Reply to "Comment on 'Bias Correction, Quantile Mapping, and Downscaling: Revisiting the Inflation Issue'"

DOUGLAS MARAUN

GEOMAR Helmholtz Centre for Ocean Research Kiel, Kiel, Germany

(Manuscript received 5 June 2013, in final form 8 November 2013)

ABSTRACT

In his comment, G. Bürger criticizes the conclusion that inflation of trends by quantile mapping is an adverse effect. He assumes that the argument would be "based on the belief that long-term trends and along with them future climate signals are to be large scale." His line of argument reverts to the so-called inflated regression. Here it is shown, by referring to previous critiques of inflation and standard literature in statistical modeling as well as weather forecasting, that inflation is built upon a wrong understanding of explained versus unexplained variability and prediction versus simulation. It is argued that a sound regression-based downscaling can in principle introduce systematic local variability in long-term trends, but inflation systematically deteriorates the representation of trends. Furthermore, it is demonstrated that inflation by construction deteriorates weather forecasts and is not able to correctly simulate small-scale spatiotemporal structure.

In a recent contribution, I identified shortcomings of quantile mapping and other deterministic bias correction approaches that are designed to correct—in addition to the mean—the variability of a numerical model simulation (Maraun 2013): if applied to downscale processes with high spatial variability to smaller scales, these approaches overestimate the spatial extent of events in the extreme tails and incorrectly modify trends. The underlying reason is that these correction approaches do not produce random small-scale variability that is not explained by the grid box simulated value. Instead, the simulated gridbox variability is inflated to match the total small-scale variability. As such, quantile mapping and related approaches—if used for downscaling—are similar to the method of inflation, a concept that has been known to be flawed for a long time (Glahn and Allen 1966; von Storch 1999). For the simulation of local time series, von Storch (1999) and Maraun (2013) propose "randomization" instead; that is, to add stochastic noise.

In his comment, Bürger (2014) mainly raises two points. First, he disagrees with my conclusions about trends: they would imply that local trends were representative of large-scale trends. He also states that inflation would not affect trends. Second, he refers back to the inflation discussion in perfect prognosis ("prog") statistical downscaling, criticizes the concept of randomization, and advocates the use of inflation. In the following, I will refute Bürger's criticism and demonstrate that the reasoning underlying the concept of inflation is based on an imprecise understanding of explained versus unexplained variability and prediction versus simulation. The criticism can most precisely be addressed referring to regression models. Therefore, I will begin with briefly reviewing the concept of regression models and the proposed method of inflated regression. Thereafter, I will address the points Bürger (2014) raised about regression in general and finally focus on his arguments about trends in particular.

Assume the following linear relationship between a predictor x_i and a predictand Y_i (Davison 2003):

$$Y_i = ax_i + b + \eta_i. \quad (1)$$

The x_i are treated as known values, the $\eta_i \sim \mathcal{N}(0, \sigma^2)$ are treated as a sequence of normally distributed random variables, and thus Y_i is also a normally distributed

Corresponding author address: Douglas Maraun, GEOMAR Helmholtz Centre for Ocean Research Kiel, Düsternbrooker Weg 20, 24105 Kiel, Germany.
E-mail: dmaraun@geomar.de

DOI: 10.1175/JCLI-D-13-00307.1

random variable with variance σ^2 and time-varying mean $\mu_i = ax_i + b$. A key idea of regression models is that the predictor does not fully determine the predictand; that is, the variance of the y_i is larger than the variance of the ax_i. The role of the η_i is then to represent all the variance not explained by the predictor as random noise. Given N pairs of observations (x_i, y_i), $i = 1, \ldots, N$, of predictor and predictand, the unknown parameters a, b, and σ^2 can be estimated by maximum likelihood.[1] In climate science the y_i could, for example, represent observed local-scale temperature observations. In a typical perfect prognosis context, the x_i could represent the large-scale air mass and atmospheric circulation, and the a and b would model the influence of the predictors on local temperature, accounting for systematic local effects such as orography. The η_i would represent mesoscale and small-scale random variability not determined by the predictors. In a typical model output statistics context, the x_i could represent simulated gridbox precipitation in a high-resolution reanalysis. Then a and b correct systematic biases between the simulated temperature and the observed temperature, and η_i models the small-scale variability that is not explained by the gridbox scale predictor. This model can easily be generalized to include nonlinear predictor influences (Davison 2003), non-Gaussian distributed predictands (Dobson 2001), or predictor influences not only on the mean but also on, for example, the variance (Yee and Wild 1996).

Figure 1 illustrates the regression model. For the sake of simplicity it is assumed that $b = 0$. The parameters a and σ^2 are estimated from simulated data pairs (x_i, y_i), marked as gray dots. The best-fit model is depicted by the thick blue line. Remaining residuals, distributed according to η_i, are shown as thin blue lines. By construction, the sum of squared residuals is minimized by the least squares estimator. The blue shading indicates the range of variability η_i not explained by ax_i.

Now two cases have to be distinguished, serving completely different purposes. The first case is prediction. Given a predictor value x, one can of course not precisely predict a future observation y because of the noise η_i. Instead one can only predict the mean ax and the distribution around the mean for a given x. Thus, predictions based on regression models are actually probabilistic predictions: the blue line does not predict a particular value but rather the mean of the distribution, according to which a value should be observed. The blue shading represents possible prediction intervals for chosen levels of confidence (Wilks 2006; Davison 2003).

[1] That is, least squares in the case of normally distributed noise.

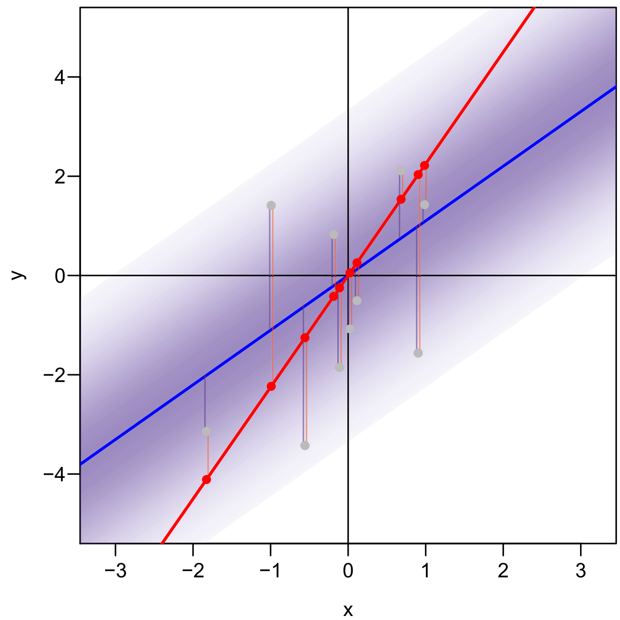

FIG. 1. Regression and inflation. Gray dots represent data pairs; the thick blue line represents the best-fit model; the thick red line represents inflated regression; the thin lines represent residuals; and the blue shading represents normal distribution according to the unexplained variance.

The second case is simulation. If the regression model is used to produce time series of Y_i given a series x_i, for example, to drive a climate impact model, one is of course interested neither in the time series of expected values ax_i nor in a predicted probability distribution. In most cases, one rather aims to simulate time series with the same statistical properties as Y_i; that is, a time series additionally containing the unexplained variability represented by η_i. In this context, it has been claimed that statistical downscaling is ‘‘smoothing’’ or ‘‘underpredicting’’ local-scale variability (e.g., Wilby et al. 2004). This is of course only the case if the predicted mean ax_i is taken as complete representation of the process at the local scale. However, the regression contains information on the unexplained local-scale variability as well, given by the distribution of η_i. Consequently, to simulate the total variability, one has to add realizations of η_i to the prediction ax_i; that is, by drawing random numbers. This standard procedure is known as simulation in statistics (e.g., Davison 2003). In climate science, it has been coined randomization (von Storch 1999) and is done by every weather generator[2] (e.g., Maraun et al. 2010).

[2] Although weather generators in general additionally use Markov processes to account for autocorrelated η_i.

In the following, I will show that inflation is appropriate for neither prediction nor simulation. In the early years of numerical weather forecasting, a tendency of regression methods to "not forecasting the extremes as often as they are observed" (Klein et al. 1959) had been noted, basically because erroneously only the predicted mean ax had been considered for the forecast but not the distribution around it. As a "correction" inflation has been suggested [e.g., Klein et al. 1959; see also Bürger (2014), and references therein]. The slope of the regression a is increased to $a_{\mathrm{inf}} = \mathrm{sd}(y_i)/\mathrm{sd}(ax_i)a$, where sd() refers to standard deviation, such that the variance of the prediction $a_{\mathrm{inf}}x_i$ (the red points) matches the total variance. This inflated regression is depicted in Fig. 1 by the red line. The inflated model obviously strongly diverges from the best-fit line (the more, the less variance is explained by the predictors). The fact that inflation thus increases the root-mean-square error between prediction and observations has been highlighted already by Glahn and Allen (1966) and later by von Storch (1999). In the language of modern forecast verification (Wilks 2006; Jolliffe and Stephenson 2003), an inflated forecast has a low accuracy (as the root-mean-square error is unnecessarily high) and also a low reliability: the conditional bias $E[Y\,|\,aX = y_f] - y_f$ is larger than that of the uninflated regression (which is zero in the idealized case of a perfectly linear relationship); that is, the expected value of the observations for a given forecast y_f is different from the forecast by $(a_{\mathrm{inf}}/a - 1)\,y_f$. The miscalibration also affects the overall skill of the forecast. To illustrate the effect of inflation on the prediction of threshold excesses, 100 000 realizations of the discussed example with prescribed correlations between X and Y of 0.5, 0.7, and 0.9 (representing predictors of different quality) have been simulated.

The accuracy of probabilistic predictions of threshold excesses can be quantified by the Brier score (Wilks 2006; Jolliffe and Stephenson 2003; Friederichs and Thorarinsdottir 2012): $1/N\sum(p_i - o_i)^2$, where p_i denotes the probability of exceedance and o_i is one for an exceedance and zero otherwise. In case of deterministic forecasts, p_i is either one or zero. Brier skill scores have been estimated with the climatology as reference forecast (i.e., a stationary normal distribution fitted to the observations). Results for two thresholds, the median and the 95th percentile of the observations y_i, are presented in Table 1. It can be seen that, for predicting excesses of the median, inflation is by construction (because the regressions intersect) as good as the uninflated regression, interpreted as a deterministic forecast. However, for the 95th percentile (and all other quantiles as well; not shown), inflated regression actually has less skill of correctly predicting excesses (because too many excesses are predicted where none occurred). Also by construction, the reduction in skill is of course lowest for high correlations where inflation is weak. For strong inflation (low correlations), the skill of inflation is even worse than that of a climatological forecast (negative values). For low correlations, also the uninflated predicted mean becomes worse than the climatological forecast. In every case, however, the correct interpretation of the regression model as a probabilistic forecast yields consistently best results. In other words, inflation hedges (Wilks 2006; Jolliffe and Stephenson 2003) the optimal deterministic forecast (the predicted mean) to improve the hit rate. Proper scores (e.g., the Brier score; Jolliffe and Stephenson 2003) identify that this "cheating" works only to the expense of consistently overpredicting (underpredicting) high (low) values. Thus, ironically, inflation yields worse forecasts (in terms of the major aspects of forecast quality: accuracy, reliability, and skill) for extremes; that is, the opposite of what it was initially designed for.

TABLE 1. Brier skill score for forecasts of the median and 95th percentile of Y, for different strengths of the correlation (corr) between predictor and predictand. Predictions are based on the predicted mean of the regression (PM), inflated regression (inf), and the correct probabilistic interpretation of the regression (prob). As a reference forecast, the climatology has been used: that is, a stationary normal distribution.

	Median			95th percentile		
Corr	PM	Inf	Prob	PM	Inf	Prob
0.9	0.43	0.43	0.60	0.31	0.25	0.49
0.7	−0.02	−0.02	0.32	−0.01	−0.28	0.20
0.5	−0.32	−0.32	0.16	−0.05	−0.61	0.07

Some authors have proposed to use inflation for simulating local-scale variability (e.g., Karl et al. 1990; Bürger 2014, and references therein). A simulated time series is a "surrogate" of missing (e.g., future) observations with, ideally, the same statistical properties as the observations, conditional on the prediction ax. In the following, I will show that inflation does not correctly represent the desired statistical properties but misrepresents the temporal and spatial structure as well as trends. The true correlation between predictor and predictand is given analytically by $[a^2/(a^2 + \sigma^2)]^{1/2} < 1$ for $\sigma > 0$. For a correct representation of the predictand, the correlation between predictor and simulation thus has to be deliberately decreased to this value by adding random variability. For the inflated time series, however, the correlation between prediction and simulation is by construction identical to one. Thus, the temporal structure of the inflated time series is wrong: it is a purely rescaled version of the predictor structure but does not represent any unexplained local variability. The same argument holds also in space: if the same predictor is chosen to downscale to station series y_A and y_B, the resulting simulations will be perfectly correlated, which is wrong for any sensible meteorological setting. Finally, inflation misrepresents trends. Assume that the predictor x_i in Eq. (1) includes a linear trend that well describes a trend in the y_i, plus a stationary component: $x_i = \alpha_x t + x_{\mathrm{stat},i}$. The correct trend in y_i is then modeled as $a\alpha_x$. Inflation by the factor a_{inf} increases the predicted (or simulated) trend in y_i to $a_{\mathrm{inf}} a \alpha_x$, which is obviously wrong exactly by the inflation factor.

The key argument of Bürger (2014) relates to trends and applies to both quantile mapping and inflated regression. He argues that my criticism of inflated trends in quantile mapping goes along with the idea ''that long-term temporal features go hand in hand with large-scale spatial features.'' This reasoning is flawed. Bürger (2014) correctly states that trends can be spatially diverse at subgrid scale: for example, because of complex interactions with the topography.[3] Now one can think of two cases: First, trends might vary locally but might still be fully determined by the large-scale predictor. Here, the local trend variation would be captured by a locally varying regression parameter a in Eq. (1). In this case, randomization could in principle correctly represent local trends. Inflation (and quantile mapping), however, would create all the problems discussed above: in particular, it would artificially increase the correctly predicted trend. This is the situation I discussed in my original manuscript. Second, trends might vary locally but because of local interactions, which cannot be predicted from known large-scale predictors. Here, the trend predicted by any downscaling model would be wrong. Also, in this case, inflation would not improve the prediction. It would again create the problems discussed above, including the inflation of the already wrong trend.

For the example given by Bürger (2014) that slight changes in the general flow over complex terrain might strongly influence the local precipitation, this means the following: if one has established a relationship between variations in the general flow and local orography and can correctly project the general flow into the future, one might be able to capture future precipitation changes (as long as potential extrapolations are physically sound). Systematic local variations in trends can in principle be accounted for by the parameters of the regression model. If the relationship is too complex to be statistically modeled, if no predictor representing the general flow is available (e.g., because the dynamical model does not correctly simulate it), or if the trend is of subscale nature and not related to large-scale predictors, one will not be able to correctly capture future changes in precipitation. In any case, inflation will deteriorate the prediction of trends. It does not alter trends according to any physical reason but simply inflates them artificially according to the unexplained variance.

To summarize, inflated regression does not improve predictions. The apparent improvement of an increased hit rate for extreme events is a mere hedging effect that deteriorates major aspects of forecast quality. Furthermore, inflated regression is not suitable for simulating local-scale time series. It rather introduces systematic biases conditional on the predictor and does not correctly represent either the observed temporal and spatial structure or the trends. The statement that refuting the inflation effect on trends goes along with the assumption that local-scale trends should be the same as large-scale trends is wrong. In short, inflation is not a feasible approach. In the statistical sciences, it is not an accepted concept and there has also not been any controversy about it.

[3] Fig. 5 in Maraun (2013) is not necessarily an example for such a case, as claimed by Bürger (2014). The large-scale trend is from the model, and the local trend is observed.

REFERENCES

Bürger, G., 2014: Comment on ''Bias correction, quantile mapping, and downscaling: Revisiting the inflation issue.'' *J. Climate,* **27,** 1819–1820.

Davison, A. C., 2003: *Statistical Models.* Cambridge University Press, 726 pp.

Dobson, A. J., 2001: *An Introduction to Generalized Linear Models.* Chapman and Hall, 240 pp.

Friederichs, P., and T. L. Thorarinsdottir, 2012: Forecast verification for extreme value distributions with an application to probabilistic peak wind prediction. *Environmetrics,* **23,** 579–594.

Glahn, H., and R. A. Allen, 1966: A note concerning the ''inflation'' of regression forecasts. *J. Appl. Meteor.,* **5,** 124–126.

Jolliffe, I. T., and D. B. Stephenson, Eds., 2003: *Forecast Verification: A Practitioner's Guide in Atmospheric Science.* Wiley, 254 pp.

Karl, T. R., W.-C. Wang, M. E. Schlesinger, R. W. Knight, and D. Portman, 1990: A method of relating general circulation model simulated climate to the observed local climate. Part I: Seasonal statistics. *J. Climate,* **3,** 1053–1079.

Klein, W. H., B. M. Lewis, and I. Enger, 1959: Objective prediction of five-day mean temperatures during winter. *J. Meteor.,* **16,** 672–682.

Maraun, D., 2013: Bias correction, quantile mapping, and downscaling: Revisiting the inflation issue. *J. Climate,* **26,** 2137–2143.

——, and Coauthors, 2010: Precipitation downscaling under climate change. Recent developments to bridge the gap between dynamical models and the end user. *Rev. Geophys.,* **48,** RG3003, doi:10.1029/2009RG000314.

von Storch, H., 1999: On the use of ''inflation'' in statistical downscaling. *J. Climate,* **12,** 3505–3506.

Wilby, R. L., S. P. Charles, E. Zorita, B. Timbal, P. Whetton, and L. O. Mearns, 2004: Guidelines for use of climate scenarios developed from statistical downscaling methods. IPCC Task Group on Data and Scenario Support for Impact and Climate Analysis (TGICA) Rep., 27 pp.

Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences.* 2nd ed. Academic Press, 627 pp.

Yee, T. W., and C. J. Wild, 1996: Vector generalized additive models. *J. Roy. Stat. Soc.,* **58B,** 481–493.