Probabilistic Characterization of Flood Hazard Using Bivariate Analysis Based on Copulas

A. Candela[1] and G. T. Aronica[2]

[1]Dipartimento di Ingegneria Civile, Ambientale, Aerospaziale, dei Materiali, Università di Palermo, Viale delle Scienze - Parco d'Orleans 90128 Palermo, Italy. PH (+39) 091-23896559; FAX (+39) 091-23860810; email: angela.candela@unipa.it
[2]Dipartimento di Ingegneria Civile, Informatica, Edile, Ambientale e Matematica Applicata Università di Messina, Villaggio S.Agata, 98166 Messina, Italy PH (+39) 090-3977164; FAX (+39) 090- 3977480; email: garonica@unime.it

ABSTRACT

This study presents an innovative approach to obtain flood hazard maps where hydrological input (synthetic flood design event) to a 2D hydraulic model has been defined by generating flood peak discharges and volumes from a bivariate statistical analysis, through the use of copulas. Synthetic hydrographs were generated by means two different approaches: an indirect one, where rainfall were generated by a stochastic bivariate rainfall generator to be entered a distributed conceptual rainfall-runoff model that consisted of a soil moisture routine and a flow routing routine; and a direct one, where stochastic generation of flood peaks and flow volumes have been obtained via copulas, which describe and model the correlation between these two variables independently of the marginal laws involved fitted on flood recorded data.

Finally, to highlight the advantages of the presented approach, flood hazard maps derived by bivariate analisys are compared to maps from conventional univariate analysis. The procedure is applied to a real case study located in the southern part of Sicily, Italy, where flood hazard maps have been obtained and compared.

INTRODUCTION

Floods are a global problem and are considered the most frequent natural disaster world-wide. Many studies show that the severity and frequency of floods have increased in recent years and underline the difficulty to separate the effects of natural climatic changes and human influences as land management practices, urbanization etc.

Flood risk analysis and assessment is required to provide information on current or future flood hazard and risks in order to accomplish flood risk mitigation, to propose, evaluate and select measures to reduce it. Thus, the European Parliament has adopted the new Directive 2007/60/EC (Flood Directive) that requires Member State to assess and to recognize areas at risk from flooding and take measures to reduce the evaluated risk. According to the Directive flood hazard maps and flood risk maps will identify areas which could be flooded according to three scenarios:

floods with a low probability (RP • 300 years), floods with a medium probability (RP • 100 years) and floods with a high probability (RP < 100 years).

Flood risk is, generally, defined as the product of the probability of an event (hazard) and the consequences of that event (vulnerability). Both components of risk can be mapped individually and are affected by multiple uncertainties as well as the joint estimate of flood risk (Beven et al., 2011). Major sources of uncertainty in producing flood risk maps include statistical analysis of extremes events from short time series, incorporating themselves measurement errors, the definition of the floodplain topography, the choice of effective hydraulic roughness coefficients and hydraulic model to use (Aronica et al., 1998b; Di Baldassarre et al., 2010; Montanari, 2007). The usual procedure to estimate flood discharge for a chosen probability of exceedance is to deal with a rainfall-runoff model associating to risk the same return period of original rainfall, in accordance with the iso-frequency criterion. Alternatively, a flood frequency analysis to a given record of discharge data is applied, but again the same probability is associated to flood discharges and respective risk.

Moreover, since flood peaks and corresponding flood volumes are variables of the same phenomenon, they should be, directly, correlated and, consequently, multivariate statistical analyses must be applied. In general, multivariate probability models have been limited by mathematical difficulties due to the generation of consistent joint laws with ad hoc marginals. Actually, copulas has overcame many of these problems (Favre et al., 2004; Vandenberghe et al., 2012; Salvadori et al., 2007), as they are able to model the dependence structure independently of the marginal distributions (Nelsen, 1999).

METHODOLOGY

This section describes in details the procedure to derive flood hazard maps in a stochastic framework. Particularly, the process here developed has a modular structure consisting of different modules: synthetic hydrographs definition to gain the hydrological input to the hydraulic model by means two different approaches, an indirect approach, and a direct one, both based on copulas theory, transformation of flood discharge to inundated area through a two-dimensional hydraulic model, the flood hazard mapping.

Hydrological input assessment: indirect approach. The hydrological input was derived by using an indirect approach described, more in details, in Candela et al., (in press). The layout of the procedure can be resumed as follows: 1) stochastic generation via copulas of rainfall to derive rainfall events; 2) rainfall-runoff modeling for estimating the hydrological response at the outlet of the watershed using a conceptual fully distributed model; 3) derivation of design hydrographs (for given design return period) by bivariate analysis (copulas) of rainfall-runoff outputs.
Once extracted independent rainfall events from the available series of sub-hourly rainfall data, firstly, storm characteristics, in terms of storm duration and average intensity or rainfall volumes have been generated using a multivariate model since

they are variables of the same phenomenon. These variables have to be analyzed jointly through theory of copulas.

In the present study, Frank's family class of 2-copulas has been considered. It is a one parameter Archimedean copula:

$$C(u_1, u_2) = -\frac{1}{\theta} \ln \left[1 + \frac{(\exp(-\theta u_1) - 1)(\exp(-\theta u_2) - 1)}{\exp(-\theta) - 1} \right] \tag{1}$$

where θ is the parameter of copula function that is related to the Kendall's coefficient of correlation τ between X and Y through the expression $\tau = 1 - 4[\Delta_1(-\theta) - 1/\theta]$ where Δ_1 is the first order Debye function (Salvadori et al, 2007).

According to the nonparametric method, the first step in determining a copula is to obtain its generating function from bivariate observations using the procedure described by Genest and Rivest (1993). Regarding marginal distributions based on univariate data, here Exponential, Gamma, Weibull and Lognormal ones, were considered. The temporal patterns of rainfall for each event have been defined using mass curves derived by the hystorical events collected in the study area. The normalized events obtained are the input for selecting an appropriate probability function for the hyetograph shape. Here the choice has been orientated towards the Beta distribution because it is a very simple model that fits reasonably well the rainfall data.

Then, a conceptual fully distributed model with climatic dependencies for the flow routing is used (Candela et al., in press). It is a conceptual fully distributed model based on the representation in the form of linear kinematic mechanism of transfer of the full outflows coming from different contributing areas of the basin through the definition of a distributed hydrological response array with climatic characteristics. Rainfall inputs are, also, distributed in space and time-varying. The SCS-CN method (1986) is used here to transform the gross rainfall in effective rainfall. and a distributed unit hydrograph with climatic dependencies for the flow routing. Travel time computation, based on the definition of a distributed unit hydrograph, has been performed, implementing a procedure using flow paths determined from a Digital Elevation Model and roughness parameters obtained from distributed geographical information.

Hydrological input assessment: direct approach. Direct stochastic input of hydrograph was preferred to by pass the uncertainties related to rainfall–runoff modelling. Flood peaks and flood volumes should be analyzed through copula theory (Aronica et al., 2012). Firstly, generation of flood peaks and flow volumes has been obtained via copulas, then, the shape of hydrograph has been generated on the basis on significant flood events, via cluster analysis.

In the present study Gumbel–Hougaard family has been chosen. It is a one parameter Archimedean copula with generation function $\phi(t) = (-\ln(t))^{\theta}$ and $t = u_1$ or u_2:

$$C(u_1, u_2) = \exp\left(-\left((-\ln u_1)^{\theta} + (-\ln u_2)^{\theta}\right)^{1/\theta}\right) \tag{3}$$

where θ is the parameter of copula function that is related to the Kendall's coefficient of correlation between two random variables, X and Y, through the closed-form relation $\tau = 1 - \theta^{1}$. Once copula is known, Monte Carlo generation of pairs of discharge and volume (Q_{max}, V_{max}) can be carried out. The determination of flood hydrographs requires the knowledge of the shape to assign to pairs of Q_{max} and V_{max}. In this study flood hydrograph shapes were generated coping with non-dimensional hydrographs and cluster analysis. Finally, FDH have been obtained for a fixed RP by merging the non-dimensional hydrographs (for specific probability) and the generated peak-volume pairs derived using copula.

Flood propagation. For flood propagation the MLFP-2D (Multi Level Flood Propagation 2-D) model (Aronica et al., 1998a) was used. It is a hyperbolic model based on DSV equations when convective inertial terms are neglected. The conservative mass and momentum equations for two-dimensional shallow-water flow can be written as follows:

$$\frac{\partial H}{\partial t} + \frac{\partial p}{\partial x} + \frac{\partial q}{\partial y} = 0 \tag{4a}$$

$$\frac{\partial p}{\partial t} + gh\frac{\partial H}{\partial x} + ghJ_x = 0 ; \qquad \frac{\partial q}{\partial t} + gh\frac{\partial H}{\partial y} + ghJ_y = 0 \tag{4b}$$

where $H(t, x, y)$ is the free surface elevation; $p(t, x, y)$ and $q(t, x, y)$ are x- and y-components of the unit discharge; h is water depth; g the gravitational acceleration; and J_x and J_y are hydraulic resistances in the x- and y- directions.
The hydraulic resistances are parameterised by the Manning-Strickler formulation and can be expressed as:

$$J_x = \frac{n^2 p \sqrt{p^2 + q^2}}{h^{10/3}} ; \quad J_y = \frac{n^2 q \sqrt{p^2 + q^2}}{h^{10/3}} \tag{5}$$

where n ($m^{-1/3} \cdot s$) is the Manning's roughness factor. Equations (4) were solved by using a finite element technique with triangular elements. The free surface elevation is assumed to be continuous and piece-wise linear inside each element, where the unit discharges in the x and y directions are assumed to be piece-wise constant.

Global hazard indexes definition. Hazard analysis gives informations about the extent and the intensity of flood scenarios associating to an exceedance probability to it. It depends on a lots of parameters such as water depth, flow velocity, duration of flooding, product of water depth by flow velocity, rate of water rise, concentration of sediments, etc. (Merz et al., 2007). The former Italian Law 267/98 requires flood hazard and risk maps according to several probabilities: low probability (RP in the range 20-50 years), medium probability (RP in the range 100-200 years), high probability (RP in the range 300-500 years). In particular, the Sicilian Flood Risk Plan set RP equal to 50, 100 and 300 years. Here, global hazard indexes proposed by Aronica et al. (2012) are used (Table1) taking into account water depth, flow velocity and their product.

Table 1. Hazard classes based on depth of flooding (h) and return period (RP).

	Return period in years		
	50	*100*	*300*
h < 0.1 m	H_1	H_1	H_1
0.3 < h < 1.0 m	H_2	H_2	H_2
1.0 < h < 1.5 m	H_4	H_3	H_2
h > 1.5 m	H_4	H_4	H_4

CASE STUDY

The procedure is applied to a real case study where flood hazard and risk maps have been obtained and compared. The Imera catchment with an area of about 2000 km^2 is located in the south-western part of Sicily, Italy. The study was focused on the sub-catchment of Imera basin with an area of 1789 km^2 and delimited downstream by a flowgauge station named Imera at Drasi. The main watercourse is about 150 km in length and it winds from the central part of the island to the Mediterranean Sea near to the city of Licata (Figure 1). The measurement network (Fig. 1) managed by the Regional Hydrographic Service from 1960, consists of eight raingauges (Canicattì, Caltanissetta, Delia, Mazzarino, Enna, Riesi, Petralia Sottana, Polizzi Generosa), located within the catchment and characterised by a temporal resolution of 10 minutes, and of one level gauge (Drasi).
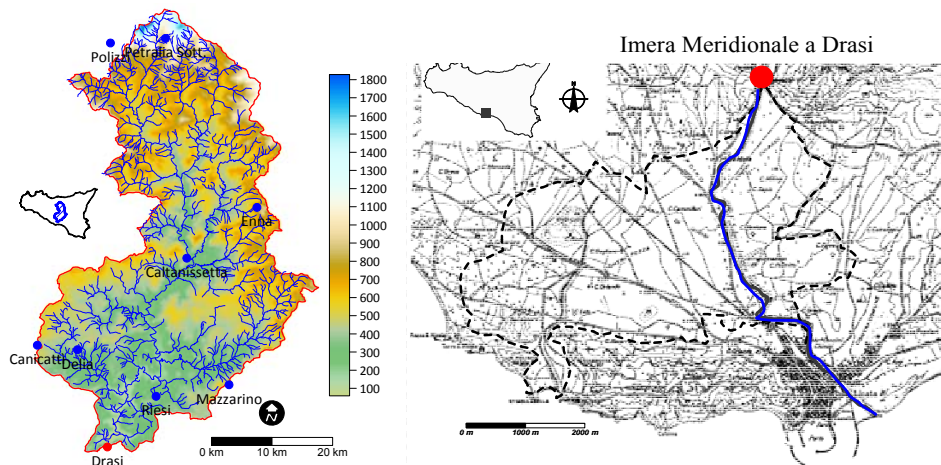

Figure 1. (left) Imera catchment layout, (right) Piana di Licata.

The alluvial areas (Piana di Licata), downstream the flowgage, are used intensively for agriculture activities with extended irrigation and many important transport facilities (railways, main roads, etc.). Residential areas and tourist accommodation are also established within the area at specific sites.

The plain has been subject to floods several times in the past decades with significant damages. For this reason, many studies considered this specific site (Aronica et al., 2002, 1998a,b) and, here, some results from these studies were used as basis for the application of the proposed procedure. Particularly, the latest severe inundation

occurred in October 1991, in that event, the rainfall duration was about 20 h, the total rain depth was 229 mm and the maximum intensity was 56 mm/h with a total rain volume of 225 Mm^3. The inundated area reported in Fig. 1 (right) was delimited with a field survey carried out a few days after the event.

RESULTS

Hydrological input. Regarding the direct approach, the hydrological input was derived using annual maximum discharges from 1960 to 2003 and 18 historical hydrographs available from Drasi station (Aronica et al., 2012). Firstly, θ parameter equal to 4.78 for the Gumbel-Hougaard copula and the generating function have been derived with a Kendall's coefficient of correlation τ equal to 0.79. Different goodness-of-fit criteria (AIC criterion, the Relative Root Mean Square Error – RRMSE, and the Anderson-Darling test) return LogNormal as best marginal distribution for flood volumes and GEV for peak discharges.
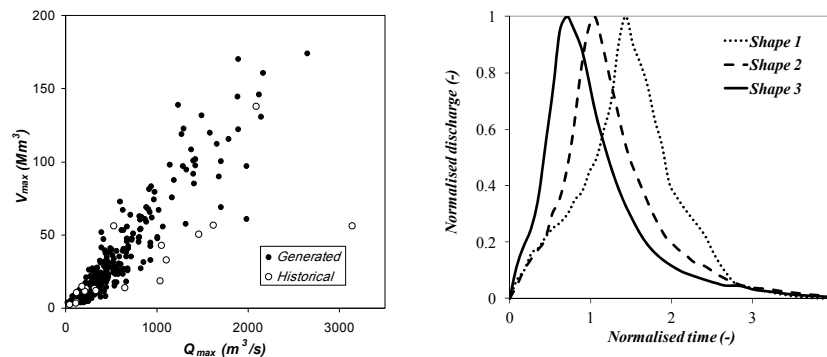

Figure 2. Q_{max}-V_{max} correlation (left) and non-dimensional clustered hydrographs (right). (adapted from Aronica et al., 2012)

Monte Carlo generation furnished 5000 pairs (Q_{max}, V_{max}), in Fig. 2(left) scatter plot of historical and generated values of peaks and volumes is shown. Shapes to assign to pairs of Q_{max} and V_{max} have been generated coping with non-dimensional hydrographs, generated on the basis on 18 significant flood events registered at Drasi station, and cluster analysis with Ward method (1963). Empirical hydrographs have been normalised that a unit peak flow and a unit flood volume have been resulted. The normalised hydrographs have been, then, grouped in various clusters according with Ward method (minimum variance algorithm that minimizes increments in sums of squares of distances of any two clusters that can be formed at each step). The results of this cluster analysis are the three shapes of hydrograph showed in Fig. 2(right). In relation to the number of hydrographs belonging to each cluster, a probability of about 0.11 (*Shape 1*), 0.5 (*Shape 2*) and 0.39 (*Shape 3*) has been assigned to these shapes.

Finally, three input hydrographs corresponding to a RP of 50, 100 and 300 years, for the propagation model, were obtained by merging the non-dimensional hydrographs (for specific probability) and the generated peak-volume pairs derived using copula (fig. 4a).

The capability of the indirect procedure was tested in reproducing the joint statistics of both peak discharges and corresponding discharge volumes through the generation of synthetic hydrographs starting from synthetic rainfall events of assigned shape, average intensity and duration obtained, in turn, by bivariate copula theory. Firstly, θ parameter for the Frank copula and the generating function have been derived in according to the procedure described in the previous paragraph, obtaining a θ value equal to -3.7573 and a Kendall's coefficient of correlation τ equal to -0.381.

The parameters of the marginal distributions considered were estimated using the Maximum Likelihood method and the best fitted distribution was selected using various criteria (AIC criterion, the Relative Root Mean Square Error – RRMSE, and the Anderson-Darling test). Lognormal probability and Weibull distributions represent best marginal distributions, respectively, for average storm intensity and for storm duration. Finally, the Beta distribution has been fitted to adimensional shape sample of each event. Model parameters are estimated by maximum likelihood (ML), while the goodness of fit was verified by the Pearson test. Finally, 5000 events were generated using Monte Carlo procedure; these events have shown an excellent reproducibility of historical events characteristics both in terms of duration-intensity correlation and in terms of adimensional shapes (Candela et al., in press).

The rainfall-runoff model calibration is only required for three parameters: c and CN_{II} for the effective rainfall module, and the hydraulic roughness k for the transfer module and has been carried out comparing observed and predicted discharges for the event of 21 December 1976, registered at Drasi flowgauge.

The latter two parameters are spatially distributed and their calibration has been carried out by considering both values as spatial correlation.

The resulting calibration parameter values were c equal to 0.68, CN_{II} equal to 87 and 75.8 $m^{-1/3}s$ for k. More details of calibration procedure are reported in Candela et al. (in press).
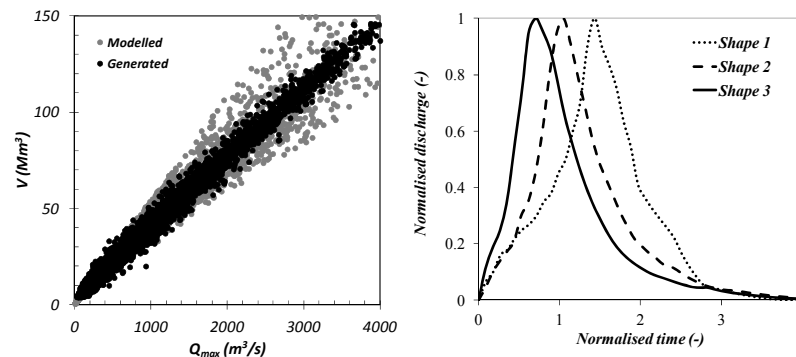

Figure 3. Q_{max}-V_{max} correlation (left) and non-dimensional clustered hydrographs (right).

Finally, Flood Design Hydrographs (FDH) via synthetic generation by using the output from the rainfall-runoff model, have been derived. It was carried out by following procedure: a) modeling of the statistical correlation between flood peak and volume pairs generated by the R-R model via copulas; b) definition of the normalized hydrograph shape in probabilistic form; c) final derivation of the FDH by rescaling the selected shape (i.e., for a fixed RP) given the synthetic flood peak and volume

values. Particularly, in fig. 4b flood design hydrographs corresponding to a Return Period (RP) of 50, 100 and 300 years are reported.

Monte Carlo generation furnished 5000 pairs (Q_{max}, V_{max}), in Fig. 3 (left) scatter plot of historical and generated values of peaks and volumes is shown. Shapes to assign to pairs of Q_{max} and V_{max} have been generated coping with non-dimensional hydrographs in the same way of the direct approach (fig.3 right).

Finally, three input hydrographs corresponding to a RP of 50, 100 and 300 years, for the propagation model were obtained by merging the non-dimensional hydrographs and the generated peak-volume pairs derived using copula (fig. 4b).
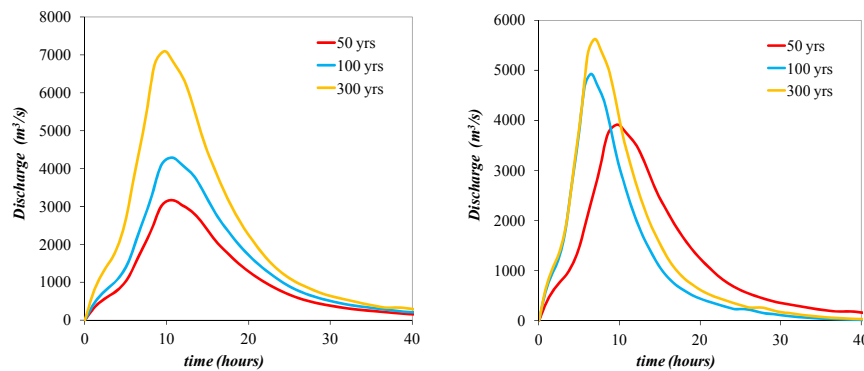

Figure 4. Flood Design Hydrograph corresponding to a RP = 50 (a), 100 (b) and 300 (c) years obtained from bivariate (a) direct approach a and (b) indirect approach.

Flood propagation. Model simulations were carried out using a triangular mesh that covers an area of about 22 km^2 discretized in 7168 triangular element. Elevations were derived from a DEM at 10 m resolution available for the study area. Vertical discontinuities have been placed along the river and the embankments of all the roads and the railway, with interruption at the culverts.

The unique parameter involved in the flood propagation model is the Manning's roughness coefficient; the domain was divided into two principal regions, floodplain and river and for both of these calibrated average roughness coefficient was assumed. In Aronica et al. (1998b) details about model calibration are reported. The resulting roughness coefficients values were, respectively, 0.037 m$^{1/3}$/s for the channel, and 0.051 m$^{1/3}$/s for the floodplain.

In Figure 5 hazard maps derived using the two hydrological inputs are reported. The map reports four hazard classes (Table1) calculated from hydrodynamic variables given by the model runs. The same figure also reports the hazard map derived using a univariate approach where flood hydrographs has been obtained by forcing the rainfall-runoff model with synthetic hyetographs from IDF curves with given return period and duration equal to the concentration time of the catchment.
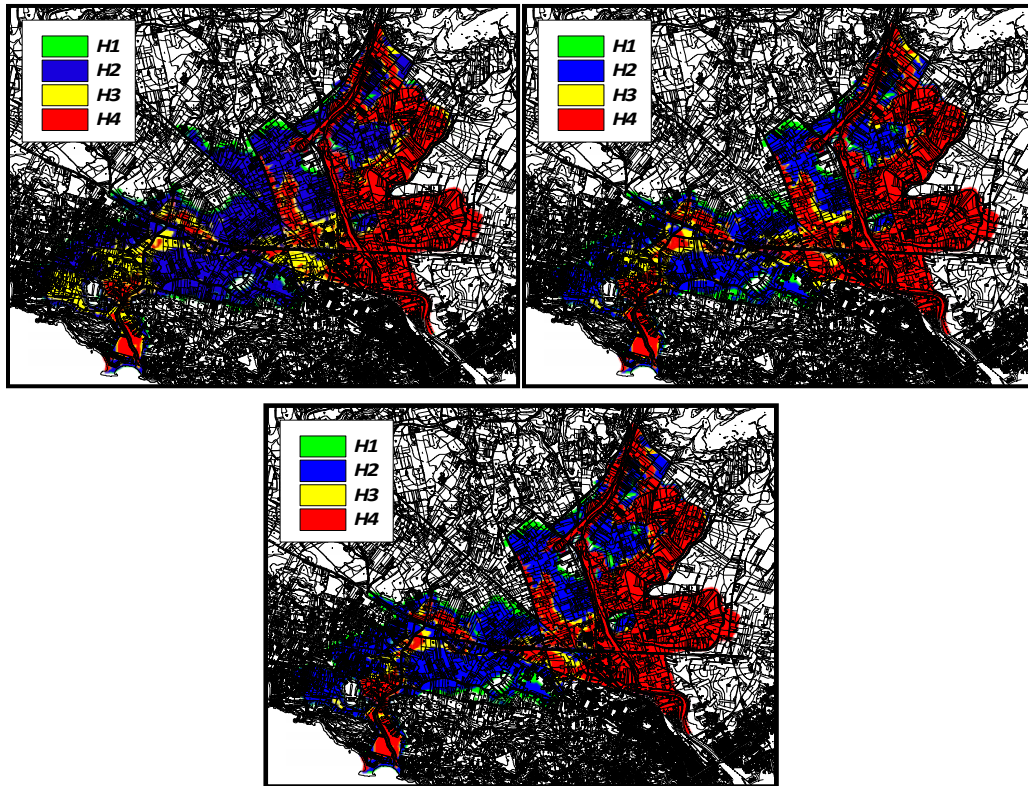
Figure 5. Hazard maps obtained using bivariate direct (top left), bivariate indirect (top right) and univariate (bottom) approach

To highlight the differences in the three approaches a comparison between hazard classes percentage are reported in Table 2. From these figure can be easily noticed the difference the bivariate and the univariate approach, the latter underestimates the H3 and H2 classes and overestimates the H4 class.

Table 2. Hazard classes percentage on flooded area

	Hazard			
	H1	*H2*	*H3*	*H4*
Bivariate direct	5.37	41.03	15.23	38.37
Bivariate indirect	6.88	33.74	10.09	49.29
Univariate	7.09	39.22	6.49	47.20

CONCLUSION

In this paper a new procedure to derive flood hazard maps including a complete probabilistic description of the hydrological inputs is presented. The novelty of the approach lies in considering the return time of the entire hydrograph and not only, a s usually, of the peak discharge. The results show the importance of evaluate flood hazard in a fully probabilistic framework and more research should be carried out also to include uncertainty, being these maps an essential tool to aid planning decisions for flood risk mitigation.

REFERENCES

Aronica, G. T., Tucciarelli, T., Nasello, C. (1998a). "2D multilevel model for flood wave propagation in flood-affected areas". *J Wat Res Plan Manag*, 124 (4), 210–217.

Aronica, G. T., Hankin, B., Beven, K. (1998b). "Uncertainty and equifinality in calibrating distributed roughness coefficients in a flood propagation model with limited data, *Ad. Wat. Resour.* 22 (4), 349–365.

Aronica, G. T., Bates, P. D., Horritt, M. S., (2002). "Assessing the uncertainty in distributed model predictions using observed binary pattern information within GLUE". *Hydrol. Proc.* 16, 2001–2016.

Aronica, G. T., Candela A., Fabio P. and Santoro M., (2012). "Estimation of flood inundation probabilities using global hazard indexes based on hydrodynamic variables", Phys Chem Earth, 42–44, 119–129.

Beven, K., Leedal, D., McCarthy S. (2011). "Framework for Assessing Uncertainty in Fluvial Flood Risk Mapping" *FRMRC Research Report SWP1.7*

Candela, A., Brigandì, G., Aronica, G. T., (in press). "Estimation of Flood Design Hydrographs using bivariate analysis (copula) and distributed hydrological modeling" *Nat. Hazards and Earth System Sciences*.

Di Baldassarre, G., Schumann, G., Bates, P., Freer, J., Beven, K.J. (2010). Floodplain mapping: a critical discussion of deterministic and probabilistic approaches, Hydrol. Sci. J., 55(3), 364-376.

Favre, A. C., El Adlouni, S., Perreault, L., Thiemonge, N. Bobee, B., (2004) "Multivariate hydrological frequency analysis using copulas", *Wat. Resour. Res.*, 40, W01101.

Genest, C., L.-P. Rivest (1993). "Statistical inference procedures for bivariate Archimedean copulas", *J. Am. Stat. Assoc.,* 88(423), 1034–1043.

Merz, B., Thieken, A. H., Gocht, M. (2007). "Flood risk mapping at the local scale: concepts and challenges", In: Begum, S., Stive, M.J.F., Hall, J.W., (Eds.), Flood Risk Management in Europe: Innovation in Policy and Practice, vol. 25. Series: *Advances in Natural and Technological Hazard Research*, vol. 13. Springer, Dordrecht, 231–251.

Montanari, A.(2007). "What do we mean by 'uncertainty'? The need for a consistent wording about uncertainty assessment in hydrology", *Hydrol. Proc.*21, 841-845.

Nelsen, R. B. (1999). "An introduction to copulas" *Lecture notes in statistics*. Springer-Verlag New York.

Salvadori, G., De Michele, C., Kottegoda, N. T., Rosso, R. (2007). "Extremes in Nature. An approach using copulas" *Wat. Sci. and Techn. Libr.* 56. Springer.

US Department of Agricolture, Soil Conservation Service (1986) National Engineering Handbook, Hydrology. Vol.4, Washington DC.

Vandenberghe, S., van den Berg, M. J., Gräler, B., Petroselli, A., Grimaldi, S., De Baets, B., and Verhoest, N.E.C. (2012). "Joint return periods in hydrology: a critical and practical review focusing on synthetic design hydrograph estimation", *Hydrol. Earth Syst. Sci. Discussion,* 9, 6781–6828.

Ward, J. E. (1963). "Hierarchical grouping to optimize an objective function", *J. American Statistical Assoc.* 58, 236–244.