ASSESSING SPATIAL UNCERTAINTY OF REFERENCE EVAPOTRANSPIRATION USING STOCHASTIC SIMULATION IN SOUTHERN ITALY (CALABRIA REGION)

International Association of Hydrological Science
STAHY-WG
STAHY Statistics in Hydrology Working Group
IAHS AISH

International Workshop
"Advances in statistical hydrology"
May 23-25, 2010
Taormina, Italy

Gabriele Buttafuoco[1], Tommaso Caloiero[2], Roberto Coscarelli[2]

1) Institute for Agricultural and Forest Systems in the Mediterranean, National Research Council, Rende CS (Italy)
2) Research Institute for Geo-Hydrological Protection, National Research Council, Rende CS (Italy)
(Ref: g.buttafuoco@isafom.cs.cnr.it)

1. INTRODUCTION

Environmental management decisions require an accurate computation of water balance. Evapotranspiration is one of the major components of the water balance and has been identified as a key factor in hydrological modelling. For this reason, several methods have been developed to calculate the reference evapotranspiration (ET_0). Whatever model is used, the errors in the input will propagate to the output of the calculated ET_0. Neglecting information about estimation uncertainty, however, may lead to improper decision-making and water resources management. One geostatistical approach to spatial analysis is stochastic simulation, which draws alternative, equally probable, realizations of a regionalized variable. Differences between the realizations provide a measure of spatial uncertainty and allow to carry out an error propagation analysis. Among the evapotranspiration models, the Hargreaves-Samani model was used.

The aim of this paper was to assess spatial uncertainty of a monthly reference evapotranspiration model resulting from the uncertainties in the input parameters (mainly temperature) in southern Italy (Calabria region). Temperature data were simulated by using Turning Bands simulation with elevation as external drift.

The ET_0 was then estimated for each set of the 500 realizations of the input variables, and the ensemble of the model outputs was used to infer the reference evapotranspiration probability distribution function. This approach allowed to delineate the areas characterized by greater uncertainty, to improve supplementary sampling strategies and Et_0 value predictions.

2. STUDY AREA AND TEMPERATURE DATA

The Calabria region covers the southern part of the Italian peninsula (Figure 1) with a surface of 15,080 km^2 and a coastline of 738 km on the Ionian and Tyrrhenian seas. In the North, it borders Basilicata region for 80 km. The Calabria region has an oblong shape with a length of 248 km and a width ranging between 31 and 111 km. Although Calabria does not have many high summits, it is one of the most mountainous regions in Italy (Figure 1): 42% of the land is mountainous, 49% hilly and only 9% is flat. The maximum elevation is 2267 m a.s.l., while the average elevation is 597 m a.s.l. Temperatures have been measured daily at 134 meteorological stations of the Italian Hydrographic Service (at present "Centro Funzionale Multirischi della Calabria" of the "Agenzia Regionale per la Protezione dell'Ambiente della Calabria (Arpacal)") during the period 1924-2009 (Figure 1). Only temperature data for June have been considered in this work. Temperature series having less than 30 years of observation were discarded and eventually 35 series have been selected. The temperature station mean density was 1 out 430 km^2.

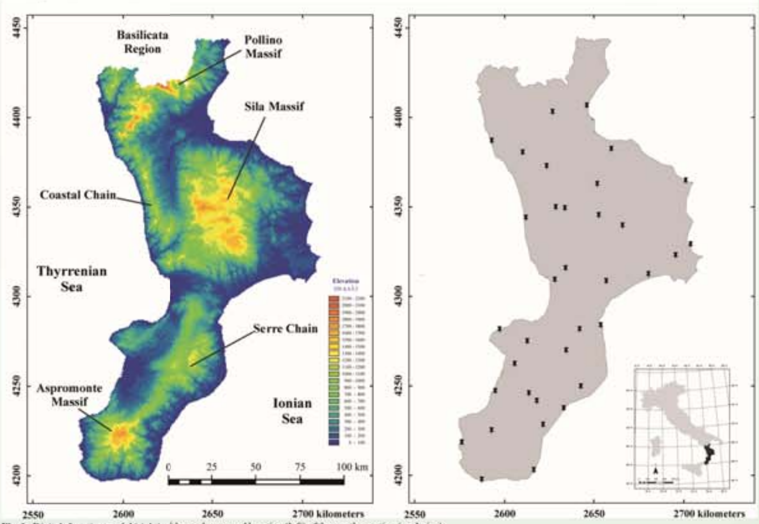

Fig. 1 - Digital elevation model (right) of the study area and location (left) of the weather station (pushpins).

3. THE HARGREAVES-SAMANI METHOD

The Hargreaves equation (Hargreaves and Samani, 1985) can be written as:

$$ET_0 = 0.0023 R_a (T + 17.8)\sqrt{(T_{max} - T_{min})}$$

where ET_0 is the computed reference evapotranspiration (mm d^{-1}); R_a is the water equivalent of the extraterrestrial radiation (mm d^{-1}) computed according to Allen et al. (1998); T_{max}, T_{min} and T are the daily maximum, minimum and mean air temperature (°C), with T calculated as the average of T_{max} and T_{min}. 0.0023 is the original empirical coefficient proposed by Hargreaves and Samani (1985).

4. STOCHASTIC SIMULATION

To assess the model output error resulting from the uncertainties in the input parameters, a Monte Carlo analysis (Heuvelink, 1998) was used, which consists in the generation of an adequate random input data set realizations which consider the joint distribution of all input variables. The model was run for each single set of realizations of the input variables and the ensemble of model outputs was used to infer the output probability function. A single Monte Carlo simulation consists in model running at all locations of a fine grid covering the interest region. We applied joint multivariate stochastic simulation (Gomez-Hernandez and Journel, 1992; Goovaerts, 1997), which is aimed at making predictions of cross-correlated variables. Such prediction is accomplished using a variogram matrix which includes not only spatial autocorrelation, but also spatial cross-correlation between variables. The latter information is expected to improve the spatial prediction of ET_0 by reducing its uncertainty which compared with traditional Monte Carlo simulations. Stochastic simulation actually allows to estimate cell-specific probability distribution functions which reflect the location of known data points and the spatial correlation structures of the variables.

Among the simulation techniques, the turning bands method with external drift was chosen because the weather stations having temperature data with more than 30 years of observations were only 35. Elevation is additional and denser information easily available, which can improve the temperature estimation. In the scope of simulation with external drift, the variable of interest $Z(\mathbf{x})$ comprises deterministic and stochastic components, then it can represent the combination by the model:

$$Z(\mathbf{x}_\alpha) = m(\mathbf{x}) + \varepsilon(\mathbf{x}) \quad \text{and} \quad E[Z(\mathbf{x})] = m(\mathbf{x})$$

where $\varepsilon(\mathbf{x})$ is the stochastic component with zero mean and variogram $\gamma_\varepsilon(\mathbf{x})$ and $m(\mathbf{x})$ is the drift which is usually modelled as a linear function of a smoothly varying secondary (external) variable $y(\mathbf{x})$:

$$m(\mathbf{x}) = a_0(\mathbf{x}) + a_1(\mathbf{x})\, y(\mathbf{x})$$

The principle is to simulate a target variable using an auxiliary linear correlated variable known at the grid nodes of the result grid file. The auxiliary correlated variable in this case study was elevation because there is a good linear correlation between both mean minimum (-0.90) and mean maximum (-0.87) temperature data. The value of the secondary variable (elevation) must be known at all primary data locations $\mathbf{x}_\alpha$ ($\alpha = 1, \ldots, n$) and at all locations $\mathbf{x}_0$ being estimated. Moreover, the secondary variable should vary smoothly in space to avoid instability of the kriging with external drift system (Goovaerts, 1997). The simulation algorithm generates a 2D simulation from the 1D simulations along the lines.

As the turning bands method is a Gaussian simulation technique, it requires a multi-Gaussian framework. Therefore, each variable has been transformed into a normal distribution beforehand and the simulation results have then been back transformed to the raw distribution afterwards.

Each geostatistical multivariate simulation has been used as input to the ET_0 model. Probabilistic information has been extracted from the set of simulated images. By averaging the simulated values at each cell, two different maps have been produced: the map of the expected value at any given location (E-type or Expected-value estimate; Journel, 1983) and the one of its standard deviation. The uncertainty in model predictions has been quantitatively evaluated from the replicate stochastic images.

REFERENCES

Allen, R.G., Pereira, L.S., Raes, D., Smith, M. (1998). Crop evapotranspiration guidelines for computing crop water requirements. FAO Irrigation and Drainage, Paper No. 56, Rome, Italy, 328 pp.
Gomez-Hernandez, J.J., Journel, A.G (1992). Joint Sequential Simulation of Multigaussian Fields. In Geostat Troia 1992, Kluwer Publ. Dordrecht, Holland.
Goovaerts, P. (1997). Geostatistics for Natural Resources Evaluation. Oxford University Press, New York, USA, pp 483.
Hargreaves, G.H., Samani, Z.A. (1985). Reference crop evapotranspiration from temperature. Applied Engineering in Agriculture, 1: 96-99.
Heuvelink, G.B.M. (1998). Error Propagation in Environmental Modeling with GIS. Taylor, Francis, London, UK, 127 pp.
Journel, A.G (1983). Non-parametric estimation of spatial distributions. Mathematical Geology, 15: 445-468.

4. RESULTS AND DISCUSSION

The summary statistics of the three variables are reported in the Table 1. The assumption of normal distribution was not accepted for both mean maximum (T_{max}) and mean minimum temperature (T_{min}) data at a probability level $p > 0.10$ and the data distributions showed long negative tails. Therefore, before conducting joint Gaussian simulation, we applied a Gaussian transformation to T_{min} and T_{max} data and fitted a linear model of coregionalization (LMC) to all direct and cross-variograms of the Gaussian transformed variables (Goovaerts, 1997). The LMC includes two different structures: a nugget effect and an exponential structure with a practical range of 54500 m. The goodness of fitting was evaluated by a cross-validation test, whose results in terms of mean experimental error and variance of standardised error were close to 0 and 1, varying between -0.05 and -0.03, and 0.94 and 0.96, respectively. The above LMC was used to produce the 500 simulations of T_{min} and T_{max}, then the expected values of T_{min} and T_{max} and their standard deviation were computed and mapped (Figure 2 and Figure 3).

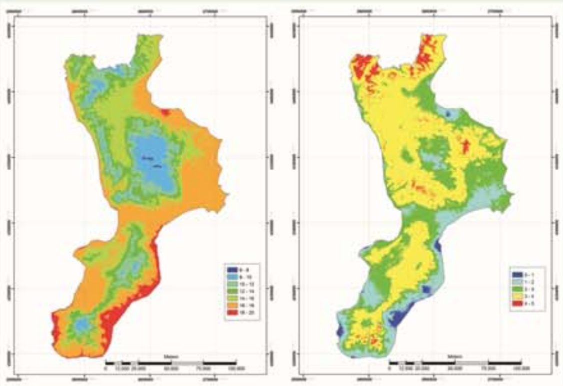

Fig. 2 - Mean and standard deviation maps of simulations for the mean minimum temperature (T_{min}).

	T_{max}	T_{min}	Elevation
Mean	27.0	15.9	348.89
Stand. Dev.	2.6	3.2	403.83
Maximum	30.9	19.7	1315.00
Upper quartile	28.7	18.0	516.00
Median	27.9	17.1	242.00
Lower quartile	26.0	15.2	15.00
Minimum	20.1	7.7	5.00
Skewness	-1.1	-1.4	1.29
Kurtosis	3.6	4.0	3.51

Table 1 - Basic statistics of mean maximum (T_{max}) and mean minimum temperature (T_{min}) data (Celsius degrees) for June and elevation data (m a.s.l.).

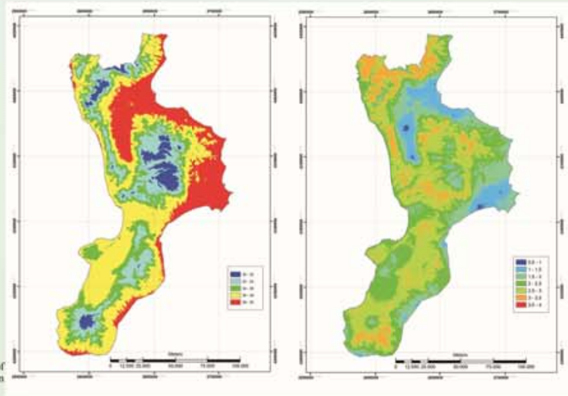

Fig. 3 - Mean and standard deviation maps of simulations for the mean maximum temperature (T_{max}).

Figure 2 and Figure 3 present a way to treat the co-simulated images of the two variables jointly, by calculating the mean and the standard deviation, respectively, of the 500 simulations at each grid node and then mapping the results for each variable. The maps of the mean show the complexity in spatial distribution of temperatures. The maps of the standard deviation, obtained by post-processing of simulations, have allowed to assess the uncertainties of non-Gaussian variables and to overcome the drawback of kriging variance of its independence from actual sample values. From a visual inspection, it shows clearly how the uncertainty distributions of temperature are mostly related to the density of the sample data (Figure 1). The 500 realizations were used as input for ET_0 model and then the maps of the mean and standard deviation of ET_0 (Figure 4) were obtained in a similar way to the maps of Figure 2 and Figure 3.

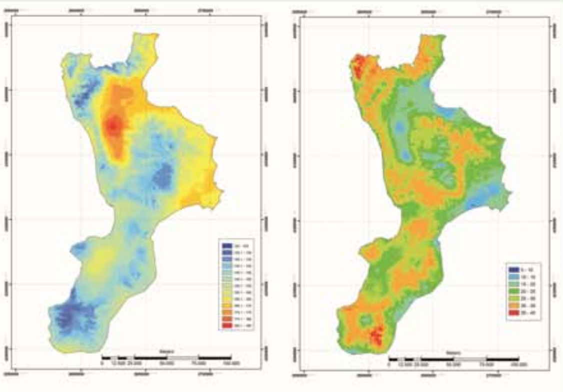

Fig. 4 - Mean and standard deviation maps of simulations for Et_0.

A visual inspection shows clearly where the uncertainty in ET_0 is high. The Monte Carlo simulation has shown how the previous uncertainties in input variables can affect the predictions of ET_0 model. The objective, however, was not to validate the model or assess the errors associated with the model type and coefficients, but rather to evaluate how the variability of inputs affects uncertainty of model prediction. This approach has demonstrated that it is possible to produce maps of uncertainty, which are more useful than the simple extrapolations of the point estimation. Of course, it is important to evaluate how well the model approximates reality, i.e. model uncertainty. If the model has high uncertainty, a difference in model output may not indicate a real change and could thus be meaningless. Therefore, it is essential to know model uncertainty for operational decision-making. Looking at the standard deviation map of ET_0 (Figure 4), only a weak spatial pattern can be distinguished; the standard deviation do not appear correlated with the estimates of ET_0, but with the density of weather stations.

5. CONCLUSIONS

The results of a spatial uncertainty analysis have shown that the prediction quality depends on the uncertainties of the data used in the analysis; therefore map makers should convey the accuracy of the maps they produce (Heuvelink, 1998). A complete characterisation of the accuracy of spatial data should also include the spatial correlation of the attributes used for estimation and stored in a GIS. In the past, a single root mean squared error was sufficient to assess spatial accuracy, but now it is no longer sufficient and much more information should be provided to characterise the quality of a map.

Finally, it is worth pointing out the consequences of estimation uncertainty in the context of decision-making and water resources management. There is currently some reluctance to perform error recognition, possibly because of the greater analysis required. However, studying uncertainty leads to increased understanding of the roles played by the different input parameters, which also allows to evaluate the relative costs and benefits of using different scenarios.