Ocean Sci. Discuss., 8, C899–C904, 2012
www.ocean-sci-discuss.net/8/C899/2012/

Ocean Science Discussions

Interactive comment on "Towards an improved description of ocean uncertainties: effect of local anamorphic transformations on spatial correlations" *by* J.-M. Brankart et al.

J.-M. Brankart et al.

Jean-Michel.Brankart@hmg.inpg.fr

Received and published: 16 January 2012

We thank Laurent Bertino for his careful reading of our paper, and for his constructive suggestions to improve the quality of the manuscript. They have been carefully taken into account as explained below.

First of all, we agree that the theroretical basis for the effect of anamorphosis described in our paper is well known (see answer to reviewer 1: the reference to Chilès and Delfiner, 1999, has been added in the subsection written to answer his comments). Our purpose is indeed to evaluate the importance of this effect in ocean applications

using a set of illustrative examples. And we also agree that the discussion of this effect is not the only originality of this paper, which is also written to show that an accurate approximation for the anamorphosis transformations can be obtained using a technically simple and efficient algorithm.

Second, we fully agree that an appropriate screening of statistical data using scatterplots is always needed to see if a simple model (e.g. a Gaussian model, with or without anamorphosis transformations) may be adequate to describe the data, or if a more general model is needed (for instance if the probability distribution is multimodal). The need for actually looking at the data is also pleasantly illustrated in a famous paper by Anscombe (1973), whose short abstract is simply: *"Graphs are essential to good statistical analysis. Ordinary scatterplots and triple scatterplots are discussed in relation to regression analysis."* However, we already presented a lot of scatterplots (a total of 38 scatterplots in 6 figures) in a previous paper (Béal et al., 2010) to discriminate the situations in which (i) the Gaussian assumption is sufficient, (ii) anamorphic transformations improve the description of the data, and (iii) anamorphic transformations do not help (even if they never introduce spurious correlations, and almost never remove meaningful correlations). On the contrary, in the present paper, the effect of the anamorphosis transformation on the scatterplot is always the same: transforming the regression curve (i.e. the line of maximum conditional probability) into a straight line, thus increasing the value of the linear correlation coefficient. (The only exception concerns the removal of spurious linear correlations in Fig. 12, see answer to minor comment below). This is why we have shown only two scatterplots (in Fig. 5) to illustrate this effect (also illustrated in the Fig. 9 of Béal et al., 2010), and used most figures to illustrate the effect of the transformation on the spatial correlation structure (not shown in Béal et al., 2010). However, it is true that leaving the underlying explanation behind the scene for the rest of the paper may be somewhat confusing. We have thus added 4 scatterplots in a new Fig. 15, providing explanations for Figs. 8 and 14 (as suggested in the first minor comment below).

Answer to minor comments:

1. Scatterplots have been added (in Fig. 15) to provide explanations for Figs. 8 and 14, together with an additional paragraph:

   This difference of behaviour between Figs. 8 and 14 can be better illustrated using scatterplots of PHY at the reference point (20° W 35° N) vs NO_3 at some distance from the reference (20° W 33° N), as shown if Fig. 15 for the correlation structure of Fig. 8 (top panels) and Fig. 14 (bottom panels), without anamorphosis (left panels) and with anamorphosis (right panels). In the first situation (corresponding to Fig. 8), the effect of wind perturbations is to introduce more or less mixing in the water column, so that the resulting perturbation of PHY and NO_3 tend to be anticorrelated (because of their opposit vertical gradient). And in the second situation (corresponding to Fig. 14), the model variability tends to positively correlate the PHY and NO_3 fluctuations. However, in both cases, we can observe in the scatterplots that the effect of the anamorphic transformations (giving the same normalized Gaussian distribution to all marginal distributions) is to produce a scatterplot with a more elliptical shape, which is a good indication that the joint distribution is also closer to a bi-Gaussian distribution. In these cases, it can be seen that the modification of the scatterplots results from the two properties of anamorphosis that were introduced in section 2.: (a) the linearization of a non-linear dependence between the two variables, and (b) the reduction of the effect of outliers (resulting here from occasional extreme behaviours). In both cases, these two properties explain the increase of linear correlation from $|\rho_{X_1X_2}| = 0.07$ to $|\rho_{Z_1Z_2}| = 0.43$ in the top panels, and from $|\rho_{X_1X_2}| = 0.24$ to $|\rho_{Z_1Z_2}| = 0.38$ in the bottom panels.

2. The large correlation in Fig. 10 between the Loop current and the Western coast of the Gulf of Mexico is due to the very simplistic assumption that is made to

generate the ensemble: it is assumed that the only source of error comes from 3 parameters in the ecosystem model and that this error is piecewise constant over the whole domain. And one of the region over which the error on the parameters is assumed constant covers the totality of the Gulf of Mexico (inside the black line in Fig. 10). With such a simple assumption, it is not really surprising that the error generated by the same parameter perturbations in the Loop Current and along the Western coast of the Gulf of Mexico can be well correlated. Anyway, in view of the size of the ensemble (200 members) and the value of the correlation (above 0.8), it makes no doubt that the observed correlations are meaningful. However, it is indeed very unlikely that this simple assumption correctly represent the *real* model uncertainties in the region, and we agree that this would have deserved a word of caution, which we have added in the paper:

   "Here, it must be remembered that, even if these large long-range correlations are certainly meaningful, they cannot be expected to describe real model errors, because they correspond to a very simple assumption, in which parameter errors are assumed constant over the whole Gulf of Mexico."

3. Yes, it is true that this is also an important aspect of the problem. Anamorphic transformations tend to increase the robustness of correlations by removing many spurious correlations (see also answer to reviewer 1). But we believe that, in this example, this can be explained without an additional figure. Because in the scatterplot, all ice fractions at any arbitrary point North of Iceland are equal to zero, except for a few outliers which produce the spurious correlation. This is like example 4 in Anscombe's quartet (Anscombe, 1973). Consequently, we have added the following text explaining the importance of this phenomenon and descring the shape of the scatterplot:

   "In the exterior of the ice pack, nearly all ice fractions are indeed equal to zero, so that the scatterplot with a point inside of the ice pack consist in a set of points aligned at $f = 0$, except for a few outliers, which produce the spurious correlation.

The shape of the scatterplot is thus like the example 4 in Anscombe's quartet (Anscombe, 1973, Fig. 4). And it is another well-known advantage of anamorphic transformations to produce more robust correlations, that are less influenced by the presence of outliers (see section 2.4)."

4. Strictly speaking, to compute the anamorphic transformation, the cdf must be invertible. And in this example, it is not since it makes a step at $f = 0$. Thus, to compute an approximate anamorphic transformation, we have had to replace the step by a steep slope. What we say in this sentence is that it would have been better to deal with the exact cdf (without approximation, and thus using a more sophisticated method than anamorphic transformations), but that having a transformation which transforms the variable into a approximate Gaussian variable is anyway often useful (even if approximate), because assimilation schemes dealing with Gaussian variables are much more efficient in practice. We have tried to clarify the sentence as follows:

   "It would of course be better to avoid any kind of approximation and to keep the (...), but this is impossible with anamorphic transformations, and it is anyway useful (...)."

5. The caption has been corrected.

6. Typo corrected.

Additional references

Anscombe F. J.: Graphs in Statistical Analysis, American Statistician, 27(1), 17–21, 1973.

Chilès J.-P., and Delfiner, P.: Geostatistics: Modeling Spatial Uncertainty, Wiley, 1999.

Interactive comment on Ocean Sci. Discuss., 8, 2147, 2011.

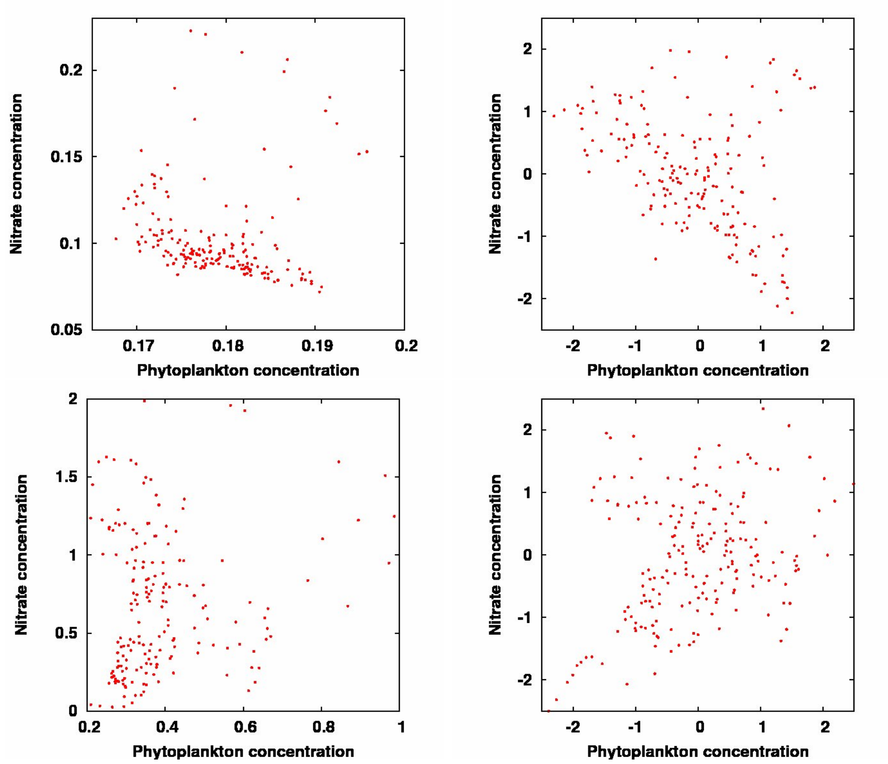

Fig. 1. New figure (Fig. 15) with scatterplots providing explanations for Figs. 8 and 14