

Interactive
Comment

***Interactive comment on “Application of the
Gaussian anamorphosis to assimilation in a 3-D
coupled physical-ecosystem model of the North
Atlantic with the EnKF: a twin experiment” by
E. Simon and L. Bertino***

E. Simon

ehouarn.simon@nersc.no

Received and published: 15 June 2009

Application of the Gaussian anamorphosis to assimilation in a 3D coupled physical-ecosystem model of the North Atlantic with the EnKF: a twin experiment.

E. Simon and L. Bertino

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper



1- General comments

- *The English is not perfect, and needs to be corrected. I suggest that the authors have their paper read by a native English speaker (two French authors and at least one French reviewer might be misleading). The sentences are sometimes too long. Also, there is an excessive use of logical articulation words.*

A rereading of the revised manuscript has been done by a a native English speaker.

- 1 *What makes the anamorphosis so successful ? Is it essentially because it constraints the variables to be positive, or is it more ? In atmospheric pollutant dispersion, one can show that the positiveness constraint is the major contribution to the non-Gaussian nature of the analysis. If so, would a truncated Kalman filter (Lauvernet et al., 2009, cited in the manuscript) just as useful as the anamorphosis ?*

The important success of the EnKF with Gaussian anamorphosis in our twin experiments results from the combination of two reasons. First the respect of the assumptions ensuring the optimality of the linear estimation. Analysis steps are made in a transformed space where the variables and the errors have a Gaussian distribution. It leads to analysis steps which never damaged the solution (the mean of the ensemble) while analysis with the plain EnKF were damaging the solution in winter time (first 3 months of the experiments). Secondly the ecosystem model NORWECOM does not allow to compensate (or "hide") little assimilation bias. On the contrary the strong non-linearities of the model tend to amplify them.

Naturally the positiveness constraint is an important contribution to the non-Gaussian nature of the analysis. Nevertheless, this is not the only one. Let

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper



assume that an ecosystem variable follows a truncated Gaussian distribution. According to the propriety that a Gaussian anamorphosis function of a Gaussian variable is linear, the shape of the anamorphosis function of the truncated Gaussian variable should be linear at least for a domain including the high Gaussian values. This is clearly not the case when looking at the anamorphosis functions of the nutrient variables (see figure 1 for example).

Furthermore we note that the shape of the anamorphosis functions of the chlorophyll-a and the two phytoplanktons are quite similar. The anamorphosis presents a curvature in the interval $[-1, 1]$ of the Gaussian space, affecting around 65% of the values (the transformed variables have a Normal distribution $\mathcal{N}(0, 1)$). Had the distribution been a truncated-Gaussian, the anamorphosis would have been a straight line, intersecting the abscissa.

The truncated Kalman filter of Lauvernet et al could be an useful method for data assimilation in ecosystem model. Nevertheless the algorithm appears to be more complicated than the simple variable transformations (anamorphosis function) that Bertino et al suggested. Furthermore the truncated filter needs theoretically to use a Gibbs sampler to sample the truncated Gaussian distribution and then an estimation of the location vector and the scale matrix (mean and covariance matrix of the truncated-Gaussian). Such algorithms can be very expensive for large systems. Unfortunately approximations suggested by Lauvernet et al to reduce the cost of the filter seem to be unsuitable for ecosystem models. For example the distribution of ecosystem variables of the NORWECOM model do not correspond to a truncation far tails of a Gaussian distribution. It means that we must compute the estimate of the parameters of the Gaussian distribution that we assume to truncate.

The formulation of the EnKF with Gaussian anamorphosis (Bertino et al) is simpler and relies on weaker assumptions on the distribution of the variables (transformation are computed from the empirical marginal distributions), making this

[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)[Discussion Paper](#)

algorithm more attractive.

- 2 *The localisation of the anamorphosis in time and space seems an important source of improvement. The idea put forward in the conclusion could be elaborated slightly more.*

We agree. The assumption of identically distributed variables is too strong. For example the vertical distribution of phytoplankton shows important differences between the upper part of the ocean (high concentration) and the deeper part (very low concentration due to the lack of light). Horizontal distribution are also different from an area to another. The benefits of a spatial refinement of anamorphosis function would be to improve the Gaussianity of the transformed ensemble used in each local analysis.

Nevertheless the way to do it is not clearly defined now. A solution could be the use of statistical classification tools in order to be able to build anamorphosis functions suitable for a set of local events. The operational framework adds also constraints on the efficiency and the costs of these approaches. In short, it is an open issue.

A short sentence has been added in conclusion.

- 3 *The observations are perturbed by a log-normal law, which presumably (?) lend itself well to anamorphosis. How do you expect the anamorphosis outcome to degrade with real errors, possibly multimodal ?*

We did not try other distributions than log-normal but we agree that the method is likely sensitive to this choice.

Real errors may be not Gaussian or not have a log-normal distribution. Nevertheless for the specific case of ecosystem models, the log-normal law appears to be suitable for modelling errors of ocean biogeochemical variables like the chlorophyll (Campbell, 1995). For the general case, the use of anamorphosis functions

based on the empirical marginal distributions of the variables can lead to more general observation errors than the simple log-normal distribution, as we assume that the variables have a continuous distribution, not necessary log-normal.

Campbell J.W. : The lognormal distribution as a model for bio-optical variability in the sea, *Journal of Geophysical Research*, 100 (C7), 13237-13254, 1995.

- 4 *A short appendix on the mathematical details of the anamorphosis would be helpful to most readers (see remarks below).*

As suggested, an appendix has been added in the revised manuscript.

2- Minor comments

- *p.619, l.29: "stochastic" → "probabilist"*

Done.

- *p.621, l.2: Reference for EnKF: Evensen could have been cited earlier at the first occurrence of EnKF*

Done.

- *p.621, l.9-l20: Other methods have been tested in geophysics in the context of data assimilation. For instance Hólm ("Assimilation and modelling of the hydrological cycle: ECMWF' status and plans", E. Hólm et al., ECMWF Technical Memorandum 383, September 2002, see specifically p.32-36) used a Gaussian anamorphosis for humidity in meteorological models. Bocquet uses a fully non-Gaussian data assimilation framework in air pollution modelling (e.g. "Inverse modelling of atmospheric tracers: Non-Gaussian methods and second-order sensitivity analysis" M. Bocquet, *Nonlin. Processes Geophys.*, 15, 127-143, 2008.)*

[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)[Discussion Paper](#)

The method in Bocquet seems relevant to our problem, but in a static context. Its implementation into a sequential data assimilation method is not obvious. Perhaps using the concept of error variance instead of relative entropy would make it easier to relate to the field. We decided not to include the reference until the link is clearer.

Hólm did not use really a Gaussian anamorphosis to deal with the non-Gaussianity of the humidity errors. He mentioned the possibility to do it in order to obtain a Gaussian control variable. But he did not use it in practice, the variables being "already close to Gaussian". He defined his control variable by a suitable normalization of forecast differences of relative humidity (what he called "symmetric δRH "). In that way, it is not really a reference of use of Gaussian anamorphosis in data assimilation.

- *p.621, l.26: "two last" → "last two"*

Done.

- *p.621, l8-9: The paragraph break should be avoided here.*

Removed.

- *p.619, l.13: "An other" → "Another"*

Done.

- *p.619, l.17: "dependent of the" → "dependent on the"*

Done.

- *p.619, l.8: "in comparison of" → "in comparison to"*

Done.

- *p.618, l.12: What is a Gaussian space ? This beginning of an explanation is only given on p.622, l.11 and further on.*

An explanation has been added.

- *p.618, l.7: "physical/biological limitations": the ellipsis is awkward.*

The ellipsis has been removed.

- *There is a lack of general references in the beginning of the introduction, before the data assimilation concepts are presented.*

References have been added.

- *p.622, l.21: "is based of" → "is based on"*

Done.

- *p.622, l.25: \mathbb{N}_N is not a standard definition for 1, 2, ..., N. Please define it.*

Replaced.

- *p.623, l.18: "By assuming that H is linear" is surprising but later discussed. Maybe you should tell the reader that a discussion follows.*

A remark has been added.

- *p.623, l.22: "a Normal law" → "a normal law"*

Done.

- *p.623, l.23: "The return" → "The pull-back" (dedicated mathematical terminology)*

Done.

- *p.624, l.3, p.626. l.1, and p.630, l.20: "Remarks": please make sentences.*

We did not change the subtitle "Remarks", as we think that they are enough explicit.

[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)[Discussion Paper](#)

- p.624, l.13: "In the case when H is extracting measurements from the state vector, this is not an issue." Why so ?

Assuming that H is extracting measurement. For example we assume that we observe the diatoms. We can choose $\chi = \phi$, where ϕ is the anamorphosis function of diatoms. Then $\tilde{H} = \phi \circ H\phi^{-1}$. As $\phi H = H\phi$ (H just extracts the diatoms component), one obtains $\tilde{H} = H$, and so \tilde{H} is linear.

The sentence has been rewritten.

- p.624, l.3: *It is certain that marginal distributions will not capture the full non-Gaussianity, especially from cross-correlations. I have personally checked that on a very simple but highly non-linear (and therefore non-Gaussian, as far as pdf are concerned) Lorenz-63 model. But dealing with the problem with marginals is perfectly acceptable for high-dimensional problem, and can hardly been avoided. A way out could be independent component analysis but could be numerically too expensive.*

We agree with the referee. Thanks for the information about the independent component analysis.

- p.625, l.11-15: *It would be good to write some of the details in a short appendix. Especially you could tell the readers that do not know anamorphosis what it implies in terms of mathematics.*

Done.

- p.625. l28: "summarized Fig.1" → "summarized in Fig.1"

Done.

- p.627, l.7: "with a minimum thickness of the top layer of 3 m" → "with a minimum thickness of 3m at the top layer"

Done.

- *p.627, l.21: Drop the "So", because the reader cannot infer that the exact number of variables is 7 from what you have just explained*

Done.

- *p.627, l.22: "is illustrated Fig.2" → "is illustrated in Fig.2"*

Done.

- *p.628, l.1: Drop the "it means"*

Done.

- *p.629, l.9-14: What is the bias reduction for exactly ? I suppose it is to debias the observation error ($E[y_n - y_n^t] = 0$) ? But would it be the same in Gaussian space ? Without the correction term, you would debias the median, so that there would be no bias in Gaussian space (?). The objective is in any case, to match the BLUE requirements for the EnKF analysis.*

We agree with the referee. The bias reduction term has been added to remove the mean of the noise added to the true state in the construction of the observations. It means that in the physical space, the mean of the observations is equal to the true state. As stated it is no longer true in the Gaussian space, where there could be a low residual bias in the observations. Nevertheless in these twin experiments the EnKF with Gaussian anamorphosis works much better than the plain EnKF which is not affected by this potential low bias.

- *p.629, l.15-24: Couldn't you write in one sentence, that you generate the biological fields with this dynamical ensemble ?*

A sentence has been added.

- *p.629, l.25: Mitchell and Houtekammer were the first to introduce localisation in EnKF. You could cite them too.*

[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)[Discussion Paper](#)

We did not cite Mitchell and Houtekammer as the localization procedure that we use (Evensen, 2003) does not correspond to their procedure. In Mitchell and Houtekammer, the localization is realized by multiplying by a Schur product the covariance error matrix and the "localization" function (or taper function). In Evensen, the localization is realized by defining sublocal assimilation problem: selected and weighted (with a taper function) local observations are assimilated on each grid point.

- *p.631, l.13: "histogram of the transformed" is making reference to the anamorphosis procedure, which could be confusing for a reader unaware of its details. That is why again a short appendix on the maths of the method would be useful.*

Added.

- *p.631, l.14: "gaussians" → "Gaussians"*

Done.

- *p.632, l.3, Eq.(9): It seems that you do not make use of STD later in the text. If so, do not introduce it.*

STD appears in several figures and in the text of revised manuscript.

- *p.634, l.8: "as said previously" → "as stated previously"*

Done.

- *p.634, l.13: "and the data assimilation" → "and data assimilation"*

Done.

- *p.634, l.14: "can not" → "cannot"*

Done.

- p.634, l.15: *"the solution issued from"* → *"the solution obtained from"*
Done.

Interactive comment on Ocean Sci. Discuss., 6, 617, 2009.

OSD

6, C139–C150, 2009

Interactive
Comment

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper



Interactive
Comment

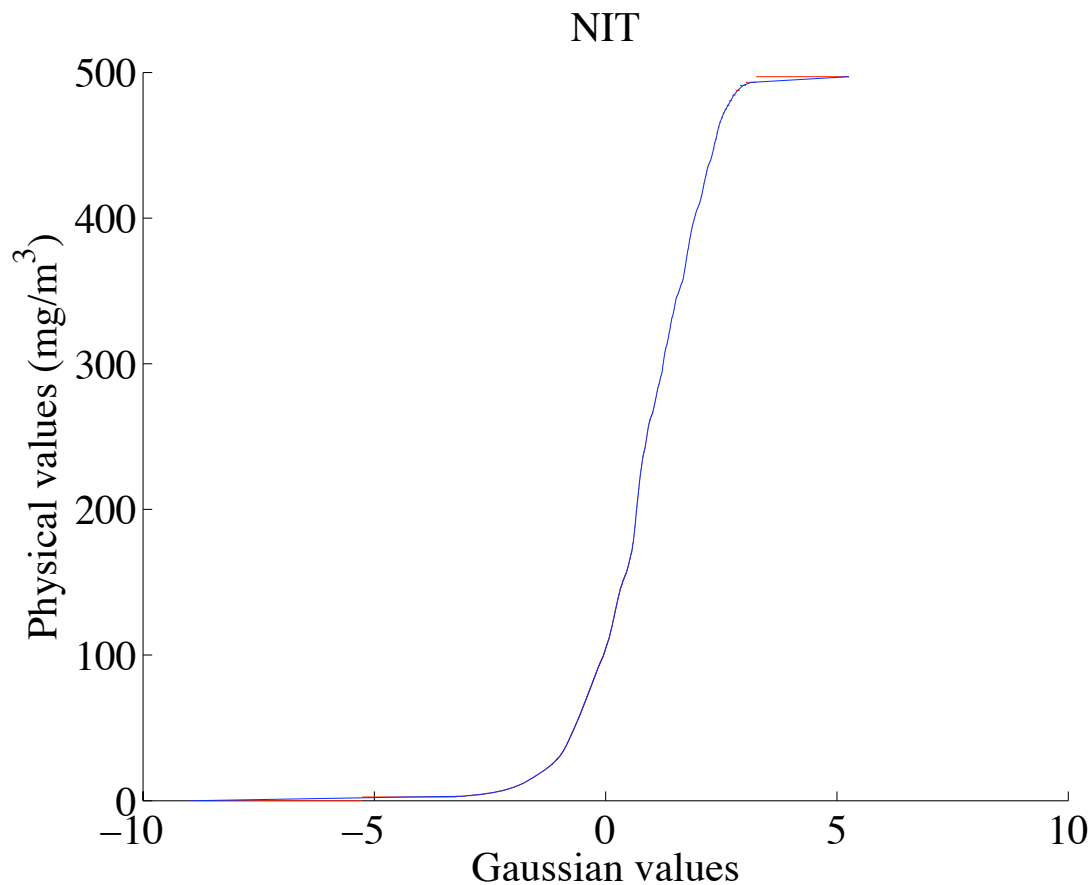


Fig. 1.

C150

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

