Nonlinear Processes
in Geophysics
Open Access
Discussions

Interactive comment on "An improved ARIMA model for hydrological simulations" *by* H. R. Wang et al.

R. V. Donner (Referee)

reik.donner@pik-potsdam.de

Received and published: 21 July 2014

In their manuscript, the authors present a statistical model for describing and potentially predicting monthly precipitation values from available time series. Their approach is based on a combination of cluster analysis and ARIMA modeling for each calendar month, thereby accounting for seasonality in the underlying data. Their approach is demonstrated to be superior to classical seasonal ARIMA models in a case study of precipitation data from some Chinese meteorological station.

Positively speaking, I don't find any significant scientific errors standing against a publication of this work. However, I am somewhat reluctant against recommending it for publication in NPG, given the aims and scopes of the journal as stated on its website: "...furthering knowledge on nonlinear processes in all branches of Earth, planetary

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

and solar system sciences" and/or applying "nonlinear analysis methods to both models and data". I think that the present contribution fails to address any of these two aspects - the authors neither provide any discussion of nonlinear processes, nor they apply any nonlinear methods. What they do is using rather classical linear statistical methods to the problem of precipitation forecasting - the latter could be considered a nonlinear problem, but nonlinearity of specific processes do not play any role in the present manuscript. Therefore, I doubt that the contents of the manuscript fully meet the scope of this journal. In any case, this is a question that needs to be finally addressed by the editor.

Beyond this general impression, I have several concerns regaring the presentation of the material that would call for some major revision of the text before it could eventually be considered ready for publication.

1. The authors speak about "hydrological simulations" in the title of their manuscript, but exclusively deal with precipitation, which would be a meteorological rather than hydrological variable. Of course, rainfall is important for hydrology, but the present title is not sufficiently specific and does not clearly reflect the contents of the paper.

2. Many aspects regarding the methods and data considered in this study are insufficiently explained, or statements are too vague to be actually convincing. For details, see below.

3. Sometimes, the authors are not specific enough when using statistical terminology (e.g., regarding the use of ARMA vs. ARIMA, stationarity, cluster analysis, etc.). In several cases, additional explanations need to be added in order to allow for a fair and complete evaluation.

Specific comments:

- p.843, ll.11-12: Do you mean an additive decomposition into stationary and non-stationary components here (e.g., obtaining stationary residuals after removing trends and seasonality in the mean)? I somehow doubt that such a decomposition always exists, especially in case of non-additive superpositions of different variability components.

- p.843, l.17: This is an ARMA model, not ARIMA.

- p.844, l.7: I recommend modifying the notation to differentiate between single observations and sequences, e.g., y_t vs. $\{y_t\}$.

- p.844, l.13: "quasi-periodic" has a clearly different meaning in the context of non-linear geophysics; seasonal variation is by definition periodic, not quasi-periodic.

- p.844, ll.18-23: Please explain the meaning of S and D in equation (6).

- p.844, l.22: What do you mean by "seasonal autocorrelation coefficient" - this term is not clear.

- p.845, l.2: Statistical models can only be "estimated", not "calculated".

- p.845, ll.4-6: Which notion of stationarity is used here? Why don't the authors apply any statistical tests for stationarity if this is a crucial point, but restrict themselves to rather "qualitative" and "visual" evaluation of stationarity?

- p.845, l.11: Reference Chen et al. (2004) seems to deal with ARFIMA models. Of course, ARIMA is a subset of ARFIMA, but I wonder if there are no better references. In any case, some more details on the estimator should be given.

- p.845, l.12: Which "white noise test" is used in the proposed procedure?

- p.845, l.16: The mathematical statement should rather read: $\exists k_0,\ t$ such that $corr(...) \neq 0$

- p.845, l.20: Number of cases of what?

Interactive
Comment

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

- p.845, ll.20-21: This appears just to be convention rather than a strict requirement. Is this correct? If so, please emphasize this fact more clearly.

- p.846, l.7: What is "linear least variance" - do you mean "linear least squares method"?

- p.846, l.8: What is L?

- p.846, ll.14-15: Please explain why information on inter-monthly variation may be lost in seasonal ARIMA models. There is still the auto-regressive part of the model that should account for such dependencies.

- p.847, ll.7-8: I suppose that the number of groups may depend on the specific cluster analysis approach utilized, as well as on the corresponding "model" selection criterion. Which methodological specifications are used in this work (should probably be addressed explicitly in Section 4.3).

- p.847, ll.9-10: Be more specific: maximum, minimum, truncated mean of what? linear regression model of what as a function of what?

- p.847, ll.11-12: By using such a strict classification according to the results of cluster analysis, information on similarities between months from different classes is completely neglected. However, depending on the specific clustering algorithm, there can still be significant similarities. Wouldn't it make more sense to replace the clustering step at all by some alternative approach taking into account the degree of similarity between all pairs of calendar months as weight factors in estimating statistical models. Naively, I would expect that such an approach could further reduce the model error substantially.

- p.848, l.10: How did you determine that 1/3 is a proper order of the power-law transformation?

- p.848, l.11: I disagree that the data become stationary due to a power-law transform, since this is a monotonous transform (in the case of the exponent used here). That is, if the original data are (non-)stationary, so are the transformed

data. This is probably related to a previous comment of mine: it is not clear which concept of stationarity is considered here and if it is rigorously tested for (apparently not) or just visually inspected.

- p.848, l.15: What kind of model selection approach has been followed here to determine the order of the ARIMA model in equation (10)? This is crucial information and must be provided in order to allow for a fair assessment of this manuscript.
- p.849, ll.15-16: In fact, the models listed in Tab. 4 are ARMA, not ARIMA.
- p.850, ll.4-5: Why is the offset from zero only considered for $\alpha \neq 0$?
- p.850, l.6: I disagree that the data in Fig. 11 are symmetric at all.
- p.850, ll.7-8: Which clustering approach is used here (see comment to p.847, ll.7-8)?
- p.850, l.12: Again - how has the presence of stationarity been tested for?
- p.850, ll.12-13: Please further justify why you expect to be able to extract more information from the data set of maxima. Stationarity alone appears no convincing justification.
- p.851, l.12: Please rephrase. If you would use information of the predicted month as well, this would not be a prediction anymore.
- p.855, Tab.2: The models considered here are seasonal ARIMA, not ARIMA - please be more specific. The content of the last six columns is not understandable from the table header.
- p.857, Tab. 4: These models are ARMA, not ARIMA. Since you work with precipitation, this is probably not too surprising, since memory of precipitation is commonly relatively short (less than one year, and so an integration term is probably not needed).
- p.858, Tab.5: From the caption and table header, it is not possible to infer the contents of the table completely.

Interactive
Comment

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

- pp.865-871: Terms like "line graph", "columnar section" and "function gram" (what is this?) are somewhat awkward in a caption and should be removed or rephrased.
- In general, careful proofreading is recommended. There are still some problems with the English, especially regarding missing articles and the improper use of participles.

Interactive comment on Nonlin. Processes Geophys. Discuss., 1, 841, 2014.