The treatment of public participation data in environmental impact assessment: setting up smart systems for the synthesis and mapping of vague definitions

Co-authors: Sanchez-Trigueros, F.[1]; Carver, S.[1]; Huck, J.[2]; Matt, R.[3], McBride, B. B.[4] ; Stumpff, L. M.[5] and Watson, A.[4]

Speaker: Sanchez-Trigueros, F.[1]

[1] School of Geography, Faculty of the Environment, University of Leeds.
[2] Imagination Lancaster, LICA, Lancaster University.
[3] Forestry Department, The Confederated Salish and Kootenai Tribes of the Flathead Nation.
[4] College of Forestry and Conservation, The University of Montana.
[5] Evergreen State College.
[6] Aldo Leopold Wilderness Research Institute, USDA Forest Service.

Corresponding author: Fernando Sanchez-Trigueros (F.SanchezTrigueros@leeds.ac.uk)

1. Introduction

This paper describes a research protocol and software implementation for the synthesis, retrieval and mapping of georeferenced documents. The model is based on the addition of written text to events located in space, laying theoretical foundations for Information Systems with both geospatial and text mining functionality. We focus on human languages and describe a routine for the treatment of written text provided from spatial surveys, in the context of crowd-sourcing environmental information.

Fig. 1 depicts the outline of the research protocol. We consider Stages 1 and 5 to be of universal application in studies founded on principles of the scientific method, whereas Stages 2 and 3 have already been treated before (e.g. Carver *et al.*, 2009; Gunderson, 2006; Gunderson and Watson, 2007). The paper hence centres on the technical basis for the execution of Stage 4, demonstrating the full research process, however, with a study on public perceptions of environmental change on the Flathead Indian Reservation (Montana).

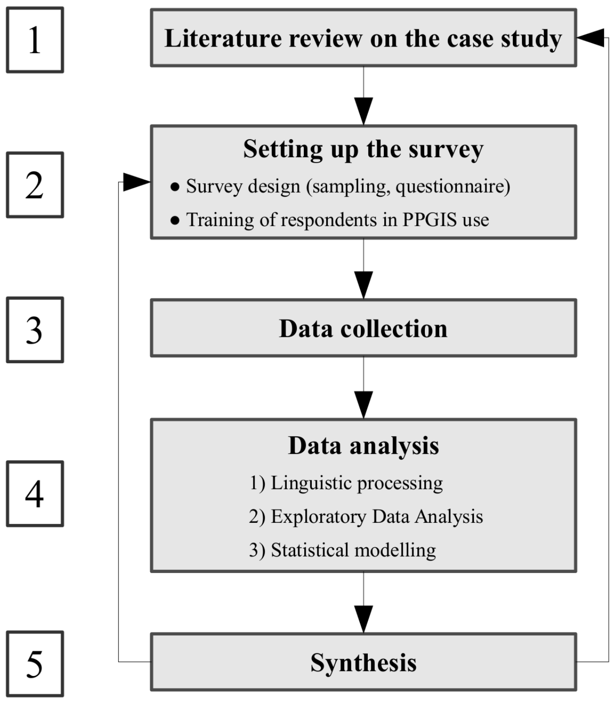


Fig. 1. Suggested workflow for the study of spatiotemporally referenced text datasets.

2. The role of Public Participation GIS in environmental impact assessment

Changes in local ecosystems, as documented throughout the planet over the last decades, have been associated with a steep rising of the Earth's average surface temperature on a global scale (Solomon *et al.*, 2007). Deciding what strategies to take so that habitats and human communities are resilient to these impacts has hence become a significant challenge, in the context of which risk and monitoring have taken a major role in setting the pace of policy and planning agendas (e.g. the UN's Millenium Development Goals[1] or the EU's Europe 2020[2] strategy; see also Field *et al.*, 2012; Parry *et al.*, 2007; Metz *et al.*, 2007). In this respect we believe in the positive effects of including the social actors, their interests and perceptions in the process of policy-making, planning and implementation, given the capacity of public engagement to catalyze the consolidation of decisions taken at the societal scale (Pickles, 1995). Public Participation Geographical Information Systems (PPGIS) have the potential to become instrumental in documenting such interests and perceptions,

1 http://www.un.org/millenniumgoals. [Accessed on 17/12/2013]
2 http://ec.europa.eu/europe2020/index_en.htm. [Accessed on 17/12/2013]

by providing technologies for the involvement of communities in the description of space (Kingston, 2007). When coupled with text data, PPGIS can additionally allow the segregation of landscape properties on the basis of the meanings people ascribe to locations, and hence lead to a better understanding of spatial relationships between tangible and intangible elements of humanized ecosystems (Carver *et al.*, 2001). For instance, PPGIS platforms have been used to describe individual perceptions of wilderness among traditional communities (Carver *et al.*, 2009).

3. The problem of vagueness in the qualification of data

A major impediment to the retrieval of information from text data is the allocation of meaning to vague descriptions and the uncertain definitions that derive from it. Uncertainty *sensu lato* is a property of decision-making processes that arises from imperfect knowledge about the initial formation and following development of systems. The notion of vagueness can thus be placed within a general taxonomy of uncertainty where specific properties of qualitative and quantitative data types are distinguished (Table 1). Following this ontology, the paper considers the effects of ambivalence, or more specifically of semantic ambivalence as it is understood after a closer consideration of the concept (Table 2), on taking account of qualitative properties in text data.

4. Solutions to information retrieval from the automated processing of human languages

Retrieval of semantic information is implemented by Natural Language Processing (NLP), a main component of Artificial Intelligence where computer science and linguistics converge for the development of verbalized human–computer interaction systems (Chowdhury, 2003; Joshi, 1991). NLP makes use of machine learning, data mining, computational linguistics and, more broadly, is founded on principles of statistical linguistics and cognitive linguistics. Synthesis and retrieval of semantic properties out of text data hence involves an analytical problem where application of NLP-based procedures naturally emerges, especially of those aiming at morphological segmentation (e.g. for the generalization of words as neutral forms), named entity recognition, word sense disambiguation, co-reference resolution (e.g. anaphora resolution), part-of-speech tagging, sentence breaking, syntactic analysis, text simplification, automatic summarization, natural language understanding (e.g. through first-order logic), sentiment analysis and topic segmentation (Jurafsky and Martin, 2008; Manning and Schütze, 1999).

Table 1. A taxonomy of uncertainty.

Type of uncertainty	Explanation	Type of data	
		Quantitative	Qualitative
Classification ambiguity	Variable systems to allocate values to classes.	●	●
Accuracy	Measurable difference between the observed value and the real one.	● (Metric)	● (Topological)
Ambivalence	Feasibility that several values may be allocated to a single event.	●	●
Completeness	Occurrence of missing data.	●	●

Table 2. A taxonomy of ambivalence.

Type of ambivalence	Explanation	Type of data	
		Quantitative	Qualitative
Decimal redefinition	Re-expansion of the decimal digits of a previously rounded or truncated quantity due to numerical generalization.	●	
Semantic ambivalence (aka semantic vagueness)	Interpretation of a data unit with alternative meanings.		
a) Subclass redefinition	Redefinition of a superclass term with one of its subclasses.		●
b) Homonymy	Redefinition of a term having several meanings (e.g. pronoun precision and polysemy).		●
c) Semantic inference	Alternative meanings can be inferred from a text when testing textual entailments.		●
Random ambivalence (aka precision)	The value of an event is modelled as a stochastic realization of a probability model. The dispersion of feasible values gives the precision of the measurement.	●	●
Procedural ambivalence	Different results are obtained on applying alternative, suitable methods.	●	●
Stage ambivalence:	An event may have had different stages along a defined time span.		
a) Inclusive	All stages have actually occurred.	●	●
b) Exclusive	The actual occurrence of a stage invalidates the possibility that others have occurred.	●	●

5. Setting up intelligent geographical information systems: the multimodal database

In the present study the concept of *multimodality* (Jewit, 2009; Kress, 2010) is used to refer to objects that, potentially, may have spatiotemporal coordinates, verbalized expressions (typically written text) and co-located attributes (that is, any geospatial process that coincides with the object in space-time, e.g. unemployment rates, pollution indexes or named urban entities), all deriving from common or different data sources. As shown in Fig. 1, Stage 5 is actually a three-step process that (1) starts with linguistic synthesis and organization of text data (e.g. by the summary and clustering of linguistic properties), (2) proceeds to Exploratory Data Analysis for description of multidimensional patterns in the multimodal dataset and (3) concludes with the statistical modelling of such patterns, paying special attention to the testing of spatial processes explaining the distribution of verbalized information. Ultimately, the analysis aims at the retrieval of meanings provided by data sources, allowing the mapping of human perceptions and the testing of associations between messages and the spatiotemporal context where such messages were produced.

Central instruments in this analysis are (i) a logical data structure for the integral storage of multimodal data and (ii) an interface that enables their retrieval and examination. In our data ontology each object in the multimodal database is a *feature* that may contain spatiotemporal, linguistic and measure-based properties. For multimodal data storage we thus consider an Object-Oriented database where events may have all or some of the above types and related attributes – geospatial (geometry, reference system, location), textual (lexicon, part-of-speech syntax, textual entailments) and co-locational. Time is expressed by grouping all contemporaneous features into the same *time* layer of a multimodal dataset (Fig. 2).

Semantic retrieval and mapping is done by an *ad hoc* extension of Structured Query Language that implements NLP functionality through a naturalized query language, in keeping with the Natural Language Programming paradigm (Veres, 2008). The present development aims at testing implicit meanings, e.g. by the syntax `select [those] features where [their] text entails` followed by the statement to be evaluated, as well as the assessment of semantic vagueness by commands such as `select [those] features where [their] text has [the] noun heather or [its] superclass` (words between brackets can be omitted), referring to lexical hierarchies like those of the WordNet corpus[3].

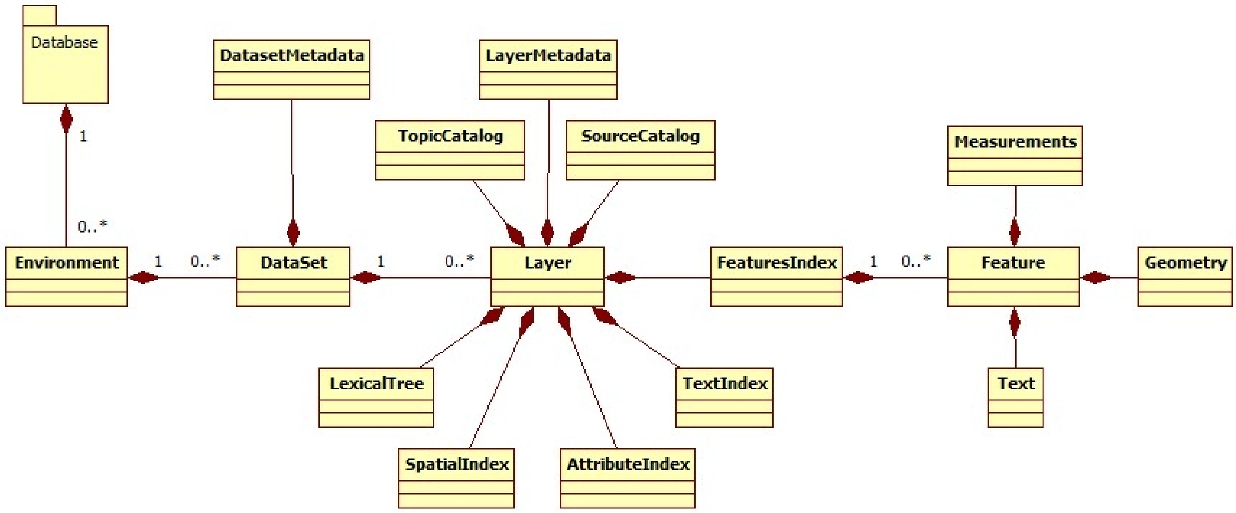


Fig. 2. Class diagram of the integral database. All associations between classes are One-to-One compositions of the 'has a' type (i.e. the owner class, the one with the filled diamond attached, has strictly one instance of the owned class) unless otherwise stated.

6. Case study: fire ecosystems and traditional ecological knowledge in the northern Rocky Mountains

We show some capabilities of this information system with a PPGIS case study on public perceptions of environmental change in fire ecosystems of the Flathead Indian Reservation (Montana, USA) (Fig. 3), within the wider topic of investigating interrelations between science, technology and Traditional Ecological Knowledge (TEK). A major source of concern on the reservation is the effect of fire suppression on the spatial structure and health of tribal woodlands and related habitats. The widespread implementation of federal policies against wildfires over the last century has led to forest ecosystems afflicted by oversized tree communities, anomalous accumulations of dead wood on the forest floor, dense understories of brush and young trees and closed forest canopies (CSKT, 2005), which has largely contributed to an increase in seasonal episodes of uncontrollable wildfire, the reduction of soil moisture, a decrease in sunlight to the forest floor and a proliferation of plant pathogens and disease (CSKT, 2013).

To provide a better depiction on how local communities perceive the current state of tribal

3 http://wordnet.princeton.edu. [Accessed on 18/12/2013]

woodlands, the Confederated Salish and Kootenai Tribes have closely collaborated with the Aldo Leopold Wilderness Research Institute in a novel approach that seeks for synergies between academia and native communities in knowledge management. Within this framework of tribal involvement a semi-structured PPGIS survey was conducted under tribal supervision among a sample of twenty-nine key informants representing tribal and non-tribal residents of the Jocko landscape unit, with emphasis placed on (i) residents' mapping of the locations of recent changes in their local environment, (ii) residents' attachment of meanings to those locations, and (iii) residents' perceptions about the potential application of TEK to help promote recovery and resilience of those locations.

The survey comprised a demographic questionnaire, a thematic questionnaire on local environmental knowledge, and a "spray and say" questionnaire with fuzzy tagging of places and subquestions about meanings attached to those places (Appendix A). Data collection was carried out with the Map-Me ("Mapping Meanings") PPGIS tool. A distinguishing feature of Map-Me is its ability to document vague locations of geographical features by a fuzzy tagger (the "spray-can" tool) following the initial approach of Waters and Evans (2003) and Evans and Waters (2007) for raster grids, and more recently Huck *et al*. (2013) for vector models, which enables a suitable means to take account of the vagueness of people's perceptions on expressing spatial features of their environment (Fig. 4).

Twenty-eight respondents provided *spray patterns* in response to the question "Indicate an area that you believe has changed over the years" (Fig. 5). Spray patterns were then reprojected to the Coordinate Reference System "NAD83 (NSRS2007) / Montana" (EPSG: 3604) at the beginning of the data analysis stage. In order to standardize respondents' behaviour on tagging space, a frequency model deriving from overlapping rasterized spray patterns were used instead of the raw multipoint objects. In this model, for each individual spray pattern the algorithm rasterizes the raw multipoint distribution as an *indicator surface,* such that each cell of the surface is given the value of 1 if one or more points occurs in it, and 0 otherwise. Indicator surfaces from all spray responses are next added and a *relative frequency surface* is computed with regard to how frequently every cell is sprayed, by the transformation *Cell value = Number of overlapping indicator surfaces tagging the cell / Total number of spray patterns in the sample*.

The maximum frequency of overlapping spray patterns in a given cell is 36 % of the total number of spray patterns (Fig. 6). Spots where highest frequencies of respondents agree on having observed environmental changes are found along the Jocko basin and in a primitive area of the Mission Mountains, in the eastern part of the Jocko landscape unit. At present, responses are also being processed by the multimodal software implementation here presented, with the goal of grouping and mapping spray patterns according to the semantic properties of their attached comments. In relation to the query `select features where use of prescribed fire is good`, 31 % of spray patterns support the use of prescribed fire in tagged spots, with the maximum frequency of overlapping spray patterns being 12 % of all spray patterns (Fig. 7). Sites most frequently tagged by non-tribal people stretch along the Middle Jocko and in herbaceous lands between the basin and higher woodlands, whereas spots most frequently tagged by tribal members concentrate in the mountain woodlands and on summits of the eastern primitive mountain area.

7. Discussion and conclusions

An initial conclusion from Fig. 7 is that, if demographic attributes are disregarded, a subsample of both tribal and non-tribal residents agree on a common idea, that restoring prescribed fire in the Jocko landscape unit would have a positive effect on local environments. However, internal differences emerge when geospatial and demographic properties are jointly taken into account. Clearly, opinions differ about where prescribed fire should be introduced. Besides, these seem to be somehow linked to cultural identity based on tribal membership.

A comprehensive interpretation of this pattern still needs further exploratory analysis, nevertheless, as well as testing of spatial association hypotheses. At least two questions should be examined: (i) what is the probability of the observed associations between spray patterns and geographical location to be the product of chance, and (ii) what substratum covariates may explain the observed relative frequencies at the local scale. Candidate covariates might derive from variables about land use, the evolution of fire regimes and the recent history of land status. In addition, methods of analysis at the inferential level could consist (a) in geosimulation, replicating the use of the fuzzy tagger by respondents according to some spatial association hypothesis (e.g. testing that tagging behaviour relates to specific landscape attributes), (b) in resampling the assignment of spray patterns to cultural labels (i.e. to respondents) by bootstrapping-based techniques, and (c) in the fitting of spray patterns to spatial regression models.

Even though this may not be the first time that coupling verbalized expressions with geospatial and demographic properties enhances information outputs (if compared to taking these components separately), it confirms the perception that treating and analyzing multimodal phenomena does involve a significant leap with regard to reductive perspectives. The “spray and say” approach enables a more inclusive collection of people's spatial perceptions and ways of expression, but it is the processing of multimodal datasets that allows an holistic treatment of multiple modes and dimensions that objects may take. Ultimately, this processing would improve with comprehensive multimodal systems able to couple a large number of modes simultaneously (e.g. spatiotemporal, verbal, pictorial and measure-based, to name the most evident ones), clearing the path towards more versatile models of cognition.

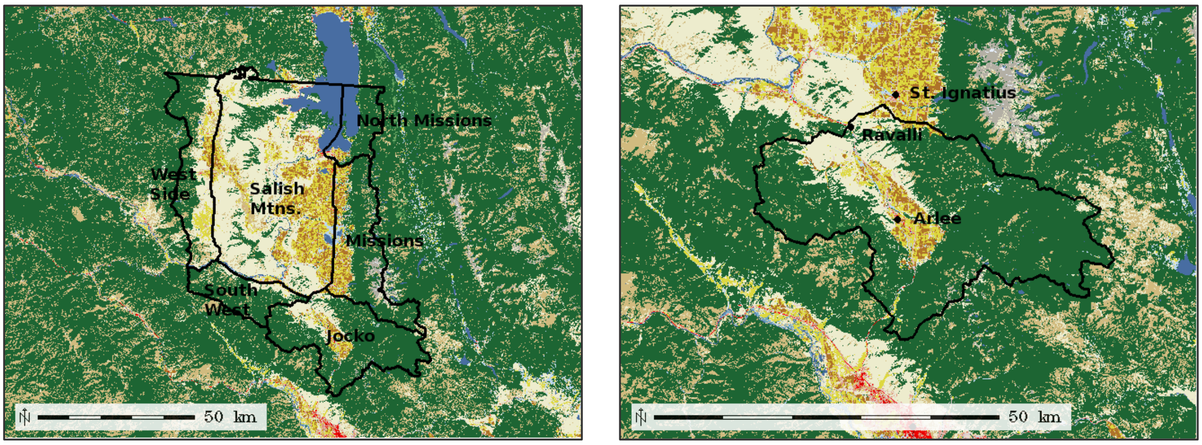


Fig. 3. The Flathead Reservation (left) and the Jocko Landscape Unit (right).

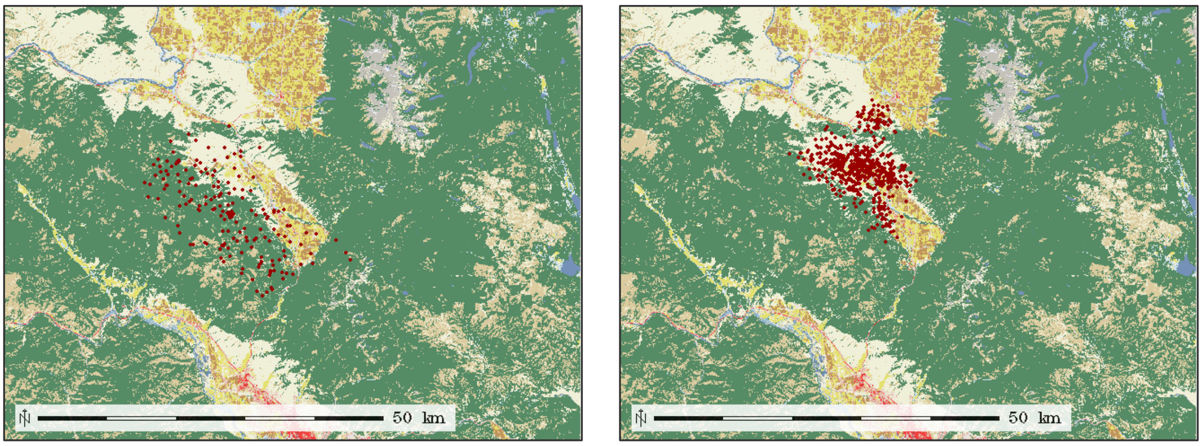


Fig. 4. Use of the fuzzy tagger. Each of these two examples of a spray pattern comprises a whole 'coat of spray paint' drawn by a different respondent with regard to a common question, so they each link to a single text object (in this case, a sequence of comments provided by the respondent) that qualifies the full coat.

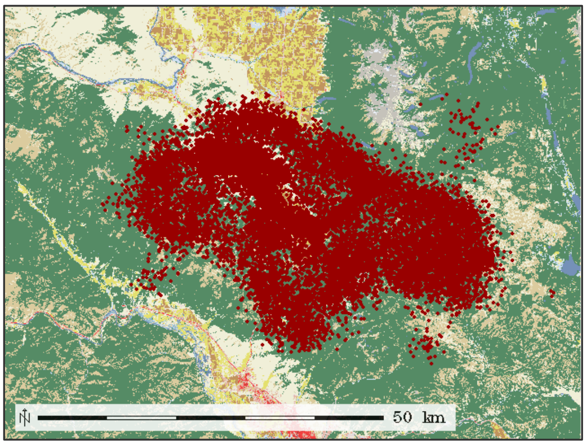
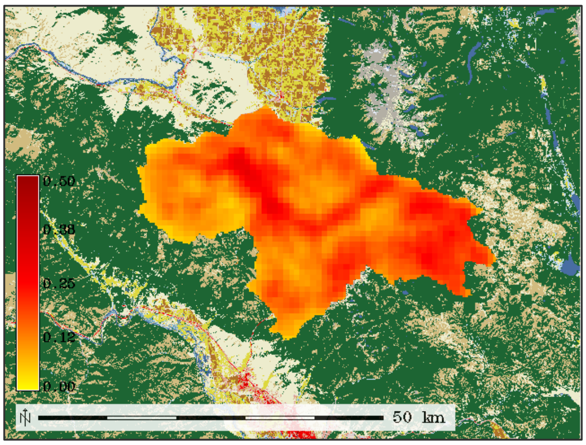


Fig. 5. Joint mapping of all spray patterns tagging observed environmental changes. *N*(spray patterns) = 28. *N*(dots) > 20,000.



Fig, 6. Heat map of Fig. 5, as the relative frequency that a location is tagged as a spot having undergone environmental changes.

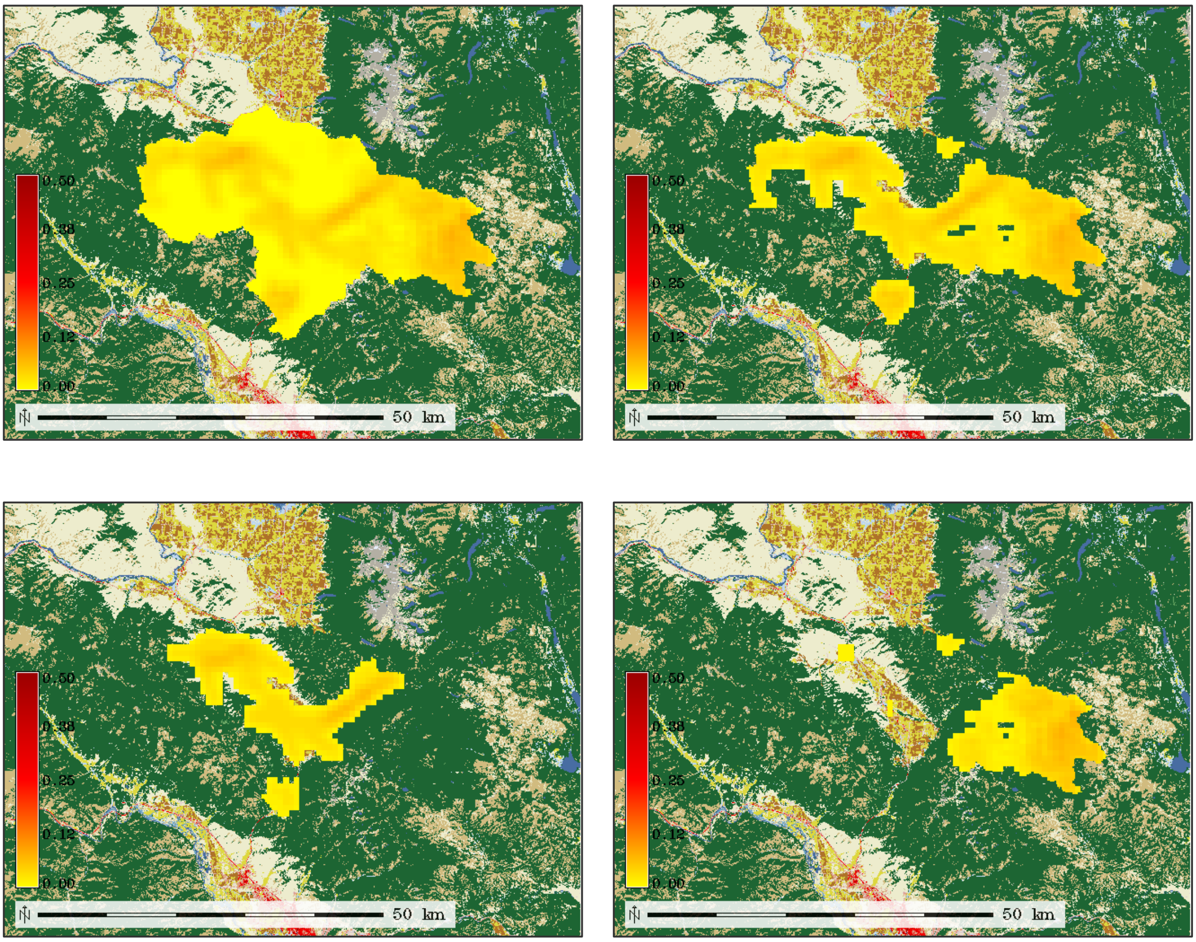


Fig. 7. Relative frequency that tagged locations contain comments supporting *in situ* use of prescribed fire. Upper left: all places. Upper right: places where the frequency is higher than 0. Lower left: answers provided by non-tribal residents. Lower right: answers provided by tribal residents.

REFERENCES

Carver, S.; Evans, A.; Kingston, R. and Turton, I. (2001). Public participation, GIS and cyberdemocracy: evaluating on-line spatial decision support systems, *Environment and Planning B*, 28(6): 907-921.

Carver, S.; Watson, A.; Waters, T.; Matt, R.; Gunderson, K. and Davis, B. (2009). Developing Computer-Based Participatory Approaches to Mapping Landscape Values for Landscape and Resource Management (2009). In S. Geertman, and J. Stillwell (eds.), *Planning Support Systems Best Practice and New Methods*, pp. 431-448. Springer.

Chowdhury, G. (2003). Natural language processing. *Annual Review of Information Science and Technology*, 37: 51-89.

Confederated Salish and Kootenai Tribes (CSKT) (2005). *Ecosystem Management on the Flathead Indian Reservation*. Confederated Salish and Kootenai Tribes Forestry Department: Pablo, MT.

Confederated Salish and Kootenai Tribes of the Flathead Reservation (CSKT) (2013). *Climate Change Strategic Plan*. Confederated Salish and Kootenai Tribes: Pablo, MT.

Evans, A.J. and Waters, T. (2007) Mapping vernacular geography: web-based GIS tools for capturing “fuzzy” or “vague” entities, *International Journal of Technology, Policy and Management*, 7(2): 134-150.

Field, C.B.; Barros, V.; Stocker,T.F.; Qin, D.; Dokken, D.J.; Ebi, K.L.; Mastrandrea, M.D.; Mach, K.J.; Plattner, G.-K.; Allen, S.K.; Tignor, M. and Midgley, P.M. (2012). *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation*. Cambridge University Press: Cambridge.

Gunderson, K. (2006). Understanding place meanings for wilderness: personal and community values at risk, *International Journal of Wilderness*, 12(1): 27-31.

Gunderson, K. and Watson, A. (2007). Understanding place meanings on the Bitterroot National Forest, Montana, *Society and Natural Resources*, 20: 705-721.

Huck, J.J.; Whyatt, J.D.; Yielding, S.; Stanford, H. and Coulton, P. (2013). Development and application of a “spray-can” tool for fuzzy geographical analysis. *21st GIS Research UK Conference*.

Jewitt, C. (ed.) (2009). *The Routledge Handbook of Multimodal Analysis*. London: Routledge.

Joshi, A. K. (1991). Natural Language Processing. *Science*, 253 (5025): 1242-1249.

Jurafsky, D. and Martin, J. H. (2008). *Speech and Language Processing, 2nd edition*. Pearson-Prentice Hall: Upper Saddle River, NJ.

Kress, G. (2010). *Multimodality: A Social Semiotic Approach to Contemporary Communication*. New York: Routledge.

Kingston, R. (2007). Public participation in local policy decision-making: the role of web-based mapping. *The Cartographic Journal*, 44(2): 138-144.

Manning, C. D. and Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*. The MIT Press: Massachusetts.

Metz, B.; Davidson, O.R.; Bosch, P.R.; Dave, R. and Meyer, L.A. (2007). *Climate Change 2007: Working Group III: Mitigation of Climate Change. Contribution of Working Group III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press: Cambridge and New York.

Parry, M.L.; Canziani, O.F.; Palutikof, J.P.; van der Linden, P.J. and Hanson, C.E. (2007). *Climate Change 2007: Working Group II: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press: Cambridge and New York.

Pickles, J., ed. (1995). *Ground Truth: The social implications of geographic information systems*. Guilford: New York.

Solomon, S.; Qin, D.; Manning, M.; Chen, Z.; Marquis, M.; Averyt, K.B.; Tignor, M. and Miller H.L. (2007). *Climate Change 2007: Working Group I: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* . Cambridge University Press: Cambridge and New York.

Veres, S. M. (2008). *Natural Language Programming of Agents and Robotic Devices: publishing for agents and humans in sEnglish.* SysBrain: London.

Waters, T. and Evans, A.J. (2003). Tools for web-based GIS mapping of a "fuzzy" vernacular geography. *Proceedings of the 7th International Conference on GeoComputation.* [Available at http://geocomputation.org/2003]

Appendix A. Questionnaire of the PPGIS survey on the Flathead Reservation

Demographic questions

- Are you an enrolled member of the Confederated Salish and Kootenai Tribes (CSKT)?
- What is your age group?
- What is your gender?
- What is your occupation?
- If retired, what was your occupation?
- What is your role in Tribal affairs (e.g. official position, committee participation, Elder)?

Traditional Ecological Knowledge questions

- How familiar are you with the Jocko Landscape?
- How worried are you about climate change?
- How much do you think climate change will affect native plants in the Jocko Landscape (such as from rivers flooding, wildfires, droughts, damage from bark beetles, temperature increases, etc.)?
- How do you feel, if at all, that traditional knowledge can help solve these impacts or problems for this area? Please give an example.
- How much do you think climate change will affect wild animals in the Jocko Landscape (such as rivers flooding, wildfires, droughts, damage from bark beetles, temperature increases, etc.)?
- How do you feel, if at all, that traditional knowledge can help solve these impacts or problems? Please give an example.

- How much do you think climate change will affect fisheries in the Jocko Landscape (such as rivers flooding, wildfires, droughts, damage from bark beetles, temperature increases, etc.)?
- How do you feel, if at all, that traditional knowledge can help solve these impacts or problems? Please give an example.
- How much do you think climate change will affect crops in the Jocko Landscape (such as rivers flooding, wildfires, droughts, damage from bark beetles, temperature increases, etc.)?
- How do you feel that, if at all, that traditional knowledge can help solve these impacts or problems? Please give an example.
- How much do you think climate change will affect livestock in the Jocko Landscape (such as rivers flooding, wildfires, droughts, damage from bark beetles, temperature increases, etc.)?
- How do you feel, if at all, that traditional knowledge can help solve these impacts or problems? Please give an example.
- How much do you think climate change will affect outdoor recreation places (such as parks, beaches, lakes, rivers, forests, etc.) in the Jocko Landscape (such as rivers flooding, wildfires, droughts, damage from bark beetles, temperature increases, etc.)?
- How do you feel, if at all, that traditional knowledge can help solve these impacts or problems? Please give an example.
- How much do you think, if at all, climate change will affect water quality, quantity or location in the Jocko Landscape (such as rivers flooding, wildfires, droughts, damage from bark beetles, temperature increases, etc.)?
- How do you feel, if at all, that traditional knowledge can help solve these impacts or problems? Please give an example.
- How much do you think climate change will affect trees, shrubs, and grasses in the Jocko Landscape (such as rivers flooding, wildfires, droughts, damage from bark beetles, temperature increases, etc.)?
- How do you feel, if at all, that traditional knowledge can help solve these impacts or problems? Please give an example.

- How much do you think climate change will influence wildland fires in the Jocko Landscape (such as rivers flooding, wildfires, droughts, damage from bark beetles, temperature increases, etc.)?
- How do you feel, if at all, that traditional knowledge can help solve these impacts or problems? Please give an example.
- How well do you feel that these questions have allowed you to express your beliefs about climate change?

"Spray and say" questions

Tag question:

- The area outlined in red is the Jocko Landscape. Please indicate an area that you believe has changed over the years.

Subquestions about tagged spots:

- What did this area used to be like and what is the source of your knowledge?
- What is the area like now and what do you believe has caused the change from what it used to be like?
- What would you like this area to be like in the future and why?
- What actions need to be taken and what do you believe will be the primary obstacle to achieving this end state?