**Climate
of the Past
Discussions**

Interactive
Comment

# *Interactive comment on* "On the verification of climate reconstructions" *by* G. Bürger

**Anonymous Referee #2**

Received and published: 22 March 2007

1. Objective and results:

The main goal of the author is to compare the performance of past NH temperature reconstruction techniques. Assessing the skill of statistical techniques is an important topic for the reconstruction community. This objective of is worthwhile to pursue and it should be interest of the CP readership. The two tools used by the author are a set of two statistics, Reduction of Error (RE) and Coefficiency of Error (CE), and a multi-crossvalidation method. The author has mainly two results. First, he showed that RE and CE are linearly linked. Second, RE is classically overestimated because of the presence of trends, but still methods based on proxy data still outperformed nonsense predictors.

2. Novelty/originality:

Although a lot of work on studying skills, scores, cross-validation techniques has al-

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

EGU

ready been done in statistics and weather forecasts, such issues have not been very much explored in the realm of reconstruction studies. Consequently, this work is a welcome addition to the reconstruction community. Neither new statistical techniques nor novel reconstructions were not derived in this paper. Consequently, I believe that this manuscript could be viewed as an overview (or review) of past skills techniques that have been used in other fields.

3. Organization, clarity and technical correctness:

This paper is rather difficult to read and this lack of clarity is the main drawback of this work. May be, by trying to make things too general, the author takes the risk to lose his/her reader very quickly. This is particularly true for the public of PC that may be not well versed in statistics.

For example, in weather forecasting literature, there are many possible definitions of skill (Odds Ratio Skill Score, Pierce Skill Score, etc). But in this paper, the term "skill" is never clearly defined. There are many instances of unclear, obscure and complex sentences in this manuscript. Instead of giving an exhaustive list, I am going to illustratre my point by just taking two sentences from the manuscript (see p251): *"Calibrating is done,..., by optimizing the model skill for a selected sample (the calibration set) and is almost affected by the presence of sampling noise. This renders the model imperfect, and its true skill is bound to shrink. But it is this skill that is relevant when independent data are to be predicted".* The expression "by optimizing the model skill" is rather puzzling because (1) there is no "the skill" but "a" skill among many possible skills, and (2) the parameters of a given statistical model are estimated by optimizing a given criterion (Mean Squared Error, likelihood function, etc) that can be very different for a chosen skill score that has been calculated after the parameter estimation step. Then, the author says that noise "renders the model imperfect". A statistical model like a regression model is always written as: observations = f(proxies) + noise where the noise is classically Gaussian and iid. Hence, the noise is included in the model, so it does not make the model imperfect. What does make the model imperfect? It can incorrectly

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

EGU

specified (e.g. linear instead of log). The noise is red but it is assumed to be white, etc. I think (but I am not sure) that the author means that an estimator of a skill is a random variable and that the bias and the variance of this estimator can change with the variance of the noise (or more if it is a red noise). With respect to the expression "But it is this skill that is relevant when independent data are to be predicted", it is also very misleading because the definition of independent data is that they have no memory, and consequently they can not predicted!! The same lack of clarity can be made for the term "model". For example, "empirical models of this kind" at the bottom of page 252 is a very vague expression. I have the feeling that the author does not want to make the difference between the statistical model (e.g. linear models) and the estimation procedure to estimate the parameters of the chosen models. Bootstrapping and cross-validation are not models!! They are estimation procedure. My suggestion will be to: (0) recall the definitions of different skills, (1) pick up one early on, (2) define clearly the statistical models that will be used in the paper, (3) define clearly the estimation procedures that will be used in the paper. Overall, I think that the author is handling a lot of complex statistical concepts. The introduction and the abstract should be completely rewritten. The concepts should be illustrated by simple examples and clearly defined. I advise the author to read (or reread) the paper by Thornes and Stephenson (2001, Meteo. Appl.) that presents a simple and elegant overview of skill techniques. If the author could follow the same pedagogical approach, that will greatly improved the scope of his work. From a more technical point of view, the author derived the relationship between CE and RE by assuming that the validation is done using the entire population. I would like to know how this relationship changes when this assumption does not hold. The author computed his CE and RE for existing NHT reconstruction techniques. But, working with real observations does not allow the statistician to know if the assumed statistical model is the real one. It would have been possible to compare the different skill procedures by fixing know statistical models with different type of noise and by simulating data from them. In this context, it would have been clearer to assess the quality of reconstruction techniques. Indeed, I believe that the focus of this paper

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

EGU

is not NHT per se but rather assessing and discussing the quality of reconstructions techniques.

## 4. OVERALL RECOMMENDATION:

In summary, the statistical study of assessing the quality reconstruction procedures is essential to the climate community. The approach developped by the author is a valuable attempt and represents a lot of work. But the paper is rather unfocused, complicated and could be simplified if the basics concepts were better explained. I believe that a revised version that takes my comments into account should be suitable for publication.

Interactive comment on Clim. Past Discuss., 3, 249, 2007.

EGU