Hydrol. Earth Syst. Sci. Discuss., 2, S320–S325, 2005
www.copernicus.org/EGU/hess/hessd/2/S320/
European Geosciences Union

Hydrology and Earth System Sciences Discussions

HESSD

2, S320–S325, 2005

Interactive Comment

Interactive comment on "Using a spatio-temporal dynamic state-space model with the EM algorithm to patch gaps in daily riverflow series, with examples from the Volta Basin, West Africa" *by* B. A. Amisigo and N. C. van de Giesen

B. A. Amisigo and N. C. van de Giesen

Received and published: 24 June 2005

We thank both reviewers for their constructive comments. We have tried to accomodate all comments as much as possible. Specifically, we tried to balance the theoretical treatment and the application side. Because we want to provide the readers with a complete overview of the theory, this meant that the article has become quite a bit longer than the original paper. Our detailed responses are summarized below.

Reviewer 1

Comment Response

1 The Bayesian analysis is a general procedure for estimating quantities such as sys-

tem states, cast in a statistical framework. It requires specification of prior conditional probabilities of the states being estimated given the parameters and observations. If the dynamic system is considered linear with Gausian initial system state and errors, then the Bayesian analysis can be reduced to the Kalman Filter and smoother. This simplifies the analysis to the estimation of the mean and covariance of the system states instead of the estimation of the conditional probabilities of the states at each time step. This is the simplification introduced in the data infilling procedure we adopted. This simplification is due, in part, to the difficulty in specifying an appropriate prior distribution for the system states (or surrogate states, such as catchment daily runoff) in our analysis. This difficulty has now been clarified in the text.

2 The model parameters are not of specific interest in this particular hydrological application and, therefore, we paid no attention to the associated uncertainties. For the sake of completeness, we have now added remarks in the text that point to how, if needed, parameter uncertainty estimates may be obtained by including parameters in the state vector or by means of Markov Chain Monte Carlo simulations.

3 The present simulations were run with unconstrained Q. We also trried to unconstrain R, but this resulted in non-convergence. We added remarks to this effect in the text. The probable reason for the non-convergence is that since the target station has missing runoff values, the information content of the remaining runoff at the station used in the analysis may not be sufficient.

4 The objective of the paper is twofold:

1) Introduce the Kalman smoother/EM approach to the hydrological public, and

2) Show the usefulness of this approach when it is applied to filling data gaps in runoff series.

We emphasize these dual objectives now a bit more in our introduction. Because of the first objective, we give a relatively extensive overview of the theory so people can

HESSD

2, S320–S325, 2005

reproduce it on the basis of the information given here. We agree that in first instance there was no balance between theory and application. Therefore, we expanded the analysis/results part by performing quite a bit of extra analysis with Lawra runoff to better explain the strengths of the model. Minor comment The Euclidean norm was used for the convergence tests (remark to this effect has been added to the text).

Reviewer 2

Comment Response

1 We have now added a simple gap-filling method (linear regression) for comparison, which shows the advantage of the proposed method. The modified write-up of the results and discussion section with the extra results from the Lawra analyses show the strength of the model. It is shown that where the spatial correlation is not so good, the model can shift more to the information contained in the non-missing runoff part of the hydrograph at the target station and produce good estimates of missing values.

We used linear regression instead of approach suggested by Wang (2001) because the latter is meant for record augmentation and uses only summary statistics. Comparing with Wang (2001) would, therefore, not really have been "fair". A second advantage, that we had not mentioned before, is that the proposed method produces standard error estimates, something other methods would not be very good at.

2 We agree that the theoretical treatment is somewhat extensive. In order to remediate this we did the following:

- The goal of the article is twofold: a) Introduction of this technique to the field of hydrology, and b) give a practical application to show its usefulness for a problem the authors encountered (missing data in Volta record). This has been emphasized now in the introduction. Because of the first goal, we feel it would be worthwhile to present the complete theory, in order for hydrological readers to be able to apply the theory without having to have recourse to (many) other articles.

- To improve the readability we adjusted the headings as suggested by the reviewer. We also added a flow-diagram that should aid the reader in following the reasoning.

- We added the comparison with a simple model (linear regression, as stated above) which deepened the application side of the article and giving it more balance.

3 The research arose as a consequence of us working on the hydrology of the Volta Basin. Specifically, we are trying to develop data-driven modeling approaches, which are greatly hindered by even small gaps in the runoff series. We would really prefer to remain "faithful" to our original study object, the Volta. The application side has now been enriched with a simple model thereby better showing the advantages of what seem a relatively elaborate approach.

Specific comments

Title We shortened the title as suggested.

Abstract Word added as suggested.

Page 2 -third sentence We agree and adjusted the text. For data-driven model development (an activity we are presently undertaking), complete time series are very convenient.

Page 2 - sentence 8 All methods refered to are described in Gyau-Boakye and Schultz (1994). This has now been made clear by adding the word "they".

Page 3 - 3rd sent. The Kalman smoother guarantees optimal use of available data when the underlying assumptions (Gaussian error distribution, Gaussian initial state vector) are true. Because we can not prove this to be true, we weakened our statement from "optimal" to "good"

Page 4 - “to be” removed

- Sub-heading positions proposed in the revised document

HESSD

2, S320–S325, 2005

Interactive Comment

- Alpha and beta stand for spatial locations; alpha for the x locations and beta for the y locations (text adjusted accordingly)

- “marks” replaced by “denotes”

Page 5 - Sentence rephrased

- The state’s convariance matrix is also called dispersion matrix and measures the uncertainty in the estimate of the state variable.

- Xt (with a subscript only) is always the actual state value; with also a superscript ( ) it is an estimated value. Yt is always observed, not simulated. This is clarified in the revised text.

- When the log-likelihood is used, as in eq 3, the parameters are estimated values and marked as such. For the Kalman-smoother, it is assumed that the parameters are known and the state vector is being estimated. Thus the parameters are supposed to be “actual” values and are not marked as estimators.

Page 7 - This is indeed a confusing issue. The model parameters are the parameters of the state-space model and should not be considered the same as parameters in "normal" hydrological models. This is now clarified in the text.

- The E-M algorithm for each r is run for all t - batch mode. This is now indicated in the revised text.

Page 12 This is indeed correct, sorry for the confusion. Runoff is being used as surrogate for catchment wetness, which is a state vector. Clarified with justification in the revised text.

Pages 13-14 Sec 6 Results and discussion extended in revised text.

Pages 13-14 Sec 7 Corrected.

Page 15 The need for high spatial correlation between the runoffs of the stations used has been de-emphasized in revised text since the Lawra data show that this may not be necessary for good results if the information content of the non-missing runoff at the target station is good enough. This is re-emphasized in the extended discussion.

Interactive comment on Hydrology and Earth System Sciences Discussions, 2, 449, 2005.