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

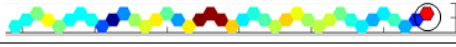**Fig. 8** Comparison of ranks generated by the SOM based site ranking tool and ordered position of BMUs in the hexagonal SOM generated by GeoSOM toolbox

The verification of the SOM based site ranking tool demonstrates that the tool can be used with confidence in multicriteria site selection and ranking problems. Since the site ranking is based on the naturally found clusters in the data, therefore the risk of user's personal judgement and choice can be reduced to minimum.

5.2 Application of the site ranking by neighbourhood analysis tool

As explained in Section 4, a new technique, i.e. the Criterion Sorting Mechanism (CSM) has been introduced in the Site ranking by neighbourhood analysis tool. Tool can be used to rank any number of candidate sites but for demonstration, an application of ranking 3 candidate sites is used here. The problem involves ranking three land parcels A, B and C with respect to environmental and economic criteria as discussed in [2]. The Economic objective has only one criterion, i.e. the Price. The Environment objective is dependent on two criterions, i.e. Slope and Views. The values of these criterions for each parcel are given in Table 1. A fishnet (500×500) was generated over the parcels and attribute information was assigned to each cell. Another Shapefile was created that contains three candidate sites, one in each parcel.

Table 1 Values of Price, Slope and Views criterion for the three candidate sites - Adopted from the site suitability problem discussed by Malczewski [2]

Candidate sites (land parcels)	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

Maximum score procedure was used to commensurate the data. All three criterions used are "Cost" criterions, i.e. the less the better, therefore the Site with rank of 1 is considered as the best site. A buffer of 3km was applied around the sites for site neighbourhood analysis. Average value of each criterion in the defined neighbourhood was used for the ranking purpose. For verification of the CSM technique, TOPSIS method was used to rank the three sites. The tool generated two types of ranks for each site using the CSM method: I) Site ranks based on individual criterions and II) Cumulative site ranks based on all criterions. Site A has the lowest value for "Views" criterion hence it is assigned rank-1. Site B has the lowest value for the "Price" criterion and therefore it is assigned rank-1 for this criterion. Site C has the lowest value for the "Slope" and therefore it is assigned rank 1 for this criterion. Fig. 9 shows that the overall site ranks produced by TOPSIS and CSM methods are the same, which verifies the accuracy of the CSM method.
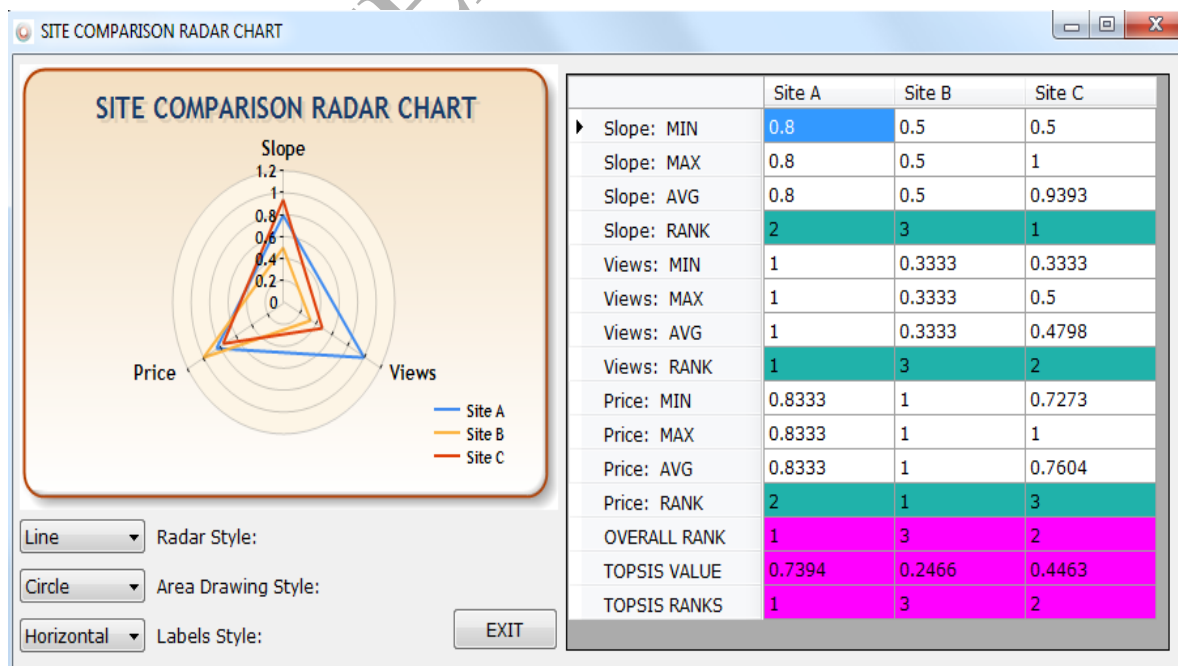

Fig. 9 Comparison of site ranks generated by TOPSIS and CSM methods for the three candidate sites

5.3 Application of the GRNN based regression analysis tool

As explained in Section 5, a modification to the original GRNN algorithm has been implemented in this tool for GWR analysis. A complete verification of the tool for semi-parametric GWR model has been presented by Irfan et al. [26]. In order to further examine the accuracy of prediction capability of the GRNN tool, a demo application is provided here.

For this purpose, the rate of cancer incidence per 100,000 of population in Wales (from 2000-2009) at the LSOA level is used as the dependent variable. Socio-economic criterions such as housing standards and household income are used as independent variables. Percentage of dwellings under different council tax bands was used for housing standard. For household income, CACI's PayCheck gross household income data [32] is used. Apart from these independent variables, the spatial distance between different LSOAs has been also used as one of the independent variables. For prediction analysis, one LSOA was randomly selected out of the dataset. A prediction was then made at this point using the two codes and the error was calculated. Data has been scaled between 0 and 1 and the same σ is used for each variable. Different values for σ were tested in both codes, i.e. 0.1, 0.2, 0.5 and 1.0.

The prediction capability of the tool is compared to the NewGRNN tool in Matlab. As shown in table 2, the values of Y parameter (dependent variable) is calculated using the GRNN tool without using spatial parameters, with a fixed kernel size covering the entire study area and then finally with a spatially adaptive kernel size of 20 neighbouring regions. The two codes have produced very similar results using the same value of σ. The slight difference can be due to the fractional changes in rounding off errors, or the way the distance in calculated in the two codes.

Table 2 Verification results of GRNN based regression analysis tool using Matlab

Sigma parameter value	Y	Ý-Matlab	Ý-GRNN prediction tool		
			Without spatial parameter	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 average value of the dependent variable (Rate of cancer incidence) is 570.5 in the entire study area. As the value of σ is increased, the prediction becomes close to the average value. This is an expected trend of the GRNN as described earlier. Very small σ can also result in over fitting which can be avoided by using the Holdout Method for the entire dataset and finding the most suitable Sigma parameters with least RMSE values.

6 Conclusions

A new generic and integrated SDSS has been developed and presented, which is capable of addressing commonly faced multicriteria spatial decision problems, such as site selection, site ranking, site impact assessment and spatial knowledge discovery. It facilitates the decision making process in a quantitative and evidence based approach. Due to the generic nature of the system, it can be applied in a wide range of applications involving these spatial decision problems. The developed system can serve as a useful modern decision assisting tool, incorporating key environmental, public health, socio-economic and technical factors into decision making and providing optimal solutions for critical questions.

The SDSS has been developed using DotSpatial and SpatiaLite open source technologies, in Microsoft .NET C# programming language. A modular design has been adopted, which allows the capabilities of the system to be extended without structural changes. The system utilises a number of AI and MCDA techniques to provide an adequate model-base for spatial decision making. A new AHP based site selection module is developed that also facilitates group decision making by allowing decision makers inputs to be stored in the database as themes. A number of functions are provided in this module for data commensuration, filtration, criterion's weight selection, sensitivity analysis and visualisation of results. Two new site ranking techniques have been presented which minimise the user's personal choice and preferences in the overall site selection process. The first technique utilises the multivariate ordering capability of a one-dimensional SOM. The second technique ranks the sites based on a systematic comparison of key criterions in the neighbourhood of candidate sites. Using SOM and neighbourhood comparison tools together enables the sites to be ranked and prioritised based on evidence. The potential decision maker's choice and judgment is therefore minimised.

A spatial variant of the RIAM technique has been developed which links spatial data with RIAM impact components. This facilitates an evidence based comparison of potential sites by analysing the existing situation of key criterions in their surroundings. A modified approach for the GRNN has been introduced to support GWR. For spatial knowledge discovery, a SOM based clean correlation finding tool has been developed. Similarly a Parallel coordinates Plot (PC) based tool has been developed for exploratory data analysis and geo-visual analytics. These tools can be used independently or in any combination, as required by the application.

Demo applications of different analytical modules are presented and verification is carried out using alternative reliable software such as Matlab. For this purpose geodatabase has been populated with environmental, public health and, socio-economic dataset for Wales, UK. Verification results provides further confidence on the accuracy of the analytical modules developed based on the results achieved.

Acknowledgements

REFERENCES

1. Karnatak HC, Saran S, Bhatia K, Roy PS (2007) Multicriteria spatial decision analysis in web GIS environment. Geoinformatica 11 (4):407-429

2. Malczewski J (1999) GIS and Multicriteria Decision Analysis. Wiley,

3. Sugumaran R, Degroote J (2010) Spatial Decision Support Systems: Principles and Practices. CRC Press,

4. Stewart RN, Purucker ST (2011) An environmental decision support system for spatial assessment and selective remediation. Environmental Modelling & Software 26 (6):751-760

5. Carlon C, Critto A, Ramieri E, Marcomini A (2007) DESYRE: Decision support system for the rehabilitation of contaminated megasites. Integrated environmental assessment and management 3 (2):211-222

6. Escalante H, Castro L, Gauthier-Maradei P, De La Vega RR (2016) Spatial decision support system to evaluate crop residue energy potential by anaerobic digestion. Bioresource Technology 219:80-90

7. Vafaeipour M, Zolfani SH, Varzandeh MHM, Derakhti A, Eshkalag MK (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 86:653-663

8. Zanuttigh B, Simcic D, Bagli S, Bozzeda F, Pietrantoni L, Zagonari F, Hoggart S, Nicholls RJ (2014) THESEUS decision support system for coastal risk management. Coastal Engineering 87:218-239

9. Comino E, Bottero M, Pomarico S, Rosso M (2014) Exploring the environmental value of ecosystem services for a river basin through a spatial multicriteria analysis. Land use policy 36:381-395

10. Gorsevski PV, Cathcart SC, Mirzaei G, Jamali MM, Ye X, Gomezdelcampo E (2013) A group-based spatial decision support system for wind farm site selection in Northwest Ohio. Energy Policy 55:374-385

11. Cothren J, Thoma G, DiLuzio M, Limp F (2013) Integration of Water Resource Models with Fayetteville Shale Decision Support and Information System. University Of Arkansas System,

12. Ruiz M, Romero E, Pérez M, 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 22:320-333

13. McIntosh BS, Ascough J, Twery M, Chew J, Elmahdi A, Haase D, Harou JJ, Hepting D, Cuddy S, Jakeman AJ (2011) Environmental decision support systems (EDSS) development–challenges and best practices. Environmental Modelling & Software 26 (12):1389-1402

14. DotSpatial Library. http://dotspatial.codeplex.com. Accessed 16 March 2017

15. SpatiaLite. https://www.gaia-gis.it/fossil/libspatialite/index. Accessed 16 March 2017

16. Irfan M (2014) An integrated, multicriteria, spatial decision support system, incorporating environmental, social and public health perspectives, for use in geoenergy and geoenvironmental applications. Cardiff University,

17. Irfan M, Koj A, Sedighi M, Thomas HR GIS Based Site Ranking using Neighbourhood Analysis and Comparison. In: Geoprocessing2014, Barcelona, Spain, March 23 - 27, 2014 2014. IARIA, pp 51-57

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

19. Nefeslioglu HA, Sezer EA, Gokceoglu C, Ayas Z (2013) A modified analytical hierarchy process (M-AHP) approach for decision support systems in natural hazard assessments. Computers & Geosciences 59:1-8

20. Kohonen T (2001) Self-Organizing Maps. Springer Berlin Heidelberg,

21. Kohonen T (1990) The self-organizing map. Proceedings of the IEEE 78 (9):1464-1480

22. Huang J-J, Yoon K (2011) Multiple attribute decision making: methods and applications. Chapman and Hall/CRC,

23. Pastakia CM, Jensen A (1998) The rapid impact assessment matrix (RIAM) for EIA. Environmental Impact Assessment Review 18 (5):461-482

24. Specht DF (1991) A general regression neural network. IEEE transactions on neural networks 2 (6):568-576

25. Fotheringham AS, Brunsdon C, Charlton M (2003) Geographically weighted regression: the analysis of spatially varying relationships. John Wiley & Sons,

26. Irfan M, Koj A, Thomas H, Sedighi M Geographical General Regression Neural Network (GGRNN) tool for geographically weighted regression analysis. In: Geoprocessing2016, Venice, Italy, April 24 - 28, 2016 2016. IARIA, pp 154-159

27. Corona F, Lendasse A (2005) Input Selection and Regression using the SOM. Helsinki University of Technology, Finland

28. Vesanto J, Ahola J Hunting for correlations in data using the self-organizing map. In: International ICSC Congress on Computational Intelligence Methods and Applications: CIMA'99, Rochester, N.Y., USA, 1999 1999. ICSC, pp 279-285

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

30. WIMD. 2011. Welsh Index of Multiple Deprivation. Welsh Government. http://gov.wales/statistics-and-research/welsh-index-multiple-deprivation/?lang=en. Accessed 17 February 2015

31. Bação F, Lobo V, Painho M (2005) The self-organizing map, the Geo-SOM, and relevant variants for geosciences. Computers & Geosciences 31 (2):155-163

32. CACI-PayCheck household income. http://www.caci.co.uk/products/product/paycheck. Accessed 16 March 2017

33. GRC SEREN Project. http://grc.engineering.cf.ac.uk/research/seren/. Accessed 16 March 2017