Atmos. Chem. Phys. Discuss., 9, C11841–C11858, 2010
www.atmos-chem-phys-discuss.net/9/C11841/2010/
© Author(s) 2010. This work is distributed under
the Creative Commons Attribute 3.0 License.

Atmospheric Chemistry and Physics Discussions

ACPD

9, C11841–C11858, 2010

Interactive Comment

Interactive comment on "Technical Note: Trend estimation from irregularly sampled, correlated data" *by* T. von Clarmann et al.

T. von Clarmann et al.

thomas.clarmann@kit.edu

Received and published: 31 March 2010

The authors like to thank both reviewers for the careful assessment of our manuscript and the detailed comments. In our reply, we have inserted the original review in bold face in order to avoid excessive cross-referencing. Our reply is printed in normal face.

Reply to reviewer 1:

General
The subject of this Technical Note (TN) is very important for the atmospheric science community and is certainly worth of publication in an ACP Technical Note.

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

Correlations among measurements are too often disregarded, unnecessarily growing the error budgets of the derived parameters / trends. Despite the good scientific significance of the subject treated, I have, however, serious concerns regarding the scientific quality of the approach used. In particular, in the TN the discussion is always very theoretical and there is no occasion in which the presented theory is applied to an even small test dataset. If the authors plan to postpone the presentation of examples of application of their theory to a second forthcoming paper I do not agree with them for several reasons.

We are happy that the reviewer appreciates the importance of the topic. The reason that no case studies were included in the original manuscript was that it is difficult to find cases which are realistic (based on real data) and at the same time suitable for demonstration purposes. Most real data require consideration of too many aspects and their treatment thus is far too complicated to use these data to demonstrate applicability and basic characteristics of a method. Nevertheless, we have meanwhile identified a dataset suitable for case studies to demonstrate the applicability of the methods proposed. The case studies will be included in the revised version of the manuscript.

I have found some (too many !) errors / imperfections in the presented equations, and I cannot guarantee that I found "all" of them. I have seen that in some occasions the main author of this paper delegated the reviewers to do an extensive debugging of his work. I do not agree with this approach. Instead, together with the theory, the authors should present a set of convincing examples in which their theory is successfully applied to test data. On their own, numerical examples are a kind of validation, increasing the trust of the reader on the presented theory.

ACPD

9, C11841–C11858, 2010

Interactive Comment

We regret the errors and inaccuracies and we are confident that the formalism will be sufficiently accurate in the revised version. Applicability of the schemes proposed will be demonstrated on the basis of numerical case studies based on real measurement data.

The presented theory could easily embed unpredictable bottlenecks or problems that show-up only when the algorithm is implemented into a computer program and applied to data (e.g. convergence, stability problems, etc., see the specific comments below).

In principle we agree; tests show that the methods are robust. We have performed tests with synthetic data which show that the methods work as they are supposed to work, and we have performed tests with real data to demonstrate robustness and applicability. Examples of the latter will be included in the revised paper.

The authors claim that neglecting correlations has an impact on the derived trend. I agree, however, how large is the effect? If the effect is small one could decide that the complication of using the full error covariance is not worthwhile. The authors should quantify the roughness of the approximation of neglecting correlations with an example.

We disagree in this point. The roughness of the approximation of neglecting correlations depends on the example chosen, particularly how the data set is clustered; we cannot just report the importance as a function of the correlation length or correlation coefficient, because the importance of the correlations depends on the x-values of the correlated data subset. To judge whether the complication of using the full error covariance is worthwhile or not always needs the rigorous method as a reference, even if in a particular case the result may be that a simplified method might be sufficient. In this case the rigorous method is by no means obsolete but necessary to justify the simplified approach. What we can and will do is to give examples to prove that there exist cases where the complication of using the full error covariance is necessary. It would, however, not be appropriate to infer a general rule from these examples.

To conclude, I do not recommend publication of this TN in ACP due to the poor scientific quality of the approach used. Of course I am ready to change opinion if the authors decide for a major revision of their paper, with inclusion of several numerical examples addressing the above specified concerns and the specific comments outlined below.

We are improving the manuscript by inclusion of numerical examples. Suitable data sets have already been identified.

Specific comments

1. p.27679, l.1-3: it is not clear how the vector x is set-up.

Here x is a scalar representing the time of a single measurement. In compliance with the ACP format we use italic font for scalars, bold face italics for vectors, and bold face roman face for matrices. Eq. (3) is a scalar equation.

2. p.27679, l.15: .... =0

agreed, this will be corrected;

3. p.27679, l.16: the first “=” does not hold.

agreed, this will be corrected;

4. p.27680, l.12: this equation can be simplified to get the same form as the equation for a.

(should probably read “ same form as the equation for b”); the simplified form has been derived and will be included.

5. p.27681, Eq.(13): there are two serious errors showing that this formula was never tested in a practical case. Please check matrix products, some of them are not possible due to dimension mismatch. The transpose sign is in the wrong place.

The equation will be corrected. Since the error refers to the product of a matrix with a scalar, our computer code was not sensitive to this error.

6. p.27681, l.13-18: This sentence is too long, I was not able to follow it.

This sentence will be split.

7. p.2782: I was not able to follow the discussion here. Long sentences and lack of practical examples do not help the reader.

ACPD
9, C11841–C11858, 2010

CC BY

The text will be rewritten and examples will be added to illustrate the contents.

If u is a global or multi-site field, the covariance matrix of Eq.(14) characterizes the spatial variability of u within the considered sample of satellite measurements. Wouldn't it be reasonable to scale this matrix according to the actual distance between the two stations mentioned at l.2 of the same page?

With global satellite data, the correlations can be evaluated for the actual distances directly without any generalized assumption on their dependence on the distance between sites.

8. p.27683, ll.1-18: This reasoning is difficult to follow and rather speculative if not supported by a practical example.

An example will be added.

9. p.27684, ll.4-14: I do not see the reason why these derivatives are introduced here. At the end, the final analytical solution is not given and the lazy reader is invited to solve numerically the minimization problem. I would either remove equations from 18 to 21 or present also the final analytical solution (seems not too difficult to get). The user could also be interested to know if secondary minima of this χ^2 are expected and if the inversion is well-posed / conditioned. This could help in the choice of the minimization algorithm. Here I got, once again, the impression that the authors themselves never tried the approach they are proposing.

ACPD

9, C11841–C11858, 2010

CC BY

Eqs. (18-21) define a set of 4 linear equations with four unknowns (a...d) which, unless degenerated, can be solved by any linear algebra package. A dedicated minimization algorithm is not required. Since there is a *linear* system of equations to be solved, secondary minima can be excluded (c.f. reply to #16). Given the large number of variants of this approach (e.g. superposition of periodic functions of different wavelengths) it seems not helpful to have the (lengthy!) closed-form expression for a selected subset. The approach to set the derivatives zero and to solve the resulting set of linear equations to get the coefficients is quite general and can easily adopted for variants of the regression model. In the revised version we will present more details of the path towards the solution which we thought were self-evident. The robustness of the approach will be demonstrated in a case study.

10. p.27685, l.1-5: Again, this reasoning seems rather speculative if an example is not provided.

Meanwhile we consider the term "overtones" inappropriate because it suggests the periodic length of the second oscillation to be an integer fraction of the largest periodic length, which often is not the case; e.g. the QBO is not a exactly bi-annual oscillation. However, to come back to the point of the reviewer, we don't quite see why the extension of the method proposed towards multiple periodics might be speculative, once we'll have shown that the method works for a single superimposed periodic function.

11. p.27685, Eq.(22): Is this a scalar or vector equation? Please clarify and use consistent symbols, see also below.

ACPD

9, C11841–C11858, 2010

This is, as with all regression models presented in this manuscript, a scalar equation. In compliance with ACP format we use normal face italic for scalar variables, bold face italic for vectors, and bold face roman for matrices. The step from scalar regression models to vector equations for calculation of χ^2 is described after Eq. (4).

12. p.27685, Eq.(23): in this equation a scalar is added to a vector ... please correct as appropriate.

Indeed in Eqs. (23), (25), (28), and (29) it should read $a\vec{e}$ instead of a. This will be corrected.

How do you define exactly the vector c of this equation? It should be linked with cmonth(x) of Eq.(22), however the symbol used is different.

c_{month} is a scalar, containing, e.g. the typical difference between the average temperature in one particular month and temperature averaged over all months. $\vec{c}$ (bold face in Eq. (23); Attention: the font/format for 'vector' is different in this reply compared to the ACP style-file) is a vector (of length 12 in this example) containing all these monthly corrections for January ...December. This definition will be added to the text. The selection matrix **U** selects the applicable correction term from vector $\vec{c}$. E.g. if the time of the measurement is in January, the entry in **U** will be 1 in the column representing January and 0 in the other columns. Regrettably, the algebra in Eq. (23) (and 25, 28. and 29) is not consistent with definitions in the text. With $\vec{c}$ being a column vector, and **U** set up such that each column represents a month, and each row represents a data point, it should read $\mathbf{U}\vec{c}$. Linear dependence between equations involving $\partial\chi^2/\partial a$ and $\partial\chi^2/\partial c_j$ is remedied by constraining $c_1 + \ldots + c_J$ to zero.

ACPD

9, C11841–C11858, 2010

Interactive Comment

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

13. p.27685, l.17: Binning of what? In which domain?

Binning of data points in the time domain, in a sense that, e.g. the same 'January' correction is applied to each data point in January, regardless if it is January 1st, 15th or 31st. This context should be clear from the text above Eq. (22), where the term 'binning' was introduced and used in the context of monthly mean). Nevertheless we have specified it in the revised version when used the first time. Here we refer to the definitions and conventions introduced there.

Under which circumstances binning or averaging is to be avoided? Please provide an example.

If the data set available is composed of monthly means, binning is inherent. If the time resolution of the data set is better, and the difference of the corrections associated with subsequent months is large, binning is to be avoided. We will include a note on this in the text.

Here c became a scalar, in Eq.(23) it was a vector, please make a decision and then use consistent notations.

Whenever c is printed in normal italic face, it is a scalar representing one single correction. We use this notation in all proposed regression models. Whenever c is printed in bold italic face, it is a vector, being composed of all monthly corrections. This representation is used in all χ^2 expressions.

14. p.27685, l.24: month = month + 12 ... therefore I conclude 0 = 12 ?? Please

ACPD

9, C11841–C11858, 2010

Interactive Comment

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

correct.

This will be corrected.

15. p.27685, Eq.(25): same problem of Eq.(23).

See above; will be corrected.

16. p.27686, ll.7,8: In my view there is no guarantee of success here. There could be multiple minima of the cost function...

We do not see why the minimum here should be more ambiguous than with any other regression model. The only difference compared to Eq. (23) is, that here the applicable correction for, say, January 31st, is approximately half way between the January and the February correction. The discrete application of the correction is replaced by a continuous time-dependence. By allowing this interpolation in the regression function, the number of fit variables is not increased, nor is any non-linearity added to the problems. We have included an example to demonstrate that, after inclusion of the zero-mean constraint for the monthly corrections as with the binning approach, this approach is sufficiently robust.

The χ^2 function is the square of the residual of a *linear* model and the measurements weighted by the inverse measurement covariance matrix. Minimization leads to a 14×14 set of linear equations, which has an unambiguous solution unless the equations are linearly dependent (see below). Multiple minima are a characteristics of nonlinear models. The regression model used here is linear in its coefficients. The V

ACPD

9, C11841–C11858, 2010

Interactive Comment

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

formalism does not introduce any non-linearity.

... or the problem could be ill-posed.

The characteristics of the **U** formalism is that it adds J orthogonal equations to the problem. The **V** formalism is closely related to the **U** formalism. Ill-posedness as well as the underdetermined case are expected only if the correction for a certain month is tried for which no data are available. Ill-posedness is not an issue for typical applications of the methods proposed. We will demonstrate this with a typical example. Ill-posedness seems not to be inherent in the problem as such but seems to associated with the choice of inappropriate regression functions.

The authors should illustrate an example of successful application of this method to real data. This would give some confidence that the proposed approach is feasible, at least in some cases.

Such an example will be included.

17. p.27686, ll.12,13: is there some physical justification for smoothing the (scalar ?) function c? It will depend on what is c. Please illustrate a practical test case.

18. p.27686, Eq.(28): same problem of Eq.(23).

19. p.27686, ll.20,21: How to choose γ? The second order cyclic differences matrix would seem more appropriate here. Why do you suggest the first order differences matrix

20. p.27687, Eq.(29): same problem of Eq.(23). How would you determine D?

ACPD

9, C11841–C11858, 2010

CC BY

Have you ever tried this implementation or is just a speculation?

In none of our examples we have found evidence that regularization is actually necessary. Therefore, and because the inclusion of examples adds considerable length to the Technical Note anyway, we have decided to remove the section on regularization.

21. p.27687, Eqs(30-31): These equations are correct only if the derivatives appearing therein (vectors) are defined in a very unusual way. Please provide an appropriate definition of the derivative vectors and correct the equations if necessary.

The definition will be provided.

Note also that the usefulness of these equations depends on the method used for the minimization of the cost function. If a stochastic method is used (e.g. to avoid secondary minima) these expressions are useless.

None of our approaches includes a stochastic method, except those involving the Tikhonov approach which can be interpreted as stochastic. The latter will no longer be included in the revised manuscript. Thus we consider these expressions valid and useful.

22. p.27687, l.18: surprising small uncertainties... please show the example you have in mind here

It is well known that a constant bias in the data used for a time series does not affect the trend at all. This would be the extreme case of correlations. We have encountered

ACPD
9, C11841–C11858, 2010
Interactive Comment

this behaviour of uncertainties in tendency when the correlated fraction of the error became dominant. However, a lengthy explanation of this well-known issue will add unnecessary complication to the text, so we might consider to remove this statement.

23. p.27688, l.8: is the hypothesis of normally distributed errors really necessary?

Our application is the “Chi-square test for variance in a normal population”. In Rodgers ‘Inverse Methods for Atmospheric Sounding’, World Scientific, 2000, page 187 it is stated: “The χ^2 test is a way of testing whether a particular random vector belongs to a given Gaussian distribution...”. In our application the random vector is the vector of residuals. In general, the χ^2 test can also be used for other distributions, but the χ^2 formulation we have chosen (where the squared residual vector is weighted by the inverse covariance matrix of measurement error) is more specific, since it is based on the Gaussian distribution.

24. You state that if L is a cyclic first order differences smoothness constraint:
p.27688, l.14: the number of the degrees of freedom (of the χ^2, I guess) is the rank of the regularization matrix L^{T}L;
p.27688, l.15: rank[L^{T}L] = n $-$ i$_u$;
I think none of these statements is correct. Please include an analytical proof or provide numerical evidence of their validity. In alternative, if references exist where these statements are demonstrated, please cite them. I am sorry that I can not suggest what are the “correct” expressions here because Eqs. (28-29) need first to be clarified (as already mentioned above).

We have decided to skip this paragraph for the following reason: As written in the

ACPD

9, C11841–C11858, 2010

Interactive Comment

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

manuscript, χ^2 statistics leads to meaningful results only if the regularization term represents the true statistics of the differences of the state variables between subsequent months. In practice, the strength of the Tikhononv regularization usually is some kind of ad hoc choice but not based on stochastics. Thus this paragraph is obsolete anyway.

25. p.27688-27689, Sect.6, Application areas. Again: this section looks purely speculative if not supported by at least one pertinent example.

After inclusion of examples, this paragraph might be obsolete. We consider to delete it because the discussion of examples itself will provide a sufficient link to application.

Reply to reviewer 2 :

This paper describes a metholodololody for an accurate calculation of trends usinga dataset that has its own representation biases. A particular focus is the problem of using a dataset composed of observations from various sites, inhomogeneously representing the global distribution, with samples that do not cover the same time periods. This aspect is critical to avoid possible biases associated with trend estimates. Therefore, I would recommend publication in ACP but provided the following general comments be addressed, in particular that an illustrative example be added to the discussion.

We are happy that the reviewer appreciates the importance of the topic and we will include examples.

General comments: The paper is clearly written and the method and calculations

ACPD
9, C11841–C11858, 2010

CC BY

are presented in an educational way (although some errors should be corrected and some points could be clarified – see reviewer 1 and specific comments).

we will correct the errors and provide clarification where needed.

However, I fully agree with reviewer 1 that, even though this paper aims at presenting a methodology, it needs to include a practical example. An illustration would help the reader through this very technical paper but also – and probably more importantly – demonstrate the quality of (and the need for) this approach. This would not need to be a complicated trend analysis but rather a direct application on a simple case study along with sensitivity tests.

Useful data sets have already be identified and test cases will be included in the revised manuscript.

Another general comment is the possibility to use satellite observations for spatial variability evaluations (and hence correlations) to be included in the Sy covariance matrix. While it is true that satellite observations allow quasi global coverage with quite long records for several parameters (allowing applicability of the methodology proposed at different time periods), the question of vertical resolution should be mentioned. For temperature or water vapor, but also trace gases, etc., the vertical resolution in the troposphere is at best several kilometers, with lack of sensitivity to the surface layer for several sensors. Limb sounders allow better vertical resolution but with sampling only the ULTS and higher. Since in situ observations are often done at the surface, this may cause a series of additional difficulties. This should be discussed. Again, a specific example of application would be most helpful.

ACPD
9, C11841–C11858, 2010

We will discuss this issue in the revised paper.

Some more specific comments: I will here only detail the points which have not been raised by reviewer #1. p. 27677 – last sentence (first of following page): not clear although it is were the objectives of the paper is presented... I would recommend cutting this long sentence in 2.

We agree.

p. 27678, l. 18: again, to demonstrate this, an applied example of this effect would be very interesting! Has it already been highlighted? If so, it should be mentioned and a reference should be provided.

We are not aware of a published example but will present an example to demonstrate this.

p. 27681, eq 13: an introduction on how this equation was derived might be helpful.

As discussed above, two e-vectors in the denominators were missing in the original manuscript. This error will be corrected. The equation is just the derivative of b with respect to the measured data $\vec{y}$ multplied from the left and from the right onto the measurement covariance matrix, strictly following Gaussian error propagation.

p. 27682: the authors should discuss what parameters should be included

ACPD

9, C11841–C11858, 2010

Interactive Comment

in Ssat, and how the incompatibilities in vertical resolutions between surface measurements and satellite measurements may be accounted for.

Agreed, this issue will be raised and recommendations will be made.

p. 27682: References on how this is done in the literature would be useful.

I am not aware of examples on this in the literature.

p. 27684-86: this part is rather difficult to follow. An example and illustrative figures (on the methodology and the results) is, again, necessary. As already mentioned by reviewer #1, the notations to differenciate scalars/vectors/matrices should be reviewed.

An example will be added. Scalar, vector and matrix notation complies with the ACP guidelines.

p. 27687, l. 5: two 'and'...

This will be corrected.

Planned modifications to the manuscript beyond those requested by the reviewers:

1. We will avoid the term 'overtones', because it is misleading, as stated above.

ACPD

9, C11841–C11858, 2010

2. We will adjust the algebra of the term involving the selection matrices to the definition in the text in Eqs. (23), (25), (28) and (29).

3. In the original manuscript we considered periodic variations of constant amplitude only. In the revised manuscript we will present a method which considers also periodic variations of which the amplitude is a function of the state value itself. An example will be added to demonstrate the robustness of the proposed method.

Interactive comment on Atmos. Chem. Phys. Discuss., 9, 27675, 2009.

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper