

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <https://orca.cardiff.ac.uk/id/eprint/133134/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Arts, Koen, Macleod, Christopher J. A., Ioris, Antonio A. R. , Han, Xiwu, Sripada, Somayajulu, Braga, João F., Maffey, Georgina, Jekjantuk, Nophadol, Zeng, Cheng and van der Wal, René 2019. Towards more effective online environmental information provision through tailored natural language generation: profiles of Scottish river user groups and an evaluative online experiment. *Science of the Total Environment* 673 , pp. 643-655. 10.1016/j.scitotenv.2019.03.440

Publishers page: <https://doi.org/10.1016/j.scitotenv.2019.03.440>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See <http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.





# Towards more effective online environmental information provision through tailored Natural Language Generation: Profiles of Scottish river user groups and an evaluative online experiment

Koen Arts <sup>a,c,\*</sup>, Christopher J.A. Macleod <sup>b</sup>, Antonio A.R. Ioris <sup>c,d</sup>, Xiwu Han <sup>c</sup>, Somayajulu Sripada <sup>c</sup>, João F. Braga <sup>c</sup>, Georgina Maffey <sup>c</sup>, Nophadol Jekjantuk <sup>c</sup>, Cheng Zeng <sup>c</sup>, René van der Wal <sup>c</sup>

<sup>a</sup> Wageningen University and Research, Post box 47, 6700 AA Wageningen, the Netherlands

<sup>b</sup> The James Hutton Institute, Craigiebuckler, Aberdeen AB15 8QH, Scotland, United Kingdom

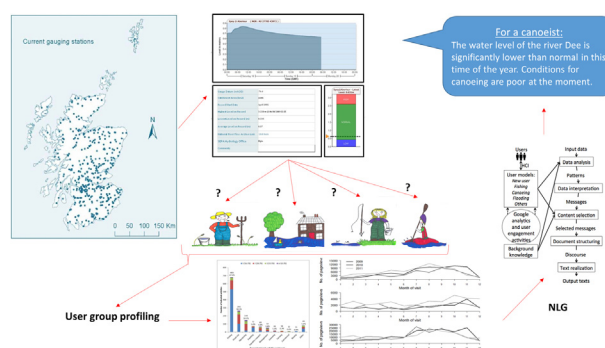
<sup>c</sup> The University of Aberdeen, King's College, Aberdeen AB24 3FX, Scotland, United Kingdom

<sup>d</sup> Cardiff University, School of Geography and Planning, Cardiff CF10 3WA, Wales, United Kingdom

## HIGHLIGHTS

- Online provision of dynamic information by authorities may lack focus.
- Profiling of user groups is an important step in improved information provision.
- Natural Language Generation (NLG) can aid user group specific information provision.
- Through NLG, key interpretation can be offered to dynamic information.

## GRAPHICAL ABSTRACT



## ARTICLE INFO

### Article history:

Received 21 November 2018

Received in revised form 26 March 2019

Accepted 27 March 2019

Available online 28 March 2019

Editor: Sergi Sabater

### Keywords:

Environmental communication

Environmental regulator

Fishing

NLG (Natural Language Generation)

Public authorities

River level

## ABSTRACT

As a result of societal transformations, political governance shifts, and advances in ICT, online information has become crucial in efforts by public authorities to make citizens better stewards of the environment. Yet, their environmental information provision may not always be attuned to end users' rationales, behaviours and appreciations. This study revolves around dynamic river level information provided by an environmental regulator – updated once a day or more, and collected by a sensor network of 333 gauging stations along 232 Scottish rivers. Employing an elaborate mixed methods approach with qualitative and quantitative elements, we examined if profiling of web page user groups and the subsequent employment of a specially designed Natural Language Generation (NLG) system could foster more effective online information provision. We identified profiles for the three main user groups: fishing, flood risk related, and paddling. The existence of well-distinguishable rationales and characteristics was in itself an argument for profiling; the same river level information was used in entirely different ways by the three groups. We subsequently constructed an advanced online experiment that implemented NLG based on live river level data. We found that textual information can be of much value in translating dynamic technical information into straightforward messages for the specific purposes of the user groups. We conclude that tailored NLG could be widely used in more effective online

\* Corresponding author at: Wageningen University and Research, Post box 47, 6700 AA Wageningen, the Netherlands.

E-mail address: [koen.arts@wur.nl](mailto:koen.arts@wur.nl) (K. Arts).

environmental information provision, and we provide five practical recommendations for public authorities and other information providers.

© 2019 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Public authorities collecting data on the environment have an increased obligation to offer online information access to relevant audiences. This is in line with a global promotion to open public data (Mathur, 2009; Shadbolt et al., 2012) and with broader ‘political modernisation’ aspirations to replace ‘command and control’ regulation with ‘command and covenant’ stewardship (Arts and Leroy, 2006). The latter implies new societal roles for public authorities, with environmental information becoming a vehicle to generate citizen engagement with, and co-governance of, the natural environment (Bäckstrand, 2003; Fleischhauer et al., 2012).

Informational governance (Mol, 2008) examines new forms of governing through information, and transformative changes in governance institutions due to new information flows (Soma et al., 2016b). Scholars interested in informational governance have called for research into the relationships between environmental information and its use by government bodies and wider society (Soma et al., 2016a). Environmental information is increasingly playing a key role in the governance of natural resources, e.g. Arctic marine governance (Lamers et al., 2016), and government portals designed to provide environmental information and associated services can be seen as tools to exchange information between government bodies and other social actors (Sandoval-Almazan and Gil-Garcia, 2012). The importance of tailoring environmental information on such portals to meet the needs of users, is likewise increasingly recognised (see e.g. Christel et al., 2018 for the importance of differential provision of climate information to different sectors).

Opportunities for improved information provision are strongly influenced by rapid developments in Information and Communications Technology (ICT), which have been shaping many domains of contemporary societies (Castells, 2010) including that of natural resource management (Arts et al., 2015b; Conde-Clemente et al., 2017a; Mol, 2008). The rapid increase in accessible digital technologies and data science tools is transforming understanding and management of key natural resources. For example, cloud-based tools for geospatial analysis and earth observation data are now freely available (Gorelick et al., 2017). Social media platforms are increasingly the default source of information during natural catastrophes, including flood events, thus improving two-way communication (Kryvasheyev et al., 2016).

Content determination is an important step in deciding what information is to be communicated in a generated text or visual message. Indeed, content determination is one of the first stages in a Natural Language Generation (NLG) system, i.e. software developed to produce texts in human language from computer-based input data (Reiter and Dale, 2000). NLG texts are nowadays used for all manner of communication purposes ranging from textual weather forecasts to deforestation reports (Ramos-Soto et al., 2013). Increasingly, automatic linguistic descriptions are linked to big data, allowing for the communication of dynamic phenomena (Conde-Clemente et al., 2017c; Siddharthan et al., 2019).

Data science requires harnessing statistical, computational and human components (Blei and Smyth, 2017). In addition to covering specific events, social media can be combined with river level information for national-scale assessment (Barker and Macleod, 2019). The management of water bodies provides a rich setting to study more effective information provision. Water managers seem relatively quick in piloting and implementing Information and Communication Technologies (ICT) (Hannah et al., 2011; Mackay et al., 2015; Montanari et al.,

2013). This may be because water management relates to various vital societal concerns including drinking water supply, climate change mitigation, and flood risk control. Developments in sensor networks and other geospatial cyberinfrastructures have been transforming data collection, and thus feed into novel possibilities for information provision and communication (Campbell et al., 2013; De Longueville Bertrand, 2010). The use of web applications, such as digital observatories (Mackay et al., 2015; Vitolo et al., 2015), has been proposed to aid land and water management. More recently, the potential of information sharing platforms for connective action in rural Africa has been demonstrated (Cieslik et al., 2018).

Despite the opportunities and improvements provided by novel ICT and the open data movement (Janssen et al., 2012), there remain many factors that result in ineffective or even absent communication or information provision by public authorities (Kamal, 2006; Loroño-Leturiondo et al., 2019). These factors include liability issues regarding consistency and quality of the provided information (Arts et al., 2015a), and conceptual barriers related to diverging understandings of how ICTs should be used (Arts et al., 2016). Moreover, a review of the spread of ICT to enable public participation in urban water governance found that although these tools enable many people to be better informed, they provided few opportunities for discussion and deliberation (Mukhtarov et al., 2018).

Rivers can be understood as common pool resources, and their users are usually highly heterogeneous and express a plurality of values and interests. This plurality results in different levels of demand with respect to data, specific data format, and level of detail (Paul et al., 2018). In addition, interpretation issues related to discrepancies between science-based expert and layperson understandings may be present, and prior to that, lack of knowledge or vision on who would use the data and for what purposes (Arts et al., 2013; Hertzum, 1999).

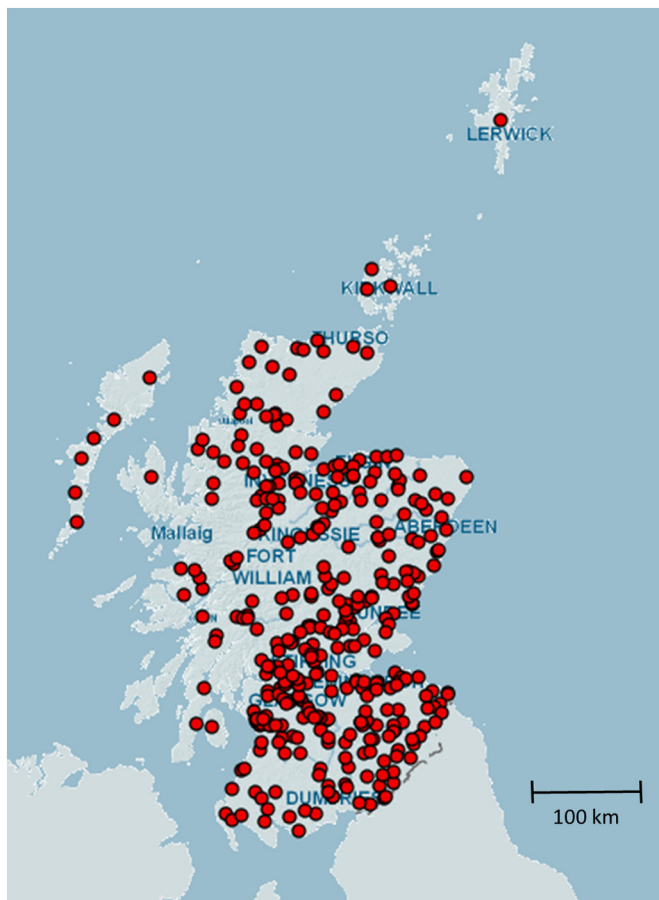
In this paper we ask if profiling of web page user groups (phase 1) and the subsequent employment of a specially designed NLG system (phase 2), could be steps towards more effective online information provision. We ask these research questions in the context of a regulator's web pages of a national network of river level sensors, and employed an elaborate mixed methods approach. We thus present an interdisciplinary study – in between environmental, social and computing science – that brings together several spheres of the total environment, including the hydrosphere (rivers) and anthroposphere (different user groups).

## 2. Materials and methods

### 2.1. Materials

We describe a study of the profiling of user groups as input for a NLG knowledge base (Reiter and Dale, 2000). Our focus is on the users of river level webpages developed and hosted by an environmental regulator in the United Kingdom. The Scottish Environment Protection Agency (SEPA) is an executive non-departmental public body of the Scottish Government, and the main public authority on environmental regulation in Scotland (Ioris, 2008). The river level web pages (<http://apps.sepa.org.uk/waterlevels/> – hereafter ‘the webpages’, see Fig. 1) are one of the most visited parts of SEPA's entire website (ascertained through Google Analytics, see Section 2.2.1) and represent a flagship of the organisation's digital information supply (Arts et al., 2015a; Macleod et al., 2012). The webpages provide dynamic river level





**Fig. 1.** Map of Scotland with online river level gauging stations.  
(Source: SEPA website - <http://apps.sepa.org.uk/waterlevels/default.aspx?sm=t>, accessed 3-3-2019.)

information (updated once a day or more often) collected by a sensor network of (in 2019) 359 online gauging stations, along 232 rivers in 107 catchments across Scotland (cf. Black and Cranston, 1995) (Fig. 1).

While SEPA is not legally required to communicate river level information to the general public, SEPA has an open data plan since 2016, and is statutorily obliged to provide flood warning service to citizens. Moreover, as Arts et al. (2016) show, in the context of new water regulation and attempts to foster public engagement, effective communication of river level information is all the more important. Similar platforms by environmental authorities can be found elsewhere, such as in Australia, <http://www.bom.gov.au/australia/flood/?ref=ftr>, England ([www.environment-agency.gov.uk](http://www.environment-agency.gov.uk)), the Netherlands (<https://waterinfo.rws.nl/#!/kaart/waterhoogte-t-o-v-nap/>), Norway (<https://www.nve.no/hydrology/>), Spain ([www.saihebro.com](http://www.saihebro.com)) and the USA (<https://water.weather.gov/ahps/index.php>). In the Spanish context, hydrological data has been used to create NLG news stories for general users (Molina, 2012; Molina and Flores, 2012). Contrary to our approach, this was based on data journalism, focusing on ‘newsworthy’ changes in time series (see also Novák, 2016).

## 2.2. Methods for Phase 1

A mixed-methods approach was employed, using qualitative and quantitative elements. The research comprised two stages with the following aims:

- Phase 1: User group profiling – To identify the webpage user groups, to find out how the key user groups engage with and value the webpages, how the webpages are used in relation to other

information sources, and how they feed into decision-making in relation to river activities or interests;

- Phase 2: NLG experiment – To ascertain if supporting textual information (through NLG) for specific user groups aids towards more effective information provision. We selected the ‘fishing’ user group as our focus for the experiment given the high number of users and therefore a potential high number of experiment participants.

Five methods fed into the profiling (phase 1) part of our longitudinal study that commenced in 2012.

### 2.2.1. Google Analytics

This functionality was enabled by SEPA for its webpages in 2007. Traffic to river level pages was initially reported from 2008; however it was not until 2009 that a substantial number of stations (104) became digitally available. In 2012, Google modified how it calculated page visits; thus, to be able to compare traffic volume across multiple years we constrained our analyses and comparisons to traffic data from the three-year period 2009–2011. The web analytics data were collected using R package *ganalytics* v0.1 for R v3.0, which uses the Google Analytics Application Programming Interface (API) protocols. In total, the webpages received 3,449,954 visits, generating 13,538,626 ‘page views’ between 2009 and 2011. In 2011, more than 25% of all visits to SEPA’s entire website concerned the river level webpages.

### 2.2.2. Online survey

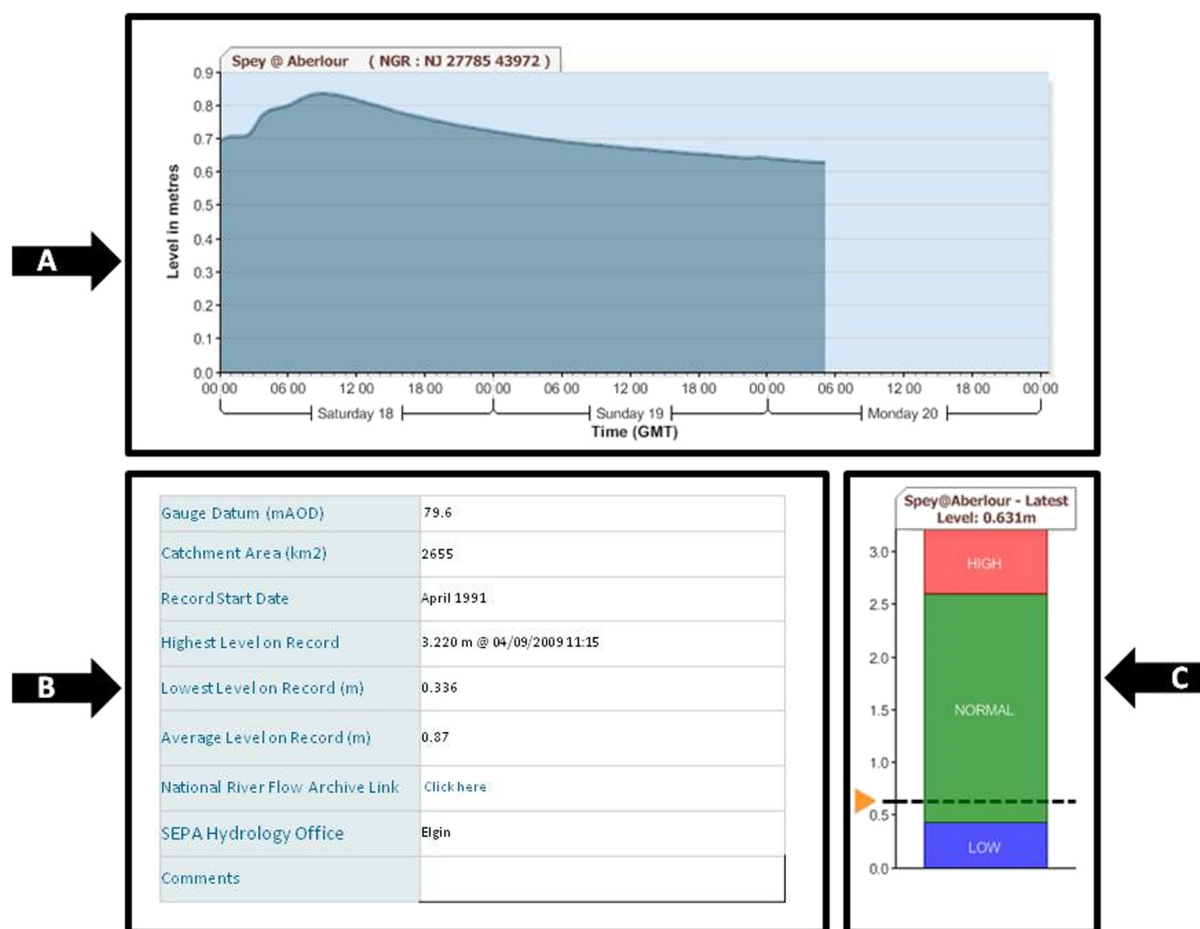
In 2012 and 2013 we ran an online survey targeting the main users of the SEPA webpages. The survey was conducted with SurveyMonkey; it was tested with SEPA staff and independent users, and we advertised the survey by means of a pop-up banner posted by SEPA on its webpages. The banner popped up once per opening period for a returning visitor, unless web browser cookies had been deleted. The survey included 14 main questions, predominantly multiple-choice but with an optional text box for each question. To counter potential seasonal bias (and to include main holidays in the United Kingdom), the survey was opened twice for a prolonged period: December 2012 to April 2013 and June 2013 to October 2013, totalling 125 and 97 days respectively. The survey was also promoted by means of emails sent to various UK organisations with an interest in Scottish rivers. These organisations were primarily identified through Google Analytics of the webpages (Section 3.1.1). A total of 1923 respondents opened the survey, resulting in 1264 unique responses that followed through from the first to the last question (questions could be skipped).

### 2.2.3. Interviews

Following on from analysis of the survey results we targeted representatives of the identified user groups. The interviewees were randomly sampled, stratified by user group from the pool of survey participants who declared willing to take part in further research and provided contact details. A total of 32 phone interviews (paddling  $n = 9$ , fishing  $n = 10$ , flood-risk-related  $n = 6$ ) were conducted. Randomly selected users (i.e. members of the public) that did not fall in either of the focal groups were also interviewed to collect additional perceptions ( $n = 7$ ). All interviews were conducted in 2013, recorded (mean duration 14 min) and verbatim transcribed.

### 2.2.4. Workshops

To explore the potential role and form of supporting textual information, two workshops with 15 participants in total were organised in 2014, with representatives of the user groups (cf. user-centred design – Bevan and Curson, 1999). Participants were invited on the basis of preparedness to participate in further research as indicated in the online survey, but stratified by the AB postcode (Aberdeen and Shire) to minimise travel for participants. Participants were divided



**Fig. 2.** The three main elements of our focal Scottish Environment Protection Agency (SEPA) river level webpages: A) 'Water Level graph' showing the changes in the water levels at one gauging station (here along the river Spey) over 72 h; B) overview of semi-static information for this station; C) 'Current water level indicator' which puts the last recorded level in the context of previously recorded levels at the station.

into two groups according to their primary interests in the river level pages (evening 1: paddling and fishing; evening 2: flood-risk related, and fishing and other). Four group exercises followed, moderated by a facilitator for each group. First, as an introduction exercise, groups were given printed examples of the current webpages (Fig. 2), and were asked to discuss: what the most important pieces of information on the webpage were; why this was the case; and how long they spent on the webpages. Second, to explore textual forms of how river level information could be presented, participants were asked to describe river level trends (verbally and written), using three printed examples of the river level webpages. Each example was chosen to show distinct hydrographs (e.g. a rapid rising limb). Third, to explore where textual information might be presented, participants were asked to consider the design of the webpage itself. Using printed examples of the river level webpages, groups discussed and annotated the pages to indicate how they felt information could be differently presented. Fourth, groups explained the outcomes to each other, followed by plenary discussion (which was recorded).

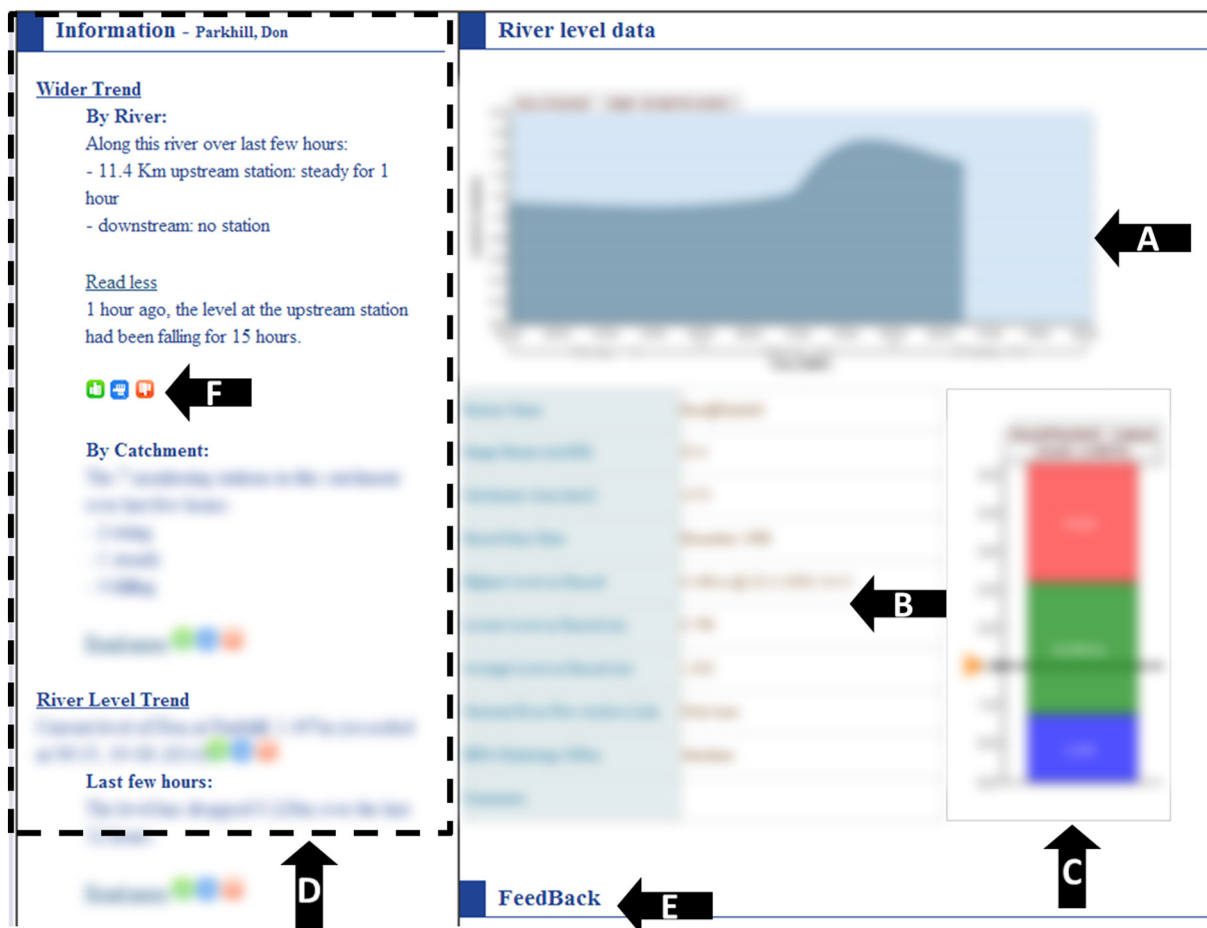
#### 2.2.5. Exploration of other online environmental information providers

To gain insight into the information that user groups used in conjunction with the river level information (but which was not to be found on the river level webpages), we studied commercial and non-commercial websites, and other digital sources such as newsletters provided by organisations with a stake in Scottish rivers. These were selected following references from text box suggestions in the

online survey and the interviews. Two informal interviews were also conducted with other river level service providers. This interview material was analysed on key features (content, interface) that were not, or differently, provided by SEPA's river level webpages.

#### 2.3. Method for Phase 2

For Phase 2, an online NLG experiment was conducted over 49 days in 2014 with 33 participants from the fishing user group. The aim of this experiment was to understand if the employment of a NLG system based on a user group profile, could help towards more effective online information provision. A copy of the SEPA webpages was built (including parts A, B and C in Fig. 2) by the research team and linked to SEPA's river level databases to provide identical information for all 333 monitoring stations in 2014. The copy of the website was hosted on a separate internet domain. Additional textual information was provided on the same screen (Fig. 3D, Table 1). All text categories (see Table 1) randomly changed order upon each new opening of the webpage by a user to pre-empt bias related to position. Except for the 'regional weather forecast' (provided through an API from the Met Office – [www.metoffice.gov.uk](http://www.metoffice.gov.uk)), text was created through NLG on the basis of the dynamic SEPA river level data. The steps carried out, i.e. the NLG 'pipeline', are summarized in Table 1.



**Fig. 3.** Example of one of 333 river level webpages which were automatically generated and presented to experiment participants. The blurred sections were part of the experimental set-up (one-click activation). Parts A–C were identical to the SEPA webpages (Fig. 2). Textual information (D), the feedback box (E) and ‘(dis)like’ buttons (F) were experimental components.

Participants were invited through their email addresses optionally entered in the online survey. Experiment participants had to create usernames and passwords and were encouraged not to use the SEPA webpages during the time of the experiment. Four forms of data collection were conducted.

- I. *Pre-experiment surveys* for research permission and baseline information, and *post-experiment surveys* with four evaluative questions related to: a) whether the additional texts had affected personal decision-making about the user group activity; b) how the participant valued the different elements of the new information (text categories in Table 1); c) how the additional texts could be improved upon; and d) if river level information provision tailored to user groups was deemed desirable.
- II. ‘Like’ buttons, to make possible rapid evaluation of new information sections by users; each text category (Table 1) was accompanied by three buttons: ‘thumbs up’, ‘thumbs horizontal’ and ‘thumbs down’ (Fig. 3F). One option could be chosen per section per visit.
- III. *Feedback box*, allowing participants to provide feedback at any time, on a topic of their choice.
- IV. *Website visit behaviour* (tacit user feedback), assessed through mouse clicks or finger tabs required to activate a blurred information section. This blurring technique was used to verify which parts of the webpages were actually viewed (and for how long). Clicking on a new section resulted in a blurring of the previous section. To ensure that participants would not

have to guess what information was available where, the headers of each textual information section were readable throughout.

For all methods, free, prior, and informed consent was obtained from all participants regarding the research purpose, methods used, use and storage of data, their rights as participants, implications of their participation, and access to research materials and outputs. Anonymity and confidentiality were ensured in all stages.

### 3. Results

#### 3.1. Users group profiles

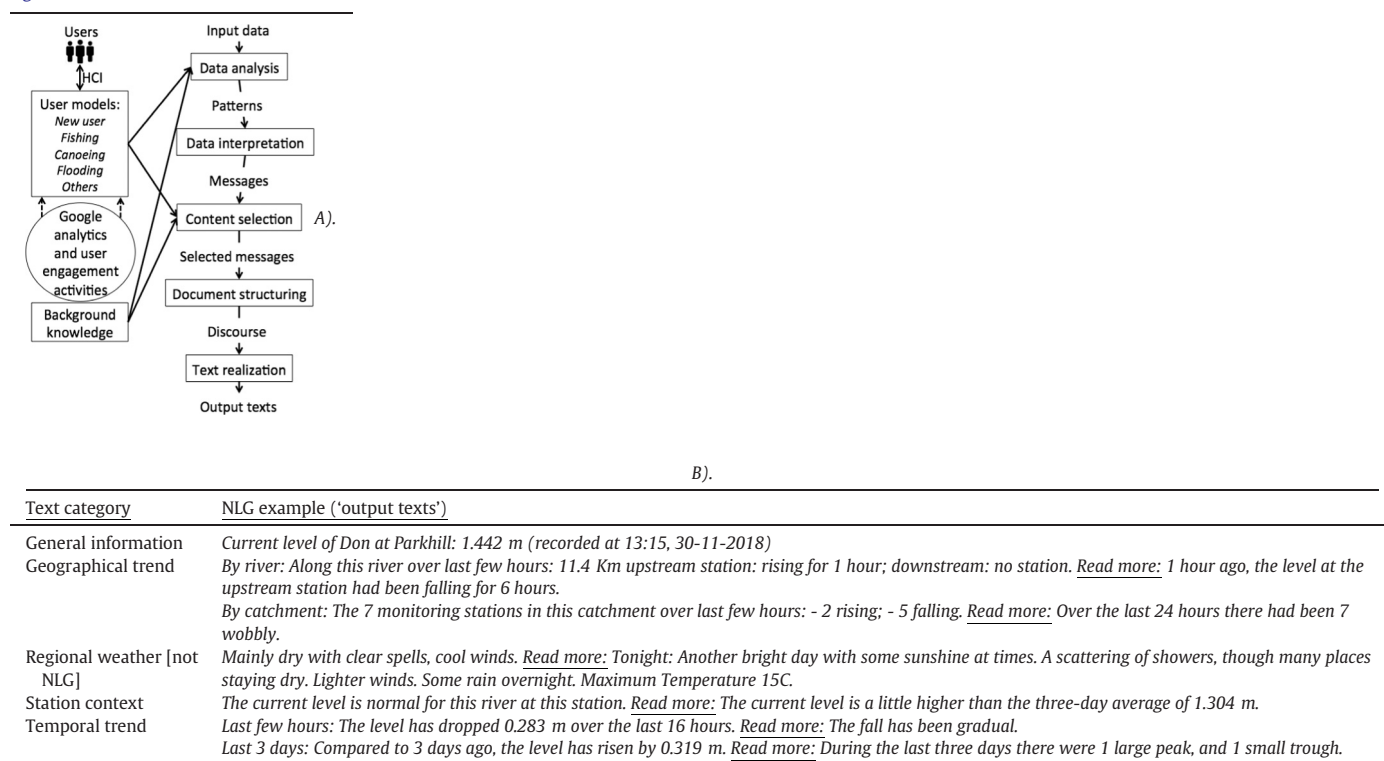
##### 3.1.1. Google Analytics

From Google Analytics it emerged that the number of ‘page views’ throughout each calendar year was seasonal, with peaks during summer months (July, August) and early autumn (September). Through analyses of traffic sources we identified five major streams of traffic: direct (39%); free search generated (34%); fishing related (14%); boat related (6%); and weather/flood risk related traffic (1%); the remaining 6% comprised miscellaneous traffic sources.

We examined the temporal nature of the traffic from each stream and found distinct patterns (Fig. 4).

Fishing-related sources generated traffic with a rather strong seasonality effect, concerning a gradual build-up of visitation rates from

**Table 1**  
A). NLG 'pipeline' showing the different steps for the data-driven automatised generation of 'output texts' NLG 'pipeline'. B). Overview of NLG text categories and examples as presented in Fig. 3D.



very low in winter to much higher and sustained rates during summer and autumn (Fig. 4A). Traffic coming from boat-related sources was relatively evenly spread across each year, with somewhat raised website use in autumn (Fig. 4B). The number of visitors from weather/flood risk related sources was low and constant, with the exception of August and December 2011 (Fig. 4D). Focusing on the two major traffic sources, search (Fig. 4C) and direct traffic (Fig. 4E), we observed strong seasonality with monthly page views more than double those during spring and winter. The different patterns presented by each cluster indicate a seasonal demand for information according to the respective group interest.

While it was not possible to determine the reasons for use of direct traffic and search traffic, the high Spearman correlation coefficients with fish-related page views values across all years (0.83–0.93; Table 2), and indeed rather similar shaped seasonal pattern of Fig. 4A, C and E, suggest that most direct and free search-generated traffic was driven by an interest in fishing.

### 3.1.2. Survey

Subsequent profiling based on the online survey helped towards the further delineation of actual user groups. Yet, it emerged, from the online survey, that the various river related activities or interests were not necessarily mutually exclusive. Analysis of answers to the online survey question 'For which of the following activities do you mainly use the web pages?' showed that 56% of respondents chose one activity (or interest in the case of 'flood risk'), and thus just under half of respondents highlighting one or more main activities or interests (Fig. 5).

Two activities were indicated by 28%, three by 125% and four or more by 4% of survey respondents. These figures underpinned the rationale for SEPA webpage user group profiling. For our purposes, we focused on profiling users related to the three largest well-delineated groups: 'fishing', 'flood risk' and 'paddling'. 'Monitoring' was excluded because it emerged to be a 'secondary' activity underpinning 'primary' activities such as fishing and paddling. Paddling was here defined as a

combination of 'canoeing' and 'kayaking', partly because many information provision websites address both groups combined or in the same interface. However, special attention was given to key differences in webpage usage and decision making around these two activities.

The 'scientific research' respondents appeared to use the webpages generally in a more technical way – for instance by downloading river level data in CSV files for their own analyses (occasionally provided by SEPA through use of hyperlink in section Fig. 2B) – and were therefore less interesting as a user group for potentially relatively basic additional textual communication.

The 'other' category made up 6% of all respondents, and included activities such as swimming, river crossing for a hill race, historical research, mink raft volunteering, diving, teaching, community council activities, mineral panning, freshwater pearl mussel surveying, photographing, path access walking, and simply "being interested". This diversity of users acted as a reminder that tailored information provision should not be presented at the cost of general information provision, but instead be an addition.

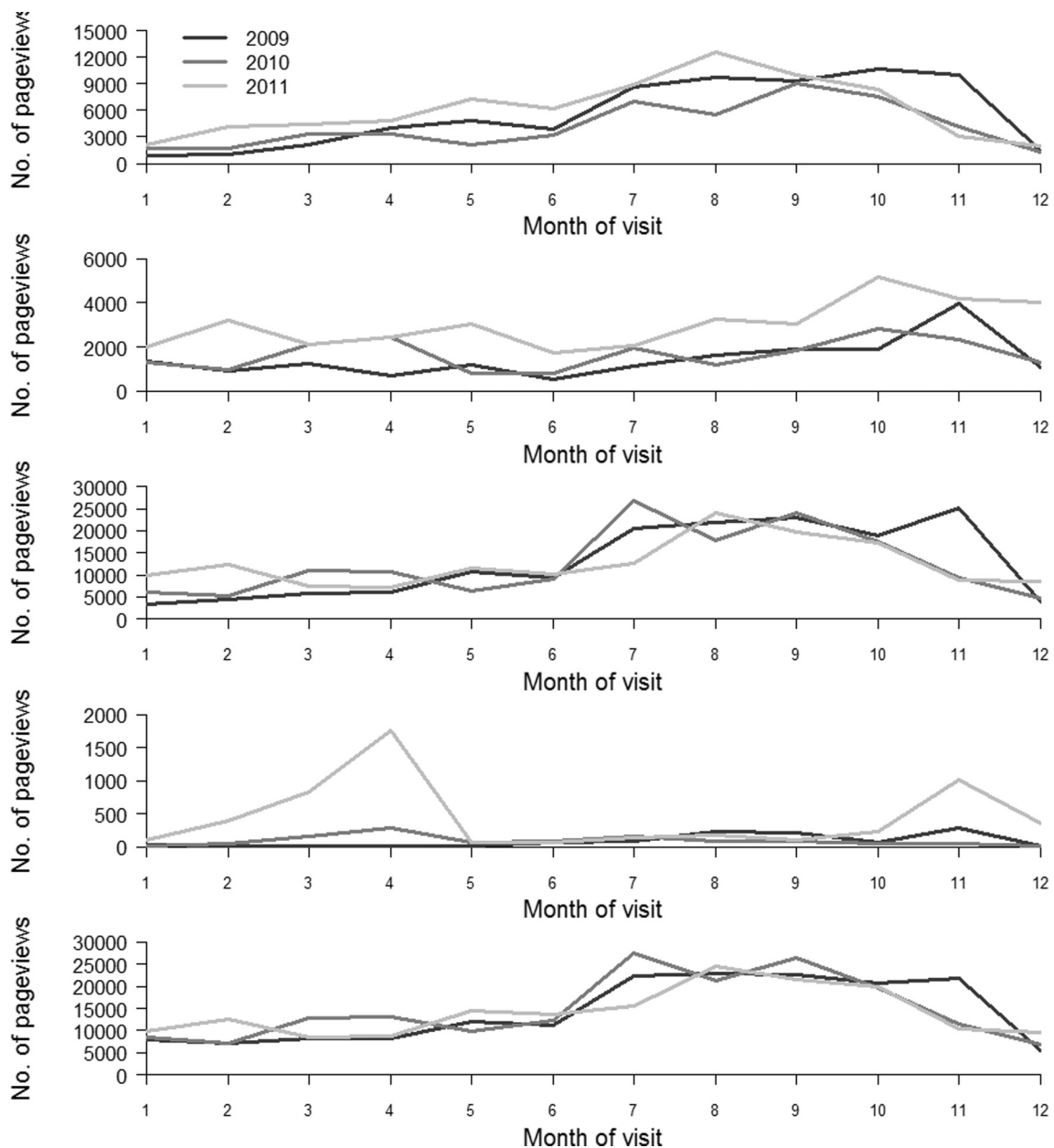
### 3.1.3. User group profiles for 'fishing', 'flood risk related' and 'paddling'

This section provides descriptions of user group profiles of 'fishing', 'flood risk' and 'paddling'. Each profile development was based on the data generated by the five methods in phase 1 of this research (Google analytics, online survey, interviews, workshops, and exploration of other online environmental information providers). Profiles are structured around three themes: webpage use; decision making about the activity; and use of other information sources.

#### 3.1.3.1. User group profile 'fishing'

**3.1.3.1.1. Webpage use.** A large proportion of all respondents to the online survey indicated that their main use of the webpages was related to fishing (Fig. 5). Almost three quarters of these respondents visited the pages every time in a personal capacity. That this concerned recreational fishing was confirmed through the interviews and the





**Fig. 4.** Monthly distribution of pageviews generated from fishing related sources (A), boat related sources (B), search related sources (C), weather/flood risk related sources (D), and direct traffic (E) to the river level webpages. The traffic volume related to the different clusters varied in one order of magnitude.

exploration of other online information providers. Almost half deemed the information on the webpages ‘very important’, and about a third ‘extremely important’. The graph (Fig. 2a) was indicated as the most relevant part of the webpage information items. In terms of potential improvements of the current information on the webpages, 110 respondents mentioned the expansion of the graph’s timescale and 86

respondents desired more frequent river level updates: “*more regular updates of river height levels would save me a lot of petrol money on wasted journeys. I know that all travelling salmon anglers, especially those travelling up from England to Scotland, would greatly appreciate more frequent updates.*” A conversion option from metric to imperial units (river level in feet and inches) was also suggested ( $n = 28$ ). The top five most visited river webpages were respectively the Tay, Clyde, Spey, Tweed and Dee. From the interviews it emerged that the Spey, Tay, Dee and Tweed are known as the ‘big four’ Scottish game fish rivers. Indicated seasonal visits dropped steeply in winter, but remained high (more than 1 visit a day) compared to other user groups.

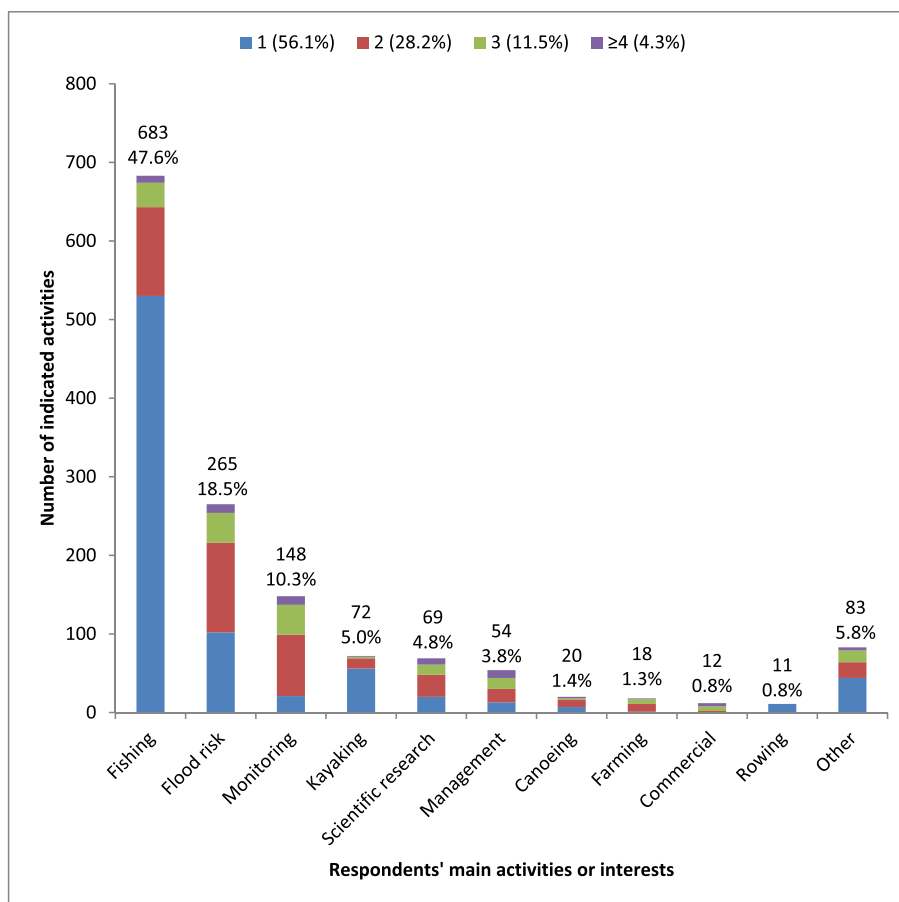
**3.1.3.1.2. Decision making about activity.** Recreational fishers along Scottish rivers fell roughly in two groups: freshwater game fish (salmon and trout) and coarse fish (all other species), and may use different techniques (e.g. fly fishing, spin fishing, legering, float fishing, of which the former two seem most popular). Most commonly fished for species were Atlantic salmon ( $\pm$ January–November), trout (prime

**Table 2**

Spearman correlation values between direct and search traffic sources with fishing and boat related traffic sources.

Correlations	2009	2010	2011
Direct vs. fishing related	0.87	0.93	0.87
Direct vs. boat related	0.56	0.68	0.53
Search vs. fishing related	0.91	0.93	0.83
Search vs. boat related	0.52	0.69	0.58





**Fig. 5.** Number of indications against each of the main activities (or interests), as generated by respondents to the online survey ( $n = 1264$  in total,  $n = 1076$  for this question with 1435 indications). The legend shows the number of different activities as indicated by respondents. For example, the red part in the 'fishing' category denotes that 113 respondents indicated two main activities, one of which was fishing.

months July and August, also by night fishing), and grayling (may be caught year round). Generally, in order to fish for game fish, advanced booking is required for access to a certain river 'beat'. Such fishers may include those who travelled from afar, arranged accommodation and are thus generally less flexible. This contrasted with local fishers who may go out when the conditions are right. Still, most fishers tended to construct a decision-making frame with many variables about where along the beat, and when, to fish. For salmon (seemingly the most fished and paid for species), interviewees described the ideal circumstances as a 'nice' river flow, with river levels neither 'too high' (prevents salmon 'running', i.e. swimming, because of high energy cost), nor 'too low' (prevents salmon running too), the level not rising too quickly (coloured water makes lure invisible to salmon), not too cold or too warm (leads to inactive behaviour - "about 57 and 58 ° F [ $\sim 14$  ° C] is ideal"). It was noted that ideal circumstances also depended on the specifics of the river and the beat, e.g. a certain river level may create favourable conditions along some beats. Other important factors included recent catches, beat availability, weather (affecting fish behaviour, but also influencing the practice of fishing, with wind direction being important for casting) time of the year, etc. Overall though, circumstances were deemed best during a gradual fall of the river level after a high: "salmon will take a fly more readily on a falling river". The most important information that fed into decision making was related to the river levels and the weather.

**3.1.3.1.3. Use of other information sources.** According to the online survey, more than half of the fishers used the webpages in combination with weather information. Other, less commonly used, sources were webcams, tidal information, Facebook updates from ghillies, angling club newsletters, and catch reports. If SEPA could provide additional

information, a vast majority of the fishing-related respondents would request water temperature, more than half rainfall information, and about a third water flow and historical river level info. From the other materials it emerged that webcams were used widely by fishers, although the coverage in Scotland is not deemed very good (yet increasing). Summer low river level indications (as opposed to all-time lows) would make a better benchmark for fishers to indicate river level (i.e. 'above normal summer low'). In contrast to the webpages, a very popular commercial website ([www.fishpal.co.uk](http://www.fishpal.co.uk)) provided: river levels in imperial units, and 'above normal summer lows'; trend (steady, falling, rising); and river level graphs also for the last 28 days and year so far. From other fishing-related information providers, it emerged that alert services for desirable river conditions seemed to grow bigger. Mobile texts alerts (through commercial providers) indicated when a river level hit the desired height, or the general trend of the river level.

### 3.1.3.2. User group profile 'flood risk related'

**3.1.3.2.1. Webpage use.** 18.5% ( $n = 265$ ) of survey respondents indicated a flood risk related interest in the webpages (Fig. 5). Of these, half mentioned using them every time in a personal capacity (which was considerably lower than the other two user groups). Information for flood risk interest was deemed mostly 'very' and 'extremely' important. To most respondents, the graph (Fig. 2A) was the most important part of the webpages (almost three quarters considered it extremely relevant), but the bar indicator (Fig. 2C) was also highly valued. 35 respondents would like to see an expansion of the graph's timescale and 32 respondents more frequent and regular updates (i.e. less delay). As one online survey respondent wrote: "As a household that has experienced flooding, and still has unresolved problems with the SEPA flood

warning system, I prefer to be able to make my own judgements on the likelihood of high water, but I can only do that when SEPA keeps its river level data current". Webpages for the larger Scottish rivers were the most visited ones. They were visited throughout the year (with less of a drop in the winter as compared to the other user groups) and with generally frequent checks (more than one a day).

**3.1.3.2.2. Decision making about interest.** SEPA has a designated flood warning system which means that the residents living in areas in danger of flooding can request being contacted by SEPA, usually by phone or text, when conditions of actual flood risk appear. Many of the 'flood risk related' user group seemed to have registered for this system. However, it emerged that the system was unsatisfactory to many webpage users; it was deemed to be too general for local geographies, and sometimes too slow. Members of this user group used the river level information on different rivers within the same river system or catchment either from upstream stations or nearby stations. They did so (in combination with weather information, particularly rainfall) to anticipate potential floods in their local area. "We get quite a lot of flood warnings, (...) and really the only way for us to tell how much it is likely to affect us particularly is from the rate of rise at the nearest monitoring station". The bar indicator received regular criticism (since flooding events often fell into the 'normal' range). An integration of the bar and graph (cf. the English Environment Agency) would make the information to some users more useful as it would allow for quicker interpretation.

**3.1.3.2.3. Use of other information sources.** From the survey respondents, almost three quarters used the webpages in combination with information about the weather, about a third with flooding-specific information, and a fifth with a web cam. If SEPA could provide additional information, rainfall information would have priority, followed by water flow and historical river level information. Weather information in relation to snow melt was also deemed important, and overviews of previous years to compare for each month would help to better contextualise the values in the bar indicator which were seen as not very useful.

### 3.1.3.3. User group profile 'paddling'

**3.1.3.3.1. Webpage use.** An outcome of the online survey was that paddlers accessed the webpages also from their smartphone and sometimes near a river. The graph (Fig. 2A) was deemed the most relevant information on the webpages. Amongst the most given suggestions for improvement of the current webpages were the expansion of the graph's timescale (towards weeks, months or even years as opposed to three days); improvement of the bar indicator (Fig. 2C) – the 'normal' category was found to be too crude and thus unhelpful. More frequent river level updates and a better map that included catchments were also returning suggestions. In contrast to kayakers, canoeists visited mostly the webpages for the larger, 'flatter' rivers such as the Tay and the Spey. The webpages were used throughout the year (with most frequent visits either in or just before the weekend), but with slightly more use for kayakers in autumn (i.e. September–November). For canoeists there was a clear dip in winter (December–February) which was not present for kayakers.

**3.1.3.3.2. Decision making about activity.** From the sources it emerged that the graph was primarily used by paddlers for two insights: 1. How much water is in the river (too low: scrape ground, too high: danger from overhanging trees); and 2. Whether river levels are falling or rising. For kayakers, rising water often meant challenging conditions, and this was usually best predicted by "big spikes" on the graph. Many experienced kayakers were 'white water kayakers' and they described themselves as 'rain chasers'. 'When?' and subsequently 'where?' were the underlying dimensions of white water kayakers' rationales; they actively looked for small rivers in spate that offered rough white water and rapids – and were prepared to travel and adjust plans last minute. An interviewee said: "Some of these [rivers] are rising and dropping and in a couple of hours, you have missed it. Especially the rarer ones that you get every, maybe once a year or once every two years". The best river

conditions tended to be in winter – more rainfall, although sometimes low upstream rivers in winter because of frozen hills – and spring (melting snow). If the summer was dry there would be very little paddling, despite summer holidays. White water kayakers did not necessarily paddle entire stretches of rivers or burns, but looked for shorter and intense runs with fast water or steep descent, often in small groups of two to six paddlers. Compared to white water kayakers, the webpages were less used by canoeists for finding rivers, and more for finding out about the conditions of the targeted river. This was mainly because fewer Scottish rivers were deemed suitable for canoeing; canoeists were generally looking for relatively flat and wide rivers (downstream) to cover larger stretches (i.e. 'touring'), and it was deemed important to know about potential rapids or shoots in such rivers. Canoeists would take less risk due to the open boat design (avoidance of strong currents), and would generally go out less in winter. Indeed, water temperature had to be 'bearable' in case of capsizing. One interviewee said that canoes were "much easier to sink and the consequences of a swamping are much more difficult to deal with. So canoeists generally tend to be more conservative than white water kayakers who probably go out more for thrills".

**3.1.3.3.3. Use of other information sources.** Weather information would be the most helpful additional information to kayakers and canoeists according to the survey. For canoeist, this did not just include rainfall but also wind, as strong wind makes steering of a canoe more challenging. Paddlers would also welcome water flow, water temperature, and historical river level information. Some paddlers also looked for dam release information and webcams. But rainfall information stood out as being most important, the reason being that it allowed users to better predict the river attributes level and flow. Important here for kayakers was that many of the steep, smaller (upstream) rivers did not have gauging stations. They overcame this by looking at the trend in the wider catchment of the particular stream. The assumption was that if adjacent rivers go up, there was good chance the river of choice could rise too. This involved taking some risk, explained an interviewee: "I have turned up at rivers hundreds of times and I have lost count the amount of times I have turned up at a river and there has been nothing in it, when you think, 'there has got to be water in it!'". It was said that the risk may be better estimated on the basis of additional information such as rainfall and water flow. Dam releases and webcams were also mentioned in this respect. Kayaking and canoeing communities often used their own calibrations for river level and river trend (sometimes on the same site and interface). Kayaking websites frequently offered tools to match rivers to personal condition preferences, or easy-to-view indications of whether rivers volumes are e.g. 'huge, very high, high, medium, low, scrapeable, empty', and their water level trends ('going up, going down, steady').

### 3.1.4. Potential for NLG

It emerged from phase 1 that the graphical presentation of the river level data was clear and understandable, and that this should remain the basis for any improved information provision. In line with the outcomes of the profiling exercises, workshop participants confirmed that their interest in river level data, and subsequent potential for improvement, was linked with the specific interests of each user group. For example, for fishers it was important to establish whether there were "angling opportunities", while kayakers were keen to know "whether parts of the river are passable". For the fishing user group there were three common elements to the descriptions of river level trends given by participants, and these were subsequently used for the creation of NLG:

- 1) The rate of the river level trend (e.g. rapidly falling);
- 2) Characteristics of the graphs i.e. steady periods, peaks, and to a lesser extent troughs, were described relative to the wider trend in the graph;
- 3) The general water level of the river.

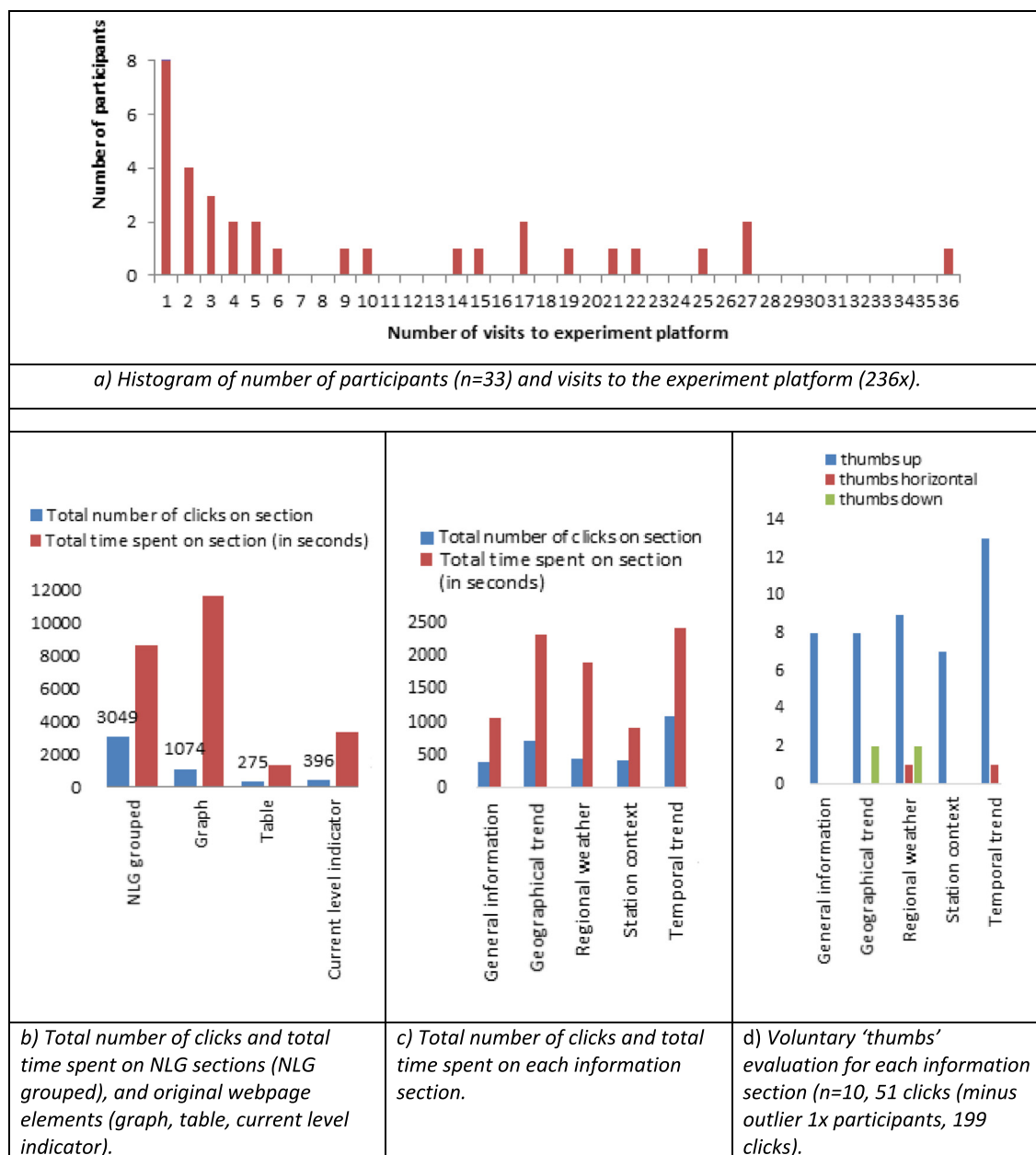


Fig. 6. Graphs from the online NLG experiment.

### 3.2. Results online NLG experiment

'Fishing' had emerged as by far the largest user groupn (Fig. 5), and so it was decided to target this group for the experiment. This generated 33 participants of this user group, resulting in 235 platform visits. Ten participants visited the website once, 23 participants twice or more (Fig. 6a). Of all the experiment elements, the graph was the most looked at element (Fig. 6b). This may be surprising because it had no novelty factor, yet it was found in phase 1 that the graph was a trusted and highly appreciated feature of the SEPA webpages for the fishing user group. Still, participants were generally positive about the 'general information' NLG text category (see Table 1, and Fig. 3D). Although offering information also visible in the graph ("should be obvious in the graphics"), it served a function for some, for example "to easily read the precise river level". Such appreciation of the 'general [river level] information' category was also reflected in Fig. 6c,d. The 'temporal trend' NLG text category scored highest in the voluntary 'thumbs' evaluation

(Fig. 6d) and received the highest amount of clicks and time spent on it (Fig. 6c). From the written feedback it emerged that the main disadvantage of this category was the lack of precision in its descriptions. 'Station context' was not valued highly by fishers and least time was spent on it (Fig. 6c); it was deemed "too general" and not adding to the graphed information. 'Regional weather' scored fairly high on all fronts (Fig. 6c,d); while not as elaborate as weather websites, it was valued by several participants in that it "saves having to access a separate weather site". 'Geographical trend' received a relatively high amount of clicks and time spent on it (Fig. 6c), but a low score from the 'thumbs' category. While it was deemed useful for obtaining a bigger picture, the mixed appreciation of the 'geographical trend' category may have reflected that the primary focus of fishers was generally on local 'micro-conditions' of the river at a particular river beat. To the question if the texts affected decision making or insight in relation to fishing, mixed replies were given. On the one hand it was felt that the information overlapped quite a bit with the information obtainable from the

graph. On the other hand they were valued as “helpful”, and one participant said: “*I really like the additional data given in the left-hand column. This is great contextual information for fishers and adds significant value to the original, pretty good, SEPA data*”. A lot of respondents saw the potential of user group tailored river level information provision. In the words of another respondent: “*yes it would be a step forward, as myself as a salmon fisherman has different needs as oppose to farmer etc.*” Other positive comments included “*set up looked very good*” and “*from a fisherman's point of view I found the experiment most helpful*”.

## 4. Discussion

### 4.1. User group profiling and NLG

Our triangulated results suggest that more basic profiling exercises for managerial purposes by public authorities may suffice as a foundation for improved information provision. The articulation of user groups through profiling and the experiment has importance not only for a river level or information tailoring context, but also for wider river and catchment planning and management. To our knowledge, we present the most elaborate description of main (online) river user groups at a national level (Section 3.1.3). Moreover, traditional stakeholder analyses tend to approach communities in terms of power relations and (conflicting) interests (cf. Reed, 2008); we considered stakeholder positions in the context of online information use, thus providing insight into relationships between offline outdoor spheres related to river use, and online spheres related to information.

When considering the evaluation methods in the experiment, the (NLG generated and thus textual) category of temporal trends came out as the most important one. This was perhaps surprising in the sense that this information was essentially ‘available’ in the graph too (as opposed to geographical trend or weather forecast). While acknowledging the important role of visualisation in information communication (Grainger et al., 2016; Levontin et al., 2017; McInerney et al., 2014), our investigation indicates that there is added value of textual information provision in addition to visual information provision. Indeed, a combination of visual and textual content appears a fruitful route (Lazard and Atkinson, 2015; Siddharthan et al., 2019). The additional layer of explanation offered by text was valued by various experiment participants. From our results we argue that NLG has the potential to play a much bigger role in tailored or dynamic environmental information provision. This aligns with conclusions by others in different domains (Conde-Clemente et al., 2017a,b; Gkatzia et al., 2017). In addition, in contexts where language may have an edge over visual representations – for instance because of limited data transfer speed, poor coverage, or older hardware or software (Banks, 2013; Maffey et al., 2015) – NLG will be of much value too. Connection to particular social media may be opportune here as well, e.g. tweeting river level stations (see e.g. <https://twitter.com/riverlevelsuk> although not using NLG). It is likely that artificial intelligence and machine learning will play a bigger role on this front too (Ayturan et al., 2018; Bellinger et al., 2017; Mosavi et al., 2018). For instance, advanced datamining can be used to create chatbots, which engage in dialogue with website users.

### 4.2. Barriers and opportunities for tailored information provision by public authorities

Online environmental information provided by public authorities is often non-specific. This study shows that, at least with respect to river level information, it may be feasible to identify and profile main user groups comprising the bulk of users (Fig. 5, Section 3.1.3), and provide tailored information. However, two main concerns should accompany such an exercise.

First, while most users indicated one focal activity or interest, almost half indicated multiple activities (Fig. 5). A challenge not addressed in

this work is how textual tailoring could support multiple activities of a user (see Webster et al., 2014).

Second, creating categories of inclusion simultaneously creates exclusions, which in turn may have implications for creating or perpetuating digital divides and social inequalities (Arts et al., 2018; Castells, 2010). Public authorities particularly need to operate conscientiously on this front and have clear responsibilities with regard to informational governance in the Information Age (Loroño-Leturiondo et al., 2019; Mol, 2008).

Bearing these two points in mind, several arguments can be made in favour of tailored approaches to information provision. This holds especially true when, as this study shows with regard to river level information, the same information is used in entirely different ways by clearly identifiable user groups.

First, it aligns with society-wide trends towards user-centric approaches of environmental governance (Zulkafli et al., 2017). If public authorities take seriously the normative components related to the late-modern governance shifts in political modernization processes (e.g. Arts and Leroy, 2006), then effective information provision – and perhaps information co-production (Loroño-Leturiondo et al., 2019; Hewitt and Macleod, 2017) – is a key vehicle towards ‘command and covenant’ stewardship and interaction with members of the public. A key question related to our case though, is to what degree mere information provision is desirable, as opposed to two-way communication streams with double or triple feedback loops. Examples of the latter reside in developments related to digital catchment observatories and polycentric environmental resources management (Mackay et al., 2015; Zulkafli et al., 2017) in which stakeholders play a central role in the constitution and operation of the information network (Macleod, 2015).

Second, consistency in non-specific information provision by public authorities may be implicitly sought but not necessarily achieved. For instance, SEPA's separate flood warning system targets residents vulnerable to flooding. Yet, the ‘flood risk related’ group pursued independent information interpretation because the flood warning system was deemed too general and sometimes too slow. Opportunities provided by access to non-specific information could thus be seen as an argument against tailored information provision. On the other hand, tailoring does not preclude additional provision of raw data or basic information, and in this case the argument by the flood risk group was not against profiling, but followed from discontent with a different service. The general point here is that tailored approaches to information provision may address differentiation in access and use amongst user groups.

## 5. Conclusion and recommendations

We set out to address if profiling of Scottish river level web page user groups (phase 1), and the subsequent associated employment of a specially designed NLG system (phase 2), could be steps towards more effective (tailored) online information provision. With regard to phase 1, we identified and described very specific profiles for the three main user groups: fishing, flood risk related and paddling. The clear delineation and existence of well-distinguishable rationales is an argument for profiling; the same river level information was used in entirely different ways by the three groups. This insight provides a strong basis for steps towards the tailoring of environmental information. Our triangulated results indicate that basic profiling exercises for managerial (non-academic) purposes by public authorities may suffice as a foundation for improved information provision.

In terms of provided information categories for the fishing user group through NLG based on live and dynamic river level data (phase 2), the category of temporal trend came out as the most important one. It demonstrated that besides visual information, textual information provision can be of value to users. The additional layer of explanation, which is arguably more difficult to provide visually, is valued by



many and plays an important role in translating dynamic technical information to plain messages for the specific purposes of the user groups.

Finally, we list five key recommendations for public authorities and other information providers (for instance operating in the realm of climate services – Vaughan and Dessai, 2014):

1. If transformations towards 'command and covenant' stewardship are taken seriously by public authorities (Arts et al., 2016), more effective information provision and communication is a key dimension, often in need of improvement.
2. Many public authorities and other information providers do not have full insight into who uses the provided information, for what reasons, and how it is interpreted. Profiling user groups is a fruitful and relatively straightforward way of finding out about a range of uses and demands for information.
3. Effective information provision may take many shapes, but since environmental information is used in entirely different ways by diverse user groups, the tailoring of information is an important step towards more targeted information provision and more effective communication.
4. There is value in textual information provision (for instance provided through NLG) alongside visual information provision.
5. Collaborations between public authorities who provide information and (interdisciplinary) research groups are needed to lead to researched applications of digital, social and natural science knowledge and methodologies, to provide solid insights for all involved and beyond.

## Acknowledgements

The authors thank SEPA staff, particularly Bruce Eriksen, as well as all research informants and experiment participants for their time and effort. The research described here was supported by an award (EP/G066051/1) made by the Research Councils United Kingdom (RCUK) Digital Economy programme to the dot.rural Digital Economy Hub at the University of Aberdeen.

## References

- Arts, B., Leroy, P., 2006. Institutional processes in environmental governance: lots of dynamics, not much change? In: Arts, B., Leroy, P. (Eds.), *Institutional Dynamics in Environmental Governance*. Springer, Dordrecht, the Netherlands, pp. 267–283.
- Arts, K., Webster, G., Sharma, N., Melero, Y., Mellish, C., Lambin, X., Van der Wal, R., 2013. Capturing mink and data: interacting with a small and dispersed environmental. Framework for Responsible Research and Innovation in ICT FRRRICT-Case Study, pp. 1–5.
- Arts, K., Ioris, A.A.R., MacLeod, C.J.A., Han, X., Sripada, S., Braga, J.R.Z., Van der Wal, R., 2015a. Supply of online environmental information to unknown demand: the importance of interpretation and liability related to a national network of river level data. *Scott. Geogr. J.* 1–8 <https://doi.org/10.1080/14702541.2014.978809>.
- Arts, K., Rabelo, M., De Figueiredo, D., Maffey, G., Ioris, A., Girard, P., 2018. Online and offline representations of biocultural diversity: a political ecology perspective on nature-based tourism and indigenous communities in the Brazilian Pantanal. *Sustainability* 10. <https://doi.org/10.3390/su10103643>.
- Arts, K., Ioris, A.A.R., Macleod, C.J.A., Han, X., Sripada, S.G., Braga, J.R.Z., van der Wal, R., 2016. Environmental communication in the information age: institutional barriers and opportunities in the provision of river data to the general public. *Environ. Sci. Pol.* 55, 47–53. <https://doi.org/10.1016/j.envsci.2015.08.011>.
- Arts, K., Rabelo, M., De Figueiredo, D., Maffey, G., Ioris, A., Girard, P., 2018. Online and offline representations of biocultural diversity: a political ecology perspective on nature-based tourism and indigenous communities in the Brazilian Pantanal. *Sustainability* 10. <https://doi.org/10.3390/su10103643>.
- Ayturan, Y.A., Ayturan, Z.C., Altun, H.O., 2018. Air pollution modelling with deep learning: a review. *Int. J. Environ. Pollut. Environ. Model.* 1, 58–62.
- Bäckstrand, K., 2003. Civic science for sustainability: reframing the role of experts, policy-makers and citizens in environmental governance. *Glob. Environ. Polit.* 3, 24–41. <https://doi.org/10.1162/152638003322757916>.
- Banks, K., 2013. *The Rise of the Reluctant Innovator*. 1st edition. London Publishing Partnership, London.
- Barker, J.L.P., Macleod, C.J.A., 2019. Development of a national-scale real-time Twitter data mining pipeline for social geodata on the potential impacts of flooding on communities. *Environ. Model. Softw.* 115, 213–227. <https://doi.org/10.1016/j.envsoft.2018.11.013>.
- Bellinger, C., Mohamed Jabbar, M.S., Zaiane, O., Osornio-Vargas, A., 2017. A systematic review of data mining and machine learning for air pollution epidemiology. *BMC Public Health* 17, 1–19. <https://doi.org/10.1186/s12889-017-4914-3>.
- Bevan, N., Curson, I., 1999. Planning and implementing user-centred design. CHI'99 Extended Abstracts on Human Factors in Computing Systems. ACM ISBN, pp. 137–138.
- Black, A.R., Cranston, M.D., 1995. River flow gauging station data usage and network evolution in Scotland. In: Black, A.R., Johnson, R.C. (Eds.), *Fifth National Hydrology Symposium*, Edinburgh. British Hydrological Society, Wallingford, pp. 6.19–6.25.
- Blei, D.M., Smyth, P., 2017. Science and data science. *Proc. Natl. Acad. Sci.* 114, 8689–8692. <https://doi.org/10.1073/pnas.1702076114>.
- Campbell, J.L., Rustad, L.E., Porter, J.H., Taylor, J.R., Dereszyński, E.W., Shanley, J.B., Gries, C., Henshaw, D.L., Martin, M.E., Sheldon, W.M., Boose, E.R., 2013. Quantity is nothing without quality: automated QA/QC for streaming environmental sensor data. *Bioscience* 63, 574–585. <https://doi.org/10.1525/bio.2013.63.7.10>.
- Castells, M., 2010. *The Information Age: Economy, Society, and Culture. Volume I: The Rise of the Network Society*. Wiley-Blackwell, Chichester.
- Christel, I., Hemment, D., Bojovic, D., Cucchiatti, F., Calvo, L., Stefaner, M., Buontempo, C., 2018. Introducing design in the development of effective climate services. *Clim. Serv.* 9, 111–121. <https://doi.org/10.1016/j.cliser.2017.06.002>.
- Cieslik, K.J., Leeuwis, C., Dewulf, A.R.P.J., Lie, R., Werners, S.E., Van Wessel, M., Feindt, P., Struijk, P.C., 2018. Addressing socio-ecological development challenges in the digital age: exploring the potential of Environmental Virtual Observatories for Connective Action (EVOCA). *NJAS-Wagen J Life Sci.* 86–87, 2–11. <https://doi.org/10.1016/j.njas.2018.07.006>.
- Conde-Clemente, P., Alonso, J.M., Nunes, É.O., Sanchez, A., Trivino, G., 2017a. New types of computational perceptions: linguistic descriptions in deforestation analysis. *Expert Syst. Appl.* 85, 46–60. <https://doi.org/10.1016/j.eswa.2017.05.031>.
- Conde-Clemente, P., Alonso, J.M., Trivino, G., 2017b. rLDCP: R package for text generation from data. 2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), pp. 1–6 <https://doi.org/10.1109/FUZZ-IEEE.2017.8015487>.
- Conde-Clemente, P., Trivino, G., Alonso, J.M., 2017c. Generating automatic linguistic descriptions with big data. *Inf. Sci.* 380, 12–30.
- De Longueville Bertrand, B., 2010. Community-based geoportals: the next generation? Concepts and methods for the geospatial Web 2.0. *Comput. Environ. Urban. Syst.* 34, 299–308. <https://doi.org/10.1016/j.compenvurbysys.2010.04.004>.
- Fleischhauer, M., Greiving, S., Flex, F., Scheibel, M., Stickler, T., Sereinig, N., Koboltschnig, G., Malvati, P., Vitale, V., Grifoni, P., Firus, K., 2012. Improving the active involvement of stakeholders and the public in flood risk management - tools of an involvement strategy and case study results from Austria, Germany and Italy. *Nat. Hazards Earth Syst. Sci.* 12, 2785–2798. <https://doi.org/10.5194/nhess-12-2785-2012>.
- Gkatzia, D., Lemon, O., Rieser, V., 2017. Data-to-text generation improves decision-making under uncertainty. *IEEE Comput. Intell. Mag.* 12, 10–17. <https://doi.org/10.1109/MCI.2017.2708998>.
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth Engine: planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* 202, 18–27. <https://doi.org/10.1016/j.rse.2017.06.031>.
- Grainger, S., Mao, F., Buytaert, W., 2016. Environmental data visualisation for non-scientific contexts: literature review and design framework. *Environ. Model. Softw.* 85, 299–318. <https://doi.org/10.1016/j.envsoft.2016.09.004>.
- Hannah, D.M., Demuth, S., van Lanen, H.A.J., Looser, U., Prudhomme, C., Rees, G., Stahl, K., Tallaksen, L.M., 2011. Large-scale river flow archives: importance, current status and future needs. *Hydrol. Process.* 25, 1191–1200. <https://doi.org/10.1002/hyp.7794>.
- Hertzum, M., 1999. A review of museum web sites: in search of user-centred design. *Arch. Museum Informatics Cult. Herit. Informatics Q.* 12, 127–138. <https://doi.org/10.1023/A:1009009104685>.
- Hewitt, R., Macleod, C., 2017. What do users really need? Participatory development of decision support tools for environmental management based on outcomes. *Environ.* 4, 88. <https://doi.org/10.3390/environments4040088>.
- Ioris, A.A.R., 2008. Water institutional reforms in Scotland: contested objectives and hidden disputes. *Water Altern.* 1, 253–270.
- Janssen, M., Charalabidis, Y., Zuiderwijk, A., 2012. Benefits, adoption barriers and myths of open data and open government. *Inf. Syst. Manag.* 29, 258–268. <https://doi.org/10.1080/10580530.2012.716740>.
- Kamal, M.M., 2006. IT innovation adoption in the government sector: identifying the critical success factors. *J. Enterp. Inf. Manag.* 19, 192–222.
- Kryvasheyeu, Y., Chen, H., Obradovich, N., Moro, E., Van Hentenryck, P., Fowler, J., Cebrían, M., 2016. Rapid assessment of disaster damage using social media activity. *Sci. Adv.* 2, e1500779. <https://doi.org/10.1126/sciadv.1500779>.
- Lamers, M., Pristupa, A., Amelung, B., Knol, M., 2016. The changing role of environmental information in Arctic marine governance. *Curr. Opin. Environ. Sustain.* 18, 49–55. <https://doi.org/10.1016/j.cosust.2015.08.015>.
- Lazard, A., Atkinson, L., 2015. Putting environmental infographics center stage: the role of visuals at the elaboration likelihood model's critical point of persuasion. *Sci. Commun.* 37, 6–33. <https://doi.org/10.1177/1075547014555997>.
- Levontin, P., Baranowski, P., Leach, A.W., Bailey, A., Mumford, J.D., Quetglas, A., Kell, L.T., 2017. On the role of visualisation in fisheries management. *Mar. Policy* 78, 114–121. <https://doi.org/10.1016/j.marpol.2017.01.018>.
- Lorofo-Leturiondo, M., O'Hare, P., Cook, S.J., Hoon, S.R., Illingworth, S., 2019. Building bridges between experts and the public: a comparison of two-way communication formats for flooding and air pollution risk. *Geosci. Commun.* 2 (39–53), 2019. <https://doi.org/10.5194/gc-2-39-2019>.
- Mackay, E.B., Wilkinson, M.E., Macleod, C.J.A., Beven, K., Percy, B.J., Macklin, M.G., Quinn, P.F., Stutter, M., Haygarth, P.M., 2015. Digital catchment observatories: a platform for engagement and knowledge exchange between catchment scientists, policy makers, and local communities. *Water Resour. Res.* 51, 4815–4822. <https://doi.org/10.1002/2014WR016824>.
- Macleod, C.K.J.A., 2015. Collaborative knowledge in catchment research networks: integrative research requirements for catchment systems science. *Collaborative Knowledge in Scientific Research Networks*. IGI Global, pp. 214–236.
- Macleod, C., Sripada, S., Ioris, A., Arts, K., Van der Wal, R., 2012. Communicating river level data and information to stakeholders with different interests: the participative

- development of an interactive online service. In: Seppelt, R., Voinov, A., Lange, S., D. B. (Eds.), *International Congress on Environmental Modelling and Software Managing Resources of a Limited Planet—Sixth Biennial Meeting*. International Environmental Modelling and Software Society (iEMSs), Leipzig.
- Maffey, G., Homans, H., Banks, K., Arts, K., 2015. Digital technology and human development: a charter for nature conservation. *Ambio* 44, 527–537. <https://doi.org/10.1007/s13280-015-0703-3>.
- Mathur, P., 2009. Environmental communication in the information society: the blueprint from Europe. *Inf. Soc.* 25, 119. <https://doi.org/10.1080/01972240802701676>.
- McInerny, G., Chen, M., Freeman, R., Gavaghan, D., Meyer, M., Rowland, F., Spiegelhalter, D., Stefaner, M., Tassarolo, G., Hortal, J., 2014. Information visualisation for science and policy: engaging users and avoiding bias. *Trends Ecol. Evol.* 29, 148–157. <https://doi.org/10.1016/j.tree.2014.01.003>.
- Mol, A.P.J., 2008. *Environment Reform in the Information Age: The Contours of Information Governance*. University Press, Cambridge <https://doi.org/10.1068/c0508j>.
- Molina, M., 2012. Simulating data journalism to communicate hydrological information from sensor networks. *Adv. artif. intell. – IBERAMIA 2012. Lect. Notes Comput. Sci* 7637, 722–731.
- Molina, M., Flores, V., 2012. Expert systems with applications generating multimedia presentations that summarize the behavior of dynamic systems using a model-based approach. *Expert Syst. Appl.* 39, 2759–2770. <https://doi.org/10.1016/j.eswa.2011.08.135>.
- Montanari, A., Toth, E., Young, G., Savenije, H.H.G., Hughes, D., Wagener, T., Ren, L., Koutsoyiannis, D., Cudennec, C., Grimaldi, S., Blöschl, G., Rogger, M., Viglione, A., Sivapalan, M., Beven, K., Gupta, H., Hipsey, M., Schaeffli, B., Arheimer, B., Boegh, E., Schymanski, S.J., Di Baldassarre, G., Yu, B., Hubert, P., Huang, Y., Schumann, A., Post, D., Srinivasan, V., Harman, C., Thompson, S., McMillan, H., Characklis, G., Pang, Z., Belyaev, V., 2013. “Panta Rhei—everything flows”: change in hydrology and society – the IAHS scientific decade 2013–2022. *Hydrol. Sci. J.* 58, 1256–1275. <https://doi.org/10.1080/02626667.2013.809088>.
- Mosavi, A., Ozturk, P., Chau, K.W., 2018. Flood prediction using machine learning models: literature review. *Water (Switzerland)* 10, 1–40. <https://doi.org/10.3390/w10111536>.
- Mukhtarov, F., Dieperink, C., Driessen, P., 2018. The influence of information and communication technologies on public participation in urban water governance: a review of place-based research. *Environ. Sci. Pol.* 89, 430–438. <https://doi.org/10.1016/j.envsci.2018.08.015>.
- Novák, V., 2016. Mining information from time series in the form of sentences of natural language. *Int. J. Approx. Reason.* 78, 192–209. <https://doi.org/10.1016/j.ijar.2016.07.006>.
- Paul, J.D., Buytaert, W., Allen, S., Ballesteros-Cánovas, J.A., Bhusal, J., Cieslik, K., Clark, J., Dugar, S., Hannah, D.M., Stoffel, M., Dewulf, A., Dhital, M.R., Liu, W., Nayaval, J.L., Neupane, B., Schiller, A., Smith, P.J., Supper, R., 2018. Citizen science for hydrological risk reduction and resilience building. *WIREs Water* 2018 (5), e1262. <https://doi.org/10.1002/wat2.1262>.
- Ramos-Soto, A., Bugarin, A., Barro, S., Taboada, J., 2013. Linguistic descriptions for automatic generation of textual short-term weather forecasts on real prediction data. *Flex. query answering syst. Lect. Notes Comput. Sci* 8132, 269–280. <https://doi.org/10.1109/TFUZZ.2014.2328011>.
- Reed, M.S., 2008. Stakeholder participation for environmental management: a literature review. *Biol. Conserv.* 141, 2417–2431. <https://doi.org/10.1016/j.biocon.2008.07.014>.
- Reiter, E., Dale, R., 2000. *Building Natural Language Generation Systems*. 1st edition. University Press, Cambridge.
- Sandoval-Almazan, R., Gil-Garcia, J.R., 2012. Are government internet portals evolving towards more interaction, participation, and collaboration? Revisiting the rhetoric of e-government among municipalities. *Gov. Inf. Q.* 29, S72–S81. <https://doi.org/10.1016/j.giq.2011.09.004>.
- Shadbolt, N., Hara, K.O., Berners-lee, T., Gibbins, N., Glaser, H., 2012. *Linked open government data: lessons from Data.gov.uk*. *IEEE Intell. Syst.* 27, 16–24.
- Siddharthan, A., Ponnampereuma, K., Mellish, C., Zeng, C., Heptinstall, D., Robinson, A., Benn, S., Van der Wal, R., 2019. Blogging birds: telling informative stories about the lives of birds from telemetric data. *Commun. ACM* 62, 68–77. <https://doi.org/10.1145/3231588>.
- Soma, K., MacDonald, B.H., Termeer, C.J.A.M., Opdam, P., 2016a. Introduction article: informational governance and environmental sustainability. *Curr. Opin. Environ. Sustain.* 18, 131–139. <https://doi.org/10.1016/j.cosust.2015.09.005>.
- Soma, K., Termeer, C.J.A.M., Opdam, P., 2016b. Informational governance – a systematic literature review of governance for sustainability in the information age. *Environ. Sci. Pol.* 56, 89–99. <https://doi.org/10.1016/j.envsci.2015.11.006>.
- Vaughan, C., Dessai, S., 2014. Climate services for society: origins, institutional arrangements, and design elements for an evaluation framework. *WIREs Clim. Change* 5, 587–603. <https://doi.org/10.1002/wcc.290>.
- Vitolo, C., Elkhatib, Y., Reusser, D., Macleod, C.J.A., Buytaert, W., 2015. Web technologies for environmental big data. *Environ. Model. Softw.* 63, 185–198. <https://doi.org/10.1016/j.envsoft.2014.10.007>.
- Webster, G., Sripada, S., Mellish, C., Melero, Y., Arts, K., Lambin, X., Van der Wal, R., 2014. Determining content for unknown users: lessons from the MinkApp case study. *Proceedings of the 8th International Natural Language Generation Conference (INLG)* 19–21 June 2014 Philadelphia, USA. <http://www.aclweb.org/anthology/W14-4417>.
- Zulkafli, Z., Perez, K., Vitolo, C., Buytaert, W., Karpouzoglou, T., Dewulf, A., De Bièvre, B., Clark, J., Hannah, D.M., Shaheed, S., 2017. User-driven design of decision support systems for polycentric environmental resources management. *Environ. Model. Softw.* 88, 58–73. <https://doi.org/10.1016/j.envsoft.2016.10.012>.