MODEL-ERROR ADAPTIVE PARAMETER DETERMINATION OF A CONCEPTUAL RAINFALL PREDICTION MODEL

Konstantine P. Georgakakos

Reprinted from IEEE 1984 PROCEEDINGS OF THE SIXTEENTH SOUTHEASTERN SYMPOSIUM ON SYSTEM THEORY

IEEE COMPUTER SOCIETY
1109 Spring Street, Suite 300
Silver Spring, MD 20910

IEEE COMPUTER SOCIETY PRESS

MODEL-ERROR ADAPTIVE PARAMETER DETERMINATION OF A CONCEPTUAL RAINFALL PREDICTION MODEL

Konstantine P. Georgakakos
National Research Council - NOAA Research Associate

Hydrologic Research Laboratory, NWS, NOAA

ABSTRACT

Parameter estimation of a stochastic precipitation model is based on the study of contour maps of various performance indices in the parameter space. For each parameter set, a stochastic approximation algorithm is used to determine the best model-error spectral density. Results show the robustness of the parameterization to changes in storm type, topographic regime and performance criteria.

1. INTRODUCTION

A one-dimensional stochastic rainfall prediction model is analyzed in the following. The state-space form of the model equations were developed[1] based on atmospheric thermodynamics and cloud microphysics principles. Using surface pressure, temperature and dew-point temperature, the model gives as an output the average precipitation rate in the area characterized by the temperature and pressure input. The model formulation is based on pseudo-adiabatic ascent of the air-masses and on simplified cloud microphysics with exponential particle-size distribution and linear dependence of the particle terminal fall-velocity on the particle diameter. Evaporation of the falling particles, for unsaturated sub-cloud layer is explicitly taken into account by the model.

This work presents a procedure for the identification of the two free model parameters. Contours of various performance indices are examined in the parameter space. A stochastic approximation of the appropriate model-error spectral density Q is computed for each parameter set in the parameter space examined. In this way, adjustment of the model error statistics is done automatically for different parameter sets so that, for every parameter set, the Q-optimized value of each performance index is computed.

Following the presentation of the model equations in the next section, section 3 presents the parameter identification strategy and calibration data. In section 4 the results of parameter estimation are discussed. Verification is offered in section 5 where the calibrated model is used in real time to predict hourly precipitation rates. Section 6 presents conclusions drawn from this study.

2. MODEL FORMULATION

The scalar model equations are:

$$\frac{dX(t)}{dt} = f_t(\underline{u}) + A_t(\underline{u}) \quad X(t) + W(t) \tag{1}$$

$$Z(t_k) = H_k(\underline{u}) \quad X(t_k) + V(t_k); \; k=1,2,... \tag{2}$$

where,

$X(t)$: Mass of the condensed liquid water equivalent in the area characterized by the temperature and pressure input $\underline{u}$ at time t (model state).

$\underline{u}$: Model input vector such that $\underline{u}^T = [T_o, p_o, T_d]$, with T_o, p_o, T_d, the current surface temperature, pressure and dew-point temperature, and upperscript T denoting transpose of a vector or matrix quantity.

$f_t(\underline{u})$: Rate of moisture input to the clouds computed from the pseudo-adiabatic ascent of surface air characterized by T_o, p_o, and T_d.

$A_t(\underline{u}) \quad X(t)$: Input dependent cloud-moisture depletion rate.

$Z(t_k)$: Observations of the spatially averaged precipitation rate in the area characterized by $\underline{u}$. The observations take place at regular intervals $\Delta t = t_{k+1} - t_k$.

$H_k(\underline{u}) \quad X(t_k)$: Model-predicted average precipitation rate in the area characterized by $\underline{u}$ at time t_k.

$W(t)$: White-noise error input with zero mean and spectral density Q.

$V(t_k)$: White-noise error sequence with zero mean and variance R.

Expressions for $f_t(\underline{u})$, $A_t(\underline{u})$ and $H_k(\underline{u})$ have been derived.[1]

The two free model parameters are:

0094-2898/84/0000/0111$01.00

1) The ratio EPS1 of the updraft velocity to the square root of the potential thermal energy per unit mass of the ascending air at the height of average updraft velocity, and

2) The time- and storm-constant cloud-particle average diameter denoted by EPS4.

The function $f_t(\underline{u})$ depends on EPS1, while $A_t(\underline{u})$ and $H_k(\underline{u})$ depend both on EPS1 and EPS4.

As EPS1 increases, the updraft velocity increases giving a higher moisture input rate $f_t(\underline{u})$.

The primary mechanism that determines $A_t(\underline{u})$ and the $H_k(\underline{u})$ is the fall of the cloud particles in the updraft field. Therefore, the larger the particles (consequently heavier) the higher the precipitation rate is. On the other hand, the greater the updraft, the lower the precipitation rate becomes. As a consequence, when EPS1 increases, $A_t(\underline{u})$ and $H_k(\underline{u})$ tend to decrease, while, when EPS4 increases, $A_t(\underline{u})$ and $H_k(\underline{u})$ tend to increase.

Note that the behavior of $H_k(\underline{u})$ when EPS1 and EPS4 are varying may lead to non-observability conditions for the precipitation model. For high EPS1 and low EPS4, $H_k(\underline{u})$ tends to zero exponentially, making the model state unobservable from observations of the precipitation rate.

The model error noise W(t) in Eq. (1) represents 1) random errors in the model structure due to simplification of the physical mechanisms, 2) random errors due to inaccurate specification of the model parameters and 3) random errors in the input variables.

The observation noise $V(t_k)$ in Eq. (2) represents: 1) random errors in the input variables, and 2) random errors in the observations of the precipitation rate.

The continuous dynamics-discrete observations Extended Kalman filter is used as the state estimator.[1] Important for the present study is the fact that the filter parameters Q and R need to be determined together with the optimal set of parameters EPS1 and EPS4.

Once EPS1, EPS4, Q and R have been determined, the prediction step of the Extended Kalman filter can be used to yield a forecast of the model state and of the precipitation rate in real time.

3. PARAMETER IDENTIFICATION STRATEGY

The calibration data consists of storm-data from the meteorological station at Logan Airport, Boston, Mass. Hourly values of T_o, p_o, and T_d together with hourly, average precipitation rate for several storms are used. The storms were divided into two groups:

-- A Convective Group (CG) consisting of a line-storm and a tropical storm with a total of 110 wet-hours.

-- A Stratiform Group (SG) consisting of low-pressure frontal storms with a total of 125 wet-hours.

The CG group had a time-average of 2.4 mm/hour and a time-standard deviation of 2.7 mm/hour. The SG had a time-average of 1.8 mm/hour with a time-standard deviation of 2.5 mm/hour.

Given that the observation errors associated with T_o, p_o, and T_d were rather small with respect to other errors modeled by W(t) and $V(t_k)$, the input was assumed error free. A time-constant standard error of 1 mm/hour was used for the precipitation rate observation (R = 1 mm/hour).

Since a large part of the noise level in Q is due to the errors from inaccurate parameter specification, an adaptive scheme was used to obtain a reasonable Q-value for a certain parameter set (EPS1, EPS4).

The stochastic approximation iteration[2] is given next (k = 0,1,2,...):

$$\hat{Q}(t_{k+1}) = \hat{Q}(t_k) + \frac{S(t_k)\quad g(t_{k+1})}{g^2(t_{k+1})\quad S(t_k) + \sigma^2(t_{k+1})}$$

$$\left[r^2(t_{k+1}) - H^2_{k+1}\ (\underline{u})\quad P(t_{k+1}|t_k) - R\right] \qquad (3)$$

and

$$S(t_{k+1}) = \left[\frac{g^2(t_{k+1})}{\sigma^2(t_{k+1})} + \frac{1}{S(t_k)}\right]^{-1} \qquad (4)$$

with

$$\sigma^2(t_{k+1}) = 2\left[H^2_{k+1}\ (\underline{u})\quad P(t_{k+1}|t_k) + R\right]^2 \qquad (5)$$

$$g(t_{k+1}) = H^2_{k+1}\ (\underline{u})\quad \frac{\partial P(t_{k+1}|t_k)}{\partial Q} \qquad (6)$$

$$\frac{\partial P(t_{k+1}|t_k)}{\partial Q} = \frac{\exp\{2A_k\ (\underline{u})\ \Delta t\}}{\left[\dfrac{H^2_k\ (\underline{u})\quad P(t_k|t_{k-1})}{R} + 1\right]^2}$$

$$\frac{\partial P(t_k|t_{k-1})}{\partial Q} + \frac{\exp\{2A_k\ (\underline{u})\ \Delta t\} - 1}{2A_k\ (\underline{u})} \qquad (7)$$

$$A_k(\underline{u}) \triangleq A_{t_k}(\underline{u}) \tag{8}$$

$P(t_{k+1}|t_k)$ is the predicted state variance at time t_{k+1} given information up to and including time t_k, and $r(t_{k+1})$ is the 1-step predicted residual at time t_{k+1}. $\hat{Q}(t_k)$ represents the current, at time t_k, estimate of Q.

The stochastic approximation algorithm presented forces compatibility between the filter-predicted variance of the innovations and the actual residuals variance.

Initial conditions used for the initiation of iterations were:

$$\hat{Q}_o = \frac{1}{\Delta t} \left[f_t^2(\underline{u}) \mid A_t^2(\underline{u}) \right] \Bigg|_{t = t_o} \tag{9}$$

$$S_o = (o.9 \times \hat{Q}_o)^2 \tag{10}$$

$$\frac{\partial P_o}{\partial Q} = \frac{\exp\{2A_o(\underline{u})\,\Delta t\} - 1}{2A_o(\underline{u})} \tag{11}$$

The choice of $\hat{Q}_o$ presented scales the initial estimate of $\hat{Q}_o$ according to the particular $f_t(\underline{u})$ and $A_t(\underline{u})$ functions for each choice of EPS1, EPS4. The initial estimate S_o confirms the large uncertainty associated with $\hat{Q}_o$.

The parameter estimation strategy was to construct contours of several performance indices in the parameter space (EPS1, EPS4), for each of the storm groups CG and SG. The Q matrix was estimated for each parameter set by the stochastic approximation algorithm developed.

Three performance criteria were used in an effort to examine different aspects of the model performance.

Errors in the total mass of each storm-group precipitation were represented by the absolute proportional error (APME). This criterion is the absolute value of the ratio of the 1-step predicted residuals mean to the mean of the corresponding observations for the period under study. A value of zero represents optimal performance with respect to this criterion.

The standard least-squares criterion is represented by the proportional standard error (PSE). It is the ratio of the 1-step predicted residuals standard deviation to the standard deviation of the corresponding observations. It gives the proportion of the observations standard deviation unexplained by the model. A value of zero corresponds to perfect performance with respect to PSE.

Maximum likelihood estimation is represented by the average value of the log-likelihood (ALL) over the period of interest. The greater the value of this criterion the better the model performance is. Optimization with respect to ALL gives the parameter values with the highest probability of generating the observed sequence under the assumption that the model structure is the true one.

4. PARAMETER IDENTIFICATION

The space of the two free model parameters was divided in grids and the value of each performance criterion was computed for each model grid-point for each of the two storm groups (convective and stratiform). The discretization intervals were $0.24\text{x}10^{-3}$ for EPS1 and $0.13\text{x}10^{-4}$m for EPS4. Parameter EPS1 ranged from 10^{-4} to $0.5\text{x}10^{-2}$ while parameter EPS4 ranged from 10^{-5}m to $0.2\text{x}10^{-3}$ m. The intervals chosen contain physically reasonable values for the two parameters.

Table 1 presents the optimal parameter sets and associated criterion values for all cases. Also shown are the intervals ΔEPS1 and ΔEPS4 of EPS1 and EPS4 that contain the "optimal" set of parameters for each performance criterion and storm group, and for which: APME was less or equal to 0.1, and PSE, ALL were within 5 percent of their optimal value.

Characteristic of the values in Table 1 is the fact that at the 5% level of the performance indices PSE and ALL and for values of APME less than 0.1 considerable overlapping of the "optimal" parameter regions occurs. This suggest robust model structure and parameterization with respect to storm-types and performance criteria. The cross-section of the "optimal" parameter regions for all cases was:

$$\text{ESP1:}\quad 14\text{x}10^{-3} \text{ to } 2.1\text{x}10^{-3}$$

$$\text{EPS4:}\quad 0.45\text{x}10^{-4}\text{m to } 0.6\text{x}10^{-4}\text{m} \tag{12}$$

Table 1 also shows that similar performance-criterion values resulted for the two storm-types.

Contour plots of the final Q values and of the coefficient of variation of Q during the adaptation period suggested:

1) Dependence of Q on the parameter set, with final Q values ranging from 35 down to 0.005 $\text{kg}^2/(\text{m}^4 \text{ x sec})$, and

2) Coefficients of variation ranging from about 1.5 down to 0.1 with a parameter-space spatial-average value of 0.4 for both storm groups.

Table 1

SUMMARY OF PARAMETER-SEARCH RESULTS

	APME		PSE		ALL	
	CG	SG	CG	SG	CG	SG
OPTIMAL CRITERION VALUE:	0.002	0.005	0.74	0.78	-2.02	-2.05
OPTIMAL EPS1 ($x10^3$):	1.32	1.08	2.31	4.5	1.57	1.32
OPTIMAL EPS4 ($x10^4$):	1.49	2.0	0.73	0.48	0.61	0.48
ΔEPS1 ($x10^3$):	1.05 TO 1.82	0.84 TO 1.82	0.84 TO 4.50	0.84 TO 5.0	0.84 TO 3.27	0.84 TO 3.27
ΔEPS4 ($x10^4$):	0.29 TO 2.0	0.35 TO 2.0	0.48 TO 1.05	0.42 TO 0.67	0.42 TO 1.05	0.42 TO 0.73

These observations confirm the necessity of the use of different Q values for different parameter sets if one is going to draw meaningful conclusions from contour plots of parameter-estimation performance criteria. The low coefficient of variation of Q for the adaptation period and for both storm groups indicates successful choice of the initial conditions for the stochastic approximation algorithm. Given the exponential decay of $S(t_k)$ in Eq. (4), it also suggests short adaptation period.

For the "optimal" region of parameters in Eq. (12) (SI units) the adaptation algorithm gave a final Q value:

$$\hat{Q}_{Final} = 0.01 \quad Kg^2|(m^4 x \text{ sec}) \qquad (13)$$

with the $\hat{Q}_{Final}$ remaining within 10% of the value in Eq. (13) for all the sets (EPS1, EPS4) within "optimal" parameter region.

5. VERIFICATION

Several storms from the meteorological station at the International Airport in Tulsa, Oklahoma, formed the verification data. This location represents a radically different climatic and topographic regime from Boston, Mass., used in model calibration.

Hourly forecasts of the precipitation rate were obtained from the stochastic precipitation model with parameter values:

$$EPS1 = 1.65x10^{-3}$$

$$EPS4 = 5.5x\ 10^{-5}m$$

and

$$Q = 0.01\ kg^2/(m^4 x \text{ sec})$$

Figure 1 shows the hourly forecasts of the precipitation rate in mm/hour (dashed line) together with the corresponding observations.

For the verification run, APME took the value 0.21 and PSE was equal to 0.83. The variance of the normalized residuals (residuals divided by the filter predicted variance) was equal to 1.2 and the log-1, log-2, and log-3 correlation coefficients of the normalized-residual time-series were equal to -0.13, 0.02, -0.03, indicating near optimal filter performance.

The model performance was compared to simple persistence (predictions are the current observations of precipitation rate) and linear extrapolation (using the last two observations of precipitation rate to predict one-step ahead) models. The ratios of the stochastic model residual variance to the variances of the persistence and extrapolation models were 0.92 and 0.38 showing improvement over those prediction models.

Reasonable values of the physical model components were observed in the verification run. For example, the updraft velocity had an average value of 0.12 m/sec, the average liquid water content of the cloud column was 1.39 grams/m^3 and the cloud height averaged 5.86 km.

Noteworthy is the fact that the verification values of the performance indices APME and PSE are close to the optimal values obtained during the calibration stage for the climatic regime of Boston, Mass. This supports the conclusion that the model structure and parameter values are robust with respect to changes in climatic and topographic regime.

6. CONCLUSIONS

The two free model parameters of a scalar stochastic precipitation model[1] were estimated for a wide variety of storms. Simultaneous estimation of the model error spectral density was performed based on a stochastic-approximation iterations scheme that used parameter dependent initial conditions for the iterations.

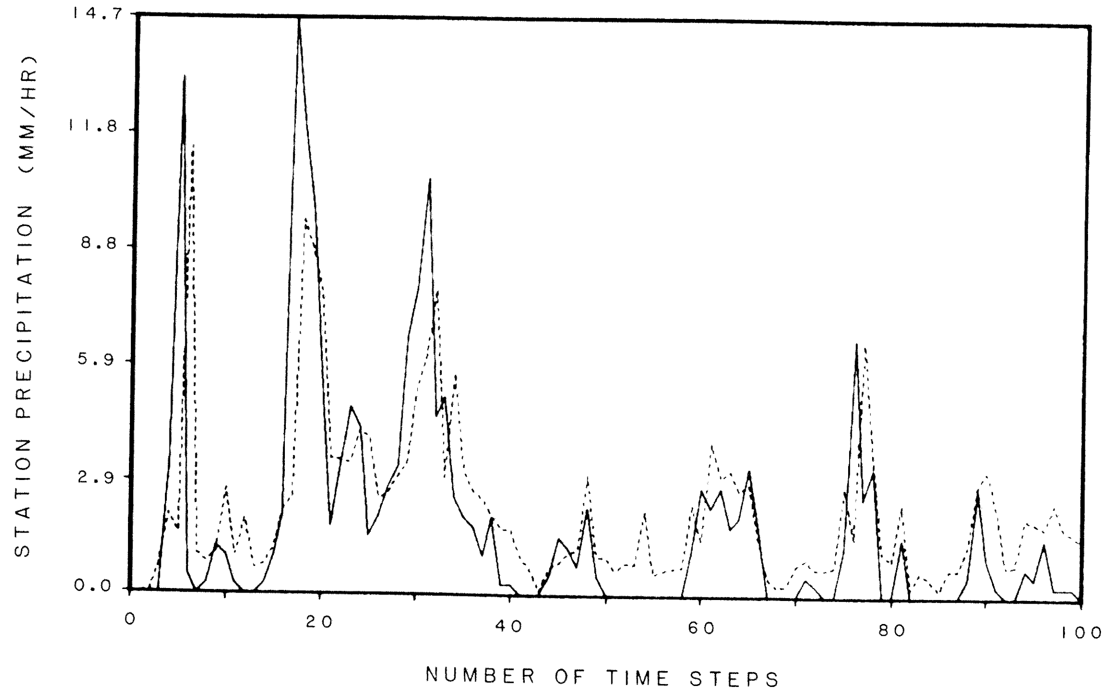

Figure 1. Forecasts (dashed line) vs. observations (solid line) for the verification storm group (TULSA, OK). One-hour time steps.

The contour maps of several performance criteria in the parameter space showed robust model structure in the forecast of hourly precipitation rates for several performance criteria and various storm types (convective and stratiform).

A verification run with data from a different climatic and topographic regime indicated that the model parameters are reasonably location independent.

Therefore, the model does not appear to require recalibration for different storms and locations. This is especially convenient for real-time forecasting uses.

Contour maps of Q showed dependence of the optional Q-value on the model free parameters. This suggests that determination of the appropriate Q for each parameter set in the parameter space is vital, if one wants to draw meaningful conclusions from contour maps of performance criteria in the parameter space.

Acknowledgements: This work was done with the support of the National Research Council and the National Oceanic and Atmospheric Administration.

REFERENCES

[1] Georgakakos, K. P., and R. L. Bras, A precipitation Model And Its Use In The Real-Time River Flow Forecasting, Ralph M. Parsons Laboratory, Hydrology and Water Resources, MIT Technical Report No. 286, July 1982.

[2] Kitanidis, P. K., and R. L. Bras, Real-Time Forecasting of River Flows, Ralph M. Parsons Laboratory, Hydrology and Water Resources, MIT Technical Report No. 235, June 1978.