On the Use of ''Inflation'' in Statistical Downscaling

HANS VON STORCH

Institute of Hydrophysics, GKSS Research Centre, Geesthacht, Germany

8 March 1999 and 22 June 1999

ABSTRACT

The technique of ''inflating'' in downscaling, which makes the downscaled climate variable have the right variance, is based on the assumption that all local variability can be traced back to large-scale variability. For practical situations this assumption is not valid, and inflation is an inappropriate technique. Instead, additive, randomized approaches should be adopted.

1. Statistical downscaling

The basic idea of statistical downscaling is to build an empirical model,

$$\mathbf{y} = \mathcal{F}(\mathbf{x}), \tag{1}$$

for a small-scale feature $\mathbf{y}$, not adequately described in GCMs, and large-scale features $\mathbf{x}$, well resolved (e.g., von Storch 1995). The technique has become popular in the past years. As predictands, $\mathbf{y}$ has been used as weather variables, such as daily temperatures, and climatic statistics such as intramonthly 95th percentiles of significant wave height (WASA Group 1998) or as monthly precipitation amounts. The predictor $\mathbf{x}$ has often been chosen as characteristics of the circulation, such as indices or principal components of air pressure fields, or of temperature, or combinations of both (Gyalistras et al. 1994).

The approach behind (1) is closely related to synoptic climatology, relating synoptic situations to local weather (Klein and Glahn 1974). Indeed, statistical downscaling is formally identical to ''perfect prog'' in weather forecasting. An important difference is that in downscaling the ''predictor'' is requested to be well simulated by climate models. From synoptic experience, it is well known (Starr 1942) that different $\mathbf{y}$'s are consistent with the same $\mathbf{x}$. That is, Eq. (1) must be understood as a stochastic equation. Feature $\mathbf{y}$ is a random vector, and $\mathcal{F}$ is a random function, conditioned by the realization $\mathbf{x}$ (von Storch 1999). The predictor $\mathbf{x}$ is also variable.

In many applications, the function $\mathcal{F}$ is simply linear. A prototype is linear regression. For the sake of simplicity we assume

$$\mathcal{F}(\mathbf{x}) = \alpha\mathbf{x} + \epsilon, \tag{2}$$

with ϵ drawn from a normal distribution with zero mean and standard deviation σ, and $0 < \alpha < 1$. The variations in ϵ are assumed to be independent from $\mathbf{x}$. In this setting, the randomness in $\mathbf{y}$ stems from the randomness in ϵ. In almost all applications, it is not the stochastic formulation (2) that is used but the deterministic version

$$\hat{\mathbf{y}} = \alpha\mathbf{x}. \tag{3}$$

Then, $\hat{\mathbf{y}}$ is an unbiased estimator of the conditional expectation of $\mathbf{y}$ given $\mathbf{x}$. In terms of the mean square error, it is an optimal estimator. Any specification different from (3) returns larger mean square errors.

However, the downscaled values $\hat{\mathbf{y}}$ have smaller variance than the local values $\mathbf{y}$. From $\text{var}(\hat{\mathbf{y}}) = \alpha^2\,\text{var}(\mathbf{x})$ and $\text{var}(\mathbf{y}) = \alpha^2\,\text{var}(\mathbf{x}) + \sigma^2$ follows $\text{var}(\hat{\mathbf{y}}) < \text{var}(\mathbf{y})$. That is meaningful, as the predictor $\mathbf{x}$ does not completely specify $\mathbf{y}$. For instance, as demonstrated by Roebber and Bosart (1998), two very similar synoptic situation produced markedly different precipitation distributions. Also, when a regional climate model is run with identical large-scale forcing but slightly different initial conditions, both runs deviate from each other for infinite times, even though the synoptic situations are very similar (e.g., Ji and Vernekar 1997; Rinke and Dethloff 1999).

2. Inflation and randomization

In many climate compact studies, time series $\tilde{\mathbf{y}}$ are needed with $\text{var}(\tilde{\mathbf{y}}) = \text{var}(\mathbf{y})$. To meet this requirement, it has been proposed to *inflate* $\hat{\mathbf{y}}$ by setting (Karl et al. 1990)

Corresponding author address: Dr. Hans von Storch, Institute of Hydrophysics, GKSS Research Centre, Max-Planck-Strasse 1, 21502 Geesthacht, Germany.
E-mail: storch@gkss.de

$$\tilde{\mathbf{y}} = \beta\hat{\mathbf{y}} = \sqrt{\frac{\mathrm{var}(\mathbf{y})}{\mathrm{var}(\hat{\mathbf{y}})}} \cdot \hat{\mathbf{y}}. \tag{4}$$

This technique has been used in the downscaling literature to some extent (e.g., Huth 1999). A technically more sophisticated albeit conceptually closely related canonical correspondence analysis (CCA)–based technique named ''expanded downscaling'' has been proposed (Bürger 1996) and tested (Dehn et al. 1999).

We claim that this approach is not meaningful. First, it makes the implicit assumption that all variability in $\mathbf{y}$ would be related to variability in $\mathbf{x}$, which is certainly not the case as outlined above.

Second, the mean square error of the inflated estimator is larger than that of the original estimator. To show this claim, we assume for the sake of simplicity that $\mathrm{var}(\mathbf{x}) = \mathrm{var}(\mathbf{y}) = 1$. Then $\mathrm{var}(\hat{\mathbf{y}}) = \alpha^2$ and $\beta = 1/\alpha$. For the mean square error we find $\mathrm{var}(\hat{\mathbf{y}} - \mathbf{y}) = \sigma^2$ and

$$\begin{aligned}\mathrm{var}(\tilde{\mathbf{y}} - \mathbf{y}) &= \mathrm{var}[(1 - \alpha)\mathbf{x} + \epsilon] \\ &= (1 - \alpha)^2 + \sigma^2 > \sigma^2.\end{aligned} \tag{5}$$

The smaller α, the larger β and the larger the increase of the error, as is intuitively to be expected.

An alternative to the inflation technique is to treat the ''unexplained'' part as ''existent, irregular, and unexplained''—namely, by adding noise. Thus, it is recommended to use a randomized downscaling specification,

$$\mathbf{y}^* = \hat{\mathbf{y}} + \mathrm{noise}, \tag{6}$$

thus satisfying the initial assumption of Eq. (1). There is no need for the noise to be white in time or space; more sophisticated formulations will in many practical situations be required. For instance, the variance and the auto correlation of the noise could depend on $\mathbf{x}$.

In important difference between ''inflation'' and ''randomization'' is that the space–time statistics of $\tilde{\mathbf{y}}$ are entirely controlled by the variations of $\mathbf{x}$, whereas the randomized $\mathbf{y}^*$ exhibits spatial and temporal variability only partially controlled by the large-scale state $\mathbf{x}$. Because of this feature, externally introduced signals, represented by $\mathbf{x}$, are overspecified by inflation.

The approach (6) has successfully been implemented in a study on landslides (Buma and Dehn 1999; Dehn and Buma 1999). However, a direct comparison of the two approaches has not yet been conducted.

REFERENCES

Buma, J., and M. Dehn, 1999: The impact of climate change on a landslide in south east France, simulated using different GCM-scenarios and downscaling methods for local precipitation. *Climate Res.,* in press.

Bürger, G., 1996: Expanded downscaling for generating local weather scenarios. *Climate Res.,* **7,** 111–128.

Dehn, M., and J. Buma, 1999: Modelling future landslide activity based on general circulation models. *Geomorphology,* in press.

——, G. Bürger, J. Buma, and P. Gasparetto, 1999: Impact of climate change on slope stability using expanded downscaling. *Eng. Geol.,* in press.

Gyalistras, D., H. von Storch, A. Fischlin, and M. Beniston, 1994: Linking GCM simulated climatic changes to ecosystems models. Case studies of statistical downscaling in the Alps. *Climate Res.,* **4,** 167–189.

Huth, R., 1999: Statistical downscaling in central Europe: Evaluation of methods and potential predictors. *Climate Res.,* in press.

Ji, Y., and A. D. Vernekar, 1997: Simulation of the Asian summer monsoons of 1987 and 1988 with a regional model nested in a global GCM. *J. Climate,* **10,** 1965–1979.

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 observed local climate. Part I: Seasonal statistics. *J. Climate,* **3,** 1053–1079.

Klein, W. H., and H. R. Glahn, 1974: Forecasting local weather by means of model output statistics. *Bull. Amer. Meteor. Soc.,* **55,** 1217–1227.

Rinke, A., and K. Dethloff, 1999: The influence of initial and boundary conditions on the climate of the Arctic in a regional climate model. *Climate Res.,* in press.

Roebber, P. J., and L. F. Bosart, 1998: The sensitivity of precipitation to circulation details. Part I: An analysis of regional analogs. *Mon. Wea. Rev.,* **126,** 437–455.

Starr, V. P., 1942: *Basic Principles of Weather Forecasting.* Harper Brothers Publishers, 299 pp.

von Storch, H., 1995: Inconsistencies at the interface of climate impact studies and global climate research. *Meteor. Z.,* **4,** 72–80.

——, 1999: The global and regional climate system. *Anthropogenic Climate Change,* H. von Storch and G. Flöser, Eds., Springer Verlag, 3–36.

WASA Group, 1998: Changing waves and storms in the Northeast Atlantic? *Bull. Amer. Meteor. Soc.,* **79,** 741–760.