Environmental Modelling & Software 84 (2016) 85–98

Contents lists available at ScienceDirect

Environmental Modelling & Software

journal homepage: www.elsevier.com/locate/envsoft

Downscaling near-surface atmospheric fields with multi-objective Genetic Programming

Tanja Zerenner*, Victor Venema, Petra Friederichs, Clemens Simmer

Meteorological Institute, University of Bonn, Auf dem Huegel 16, 53121 Bonn and Meckenheimer Allee 176, 53115 Bonn, Germany

ARTICLE INFO

Article history:
Received 17 December 2015
Received in revised form
27 April 2016
Accepted 11 June 2016
Available online 5 July 2016

Keywords:
Statistical downscaling
Disaggregation
Evolutionary computation
Machine learning
Pareto optimality
Coupled modeling

ABSTRACT

We present a new Genetic Programming based method to derive downscaling rules (i.e., functions or short programs) generating realistic high-resolution fields of atmospheric state variables near the surface given coarser-scale atmospheric information and high-resolution information on land surface properties. Such downscaling rules can be applied in coupled subsurface-land surface-atmosphere simulations or to generate high-resolution atmospheric input data for offline applications of land surface and subsurface models. Multiple features of the high-resolution fields, such as the spatial distribution of subgrid-scale variance, serve as objectives. The downscaling rules take an interpretable form and contain on average about 5 mathematical operations. The method is applied to downscale 10 m-temperature fields from 2.8 km to 400 m grid resolution. A large part of the spatial variability is reproduced, also in stable nighttime situations, which generate very heterogeneous near-surface temperature fields in regions with distinct topography.

1. Introduction

With growing computational power integrated modeling platforms coupling atmospheric, land surface and subsurface models are increasingly used to account for interactions and feedbacks between the different components (e.g., Shrestha et al., 2014). The feedback processes are largely driven by the turbulent exchange fluxes of energy, moisture and momentum at the interface between land surface and atmosphere. The use of spatially averaged parameters or state variables at the land surface or the lower atmospheric boundary layer (ABL) can introduce biases in the flux estimation. In current atmospheric models for numerical weather prediction, which are typically applied at scales of few kilometers, heterogeneities at smaller scales are mostly neglected.

Subgrid-scale parameterization of the land surface like tile, mosaic or mixture approaches significantly improve the estimation of the surface fluxes (e.g., Avissar and Pielke, 1989; Koster and Suarez, 1992; Leung and Ghan, 1995; Schlünzen and Katzfey, 2003). Shao et al. (2001) showed that also the representation of the subgrid-scale atmospheric heterogeneity improves the flux estimates.

The explicit subgrid approach by Seth et al. (1994) allows to combine the subgrid representation of the land surface with downscaled atmospheric forcings. In the explicit approach each atmospheric model grid box covers $N \times N$ land surface columns, i.e., a higher resolution land surface scheme is nested into a coarser resolution atmospheric model (see also Giorgi et al. (2003) and Ament and Simmer (2006) for discussion). This is analogue to coupling a coarser atmospheric model with a high-resolution land surface model, as it is often done in the aforementioned integrated modeling platforms. This approach is feasible because of the comparatively low computational cost of land surface and subsurface models.

Besides the potential to improve the estimation of the turbulent exchange fluxes, downscaling of the near-surface atmospheric state variables can provide better forcing data for land surface, subsurface and agricultural models. This is important as besides the turbulent exchange coefficients also many processes at the earth's surface, e.g., related to vegetation, are nonlinear. Furthermore, the representation of runoff production or snow melt, which are threshold dependent, would benefit from taking subscale atmospheric variability into account.

Seth et al. (1994) introduced a simple atmospheric downscaling for the global climate scale (from 3.0° to 0.5° ≈ 50 km), which for instance corrects near-surface temperature using the model simulated ground temperature or topographic height at the high resolution. Fiddes and Gruber (2014) presented a more advanced

* Corresponding author.
E-mail addresses: tanjaz@uni-bonn.de (T. Zerenner), victor.venema@uni-bonn.de (V. Venema), pfried@uni-bonn.de (P. Friederichs), csimmer@uni-bonn.de (C. Simmer).

http://dx.doi.org/10.1016/j.envsoft.2016.06.009

physically based downscaling scheme, TopoSCALE, to create high-resolution forcing data for land surface models from global climate reanalysis (from 0.75° to ≤100 m), using fine-scale topography information from a high-resolution digital elevation model. A similar approach was taken by Schomburg et al. (2010, 2012), who developed a downscaling scheme at the mesoscale (from 2.8 km to 400 m) by statistically evaluating high-resolution atmospheric model runs. The scheme leads to improvements for certain variables (e.g., near-surface pressure) and weather conditions (e.g., near-surface temperature in unstable atmospheres). As processes in the lower ABL can be complex and highly nonlinear, the conditional linear regression approach used in Schomburg et al. (2010) appears not to be sufficient to capture many of the processes acting in the lower ABL.

In this study we introduce a more flexible approach to detect relations (downscaling rules) that generate high-resolution atmospheric fields from coarse atmospheric information and high-resolution information on land surface characteristics. We employ Genetic Programming (GP), a machine learning method from the area of evolutionary computation (e.g., Koza, 1992; Banzhaf et al., 1997). Like artificial neural networks GP allows to flexibly model complex nonlinear and multivariate relations with the advantage that the downscaling rules take the form of equations or program code, which is readable, and thus can be checked for physical consistency.

Coulibaly (2004), Liu et al. (2008) and Hashmi et al. (2011) employed GP based methods to downscale temperature and/or precipitation from global climate model output to a station or catchment mean. The results were compared to the Statistical Down-Scaling Model (SDSM) by Wilby et al. (2002). In all three studies the GP based methodologies performed better than the SDSM. In Coulibaly (2004) and Hashmi et al. (2011) the downscaling models resulting from GP and SDSM were explicitly compared showing that the GP model not only performed better, but also required less predictor variables. Liu et al. (2008) additionally compared the GP results against a feed forward neural net. Both methods performed about equally well.

Unlike previous studies which employ GP for atmospheric downscaling, we aim at the downscaling of coherent spatial fields. To this goal we employ a multi-objective approach, because a regression aiming solely at the minimization of the root mean square error (RMSE) is known to underpredict variance. The multi-objective approach allows to consider different characteristics of the fine-scale atmospheric fields, for instance spatially distributed variance, during the learning procedure.

This article introduces multi-objective Genetic Programming for the downscaling of atmospheric fields. As a first application we present the downscaling of near-surface temperature fields, which can exhibit very complex fine-scale patterns depending on atmospheric stability and thus offers a problem of sufficient complexity for testing the method. We build upon the same data set as used by Schomburg et al. (2010), which is introduced in Section 2. In Section 3 the methodology is explained in detail. Section 4 describes setup and results of downscaling 10 m-temperature, which are discussed in Section 5. Application to other atmospheric state variables, as well as the implementation of the downscaling scheme within a coupled modeling framework for the soil-vegetation-atmosphere system is part of ongoing work. Details on future plans are provided in Section 6.

2. Data

The downscaling rules are derived using the output of high-resolution simulations with the COSMO model (Baldauf et al., 2011) provided by Schomburg et al. (2010). The simulations have a grid spacing of 400 m and a time step of 4 s to satisfy the Courant-Friedrich-Levy stability criterion. The domain covers 168 km × 168 km centered over the Rur catchment in western Germany, which is the main investigation area of the Transregional Collaborative Research Centre 32 (TR32) on 'Patterns in Soil-Vegetation-Atmosphere-Systems' (Vereecken et al., 2010; Simmer et al., 2015), within which this study has been carried out. The data set contains hourly output for 8 simulation periods with a length of 1–2 days governed by different weather conditions (see Table 1). We consider only the inner 112 km × 112 km of the domain (i.e., 280 × 280 grid points) to exclude nesting effects. To reduce computational cost we extract single days and time steps to create our training data set. The scheme by Schomburg et al. (2010) has been initially developed for the downscaling from 2.8 km to 400 m grid resolution. In this study we consider the same scales, i.e., we aim at a downscaling by a factor of seven.

3. Methods

We downscale near-surface atmospheric fields by establishing a statistical relation (downscaling rule) between the coarse atmospheric model output and the high-resolution atmospheric fields using quasi-static high-resolution land surface information. Thus, we assume that the structure of the atmospheric boundary layer near the surface is significantly influenced by land surface heterogeneity.

A rule search algorithm based on Genetic Programming is set up, which can potentially detect multivariate and nonlinear downscaling rules. Such rules are much less complex than running the full 3D-model at high resolution. It is not expected that the downscaling rules reproduce the exact high-resolution references. Due to turbulence for instance, there will always be a remaining component of the fine-scale fields that cannot be reconstructed.

We take a multi-objective approach that allows multiple characteristics of the fine-scale fields to be incorporated during the fitting of the regression model. Minimizing only the root mean square error (RMSE) would result in downscaling rules predicting the expected value of the temperature anomalies given surface characteristics and coarse atmospheric state. Such an estimator is known to have too small variance (e.g., Hastie et al., 2009). Instead of aiming at predicting the expected value, we aim at downscaling rules returning realizations from an unknown multivariate probability density function (PDF). We do not optimize solely the RMSE, but also objectives that quantify the spatial variance on the subgrid-scale and the cumulative density functions (CDFs) of the full fields.

When we formulate the downscaling problem as a multi-objective optimization problem, we face, however, the following problems. Minimizing the sum of different objectives is problematic, since they may have different units and ranges. Even with an appropriate scaling procedure there is a risk of treating the objectives unequally or getting trapped in a local minimum. Firstly, we can never know, what is the minimum value of each objective that can be achieved by the regression. Thus, designing an appropriate scaling procedure is difficult and one would need to decide on the relative importance of the different objectives in advance. Secondly, adding multiple, conflicting objectives very likely results in a fitness function with multiple local minima, which makes optimization more difficult. To avoid these problems, we have implemented fitness calculation according to the Strength Pareto Evolutionary Algorithm (SPEA) by Zitzler and Thiele (1999), instead of using a single (weighted) fitness or cost function. Approaches for multi-objective optimization like SPEA are widely used in evolutionary computation. In SPEA the fitness calculation during the fitting procedure is based on an intercomparison of the different models.