

Bayesian models of thermal and pluviometric time series in the Fucino plateau

Adriana Trabucco,¹ Federico M. Stefanini²

¹Regione Abruzzo, Agenzia Regionale per i Servizi di Sviluppo Agricolo, Vasto; ²Dipartimento di Statistica G. Parenti, Università di Firenze, Italy

Abstract

This work was developed within the Project *Metodologie e sistemi integrati per la qualificazione di produzioni orticole del Fucino* (Methodologies and integrated systems for the quality improvement of horticultural products in the Fucino plateau), sponsored by the Italian Ministry of Education, University and Research, Strategic Projects, Law 448/97. Agro-system managing, especially if necessary to achieve high quality in speciality crops, requires knowledge of main features and intrinsic variability of climate. Statistical models may properly summarize the structure existing behind the observed variability, furthermore they may support the agronomic manager by providing the probability that meteorological events happen in a time window of interest. More than 30 years of daily values collected in four sites located on the Fucino plateau, Abruzzo region, Italy, were studied by fitting Bayesian generalized linear models to air temperature maximum/minimum and rainfall time series. Bayesian predictive distributions of climate variables supporting decision-making processes were calculated at different timescales, 5-days for temperatures and 10-days for rainfall, both to reduce computational efforts and to simplify statistical model assumptions. Technicians and field operators, even with limited statistical training, may exploit the model output by inspecting graphs and climatic profiles of the cultivated areas during decision-making processes. Realizations taken from predictive distributions may also be used as input for agro-ecological models (e.g. models of crop growth, water balance). Fitted models may be exploited to monitor climatic changes and to revise climatic profiles of interest areas,

periodically updating the probability distributions of target climatic variables. For the sake of brevity, the description of results is limited to just one of the four sites, and results for all other sites are available as supplementary information.

Introduction

This work was developed within the *Metodologie e sistemi integrati per la qualificazione di produzioni orticole del Fucino* (Methodologies and integrated systems for the quality improvement of horticultural products in the Fucino plateau), Italian Ministry of Education, University and Research, Strategic Projects, Law 449/97. The Fucino (Abruzzo region, Italy), was an old Apennines lake, that because of its shallowness used to show extensive marsh areas. It was definitively drained during the second half of 1800, to increase the availability of areas usable for agricultural purposes. The drainage possibly caused significant climatic changes over the Fucino basin and on the surrounding region (Tomassetti *et al.*, 2003). Today it is an area of about 13,000 ha, on a plateau above 700 m a.s.l., completely flat and flanked by peaks reaching even 2500 m a.s.l., where farms cultivate mainly open-field vegetable crops such as potatoes, carrots, radicchioes, fennels, cauliflowers. The main project objective was to improve the quality of products to meet the consumers needs, and checking the possibility of introducing new crops. In this area, climate features, suitability of soil, depth to which the groundwater is maintained in the reclaimed area and natural resource management are main factors that largely control and limit the agricultural choices. The research developed in the last 30-40 years in the agricultural field, confirms that the interdisciplinary approach and the availability of tools for decision support are basic to correctly handle the complexity of agro-systems. The improvement of knowledge base within plant physiology and agricultural meteorology closely related to the evolution of computer technologies, led to development of mathematical descriptions of plant growth and soil water balance and to build agronomic models (crop models) and decision support systems. To incorporate a range of environmental concerns, and to calculate effects of tillage systems and other management practices, functions for nutrient cycling, soil carbon dynamics, have been also developed. The research in this area, also confirms that the availability of climatic information is critical for the effectiveness of decision support systems (Steiner *et al.*, 2008). Many methods to characterize climate, with the main purpose to identify key indices to summarize climatic resources and the potential ability to sustain animal and plant life of different areas of the Planet, were developed. Despite the undoubted validity of these methods (the first, developed by Köppen and Geiger in 1928, remains in use today), due to their summarizing goal, they are not also so effective in supporting agronomic decision making, whereas uncertainties of those summaries must be taken into account. The climatic characterization of a territory is based on the observation and the systematic collecting of data on atmospheric conditions for long periods of time, which cor-

Correspondence: Dr. Federico M. Stefanini, Dipartimento di Statistica "G. Parenti", Università di Firenze, viale Morgagni 59, 50134 Firenze, Italy. Tel. +39.055.4237211 - Fax: +39.055.4223560. E-mail: stefanini@ds.unifi.it

Key words: agroclimatology, Bayesian models, Fucino plateau.

Acknowledgements: the authors would like to thank ARSSA for providing data from the Historical Meteorological Data Bank of the Abruzzo Region. The Digital Elevation Model is provided by Services for Territorial and Telematic Information, Aquila, Abruzzo region. The authors would also thank Emanuele Bonfitto for technical insights on agricultural decision making.

Received for publication: 20 January 2011.

Accepted for publication: 10 August 2011.

©Copyright A. Trabucco and F.M. Stefanini, 2011

Licensee PAGEPress, Italy

Italian Journal of Agronomy 2011; 6:e29

doi:10.4081/ija.2011.e29

This work is licensed under a Creative Commons Attribution NonCommercial 3.0 License (CC BY-NC 3.0).

responds to at least 30 years, as established by the World Meteorological Organization (WMO). Measurements on the meteorological variables, such as air temperature, rain amount and wind speed, are unlikely to take the same values over the years, even considering the same place and the same season of the year, therefore the analysis of these variables necessarily has to face the inherent uncertainty associated to them. The presence of uncertainty is well known by climatologists, among which we mention one of the fathers of Italian agroclimatology, the physicist Ezio Rosini, who strongly supported the substitution of pseudo-deterministic approaches based on point estimates with an approach based on intervals at a fiducial level considered acceptable for the specified goals to reach (Rosini, 1988). The specialized literature has long since agreed on the necessity of using probabilistic models to characterize climatologic variables, for example to find the structural components describing the cyclic regular behaviour (seasons, years) of time series of temperatures even if buried into observed data affected by noise. From a methodological point of view, statistical models have been developed for climatological variables ranging from ARIMA and state-space models (Sneyers, 1990; Brockwell and Davies, 1998) to hidden Markov (Bellone *et al.*, 2000) and Poisson processes (Cowpertwait, 1994). A model may be more or less useful depending on how much it captures about structure of the phenomenon under study, but there is almost never a *true* model to search for. Therefore the challenge is about the development of a statistical model which is simple enough to be used by agronomic managers without too much statistical training but rich enough and structured to catch main climatic features involved into decision making. Given that Bayesian methods are particularly suitable for forecasting and decision making (Bernardo and Smith, 1994; Lindley, 2000), in this paper some Bayesian models are developed to characterize rainfall and temperature in the Fucino basin.

Materials and Methods

In this section data collection and statistical models are described. Time series, whose main geographic features are summarized in Table 1, were collected in four sites situated inside the Fucino plateau, that is the agricultural area involved by the project reported in the introduction. Years from 1951 to 2007 were considered, although some series are shorter due to some years of delay in starting to collect data. At each meteorological station, minimum/maximum temperatures and rainfall were collected on a daily basis. The choice of aggregation data on a appropriate time scale was performed to fulfil different requirements. The aggregation of data into intervals may be performed by the agricultural manager provided that the interval is appropriate and reasonable to create relationships between meteorological variables and the main aspects of crop management. Also, a reasonable degree of aggregation makes computation and model assumptions lighter, for example for the possibility of exploiting the Central Limit Theorem. For these reasons, Bayesian generalized linear models were fitted to time series aggregated into 5-days for temperatures (pentadal averages) and 10-days for rainfall (dekadal amount).

Table 1. Main features of locations where time series were recorded.

Sites	Longitude (UTM 32)	Latitude (UTM 32)	Height (a.s.l.)	Thermometry (years of survey)	Pluviometry (years of survey)
San Benedetto dei Marsi (province of L'Aquila)	4663436.10	882915.42	687	56	55
Borgottomila (province of L'Aquila)	4659284.07	877795.62	650	56	59
Ortucchio (province of L'Aquila)	4655086.98	866073.07	680	31	32
Avezzano (province of L'Aquila)	4662521.71	866529.10	698	33	34

Air temperature

Model

Regarding temperatures, the pentadal mean value y_t at time t is explained by a cyclic component (seasonal component) that captures mean recurrence of temperature and an autoregressive component made by the observation performed one or more time units back, that is y_{t-1}, y_{t-2}, \dots , so that shifts of temperature due to prolonged external events (for example rain) on few time units are taken into account. The class of models we considered is defined by the formula below:

$$y_t = \mu + \sum_{k=1}^K \{ \beta_k \sin(k2\pi t / T) + \alpha_k \cos(k2\pi t / T) \} + \sum_{d=1}^D \{ \delta_d (y_{t-d} - \mu) \} + \varepsilon_t \quad (1)$$

where ε_t are identically and independently distributed normal errors with mean 0 and variance σ^2 ; μ is the overall mean, β_k and α_k are coefficients of the trigonometric representation of the cyclic component and δ_d are the auto-regressive terms; T is the period length of the cyclic component. Model search, that is optimal values of K and D , was performed by minimizing the Bayesian Information Criterion (often called BIC). Model checking was performed by graphically inspecting model residuals (McCullagh and Nelder, 1989). The initial distribution of the above model parameters was defined to be weakly informative, that is a zero mean overdispersed normal distribution for parameters in the linear predictor and an inverse gamma distribution for the variance of the error term, with parameters both equal to 0.01. The final joint distribution of model parameters was approximated by Monte Carlo simulation (Bernardo and Smith, 1994).

Rain occurrence and rainfall amount

Model

Rainfall was modelled by decomposing the time series into a binary variable Z representing the dekadal occurrence of precipitation and a dekadal intensity of precipitation X which is identically null if Z is null and is strictly positive if Z is equal to one (it does rain). A logit link function was adopted for the occurrence variable, thus the expected value after transformation is explained by a linear model, here in matrix form:

$$\ln(E[Z_t] / (1 - E[Z_t])) = W_t \gamma \quad (2)$$

with W_t the design matrix and γ the vector of parameters. Similarly, the natural logarithm of the expected value of X is modelled by a linear model:

$$\ln(E[X_t]) = B_t \psi \quad (3)$$

with B_t the design matrix and ψ the vector of parameters. The two design matrices contains trigonometric expansions of the cyclic (seasonal) component and autoregressive terms. For the sake of brevity, the class of models based on (2) and (3) are not further detailed, but a

comprehensive account on generalized linear models is given by McCullagh and Nelder (1989), while Stern and Coe (1984) provide a full account on how to define the likelihood function for rainfall time series. Weakly informative prior distributions were selected for the above models. Note that in equations (2) and (3) explicit conditioning is omitted to simplify notation. The model output for dekadal rainfall and pentadal temperatures includes the predictive distribution of variables over a time window of interest, for example one year. Such output is used both to check model predictions within an observed year and to help in decision making for the next unobserved month. All the computations were performed in the open source R environment (R Development Core Team, 2005) and tailored functions were developed to make possible batch processing.

Results and Discussion

The best model for each variable was obtained by BIC optimization. Top scored models always contain autoregressive terms, therefore the information about an event observed in a given time interval is also informative for subsequent time intervals, a result which is not straightforward to state without model search and fitting. These model components exert a relevant effect while performing short-to-medium forecasting of rainfall and temperatures: uncertainty increases moving from observed time intervals towards time intervals far in the future. Thus, the models here presented are best suited for short term forecasting or to compare seasonal components belonging to time series of different locations so to characterize structural agro-climatic features of different fields. The posterior distributions of model parameters and the Bayesian predictive distributions on selected pentadal/dekadal time intervals have been obtained for all the climatological variables at each location. For the sake of brevity, we decided to present results of just one site, called Borgottomila, and to select the worst fitted years (2005 for temperatures and 2007 for rainfall) so to make clear the operating limits that may result from model fitting in a relatively simple class of models. Results are followed by a discussion on basic climatic features of the Fucino plateau, that are also evident on the chosen site. Tables of results for all other sites are available as supplementary information (from Table 3 to Table 23).

Air temperature

In Figures 1 and 2 the central continuous line represents the expected value provided by the model for Borgottomila site, and the two dashed lines represent the estimated 95% Bayesian credibility interval, thus the probability of observing a pentadal mean value outside the dashed interval, is 0.05. The figures also show, displayed with dotted lines, pentadal means of minimum and maximum temperatures (respectively T_{min} and T_{max}) recorded at the same site in a unusual year (2005).

In general, values of minimum air temperature show particularly cold weather conditions from mid-November to end of March and cool temperatures during summer night. The thermal minimum oscillations of pentadal means are comprised in a range of $\pm 4^{\circ}\text{C}$ with a credibility level of 95%. Year 2005 was characterized by extraordinary snowing and persistent freezing conditions at the end of winter, an event that actually involved the whole Italian peninsula. Figures 1 and 2 highlight that at least two persistent cold episodes characterized the winter of 2005 from the end of January to mid-March and that they lasted about 15 days each; the pentadal mean values of the minimum temperature decreased below the expected values till reaching $-8/-10^{\circ}\text{C}$; those values are outside the credibility interval or close to the edge of the chosen credibility interval (level 95%). Memorable cold winters are present in the time series: at Borgottomila site, in the third pentad of February 1956, the mean of min-

imum temperatures reached -21°C with an absolute minimum below -30°C . The pentadal maximum temperature is at the lowest value in the month of January, therefore this is the coldest month of the year. The hottest pentad, with expected values around 30°C , is reached in the period between 15th of July and 15th of August. The pentadal thermal oscillation is in a range of about $\pm 6^{\circ}\text{C}$. In the end of winter 2005, due to the previously described episodes, pentadal values of maximum temperatures were near or under the lowest 2.5% credibility limit, and in the coldest pentad five consecutive ice days occurred.

Probability of frost and of exceeding some thermic thresholds of agricultural interest

Agronomic managers may exploit the output in several ways, for example model predictions may be summarized on other coarser time scales (e.g., 10 days) or they can be conditioned with respect to events representing a scenario of interest (what-if analysis), which the decision maker needs to address to. As life of the different species, phenological phases of plants and the epidemiology of their parasites, are regulated by optimal and critical thermal thresholds, agronomic manager may also be supported by the estimate of probability of meaningful threshold events. For example, the appropriate date for an agronomic practice, as seeding or transplanting, may be decided relying on model predictions after selecting a sustainable risk of failure, which

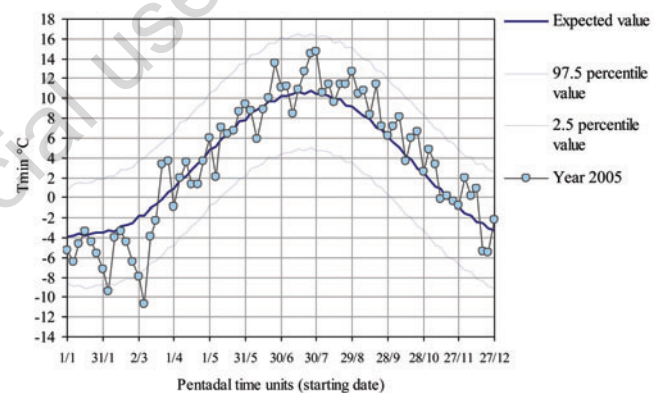


Figure 1. Borgottomila site, minimum temperatures. Bayesian predictive distribution of pentadal averages and observed values of year 2005.

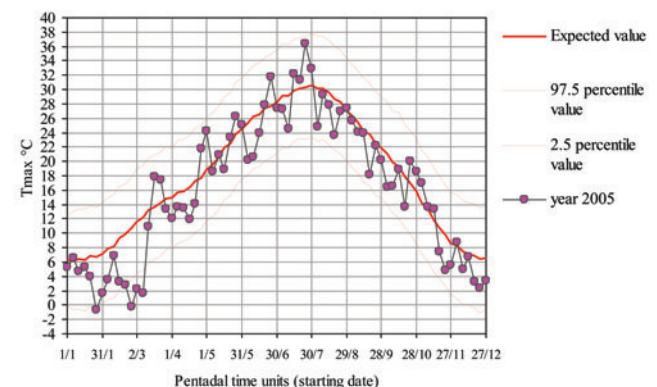


Figure 2. Borgottomila site, maximum air temperatures. Bayesian predictive distribution of pentadal averages and observed values of year 2005.

coincides with the probability that temperature drops under the thermal threshold required. In Figures 3 and 4, probability values of some useful threshold events, are reported. Among the most interesting risks for agricultural activity in this area, the spring temperature range is of particular importance therefore in Figure 5 an overview on Fucino thermometry around the entrance of the spring season is shown. From comparative analyses of the maximum and minimum temperatures at some thermal thresholds, like $T_{max} > 10^{\circ}\text{C}$ and $T_{min} < 0^{\circ}\text{C}$, it can be inferred that at entrance of the spring season in all sites analyzed the maximum temperature is stabilizing above 10°C (an effective feature in promoting vegetative work) but there is still a substantial risk of late frost, with probability of such event above 50%. At all sites, as shown in Figure 3 at Borgottomila, it is necessary to wait until the end of April to infer that the freeze probability approaches zero, although intense cold with persistence in 5 consecutive days may return until the following month, even if with low probability (less than 5%).

Rainfall amount

In the study of the pluviometric regime of sites, rainfall amount has been analysed, with probabilistic dekadal scale models. The peculiar distribution of the rainfall variable makes it not very useful to summarize dekadal values in terms of mean intensity: the presence of dekadal time units with null intensity is obscured by the positive intensity of the same time interval in different years. After model fitting, annual, seasonal and monthly estimates were obtained and results used to

characterize the annual rainfall trend of sites. The rainfall amount has been characterized by some quantiles of the predictive distribution: 5%, 25%, 50%, 75% and 95%. The median (quantile 50%) summarizes the central location of the distribution, while its variability is captured by extreme quantiles or by differences of quantiles, for example 5% and 95%. The rainfall recorded in the 2007 at the mentioned site was considered with the purpose of showing how the Bayesian model performs in one of the driest years of the time series. Borgottomila site was chosen to summarize basic rainfall features common to all the Fucino plateau (Table 2, Figure 6). The annual rainfall amount, with expected values between 620-700 mm, situates the sites of plateau among the dry ones of the Abruzzo region. Spring and summer seasons are characterized by a small rain supply while autumn and winter seasons assure more than 60% of the yearly rainfall.

The driest months are July and August, while the rainiest ones range from October to December. On month time scale (Figure 6), the graphic displays pronounced asymmetries above the expected values with quite large range intervals. The highest rainfall amount values are more uncertain and highly dependent on the realization of the above mentioned extreme events.

Rainfall occurrence and probability of exceeding some thresholds of agricultural interest

The analysis of rainfall occurrence (at least one rain event in the considered time interval) and probability values of exceeding some

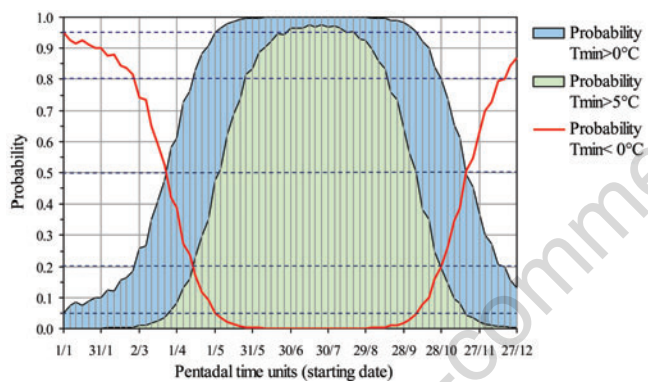


Figure 3. Borgottomila site, minimum temperatures. Probability of freezing and of exceeding some critical thresholds of agricultural interest.

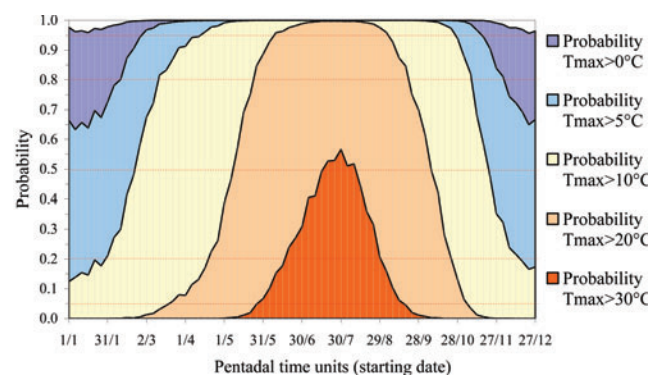


Figure 4. Borgottomila site, maximum temperatures. Probability of exceeding some thermal thresholds of agricultural interest.

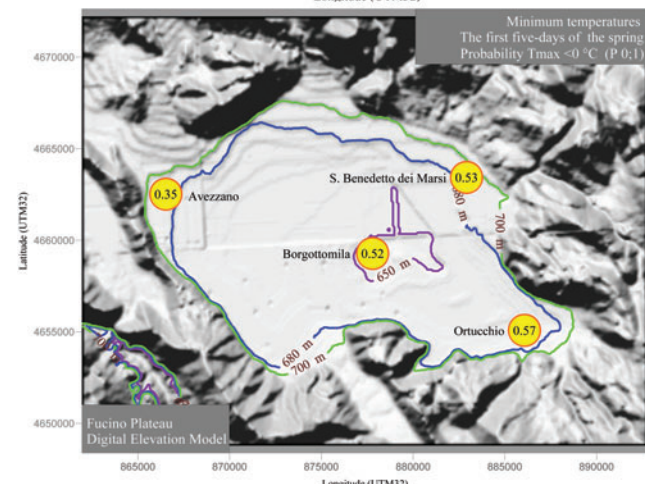
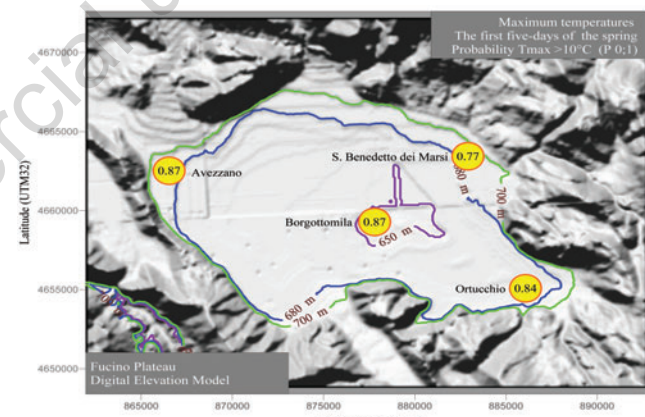


Figure 5. Fucino plateau at the entrance of spring season. Probability of exceeding a thermal threshold of agricultural interest and late frost probability, at the four sites analyzed.

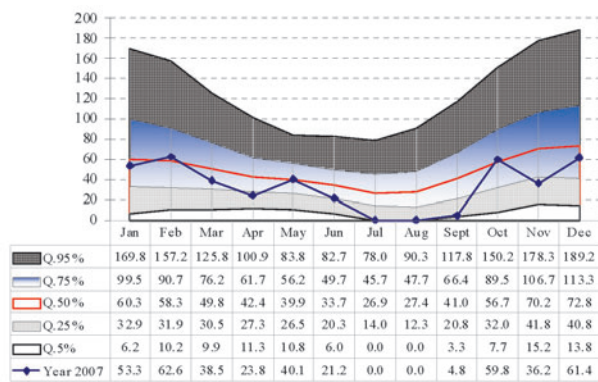


Figure 6. Analysis of monthly rainfall amount (millimetres) probability at Borgottomila site. Quantiles of the predictive distribution are shown by column. In the last line from top, the observed values of year 2007 are shown.

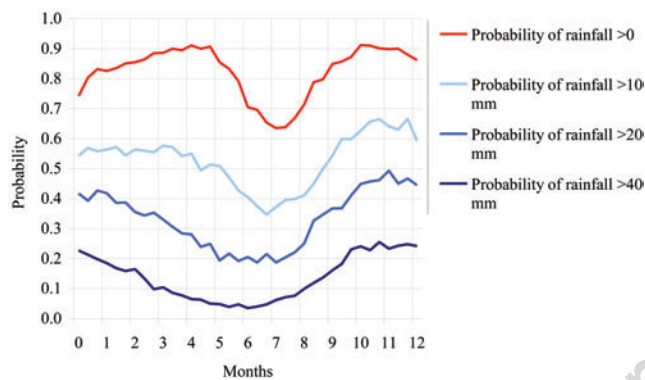


Figure 7. Probability of rainfall (at least one rain event) and of exceeding some rainfall amount of agricultural in Borgottomila site.

thresholds (10 mm, 20 mm, 40 mm) of interest on dekadal scale have been analysed. These thresholds, in general, have different uses depending on periods of the year and the subject of interest. For example, in assessing the amount of rainfall available for crop requirement in the summer dekads, the low probability associated with the threshold above 40 mm makes it of little interest. On the other hand, an in-depth study of tail area probability values could be useful elsewhere, like for flood safety in the field of reclaim, where the focus is shifted principally on precipitation episodes of very high intensity (extreme events), that is on rare events. For this type of estimates we recommend specifically tailored statistical models (Coles, 2001). At all the studied sites, for example Borgottomila (Figure 7), occurrence of precipitation (at least one rain event) is an event with probability above

60% in all the dekads, with peaks of 90% in some months such as April, May, November, December.

Appreciable rainfall amount, from 20 mm to 40 mm, has a smaller probability value in all the dekads. The exceeding of the first threshold (20 mm) is an event with probability value about 30% to 45% in autumn and winter and about 20% in warmer seasons, where probability values decrease under 10% for the 40 mm threshold. The selected time scale aggregation may lead to a useful analysis of water requirements for several cultivated species and to the optimization of decisions regarding water resource use for irrigation. Fitted models may be used to track the onset of sufficient rain to carry out sowing / transplanting operations given the joint probability of two or more events, for example the event *first dekad with at least 25 mm of rainfall, the second dekad summing to at least 20 mm and the third dekad summing to at least 20 mm*. During warmer months, where the evapotranspiration flux reaches highest values and the probability of an appreciable rainfall (at least 20-40 mm for dekad) is indeed very low, most of cultivated species should be efficiently supported by adequate irrigation. Water balance models exploiting as input the expected values of precipitations and temperatures, together with their Bayesian credibility intervals, may suitably support irrigation planning of crops by taking care of the inter annual natural variability of climatological processes.

Conclusions

Bayesian statistical models developed in this work allow not only to characterize the climatic variables here considered on a suitable aggregated scale, pentadal or dekadal time scales, but above all to describe in quantitative terms the structure and the intrinsic variability of time series. Short term forecasting allowed to evaluate the risk of climate events which are relevant in agricultural decision making by accounting for the inherent uncertainty. Summaries based on the inferred cyclic components make possible the climatic comparison of different locations, for example during enterprise activity planning and investments, and the support to decision making by short time forecasting. Most notably, all quantities of interest are related to suitable measures of uncertainty, so that consequences due to uncertain previsions or overstated accuracy may be promptly considered. Indeed limitations of the developed models rest in the class we considered, that is generalized linear models (McCullagh and Nelder, 1989) which are conceptually related to linear regression. While this choice makes the model fitting and output understandable with limited statistical training, it provides a quite stiff class of models for detailed predictions. A natural extension could include the possible departure of the cyclic component from the average behaviour, for example by mean of a hierarchical model, at the price of heavier computations. Another approach widely considered in the literature is the class of statespace models (Brockwell and Davies, 1998, chapter 12), which are able to make estimated parameters local in time: here rainfall is more difficult than tem-

Table 2. Quantiles of the predictive distribution of annual and seasonal rainfall amount at Borgottomila site. On the last column the observed values of year 2007 are shown.

Quantiles of rainfall amount	Q. 5%	Q. 25%	Q. 50%	Q. 75%	Q. 95%	Observed values (year 2007)
mm/year	474.4	592.5	679.1	883.1	936.6	401.8
mm/winter season	79.1	138.2	182.1	239.6	348.0	151.6
mm/spring season	64.2	96.7	122.4	155.5	202.6	93.5
mm/summer season	34.6	75.8	105.5	143.9	212.3	3.4
mm/autumn season	101.2	162.5	219.7	279.9	402.1	153.2

Q, quantiles.

perature to fit. Statespace models are also suited for the joint study of spatial variability and temporal dynamic, and they may incorporate rich sources of information, like radar records located in the surroundings of the weather stations. The case study we described showed that, even in the worst fitted year, our models are able to support agricultural decision making, and they require a limited degree of statistical training. Among the remarkable potential extensions rooted into a probabilistic statistical paradigm we mention dynamic Bayesian networks because they might become the entry point towards the creation of statistical experts systems for regional agriculture.

References

- Bellone E., Hughes J.P., Guttorp P., 2000. A hidden Markov model for downscaling synoptic atmospheric patterns to precipitation amounts. *Clim. Res.* 15:1-12.
- Bernardo J.M., Smith A.F.M., 1994. *Bayesian theory*. Wiley, New York, NY, USA.
- Brockwell P.J., Davies R.A., 1998. *Time Series: Theory and Methods*. Springer Verlag, New York, NY, USA.
- Coles S., 2001. *An Introduction to Statistical Modelling of Extreme Values*. Springer Verlag, London, UK.
- Cowpertwait P., 1994. A generalized point process model of rainfall. *P. R. Soc. London Series A* 447:23-37.
- Lindley D., 2000. The philosophy of statistics. *Statistician* 49:293-337.
- McCullagh P., Nelder J.A., 1989. *Generalised linear models*. Chapman & Hall, New York, NY, USA.
- R Development Core Team, 2005. Available from: <http://www.R-project.org>
- Rosini E., 1988. *Introduzione all'agroclimatologia*, vol. 1, 2, 3. ERSASMR Ed., Bologna, Italy.
- Sneyers R., 1990. On the statistical analysis of series of observations. *Tech. Note 143, WMO No. 415*, Geneva, Switzerland.
- Steiner J.L., Hatfield J.L., 2008. Winds of change: a century of agroclimate research, celebrate the centennial. *Agron. J.* 100:(Suppl.3):132-152.
- Stern R.D., Coe R., 1984. A model fitting analysis of daily rainfall data. *J. R. Stat. Soc. Ser. A* 147:1-34.
- Tomassetti B., Giorgi F., Verdecchia M., Visconti G., 2003. Regional model simulation of the hydrometeorological effects of the Fucino Lake on the surrounding region. *Ann. Geophys.* 21:2219-2232.