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,. . .,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
1125-7865/

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-

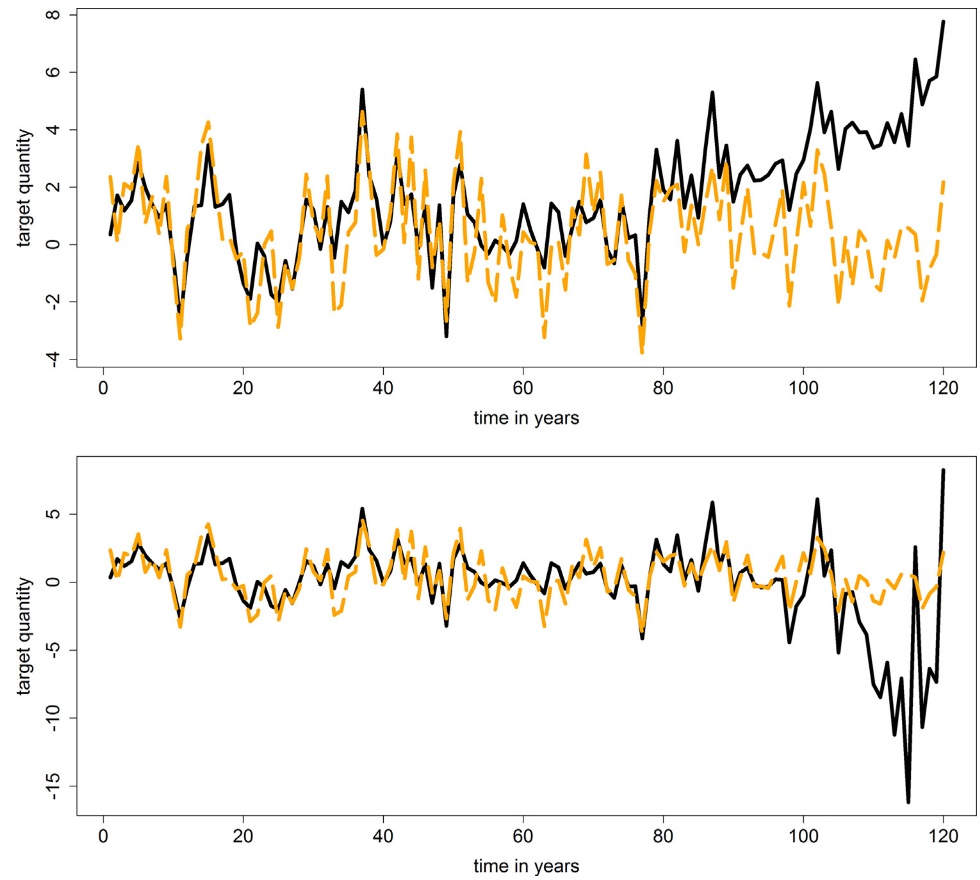

Fig 1. Examples of z-transformed pseudo-proxy datasets for most extreme cases of scenario I, i.e. trend-scenario (top) and II, i.e. noise-scenario (bottom). Solid black lines refer to the predictor (tree-ring parameter), dashed orange lines to the predictand (climate parameter). The term target quantity refers to any kind of variable that is to be predicted, e.g. mean monthly temperature. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

plementary Figs.). Each of the possible 210 combinations among time span, noise intensity, and level of instability was repeated 100 times for randomly subsampled data thereby obtaining each 21,000 pseudo-proxy sets for each of the scenarios I and II, thus altogether 42,000 sets. Fig. 1 illustrates each one example for scenario I and II, respectively with maximum time span, maximum noise, and maximum level of instability. The spectrum of r^2-values between predictor and predictand over all pseudo-proxy sets depends on the settings of noise and instability and varied from 0.28 to 0.88.

We are well aware that adding noise to the predictor variable – as done for the instable pseudo-proxy sets – violates the assumption of OLS regression of a noiseless predictor variable (e.g. Kutzbach et al., 2011). Consequently, model slopes will be underestimated, this also affecting model intercepts. However, as this violation occurs systematically and both BTFS and CCV rely on OLS, it will have no effect on the comparison between both. Nevertheless, we decided to use OLS since this is a frequently used approach in dendroclimatology. As mentioned above, it is possible to apply BTFS to other regression types such as inverse OLS and RMA. For details on alternative approaches with different noise assumptions see e.g. Kutzbach et al. (2011).

2.3. Comparative evaluation

In terms of validation, BTFS was applied to each of the 21,000 pseudo-proxy sets of scenarios I and II. For each of the pseudo-proxy sets we also computed standard cross-calibration verification (CCV) statistics (see introduction). Furthermore, to test the performance of bootstrapped RE and CE as proposed in the introduction, CCV was bootstrapped over i iterations from which respective ECDFs were derived (we call this approach BCCV). These ECDFs were used in a similar manner as for BTFS regression parameters, to determine whether the 95% confidence interval includes positive values. If not, BCCV would indicate an invalid transfer function. BCCV and CCV were considered instable if at least one of the four RE and CE values was (significantly) negative. The proportion of detected instabilities among the i iterations for each of the 210 combinations of time span, noise intensity, and level of instability was calculated for BTFS, CCV, and BCCV and compared among each other.

In terms of statistical comparison, we fitted two-parameter logistic regressions to the frequency of detected instabilities of BTFS and CCV for both scenarios. In these models, the frequency of detected instabilities was the dependent response variable and the level of instability the independent explanatory variable. Models were fitted individually for each of the possible 35 combinations of span (5 levels) and noise (7 levels) to account for a differing sensitivity of BTFS over these combinations (for an explanation see Section 4.2). From the resulting 35 model predictions per scenario (I vs. II) and approach (BTFS vs. CCV) – thus altogether 140 models – average prediction as well as the corresponding 95% confidence interval were computed to compare between BTFS and CCV models. Due to a generally weaker performance, BCCV was not considered in this comparison.

2.4. Implementation

All analyses were carried out using R version 3.2.2 (R Core Team, 2015) extended by the packages dplR (Bunn et al., 2015), lattice

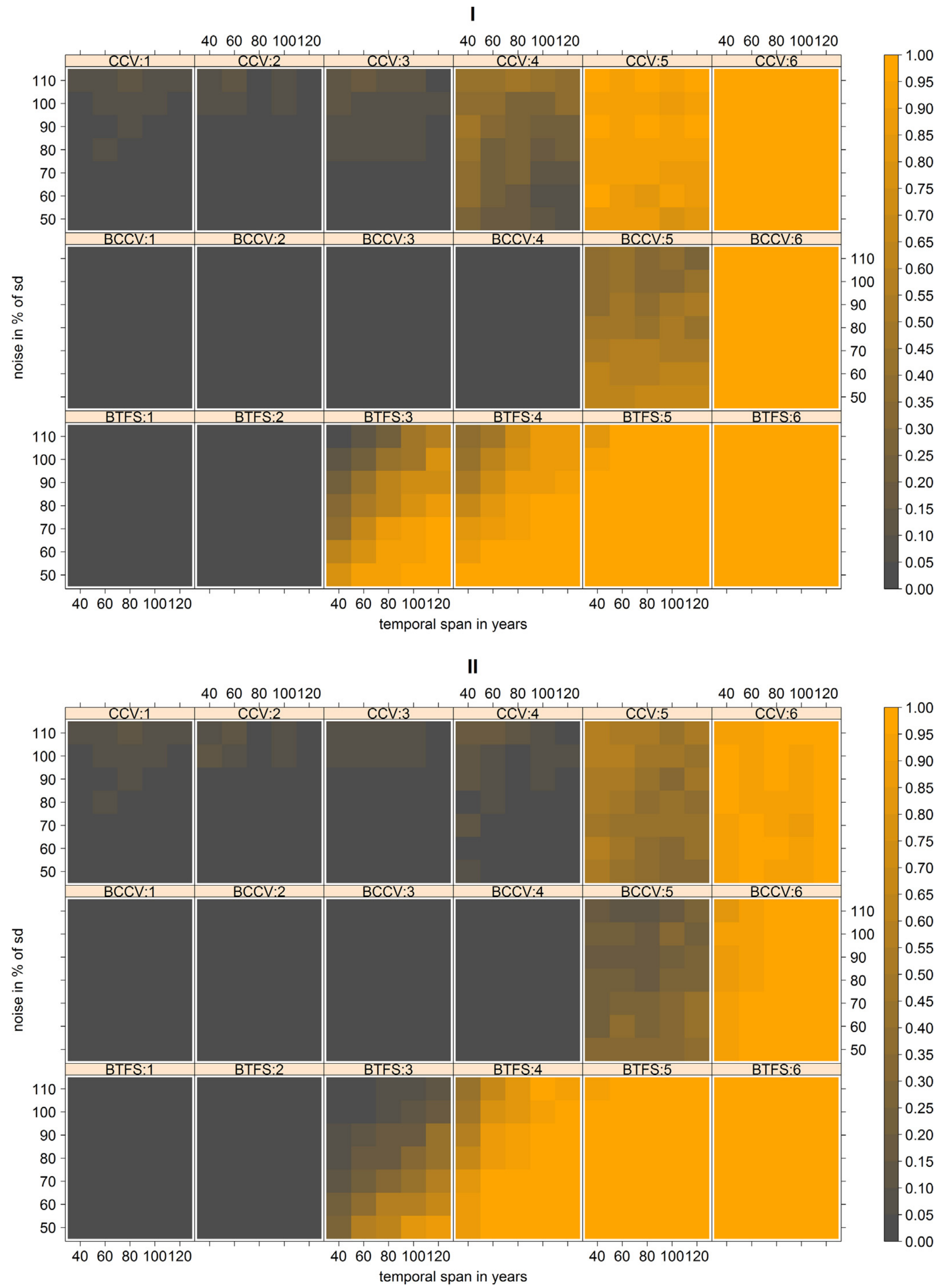

Fig. 2. Frequency of detected instabilities for scenario I (trend-scenario, top) and scenario II (noise-scenario, bottom) based on CCV (upper panels), BCCV (mid panels) and BTFS (lower panels) for different time spans (x-axis), noise intensities (y-axis), and strength of instabilities (increasing from left = stable to right = instable). It can clearly be seen that BTFS already responded at lower levels of instability than CCV and particularly BCCV for both scenarios.

(Sarkar, 2008), and ncdf (Pierce, 2014). The release of a BTFS R-package is intended.

3. Results

With respect to the trend-scenario (I), a clear difference was observed among BCCV, CCV and BTFS (Fig. 2, top). For the two lowest levels of instability, CCV detected instabilities in a few cases for pseudo-proxies having low r^2. Considering instability levels three through five, BTFS detected instability much more frequently than CCV and particularly BCCV. For BTFS, noise and temporal span significantly affected the number of detected instabilities (average r^2 of a respective additive model = 0.90, $p < 0.001$; lower mid and right panels, Fig. 2, top). RE and CE significantly decreased with increasing instability ($r = -1$, $p < 0.001$).

Regarding the noise-scenario (II), a similar behavior as for the trend-scenario was observed. That is, CCV again detected a low number of instabilities at low instability levels with low r^2. For

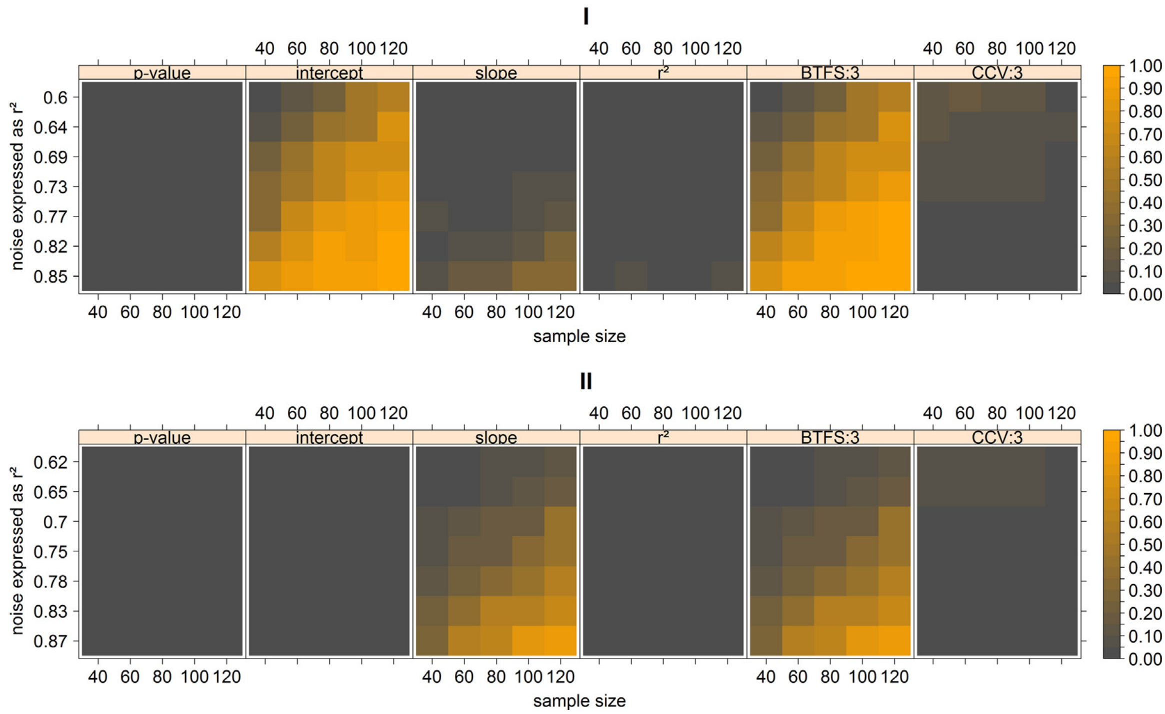

Fig. 3. Frequency of detected instabilities for single bootstrapped regression parameters (p-value, intercept, slope, r^2), the comprised evaluation of all four parameters (BTFS), and CCV for the third instability level of scenario I (upper panels) and II (lower panels). The two rightmost panels correspond with the respective panels 3 of Fig. 2. In contrast to Fig. 2 the values on the y-axis here demarcate the average r^2 for the different noise intensities. For similar figures representative of instability levels four to six please see Supplementary Figs. S1–S3.

higher instability levels, BTFS appeared more sensitive than CCV and was significantly affected by noise and temporal span (average r^2 of a respective additive model = 0.87, $p < 0.001$; Fig. 2, bottom). Again, BCCV performed worse compared to CCV. RE and CE significantly decreased with increasing instability ($r = -1$, $p < 0.001$).

Analysis of the single BTFS parameters revealed that the intercept responded most sensitively at moderate instability levels for the trend-scenario, but was superseded by model slopes at higher instability levels (Fig. 3 upper panel and Figs. S1–S3). In contrast to the trend-scenario, model slopes appeared to be most sensitive towards instability throughout all instability levels of the noise-scenario (Fig. 3 lower panel and Figs. S1–S3).

Two parameter logistic regressions were successfully fitted to the frequencies of detected instabilities, as indicated by r^2-values generally above 0.98 ($p < 0.001$) for all of the 140 computed models. For the two lowermost instability levels CCV models predicted a higher frequency of instability. Nevertheless, a clear response to increasing instability on average occurred at lower instability levels for BTFS, however with slightly higher confidence intervals (Fig. 4).

4. Discussion

4.1. Comparative evaluation

In our comparative evaluation of BTFS against CCV, we found that BTFS was more sensitive towards instability. For both scenarios BTFS responded with a higher frequency of detected instabilities at lower instability levels compared to CCV and BCCV.

Analyzing the CCV behavior along the gradient of instability reflects that CE and also RE are sensitive towards instability but stay positive until a certain point has been reached (fourth to fifth level of instability). In other words, below the fourth instability level, the residuals of the transfer function in most cases were smaller than the residuals of the null model, whereas they mostly were larger above those levels. Thus, the lower sensitivity of CCV has to be explained by the rigorous threshold of zero for RE and CE which does not account for the steady and significant decrease of RE and CE over increasing instability. In contrast, CCV detected a low number of instabilities at instability levels one to four with low r^2. As level one represents stable conditions, we interpret these detections as 'false positives', probably caused by influential outliers in pseudo-proxies with low r^2. This observation reflects the already known sensitivity of RE and CE towards outliers (Cook et al., 1994).

A possibility to cope with the sharp threshold of CCV is either to increase the threshold or to bootstrap RE and CE. However, bootstrapping of RE and CE (as done in BCCV) resulted in even fewer instability detections on the fifth and sixth level of instability. This is understandable, as 95% confidence intervals will contain positive values when the average RE and CE are just below zero. In such cases the transfer function fails for CCV but passes BCCV. Reconsideration of the critical value zero would be an option to cope with such effects, but any choice would arguably be more arbitrary than the original choice.

We explain the higher sensitivity of BTFS in comparison to CCV with BTFS relying on bootstrapped test-statistics of four model parameters (ratios of intercept, slope, and r^2 as well as least significance of regression) instead of one which only indirectly accounts for possibly changing model parameters (residuals in CCV and BCCV). The derived empirical cumulative distribution functions are advantageous because they allow for defining confidence intervals and deriving p-values. Thereby, robust estimates of the stability of transfer function parameters are achieved. At the same time, bootstrapped ratios of model parameters may serve as quantifiers for transfer function stability. In contrast, CCV is insensitive against changing model parameters as long as the residuals of the prediction stay below the residuals of the null model, which is reflected in its later response to instable transfer functions within our experiments.

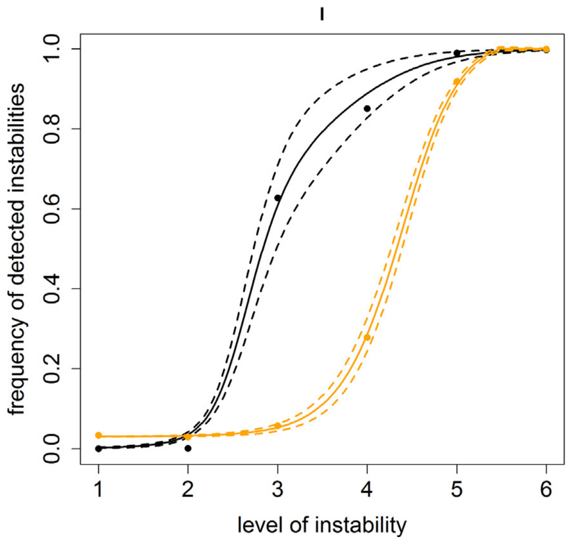

Fig. 4. Comparison of two parameter logistic regressions on the frequency of detected instabilities for BTFS (black) and CCV (orange) regarding the two scenarios. Solid lines represent the prediction mean, dashed lines the respective 95% confidence intervals of predictions. Respectively colored dots indicate the mean frequency of detected instabilities as obtained from our pseudo-proxy experiments. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4.2. BTFS performance in detail

Comprising fluctuations in intercept, slope, r^2, and representing model significance, BTFS covers several possibilities of instability. These are I) a changing offset (intercept) over time (main cause of weak instability in the trend scenario), II) a changing slope over time (main cause of instability in the noise scenario and strong instability in the trend-scenario), III) a change of explained variance over time (both scenarios), and IV) a change of significance of regression models over time (not covered by our scenarios, but theoretically possible).

The observed behavior of the single bootstrapped model parameters for the trend-scenario has a theoretical explanation. Model slopes are estimated using the covariance: $b = cov(x, y)/var(x)$, whereas model intercepts are derived from model slopes: $a = \bar{y} - b \cdot \bar{x}$. Covariance is only marginally affected for low trend instabilities, wherefore the increasing offset between predictor and predictand mainly is accounted for by adjusting the intercept. For stronger trend instabilities, the effect on covariance increases (also being reflected in more significant changes of r^2) leading to more significant changes of model slopes. As model slopes partially account for the offset between predictor and predictand, the change of intercept between calibration and verification period decreases, this resulting in more significant slope ratios than intercept ratios for higher levels of trend instability. The higher importance of slopes throughout all noise-scenarios we explain by the decreasing covariance between predictor and predictand while the variance of the predictor increases. Both effects lead to lower model slopes. Due to the negligible change of mean between calibration and verification period in the noise-scenario the intercept is only marginally affected.

The lower BTFS sensitivity for shorter periods and stronger noise we explain by increasing variability of bootstrapped model parameter ratios (intercept, slope and r^2). That is, for strong noise and/or short periods, the influence of single values gains a stronger weight, which increases the variability of parameter estimates over the iterations. The higher variability results in higher p-values derived from the empirical cumulative distribution function, this falsely indicating a stable transfer function ('false negative', i.e. type II error in statistics). A possibility to cope with this effect could be to increase the significance level. This however increases the probability of committing a type I error ('false positive'), i.e. detect instability where none is present. Instead of increasing the significance level, we therefore recommend reporting the respective p-value along with the temporal span, this indicating the reliability in the final decision of accepting the null-hypothesis. For combinations of low r^2 and short calibration periods, low BTFS p-values (e.g. $0.05 < p < 0.2$) should be interpreted carefully. For high p-values type II errors are less likely.

Despite the mentioned effects, our pseudo-proxy scenarios have shown that BTFS is more sensitive with respect to instability detection than CCV, thus has a comparably lower risk of type II errors than CCV over all levels of instability. It is well-known from statistics that the probability of committing type I errors is defined by the significance threshold, which for BTFS was set to 0.05 as relying on the bootstrapped 95% confidence interval. No such results can be given for the standard CCV approach which does not incorporate statistical tests.

4.3. Possible limitations of our experiments

It is important to stress that the behavior observed in our two scenarios largely depends on the specific data features of our pseudo-proxies and therefore cannot be generalized, since trends or noise may vary over time with properties differing from our pseudo-proxies. Thus, for real data one cannot straightforwardly assign instable transfer functions to one of our theoretical examples. Nevertheless, inspection of the single behavior of intercept, slope, and r^2 may give some insight into the causes of instability, if interpreted carefully in the scope of our results based on theory-motivated pseudo-proxies and their discussion.

In this context we want to highlight, that our experiments did not cover specific cases. That is, if predictor and predictand have shared low-frequency variance but differing high-frequency variance. For example one may imagine the predictor to be a straight line or a smooth sinusoidal wave whose low-frequency patterns (trend or wave) is reflected in the predictor, too, but obscured by a high-frequency noise. Real-world examples may be the reactivity of tree-growth on the long-term to solar cycles and atmospheric modes such as the Pacific Decadal Oscillation (e.g. D'Arrigo and Jacoby, 1991; Trouet and Taylor, 2010). If in these examples the low frequency signals dominate the variance in predictor and predictand and the relationship between both is stable over time, CCV and BTFS will indicate valid transfer functions. However, they will not reflect whether the high-frequency variations of predictor and predictand are synchronous or asynchronous. In such cases, an annually resolved reconstruction would suggest a false temporal precision wherefore a respective reconstruction should be low-pass filtered. To identify such cases, additional statistics to those provided by CCV and BTFS are needed. For instance, Gleichläufigkeit (Buras and Wilmking, 2015; Eckstein and Bauch, 1969) between predictor and predictand may be of value, as it reflects high-frequency synchronicity. In this context, sign-test (Fritts, 1976) is frequently used to quantify the shared high-frequency variation of predictor and predictand.

Finally, as mentioned in Section 2.2, using noisy predictors violates OLS assumptions, thereby leading to systematically underestimated model slopes (Kutzbach et al., 2011). This effect is of particular interest in the noise scenarios, where the underestimation of model slopes is pronounced for the period with enhanced noise. Though not being mathematically accurate, this effect will not affect the comparative evaluation of BTFS and CCV, as both rely on OLS regression which is commonly used in dendroclimatology despite the predictor variables (tree-ring parameters) likely being noisy. Here the incorporation of more sophisticated regression techniques such as SINOMA (Buras et al., 2014; Thees et al., 2014) into BTFS would allow for more accurate estimation of model slopes. Respective implementations are intended for future studies.

5. Conclusion

A comparative evaluation of 42,000 pseudo-proxy datasets revealed that the Bootstrapped Transfer Function Stability test (BTFS) supersedes the yet frequently used cross-calibration-verification and an alternative CCV approach based on bootstrapping (BCCV). BTFS was more sensitive towards simulated instabilities and provides users with probability estimates which can be reported as a measure of confidence in the transfer functions. We therefore recommend using BTFS as a new and more robust tool for testing the stability of transfer functions.

Acknowledgements

AB was funded by the Bavarian Staatsministerium für Ernährung, Landwirtschaft und Forsten (StMELF). This study is a contribution to the Virtual Institute of Integrated Climate and Landscape Evolution Analysis – ICLEA – of the Helmholtz Association, grant number VH-VI-415. We are grateful for the valuable comments by two anonymous reviewers.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.dendro.2017.01.005.

References

Bunn, A., Korpela, M., Biondi, F., Campelo, F., Mérian, P., Qeadan, F., and Zang, C., 2015. dplR: Dendrochronology Program Library in R. R package version 1.6.3. http://CRAN.R-project.org/package=dplR.

Buras, A., Wilmking, M., 2015. Correcting the calculation of Gleichläufigkeit. Dendrochronologia 34, 29–30.

Buras, A., Thees, B., Czymzik, M., Dräger, N., Kienel, U., Neugebauer, I., Ott, F., Scharnweber, T., Simard, S., Slowinski, M., Slowinski, S., Tecklenburg, C., Zawiska, I., Wilmking, M., 2014. SINOMA – a better tool for proxy based reconstructions. Geophys. Res. Abstr. 16, 1708–1701.

Cook, E.R., Kairiukstis, L.A., 1990. Methods of Dendrochronology: Applications in the Environmental Sciences. Kluwer Academic, Dordrecht, 393 pp.

Cook, E.R., Briffa, K.R., Jones, P.D., 1994. Spatial regression methods in dendrochronology: a review and comparison of two techniques. Int. J. Climatol. 14, 379–402.

D'Arrigo, R.D., Jacoby, G.C., 1991. A 1000-year record of winter precipitation from northwestern New Mexico, USA : a reconstruction from tree-rings and its relation to El Niño and the Southern Oscillation. Holocene 1, 95–101.

Eckstein, D., Bauch, J., 1969. Beitrag zur Rationalisierung eines dendrochronologischen Verfahrens und zur Analyse seiner Aussagesicherheit. Forstwissenschaftliches Centralblatt 88, 230–250.

Fritts, H.C., 1976. Tree Rings and Climate. Academic Press, London, 567 pp.

Kutzbach, L., Thees, B., Wilmking, M., 2011. Identification of linear relationships from noisy data using errors-in-variables models – relevance for reconstruction of past climate from tree-ring and other proxy information. Clim. Change 105, 155–177.

Pierce, D., 2014. ncdf: Interface to Unidata netCDF data files. R package version 1.6.8. http://CRAN.R-project.org/package=ncdf .

R Core Team, 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria https://www.R-project.org/.

Sachs, H., 1977. Palaeoecological reponse functions. Ann. Rev. Earth Planet. Sci. 5, 159–178.

Sarkar, D., 2008. Lattice: Multivariate Data Visualization with R. Springer, New York, ISBN 978-0-387-75968-5.

Thees, B., Buras, A., Jetschke, G., Kutzbach, L., Zorita, E., Wilmking, M., 2014. SINOMA – a new iterative statistical approach for the identification of linear relationships between noisy time series. Geophys. Res. Abstr. 16, 1714.

Trouet, V., Taylor, A.H., 2010. Multi-century variability in the Pacific North American circulation pattern reconstructed from tree rings. Clim. Dyn. 35, 953–963.

Wahl, E.R., Smerdon, J.E., 2012. Comparative performance of paleoclimate field and index reconstructions derived from climate proxies and noise-only predictors. Geophys. Res. Lett. 39, L06703.

Wilson, R., D'Arrigo, R., Buckley, B., Büntgen, U., Esper, J., Frank, D., Luckman, B., Payette, S., Vose, R., Youngblut, D., 2007. A matter of divergence: tracking recent warming at hemispheric scales using tree ring data. Clim. Dyn. 112, D17103.