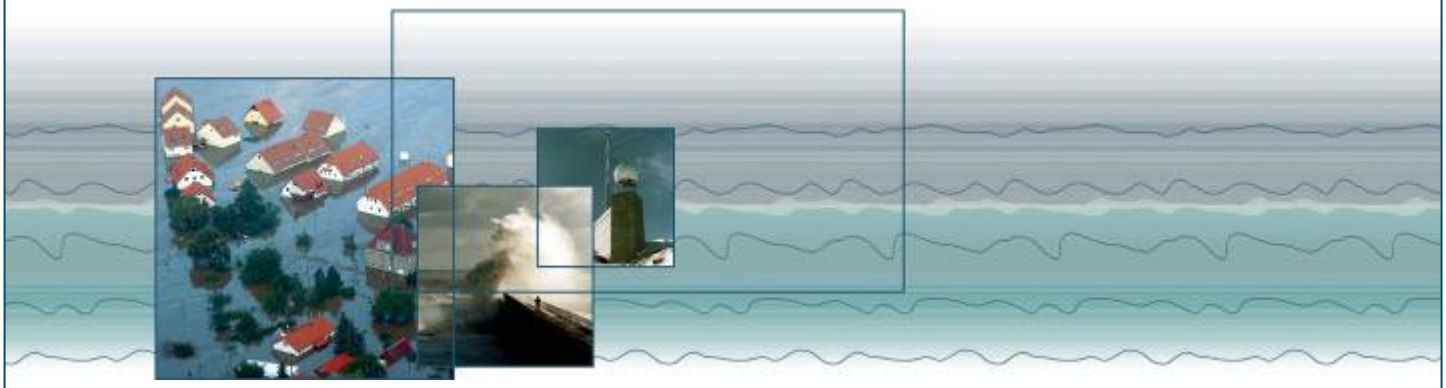
Integrated Flood Risk Analysis
and Management Methodologies



Realtime Guidance for Flash Flood Risk Management

Date March 2008

Report Number T16-08-02 Ed. 2
Revision Number 2_0_P01

Deliverable Number: D16.1
Due date for deliverable: February 2008
Actual submission date: March 2008
Task Leader UniPad

FLOODsite is co-funded by the European Community
Sixth Framework Programme for European Research and Technological Development (2002-2006)
FLOODsite is an Integrated Project in the Global Change and Eco-systems Sub-Priority
Start date March 2004, duration 5 Years

Document Dissemination Level

PU	Public	PU
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Co-ordinator: HR Wallingford, UK
Project Contract No: GOCE-CT-2004-505420
Project website: www.floodsite.net

SIXTH FRAMEWORK PROGRAMME

DOCUMENT INFORMATION

Title	**Real time guidance for flash flood risk management**
Lead Author	**Marco Borga (UniPad)**
Contributors	**Eric Gaume (ENPC), Mario Martina (UniBo), Jutta Thielen (JRC)**
Distribution	**Public**
Document Reference	**T16-08-02**

DOCUMENT HISTORY

Date	Revision	Prepared by	Organisation	Approved by	Notes
13/07/07	1_0_P16	Marco Borga	UniPad		First draft
05/09/07	1_1_P16	Marco Borga	UniPad		Revised Table of Contents and introduction
06/02/08	1_3_P16	Marco Borga	UniPad	All	Completed first draft
21/03/08	1_4_P16	Marco Borga	UniPad	All	Completed revised draft
06/04/08	1_5_P16	Marco Borga	UniPad		Revised draft ready for Theme leader revision
08/04/06	1_6_P01	Paul Samuels	HR Wallingford		Final Formatting, renamed file
10/02/09	1_7_P16	Marco Borga	UniPad		Final revision
19/05/09	2_0_P01	Jemima Rance	HR Wallingford		Formatting for publication

ACKNOWLEDGEMENT

The work described in this publication was supported by the European Community's Sixth Framework Programme through the grant to the budget of the Integrated Project FLOODsite, Contract GOCE-CT-2004-505420.

DISCLAIMER

This document reflects only the authors' views and not those of the European Community. This work may rely on data from sources external to the members of the FLOODsite project Consortium. Members of the Consortium do not accept liability for loss or damage suffered by any third party as a result of errors or inaccuracies in such data. The information in this document is provided "as is" and no guarantee or warranty is given that the information is fit for any particular purpose. The user thereof uses the information at its sole risk and neither the European Community nor any member of the FLOODsite Consortium is liable for any use that may be made of the information.

© Members of the FLOOD*site* Consortium

SUMMARY

A flash flood is a flood that follows the causative event in a short period of time and often is characterized by a sudden increase in level and velocity of the water body. The term "flash" reflects a rapid response to the causative event, with rising water levels in the drainage network reaching a crest within minutes to a few hours of the onset of the flood event, leaving extremely short time for warning. Thus, flash floods are localized phenomena that occur in watersheds with maximum response times of a few hours—that is, at spatial scales of approximately 3-4,000 km^2 or less, depending on the catchment characteristics. Most flash floods occur in streams and small river basins with a drainage area of a few hundred square kilometres or less.

The specific space and time scales which characterise the flash flood hazard have a marked influence on the dimensions of the flash flood vulnerability. This is typically represented by dispersed urbanization, transportation, tourism structures, as well as urbanised areas downstream of small basins (particularly along the Mediterranean coast). The dispersed spatial character of the vulnerability makes it difficult (and often unsustainable in ecological or economic terms) the management of the flash flood risk by means of structural measures, which aims to reduce flood volumes and peaks. Non structural measures, and more specifically flash flood forecasting and warning, are, by their nature, more suitable to cope with flash flood risk.

However, flash flood characteristics provide a challenge to traditional procedures for flood forecasting, mainly because of:

(i) the downscaling problem due to the incoherent space and time scales between atmospheric models and the flash-flood triggering processes;
(ii) the need to provide spatially distributed forecasts over the river network, rather than just at a few outlet river sections, and
(iii) the ungauged basin problem due to the fact that the small basins prone to flash-floods are rarely gauged and must be modelled without calibration.

It is therefore unrealistic to expect high levels of forecast reliability for localized thunderstorms occurring on small and medium size watersheds. However, flash flood forecasts do not necessarily need to be accurate to be effective. Indeed, effective flash flood warnings have been issued on the basis of flood warning systems which lack rainfall-runoff forecast models. However, without a rainfall-runoff model it is difficult to forecast the flood potential of storms that have complex space-time texture, particularly when the storm is near the flood/no flood threshold. Therefore, a problem exists about the most pertinent hydrologic model structure for flash flood forecast.

Two approaches are currently available which have been developed specifically for flash flood forecasting. In this report we aim to evaluate these approaches according with specific European hydroclimatic conditions.

The first approach is based on the concept of the statistical distributed (SD) model. The essence of this approach is to use the ensemble of antecedent model predictions (obtained from previous events) to rank the severity of the predictions of interest. The chosen (generic) distributed model is run by using archived radar-rainfall grids to derive flood probability characteristics of simulated flows for all cells in the distributed model. When subsequently the distributed model is run in forecast mode, the flooding flow threshold for each grid cell is defined in terms of a flood probability level rather than in an absolute value. In this manner, the flooding flow computed from simulated data is different than that computed from observed data because it takes into account the hydrologic model uncertainty. Because of this, the method:

- can be implemented initially using a priori rainfall-runoff parameters;
- can potentially provide benefits without requiring extensive model calibration;
- is particularly well suited for application in ungauged basins.

The second approach is focused on the representation of the soil moisture initial conditions across a landscape segmented into a network of small-size catchments. Local initial soil moisture is obtained by running continuous conceptual hydrological models, whose parameters are specified for the single (potentially ungauged) catchment by way of parameter regionalisation techniques. Based on these soil moisture estimates, the rainfall depth of a given duration which brings the basin to flooding is computed. This is the basis of Flash Flood Guidance (FFG), a methodology pioneered in the United States which provides flash flood forecasting and warning without explicitly modelling the event dynamics. The methodology allows one to identify with enough lead time the portions of a region which are characterised by high flash flood potential, given their high soil moisture conditions. Once the areas most exposed to potential flash flood occurrence have been identified, it is easier to develop a spatially articulated cascade of hazardous weather messages (Outlook, Watch, and Warning) linked to various lead times and pre-defined actions.

This Task provides
i) guidelines on effectiveness and limitations of use of both approaches in flash flood conditions,
ii) analysis of relevance to practice of results obtained, and
iii) indications of remaining gaps in knowledge.

CONTENTS

Document Information ii
Document History ii
Acknowledgement ii
Disclaimer ii
Summary iii
Contents v

1. Introduction ... 1
   1.1 Structure of the report ... 3

2. Evaluation of rainfall-runoff models for flash-flood forecasting ... 4
   2.1 Introduction: objectives and approach ... 4
      2.1.1 Rainfall-runoff modelling: a requirement for flash-flood forecasting ... 4
   2.2 The upper Loire river pilot area ... 4
   2.3 Performances of RR models in simulation and forecasting modes ... 6
      2.3.1 Introduction: the proposed approach ... 6
      2.3.2 Intercomparison of the conceptual RR models ... 10
      2.3.3 Comparison with the alternative models ... 11
      2.3.4 Forecasting performances of the tested models ... 13
      2.3.5 Conclusions ... 19
   2.4 Performances of RR models in simulation and forecasting modes ... 19
      2.4.1 Introduction ... 19
      2.4.2 Hourly point and spatial rainfall estimation error model ... 21
      2.4.3 Correlation of estimation errors over time ... 23
      2.4.4 Propagation of the estimation errors into RR models ... 25
      2.4.5 Conclusions ... 26

3. Flash-flood prediction based on discharge thresholds and high-resolution operational weather forecasts ... 27
   3.1 Introduction ... 27
   3.2 Hydrological model, input data and methodologies ... 27
      3.2.1 LISFLOOD model and set-up for this study ... 27
      3.2.2 Input data for the study ... 27
      3.2.3 Methodology of discharge threshold exceedances ... 27
   3.3 Brief description of the two case studies ... 28
      3.3.1 The 8-9th September 2002 event in the Cevennes-Vivarais region (France) ... 28
      3.3.2 The 29th August 2003 case in the Eastern Italian Alps ... 29
   3.4 Results ... 29
      3.4.1 Long-term simulations and calculation of thresholds ... 29
      3.4.2 Forecasting the 8-9th September 2002 event ... 31
      3.4.3 Forecasting the 29th August 2003 event in Italy ... 33
      3.4.4 Six months assessment of forecast performance for the French case study ... 35
   3.5 Summary and Conclusions ... 35

4. Flash flood warning based on rainfall depth-duration thresholds and soil moisture conditions: An assessment under European conditions ... 37
   4.1 Introduction ... 37
   4.2 Study Areas and Data ... 38

4.2.1 Rationale for basins selection ... 43
4.3 The threshold-base technique ... 43
4.3.1 Description of the hydrological model ... 43
4.4 Assessment methodology ... 45
4.4.1 Model application ... 45
4.4.2 FFG assessment ... 48
4.4.3 Threshold flooding conditions ... 50
4.5 Results ... 50
4.5.1 Model application ... 50
4.5.2 FFG assesment ... 52
4.6 Conclusions and future works ... 58

5. Flash flood warning based on Bayesian rainfall thresholds ... 60
5.1 Introduction ... 60
5.2 The Bayesian Rainfall Thresholds using Monte-Carlo (BRT-MC) ... 60
5.2.1 Phase 1 - The rainfall thresholds estimation ... 60
5.2.2 Phase 2 - Operational use of the rainfall threshold approach ... 64
5.3 The Bayesian Rainfall Thresholds using the Normal Quantile Transform (BRT-NQT) ... 64
5.4 Application on the Posina Catchment of the BRT-MC ... 65
5.4.1 The "risk- averse" case ... 68
5.4.2 The "risk- prone" case ... 70
5.4.3 The "real" case ... 71
5.5 Considerations ... 73

6. Conclusions and recommendations ... 75

7. References ... 78

8. Appendix 1: Presentation of the various Rainfall-Runoff models used in Section 2 ... 83
8.1 CREC model ... 83
8.2 GR4J model ... 83
8.3 HBV model ... 83
8.4 IHAC model ... 84
8.5 Artificial Neural networks ... 84

Tables

Table 2-1: List and characteristics of the upper Loire river gauging stations 5
Table 2-2: Comparison of the Characteristic time of forecast and the estimated times of concentration of the watersheds. 18
Table 2-3: Percentage of computed interpolation errors comprised in various theoretical confidence intervals for four test rain gauges. 23
Table 2-4: Standard deviations of the estimation relative error of mean intensities over various durations: 1) observed and 2) modelled. Constant correlation coefficient equal to 0.6. 24
Table 2-5: Proportion of the observed values comprised in the 90% simulated confidence interval, with and without a 20% tolerance. 26
Table 3-1: Definition of contingence table 30
Table 4-1: Basins characteristics 41
Table 4-2: Periods with data available: parent basins 41
Table 4-3: Periods with data available: interior points 42
Table 4-4: Relationship among parent and nested basins 42
Table 4-5: Four-cell contingency table used in the study 49
Table 4-6: Model validation and calibration results. Parent basins 51

Table 4-7: Model validation results. Interior Points 51
Table 5-1: Parameters of the cost functions in the different cases 67
Table 5-2:Two-by-two contingency table for the assessment of a threshold based forecasting system 67

Figures

Figure 2-1: Upper Loire Watershed upstream the Bas-en-Basset streamgauge and location of the hourly rain gauges and the stream gauges. 5
Figure 2-2: Presentation of the POD, FAR and CSI statistics 9
Figure 2-3: Nash criterions obtained on the two validation periods with the six tested conceptual RR models: median, maximum and minimum Nash values for the 11 watersheds. RR model without update or with an update based on the past measured discharges. 10
Figure 2-4: Calibration and validation Nash criterions for neural networks on the various watersheds. 1-hour prediction. Median, maximum and minimum Nash values for 10 calibration trials. N is the number of hidden neurones. 12
Figure 2-5: Calibration and validation Nash criterions for neural networks on the various watersheds. 3-hour prediction. Median, maximum and minimum Nash values for 10 calibration trials. N is the number of hidden neurones. 13
Figure 2-6: Comparison between the discharge variation forecasting efficiency (Nash criterions on discharge variations) obtained with the 3 tested types of models for 3 forecasting horizons. Blue (positive Nash), red (negative Nash) 14
Figure 2-7: Some examples of forecasted hydrographs 15
Figure 2-8: Average, maximum and minimum forecasting performances of the various tested RR models for the 11 upper Loire river watersheds. Rainfall is supposed to be perfectly known in advance. 17
Figure 2-9: Evolution of the forecasting NSE criterions with the lead-time and impact of the uncertainties concerning the future rainfalls. Conceptual RR model (dots and thick black line). 18
Figure 2-10: Estimated kriging interpolation relative standard error for hourly and daily rainfall amounts over the upper Loire river area. 20
Figure 2-11: Evolution of the percentage of the upper Loire river area where the theoretical kriging relative standard error is lower than 0.5 or 0.7. 21
Figure 2-12: Example of a comparison between the theoretical and the observed distributions of the standardized rainfall estimation errors for four rain gauge of the network. 22
Figure 2-13: Example of comparison between observed (crosses and triangles) and simulated (lines) interpolation error quantiles for rainfall accumulations over 1 to 24 hours. Same rain gauge and two tested correlation coefficients. 23
Figure 2-14: Two examples of observed flood hydrographs (red lines), 90% confidence interval for the simulated discharges (blue lines) and limits of the simulated discharge values obtained after 20 Monte Carlo runs. 25
Figure 3-1: Map of the study area and zoom into the Cevennes region with the catchments used for this study. 29
Figure 3-2: Map of thr Eastern Italian Alps with the Tagliamento river basin highlighted. The station Udinese used in the report is highlighted by a red circle 29
Figure 3-3: Scatterplots for stations in the Ardeche (a) and Gard (b) with observed discharges in m3/s on the x-axis and simulated discharges in m3/s based on synop data on the y-axis. 30
Figure 3-4: Contingence tables of hit, false, and missed threshold exceedances for 1 station in the Ardeche (gray), and Gard (shaded) 31
Figure 3-5: Hydrographs of observed and forecasted discharges in m3/s (y-axis) for the station Gard 104 in the Gard river from the 9th-12th Sep 2002 starting at 12:00 in hourly time steps (x-axis). The flood forecasts based on DWD forecasts start on 0906 00:00 until 0909 00:00 in 12h time intervals. 32

Figure 3-6: Visualisation of threshold exceedances in hourly time steps for the flood forecasts based on radar data and 48 hours DWD Lokalmodell forecasts from 20020907 00:00, 20020907 12:00, 20020908 00:00, 20020908 12:00 and 20020909 00:00. The time is indicated in 6 hour intervals at the top. The exceeded thresholds are colour coded as purple (severe), red (high), yellow (medium) and green (low). 32
Figure 3-7: Summary threshold exceedance maps showing the highest threshold exceeded during the 48 h forecasting time for flood forecasts based on the DWD Lokallmodell weather forecasts from 20020907 00:00 and in 12 hourly steps until 20020909 12:00. The threshold exceedances are colour coded with purple (severe), red (high), yellow (medium) and green (low). 33
Figure 3-7 illustrates the spatial development of forecasted event. Each panel shows the maximum alert threshold exceeded during the forecasting period in 12 hourly steps. The panel clearly illustrates that the event is first forecasted on the 07th 12:00 to take place in the upstream areas of all 4 river basins. In the next forecasts the emphasis is mostly in the Gard and Ceze rivers and less in Ardeche and Virdourle. The panel shows how the flooding is forecasted to affect almost the whole basins well exceeded the severe thresholds over large parts of the river basins. Downstream, towards the outlets, mostly only high thresholds are exceeded. 33
Figure 3-8: Summary threshold exceedance maps showing the highest threshold exceeded during the 48 h forecasting time for flood forecasts based on the DWD Lokallmodell weather forecasts from 20030827 00:00 and in 12 hourly steps until 20030829 00:00. The threshold exceedances are colour coded with purple (severe), red (high), yellow (medium) and green (low). 34
Figure 3-9: Hydrographs of observed and forecasted discharges in m3/s (y-axis) for the station Udinese in the Tagliamento basin. The flood forecasts based on DWD forecasts start on 20030825 00:00 until 20030630 00:00 in 12h time intervals. 35
Figure 4-1: Study basins and their location in France and Italy 40
Figure 4-2: Plot of mass balance data from the study basins on the Budyko curve. 42
Figure 4-3 a-e: One year (01.01.1993-31.12.1993) of hourly results at the outlet for the Cordevole river at Vizza (a) simulation results with model parameter transposition from Cordevole at Saviner; (b) simulation residuals, (c) simulation results with model parameter and soil moisture transposition from Cordevole at Saviner; (d) simulation residuals (e) relative soil moisture content of the PDM storage obtained from model parameter transposition at Vizza and from model simulation at Saviner. 48
Figure 4-4a-c. Evaluation of the threshold-based technique for: a) parent basin with model calibration; b) interior basins with model parameter transposition; c) interior basins with both model parameter ans soil moisture status transposition. POD and FAR scores are plotted. 53
Figure 4-5a-c. Evaluation of the threshold-based technique for: a) parent basin with model calibration; b) interior basins with model parameter transposition; c) interior basins with both model parameter and soil moisture status transposition. CSI scores are plotted. 54
Figure 4-6a-c: Overall assessment of the threshold-based technique for: a) parent basin with model calibration; b) interior basins with model parameter transposition; c) interior basins with both model parameter and soil moisture status transposition. POD and FAR scores are plotted. 56
Figure 4-7a-c: Overall assessment of the threshold-based technique for: a) parent basin with model calibration; b) interior basins with model parameter transposition; c) interior basins with both model parameter and soil moisture status transposition. CSI scores are plotted. 57
Figure 5-1: Schematic representation of the proposed methodology. (1) Subdivision of the three synthetic time series according to the soil moisture conditions (AMC); (2) Estimation of the joint pdfs between rainfall volume and water stage or discharge; (3) Estimation of the “convenient” rainfall threshold based on the minimisation of the expected value of the associated utility function 61

Figure 5-2: The synthetic time series and the three time values to be defined 62
Figure 5-3: Expected value of the utility $E(U(q,v|T,T))$ for a given accumulation time T. 63
Figure 5-4: Rainfall thresholds plotted as a function of the rainfall accumulation time. 63
Figure 5-5: Evaluation of the different FFG methods. 65
Figure 5-6: Utility/cost function for the "risk-averse" case 69
Figure 5-7: the rainfall threshold for the "risk-averse" case 69
Figure 5-8: Utility/cost function for the "risk-prone" case 70
Figure 5-9:The rainfall threshold for the "risk-prone" case 71
Figure 5-10: Utility/cost function for the "real" case 72
Figure 5-11: The rainfall threshold for the "real" case 73
Figure 5-12: Comparison of the rainfall thresholds for the examined cases 74

1. Introduction

Effective response to the problem posed by flash floods requires a clear characterisation of the **physical** and **social** components of the risk. Flash flood data collected recently for different hydro-climatic regions in Europe[1] and for the last decades (Gaume et al., 2007) show that flash floods have the following characteristics:

- they occur suddenly, with water levels in the drainage network reaching a crest within 1 to 10 hours after the onset of the rain event, leaving extremely short time for warning;
- they are generated by intense rainfall that trigger the dispersed response of watersheds of some tens to some hundred kilometres;
- they are characterised by high kinetic energy, resulting in a high threat to life and severe specific (per unit of area) damages to property and infrastructure (see for instance Ruin et al. (2007) for a description of the loss of life circumstances in flash floods);
- they are locally rare (Borga et al., 2007; Norbiato et al., 2007) and poorly documented, both in terms of physical processes involved and in terms of losses and social vulnerability analysis.

The examination of flash flood regimes across Europe shows that space and time scales of flash floods change systematically when moving from Continental to Mediterranean regions, while seasonality shifts accordingly from summer to autumn months (Bain et al., 2008). This has several hydrological implications, which need to be considered, for example, when examining potential effects of land use (urbanisation, deforestation, afforestation) and climate change on flash floods.

The above physical factors depending on exposure interact with the social, economical and environmental dimensions of the flash flood vulnerability. This is typically represented by dispersed urbanization, transportation, tourism structures, as well as urbanised areas downstream of small basins (particularly along the Mediterranean coast) (De Marchi et al., 2007). The variability in flash flood regimes combines with differences in social, economical and institutional characteristics at local, regional and national level to generate a complex pattern of impact, response and mitigation policies to flash flood risk across Europe.

Several important consequences arise in term of risk management strategies from the characterisation of hazard and vulnerability of flash floods:

- Because the elements at risk are highly dispersed, the management of the flash flood risk by means of structural measures aiming to reduce flood volumes and peaks is difficult (and often unsustainable in ecological or economic terms). Region-wide river training measures and land-use planning have an important role in flash flood risk management, particularly when it is associated to debris flows. Even in these cases, however, an often unquantified degree of residual risk will remain, which require 'acceptable' flood risk to be determined and mitigation solutions to be implemented.
- Flash flood forecasting, warning and emergency management are, by their nature, suitable to cope with the residual risk (Romang et al., 2008). Advancements in precipitation and flash flood forecasting are essential to improve emergency management. However, a focus on advances in forecasting alone will not be sufficient to reduce casualties and damages. A better understanding of the behaviour of people exposed to risk and of the organisation of warning and rescue services is also essential. This understanding includes perceptions of risk and how such risk is communicated (Handmer, 2001).

[1] Three European hydroclimatic regimes were considered: Mediterranean (southern Europe and Mediterranean islands, no Alpine influence); Alpine Mediterranean (southern Europe); inland continental (central and eastern Europe).

- Specific preparedness strategies are necessary. The local characteristics and sudden nature of occurrence of flash floods are best managed by the local authorities with active and effective involvement of the people at risk and with effective coordination between local, regional and national level. The time available for communication is very limited and typically there is no time for learning as the flood develops. The preparedness strategies must capitalise on improvements in flash flood forecasting and warning and, at the same time, to adapt to the large uncertainties affecting these forecasts (Parker et al., 2007).
- Long term risk management for flash flood needs to address the tensions between risk management and economic development. Embracing scenarios of the future (including climatic, demographic and socio-economic changes) within the decision making process is required to identify precautionary, sustainable and adaptable risk mitigation policies and strategies. Given the unique characteristics of flash floods and the large uncertainties affecting the long term risk assessment, a specific framework should be developed for risk communication with local stakeholders (Faulkner et al., 2007).

This report focuses on development and evaluation of flash flood forecasting and warning techniques, within the broader framework of flash flood risk management. Linkages to integrated flash flood risk management are developed in Task 23 of FLOODsite.
The observations reported above lead to characterisation of flash flood forecasting with respect to plain flood forecasting. In essence, the twofold consequences of the space/time scales of occurrence of flash flood are that forecasting of flash-floods:

- **depends critically on meso-scale storm forecasting, with a specific attention to the processes leading to slow movement of the precipitation system;**
- **necessitates real time hydrological modelling, with a specific attention to the runoff generation processes over a wide range of scales.**

Although they are seldom all deployed at the same time, the technical requirements for a hydrometeorological flash flood forecasting system include:

- a numerical weather prediction (NWP) model, capable to provide short-range Quantitative Precipitation Forecasts (QPF),
- a remote sensing based (radar, satellites) precipitation detection system, for storm monitoring and for the possible initialization and conditioning of the NWP model, and
- a hydrological-hydraulic forecasting model, capable to forecast the stream response from the rain input.

Whereas progress has been made in the last decade in the integration of meteorological forecasts and radar observations in flash flood surveillance, lack of observations hamper advances on understanding the hydrological processes at work during flash floods, and, consequently, on forecasting the stream response to extreme precipitations. Due to this reason, this report focuses on the hydrological aspects of flash flood forecasting, and in particular on examination of results obtained with flash flood forecasting on small scale catchments which are considered ungauged or poorly gauged. **It should be reminded here that assessment of results obtained from the integrated forecasting chain, including output from NWP models and rainfall nowcasting techniques, is provided in Task 23 for a number of case studies**.

Flash flood forecasting challenges traditional forecasting procedures because:

- the short lead time available, which implies both the integration of meteorological and hydrologic forecast, and the development of procedures which afford early identification of the areas, within broader regions, which are more susceptible to flash flood development;
- the need to provide spatially distributed forecasts over river networks, rather than just at a few river sections, and
- the ungauged basin problem due to the fact that the small basins prone to flash flood are rarely gauged and must be modelled without calibration.

Two approaches are currently available which have been developed specifically for flash flood forecasting. In this report we aim to evaluate these approaches according with specific European hydroclimatic conditions.

The first approach is based on the concept of the statistical distributed (SD) model. **The essence of this approach is to use the ensemble of antecedent model predictions (obtained from previous events) to rank the severity of the predictions of interest.** The chosen (generic) distributed model is run by using archived radar-rainfall grids to derive flood probability characteristics of simulated flows for all cells in the distributed model (Reed et al., 2007). When subsequently the distributed model is run in forecast mode, the flooding flow threshold for each grid cell is defined in terms of a flood probability level rather than in an absolute value. In this manner, the flooding flow computed from simulated data is different than that computed from observed data because it takes into account the hydrologic model uncertainty. Because of this, the method:

- can be implemented initially using a priori rainfall-runoff parameters;
- can potentially provide benefits without requiring extensive model calibration;
- is particularly well suited for application in ungauged basins.

The second approach is focused on the representation of the soil moisture initial conditions across a landscape segmented into a network of small-size catchments. Local initial soil moisture is obtained by running continuous conceptual hydrological models, whose parameters are specified for the single (potentially ungauged) catchment by way of parameter regionalisation techniques (Norbiato et al., 2008). Based on these soil moisture estimates, the rainfall depth of a given duration which brings the basin to flooding is computed. This is the basis of Flash **Flood Guidance (FFG)**, a methodology pioneered in the United States (Carpenter et al., 1999; Reed et al., 2004; Georgakakos, 2004, 2006) which provides flash flood forecasting and warning without explicitly modelling the event dynamics. The methodology allows one to identify with enough lead time the portions of a region which are characterised by high flash flood potential, given their high soil moisture conditions. Once the areas most exposed to potential flash flood occurrence have been identified, it is easier to develop a spatially articulated cascade of hazardous weather messages (Outlook, Watch, and Warning) linked to various lead times and pre-defined actions.

1.1 Structure of the report

Given the objectives stated in the Introduction, the structure of the report is articulated as follows.

Section 2 provides results from application of conceptual semi-distributed rainfall-runoff models under flash flood conditions. These models are frequently used by operational services for flood forecasting. It is therefore of considerable practical importance to examine the suitability of this type of models for flash flood forecasting, and to identify the major sources of uncertainty.

Section 3 reports about the development and application of a threshold-based approach, which aims to make best use of model results even modelled flows are not a perfect match for reality and to use in the most efficient way of the few data available in flash flood forecasting. The approach requires running a distributed model using archived radar-rainfall grids to derive flood probability characteristics of simulated flows for all cells in the distributed model (Reed et al., 2004). When subsequently running the distributed model in forecast mode, the flooding flow threshold for each grid cell is defined in terms of a flood probability level rather than an absolute value of flow. In this manner, the flooding flow computed from simulated data is different than that computed from observed data because it takes into account the hydrologic model uncertainty.

Section 4 evaluates a threshold-based flash flood warning approach, by considering a wide range of climatic and physiographic European conditions, and by focusing on ungauged basins. The system is derived from the Flash Flood Guidance (FFG, hereafter) approach. The FFG is the depth of rain of a given duration, taken as uniform in space and time on a certain basin, necessary to cause minor flooding at the outlet of the considered basin. This rainfall depth, which is computed based on a lumped hydrological model, is compared to either real time-observed or forecasted rainfall of the same duration and on the same basin. If the nowcasted or forecasted rainfall depth is greater than

the FFG (i.e., the Flash Flood Threat – FFT - is greater than zero or greater than a given threshold), then flooding in the basin is considered likely.

Section 5 aims to develop and evaluate a threshold-based technique based on the minimisation of a Bayesian Loss Function of the discharge conditional upon the state of saturation of the catchment.

Specific conclusions are reported at the end of each Section. Indications for future works are reported as general conclusion of the report.

2. Evaluation of rainfall-runoff models for flash-flood forecasting

2.1 Introduction: objectives and approach

2.1.1 Rainfall-runoff modelling: a requirement for flash-flood forecasting

Heavy rainfall accumulation is a necessary but not sufficient condition for flash floods, since hydrology critically controls flash-flood-triggering. Without hydrological analysis, it is impossible to evaluate the flood potential of storms, particularly in the fringe of the flood/no flood threshold.
Semi-distributed, conceptual rainfall-runoff models are often used for flood forecast by operational services. It is therefore of considerable practical importance to examine the suitability of this type of models for flash flood forecasting, and to identify the major sources of uncertainty. Here, an evaluation under flash flood conditions is reported for six different continuous models, rather parsimonious in structure (number of parameters ranging from 4 to 8). The models have been applied on 11 sub-basins of the Loire basin, in central-southern France, with drainage area ranging from 20 to 3234 km^2. The conditions are typical of flash flood basins, with a rather good raingauge density (1 station in average per 80 km^2).

It should be recognized that only one of the flash flood requirements is considered here: application to relatively small-scale catchments with relatively intense storm events. Other requirements concerning capability to provide spatially distributed flow forecast and capability to deal with ungauged basins are not considered in this section.

2.2 The upper Loire river pilot area

The upper Loire River is located in the northern part of the Cevennes-Vivarais Hydro-Meteorological observatory. It is an upland, mainly rural area with a bedrock dominantly composed of plutonic and metamorphic formations and locally volcanic stones. The soils are relatively shallow. The upper Loire area is regularly exposed to severe flash floods. The September 1980 flash flood, the major flood of the upper Loire for the last 50 years, killed for instance 8 people in Brives Charensac, just downstream the Coubon gauging station (Table 2-1). Its estimated peak discharge in Coubon is 2000 m^3/s (about 2,3 $m^3/s/km^2$ for 865 km^2). This value is close to the maximum ever reported peak discharge values in France for watersheds of similar areas (3000 m^3/s for the Erieux (850 km^2) flood of the 10/09/1957 or 3000 to 3500 m^3/s for the Gardon flood in Anduze (540 km^2) on the 30/09/1958 and the 09/09/2002).

The upper Loire river watershed at Bas-en-Basset covers 3234 km^2. The river gauging network, established for flood forecasting purposes, counts 11 active stations. The density of the rain gauge network has progressively increased over the years: the number of hourly rain gauges has grown from 6 in the 70's to 40 nowadays. The hourly rain gauge network is relatively dense (about 1/80 km^2) if compared to the average density of hourly rain gauges over France 1/500 km^2.

A meteorological Radar, located in Sembadel in the North of the upper Loire watershed, has been in operation since 1996. Due to technical problems, especially elevation of the Radar and masking

of the Radar beam by surrounding trees, it has not been possible until now to use this Radar in a quantitative way to estimate rainfall rates. Spatial average rainfall amount estimations can therefore only rely on the rain gauge network in this area. Kriging has been used for the spatial interpolation of the measured point hourly rainfall rates and the estimation of spatial averaged hourly rainfall amounts on the sub-watersheds, inputs of the rainfall-runoff models (see Section 2.4).

Table 2-1 summarizes the main characteristics of the 11 stream gauges, of their upstream watersheds and of the available discharge series. The data set covers a large range of watershed types and sizes. The Mediterranean storms inducing flash-floods affect the south-east part of the area (Tauron, Gazeille, upper Loire, Lignon). This Mediterranean influence explains the very high decennial specific peak discharge values of these watersheds. While the north-western part of the upper Loire river (Gagne, Borne, Dunière) is under a typical oceanic climatic influence with moderate floods. Snowfall and snowmelt are important factors of the hydrologic budget on the south-east part of the upper-Loire but have little influence on the major floods that predominantly occur in autumn. They will therefore, for sake of simplicity, not be taken into account in the tested rainfall-runoff models.

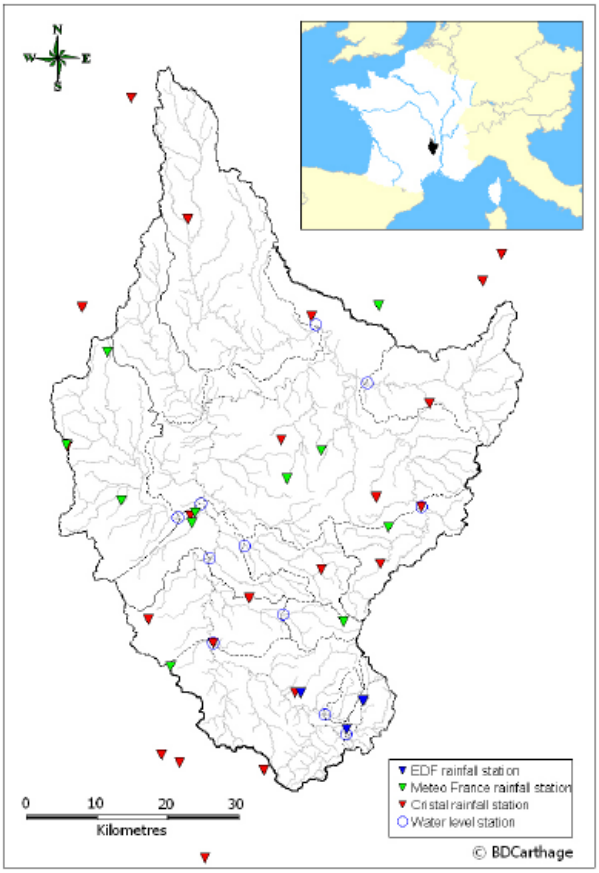


Figure 2-1: Upper Loire Watershed upstream the Bas-en-Basset streamgauge and location of the hourly rain gauges and the stream gauges.

Table 2-1: List and characteristics of the upper Loire river gauging stations

River	Gauging station	Watershed area (km^2)	Approx. time of conc. (hours)	Period of record	Runoff (mm)	Decennial peak discharge ($m^3/s/km^2$)
Tauron	Cros (**Cros**)	20	2-3	1983-2002	1100	4.5
Gazeille	Besseyre (**Bess**)	51	2-4	1990-2003	900	1.4

Loire	Rieutord (**Rieu**)	62	3-5	1983-2002	1380	5.6
Gagne	Pandreaux (**Pand**)	107	3-5	1998-2003	470	
Lignon	Chambon (**Cham**)	139	5-7	1977-2003	705	1.4
Duniere	Vaubarlet (**Vaub**)	228	8-10	1991-2003*	430	0.3
Borne	Espaly (**Espa**)	375	10-13	1985-2003*	305	0.3
Loire	Goudet (**Goud**)	432	5-8	1982-2003	410	0.9
Loire	Coubon (**Coub**)	732	8-12	1998-2003	350	
Loire	Chadrac (**Chad**)	1310	10-13	1977-2003	420	0.5
Loire	Bas-en-Basset (**Bas**)	3234	20-24	1977-2003	370	0.4

** mostly average daily discharges available except for the major flood events.*

The assessment of the efficiency and the improvement of the existing flood forecasting tools is a major concern for the local authorities. Efficient tools would be essential to protect the local stakes and people as illustrated by the 1980 flood and more recently by the 1996 and 2003 floods. They are also needed to manage efficiently the Villerest flood attenuation dam located downstream Bas-en-Basset that protects major cities as Nevers, Orléans or Tours.
Moreover, the upper Loire hydrologic network is considered as one of the very well maintained operational network in France.

For the these three reasons - i.e. a region exposed to flash floods, high operational stakes and relatively good quality data – the upper Loire appeared to be a particularly well suited area to test the efficiency of rainfall-runoff models in forecasting flash floods.

2.3 Performances of RR models in simulation and forecasting modes

2.3.1 Introduction: the proposed approach

The tested conceptual rainfall-runoff models

A large body of scientific literature has been devoted to rainfall-runoff modeling. Most of the authors came to the conclusion that the data sets available in hydrology can only support the development of models with limited complexity – i.e. the calibration of models with a limited number of parameters, typically 4 to 8 (Perrin et al., 2001; Gaume et al., 1998; Jakeman & Hornberger, 1993).

Based on this, we decided to test simple lumped conceptual rainfall-runoff models running on a continuous basis. The possible usefulness of distributed hydrologic models will be discussed at the end of Section 2.4. Six different models, with numbers of parameters ranging from 4 to 8, have been selected:

- The CREC model already applied on the Loire river (EDF, 1972),
- The GR4 model, widely used in France (Perrin et al., 2003),
- The HBV model (Bergstrom, 1995) and IHAC model (Jakeman et al., 1990),
- Two versions of the Topmodel (Beven & Kirkby, 1979).

The diagrams of some of these models are given in the Appendix 1. An inter-annual averaged decadal potential evapo-transpiration has been used as input for the models, since evapo-transpiration is known to have little influence on the performances of RR models (Oudin et al., 2005).

The rainfall-runoff models were used in three different modes:

- **Mode 1:** the **standard simulation mode**, the input of the model being the rainfall amounts and evapotranspiration values up to the date and time of the forecast.

- **Mode 2: Forecasting mode with updating and known rainfall**. To be efficient forecasting tools, RR models have to be updated based on the last known and simulated discharge values. A large variety of updating procedures have been proposed: the model parameters or its state variables or even its possible input variable correction factors can be updated taking into account the last forecasting errors, or a model forecasting the future modeling errors can be added to the RR model. The results obtained with these various approaches seem to be generally similar (Yang & Michel, 2000). We chose here the simplest updating procedure: the last known error value is added to the simulated discharge. It is equivalent to consider that the RR model forecasts not directly the discharge over the forecasting horizon but the discharge variation.

$$Q^f_{t+i} = Q^s_{t+i} + (Q_t - Q^s_t) = Q_t + (Q^s_{t+i} - Q^s_t) \quad (2\text{-}1)$$

  where Q^f_{t+i} stands for the forecasted discharge for time $t+i$, Q^s_{t+i} is the discharge simulated by the RR model for time t+i and Q_t is the observed discharge at time t, i being the forecasting horizon.

- **Mode 3: Forecasting mode with updating and unknown rainfall**. This is the same as previously but the future rainfall intensity values over the forecasting horizon t to t+i are considered unknown and set equal to zero.

The tested alternative models

Three alternative types of models have also been considered as references for the evaluation of the performances of the tested lumped conceptual rainfall-runoff models.

The first one is the simplest possible forecasting model, sometimes referred to as the "persistence" model. This model can be used in the absence of forecasts about the future evolution of the discharges. It consists in repeating the last observed discharge value:

$$Q^s_{t+i} = Q_t \quad (2\text{-}2)$$

where Q^s_{t+i} stands for the simulated (forecasted) discharge for time $t+i$ and Q_t is the observed discharge at time i, i being the forecasting horizon.

The second type of considered forecasting models is the linear model class. Easy to use and calibrate, linear models are sometimes still in use in some operational hydrological forecasting services. Rainfall-runoff dynamics are well known to be highly non-linear. Linear models are therefore of relatively poor standard and adapted model should significantly out-perform them. Supposing the future rainfall rates are perfectly known, the linear models have the following form:

$$Q^s_{t+i} = \alpha_0 + \sum_{k=1}^{n} \alpha_k Q_{t-k} + \sum_{j=1}^{m} \beta_j P_{t+i-j} \quad (2\text{-}3)$$

where α_k and β_j are the parameters of the model. n and m the number of previous discharge and rainfall intensity values considered where adjusted. They correspond to the values leading to the best criterion obtained on validation sets.

Finally, artificial neural networks (ANN) have also been considered as possible discharge forecasting models. A large scientific literature exists on the use of ANN for flood forecasting issues. Most of the published results show a superiority of ANN over other types of models including linear models and sometimes RR conceptual models. A thorough evaluation of the performance of flood forecasting models should therefore almost necessarily include ANN. A large variety of programs or libraries exist that facilitate the implementation of ANN. The Matlab library has been used in the present case study. Feed-forward neural networks with one hidden layer and a sigmoid activation function (see appendix for the definition of these terms) where tested. These are, by far, the most frequently used ANN especially in hydrology. The Levenberg-Marquardt algorithm has been chosen for the calibration of the ANN with a stopping criterion based on the mean square error gradient.

Various strategies have been proposed to calibrate ANN especially to avoid "over-calibration": 1) "cross-validations" are sometimes conducted, the calibration being stopped when the performances of the ANN begins to decrease on an alternative validation data set (Anctil et al., 2004 ; Maier & Dandy, 2000 ; Imrie et al., 2000 ; Coulibaly et al., 1999), or 2) "pruning algorithms" (Hassoun, 1995) consisting in eliminating after calibration the weights of the ANN close to zero to reduce the number of its parameters. The evaluation of these more or less complex implementation strategies is still a matter of research and scientific controversy (Hassoun, 1995). The results obtained may largely depend on the case study and the skill of the model user. A complete analysis of these strategies was beyond the range of the present study. We chose here therefore a relatively simple implementation and evaluation procedure inspired by the work of Gosset & Gaume (2003) and based on a calibration-validation approach. The optimization of the ANN implementation could have improved at least slightly the obtained results. Nevertheless, the selected approach gives probably a good image of what performances a non-specialist of ANN is likely to obtain.

For each tested ANN structure defined by a set of input variables and the number of hidden neurons, ten calibration trials were realized with initial values randomly chosen and the validation criterions computed for the ten calibrated ANN. The median criterion, as well as the maximum and minimum obtained values are shown for each watershed and for each tested ANN structure.

The inputs of the neural networks have been adjusted. They contain the three previous known discharges (the three previous discharges for a one step ahead prediction) and the L previous rainfall intensities, L being equal to half of the time of concentration of the watershed in hours plus two. It corresponds more or less to the inputs of the optimal linear models.

The evaluation criterion

Three criterions will be used to evaluate the performances of the tested models.

- The well known Nash-Sutcliffe efficiency criterion (NSE) (Nash & Sutcliffe, 1970), often used to evaluate RR models that approximately represents the percentage of the variance of a given series explained by the model.

$$NSE = 1 - \frac{\sum_{k=1}^{n} (Q_k^s - Q_k)^2}{\sum_{k=1}^{n} (Q_k - \bar{Q})^2} \qquad (2\text{-}4)$$

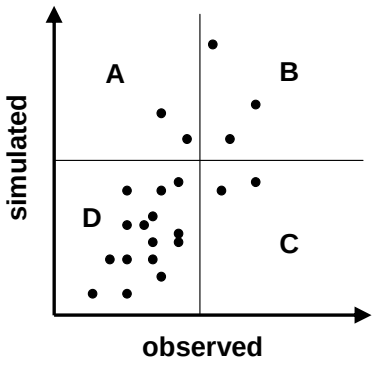
  where $\bar{Q}$ the average observed discharge over the n time steps of the series, k is the time step index, Q_k is the observed discharge at time step k and Q_k^s is the simulated discharge.

- The Nash-Sutcliffe criterion computed on the discharge variations over the forecasting horizon i. The mean square error is minimized (i.e. the NSE is minimized) for the model

calibration. This second criterion will reveal if the calibrated models are able to forecast not only the discharges but also the discharge evolutions, the tendencies. Is the model only an interpolator or has it cached some of the dynamics of the RR process?

$$CRIT = 1 - \frac{\sum_{k=i}^{n}\left[(Q_k^s - Q_{k-i}^s) - (Q_k - Q_{k-i})\right]^2}{\sum_{k=i}^{n}\left[(Q_k - Q_{k-i}) - \overline{(Q_k - Q_{k-i})}\right]^2} \quad (2\text{-}5)$$

- The last criterions have been specifically developed to evaluate forecasting tools, especially meteorological models. Given a discharge threshold value over which an alarm is triggered, they evaluate the proportion of times where the forecasting model leads to the delivery of false alarms and to miss alarms. A probability of detection (POD), a False Alarm Ratio (FAR) as well as a CSI (critical success index) mixing both will be computed for various forecasting horizons (cf. Figure 2-2).



A = false alarms
B = correct alarms
C = missed alarms

$$POD = \frac{B}{B+C}$$

$$FAR = \frac{A}{A+B}$$

$$CSI = \frac{B}{\frac{1}{1-FAR} + \frac{1}{POD} - 1}$$

- POD: probability of detection
- FAR: false alarm ratio
- CSI: critical success index

Figure 2-2: Presentation of the POD, FAR and CSI statistics

The calibration-validation approach

The available discharge and rainfall intensity series have been divided into three subsets of equivalent length. The first sub-set is then used to calibrate the models and the two other serve as validation sets. The length of these series depends on the watershed (see Table 2-1). It ranges from 2 years for Coubon and Pandreaux which is extremely short to 8 years for Chambon, Chadrac and Bas-en-Basset which is a relatively long period if compared to previous RR applications. It is foreseen that the length of the calibration period will affect the extrapolation performance of the models, especially of the conceptual models and neural networks.

Another important fact which may affect the validation results is the progressive increase of density of the rain gauge network over the considered period. The nature of model inputs and especially the level of uncertainty attached to these inputs may differ from one subset to another (see Section 2.4), the model being calibrated on the first sub-period, a priori the period with the lowest quality inputs. As the input and their uncertainty change, the optimal parameter sets of the models may also change over time due to compensation effects for non-linear models (Huard & Mailhot, 2006).

The split-sample tests (Klemes, 1986) reverse the calibration and validation series. They have shown that the influence of the density of the rain gauge network on the calibrated models and their performance is relatively limited. The calibrated parameter values of the conceptual models are relatively stable from one period to another and the validation performance of all the models on a given sub-series does not vary significantly with the sub-set used for their calibration.

We chose here to show the results obtained in the case study closest to the operational situation: calibration of the model on the oldest data set.

2.3.2 Intercomparison of the conceptual RR models

Figure 2-3 shows a synthesis of the validation results obtained with the conceptual models on the 11 watersheds. The validation results without update (mode 1: pure simulation) are fluctuating from one watershed to another as indicated by the distance between maximum and minimum values. Four major conclusions can be drawn from these first results:

- Apparently, the second validation series, the most recent one, is better reproduced by the models than the first one, despite that the rainfall inputs of the first validation series are closer to the calibration rainfall inputs. This indicates that the Nash criterion is much more influenced by the specificities of the hydrological series (dry or wet periods, presence of extreme floods…), than by the accuracy of the spatial rainfall estimates.

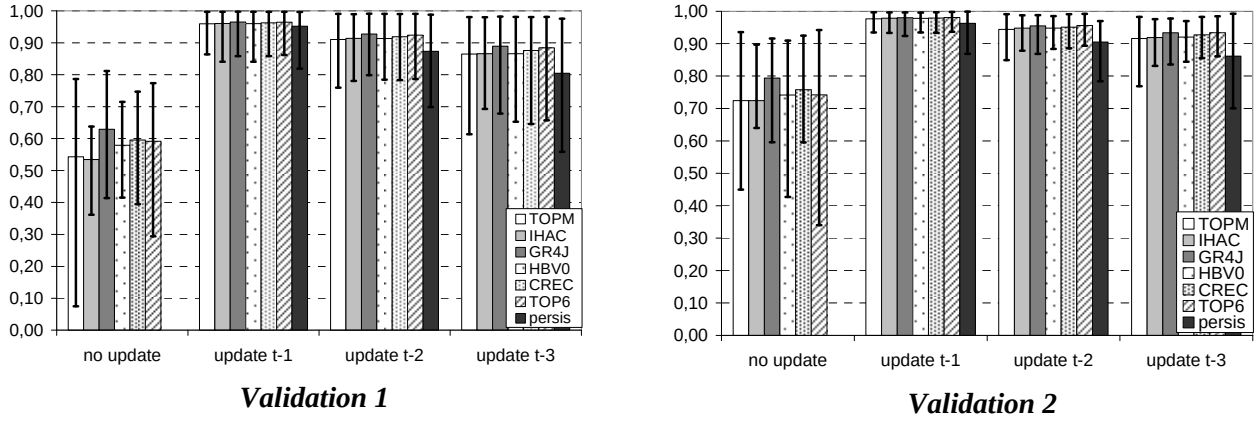


Figure 2-3: Nash criterions obtained on the two validation periods with the six tested conceptual RR models: median, maximum and minimum Nash values for the 11 watersheds. RR model without update or with an update based on the past measured discharges.

- The Nash values obtained on hourly series, between 0.1 and 0.95 depending on the period, the watershed and the model, with a median value of 0.65, may not appear particularly high. They nevertheless lie in the ranges of values generally obtained in validation on daily data with conceptual RR models according to the scientific and technical literature (Perrin et al., 2001). The present case study can therefore be considered as typical of an operational application of RR models: the Nash criterions are neither particularly low nor exceptionally high.

- Updating is absolutely necessary to beat the persistence model.

- The results obtained with the various tested conceptual RR models are very similar. The choice of the RR model does not appear to be a significant issue. The forecasting performance is much more influenced by the accuracy of the available data (see Section 2.4) and by the way the models are implemented (updating) than by the models themselves.

GR4J, the most parsimonious model, which has the slightly better results, will be used in the rest of the report.

Whereas conceptual RR models are often used as forecasting tools in an operational context, a large amount of scientific literature has been devoted to other possible flood forecasting models, especially neural networks. Let us now evaluate the relative performances of these models.

2.3.3 Comparison with the alternative models

The calibration results are logic: Figure 2-4 and Figure 2-5 illustrate the results obtained for the 1-hour and 3-hours ahead prediction models. The Nash values grow with the number of parameters, i.e. the number of degrees of freedom, of the tested models. Even the simplest ANN, with one hidden neurone, outperforms the linear and, of course, the persistence model indicating a larger flexibility of the ANN when compared to the linear model. The spread of Nash values of the calibrated models, noticeable on the Cros and Rieutord watersheds indicate calibration difficulties for ANN with more than one hidden neurone. Note that all the models exhibit poorer Nash values on these two watersheds having the shortest time of concentration.

In contrast, the patterns of the Nash criterions obtained on the validation data sets appear highly dependent on the watershed and the series. Generally, but not always, the best ANN median Nash as well as the linear model Nash exceed the persistence model Nash values for the validation sets. But the differences are relatively modest: i.e. the persistence model is hard to beat. Moreover, the ANN performances are fluctuating from one watershed to another and from one validation set to another. Except for some specific cases (Chambon-sur-Lignon and Espaly watersheds), the best median validation performances are obtained with the simplest ANN model (1 hidden neurone). Moreover, the validation Nash values of the ten calibrated ANN with more than one hidden neurone are highly variable (see the difference between the minimum and maximum Nash value on the figure) and the median Nash value of the best ANN significantly exceeds the Nash value of the linear model for only a limited number of case studies.

These disappointing validation results obtained with ANN may be partly due to the simple implementation procedure used that does not, for instance, limit the risk of over-fitting. Figure 2-4 and Figure 2-5.d show the Nash values on the validation period 2 of the ANN having the best Nash value on the validation period 1. This simplified cross-validation procedure eliminates the ANN having the worst validation results. In a larger amount of cases, the best validation results are obtained, this time, with ANN with more than one hidden neurone: i.e. cross-validation procedures are necessary for the implementation of complex ANN. But the general validation pattern remains unchanged. For the 1 hour ahead predictions, ANNs significantly outperform the linear model in only two case studies (Rieutord and Espaly) and for half of the cases the validation Nash of the best ANN is lower than the Nash of the linear model. The situation is only marginally better for the 3 hours ahead predictions (Figure 2-5).

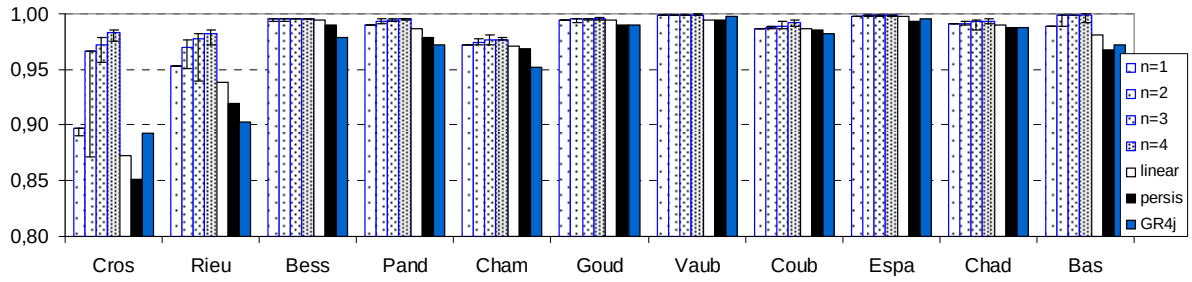
It is likely that the ANN performances could be improved slightly with a more sophisticated implementation procedure, but the general conclusions would probably remain unchanged. ANNs appear to be efficient simulation tools if compared to the linear model in some case studies, the Rieutord watershed example here. But the efficiency of the ANN and their implementation procedure appear to be case dependent and highly fluctuating.

The Nash criterions obtained with the conceptual RR model GR4J are close to ones of the linear model and the ANN, but GR4J appears more robust:

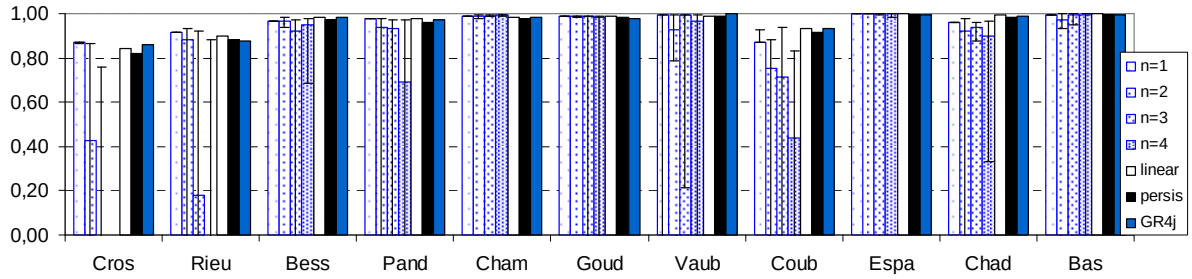
- The decrease of the Nash criterion between calibration and validation is less pronounced than with the two other types of models.

- The distance between the conceptual model and the other tested models appears to grow in favor of the conceptual model as the lead time augments (see Figure 2-5). This reveals a specificity of the conceptual model: it makes better use of the rainfall information which becomes increasingly important as the forecasting lead-time grows.

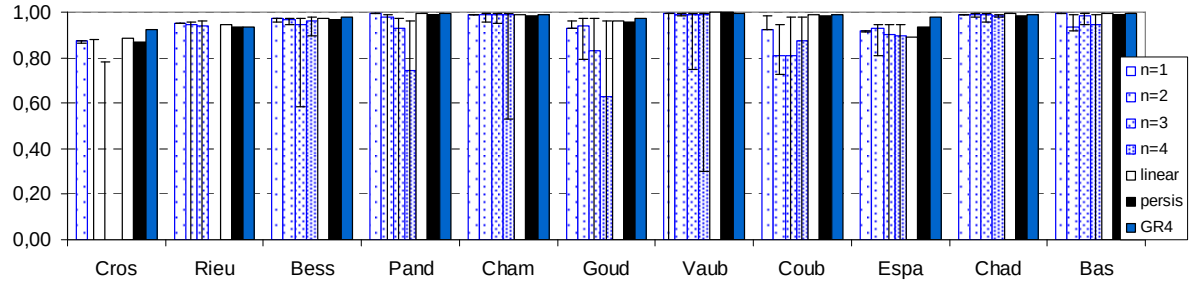
We will have some further illustration of this last idea in the next sections.



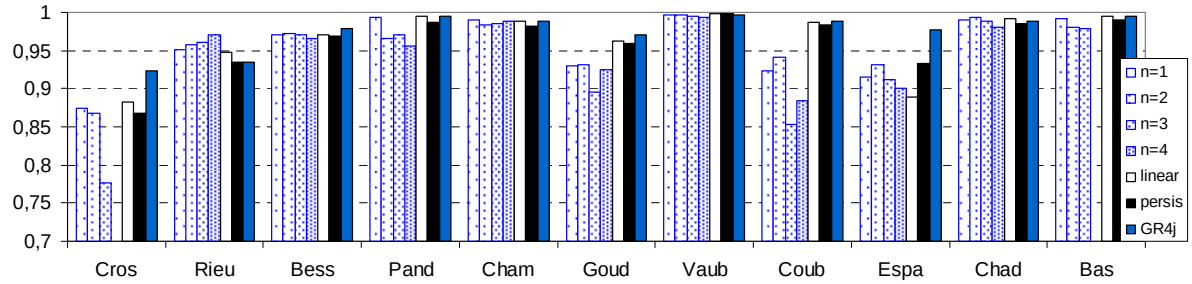
(a) Calibration



(b) Validation 1

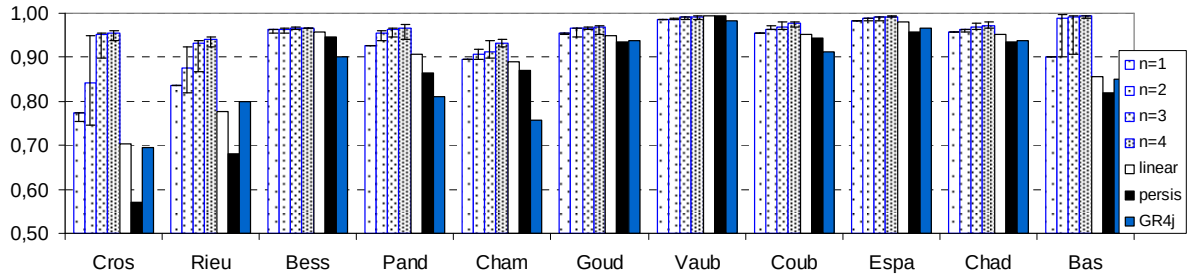


(c) Validation 2

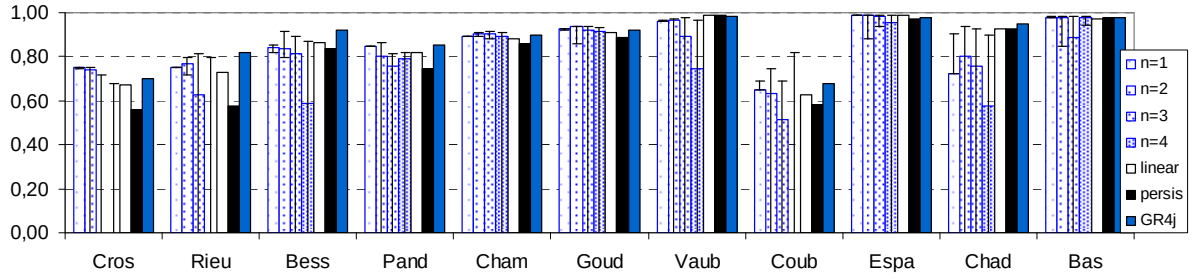


(d) Validation 2 second approach

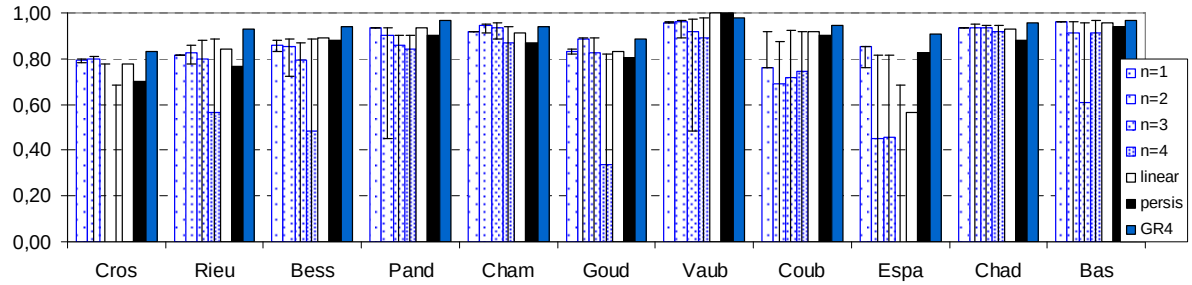
Figure 2-4: Calibration and validation Nash criterions for neural networks on the various watersheds. 1-hour prediction. Median, maximum and minimum Nash values for 10 calibration trials. N is the number of hidden neurones.



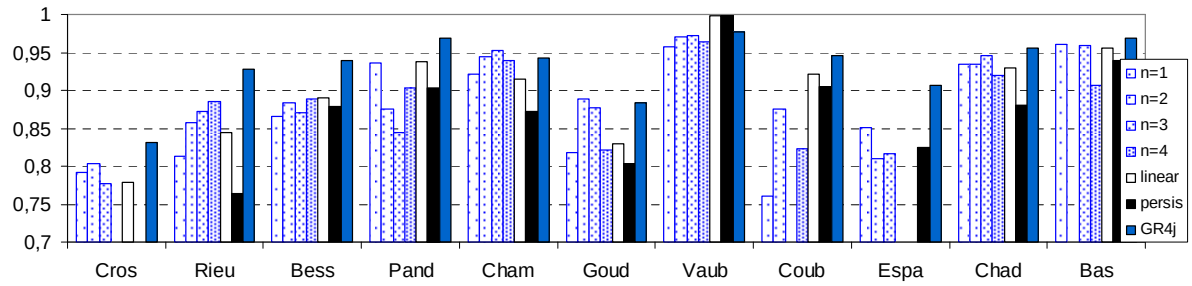
(a) Calibration



(b) Validation 1



(c) Validation 2



(d) Validation 2 second approach

Figure 2-5: Calibration and validation Nash criterions for neural networks on the various watersheds. 3-hour prediction. Median, maximum and minimum Nash values for 10 calibration trials. N is the number of hidden neurones.

2.3.4 Forecasting performances of the tested models

The computed Nash criterions computed with the future rainfall supposed to be perfectly known do not really inform about the usefulness of the models for an operational forecasts. This section gives three additional point of view which help to better assess the potential and the limits of the tested models.

- It will be firstly tested if the models are able to predict the future evolution of the discharge over the next few hours. Usually, forecasters are aware of the actual state of the rivers and they need information about the trends. For short lead-time, prediction of trends is as important for the forecasters than the prediction of the future discharge.

- Aggregate criterions like the Nash value do not completely inform about the efficiency of the models. The forecasts of the various models will be compared in detail for some floods looking at the measured and forecasted flood hydrographs.

- All the evaluations were conducted with the hypothesis that the future rainfall is perfectly known. This helps to isolate the prediction errors due to the model and the errors due to uncertainties of future rainfall. How are the forecasts affected if the future rainfall is unknown? This will be tested for various lead-times and models.

Ability of the models to forecast the discharge evolutions

Figure 2-6 compares the validation Nash values computed on the discharge variations for the three tested types of models and three lead-time.

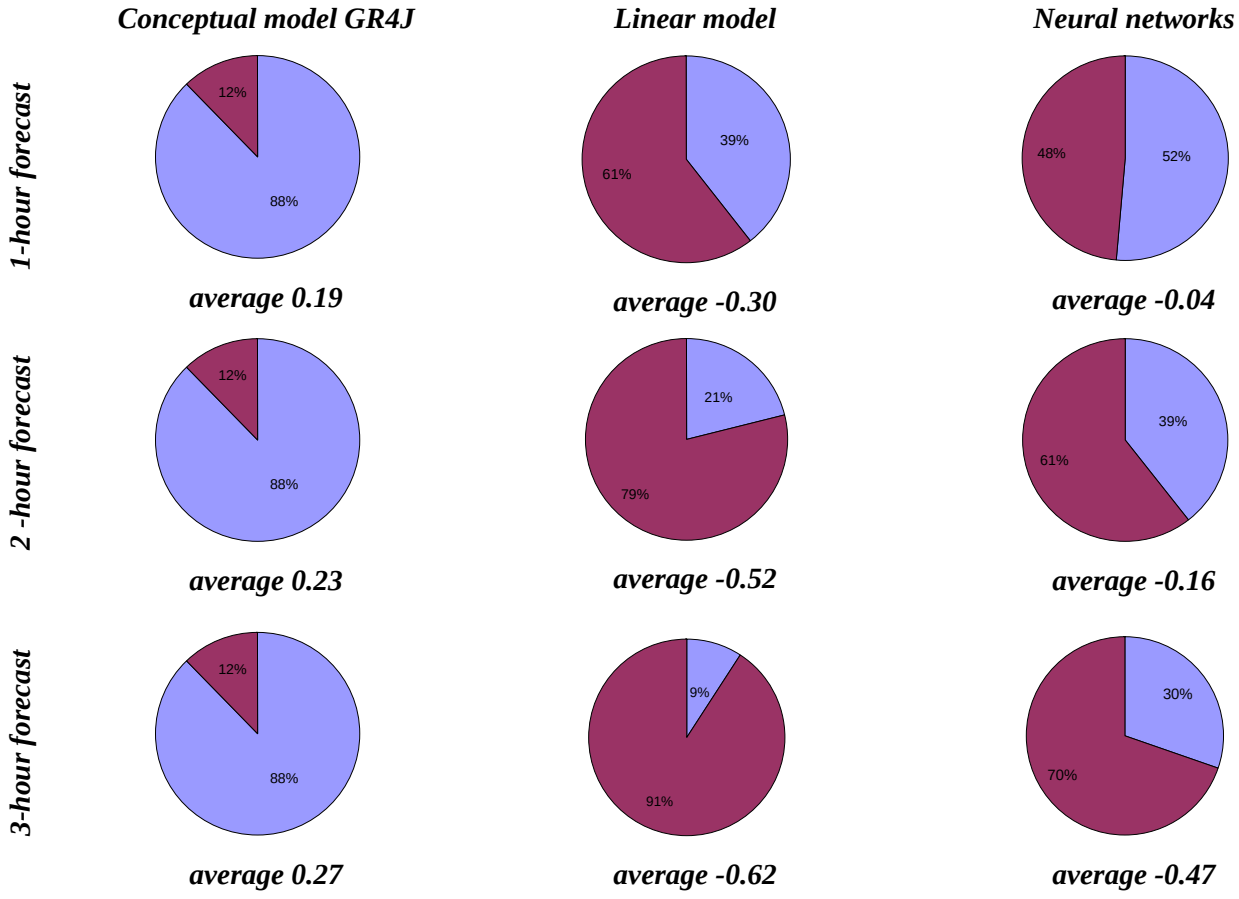


Figure 2-6: Comparison between the discharge variation forecasting efficiency (Nash criterions on discharge variations) obtained with the 3 tested types of models for 3 forecasting horizons. Blue (positive Nash), red (negative Nash)

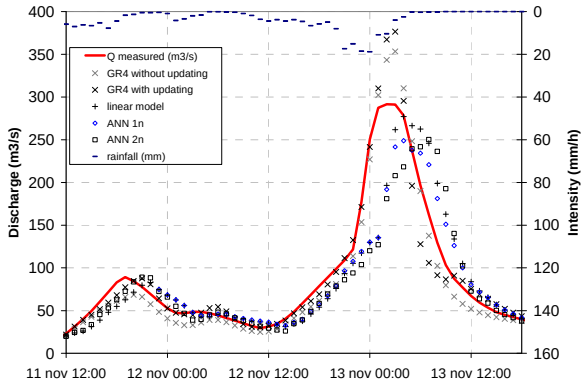
This figure reveals a clear difference between the models and especially between the GR4J model and the two others. In the majority of the cases (watersheds and validation series), the Nash criterion of the GR4J model is positive: the conceptual model explains one part of the observed

variations of the discharge trends. It is not the case for the linear model and to a lower extent for the Neural networks. In both cases, the average Nash is negative. Both types of models behave as pure interpolation models. They perform well on the criterion on which they were calibrated, the mean square error or NSE, but did not catch the other characteristics of the observed series. In other words, they do not fit to the dynamics of the simulated process.

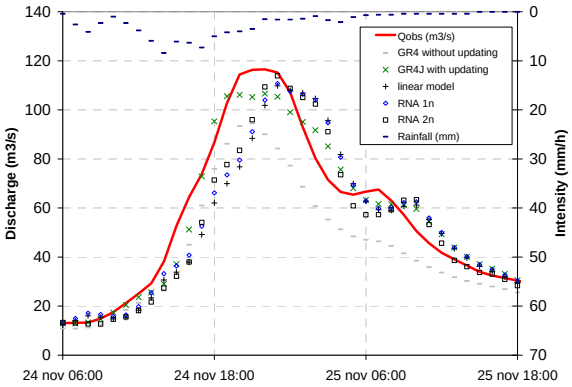
The evolution of the criterion values with the forecasting lead time differs also. While the criterion has a tendency to decrease as the lead time increases for the linear model and the neural networks, it grows for the conceptual model. This reveals again the variable capacities of the model to valuate the rainfall information which becomes increasingly important as the lead time grows. A large part of the forecasting efficiency of the linear models relies on the autocorrelation of the discharges. This is clearly revealed by the values of their calibrated parameters: the weights affected to past rainfall have much lower values than the weights affected to past discharges. It is much more complex to diagnose for Neural networks due to the complexity of their structure, but the comparison between the behaviours of linear models indicates that they encounter the same difficulty to valuate rainfall data.

Finally, even if the Nash criterion of the conceptual model is positive on average, it is quite low. Only 20 to 30% of the discharge trend variations are explained by the models. This explains why despite the fact that conceptual RR models have been developed for 40 years, they still are seldom used by operational flood forecasting services.

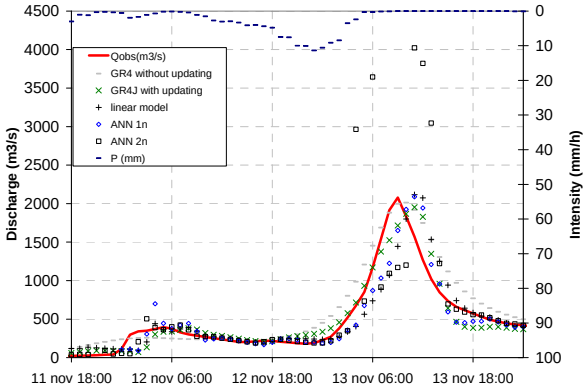
Detailed analysis of the forecast on some events



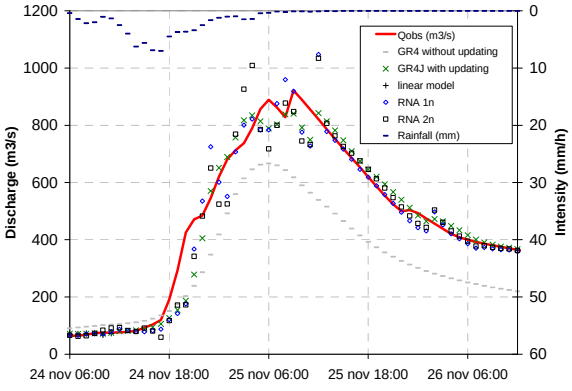
a) Chambon, 13 Nov. 1996, 3 hours ahead



b) Chambon, 25 Nov. 2002, 3 hours ahead



c) Bas-en-Basset, 13 Nov. 1996, 3 hours ahead



d) Bas-en-Basset, 25 Nov. 2002, 3 hours ahead

Figure 2-7: Some examples of forecasted hydrographs

A detailed analysis of the forecasted flood hydrographs gives another insight into the differences between the models. Some flood hydrographs are presented in Figure 2-7 for two watersheds with

different times of concentrations and for three hours ahead forecasts. The specificities of the three types of tested forecasting models are especially noticeable on the rising limb of the flood hydrographs. The conceptual RR model is the only one that really anticipates the changes in the discharge evolution trends due to the rainfall events. This is clear on the 13 November 1996 flood hydrograph at Chambon. It is less evident for the same flood at Bas-en-Basset. Due to the high time of concentration of the upstream watershed at Bas-en-Basset, the flood hydrograph is less peaky and the interpolation models (Artificial neural networks and linear models) have the opportunity to adjust themselves.

Clearly, the ANNs and the linear model fail where forecasting tools are particularly needed, i.e. in detecting in advance the changes in the discharge evolution trends. These changes are induced by rainfall. This default of the models indicates without doubt that they do not efficiently evaluate the input information on rainfall intensities. Timing error is a well known problem of ANNs (Abrahart et al., 2007). It may be partly solved by modifying the criterion used for their calibration including for instance a penalty for timing errors, but at the price of a reduction of their Nash criterion.

The models specifically developed for RR modelling are therefore superior to ANN and linear models on the condition that their timing is correct. This is the case for the simulated floods at Chambon but not at Bas-en-Basset. Looking in detail at the results for the various watersheds, it appears that the timing is generally pretty well reproduced by the model for the smaller watersheds and that difficulties arise for the largest watersheds (Chadrac and Chambon). The fact that the timing error is varying from one event to the other (compare Figure 2-7.c and Figure 2-7.d) indicates that it is not a problem linked to the parameterization of the RR model but to the specificities of the rainfall event not taken into account in the model and especially the spatial distribution of rainfall: rainfall concentrated on the downstream part of the watershed will lead to a more rapid reaction of the watershed and vice-versa. Moreover, due to the non-linearity of the RR processes and rainfall amount concentrated on a sub-area of the watershed will not have the same impact than the same amount homogeneously distributed over the whole watershed. For the largest watersheds, the use of distributed hydrological models, able to take into account the spatial distribution of rainfall may lead to an improvement of the forecasting performances, while it does not appear to be the case for watersheds up to 500 km^2 on the upper Loire case study.

The results obtained with the ANN with 2 hidden neurones for the November 1996 flood at Bas-en-Basset suggest a last comment (Figure 2-7.c). The model produces unrealistic predictions for this event with extremely high discharge values and high frequency oscillations at the beginning of the event. This flood is the most important of the available series. It corresponds for the models to an extrapolation case and ANN, especially ANN with high complexity (high number of hidden neurones), are well known to be inappropriate for extrapolating. The main requirement of flood forecasting services is precisely to be able to deliver accurate forecasts for large floods. They need therefore to extrapolate. There is hence a significant question regarding whether ANN are really appropriate for flood forecasting.

POD, FAR and CSI statistics

The previous section has been mainly focused on short term forecasts over a few hours. But the RR models can be used in combination with meteorological forecasts for longer term forecasting – typically 6 to 24 hours. For such lead times, the objective is generally to detect in advance that a given discharge threshold will be exceeded and specific evaluation criterions have been proposed (POD, FAR and CSI). This will be the main focus of the second part of this report devoted to FFG. To have some elements of comparison, POD, FAR and CSI statistics of the tested models were also computed. Of course here, the input of the RR models are not forecasted rainfall amounts but actual measured rainfall rates.

Four different thresholds have been considered for the computation of the criterions:

- Threshold 1: Vigilance stage - mobilization of flood forecasting services.
- Threshold 2: Pre-warning stage – mobilization of local authorities and emergency services.
- Threshold 3: Warning stage – a warning message is issued.
- Threshold 4: Decennial flood.

The existing threshold values of stages and corresponding discharges were used for the six gauging stations included in the flood warning system: Goudet, Espaly, Chadrac, Chambon, Vaubarlet and Bas-en-Basset. For the five other stations (Rieutord, Cros, Pandereau, Bessère, Coubon), the same ratios between the mean annual discharge and the thresholds observed in Goudet have been selected: i.e. 10 for threshold 1, 25, 33 and 100. These ratios have a general tendency to decrease with the size of the watershed and the reduction of the variability of the discharges. They are for instance equal to 7, 8, 12 and 50 at Bas-en-Basset.

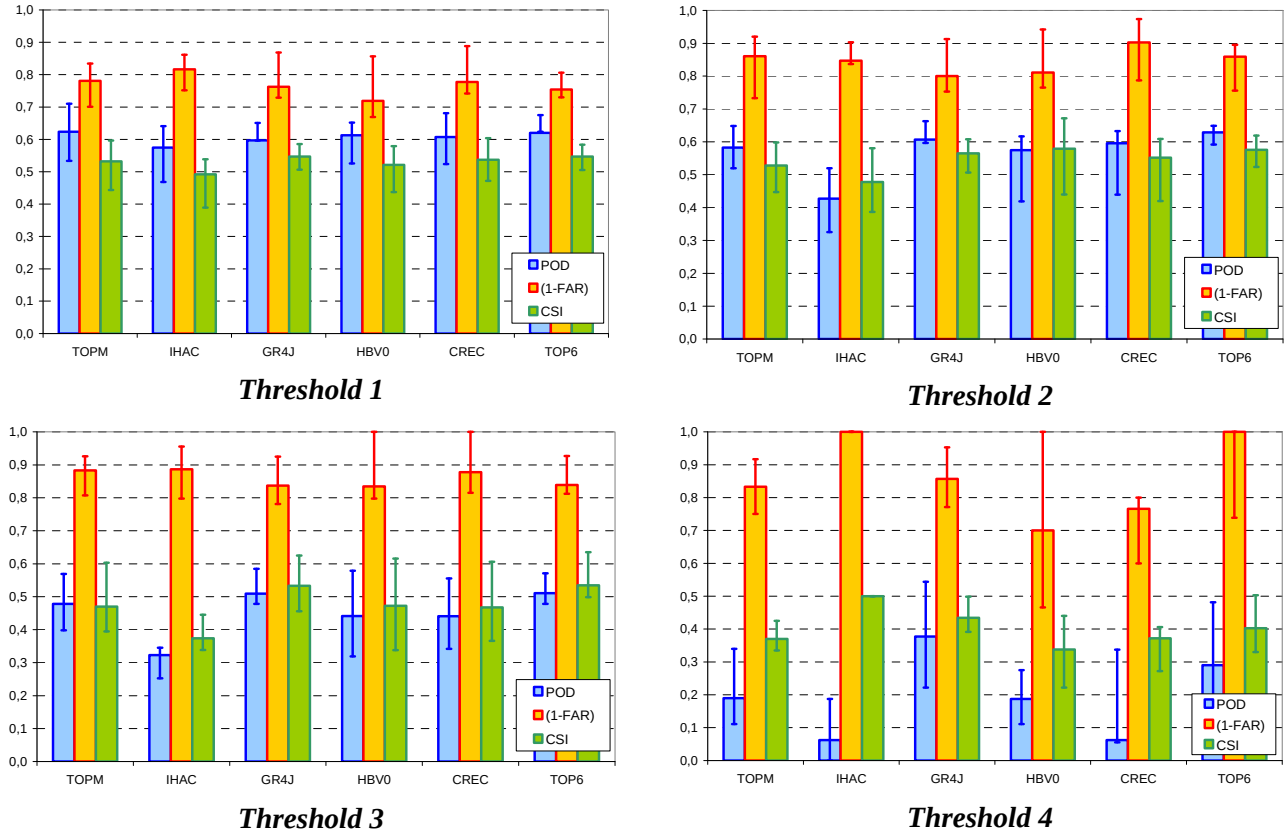


Figure 2-8: Average, maximum and minimum forecasting performances of the various tested RR models for the 11 upper Loire river watersheds. Rainfall is supposed to be perfectly known in advance.

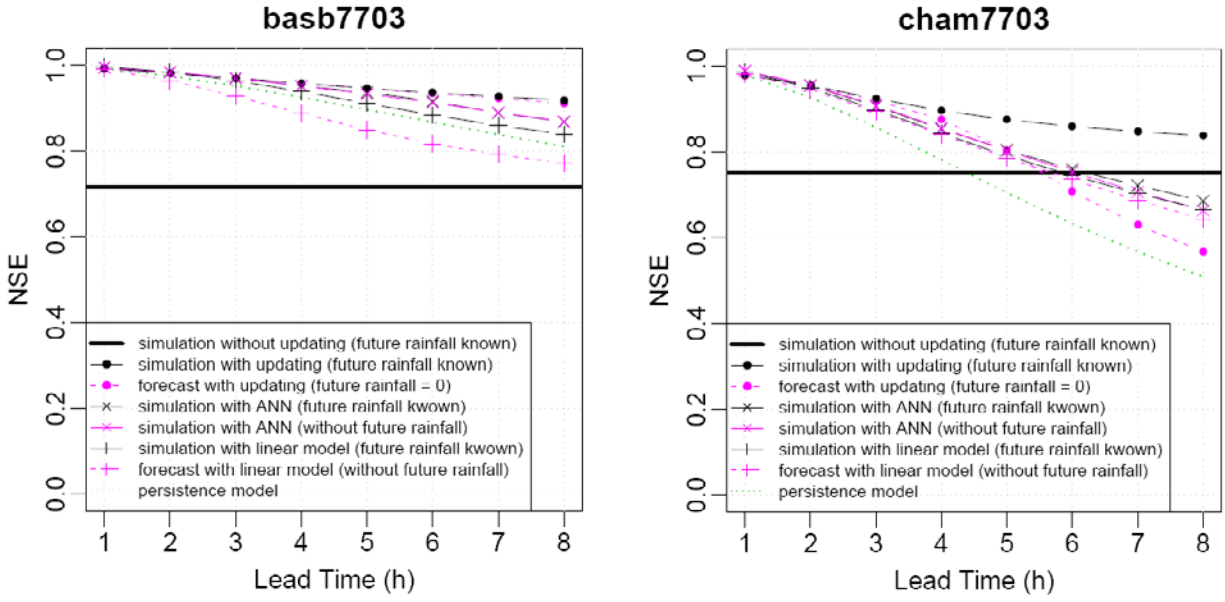
The results obtained are, as it was the case for the other evaluation criterions, not much dependent on the model. They also appear to be similar to the results obtained with other types of models on the same data sets (see next sections for complementary results and discussion on the comparison between RR models and Flash Flood Guidances).

Impact of uncertainties concerning future rainfall

All the previous runs have been realized with the hypothesis that the rainfall rates were perfectly known over the forecasting lead-time. This corresponds to an ideal case: perfect rainfall forecast. How are the forecasting results affected if the future rains are unknown? The answer depends on the lead-time and the response time of the watershed.
To test it, a second series of runs were conducted:

- New linear models and ANNs were calibrated and validated with inputs including only the discharges and the rainfall rates measured before the current time step,
- The conceptual RR models were not calibrated again but the rainfall rate values set to zero after the current time step to produce the forecasts. This corresponds to the most extreme future rainfall scenario: the rainfall ceases after the current time step.

The average results for the two validation periods and for various forecasting lead-time are summarized in Figure 2-9.



a) Bas-en-Basset **b) Chambon**

Figure 2-9: Evolution of the forecasting NSE criterions with the lead-time and impact of the uncertainties concerning the future rainfalls. Conceptual RR model (dots and thick black line).

Table 2-2: Comparison of the Characteristic time of forecast and the estimated times of concentration of the watersheds.

Gauging station	Characteristic time (hours)	Approx. time of conc. (hours)
Cros (**Cros**)	2	2-3
Besseyre (**Bess**)	2	2-4
Rieutord (**Rieu**)	2	3-5
Pandreaux (**Pand**)	3	3-5
Chambon (**Cham**)	3	5-7
Vaubarlet (**Vaub**)	6	8-10
Espaly (**Espa**)	>8	10-13
Goudet (**Goud**)	4	5-8
Coubon (**Coub**)	6	8-12
Chadrac (**Chad**)	5	10-13

Bas-en-Basset (**Bas**)	>8	20-24

They confirm the tendency observed on Figure 2-5: the distance between the conceptual RR model and the other models grows when the lead-time increases if the future rainfalls are known.

The efficiencies of all the forecasting models are reduced if the future rainfalls are not taken into account. In the case of the GR4J model, this reduction is very limited for a certain range of lead-times and increases rapidly when a certain "characteristic time", linked to the time of concentration of the watershed (see Table 2-2), is exceeded. This characteristic time appears to be equal to 1/3 to 1/2 of the estimated time of concentration (delay between the moment the rainfall ceases and the inflection of the decreasing limb of the flood hydrograph).

Under this characteristic time, the uncertainties on future rainfall have little influence on the conceptual model RR simulations. The conclusions drawn for Bas-en-Basset and Chambon for the 1-hour and 3-hours lead-times would for instance not be modified if the uncertainty on future rainfall would be taken into account. Over this characteristic time, the hydrological forecast becomes sensitive to the rainfall forecast, which rapidly becomes the critical factor determining the accuracy of the overall forecast.

2.3.5 Conclusions

The major conclusions that can be drawn from this detailed analysis of the forecasting performances of the rainfall-runoff models on the upper Loire area are the following:

- ANNs are difficult tools to use. The risk of over-fitting or over-parameterisation limits the complexity (number of hidden neurones) of the models that can be calibrated on the existing data set and limits thereby the usefulness of ANN. Their efficiency appears to be highly variable from one case study to the other and it can hardly been anticipated in which case they will be efficient. Furthermore, ANNs like the linear model appear to be essentially interpolation models that do not take a real advantage of the rainfall information. For all these reasons, conceptual RR models, specifically developed to simulate the rainfall-runoff process, appear to be better suited flood forecasting tools.

- Over a certain watershed area, distributed hydrologic models may be more efficient than the tested lumped models, reproducing more accurately the timing of the floods which is an important factor affecting the quality of the predictions especially when an updating procedure is used.

- Overall, the performance of the forecasting models is relatively modest. Only 20 to 30% of the discharge trend variance is explained by the models even after a large calibration and validation effort. This explains why, despite the fact that conceptual RR models have been developed for 40 years, they are seldom used for flash flood forecastings. This relatively disappointing result may be due to the limitations of the models but also to the various sources of uncertainties and particularly the uncertainties on the actual rainfall amounts. The impact of the rainfall estimation uncertainties will be studied in detail in the next part of this report.

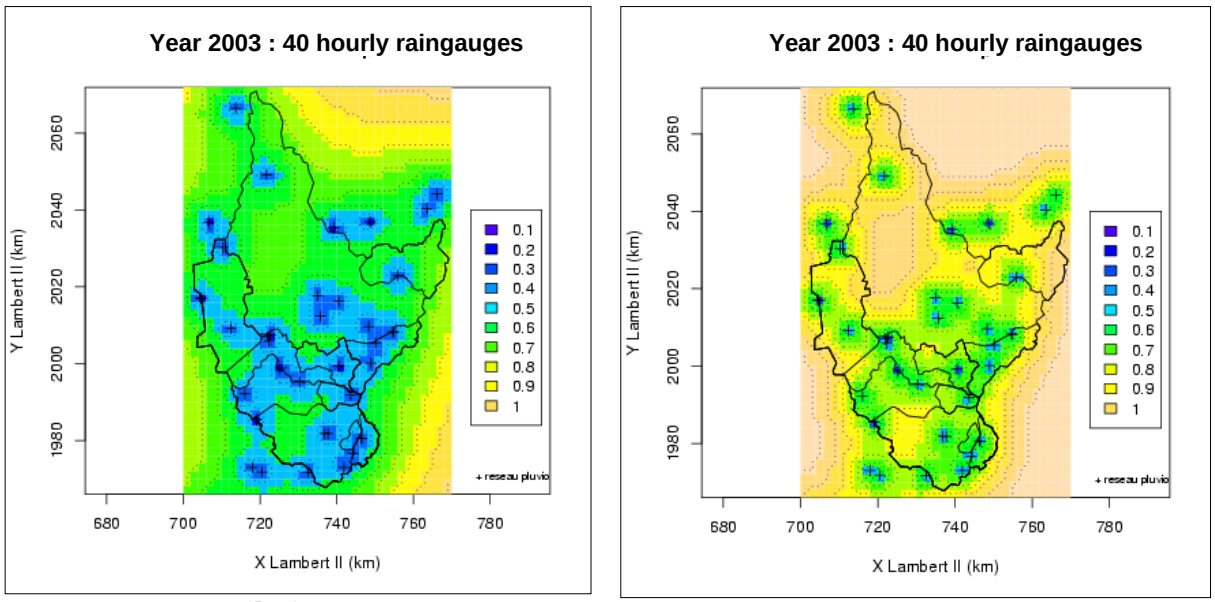
2.4 Performances of RR models in simulation and forecasting modes

2.4.1 Introduction

The previous section has exposed the limits of the rainfall-runoff models when used in an operational forecasting context. The inaccuracy of the forecasts based on RR simulations may be attributed to the simplicity of the models used that are probably not able to represent the rainfall-runoff dynamics in all its complexity. But we just have also seen that the information content of the

datasets available for the calibration of the RR models drastically limits their possible number of parameters and hence their complexity. We must cope wit this unavoidable state of facts.

Another major factor limiting the RR simulation accuracy is the uncertainties in their input (rainfall) and output (discharge) variables. If the errors in the measured discharges are well known and generally limited – plus or minus 10 to 20% - mean areal rainfall rates computed on the basis of point rainfall measurements from a rain gauge network or even radar data are generally much more inaccurate as illustrated by Figure 2-10 and Figure 2-11. Despite the densification of the rain gauge network in the upper Loire region during the last 20 years, high uncertainties remain on the rainfall rates that can be estimated through spatial interpolation in many parts of the region, especially when short time steps are considered (hourly rainfall rates). The estimated kriging interpolation relative standard error for hourly rain rates is lower than 50% for less than 50% of the territory. In other words, poor or even no information on the hourly rain rates is available on the majority of the area.



Daily time step ***Hourly time step***

Figure 2-10: Estimated kriging interpolation relative standard error for hourly and daily rainfall amounts over the upper Loire river area.

This certainly has a tremendous impact on the RR simulation efficiency. It appeared interesting, before closing the discussion on the RR model usefulness for flash flood forecasting, to study in more details the areal rainfall rate estimation uncertainties and their impact on the RR simulations. This additional work has two main objectives:

1. It can give a measure of the possible gains that could be obtained through an improvement of the rainfall measuring techniques especially the radar system. The results presented hereafter plead in favour of the pursuit of the efforts to improve the quantitative radar data valuation, but also to evaluate the radar rainfall estimation uncertainties.

2. Knowing the rainfall estimation uncertainty level and being able to evaluate it, could help the operational flood forecasting services to turn from the standard deterministic approach, disappointing since it frequently fails to deliver correct forecasts (see previous section), to a stochastic approach taking into account all the possible discharge evolutions given the uncertainties about actual rainfall amounts.

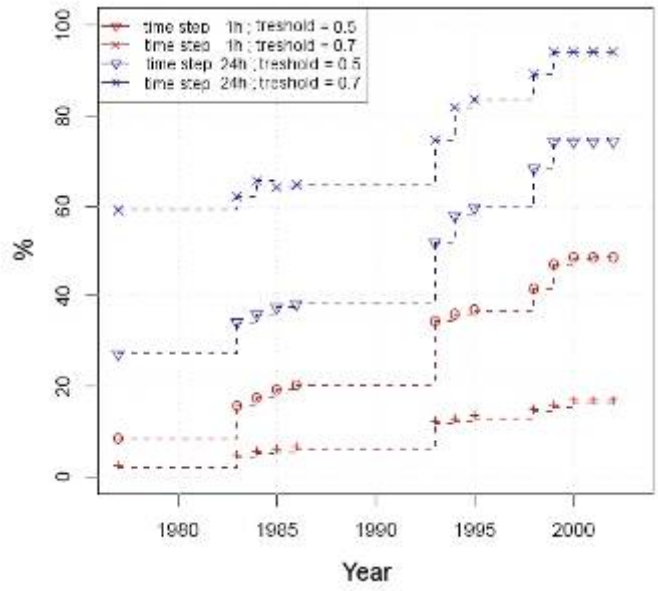


Figure 2-11: Evolution of the percentage of the upper Loire river area where the theoretical kriging relative standard error is lower than 0.5 or 0.7.

We will hereafter proceed in three steps to evaluate the mean areal rainfall estimation errors and their impact on the simulated discharges.

1. Step 1 : calibrate and validate an hourly rainfall error model.

2. Step 2 : calibrate and validate a temporal dependence model for these errors.

3. Step 3 : use Monte Carlo simulations of rainfall scenarios based on the calibrated error model and propagate these scenarios into the selected RR model (the GR4j model).

The hereafter shortly presented developments around the definition and validation of a rainfall estimation error model (steps 1 and 2) may appear a little bit sophisticated. This sophistication is nevertheless not a scientific gadget: the realisticness of the error model is a necessary condition to draw any valuable conclusion from the propagation of these errors into rainfall-runoff RR models.

2.4.2 Hourly point and spatial rainfall estimation error model

Geostatistical methods and especially kriging are certainly the most popular approaches in hydrology for the spatial interpolation of rainfall (Lebel and Labordes, 1988). They have been widely used in the past and tested and appear to deliver reasonably good rainfall estimates. An interpolation model based on kriging delivers not only interpolated values but also an error model for these interpolated values. This is the part which is particularly important in the present case study. The objective of this section is to test if the theoretical error model is in accordance with the observed errors. This will be tested for point rainfall estimates for which reference values are available and enable an error computation. The validity of the error model given that the interpolation model as well as the point rainfall error model have been validated will be supposed.

A standard cross testing method will be used. It consist in withdrawing in turn one rain gauge from the network and to estimate the rain rates at this gauge for the various time steps of the available rainfall events on the basis of the other measured rainfall intensities on the network.

The main adjustment factor of the kriging and corresponding interpolation model is the variogram. The following hypothesis were done: a) the variogram is independent on the time step and the

rainfall event (climatological kriging), b) isotropy of the variogram, c) the possible influence of altitude, exposition … is neglected for the interpolation (i.e. the interpolation is only based on inter-distances), d) a spherical variogram is used with an estimated decorrelation distance of 20 kilometers for hourly rain rates.

With such hypotheses, the point rainfall interpolation standard deviation of the interpolation errors σ depends only on the structure of the surrounding rain gauge network and on the standard deviation of the rainfall field *SD*. The relative theoretical standard deviation $\sigma_r = \sigma / SD$ is constant for a given network structure.

Resulting observed and theoretical[2] interpolation relative error distributions (relative: divided by the standard deviation of the rainfall field SD) are compared in Figure 2-12 for various rain gauge of the upper Loire area network.

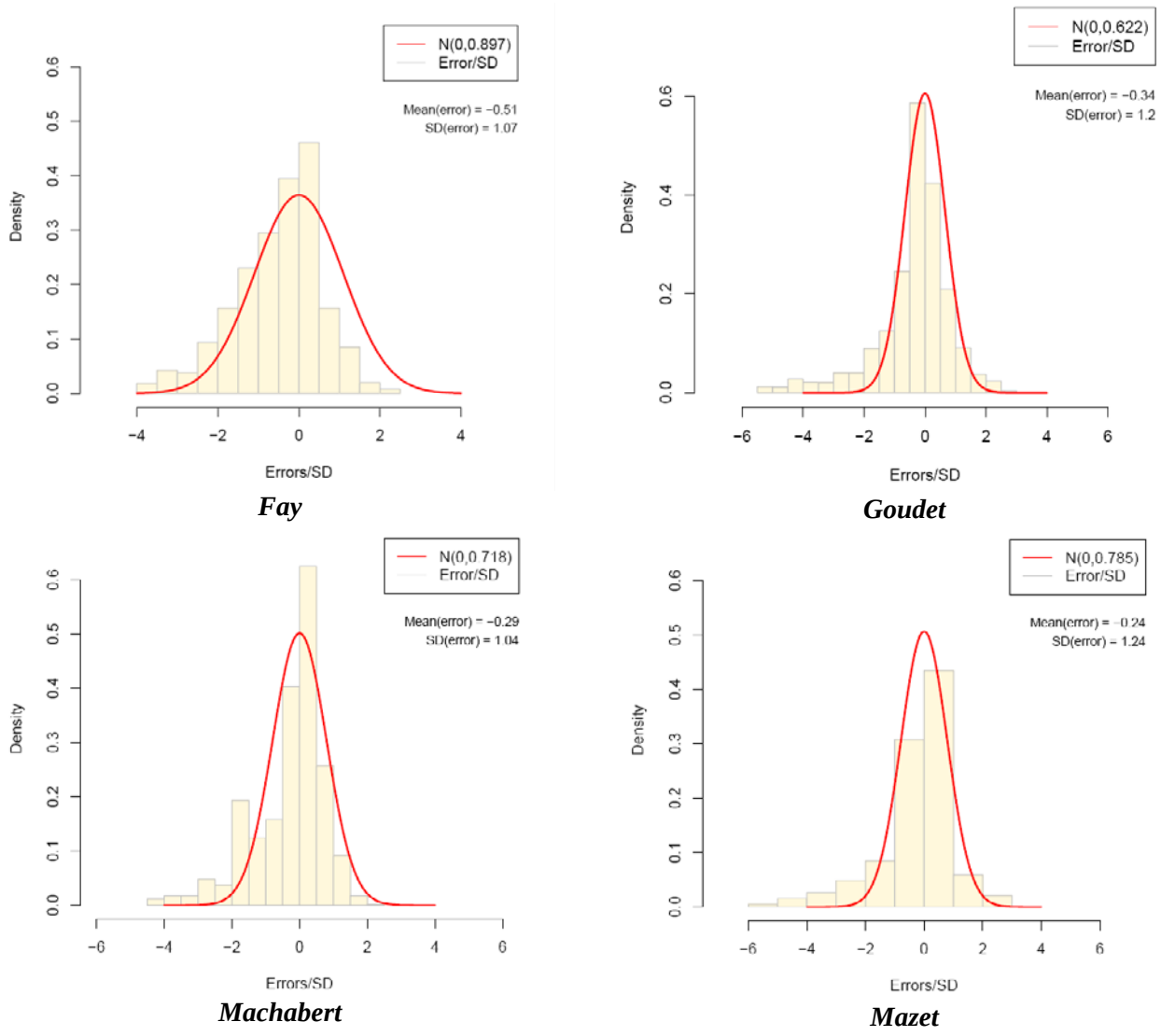


Figure 2-12: Example of a comparison between the theoretical and the observed distributions of the standardized rainfall estimation errors for four rain gauge of the network.

Considering the simplicity of the interpolation and error models – linear interpolation with one parameter which is the decorrelation distance, Gaussian error distributions – the validation results,

[2] Gaussian distribution with 0 mean and a standard deviation σ_r

still far from perfect, are relatively satisfactory. The bias, foreseen for interpolated values with a lower bound equal to zero, is generally limited except for the Fay rain gauge located on the relief and more frequently exposed to severe Mediterranean thunderstorms. The standard deviation of the observed interpolation errors is always larger than the theoretical one due to higher density of largely underestimated values in the observed distribution if compared to the theoretical Gaussian one.

Table 2-3: Percentage of computed interpolation errors comprised in various theoretical confidence intervals for four test rain gauges.

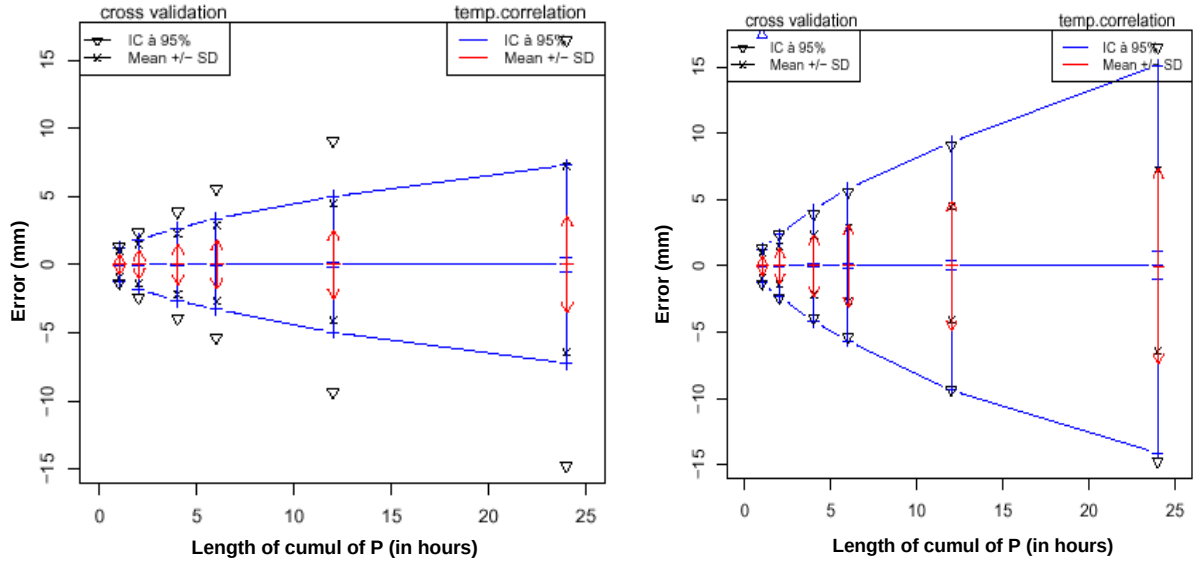
Rain Gauge	68 % CI (±σ)	95% CI (±2σ)	99.7 % CI (±3σ)
Fay	53.5	77.3	89.8
Goudet	59.4	80.7	90.3
Machabert	63.1	84.8	93.4
Mazet	59.8	81.8	91.0
Theoretical value	**68**	**95.0**	**99.7**

Overall, the theoretical interpolation error model will lead to a slight under-estimation of the error ranges or quantiles (Table 2-3).

2.4.3 Correlation of estimation errors over time

Successive interpolation errors may be dependent in time. Errors are due to the spatial heterogeneity of the rainfall that is not captured by the rain gauge network, i.e. at lower space scales, typically to the presence of rain cells in some areas. The cells may be developing and progressing slowly over the area and the same type of error repeated more than one hourly time step.

This temporal dependence has to be taken into account to produce realistic rainfall interpolation error scenarios.



correlation coefficient 0 ***correlation coefficient 0.6***

Figure 2-13: Example of comparison between observed (crosses and triangles) and simulated (lines) interpolation error quantiles for rainfall accumulations over 1 to 24 hours. Same rain gauge and two tested correlation coefficients.

A simple time dependence model has been tested for the point and average spatial interpolation errors: a first order auto-regressive model. This model applies to the interpolation relative error whose distribution is theoretically independent on the time step according to the selected interpolation method.

$$\eta_{i+1} = \rho\eta_i + \sqrt{(1-\rho^2)}\sigma_r\varepsilon_{i+1} \quad (2\text{-}6)$$

where η_i is the relative interpolation error at time step *i*, ρ is the correlation coefficient between two successive errors, σ_r is the theoretical standard deviation of the interpolation relative error distribution, and ε_{i+1} a random variable with a standard normal distribution.

It is straightforward that if the time dependence model holds for point rainfall estimation through spatial interpolation and if the correlation coefficient does not depend on the location in space (i.e. on the structure of the surrounding rain gauge network), than the same time dependence model with the same correlation coefficient value holds for errors done on estimated average spatial rainfall amounts.

Of course the correlation coefficient probably depends on the density and proximity of the surrounding rain gauges. We will nevertheless test, using the available point rainfall measurements,

1. If the simple proposed time dependence model is pertinent, i.e. if it generates point rainfall estimation error series coherent with the observed ones and especially if it is able to explain the evolution of the interpolation error distributions for rainfall amounts cumulated over 1 to 24 hours.

2. If the suited correlation coefficient does not depend too much on the location.

If, according to the results of the cross validation tests, these two conditions appear to be reasonable, the combination of the kriging and autoregressive models will probably deliver good approximations of the errors made in the approximation of average spatial rainfall rate over watersheds, which is what we are looking for.

Table 2-4: Standard deviations of the estimation relative error of mean intensities over various durations: 1) observed and 2) modelled. Constant correlation coefficient equal to 0.6.

Time step (hours)	Fay		Goudet		Machabert		Mazet	
	observed	modelled	observed	modelled	observed	modelled	observed	modelled
1	1.05	0.74	0.76	0.66	0.76	0.77	0.97	0.69
2	0.95	0.65	0.66	0.58	0.69	0.68	0.72	0.61
4	0.79	0.54	0.55	0.49	0.55	0.56	0.57	0.51
6	0.70	0.47	0.49	0.42	0.48	0.49	0.47	0.44
12	0.58	0.36	0.39	0.33	0.40	0.38	0.36	0.34
24	0.49	0.26	0.29	0.24	0.32	0.27	0.28	0.24

Figure 2-13 and Table 2-4 illustrate some of the cross-correlation results. The impact of the correlation coefficient on the error series structure and especially on the error distributions of rainfall amounts cumulated over more than one hour is clearly noticeable. The comparison with the distributions of observed errors reveals clearly the necessity to take into account the dependence in time of interpolation errors and the relative adequacy of the proposed time dependence model.

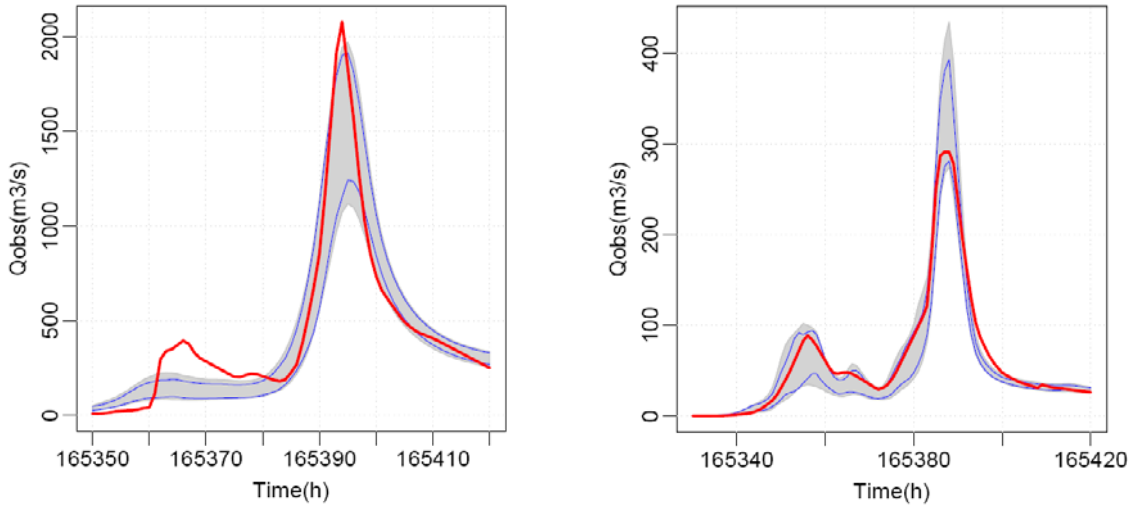
The best suited correlation coefficient value seems not to depend too much on the location (see Table 2-4) which is the second very satisfactory result of the cross-correlation. The proposed model

reproduces quite well the evolution interpolation error distributions for rainfall amounts cumulated over a large range of time steps and for various locations of the rain gauge network. Of course, it is not perfect. The selected correlation coefficient leads generally to slightly underestimate the variance of the estimation errors (see Table 2-4).

The interpolation error model being chosen and at least verified if not perfectly validated – its validity for average spatial rainfall estimation could not be completely tested – let us now use Monte Carlo runs to simulate various scenarios of possible average spatial hourly rainfall amount series corresponding to the available point rainfall measurements and propagate these scenarios into a calibrated RR model to evaluate the impact of the rainfall estimation uncertainties on rainfall-runoff simulation results and hence on RR forecasting efficiency.

2.4.4 Propagation of the estimation errors into RR models

Figure 2-14 shows two examples of the propagation results of the rainfall estimation uncertainties for the major flood observed during the recent period in the upper Loire river area. The RR model used here is the GR4J model calibrated on the first third of the series available on each watershed. The impact of the mean areal rainfall uncertainties on the RR simulations appears impressive, and this for all the watersheds, and particularly all the watershed areas, and all the floods. Generally, the upper bound of the 90% simulated uncertainty range is about 1.5 higher than the lower bound. Even if the RR model were perfect, which of course they are far from, mean areal rainfall estimation uncertainties set a relatively low limit to the accuracy of RR simulations or forecasts.



Bas-en-Basset (10/11/1996) ***Chambon-sur-Lignon (10/11/1996)***

Figure 2-14: Two examples of observed flood hydrographs (red lines), 90% confidence interval for the simulated discharges (blue lines) and limits of the simulated discharge values obtained after 20 Monte Carlo runs.

Table 2-5 gives another insight into the impact of rainfall estimation uncertainties. The computed 90% uncertainty bounds appear to contain a large proportion of the measured discharges. This is particularly through for the flood periods (observed discharge greater than 10 times the average annual discharge) or if a tolerance of 20% is considered. This tolerance represents both: the discharge measurement uncertainties and the reasonable efficiency goal of the forecasters.

Table 2-5: Proportion of the observed values comprised in the 90% simulated confidence interval, with and without a 20% tolerance.

	Rieutord		Chambon		Bas-en-Basset	
	Qobs	**Qobs +/- 20%**	**Qobs**	**Qobs +/- 20%**	**Qobs**	**Qobs +/- 20%**
All data	18.4	99.8	15.6	99.7	8.4	79.3
Qobs>10 Qmean	31.0	87.3	22.2	64.9	16.7	59.0

For the smallest watersheds (Rieutord, but also Cros, Beyssère or Pandereau), the simulated 90% confidence interval contains almost 90% of the measured discharge values when the tolerance factor is considered. In other words, rainfall estimation uncertainties may completely explain the differences between measured and simulated discharges. Clearly, rainfall estimation uncertainties appear in this case as one of the major factors limiting the RR simulation accuracy if not the major one.

This is not completely through for the largest watersheds (typically watershed areas greater than 500 km^2) where other factors seem to play also a role: the proportion of measured discharge values included in the 90% simulated confidence interval remains important when a tolerance is considered but is significantly lower than 90% and seems to decrease as the size of the watershed increases.

A detailed analysis of the simulated and observed hydrographs reveals delays and apparent fluctuations in runoff rates between events that undoubtedly can be attributed to the spatial heterogeneity and repartition of the rainfall. Without surprise, the tested lumped modelling approach is not completely well adapted when the watershed sizes increase.
Distributed hydrological model may then perform better with the limit exposed earlier on the number of parameters of the models. The possible gain of distributed models may be reduced by an increase of the number of their calibration parameters…

2.4.5 Conclusions

The major conclusion of the two previous sections is that RR simulations remain uncertain even in an optimal situation: good quality and long datasets, intensive effort for RR model selection and calibration. These uncertainties have, for the moment, limited their use by operational forecasting services used to deliver relatively accurate forecasts (now-casts) on large rivers based on flood wave propagation models.

A large part of the RR simulation errors can hardly be reduced. It is explained by rainfall estimation uncertainties, except on large watersheds (typically areas over 500 km^2) where the shape of the hydrograph can be influenced by the spatio-temporal pattern of the rainfall event and where distributed RR models may bring a slight improvement if compared to the tested lumped models.
From a practical point of view, rainfall estimation uncertainties limit drastically the possible accuracy of RR simulations. Operational forecasting services should be aware of this limit to efficiently use the RR models and if possible evaluate these uncertainties in real time to be able to deliver confidence intervals along with their traditional deterministic forecasts. Ensemble or Monte Carlo forecasts are now used routinely in meteorological forecasting; there is no reason why they should be disregarded by hydrologists. The error scenario simulation model developed in this section could help to build such ensemble forecasts in the case where mean areal rainfall amounts are estimated through a rain gauge network. The same type of model is still to be developed for the case where quantitative radar estimations are used.

3. Flash-flood prediction based on discharge thresholds and high-resolution operational weather forecasts

3.1 Introduction

The aim of this section is to propose and test a regional approach for early flashflood warning applicable to ungauged river basins. The approach is based on the concept of the statistical distributed (SD) model. **The essence of this approach is to use the ensemble of antecedent model predictions (obtained from previous events) to rank the severity of the predictions of interest.** The chosen (generic) distributed model is run by using archived radar-rainfall grids to derive flood probability characteristics of simulated flows for all cells in the distributed model. When subsequently the distributed model is run in forecast mode, the flooding flow threshold for each grid cell is defined in terms of a flood probability level rather than in an absolute value. In this manner, the flooding flow computed from simulated data is different than that computed from observed data because it takes into account the hydrologic model uncertainty. Because of this, the method:

- can be implemented initially using a priori rainfall-runoff parameters;
- can potentially provide benefits without requiring extensive model calibration;
- is particularly well suited for application in ungauged basins.

The method has been verified on long-time historic series, two case studies of extreme flashfloods, and a 6 month flashflood forecasting period.

3.2 Hydrological model, input data and methodologies

3.2.1 LISFLOOD model and set-up for this study

The hydrological model used for this study is the LISFLOOD model. It is a hybrid between conceptual and physical model combined with a routing module in the channel (de Roo, 1999; van der Knijff & de Roo, 2005). A detailed description of LISFLOOD is given by van der Knijff (2006). If possible, input parameters of LISFLOOD are estimated from existing datasets. Five parameters cannot be a priori estimated and should be calibrated.

For this study the model has been set-up with 1 km grids. Time steps are adapted to the resolution of the input variables. For the long term simulations these are daily and for the detailed case study calculations and all forecasts hourly. Since the aim of the study is to test if the approach can be used in ungauged river basins, the available discharge data have been used for comparison and validation only but not for calibration. Instead the default values for the parameters have been used throughout the study.

3.2.2 Input data for the study

Observed rainfall and temperature data have been extracted from the meteorological archive of the AgriFish unit at the DG Joint Research Centre. This archive holds data from about 2000 stations across Europe. The density of available stations in the study areas is low. For the French case study radar data from the Bollène 2002 experiment are also available (Chapon, 2006; Boudevillain et al., 2006).

High-resolution operational weather forecasting data are provided by the Lokalmodell of the German National weather service (DWD). In 2002 the Lokalmodell of the DWD had a spatial resolution of 7km and an hourly temporal resolution. The forecasting period was 48 hours. The DWD forecasts are provided every 12 hours starting at 00UTC and 12UTC.

3.2.3 Methodology of discharge threshold exceedances

Simulated hydrographs per se do not constitute a flood forecast –in addition a decision making element needs to be incorporated: is the discharge going to exceed a critical threshold or not? The

determination of the critical thresholds is crucial, in particular when dealing with watersheds where only few or no discharge measurements are available. The following model consistent approach is proposed:

1. Simulate a long timeseries of discharge based on observed meteorological data. Obviously, the denser the station network, the better rainfalls and subsequent discharge peaks can be captured.

2. Calculate thresholds from the simulated discharges. Due to the relatively short time series for which reliable meteorological data are available for this study, a quantiles approach is used here. Discharges are ranked from highest to lowest and certain cut-off quantiles are chosen as thresholds. The highest discharge simulated defines the severe threshold level, the 99th percent highest discharge the high threshold level, the 98th percent medium threshold and the 97th percent the low threshold. Alternative methods such as estimation of return periods could also be used if sufficient long time series are available.

3. The performance of the simulations is assessed through exceedances of critical thresholds, e.g. is Qobs > Qcritical obs when Qsim > Qcritical sim.

The major advantage of this approach is that systematic over- or under-prediction of the model is compensated for by looking at relative differences only. If the model systematically overestimate discharges in a given river reach because of a non-optimised parameterization or lack of processes such irrigation or reservoir operations, this is reflected in the thresholds as well as in the forecasts. The disadvantage is that this approach may produce reasonable results in terms of threshold exceedance while the model results are seriously offset from the observed hydrographs.

3.3 Brief description of the two case studies

3.3.1 The 8-9th September 2002 event in the Cevennes-Vivarais region (France)

The Cévennes-Vivarais region is situated Southeast of the Massif Central in France (Figure 3-1:). The main Cévennes rivers expose a typical Mediterranean hydrological regime with very low levels of water in summer and floods occurring mainly during the autumn. The 8-9 September 2002 heavy precipitation event was responsible for one of the most important floods ever recorded in the Cévennes – Vivarais region. It caused 24 casualties and an economic damage estimated at 1.2 billions euros (Huet et al., 2003). A detailed description of the case study can be found in Delrieu et al. (2005), Chancibault et al. (2006) and Nuissier et al. (2007).

In this region, the observation system usually comprises: (i) two networks of about 400 water level stations and about 180 hourly rain gauges; (ii) a network of about 45 water level stations, and (iii) two weather radar of the ARAMIS network from Météo-France. For this study, however, only those meteorological and hydrological stations with long-term records from 1990 to 2002 were chosen, reducing the number of meteorological stations to 14 stations in total of which only 3 were located directly within the study area. The number of discharge stations used for analysis is 14. Those stations with strong reservoir influence were not used.

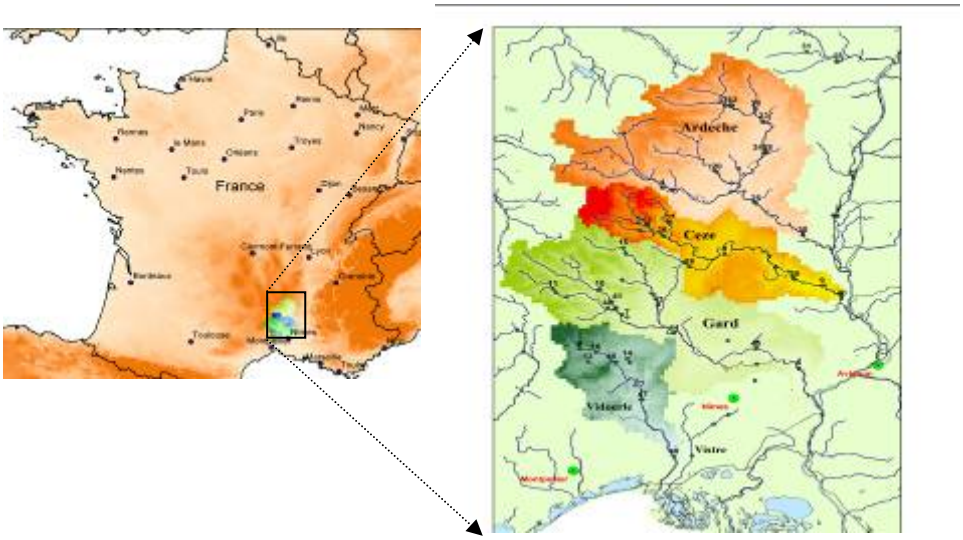


Figure 3-1: Map of the study area and zoom into the Cevennes region with the catchments used for this study.

3.3.2 The 29th August 2003 case in the Eastern Italian Alps

On the 29 August 2003, an extreme storm developed over the Fella catchment, covering an area of 730km^2 (Val Canale), in the Upper Tagliamento river basin in Northeastern Italy. The region received rainfall amounts up to 400 mm in 6 hours, with return times exceeding 500 years. This extreme rainfall event triggered one of the century's most significant floods in the Tagliamento river basin and produced remarkable flash floods, landsliding and debris flows in some of its tributaries. The event resulted in 2 casualties and an economic damage around 1 billion Euros (Borga et al., 2006). The same meteorological input data sources as for the French case study have been used.

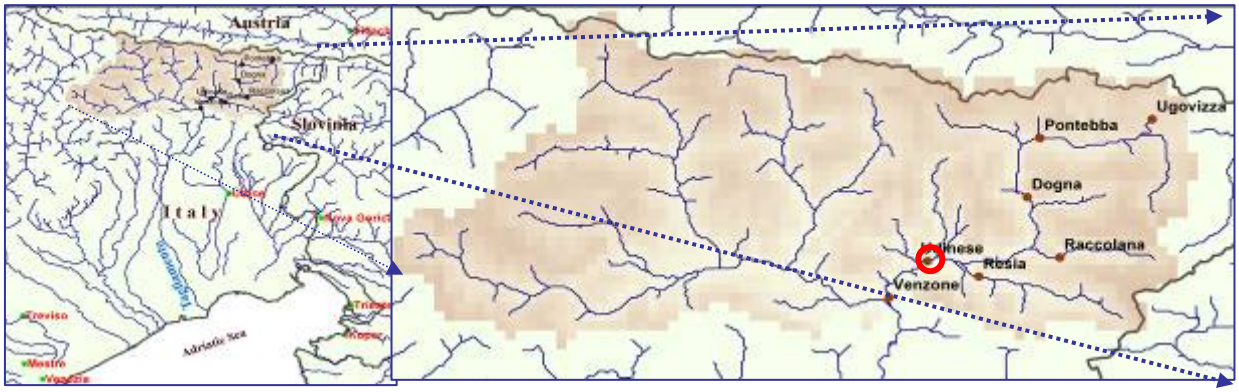


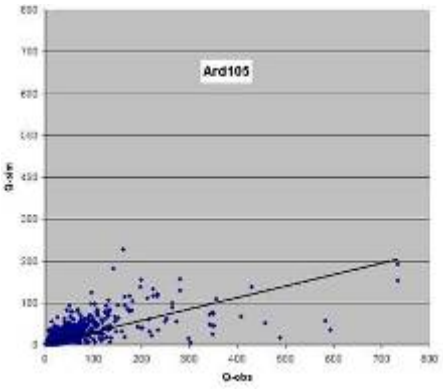
Figure 3-2: Map of thr Eastern Italian Alps with the Tagliamento river basin highlighted. The station Udinese used in the report is highlighted by a red circle

3.4 Results

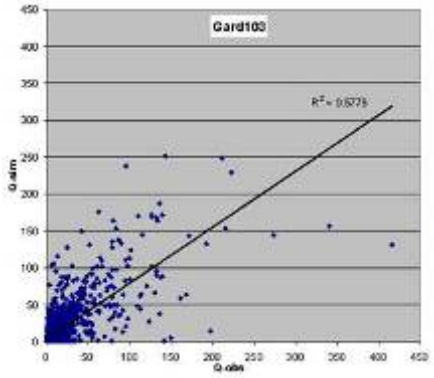
3.4.1 Long-term simulations and calculation of thresholds

Figure 3-3 shows examples of scatterplots for observed discharges (x-axis) and simulated discharges (y-axis) for a period from 1990-2002 for a station in the Ardeche (FR) and the Gard (FR). While for the Gard the model has a slight tendency to underestimate the discharges, the underestimation is important in the Ardeche station. Taken into account that the hydrological model has not been calibrated on observed discharges and considering the low resolution of the

meteorological input data, it is not surprising that the simulated discharges are not reproducing well the observations.



a) Ardeche (ARD105), 636 km2 ups



b) Gard (Gar103), 263 km2 ups

Figure 3-3: Scatterplots for stations in the Ardeche (a) and Gard (b) with observed discharges in m3/s on the x-axis and simulated discharges in m3/s based on synop data on the y-axis.

From the long-term time series the four critical thresholds severe, high, medium and low have been established through ranking procedure for each pixel. The same method is applied to observed discharges to obtain the corresponding critical values Qc obs. An event is defined if any of the threshold levels has been exceeded – in which case it is sufficient to make the analysis only for the lowest category. With this definition of an event contingence tables (Table 3-1) for hits, false alarms, misses and positive rejections have been calculated (Figure 3-4)

Table 3-1: Definition of contingence table

	$Q_{obs} \geq Q_{c\ obs}$	$Q_{obs} < Q_{c\ obs}$
$Q_{sim} \geq Q_{c\ sim}$	H (Hit)	F (False)
$Q_{sim} < Q_{c\ sim}$	M (Miss)	PR (Positive Rejection)

Figure 3-4 shows for the same stations as in Figure 3-3 the number of hits, false alarms and misses. Positive rejections, the vast majority of the cases, are not plotted to avoid distortion of the graphs.



Figure 3-4: Contingence tables of hit, false, and missed threshold exceedances for 1 station in the Ardeche (gray), and Gard (shaded)

For both basins there are generally more hits than false alarms or misses but the hit rate is slightly lower than in the Gard. Obviously flashfloods are events that can be very localized and there is a high probability that the synoptic station network – the input data for these simulations - misses the event - hence a high number of misses. Equally, if one station captures heavy precipitation, the event may have been quite localized but the interpolation smears out the rainfall to larger areas, producing high accumulated rainfalls leading to simulated flashfloods where there were none. The high number of false alarms can also be partially explained with the daily time steps of the model, which can introduce a 1 day time shift of the peak leading to a miss and/or false alarm.

Nevertheless, the results show that even for stations like in the Ardeche, where the model systematically simulates far too low discharges – and with that would never reach flooding level - the number of detected events through threshold exceedance is larger than the number of missed events. Obviously the better the overall simulations, e.g. in the Gard, the better the distribution of hits to false alarms.

3.4.2 Forecasting the 8-9th September 2002 event

Figure 3-5 shows observed and forecasted hydrographs together with the simulated criticial thresholds for a station in the Gard (Gar104) river. The thresholds are colour coded as purple (severe), red (high), yellow (medium) and green (low). This graph nicely illustrates the principle of the threshold exceedance. While the simulated discharges – even with high resolution radar data – are a factor 3 lower than the observed discharges, the high threshold is being exceeded with all forecasts from the 7th September 12:00 forecast onwards. The timings of the forecasted peaks agree well with the observed peak on 9th September at 6:00 o'clock +/- 1-2 hours. This corresponds to an early warning time of 42 hours. Estimating the runtimes of the meteorological and hydrological models as well as data transfer and preparation time of the data to about 6 hours, the leadtime is still of the order of 1 day and more. The severe threshold is exceeded with radar data only.

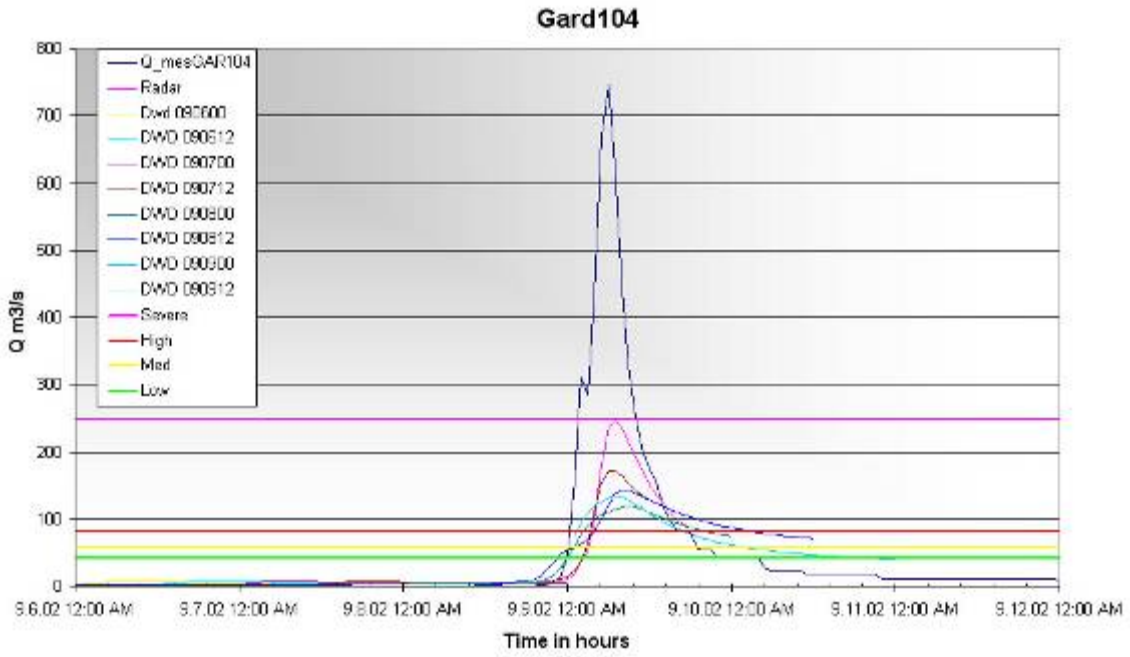


Figure 3-5: Hydrographs of observed and forecasted discharges in m3/s (y-axis) for the station Gard 104 in the Gard river from the 9th-12th Sep 2002 starting at 12:00 in hourly time steps (x-axis). The flood forecasts based on DWD forecasts start on 0906 00:00 until 0909 00:00 in 12h time intervals.

A different visualization (Figure 3-6) of the flood threshold exceedances, described by Ramos et al. (2007), gives a quick and easy overview of the forecasted exceedances.

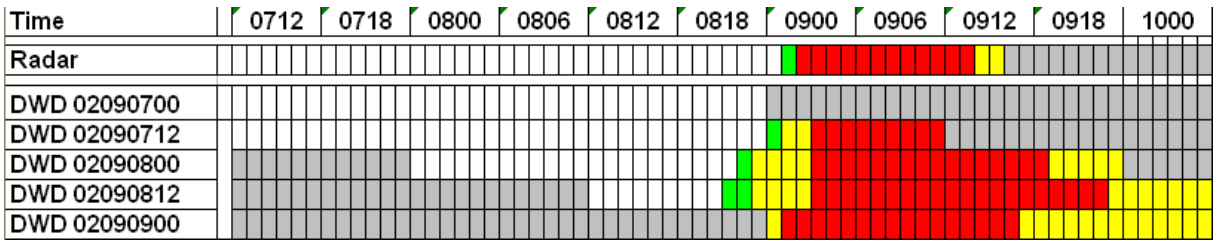


Figure 3-6: Visualisation of threshold exceedances in hourly time steps for the flood forecasts based on radar data and 48 hours DWD Lokalmodell forecasts from 20020907 00:00, 20020907 12:00, 20020908 00:00, 20020908 12:00 and 20020909 00:00. The time is indicated in 6 hour intervals at the top. The exceeded thresholds are colour coded as purple (severe), red (high), yellow (medium) and green (low).

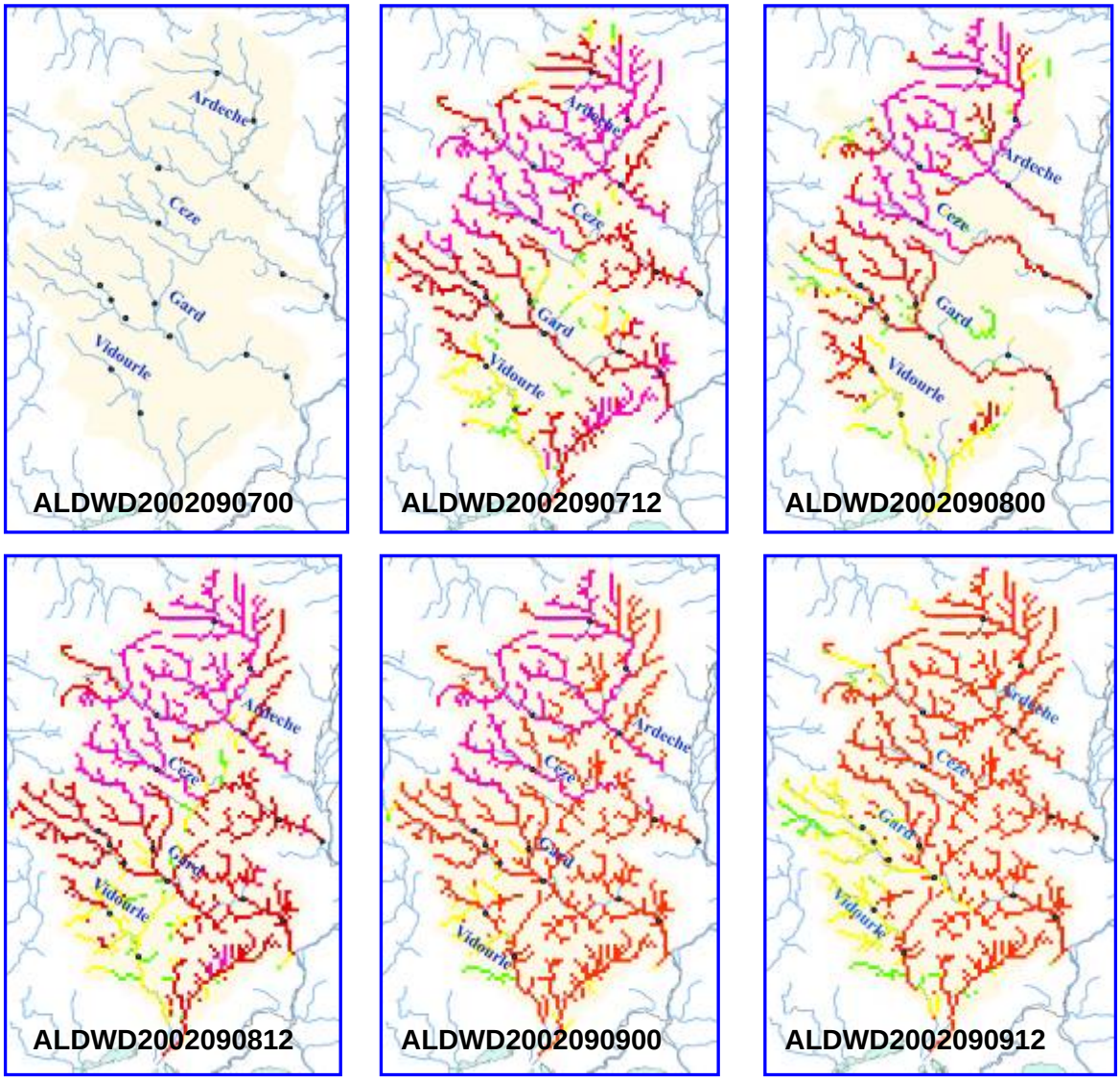


Figure 3-7: Summary threshold exceedance maps showing the highest threshold exceeded during the 48 h forecasting time for flood forecasts based on the DWD Lokallmodell weather forecasts from 20020907 00:00 and in 12 hourly steps until 20020909 12:00. The threshold exceedances are colour coded with purple (severe), red (high), yellow (medium) and green (low).

Figure 3-7 illustrates the spatial development of forecasted event. Each panel shows the maximum alert threshold exceeded during the forecasting period in 12 hourly steps. The panel clearly illustrates that the event is first forecasted on the 07th 12:00 to take place in the upstream areas of all 4 river basins. In the next forecasts the emphasis is mostly in the Gard and Ceze rivers and less in Ardeche and Virdourle. The panel shows how the flooding is forecasted to affect almost the whole basins well exceeded the severe thresholds over large parts of the river basins. Downstream, towards the outlets, mostly only high thresholds are exceeded.

3.4.3 Forecasting the 29th August 2003 event in Italy

Figure 3-8 illustrates the spatial distribution of the forecasted highest exceeded flood thresholds for the Italian case study. The forecasts started on the 27th Aug 2003 and are then shown in 12 hourly time steps until the 29th Aug 2003. Also for this event the time of the peak was well captured and forecasted more than 24 h in advance (Figure 3-9). Almost all simulations exceeded the high threshold alert. Similar to the French case, it seems that the spatial event of the floods may have

been overpredicted. However, without information in the other river basins no clear conclusions can be drawn.

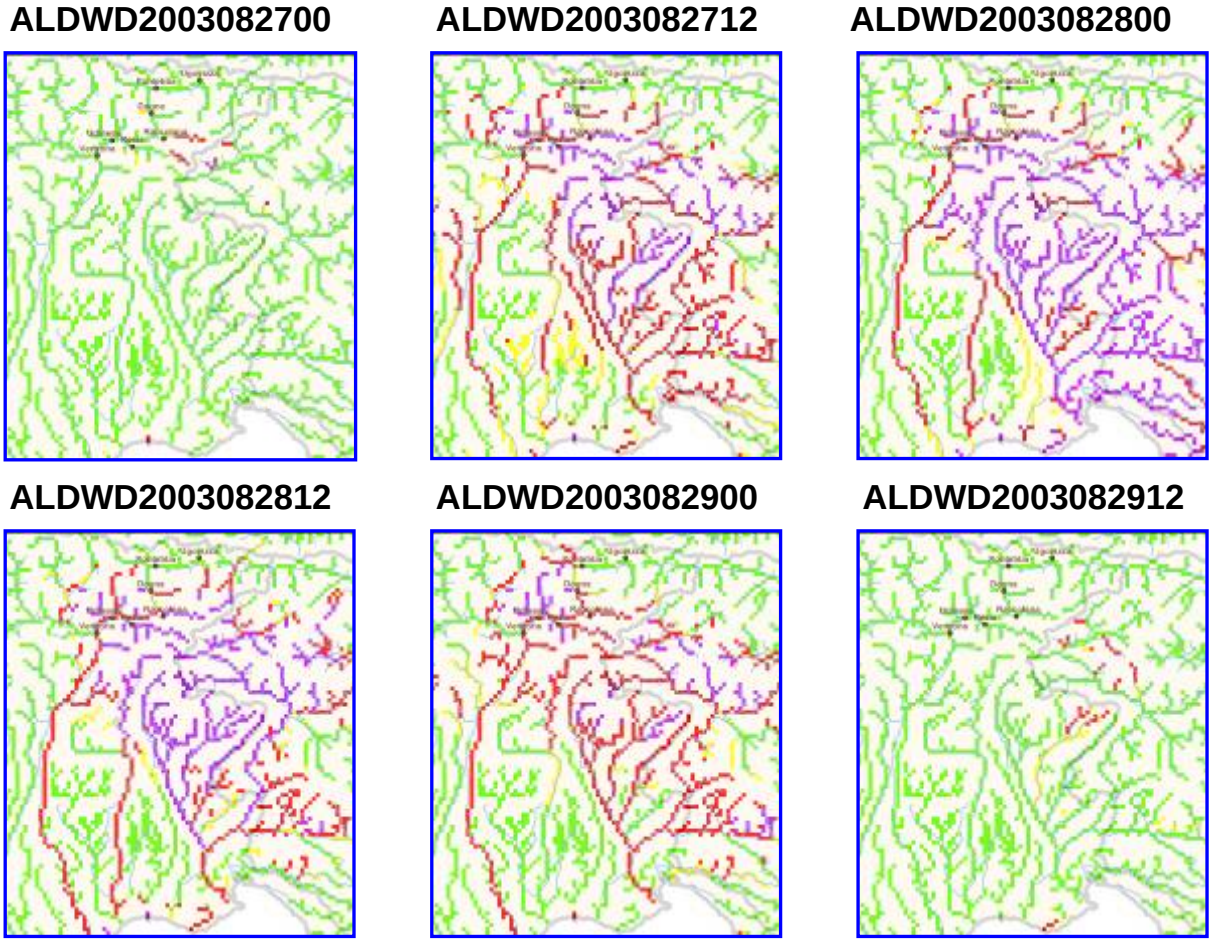


Figure 3-8: Summary threshold exceedance maps showing the highest threshold exceeded during the 48 h forecasting time for flood forecasts based on the DWD Lokallmodell weather forecasts from 20030827 00:00 and in 12 hourly steps until 20030829 00:00. The threshold exceedances are colour coded with purple (severe), red (high), yellow (medium) and green (low).

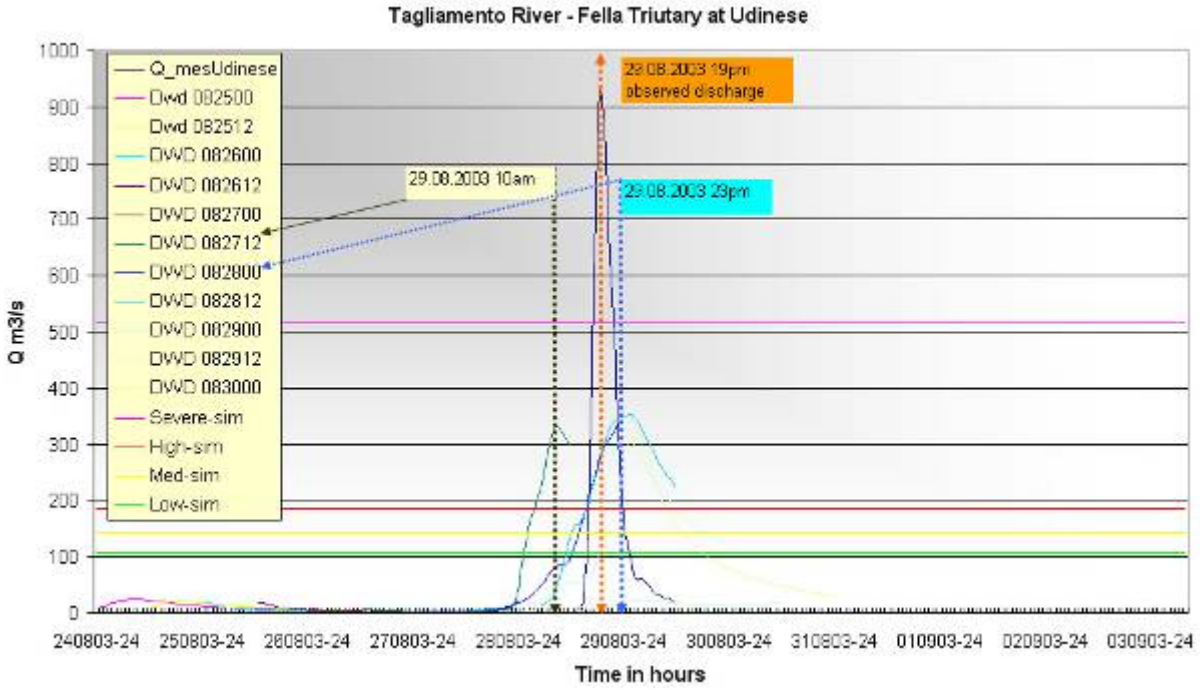


Figure 3-9: Hydrographs of observed and forecasted discharges in m3/s (y-axis) for the station Udinese in the Tagliamento basin. The flood forecasts based on DWD forecasts start on 20030825 00:00 until 20030630 00:00 in 12h time intervals.

3.4.4 Six months assessment of forecast performance for the French case study

An important control study is the long-term assessment of the distribution of hits, false alarms and misses. For this study flood forecasts from 5th June 2002 to 31st December 2002 have been carried out and the exceedance of the thresholds compared with observations. Unfortunately little information about flood events is known. Since the station information from one gauge to another may not be independent, 4 representive stations, one in each catchment, have been chosen for the analysis. Defining an event when at least the lowest alarm level has been exceeded, 4 events took place in the Gard from the 5th June 2002 to the 31st December 2002. All 4 events were predicted based on the DWD forecasts. In the Ardeche 3 events were observed but only 1 correctly forecasted, 1 missed, and 2 falsely forecasted. In the Ceze 4 observed events took place and all predicted as well as 1 false alarm. In the Virdourle 2 events took place and were correctly predicted but also 3 falsely predicted. In summary, over the 4 river basins the forecasting model would have produced 11 hits, 3 missed events and 5 false alarms. In particular, an event took place in the region on 11-15th December was spatially shifted in the forecasts producing a “missed” event in one basin and a “false alert” in another basin. It is planned to repeat this preliminary analysis taking into account more stations and possibly longer time periods. In this context also the impact of reservoirs on the events will be taken into account which was not considered in this preliminary analysis.

3.5 Summary and Conclusions

In this paper the possibility of interfacing short-range numerical weather forecasts with a spatially distributed rainfall-runoff model for early flashflood warning in ungauged river basins has been explored. The methodology is based on flood threshold exceedances where the thresholds are derived from long-term simulations with an essentially uncalibrated hydrological model. The same model is then used with weather forecast data and the model consistent thresholds applied for the analysis.

The proposed forecasting strategy addresses a number of shortcomings typically present in flashflood forecasting, namely coarse meteorological station networks and few or no discharge station data. High-quality radar data has been used for this study as the true rainfall but often these are not available in the presented quality.

Results of the study show that by looking at relative differences and model consistent thresholds early warning for flashfloods can be given with leadtimes exceeding 24 hours. In both case studies the weather forecasts captured the event well allowing to predict peak, magnitude and spatio-termporal distribution of the event well with an absolute leadtime of more than 36 hours. Taking into account computing, processing and analyses time the effective leadtime could still have been of the order of 24 hours which would be sufficient to allow authorities to take precautionary measures and have more time to act once monitoring devices confirm the event.

The results also show, however, that the principle can also be useful for those areas where such data do not exist and where the approach could greatly contribute to the preparedness for flashflood events in terms of awareness, identification of regions at risk, potential magnitude and timing of the event.

4. Flash flood warning based on rainfall depth-duration thresholds and soil moisture conditions: An assessment under European conditions

4.1 Introduction

Given the specific space-time scales of flash flood events, at least two features characterise flash flood forecasting with respect to riverine flood forecasting and point out to their larger uncertainty. These are: i) the short lead time, which implies both the integration of meteorological and hydrologic forecast, and the difficulties of using data assimilation procedures based on real time observed discharges to reduce uncertainty in hydrologic predictions, and ii) the need to provide local forecasts, which means that, on one hand, the rainfall must be monitored and forecasted on a wide range of space/time scales, and, on the other hand, every tributary of a monitored basin can be considered as a potential target for flood warning. In this sense, flash flood forecasting exemplifies the ungauged basin prediction problem under extreme conditions.

The assessment of the susceptibility to flash flood, by taking time-varying hydrologic characteristics such as soil moisture status and snowcover into account before the potential event, is a critical step to anticipate the locations of the river system which may be hitted by the flood. Even though the occurrence, location and (or) timing of the flash flood is still uncertain, this information may provide enough lead time so that flash flood mitigation measures can be planned and managed in an anticipatory rather than responsive manner. Lead time in these cases may be comprised between one day and few hours. The provision of information about susceptibility to flash flood is one of the objectives of the Flash Flood Guidance system, which is operating in the United States since 1970s (Mogil et al., 1978). According to Georgakakos (2006), the US National Weather Service relies routinely on Flash Flood Guidance (FFG, hereinafter) computations to produce flash flood watches and warnings. FFG is the depth of rain of a given duration, taken as uniform in space and time on a certain basin, necessary to cause minor flooding at the outlet of the considered basin. This rainfall depth, which is computed by running in inverse mode a lumped hydrological model, is compared to either real time-observed or forecasted rainfall of the same duration and on the same basin. If the nowcasted or forecasted rainfall depth is greater than the FFG, then flooding in the basin is considered likely. Georgakakos (2006) provided the theoretical basis of developing operational flash flood guidance systems by using analytical methods. Ntelekos et al. (2006) analysed uncertainty propagation within a simplified FFG system.

Apart from their extensive use in the United States (Georgakakos, 2006) and in Central America (Georgakakos, 2004; Sperfslage et al., 2004), in Europe, the Integrated Project FLOODSite (http://www.floodsite.net) among others aims at assessing the advantage for using the rainfall threshold approach as an alternative to the traditional ones in the case of flash floods.

Alternatives to the FFG have been proposed in the last years, generally taking advantage from the development of spatially distributed hydrological models (Moore et al., 2006; Reed et al., 2007; Bloeschl et al., 2007). However, development of the FFG concept is still ongoing. It is recognised that the FFG provides a useful concept that simplifies communication about the hydrological status of basins from hydrologists to meteorologists and that it represents a potential benchmark for further development and intercomparison purposes.

The objective of this section is to evaluate a threshold-based flash flood warning approach based on FFG by considering a wide range of climatic and physiographic European conditions, and by focusing on ungauged basins. More specifically, system results have been evaluated at gauged interior sites that were not used to calibrate the lumped hydrological model. At these sites, the model parameters were regionalised by transposition from the parent basin, and by using GIS-based techniques to derive basin-specific parameters for simulation of snow melt dynamics and

flow routing. Results derived in this way are considered indicative of expected performance at ungauged locations, under the conditions that model calibration may be carried out at larger spatial scales and that computed parameters may be transposed at smaller spatial scales.

Two alternative ways to compute soil moisture status were considered. The first consists on transpose soil moisture status, further than model parameters, from parent basins to interior basins. This technique has obvious advantages in terms of operational implementation, by reducing computational efforts to obtain FFG at several interior sites. However, soil moisture status may be biased when used at the scale of the specific interior basin, due to use of precipitation and evaporation estimates which are representative at the scale of the parent basin but potentially not at the scale of the interior basins. As a second alternative, the use of time-constant soil moisture status as an input to the threshold-based flash flood warning system was evaluated. This alternative is representative of time-constant precipitation depth-durations thresholds and it is obtained by setting the values of the model soil moisture status on their time-average value. Assessment of this procedure allows to evaluate the decrease in accuracy associated to lack of information about the temporal variation of soil moisture status before the flood event.

The simulation experiments described in this section are designed to understand the potential benefits and limitations of the flash flood warning approach and to guide further development. More specifically, modelling experiments address the four questions below.

1. Which is the impact on FFG technique accuracy due to of time-uniform precipitation? This issue is approached by comparing threshold-based results with those obtained by using the hydrological models with the observed precipitation input.

2. How simulation accuracy at ungauged interior points, simulated by using transposed parameters from parent basins, compare with results obtained for parent basins where calibration has been carried out?

3. Which is the decrease in simulation accuracy associated to transposing both parameter and soil moisture from the larger scale parent basins?

4. How technique performance degrades when a time-constant soil moisture status is used?

4.2 Study Areas and Data

Data from two distinct European regions were used in this study: north eastern Italy (with eight basins) and central France (with three basins). Figure 4-1 shows the location of the basins. Table 4-1 provides more detailed basin information, with Table 4-2 and * Discharge data available since 1/10/1993

Table 4-3 providing information on the length period with hourly data available and division among calibration and validation period. Table 4-4 provides the information about the topological connection between parent basins and nested basins. Two parent basins contain two nested basins each, and other two parent basins contain one nested basin each.

Drainage area is comprised between 116 km^2 and 3244 km^2 for the parent basins, and between 7.3 km^2 and 233 km^2 for the nested basins. The size of the largest basin (Loire river at Bas-en Basset, with 3244 km^2) is at the limit of the spatial scales usually met for flash flood analysis. This case has been used here to assess the performance of the system with increasing the basin scale. The second largest basin is the Dunierés at Vauberlet, with 233.4 km^2.

The topography of these basins is in general rather complex, with some high altitude basins (Ridanna, Cordevole at Saviner, Cordevole at Vizza), characterised by top altitudes exceeding 3000 m a.s.l., and elevation range comprised between 800 and 2000 m. Floods in some of these basins may be influenced by snow accumulation and melt. Even though only rarely these floods are short and intense enough to be characterised as flash floods, snow-related processes are important elements for characterisation of the seasonal hydrological balance. It has been decided therefore to include dynamics of snow accumulation and melt in the modelling strategy.

In general, the river regime of these river systems is altered in a negligible way by management activities, such as artificial reservoirs and diversions. However, the regime of the Brenta river is influenced by two relatively large natural lakes (Caldonazzo and Levico), with 77 km^2 area drained by the lakes. In this case, the influence of the natural lakes on the river regime has been taken into account by subdividing the basin into subunits and simulating the effects of the lakes. Artificial reservoirs exist on the Loire river. The 200 km^2 upstream part of the Loire watershed is equipped with dams for hydropower plants and 350 km^2 of the Lignon watershed downstream Le Chambon sur Lignon is controlled by a dam built for the water supply of the City of Saint Etienne. These reservoirs are managed to be as "transparent" as possible during floods; as a consequence, their influence on flood flows is very limited especially at Bas-en-Basset. No significant dams exist on the Gagne and Dunieres river systems.

Karstified aquifer influence the runoff response for specific portions of the Posina and Brenta river systems.

Annual runoff coefficient range from rather low values (around 0.42) for the French basins to relatively high values for some Italian high altitude basins (0.8 for Ridanna). The Budyko's climatic classifications scheme (Budyko, 1974) has been applied to compare and contrast the climatic characteristics of these basins. This is achieved by presenting the specific response of each of these basins on the Budyko curve (Figure 4-2), which is a plot that expresses E/P, the ratio of average annual actual evapotranspiration (E) to average annual precipitation (P) as a function of EP/P, the ratio of average annual potential evapotranspiration (EP) to average annual precipitation (P). Actual evapotranspiration (E) for each basin was derived as the long-term difference between P and R (runoff) for the basins. Figure 4-2 shows clearly that the Italian basins (2, 3, 4, 5, 8, 10) represent a wet climate, whereas the French basins (1, 6, 7) have a medium climate.

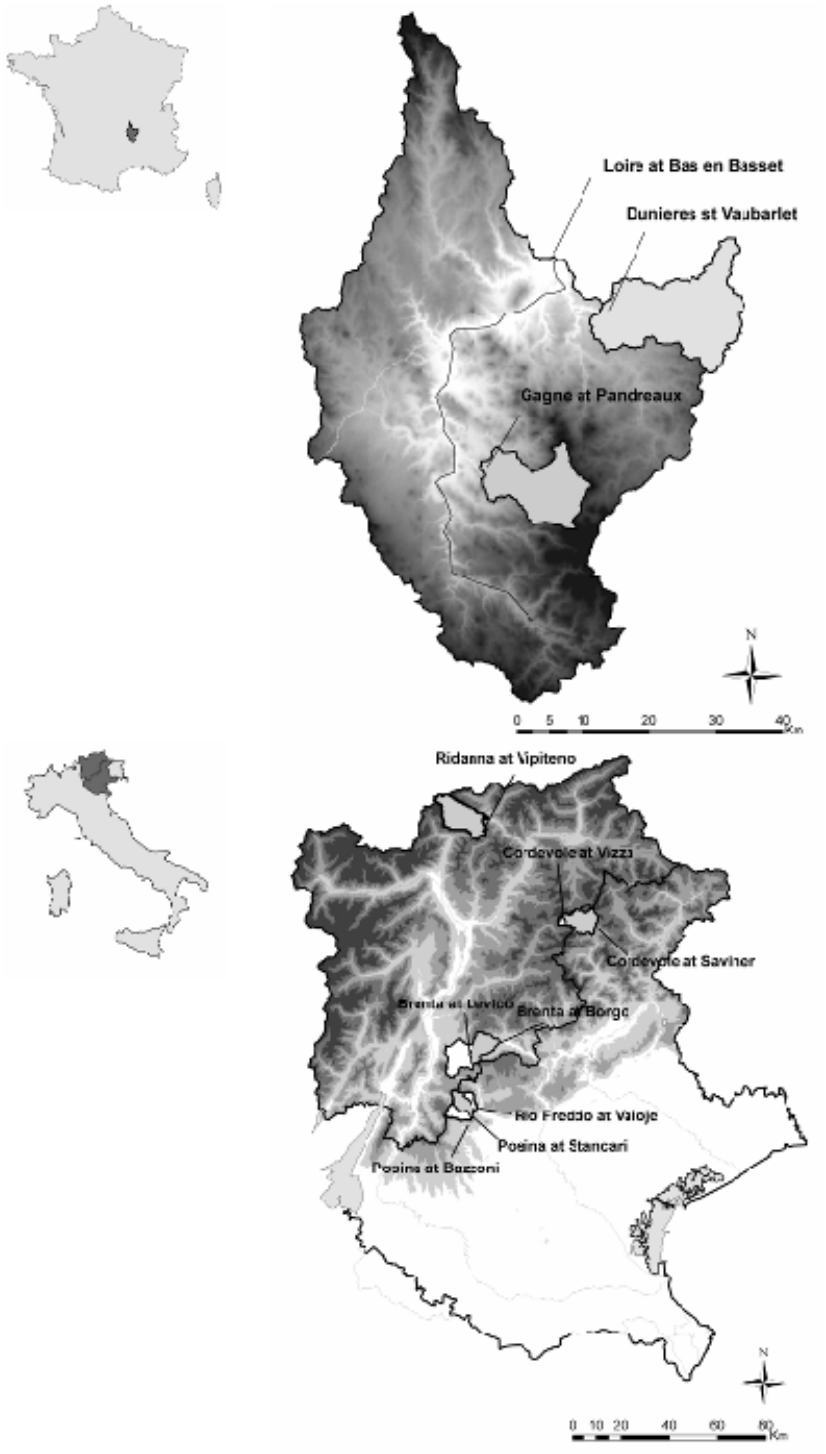


Figure 4-1: Study basins and their location in France and Italy

Table 4-1: Basins characteristics

Station name	Basin number	Area (km^2)	Elevation range (m)	Mean Annual Rainfall (mm)	Mean Annual Runoff (mm)	Runoff Coefficient
Loire at Bas-en-Basset	1	3244	450 – 1800	888	361	0.41
Cordevole at Saviner	2	109	1025 - 3200	1110	770	0.69
Posina at Stancari	3	116	388 – 2300	1645	1000	0.61
Brenta at Borgo	4	213.7*	380 - 2400	1068	650	0.61
Ridanna at Vipiteno	5	210.2	940 - 3600	1271	1016	0.80
Gagne at Pandreaux	6	121.9	600 – 1500	909	393	0.43
Dunieres at Vauberlet	7	233.4	584 - 1400	939	391	0.42
Cordevole at Vizza	8	7.3	1810 - 3200	1145	880	0.77
Posina at Bazzoni	9	38.8	453 - 2300	1717	1086	0.63
Rio Freddo at Valoje	10	22.2	390 - 2000	1549	835	0.54
Brenta at Levico	11	113*	435 - 2000	1088	-	-

* In the case of Brenta basin 77km^2 of the area is drained by natural lakes.

Table 4-2: Periods with data available: parent basins

Station name	Basin number	Periods with hourly data available	Calibration period	Validation period
Loire at Bas-en-Basset	1	1/10/1990 1/10/2003	1/10/1997 1/10/2003	1/10/1990 30/09/1997
Cordevole at Saviner	2	1/10/1992 1/10/2000*	1/10/1993 30/09/1997	1/10/1997 1/10/2000
Posina at Stancari	3	1/10/1992 1/10/2000	1/10/1992 30/09/1996	1/10/1996 1/10/2000
Brenta at Borgo	4	1/10/1994 1/10/2005	1/10/1994 30/09/2001	1/10/2001 1/10/2005
Ridanna at Vipiteno	5	1/10/1992 1/10/2004	1/10/1992 30/09/1997	1/10/1997 1/10/2004

* Discharge data available since 1/10/1993

Table 4-3: Periods with data available: interior points

Station name	Basin number	Periods with hourly data available
Gagne at Pandreaux	6	1/10/1998 - 1/10/2003
Dunierés at Vauberlet	7	1/10/1990 - 1/10/2003
Cordevole at Vizza	8	1/10/1992 - 1/10/2000
Posina at Bazzoni	9	1/10/1993 - 1/10/2000
Rio Freddo at Valoje	10	1/10/1993 - 1/10/2000
Brenta at Levico	11	1/10/1994 - 1/10/2005 (only flood data)

Table 4-4: Relationship among parent and nested basins

Parent basins	Nested basins
1	6, 7
2	8
3	9, 10
4	11
5	-

Figure 4-2 shows also significant climate variability among parent and nested basins. This is the case for Cordevole at Saviner (2) and at Vizza (8), where differences are due mainly to different elevation ranges among parent and interior basin, and for Posina at Stancari (3) and Rio Freddo (10), where differences are mainly due to the differentiated impact of the karstified aquifer.

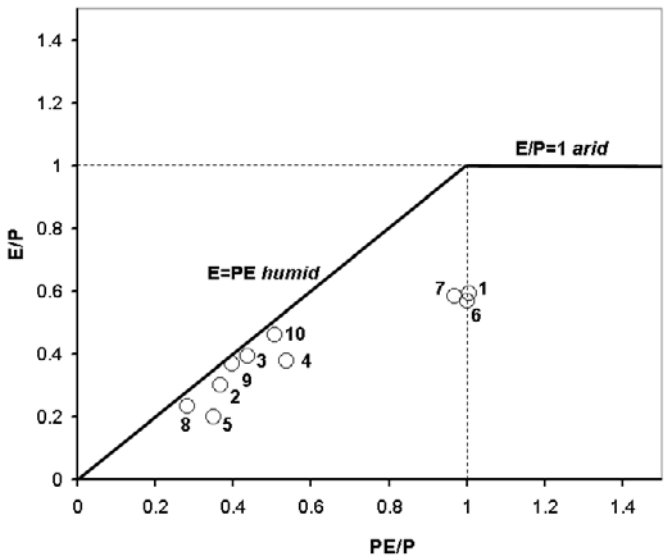


Figure 4-2: Plot of mass balance data from the study basins on the Budyko curve.

The length of the hourly record of streamflow, precipitation and temperature data ranges from 5 to 13 years, with a total of 101 years. The data were quality controlled and as a results part of the record was set to missing. Basin-averaged precipitation estimates were obtained based on rain gauge stations by using a Thiessen technique, with densities ranging from 1 station per 15 km^2 (Brenta river basin) to 1 station per 140 km^2 (Loire river basin).

The stage-discharge relationship at the interior streamflow gauge of Brenta at Levico is considered exceedingly uncertain for low flows. Due to this reason, only flood discharge data were used.

Digital Elevation Model (DEM) at four different resolutions were used: 75 m for the French basins, 30 m for Brenta and Ridanna, 25 m for Cordevole and 20 m for Posina.

4.2.1 Rationale for basins selection

The study basins in Figure 4-1 were selected for several reasons. First, these basins had the data required to conduct the intercomparison, with concurrent time series of hourly rainfall, temperature and discharge data made available for the basin outlets and selected interior points. The quality of the data available is representative of operational conditions, subject to complexities due to rough orography and high space-time variability. Lack of significant modification of the streamflow due to reservoirs and diversions simplifies the intercomparison study.

A second critical criterion for selection is the observation of past flash flood events in these basins and their representativeness of conditions leading to flash flood.

Finally, the selected parent basins contain internal points having observed streamflow data, allowing to develop study questions regarding the prediction of interior hydrologic processes.

The Ridanna basin has no interior gage locations. This basin represents an additional case for testing the threshold methodology over a high-altitude alpine basin frequently hitted by small scale flash flood, usually triggering shallow landsliding and debris flows.

Lastly, the hydrometeorology of flash flooding in these areas has been widely studied. Borga & Vizzaccaro (1997) and Dinku et al. (2000) analysed estimation uncertainties of flood-generating storms based on raingauges and radar observations for the upper Astico river system (Posina). In the same region, Borga et al. (2000) and Hossain et al. (2004) examined the impact of errors in radar-based rainfall estimates on flood prediction uncertainty.

4.3 The threshold-base technique

The threshold-based technique applies to a given basin and for a given rainfall duration, and it is based on the comparison between the FFG and the either real time-observed or forecasted rainfall. If the nowcasted or forecasted rainfall depth is greater than the FFG, then flooding in the basin is considered likely. It is important to recognise that the FFG technique does not predict flash flood timing, but only that a flood threat is imminent. The major effort within the FFG is the correct assessment of the flood magnitude, while the correct timing forecast is left to the monitoring activity triggered by the flash flood alert.

In this study, the FFG is obtained by running a lumped hydrological model for several hypothetical rainfall amounts for the selected rainfall duration and given current soil moisture conditions. Five rainfall durations are considered: one, three, six, twelve and twenty four hours. The lumped soil-moisture accounting hydrological model used in the methodology is described next.

4.3.1 Description of the hydrological model

The model used in this study is a semi-distributed conceptual rainfall-runoff model, following the structure of the PDM (Probability Distributed Moisture) model (Moore, 1985). The model runs on a hourly time step and consists of a snow routine, a soil moisture routine and a flow routing routine.

The snow routine represents snow accumulation and melt by using a distribution function approach based on a combined radiation index degree-day concept (Cazorzi & Dalla Fontana, 1986). Snow melt is computed as follows:

$$M = \begin{cases} f_m EI\,(T_h - T_0) & T_h > T_0 \\ 0 & T_h \le T_0 \end{cases} \tag{4.1}$$

where M is the melt rate [mm h^{-1}], T_h is the hourly mean temperature [°C], T_0 [°C] is a threshold temperature beyond which melt is assumed to occur, fm is a melt factor and EI [J m^{-2} h^{-1}] is an energy index which represents the potential radiation energy (variable in time) for a given site in the basin. EI is computed for each topographic element of the basin taking into account solar altitude angle, optical depth of the atmosphere, elevation, aspect, slope and shading effects. The basin is subdivided into temperature bands (generally ranging 200 m in elevation), and for each band the empirical distribution of the energy index is used in the lumped form of the snow melt module.

Catch deficit of the precipitation gauges during snowfall is corrected by a snow correction factor, *SCF*. A threshold temperature interval $T_R - T_S$ is used to distinguish between rainfall, snowfall and a mix of rain and snow. The model includes routines able to compute runoff produced during rain-on-snow events.
Potential evapotranspiration is estimated by using the Hargreaves and Samani method (Hargreaves and Samani, 1982).

The soil moisture routine uses a probability distribution to describe the spatial variation of water storage capacity across a basin. Saturation excess runoff generated at any point in the basin is integrated over the basin to give the total direct runoff entering the fast response pathways to the basin outlet. Drainage from the soil is subdivided into a subsurface and a baseflow component. Storage representations of the fast, medium(subsurface flow) and slow response pathways yield a fast, medium and slow response at the basin outlet which, when summed, gives the total basin flow. The PDM model configuration used here employs a Pareto distribution of storage capacity, c. This has the distribution function

$$F(c) = 1 - \left[1 - \left(c / c_{\max}\right)\right]^b \tag{4.2}$$

where *cmax* [mm] is the maximum storage capacity in the basin and the parameter *b* [-] controls the degree of spatial variability of storage capacity over the basin. The instantaneous rate of fast runoff generation from the basin is obtained by multiplying the rainfall rate by the proportion of the basin which is saturated

Losses due to evaporation are calculated as a function of potential evaporation and the status of the soil moisture store. The dependence of evaporation loss on soil moisture content is introduced by assuming the following simple function between the ratio of actual to potential evaporation, *E/EP*, and soil moisture deficit, *Smax-S(t)*:

$$\frac{E(t)}{EP(t)} = 1 - \left\{ \frac{\left[S_{\max} - S(t)\right]}{S_{\max}} \right\}^{b_e} \tag{4.3}$$

where *Smax* [mm] is the total available basin storage, be [-] is an exponent coefficient and *S(t)* [mm] is the basin moisture storage at time t. Drainage to the slow flow path, d [mm h^{-1}], is represented by a function of basin moisture storage *S(t)* such that

$$d = \left(k_g\right)^{-1}\left[S(t) - S_t\right]^{b_g} \tag{4.4}$$

where the parameters are a time constant *kg* [h mm bg-1], an exponent coefficient *bg* [-] and a threshold storage *St* [mm] below which there is no drainage.

The slow or base flow component, q_s [mm h^{-1}], of the total runoff is assumed to be routed through an cubic storage which is in general more suitable to represent the groundwater emptying (Moore et al., 2002) such that

$$q_b = kS^3 \tag{4.5}$$

where S_s [mm] is the depth of storage and k_s [$h^{-1}mm^{-2}$] is the decay parameter of the store.

Direct runoff from the proportion of the basin where storage capacity has been exceeded is routed by means of a geomorphology-based distributed unit hydrograph. With this procedure, a geomorphologic filter based on a threshold drainage area (A_{th}) is used to distinguish hillslopes and channel network starting from the space-filling representation of the drainage system directly obtainable from DEMs (Montgomery & Foufoula-Georgiou, 1993; Da Ros & Borga, 1997a). The routing time of each site in the basin is evaluated assigning different typical velocity values in each pixel pertaining to the basin and classified as hillslope or channel. The two velocities, v_h and v_c, used to describe the flow routing process in each of the two components of the drainage system are assumed here constant; they maintain a physical meaning as the average velocities on hillslopes and in channel network. Total runoff is computed as the sum of slow and fast runoff.

4.4 Assessment methodology

4.4.1 Model application

Three different strategies were considered to implement the methodology: i) model parameter calibration; ii) model parameter transposition from parent basin to interior sites; iii) model parameter and soil moisture status transposition from parent basins to interior points. These strategies can be described as follows.

Model parameter calibration

The goal of calibration is to adjust the model's parameters to decrease the difference between observed and simulated streamflow values. The closeness of fit can be checked qualitatively (e.g. plots of observed and simulated hydrographs) or quantitatively (residual statistics such as the Bias, Nash-Sutcliffe efficiency, etc.). In this study, the Shuffled Complex Evolution-University of Arizona (SCE-UA, Duan et al., 1992) global optimization algorithm was used for calibration of the hydrological model parameters over the five parent basins. The SCE-UA global search procedure is based on the downhill simplex method (Nelder & Mead, 1965), combined with a random search procedure and the idea of complex shuffling. Even though the final objective of the study is to obtain a good description of flood events, equal weight was placed to the reconstruction of low flows, in an effort to improve the description of soil moisture conditions before the flood events.
The following objective functions were used during the optimization process for this study:

1. the Nash and Sutcliffe (1970) coefficient of efficiency defined as:

$$E_{NS} = 1 - \frac{\sum_{i=1}^{n}(O_i - S_i)^2}{\sum_{i=1}^{n}(O_i - O_{ave})^2} \tag{4.6}$$

where O_i is the hourly *i-th* observed discharge, S_i is the simulated discharge, and O_{ave} is the mean value of the observed discharges. The coefficient of efficiency was selected because it is dimensionless and is easily interpreted. If the model predicts observed streamflow with perfection then $E_{NS}=1$. If $E_{NS}<0$ then the model's predictive power is worse than simply using the average of the observed values.

2. the relative bias (RB) defined as

$$RB = \frac{\sum_{i=1}^{n}(S_i - O_i)}{\sum_{i=1}^{n} O_i} \tag{4.7}$$

RB is a measure of total volume difference between observed and simulated streamflows, and is important in the evaluation of simulations from continuous hydrologic models.

A simple split sample test (Klemes, 1986) was considered for calibration and validation of the hydrological model. The test involves dividing the available data into two sets, one used for parameter estimation (calibration period) and the other for validation (validation period).

Model parameter transposition

The process of transferring information (such as basin model parameter values) from neighbouring basins to the basin of interest is generally referred to as hydrological regionalisation (Bloeschl & Sivapalan, 1995). Numerous regionalisation methods have been proposed in the literature for the case of basin model parameters (Bloeschl, 2005). Merz & Bloeschl (2004) examined the performance of various methods of regionalising the parameters of a conceptual basin model in 308 Austrian basins. They concluded that the methods based on spatial proximity performed better than those based on physiographic basin attributes. Similar finding were reported by Kokkonen et al. (2003) and Parajka et al. (2005). Kokkonen et al. (2003, p. 2219), concluded that "when there is a reason to believe that, in the sense of hydrological behaviour, a gauged basin resembles the ungauged basin, then it may be worthwhile to adopt the entire set of calibrated parameters from the gauged basin instead of deriving quantitative relationships between basin descriptors and model parameters". One of the advantages of the similarity approach may be that the complete set of model parameters is transposed from a donor basin.

In this study, a similarity approach was used for the parameter estimation of the interior gauges based on transposing the complete set of parameters from the parent basin to the interior basin. This verification strategy aims to obtain results which are indicative of expected performance at ungauged locations, under the conditions that model calibration may be carried out at larger spatial scales and that computed parameters may be transposed at smaller spatial scales.

Model parameter and model soil moisture status transposition

With this strategy, FFG values are obtained at interior sites based on transposing both model parameters and soil moisture status from the parent basin. This implies that the model is run based on input data (precipitation and temperature) for the parent basin. This methodology has obvious advantages in terms of operational implementation, by reducing computational efforts to obtain FFG at several interior sites. The model is run only at the level of the parent basin, and FFG computations are carried out for the specific interior sites. On the other hand, the quality of the model-based soil moisture status estimates obtained in this way may be altered. Biases in rainfall and temperature accumulate over weeks and months and soil moisture status are not as accurate as those obtained by running the model over the specific interior basins.

Figure 4-3 shows one year of simulation results at the outlet for the Cordevole basin at Vizza. Figure 4-3(a) shows simulation results with model parameter transposition from Cordevole at Saviner, whereas Figure 4-3(c) shows simulation results with model parameter and soil moisture status transposed from Cordevole at Saviner. Figure 4-3(b) and Figure 4-3(d) report simulation residuals in the two cases, respectively. These figures show that for this basin the two approaches yield hydrologically acceptable representations of the watershed behavior. At this scale, the two hydrographs appear visually similar. Only small differences can be seen, e.g. model with parameter transposition is biased low during the autumn floods (5900-6500), whereas model with parameter and soil moisture status is biased high during the same period. Figure 4-3(e) shows relative soil moisture content of the PDM storage obtained from model parameter transposition at Vizza and from model simulation at Saviner. This figure shows clearly that there is a large difference between the two soil moisture statuses during the period from October (6500), when snow accumulation starts on the basin, to late March (1800) when snowmelt begins. This difference is due to the different impact that solid precipitation has on the hydrological behavior. For Vizza, almost all precipitation after October fall in solid phase, providing negligible input to the soil moisture store. On the contrary, on the lower basin closed at Saviner most of the precipitation falls in liquid form and feeds the PDM storage.

However, this bias has apparently a negligible impact on the simulation of the summer and fall floods, since the hydrological status of the two basins is reset during snowmelt. A smaller bias can be identified during the summer period, this being explained by the different precipitation and evaporation accumulations on the two basins.

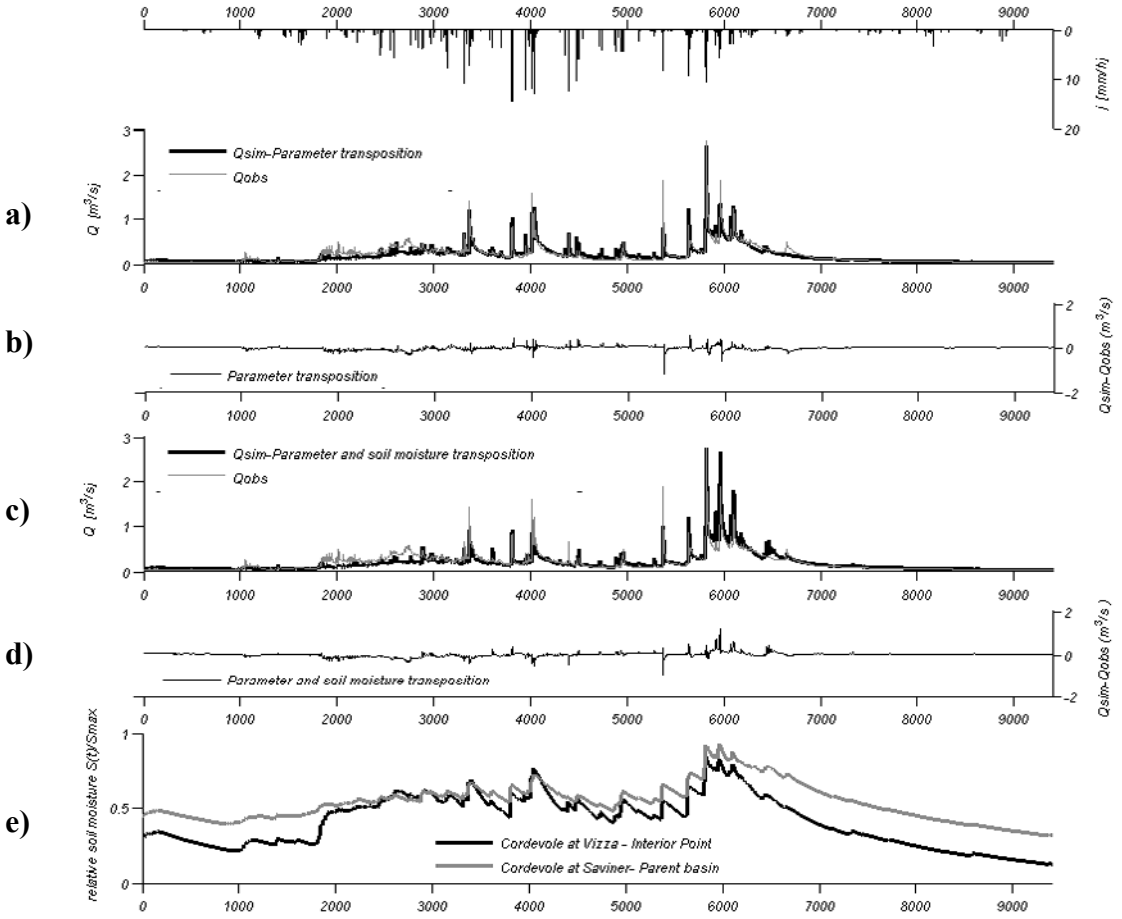


Figure 4-3 a-e: One year (01.01.1993-31.12.1993) of hourly results at the outlet for the Cordevole river at Vizza (a) simulation results with model parameter transposition from Cordevole at Saviner; (b) simulation residuals, (c) simulation results with model parameter and soil moisture transposition from Cordevole at Saviner; (d) simulation residuals (e) relative soil moisture content of the PDM storage obtained from model parameter transposition at Vizza and from model simulation at Saviner.

4.4.2 FFG assessment

Five rainfall durations are considered for computation of the FFG: one, three, six, twelve and twenty four hours. The model is run continuously in time, and five values of FFG are computed each day (at 12:00) for each considered basin. Selection of the time during the day when FFG is computed has been shown to have negligible impact on final results. For the considered day, the FFG values are compared with the maximum estimated areal precipitation over the corresponding five durations. The technique predicts the exceedance of the threshold flooding (i.e.,a flash flood warning would be issued) when estimated precipitation exceeds FFG for at least one precipitation duration.

Assessment of the quality of flash flood forecasts based on FFG estimates is obtained by using contingency tables. Contingency tables are highly flexible methods that can be used to estimate the quality of a deterministic forecast system (Mason and Graham, 1999) and, in their simplest form, indicate its ability to anticipate correctly the occurrence or non occurrence of predefined events. A four-cell contingency table can be constructed which depicts the relationship between the forecasts and the events. Consider a set of forecasts that can have only two alternatives (e.g., yes, no) (Table 4-5). Let:

- X denote the number of positive forecasts that correspond to an occurrence of the event (hits).
- Y denote the number of events that occurred in conjunction with a negative forecasts (missed events).
- Z denote the number of positive forecasts that were not accompanied by an event (false alarms).
- W denote the number of negative forecasts that did not have any associated events.

Table 4-5: Four-cell contingency table used in the study

	FORECASTS	
EVENTS	**YES**	**NO**
YES	*X*	*Y*
NO	*Z*	*W*

These statistics can be used to summarise the contingency table:

- the probability of detection (POD). It is the ratio of correctly forecasted events to the total number of events

$$POD = \frac{X}{X+Y} \tag{4.8}$$

the range of values for POD goes from 0 to 1, the latter value being desirable. A POD of one means that all occurrences of the event were correctly forecast.

- the false alarm rate (FAR). It is the ratio of the number of false alarms to the total number of predicted events:

$$FAR = \frac{Z}{X+Z} \tag{4.9}$$

The range of values for FAR goes from 0 to 1, the former value being desirable. A FAR of zero means that in the verification sample, no non-occurrences of the event were forecast to occur.

Neither POD nor FAR can give a complete picture of forecasting success; it is therefore desirable to include a statistic depending on both POD and FAR. This is the critical success index (CSI) (Schaefer, 1990; Wilks, 1995). The CSI is the ratio of correctly forecasted events to the total number of event forecasts that were either made (X+Z) or needed (Y):

$$CSI = \frac{X}{Y+X+Z} = \frac{1}{POD^{-1} + (1-FAR)^{-1} - 1} \tag{4.10}$$

For either a zero POD or a unit FAR, the value of CSI is uniquely equal to zero, since there are no hits. The range of values for CSI goes from 0 to 1, the latter value being desirable.

Results obtained from the threshold-based methodology are compared with corresponding results from two alternatives methodologies. With the first alternative, FFG is contrasted with the temporal-detailed hydrological model (*Model*, hereafter). This provides the study with an evaluation of the assumption of time-uniform rainfall implied by the FFG. Scores statistics are obtained by comparing streamflow predicted by the model with the observed events, for each day, in terms of exceedance of the flooding threshold.

With the second alternative, FFG is contrasted with use of a time-constant soil moisture status as an input to the threshold-based flash flood warning system. This alternative is representative of time-constant depth-durations precipitation thresholds (*Constant*, hereafter). These constant depth-duration precipitation values are derived by setting the values of the model soil moisture status to those corresponding to the annual average discharge value. Assessment of this procedure allows the study to evaluate the degradation of accuracy associated to the loss of information about the temporal variation of antecedent soil moisture status.

4.4.3 Threshold flooding conditions

In this study the FFG is computed based on two different threshold flooding conditions. The first one (called hereafter High Threshold – HT) is based on the bankfull flow, characterised by 2-year return time. Carpenter et al. (1999) suggest that a 2-year flood is a reasonable threshold to use for flood warnings given that the flood flow associated with damage or hazard is often a little higher than bankfull flow. Use of this definition led to identification of 55 flood events exceeding the basin-specific thresholds, over the whole archive of streamflow data. However, use of this definition may give rise to sampling problems for the basins characterised by short data record length, due to the small number of local flood events. Owing to this reason, we use also a Low Threshold (LT), characterised by a return time around 0.5 year, corresponding to 223 flood events exceeding the threshold.

The assessment strategy comprises therefore three different procedures for model implementation, three different procedures for FFG assessment, and two threshold flooding conditions.

In the assessment procedure report below, system performances obtained on interior basins are compared with those obtained on parent basins. This implies that system performances on parent basins is comparable to the one that would have been attained on interior basins, in the case of model parameter calibration. Our results, which are not reported here for the sake of brevity, show that score statistics obtained on interior basins are within the range of those of parent basins, when model parameter calibration is carried out.

4.5 Results

4.5.1 Model application

Results on the three different methodologies used for model implementation are reported in Table 4-6 and

Table 4-7 where the coefficient of efficiency (E_{NS}) and the relative bias (RB) are reported for the whole data period. For the parent basins, where the model has been calibrated, results concerning both the calibration and the independent validation period are also reported. Efficiency values for calibration and validation are relatively homogeneous, with the exception of Brenta at Borgo, where the validation period was considerably wetter than the calibration period. Efficiency is lower than average for the largest basin considered in the study (Loire at Bas-en-Basset, (1)) and for Brenta at Borgo (4). For the Loire basin, this suggests that there may be a mismatch between time and space scales of the hydrological model representation for this basin. In other words, the size of this basin and its inherent spatial variability are such that a lumped representation of precipitation and hydrological processes does not ensure a correct description of hourly streamflow

temporal variability. In the case of the Brenta river, relatively poor model accuracy is due to the combined influence of lake storage, on one hand, and karstified aquifer, on the other.

Inspection of results reported for the interior gauges by using transposition of model parameters shows that efficiencies are generally degraded. Comparing overall efficiency computed on the whole simulation period shows that the coefficient of efficiency decreases by 23%, from 0.74 to 0.57. Ranking of efficiencies is not always respected when moving from parent to interior points: efficiency of model application at interior points with high (low) parent efficiency, is not always high (low). For instance, Rio Freddo at Valoje (10) is characterised by the lowest efficiency (0.20), whereas the parent (Posina at Stancari, (3)) has the highest efficiency (0.86). This may be due to the effect of the karstified aquifer, which influences Rio Freddo more significantly than its parent. The bias (both high and low) is also inflated when moving from parent to interior points.

Table 4-6: Model validation and calibration results. Parent basins

Parent basins	Calibration period		Validation period		Whole simulation period	
	E_{NS}	RB (%)	E_{NS}	RB (%)	E_{NS}	RB (%)
1	0.72	7	0.65	-4.25	0.66	2.33
2	0.72	5.35	0.75	0.78	0.73	3.42
3	0.76	1.1	0.77	2.2	0.86	2.31
4	0.71	2.64	0.54	8.74	0.64	5.02
5	0.80	1.36	0.78	-2.45	0.8	3.56

Table 4-7: Model validation results. Interior Points

Interior Points	parameter transposition		parameter and soil moisture transposition	
	E_{NS}	RB (%)	E_{NS}	RB (%)
6	0.64	-4.3	0.61	-13.2
7	0.46	-8.1	0.40	-15.2
8	0.69	1.4	0.75	-8.2
9	0.65	-6.2	0.68	-12.1
10	0.4	9.3	0.20	18.2
11	-	-	-	-

In spite of these observations, it is interesting to note that for three interior basins efficiency is larger than 0.6. This supports the view than transposing parameters from a donor to a similar basin

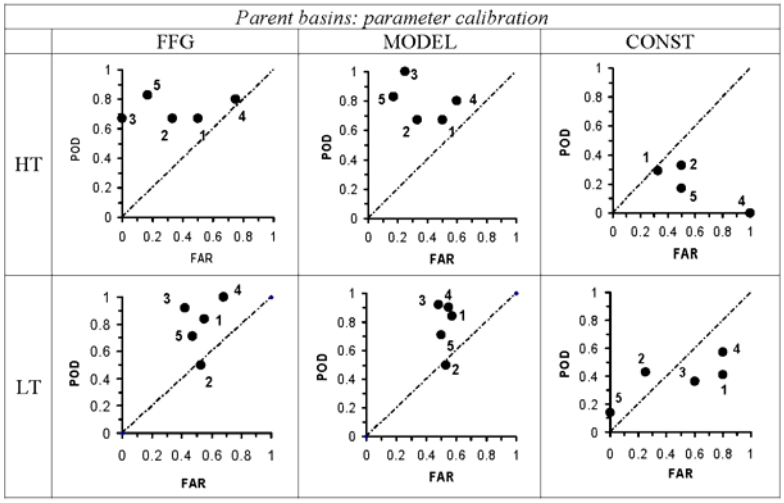
has the potential to ensure reasonable performances in regionalisation efforts, even at the hourly time step used in this study.

Transposition of both model and soil moisture status values parameters is associated to a further slight degradation of efficiency (from 0.57 to 0.53) and to a large inflation of bias, with RB values up to 18.2%. This is clearly an effect of using biased soil moisture values in the model framework.

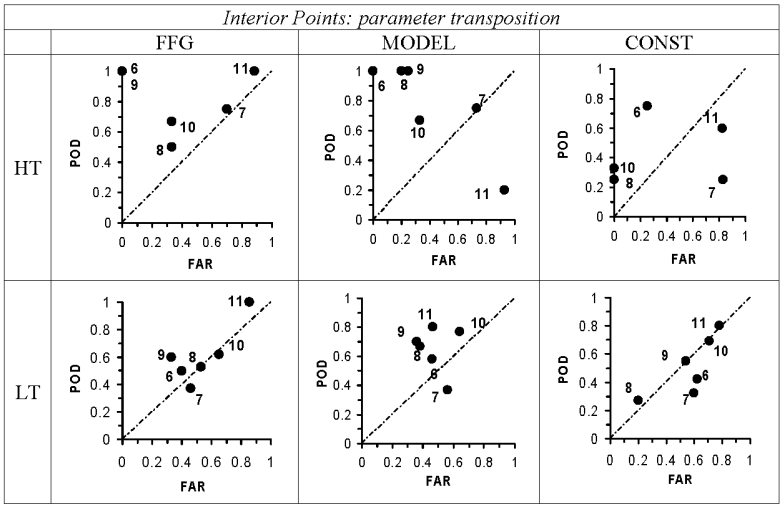
4.5.2 FFG assesment

Results concerning the threshold-based methodology are reported by using FAR and POD scores for each basin in Figure 4-4a-c, for i) parent basins, ii) interior points with model parameters transposition, and iii) interior points with model parameter and soil moisture status transposition, respectively. For each type of model application, the three different strategies of FFG assessment and the two flooding thresholds are considered. Corresponding CSI values are reported in Figure 4-5a-c, which report scatter plot of CSI values from FFG versus CSI values obtained from model application and use of constant threshold.

a)



b)



c)

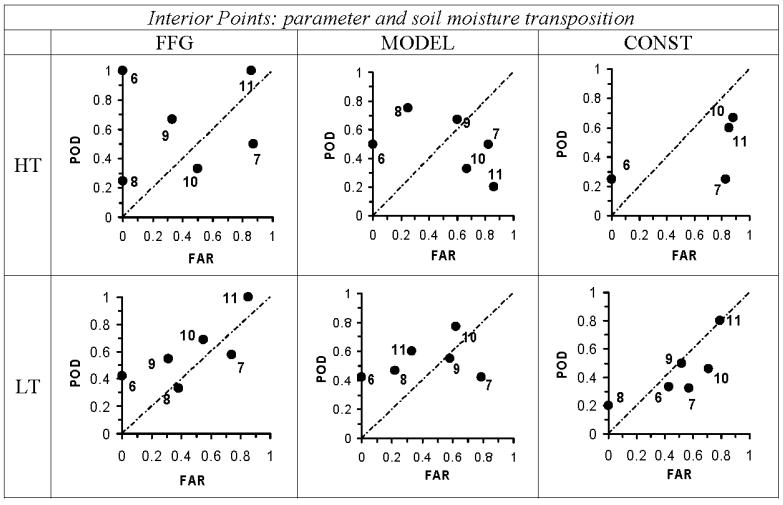
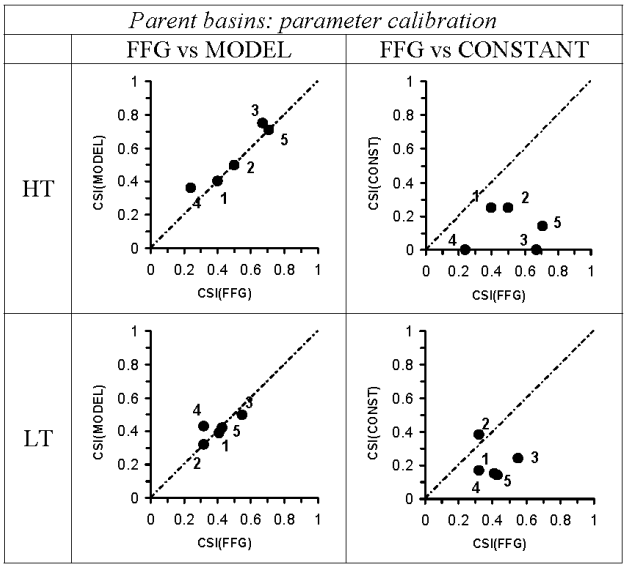
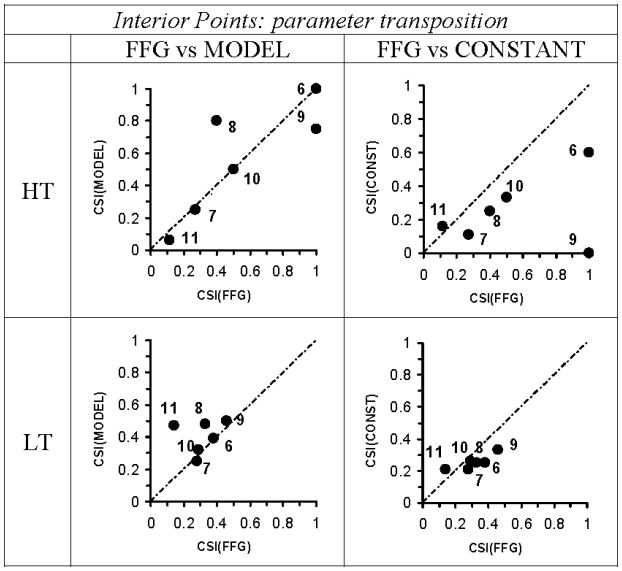


Figure 4-4a-c. Evaluation of the threshold-based technique for: a) parent basin with model calibration; b) interior basins with model parameter transposition; c) interior basins with both model parameter ans soil moisture status transposition. POD and FAR scores are plotted.

a)



b)



c)

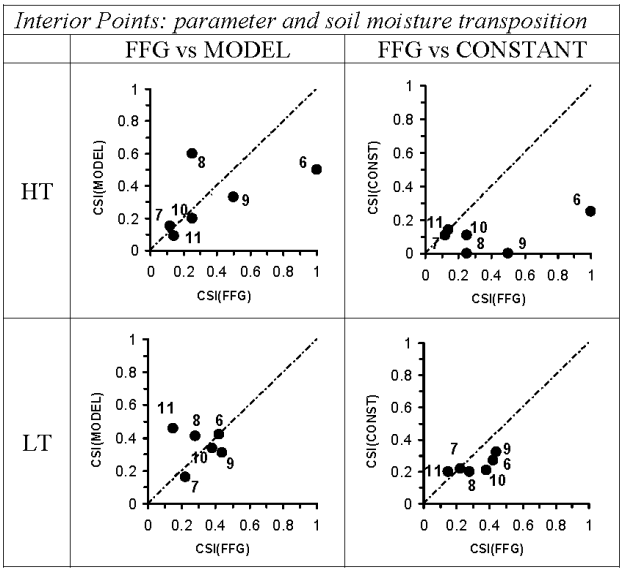


Figure 4-5a-c. Evaluation of the threshold-based technique for: a) parent basin with model calibration; b) interior basins with model parameter transposition; c) interior basins with both model parameter and soil moisture status transposition. CSI scores are plotted.

Parent basin: model parameter calibration

Results reported in Figure 4-4(a) with application of the Flash Flood Guidance shows that good results are generally obtained for the high flooding threshold. In this case, POD is always higher than 0.6, and FAR is less than 0.5 (with the exception of Brenta, with FAR equal to 0.72). For two basins (Posina at Stancari (3) and Ridanna (5)), POD is larger than 0.64 and FAR is less than 0.19. Performances are slightly degraded for some basins when using the low flooding threshold. This is the case for Posina (3) and Ridanna (5), with an increase of FAR, and for Cordevole (2), with a decrease of POD. However, this does not occur for the Loire (1) and for Brenta at Borgo (4). Inspection of model simulations (not reported here) shows that this is an effect of modelling uncertainties of relatively small rain-on-snow events during the snowmelt season. As such, these effects have more impact on basins more affected by snow-related processes. The influence of these events is larger for the LT scenario than for the HT scenario.

Comparison of score statistics obtained by using the Flash Flood Guidance with those resulting from direct model application shows a slight degradation of system performance. This provides an indication that the use of time-uniform rainfall in the FFG context has a limited impact on forecast accuracy, at least for the parent basin. CSI values in Figure 4-5(a) for Model application range between 0.4 and 0.8, whereas corresponding values for FFG range between 0.22 and 0.7. It is interesting to note that the CSI ranking is very similar to the ranking of the same basins in terms of model efficiencies, with Posina (3) and Ridanna (5) ranking high and Loire (1) and Brenta (4) ranking low. This means that, as expected, high (low) efficiency in model application translates into high (low) CSI values.

System performances degrade considerably when a constant depth-duration precipitation threshold is used. In this case, POD is always less than 0.6, and FAR may be as high as one. CSI values are less than 0.4, with most of the basins comprised between 0 and 0.3. This shows, as expected, that information on antecedent soil moisture status is essential for flash flood forecasting in these humid to medium climate basins.

Interior points: model parameter transposition

Results reported for the interior points by using transposition of the model parameters from the parent basins show that the variability in performances (for instance, CSI) among the various basins increases considerably, with respect to parent basins. Whereas this may be due to differences between parent and interior basins, we note here that absence of local calibration may introduce random error and bias which translate into inflated variability in performances between basins. For two interior basins (Dunières (7) and Brenta at Levico (11)) CSI is rather low, with values less than 0.3 for both FFG and Model. One of these basins, Dunières, ranks low also in terms of model efficiency. The three basins which rank high in terms of model efficiency (Cordevole at Vizza (8), Posina at Bazzoni (9) and Gagne (6)) have also relatively high CSI. In general, CSI at interior points with high (low) parent CSI, is also high (low).

Basin size and climate have apparently no impact on CSI. Medium size basins, characterised by humid and medium climate, may exhibit high (Gagne (6) and Posina at Bazzoni (9)) and low (Brenta at Levico (11) and Dunières (7)) CSI values. A small size basin like Cordevole at Vizza (8) range in between these two extremes.

For interior points, results obtained from FFG are generally comparable with those obtained the model, as already observed for parent basins. However, one exception can be noted: Cordevole at Vizza (8). This interior points have the smallest basin area compared to the other basins. Since small spatial scale implies reduced response times, this means that safe implementation of the FFG concept, as described in this study, may be limited by scale considerations, at least with comparison to model application.

As observed for parent basins, CSI degrades markedly when using a constant depth-duration precipitation threshold.

Interior points: model parameter and soil moisture status transposition

Use of soil moisture status transposition from parent basins (added to use of parameter transposition) produces a remarkable deterioration of system performance with respect to parameter transposition, as can be noted in the CSI plots. This is due to the biased character of these estimates, which are obtained by using estimates of precipitation and temperature from the parent basins.

Even in this case, it is observed that CSI degrades markedly when using a constant depth-duration precipitation threshold. This suggests that even a poor estimate of temporal variability of soil moisture, as the one derived from the parent basins, may improve markedly above the condition of no-information on antecedent soil moisture status.

Overall score statistics

Analysis of results reported in the previous sections show that score values obtained for the high flooding threshold exhibit always more dispersion than corresponding scores obtained for the low threshold. This is clearly an effect of the sampling problem which arises in these computations. Use of a high flooding threshold generally results in the analysis of a small sample of events which may prevent reliable characterization of the system performance. Lowering the threshold incorporates more events into the analysis, hence mitigating the sampling problem. However, this choice exposes one to the risk of including small flood events which are not representative of the dynamics under study. To increase reliability of high threshold statistics, we analyse overall score statistics computed on all the considered basins. Overall score statistics are computed based on one overall contingency table generated from all basins considered in the study. The overall score statistics are reported in Figure 4-6a-c and Figure 4-7a-c, for POD/FAR and CSI, respectively.

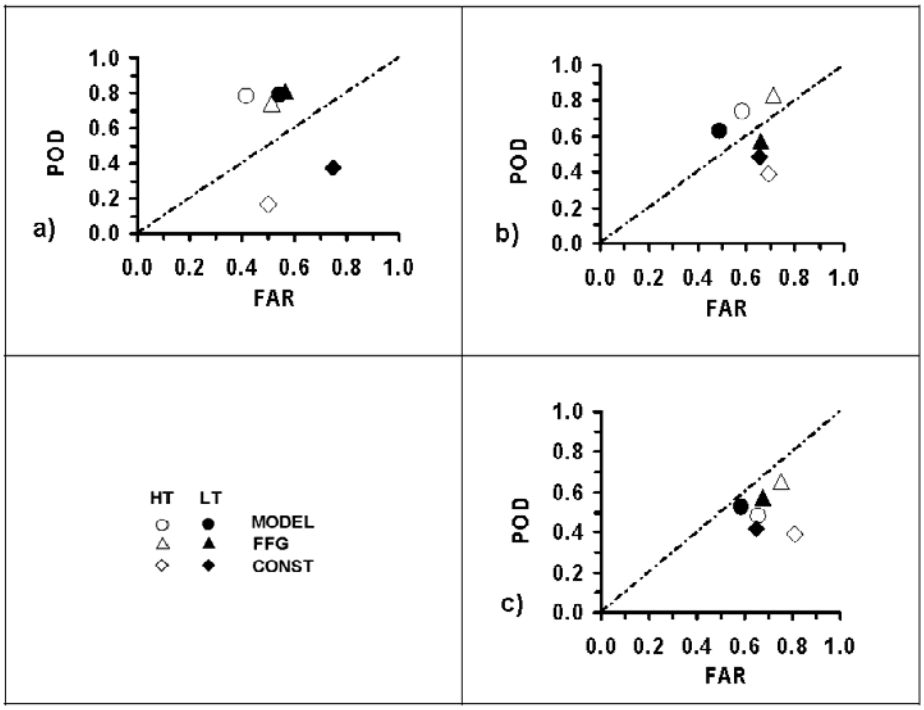


Figure 4-6a-c: Overall assessment of the threshold-based technique for: a) parent basin with model calibration; b) interior basins with model parameter transposition; c) interior basins with both model parameter and soil moisture status transposition. POD and FAR scores are plotted.

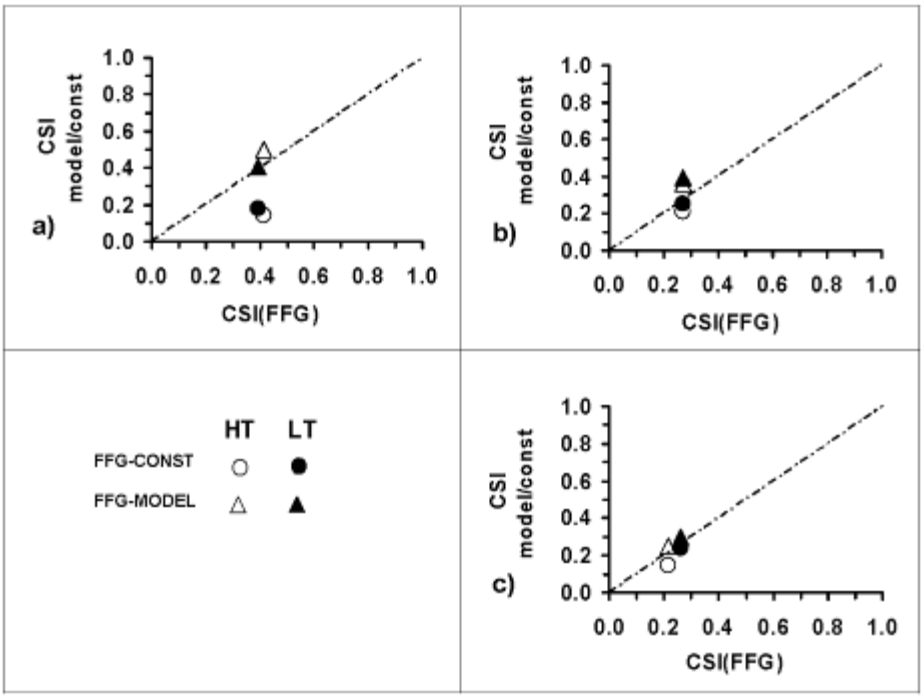


Figure 4-7a-c: Overall assessment of the threshold-based technique for: a) parent basin with model calibration; b) interior basins with model parameter transposition; c) interior basins with both model parameter and soil moisture status transposition. CSI scores are plotted.

For parent basins and the high threshold, FFG is characterised by a POD of 0.76 with FAR of 0.48. In this case, CSI is equal to 0.43. For interior basins and high threshold, with parameters transposed by parent basins, POD increases to 0.85, but at the expenses of increasing FAR to 0.68. The overall CSI is equal to 0.28, in this case. This shows that the deterioration of performances following application of FFG to ungauged basins (with parameter transposition) is not negligible, and amounts to 35% (assuming that overall system performances on parent basins is comparable to the one that would have been attained on interior basins, in the case of model parameter calibration).

For interior basins, with parameters and soil moisture status transposed by parent basins, POD reduces to 0.64, while FAR increases to 0.73. The overall CSI is equal to 0.22 in this case and shows a decrease of 21% with respect to the case of parameter transposition.

Differences between FFG and direct model application are rather modest, and decrease with decreasing the accuracy of model application. The percent difference amounts to 18% for the parent basins, to 15% for interior basins with parameter transposition, and to 12% for interior basins with parameter and soil moisture status transposition. This is not unexpected, showing that the impact on system performances due to the use of time-uniform precipitations reduces when other sources of uncertainties, related to lack of calibration and biases in the soil moisture estimations, become more significant.

Performance differences between FFG and use of constant depth-duration precipitation threshold are very high for the parent basins and decrease with decreasing the model accuracy. The percent difference amounts to 53% for the parent basins, to 25% for interior basins with parameter transposition, and to 19% for interior basins with parameter and soil moisture status transposition.

It is noted that differences between results obtained for the high and the low threshold are relatively low for the case of parent basins and for interior basins with parameter transpositions. These differences become comparable to those which arise among the various procedures for the case of model parameter and soil moisture status transposition.

4.6 Conclusions and future works

A threshold-based flash flood warning approach has been developed and tested on a wide range of climatic and physiographic European conditions, and by focusing on ungauged basins. The system is derived from the Flash Flood Guidance approach. The FFG is the depth of rain of a given duration, taken as uniform in space and time on a certain basin, necessary to cause minor flooding at the outlet of the considered basin. This rainfall depth, which is computed based on a lumped hydrological model, is compared to either real time-observed or forecasted rainfall of the same duration and on the same basin. If the nowcasted or forecasted rainfall depth is greater than the FFG, then flooding in the basin is considered likely.

The study provides an assessment of this technique based on operational quality data from eleven basins (six nested included in five larger parent basins) located in two European regions: north-eastern Italy and central France. The model used in this study is a semi-distributed conceptual rainfall-runoff model, following the structure of the PDM (Probability Distributed Moisture) model. System performances are evaluated by means of categorical statistics, such as the Critical Success Index (CSI).

Four questions posed in the Introduction were investigated to help understand the potential limitations and benefits of the FFG approach and guide further research and development. The experiments yielded the following answers:

1. Comparison of score statistics obtained by using the Flash Flood Guidance with those resulting from direct model application shows a slight degradation of system performance. Differences between FFG and direct model application are rather modest, and decrease with decreasing the accuracy of model application. The percent difference amounts to 18% for the parent basins, to 15% for interior basins with parameter transposition, and to 12% for interior basins with parameter and soil moisture status transposition. This is not unexpected, showing that the impact on system performances due to the use of time-uniform precipitations reduces when other sources of uncertainties, related to lack of calibration and biases in the soil moisture estimations, become more significant.

2. Results show that overall CSI is equal to 0.43 for the parent basins, where the hydrological model has been calibrated. CSI reduces to 0.28 for the interior basins, when model parameters are transposed from parent basins. This shows that the deterioration of performances following application of FFG to ungauged basins (with parameter transposition) is not negligible, and amounts to 35% (assuming that overall system performances on parent basins is comparable to the one that would have been attained on interior basins, in the case of model parameter calibration).

3. For interior basins, with parameters and soil moisture status transposed by parent basins, CSI reduces to 0.22 and shows a decrease of 21% with respect to the case of parameter transposition.

4. Performance differences between FFG and use of constant depth-duration precipitation threshold are very high for the parent basins and decrease with decreasing the model accuracy. The percent difference amounts to 53% for the parent basins, to 25% for interior basins with parameter transposition, and to 19% for interior basins with parameter and soil moisture status transposition. This suggests that even a poor estimate of temporal variability of soil moisture, as the one derived from the parent basins, may improve markedly above the condition of no-information on antecedent soil moisture status.

These results show clearly that the performance of the FFG system hinges on the accurate representation of the initial soil moisture conditions. This shows that system improvements could

be expected by additional work on real-time updating of model status. A natural choice would be to adjust the basin soil moisture state by making use of runoff data in a real time mode. The rationale of this is that runoff is usually an excellent indicator of the basin soil moisture state. An updating method widely used is the Kalman Filter which consists of weighting measurements and simulation, the weight (or Kalman gain) being a function of the measurement error and the model error (Da Ros & Borga, 1997b). Additional work should focus on the value of real-time updating for ungauged basins (by transposing updated soil moisture status to interior points) and at various spatial scales.

Further work should also focus on use of simpler modelling concepts that makes more proper use of constant precipitation depth thresholds (Martina et al., 2006) and which incorporate knowledge about flash flood related damages. Along this line, a research effort is now underway to evaluate the effectiveness of using model-based constant precipitation depth thresholds (variable on a monthly basis) which are based on monthly averaged soil moisture status and which incorporate a Bayesian utility function minimization.

5. Flash flood warning based on Bayesian rainfall thresholds

5.1 Introduction

The aim of this section is to determine the FFG rainfall depth and compute FFG-values based on the minimisation of a Bayesian Loss Function of the discharge conditional upon the state of saturation of the catchment.

Rainfall thresholds are here defined as the cumulated volume of rainfall during a storm event which can generate a critical water stage (or discharge) at a specific river section. When the rainfall threshold value is exceeded, the likelihood that the critical river level (or discharge) will be reached is high and consequently it becomes appropriate to issue a flood alert; alternatively, no flood alert is going to be issued when the threshold level is not reached. In other words the rainfall thresholds must incorporate a "convenient" dependence between the cumulated rainfall volume during the storm duration and the possible consequences on the water level or discharge in a river section. The term "convenient" is here used according to the meaning of the decision theory under uncertainty conditions, namely the decision which corresponds to the minimum (or the maximum) expected value of a Bayesian cost utility function.

There are two possible approaches for the same methodology: (a) using the Monte-Carlo simulations or (b) using the Normal Quantile Transform. The main difference of the two is the requirements in terms of data, i.e. the time series of rainfall and discharge.

5.2 The Bayesian Rainfall Thresholds using Monte-Carlo (BRT-MC)

This approach was developed by Martina et al. (2006) and it can be referred to that paper for a more detailed description. In order to ease the description of the methodology, illustrated in Figure 5-1, two phases are here distinguished: (1) the rainfall thresholds estimation phase and (2) the operational utilization phase. The first phase includes all the procedures aimed at estimating the rainfall thresholds related to the risk of exceeding a critical water stage (or discharge) value at a river section. These procedures are executed just once for each river section of interest. The second phase includes all the operations to be carried out each time a significant storm is foreseen, in order to compare the precipitation volume forecasted by a meteorological model with the critical threshold value already determined as in phase 1.

5.2.1 Phase 1 - The rainfall thresholds estimation

Given the different initial soil moisture conditions, which can heavily modify the runoff generation in a catchment, it is necessary to clarify that it is not possible to determine a unique rainfall threshold for a given river section. It is well known that the water content into the soil strongly affects the basin hydrologic response to a given storm, with the consequence that a storm event considered irrelevant in a dry season can be extremely dangerous in a wet season when the extent of saturated areas may be large. This implies the necessity of determining several rainfall thresholds for different soil moisture conditions. A useful indicator for discriminating among soil moisture classes, the Antecedent Moisture Condition (AMC) leads to the following three classes of soil moisture AMC I (dry soil), AMC II (moderately saturated soil) and AMC III (wet soil). Since each AMC class will condition the magnitude of the rainfall threshold, three threshold values have to be determined. Phase 1 of the proposed methodology for deriving the rainfall thresholds follows the three steps.

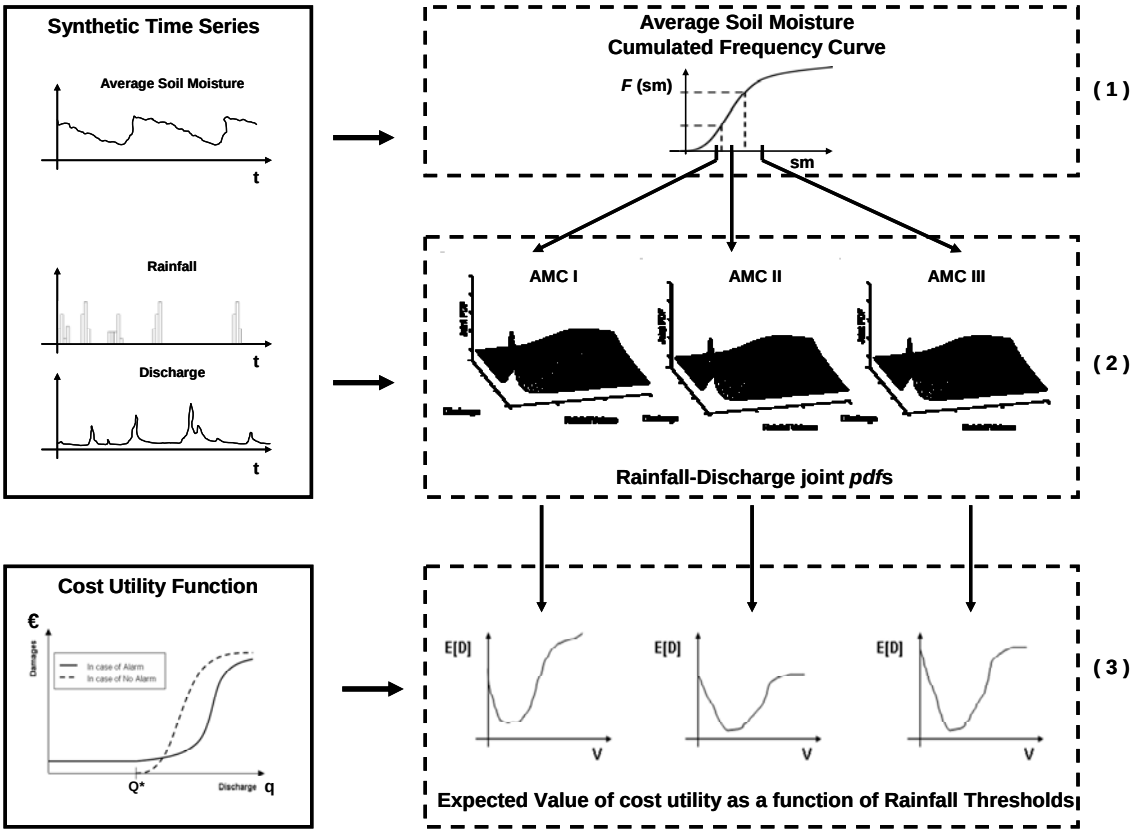


Figure 5-1: Schematic representation of the proposed methodology. (1) Subdivision of the three synthetic time series according to the soil moisture conditions (AMC); (2) Estimation of the joint pdfs between rainfall volume and water stage or discharge; (3) Estimation of the "convenient" rainfall threshold based on the minimisation of the expected value of the associated utility function

Step 1: Sorting the time series according to the AMC classes

In order to account for the different soil moisture initial conditions, it is necessary to divide the three synthetic records, namely the stochastically generated rainfall, the soil moisture conditions and the water levels (or discharges) obtained via simulation, in three subsets, each corresponding to a different AMC class.

This subdivision is performed on the basis of the AMC value relevant to the soil moisture condition preceding a storm event. According to this value, the corresponding rainfall and discharge time series will be grouped in the appropriate AMC class.
With reference to Figure 5-2, three time values are defined:

t_0 the storm starting time

T the rainfall accumulation time

T_C the catchment concentration time

The latter of which can be estimated from empirical relationships based on the basin geomorphology or from time series analysis, when long records are available. As it emerged from the sensitivity analysis of the proposed methodology, in reality there is no need of great accuracy in the determination of T_C.

On the basis of the above defined time values, the rainfall volume V_T (or rainfall height) cumulated from t_0 to $t_0 + T$ and the maximum discharge value Q (or the maximum water stage) falling in

the time interval from $t_0 + T$ to $t_0 + T_C$ are retained and grouped in one of the classes according to the AMC value at t_0 .

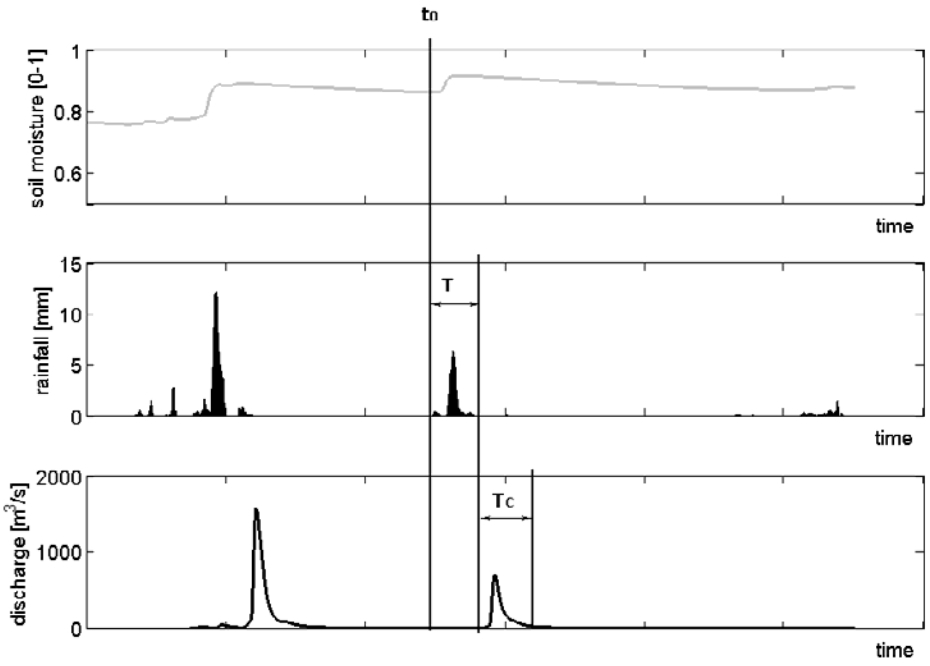


Figure 5-2: The synthetic time series and the three time values to be defined

Step 2: Fitting the joint probability density function

Once the corresponding couple of values (the rainfall total and the relevant maximum water stage or discharge) have been sorted into the three AMC classes, for each class one can use these values to determine the joint probability density functions (jpdf) between the rainfall total and the relevant maximum discharge (or the water stage), to be used in step three. .

Step 3: Estimation of the most convenient rainfall threshold

Following the Bayesian decision theory the concept of "convenience" is introduced as the minimum expected cost under uncertainty. The term "cost" does not refer to "actual costs" of flood damages that are probably impossible to be determined, but rather a Bayesian utility function describing the damage perception of the stakeholder, which may even include the non commensurable damages due to "missed alert".

The searched most "convenient" rainfall threshold value can thus be determined as the one that minimizes the expected utility cost, namely:

$$V_T^* = \underset{V_T}{Min} \left\langle E\left\{U\left(q,v|V_T,T\right)\right\}\right\rangle = \underset{V_T}{Min} \left\langle \int_0^{+\infty}\int_0^{+\infty} U\left(q,v|V_T,T\right) f\left(q,v|T\right) dq\, dv \right\rangle \tag{5.1}$$

where $U\left(q,v|V_T,T\right)$ is the utility cost function, $f\left(q,v|T\right)$ is the joint probability distribution function of the rainfall volume and the discharge peak value. One rainfall threshold value V_T^* will be derived for each accumulation time T. In Figure 5-4 one can see the typical shape of the expected value of the utility $E\left(U\left(q,v|T,T\right)\right)$ for a given accumulation time T. All the values of

the rainfall thresholds obtained for each AMC class, can be plotted as a function of the rainfall accumulation time (Figure 5.4).

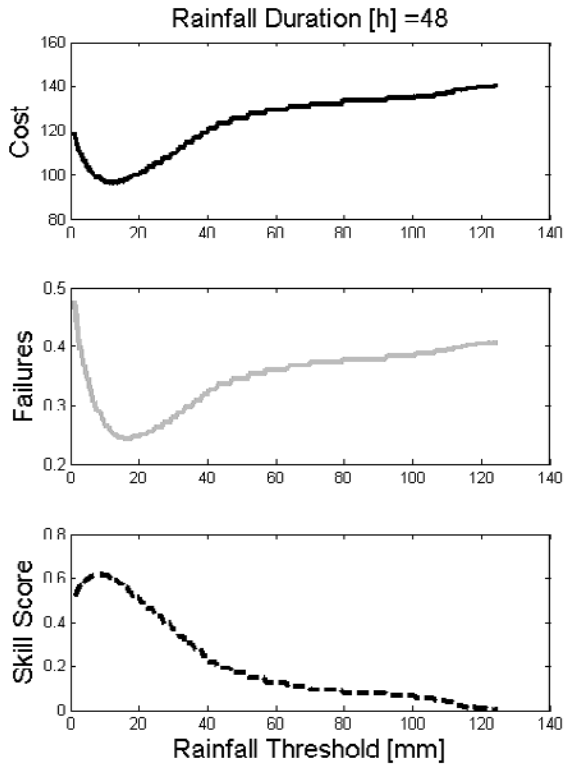


Figure 5-3: Expected value of the utility $E\left(U\left(q,v\middle|T,T\right)\right)$ *for a given accumulation time* T.

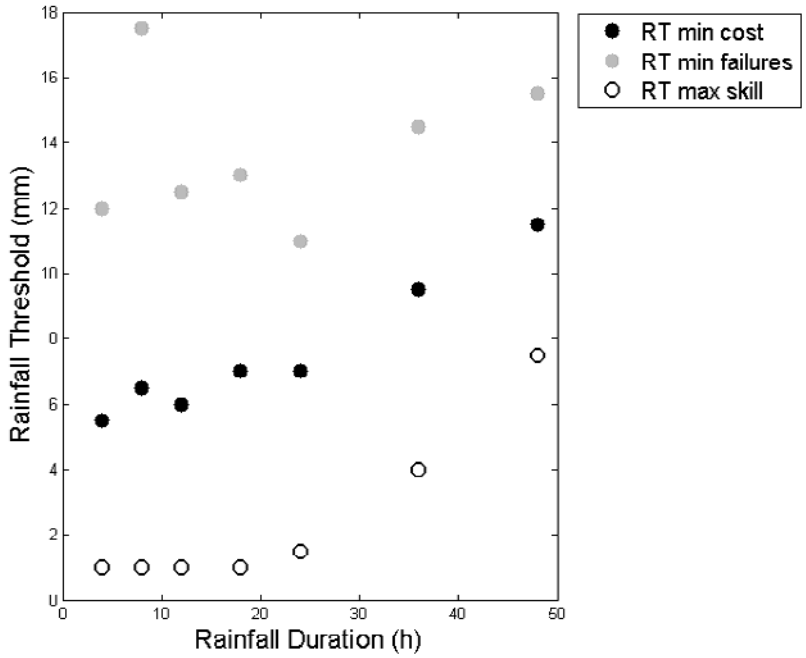


Figure 5-4: Rainfall thresholds plotted as a function of the rainfall accumulation time.

5.2.2 Phase 2 - Operational use of the rainfall threshold approach

A methodology, which makes use of the long synthetic time series, already obtained for the thresholds derivation, is here proposed, to be applied only once in phase 1. It is possible to determine on a monthly basis, the mean soil moisture as a function of the cumulated rainfall volume over the previous n days. More in detail, the mean soil moisture of the month vs the precipitation volume cumulated over the previous 72 hours. These results were obtained by using the 10,000 years synthetic rainfall and the corresponding simulated soil moisture series. These graphs will be referred to as the "AMC Calendar".

When a storm is forecasted, using the rainfall thresholds together with the AMC Calendar, it is possible to:

1. Determine the mean catchment soil moisture and the correct AMC class, entering into to the monthly graph with the cumulated rainfall volume recorded in the previous n hours.

2. Choose the rainfall threshold corresponding to the identified AMC class.

3. Add the forecasted cumulated rainfall to the observed rainfall volume.

4. Issue a flood alert if the identified threshold is overtopped.

5.3 The Bayesian Rainfall Thresholds using the Normal Quantile Transform (BRT-NQT)

One of the limits of the method described in Martina et al. (2006) is represented by the excessive data requirement. As a matter of fact, the minimization of the expected cost of flood warning upon which the rainfall threshold is computed needs an estimation of the joint probability density function between the rainfall volume at a fixed duration, V(T), and the peak of the discharge following a time interval equal to the concentration time of the catchment, Qp. The estimation of the empirical (i.e. based on the data) joint probability density function requires a large amount of data and therefore long rainfall and discharge time series (e.g. thousands of years). Since it is not possible to find such observed long time series, they are synthetically reproduced by a rainfall-runoff model calibrated on the observed discharge data and fed by a stochastic rainfall model calibrated on the observed rainfall data. This chain of models can then replicate rainfall and discharge time series as long as required. The joint probability density function is then estimated following a classical Monte Carlo approach. This methodology will be hereafter referred to as BRTMC (Bayesian Rainfall Threshold using a Monte Carlo approach).

To overcome the limit of the BRTMC methodology, a second methodology has been recently developed by UniBo hereafter referred to as BRTNQT (Bayesian Rainfall Threshold using the Normal Quantile Transform). The difference with BRTMC consists in the inference of the joint probability density. Instead of the classical Monte Carlo approach the inference of the joint PDF is performed by transforming the two variables, V(T) and Qp, into two standard normally distributed variables by means of the Normal Quantile Transform (NQT). This procedure ensures, by construction, that the marginal distribution of the variables are standard normal, but does not guarantee that the joint PDF is multivariate standard normal distribution. Therefore generally the normality of the joint distribution must be tested by comparing the empirical (based on the data) distribution with the theoretical form or existing goodness of fitting test. The NQT leads to the inference of the meta-Gaussian joint PDF which can be performed using a much smaller amount of data then the BRTMC (e.g. tens of years).

During the third year, the BRTNQT has been developed and implemented such that the only data requirement is the rainfall and discharge time series (for the joint PDF inference) and the average soil moisture time series which conditions the rainfall thresholds. The average soil moisture conditions, which often necessitates to be simulated by a rainfall-runoff model, can be substituted

with a reliable antecedent conditions index (such as API, AMC, etc..) computed based on the precipitation. A computer program which performs the BRTNQT methodology has been developed and implemented on the pilot basins.

Moreover, UniBo developed a procedure for the evaluation of the different FFG methods. This procedure has three options: the evaluation is performed on the basis of (i) the expected cost of the flood warning system defined by a utility cost function or (ii) the score of the contingency table or (iii) some forecast skill scores such as the hit rate and the false-alarm rate (Figure 5-5).
In summary the BRT-NQT methodology:

- It computes the "optima" rainfall thresholds according to a Bayesian cost function (BCF).
- It is based on the convolution of the meta-Gaussian joint probability distribution function (mG-jPDF) between rainfall storm volume and peak flow with the BCF.
- It requires the time series of rainfall, discharge and average soil moisture (or another antecedent soil coindition index) of moderate length (e.g. 10 years).
- It does not require a rainfall stochastic model and may not require a RR model.

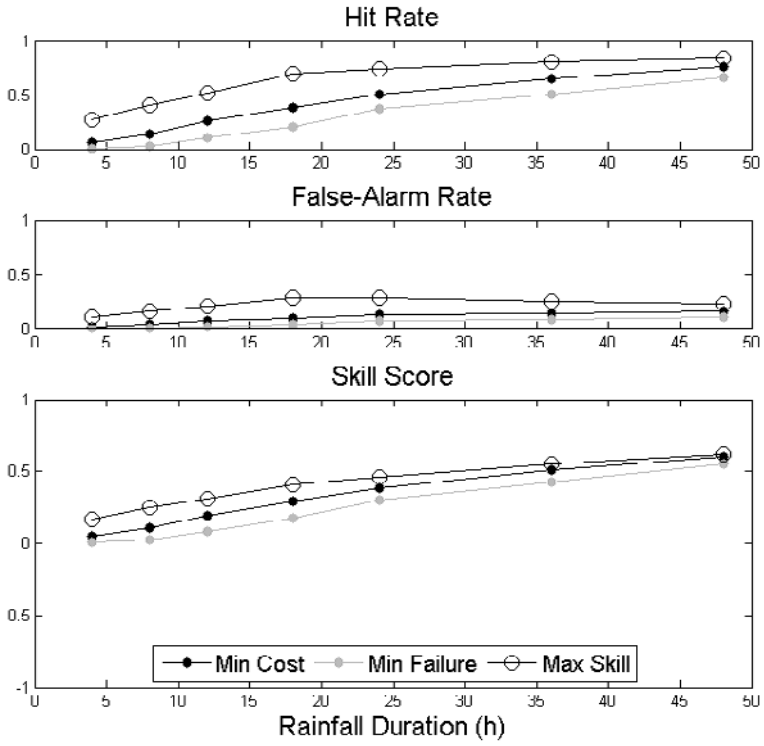


Figure 5-5: Evaluation of the different FFG methods.

5.4 Application on the Posina Catchment of the BRT-MC

The Posina catchment (116 km^2) is located in nort-eastern Italy. Mean annual precipitation is around 1600-1800 mm. Altitude ranges from 2230 m a.s.l. to 390 m a.s.l. at the outlet. The topography is rugged with an average slope of 28°. Forests cover 68% of the basin. The remaining

area is devoted to agriculture with mixed farming and pasture. Ten years of precipitation/temperature/discharge data are available at hourly time accumulation for this catchment.

In order to test the capability of the methodology at reproducing the criteria of the decision maker by means of the Cost/Utility functions, it has be designed an experiment which reproduces different "attitudes" or different decision criteria for the alarm management. This could be done by selecting appropriate parameters of the Utility/Cost function. The Utility/Cost function has be defined as (Martina et al., 2006):

$$U\left(q,v\middle|V_T,T\right)=\frac{a}{1+be^{-c\left(q-Q^*\right)}} \quad \text{when} \quad v\le V_T \text{ and no alert is issued} \tag{5.2a}$$

$$U\left(q,v\middle|V_T,T\right)=D+\frac{a'}{1+b'e^{-c'\left(q-Q^*\right)}} \quad \text{when} \quad v\le V_T \text{ and no alert is issued} \tag{5.2b}$$

with T the time of rainfall depth accumulation, v the forecasted volume and V_T the rainfall threshold value, while a, b, c and a', b', c' are appropriate parameters. Due to the fact that the utility functions are only functional to the final objective of providing the decision makers with tools reflecting their risk perception, the shape of such functions, as well as the relevant parameter values, can be jointly assessed, by analysing the relevant effects on the decision process over past events.

$U\left(q,v\middle|V_T,T\right)$ is the utility cost function, which if $v\le V_T$ expresses the perception of damages when no alert is issued no costs will occur if the discharge q will remain smaller than a critical value Q^*, while damage costs will grow noticeably if the critical value is overtopped. On the contrary, if $v>V_T$ it expresses the perception of damages when the alert is issued a cost which will be inevitably paid to issue the alert (evacuation costs, operational cost including personnel, machinery etc.), and damage costs growing less significantly when the critical value Q^* is overtopped and the flood occurs. The utility function to be used will differ depending on the value of the cumulated rainfall forecast v and the rainfall threshold V_T. If the forecast precipitation value is smaller or equal to the threshold value, the alert will not be issued; on the contrary, if the forecasted precipitation value is greater than the threshold value, an alarm will be issued.

Three different attitudes of the decision maker have been selected which generated three different cases for the application of the utility/cost function: (a) the "risk-averse" case, (b) the "risk-prone" case, (c) the "real" case. The first two cases have been chosen in order to represent some sort of extreme attitudes of the decision makers, while the third case has been designed on the basis of real experience. It is important to say that, especially for the first two cases, the utility/cost function represents more the "perception" of the costs and of the damages rather than the real costs. This means that not only the real costs are important but also the weights which each decision maker attribute to them. The reasons for that are: (1) there are some costs which are not valuable as the credibility loss; (2) there is a different perception of the costs in terms of social and psychological impacts than only the economic one.

The definition of the cases is synthesized in the following Table 5-1 where are reported the parameters of the equations (5.2) and (5.3) which together define the utility/cost function.

Table 5-1: Parameters of the cost functions in the different cases

Case	Alarm			Not Alarm			
	a (€)	b	c	a' (€)	b'	c'	D (€)
a) The "risk-averse" case	$10*10^6$	2	0.02	$3*10^6$	60	0.03	$7*10^3$
b) The "risk-prone" case	$10*10^6$	500	0.023	$7*10^6$	1000	0.027	$15*10^3$
c) The "real" case	$10*10^6$	200	0.025	$5*10^6$	800	0.03	$10*10^3$

In order to compare the described cases with some references, have been defined also two criteria independent by the cost function.

The first is the "Maximum Skill Score" criteria. This criteria is based on the Skill Score coefficient which is defined by the contingency table. Contingency tables are highly flexible methods that can be used to estimate the quality of a deterministic forecast system (Mason and Graham, 1999) and, in their simplest form, indicate its ability to anticipate correctly the occurrence or non occurrence of predefined events. A warning *W* is defined as the forecast of the occurrence of an event *E* (in this case the overtopping of a threshold). A two-by-two contingency table can be constructed as illustrated in Table 5-1.

Table 5-2:Two-by-two contingency table for the assessment of a threshold based forecasting system

		Forecasts	
Observations	*Warning* *W*	*No Warning* *W'*	*Total*
Event, E	*h*	*m*	*e*
Non Event, E'	*f*	*c*	*e'*
Total	*w*	*w'*	*n*

From a total number of *n* observations, one can distinguish the total number of event occurrences (*e*) and that of non-occurrences (*e'*); the total number of warnings is denoted as *w*, and that of no-warnings as *w'*. The following outcomes are possible: a hit, if an event occurred and a warning was issued (with *h* the total number of hits); a false alarm, if an event did not occur but a warning was issued (with *f* the total number of false alarms); a miss, if an event occurred but warning was not issued (with *m* the total number of misses); a correct rejection, if an event did not occur and a warning was not issued (with *c* the total number if correct rejections).

The skill of a forecasting system can be represented on the basis of the hit rate and the false-alarm rate. Both ratios can be easily evaluated from the contingency table (Mason, 1982):

$$\begin{cases} \text{hit rate} & = \dfrac{h}{h+m} = \dfrac{h}{e} \\ \text{false - alarm rate} & = \dfrac{f}{f+c} = \dfrac{f}{e'} \end{cases} \tag{5.3}$$

The hit and false-alarm rates (Eq. 5.3), indicate respectively the proportion of events for which a warning was provided correctly, and the proportion of non events for which a warning was provided incorrectly. The hit rate is sometimes known as the probability of detection and provides an estimate of the probability that an event will be forewarned, while the false-alarm rate provides an estimate of the probability that a warning will be incorrectly issued (Eq. 5.4).

$$\begin{cases} \text{hit rate} & = p(W|E) \\ \text{false - alarm rate} & = p(W|E') \end{cases} \tag{5.4}$$

For a system that has no skill, warnings and events are by definition independent occurrences, therefore, the probability of issuing an alert does not depend upon the occurrence or non occurrence of the event, namely:

$$p(W|E) = p(W|E') = p(W) \tag{5.5}$$

This equality occurs when warnings are issued at random. When the forecast system has some skill, the hit rate exceeds the false-alarm rate; a bad performance is indicated by false-alarm rate exceeding the hit rate. Because of the equality of the hit and false-alarm rates for all forecasts strategies with no skill, the difference between the two rate indexes can be considered an equitable skill score *ss* (Gandin & Murphy, 1992).

$$ss = p(W|E) - p(W|E') \tag{5.6}$$

The second criteria is the "minimum failure" criteria which consists in a simplification of the previous one. Instead of looking at the overall skill score of the system, in this case it is minimized the number of the failures as the sum of the "missed alarms" (*m*) and of the "false alarm" (*f*).

The choice of these criteria is justified by the need to have to different approaches to the problem of the rainfall thresholds. The first is the approach which weights in the same way the failures and the success of the system with no regard to the impact than a failure can have in function of the magnitude of the event. The latter instead is the approach does not consider the benefits of the alarm in terms of reduction of costs. These two approaches should determine very different rainfall thresholds which should always be respectively higher and lower of the cases, under examination, which make use of the utility function.

5.4.1 The "risk- averse" case

This case was designed in order to simulate the behaviour of decision-makers very reluctant to the consequences, both economic and social, of a flood. This means that rather than to risk a missed alarm they can accept many cases of false alarm with the associated costs.

False Alarm costs.

There are practically not false alarm costs. As a matter of fact this represents the perception of the costs, since in reality they will of course exist, but they are accepted or considered unavoidable.

Missed Alarm costs (flood damages in case of missed alarm)

These costs are very large compared to the others. They can be or not be link to the magnitude of the flood. In the latter case they grow very rapidly with the discharge starting from low values of discharge over the critical thresholds and they reach relatively soon the maximum probable damage.

Correct Alarm costs (flood damages in case of alarm)

The damages associated with a flood correctly forecasted although not negligible, are relatively low compared to those of the previous case. This is due to the perception of the effects of the protection

measures (e.g. the evacuation) on the flood damages. In this case the benefits of the alarm are overemphasized in order to reproduce the behaviour of the decision-maker.

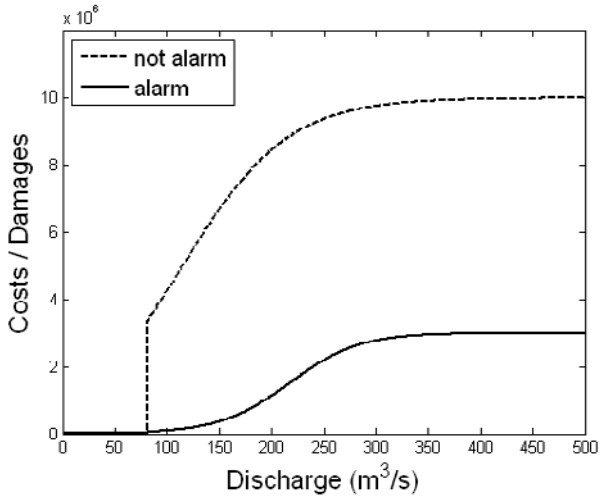


Figure 5-6: Utility/cost function for the "risk-averse" case

With these requirements the utility function appears as shown in the Figure 5-6 above. Particularly it can be noted as the "not alarm" costs increase rapidly in the first part of the curve and the difference between the "not alarm" and "alarm" maximum losses is high.

With that utility function the determination of the rainfall threshold is as it is shown in the graph below. As the Figure 5-7 shows, the rainfall thresholds determined in this case are very close to the "max skill" criteria. The alarm is issued very frequent given that the cost for missed alarm is very low and the reluctance for the risk of a flood due to the high cost even for low magnitude events.

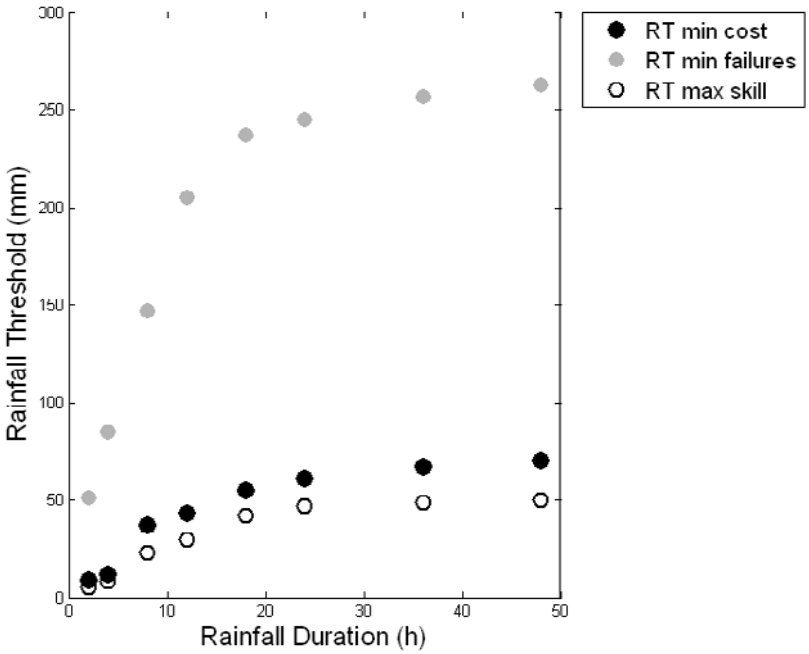


Figure 5-7: the rainfall threshold for the "risk-averse" case

5.4.2 The "risk- prone" case

This case was designed in order to simulate the behaviour of decision-makers who can accept to some level the costs due to a flood rather than those due to a system failure such as a false alarm. This could be also the case of areas of not economically valuable to the proximity of the river, while areas of important social and economic value are remote but, with extremely high events, can be affected by a flood.

False Alarm costs.

The costs are considered very high in order to reproduce the reluctance of the decision maker to this case. This could be justify not only by the direct costs of a false alarm such as an evacuation, but also by the imagine damage or loss of credibility for the decision maker.

Missed Alarm costs (flood damages in case of missed alarm)

These costs are large but not compared to the other cases. They are very low for low discharge value in order to reproduce the attitude of the decision makers at accepting the low severity – high frequency damages, while the costs increases for more severe floods.

Correct Alarm costs (flood damages in case of alarm)

The damages associated with a flood correctly forecasted are sensibly lower than those in the missed alarm costs, but not as low as those in the "risk-averse" case. This is to simulate the perception of the decision maker who does not emphasize the benefits of the protection measures.

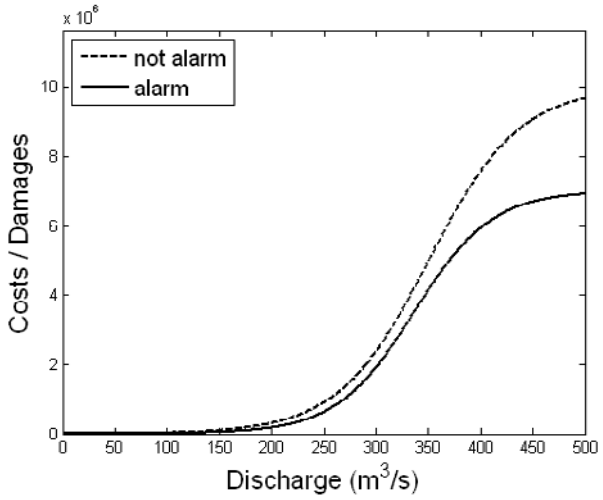


Figure 5-8: Utility/cost function for the "risk-prone" case

With these requirements the utility function appears as shown in the Figure 5-8 above. Particularly it can be noted that the differences between the "alarm" and the "not alarm" costs is not high due to the perception of the benefits of the alarm especially for event of moderate magnitude. Also it can be noted as the function increases slower than that in the previous case in order to simulate the attitude of the decision maker to "accept" the high-frequent low-magnitude risks.

With that utility function the determination of the rainfall threshold is as it is shown in the graph below. As the Figure 5-9 shows, the rainfall thresholds determined in this case are very close to the "min costs" criteria. The alarm is not issued very frequent given that the costs for missed alarm are high compared to that for low-magnitude events. Thus while it is very frequent to have an event of

low magnitude, it is very rare to issue an alarm. The alarm is issued only with events of high magnitude and therefore of low frequency.

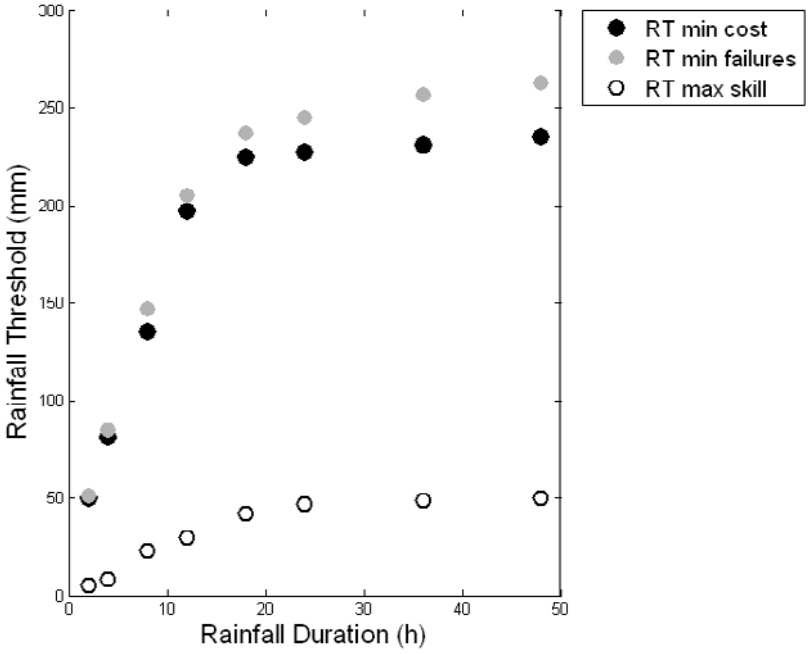


Figure 5-9:The rainfall threshold for the "risk-prone" case

5.4.3 The "real" case

This case was designed on the basis of real experience collected from decision-makers and operators of the civil protection involved in the management of the alarms and evacuation procedure due to a general natural hazard which could comprises flood, landslides and debris flow.

False Alarm costs.

From some experience, especially those of the last evacuation in the 1999, the evacuated areas are first of all those exposed to debris-flow risk in the upper part of the catchment close to Altiero and other small villages of the rest of the catchment areas. Generally these costs can be estimated as the costs associated with the activity of 50 people for a workday and of the connected machineries.

Missed Alarm costs (flood damages in case of missed alarm)

These costs grow very rapidly with the flow values greater than the discharge with 50 years of return period reaching the maximum around a discharge of 100 years, while for values lower than the 50 years discharge the costs are not important. This is due to the infrastructures built in the last years along the main reach of the Posina river, since here the river crosses an area extremely vulnerable at landslides.

Correct Alarm costs (flood damages in case of alarm)

The damages associated with a flood correctly forecasted, and thus in case of alarm, are similar as trend to the previous case but it can be estimated a rough reduction of 50% of the damages. It is anyway very difficult to make this estimation since the damages due to the flood can not be distinguished from those due to the landslides and debris flows occurring during the same event.

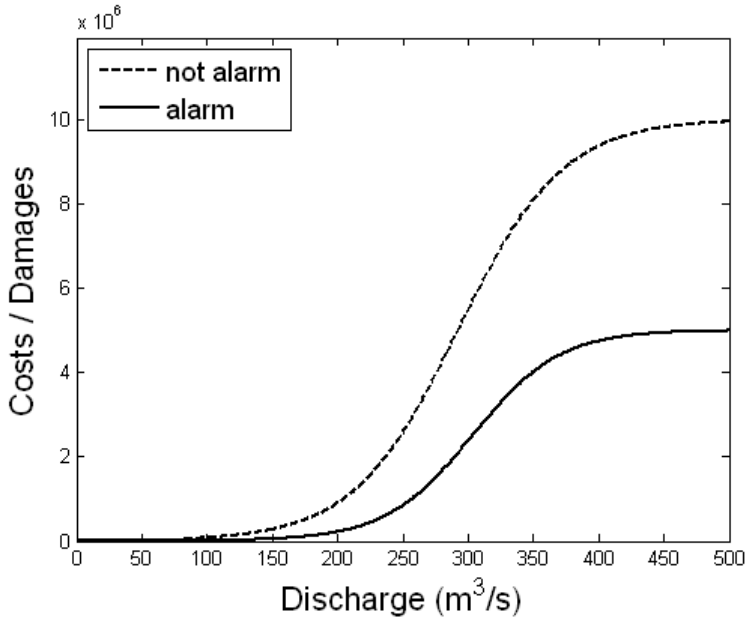


Figure 5-10: Utility/cost function for the "real" case

With these requirements the utility function appears as shown in the Figure 5-10 above. Particularly it can be noted that the differences between the "alarm" and the "not alarm" costs is higher than the "risk-prone" case, but not as high as the "risk-averse" case. This is due to the perception of the benefits of the alarm especially for event of moderate magnitude. Also it can be noted as the function increases slow in the lower part for the impacts of the hydraulic infrastructures. Also the curve in the "not alarm" try to simulate the attitude of the decision maker to "accept" the high-frequent low-magnitude risks.

With that utility function the determination of the rainfall threshold is as it is shown in the graph below. As the Figure 5-11 shows, the rainfall thresholds determined in this case are closer to the "min costs" criteria than to the "max skill" criteria. The alarm is not issued very frequent given that the costs for missed alarm are high compared to that for low-magnitude events. Thus while it is very frequent to have an event of low magnitude, it is very rare to issue an alarm. The alarm is issued only with events of high magnitude and therefore of low frequency.

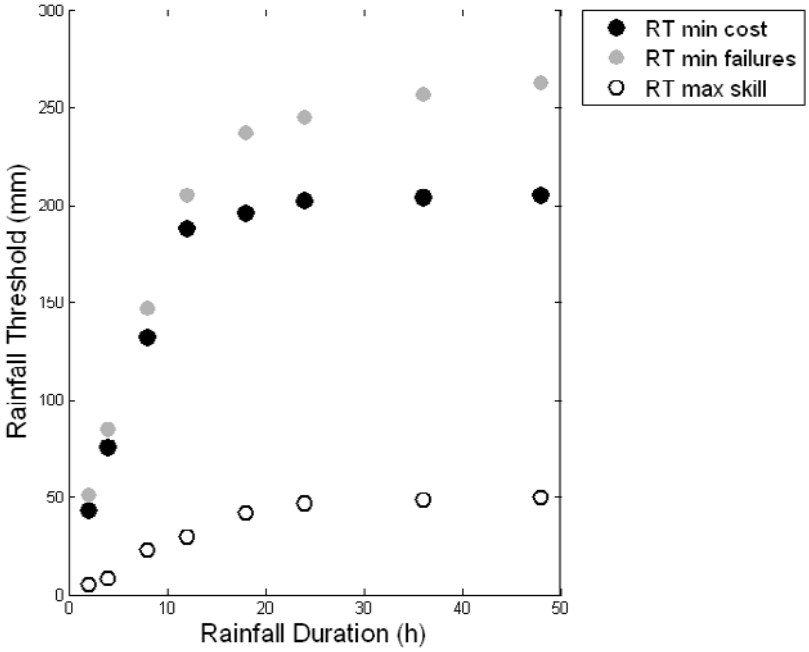


Figure 5-11: The rainfall threshold for the "real" case

5.5 Considerations

From the comparison of the three cases the following observations can be drawn.

1. The methodology is very sensitive to the definition of the utility/cost function as it can be seen by the previous figures. As a matter of fact, the Figure 5-12 reports all the three cases. Examination of the figure shows that - even using the same joint probability distribution function between the peak of the discharge and the rainfall volume and the same definition of critical discharge - the rainfall thresholds can differ based on the different utility functions. This is a positive feature of the method and indicates that the rainfall threshold can represent effectively the different decision criteria adopted.

2. The "real" case under examination (Posina) is much closer to the Risk-prone case in terms of the resulting rainfall threshold. However this is not a general rule. Indeed this heavily depends on the chosen parameters and on the features of the catchment. Nonetheless if the utility/cost function represent the decision criteria, the conclusion that can be derived are very interesting.

3. The use of the rainfall threshold as a direct alarm system could be not reliable. Although the simulations where purposely design to highlight the features of the system, a critical view of the methodology from an operational perspective can raise some limits. Indeed operationally in case of a critical event there is not a unique level of protection as it has been used here. For instance if a critical event is predicted 48 hours in advance, the reaction could be only a more accurate monitoring of the river flow/stage and only after the confirmation of the critical conditions the emergency procedure could be activated and the alarm issued. This limit is reflected in the adoption of a unique rainfall threshold to decide whether do nothing ("not alarm") or do everything is planed for an emergency ("alarm") without any respect of the magnitude of the predicted event.

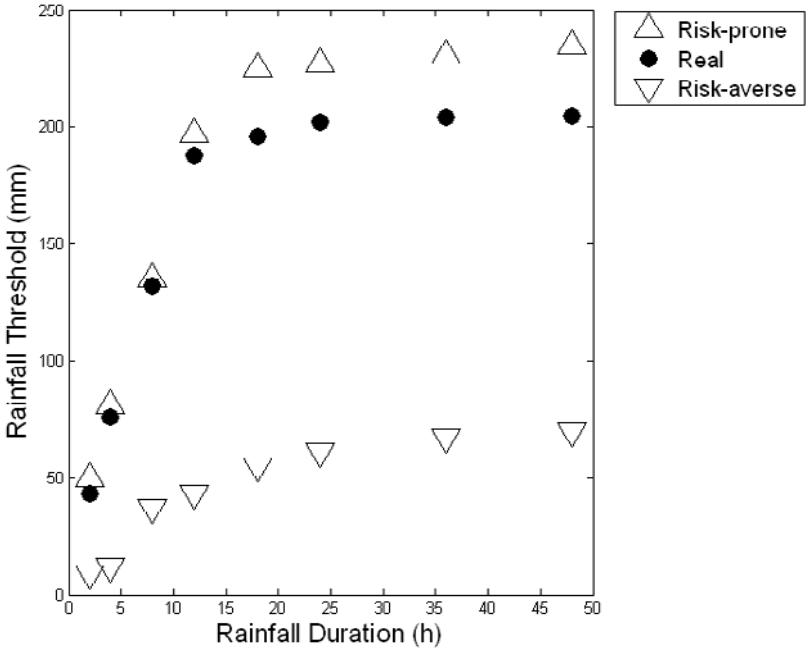


Figure 5-12: Comparison of the rainfall thresholds for the examined cases

6. Conclusions and recommendations

Conclusions are reported below with reference to the main Sections of the report.

General capabilities of semi-distributed, conceptual rainfall-runoff models in the case of flash floods

Overall, the performance of the forecasting models is relatively modest. Only 20 to 30% of the discharge trend variance is explained by the models even after a large calibration and validation effort. This relatively disappointing result may be due to the limits of the models but also to the various sources of uncertainties and particularly the uncertainties on the actual rainfall amounts.

The impact of the rainfall estimation uncertainties was analysed in a second part of the study. This analysis was carried out by describing the sampling uncertainty in rainfall estimation by means of a stochastic rainfall model. This study has shown that, for basin up to 500 km^2, a large part of the rainfall-runoff simulation errors is explained by rainfall estimation uncertainties. For larger basins the shape of the hydrograph can be influenced by the spatio-temporal pattern of the rainfall event and distributed RR models may bring a slight improvement if compared to the tested lumped models.

From a practical point of view, this shows that rainfall estimation uncertainties limit drastically the possible accuracy of RR simulations. Operational forecasting services should be aware of this limit to efficiently use the RR models and if possible evaluate these uncertainties in real time to be able to deliver confidence intervals along with their traditional deterministic forecasts. Ensemble or Monte Carlo forecasts are now used routinely in meteorological forecasting; there is no reason why they should be disregarded by hydrologists. The error scenario simulation model developed in this chapter could help to build such ensemble forecasts in the case where mean areal rainfall amounts are estimated through a rain gauge network. The same type of model is still to be developed for the case where quantitative radar estimations are used.

Assessment of a statistical-distribute model for flash flood forecasting and warning

Results of the study show that by looking at relative differences and model consistent thresholds early warning for flashfloods can be given with leadt imes exceeding 24 hours. In both the analysed case studies the weather forecasts captured the event well allowing to predict peak, magnitude and spatio-termporal distribution of the event well with an absolute leadtime of more than 36 hours. Taking into account computing, processing and analyses time the effective leadtime could still have been of the order of 24 hours which would be sufficient to allow authorities to take precautionary measures and have more time to act once monitoring devices confirm the event.

The results also show, however, that the principle can also be useful for those areas where such data do not exist and where the approach could greatly contribute to the preparedness for flashflood events in terms of awareness, identification of regions at risk, potential magnitude and timing of the event.

The results of this study suggest that we should be able to improve hydrologic guidance provided to forecasters in flash flood situations by running an available distributed hydrologic model with available parameter estimation techniques. The successes of both an uncalibrated distributed model and a distributed model with minimal, simplistic calibration (only at parent basin outlets) suggest that the models reasonably represent interior basin processes. We anticipate that future enhancements to distributed model calibration techniques and data assimilation will improve results even further.

Assessment of the Flash Flood Guidance concept

The assessment was carried out on a large database including catchments from Italy and France. The study yielded the following indications:

1. Comparison of score statistics obtained by using the Flash Flood Guidance with those resulting from direct model application shows a slight degradation of system performance. Differences between FFG and direct model application are rather modest, and decrease with decreasing the accuracy of model application. The percent difference amounts to 18% for the parent basins, to 15% for interior basins with parameter transposition, and to 12% for interior basins with parameter and soil moisture status transposition. This is not unexpected, showing that the impact on system performances due to the use of time-uniform precipitations reduces when other sources of uncertainties, related to lack of calibration and biases in the soil moisture estimations, become more significant.
2. Results show that overall CSI is equal to 0.43 for the parent basins, where the hydrological model has been calibrated. CSI reduces to 0.28 for the interior basins, when model parameters are transposed from parent basins. This shows that the deterioration of performances following application of FFG to ungauged basins (with parameter transposition) is not negligible, and amounts to 35% (assuming that overall system performances on parent basins is comparable to the one that would have been attained on interior basins, in the case of model parameter calibration).
3. For interior basins, with parameters and soil moisture status transposed by parent basins, CSI reduces to 0.22 and shows a decrease of 21% with respect to the case of parameter transposition.
4. Performance differences between FFG and use of constant depth-duration precipitation threshold are very high for the parent basins and decrease with decreasing the model accuracy. The percent difference amounts to 53% for the parent basins, to 25% for interior basins with parameter transposition, and to 19% for interior basins with parameter and soil moisture status transposition. This suggests that even a relatively poor estimate of temporal variability of soil moisture, as the one derived from the parent basins, may improve markedly above the condition of no-information on antecedent soil moisture status.

Results show that overall CSI is equal to 0.43 for the parent basins, where the hydrological model has been calibrated. CSI reduces to 0.28 for the interior basins, when model parameters are transposed from parent basins, and to 0.21, when both model parameters and soil moisture status is transposed from parent basins. Performance differences between FFG and use of time-constant soil moisture status are very high for the parent basins and decrease with decreasing the system accuracy. The percent difference amounts to 53% for the parent basins, to 25% for interior basins with parameter transposition, and to 19% for interior basins with parameter and soil moisture status transposition.

These results improve our understanding of the applicability and reliability of this technique at various scales and under various scenarios of data availability.

Remaining gaps in knowledge

Based on the conclusions reported above, a number of interesting topics for further research can be identified:

- Verification is required to measure current performance and to establish a baseline to be used to quantify the effectiveness of future enhancements. Demonstrated improvements in flash flood forecasting will lead to improved flash flood risk management. However, as flash floods are locally rare event and are difficult to capture with existing conventional

hydrometeorological networks, it is difficult to establish consistent verification of flash flood forecasting and warning accuracy with existing data archives. Efforts should be directed to i) establish protocols for flash flood observations, and ii) develop archives of flash flood events across Europe.

- Verification of Flash Flood Guidance has been done here with reference to perfectly known rainfall scenarios. Including uncertainty in rainfall forecast is required to produce an overall evaluation of this kind of approach.

- Increasing the accuracy of rainfall estimation at the flash flood event scales is critical to improve the quality of the forecast. However, a significant effort should be focused on developing better methods for defining soil moisture conditions at the onset of the event. Retrieval of soil moisture status by means of satellite observations may represent a promising data source for this purpose, at least for some European hydro-climatic environments.

- Further work should focus on developing threshold-based approaches that incorporate use of constant precipitation depth thresholds together with knowledge about flash flood related damages in a Bayesian utility function minimization. This may have significant impact on effective incorporation of stakeholders risk perceptions in defining warning thresholds.

7. References

1. ABRAHART R J, HEPPENSTALL A J, SEE L M (2007), *Timing error correction procedure applied to neural network rainfall-runoff modeling.* Hydrological sciences journal, 52, pp414-431.

2. ANCTIL F, PERRIN C, ANDRÉASSIAN V (2004), *Impact of the length of observed records on the performance of ANN and of conceptual parsimonious rainfall-runoff forecasting models*. Environmental Modelling and Software, 19(4), pp357-368.

3. BAIN, V. et AL., (2008). Methods of Post Flood Field Investigation: Lessons learned from trial on Selščica Sora River, Slovenia. *Geophisical Research Abstracts. EGU 2008.*

4. BERGSTRÖM S (1995), *The HBV model,* in Computer Models of Watershed Hydrology, V. Singh (ed.), Water Resources Publications, pp443-476.

5. BEVEN K, KIRBY M (1979), *A physically-based, variable contributing area model of basin hydrology* ,Hydrol. Sci. Bull., 23, pp419-438.

6. BOUDEVILLAIN B, DELRIEU G, CHAPON B, KIRSTETTER P E , NICOL J, ANDRIEU H (2006), *The Bollène-2002 experiment: innovative algorithms and evaluation of processing strategies for radar QPE in the Cévennes – Vivarais region,* Proc. 4th Europ. Conf. on Radar in Met.& Hydr, Barcelona, pp157-160.

7. BLOESCHL G, SIVAPALAN M (1995), *Scale issues in hydrological modelling – a review*. Hydrol. Proc., 9, pp251-290.

8. BLOESCHL G (2005), *Rainfall-runoff modelling of ungauged basins.* Article 133 in: Encyclopedia of Hydrological Sciences, M. G. Anderson (Managing Editor), J. Wiley & Sons, Chichester, pp. 2061-2080.

9. BLÖSCHL G, RESZLER C, KOMMA J (2007), *A spatially distributed flash flood forecasting model.* Submitted to Environmental Modelling & Software.

10. BORGA M (2006), HYDRATE Project: Description of work. University of Padova, 256 pp.

11. BORGA M, VIZZACCARO A (1997), *On the interpolation of hydrologic variables: formal equivalence of multiquadratic surface fitting and kriging.* Journal of Hydrology., 195, pp160-171.

12. BORGA M, ANAGNOSTOU E N, FRANK E (2000), *On the use of real-time radar rainfall estimates for flood prediction in mountainous basins.* Journal of Geophysical Research., 105, D2, pp2269-2280.

13. BORGA, M., DEGLI ESPOSTI, S., NORBIATO, D. (2006). *Influence of errors in radar rainfall estimates on hydrological modelling prediction uncertainty*. Water Resources Research, 42, W08409, doi:10.1029/2005WR004559.

14. BORGA, M., BOSCOLO, P., ZANON, F., SANGATI, M. (2007). *Hydrometeorological analysis of the August 29, 2003 flash flood in the eastern Italian Alps.* J. Hydrometeorology, 8(5), 1049-1067.

15. BUDYKO M I (1974), *Climate and life.* Academic Press, New York.

16. CARPENTER T M, SPERFSLAGE J A, GEORGAKAKOS K P, SWEENEY T, FREAD D L (1999), *National threshold runoff estimation utilizing GIS in support of operational flash flood warning systems,* Journal of Hydrology, 224, pp21-44.

17. CAZORZI F, DALLA FONTANA G (1996), *Snowmelt modelling by combining air temperature and a distributed radiation index.* Journal of Hydrology, 181, pp169-187.

18. CHANCIBAULT K, ANQUETIN S, DUCROCQ V, SAULNIER G M (2006), *Hydrological evaluation of high-resolution precipitation forecast of the Gard flash-flood event (8-9 September 2002),* Q. J. R. Meteorol. Soc, 132, pp 1091-1117.

19. CHAPON B (2006), *Etude des pluies intenses dans la région Cévennes-Vivarais à l'aide du radar météorologique. Régionalisation des traitements radar et analyse granulométrique des pluies au sol*, PhD Thesis (in French) University Joseph Fourier, Grenoble, 197 p.

20. COLLIER C (2007), *Flash flood forecasting : What are the limits of predictability?,* Quarterly Journal of the Royal Meteorological Society. Q. J. R. Meteorol. Soc., vol. 133, no 622A, pp. 3-23 .

21. COULIBALY P, ANCTIL F, BOBEE B (1999), *Prévision hydrologique par réseaux de neurones artificiels : état de l'art.* Canadian Journal of Civil Engineering, 26, pp293-304.

22. CREUTIN J D & BORGA M (2003), *Radar hydrology modifies the monitoring of flash flood hazard*, Hydrological Processes, 17, 7, pp1453-1456, doi 10.10002/hyp.5122.

23. DA ROS D & BORGA M (1997a), *Use of digital elevation model data for the derivation of the geomorphologic instantaneous unit hydrograph.* Hydrological Processes, 11, pp13-33.

24. DA ROS D & BORGA M (1997b), *Adaptive use of a conceptual model for real time flood forecasting.* Nordic Hydrology, 28(3), pp169-188.

25. DE MARCHI, B., SCOLOBIG, A. et al. (2007). Risk Construction and Social Vulnerability in Italian Alpine Region. *Tast 11 Report, FLOODsite Integrated Project. Gorizia, Institute for International Sociology.*

26. DELRIEU G, DUCROCQ V, GAUME E, NICOL J, PAYRASTRE O, YATES E, KIRSTETTER P E, ANDRIEU H, AYRAL P A, BOUVIER C, CREUTIN J D, LIVET M, ANQUETIN S, LANG M, NEPPEL L, OBLED C, PARENT-DU-CHÂTELET J, SAULNIER G M, WALPERSDORF A, WOBROCK W (2005), *The catastrophic flash-flood event of 8-9 September 2002 in the Gard region, France,* J. Hydrometeorology, 6, pp 34-52.

27. DINKU T, ANAGNOSTOU E N, BORGA M (2002), *Improving radar-based estimation of rainfall over complex terrain*, Journal of Applied Meteorology, 41(12), pp1163–1178.

28. DUAN Q, SOROOSHIAN S, GUPTA V K (1992), *Effective and efficient global optimization for conceptual rainfall-runoff models,* Water Resources Research, 28(4), pp265-284.

29. ELECTRICITE DE FRANCE (1972), Etude hydrologique du bassin de la Haute-Loire. Etudes des relations globales pluie-débit sur le haut bassin de la Loire à Bas-en-Basset, EDF-DTG & Laboratoire National d'Hydraulique (LNH), Technical report.

30. FAULKNER, H., PARKER, D., GREEN, C., BEVEN, K. (2007). *Developing a Translational Discourse to Communicate Uncertainty in Flood Risk between Science and the Practitioner*. Ambio Vol. 36, No. 7, December 2007.

31. GANDIN L S, MURPHY A H (1992), *Equitable skill scores for categorical forecasts.* Mon. Weather Rev., 120, pp361-370.

32. GAUME, et al. (2007). *First Year HYDRATE Report Project, WorkPackage 1 on* Flash Flood Primary data analysis.

33. GAUME E**,** GOSSET R (2003), *Overparametrization a major obstacle to the use of neural networks in hydrology,* Hydrology and Earth System Sciences, 7(5), pp693-706.

34. GAUME E, VILLENEUVE J P, DESBORDES M (1998), *Uncertainty assessment and analysis of the calibrated parameter values of an urban storm water quality model,* Journal of Hydrology, 210, pp38-50.

35. GEORGAKAKOS K P (1992), *Advances in forecasting flash floods,* In Proceedings of the CCNAA-AIT Joint Seminar on Prediction and Damage Mitigation of Meteorologically Induced Natural Disasters, 21-24 May 1992, National Taiwan University, Taipei, Taiwan, pp280-293.

36. GEORGAKAKOS K P (2004), *Mitigating adverse hydrological impacts of storms on a global scale with high resolution, global flash flood guidance,* Abstracts Volume of International Conference on Storms/AMOS-MSNZ National Conference, 5–9 July. Autralian Meteorological Society, Brisbane, Australia pp23-30.

37. GEORGAKAKOS K P (2006), *Analytical results for operational flash flood guidance,* Journal of Hydrology, 317 (1-2), pp81-103, doi:10.1016/j.jhydrol.2005.05.009.

38. HANDMER, J.W. (2001). Improving flood warnings in Europe. *Environmental Hazards.* 3,19-28.

39. HARGREAVES G H, SAMANI Z A (1982), Estimating potential evapotranspiration. J. Irrig. Drain. Div., 108, pp225-230.

40. HOSSAIN F, ANAGNOSTOU E N, DINKU T, BORGA M (2004), *Hydrological model sensitivity to parameter and radar rainfall estimation uncertainty.* Hydrological Processes, pp18-17, 3277-3291, 10.1002/hyp.5659.

41. HASSOUN M H (1995), *Fundamentals of artificial neural networks.* MIT Press, Cambridge.

42. HUARD D, MAILHOT A (2006), *A Bayesian perspective on input uncertainty in model calibration: Application to hydrological model abc.* Water Resources Research, 42, W07416, doi:10.1029/2005WR004661.

43. HUET P, MARTIN X, PRIME J L, FOIN P, LAURAIN C, CANNARD P (2003), *Retour d'expérience des crues de septembre 2002 dans les departments du Gard, de l'Hérault, du Vaucluse, des Bouches-du-Rhône, de l'Ardèche et de la Drôme.* Technical report, Ministère de l'Ecologie et du Développement Durable, Republique Française, 133pp.

44. IMRIE C E, DURUCAN S, KORRE A (2000), *River flow prediction using artificial neural networks : generalisation beyond the calibration range.* Journal of Hydrology, 233, pp138-153.

45. JAKEMAN A J, HORNBERGER G M (1993), *How much complexity warranted in a rainfall-runoff model?*, Water resources research, 29(8), pp2637-2649.

46. JAKEMAN A, LITTLEWOOD I, WHITEHEAD P (1990), *Computation of the instantaneous unit hydrograph and identifiable component flows with application to two small upland catchments,* Journal of Hydrology, 117, pp275-300.

47. KLEMES V (1986), *Operational testing of hydrological simulation models.* Hydrol. Sci. J., 31, pp13-24.

48. KOKKONEN T S, JAKEMANN A J, YOUNG P C, KOIVUSALO H J (2003) *Predicting aily flows in ungauged basins: model regionalisation from basin descriptors at the Coweeta Hydrologica Laboratory, North Carolina,* Hydrol. Processes, 17, pp2219-2238, DOI:10.1002/hyp.1329,2003.

49. LEBEL T, LABORDES J P (1988), *A geostatistical approach for areal rainfall statistics assessment.* Stochastic Environmental Research and Risk Assessment 2(4), pp245-261

50. MAIER H R & DANDY G C (2000), *Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications.* Environmental Modelling and Software, 15, pp101-124.

51. MARTINA M L V, TODINI E, LIBRALON A (2006), *A Bayesian decision approach to rainfall thresholds based flood warning,* Hydrol. Earth Syst. Sci., 10, pp413-426.

52. MASON S J (1982), *A model for assessment of weather forecasts.* Aust. Meteor. Mag., 30, pp291-303.

53. MASON S J, GRAHAM N E (1999), Conditional Probabilities, Relative Operating Characteristics and Relative Operating Levels, Weather and Forecasting, 14, pp713-725.

54. MERZ R & BLÖSCHL G (2004), *Regionalisation of basin model parameters,* Journal of Hydrology, 287, pp95-123, DOI:10.1016/j.jhydrol.2003.09.028.

55. MOGIL H M, MONRO J C, GROPER H S (1978), *NWS's flash flood warning and disaster preparedness programs,* Bulletin of the American Meteorological Society, 59, pp690-699.

56. MONTGOMERY D R & FOUFOULA-GEORGIOU E (1993), *Channel network source representation using digital elevation models*, Water Resources Research, 29(12), pp3925–3934.

57. MOORE R J (1985), *The probability-distributed principle and runoff production at point and basin scales*. Hydrol. Sci. J., 30, pp273-297.

58. MOORE R J & BELL V A (2002), *Incorporation of grownd water losses and well data in rainfall-runoff models illustrated using PDM.* Hydrology and hearth System Sciences, 6 (1), pp25-38.

59. MOORE R J, COLE S J, BELL V A, JONES D A (2006), *Issues in flood forecasting: ungauged basins, extreme floods and uncertainty.* In: K.N.N. Tchiguirinskaia andT.P. Hubert (Editors), 8th Kovacs Colloquium. Frontiers in flood research IAHS, Paris, pp103-122.

60. NASH J E & SUTCLIFFE J E (1970), *River flow forecasting through conceptual models, Part I: A discussion of principles,* Journal of Hydrology, 10, pp282–290.

61. NELDER J A & MEAD R (1965), *A simplex method for function minimization,* Computer Journal, 7(4), pp308-313.

62. NORBIATO, D., BORGA, M., SANGATI, M. & ZANON, F. (2007). *Regional Frequency Analysis of Extreme Precipitation in the eastern Italian Alps and the August 29, 2003 Flash Flood*. J. Hydrology, 345, 3-4, 149-166.

63. NORBIATO, D., BORGA, M., DEGLI ESPOSTI, S., GAUME, E. & ANQUETIN, S. (2008). *Flash flood warning based on rainfall depth-duration thresholds and soil moisture conditions: An assessment under European conditions*. J. Hydrology, under revision.

64. NTELEKOS A, KRAJEWSKI W F, GEORGAKAKOS K P (2006), *On the uncertainties of flash flood guidance. Towards probabilistic forecasting of flash floods.* Journal of Hydrometeorology, pp896–915.

65. NUISSIER O, DUCROCQ V, RICARD D, THOUVENAIN T, ANQUETIN S (2006), *A numerical study of three catastrophic precipitating events over western Mediterranean region (Southern France). Part I: Numerical framework and synoptic ingredients.* Q. J. R. Meteorol. Soc. (revision)

66. OUDIN L, MICHEL C, ANCTIL F (2005), Which potential evapotranspiration input for a lumped rainfall-runoff model? Part 1 - Can rainfall-runoff models effectively handle detailed potential evapotranspiration inputs? Journal of Hydrology, 303(1-4), pp275-289.

67. PARKER, D. J., TAPSELL, S., MCCARTHY, S. (2007). *Enhancing the human benefits of flood warning*, Natural Hazards, 43, 397-414.

68. PARAJKA J, MERZ R, BLÖSCHL G (2005), *A comparison of regionalisation methods for basin model parameters,* Hydrology and Earth System Sciences, 9, pp157-171.

69. PERRIN C, MICHEL C, ANDRÉASSIAN V (2003), *Improvement of a parsimonious model for streamflow simulation* Journal of Hydrology, 279, pp275-289.

70. PERRIN C, MICHEL C, ANDREASSIAN V (2001), *Does a large number of parameters enhance model performance ? Comparative assessment of common catchment model structures on 429 catchments.* Journal of Hydrology, 242, pp275-301.

71. RAMOS M H, BARTHOLMES J, THIELEN-DEL POZO J (2007), *Development of decision support products based on ensemble forecasts in the European Flood Alert System.* Submitted to Atmospheric Science Letters, Elsevier(minor revision).

72. REED S, SCHAAKE J, ZHANG Z (2007), *A distributed hydrologic model and threshold frequency-based method for flash flood forecasting at ungauged locations.* Journal of Hydrology, 337 (3), pp402-420.

73. ROMANG, H., WILHELM, C., FREDE, V., BEROD, D., HEGG, C., RHYNER, J. (2008). Emergency management tools for floods and debris flow events in smaller catchments. *In: Interpraevent: Tagungsband Internationales Symposium Interpraevent,* 2008, Dornbirn.

74. RUIN, I., GAILLARD, J. C., LUTOFF,, C. (2007). Drivers' risk perception of severe storms hazards in Southern France. *4th European Conference on Severe Storms 10 - 14 September 2007 – Trieste.*

75. SCHAEFER J T (1990), *The Critical Success Index as an indicator of warning skill.*, Weather and Forecasting, 5, pp570-575.

76. SWEENEY T L (1992), *Modernized areal flash flood guidance.* NOAA Technical Report NWS HYDRO 44, Hydrologic Research Laboratory, National Weather Service, NOAA, Silver Spring, MD, October, 21 pp. and an appendix.

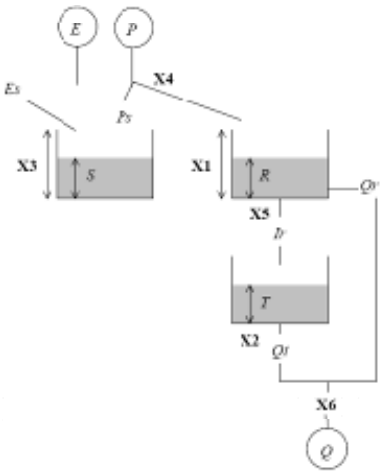
77. SPERFSLAGE J A, GEORGAKAKOS K P, CARPENTER T M, SHAMIR E, GRAHAM N E, ALFARO R, SORIANO L (2004), *Central America Flash Flood Guidance (CAFFG) User's Guide.* HRC Limited Distribution Report No. 21. Hydrologic Research Center, San Diego, CA, 82 pp.

78. WILKS D S (1995), Statistical Methods in the Atmospheric Sciences, An Introduction. Academic Press, New York, pp238-250.

79. YANG X, MICHEL C (2000), *Flood forecasting with a watershed model : a new method of parameter updating,* Hydrological Sciences Journal, 45 (4), pp537-546.

8. Appendix 1: Presentation of the various Rainfall-Runoff models used in Section 2

8.1 CREC model



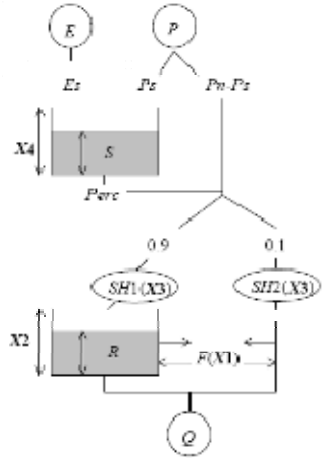
3 conceptual reservoirs and 6 parameters.

X3 and X4 have an impact on the « soil » reservoir, the evapotranspiration and the annual water budget,

X1, X5, X2 control the partition between slow and rapid flows and the shape of recession curves,

X6, parameter of a unit hydrograph, controls the shape of flood hydrographs.

8.2 GR4J model



2 conceptual reservoirs and 4 parameters

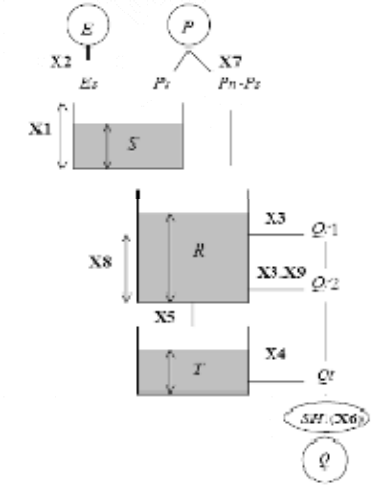
X4 parameter of the soil reservoir, controls the annual water budget,

X1 gain or losses due to groundwater exchanges with surround ding watersheds,

X3, unit hydrograph parameter, controls the shape of flood hydrographs

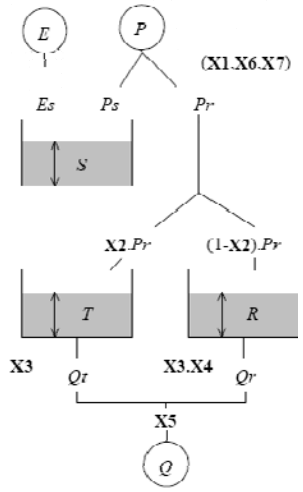
X2 controls the shape of recession curves

8.3 HBV model



3 conceptual reservoirs and 9 parameters.

8.4 IHAC model



3 conceptual reservoirs and 7 parameters

8.5 Artificial Neural networks

The networks used in this study are three-layer feed-forward neural networks (FNN see **Error! Reference source not found.**). These models have been widely used for hydrological modeling, because three layers are considered sufficient to generate arbitrarily complex output signals. Each of these three layers has a precise role. The first input or passive layer is dedicated to the capture of the external inputs (Xi) and to their delivery to each of the neurons of the next layer. The second, also called the hidden layer, performs complex non-linear mapping of the input data, in order to simulate the relationship between inputs and outputs (y)of the model. The outputs of the hidden layer are gathered and processed by the last or output layer, which delivers the final output of the network. The FNNs are only one example of the many possible structures of Artificial Neural Networks (ANNs).

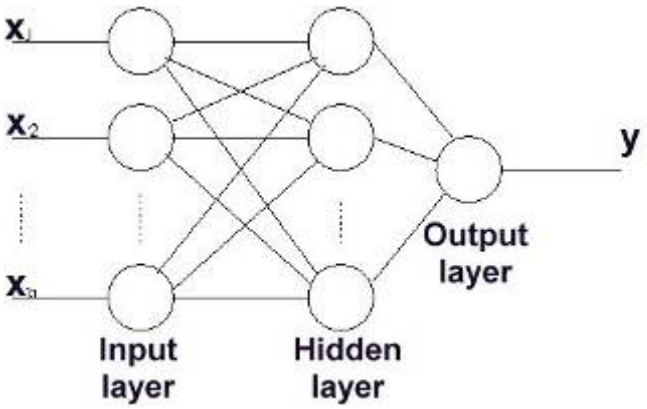


Figure 8-1 : Structure of a Feed-forward neural network

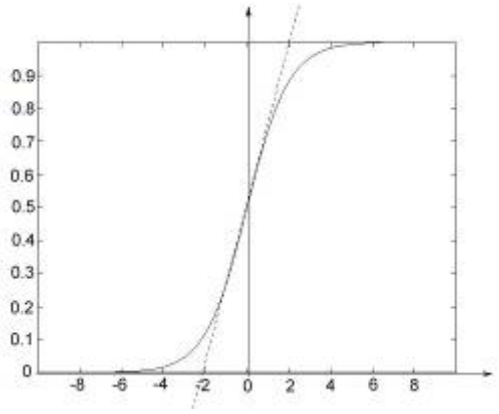


Figure 8-2: shape of the sigmoid function

A neuron is a processing unit with n inputs ($x_1,x_2,\ldots,x_n$) and only one output (y), with

$$y = f(x_1, x_2, ..., x_n) = A\left[w_0 + \sum_1^n w_i x_i \right]$$

where the w_i are the weights of the neuron and A is the so called activation or transfer function. In a FNN, the outputs of the neurons of a layer are the inputs of the neurons of the next layer. A sigmoid activation function (Figure 8-2) is generally chosen for the neurons of the hidden layer, the identity function being used for the input and output neurons.