

A methodological critique on using temperature-conditioned resampling for climate projections as in the paper of Gerstengarbe et al. (2013) winter storm- and summer thunderstorm-related loss events in *Theoretical and Applied Climatology* (TAC)

Frank Wechsung¹ · Maximilian Wechsung²

Received: 30 July 2014 / Accepted: 29 July 2015 / Published online: 16 August 2015
© Springer-Verlag Wien 2015

Abstract The STatistical Analogue Resampling Scheme (STARS) statistical approach was recently used to project changes of climate variables in Germany corresponding to a supposed degree of warming. We show by theoretical and empirical analysis that STARS simply transforms interannual gradients between warmer and cooler seasons into climate trends. According to STARS projections, summers in Germany will inevitably become dryer and winters wetter under global warming. Due to the dominance of negative interannual correlations between precipitation and temperature during the year, STARS has a tendency to generate a net annual decrease in precipitation under mean German conditions. Furthermore, according to STARS, the annual level of global radiation would increase in Germany. STARS can be still used, e.g., for generating scenarios in vulnerability and uncertainty studies. However, it is not suitable as a climate downscaling tool to access risks following from changing climate for a finer than general circulation model (GCM) spatial scale.

The recently published paper of Gerstengarbe et al. (2013) in TAC presents the STatistical Analogue Resampling Scheme (STARS) which uses the temperature-conditioned resampling (TCR) in order to project changes of covariables (ΔY) for a given temperature increase ($+\Delta T$). The TCR algorithm

STARS described in Orłowsky et al. (2008) draws with replacement sections from past weather records. At the first step (annual level), full yearly data records are resampled. At the second step (12-day block level), daily weather sequences of 12 days are drawn. The resampling is restricted (i.e., conditioned) at both levels by a supposed temperature increase. Due to the low variance of mean annual temperatures, climate-relevant temperature increases (+1 to +3 K) cannot be achieved during the first step. Thus, the second step is used to build the needed temperature (T) trend, and it is possible due to the much higher temperature variability at this level. Further details of the STARS algorithm are given by Gerstengarbe et al. (2013) and Orłowsky et al. (2008).

The authors of the recent paper in TAC claim for resampled time series for Germany that the similarity between the mean statistics of the observed and resampled values proves the validity of the STARS algorithm. Furthermore, they suggest that TCR as realized in STARS could be used for the temperature conditioned climate scenario projections. As will be shown below, both statements are wrong.

In general, a good representation of mean statistical estimators (e.g., mean, standard error, percentiles) can be assumed by “any” resampling algorithm as long as the mean climatic characteristics of the sample period (i.e., 1951–1975 in Gerstengarbe et al. (2013)) are close to that of the validation period (i.e., 1976–2000 in Gerstengarbe et al. (2013)). Furthermore, covariable Y should change with increasing T following the covariances between Y and T in the data. Expressed in a simplified manner, the change in Y , ΔY , for a given ΔT should result from $\Delta Y = \frac{\Delta Y}{\Delta T} \cdot \Delta T$.

The similarity of statistical measures of the original and resampled data follows from the characteristics of bootstrapping procedures (Efron 1979; Liu 1988; Liu and Singh 1995; Kunsch 1989). In the discussed paper, the sample

✉ Frank Wechsung
wechsung@pik-potsdam.de

Maximilian Wechsung
maximilian.wechsung@uni-jena.de

¹ Potsdam Institute for Climate Impact Research,
Potsdam, Brandenburg, Germany

² Friedrich-Schiller-Universität Jena, Jena, Thuringia, Germany

and resampling periods (1951–75, 1976–2000) differ only slightly in average T (about 0.5 K, own calculation). Consequently, the reported general similarities between the observed and resampled data for 1976–2000 are not surprising and do not prove the validity of the algorithm. Nevertheless, if the covariances in the data form an effective $\Delta Y/\Delta T$ that differs significantly from zero, the validation could even reveal some discrepancies for a small increase of mean temperature. Indeed, this is the case for the Y variable precipitation in the described study. The resampled precipitation mean shows a major discrepancy compared with the observed one in 1976–2000: the precipitation mean was underestimated by -36.92 mm (see Table 1 in Gerstengarbe et al. (2013)). In our opinion, the fact was not sufficiently acknowledged by the authors. It is significant and occurs for an only small mean temperature difference between the two periods. Normalized for a 1-K temperature change, the bias would amount to about -74 mm/K. Following Fig. 4 in Gerstengarbe et al. (2013), this more than outweighs the extent of the presented precipitation changes per degree warming during summer and winter from simulations of STARS for the period 2010–2060. Thus, the reported proximity in $\Delta P/\Delta T$ between general circulation model (GCM) and STARS simulations might disappear when this bias was taken into account. As will be shown below, the STARS bias is not only empirically significant but also conceptually consistent with the unsuitability of TCR for climate projections.

The unsuitability of TCR for climate projections follows from the formal inequality of the interannual temperature sensitivity of Y ($\Delta Y/\Delta T$) and the temperature trend sensitivity ($\Delta Y/\Delta t/\Delta T/\Delta t$), and from the empirical divergence of both sensitivities. A theoretical analysis of this inequality was recently presented by the authors (Wechsung and Wechsung 2014).

The interannual temperature sensitivity and the temperature trend sensitivity are only equivalent for the singular case of deterministic relationships (i.e., correlations equal one) between T and Y on one hand and T and t and P and t on the other hand (Wechsung and Wechsung 2014). Under such conditions, any interannual change in Y and T would be an expression of the long-term climate change. However, such a trivial case would have resolved a long time ago questions about possible future climates.

Due to the fact that TCR in STARS is conditioned on ΔT and not on $\Delta T/\Delta t$, it will follow the interannual correlations $\rho(Y, T)$ and not the trend relationships $\rho(\Delta Y/\Delta t, \Delta T/\Delta t)$. Thus, STARS is not able to extend past climate trends. Instead, it turns interannual variability into climate trends.

Nevertheless, an empirical plausibility of the STARS approach could be still given if the interannual correlations between temperature and covariables, $\rho(Y, T)$, and the correlations between the trend changes $\rho(\Delta Y/\Delta t, \Delta T/\Delta t)$ did not diverge meaningfully at the two resampling levels (years, 12-day blocks). We can test this in real datasets using the interannual

correlation $\rho(Y, T)$ and the temporal correlation $\rho(Y, t)$. The latter should at least diverge from zero and follow the sign direction of $\rho(Y, T)$, if we assume an existing warming trend.

In Fig. 1a, b, we present for the two Y variables precipitation (P) and global radiation (R) the correlations $\rho(Y, T)$ and $\rho(Y, t)$ for 12-day blocks ordered by month based on averaged data for Germany. The correlations $\rho(Y, T)$ were calculated for the first 25 years, 1951–1975 and the whole period, 1951–2010 (shown as black boxes). The correlations $\rho(Y, t)$ were determined for the whole period (gray boxes). The correlations $\rho(Y, T)$ for the first 25 years and the whole period follow the same monthly pattern. Thus, long-term changes between Y and T were not affecting the correlations. This is not surprising because the temporal correlations of both P and R for the whole period fluctuated mostly in the band of insignificant changes, which is consistent with $\rho(\Delta Y/\Delta t, \Delta T/\Delta t) \sim 0$. From $\rho(P, T) > 0$ in winter and $\rho(P, T) < 0$ in summer (Fig. 1a), it follows that STARS will tend to generate dryer summer and wetter winter months in the second resampling step, while the long-term trends do not indicate major changes.

Both tendencies are unlikely to cancel out each other under German conditions. First of all, there is a dominance of months with negative correlations between P and T (April to October) over the months (November–March) with positive correlation. The absolute values for the mean 12-day sensitivity to temperature change (-0.13 mm/1 K***, $+0.12$ mm/K***, *** significant at $p \leq 0.01$ level) are similar for both periods. From the dominance of negative correlations at the given temperature sensitivity values alone, a negative annual net change in precipitation can be expected. This tendency will be further strengthened by STARS date flexibility during the resampling process. Warmer days will be not only taken exactly from the same date within a year but from dates that surround the actual sample date (e.g., ± 50 days). As a consequence, warmer spring and autumn days with negative interannual correlations between P and T will drift into the winter season (Wechsung and Wechsung 2014).

The sensitivity of precipitation to annual temperature changes was small for Germany in 1951–1975. Thus, while the first resampling step will generate no specific changes in P , the second resampling step will tend to result in an annual precipitation decrease with increasing temperatures, and exactly this was realized in the reported STARS scenario for Germany (Gerstengarbe et al. 2013). The precipitation bias reported above is a logical consequence of the interannual correlations between P and T at the 12-day block level that control the second resampling step.

The principal flaw of the STARS algorithm can be even more convincingly demonstrated by a model intercomparison experiment (Wechsung and Wechsung 2014). We used a climate simulation run of the MPI-ESM-LR climate model for the RCP 8.5 scenario (Giorgetta et al. 2013). The simulation

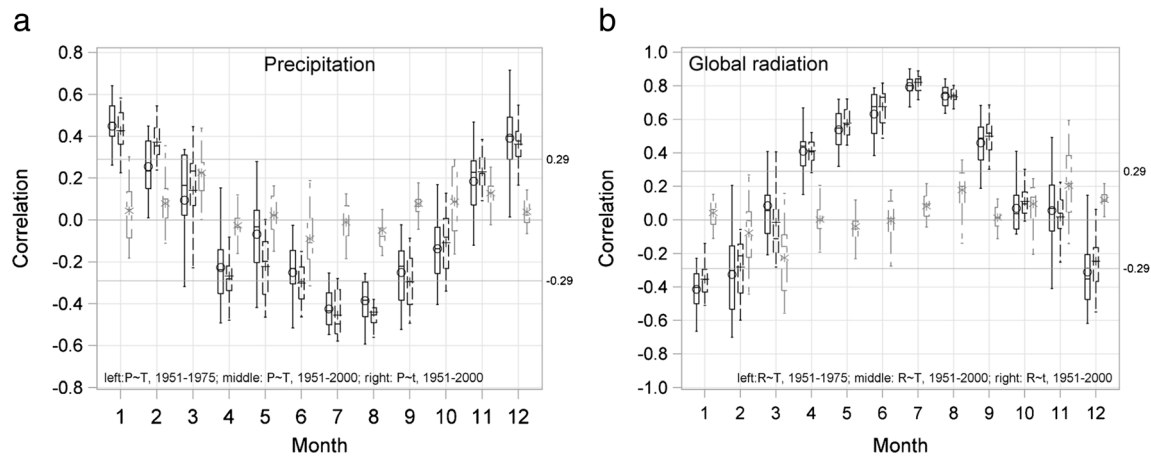


Fig. 1 Monthly ordered running 12-day correlations of **a** precipitation (P) and **b** global radiation (R) to temperature for Germany. In both figures, the *left* two boxes per month (*black*) represent the interannual correlations with temperature (T) in the periods 1951–1975 (*solid*) and 1951–2010 (*dashed*). The *right* boxes per month (*dashed gray*) show the temporal correlations of P and R for the period 1951–2010. The single

correlations are based on spatial means of 12-days running averages derived from daily weather records of 1218 German weather stations. The 0.95 levels for correlations significantly different from zero are represented by corresponding values and gray reference lines. They relate to the *middle* and *right* boxes

was carried out for the AR5 report of the IPCC. The data from 1959 to 2100 and German grid cells were spatially averaged by day. The period 1959–2005 was used as a training period for STARS. The sensitivities of mean 12-day precipitation to interannual temperature change are -0.15 mm/K^{***} and $+0.15 \text{ mm/K}^{***}$ for the periods April–October and November–March, respectively. The interannual sensitivities for global radiation are 10.1 W/m^2^{***} and $-0.85 \text{ W/m}^2^{***}$ for the same periods. The simulated temperature has a significant linear trend during the training period ($+3 \text{ K/100 years}^{***}$). In contrast, the time series of global radiation and precipitation do not show any trend during that period.

The STARS simulations were carried out in two steps firstly for the period 2006–2052 and secondly for the period 2054–2100. The length of the two periods equals that of the sample period. Taking the linear annual temperature trend of the original time series for the two periods (from 9.23 to 10.8 °C, from 10.87 to 12.44 °C), the climate for the periods 2006–2052 and 2054–2100 was simulated in 100 statistical realizations and compared with the original MPI-ESM-LR simulations.

If STARS would work properly as a statistical downscaling tool, the STARS output for 2006–2100 should be more or less randomly distributed around the MPI-ESM-LR data series. For temperature, this is the case (Fig. 2a) as STARS outputs for T follow the imposed trend. In contrast, the simulated 100-year trends for global radiation ($+32 \text{ W/m}^2^{***}$) and precipitation (-102 mm^{***}) diverge markedly from those in the original MPI-ESM-LR time series, which do not show significant trends for both variables (Wechsung and Wechsung 2014, Fig. 2b, c). The divergence develops into the expected directions.

The global radiation increases because the warming during most parts of the year can be only realized by a preferential

resampling of days with higher global radiation levels. As one consequence, the ratio between the short- and longwave downward radiation increases over time in STARS resamples instead to decrease as expected from the physics of global warming and simulated by MPI-ESM-LR (Fig. 2b). Furthermore, the preference for brighter days particularly during summer months implies a preference for stiller days with less eastward winds (Fig. 2c) and a decrease in the maximum daily precipitation during summer (Fig. 2d). Both should be highly relevant for future hail and thunderstorm risk.

In summary, the TCR model STARS transforms interannual gradients between warmer and cooler 12-day weather means into climate gradients. This conceptual flaw does not exclude chances for possible agreements between some STARS and GCM outputs. Such a similarity occurs when comparing the seasonal temperature sensitivity of precipitation trends as depicted by Fig. 4 in Gerstengarbe et al. (2013). However, this similarity could be produced with any weather record subsampled from the last 200 years following Casty et al. (2007). During that time, warmer (cooler) summer was associated with dryer (wetter) weather and warmer (cooler) winter correlated with wetter (drier) weather. Consequently, STARS would project dryer summer and wetter winters from any 40-year sample period of the last 200 years for an imposed temperature increase. In this context, it is interesting to note that the rightly calculated trend sensitivities of precipitation to temperature change ($\Delta P/\Delta t/\Delta T/\Delta t$ supplied by (Menz 2014)) are reported as simple differentials $\Delta P/\Delta T$ in Gerstengarbe et al. (2013). That is an explicit indication of the misunderstanding discussed above.

The problems discussed here are not unique to STARS. They apply also to other TCR-based algorithm as the k- nn algorithm (Yates et al. 2003). The latter was originally used

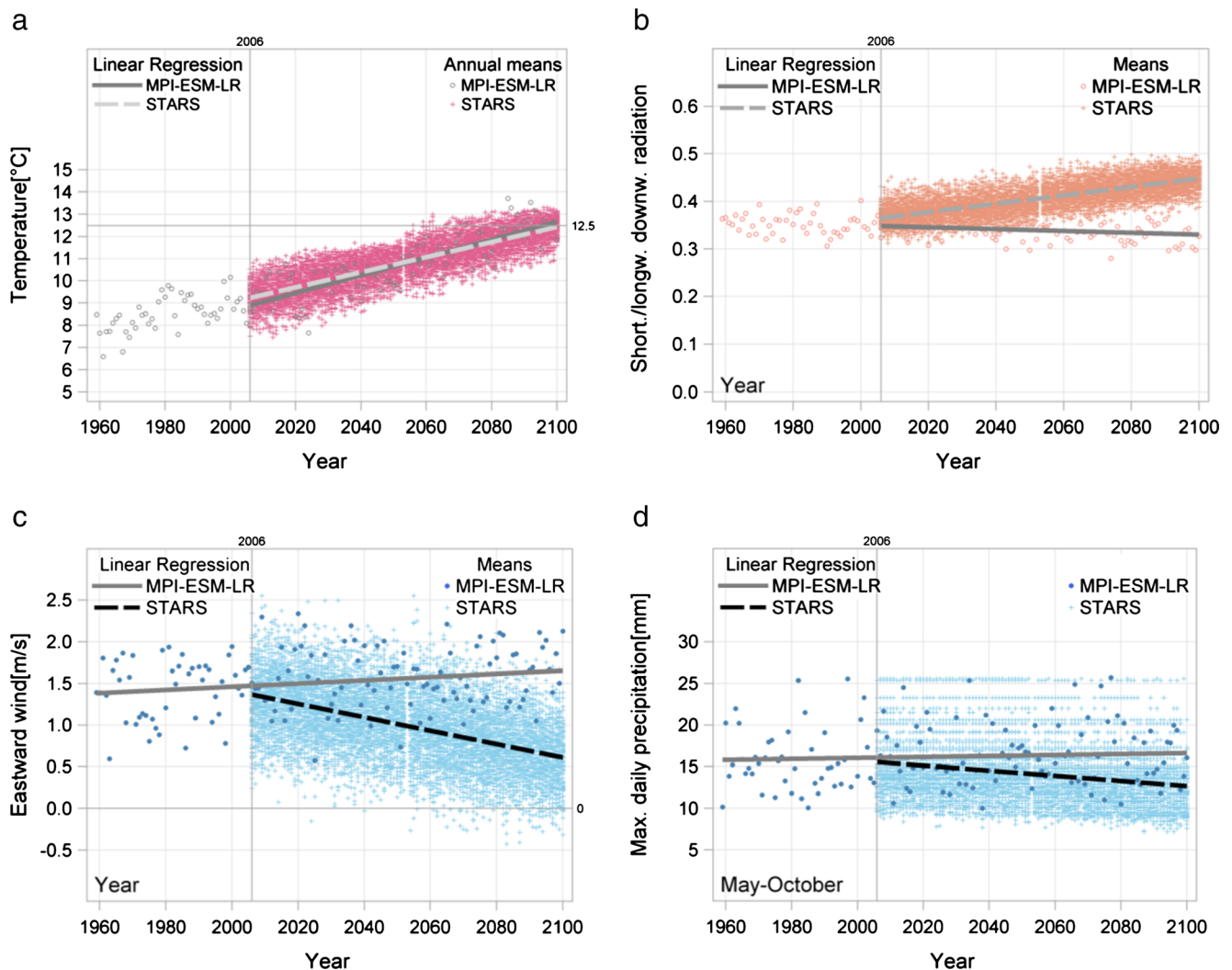


Fig. 2 Application of the STARS algorithm to MPI-ESM-LR data: higher scenario levels in annual temperature (**a**) request the preferential resampling of warmer, brighter, and stiller 12-day blocks in spring,

summer, and autumn leading to increasing ratios between short- and longwave downward radiation (**b**), a decrease of eastward winds (**c**), and a decrease in the maximum daily precipitation during summer (**d**)

to extend stationary weather records. Yates et al. (2003) used the same algorithm also to generate climate scenarios for an imposed temperature increase similar to STARS. However, the authors were well aware that the generated tendencies for covariables reflect the interannual correlations between those and temperature. Thus, the application was limited to vulnerability studies only. Such analog scenarios have been shown to be valuable tools to explore the vulnerability of the water cycle to possible climate change (Wechsung et al. 2008).

However, in the German context, STARS is suggested and still can be used to generate climate scenarios with warmer summer and wetter winter which are not triggered by an increase of greenhouse gasses but by changes in the global radiation. As explained above, STARS is not a suitable tool for downscaling coarser scenario results from global circulation models. Therefore, interpretations

in Gerstengarbe et al. (2013) based on STARS simulations of future climate can be strongly misleading. Although the suggested increase of hail events seems plausible, the scientific basis for this conclusion is not. The coincidence between STARS outcome and expectation could be due to a dominating influence of the increasing vapor pressure on the results. A detailed discussion on this issue is not possible on the basis of the presented information in the Material and Method section of Gerstengarbe et al. (2013).

This methodological critique does not generally apply to statistical methods that are used to downscale climate scenarios of a coarser resolution to a finer scale. However, the chosen analytical method might be helpful also to test other downscaling techniques for their feasibility. As was shown above, possible fundamental deficits might be easily overseen due to seemingly good validation results.

Acknowledgement The authors like to thank one anonymous reviewer for his/her idea of a model intercomparison in order to demonstrate the characteristics of STARS. They also thank Christopher Menz for carrying out the model intercomparison experiment. The authors are grateful to Valentina Krysanova and Zbyszek Kundzewicz for their comments on previous draft versions.

We finally like to acknowledge the role STARS simulations have played in the past for developing impact assessments that used an ensemble of climate scenario realizations instead of a single one at times when ensembles of RCM and GCM simulations were not as easily available as today.

References

- Casty C, Raible CC, Stocker TF, Wanner H, Luterbacher J (2007) A European pattern climatology 1766–2000. *Climate Dynamics* 29(7–8):791–805
- Efron B (1979) 1977 Rietz lecture—bootstrap methods—another look at the jackknife. *Ann Stat* 7 (1):1–26. doi:DOI 10.1214/aos/1176344552
- Gerstengarbe F-W, Werner P, Österle H, Burghoff O (2013) Winter storm- and summer thunderstorm-related loss events with regard to climate change in Germany. *Theoretical and Applied Climatology* 114(3–4): 715–724. doi:10.1007/s00704-013-0843-y
- Giorgetta MA, Jungclaus J, Reick CH, Legutke S, Bader J, Böttinger M, Brovkin V, Crueger T, Esch M, Fieg K, Glushak K, Gayler V, Haak H, Hollweg H-D, Ilyina T, Kinne S, Kornblueh L, Matei D, Mauritsen T, Mikolajewicz U, Mueller W, Notz D, Pithan F, Raddatz T, Rast S, Redler R, Roeckner E, Schmidt H, Schnur R, Segsneider J, Six KD, Stockhouse M, Timmreck C, Wegner J, Widmann H, Wieners K-H, Claussen M, Marotzke J, Stevens B (2013) Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project phase 5. *Journal of Advances in Modeling Earth Systems* 5(3):572–597. doi:10.1002/jame.20038
- Kunsch HR (1989) The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics* 17(3):1217–1241
- Liu RY (1988) Bootstrap procedures under some non-lid models. *Ann Stat* 16 (4):1696–1708. doi:DOI 10.1214/aos/1176351062
- Liu RY, Singh K (1995) Using iid bootstrap inference for general non-lid models. *Journal of Statistical Planning and Inference* 43 (1–2):67–75. doi:Doi 10.1016/0378-3758(94)00008-J
- Menz C (2014) Calculation of dP/dT for the intercomparison of German GCM and STARS simulations. Personal Communication
- Orlowsky B, Gerstengarbe FW, Werner PC (2008) A resampling scheme for regional climate simulations and its performance compared to a dynamical RCM. *Theoretical and Applied Climatology* 92(3–4): 209–223
- Wechsung F, Kaden S, Behrendt H (2008) Klöcking B (eds). *Integrated analysis of the impacts of global change on environment and society in the Elbe Basin Weißensee Verlag, Berlin*
- Wechsung F, Wechsung M (2014) Drier years and brighter sky—the predictable simulation outcomes for Germany’s warmer climate from the weather resampling model STARS. *International Journal of Climatology*:n/a-n/a. doi:10.1002/joc.4220
- Yates D, Gangopadhyay S, Rajagopalan B, Strzepek K (2003) A technique for generating regional climate scenarios using a nearest-neighbor algorithm. *Water Resources Research* 39(7)