Following Flood Dynamics by SAR/Optical Data Fusion

A. D'Addabbo, A. Refice, G. Pasquariello and F. Lovergine
CNR-ISSIA
Via Amendola 122/D
70125 Bari, Italy
Email: daddabbo@ba.issia.cnr.it

S. Manfreda
DICEM, University of Basilicata
Via Lazazzera,
75100 Matera, Italy

***Abstract*—Synthetic aperture radar (SAR) acquisitions are particularly useful to produce flood maps thanks to their all-weather and day-night capabilities. However, repetition intervals of radar instruments are in the order of several days for routine operations, reaching daily or higher frequencies only in tasked conditions. Therefore, to follow flood dynamics, images acquired by different sensors at different times may be beneficial. In the present work, multi-temporal SAR intensity, InSAR coherence and optical data are considered to describe a flood event occurred in the Basilicata region (southern Italy) on December 2013. In this case study, optical data have a twofold role: they allow to follow the flood dynamics (because SAR and optical data have been acquired in different dates during the inundation event), and they add information concerning the land cover of the analyzed area. The data fusion approach is based on Bayesian Networks (BNs). It is shown that the synergetic use of different information layers can help detect more precisely the areas affected by the flood, reducing false alarms and missed identifications which may affect algorithms based on data from a single source. The produced flood maps are compared to reference maps, independently obtained; the comparison indicates that the proposed methodology is able to reliably follow the temporal evolution of the phenomenon, assigning high probability to areas most likely to be flooded, reaching accuracies of up to 89%.**

I. Introduction

Flooding is one of the most dramatic and frequent natural disaster that affects several areas in the world. Producing accurate flood maps and following event dynamics are crucial issues in hydrogeological studies. A valuable information source for these tasks can be remotely sensed imagery, providing a synoptic overview of flood extension at each image acquisition time. Synthetic aperture radar (SAR) acquisitions are particularly useful to produce flood maps thanks to their all-weather and day-night capability [1], [2]; moreover, new satellite missions supply a large amount of daily acquired radar images, characterized by high spatial resolution, useful in flood detection problems. Nevertheless, repetition intervals of radar instruments are in the order of several days for routine operations, reaching daily or higher frequencies only in tasked conditions, typically under requests from civil protection or other public entities in case of severe events, and for limited times. Therefore, to follow flood dynamics, images acquired by different sensors at different times could be beneficial. In the present work, multi-temporal SAR intensity images, multi-temporal InSAR coherence data and optical data are considered to describe a flood event occurred in the Basilicata region (southern Italy) on December 2013.

Flood scenarios are typical examples of complex situations in which different factors have to be considered to provide accurate and robust interpretation of the situation on the ground [3]. Recently, the fusion of multi-temporal, multi-sensor and/or multi-platform Earth observation image data, together with other ancillary information, seems to have a key role in the pursuit of a consistent interpretation of flooded scenes [4]–[6]. In this case study, optical data have a twofold role: they allow to follow the flood dynamics (because SAR and optical data have been acquired in different dates during the inundation event), and, moreover, they add information concerning the land cover of the analyzed area.

The data fusion approach, used in the present work to produce flood maps and follow flood dynamics, is based on Bayesian Networks (BNs), a tool recently proven to be a viable means to integrate multi-temporal SAR intensity images and InSAR coherence data, with ancillary information [6]. It is shown that the synergetic use of different information layers can help to detect more precisely the areas affected by the flood, reducing false alarms and missed identifications which may affect algorithms based on data from a single source. In Section II some information on the study area and the considered flood event is reported, and the available data are described. In Section III the data fusion approach is discussed in detail. The experimental results are presented in Section IV. Finally, in Section V, some conclusions are drawn.

II. Study area and data set

We analyze a flood event occurred in December, 2013, in the Basilicata region, Southern Italy, involving the Bradano river basin. It was due to rainfall starting from November 30 and continuing until December 5. In particular, we consider an area around the Bradano river, for which the peak flow (measured at the gauging station SS106 close to the outlet) was recorded in the evening of December 1, with the discharge reaching about 800 m^3/s, and caused an inundation that propagated in the surrounding areas, producing floods during the following four days. Another, less relevant peak of about 300 m^3/s was registered on December 2. Thereafter, the rainfall event was exhausted and the dams released some of the stored flood

volume. In particular, as officially reported in the data of Basilicata Civil Protection, on Dec. 2 the Bradano river provoked inundations that affected large areas concentrated along the river banks. In the following days, the phenomenon was in a decreasing phase and the flood effects were propagating toward the coast, interesting smaller areas progressively farther away from the river [6].

A series of 3 COSMO-SkyMed stripmap SAR images was available over the analyzed area, with a ground pixel size of approximately 3×3 m^2, acquired in the same geometry, polarization, and incidence angle, so that InSAR processing could be performed. The acquisition dates are November 16 and December 2 and 3, 2013. As can be noted, the two SAR acquisitions of Dec. 2 and 3 (depicted in fig. 1-(a) and -(b)) offer a useful observation data set to follow the temporal evolution of the flood wave phenomenon, that was in a crucial phase on December 2 and in a decreasing one on the following day. The InSAR coherence computed between these two images was also considered (fig. 1-(c)). Other 7 COSMO-SkyMed stripmap SAR images were considered over the same area, acquired in absence of flood events, respectively on October 2, 3 and 10, 2010, January 17, 18 and 25, and February 18, 2011. InSAR coherence between the Oct. 2 and 3, 2010, and Jan. 17 and 18, 2011, was considered as well. These additional SAR data act here as a reference data set, that allows to capture multi-temporal backscattering and coherence trends, aiding in the detection of flood.

We also consider a Pleiades-1B optical image, at a ground resolution of 2 m, acquired on Dec. 5, 2013, which presents sufficient cloud-free information around the study area. The image is acquired in 4 spectral bands, of which the ones corresponding to the red and near infrared intervals were used to obtain the normalized difference vegetation index (NDVI) according to the formula:

$$\text{NDVI} = \frac{L_{\text{IR}} - L_{\text{R}}}{L_{\text{IR}} + L_{\text{R}}}, \tag{1}$$

where L_{IR} and L_{R} are the IR and red radiance values, respectively. Use of the radiances instead of the reflectances in 1 is justified by assuming constant atmospheric and other effects on the two close spectral bands. Positive NDVI values indicate the presence of vegetation, while negative values are associated with the presence of water. In fig. 1(d), a RGB composition of Pleiades bands 4, 3 and 2 is depicted.

III. Methods

We consider a BN as represented in Fig. 2: it is designed as a Direct Acyclic Graph (DAG) [7], where the circles correspond to random variables and the arrows describe causal relations between them, modelled with conditional probabilities. The proposed BN combines information extracted by the set of SAR intensity images, the set of InSAR coherence images, the optical image and some ancillary information about flood distribution in the analyzed area. In detail, the random variable F is discrete and consist of only two states: flood (f) and no flood (nf), so that $P(F = f) = 1 - P(F = nf)$.

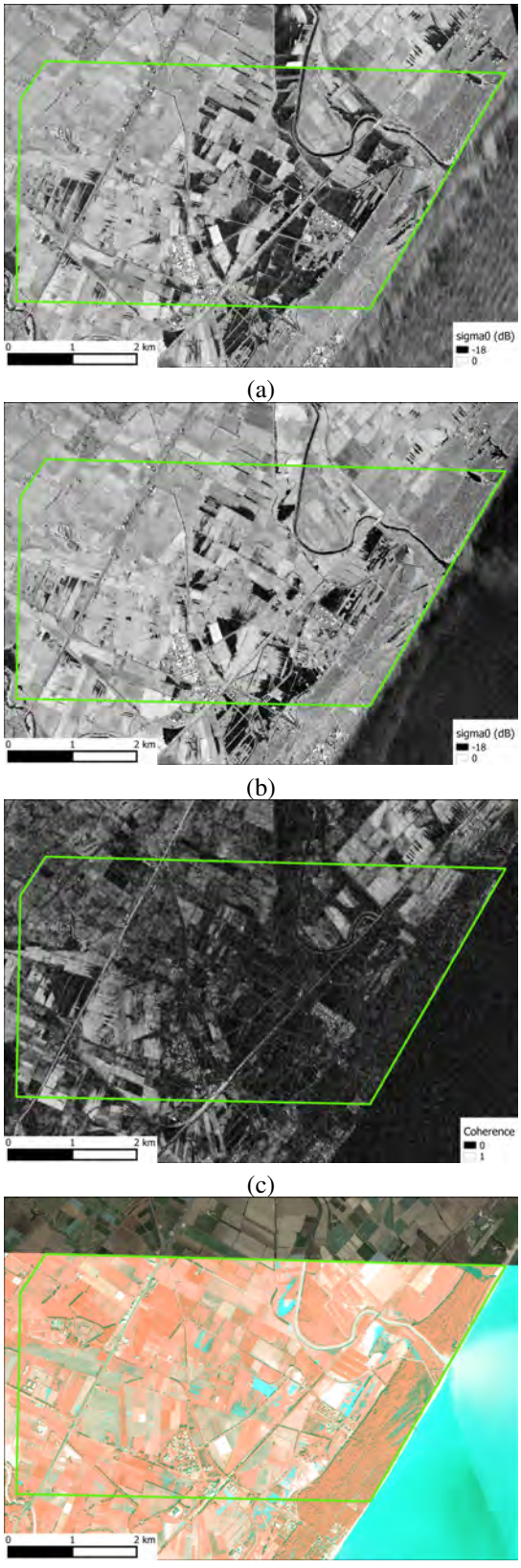

Fig. 1. Input data: backscattering coefficient σ^0 image acquired on (a) Dec. 2 and on (b) Dec. 3, (c) coherence image computed between the Dec. 2 and 3 acquisitions, (d) a RGB composition of bands 4,3,2 of the Pleiades image acquired on Dec. 5.

$$p(F=f|\sigma^0=\sigma^{0*},\gamma=\gamma^*,o=o^*,t=t^*)=\frac{p(F=f,\sigma^0=\sigma^{0*},\gamma=\gamma^*,o=o^*,t=t^*)}{p(\sigma^0=\sigma^{0*},\gamma=\gamma^*,o=o^*,t=t^*)}=$$

$$=\frac{[\sum_{C_{\sigma^0}}p(\sigma^0=\sigma^{0*}|C_{\sigma^0})p(C_{\sigma^0}|F=f)][\sum_{C_\gamma}p(\gamma=\gamma^*|C_\gamma)p(C_\gamma|F=f)][\sum_{C_o}p(o=o^*|C_o)p(C_o|F=f)]p(F=f|t=t^*)}{\sum_F[\sum_{C_{\sigma^0}}p(\sigma^0=\sigma^{0*}|C_{\sigma^0})p(C_{\sigma^0}|F)][\sum_{C_\gamma}p(\gamma=\gamma^*|C_\gamma)p(C_\gamma|F)][\sum_{C_o}p(o=o^*|C_o)p(C_o|F)]p(F|t=t^*)}$$

This is the variable that we want to estimate by statistical inference. The variables σ^0,γ and o correspond respectively to the n-dimensional vector obtained from the multi-temporal SAR intensity imagery, the m-dimensional vector obtained from the multi-temporal InSAR coherence imagery, and the 1-dimensional vector obtained from the optical NDVI image. Finally, the variable t represents ancillary information which can be added to the data, in this case a geomorphic flooding index (GFI) derived from a Light Detection And Ranging (LiDAR)-derived DEM at 3 m posting. The t variable is computed from the definition in [8]:

$$t=\ln(h_r/H),$$

where h_r is a function of the contributing area A_r in the nearest section of the drainage network hydrologically connected to the point under exam, and H is the elevation difference between the cell under exam and the final point of the above-identified path. In this index, the parameter h_r is representative of the probable water level in a cross section of the river hydraulically connected to the point, and it is computed as a power law of A_r with exponent set equal to 0.3. The variable t is introduced to assign correct a priori flood probability to zones, even relatively far from the river course, that may be flooded due to the presence of dense channel networks, that become a preferential vector for the flow during a flood. The t layer is only available on an area which is contained within the green polygon shown in Figs. 1 and Fig. 4, overlayed on the various maps.

Since the flood state often does not exhibit a simple "causal" relation with the imagery, the intermediate variable classes C_{σ^0}, C_γ and C_o are introduced [9]. They consist each of several possible states, that can be determined either by using some a priori knowledge on the scene, or by extracting them automatically from the imagery. The relations between variables are modelled as conditional probabilities and the conditional independence assumptions hold, i.e. if two variables are not directly connected in the graph they are independent. We want to infer the value in eq. 1, where $\{\sigma^{0*},\gamma^*,o^*,t^*\}$ are observed values for the random variables σ^0,γ,o and t, respectively. In particular, each conditional probability $p(g|C_g)$, with $g\in\{\sigma^0,\gamma,o\}$ is given as a multi-dimensional probability function generated by a mixture of K_g Gaussian distributions, whose parameters $\{\mu_g,\Sigma_g\}$ are automatically computed by applying a K-means algorithm [10]. The number K_g of clusters C_g are determined by a trial-and-error procedure, in order to find the value that provides a good representation of the different classes believed to be present in the analyzed scene. The probabilities $p(C_g|F)$ are computed by applying the Bayes rule: $p(C_g|F)=p(F|C_g)p(C_g)/p(F)$, where the $p(C_g)$ terms are derived from the K-means algorithm, and the $p(F|C_g)$ are assigned by the user. The conditional probability $p(F|t)$ can be modelled in different ways. In the following, we consider $p(F|t)$ modelled as a sigmoid function.

IV. Experimental results

Flood maps have been produced by applying eq. (1). In particular, the conditional probabilities $p(g|C_g)$ have been computed as multi-dimensional probability functions generated, respectively, by a mixture of $K_{\sigma^0}=32,K_\gamma=8,K_o=16$ Gaussian distributions. Their parameters have been automatically computed by applying the K-means algorithm [10]. The probabilities $P(F|C_g)$ have been assigned by the user by considering the centroid value μ_{C_g} of each intermediate class. The probability $p(F|t)$ has been computed as a sigmoid function of the variable t, with parameters which have been assigned by the user by considering information on the flood evolution.

The flood maps depicted in fig. 3 have been obtained by considering a threshold of 0.8 to each final $P(F=\mathrm{f}|g)$ map. The BN maps have been compared with reference flood maps independently obtained from the Flora2D hydraulic model [11] and depicted in Fig. 4. The flood and no-flood instances are unbalanced on the maps, so the true positive rate (TPR) and false positive rate (FPR) values are also considered to give more objective evaluations. They are defined as:

$$\mathrm{TPR}=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}};$$

$$\mathrm{FPR}=\frac{\mathrm{FP}}{\mathrm{FP}+\mathrm{TN}};$$

where TP is the number of True Positives, i.e. the actual flood data that are correctly classified, FP is the number of

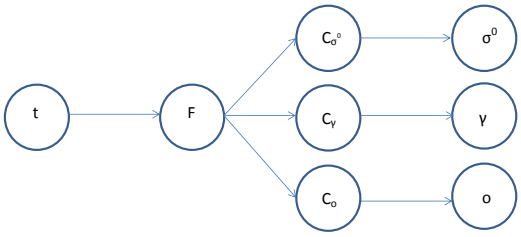

Fig. 2. BN scheme

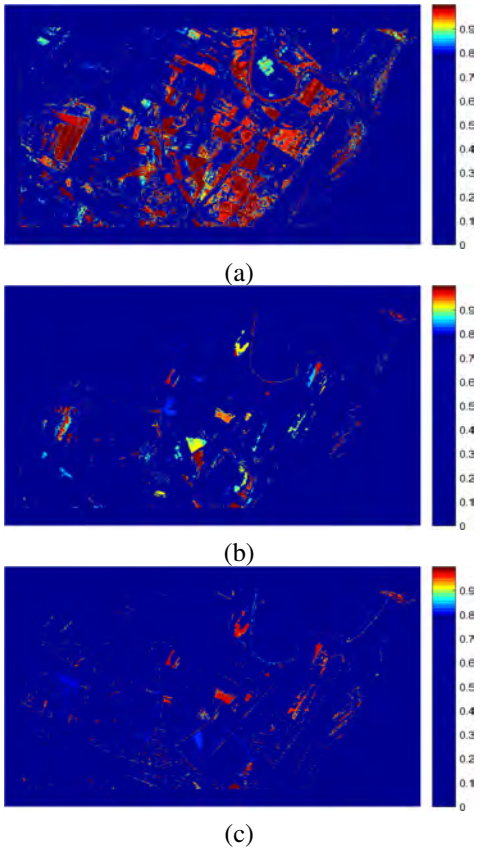

Fig. 3. Flood maps produced by BN for (a) Dec. 2, (b) Dec. 3, and (c) Dec. 5 dates. Colors are proportional to the final $p(F)$ probability.

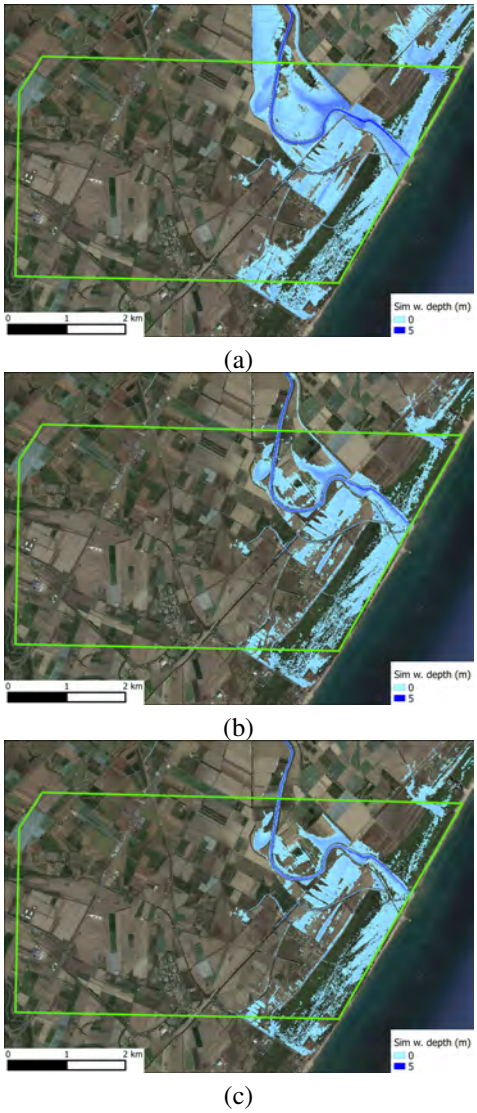

Fig. 4. Reference maps obtained from Flora 2D hydraulic model for (a) Dec. 2, (b) Dec. 3, and (c) Dec. 5 dates. Here, colors are proportional to simulated water depth.

False Positives, i.e. negative data classified as positive, TN is the number of True Negatives, i.e. the actual no flood data that are correctly classified, and FN is the number of False Negatives, i.e. positive data classified as negative. In Fig. 5, the corresponding receiver operator characteristic (ROC) curves are reported: a relative agreement with the reference maps has been obtained, with areas under the curve (AUC) values for the Dec. 2, Dec. 3 and for the Dec. 5 map equal to 0.61, 0.60 and 0.60, respectively. The best overall accuracies, obtained by varying the probability threshold values, are respectively equal to 80%, 87% and 89%, for Dec. 2, 3 and 4. It is worth noting that the test area lies between the Bradano and Basento rivers and could thus be potentially interested by flood due to the cumulative action of both rivers, while the reference data consider only the Bradano contribution. Moreover, some areas could be inundated for rain accumulation. For this reason, the Dec. 2 map shows several flooded areas which are not visible in the reference map: these decrease the AUC of the ROC curves. In the future, the flooded areas not affected by the river flow should be recognized and removed from the comparison. Nevertheless, the BN framework allows to reduce false alarms and missed identifications by performing data fusion. In fact, by considering the pixels exhibiting high values of backscattering intensity on Dec. 2, we found that these are almost 59% of the total analyzed pixels and are scattered across the whole scene. By considering also the InSAR coherence information, we found that only about half of these points (corresponding to about 32% of the total analyzed pixels) exhibit also a significant change in the coherence value. Moreover, we observed that, by adding both NDVI information and ancillary information, they are reduced to only 21%. In this way, a useful reduction of false alarms can be obtained. On the other hand, some areas, that we know were flooded, but that exhibited high values of backscattering intensity, show high values of posterior probability, indicating that the data fusion performed through the BN framework allows to overcome the missed identification which would be obtained by using only one data source. Similar consideration are true for the other two maps.

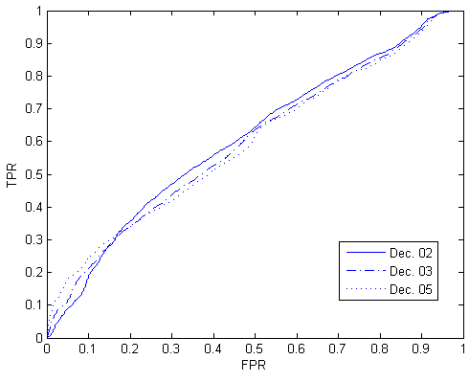

Fig. 5. ROC curves obtained by comparing the Probability Maps with the reference maps for Dec. 2, Dec. 3 and Dec. 5.

V. Conclusion

We have shown a data fusion approach, based on Bayesian Networks (BNs), to detect flooded areas and following flood dynamics. It opportunely combines SAR intensity, InSAR coherence and optical imagery with ancillary data. Results show the advantage of integrating heterogeneous sources of information in order to reduce uncertainties in the mapping of the presence of water on different land cover types. In particular, it allows to both mitigate the false alarms and to correctly identify flooded areas in events characterized by complex land cover ground conditions and time evolution.

Acknowledgment

COSMO-SkyMed images are courtesy of Italian Space Agency. Ing. L. Candela, ASI, is kindly acknowledged for support in data acquisition. InSAR processing was performed by Dr. D. O. Nitti of GAP s.r.l.

References

[1] L. Pulvirenti, M. Chini, N. Pierdicca, L. Guerriero, P. Ferrazzoli, "Flood monitoring using multi-temporal COSMO-SkyMed data: image segmentation and signature interpretation," *Remote Sens. Environ.*, Vol. 115, pp. 990-1002, 2011.

[2] P. Matgen, R. Hostache, G. Schumann, L. Pfister, L. Hoffmann, H.H.G. Savenije, "Towards an automated sar-based flood monitoring system: lessonslearned from two case studies," *Physics and Chemistry of the Earth*, vol. 36, pp. 241–252,2011.

[3] N. Pierdicca, L. Pulvirenti, M. Chini, L. Guerriero, L. Candela, "Observing floods from space: Experience gained from COSMO-SkyMed observations," *Acta Astronautica*, Vol. 84, pp. 122-133, 2013.

[4] A. Refice, D. Capolongo, G. Pasquariello, A. D'Addabbo, F. Bovenga, R. Nutricato, F. P. Lovergine, and L. Pietranera, "SAR and InSAR for flood monitoring: examples with COSMO-SkyMed data," *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.*, Vol. 7, No. 7, pp. 2711-2722, 2014.

[5] L. Pulvirenti, M. Chini, N. Pierdicca, G. Boni, "Use of SAR Data for Detecting Floodwater in Urban and Agricultural Areas: The Role of the Interferometric Coherence," *IEEE Trans. Geosci. Rem. Sens.*, Vol. PP, pp. 1-13, 2015.

[6] A. D'Addabbo, A. Refice, G. Pasquariello, F. Lovergine, D. Capolongo, S. Manfreda, "A Bayesian Network for Flood Detection Combining SAR Imagery and Ancillary Data", *IEEE Trans. Geosci. Rem. Sens.*, in press.

[7] D. Barber, *Bayesian Reasoning and Machine Learning*, Cambridge University Press, 2012.

[8] S. Manfreda, C. Samela, A. Sole, M. Fiorentino "Flood-Prone Areas Assessment Using Linear Binary Classifiers based on Morphological Indices," *Vulnerability, Uncertainty, and Risk*, pp. 2002-2011, 2014.

[9] D. Frey, M. Butenuth, and D. Straub, "Probabilistic graphical models for flood state detection of roads combining imagery and DEM," *IEEE Geosci. Rem. Sens. Letters*, Vol. 9, No. 6, pp. 1051-1055, 2012.

[10] C.M. Bishop, *Neural Networks for Pattern Recognition*, Oxford University Press, 1995.

[11] S. Manfreda, C. Samela, A. Gioia, G. Consoli, V. Iacobellis, L. Giuzio, A. Cantisani, A. Sole, "Flood-Prone Areas Assessment Using Linear Binary Classifiers based on flood maps obtained from 1D and 2D hydraulic models," *Natural Hazards*, vol. 79, no. 2, pp. 735754, 2015.