ASSESSMENT OF BAYESIAN MODELS FOR RAINFALL FIELD RECONSTRUCTION

Francesca Bruno, Fedele Greco, Elena Scardovi

Abstract We develop a Bayesian hierarchical model for predicting rain occurrence and amount on a fine-pixel grid, by modelling the relationship between rain gauge and radar data. An enrichment consists in including neighbourhood information as a covariate for dealing with spatial misalignment. A focus of this work is on evaluation of the predictions, which can consist in point or probabilistic forecasts. In the first case, competing models can be assessed and ranked on the basis of consistent scoring functions. On the other hand, probabilistic forecasts consist in full predictive distributions and ought to be calibrated and sharp. Suitable adaptation to the zero-inflated model of standard tools for evaluation are applied to the prediction of hourly rainfall in Emilia-Romagna region.

Abstract *In questo lavoro viene presentato un modello gerarchico Bayesiano per la previsione della presenza e della quantitá di pioggia, che considera la relazione tra misurazioni da pluviometro e da radar; informazioni spaziali intorno al sito pluviometro permettono di correggere il misallineamento. Le previsioni sono ottenute in forma probabilistica, fornendo l'intera distribuzione predittiva, e sono riassunte, tramite l'applicazione di un funzionale, in previsioni puntali. Strumenti per la valutazione e il confronto tra modelli zero-inflated sono proposti e applicati alla previsione di pioggia oraria in Emilia-Romagna.*

Key words: Hierarchical Bayesian model, zero-inflated distribution, calibration, consistent scoring functions, proper scoring rules

---

Francesca Bruno
Department of Statistical Sciences "Paolo Fortunati" - University of Bologna, Via delle Belle Arti, 41 - Bologna, e-mail: francesca.bruno@unibo.it

Fedele Greco
Department of Statistical Sciences "Paolo Fortunati" - University of Bologna, Via delle Belle Arti, 41 - Bologna, e-mail: fedele.greco@unibo.it

Elena Scardovi
Department of Statistical Sciences "Paolo Fortunati" - University of Bologna, Via delle Belle Arti, 41 - Bologna, e-mail: elena.scardovi2@unibo.it

1 Motivating example: Emilia-Romagna rainfall data

Rainfall measurements are essential for public authorities, being the basis for hydrological models and risk monitoring; knowledge of rainfall amounts with high spatial resolution can be useful for water resource planning and management. Direct measurements are provided by rain gauges in sparsely distributed locations. Radar data are available on fine-pixel grids, thus overcoming the problem of sparseness of the rain gauge network. However they consist in indirect measurements and are not reliable for the assessment of rain amounts, despite accurate pre-processing performed by removing systematical and occasional biases. We focus on hourly rainfall data in Emilia-Romagna Region, in Italy. Radar data consist of hourly rainfall maps with 1x1 km grid cell resolution, within a circle of 125 km radius, centered near Bologna; in this area, about 300 rain gauges are available. The data are heterogeneous in space and time, and many zero values are present, corresponding to dry hours. Table 1 reports summaries of 8 rain events of September-October 2010, analysed in a previous paper ([1]), characterised by different duration and features.

Table 1 Descriptive statistics of 8 rainfall events in September-October 2010: event ID (Event), number of hours characterizing the event (H.rs), percentage of zero measurements (Zeros), first (Q1), second (Med) and third (Q3) quartiles of positive amounts, maximum (Max) and correlation between rain gauge and radar measurements (Corr)

Event	H.rs	Zeros	Q1	Med	Q3	Max	Corr
E1	6	24	0.4	1.0	2.4	19.4	0.81
E2	9	49	0.6	2.0	4.4	31.2	0.80
E3	10	30	0.4	0.8	2.2	34.4	0.55
E4	6	49	0.6	1.8	3.8	27.0	0.59
E5	7	53	0.2	0.8	3.4	37.2	0.76
E6	5	61	0.2	0.6	1.8	12.4	0.79
E7	11	67	0.2	0.8	2.0	15.0	0.46
E8	16	43	0.4	0.8	1.6	10.8	0.43

2 Model specification for spatial prediction

A flexible choice for dealing with the remarkable quantity of zero measurements is proposed in ([8]), consisting in a zero-inflated model:

$$p(Y_{st}|R_{st},\pi_{st}) = \pi_{st} I_{Y_{st}=0} + (1-\pi_{st}) p(X_{st}) I_{Y_{st}>0} \quad s \in S_G \quad t = 1,\dots,T$$

where π_{st}, Y_{st} and R_{st} are the probability of zero and the rain gauge and radar measurements at location s and time t respectively, $p(X_{st}) = p(Y_{st}|Y_{st} > 0)$, and S_G is the set of rain gauge locations (when dealing with radar, the pixel containing the location s is considered in this first formulation).

Both rain occurrence and rain accumulation are modeled by exploiting radar as a covariate, and the spatial information is captured by normal random effects whose correlation decreases exponentially with the euclidean distance d between locations. Previous work ([1]) encourages separate modeling of different hours and suggests a slight preference for the gamma distribution over the competing lognormal for modeling positive rain amounts. This is our "base" model:

•Rain occurrence: $$\text{probit}(1-\pi_{st}) = \gamma_{1t} + \gamma_{2t}\log(R_{st}) + \varepsilon_{st} \tag{1}$$

$$\text{where} \quad \varepsilon_t|\sigma^2_{\varepsilon t}, \phi_\varepsilon \sim MVN(\mathbf{0}, \sigma^2_{\varepsilon t}\Sigma_\varepsilon) \quad \text{and} \quad \Sigma_\varepsilon(s,s') = \exp(-\phi_\varepsilon\, d_{ss'}) \tag{2}$$

• Rain accumulation:

$$X_{st}|\mu_{st}, \tau_t \sim Gamma(\tau_t, \tau_t/\mu_{st}) \quad \text{where} \quad \log\mu_{st} = \beta_{1t} + \beta_{2t}\log(R_{st}) + \alpha_{st} \tag{3}$$

$$\alpha_t|\sigma^2_{\alpha t}, \phi_\alpha \sim MVN(\mathbf{0}, \sigma^2_{\alpha t}\Sigma_\alpha) \quad \text{and} \quad \Sigma_\alpha(s,s') = \exp(-\phi_\alpha\, d_{ss'}). \tag{4}$$

Noninformative hyperpriors are exploited for hyperparameters.

In order to enrich the model and to address misalignment we consider the inclusion of the mean $R_{\bar{s}t}$ of the radar pixels around the rain gauge location $s \in S_G$ as a further covariate when modelling both rain occurrence and rain accumulation; moreover, as suggested in ([8]), we added an indicator function detecting when the radar pixel containing location s measures less then 0.2 mm (equivalent to rain gauge sensibility). In the following, we use the denomination "neigh" for referring to the model in which equations (1) and (3) are modified in the following way:

$$\text{probit}(1-\pi_{st}) = \gamma_{1t} + \gamma_{2t}\log(R_{st}) + \gamma_{3t}\log(R_{\bar{s}t}) + \gamma_{4t}I_{\{R_{st}>0.2\}} + \varepsilon_{st} \tag{5}$$

$$\log\mu_{st} = \beta_{1t} + \beta_{2t}\log(R_{st}) + \beta_{3t}\log(R_{\bar{s}t}) + \beta_{4t}I_{\{R_{st}>0.2\}} + \alpha_{st}. \tag{6}$$

Parameter estimation is performed through Markov chain Monte Carlo algorithms, implemented in C for real-time rainfall field reconstruction.

3 Evaluation of the predictions

Point forecasts are required in many practical situations. We evaluate and rank our predictions from models "base" and "neigh" according to the mean absolute error and the mean squared error, which are consistent with respect to the median and mean respectively (see [4]).
Predictive intervals are often associated with point forecasts for providing information about their uncertainty. If we are looking for centered credibility intervals, when their left extremes coincide with zero, precautions for calculating the coverage are needed.

To realize their full potential, forecasts ought to be probabilistic in nature, taking the form of probability distributions ([3]). When probabilistic forecasts are available, the assessment of their adequacy should address calibration and sharpness. The former refers to the statistical consistency between the probabilistic forecasts and the observations; the latter corresponds to the concentration of the predictive distributions and is a property of the forecasts only. The more concentrated the predictive distributions, the sharper the forecasts, and the sharper the better, subject to calibration (see [5]).
The most common tool for assessing probabilistic calibration is the histogram of the Probability Integral Transform (PIT). In the case of precipitation, the distribution is zero-inflated, thus requiring a correction in correspondence to zero observations. Literature proposes two alternative approaches: the predictive CDF F can be randomized in zero by multiplying it for a standard uniform variable, or a non-randomized method can be adopted, as introduced in [2] for the case of count data. The second approach deals with discrete data and is thus not directly applicable in our case, due to the continuous nature of the distribution on positive amounts; in this work we propose an adjustment for zero-inflated distributions with continuous modeling of the positive values. The procedure consists in:

1 fixing a number J of bins in which [0,1] will be divided
2 calculating the CDF of the PIT, conditioning on the observed value, in $u_j = j/J,\ j = 1,\dots,J$

$$F^{PIT}(u_j|y=0) = \begin{cases} 0 & u \le \pi \\ u/\pi & 0 \le u \le \pi \\ 1 & u \ge \pi \end{cases} \qquad F^{PIT}(u_j|y, y \neq 0) = \begin{cases} 0 & u < F(y) \\ 1 & u \ge F(y) \end{cases} \tag{7}$$

(we dropped the subscripts s and t from y, π and F for simplicity of notation)
3 calculating the mean $\bar{F}^{PIT}(u_j)$ over the observations $y_1,\dots,y_n$ for each $u_j = j/J,\ j = 1,\dots,J$;
4 computing $f_j = \bar{F}^{PIT}(u_j) - \bar{F}^{PIT}(u_{j-1}),\ j = 1,\dots,J$;
5 drawing a PIT histogram with heights f_j.

If the distribution of all the potential predictive values is considered, aggregating over the individuals, a check for marginal calibration can be done, verifying its statistical compatibility with the empirical distribution of all the observations ([5]).

Summary measures of predictive performance addressing calibration and sharpness simultaneously are provided by scoring rules. They consist in functions $S(F^{pred}, y^{obs})$ of the predictive distribution F^{pred} and of the observation y^{obs} quantifying a penalty the forecaster aims to minimize, and they should be proper ([5]). Comparisons are made on the means of the scoring rule values over the observations. The Brier Score (BS) is a useful and proper scoring rule when the interest is on the exceedance of a certain threshold; in particular, we focus on the detection of rainfall occurrence (i.e. zero threshold). The continuous ranked probability score (CRPS) is the integral of

the BS over the thresholds, and is a wide used proper scoring rule; [6] provides a useful formulation for computing it when a big sample from the predictive distribution is available, as happens with Markov chain Monte Carlo outputs.

4 Assessment and comparison between two model specifications

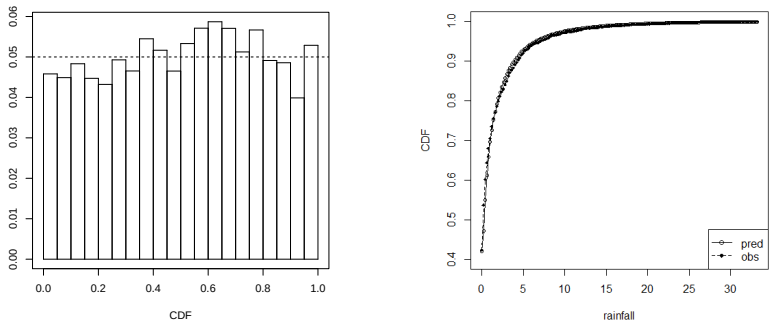
The competing models “base” and “neigh” are compared in terms of predictive performances at a random set of 50 validation sites. When assessing point forecasts, we also compare the predictions with the raw radar data. Results are reported in Table 2, showing both models succeed in correcting radar data. The additional covariates (model “neigh”) improve point forecasts, according both to MAE and MSE. Coverages of the centered 50% and 90% predictive intervals are close to the nominal levels (not shown). Also CRPS is slightly lower for the second model (“neigh”). Graphical displays assessing calibration look very similar. In Fig. 1 we show probabilistic and marginal calibration displays for model “base”; the corresponding plots for model “neigh” look very similar. Non-randomized PIT histogram is almost uniform in both cases, meaning that probabilistic calibration is achieved. The marginal calibration plot reports the predictive and empirical CDFs obtained pooling together all the observations; only small discrepancies are detected.

Table 2 Assessment of point forecasts via MAE and MSPE, of probabilistic forecasts via CRPS, and of the ability in predicting the probability of rain via BS (with respect to zero threshold)

Event	MAE			MSE			CRPS		BS	
	base	neigh	radar	base	neigh	radar	base	neigh	base	neigh
E1	0.33	0.32	0.94	0.34	0.32	2.57	0.25	0.25	0.08	0.07
E2	0.90	0.89	1.40	5.4	5.32	8.57	0.67	0.66	0.07	0.07
E3	0.58	0.57	1.38	2.05	2.18	5.06	0.42	0.42	0.10	0.10
E4	0.79	0.73	1.86	4.27	4.62	13.40	0.57	0.54	0.07	0.07
E5	0.95	0.92	1.54	7.12	6.36	11.95	0.69	0.66	0.11	0.11
E6	0.23	0.23	0.68	0.53	0.44	1.48	0.18	0.18	0.09	0.09

Acknowledgements We wish to thank ARPA-SIMC Emilia Romagna for introducing us to the problem, providing data and performing data pre-processing. Acknowledgements to Tilmann Gneiting for the joint work on the evaluation of predictions. This research work was funded by an ARPA-SIMC / Department of Statistical Sciences agreement (nr. 20/2012) and by a FIRB 2012 grant (project no. RBFR12URQJ; title: Statistical modelling of environmental phenomena: pollution, meteorology, health and their interactions) for research projects by the Italian Ministry of Education, Universities and Research.

Fig. 1 Graphical tools for the assessment of probabilistic forecasts



Non-randomized PIT histogram

Marginal calibration plot

References

1. Bruno F., Cocchi D., Greco F., Scardovi E.: Spatial reconstruction of rainfall fields from rain gauge and radar data. Stoch. Environ. Res. Risk Assess. doi: 10.1007/s00477–013–0812–0 (2013)
2. Czado C., Gneiting T., Held L.: Predictive model assessment for count Data. Biometrics. **65**, 1254-1261 (2009)
3. Dawid, A.P.: Statistical theory: The prequential approach (with discussion). J. R. Stat. Soc. A, **147**, 278–292 (1984)
4. Gneiting T.: Making and evaluating point forecasts. J. Amer. Stat. Assoc. **106** No. 494, 746–762 (2011)
5. Gneiting T., Balabdaoui F., Raftery A.: Probabilistic forecasts, calibration and sharpness. J. R. Stat. Soc. B, Statistical Methodology. **69**, 243-268 (2007)
6. Gneiting T., Raftery A.: Strictly proper scoring rules, prediction, and estimation. J. Amer. Stat. Assoc. **102**, No. 477, 359–378 Review Article. (2007)
7. Scheuerer M.: Probabilistic quantitative precipitation forecasting using Ensemble Model Output Statistics. Quart. J. R. Meteor. Soc. doi: 10.1002/qj.2183 (2013)
8. Sloughter J., Raftery A.E., Gneiting T., Fraley C.: Probabilistic quantitative precipitation forecasting using Bayesian Model Averaging. Mon. Weath. Rev. **135**, 3209–3220 (2007)