RAINFALL FORECASTING USING LIMITED AREA MODELS AND STOCHASTIC MODELS

L. DI TRIA[a] , S. GRIMALDI[a] , F. NAPOLITANO[b] , L. UBERTINI[a,b]

[a] *University of Rome "La Sapienza", D.I.T.S. n° 37 - Dept. of Hydraulics, Transportation and Roads, Via Eudossiana, 18 00184 Rome – Italy*
e-mail: grimaldi@dits.ing.uniroma1.it
[b] *CNR-National Group for Hydrogeological Hazards Prevention (GNDCI), Via Madonna Alta, 126 06128 Perugia – Italy*
e-mail: fan@cidra.ing.uniroma1.it

ABSTRACT

The research is aimed to analysing some features of stochastic, with ARMA models, and deterministic rainfall forecasts, provided by limited area meteorological models, better known as LAMs.
First of all stochastic forecast procedure is developed, conditional upon the latest rain gauge measurements, by means an autoregressive moving average model, being the autocovariance structure of some clustered point processes used in rainfall simulation equivalent to the autocovariance structure of certain low-order ARMA processes. Next the LAMs rainfall field forecasts, analysed in order to evaluate the mean areal precipitation error with the observed values recorded in ground measurements, highlighted the presence of a bias in forecasted values. The coupling, in order to have better operative results, between the LAMs output based forecasting and the statistical ARMA based forecasting, has shown that, in the analysed case studies (some rainfall events observed in a small sub-basin of Tiber basin in central Italy), the use of LAMs forecast, which can be considered as an error affected measure of future precipitation, improves the 6 hours rainfall forecasting.

1 INTRODUCTION

Among natural hazards flooding and landslide occurrence, often referred to as hydrogeological hazard, are highly dangerous. In many countries the economic losses and casualties due to floods and landslides are greater than recognized, and generate a yearly loss of property larger than that from other natural disasters, including earthquakes, volcanic eruptions and windstorms.

In the last few years catastrophic rainfall events have occurred in the Mediterranean area, leading to floods, flash floods and shallow landsliding (debris and mud flow).

These recent events have outlined the urgent need for:

- the implementation of forecasting systems able to predict meteorological conditions leading to disastrous runoff occurrences;
- some policies for issuing warnings, or alarms, to local authorities and the population.

Indeed, early warning systems in urban areas appear to be the only non structural measure suitable for reducing flood risk, if diffused with enough lead time and adequate reliability. Flood and flash flood forecasting requires both the evaluation of the predictability of ground effects of large, or extreme, rainstorms, as well as the evaluation of the social response to an early warning message. Nowadays, one of the major objective in applied hydrology is the hydraulic protection of areas prone to flood risk, in order to decrease the probability of inundation and the reduction of damages on the territory.

This goal can be achieved by means structural measures (like river engineering works which allow to increase the hydraulic conveyance of the river or to temporary store water in some retention basins), or by means non-structural measures by using real-time flood forecasting systems.

With the arrival of new measuring systems (Meteorological Radar and Meteosat Satellites), and, especially, with the new possibilities offered by Limited Area Models (LAM's), it has finally become apparent that it is possible to create complex forecasting systems on a series of different time scales (see Todini, 1998a).

At present, there are essentially three basic systems for providing precipitation measurements, which can be used for real time flood forecasting:

1. conventional ground based tele-metering raingauges, with long historical records generally available for calibrating the rainfall-runoff models;
2. weather radar system, which importance has grown in the last decade, with a finer spatial description of the precipitation field and the possibility of observing approaching storms sometimes before arriving over the catchment of interest;
3. remote sensors such geo-stationary (e.g. Meteosat) or microwave polar satellites.

A further advance in the representation of rainfall fields was achieved with the advent of the so-called LAM's. At the present time, atmospheric forecasting models must be viewed (Brath, 1997; Bongioannini Cerlini & Todini, 1998, Todini, 1998a,b) as valid qualitative-quantitative rainfall forecasting tools at 24, 48 and 72 hours (of course, at these forecasting horizons it is required not an absolute precision, but rather an order of magnitude) for events of great intensity and when these phenomena occur on a considerable scale and size (e.g. the flood event of November 1994 on the Po river in Italy, Brath & Maione, 1997); nevertheless, they cannot yet be regarded as providers of quantitative rainfall forecasts in the short term (6-12 hours) to be used directly for flood forecasting purposes as an exciting force on hydrological models, since the quantitative forecasting of precipitation, on the time and space scales commensurate with the dynamic of the hydrological phenomena, has not yet achieved that degree of precision necessary to avoid on the one hand the non-forecasting of exceptional small-scale situations and, on the other, the issuance of unwarranted alarms.

In order to perform short term real-time flood forecasts for small basins (i.e. with size ranging 100-1,000 km^2), for which the flood forecast can be made only when a precipitation forecast is available, it has been pointed out that it is necessary to have, at least, an advance warning time of six-eight hours.

The aim of this paper is to use a combination of stochastic processes (e.g. ARMA models) based rainfall forecasts and LAM's quantitative rainfall predictions in order to bypass some disadvantages that each method displays, and to improve short term flood forecasting for medium-small river basins.

In section 2 the stochastic process used are highlighted, in section 3 some considerations about LAM's are presented, and in section 4 some analysed case studies are illustrated, finally in section 5 we discuss the results and the future needs on the subject.

2 RAINFALL FORECASTING WITH STOCHASTIC PROCESSES

Stochastic processes are widely used in hydrological variables (e.g.: rainfall, flows and temperature) forecasting (see Burlando et al., 1993 and references cited therein; Brath et al., 1997 and 1998).

Using ARMA family models (Box & Jenkins, 1976; Brockwell & Davis, 1987; Hipel & McLeod, 1994) it is possible to consider a time series as one of the potential realisations of a stochastic processes $X(t,\omega_o)$ (where $t \in T$: parametric space, $\omega_o \in \Omega$: event space), this allows us to define a linear model $\phi(B)X_t=\theta(B)a_t$, where X_t is the realisation of the process, $\phi(B)$ and $\theta(B)$ are the polynomial functions of the backshift operator B and a_t is the white noise. The polynomials represent the correlation structure of analysed time series, such as the filter $\theta(B)[\phi(B)]^{-1}$ applied to the process gives, as residual, an independent stochastic process a_t. In order to modelling a stochastic process the next logical steps are to be followed:

1) *preliminary analysis*: interpretation of the results of the Box and Cox transformations and of some differencing on the series performed to satisfy the stationary conditions (lack of trend, homoscedasticity);
2) *identification:* definition of the model type assuming the number and the order of the parameters that characterise the polynomials of the general model. This is achieved interpreting the trends of the autocorrelation function (ACF) and the partial autocorrelation function (PACF);
3) *fitting:* computing of the parameters values; this is achieved with the application of the maximum likelihood estimation method in the time domain developed with the Marquardt non-linear algorithm;
4) *control:* application of some tests on residuals to verify the properties of randomness, normality and stationarity.

The "*optimal*" model obtained with this procedure can be used both in simulation and in forecasting.

In order to perform a forecast of a stochastic process $X_{t=n}$ with k-terms ahead it is need the definition of statistical characteristics of random variable $X_{t=n+k}$, with the ensemble of information $I_n = \{X_{n-j},\ j=1,2,.....\}$, due to the fact that the forecast, $F_{n,k}$, that minimises the mean squared error, is obtained with the mean of X_{n+k} conditioned upon the I_n: $F_{n,k}=E(X_{n+k}|I_n)$.

Temporal rainfall models based on point processes theory, namely the Neyman-Scott Rectangular Pulses (NSRP), have shown a good agreement with historical rainfall series (Calenda & Napolitano, 1999 and references cited therein); Obeysekera et al. (1987) showed that the correlation structure of the average rainfall process over non-overlapping time intervals derived from such point process models is equivalent to the correlation structure of an ARMA(2,2) process. These similarities of the correlation structure of this ARMA process allow us to by-pass some steps of the procedure, precisely the steps 2 and 4, *identification* and *control*, that in the real-time forecasting case are the more difficult. In fact in this kind of application it is need to proceed with an adaptive parameters estimation procedure as illustrated in the work of Burlando et al. (1993). Briefly the adaptive procedure can be summarised as follows:

- definition of minimum number of data that assures the convergence of the parameter estimation algorithm;
- forecasting of a d number of data ahead;
- updating of parameters the time series after the d time;
- back to the first step.

In the ARMA(2,2) model the forecasting equation for the first term is:

$$X_t(1) = \phi_1 X_t + \phi_2 X_{t-1} - \vartheta_1 \hat{\varepsilon}_t - \vartheta_2 \hat{\varepsilon}_{t-1} \qquad (1)$$

3 RAINFALL FORECAST BASED ON LIMITED AREA MODELS

In recent years, considerable advances have been achieved in the mathematical modeling of the atmospheric phenomena, mainly by virtue of the rapid and continuous development of computer facilities with increasing computation capabilities. Accordingly, the spatial and temporal resolutions of the rainfall forecasts provided by numerical weather prediction models have been significantly increased. Therefore, one can expect that in a next future these forecasts could fully meet the requirements necessary for being used in real-time flood forecasting systems for medium and small size basins, affected by flash floods. Nowadays, rainfall forecasts are provided by the numerical implementation of several deterministic weather prediction models. As well known these models can be divided into General Circulation Models (GCM), or Mesoscale models, whose domain of integration is represented by the entire earth, and Limited Area Models (LAM), which operate on smaller domains (limited areas). Mesoscale models (with mesh sizes of 100x100 km^2) were basically conceived as descriptive models of climate and weather. As these models evolved, the *"precipitation" variable* was added to them, which is not however one of those quantities which model designers call *"state variable",* i.e. a variable which is vital for the description of the physical state of the system.

A further advance in the representation of rainfall fields was achieved with the Limited Area Models (initially and boundary conditions are assumed by the results provided by a GCM), with mesh sizes as small as the design limit of representation for hydrostatic models of 10x10 km^2, which on the one hand allow the orography to be introduced in greater detail and on the other allow a finer discretisation of the forecast precipitation quantities, and therefore a more realistic spatial variability.

Among the mesoscale models, one of the most advanced in the world is the GCM operated at the European Center for Medium range Weather Forecasting (ECMWF) in Reading, United Kingdom (see ECMWF 1995). In Italy, one of the most advanced models for operational precipitation forecasting is that operated in Bologna by the Meteorological Service of Emilia-Romagna Region (LAMBO model). The initial and boundary condition necessary for the operation of the LAM are derived from the ECMWF data.

The use of limited area models would allow to partially overcome the limitations related to model resolution. However, these limitations could have particularly severe effects for Italy, where the presence of important mountain barriers (both the Alps arid the Apennines) would require a much more limited grid size in order to get an enough accurate representation of the topography.

The development of mesoscale meteorological models on areas having complex orography has been recognized to be a major challenge of research for the future by meteorologists and hydrologists.

Some researches (Brath & Maione, 1997; Bongioannini Cerlini & Todini, 1998; Todini, 1998a,b; Cosentino, 1999) on the possibility of using directly LAM's results in a flood forecasting systems have shown, as cited in section 1, an inability, due their limitations about physical processes depiction and their spatial resolution, to quantify rainfall with a good degree of approximation without a post-processing of data. The confirm of this point is highlighted in figures 7a and 8a, where the LAM's outputs rainfall forecasting is shown for two events of the case study.

4 THE CASE STUDY

Some rainfall events, recorded in the Paglia basin in Central Italy from 1991 to 1998, has been considered.

The adaptive estimation procedure has been applied to the time series for performing the forecasting with time in advance of 1, 2 and 3 hours.

The figures 1, 2 and 3 show the comparison between the observed values and the forecasted values for every events analysed for the three time in advance considered: it is evident that from the third step of forecast the dispersion of the points is widespread and, as shown in figure 4, for a one event, the performance of the forecasting sharply decays.

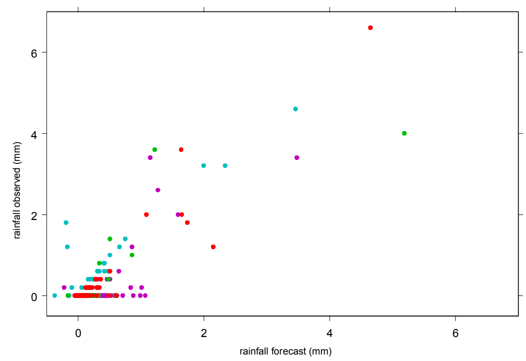


Figure 1 – Observed-Forecasted values for 1h time in advance

Observed and forecasted, *1h* and *2h* ahead, series are plotted in figures 5-6 for certain events considered: the initial temporal shift represents the minimum values number needed for the adaptive estimation procedure to converge.

With the aim of developing a reliable 6 hours ahead forecasting, LAM's outputs have been introduced, as a preliminary approach, in the forecast procedure.

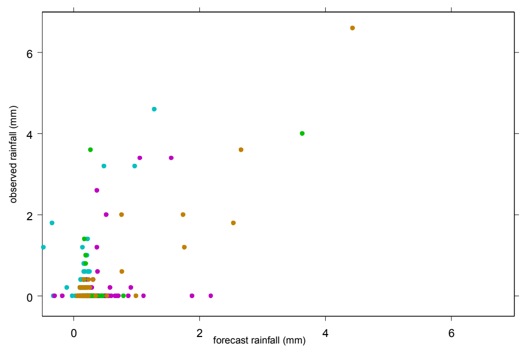


Figure 2 – Observed-Forecasted values for 2h time in advance

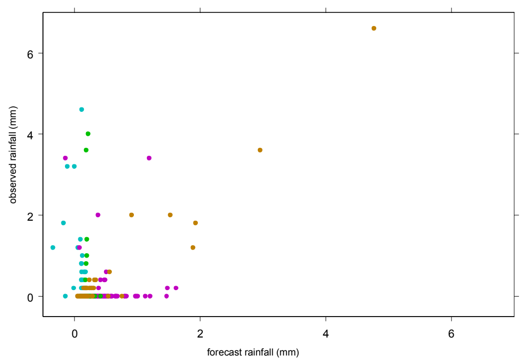


Figure 3 – Observed-Forecasted values for 3h time in advance

LAM's output for the events have been supplied by the Meteorological Service of Emilia Romagna Region.

At the present, as a first approximation, the coupling of stochastic processes based forecasting with the LAM's outputs is very rough. The simple idea is, with the aim to furnish a simple operative method, the following:

- the stochastic forecasting for two hours ahead is considered;
- the six hours ahead LAM's outputs are considered as a first approximation cumulative rainfall in the next future;

- the amount of the first two hours values forecasted are subtracted from the total amount of LAM's predictions;
- the last values to forecast from the 3rd hour to the 6th hour are obtained by the residual of LAM's outputs downscaled uniformly.

Figures 7 and 8 show the improvement of the forecasting after the combined procedure proposed above.

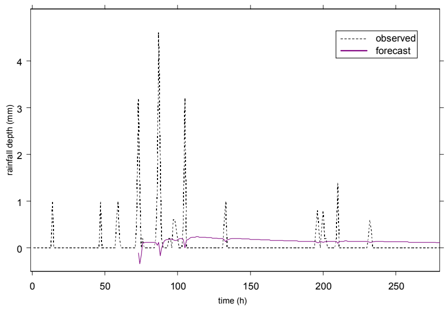
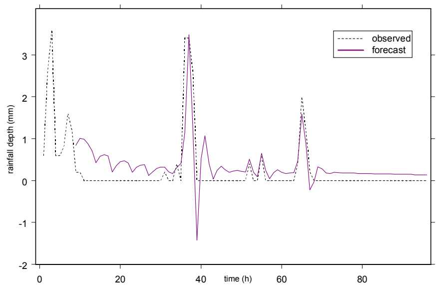
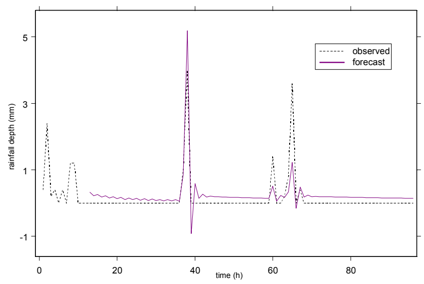


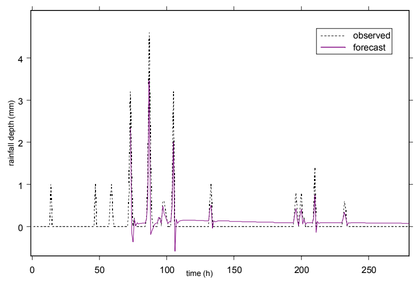
Figure 4 – Forecasted values for 3h time in advance for an event



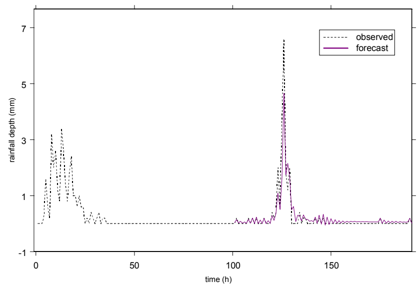
(a)



(b)



(c)



(d)

Figure 5 – Forecasted values for 1h time in advance for 4 events, a, b, c and d.

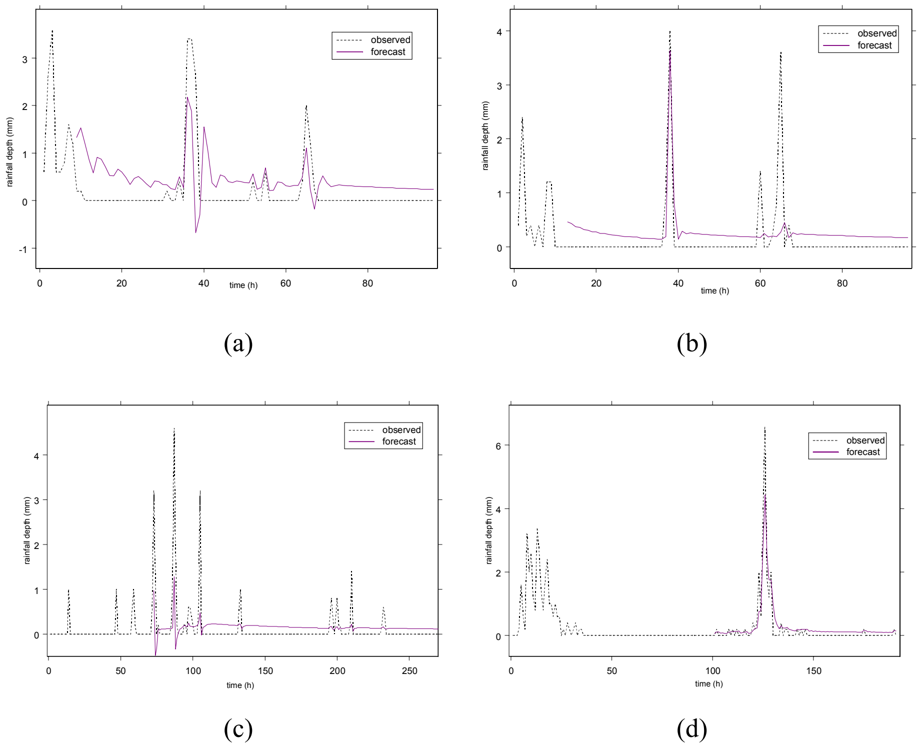


Figure 6 – Forecasted values for 2h time in advance for the same events, a, b, c and d.

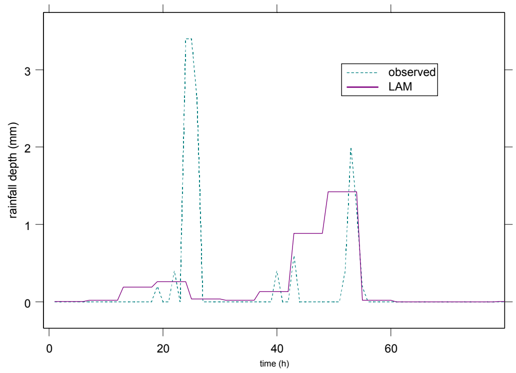


Figure 7a – Forecasting values (LAM) vs. observed value for an event

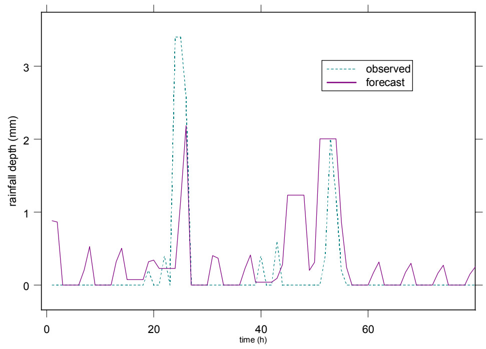


Figure 7b – Forecasting combined values (LAM & stochastic) vs. observed values

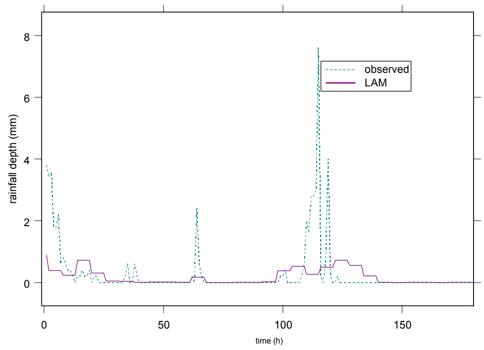


Figure 8a – Forecasting values (LAM) vs. observed value for an event

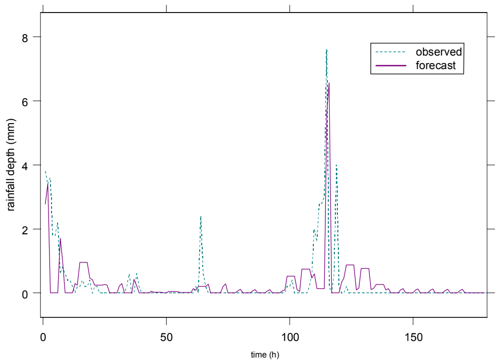


Figure 8 b - Forecasting combined values (LAM & stochastic) vs. observed values

5 CONCLUSIONS

Although very preliminary the results of the research have shown that, at present time, both hourly rainfall statistical forecasting and LAM's predictions are not, alone, capable to solve the problem of short term flood forecasting in small basins with concentration time less than 12 hours.

The statistical forecasting is satisfactory, when an adaptive procedure for parameter calibration is used, only for 2 terms ahead whereas for more than 2 terms is unsatisfactory.

A simple procedure of coupling stochastic forecasts with the LAM's outputs has shown that it is possible to overcome partially this troubles, even if the results, at present time, depend on the intrinsic limitation in the LAM's physical processes depiction and in the spatial resolution of rainfall measurements with traditional raingauges.

Some future research improvements are needed, both in the stochastic forecasting, for example considering also the non linear models, that are a very adequate representation for intermittent process like rainfall, and in the LAM's spatial resolution and physical processes schematisation.

ACKNOWLEDGMENTS

Authors thank Prof. Stefano Tebaldi, Meteorological Service of Emilia Romagna region, for providing the LAMBO data; Mauro Colantonio and Monica Cianfa for their fruitful help in processing data.

REFERENCES

Bongioannini Cerlini P. and Todini E.: "Studio di previsioni di piena mediante l'uso di modelli meteorologici e statistici", Proc.of XXVI Convegno di Idraulica e Costruzioni Idrauliche, Catania, Italy, 1998.

Box G.E.P. & Jenkins G.M., "Time Series Analysis: Forecasting and Control", Holden-Day, San Francisco, California. 1976.

Brath A.: "On the role of numerical weather prediction models in real time flood forecasting", Proc. Intern. RIBAMOD Workshop on Integrated Systems for Real-Time Flood Forecasting and Warning, EU DGXII Com., 1997.

Brath A. and Maione U.: "Sulla capacità predittiva dei modelli meteorologici ai fini della prevenzione delle catastrofi idrogeologiche " Proc. of int.

Workshop "La prevenzione delle catastrofi idrogeologiche: il contributo della ricerca scientifica", Alba, Italy, 1997.
Brath A., Franchini M., Montanari A. and Toth E.: "Coupling deterministic and stochastic models for real-time flood forecasting", Proc. Intern. RIBAMOD Workshop on Integrated Systems for Real-Time Flood Forecasting and Warning, EU DGXII Com., 1997.
Brath A., Montanari A. and Toth E.: "Su alcune tecniche stocastiche per il miglioramento delle prestazioni del preannuncio di piena", Proc.of XXVI Convegno di Idraulica e Costruzioni Idrauliche, Catania, Italy, 1998.
Brockwell P.J. Davis R.A, "Time Series. Theory and Methods", Springer Verlag, New York, 1987.
Burlando P., Rosso R., Cadavid L.G. and Salas J.D.: "Forecasting of short-term rainfall using ARMA models", *J. of Hydrol.*, 144, 193-211, 1993.
Calenda G. and Napolitano F.: "Parameter estimation of neyman-scott processes for temporal point rainfall simulation", in press on *J. of Hydrol.*, 1999.
Cosentino C.: "Previsione in tempo reale del rischio idrogeologico" Ph.D. Thesis University of Romatre, 1998.
ECMWF (European Center for Medium-Range Weather Forecasts):" The description of the ECMWF/WCRP level III-A global atmospheric data archive", Reading, UK, 1995.
Hipel K.W. McLeod A.I., "Time Series Modeling of Water Resources and Enviromental Systems", Elsevier Science, 1994.
Obeysekera J.T.B., Tabios III G.Q. and Salas J.D.: "On parameter estimation of temporal rainfall models", *Water Resour. Res*., 23-10, 1837-1850, 1987.
Todini E.: "FLOODSS – Un sistema operativo di supporto decisionale per la mappatura delle aree a rischio di alluvione e la gestione delle emergenze", Proc.of XXVI Convegno di Idraulica e Costruzioni Idrauliche, Catania, Italy, 1998.
Todini E.: " Real time flood forecasting: problems, possibilities and proposed solutions", Springer-Verlag, 1998.