Kalman filtering approach in the calibration of radar rainfall data

Marco Costa[1], Magda Monteiro[2], A. Manuela Gonçalves[3]

[1] Escola Superior de Tecnologia e Gestão de Águeda - Universidade de Aveiro, Portugal and CMAF–UL
[2] Escola Superior de Tecnologia e Gestão de Águeda - Universidade de Aveiro, Portugal and CIDMA-UA
[3] Departamento de Matemática e Aplicações, Universidade do Minho, Portugal and CM-UM

E-mail for correspondence: `marco@ua.pt`

Abstract: This work presents a comparative study of some models to estimate radar rainfall in real time using the Kalman filtering approach. This comparison adresses the parameters estimation, the assessment of the accuracy estimates obtained by each model and the impact of the number of rain gauges used in the improvement of radar calibration estimates.

Keywords: Kalman filter, state space model, rainfall estimates, weather radar, calibration

1 Introduction

The weather radar provides precipitation data in a large area, for instance in a radial distance from the radar of 300Km (Figure 1). One of the advantages of radar rainfall over rain gauges is the provision of continuous measurements in real-time, which is unachievable even in a dense telemetered rain gauges network, since there is a large space-time variability of precipitation. However, their estimates have a poor performance, when comparing with gauges estimates, due to errors of either meteorological or instrumental nature which need to be reduced. Having this into account, in the recent years several approaches have been proposed to minimize radar errors, among which is included the combination of radar and gauges measurements, through a state space representation associated to the Kalman filter. This paper aims to discuss and compare different state space formulations through its application to the same data set. This comparison adresses the parameters estimation, the assessment of the accuracy estimates obtained by each model and the impact of the number of rain gauges used in the improvement of radar calibration estimates. It is also important to analyse the behaviour of different state space representations associated to different rain gauges network densities.

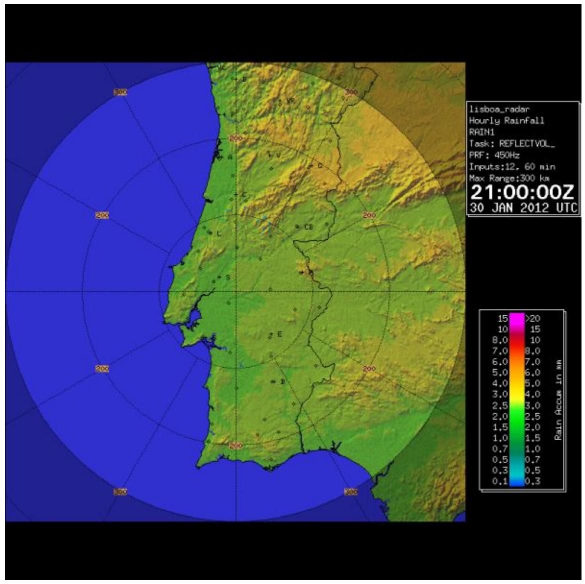

FIGURE 1. Radar umbrella of the weather radar located in Cruz de Leão, Coruche, Portugal

2 State space models and the Kalman filter

The Kalman filter approach provides a real-time scheme to calibrate radar rainfall estimates based on the rain gauge measurements. It is applied to a class of models that admits a state space representation of the form

$$\begin{aligned} A_t &= \beta_t B_t + e_t & (1) \\ \beta_t &= \mu + \phi(\beta_{t-1} - \mu) + \varepsilon_t. & (2) \end{aligned}$$

Equation (1) is the measurement equation and relates the observable variable A_t with the unobservable variable β_t, called the state, while Eq. (2) is the transition or state equation. B_t is a known coefficient and e_t is the measurement error which is a white noise, with variance σ_e^2. The state β_t is a stationary AR(1) process with mean μ, $|\phi| < 1$, where ε_t is a white noise with variance σ_ε^2. Furthermore no assumption is made about the distributions of the disturbances e_t and ε_t, only that they are uncorrelated.
Assuming that parameters of the state-space model are known, the Kalman filter is an iterative algorithm that produces, at each time t, an estimator of the state vector β_t, which is the orthogonal projection of the state vector onto the observed variables up to that time. Let $\widehat{\beta}_{t|t-1}$ represent the predictor of β_t based on the information up to time $t-1$ and let $P_{t|t-1}$ be its mean square error (MSE). The recursive process needs initial values for the state $\beta_{1|0}$ and for its variance $P_{1|0}$, which in this case, as

the state process is assumed to be a stationary AR(1) process it is taken $\widehat{\beta}_{1|0} = \mu$ and $P_{1|0} = \sigma_\beta^2 = \frac{\sigma_\varepsilon^2}{1-\phi^2}$. When the parameters of the model are unknown they have to be estimated and plugged in into the Kalman filter recursive equations. Since precipitation data deviates, in general, from the normal curve it will be considered non parametric methods to estimate the parameters, namely the consistent parameters estimators proposed by Costa and Alpuim (2010). The estimation for the mean of the state process $\{\beta_t\}$, μ, is the average of the ratios A_t/B_t and the remaining parameters of the state process $\{\beta_t\}$ are estimated based on the autocovariance structure of an AR(1) stationary process. The estimator of ϕ is obtained by least square method taking the autocovariances $\widehat{\gamma}_k$, with $k = 1, ..., \ell$, where ℓ is choose having into account the sample dimension. σ_ε^2 is estimated using $\widehat{\sigma}_\varepsilon^2 = \frac{1-\widehat{\phi}^2}{\widehat{\phi}}\widehat{\gamma}_1$ and the variance of the measurement equation is done through the relationship $var(\frac{A_t}{B_t}) = \sigma_\beta^2 + B_t^{-2}\sigma_e^2$.

3 Models

3.1 Linear calibration (LC)

The linear calibration model was proposed by Alpuim and Barbosa (1999) and Costa and Alpuim (2010) and relates rain gauges and radar measurements through a multiplicative factor of calibration, as follows

$$\begin{aligned} G_t &= \beta_t R_t + e_t \\ \beta_t &= \mu + \phi(\beta_{t-1} - \mu) + \varepsilon_t. \end{aligned}$$

G_t is the rain gauge observation in time t, R_t is the radar measurement at the same time and location and β_t is the respective calibration factor. The LC model does not impose any restrictions to the radar or rain gauges measurements unlike other that will be presented.

3.2 Mean field radar rainfall logarithmic bias modelling (FB)

The mean field radar rainfall logarithm bias model was proposed in Chumchean et al. (2006) and is based on the assumption that there are a consistent bias between radar and rain gauges measurements, that is,

$$Y_t = \frac{1}{k}\sum_{i=1}^{k} \log_{10}\left(\frac{G_{i,t}}{R_{i,t}}\right)$$

where k is the number of radar-gauge pairs data available in time t, and $G_{i,t}$ and $R_{i,t}$ are the rainfall and unfiltered radar rainfall at time t at location i. The temporal evolution of the mean field logarithm bias is modeled through the state space model

$$\begin{aligned} Y_t &= \beta_t + e_t \\ \beta_t &= \mu + \phi(\beta_{t-1} - \mu) + \varepsilon_t. \end{aligned}$$

3.3 Power law modelling (PL)

Brown et al. (2001) make the assumption that gauge and radar reflectance measurements can be related through a power law, $G_t = bR_t^{\alpha}$. The authors consider a linearization of the power law where the parameters α and b are not necessarily fixed quantities but may vary stochastically over time. However they concluded that α could be treated as if it is constant, which result in

$$\begin{aligned} Y_t &= \alpha U_t + \beta_t + e_t \\ \beta_t &= \mu + \phi(\beta_{t-1} - \mu) + \varepsilon_t, \end{aligned}$$

where $Y_t = \log_{10}(G_t)$, $U_t = \log_{10}(R_t)$ and e_t is a white noise error. α is previously estimated by the method of least squares as the slope of the usual linear regression between Y_t and U_t. Note that PL and FB models assume that both radar and gauges measurements are nonzero due to the logarithmic function. Another note to point out is that modelling procedure of LC and PL models is based on single-site approach, and it will be necessary to interpolate the predicted calibration factors β_t to other locations where it is intended to correct the radar measurements.

4 COMPARATIVE STUDY

It is used a data set of 17 stratiform storms between September of 1998 and November of 2000 (in a total of 178 hourly precipitation estimates) in a 10×14 Km^2 area, located around 40 Km north of Lisbon at a distance from 31 to 44 Km from the weather radar in *Cruz do Leão*. This area has five rain gauges: Merceana (Mr), Meca (M), Olhalvo (O), Penedos (P) and Abrigada (A) and it has the highest gauge density under the radar umbrella ($\sim$ 1 gauge/28Km^2). The performances of the calibration of the three models are compared in a set of scenarios considering all the combinations using 1, 2, 3 or 4 rain gauges to calibrate the radar estimates in the remaining gauges not used in the parameters estimation procedure in a total of 30 scenarios.

4.1 Models specification and calibration procedure

In order to ensure the independence between parameters estimation and the calibration modelling, three storms occurred in 13 of January, 28 of April and 19 of October of 2000 (62 hours) are used to estimate the models parameters, while the remaining storms are used in the assessment of the performance of the calibration. The radar calibration procedure focus on the remaining fourteen storms not used in the parameter estimation (116 hourly measurements). The calibration procedure in the scenarios with more than one rain gauges needs interpolating its calibration factors to

TABLE 1. Square roots of the empirical mean square errors of the three models. In brackets is the % of RMSE reduction comparing to the reference value.

number of rain gauges	Model LC	FB	PL
1	1.38 (-3%)	1.30 (-9%)	1.19 (-16%)
2	1.21 (-16%)	1.23 (-14%)	1.10 (-23%)
3	1.16 (-19%)	1.16 (-19%)	1.11 (-22%)
4	1.09 (-23%)	1.07 (-25%)	1.11 (-22%)
global	1.21 (-15%)	1.19 (-16%)	1.13 (-21%)

other locations for models LC and PL. In this case it is considered the inverse square distance method which takes into consideration all available rain gauges to calibrate the radar estimates. For each scenario it was implemented the Kalman filter equations in order to predict the state β_t at each hour t. As it is considered a real-time procedure, the filtered prediction $\widehat{\beta}_{t|t}$ is used to estimate β_t.When the calibration procedure includes only a single rain gauge, the process to extend the calibration to other location is a straightforward procedure. This remains true even when the model applied is the FB since this model assumes a single mean field bias of calibration. Note that for LC model the radar calibration is obtained by multiplying the radar estimate R_t by the filtered calibration factor $\widehat{\beta}_{t|t}^{(LC)}$, while in FB and PL models it is necessary to convert the respective $\widehat{\beta}_{t|t}$ into B_t.

4.2 Performance assessment of models

The models performance assessment is done according to the empirical square root of the mean square error of point prediction using the fourteen storms (116 hourly observations) kept for this purpose. It is compared the gauges rainfall estimates G_t with the calibrated radar cell measurement $\widehat{R}_t^{(m)}$, with $m =$ LC, FB and PL, at the same location. As it are available five rain gauges in the area under study, it is considered systems with 1, 2, 3 or 4 gauges to the calibration process and for each of these schemes are computed the empirical square root of the mean square error for all combination with each number of gauges. The empirical square root of the mean square error $\text{RMSE}_k^{(m)}$ for the scheme modelled based on k rain gauges with the model m is computed by

$$\text{RMSE}_k^{(m)} = \sqrt{\frac{1}{116(5-k)} \sum_{i}^{5-k} \sum_{t}^{116} \left(G_t^i - \widehat{R}_t^{(m),i} \right)^2}.$$

Table 1 presents the RMSE for the three models considering different numbers of rain gauges in the calibration process. The pre-calibration RMSE of the five rain gauges is taken as reference value to analyse the impact of the calibration procedures. For data set under the calibration procedure (the fourteen events) this value is 1.43. Thereby, it is possible to compare the models performance in view of the percentage of the reference value reduction (indicated in Table 1 in brackets). It can be state that the models lead to a reduction in the error of the radar rain estimates. Neveryteless, the model PL is less sensitive to the number of rain gauges used in the calibration process. Both RMSE of models LC and FB decrease significantly when it is added more gauges to the calibration process. When the rain gauges density is the lowest (1 gauge per 140Km^2) the PL model performed the largest reduction of the RMSE with a strong difference to the other models. On the other hand, when it is considered the highest density (1 gauge per 35Km^2), models have similar performances, nevertheless the FB model produces the greatest reduction of the RMSE.

Acknowledgments: Marco Costa was partially supported by Portuguese Foundation for Science and Technology ("FCT–Fundação para a Ciência e a Tecnologia"), PEst OE/MAT/UI0209/2011. Magda Monteiro was partially supported by *FEDER* funds through *COMPETE* – Operational Programme Factors of Competitiveness ("Programa Operacional Factores de Competitividade") and by Portuguese funds through the *Center for Research and Development in Mathematics and Applications* (University of Aveiro) and the FCT, within project PEst-C/MAT/UI4106/2011 with COMPETE number FCOMP-01-0124-FEDER-022690. A. Manuela Gonçalves was partially financed by FEDER Funds through COMPETE and by FCT, within the Project Est-C/MAT/UI0013/2011.

References

Alpuim T., Barbosa S. (1999). The Kalman filter in the estimation of area precipitation. *Environmetrics*, **10**, 377 – 394.

Brown M., Diggle P., Lord M., Young P. (2001). Space-Time Calibration of Radar Rainfall Data. *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, **50(2)**, 221 – 241.

Chumchean S., Seed A., Sharma A. (2006).Correcting of real-time radar rainfall bias using a Kalman filtering approach. *Journal of Hydrology*, **317**,123 – 137.

Costa M., Alpuim T. (2010). Parameter estimation of state space models for univariate observations. *Journal of Statistical Planning and Inference*, **140**, 1889 – 1902.

Costa M., Alpuim T. (2011). Adjustment of state space models in view of area rainfall estimation. *Environmetrics*, **22**, 530 – 540.