Chapter

KALMAN FILTERING APPROACH IN THE CALIBRATION OF RADAR RAINFALL DATA - A COMPARATIVE ANALYSIS OF STATE SPACE REPRESENTATIONS

Marco A. S. Costa[1*]***, Magda S. V. Monteiro***[2]
and A. Manuela Gonçalves[3]
[1] Escola Superior de Tecnologia e Gestão de Águeda,
Universidade de Aveiro, Portugal
Centro de Matemática e Aplicações Fundamentais da Universidade de Lisboa
[2] Escola Superior de Tecnologia e Gestão de Águeda,
Universidade de Aveiro, Portugal
Unidade de Investigação de Matemática e Aplicações
[3] Departamento de Matemática e Aplicações, Universidade do Minho, Portugal
Centro de Matemática da Universidade do Minho

Abstract

In this chapter it is presented a comparative study of some methods to estimate radar rainfall in real time. Radar rainfall estimates have a poor performance comparatively to rain gauge estimates, due to errors of either meteorological or instrumental nature. Nevertheless, weather radar presents several advantages over rain gauges, namely by providing continuous measurements in real time, which it is not possible even in

[*]E-mail address: marco@ua.pt

a dense telemetered rain gauge network, due to the large space-time variability of precipitation. Given these advantages, several approaches have been proposed to minimize radar errors. Namely, the combination of these two types of measurements via a state space representation associated to the Kalman filter has been investigated in recent years. However, recent literature presents different state space representations, and therefore their results are not directly comparable. This work intends to discuss and compare different state space formulations based on a same data set; for instance, the comparison between the modeling of the mean field radar rainfall logarithmic bias (Chumchean et al., 2006), a linear radar-rain gauge calibration model (Alpuim & Barbosa, 1999; Costa & Alpuim, 2011) and a power law model (Brown et al., 2001). This investigation takes into account some issues associated to the state space approach: for instance, parameters estimation, the assessment of the accuracy estimates obtained by each model. Another question worth investigating is the impact of the number of rain gauges used in the improvement of radar calibration estimates. It is useful to analyze the behavior of different state space representations associated to different rain gauge network densities. The data set consists of a set of storms between September 1998 and November 2000 in an area located around 40Km upnorth of the weather radar located in Cruz do Leão, in the Lisbon region. This region has been used in several works of the Portuguese Institute of Meteorology since this area presents the highest rain gauge density under the radar umbrella.

Key Words: Kalman filter, state space model, rainfall estimates, weather radar, calibration, nonparametric estimation.

AMS Subject Classification: 60G35, 62M10, 62M05.

1. Introduction

The weather radar allows the monitoring of the rainfall in area, thus constituting one of its main advantages. However, precipitation radar estimates contain errors associated with the reflectivity measurement errors or with the reflectivity rainfall rate (Z-R) conversion errors (Jordan et al., 2000). Nevertheless, even if these two sources of error have been corrected based on an understanding of the physical processes, there remain errors in the radar rainfall estimates when compared with rain gauge measurements (Chumchean et al., 2006). A suitable mechanism to reduce errors is to combine both weather radar and rain gauge network data. On the one hand, the weather radar pro-

vides precipitation data in a large area, for instance in a 300Km radial distance from the radar (Figure 1). On the other hand, a rain gauge network gives accurate point measurements in its location. The differences between radar and rain gauges are understandable because these are two different types of rainfall measuring sensors.

There are several ways to tackle with radar estimates bias when compared with accurate rain gauge estimates. Krajewski (1987) and Severino & Alpuim (2005) apply an optimal interpolation method based on Kriging and CoKriging. An alternative way to merge the radar and rain gauge measurements is to relate the two types of measurements through a state space representation associated to the Kalman filter algorithm. In this chapter it is performed a discussion of three models based on the Kalman filter framework. Chumchean et al. (2006) formulated a model with the assumption-which was also considered in Anagnostou et al. (1998) and Krajewski & Smith (2002)-that radar errors produce a uniform bias in radar rainfall estimation when compared to rain gauge data. Brown et al. (2001) proposed a single-site modeling approach with a power law model to describe the relationship between gauge and radar data, which needs a linearization procedure with the logarithmic function. Alpuim & Barbosa (1999) and Costa & Alpuim (2011) used a multiplicative calibration factor model. Whereas Alpuim & Barbosa (1999) compared the mean area prediction in a univariate and multivariate point of view, Costa & Alpuim (2010) evaluated the models performance when its parameters are estimated by Gaussian likelihood and nonparametric estimators.

The comparison of the three models is achieved based on a single data set. However, the data set is divided in two subsets: one for the parameters estimation and the remaining data for the assessment procedure. The parameter estimation is performed by using the nonparametric estimators suggested in Costa & Alpuim (2010), which had achieved good results in the radar calibration (Costa & Alpuim, 2011).

2. Kalman Filtering Approach

The Kalman filtering approach has the main advantage of providing a real-time scheme to calibrate radar rainfall estimates based on rain gauge measurements. This algorithm is an iterative procedure to update unknown stochastic variables predictions taking into account the observed variables at each time, and provides an estimate of the error variance in the computed prediction. This potential has encouraged the application of the Kalman filter techniques in the

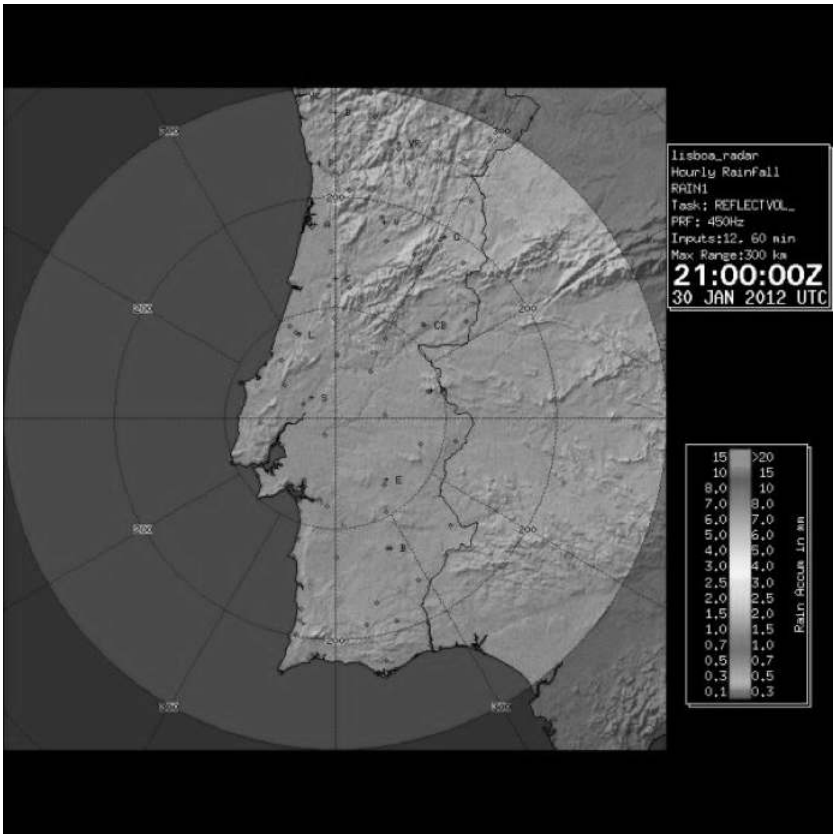

Figure 1. Radar umbrella of the weather radar located in Cruz do Leão, Coruche, Portugal.

radar rainfall estimates calibration in several forms. For simplicity, in this chapter the modeling procedure of the relation between rain gauge and radar estimates of the precipitation is based on a single site modeling approach associated to interpolation methods, if needed.

2.1. State space models and the Kalman filter

The Kalman filter (Kalman, 1960) is applied to a class of models that admits a state space representation. In this chapter, it will be considered, in general,

the state space representation as follows

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

Equation (1) is called the measurement equation and relates the observable variable A_t with the unobservable variable β_t, called the state. The coefficient B_t is known and e_t is a white noise, called the measurement error, with variance σ_e^2. Furthermore, the state β_t varies in time according to Eq. (2), the transition or state equation, i.e., a stationary AR(1) process with mean μ. Here, ϕ is an autoregressive coefficient, $|\phi| < 1$, and ε_t is a white noise with variance σ_ε^2. The disturbances e_t and ε_t are assumed to be uncorrelated, that is, $E(e_t\varepsilon_s) = 0$, for all t and s. The model (1)-(2) can be seen as a dynamic linear regression where the slope is an AR(1) process.

Notice also that no assumption is made about the errors distributions. Indeed, it is usual to consider the model (1)-(2) as a standard linear Gaussian state space time series model, that is, the errors e_t and ε_t follow Gaussians distributions. However, in general precipitation data deviates from the normal curve. Moreover, in Brown et al. (2001) it is indicated that the distribution of residuals is somewhat longer tailed than the normal distribution, which reinforces that the assumption of the normality of the errors may not be a good choice. Alternatively, in this chapter, are considered nonparametric estimators for the parameters, following the work of Costa & Alpuim (2010).

Assuming that parameters of a state-space model are known, the Kalman filter recursions give the best linear predictors to filter, forecast, and smoothing the prediction of vector of states. 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). As the orthogonal projection is a linear estimator, the predictor for the observable at time t, A_t, is given by

$$\widehat{A}_{t|t-1} = B_t\widehat{\beta}_{t|t-1}.$$

When, at time t, A_t is available, the prediction error or innovation, $\eta_t = A_t - \widehat{A}_{t|t-1}$, is used to update the estimate of A_t, through the equation

$$\widehat{\beta}_{t|t} = \widehat{\beta}_{t|t-1} + K_t\eta_t$$

where K_t is called the Kalman gain matrix and is given by

$$K_t = P_{t|t-1}A_t\left(A_t^2P_{t|t-1} + \sigma_e^2\right)^{-1}.$$

Furthermore, the MSE of the update predictor $\widehat{\beta}_{t|t}$ verifies the relationship

$$P_{t|t} = P_{t|t-1} - K_t A_t P_{t|t-1}.$$

In turn, at time t, the forecast for the state vector β_{t+1} is given by the equation

$$\widehat{\beta}_{t+1|t} = \mu + \phi(\widehat{\beta}_{t|t} - \mu)$$

with MSE $P_{t+1|t} = \phi^2 P_{t|t} + \sigma_\epsilon^2$. 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 given by

$$\widehat{\beta}_{1|0} = \mu \text{ and } P_{1|0} = \sigma_\beta^2 = \frac{\sigma_\varepsilon^2}{1-\phi^2}.$$

2.2. Models

In this chapter are considered three models to perform the radar rainfall estimates calibration proposed in the Kalman filtering approach literature. Although the three models are real-time procedures based on Kalman filter recursions, different models are taken into consideration to relate rain gauges and radar measurements. The next sections briefly present the models as well as their assumptions.

2.2.1. Linear calibration (LC)

Alpuim & Barbosa (1999) and Costa & Alpuim (2010) proposed a state space representation that relates rain gauges and radar measurements by a multiplicative factor of calibration between these two estimates, as follows

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

where 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 corresponding calibration factor for each pair of values. The model has the assumptions of the general model (1)-(2), with the necessary adaptations. This formulation assumes a local linear relation between radar and rain gauge estimates since it is considered, for each time, that rain gauge measurement is proportional to radar observation added to an error.

The linear calibration model has the advantage of not presenting restrictions to the radar and rain gauge measurements, that is, it can be formulated even in the cases when radar or rain gauges present null estimates. As will be seen later on, other models present this type of restriction.

As the modeling procedure is the single-site approach, it will be necessary to interpolate the predicted calibration factors β_t to other sites where it is intended to correct radar measurements.

2.2.2. Mean field radar rainfall logarithmic bias modeling (FB)

The second model, designated by mean field radar rainfall logarithm bias, was proposed in Chumchean et al. (2006) and is based on the assumption that there is a consistent bias between radar and rain gauge measurements. This assumption is supported by Austin (1987), who concludes that when the radar adjustment schemes force agreement at a few gauges sites these are inappropriate due to large random discrepancies between radar and gauge measurements as a result of sampling errors. However, Costa & Alpuim (2010), in a comparative analysis between one calibration factor and a single site approach, concluded that the single site approach produces better results (minimum mean square error).

Assuming that is reasonable to admit a single bias, Chumchean et al. (2006) defined the mean field radar rainfall logarithmic bias at time t as

$$\beta_t = \frac{1}{k}\sum_{i=1}^{k} \log_{10}\left(\frac{G_{i,t}}{R_{i,t}}\right) \tag{3}$$

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 logarithmic mean field bias is modeled through the state space model

$$Y_t = \beta_t + e_t \tag{4}$$

$$\beta_t = \mu + \phi(\beta_{t-1} - \mu) + \varepsilon_t. \tag{5}$$

where Y_t is the observed mean field logarithmic bias at time t and is computed by Eq. (3). While in the original work of Chumchean et al. (2006) the state β_t is taken as a stationary AR(1) process with zero mean, in this chapter this assumption is not considered, since it seems that there is no reason for imposing a zero mean for the state process, therefore it is considered the mean μ.

Indeed, if the state with zero mean is imposed, it would imply admitting that the ratios G/R are, in average, equal to one, that is, $G = R$. However, as is largely accepted, in general the weather radar underestimates precipitation intensity. The model (4)-(5) assumes that both radar and gauges measurements are nonzero due to the logarithmic function.

2.2.3. Power law modelling (PL)

Brown et al. (2001) make the assumption that the relationship between gauge and radar reflectance measurements can be described by a power law, as previously mentioned in Collinge & Young (1993),

$$G_t = bR_t^a \tag{6}$$

In a first proposition, Brown et al. (2001) consider a linearization of Eq. (6) where the parameters a and b are not necessarily fixed quantities but may vary stochastically over time. That is,

$$\begin{aligned} Y_t &= \alpha_t U_t + \beta_t + e_t \\ \alpha_t &= \mu_\alpha + \phi_\alpha(\alpha_{t-1} - \mu_\alpha) + \varepsilon_{\alpha,t} \\ \beta_t &= \mu_\beta + \phi_\beta(\beta_{t-1} - \mu_\beta) + \varepsilon_{\beta,t} \end{aligned}$$

where $Y_t = \log_{10}(G_t)$, $U_t = \log_{10}(R_t)$ and e_t is a white noise error. However, Brown et al. (2001) concluded that the prediction intervals on the parameter α remain quite wide most of the time, suggesting that it is poorly identified; therefore, they treated this parameter as if it is constant.

Thus, in this chapter will be taken the model

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

where α is previously estimated by the method of least squares as the slope of the usual linear regression between Y_t and U_t. As this approach is a single-site modeling procedure, it also needs interpolation methods to calibrate radar estimates in other sites.

2.3. Estimation of the parameters

As mentioned before, the estimation of the parameters is performed by non-parametric estimators based on the method of moments, which does not assume any distribution for the errors. Costa & Alpuim (2010) proposed consistent estimators for the parameters of the model (1)-(2) whose performance

was compared to the usual maximum likelihood estimators considering Gaussian errors in a Monte Carlo study. The nonparametric estimators presented a good performance even when errors were simulated with normal distribution. It is possible to obtain the maximum likelihood estimates by maximizing the log-likelihood function in order to estimate unknown parameters by using numerical methods, namely the EM algorithm (Shumway & Stoffer, 1982), or the Newton-Raphson algorithm (Harvey, 1996). However, the maximum likelihood method may not converge at all and produces estimates that fall outside the parameters space more frequently than the non-parametric estimators, particularly for small values of the autoregressive coefficient (Costa & Alpuim, 2010).

The estimation of the mean μ of the state process $\{\beta_t\}$ is performed as the average of the ratios A_t/B_t, that is

$$\widehat{\mu} = \frac{1}{n}\sum_{t=1}^{n}\frac{A_t}{B_t}.$$

The remaining parameters of the state process $\{\beta_t\}$ are estimated based on the autocovariance structure of an AR(1) stationary process. Indeed, if $\{\beta_t\}$ follows an AR(1) stationary process, it verifies the recursive equation $\gamma_k = \phi\gamma_{k-1}$, for any $k \geqslant 2$, where γ_k is the autocovariance function, whose empirical estimator is

$$\widehat{\gamma}_k = \frac{1}{n}\sum_{t=1}^{n-k}\left(\frac{A_{t+k}}{B_{t+k}} - \widehat{\mu}\right)\left(\frac{A_t}{B_t} - \widehat{\mu}\right).$$

The estimator for the autoregressive parameter ϕ is obtained by the least square method taking into account the autocovariances $\widehat{\gamma}_k$, with $k = 1, ..., \ell$, that is, considering $\ell - 1$ equations of the form $\gamma_k = \phi\gamma_{k-1}$, which produce the estimator

$$\widehat{\phi} = \sum_{k=2}^{\ell}\widehat{\gamma}_k\widehat{\gamma}_{k-1}\left(\sum_{k=2}^{\ell}\widehat{\gamma}_k^2\right)^{-1}.$$

The choice of ℓ depends on sample dimension, as proposed in Costa & Alpuim (2010). Notice that $\{\beta_t\}$ is an AR(1) stationary process, so the noise variance in the state equation σ_ε^2 is estimated using the explicit formula

$$\widehat{\sigma}_\varepsilon^2 = \frac{1-\widehat{\phi}^2}{\widehat{\phi}}\widehat{\gamma}_1.$$

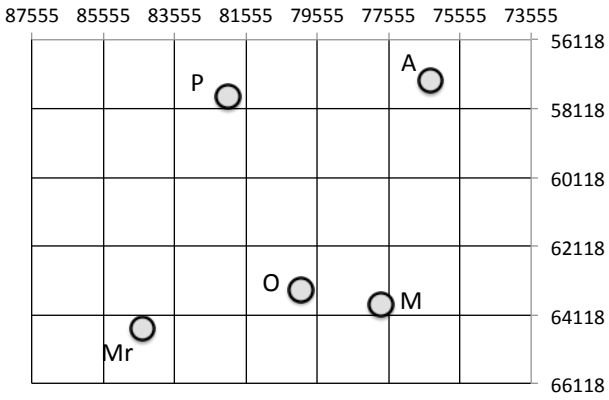

Figure 2. Location of the five rain gauges in the Portuguese system of coordinates and in the grid of 35 cells used in this chapter.

The variance of the measurement equation is done by using the relationship $var(A_t/B_t) = \sigma_\beta^2 + B_t^{-2}\sigma_e^2$, where $\sigma_\beta^2 = \frac{\sigma_\varepsilon^2}{1-\phi^2}$, that produces the estimator

$$\widehat{\sigma}_e^2 = \left[\sum_{t=1}^{n}\left(\frac{A_t}{B_t} - \widehat{\mu}\right)^2 - n\widehat{\sigma}_\beta^2\right]\left(\sum_{t=1}^{n} B_t^{-2}\right)^{-1}.$$

3. Comparative Study

The models have been proposed in the literature accompanied by case studies to illustrate their benefits and advantages. However, as in each case is used a different data set, it is not possible to compare their performances. In this chapter it is performed a modeling work based on a data set of radar and gauge estimates, where the three models above described are applied. This approach allows to compare the performances of these models based on the same data and with the same conditions, namely, the same density gauge/Km^2.

3.1. Experiment description

The models comparison is done considering a data set of 17 stratiform storms between September 1998 and November 2000 (in a total of 178 hourly precipitation estimates) in a 10×14 Km^2 area (Figure 2), including the Alenquer River basin, located around 40 Km north of Lisbon and between 31 and 44 Km

Table 1. Descriptive statistics of radar and rain gauges data in the five locations

descriptives	Rain gauges					Radar				
	Mr	M	O	P	A	Mr	M	O	P	A
average	1.60	1.45	1.51	1.64	1.60	1.06	0.95	0.95	0.91	1.04
median	1.00	0.85	1.05	1.00	1.00	0.52	0.45	0.50	0.43	0.49
std. deviation	1.78	1.73	1.62	1.86	1.72	1.48	1.22	1.24	1.25	1.34
kurtosis	5.67	6.69	4.32	4.03	2.96	10.61	4.68	6.08	8.02	5.10
skewness	2.03	2.27	1.91	1.89	1.61	2.88	2.07	2.28	2.58	2.12

of distance from the weather radar in Cruz do Leão. This area has been studied by meteorologists of the Institute of Meteorology of Portugal due to the fact that there are five rain gauges located in this area: Merceana (Mr), Meca (M), Olhalvo (O), Penedos (P) and Abrigada (A). These rain gauges correspond to a reasonable high density ($\sim$ 1 gauge/28Km^2) and present the highest gauge density under the radar umbrella. Furthermore, this area is associated to a very low concentration time (about 3 h), which makes this region particularly prone to flash floods, making it especially appropriated to study models. Since about 90% of the region is below 400 m, it might be considered homogeneous from the point of view of the climatology of precipitation, and it was not deemed necessary to explicitly correct the orography effect on the precipitation values (Alpuim & Barbosa, 1999).

Previous works in this area have considered smaller or larger areas (Alpuim & Barbosa, 1999; Severino & Alpuim, 2005; Costa & Alpuim, 2011) but without significant differences in the area size. This small river basin, prone to flash floods and intensively instrumented, was selected for studies of precipitation measurement by radar, as well as for other hydrometeorological and hydrological studies and for testing different hydrological models by using radar data (Rossa et al., 2005). Costa & Alpuim (2011) used this data set in order to compare the impact between the nonparametric estimator and the Gaussian maximum likelihood estimators in view of the area rainfall estimation. As usual in this type of studies, the radar resolution is 2×2 Km; however, the nominal resolution of the radar is 1×1 Km that was transformed in the resolution 2×2 Km. The radar and rain gauge estimates were converted to hourly measurements, in this case to each cell of 2×2 Km.

The available data from Cruz do Leão weather radar was collected in an important period time testing since this radar was installed in 1995. Therefore, it is expected to observe a significant variability in this initial period of the radar's work.

Table 2. Combination of rain gauges considered for each number of rain gauges taken in the calibration procedure

number of rain gauges					
1	2		3		4
Mr	MrM	MP	MrMO	MrPA	MOPA
M	MrO	MA	MrMP	MOP	MrOPA
O	MrP	OP	MrMA	MOA	MrMPA
P	MrA	OA	MrOP	MPA	MrMOA
A	MO	PA	MrOA	OPA	MrMOP

Table 1 shows the main descriptive statistics of radar and gauges measurements. In a preliminary analysis, it is clear that radar measurements underestimate, since radar samples have lower averages than the respective gauge averages. Another relevant conclusion from Table 1 is the confirmation that the precipitation data is not Gaussian distributed. Indeed, the sample kurtosis and skewness clearly deviated from the excepted Gaussian values for these statistics. On the other hand, the same conclusion is achieved through the analysis of the histograms, Q-Q plots and Kolmogorov-Smirnov test, which are omitted for simplicity.

Additionally, Figure 3 shows the boxplots of gauge and radar measurements, ratios G_t/R_t and $\log_{10}(G_t/R_t)$. The transformation of ratios G_t/R_t through the logarithmic function clearly normalized data; however, as it will be mentioned later on, the model based on this procedure does not produce the best results. As expected, the data set has several outliers, which is very common in this type of data. Only the severe outliers were not considered in the parameters estimation procedure because they may distort their estimates. Nevertheless, the remaining outliers were kept because they are important, mainly to estimate correctly the variances of the errors.

In order to ensure the independence between parameters estimation and the calibration modeling, three storms occurred in January 13, April 28 and October 19, 2000 (62 hours) are used to estimate the models parameters, while the remaining storms are used in the assessment of calibration performance. The calibration performances of the three models are compared in a set of scenarios considering 1, 2, 3 or 4 rain gauges to calibrate the radar estimates in the remaining gauges not used in the parameters estimation procedure. Table 2 shows all scenarios considered in the calibration procedure.

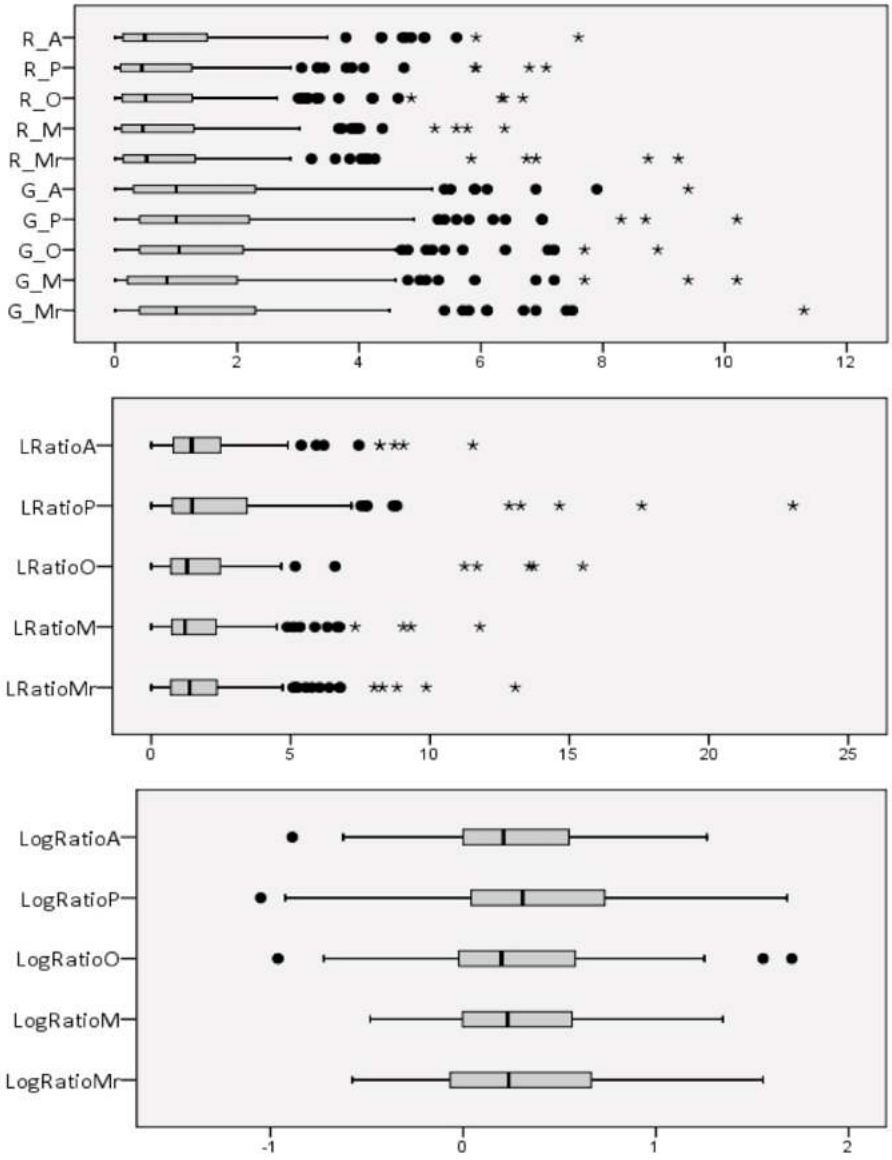

Figure 3. (Top) Boxplots of radar (R) and gauges (G) measurements of the precipitation in the five locations with rain gauges; (Center) Boxplots of linear ratios (LRatios) G_t/R_t; (Bottom) Boxplots of the $\log_{10}(G_t/R_t)$ values.

3.2. Models specification

Before the prediction of the calibrated radar estimates, it is necessary to fit each of the three models to the data relatively to the five locations where the rain gauges are located. This adjustment is performed by using only the three storms mentioned above in order to separate the parameters estimation from the radar calibration procedure.

Notice that, as it was previously seen, precipitation data is usually not normally distributed. This puts the assumption of normality into question if it is intended to obtain Gaussian maximum likelihood estimates. Furthermore, the data set used in this chapter consists of a set of disjoint storms with a short length in general. This data structure is not the most suitable structure to implement numerical algorithms since, in practice, the log-likelihood is the sum of log-likelihood of small samples, which may also contain missing values. Thus, the selected nonparametric estimators seem to be adequate for this type of data because they are less influenced by missing values or by non-sequential data. Hence, the nonparametric estimators for the state equation's parameters are constructed on the lags of observations: for instance, the autocovariance γ_k is computed based on the observed lags k, which may not present the same number of observations than the other γ_s.

Table 3 presents the parameters estimates for the single site models of each of the three models. Notice that it was followed the suggestion of Costa & Alpuim (2010) to choose ℓ, in this case $\ell = 15$, since the storms are very short. Both LC and FB models confirm that, in average, radar measurements are lower than gauges estimates, but in a different magnitude. It is possible to analyze and interpret the estimates of averages μ in models LC and FB. These estimates can be compared approximately if the conversion of the estimate of μ of the model FB is performed as 10^{μ}. By performing this conversion, the results are 1.406, 1.285, 1.211, 1.268 and 1.396, respectively, which represent that model FB tends to produce lower calibration factors between G_t and R_t than LC model, that is, the first model is more conservative.

The LC model presents an advantage in the interpretation of errors variance. Indeed, the variance σ_e^2 can be interpreted as the variability associated to the rain gauge, while the variance σ_ε^2 can be interpreted as a component of the variability associated to the radar. Considering this interpretation, the errors variances estimates obtained in the LC model indicate that the rain gauges located in Merceana (Mr) and Penedos (P) have the highest variability. This is possibly due to the fact that these two rain gauges are of the bascule type, while

Table 3. Parameters estimates for the single-site models for each of the three models LC, FB and PL

model			LC		
gauge	μ	ϕ	σ_ε^2	σ_e^2	
Mr	1.76	0.608	0.863	0.00372	
M	1.43	0.691	0.229	0.00112	
O	1.33	0.505	0.187	0.00112	
P	1.63	0.657	0.650	0.01110	
A	1.53	0.768	0.325	0.00126	
model			FB		
gauge	μ	ϕ	σ_ε^2	σ_e^2	
Mr	0.148	0.247	0.0640	0.0619	
M	0.109	0.751	0.0098	0.0940	
O	0.083	0.485	0.0078	0.0917	
P	0.103	0.552	0.0335	0.1192	
A	0.145	0.714	0.0093	0.0599	
model			PL		
gauge	α	μ	ϕ	σ_ε^2	σ_e^2
Mr	0.695	0.116	0.172	0.0418	0.0643
M	0.660	0.123	0.866	0.0129	0.0735
O	0.662	0.078	0.609	0.0066	0.0819
P	0.235	0.175	0.591	0.0561	0.0611
A	0.712	0.158	0.576	0.0239	0.0356

the remaining gauges are of the syphon type. The models residuals were analyzed in order to perform a diagnostic checking of the models adjustment. As it is assumed that errors are a white noise process, both autocorrelations (ACF) and partial autocorrelations (PACF) functions were drawn. For instance, Figure 4 shows samples ACF and PACF functions for the Abrigada rain gauge. In general, samples ACF and PACF functions have the expected behaviour of a white noise with no significant serial correlation, which validates this initial assumption. On the other hand, this check supports that the nonparametric estimators are proper to this type of data.

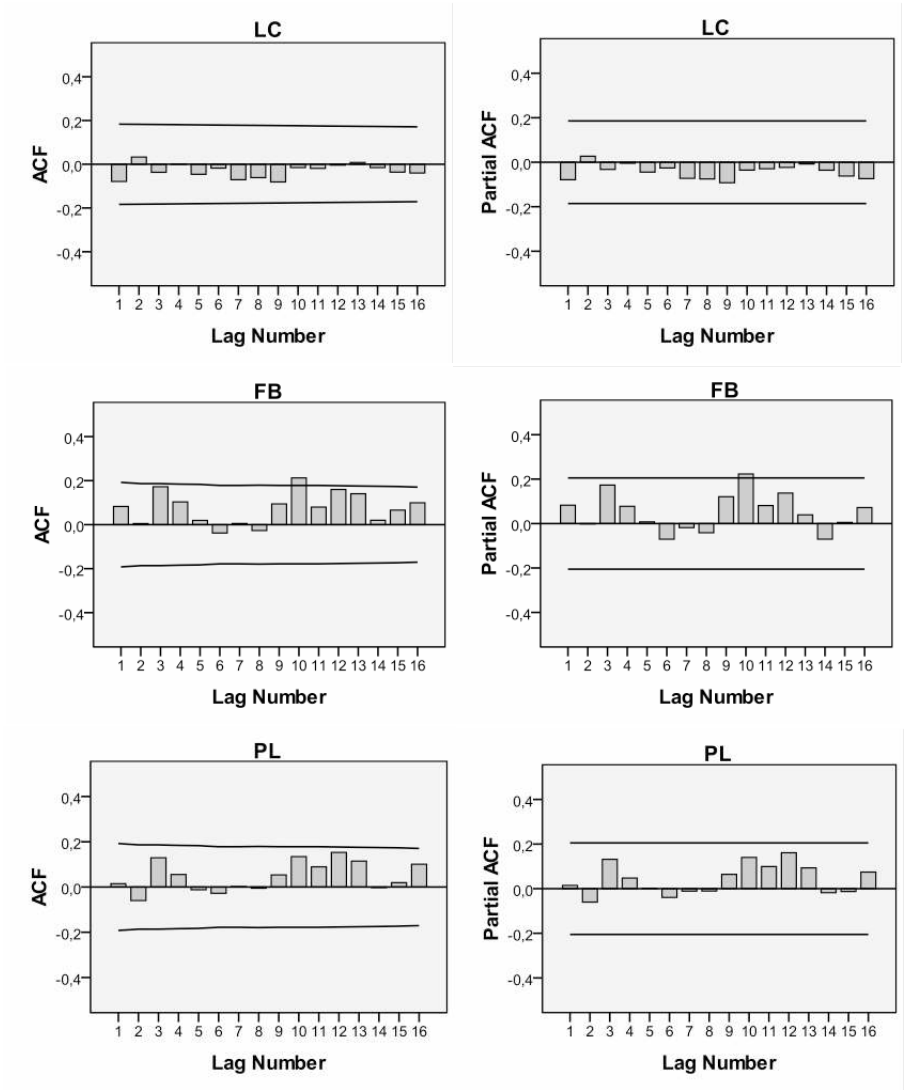

Figure 4. Estimates of autocorrelation function (ACF) and partial autocorrelation function for the rain gauge located in Abrigada.

3.3. Calibration procedure

The radar calibration procedure focus on the remaining fourteen storms not used in the parameter estimation (116 hourly measurements). The calibration

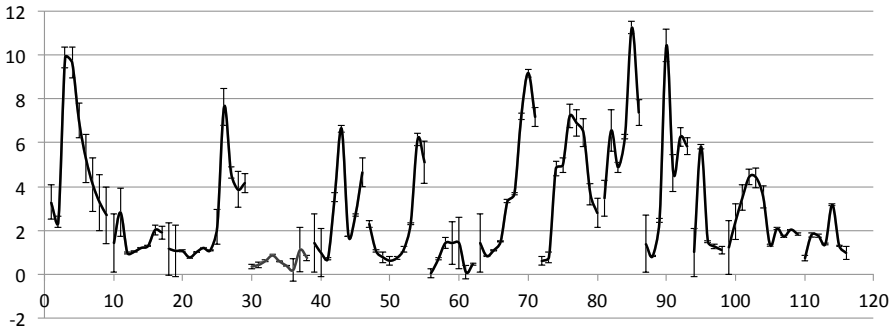

Figure 5. Calibration factors $\widehat{\beta}_{t|t}$ with the empirical pointwise intervals $\pm 2\sqrt{P_{t|t}}$ relatively to the rain gauge located in Meca using the single-site model LC.

procedure in the scenarios with more than one rain gauge requires the interpolation of its calibration factors to other locations. The FB model does not need interpolation procedures since it assumes a single mean field bias in the radar umbrella.

The Kalman filter equations were implemented for each scenario of Table 2 in order to predict the state β_t at each hour t. Since the aim is to consider a real-time procedure, the needed prediction is the filtered prediction of β_t, that is, $\widehat{\beta}_{t|t}$. Notice that in the beginning of each storm the Kalman filter algorithm is initialized with the initial values $\widehat{\beta}_{1|0}$ and $P_{1|0}$. For instance, Figure 5 shows the filtered calibration factors $\widehat{\beta}_{t|t}$, with the empirical pointwise intervals with semi-amplitude $\pm 2\sqrt{P_{t|t}}$ for the single-site model of Meca by using the model LC for the fourteen storms. When the calibration procedure includes only a single rain gauge, the process of extending the calibration to other location is a straightforward procedure. This remains true even if the model applied is the FB, since this model assumes a single mean field bias of calibration. In this case, Chumchean et al. (2006) used the equation $B_t = 10^{\widehat{\beta}_{t|t}^{(FB)} + 0.5P_{t|t}}$ to convert the predicted state $\beta_t^{(FB)}$ into B_t, which is the calibration factor used for correcting the radar measurements in other locations, that is,

$$\widehat{R}_t^{(FB)} = R_t 10^{\widehat{\beta}_{t|t}^{(FB)} + 0.5P_{t|t}}.$$

In the case of model LC, the radar calibration is obtained by multiplying

the radar estimate R_t by the filtered calibration factor $\widehat{\beta}_{t|t}^{(LC)}$, that is,

$$\widehat{R}_t^{(LC)} = R_t \widehat{\beta}_{t|t}^{(LC)}.$$

The prediction of the states of model PL, $\beta_{t|t}^{(PL)}$ allows to obtain the radar calibration by using the expression

$$\widehat{R}_t^{(PL)} = 10^{\widehat{\beta}_{t|t}^{(PL)}} R_t^{\widehat{\alpha}}.$$

Interpolation methods are needed when the correction of radar measurements is based on more than one rain gauge and the modeling procedure assumes one of the models LC or PL. When the set of rain gauges is sufficiently representative to fit a model of spatial continuity into the spatial process of calibration factors, it is suitable to use stochastic methods of interpolation in space, as the Kriging method. However, the available data has few rain gauges to allow the adjustment of spatial continuity models. In the literature there are several alternative methods such as the Thiessen polygons, the multiquadratic interpolation or the inverse square distance methods (for instance, Haberlandt, 2007; Babak & Deutsch, 2008). In this case it is considered the inverse square distance method, which takes into consideration all available rain gauges to calibrate radar estimates. In this method, the estimate of the surface function $f(\cdot)$ in a point s_0, knowing the function in the points s_i, with i =1, 2,..., k, is the weighted average.

$$\widehat{f}(s_0) = \frac{\sum_{i=1}^{k} d^{-2}(s_0, s_i) f(s_i)}{\sum_{i=1}^{k} d^{-2}(s_0, s_i)}$$

where $d(s_0, s_i)$ is the Euclidian distance between the points s_0 and s_i.

3.4. Models performance assessment

The models performance assessment is done according to the empirical square root of the mean square error of point prediction by using the fourteen storms data (116 hourly observations) kept for this purpose. That is, for each gauge not used in the calibration procedure, it is compared the gauge rainfall estimates G_t with the calibrated radar cell measurement $\widehat{R}_t^{(m)}$, with m = LC, FB

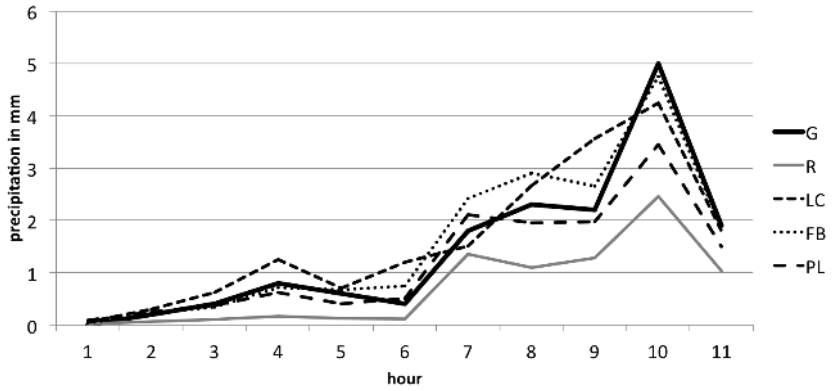

Figure 6. Validation rain gauge of Meca calibrated using Olhalvo and Abrigada gauges for the storm of November 2, 2000.

and PL, at the same location. As there are five rain gauges available in the area under study, it is considered systems with 1, 2, 3 or 4 gauges for the calibration process, and for each of these schemes is computed the empirical square root of the mean square error for all combination of each number of gauges.

The empirical square root of the mean square error $\text{RMSE}_k^{(m)}$ for the scheme modeled based on k rain gauges with the model m is computed by

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

Table 4 compares the square roots of the empirical MSE between the three models by considering different numbers of rain gauges in the calibration process. The pre-calibration RMSE of the five rain gauges is taken as the reference value to analyze the impact of the calibration procedures. For the 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 4 in brackets). These results have to be interpreted as a global measure of the models performance since these values do not take into consideration the relative position of both rain gauges of calibration and calibrated radar predictions.

It can be stated that the models lead to a significant reduction in the error of the radar estimates of precipitation. However, an interesting conclusion

Table 4. Comparing square roots of the empirical mean square errors of the three models. In brackets is indicated the % of RMSE reduction in comparison 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%)

that may be drawn is that the model PL is less sensitive to the number of rain gauges used in the calibration process. RMSE of both models LC and FB decrease significantly when more gauges are added to the calibration process. When the rain gauge density is the lowest (1 gauge per 140Km^2) the PL model performed the largest RMSE reduction with a strong difference to other models. On the other hand, when the higher density (1 gauge per 35Km^2) is considered, the models present similar performances. Nevertheless, the FB model produces the greatest RMSE reduction.

4. Conclusion

It can be stated that the Kalman filter approach with any of the three models leads to a significant reduction of point estimation of the precipitation. The state space models employed in this approach, have shown to be a useful tool in the improvement of radar data accuracy. The combination of the two sensors (radar and rain gauges) allows conjugating the advantage of the area estimation of the radar with the accuracy of the point prediction of the rain gauges.

The primary objective in this chapter has been to provide a fair comparison between three alternative models that have been applied in the calibration radar estimates of precipitation by using the Kalman filter algorithm. Taking a single data set in the modeling process for all models allows to assess the models performances with equal conditions. Both the linear calibration model and the mean field radar rainfall logarithmic bias model were more sensitive to gauge density. These models performances have improved with the increase

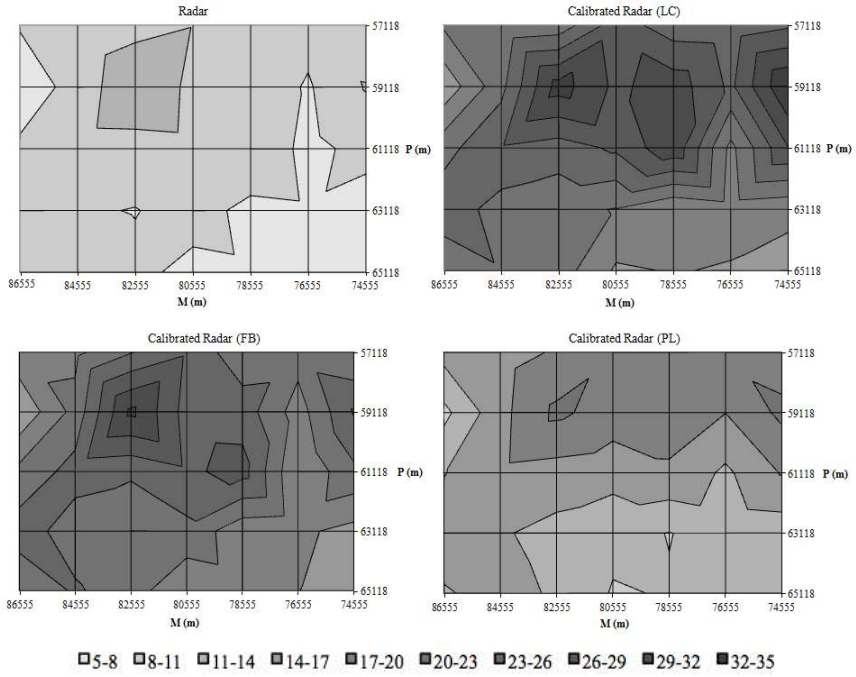

Figure 7. Accumulated precipitation isopleths, in mm, for the storm of November 2, 2000 for the radar non calibrated estimates and for the calibrated estimates considering the five rain gauges and the three models LC, FB and PL.

in the number of gauges in the calibration procedure. On the other hand, the performance of the power law model was smoother as the gauge density was increased. Figure 7 illustrates the radar calibration based on the accumulated precipitation of one storm by using all available rain gauges in the area.

In summary, the use of state space models associated to the Kalman filter algorithm is an efficient approach to increase the accuracy of weather radar estimates. Moreover, this approach may be implemented in a real-time scheme where models are established based on an independent data set of the calibration process, which continues to develop as new data is available.

Acknowledgements

Marco Costa was partially supported by Fundação para a Ciência e a Tecnologia, PEst OE/MAT/UI0209/2011. A. Manuela Gonçalves was partially financed by FEDER Funds through "Programa Operacional Factores de Competitividade – COMPETE" and by Portuguese Funds through FCT - "Fundação para a Ciência e a Tecnologia", within the Project Est-C/MAT/UI0013/2011. 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 Portuguese Foundation for Science and Technology ("FCT–Fundação para a Ciência e a Tecnologia"), within project PEst-C/MAT/UI4106/2011 with COMPETE number FCOMP-01-0124-FEDER-022690.

References

[1] Alpuim T. & Barbosa S. (1999). The Kalman filter in the estimation of area precipitation. *Environmetrics*, 10:377-394.

[2] Anagnostou, E.N. & Krajewski, W.F. (1998). Calibration of the WSR-88D precipitation processing system. *Weather Forecasting*, 13:396-404.

[3] Austin, P. (1987). Relation between measured radar reflectivity and surface rainfall. *Monthly Weather Review*, 115: 1053-1071.

[4] Babak, O. & Deutsch C.V. (2008). Statistical approach to inverse distance interpolation. *Environmental Research and Risk Assessment*, 23: 543-553.

[5] 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.

[6] 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.

[7] Collinge V.K. & Young P.C. (1993). The calibration of weather radars using a time series approach. In *Concise Encyclopaedia of Environmental Systems* (ed. P.C. Young), 706-710. Oxford: Pergamon.

[8] Costa M. & Alpuim T. (2010). Parameter estimation of state space models for univariate observations. *Journal of Statistical Planning and Inference*, 140: 1889-1902.

[9] Costa M. & Alpuim T. (2011). Adjustment of state space models in view of area rainfall estimation. *Environmetrics*, 22:530-540.

[10] Haberlandt, U. (2007). Geostatistical interpolation of hourly precipitation from rain gauges and radar for a large-scale extreme rainfall event. *Journal of Hydrology*, 332:144-157.

[11] Harvey, A.C. (1996). *Forecasting Structural Time Series Models and the Kalman Filter*. Cambridge University Press: Cambridge.

[12] Jordan P., Seed A. & Austin G. (2000). Sampling errors in radar estimates rainfall. *Journal of Geophysical Research*, 105(D2):2247-2257.

[13] Kalman, R.E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering, Transactions ASME*, 82(Series D):35-45.

[14] Krajewski W. (1987). Cokriging radar-rainfall and rain gauge data. *Journal of Geophysical Research*, 92(D8):9571-9580.

[15] Krajewski W.F. & Smith, J.A. (2002). Radar hydrology: rainfall estimation. *Advances in Water Resources*, 25:1387-1394.

[16] Rossa A., Bruen M., Fruehwald D., Macpherson B., Holleman I., Michelson D. & Michaelides S. (2005). Use of Radar Observations in Hydrology and NWP models. Document external project: 2005, COST 717 Final Report, EUR 21954, 292 pp, Brussels.

[17] Severino J. & Alpuim T. (2005). Spatiotemporal models in the estimation of area precipitation. *Environmetrics*, 16:773-802.

[18] Shumway R., Stoffer, D. (1982). An approach to time series smoothing and forecasting using EM algorithm. *Journal of Time Series Analysis*, 3:253-264.