BONNER METEOROLOGISCHE ABHANDLUNGEN

Heft 80 (2017) (ISSN 0006-7156)
Herausgeber: Andreas Hense

Tanja Zerenner

ATMOSPHERIC DOWNSCALING USING MULTI-OBJECTIVE GENETIC PROGRAMMING

BONNER METEOROLOGISCHE ABHANDLUNGEN

Heft 80 (2017) (ISSN 0006-7156)
Herausgeber: Andreas Hense

Tanja Zerenner

ATMOSPHERIC DOWNSCALING USING MULTI-OBJECTIVE GENETIC PROGRAMMING

Atmospheric Downscaling using Multi-Objective Genetic Programming

Dissertation
zur
Erlangung des Doktorgrades (Dr. rer. nat.)
der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von
Dipl.-Met. Tanja Zerenner
aus
Stuttgart

Bonn, Dezember, 2016

Diese Arbeit ist die ungekürzte Fassung einer der Mathematisch-Naturwissenschaftlichen Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn im Jahr 2016 vorgelegten Dissertation von Tanja Zerenner aus Stuttgart.

This paper is the unabridged version of a dissertation thesis submitted by Tanja Zerenner born in Stuttgart to the Faculty of Mathematical and Natural Sciences of the Rheinische Friedrich-Wilhelms-Universität Bonn in 2016.

Anschrift des Verfassers: Address of the author:

Tanja Zerenner
Meteorologisches Institut der
Universität Bonn
Auf dem Hügel 20
D-53121 Bonn

1. Gutachter: Prof. Dr. Clemens Simmer, Rheinische Friedrich-Wilhelms-Universität Bonn
2. Gutachter: PD Dr. Petra Friederichs, Rheinische Friedrich-Wilhelms-Universität Bonn

Tag der Promotion: 21. Juni 2017

Zusammenfassung

Numerische Modelle, welche für Wettervorhersagen und Klimaprojektionen verwendet werden, simulieren das Zusammenspiel physikalischer Prozesse in der Atmosphäre. Bedingt durch den hohen Rechenaufwand atmosphärischer Modelle treten jedoch häufig Diskrepanzen zwischen benötigter und verfügbarer Auflösung atmosphärischer Daten auf. Ein möglicher Ansatz, höher aufgelöste atmosphärische Daten aus vergleichsweise grobem Modelloutput zu generieren, ist statistisches Downscaling.

Die vorliegende Arbeit stellt multi-objektives Genetic Programming (MOGP) als Methode für das Downscaling atmosphärischer Daten vor. MOGP wird verwendet, um Downscaling Regeln (statistische Beziehungen) zu generieren, welche grobskalige atmosphärische Daten auf die Punktskala oder ein höher aufgelöstes Gitter abbilden. Im Gegensatz zu klassischen Regressionsansätzen, in welchen die Struktur des Regressionsmodells vorgegeben wird, entwickelt MOGP Modellstruktur und Modellparameter simultan. Dieses erlaubt es, auch nicht lineare und multivariate Beziehungen zwischen Prädiktoren und Prädiktand zu berücksichtigen. Ein klassisches lineares Regressionsmodel schätzt den Erwartungswert des Prädiktanden, eine Realisierung von Prädiktoren gegeben, und minimiert somit den mittleren quadratischen Fehler (root mean square error, RMSE), aber unterschätzt im Allgemeinen die Varianz. Mit einem multi-objektiven Ansatz können multiple Kostenfunktionen berücksichtigt werden, welche nicht ausschließlich auf die Minimierung des RMSE ausgelegt sind, sondern simultan auch Varianz und Wahrscheinlichkeitsverteilung berücksichtigen.

In dieser Arbeit werden zwei verschiedene Anwendungen von MOGP für atmosphärisches Downscaling präsentiert: Das Downscaling mesoskaliger oberflächennaher atmosphärischer Felder von einem 2.8 km auf ein 400 m Gitter und das Downscaling von Temperatur- und Niederschlagszeitreihen von globalen Reanalysedaten auf lokale Stationen.
(1) Mit wachsender Rechenleistung werden integrierte Modellplattformen, welche Atmosphärenmodelle mit Landoberflächenmodellen und hydrologischen Bodenmodellen koppeln, immer häufiger verwendet, um auch die Interaktionen und Feedbacks zwischen den Komponenten des Boden-Vegetations-Atmosphären Systems zu berücksichtigen. Aufgrund kleinskaliger Heterogenitäten in Landoberfläche und Boden benötigen die Landoberflächen- und Bodenmodelle eine hohe Gitterauflösung. Für atmosphärische Modelle hingegen ist eine solch hohe Auflösung rechnerisch nicht praktikabel. Daher findet sich typischerweise ein Skalenunter-

schied zwischen atmosphärischer und Landoberflächen-/hydrologischer Modellkomponente. Solch ein Skalensprung kann jedoch zu Problemen bei der Schätzung der turbulenten Flüsse zwischen Atmosphäre und Boden führen, da die turbulenten Flüsse in nichtlinearer Weise vom Zustand des Bodens und der bodennahen Atmosphäre abhängen. Die mit MOGP entwickelten Downscaling Regeln verwenden grob aufgelöste atmosphärische Daten und hoch aufgelöste Landoberflächen-Informationen, um hoch aufgelöste Felder verschiedener bodennaher atmosphärischer Variablen (Temperatur, Windgeschwindigkeit etc.) generieren. Die Regeln basieren somit auf der Annahme, dass die bodennahe atmosphärische Grenzschicht signifikant von der Heterogenität der Landoberfläche beeinflusst wird. Zwar erreicht MOGP für diese Anwendung nur selten eine signifikante Reduktion des RMSE gegenüber einer reinen Interpolation, jedoch kann, abhängig von der betrachteten atmosphärischen Variablen, ein großer Teil der räumlichen Variabilität wiederhergestellt werden ohne oder mit nur sehr geringem Anstieg des RMSE.
(2) Studien zur Auswirkung des Klimawandels benötigen oft hochaufgelöste oder lokale atmosphärische Daten. Der Output globaler Klimamodelle, mit Hilfe derer Klimaprojektionen erstellt werden, ist gemeinhin zu grob. MOGP wird verwendet, um Tagesmaximum, -minimum und -mittel der Temperatur sowie den täglich akkumulierten Niederschlag an lokalen Stationen in Europa zu schätzen. Die Resultate werden mit linearen Regressionsmethoden verglichen. Für das Downscaling von Temperatur liefert eine klassische lineare Regression bereits sehr gute Resultate, welche MOGP im Allgemeinen an Qualität übertreffen. Für Niederschlag hingegen sind die MOGP Resultate vielversprechend, auch im Vergleich zu generalisierten linearen Modellen. Insbesondere die Repräsentation von Niederschlagsextremen und räumlicher Korrelation (letzteres ist nicht Bestandteil der Kostenfunktionen) sind vielversprechend.

Abstract

Numerical models are used to simulate and to understand the interplay of physical processes in the atmosphere, and to generate weather predictions and climate projections. However, due to the high computational cost of atmospheric models, discrepancies between required and available spatial resolution of modeled atmospheric data occur frequently. One approach to generate higher-resolution atmospheric data from coarse atmospheric model output is statistical downscaling.

The present work introduces multi-objective Genetic Programming (MOGP) as a method for downscaling atmospheric data. MOGP is applied to evolve downscaling rules, i.e., statistical relations mapping coarse-scale atmospheric information to the point scale or to a higher-resolution grid. Unlike classical regression approaches, where the structure of the regression model has to be predefined, Genetic Programming evolves both model structure and model parameters simultaneously. Thus, MOGP can flexibly capture nonlinear and multivariate predictor-predictand relations. Classical linear regression predicts the expected value of the predictand given a realization of predictors minimizing the root mean square error (RMSE) but in general underestimating variance. With the multi-objective approach multiple cost/fitness functions can be considered which are not solely aimed at the minimization of the RMSE, but simultaneously consider variance and probability distribution based measures.

Two areas of application of MOGP for atmospheric downscaling are presented: The downscaling of mesoscale near-surface atmospheric fields from 2.8 km to 400 m grid spacing and the downscaling of temperature and precipitation series from a global reanalysis to a set of local stations.
(1) With growing computational power, integrated modeling platforms, coupling atmospheric models to land surface and hydrological/subsurface models are increasingly used to account for interactions and feedback processes between the different components of the soil-vegetation-atmosphere system. Due to the small-scale heterogeneity of land surface and subsurface, land surface and subsurface models require a small grid spacing, which is computationally unfeasible for atmospheric models. Hence, in many integrated modeling systems, a scale gap occurs between atmospheric model component and the land surface/subsurface components, which potentially introduces biases in the estimation of the turbulent exchange fluxes at the surface. Under the assumption that the near surface atmospheric boundary layer is significantly influenced by land surface heterogeneity, MOGP is used to evolve downscaling rules that recover

high-resolution near-surface fields of various atmospheric variables (temperature, wind speed, etc.) from coarser atmospheric data and high-resolution land surface information. For this application MOGP does not significantly reduce the RMSE compared to a pure interpolation. However, (depending on the state variable under consideration) large parts of the spatial variability can be restored without any or only a small increase in RMSE.
(2) Climate change impact studies often require local information while the general circulation models used to create climate projections provide output with a grid spacing in the order of approximately 100 km. MOGP is applied to estimate the local daily maximum, minimum and mean temperature and the daily accumulated precipitation at selected stations in Europe from global reanalysis data. Results are compared to standard regression approaches. While for temperature classical linear regression already achieves very good results and outperforms MOGP, the results of MOGP for precipitation downscaling are promising and outperform a standard generalized linear model. Especially the good representation of precipitation extremes and spatial correlation (with the latter not incorporated in the objectives) are encouraging.

Contents

1 Introduction

The dynamical processes in the atmosphere (and at the land surface and in the subsurface) act at intrinsic spatial and temporal scales. Turbulent eddies have a spatial extension of several centimeters to a few hundred meters and a life span between seconds and a few minutes; convective events occur over a wide range of scales from several meters to several kilometers with durations between minutes and hours; large scale oscillation patterns, such as the El Niño and La Niña phases of the El Niño Southern Oscillation ENSO typically persist over several months (cf. Fig. 1.1).

Numerical models are used to simulate the dynamical and physical processes in the atmosphere (and at the land or subsurface) in order to make predictions of future *weather and climate* conditions [1] and to improve understanding of the interplay of the processes involved. Atmospheric models typically rely on a set of prognostic hydro-thermodynamic partial differential equations and a set of diagnostic equations, which together describe the atmospheric state and its change in time. The set of equations is discretized using finite differences, finite elements etc. or via a spectral approach and solved by numerical integration schemes. Solving the equations is computationally expensive.

General Circulation Models (GCMs), which cover the whole globe, are therefore restricted to a relatively coarse grid spacing (typically in the order of 100 kilometers) by the available computational power. In climate modeling, regional climate models (RCMs) use a smaller grid spacing as these are nested into the GCMs and applied over smaller (regional) domains (e.g., continental Europe with a grid spacing of 7 km). The grid spacing can be decreased further by carrying out multiple nesting steps (e.g., up to 3 km for the Alpine region) (cf. Fig. 1.2).

Also in operational numerical weather prediction (NWP) such model chains are common. GCMs provide initial conditions (taken from an analysis run) and boundary conditions to limited area model NWP models with a smaller grid spacing typically for continental domains (e.g., COSMO-EU covering a European domain with a 7 km grid spacing). Even more local models with smaller grid spacing (e.g., COSMO-DE covering Germany with a grid spacing of 2.8 km) receive initial and boundary conditions from the continental scale models.

Atmospheric processes that are not resolved by the model resolution - note that for numer-

[1] *Weather* refers to the short term atmospheric state. *Climate* refers to long term (30 years or longer) statistics of (daily) weather.

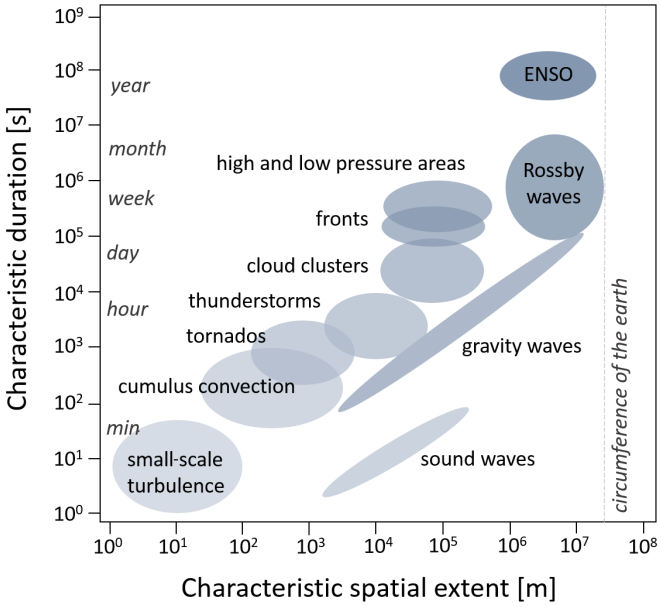

Figure 1.1.: Important atmospheric processes and their intrinsic spatial and temporal scales (following e.g., *Laing and Evans*, 2011).

ical reasons the effective resolution of a model corresponds to at least four grid boxes - are accounted for by parameterization schemes. GCMs incorporate cloud schemes, most RCM and NWP applications incorporate convection schemes, and even high-resolution simulations with grid spacings of a few hundred meters still require turbulence schemes. In large eddy simulations (LES) of the atmospheric boundary layer, which is feasible for very small domains only, the large eddies, which are responsible for the major part of the turbulent exchange of mass, energy and momentum, are resolved. However, the smaller eddies, typically in the inertial subrange, are parameterized even in LES simulations.

With the atmosphere being a *(deterministic) chaotic* system[2] its numerical modeling is inevitably subject to large uncertainties. Overall uncertainty results from uncertainties of initial and boundary conditions, and uncertainties induced by numerical discretization and parameterization schemes. Especially the latter are known to constitute a major source of uncertainty. GCMs often have difficulties with the representation of clouds and precipitation. In higher resolution models the representation of convection and turbulent exchange of energy, moisture and momentum are still challenging.

Modeled atmospheric data is of use to many communities. The value of modeled atmospheric data to other communities is limited by uncertainty (and potentially systematic errors) and by the available grid spacing. Hydrological models, land surface models, agricultural models etc. require atmospheric data often at high spatial resolutions as many hydrological and land surface modeling applications use a small grid spacing (100 m or less) to account for the heterogeneity of land surface and soil at small scales. Using coarser-scale atmospheric data as forcing for such simulations can induce biases for example in the estimation of the turbulent

[2] *Deterministic chaos* refers to systems for which any small change in the initial state can (after some time) lead to a completely different development of the system state.

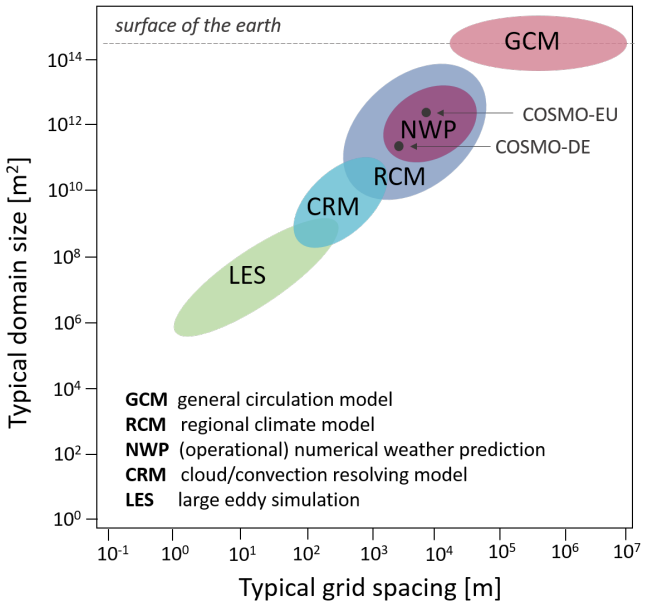

Figure 1.2.: Domain extent and grid spacing of common types of atmospheric models.

fluxes as the fluxes are described by nonlinear functions of the state of the land surface and the lowermost atmosphere. With increasing computational power integrated modeling platform coupling subsurface, land surface and atmospheric models are used more and more frequently. For computational reasons and to improve the simulation of the turbulent exchange fluxes such platforms often employ mosaic approaches (assigning several land and subsurface grid boxes to one atmospheric model column). An appropriate atmospheric downscaling might further improve the flux estimation as well as the simulation of threshold dependent processes such as snow melt or soil freezing. Moreover, climate change impact modeling (vegetation modules, crop modules, hydrological modules etc.) requires local (up to point scale) atmospheric information.

To match the requirements of the users, modeled atmospheric data is often processed using model output statistics, e.g., for bias correction, to reduce systematic errors. The discrepancy between required resolution and available resolution from the models (representativeness problem) is addressed by downscaling techniques[3]. In practice many downscaling techniques combine the correction of model errors and representativeness problem. For regionalized or local climate projections, a vast number of empirical-statistical downscaling techniques have emerged over the past decades.

The multi-objective Genetic Programming (MOGP) approach presented in this study differs from the majority of downscaling techniques in two ways: (1) Genetic Programming allows the implementation of symbolic regression. That is, model structure and model parameters are evolved simultaneously, such that nonlinear and multivariate predictor-predictand relations can be flexibly accounted for. For most present techniques (e.g., most regression approaches,

[3]The nesting of regional climate models into general circulation models is referred to as *dynamical downscaling*. Empirical-statistical approaches are referred to as *empirical-statistical downscaling*.

analog techniques) the model structure has to be predefined and only the model parameters are optimized. (2) Most existing techniques either aim at matching modeled and reference probability distributions (e.g., quantile mapping or stochastic weather generators) or aim at a pointwise match of the reference data (e.g., most regression techniques including neural networks). The multi-objective approach allows to incorporate multiple fitness/cost function when fitting the downscaling model. That is, distribution based measures, such as the integrated quadratic distance, can be considered together with measures comparing reference and prediction pointwise, such as the root mean square error.

The presented work has been carried out in the framework of the *Transregional Collaborative Research Center 32* on *Patterns in Soil-Vegetation-Atmosphere-Systems*. Initial motivation of the work is the development of an improved atmospheric downscaling scheme to be applied in fully coupled subsurface-land surface-atmosphere simulations with the *Terrestrial Systems Modeling Platform* (TerrSysMP) developed within TR32. During the work on this thesis the COST action VALUE on *Validating and Integrating Downscaling Methods for Climate Change Research* has set up a set of downscaling experiments for climate data aiming at a comprehensive intercomparison of existing downscaling techniques which has motivated an additional application of MOGP.

The Chapters 2-4 offer background information on the different aspects involved in this work. Chapter 2 introduces Genetic Programming and can be skipped by readers familiar to GP. Chapter 3 provides background information on the integrated modeling of the soil-vegetation-atmosphere system focusing on the representation of land-surface heterogeneity, existing atmospheric disaggregation approaches and introduces the TerrSysMP. Furthermore, the atmospheric component model of TerrSysMP, COSMO, which has been used for this study, is described in more detail. Downscaling approaches, mainly designed for downscaling of climate data, are reviewed in Chapter 4. The MOGP downscaling approach is introduced in Chapter 5. The detailed setup and the results of MOGP for downscaling near-surface atmospheric fields (from 2.8 km to 400 m grid spacing) are described in Chapter 6. For this MOGP application high-resolution modeled data serves as reference. Thus, the downscaling aims to account purely for the representativeness problem ignoring potential model errors. In Chapter 7 we apply MOGP to the first experiment set up by COST-VALUE, which considers the downscaling of climate data time series from GCM scale to point scale. For this application observation data serves as reference. Hence, the downscaling accounts for both, the representativeness problem and potential model errors, simultaneously without distinguishing between the two. The major results of this thesis and their implications are summarized and discussed in the conclusion in Chapter 8. Parts of Chapter 5, 6 and Appendix A are published in *Simmer et al.* (2015) and *Zerenner et al.* (2016).

2 Genetic Programming

Genetic Programming addresses the problem of automatic programming, namely, the problem of how to enable a computer to do useful things without instructing it, step by step, how to do it (John R. Koza in *Banzhaf et al.*, 1997).

Genetic Programming (GP) automatically creates program code to solve user defined tasks requiring only a minimum of information by the user. In particular the user is not required to prescribe the size and shape of the solution. GP belongs to the evolutionary computation techniques, which are based on the Darwinian concept of *survival of the fittest*.
For a given problem a generation of initial solutions is created randomly (or incorporating prior knowledge). Each of these candidate solutions is applied to the problem and evaluated. The solutions from the existing generation are then modified to form a new generation (cf. Fig. 2.1). The better a candidate solution performs, the more likely it contributes to the successive generation. The evolution is stopped and the best solution found returned when a certain number of generations defined by the user is reached (or when a solution is found that performs sufficiently well).
Evolutionary computation contains not only GP, but also many related techniques such as genetic algorithms (GA) or gene expression programming (GEP). In this work we stick to the classical (tree-based) Genetic Programming. The content of this chapter largely relies on the textbooks by *Banzhaf et al.* (1997), *Mitchell* (1998), *Poli et al.* (2008), *Affenzeller et al.* (2009) and the pioneering work by *Koza* (1992). After introducing the fundamental elements

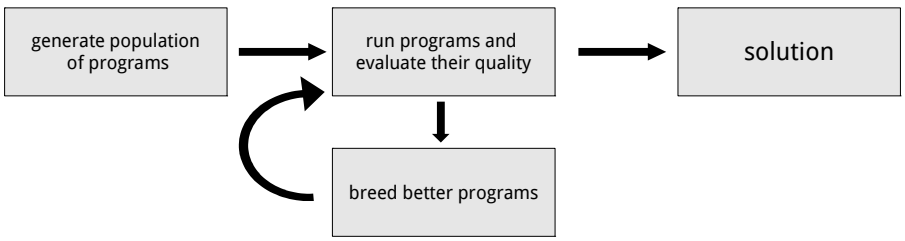

Figure 2.1.: Concept of Evolutionary Computation: Given a problem to solve, a population of potential solutions is frequently tested and updated until the termination criterion is met (Figure adapted from *Poli et al.*, 2008).

of a GP system and the associated terminology (Sec. 2.1), we walk through a standard GP algorithm step by step (Sec. 2.2). Finally some real world applications of GP are presented (Sec. 2.3).

2.1. Terminology and Definitions

Parse tree In classical GP the solutions (individuals) are represented by parse trees, which consist of functions and terminals. Each element of a parse tree is also referred to as a node. Figure 2.2 shows an example of a parse tree representing a simple equation. The tree consists of 9 nodes arranged on 4 levels. Parse trees are read bottom to top. That is, the parse tree in Figure 2.2 is evaluated as follows:

(1) 4 is multiplied with c,

(2) 2 is divided by 3,

(3) the result of (2) is multiplied with b,

(4) the results of (1) and (3) are subtracted.

Hence, the parse tree represents the equation $4c - \frac{2}{3}b$. This is a simple example of a parse tree containing only arithmetic functions (*minus*, *times*, *divide*), variables (c, b) and constants (2, 3, 4). Dependent on the problem to solve, parse trees can be much larger and much more complex.

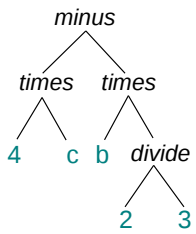

Figure 2.2.: Example of a parse tree representing the simple equation $4c - \frac{2}{3}b$.

Terminal set The terminal set $\mathcal{T}$ provides the basic input to the parse trees. Thus, terminals terminate the branches of the tree. The terminal set can contain numerical constants, variables (for instance in regression problems) and any kind of zero-argument functions. In the example parse tree shown in Figure 2.2 the terminals used are the numerical constants 2, 3 and 4, and the variables b and c.

Function set The function set $\mathcal{F}$ contains all of the functions and statements available to the GP system. The type of functions can be very diverse. Examples of possible functions and statements are:

- **Arithmetic functions**, i.e., plus, minus, multiply, divide,
- **Transcendental functions**, e.g., trigonometric functions and logarithmic functions,

- **Boolean functions**, i.e., AND, OR, NAND, NOR,
- **Conditional statements**, e.g., IF ... THEN ... ELSE ...,
- **Loop statements**, e.g., FOR ... DO ...,
- **Subroutines**.

In the example parse tree shown in Figure 2.2 the function used are subtraction (*minus*), multiplication (*times*) and division (*divide*).

Arity The arity of a function is the number of input arguments. All elements of the terminal set have arity zero. The elements of the function set have an arity of at least one.

Genetic Operators During the evolutionary process, again and again new populations of candidate solutions are created from the already existing ones by applying genetic operators. The three standard genetic operators commonly used in tree-based GP are:

(1) **Reproduction** The reproduction operator is the most straightforward. An individual is selected from the current generation, copied and inserted into the new population (cf. Fig. 2.3).

(2) **Crossover** The crossover operator combines two individuals (parents). At first two individuals from the current generation are selected to serve as parents. From each parent a subtree is chosen randomly. The subtrees are exchanged (cf. Fig. 2.3). Crossover transforms two existing individuals into two new individuals.

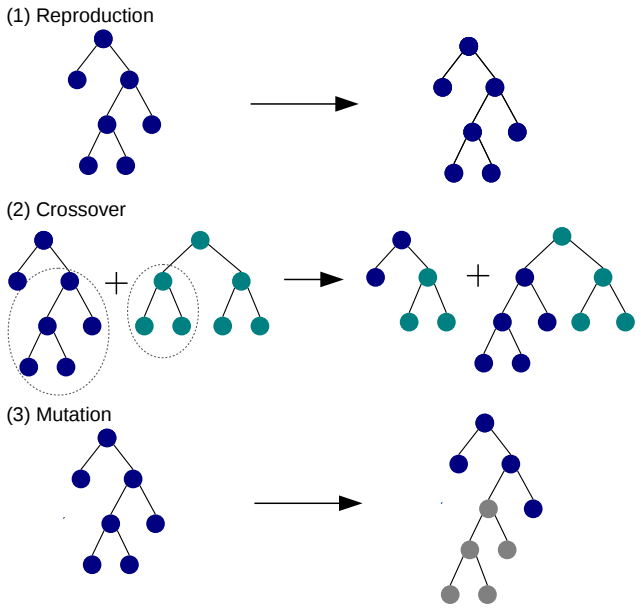

Figure 2.3.: The three common genetic operators used in standard (tree-based) Genetic Programming.

(3) **Mutation** Mutation operates on one individual only. There are several variants of the mutation operator. In our algorithm we use the standard subtree-mutation. An individual from the current generation is selected. From this individual a randomly chosen subtree is cut off and replaced by a new randomly generated subtree (cf. Fig. 2.3). The mutation operator allows new program sequences to enter the evolutionary process.

Fitness function The fitness or fitness function is the measure used in GP to quantify how well a candidate solution solves the problem given and is used to evaluate the candidate solutions. During the evolutionary process the fitness also determines how likely an individual is selected to serve as parent in the creation of a new generation. The most simple way is to set the selection probability of the individuals proportional to their fitness.

2.2. Preparing and Running GP

Preparation Figure 2.4 illustrates input and output of a standard GP system. The user provides function and terminal sets. An appropriate selection of functions and terminals is of great importance to successfully employ GP to solve a given problem. Further, a sufficient set of training data has to be supplied. The outcome is evaluated by the fitness function. Like the definition of functions and terminals, also an appropriate fitness function is essential to a successful GP setup. The user can specify some additional run parameters, such as the size of each population, maximum number of levels or nodes for the parse trees or the selection properties of reproduction, crossover and mutation operators. Finally, a termination criterion has to be provided that defines when to stop the GP run. The termination criterion can be reaching a certain fitness value or a certain number of generations.

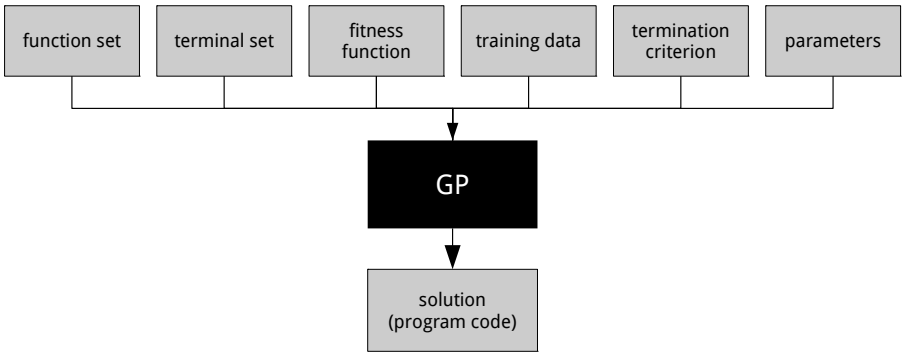

Figure 2.4.: Illustration of input and output to Genetic Programming. What happens inside the *black box* is shown in Fig. 2.5 and explained in the text.

Execution Figure 2.5 shows the typical structure of a GP system. The number of generations run so far is saved as the variable gen. The termination criterion is indicated by gen_{max}, the maximum number of generations to run. The number of individuals in the current generation is saved as the variable ind. The number of individuals in each generation is ind_{gen}.

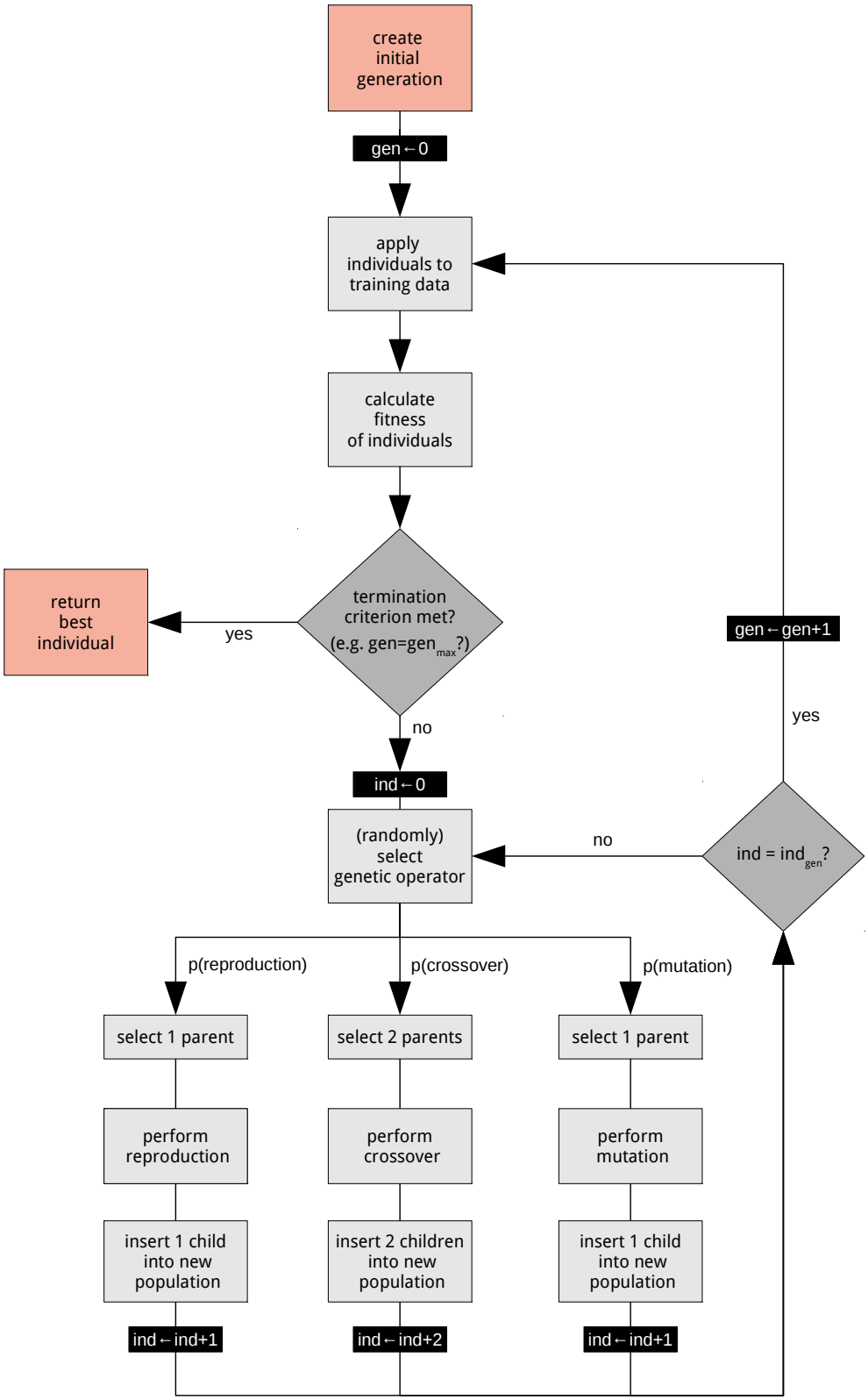

Figure 2.5.: Flowchart of the main steps of Genetic Programming. A detailed description can be found in the text.

(1) First an initial population (generation 0) of candidate programs is generated. The variable *gen* is set to zero, i.e., *gen* ←0. The initial generation can be created either randomly or include already known approximate solutions of the given problem.

(2) The individuals are applied to the training data provided by the user.

(3) The outcome of each individual is evaluated according to the fitness function.

(4) If the termination criterion is met (e.g., if the maximum number of generations is reached or if a satisfactory solution is found) the algorithm stops and and the best individual is returned. If the termination criterion is not met, the algorithm continues with the creation of a new, hitherto empty (ind←0), generation:

   (a) A genetic operator is randomly selected based on the selection property p.

   (b) Depending on the operator selected one or two parents are drawn from the current generation.

   (c) The genetic operator is applied to the parent(s).

   (d) The offspring is inserted into the new population. That means the number of individuals in the new population is increased by one (ind←ind+1) for mutation or reproduction, or by two (ind←ind+2) for crossover.

   (e) Starting from (a) this sequence is repeated until enough individuals for the new generation have been created (i.e., until ind=ind$_{gen}$).

(5) After the new generation is created the algorithm continues at (2).

2.3. Real World Applications

Over the years GP has been applied to a variety of real world problems. Based on thousands of GP applications over the last decades, some criteria have emerged indicating if GP is likely to be a suitable method for a problem (e.g., *Poli et al.*, 2008). GP is likely to perform well

- when the relation between the relevant variables is either unknown or at least not fully understood,
- when finding the shape and size of the solution is a mayor part of the problem,
- when sufficient amounts of training/test data are available,
- when there are good possibilities to test the performance of a candidate solution, but poor chances to directly obtain a sufficient solution,
- when conventional mathematics do not (or can not) provide an analytic solution,
- when an approximate solution is acceptable (or the only solution that is ever likely to be obtained),
- when small improvements in performance may have large impacts for applications.

Symbolic Regression Many real world applications of GP are symbolic regression problems. Symbolic regression refers to the fitting of observed data with the structure of the regression model unknown. Symbolic regression is often used for data where the underlying process is not known or not yet understood well enough to describe it in terms of a mathematical formula. Symbolic regression has been one of the earliest applications of GP (*Koza*, 1992). Common fitness functions of symbolic regression problems are the mean error or the root mean square error between the output of the GP solutions and the desired outcomes as contained in the training data set.
Regression problems occur in almost any scientific area (and not only there). Also the detection of downscaling rules considered in the later chapters of this work ranks among the symbolic regression problems. Here, mappings are established which predict the high-resolution data from coarser-resolution information using observed or modeled data at high-resolution for training.

Image and Signal Processing In the area of image and signal processing GP has been used, for instance, to visually classify objects (e.g., *Smart and Zhang*, 2003), for content based image-retrieval (e.g., *Torres et al.*, 2009) or to detect certain image features (e.g., *Tackett*, 1993).

Compression and Data Mining So called programmatic compression has been already considered in *Koza* (1992). Programmatic compression treats an image as a function of row and column index of each pixel. Such functions can be derived using GP and serve as a lossy[1] compressed version of an image. The technique of programmatic compression has been further studied and applied to both images and sounds in *Nordin and Banzhaf* (1996). Lossless[2] image compression using GP has been first considered in *Fukunaga and Stechert* (1998) who evolve non-linear models predicting the value of a pixel from a subset of neighboring values. *Kattan and Poli* (2008) proposed a lossless data compression in which GP combines well known compression algorithm such that optimal reduction of the file length is achieved.

Bioinformatics and Medicine A large number of studies considers classification and data mining for large biomedical data sets, such as gene microarray data, by means of GP (e.g., *Hong and Cho*, 2006; *Yu et al.*, 2007).

Economic Modeling In the economic sector GP has been, for instance, employed for evolving trading rules (e.g., *Yu et al.*, 2005) or to predict stock indices (e.g., *Chen et al.*, 1999).

Geosciences Compared to other areas, such as bioinformatics or economics, GP has been rarely applied to geoscientific tasks. *Parasuraman et al.* (2007) and *Kim and Kim* (2008), for

[1] In *lossy* file compression certain information is permanently eliminated (especially redundant information). When the file is uncompressed, only a part of the original information is available. Applied for instance to a graphic lossy compression typically reduces the resolution the graphic.
[2] In *lossless* file compression no information from the original file is lost. When the file is uncompressed, the original file is fully restored.

instance, use GP (or genetic algorithms) for evapotranspiration modeling. *Wang* (1991) employ genetic algorithms for calibrating conceptual run-off models. The few studies employing GP to the downscaling of general circulation model output (*Coulibaly*, 2004; *Liu et al.*, 2008; *Hashmi et al.*, 2011) are reviewed in detail in Section 4.4.

3

Integrated Modeling of the Soil-Vegetation-Atmosphere System

The Transregional Collaborative Research Center 32 studies *Patterns in Soil-Vegetation-Atmosphere Systems* using monitoring, modeling and data assimilation (*Vereecken et al.*, 2010; *Simmer et al.*, 2015). Processes in soil, vegetation and atmosphere act over a large range of spatial and temporal scales. The land surface is strongly heterogeneous with respect to topography and texture with especially the latter being strongly affected by human usage. Typically atmospheric models are computationally significantly more expensive than land surface and subsurface models and therefore applied with comparatively coarse grid spacings. Effects of subscale land surface heterogeneity on the turbulent exchange between land surface and atmosphere can be partly accounted for by parameterization schemes. An overview of the treatment of land surface heterogeneity in atmospheric modeling is given in Section 3.1.
Not only the land surface is strongly heterogeneous but also the lower atmospheric boundary layer, which is significantly influenced by land surface heterogeneity. Atmospheric disaggregation or subgrid-scale parameterizations aim to account for the spatial variability in the lowermost atmosphere with the ultimate goal of scale-consistent coupling between atmospheric models and surface (and subsurface) schemes or models. In Section 3.2 existing approaches are reviewed with a special focus on the 3-step downscaling scheme by *Schomburg et al.* (2010), which has been developed in the first phase of TR32. The 3-step scheme has been designed for downscaling mesoscale near-surface atmospheric fields from 2.8 km to 400 m grid spacing. In the central step of the scheme, which aims to reconstruct the fine-scale structures of the fields, a linear regression utilizing high-resolution land surface information is applied. The 3-step scheme performs well for a subset of the atmospheric state variables required by land surface and subsurface models and/or under certain weather conditions. By applying MOGP to the downscaling of atmospheric near-surface fields at the mesoscale, we aim to account also for complex and nonlinear processes in the lower atmospheric boundary layer, which cannot be captured by a simple linear regression (cf. Chap. 6).
During the second phase of the TR32 the integrated *Terrestrial Systems Modeling Platform* (TerrSysMP) has been set up by *Shrestha et al.* (2014). TerrSysMP offers a highly modular platform coupling the atmospheric model COSMO, the land surface model CLM and the subsurface model ParFlow via an external coupling interface. In Section 3.3 TerrSysMP and its

components are briefly introduced. In the current setup TerrSysMP optionally includes the 3-step downscaling algorithm by *Schomburg et al.* (2010).
To improve the downscaling algorithm a new and larger reference data set has been created by carrying out high-resolution simulations (400 m grid spacing) with the most recent version of the COSMO model. The COSMO model is thus reviewed in more detail in Section 3.4 focusing on the COSMO-DE configuration, which has been adapted for the high-resolution simulations (cf. Sec. 6.1.1).

3.1. Land Surface Heterogeneity in Earth System Modeling

The land surface is an important constituent of the earth system as it represents the interface between atmosphere, biosphere and subsurface hydrology. The land surface exchanges momentum, energy, and water and other constituents, such as CO_2, with the atmosphere and thus impacts weather and climate. In atmospheric modeling so-called soil-vegetation-atmosphere transfer schemes (SVATs) are used to capture these interactions. Typical SVATs are composed of sub-models (soil modules, vegetation modules, snow modules, land surface hydrology modules) interacting with the atmosphere and with each other. SVATs calculate the surface-atmosphere exchange fluxes of momentum, energy (radiation, sensible heat, latent heat), moisture and other constituents, such as CO_2, as lower boundary condition to the atmospheric models.
The accurate representation of the turbulent exchange fluxes in models is challenging as the fluxes result from an interacting chain of parameterized processes above and below the land surface (e.g., *Schomburg*, 2011). The turbulent exchange fluxes are typically parameterized following the Monin-Obhukov similarity theory (e.g., *Stull*, 2012). Monin-Obhukov theory describes turbulent motion within the lower atmosphere above homogeneous terrain. In reality the land surface is not homogeneous, but heterogeneous with respect to many parameters and over a wide range of spatial scales. Land surface heterogeneity is created by:

- vegetation cover and surface type (vegetation, bare soil, urban area, etc.),
- terrain morphology (elevation, slope, orientation),
- soil characteristics,
- variability of climatic forcings (e.g., spatial precipitation patterns)

(*Giorgi and Avissar*, 1997). The heterogeneity of the land surface affects the land-atmosphere exchange of energy, momentum, moisture and other constituents and thus impacts energy and moisture budgets. The turbulent exchange over forested areas is typically much stronger than over bare soil areas. This is, firstly, due to the much larger roughness of forests leading to a more pronounced coupling between land surface temperature and near-surface air temperature. Secondly, the roots of the trees can access deep soil moisture reservoirs leading to moister conditions in the lower atmosphere compared to bare soil areas caused by transpiration (*Schomburg*, 2011). This is just one example of how the turbulent exchange fluxes are affected by land surface characteristics. Idealized simulations with a mesocale model by

Avissar and Pielke (1989) have shown, for instance, strong differences for (maximum) latent and sensible heat fluxes over different surfaces (460 W/m^2 difference between sensible heat fluxes over water bodies and urban areas; 610 W/m^2 difference between built up areas and cropland). Measurement campaigns have confirmed large differences in the turbulent fluxes for different land surfaces (e.g., LITFASS *Lindenberg Inhomogeneous Terrain - Fluxes between Atmosphere and Surface*, *Beyrich and Mengelkamp*, 2006 ; EVA-GRIPS *Evaporation at Grid and Pixel Scale*, *Mengelkamp et al.*, 2006).

Atmospheric models are computationally expensive and therefore limited in grid spacing. Even with increasing computational power (and smaller grid spacing) the land surface remains heterogeneous at the subgrid-scale. Many effects induced by subgrid-scale land surface heterogeneity can be parameterized (e.g. *Avissar and Pielke*, 1989; *Avissar*, 1992; *Koster and Suarez*, 1992; *Seth et al.*, 1994; *Leung and Ghan*, 1995; *Schlünzen and Katzfey*, 2003; *Heinemann and Kerschgens*, 2005; *Ament and Simmer*, 2006).

The effects of land surface heterogeneity are often split into aggregation and dynamical effects (e.g., *Giorgi and Avissar*, 1997). Aggregation effects occur when two land surface processes F and G are nonlinearly dependent on the heterogeneous surface variables x, y such that the grid box averaged effect of x on F cannot be represented by averaging x over the grid box and than applying F, i.e.,

$$\overline{F(x)} \neq F(\bar{x}); \overline{G(x)} \neq G(\bar{x})$$

and such that combined effects of heterogeneity of x on F and y on G cannot be calculated from the averaged effects of $F(x)$ and $G(y)$, i.e.,

$$\overline{F(x)G(y)} \neq \overline{F(x)}\ \overline{G(y)} \neq F(\bar{x})G(\bar{y}).$$

Aggregation effects have been shown to affect latent and sensible heat fluxes as well as the simulation of snow, soil moisture dynamics and runoff as all these processes exhibit a nonlinear dependency on land surface characteristics and/or state variables at the surface and/or the lower atmosphere. The use of averaged parameters and state variables can introduce significant biases when simulating such processes. Aggregation effect models aim to reduce these biases by calculating the contribution of subgrid-scale heterogeneity to the grid box average of land-atmosphere exchange fluxes, water budgets and so on.

Dynamical effects are associated with micro- and mesoscale circulations induced by land surface heterogeneity. These can influence the boundary layer structure and the vertical transport of momentum, energy and water. Models of dynamical effects attempt to simulate the relevant impacts of the land surface heterogeneity induced micro- and/or mesoscale circulations. In coarser-scale models with a grid spacing in the order of 10-100 km both dynamical and aggregation effects are not sufficiently resolved. In smaller-scale models, such as the COSMO-DE (2.8 km grid spacing), dynamical effects are explicitly modeled to some degree. However, the aggregation effects remain relevant also for grid spacings of few kilometers, as surface heterogeneity is present down to very small scales.

Aggregation methods seek to parameterize aggregations effects and can be split into two classes, the discrete methods and the PDF methods. The discrete methods (tile and mosaic

approaches) divide each model grid box into a number of homogeneous subregions denoted as tiles or patches. The surface calculations are carried out separately for each tile and afterwards aggregated to the full model grid box by computing an area weighted average over all tiles/patches. In the PDF methods the parameters and variables which are heterogeneous on the subgrid scale are described by either analytically or empirically derived PDFs. Aggregation is then carried out in the phase space spanned by the parameter's PDFs.
Tile approaches (e.g., *Avissar and Pielke*, 1989; *Koster and Suarez*, 1992)[1] divide the model grid box into a number of homogeneous tiles that exchange fluxes with the atmosphere directly and independent of each other. The subgrid-scale tiles can be defined based on vegetation types (*Koster and Suarez*, 1992), topographic elevation (*Leung and Ghan*, 1995) or via a combination of different land surface characteristics such as vegetation, soil, slope orientation and so on as in *Avissar and Pielke* (1989). Tile schemes do not keep track of exact location of the tiles.
In the explicit subgrid approach by *Seth et al.* (1994) each model box is divided into N^2 subgrid elements (cf. Fig. 3.3), i.e., a higher-resolution land surface scheme is nested into the coarser-resolution atmospheric model. In the following the explicit subgrid scheme is also referred to as mosaic scheme. In *Seth et al.* (1994) each subgrid cell is governed by a single vegetation type (either the most frequent one or a type whose characteristics match the characteristics averaged over all surface types) and bare soil. The main difference between tile and mosaic approach is that in the mosaic approach each subgrid cell is assigned a specific location. Thus, climatic forcing can be explicitly distributed over the subgrid cells. The basic assumption motivating the discrete mosaic is that subgrid-scale climatic forcing experienced by the land surface is important for the calculation of net exchange of heat, moisture, momentum. Still, the subgrid-scale heterogeneity of the surface does not penetrate vertically above the surface layer. The atmosphere only sees the effective fluxes and dynamical effects are not accounted for.

3.2. Atmospheric Disaggregation

In most studies applying tile or mosaic approaches (e.g., *Avissar and Pielke*, 1989; *Koster and Suarez*, 1992) grid box averaged atmospheric forcing is applied to each subgrid tile or grid cell, i.e., no atmospheric disaggregation is carried out. However, the lower atmospheric boundary layer is known to be highly heterogeneous especially above heterogeneous land surfaces even down to small scales (below 1 km).
Pitman et al. (1993) have investigated the effects of the assumption of constant precipitation over the subgrid-scale tiles in GCM simulations. The authors have compared simulations with constant precipitation over each grid box with simulations where precipitation has been disaggregated such that only a fraction of each grid box is governed by precipitation. The intensity has been disaggregated as the grid box average precipitation divided by the precipitation fraction. It has shown that precipitation disaggregation can have a huge impact

[1] *Tile approaches* are sometimes also referred to as *mosaic approaches*. When we talk of mosaic approaches in this study, we refer to the the explicit subgrid approach as formulated in *Seth et al.* (1994).

on runoff simulation. The simulated water budget changed from evaporation dominated to runoff dominated. Other studies showed less sensitivity to fractional precipitation disaggregation (*Giorgi*, 1997a,b), which might be due to different runoff parameterizations. Hydrological models often operate at even higher resolutions than land surface schemes. Several studies, for instance by *Singh* (1997) and *Segond et al.* (2007), have confirmed the importance of a realistic distribution of precipitation for evaporation and runoff simulation.
Not only the subgrid-scale variability of precipitation affects the simulation of land surface and subsurface processes, but also the subgrid-scale of near surface temperature, humidity, wind speed and incoming radiation affects land and subsurface. *Shao et al.* (2001) have examined the effects of both land surface heterogeneity and near surface atmospheric heterogeneity on the simulation of the surface energy and momentum fluxes with a mesoscale model. A series of numerical experiments has been carried out over a domain of 40×40 km centered around Cologne. In the simulations different grid spacings (1 km, 2 km and 4 km) for atmosphere and land surface have been used. It has shown that not only an increased grid resolution of the land surface for a given atmospheric grid resolution improves the simulation of the fluxes, but also an increased atmospheric grid resolution for a given land surface grid resolution leads to an improved flux estimation. While the former is widely agreed on, the latter contradicts the often prevailing view that atmospheric subgrid variability especially on smaller scales (e.g., in meso-γ scale weather prediction) only plays a minor role compared to the land surface heterogeneity itself. Subgrid atmospheric motions (and thus variability) might be an important factor to be included in subgrid closure schemes of atmospheric models.

3.2.1. Early Approaches

The discrete (mosaic) scheme (*Seth et al.*, 1994) allows the usage of explicitly distributed atmospheric forcings. *Seth et al.* (1994) have applied a simple disaggregation for temperature, humidity and convective clouds and precipitation. Temperature and humidity have been either downscaled proportional to soil temperature or soil moisture anomalies or based on topographic height anomalies. The simulations have been carried out with very coarse grid spacing of approximately 30° ($\approx$ 300 km) for the atmosphere and 5° ($\approx$ 50 km) for the land surface. For simulations over 20 years the atmospheric disaggregation has changed the heat fluxes up to 15% and runoff up to 33%.
Giorgi et al. (2003) have adapted the approach of *Seth et al.* (1994) for simulations with a regional climate model over the Alpine region with a 60 km grid spacing for the atmosphere and a 10 to 15 km grid spacing for the land surface. Near-surface atmospheric temperature and specific humidity have been downscaled based on topographic height. Convective precipitation has been distributed over one randomly chosen third of the subgrid-scale pixels. The analysis of an 11 months simulation has shown an improved near-surface temperature over the complex Alpine terrain and a more realistic representation of snow patterns, which may lead to a better simulation of the seasonal evolution of the surface hydrology.
Molod et al. (2003) have presented a technique called extended mosaic. In the extended mosaic not only the land surface processes are simulated in the subgrid-scale tiles, but also turbulent motion in the boundary layer. A comparison of GCM simulations with the standard land

surface mosaic and the extended mosaic have shown large differences for various regions all over the globe (*Molod et al.*, 2004).

3.2.2. The Schomburg 3-Step Scheme

The downscaling scheme of *Schomburg et al.* (2010) has been developed for downscaling near-surface atmospheric fields from 2.8 km to 400 m scale and consists of three steps. The three steps can be applied consecutively or individually for stand-alone offline simulations as well as in a fully coupled model system, such as TerrSysMP (*Shrestha et al.*, 2014). In the first step, a biquadratic spline interpolation is used to interpolate the coarse-scale atmospheric data to a higher resolution while conserving mean and horizontal gradients of the coarse field. In the second step, deterministic downscaling rules are applied to the interpolated field. The rules are based on relations between atmospheric variables and the high resolution surface information. In the last step, autoregressive noise is added to the field to restore the high resolution variance of the fields (cf. Fig. 3.1).

Step 1: Spline interpolation

The biquadratic spline interpolation smooths the coarse field and can be written as

$$y_{ij} = y(i,j) = a_1 + a_2 i + a_3 j + a_4 i^2 + a_5 j^2, \tag{3.1}$$

with y denoting an atmospheric variable, for instance temperature, (i,j) the grid point on the fine-scale, and $a_1, ..., a_5$ denoting the regression coefficients. To estimate the regression coefficients five constraints are introduced: The derivatives of Equation 3.1 at the four edges of the coarse pixel are required to equal the gradient between the coarse pixel and the corresponding neighboring pixel. Further, the coarse pixel mean is conserved.

Step 2: Deterministic downscaling rules

The deterministic downscaling rules are applied to estimate the high-resolution anomalies on top of the interpolation. High-resolution surface information serves as predictors in a linear regression. Some of the near-surface atmospheric variables can be downscaled exploiting known physical relationships:

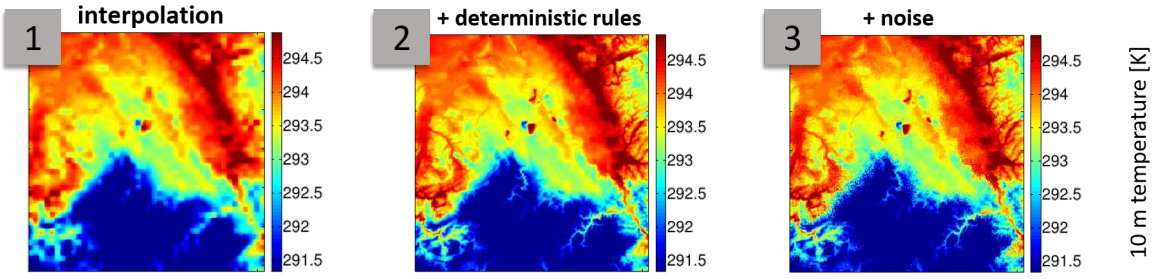

Figure 3.1.: The three steps of the downscaling scheme by *Schomburg et al.* (2010) applied to a 10 m temperature field on May 12th 2008, a clear sky day, at 10 UTC (Figures by A. Schomburg).

- Surface **pressure** anomaly Δp is estimated using relief height Δz as predictor in the hydrostatic equation

$$\Delta p = -\rho g \Delta z \tag{3.2}$$

with the assumption of a constant air density of $\rho = 1.19$ kg/m^3.

- For cloud free skies the diffuse part of upwelling **shortwave radiation** $S_{dif} \uparrow$ is downscaled using surface albedo for direct α_{dir} and diffuse α_{dif} radiation by

$$S_{dif} \uparrow = \alpha_{dir} S_{dir} \downarrow + \alpha_{dif} S_{dif} \downarrow \tag{3.3}$$

with $S_{dir} \downarrow$ and $S_{dif} \downarrow$ denoting direct and diffuse downwelling radiation.

For the remaining five variables (temperature, wind speed, specific humidity, longwave radiation, precipitation) there are no known direct relationships to the surface characteristics. Therefore, the training data set has been evaluated for possible correlations, which usually depend on the prevailing weather conditions. An automatic rule detection system has been set up to find the best predictors, indicators and thresholds. The system calculates the correlations between the 5 predictands and 16 possible predictors given by the surface parameters and fields derived therefrom for different subsets of the training data. The selection of the data subsets is based on 24 different indicators, e.g., vertical temperature gradients or wind speed below certain thresholds. The system selects only rules achieving correlations above 0.7 and applicable to at least 10% of the data:

- For the near surface **temperature** several rules have been found. The best rule found uses orographic height information for downscaling when the temperature gradient of the lowest 105 m is smaller than 0.0057 Km^{-1}.

- The **longwave net radiation** can be disaggregated using ground temperature as predictand when the cloud cover is below 43% or when the longwave net radiation is less than -82.5 Wm^{-2}. These two indicators are almost equivalent.

For other weather situations and the remaining variables (specific humidity, wind speed and precipitation) no applicable rules could be found.

Step 3: Noise generation

As many processes at the surface are nonlinear, also the reproduction of variance can be important to reduce biases. Except for the near-surface pressure, steps 1 and 2 do not reproduce all fine-scale variability. Therefore, in step 3 autoregressive Gaussian noise is added to the fields resulting from step 1 and 2. For this a stepwise multiple linear regression has been applied to predict the fine-scale standard deviation from the coarse-scale standard deviation (of the surrounding 3×3 coarse pixels) of the respective variable and other atmospheric parameters. The autoregression coefficients are obtained from several high-resolution model runs. In the third step spatial correlations are ignored.
The coarse pixel mean is conserved by subtracting the difference between the coarse value and

the mean of the downscaled values for each coarse grid cell. This is important to assure that the conservation of energy and mass is not violated by the downscaling. In case the downscaling predicts unphysical values (e.g., negative wind speeds or negative values for shortwave radiation) the respective values are set to zero or to the coarse pixel mean (for wind speed). In such cases the conservation of the mean is ensured by multiplying the subgrid values by the fraction of the coarse mean before and after correcting the unphysical values.
Precipitation is treated differently from all other variables since the assumption of Gaussian noise does not model the distribution of the precipitation anomalies well. For precipitation the Gaussian noise term is transformed via an exponential function to match the distributions estimated from two high-resolution model runs that have generated precipitation. The transformed noise terms are multiplied by the coarse pixel mean precipitation.

3.2.3. TopoSCALE

Fiddes and Gruber (2014) have presented a physically based and computationally efficient downscaling scheme called TopoSCALE for gridded climate data in complex terrain. TopoSCALE is foremost aimed at creating high-resolution forcing data for land surface models ($\leq$ 100 m) from general circulation model output (50-100 km) by using fine-scale topography information from a high-resolution digital elevation model.

- **Temperature, wind speed and specific humidity** are interpolated according to the subgrid-scale topographic height and the vertical gradients of the respective variables, the latter being obtained from the coarse-scale atmospheric model output. For wind speed an additional topographic correction according to *Liston and Sturm* (1998) is optionally applied.

- **Shortwave radiation** is downscaled using multiple steps. First, shortwave radiation is partitioned into direct and diffuse components. Direct shortwave radiation can then be downscaled according to the difference of the optical path length determined from the topographic heights at grid- and subgrid-scale. Topographic corrections are applied to both diffuse and direct radiation at subgrid-scale to account for shadowing effects occurring especially within complex terrain.

- **Longwave radiation** is downscaled in multiple steps. First clear sky emissivity is determined at the grid and subgrid-scale using grid-scale and downscaled (i.e., subgrid-scale) temperature and specific humidity. Next all-sky emissivity at grid scale is determined from the grid-scale longwave radiation and temperature utilizing the Stefan-Boltzmann equation. The difference between grid-scale clear sky and all sky emissivity provides the correction factor for the subgrid-scale longwave radiation obtained from subgrid-scale temperature according to the Stefan-Boltzmann equation. This approach assumes that cloud emissivity at grid and subgrid elevations are the same, but accounts for the reduction of clear-sky emissivity with height. After the elevation correction, terrain effects are accounted for by multiplication with the sky-view factor, i.e., the fraction of the sky visible from subgrid-scale pixel.

- **Precipitation** is downscaled assuming a simple nonlinear lapse and optionally utilizing precipitation climatologies.

TopoSCALE has been tested and compared with unprocessed coarse GCM data and a set of simple disaggregation methods (e.g., assuming a fixed lapse rate for temperature). A comparison for up to 210 local stations in the Swiss Alps has shown significant improvements for TopoSCALE for air temperature, relative humidity and incoming longwave radiation compared to the reference methods.

3.2.4. VERTEX

Recent studies by *de Vrese and Hagemann* (2016) and *de Vrese et al.* (2016) suggest a conceptually different approach called *VERtical Tile EXtension* (VERTEX), which can be seen as an advancement of *Molod et al.* (2003). VERTEX expands the concept of the tile approach into the vertical. Horizontal homogeneity is thus explicitly considered not only at the land surface, but also within the lower atmospheric model layers, where turbulent mixing is calculated for the single tiles defined by the different land surfaces. In addition to *Molod et al.* (2003) also horizontal turbulent exchange between the subgrid-scale tiles is explicitly included in the scheme.
In *de Vrese et al.* (2016) single-column simulations at the GCM scale incorporating the VERTEX scheme and employing a simple flux-aggregation scheme have been carried out. It has shown that the vertical turbulent transport can largely differ between the subgrid tiles. Further, the comparison of the simulations with and without VERTEX has shown that the horizontal disaggregation of the turbulent mixing process considerably impacts the mean state of the grid box. In the simulations the impact of the explicit subgrid-scale representation of the lower ABL has been, roughly approximated, half as large as for the explicit representation of land surface heterogeneity.
The VERTEX technique is still in an early stage of development, but first tests suggest that it might offer a promising approach to improve the representation of aggregation effects in coupled land surface-atmospheric simulations. In the first application (without any computational optimization) a model set up using 14 tiles for both, land surface and atmosphere, models with the VERTEX scheme take almost 40% longer than simulations with a standard flux-aggregation scheme. Thus, computational optimization will be a crucial to make the VERTEX scheme applicable for long term GCM simulations. Studies with smaller grid spacings (e.g., for RCMs or mesoscale NWP models) have, to our current knowledge, not yet been carried out.

3.3. The Terrestrial Systems Modeling Platform TerrSysMP

The Terrestrial Systems Modeling Platform (TerrSysMP) by *Shrestha et al.* (2014), which has been developed within the Transregional Collaborative Research Center 32, offers a highly modular framework for simulations of the soil-vegetation-atmosphere system. TerrSysMP consists of the atmospheric model COSMO, land surface model CLM and the groundwater

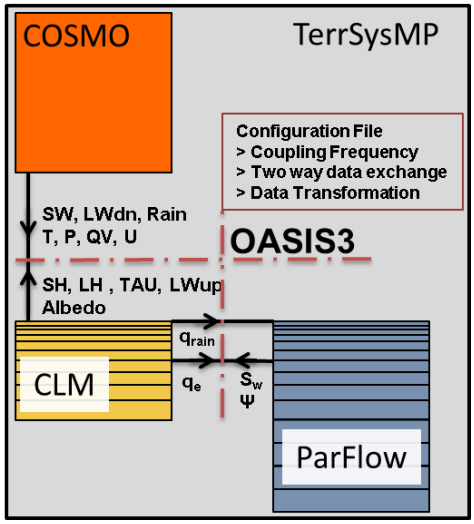

Figure 3.2.: Schematic diagram of TerrSysMP from *Shrestha et al.* (2014). Atmospheric model COSMO, land surface model CLM and subsurface model ParFlow are coupled via the external coupler OASIS3, which manages the data exchange between the component models ©American Meteorological Society (AMS).

model ParFlow. The three component models are coupled using the OASIS coupler, which allows for a separation of model grid spacing, time stepping and coupling frequency between the component models (cf. Fig. 3.2).

The lower boundary of atmospheric models is commonly represented by soil-vegetation-atmosphere transfer (SVAT) schemes. From an atmospheric modelers perspective state-of-the-art land surface and groundwater models can offer an improved parameterization of the lower atmospheric boundary and thereby improve the representation of the exchange fluxes of energy, moisture and momentum. This is important as these fluxes largely affect the evolution of the atmospheric boundary layer. A better understanding and representation of surface fluxes might ultimately lead to better atmospheric and hydrological predictions (e.g., *Avissar and Pielke*, 1989; *Betts et al.*, 1996; *Walko et al.*, 2000; *Ament and Simmer*, 2006). Recent studies further suggest that the surface fluxes can be strongly coupled to groundwater table dynamics (e.g., *Maxwell et al.*, 2007; *Miguez-Macho et al.*, 2007; *Anyah et al.*, 2008; *Maxwell and Kollet*, 2008; *Kollet and Maxwell*, 2008; *Rihani et al.*, 2010).
Increasing computational power allows for the use of integrated modeling approaches, such as TerrSysMP, simulating interaction and feedback processes between soil, vegetation and atmosphere. Idealized simulations with TerrSysMP (*Shrestha et al.*, 2014) support previous studies showing a strong linkage between groundwater dynamics, biogeophysical processes and the atmospheric boundary layer. Simulations over the Rur catchment exhibit patterns in root zone soil moisture, which impact turbulent exchange fluxes and thus boundary layer evolution. *Rahman et al.* (2015) have found evidence that, especially under strong convective conditions, groundwater table dynamics can affect atmospheric boundary layer height, convective available potential energy, and potentially precipitation. Thus, it appears that not only a sufficient representation of land surface processes is crucial to atmospheric modeling, but also the representation of subsurface processes.

As the land surface (topography, land cover) and the soil (soil type) are highly heterogeneous at small scales and strongly affect the spatial variability of soil moisture and groundwater table depth, land surface and subsurface/hydrological models are typically applied with a relatively small grid spacing. Grid spacing of typical atmospheric mesoscale models is too large to sufficiently represent small-scale variability of land surface and soil. Thus, a scale gap between atmosphere and land surface occurs. In modular platforms, such as TerrSysMP, it is common to apply land surface and subsurface models with a higher spatial resolution than the atmospheric model (following the mosaic approach; cf. Fig. 3.3). Not only the use of averaged state parameters at the land surface, but also for the lowermost atmosphere can introduce biases in flux estimation. Downscaling/disaggregation schemes seek to overcome this issue aiming at a more realistic representation of surface fluxes and their spatial variability.
In the following the component models are briefly introduced together with the OASIS coupling interface which, for TerrSysMP, optionally incorporates the 3-step downscaling scheme by *Schomburg et al.* (2010) (cf. Sec. 3.2.2). The COSMO model, which has been applied for this thesis (cf. Chap. 6), is introduced in more detail in the Section 3.4.

COSMO

The COSMO model is a non-hydrostatic limited area model for atmospheric predictions. The basic version of the model has been developed at the *German Weather Service* (*"Deutscher Wetterdienst"*, DWD). The *Consortium of Small-Scale Modeling* (short COSMO), which is comprised of several mainly European weather services, coordinates improvements, maintenance and applications. The TerrSysMP setup of *Shrestha et al.* (2014) contains the convection permitting model configuration COSMO-DE (*Baldauf et al.*, 2011), which has in its operational setup a grid spacing of 2.8 km.

CLM

The *Community Land Model* (CLM; *Oleson et al.*, 2004, 2008) is the land surface scheme of TerrSysMP. In the coupled setup, CLM serves as an interface between atmospheric and subsurface model and calculates the fluxes of energy, momentum, moisture and carbon between soil, vegetation and atmosphere. The CLM is a complex land surface model comprised of biogeophysical, biogeochemical, (dynamic) vegetation and simple hydrological components. Biogeophysical processes modeled by CLM include absorption, reflection and transmittance of solar radiation; absorption and emission of longwave radiation; momentum, sensible and latent heat transfer from canopy and soil surface; plant physiology and photosynthesis; and canopy, snow and soil hydrology.
The CLM allows for the representation of subgrid-scale surface heterogeneity by a tiling approach with thee subgrid levels: Each CLM grid cell can be comprised of up to five different land cover types (glacier, lake, wetland, urban, vegetation). Each of those land units can have multiple soil columns with distinct soil types. Each soil column can have up to 4 (out of 16 possible) plant functional types (PFTs) with each PFT having its own distinct set of plant physiological parameters.

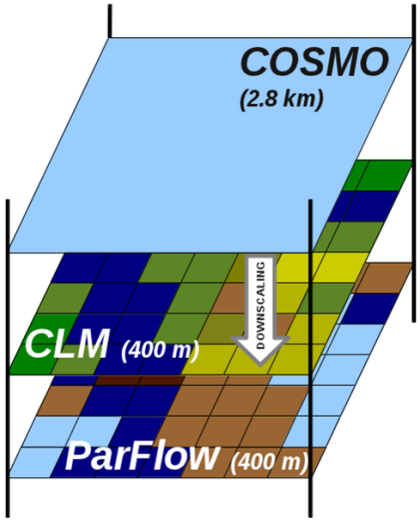

Figure 3.3.: Illustration of the scale gap occurring between the atmospheric component model (COSMO), and the land surface (CLM) and subsurface (ParFlow) component models of TerrSysMP. The grid spacings of the single component models can be flexibly chosen by the user of TerrSysMP. Using a smaller grid spacing for land and subsurface than for the atmosphere is often reasonable as discussed in the text.

ParFlow

ParFlow is a variably saturated 3D Richardson equation based groundwater flow model with two dimensional overland flow (*Kollet and Maxwell*, 2006). Recently a terrain following grid transformation with a variable vertical discretization has been implemented (*Maxwell*, 2013), which leads to significant reductions in computational cost, especially for regions with distinct topography, and thus allows the use of finer grid resolutions near the surface and within the root zone leading to an improved representation of land surface fluxes.

OASIS

The OASIS3 coupler (*Valcke*, 2013) serves as the coupling interface of TerrSysMP. OASIS allows a highly modular usage of the modeling platform. The TerrSysMP user can not only run the fully coupled system (i.e., COSMO-CLM-ParFlow), but also stand-alone versions of the component models, COSMO only coupled with CLM or CLM coupled with ParFlow with offline atmospheric forcing.
Model definition (model grid, input/output variables, ...) initialization and termination as well as the data exchange between the component models are all organized via the OASIS interface. Coupling frequency, names and spatial grid of the coupling fields as well as potential transformations of the coupling fields are specified in the OASIS configuration file. Time-integration-averaging and spatial interpolation operators are available to transform the 2D coupling fields before transferring them between the component models. Time integration and averaging is used when the coupling time step is larger than a component model time step. In *Shrestha et al.* (2014) precipitation send from COSMO to CLM is integrated in time and the remaining atmospheric variables are temporally averaged. Spatial interpolation is required

when the coupling variables are defined on different grids. COSMO for instance uses a rotated geographical coordinate system, while CLM variables are defined on a regular geographical grid. Thus, the coupling fields exchanged between COSMO and CLM need to be interpolated to the respective grids.

COSMO-CLM At the interface between the atmospheric model COSMO and the land surface model CLM atmospheric forcing terms and surface fluxes are exchanged. The state variables of the lowest COSMO layer at the current time step serve as forcing for CLM. The COSMO model sends temperature T, horizontal wind speed w_h, specific humidity q, (convective and grid scale) precipitation P, pressure p, incoming direct and diffuse shortwave SWR_{dir} and SWR_{dif} and incoming longwave radiation LWR_{dif}. CLM then calculates the surface fluxes of latent and sensible heat LH and SH, momentum fluxes τ, albedo α and outgoing radiation LW and sends them back to COSMO (cf. Fig. 3.2). The estimated fluxes are used to update the (dimensionless) transfer coefficients of heat and momentum at the surface in COSMO.
In TerrSysMP the existing tile approach for CLM is not used. Instead a mosaic approach is employed where each atmospheric grid cell contains multiple land surface grid cells. In the mosaic each CLM grid cell is homogeneous. That is, each grid cell is assigned one PFT, one soil column and one land unit. The mosaic approach allows to combine the simulation of subsurface hydrodynamics with overland flow as done by ParFlow. Furthermore, the mosaic allows the use of distributed atmospheric forcings. In the current version of TerrSysMP atmospheric downscaling according to *Schomburg et al.* (2010) is available as an optional component of the OASIS interface.

CLM-ParFlow In the coupled mode ParFlow, with its 3D variably saturated groundwater representation and its free surface overland flow boundary condition, computes surface runoff and subsurface hydrodynamics, which allows for a flexible, potentially 3D, redistribution of soil moisture and groundwater. The 1D column soil moisture representation of CLM is replaced by ParFlow (in 3D or 1D formulation). The OASIS interface allows for dynamical interactions between CLM and the 3D groundwater model. ParFlow sends relative saturation S_w and pressure head Ψ of the upper 10 soil layers to CLM. CLM sends the source and sink terms for soil moisture, top soil moisture flux q_{rain} and soil evapotranspiration q_e, of the upper 10 layers to ParFlow (cf. Fig. 3.2).

3.4. The COSMO Model

The COSMO model is based on the primitive (nonhydrostatic) hydro-thermodynamic equations and is designed for both scientific applications and operational weather forecasts. The operational forecast chain of the German Weather Service (DWD) consists of three models (Fig. 3.4). The global model ICON has a grid spacing of 13 km and provides initial and boundary conditions for ICON-EU. ICON-EU has a grid spacing of 6.5 km and provides initial and boundary conditions for COSMO-DE, which has a grid spacing of 2.8 km and 50 vertical layers. For the 400 m simulations in this thesis (cf. Sec. 6.1) COSMO-DE had to be adapted.

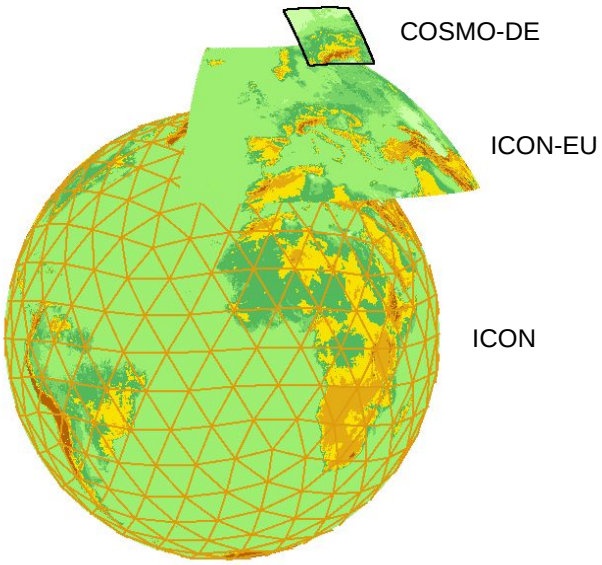

Figure 3.4.: The current operational model chain of the German Weather Service (© DWD).

A detailed description of the COSMO model is provided in the model documentation (*Doms and Baldauf* 2015; *Doms et al.* 2011; *Schaettler et al.* 2015).

Coordinate System and Domain

In the horizontal the COSMO model uses rotated latitude/longitude (λ, φ) coordinates, which are obtained from the geographical coordinates by tilting the north pole such that the equator of the rotated grid lies in the center of the model domain. This way the model grid is approximately regular. In the vertical generalized terrain following coordinates (ς) are used. Unlike pressure-based vertical coordinates ς is fixed in time.
In COSMO-DE the north pole of the model grid is located at 40.0° N and 170° W. The model domain covers whole Germany including parts of adjacent countries and spans an area of approximately 1300×1200 km^2.

Basic State and Model Equations

The basic equations are formulated using a time-independent base state given by a horizontally homogeneous, hydrostatically balanced and resting atmosphere. For computational reasons, the thermodynamic quantities pressure p, temperature T and density ρ are formulated as the sum of the height-dependent base state (e.g., T_0) and a location and time dependent deviation (e.g., T')

$$T(x, y, z, t) = T_0(z) + T'(x, y, z, t).$$

The height-dependent base states $p_0(z)$, $T_0(z)$ and $\rho_0(z)$ are related via the hydrostatic equation $\partial p_0/\partial z = -\rho_0 g$ and the equation of state $p_0 = \rho_0 R_d T_0$, with R_d being the gas constant for dry air and and g the gravitational acceleration. A constant rate β for the temperature increase with the logarithm of pressure is used ($\partial T_0/\partial \ln p_0 = -\beta g$). The full base state can then be defined by three values: β, which is set to 42 K, and the pressure and temperature at

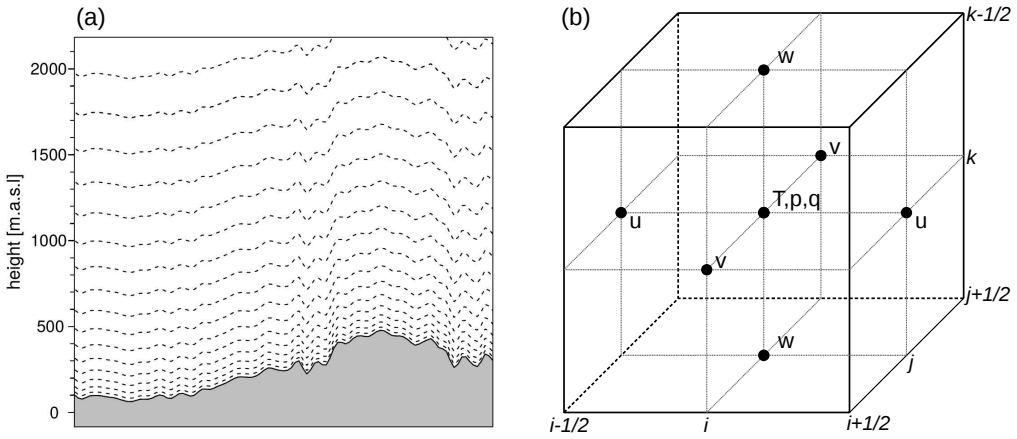

Figure 3.5.: Grid structure of the COSMO model: (a) shows the heights (in meters above see level) of the lowest vertical layers of COSMO-DE over hilly terrain. The dashed line corresponds to the half levels. The solid line indicates the lowermost half level, which corresponds to the topography. (b) shows one model grid box of COSMO ($\Delta V = \Delta\lambda\Delta\varphi\Delta\varsigma$) with the Arakawa-C/Lorenz staggering of the prognostic variables.

sea level height, which are set to $p_{SL} = p(z = 0) = 1000$ hPa and $T_{SL} = T(z = 0) = 288.15$ K, respectively.

With the definition of the base state and the coordinate transformation described above, the primitive hydro-thermodynamic equations are transformed into a set of prognostic equations for the vector of wind speed $\vec{v} = (u, v, w)$, pressure deviation p', temperature deviation T', specific humidity q^v, specific cloud water content, specific cloud ice content and the specific water contents of rain, snow and graupel. Effects of subgrid-scale processes on the prognostic variables are determined from paramterization schemes.

Discretization and Integration

The model equations are solved numerically with using finite differences. That means, the prognostic variables are defined at grid points and the spatial differential operators are approximated by finite differences. The temporal integration is also carried out in a discrete form with a fixed time step Δt.

For spatial discretization a regular latitude/longitude-grid with grid spacings $\Delta\lambda = \Delta\varphi = 0.025° \approx 2.8$ km and a fixed terrain following vertical coordinate with a vertical level spacing $\Delta\varsigma$ are used. The height of the lowest layers of COSMO-DE is shown in Fig. 3.5(a).

The grid point in the center of each model grid box is denoted by (i, j, k) where i relates to the λ-direction, j relates to the φ-direction and k relates to the ς-direction. The prognostic variables are staggered on an Arakawa-C grid (Fig. 3.5(b)). All prognostic variables, except wind speed and turbulent kinetic energy, are defined at the grid box centers. The wind speed components (u, v, w) and the turbulent kinetic energy TKE are defined at the boundaries of the grid box.

The centers of the vertical model layers, denoted by k, are called main levels. The half levels, denoted by $k + / - 1/2$, are shifted up- or downwards by $\Delta\varsigma/2$. If the atmosphere is vertically

discretized into n_k layers, we have n_k main levels and $n_k + 1$ half levels. The lowest half level corresponds to the lower model boundary given by the orography. The uppermost half level corresponds to the upper model boundary.
The nonhydrostatic model equations describe a compressible atmosphere. Thus, fast propagating sound waves are part of the solution. To decrease computational cost and still ensure numerical stability the terms describing the propagation of the fast waves are treated separately. The slow processes are integrated from time t to time $t + \Delta t$, i.e., using the model time step Δt. The fast modes are integrated using a smaller time step Δt_s. In COSMO-DE, i.e., for a grid resolution of 2.8 km, a time step of $\Delta t = 30$ s is sufficient. The integration is carried out using a Runge-Kutta scheme.

Physical Parameterizations

Physical processes that are not resolved by the grid spacing are parameterized. In the following a short summary of the parameterization schemes used in COSMO-DE is given.

- The **turbulence** parameterization estimates the effect of subgrid-scale turbulent diffusion. The subgrid-scale flux of a variable is related to its gradient via a diffusion coefficient K. The diffusion coefficients for heat and momentum, K_h and K_m, are determined in the parameterization scheme. The turbulence scheme distinguishes between a transfer layer close to the surface, which ranges from the land surface to the lowest main level, the planetary boundary layer and the free atmosphere.
  In the planetary boundary layer and in the free atmosphere the calculation of the turbulent exchange coefficients is based on a prognostic equation for the turbulent kinetic energy. The exchange coefficients depend on atmospheric stability and vertical wind sheer.
  For the parameterization of the turbulent exchange fluxes at the surface, the transfer layer is subdivided into a Prandtl layer, a roughness layer and a laminar layer just above the surface. In the laminar layer the resistance is a linear function of height. In the roughness layer the resistance is an exponential function of height and depends on the roughness length and the leaf area index (LAI). Roughness length and (maximum) LAI are prescribed by an external data set. Over water the roughness length is calculated from the Charnock equation (*Charnock*, 1955). In the Prandtl layer the resistance is a logarithmic function of height and dependens on stability.

- **Soil Processes** are parameterized by a soil module that predicts soil temperature and water content. The temperature and humidity at the land surface serve as lower boundary conditions for COSMO and are important to determine the exchange fluxes of heat, moisture and momentum. The multi-layer soil module of COSMO (TERRA-ML) contains 7 layers to describe the soil processes. The 8th and lowest layer in approximately 15 m depth is ascribed climatological temperatures. Within the upper 6 layers hydrological processes are described by a diffusion equation for the water transport. In the lower layers (below approximately 2.5 m depth) capillary transport is neglected and the gravitational flux acts as runoff/drainage. Most of the parameters required by the soil

module such as heat capacity are dependent on the soil texture. All grid elements with a fraction of land larger than 50% are considered land pixels. All other grid elements are considered water pixels. The temperature for water pixels is obtained from an analysis of the sea surface temperature and kept constant over the simulation period.

- **Radiation** is parameterized according to *Ritter and Geleyn* (1992). The paremeterization is based on a δ-two-stream approximation of the radiative transfer equation and considers three shortwave (solar) and five longwave (terrestrial) spectral intervals. Clouds, aerosols, water vapor and other gaseous tracers affect radiative transfer by emission, absorption and scattering. In COSMO-DE radiative transfer calculations are carried out once in a time interval of 15 minutes, i.e., only every 36th model time step. The calculated warming and cooling rates are kept constant until the parameterization is called the next time. In the operational COSMO-DE setup the radiation calculations are carried out on a coarser grid (by a factor of 2) than the actual model grid to reduce computational cost.

- Grid-scale **precipitation** and clouds are parameterized by a Kessler-type bulk formulation (*Kessler*, 1969) that categorizes atmospheric water particles into different classes. In COSMO-DE water vapour and five different classes of hydrometeors are considered, namely cloud water, cloud ice, rain, snow and graupel particles. The particles belonging to the different classes interact through various microphysical processes, which are parameterized by transfer rates.

- In COSMO-DE only shallow **convection** is parameterized. Deep convection (thunderstorms, ...) is resolved at a grid spacing of 2.8 km. Shallow convection is parameterized according to *Tiedtke* (1989). The mass flux is assumed to be proportional to the vertically integrated moisture convergence between surface and cloud base. Shallow convection does not directly generate precipitation. There is no (parameterized) convective precipitation in COSMO-DE.

- The parameterization schemes for grid-scale clouds and precipitation require saturation equilibrium to determine the condensation rate. Thus, grid-scale clouds are only generated in grid boxes with a relative humidity of 100%. For the radiative transfer calculation and for many post-processing applications, also the representation of **partial cloudiness** is important, i.e., the cloudiness in grid boxes with a saturation below 100%. Partial cloudiness is determined as an empirical function of relative humidity, model layer height and potentially convection.

External, Initial and Boundary Data

The COSMO model requires the following external parameter fields: Orographic height, fraction of land, roughness length, soil type, root depth, plant cover and leaf area index. The external parameter fields are kept constant during a model run. In COSMO-DE the orography is smoothed to ensure numerical stability.

Initial conditions for experimental forecasts and simulations with COSMO-DE can be derived from the COSMO-EU (predecessor of ICON-EU) analysis. The interpolation program (INT2LM) interpolates the 7 km data to the finer 2.8 km grid spacing to obtain an initial state for the COSMO-DE. Because of the different horizontal and vertical resolution of the driving model the initial state is not well defined and COSMO-DE requires a spin-up period of about 3-6 hours to adjust to the higher-resolution topography and develop clouds etc.
As a limited area model COSMO-DE requires boundary conditions. Typically these are also obtained from COSMO-EU and mapped to the COSMO-DE boundaries using INT2LM. The boundary conditions from COSMO-EU are available in time intervals of 1 hour. Between the full hours the boundary information is derived by applying a linear interpolation in time between two sets of boundary conditions.

4 Downscaling of General Circulation Model Simulations

Projections of future climate are generated using general circulation models (GCMs) (e.g. *Taylor et al.*, 2012). The usefulness of GCMs for impact studies is yet restricted by their coarse spatial and temporal resolution:

1. Gridded GCM output represents area means rather than point values.

2. Effects of subgrid-scale heterogeneity, e.g., local and regional topography, land sea distribution and land use, on regional climates can not be explicitly modeled but only through parameterizations.

Downscaling aims to bridge the gap between GCM output and the requirements of impact modelers by relating regional or local climate variables to the larger scale climate simulations. The fundamental assumption behind all downscaling techniques is that relationships can be established between atmospheric processes occurring at disparate spatial and temporal scales

$$local\ climate\ response = \mathcal{F}(external, larger\ scale\ forcing). \tag{4.1}$$

That is, the local climate is assumed to relate to the external and larger scale forcing via some, stochastic or deterministic, function(s) $\mathcal{F}$ (*von Storch et al.*, 2000). Regional climate is not fully determined by the large-scale state, but may be treated as a random process conditioned on the larger-scale climate. $\mathcal{F}$ can be modeled in two conceptually different ways: dynamically through regional climate models (RCMs) or empirically from observational or modeled data sets as a deterministic or stochastic function.
This chapter provides an introduction to the downscaling of climate model output. The chapter relies in large parts on the review articles by *Wilby and Wigley* (1997), *Xu* (1999), *von Storch et al.* (2000), *Fowler et al.* (2007) and *Maraun et al.* (2010). Dynamical and empirical-statistical are reviewed in Section 4.1 and 4.2. Advantages, disadvantages and open challenges are discussed in Section 4.3. Also Genetic Programming ranks among the empirical-statistical downscaling approaches. Studies employing GP for downscaling GCM output are reviewed in Section 4.4.

4.1. Dynamical Downscaling

In dynamical downscaling a regional climate model (RCM) is embedded within the GCM. The RCM covers a subdomain of the globe and receives initial and boundary conditions from the driving GCM. The smaller domain size allows to use a smaller grid spacing. RCMs can be computationally as expensive as GCMs due to the higher spatial and temporal resolution. In principle RCMs model atmospheric processes in a similar manner as GCMs, but the higher grid resolution/smaller time step allows an improved treatment of physical and dynamical processes. This aims at generating realistic mesoscale patterns not resolved by GCMs. For instance, orographically induced precipitation can be much better captured with a higher resolved land surface. *Giorgi and Marinucci* (1996) have found evidence that the skill of RCMs concerning the simulation of spatial patterns and temporal characteristics may increase with increasing grid resolution (see also *Prein et al.*, 2015). With increasing computational power several projects generating ensemble regional climate simulations have been initiated: PRUDENCE (*Christensen and Christensen*, 2007) and ENSEMBLES (*Van der Linden and Mitchell*, 2009) in Europe; NARCCAP (*Mearns et al.*, 2013) in North America; and the Coordinated Regional Downscaling Experiment (CORDEX) considering several regional domains all over the globe (*Giorgi et al.*, 2009).
RCMs provide spatially and temporally coherent data for multiple variables. RCM simulations are typically consistent with the driving GCM, which implies that RCMs can inherit biases from the driving GCM (e.g., *Dosio and Paruolo*, 2011). In climate change studies such biases may largely cancel out when the difference between control runs (past) and projections (future) is considered. However, model biases might change in a changing climate (*Christensen et al.*, 2008). Moreover, the performance of a GCM strongly depends on domain size and location (e.g., *Miguez-Macho et al.*, 2004).
As for the GCMs, RCM grid box values represent area averages. The effective resolution of an RCM (as for every atmospheric model) is larger than the grid resolution, meaning that RCMs only provide meaningful information of precipitation extremes on a larger scale than the grid spacing. *Skamarock* (2004) for instance have found the effective resolution to be about seven times larger than the model grid spacing (in x- and y-direction) for the atmospheric model WRF in a numerical weather prediction setup. For impact studies requiring station or site specific information additional (empirical-statistical) downscaling might be needed.

4.2. Empirical-Statistical Downscaling

The most simple statistical approach is to provide some kind of mapping between the larger scale predictors X and the expected value of predictand Y at the local scale. That is, $E(Y|X) = \mathcal{F}(X, \beta)$ with the mapping function $\mathcal{F}$ and a set of unknown parameters β to be estimated when calibrating a downscaling model.
Wilby and Wigley (1997) have categorized the empirical-statistical downscaling methods into *regression methods, weather type approaches* and *stochastic weather generators*. *Maraun et al.* (2010) have used a categorization into *perfect prognosis* (PP) approaches and *model output*

statistics (MOS) based on the chosen predictors. PP methods exploit predictors that are well simulated by the models assuming that the model offers a perfect prognosis of those variables. MOS on the other hand explicitly aims at a statistical correction of model errors.
After a short note on predictands and predictors considered in empirical-statistical downscaling, the different downscaling approaches are introduced in more detail. Many applied downscaling methods combine different approaches.

Predictors and Predictors

The list of predictands in literature ranges from actual values of climate model variables (e.g., precipitation, temperature, snow pack) to their monthly or yearly statistics (e.g., wind speed distributions) and impact variables (e.g., frequency of land slides). Most relevant variables for impact studies and therefore the most considered predictands are precipitation and temperature with precipitation being much more difficult to model due to its high spatial and temporal variability and its, over a wide range of scales, non Gaussian distribution.
The choice of predictors significantly influences the results of a downscaling method. Appropriate predictors need to

- be variables containing information on the predictands,
- have a stationary predictor - predictand relation (not changing in changing climate),
- fully capture the effect of climatic changes.

GCMs offer a large number of 2D and 3D fields of candidate predictors. Some are known to be more reasonably simulated than others. While tropospheric quantities such as temperature and geopotential height are considered to be skillfully simulated by GCMs, quantities like water vapor, clouds and precipitation are considered to be among the least accurately simulated variables in GCMs, which makes precipitation downscaling even more challenging. Instead of (only) taking grid box values or the high-dimensional fields directly as 'raw' predictors, many downscaling methods derive (additional) physical meaningful predictors by applying dimensionality reduction methods, such as principle component analysis (PCA), to the GCM fields (e.g., *Simon et al.*, 2013). Other derived predictors are for instance airflow direction and strength instead of zonal and meridional wind components as returned by the GCMs or physically motivated indices, such as the North Atlantic Oscillation (NAO) index derived from the North Atlantic pressure field.

Regression Approaches/Transfer Functions

Regression techniques are among the earliest downscaling approaches. Regression establishes a linear or nonlinear relationships between subgrid-scale (e.g., single site) predictands and coarser-scale (e.g., GCM grid-scale) predictor variables.
For quantities that can be well approximated by a Gaussian distribution (e.g., temperature) already a simple (multiple) linear regression (cf. Appendix B) can yield reasonable results (cf. Section 7.3). Still, early methods purely applying (multiple) linear regression are known

to underrepresent variance as standard regression estimates the conditional mean of the predictand given the predictors.
In *Karl et al.* (1990) missing variance is restored by multiplying the downscaled prediction with a suitable factor. This technique is known as variance inflation. A more sophisticated but conceptually similar technique is the expanded downscaling proposed by *Bürger* (1996), which links the covariance of local weather variables to the covariance of the global circulation model output. However, *von Storch* (1999) suggests to rather add an explicit noise term, because the variation of a predictand variable is in fact not fully explained by the predictors.
Precipitation downscaling is typically more difficult than temperature downscaling. While for annual precipitation totals the Gaussian assumption may be feasible, it is surely not on shorter time scales. Daily precipitation is commonly modeled with a gamma distribution (e.g., *Katz*, 1977). Regression for non-Gaussian quantities can be accomplished with generalized linear models (GLMs). In a GLM the predictand might follow any distribution from the exponential family. Like multiple linear regression a GLM models the conditional mean of the chosen distribution (i.e., the expected value of the predictand given the predictors) by a linear function of a set of predictors, but in contrast to a linear model (LM) the conditional mean μ may be transformed by a (e.g., logarithmic) link function. (cf. Appendix B).
In a GLM (or LM) framework variance can be obtained by sampling from the modeled distribution with the variability of the predictand represented by the scale parameter (or standard deviation). Most applications of GLMs are in fact weather generators. GLMs can capture the conditional mean of a wide range of distributions.
Nonlinear regression approaches used for downscaling include Artificial Neural Networks (ANNs e.g., *Schoof and Pryor*, 2001; *Coulibaly et al.*, 2005) and Genetic Programming (GP) (e.g., *Coulibaly*, 2004; *Hashmi et al.*, 2011). The main motivation behind this approaches is to capture nonlinear and non additive predictor-predictand relationships. Further, no specific distribution has to be assumed for the predictand variables. The state of the art in using GP for GCM downscaling is reviewed in Section 4.4 in more detail.

Weather Pattern Approaches/ Weather Typing Schemes

Weather typing schemes relate a local variable, such as precipitation, to a set of weather classes by means of a weather classification scheme. The weather classes can be defined objectively (e.g., based on principal component analysis) or subjectively (e.g., synoptic categories like the "Europaeische Großwetterlagen"). Local weather variables such as precipitation are conditioned on the weather classes. This approach is a special case of a (LM or GLM) regression model with a discrete set of weather types used as predictor instead of continuous fields, i.e., $\mu = \mu(X_k)$ with k being the index of a weather type and $\mu(X_k)$ being the local mean of the predictand variable for weather type k. Weather typing approaches can be extended in a similar manner as LMs and GLMs. A weather type dependent noise term can be added or one can sample from the conditional PDFs instead of taking the expected value for each weather class as prediction. In fact weather typing schemes are most often applied to condition stochastic weather generators (see below). Studies incorporating weather typing into downscaling methodologies can be found for instance in *Enke and Spekat* (1997); *Goodess and*

Palutikof (1998); *Conway and Jones* (1998); *Boé et al.* (2006) and *Vrac et al.* (2007).

Analog Method

The analog method is in principle a special case of a weather typing scheme. Though the analog method has initially been developed for short term weather forecasting (*Lorenz*, 1969), it is widely known and used in the downscaling community. The GCM (or RCM) predictions are compared to the observational record to identify the most similar large scale pattern. This is done by introducing a similarity metric on the global (or regional scale). The corresponding observations Y serve as prediction at the local scale $Y_i = Y(analog(X_i))$ (*Zorita et al.*, 1995; *Zorita and von Storch*, 1999).
Prerequisite for applying an analog technique is a large observational record of global and local scale weather. The selection of the similarity metric is not straightforward. Both the type of the metric (e.g., the Euclidian distance or some weighted form thereof) and the predictor variables incorporated in the metric have to be selected. Known problems of the analog method are systematic over- and underselection of certain observational periods which can lead to an underestimation of temporal variance. For this reason some implementations of the analog method chose randomly between the k most similar situations instead of always picking the most similar one. Further, the standard analog method can not produce values outside the range covered by the observations.

Stochastic Weather Generators

Stochastic weather generators (WGs) are random number generators yielding local scale weather time series that resemble the statistical properties of the observed weather. The core of a typical weather generator consists of a first (or multi-order) Markov process, i.e., for each day precipitation occurrence and possibly precipitation amount is governed by the outcomes on the previous day(s). Precipitation amount is commonly modeled in a second step often assuming a gamma distribution for daily precipitation amounts (e.g., *Richardson and Wright*, 1984). Remaining required variables such as temperature can then be conditioned on the estimated precipitation series.
Weather generators are in an unconditional form not ranked among the downscaling approaches. When WGs are applied for downscaling the WG parameters need to be adjusted based on the GCM projections. This can be done by introducing multiplicative (for precipitation) or additive (for temperature) change factors, which relate the change in climate at the local scale to the larger scale climatic change as predicted by the GCMs (*Kilsby et al.*, 2007). In other WGs the parameters are conditioned on large scale weather (e.g., *Wilks and Wilby*, 1999). The latter approach assumes that the climate change signal is sufficiently captured by the change in the occurrence frequencies of the large scale weather patterns and thus allows for a direct application under changing climate. WGs can reach high degrees of complexity when multiple variables have to be generated (in a consistent way) and/or spatially consistent data (gridded or at a set of local stations) is required. For more details on such approaches see *Maraun et al.* (2010).

Model Output Statistics

Model output statistics (MOS) is foremost applied to correct RCM errors. Most MOS approaches have initially been designed for numerical weather prediction (*Klein and Glahn*, 1974). With the increasing availability of RCM scenarios MOS has become more popular for downscaling. MOS typically establishes relations between the RCM simulation of a variable and the local scale observation of the same variable and thus corrects RCM errors. While the potentially large deviation of modeled precipitation from the 'true' precipitation makes it a questionable predictor in perfect prognosis approaches. However, the simulated precipitation may still contain information about the real precipitation. This is used in MOS where the simulated time series (or intensity distribution) serves as predictor for the local precipitation time series (or properties of the local scale intensity distribution).
Let us consider precipitation as predictand variable. A widely used MOS method is the scaling or direct method (e.g., *Widmann et al.*, 2003; *Lenderink et al.*, 2007). Here the local precipitation at time t (within the projection period) P_t^D is given as the simulated precipitation $P_t^{projection}$ from the projection run corrected by the ratio between mean observed precipitation and mean modeled precipitation over the control period,

$$P_t^D = P_t^{projection} \frac{\bar{P}^{obs}}{\bar{P}^{control}}. \tag{4.2}$$

The above method corrects mean and variance by the same factor, i.e., the ratio between mean and variance (also known as coefficient of variation) is not changed.
A more generalized approach to MOS is quantile mapping (e.g., *Themeßl et al.*, 2011). Quantile mapping considers different precipitation intensities separately. For the calibration period a mapping is established which relates the observational CDF of the calibration period to the respective simulated CDF. The mapping can be established using the empirical CDFs from the calibration period or using theoretical distributions fitted to observational and modeled CDFs. Similar to the analog method standard quantile mapping can not produce values beyond the observed range. If this is desired adjustments are necessary (e.g., *Boé et al.*, 2007). In the context of climate change MOS is mainly used to correct RCMs. However, when treated with caution, it might be also useful in areas where no RCM data is available. Unlike PP methods, MOS approaches have to be individually trained for each RCM (or GCM) as they explicitly aim at correcting (model dependent) model errors.

4.3. Advantages, Disadvantages and Challenges

The most significant advantages and disadvantages of dynamical and empirical-statistical downscaling are listed in Table 4.1. Dynamical downscaling provides physically consistent 3D fields of multiple climate variables by explicitly modeling the underlying physical processes. Forced by GCM data at the boundaries, the RCM simulations are generally consistent with the driving GCMs. However, the RCM skill is highly dependent on the skill of the driving model. RCMs may inherit biases from the driving GCMs. Dynamical downscaling is computationally expensive and still, if impact studies require local point data, e.g., at single stations,

additional empirical-statistical downscaling is needed.

Empirical and statistical downscaling techniques can provide point-scale information right away. Statistical approaches can be used for quantities not modeled in GCMs and are computationally cheap. Like dynamical downscaling also the skill of statistical approaches can depend on the skill of the GCM (or RCM) to which they are applied. Spatial and inter-variable consistency is not guaranteed by many statistical approaches. Results may further depend on the choice of predictor variables. To appropriately calibrate a statistical downscaling model, a sufficient observational data is required. Finally, a crucial assumption behind all empirical-statistical models is that the link established when calibrating the downscaling method also holds for a changing climate.

Many studies have reported similar skill for dynamical and statistical downscaling methods (e.g., *Kidson and Thompson*, 1998; *Murphy*, 1999; *Mearns et al.*, 1999) under present climate conditions. *Schmidli et al.* (2006, 2007) have compared different RCMs against statistical models incorporating regression approaches, weather typing methods, a weather generator, and a bias correction and spatial dissaggregation approach. RCMs and statistical models exhibit similar biases, but the low frequency (year to year) variability and the convection in complex terrain during the summer months has been significantly better captured by the RCMs. It is however debatable if the high computational cost of dynamical downscaling compared to empirical-statistical approaches is merited (*von Storch et al.*, 2000).

Studies addressing the intercomparison of different empirical and statistical downscaling methods have been carried out more and more frequently during the past two decades. A study by *Wilby et al.* (1998) has compared the performance of two weather generators, two artificial neural nets and two vorticity/circulation based approaches for the downscaling of daily precipitation. When considering daily statistics the WGs have performed superior compared to the circulation based methods which in turn perform superior to the ANNs. However, the WGs capture low frequency variability, here measured by the standard deviation of monthly rainfall totals, rather badly. For evaluating monthly totals the above ranking of methods is approximately opposite with the ANNs performing slightly superior to the vorticity based methods which in turn perform superior to the WGs.

Table 4.1.: Advantages (+) and disadvantages (-) of dynamical and empirical-statistical downscaling techniques (adapted from *Wilby et al.*, 2002; *Fowler et al.*, 2007).

Dynamical Downscaling	**Empirical-Statistical Downscaling**
+ provides consistent 3D fields	+ can provide point-scale information
+ consistent with GCM driving fields	+ computationally cheap
+ accounts for physical processes at respective scale	+ can be used for variables not modeled by GCMs/RCMs
- skill depends on GCM	- skill depends on GCM
- computationally expensive	- skill depends on predictor choice
- skill may depend on domain location and domain size	- require long, reliable, consistent historical calibration data
- skill depends on grid spacing (possibly additional statistical downscaling needed)	- potential non-stationarity of predictor-predictand relation
	- spatial, inter-variable, physical consistency not ensured

In *Zorita and von Storch* (1999) the performance of an analog method, a linear method based on canonical correlations and an artificial neural network are compared. The study shows that in case of normally distributed variables linear methods should be preferred as they yield satisfactory results (right level of variability can be easily ensured by adding noise) and provide a direct physical interpretation. For non-normally distributed variables the analog method should be preferred. The concept behind the analog method is simple and it achieved a comparable or even better performance than the more complicated ANNs. However, the authors themselves point out that other studies report a more successful implementation of ANNs for downscaling, e.g., monthly precipitation (*Hewitson and Crane*, 1992).
The results of the above studies illustrate some difficulties faced in intercomparisons. Different dowsncaling approaches typically have different advantages and disadvantages. The skill of a downscaling method depends on the downscaling problem (e.g., considered area, spatial and temporal resolution to be obtained) and the implementation of the method (e.g., predictor choice, optimization techniques) (e.g., *Goodess et al.*, 2007; *Maurer and Hidalgo*, 2008; *Gutmann et al.*, 2012).
Nevertheless some drawbacks seem to be common. Downscaling approaches seldom capture climate variability at all spatial and temporal scales. According to *Maraun et al.* (2010) especially the representation of precipitation on subdaily scales and high-resolution spatial scales is not sufficiently studied. Further, the representation of extremes, especially extreme convective summer precipitation, is still challenging for both dynamical and statistical approaches as convection is hard to predict and typically occurs locally.
Ongoing work in the framework of the COST action *Validating and Integrating Downscaling Methods for Climate Change Research* (VALUE) aims at a comprehensive intercomparison of downscaling methods and coordinates a series of statistical downscaling experiments with different prerequisites and focuses (www.cost-value.eu; *Maraun et al.*, 2015).

Validation

A sufficient validation of a downscaling method needs to demonstrate the robustness of a *downscaling link* under climate change. The predictor-predictand relation needs to hold under changing climate and the predictors need to sufficiently represent the climate change signal. The classical validation approach splits the observational record into a training period to specify the method and an observational record ideally with different characteristics. For instance *Wilks* (1999) split the observation record into dry and wet season for training and validation. Although such approaches can not fully determine if a downscaling method holds under climate change, they provide indications if a method may do so.

Uncertainty

Quantifying uncertainties in regional/local climate projections is difficult as several sources of uncertainty are involved in the generation of the data. The first source of uncertainty is the scenario of future emissions. Many climate projections consider different kinds of emission scenarios to quantify inherent uncertainties. Second, the models (GCMs and RCMs) are sub-

ject to errors and uncertainties. Ensemble simulations with different models, different model setups and/or perturbed initial (and boundary) conditions help to obtain uncertainty estimates for the scenario simulations. Also the regionalization methods (dynamical downscaling and/or statistical downscaling) are subject to errors and uncertainties. Finally, the natural variability of the climate system adds further uncertainty (*von Storch et al.*, 2000). Obtaining reasonable uncertainty quantifications for downscaled climate projections is thus challenging.

4.4. Genetic Programming for GCM Downscaling

The first study (known to us) that applies Genetic Programming (GP) to the downscaling of GCM output can be found in *Coulibaly* (2004). In this study GP is used to simulate local daily maximum and minimum temperatures at the Chute-du-Diable weather station in north-eastern Canada. The forty years of daily temperature records available at the station are split into a training data set (first 30 years) and a validation data set (remaining 10 years). 14 large-scale variables from the NCEP reanalysis (*Kalnay et al.*, 1996) from the grid box closest to the station serve as potential predictors.
For comparison the Statistical Downscaling Model (SDSM) by *Wilby et al.* (2002) has been applied to the same downscaling problem. The SDSM combines regression based methods, weather typing and parts of a stochastic weather generator. Dependent on its setup the SDSM first carries out a regression between large-scale circulation patterns and local-scale variables and second uses stochastic approaches to artificially increase the variance of the downscaled series to better match the variance of the observations.
For the described task GP has performed slightly better than the SDSM with an RMSE of 3.59°C (GP) compared to 4.07°C (SDSM) for downscaling daily maximum temperature and an RMSE of 4.57°C (GP) compared to 5.15°C (SDSM) for daily minimum temperature. Furthermore, the GP solutions contain only two of the 14 suggested predictor variables, namely daily mean temperature and 500 hPa geopotential height to predict daily maximum temperature, and daily mean temperature and vorticity at 500 hPa height to predict daily minimum temperature. The SDSM requires six predictors for each predictand variable (cf. Table 4.2).
Hashmi et al. (2011) have applied Gene Expression Programming (GEP), a variant of GP, for downscaling daily precipitation at the Clutha River watershed in the South Island of New Zealand. As no significant correlation between large scale predictors and precipitation at single sites has been found, the areal average precipitation over 23 weather stations distributed

Table 4.2.: Results of the studies on GP based downscaling by *Coulibaly* (2004) and by *Hashmi et al.* (2011).

Study	**Predictand**		**RMSE**		**No. of Predictors**	
			GP/GEP	SDSM	GP/GEP	SDSM
Coulibaly (2004)	T_{max}	training	3.54°C	-	2	6
		testing	3.59°C	4.07°C		
	T_{min}	training	4.65°C	-	2	6
		testing	4.57°C	5.14°C		
Hashmi et al. (2011)	P	training	5.23 mm	5.61 mm	7	10
		testing	5.35 mm	6.03 mm		

over the catchment is considered as predictand variable. As in *Coulibaly* (2004) the observational record is split into a 30 year training period and a 10 year validation period. 26 different variables from the NCEP reanalysis taken at the gridbox closest to the catchment serve as predictors.
The results of the GEP based downscaling have been again compared to the performance of the SDSM. GEP has performed slightly better than the SDSM with an RMSE of 5.23 mm (GEP) compared to 5.61 mm (SDSM) for the training period and an RMSE of 5.35 mm (GEP) compared to 6.03 mm (SDSM) for the validation period. As in *Coulibaly* (2004) the evolutionary approach has required less predictor variables than the SDSM (cf. Table 4.2).
Liu et al. (2008) employ evolutionary polynomial regression (based on GP), an artificial neural net and the SDSM to the downscaling of daily maximum and minimum temperature and precipitation at the Chute-du-Diable catchment and come to similar conclusions as the above studies. The evolutionary technique and the neural net have performed about equally well and both have performed better than the SDSM.
Hassanzadeh et al. (2013) take a different approach utilizing GP for the quantile-based downscaling of intensity-duration-frequency (IDF) curves for the city of Saskatoon in Canada. IDF curves for precipitation are commonly used in engineering, planning and design to provide information on the frequency, duration and intensity of extreme events. The authors have concluded that Genetic Programming is a promising tool for extracting mathematical mappings between extreme rainfall quantiles at GCM scales and local scales.
Pour et al. (2014) have presented a first study on downscaling of extreme precipitation events by means of GP. The authors have predicting days exceeding the 90% quantile as well as the yearly maximum numbers of consecutive wet and dry days on the East Coast of Malaysia. The performance of the GP models has been compared to an ANN and the SDSM with GP yielding overall favorable results.
The application of multi-objective Genetic Programming (MOGP) to the downscaling of GCM output (in our application ERA-Interim reanalysis data) is described and discussed in Chapter 6. We consider both temperature (daily maximum, minimum and mean) and daily accumulated precipitation at 86 stations distributed over Europe. Considering multiple stations helps to assess the generalizeability of the GP results as the skill of downscaling methods depends on the considered stations. Further, we downscale local daily accumulated precipitation at the weather stations without applying any spatial averaging. The results are assessed not only on the basis of te RMSE, but also considering distribution-based measures, spatial and temporal correlation and the representation of extremes.

5

The Multi-Objective Genetic Programming Downscaling Methodology

We employ Multi-Objective Genetic Programming (MOGP) to the downscaling of atmospheric data. MOGP serves as a statistical downscaling technique to establish statistical relations (in the following referred to as downscaling rules) between predictors and predictands. The detection of downscaling rules is treated as a symbolic regression problem (cf. Sec. 2.3). Using GP one does not need to prescribe the structure of the regression model. GP evolves both model structure and model parameters and thereby allows to account for potential non-linear and/or multivariate predictor-predictand relations.
By only minimizing the root mean square error (RMSE) downscaling rules would aim at predicting the expected value of the predictand given the predictors (when assuming the residuals to be normally distributed). Such estimators are known to have a too small variance (e.g., *Hastie et al.*, 2009). We seek downscaling rules returning realizations from the respective (unknown) multivariate probability density function (PDF). Therefore we do not optimize solely w.r.t. the RMSE, but also concerning other parameters such as spatial or temporal variance and/or PDF. We thus deal with multiple objectives.

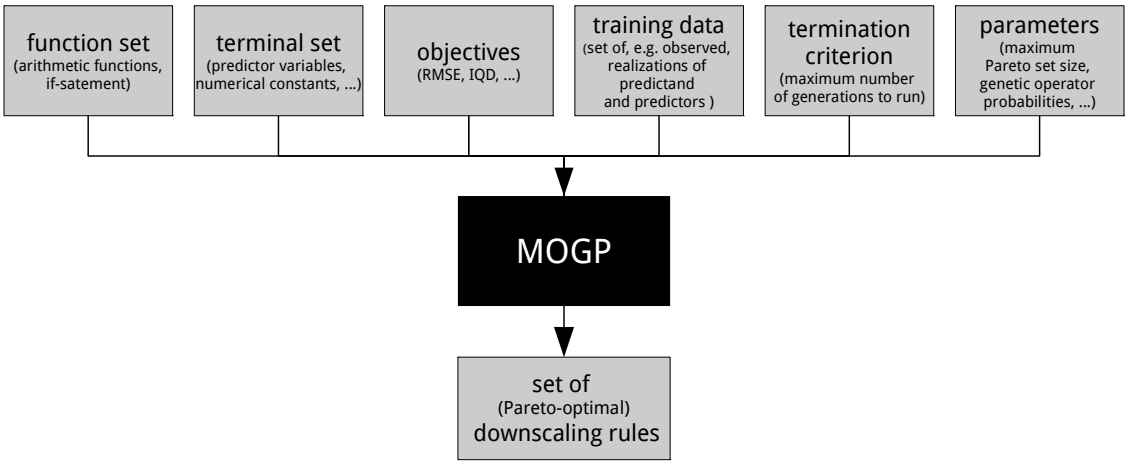

Figure 5.1.: Illustration of input and output to multi-objective Genetic Programming for the detection of downscaling rules.

When we formulate the downscaling problem as a multi-objective optimization problem, we face some difficulties. Scaling the objectives is difficult since we usually not a priori know the minimum achievable value for each objective. Thus, one would need to decide on the relative importance of the different objectives in advance. Further, 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 the 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 and ranking of the different models of a generation rather than on values of the objectives alone. The result is a finite set of so called Pareto optimal models (downscaling rules) which allows the user to select a model based on a trade-off between the different objectives in hindsight.
The input to be provided by the user to run MOGP for the discovery of downscaling rules is illustrated in Fig. 5.1. The user provides:

- the function set - in the following the arithmetic functions and an if statement,
- the terminal set - in the following mostly the predictor variables and a set of numerical constants,
- training data - in the following a set of realizations of predictand and predictor variables from models or observations,
- a termination criterion - in the following reaching a number of generations defined in advance,
- run parameters (genetic operators and their probabilities, population size, maximum Pareto set size).

Detailed settings (predictors etc.) for the two different MOGP applications following this chapter are provided in the respective chapters. For the first application, the downscaling of mesoscale near-surface atmospheric fields, the MOGP setup is described in Section 6.2. For the second application, the downscaling of climate reanalysis data to local stations, the MOGP setup is described in Section 7.2.
The current chapter is structured as follows. After introducing the concept of Pareto Optimality, the multi-objective Genetic Programming (MOGP) algorithm is described. Fitness assignment and clustering procedure are explained and we walk through MOGP step by step. The downscaling methodology is part of *Zerenner et al.* (2016). Our GP code is in large parts based on the GPLAB (*Silva and Almeida*, 2003), a Genetic Programming toolbox for MATLAB.

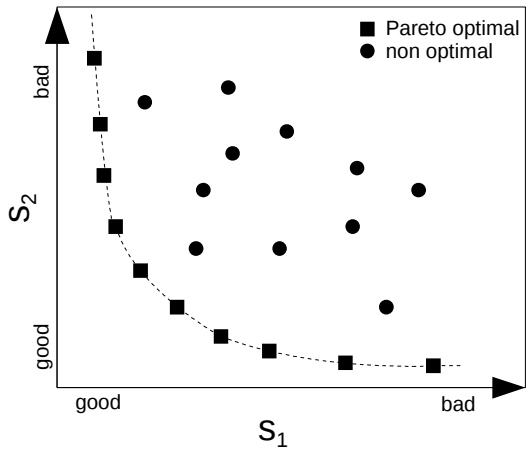

Figure 5.2.: Example of a minimization problem with two objectives (s_1 and s_2). The squares correspond to the Pareto optimal solutions; the circles to the non-optimal solutions. The dashed line is the so-called Pareto front. The term Pareto front is often used synonymously for the set of Pareto optimal solutions.

5.1. Pareto Optimality

The term Pareto optimality originates in economics. The state of an economic system is called Pareto optimal when economic resources are distributed in such a way that it is impossible to improve the situation of one person without deteriorating the situation of at least one other person.

For optimization problems that involve multiple, sometimes conflicting objectives there is no single optimal solution. Usually there exists a set of alternative solutions in which no solution is optimal in the sense that it is superior to all other solutions for all objectives. The solutions which are optimal in the sense that there exists no other solution which is better concerning all objectives are called Pareto optimal.

Figure 5.2 shows an example of a minimization problem with two objectives s_1 and s_2. The squares constitute the set of Pareto optimal solutions as for each square there is no solution that is better with respect to both objectives (i.e., no solution with smaller s_1 and s_2). The circles correspond to the non-optimal solutions. For each circle exists at least one solution that is better with respect to both objectives.

The (multiple) objectives correspond to different quality criteria of the desired solution. We denote the objective space containing all objective functions as $\mathcal{O}$ and the solution space containing all potential solutions as $\mathcal{Q}$. An objective $s_i \in \mathcal{O}$ is calculated by comparing prediction (downscaling) $\mathbf{y}^D$ and reference $\mathbf{y}^R$. The prediction results from the solution (downscaling rule) $\alpha \in \mathcal{Q}$ applied to the predictors $\mathbf{x}$. Thus, incorporating all dependencies, we can write $s_i(\mathbf{y}^D, \mathbf{y}^R) = s_i(\alpha, \mathbf{x}, \mathbf{y}^R)$. For simplicity in the following we only include the dependency on the solution α explicitly.

Be $\mathbf{s}(\alpha) = (s_1(\alpha), s_2(\alpha), ..., s_m(\alpha))^T$ the objective vector, i.e., the vector containing all m objectives. Let us consider two solutions $\alpha, \beta \in \mathcal{Q}$. The solution α is said to dominate β

$(\alpha \succ \beta)$ if and only if

$$\begin{aligned} &\forall i \in \{1, 2, ..., m\} : s_i(\alpha) \leq s_i(\beta) \\ &\wedge \exists j \in \{1, 2, ..., m\} : s_j(\alpha) < s_j(\beta). \end{aligned} \tag{5.1}$$

Or in words, α dominates β if α is at least as good as β with respect to all objectives and there exists at least one objective for which α is better than β. The solutions that are not dominated by any element in the solution space $\mathcal{Q}$ are Pareto optimal.
A solution α is said to cover β $(\alpha \succeq \beta)$ if $\alpha \succ \beta$ or $\mathbf{s}(\alpha) = \mathbf{s}(\beta)$, i.e., either α dominates β or they both perform equally well concerning all objectives.

5.2. MOGP Algorithm

The multi-objective fitness assignment according to SPEA requires two main changes compared to traditional GP. Firstly, each generation is split into two sets called populations $\mathcal{P}$ and $\mathcal{P}'$. The population $\mathcal{P}$ is evolving over time as in traditional GP, whereas the second population $\mathcal{P}'$, the so-called Pareto set, contains all Pareto optimal solutions. Secondly, the fitness calculation for individuals in both populations $\mathcal{P}$ and $\mathcal{P}'$ is based on a comparison between the individuals (i.e., based on the number of individuals dominated by or dominating a solution) rather than on the absolute performance.

Fitness Assignment

In SPEA the fitness assignment consists of two steps (cf. Fig 5.3):

(1) To each solution in the Pareto set $\alpha \in \mathcal{P}'$ a real value called fitness $f'(\alpha) \in [0, 1)$ is assigned. The better the performance of a downscaling rule, the smaller the fitness. The fitness $f'(\alpha)$ is proportional to the number n of individuals $\beta \in \mathcal{P}$ that are covered by α, i.e., $\alpha \succeq \beta$. Let N be the total number of individuals in $\mathcal{P}$. Then $f'(\alpha)$ is defined as

$$f'(\alpha) = \frac{n(\beta | \alpha \succeq \beta, \alpha \in \mathcal{P}', \beta \in \mathcal{P})}{N + 1}. \tag{5.2}$$

To clearly separate between the fitness of individuals in $\mathcal{P}$ and $\mathcal{P}'$, the fitness of the individuals in $\mathcal{P}'$ is also called strength, hence the name *Strength Pareto Evolutionary Algorithm* (in the following also referred to as Strength Pareto Approach).

(2) The fitness $f(\beta)$ of an individual in the population $\mathcal{P}$ $(\beta \in \mathcal{P})$ is calculated as the sum over the fitness of all individuals in the Pareto set $\alpha \in \mathcal{P}'$ that cover β, i.e.,

$$f(\beta) = 1 + \sum_{\alpha, \alpha \succeq \beta} f'(\alpha), \tag{5.3}$$

where $f(\beta) \in [1, N) \subset \mathbb{R}$. One (1) is added to the sum to ensure that the individuals in the Pareto set $\mathcal{P}'$ have better fitness than those in $\mathcal{P}$.

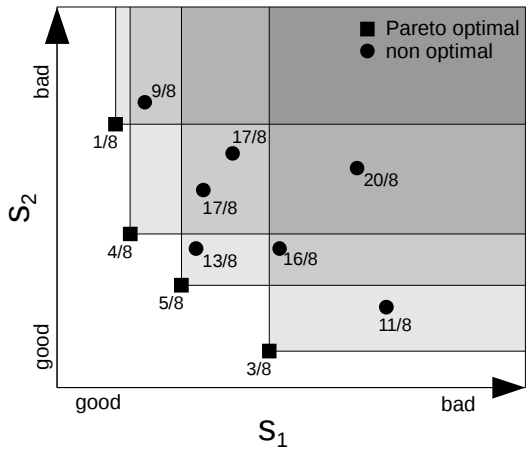

Figure 5.3.: Example of a minimization problem with two objectives (s_1 and s_2). The squares correspond to the Pareto optimal solutions; the circles to the non-optimal solutions. The number associated with each solution gives the fitness according to the Strength Pareto Approach. The darker the area a non-optimal solution is located in, the more solutions from the Pareto set dominate it, hence the worse it is ranked. The figure is adapted from *Zitzler and Thiele* (1999).

Figure 5.3 illustrates one possible scenario of a minimization problem with two objectives s_1 and s_2. The lowest point in Fig. 5.3 shows an individual contained in the Pareto set $\mathcal{P}'$ that dominates 3 out of the 7 individuals in $\mathcal{P}$. Therefore its fitness f' equals $3/(7+1) = 3/8$. The next lowest point represents an individual from the population $\mathcal{P}$ which is dominated only by one individual with a fitness f' of $3/8$. Hence its fitness f calculates as $1 + 3/8 = 11/8$.

Clustering Procedure

If no constraint is applied, the Pareto set can grow very large. In SPEA a clustering procedure is applied when the size of the Pareto set exceeds a given limit. Clustering algorithms group individuals into clusters according to their similarity. We use an agglomerative hierarchical clustering (e.g., *Hastie et al.*, 2009), where initially each individual represents its own cluster. The individuals (or later clusters) which are closest according to some distance metric (we use the Euclidian distance in the objectives space) are stepwise grouped together to form new, larger clusters. The clustering procedure is stopped when the desired number of clusters, i.e., the maximum size of the Pareto set, is reached. From each cluster, the member closest to the center of the respective cluster is to become part of the new, pruned Pareto set.
To assure that all objectives are considered equally the values are normalized before the clustering is applied. Since our objectives s_i decrease with better fit, the objectives are scaled via

$$s_i^{sc}(\alpha) = \frac{s_i(\alpha) - \min\limits_{\beta \in P'}(s_i(\beta))}{\max\limits_{\gamma \in P'}(s_i(\gamma))}. \tag{5.4}$$

That is, for each objective s_i we first subtract the minimum value occurring in the current Pareto set from all values. And second, from the new values we take the maximum value of s_i and divide all values by that maximum. The resulting s_i^{sc} are then restricted to the interval $[0, 1]$. Since fitness results from a greater than/smaller than comparison, scaling does not

effect the fitness. Note that scaling is only applied for the clustering. In contrast to randomly drawing a number of individuals a clustering procedure helps to preserve the diversity of the solutions while shrinking the Pareto set.

The Algorithm

Figure 5.4 shows a basic flowchart of the GP algorithm incorporating fitness assignment according to SPEA, which is in the following explained step by step.

(1) An initial population $\mathcal{P}$ of individuals (downscaling rules) is generated. The initial population is created randomly using the *Ramped Half-and-Half method*[1] as implemented in the GPLAB.

(2) Each individual is applied to the training data set.

(3) From the result of (2) the objectives are calculated.

(4) The Pareto set $\mathcal{P}'$ is updated: All individuals in population $\mathcal{P}$ that are not dominated within $\mathcal{P}$ are moved to the Pareto set $\mathcal{P}'$. The individuals in $\mathcal{P}'$ that are covered by another member of $\mathcal{P}'$ are removed. In case the number of individuals stored in $\mathcal{P}'$ exceeds the allowed maximum, the number of individuals in $\mathcal{P}'$ is reduced by hierarchical clustering (see Clustering Procedure).

(5) The fitness of each individual in $\mathcal{P}$ and $\mathcal{P}'$ is calculated by comparing the individual's performances (see Fitness Assignment).

(6) If the stopping criterion is met, the final Pareto set is returned. If the stopping criterion is not met, the algorithm continues with (7).

(7) The next generation is created by combining and mutating individuals from the current $\mathcal{P} + \mathcal{P}'$. The creation of the new generation consists of two steps. First, a sampling procedure is applied to determine the parents. Second, genetic operators (crossover, mutation) are applied to create new individuals.

   (a) For sampling we use the lexicographic parsimony pressure (*Luke et al.*, 2002) as it is implemented in GPLAB. A number of individuals is randomly drawn from the current $\mathcal{P} + \mathcal{P}'$. The individual drawn with the best fitness is to become parent. In case several individuals are equally fit the smallest one, i.e., the one consisting of the least number of nodes, is chosen.

[1] In the *Ramped Half-and-Half* method an equal number of individuals are initialized for each depth between two and the maximum allowed tree depth. For each depth, half of the individuals are initialized using the *Full* method, the other half using the *Grow* method. The parse trees are generated iteratively starting with the root note (i.e., the single node at the first level of a parse tree). In the *Full* method non terminal nodes are added until the maximum tree depth is reached and only the last level contains terminal nodes. In the *Grow* method each node, except for the root node, is randomly chosen between terminals and non-terminals. The population resulting from the *Ramped Half-and-Half* initialization is thus very diverse in size.

(b) The genetic operators are applied as follows: Crossover recombines two parents. The parent parse trees are cut at randomly chosen nodes and the separated subtrees are exchanged. (Subtree)-mutation cuts a randomly chosen subtree from the parent and replaces it by a new randomly created subtree. Parent selection and application of genetic operators are repeated until the new generation is full, i.e., until the population size defined in the settings is reached. For a detailed explanation see Section 2.1.

Starting from (2) the succeeding steps are iteratively repeated until the stopping criterion is met (see (6)).

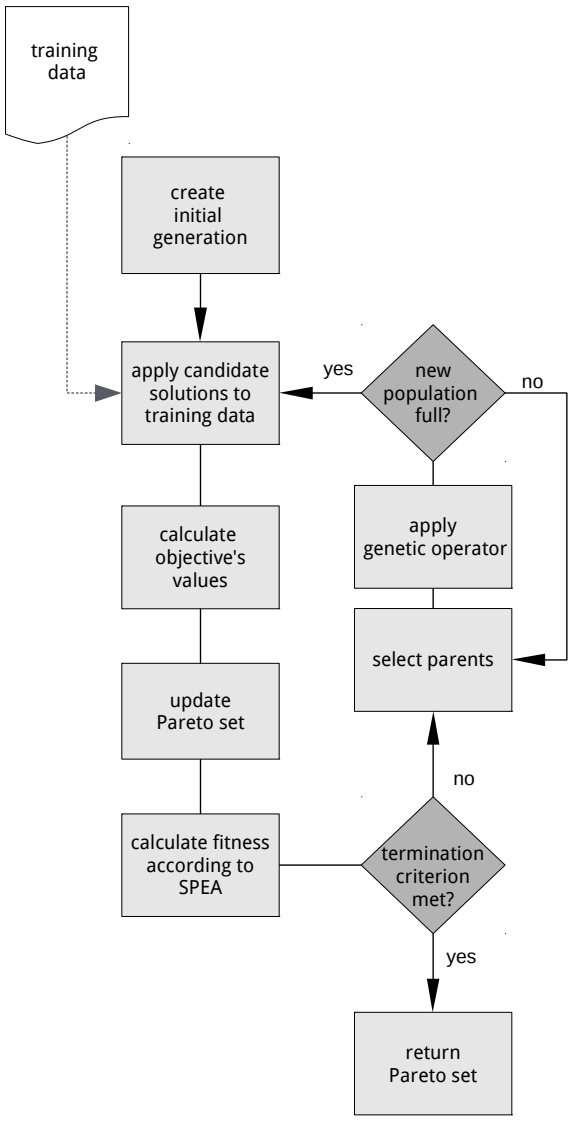

Figure 5.4.: Flowchart showing the essential steps of Genetic Programming with multi-objective fitness assignment according to the Strength Pareto Evolutionary Algorithm by *Zitzler and Thiele* (1999).

6 Downscaling Mesoscale Near-Surface Fields using MOGP

This chapter describes the application of MOGP to the downscaling of atmospheric mesoscale fields near the land surface from 2.8 km to 400 m grid spacing. This is aimed at an improved downscaling scheme for implementation in the TerrSysMP (*Shrestha et al.*, 2014). Thus, we aim to downscale all variables that are transferred from the coarse atmospheric component model COSMO to the higher-resolution land surface and subsurface component models CLM and ParFlow (cf. Sec. 3.3): pressure, temperature, specific humidity and horizontal wind speed at the lowest COSMO model layer in $\approx$ 10 m height, incoming longwave radiation and direct and diffuse shortwave radiation at the land surface, and finally instantaneous precipitation.
An atmospheric downscaling as described in this study can pursue different goals. Ideally it leads to an improved representation of mass and energy fluxes, but it can also enhance the representation of threshold dependent processes, such as runoff and snow melt, within fully-coupled simulations as well as in stand-alone land surface/subsurface simulations. Further, the downscaled near-surface fields might also offer valuable input for agricultural models for which for instance a good representation of night frost is important.
As in *Schomburg et al.* (2010) we first smoothen the coarse predictand field using a biquadratic spline-interpolation (cf. Sec. 3.2.2). Second, the anomalies are predicted by MOGP derived downscaling rules and added to the smooth predictand field to obtain the final downscaled field (cf. Fig. 6.1). The downscaling rules in this chapter use the coarse atmospheric model output and quasi-static high-resolution land surface information as predictors. That is, we

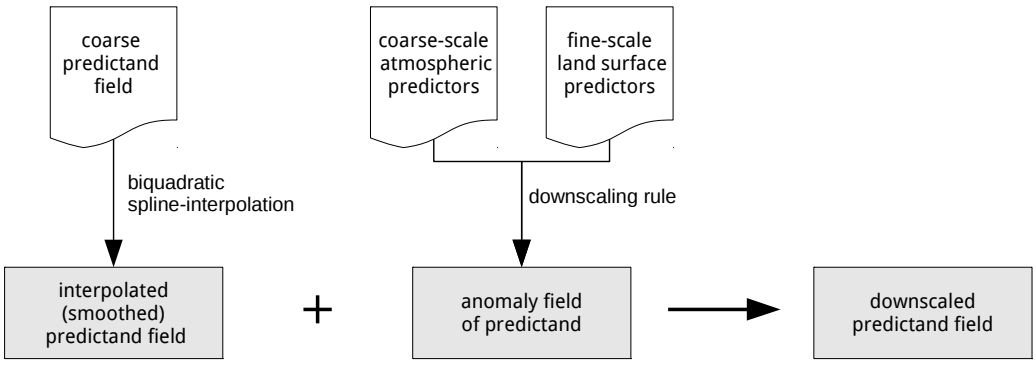

Figure 6.1.: Sketch of the full downscaling concept.

assume that the structure of the atmospheric boundary layer near the surface is significantly influenced by land surface heterogeneity.
Even with a flexible regression approach like MOGP the downscaling rules are much less complex and computationally a lot cheaper than running the full atmospheric model in 3D at a high resolution. Downscaling rules can never reproduce the exact high-resolution references; due to turbulence for instance, there will always be a component of the fine-scale fields that cannot be reconstructed.
To train the downscaling rules a training data set has been created from high-resolution COSMO model simulations with a horizontal grid spacing of 400 m. Model setup and simulation periods are introduced in Section 6.1. The MOGP setup is described in Section 6.2. Results are presented in Section 6.3 and discussed in Section 6.4.

6.1. Data

6.1.1. Setup of the 400 m COSMO Simulations

The downscaling algorithm by *Schomburg et al.* (2010) has been designed for downscaling from 2.8 km to 400 m grid spacing. In order to develop and validate the downscaling algorithm *Schomburg et al.* (2010) have carried out high-resolution COSMO simulations with a grid spacing of 400 m. In this work the same scales are considered, i.e., we aim at a downscaling by a factor of 7. For the first tests of MOGP the data from the high-resolution simulations from *Schomburg et al.* has been used (cf. Appendix A). These have later been replaced by new high-resolution simulations with the most recent version of the COSMO model (COSMO 5.3).
A grid spacing of 400 m is considerably smaller than the highest operationally used grid spacing of 2.8 km for COSMO-DE. Thus, for the high-resolution simulations a high-resolution external parameter set for the land surface is required and some configurations in the model setup need to be adapted. In the following the setup of the 400 m simulations is described focusing on the differences compared to the operational COSMO-DE. The model setup corresponds largely to the setup of *Schomburg et al.* (2010)

Domain

The model domain covers an area of 168 × 168 km (i.e., 420 × 420 grid points at a 400 m grid spacing) and covers the western part of North Rhine-Westphalia as well as parts of Luxembourg, Belgium and the Netherlands (Fig. 6.2). The domain is centered over the two small rivers Rur and Erft, which is the main investigation area of the Transregional Collaborative Research Centre 32 (*Simmer et al.*, 2015).

External, Initial and Boundary Data

For a grid spacing of 400 m the operationally used surface data from DWD is too coarse. *Schomburg et al.* have created an external data set with 400 m grid spacing.

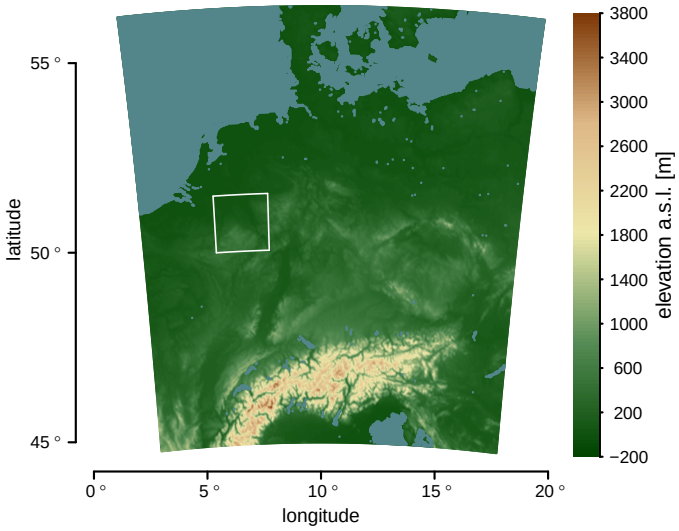

Figure 6.2.: Location of the 168 × 168 km domain of the COSMO simulations with 400 m grid spacing (white box) within the operational COSMO-DE domain.

- **Topographic height** information has been obtained from the Shuttle Radar Topography Mission (*Farr et al.*, 2007). The original data set has a grid resolution of 90 m and has been averaged to 400 m. In a boundary area, 20 km around the edges, the orography has been slightly smoothed to ensure a gradual transition to the coarser COSMO-DE boundary forcing. The domain contains flat areas in the north and northwest as well as mountainous regions in the south and the northeast (cf. Fig. 6.3(a)).

- **Soil type** information has been obtained by merging data sets from North Rhine-Westphalia (IS BK50) and Rhineland-Palatinate (BÜK200). The soil type classes have been mapped onto soil type categories used in the COSMO model. For the western part of the domain, which covers parts of the Netherlands, Luxembourg and Belgium, no high-resolution soil information was available. Thus here the information from the coarser operationally used external data set of the DWD has been used. The dominant soil type for the domain is loam followed by sand and sandy loam (cf. Fig. 6.3(b)).

- **Land use** characteristics are obtained from the CORINE data set (*EEA*, 2000).The original data has a grid spacing of 100 m and distinguishes 44 different land use classes. The data has been upscaled to 400 m grid spacing according to the dominant land cover class and contains 29% forests, 22% irrigated arable land, 14% pastures, 13% urban fabric areas, and 11% complex cultivation patterns (cf. Fig. 6.3(c))

The remaining external parameter fields **fraction of land, plant cover, leaf area index, roughness length, root depth** and the **soil texture parameters** for TERRA have been determined from topography, soil type and land use information mainly via look-up tables (for details see *Schomburg* (2011)).

Initial and lateral boundary data for the simulations are taken from the COSMO-DE analyses of the DWD. Analyses runs use data assimilation schemes, which nudge the model state

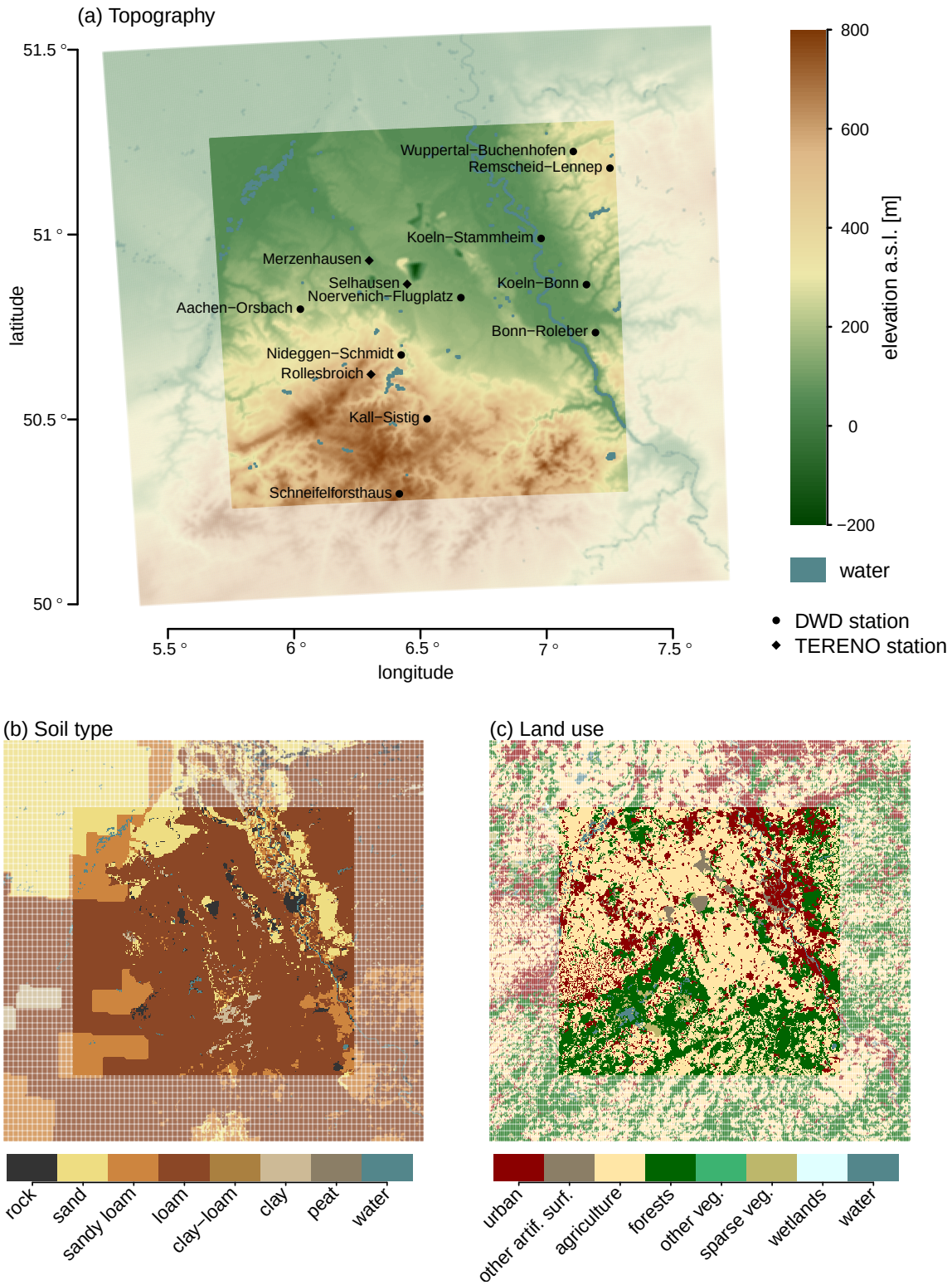

Figure 6.3.: The domain of the 400 m simulations in detail. The full domain covers 168 × 168 km (i.e., 420 × 420 grid points). To exclude boundary effects only the inner 280 × 280 grid points are used for training and validating the downscaling rules. The excluded area is indicated by the transparent margin areas in the figures. (a) shows the topography and the location of DWD and TERENO stations (cf. Figs. 6.5). (b) shows the high-resolution soil type information (from IS BK50 and BÜK200) and (c) shows the high-resolution land cover information (from CORINE).

towards the measurements. At a grid resolution of 400 m broad rivers such as the Rhine are resolved as connected water pixels. Water pixels require the attribution of a water temperature. In COSMO water temperature is typically estimated from the sea surface temperature. As the domain of the 400 m simulations does not contain ocean, the river temperatures are assigned according to the mean annual cycle of the water temperature as measured near Karlsruhe.

Model Configuration

For the 400 m simulations some settings have do be altered compared to the operational COSMO-DE setup(Table 6.1). The time stepping is adjusted from 25 sec to 4 sec to obey the Courant-Friederich-Levy (CFL) stability criterion. The increase of the grid resolution leads to a major increase of computation time. The vertical grid spacing has not been altered (i.e., 51 vertical layers as in COSMO-DE).

Turbulence Turbulence parameterizations are typically developed for certain scales and are not readily applicable to other scales. Usually the horizontal grid spacing is large compared to the vertical grid spacing and it is thus assumed, that the turbulent exchange is much larger in the vertical than in the horizontal. Horizontal exchange coefficients are often neglected. With a grid spacing of a few hundred meters in the horizontal it is questionably if the assumption of negligible horizontal turbulent exchange is still valid.
In COSMO different turbulence schemes can be used. The default is a prognostic TKE scheme which neglects the horizontal turbulent exchange coefficients. An alternative scheme has been developed for very high-resolution simulations ($\approx$ 100 m grid spacing) with the Litfass Local Model (LLM), a variant of the Local Model (LM - the predecessor of the COSMO model). The LLM model has been set up in the framework of the LITFASS project of DWD (*Beyrich et al.*, 2002). The LLM turbulence scheme estimates the horizontal exchange coefficients from the vertical coefficients by scaling them according to the horizontal grid spacing. Thus, the LLM scheme automatically adapts to the model resolution while the default TKE-scheme has tuning parameters, such as the maximum turbulent length scale, and is thus not straightforwardly adaptable to higher resolutions. For these reasons *Schomburg* (2011) used the LLM-scheme for the 400 m COSMO simulations.
In the latest COSMO versions (COSMO 5.1 and beyond) the TKE-scheme has been improved

Table 6.1.: Summary of adjustments made to set up the high-resolution COSMO simulations (compared to the operational COSMO-DE setup).

Setting (Namelist Parameter)	**COSMO-DE**	**COSMO-400m**
grid spacing	2.8 km	400 m
model time step	25 s	4 s
spatial averaging for rad. calc. (lrad_avg)	yes (TRUE)	no (FALSE)
turbulence parametrization (itype_turb)	TKE-scheme (3)	LLM-scheme (7)
3 dimensional turbulence (l3dturb)	no (FALSE)	yes (TRUE)

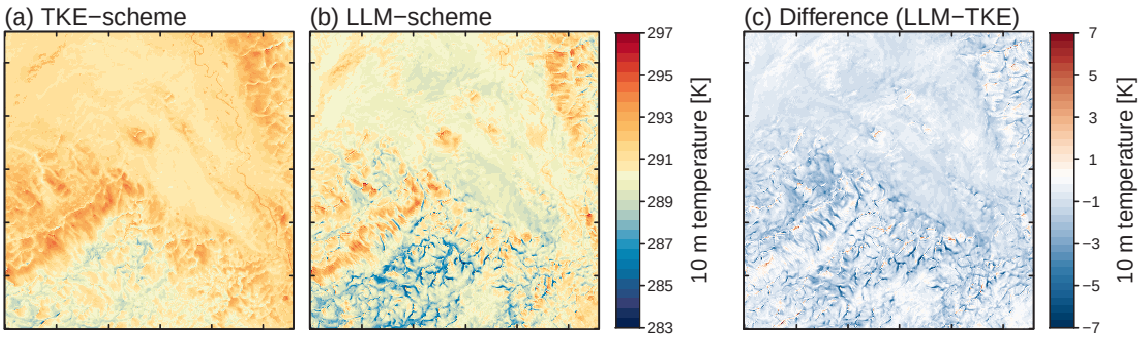

Figure 6.4.: Simulated 10 m temperature fields on July 18th 2014 at 3:00 UTC with (a) the standard TKE-scheme and (b) the LLM-scheme for turbulence. (c) shows the difference between both.

and can now be used as a 3D scheme, which improves model results mainly at horizontal resolutions below 1 km grid spacing. In order to find the suitable turbulence scheme for our simulations, we have simulated two test cases with the new 3D TKE-scheme and the 3D LLM-scheme. Though both schemes consider three dimensional turbulence, the simulations differ especially for the near surface temperature during clear sky nights (Fig. 6.4). The LLM-scheme simulates generally lower nighttime temperatures than the TKE-scheme while the spatial patterns of the temperature fields are similar; both fields show channel structures formed by cold air draining into the valleys due to the stable atmospheric layering, which is typical of nights with strong radiative cooling. However, the simulation with the LLM-scheme gives up to 5 K lower temperatures in the valleys of the Eifel than the simulations with the TKE-scheme.

We have compared the simulated temperatures against measurements from 10 DWD and 3 TERENO stations located within the model domain (for station locations see Fig. 6.3). The time series at four selected stations are shown in Fig. 6.5. "Noervenich Flugplatz" is located in flat terrain. "Kall-Sistig" is located at a hillside. "Schneifelforsthaus" is located near the peak of a mountain and "Wuppertal Buchenhofen" is located in a valley (between two low mountain ranges).

The differences in the 2 m temperature between TKE- and LLM-scheme are rather small compared to the differences to the observations. The range of the diurnal cycle of the 2 m temperature is generally underestimated by the simulations compared to the observations. The difference is most pronounced at the stations in the lower altitudes (cf. Fig. 6.5(a) and (d)). For the stations at higher altitudes the simulations match the observations more closely (cf. Fig. 6.5(b) and (c)). For all four stations the simulated daily maximum temperature is to small compared to the observations. Daily observed minimum temperatures are better captured by the simulations. However, for the stations at low altitudes the minimum temperatures tend to be overestimated compared to the observations. For the minimum temperature the LLM-scheme is in general closer to the observations than the TKE-scheme. For the daytime temperature the TKE-scheme appears to be closer to the observations though only marginally.

Based on these comparisons and the reasoning in *Schomburg* (2011) we use the LLM-scheme. The more reasonable representation of near-surface temperatures during clear sky nights is

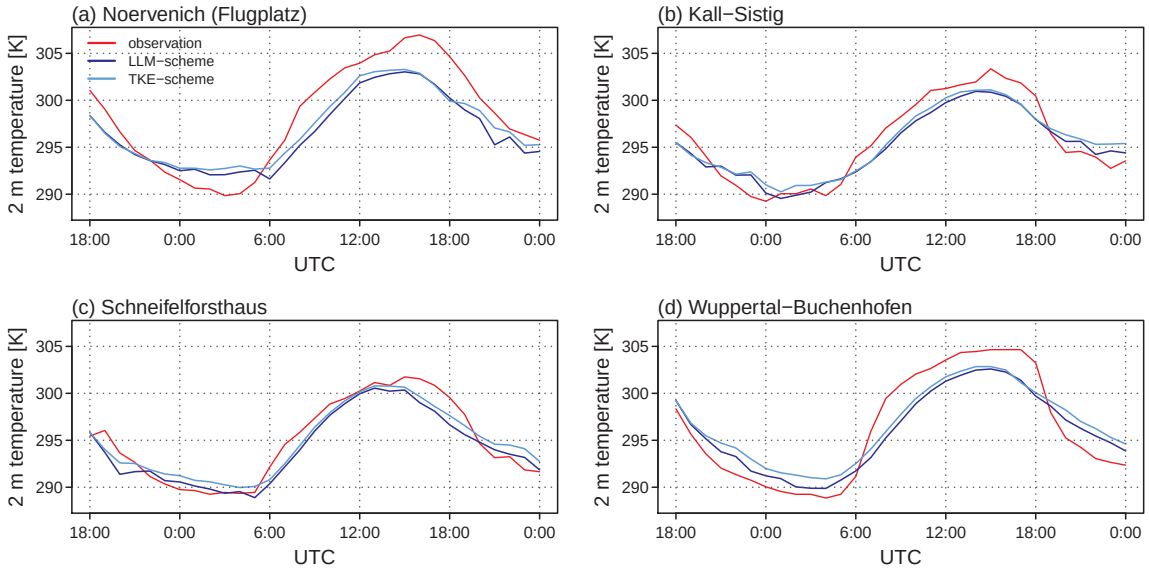

Figure 6.5.: 2 m temperature from July 17th 2014 18:00 UTC to July 19th 2014 0:00 UTC. The red line shows station measurements. The blue lines show the simulation results at the grid point closest to the respective station either with the TKE-scheme (light blue) or with the LLM-scheme (dark blue). The location of the stations is shown in Fig. 6.3.

potentially important for training the downscaling scheme as the resulting channel structures lead to very pronounced fine-scale variability and thus comparatively large temperature anomalies.

Radiation In COSMO-DE radiation calculations are carried out every 15 min. Since clouds can cross several 400 m grid boxes in few minutes the radiation calculations are carried out every 3 minutes in the 400 m simulations. Moreover, radiation calculations are carried out for every single grid box and not averaged over grid-boxes as in COSMO-DE. As the radiation parameterization is computationally expensive also these changes contribute significantly to the increase of computing time. In the 400 m simulations a model time step calling the radiation scheme takes approximately three times as long as a model time step not calling the radiation scheme.

6.1.2. Simulation Periods

In total 30 two-day periods have been simulated, half for training the downscaling rules, half for validation. Instead of using the (hourly) output for the complete 48 hour simulations, single time steps have been extracted for training and validation to reduce computational cost (cf. Table 6.2).

Table 6.2.: Dates and prevailing weather conditions of the 400 m COSMO simulations. The full data set contains 30 simulation periods of two days each. The data has been split into two parts of 15 simulations each. The first part is used for training, the second for the validation of the downscaling rules. The three right hand columns list the time steps extracted from the data to reduce computational cost. For pressure p, temperature T, specific humidity q and wind speed w_h at 10 m height, as well as incoming longwave radiation at the surface LWR the same time steps have been extracted. For incoming direct and diffuse shortwave radiation SWR_{dir}, SWR_{dif} and precipitation P only time steps where the respective variable is unequal zero for at least one grid box are selected.

(a) **Training**

Date	**Weather**	$p,\ T,\ q,\ w_h,\ LWR$	$SWR_{dir},\ SWR_{dif}$	P
11/12 Mar 2013	snow	12-02:00, 12-15:00	11-16:00, 12-15:00	11-08:00, 11-10:00, 12-12:00, 12-12:00
26/27 Jan 2014	clouds, precipitation	27-02:00, 27-14:00	26-08:00, 27-14:00	26-21:00, 26-23:00, 27-02:00, 27-03:00
01/02 May 2014	wind, clouds, heavy local precipitation	01-13:00, 02-22:00	01-13:00, 02-18:00	01-18:00, 01-19:00, 01-22:00
11/12 May 2014	strong winds, convective precipitation	11-12:00, 12-21:00	11-12:00, 12-19:00	11-09:00, 12-11:00, 12-21:00
28/29 Sept 2014	fair weather conditions	28-05:00, 29-10:00	28-07:00, 29-10:00	-
18/19 Oct 2014	fair weather conditions	18-20:00, 19-01:00	18-15:00, 19-09:00	-
03/04 Nov 2014	stratiform precipitation	03-18:00, 04-08:00	03-15:00, 04-08:00	04-07:00, 04-08:00, 04-20:00
01/02 Jan 2015	wind, clouds and precipitation	02-03:00, 02-16:00	02-10:00, 02-15:00	01-22:00, 02-03:00, 02-04:00
30/31 Jan 2015	wind, clouds and precipitation	30-23:00, 31-07:00	30-12:00, 31-13:00	30-23:00, 31-00:00, 31-07:00, 31-09:00
12/13 Feb 2015	mostly clear sky	12-11:00, 13-19:00	12-11:00, 13-16:00	-
08/09 Mar 2015	first clear sky, later clouds, weak precip.	08-09:00, 09-04:00	08-09:00, 09-07:00	-
01/02 July 2015	clear sky, hot	02-03:00, 02-15:00	02-09:00, 02-15:00	-
10/11 July 2015	clear sky, hot	10-06:00, 11-00:00	10-06:00, 11-18:00	-
25/26 July 2015	stratiform precipitation, strong winds	25-09:00, 26-17:00	25-09:00, 26-17:00	25-09:00, 26-15:00, 26-17:00
04/05 Aug 2015	stratiform precipitation	04-17:00, 05-07:00	04-17:00, 05-07:00	04-10:00, 04-12:00, 04-14:00

(b) **Validation**

Date	**Weather**	p, T, q, w_h, LWR	SWR_{dir}, SWR_{dif}	P
06/07 May 2014	stratiform precipitation	06-18:00, 07-08:00	06-15:00, 07-08:00	06-22:00, 07-03:00, 07-04:00
17/18 Sept 2014	fair weather conditions	17-10:00, 18-01:00	17-15:00, 18-09:00	02-02:00, 02-07:00, 02-08:00
22/23 May 2015	partly cloudy, weak wind	22-09:00, 23-04:00	22-09:00, 23-07:00	-
04/05 June 2015	fair weather conditions	04-06:00, 05-00:00	04-06:00, 05-18:00	-
12/13 June 2015	convective clouds and precipitation	12-12:00, 13-21:00	12-12:00, 13-19:00	12-12:00, 12-15:00, 12-17:00, 12-19:00
12/13 July 2015	stratiform (later convective) precip.	13-03:00, 13-06:00	13-10:00, 13-12:00	12-23:00, 13-00:00, 13-07:00
17/18 July 2015	clear sky	18-03:00, 18-15:00	18-09:00, 18-15:00	-
01/02 Aug 2015	weak winds, some clouds later clear sky	01-17:00, 02-07:00	01-17:00, 02-07:00	-
01/02 Jan 2014	moderate winds, clouds, precipitation	01-13:00, 02-22:00	01-13:00, 02-11:00	01-18:00, 01-19:00, 01-22:00, 01-23:00
18/19 Dec 2014	clouds and precipitation	19-02:00, 19-14:00	18-08:00, 19-14:00	18-21:00, 18-23:00, 19-03:00
09/10 Jan 2015	cloudy, weak winds, strong precipitation	10-02:00, 10-15:00	09-14:00, 10-15:00	09-08:00, 09-10:00, 10-12:00, 10-14:00
19/20 Feb 2015	first clear sky, later cloudy	19-05:00, 20-10:00	19-07:00, 20-10:00	-
01/02 Mar 2015	front crossing	01-23:00, 02-07:00	01-12:00, 02-13:00	01-17:00, 01-22:00, 02-02:00
29/30 Mar 2015	moderate winds, stratiform precipitation	29-09:00, 30-17:00	29-09:00, 30-17:00	29-12:00, 29-18:00, 30-00:00
09/10 Apr 2015	fair weather conditions	09-11:00, 10-19:00	09-11:00, 10-16:00	-

6.2. MOGP Setup

MOGP has been applied to all variables required by the CLM model as atmospheric input. For each variable (except for p) MOGP is run at least twice. In the first run (MOGP_{full}) a large set of predictor variables is considered. The atmospheric and surface variables included in the full training data sets are selected based on our understanding of physical processes potentially affecting a predictand variable. In the second MOGP run (MOGP_{sub}) the terminal set comprises only a subset of the predictors from the first run, namely those that have been used in at least 10% of the Pareto optimal downscaling rules returned by the first MOGP run (cf. Tables 6.4, 6.5). For some predictand variables, additional MOGP runs are carried out. MOGP_{soil} incorporates variables (soil temperature, soil water content) from the soil model TERRA as predictors at high resolution, as those quantities might affect temperature and humidity in the lower atmosphere. MOGP_{noise} contains a simple noise generator in the terminal set as it is expected that the subgrid-scale variability of the atmospheric variables is only in parts determined by the land surface heterogeneity.
Note, that the downscaling rules predict anomalies, i.e., the differences between the spline-interpolated fields and the high-resolution reference (cf. Fig. 6.1). To avoid biases we conserve the coarse pixel mean by subtracting the mean anomaly predicted over a coarse pixel from the predicted anomalies at every pixel on the fine scale,

$$y^D_{tij} = \tilde{y}^D_{tij} - \frac{1}{7 \times 7} \sum_{ij \in X(p,q)} \tilde{y}^D_{tij}, \tag{6.1}$$

with $\tilde{y}^D_{tij}$ being the original anomaly at time step t and grid point (i, j) predicted by a downscaling rule and $X(p, q)$ being a coarse pixel containing 7×7 pixels at the fine scale (cf. Fig. 6.6).

6.2.1. Objectives

MOGP optimizes a trade-off between several objectives defined by the user. The first three objectives chosen - the root mean square error, the error of the subgrid scale standard deviation, and the mean integrated quadratic distance - quantify the quality of the downscaling rules. The complexity of a downscaling rule serves as fourth objective. In our case, each objective is negatively oriented in the sense that it decreases the better the fit. The motivation and definition of the four objectives are given in the following.

Root Mean Square Error The root mean square error (RMSE) between downscaled y^D and reference anomalies y^R is given by

$$RMSE = \sqrt{\frac{1}{n_t n_i n_j} \sum_{i,j,t} (y^R_{tij} - y^D_{tkl})^2}, \tag{6.2}$$

where n_i and n_j denote the number of pixels in x- and y-direction and n_t the number of training data fields (cf. Fig. 6.6).

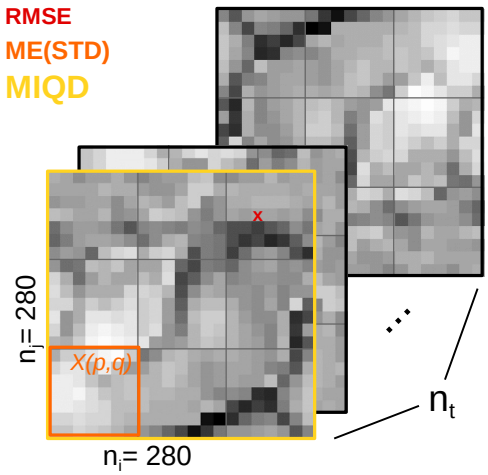

Figure 6.6.: Sketch of the scales at which the different objectives are defined. The root mean square error RMSE is defined on the fine pixel scale. The standard deviation considered by the mean error of the standard deviation ME(STD) is defined on the coarse pixel scale. The integrated quadratic distance considered in the mean integrated quadratic distance MIQD is defined on the full fields.

Mean Error of Standard Deviation The mean error of the subgrid-scale standard deviation ME(STD) is aimed at the reconstruction of the subgrid-scale spatial variability and defined as follows. Let $X(p,q)$ denote the pixels on the coarse scale containing 7×7 fine-scale pixels (cf. Fig. 6.6). The coarse pixel standard deviation $\sigma(y^R_{tpq})$ of the reference field t reads

$$\sigma(y^R_{tpq}) = \sqrt{\frac{1}{7\times 7-1}\sum_{i,j\in X(p,q)}(y^R_{tij}-\bar{y}^R_{tpq})^2} = \sqrt{\frac{1}{7\times 7-1}\sum_{i,j\in X(p,q)}(y^R_{tij})^2}, \tag{6.3}$$

with $\bar{y}^R_{tpq}$ denoting the coarse pixel mean. Again y represents the anomalies. As the spline-interpolation conserves the coarse pixel mean, the coarse pixel mean anomaly (high-resolution minus spline-interpolated field) equals zero, $\bar{y}^R_{tpq} = \bar{y}^D_{tpq} = 0$. Thus, the second objective can be defined as

$$ME(STD) = \frac{7\times 7}{n_p n_q n_t}\sum_{t,p,q}\left|\sigma(y^R_{tpq}) - \sigma(y^D_{tpq})\right|. \tag{6.4}$$

Mean Integrated Quadratic Distance The mean integrated quadratic distance MIQD aims at the minimization of the differences between the empirical cumulative distribution functions (CDFs) of the downscaled and the reference field. The integrated quadratic distance IQD between two CDFs F_1 and F_2 is defined as (*Thorarinsdottir et al.*, 2013):

$$IQD = \int_{-\infty}^{\infty}(F_1(x) - F_2(x))^2 dx. \tag{6.5}$$

Let us denote the empirical CDF of a downscaled field at time t with $F^D_t(y)$ and the CDF of the reference field with $F^R_t(y)$ respectively. From the CDFs we calculate the IQD for each field

(time step) separately (cf. Fig. 6.6) and take the mean over all time steps (i.e., the complete training data set) as objective

$$MIQD = 1/n_t \sum_t \int_{-\infty}^{\infty} (F_t^R(y) - F_t^D(y))^2 dy. \tag{6.6}$$

Solution Size The fourth objective is the size of the solutions, i.e., the number of nodes of a parse tree/downscaling rule. Smaller solutions can be checked for physical consistency more easily, are computationally less expensive and less prone to overfitting. Incorporating the size as an objective further provides information on the dependency between the quality and the complexity of the downscaling rules.

6.2.2. Parameters

The MOGP setup is summarized in Table 6.3. The function set contains the four arithmetic functions and an if statement (*if* A>B *do* C *else do* D). The terminal set contains constants, numbers randomly drawn from the interval $[0, 1]$ and some decimal powers, and variables. The set of predictor variables comprises coarse-scale atmospheric information (cf. Table 6.4) and high-resolution land surface information (cf. Table 6.5). The predictors are not further preprocessed (i.e., no scaling etc. is applied).
Each MOGP run evolves 200 generations with 100 individuals each and a maximum Pareto set size of 50. The genetic operators for (subtree-)mutation and crossover are applied with a probability of 50% each. To keep the downscaling rules readable we set the maximum tree depth to 5 levels. With the above settings a single MOGP run (for one predictand) takes approximately 5 days on a normal PC.

Table 6.3.: Summary of the GP parameters. (Protected division means that division by zero returns the dividend not an error.)

Parameter	Value
function set	+ , - , × , protected /, if
terminal set	random numbers [0,1], 10, 10^2, 10^3, predictor variables (Tables 6.4, 6.5); random fields for MOGP_{noise} only
generations	200
population size	100
max. Pareto set size	50
genetic operators	(subtree-)mutation, crossover
max. tree levels	5

Atmospheric Predictors

The atmospheric predictors are described in Table 6.4. Some of the parameters have been derived from the COSMO model output, such as the vertical gradients (gr) and the coarse-scale 3×3 pixel standard deviation (sd3×3). The pressure gradient between the two lowest

2 main levels at approximately 10 m and 35 m height is for instance given by

$$p_{gr25} = \frac{p(2) - p(1)}{h(2) - h(1)} = \frac{p_{35} - p_{10}}{35 - 10}. \tag{6.7}$$

At each fine-scale pixel the coarse-scale 3×3 standard deviation is the standard deviation within the 3×3 coarse pixels surrounding the respective fine-scale pixel. This is aimed at reproducing spatial variability assuming that coarse and fine-scale variability are related.
The specific humidity of an air parcel is the ratio between the water vapor mass m_w and the air parcels total (dry + moist) mass m_t, i.e, $q = m_w/m_t$. The solar zenit angle Ω is the angle between the zenit and the sun center. The total precipitable water TPW is the amount of all water in an atmospheric column and typically given in millimeters of liquid water. The total precipitable water can be imagined as the height of water in an atmospheric column if all water in the vertical column were precipitated as rain.
The last two predictors are the convective available potential energy of a mean surface layer parcel $CAPE_{ml}$ and of the most unstable parcel $CAPE_{mu}$. If an air parcel is lifted it can, depending on the atmospheric stability, reach a level of free convection where it experiences buoyancy such that it accelerates further until it reaches equilibrium with its surroundings. $CAPE$ is calculated by vertically integrating the local buoyancy of a parcel from the level of free convection z_f to the equilibrium level z_n,

$$CAPE = g \int_{z_f}^{z_n} \frac{T_{v,par} - T_{v,env}}{T_{v,env}} dz \tag{6.8}$$

with $g = 9.81$ m/s being the gravitational acceleration and T_v the virtual temperature of the considered air parcel (par) and its environment (env). The virtual temperature of a moist air parcel is the temperature of the corresponding dry parcel with same pressure and density. A high $CAPE$ suggests potential for strong convection and is thus an indicator for potentially severe weather.

Surface Predictors

The surface predictors are listed in Table 6.5. The anomalies are defined with respect to the corresponding spline-interpolated fields [1]. The four tp predictands describe the topography at each grid point with respect to its direct four neighbours. tp_1 is the mean height difference to neighboring grid points, tp_1^* is the anomaly of tp_1 again with respect to the spine-interpolated field. tp_2 and tp_3 are the local slopes in the direction of lowest and highest neighbouring grid point.
The topographic index TI and the contributing area CA are parameters widely used in hydrological models and studies. The contributing area of a pixel is the catchment area from which water drains to the pixel. The topographic index (or topographic wetness index) is

[1] The anomaly is determined in two steps. First, the high-resolution field is coarsened to a 2.8 km grid by spatial averaging. Second, the coarse field is interpolated back to the high resolution. The difference between original high-resolution and interpolated field gives the anomaly.

Table 6.4.: Atmospheric predictors at coarse (2.8 km) grid resolution incorporated in the terminal sets for the different predictand variables. The predictors are further described in the text. The crosses indicate whether a certain predictor (row) is contained in the MOGP terminal set for a certain downscaling predictand (column). The different colors, black and red, relate to the two-step procedure of the MOGP runs. In a first MOGP run (MOGP$_{full}$) the full set of predictors (red crosses and black bullets) is used. In the second run (MOGP$_{sub}$) only predictors (red crosses) are incorporated in the terminal set that are used in at least 10% of the Pareto optimal downscaling rules returned by MOGP$_{full}$.

		PREDICTANDS							
PREDICTORS		p^*	T^*	w_h^*	q^*	$LWR_{dif}{}^*$	SWR_{dir}^*	SWR_{dif}^*	P^*
p_{10}	pressure at lowest model layer ($\approx$ 10 m)	X							
p_{gr25}	vert. p gradient of lowest 2 layers ($\approx$ 25 m)	X							
p_{gr60}	vert. p gradient of lowest 3 layers ($\approx$ 60 m)	X							
p_{gr110}	vert. p gradient of lowest 4 layers ($\approx$ 110 m)	X							
$p_{sd3\times3}$	standard deviation of p_{10} in 3×3 coarse pixels	X							
T_{10}	temperature at lowest model layer ($\approx$ 10 m)		X		•				
T_{gr25}	vert. T gradient of lowest 2 layers ($\approx$ 25 m)		•	•	•				•
T_{gr60}	vert. T gradient of lowest 3 layers ($\approx$ 60 m)		X	•	•				X
T_{gr110}	vert. T gradient of lowest 4 layers ($\approx$ 110 m)		•	•	X				•
$T_{sd3\times3}$	standard deviation of T_{10} in 3×3 coarse pixels		X						
q_{10}	specific humidity at lowest model layer ($\approx$ 10 m)				X				
q_{gr25}	vert. q gradient of lowest 2 layers ($\approx$ 25 m)				X				
q_{gr60}	vert. q gradient of lowest 3 layers ($\approx$ 60 m)				X				
q_{gr110}	vert. q gradient of lowest 4 layers ($\approx$ 110 m)				X				
q_{m1660}	sum of q over lowest 15 layers			•	•	•	X	•	
$q_{10,div}$	horizontal divergence of the q_{10} flux			•				•	
$q_{m1660,div}$	horizontal divergence of the q_{m1660} flux			•					
$q_{sd3\times3}$	standard deviation of q_{10} in 3×3 coarse pixels			•	•	•	•	•	
$w_{v,20}$	vertical wind speed ($\approx$ 20 m)		•	•	•	•	•	•	•
$w_{h,10}$	horizontal wind speed at lowest model layer ($\approx$ 10 m)		•	X	•	•	X	•	•
$w_{h,gr25}$	vert. wind sheer of lowest 2 layers ($\approx$ 25 m)			•					
$w_{h,gr60}$	vert. wind sheer of lowest 3 layers ($\approx$ 60 m)			•					
$w_{h,gr110}$	vert. wind sheer of lowest 4 layers ($\approx$ 110 m)			•					
$w_{h,sd3\times3}$	standard deviation of $w_{h,10}$ in 3×3 coarse pixels			X					
TKE_{20}	turbulent kinetic energy ($\approx$ 20 m)		•	X	•	•	•	•	•

PREDICTORS		PREDICTANDS							
		p^*	T^*	w_h^*	q^*	${LWR_{dif}}^*$	SWR_{dir}^*	SWR_{dif}^*	P^*
$R_{net,s}$	net radiation at the surface		•		•	•	•	X	•
$SWR_{net,t}$	net downward shortwave radiation at the TOA					•	•	X	
$LWR_{net,t}$	net downward longwave radiation at the TOA					•	•	•	
$SWR_{dir,s}$	direct incoming SWR at the surface					X	•	•	
$SWR_{dif,s}$	diffuse incoming SWR at the surface					X	•	•	
$LWR_{dif,s}$	incoming LWR at the surface					X	•	•	
Ω	cosine of solar zenit angle		•		•	•	X	•	
$LWR_{dif,s,sd3\times3}$	standard deviation of $LWR_{dif,s}$ in 3×3 coarse pixels					•			
$SWR_{dir,s,sd3\times3}$	standard deviation of $SWR_{dir,s}$ in 3×3 coarse pixels						•		
$SWR_{dif,s,sd3\times3}$	standard deviation of SWR_{dif_s} in 3×3 coarse pixels							•	
$CAPE_{ml}$	CAPE of mean surf. layer parcel					•	•	•	•
$CAPE_{mu}$	CAPE of most unstable parcel					•	X	•	•
CLC_t	total cloud cover		•		•	X	•	X	•
CLC_l	cloud cover of low clouds (surface to 800 hPa)		•		•	•	X	•	•
CLC_m	cloud cover of medium high clouds (800 hPa to 400 hPa)				•	•	X	•	•
CLC_h	cloud cover of high clouds (400 hPa to TOA)				•	•	•	•	X
TCW	vertical integrated cloud water				•	•	•	X	X
TCI	vertical integrated cloud ice				•	•	•	•	X
$CLC_{t,sd3\times3}$	standard deviation of CLC_t in 3×3 coarse pixels					•	•	•	
$CLC_{l,sd3\times3}$	standard deviation of CLC_l in 3×3 coarse pixels					•	X	X	
P	instantaneous precipitation			•	•	•	•	•	X
TPW	precipitable water				•	•	•	•	X
TP	vertically integrated water content of P				•	•	•	•	X
$P_{sd3\times3}$	standard deviation of P in 3×3 coarse pixels								X

Table 6.5.: Same as in Table 6.4, but for the surface predictors at the fine (400 m) grid resolution incorporated in the terminal sets for the different predictand variables. The green crosses in brackets indicate additional predictors for MOGP_{soil}.

PREDICTORS		PREDICTANDS							
		p^*	T^*	w_h^*	q^*	LWR_{dif}^*	SWR_{dir}^*	SWR_{dif}^*	P^*
h	topographic height	X	X	X	X	X	X	X	X
h^*	topographic height anomaly	X	X	X	•	X	X	•	•
tp_1	mean height difference to neighboring grid points		X	X	•	•	X	X	•
tp_1^*	anomaly of tp_1		X	X	•	•	X	X	•
tp_2	slope to lowest neighboring grid point		X	X	X	•	X	•	•
tp_3	slope to highest neighboring grid point		X	X	•	•	•	•	•
TP	topographic index		•	•	•	X	•	•	•
TP^*	anomaly of topographic index		X	•	X	X	X	•	•
CA	contributing area		•	•	X	•	•	•	
CA^*	anomaly of contributing area		•	•	•	•	•	•	•
PLC	plant cover		•	X	X	X	•	X	X
PLC^*	anomaly of plant cover		X	X	•	•	•	•	X
z_0	roughness length		X	X	•				X
z_0^*	anomaly of roughness length		X	X	•				X
α	albedo		•		•	X	•	X	
α^*	anomaly of albedo		•		X	X	X	X	
LAI	leaf area index				X	X			X
LAI^*	anomaly of leaf area index				X	•			•
ST	soil type				X	X			
FR_{land}	fraction of land		X	X	X	X	•	X	
$h_{gr,w,c}$	gradient of h in wind direction (of coarse pixel)	X		•					•
$z_{0,gr,w,c}$	gradient of z_0 in wind direction (of coarse pixel)			X					X
$h_{gr,w,f}$	gradient of h in wind direction (of fine-scale pixel)	X		X					•
$z_{0,gr,w,f}$	gradient of z_0 in wind direction (of fine-scale pixel)			•					•
T_{so}	soil temperature of uppermost soil layer ($\approx$ 5 cm depth)		(X)		(X)				
T_{so}^*	anomaly of T_{so}		(X)		(X)				
W_{so}	soil water content of uppermost soil layer ($\approx$ 5 cm depth)		(X)		(X)				
W_{so}^*	anomaly of W_{so}		(X)		(X)				
$W_{so,7}$	water content of soil column between surf. and $\approx$ 7 m depth		(X)		(X)				
$W_{so,7}^*$	anomaly of $W_{so,7}$		(X)		(X)				

defined as

$$TI = ln\frac{CA}{tan(b)}, \tag{6.9}$$

with b the local slope. The topographic index gives an estimate how likely water gathers at a pixel.
The plant cover PLC is the percentage of a surface pixel covered by plants. The roughness length z_0 quantifies the friction of the air flow close to the land surface. The albedo α is the ratio between reflected and incoming radiation at the respective surface. The leaf area index LAI is given by the (one-sided) area of all leaf surfaces above a land surface pixel divided by the area of the respective pixel and is an important parameter for transpiration and CO_2 budget. The fraction of land FR_{land} indicates the percentage of a surface pixel covered by land (and not water).
The parameters $h_{gr,w,c}$ and $z_{0,gr,w,c}$ are the horizontal gradients of the coarse topographic height and roughness length fields in the near-surface wind direction (at coarse scale). $h_{gr,w,f}$ and $z_{0,gr,w,f}$ are the horizontal gradients of the fine-scale topographic height and roughness length fields in the near surface wind direction (at coarse scale).

6.3. Results

We have applied MOGP to the detection of downscaling rules for pressure p, temperature T, specific humidity q and horizontal wind speed w_h at 10 m height, incoming diffuse longwave radiation LWR_{dif}, incoming direct shortwave radiation SWR_{dir} and incoming diffuse shortwave radiation SWR_{dif} at the land surface and precipitation P from 2.8 km to 400 m grid spacing. For all variables except pressure and precipitation several MOGP runs have been carried out with slightly different setups:

- MOGP$_{full}$ uses the full set of predictors (cf. Tables 6.4, 6.5)
- MOGP$_{sub}$ uses only the 10-16 predictors most frequently used by MOGP$_{full}$ (cf. Tables 6.4, 6.5)
- MOGP$_{soil}$ uses predictors from MOGP$_{sub}$ together with some additional fine-scale predictors from the land-surface/subsurface scheme like soil temperature and soil water content (cf. Tables 6.4, 6.5).
- MOGP$_{noise}$ uses the predictors from MOGP$_{sub}$ and additionally incorporates a random number generator (in the terminal set) which randomly draws numbers form a uniform distribution in the interval [0,1]. The random number generator is applied for each grid point of each field each time a parse tree is evaluated. (The random constants included in the terminal set for all MOGP setups are the same for each grid point and field and are drawn only once.) The random number generator provides a possibility for GP to generate variability that is not directly related to the land surface predictors.

The evaluation of the downscaling performance relies for the most part on the objective functions RMSE, ME(STD) and MIQD defined in Section 6.2.1. All objectives are formulated

as penalties, i.e., the smaller the objective the better. Thus, we define the relative reduction of an objective s_i by a downscaling rule α as

$$\tilde{s_i}(\alpha) = 1 - s_i(\alpha)/s_i(0), \tag{6.10}$$

where $s_i(0)$ is the value of an objective when predicting zero anomalies, which corresponds to the spline-interpolated field. Note that $\tilde{s_i}(\alpha) \in (-\infty, 1]$.
The relative reduction is analogue to a skill score, with the objective for a perfect solution $s_i(\alpha)$ equal to zero. A positive $\tilde{s_i}(\alpha)$ indicates that the downscaled field is better than the spline-interpolated field concerning objective s_i; for a perfect downscaling $\tilde{s_i}(\alpha) = 1$; for a downscaling that performs as good as the interpolated field $\tilde{s_i}(\alpha) = 0$; for a downscaling worse than the interpolated field $\tilde{s_i}(\alpha) < 0$.
The importance of the predictors for downscaling a certain variable is quantified by the frequency with which the predictor x_i is contained in the Pareto set returned by MOGP,

$$h(x_i) = \frac{n(\alpha \in \mathcal{P}' | x_i \in \alpha)}{n(\alpha \in \mathcal{P}')}, \tag{6.11}$$

or in words, the frequency $h(x_i)$ is given by the number n of rules α in Pareto set $\mathcal{P}'$ containing predictor x_i at least once divided by the total number of rules in the Pareto set.

6.3.1. Pressure

Overview

Figure 6.7 gives an overview of the performance of MOGP for downscaling pressure at 10 m height. Only few rules reduce the RMSE, but one downscaling rule that reduces the RMSE by

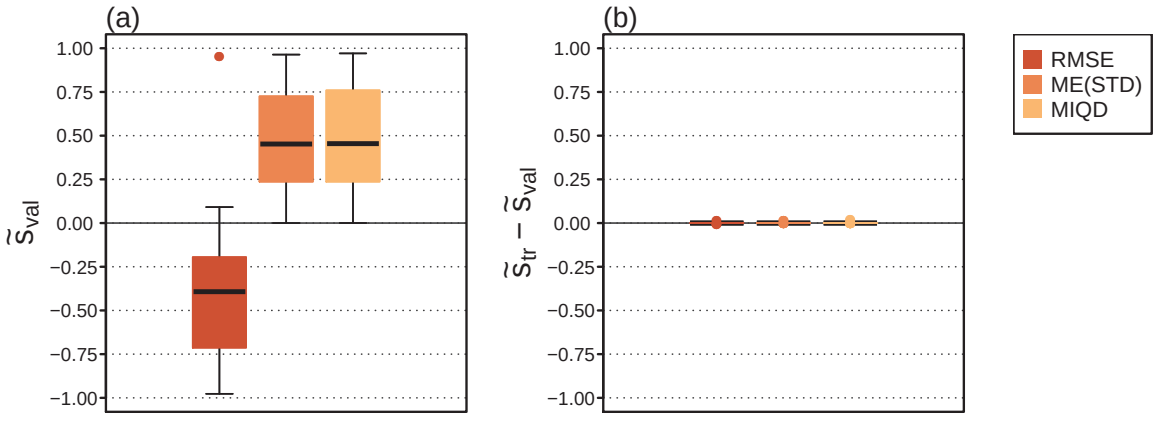

Figure 6.7.: Overview of the MOGP performance for downscaling 10 m pressure. (a) relative reduction of RMSE, ME(STD) and MIQD for the 50 rules from the Pareto set applied to the validation (val) data set; (b) difference in relative reduction between training (tr) and validation data set. The horizontal line within the boxes is the median. The upper and lower boundaries of the boxes correspond to the 75%- and 25%-quantiles. The whiskers indicate the range spanned by maximum and minimum. The length of the whiskers is restricted to 1.5 times the interquartile range (i.e., range between 25%- and 75%-quantile). Values outside this range are considered outliers and shown as circles.

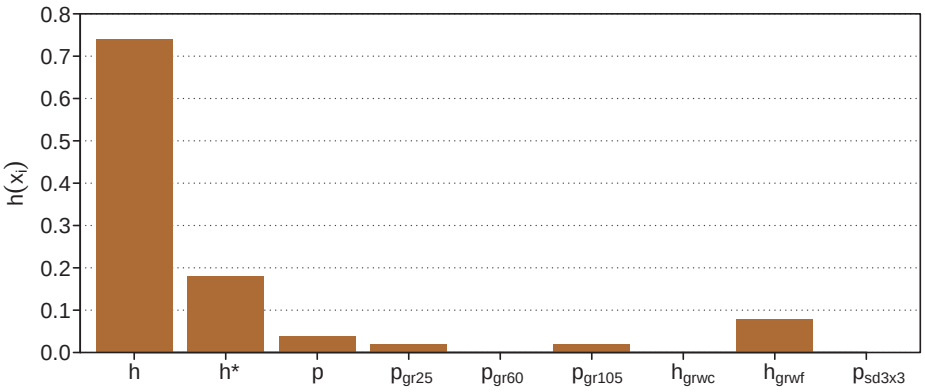

Figure 6.8.: Usage frequency $h(x_i)$ of the predictors (cf. Tables 6.4, 6.5) for downscaling 10 m pressure.

more than 90% sticks out. The ME(STD) and MIQD are reduced by up 99%. No overfitting tendencies occur (Fig. 6.7(b)).
Surface height h and surface height anomaly h^* are the most frequently used predictors followed by $h_{gr,w,f}$, near-surface pressure p and near-surface vertical pressure gradients (Fig. 6.8). Physics suggest a strong relation between height z and pressure p given by the hydrostatic equation,

$$\frac{\partial p}{\partial z} = -\rho g, \tag{6.12}$$

where ρ is the air density and $g = 9,81$ m s^{-2} the gravitational acceleration.

Selected Downscaling Rules

The rule returned by MOGP that yields the best reduction of RMSE (97%) and a reduction of 97% to 99% for ME(STD) and IQD (cf. Table 6.6) is shown in Fig. 6.9. Translating the parse tree into an equation the rule reads

$$p^* = h_{gr,w,f} + h^* \; p_{gr25} \approx h^* \; p_{gr25} \qquad \text{(MOGP)}.$$

That is, the pressure anomaly p^* w.r.t. the spline-interpolated field is given by the surface height anomaly h^* times the pressure gradient of the lowest 25 m p_{gr25} added to the horizontal gradient of the (fine-scale) surface height in the (coarse-scale) wind direction $h_{gr,w,f}$. The first

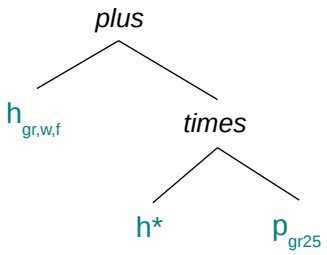

Figure 6.9.: One downscaling rule returned by MOGP for downscaling 10 m pressure.

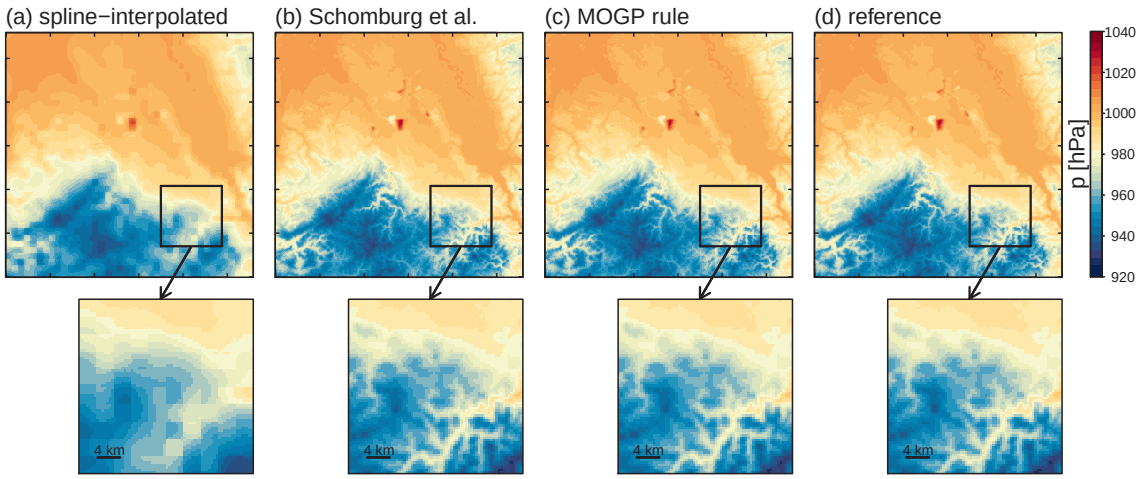

Figure 6.10.: 10 m pressure field on May 7$^{\text{th}}$ 2014 8:00 UTC. (a) shows the spline-interpolated field; (b) shows the downscaled field resulting from the downscaling rule from *Schomburg et al.* (2010); (c) shows the downscaled field resulting from the MOGP downscaling rule (cf. Fig. 6.9); (d) shows the reference field from the high-resolution model run.

term $h_{gr,w,f}$ is, however, much smaller than $h^* p_{gr25}$ and can be neglected without significantly affecting the results.
The downscaling rule for pressure from *Schomburg et al.* (2010) is based on the hydrostatic equation and assumes a constant air density of $\rho = 1.19$ kg m^{-3}, i.e.,

$$p^* = -11.67\ h^* \qquad (\textit{Schomburg} \text{ et al.}).$$

Also the downscaling rule from *Schomburg et al.* (2010) performs well (cf. Table 6.6), but the MOGP rule performs slightly better with a relative reduction that is about 1% larger for the three objectives. Figure 6.10 shows the downscaled pressure field at 10 m height on May 7$^{\text{th}}$ 2014 8:00 UTC. The strong dependency of the 10 m pressure on surface height is obvious (cf. Fig. 6.3 (a)). By eye one can not distinguish between the high-resolution reference and the downscaled pressure fields. The difference in performance between the two rules is small and both are simple and physically reasonable, but for 29 out of 30 near-surface pressure fields contained in the validation data set the MOGP rule performs better then the rule from *Schomburg et al.* (2010). Thus, the use of the MOGP rule is suggested.

Table 6.6.: Relative reduction of the objectives by the MOGP downscaling rule for 10 m pressure (Fig. 6.9) and by the downscaling rule from *Schomburg et al.* (2010) (see text) applied to the validation data set.

	$\widetilde{\text{RMSE}}$	$\widetilde{\text{ME(STD)}}$	$\widetilde{\text{MIQD}}$
MOGP	0.97	0.97	0.99
Schomburg et al.	0.96	0.97	0.98

6.3.2. Temperature

Figure 6.11 summarizes the performance of MOGP for downscaling 10 m temperature. ME(STD) and MIQD are reduced by up to about 70% and 85%, respectively, by all four MOGP setups (full, sub, soil, noise). The RMSE is reduced only by some rules. On average the RMSE is slightly increased. No systematic overfitting occurs, but the RMSE is occasionally better reduced for the training data set (by up to 15%) for MOGP$_{full}$ and MOGP$_{soil}$ (Fig. 6.11(b)). The performance of the different MOGP setups differs only slighly. As MOGP is non-deterministic and only one run has been carried out with each setup, any interpretation of the differences should be treated with caution. Still, for temperature downscaling it seems not to be beneficial to reduce the predictor set from the full set to a subset. The use of the soil predictors (soil temperature and soil water content) leads to a slight improvement. Using a basic noise generator does not affect the overall performance of MOGP much.
Figure 6.12 shows scatter plots for the Pareto sets from all four MOGP setups and incorporates all four objectives. Except for some outliers for the RMSE, MOGP$_{full}$ and MOGP$_{soil}$ yield the best trade-off between the different objectives. The performance concerning RMSE, ME(STD) and MIQD is only slightly dependent on the solution SIZE, but the very small solutions with ≤ 5 nodes cannot compete with the more complex ones. The solutions with about 10 nodes perform as well as more complex solutions with 20 nodes or more.
The overall most used predictors are topographic height anomaly h^*, the parameters derived from the topography field tp_1, tp_1^*, tp_2 ,tp_3, the vertical temperature gradient of the lowest 60 m T_{gr60} and the 3×3 coarse pixel standard deviation of 10 m temperature $T_{sd3\times3}$ (Fig. 6.13). From the six soil predictors the anomaly of the soil water content in the uppermost soil layer W_{so}^* is by far the most frequently used. The use of the topographic height anomaly and a vertical temperature gradient is physically reasonable. The downscaling rule from *Schomburg et al.* (2010) uses very similar predictors, namely the temperature gradient of the lowest 105 m T_{gr105} and the topographic height anomaly h^*,

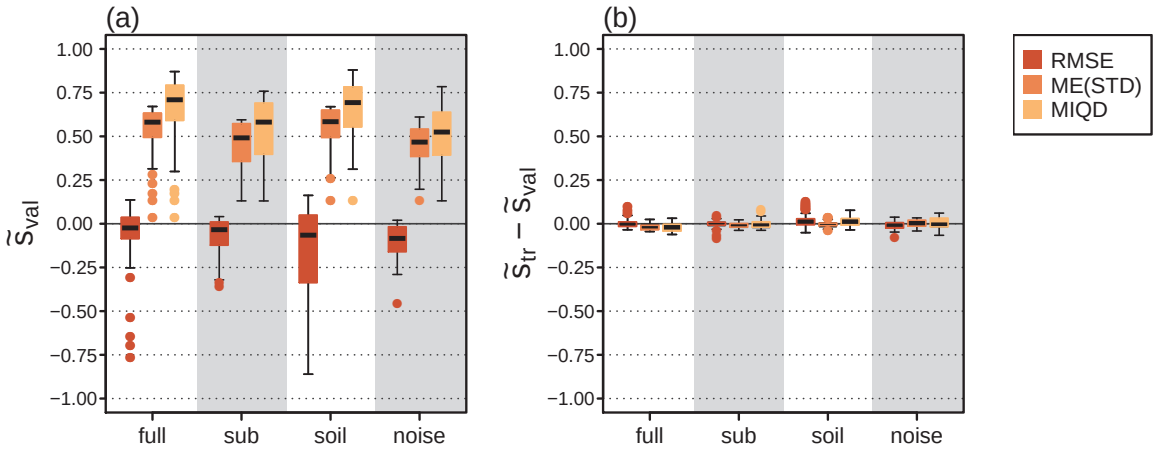

Figure 6.11.: Overview of the MOGP performance for downscaling 10 m temperature. (a) relative reduction of RMSE, ME(STD) and MIQD for the 50 rules from the Pareto set rules applied to the validation (val) data set; (b) difference in relative reduction between training (tr) and validation data set.

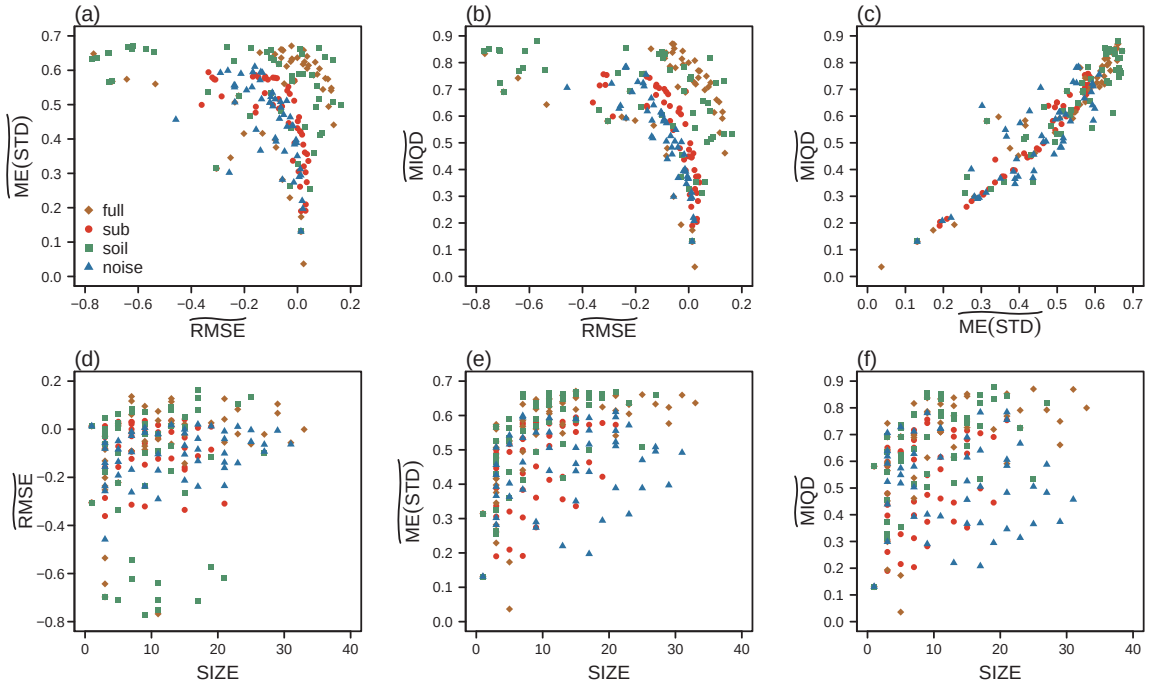

Figure 6.12.: Scatter plots of the relative reduction of RMSE, ME(STD) and IQD for the Pareto sets for downscaling 10 m temperature when applied to the validation data. The bottom row additionally considers the solution SIZE (i.e., number of nodes) as fourth objective.

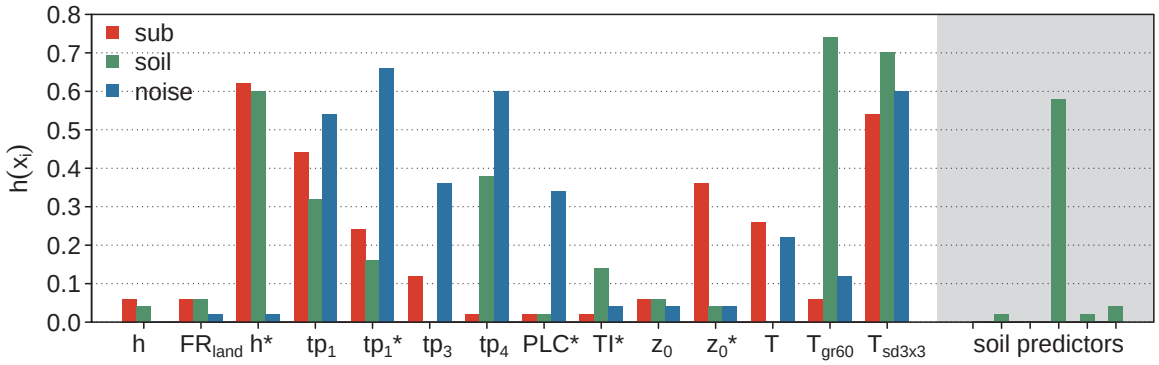

Figure 6.13.: Usage frequency $h(x_i)$ of the predictors (cf. Tables 6.4, 6.5) for downscaling 10 m temperature. The soil predictors are from left to right: T_{so}, T^*_{so}, W_{so}, W^*_{so}, $W_{so,7}$ and $W_{so,7}$.

IF $T_{gr105} < 0.0058\ Km^{-1}$ (Schomburg et al.)

$T^* = -0.0084\ Km^{-1}\ h^*$

ELSE

$T^* = 0.$

That is, for unstable conditions in the lower boundary layer, indicated by the temperature gradient of the lowest 105 m being lower than 0.0058 Km^{-1}, the temperature anomaly is approximately proportional to the topographic height anomaly h^*. For stable conditions this relation is not valid. Here the rule returns zero anomalies.

Also the coarse scale standard deviation of 10 m temperature T_{sd3x3} is used in the downscaling

scheme by *Schomburg et al.* (2010). It serves as a predictor to estimate the amplitude of the noise added in the third and final step of the scheme. The frequent use of this parameter also by MOGP implies that spatial variability at coarse (2.8 km) and fine (400 m) scale are related for the considered model domain. The *tp* predictors are foremost used by MOGP_{noise}. These predictors contain information on the local topography heterogeneity.

Selected Downscaling Rules

A comparison of the downscaling rule from *Schomburg et al.* (2010) and an MOGP based downscaling rule (*Zerenner et al.*, 2016) revealed a slightly favorable performance for the rule from *Schomburg et al.* for unstable atmospheres. For stable conditions, especially during clear sky nights, the MOGP based downscaling recovers large parts of the subgrid-scale variability (with only little increase in RMSE) while the rule from *Schomburg et al.* is not applicable, i.e., returns zero anomalies.

In the following two well performing MOGP downscaling rules are selected and analyzed in more detail. To select these two downscaling rules the sum of the relative reduction of RMSE, ME(STD) and MIQD is calculated. The rules which give the largest overall reduction are selected (Fig. 6.14). Rule 1 is from MOGP_{soil}. Rule 2 stems from MOGP_{full}.

Both rules show only minor difference in performance between training and validation data

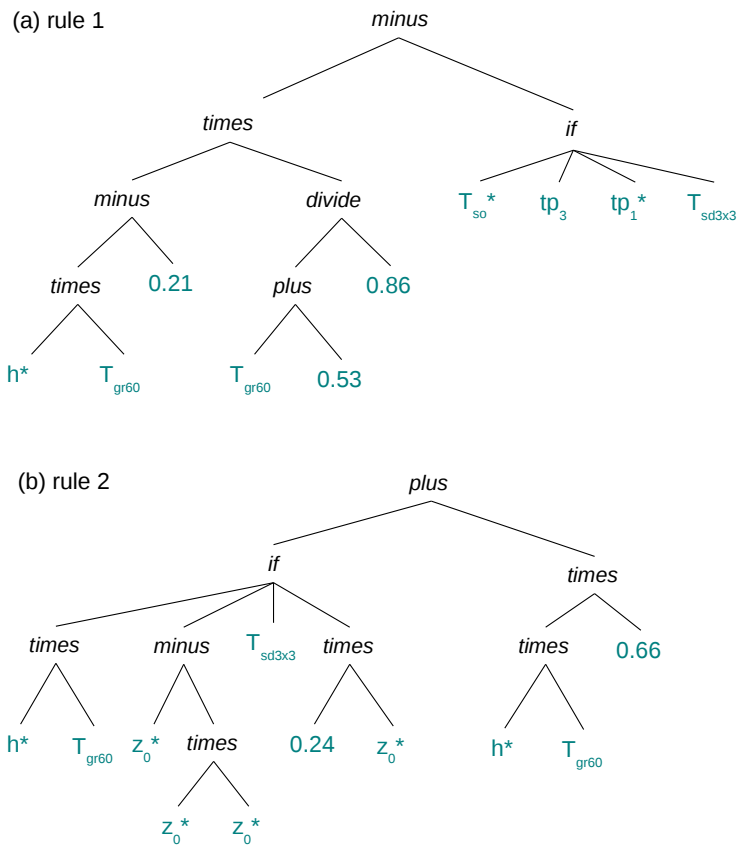

Figure 6.14.: Two downscaling rules returned by MOGP for downscaling 10 m temperature. Rule 1 is from the Pareto set returned by MOGP_{full}. Rule 2 is from the Pareto set returned by MOGP_{soil}.

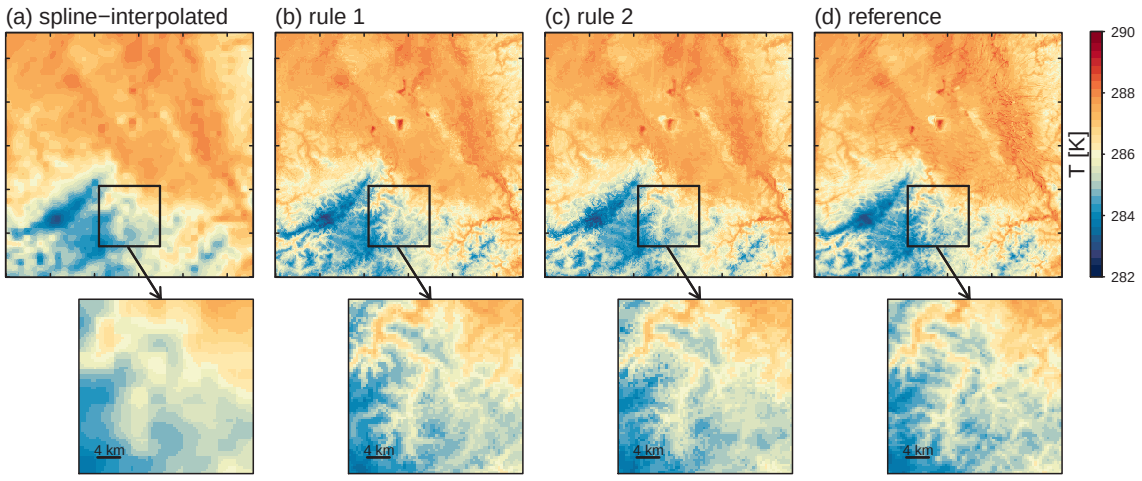

Figure 6.15.: 10 m temperature field on May 22nd 2015 9:00 UTC. (a) shows the spline-interpolated field; (b) and (c) show the downscaled fields resulting from rules 1 and 2 (cf. Fig. 6.14); (d) shows the reference field from the high-resolution model run.

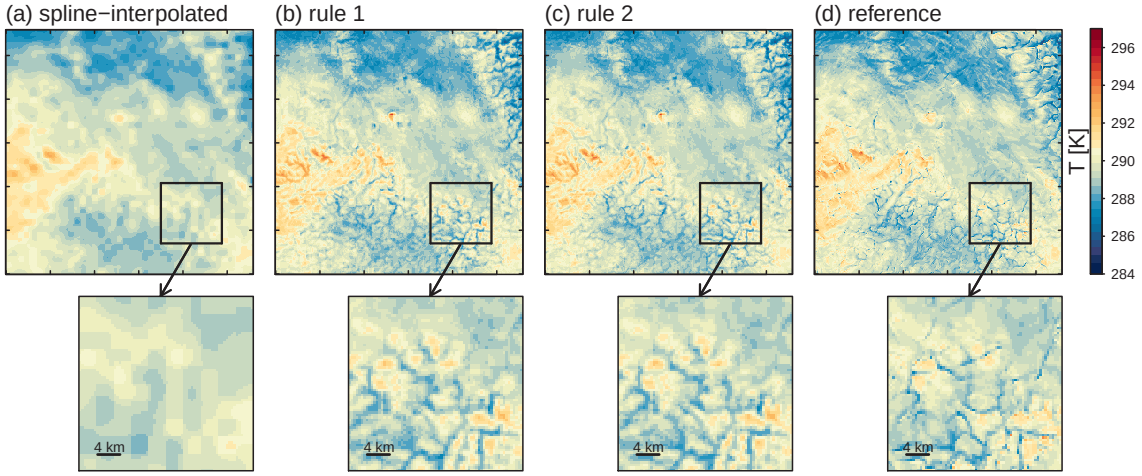

Figure 6.16.: 10 m temperature field on September 18th 2014 1:00 UTC. (a) shows the spline-interpolated field; (b) and (c) show the downscaled fields resulting from rules 1 and 2 (cf. Fig. 6.14); (d) shows the reference field from the high-resolution model run.

set (Table 6.7). The CDF and the subgrid-scale standard deviation of the high-resolution reference fields are well reproduced by both downscaling rules. For the ME(STD) the relative reduction amounts to 0.66. For the MIQD a relative reduction of ≥ 0.82 is achieved. Rule 1 slightly increases the RMSE. Rule 2 yields a minor reduction of 1%.

Rule 1 from Fig. 6.14 can be simplified. Adding or subtracting coarse-scale predictors and/or numerical constants does not have any effect on the final downscaled field as the coarse pixel mean is conserved. This means that terms such as $-T_{sd3x3}$ can be neglected. After removing those terms and summarizing the coefficients rule 2 reads:

IF $T^*_{so} > tp_3$ (rule 1)

$T^* = 0.61\ h^*\ T_{gr60} + 1.16\ h^*\ T^2_{gr60} - tp^*_1$

ELSE
$T^* = 0.61\ h^*\ T_{gr60} + 1.16\ h^*\ T_{gr60}^2$.

The most important term, a multiplication of topographic height anomaly and the vertical temperature gradient in the lowest three atmospheric layers $h^* T_{gr60}$, is physically intuitive. As $\mathcal{O}(T_{gr60}) \approx 10^{-2}$ K/m, the term that is quadratic in T_{gr60} is typically at least one order of magnitude smaller than the one linear in T_{gr60}. In cases where the soil temperature anomaly of the uppermost soil layer is larger than tp_3, tp_1^* is subtracted. This is for most situations and most fine-scale pixels the case. The topographic parameter tp_1^* ranges between -0.42 and 0.34 for the considered domain, with the largest (positive or negative values) in the mountainous areas. For valley pixels tp_1^* is large positive, for mountain tops large negative. That means subtracting tp_1^* tends to increase the predicted anomalies for calm, clear sky nights, where negative anomalies occur in the valleys and positive anomalies in higher altitudes. During day time situations with a well mixed boundary layer the effect is inverse, i.e, the magnitude of the temperature anomalies is reduced by the term. This is reasonable as the most pronounced subgrid-scale variability in the near surface temperature fields is observed during calm, clear sky nights.
Translating the parse trees from Fig. 6.14 into an equation, rule 2 reads:

IF $h^*\ T_{gr60} > -z_0^*$ (rule 2)
$T^* = 0.66\ h^*\ T_{gr60}$
ELSE
$T^* = 0.24\ z_0^* + 0.66\ h^*\ T_{gr60}$.

As in rule 1, also in rule 2 a multiplication between topographic height anomaly h^* and vertical temperature gradient in the lowest three atmospheric layers T_{gr60} constitutes the most important term.
In cases where the roughness length anomaly z_0^* is large compared to $h^*\ T_{gr60}$, the term $0.24\ z_0^*$ is added. This is less intuitive. A large roughness length implies stronger turbulence and smaller horizontal wind speeds. According to the rule large positive roughness length anomalies are accompanied by positive temperature anomalies, negative roughness length

Table 6.7.: Relative reduction of the different objectives by the MOGP downscaling rules for 10 m temperature from Fig. 6.14 applied to training and validation data sets. The values for the training data set are given in brackets. The bottom two rows provide the relative reduction for the two fields in Figs. 6.15 (May 22nd 2015 9:00 UTC) and 6.16 (September 18th 2014 1:00 UTC).

(a) rule 1

	$\widetilde{\text{RMSE}}$	$\widetilde{\text{ME(STD)}}$	$\widetilde{\text{MIQD}}$
	-0.02 (-0.05)	0.67 (0.66)	0.84 (0.84)
2015/05/22 9:00	0.28	0.75	0.97
2014/09/18 1:00	-0.15	0.67	0.94

(b) rule 2

	$\widetilde{\text{RMSE}}$	$\widetilde{\text{ME(STD)}}$	$\widetilde{\text{MIQD}}$
	0.01 (0.08)	0.67 (0.66)	0.82 (0.83)
2015/05/22 9:00	0.22	0.72	0.90
2014/09/18 1:00	-0.03	0.67	0.96

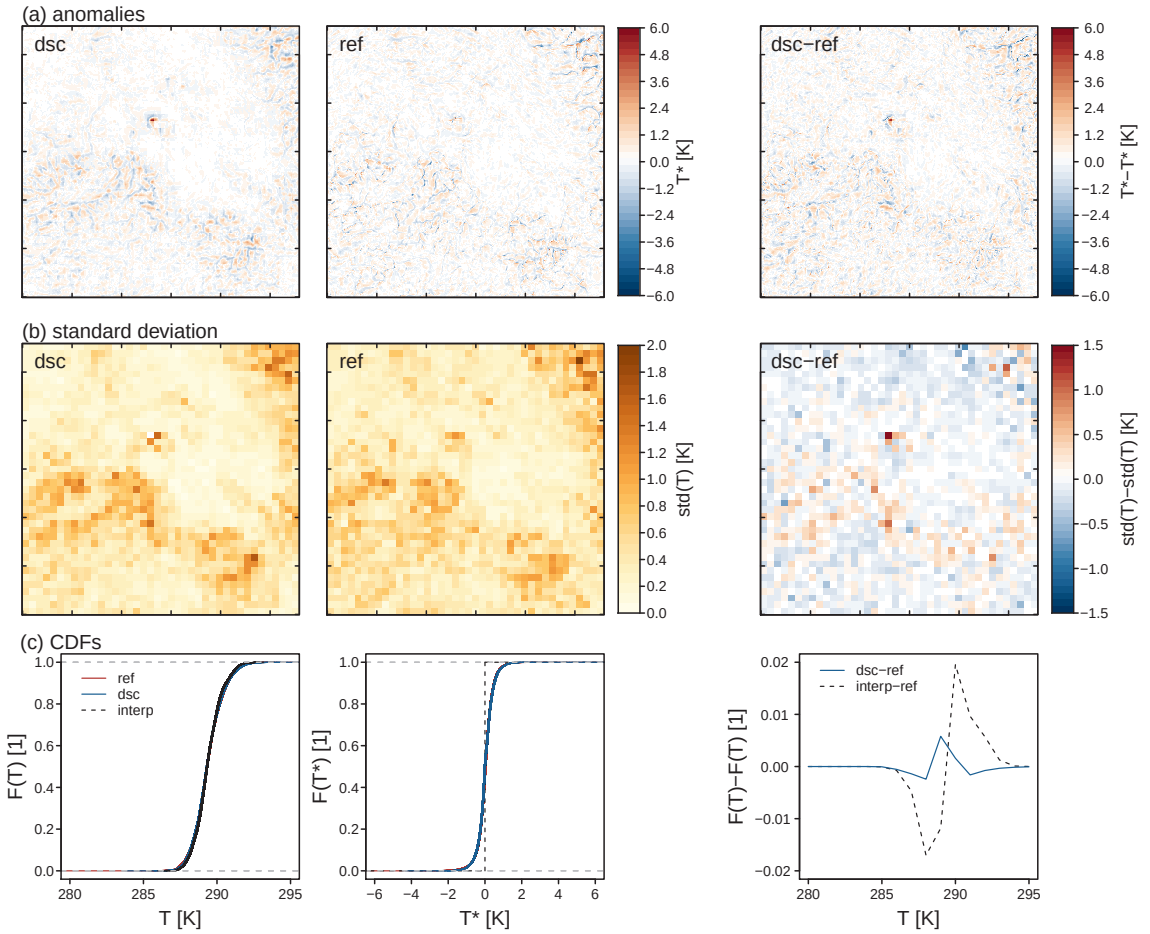

Figure 6.17.: Detailed evaluation of the MOGP performance for downscaling the 10 m temperature field on September 18$^{\text{th}}$ 2014 1:00 UTC (cf. Fig. 6.16): (a) illustrates the performance concerning the RMSE and provides the temperature anomalies T^* for downscaled (dsc) and reference field (ref) and the respective difference plot; (b) illustrates the performance concerning the ME(STD) and provides the fine-scale standard deviation of downscaled and reference temperature fields T together with the respective error; (c) shows the CDFs of downscaled and reference temperature (and anomaly) fields and the difference in CDFs.

anomaly by negative temperature anomalies.

Figures 6.15 and 6.16 illustrate the performance of the two rules for two 10 m temperature fields from the validation data set. Figure 6.15 shows the 10 m temperature field on May 22$^{\text{nd}}$ 2015 9:00 UTC, a calm day with few clouds and low wind speeds. The overall fine-scale structure of the temperature field is well captured by both rules. As already shown by *Schomburg et al.* (2010) in unstable atmospheres the near-surface temperature anomalies are closely related to the topographic height anomaly. The RMSE is reduced by 28% and 22% by rule 1 and rule 2, respectively (Table 6.7). The two objectives quantifiying the fine-scale spatial variability are reduced by at least 70%.

Figure 6.16 shows the near-surface temperature field on September 18$^{\text{th}}$ 2014 1:00 UTC, a mostly clear sky night following a warm, clear sky day. In clear sky nights temperature anomalies can grow very large in regions with strong topography due to strong radiative cooling leading to a stable or even inverse layered atmosphere. Cold air accumulates in the valleys

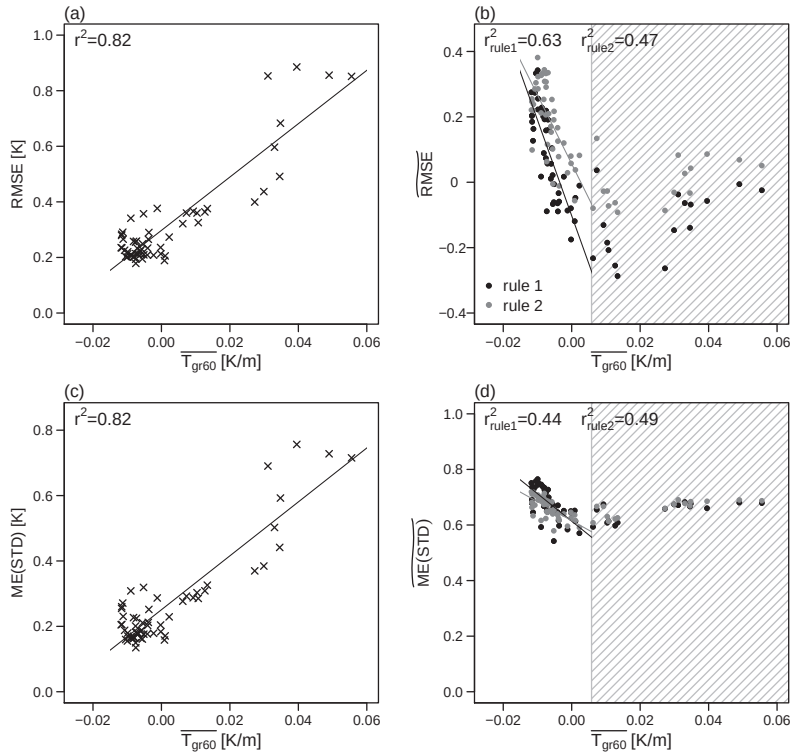

Figure 6.18.: Left column: RMSE and ME(STD) of spline-interpolated 10 m temperature fields compared to high-resolution reference fields versus the spatially averaged temperature gradient in the lowest three COSMO layers $\overline{T_{gr60}}$. Each cross corresponds to one field from training and validation data set. Right column: Relative reduction of RMSE and ME(STD) for rule 1 and 2. Each point corresponds to one downscaled field. The lines indicate a standard linear regression fit with r^2 denoting the coefficient of determination. For the right column only fields with $\overline{T_{gr60}} < 0.0058$ K/m are considered by the regression. Values within the shaded area are neglected.

and creates pronounced channel structures in the near-surface temperature field. The channel structures in the temperature field are well reproduced by both downscaling rules. However, the downscaled field is slightly smoother than the reference from the high-resolution model run.

For the nighttime field the objectives are less well reduced than for the daytime field (Table 6.7). This is not surprising as clear sky nighttime conditions create a more complex and temporally more variable fine-scale pattern compared to the approximately linear relation between topography and near-surface temperature for unstable atmospheres. ME(STD) and MIQD are reduced by $\geq 67\%$ for the nighttime situation. The RMSE is slightly increased by both downscaling rules, which is, however, barely visible (Fig. 6.16). Rule 2 performs slightly favorable for the nighttime temperature field especially concerning the RMSE (Table 6.7).

Figure 6.17 gives a more detailed look at the performance concerning different objectives by downscaling rule 1 for September 18$^{\text{th}}$ 2014 1:00 UTC. Figure 6.17 (a) shows that the overall pattern of temperature anomalies estimated by the downscaling rule 1 matches the reference pattern. However, the anomalies of the downscaled field are too smooth and/or slightly displaced compared to the reference. Hence, no reduction in RMSE is achieved. The subgrid-scale standard deviation is well reproduced (Fig. 6.17 (b)). The largest errors occur at Hambach (a brown coal surface mine) and at the outer mountain ranges of the Eifel region.

Here the downscaled field shows a larger subgrid-scale standard deviation than the reference. In the flat areas the downscaled field tends to show a smaller subgrid-scale standard deviation than the reference. Fig. 6.17 (c) illustrates the performance concerning the IQD. The difference plot on the right shows that the interpolated field has too few pixels with temperatures less than about 288 K as the cold air channels are not resolved. This difference in the CDFs is well reduced by the downscaling.
The downscaling performance depends on the atmospheric stability of the lower boundary layer (Fig. 6.18). As expected RMSE and ME(STD) between interpolated and high-resolution 10 m temperature are larger the larger the vertical temperature gradient. For the considered data set the relation is well approximated by a linear fit. For $\overline{T_{gr60}} < 0.0058$ K/m we find a strong relation between average vertical temperature gradient $\overline{T_{gr60}}$ and the relative reduction of the RMSE with a *coefficient of determination*[2] of $r^2 = 0.82$. The smaller the temperature gradient the larger the relative reduction of the RMSE. For larger positive temperature gradients, i.e., strong temperature inversions, the relative reduction of the RMSE stays approximately constant. For rule 1 some outliers occur. For the relative reduction of the ME(STD) similar tendencies as for the RMSE are observed, but the dependency between relative reduction of ME(STD) and temperature gradient is less strong. Both downscaling rules recover on average about 70% of the spatial variability for 10 m temperature.
The downscaling rule from *Schomburg et al.* (2010) has shown to perform well for unstable atmospheres (see also *Zerenner et al.*, 2016). We thus suggest the following combination of the *Schomburg* downscaling rule and rule 2 detected by MOGP:

IF $T_{gr105} < 0.0058\ Km^{-1}$
  $T^* = -0.0084 h^*$
ELSE
  IF $h^*\ T_{gr60} > -z_0^*$
    $T^* = 0.66\ h^*\ T_{gr60}$
  ELSE
    $T^* = 0.24\ z_0^* + 0.66\ h^*\ T_{gr60}$.

6.3.3. Specific Humidity

Overview

For downscaling 10 m specific humidity the relative reduction of the objectives is smaller than for temperature and pressure downscaling (Fig. 6.19(a)). The RMSE is on average increased by about 20% and only few downscaling rules achieve any reduction of the RMSE. The ME(STD) is on average reduced by about 25%. The MIQD is on average reduced by 30% to 50%, but some rules even increase the MIQD. For the RMSE the Pareto set tends to perform better for the training data set than for the validation data set (Fig. 6.19(b)). For the other objectives no systematic difference in performance is obvious.

[2] The *coefficient of determination* r^2 is the portion of the variation of the dependent variable that is explained by a fitted linear model. In a univariate linear regression, r^2 equals the squared Pearson correlation coefficient. For details on linear regression see Appendix B.1.

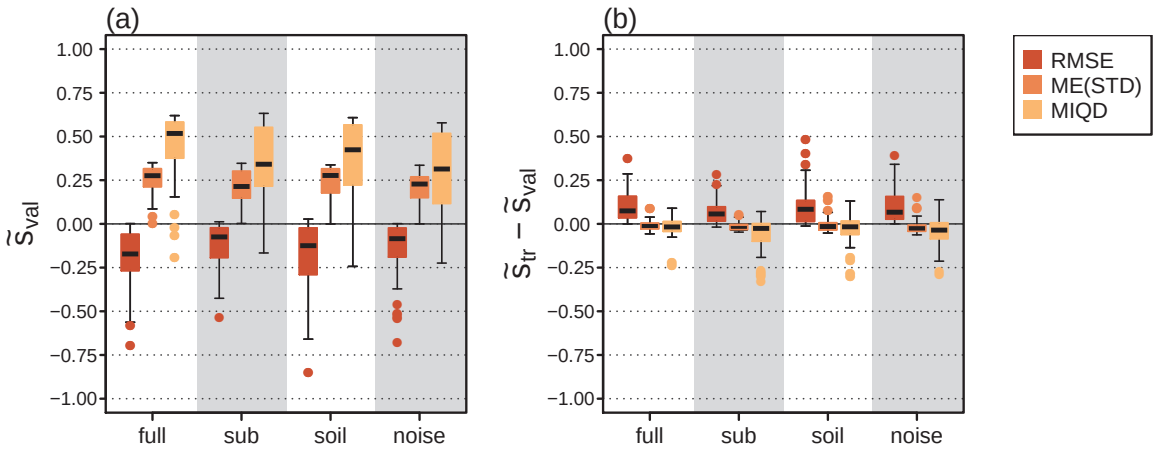

Figure 6.19.: Overview of the MOGP performance for downscaling 10 m specific humidity. (a) relative reduction of RMSE, ME(STD) and MIQD for the 50 rules from the Pareto set applied to the validation (val) data set; (b) difference in relative reduction between training (tr) and validation data set.

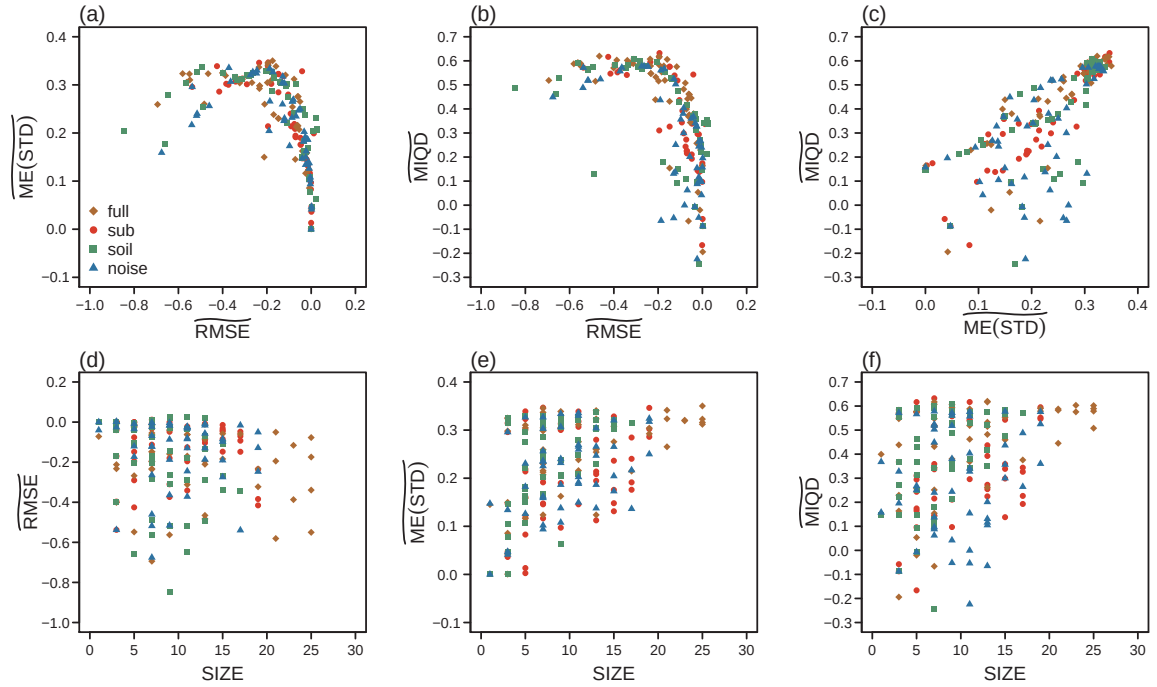

Figure 6.20.: Scatter plots of the relative reduction of RMSE, ME(STD) and MIQD for the Pareto sets for downscaling 10 m specific humidity when applied to the validation data. The bottom row additionally considers the solution SIZE (i.e., number of nodes) as fourth objective.

The difference between the results of the different MOGP setups is small (Fig. 6.19). Also when considering a trade-off between RMSE, ME(STD) and MIQD, none of the four MOGP setups is favorable (Fig. 6.20). MOGP$_{full}$ and MOGP$_{sub}$ perform slightly favorable for ME(STD) and MIQD, but at the cost of a larger RMSE. No distinct relation between the reduction of RMSE, ME(STD) and MIQD and the solution SIZE is found. The overall largest and thereby most complex downscaling rules are generated by MOGP$_{full}$.

The overall most used predictors are the vertical humidity gradients q_{gr}, the vertical temper-

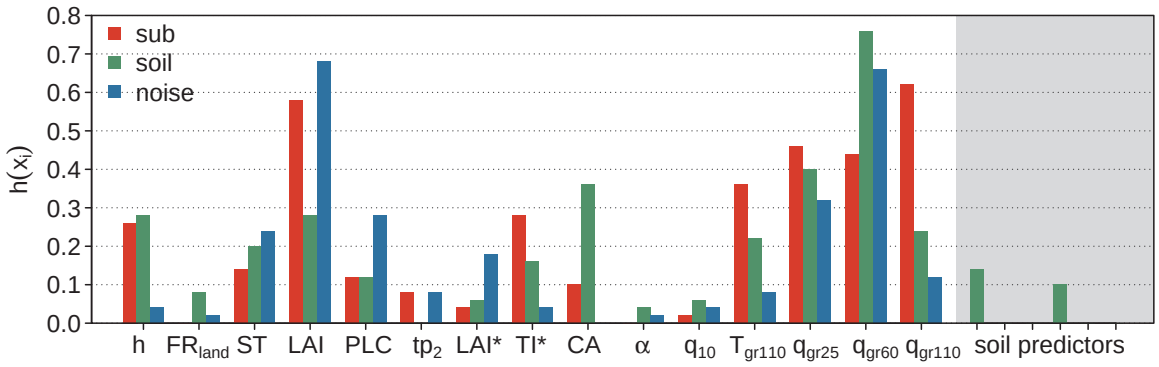

Figure 6.21.: Usage frequency $h(x_i)$ of the predictors (cf. Tables 6.4, 6.5) for downscaling 10 m specific humidity. The soil predictors are from left to right: T_{so}, W_{so}, W^*_{so}, T^*_{so}, $W_{so,7}$ and $W_{so,7}$.

ature gradient over the lowest 4 layers T_{gr110} as an indicator for the near-surface stability and the leaf area index LAI (Fig. 6.21). Especially the use of variables like leaf area index, plant cover and soil type is reasonable as evaporation from the soil and transpiration from plants affects the humidity near the surface.
From all soil predictors only T_{so} and W^*_{so} are used, but only by about 10% of the downscaling rules from MOGP$_{soil}$. As already suggested by Fig. 6.19 and 6.20 the soil predictors appear not to provide valuable additional information on the near-surface fine-scale specific humidity.

Selected Downscaling Rules

The parse trees of two downscaling rules selected based on their overall performance are shown in Fig. 6.22. Rule 1 is from MOGP$_{sub}$. Rule 2 is from MOGP$_{soil}$. For both rules the relative reduction of the objectives differs only slightly between training and validation (Table 6.8). Only for the RMSE we find a noticeable difference for rule 1 with a relative reduction of -0.09 for the training data set and -0.15 for the validation data set. Rule 2 shows a relative reduction between 0.02 and 0.04 for the RMSE. ME(STD) and IQD are reduced by 19% to 58% with rule 1 yielding a stronger reduction than rule 2.
Translating the parse tree of rule 1 (Fig. 6.22) into an equation we get:

$$q^* = q_{gr110}/CA \qquad \text{(rule 1)}.$$

As the vertical temperature gradient T_{gr110} is constant over each coarse pixel, adding T_{gr110} has no effect on the predicted specific humidity anomaly. Thus, rule 1 is the fraction of the vertical specific humidity gradient in the lowest four atmospheric layers q_{gr110} and the contributing area CA.
Rule 2 is slightly more complex than rule 1. Translated into an equation rule 2 (Fig. 6.22) reads:

$$q^* = T_{so}\ q_{gr25} + 10\ FR_{land}\ q_{gr60}\ T^*_{so} \qquad \text{(rule 2)}.$$

Here the vertical humidity gradients of the lowest 25 and 60 m, q_{gr25} and q_{gr60}, the frac-

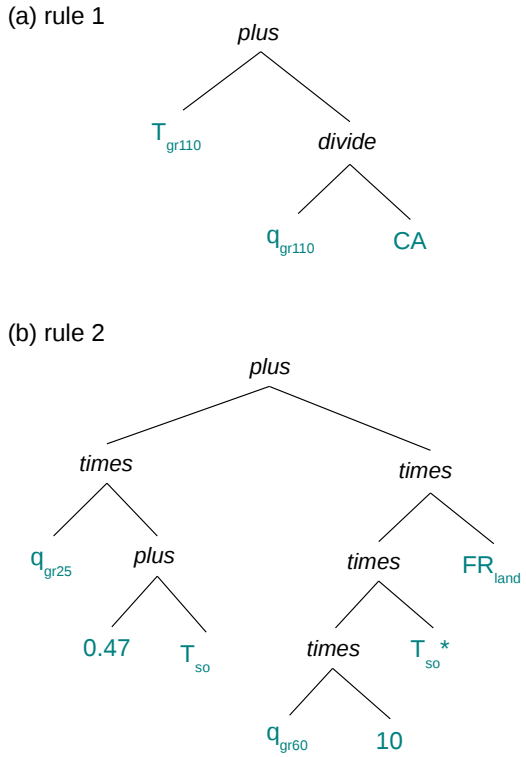

Figure 6.22.: Two downscaling rules returned by MOGP for downscaling 10 m specific humidity. Rule 1 is from the Pareto set returned by MOGP$_{sub}$. Rule 2 is from the Pareto set returned by MOGP$_{soil}$.

Table 6.8.: Relative reduction of the different objectives by the MOGP downscaling rules for 10 m specific humidity (Fig. 6.22) applied to training and validation data sets. The values for the training data set are given in brackets. The bottom two rows provide the relative reduction for the two fields in Figs. 6.23 (August 2nd 2015 07:00 UTC) and 6.24 (July 18th 2015 03:00 UTC).

(a) rule 1

	$\widetilde{\text{RMSE}}$	$\widetilde{\text{ME(STD)}}$	$\widetilde{\text{MIQD}}$
	-0.15	0.32	0.58
	(-0.09)	(0.31)	(0.56)
2015/08/02 7:00	-0.16	0.53	0.65
2015/07/18 3:00	-0.02	0.19	0.36

(b) rule 2

	$\widetilde{\text{RMSE}}$	$\widetilde{\text{ME(STD)}}$	$\widetilde{\text{MIQD}}$
	0.02	0.22	0.35
	(0.04)	(0.19)	(0.33)
2015/08/02 7:00	0.04	0.28	0.33
2015/07/18 3:00	0.01	0.06	0.17

tion of land FR_{land}, and soil temperature its anomaly, T_{so} and T^*_{so}, are used as predictors. Both downscaling rules are difficult to interpret. However, larger (positive or negative) vertical humidity gradients are accompanied by larger (positive or negative) humidity anomalies, which suggests that vertical and horizontal humidity variability are related. Still the effect of the available predictors on the near-surface specific humidity (anomalies) is only minor as indicated by the rather weak performance of the downscaling rules compared to near-surface temperature and pressure.

Figures 6.23 and 6.24 show the performance of the selected downscaling rules for two 10 m specific humidity fields extracted from the validation data set. Figure 6.23 shows the specific

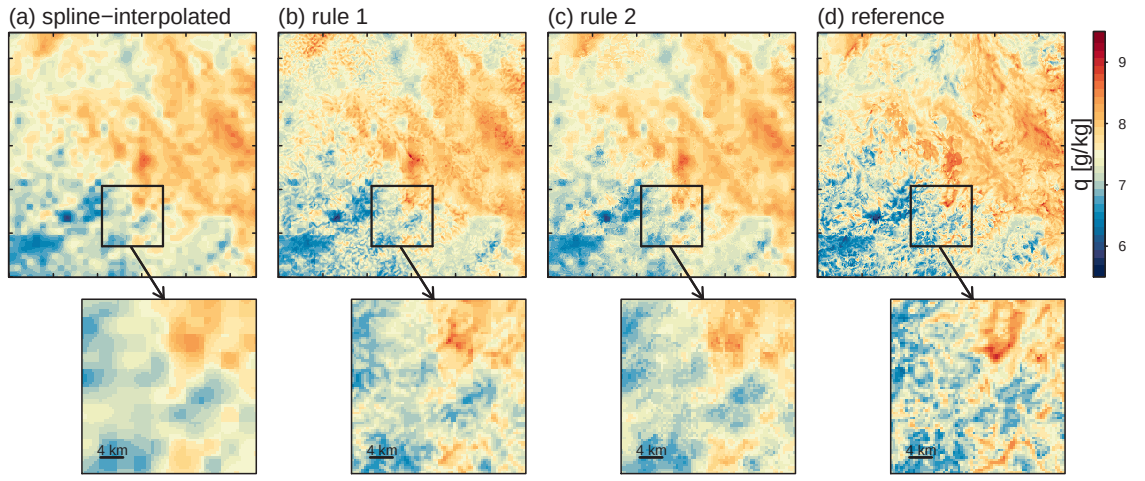

Figure 6.23.: 10 m specific humidity field on August 2nd 2015 07:00 UTC. (a) shows the spline-interpolated field; (b) and (c) show the downscaled fields resulting from rules 1 and 2 (cf. Fig. 6.22); (d) shows the reference field from the high-resolution model run.

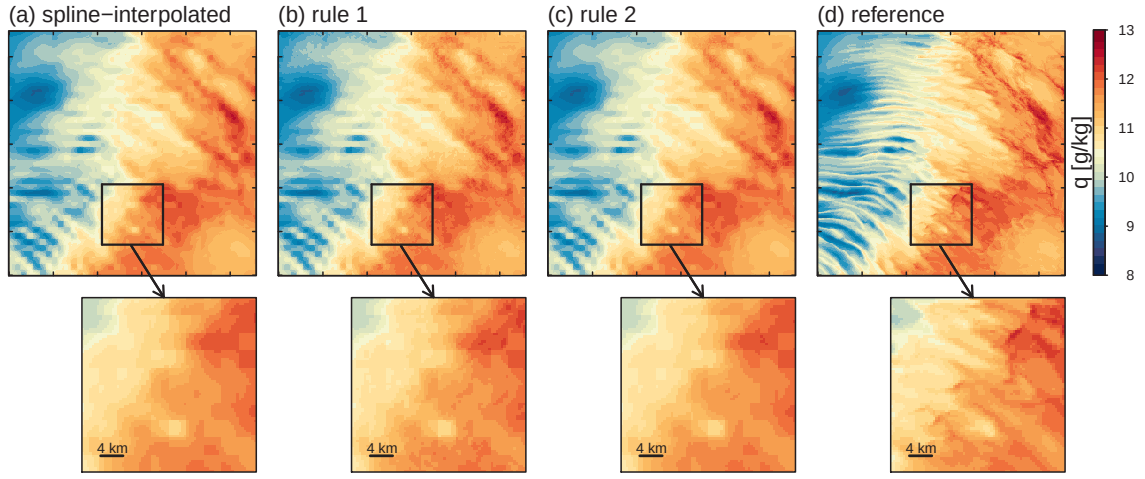

Figure 6.24.: 10 m specific humidity field on July 18th 2015 03:00 UTC. (a) shows the spline-interpolated field; (b) and (c) show the downscaled fields resulting from rules 1 and 2 (cf. Fig. 6.22); (d) shows the reference field from the high-resolution model run.

humidity in 10 m height on August 2nd 2015 07:00 UTC, an overall calm day with weak winds and little cloud cover. Note that, like for the other variables, red color indicates high values, i.e., high specific humidity, and blue colors indicate low values, i.e., little specific humidity. In the flat areas of the domain relatively high humidity is observed on August 2nd 2015 07:00 UTC. Over the mountainous Eifel the specific humidity is lower. We see that both downscaling rules recover some subgrid-scale variability especially over the Eifel where the overall fine-scale variability is highest. It is, however, also obvious that both downscaling rules underestimate the subgrid-scale variability. For rule 1 the relative reduction of ME(STD) and IQD equals 0.53 and 0.65, respectively (Table 6.8). For rule 2 the relative reduction is even smaller. The stronger reduction of ME(STD) and MIQD by rule 1 comes with a 16% increase of the RMSE, while rule 2 slightly reduces the RMSE by about 4%.

Figure 6.24 shows the 10 m specific humidity on July 18th 2015 03:00 UTC, an overall clear

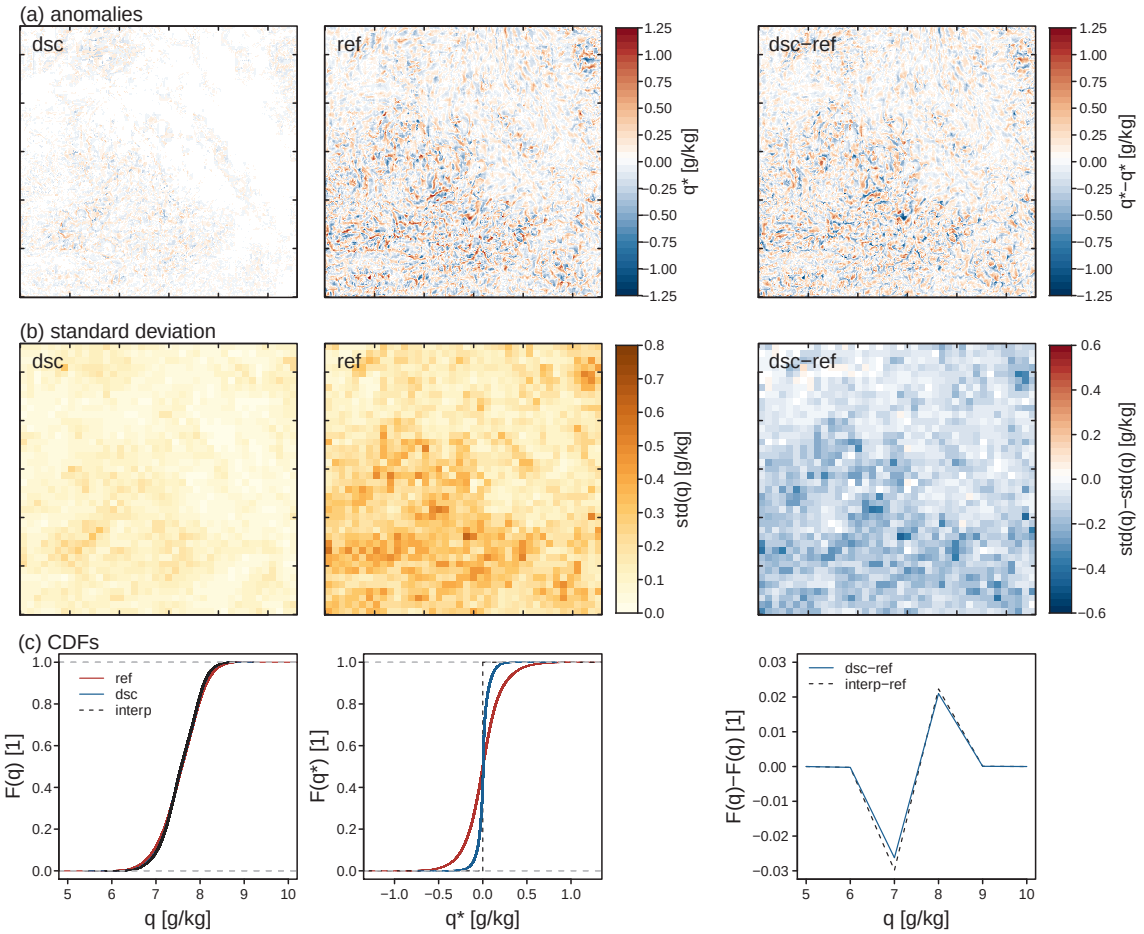

Figure 6.25.: Detailed evaluation of the MOGP performance for downscaling the 10 m specific humidity field on August 2nd 2015 07:00 UTC (cf. Fig. 6.23): (a) illustrates the performance concerning the RMSE and provides the specific humidity anomalies q^* for downscaled (dsc) and reference field (ref) and the respective difference plot; (b) illustrates the performance concerning the ME(STD) and provides the fine-scale standard deviation of downscaled and reference humidity fields q together with the respective error; (c) shows the CDFs of downscaled and reference humidity (and anomaly) fields and the difference in CDFs.

sky day. The wavelike structure in the reference field shows how gravity waves influence the humidity field by vertical and horizontal transport of air masses. Such wave structures, though potentially generated by topography, can not be directly (locally and instantaneously) predicted from the surface heterogeneity. Here both downscaling rules generate only little subgrid-scale variability with a relative reduction of ME(STD) of only 0.19 and 0.06 for rule 1 and rule 2, respectively (Table 6.8).

Figure 6.25 provides a more detailed look at performance concerning the different objectives by downscaling rule 1 for the example field from Fig. 6.23. Large anomalies in reference and downscaled fields occur in the same areas, preferably areas with pronounced topography (Fig. 6.25 (a)). The magnitude of the anomalies and thereby the subgrid-scale standard deviation are, however, strongly underestimated by the downscaling (Fig. 6.25 (b)). The effect of the downsclaling on the CDFs is hardly visible (Fig. 6.25 (c)).

Looking at the different fields of the validation data set has shown that larger 10 m wind

speeds typically go along with less subgrid-scale variability in the 10 m relative humidity field, i.e., a smaller RMSE and ME(STD) when comparing high-resolution reference and spline-interpolated humidity fields. The performance of the downscaling rules shows only a small dependency on near-surface wind speed, but the relative reduction of the ME(STD) is by tendency larger for calm wind conditions, i.e., for fields with stronger fine-scale variability. Comparing rules 1 and 2 strong outliers concerning the relative reduction of the RMSE are found for rule 1 (up to -0.5) while for rule 2 the relative reduction of the RMSE ranges between -0.1 and 0.1. For the majority of fields from the validation data set rule 2 reduces the RMSE. Thus, the usage of rule 2 is suggested. rule 2 on average recovers only 20% to 30% of the subgrid-scale variability, but on average without any increase in RMSE.

6.3.4. Wind Speed

Overview

For wind speed the ME(STD) and the MIQD are reduced by up to 65% and 85%, respectively (Fig. 6.26). Only few downscaling rules reduce the RMSE. The best rules for the RMSE achieve a reduction of about 10%. For MOGP$_{sub}$ and MOGP$_{soil}$ the performance concerning ME(STD) and MIQD is found to be slightly favorable over MOGP$_{full}$. For wind speed downscaling using the smaller predictor set appears to be beneficial. The inclusion of a basic noise generator has no visible effect on the performance. For MOGP$_{full}$ a few rules perform better for the training data set than for the validation data set. However, none of the three MOGP setups shows any systematic overfitting.
MOGP$_{sub}$ and MOGP$_{noise}$ yield a better trade-off between the different objectives than MOGP$_{full}$ (Fig. 6.27). The performance concerning RMSE, ME(STD) and MIQD is only slightly dependent on the solution SIZE. Solutions with about 10 nodes perform as well as more complex solutions containing more nodes. The most complex rules are generated by MOGP$_{noise}$.

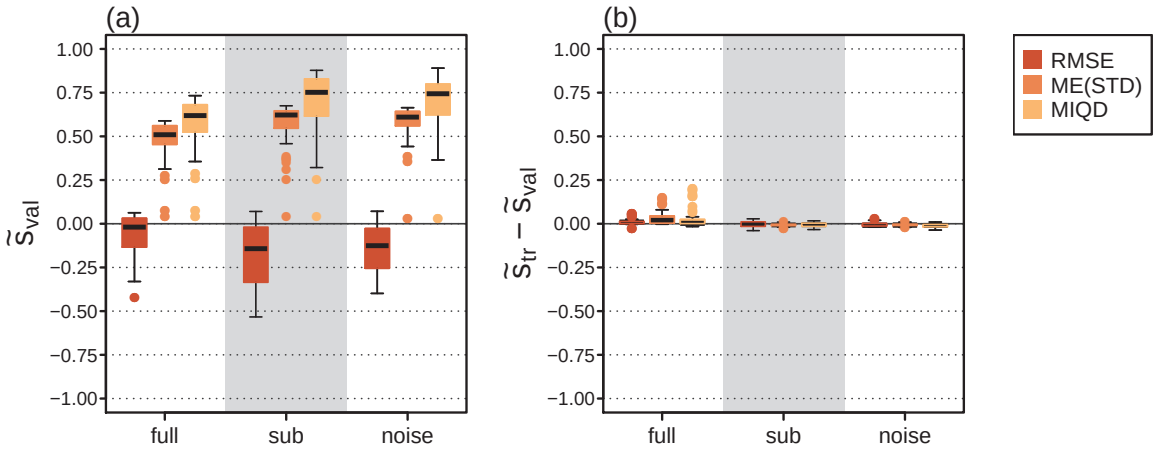

Figure 6.26.: Overview of the MOGP performance for downscaling 10 m horizontal wind speed. (a) relative reduction of RMSE, ME(STD) and MIQD for the 50 rules from the Pareto set rules applied to the validation (val) data set; (b) difference in relative reduction between training (tr) and validation data set.

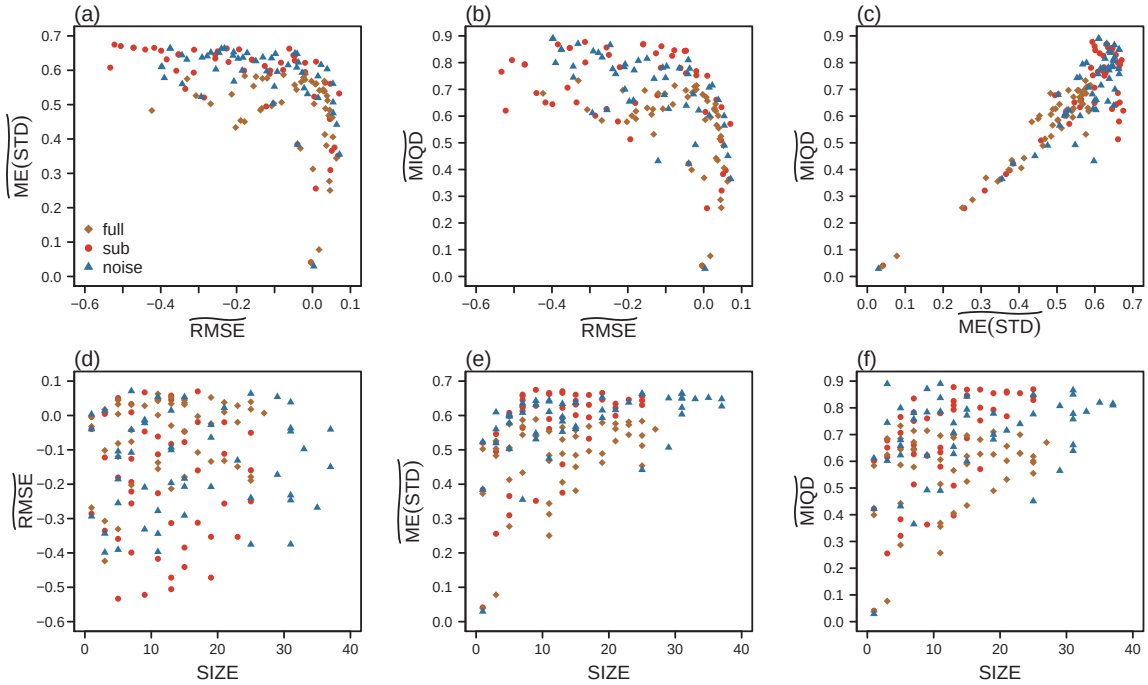

Figure 6.27.: Scatter plots of the relative reduction of RMSE, ME(STD) and MIQD of the Pareto sets for downscaling 10 m horizontal wind speed applied to the validation data. The bottom row additionally considers the solution SIZE (i.e., number of nodes) as fourth objective.

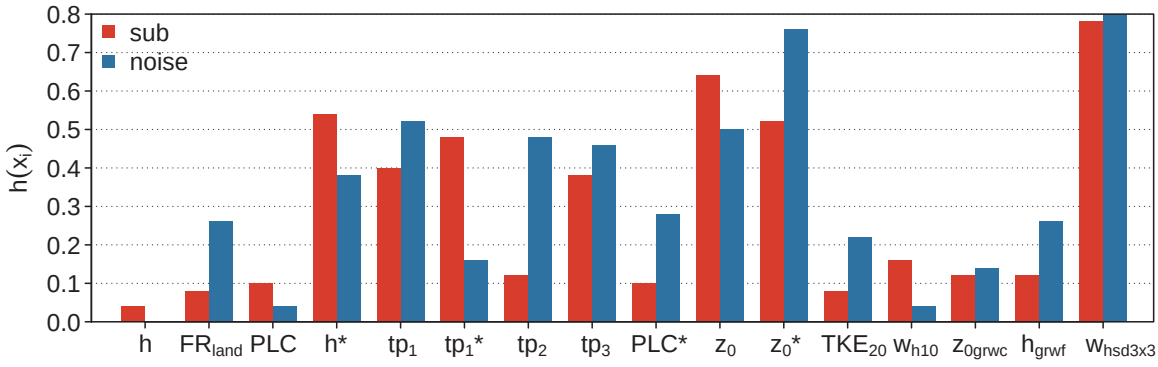

Figure 6.28.: Usage frequency $h(x_i)$ of the predictors (cf. Tables 6.4, 6.5) for downscaling 10 m horizontal wind speed.

The most frequently used predictors are the coarse-scale standard deviation of the 10 m horizontal wind speed $w_{h,sd3\times3}$, the roughness length z_0 and the roughness length anomaly z_0^* (Fig. 6.13). This is physically reasonable. Over rough surfaces (e.g., woods, cities) the near-surface wind speed is smaller than over more even surfaces (e.g., low grassland, sea). The frequent usage of the coarse scale standard deviation implies that there is a relation between the spatial variability on the coarse and on the fine scale as already assumed and used by *Schomburg et al.* (2010). Further frequently used predictors are topographic height anomaly h^* and the various predictors derived from topography, which is reasonable as also the location of a pixel within the terrain (e.g., luv or lee, mountain top or valley) effects the local wind speed near the surface.

Table 6.9.: Relative reduction of the different objectives achieved by the MOGP downscaling rules for 10 m horizontal wind speed from Fig. 6.29 applied training and validation data sets. The values for the training data set are given in brackets. The bottom two rows provide the relative reduction for the two fields in Figs. 6.30 (January 10th 2015 0:00 UTC) and 6.31 (June 5th 2015 15:00 UTC).

(a) rule 1

	$\widetilde{\text{RMSE}}$	$\widetilde{\text{ME(STD)}}$	$\widetilde{\text{MIQD}}$
	-0.08	0.62	0.81
	(-0.07)	(0.61)	(0.78)
2015/01/10 15:00	-0.02	0.60	0.85
2015/06/05 0:00	-0.18	0.61	0.77

(b) rule 2

	$\widetilde{\text{RMSE}}$	$\widetilde{\text{ME(STD)}}$	$\widetilde{\text{MIQD}}$
	-0.05	0.64	0.84
	(-0.07)	(0.64)	(0.84)
2015/01/10 15:00	0.03	0.61	0.72
2015/06/05 0:00	-0.15	0.66	0.88

Selected Downscaling Rules

In the following the two downscaling rules are selected and analyzed that yield the highest sum of the relative reduction of RMSE, ME(STD) and MIQD (Fig. 6.29). Rule 1 is from MOGP$_{full}$. Rule 2 from MOGP$_{sub}$. Both downscaling rules perform about equally well for training and validation data sets (Table 6.9). The ME(STD) is reduced by $\geq 0.61\%$. The relative reduction of the MIQD ranges between 0.78 and 0.84. The RMSE is with a relative reduction between -0.05 and -0.08 on average slightly increased.

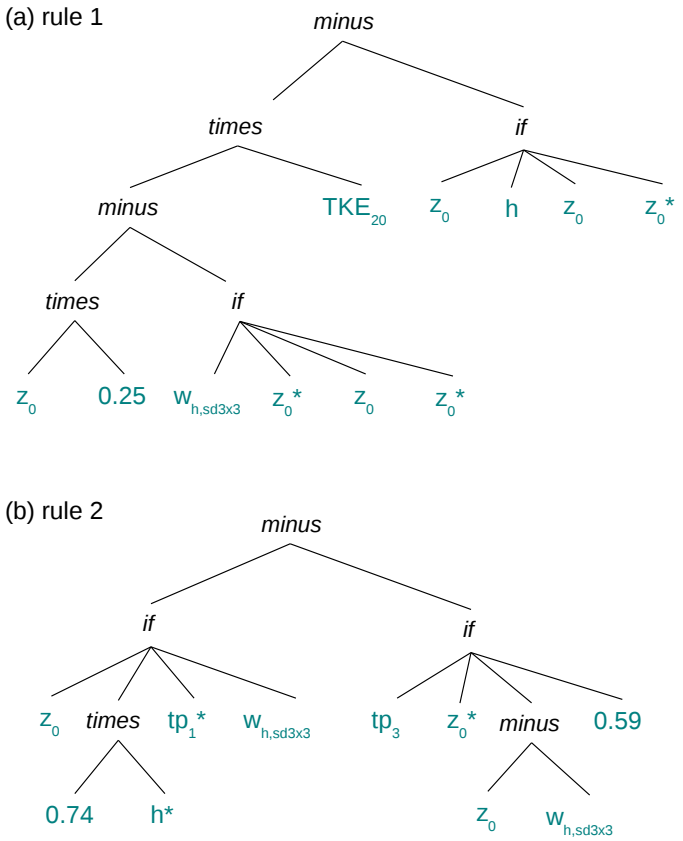

Figure 6.29.: Two downscaling rules returned by MOGP for downscaling 10 m horizontal wind speed. Rule 1 is from the Pareto set returned by MOGP$_{full}$. Rule 2 is from the Pareto set returned by MOGP$_{sub}$.

Translating the parse tree of rule 1 (Fig. 6.29) into equations gives:

IF $w_{h,sd3\times3} > z_0^*$ AND $z_0 > h$ (rule 1)
$w_h^* = (-0.75\ TKE_{20} - 1)\ z_0$
ELSE IF $w_{h,sd3\times3} > z_0^*$ AND $z_0 \leq h$
$w_h^* = -0.75\ z_0\ TKE_{20} - z_0^*$
ELSE IF $w_{h,sd3\times3} \leq z_0^*$ AND $z_0 > h$
$w_h^* = (0.25\ TKE_{20} - 1)\ z_0 - TKE_{20}\ z_0*$
ELSE
$w_h^* = 0.25\ TKE_{20}\ z_0 - (1 + TKE_{20})\ z_0^*$.

With two intricate if-statements the full downscaling rule appears quite complex at first. Still the predictors used, the roughness length z_0, roughness length anomaly z_0^* and the near-surface turbulent kinetic energy TKE_{20}, appear reasonable. The conditional statements further incorporate the coarse-scale standard deviation of the near-surface wind speed $w_{h,sd3\times3}$ and the topographic height h.
The first or third equation is used when the roughness length is larger than the topographic height ($z_0 > h$) which is rarely the case (exception being the brown coal surface mine Hambach, cf. Fig. 6.3). For the most part of the domain the second or fourth equation is used depending on the coarse-scale standard deviation of the horizontal wind speed $w_{h,sd3\times3}$. For $w_{h,sd3\times3} > z_0^*$ the second equation is used, where large positive roughness lengths z_0 and roughness length anomalies z_0^* lead to large negative wind speed anomalies, which is physically reasonable. A small roughness length z_0 and a large negative z_0^* lead to positive wind speed anomalies. For $w_{h,sd3\times3} \leq z_0^*$, i.e., for situations and areas with little spatial variability of the near-surface wind on the coarse scale, the fourth equation is used. Again positive roughness length anomalies z_0^* lead to negative wind speed anomalies. The relation in z_0 is not intuitive at first glance. Here large z_0 appears to be accompanied by positive wind speed anomalies. However, given that the coarse pixel mean of the wind speed is conserved during the downscaling, this term should be rather interpreted as reducing the anomalies resulting from the previously discussed term incorporating z_0^*. For situations with little variance on the coarse scale less fine-scale variability is generated, which again appears reasonable.
After simplification rule 2 (Fig. 6.29) reads:

IF $z_0 > 0.74\ h^*$ AND $tp_3 > z_0*$ (rule 2)
$w_h^* = tp_1^* - z_0$
ELSE IF $z_0 > 0.74\ h^*$ AND $tp_3 \leq z_0*$
$w_h^* = tp_1^*$
ELSE IF $z_0 \leq 0.74\ h^*$ AND $tp_3 > z_0*$
$w_h^* = -z_0$
ELSE
$w_h^* = 0$.

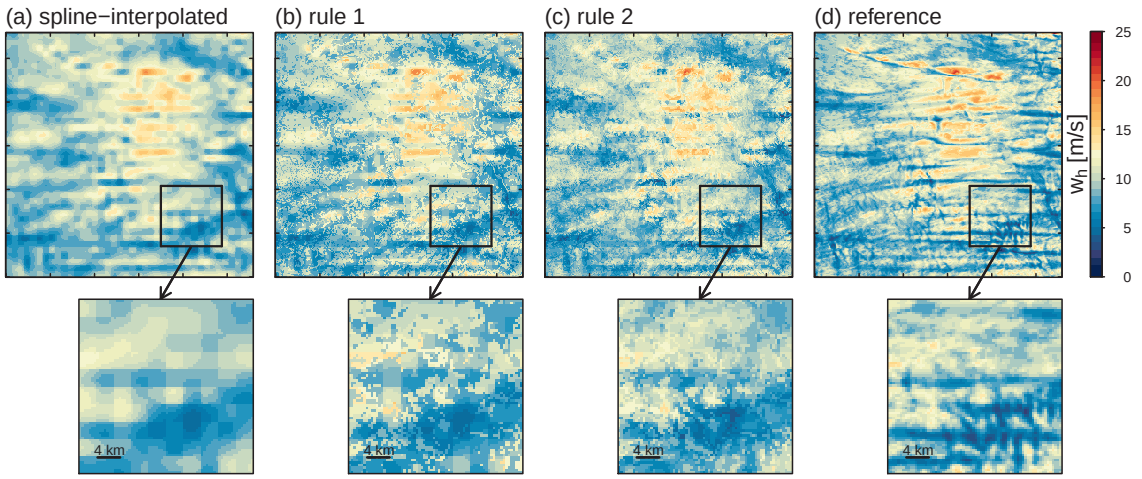

Figure 6.30.: Horizontal wind speed in 10 m height on January 10$^{\text{th}}$ 2015 15:00 UTC. (a) shows the spline-interpolated field; (b) and (c) show the downscaled fields resulting from rules 1 and 2 (cf. Fig. 6.29); (d) shows the reference field from the high-resolution model run.

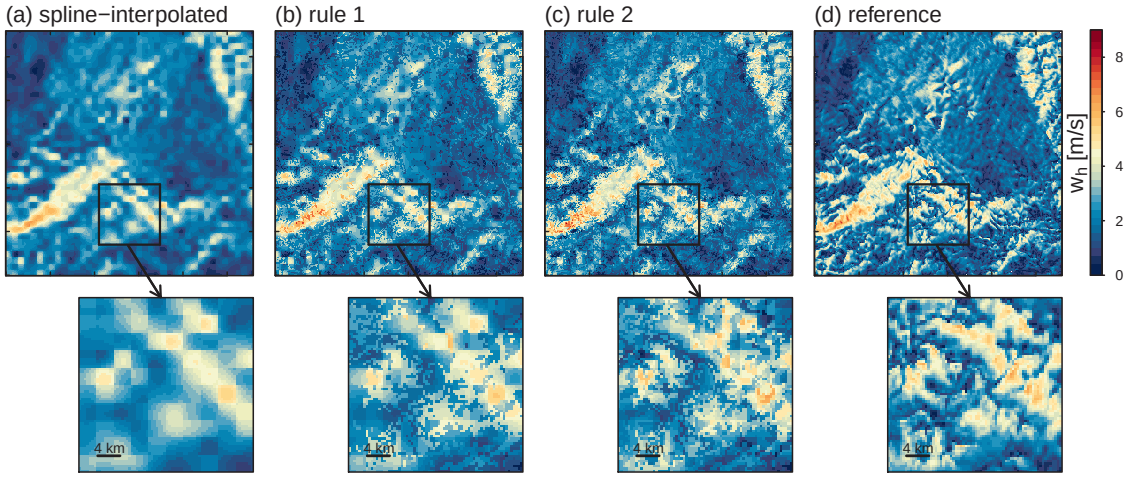

Figure 6.31.: Horizontal wind speed in 10 m height on June 5$^{\text{th}}$ 2015 0:00 UTC. (a) shows the spline-interpolated field; (b) and (c) show the downscaled fields resulting from rules 1 and 2 (cf. Fig. 6.29); (d) shows the reference field from the high-resolution model run.

While the if-statements incorporating topography predictors tp and roughness length z_o are difficult to interpret, the different downscaling equations are simple. A positive tp_1^* (pixels which are higher than the average direct neighboring pixel) leads to a positive wind speed anomaly. A negative tp_1^* leads to a negative wind speed anomaly. This is physically reasonable. The near surface wind speed in a valley, for instance, is typically lower than for the surrounding pixels. Further, the larger the roughness length z_0 the larger the negative wind speed anomaly.

Figures 6.31 and 6.30 show the performance of the selected downscaling rules for 10 m wind fields extracted from the validation data set. Figure 6.30 shows the 10 m wind speed on January 10$^{\text{th}}$ 2015 15:00 UTC, a cloudy and windy day with strong precipitation. The gravity waves, which generate the major part of the spatial variability of the horizontal wind speed, are well resolved in the reference field, but not captured by the downscaling. This is not surprising.

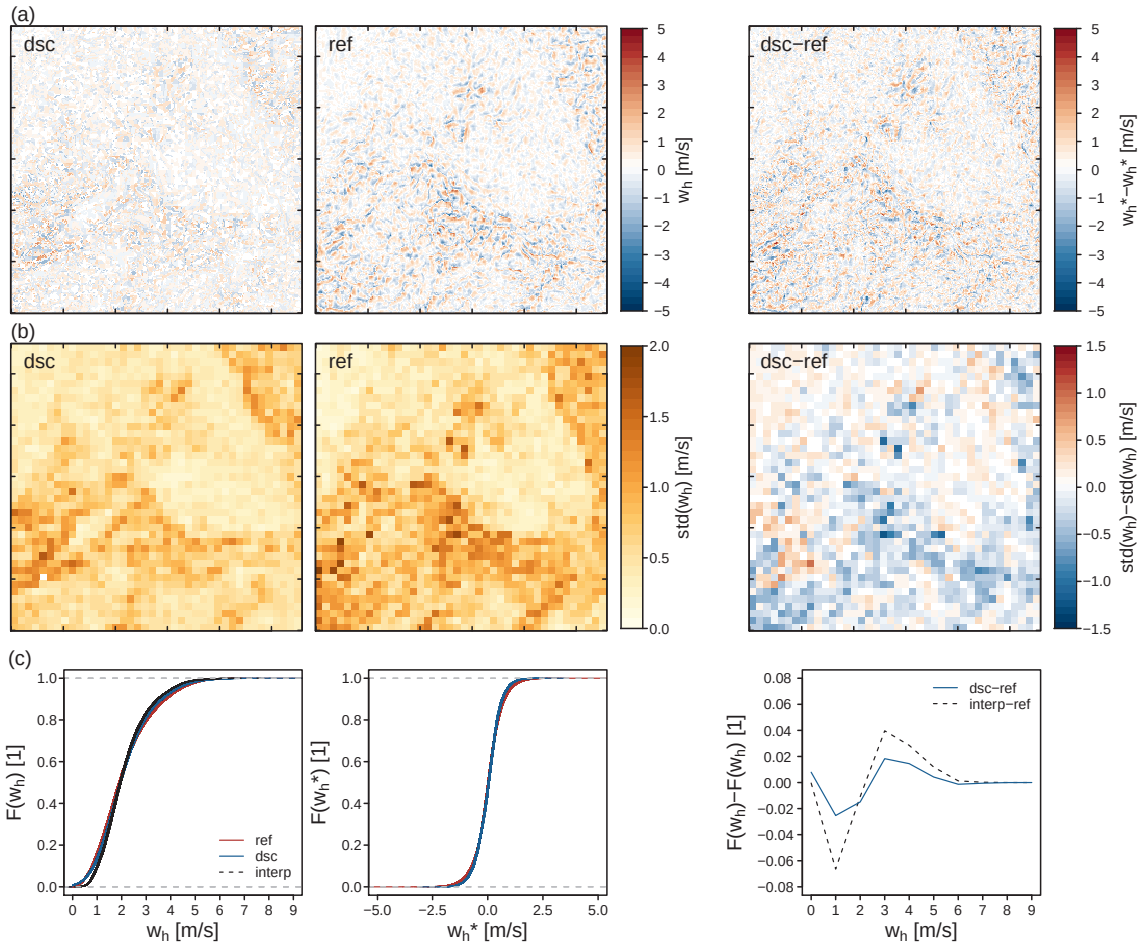

Figure 6.32.: Detailed evaluation of the MOGP performance for downscaling the 10 m horizontal wind speed on June 5$^{\text{th}}$ 2015 0:00 UTC (cf. Fig. 6.31): (a) illustrates the performance concerning the RMSE and provides the wind speed anomalies w_h^* for downscaled (dsc) and reference field (ref) and the respective difference plot; (b) illustrates the performance concerning the ME(STD) and provides the fine-scale standard deviation of downscaled and reference wind speed w_h together with the respective error; (c) shows the CDFs of downscaled and reference wind speed (and wind speed anomalies) and the difference in CDFs.

Gravity waves can be generated by surface heterogeneity, but propagate in space and time. Thus, no steady relation between instantaneous wind fields and land surface properties can be established. The anomalies predicted by the downscaling rules appear nevertheless physically meaningful as a significant part of the fine-scale variability (relative reduction of ME(STD) and MIQD ≥ 0.6) is reproduced without increasing the RMSE (Table 6.9).

Figure 6.31 shows the 10 m wind speed on June 5$^{\text{th}}$ 2015 0:00 UTC, a rather warm, calm and mostly clear sky day. In the flatter areas of the model domain the wind speed is mostly < 4 m/s. Only in the mountainous regions slightly larger wind speeds are observed. Both downscaling rules generate some subgrid-scale variability. The fine-scale pattern, however, visibly differs from the reference. The RMSE is increased by 18% and 15% by rule 1 and rule 2, respectively. Concerning the two objectives quantifying the spatial variability of the 10 m wind a relative reduction ≥ 0.6 is observed (Table 6.9).

Figure 6.32 illustrates the downscaling performance of rule 2 concerning the different objec-

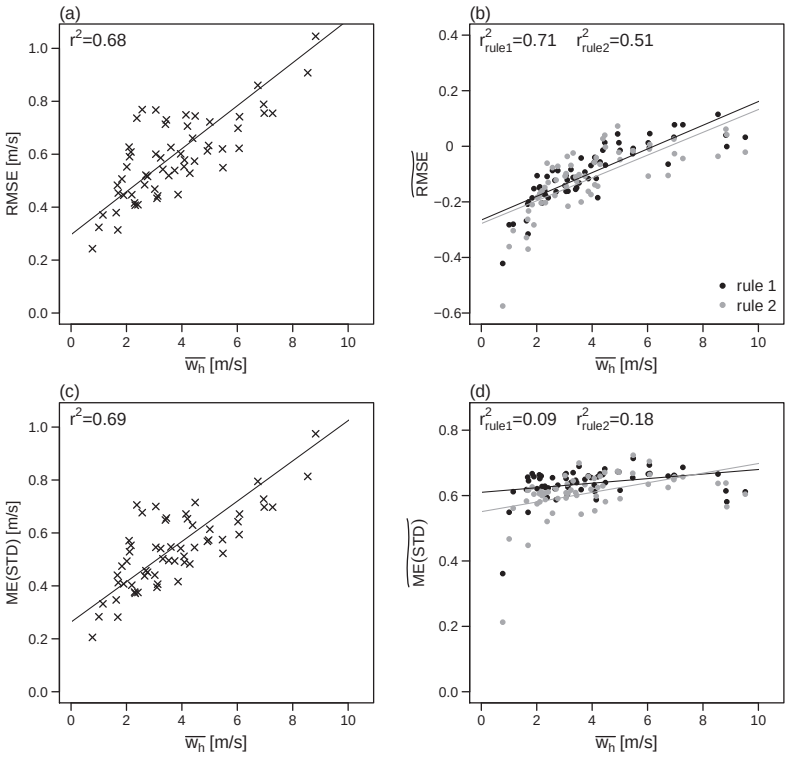

Figure 6.33.: Left column: RMSE and ME(STD) of the spline-interpolated wind fields compared to the high-resolution reference fields versus the spatially averaged wind speed $\overline{w_h}$. Each cross corresponds to one field from training or validation data set. Right column: Relative reduction of RMSE and ME(STD) for rule 1 and 2. Each point corresponds to one downscaled field. The lines indicate a standard linear regression fit with r^2 denoting the coefficient of determination.

tives for the wind fields from Fig. 6.31. Large anomalies are present in the same areas in downscaled and reference field (Fig. 6.32(a)). However, as already seen in Fig. 6.31 the fine-scale patterns clearly differ. In the mountainous regions rule 2 underestimates the subgrid-scale standard deviation (Fig. 6.32(b)). The largest errors occur at the outer mountain ranges of the Eifel region. A clear reduction of the MIQD is obvious from the difference plot between the CDFs (Fig. 6.32(c) right).

The downscaling performance depends on the overall wind conditions (Fig. 6.33). As expected RMSE and ME(STD) between interpolated and high-resolution wind speed fields are larger, the larger the average wind speed (Fig. 6.33(a),(c)). For the considered data the relation is found to be approximately linear with $r^2 = 0.68$ and $r^2 = 0.69$, respectively. The relative reduction of the RMSE depends on the average wind speed (Fig. 6.33(b)). For both downscaling rules the larger the wind speed, the more likely the RMSE is reduced. The relative reduction of the ME(STD) is largely independent of the average wind speed. Both downscaling rules predict a portion of $\approx 60\%$ of the spatial variability for all fields considered (Fig. 6.33(d)).

Based on Fig. 6.33 rule 1 appears to be all in all slightly favorable as it shows less strong outliers. However, for lower wind speeds $\overline{w_h} < 4$ m/s an increase in RMSE by about 10% to 20% has to be accepted. For higher wind speeds the RMSE is kept constant by the downscaling or even reduced by up to 10%. For applications the downscaling rule 1 can be simplified to

IF $w_{h,sd3\times3} > z_0^*$
$w_h^* = -0.75\ z_0\ TKE_{20} - z_0^*$
ELSE
$w_h^* = 0.25\ TKE_{20}\ z_0 - (1 + TKE_{20})\ z_0^*$.

with only marginal impact on the performance as discussed in detail above.

6.3.5. Radiation

Overview

The incoming radiation at the land surface is the result of the complex process of radiative transport in the atmosphere. The δ-two stream approximation used in COSMO and many other atmospheric models only considers the vertical propagation of radiation. Potential horizontal components are neglected. The radiative transfer parameterization calculates to the propagation of radiation through an atmospheric column from the model top to the land-surface. Direct effects of topographic shadowing or topographic slope and aspect on incoming radiation are neglected in the operational COSMO setup and, to stay consistent, also in the 400 m COSMO simulations. The relation between surface heterogeneity and incoming radiation at the surface is thus expected to be small.
Figure 6.34 provides an overview of the MOGP performance for downscaling incoming longwave radiation LWR, incoming direct shortwave radiation SWR_{dir} and incoming diffuse radiation at the surface SWR_{dif}. For incoming longwave radiation the MIQD is reduced by up to 55% and the ME(STD) by up to 35%. The RMSE is on average increased. No downscaling rules achieves any visible reduction in RMSE. Also for the shortwave radiation no downscaling rule visibly reduced the RMSE. For the shortwave radiation the relative reduction of

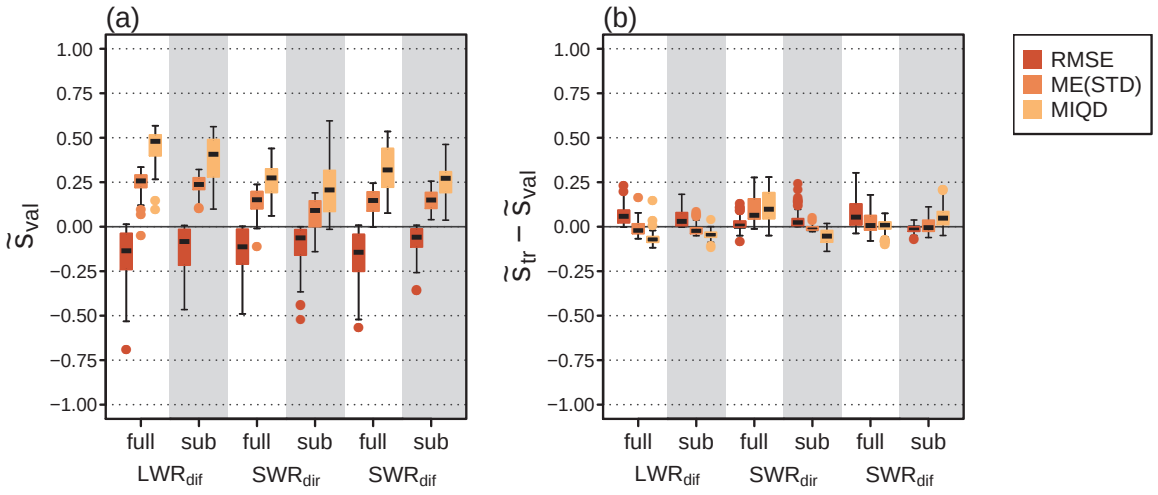

Figure 6.34.: Overview of the validation of the Pareto sets returned by MOGP for downscaling incoming longwave radiation SWR_{dif}, incoming direct shortwave radiation SWR_{dir} and incoming diffuse radiation at the surface SWR_{dif}: (a) relative improvement of RMSE, ME(STD) and MIQD for the Pareto set rules applied to the validation (val) data set; (b) difference in relative improvement between training (tr) and validation data set.

ME(STD) and MIQD is overall smaller than for the longwave radiation. The differences in performance between training and validation data set are for radiation larger than for most other variables considered so far (Fig. 6.34(b)). Still, with the exception of MOGP$_{full}$ for downscaling incoming direct short wave radiation, no systematic overfitting tendencies are found.
For downscaling incoming diffuse longwave radiation the most frequently used predictors are soil type ST and leaf area index LAI, followed by topographic index anomaly TI^*, topographic height anomaly h^* and plant cover PLC. For incoming direct shortwave radiation the most frequently used predictors are the coarse-scale incoming direct shortwave radiation at the surface $SWR_{dir,s}$ and the cloud cover of medium high clouds CLC_m followed by topography related predictors tp_2, tp_1^*, topographic height h, topographic height anomaly h^* and albedo anomaly α^*. For incoming diffuse shortwave radiation the most frequently used predictors are the coarse-scale incoming diffuse shortwave radiation at the surface $SWR_{dif,s}$, the net radiation at the surface R_{net}, topography parameter tp_1^*, the total cloud cover CLC_t and the albedo anomaly α^*. Especially the usage of surface albedo α, i.e., the portion of incoming solar radiation reflected by the surface, and the use of the cloud cover predictors CLC appear reasonable as those variables physically affect the radiation partitioning in an atmospheric column.

Selected Downscaling Rules

In the following we take a closer look at the downscaling rules with the largest sum of relative reduction of RMSE, ME(STD) and MIQD for incoming diffuse longwave radiation, direct shortwave radiation and diffuse shortwave radiation at the land surface. The parse trees of these downscaling rules are shown in Figs. 6.35 and 6.38. The values of the relative reduction of the objectives are listed in Tables 6.10 and 6.11.

Longwave Radiation For incoming longwave radiation two downscaling rules were selected. Both downscaling rules cause an increase in RMSE by 7% and 3%, respectively. Concerning ME(STD) and IQD the relative reduction ranges between 0.26 and 0.44. The difference in performance between training and validation data set is overall small.
After removing all unnecessary terms and evaluating the constants, rule 1 (Fig. 6.35) reads:

$$LWR^* = (110 - 0.62\ CAPE_{mu} - 2\ h^*))/\alpha \qquad \text{(rule 1)}$$

This implies that large positive topographic height anomalies h^* lead to large negative longwave radiation anomalies LWR^*. Further, a small (shortwave) albedo α is accompanied by larger, positive or negative, LWR^*.
After simplification rule 2 (Fig. 6.35) reads:

$$LWR^* = (CLC_{t,sd3\times3} - h^*)/\alpha + ST - 100/TI \qquad \text{(rule 2)}$$

The first term is similar to rule 1. Instead of $CAPE$ here the coarse-scale variability of

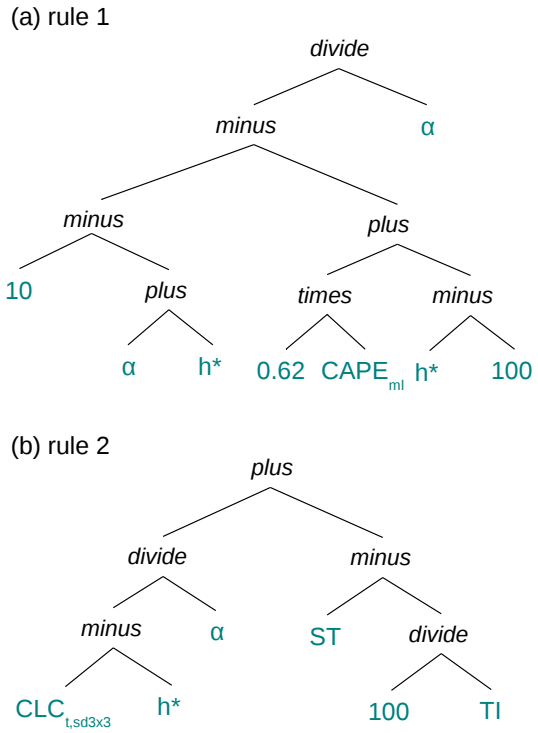

Figure 6.35.: Two downscaling rules returned by MOGP for downscaling incoming longwave radiation at the land surface. Both rules are from MOGP$_{full}$.

the total cloud cover $CLC_{t,sd3\times3}$ is used with a large cloud cover variability being associated with large positive longwave radiation anomalies LWR^*. The third term $-100/TI$ implies that a small topographic index TI is accompanied by a large negative longwave radiation anomalies.

Figures 6.36 and 6.37 show the performance of the selected downscaling rules for two longwave radiation fields extracted from the validation data set. Figure 6.36 shows the incoming longwave radiation on August 2$^{\text{nd}}$ 2015 07:00 UTC, a calm day with few clouds and weak winds. The incoming longwave radiation ranges between 300 and 420 W/m^2. The largest variability is present in the areas with spatially variable cloud cover. Especially in the north eastern part of the domain and also in the south eastern part the cloud structures are clearly

Table 6.10.: Relative reduction of the objectives by the MOGP downscaling rules for incoming longwave radiation (Fig. 6.14) applied to training and validation data sets. The values for the training data set are given in brackets. The bottom two rows provide the relative reduction of the objectives for the two fields in Figs. 6.36 (August 2$^{\text{nd}}$ 2015 07:00 UTC) and 6.37 (June 13$^{\text{th}}$ 2015 21:00 UTC).

(a) rule 1

	$\widetilde{\text{RMSE}}$	$\widetilde{\text{ME(STD)}}$	$\widetilde{\text{MIQD}}$
	-0.07 (-0.06)	0.26 (0.27)	0.44 (0.48)
2015/08/02 7:00	-0.12	0.51	0.63
2015/06/13 12:00	-0.70	-0.14	-0.17

(b) rule 2

	$\widetilde{\text{RMSE}}$	$\widetilde{\text{ME(STD)}}$	$\widetilde{\text{MIQD}}$
	-0.03 (-0.03)	0.30 (0.30)	0.40 (0.41)
2015/08/02 7:00	-0.02	0.47	0.48
2015/06/13 21:00	-0.03	0.23	0.30

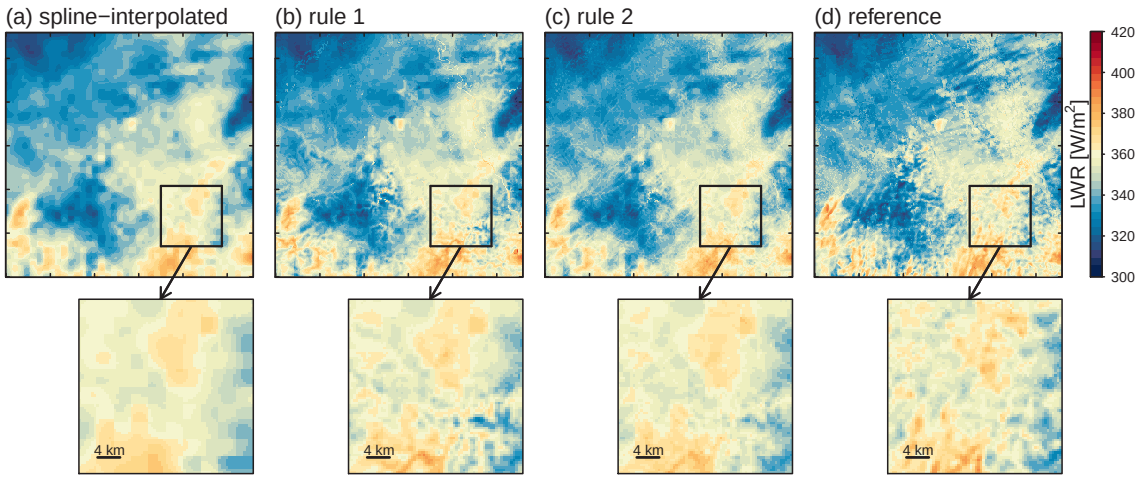

Figure 6.36.: Incoming longwave radiation at the surface on August 2nd 2015 07:00 UTC. (a) shows the spline-interpolated field; (b) and (c) show the downscaled fields resulting from rules 1 and 2 (cf. Fig. 6.35); (d) shows the reference field from the high-resolution model run.

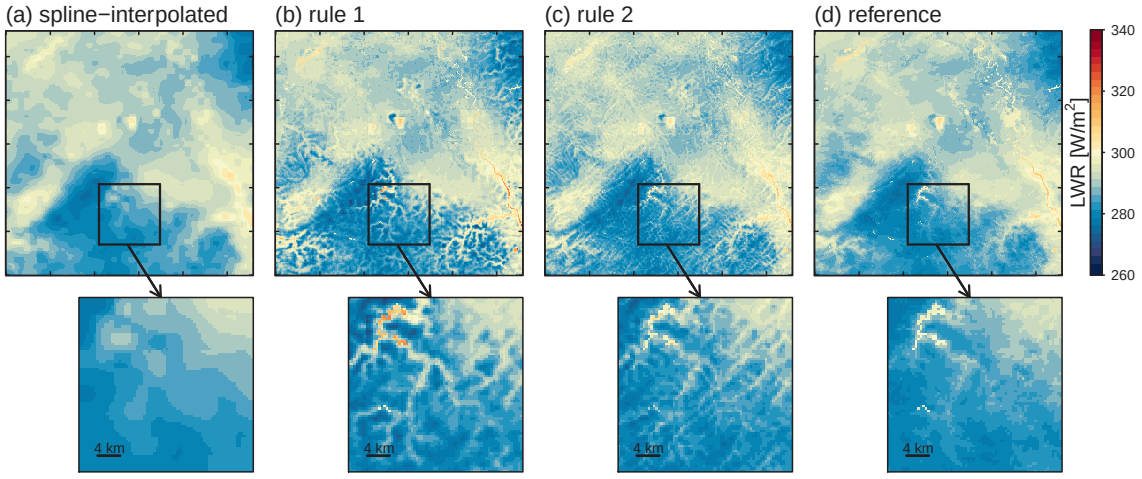

Figure 6.37.: Incoming longwave radiation at the surface on June 13th 2015 21:00 UTC. (a) shows the spline-interpolated field; (b) and (c) show the downscaled fields resulting from rules 1 and 2 (cf. Fig. 6.35); (d) shows the reference field from the high-resolution model run.

visible looking at incoming longwave radiation field. Both downscaling rules recover a portion of the subgrid-scale variability. ME(STD) and IQD are reduced by about 50% (Table 6.10). The RMSE is increased by 12% for rule 1 and by 2% for rule 2. The fine-scale pattern of the downscaled field, however, does not resemble the patterns induced by the variable cloud cover.

Figure 6.37 shows the incoming longwave radiation on June 13th 2015 21:00 UTC, a day with convective clouds and precipitation. The incoming longwave radiation ranges between 260 and 340 W/m^2. The high-resolution reference field shows little fine-scale variability. Rule 1 strongly overestimates the spatial variability increasing the RMSE by about 70%, the ME(STD) by 14% and the IQD by 17%. This is most obvious in the Eifel region. In the flatter areas of the model domain the downscaled field from rule 2 well resembles the reference. Rule 2 reduces the ME(STD) and MIQD by 23 and 30%, respectively, and increases

the RMSE only slightly by about 3% (Table 6.10).
The performance of rule 2 is overall favorable compared to rule 1. However, looking closely at the different training and validation data fields, we find the spatial variance for situations with very little fine-scale variability in the reference field to be systematically overestimated. For some fields (3 out of 30 in the validation data set) ME(STD) and MIQD are increased by the downscaling. The applicability of the downscaling rule for incoming longwave radiation needs to be carefully tested to see if positive or negative effects dominate.

Shortwave Radiation Already in the overview in Fig. 6.34 only a small reduction of the objectives has been found for both diffuse and direct components of the incoming shortwave radiation at the surface. It is, however, not surprising that the incoming shortwave radiation can hardly be related to surface properties. We will anyhow take a brief look at two example downscaling rules returned by MOGP for direct and diffuse incoming shortwave radiation.
The rule for direct shortwave radiation SWR_{dir} (Fig. 6.38) increases the RMSE by 30% for the validation data set (Table 6.11). ME(STD) and MIQD are reduced by 17% and 37%, respectively. For all three objectives there is a strong difference between the performance for training and validation data set, with the performance being clearly worse for the validation data.
The rule for downscaling diffuse shortwave radiation SWR_{dif} (Fig. 6.38) shows a slightly better performance. The RMSE is increased by 14% (Table 6.11). ME(STD) and MIQD are reduced by 26% and 52%, respectively. For the downscaling rule for the diffuse shortwave

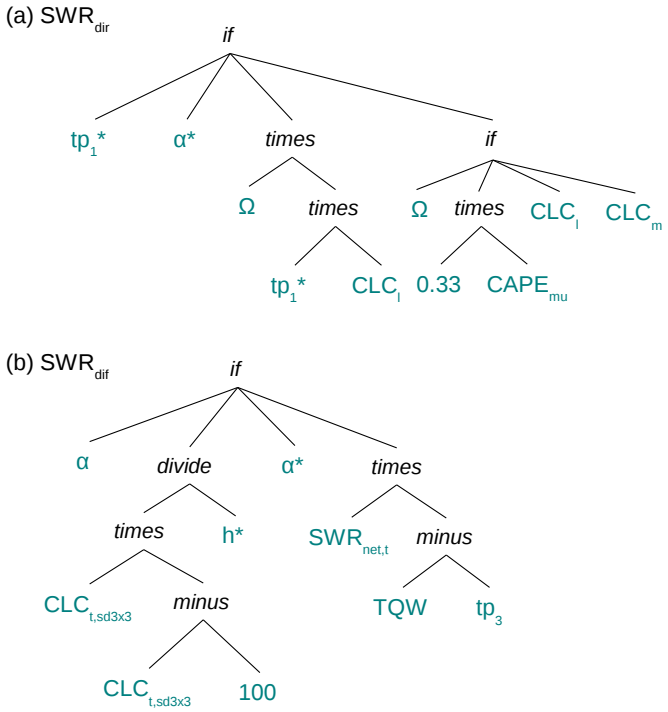

Figure 6.38.: Downscaling rules returned by MOGP for downscaling incoming direct and diffuse shortwave radiation at the land surface. Both rules are from MOGP$_{full}$.

Table 6.11.: Relative reduction of the objectives by the MOGP downscaling rules for incoming direct and diffuse shortwave radiation (Fig. 6.38) applied to training and validation data sets. The values for the training data set are given in brackets. The bottom row provide the relative improvements for the fields in Figs. 6.39 (May 22$^{\text{nd}}$ 2015 9:00 UTC) and 6.40 (May 7$^{\text{th}}$ 2014 08:00 UTC), respectively.

(a) SWR_{dir}

	$\widetilde{\text{RMSE}}$	$\widetilde{\text{ME(STD)}}$	$\widetilde{\text{MIQD}}$
	-0.30	0.17	0.37
	(-0.19)	(0.36)	(0.52)
2015/05/22 9:00	-0.19	0.44	0.61

(b) SWR_{dif}

	$\widetilde{\text{RMSE}}$	$\widetilde{\text{ME(STD)}}$	$\widetilde{\text{MIQD}}$
	-0.14	0.26	0.52
	(-0.15)	(0.29)	(0.50)
2014/05/07 8:00	-0.12	0.39	0.53

radiation no systematic difference in performance between training and validation data set is observed.

After simplification the rule for downscaling direct shortwave radiation (Fig. 6.38) reads:

IF $tp_1^* > \alpha^*$
$\quad SWR_{dir}^* = \Omega\ tp_1^*\ CLC_l$
ELSE
$\quad SWR_{dir}^* = 0.$

That is, the fine-scale pattern of the direct incoming shortwave radiation is related to the topography parameter tp_1*. Meaning for valleys negative anomalies are predicted; for mountain pixels, positive anomalies are predicted. One might argue that a longer path through the atmosphere might go along with a smaller direct incoming radiation. This effect is, however, small for neighboring pixels. The composition of the atmospheric column (clouds, water vapor, aerosols) has a much stronger impact.

The downscaling rule for diffuse shortwave radiation (Fig. 6.38) reads:

IF $\alpha > CLC_{t,sd3\times3}\ (CLC_{t,sd3\times3} - 100)/h^*$
$\quad SWR_{dif}^* = \alpha*$
ELSE
$\quad SWR_{dif}^* = -SWR_{net,t}\ tp_3.$

The fine-scale pattern of the incoming diffuse shortwave radiation is related to the albedo anomaly α^* and the topography parameter tp_3. Using the albedo anomaly appears reasonable as a large albedo means that a great portion of the incoming solar radiation is reflected by the land surface rather than absorbed. This reflected solar radiation propagates upwards, is potentially scattered at cloud droplets and thereby potentially increases the downward propagating component of the diffuse shortwave radiation.

Figure 6.39 shows the direct incoming shortwave radiation on May 22$^{\text{nd}}$ 2015 9:00 UTC, a partially cloudy day with weak winds. Figure 6.40 shows the diffuse incoming shortwave radiation on May 7$^{\text{th}}$ 2014 08:00 UTC a cloudy day with light to moderate precipitation. For both fields the respective downscaling rules generate some subgrid-scale variability. However, the

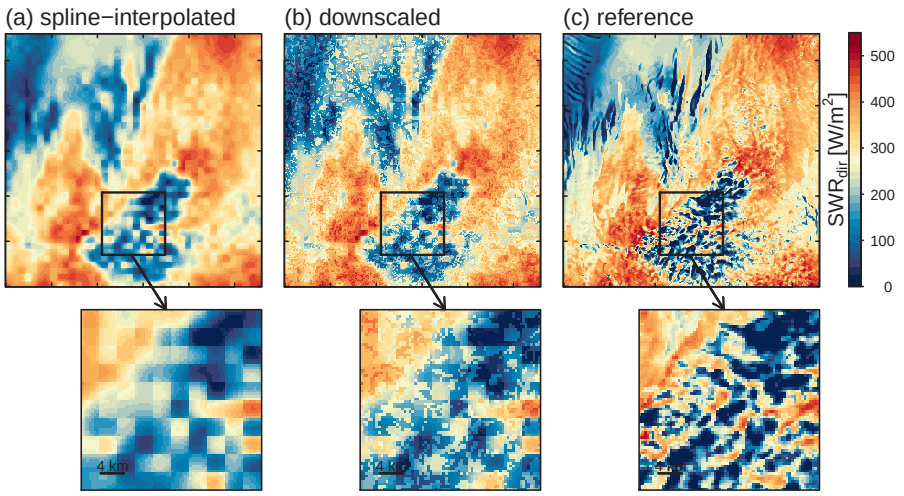

Figure 6.39.: Incoming direct shortwave radiation at the surface on May 22nd 2015 9:00 UTC. (a) shows the spline-interpolated field; (b) shows the downscaled field resulting from the downscaling rule in Fig. 6.38 (a); (c) shows the reference field from the high-resolution model run.

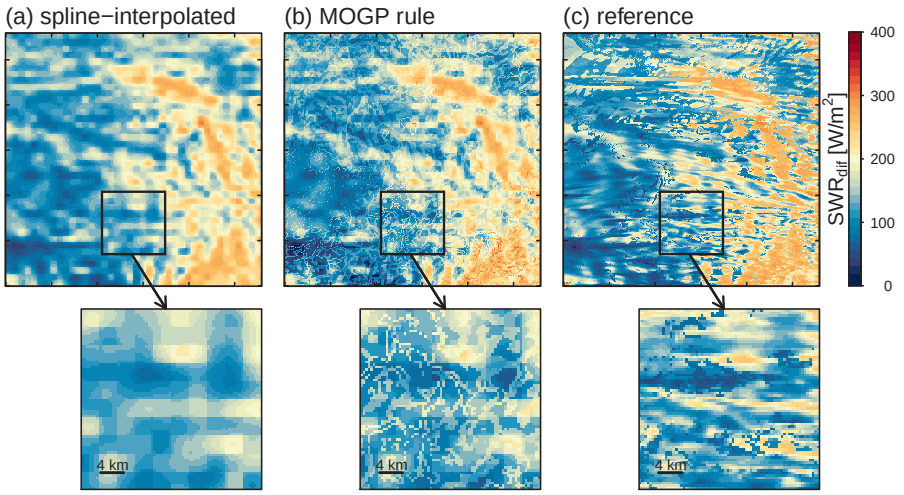

Figure 6.40.: Incoming diffuse shortwave radiation at the surface on May 7th 2014 08:00 UTC. (a) shows the spline-interpolated field; (b) shows the downscaled field resulting from the downscaling rule in Fig. 6.38 (a); (c) shows the reference field from the high-resolution model run.

actual fine-scale pattern, which is mostly generated by the cloud cover, is missed. For about one third of the validation fields not only the RMSE, but also the ME(STD) and MIQD are increased by the above downscaling rules. The usage of the downscaling rules for direct and diffuse shortwave radiation is therefore not suggested.

6.3.6. Precipitation

In its current setup MOGP is not able to evolve downscaling rules that reconstruct the subgrid-scale variability of instantaneous precipitation fields (Fig. 6.41). This is, however, expected. It would be surprising if a relation between surface heterogeneity and instantaneous precipitation amounts could be established. For longer term precipitation statistics and/or larger time

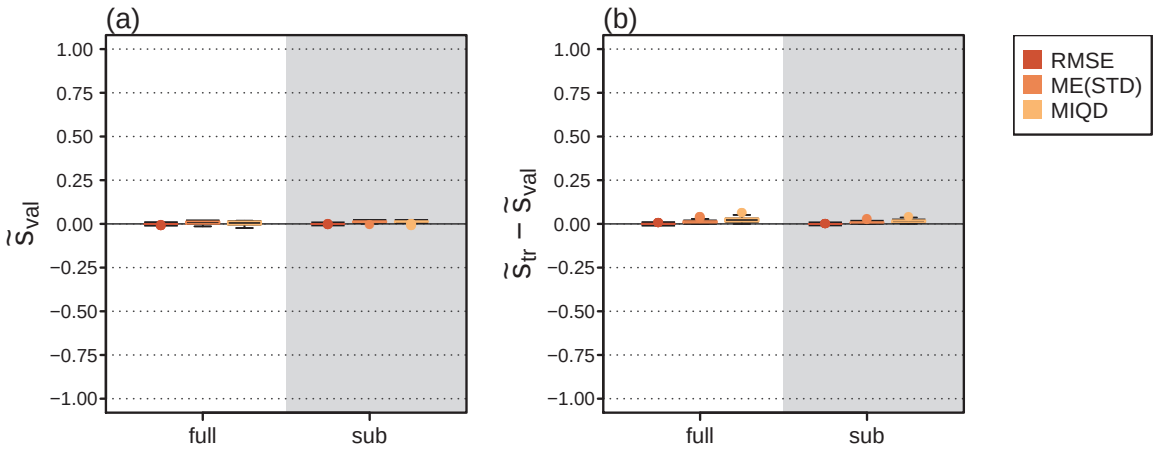

Figure 6.41.: Overview of the validation of the Pareto sets returned by MOGP for downscaling instantaneous precipitation (a) relative improvement of RMSE, ME(STD) and MIQD for the Pareto set rules applied to the validation (val) data set; (b) difference in relative improvement between training (tr) and validation data set.

scales the situation might be different. For the scales considered in this thesis we suggest to stick with the approach of *Schomburg* (2011), which restores subgrid-scale variability of precipitation adding autogregressive noise such that the PDF and the autocorrelation function of precipitation (from a set of high-resolution reference model runs) are reproduced as closely as possible.

6.3.7. Summary

In the following the downscaling rules suggested for future tests (see also Sec. 6.4) are summarized.

Pressure 10 m pressure can be downscaled as

$$p^* = h^* \; p_{gr25},$$

i.e, the pressure anomaly p^* with respect to the spline-interpolated pressure field is given by the topographic height anomaly times the vertical pressure gradient in the lowest 2 layers of the COSMO model ($\approx$ 25 m). RMSE, ME(STD) and MIQD are reduced by $\geq$ 97% by such downscaling.

Temperature For 10 m temperature the following combination of the downscaling rule from *Schomburg et al.* (2010) and MOGP is suggested:

IF $T_{gr105} < 0.0058 \; Km^{-1}$
  $T^* = -0.0084 \; h^*$
ELSE
  IF $h^* \; T_{gr60} > -z_0^*$
    $T^* = 0.66 \; h^* \; T_{gr60}$

ELSE
$T^* = 0.24\ z_0^* + 0.66\ h^*\ T_{gr60}$

That is, depending on the atmospheric stability the temperature anomaly T^* with respect to the spline-interpolated field is related to the topographic height anomaly h^*, the temperature gradient of the lowest three layers T_{gr60} and the roughness length anomaly z_0^*. For unstable atmospheres the downscaling rule from *Schomburg et al.* (2010) has shown to work well. For stable situation (mostly clear sky nights) the MOGP rule recovers about 70% of the subgrid-scale variability with no or only marginal increase in RMSE.

Specific Humidity 10 m specific humidity can be downscaled as

$q^* = T_{so}\ q_{gr25} + 10\ FR_{land}\ q_{gr60}\ T_{so}^*,$

i.e., the specific humidity anomaly gradient is related to the vertical humidity gradients of the lowest 25 and 60 m, q_{gr25} and q_{gr60}, the fraction of land FR_{land}, and the soil temperature its anomaly, T_{so} and T_{so}^*. The presented downscaling rule only recovers about 20-30% of the subgrid-scale variability, but without any increase in RMSE.

Wind speed The horizontal wind speed in 10 m height can be downscaled as

IF $w_{h,sd3\times3} > z_0^*$
$w_h^* = -0.75\ z_0\ TKE_{20} - z_0^*$
ELSE
$w_h^* = 0.25\ TKE_{20}\ z_0 - (1 + TKE_{20})\ z_0^*,$

i.e., the wind speed anomaly w_h^* with respect to the spline-interpolated field is related to roughness length z_0, roughness length anomaly z_0^* and 20 m turbulent kinetic energy TKE_{20} dependent on the coarse-scale horizontal wind speed variability $w_{h,sd3\times3}$. About 60% of the subgrid-scale variability is restored by this rule. For low wind speeds ($\leq$ 5 m/s) an increase in RMSE of 10-20% has to be accepted. For larger wind speeds ($>$ 5 m/s) the RMSE is kept constant or even reduced by up to 10%.

Radiation The best rule found for downscaling incoming longwave radiation at the surface is:

$LWR^* = (CLC_{t,sd3\times3} - h^*)/\alpha + ST - 100/TI.$

That is, the longwave radiation anomaly LWR^* with respect to the spline-interpolated field is related to the coarse-scale cloud cover variability $CLC_{t,sd3\times3}$, topographic height anomaly h^*, albedo α, soil type ST and topographic index TI. About 30% of the subgrid-scale vari-

ability is restored by the rule with an increase in RMSE of 3% on average. However, the rule tends to strongly overestimate subgrid-scale variability in some case, namely when the fine-scale variability on the high-resolution reference field is very small. The application of the downscaling rule for longwave radiation should therefore be treated with caution.
For incoming direct and diffuse shortwave radiation no applicable rules have been detected.

Precipitation Also for precipitation no satisfactory downscaling rules have been detected.

6.4. Discussion and Outlook

Comparison to the Schomburg Downscaling Approach

Compared to the downscaling rules from *Schomburg et al.* (2010) the MOGP based downscaling rules cover more variables (e.g., wind speed) and for some variables a larger variety of atmospheric conditions (e.g., temperature in stable atmospheres). This is, however, most likely only partly due to using Genetic Programming, but also due to the multi-objective approach. MOGP does not solely aim at a pixelwise reconstruction of the high-resolution reference fields, which is for most variables (except for 10 m pressure and temperature in a well mixed boundary layer) rarely possible, but also values the reconstruction of subgrid-scale variability and PDF. In *Schomburg et al.* (2010) subgrid-scale variability not recovered by the downscaling rules in step 2 is optionally added as noise in step 3, but always at the cost of a larger RMSE.

Benefits and Limitations of MOGP

To detect downscaling rules in the described setup existing relations between surface properties (topography, roughness length, etc.) and atmospheric variables at the high resolution are required. From physics it is expected that such relations are much stronger for certain atmospheric variables than for others. The 10 m pressure for instance is approximately related to the topographic height by the hydrostatic equation. The 10 m temperature anomalies are also related to the high-resolution topography, but the relation depends on atmospheric stability. In an unstable or well mixed boundary layer adiabatic temperature profiles are expected, while under stable conditions cold air accumulates in the valleys. Also near-surface wind speed is affected by topography (luv or lee, mountain or valley) and the roughness length of the land surface. Temperature and pressure anomalies associated with topography as well as wind speed anomalies associated with roughness length or topography are well reconstructed by the MOGP downscaling rules with the largest effect of the downscaling rules being observed in the mountainous parts of the model domain.
Subgrid-scale variability generated by gravity waves, turbulence or subgrid-scale variability of the cloud cover is not or only barely recovered by the MOGP downscaling rules. Although gravity waves are initiated by surface heterogeneity, they propagate in space and time. Thus, no direct link can be established between land surface properties and subgrid-scale variability induced by gravity waves. Also the effect of turbulence is not predictable on the subgrid

scale. Radiation and potentially precipitation downscaling could benefit from subgrid-scale information about the cloud cover. Such information might become available with subgrid-scale cloud schemes. A terrain correction for incoming radiation at the surface may be applied to account for shadowing effects and topographic slope and aspect. At the grid scale such an algorithm is already available in COSMO. An adaptation of this scheme to the subgrid scale is straightforward given high-resolution information on topography. For precipitation we do not expect any noticeable relation with surface properties at the considered space and time scales. Here, a purely stochastic approach might be the only option to restore subgrid-scale variability.

Improving the Performance of MOGP

A known difficulty in Genetic Programming is the evolution of numerical constants. In an early stage of this study we have included an additional step either varying or optimizing all numerical constants after the creation of each new generation. The downscaling rules evolved with or without the variation/optimization step have not shown any significant differences in performance. Incorporating the optimization step GP converges faster when counting the number of generations evolved, but not when considering run time. In the future more advanced approaches such as a *numeric mutation* operator (*Evett and Fernandez*, 1998) could be employed[3] to improve the representation of numerical constants in our GP algorithm.
Also the representation of noise might be reconsidered. The random number generator tested so far has not visibly improved the results. Advanced algorithms to generate spatially and/or temporally correlated noise might be more appropriate to restore fine-scale variability caused by turbulence, subgrid-scale cloud cover variability etc.
The occurrence of bloat, i.e., the tendency of the downscaling rules to contain unnecessary terms, is a known problem in GP. In our GP algorithm bloat is reduced by the size constraints applied to the parse trees and by incorporating the solution size in the objectives, but many of the generated downscaling rules still contain unnecessary terms, which have been removed manually. More advanced techniques to control bloat exist and should be tested (e.g., *Luke and Panait*, 2006). Further, the simplification of the downscaling rules (scale analysis etc.) may be automated.
The efficiency of the MOGP algorithm can be improved by adjusting run dimensions (population size, Pareto set size, total number of generations), function and terminal sets, objective functions, genetic operator probabilities, and the maximum tree size. Also the rule selection from the Pareto set could be reconsidered. For now, the downscaling rules that perform best when considering the sum of the relative reduction of RMSE, ME(STD) and MIQD have been selected, but also other selection criteria assigning more weight to selected objectives are possible.
The performance of MOGP for downscaling is like any other method limited by the information contained in the available predictors. No statistical method will exactly reconstruct

[3] *Numeric mutation* is a genetic operator (like mutation or crossover), which replaces all numeric constants of an individual with new numeric constants chosen at random from a uniform distribution within a certain range specified as the current value of the constant plus or minus a so-called temperature factor.

the high-resolution reference fields generated by a complex 3D atmospheric model at high resolution.

Effect of Downscaling on Surface Fluxes

The main goal of our downscaling effort is to offer more realistic atmospheric input to land- and subsurface models employed at higher resolution than the driving or coupled atmospheric model. The derived downscaling rules, however, do not necessarily improve the resulting surface exchange fluxes, because our objectives do not account for covariances between atmospheric and surface state variables. Especially for temperature this can be problematic as the turbulent exchange coefficients are nonlinear functions of the near-surface atmospheric stability, which is determined by the difference between the atmospheric temperature in the lowest model layer and the temperature at the land surface.
A first investigation of the effect of the MOGP based temperature downscaling on the turbulent exchange of latent and sensible heat fluxes during a clear sky day and night is presented in *Zerenner et al.* (2016). For daytime situations the downscaling moved the fluxes closer to the reference. For nighttime situations, where the RMSE is slightly increased by our downscaling, the resulting fluxes differ more strongly from the reference. Small shifts between the fine-scale fields of 10 m temperature and ground surface temperature are probably the main reason. In reality like in coupled simulations the ground surface temperature reacts to the atmospheric near-surface temperature. A large temperature difference leads to stronger fluxes which then reduce the difference. Indeed, we see that in the described test the fluxes are by tendency overestimated in their absolute value when the downscaling is applied. When driving a land-surface model and the ground surface temperature reacts to the downscaled atmospheric forcing, we expect the overestimation tendencies to be reduced or to disappear. In tests employing the COSMO model with a 2.8 km grid spacing and the land-surface scheme TERRA with a 400 m grid spacing *Schomburg* (2011) have found only little effect of their downscaling on latent and sensible heat fluxes compared to the mosaic approach. One reason for this is the error correlation between near-surface atmosphere and surface parameters observed for instance for temperature. *Schomburg* (2011) have found that in the standard mosaic with coarse-scale atmospheric forcing an overestimation in atmospheric temperature (w.r.t. to observations) is accompanied by a similar overestimation in surface temperature. Thus, when estimating the fluxes the errors tend to largely cancel out. Therefore, the effect of the downscaling on the turbulent exchange fluxes is comparably small although a downscaling provides more realistic input to the land-surface scheme. Similar effects will most likely occur when applying and testing the MOGP based downscaling rules.

Scale and Model Dependency

The effect of atmospheric downscaling is scale dependent. *Giorgi et al.* (2003) for instance have carried out simulation at much larger scales (grid spacing of 50-100 km, longer simulation periods of several months) and report positive effects of atmospheric downscaling, such as a better representation of the hydrological cycle through a better representation of snow cover

in mountainous areas.
The MOGP rules are trained for downscaling from 2.8 km to 400 m grid spacing. The rules may be valid for larger scale ranges, but they are likely not directly transferable to scales of 50 km and larger, except for pressure and temperature (at least for unstable atmospheres). Downscaling at larger scales will require rerunning MOGP with an appropriate training and validation data set.
Despite the potential scale dependency of the downscaling rules also a model dependency is likely. Already changing the turbulence parameterization in COSMO significantly affects for instance the simulated 10 m temperature field (cf. Fig. 6.4). Thus, using another model setup or a completely different atmospheric model likely requires different downscaling rules at least if the user aims to stay consistent with the model used.

Suggestions for Future Work

The most important next step is the testing of the MOGP based downscaling scheme. In a first test land and/or subsurface models will be driven by the downscaled atmospheric fields. A comparison to land and subsurface model runs driven by coarse-scale atmospheric data as well as atmospheric data from high-resolution COSMO model runs allows for a direct assessment of the effect of the atmospheric downscaling on land and subsurface simulations and the estimated turbulent fluxes.
The validation of the atmospheric downscaling within fully coupled subsurface-land surface-atmosphere simulations is more difficult. Running an atmospheric model with different grid spacing leads to a relatively fast divergence between the different model runs which makes it unfeasible to directly compare a simulation incorporating atmospheric downscaling with a simulation where the atmospheric model is applied at high-resolution (*Schomburg et al.*, 2012).
The MOGP performance may be improved by optimizing the MOGP setup for reproducing the turbulent exchange fluxes. To this goal the temperature difference between surface and lowest atmospheric layer can be introduced as an additional objective. Also other inter-variable relationships can be considered by the objectives, such as the correlation between topographic height and temperature anomaly.
Finally, multi-objective GP generates a set of Pareto optimal solutions and not just one downscaling rule. The Pareto set may be used to generate a downscaling ensemble, which may be realized either by using different rules for each ensemble member or by randomly switching between downscaling rules. Such an ensemble approach may help to estimate the sources of uncertainty induced by the downscaling procedure.
Also conceptionally different approaches to assess the scale gap between atmospheric and land surface/subsurface models are conceivable. The horizontal homogeneity of atmospheric parameters typically increases with height. This suggests running the atmospheric boundary layer, i.e., the part of the atmosphere that is most significantly influenced by the land surface, at a higher resolution than the higher atmospheric layers. Such an approach is, however, not straightforward and would require major adjustments in the model code and significantly increase the run time. Still, variable grid approaches are already used in geoscientific models

(e.g., *Burstedde et al.*, 2010, 2013) and with increasing computational power such an approach may become a feasible compromise between running the full atmospheric model at coarse or at high resolution. For pressure, temperature, specific humidity and wind speed such an approach would be a way to obtain high-resolution fields near the surface that are consistent with the overall model used.

Summary

- The MOGP derived downscaling rules cover more variables (e.g., wind speed) and for some variables a larger variety of atmospheric conditions (e.g., temperature in stable atmospheres) than the rules from *Schomburg et al.* (2010).
- The next step will be the application and evaluation of the MOGP based downscaling rules for driving land/subsurface simulations and within the fully coupled (soil-vegetation-atmosphere) TerrSysMP.
- The representation of subgrid-scale variability induced by gravity waves, cloud cover etc. may be improved by incorporating spatially and temporally correlated noise into MOGP.
- Transforming the Pareto sets into a downscaling ensemble might allow for an estimation of the uncertainty induced by the downscaling procedure.
- With increasing computational power a variable grid approach using a smaller grid spacing for the atmospheric boundary layer might become computationally feasible.

7 Downscaling Climate Reanalysis Data to Stations using MOGP

The downscaling of GCM output to local stations (or a higher resolution grid) is a frequent task for empirical/statistical downscaling methods and investigated in a large number of studies (cf. Chapter 4). Temperature and precipitation are the variables most commonly considered as they constitute important quantities for climate change impact studies. While for temperature linear methods often yield satisfactory results, precipitation downscaling is challenging and still subject of ongoing research.

The *European Cooperation in Science and Technology* (COST) action on *Validating and Integrating Downscaling Methods for Climate Change Research* (VALUE) defines and coordinates an intercomparison of statistical downscaling methods. The experiments are open not only to members of COST-VALUE, but also to external contributors. The data and experiment description is provided by COST-VALUE (www.value-cost.eu; *Maraun et al.* (2015)). Contributors can download the data from the web page, apply their downscaling methods and upload the results onto a validation portal, which is used for the centralized validation and intercomparison.

We have applied the multi-objective Genetic Programming downscaling method to the first two experiments set up by COST-VALUE. After introducing experiment design and data in Section 7.1, Section 7.2 describes the MOGP setup for the downscaling experiments in detail. Finally, results are presented in Section 7.3 and discussed in Section 7.4.

7.1. Experiment Design and Data

Experiment 1(a) - Perfect Prognosis, Station Data

Experiment 1(a) has been designed to test how well a downscaling method is able to represent point data (*Maraun et al.*, 2015) and considers the downscaling of global reanalysis data to a set of weather stations distributed over Europe. The predictors are taken from the ERA-Interim reanalysis (*Dee et al.*, 2011) of the European Centre for Medium-Range Weather Forecasts (ECMWF). The data is available with a grid spacing of approximately 80 km in the horizontal and 60 levels in the vertical ranging from the surface up to 0.1 hPa. The

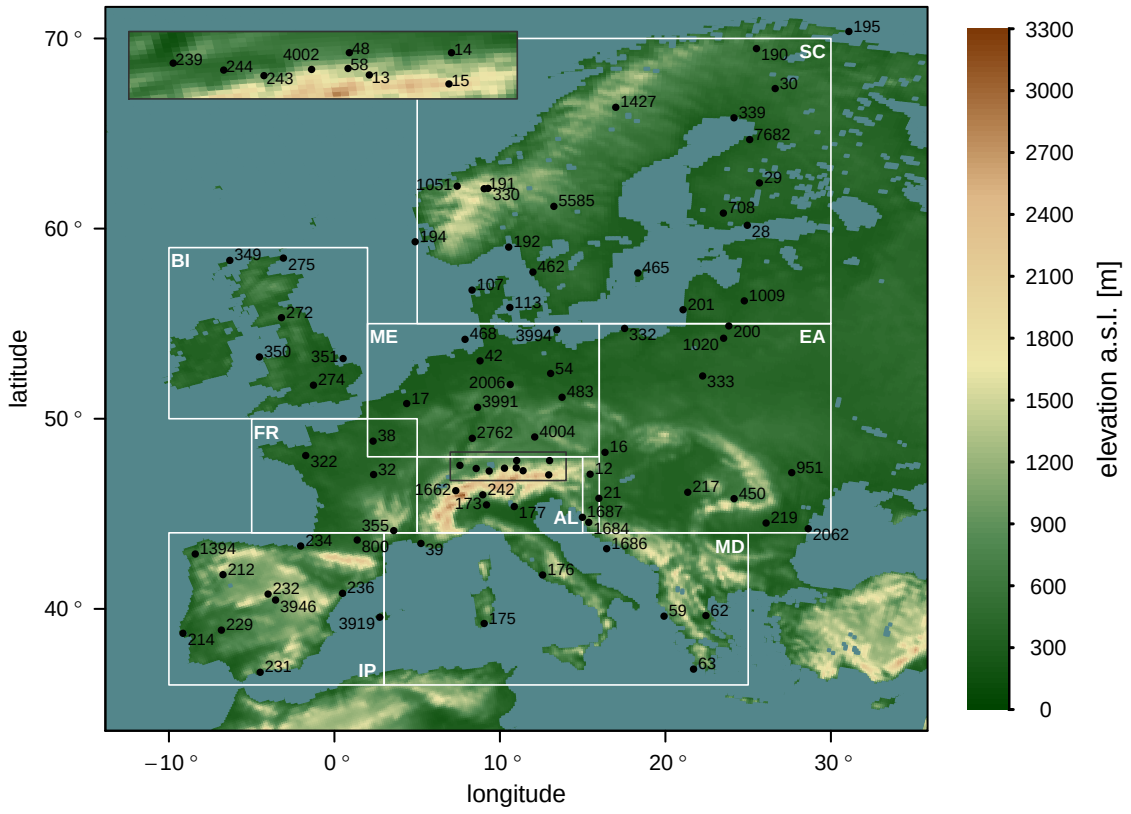

Figure 7.1.: Location and IDs of the 86 weather stations that have been selected for downscaling experiment 1(a) of COST-VALUE. A detailed experiment description can be found in the text. The box in the upper left corner shows a zoom into the northern Alpine region where the station density is comparably high. The white boxes indicate the PRUDENCE regions after *Christensen and Christensen* (2007): British Isles (BI), Iberian Peninsula (IP), France (FR), Mid-Europe (ME), Scandinavia (SC), Alps (AL), Mediterranean (MD) and Eastern Europe (EA). The regions are not relevant for the experiment design and setup, but are used to categorize the stations for the analysis of the results.

ERA-Interim reanalysis starts in the year 1979 and is until now continuously updated.
Atmospheric reanalyses, such as ERA-Interim, aim to provide the best possible, spatially and temporally consistent, estimate of the state of atmosphere in the past. Global reanalyses typically rely on general circulation models (GCMs) which - in the reanalysis setup - assimilate large observational data sets. Accordingly, reanalysis data is subject to uncertainties and errors, resulting from model errors, errors in the data assimilation schemes or a lack of usable observational data for specific regions and/or time spans.
In downscaling experiment 1(a) the predictors to offer a perfect prognosis (short perfect prog. or PP). This means the reanalysis data is assumed to offer a perfect estimate of the atmospheric state on the coarse scale. It is thus not distinguished between errors of reanalysis data set and the representativeness problem, i.e., the discrepancy between observations and reanalysis originating from the scale gap between a single grid point and a reanalysis grid box.

The predictand data is taken from ECA&D, the European Climate Assessment and Data set (*Klein Tank et al.*, 2002). 86 weather stations have been selected to form a data set representative for the different climates and local characteristics present in Europe. Figure 7.1

Table 7.1.: List of the 86 weather stations selected for downscaling experiment 1(a) of COST-VALUE: Country where the station is located, station ID according to ECA&D, lat/lon-coordinates, altitude (alt.) and Koeppen-Geiger climate classification (*Köppen*, 1884, 1918). The abbreviation for the climate classes are: Arid - Steppe - cold (BSk); Temperate - without dry season - hot summer (Cfa); Temperate - without dry season - warm summer (Cfb); Temperate - dry summer - hot summer (Csa); Cold/Continental - without dry season - warm summer (Dfb); Cold/Continental - without dry season - cold summer (Dfc); Polar - Tundra (ET).

Country	ID	Name	Lon [°]	Lat[°]	Alt.	Climate
Austria	12	Graz	15.45	47.08	366	Cfb
	13	Innsbruck	11.40	47.27	577	Cfb
	14	Salzburg	13.00	47.80	437	Cfb
	15	Sonnblick	12.95	47.05	3106	ET
	16	Wien	16.35	48.23	198	Cfb
Belgium	17	Uccle	4.36	50.80	100	Cfb
Croatia	21	Zagreb-Gric	15.98	45.82	156	Cfa
	217	Arad	21.35	46.13	116	Cfb
	1684	Gospic	15.37	44.55	564	Cfb
	1686	Hvar	16.45	43.17	20	Csa
	1687	Zavizan	14.98	44,82	1594	Dfc
Denmark	107	Vestervig	8.32	56.77	18	Cfb
	113	Tranebjerg	10.60	55.85	11	Cfb
Finland	28	Helsinki-Kaisaniemi	24.95	60.18	4	Dfb
	29	Jyvaskyla-Lentoasema	25.68	62.40	139	Dfc
	30	Sodankyla-Lapin-Ilmatiet	26.63	67.36	179	Dfc
	708	Jokioinen-Jokioisten	23.50	60.81	104	Dfb
	7682	Sikajoviki-Revonlahti	25.09	64.68	48	Dfc
France	32	Bourges	2.37	47.07	161	Cfb
	38	Paris-14E	2.34	48.82	75	Cfb
	39	Marseille-Marignane	5.23	43.44	5	Csa
	322	Rennes	-1.73	48.07	36	Cfb
	355	Mont-Aigoual	3.58	44.12	1567	Cfb
	800	Toulouse-Blagnac	1.38	43.62	151	Cfa
Germany	42	Bremen	8.80	53.04	4	Cfb
	48	Hohenpreissenberg	11.01	47.80	977	Cfb
	54	Potsdam	13.06	52.38	81	Cfb
	58	Zugspitze	10.99	47.42	2964	ET
	468	Helgoland	7.89	54.18	4	Cfb
	483	Dresden-Klotzsche	13.76	51.13	227	Cfb
	2006	Brocken	10.62	51.80	1142	Dfc
	2762	Rheinstetten	8.33	48.73	116	Cfb
	3991	Giessen-Wettenberg	8.65	50.60	203	Cfb
	3994	Arkona	13.44	54.68	42	Cfb
	4002	Oberstdorf	10.28	47.40	806	Cfb
	4004	Regensburg	12.10	49.04	365	Cfb
Greece	59	Corfu	19.92	39.62	11	Csa
	62	Larissa	22.45	39.65	72	BSk
	63	Methoni	21.70	36.83	51	Csa
Italy	173	Milan	9.19	45.47	150	Cfa
	175	Cagliari	9.05	39.23	21	Csa
	176	Roma-Ciampino	12.58	41.78	105	Csa
	177	Verona-Villafranca	10.87	45.38	68	Cfa
Lithuania	200	Kaunas	23.83	54.88	77	Dfb
	201	Klaipeda	21.07	55.73	6	Cfb
	1009	Birzai	24.77	56,20	60	Dfb
	1020	Lazdijai	23.52	54.23	133	Dfb

Country	ID	Name	Lon [°]	Lat[°]	Alt.	Climate
Norway	190	Karasjok	25.50	69.47	129	Dfc
	191	Kjoermsgrende	9.05	62.10	626	Dfc
	192	Faerder	10.53	59.03	6	Cfb
	194	Utsira-Fyr	4.88	59.31	55	Cfb
	195	Vardoe	31.08	70.37	14	ET
	330	Fokstua	9.28	62.12	952	Dfc
	1051	Tafjord	7.42	62.23	15	Csb
Poland	332	Lebs	17.53	54.75	2	Dfb
	333	Siedlce	22.25	52.25	152	Dfb
Portugal	212	Braganca	-6,73	41.80	690	Csb
	214	Lisboa-Geofisica	-9.15	38.72	77	Csa
Romania	219	Bucaresti-Banasa	26.08	44,52	90	Cfa
	450	Sibiu	24.15	45.80	444	Cfb
	951	Iasi	27.63	47,17	102	Cfa
	2062	Contanta	28.63	44.22	13	Cfa
Spain	229	Badajoz/Talavera-La-Real	-6.83	38.88	185	Csa
	231	Malaga	-4.49	36.67	7	Csa
	232	Navacerrada	-4,01	40.78	1894	Csb
	234	San-Sebastian-Igueldo	-2,04	43.31	251	Cfb
	236	Tortosa-Obervation-del-Ebro	0.49	40.82	44	Csa
	1394	Santiago de Compostela	-8.41	42.89	370	Cfb
	3919	Palma-de-Mallorca	2.74	39.56	8	BSk
	3946	Madrid-Barajas	-3.56	40.47	609	BSk
Sweden	339	Haparanda	24.14	65.83	5	Dfc
	462	Goteborg	11.99	57.72	5	Cfb
	465	Visby	18.33	57.67	42	Cfb
	1427	Jackvik	17.00	66.38	430	Dfc
	5585	Salen	13.26	61.17	360	Dfc
Switzerland	239	Basel-Binnigen	7.58	47.55	316	Cfb
	242	Lugano	8.97	46.00	300	Cfa
	243	Saentis	9.35	47.25	2502	ET
	244	Zuerich	8.57	47.38	556	Cfb
	1662	Sion-2	7.35	46.22	482	Cfb
United Kingdom	272	Eskdlamuir	-3.20	55.32	242	Cfb
	274	Oxford	-1.27	51.77	63	Cfb
	349	Stornoway	-6.32	58.33	9	Cfb
	350	Valley	-4.53	53.25	11	Cfb
	351	Waddington	0.52	53.17	68	Cfb
	275	Wick	-3.08	58.45	36	Cfb

shows the location and IDs of the stations. The stations are distributed over Europe with a comparably high station density in the Alpine region. This region is of special interest due to the distinct topography. Here, large differences between reanalysis and station data can be expected due to the impact of topography which is not resolved in the reanalysis. Also the differences between the stations, though close, are comparably large due to their different heights and location within the terrain. Daily maximum, minimum and mean temperature as well as daily accumulated precipitation amounts are considered as predictand variables in experiment 1(a). Detected outliers for maximum and minimum temperatures as well as inconsistent data with $T_{min} > T_{max}$ have been removed from the data set.

The experiment is set up as a cross-validation. The entire time span of the experiment ranges from January 1st 1979 to December 31st 2008. To perform a 5-fold cross-validation the time

span has been split into 5 subperiods of 6 years each (i.e., 1979 - 1984; 1985 - 1990; 1991 - 1996; 1997 - 2002; 2003 - 2008). For the cross-validation the calibration/validation process of the downscaling method has to be repeated five times. Each of the 5 subperiods is (in succession) considered as the test/validation data set and the remaining 4 subperiods are used for calibration/training of the downscaling model.

Experiment 1(b) - Perfect Prognosis, Gridded Data

Experiment 1(b) is identical to experiment 1(a) except for the predictand data set which comes from E-OBS (*Haylock et al.*, 2008). E-OBS is a gridded data set with a grid spacing of 0.22° based on the ECA&D observational data. For experiment 1(b) the data from the gridboxes closest to the 86 locations used in experiment 1(a) is chosen. When considering observational data that have been regridded to the predictor grid size, the representativeness problem does not occur. Hence, the downscaling performance w.r.t. model errors, such as those inherent in parameterization schemes or resulting from the omission of subgrid-scale topography etc, can be directly assessed.

7.2. MOGP Setup

We apply MOGP to detect downscaling rules for each of the 86 stations separately. As already mentioned, the experiments are set up as a cross-validation with 5 successive validation periods. All 4 predictand variables suggested by VALUE are downscaled, i.e., daily maximum, minimum and mean temperature and daily accumulated precipitation. Thus, in total $86 \times 5 \times 4$ MOGP runs have been carried out.

The downscaling rules are trained to predict anomalies with respect to ERA-Interim. For precipitation the rules predict the anomaly w.r.t. the closest ERA-Interim grid point. For temperature the anomaly w.r.t. to the linearly interpolated maximum, minimum and mean temperature in 2 m height is predicted. The objectives are calculated from the absolute values, i.e., by comparing the sum of the ERA-Interim predictions and downscaled anomalies to the station observations.

The set up of the MOGP is very similar to the setup in Section 6.2. The main difference lies in the definition of the objectives. While in Section 6.2 the objectives have been foremost aimed at reproducing spatial variability and do not explicitly consider temporal variability, the focus is now placed on the temporal variability of the time series at each particular station. The spatial variability between stations is not explicitly considered.

For temperature downscaling MOGP can be readily applied. For precipitation is has to be ensured that no negative values are returned by the downscaling. This is achieved by setting all negative values predicted by a downscaling rule to zero before applying the objective functions to the downscaled series.

7.2.1. Objectives

All objectives are formulated as penalties, i.e., the smaller the better. The root mean square error is the only objective which compares downscaled and reference time series pointwise. All other objectives evaluate statistics calculated from the whole training time series.

Root Mean Square Error The root mean square error between predicted (downscaled D) y^D and observed (reference R) time series y^R

$$RMSE = \sqrt{\frac{1}{n}\sum_t (y_t^R - y_t^D)^2}, \tag{7.1}$$

where n is the number of observations at a station. As discussed before, solely minimizing the RMSE leads to a prediction of the expected value given the predictors. Restoring variability requires further objectives.

Absolute Error of Standard Deviation The absolute error of the standard deviation is the second objective and defined as

$$AE(STD) = |\sigma(y^R) - \sigma(y^D)| = \sqrt{\frac{1}{n-1}}\left|\sqrt{\sum_t (y_t^R - \bar{y}_t^R)^2} - \sqrt{\sum_t (y_t^D - \bar{y}_t^D)^2}\right|. \tag{7.2}$$

That is, first the standard deviations σ of downscaled and reference time series are calculated and the absolute value of the difference in standard deviation serves as objective.

Mean Error of Quantiles The mean error of a set of selected quantiles serves as third objective. This objective aims to restore the probability densities of temperature and precipitation at the respective station and is defined as

$$ME(Q) = \frac{1}{n_q}\sum_{Q_i \in \mathcal{Q}} |Q_i^R - Q_i^D| \tag{7.3}$$

with $\mathcal{Q}$ being the set of selected quantiles for temperature T and precipitation P,

$$\mathcal{Q}_T = \{Q_{0.001}, Q_{0.01}, Q_{0.05}, Q_{0.25}, Q_{0.5}, Q_{0.75}, Q_{0.95}, Q_{0.99}, Q_{0.999}\}$$

$$\mathcal{Q}_P = \{Q_{0.5}, Q_{0.6}, Q_{0.7}, Q_{0.8}, Q_{0.9}, Q_{0.95}, Q_{0.99}, Q_{0.999}\},$$

and n_q the number of elements in $\mathcal{Q}$. For precipitation, the low quantiles are expected to approach zero and are thus neglected as the frequency of precipitation occurrence is accounted for by the absolute error of the precipitation frequency AE(F) (see below). For precipitation, we choose to consider foremost high quantiles in order to obtain a good representation of extremes. For temperature, moderate as well as very high and very low quantiles are considered.

Absolute Bias One central goal of empirical-statistical downscaling of GCM output is to reduce biases, i.e., over- or underestimation of a quantity by a GCM. Therefore the bias serves as fourth objective. To formulate the bias as a penalty, we take the absolute value of the difference in mean $\bar{y}$ between reference and downscaled time series as objective, i.e.,

$$AB = |\bar{y}^R - \bar{y}^D| = \left| \frac{1}{n} \sum_t y_t^R - \frac{1}{n} \sum_t y_t^D \right| = \frac{1}{n} \left| \sum_t (y_t^R - y_t^D) \right|. \tag{7.4}$$

A bias in a downscaled time series also shows in the RMSE. However, a large RMSE does not necessarily result from a large bias. Thus, it appeared reasonable to introduce the bias as an additional objective.

Absolute Error of Occurrence Frequency (precipitation only) Daily precipitation series contain large number of zero values. Thus, for precipitation downscaling an additional objective is introduced in order to quantify the ability of a downscaling rule to capture the frequency of precipitation occurrence. We determine the absolute error of the precipitation occurrence frequency as

$$AE(F) = \frac{1}{n} |n(y^R > 0) - n(y^D > 0)|, \tag{7.5}$$

where $n(y > 0)$ is the number of values larger than 0. This definition compares the total number of precipitation days in reference and downscaled time series. The timing of the precipitation events is not considered by AE(F).

Solution Size The size of the solutions, i.e., the number of nodes a downscaling rule consists of, is our last objective. This aims at obtaining a reasonable trade-off between the quality and complexity of the downscaling rules.

7.2.2. Parameters

The settings for GP parameters are specified in Table 7.2. The function set contains the four arithmetic functions and an if statement (*if* A>B *do* C *else do* D). The terminal set contains random constants drawn from the interval $[0, 1]$, some decimal powers and the predictor variables.

We use the standard predictor variables suggested by COST-VALUE as listed in Table 7.3. The list contains surface variables as well as a set of variables at selected pressure levels. All predictors except for precipitation are bilinearly interpolated from the ERA-Interim grid to the location of the respective station. Precipitation is taken from the nearest neighboring ERA-Interim grid point. The predictors are not further pre-processed, i.e., no removal of seasonal cycle, scaling etc. is applied.

MOGP is run for 200 generations (for each cross-validation period and each station) with 100 individuals each generation and a maximum Pareto set size of 100. The genetic operators for (subtree-)mutation and crossover are applied with a probability of 50% each. To keep the downscaling rules readable the maximum tree depth is set to 6 levels. With the above setting,

Table 7.2.: Summary of the GP parameters. Protected division means that division by zero returns the dividend not an error.

Parameter	Value
function set	+ , - , ×, protected /, if
terminal set	random numbers [0,1], 10, 10^2, 10^3, predictor variables (Table 7.3)
generations	200
population size	100
max. Pareto set size	100
genetic operators	(subtree-)mutation, crossover
max. tree levels	6

Table 7.3.: List of commonly used ERA-Interim predictors for experiment 1(a) and (b) according to www.value-cost.eu.

Predictor		Levels [hPa]	Units	Aggregation
u	zonal wind component	250 500 700 850 1000	m/s	daily mean
v	meridional wind component	250 500 700 850 1000	m/s	daily mean
z	geopotential height	250 500 700 850 1000	m^2/s^{-2}	daily mean
p_{sl}	sea level pressure	-	Pa	daily mean
T	temperature	250 500 700 850 1000	K	daily mean
T_{mean}	mean temperature (2 m)	-	K	daily mean
T_{min}	minimum temperature (2 m)	-	K	daily minimum value
T_{max}	maximum temperature (2 m)	-	K	daily maximum value
q	specific humidity	250 500 700 850 1000	kg/kg	daily mean
P	precipitation	-	m	daily accumulated value

a single MOGP run takes approximately 10 min on a normal PC (using one node), i.e., the entire runs for one experiment (4 predictands at 86 stations with 5 different training periods) take approximately 12 days (when carried out sequentially). Due to the comparably small memory requirements 4 runs can easily be performed simultaneously without any noticeable deceleration.

7.3. Results

In the following the MOGP results for downscaling daily maximum, minimum and mean temperature and daily accumulated precipitation are described in detail. The results for both experiments were submitted to COST-VALUE. In the present thesis we restrict ourselves to the results of experiment 1(a).
The MOGP results are compared against the raw data from the ERA-Interim reanalysis. Standard downscaling approaches (in the following also referred to as reference methods) based on linear regression (for temperature) and generalized linear models (for precipitation) serve as further reference to evaluate the performance of MOGP.

MOGP Variants

As explained in Chapter 5 MOGP returns a set of Pareto optimal downscaling rules. To condense the results, for each station and each cross-validation period three Pareto optimal rules are selected according to different selection criteria.

MOGP$_{RMSE}$ From each Pareto set (for the 86 stations and 5 cross-validation periods) we pick the downscaling rule yielding the smallest RMSE (for the training data). These downscaling rules are in the following referred to as MOGP$_{RMSE}$.

MOGP$_Q$ From each Pareto set the downscaling rule with the smallest ME(Q) (for the training data) is selected. These downscaling rules are in the following referred to as MOGP$_Q$.

MOGP From each Pareto set one downscaling rule is selected that yields a trade-off between all objectives. So far the selection is carried out subjectively as no definite selection criterion for this rules has yet been defined (see Sec. 7.4).

As explained in Chapter 5 the idea behind the multi-objective approach is to fit a trade-off between different, potentially conflicting objectives. That is for judging the performance of MOGP only the third variant which is fitting a trade-off between the objectives should be considered. The results for MOGP$_{RMSE}$ and MOGP$_Q$ are given to estimate the advantages of the multi-objective setup compared to using GP to minimize foremost the RMSE or foremost aimed at matching the observed PDF.

Performance Measures

To evaluate the performance of MOGP and reference methods we consider some of the objectives introduced in Section 7.2.1 (some objectives have been slightly adjusted) and some additional performance measures. Precisely, the following quantities have been calculated to assess the downscaling performance.

Bias B The bias is the difference in mean $\bar{y}$ between predicted (downscaled D) y^D and observed (reference R) y^R, i.e.,

$$B = \bar{y}^D - \bar{y}^R = \frac{1}{n}\sum_t (y_t^D - y_t^R), \tag{7.6}$$

where n is the number of observations (i.e., the length of the series neglecting any missing data). A positive bias indicates an overestimation of temperature or precipitation on average. A negative bias indicates an underestimation.

Root Mean Square Error RMSE The RMSE compares predicted and observed time series pointwise and is given by

$$RMSE = \sqrt{\frac{1}{n}\sum_t (y_t^D - y_t^R)^2}. \tag{7.7}$$

Pearson Correlation Coefficient ρ The Pearson correlation coefficient serves as a measure for the linear relation between predicted and observed series. It is defined as

$$\rho = \rho(y^D, y^R) = \frac{cov(y^D, y^R)}{\sigma(y^D)\sigma(y^R)} = \frac{\sum_t (y_t^D - \bar{y}_t^D)(y_t^R - \bar{y}_t^R)}{\sqrt{\sum_t (y_t^D - \bar{y}_t^D)^2 \sum_t (y_t^R - \bar{y}_t^R)^2}}. \tag{7.8}$$

Error of the Standard Deviation E(STD) The error of the standard deviation is given by the difference between the standard deviation of the observed series $\sigma(y^R)$ and the predicted series $\sigma(y^D)$, i.e.,

$$E(STD) = \sigma(y^D) - \sigma(y^R) = \sqrt{\frac{1}{n-1}} \left(\sqrt{\sum_t (y_t^D - \bar{y}_t^D)^2} - \sqrt{\sum_t (y_t^R - \bar{y}_t^R)^2} \right). \tag{7.9}$$

A positive E(STD) indicates an overestimation of the standard deviation (i.e., a higher variance for the downscaled series compared to the observations). A negative E(STD) indicates an underestimation.

Integrated Quadratic Distance IQD The integrated quadratic distance measures the difference between the empirical cumulative distribution functions G of the downscaled and the reference series,

$$IQD = \int_{-\infty}^{\infty} (G^D(y) - G^R(y))^2 dy. \tag{7.10}$$

The empirical CDF for the reference series G^R calculates as

$$G^R(y) = \frac{n(y^R \leq y)}{n}. \tag{7.11}$$

That is, the empirical CDF at each value of y is given by the fraction of the numbers of observations less or equal to y divided by the total number of observations. The definition of the downscaled CDF G^D is analogous to the definition of G^R .

Error of Quantiles E(Q_i) For selected quantiles the respective error is determined as

$$E(Q_i) = Q_i^D - Q_i^R. \tag{7.12}$$

If E(Q_i) is positive, this means Q_i is overestimated by the respective downscaling method. A negative E(Q_i) means Q_i is underestimated.

Error of Precipitation Frequency E(F) The error of the rain day frequency calculates as

$$E(F) = \frac{1}{n} \left(n(y^D > 0) - n(y^R > 0) \right), \tag{7.13}$$

where $n(y > 0)$ is the total number of values larger than 0. A positive E(F) means the respective downscaling method gives too many days with precipitation. A negative E(F)

indicates that the respective method gives too little precipitation days.

Error of lag-i Autocorrelation E(AC$_i$) The error of autocorrelation at lag-i is calculated as

$$E(AC_i) = \rho(y_t^D, y_{t+i}^D) - \rho(y_t^R, y_{t+i}^R) \tag{7.14}$$

with ρ denoting the Pearson correlation coefficient (cf. Eq. 7.8). A positive E(AC$_i$) indicates an overestimation of the autocorrelation at the respective lag. A negative E(AC$_i$) indicates an underestimation.

RMSE and Bias of the Spatial Correlation RMSE(SC), B(SC) The RMSE of the spatial or inter-station correlation between two stations i, j calculates as

$$RMSE(SC) = \sqrt{\frac{2}{n_s^2 - n_s} \sum_{i,j<i} (\rho(y_i^D, y_j^D) - \rho(y_i^R, y_j^R))^2}. \tag{7.15}$$

with n_s denoting the number of stations. The bias of the spatial correlation is defined as

$$B(SC) = \overline{\rho(y_i^D, y_j^D)} - \overline{\rho(y_i^R, y_j^R)} \tag{7.16}$$

with the bar denoting the mean over all stations, i.e., $\overline{\rho(y_i^R, y_j^R)} = \frac{2}{n_s^2 - n_s} \sum_{i,j<i} \rho(y_i^R, y_j^R)$. $\overline{\rho(y_i^R, y_j^R)}$ is defined analogous.

For quantities which can take positive and negative values (i.e., B, E(STD), E(Q$_i$), E(F), B(SC), E(AC$_i$)) also the absolute values are used if this appears reasonable. These are denoted by AB, AE(STD), AE(Q$_i$), AE(F), AE(AC$_i$) and AB(SC), respectively.

7.3.1. Temperature

Reference Methods

Two standard methods for temperature downscaling (e.g., *Huth*, 2002) serve as reference to evaluate the performance of MOGP. First reference is a multiple linear regression or linear model (LM). The predictors used are the same as offered to MOGP (cf. Table 7.3), but standardized to zero mean and unit variance. No predictor selection is performed, i.e., all predictors are incorporated in the linear model. A more detailed description of multiple linear regression is given in Appendix B. The second reference method is a simple weather generator (WG). Instead of taking the series of expected values predicted by the linear model, we sample from the underlying Gaussian distribution to account for the variance not explained by the predictors. Or in other words, Gaussian noise with zero mean and constant variance is added to the prediction of the LM to recover the variance of the observed series.

Table 7.4.: Objective functions and related measures for ERA-Interim and downscaled T_{max}: Bias B, root mean square error RMSE, error of standard deviation E(STD), Pearson correlation coefficient ρ, integrated quadratic distance IQD, error of selected quantiles $E(Q_i)$. The bar denotes the mean over all stations. *max* and *min* refer to the maximum and minimum over all stations. Values discussed in the text are highlighted in bold font.

	ERA	**LM**	**WG**	**MOGP**$_{RMSE}$	**MOGP**$_{Q}$	**MOGP**
$\overline{\text{AB}}$ [K]	**2.03**	**0.01**	**0.02**	**0.11**	**0.09**	**0.06**
$\overline{\text{B}}$	-0.158	-0.001	0.000	0.030	0.017	0.006
B_{min}	-6.08	-0.05	-0.06	-0.36	-1.21	-0.19
B_{max}	16.02	0.04	0.04	0.57	0.34	0.23
$\overline{\text{RMSE}}$ [K]	**3.02**	**1.41**	**1.97**	**1.74**	**3.10**	**2.13**
RMSE_{max}	16.45	2.32	3.24	3.24	25.93	4.15
$\overline{\text{AE(STD)}}$ [K]	**0.65**	**0.13**	**0.01**	**0.36**	**0.49**	**0.10**
$\overline{\text{E(STD)}}$	-0.42	-0.13	0.00	**-0.28**	**0.36**	-0.06
ME(STD)_{min}	-1.97	-0.45	-0.05	-1.43	-0.28	-0.78
ME(STD)_{max}	1.54	-0.03	0.03	0.53	19.89	0.53
$\overline{\rho}$ [1]	0.97	0.98	0.97	0.97	0.92	0.96
ρ_{min}	0.88	0.92	0.85	0.89	0.24	0.82
$\overline{\text{IQD}}$ [K^2]	0.0400	0.0001	0.0001	0.0006	0.0003	0.0002
IQD_{max}	0.629	0.001	0.009	0.005	0.002	0.0001
$\overline{\text{AE}(Q_{0.01})}$ [K]	2.19	**0.50**	**0.43**	1.12	**0.29**	**0.52**
$\overline{\text{E}(Q_{0.01})}$	0.98	0.41	-0.00	0.87	0.07	0.25
$\text{E}(Q_{0.01})_{min}$	-7.84	-0.60	-0.90	-1.80	-0.84	-1.24
$\text{E}(Q_{0.01})_{max}$	16.50	2.80	2.40	7.70	1.60	3.16
$\overline{\text{AE}(Q_{0.25})}$ [K]	1.72	**0.14**	**0.15**	0.34	0.28	**0.21**
$\overline{\text{E}(Q_{0.25})}$	0.08	0.07	0.02	0.18	0.07	0.03
$\text{E}(Q_{0.25})_{min}$	-5.70	-0.60	-1.10	-1.50	-1.80	-1.40
$\text{E}(Q_{0.25})_{max}$	13.50	0.50	0.40	1.20	1.00	0.70
$\overline{\text{AE}(Q_{0.5})}$ [K]	2.08	**0.09**	**0.11**	0.19	0.19	**0.17**
$\overline{\text{E}(Q_{0.5})}$	-0.19	-0.03	-0.01	-0.03	0.06	-0.00
$\text{E}(Q_{0.5})_{min}$	-6.30	-0.40	-0.50	-1.10	-1.20	-1.00
$\text{E}(Q_{0.5})_{max}$	16.10	0.40	0.30	0.50	0.60	0.60
$\overline{\text{AE}(Q_{0.75})}$ [K]	2.46	**0.10**	**0.11**	0.31	0.26	**0.20**
$\overline{\text{E}(Q_{0.75})}$	-0.39	-0.07	-0.06	-0.14	-0.03	-0.04
$\text{E}(Q_{0.75})_{min}$	-7.50	-0.30	-0.30	-1.10	-0.80	-0.60
$\text{E}(Q_{0.75})_{max}$	18.10	0.40	0.30	1.02	1.10	0.80
$\overline{\text{AE}(Q_{0.99})}$ [K]	3.25	**0.49**	**0.47**	0.84	**0.23**	**0.35**
$\overline{\text{E}(Q_{0.99})}$	-1.25	-0.32	0.15	-0.66	-0.01	-0.08
$\text{E}(Q_{0.99})_{min}$	-10.00	-3.10	-1.90	-4.30	-1.30	-2.66
$\text{E}(Q_{0.99})_{max}$	18.70	0.70	1.10	1.39	0.94	1.14

Overview

Figure 7.2 shows an overview of the performance of the different downscaling approaches for downscaling daily maximum temperature T_{max}. Table 7.4 provides the mean, maximum and minimum values (over all 86 stations) of the measures from Fig. 7.2.
The bias is well reduced by all downscaling techniques, but linear model (LM) and weather generator (WG) yield an on average smaller (absolute) bias than MOGP (0.01 to 0.02 K for LM and WG compared to 0.06 K for MOGP). MOGP$_{RMSE}$ and MOGP$_Q$ show an even larger bias which is not surprising as the performance concerning the bias is not considered when selecting the downscaling rules of these two variants.
Also concerning the RMSE, the reference methods perform on average better than MOGP. The LM well reduces the RMSE from on average 3.02 K for ERA to 1.41 K. For WG the RMSE is with 1.97 K slightly larger as recovering variance increases the RMSE. MOGP$_{RMSE}$ ranks between the two reference methods. MOGP and MOGP$_Q$ both exhibit a larger RMSE than WG, and MOGP$_Q$ shows some strong outliers which are caused by MOGP$_Q$ predicting unphysically large temperatures for single days. With MOGP being able to fit nonlinear downscaling rules there is a risk of such effects, when the range of predictors and predictand

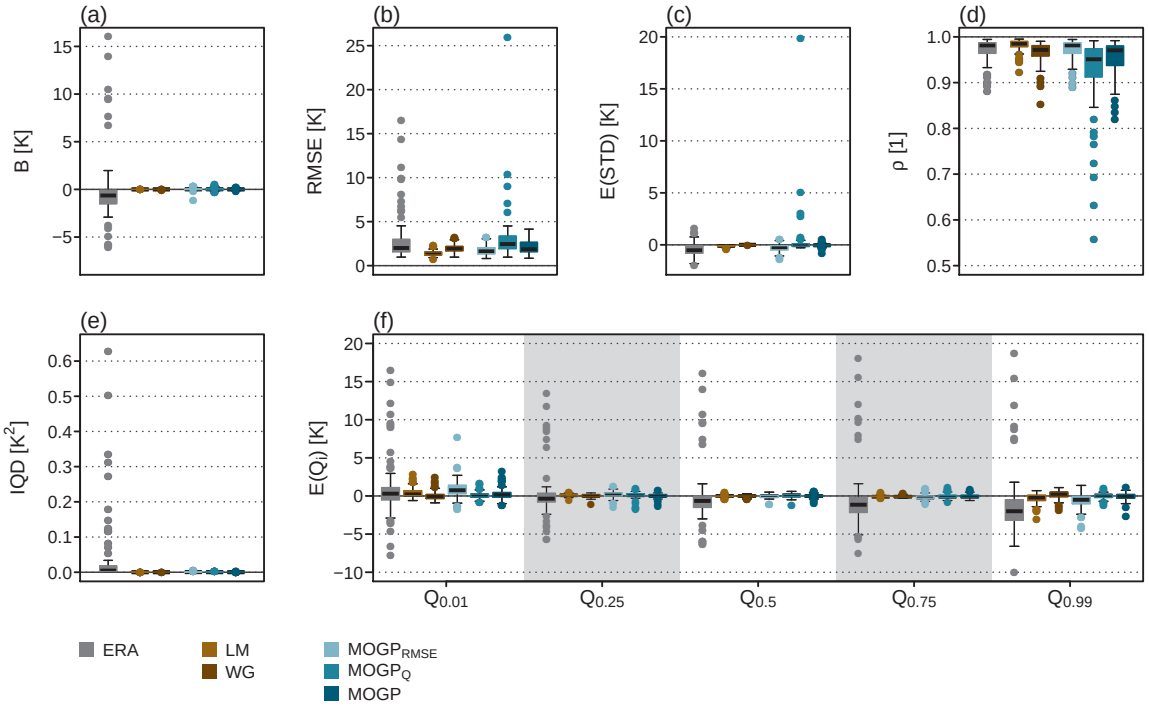

Figure 7.2.: Boxplots of objective functions and related quantities for ERA-Interim and downscaled daily maximum temperature T_{max}: Bias B, root mean square error RMSE, error of standard deviation E(STD), Pearson correlation coefficient ρ, integrated quadratic distance IQD, error of selected quantiles E(Q_i). Each boxplot corresponds to 86 values (for the 86 stations considered) and is obtained from the full downscaled time series (1979-2008). The horizontal line within the boxes is the median, the upper and lower boundaries of the boxes correspond to the 75%- and 25%-quantiles. The whiskers indicate the range spanned by maximum and minimum. The length of the whiskers is restricted to 1.5 times the interquartile range (i.e., the range between 25%- and 75%-quantile). Values outside this range are considered outliers and shown as circles.

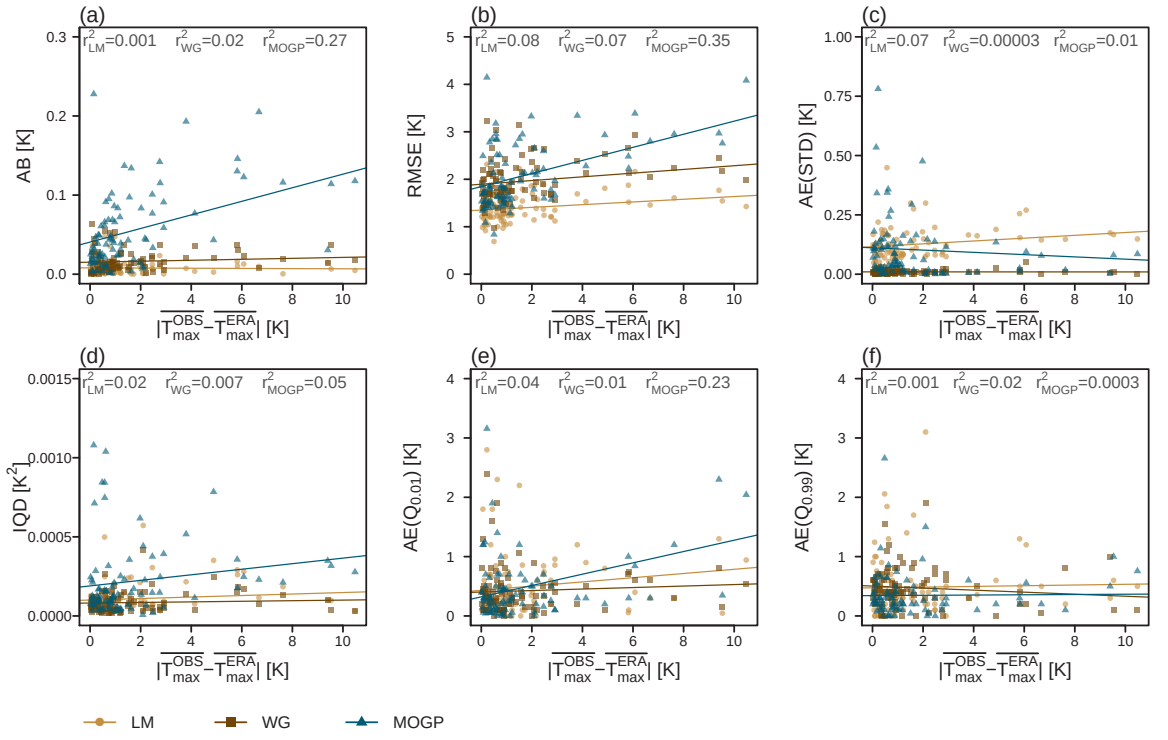

Figure 7.3.: Scatterplots between selected quantities from Fig. 7.2 and absolute ERA-Interim bias for daily maximum temperature T_{max}. Each point corresponds to one station and is obtained from the full downscaled time series (1979-2008). To better distinguish between the point clouds, regression lines (standard linear regression) have been fitted.

covered by the training data is not sufficient. For MOGP and MOGP$_{RMSE}$ we do not observe such problems. Largely neglecting the RMSE during the fitting, as done by MOGP$_Q$, appears to be critical. Unphysical outliers, as long as they are rare, are not punished severely by AE(Q).

Concerning the error of the standard deviation the performance of MOGP ranks on average between LM and WG. The WG reproduces the observed variance best with an AE(STD) of only 0.01 K. MOGP yields an AE(STD) of 0.1 K. As expected MOGP$_{RMSE}$ underestimates temporal variance by tendency. MOGP$_Q$ overestimates the variance on average, mostly due the outliers.

The IQD is well reduced by all methods, which is in large parts due to the reduction in bias discussed above. To get a better idea how well the downscaling methods recover the observed PDFs, we take a look at the representation of selected quantiles. 25%-, 50%- and 75%-quantile are well reproduced by all methods, again with LM and WG yielding slightly smaller errors than MOGP (on average ≈ 0.1 K compared to ≈ 0.2 K). For the tails of the PDF, here represented by the 1%- and 99%-quantiles, we find larger errors. Despite few outliers, MOGP$_Q$ and MOGP perform comparable to and in some cases even better than the reference methods. According to Table 7.4 for most of the considered quantiles MOGP performs better than MOGP$_Q$. This seems surprising as in MOGP$_Q$ the predictors are used to closely resemble the observed PDF for the training period. However, when neglecting the RMSE, the respective downscaling rule appears not to work as well for the validation period. Still for the 1%- and

99%-quantiles MOGP_Q performs on average best of all methods. A low RMSE and a good representation of the distribution tails are hard to achieve with a single downscaling rule.
This overview suggests the reference methods as overall favorable compared to MOGP for downscaling daily maximum temperature. Evaluating daily minimum and daily mean temperature substantially leads to the same conclusion.
Scatter plots between the performance measures and the absolute bias of ERA-Interim w.r.t. the observations at each station (cf. Fig. 7.3) reveal that absolute bias and RMSE for MOGP are by tendency larger, the larger the ERA bias. For the RMSE a linear fit gives a *coefficient of determination*[1] of 0.35. For LM and WG no relation between ERA bias and downscaling performance is visible.

Selected Stations

We take a closer look at the performance of the different downscaling approaches for the stations Salzburg and Saentis. Salzburg is located at approximately 450 m.a.s.l and surrounded

[1] The *coefficient of determination* r^2 is the portion of the predictand variation that is explained by a fitted linear model. In a univariate linear regression, r^2 equals the squared Pearson correlation coefficient of predictand and predictor variable.

Table 7.5.: Objective functions and related measures for daily maximum temperature T_{max} predicted by the different downscaling methods at Salzburg (station 14) and Saentis (station 243). The quantiles of the observed CDF are $Q_{0.01} = -4.6$ K, $Q_{0.25} = 7.0$ K, $Q_{0.5} = 15.0$ K, $Q_{0.75} = 21.9$ K and $Q_{0.99} = 31.8$ K at Salzburg and $Q_{0.01} = -14.3$ K, $Q_{0.25} = -3.0$ K, $Q_{0.5} = 1.4$ K, $Q_{0.75} = 6.5$ K and $Q_{0.99} = 15.5$ K at Saentis.

(a) Salzburg

	ERA	**LM**	**WG**	$\textbf{MOGP}_{RMSE}$	$\textbf{MOGP}_Q$	**MOGP**
B [K]	-2.02	0.01	0.02	0.01	-0.12	-0.08
RMSE [K]	2.74	1.69	2.36	1.88	3.13	2.14
ρ [1]	0.98	0.98	0.97	0.98	0.94	0.97
E(STD) [K]	-0.62	-0.15	-0.01	-0.50	-0.16	-0.06
IQD [K^2]	0.015	0.0001	0.0001	0.0007	0.0003	0.0001
$E(Q_{0.01})$ [K]	0.10	0.40	0.00	1.80	0.20	0.50
$E(Q_{0.25})$ [K]	-1.80	0.00	0.00	0.20	0.20	0.00
$E(Q_{0.5})$ [K]	-2.35	-0.10	-0.10	-0.30	-0.50	-0.30
$E(Q_{0.75})$ [K]	-2.50	-0.10	-0.10	-0.50	-0.40	-0.20
$E(Q_{0.99})$ [K]	-2.50	0.26	0.56	-0.34	0.16	0.26

(b) Saentis

	ERA	**LM**	**WG**	$\textbf{MOGP}_{RMSE}$	$\textbf{MOGP}_Q$	**MOGP**
B [K]	10.48	0.00	0.02	0.32	-0.22	-0.12
RMSE [K]	11.13	1.43	1.98	3.23	3.89	4.08
ρ [1]	0.88	0.98	0.96	0.90	0.85	0.82
E(STD) [K]	1.21	-0.14	-0.001	0.52	0.43	0.09
IQD [K^2]	0.34	0.00003	0.00003	0.002	0.0005	0.0003
$E(Q_{0.01})$ [K]	10.70	0.94	0.54	1.84	1.54	2.04
$E(Q_{0.25})$ [K]	8.50	0.00	-0.20	-1.00	-1.50	-1.10
$E(Q_{0.5})$ [K]	10.70	0.10	0.20	0.50	0.20	0.30
$E(Q_{0.75})$ [K]	12.03	-0.07	0.03	1.03	0.62	0.43
$E(Q_{0.99})$ [K]	11.90	-0.50	-0.10	1.20	-0.30	-0.76

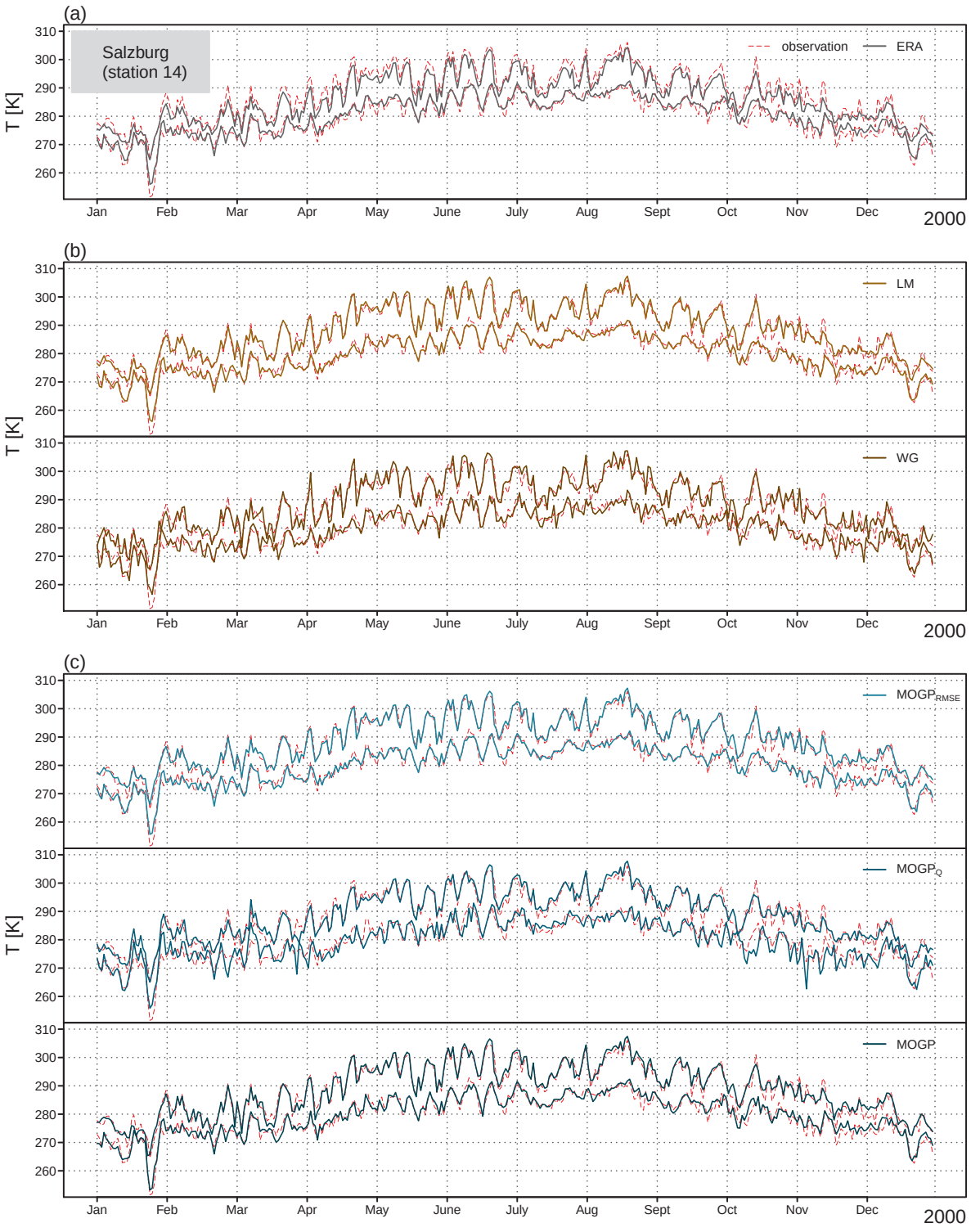

Figure 7.4.: Time series of daily maximum and minimum temperature (T_{max} and T_{min}) as observed, from ERA-Interim and from the different downscaling techniques at Salzburg (station 14) in the year 2000. The dashed red lines show the observations. The solid lines show the temperatures from ERA-Interim and from the different downscaling approaches.

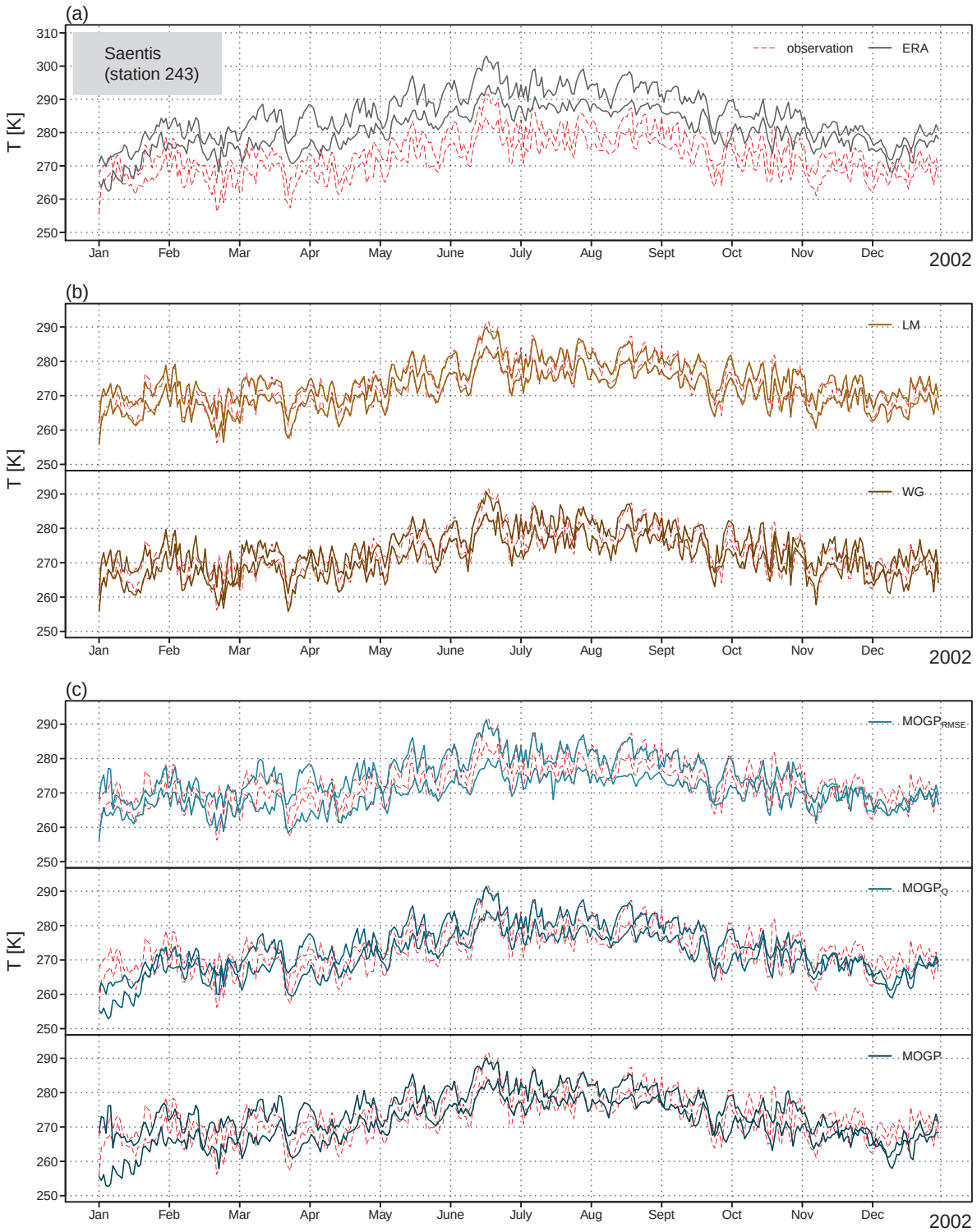

Figure 7.5.: Time series of daily maximum and minimum temperature (T_{max} and T_{min}) as observed, from ERA-Interim and from the different downscaling techniques at Saentis (station 243) in the year 2002. The dashed red lines show the observations. The solid lines show the temperatures from ERA-Interim and from the different downscaling approaches.

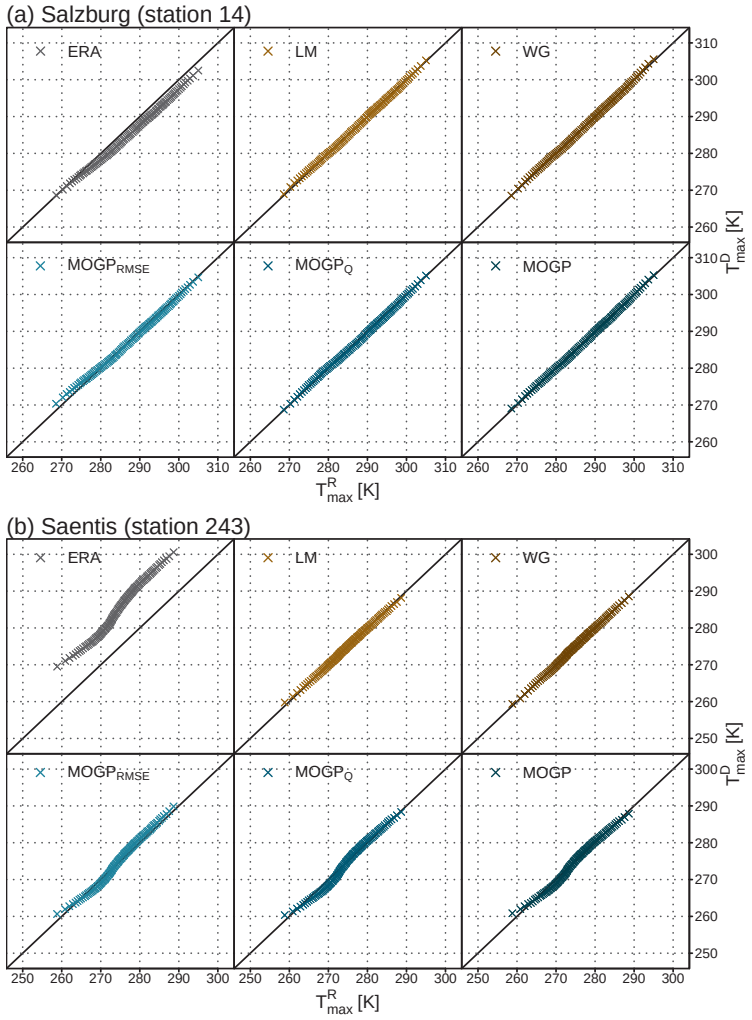

Figure 7.6.: Quantile-quantile plots for daily maximum temperature T_{max} at Salzburg (station 14) and Saentis (station 243) from ERA-Interim and estimated by the different downscaling techniques. The x-axis corresponds to the observed quantiles; the y-axis corresponds to the estimated quantiles.

by mountains leading to relatively dry conditions and moderate temperatures (cf. Figs. 7.4, 7.13). For Salzburg the bias of ERA-Interim compared to the observation equals -2.02 K for daily maximum temperature. Saenits is a mountain station in Switzerland located approximately 2500 m.a.s.l and shows large precipitation amounts throughout the year and comparably low temperatures (cf. Figs. 7.5, 7.14). Saentis has the largest ERA-Interim bias (10.48 K) of all stations considered.

Table 7.5 lists the objective functions and related measures for downscaled daily maximum temperature at Salzburg und Saentis. For Salzburg the performance of MOGP is similar to the performance of LM and WG. For Salzburg one can hardly tell any differences between the downscaling approaches (cf. Fig. 7.4). For Saentis LM and WG achieve a closer match of the observations than MOGP (cf. Fig. 7.5).

Quantile-quantile plots (Fig. 7.6) confirm these findings. For Salzburg the PDF of the daily maximum temperature of ERA-Interim deviates only slightly from the observations. All downscaling approaches restore the observed PDF well. Only MOGP$_{RMSE}$ shows some drawbacks

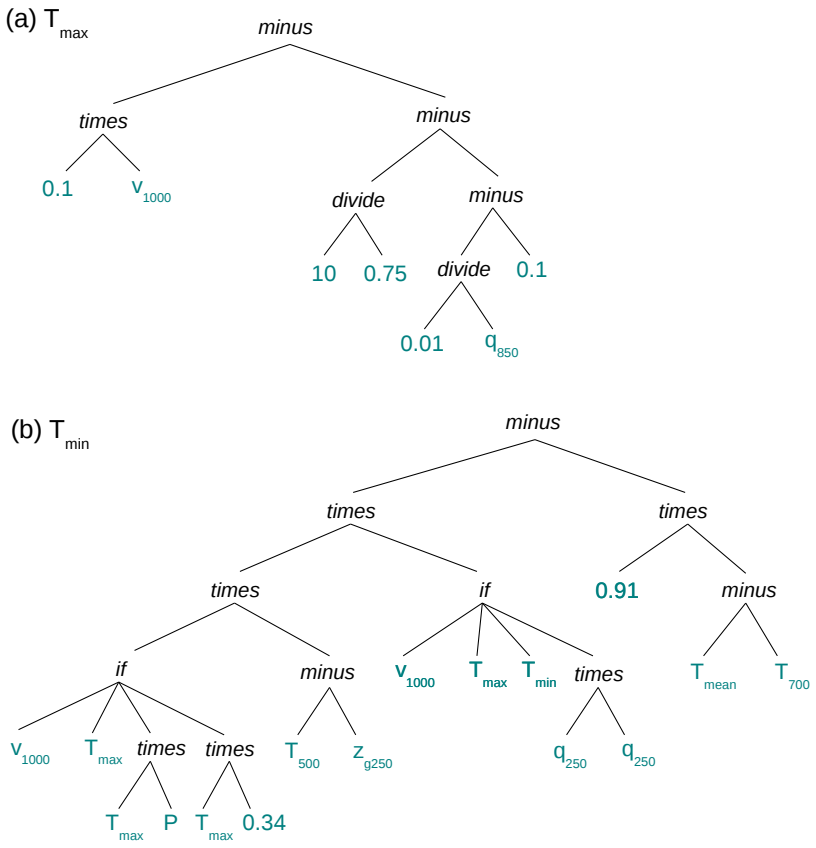

Figure 7.7.: Two downscaling rules as returned by MOGP for daily maximum and minimum temperature (T_{max} and T_{min}) at Saentis (station 243).

and overestimates the low quantiles. All other methods (LM, WG, MOGP$_Q$ and MOGP) yield an IQD of ≤ 0.0003 K^2. For Saentis the situation is different. ERA-Interim strongly overestimates the daily maximum temperature. LM and WG achieve a good fit of the observed PDF with an IQD of 0.00003 K^2. Though the MOGP methods well reduce the bias, they yield less satisfactory results concerning the PDF. It appears that MOGP is in this case not capable of capturing the shape of the PDF properly. The PDF appears to be essentially only shifted.

Figure 7.7 shows two example downscaling rules for maximum and minimum temperature at Saentis. The two rules are extracted from the Pareto set returned for the fourth cross-validation period (validated for 1997-2002) and yield the lowest RMSE of all Pareto optimal rules (MOGP$_{RMSE}$). The rule for the maximum temperature is rather simple and amounts to an almost linear regression with only two predictors. After evaluating the coefficients and some simplification it reads:

$$T^D_{max} \;=\; T^{ERA}_{max} + 0.1\; v_{1000} - \frac{0.01}{q_{850}} - 13.23.$$

A large regression constant of -13.23 accounts for the large bias between the ERA-Interim reanalysis and observations at Saentis. According to the downscaling rule dry air conditions, here represented by low specific humidity q in 850 hPa, are accompanied by a larger differ-

ence between ERA-Interim and observed maximum temperature. Near-surface wind from the south (i.e., positive v_{1000}) reduces the difference; near-surface wind from the north (i.e., negative v_{1000}) further increases the difference.
The rule for the daily minimum temperature is much more complex, involving if statements and including 10 different predictors. It reads

IF $v_{1000} > T_{max}$
$T_{min}^{D} = T_{min}^{ERA} + T_{max}\ T_{min}\ P\ (T_{500} - z_{g250}) - 0.91\ (T_{mean} - T_{700})$
ELSE
$T_{min}^{D} = T_{min}^{ERA} + 0.34\ T_{max}\ q_{250}^{2}\ (T_{500} - z_{g250}) - 0.91\ (T_{mean} - T_{700}).$

Since v_{1000} (in m/s) is never larger than T_{max} (in Kelvin), in fact only the second equation is used. The phenomenon of unnecessary terms occurring in equations or program code generated by GP is called bloat (e.g., *Poli et al.*, 2008). There are many approaches to reduce bloat. In our simple GP system we have used the tree size as an additional objective and have applied a limitation to the maximum number of tree levels to reduce bloat. There are numerous more advanced approaches, but how to fully prevent bloat is a topic of ongoing research.
The downscaling equation consists, despite of T_{min}^{ERA}, of two terms. A scale analysis based on the training/calibration data set shows that the first term $0.34\ T_{max}\ q_{250}^{2}\ (T_{500} - z_{g250})$ is one to two orders of magnitude smaller than the second term $-0.91\ (T_{mean} - T_{700})$. While the first term is difficult to understand, the second term has a simple physical meaning. $T_{mean} - T_{700}$ gives the daily average vertical temperature gradient between 2 m height and 700 hPa height. In most situations temperature decreases with height, i.e., the whole term including the factor of -0.91 is for most cases negative. Thus (according to the downscaling rule) ERA-Interim overestimates T_{min} the stronger, the larger the average decrease of temperature with height which is physically reasonable as Saentis is located in approximately 2500 m.a.s.l.

Spatial and Temporal Correlation

Spatial and temporal correlation are not directly considered by the MOGP objectives and also by none of the reference methods. For ERA-Interim the autocorrelations are, except for a few stations, smaller than for the observations (cf. Fig 7.8). This is for the most part due to ERA-Interim representing area averages which tend to have a smaller variance in space and time compared to point observations. The larger the considered lag, the larger this effect appears to be. The LM reduces the autocorrelation error (most obvious for the outlier stations), but not sufficiently which is largely due to the LM underestimating temporal

Table 7.6.: Bias B and root mean square error RMSE of the Pearson correlation coefficients of daily maximum temperature T_{max} between all 86 station.

	ERA	**LM**	**WG**	**MOGP**$_{RMSE}$	**MOGP**$_{Q}$	**MOGP**
B(SC) [1]	0.045	0.021	-0.007	0.032	-0.055	0.009
RMSE(SC) [1]	0.061	0.024	0.012	0.039	0.136	0.038

variance. Accounting for variability, as done by the WG, reduces the autocorrelation. For some stations the autocorrelation of the obervations is well reproduced, for the majority of stations, however, the WG underestimates the autocorrelation. MOGP$_{RMSE}$ shows a similar behavior as the LM with similar errors at lag-1 and slightly larger errors at lag-2 and 3. MOGP$_Q$ shows the largest errors and a clear tendency to underestimate the autocorrelation. For MOGP no general tendency to either under- or overestimate autocorrelation is found. The magnitude of the error is, however, similar to LM and WG for lag-1 and increases for the larger lags.

Spatial correlation can be demonstrated by (Pearson) correlation matrices calculated from the temperature time series at the different stations. To keep the correlation matrix at reasonable dimensions, Fig. 7.9 shows the correlation matrix for a subset of stations located in the Alpine region. Table 7.6 provides bias and RMSE of the spatial correlation for the full 86×86 matrices. The correlation matrix of the observed series shows a distinct pattern, which is largely related to the distance between the respective stations and their altitude. The ERA-Interim series are in general too strongly correlated which is mainly caused by ERA-Interim representing area means. LM and WG well restore the spatial correlation pattern and reduce the RMSE from 0.061 for ERA-Interim to 0.024 for the LM and 0.012 for the WG. As for the temporal correlation, LM overestimates and WG underestimates spatial correlations slightly with a bias of 0.021 and -0.07, respectively. For MOGP the bias is with 0.009 comparable to LM and WG. However, the reference methods obtain a smaller RMSE. For the Alpine stations MOGP clearly overestimates the spatial correlation (Fig. 7.9). For Mid-Europe, for instance, the performance of MOGP is much closer to the reference methods. The comparably bad performance for the Alpine region is largely related to the bad performance of MOGP for stations with a large ERA-Interim bias compared to the observations.

Predictor Statistics

To investigate the importance of the different predictors for downscaling daily maximum temperature we analyze how the different predictors are used by LM and MOGP (Fig. 7.10). For the LM the regression coefficients β_i (cf. Appendix B.1) serve as indicator for the importance of each predictor x_i for downscaling daily maximum temperature. For MOGP the importance of the different predictors is estimated by the frequency h_i with which the different predictors x_i occur in the final Pareto sets, i.e.,

$$h_i = \frac{n(\alpha \in \mathcal{P}' | x_i \in \alpha)}{n(\alpha \in \mathcal{P}')}, \tag{7.17}$$

or in words h_i is given by the number n of downscaling rules α in Pareto set $\mathcal{P}'$ containing predictor x_i at least once divided by the total number of rules in the Pareto set. The $\overline{\beta_i}$ and $\overline{h_i}$ shown in Fig. 7.10 are given by the h_i and β_i averaged over the 5 Pareto sets for the 5 cross-validation periods. The quantities $\overline{\beta_i}$ and $\overline{h_i}$ thus have a different meaning and should not be compared quantitatively, but only qualitatively. Though, MOGP may potentially include nonlinear predictor-predictand relations one would in general expect the predictors with a large regression coefficient (in the LM) to be frequently used by MOGP.

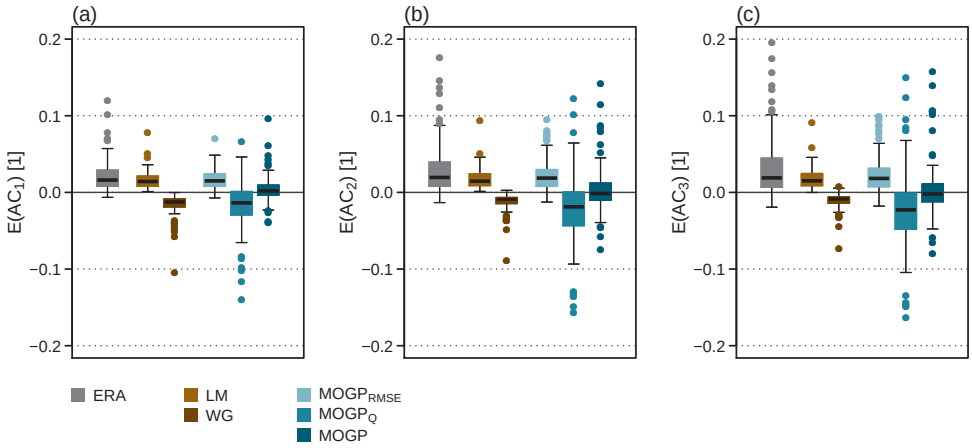

Figure 7.8.: Error of the lag-1 (a), lag-2 (b) and lag-3 (c) (Pearson) autocorrelation of the daily maximum temperature T_{max} series predicted by the different downscaling techniques. The observed autocorrelation averaged over all stations at different lags is $AC_1 = 0.94$, $AC_2 = 0.88$ and $AC_3 = 0.85$.

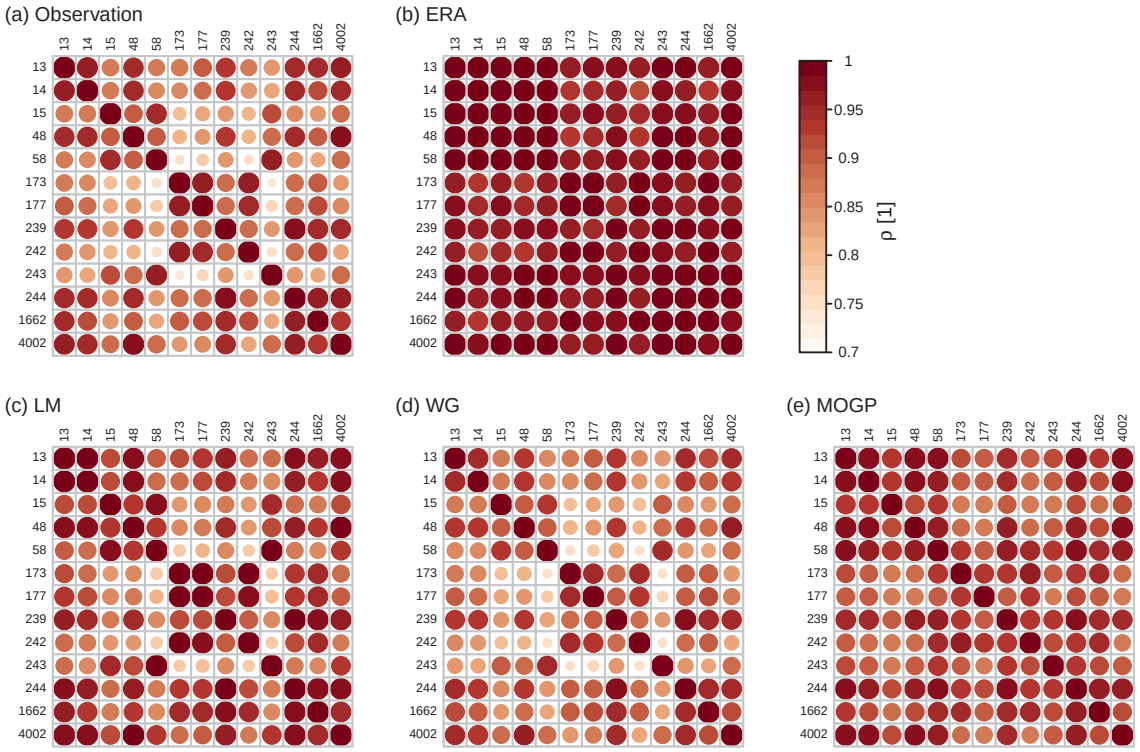

Figure 7.9.: Pearson correlation matrix of daily maxium temperature T_{max} for all stations within the Alpine (AL) PRUDENCE region - cf. Fig. 7.1.

For the LM the geopotential and pressure predictors are most important followed by temperature predictors and, with a larger gap, humidity, wind speed and finally precipitation. It is expected that the temperature from ERA-Interim contains information about the local station temperatures. The geopotential is an indicator for high/low pressure systems describing the overall weather conditions. For MOGP the ranking of the predictors is approximately inversed with wind speed, specific humidity and precipitation being used much more frequently than temperature and geopotantial height. Wind speed and humidity may of course contain some

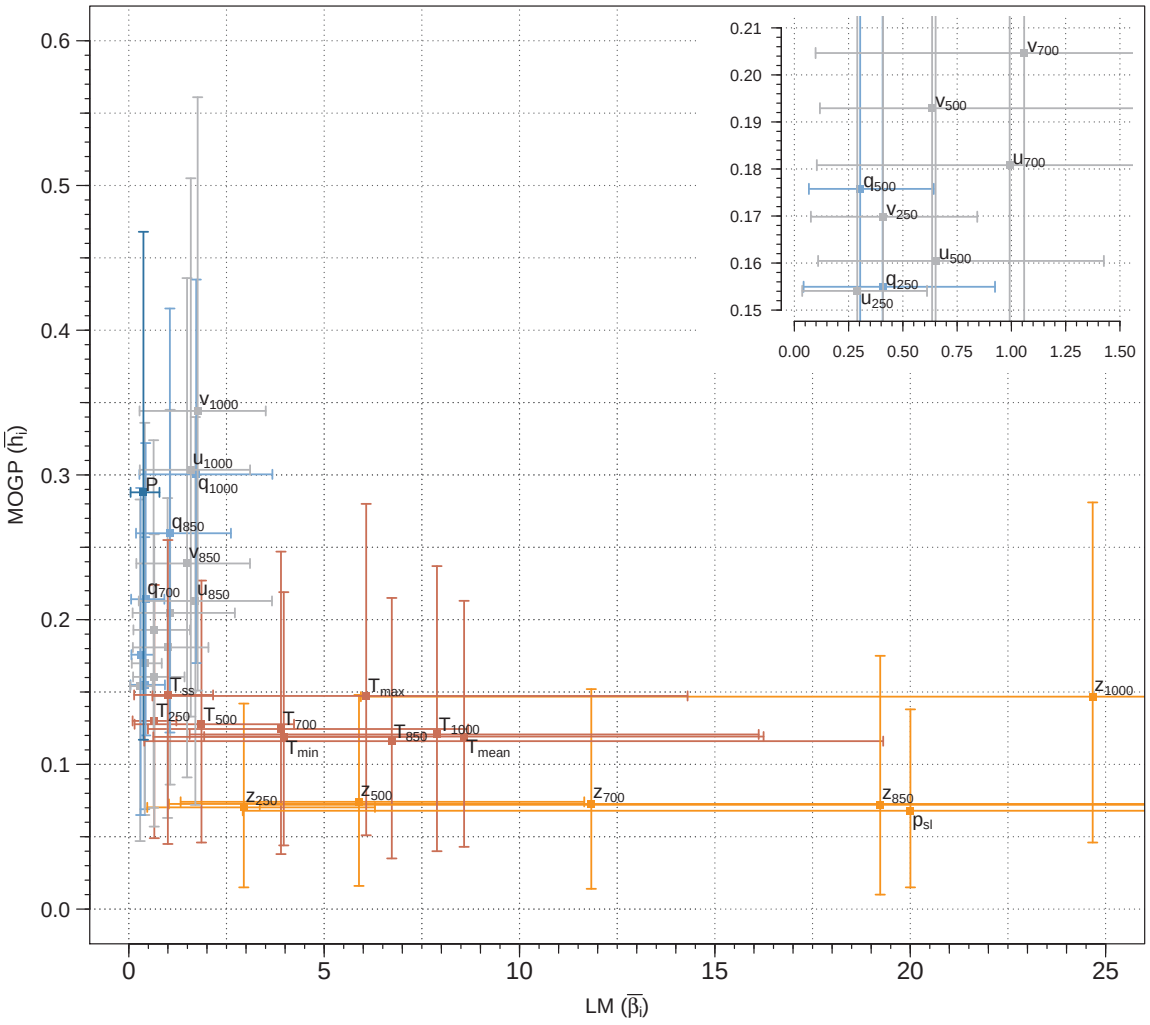

Figure 7.10.: Usage of the different predictors by multi-objective Genetic Programming (MOGP) and multiple linear regression (LM) for downscaling daily maximum temperature T_{max} at 86 stations. For the LM the regression coefficients $\overline{\beta_i}$ for each station averaged over all 5 cross-validation periods are given. For MOGP the frequency $\overline{h_i}$ with which the predictors x_i occur in the Pareto sets is given, again for each of the 86 stations averaged over all cross-validation periods. Details on the calculation are provided in the text. The points indicate the mean values over all stations. The bars indicate the 10%-and 90%-quantiles. The colors emphasize the different types of predictors considered. Temperature predictors T are shown in red; geopotential height z or pressure p based predictors are orange; wind speed predictors u, v are grey; specific humidity q is shown in light blue and precipitation P in darker blue. For z_{700}, z_{850}, z_{1000} and p_{sl} the 90%-quantile of $\overline{\beta_i}$ lies outside the plotted range, $Q_{90}(z_{700}) = 26.6$, $Q_{90}(z_{850}) = 44.2$, $Q_{90}(z_{1000}) = 55.1$, $Q_{90}(p_{sl}) = 45.6$. The smaller figure top right is a zoom into the main figure; the area is indicated by the axes.

information on the deviation between coarse-scale ERA-Interim temperatures and observed temperatures. Still it is surprising that geopotential and temperature predictors are only little used by MOGP[2]. The predictor ranking for MOGP becomes better understandable by looking at the typical scales of the different predictors:

$$\mathcal{O}(z) \approx 10^2 - 10^5 \text{ m},$$

$$\mathcal{O}(T) \approx 10^2 \text{ K},$$

$$\mathcal{O}(u) \approx \mathcal{O}(v) \approx 10^0 - 10^1 \text{ m/s},$$

$$\mathcal{O}(q) \approx 10^{-3} - 10^{-2} \text{ kg/kg},$$

$$\mathcal{O}(P) \approx 10^0 - 10^1 \text{ (rarely } 10^2\text{) mm}.$$

$\mathcal{O}$ denotes the order of magnitude. MOGP tends to use predictors with a smaller order of magnitude, i.e., wind speed, humidity and precipitation, more frequently. In comparison geopotential and also temperature predictors are rarely used, though these quantities may contain important information as they constitute important predictors for the LM. For both approaches the importance of the different predictors strongly depends on the considered station as indicated by the bars which correspond to the 10%-and 90%-quantiles of $\overline{\beta_i}$ and $\overline{h_i}$, respectively. The larger $\overline{\beta_i}$ or $\overline{h_i}$, the larger the range of the bars tends to be. For the LM this tendency is more pronounced than for MOGP.

We suppose that the predictor selection in MOGP is largely driven by the order of magnitude of the predictors due to potential problems with the evolution of numerical constants. The evolution of exact numerical constants is difficult to achieve with GP and related techniques (e.g., *Evett and Fernandez*, 1998). This issue is further discussed in Section 7.4. Despite the discussed differences the predictor usage of LM and MOGP also shows some similarities. For both methods the variables at lower levels appear to be more important than at higher levels. This holds for all types of variables, i.e., wind speed, specific humidity, temperature and geopotential height. As we are downscaling the daily maximum temperature in 2 m height this is physically reasonable for both approaches.

Looking at the considered predictors potential reasons for LM and WG outperforming MOGP (cf. Figs. 7.2-7.8) may be that MOGP is not able to sufficiently make use of the information contained in the coarse-scale temperature and geopotential height. Further, each predictor is on average contained in at maximum 25% of the MOGP based downscaling rules, while the LM exploits the full predictor set. A more fair comparison between methods would restrict both methods to the same numbers of predictors. This issue is further discussed in Section 7.4.

[2]In the current setup MOGP predicts the temperature anomalies, i.e., the ERA-Interim daily maximum temperature is actually used by the final downscaling model.

Summary

- For temperature downscaling standard linear regression based methods achieve a good match of the observations.
- MOGP_{RMSE} is throughout outperformed by LM. MOGP_{Q} is (except for the representation of extremes) outperformed by WG. For the most part also MOGP is outperformed by LM and WG, but not concerning the representation of extremes and (for the majority of stations) the representation of autocorrelation.
- The difference between the performance of MOGP and the reference methods tends to be larger for stations with a large bias between ERA-Interim and observation. For stations with a small ERA bias the performance of MOGP and reference methods is comparable.
- The frequency with which the different predictors are used by MOGP appears to be strongly dependent on their magnitude which is presumably caused by difficulties when evolving numerical constants for scaling the predictors.

7.3.2. Precipitation

Reference Methods

Precipitation downscaling is known to be more difficult than temperature downscaling, due to its non-Gaussian distribution. Three methods based on generalized linear models (GLMs; cf. Appendix B.2), a Poisson GLM, a Gamma GLM and a simplistic gamma distribution based weather generator (WG), are used as reference to evaluate the performance of MOGP for downscaling daily accumulated precipitation. All three reference methods employ a two step procedure modeling precipitation occurrence and precipitation amounts separately. For the reference method the predictors are the same as offered to MOGP (cf. Table 7.3), but standardized to zero mean and unit variance

In the first step precipitation occurrence is modeled using logistic regression. Logistic regression estimates the probability of precipitation occurrence on each day. The series of probabilities is transformed into a Boolean series (i.e., yes/no or 1/0) such that predicted precipitation probability of 70%, is transformed into a 1 with a 70% chance and into a zero with a probability of 30%. The resulting Boolean series is thus one possible realization.

In the second step precipitation amounts are estimated. The Poisson distribution based generalized linear model (PGLM), as for instance employed in *Schoof and Pryor* (2001), assumes daily precipitation amounts to follow a Poisson distribution. The predictors enter the Poisson

GLM via a logarithmic link function.
The gamma generalized linear model (GGLM) assumes daily precipitation amounts to follow a gamma distribution, which is frequently done in literature, for instance in *Coe and Stern* (1982); *Stern and Coe* (1984); *Chandler and Wheater* (2002) and *Fealy and Sweeney* (2007). The shape parameter of the gamma distribution is assumed to be temporally constant, i.e., for each station and cross-validation period the daily precipitation values are assumed to have a constant coefficient of variation (e.g., *Chandler and Wheater*, 2002). The predictors enter the Gamma GLM via a logarithmic link function.
Also for the weather generator (WG) a gamma distribution with a constant coefficient of variation is assumed, but instead of taking the expected value predicted by the gamma GLM, the WG prediction is obtained by drawing from the estimated Gamma distributions. That is, for each day the predicted precipitation amount is drawn from the respective distribution, i.e., a Gamma distribution with the expected value for the respective day and a temporally constant coefficient of variation. The autocorrelation of the precipitation series is not explicitly modeled by either of the reference approaches as it is has not been incorporated in the MOGP objectives.
For the four stations Vestervig (107), Fokstua (51), Mont-Aigoual (355) and Birzai (1009) at least one of the reference methods has occasionally generated unphysical values when applied to the validation data set. This is most likely caused by predictor values in the validation period being beyond the training data range. A careful predictor selection might improve the results of the reference methods for the mentioned stations. However, as the GLM based methods shall only provide a very first reference to assess the MOGP performance, we decided to neglect those stations to keep the reference methods simple and consistent in their setup. Further, we faced some problems with data (re)acquisition when setting up the reference methods causing us to further neglect the two stations Karasjok (190) and Vardoe (195). This leaves us with a set of 80 stations for evaluating precipitation downscaling. (Note that MOGP has been applied to all 86 stations and the respective results were submitted to COST-VALUE.)

Overview

Figure 7.11 provides an overview of the performance of the different downscaling approaches for daily accumulated precipitation P. Table 7.7 provides the mean, maximum and minimum values of the performance measures from Fig. 7.11.
The mean absolute error of the precipitation frequency is reduced from 0.30 (ERA-Interim) to 0.01 for the three reference methods. MOGP, MOGP$_{RMSE}$ and MOGP$_Q$ show larger errors of 0.19 and 0.10, respectively. However, MOGP estimates both precipitation occurrence and precipitation intensity with one downscaling rule while the reference methods implement a two step procedure.
The PGLM best reduces the mean absolute bias from 0.68 mm to 0.03 mm, followed by MOGP, with a mean absolute bias of 0.09, and then GGLM, WG and the other two MOGP variants. At most stations MOGP$_Q$ shows a clear tendency to overestimate average precipitation. GGLM and WG show a similar tendency, but less pronounced except for a few outliers.

Table 7.7.: Objective functions and related measures for ERA-Interim and downscaled precipitation P: Error of precipitation occurrence frequency E(F), bias B, root mean square error RMSE, error of standard deviation E(STD), Pearson correlation coefficient ρ, integrated quadratic distance IQD, error of selected quantiles E(Q_i). The bar denotes the mean over all stations. *max* and *min* refer to the maximum and minimum over all stations. Values discussed in the text are highlighted in bold font.

	ERA	**PGLM**	**GGLM**	**WG**	**MOGP$_{RMSE}$**	**MOGP$_Q$**	**MOGP**
$\overline{\text{AE(F)}}$ [1]	**0.30**	**0.01**	**0.01**	**0.01**	**0.19**	**0.10**	**0.04**
$\overline{\text{E(F)}}$	0.30	-0.01	-0.01	-0.01	0.16	0.03	-0.01
E(F)$_{min}$	0.04	-0.05	-0.04	-0.06	-0.15	-0.26	-0.13
E(F)$_{max}$	0.49	0.01	0.01	0.01	0.74	0.47	0.11
$\overline{\text{AB}}$ [mm]	**0.68**	**0.03**	**0.19**	**0.18**	**0.39**	**5.11**	**0.09**
$\overline{\text{B}}$	-0.03	**-0.01**	**0.18**	**0.16**	**-0.35**	**5.04**	**0.02**
B$_{min}$	-3.94	-0.61	-0.41	-0.41	-1.96	-0.51	-0.54
B$_{max}$	1.88	0.29	1.29	1.13	0.46	249.93	0.34
$\overline{\text{RMSE}}$ [mm]	**4.83**	**4.96**	**8.96**	**10.76**	**5.10**	**411.55**	**5.80**
RMSE$_{max}$	12.51	11.91	72.79	60.28	11.22	18971.81	13.40
$\overline{\text{AE(STD)}}$ [mm]	**1.85**	**1.63**	**4.51**	**4.96**	**1.86**	**405.24**	**0.30**
$\overline{\text{E(STD)}}$	-1.79	**-1.55**	**3.22**	**4.94**	**-1.86**	**405.12**	**-0.28**
ME(STD)$_{min}$	-9.22	-3.31	-2.87	-0.31	-6.13	-0.84	-1.19
ME(STD)$_{max}$	1.06	3.04	66.51	54.11	0.02	18966.83	0.29
$\overline{\rho}$ [1]	0.57	0.49	0.40	0.29	0.45	0.21	0.43
ρ_{min}	0.24	0.30	0.07	0.03	0.06	-0.01	0.16
$\overline{\text{IQD}}$ [mm^2]	**0.0118**	0.0069	0.0055	**0.0003**	0.0070	**0.0022**	**0.0004**
IQD$_{max}$	0.0473	0.0220	0.0220	0.0031	0.0440	0.0154	0.0019
$\overline{\text{AE(Q}_{0.5})}$ [mm]	0.71	0.42	0.39	**0.03**	0.42	**0.18**	**0.08**
$\overline{\text{E(Q}_{0.5})}$	0.67	0.41	0.38	0.00	0.41	0.09	0.05
E(Q$_{0.5}$)$_{min}$	-0.90	-0.10	-0.10	-0.20	-0.10	-1.50	-0.30
E(Q$_{0.5}$)$_{max}$	2.30	4.10	4.20	0.30	1.90	2.10	1.00
$\overline{\text{AE(Q}_{0.75})}$ [mm]	1.55	1.50	1.35	**0.21**	0.77	**0.76**	**0.35**
$\overline{\text{E(Q}_{0.75})}$	0.96	1.49	1.33	-0.05	0.27	0.40	0.26
E(Q$_{0.75}$)$_{min}$	-4.10	-0.30	-0.40	-1.60	-2.20	-2.00	-1.30
E(Q$_{0.75}$)$_{max}$	4.10	4.90	4.50	0.60	4.00	3.80	1.80
$\overline{\text{AE(Q}_{0.95})}$ [mm]	3.29	3.40	2.87	**0.73**	3.65	**0.86**	**0.78**
$\overline{\text{E(Q}_{0.95})}$	-2.81	-3.33	-2.81	-0.66	-3.58	0.32	-0.28
E(Q$_{0.95}$)$_{min}$	-23.40	-10.40	-9.10	-3.38	-15.90	-2.48	-2.80
E(Q$_{0.95}$)$_{max}$	4.21	1.28	1.10	1.49	1.20	3.40	3.88
$\overline{\text{AE(Q}_{0.99})}$ [mm]	8.97	10.16	6.93	**3.74**	9.01	**1.38**	**1.82**
$\overline{\text{E(Q}_{0.99})}$	-8.81	-10.11	-6.62	3.37	-9.01	-0.96	-1.62
E(Q$_{0.99}$)$_{min}$	-42.86	-20.26	-17.57	-2.44	-28.86	-7.64	-6.74
E(Q$_{0.99}$)$_{max}$	3.76	2.34	10.30	24.52	0.09	5.12	2.01
$\overline{\text{AE(Q}_{0.995})}$ [mm]	12.00	13.16	8.25	**9.20**	11.56	**2.27**	**2.52**
$\overline{\text{E(Q}_{0.995})}$	-11.78	-12.82	-6.49	8.71	-11.55	-1.73	-2.22
E(Q$_{0.995}$)$_{min}$	-47.95	-23.00	-19.72	-4.24	-31.71	-11.53	-8.74
E(Q$_{0.995}$)$_{max}$	4.54	13.36	33.66	66.24	0.31	7.77	2.60

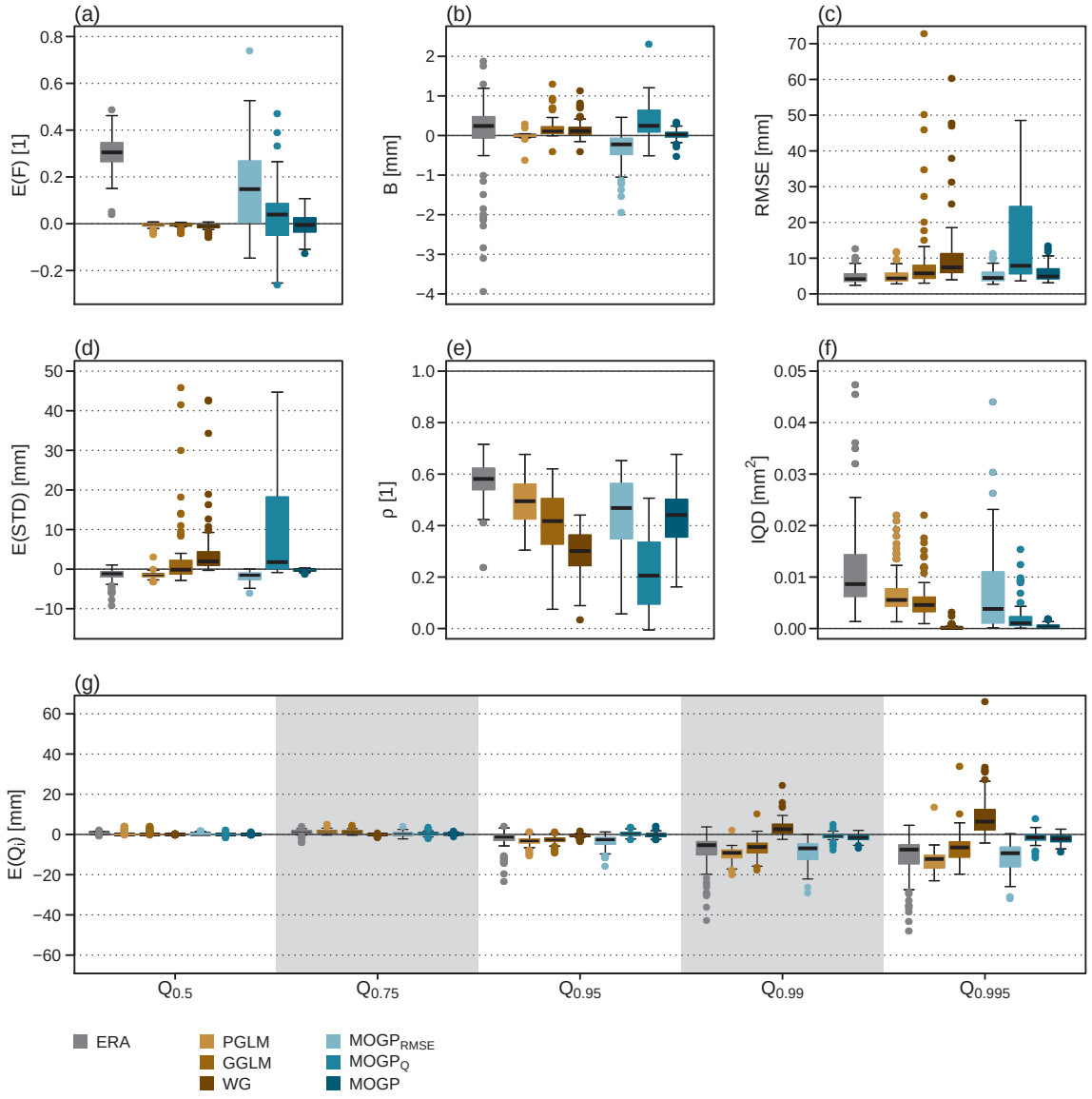

Figure 7.11.: Boxplots of objective functions and related quantities for ERA-Interim and downscaled accumulated daily precipitation P: Error of precipitation occurrence frequency E(F), bias B, root mean square error RMSE, error of standard deviation E(STD), Pearson correlation coefficient ρ, integrated quadratic distance IQD, error of selected quantiles E(Q_i). Each boxplot corresponds to 80 values (for the 80 stations considered) and is obtained from the full downscaled time series (1979-2008). For Concerning E(STD) and RMSE MOGP$_Q$ produces some strong outliers beyond the range covered by subfigures (c) and (d)[3]. For a general explanation of a boxplot see Fig. 7.2.

MOGP$_{RMSE}$ underestimates average precipitation for most stations.

None of the considered downscaling methods achieves a reduction of the RMSE for the majority of stations which illustrates the difficulty of downscaling precipitation compared to temperature. PGLM and MOGP$_{RMSE}$ keep the RMSE almost constant. MOGP slightly increases the RMSE. GGLM, WG and, most pronounced, MOGP$_Q$ increase the RMSE considerably.

The variance is for most stations only slightly underestimated by ERA-Interim. For few stations the difference in variability is, however, large exceeding 5 mm and for the most extreme

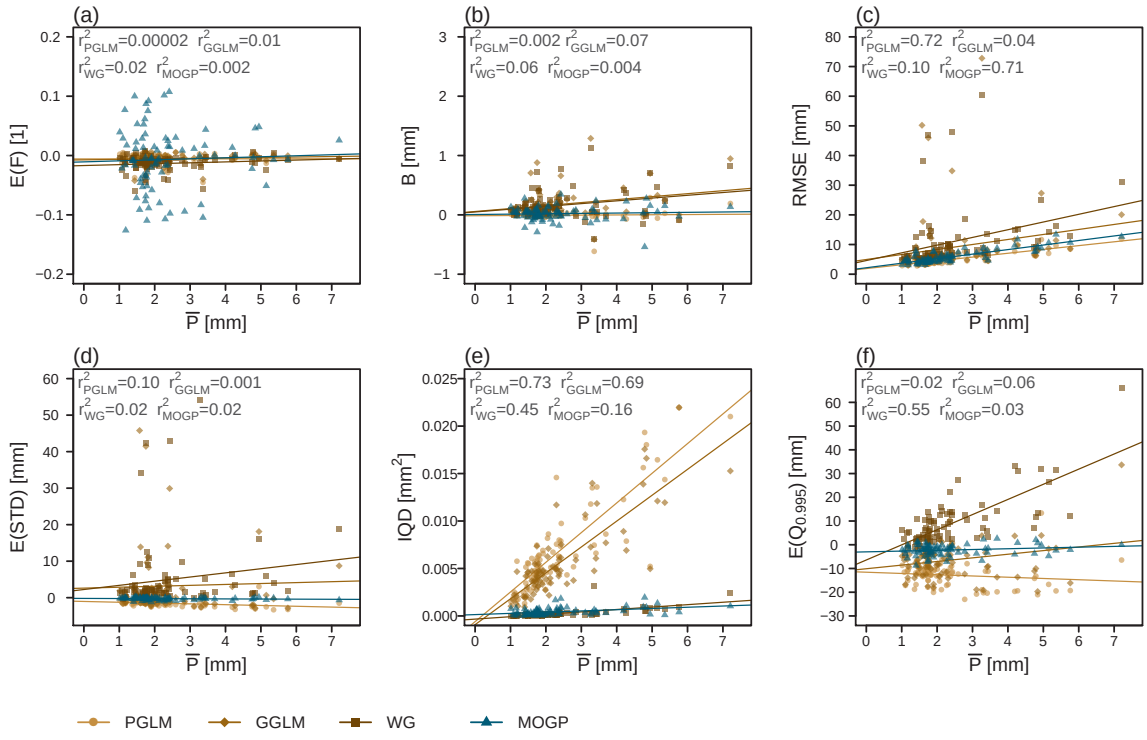

Figure 7.12.: Scatterplots between selected quantities from Fig. 7.11 and average daily precipitation $\bar{P}$ at the 80 stations. Each point corresponds to one station and is obtained from the full downscaled time series (1979-2008). To better distinguish between the point clouds, regression lines (from a standard linear regression) have been fitted and added to the figures.

stations even 9 mm. MOGP best matches the observed variance with a mean absolute error of standard deviation of 0.30 mm. All other approaches are less satisfying. While the PGLM clearly reduces the error for the outlier stations, it achieves only a small reduction of error when considering the average over all stations. Further, the PGLM slightly underestimates the temporal variance on average, as does MOGP$_{RMSE}$. GGLM, WG and especially MOGP$_Q$ strongly overestimate the temporal variance for a few stations.

The integrated quadratic distance is reduced by all techniques. WG and MOGP obtain the closest match between downscaled and observed PDF reducing the IQD from 0.0118 (for ERA-Interim) to 0.0003 and 0.0004, respectively; MOGP$_{RMSE}$ achieves only a marginal reduction. PGLM and GGLM yield a stronger reduction, but are clearly outperformed by WG, MOGP$_Q$ and MOGP.

To investigate the downscaled PDFs further, the representation of selected quantiles are compared. The 50%-quantile is well reproduced by all methods. Already for the 75%-quantile differences become apparent. WG and MOGP achieve the closest match. Also MOGP$_{RMSE}$ and MOGP$_Q$ give overall good results, but with a few outliers. PGLM and GGLM slightly overestimate the 75%-quantile on average. The higher the quantile, the larger the errors and the larger the differences between the downscaling techniques. Solely MOGP and MOGP$_Q$

[3] In Fig. 7.11 (c) and (d) the stations Paris (38), Kjoeremsgrende (191), Klaipeda (201), Basel-Bininngen (239), Dresden-Klotzsche (483), Hvar (1686), Giessen-Wettenberg (3991) are not shown for MOGP$_Q$. For these stations the RMSE ranges up to $\mathcal{O}(RMSE) = 10^5$ for MOGP$_Q$ and the E(STD) up to $\mathcal{O}(E(STD)) = 10^4$.

achieve satisfactory results up to the 99.5%-quantile and provide potentially useful information on frequency and intensity of extreme events.
Figure 7.12 reveals the relation between the downscaling performance and average station precipitation. For PGLM and MOGP a strong linear relation between RMSE and average station precipitation is found ($r^2 > 0.7$). The larger the average station precipitation, the larger the RMSE tends to be. It is not surprising that for stations with stronger precipitation by tendency larger errors occur. At first glance the linear regression fit indicates a similar tendency also for GGLM and WG, but with a coefficient of determination of $r^2 \leq 0.1$. For GGLM and WG few stations with low to moderate average precipitation show strong outliers w.r.t. the RMSE. Also for the IQD some relation between average station precipitation and the performance of the downscaling methods is obvious. For PGLM and GGLM, the larger the average station precipitation, the larger the IQD ($r^2 \geq 0.69$). For the WG similar tendencies exist, but much less pronounced. Only for MOGP, the IQD appears to be largely independent of the average station precipitation. Also concerning the 99.5%-quantiles MOGP performs well largely independent of the average station precipitation while PGLM and GGLM underestimate the high quantiles for most stations and the WG tends to overestimate extremes the stronger, the larger the average station precipitation ($r^2 = 0.55$). This is, however, not necessarily due to the average precipitation amount at a station, but rather due to the shape of the PDF.

Selected Stations

The results for the stations Salzburg and Saentis are examined in more detail. Saenits, a mountain station, shows large precipitation amounts throughout the year. Salzburg, located in lower altitudes and surrounded by mountainous areas, is comparably dry.
Table 7.8 lists objective functions and related measures for Salzburg and Saentis. For Salzburg the differences between the methods are comparably small. Concerning precipitation frequency, bias and RMSE PGLM, GGLM and WG perform slightly better than MOGP. Concerning standard deviation and extreme quantiles MOGP yields lower errors than the reference methods. For Santis MOGP performs best when considering all objectives. Only concerning the precipitation frequency the reference methods perform slightly better. Concerning RMSE and correlation the PGLM performs slightly better than MOGP. However, the PGLM underestimates temporal variance and extremes. GGLM and WG strongly overestimate temporal variance and extremes. This is in accordance to Fig. 7.12 where we already found the performance of MOGP in general favorable compared to the reference methods for stations with large precipitation amounts. While for Salzburg all downscaled precipitation series appear physically reasonable at first glance (cf. Fig. 7.13), for Saentis the overestimation of extremes by GGLM and WG is immediately obvious (cf. Fig. 7.14). For Saentis the assumption of a gamma distribution appears not to be valid especially for the distribution tails.
The quantile-quantile plots shown in Figs. 7.15 and 7.16 confirm the above findings. Figure 7.15 shows the 1%- to 99%-quantiles. For Salzburg already the ERA-Interim precipitation matches the observed distribution quite closely. PGLM and GGLM overestimate the low quantiles by up to 5 mm and strongly underestimate the higher quantiles by up to 15 mm.

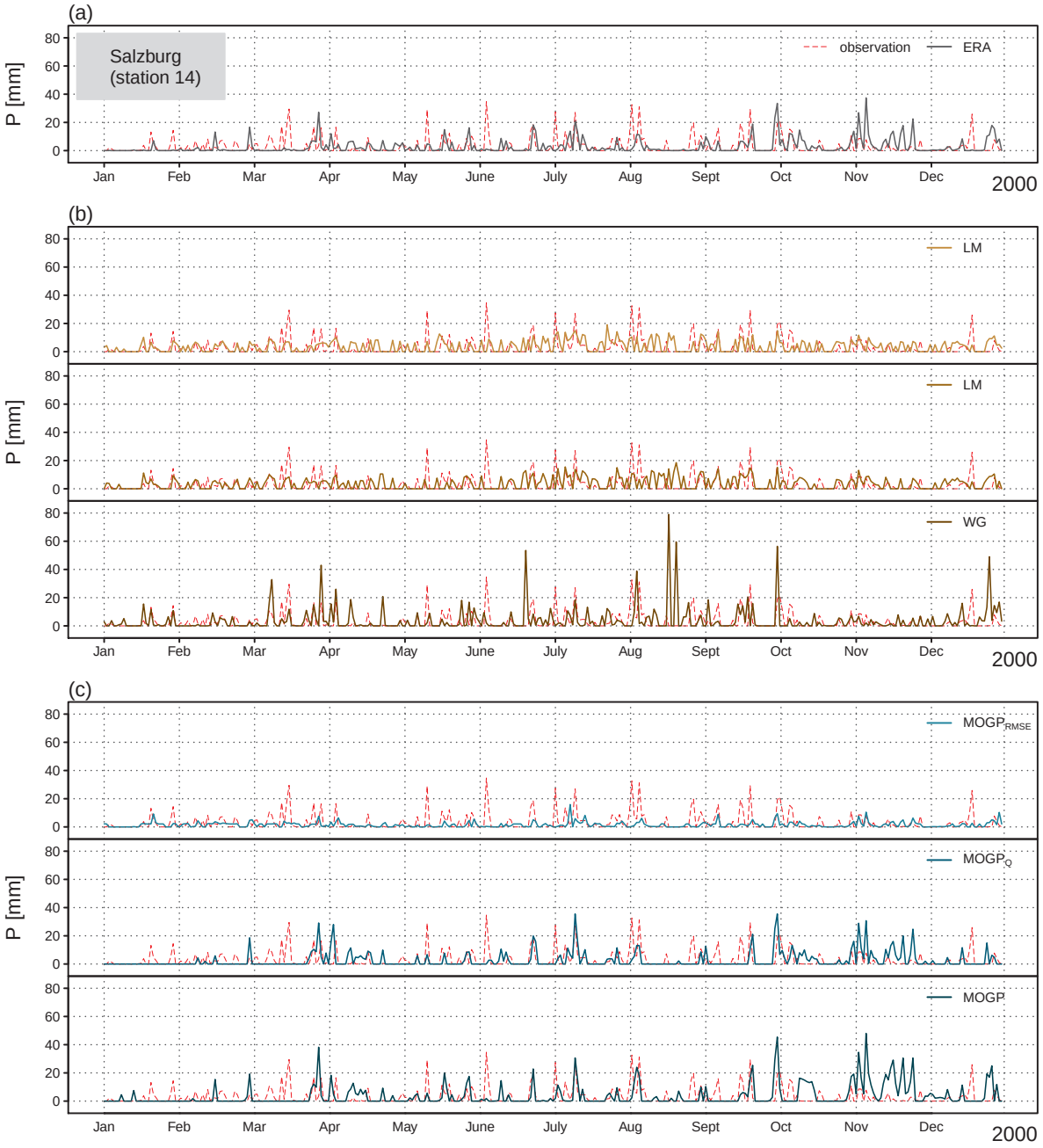

Figure 7.13.: Time series of daily accumulated precipitation P as observed, from ERA-Interim and from the different downscaling techniques at Salzburg (station 14) in the year 2000. The dashed red lines show the observations. The solid lines show the precipitation from ERA-Interim and from the different downscaling approaches.

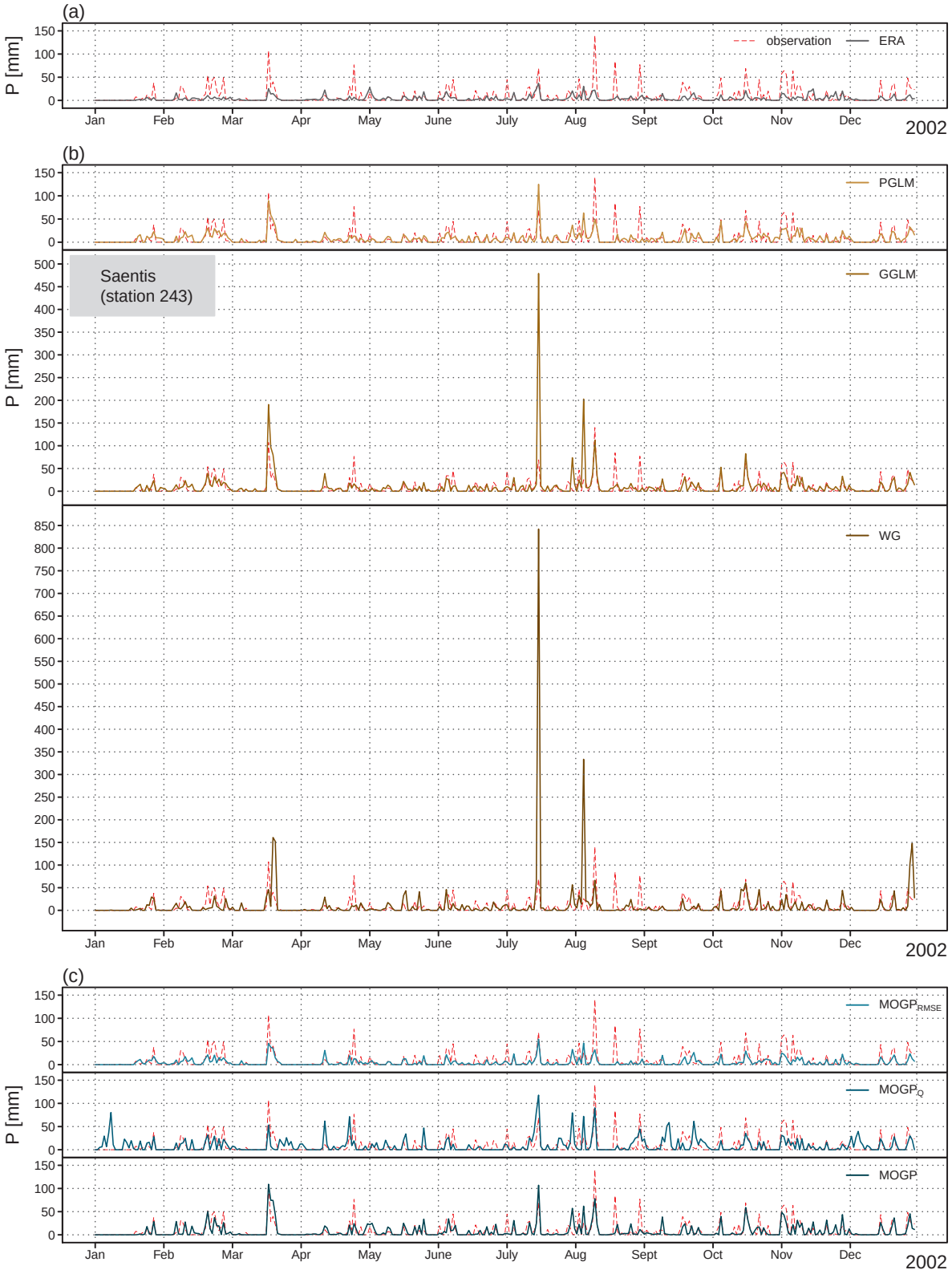

Figure 7.14.: Time series of daily accumulated precipitation P as observed, from ERA-Interim and from the different downscaling techniques at Saentis (station 243) in the year 2002. The dashed red lines show the observations. The solid lines show the precipitation from ERA-Interim and from the different downscaling approaches.

Table 7.8.: Objective functions and related measures for daily accumulated precipitation P predicted by the different downscaling methods at Salzburg (station 14) and Saentis (station 243): Error of precipitation occurrence frequency E(F), Bias B, root mean square error RMSE, error of standard deviation E(STD), Pearson correlation coefficient ρ, integrated quadratic distance IQD, error of selected quantiles E(Q_i). The quantiles of the observed CDF are $Q_{0.5}$ = 0.1 mm, $Q_{0.75}$ = 3.5 mm, $Q_{0.95}$ = 16.8 mm, $Q_{0.99}$ = 31.9 mm and $Q_{0.995}$ = 37.5 mm at Salzburg and $Q_{0.5}$ = 0.5 mm, $Q_{0.75}$ = 8.5 mm, $Q_{0.95}$ = 35.5 mm, $Q_{0.99}$ = 64.0 mm and $Q_{0.995}$ = 73.0 mm at Saentis.

(a) Salzburg

	ERA	**PGLM**	**GGLM**	**WG**	**MOGP$_{RMSE}$**	**MOGP$_Q$**	**MOGP**
E(F) [1]	0.15	-0.02	-0.00	-0.02	-0.07	-0.01	-0.08
B [mm]	-0.37	-0.02	0.07	0.05	-1.94	-0.25	0.10
RMSE [mm]	7.84	6.66	6.78	8.97	7.30	9.35	8.89
ρ [1]	0.24	0.35	0.33	0.20	0.19	0.20	0.16
E(STD) [mm]	-1.08	-2.72	-2.59	0.43	-3.42	0.98	0.03
IQD [mm^2]	0.001	0.013	0.014	0.0001	0.017	0.0004	0.0004
E($Q_{0.5}$) [mm]	0.40	-0.10	-0.10	-0.10	-0.10	-0.10	-0.10
E($Q_{0.75}$) [mm]	-0.30	2.60	2.60	0.10	-2.20	-0.40	0.60
E($Q_{0.95}$) [mm]	-2.59	-5.90	-5.59	-0.90	-9.70	-1.00	0.70
E($Q_{0.99}$) [mm]	-2.90	-16.50	-15.76	2.44	-14.86	-1.11	0.70
E($Q_{0.995}$) [mm]	-4.20	-20.22	-19.14	6.90	-14.76	-1.38	2.30

(b) Saentis

	ERA	**PGLM**	**GGLM**	**WG**	**MOGP$_{RMSE}$**	**MOGP$_Q$**	**MOGP**
E(F) [1]	0.27	-0.01	-0.01	-0.01	0.27	-0.18	0.03
B [mm]	-3.94	0.19	0.95	0.82	-1.96	-0.31	0.14
RMSE [mm]	12.51	11.91	20.06	31.15	11.22	43.27	12.68
ρ [1]	0.55	0.59	0.47	0.32	0.60	0.10	0.55
E(STD) [mm]	-9.22	-1.50	8.72	18.87	-6.13	28.68	-0.64
IQD [mm^2]	0.046	0.021	0.015	0.002	0.022	0.003	0.001
E($Q_{0.5}$) [mm]	1.10	4.00	3.45	0.10	1.70	-0.50	0.50
E($Q_{0.75}$) [mm]	-3.90	2.40	1.70	-1.60	-1.10	-0.40	1.30
E($Q_{0.95}$) [mm]	-23.40	-10.40	-7.80	-1.40	-15.90	-2.49	-2.40
E($Q_{0.99}$) [mm]	-42.86	-13.70	10.30	24.52	-28.86	-4.41	-2.90
E($Q_{0.995}$) [mm]	-47.95	-6.44	33.66	66.24	-31.03	-2.59	-0.02

The distribution estimated by MOGP resembles the observed distribution well, as do MOGP$_Q$ and WG. MOGP$_{RMSE}$ strongly underestimates the higher quantiles. Concerning the extremes indicated by the quantile-quantile plots for 99%- to 99.99%-quantiles (cf. Fig. 7.16) larger differences between the methods are apparent. The higher the considered quantile the stronger the underestimation for PGLM and GGLM. The WG achieves a better fit, but throughout overestimates the high quantiles by up to about 10 mm. The best match is achieved by MOGP$_Q$ and MOGP. Up to the 99.6%-quantile the MOGP predicted distributions closely match the observations. More extreme quantiles are underestimated. However, estimating such high quantiles from limited time series is highly uncertain anyhow.

For Saentis the difference between observed and modeled distributions is much larger. ERA-Interim generally underestimates the observed precipitation. PGLM and GGLM correct this mismatch in large parts, but again both GLMs overestimate low and underestimate the higher quantiles (cf. Fig. 7.15). The WG achieves a good match up to the 98%-quantile. MOGP$_{RMSE}$ throughout underestimates the quantiles. As for Salzburg, MOGP$_Q$ and MOGP

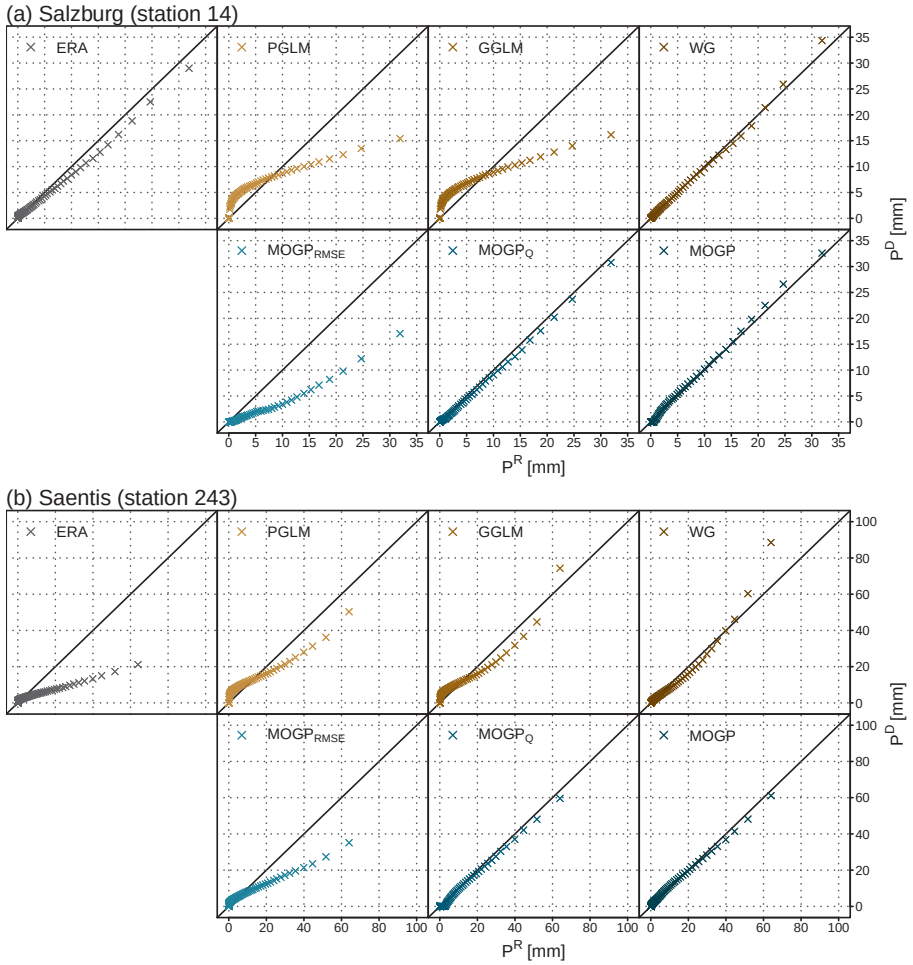

Figure 7.15.: Quantile-quantile plots of 1%- to 99%-quantiles for daily accumulated precipitation P at Salzburg (station 14) and Saentis (station 243) from ERA-Interim and estimated by the different downscaling techniques. The x-axis corresponds to the observed quantiles; the y-axis corresponds to the estimated quantiles. Shown are the 1%- to 99%-quantiles.

yield the best results also concerning the extremes (cf. Fig. 7.16). Also the PGLM performs reasonably well for 99%- to 99.99%-quantiles compared to ERA-Interim. GGLM and WG largely overestimate the extremes.

The downscaling rule for Saentis shown in Figure 7.17 reads:

IF $v_{1000}/u_{700} > 0.71$ and $u_{500} > 0.48$

$$P^D = P^{ERA} - q_{850}\ v_{250}\ (u_{500} + 0.01) + v_{1000} + 0.1$$

ELSE

$$P^D = 2\ P^{ERA} - q_{850}\ v_{250}\ (u_{500} + 0.01) + 0.1.$$

Depending on wind conditions (u_{500}, u_{700}, v_{1000}) MOGP uses slightly different equations to calculate the downscaled precipitation. Depending on the wind direction an air mass reaching Saentis has already traveled across the Alps potentially raining out. Particularly for winds from south-south-west in the lower atmosphere (i.e., $u_{700} > 0$ and $v_{1000} > 0$ and $v_{1000} > 0.7\ u_{700}$), the first equation is chosen. The equations differ only concerning two terms:

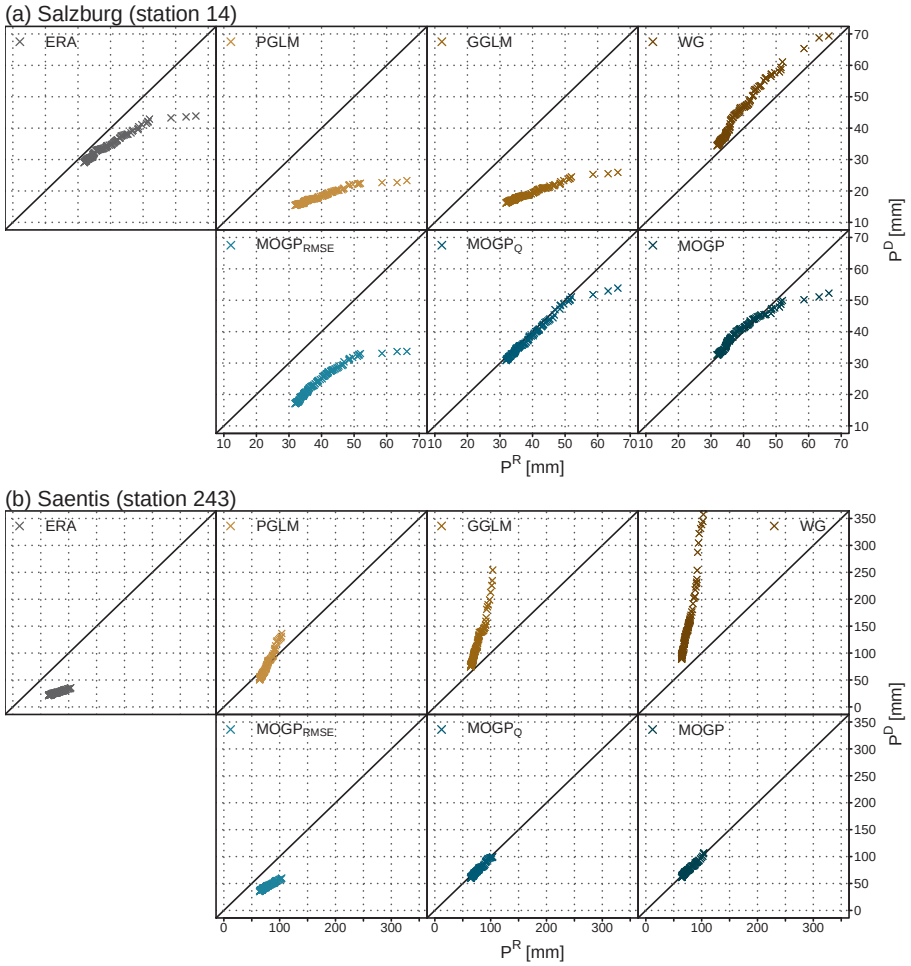

Figure 7.16.: Quantile-quantile plots of 99%- to 99.99%-quantiles for daily accumulated precipitation at Salzburg (station 14) and Saentis (station 243) from ERA-Interim and estimated by the different downscaling techniques. The x-axis corresponds to the observed quantiles; the y-axis corresponds to the estimated quantiles.

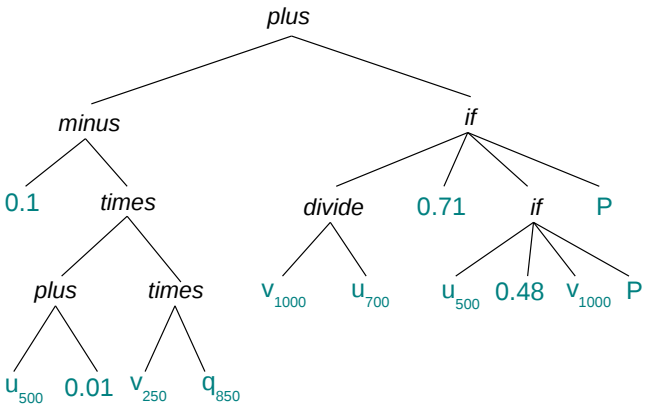

Figure 7.17.: An example downscaling rule as returned by MOGP for daily accumulated precipitation P at Saentis (station 243).

In the first equation v_{1000} is added. In the second equation there is a factor of 2 applied to the ERA-Interim precipitation, which has a strong impact on the downscaled precipitation especially for days with strong precipitation in ERA-Interim. Such a factor is not surprising as the precipitation amounts observed at Saentis are on average much larger than for the respective ERA grid box.
The term $-q_{850}\ v_{250}\ (u_{500} + 0.01)$ occurs in both equations. Specific humidity is multiplied with the meridional and zonal wind components at 250 hPa and 500 hPa geopotential height, respectively. For winds from the north-west or south-east the full term becomes positive otherwise it is negative. The higher the specific humidity the larger are both positive and negative effects. An exact and complete explanation of the downscaling rule is difficult. Any interpretation should be treated with care. With large-scale predictors such as EOF coefficients or weather classes one might obtain more easily interpretable rules than with grid box predictors only.

Spatial and Temporal Correlation

The autocorrelation of daily accumulated precipitation at lags of one to three days is shown in Figure 7.18. As observed for temperature, also the autocorrelation of the precipitation series from ERA-Interim are for the majority of stations larger than for the observations. This is foremost caused by ERA-Interim representing area means with by tendency smaller temporal variance compared to the observations. Precipitation series typically have a much smaller autocorrelation than temperature series. The average lag-1 autocorrelation of the observed precipitation equals 0.22. Thus, the autocorrelation error of up to 0.2 as found for all considered downscaling methods are large. PGLM, MOGP$_{RMSE}$ and MOGP overestimate autocorrelation for most stations. GGLM and MOGP$_Q$ show large errors in both positive and negative directions. The WG underestimates autocorrelation for most stations. At lag-1 the errors are largest for all methods. For the larger lags the over- and underestimation tendencies persist, but the magnitude of the errors decreases for all methods. However, also the observed

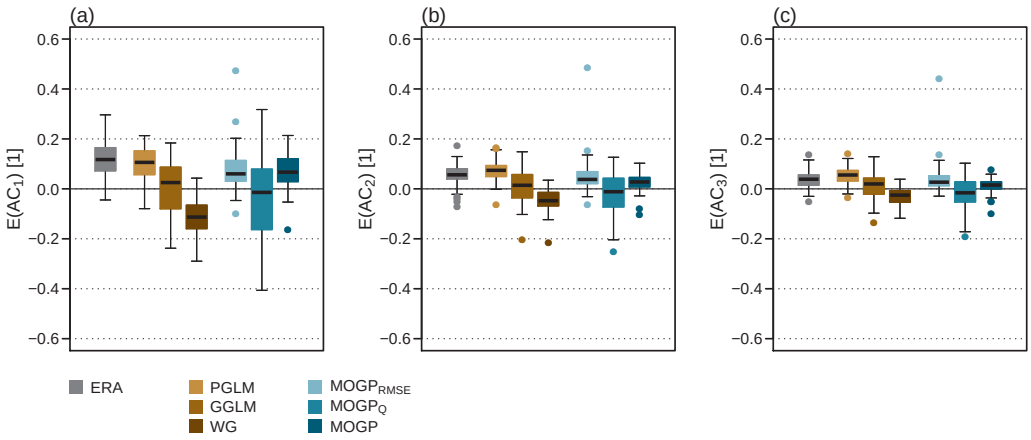

Figure 7.18.: Error of the lag-1 (a), lag-2 (b) and lag-3 (c) (Pearson) autocorrelation of the daily accumulated precipitation P series predicted by the different downscaling techniques. The observed autocorrelation averaged over all stations at different lags is $AC_1 = 0.22$, $AC_2 = 0.09$ and $AC_3 = 0.07$.

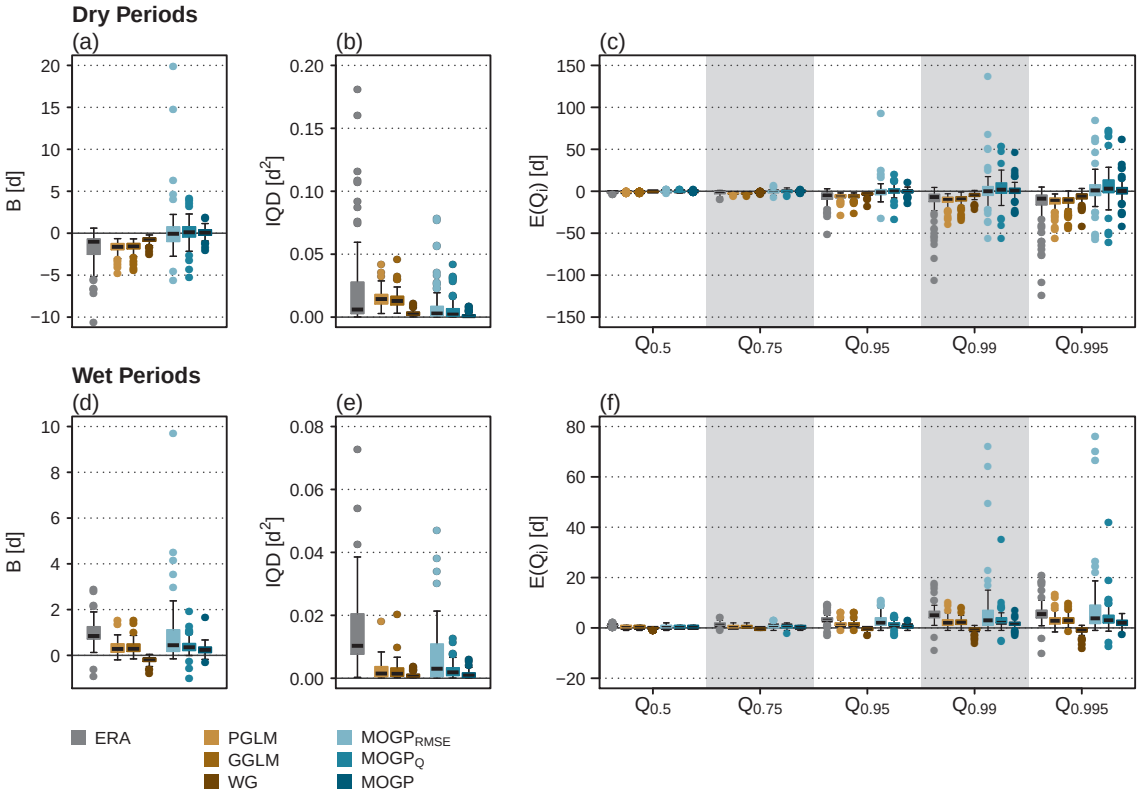

Figure 7.19.: Bias, integrated quadratic distance IQD and error of selected quantiles $E(Q_i)$ for the probability density functions of dry and wet period duration in days [d]. The mean quantiles of the obervations are $Q_{0.5} = 3.2$ d, $Q_{0.75} = 6.8$ d, $Q_{0.95} = 18.0$ d, $Q_{0.99} = 31.6$ d and $Q_{0.995} = 37.6$ d for the length of periods and $Q_{0.5} = 1.4$ d, $Q_{0.75} = 2.5$ d, $Q_{0.95} = 5.0$ d, $Q_{0.99} = 8.0$ d and $Q_{0.995} = 9.2$ d for the length of wet periods.

autocorrelation decreases with larger lags.

Note that no method explicitly models autocorrelation. The performance of the WG concerning temporal correlation might be improved by accounting for temporal correlation when sampling by conditioning each value on the preceding value(s). The performance of MOGP might be improved by adding an objective quantifying autocorrelation and by adding a possibility to incorporate noise in the downscaling rules (cf. Sec. 7.4).

The length of wet and dry periods is closely linked to the autocorrelation of precipitation series. From the full precipitation series at each station, we have calculated the empirical CDFs of the length of wet and dry periods, i.e., the number of consecutive wet or dry days, at the respective station. The bias, the integrated quadratic distance and the error of selected quantiles for the length of wet and dry periods are shown in Figure 7.19.

ERA-Interim has too many days with small precipitation amounts compared to the point observations leading to an underestimation of the dry period lengths and an overestimation of the wet period lengths for most stations. This is expected due to ERA-Interim representing area averages. PGLM, GGLM, WG and MOGP clearly reduce the bias. PGLM and GGLM systematically underestimate dry period lengths and overestimate wet period lengths. The WG further reduces the bias, but underestimates both average dry and wet period lengths slightly. The absolute bias for MOGP and WG is similar, but MOGP shows less systematic

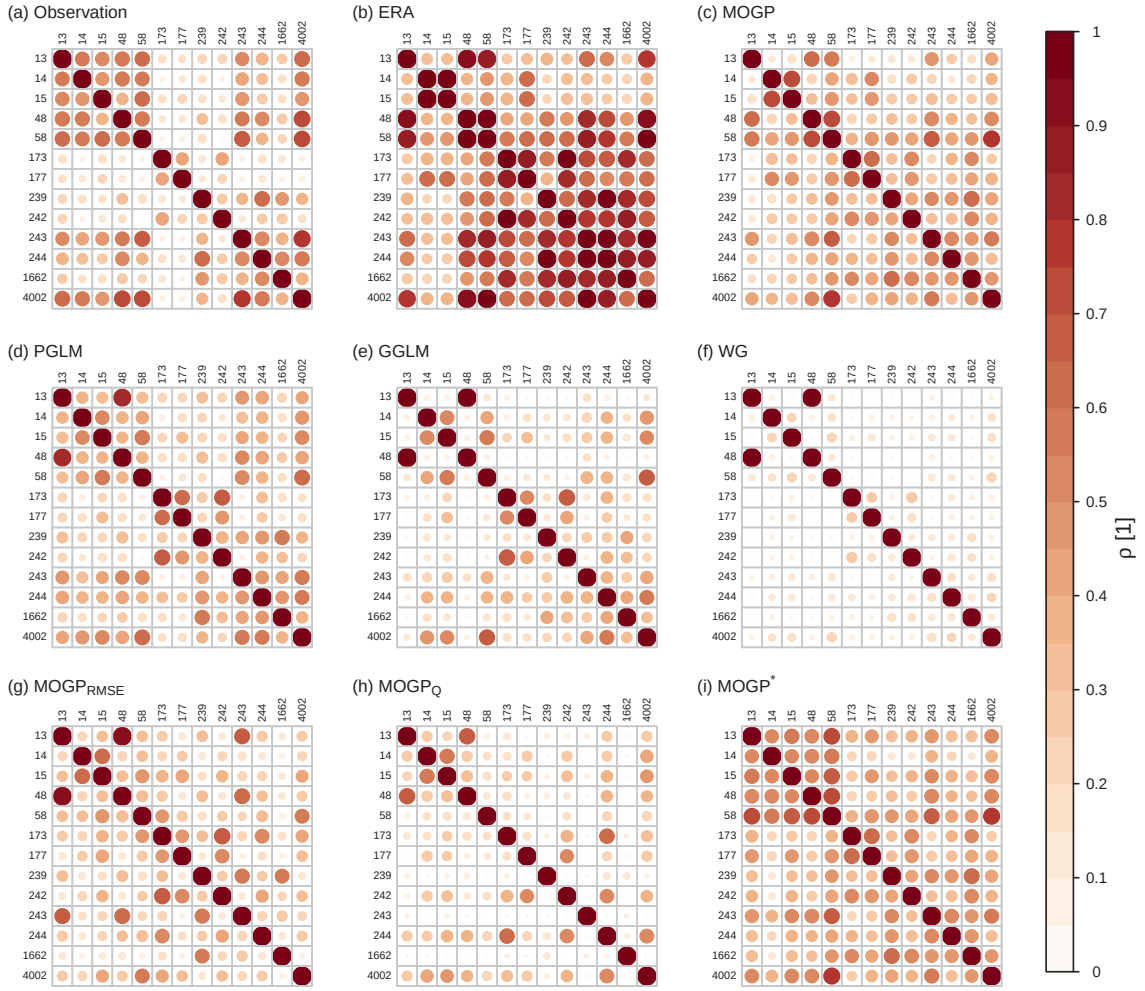

Figure 7.20.: Pearson correlation matrix of daily accumulated precipitation P for all stations within the Alpine (AL) PRUDENCE region (cf. Fig. 7.1).

under-/overestimation tendencies, especially for the dry periods.

WG and MOGP best reduce the IQD. MOGP performs slightly better for the dry period lengths. The WG performs slightly better for the wet period lengths. For the high quantiles ($Q_{0.99}$ and above) the WG shows a slight tendency to underestimate the respective quantiles. That is, the longest dry and wet periods for WG are too short compared to the observations. For MOGP the absolute errors of the high quantiles are comparable to the WG, but MOGP shows no systematic over-/or underestimation for the dry period lengths. For the wet period length MOGP overestimates the high quantiles on average, i.e., the longest wet periods predicted by MOGP are too long compared to the observations. Especially the very long dry and wet spells are not adequately represented. As argued above incorporating the representation of autocorrelation or the CDF of wet and dry period lengths into the MOGP objectives might improve its performance.

Spatial correlation is investigated by calculating Pearson correlations between the 80 stations. The bias and RMSE of the spatial correlation for all stations and for selected regions is provided in Table 7.9. The correlation matrices for the selected regions are shown in Figs. 7.20 to

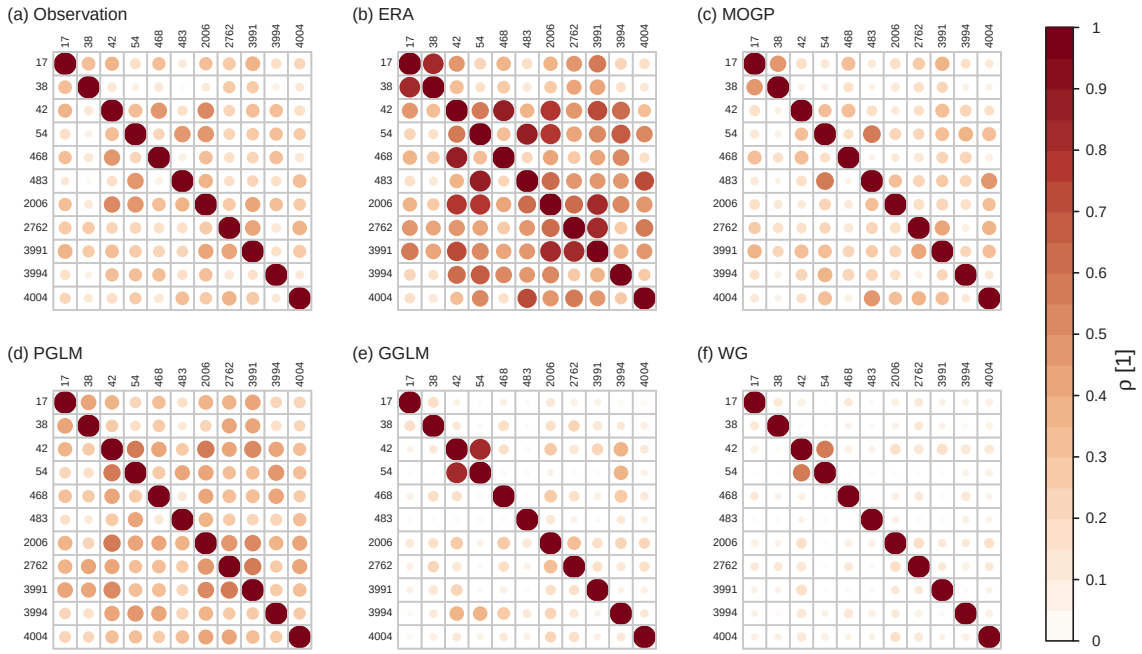

Figure 7.21.: Pearson correlation matrix of daily accumulated precipitation P for all stations within the Mid-Europe (ME) PRUDENCE region (cf. Fig. 7.1).

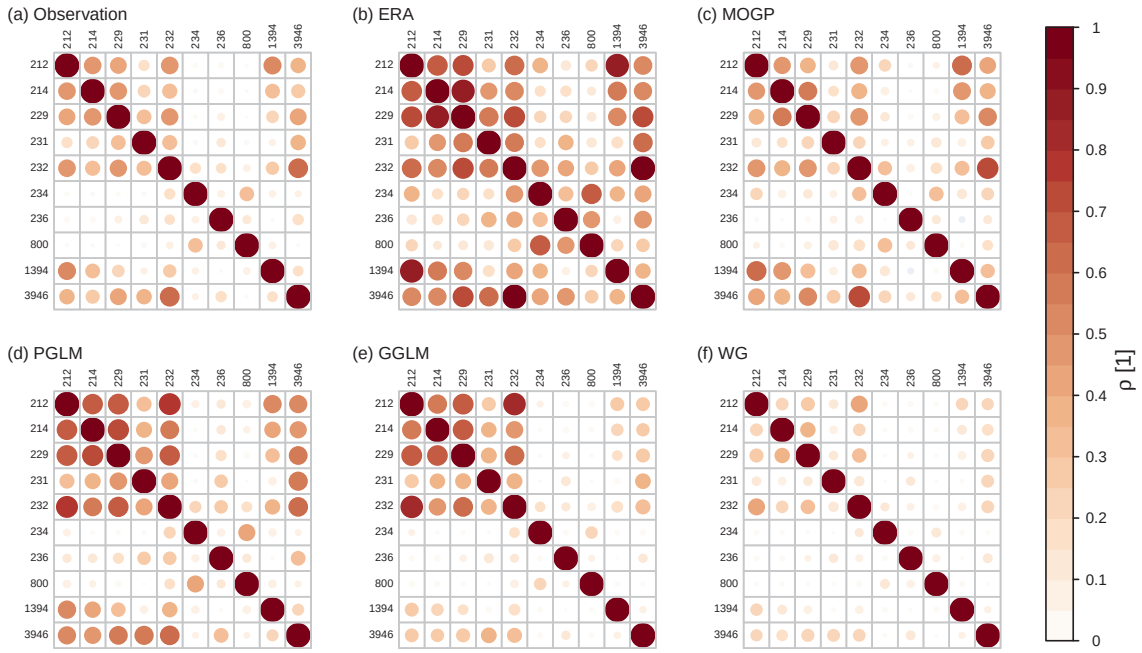

Figure 7.22.: Pearson correlation matrix of daily accumulated precipitation P for all stations within the Iberian Peninsular (IP) PRUDENCE region (cf. Fig. 7.1).

Table 7.9.: Bias B and root mean square error RMSE of the Pearson correlation coefficients for daily accumulated precipitation P between all 80 station considered and for three station subsets according to the PRUDENCE regions (cf. Fig. 7.1).

	ERA	**PGLM**	**GGLM**	**WG**	**MOGP**$_{RMSE}$	**MOGP**$_{Q}$	**MOGP**
B(SC)	0.039	0.038	0.005	-0.021	0.009	-0.025	0.003
RMSE(SC)	0.117	0.074	0.075	0.077	0.079	0.095	0.059
				Alps (AL)			
B(SC)	0.274	0.008	-0.098	-0.223	-0.030	-0.145	0.057
RMSE(SC)	0.359	0.109	0.222	0.289	0.219	0.278	0.195
				Mid-Europe (ME)			
B(SC)	0.196	0.084	-0.119	-0.169	-0.020	-0.210	-0.033
RMSE(SC)	0.227	0.105	0.188	0.207	0.125	0.243	0.010
				Iberian Peninsula (IP)			
B(SC)	0.230	0.096	0.007	-0.083	0.074	-0.146	0.037
RMSE(SC)	0.244	0.123	0.109	0.125	0.135	0.206	0.094

7.22. Note that the inter-station correlation has not been taken into account when evolving the MOGP downscaling rules and neither when extracting single rules from the Pareto sets. Keeping this in mind, the correlation between the stations is satisfactorily reproduced with a marginal bias of 0.003 and an RMSE of 0.059, which is clearly smaller than the respective values for the reference methods and the MOGP variants (cf. Table 7.9). The error values vary when considering only selected regions.

The correlation matrix for the Alpine (AL) stations (cf. Fig. 7.20) of the observed series shows a distinct pattern, which is largely related to the distance between the stations and their location within the terrain. For the Alpine region spatial correlation is, with an RMSE of 0.109, best modeled by the PGLM. MOGP tends to overestimate the correlations for stations with low observed correlations leading to a positive bias of 0.057. With a slightly different selection of rules from the Pareto set (MOGP*) the observed correlation pattern can be better reproduced (cf. Fig. 7.20(h)). Still, for the weakly correlated station pairs some overestimation tendency remains. As our MOGP system currently does not contain noise, this is not surprising. The Alpine stations are located within a small area and all (temporal) variance at the single stations is modeled using the coarse-scale predictors, which may be too strongly correlated to obtain low enough correlations between some station pairs.
For the Mid-European stations (cf. Fig. 7.21) the PGLM overestimates spatial correlation on average, while GGLM and WG again systematically underestimate spatial correlation. MOGP reproduces the observed correlation pattern in large parts. Only the correlation between the station 2006 (Brocken) and the remaining stations is clearly underestimated.
For the Iberian Peninsula the observed correlation pattern is comparatively well resembled by all of the methods (cf. Fig. 7.22). However, again the PGLM overestimates while the WG underestimates the spatial correlation. GGLM and MOGP closely match the observed correlation pattern.

Predictor Statistics

Figure 7.23 shows how the different predictors are used by GGLM and MOGP for downscaling daily accumulated precipitation. For the GGLM the importance of each downscaling predictor is indicated by the respective regression coefficient $\bar{\beta}_i$ averaged over all 5 cross-validation periods. For MOGP the importance of the different predictors is quantified by the average frequency $\overline{h_i}$ with which the different predictors occur in the final Pareto sets (cf. Eq. 7.17). Similar to temperature downscaling (cf. Fig. 7.10) the importance of the different predictors strongly depends on the station under consideration. The larger the $\overline{\beta_i}$ and $\overline{h_i}$ averaged over all stations (indicated by the squares), the wider the respective bars denoting the range of 10%- and 90%-quantiles. For the GGLM this tendency is much more pronounced than for MOGP. The ranking of the different predictor quantities is approximately the same as for downscaling daily maximum temperature. The clustering of the predictor variables (different colors) is even more pronounced for precipitation downscaling. The largest $\overline{\beta_i}$ are found for the geopotential predictors followed by temperature, specific humidity, wind speed and finally precipitation. The largest $\overline{h_i}$ are found for precipitation followed by wind speed, humidity, temperature and finally geopotential height (with the exception of z_{1000}). Quantities taken at lower levels are on average of higher importance for both methods than the respective quantities taken at higher levels. This is most apparent for the 1000 hPa geopotential height, which is a very important predictor for both methods. The frequent usage of P and z_{1000} as predictors for MOGP is physically reasonable, but might be largely due to their order of magnitude and not purely due to their physical meaning.
Compared to temperature, the MOGP rules for precipitation contain slightly more predictors on average. Still even the most frequently used predictors, P and z_{1000}, are on average contained in less than 35% of the Pareto optimal rules while the GGLM exploits the full predictor set. This hints at difficulties with the evolution of numerical constants and shows room for improvements. This is revisited in some more detail in Section 7.4.

Summary

- The reference methods (two GLMs and a simple weather generator WG) well predict the number of precipitation days and reduce the ERA-Interim bias. Also, the IQD is reduced well, especially by the WG.
- MOGP outperforms the reference methods concerning standard deviation and most significantly concerning the representation of extremes (frequency and intensity).
- No downscaling approach yields any significant reduction in RMSE. The smallest increase in RMSE observed for PGLM, MOGP$_{RMSE}$ and MOGP.

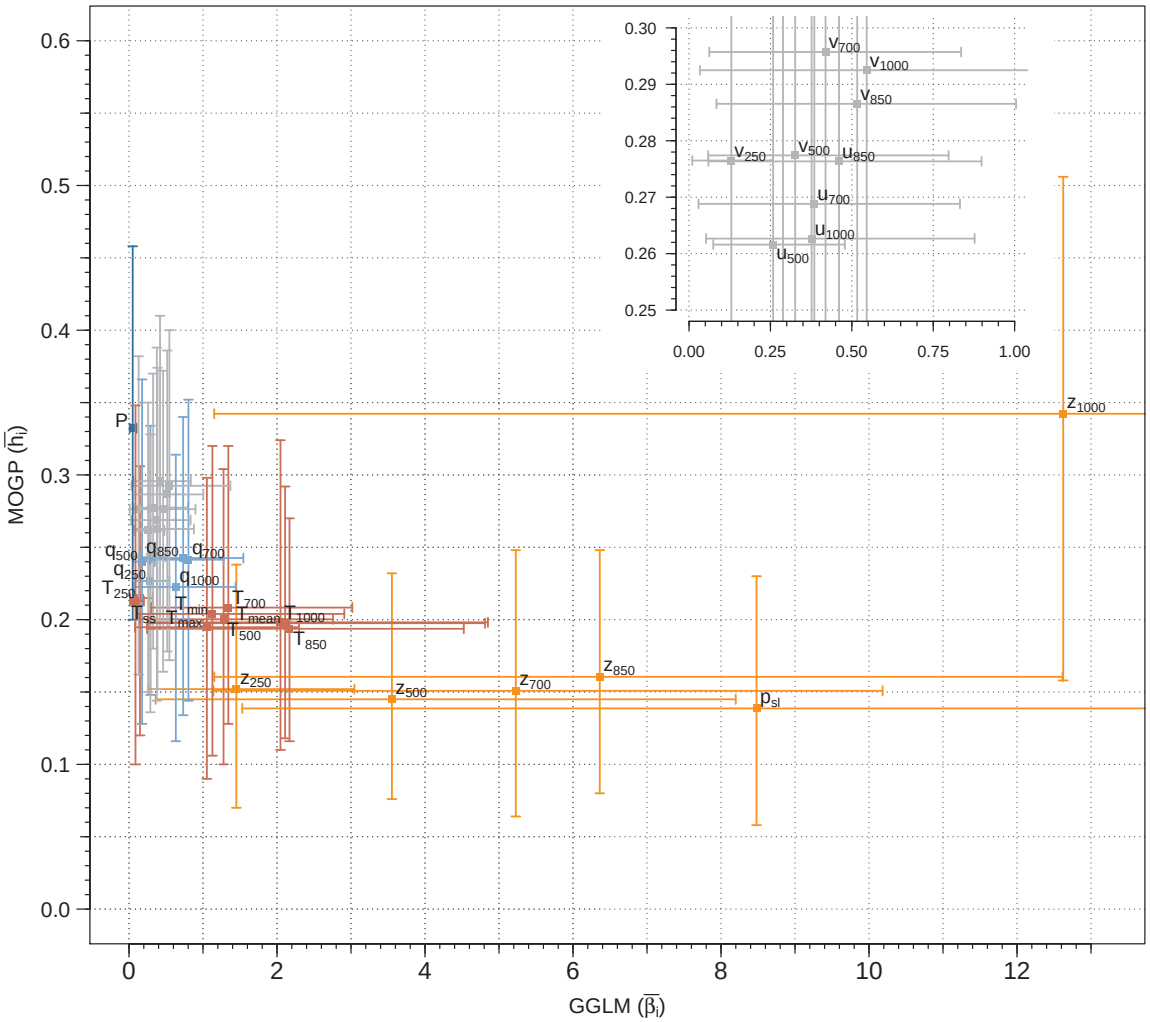

Figure 7.23.: Usage of the different predictors by multi-objective Genetic Programming (MOGP) and gamma generalized linear model (GGLM) for downscaling daily accumulated precipitation P at 80 stations. For the GGLM the regression coefficients β_i for each station averaged over all 5 cross-validation periods are given. For MOGP the frequency h_i with which the predictors x_i occur in the Pareto sets is given for all 80 stations and again averaged over all cross-validation periods. Details on the calculation are provided in the text. The squares indicate the mean values over all stations. The bars indicate the 10%-and 90%-quantiles. The colors emphasizse the different types of predictors considered. Temperature predictors T are shown in red; geopotential height z and pressure p based predictors are orange; wind speed predictors u, v are grey; specific humidity q is shown in light blue and precipitation P in darker blue. For z_{1000} and p_{sl} the 90%-quantile lies outside the plotted range ($Q_{90}(z_{1000}) = 24.7$, $Q_{90}(p_{sl}) = 16.7$). The smaller figure top right is a zoom into the main figure; the area is indicated by the axes.

- The differences between MOGP and the reference methods are larger for stations with high average daily precipitation. Especially for stations with high average daily precipitation MOGP outperforms the GLM based methods.
- The length of dry and wet periods are best captured by MOGP and WG. The representation of long wet and dry spells is not satisfying.
- Considering all 80 stations MOGP best reproduces spatial/inter-station correlation.
- As already observed for temperature downscaling, the frequency of predictor usage in MOGP appears to be strongly dependent on their order of magnitude presumably caused by difficulties when evolving numerical constants.

7.4. Discussion and Outlook

Temperature

In the current setup MOGP is outperformed by multiple linear regression for temperature downscaling. There are several reasons for this. The first reason is the size restriction applied to the MOGP parse trees. In our setup the parse trees are restricted to 6 levels, which (when no if statement is contained in the tree) amounts to a maximum of 32 terminal nodes. In case a parse tree represents a multiple linear regression equation (i.e., each predictor is assigned a numerical coefficient) the size restriction corresponds to a maximum of 16 predictors. This means the MOGP rules are restricted w.r.t their size and thus also the number of predictors while the multiple linear regression (in the current setup without any predictor selection) exploits the full predictor set.
Second, evolving appropriate numerical constants for the terminal nodes is a known weakness of Genetic Programming. Genetic operators such as *numeric mutation*[4] have shown some ability to improve GP in this respect (*Evett and Fernandez*, 1998). Such techniques have not been used in this thesis.
Finally, the number of individuals per generation as well as the total number of generations created and evaluated during an MOGP run might limit its performance. Additional longer runs with a total number of 1000 generations, a population size of 300 and a maximum Pareto set size of 200 individuals have been carried out for the stations Salzburg and Saentis for all 5 cross-validation periods. To allow for more complex solutions potentially containing larger numbers of predictors the maximum tree size is set to 20 levels for the longer runs. The minimum values for the mean error of standard deviation and the error of quantiles occurring in the Pareto sets can be significantly reduced by the longer runs/more complex solutions. For Salzburg no RMSE reduction has been achieved. For Saentis the best rules clearly reduce

[4] *Numeric mutation* is a genetic operator (like mutation or crossover), which replaces all numeric constants of an individual. The new numeric constants are chosen at random from a uniform distribution within a certain range specified as the old value of that constant plus or minus a so-called temperature factor.

RMSE [mm]						
ERA	LM	WG	$MOGP_{RMSE}$	$MOGP_{Q}$	MOGP	$MOGP'_{RMSE}$
11.13	1.42	1.98	3.23	3.89	4.08	1.76

Table 7.10.: Root mean square error (RMSE) for ERA-Interim and downscaled series of daily maximum temperature at Saentis. All values except for $MOGP'_{RMSE}$ refer to the whole time series (1979-2008) and have resulted from the MOGP setup as described in Sec. 7.2. $MOGP'_{RMSE}$ gives the minimum value of RMSE achieved by a downscaling rule of the Pareto set for one of the cross-validation periods as returned by one of the longer MOGP runs. The detailed setup of these runs is given in the text.

the RMSE by about 40-50% for 3 of the 5 cross-validation periods. For the other two runs no reduction in RMSE has been observed. This is most likely not caused by the different training data sets, but rather an effect of the nondeterministic nature of GP. Changes in the MOGP setup can, thus, improve the MOGP results (cf. Table 7.10). However, the longer the MOGP runs and the more complex the downscaling rules, the more prone MOGP is to overfitting. Which run dimensions (number of individuals evaluated and maximum tree size allowed) provide a reasonable setup, should be carefully assessed especially when carrying out very large MOGP runs.
Overall a linear predictor-predictand relation appears reasonable for daily temperature. MOGP cannot evolve better solution structures than provided by multiple linear regression. The MOGP performance might be improved by optimizing run dimensions and incorporating numeric mutation or similar operators. Considering the complexity and higher computational effort of MOGP a standard multiple linear regression is the favorable approach for the temperature downscaling problem described.

Precipitation

For precipitation the MOGP based downscaling is promising as MOGP outperforms the reference methods for most performance measures considered, most dominantly for the representation of precipitation extremes (w.r.t. event intensity and return frequency) and the spatial correlations between the stations. The reference methods outperform MOGP only for the frequency of precipitation days. However, the reference methods model precipitation occurrence and precipitation amount separately while MOGP models both with the same equation with negative values predicted being set to zero (representing dry days). In the future one might introduce a two step procedure also for MOGP. Or, as the logistic regression seems to work well for modeling precipitation occurrence, one might combine logistic regression and MOGP. As for temperature, the MOGP downscaling rules use less predictors than the generalizd linear models which exploit the full predictor set. In case of precipitation this appears not to be disadvantageous for MOGP. Precipitation downscaling is considered more difficult than temperature downscaling due to its non-Gaussian distribution. For daily precipitation often a gamma distribution is assumed like in two of the reference methods considered. Especially for many stations with large observed precipitation amounts, the assumption of a gamma distribution is questionable leading - in our setup - to a strong overestimation of extremes. The reference methods should not be regarded as state-of-the-art, but especially the gamma

distribution based methods constitute well known and widely applied approaches. For precipitation MOGP succeeds to evolve model structures better suited for precipitation than standard generalized linear models. As for temperature the MOGP performance might be improved by incorporating numeric mutation or similar operators, by running larger number of generations containing higher numbers of individuals each or by relaxing the size constraints of the solutions. Due to the promising performance for precipitation, the effect the MOGP setup (tree size constraints, Pareto set size etc.) on the performance should be further investigated.

What is new?

The presented application adds to earlier studies on downscaling reanalysis data by means of Genetic Programming (cf. Sec. 4.4) in two ways: (1) The Genetic Programming based downscaling is applied to a set of 86 stations rather than one (or an average over multiple stations). (2) A multi-objective approach is introduced to more closely reproduce the temporal variability of the station time series.
We have seen that the performance of MOGP evaluated in comparison to standard reference methods differs for from station to station for both temperature and precipitation. For temperature the current MOGP setup performs best (relative to the reference methods) when the temperature bias of ERA-Interim for a station is small. For precipitation MOGP performs best (relative to the reference methods) when there are large precipitation amounts observed at a station. This underlines the importance of considering multiple stations with different regions and climates when evaluating downscaling techniques.
The described multi-objective optimization approach, which is popular among researchers in evolutionary computation, has to our knowledge not been utilized for downscaling problems before. The MOGP downscaling rules represent a trade-off between minimizing RMSE and reproducing variability. As illustrated by the comparison with the two variants MOGP_{RMSE} (foremost minimizing the RMSE) and MOGP_{Q} (foremost optimizing the representation of temporal variability), the value of the multi-objective optimization (especially for precipitation downscaling) lies in the concept of fitting a trade-off between several, pointwise (RMSE) and distribution based (E(Q), IQD, E(STD)), objectives. The two variants perform clearly worse than MOGP when considering all objectives. For precipitation only a slight increase in RMSE can be traded for a strong decrease of the IQD (compare MOGP_{RMSE} and MOGP). A simple approach to add noise by sampling from a gamma distribution (compare MOGP and WG) leads to strong increase of RMSE for almost all stations considered.

Appropriate Representation of Variance

The problem of how to meaningfully reproduce variance with downscaling approaches has been addressed in several studies. Most approaches first apply a downscaling technique that minimzes the RMSE and afterwards add variability to match the variance of the observations more closely. In inflation (*Karl et al.*, 1990) and similar approaches (e.g., *Bürger*, 1996) all variance is deterministically related to the predictor variables. In the simplest case this is

done by multiplication with a (constant) factor. Also in the current MOGP setup all variance is deterministically related to the predictors meaning each time a downscaling rule is applied to a certain data set it will return exactly the same downscaled series.
According to *von Storch* (1999) (see also *Maraun* (2013) for discussion) fully deterministic approaches are physically unreasonable for two reasons. First, because the large scale weather does not fully determine local weather. Second, because variance inflation increases the RMSE. Especially the first point also applies to MOGP. The increase of RMSE is, however, minor. In *von Storch* (1999) a randomized downscaling approach combining a deterministic and a stochastic term is suggested, i.e.,

$$y = \hat{y} + y^*, \tag{7.18}$$

where the downscaled series y is given by the sum of $\hat{y}$, which deterministically depends on the predictors, and a noise term y^*. The noise does not need to be white in time (or space). The variance and the autocorrelation of the noise may depend on the predictor variables. Such an approach can also be used in MOGP. The full parse tree (downscaling rule) can be decomposed into two or three smaller trees as illustrated by Fig. 7.24. Tree 1 provides the deterministic estimate and variability not modeled by tree 1 can be added using additive or multiplicative noise. In the simplest case the noise is randomly drawn from a given probability distribution and (only) the variance or shape parameter is defined by tree 2. One can easily offer several different PDFs to generate noise to MOGP. Further, one can introduce an autocorrelation parameter, for instance simply the lag-1 autocorrelation, in a third tree. When doing so it would be necessary to include the autocorrelation also in the objective functions. Adding noise is not only physically reasonable, but might also improve the representation of spatial correlation.
The subtrees in Fig. 7.24 can be evolved in two conceptually different ways. They can be evolved simultaneously using a multi-objective GP setup such that MOGP 'decides' which portion of the full variability is deterministically related to the predictors (tree 1) and which portion is generated using the noise function (tree 2). Another way would be to set up a two-step procedure. That is, to evolve tree 1 and tree 2 in separate and consecutive steps. In the first step a (single-objective) GP algorithm can be used to minimze the RMSE (tree 1)

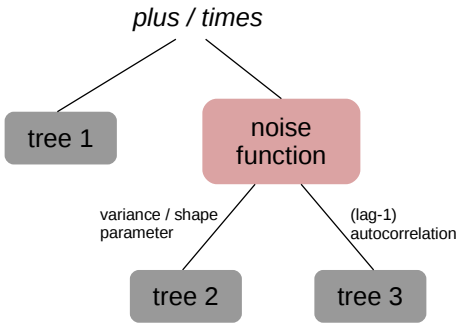

Figure 7.24.: Possible prescribed structure of a parse tree incorporating a noise generation function.

and in the second step GP is used to specify the noise. A useful objective for the second step (tree 2) might be the IQD, but also a multi-objective setup incorporating additional objectives quantifying the representation of extremes might be useful. The autocorrelation parameter might be optimized by using the autocorrelation, for instance simply at lag-1, as objective.

Suggestions for Future Work

The application of MOGP described in the present chapter should be regarded as an extensive first test. At the current stage the algorithm shall not be blindly applied to map climate projections to the local scale. However, the results for precipitation downscaling appear promising. Extremes are well represented and also spatial correlation is reasonably reproduced considering this is not explicitly contained in the objectives. Both representation of extremes and spatial correlation are knowingly difficult tasks in downscaling. We thus suggest further tests and adaptations for MOGP for precipitation downscaling.

The presented MOGP rules incorporate a large set of grid point predictors, i.e., atmospheric variables taken from the ERA-Interim grid box closest to the respective station. Several results (the comparatively bad performance for temperature downscaling and the statistics of downscaling predictors selected by MOGP) hint at known problems with an efficient evolution of numerical constants. In future MOGP applications predictors may be normalized before offered to the algorithm. It appears reasonable to repeat the MOGP runs for the presented downscaling experiment with a normalized predictor set and with predictor subsets selected by the user in advance. Such tests will help to find the most efficient way to set up MOGP. The multi-objective approach might be also applicable for model output statistics (MOS), which commonly uses solely GCM (or RCM) precipitation as predictor. Further, spatial predictors, for instance derived from principal component analysis, can provide a better physical basis for MOGP as predictors derived from the full 3D model output to capture the large-scale atmospheric weather patterns, which are neglected when using grid point predictors only.
However, not only the predictor set might be adapted, but also the function set offered to MOGP. Runs including logarithmic or exponential functions can be carried out. Further, the run dimensions (i.e., number of generations, population size and Pareto set size) and the size limitations applied to the downscaling rules can be optimized. Finally, the objective functions can be adapted. For instance the error of standard deviation might not be necessary when incorporating the integrated quadratic distance.

Downscaling methods are developed to map climate projections to smaller, regional or local, scales in order to estimate regional or local climate change and its impact. Whether a statistical downscaling link derived from a reanalysis and past observations holds under changing climate can not be fully determined. Future climate conditions may lie outside the range covered by the observational period. To get a first hint if there is any potential for a downscaling approach to hold under future conditions it is common to spit the observational record into periods with different characteristics (i.e., wet/dry years, hot/cold years) rather than coherent time periods (e.g., *Gutiérrez et al.*, 2013). If such a test gives satisfactory re-

sults, one might consider additional predictand variables like daily maximum wind speed, for instance.

In order to efficiently evaluate the large set of MOGP runs with differing set ups, the selection of downscaling rule(s) from the full Pareto set should be automated. A first idea how to obtain a subset of applicable Pareto optimal rules $\widetilde{\mathcal{P}'}$ from the full Pareto set $\mathcal{P}'$ is by neglecting rules that only yield a reasonable performance concerning a subset of the objectives

$$\widetilde{\mathcal{P}'} = \{\alpha \in \mathcal{P}' \mid \forall i,\ s_i(\alpha) \leq (1+\delta) \min_{\beta \in \mathcal{P}'}(s_i(\beta))\}. \tag{7.19}$$

In words, all downscaling rules that do not deviate more than δ concerning their performance w.r.t. the objectives s_i from the best values present in the Pareto set become part of $\widetilde{\mathcal{P}'}$. The tree size should probably be neglected here. All other objectives can potentially be included. In this way one may use the last rule remaining in $\widetilde{\mathcal{P}'}$ when increasing δ as an optimum trade-off between the objectives. It is further possible to assign more weight to certain objectives by assigning each objective s_i its own δ_i.
Up to now we have selected and evaluated single downscaling rules from the Pareto set returned by MOGP. In the future the Pareto optimal set of downscaling rules might be used to generate a *downscaling ensemble*. An ensemble approach may provide a possibility to assign uncertainty estimates to the downscaling results. In the current MOGP setup not all downscaling rules contained in the Pareto set are potentially physically reasonable. Thus not all rules constitute potential ensemble members. Some downscaling rules show an intolerable large RMSE (while yielding very good results for ME(Q) and/or ME(STD)); some show intolerably bad performance concerning several objectives (while being very small and thus performing very well concerning the solution size). Hence, we do not expect the Pareto set returned by MOGP to directly provide a reasonable ensemble. Similar to the idea for rule selection described above, the δ in Eq. 7.19 can be defined such that $\widetilde{\mathcal{P}'}$ contains the desired number of ensemble members. To obtain a sufficient set of reasonable rules the Pareto set size might have to be increased. Another option would be to adapt the clustering procedure within the MOGP algorithm such that the rules representing a trade-off between the objectives are more likely to remain in the Pareto set than those optimizing single objectives. However, this would be more complicated than the approach suggested above and may further hinder the evolutionary process.
In atmospheric sciences ensemble approaches are a common way to assess the uncertainty of weather and climate model predictions. In an ideal setup an ensemble would provide a random sample of the underlying PDF. When such an ensemble has a small spread (meaning all ensemble members are close to each other) this indicates a low forecast uncertainty. A large spread in return indicates a high uncertainty (e.g., *Wilks*, 2011). In reality, however, it is often impossible to ensure a truly random sample. Dynamical models contain errors from various sources. When aiming to generate a Pareto optimal downscaling ensemble from the MOGP predictions one faces a conceptually similar problem. It is nowhere ensured that all ensemble members are equally probable realizations. Thus, the estimation of the desired

underlying PDF has to be treated with caution. The most pragmatic approach for generating a Pareto optimal downscaling ensemble would be to select the ensemble members such that the resulting ensemble is well calibrated. If this is practical or would require too large and computationally demanding Pareto set sizes needs to be tested.

Summary

- For temperature downscaling MOGP does not provide obvious advancements compared to standard linear regression based downscaling approaches.
- For precipitation downscaling MOGP shows a strong potential to evolve reasonable trade-offs between minimizing (pointwise) RMSE and distribution based measures. Especially the representation of extremes is promising. Still some aspects need to be carefully investigated when aiming to make MOGP applicable for downscaling climate projections:
  - Optimize MOGP setup: run dimensions (population size, Pareto set size, total number of generations), function and terminal sets, objective functions, genetic operator probabilities, etc.
  - Carry out physically based predictor (pre-)selction; derive spatial predictors (e.g., from PCA analysis); predictor scaling.
  - Implement the ability to add noise to account for variability that is not deterministically related to the predictors.
  - Test the potential of MOGP based downscaling rules to hold under changing climate.
- In the long run MOGP might serve as an automatic and globally applicable downscaling algorithm for precipitation (and potentially also other non-Gaussian distributed atmospheric variables).
- The Pareto set returned by MOGP may be converted into a Pareto optimal ensemble in order to to obtain estimates of the uncertainties induced by the downscaling procedure.

8 Conclusion

This thesis introduces multi-objective Genetic Programming (MOGP) for the downscaling of atmospheric data. Genetic Programming (GP), an evolutionary computation technique, generates models (equations or program code) by analogy with evolution in nature (*Koza*, 1992). A population of candidate models is initialized and evolves over several generations following the principle of survival of the fittest. GP carries out a symbolic regression, i.e., a type of regression where both model structure and model parameters are a priori unknown. GP generates model structure and model parameters simultaneously to account for potential nonlinear and multivariate relations between predictand and predictor variables. The generated models, downscaling rules in our application, are coded as parse trees. They are therefore easily readable in contrast to artificial neural networks, which are currently the dominant approach for nonlinear regression problems.

The derived multi-objective approach following *Zitzler and Thiele* (1999) allows for the simultaneous optimization of root mean square error (RMSE), objectives quantifying spatial and/or temporal variability, and the complexity of the downscaling rules. To our knowledge, a multi-objective approach as described in this thesis and in *Zerenner et al.* (2016) has previously not been used for downscaling.

For the application of MOGP to the downscaling of spatial fields (Chapter 6) the objectives comprise the RMSE between downscaled and reference fields, the mean error of the subgrid-scale standard deviation, the integrated quadratic distance between the cumulative densities of downscaled and reference fields, and the size of the parse trees as an indicator for the complexity of the downscaling rules. For the application to the downscaling of global reanalysis data to temperature and precipitation series at local stations (Chapter 7) the objectives comprise, besides the RMSE, the bias between downscaled and reference time series, the error of the temporal standard deviation, the difference between the cumulative densities of downscaled and reference series, the error of selected quantilies (with a focus on extreme quantiles), the size of the parse trees, and for precipitation the error of the frequency of precipitation days. In the multi-objective setup MOGP returns not one downscaling model, but a set of Pareto optimal models. At the current stage one or two of these downscaling rules have been selected by hand and analyzed in more detail.

The application of MOGP to discover rules for downscaling near-surface atmospheric fields from 2.8 km to 400 m grid spacing is described in Chapter 6. This application aims at a scale-consistent two-way coupling between land surface models and atmospheric models as the latter are often applied with a larger grid spacing mainly for computational reasons. The subgrid-scale variability of the atmospheric fields is estimated from high-resolution information on land surface properties. For some atmospheric variables pixelwise relations between the atmospheric fields and the surface are found: near-surface pressure, near-surface temperature for unstable atmospheres, near-surface wind speed for large wind speeds and near-surface specific humidity. For these variables and conditions the downscaling approach reduces the RMSE and recovers between 30% (for specific humidity) and 97% (for pressure) of the subgrid-scale standard deviation. Also for stable atmospheres the subgrid-scale variability of temperature is well restored by the downscaling, but at the cost of a slightly increased RMSE. The same is observed for wind speed downscaling under calm conditions. For precipitation and incoming radiation no suitable downscaling rules were detected. However, for incoming radiation the terrain correction accounting for terrain slope, aspect and shadowing effects that is available within the COSMO model could be adapted to the subgrid scale. For precipitation downscaling a statistical approach, such as in *Schomburg et al.* (2010), can be used to recover subgrid-scale variability. The effect of the downscaling rules on the simulated exchange fluxes of mass, energy and momentum is the central aspect of future work. A more detailed discussion is provided in Section 6.4.
The application of MOGP to the downscaling of reanalysis data to the locations 86 stations in Europe is described in Chapter 7. Temperature (daily maximum, minimum, mean) and daily accumulated precipitation were downscaled using MOGP and several linear methods for comparison. For temperature a multiple-linear regression works well and outperforms MOGP. For precipitation MOGP outperforms all three linear methods considered (a Poisson distribution based generalized linear model, a gamma distribution based generalized linear model and a basic gamma distribution based weather generator) for the majority of the performance measures. Especially the representation of extreme events by MOGP is promising. A more detailed discussion is given in Section 7.4.

The high flexibility of MOGP also entails some disadvantages. Finding optimal run dimensions (population size, Pareto set size etc.) is not straightforward and may require tests for new applications and data sets. Another negative feature is the occurrence of bloat, i.e., the tendency of the downscaling rules to contain unnecessary terms, which is a common problem in GP. Bloat is reduced by applying size constraints to the parse trees and by incorporating the solution size in the objectives, but the generated rules occasionally still contain unnecessary terms. Such terms can be removed manually, for instance via a scale analysis. Moreover, the detection of numerical constants is a know difficulty in GP which is most likely a major reason for multiple-linear regression outperforming MOGP for estimating local temperature time series from reanalysis data.
The MOGP algorithm used in this thesis, which is based on the GPLAB, a Genetic Programming toolbox for Matlab (*Silva and Almeida*, 2003), may be optimized with respect to the

detection of numerical constants and bloat reduction in the future. The theory of Genetic Programming is a scientific field of it own constantly leading to new developments and advancements.

Despite these technical aspects three major points are suggested for further developments of multi-objective Genetic Programming for atmospheric downscaling:

1. Representation of Noise
   In the current setup the spatial and/or temporal variance of the predictand is (with the exception of few MOGP runs testing a basic noise generator) deterministically linked to the predictors. This is problematic as in reality the variability of the predictand is not fully determined by the predictors (*von Storch*, 1999; *Maraun*, 2013). Especially the simulation of processes such as turbulence or convection is subject to large uncertainties. Thus, variability generated by such processes is more appropriately represented by stochastic techniques. We thus suggest to split the parse tree into a deterministic and a stochastic subtree. The latter may link the amplitude of the noise and potentially also its spatial and temporal correlation to the predictands (cf. Sec. 6.4, 7.4).

2. MOGP as an Automated Downscaling System
   For climate impact studies, which require local climate information, automated downscaling algorithms are important. MOGP can be expanded to a fully automated downscaling system which requires the user to only to provide a sufficient set of training and testing data and, optionally, the objectives to be optimized (RMSE, spatial/temporal variance, extremes, autocorrelation, etc.). This requires two main steps. First, the selection of the downscaling rule(s) from the Pareto set needs to be automated. Second, the selected downscaling rules should be simplified (unnecessary terms removed), and tested automatically to make sure no unphysical values are generated. Such a fully automated algorithm would allow the user to easily apply the MOGP based downscaling algorithm to new data sets and for a large number of stations.

3. Downscaling Ensemble
   The set of Pareto optimal downscaling rules may be used to generate a downscaling ensemble. In atmospheric sciences ensemble approaches are a common way to assess the uncertainty of weather predictions and climate projections. An ensemble approach may provide an uncertainty estimate for the downscaling procedure. However, the Pareto set returned by MOGP is not expected to directly provide a reasonable ensemble. Thus an appropriate algorithm for selecting the ensemble from the full Pareto set needs to be defined (cf. Sec. 7.4). The generated ensemble may further require calibration. A similar approach is conceivable to estimate the sources of uncertainty that is induced by the downscaling procedure in a fully coupled atmosphere-land surface-subsurface simulation (cf. Sec. 6.4).

It can never be fully ensured that a statistical downscaling model trained using reanalysis data and observations holds under changing climate. As downscaling aims at deriving local

information from projections of future climate conditions, this is problematic for any statistcal downscaling technique and can strongly depend not only on the downscaling approach, but also on the considered predictors. Some careful testing with respect to this question should be carried out (cf. Sec. 7.4).
One may further argue that with increasing computational power and grid resolutions the need for downscaling techniques decreases. This is, however, a slow process as the computational cost quickly increases for smaller grid spacings. Moreover, whenever point data is required statistical downscaling is inevitable. MOGP and other statistical techniques are computationally much less demanding than high-resolution atmospheric modeling. Not every researcher has access to state-of-the-art computer clusters required for running atmospheric models at a high resolution.
Also for integrated subsurface-land surface-atmosphere simulations domain sizes will increase, grid spacing decrease with increasing computational power. Thus, the downscaling scheme may need to be retrained for smaller scales. However, as heterogeneities in soil, land surface and lower boundary layer are present up to very small scales, subgrid-scale variability and thereby a need for downscaling remains. Finally, MOGP can not only be used for downscaling, but for all kinds of regression problems where capturing the variability of the predictand is of high importance and the predictor-predictand relations are complex (i.e., not well captured by standard regression techniques) or unknown.

Appendices

A

Preliminary MOGP Runs

This section describes two tests we have carried out with MOGP before applying the methodology to the different atmospheric state variables (see Chapter 6). For the tests, near-surface temperature serves as predictand as it can exhibit very complex fine-scale patterns depending on atmospheric stability and thus offers a problem of sufficient complexity for testing the method. The first test (Test I) is aimed at finding a reasonable number of generations for the MOGP runs. As a second test (Test II) a cross-validation (i.e., leave-one-out) experiment has been carried out to test for overfitting. The results of the cross-validation experiment are in more detail described in *Zerenner et al.* (2016).

A.1. MOGP Setup

Like in Chapter 6 we estimate anomalies, i.e., the differences between spline-interpolated fields and high-resolution reference. Further, the coarse pixel mean is conserved by subtracting the mean anomaly over a coarse pixel that is predicted by a downscaling rule.

A.1.1. Objectives

Some objectives have been calculated slightly different for the preliminary MOGP runs compared to Chapter 6.
In the preliminary runs we have tested a neighborhood based root mean square error (RMSE) aiming not to punish small displacements between predicted y^D and observed anomalies y^R. The neighborhood based RMSE calculates as

$$RMSE = \sqrt{\frac{1}{n_t n_i n_j} \sum_{i,j,t} \min_{k,l \in U(i,j)} (y^R_{tij} - y^D_{tkl})^2} \tag{A.1}$$

where neighborhood $U(i,j) = \{(i,j),(i,j+1),(i,j-1),(i+1,j),(i-1,j)\}$ contains the the grid point (i,j) and its four direct neighbors and n_i and n_j denote the total number of pixels in x- and y-direction and n_t the number of training fields.
The mean error of standard deviation ME(STD) has been calculated analogue to Chapter 6. The mean integrated quadratic distance MIQD for the preliminary runs is calculated similar

to Chapter 6, but from the CDFs of discretized temperature distribution with a bin width of 0.25 K. The fourth objective, the SIZE of the solutions, is analogue to Chapter 6.

A.1.2. Parameters

For both tests we use the high-resolution COSMO model simulations by *Schomburg et al.* (2010) for training and validation. For computational reasons, single time steps have been extracted from the 8 simulation periods to create training and validation data sets (cf. Table A.1). For test I, one MOGP run has been carried out incorporating all days except for October 14$^{\text{th}}$ 2007 in the training data set. The remaining day can then potentially be used for a preliminary validation. For test II, eight MOGP runs have been carried out each omitting a different day in the training step (i.e., leave-one-out). Except for the total number of

Table A.1.: Dates and prevailing weather conditions of the high-resolution COSMO simulations by *Schomburg et al.* (2010) used to create the training (and validation data sets) for the MOGP tests. The original data set contains simulation periods of 1-2 days. To weight all simulation periods equally, we only include one day of each simulation period in the GP training data sets. The right column lists the time steps we have extracted for our training data set to reduce computational cost.

Date	Weather	Time Steps
27 Aug. 2007	varying cloud cover, no precipitation	03:00-04:00, 15:00-16:00
14 Oct. 2007	clear sky	11:00-12:00, 23:00-24:00
10 Mar. 2008	strong winds, variable clouds and precipitation	10:00-11:00, 22:00-23:00
2 May 2008	clouds and precipitation	00:00-01:00, 12:00-13:00
10 May 2008	clear sky	01:00-02:00, 13:00-14:00
7 June 2008	convective clouds and precipitation	05:00-06:00, 17:00-18:00
21 July 2008	synoptically driven stratiform rainfall	09:00-10:00, 21:00-22:00
28 Aug. 2008	cloudy, some rain	07:00-08:00, 19:00-20:00

Table A.2.: Predictors contained in the terminal set for downscaling near-surface temperature (Tests I and II). The fields of the atmospheric state variables are given at coarse resolution (i.e., 2.8 km), the quasi-static surface property fields are given at high resolution (i.e., 400 m). tp_1^* has been calculated in two steps: (1) averaging tp_1 to the coarse scale; (2) calculating the difference between the original tp_1 and the coarsened field.

Atmospheric Information (coarse)	
T_{10}	temperature at the lowest model layer (10 m)
T_{gr25}	vert. temp. gradient of lowest 2 layers ($\approx$ 25 m)
T_{gr60}	vert. temp. gradient of lowest 3 layers ($\approx$ 60 m)
T_{gr110}	vert. temp. gradient of lowest 4 layers ($\approx$ 110 m)
w_v	vertical wind speed at $\approx$ 20 m
w_h	horizontal wind speed at $\approx$ 10 m
$R_{net,s}$	net radiation at the surface
Surface Information (high-resolution)	
h	topographic height
h^*	topographic height anomaly
tp_1	mean height difference to neighboring grid points
tp_1^*	anomaly of tp_1
tp_2	slope to lowest neighboring grid point
tp_3	slope to highest neighboring grid point
tp_4	number of direct neighbors lower than grid point
PLC	plant cover
z_0	roughness length
α	albedo

Table A.3.: Summary of the MOGP settings for tests I (II). (Protected division means that division by zero returns the dividend not an error.)

Parameter	Value
objectives	RMSE (neighborhood based), ME(STD), MIQD, SIZE
function set	+, -, ×, protected /, if
terminal set	random numbers [0,1], variables (Table A.2)
generations	500 (200)
population size	100
max. Pareto set size	50
genetic operators	(subtree-)mutation, crossover
max. tree levels	5

generations created and evaluated the MOGP setup of tests I and II is basically the same (Table A.3). The potential predictors are selected based on our understanding of atmospheric processes, which influence near-surface temperature (Table A.2). The number of individuals per generation is set to 100. We run 500 (Test I) or 200 (Test II) generations, i.e., each run evaluates 50000 (Test I) or 20000 (Test II) potential downscaling rules. The maximum Pareto set size is set to 50, which provides a sufficient coverage of the Pareto front while not requiring too much memory space. For computational reasons and to keep the solutions readable, we furthermore limit the tree size to 5 levels. Besides the predictors described above, the terminal set contains random numbers drawn from the interval [0,1]. The function set contains the arithmetic functions with two input arguments each and an if statement with four input arguments (i.e., if $a > b$ do c else do d).

A.2. Results

A.2.1. Test I

Figure A.1 shows the performance of the Pareto set for each generation and the Pareto set size. Concerning the RMSE the Pareto set mean decreases only little with time. The Pareto set minimum, however, visibly decreases during the first approximately 60 generations. Concerning ME(STD) and MIQD a similar behavior is found. For the first 70 generations the Pareto set mean of ME(STD) and MIQD decreases rapidly. During the later generations the mean is barely reduced further. The SIZE of the downscaling rules increases with the number of generations evolved. The Pareto set size increases quickly up to the allowed maximum of 50 rules and after some oscillation stays constant during the later generations. The drops at approximately 40 and 60 generations are caused by one or more newly discovered downscaling rules outperforming several (here $\approx$ 20 to 30) rules of the Pareto set from the previous generation. The dominated rules are removed from the Pareto set and the Pareto set size shrinks. Summing up, a maximum number of generations of about 100 appears sufficient. For the following MOGP runs the maximum Pareto set size is, however, set to 200. Firstly, because of the decrease of the Pareto set minimum of ME(STD) after generation 100. Secondly, to make sure that the run dimensions are sufficient also for other MOGP setups (other predictands, more predictors, larger training data sets).

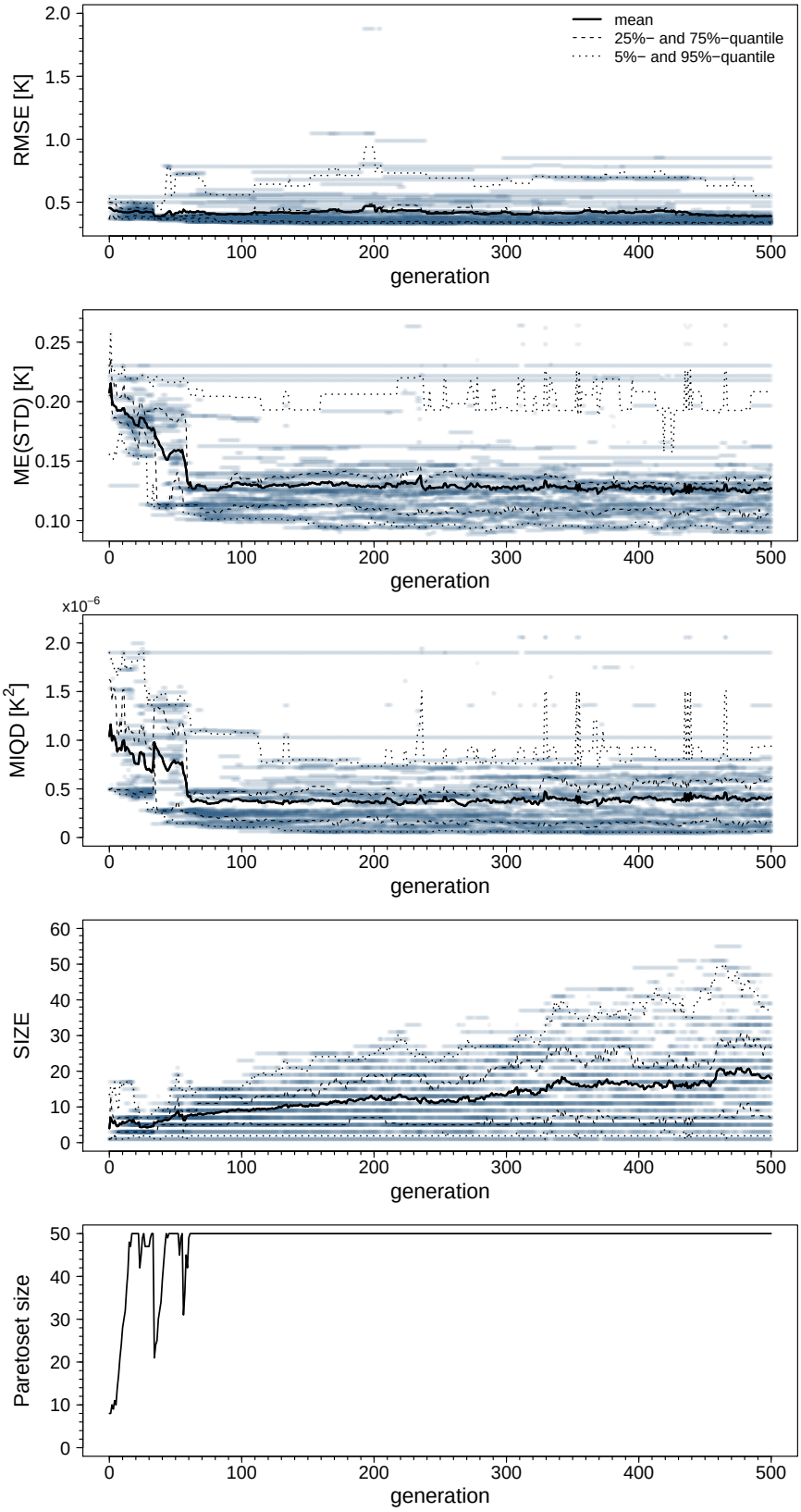

Figure A.1.: Performance of the Pareto sets from test I concerning root mean square error (RMSE), mean error of subgrid-scale standard deviation (ME(STD)) and mean integrated quadratic distance (MIQD) and solution SIZE as well as the Pareto set size during the MOGP run. Each blue point corresponds to one individual from the Pareto set of a generation. The lines indicate the mean over the Pareto set and selected quantiles.

A.2.2. Test II

Figure A.2 shows the difference of the relative reduction of the different objectives between training and validation data set ($\tilde{s}_{tr} - \tilde{s}_{val}$). If a box is located above the zero line, this indicates that the relative reduction is in general larger for the training data than for the validation data, i.e., potentially overfitting occurs. For the majority of cases the median is close to zero. With the exception of May 10^{th} 2008, the medians are spread about equally into positive and negative directions, which indicates that no systematic overfitting takes place. The run excluding and validated on May 10^{th} 2008 sticks out in Fig. A.2. The extraordinary clear sky conditions on this day led to very pronounced fine-scale structures in the near-surface temperature field. Thus, the exclusion from the training data set causes the bad performance. In this case the training data set excluding May 10^{th} is not sufficient. Accordingly the downscaling rules need to extrapolate leading to the bad results for this case. For most of the 8 cases there are a few outliers (≈ 2 to 6 out of 50) for which the performance on the validation data set is clearly worse compared to the training data set. These outliers are most apparent for the RMSE. The corresponding solutions are very small (sometimes consisting of only one node). They are part of the Pareto set due to their good performance concerning the 4th objective, the solution SIZE.

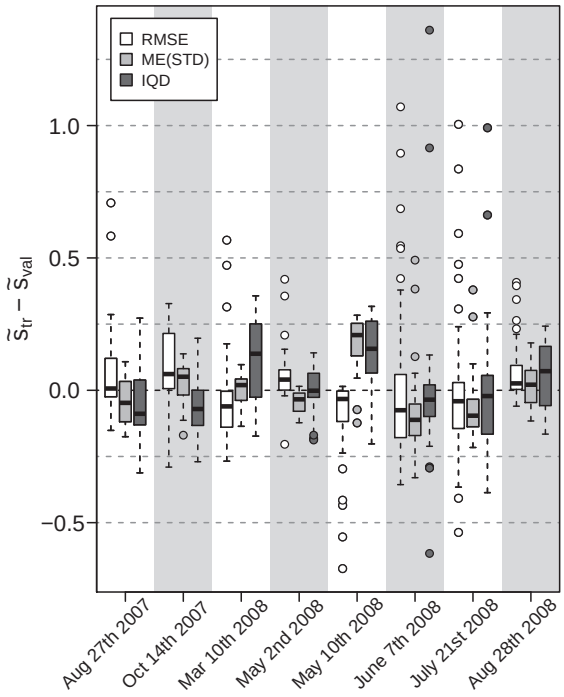

Figure A.2.: Difference in relative reduction of root mean square error (RMSE), mean error of subgrid-scale standard deviation (ME(STD)) and mean integrated quadratic distance (here IQD) between training and validation data set ($\tilde{s}_{tr} - \tilde{s}_{val}$) for all 8 runs. Each box results from 50 values, one for each solution from the Pareto set. The horizontal line within the boxes is the median, the upper and lower boundaries of the boxes correspond to the 75%- and 25%-quantiles. The whiskers indicate the range spanned by maximum and minimum. The length of the whiskers is restricted to 1.5 times the box size. Values outside this range are considered outliers and shown as circles. Figure is taken from *Zerenner et al.* (2016).

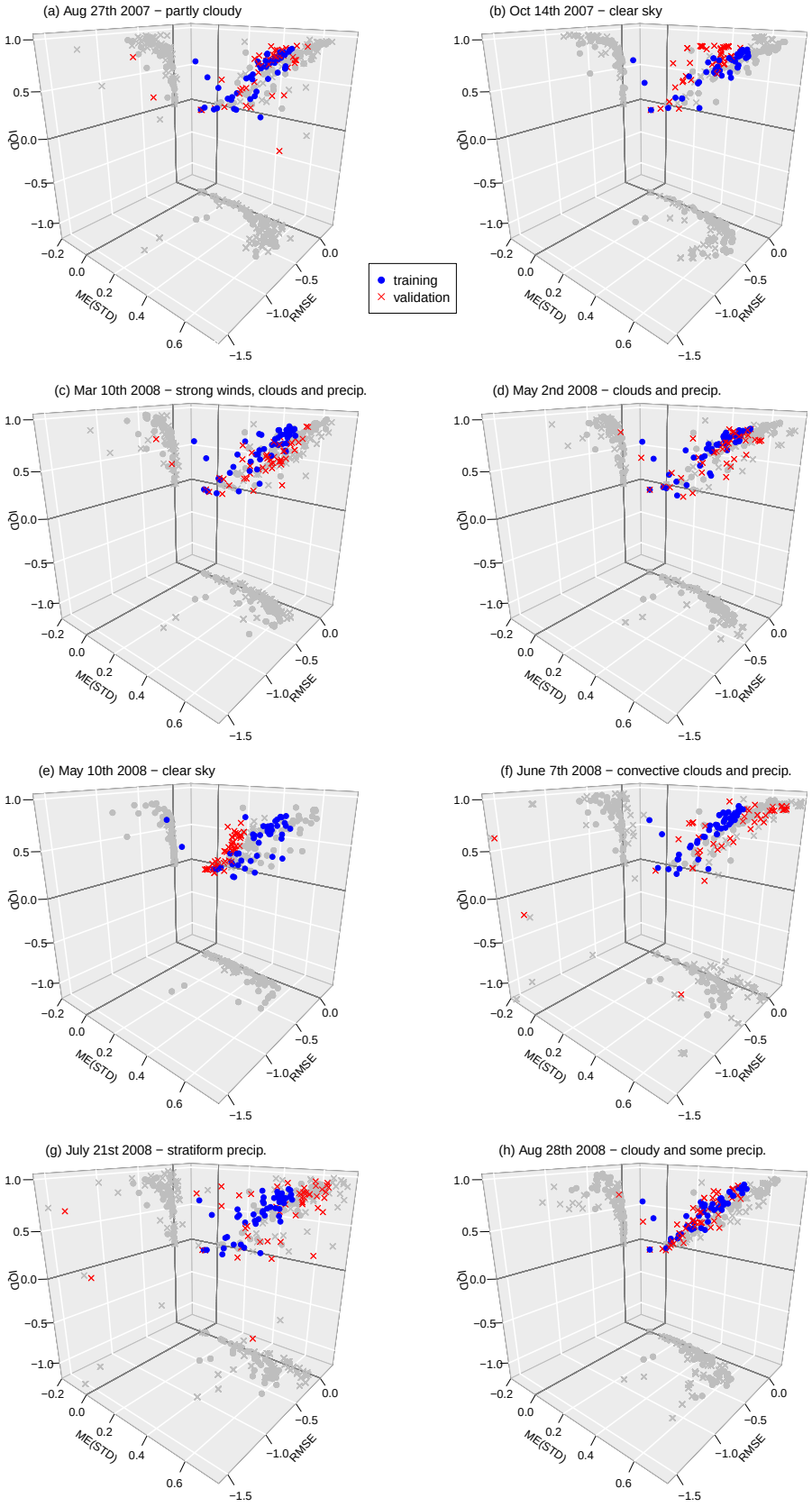

Figure A.3.: Relative reduction of root mean square error (RMSE), mean error of subgrid-scale standard deviation (ME(STD)) and mean integrated quadratic distance (here IQD) for the 8 Pareto sets. The larger the value for the relative reduction, the better the performance concerning the respective objective. The 8 subfigures show the Pareto sets returned by the 8 GP runs (each omitting a different day in the training). The blue circles indicate the performance for the training data; the red crosses indicate the performance for the validation data (i.e., the day omitted in the training): (a) shows the results of the GP run where August 27$^{\text{th}}$ 2007 was omitted for training and so on. Figure is taken from *Zerenner et al.* (2016).

Figure A.3 shows the relative reduction of RMSE, ME(STD), and MIQD for the solutions from the 8 Pareto sets. For the three objectives very different reductions are achieved. Both MIQD and ME(STD) are improved by the GP solutions compared to the spline-interpolation (i.e., positive relative reduction). The relative reduction of the MIQD amounts on average to $50 - 60\%$ with a maximum of about 90%. The relative reduction of the ME(STD) is slightly lower with an average of about to $40 - 50\%$ with a maximum of about 70%. In contrast, the RMSE is on average increased by about 10% compared to the spline-interpolation. At first glance it is disappointing that the RMSE is not decreased by most of the downscaling rules. Also with the nonlinear regression a pixelwise reproduction of the the high-resolution fields appears to be impossible. Nevertheless, it is possible recover of the spatial variability on the subgrid-scale as ME(STD) and MIQD are clearly reduced by almost all of the downscaling rules. Most of outliers discussed above already show a less satisfying performance on the training data sets (e.g., Fig. A.3(a)).

B

Regression Techniques

Regression analysis can be used to model a (univariate) response variable y (also called dependent variable or predictand) given a vector of covariates (also called explanatory variables, dependent variables, predictors). A regression model is typically derived from a set of samples, which consists of n observations of predictand variable y_i and predictors $\boldsymbol{x_i} = (x_{i1}, .., x_{ip})^T$. Many quantities (such as temperature) are quasi-normally distributed and can be modeled sufficiently using a simple linear model (LM) obtained from multiple linear regression (MLR) (Section B.1). Many quantities not following a normal distribution (such as precipitation amounts) can be treated more appropriately by generalized linear models (GLM) (Section B.2). Both LM and GLM are based on the assumption that the y_i are conditionally independent given $\boldsymbol{x_i}$. Further, both approaches require certain residual distributions and, as the name implies, rely to some extent on linear relations between predictors and predictand variable (GLMs allow for tranformations of the linear predictor, see below).

B.1. Multiple Linear Regression

Multiple linear regression models the relationship between the conditional mean of the response variable and two or more explanatory variables by fitting a linear equation to a given set of samples,

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}. \tag{B.1}$$

with $\boldsymbol{\beta}$ denoting the (unknown) regression coefficients and $\boldsymbol{\varepsilon}$ denoting the residuals, which are assumed to be statistically independent and normally distributed with constant variance. The matrix $\mathbf{X}$ is called design matrix. In the most straightforward case the first column of $\mathbf{X}$ is given by the vector $(1, \cdots, 1)^T$ such that β_0 defines the intercept of the regression model and the remaining columns are filled with the vectors of predictors such that each β_j is the regression coefficient for the j-th predictor. A design matrix can also contain quite complex

predictor transformations. (We have not made use of the option. Therefore, we denote both, predictors (or samples thereof) and the entries of the design matrix by the letter x as these are equivalent when defining the design matrix as described.).
Given a set of samples the vector of regression coefficients $\boldsymbol{\beta}$ can be estimated using an (ordinary) least squares estimator, i.e., by minimizing the sum of the squared residuals $\boldsymbol{\varepsilon}$. Expressing the residual sum of squares S as a function of $\boldsymbol{\beta}$ we get

$$S(\boldsymbol{\beta}) = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}). \tag{B.2}$$

The global minimum of $S(\boldsymbol{\beta})$ can be obtained by differentiating with respect to $\boldsymbol{\beta}$

$$\frac{dS}{d\boldsymbol{\beta}^T}(\hat{\boldsymbol{\beta}}) = \frac{d}{d\boldsymbol{\beta}^T}\left(\mathbf{y}^T\mathbf{y} - \boldsymbol{\beta}^T\mathbf{X}^T\mathbf{y} - \mathbf{y}^T\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}^T\mathbf{X}^T\mathbf{X}\boldsymbol{\beta}\right)\Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} = -2\mathbf{X}^T\mathbf{y} + 2\mathbf{X}^T\mathbf{X}\hat{\boldsymbol{\beta}} = 0. \tag{B.3}$$

Solving for $\hat{\boldsymbol{\beta}}$ we get

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}. \tag{B.4}$$

B.2. Generalized Linear Models

Generalized linear models (GLMs) provide a flexible generalization of classical linear regression. This section on GLM relies in large parts on the textbook *Multivariate Statistical Modelling based on Generalized Linear Models* by *Fahrmeir and Tutz* (2001). GLMs have been formulated by *Nelder and Wedderburn* (1972) to establish a unified framework for various statistical models, including linear regression, logistic regression and Poisson regression. The first generalization in GLMs relates to the underlying probability density function. In the GLM framework the residuals do not need to follow a Gaussian distribution, but any probability distribution belonging to the exponential family. For univariate GLMs the probability density functions of the predictands y_i can be written as

$$f(y_i|\theta_i, \phi) = exp\Big\{\frac{y_i\theta_i - b(\theta_i)}{\phi} + c(y_i, \phi)\Big\} \tag{B.5}$$

with θ_i called natural parameter and ϕ called scale or dispersion parameter. The functions b and c are specified by the type of the exponential family. Important members of the exponential family are the normal, binomial, Poisson and gamma distribution. Expectation $\mu_i = \mathrm{E}(y_i)$ and variance $\mathrm{Var}(y_i)$ are related to b as

$$\mathrm{E}(y_i) = \frac{\partial b(\theta_i)}{\partial \theta_i}, \quad \mathrm{Var}(y_i) = \phi\frac{\partial^2 b(\theta_i)}{\partial \theta_i^2}. \tag{B.6}$$

The second generalization of GLMs compared to linear regression is the usage of a so called link function. In a GLM the linear predictor $\eta_i = \boldsymbol{x_i}\boldsymbol{\beta} = \beta_0 + \beta_1 x_{i1} + ... + \beta_p x_{ip}$ is related to the expectation of the response variable μ_i via a link function g, such that

$$\eta_i = g(\mu_i), \quad \mu_i = h(\eta_i) \tag{B.7}$$

with the response function h being the inverse of the link function, i.e., $h = g^{-1}$.
Each probability distribution has a canonical (or natural) link function. The natural link function relates the natural parameter θ_i directly to the linear predictor η_i such that with g being the natural link function $g(\mu_i) = \theta(\mu_i) = \eta_i$. However, for certain applications also non-natural link functions can be appropriate. A specific GLM is hence fully characterized by the type of the exponential family, the response (or link) function and the design matrix.

Normal Distribution For the normal distribution the canonical link function is given by the identity. With

$$\theta(\mu_i) = \mu_i, \quad \phi = \sigma^2, \quad b(\theta_i) = \frac{1}{2}\theta_i^2, \quad c(y_i, \phi) = -\frac{1}{2}\Big(\frac{y_i^2}{\phi} + \ln(2\pi\phi)\Big) \tag{B.8}$$

Eq. B.5 translates into the normal distribution

$$f_{norm}(y_i|\mu_i, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\big(-\frac{1}{2}\Big(\frac{y_i - \mu_i}{\sigma}\Big)^2\Big). \tag{B.9}$$

with expectation μ_i and variance σ^2

$$\mathrm{E}(y_i) = \mu_i, \quad \mathrm{Var}(y_i) = \sigma^2. \tag{B.10}$$

With the natural link function $g(\mu_i) = \mu_i = \eta_i = \boldsymbol{X}\boldsymbol{\beta}$ the GLM transitions into the standard linear model.

Gamma Distribution The gamma distribution is useful for regression of non negative variables such as precipitation amounts. The natural link function of the gamma distribution is given by the reciprocal. With

$$\theta_i(\mu_i) = -\frac{1}{\mu_i}, \quad \phi = \frac{1}{\nu}, \quad b(\theta_i) = -log(-\theta_i), \quad c(y_i, \phi) = -log(\Gamma(\nu)) + (\nu - 1)log(y_i) + \nu log(\nu) \tag{B.11}$$

Eq. B.5 translates into the gamma distribution, i.e.,

$$f_{gamma}(y_i|\nu, \mu_i) = \frac{1}{\Gamma(\nu)}\Big(\frac{\nu}{\mu_i}\Big)^{\nu} y_i^{\nu-1} \exp\Big(-\frac{\nu}{\mu_i}y_i\Big), \quad y_i \geq 0 \tag{B.12}$$

with expectation and variance given by

$$\mathrm{E}(y_i) = -\frac{1}{\theta_i} = \mu_i, \quad \mathrm{Var}(y_i) = \frac{\mu_i^2}{\nu}. \tag{B.13}$$

That is, $\mu_i > 0$ denotes the mean. The parameter $\nu > 0$ is referred to as shape parameter. Further important link functions for gamma distribution based GLMs are the identity and the logarithmic function.

Poisson Distribution The Poisson distribution is used for regression of discrete non-negative quantities. In the following paragraph y denotes a discrete non-negative quantity. The natural

link function of the Poisson distribution is the logarithm. With

$$\theta_i(\mu_i) = log\lambda_i, \quad \phi = 1, \quad b(\theta_i) = \exp(\theta_i), \quad c(y_i, \phi) = -log(y_i!), \tag{B.14}$$

where $\lambda_i > 0$, Eq. B.5 gives the Poisson distribution. That is,

$$f_{poisson}(y_i|\lambda_i) = \frac{\lambda^{y_i} e^{-\lambda_i}}{y_i!} \tag{B.15}$$

with the positive real number λ_i being equal to mean and variance of the respective distribution

$$\mathrm{E}(y_i) = \mathrm{Var}(y_i) = \lambda_i. \tag{B.16}$$

Logistic Regression (Bernoulli Distribution) Logistic regression is an approach for modeling discrete (often *dichotomous*[1]) predictands. In the following we consider logistic regression that assumes a Bernoulli distribution (Eq. B.18). If the predictand series is coded as 0 and 1, the prediction of logistic regression can be interpreted as the probability of the predictand taking the value 1, i.e., $\mu_i = p_i(y_i = 1)$. To make sure the logistic regression predicts only values in the interval (0, 1) the dependent variable is transformed using the natural link function of the Bernoulli distribution given by the logit function. With

$$\theta_i(p_i) = log\left(\frac{p_i}{1-p_i}\right), \quad \phi = 1, \quad b(\theta_i) = \log(1+\exp(\theta_i)), \quad c(y_i, \phi_i) = 0 \tag{B.17}$$

Eq. B.5 translates into the Bernoulli distribution

$$f_{bernoulli}(y_i|p_i) = p_i^{y_i}(1-p_i)^{1-y_i} \tag{B.18}$$

with

$$\mathrm{E}(y_i) = p_i, \quad \mathrm{Var}(y_i) = p_i(1-p_i). \tag{B.19}$$

Inserting into the response function $h = g^{-1}$ gives,

$$h(\eta_i) = \frac{\exp(\eta_i)}{1+\exp(\eta_i)} = \frac{\exp(\beta_0 + \beta_1 x_{i1} + ... + \beta_p x_{ip})}{1+\exp(\beta_0 + \beta_1 x_{i1} + ... + \beta_p x_{ip})} = p(y_i = 1|\boldsymbol{x_i}). \tag{B.20}$$

Thus, the linear predictor which can take values $\eta_i \in (-\infty, \infty)$ translates into $p(y_i = 1) \in (0, 1)$.

Maximum Likelihood Estimation Generalized linear models are typically inferred from data using maximum likelihood estimation. Given a set of n samples (and a design matrix) the maximum likelihood estimator (MLE) of the unknown parameter vector $\boldsymbol{\beta}$ of the model $\mathrm{E}(y_i|\boldsymbol{x_i}) = \mu_i = h(\boldsymbol{x_i}\boldsymbol{\beta})$ is obtained by maximizing its logarithmic likelihood. The contri-

[1] A *dichotomous* partitioning of a set into two groups (A and B) suffices the following two conditions: (1) All elements can be grouped either into A or B; (2) No element can be grouped into both groups. A precipitation series can be transformed to a dichotomous quantity for instance by a transformation into a binary series with 0 denoting no precipitation and 1 denoting precipitation occurrence.

bution of a single observation y_i to the log-likelihood calculates (up to an additive constant) as

$$l_i(\theta_i) = log(f(y_i|\theta_i, \phi)) = \frac{y_i\theta_i - b(\theta_i)}{\phi}. \tag{B.21}$$

After inserting $\theta_i = \theta(\mu_i)$ and $\mu_i = h(\eta_i) = h(\boldsymbol{x_i}\boldsymbol{\beta})$ the log-likelihood of the full sample can be written as

$$l(\boldsymbol{\beta}) = \sum_i l_i(\boldsymbol{\beta}). \tag{B.22}$$

Instead of directly searching for a global maximum of the above function the MLE is typically derived by minimizing the score function $s(\boldsymbol{\beta})$ which is given by the derivative of the likelihood with respect to $\boldsymbol{\beta}$

$$s(\boldsymbol{\beta}) = \frac{\partial l}{\partial \boldsymbol{\beta}}. \tag{B.23}$$

That is, one searches for a parameter vector $\hat{\boldsymbol{\beta}}$ such that $s(\hat{\boldsymbol{\beta}}) = 0$. The actual estimation of $\hat{\boldsymbol{\beta}}$ is typically carried out by an iterative approach, such as the iteratively reweighted least squares method (e.g., *Holland and Welsch*, 1977) which is also the default method for MLE estimation of GLM provided by R (*R Development Core Team*, 2013).

List of Abbreviations

ABL	Atmospheric Boundary Layer
ANN	Artificial Neural Network
a.s.l.	above sea level
CAPE	Convective Available Potential Energy
CCA	Canonical Correlation Analysis
CDF	Cumulative Density Function
CLM	Community Land Model
CO_2	Carbon Dioxide
CORDEX	Coordinated Regional Climate Downscaling Experiment
COSMO	Consortium of Small-Scale Modeling
COSMO-DE	COSMO-model Deutschland (Germany)
COSMO-EU	COSMO-model Europe
COST	European Cooperation in Science and Technology
CRM	Cloud/Convection Resolving Model
DWD	German Meteorological Service (Deutscher Wetterdienst)
EA	Evolutionary Algorithm
ECA&D	European Climate Assessment & Dataset
ECMWF	European Center for Medium-Range Weather Forecast
ENSO	El Niño Southern Oscillation
E-OBS	ENSEMBLES Observations Gridded Dataset
EOF	Empirical Orthogonal Function
ERA	ECMWF Reanalysis
GA	Genetic Algorithm
GCM	General Circulation Model
GEP	Gene Expression Programming
GGLM	Gamma Generalized Linear Model
GLM	Generalized Liner Model
GP	Genetic Programming
ICON	Icosahedral Non-Hydrostatic Model
ICON-EU	ICON Europe

LAI	Leaf Area Index
LAM	Limited Area Model
LES	Large Eddy Simulation
LLM	Litfass Local Model
LM	Linear Model
MLE	Maximum Likelihood Estimator
MOGP	Multi-Objective Genetic Programming
MOS	Model Output Statistics
NAO	North Atlantic Oscillation
NARCCAP	North American Regional Climate Change Assessment Program
NWP	Numerical Weather Prediction
OASIS	Ocean Atmosphere Sea Ice Soil
ParFlow	Parallel Watershed Flow Model
PCA	Principal Component Analysis
PDF	Probability Density Function
PFT	Plant Functional Type
PGLM	Poisson Generalized Linear Model
PP	Perfect Prognosis
RMSE	Root-Mean-Square Error
RCM	Regional Climate Model
SDSM	Statistical Downscaling Model (*Wilby et al.*, 2002)
SPEA	Strength Pareto Evolutionary Algorithm
STD	Standard Deviation
SVAT	Soil-Vegetation-Atmosphere Transfer
TerrSysMP	Terrestrial Systems Modeling Platform
TERENO	Terrestrial Environmental Observatories
TKE	Turbulent Kinetic Energy
TOA	Top of the Atmosphere
UTC	Universal Time Coordinated
VERTEX	Vertical Tile Extension
WG	Weather Generator
WRF	Weather Research and Forecasting (Model)

List of Symbols

In the following the frequently used symbols and variables are listed.

Chapter 5 - Multi-objective Genetic Programming Downscaling Method

α	individual (downscaling rule)
β	individual (downscaling rule)
γ	individual (downscaling rule)
f	fitness of individual (downscaling rule) in $\mathcal{P}$
f'	fitness of individual (downscaling rule) in $\mathcal{P}'$
i	index
j	index
N	number of individuals (downscaling rules) in $\mathcal{P}$
$\mathcal{P}$	population
$\mathcal{P}'$	Pareto set
$\mathbf{s}$	vector of objectives
s_i	objective (function)
s_i^{sc}	scaled objective
m	number of objectives
$\mathcal{O}$	objective space
$\mathcal{Q}$	solution space
$\mathbf{x}$	predictors
$\mathbf{y}^D$	downscaled (D) predictand vector
$\mathbf{y}^R$	reference (R) predictand vector
y_{tij}^D	downscaled (D) predictand anomaly at grid point (i, j) and time t
y_{tij}^R	high-resolution reference (R) predictand anomaly at (i, j) and time t

Chapter 6 - Downscaling Mesoscale Near-Surface Fields using MOGP

A list of all predictors used in Chapter 6 is given in Tables 6.4 and 6.5.

F	cumulative density function
h_i	frequency of predictor x_i in the Pareto set

i	index of fine-scale grid point
IQD	integrated quadratic distance
j	index of fine-scale grid point
LWR	incoming (diffuse) longwave radiation and the land surface
$ME(STD)$	mean error of standard deviation
$\widetilde{ME(STD)}$	relative reduction of the ME(STD)
$MIQD$	mean integrated quadratic distance
$\widetilde{MIQD}$	relative reduction of the MIQD
n_t	number of time steps
n_i	number of (fine-scale) grid points in x-direction
n_j	number of (fine-scale) grid points in y-direction
p	pressure and, as a subscript, index of coarse-scale grid point
P	instantaneous precipitation
q	specific humidity and, as subscript, index of coarse-scale grid point
r^2	coefficient of determination
$RMSE$	root mean square error
$\widetilde{RMSE}$	relative reduction of the RMSE
σ	standard deviation
SWR_{dif}	incoming diffuse shortwave radiation at the land surface
SWR_{dir}	incoming direct shortwave radiation at the land surface
t	time index
T	temperature
w_h	horizontal wind speed
x_i	predictor
$X(p,q)$	coarse scale pixel
y^D_{tij}	downscaled (D) predictand anomaly at grid point (i,j) and time t
y^R_{tij}	high-resolution reference (R) predictand anomaly at (i,j) and time t

Chapter 7 - Downscaling Climate Reanalysis Data to Stations using MOGP

A list of all predictors used in Chapter 7 is given in Table 7.3.

α	downscaling rule
AB	absolute bias
$AE(F)$	absolute error of precipitation frequency
$AE(STD)$	absolute error of standard deviation
β_i	regression coefficient for predictor i
B	bias
$B(SC)$	bias of spatial (i.e., station) correlation
$E(F)$	error of precipitation frequency
$E(STD)$	error of standard deviation
$E(Q_i)$	error of quantile Q_i
G	cumulative density function

h_i	frequency of predictor x_i in the Pareto set
IQD	integrated quadratic distance
$ME(Q)$	mean error of quantiles
n	number of time steps (length of time series)
n_s	number of stations
P	daily accumulated precipitation
$\mathcal{P}'$	Pareto set
Q_i	i%-quantile
ρ	Pearson correlation coefficient
r^2	coefficient of determination
$RMSE$	root mean square error
$RMSE(SC)$	root mean square error of spatial (i.e., station) correlation
σ	standard deviation
t	time index
T_{max}	daily maximum temperature
T_{mean}	daily mean temperature
T_{max}	daily minimum temperature
y_t^D	downscaled (D) predictand anomaly at time t
y_t^R	high-resolution reference (R) predictand anomaly at time t

Appendix A - Preliminary MOGP Runs

A list of all predictors used in Appendix A is given in Table A.2.

i	index of fine-scale grid point
j	index of fine-scale grid point
k	index of fine-scale grid point
l	index of fine-scale grid point
$ME(STD)$	mean error of standard deviation
$MIQD$	mean integrated quadratic distance
n_t	number of time steps
n_i	number of (fine-scale) grid points in x-direction
n_j	number of (fine-scale) grid points in y-direction
$RMSE$	root mean square error
$\tilde{s}_{tr}$	relative reduction objective s for the training data set
$\tilde{s}_{val}$	relative reduction objective s for the validation data set
t	time index
$U(i,j)$	(fine-scale) grid point (i,j) and its four direct neighbors
y_{tij}^D	downscaled (D) predictand anomaly at grid point (i,j) and time t
y_{tij}^R	high-resolution reference (R) predictand anomaly at (i,j) and time t

Appendix B - Regression Techniques

$\boldsymbol{\beta}$	vector of (unknown) regression coefficients
$\hat{\boldsymbol{\beta}}$	estimated vector of regression coefficients
β_j	regression coefficient for the j-th predictor
$b(\theta_i)$	function defining the type of the exponential family
Γ	gamma function
$c(y_i, \phi)$	function defining the type of the exponential family
$\boldsymbol{\varepsilon}$	vector of residuals
$\mathrm{E}(y_i)$	expected value of y_i
f	probability density function
$f_{bernoulli}$	Bernoulli distribution
f_{gamma}	gamma distribution
f_{norm}	normal distribution
$f_{poisson}$	Poisson distribution
$g(\mu_i)$	link function
η_i	linear predictor
$h(\eta_i)$	response function (inverse of the link function g)
λ	mean and variance parameter of the Poisson distribution
l	log-likelihood function
μ_i	expected value (parameter of normal and gamma distribution)
ν	shape parameter of gamma distribution
n	number of observations (realizations)
ϕ	scale or dispersion parameter
p	number of predictors
p_i	probability (parameter of the Bernoulli distribution)
$p(y_i = 1 \vert \boldsymbol{x_i})$	probability that $y_i = 1$ given $\boldsymbol{x_i}$
s	score function
S	residual sum of squares
θ_i	natural parameter
$\mathrm{Var}(y_i)$	variance of y_i
$\mathbf{X}$	design matrix
$\boldsymbol{x_i}$	vector of the i-th realization of the p predictors
x_{ij}	i-th realization of j-th predictor (entries of $\mathbf{X}$)
$\boldsymbol{y}$	vector of predictand realizations
y_i	i-th realization of the predictand

Bibliography

Affenzeller, M., S. Wagner, S. Winkler, and A. Beham (2009), *Genetic algorithms and genetic programming: Modern concepts and practical applications*, Crc Press.

Ament, F., and C. Simmer (2006), Improved representation of land-surface heterogeneity in a non-hydrostatic numerical weather prediction model, *Boundary-Layer Meteorology*, *121*(1), 153–174.

Anyah, R. O., C. P. Weaver, G. Miguez-Macho, Y. Fan, and A. Robock (2008), Incorporating water table dynamics in climate modeling: 3. simulated groundwater influence on coupled land-atmosphere variability, *Journal of Geophysical Research: Atmospheres*, *113*(D7).

Avissar, R. (1992), Conceptual aspects of a statistical-dynamical approach to represent landscape subgrid-scale heterogeneities in atmospheric models, *Journal of Geophysical Research: Atmospheres*, *97*(D3), 2729–2742.

Avissar, R., and R. A. Pielke (1989), A parameterization of heterogeneous land surfaces for atmospheric numerical models and its impact on regional meteorology, *Monthly Weather Review*, *117*(10), 2113–2136.

Baldauf, M., A. Seifert, J. Förstner, D. Majewski, M. Raschendorfer, and T. Reinhardt (2011), Operational convective-scale numerical weather prediction with the COSMO model: Description and sensitivities, *Mothly Weather Review*, *139*(12), 3887–3905.

Banzhaf, W., P. Nordin, R. Keller, and F. Francone (1997), *Genetic programming: An introduction: On the automatic evolution of computer programs and its applications (The Morgan Kaufmann Series in Artificial Intelligence)*, Morgan Kaufmann Publishers, San Francisco, CA, USA.

Betts, A. K., J. H. Ball, A. Beljaars, M. J. Miller, and P. A. Viterbo (1996), The land surface-atmosphere interaction: A review based on observational and global modeling perspectives, *Journal of Geophysical Research: Atmospheres*, *101*(D3), 7209–7225.

Beyrich, F., and H.-T. Mengelkamp (2006), Evaporation over a heterogeneous land surface: EVA_GRIPS and the LITFASS-2003 experiment - an overview, *Boundary-Layer Meteorology*, *121*(1), 5–32.

Beyrich, F., H.-J. Herzog, and J. Neisser (2002), The LITFASS project of DWD and the LITFASS-98 experiment: The project strategy and the experimental setup, *Theoretical and Applied Climatology*, *73*(1-2), 3–18.

Boé, J., L. Terray, F. Habets, and E. Martin (2006), A simple statistical-dynamical downscaling scheme based on weather types and conditional resampling, *Journal of Geophysical Research: Atmospheres*, *111*(D23).

Boé, J., L. Terray, F. Habets, and E. Martin (2007), Statistical and dynamical downscaling of the seine basin climate for hydro-meteorological studies, *International Journal of Climatology*, *27*(12), 1643–1655.

Bürger, G. (1996), Expanded downscaling for generating local weather scenarios, *Climate Research*, *7*(2), 111–128.

Burstedde, C., O. Ghattas, M. Gurnis, T. Isaac, G. Stadler, T. Warburton, and L. Wilcox (2010), Extreme-scale AMR, in *Proceedings of the 2010 ACM/IEEE International Conference for High Performance Computing, Networking, Storage and Analysis*, pp. 1–12, IEEE Computer Society.

Burstedde, C., G. Stadler, L. Alisic, L. C. Wilcox, E. Tan, M. Gurnis, and O. Ghattas (2013), Large-scale adaptive mantle convection simulation, *Geophysical Journal International*, *192*(3), 889–906.

Chandler, R. E., and H. S. Wheater (2002), Analysis of rainfall variability using generalized linear models: A case study from the west of ireland, *Water Resources Research*, *38*(10), 1192.

Charnock, H. (1955), Wind stress on a water surface, *Quarterly Journal of the Royal Meteorological Society*, *81*(350), 639–640.

Chen, S.-H., H.-S. Wang, and B.-T. Zhang (1999), Forecasting high-frequency financial time series with evolutionary neural trees: The case of heng-sheng stock index, in *IC-AI*, pp. 437–443.

Christensen, J. H., and O. B. Christensen (2007), A summary of the PRUDENCE model projections of changes in european climate by the end of this century, *Climatic Change*, *81*(1), 7–30.

Christensen, J. H., F. Boberg, O. B. Christensen, and P. Lucas-Picher (2008), On the need for bias correction of regional climate change projections of temperature and precipitation, *Geophysical Research Letters*, *35*, L20,709.

Coe, R., and R. Stern (1982), Fitting models to daily rainfall data, *Journal of Applied Meteorology*, *21*(7), 1024–1031.

Conway, D., and P. Jones (1998), The use of weather types and air flow indices for gcm downscaling, *Journal of Hydrology*, *212*, 348–361.

Coulibaly, P. (2004), Downscaling daily extreme temperatures with genetic programming, *Geophysical Research Letters*, *31*, L16,203.

Coulibaly, P., Y. B. Dibike, and F. Anctil (2005), Downscaling precipitation and temperature with temporal neural networks, *Journal of Hydrometeorology*, *6*(4), 483–496.

de Vrese, P., and S. Hagemann (2016), Explicit representation of spatial subgrid-scale heterogeneity in an esm, *Journal of Hydrometeorology*, *17*(5), 1357–1371.

de Vrese, P., J.-P. Schulz, and S. Hagemann (2016), On the representation of heterogeneity in land-surface–atmosphere coupling, *Boundary-Layer Meteorology*, *160*, 157––183.

Dee, D., S. Uppala, A. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M. Balmaseda, G. Balsamo, P. Bauer, et al. (2011), The ERA-interim reanalysis: Configuration and performance of the data assimilation system, *Quarterly Journal of the Royal Meteorological Society*, *137*(656), 553–597.

Doms, G., and M. Baldauf (2015), A description of the nonhydrostatic regional COSMO-model - part I: Dynamics and numerics, *Tech. rep.*, Deutscher Wetterdienst.

Doms, G., J. Förstner, E. Heise, H.-J. Herzog, D. Mironov, M. Raschendorfer, T. Reinhardt, B. Ritter, R. Schrodin, J.-P. Schulz, and G. Vogel (2011), A description of the nonhydrostatic regional COSMO-model - part II : Physical parameterization, *Tech. rep.*, Deutscher Wetterdienst.

Dosio, A., and P. Paruolo (2011), Bias correction of the ENSEMBLES high-resolution climate change projections for use by impact models: Evaluation on the present climate, *Journal of Geophysical Research: Atmospheres*, *116*(D16).

Enke, W., and A. Spekat (1997), Downscaling climate model outputs into local and regional weather elements by classification and regression, *Climate Research*, *8*(3), 195–207.

Evett, M., and T. Fernandez (1998), Numeric mutation improves the discovery of numeric constants in genetic programming, pp. 66–71, Morgan Kaufmann.

Fahrmeir, L., and G. Tutz (2001), *Multivariate Statistical Modelling Based on Generalized Linear Models*, Springer.

Farr, T. G., P. A. Rosen, E. Caro, R. Crippen, R. Duren, S. Hensley, M. Kobrick, M. Paller, E. Rodriguez, L. Roth, et al. (2007), The shuttle radar topography mission, *Reviews of Geophysics*, *45*, RG2004.

Fealy, R., and J. Sweeney (2007), Statistical downscaling of precipitation for a selection of sites in ireland employing a generalised linear modelling approach, *International Journal of Climatology*, *27*(15), 2083–2094.

Fiddes, J., and S. Gruber (2014), TopoSCALE v. 1.0: Downscaling gridded climate data in complex terrain, *Geoscientific Model Development*, *7*(1), 387–405.

Fowler, H. J., S. Blenkinsop, and C. Tebaldi (2007), Linking climate change modelling to impacts studies: Recent advances in downscaling techniques for hydrological modelling, *International Journal of Climatology*, *27*(12), 1547–1578.

Fukunaga, A., and A. Stechert (1998), Evolving nonlinear predictive models for lossless image compression with genetic programming, pp. 95–102.

Giorgi, F. (1997a), An approach for the representation of surface heterogeneity in land surface models. part ii: Validation and sensitivity experiments, *Monthly Weather Review*, *125*(8), 1900–1919.

Giorgi, F. (1997b), An approach for the representation of surface heterogeneity in land surface models. part i: Theoretical framework, *Monthly Weather Review*, *125*(8), 1885–1899.

Giorgi, F., and R. Avissar (1997), Representation of heterogeneity effects in earth system modeling: Experience from land surface modeling, *Reviews of Geophysics*, *35*(4), 413–437.

Giorgi, F., and M. Marinucci (1996), Improvements in the simulation of surface climatology over the european region with a nested modeling system, *Geophysical Research Letters*, *23*(3), 273–276.

Giorgi, F., R. Francisco, and J. Pal (2003), Effects of a subgrid-scale topography and land use scheme on the simulation of surface climate and hydrology. part i: Effects of temperature and water vapor disaggregation, *Journal of Hydrometeorology*, *4*(2), 317–333.

Giorgi, F., C. Jones, G. R. Asrar, et al. (2009), Addressing climate information needs at the regional level: The CORDEX framework, *World Meteorological Organization (WMO) Bulletin*, *58*(3), 175–183.

Goodess, C., C. Anagnostopoulou, A. Bárdossy, C. Frei, C. Harpham, M. Haylock, Y. Hundecha, P. Maheras, J. Ribalaygua, J. Schmidli, et al. (2007), An intercomparison of statistical downscaling methods for europe and european regions - assessing their performance with respect to extreme temperature and precipitation events, *Climatic Research Unit Research Publication (CRURP), University of East Anglia, UK*, (11).

Goodess, C. M., and J. P. Palutikof (1998), Development of daily rainfall scenarios for southeast spain using a circulation-type approach to downscaling, *International Journal of Climatology*, *18*(10), 1051–1083.

Gutiérrez, J. M., D. San-Martín, S. Brands, R. Manzanas, and S. Herrera (2013), Reassessing statistical downscaling techniques for their robust application under climate change conditions, *Journal of Climate*, *26*(1), 171–188.

Gutmann, E. D., R. M. Rasmussen, C. Liu, K. Ikeda, D. J. Gochis, M. P. Clark, J. Dudhia, and G. Thompson (2012), A comparison of statistical and dynamical downscaling of winter precipitation over complex terrain, *Journal of Climate*, *25*(1), 262–281.

Hashmi, M. Z., A. Y. Shamseldin, and B. W. Melville (2011), Statistical downscaling of watershed precipitation using gene expression programming (gep), *Environmental Modelling and Software*, *26*(12), 1639–1646.

Hassanzadeh, E., A. Nazemi, and A. Elshorbagy (2013), Quantile-based downscaling of precipitation using genetic programming: Application to idf curves in saskatoon, *Journal of Hydrologic Engineering*, *19*(5), 943–955.

Hastie, T., R. Tibshirani, and J. Friedman (2009), *The elements of statistical learning*, vol. 2, Springer.

Haylock, M., N. Hofstra, A. Klein Tank, E. Klok, P. D. Jones, and M. New (2008), A european daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006, *Journal of Geophysical Research: Atmospheres*, *113*(D20).

Heinemann, G., and M. Kerschgens (2005), Comparison of methods for area-averaging surface energy fluxes over heterogeneous land surfaces using high-resolution non-hydrostatic simulations, *International Journal of Climatology*, *25*(3), 379–403.

Hewitson, B. C., and R. G. Crane (1992), Large-scale atmospheric controls on local precipitation in tropical mexico, *Geophysical Research Letters*, *19*(18), 1835–1838.

Holland, P. W., and R. E. Welsch (1977), Robust regression using iteratively reweighted least-squares, *Communications in Statistics-Theory and Methods*, *6*(9), 813–827.

Hong, J.-H., and S.-B. Cho (2006), The classification of cancer based on dna microarray data that uses diverse ensemble genetic programming, *Artificial Intelligence in Medicine*, *36*(1), 43–58.

Huth, R. (2002), Statistical downscaling of daily temperature in central europe, *Journal of Climate*, *15*(13), 1731–1742.

Kalnay, E., M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, et al. (1996), The NCEP/NCAR 40-year reanalysis project, *Bulletin of the American Meteorological Society*, *77*(3), 437–471.

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, *Journal of Climate*, *3*(10), 1053–1079.

Kattan, A., and R. Poli (2008), Evolutionary lossless compression with gp-zip, in *Proceedings of the 10th annual conference on Genetic and evolutionary computation*, pp. 1211–1218, ACM.

Katz, R. W. (1977), Precipitation as a chain-dependent process, *Journal of Applied Meteorology*, *16*(7), 671–676.

Kessler, E. (1969), On the distribution and continuity of water substance in atmospheric circulation, *Meteorological Monographs*, *10*(32), 88.

Kidson, J. W., and C. S. Thompson (1998), A comparison of statistical and model-based downscaling techniques for estimating local climate variations, *Journal of Climate*, *11*(4), 735–753.

Kilsby, C., P. Jones, A. Burton, A. Ford, H. Fowler, C. Harpham, P. James, A. Smith, and R. Wilby (2007), A daily weather generator for use in climate change studies, *Environmental Modelling and Software*, *22*(12), 1705–1719.

Kim, S., and H. S. Kim (2008), Neural networks and genetic algorithm approach for nonlinear evaporation and evapotranspiration modeling, *Journal of Hydrology*, *351*(3), 299–317.

Klein, W. H., and H. R. Glahn (1974), Forecasting local weather by means of model output statistics, *Bulletin of the American Meteorological Society*, *55*(10), 1217–1227.

Klein Tank, A., et al. (2002), Daily dataset of the 20th-century surface air temperature and precipitation for the european climate assessment, *Internationl Journal of Climatology*, *22*(12), 1441–1453.

Kollet, S. J., and R. M. Maxwell (2006), Integrated surface - groundwater flow modeling: A free-surface overland flow boundary condition in a parallel groundwater flow model, *Advances in Water Resources*, *29*(7), 945–958.

Kollet, S. J., and R. M. Maxwell (2008), Capturing the influence of groundwater dynamics on land surface processes using an integrated, distributed watershed model, *Water Resources Research*, *44*(2).

Köppen, W. P. (1884), Die Wärmezonen der Erde, nach der Dauer der heissen, gemässigten und kalten Zeit und nach der Wirkung der Wärme auf die organische Welt betrachtet, *Meteorologische Zeitschrift*, pp. 215–226.

Köppen, W. P. (1918), *Klassifikation der Klimate nach Temperatur, Niederschlag und Jahreslauf*, Perthes.

Koster, R. D., and M. J. Suarez (1992), Modeling the land surface boundary in climate models as a composite of independent vegetation stands, *Journal of Geophysical Research: Atmospheres*, *97*(D3), 2697–2715.

Koza, J. (1992), *Genetic Programming, On the Programming of Computers by Means of Natural Selection*, MIT Press, Cambridge, MA, USA.

Laing, A., and J. Evans (2011), Introduction to tropical meteorology, http://kejian1.cma.gov.cn/vod/comet/tropical/textbook_2nd_edition/print_versions.htm.

Lenderink, G., A. Buishand, and W. v. Deursen (2007), Estimates of future discharges of the river rhine using two scenario methodologies: Direct versus delta approach, *Hydrology and Earth System Sciences*, *11*(3), 1145–1159.

Leung, L. R., and S. Ghan (1995), A subgrid parameterization of orographic precipitation, *Theoretical and Applied Climatology*, *52*(1-2), 95–118.

Liston, G. E., and M. Sturm (1998), A snow-transport model for complex terrain, *Journal of Glaciology*, *44*(148), 498–516.

Liu, X., P. Coulibaly, and N. Evora (2008), Comparison of data-driven methods for downscaling ensemble weather forecasts, *Hydrology and Earth System Sciences Discussions*, *12*(2), 615–624.

Lorenz, E. N. (1969), Atmospheric predictability as revealed by naturally occurring analogues, *Journal of the Atmospheric Sciences*, *26*(4), 636–646.

Luke, S., and L. Panait (2006), A comparison of bloat control methods for genetic programming, *Evolutionary Computation*, *14*(3), 309–344.

Luke, S., L. Panait, et al. (2002), Lexicographic parsimony pressure., in *GECCO*, vol. 2, pp. 829–836.

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

Maraun, D., F. Wetterhall, A. M. Ireson, R. E. Chandler, E. J. Kendon, M. Widmann, S. Brienen, H. W. Rust, T. Sauter, M. Themeßl, et al. (2010), Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user, *Reviews of Geophysics*, *48*(3).

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

Maurer, E. P., and H. G. Hidalgo (2008), Utility of daily vs. monthly large-scale climate data: An intercomparison of two statistical downscaling methods, *Hydrology and Earth System Sciences*, *12*(2), 551–563.

Maxwell, R. M. (2013), A terrain-following grid transform and preconditioner for parallel, large-scale, integrated hydrologic modeling, *Advances in Water Resources*, *53*, 109–117.

Maxwell, R. M., and S. J. Kollet (2008), Interdependence of groundwater dynamics and land-energy feedbacks under climate change, *Nature Geoscience*, *1*(10), 665–669.

Maxwell, R. M., F. K. Chow, and S. J. Kollet (2007), The groundwater–land-surface–atmosphere connection: Soil moisture effects on the atmospheric boundary layer in fully-coupled simulations, *Advances in Water Resources*, *30*(12), 2447–2466.

Mearns, L., I. Bogardi, F. Giorgi, I. Matyasovszky, and M. Palecki (1999), Comparison of climate change scenarios generated from regional climate model experiments and statistical downscaling, *Journal of Geophysical Research: Atmospheres*, *104*(D6), 6603–6621.

Mearns, L., S. Sain, L. Leung, M. Bukovsky, S. McGinnis, S. Biner, D. Caya, R. Arritt, W. Gutowski, E. Takle, et al. (2013), Climate change projections of the north american regional climate change assessment program (NARCCAP), *Climatic Change*, *120*(4), 965–975.

Mengelkamp, H.-T., F. Beyrich, G. Heinemann, F. Ament, et al. (2006), Evaporation over a heterogeneous land surface: The EVA-GRIPS project, *Bulletin of the American Meteorological Society*, *87*(6), 775–786.

Miguez-Macho, G., G. L. Stenchikov, and A. Robock (2004), Spectral nudging to eliminate the effects of domain position and geometry in regional climate model simulations, *Journal of Geophysical Research: Atmospheres*, *109*(D13).

Miguez-Macho, G., Y. Fan, C. P. Weaver, R. Walko, and A. Robock (2007), Incorporating water table dynamics in climate modeling: 2. formulation, validation, and soil moisture simulation, *Journal of Geophysical Research: Atmospheres*, *112*(D13).

Mitchell, M. (1998), *An introduction to genetic algorithms*, MIT press.

Molod, A., H. Salmun, and D. W. Waugh (2003), A new look at modeling surface heterogeneity: Extending its influence in the vertical, *Journal of Hydrometeorology*, *4*(5), 810–825.

Molod, A., H. Salmun, and D. W. Waugh (2004), The impact on a gcm climate of an extended mosaic technique for the land-atmosphere coupling, *Journal of Climate*, *17*(20), 3877–3891.

Murphy, J. (1999), An evaluation of statistical and dynamical techniques for downscaling local climate, *Journal of Climate*, *12*(8), 2256–2284.

Nelder, J., and R. Wedderburn (1972), Generalized linear models, *Journal of the Royal Statistical Society. Series A (General)*, *135*(3), 370–384.

Nordin, P., and W. Banzhaf (1996), Programmatic compression of images and sound, in *Proceedings of the 1st annual conference on genetic programming*, pp. 345–350, MIT Press.

Oleson, K., G.-Y. Niu, Z.-L. Yang, D. Lawrence, P. Thornton, P. Lawrence, R. Stöckli, R. Dickinson, G. Bonan, S. Levis, et al. (2008), Improvements to the community land model and their impact on the hydrological cycle, *Journal of Geophysical Research: Biogeosciences*, *113*, G01,021.

Oleson, K. W., Y. Dai, G. Bonan, M. Bosilovich, R. Dickinson, P. Dirmeyer, F. Hoffman, P. Houser, S. Levis, G.-Y. Niu, et al. (2004), Technical description of the community land model (CLM), *Tech. rep.*, NCAR Technical Note, National Center for Atmospheric Research, Boulder, CO, USA.

Parasuraman, K., A. Elshorbagy, and S. K. Carey (2007), Modelling the dynamics of the evapotranspiration process using genetic programming, *Hydrological Science Journal*, *52*(3), 563–578.

Pitman, A. J., Z.-L. Yang, and A. Henderson-Sellers (1993), Sub-grid scale precipitation in ALCMs: Re-assessing the land surface sensitivity using a single column model, *Climate Dynamics*, *9*(1), 33–41.

Poli, R., W. B. Langdon, N. F. McPhee, and J. R. Koza (2008), A field guide to genetic programming, http://www.gp-field-guide.org.uk.

Pour, S. H., S. B. Harun, and S. Shahid (2014), Genetic programming for the downscaling of extreme rainfall events on the east coast of peninsular malaysia, *Atmosphere*, *5*(4), 914–936.

Prein, A. F., W. Langhans, G. Fosser, A. Ferrone, N. Ban, K. Goergen, M. Keller, M. Tölle, O. Gutjahr, F. Feser, et al. (2015), A review on regional convection-permitting climate modeling: Demonstrations, prospects, and challenges, *Reviews of Geophysics*, *53*(2), 323–361.

R Development Core Team (2013), R: A language and environment for statistical computing, http://www.R-project.org.

Rahman, M., M. Sulis, and S. Kollet (2015), The subsurface–land surface–atmosphere connection under convective conditions, *Advances in Water Resources*, *83*, 240–249.

Richardson, C. W., and D. A. Wright (1984), WGEN: a model for generating daily weather variables.

Rihani, J. F., R. M. Maxwell, and F. K. Chow (2010), Coupling groundwater and land surface processes: Idealized simulations to identify effects of terrain and subsurface heterogeneity on land surface energy fluxes, *Water Resources Research*, *46*, W12,523.

Ritter, B., and J.-F. Geleyn (1992), A comprehensive radiation scheme for numerical weather prediction models with potential applications in climate simulations, *Monthly Weather Review*, *120*(2), 303–325.

Schaettler, U., G. Doms, and C. Schraff (2015), A description of the nonhydrostatic regional COSMO-model - part VII : User's guide, *Tech. rep.*, Deutscher Wetterdienst.

Schlünzen, K. H., and J. J. Katzfey (2003), Relevance of sub-grid-scale land-use effects for mesoscale models, *Tellus A*, *55*(3), 232–246.

Schmidli, J., C. Frei, and P. L. Vidale (2006), Downscaling from gcm precipitation: A benchmark for dynamical and statistical downscaling methods, *International Journal of Climatology*, *26*(5), 679–689.

Schmidli, J., C. Goodess, C. Frei, M. Haylock, Y. Hundecha, J. Ribalaygua, and T. Schmith (2007), Statistical and dynamical downscaling of precipitation: An evaluation and comparison of scenarios for the european alps, *Journal of Geophysical Research: Atmospheres*, *112*(D4).

Schomburg, A. (2011), Improving the simulation of small-scale variability in radiation and land-surface parameterizations in a mesoscale numerical weather prediction model, Ph.D. thesis, Universitäts-und Landesbibliothek Bonn.

Schomburg, A., V. Venema, R. Lindau, F. Ament, and C. Simmer (2010), A downscaling scheme for atmospheric variables to drive soil–vegetation–atmosphere transfer models, *Tellus B*, *62*(4), 242–258.

Schomburg, A., V. Venema, F. Ament, and C. Simmer (2012), Disaggregation of screen-level variables in a numerical weather prediction model with an explicit simulation of subgrid-scale land-surface heterogeneity, *Meteorology and Atmospheric Physics*, *116*(3-4), 81–94.

Schoof, J. T., and S. Pryor (2001), Downscaling temperature and precipitation: A comparison of regression-based methods and artificial neural networks, *International Journal of Climatology*, *21*(7), 773–790.

Segond, M.-L., H. S. Wheater, and C. Onof (2007), The significance of spatial rainfall representation for flood runoff estimation: A numerical evaluation based on the lee catchment, uk, *Journal of Hydrology*, *347*(1), 116–131.

Seth, A., F. Giorgi, and R. E. Dickinson (1994), Simulating fluxes from heterogeneous land surfaces: Explicit subgrid method employing the biosphere-atmosphere transfer scheme (BATS), *Journal of Geophysical Research: Atmospheres.*, *99*(18), 651–18.

Shao, Y., M. Sogalla, M. Kerschgens, and W. Brücher (2001), Effects of land-surface heterogeneity upon surface fluxes and turbulent conditions, *Meteorology and Atmospheric Physics*, *78*(3-4), 157–181.

Shrestha, P., M. Sulis, M. Masbou, S. Kollet, and C. Simmer (2014), A scale-consistent terrestrial systems modeling platform based on COSMO, CLM and ParFlow, *Monthly Weather Review*, *142*(9), 3466–3483.

Silva, S., and J. Almeida (2003), GPLAB-a genetic programming toolbox for MATLAB, in *Proceedings of the Nordic MATLAB conference*, pp. 273–278, Citeseer.

Simmer, C., M. Masbou, I. Thiele-Eich, W. Amelung, H. Bogena, S. Crewell, B. Diekkrüger, F. Ewert, H.-J. H. Franssen, J. A. Huisman, et al. (2015), Monitoring and modeling the terrestrial system from pores to catchments-the transregional collaborative research center on patterns in the soil-vegetation-atmosphere system, *Bulletin of the American Meteorogicyl Society*, *96*(10), 1765–1787.

Simon, T., A. Hense, B. Su, T. Jiang, C. Simmer, and C. Ohlwein (2013), Pattern-based statistical downscaling of east asian summer monsoon precipitation, *Tellus A*, *65*, 19,749.

Singh, V. (1997), Effect of spatial and temporal variability in rainfall and watershed characteristics on stream flow hydrograph, *Hydrological Processes*, *11*(12), 1649–1669.

Skamarock, W. C. (2004), Evaluating mesoscale NWP models using kinetic energy spectra, *Monthly Weather Review*, *132*(12), 3019–3032.

Smart, W., and M. Zhang (2003), Classification strategies for image classification in genetic programming, in *Proceeding of image and vision computing conference*, pp. 402–407, Palmerston North, New Zealand.

Stern, R., and R. Coe (1984), A model fitting analysis of daily rainfall data, *Journal of the Royal Statistical Society. Series A (General)*, *147*(1), 1–34.

Stull, R. B. (2012), *An introduction to boundary layer meteorology*, vol. 13, Springer.

Tackett, W. A. (1993), Genetic programming for feature discovery and image discrimination, in *Proceedings of the Fifth International Conference on Genetic Algorithms*, pp. 303–311, Morgan Kaufman.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl (2012), An overview of CMIP5 and the experiment design, *Bulletin of the American Meteorological Society*, *93*(4), 485–498.

Themeßl, M. J., A. Gobiet, and A. Leuprecht (2011), Empirical-statistical downscaling and error correction of daily precipitation from regional climate models, *International Journal of Climatology*, *31*(10), 1530–1544.

Thorarinsdottir, T. L., T. Gneiting, and N. Gissibl (2013), Using proper divergence functions to evaluate climate models, *Journal of Uncertainty Quantification*, *1*(1), 522–534.

Tiedtke, M. (1989), A comprehensive mass flux scheme for cumulus parameterization in large-scale models, *Monthly Weather Review*, *117*(8), 1779–1800.

Torres, R. d. S., A. X. Falcão, M. A. Gonçalves, J. P. Papa, B. Zhang, W. Fan, and E. A. Fox (2009), A genetic programming framework for content-based image retrieval, *Pattern Recognition*, *42*(2), 283–292.

Valcke, S. (2013), The OASIS3 coupler: A european climate modelling community software, *Geoscientific Model Development*, *6*(2), 373–388.

Van der Linden, P., and e. Mitchell, JFB (2009), ENSEMBLES: Climate change and its impacts: Summary of research and results from the ENSEMBLES project.

Vereecken, H., S. Kollet, and C. Simmer (2010), Patterns in soil-vegetation-atmosphere systems: Monitoring, modeling, and data assimilation, *Vadose Zone Journal*, *9*(4), 821–827.

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

von Storch, H., B. Hewitson, and L. Mearns (2000), Review of empirical downscaling techniques, *Tech. rep.*, Regional climate development under global warming.

Vrac, M., M. Stein, and K. Hayhoe (2007), Statistical downscaling of precipitation through nonhomogeneous stochastic weather typing, *Climate Research*, *34*(3), 169–184.

Walko, R. L., L. E. Band, J. Baron, T. G. Kittel, R. Lammers, T. J. Lee, D. Ojima, R. A. Pielke Sr, C. Taylor, C. Tague, et al. (2000), Coupled atmosphere-biophysics-hydrology models for environmental modeling, *Journal of Applied Meteorology*, *39*(6), 931–944.

Wang, Q. (1991), The genetic algorithm and its application to calibrating conceptual rainfall-runoff models, *Water Resources Research*, *27*(9), 2467–2471.

Widmann, M., C. S. Bretherton, and E. P. Salathé Jr (2003), Statistical precipitation downscaling over the northwestern united states using numerically simulated precipitation as a predictor, *Journal of Climate*, *16*(5), 799–816.

Wilby, R. L., and T. M. L. Wigley (1997), Downscaling general circulation model output: A review of methods and limitations, *Progress in Physical Geography*, *21*(4), 530–548.

Wilby, R. L., T. Wigley, D. Conway, P. Jones, B. Hewitson, J. Main, and D. Wilks (1998), Statistical downscaling of general circulation model output: A comparison of methods, *Water Resources Research*, *34*(11), 2995–3008.

Wilby, R. L., C. W. Dawson, and E. M. Barrow (2002), SDSM - a decision support tool for the assessment of regional climate change impacts, *Environmental Modeling and Software*, *17*(2), 145–157.

Wilks, D. S. (1999), Multisite downscaling of daily precipitation with a stochastic weather generator, *Climate Research*, *11*(2), 125–136.

Wilks, D. S. (2011), *Statistical methods in the atmospheric sciences*, vol. 100, Academic press.

Wilks, D. S., and R. L. Wilby (1999), The weather generation game: A review of stochastic weather models, *Progress in Physical Geography*, *23*(3), 329–357.

Xu, C.-Y. (1999), From GCMs to river flow: A review of downscaling methods and hydrologic modelling approaches, *Progress in Physical Geography*, *23*(2), 229–249.

Yu, J., J. Yu, A. A. Almal, S. M. Dhanasekaran, D. Ghosh, W. P. Worzel, and A. M. Chinnaiyan (2007), Feature selection and molecular classification of cancer using genetic programming, *Neoplasia*, *9*(4), 292–303.

Yu, T., S.-H. Chen, and T.-W. Kuo (2005), Discovering financial technical trading rules using genetic programming with lambda abstraction, in *Genetic programming theory and practice II*, pp. 11–30, Springer.

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

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

Zorita, E., and H. von Storch (1999), The analog method as a simple statistical downscaling technique: Comparison with more complicated methods, *Journal of Climate*, *12*(8), 2474–2489.

Zorita, E., J. P. Hughes, D. P. Lettemaier, and H. von Storch (1995), Stochastic characterization of regional circulation patterns for climate model diagnosis and estimation of local precipitation, *Journal of Climate*, *8*(5), 1023–1042.

BONNER METEOROLOGISCHE ABHANDLUNGEN

Herausgegeben vom Meteorologischen Institut der Universität Bonn durch Prof. Dr. H. FLOHN (Hefte 1-25), Prof. Dr. M. HANTEL (Hefte 26-35), Prof. Dr. H.-D. SCHILLING (Hefte 36-39), Prof. Dr. H. KRAUS (Hefte 40-49), ab Heft 50 durch Prof. Dr. A. HENSE.

Heft 1-63: siehe `http://www.meteo.uni-bonn.de/bibliothek/bma`

64-77: open access, verfügbar unter `https://uni-bn.de/kpSDaQffe1`

Heft 64: ***Michael Weniger***: Stochastic parameterization: a rigorous approach to stochastic three-dimensional primitive equations, 2014, 148 S. + XV.

Heft 65: ***Andreas Röpnack***: Bayesian model verification: predictability of convective conditions based on EPS forecasts and observations, 2014, 152 S. + VI.

Heft 66: ***Thorsten Simon***: Statistical and Dynamical Downscaling of Numerical Climate Simulations: Enhancement and Evaluation for East Asia, 2014, 48 S. + VII. + Anhänge

Heft 67: ***Elham Rahmani***: The Effect of Climate Change on Wheat in Iran, 2014, [erschienen] 2015, 96 S. + XIII.

Heft 68: ***Pablo A. Saavedra Garfias***: Retrieval of Cloud and Rainwater from Ground-Based Passive Microwave Observations with the Multi-frequency Dual-polarized Radiometer ADMIRARI, 2014, [erschienen] 2015, 168 S. + XIII.

Heft 69: ***Christoph Bollmeyer***: A high-resolution regional reanalysis for Europe and Germany - Creation and Verification with a special focus on the moisture budget, 2015, 103 S. + IX.

Heft 70: ***A S M Mostaquimur Rahman***: Influence of subsurface hydrodynamics on the lower atmosphere at the catchment scale, 2015, 98 S. + XVI.

Heft 71: ***Sabrina Wahl***: Uncertainty in mesoscale numerical weather prediction: probabilistic forecasting of precipitation, 2015, 108 S.

Heft 72: ***Markus Übel***: Simulation of mesoscale patterns and diurnal variations of atmospheric CO_2 mixing ratios with the model system TerrSysMP-CO_2, 2015, [erschienen] 2016, 158 S. + II

Heft 73: ***Christian Bernardus Maria Weijenborg***: Characteristics of Potential Vorticity anomalies associated with mesoscale extremes in the extratropical troposphere, 2015, [erschienen] 2016, 151 S. + XI

Heft 74: ***Muhammad Kaleem***: A sensitivity study of decadal climate prediction to aerosol variability using ECHAM6-HAM (GCM), 2016, 98 S. + XII

Heft 75: ***Theresa Bick***: 3D Radar reflectivity assimilation with an ensemble Kalman filter on the convective scale, 2016, [erschienen] 2017, 96 S. + IX

Heft 76: ***Zied Ben Bouallegue***: Verification and post-processing of ensemble weather forecasts for renewable energy applications, 2017, 119 S.

Heft 77: ***Julia Lutz***: Improvements and application of the STatistical Analogue Resampling Scheme STARS, 2016, [erschienen] 2017, 103 S.

Heft 78: ***Benno Michael Thoma***: Palaeoclimate Reconstruction in the Levant and on the Balkans, 2016, [erschienen] 2017, XVI, 266 S.

Heft 79: ***Ieda Pscheidt***: Generating high resolution precipitation conditional on rainfall observations and satellite data, 2017, V, 173 S.

Heft 80: ***Tanja Zerenner***: Atmospheric downscaling using multi-objective genetic programming, 2016, [erschienen] 2017, X, 191 S.

METEOROLOGISCHES INSTITUT
MATHEMATISCH NATURWISSENSCHAFTLICHE FAKULTÄT
UNIVERSITÄT BONN

Meteorologisches Institut
Universität Bonn