

An Entropy Method for Floodplain Monitoring Network Design

E. Ridolfi^a, K. Yan^b, L. Alfonso^b, G. Di Baldassarre^b, F. Napolitano^a, F. Russo^a,
and Paul D. Bates^c

^a*Dipartimento di Idraulica, Trasporti e Strade – Università “Sapienza”, Rome, Italy, elena.ridolfi@uniroma1.it*

^b*Department of Integrated Water Systems and Governance - Hydroinformatics, UNESCO-IHE Institute for Water Education, Delft, The Netherlands*

^c*School of Geographical Sciences, University of Bristol, Bristol, UK*

Abstract. In recent years an increasing number of flood-related fatalities has highlighted the necessity of improving flood risk management to reduce human and economic losses. In this framework, monitoring of flood-prone areas is a key factor for building a resilient environment. In this paper a method for designing a floodplain monitoring network is presented. A redundant network of cheap wireless sensors (GridStix) measuring water depth is considered over a reach of the River Dee (UK), with sensors placed both in the channel and in the floodplain. Through a Three Objective Optimization Problem (TOOP) the best layouts of sensors are evaluated, minimizing their redundancy, maximizing their joint information content and maximizing the accuracy of the observations. A simple raster-based inundation model (LISFLOOD-FP) is used to generate a synthetic GridStix data set of water stages. The Digital Elevation Model (DEM) that is used for hydraulic model building is the globally and freely available SRTM DEM.

Keywords: Flood Risk, Flood Monitoring, Entropy, Network Design, Genetic Algorithm, SRTM DEM, Flood Modeling.

PACS: 92.40.Qk, *92.40.qp, 89.70.Cf

INTRODUCTION

In recent years the number and impact of disasters in Europe has increased consistently: in the period from 1998-2009 floods and storms caused 52 billion EUR and 44 billion EUR, respectively [11]. In this framework, the necessity of managing and reducing the flood impacts was evident. The interconnection between human development and flood risk was introduced by Gilbert White with the so-called “levee effect”. This false sense of safety leads people to invest in risk-prone areas because of a low flood risk perception [7;9]. The Directive 2007/60/EC, [11], imposed on EU Members States the requirement to analyze and identify territories in flood prone areas and to draw flood risk maps in order to prevent flood hazards and increase preparedness for those events. In the past, effort has been made in terms of hazard nowcasting [18] and flood forecasting [25; 21]. Flood risk has also been assessed through flood mapping, thanks to one-dimensional (1D) and two-dimensional (2D) hydraulic models used as numerical tools [3; 6; 14]. Furthermore the introduction of remote sensing data has led to important improvements in calibrating models for flood extent definition and in monitoring flood prone areas, [5; 26]. The issue of assessing and designing hydrologic data collection networks through an entropy approach has been discussed by authors in fields ranging from water-quality monitoring (e.g. 19; 20) to raingauge networks (e.g. 17; 24). The paper is developed as follows: first a simple raster-based inundation model, LISFLOOD-FP [4], is used to generate a synthetic data set of water stages in 16163 points of the studied area. In each point, it is assumed to be a sensor of the GridStix network [15], thus for each point a synthetic value of the water stage is available. The Digital Elevation Model (DEM) used for hydraulic model building is the globally and freely available SRTM DEM. Then the optimization problem consists of finding the optimal set of monitors by posing it as a Three-Objective Optimization Problem (TOOP). The three objective functions to guarantee the optimality of the set of monitors are: a) to acquire as much information as possible (maximizing the information content); b) to restrain the number of sensors over the area (minimizing their redundancy); and c) to maximize their accuracy (minimizing the entropy of the errors between the observed and modeled water stage).

NETWORK DESIGN THROUGH THE OPTIMIZATION PROBLEM

In literature, the entropy theory has been extensively applied to water resource field [28]. For instance [12] found a link between the mean elevation of a river basin and its entropy, [22] evaluated rainfall threshold value minimizing a Bayesian-entropy risk function. The marginal entropy of a Random Vector (RV) is defined as [23]:

$$H(X) = E[-\log_2 p(X)] = -\sum_i p_i \log_2 p_i \quad (1)$$

where $P\{X = x_i\} = p_i$. Considering N RVs, their joint entropy is [17; 24]:

$$JH = H(X_1, X_2, \dots, X_N) = -\sum_i \sum_j p_{i, \dots, j} \log_2 p_{i, \dots, j} \quad (2)$$

where $p_{i, \dots, j}$ is the joint probability of the N variables. For evaluating the marginal probability, all RVs components are subdivided into different classes and their frequency is assessed. In order to evaluate the joint probability of (e.g.) 2 RVs, their components are conglomerated so that the marginal entropy of the conglomerated vector is equal to the joint entropy of the 2 RVs: $H(Z) = H(A, B)$, [2; 16]. The total correlation index measures the redundancy:

$$TotC(X_1, X_2, \dots, X_N) = \sum_{i=1}^N H(X_i) - H(X_1, X_2, \dots, X_N) \quad (3)$$

where $H(X_i)$ is the marginal entropy of the i th RV and $H(X_1, X_2, \dots, X_N)$ is the joint entropy of the N RVs, [1; 29]. The case study is a 10-km reach of the River Dee, between the two gauge stations at Farndon and Iron Bridge, Fig. 1. The 2D model LISFLOOD-FP, [4], is used to generate a synthetic GridStix data set of water stages.



FIGURE 1. The River Dee (black lines), the two Environment Agency of England and Wales (EA) gauge stations at Farndon and Iron Bridge. The figure also shows the area inundated by the flood of December 2006, derived by high-resolution satellite image (ERS-2 SAR).

The Digital Elevation Model (DEM) used for hydraulic model building is the globally and freely available SRTM DEM. The boundary condition is the low-magnitude (return period of around 2 years) flooding of December 2006, [8]. Because of the Manning's coefficient of the floodplain is the only parameter that affects the results [8], a constant n value was assigned to the channel (i.e. $0.25 \text{ s/m}^{1/3}$), while the floodplain varies between 0.02 and 0.10 ($\text{s/m}^{1/3}$). In absence of a suitable observed event, the results of a fully two-dimensional finite element model, TELEMAC-2D [13], are considered as "truth". Then, the entropy of the errors between each model and the "observed" data is evaluated for each time step. The mean entropy is evaluated over all time steps and those sensors with a mean entropy lower than the median of the whole set of monitors is disregarded in order to avoid solutions with monitors that provide a low amount of information. For designing the sensors network, the TOOP was used, representing an innovation with respect to the MOOP [2], where only redundancy and joint information were taken into account. The introduction of a third objective function guarantees that accuracy is maximized, (i.e. the mean entropy of errors between observed and simulated data). The optimization problem is then formulated as follows:

$$\begin{aligned} \text{Min}(TotC) &= \text{Min}\{C(X_{sim1}, X_{sim2}, \dots, X_{simN})\} \\ \text{Min}(JH) &= \text{Max}\{H(X_{sim1}, X_{sim2}, \dots, X_{simN})\} \\ \text{Min}(H_{mean}) &= \text{Min}\left(\frac{1}{N} \sum_{i=1}^N (X_{TELEMACi} - X_{simi})\right) \end{aligned} \quad (4)$$

where X_{sim_i} is the water stage time series simulated in the i th point (i.e. i th sensor) and $X_{TELEMAC_i}$ is the corresponding time series generated with TELEMA-2D. The TOOP was solved through the non-sorted genetic algorithm (NSGA-II), [10] and the optimal sets of solutions for different number of sensors are represented in a 3D Pareto front, Fig. 2.

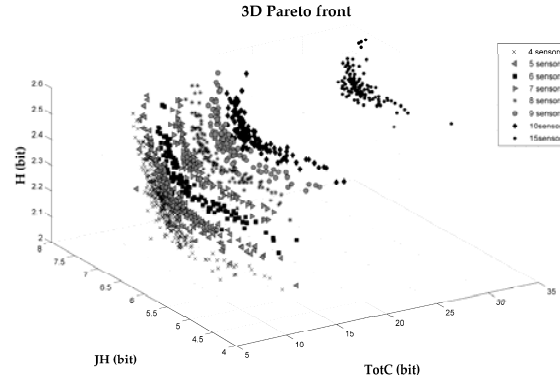


FIGURE 2. 3D Pareto front of best solutions, considering a number of variables (i.e. sensors) from 4 to 15.

From Fig.2 it is possible to observe that as the number of sensors increase, the error decreases (i.e. Hmean), JH increases marginally and the TotC increases significantly. This means that once that the optimal number of sensors is deployed, adding more sensors implies that there is no information gain compared to the high redundancy increment. Thus, it is not worthy considering a number of sensors higher than 10. Over this number the redundancy is much higher without any appreciable increase of information or precision. Observing the Pareto, the optimal number of sensors is 6, representing a good trade-off for fulfilling the three conditions: JH is higher than for fewer sensors, and although its TotC is higher than for 5 sensors, the Hmean is smaller, assuring more precision. From the original sets given by the Pareto front, were chosen only sets with Mean Absolute Error (MAE) < 0.5 m, because the DEM vertical accuracy under 0.5 m achieves good modeling results, [27]. Although the SRTM DEM is characterized by a course resolution [27], this threshold value of the MAE was assumed in order to investigate the potentiality of the DEM itself. Through a frequency analysis, the best Manning's coefficient value is equal to 0.055 s/m^{1/3} and it matches literature results [8].

CONCLUSIONS

It has been demonstrated that it is possible to evaluate the best location and the optimum number of sensors on a flood prone area using an entropy-based approach. Restraining the number of sensors from the original number allows to eliminate locations that are not significant for describing the flood extent. Results obtained using the entropy method match with those provided by literature guidelines. Furthermore, using the TOOP, several configurations can be obtained for the same number of sensors. One criteria to select one solution is by looking at the set with the lowest MAE. It is important to highlight that using the low resolution SRTM DEM, results are reliable and underline the great potential for this freely and globally available DEM.

ACKNOWLEDGMENTS

Authors are grateful to the European Space Agency (ESA) for allowing access to the flood image used in this study (Category 1 Project ID: 5739) and the EA for providing measured data of the two gauge stations.

REFERENCES

1. L. Alfonso, A. Lobbrecht, R. Price, "Information theory-based approach for location of monitoring water level gauges in polders", *Water Resour. Res.*, **46**, W03528, doi: 10.1029/2009WR008101 (2010a).
2. L. Alfonso, A. Lobbrecht, R. Price, "Optimization of water level monitoring network in polder systems using information theory", *Water Resour. Res.*, **46**, W12553, doi: 10.1029/2009WR008953, (2010b).
3. G. Aronica, P. D. Bates and Horritt, M. S., "Assessing the uncertainty in distributed model predictions using observed binary pattern information within GLUE", *Hydrol. Processes* **16** (10), 2001–2016 (2002).
4. P.D. Bates, A.P.J. De Roo, "A simple-based raster model for flood inundation simulation", *J. of Hydrol.*, **236** (1-2), 54-77(2000).
5. P. D. Bates, M. S. Horritt, G. Aronica and K. Beven, "Bayesian updating of flood inundation likelihoods conditioned on flood extent data", *Hydrol. Processes* **18**, 3347–3370 (2004).
6. G. Di Baldassarre, A. Castellarin, A. Montanari and A. Brath, "Probability weighted hazard maps for comparing different flood risk management strategies: a case study", *Natural Hazards*, doi:10.1007/s11069-009-9355-6 (2009).
7. G. Di Baldassarre, A. Montanari, H. Lins, D. Koutsoyiannis, L. Brandimarte, G. Blöschl, "Flood fatalities in Africa: From diagnosis to mitigation", *Geophys. Res. Lett.*, **37**, L22402, doi:10.1029/2010GL045467, (2010a).
8. G. Di Baldassarre, G. Schumann, P.D. Bates, J.E. Freer, K.J. Beven, "Flood-plain mapping: a critical discussion of deterministic and probabilistic approaches", *Hydrolog. Sci. J.*, **55**(3), 364-376 (2010b).
9. A. Castellarin, G. Di Baldassarre, A. Brath, "Floodplain management strategies for flood attenuation in the River Po", *River Res. Applic.* **27**, 1037–1047 (2011).
10. K. Deb, A. Pratap, S. Agarwal, T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II", *Evol. Comput.*, **6**, 182-197 (2002).
11. European Parliament, Directive 2007/60/Ec of the European Parliament and of the Council of 23 October 2007 on the assessment and management of flood risks. <http://eur-lex.europa.eu/en/index.htm> (2007).
12. M. Fiorentino, P. Claps and V. P. Singh, "An entropy-based morphological analysis of river basin networks", *Wat. Resour. Res.* **29**(4), 1215-1224 (1993).
13. J. C. Galland, N. Goutal, and J. M. Hervouet, "TELEMAC: A new numerical model for solving shallow water equations", *Adv. Water Resour.* **14**(3), 38–148 (1991).
14. M. S. Horritt, G. Di Baldassarre, P. D. Bates and A. Brath, "Comparing the performance of 2-D finite element and finite volume models of floodplain inundation using airborne SAR imagery", *Hydrol. Processes* **21**, 2745–2759 (2007).
15. D. Hughes, P. Greenwood, G. Coulson, G. Blair, F. Pappenberger, P. Smith and K. Beven, "GridStix: Supporting flood prediction using embedded hardware and next generation grid middleware", Proceeding of WoWMoM 2006, Buffalo-Niagara Falls, NY, 2006, pp. 621-626.
16. A. Kraskov, H. Stogbauer, R. G. Andrzejak and P. Grassberger, "Hierarchical clustering using mutual information", *Europhys. Lett.*, **70**, 278–284 (2005).
17. P. F. Krstanovic, V. P. Singh, "Evaluation of rainfall networks using entropy, I: Theoretical development", *Water Resour. Manag.*, **6**, 279–293, (1992).
18. F. Lombardo, F. Napolitano, F. Russo, "On the use of radar reflectivity for estimation of the areal reduction factor", *Nat. Hazards Earth Syst. Sci.*, **6** (3), 377-386 (2006).
19. Y. Mogheir, V. P. Singh, "Application of information theory to groundwater quality monitoring networks", *Water Resour. Manag.*, **16**, 37–49, 2002.
20. Y. Mogheir, J. L. M. P. de Lima, V. P. Singh, "Characterizing the spatial variability of groundwater quality using the entropy theory, I. Synthetic data", *Hydrol. Process.*, **18**, 2165–2179 (2004).
21. V. Montesarchio, F. Lombardo, F. Napolitano, "Rainfall thresholds and flood warning: an operative case study", *Nat. Hazards Earth Syst. Sci.*, **9**, 135-144, (2009).
22. V. Montesarchio, E. Ridolfi, F. Russo, F. Napolitano, "Rainfall threshold definition using an entropy decision approach and radar data", *Nat. Hazards Earth Syst. Sci.*, **11**, 2061–2074, doi:10.5194/nhess-11-2061-2011 (2011).
23. A. Papoulis, Probability, *Random Variables and Stochastic Processes*, 3rd Edn., McGraw-Hill, New York, 1991.
24. E. Ridolfi, V. Montesarchio, F. Russo, F. Napolitano, "An entropy approach for evaluating the maximum information content achievable by an urban rainfall network", *Nat. Hazards Earth Syst. Sci.*, **11**, 2075-2083 (2011).
25. F. Russo, F. Lombardo, F. Napolitano, E. Gorgucci, "Rainfall stochastic modeling for runoff forecasting", *Phys. Chem. Earth*, **31** (18), 1252–1261 (2006).
26. G. Schumann, P. Matgen, L. Hoffmann, R. Hostache, F. Pappenberger, L. Pfister, "Deriving distributed roughness values from satellite radar data for flood inundation modelling", *J. Hydrol.*, **344**, 96– 111 (2007).
27. G. Schumann, G. Di Baldassarre and P. D. Bates, "The utility of space-borne radar to render flood inundation maps based on multi-algorithm ensembles", *IEEE Trans. Geosci. Remote Sens.* **47**(8), 2801–2807 (2009).
28. V. P. Singh, "The use of entropy in hydrology and water resources", *Hydrol. Process.* **11**, 587–626 (1997).
29. S. Watanabe, "Information theoretical analysis of multivariate correlation", *IBM J. Res. Dev.*, **4**(1), 6682 (1960).