Biogeosciences Discuss., 9, C562–C569, 2012
www.biogeosciences-discuss.net/9/C562/2012/

Biogeosciences Discussions

Interactive comment on "Technical Notes: Calibration and validation of geophysical observation models" *by* M. S. Salama et al.

M. S. Salama et al.

salama@itc.nl;souseto@gmail.com

Received and published: 4 April 2012

<Responses of the authors>

First we would like to thank the reviewers for providing their suggestions and comments, which have assisted us to improve the manuscript.

The comments of the reviewers are in general positive and indicate that manuscript is of good quality, but needs some fine tuning.

Both reviewers have suggested that a more explanation is needed on the added-value of the proposed method with respect to existing procedures. In addition to some minor comments, one reviewer has raised an important point about the randomness of the optimal Calibration/Validation pairs.

We have answered all the concerns of the reviewers and revised the manuscript. The main changes in the revised manuscript are as follows, 1- A more detailed review on the existing works has been added. 2- The added value of the proposed method has been justified in view of existing works. 3- The underlying mechanism for optimal splitting of the data into Cal and Val sets has been identified.

With this response we also provide a supplement document to indicate how we have incorporated the suggestions of the reviewers and our responses to their comments in the manuscript: blue font for addition (new text) and the red font indicates deletion.

With this response we think that we have adequately answered the concerns of the reviewers and revised the manuscript to the level of BG publications.

Salama et al.

<><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><>

Anonymous Referee #1 The paper is describing a method for calibrating and validating observational models that relate remotely sensed energy fluxes to geophysical variables of land and water surfaces. This is a very interesting topic especially for all the Cal/Val activities of new sensors (e.g. SMAP). The method does not seem to be new, but, in case, a new application of already developed statistical methods. I am not particularly skilled on these statistical methods; however, I suppose that major credit should be given to persons who first developed such a type of procedures. References should be therefore improved and up-to-dated. In my opinion the application of the statistical method should be explained with more details in order to make clearer what is new and is not and to add details on the followed procedure. Since the paper has been sumbittes as a Technical note, I suggest accepting it for publication after some minor revision.

<Response> We have elaborated on the used statistical method in which the novelty of the proposed method was justified in view of exist-

ing procedures with proper credits and with up-to-date references. Further details are provided in our response to comment 1 of review 2. <><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><>

Referee #1 Minor: Caption of fig. 2: It is better ‘Determination coefficient (R2)’ instead of ‘coefficients of determination’. In any case the captions should cite the same variables of the diagrams.

<Response> It has been changed to “determination coefficient” throughout the whole manuscript.

<><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><>

Anonymous Referee #2 General comments: In this paper, the authors describe a calibration/validation method, which is inspired on the traditional cross-validation framework. The developed method is illustrated based on two datasets from the field of remote sensing. The development of techniques supporting calibration/validation and uncertainty assessment is important and should receive sufficient attention in the literature. However, I have a few questions about the proposed method in this paper:

(1) The proposed framework is claimed to be novel (and I do not necessarily disagree on that), however, it is strongly related to what is called ’bootstrapping’ in the classical statistical literature. Since its introduction by Efron in the early eighties, bootstrapping procedures have been studied extensively in the statistical literature as well as in applied domains. As such, I believe that a new method should be evaluated in comparison with this established framework. Nevertheless, the authors do not mention the existence of this well-known procedure. A better situation of the method with regard to existing techniques is essential.

<Response> Indeed, the proposed method uses the bootstrapping method of Efron without replacements (please note that Erforn’s method is with replacement). However, we combine it with the Jackknife technique (which leaves out one observation) and

adapt the sample size at each bootstrapping.

Bootstrapping and Jackknife methods are usually used to provide the standard error of the derived “plug in” estimate. This combination of (1) bootstrapping without replacement with (2) Jackknife sampling and (3) changing the sample size at each test iteration is novel and provides not only the accuracy of regressed-estimate but the full PDF of regressed-estimates (plug in) and the PDF of their errors.

There are five innovative aspects to our approach: 1- We combine both approaches: the bootstrap which resamples from the observations and the jackknife which deletes -one or - n observations;

2- Each bootstrapped Jackknife sample has also a different sample size. Thus, the effect of sample size on the accuracy of regressed-estimates is embedded within our method.

3-The application of this method to calibrate and validate observational models is new.

4- In addition to providing accuracy measures for the regressed-estimates, we also provide a confidence interval for the errors themselves.

5- With our method the underlying probability distributions of estimate and their errors can be quantified.

We have added this explanation and described our method in the framework of Efron’s works. Proper citations and explanations are now provided in the text. The added value of our procedure is now better highlighted. <><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><>

Referee #2 (2) The authors introduce a sampling procedure and use this procedure to approximate the distribution of several statistics and model parameters. In classical linear regression, the sample size is very important with respect to the distribution of a statistic. For instance, in case of a regression model, the distribution of the slope (and its confidence interval) is heavily influenced by the sample size. Typically, knowing the distribution of a statistic for a given sample size is important. The approximated distribution that arises when parameters are repeatedly estimated based on samples of varying size might not be very informative. Stated differently, I do not immediately see what kind of relevant 'statistical' question can be answered based on such a distribution.

<Response> Indeed, the accuracy of estimates is dependent on the sample size and the size of the used set lies at the heart of successful Cal/Val activities of earth observation (EO) models and products. Here, we search for the optimal division (thus, sample size) of the Cal/Val sets such that the Cal set produces EO-model coefficients that enable in generating EO products (estimated from the Val set) with an accuracy satisfying the mission requirements.

Hence, the statistical questions addressed within our manuscript is: what is the minimum sample size needed for calibrating observation models so that it produces EO products within the designed mission accuracy and within the accuracy of the calibration itself? This question can only be answered if we change the sample size and study it's effect on the accuracy of calibration and validation as shown in Fig.1 of the manuscript. Moreover only through this type of adapted sampling the shape of the underlying PDF can be revealed, therefore no assumptions on the shape and/or parameters of underlying PDF are needed. <><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><>

Referee #2 (3) The authors illustrate their method using two datasets. In both settings, they use a simple linear regression model. For these models, the distributions of the model parameters can be derived theoretically. Basically, this derivation only holds for normally distributed error terms; however, since both datasets contain at least 75 observations, validity can be assumed based on the central limit theorem. Moreover, the obtained t-distribution further supports this claim (as it can be the result of mixing a large number of Gaussians). Bearing this in mind, why do the authors make use of a more complicated and computationally demanding sampling scheme instead?

<Response> The normality assumption is not valid for linear models and the Student-t distribution is broader and therefore better suited. I quote from Singh (1988) [Economic Letter 27-1-19988 pp27-53, doi: http://dx.doi.org/10.1016/0165-1765(88)90218-2] "In many theoretical research work in linear regression analysis as well as in many applications of linear regression models to practical situations, the error terms are assumed to be normally and independently distributed, each with zero mean and common variance. However, as it is well known, error terms can have non-normal distributions [e.g. see Zellner (1976) for references to works employing non-normal errors]. As such the normality assumption on the distribution of the error terms could be misleading and its violation could have adverse effects on the inference drawn in a number of situations; see, for example, Gnanadesikan (1977). Recently, some authors [e.g. Zellner (1976), King (1979, 1980), Ullah and Zinde-Walsh (1984) and Singh (1987) have employed a broader assumption, namely, that the error terms have a joint multivariate Student-t distribution. With this assumption, the marginal distribution of each term is univariate Student-t, a distribution that includes the Cauchy and normal distributions as special cases. "

However the assumed t-Student distributions in these studies were assumed to have known degrees of freedom, I quote from the same reference:

"However, to the best of this author's knowledge, almost all the works in linear regression models employing error terms having multivariate Student-1 distribution have assumed that the degrees of freedom of the distribution is known."

Although the regression parameters and associated errors of a linear modal can be assumed to be t-distributed, the degree of freedom is generally unknown. Moreover for a non-linear model it is more complicated to un-puzzle the underlying distribution, i.e. there is no straightforward theoretical approximation of the expected PDF.

In this paper we have restricted the analysis to a linear model. Our suggested approach reproduced what the theory should predict (normal like distribution). In this regard, having the result of our sampling scheme reproducing the T-student distribution is another validation of the correctness of our model.

Thus we make use of the more complicated and computationally demanding sampling scheme because only through this approach we can quantify in an objective manner the optimal sample size and the optimal division of a data set into Cal and Val pairs. If we would follow the theory, we would have no means to justify our assumption on the underlying probability distribution and its parameters (e.g. degree of freedom for the T-Student distribution). <><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><>

Referee #2 Specific comments: - p317.17: I think this formula should be: ‘nr = n-2kmin+1’

<Response> Although nr was implemented in the code as nr=n-2Kmin+1, the +1 was dropped by accident in the text. <><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><>

Referee #2 - p318.1: The nuci becomes larger for an increasing number of cal pairs ki. Hence, matchups with a large cal set eventually contribute more to the derived distributions. What is the effect of the variable nuci on the distributions, and how would these distributions look like if a fixed nuc would be used?

<Response> Fixing the number of sampling points ki will result in PDF with lower Kurtosis, the PDF will be less peaked. That means adapting the sample size will increase the accuracy of the derived parameters (slope and intercept in this case), as the dispersion will also be reduced. The importance of adapting the size of the sample is related to the common practice in calibration and validation of earth observation products, Also consult our response to comment 2. <><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><>

Referee #2 - p319.17: The best results may possibly show randomness regarding the

number of cal/val pairs; however, I expect that the worst results will not show such randomness, and will mainly occur with unbalanced cal/val sets. Therefore, I do not agree that an optimal setup for subdividing matchups into cal/val sets cannot be defined, and that the only objective approach is to evaluate all possible combinations.

<Response> We would like to thank the reviewer for rising this point. We have further investigated the randomness and found that the best Cal/Val pair can be achieved when the first and second moments of each set are equal to that of the original data set. As such if the user is only interested to find the optimal cal/ val pair without getting the whole PDF s/he can stop the computation when the Cal/Val sets satisfy the condition mu_cal=mu_val=mu_data and sig_cal=sig_val=sig_data.

Where: _cal, _val and _data are for the calibration, validation and original data sets, respectively mu and sig refer to the mean and standard deviation of a set.

We change the conclusion in p319 line 17 in which the underlying mechanism for obtaining the optimal Cal/Val pair is now better explained. <><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><>

Referee #2 Technical corrections: - p313.22: applications ‘use’ or ‘make use of’

<Response> Corrected to “make use of”.

<><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><>

Please also note the supplement to this comment:
http://www.biogeosciences-discuss.net/9/C562/2012/bgd-9-C562-2012-supplement.pdf

Interactive comment on Biogeosciences Discuss., 9, 311, 2012.