Geophysical Research Abstracts
Vol. 14, EGU2012-2355, 2012
EGU General Assembly 2012

Automated Probabilistic Quality Assurance for Data Publication

A. Düsterhus and A. Hense
Meteorological Institute, University of Bonn, Germany (andue@uni-bonn.de)

For a researcher the publication of his work is a very important part for building up his reputation. At the moment this is mainly done by producing written publications, but the publication of raw data will become more important in the coming years. An important part of the publication process is the quality assurance of the data not only by technical means, but also by its content. For this quality assurance methods for general data are a very important part.

An important aim here is to test unknown datasets on inconsistencies, which is a challenging task. Useful tools in inconsistency detections are change point models. Those are fitted to the data and if the model indicate, that a model with a step is the most probable explanation for the data, the user looks for explanations for the steps. We use and modified here a bayesian approach developed by Dose and Menzel, which was not used for quality assurance until now.

But change point detection work mostly only on one dimensional time series. For multidimensional time series, additional tools are necessary. One tool for this is the estimation of probability densities for parts of the data and comparing them to other parts. This was done by us by using histograms and comparing them with help of the Kullback-Leibler Distance and the Earth Mover's Distance. With this it is possible to identify parts with different characteristics like variance shift, level shifts or rounding within the data.

This contribution introduce both, the identification method of inconsistencies by estimating probability densities and the bayesian change point detection method. It also give insights to some sensitivity tests and applications not only for the separate methods but also for their combination.