



Jul 1st, 12:00 AM

Spatio-temporal modelling of the spread of common reed on the Finnish coast

Anas Altartouri

Ari Jolma

Päivi Korpinen

Follow this and additional works at: <https://scholarsarchive.byu.edu/iemssconference>

Altartouri, Anas; Jolma, Ari; and Korpinen, Päivi, "Spatio-temporal modelling of the spread of common reed on the Finnish coast" (2010). *International Congress on Environmental Modelling and Software*. 297.
<https://scholarsarchive.byu.edu/iemssconference/2010/all/297>

This Event is brought to you for free and open access by the Civil and Environmental Engineering at BYU ScholarsArchive. It has been accepted for inclusion in International Congress on Environmental Modelling and Software by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.

Spatio-temporal modelling of the spread of common reed on the Finnish coast

Anas Altartouri^a, Ari Jolma^a, Päivi Korpinen^b

^a Lahti Center, Department of Civil and Environmental Engineering, School of Science and Technology, Aalto University, Finland

^b Finnish Environment Institute (SYKE), Finland,

Email addresses: anas.altartouri@tkk.fi, ari.jolma@tkk.fi, paivi.korpinen@ymparisto.fi

Abstract: The common reed (*Phragmites Australis*) has overtaken the Finnish coast of the Gulf of Finland in many places influencing the coastal ecology and reducing the recreational value of the coast. Our research aims at developing a spatio-temporal simulation model of the spread of reed. In order to account for the dynamic nature of the spread, a cellular automata model is employed. We investigate in depth the explanatory variables and their relevance to the phenomenon. It was found that water depth, proximity to river mouths, and sea openness are significant explanatory variables of the phenomenon. A predictive logistic regression model is suggested to aggregate the explanatory variables and provide transition rules for the cellular automata. The results of the regression model are presented and discussed, raising the necessity of incorporating Bayesian inference in the logistic regression model. Therefore, more insightful likelihood functions for the explanatory variables are provided and discussed.

Keywords: Bayesian inference; Cellular Automata; Gulf of Finland; Logistic regression; *Phragmites australis*

1. INTRODUCTION

The common reed (*Phragmites australis*) has overtaken the Finnish coast of the Gulf of Finland in many places causing substantial impact on the market value of coastal and archipelago properties. Reed invasion is thought to be an important factor influencing the reproduction success of both birds and fish. It reduces biodiversity and alters habitat [Minchinton, 2002], favouring certain species and negatively influencing the abundance of others. During the past decades, the common reed has spread extensively in the inner and middle archipelago areas in Finland covering now up to 8% of the area of some coastal municipalities in the Southern Finland [Pitkänen, 2006]. This makes the phenomenon of reed spread linked to three of the major environmental concerns, namely habitat change, biodiversity loss, and invasive species [Gewin, 2005]. However, our ability to quantify the consequences of such a phenomenon is restricted by the absence of adequate models for predicting and simulation it.

The present paper aims at providing a conceptual framework for spatio-temporal modelling of the spread of reed. It approaches the matter by combining different methods, including cellular automata (CA), to account for the dynamic nature of the phenomenon; logistic regression, to aggregate different environmental explanatory variables; and Bayesian inference, to incorporate our knowledge about the matter and overcome the limitations of pure statistical modelling toward more credible simulation.

Several CA models have been recently developed for a number of applications, including urban planning and growth [Kocabas and Dragicevic, 2007; Vliet *et al.*, 2009], hydrology

[Parsons and Fonstad, 2007], forest dynamics [Lett *et al.*, 1999], and epidemic propagation [Sirakoulis *et al.*, 2000]. However, the potential of the CA model in simulating the spread of invasive plants is poorly investigated [Huang *et al.*, 2008]. From another perspective, there has been effort to enhance the CA transition rules by combining other approaches, such as multi-criteria evaluation [Wu and Webster, 1998], principal component analysis [Li and Yeh, 2002], neural networks [Yeh and Li, 2003], and Bayesian Networks [Kocabas and Dragicevic, 2007]. Nonetheless, the potential of Bayesian logistic regression for providing CA transition rules is poorly explored.

We start with a short description of reed ecology (section 2) in order to understand in depth the reed behaviour and, therefore, propose explanatory environmental variables causally linked to the phenomenon. Then, section 3 provides description of the study area followed by description of the methods used (section 4). Afterwards, we present and discuss in section 5 the results of the analyses conducted, and propose the enhancement of the logistic regression model by incorporating Bayesian inference. Finally, we wrap up with some concluding remarks (section 6).

2. REVIEW OF REED ECOLOGY

Phragmites australis is a perennial vascular plant with high shoots of about 4 m. The common reed forms usually very dense stands (100-300 shoots per 1 m²) which are capable of fast spreading when conditions are favourable. Reed is one of the most spreading vascular plants globally. It can spread both with rhizomes and with seeds. However, reproduction by seeds is poor, and most of the spreading happens with rhizomes [Haslam, 1972; Bart and Hartman, 2003]. The reed beds have several competitive advantages against other plants which makes the aggressive spreading possible.

The global spread can be explained with the high tolerance of reed to different environmental conditions, making it a robust competitor in several ways [Burdick and Konisky, 2003]. It prefers nutrient-rich habitats; hence, eutrophication can be one explanation for the expansion of reed. Reed tolerates high salinity by placing its vital roots in fresh waters or deep in less saline waters where groundwater seepage occurs [Adams and Bate, 1999]. Thus, salinity cannot be accounted for as explanation of the reed spread, particularly in the Gulf of Finland where water is classified as brackish. Furthermore, reed has a high tolerance to water level fluctuations [Haslam, 1972; Lelong *et al.*, 2007]. Human impacts provide a partial explanation of how *Phragmites* competes so well. Hydrological alterations and construction activities can explain the spread of reed [Bart and Hartman, 2003; King *et al.* 2007]. There is evidence that variations in climatic conditions, particularly increased precipitation enhance the performance of reed [Minchinton, 2002].

3. STUDY AREA AND DATA

Our study area is the Finnish coast of the GOF extended to the SW-Finnish archipelago. With approximately 25000 islands larger than 500 m² [Tolvanen and Sominen, 2005], the area is considered one of the most island-rich archipelagos in the world [Ekeboom *et al.*, 2003]. The coastline of the area is extremely complex; the lengths of the coastline and shores of all the islands is about 8200 km. The average water depth in the area is 20 m. The area is characterised by brackish nature with low salinity, ranging from 5-6.5 PSU in the west of the GOF to 0-3 PSU at its easternmost end [Soomere *et al.*, 2008].

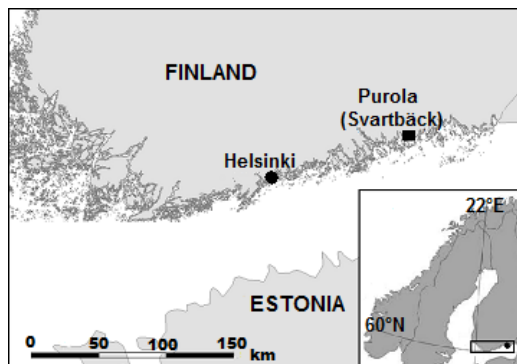


Figure 1. Study area

The analysis was carried out on a site of about 50 km² in Purola (Svartbäck) near to the outlet of River Kymijoki (one of the major rivers flowing into the GOF) at Ruotsinpyhtää (see Figure 1). Available datasets include (a) a digital elevation model (DEM) of the main land and islands along the study area with spatial resolution of 25x25m cell; (b) a digital depth model (DDM) of the seabed with spatial resolution of 100x100m cell; (c) reed coverage of the whole area in 2001; and (d) temporal reed coverage for Purola (Svartbäck) site from 2003 and 2006.

4. METHODS

In order to account for the dynamic nature of the phenomenon of reed spread, CA method was selected. Cellular Automaton is a discrete model where space is represented by a grid of cells with a finite number of states [Fonstad, 2006]. An initial state configuration is generated, then the progression of a cell from one generation to the next is determined based on (a) its initial state, (b) the initial states of its neighbourhood, and (c) a series of transition rules. The function based on which CA work can be written as follows:

$$S_{ij}(t+1) = f(N_{ij}(t), S_{ij}(t), T) \quad (1)$$

where $S_{ij}(t+1)$ denotes the state of the cell in question in the next generation, $S_{ij}(t)$ denotes the cell state at the initial time, $N_{ij}(t)$ denotes the neighbourhood state at the initial time, and T is a set of transition rules. A flow chart of the proposed CA model is illustrated in Figure 2. An initial state of a cell at time t is configured. Then, a probability vector of each possible state in the next generation is computed based on the analyses of its suitability for reed growing and the initial states of its neighbouring cells. In order to account for the random nature of the phenomenon, Monte Carlo method is used. It generates a random number to which the computed probability is compared and assigned either 0 or 1 (absence or presence of reed, respectively), representing the cell state in the next generation. Applying Monte Carlo method also allows us building different scenarios of the coverage of reed in the future generations.

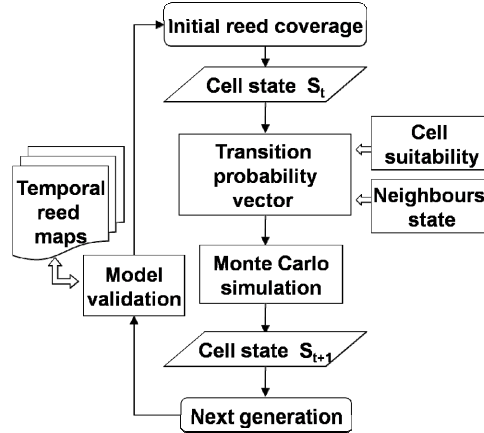


Figure 2. Flow chart of CA model

The determination of the transition rules is crucial to the CA model. The analyses needed for this purpose are in two folds; analysis of the suitability of a given cell for reed growing given the explanatory variables, and analysis of the influence of the neighbourhood of a cell on its future state.

Neighbourhood analysis aims at understanding the way by which an already reed-occupied area would influence its neighbourhood. Reproduction of reed by seeds is poor and most of the spreading happens with rhizomes. Hence, we assume that reed will smoothly expand into areas where the majority of neighbouring cells are reed-occupied. We used temporal reed coverage data from the year 2003 and 2006. Based on the comparison between their states in 2003 and 2006, cells within the study site were classified into four categories; cells were reed-free and became reed-occupied (reed growth), cells were reed-occupied and became reed-free (reed death), cells occupied by reed at both time steps, and finally, cells free of reed at both time steps. For every cell where the state has changed (reed has grown or died), the number of reed-occupied neighbours in 2003 was counted, indicating its neighbourhood state. Then, we combined information of the neighbourhood state of a

given location with its state change (growth or death) in order to provide insightful reflection their association. The analysis steps are illustrated in Figure 3.

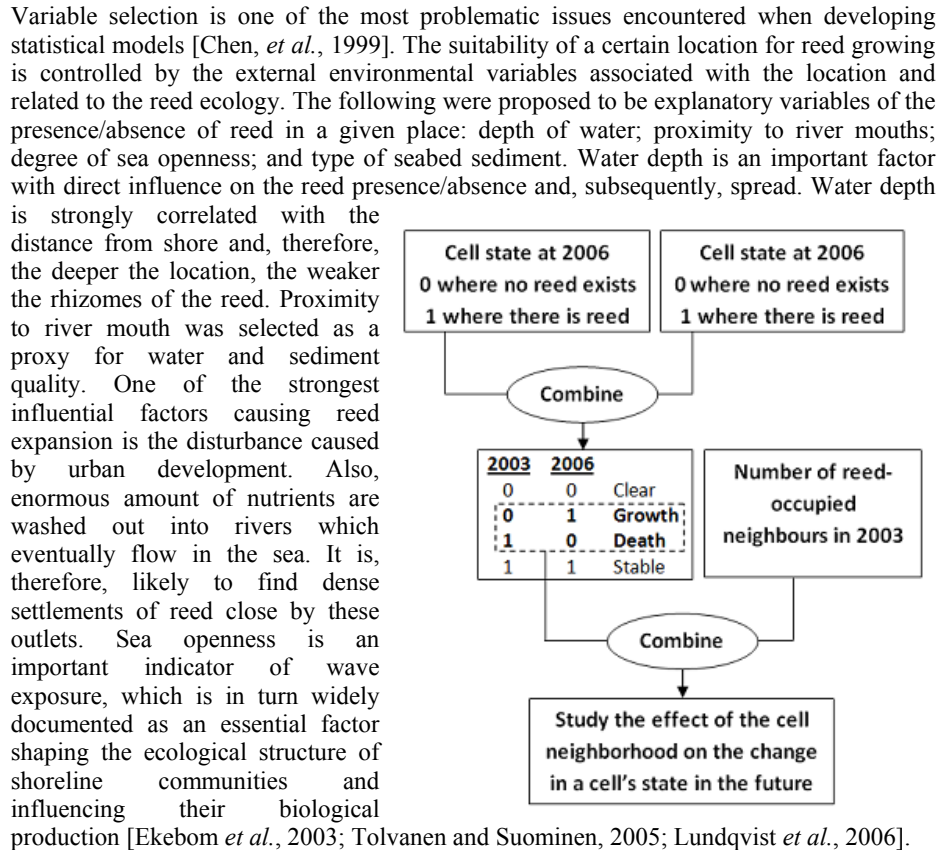


Figure 3. Neighbourhood analysis

The Euclidian distance was calculated from each cell to the nearest defined river mouth. With regard to the openness analysis, a number of methods were developed to calculate the wave exposure and fetch. These methods estimate openness based on different input parameters, ranging from fetch and abstract wind data [Lundqvist *et al.*, 2006; Pepper and Puotinen, 2009] to more detailed data about wave exposure [Tolvanen and Suominen, 2005] and wave power [Ekebom *et al.*, 2003]. However, openness can be satisfactorily estimated by calculating fetch, the stretch of water surface over which generated waves can develop freely. The method is based on summarising measured fetch lines around each point to indicate the openness of that point (see Figure 4). In order to test the correlation between the explanatory variables and the reed presence/absence data was represented in a gridded layers; a binary layer of the reed coverage, where a value of one is assigned for reed-occupied cells and zero elsewhere; a DDM representing water depth (in meters); a layer representing the distance (in meters) of each cell to the nearest river mouth; and finally, an openness layer obtained by the summation of the total length of the fetch lines originated from the corresponding cell. No sufficient data on seabed sediment was available.

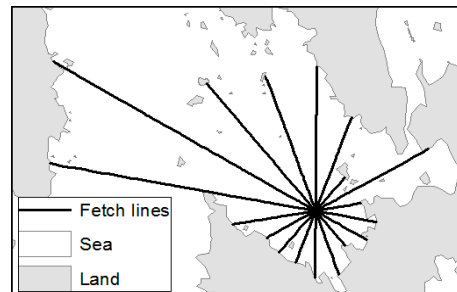


Figure 4. An example of fetch calculation

The correlation was tested using two approaches. Firstly, the correlation between the layer of reed coverage (values are either one or zero) and the layer representing each explanatory variable (continuous data with different range for each variable) was tested. As the two variables being correlated are dichotomous (reed presence/absence) and continuous (the explanatory variables), we used the Point Biserial correlation coefficient [Edwards, 1976]. The second approach to testing correlation considers the intensity of reed in relation to each explanatory variable of a given location. The term ‘reed intensity’ refers to the size of the reed settlement around a given reed-occupied cell. This approach was used to distinguish areas where the presence of reed is a pervasive phenomenon from those where it is a random occurrence. Practically, this layer was obtained using a focal function which sums the total number of reed-occupied cells within a window of 500x500m around the reed-occupied cell in question.

Logistic regression analysis was employed in order to establish relationships between our variables, and express them succinctly. Two parts can be distinguished in a regression analysis; the dependent variable representing the phenomenon (whose presence in a location is being investigated the reed existence in our case); and the independent variables, which are the attributes of the locations that influence the presence of the phenomenon, *i.e.* the explanatory variables. Logistic regression is used since we already know sites where reed exists (dichotomous response variable) and attempt to determine why it could be found in those sites, and consequently be able to predict it at an un-sampled location.

5. RESULTS AND DISCUSSION

5.1 Neighbourhood influence on the spread of reed

Results of the neighbourhood analysis point out strong influence of the neighbourhood state of a given location on the likelihood of reed spreading into that location. Cells with a majority of neighbours (> 4) being reed-occupied tend to have reed expanded into them in the future. Table 1 summarises the results of the neighbourhood analysis carried out on the temporal reed coverage data from the year 2003 and 2006. A total number of 439208 cells had changed their states between 2003 and 2006. More than two-third (68%) of those cells were reed-free in 2003 and became reed-occupied in 2006. On the other hand, 32% of the cells were covered by reed in 2003 and became clear in 2006. It should be emphasised here that number of reed-occupied neighbours in 2003 was the distinction between the two groups; of the total number of cells where reed had expanded by 2006, 96.9% had more than 4 neighbouring cells already occupied by reed in 2003. On the contrary, 99.7% of those cells where reed died had less than four neighbouring cells occupied by reed in 2003.

Table 1. Reed growth/death between 2003 and 2006

A total of 439208 cells were analysed between 2003 and 2006	
<u>32% died:</u>	<u>68% grew:</u>
99.7% had <=4 cells occupied by reed in 2003	3.1% had <=4 cells occupied by reed in 2003
0.3% had >4 cells occupied by reed in 2003	96.9% had >4 cells occupied by reed in 2003

Table 2. Percentage of reed growth/death for every neighbourhood case

In a more detailed view, Table 2 presents the percentage of cells which changed their states by 2006 (*i.e.* became reed-occupied or reed-free), categorised by the number of reed-occupied neighbouring cells in 2003. It can be confidently said that a cell surrounded by reed from three sides (that is, having five or more reed-occupied neighbours) will become reed-occupied in the future. On the contrary, reed tends to be cleared out from locations where the majority of surrounding cells are reed-free. These results support the assumption that rhizomes are the main way by which reed expands. It can also be concluded that the presence of large reed settlements in a location indicates the suitability of the surrounding areas for reed growing.

# of reed-occupied neighbouring cells in 2003	% of cells died by 2006	% of cells grew by 2006	Total # of cells
0	99.999	0.003	118662
1	99.30	0.70	5860
2	95.21	4.788	4156
3	92.92	7.08	8124
4	33.18	66.82	12774
5	1.25	98.75	16542
6	1.44	98.56	8269
7	0.16	99.84	12402
8	0.001	99.999	252419

5.2 Explanatory variables of the spread of reed

The results indicate correlation (in different levels and types) between the presence/absence of reed in a location and the suggested explanatory variables associated with that location. The correlation coefficients for the water depth, proximity to river mouths, and openness were found to be 0.22, -0.31, and -0.33, respectively. Stronger correlation resulted, however, using the second approach which takes into account the intensity of reed rather than merely its presence or absence. These were 0.26, -0.55, and -0.45 for the water depth, proximity to river mouth, and openness, respectively.

The correlation between the phenomenon and the water depth was found to be positive indicating that the deeper the water in a location, the less chance of reed to grow there. This was, however, a low correlation (0.22 and 0.26 for both approaches) and the reason might be the roughness of the DDM used (100x100m cell), which abstracts a range of depth values into their average and, subsequently, makes it difficult to detect a trend of the behaviour of reed at different depth values. Regarding the proximity to river mouths, the results show strong negative correlation. Hence, the closer a location to a river mouth, the more suitable it is for reed growing. That is, chances for reed to grow in areas distant from a river mouth become smaller than in those close to it. The correlation coefficient calculated using the second approach (-0.55) suggests that reed settlements tend to be large in locations close to a certain river mouth. More detailed discussion about this explanatory variable can be found in section 5.3. Similarly, strong correlation was found between the phenomenon and the openness of a location. With -0.33 and -0.45 correlation coefficients (for both approaches), the results indicate that relatively open sea is a non-suitable environment for reed growing. This implies, on the contrary, that relatively close shores and inlets are more susceptible to reed occupation. This factor is of importance in our case given the complex nature of the archipelago's shoreline.

5.3 Logistic regression model

Ordinary Least Squares method was run to obtain parameters of the regression model. Input data was the water depth, proximity to river mouths, and the openness index associated with every location where reed exists. An assessment of the overall model performance, R^2 was found to be 20%. Although this does not reflect high performance, it is important to consider other diagnostic tests for more understanding of the model performance. It is important here to point out that the obtained value of R^2 does not necessarily contradict with the abovementioned results which indicated correlation between the explanatory variables and the spread of reed. It rather suggests that these explanatory variables do not provide a complete representation of the phenomenon and, therefore, more explanatory variables would result in a better prediction model. In order to assess the

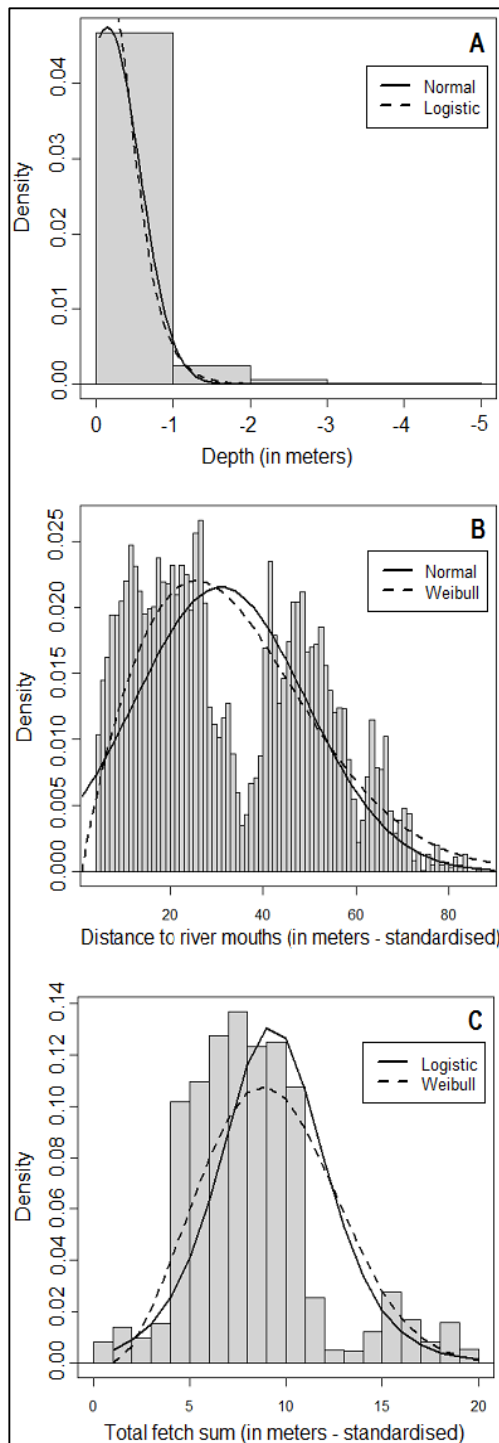
statistical significance of the explanatory variables for the model, student's t test was consulted, indicating that the variables were significant ($P < 0.05$). Additionally, the Variance of Inflation Factor (FIV) was used to test redundancy in the explanatory variables. Results of this test (all < 1.5) did not imply problems of collinearity among the explanatory variables. Furthermore, Joint Wald statistics indicates that the model was overall significant ($P < 0.05$), that is, the explanatory variables were effective. The resultant equation of reed prediction can be written as the following:

$$y = \frac{1}{1 + \exp\left\{-\left(1.07828 + 0.11348x_1\right) - 0.000126x_2 - 0.000002x_3\right\}} \quad (2)$$

where y is the prediction of reed existence (ranges from 0 to 1, indicating very low and very high existence probability, respectively), x_1 denotes the water depth (m), x_2 denotes the distance the river mouth (m), and x_3 denotes the sum of fetch lines (m). In order to assess the reliability of the prediction formula, it was applied on another two sites of the study area where the coverage of reed is already known. The results show low correspondence between the predicted reed coverage and the reality. It was noticed that the model yielded a prediction values range from 0.3 to 0.7; no lower or higher prediction values resulted from the formula.

Whilst there was correlation found between the reed existence and the explanatory variables, there still have been outliers where reed does exist in a non-favourable location. These values negatively influence the prediction model, which is subsequently no more able to detect a trend of reed existence. Moreover, the model expresses the relationship between the phenomenon and the explanatory variables linearly which might be inaccurate. It is essential to provide a detailed expression about the influence of each variable on the reed existence taking into account different kind of relationships at different ranges of the variable values. The heavy reliance of the logistic regression on data raise up the need for adjusting the relationships according to our belief adopted from the understanding of reed ecology. According to Austin [2002, pp. 82], “*statistical modelling cannot substitute for ecological insight, appropriate environmental gradients, and knowledge of the process linking environment with biota*”. This insight, however, can be achieved by Bayes’ Theorem which incorporates prior knowledge about the phenomenon, coupled with the likelihood of the data, to measure the degree of belief about the model parameters [Ellison, 2004].

The benefits of Bayesian statistics for ecological modelling are well established [Ellison, 2004] as it allows the utilization of prior knowledge about



the parameters of the logistic regression [Accad *et al.*, 2005; Al-Awadi and Garthwaite, 2006; Denham and Mengersen, 2007; Xu and Akella, 2008; Low Choy *et al.*, 2009, James *et al.*, 2010]. Usual pattern of Bayesian analyses [Ellison, 2004; Gelman *et al.*, 2004] consists of eliciting prior distributions, formulating the likelihood functions, and, using Bayes' Theorem, quantifying posterior probability distributions of the regression model. We present here possible likelihood functions as response curves showing the influence of each environmental variable on the probability of success, that is, probability of reed existence. These functions are then adjusted by incorporating those elicited from expert knowledge. Below we discuss the likelihood function for the explanatory variables. However, prior elicitation from expert knowledge is not addressed in this study.

Figure 5. Histograms and PDFs of the explanatory variables

In order to obtain likelihood functions of each explanatory variable, we need to find the distribution function that best fits each variable. It is helpful to start with graphical exploration of data. Therefore, histograms were drawn where x-axes represent the range of values corresponding to the variable in question, and y-axes represent the occurrence frequency (converted later on to probability density) of these values. Figure 5 (A, B, and C) shows the obtained histograms for the water depth, proximity to river mouths, and total fetch, respectively. It should be noticed that in the case of proximity to river mouth and total fetch distance, x-axes represent the normalized values of these covariates. Another issue to be pointed out here is that frequencies represent the number of occurrence of the corresponding variable values only in those locations where reed exists.

Histograms can provide insights on the behaviour of the dependent variable at different ranges of covariate values. This can clearly be noticed in the variable representing the proximity to river mouths; when testing the correlation of this variable's values with the data of reed existence, the result indicated a strong negative correlation (-0.55). This result might lead to the conclusion that the closer a location to a river mouth, the more suitable it is for reed. However, looking at the histogram of this variable (see Figure 5B), a slightly different, and more detailed, image is revealed; it can be noticed that the previous conclusion is relevant only on the range of the covariate values beyond a certain buffer around a river mouth. That is, locations right at a river mouth or very close to it tend to have very little reed coverage. Only after a certain distance, however, the statement that closer locations to a given river mouth tend to have dense reed coverage becomes valid. In fact, one can see that such a conclusion is sensible as sea water at the river outlets, and a certain buffer around them, is subjected to the dynamicity of the river flow and, therefore, considered less suitable for reed growing.

Having obtained the histogram for each variable, it is possible to find a mathematical function which represents in a good way each variable, *i.e.* a fitting probability distribution. Using R statistical package [R Development Core Team, 2009], different continuous distributions were fitted on the histograms, including Normal (μ , σ), Logistic (μ , σ), and Weibull (α , β , γ). The histogram and two alternative probability density functions (PDFs) of each covariate are illustrated in Figure 5 (A, B, and C). The goodness-of-fit of each PDF was tested using Kolmogorov-Smirnov test indicating adequate fitting ($P > 0.05$). It should be mentioned here that it is out of the scope of this study to investigate all alternatives of the PDFs and find the best fitting for each variable.

6. CONCLUDING REMARKS

Modelling of environmental phenomena is a complicated task due to their complexity and uncertainty, making it challenging to develop realistic and credible models. The study presented a framework for a simulation model of the spread of reed, a phenomenon causing negative impacts on the environment. Depth of water, proximity to river mouths, and degree of openness were found to be significant explanatory variables of the reed spread. However, additional variables can enhance the model performance. It was also found that the state of neighbourhood of a location highly influences its future state in terms of reed

coverage, which provides strong explanation for the model. For this reason, as well as to adapt to the dynamic nature of the phenomenon, CA model is found to be of high potential for simulating the spread of reed. Logistic regression can provide the transition rules for the CA. Nonetheless, in order to enhance these rules and cope with the limitations of the logistic regression model, Bayesian inference can be employed. This involves combining the likelihood obtained from data, and prior distribution elicited from expert knowledge, in order to determine the posterior distribution of each covariate which summarises our knowledge about the way each covariate influence the phenomenon. Incorporating this information into the logistic regression model would result in a more insightful and accurate modelling of the spread of reed.

ACKNOWLEDGMENTS

The authors wish to thank Xiaojie Chen and Ioan Ferencik from Lahti Center of Aalto University, Samu Mäntyniemi from the Department of Environmental Science of University of Helsinki, and Paula Väänänen from the Finnish Environment Institute (SYKE) for the for their comments, suggestions, and support.

REFERENCES

- Accad, A., S.J. Low Choy, D. Pullar, and W. Rochester, Bioregional classification via model-based clustering. In: A. Zenger, and R.M. Argent (Eds.), *MODSIM 2005 International Congress on Modelling and Simulation*. Modelling and Simulation Society of Australia and New Zealand: 1326-1332, 2005.
- Adams, J.B., and G.C. Bate, Growth and photosynthetic performance of *Phragmites australis* in estuarine waters: a field and experimental evaluation, *Aquatic Botany* 64, 359-367, 1999.
- Al-Awadi, S.A., and P.H. Garthwaite, Quantifying expert opinion for modelling fauna habitat distributions, *Computational Statistics* 21(1), 121-140, 2006.
- Austin, M.P., Case studies of the use of environmental gradients in vegetation and fauna modelling: theory and practice in Australia and New Zealand. In: *Predicting Species Occurrences: Issue of Accuracy and Scale*, J.M. Scott, P.J. Heglund, and M.L. Morrison. Washington, Island Press, 73-82.
- Bart, D., and J.M. Hartman, The Role of Large Rhizome Dispersal and Low Salinity Windows in the Establishment of Common Reed, *Phragmites australis*, in Salt Marshes: New Links to Human Activities, *Estuaries* 26(2B), 511-521, 2003.
- Burdick, D., and R. Konisky, Determinants of Expansion for *Phragmites australis*, Reed, in Natural and Impacted Coastal Marshes, *Estuaries* 26(2B), 407-416, 2003.
- Chen, M.H, J.G. Ibrahim, and C. Yiannoutsos, Prior elicitation, variable selection and Bayesian computation for logistic regression models, *Journal of Royal Statistics Society* 61(1), 223-242, 1999.
- Denham, R., and K. Mengersen, Geographically assisted elicitation of expert opinion for regression models, *Bayesian Analysis* 2(1), 99-136, 2007.
- Edwards, A.L., The point biserial coefficient. In: *An Introduction to Linear Regression and Correlation*, W.H. Freeman, NewYork, 72-76, 1976.
- Ekeboom, J., P. Laihonon, and T. Suominen, A GIS-based step-wise procedure for assessing physical exposure in fragmented archipelagos, *Estuarine, Coastal and Shelf Science* 57, 887-898, 2003.
- Ellison, A.M., Bayesian inference in ecology, *Ecology Letters* 7, 509-520, 2004.
- Fonstad, M.A., Cellular automata as analysis and synthesis engines at the geomorphology-ecology interface, *Geomorphology* 77(3, 4), 217-234, 2006.
- Gelman, A., J.B. Carlin, H.S. Stern, and D.B. Rubin, *Bayesian Data Analysis*, second edition, Chapman & Hall/CRC, 668 pp., 2004.
- Gewin, V., Industry lured by the gains of going green, *Nature* 436, 173, 2005.

- Haslam, S.. Biological flora of the British Isles, *Phragmites communis* Trin, *Journal of Ecology* 60(2), 585-610, 1972.
- Huang, H.M., L.Q. Zhang, Y.J. Guan, and D.H Wang, A cellular automata model for population expansion of *Spartina alterniflora* at Jiuduansha Shoals, Shanghai, China, *Estuarine, Coastal and Shelf Science* 77(1), 47-55, 2008
- James, A., S. Low Choy, and K. Mengersen, Elicitor: An expert elicitation tool for regression in ecology, *Environmental Modelling & Software* 25, 129-145, 2010.
- King, R., W. Deluca, D. Whigham, and P. Marra, Threshold Effects of Coastal Urbanization on *Phragmites australis* (Common Reed) Abundance and Foliar Nitrogen Chesapeake Bay. *Estuaries and Coasts* 30(3), 469-481, 2007.
- Kocabas, V., and S. Dragicevic, Enhancing a GIS Cellular Automata Model of Land Use Change: Bayesian Networks, Influence Diagrams and Causality. *Transactions in GIS* 11(5), 681-702, 2007.
- Lelong, B., C. Lavoie, Y. Jodoin, and F. Belzile, Expansion pathways of the exotic common reed (*Phragmites australis*): a historical and genetic analysis, *Diversity and Distributions* 13, 430-437, 2007.
- Lett, C., C. Siber, and N. Barret, Comparison of a cellular automata net work and an individual-based model for the simulation of forest dynamics, *Ecological Modelling* 121, 277-293, 1999.
- Li, X., and A.G.O. Yeh, Urban simulation using principal components analysis and cellular automata for land-use planning, *Photogrammetric Engineering and Remote Sensing* 68, 341-351, 2002.
- Low Choy, S.J., R. O'Leary, and K. Mengersen, Elicitation by design for ecology: using expert opinion to inform priors for Bayesian statistical models, *Ecology* 90(1), 265-277, 2009.
- Lundqvist, D., D. Jansen, T. Balstroem, and C. Christiansen, A GIS-Based Method to Determine Maximum Fetch Applied to the North Sea–Baltic Sea Transition, *Journal of Coastal Research* 22(3), 640-644, 2006.
- Minchinton, T., Precipitation during El Niño correlates with increasing spread of *Phragmites australis* in New England, USA, coastal marshes, *Marine Ecology Progress Series* 242, 305-309, 2002.
- Parsons, J.A., and A.M. Fonstad, A cellular automata model of surface water flow, *Hydrological Processes* 21, 2189-2195, 2007.
- Pepper, A., and M. L. Puotinen, GREMO: A GIS-based generic model for estimating relative wave exposure. In: R.S. Anderssen, R.D. Braddock and L.T.H. Newham (Eds.), 18th World IMACS / MODSIM Congress, Cairns, Australia, 13-17, 2009.
- Pitkänen, T., Missä ruokoa kasvaa? Järviruokoalueiden satelliittikartoitus Etelä-Suomen ja Viron Väinämeren rannikoilla. *Turun ammattikorkeakoulun puheenvuoroja* 29, 82, 2006 (In Finnish).
- R Development Core Team, R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0, 2009, URL <http://www.R-project.org>.
- Sirakoulis, G.C.H., I. Karafyllidis, and A. Thanakakis, A cellular automaton model the effects of population movement and vaccination on epidemic propagation, *Ecological Modelling* 133, 209-229, 2000.
- Soomere, T., K. Myrberg, M. Leppäranta, and A Nekrasov, The progress in knowledge of physical oceanography of the Gulf of Finland: a review for 1997–2007, *Oceanologia* 50(3), 287-362, 2008.
- Tolvanen, H., and T. Suominen, Quantification of openness and wave activity in archipelago environments, *Estuarine, Coastal and Shelf Science* 64, 436-446, 2005.
- Vliet, J.V., R. White, and S. Dragicevic, Modeling urban growth using a variable grid cellular automaton, *Computers, Environment and Urban Systems* 33(1), 35-43, 2009.
- Wu, F., and C.J. Webster, Simulation of land development through the integration of cellular automata and multicriteria evaluation, *Environment and Planning B-Planning and Design* 25, 103-126, 1998.
- Xu, Z., R. Kella, A Bayesian logistic regression model for active relevance feedback, In: S.H. Myaeng, D.W. Oard, F. Sebastiani, T.S. Chua, and M.K. Leong (Eds.), Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval, Singapore, 227-234, 2008.

Yeh, A.G.O., and X. Li, Simulation of development alternatives using neural networks, cellular automata, and GIS for urban planning. *Photogrammetric Engineering and Remote Sensing* 69, 1043-1052, 2003.