Deterministic and Stochastic Precipitation Downscaling using Multi-Objective Genetic Programming

Tanja Zerenner
Meteorological Institute, University of Bonn
Bonn, Germany
tanjaz@uni-bonn.de

Victor Venema
Meteorological Institute, University of Bonn
Bonn, Germany
victor.venema@uni-bonn.de

Petra Friederichs
Meteorological Institute, University of Bonn
Bonn, Germany
pfried@uni-bonn.de

Clemens Simmer
Meteorological Institute, University of Bonn
Bonn, Germany
csimmer@uni-bonn.de

ABSTRACT

Symbolic regression is used to estimate daily time series of local station precipitation amounts from global climate model output with a coarse spatial resolution. Local precipitation is of high importance in climate impact studies. Standard regression, minimizing the RMSE or a similar point-wise error, by design underestimates temporal variability. For impact studies realistic variability is crucial. We use multi-objective Genetic Programming to evolve both deterministic and stochastic regression models that simultaneously optimize RMSE and temporal variability. Results are compared with standard methods based on generalized linear models.

CCS CONCEPTS

• **Computing methodologies → Genetic programming**; • **Applied computing → Earth and atmospheric sciences**; *Mathematics and statistics*;

KEYWORDS

geosciences, atmospheric sciences, climate, symbolic regression, variability, noise, time series

ACM Reference Format:

Tanja Zerenner, Victor Venema, Petra Friederichs, and Clemens Simmer. 2018. Deterministic and Stochastic Precipitation Downscaling using Multi-Objective Genetic Programming. In *GECCO '18 Companion: Genetic and Evolutionary Computation Conference Companion, July 15–19, 2018, Kyoto, Japan.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3205651.3208778

1 BACKGROUND AND MOTIVATION

Empirical-statistical downscaling relates local variables such as precipitation to the larger-scale atmospheric conditions (provided by a global or regional climate model) via a stochastic or deterministic function $\mathcal{F}$,

$$\textit{local climate response} = \mathcal{F}(\textit{larger-scale forcing})$$

This shall not imply that local climate is fully determined by the large-scale atmospheric state, but it may be treated as a stochastic process conditioned on the larger-scale climate [6]. Downscaling of precipitation is particularly challenging due to its strong spatial and temporal variability over a wide range of scales, and its non-Gaussianity.

Standard regression estimates the conditional expectation E of a local variable Y, in the following precipitation, given the larger-scale atmospheric state $\mathbf{x}$, i.e., $Y_x = E(Y|\mathbf{x}) + \epsilon$, with the residual ϵ denoting the component of Y that is not directly determined by $\mathbf{x}$. $E(Y|\mathbf{x})$ as an estimate of Y thus by design underpredicts variance. Impact models, however, typically require local climate information with realistic variability. Appropriate techniques are needed to model the variability contained in ϵ. The validity purely deterministic models to describe this variability is still a topic of discussion [2, 5].

2 METHOD

We follow the downscaling intercomparison experiment 1(a) designed by the COST Action VALUE [3]. Daily accumulated precipitation at European weather stations is estimated from coarse resolution ERA-Interim reanalysis data [1]. The experiment is set up as a 5-fold cross-validation (24 years of data are used for training, 6 years for validation).

We use multi-objective Genetic Programming (MOGP) to evolve symbolic regression models of both deterministic and stochastic nature which optimize a trade-off between a low RMSE and realistic temporal variability. The different versions of MOGP used are summarized in Table 1. Temporal variability is quantified by the integrated quadratic distance (IQD) between the empirical cumulative distribution functions of the downscaled and observed precipitation series. The objective IQD is calculated from the empirical distributions of the full time series. To obtain IQD_{sub} the

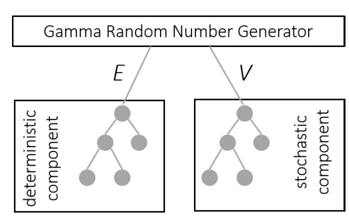

Figure 1: Grammar for MOGP setups S and S_{sub}.

GECCO '18 Companion, July 15–19, 2018, Kyoto, Japan

ACM ISBN 978-1-4503-5764-7/18/07.
https://doi.org/10.1145/3205651.3208778

Table 1: MOGP versions and reference methods.

a. MOGP versions

name	method	objective
D	deterministic	RMSE, IQD
D_{sub}	deterministic	RMSE, IQD_{sub}
S	stochastic	RMSE, IQD, RMSE-det
S_{sub}	stochastic	RMSE, IQD_{sub}, RMSE-det

b. reference methods

gglm	gamma generalized linear model
D_{ref}	gglm with deterministic variability
S_{ref}	gglm with stochastic variability

full, multi-year time series is split into several three month long subseries, then the IQDs are calculated separately for the subseries and averaged. Thus in contrast to IQD, IQD_{sub} takes into account that precipitation amounts and temporal variability typically vary with season.

In the *deterministic* versions all predictand variability is deterministically derived from the predictors. In the *stochastic* versions a simple grammar is employed to fix the root node of the parse trees to a conditional random number generator with arity 2 that draws numbers from gamma distributions with expected value E and variance V (Fig. 1). For the stochastic setups the RMSE of the result of the deterministic subtree (RMSE-det) serves as an additional objective. RMSE-det is small when the major part of the ϵ variability is not generated by the deterministic, but by the stochastic component of the regression model.

The multi-objective approach for downscaling was introduced in [7]. The MOGP code used is based on SPEA [8] and the GPLAB [4].

3 RESULTS

For the evaluation of the estimated time series we calculate the relative reduction of RMSE and IQD with respect to the raw precipitation series from the closest reanalysis grid box via

$$rel.reduction\ of\ RMSE = 1 - RMSE(dsc,obs)/RMSE(raw,obs),$$

where *dsc* denotes the downscaled series, *obs* the observed series, and *raw* the grid box precipitation from the reanalysis (which corresponds to no downscaling). A reduction of 1 is the optimum.

For the majority of stations each downscaling model from the reference methods is dominated by one or more from the MOGP downscaling models concerning the optimized objectives. MOGP is most beneficial for stations with strong precipitation and distinct variability (not shown).

The performance of a downscaling model should not significantly differ between training and validation. Concerning this criterion we find considerable differences between the methods for some cases (stations/cross-validation periods), e.g., for Salzburg Austria. To compare with the reference methods D_{ref} and S_{ref}, we select those MOGP models from the Pareto sets with the best IQD for the training period. These models offer the best representation of variability found while keeping the RMSE as low as possible. When comparing the difference in IQD between training and validation for these MOGP models as well as the reference methods, we find that only S_{sub} produces potentially useful models. The combination

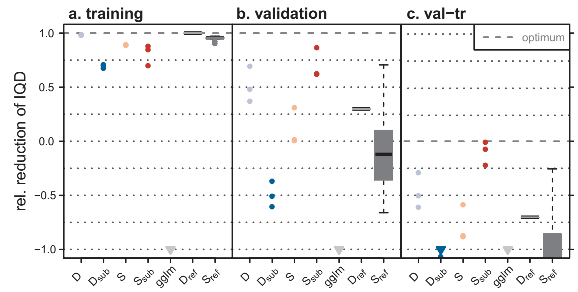

Figure 2: Best three MOGP models from Pareto sets w.r.t. IQD for the training period and 10 realizations[1] of each reference method (for exemplary station Salzburg Austria, training period 1979-2003, validation period 2003-2008). A triangle indicates values below plotted range.

of the stochastic technique and with the IQD_{sub} objective appears to be beneficial for some cases.

4 CONCLUSION AND OUTLOOK

We cannot conclude that a stochastic representations of variability is in general superior to the deterministic models, but we have seen that the choice of an appropriate grammar and an appropriate set of objectives is crucial when evolving downscaling models with MOGP. Evaluation of results for up to 86 stations all over Europe will allow a statistical analysis of the performance aiming to (1) find the best possible MOGP setup and to (2) understand under which conditions certain approaches work/fail.

ACKNOWLEDGMENTS

This work has been carried out within the Transregional Collaborative Research Center 32 funded by the German Research Foundation. The authors further acknowledge the COST Action VALUE.

REFERENCES

[1] D. P. Dee, S. M. Uppala, A. J. Simmons, et al. 2011. The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society* 137, 656 (2011), 553–597.

[2] D. Maraun. 2013. Bias correction, quantile mapping, and downscaling: Revisiting the inflation issue. *Journal of Climate* 26, 6 (2013), 2137–2143.

[3] D. Maraun, M. Widmann, J. M. Gutiérrez, S. Kotlarski, R. E. Chandler, E. Hertig, J. Wibig, R. Huth, and R. A. I. Wilcke. 2015. VALUE: A framework to validate downscaling approaches for climate change studies. *Earth's Future* 3, 1 (2015), 1–14.

[4] S. Silva and J. Almeida. 2003. GPLAB-a genetic programming toolbox for MATLAB. In *Proceedings of the Nordic MATLAB conference*. Citeseer, 273–278.

[5] H. von Storch. 1999. On the use of inflation in statistical downscaling. *Journal of Climate* 12, 12 (1999), 3505–3506.

[6] H. von Storch, B. Hewitson, and L. Mearns. 2000. *Review of empirical downscaling techniques*. Technical Report. Regional climate development under global warming.

[7] T. Zerenner, V. Venema, P. Friederichs, and C. Simmer. 2016. Downscaling near-surface atmospheric fields with multi-objective Genetic Programing. *Environmental Modeling and Software* 84 (2016), 85–98.

[8] E. Zitzler and L. Thiele. 1999. Multiobjective evolutionary algorithms: A comparative case study and the strength pareto approach. *IEEE transactions on Evolutionary Computation* 3, 4 (1999), 257–271.

[1] For the reference methods precipitation occurrence is modeled separately from precipitation amount with a stochastic approach. Therefore we can determine more than one realization for gglm and D_{ref}.