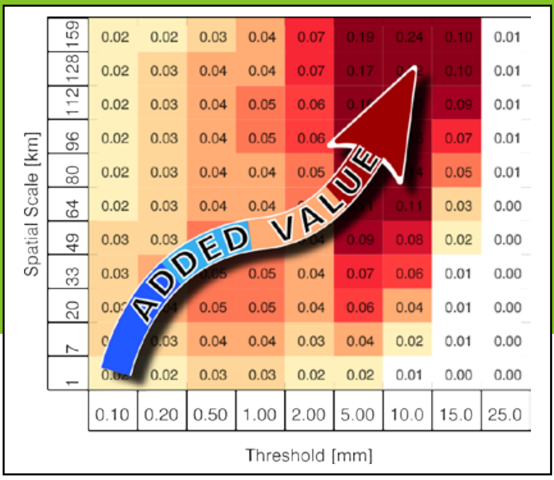
Wegener Center for Climate and Global Change
University of Graz



Scientific Report No. 39-2011

NHCM-1: Non-hydrostatic climate modelling

Part I

Defining and Detecting Added Value in Cloud-Resolving Climate Simulations

Andreas F. Prein
Andreas Gobiet

February 2011

Wegener Center
www.wegcenter.at

UNI GRAZ

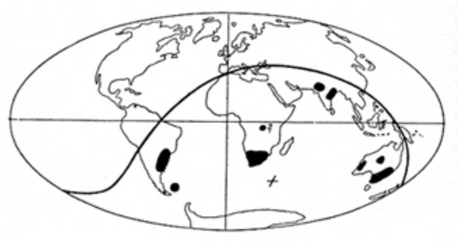
kindly supported by:

FWF

Der Wissenschaftsfonds.

The **Wegener Center for Climate and Global Change** combines as an interdisciplinary, internationally oriented research center the competences of the University of Graz in the research area „Climate, Environmental and Global Change". It brings together, in a dedicated building close to the University central campus, research teams and scientists from fields such as geo- and climate physics, meteorology, economics, geography, and regional sciences. At the same time close links exist and are further developed with many cooperation partners, both nationally and internationally. The research interests extend from monitoring, analysis, modeling and prediction of climate and environmental change via climate impact research to the analysis of the human dimensions of these changes, i.e., the role of humans in causing and being effected by climate and environmental change as well as in adaptation and mitigation. (more information at www.wegcenter.at)

The research documented in this report was supported by the Austrian Science Fund.



Alfred Wegener (1880-1930), after whom the Wegener Center is named, was founding holder of the University of Graz Geophysics Chair (1924-1930) and was in his work in the fields of geophysics, meteorology, and climatology a brilliant, interdisciplinary thinking and acting scientist and scholar, far ahead of his time with this style. The way of his ground-breaking research on continental drift is a shining role model — his sketch on the relationship of the continents based on traces of an ice age about 300 million years ago (left) as basis for the Wegener Center Logo is thus a continuous encouragement to explore equally innovative scientific ways:
paths emerge in that we walk them (Motto of the Wegener Center).

Wegener Center Verlag • Graz, Austria

ISBN 978-3-9502940-6-4

February 2011

Contact: Andreas F. Prein.
andreas.prein@uni-graz.at

Wegener Center for Climate and Global Change
University of Graz
Leechgasse 25
A-8010 Graz, Austria
www.wegcenter.at

NHCM-1:
Non-hydrostatic climate modelling

Part 1

Defining and Detecting Added Value in Cloud Resolving Climate Simulations

Andreas Prein[1], Andreas Gobiet[1]

[1] Wegener Center for Climate and Global Change and Insitute for Geophyics, Astrophysics and Meteorology, Institute of Physics, University of Graz, Graz, Austria

15 February, 2011

Contents

Preface

This document describes progress and results of the FWF funded project NHCM-1 (project ID P19619). It is part one of a three document collection. The other parts are:

- M. Suklitsch, H. Truhetz, A. Prein, A. Gobiet (2011): *Current state of selected cloud resolving regional climate models and their error characteristics*, WegCenter Report, **40**, WegCenter Verlag, Graz, Austria, ISBN 978-3-9502940-7-1.

- A. Prein, H. Truhetz, M. Suklitsch, A. Gobiet (2011): *Evaluation of the Local Climate Model Intercomparison Project (LocMIP) simulations.* WegCenter Report, **41**, WegCenter Verlag, Graz, Austria, ISBN 978-3-9502940-8-8.

Further results have been or will be published in peer reviewed literature:

- Awan, N. K., A. Gobiet, M. Suklitsch, 2011: *Simulating climate extremes: Performance of high resolution regional climate models in European Alpine region*, J. Geophys. Res., submitted

- Awan, N. K., A. Gobiet, H. Truhetz, 2010: *Parameterization induced error-characteristics of MM5 and WRF operated in climate mode over the Alpine Region: An ensemble based analysis*, J. Climate, accepted

- Suklitsch, M., A. Gobiet, H. Truhetz, N. K. Awan, H. Göttel, D. Jacob, 2010: *Error Characteristics of High Resolution Regional Climate Models over the Alpine Area*, Clim. Dyn., doi: 10.1007/s00382-010-0848-5, published "online first" (http://www.springerlink.com/content/qj813x0255712826/)

- Suklitsch, M., A. Gobiet, A. Leuprecht, C. Frei, 2008: *High Resolution Sensitivity Studies with the Regional Climate Model CCLM in the Alpine Region*, Meteorol. Z., **17**, 4, 467 – 476

Additionally, there are plans for three peer reviewed publications about the topics covered by the three documents in this collection.

Abstract

convection resolving climate simulations (CRCS) have high potential to improve multiple error sources in state of the art regional climate models (RCMs) by explicitly resolving deep convection and by representing orography and land cover with high accuracy. However, the added value of CRCSs compared to simulations on coarser grids is difficult to assess by traditional statistical methods. In order to assess added value in a systematic way, we separate four categories (mean climate-, spatial-, temporal-, and event based added value) which can be evaluated separately. For each category several statistical methods are introduced which make different aspects of added value visible. They include traditional statistics like biases, correlation coefficients, or root mean squared errors (RMSEs), but also comprise methods which were especially developed to evaluate highly resolved simulations on grid point basis. The latter methods are of particular interest with regard to convection resolving simulations, since they are suited to analyze the model performance at small spatial scales. A specific focus is also on the evaluation of simulations with regard to small scale extreme events (e.g., heavy convective precipitation) and other local weather and climate phenomena like foehn and inversions. Each introduced method is designed to evaluate and quantify a certain aspect of simulations and sometimes also certain parameters. The presented categories and methods can be used as a guideline for the evaluation of high resolution climate simulations and are basis for the design of a standard evaluation scheme which is used in the Local Climate Model Intercomparison Project (LocMIP), which is described in the third part of this report series.

1 Introduction

The impacts of climate change on society and ecosystems are predominately related to the regional and local scale characteristics of climate change. Even if General Circulation Models (GCMs) are able to simulate the large scale circulation properly, they have much too coarse grid scales to resolve for example the high-resolution patterns of precipitation, regional wind fields, or important local atmosphere surface interactions. This calls for methods to transfer the results from GCMs to the regional and local scale.

The most rigorous of all downscaling techniques is the dynamical downscaling with regional climate models (RCMs) (e.g., Dickinson et al. (1989), Giorgi and Bates (1989), Wang et al. (2004), and Rummukainen (2010)). Thereby a RCM is nested into a global model (or another RCM simulation) to produce climate simulations with resolutions which cannot be achieved by GCMs because of computational constraints. State of the art regional climate simulations have horizontal grid distances between 10 km and 50 km (e.g., Christensen and Christensen (2007), Linden and Mitchel (2009), Loibl (2010)). However, many important phenomena are beyond these scales and have to be parameterized. One prominent example is moist convection, which has to be parameterized at grid-scales above about 3 km and is regarded as one major error source in climate simulations.

Therefore, climate simulations at convection resolving scales have two main advantages: First, they are able to resolve the topography and surface fields more accurately and second, deep convection can be simulated explicitly without parameterizations. In numerical weather prediction (NWP) convection resolving simulations have already proven to yield a more realistic precipitation pattern, especially for cases with moist convection over orographically complex areas (Mass et al. 2002).

However, convection resolving climate simulationss (CRCSs) have also disadvantages like high computational costs, and the need of large data storage space. Furthermore, the evaluation of convection resolving simulations poses difficulties because of the lack of highly resolved reference datasets. But even if such datasets are available, particularly the evaluation of highly resolved precipitation fields is challenging, because of the so called 'double penalty' problem. If a precipitation object like a convective cell is misplaced only by a few grid points traditional statistical methods (like the bias, the root mean squared error (RMSE), or the correlation coefficient) will treat the misplaced object as falls alarm and the observed object as a missed event and therefore refer a bad performance to the simulation even if the spatial patterns are perfectly reproduced. Small temporal or spatial misplacements of precipitation are however inevitable since

Table 1.1: Overview of the RCMs, institutes, and acronym of the LocMIP ensemble.

RCM	Version	Institute	abbreviation
CCLM	4.0	Wegener Center	W_CLM
CCLM	4.8	BTU Cottbus	B_CLM
MM5	3.7.4	Wegener Center	W_MM5
WRF	2.2.1	Wegener Center	W_WRF

the predictability limit falls of rapidly toward small scales (1 km – 100 km) because of upscale error propagation (Wernli et al. 2008). To avoid these 'double penalty' problem, multiple methods where developed especially in the NWP community.

In this report the following four main categories are defined, in which added value of CRCSs could get visible:

- mean climate,
- spatial characteristics,
- temporal characteristics,
- event based.

Specific statistical methods are introduced for each category to make different aspects of added value visible. Some methods which are focusing on climatologically means and variabilities are generally applicable for hindcasts and control runs (e.g., bias maps, Taylor statistics, ...), while others evaluate the performance of the simulations to reproduce the weather of the past and therefore need hindcast simulations (e.g., spatial and temporal highly resolved precipitation evaluations).

Examples for the described methods are given from published literature and from the evaluation of the Local Climate Model Intercomparison Project (LocMIP) ensemble (Prein et al. 2011). The LocMIP ensemble consists of four different RCMs (or versions of RCMs; see Table 1.1) and covers the eastern Alpine region within the periods June, July, and August (JJA) 2007 and December, January, and February (DJF) 2008. The acronyms of the simulations are build of a capital letter indicating the institution in which the simulations were performed (W for the Wegener Center for Climate and Global Change (WegCenter) and B for the Brandenburg University of Technology Cottbus (BTU)), the name of the RCM, and a number which describes the horizontal grid spacing of the simulation (10 for ∼10 km grid spacing and 3 for ∼3 km grid spacing). The major difference between the two resolutions is that in the 10 km simulations convection is parameterized while in the 3 km simulations deep convection is resolved explicitly. For more details about the LocMIP ensemble and its evaluation see Prein et al. (2011).

2 Mean Climate

In the atmosphere energy and mass is transferred from large scales to small ones and vice versa. Thereby, small scale features can affect large scale effects significantly. For example, in a regional climate model (RCM) with smaller grid spacing the orography is better represented which might not only have effects on regional precipitation but can also effect large scale rainfall patterns through shadowing effects of mountains. Added value of this category is not easily obtained, since it has to be transferred from the smaller scales where it resides, to the larger scales. However, it is easy to analyze, since standard techniques can be used for the evaluation of mean climate data, which is the focus of this chapter.

Three methods, bias maps, annual cycles, and conditional quantile plots, are described. In bias maps, the spatial distribution of temporal averaged biases is shown on grid-point basis, while in annual cycles spatial averages are calculated for climatologically means of each month. In conditional quantile plots climatologically conditional biases are calculated for quantiles of the spatial-temporal distribution of the reference dataset.

2.1 Bias Maps

Biases are systematic errors which can be calculated by subtracting observational values at a given time or time span from the simulated values. For deriving a map this has to be done point-wise on a common grid:

$$b_{jk} = \frac{1}{n}\sum_{i=1}^{n}\left(x_{ijk} - o_{ijk}\right), \tag{2.1}$$

where x_{ijk} is the simulated value and o_{ijk} is the observed value at a given time i, given longitude j, and given latitude k. The total time n, over which is averaged has to be long enough to capture decadal variability to derive a meaningful climatological bias of the simulation. It is advisable to draw bias maps not only on annual basis but also on seasonal, because error characteristics in climate models are often differing with different times of a year (see Section 2.2).

In Figure 2.1 the differences between the Fifth-Generation Mesoscale Model (MM5) 10 km (panel a) and 3 km (panel b) simulations minus the Integrated Nowcasting through Comprehensive Analysis (INCA) are displayed as mean bias of 2 m temperature for June,

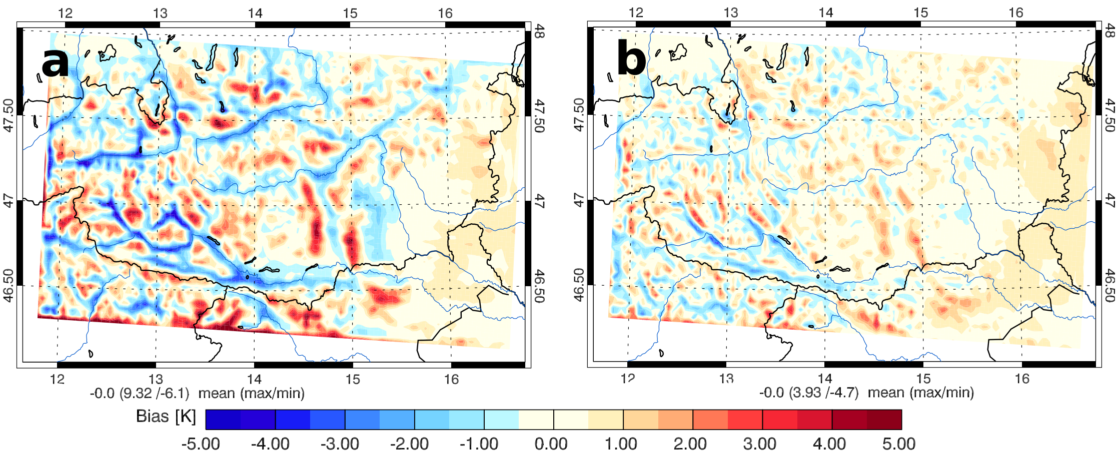


Figure 2.1: Bias maps of the 10 km (panel a) and 3 km (panel b) MM5 LocMIP simulations minus INCA. Shown are the mean differences of 2 m temperature for JJA 2007. Below the maps, the mean, the maximum, and the minimum difference is displayed.

July, and August (JJA) 2007. In both cases the spatial mean bias is zero, but the maximum and minimum biases are larger in the 10 km gridded simulation.

The major part of the improvement can be attributed to a more accurate representation of the orography in the 3 km simulation which leads to smaller biases in valleys and mountain ridges in orographically complex regions. In flat or hilly areas like south eastern Styria, northern Slovenia, or western Hungary biases do not change.

2.2 Annual Cycle

Annual cycles are a basic pattern of climate which are caused by the changing orbital position of the earth's during the course of a year. Thereby atmospheric parameters are influenced by the orbital position either directly by the variation of incoming solar radiation or indirectly by changes in the synoptic circulation (e.g. monsoon systems, strength of westerlies). Climate models typically have different error characteristics during different seasons because of the changing atmospheric and surface processes. A good example is the predominance of convective precipitation in mid latitude summers and the mostly frontal precipitation during winters.

Looking at biases in the annual cycle can reveal insights in weaknesses of the representation of physical processes within a climate model. Thereby, data are typically spatially and temporally averaged for each month of the year. The biases between the simulated minus the reference values can then be displayed as a time line representing the average monthly climatologically situation over a specific area.

2.3 Conditional Quantile Plots

The joint distribution of observed and simulated variables can give valuable insights in the simulation performance and in the statistical characteristics of the simulation and observation (Murphy and Winkler 1987). If the considered parameters are not already discrete variables they have to be rounded to a finite set of values. Considering the simulation x which can take any value x_i with $i = 1,\ldots, I$; and the observation o with values o_j $(j = 1,\ldots J)$ then the joint distribution can be denoted as:

$$p(x_i, o_j) = Pr\{x_i, o_j\} = Pr\{x_i \cap o_j\}\,; i = 1, \ldots, I; j = 1, \ldots, J. \tag{2.2}$$

Equation 2.2 shows the bivariate discrete probability distribution of the simulation and the forecast containing a probability for each $I \times J$ possible combinations. This joint distribution can be factorized as follows:

$$p(x_i, o_j) = p(o_j \mid x_i)\, p(x_i)\,; i = 1, \ldots, I; j = 1, \ldots, J. \tag{2.3}$$

This is called calibration-refinement factorization (Murphy and Winkler 1987). The first part in Equation 2.3 $p(o_j \mid x_i)$ consist of I conditional distributions which account for the probabilities that the observation measure a particular value o_j when the value x_i has been simulated. The second part in Equation 2.3 refers to the marginal distribution and shows the relative frequency of the simulated values x_i.

Conditional Quantile plots offer an informative way to display the properties of the joint distribution by showing the two parts of Equation 2.3 separately in a single plot Figure 2.2. The conditional distributions $p(o_j \mid x_i)$ are represented by their median and the lower and upper quartiles (Q25 and Q75). In this image a perfect forecast would exactly lie on the 1:1 diagonal. Here it can be seen, that there is a strong cold bias at warmer simulated temperatures. However, the most frequently simulated temperatures feature relatively small biases, which results in a moderately cold overall bias of -3.6 K in the 10 km (CLM34_D2), -3.5 K in the 3 km (CLM50_D3), and -3.1 K in the 1 km grid spacing simulation (CLM50_D4a). The distributions in the lower part of Figure 2.2 show the marginal distributions $p(x_i)$ of the simulations and the observation. A strong zero degree peak is visible in all COSMO model in CLimate Model (CCLM) simulations which flattens with increasing grid spacing.

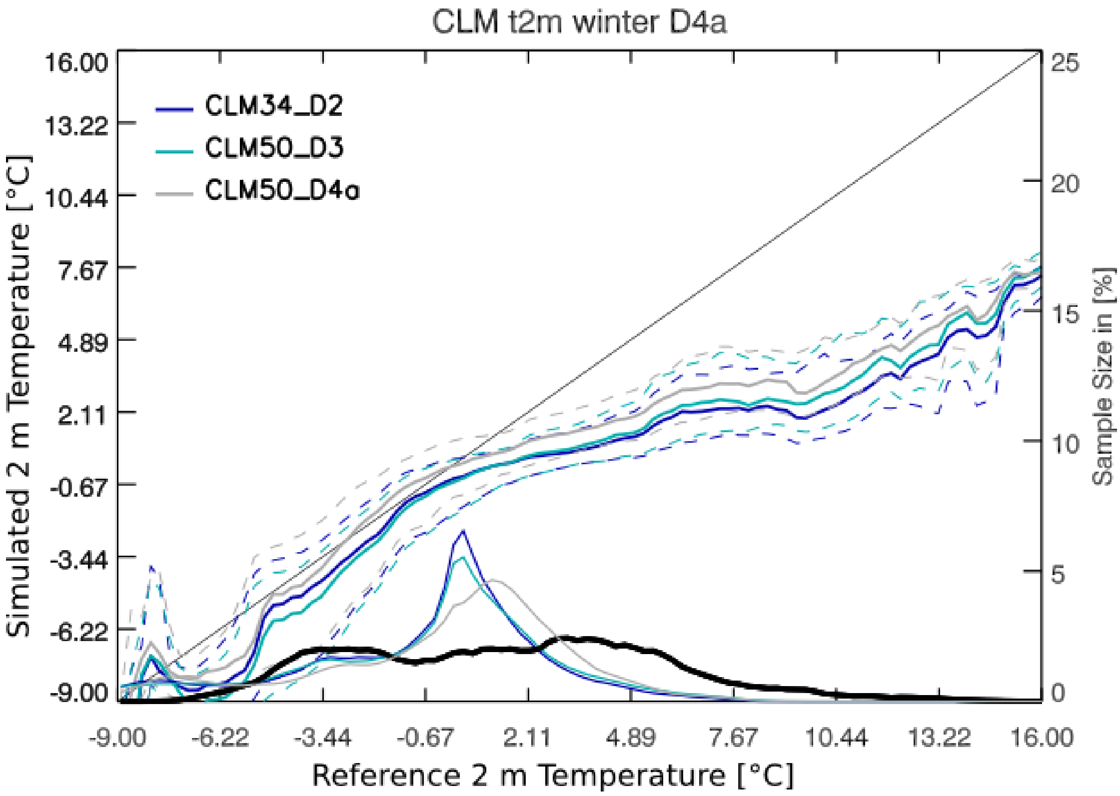


Figure 2.2: Conditional Quantile plot for 2 m temperature of three CCLM simulations with 10 km (dark blue), 3 km (light blue), and 1 km (grey) grid spacing compared with INCA in south-eastern Styria. The solid lines are the median values while the dashed lines are the 25 % and 75 % quartiles. The relative frequencies of the simulated and observed values correspond to the right y Axis. The black line displays the distribution of INCA.

3 Spatial Characteristics

In this chapter, we focus on the quality of spatial variability and spatial patterns of temporally averaged simulated fields. Since higher resolution directly relates to the potential ability of a model to more realistically represent fine-scale spatial patterns, one can expect to find added value here much more easily than in spatial averages, where only the net-effect of fine-scale improvements is visible. In our applications the model fields are always interpolated to the finer resolved grid of the reference dataset. This allows, e.g., to identify also added value in the finer scale simulations due resolved spatial structures which are simply smoothed out in the coarser scale simulations. Added value of this kind wouldn't necessarily imply that the coarser scale simulations are performing worse (they might be even perfect on their appropriate scale), but it would still demonstrate additional useful information in the finer scale simulations.

3.1 Spatial Correlation

The correlation r_{xo} is a measure for the linear connection between two variables and is defined as the standardized covariance cov_{xo}:

$$r_{xo} = \frac{cov_{xo}}{s_x \cdot s_o} \tag{3.1}$$

$$cov_{xo} = \frac{\sum_{i=1}^{N} (x_i - \bar{x}) \cdot (o_i - \bar{o})}{N - 1} \tag{3.2}$$

$$s_x = \sqrt{\frac{\sum_{i=1}^{N} (x_i - \bar{x})^2}{N}} \tag{3.3}$$

$$s_o = \sqrt{\frac{\sum_{i=1}^{N} (o_i - \bar{o})^2}{N}} \tag{3.4}$$

$$r_{xo} = \frac{\sum_{i=1}^{N} (x_i - \bar{x}) \cdot (o_i - \bar{o})}{(N - 1) \cdot s_x \cdot s_o}. \tag{3.5}$$

Thereby s_x and s_o are the standard deviations of the simulation and the observation and a bar above a variable means averaging. To obtain the spatial correlation coefficient the observation and the simulation has to be given at the same points i. A very useful tool for displaying the correlation coefficient together with the pattern root mean squared error (RMSE) E' and the standard deviations σ_o and σ_x is the Taylor diagram (Taylor 2001). The mathematical description looks as follows:

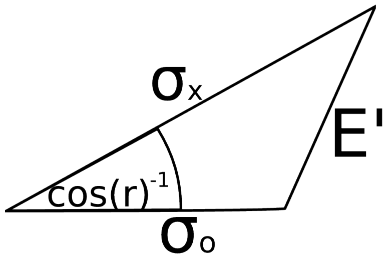


Figure 3.1: Geometrical relationship between the correlation coefficient r, the pattern RMSE E' and the standard deviations σ_o and σ_x.

$$E^2 = \bar{E}^2 + E'^2 \tag{3.6}$$

$$\bar{E} = \bar{x} - \bar{o} \tag{3.7}$$

$$E' = \left\{ \frac{1}{N} \sum_{i=1}^{N} [(x_i - \bar{x}) - (o_i - \bar{o})]^2 \right\}^{1/2} \tag{3.8}$$

$$E'^2 = \sigma_x^2 + \sigma_o^2 - 2\sigma_x\sigma_o r. \tag{3.9}$$

Hereby, the RMSE E^2 is the sum of the squared overall bias $\bar{E}^2$ and the squared pattern RMSE E'^2. Equation 3.9 can be geometrically interpreted by using the Law of Cosines $c^2 = a^2 + b^2 - 2ab\cos\phi$ where a, b, and c are the sides of a triangle, and ϕ is the angle opposite of c. The geometrical relationship of r, E', σ_x and σ_o is displayed in Figure 3.1.

In Figure 3.2 an example Taylor plot is shown for the Local Climate Model Intercomparison Project (LocMIP) regional climate model (RCM) simulations. Displayed are the spatial correlation coefficients of 2 m temperature over the eastern Alps for the mean 2007 June, July, and August (JJA) period. As expected, a general improvement is gained in terms of 2 m temperature when one moves from a 10 km (darker colors) to a 3 km (brighter colors) model grid. However, for other variables which are less directly related to the altitude of the orography than temperature, such improvements cannot be automatically expected.

A different way to look at the same data as in Figure 3.2 is to calculate r, E', σ_x, and σ_o for every time slice and to show them as point cloud or as contour plot (Figure 3.3) in the Taylor diagram. Thereby, the spread of the statistical parameters gets visible and structures in their temporal deviations appear. It should be noted that calculating the average of r, E', σ_x, and σ_o for every time step leads to a different result as calculating the quantities from the time averaged spatial field.

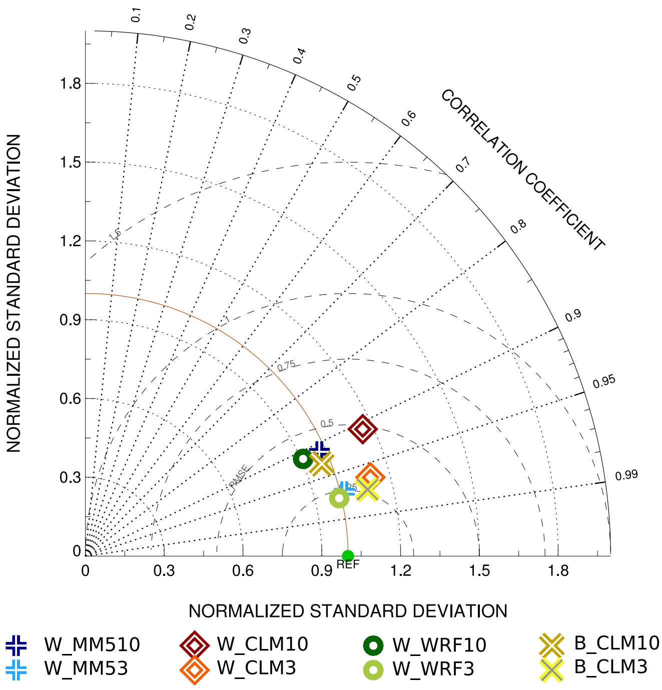


Figure 3.2: Taylor plot for the 2 m temperature of the four LocMIP RCMs in an eastern Alpine domain. The green point marks the reference field REF (here INCA). The distance to this point shows the pattern RMSE while the radial distance from the origin shows σ_x/σ_o. The correlations between the two fields display the azimuthally position of the simulated field.

3.2 Spatial Rainfall verification

Verifying spatial rainfall simulations is one of the most challenging verification tasks (Stevenson 2006). There are three problems that occur in spatial rainfall fields. First, precipitation fields are highly discontinuous and values can change on very small spatial distances. Second, there is often a mismatch between simulations and observations since simulations are generally given on some clearly defined grid while observations are taken on specific points or must be estimated indirectly from radar observations. Third, the distribution of precipitation is highly skewed which excludes statistical methods which demand Gaussian assumptions.

By decreasing the grid spacing of RCMs the above mentioned problems get even more severe, if precipitation fields are considered in a high temporal and spatial resolution.

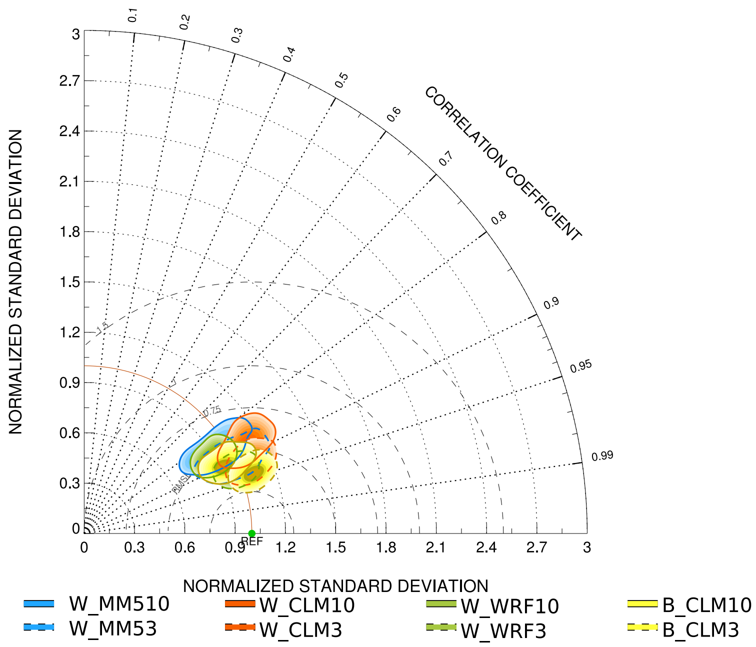


Figure 3.3: Same as in Figure 3.2 but for hourly data shown as contour plot. Intensive colours mean a clustering of points while different colours show different RCMs. 10 km simulations have solid contour lines while 3 km simulations have dashed.

This is because the chance of a 'double penalty' is increased if highly resolved simulations are evaluated by considering short time periods (e.g., hours). This 'double penalty' occurs because even slight misplacements of precipitation in the simulation can lead to numerous missed events and false alarms when traditional statistical methods are used. Even when the simulation represents the large-scale precipitation accurate, the small-scale errors dominate the total error (Mass et al. 2002). In general, small misplacements cannot be avoid because upscale error propagation lead to a decreasing predictability limit toward small scales (1 km – 100 km) (Wernli et al. 2008).

Multiple, approaches have been developed to overcome the above mentioned problems. Those methods do not require a perfect fit of the simulation and the observation at the fine scale. In the following sub chapters an overview of commonly used methods is given.

3.2.1 Filtering Approaches

The common feature of filtering methods is that they separate the spatial structures in different scales and compare them with the observation. Thereby, filtering approaches can further be separated in neighborhood methods and scale separation methods.

Neighborhood Methods

Neighborhood or fuzzy verification methods give credits to simulated events which are close to the observation. Ebert (2008) provides a good overview on 10 fuzzy verification methods which are used in the verification of numerical weather predictions. Ament et al. (2008) presented an evaluation of 12 fuzzy verification methods. They found three statistics, which perform very good on detecting a broad range of forecast errors. Those three methods are in detail discussed in the following paragraphs while additionally seven methods are introduced more briefly with respect to their references.

Fractional Skill Score Roberts and Lean (2008) developed a verification method which shows how the skill of a simulation varies with spatial scale. The basic idea behind the Fractional Skill Score (FSS) is that a simulation is useful if the spatial frequency of events is similar in the forecast and in the observation. The precondition for using this method is that the observation and the simulation are given on the same grid. In the first step, the originally observed and simulated fields are transferred to binary fields (I_o and I_x) by choosing a set of thresholds q (e.g., q=0.5 mm/d, 1 mm/d, 2 mm/d, and 4 mm/d) and setting all grid-cells exceeding the threshold to 1 and all others to 0,

$$I_o = \begin{cases} 1 & o \geq q \\ 0 & o < q \end{cases} \tag{3.10}$$

and

$$I_x = \begin{cases} 1 & x \geq q \\ 0 & x < q. \end{cases} \tag{3.11}$$

Secondly, for all grid points in the binary fields the spatial density of ones compared to zeros is calculated for a given squared neighborhood of length n:

$$O_{(n)}(i,j) = \frac{1}{n^2} \sum_{k=1}^{n} \sum_{l=1}^{n} I_O \left[i+k-1-\frac{(n-1)}{2}, j+l-1-\frac{(n-1)}{2} \right], \tag{3.12}$$

$$X_{(n)}(i,j) = \frac{1}{n^2} \sum_{k=1}^{n} \sum_{l=1}^{n} I_X \left[i+k-1-\frac{(n-1)}{2}, j+l-1-\frac{(n-1)}{2} \right]. \tag{3.13}$$

In Equation 3.12 $O_n(i,j)$ is the field of observed fractions of values exceeding the threshold for a square of length n and $X_n(i,j)$ accordingly is the field for the simulation assessing the spatial densities in the binary fields. Thereby, i goes from 1 to N_x and j goes from 1 to N_y, where N_x corresponds to the numbers of columns in the domain and N_y to the number of rows. The fractional fields $O_n(i,j)$ and $X_n(i,j)$ are generated for different spatial scales by changing the value of n from 1 to maximum $2N-1$, whereby N is $max(N_x,N_y)$. If neighborhood points lie outside the domain, their value is assumed as zero.

After the fractional fields $O_n(i,j)$ and $X_n(i,j)$ are known, the third step is to calculate fraction skill scores. Therefore, the mean squared error (MSE) is calculated:

$$MSE_{(n)} = \frac{1}{N_x N_y} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \left[O_{(n)}(i,j) - X_{(n)}(i,j) \right]^2 . \tag{3.14}$$

$$MSE_{(n)ref} = \frac{1}{N_x N_y} \left[\sum_{i=1}^{N_x} \sum_{j=1}^{N_y} O_{(n)i,j}^2 + \sum_{i=1}^{N_x} \sum_{j=1} X_{(n)i,j}^2 \right] \tag{3.15}$$

From Equation 3.14 the FSS can be calculated as a MSE skill score:

$$FSS_{(n)} = \frac{MSE_{(n)} - MSE_{(n)ref}}{MSE_{(n)perfect} - MSE_{(n)ref}} = 1 - \frac{MSE_{(n)}}{MSE_{(n)ref}}, \tag{3.16}$$

$MSE_{(n)perfect}$ is the MSE of a perfect simulation and therefore 0 for a given neighborhood length n. $MSE_{(n)ref}$ is the largest obtainable MSE from the given simulation and reference dataset. Therefore, a FSS of 1 indicates the best possible simulation.

An example is visible in Figure 3.4, where the FSS of the LocMIP Wegener Center for Climate and Global Change (WegCenter) COSMO model in CLimate Model (CCLM) simulations with 3 km grid spacing (left plot) is compared to the corresponding 10 km grid spacing mother simulation (middle plot). The right plot shows the difference between the 3 km minus the 10 km simulation. It gets visible that the simulation with the 3 km grid spacing has a higher FSS especially for medium and strong rainfall events and for large neighborhood sizes.

Upscaling The upscaling verification method was first published by Zepeda-Arce et al. (2000) and is conceptionally build on the assumption that a useful simulation resembles the observation when averaged to a coarser scale. In the upscaling method the threat score (TS) is calculated as a measure of scale. The TS compares the area of precipitation above a threshold between a simulation and an observation (e.g., see Wilks (2005)) and is defined as:

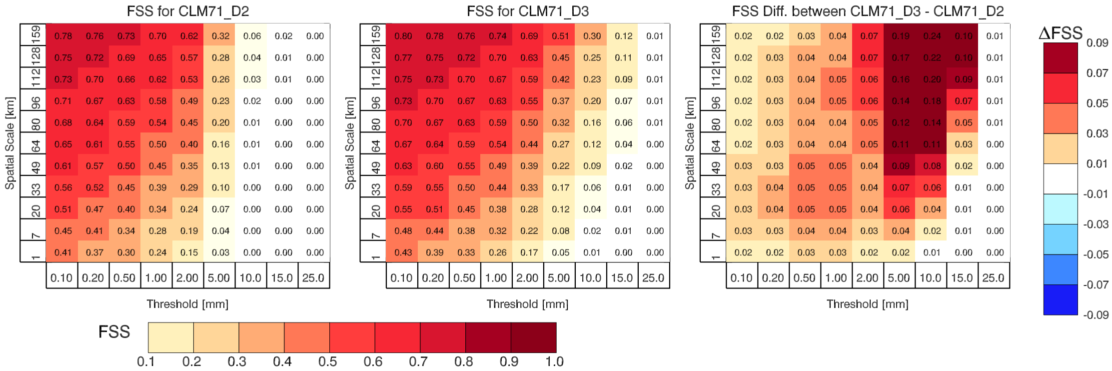


Figure 3.4: The left and the middle plot are the FSSs of the LocMIP WegCenter CCLM simulations with 3 km and 10 km grid spacing. The right plot shows the difference between the FSSs of the 3 km simulation minus the FSSs of the 10 km simulation. In rows different thresholds and in columns different neighborhood sizes are displayed.

$$TS = \frac{A_c}{A_o + A_x - A_c}, \tag{3.17}$$

thereby, A_c is the area where the simulation correctly produced precipitation above the threshold, A_o is the total observed area, and A_x the total simulated area. The best possible TS is one whereas the worst is zero. The TS is scale depended and gets typically higher with increasing scale. Furthermore, the TS can be expressed as a function of spatial scale and precipitation intensity by regrinding the simulated and observed fields on grids with different spacings and by using different precipitation thresholds.

Intensity-Scale Casati et al. (2004) uses a method which gives credits to a simulation which has more accurate structures than a random arrangement of the observation. As a first step, simulated and observed data has to be preprocessed. Therefore, all non-zero precipitation values are dithered by adding uniformly distributed noise with a magnitude of $\pm 1/64$ mm/h. Thereafter, the precipitation values are normalized with a (base 2) logarithmic transformation and the pixels with zero precipitation are set to -6. The normalization is done to produce more normally distributed data and to remove skewness. This data is then recalibrated with the following transformation function:

$$X' = F_O^{-1}(F_X(X)), \tag{3.18}$$

where each value of the simulated field X is substituted with the value of the observed field O having the same empirical cumulative probability F_O and F_X. With this procedure biases in the marginal distribution of the simulated precipitation are erased.

After the preprocessing the intensity-scale verification can begin. Therefore, the simulated and observed fields are converted into a binary image using thresholds q =0 mm/h, 1/32 mm/h, 1/16 mm/h, ..., 128 mm/h:

$$I_O = \begin{cases} 1 & O > q \\ 0 & O \leq q \end{cases} \tag{3.19}$$

and

$$I_{X'} = \begin{cases} 1 & X' \geq q \\ 0 & X' < q. \end{cases} \tag{3.20}$$

Then the binary error Z is calculated:

$$Z = I_{X'} - I_O. \tag{3.21}$$

With a two-dimensional discrete Haar wavelet decomposition the binary error can be expressed as the sum of components on different spatial scales:

$$Z = \sum_{l=1}^{L} Z_l, \tag{3.22}$$

where l is referring to the spatial scale of the error and not to the scales of the precipitations features or their displacements. A detailed description of the two dimensional Haar wavelet decomposition can for example be found in Mallat (1999) or Nievergelt (1999).

With the mean squared Z values the MSE of the binary image is calculated:

$$MSE = \bar{Z^2}. \tag{3.23}$$

Since the components of a discrete wavelet transformation are orthogonal $\bar{Z_l Z_{l'}} = 0 l \neq l'$ the MSE can be written as:

$$MSE = \sum_{l=1}^{L} \sum_{l'=1}^{L} \bar{Z_l Z_{l'}} = \sum_{l=1}^{L} \bar{Z_l^2} \tag{3.24}$$

$$\text{and} \tag{3.25}$$

$$MSE = \sum_{l=1}^{L} MSE_l, \tag{3.26}$$

where $MSE_l = \bar{Z_l^2}$. Thereby, MSE_l depends on the spatial scale l and the threshold u which enables the evaluation of precipitation on different scales and intensities. With the now obtained data the MSE skill score can be calculated for every scale and threshold:

$$SS = \frac{MSE - MSE_{random}}{MSE_{best} - MSE_{random}} = 1 - \frac{MSE}{2\varepsilon(1-\varepsilon)}, \tag{3.27}$$

where $MSE_{best} = 0$ is a perfect simulation, $MSE_{random} = 2\varepsilon(1-\varepsilon)$ is the MSE of random created binary field and ε is the fraction of rain-pixels in the observation. By assuming that the random observation and simulation binary fields are Bernoulli distributed variables $I_X \sim Be(\varepsilon)$ and $I_O \sim Be(\varepsilon)$ which have (unbiased) means $E(I_{X'}) = E(I_O) = \varepsilon$ and variances $\sigma^2_{I_O} = \sigma_{I^2_{X'}} = \varepsilon(1-\varepsilon)$ it can be shown that the binary error (Equation 3.21) has a mean $E(Z) = 0$ and a variance $\sigma^2_Z = \sigma^2_{I_O} + \sigma_{I^2_{X'}} = 2\varepsilon(1-\varepsilon)$. The expected value of MSE is then $MSE_{random} = E(Z^2) = E(Z - E(Z))^2 = \sigma^2_Z = 2\varepsilon(1-\varepsilon)$. With the assumption that the MSE is uniformly distributed over all scales SS can be written as the sum of its means over all scales by using Equation 3.24 and Equation 3.27:

$$SS_l = 1 - \frac{MSE_l}{2\varepsilon(1-\varepsilon)/L}. \tag{3.28}$$

Minimum coverage The minimum coverage verification assumes that a skill-full simulation produces an event over a minimum coverage of a region of interest. This method was presented by Damrath (2004) on the International Verification Methods Workshop in Montreal (15 – 17 September 2004). The presentation is available under http://www.cawcr.gov.au/projects/verification/Workshop2004/presentations/5.3_Damrath.pdf (15 February, 2011).

Fuzzy logic, joint probability The idea behind the fuzzy logic method (Damrath 2004) is to use neighborhood events to generate probabilities. Thereby, the forecast as the observation have likelihoods to become an event or a non event. If greater parts of a forecast are correct than incorrect, the forecast is useful. For the verification the probability of detection (POD), the false alarm ratio (FAR), and the equitable thread score (ETS) are used (e.g., Wilks (2005)). Ebert (2002) presented a variation of this method at the NCAR/FAA Verification Workshop in Boulder, Colorado from the 30 July – 1 August 2002. The presentation is available under http://www.rap.ucar.edu/research/verification/verification_wkshp_2002/pdfs/FuzzyVerificatio_ebert.pdf (15 February, 2011).

Multi-event contingency table Atger (2001) suggested the method of multi-event contingency table which is based on the assumption that an useful simulation produces at least one event which is close to an observed event. Therefore, the traditional 2×2 contingency table gets expanded. In contingency tables a discrete amount of possible combinations of simulated and observed event pairs are compared. In the most easy case of a 2×2 contingency table the probabilities of predicting an event (e.g., an Hurricane) when there is one observed, to predict an event when there is non observed, to predict non event when there is one observed, and to predict none event when there occurs none are shown in a table (a more detailed description can be found in e.g., Wilks (2005) or Stevenson (2006)).

Atger (2001) expands the 2×2 contingency table not only to several intensity thresholds, but also introduces the possibility of adding additional thresholds like temporal or spatial closeness of an event.

Pragmatic approach The method of producing a probabilistic *pseudo-ensemble* with a deterministic forecast within a neighborhood of the reference for verifying the forecast quality was published by Theis et al. (2005). They suggested that instead of evaluating the simulation in a neighborhood with the reference in the same neighborhood, it should be compared with the central grid box of the reference. For the evaluation, they used the Brier score and Brier skill score with the decision model, that a skilful simulation has a high probability of detecting events and non events. For a definition of the Brier score see for example Wilks (2005) or Stevenson (2006).

Practical perfect hindcast Brooks et al. (1998) had the idea that a skillful forecast is resembling a practically perfect hindcast which is a forecast achieved by having knowledge of the observation beforehand. This fuzzy verification method was particularly developed for rare events. For the evaluation the TS ore critical success index (CSI) is used (see e.g., Wilks (2005) or Stevenson (2006))). To generate a practically perfect forecast a Gaussian kernel filter is applied to the observation to obtain a probability field for the event.

Conditional square root for ranked probability score (RPS) In the method of Germann and Zawadzki (2004) precipitation frequencies are used in logarithmic increasing intervals to compute the ranked probability score (RPS):

$$RPS = \frac{1}{M-1} \sum_{m=1}^{M} (CDF_{x,m} - o_m)^2, \tag{3.29}$$

where M is the number of event categories and $CDF_{x,m}$ is the cumulative probability of the simulation to exceed a particular threshold for event category m. o_m gets one if the observed value exceeds the threshold for category m and zero if not. Taking the square root of RPS reveals the standard error of the simulation probability in probability space. To compare results over different kind of events, $\sqrt{RPS}$ is normalized by the observed rain fraction which leads the conditional square root of the RPS.

Areal-related RMSE The basic idea behind the verification method of Rezacova et al. (2007) is that a useful simulation has a similar intensity distribution as the observation. Therefore, precipitation values are compared within a certain neighborhood to obtain a scale dependency of the RMSE. The observed and simulated values within the neighborhood have to be ordered from the smallest to the largest and then the RMSE is computed from the ordered series.

Scale Separation Methods

The goal of the here presented methods is, to examine performance as a function of spatial scale. For this reason, Fourier or wavelet transformations are a common tool to decompose atmospheric fields and look at different scales separately.

The Discrete Cosine Transformation (DCT) Using Furrier transformations enables to decompose a periodic function into its wavenumbers of partial frequencies (Peixoto and Oort 1992). Denis et al. (2002) where the first who used the 2D discrete cosine transformation (DCT) for limited areas. Therefore, the precipitation field has to be mirrored at the position $i = j = -1/2$ to make it symmetric. Thereafter, the Fourier transformation can be applied, centered on $i = j = -1/2$. This is a special case of a Fourier transformation which is called DCT because the sine components of the Fourier series are zero for symmetric functions. Concerning a 2D field f_{ij} of N_i by N_j grid points, the direct and inverse DCT are defined as:

$$F(m,n) = \beta(m)\beta(n) \sum_{i=0}^{i=N_i-1} \sum_{j=0}^{j=N_j-1} f(i,j) \cos\left[\pi m \frac{(i+1/2)}{N_i}\right] \cos\left[\pi n \frac{(j+1/2)}{N_j}\right] \tag{3.30}$$

$$f(i,j) = \sum_{m=0}^{m=N_i-1} \sum_{n=0}^{n=N_j-1} \beta(m)\beta(n) F(m,n) \cos\left[\pi m \frac{(i+1/2)}{N_i}\right] \cos\left[\pi n \frac{(j+1/2)}{N_j}\right] \tag{3.31}$$

$$\beta(m) = \begin{cases} \sqrt{\dfrac{1}{N_i}}, & \text{for } m = 0 \qquad (3.32\text{a}) \\ \sqrt{\dfrac{2}{N_i}}, & \text{for } m = 1, 2, \ldots, N_i - 1 \qquad (3.32\text{b}) \end{cases}$$

$$\beta(n) = \begin{cases} \sqrt{\dfrac{1}{N_j}}, & \text{for } n = 0 \qquad (3.33\text{a}) \\ \sqrt{\dfrac{2}{N_j}}, & \text{for } n = 1, 2, \ldots, N_j - 1 \qquad (3.33\text{b}) \end{cases}$$

Thereby, f_{ij} is the value of the field at grid point number (i, j), and F_{mn} is the real spectral coefficient corresponding to the 2D-wavenumber at (m, n). A more detailed derivation of the DCT can be found in the appendix of (Denis et al. 2002) and an application of this method on high resolved RCM simulations can be seen in Kapper (2009).

To generate variance spectra of a 2D field, the variances have to be connected to a specific wavelength. To do so, the method of *binning* was suggested by Denis et al. (2002). It is based on dividing the wavenumber field into multiple quarters of ellipses. The space between two ellipses can be connected to a specific wavenumber for which the variances are summed up. For a detailed description see Denis et al. (2002).

As an example for the application of the DCT Figure 3.5 shows the power spectra of 700 hPa wind speed in January 2008 over south-east Styria for Fifth-Generation Mesoscale Model (MM5), CCLM, and Weather Research and Forecasting model (WRF) simulations on 10 km, 3 km, and 1 km grid spacing. Evident, is the increasing variance of small grid spacing simulations in smaller wavelengths. Thereby, WRF shows always the lowest variances, whereas CCLM has always the highest. The effective resolution of the simulations can be estimated in Figure 3.5 by looking at the points where the lines of the coarser gridded simulations start to differ from those of the finer gridded one. For example, the 3 km simulations start to differ at approximately 30 km wavelength from the 1 km simulations, which means that the effective resolution of the wind field in 700 hPa is ∼10 times the grid spacing.

A higher variance in smaller wavelengths however, does not directly indicate that there is added value in RCM simulations with smaller grid spacings. It is a necessary but no sufficient feature because it has to be proven that the higher variance also carries additional information and not only random noise.

Intensity-scale The intensity scale method, published by Casati et al. (2004) was already introduced in Section 3.2.1. By using Haar wavelets, reference and simulated precipitation fields are separated in different scales which than are evaluated by a MSE skill score. Since the intensity-scale method uses spectral decomposition and skill scores on different scales and intensities it can be assigned to fuzzy and scale separation methods.

Variogram The basic idea of evaluating fields with the variogram method is related to publications of Gebremichael and Krajewski (2004), Germann and Joss (2001), Harris et al. (2001), and Zepeda-Arce et al. (2000). The method itself was proposed by Marzban and Sandgathe (2009) and compares two fields in terms of their covariance structures. Thereby, the fields can be compared in two different ways, where one accounts for displacements and intensity errors, while the other is only sensitive to intensity errors. The analogy, which is insensitive to intensity and only accounts for displacements can be also

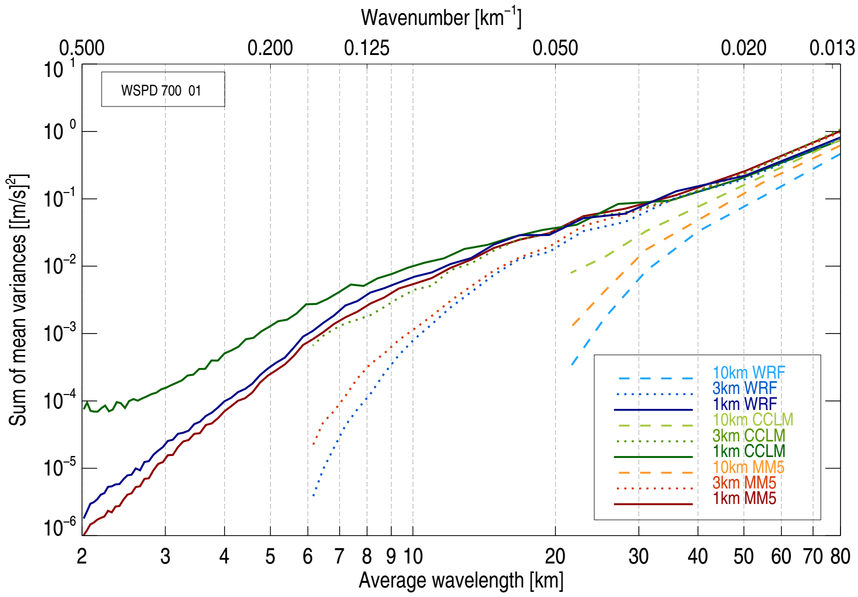


Figure 3.5: Variance spectra of January 2008, 700 hPa wind speed in south-east Styria for three different RCM (different colours) outputs with 10 km, 3 km, and 1 km grid spacing (different line-styles) (Kapper et al. 2010).

calculated and is called *correlogram*.

As the name suggests, the central quantity in the variogram method is the variogram which is defined as the RMSE between two points of the field as a function of the distance between the points (Noel 1993). Thereby, all points within the same distance are averaged. The variogram detects variations within the field as a function of scale and so summarizes the covariance structure in a spatial field.

As mentioned above three different methods of the variogram where introduced by Marzban and Sandgathe (2009). In the first, all grid points are considered to calculate the variogram. The second should be used for fields where continuous and discrete quantities occur (e.g., precipitation). In this case only the non-zero grid-points are used to generate the variogram. It can be shown, that the first method accounts for displacements and intensity errors while the second does not account for global displacements (e.g., shifts) any more. Finally, the differences between the reference variogram and the simulated variogram are calculated which is then called 'delta variogram'. A perfect forecast is denoted with zero differences at all scales.

As already mentioned, the third method is called correlogram, which is insensitive to intensity errors. It is build of the generalized Pearson correlation coefficient to a 2-dimensional case. Therefore, the differences between two correlograms, called 'delta correlogram' show only the displacement errors.

An application and a comparison of the variogram method with two other spatial verification methods can be found in Marzban et al. (2009).

3.2.2 Displacement Approaches

Feature Based Approaches

The basic idea behind object respectively feature based spatial verification methods is to identify relevant features in the simulated and observed field and compare characteristic attributes of both fields with each other. Some example methods will be discussed in the following paragraphs.

SAL Verification Method Wernli et al. (2008) introduced an object-based quality measure which considers three components accounting for the structure, amplitude, and location (SAL) of a precipitation field. The SAL measure aims to address the following issues:

- quantify the quality of a simulated precipitation field over a fixed area (e.g., a river catchment),
- considering the structure of the precipitation field (e.g., scattered convective cells, frontal rain, ...), and
- a one to one matching between the reference and the observed field is not required.

As a first step individual precipitation objects have to be identified for calculating the location and structure components. Therefore, a precipitation threshold is chosen:

$$R^* = fR^{max}. \tag{3.34}$$

In Equation 3.34 R is the precipitation field, R^* is the precipitation threshold, and R^{max} is the maximum rainfall within the considered domain. For the constant f, Wernli et al. (2008) suggested a value of $f = 1/15$ because this factor leads to contours which are reasonable with contours identified by eye. Grid-cells cluster to an object if one of the neighborhood cells is above the threshold f. A possible algorithm to identify objects can be found in Wernli and Sprenger (2007).

The amplitude component A is calculated by using the normalized differences of the average precipitation values:

$$A = \frac{D(R_{mod}) - D(R_{obs})}{0.5\,[D(R_{mod}) + D(R_{obs})]}. \tag{3.35}$$

$D(R)$ denotes the domain average precipitation in the observed (obs) or simulated (mod) field.

$$D(R) = \frac{1}{N} \sum_{(i,j)\in D} R_{ij}, \tag{3.36}$$

where R_{ij} are the grid point values of the precipitation amount. A is a simple quantity, showing information of the domain wide amount of precipitation by ignoring the field‘s sub regional structure. It has values between [-2 ... +2] whereby a perfect simulations in terms of amplitude leads to $A = 0$. An A value of ± 1 corresponds with an over- or underestimation of precipitation by the factor of 3.

The location component L consist of two additive parts: $L = L_1 + L_2$:

$$L_1 = \frac{\mid \mathbf{x}(R_{mod} - \mathbf{x}(R_{obs}) \mid}{d}, \tag{3.37}$$

where d is the maximum distance between two boundary points of domain D and $\mathbf{x}(R)$ corresponds to the center of mass of the precipitation field R. L_1 has values between 0 and 1 and gives a first order estimation of the precipitation distribution in the considered region whereby $L_1 = 0$ if the centers of mass are at the same position. However, many different fields can have equally centers of mass which makes the second quantity L_2 necessary. It accounts for the distances between the center of mass of the total precipitation field and single precipitation objects. Therefore the total precipitation amount is calculated for every object:

$$R_n = \sum_{(i,j)\in R_n} R_{ij}. \tag{3.38}$$

Then the averaged and waited distances between the centers of mass of the individual objects are calculated:

$$r = \frac{\sum_{n=1}^{M} R_n \mid \mathbf{x} - \mathbf{x}_n \mid}{\sum_{n=1}^{M} R_n}, \tag{3.39}$$

whereby, M is the total number of precipitation objects. r can have a maximum value of $d/2$ (half of the maximum distance between two grid points in the domain). It is zero if there is only one object in the domain. It should be noted that $\sum_{n=1}^{M} R_n$ is not equal to $\sum_{(i,j)\in D} R_{ij}$ (in Equation 3.36) because the former only considers grid points above the threshold R^*. L_2 is now calculated as follows:

$$L_2 = 2\left[\frac{\mid r(R_{mod}) - r(R_obs) \mid}{d}\right]. \tag{3.40}$$

L_2 is only greater than zero if either the observation or the simulation has more than one object in the domain. L_2 can range from zero to one which means that L can have values between zero and two. Zero indicates that the total center of mass as the averaged distances of the objects and the center of mass are the same in the observation and in the simulation. However, this does not mean a perfect match between the two field, because the L value is for example invariant to rotations around the center of mass.

The last missing component in the SAL method is the structure (S) component in which the volumes of the precipitation objects are compared and which contains information about the mass and the shape of the objects. Therefore, first a scaled volume V_n is calculated for every object:

$$V_n = \sum_{(i,j)\in R_n} R_{ij}/R_n^{max} = R_n/R_n^{max}. \tag{3.41}$$

In Equation 3.41 R_n^{max} stands for the maximum precipitation within the object n and has to be $R_n^{max} \leq R^{max}$. Then V, the weighted mean of all objects scaled precipitation volume is computed for the reference and the simulated field:

$$V(R) = \frac{\sum_{n=1}^{M} R_n V_n}{\sum_{n=1}^{M} R_n}. \tag{3.42}$$

Similar to the A component, the S component is the normalized difference in V:

$$S = \frac{V(R_{mod}) - V(R_{obs})}{0.5\,[V(R_{mod}) + V(R_{obs})]}. \tag{3.43}$$

S becomes negative if too small or too peaked objects are simulated or positive if widespread precipitation is modeled but small convective cells are observed. Two example SAL diagram from Wernli et al. (2008) are shown in Figure 3.6. Panel a) displays daily precipitation forecasts of the COSMO-aLMo and panel b) those from the ECMWF model during the summer seasons 2001–04 in the German part of the Elbe catchment.

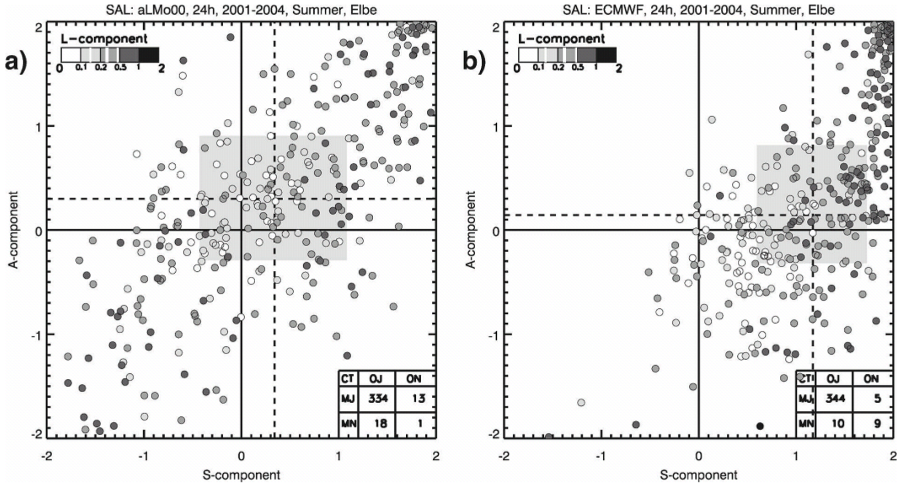


Figure 3.6: SAL diagrams for daily summertime precipitation forecasts from the COSMO-aLMo (panel a) and the ECMWF model (panel b) over the German part of the Elbe catchment during 2001–04. The dots correspond to the three values in the SAL statistic for a particular day. The L-component is shown as the color of the circles, while the values of the A component are displayed on the y-axis and those of the S components on the x-axis. The median A and S values are shown as dotted lines and the grey box displays the 25th and 75th percentile. The contingency tables in the right bottom corner of each plot display the number of falls alarms (cell MJ-ON), missed events (cell MN-OJ), dry days (MN-ON), and wet days (MJ-OJ) in the simulation (M) and observation (O) (Wernli et al. 2008).

Cluster Analysis The statistical method of cluster analysis was used by Marzban and Sandgathe (2006) to identify clusters in precipitation fields whose members are more similar to one another as to members of the other clusters. The specific kind of cluster analysis used is called agglomerative hierarchical cluster analysis (CA). Details of cluster analysis in general and for the CA method in particular can for example be found in Everitt et al. (2001). Generally, a cluster is an event or an object in a gridded field. The method is used to identify objects in the observation and simulation field. Afterwards, the two fields are compared with respect to the clusters fund in each of them.

The result from the cluster analysis is not a single error value, but an array of error values for different numbers of clusters in the two fields. Marzban and Sandgathe (2006) suggest that the outcome should be viewed as error surface in a three dimensional space, whereby the x and y dimension are the number of clusters in the observation and forecast field (NC_o and NC_f) and the height of these surface shows the error of the forecast at the correspondent scale.

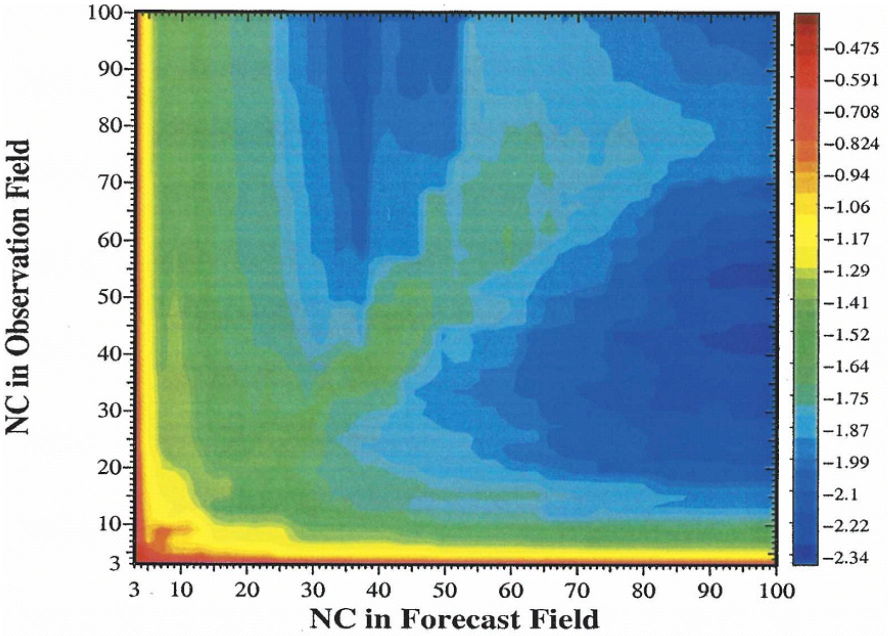


Figure 3.7: Cluster analysis performed on Coupled Ocean Atmospheric Mesoscale Prediction System data compared with reference data from the River Forecast Center. On the x-axis the number of clusters (NC) in the forecast field and on the y-axis those on the observed field are displayed. The colors of the contour indicate the calculated error from the CA analyses (Marzban and Sandgathe 2006).

Figure 3.7 shows the results of a cluster analysis performed by Marzban and Sandgathe (2006). They compared precipitation fields from the Coupled Ocean Atmospheric Mesoscale Prediction System with reference data from the River Forecast Center. In general the forecast errors are large (red through yellow) if the number of clusters is small, either in the reference or in the simulation. This means that the error between both fields is generally large if there are only a few large clusters in one field whereby there are multiple small clusters in the other field. The ridge along the diagonal shows also increased error values which occur when equal numbers of clusters are 'forced' to occur in both fields when the 'true' numbers are unequal. The symmetric structure along the ridge is a sign for the quality of the simulation because if it would be absent the forecast would have the wrong scale. Here, the error surface has smaller values when the number of clusters in the forecast field is greater than in the observed field. This is because the forecast has generally to many clusters compared to the observed field.

Method for Object-based Diagnostic Evaluation (MODE) Davis et al. (2006) suggested an object based rainfall verification method, where precipitation objects are defined in the forecast and the observation with a smoothing and thresholding process. The basic verification quantity used is the rainfall distribution rather than the rainfall amount. This has multiple reasons but was especially chosen because observations (e.g., radar images) are more precise in the location of at least mesoscale precipitation areas than at the amount of rainfall within them.

After objects are identified in both fields, objects from the observation have to be linked with objects from the simulation to compute error statistics. Therefore, matching conditions have to be described that are not too restrictive on the location of the objects but also do not compare objects that are not related by general physical process.

The following properties are calculated for each object.

- Intensity: to account for the rainfall intensity the 25th and 90th percentiles are evaluated.
- Area: is a simple measure of object size
- Centroid: comparing the centers of mass give an estimation of rainfall displacement.
- Axis angle: a line through an object to identify the orientation of an object.
- Aspect ratio: the ratio between the major and the minor axes of an object.
- Curvature: a circular arc is fitted to the object to show the objects general deviation from straightness. The corresponding radius of the arc is used to describe the objects curvature.

A practical application of the Method for Object-based Diagnostic Evaluation (MODE) on numerical forecasts can be found in Davis et al. (2009).

Contiguous Rain Area (CRA) The Contiguous Rain Area (CRA) method was published by Ebert and McBride (2000) and evaluates the location, the size, the intensity, and the fine scale patterns of features in precipitation systems. Similar than in the MODE method objects from the observation have to be linked to objects in the simulation to evaluate the performance of the simulation. The displacement is thereby calculated by translating the simulated rainfall field until the squared difference is minimized.

Procrustes Micheas et al. (2007) suggested the usage the Procrustes shape analysis methods for verifying precipitation forecast. With this method the forecast error can be decomposed into any number of components like displacement, shape, size, orientation, and intensity. Furthermore, every error component can get a specific weight so that the quality of the forecast can be calculated.

A further application of these method can be found in Lack et al. (2010).

Field Deformation

Field deformation approaches have in general, that they evaluate how much a field has to be transformed to match the observation. This is in particular interesting for precipitation because models often produce phase errors and displace weather systems in space or time.

Displacement Amplitude Score The displacement and amplitude score was published by Keil and Craig (2007) and in an revised version by Keil and Craig (2009). An optical flow technique is used to avoid problems in identifying features and problems occurring through linking specific objects in the simulation to objects in the observation.

The optical flow method is based on a pyramid algorithm where the fields first are re-gridded on a coarser grid where 2^F pixels are averaged on one pixel element. F is thereby called the sub sampling factor. On the coarse grid for every grid cell of the simulation the distance of the minimum squared error (compared to the observation) is searched in a neighborhood of $\pm$ 2 grid cells. The obtained vector field of displacements is then applied to the original simulated image which generates an intermediate image accounting for large scale displacements. Afterwards, the intermediate image is coarse grained by averaging 2^{F-1} pixels which is the next pyramid level and the displacement vector field is calculated and applied as mentioned above. This algorithm is repeated until the full resolution is obtained. After the total displacement vector field, which morphs the simulation to the observation, is defined, the displacement vector field, which morphs the observation to the simulation, is calculated with the same algorithm.

The resulting displacement vector fields are the sum of the vector fields at each pyramid level and are used to build the final morphed images. More details on the algorithm can be found in Keil and Craig (2007) and Zinner et al. (2008).

The displacement and amplitude score is then calculated by considering two quantities. The first accounts for the displacement error which is the magnitude of the displacement vector giving the distance of a simulation to an observed object (if any). The second accounts for the amplitude error which is the difference between the observed and the morphed simulation field. Accordingly, the same quantities are calculated for the morphed observation onto the simulation.

If two objects in the simulation and the observations are separated by more than the maximum search distance, they are treated as two independent objects and the amplitude error accounts for one missed event and one false alarm.

Finally, the displacement and amplitude errors are combined to yield the displacement amplitude score. Thereby, the two components are weighted so that the maximum possible displacement error between two objects equals the amplitude error for the same two objects that would occur if the distance between the two objects would have been larger as the maximum search distance.

To visualize the above described formalism, Figure 3.8 shows an idealized example of an observation (panel a) and an identical forecast field (panel b) which is 50 pixels shifted to the right. The resulting vector field (forecast is transformed to observation) is shown in panel b. Panel c displays the morphed forecast field and panel d and e show the displacement and amplitude error fields.

Forecast Quality Index (FQI) Venugopal et al. (2005) suggested a metric called forecast quality index (FQI), which combines the measure of errors in amplitude and displacement of spatial rain intensity fields. The distant based measurement is a non-linear metric is then called the *Hausdorff Distance* which was modified by Venugopal et al. (2005) to improve the treatment of outliers and is called Partial Hausdorff Distance (PHD). The amplitude-based metric was new developed by them and is called the *universal image quality index*.

Image Warping Image warping can give an estimation of the spatial error by accounting for the warping deformation of a simulation. There are multiple publications witch described methods using image warp techniques to verify the accuracy of precipitation forecasts (e.g., Dickinson and Brown (1996), Alexander et al. (1999), Åberg et al. (2005), or Gilleland et al. (2010)).

The basic method of the image warping technique can be separated in three parts. First, a warping function is applied to the simulated field that controls its deformation. Second, an error function is needed that accounts for the intensity deviation between the morphed simulation and the reference field. Third, to penalize unrealistic deformations of the simulated field a smoothness prior has to be used.

Figure 3.9 shows the ideal test case pert007 from the ICP (`http://www.ral.ucar.edu/projects/icp/`) on which image warping was applied by Gilleland et al. (2010). On the left, the forecast field is shown, which is displaced by 48 km too far to the east and 80 km too far to the south. Additionally, the rainfall intensity is reduced by 1.27 mm from the original observed field (shown in the middle). Both fields are shown only above a threshold of 75 % to remove frequency biases and scattering. In the figure of virtually forecasted rainfall (left pannel), also vectors indicating the warp movement are shown. On the right the final deformed forecast field is displayed.

3.2.3 Overview of the Ability of Spatial Rainfall Verification Methods

The spatial verification methods introduced in the subsections above provide great opportunities for better interpretable and more accurate precipitation evaluations. Thereby,

Attribute	Traditional	Feature Based	Neighbor-hood	Scale	Field Deformation
Performance at different scales	Indirect	Indirect	Yes	Yes	No
Location error	No	Yes	Indirect	Indirect	Yes
Intensity errors	Yes	Yes	Yes	Yes	Yes
Structure errors	No	Yes	No	No	Yes
Hits, etc.	Yes	Yes	Yes	Indirect	Yes

Table 3.1: Attributes measured by the traditional and new spatial precipitation verification methods (Brown et al. 2009).

each method is useful in certain situations and to answer certain questions. However, all of them have also limitations.

The major limitation of feature based approaches (scale separation and feature based) is that they do not clearly isolate different kinds of errors (e.g., amplitude, displacement). In case of the displacement methods (feature based and field deformation) the matching criteria are somehow arbitrary and many parameters have to be fitted.

Advantages of feature based approaches are that they account for uncertainties in the simulation and the observation and are able to deal with unpredictable scales. They give scale dependent information and are mostly simple and easy to interpret. The displacement approaches are able to measure the displacement and give credits to close precipitation fields. Furthermore, feature based approaches are able to measure the structure of precipitation while field deformation methods are able to account for aspect ratio and orientation of the rainfall objects.

In table Table 3.1 a quick overview is given on the attributes that are measured by different methods.

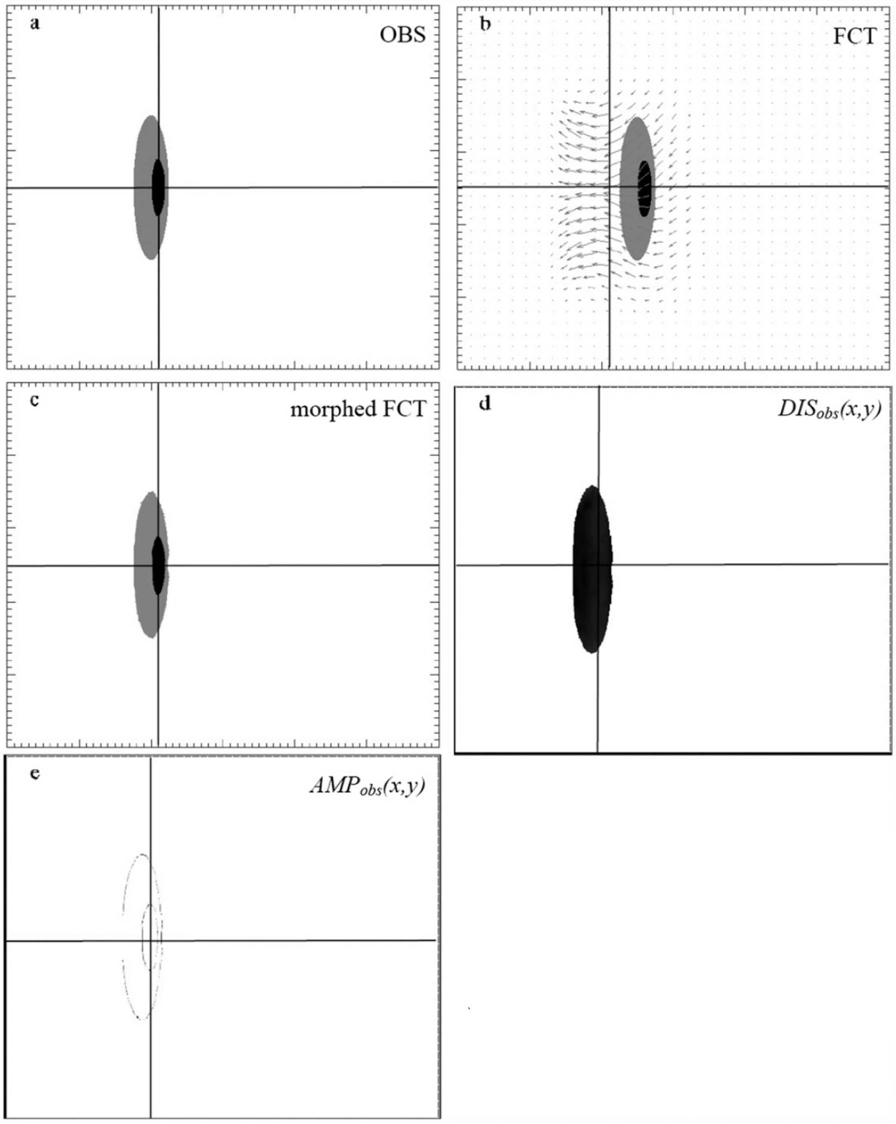


Figure 3.8: An idealized example of an observed object (OBS) (panel a) and a forecasted (FCT) object (panel b) which is 50 grid boxes shifted to the right. Here the simulation is shifted towards the observation. Panel b shows the vectors of the displacement array. Panel c displays the morphed forecast while panel d shows the observation space displacement error field and panel e the amplitude error field (Keil and Craig 2009).

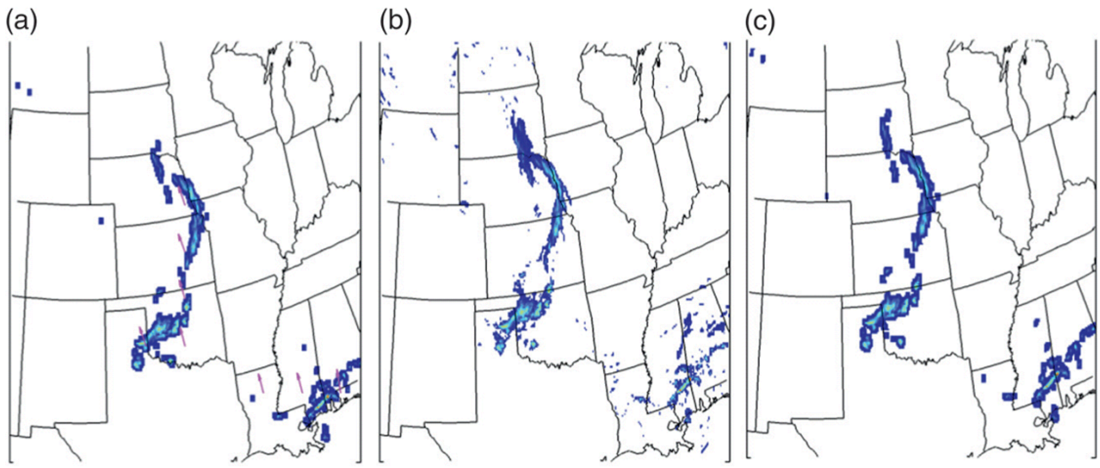


Figure 3.9: Example of an image warp for an ideal test case. Left is the original forecast field, in the middle is the verification field and right is the deformed forecast field (Gilleland et al. 2010).

4 Temporal Characteristics

In this chapter, methods to analyze and evaluate the temporal characteristics of simulations are introduced and discussed. First, diurnal cycles are considered because a correct representation of a diurnal cycle needs an interaction of many different physical processes and can therefore give valuable insights in model physics. The temporal correlation is used to account for linear relations between simulations and observations. The cross correlation analyze is introduced to find temporal delays between the simulation and the reference datasets. Furthermore, to analyze differences in the information contained in the temporal frequencies spectra, a frequency decomposition method is discussed.

4.1 Diurnal Cycle

The diurnal cycle is one of the most basic climate patterns. The most common is probably the diurnal cycle of temperature which occurs because of the varying energy input from the sun during different times of the day. For a time series $\vec{X}_t$ with k time-slices per day the calculation of the diurnal cycle is a simple average over every $t = [i, i+k, i+2k, \ldots, i+N]$ element of the time vector where $i = 1, 2, \cdots, k$ and N is the number of elements in $\vec{X}_t$ divided by k.

The mean diurnal cycle can give valuable insights in the representation of physical processes in a simulation. Therefore, it is also interesting to average diurnal cycles for e.g., seasons, days with convective precipitation, dry days.

An example diurnal cycle of precipitation in June, July, and August (JJA) 2007 can be seen in Figure 4.1 for the Local Climate Model Intercomparison Project (LocMIP) simulations. Clearly visible is the afternoon precipitation maximum at 5 pm which occurs because of the convective precipitation activities during this time. Especially the Weather Research and Forecasting model (WRF) and the W_CLM3 and B_CLM3 simulations overestimate this maximum notably.

4.2 Temporal Correlation

For regional climate model (RCM) hindcast simulations the temporal correlation gives information about linear relationship of two variables. The mathematical formulation to calculate the temporal correlation coefficient is the same as described for the spatial

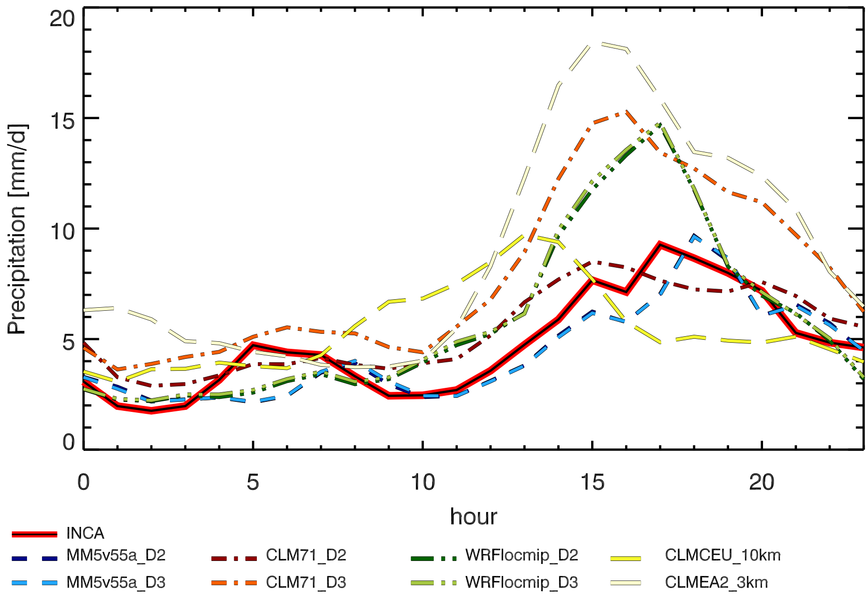


Figure 4.1: Diurnal cycle of JJA precipitation over the Hohe Tauern national park. Shown is the reference dataset INCA, and the eight simulations from the LocMIP experiment.

correlation in Section 3.1. The values of the correlation coefficients are allays between -1 and +1 and therefore well suited to compare the performance of different simulations.

Similar to the spatial correlation evaluations, also the temporal correlation can be displayed with in a Taylor diagram. Example Taylor diagrams for precipitation in three regions in JJA 2007 are displayed in Figure 4.2. In this diagram the statistical values for hourly spatial averaged fields are shown. Clearly visible is, that the correlations in the entire eastern Alpine region (panel a) are generally higher than in the small sub regions of panel b and c. This is because displacement errors of precipitation objects (e.g., convective cells) can be increasingly larger and biases are cancelling out more easily when the spatial averaged domain size enlarges.

A further very informative index which is visible in Taylor diagrams is the temporal standard deviation ratio of the simulated time series divided by those of the reference time series. As can be seen in Figure 4.2 the correlation coefficients are very similar between 3 km simulations and their correspondent 10 km mother simulations. However, the temporal variability in the 3 km COSMO model in CLimate Model (CCLM) simulations, is much higher than in the 10 km simulations which indicates more intensive or more frequent rainfall in the higher resolved simulations.

Instead of averaging the spatial data on an hourly basis, it is also possible to compute

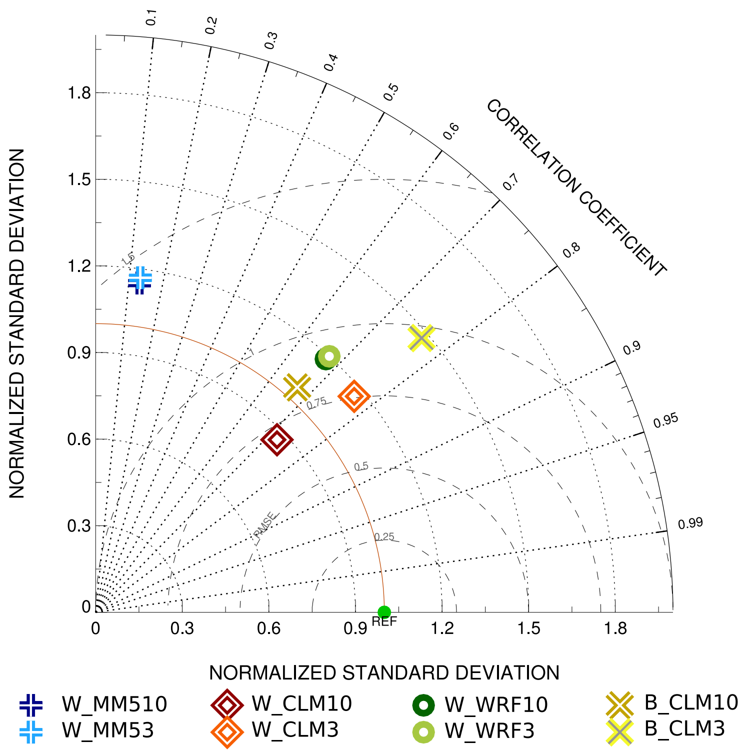


Figure 4.2: Taylor diagram for spatially mean hourly temporal values for precipitation in summer 2007 in the eastern Alps. Different RCMs have different colors and symbols. The 10 km grid spacing simulations have intensive colors, while the 3 km simulations have bright colors.

the temporal correlation coefficients for every grid cell of the considered domain. This method gives information about the temporal performance of a simulation dependent on the location. In Figure 4.3 this is shown for global radiation in December, January, and February (DJF) 2007/08 for the LocMIP simulations above an eastern Alpine domain.

4.3 Cross Correlation

A very common mistake in RCM hindcast simulations are earliness or delays in the time series (e.g., the passage of a front). A useful method to evaluate such errors is the usage of cross correlations. Therefore, the correlation coefficient r is calculated for every possible delay d:

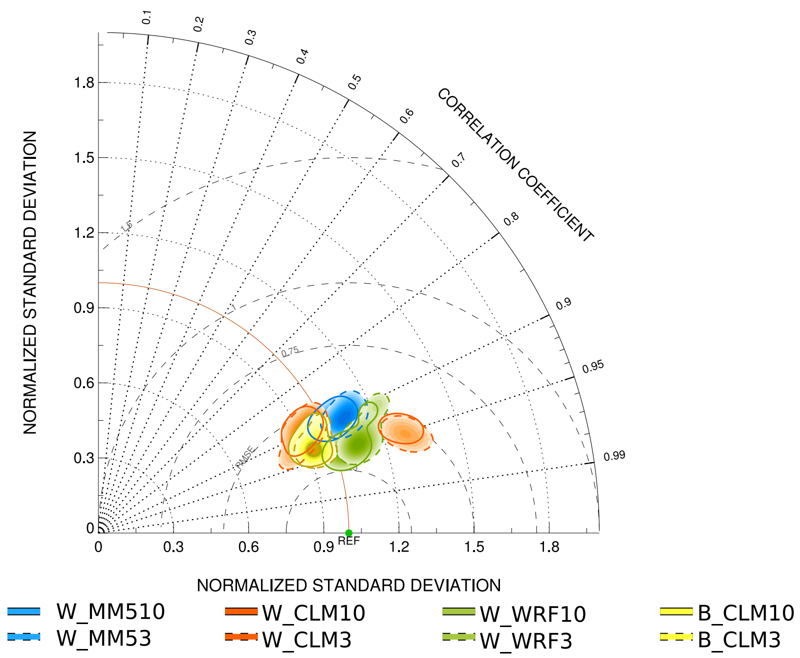


Figure 4.3: Taylor plot which shows the grid point wise temporal values of the hourly fields of global radiation in winter 2007/08 from the LocMIP simulations in the eastern Alpine region as contours. Different RCMs have different colors. The brightness of the color represents the density of data points in this area. 10 km grid spacing simulations have solid lines around the outermost contour while 3 km simulations have dashed lines.

$$r_d = \frac{\sum_{i=0}^{N} (o_i - \bar{o})(x_{i-d} - \bar{x})}{\sqrt{\sum_{i=0}^{N} (o_i - \bar{o})^2} \sqrt{\sum_{i=1}^{N} (x_{i-d} - \bar{x})^2}}, \tag{4.1}$$

where the variables are the same as in Equation 3.1. r_d is computed for all delays $d = 0, 1, 2, \ldots, N-1$ and the result is a cross correlation series which is twice as long as the original series. Most commonly indexes in the series which are lower than zero or greater than N ($i - d < 0$ or $i - d \geq N$) are treated as zero or ignored.

In the application on delays in time series between the reference and simulated atmospheric parameters it is not necessary to compute the entire possible range between

$d = 0, 1, 2, \ldots, N-1$ because only the same physical phenomena should be linked to each other. Therefore, it is sufficient to calculate cross correlations only between a time slice of a few hours.

A limitation of this method is, that temporally high resolved simulation and observation data (time slices should be at least equal or smaller than one hour) have to be used to get meaningful information about delays.

4.4 Evaluating Timing Errors With SAL

A timing error verification method for precipitation was proposed by Zimmer and Wernli (2009). They suggested, to use the structure, amplitude, and location (SAL) method (see Section 3.2.2) which does not need exact matches between the simulation and observed fields. The minimum location component L is searched by computing it for the t=[-3,-2,...,2,3] hourly time slices in the simulation. The time slice t where L gets minimum is then considered as the timing error of the model.

4.5 Frequency Decomposition

Frequency decompositions can similarly to its application in spatial verification (see Section 3.2.1) also be used in temporal verification. Spatial scale separation methods like the discrete cosine transformation (DCT) show that there is more variability in small spatial scales in simulations with higher grid spacings than in simulations with coarse grid spacing (e.g., see Figure 3.5). Similar to this approach, here it should be investigated if simulations with fine grid spacing have more variability in short time ranges than simulations with coarse grids since fine gridded simulations are able to resolve phenomena with smaller scale than coarse gridded simulations.

For this reason a discrete Fourier Transformation is applied to the time series of each grid point. The mathematical formulation looks as follows:

$$X(k) = \sum_{n=-\infty}^{N-1} x(n) e^{-jk\frac{2\pi}{N}}, k = 0, 1, 2, \ldots, N-1. \tag{4.2}$$

In Equation 4.2 $X(k)$ are the discrete Fourier coefficients, n is the time variable, and k is the index of the frequencies. For real numbers $x(n)$ the discrete Fourier coefficients $X(k)$ are in general complex numbers and can be interpreted as a sum of sinus and cosine functions. In generally we are interested in the absolute value of these coefficients rather than in its compounds because the absolute value shows the total amount of information contained at a certain frequency. The total amount is calculated as the square of the $X(k)$ values and is called the power of the signal. It should be reminded, that the amount of a complex number is the distance of the complex number to the origin. The power of the signal at each frequency is called the power spectrum of the signal.

After calculating the power spectra for each grid point it can be averaged over the whole domain or specific regions of interest. The power spectra of different simulation and those of the reference can then be compared in a simple diagram where on the x-axis the frequencies and on the y-axis the power (most commonly in dB) is shown.

The experience from the LocMIP evaluations showed that there are no differences between e.g., 3 km and 10 km resolution simulations if they are considered on an hourly basis. So for this evaluation technique temporal resolutions of at least lower than one hour have to be considered.

5 Characteristics of Specific Weather and Climate Phenomena

In the former chapters entire time series where considered to analyze the performance of climate simulations. Here single events or a collection of events are considered to evaluate the skill of the models to simulate particularly phenomena in the atmosphere. This approach is especially important for the understanding of atmospheric processes and to find shortcomings in the models and therefore support model development.

5.1 Extreme Events

Extreme events are defined statistically as events that deviate strikingly from the statistical mean and are those processes in the atmosphere which have the greatest impact on the society and the environment. Therefore, the demand for accurate predictions of the development of extremes in future climate change is large. Extreme events are often small scale phenomena (e.g., convective cells, wind gusts, tornadoes) which makes the grid spacing of the simulations particularly important. To get statistically significant and meaningful results, the evaluation of extreme events demands long time series in both, the observation and the simulation. In the following section an overview on different methods to evaluate extreme events in climate simulations will be given.

5.1.1 Quantile-Quantile (QQ) Plots

The quantile-quantile (Q-Q) plot is a scatter plot and compares the marginal cumulative probability distribution of a sample (e.g., a simulation) with those of a theoretical distribution or another sample of data (e.g., an observation). Therefore, the empirical quantiles of the observation are plotted against the empirical quantiles of the simulation. If the data points lie on the diagonal, a perfect fit exists between the observed and the simulated distributions.

If extreme values are considered, the quantiles lower than e.g., 10 % or higher than 90 % are most interesting because they provide information about the tails of the distributions.

An example Q-Q plot for observed precipitation compared with an theoretical Gamma and Gaussian distribution can be seen in Figure 5.1.

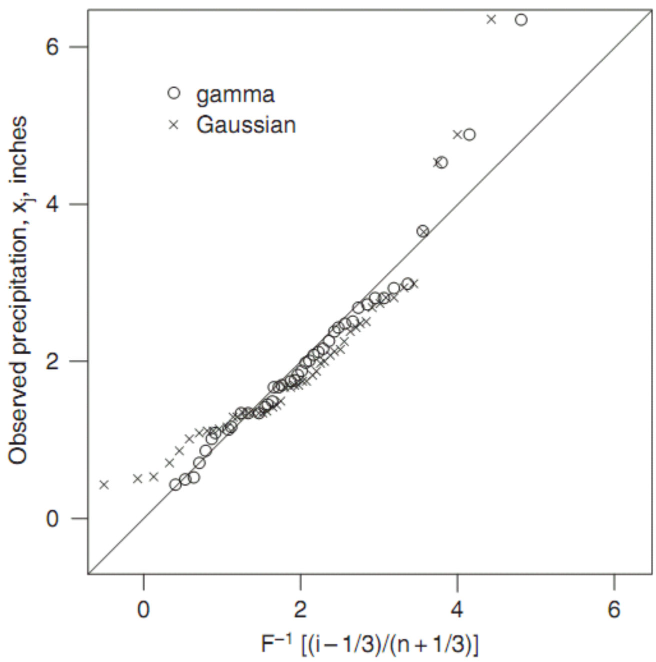


Figure 5.1: Q-Q plots for 1933 – 1982 January precipitation in Ithaca compared with a Gamma (circles) and a Gaussian (crosses) distribution. Observed precipitation amounts are on the y-axis while the amounts from the fitted distributions are displayed on the x-axis (Wilks 2005).

A modification of the Q-Q plot is the probability-probability (P-P) plot which shows the empirical cumulative probability of simulated data compared to reference or theoretical data. P-P plots are less frequently used than Q-Q plots, maybe because comparing dimensional data is more easy to interpret than comparing cumulative probabilities. Furthermore, differences in the extreme tails of the distribution are also less pronounced in the P-P plots. Further information on Q-Q or P-P plots can for example be found in Wilks (2005) or Stevenson (2006).

5.1.2 Extreme-Value Distribution

Extreme values are defined statistically as values that are differing strikingly from the statistical mean. Therefore extreme data are rare and either unusually large or small. A common example for extreme values are a collection of annual maximums or block maxima (the maximum value from a block of m elements). An extreme-value dataset is derived for example by collecting the hottest day of each year out of a time series of 30 years.

In extreme-value statistics, it can be shown that the distribution of such an extreme-value dataset will fit a known distribution increasingly close with increasing m, independent of the distributions of the observation or simulation (e.g., (Leadbetter 1983), (Coles 2001)). This result is called the Extremal Types Theorem or Fisher-Tippett-Theorem (Fisher and Tippett 1928) and is the equivalent in the extreme value theory to the Central Limit Theorem which says that distributions of sums are converging to a Gaussian distribution.

The distribution derived is called general extreme value (GEV) distribution which has the following probability density function (PDF):

$$f(x) = \frac{1}{\beta}\left[1+\frac{\kappa(x-\zeta)}{\beta}\right]^{1-1/\kappa} exp\left\{-\left[1+\frac{\kappa(x-\zeta)}{\beta}\right]^{(}-1/\kappa)\right\}, 1+\kappa(x-\zeta)/\beta > 0. \tag{5.1}$$

In Equation 5.1 ζ is a location parameter, β is a scale parameter, and κ is a shape parameter. By integrating Equation 5.1 the cumulative density function (CDF) is derived:

$$F(x) = exp\left\{-\left[1+\frac{\kappa(x-\zeta)}{\beta}\right]^{(}-1/\kappa)\right\}. \tag{5.2}$$

By inverting Equation 5.2 a formulation for the quantile function can be formulated:

$$F^{-1}(p) = \zeta + \frac{\beta}{\kappa}\left\{[-ln(p)]^{-\kappa}-1\right\}. \tag{5.3}$$

The parameters κ, β, and ζ can be estimated by using the method of maximum likelihood (see e.g., Wilks (2005)) or the L-moments method (Hosking (1990), Stedinger et al. (1993)) which is more often used for small data samples.

There are three special cases of the GEV distribution:

1. Gumbel, or Fisher-Tippett Type I distribution (derived when $\kappa = 0$),
2. Frechet, or Fisher-Tippett Type II distribution (for $\kappa > 0$),
3. Weibull, or Fisher-Tippett Type III distribution (for $\kappa < 0$),

which have different properties. More information about this types can be for example found in Wilks (2005).

Often derived results of extreme value analysis are quantities of large cumulative probabilities like the value of an event with an annual probability of 0.01. However, as long as the number of years n is not very large it is not possible to directly estimate

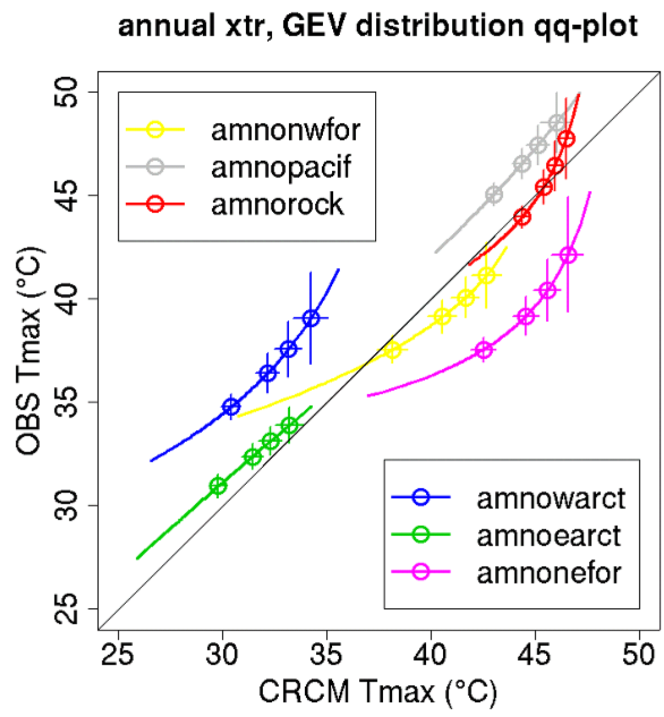


Figure 5.2: Q-Q plots for the GEV distributions of yearly maxima 2 m temperature on daily basis of observational data and simulations between 1961 – 2000 on six different domains over North America (Casati and Lefaivre 2009).

those quantities. Instead a well fitted extreme-value distribution can provide values of probabilities which are larger than $1 - 1/n$.

If extreme minima are considered (e.g., smallest of m observations in n years) the above explained formalism is equally applicable by solving the equations for $-X$.

An example for the evaluation of simulated maxima 2 m temperature on daily basis can be seen in Figure 5.2 (Casati and Lefaivre 2009). With this method it is not only possible to evaluate the performance in different regions (as shown in Figure 5.2) but also to compare the skill of different simulations with each other.

5.1.3 Peaks over Threshold (POT) and Return Periods

Pickands (1975) showed that values of the tail of a distribution are following a general Pareto distribution (GPD) asymptotically if the parent distribution belongs to one of the three above mentioned extreme value distributions. The GPD is expressed as follows:

$$F(x) = 1 - \left(1 + \frac{\kappa(x-\zeta)}{\beta}\right)^{-\frac{1}{\kappa}}, \tag{5.4}$$

and is defined as the GEV distribution by three parameters (location ζ, shape κ, and scale β). After the parameters are fitted to the data, the derived probabilities can be translated into return periods $R(x)$:

$$R(x) = \frac{1}{\omega\,[1 - F(x)]}, \tag{5.5}$$

where ω is the average sampling frequency of the sample time frame ($\omega = 1/n$). The return period is the typical time span in which an event of the magnitude x is expected to occur once. If for example, annual maximum data are considered $\omega = 1\text{yr}^{-1}$ for an event with the cumulative probability $F(x) = 0.99$ which has a probability of $1 - F(x)$ in any given year, the value of this event will be associated with a return period of 100 years and will be called the 100 year event.

Knote et al. (2010) used this method to calculate return values of 2 m temperature for a 20th century control run and a A1B scenario simulation over a domain covering Rhineland-Palatinate and Saarland of Germany as well as parts of Luxemburg (see Figure 5.3).

5.1.4 Extreme Value Indices

Extremes like the annual maximum wind speed are very often modeled with the in Subsection 5.1.2 mentioned GEV distribution. In this subsection extremes are evaluated with index values whereby only few of them can be assumed to follow a GEV distribution.

Here a selection of the most common extreme value indices is presented. A more comprehensive overview of indices can be found on the CCl/CLIVAR/JCOMM Expert Team (ET) on Climate Change Detection and Indices (ETCCDI) homepage (`http://cccma.seos.uvic.ca/ETCCDI/index.shtml`, 15 February, 2011) or on the European Climate Assessment & Dataset homepage (`http://eca.knmi.nl/indicesextremes/indicesdictionary.php`, 15 February, 2011).

Temperature

Multiple temperature indices are defined in literature (e.g., Mekins and Vincent (2005) or Vincent and Mekins (2004)) from which eight indices are selected and presented in Table 5.1. The indices are describing cold events (cold days, frost days, cold nights), warm events (warm days, summer days, warm nights), and the temperature variability (standard deviation of the daily mean temperature, diurnal temperature range). Commonly they are calculated on an annual basis.

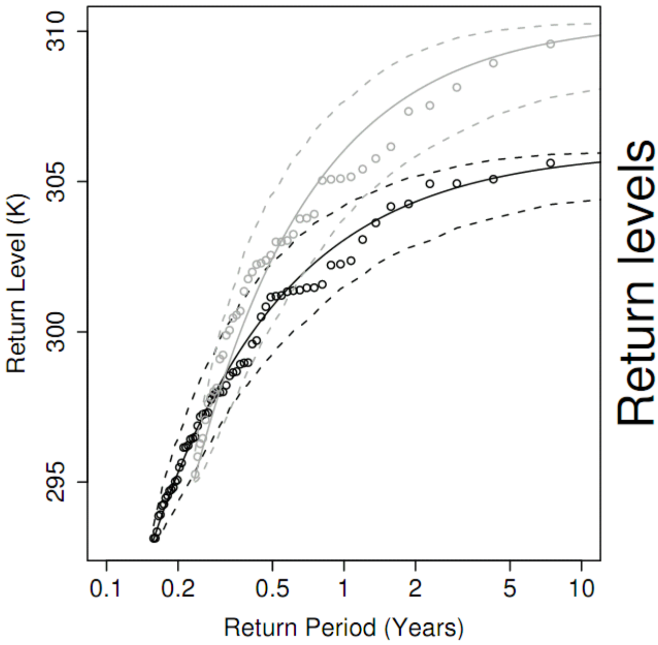


Figure 5.3: Return value versus return periods for maximum 2 m temperature from a control run (black) and the A1B scenario (grey). The solid lines show the fitted distribution while the dots show the real data. Dashed lines are corresponded with 95 % confidence intervals (Knote et al. 2010).

Precipitation

The eight precipitation extreme value indices in Table 5.2 are covering precipitation type, intensity, frequency, and extremes. The annual snowfall ratio and the snow to total precipitation ratio are showing the changes in solid precipitation which is a very important climate characteristic in elevated and high latitude regions.

Cloudiness

Clouds have am major impact on the incoming and outgoing radiation as on the surface energy budget. Therefore, a proper simulation of cloud cover is very important. In Table 5.3 three indices for cloudiness are introduced, which account for mean daily cloud cover and mostly cloudy and mostly sunny days.

Temperature Indices	Definitions	Units
Frost days	Number of days with T_{min} <0 °C	d
Cold days	Number of days with $T_{max} < 10^{th}$ percentile	d
Cold nights	Number of days with $T_{min} < 10^{th}$ percentile	d
Summer days	Number of days with T_{max} >25 °C	d
Warm days	Number of days with $T_{max} > 90^{th}$ percentile	d
Warm nights	Number of days with $T_{min} > 90^{th}$ percentile	d
Diurnal temperature range	Mean of the difference between T_{max} and T_{min}	°C
Standard deviation of T_{mean}	Standard deviation of daily mean temperature from T_{mean}	°C

Table 5.1: Definition of eight temperature indices. T_{max}, T_{min}, and T_{mean} are the daily maximum, minimum, and mean temperatures respectively.

Precipitation Indices	Definitions	Units
Annual snowfall precipitation	Annual accumulated liquid equivalent of snowfall amount	mm
Snow to total precipitation ratio	Annual accumulated snow to total precipitation ratio	%
Days with precipitation	Number of days with precipitation > trace	d
Simple day intensity index of P	Annual total precipitation/number of days with P > trace	mm/d
Max. number of consecutive dry days of P	Max. number of consecutive dry days	d
Highest 5-day precipitation amount	Maximum precipitation sum for 5-day interval	mm
Very wet days ($\geq 95^{th}$ percentile)	Number of days with precipitation $\geq 95^{th}$ percentile	d
Heavy P days ($\geq$ 10 mm)	Number of days with precipitation $\geq$ 10 mm	d

Table 5.2: Definition of eight precipitation indices. P means total precipitation. Trace is meant as the minimum measurable precipitation amount.

Cloudiness Indices	Definitions	Units
Mean of daily cloud cover	Annual average of cloud cover	%
Mostly sunny days	days with cloud cover < 16 %	d
Mostly cloudy days	days with cloud cover > 84 %	d

Table 5.3: Definition of three cloudiness indices.

Cloudiness Indices	Definitions	Units
Max. of non consecutive dry days	longest period within one year without measurable precipitation (<1 mmd)	d
Six Month Standardized Precipitation Index	details can be found in Guttman (1999)	
Potential evaporation	for more details see Allen et al. (1994*b*) and Allen et al. (1994*a*)	mm
Three Month Standardized Precipitation Index	details can be found in Guttman (1999)	

Table 5.4: Definition or references of four drought indices.

5.1.5 Drought

Drought indices are either based on precipitation alone (like the maximum non consecutive dry days index) or are derived from multiple atmospheric- and soil parameters. Droughts have major effects on agriculture and water supply and have therefore significant impact on live quality and the economically development of a region. In Table 5.4 four common drought indices are introduced. A more comprehensive overview of drought indices and their application can for example be found in Heinrich (2008) or Heinrich and Gobiet (2011).

5.2 Weather Events

In this section evaluation methods for model simulations of weather events are discussed. Most of the physical phenomena behind the events have regional or local spatial scales and are therefore hoped to be improved by high horizontal grid spacings.

5.2.1 Inversions

Generally, in an inversion the temperature increases with height. Thereby, inversions can have different origins. Subsidence inversions occur when an air body sinks gradually in the atmosphere and is thereby warmed adiabatically. Capping inversions occur when warm air is moving above cold air so that the convective boundary layer is capped by an inversion layer. Radiation inversions are built when the surface radiation budget is negative. This situation especially appears during nighttime and in winter when the angle of the sun is very low in the sky. If sufficient humidity is present in the cool layer typically fog is present below the inversion cap.

In general, all three types of inversions need atmospheric conditions with low synoptic scale forcing. Since strong solar radiation leads to strong mixing in the lower parts of

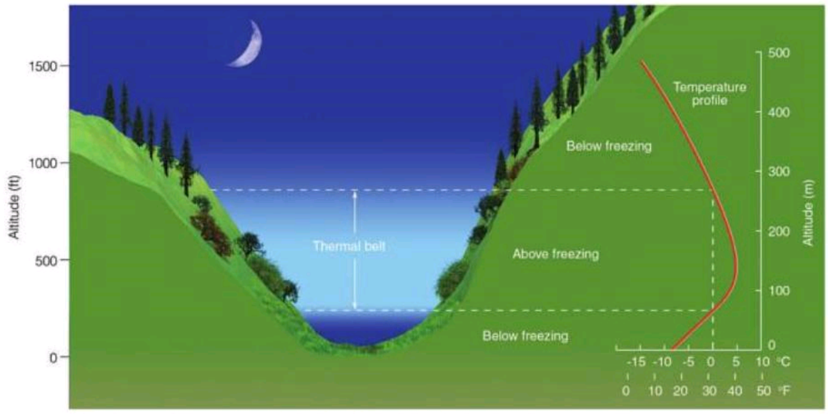


Figure 5.4: Visualization of the settling of cold air during a clear night in a valley. Thereby the bottom of the valley gets colder than its surrounding hillsides (Ahrens 2011).

the atmosphere, inversions are predominant in the winter season.

In hilly or mountainous regions it often occurs that higher elevated areas are reaching out of the inversion layer. If this happens, inversions can be seen in the 2 m temperature field because valleys and basins have then colder temperatures than the surrounding mountain ridges. A possible evaluation method for these cases is to compare the observed and simulated temperature along a cross-section of a mountain slope.

A second, more accurate way to evaluate the simulation of inversions is to use the data of radiosondes. Radiosondes are able to measure air temperature, pressure, humidity, the wind components, and other atmospheric data with high accuracy. The disadvantage of this evaluation method is the high data storage space demand because multiple atmospheric levels have to be saved to compare the simulated parameters with those of the radiosondes.

5.2.2 Spring Late Frosts

Frosts during late spring are a severe problem for agriculture because during cold nights many plants can get damaged (Ahrens 2011). Especially fruit trees during spring, citrus crops, or grapes are affected. Simulations with high horizontal grid spacing are able to capture orographically details better than coarse simulations and have therefore also the potential to improve the representation of frost in valleys (see Figure 5.4).

Scheifinger et al. (2003) uses a method in which they search the latest occurrence date of minimum temperature bellow a number of threshold values for each considered year for a selection of monitoring stations. The derived data can be compared with the model output by interpolating the grid point values to the location of the stations.

5.2.3 Local Wind Systems

In situations with weak synoptic forcing, regional and local wind systems can occur. Thereby the winds are often characterized by local conditions like topography, location, and differential heating. Oliver (2005) suggested a classification of those winds in five groups:

1. Diurnal winds covering all kind of winds which can be associated by the diurnal cycle.

2. Jet-effects including all winds which are related to the local topography.

3. Antitripic winds are those winds which arise from pressure and thermal gradients (e.g., land/sea breezes or foehn winds) or gravity (e.g., fall winds).

4. Local winds created by the overrunning of cold air (e.g., dust storms and haboobs).

5. Winds created by pressure gradients over a relatively small area, and/or the uninterrupted flow over a flat surface (e.g., winds associated with blizzards or the dessert khamsin).

Local wind circulations are often a result of differential heating of the surface. Good examples are see (and lake) breezes and the corresponding land breezes which alternate on a diurnal basis. A diurnal reversal of circulation can also be found in mountain and valley breezes.

For the analysis of such wind systems relatively weak and clear synoptic weather conditions have to be prevalent because the large scale synoptic system can completely obscure local winds. So the first step to analyze local wind systems is to search for stable high pressure systems with weak synoptic scale forcing above the region of interest. Then, the wind fields can be compared between the simulation and monitoring stations within the correspondent region. The near surface wind field can be very variable on small scales because it is strongly influenced by the topography and so a careful selection of representative monitoring stations is essential. If local winds with diurnal characteristics are of interest, high temporal resolution (at least hourly data) in simulated and observed data is required.

A further important feature of local winds is their vertical expansion. Therefore, measurements on different height levels or radiosonde measurements can help to validate the vertical wind structure in simulations.

5.2.4 Foehn

As mentioned in the former subsection, foehn is a local wind which belongs to the group of antitripic winds. It is highlighted in a separate subsection because of its importance for the Alpine region and its special dependency on the orography. A general definition of foehn can be found in the glossary of the American Meteorological Society given by Defant in 1951 (Glickman 2000): Foehn is 'a warm, dry, down slope wind descending the lee side of the Alps as a result of synoptic-scale, cross-barrier flow over the mountain range'. Since this definition could mislead that foehn is only a phenomena in the Alps (there is also 'Zonda' in the south of Chilean Andes ,'Puelche' in Argentina or 'Chinook' in the Rocky Mountains, for example) the World Meteorological Organization has generalized the definition in 1992 to: 'wind (which is) warmed and dried by descent, in general on the lee side of a mountain'.

There are several atmospheric phenomena associated with foehn like wind, temperature, humidity, wave formation, and precipitation. By considering these phenomena, the detection and the assessment of foehn in regional climate model (RCM) simulations get possible.

Wind: Typical for foehn events are strong down slope winds which can get gale intensity and rather weak upstream surface winds. This phenomena is most well known for Boulder Colorado, where its known as 'Bolder Windsorm' (Clark et al. 1994).

Temperature: The main season in which foehn occur is winter, because the development of foehn winds is supported by stable atmospheric conditions. Adiabatic compression causes the warming of the descended air body which often leads to sudden and sometimes large increases in temperature in the valleys. Oard (1993), for example, documented a temperature jump of 27 °C within 7 minutes on the 11 January 1980 at Great Falls Montana.

Relative humidity: Physically linked by the Clausius Clapeyron equation a rise of temperature lead to a decrease of relative humidity. With height the absolute humidity is decreasing and so the decent of absolute dry air with foehn winds can lead to relative humidity values of less than 10 %. Snow covers are rapidly disappearing during periods with foehn, because of the warm temperatures and low humidity values.

Clouds and wave formation: If there is enough humidity, various cloud types can form when the air flows over the mountains. On the windward side of the mountains the cold air can be topped by a flat stratiform cloud. A typical sharp edge of the cloud deck which is above the mountain ridge is called the foehn wall.

Precipitation: Precipitation is commonly present at the windward slopes because of the orographic lifting of the air, especially when the atmospheric layers on the mountain barrier is unstable. Precipitation can also occur on the foehn side of the flow which is then called 'Dimmer Foehn' (Kuhn 1989).

For the evaluation of foehn events in simulations three foehn predictors are especially suitable. The vertical difference in potential temperature $\Delta\Theta$ to indicate the decent of isentropes, the pressure difference between the both sides of the barrier Δp, and the cross-barrier wind component v.

Before the predictors can be calculated, the grid points considered for the calculation of Δp and v have to be chosen. The downstream grid point is selected by searching for the grid point with the maximum cross barrier wind component v. Since on the upstream side blocking and flow over can occur depended on the considered foehn event, v is no suitable parameter for selecting the most suitable upstream grid point. Instead, the cross barrier pressure difference can be used with respect to the already found downstream grid point. The reason for the pressure difference is the decrease of pressure at the downstream side induced by the decent of isentropes nearby the mountain ridge. That's why during foehn the pressure difference Δp is positive. From a theoretical point of view Δp and v are largest nearby the crest because the strongest decent of the isentropes and so the strongest pressure gradients occur there.

For the calculation of $\Delta\Theta$ the two above derived grid points are considered. The choice of the vertical level for the calculation of $\Delta\Theta$ should be done by searching the level where the foehn events can be best distinguished by non foehn events. Therefore the overlapping areas of the PDF of $\Delta\Theta$ between foehn and non foehn events should be minimized. Drechsel (2004) found this level roughly at 300 m above ground level. In an orographic following coordinate system, there will be a height difference between a grid point located at the crest and a downstream grid point and therefore also a difference in $\Delta\Theta$. Positive $\Delta\Theta$ values indicate no or weak decent while neutral or even negative values indicate a decent along the considered model level.

6 Conclusion

Convection resolving climate simulations (CRCS) have a big potential to improve the accuracy of climate projections especially on the regional and local scale through a better representation of orography, the explicit resolving of deep convection, and a more accurate representation of soil atmosphere interactions.

However, finding added value in convection resolving climate simulationss (CRCSs) is difficult because, if spatial scales beyond 10 km horizontal grid spacing are simulated, atmospheric phenomena like precipitation patterns are getting increasingly unpredictable and chaotic. Furthermore, the observation density is often much to coarse to resolve small scale features and reliable high resolved reference datasets for the model evaluation are hard to find.

Therefore, traditional statistical methods often fail in detecting added value of CRCSs. This report gives a guideline on where to search added value, how to find it, and how it can be quantified and displayed. The four introduced categories of added value (mean climate, spatial respectively temporal characteristics, and event based evaluations) help to classify the nature of distinctions between differently resolved simulations.

In the search for added value a best practice is to start searching at fine temporal and spatial scales. Therefore, methods which investigate different scales are often a useful and necessary tools (see Subsection 3.2.1 and Section 4.5). Averaging over time tends to remove small scales, except if strong stationary forcing is prevalent (mountains, coastlines). Particularly in those situations investigations on climatologically time-scales can also reveal potential added value (see Chapter 2). Promising is the detection of added value in temporal and spatial variability statistics. The small grid spacing enables a more accurate simulation of maxima and minima in atmospheric fields and therefore has a high potential to improve the spatial characteristics of atmospheric fields in general (see Chapter 3) and extreme events in particular (see Section 5.1). Also local weather phenomena like inversions, frost, local wind systems and so on which strongly depend on local and regional characteristics hold a lot of promises to contain added value (see Section 5.2).

During the last decades, methods where developed which provide great opportunities for more meaningful evaluations, especially in the field of spatial rainfall analysis. Thereby, each introduced method has a different scope and is able to answer specific types of questions. A general statement of added value in CRCSs can only be derived when comprehensive analyses within all introduced categories are performed.

List of Figures

List of Tables

Bibliography

Åberg, S. et al. (2005). 'An image warping approach to spatio-temporal modelling'. In: *Environmetrics* 16.8 (Dec. 2005), 833–848. See p. 35.

Ahrens, C. D. (2011). *Essentials of Meteorology: An Invitation to the Atmosphere.* 6th ed. Brooks Cole. ISBN: 9780840049339. See pp. 53, 61.

Alexander, G. D. et al. (1999). 'The effect of assimilating rain rates derived from satellites and lightning on forecasts of the 1993 superstorm'. In: *Mon. Weather Rev.* 127.7 (July 1999), 1433–1457. See p. 35.

Allen, R. G. et al. (1994*a*). 'An update for the calculation of reference evapotranspiration'. In: *ICID Bulletin* 43, 35–92. See p. 52.

– (1994*b*). 'An update for the definition of reference evapotranspiration'. In: *ICID Bulletin* 43, 1–34. See p. 52.

Ament, F. et al. (2008). *Results of fuzzy verification methods with COSMO over Swizerland and Germany.* URL: http://www.aemet.es/documentos/es/divulgacion/conferencias/prediccion/Ewglam/VER_MArpagaus.pdf. See p. 19.

Atger, F. (2001). 'Verification of intense precipitation forecasts from single models and ensemble prediction systems'. In: *Nonlinear Proc in Geoph* 8.6, pp. 401–417. DOI: 10.5194/npg-8-401-2001. See pp. 23, 24.

Brooks, H. E. et al. (1998). *Objective limits on forecasting skill of rare events.* See p. 24.

Brown, B. et al. (2009). *Spatial Verification Methods.* See pp. 36, 62.

Casati, B. and L. Lefaivre (2009). *Extreme Value Theory to analyse, validate and improve extreme climate projections.* See pp. 48, 60.

Casati, B. et al. (2004). 'A new intensity-scale approach for the verification of spatial precipitation forecasts'. In: *Meteorol. Appl.* 11.2 (June 2004), 141–154. See pp. 21, 26.

Christensen, J. and O. Christensen (2007). 'A summary of the PRUDENCE model projections of changes in European climate by the end of this century'. In: 81 (0 2007). 10.1007/s10584-006-9210-7, pp. 7–30. See p. 9.

Clark, T. L. et al. (1994). 'Two- and three-dimensional simulations of the 9 January 1989 severe Boulder windstorm: Comparison with observations'. In: *J. Atmos. Sci.* 51, 2317–2343. See p. 55.

Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values.* 1st ed. Springer. See p. 47.

Damrath, U. (2004). *Verification against precipitation observations of a high density network, what did we learn?* See p. 23.

Davis, C. A. et al. (2009). 'The Method for Object-Based Diagnostic Evaluation (MODE) Applied to Numerical Forecasts from the 2005 NSSL/SPC Spring Program'. In: *Weather Forecast.* 24.5 (Oct. 2009), 1252–1267. DOI: {10.1175/2009WAF2222241.1}. See p. 33.

Davis, C. et al. (2006). 'Object-based verification of precipitation forecasts. Part I: Methodology and application to mesoscale rain areas'. In: *Mon. Weather Rev.* 134.7 (July 2006), 1772–1784. See p. 33.

Denis, B. et al. (2002). 'Spectral decomposition of two-dimensional atmospheric fields on limited-area domains using the discrete cosine transform (DCT)'. In: *Mon. Weather Rev.* 130.7 (July 2002), 1812–1829. See pp. 25, 26.

Dickinson, R. E. et al. (1989). 'A regional climate model for western United States'. In: *Clim. Change* 15.3 (Dec. 1989), 383–422. See p. 9.

Dickinson, S. and R. A. Brown (1996). 'A study of near-surface winds in marine cyclones using multiple satellite sensors'. In: *J. Appl. Meteorol.* 35.6 (June 1996), 769–781. See p. 35.

Drechsel, S. (2004). *Objective Forecasting of South Foehn for the Wipp Valley based on ECMWF-T511 Model Data.* 1st ed. Department of Meteorology and Geophysics, University of Innsbruck. See p. 56.

Ebert, E. E. (2002). *Fuzzy verification: Giving partial credit to erroneous forecasts.* See p. 23.

– (2008). 'Fuzzy verification of high-resolution gridded forecasts: a review and proposed framework'. In: *Meteorol. Appl.* 15.1 (Mar. 2008), 51–64. See p. 19.

Ebert, E. E. and J. L. McBride (2000). 'Verification of precipitation in weather systems: determination of systematic errors'. In: *J. Hydrol.* 239.1-4 (Dec. 2000), 179–202. See p. 33.

Everitt, B. S. et al. (2001). *Cluster Analysis.* 4th ed. Wiley. ISBN: 9780470689356. See p. 31.

Fisher, R. A. and L. H. C. Tippett (1928). 'Limiting forms of the frequency distribution of the largest or smallest member of a sample'. In: *PROCEEDINGS OF THE CAMBRIDGE PHILOSOPHICAL SOCIETY* 24 (July 1928), 180–190. ISSN: 0008-1981. See p. 47.

Gebremichael, M. and W. F. Krajewski (2004). 'Assessment of the statistical characterization of small-scale rainfall variability from radar: Analysis of TRMM ground validation datasets'. In: *J. Appl. Meteorol.* 43.8 (Aug. 2004), 1180–1199. See p. 26.

Germann, U. and J. Joss (2001). 'Variograms of radar reflectivity to describe the spatial continuity of Alpine precipitation'. In: *J. Appl. Meteorol.* 40.6, 1042–1059. See p. 26.

Germann, U. and I. Zawadzki (2004). 'Scale dependence of the predictability of precipitation from continental radar images. Part II: Probability forecasts'. In: *J. Appl. Meteorol.* 43.1 (Jan. 2004), 74–89. See p. 24.

Gilleland, E. et al. (2010). 'Analyzing the Image Warp Forecast Verification Method on Precipitation Fields from the ICP'. In: *Weather Forecast.* 25.4 (Aug. 2010), 1249–1262. See pp. 35, 38, 60.

Giorgi, F. and G. T. Bates (1989). 'The climatological skill of a regional model over complex terrain'. In: *Mon. Weather Rev.* 117.11 (Nov. 1989), 2325–2347. See p. 9.

Glickman, T. S. E. (2000). *Glossary of Meteorology. Amer. Meteor. Soc.* 1st ed. Amer. Meteor. Soc. Boston. See p. 55.

Guttman, N. B. (1999). 'Accepting the standardized precipitation index: a calculation algorithm'. In: *J. Amer. Water Resources Assoc.* 35, 311–322. See p. 52.

Harris, D. et al. (2001). 'Multiscale statistical properties of a high-resolution precipitation forecast'. In: *J. Hydrometeorol.* 2.4, 406–418. See p. 26.

Heinrich, G. (2008). *Klimawandel und Trockengefährdung in der Landwirtschaft: Eine Analyse für die Steiermark.* 1st ed. Wegener Center Verlag Graz. ISBN: 13 978-3-9502308-8-8. See p. 52.

Heinrich, G. and A. Gobiet (2011). 'The future of dry and wet spells in Europe: A comprehensive study based on the ENSEMBLES regional climate models'. In: *Int. J. Climatol.* See p. 52.

Hosking, J. R. M. (1990). 'L-moments: analysis and estimation of distributions using linear combinations of order statistics'. In: *J. R. Stat. Soc. Ser. B-Methodol.* 52.1, 105–124. See p. 47.

Kapper, K. L. (2009). *Scale Sensitive Analysis of Regional Climate Models.* Vol. 1. Wegener Center Verlag Graz, p. 114. See p. 26.

Determination of the Effective Resolution of Regional Climate Models (2010). Vol. 12. Geophysical Research Abstracts, p. 1. See pp. 27, 59.

Keil, C. and G. C. Craig (2007). 'A displacement-based error measure applied in a regional ensemble forecasting system'. In: *Mon. Weather Rev.* 135.9 (Sept. 2007), 3248–3259. See p. 34.

– (2009). 'A Displacement and Amplitude Score Employing an Optical Flow Technique'. In: *Weather Forecast.* 24.5 (Oct. 2009), 1297–1308. See pp. 34, 37, 60.

Knote, C. et al. (2010). 'Changes in weather extremes: Assessment of return values using high resolution climate simulations at convection-resolving scale'. In: *Meteorol. Z.* 19.1 (Feb. 2010), 11–23. See pp. 49, 50, 60.

Kuhn, M. (1989). *Föhnstudien.* 1st ed. Wissenschaftliche Buchgesellschaft. See p. 55.

Lack, S. A. et al. (2010). 'An Object-Oriented Multiscale Verification Scheme.' In: *Weather Forecast* 25.1, pp. 79 –92. ISSN: 08828156. See p. 34.

Leadbetter, M. R. (1983). 'Extremes and local Dependence in Stationary-Sequences'. In: *Z Wahrschenilichkeit* 65.2, 291–306. See p. 47.

Linden, P. van der and J. F. B. Mitchel (2009). *ENSEMBLES: Climate Change and its Impacts: Summary of research and results from the ENSEMBLES project.* 1st ed. Met Office Hadley Center. See p. 9.

Reclip:century - a project conducting 21st century regional climate simulation runs focussing on the Greater Alpine Region (2010). Vol. 12. Geophysical Research Abstracts, p. 1. See p. 9.

Mallat, S. (1999). *A Wavelet Tour of Signal Processing, Second Edition (Wavelet Analysis & Its Applications).* 2nd ed. Academic Press. ISBN: 012466606X. See p. 22.

Marzban, C. and S. Sandgathe (2006). 'Cluster analysis for verification of precipitation fields'. In: *Weather Forecast.* 21.5 (Oct. 2006), 824–838. ISSN: 0882-8156. See pp. 31, 32, 60.

– (2009). 'Verification with Variograms'. In: *Weather Forecast.* 24.4 (Aug. 2009), 1102–1120. See pp. 26, 27.

Marzban, C. et al. (2009). 'Three Spatial Verification Techniques: Cluster Analysis, Variogram, and Optical Flow'. In: *WEATHER AND FORECASTING* 24.6 (Dec. 2009), 1457–1471. See p. 28.

Mass, C. F. et al. (2002).
'Does increasing horizontal resolution produce more skillful forecasts? The results of two years of real-time numerical weather prediction over the Pacific northwest'.
In: *Bull. Amer. Meteorol. Soc.* 83.3 (Mar. 2002), 407. See pp. 9, 18.

Precipitation and temperature related climate indices for Canada (2005).
San Diego: California. Am. Meteorol. Soc. See p. 49.

Micheas, A. C. et al. (2007). 'Cell identification and verification of QPF ensembles using shape analysis techniques'. In: *Journal of Hydrology* 343.3-4, pp. 105 –116. ISSN: 0022-1694. DOI: 10.1016/j.jhydrol.2007.05.036. See p. 33.

Murphy, A. H. and R. L. Winkler (1987).
'A General Framework for Forecast verification'. In: *Mon. Weather Rev.* 115.7, 1330–1338. See p. 13.

Nievergelt, Y. (1999). *Wavelets Made Easy.* 2nd ed. Birkhauser. ISBN: 0817640614. See p. 22.

Noel, A. C. C. (1993).
Statistics for Spatial Data (Wiley Series in Probability and Statistics). Rev Sub. Wiley-Interscience. ISBN: 9780471002550. See p. 27.

Oard, M. J. (1993).
'A method for predicting chinook winds east of the Montana Rockies'.
In: *Wea. Forecasting* 8, 166–180. See p. 55.

Oliver, J. E. (2005).
The Encyclopedia of World Climatology (Encyclopedia of Earth Sciences Series). Springer. ISBN: 9781402032646. See p. 54.

Peixoto, J. P. and A. H. Oort (1992). *Physics of climate.*
American Institute of Physics, p. 520. See p. 25.

Pickands, J. (1975). 'Statistical Inference Using Extreme Order Statistics'.
In: *Ann. Statist.* 3.1, 119–131. See p. 48.

Prein, A. F. et al. (2011).
NHCM-1: Non-hydrostatic climate modelling Part 3: Evaluation of the Local Climate Model Intercomparison Project(LocMIP) simulations. See p. 10.

Rezacova, D. et al. (2007).
'A radar-based verification of precipitation forecast for local convective storms'.
In: *Atmos. Res.* 83.2-4 (Feb. 2007). European Conference on Severe Storms (ECSS 2004), Leon, SPAIN, NOV, 2004, 211–224. See p. 25.

Roberts, N. M. and H. W. Lean (2008). 'Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events'.
In: *Mon. Weather Rev.* 136.1 (Jan. 2008), 78–97. See p. 19.

Rummukainen, M. (2010). 'State-of-the-art with regional climate models'. In: *WIREs Clim Chg* 1.1, 82–96. See p. 9.

Scheifinger, H. et al. (2003). 'Trends of spring time frost events and phenological dates in Central Europe'. In: *Theoretical and Applied Climatology* 74 (1 2003). 10.1007/s00704-002-0704-6, pp. 41–51. ISSN: 0177-798X. See p. 53.

Stedinger, J. R. et al. (1993). *Frequency analysis of extreme events. In: D.R.Maidment, ed., Handbook of Hydrology.* 1st ed. McGraw-Hill. See p. 47.

Stevenson, M. (2006). *Forecast verification: A practitioner's guide in atmospheric science.* Vol. 22. 2. PO BOX 211, 1000 AE AMSTERDAM, NETHERLANDS: Elsevier Science BV, 403–405. See pp. 17, 24, 46.

Taylor, K. E. (2001). 'Summarizing multiple aspects of model performance in a single diagram.' In: *J. Geophys. Res.-Atmos.* 106.D7, 7183–7192. See p. 15.

Theis, S. E. et al. (2005). 'Probabilistic precipitation forecasts from a deterministic model: a pragmatic approach'. In: *Meteorol. Appl.* 12.3 (Sept. 2005), 257–268. See p. 24.

Venugopal, V. et al. (2005). 'A new metric for comparing precipitation patterns with an application to ensemble forecasts'. In: *J. Geophys. Res.-Atmos.* 110.D8 (Apr. 2005). See p. 35.

Variations and trends in climate indices for Canada (2004). Seattle, Washington: California. Am. Meteorol. Soc. See p. 49.

Wang, Y. Q. et al. (2004). 'Regional climate modeling: Progress, challenges, and prospects'. In: *J. Meteorol. Soc. Jpn.* 82.6 (Dec. 2004). 2nd Workshop on Regional Climate Modeling for Monson Systems, Yokohama, JAPAN, MAR 04-06, 2003, 1599–1628. ISSN: 0026-1165. See p. 9.

Wernli, H. and M. Sprenger (2007). 'Identification and ERA-15 climatology of potential vorticity streamers and cutoffs near the extratropical tropopause'. In: *J. Atmos. Sci.* 64.5 (May 2007), 1569–1586. ISSN: 0022-4928. DOI: `{10.1175/JAS3912.1}`. See p. 28.

Wernli, H. et al. (2008). 'SAL-A Novel Quality Measure for the Verification of Quantitative Precipitation Forecasts'. In: *Mon. Weather Rev.* 136.11 (Nov. 2008), 4470–4487. DOI: `{10.1175/2008MWR2415.1}`. See pp. 10, 18, 28, 30, 31, 59.

Wilks, D. S. (2005). *Statistical Methods in the Atmospheric Sciences.* 2nd ed. Academic Press. ISBN: 0127519661. See pp. 20, 23, 24, 46, 47, 60.

Zepeda-Arce, J. et al. (2000). 'Space-time rainfall organization and its role in validating quantitative precipitation forecasts'.
In: *J. Geophys. Res.-Atmos.* 105.D8 (Apr. 2000), 10129–10146. See pp. 20, 26.

Zimmer, M. and H. Wernli (2009). *Towards Evaluating Timing Errors of Quantitative Precipitation Forecasts with the Feature-Based Technique SAL.* See p. 43.

Zinner, T. et al. (2008).
'Cb-TRAM: Tracking and monitoring severe convection from onset over rapid development to mature phase using multi-channel Meteosat-8 SEVIRI data'.
In: *Meteorol. Atmos. Phys.* 101.3-4 (Oct. 2008), 191–210. See p. 34.

Abstract:
Convection-resolving climate simulations (CRCS) have high potential to improve multiple error sources in state of the art regional climate models (RCMs) by explicitly resolving deep convection and by representing orography and land cover with high accuracy. However, the added value of CRCSs compared to simulations on coarser grids is difficult to assess by traditional statistical methods. In order to assess added value in a systematic way, we separate four categories (mean climate-, spatial-, temporal-, and event based added value) which can be evaluated separately. For each category several statistical methods are introduced which make different aspects of added value visible. They include traditional statistics like biases, correlation coefficients, or root mean squared errors (RMSEs), but also comprise methods which were especially developed to evaluate highly resolved simulations on grid point basis. The latter methods are of particular interest with regard to convection resolving simulations, since they are suited to analyze the model performance at small spatial scales. A specific focus is also on the evaluation of simulations with regard to small scale extreme events (e.g., heavy convective precipitation) and other local weather and climate phenomena like foehn and inversions. Each introduced method is designed to evaluate and quantify a certain aspect of simulations and sometimes also certain parameters. The presented categories and methods can be used as a guideline for the evaluation of high resolution climate simulations and are basis for the design of a standard evaluation scheme which is used in the Local Climate Model Intercomparison Project (LocMIP), which is described in the third part of this report series.

Zum Inhalt:
Konvektionsauflösende Klimasimulationen (CRCS) haben großes Potenzial die Fehlercharakteristik regionaler Klimamodelle (RCMs) durch explizite Darstellung von hochreichender Konvektion und durch genauere Darstellung von Orographie und Landnutzung zu verbessern. Allerdings ist der Mehrwert von CRCS im Vergleich zu Simulationen auf gröberen Gittern schwer mit herkömmlichen statistischen Methoden zu quantifizieren. Für eine systematische Analyse werden vier Kategorien definiert, die getrennt bewertet werden können. (Mehrwert in: Klimamittel, räumlichem Muster, zeitlicher Abfolge und Darstellung von Einzelereignissen). Für jede Kategorie werden verschiedene statistische Methoden beschrieben, mit denen diverse Aspekte des Mehrwerts sichtbar gemacht werden können. Darunter befinden sich nicht nur traditionelle Methoden wie die Analyse von systematischen Fehlern, Korrelationskoeffizienten oder quadratischen Fehlern (RMSE), sondern auch Verfahren die speziell entwickelt wurden um hochaufgelöste Simulationen auf Basis von Gitterpunkten zu evaluieren. Letztere Methoden sind im Hinblick auf CRCS besonders interessant, da sie geeignet sind die Güte eines Models in kleinen räumlichen Skalen darzustellen und besonders geeignet sind kleinskalige Extremereignisse (z.B. schwere konvektive Niederschläge) und andere lokale Wetter- und Klimaphänomene wie Föhn und Inversionen zu analysieren. Jedes beschriebene Verfahren dient der Evaluierung eines bestimmten Aspekts einer Simulation bzw. teilweise auch der Auswertung spezieller Parameter. Die vorgestellten Kategorien und Methoden sind ein Leitfaden für die Bewertung von CRCS und fungieren als Grundlage für die Gestaltung eines einheitlichen Beurteilungsschemas, welches im lokalen Klimamodell-Vergleichsprojekt (LocMIP) Anwendung findet (siehe Teil drei dieser Berichtsserie).

Wegener Center for Climate and Global Change
University of Graz
Leechgasse 25
A-8010 Graz, Austria
www.wegcenter.at
ISBN 978-3-9502940-6-4