Dendrochronologia 42 (2017) 56–62

Contents lists available at ScienceDirect

Dendrochronologia

journal homepage: www.elsevier.com/locate/dendro

Testing the stability of transfer functions

Allan Buras[a,*], Christian Zang[b], Annette Menzel[a,c]

[a] Technische Universität München, Ökoklimatologie, Freising, Germany
[b] Technische Universität München, Land Surface-Atmosphere Interactions, Freising, Germany
[c] Technische Universität München, Institute for Advanced Study, Garching, Germany

ARTICLE INFO

Article history:
Received 21 June 2016
Received in revised form
25 November 2016
Accepted 17 January 2017
Available online 28 January 2017

Keywords:
Dendroclimatology
Climate reconstruction
Verification
Coefficient of efficiency
Reduction of error
Bootstrapped Transfer Function Stability test

ABSTRACT

In dendroclimatology, testing the stability of transfer functions using cross-calibration verification (CCV) statistics is a common procedure. However, the frequently used statistics reduction of error (RE) and coefficient of efficiency (CE) merely assess the skill of reconstruction for the validation period, which does not necessarily reflect possibly instable regression parameters. Furthermore, the frequently used rigorous threshold of zero which sharply distinguishes between valid and invalid transfer functions is prone to an underestimation of instability. To overcome these drawbacks, we here introduce a new approach – the Bootstrapped Transfer Function Stability test (BTFS). BTFS relies on bootstrapped estimates of the change of model parameters (intercept, slope, and r^2) between calibration and verification period as well as the bootstrapped significance of corresponding models. A comparison of BTFS, CCV and a bootstrapped CCV approach (BCCV) applied to 42,000 pseudo-proxy datasets with known properties revealed that BTFS responded more sensitively to instability compared to CCV and BCCV. BTFS performance was significantly affected by sample size (length of calibration period) and noise (explained variance between predictor and predictand). Nevertheless, BTFS performed superior with respect to the detection of instable transfer functions in comparison to CCV.

1. Introduction

Transfer functions process a time-varying signal – a proxy – to yield another signal of estimates (Sachs, 1977). In dendroclimatology, the proxy is a tree-ring parameter, such as density or width, and the estimate a parameter of past climate, such as temperature or precipitation. Estimating the reliability of these transfer functions is a common and mandatory aspect of dendroclimatological reconstructions (e.g. Fritts, 1976; Cook and Kairiukstis, 1990). For this purpose, the so-called cross-calibration-verification (CCV) is frequently considered (Fritts, 1976; Cook et al., 1994). In CCV, a transfer function – e.g. the frequently used ordinary least-squares regression (OLS) – is computed for a calibration period (for instance half the period of available calibration data) and then applied to predict the target quantity (e.g. temperature) for the respective remaining period (the verification period). Subsequently, this procedure is repeated with swapped calibration and verification periods. For both calibration-verification sets the reduction of error (RE) and the coefficient of efficiency (CE) are calculated:

$$RE = 1 - \frac{\sum_i^n \left(x_i - \hat{x}_i\right)^2}{\sum_i^n (x_i - \bar{x}_c)^2} \quad (1)$$

$$CE = 1 - \frac{\sum_i^n \left(x_i - \hat{x}_i\right)^2}{\sum_i^n (x_i - \bar{x}_v)^2} \quad (2)$$

with:

- x_i being the measured target variable and $\hat{x}_i$ being the predicted target variable for $i = 1,\ldots,n$,
- $\bar{x}_c$ being the mean of the target variable for the calibration period,
- $\bar{x}_v$ being the mean of the target variable for the verification period,
- and positive CE and RE values indicating predictive skills greater than those of the respective null models (mean value of target quantity, i.e. the climatology of the calibration period for RE, climatology of the verification period for CE). In these cases, transfer functions are considered stable (Cook et al., 1994).

Abbreviations: BTFS, Bootstrapped Transfer Function Stability test; BCCV, bootstrapped cross calibration verification; CCV, cross calibration verification; CE, coefficient of efficiency; ECDF, empirical cumulative distribution function; OLS, ordinary least-squares regression; RE, reduction of error.

* Corresponding author at: Hans-Carl-von-Carlowitz Platz 2, 85354, Freising, Germany.

E-mail address: allan@buras.eu (A. Buras).

http://dx.doi.org/10.1016/j.dendro.2017.01.005

Accordingly, CE and RE focus on the residuals and predictive skill of models in the verification period. While this is an important aspect of transfer functions, the stability of regression parameters such as intercept, slope, and explained variance is only indirectly accounted for. That is, if one or several model parameters vary largely, the residuals of the prediction will be larger than those of the null model, this resulting in negative RE and CE values. However, for low variations of regression parameters, this may not be the case. Moreover, both metrics introduce a sharp threshold of 0 for classifying transfer functions as invalid (CE and RE ≤ 0) or valid (CE and RE > 0). However, this threshold neglects that both positive and negative CE and RE values close to zero indicate residuals in the same order as the null model, i.e. low predictive power. Thus, as long as the residuals of the reconstruction are lower than those of the respective null model, transfer functions will pass CCV, irrespective of the stability of regression parameters. Consequently, diverging climate-growth relationships – which are important to identify when reconstructing climate – may be overlooked if stability assessments are only based on CCV. Furthermore, CCV is known to be sensitive against outliers (Cook et al., 1994), thus may classify stable transfer functions invalid due to outliers in the calibration or verification period. Finally, since there is no parametric significance test for CE and RE available (Cook et al., 1994), stability assessments based on CE or RE traditionally cannot estimate the probability of obtaining false positives (i.e. type I error). One possibility to handle this drawback is the application of bootstrapping techniques to generate a distribution of RE and CE estimates (e.g. Wahl and Smerdon, 2012). The focus on predictive skills in contrast to stability of model parameters, however, remains true also for bootstrapped variants of CCV.

To overcome this drawback, we propose a new approach – the Bootstrapped Transfer Function Stability test (BTFS) – which aims at quantifying the stability and significance of transfer functions over time. In the following, BTFS is tested for a large variety of pseudo-proxies with known stability/instability and compared to CCV and a bootstrapped CCV.

2. Material and methods

2.1. Bootstrapped Transfer Function Stability test

Since the general intention of our approach is to test the stability of transfer functions over time, ordinary least squares linear regressions (OLS) are computed for two periods each covering 50% of the period with available calibration data. Other regression methods such as inverse OLS or reduced major axis models (RMA) can be applied to BTFS, too, but for reasons of simplicity we here focus on the frequently used OLS approach. For each of the two regressions, model intercept (a), model slope (b), and explained variance (r^2) are extracted and the respective parameter ratios calculated. Accordingly, parameter ratios of one indicate perfect stability of the respective model parameter. Bootstrapping is used to get robust estimates of model parameter ratios for a predefined number of iterations i (here: $i = 1000$). That is, the two periods are each randomly subsampled i times allowing for replacements and the corresponding models are computed to derive i ratio estimates of a, b, and r^2. Empirical cumulative distribution functions (ECDFs) are derived from the i estimates of each parameter and used to compute the 95% confidence interval of bootstrapped estimates. If this confidence interval does not contain the ratio 1, the respective parameter is considered instable. In other words, based on the ECDFs, BTFS tests the null-hypothesis that the observed ECDF could have been obtained if the true parameter ratio was one. Accordingly, if the associated probability is lower than 0.05, the true parameter ratio is unlikely to be one wherefore BTFS rejects a transfer function as instable.

In addition to these three parameters, the proposed approach also computes regression p-values for the bootstrapped periods. Consequently, for each period i estimates of the true p-value are obtained. For each iteration the maximum – thus least significant – p-value is extracted and the proportion of significant regressions ($p < 0.05$) is reported. If this proportion is below 0.05, a transfer function is considered invalid as regressions for at least one of the two periods frequently were non-significant. To account for different aspects of instability, the proposed approach comprises all four bootstrapped statistics (i.e. slope, intercept, r^2, and significance) in one assessment. If one of these statistics is significant, a transfer function is considered invalid. Being based on these four parameters, this approach covers several possibilities of transfer function instability. That is, if model parameters (slope, intercept, r^2) or model significance vary significantly over time this will be identified by the proposed approach. As testing transfer function stability and being based on bootstrapping we call this approach the Bootstrapped Transfer Function Stability test (BTFS).

2.2. Data

To validate BTFS and compare it to the commonly applied CCV and a bootstrapped CCV approach, we ran 42,000 pseudo-proxy experiments. To generate pseudo-proxies, we used a tree-ring dataset downloaded from the International Tree-Ring Data-Base (ITRDB; https://www.ncdc.noaa.gov/paleo/study/6344, Wilson et al., 2007). This data-set contains 15 tree-ring chronologies distributed around the Northern hemisphere, thus in our opinion represents a broad variety of tree-ring characteristics world-wide. We used these data to generate 42,000 pseudo-proxy data-sets. That is, for each set a randomly subsampled sequence of predefined length (specifications are given below) of a randomly selected tree-ring chronology was defined as predictor (in dendroclimatological transfer-functions the tree-ring parameter), whereas the predictand (climate parameter) was defined as predictor multiplied by 1.5 (the slope) and added by 1 (the intercept). Introducing slope and intercept to the pseudo-proxies was done to create more realistic conditions (i.e. slope and intercept not being zero) but this will not affect the performance of BTFS or CCV.

Subsequently, white noise (i.e. randomly generated values having no auto-correlation, zero mean, and not being correlated to the noiseless variable itself, see e.g. Kutzbach et al., 2011) was added to the predictand. Thereby a variable was obtained that – depending on the standard deviation of the added noise (specifications below) – was more or less correlated with the predictor. Based on this definition, the relationship between predictor and predictand is stable over time. To generate scenarios representative of instable transfer functions, the predictor was modified either by I) including a non-linearly increasing trend along the time-series, i.e. temporally increasing deviation among predictor and predictand or II) by non-linearly increasing the noise intensity along the time-series, i.e. a temporal weakening of the correlation among predictor and predictand. For each scenario, instability was represented by six different levels ranging from no instability to strong instability. Scenarios related to I) and II) in the following also will be termed 'trend-scenarios' and 'noise-scenarios'.

To represent different data qualities within a realistic range of conditions, pseudo-proxy sets varied in temporal span (40, 60, 80, 100, 120; x-axis on Figs. 2 and 3 as well as Supplementary Figs.), standard deviation of added noise (50, 60, 70, 80, 90, 100, and 110 percent of the predictand's standard deviation, y-axis on Figs. 2 and 3 as well as Supplementary Figs.), and differing strengths of temporal instability (ranging from no instability to strong instability resolved in 6 levels; different panels on Figs. 2 and 3 as well as Sup-