Reducing uncertainties in earth system sciences by optimizing model parameters and measurements

Joscha Reimer
Institute for Computer Science, University of Kiel, Germany

Uncertainty in Models

- **Models**
  Models are a fundamental concept in earth system sciences, as in many other areas. A model may help to explain a system, to study the effects of different components and to make predictions.
- **Uncertainty Regarding the Model**
  Often several models are available to describe a process and it is not known which describes it most realistic. In order to determine the most realistic model, measurements of the modeled values must be carried out. The results are then compared with the various model outputs to choose the most realistic model among them.
- **Uncertainty Regarding the Model Parameters**
  Often there is also the problem that the model contains several parameters whose values are only vague known. In order to determine these parameters, measurements are also carried out. The parameters are then adjusted so that the the model reproduces the measurement results as precisely as possible.

Optimal Measurements

- **Information Content of Measurements**
  Measurements of the modeled values are therefore necessary to reduce the uncertainty in models. The circumstances under which these measurements are carried out vary from measurement to measurement. A spatial or temporal process can be measured at different locations or at different times. Often, different measuring instruments and measurement techniques can be used. All these factors influence the information content of the corresponding measurement results.
- **Optimizing Measurements**
  The circumstances under which measurements are carried out can be optimized to gain the maximum of information. Thereby a sufficient result is reached with fewer measurements and time, money and effort can be saved. Furthermore, it can be estimated whether additional measurements are needed and how these measurements increase the accuracy of the results.
- **Optimal Measurements for Model Distinction**
  To be able to choose the most realistic model from several models, measurements are useful under the conditions under which the models predict very different values. Optimal conditions for the measurements are therefore under which the predictions of the models differ most.
- **Optimal Measurements for Parameter Optimization**
  For a parameter optimization measurements are particularly important under the conditions under which the optimal parameters can be well determined. That is, under which conditions the parameters values strongly affect the model prediction. Optimal conditions for the measurements are therefore under which the parameters can be determined most accurately.

Quantification of Closeness to Reality

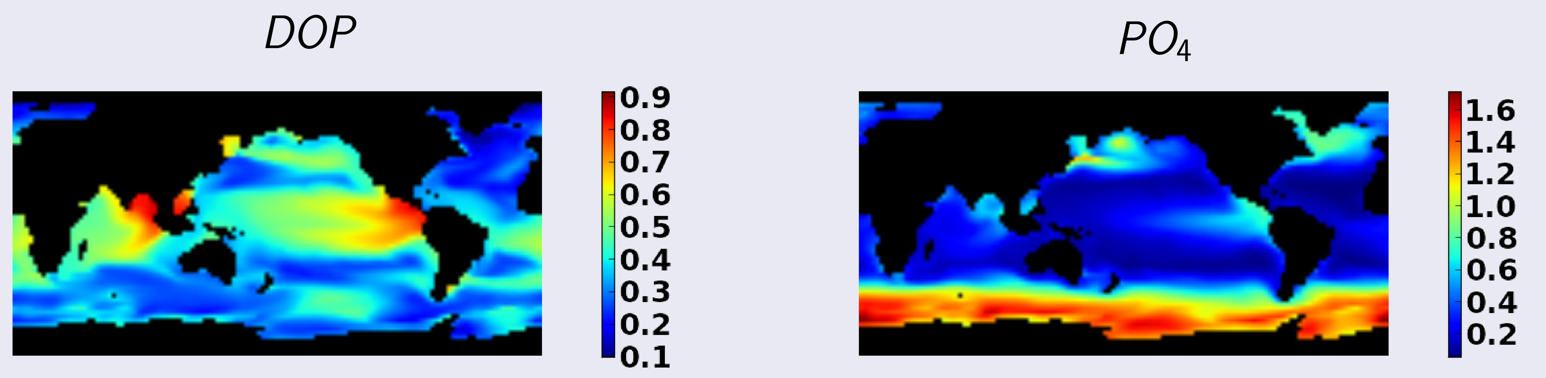
- **Misfit Function**
  To be able to chose the most realistic model or parameter values we have to quantify how close a model is to reality. This is done using a misfit function. The simplest misfit function is the sum over the (squared) differences between all measurement results and the corresponding model outputs. A more complex variant is obtained if the individual differences are weighted differently. The choice of the misfit function critically effects the quality of the results.
- **Choosing the Misfit Function**
  The choice of the misfit function should depend on the statistical properties of the measurement data. For example, the data which is based on a larger variance should be weighted less. Furthermore dependencies between the data should be considered. If for example, a spatially or temporally changing process is measured, there are automatic dependencies which are greater, the closer the measurements are to each other.

References

- T.P. Boyer H.E. Garcia R.A. Locarnini O.K. Baranova Johnson, D.R. and M.M. Zweng.
  World ocean database 2009 documentation.
  Technical report, NODC Internal Report 20, NOAA Printing Office, Silver Spring, MD, 175 pp., 2009.
- I. Kriest, S. Khatiwala, and A. Oschlies.
  Towards an assessment of simple global marine biogeochemical models of different complexity.
  Progress in Oceanography, 86(3-4):337–360, 2010.

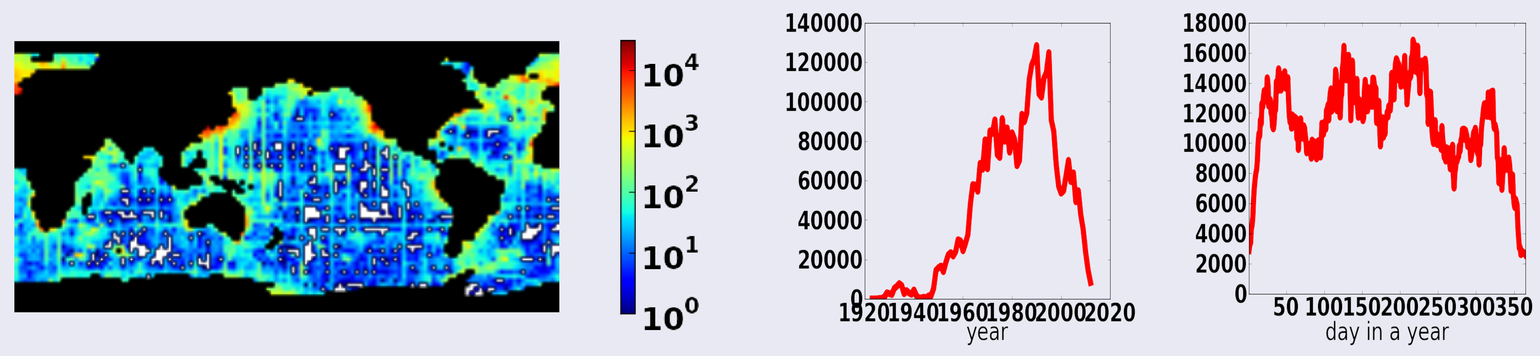
Application Example

- **The N-DOP Model**
  As application example we focus on a three-dimensional coupled marine biogeochemical-circulation model simulating the transport and biogeochemical cycles among the single nutrient (N), phosphate, and dissolved organic phosphorus (DOP). The model was presented in [2]. For this model, a parameter optimization was performed.



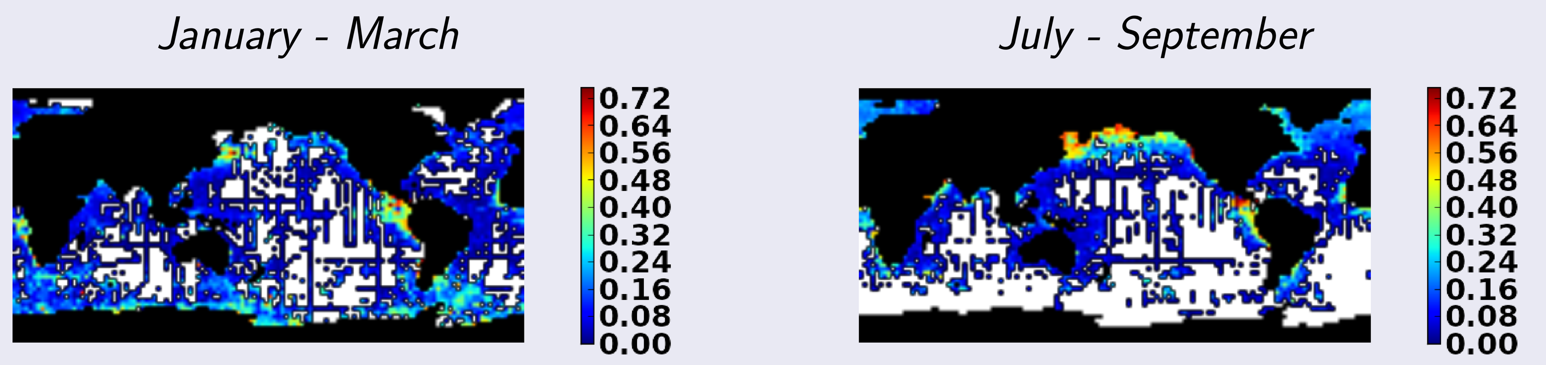
Annual averaged concentration of dissolved organic phosphorus (DOP) and phosphate (PO_4) at the surface resulting from the N-DOP model.

- **Phosphate Data**
  Through the World Ocean Database [1] over 4 million measurements for phosphate are available. For these data, a quality check was performed and implausible and incorrect data were filtered. The data date back to nearly 100 years, and were collected at different depths and at different seasons all over the world.



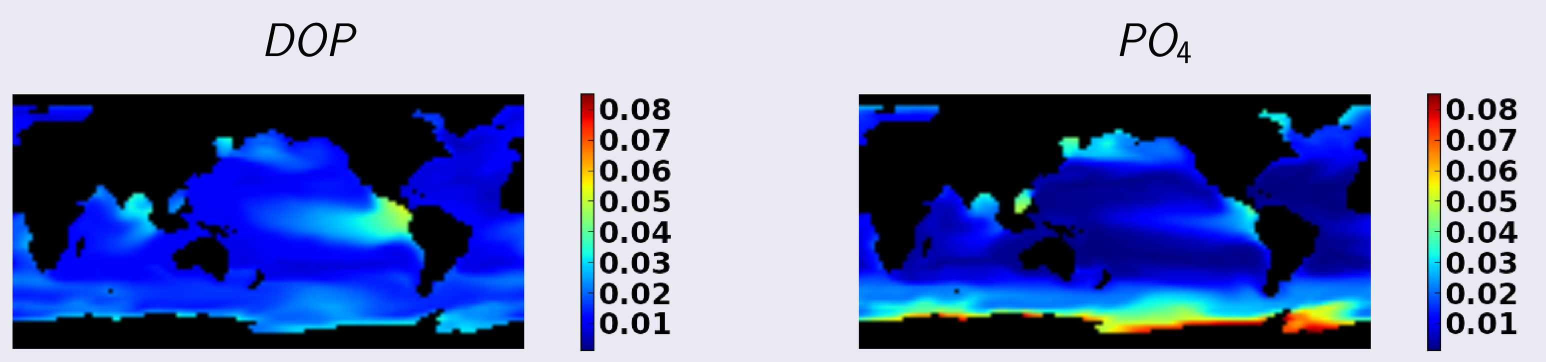
Number of observations of phosphate from the World Ocean Database.

- **Statistical Properties of Phosphate Data**
  The phosphate data must be analyzed to incorporate their distribution and dependencies in the parameter optimization of the N-DOP model. The standard deviation of the data, for example, varies greatly in space but in time only slightly.



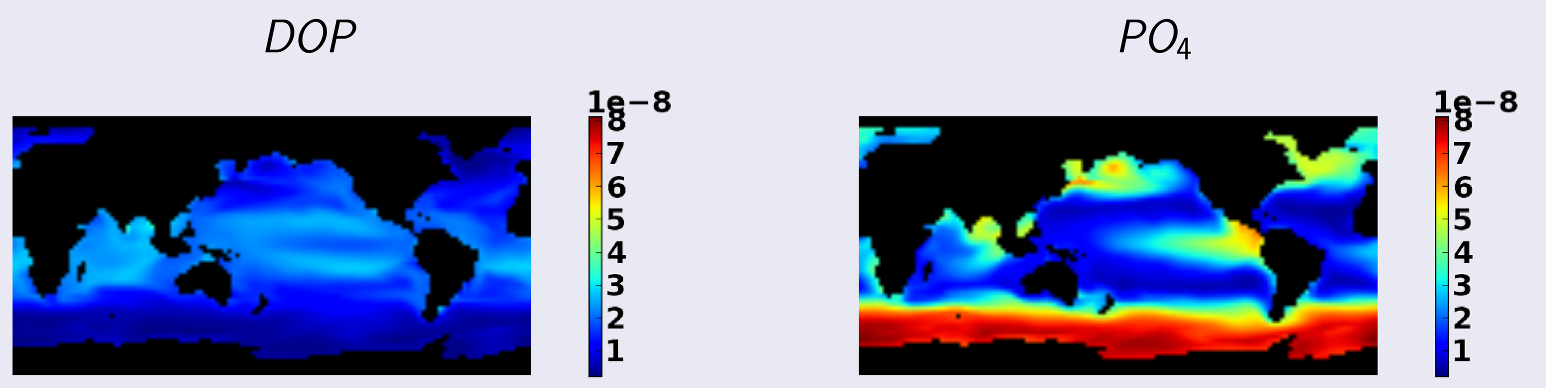
Standard deviation of observations of phosphate at the surface from the World Ocean Database.

- **Uncertainty in the N-DOP Model**
  Using the statistical properties of the measurement data it can be estimated how accurately the optimal parameters can be determined. Thus also the uncertainty which the model includes due to variability in the measurements can be estimated.



Annual averaged uncertainty of the N-DOP model at the surface resulting from the uncertainty in the measurement data.

- **Additional Measurement for the N-DOP Model**
  Next it can be estimated how additional measurements and the different measurement conditions reduce the uncertainty. The uncertainty in the N-DOP model due to measurements can only slightly by reduced by additional measurements. Measurements of phosphate are much more important than measurements of dissolved organic phosphorus and measurements of phosphate near the Antarctic are especially important.



Reduction of the (averaged) uncertainty in the N-DOP model by an additional measurement at this loaction.