COUPLING METEOROLOGICAL AND HYDROLOGICAL MODELS FOR RIVER DISCHARGE FORECASTING. PART I: A METHODOLOGICAL APPROACH

Alvise Bozzo[1], Stefano Serafin[1], Dino Zardi[1]

[1] Atmospheric Physics Group, Department of Civil and Environmental Engineering, University of Trento, Italy
E-mail: *stefano.serafin@ing.unitn.it*

Abstract: The present contribution proposes an approach for the preparation of the meteorological input needed by hydrological models to forecast the surface runoff in medium-sized river catchments (50-1000 km^2). Rather than using Limited Area Models (LAMs) to achieve a high-resolution representation of atmospheric processes, we propose to reconstruct precipitation fields at a local scale from the statistical elaboration of forecasts from a General Circulation Model (GCM, approx. resolution 20 km). The calibration of statistical methods requires the availability of a training set of historic precipitation measurements and weather analyses. With the aim of maximizing the predictive skill of the procedure, the training set can be made homogeneous by selecting a subset of similar days through an analogue search procedure: different approaches developed within synoptic climatology or borrowed from statistical data analysis can be used to the purpose. The structure of a feasible complete downscaling model is outlined in this paper, while an application is shown in the second part of this work (Bozzo et al., 2007).

Keywords: *Rainfall forecasting, Downscaling, Synoptic Classification, Hydro-meteorological models.*

1. INTRODUCTION

Several different approaches have been developed in order to forecast the river discharge in mountain catchments, ranging from empirically-based models, with a lumped structure, to physically-based spatially distributed models. Both kinds of models can be used at different temporal resolutions, e.g. for the real-time forecasting of discharge in a single event or for the long-term simulation of the water balance.

The accuracy of the predictions provided by these models is limited by a number of different sources of uncertainty. A *structural uncertainty* arises from an incomplete or simplified representation of the physics of the phenomenon, and is in principle particularly relevant in lumped models, where the processes of runoff generation and propagation are highly parameterized. Further, a *parametric uncertainty* stems from the fact that the set of parameters appearing in a given model structure cannot be considered of general validity, i.e. suited to represent a wide range of different physical conditions; this is why parameters are often calibrated on an empirical basis by minimizing a model error function. Finally, an *uncertainty in the initial conditions* is present, due to the inevitably incomplete knowledge of the initial distribution of the physical quantities relevant in the process. In the context of runoff forecasting models, the distribution of the forecast precipitation on a given area is obviously a factor of prominent importance. Here we propose an approach to limit inaccuracies in the distribution of the forecast rainfall in a given area, thereby reducing uncertainty in the initial conditions of hydrological models.

Atmospheric LAMs with a horizontal resolution of a few km are widely available nowadays, and their rainfall forecasts can in principle be used as inputs for runoff models, in a sort of one-way meteo-hydrological modelling chain. In a totally different approach, rainfall fields at a local scale can be reconstructed from the statistical elaboration of forecasts from a GCM (in this case the deterministic runs of the ECMWF Integrated Forecasting System model, http://www.ecmwf.int/research/ifsdocs/). To the purpose, eigentechniques originally developed for the downscaling of weather scenarios from climate models can be used.

This choice is mainly motivated by two considerations:

1. although LAMs generally meet the requirements needed to simulate rainfall with a horizontal resolution of the order of very few km (e.g. non-hydrostatic formulation, high-order conservative numerical schemes, proper parameterization of microphysical, boundary-layer and land-surface processes), the initialization data for such models are seldom really representative of the fine-scale variability of the atmosphere.

2. the fine-scale spatial (and temporal) structure of a rainfall field is extremely intermittent, and therefore in principle it is hardly predictable with any LAM. This happens because precipitation fields are affected by extremely localized features, which are often not entirely or not correctly captured by model grids (e.g. orography, land use, local circulations).

Essentially, we recognize that a gap exists between the scale of variability of real precipitation fields and the scale at which quantitative analyses of atmospheric variables are available; furthermore, physical factors inherently difficult to model become important precisely in this range of scales. Consequently, it is reasonable to suppose that the use of physically-based models can only have a limited success in reproducing the scale relationships between the two extremes of the range, and in providing reliable rainfall estimates. Therefore, we take the forecasts of the ECMWF model as representative of the state of the atmosphere at the scale of a few hundreds of km^2 (which we assume is a reasonable resolution limit for forecasts initialized with mesoscale observations), and seek for a suitable methodology to produce fine-scale predictions from these forecasts.

2. AN APPROACH TO THE DOWNSCALING OF DAILY PRECIPITATION

Downscaling or disaggregation methods assume that the fine-scale distribution of some quantities can be inferred from a reliable depiction of their large-scale structure, based on a statistical model of the relationships existing between the two scales. Although it is possible to choose and consider only one variable at a time (as happens in the so-called disaggregation approaches), it is also possible to build models seeking for the relationship between several large-scale predictor variables and a single small-scale predictand variable. This possibility is interesting because atmospheric simulation models tend to provide better forecasts of quantities which are affected only marginally by parameterization schemes; these include for instance geopotential height and mean sea-level pressure, but not precipitation. Knowledge of well-predicted variables then appears to be a useful addition to the information content used for the downscaling of precipitation.

The feasibility of a downscaling approach for forecasting daily rainfall at a catchment scale of approx 50-500 km^2 is investigated in this work. An extensive training set of both the predictand and predictor variables is of course needed in order to calibrate downscaling models. Data were collected and methods were developed for application in the Trentino-Alto Adige area (north-eastern Italy, approx. size 14000 km^2), roughly corresponding to the upper portion of the Adige river catchment.

Historical data of the predictand variable, i.e. daily rainfall measurements in the period 1995-2005, have been recovered for 74 weather stations owned by Meteotrentino (the weather service of the Province of Trento), 81 weather stations owned by the Hydrographic Office of the Province of Bolzano, 46 rain gauges owned or managed by the Service for Hydraulic Works of the Province of Trento, and 91 weather stations owned by the IASMA Agricultural Research Institute. This makes for a total of 292 measurement points, a rather dense observation network where each rain gauge is on average representative of an area of approx. 50 km^2.

The predictor variables are taken from ECMWF analyses of temperature, specific humidity and geopotential height at different isobaric surfaces, and from forecasts of total precipitation with lead times ranging from 24 to 120 hours. All ECMWF data cover a wide area including western Europe (20°W-30°E, 30°-60°N), with a grid-resolution of 0.25 degrees (approx. 20 and 27 km in the E-W and N-S directions respectively). The reasons for considering such a large area is that the occurrence of precipitation in a small region does not depend only on local factors, but is considerably influenced by the large-scale situation of the atmosphere.

The heart of the downscaling problem lies in finding the common patterns of variability between two distinct sets of data, a large scale predictor and a small scale predictand. These data can be visualized as tables with an equal number of rows, each referring to a sample day. Every day of the sample is then described by a number of variables (the columns of the tables). Variables in the predictor table are a set of precipitation forecasts and other atmospheric variables known at different sites, while variables in the predictand table are a set of sparse precipitation measurements. The predictor and predictand tables need to be coupled in time but not in space, i.e. they need to be referred to the same days, but they can make use of measurements or forecasts referred to locations which are not exactly coincident.

The search of joint variability patterns in the two datasets makes use of tools from multivariate statistics (Singular Value Decompostion and Canonical Correlation Analysis): a description of the algorithms leading to the identification of joint variability modes and a comparative analysis of their results has been carried out by Bretherton et al. (1992), while their forecasting application has been described by Widmann et al. (2003).

Particular care needs to be taken in the definition of the training set used in calibration. We assume that raw time-series data from weather stations are not suited to form a proper set, mostly for two reasons. The first is that precipitation is a discontinuous variable, and time-series of daily precipitation observed in the area usually consist of long periods with no precipitation, separated by rather short rainfall events. Therefore, unmodified timeseries would make for a highly inhomogeneous set of data. The second reason is that pointwise rainfall measurements often have a very localized representativity, and are influenced by a quantity of potential limiting factors, measurement errors included. From a statistical point of view, these features result in a "noisy" field with a high variance (often several times larger than the mean). On one hand, this variance cannot be entirely captured by any synthethic statistical model; on the other, it isn't either entirely physically significant.

In light of these considerations, raw data should be processed in order to: (1) form a homogeneous data set, and (2) reduce their variance by smoothing out local or noisy variability. The first aim is pursued by finding the closest historical analogues to each day for which a rainfall forecast is desired (the analogy being evaluated both in terms of synoptic-scale circulation and of rainfall amount): as a result, the downscaling training set is discontinuous and varies from day to day. The second objective is instead met by spatializing rainfall measurements on a thick regular grid, and averaging gridpoint values at the desired spatial scale. For instance, if a rainfall estimate is desired in the elementary catchments of a hydrological model, then the rainfall expected in the gridpoints comprised in each of them can be averaged together: the final predictand dataset will have as many columns as the number of catchments considered.

In summary, a strategy to provide a quantitative estimation of rainfall may consist of a sequence of steps:

1. *Determine if precipitation is expected to occur in the target day* (e.g. by using a threshold criterion).

2. *If so, find in the historical record of rainy days the closest analogues to the target day itself, and use the corresponding predictor data to build a predictor training set.* A number of different synoptic typing strategies, mostly developed in the field of synoptic climatology, can be used in order to isolate a set of similar days from a larger sample. For instance, in the Lamb approach (El Dessouky and Jenkinson, 1975) day classification is based on a series of circulation indices computed from the pressure field on a regular grid: days assigned by the synoptic typing method to the same class are treated as analogous. According to the Teweles-Wobus analogue search method (Obled et al., 2002) days are considered similar if a score quantifying their difference between the zonal and meriodional pressure gradients is close to zero. The Lund and Kirchhofer classification approaches (El-Kadi and Smithson, 1996) identify analogous days with a similar procedure: the former selects those displaying a high cross-correlation index for a given variable at a given level (e.g. sea level pressure), while the latter those having a low mean square deviation for a given autoscaled variable at a given level. Finally, hierarchical clustering algorithms can be used to the purpose (Kalkstein et al., 1987): in these methods, clusters are created by iteratively grouping days separated by the minimum distance. The distance between days can be evaluated as the Euclidean norm in the phase space of several meteorological quantitities (e.g. geopotential height, temperature and specific humidity at different levels), as well as in the space of their leading principal components. Any of these approaches finally provides a set of days sorted in order of decreasing similarity.

3. *Process rain gauge data in the set of analogue days to remove spurious variability, thus forming an appropriate predictand data set.* The rendering of sparse data on a regular grid can be performed e.g. by kriging (Wackernagel, 2003), a geostatistical interpolation method often used in hydrological applications. The value of a quantity at any point in space is evaluated as a weighted average of values measured at other points. Weights are determined by solving a linear system whose formulation descends from the requirements of unbiased estimation and optimality. Because every spatial field is considered as a particular realization of a spatially distributed random function, the first criterion results in imposing that the expected value of the estimate at a point coincides with the (genereally unknown) mean of the function at that point. The second requires instead that the variance of the estimate be minimum. The system arising from these two conditions is completely specified only once the covariance structure of the random function, represented by its variogram, is known. A variogram represents the variation of the degree of spatial

dependence of a random function with increasing lag, and needs to be evaluated based on experimental data. Measurement errors are described by variograms with a nonzero variance at a zero lag; in this case, kriging produces a smooth interpolated field with discontinuities at the measurement points. It has been verified that intrinsic kriging using a variogram estimated using the Li-Lake algorithm and fitted to an exponential function with nugget can be used to reconstruct precipitation fields varying smoothly on a regular grid from pointwise measurements. The desired predictand training set (e.g. rainfall in n different regions) can then be formed by averaging the rainfall estimate at the gridpoints included in each region.

4. *Find a statistical relationship between predictor and predictand, and use it prognostically*. Modes of joint spatial variability between the predictand field and other predictor variables can be identified by studying their cross-covariance spectrum through Singuar Value Decomposition or Canonical Correlation Analysis. Both techniques allow to represent the two data matrices as a linear combination of a series of modes sorted in order of decreasing mutual cross-covariance. Once the leading modes are isolated, two sets of time expansion coefficients (TECs) which relate them to the predictor and predictand data are computed. A cross-scale relationship between the two sets of TECs can be found either by linear regression or by nonlinear estimation based on an Artificial Neural Network approach, and then used prognostically. The predictive step consists in computing the TECs for the large-scale predictor in the target day of the forecast, and using it to estimate the TECs for the predictand through the previously found cross-scale function. The forecast field is then obtained by combining the known leading spatial modes of the predictand using the estimated TECs.

3. CONCLUSIONS

The conceptual framework of a statistical model allowing a fine-scale estimation of rainfall from the simulations of a General Circulation Model has been outlined. Various alternative approaches to the tasks of (1) searching historical analogues to the current forecast, (2) regionalizing pointwise precipitation measurements, and (3) recognizing cross-scale joint variability modes have been sinthetically described. The practical implementation of these steps is shown in Bozzo et al. (2007), along with some preliminary results.

Acknowledgements: *The authors acknowledge ECMWF for providing gridded analysis and forecast data, and Meteotrentino, the Hydrographic Office of the Province of Bolzano, the Service for Hydraulic Works of the Province of Trento and the IASMA Agricultural Research Institute for providing measured rainfall data.*

REFERENCES

Bozzo, A., S. Serafin, A. Pasetto, and D. Zardi, 2007: Coupling meteorological and hydrological models for river discharge forecasting. Part II: A case study about hydropower generation management. *Proc. 29th Int. Conf. on Alpine Meteorology, Chambéry (F), 4-8 June 2007*.

Bretherton, C. S., C. Smith, and J. M. Wallace, 1992: An intercomparison of methods for finding coupled patterns in climate data. *J. Climate*, **5**, 541–560.

El Dessouky, T. M. and A. F. Jenkinson, 1975: An objective daily catalogue of surface pressure, flow and vorticity indices for Egypt and it's use in monthly rainfall forecasting. *Meteorol. Res. Bull., Egypt*, **11**, 1–25.

El-Kadi, A. K. and P. A. Smithson, 1996: An automated classification of pressure patterns over the British Isles. *T. I. Brit. Geogr.*, **21**, 141–156.

Kalkstein, L. S., G. Tan, and J. A. Skindlov, 1987: An evaluation of three clustering procedures for use in synoptic climatological classification. *J. Climate Appl. Meteor.*, **26**, 717–730.

Obled, C., G. Bontron, and R. Garçon, 2002: Quantitative precipitation forecasts: a statistical adaptation of model outputs through and analogues sorting approach. *Atmos. Res.*, **63**, 303–324.

Wackernagel, H., 2003: *Multivariate geostatistics. An introduction with applications*. Springer, 3rd edition.

Widmann, M., C. S. Bretherton, and E. P. Salathé Jr., 2003: Statistical precipitation downscaling over the Northwestern United States using numerically simulated precipitation as a predictor. *J. Climate*, **16**, 799–816.