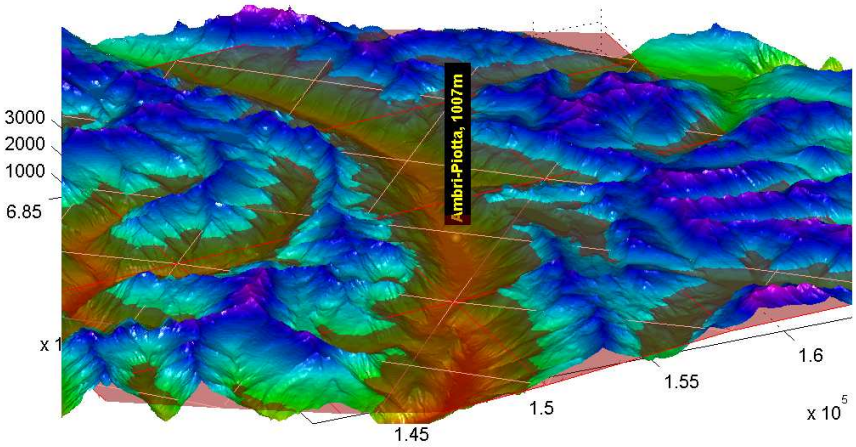
aLMo-Verification

in the

complex terrain of the Swiss Alps



Diploma Thesis by

Daniel Walker

Conducted and written at

MeteoSwiss, Zurich

Submitted to

Department of Earth Sciences, ETH Zurich

Supervisor

Dr. Mathias Rotach, MeteoSwiss

March 2005

The figure, shown on the front page presents a situation found in the Swiss alpine region near Ambri-Piotta. The view angle is along the valley-axis of the "Leventina"-valley. As data source for the "real" topography served the "DHM25"-dataset of swisstopo, schweizerisches Bundesamt fuer Landestopografie (spatial resolution: 25 m). The semi-transparent red surface represents the model-topography, now used in "aLMo" (resolution: 7 km).

Authors adress:

Daniel Walker

Schuetzenstrasse 87

8400 Winterthur

danielwalker@gmx.ch

Abstract

The output 2m-temperatures of a high resolution numerical weather prediction (NWP-) model at specified model grid points were compared with observations taken at nearby surface stations in the Swiss Alpine region. This analysis was performed due to known deficits of NWP-models in complex topography. Due to discretisation and smoothening of the real topography, great differences between the real and model-topography can be found.
In this study a semi-statistic, respectively semi-physical approach was used. On specific days, when especially large differences between modelled and observed 2m-temperatures occurred, other meteorological variables were analysed in order to find reasons for these deficits and to establish a set of criteria for identifying such critical days.

Being able to find these days by using model output variables (e.g. radiation budgets, snow temperatures etc.) a correction of the modelled 2m-temperatures based on mean observed values during such days was performed.
This correction procedure was developed, using the data of four different surface stations. We applied this method at four other stations (control group) in order to test the ability of our approach in being transferable to any other desired location inside highly complex terrain.
The impacts of this correction depended on the station and the specific model grid point. At Samedan, a member of the 'control group' where the effects of the modifications were most pronounced, the Root Mean Square difference (of a full-year dataset) was improved by over 1 K and the Mean Absolute Error by over 0.5 K. The Mean Bias on the other hand was generally less affected and usually indicated a negative impact of the applied modifications.

The influence of using different grid points for the comparison between modelled and observed daily maximum and minimum temperatures was analysed. We found that the difference in altitude between a specific model grid point and the corresponding surface station was more important in order to gain smaller deviations between the two datasets than the horizontal distance (grid point - observation).

Zusammenfassung

In dieser Diplomarbeit wurden die 2m-Temperaturen, eines hochaufgeloesten numerischen Wettermodells mit Werten von Messstationen in den Schweizer Alpen verglichen.
Anstoss fuer diesen Vergleich sind die bekannten Unzulaenglichkeiten von mesoskaligen Wettermodellen in komplexem Terrain. Die Topographie, welche im Modell verwendet wird unterscheidet sich durch Diskretisierung und Filterung enorm von der Realitaet.
In dieser Studie wurde ein semi-statistischer Ansatz gewaehlt. An einzelnen Tagen, wo speziell grosse Differenzen zwischen den vom Modell prognostizierten und den tatsaechlich gemessenen Temperaturen auftraten, wurden andere meteorologische Groessen analysiert. Dadurch konnte ein Satz von Kriterien generiert werden, welcher zur Identifizierung solch kritischer Tage diente.

Diese Identifizierung mittels Strahlungsbilanzen, Schneetemperaturen etc. ermoeglichte es uns die Korrektur der modellierten 2m-Temperaturen, basierend auf den typischen Beobachtungen an solchen Tagen. Dieser Korrekturprozess wurde an Hand von vier Stationen entwickelt und an weiteren vier alpinen Messstationen (Kontrollgruppe) getestet. Dies ermoeglichte die Bestaetigung, dass unser Ansatz nicht auf den jeweiligen Standort beschraenkt, sondern auch auf andere "Gitterpunkt-Messstation"-Paare uebertragbar ist.
Die Qualitaet des entwickelten Korrekturschemas haengt von der jeweiligen Station und des verwendeten Gitterpunktes ab.
In Samedan, einem Mitglied der Testgruppe, war der Einfluss der Korrektur am staerksten. Der Root Mean Square von den Daten eines ganzen Jahres verbesserte sich um 1 K, der Mean Absolute Error um 0.5 K. Der Mean Bias wurde generell weniger durch die Korrektur beeinflusst und zeigte eine Verschlechterung durch die angewendete Korrektur.

Wir untersuchten den Einfluss von unterschiedlichen Gitterpunkten, welche fuer den Vergleich mit den Daten der jeweiligen Messstation verwendet werden konnten. Dazu analysierten wir die Qualitaet der modellierten Tageshoechst-, bzw. Tagestiefst-Temperaturen zu den gemessen Werten an der Station. Es zeigte sich, dass der Hoehenunterschied zwischen der Station und dem verwendeten Gitterpunkt einen groesseren Einfluss hat, als die horizontale Distanz (Gitterpunkt-Messstation).

Contents

Chapter 1

Introduction

This work deals with the influence of a complex topography on the differences between a grid point variable (i.e. temperature on the 2m-level) from a numerical weather prediction (NWP) model and the observations at several meteorological surface stations. The geographic setting, chosen for this comparison, is the alpine region of Switzerland. The model in question is the Swiss high resolution mesoscale model ("aLMo").

For local weather forecasts, usually the nearest gp of the NWP-model is chosen to represent the conditions at a desired location. Even in flat terrains, this is problematic. On the one hand a NWP-model predicts the mean state of a whole grid cell, which covers in this study a total area of about 49 km^2, on the other hand a specific quantity at the specific location is a 'point information'. Because the conditions found at the desired location are strongly influenced by the local situation, for developing a relation between a certain grid point and a target location one should take the local factors into account.

In a complex terrain, these problems are even more pronounced. Especially the topography, soil types, and locally driven circulations are changing over short distances and lead to a variety of micro-climates. They are strongly influenced by the radiation budgets (Matzinger et al. (2003)). Both, the radiation budgets and the state of the atmospheric boundary layer that strongly influence the near surface temperature are difficult to display in the model due to the too coarse spatial resolution that especially in a complex terrain lead to inaccurate estimations.
For instance the short-wave radiation is subject to shading by nearby peaks thus rendering the target location in the shade when at the same time the grid cell in the much smoother topography

of the model is exposed to the sun. This may have a not negligible influence on variables like the 2m-temperature. This situation is shown in Fig. 1.1
The chart chosen as cover-picture on the front page of the present study illustrates the fact that a complex close-to-real topography (spatial resolution: 25 m) is only marginally represented by the model's orography (marked by the semi-transparent red layer).

There are various approaches to relate grid point variables to specific target locations. They usually are based on some sort of statistical analysis between the model output and observations of surface stations. Model output statistics (MOS) and neural networks are examples of such methods. MOS is essentially a multiple linear regression. The aim is to enhance the quality of the direct model output, e.g. temperatures, wind-direction and -speed etc., by means of statistics and therefore to be able to establish more accurate local weather forecast. Forecast variables and past observations serve as predictors, while the weather elements at the surface station are the predictands. Crucial for the quality of MOS is the quality and length of training data used to determine the regression coefficients. For further information about the application of neural networks, we refer to Marzban (2003) and Kretzschmar et al. (2004). The disadvantage of such statistical approaches is that they are optimised for certain cites and therefore are in general not transferable to other locations. Therefore, in this work we try to separate established model-observation differences into physically explainable and random differences. Thus an attempt is made to determine the 'physically accessible' part of these differences (occurrence of snow, radiation budgets, etc.) and suggest a correction procedure. Because such a procedure is based on the physical state, that can be found at any location it would be applicable also to any other grid point-observation pair.

However, improvements in the forecast of local variables, e.g. 2m-temperature in alpine regions are strongly needed. Especially for communities in the alpine region is tourism an important factor and therefore they depend on more accurate weather forecast. An improved forecasting quality is also important in order to enhance the quality of warnings of natural hazards, which are sensitive to different meteorological elements as to the near-surface temperature.
The fact that a major part of the total area of Switzerland lays near or within the alpine region with its mountainous topography reveals, that in these areas a significant lack of accurate forecasts must be found. In the course of this analysis we will see some examples of such wrong estimations of the NWP-model.

In the regular verification of the NWP-model ("aLMo") at MeteoSwiss, a comparison between the observations and the model's output only based on mean values of a duration of at least one month is carried out. However, in this thesis we perform a verification of aLMo at most problematic alpine stations and take a closer look at the model's capability on single days.

Several groups, some quite recently, are engaged in the topic of complex terrain and its missing implementation in most of the mesoscale forecast models (cf. Colette et al. (2003) and Mueller and Scherer (2004)) A theoretical overview which does not claim to be exhaustive about previous studies in this field is presented in "Previous Work" p. 5 et seq.
Nowadays, predicting the micro-weather gains of more importance [1]. An example of local weather forecast is a public system, already established at MeteoSwiss. This application returns a 9-days local forecast for specific ZIP-codes. This product shows the above mentioned deficits especially in the complex terrain of the Alpine region.
(By the time of writing this report, this service was found at:
www.meteoschweiz.ch/de/Prognosen/Vorhersagen/IndexVorhersagen.shtml → "Lokalprognosen Schweiz")

Additionally to the previously described problems like grid spacing and boundary-layer parameterisations, an important task was to investigate how representative 2m-temperatures at a surface station are of a specific model grid point of the model. This problem of evaluating the representativity of such observations at one point for mean values of a whole grid cell is still not entirely solved. Recent efforts in this field have been undertaken by Deng and Stull (2004) and Linder (2005). Deng investigated the spread and analysis of surface data in mountainous terrain while Linder applied this approach (of Deng & Stull) to improve the interpolation of surface temperatures in the complex topography of the Swiss Alps.
In the course of the present study, we investigate the impact of using different kinds of grid points, that differ in the horizontal and/or vertical distance between the grid point and the surface station (cf. Table 4.1 on p. 29).

The general approach adopted in this thesis starts from comparing grid-point/station pairs in order to determine the physically accessible parts of possible differences (see above). On doing this for selected sites and a year-long period (see Data, p. 14) large errors were observed that were readily identified as due to a know (and fixed) mistake in the calculation of the near surface temperature of the model. However, due to a communication problem at MeteoSwiss, the time

[1] "Predicting the micro-weather", The Economist Technology Quarterly, December 6th 2003, p. 22-23

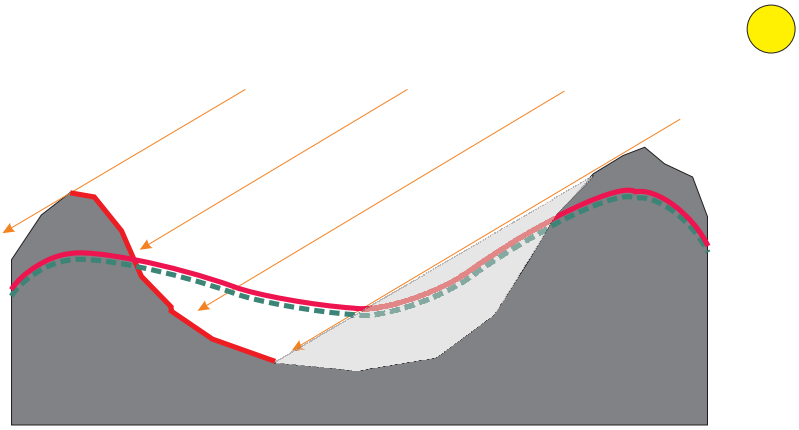


Figure 1.1: A typical situation found in the complex topography of an Alpine region. In the real topography (grey area) only the left part of the valley is exposed to the sun (red line).
A smoothed (model-) topography with a relatively coarse spatial resolution is presented by the dashed green line. Note, that even this representation includes three grid points in this section. In case of the model-topography, the whole section is exposed to direct solar radiation.

period, during which this programming error was 'active' in the model was specified inadequately and therefore the wrong time period was flagged in the data set. Therefore, the difference between model output and point observation erroneously was still dominated by this bug - and the approach developed to detect periods with large model-observation differences and to correct them based on physically accessible information, essentially re-detected this bug. Due to the time constraints for a diploma thesis it was not possible to re-design the entire analysis at the time of realising the mistake. The approach and results of this diploma thesis are hence to be taken as a 'generic' study, in which a possible approach and a 'generic solution' are developed and discussed. The specific results may only be valuable for customers who need long periods of model output for, e.g. climatological purposes of any kind.

If, for a certain reason, a 'bug corrected' 2m-temperature should be more valuable to such a customer than an entire missing period, the suggested correction procedure may be used and the accuracy of the 'corrected temperatures' can be estimated based on the present results. Furthermore, the developed procedure to identify large model-observation differences is applied to a period, when the bug was fixed after all, in order to show that indeed no such period is detected anymore. Hence the procedure may serve as an 'a posteriori justification' or success control of the bug fix that has been realised.

Chapter 2

Previous Work

In this chapter we present several publications dealing with the representation of meteorological conditions in complex terrain by NWP-models and important aspects to be considered in mountainous regions.
There are mainly two different methods found: some groups try to approach this field by using statistical means, others try to find solutions by taking a physical method into account. In the present study, the differences between 2m-temperatures, predicted by the model and observed values at alpine surface station were investigated. Since the near-surface temperatures strongly depend on the radiation budgets, the understanding of these budgets in complex terrain and their representation in the NWP-models is crucial in order to gain improved forecasts (e.g. 2m-temperatures) in alpine regions.

2.1 Approaches based on Physics

Several studies, presented in this section, deal with the radiation distribution and its relevance in NWP-models in complex terrain. While Matzinger et al. (2003) pointed out the great variability of radiation found even in a small area (cross section of a valley), in the work of Mueller and Scherer (2004) it was tried to account for this problem by improving the parameterisations in the NWP-model. Colette et al. (2003) presented a numerical simulation, showing the importance of topographic shading for the conditions found in idealised valleys, as for instance locally driven circulations or the break up of an inversion layer. In the study of Whiteman (1990) a beautiful overview of mountain meteorology, considering thermally developed wind system was presented.

Local circulations found in alpine valleys including the relevance of diurnal thermal- and pressure-gradients between the valley and the preceding plains were explained. As the previously mentioned authors, Whiteman took the different radiation budgets into account and performed a detailed analysis in the complex topography of the "Brush Creek" valley (Colorado). Deng and Stull (2004) investigated objective analysis in order to transform arbitrarily distributed observations into data at regularly spaced grid points. This method, also known as "upscaling" is especially delicate in a mountainous environment. Linder (2005) applied this approach to the complex topography of the Swiss Alpine region and performed a detailed interpolation of the near surface temperature.
In the studies, presented below, usually the radiation and the complex topography were important factors to improve the model's capability, i.e. the model output. The approach of Deng and Stull (2004) includes the difference in altitude between the grid points for instance to determine how the information can spread out in complex topography.
Another physical aspect, we think is important, is the correction of data regarding altitude. Whenever a comparison between modelled and actually measured data is performed, generally a difference in altitude between the grid point, providing the model data and the surface station, taking the measurements is present. In order to compare the two datasets properly, a correction considering this difference in altitude has to be carried out. We assume, that the problem of this fundamental correction can be easily underestimated and is crucial to the quality of a whole study. Simply using a fixed lapse rate as for instance the well known dry- or wet-adiabatic gradient usually does not satisfy the complex and dynamic stratifications, found in alpine valleys. Radiation losses leading to inversions, i.e. "cold pool"-events, or isothermal stratifications formed as a result of melting processes occurring during precipitation events in alpine valleys in a calm atmosphere are only two examples where such a standard correction method does definitely not correspond to the real conditions.
In the present analysis we found that a more distant grid point with a much smaller difference in altitude to the surface station was providing similar or even better results, in case daily minimum and maximum temperatures were considered, than a nearer one, where the altitude correction had more impact, due to a larger difference in altitude and therefore larger corrections (cf. Results, p. 35).

In order to present a brief overview, the aspects presented in these reviews have been chosen regarding the main aspects of our present study. Therefore, depending on the publication, the authors may have investigated also other topics beside the ones presented in the following summaries.

Matzinger et al. (2003) described in detail the influence of the topographic shading based on field measurements in a valley (Riviera Valley, running north-south) in the Swiss Alps. Among other things, they compared the calculated daily cycle of the extraterrestrial irradiance (solar model after Whiteman and Allwine 1986 in Matzinger et al. (2003)) on a horizontal surface with measurements taken at several locations at the valley's cross section (valley floor and several sites at the valley's slopes).
In their charts, a delay in the local sunrise and an advance in the sunset, respectively, of 2 to 3 hours compared with the astronomical sun-cycle could be observed.
Matzinger et al. stated that on the valley walls, the insolation strongly depended on the orientation of the surface to the incoming solar beam. Furthermore, the orientation of the frame of reference (i.e. horizontal or slope parallel) had a great impact on the obtained results.
As expected, by means of their field campaign, the authors showed evidence that the site-to-site contrasts were much less pronounced during overcast days when the direct beam component is absent than during clear sky conditions. Another aspect the model disregards is the local sky view factor (SVF). This limitation is caused by the coarse grid spacing. This factor stands for the hemispherical fraction of the sky (Whiteman et al. 1989, Whiteman et al. 1999, Chapman et al. 2001 in Matzinger et al. (2003)) and ranges from 0 (no sky seen) to 1 (totally flat area without any obstructions in any direction). The SVF has a strong influence on the outgoing long wave radiation and therefore on the overall net all wave radiation which in turn provides the energy responsible for turbulent sensible- and latent-heat exchange processes. A further deficit of the smoothed model topography is the wrong diurnal cycle of the mean albedo within a grid cell. Among other influences, Matzinger et al. described the dependency of the local albedo of the solar incident angle if the reflector is non-isotropic. They showed the different daily cycles of the albedo depending on the location of the measurements at the valley's cross section and the tilt of the frame of reference (horizontal or slope parallel).
The results of this work impressively demonstrate how strongly the orientation of the local topography may influence the radiation components and their budgets, respectively. Since radiation is a very important variable for many processes, they become wrong estimated by the model through

this inadequate model's description of the local topography.

In the work of Mathias Mueller (Mueller and Scherer (2004)) the impact of a complex topography on radiation was investigated. This effect was tested by implementing a new scheme into the Non-hydrostatic Mesoscale Model (NMM). In order to quantify the improvement through the new approach that describes the radiation more accurately, they compared the model output of the 2m-temperatures with hourly mean temperatures of about 400 automatic weather stations. The authors stated, that this approach was adequate because the air temperature is strongly correlated to radiation. This comparison was done once with the new radiation parameterisation and once with a control run without the new scheme.

The newly applied scheme considers the environmental effects on the radiation components in a complex terrain. At the time of submitting their publication, the authors did not know of a great number of mesoscale weather forecast models that already considered these influences. They referred to the Advanced Regional Prediction System (ARPS) where shadow-effects had been recently included and to Colette et al. (2003) who demonstrated the importance of this new implementation.

The scheme of Mueller and Scherer (2004) is mainly based on the application of a digital elevation model (DEM), which has a much higher spatial resolution than the numerical weather prediction (NWP) model. On the basis of this dataset the scheme mainly calculates the required topographic parameters according to Scherer and Parlow (1994) (in Mueller and Scherer (2004)). For their purpose, the authors calculated mean values for each grid cell of the NWP-Model. Further, in combination with the calculation of local sky view factors, the new scheme tested for each grid cell, whether the surrounding terrain allowed a direct insolation. The authors pointed out, that the major advantages of this new scheme are that, it does neither impose any impacts on the performance of the NWP-model, nor on its resolution, nor on the computational costs of the model when operationally conducted. However, they emphasised, that most of the extra computation due to this new scheme can be completed before the model is executed, so that the costs are negligible. They validated the impact of the new scheme on the 2m-temperature during clear sky summer as well as during winter conditions. Furthermore, they also investigated the performance during overcast days. It could be shown, that the impacts of the new parameterisation were due to the reduced sky view and therefore the reduced loss of energy most pronounced in complex terrain during night. Their results also showed a smaller effect of the new parameterisation during daytime.

Even on days with overcast conditions an improvement of the temperature forecast was found. Moreover, the authors discussed an important aspect of the new scheme: namely during daytime, the forecast error was even somewhat enlarged by the new parameterisation. They explained, that this was mainly a verification problem, since the surface stations are usually situated on a flat part of the grid cell (valley ground). Therefore the old unparameterised version, which also assumes a horizontal orientation of the surface, was more accurate for this specific location.
In the present author's view, however, in the total area of a grid cell all kind of orientations of the real topography can be found and it is a fact, that the model's output ought to be representative for the mean conditions in the whole grid rather than the representation of a specific surface station on the valley's ground. Therefore, by taking into account the variability of the different topographic orientations, the new parameterised version should be more representative for the mean state inside a grid cell and not for one (the flat) part of it. Carrying out such a verification is especially difficult due to the fact that the real mean state of a grid cell is not available.

In the publication of Colette et al. (2003), the influence of the valley width and depth, and the impact of topographic shading on the break up of an inversion layer in an idealised steep valley were investigated. The authors referred to three different patterns, stated by Whiteman (1982) (in Colette et al. (2003)) that may cause an inversion-layer break up in steep valleys. The authors explained that not only the heated surface, which let the convective boundary layer grow, destroys the stable layer (from below) but also a circulation, where warmed air ascending the slopes in the morning, can be recirculated and induce a break up of the inversion layer from above. Several charts of numerical simulations revealed how this recirculation with downward winds in the middle of the valley developed during daytime. Whiteman (1982) (in Colette et al. (2003)) observed that a combination of these two patterns, referred to be the third one, occurred in most of his documented cases.

Whiteman (1990) described interesting aspects of the distribution of incoming and outgoing energy in complex terrain, namely the Topographic Amplification Factor (TAF). This concept had been introduced by Wagner (1932), further investigated by Neininger (1982), and extended by Steinacher (1984) (in Whiteman (1990)).

The author explained that the incoming energy to the surface goes through an imaginary horizontal layer set above the real topography. The surface below this layer, that receives this energy, converts it into sensible heat flux and warms the air below this horizontal layer. However, the magnitude of this heating process strongly depends on the real topography below this horizontal layer. If the topography below the layer is not flat but rather mountainous (i.e. a valley), the same amount of energy crosses this imaginary horizontal layer, but the generated sensible heat flux heats a smaller volume of air than over a plain. This smaller volume is the results of the enclosures through the valley's side walls. One consequence is a larger temperature change in the enclosed valley atmosphere than in the atmosphere over a flat terrain. The same pattern can be found overnight: inside the valley, the loss of energy affects a smaller volume of air. Therefore, the cooling is more effective and the temperature drop is more pronounced than over a plain surface.

Whiteman modelled the spatial patterns of the global radiation for the Brush Creek Canyon (Colorado) by means of a digital terrain model with a 30-m grid interval for a day in late September 1984. He analysed several variables such as the distribution of the particular times of the local sunrise and sunset, the slope azimuth and elevation angle etc. Beside these variables he also took a closer look at the sky view factor (SVF). Whiteman explained that the smallest values of this factor, which the incoming long-wave and diffusive radiation depend on, can be found at the bases of the slopes (Petkovsek 1978c in Whiteman (1990)). Furthermore, while approaching the valley centre, the factor grows larger. After Whiteman, the inhomogeneous distribution of the local SVFs causes complicate spatial variations in the downward long-wave and diffusive solar radiation components. He pointed out that the higher the SVF of a location the more exposed it is to the cold radiation of the open sky. On the other hand, the deeper inside the valley the more increased are the effects of the relatively warm radiation of the adjacent side walls.
Whiteman found in his study, that on dry conditions a southwest-facing slope produced an almost equivalent daily total of sensible heat flux as at a ridge top site. Although the elevated site showed the longest daytime period and received the highest daily total of net radiation, the site on the slope was able to compensate for the shorter daytime period by its soil properties (dry, lack of vegetation) and an intensive radiation on the sloping surface in the afternoon. This example shows

that the radiation patterns not only depend on the topographic structure but also on the local soil properties and the exact orientation of the slope relating to the position of the sun.

Deng and Stull (2004) presented a tool for an improvement in objective analysis. Objective analysis, which transforms randomly spaced observed data into data at regularly spaced grid points, is often performed by combining observations and a first guess, calculated by the NWP model. After Deng and Stull (2004) one of the problems is that observations from a surface station at one location of a valley is often not representative of the conditions found up or down valley. Moreover, it is not representative of the conditions on the adjacent slopes either. This inevitably led the authors to the question, whether the first guess of the NWP-model or the measurements of a distant station is more representative in such a situation.
They separated the problem into an intra-valley and inter-valley decorrelation assumption. The authors applied a Gaussian drop off with distance from the observation to the intra-valley assumption while the inter-valley assumption took the differing elevation into account.
To calculate the spreading of temperature characteristics, the authors used a mother - daughter approach, with a so-called sharing factor that determines the fraction of information that can be shared between an observation and an analysis grid point. It is important to note that this approach required a high-resolution NWP-model, not only because the sharing factor is a function of the elevations of the analysis grid points but also because the quality of the analysis in data sparse ridges and valleys strongly depends on the quality of the first-guess of the model.

2.2 Approaches based on Statistics

In the two studies, presented in this section, the authors attempted to improve the quality of the model's prediction, i.e. direct model output of certain variables. In both examples, "Model Output Statistics" and "Neural Networks" a statistical approach was used in order to compensate the model's lack of "knowledge" of the locally determined meteorological conditions found at a specific site.
There is no denying the fact, that using the connection between different variables to determine a better forecast of another variable includes also some kind of physics. Therefore, to be more specific, it would be appropriate to call such approaches *semi-statistical*.

As the name implies, the basic principle of model output statistics (MOS) incorporates the modification of predicted variables of the NWP-model by means of statistics. MOS is essentially a multiple linear regression, where various model forecast variables and if appropriate also past observations are used to improve the quality of a specific variable on a certain location. In order to calculate the coefficients, used for the regression, the system is optimised (minimising the least squares) during a so-called training period. Here lies the weakness of this method. The quality of the whole approach is determined by the length and the quality of the training dataset. Therefore, during the application of the developed system after the training phase, the model should remain unchanged and moreover not be updated or improved. This is a problem, regarding a model used in an operational situation, where the model undergoes frequent changes. (from Kalnay (2003))
However, Hart et al. (2004) applied this concept during a sport event in wintertime. In this study, they compared the strengths and weaknesses of this approach not only with the direct model output of different NWP-Models, that were running on different spatial resolutions, but also with specifically for this region generated "human forecasts".
A neural network is similar to the principles of nonlinear regression. Marzban (2003) describes the differences to the nonlinear regression and the advantages of such a system.

In Marzban (2003) an example of post-processing of temperature forecasts by a neural network (NN) was presented. For this analysis, the Advanced Regional Prediction System (ARPS) was used (to gain more information about the ARPS the author referred to Xue et al. (2000, 2001)). Marzban explained that NNs are a generalisation of traditional statistical methods for nonlinear regression and classification. He pointed out, that there is no qualitative difference in the capabilities between polynomial regression and the application of a NN. The difference between these two approaches is rather a quantitative one, which has its source in the way the dimensionality is handled. A major difference between these two methods lies in the number of free parameters. Marzban explained that this number is crucial for the quality of the method, because the larger this number the greater the risk to overfit the data. While the number of free parameters in a linear regression model grows linearly, the same number in the polynomial regression grows exponentially. However, the author pointed out, that NNs lie between the two described models (linear, polynomial), since NNs are a nonlinear model and their number of free parameters grows linearly. Therefore, by using NNs one can combine the advantages of the linear models (e.g. a smaller number of free parameters) with the ones of the nonlinear models (e.g. capability of fitting

nonlinear data).
In this analysis, the input variables into the neural network were: Forecast hour, model forecast temperature, relative humidity, wind direction and speed, mean sea level pressure, cloud cover, and precipitation rate and amount. The single dependent variable was the observed temperature at the given station. There were 31 stations considered. For each one of those a neural network had been developed.
Marzban found an improvement of the model temperature forecasts in terms of a variety of performance measures, i.e. an average of 40% reduction in mean-squared errors across all stations which was accompanied by an average reduction in bias and variance of 70% and 20%, respectively.

In the study of Hart et al. (2004) the skill of a mesoscale-model-based model output statistics (MOS) system was elevated. For more information about MOS we refer to the Introduction p. 2 or Kalnay (2003). There were 18 sites over northern Utah evaluated, while the 2002 Winter Olympic and Paralympic Games took place.
The study revealed, that in such a complex terrain, an improvement of the model's grid spacing *did not lead* to a significant improvement on the quality of the model's outputs (surface temperature, relative humidity, wind direction, and wind speed). In the course of their work, the authors found substantially better temperature, wind and moisture forecast through the application of MOS than in case of only the direct model output was used. Even cold-pool events, poorly handled by the NWP-model (MM5) were identified by MOS.
In their study it was seen, that the application of MOS lead to a quality of the model's forecasted temperature, relative humidity, and wind speed that was equally or even better than the human-generated forecasts by the Olympic Forecast Team. The authors presented, that statistical techniques represent valuable tools to improve the model's outputs even on these high resolutions.

Chapter 3

Data

In this thesis, 2m-temperatures observed by surface stations (Automatisches Messnetz der Schweiz, "ANETZ") are compared with modelled data of the numerical Swiss weather model (aLMo). The geographic area, chosen for this analysis is located in the complex terrain of the Swiss Alps.

3.1 Time Window

The time frame we apply on our data lays between 1 June 2002 and 31 May 2003.
This choice is based on the following considerations: the need for hourly files of the analysis run and the exceptionally warm summer in 2003. These requirements are best fulfilled in a time window as described above.
Unfortunately, the time frame for this work is centred over a programming error that had most influence during the winter 2002/03. For further information about this problem, we refer to Introduction p. 3 and to Appendix p. 137.

3.2 Observational Data (surface stations)

The selected stations for this study are all situated in alpine valleys with complex topographies and are known to be problematic regarding the verification of a NWP-model. We build two groups of data, separated into a development and a control group. Each group contains the temperature measurements of four surface stations. The development group represents the base of this analysis.

The second group serves as a control instrument:

Development group	Control group
Piotta	Engelberg
Robbia	Comprovasco
Ulrichen	Samedan
Zermatt	Scuol

At the end of this analysis we test our results of the development group against the data of the control group. Table 3.1 gives an overview on the chosen stations, their coordinates and relative positions to the relevant surrounding grid points of the NWP-model.
As criteria for the selection of the stations and the building of pairs served the orientation of the valley axis, the narrowness of the valley and the approximate difference in altitude of the valley floor to a "representative" grid point of the model (definition of this "representative" grid point see Table 4.1 on p. 29).

3.3 Model data

For the comparison with the observed 2m-temperature we use the output of the numerical weather model "aLMo" (alpines Modell). There are two different types of model data, namely the model output from analysis-runs and data from forecast-runs.
In the year 2002, twice a day (00 and 12 UTC), model runs were started that calculated the forecast fields for the next 48 h. Today the same is done, but the forecast ranges to 72 h. At the same time 3h-assimilation cycles were running, producing hourly output files of the analysed fields (from Zala (2002) and De Morsier (2004)). The concept of this assimilation and the used data sources will be explained later in this chapter. However, the output fields of these analysis-runs do account for recent observations and the actual synoptic situation, while the forecast-run lack this data base.
For the present study we use the data from the analysis-runs because we did want to investigate local phenomena and not analyse the differences between modelled and observed data occurring due to a possibly inaccurately modelled synoptic situation. Thus the differences should be at minimum in the analysed fields because the actual synoptical state is merged into the model every 3 hours.

Table 3.1: Overview on the chosen surface stations. The stations are flagged with "D" (development group) and "C" (control group).
The Δl refers to the horizontal distance. α represents the angle (0°: North) from the surface station to the specific grid point. The difference in elevation (Δz) was calculated by subtracting the altitude of the model grid point from the altitude of the surface station.
To calculate (Δl) and (α) between the grid points and the surface stations, equations found in Bundesamt fuer Landestopographie (2001) have been used.
For more information about the different grid points ("Representative" and "Smallest_dz_all25") we refer to Methods, Table 4.1 on p. 29.

Station		CH-Coordinates		Elev.	Axis	Representative			Smallest_dz_all25		
		[m]	[m]	[m asl]		Δl [m]	Δz [m]	α [°]	Δl [m]	Δz [m]	α [°]
Piotta	D	694930	152500	1007	W-E	6489	-830	115	16060	-247	126
Engelberg	C	674150	186060	1035	NW-SE	4830	-507	332	11496	-3	347
Robbia	D	801850	136180	1078	N-S	4426	-1036	150	14035	-220	138
Comprovasco	C	714998	146440	575	N-S	3252	-629	185	3252	-629	185
Ulrichen	D	666740	150760	1345	SW-NE	2328	-845	183	21061	-473	138
Samedan	C	787150	156040	1705	SW-NE	4260	-512	32	20759	-251	325
Zermatt	D	624300	097575	1638	SW-NE	3195	-1163	76	21399	-502	126
Scuol	C	817130	186400	1298	SW-NE	5995	-699	69	15356	-506	53

aLMo is the operational version of the LM (Lokal-Model) at the Federal Office for Meteorology and Climatology (MeteoSwiss). In the following we briefly describe the nature of aLMo, the LM and summarise the parameterisation of the physical processes, which we think are important for the analysis of the 2m-temperature. Wherever relevant, we cite and refer to literature and the model's configuration back at the time when the analysed data was collected (2002/2003).

System Overview of the "Lokal Model" (LM)

The following part describes in detail the configurations of the "Lokal Modell", the different parameterisations, external data fields and the common principles of Data Assimilation. Note that several of these configurations and parameterisations are not valid for the "aLMo", the operational version of the "Lokal Modell" at MeteoSwiss. The "aLMo" specifications are described in a sepa-

rate part.

The limited-area model LM is a numerical weather prediction model on the meso-β and meso-γ scale as well as for various spacings from 50 km down to about 1 km. The model uses hydro-thermodynamical equations, written in the advective form, to describe a compressible non-hydrostatic flow in a moist atmosphere without any scale approximation. In 2002, the LM was driven by the new global model GME of the German Weather Service DWD (information on the GME is summarised in the *Quarterly Report of the Operational NWP-Models of the Deutscher Wetterdienst*, in Doms and Schaettler (2002)).
In case of a four dimensional data assimilation cycle (optional), which is based on a nudging analysis scheme, the initial conditions are created by the continuous LM assimilation and only the boundary data come from the GME forecast run.
(Doms and Schaettler (2002))

Parameterisations:

The quality of the parameterisations of clouds, radiation, the surface layer and the soil processes are important for the magnitude of the deviations between measured and modelled values of the 2m-temperatures. As we will learn in the course of this analysis, especially the way the model can react to a surface covered with snow is crucial for the geographical setting we study in this thesis. Table 3.2 gives an overview on the different parameterisation schemes that where in use during the chosen time period, analysed in this thesis.

External Parameters:

As external parameters we describe fields that are not generated by data assimilation or by interpolation from the driving model. These additional data sets (cf. Table 3.3 on p. 19) contain information about the mean topographical height, the roughness length, soil type, vegetation cover, land fraction, root depth and leaf area index (LAI). Due to the fact that this derivation of the data sets above by a pre-processor is very time consuming, the COSMO group (Consortium for Small-Scale Modelling where MeteoSwiss is one of today's five members) has prepared some predefined data sets with external parameters on three different domains. (Doms and Schaettler (2002))

Table 3.2: LM Model Formation: Physical Parameterisations in Doms and Schaettler (2002), Table 2.

Grid-scale clouds and precipitation:	Cloud water condensation/evaporation by saturation adjustment; precipitation formation by a bulk parameterisation including water vapour, cloud water, rain and snow (scheme HYDOR); rain and snow are treated diagnostically by assuming column equilibrium. Optional: cloud ice scheme
Subgrid-scale clouds:	Subgrid-scale cloudiness is interpreted by an empirical function depending on relative humidity and height. A corresponding cloud water content is also interpreted.
Moist convection:	Mass-flux convection scheme (Tiedtke, 1989) with closure based on moisture convergence. Optional: modified closure based on CAPE.
Radiation:	δ-two stream radiation scheme after Ritter and Geleyn (1992) for short and long wave fluxes; full cloud-radiation feedback.
Vertical diffusion:	Diagnostic K-closure at hierarchy level 2. Optional: a new level 2.5 scheme with prognostic treatment of turbulent kinetic energy; effects of subgrid-scale condensation and evaporation are included. The impact from subgrid-scale thermal circulations is taken into account as well.
Surface layer:	Constant flux layer parameterisation based on the Louis (1979) scheme. Optional: a new surface scheme including a laminar-turbulent roughness layer.
Soil processes:	Soil model according to Jacobsen and Heise (1982) with 2 soil moisture layers and Penman-Monteith transpiration; snow and interception storage are included. Climate values changing monthly (but remain fixed during forecast) in the third layer.

Table 3.3: An overview of the external datasets

Mean orography:	Source is the GTOPO30 data set (30"x30") from the USGS (US Geological Survey)
Prevailing soil type:	Source is the DSM data set (5'x5') of FAO (Food And Agricultural Organisation of the UN)
Land fraction, vegetation cover, root depth and LAI:	Source is the CORINE data set of ETC/LC (European Topic Centre on (CORINE) Land Cover)
Roughness length	Sources are the GTOPO30 and CORINE data sets

Data Assimilation:

A 3-dimensional analysis method does not allow the account for the exact observation time of asynoptic data (note that synoptic scales are mainly determined by the lateral boundary conditions that are received from the steering model (GME of DWD)) and it forces the operators to neglect most of the high-frequent data unless we apply the analysis scheme far more often which itself leads to significantly increased computational costs. On the other hand, a 4-dimensional assimilation (4DVAR) method shows the advantage that it directly includes the model dynamics in the assimilation process, but due to too large costs, this method is not used for the operational application of the LM. Due to these restrictions a scheme based on the observation nudging technique has been developed for defining the atmospheric fields.
Since data performed by the analysis run is used in this thesis, we briefly describe the data assimilation procedure which is applied at the LM and the observational data that are incorporated.
This scheme is based on an experimental nudging analysis scheme that has been developed for the DM (Deutschland Modell of DWD) and the Swiss model version (SM). The new LM-scheme has been adapted to a non-hydrostatic modelling framework. (Doms and Schaettler (2002))

Nudging-Based Assimilation Scheme

Nudging or Newtonian relaxation consists of relaxing the model's prognostic variables towards values within a predetermined time window. In the present scheme, nudging is performed towards direct observations. For this purpose a so called relaxation term is introduced into the model equations.

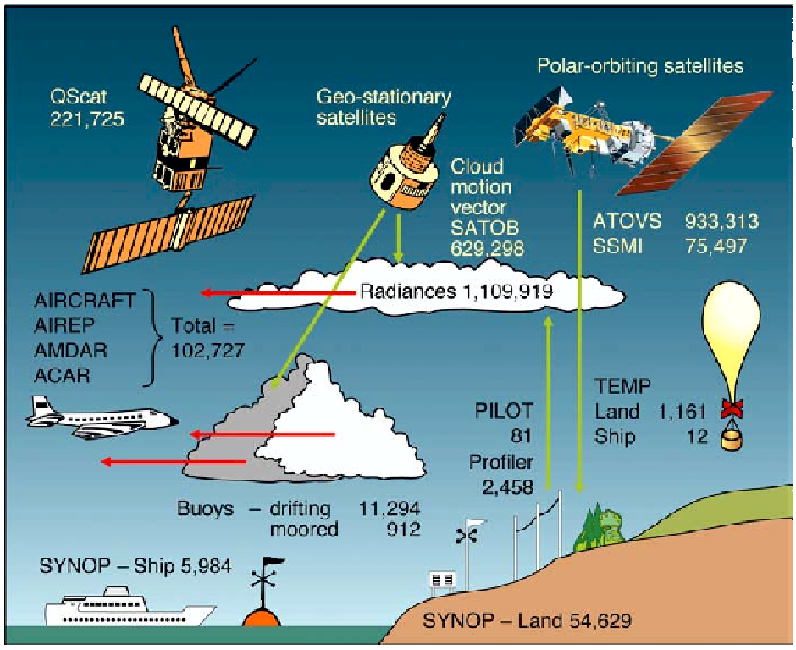


Figure 3.1: Observation used by E.C.M.W.F.'s meteorological operational system. Numbers refer to all data items received over a 24-hour period in May 2003. (Flemming et al. (2004))

For the analysis of the LM, the temperature, wind (on significant levels), humidity (up to 300 hPa, and the geopotential (to derive one pressure increment at the lowest model level) are used. The following parts of the "reports" are considered:
All data of AIRCRAFT and the station pressure, wind (stations below 100 m asl), humidity, and the 2m-temperatures (only for the soil moisture analysis) from the SYNOP, SHIP, and DRIBU reports are taken. After a vertical analysis, the variables are spread laterally along horizontal surfaces (Doms and Schaettler (2002) and Doms and Schaettler (2004) These information were valid during 2002/03 as they were in 2004).
The following listing (Source: Flemming et al. (2004)) presents the different kind of data sources (reports) and shows the incredible amount of information that are available to the "Global Monitoring of Environment and Security (GMES)" initiative at the E.C.M.W.F. In Fig. 3.1 the different observations used by the E.C.M.W.F. for the GMES are shown. (Flemming et al. (2004))

SYNOP, SHIP, DRIBU:

- SYNOP: meteorological weather stations
- SHIP: ships

- DRIBU: drifting buoys
- Observations at or near the Earth's surface are conducted by about 11000 stations, measuring meteorological parameters like atmospheric pressure, wind data, temperature and the relative humidity.
- The ships and buoys are part of the WMO Marine Program. The number of observing ships is about 7000, while approximately 40% are any time at sea. They are measuring the same variables as the surface stations and additionally the sea surface temperature. There are around 750 drifting buoys in this program, providing sea surface temperature and surface air pressure data.

TEMP, PILOT:

- TEMP: radiosondes
- PILOT: pilot balloons (track observed with a theodolite)
- There are approximately 900 upper air stations executing radiosonde ascents, most of them twice a day. In ocean areas, mainly in the North Atlantic, the observations are taken by about 15 ships.

AIRCRAFT:

- Manual reports (AIREPS)
- Automated reports with higher quality (AMDAR, ACAR, ASDAR)
- Providing data about wind and temperature, no humidity
- The total number of aircrafts measuring data like pressure, wind, and temperature during the flight is around 3000. An automated system (AMDAR) carries out high quality reports at selected flight levels as well as during the ascent and descent phases.

In order to determine more accurately the latent and sensible heat fluxes, radiative properties and screen-level temperature (2m-temperature), there are three additional analyses of the sea surface temperature (SST), the snow depth and the soil moisture in use.

The SST is analysed once a day (00 UTC) which is crucial to improve the estimation of the heat fluxes over water. Data are provided from all the ship and buoy observations of the previous 6 days. In data-poor areas additionally satellite data are called in to compensate (NCEP, National Centers for Environmental Prediction).
Knowledge about the snow depth is crucial for an improved conclusion about the radiative absorption and reflection properties of land surfaces and hence the screen-level temperature. The snow water content is a prognostic variable of the model being analysed once every six hours. Where the density of surface stations is sufficient, the observations of SYNOP stations are used. Otherwise the average snow depth increments derived from SYNOP precipitation, temperature, and weather reports as well as the model prediction are additionally applied.
Where the surface is not covered with snow, on days with clear sky conditions, the screen-level temperature and humidity is strongly influenced by the soil water content. An inaccurate specification of soil moisture can lead to major deficiencies in the forecasted temperature. The technique is based on minimising a cost function that depends on the deviations of the forecast temperature from the observed temperature and of the soil moisture from a given background state (created by 15h-forecasts from the previous day by comparing the forecasted and observed temperature). The fact that the 2m-temperature mainly depends on the soil moisture is used as well as it is assumed that (moderate) changes of soil moisture should lead to linear changes in temperature.
(Doms and Schaettler (2002) and Doms and Schaettler (2004))

Specific Information about the Alpine Model "aLMo"

Since April 2001, the LM is the operational system at MeteoSwiss, called "Alpines Modell", aLMo. This operational suite (i.e. a package of about 50 shell scripts), is the Swiss version of the LM, computed at the CSCS (Swiss Center for Scientific Computing) in Manno. The horizontal resolution amounts to 0.0625° that approximately correspond to 7 km. Nevertheless, there are plans to enhance the horizontal grid spacing up to about 2.2 km.
The aLMo domain extends from 35.11 N, 9.33 E (lower left) to 57.03 N, 23.41 E (upper right). In other words, the western boundary of the domain lies in the order of 100 km west of the Portuguese coast and extends in the east to Hungary. In the north-south direction, the domain ranges from Scotland to Sicily.
While in 2002, the driving model was the global model GME (operated at DWD), today the integrated forecast system (IFS) operated at the E.C.M.W.F. is used. During an aLMo-run, a so called LM-Package is processed which has been developed at MeteoSwiss.

Table 3.4 shows and compares the configuration of the aLMo operated at MeteoSwiss in 2002 and 2004.

(Zala (2002) and De Morsier (2004))

Table 3.4: The configuration of the aLMo at MeteoSwiss (Zala (2002) Table 9 and De Morsier (2004) Table 11)

Configuration	**in 2002**	**in 2004**
Domain Size	385 x 325 grid points	385 x 325 grid points
Horizontal Grid Spacing	0.0625°(∼ 7 km)	0.0625°(∼ 7 km)
Number of Layers	**35**, base-state pressure based hybrid	**45**, base-state pressure based hybrid
Time Step and Integrations Scheme	40 sec, 3 time-level split-explicit	40 sec, 3 time-level split-explicit
Forecast Range	**48 h**	**72 h**
Initial Time of Model Runs	00 UTC and 12 UTC	00 UTC and 12 UTC
Lateral Boundary Conditions	Interpolated from **GME** at 1-h intervals	Interpolated from **IFS** at 3-h intervals
Initial State	Nudging data assimilation cycle, no initialisation	Nudging data assimilation cycle, no initialisation
External Analyses	Merging of LM-DWD snow analysis	Merging of LM-DWD snow analysis
Special Features	Use of filtered topography	Use of filtered topography
Model Version Running	lm_f90 2.12	lm_f90 3.5+
Hardware	NEC SX5 (using **8** of **8** processors)	NEC SX5 (using **14** of **16** processors)

Assimilation Cycle in aLMo:

The data assimilation is implemented with 3-hour assimilation runs. Thereby an output file is generated every hour. The source for the observations represents the aLMo data base which is basically a copy of the ECMWF message / report data base. In 2002, during the 06-09 and 18-21h assimilation runs, ozone, vegetation and soil parameters are updated from the GME analysis (we see breaks in the plots at 6, 18UTC). Similarly, the snow analysis of GME at DWD is merged into

the initial conditions of aLMo.
Today, at 00 and 12 UTC the soil parameters are updated from the IFS analysis (Integrated Forecast System, ECMWF). For a more detailed time table of the aLMo assimilation cycle we refer to the Zala (2002), p. 32 and De Morsier (2004), p. 33, respectively.

3.4 Climatological Gradient:

In order to compare the temperature data from the model topography with the observed data measured at a surface station at a generally different altitude, the modelled temperature is vertically corrected with three different gradients.
In addition to the dry and moist adiabatic lapse rate (0.98 K/100m and 0.56 K/100m, respectively), a climatological gradient is applied. Note, Liljequist and Cehak (1984) proposed a dry adiabatic gradient of 0.98 K/100m and a wet adiabatic gradient between 0.5-0.6 K/100m.

The data sources for this calculated climatological lapse rate are homogenised observations of surface stations between 1959 and 2000. Based on this data, mean daily lapse rates have been developed by S. Scherrer within the NCCR project at MeteoSwiss. Observations of a total set of 67 surface station were available. Furthermore, for applications in the more elevated alpine regions, lapse rates have been generated taking into account only stations with elevations higher than 1000 and 1500 m asl, respectively.
For the application in the present analysis, mean lapse rates for each month have been calculated which lie between approximately -0.4 and -0.7 K/100m (cf. Fig. 3.3). Fig. 3.2 presents the mean daily climatological gradients found between 1959 and 2000. In Fig. 3.3 the monthly mean climatological gradients of all 67 surface stations are presented. The same was done for the stations above 1000 and 1500 m asl. The figure shows, that a clear seasonal cycle in the lapse rate can be found which is largest between April and August. The vertical temperature gradient becomes generally larger in case of only alpine station (e.g. elevation larger than 1000 m, and 1500 m asl, respectively) are taken into account. (by courtesy of Simon Scherrer, NCCR, MeteoSwiss)

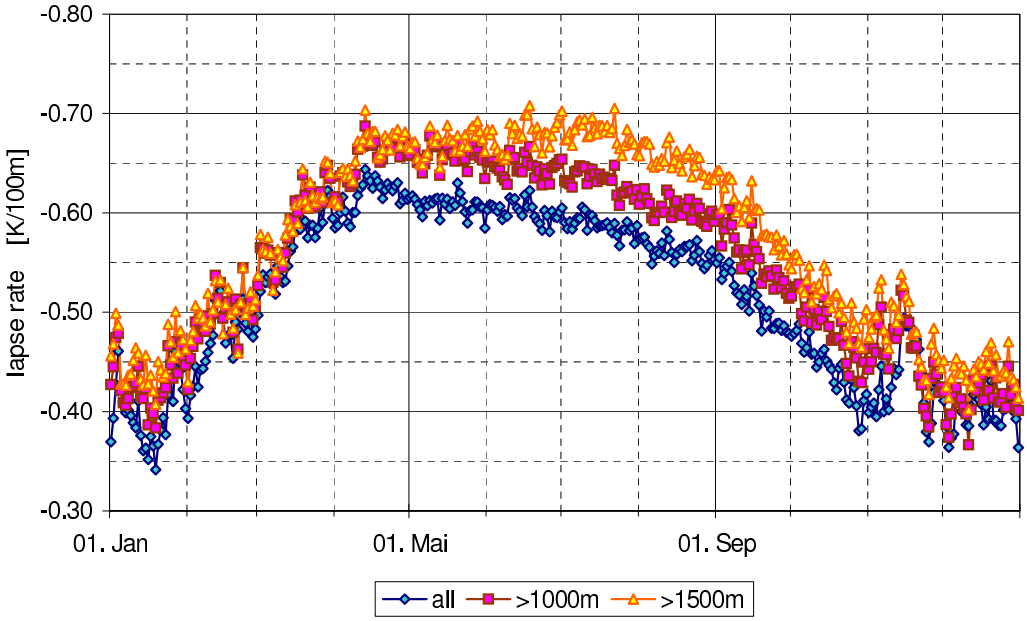


Figure 3.2: The daily mean climatological lapse rates (years: 1959-2000) for stations with the indicated hight constraints. Each data point (symbol) represents an average lapse rate (over all used stations and during the indicated time period) for the respective day.

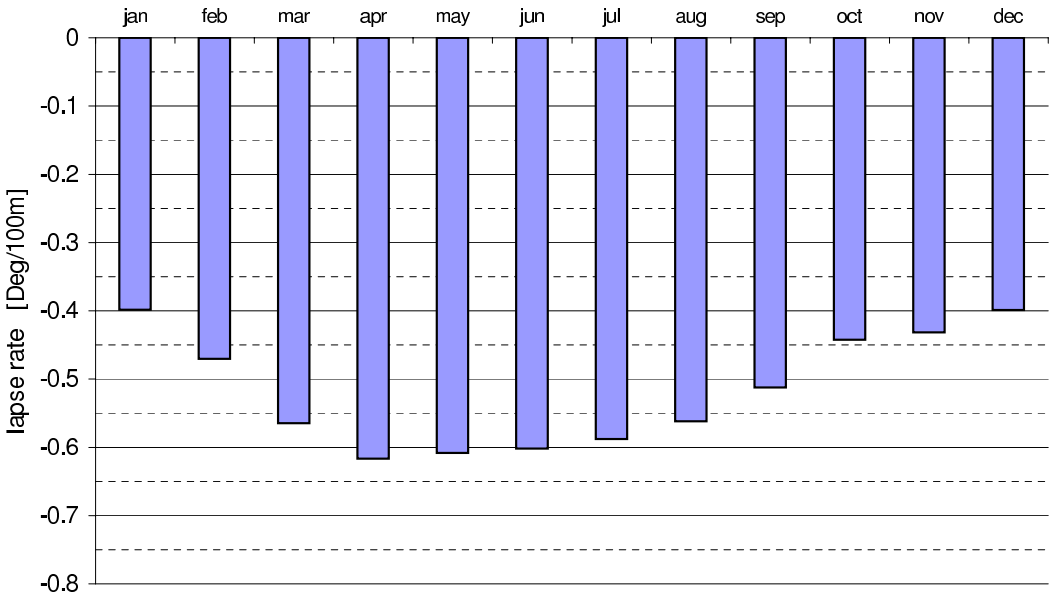


Figure 3.3: The monthly mean climatological gradients of all 67 surface stations (years: 1959-2000), based on the day-to-day information of Fig. 3.2

Chapter 4

Methods

In this chapter we describe how we processed the data. In a first step of this analysis, we simply compared measured 2m-temperatures (hourly resolution) of surface stations with modelled data from nearby grid points. We were interested, whether systematic differences between these two datasets could be found.
In scatter plots, it was possible to validate the representativity of a surface station for a model grid point close to this station. During this process we found on certain days great deviations between modelled and observed data and interesting structures in the already mentioned scatter plots.
It was a main task to develop an algorithm to find these days only by using modelled data and to design a correction method in order to lessen these great differences between the two datasets.
We describe the strategy in finding these specific days when the modelled temperatures behaved completely differently than the observed values at a surface station.

4.1 Starting Point

Figures 4.1 and 4.2 on p. 27/28 present the initial situation by the time we started with the analysis at hand. On each of these plots, the temperatures of the original model output (without any corrections) are compared with the observed data of the nearby surface station. In these examples, the grid point, where we take the model data from, is the one with the smallest difference in altitude between the grid point and the surface station of the 25 grid points surrounding the specific station. We will come back to the significance of this choice and to other possible grid points later in this chapter.

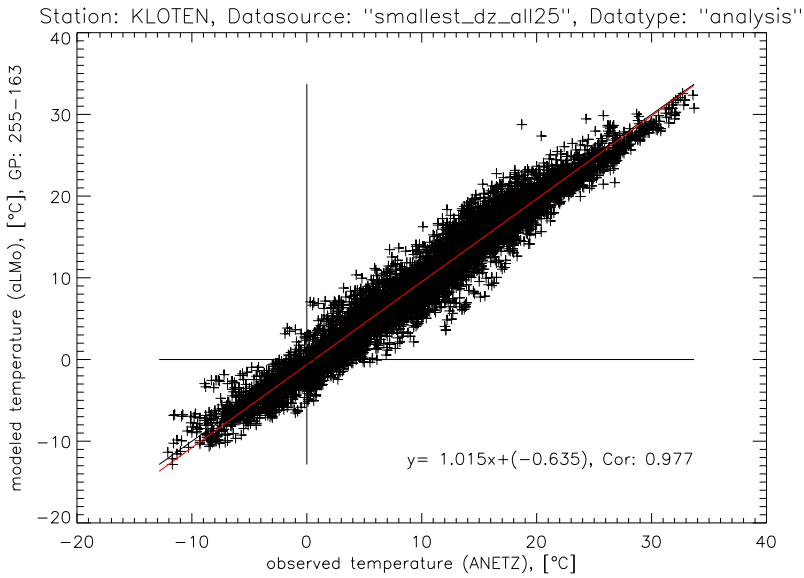


Figure 4.1: The raw full-year data of Kloten is shown. The x-axis is showing the observed temperature at the surface station. The y-axis represents the modelled temperature without any altitude correction. In the lower right corner, the equation of the linear regression and the correlation can be seen.

Note the different structures of the data: while "Kloten" exhibits a regular arrangement of the data points with most of the data close to the 45°-line, "Ulrichen" shows a completely different behaviour. The data points are not as close to the 45°-line, the correlation is much smaller and the interesting horizontal structures did not occur at Kloten.

The comparisons between Alpine stations (e.g. Ulrichen) and surface stations located in the lowlands ("Schweizerisches Mittelland") such as Kloten, aroused our motivation to find reasons for these great deviations. Especially the horizontal structures (cf. Fig. 5.5 and 5.6 on the p. 43/43) will play a major role in this analysis.

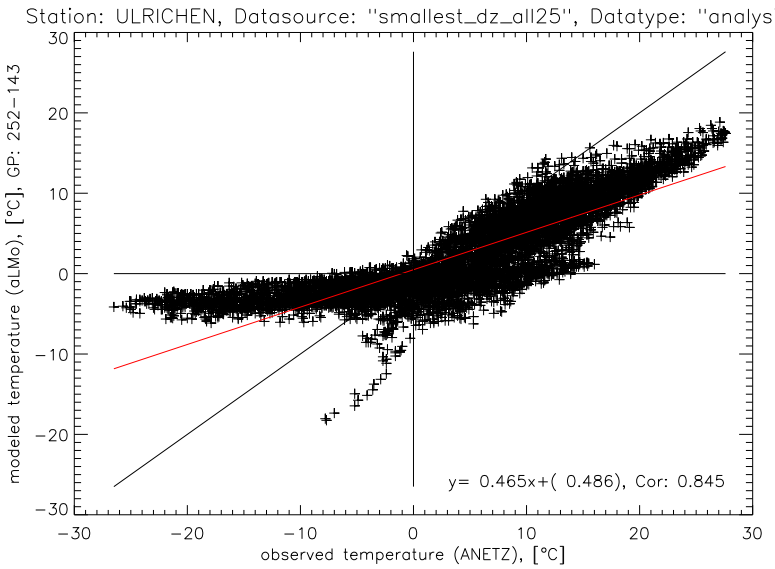


Figure 4.2: The full-year data of Ulrichen is presented. For further remarks we refer to Fig. 4.1.

4.2 Data Sources

Overview

In this study, the term "Data Source" refers to the kind of grid point, where the modelled data had been taken from. From the beginning, it was an interesting aspect of the analysis to determine the model grid point (gp), that is most appropriate for a location in the real topography.
The regular verification of aLMo by MeteoSwiss is undertaken by determining a so-called "representative" gp for every surface station. This is usually the nearest gp to the surface station. In case of a difference in altitude (hereafter referred as dz) of the nearest gp and the surface station of more than 100 m, the gp with the smallest dz of the surrounding four is believed to be most representative (cf. De Morsier (2001)). Additionally to this approach, the present analysis was carried out to a certain point using five different "data sources" that are listed in Table 4.1.

Comparison of the different Data Sources

At this point we were interested whether there was another type of gp (cf. Table 4.1) beside the standard "representative" gp that led to good correspondence between the modelled data and the observations at a certain location.
As criteria, the differences between the modelled and measured daily minimum and maximum, respectively, have been chosen. In "Analysis of the Data Source" p. 35 (Results) the impact of

Table 4.1: Overview on the different Data Sources that can be used

Name:	**Definition:**	**Abbrev.**
nearest	nearest gp to the station	
representative	nearest gp to the station if the difference in altitude between the gp and the station is less than 100 m. Otherwise this gp of the surrounding four with the least difference in altitude is chosen.	"**rep**"
smallest_dz_all25	the gp of the 25 gps next to the station with the least dz	"**min_dz25**"
smallest_dz_near4	the gp of the surrounding 4 gps with the least dz	
mean_near_4	taking the mean data and mean altitude (for the correction) of the 4 nearest gps	
mean_near_2	taking the mean data and mean altitude (for the correction) of the 2 nearest gps	

Obviously, depending on the location, two or even more of the gps above can be identical.

the different data sources is investigated in order to find the best choice when comparing modelled and observed 2m-temperatures in highly complex terrain. This analysis revealed, that the gp with the smallest dz of the 25 gps next to a station (min_dz25) showed a very good correspondence between the model and the observation (cf. Figures 5.1 and 5.2 on p. 36/37). In most cases this was even better than the "representative" gp. The further analysis was then processed by using the "representative" (rep) and "smallest_dz_all25" (min_dz25) gps.

4.3 Analysis of "Nose" / "Tail" Regions and establishing Classes

In Fig. 4.2 clusters of data points are observed, that largely deviate from the 45°-line and appear to be almost horizontal. In other words, for a given, quite large range of observed temperatures, the model seems to predict only a very narrow range of temperatures.
In Figures 5.7 (four single days) and 5.8 (few days in a row) on p. 44/45 examples of days located in such horizontal structures as found in Fig. 4.2 are shown.
These horizontal data clouds could be observed at each station of the development group. Usually, there were two specific regions, that were situated at the cold and warm "end" of the dataset, respectively. For an easier reference, we call these regions "nose" (warm) and "tail" (cold) hereafter. The two regions can easily be found on Figures 5.5 and 5.6 on the p. 43/43.
In order to find all dates, when such weak diurnal cycles of the modelled 2m-temperatures occurred while at the same time, the observation recorded a much stronger daily cycle, we analysed the whole dataset. This led us to the concept of different classes: two main and one intermediate class were created. We looked for two boundaries, that were separating a correct from an erroneous prediction of the model. Several analyses led to the decision that whenever the model is predicting a weak diurnal cycle (< 2 K) and the corresponding measured cycle was also below 2 K, the model's forecast could be assumed to be correct (concerning the diurnal cycle). On the other hand, in a case where the model was predicting the same weak daily cycle while the surface station was recording an amplitude greater than 6 K, the model's output was considered to be inaccurate. Based on these statements the two main-classes "lower than 2" and "greater than 6" were created. Table 5.2 on p. 48 presents the three different classes with their specific definitions.

In summary, the class "lower than 2" (lt_2) represents days, when the weak modelled daily cycles fairly correspond to the observed ones, i.e. both (model and observation) show an amplitude in their diurnal cycle that is lower than 2 K. The class "greater than 6" (gt_6), on the other hand, represents days, when the model again predicts a diurnal cycle of less than 2 K, while the surface station records a diurnal variation in the 2m-temperature of more than 6 K. Therefore, a major wrong estimation of the model occurs on such days.
Keep in mind that the labelling of these classes reflects the behaviour and the amplitude of the diurnal cycle of the observed data. On days of all three classes (lt_2, 2-6, and gt_6) the temperatures, predicted by the model only show a diurnal variation of less then 2 K. Therefore, it was a main task to try to identify days, which belong to the "greater than 6"-class in order to correct

the model's wrong estimations seen in Fig. 5.14 on p. 53.

4.4 Other Meteorological Variables and the development of a Set of Criteria

In the further analysis we tried to find reasons for the great differences between modelled and observed data. Therefore, beside the 2m-temperature, other meteorological variables of the model, like the snow-temperature, cloud cover, dew point, radiation budgets etc. had been analysed. By investigating these variables on days representing a specific main-class (lt_2, gt_6) we found clear differences in these parameters. This enabled us to establish criteria which identify a day of a certain class in advance only by using modelled data.

To verify the quality of these criteria (identifying days of a certain class), we applied them to the same dataset of the development group. Therefore, we were able to compare the days found through the criteria with the ones we previously used to build the specific class.
Applying the criteria on the full-year dataset, did not reveal a complete correspondence of the found days to those dates which had been used to create the specific class. It was no surprise that the better the different criteria for a certain class were chosen, the more identical were the sets of the found days and the ones actually used to define the specific class. Not only the number of days in the two sets were important, but also whether the days were corresponding to each other (cf. Table 5.4 on p. 69).
Afterward, the days, obtained by this approach, were corrected by a mean daily cycle of the observations valid for a specific class, i.e. a mean observed diurnal cycle was added to the almost flat daily cycles (< 2 K) of the model. On this stage, we applied two different kinds of corrections: First, for each class, a correction with the typical individual diurnal cycle of the specific station was applied, and second a mean daily cycle (for each class) of all four stations of the development group (Piotta, Robbia, Ulrichen, and Zermatt). Whereas the former approach is more accurate, the latter represents the only possibility for a correction in case of an absence of a surface station at a certain location.

4.5 Spreading Procedure

Because the daily cycles of the modelled temperatures for the analysed days (lt_2, gt_6) were situated inside a narrow band (cf. Fig. 5.31 on p. 84) while the observed data were much more spread out, we introduced a "spreading procedure". Even when the modelled data were corrected with a mean daily cycle, the deviations were still remarkably large due to this difference in the distribution of the two datasets. Fig. 5.31 shows this narrow band of the corrected model-data.
The "spreading"-procedure is a simple algorithm, which incorporates the multiplication of the deviation of the 00 UTC-temperature of each day to the mean 00 UTC-values with a ratio in order to gain an offset. Adding this offset (its specific value depending on the 00 UTC-deviation of this day) resulted in a much broader distribution of the model-data (cf. Fig. 5.32 on p. 85). The ratio, used in the multiplication, is calculated by dividing the standard deviation of the observed (00 UTC-values) by the standard deviation of the modelled data (00 UTC-values).

4.6 Statistical Quantities:

In the present study we used three different statistical quantities to validate different approaches, such as the ones described above. Moreover they were also applied in order to quantify the improvement of the new modifications described in this study.
Note that in the following set of equations, the subscript "m" refers to "model" and "o" to "observation".

Root Mean Square (RMS)

$$\sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n} ((T_{m,i} - T_{o,i})^2)} \tag{4.1}$$

Root Mean Square Deviation:
The square root of the second moment of a set of observations taken about some arbitrary origin, that is to say, the square root of the Mean Square Deviation or mean square error. The minimum value of the root mean square deviation occurs when the origin coincides with the arithmetic mean - it is then called the Standard Deviation.
(Kendall and Buckland (1971), p. 131)

Mean Absolute Error (MAE)

$$\frac{1}{n} \cdot \sum_{i=1}^{n} (|T_{m,i} - T_{o,i}|) \tag{4.2}$$

Mean Deviation: (*an alternative name, the author used for the MAE*)
A measure of dispersion derived from the average deviation of observations from some central value, such deviations being taken absolutely, i.e. without reference to algebraic sign. The central value may be the arithmetic mean or the median. Expressed formally the mean deviation is the **First Absolute Moment**.
(Kendall and Buckland (1971), p. 91)

Mean Bias (MB)

$$\overline{T_o} - \overline{T_m} \tag{4.3}$$

Correlation Coefficient:

In the present study we often use correlation coefficients (r) in order to compare different approaches. These coefficients were calculated by means of the IDL™ procedure "`correlate.pro`". This function computes the linear Pearson correlation coefficient of two vectors.

4.7 Final Analysis:

In this final analysis, the cognitions we found in the previous work were applied on the stations of the control group (Comprovasco, Engelberg, Samedan, and Scuol).

As in the previous analyses, we performed an altitude correction of the model data to allow a direct comparison between observed and modelled temperatures. In the course of this correction the climatological gradient was used. As a next step, by applying the set of criteria found at the development group, we extracted for each class (lt_2, 2-6, and gt_6) subsets of the full-year dataset. The 2m-temperature data of each class was modified by adding a mean daily cycle to the original diurnal cycle. Note that we processed this analysis three times with varying mean diurnal cycles in order to investigate the sensitivity of the results on these values (cf. Table 4.2). We found, that the three diurnal cycles of the 2m-temperatures and the differences between observed and modelled 2m-temperatures were similar at Piotta, Robbia, and Zermatt while Ulrichen showed a different behaviour (cf. Fig. 5.17 on p. 56. Therefore, it seemed appropriate not only to build a mean diurnal cycle of all four stations of the development group, but rather separate Ulrichen from the rest, i.e. calculate a mean daily cycle for Piotta, Robbia, and Zermatt (Mean_PioRobZer).

Table 4.2: An overview of the different daily cycles that were used in the final analysis

Type:	Definition:
Mean_all4	The mean diurnal cycle of all four stations of the development group (Piotta, Robbia, Ulrichen, and Zermatt)
Mean_PioRobZer	The mean diurnal cycle of Piotta, Robbia, and Zermatt
Ulr	The typical diurnal cycle of Ulrichen

The definitions of these different types of diurnal cycles were the direct result of the visualisation of the diurnal cycles of each station in Fig. 5.16-5.18 on p. 55 et seq.

The last analysis was processed with the spreading procedure and without it, respectively. Note that performing the spreading procedure required the standard deviation of the modelled and the observed data. To spread the data in this final analysis we used the mean standard deviations of the four stations of the development group in order to keep the corrections independent from any knowledge of data, measured at the specific surface station of the control group.

Chapter 5

Results

5.1 Analysis of the Data Source

The term "data source" is used to describe the model grid point that provided the data, we compared with the measurements of the particular surface station. For the comparison of the different kind of data sources (cf. Table 4.1 on p. 29), the mean monthly differences between the predicted daily maximum and the actually measured daily maximum temperature, and the same for the daily minimum, respectively, were analysed.

Figs. 5.1 and 5.2 on p. 36/37 show this analysis of the daily minimum and maximum temperatures, respectively, at Robbia, a member of the development group (Piotta, Robbia, Ulrichen, and Zermatt).

In Fig. 5.1, the monthly mean differences between the modelled and measured daily minimum temperatures are shown. The different bars of each set refer to different data sources (cf. Table 4.1 on p. 29). Each one of the four panels presents the results of a quarter of the year. In case of large differences between the representative ("rep") and smallest_dz_all25 ("min_dz25") gp, the latter led to better results. Only in August and October were slightly worse mean values found at the "min_dz25" than at the "rep" gp. Further, note that generally larger deviations were found during the cooler season (November until May).

Fig. 5.2 presents similar results but this time the monthly mean deviations of the daily maximum temperatures at Robbia are shown. Again, when large differences occurred, the "min_dz25" gp led to better results. Note, that only during July and August the use of the "min_dz25" gp shows slightly worse mean deviations than in case of the "representative" gp.

However, in both analyses (daily min./daily max.) one can see, that beside the "representative"

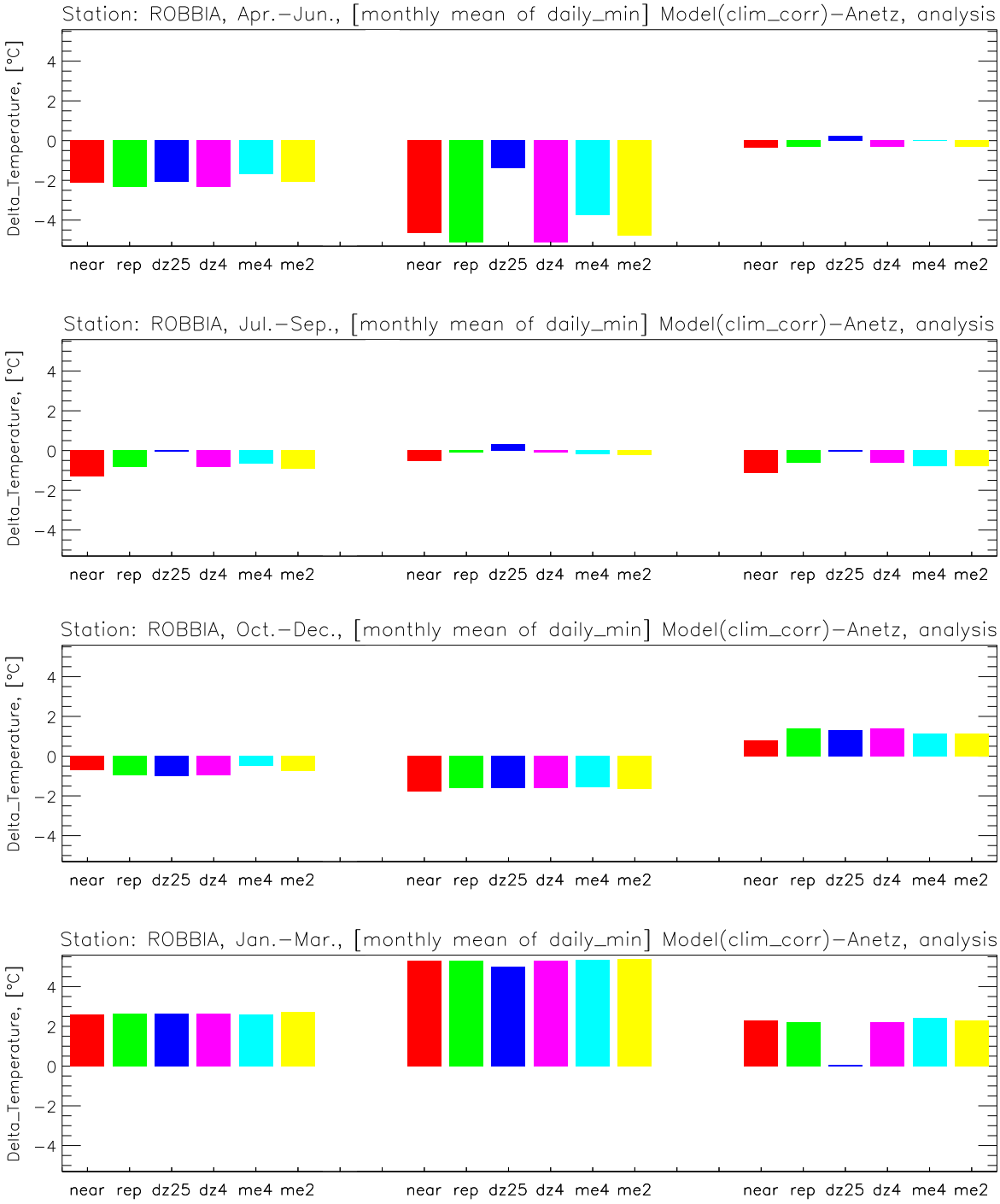


Figure 5.1: Analysis of the different data sources at the location of Robbia. For each month, the mean differences in the model predicted daily minima and the actually measured (surface station) minima are shown. The different classes of the bar plots refer to different data sources.
("near: nearest", "rep: representative", "dz25: smallest_dz_all25", "dz4: smallest_dz_near4", "me4: mean_near_4", "me2: mean_near_2")
First sub-plot: data of April, May, and June. Second sub-plot: data of July, August, and September. Third sub-plot: data of October, November, and December. Fourth sub-plot: data of January, February, and March.

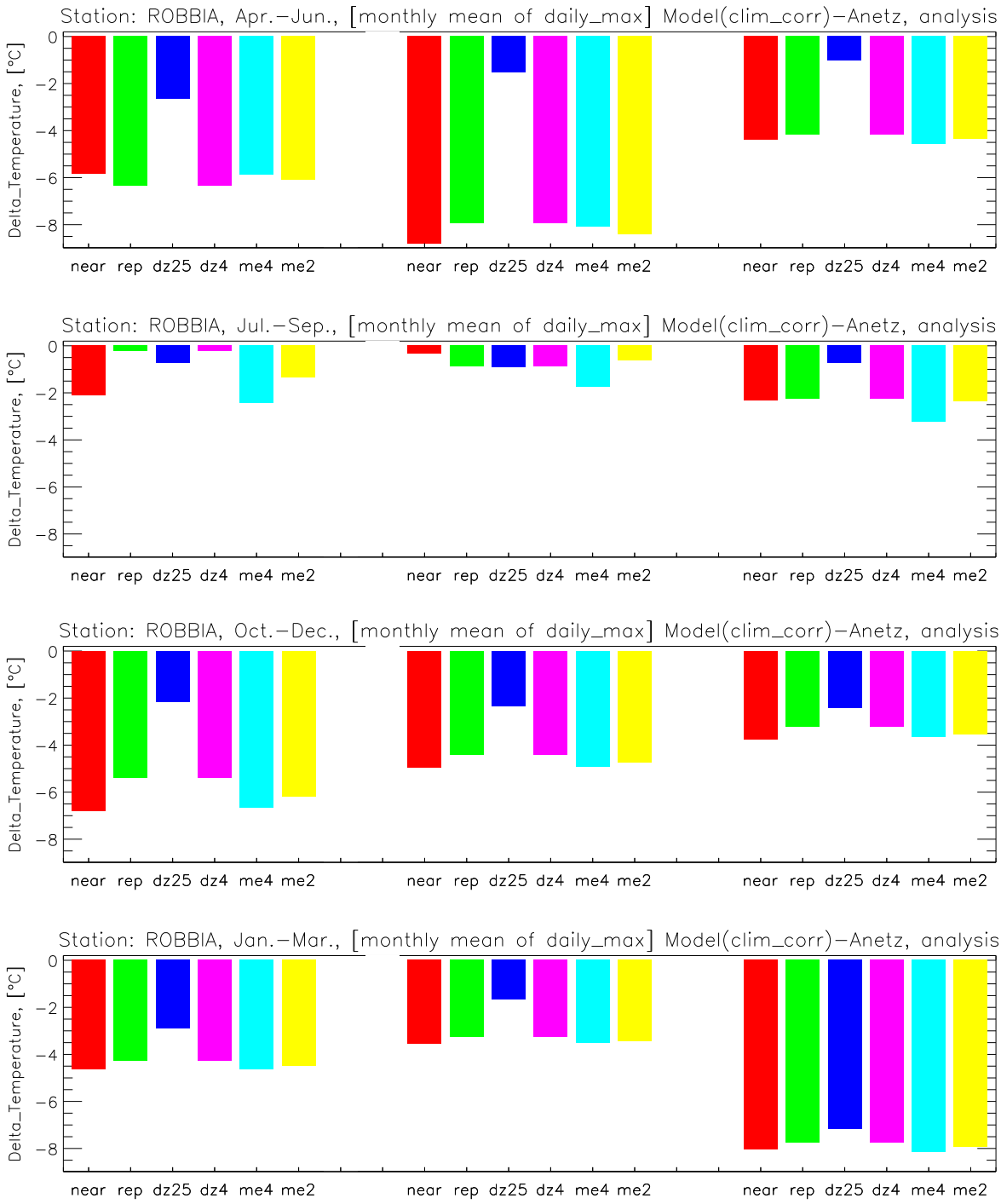


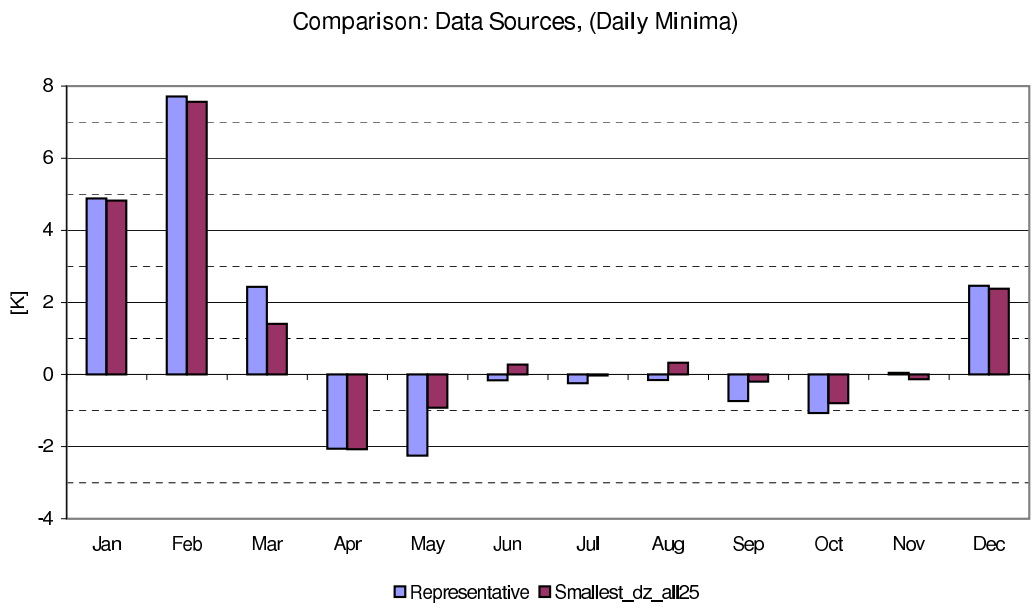
Figure 5.2: Analysis of the different data sources at the location of Robbia. For each month, the mean differences between the model predicted daily maxima and the actually measured (surface station) maxima are shown. The different classes refer to different data sources.
For further remarks we refer to Fig. 5.1.

gp, it was the "smallest_dz_all25" gp showing good results that were often even better than in case the traditional "representative" gp was used (concerning this analysis of the daily minimum and maximum temperature, respectively). This was especially true when the analysis of the daily maximum temperatures was considered.
The Figures 5.3 on p. 39 present the mean results of all four stations of the development group. In the upper panel is the analysis of the daily minimum temperatures shown. In case that only the months with large deviations ($\sim>$ 1 K) were considered (Dec - Mar), a clear improvement of the bias were found when the "min_dz25" gp was used. The lower panel, presenting the analysis of the daily maximum temperatures, shows even more distinctive results. With the exception of March, on all other months an improvement of the bias could be found.
The results in the analysis of the predicted and measured daily maximal temperatures (cf. Fig. 5.3, lower panel) are very clear for the whole year and show a clear improvement (smaller bias) in case the "min_dz25" gp was used. Although the analysed winter 2002/03 was influenced by the programming mistake, leading to wrong near-surface temperatures, the results during the summer 2002 show also a clear improvement. Therefore, we can state, that using the "min_dz25" gp led generally to better results than the "rep" gp.

5.2 Altitude Correction of the Model Data

Fig. 5.4 on p. 41 shows scatter plots of the temperature data at Ulrichen. For this comparison with the data of the surface station, the modelled data was corrected by using three different lapse rates. At the upper scatter plot, the dry lapse rate (0.98 K/100m) was applied, while at the middle and bottom panel the modelled data was corrected by using the wet (0.56 K/100m) and climatological (varying) gradient, (cf. Fig. 3.3 on p. 25), respectively. Liljequist and Cehak (1984) propose a dry-adiabatic gradient of 0.98°C/100m and a wet-adiabatic gradient between 0.5-0.6°C/100m.
The application of the climatological gradient led at all four stations of the development group to a higher correlation between the modelled and observed 2m-temperatures (cf. Table 5.1 on p. 40) and to a slope of the linear regression closer to "1".

The correlation coefficients (cf. Methods, p. 33) for of the two correction methods are shown in Table 5.1. Note that applying a constant gradient, in both cases (dry- and wet-adiabatic lapse rate) led to same correlation coefficients. Therefore, we referred to both of them simply by the abbreviation 'const'. The differences were more pronounced (up to 2.8%) at "rep" than at "min_dz25" ($<$ 1%) where a smaller vertical correction took place. Note, despite of the smaller horizontal dis-



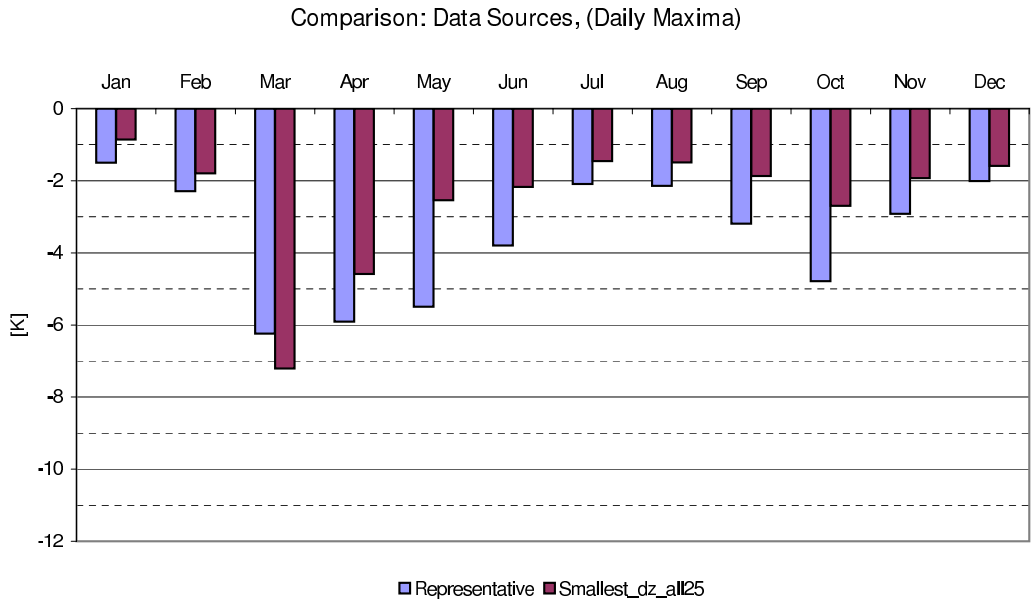


Figure 5.3: The mean monthly deviations of all four stations of the development group (Piotta, Robbia, Ulrichen, and Zermatt). On the upper chart, the analysis of the daily minimum temperatures, on the lower one, the analysis of the daily maximum temperatures is shown.

tances between the "rep"-gps and the observations, more distant gps ("min_dz25", with a smaller difference in the elevations) led to equal or even better correlation coefficients.

Therefore, we decided to use the climatological gradient for all further analyses, presented in this study.

Table 5.1: Correlation Coefficients (cf. Methods, p. 33) between observed and modelled 2m-temperatures. For both data sources, the results are presented for a constant gradient (wet/dry) and a variable gradient (climatological).
"const." refer to a constant wet/dry-gradient, while "clima." refers to (monthly) varying climatological gradients.

	Representative		Smallest_dz_all25	
	Corr. const.	Corr. clima.	Corr. const.	Corr. clima.
Piotta	0.900	0.912	0.920	0.923
Robbia	0.834	0.858	0.910	0.912
Ulrichen	0.836	0.855	0.845	0.855
Zermatt	0.870	0.894	0.897	0.905

5.3 Analysis of the Altitude Corrected Data and Definition of the "Nose" & "Tail" Regions

Creating scatter plots of the modelled vs. observed data, we recognised horizontally shaped regions that are signs for large deviations between the two datasets. Figs. 5.5 and 5.6 on p. 43 show these two regions (blue and red) at "Ulrichen" and "Robbia". Note that at Ulrichen only the most extreme part of the horizontal structure was selected, in order to characterise the most typical days. They occur at each station of the development group with different distinctiveness. For easier referring we call hereafter the red area "nose" and the blue one "tail", respectively. Obviously these two regions with their special shapes are signs for a stronger variability, respectively daily cycle of the observed data than of the modelled data.

Fig. 5.7 on p. 44 presents this behaviour of the modelled and observed 2m-temperatures during four specific days, found in the "nose"-region for "Robbia". Note the discontinuities at 6 and

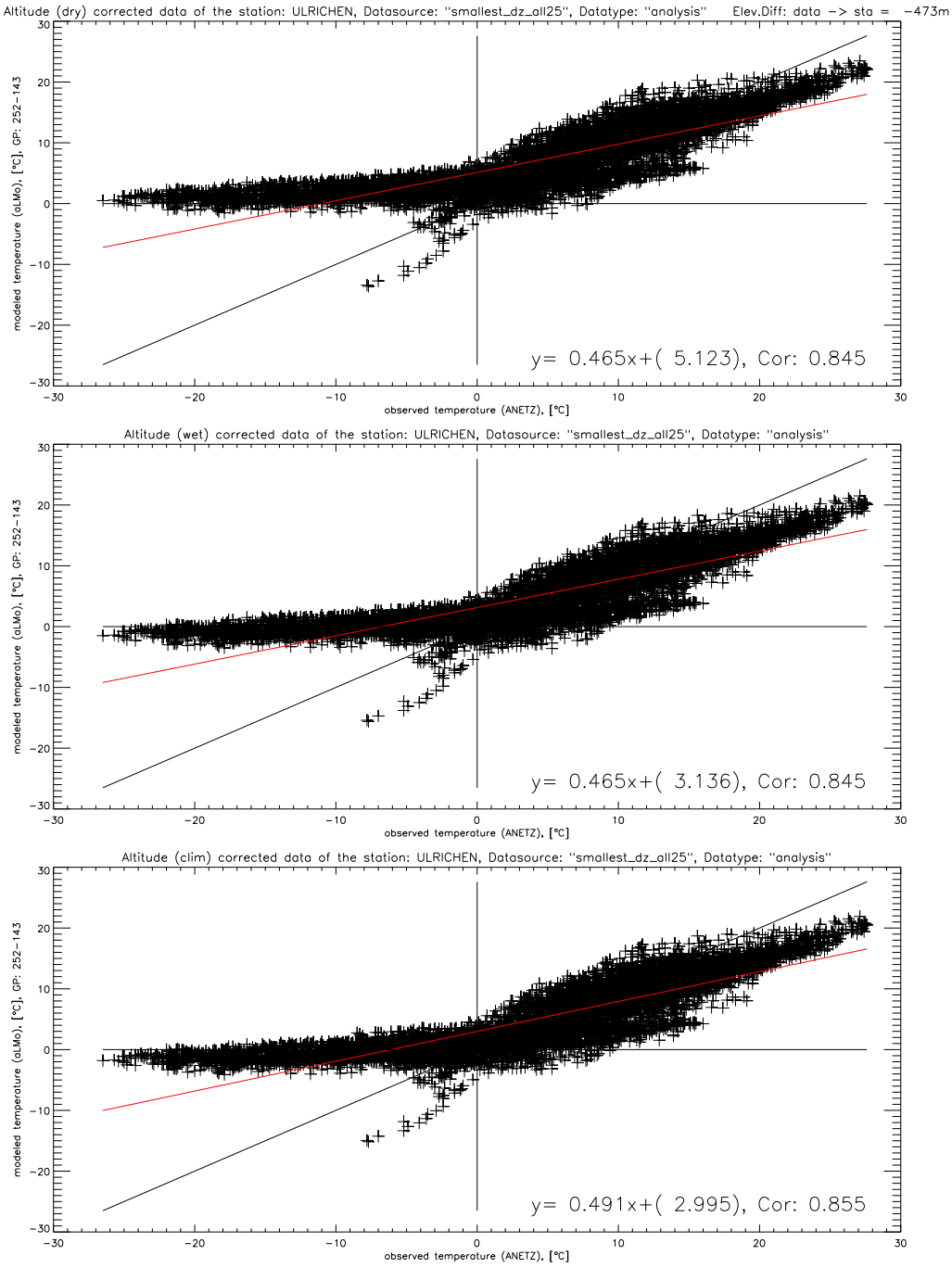


Figure 5.4: On each of the three panels, the full-year dataset of Ulrichen is shown, while on each one a different lapse rate was used to correct the modelled data. The data source for this figure is "smallest_dz_all25". In this example, the difference in altitude between the model grid point and the surface station is -473 m (surface station - model gp). In the x-direction, the observed data, in the y-direction, the modelled data is drawn.
Upper panel: the dry-adiabatic gradient, in the middle one: the wet-adiabatic, lower panel: the climatological gradient.
Note, on each panel, the equation of the linear regression and the correlation coefficient of the two datasets are shown in the lower right corner.

18 UTC and moreover the almost constant time series in the daytime (6-18 UTC) of the modelled data. Such a pattern can also be found during several days in a row, as seen in Fig. 5.8 on p. 45. The upper panel presents the 2m-temperatures during days found in the nose-region (16-30 March 2003), while on the lower one, a time period close to the one above (same season) is plotted but this time the days shown (15-29 April 2003) did not occur in the "nose"-region. Comparing these two plots, we found, that although these days lie relatively close together, the model shows a completely different behaviour.

Fig. 5.9 and 5.10 on p. 46/47 show the typical daily cycle at "Ulrichen" during days found in the "nose"- and "tail"-region, respectively.

On figures like 5.9 and 5.10 we observed, that on specific days the model was not showing any or only a very weak diurnal cycle in the 2m-temperature while the observation at the same time recorded a much stronger one. This was especially true on days found in the "tail"-region (Fig. 5.10). However, similar patterns were found for each station of the development group.

Further we saw at all stations, data sources ("rep", "min_dz25") and region-types ("nose", "tail") discontinuities occurring at 6 and 18 UTC (cf. Fig. 5.9). This included, depending on the station, an immediate drop or rise in the temperature signals of several Centigrades. These irregularities were caused by external data that entered unfiltered into the model's field. Actually, at 6 and 18 UTC, the DWD's analysis updated the soil parameters, including snow temperature and snow water content, several soil temperatures and soil moistures on different layers (cf. Zala (2002)). At present, these parameters are assimilated using the IFS's analysis (ECMWF) in the current version of "aLMo".

These large model-observation differences motivated us to find all dates in the full-year dataset when such a weak daily cycle in the modelled 2m-temperature could be found and to compare them with the observed data. Due to sometimes even large discontinuities after 6 and after 18 UTC, we used the amplitude between 7 and 18 UTC as criterion for a "weak daily cycle" in the model's 2m-temperature. A further side effect of concentrating the search on this part of the day (7-18 UTC) was, that we were able to find especially these dates, when no warming during daytime took place and ignored a possible cooling during the night.

5.4 Definition of different Classes

Fig. 5.11 on p. 49 shows the full-year temperature data of Robbia. In the three different panels, varying amplitudes in the daily cycles of the model's temperature are marked in colour. As one can see, using a daily cycle of the model's 2m-temperature of less than 2 K, we included most

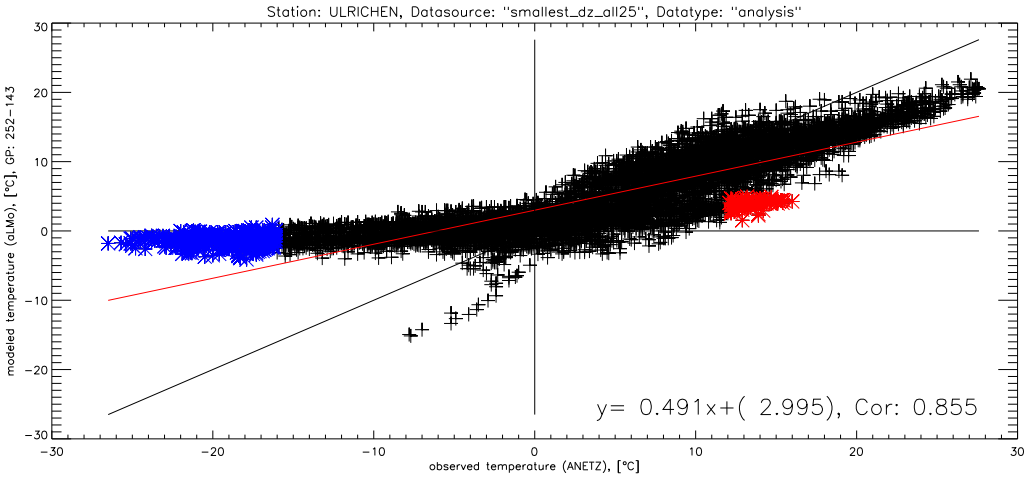


Figure 5.5: The full-year dataset of Ulrichen (hourly resolution). The x-axis is showing the observed 2m-temperatures (surface station). The y-axis presents the modelled 2m-temperatures. Note, that the modelled data were altitude-corrected (climatological gradient).

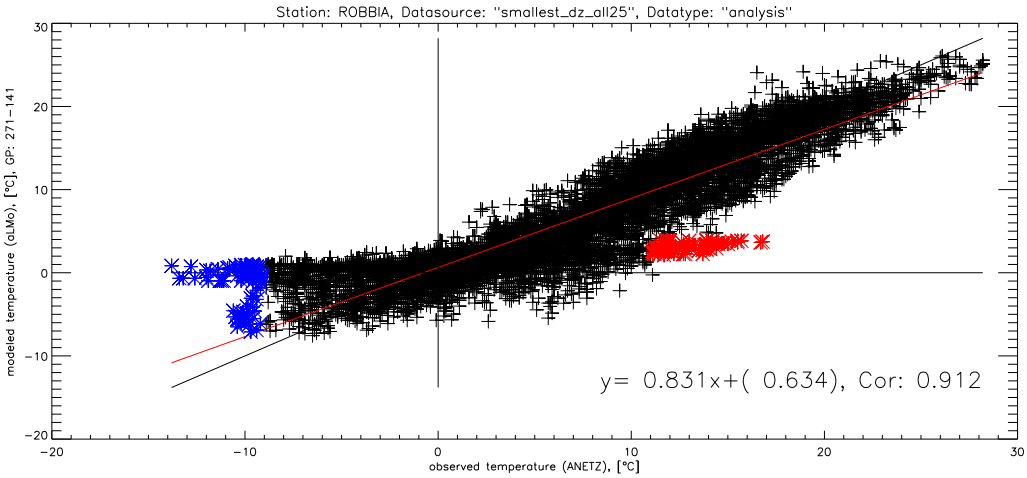


Figure 5.6: The full year dataset (2m-temperatures) of Robbia. For further remarks we refer to Fig. 5.5

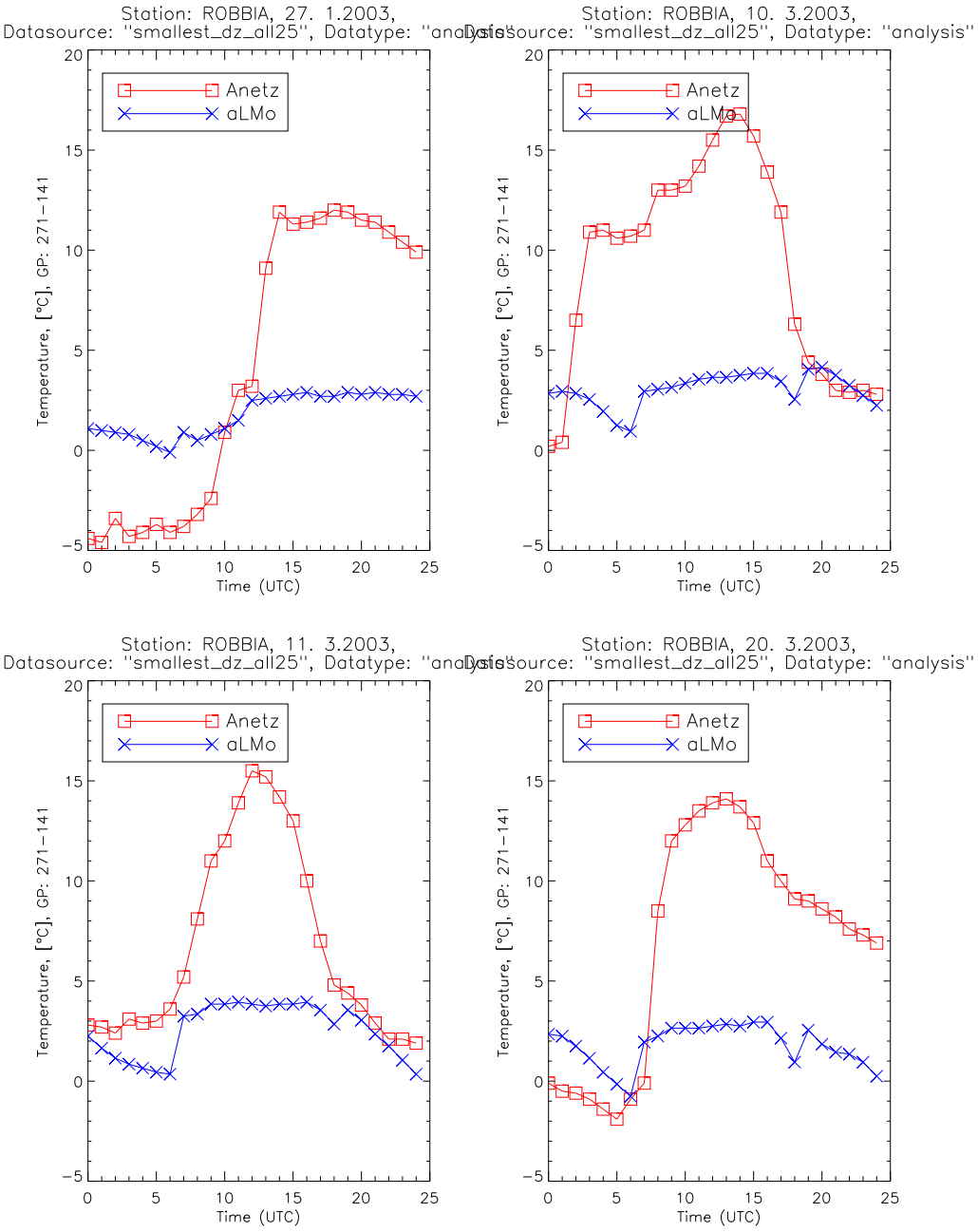


Figure 5.7: The diurnal cycles (2m-temperatures) of four days in the "nose"-region for Robbia. Red: observed data, blue: modelled data.

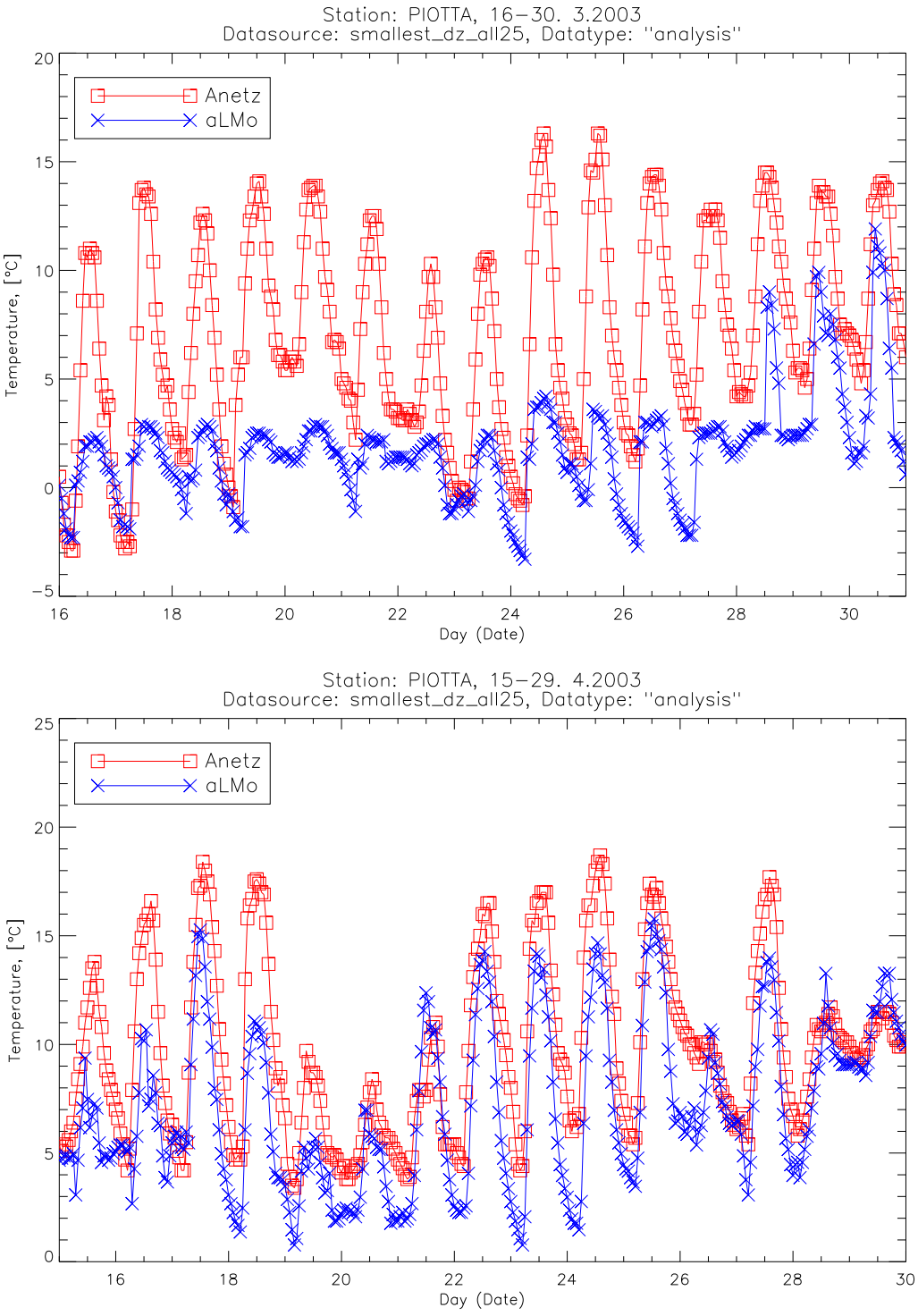


Figure 5.8: 2m-temperatures of several days in a row. The upper panel: days of the "nose"-region for Piotta. Lower panel: a period of time, choosing the same season as in the upper panel but without any days belonging to the "nose"-region for this station. Red: observed data, blue: modelled data.

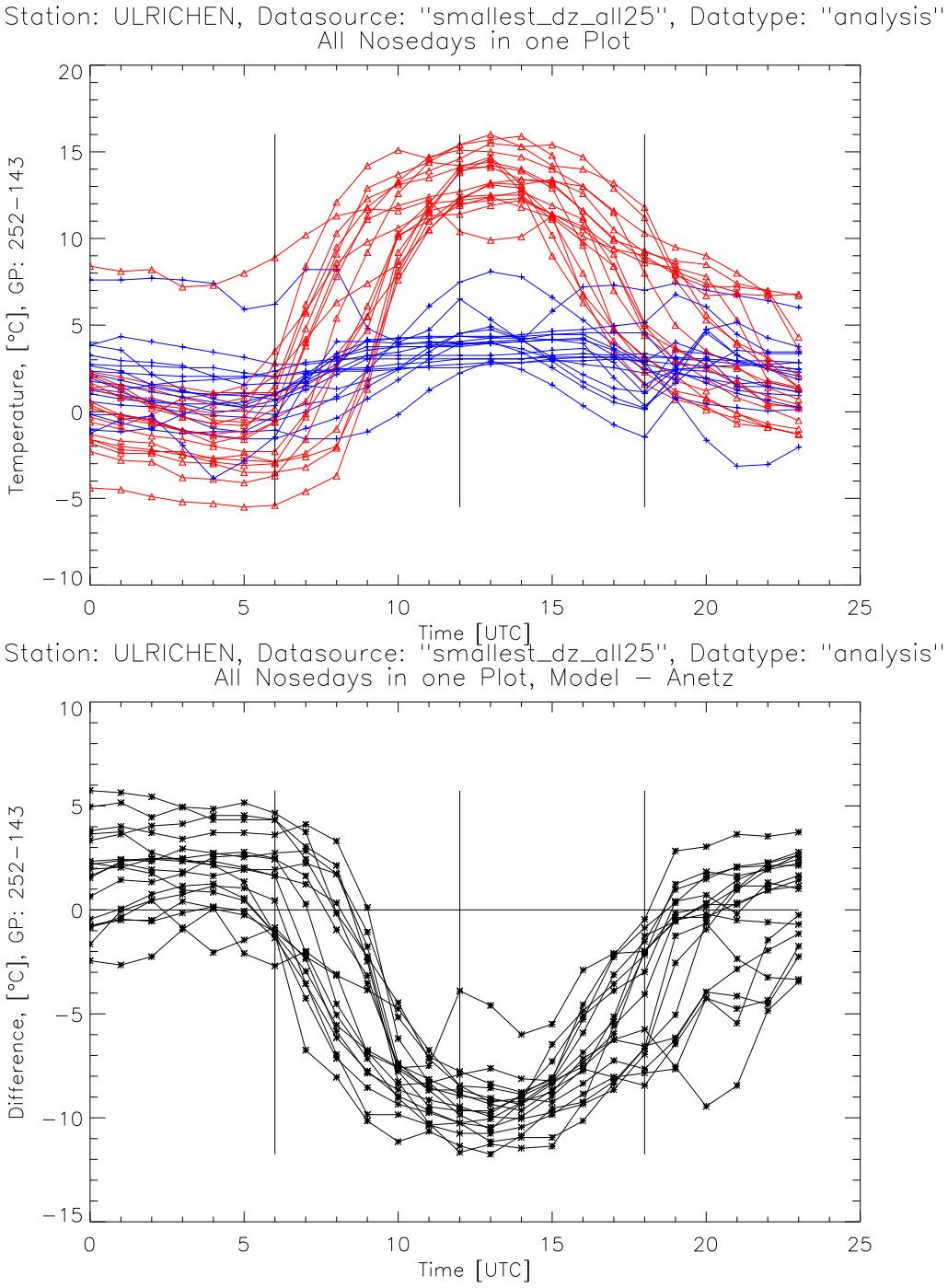


Figure 5.9: Daily cycles of the modelled and observed 2m-temperatures at Ulrichen. Data from days of the 'nose'-region. Upper panel: daily cycles (one line corresponds to one day), lower panel: "hourly" differences between the observed and the modelled data (one line represents one day, diff $= T_{model} - T_{obs}$).
Red lines: observed data, blue lines: modelled data

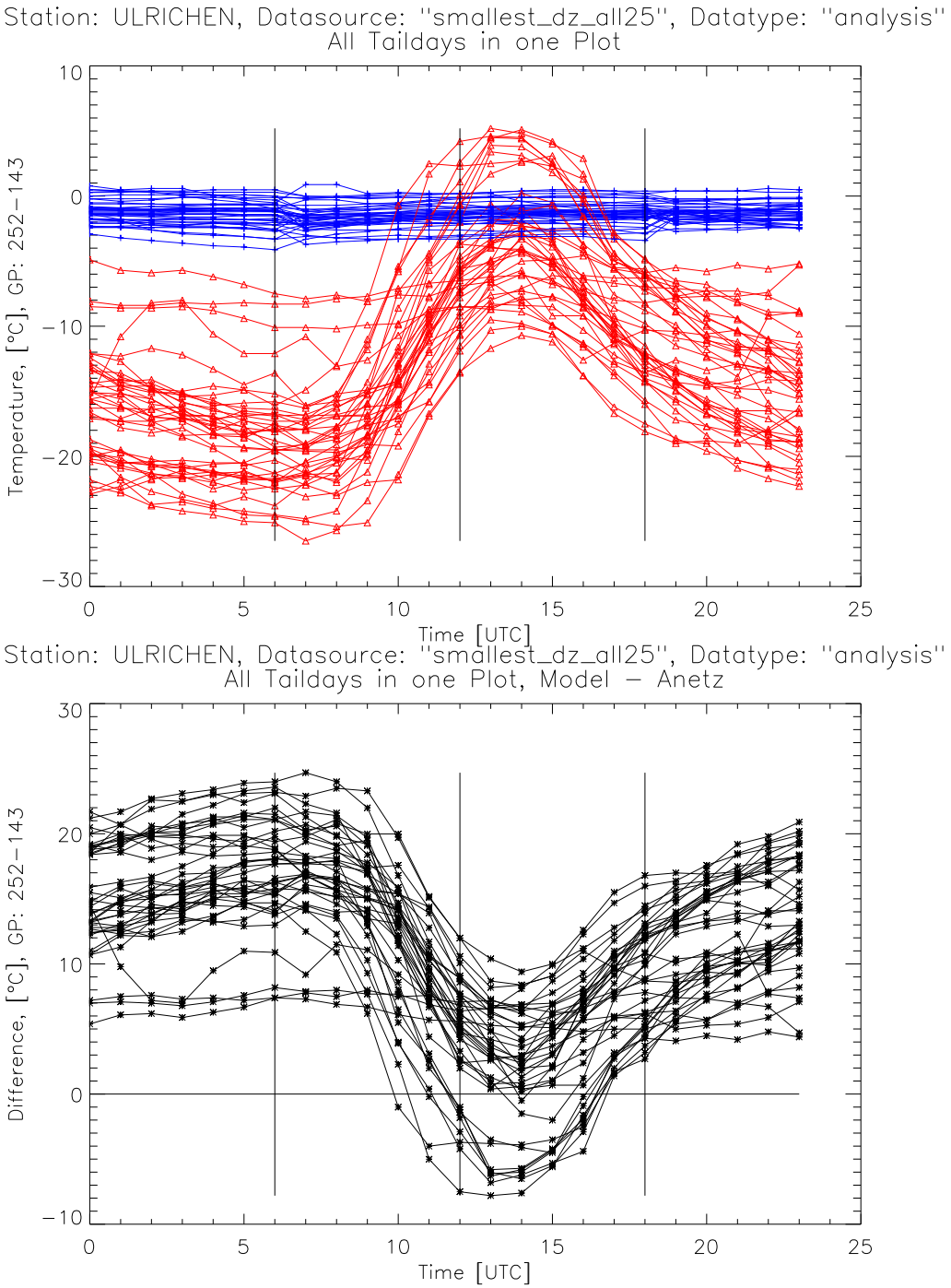


Figure 5.10: Daily cycles, found in the "tail"-region at Ulrichen. For further remarks, we refer to Fig. 5.9

of the "tail"- and a reasonable part of the "nose"-region into our analysis. We clearly see, that weak (< 2 K) and very weak (< 0.5 K) amplitudes in the model's daily cycles tended to occur at low temperatures during winter time. Another pattern we observed, was the explicit horizontal structure of the coloured clouds of data with weak modelled daily cycles.

Fig. 5.12 on p. 50 shows three different scatter plots of the full year dataset of Zermatt. The coloured points in each chart represent days when a weak daily cycle in the modelled data (< 2 K) and a varying cycle (< 2 K, 2-8 K, and > 8 K) in the observed data were found. Based on these results, we defined three classes as listed in Table 5.2. While "lower than 2" and "greater than 6" represent the main classes, the "2-6"-class is an intermediate one.

We want to point out, that the "greater than 6" class was the most important one. During days, belonging to this class, the model failed to reproduce the observed diurnal cycle that for its part was always larger than 6 K. For the same days, the model only described a diurnal variation of less the 2 K.

The "lower than 2" class constitutes a kind of 'reference' when the daily cycle of the modelled data is consistent with the one observed at the surface station.

Table 5.2: Definition of the different classes, used in this study. The amplitudes always refer to max.-min. 2m-temperatures in the period 07-18 UTC.

Class:	Properties		Abbr.
"lower than 2"	Amplitude of Anetz daily cycle:	< 2 K	"lt_2"
	Amplitude of Model daily cycle:	< 2 K	
"2.0 - 6.0"	Amplitude of Anetz daily cycle:	2 - 6 K	"2-6"
	Amplitude of Model daily cycle:	< 2 K	
"greater than 6"	Amplitude of Anetz daily cycle:	> 6 K	"gt_6"
	Amplitude of Model daily cycle:	< 2 K	

Figs. 5.13, 5.14, and 5.15 on p. 52-54 present the typical daily cycles of these three classes ("lower than 2", "greater than 6", and "2-6") considering Ulrichen as an example. Note that the model's diurnal cycles are all located in a very narrow band (cf. Fig. 5.14) and do not show any clear variation while the daily cycles of the observed temperatures were far more spread out. The measurements showed a clear and regular diurnal cycle which led especially at the "greater than 6"-class to a well-defined mean daily cycle in the differences of the two datasets (lower panel).

The Figs. 5.16-5.18 on p. 55-57 summarise the typical diurnal cycles, found at the different stations

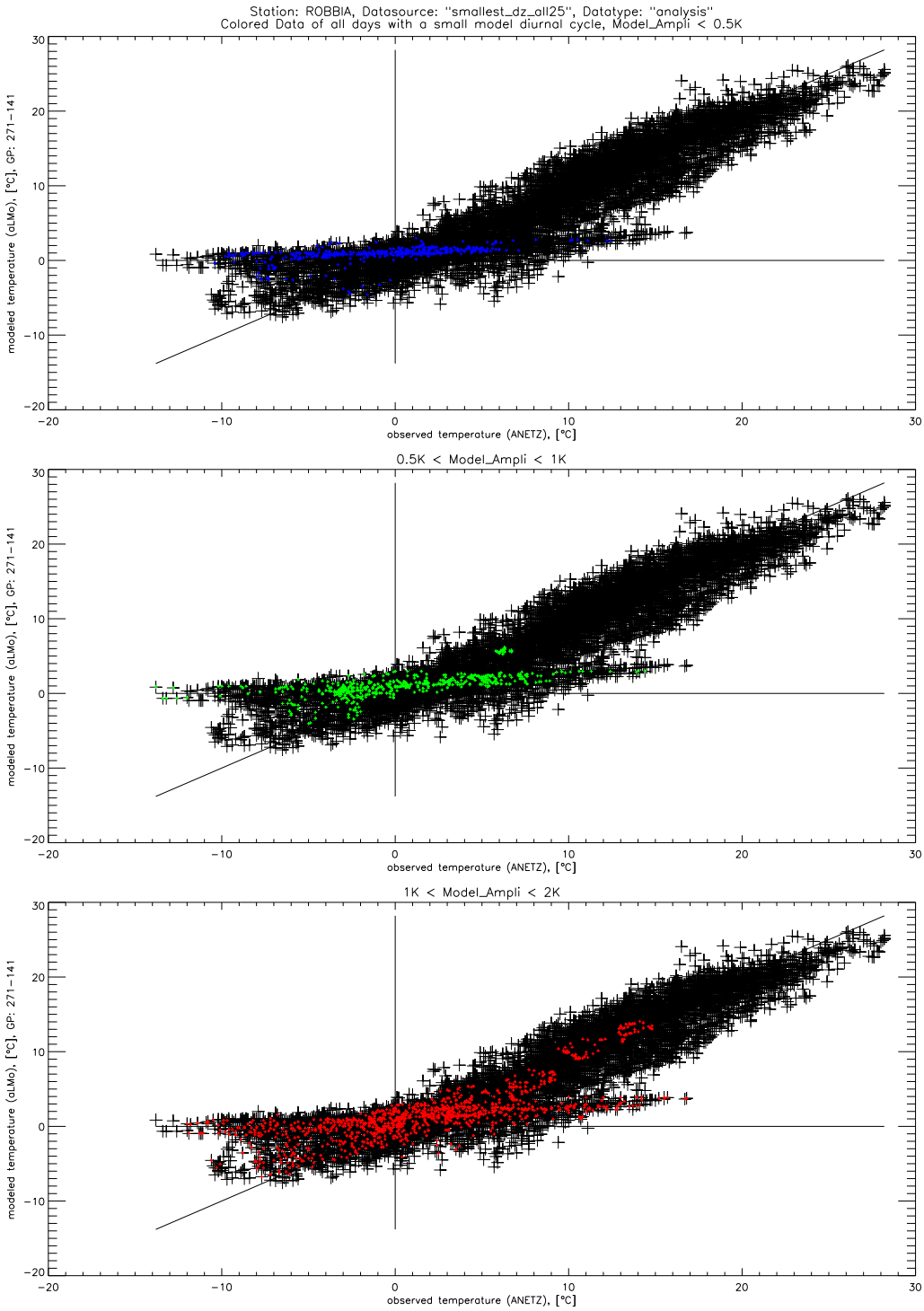


Figure 5.11: The full-year dataset of Robbia (hourly resolution). The marked data points (in different colours) refer to different amplitudes of the model's diurnal cycles. Upper panel: amplitude smaller than 0.5 K, the one in the middle: 0.5-1.0 K, and lower panel: 1.0-2.0 K.

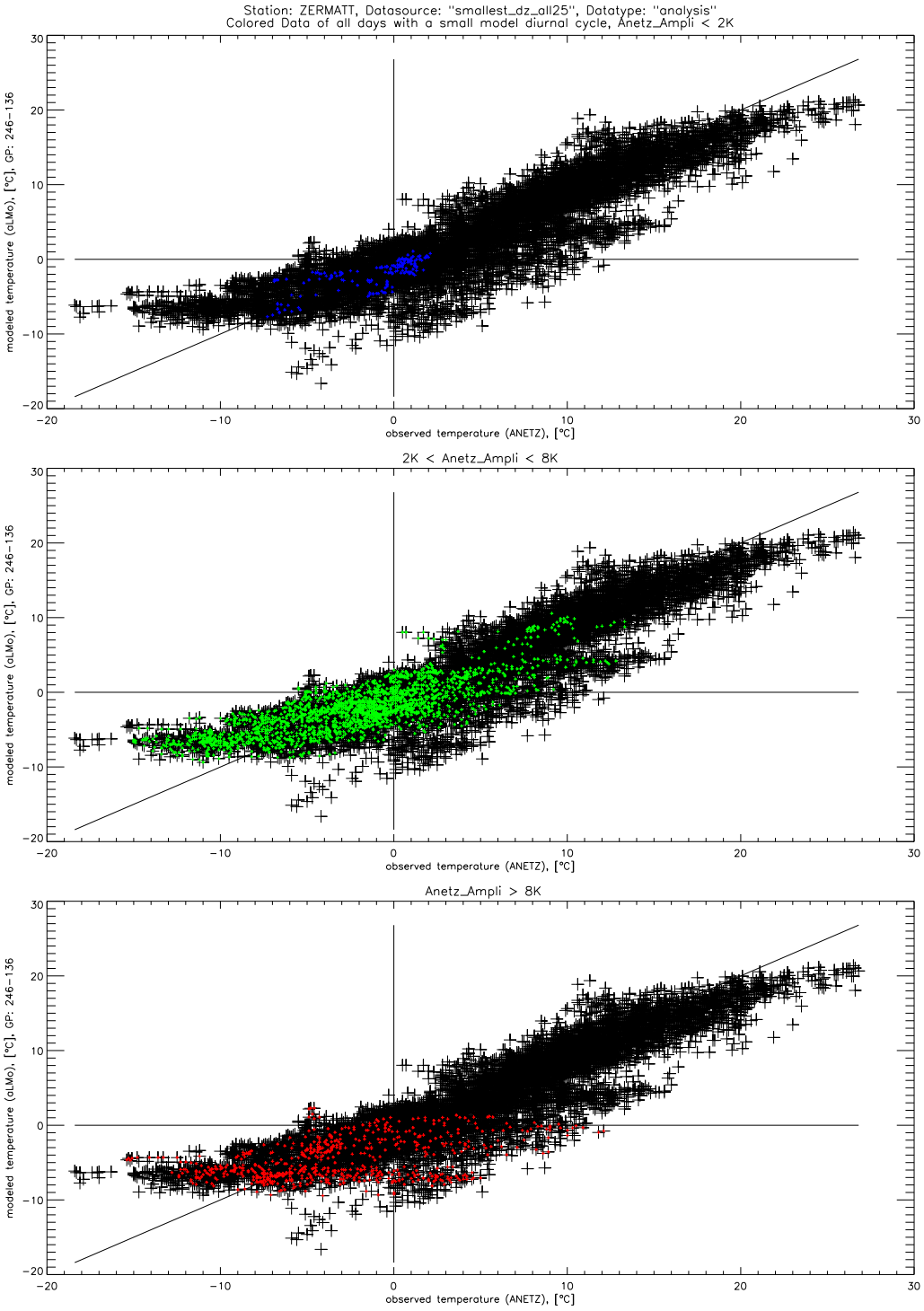


Figure 5.12: The full year dataset of Zermatt is shown (hourly resolution). The marked data points refer to different amplitudes of the diurnal cycles in the observed data at the surface station. Upper panel: observed diurnal cycle smaller than 2 K, the one in the middle: 2-8 K, and the one at the bottom: greater than 8 K.

and classes, respectively. Additionally, the mean daily cycle (inclusive mean ± standard deviation) of all 4 stations of the development group are shown.
While the data of Piotta and Robbia were close together, the mean diurnal cycles of Zermatt and Ulrichen (marked with □ and ◇) showed a different behaviour, and led to a larger standard deviation of the resulting mean daily cycle of all four stations. This was especially true at the "greater than 6"-class (Fig. 5.17).

5.5 Annual Distribution of these Classes

Figures 5.19 and 5.20 on p. 58/58 present the monthly occurrences of the days, belonging to one of the two main classes ("lower than 2", "greater than 6"). We clearly saw, that days, belonging to the "lt_2"-class (Fig. 5.19) were most common in November until January while the days of the "gt_6"-class (Fig. 5.20) were most frequent during the month December until March. Further we can state, that at "Ulrichen" the largest number of "gt_6" days were found. This corresponded well with the most distinctive "tail"-region at Ulrichen seen in the scatter plots (Fig. 5.4 on p. 41). This comparison of the number of days, belonging to the "gt_6"-class revealed, that more such days took place when the "rep" gp was used. This corresponded to the results we found regarding the snow water content (cf. "greater than 6"-days, p. 59. During most gt_6-days the surface was covered with snow. Because the "rep" gp generally exhibits a larger difference in altitude, i.e. is higher situated than the "min_dz25", the time period during which the surface in the model's topography was covered with snow is longer at the "rep" gp. According to this, a larger number of gt_6-days could be found in case the "rep" gp was used.

5.6 Analyses of other Meteorological Variables

At this point of the study we were able to identify days when the observations recorded a relatively strong diurnal cycle (> 6 K) of the 2m-temperatures while at the same time the model was not able to reproduce this pattern accurately (only daily cycle < 2 K). At the other hand we found days when a predicted weak diurnal cycle corresponded well to the measurements taken at the observations.
This motivated us to find reasons for the occurrence of such wrong estimations of the model and we focused on other model output fields. We were interested whether these variables exhibited a certain consistency to each other. A second object of this approach was to establish a set of criteria (based only on model output fields) in order to determine whether in case of a predicted

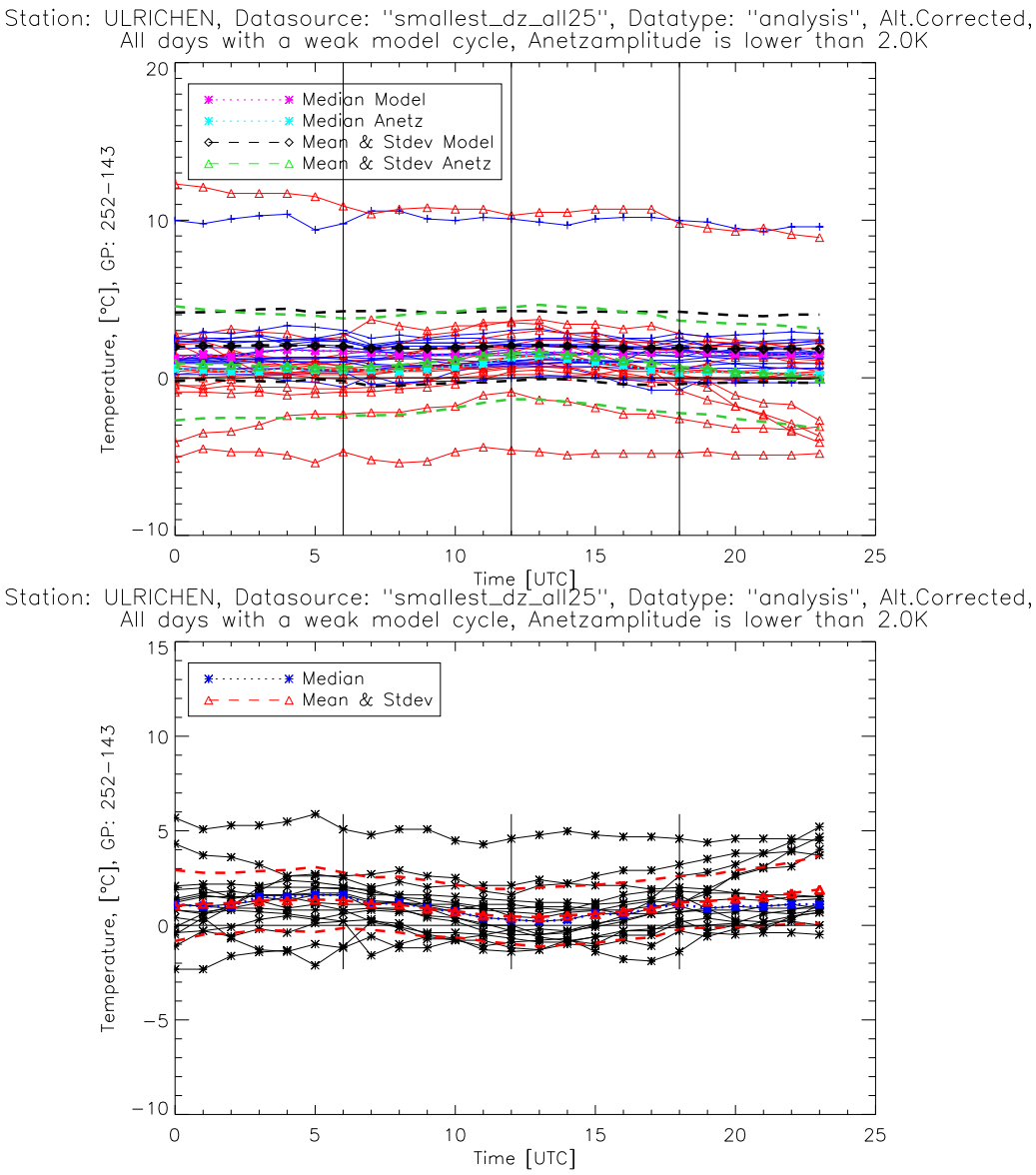


Figure 5.13: Daily cycles of the modelled and observed temperatures at Ulrichen on days of the "lower than 2"-class. Each red (observation) and blue (model) line in the upper panel represents one day belonging to this class. Thick dashed lines as indicated in the legend.
On the lower panel: the deviations between the modelled and observed data (model-obs.).

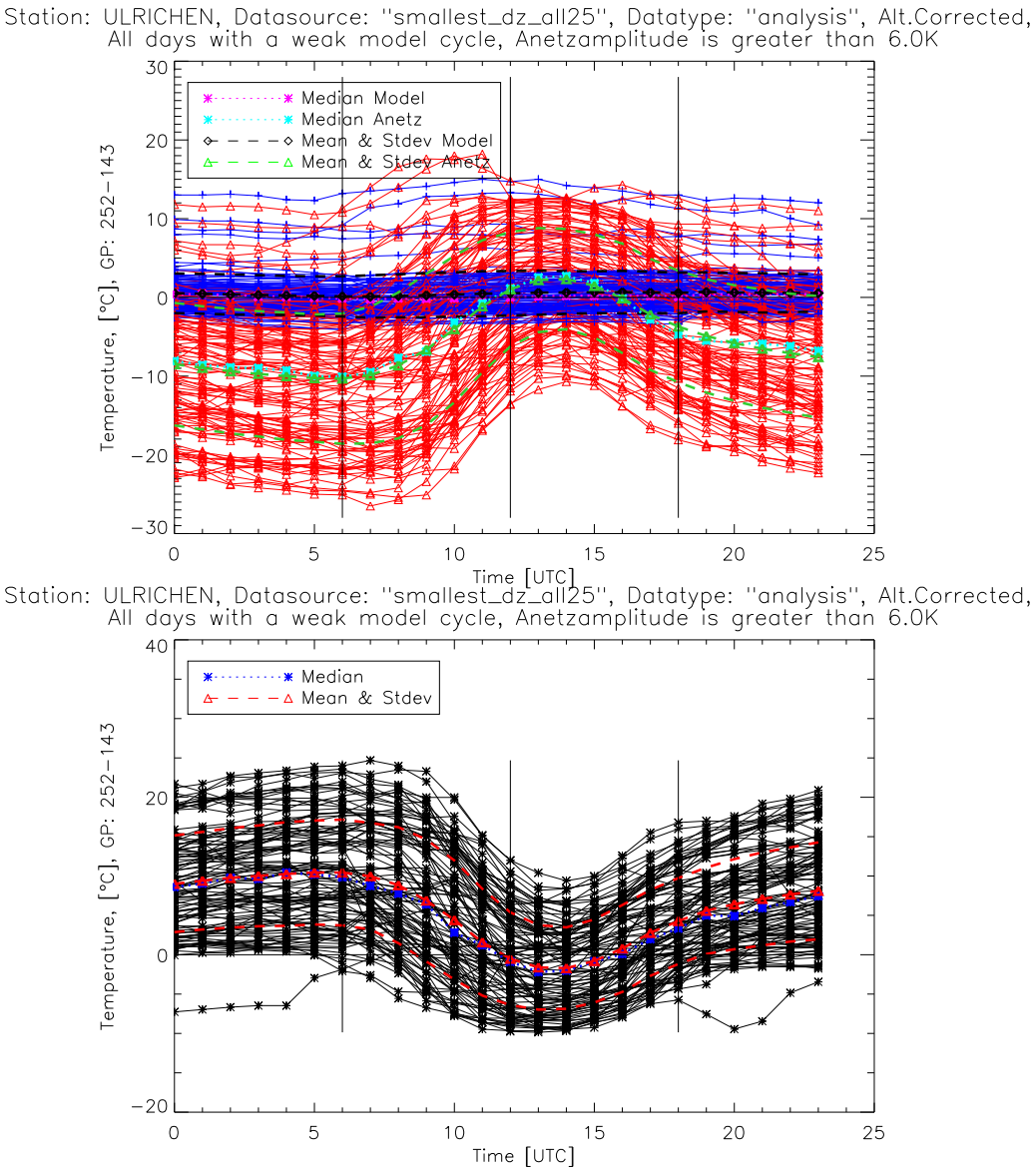


Figure 5.14: The typical daily cycles of the "greater than 6"-class at Ulrichen. For further remarks we refer to Fig. 5.13.

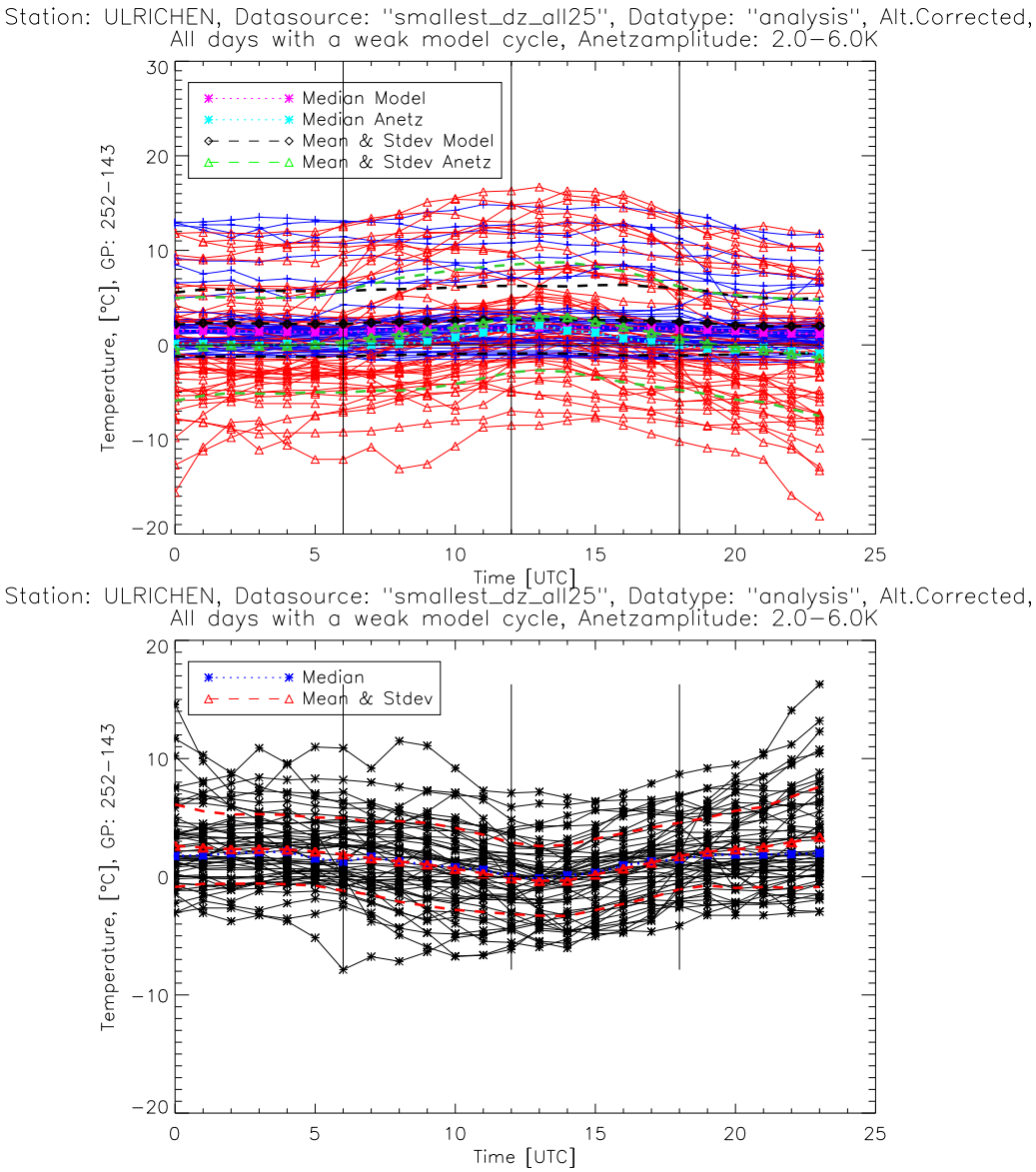


Figure 5.15: The typical daily cycles of the "2-6"-class at Ulrichen. For further remarks we refer to Fig. 5.13

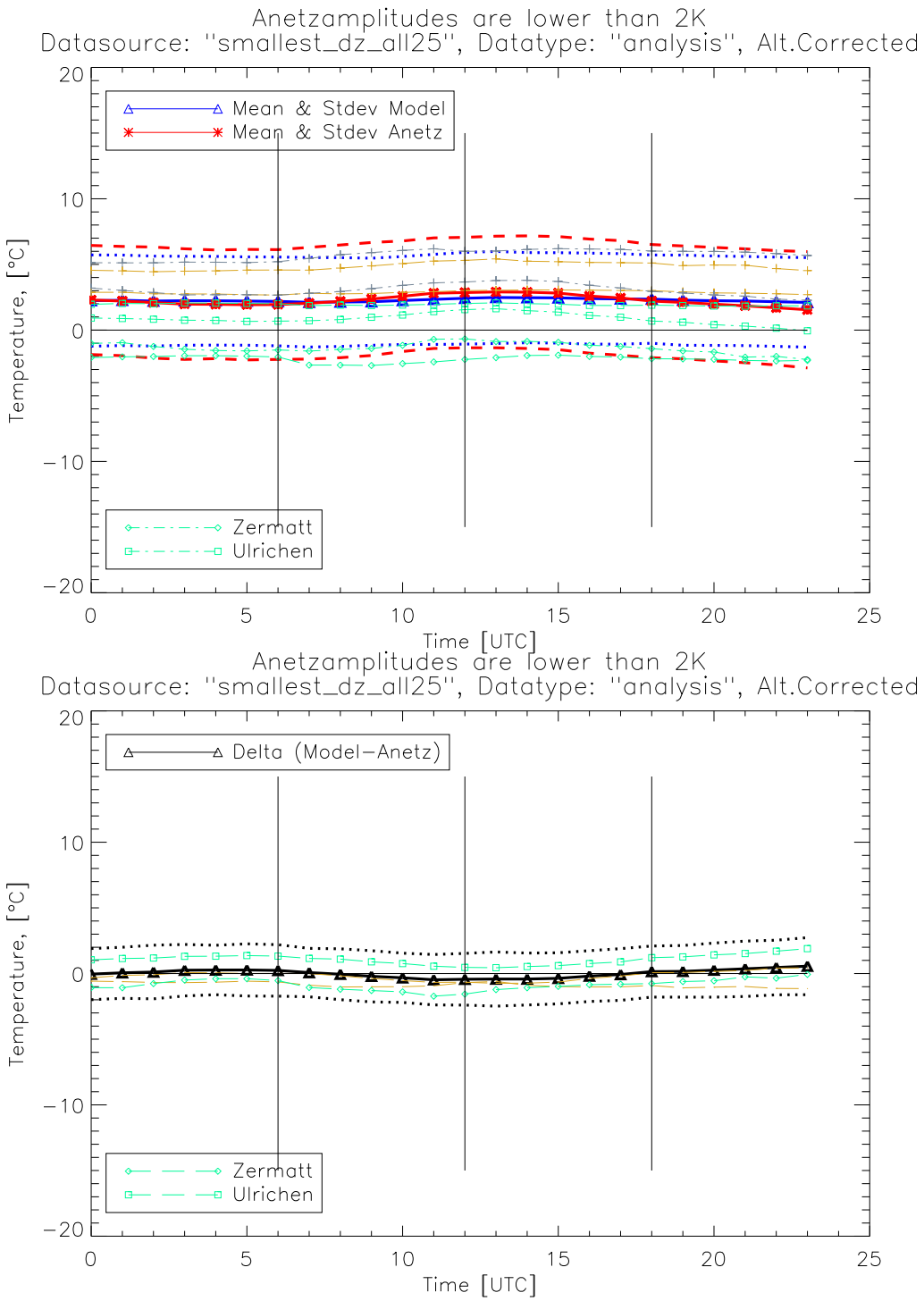


Figure 5.16: Upper panel: mean diurnal cycles at different stations of the development group for the "lt_2"-class.
"Piotta" and "Robbia" in light brown (model) and light grey (observation). "Zermatt" and "Ulrichen" (□ and ◇) in green.
Thicker blue lines (solid and dotted): the mean and mean ± standard deviation of the model, thicker red lines: the mean and mean ± standard deviation of the observation
Lower panel: the mean differences between modelled and observed temperatures.

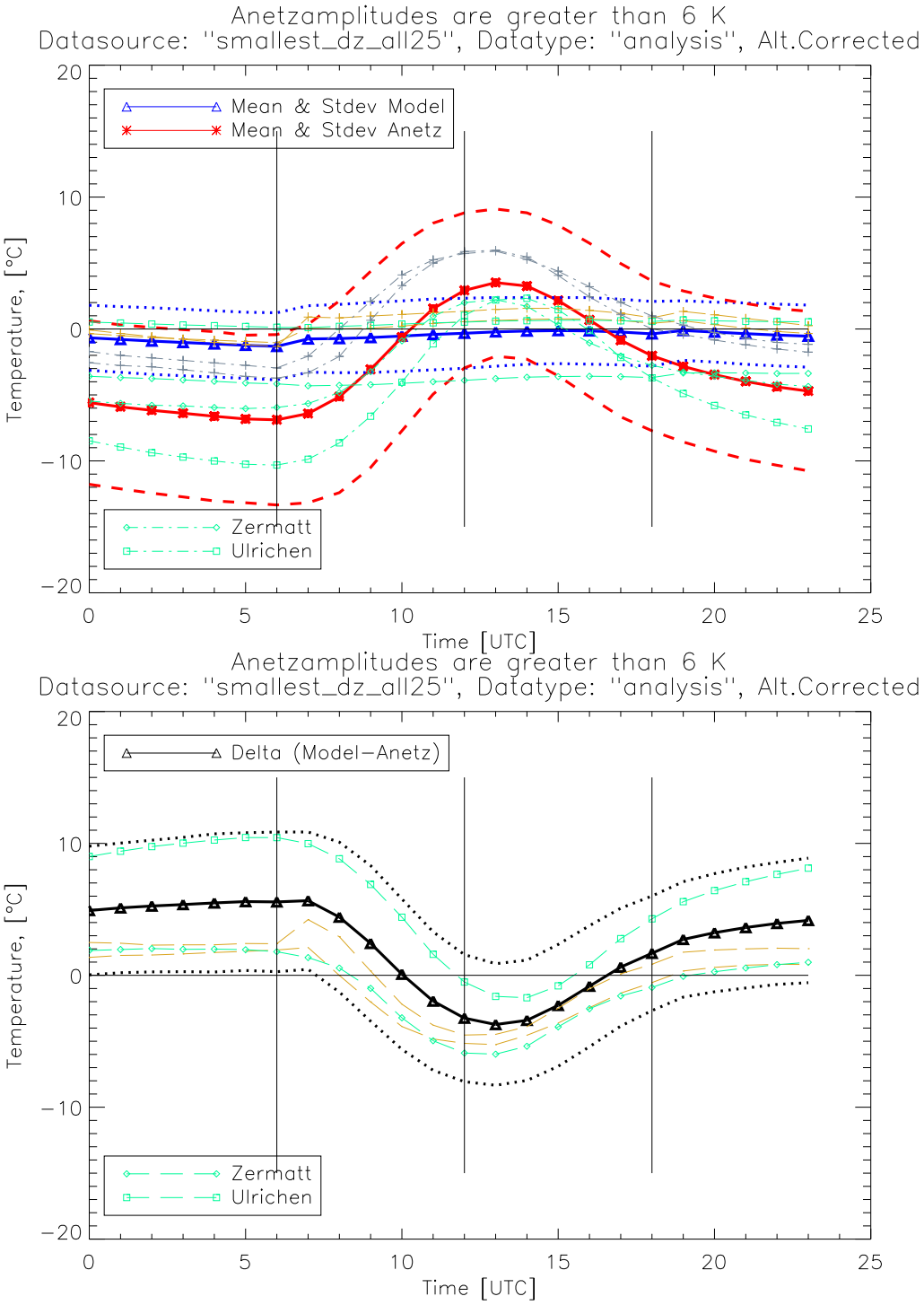


Figure 5.17: The mean diurnal cycles for the "greater than 6"-class. For further remarks we refer to Fig. 5.16

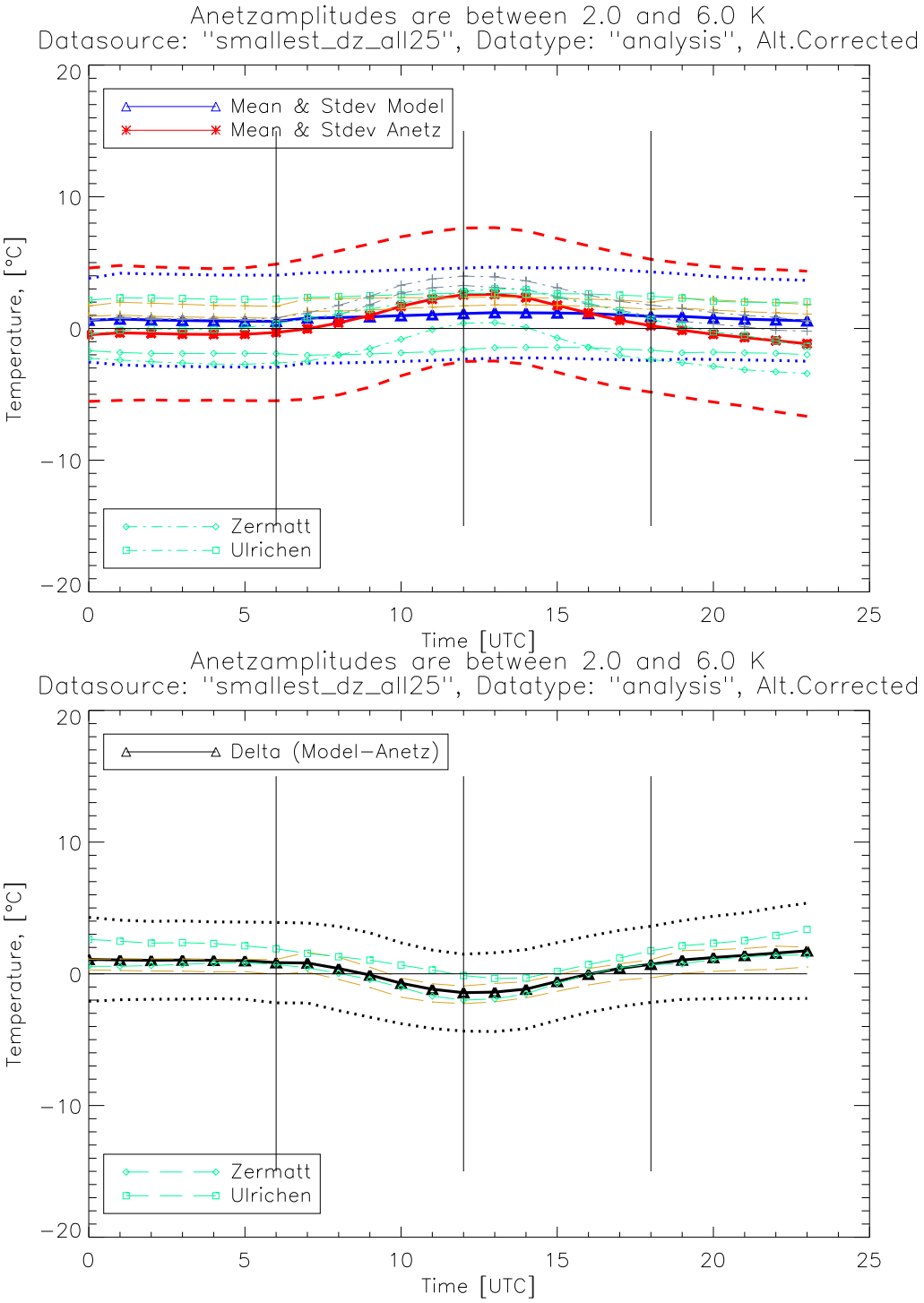


Figure 5.18: The mean diurnal cycles for the "2-6"-class are shown. For further remarks we refer to Fig. 5.16

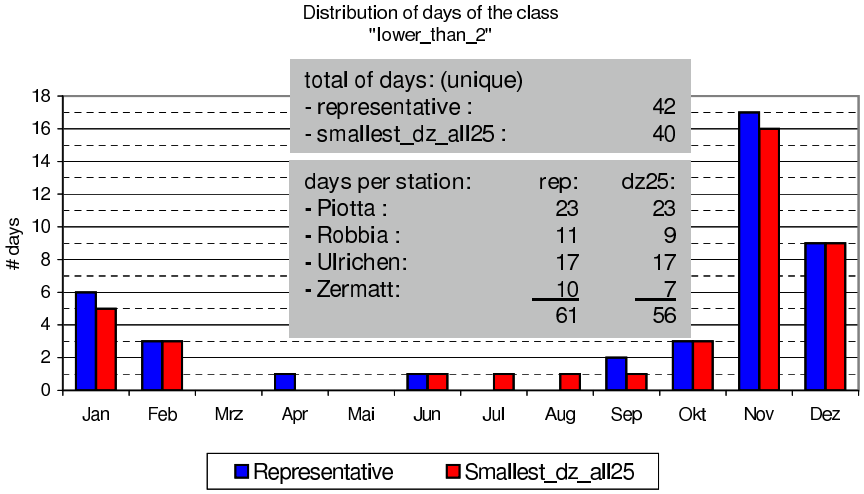


Figure 5.19: The distribution of days member of the "lower than 2"-class. Blue: the number of days at the "rep" gp. Red: the number of days at the "min_dz25" gp.
The "total of days" refers explicitly to unique days, i.e. identical dates that occur at different stations were counted only once.

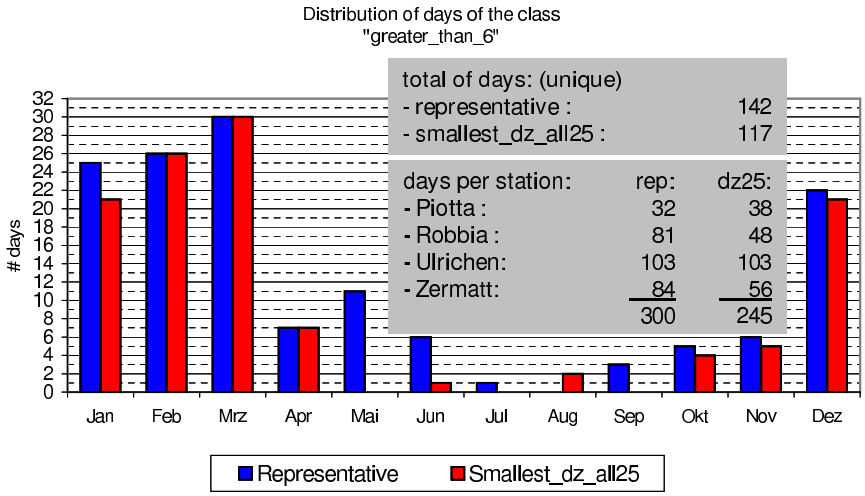


Figure 5.20: The distribution of days, belonging to the "greater than 6"-class. Blue: the number of days at the "rep" gp. Red: the number of days at the "min_dz25" gp.
The "total of days" refers explicitly to unique days, i.e. identical dates that occur at different stations were counted only once.

weak daily cycle (< 2 K) on a certain day this was an accurate forecast of not. In other words, answering the question whether a day for which a diurnal cycle of less then 2 K was predicted actually belonged to the "lt_2"-class (correct prediction) or to the "gt_6"-class (wrong prediction). It turned out, that several of these output fields behaved differently on "lt_2" and "gt_6" days, respectively. Variables like the surface radiation budgets, the cloud covers, the snow temperatures and snow water contents seemed to be these variables most sensitive to the conditions found at the two main classes. The Figs. 5.21 - 5.26 show the typical daily cycles of the above mentioned variables for each main class ("lt_2", "gt_6"). Note the black dashed and dotted lines, which represent the mean and the mean ± standard deviation. In each figure, the daily cycles of all four stations of the development group are shown.
We want to point out that at 6 and 18 UTC extreme discontinuities could be found at the snow-temperatures (up to 10 K) and smaller values at the snow water content (cf. Figures 5.25 and 5.26). In 2002 this behaviour was most likely linked to the (DWD-) analysis of both these parameters at 6 and 18 UTC.
We summarise the patterns found in the different visualisations of the above mentioned meteorological variables for each class separately.

"greater_than_6"-days

Days of the "gt_6"-class tended to be bright. This could be seen in the larger positive solar radiation budgets (cf. Fig. 5.22) and also in the stronger negative thermal radiation budgets than at "lt_2"-days. Further evidences for this assumption were found in the cloud covers (low clouds and total cloud cover, cf. Fig. 5.24) and in the spreading between 2m-temperatures and dew points. Days belonging to the "gt_6"-class exhibited generally a smaller cloud coverage and showed a larger spreading (temperature↔dew point) which indicated dryer conditions then days of the "lt_2"-class. Considering these aspects, we found that these variables were absolutely consistent with the large diurnal cycles recorded at the observations.
During most of the "gt_6"-days, the snow water contents were positive, i.e. the surface was covered with snow (cf. Fig. 5.26). This was consistent with the annual distribution of these days, showing that they were most frequent during the winter (December until March). Far more interesting were the snow-temperatures (cf. Fig. 5.26) when compared with the 2m-temperatures. Corresponding to the radiation budgets, rather strong diurnal cycles in the snow-temperatures were found (in the order up to 10 K). Although these diurnal cycles in the snow-temperatures occurred, no changes 2 m above the snow layer were recorded, i.e. 2m-temperature (diurnal cycle of the modelled 2m-

temperatures on these days ("lt_2"-, "2-6"-, "gt_6"-class) by definition < 2 K).
Further, neither the temperatures of the three different soil-layers (0, 9, and 41 cm depth) nor the temperatures on the first two model layers (45k, 44k) showed characteristic daily cycles.

"lower_than_2"-days

Days belonging to the "lt_2"-class exhibited smaller positive solar radiation budgets and at the same time less negative thermal radiation budgets (cf. Fig. 5.21) than days of the "gt_6"-class. Also the larger mean values in the cloud coverage indicated rather overcast conditions (cf. Fig. 5.23). This corresponded well to the small diurnal cycles (< 2 K) recorded at the surface stations.
On such "lt_2"-days the surface was also mainly covered with snow but the snow depth was smaller (cf. Fig. 5.25) than on "gt_6"-days. On "lt_2"-days, we found no characteristic diurnal cycles in the snow-temperatures. Again this fitted to the smaller solar radiation budgets.
All these aspects were consistent with the small diurnal variations of the 2m-temperatures observed at the surface stations.

Summary

While all these model output variables corresponded to the recorded daily cycles of the 2m-temperature at the surface stations, only the modelled 2m-temperatures in case of bright ("gt_6"-) days did not correspond to the rest of variables and not to the observed temperatures. Especially interesting were the strong diurnal cycles of the snow-temperatures during "gt_6"-days while these variations were completely missing at the 2m-level.
Note that these meteorological variables were not investigated on days belonging to the intermediate "2-6"-class (cf. Table 5.2 on p. 48).

5.7 Definition and Validation of the Criteria

In the previous section we learned that several meteorological variables showed a different behaviour in case of "lower than 2" and "greater than 6"-days, respectively. Based on these results we derived a set of criteria that could be used to identify "gt_6" and "lt_2"-days.
A typical "gt_6"-day was bright and therefore exhibited a larger positive solar radiation budget, a smaller cloud coverage, and a remarkable diurnal cycle of the snow-temperature. Additionally,

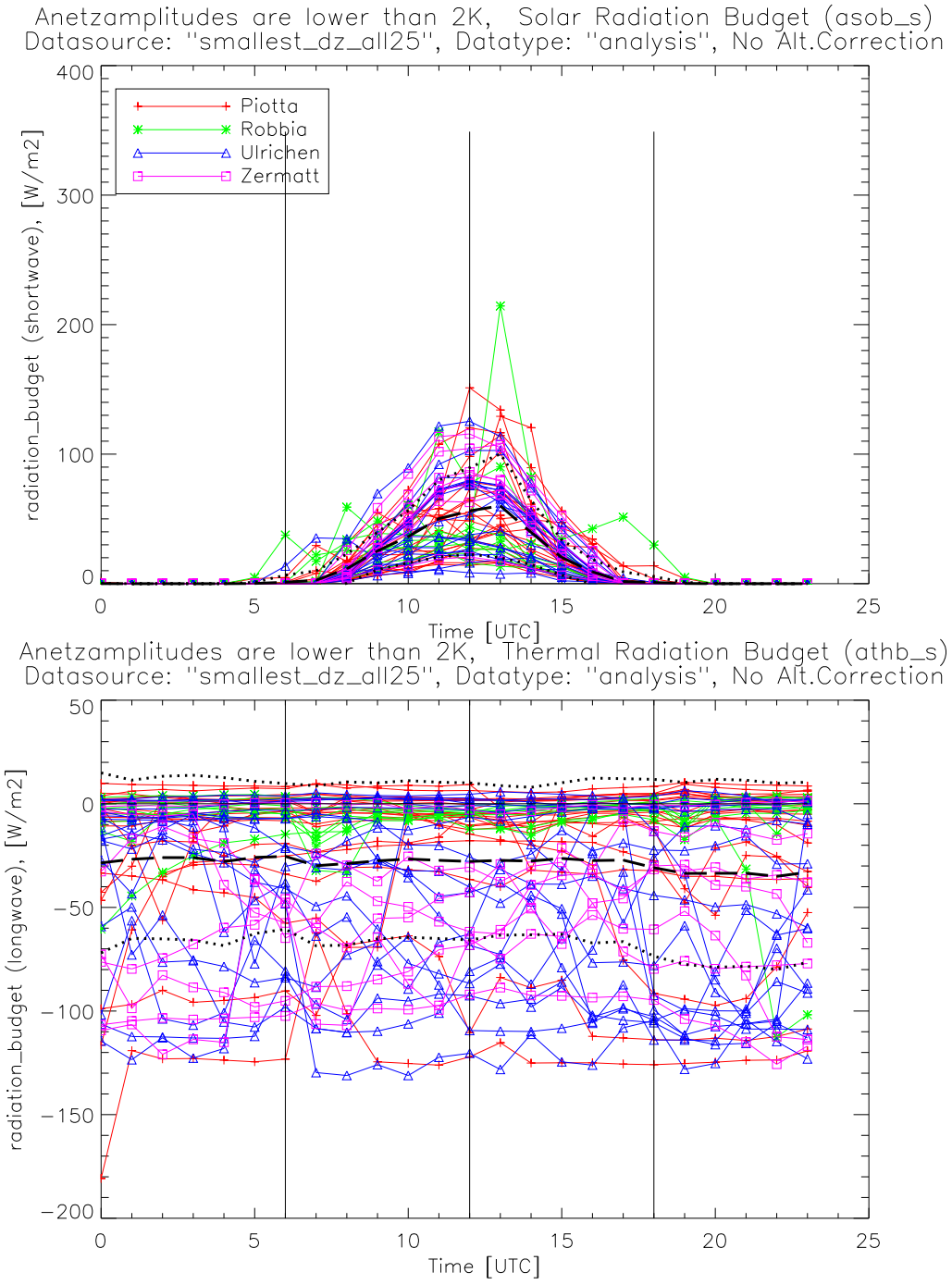


Figure 5.21: Short-wave (asob_s) and long-wave (athb_s) radiation budgets on "lower than 2"-days. Upper panel: the solar radiation budget, lower panel: the thermal radiation budget.
The different colours refer to the four surface stations. red: Piotta, green: Robbia, blue: Ulrichen, magenta: Zermatt.

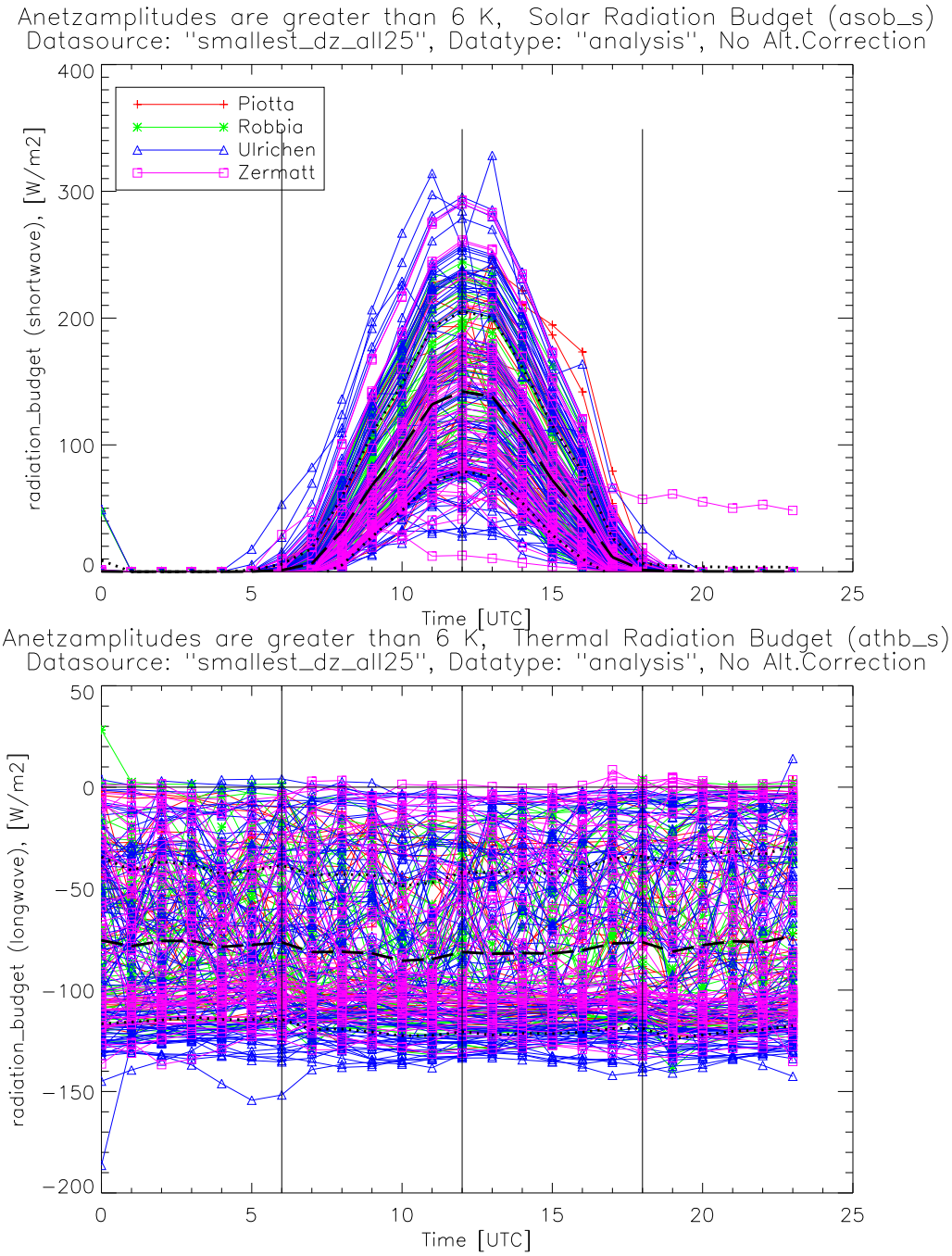


Figure 5.22: The radiation budgets on "greater than 6"-days. For further remarks we refer to Fig. 5.21

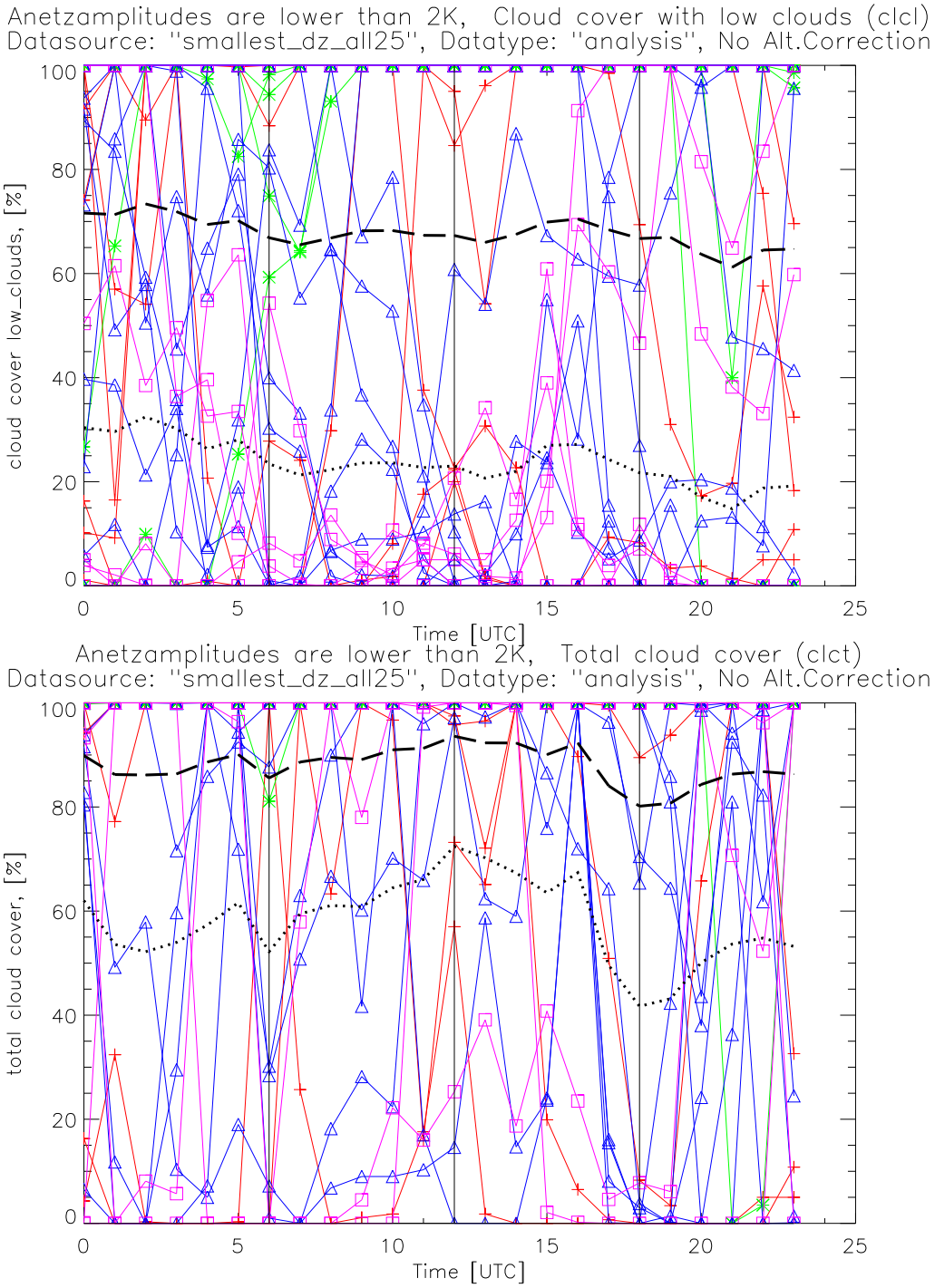


Figure 5.23: Different cloud covers (low and total clouds) on "lower than 2"-days. Upper panel: the cloud coverage by low clouds, lower panel: the coverage by all clouds on all layers (total). The y-axis specifies the cloud cover in percent.
The different colours refer to the four surface stations. red: Piotta, green: Robbia, blue: Ulrichen, magenta: Zermatt.

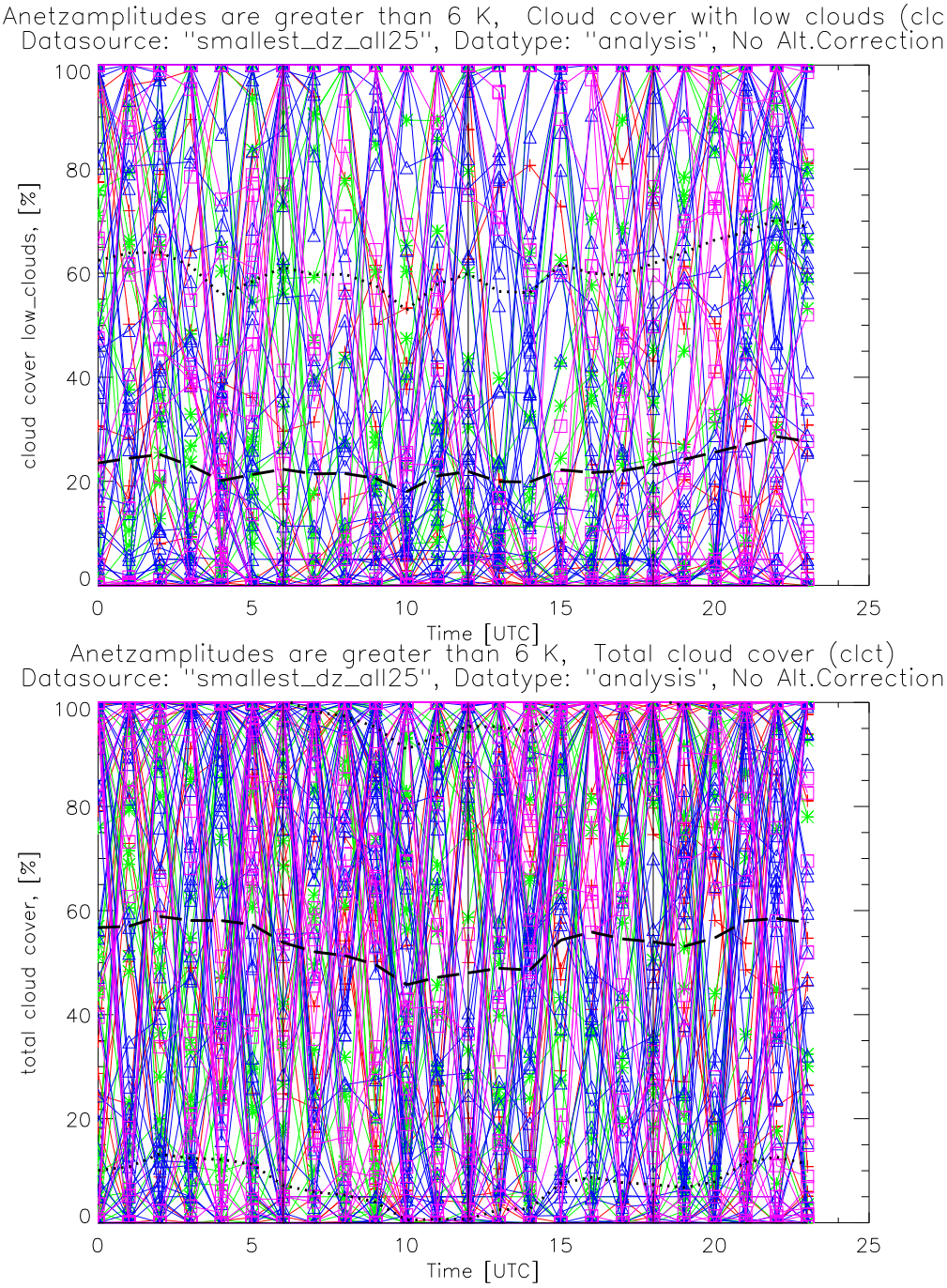


Figure 5.24: Different cloud covers (low and total clouds) on "greater than 6"-days.
Upper panel: the cloud coverage by low clouds, lower panel: the coverage by all clouds on all layers (total).

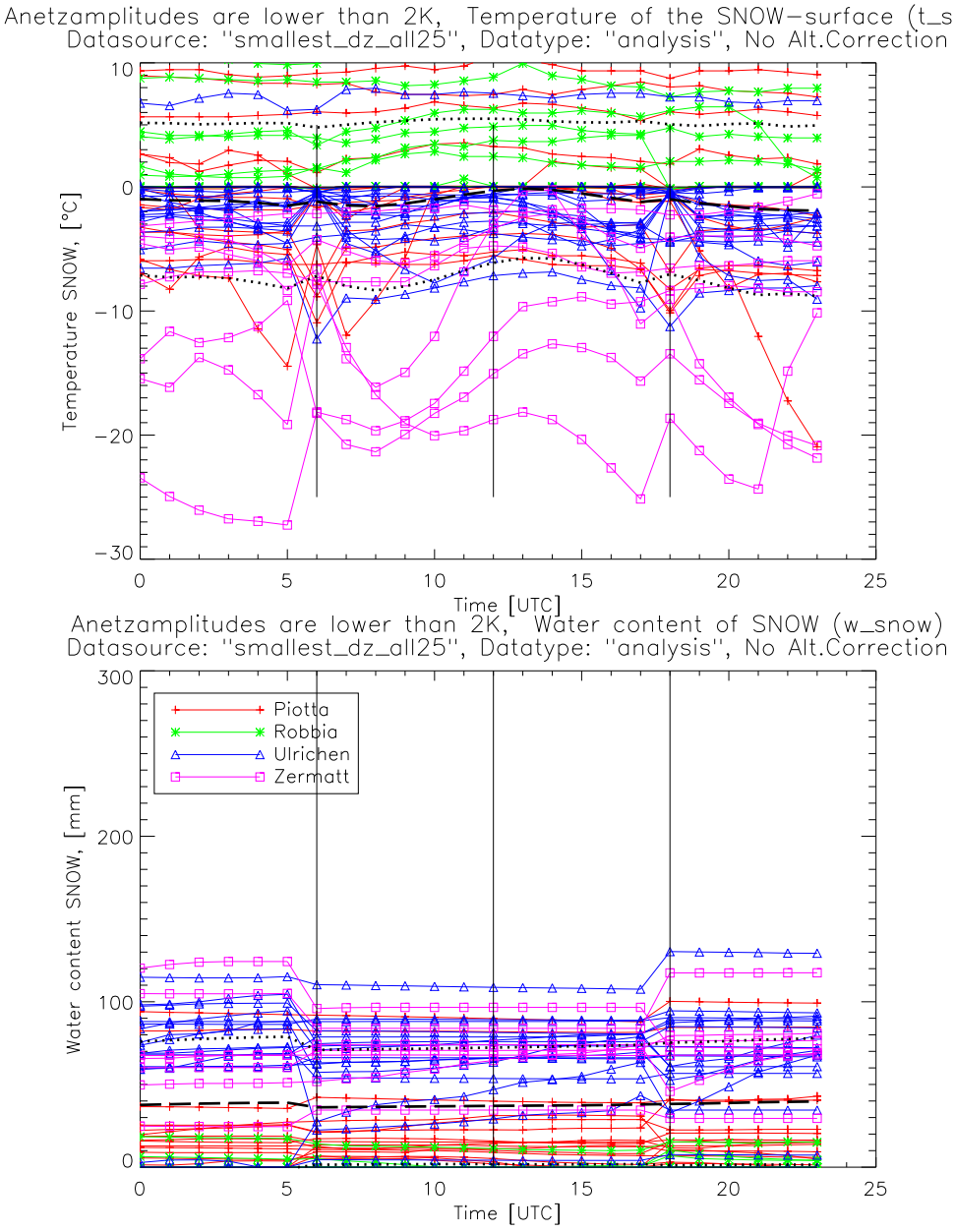


Figure 5.25: The snow temperature (upper) and the snow water content (lower panel) on "lower than 2"-days.

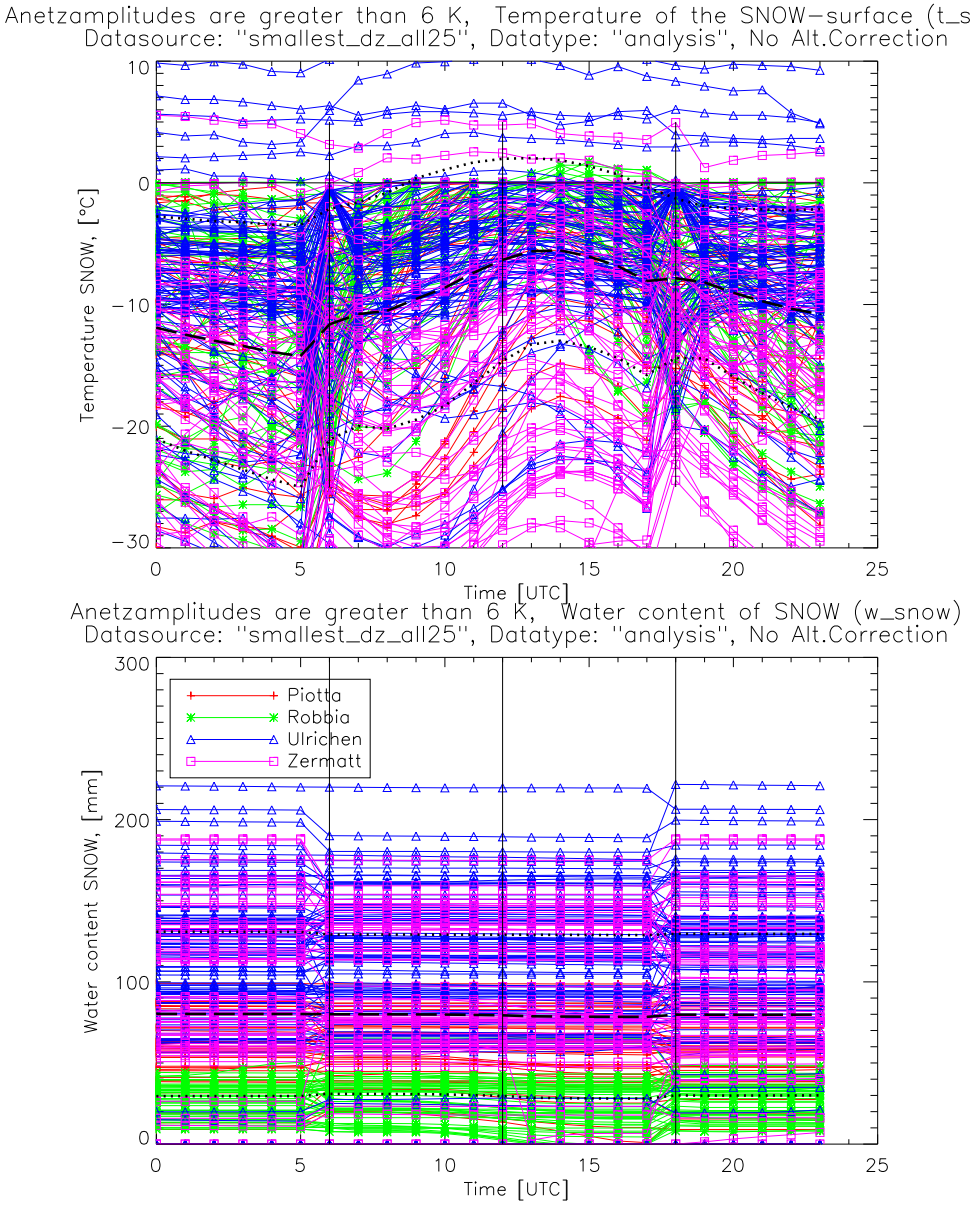


Figure 5.26: The snow temperature (upper) and the snow water content (lower panel) on "greater than 6"-days.

the soil was covered with snow and by definition the daily cycle of the 2m-temperature was lower than 2 K.
On the other hand, "lt_2"-days exhibited smaller solar radiation budgets, larger cloud coverages and no such diurnal cycles of the snow-temperature. Therefore these days were rather overcast. Again, the surface was mainly covered with snow and the diurnal variations of the 2m-temperatures were still (by definition) smaller than 2 K.

Using this knowledge, we were able to determine threshold-values for each of these variables in order to assign a specific day to one of the two main classes ("gt_6", "lt_2"). Note that the identification of days belonging to the "2-6"-class (intermediate class) was an assignment of these days, when a weak modelled diurnal cycle (< 2 K) was predicted and which did not already belonged to the "lt_2" and "gt_6"-class.

Definition of the Criteria:

Table 5.3 on p. 68 lists the different variables used as criteria and their threshold-values. In order to characterise a day with a specific variable, daily means, daily maximal values or amplitudes of diurnal cycles were used.
Note in both sets of criteria (for "rep" and "min_dz25") was exactly one variable with a threshold that explicitly separated the two classes, while the others showed overlapping ranges. In case of the "representative" gp this was the daily mean coverage through low clouds, while at the "smallest_dz_all25" gp the amplitude of the snow-temperature completely separated the two classes. Note that in case of using the amplitude of the diurnal cycle of a specific variable not the whole day (00-23 UTC) was considered. The reason for taking only a section of the data of a full day were discontinuities (around 6 and 18 UTC), most pronounced in temperature data as the ones of the 2m-level and of the snow cover.
The values, shown in Table 5.3 were found empirically by trying to 'detect' the maximum number of days in a particular class. Note that additionally to the variables, denoted to be most important in characterising a specific day, we used the daily maximal 2m-temperatures and the amplitude of the long-wave radiation budget in order to improve the ability of this method.

Table 5.3: Modelled variables characterising different classes of days with low modelled daily cycles in the 2m-temperature (< 2 K, cf. first row of each table). On the upper panel the values for the "representative" gp, in the lower one for the "smallest_dz_all25" gp are shown. Bold, the two variables that are separating the two classes ("lt_2", "gt_6").

Representative				
	lower than 2		greater than 6	
Variable	Threshold	Check for	Threshold	Check for
Amplitude in 2m-temp. (7-18 UTC)	< 2 K	lower	< 2 K	lower
Daily max. solar radiation budget	< 120 W/m^2	lower	> 70 W/m^2	higher
Daily mean snow water content	> 0 mm	higher	> 10 mm	higher
Amplitude in snow-temperature (7-17 UTC)	< 5.5 K	lower	> 2 K	higher
Daily mean cloud cover (low clouds)	> 55%	higher	< 55%	lower
Daily max. 2m-temp.	< 3 °C	lower	< 3 °C	lower
Amplitude in long-wave radiation budget	< 120 W/m^2	lower	< 140 W/m^2	lower

Smallest_dz_all25				
	lower than 2		greater than 6	
Variable	Threshold	Check for	Threshold	Check for
Amplitude in 2m-temp. (7-18 UTC)	< 2 K	lower	< 2 K	lower
Daily max. solar radiation budget	< 120 W/m^2	lower	> 60 W/m^2	higher
Daily mean snow water content	> -0.1 mm	higher	> 0 mm	higher
Amplitude in snow-temperature (7-17 UTC)	< 3 K	lower	> 3 K	higher
Daily mean cloud cover (low clouds)	> 30%	higher	< 65%	lower
Daily max. 2m-temp.	< 10 °C	lower	< 10 °C	lower
Amplitude in long-wave radiation budget	< 120 W/m^2	lower	< 130 W/m^2	lower

Table 5.4: Test on recovery rate, false alarm and hits for days with a low modelled diurnal cycle in the 2m-temperatures (i.e. < 2 K) and various observed amplitudes of daily cycles.
total: total days in this class (note, that all four stations were considered).
total found: days that were found by applying the criteria and thresholds (cf. Table 5.3).
hits: correctly identified days in the set of "total found" days.
Further information can be found on p. 69

	Representative			Smallest_dz_all25		
	lt_2	2-6	gt_6	lt_2	2-6	gt_6
total	61	288	300	56	204	245
total found	89	154	340	82	109	268
hits	34	101	224	27	60	177

Validation of the Criteria:

The previous section has shown that the two main classes ("lt_2", "gt_6") do not exhibited entirely distinctive model parameters. Still, it was interesting to analyse the 'recovery rate' when applied on the original dataset (i.e. used for the developing of the classes).
Table 5.4 shows that e.g. for the "representative" gp a total number of 300 days originally belonged to the class "gt_6", 61 to the "lt_2" and 288 to the intermediate class "2-6". Note that these numbers of days refer to the total amount of days of all four stations of the development group. Applying the criteria and thresholds (Table 5.3) yielded to 340, 89, and 154 days, respectively, i.e. a considerable amount of false alarms. In fact, only 224, 34, and 101 days, respectively were 'hits' or correctly identified for the respective class.

Table 5.4 reveals, that over 70% of the days needed of the "greater than 6"-class were found. Applying the criteria of the "lower than 2"-class we were able to find approximately half of the days needed. Keep in mind, it was the "gt_6"-class, which was most problematic and therefore deserved our main attention. The "lt_2"-class represented merely these days, when a small diurnal cycle in the modelled 2m-temperature (< 2 K) actually was correct, i.e. was also recorded at the specific surface station.
Note that the days in the "2-6"-class are not directly identified by using criteria, but rather by assigning the days not already belonging to the other two (main-) classes. In this class, the rate of the correct found days did not exceed $\sim$35% and $\sim$30% at the "representative" and "smallest_dz_all25" data source, respectively.

5.8 Applying the Corrections

In the previous sections characteristics have been recognised on those days during which the model produced almost no diurnal cycle in the 2m-temperatures while the observed values exhibited no ("lt_2") or a pronounced ("gt_6") daily variation. Clearly, for a forecast it would be desirable to correct the apparently missing daily cycle of the temperatures for the classes "gt_6" and also for the intermediate "2-6"-class. Note that we performed corrections of all three classes.
Such a correction can be effectuated as follows:

1. Based on the criteria and threshold-values of Table 5.3 attribute a specific day to one of the two classes ("lt_2","gt_6"). In case the patterns found on this day did not fit to the criteria of the two main classes, the day was attributed to the intermediate "2-6"-class.
2. Use the mean observed differences between modelled and observed 2m-temperatures (cf. Figures 5.16-5.18 on p. 55-57) for the correction of the modelled diurnal cycle of the 2m-temperature.

Note, whenever we use the term "daily cycle" or "diurnal cycle" in order to correct modelled 2m-temperatures, we refer to the process of adding the mean daily cycle of the differences (model↔observation) to the model data.

Apparently, when correcting a forecast, step 2 from above can only be performed by using the average differences. One would expect this differences to be most representative if they were only calculated from the observed data of the station of interest. However, in this case the correction method would only be applicable for this very site.
Therefore, a general correction procedure would use the mean (i.e. from all four stations of the development group) differences, with the underlying assumption that this mean correction would be applicable for any "grid point-observation"-pair in the alpine region. In order to qualify this assumption, the correction was applied twice, once using the mean differences between the observed and modelled 2m-temperatures (e.g. Fig. 5.17 on p. 56)) and once with the difference curve from the respective site alone.

In the following, the results of the different corrections and the impacts of the two data sources ("rep", "min_dz25") are described. For more detailed information we refer to the Tables A.1 - A.8 on p. 124 - 131 (Appendix). To gain clearer results we only used the data of days, when a modification at the specific class actually took place. Otherwise, using all-year data points (considering as

well these days where no modification at all was performed) led to the same but more indistinctive results.

Analysis of only Altitude-corrected Data

The statistical measures (RMS, MB, and MAE, cf. Methods, p. 32) were analysed in case of the modelled 2m-temperatures of "gt_6"-days were only altitude-corrected. These statistical measures are shown in the bar plots on Fig. 5.27 on p. 73.
The largest errors in all three statistical quantities and data sources were found at Ulrichen. At the other three stations (Piotta, Robbia, and Zermatt) the errors were smaller. While at Ulrichen RMSs ("rep","min_dz25") of more than 8 K were observed, the values at the other three stations were only around 4 K. Even greater relative differences were found in the MBs. At Ulrichen the MBs were larger than 4.5 K, at the other three stations a maximum around 1 K was seen (Piotta), i.e. a relative difference of almost 80%. Note that at Robbia and Zermatt even MBs below 0.5 K were found. The MAEs of Piotta, Robbia, and Zermatt were are below 4 K while the ones at Ulrichen exceeded 6 K.

Comparison between the different Data Sources

We analysed the statistical quantities depending on the different data sources, i.e. "representative" and "smallest_dz_all25" gp. The results of this comparison of observed and modelled 2m-temperatures of "gt_6"-days can be seen in Fig. 5.27 on p. 73).
For this analysis, the modelled data were only altitude corrected and contained these days, belonging to the "gt_6"-class. The RMSs, MBs, and MAEs are shown for all four stations of the development group.
The differences between the "rep" and "min_dz25" gp were not very large, typically smaller than 0.5 K. The "rep" gp mostly led to better results, i.e. to smaller errors. Exceptions were found in the RMS and MAE at Zermatt and in the MB at Piotta and Robbia.
In the beginning of this study, we found in general better results (concerning the daily minimum and maximum temperatures, respectively) in case the "min_dz25" gp was used (cf. p. 35). This conclusion had no validity, regarding the statistical quantities of only these days, belonging to the "gt_6"-class and including all data of the specific days instead of only the max. and min. temperatures, respectively.
Considering only "gt_6" days, the use of the "representative" gp led to better re-

sults in the selected statistical measures (RMS, MB, and MAE). The only exception is Zermatt, where reduced errors (RMS and MAE) were found in case of using the "min_dz25" gp.

The Tables A.1/A.1 and A.5/A.5 (Appendix) present the statistical analyses in detail. Using the "min_dz25" gp clearly led to worse results at Piotta and Robbia, seen in the RMS, MB, and MAE. On the other hand, at the stations of Ulrichen and Zermatt, choosing the "min_dz25" gp had a positive impact on the statistical quantities. The magnitude of the improvements depended on the station, the class and the specific statistical quantity.

Comparison between the different Mean Diurnal Cycles

Instead of using the individual daily cycle ("Individual") of a certain station (observed typical daily cycle of the differences model-obs (cf. p.70, step 2) of a specific class at a certain location), we tested as well a mean diurnal cycle of all four stations of the development group (Piotta, Robbia, Ulrichen, and Zermatt), "Mean all 4".

In Fig. 5.28 on p. 75, the impacts of using these different mean daily cycles are presented. Again, the data of days, belonging to the "gt_6"-class are shown.
All three statistical quantities showed at each station and both data sources an improvement in case the individual daily cycle was used for the correction. The differences between the "Mean all 4" and "Individual" corrections (cf. the remarks below the figure) were more pronounced at Piotta than at the other three stations, considering the RMS and MAE. Generally, the Mean Bias (MB) was at all four stations most influenced (i.e. improved) by the two different approaches (compared with the RMS and MAE). The reduction of the MBs reached values between 1.3-3.0 K. The differences found in the RMSs were smaller than 1.5 K, and in the MAEs maximal differences were around 1.4 K.

In Table 5.5 on p. 74 the statistical results for the "gt_6"-days depending on the different types of corrections are shown.
At Ulrichen, no matter which correction was applied ("Individual", "Mean all 4"), the results improved compared to the ones in case of the modelled data were only altitude-corrected. On the other hand, the mean results of Piotta, Robbia, and Zermatt showed a worsening in case of the mean diurnal cycle of all four stations of the development group ("Mean all4") was used to correct

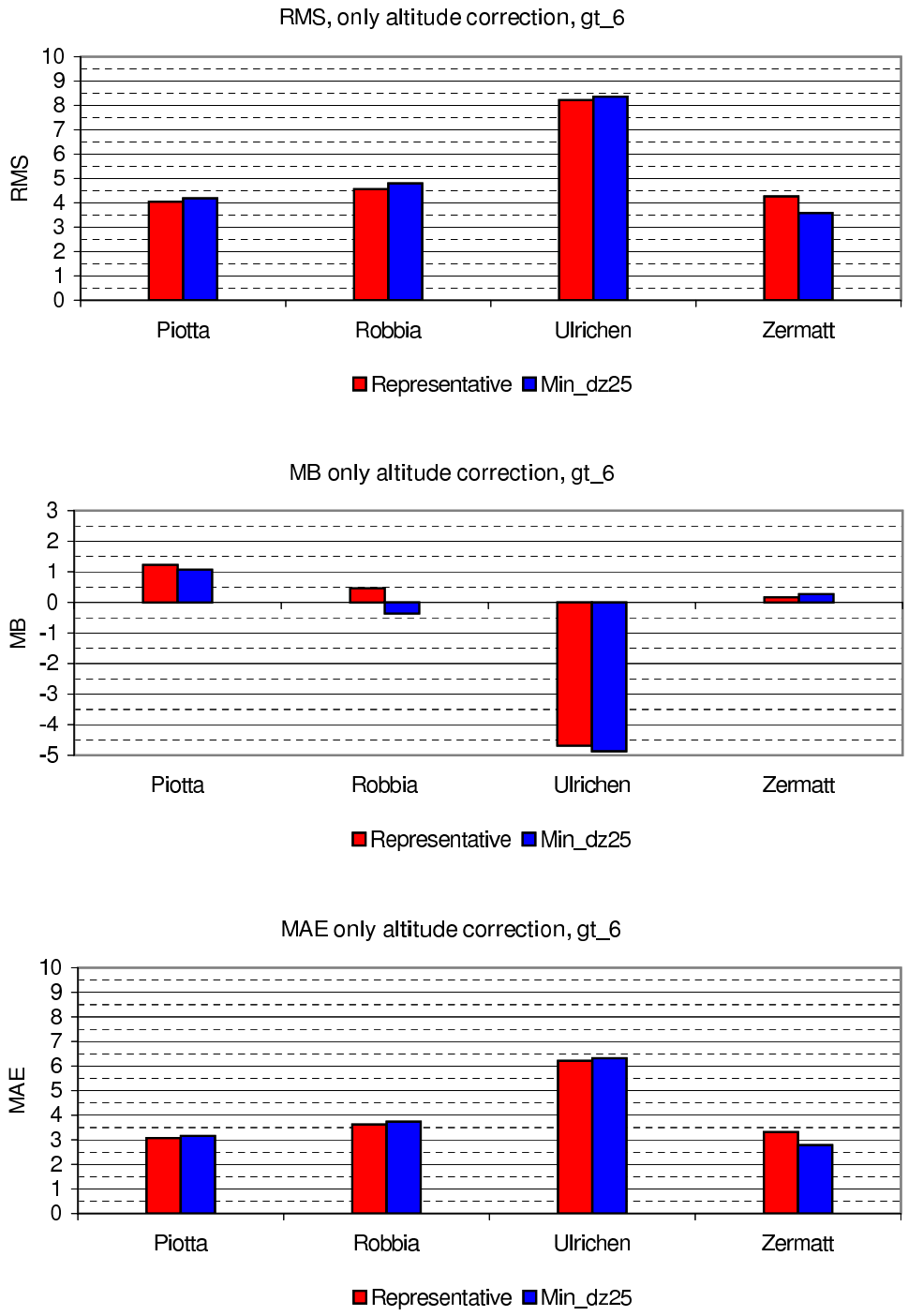


Figure 5.27: Statistical measures for the comparison between modelled and observed 2m-temperatures at the indicated stations. The modelled data was corrected by means of altitude. Different data sources ("rep", "min_dz25") are distinguished by colour as indicated. Only the data of days belonging to the "gt_6"-class were used.

Table 5.5: The statistical results depending on the different kinds of correction-types (cf. p. 72) are presented. Due to similar magnitudes of the statistical measures, mean values of the stations Piotta, Robbia, and Zermatt were built. The results at Ulrichen are presented separately.
Note that the shown values are means of both data sources ("rep", "min_dz25").
The abbreviation "only alt." refers to the results in case the modelled data were only altitude-corrected. For this analysis, the absolute values of the RMS and MB were taken.

	$\overline{Piotta, Robbia, Zermatt}$			Ulrichen		
	only alt.	mean all4	individual	*only alt.*	mean all4	individual
RMS	*4.2*	4.4	3.6	*8.3*	6.5	5.8
MB	*0.6*	2.3	0.3	*4.8*	3.0	0.6
MAE	*3.3*	3.6	2.9	*6.3*	5.0	4.9

the model data.

As one would expect, the analysis showed better results when the typical observed daily cycle of the specific station was applied instead of one mean daily cycle of all four stations of the development group.

However, this was an important result and would gain more importance whenever a correction should be carried out at locations , where no surface station is installed. Therefore, on such a place the observed "typical mean daily cycle" is not known and a mean of one or more (neighbouring) station(s) would have to be applied.

The further analysis of the development group was carried out, using the diurnal cycle of the specific station ("Individual").

Analysis of the "Individual" correction of the modelled 2m-temperatures

We present the impacts, in case the modelled temperatures were only corrected by means of an observed diurnal cycle (cf. p. 70).

Fig. 5.29 on p. 81 shows the statistical measures after the modelled 2m-temperatures were corrected with the individual observed daily cycles, i.e. data labelled with "Rep no spread" and "Dz25 no spread".

Considering the different data sources ("rep"↔"min_dz25"), the results were not uniform for all four stations. The RMS and MAE did not change much or became slightly worse at Piotta, Robbia, and Ulrichen, when the "min_dz25" gp was used compared with the "rep" gp (maximal +0.4 K at

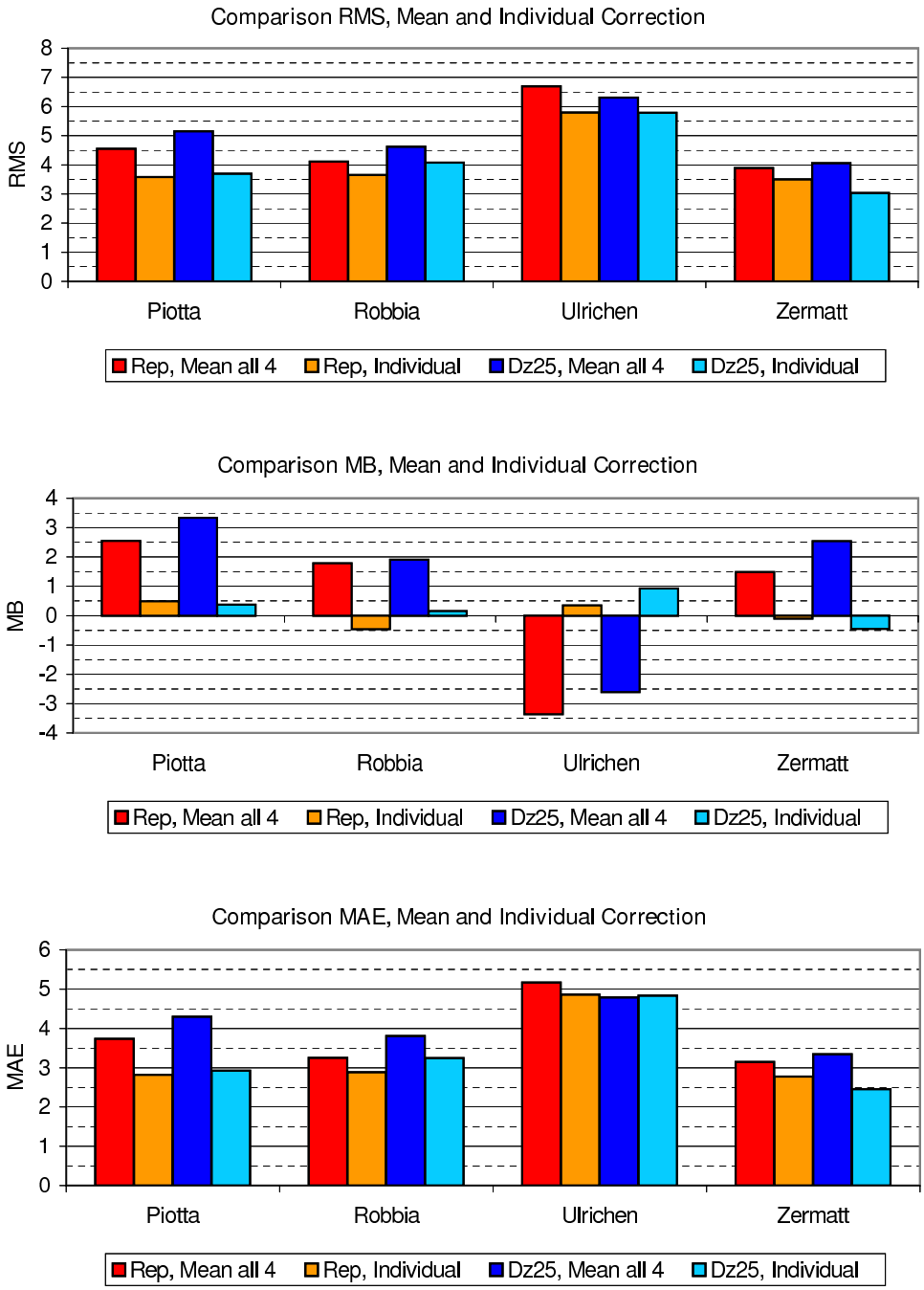


Figure 5.28: The statistical measures for different types of diurnal cycles used for the corrections.
"Mean all 4": The mean daily cycle of all four stations of the development group was used.
"Individual": The daily cycle of the specific station was applied.
"rep" gp: the bars in red and orange, "min_dz25" gp : the bars in blue.
The presented data (bar plots) only contained days, belonging to the "greater than 6"-class.

Robbia). Only at Zermatt an improvement of the RMS (-0.5 K) and of the MAE (-0.3 K) could be seen in case the "min_dz25" gp was used. The MB on the other hand got slightly better at Piotta ($\sim$ -0.1 K) and Robbia ($\sim$ -0.3 K), and showed a worsening at Ulrichen ($\sim$ +0.6 K) and Zermatt ($\sim$ +0. 4K). Note that the worsenings (positive values) at Ulrichen and Zermatt were larger than the improvements at Piotta and Robbia.

We can state that no great differences in these three statistical quantities could be found between the two data sources on "gt_6"-days (absolute differences $<$ 0.6 K).

Analysis and Impacts of the Spreading Procedure

We remember, that the day-to-day variability of the model's diurnal cycles was, compared to those of the observation, very small, i.e. in Fig. 5.14 on p. 53 we found a quite narrow band containing all diurnal cycles of the modelled values of the specific class. Therefore, we carried out a kind of data spreading to take into account these differences in the data distribution (cf. Methods, p. 32). First of all, we analysed the deviations, built by subtracting the mean model temperature (of all days of this class at 00 UTC) from the daily model's initial temperature (at 00 UTC) and the same of the observation, respectively. We observed correlations of the model's deviations and the ones of the observations. Table 5.6 on p. 79 presents the correlation coefficients, of the two sets of daily deviations.

These correlation coefficients, showing values between 0.70-0.89 at the "rep" and 0.58-0.90 at the "min_dz25" gp, indicated a connection between low (high) initial (00 UTC) temperatures of the model and low (high) initial temperatures recorded at the observations. Therefore, a spreading procedure as described in Methods on p. 32 was developed.

This "spreading" was performed within 5 steps:

1. The mean temperature and the standard deviation of the 00-UTC temperatures of all days belonging to a specific class were calculated. This was performed for the model data and the observations separately.

2. For each day, the deviations between the 00-UTC temperatures and the mean temperature of all days belonging to the specific class were computed. Note that these deviations were used to calculate the correlation coefficients shown in Table 5.6.

3. The standard deviation of the observed values was divided by the one of the modelled data in order to define a ratio.

4. This ratio, multiplied with the daily deviation between initial temperature (00-UTC) and the mean of all initial temperatures of this class led to an offset.

5. This offset, found in step 4, was added to tall (24) 2m-temperatures of the specific day.

Table 5.7 shows the correlation coefficients between the observed and modelled data for each class without and with applied spreading procedure, respectively. We found, that this spreading did not have a positive impact in general but rather depended on the individual station and class. **Overall it was seen, that this method led to worse results in the "lower than 2" and "2-6" class while the same spreading improved the results in the "greater than 6" class.**
Especially in the case of "lt_2"-class, the spreading procedure led to even worse results than the "raw", only altitude corrected data.

Fig. 5.29 on p. 81 presents the influences (i.e. statistical quantities: RMS, MB, and MAE) of the spreading procedure for the "greater than 6"-class. On these bar plots, the modelled data was modified with a diurnal cycle and where indicated with the spreading procedure (no need to say that the altitude-correction was always performed). Keep in mind that we decided for further analyses of the development group to use the "Individual" diurnal cycle (cf. p. 74).
At Piotta and Robbia, the spreading did not have a strong impact RMS and MAE ($< \sim 0{,}2$ K) while the impact at Zermatt depended on the data source. At the "rep" gp an improvement ($\sim$ -0.4 K) could be recorded, at the "min_dz25" gp a worsening ($\sim$ +0.2 K) was observed.
At Ulrichen, the spreading procedure led to better RMSs and MAEs ("rep": $\sim$ -0.8 K, "min_dz25": $\sim$ -0.5 K).
Note that the spreading procedure did not have any influence on the Mean Bias (MB).
The largest impacts and improvements occurred at Ulrichen while only little influences of the spreading procedure were found at Piotta and Robbia.

In Fig. 5.30 on p. 82 the changes in the absolute values of the statistical quantities before and after the modification are shown. The differences were built by subtracting the absolute value of the statistical quantity (RMS, MB, and MAE) before the modification (only the altitude correction was performed) from the one after the modification. This modification included a correction through the "Individual"- daily cycle and where indicated, the spreading procedure was performed. Though, positive values refer to a worsening, negative ones to an improvement through the ap-

plied corrections. The larger a "negative bar", the more effective the correction in this case. This enabled us to figure out the effectivity of the correction at a specific station and data source for "gt_6"-days.

For each station, the results are presented for both data sources, once with the spreading procedure and once without it (as indicated in the legend on Fig. 5.30). We saw a variable impact of the spreading procedure, depending on the station and data source.

The strongest impacts of the applied correction were observed at Ulrichen, no matter which data source or whether the spreading procedure was performed or not. At this station the spreading led to an improvement of the RMS and MAE at both data sources, but was more pronounced at the "rep" gp. At Ulrichen at the "rep" gp, the RMS and MAE improved another ~ 0.75 K when the spreading was performed while at the "min_dz25" gp the additional improvement was only around 0.45 K.

At Piotta and Robbia only marginal absolute differences were seen between the two different approaches (with/without spreading), i.e. RMS: $< \sim 0.2$ K, MAE: $< \sim 0.1$ K. At Zermatt at the "rep" gp, the spreading procedure exhibited a positive influence on the RMS and MAE ($\sim$ -0.4 K), but not at the "min_dz25" gp ($\sim$ +0.2 K).

To summarise, for these statistical measures the additional improvement through applying the spreading procedure was generally small compared with the impact of the correction through the "Individual" diurnal cycle.

The strongest positive impacts of the spreading were found at Ulrichen, only marginal impacts were observed at Piotta and Robbia.

The Tables A.1-A.8 (p. 124 et seq.) show the statistical analyses in more details, also for the other two classes ("lt_2" and "gt_6").

The impacts of the "Spreading Procedure" on Diurnal Cycles and Scatter Plots at Ulrichen

In the Figures 5.31 and 5.32 on p. 84/85 the typical daily cycles of Ulrichen (class: "greater than 6") without and with the spreading procedure can be seen, respectively. On the upper panel the different daily cycles are shown.

In Fig. 5.31 the altitude corrected modelled 2m-temperatures without any diurnal cycles and the modified model data, corrected with the "Individual" differences (model↔observation) are shown. All these daily cycles were situated closely together in a narrow band. This structure was successfully spread out in case the spreading procedure was performed, as seen in Fig. 5.32.

Table 5.6: Correlation coefficients of the model's deviations (cf. 5.8) and the observations. For further remarks we refer to the text on p. 76.

lower than 2	Representative		Smallest_dz_all25	
Station	Corr.	n	Corr.	n
Piotta	0.766	23	0.860	25
Robbia	0.893	22	0.806	17
Ulrichen	0.701	23	0.851	29
Zermatt	0.747	21	0.919	11

2 - 6	Representative		Smallest_dz_all25	
Station	Corr.	n	Corr.	n
Piotta	0.887	29	0.845	26
Robbia	0.865	51	0.907	19
Ulrichen	0.795	41	0.705	45
Zermatt	0.893	33	0.872	19

greater than 6	Representative		Smallest_dz_all25	
Station	Corr.	n	Corr.	n
Piotta	0.779	73	0.666	63
Robbia	0.734	74	0.583	45
Ulrichen	0.755	94	0.768	83
Zermatt	0.857	99	0.743	77

Table 5.7: The correlation coefficients of the corrected modelled data (for each class) and the observed data are shown. Note, that only these data points were taken into account, which were actually corrected, i.e. days representing the specific class.
In **bold** letters, where an improvement through the spreading resulted.
In *italic* letters, when the spreading led to even worse results than in case of only the altitude correction was performed.
The abbreviation "only alt." refers to these results in case of the modelled data was only altitude corrected.

lower_than_2	Representative			Smallest_dz_all25		
Station	only alt.	no spread	with spread	only alt.	no spread	with spread
Piotta	0.825	0.833	*0.822*	0.844	0.849	**0.854**
Robbia	0.914	0.917	*0.909*	0.787	0.785	*0.777*
Ulrichen	0.776	0.805	*0.760*	0.849	0.858	*0.822*
Zermatt	0.817	0.825	**0.830**	0.842	0.860	*0.834*

2-6	Representative			Smallest_dz_all25		
Station	only alt.	no spread	with spread	only alt.	no spread	with spread
Piotta	0.901	0.905	0.905	0.827	0.833	*0.824*
Robbia	0.853	0.890	0.874	0.893	0.905	0.900
Ulrichen	0.832	0.853	0.839	0.750	0.779	0.762
Zermatt	0.911	0.922	**0.924**	0.872	0.878	*0.862*

greater than 6	Representative			Smallest_dz_all25		
Station	only alt.	no spread	with spread	only alt.	no spread	with spread
Piotta	0.670	0.693	**0.776**	0.726	0.698	**0.710**
Robbia	0.654	0.782	**0.811**	0.630	0.708	**0.722**
Ulrichen	0.683	0.686	**0.803**	0.644	0.679	**0.783**
Zermatt	0.775	0.834	**0.879**	0.584	0.740	**0.754**

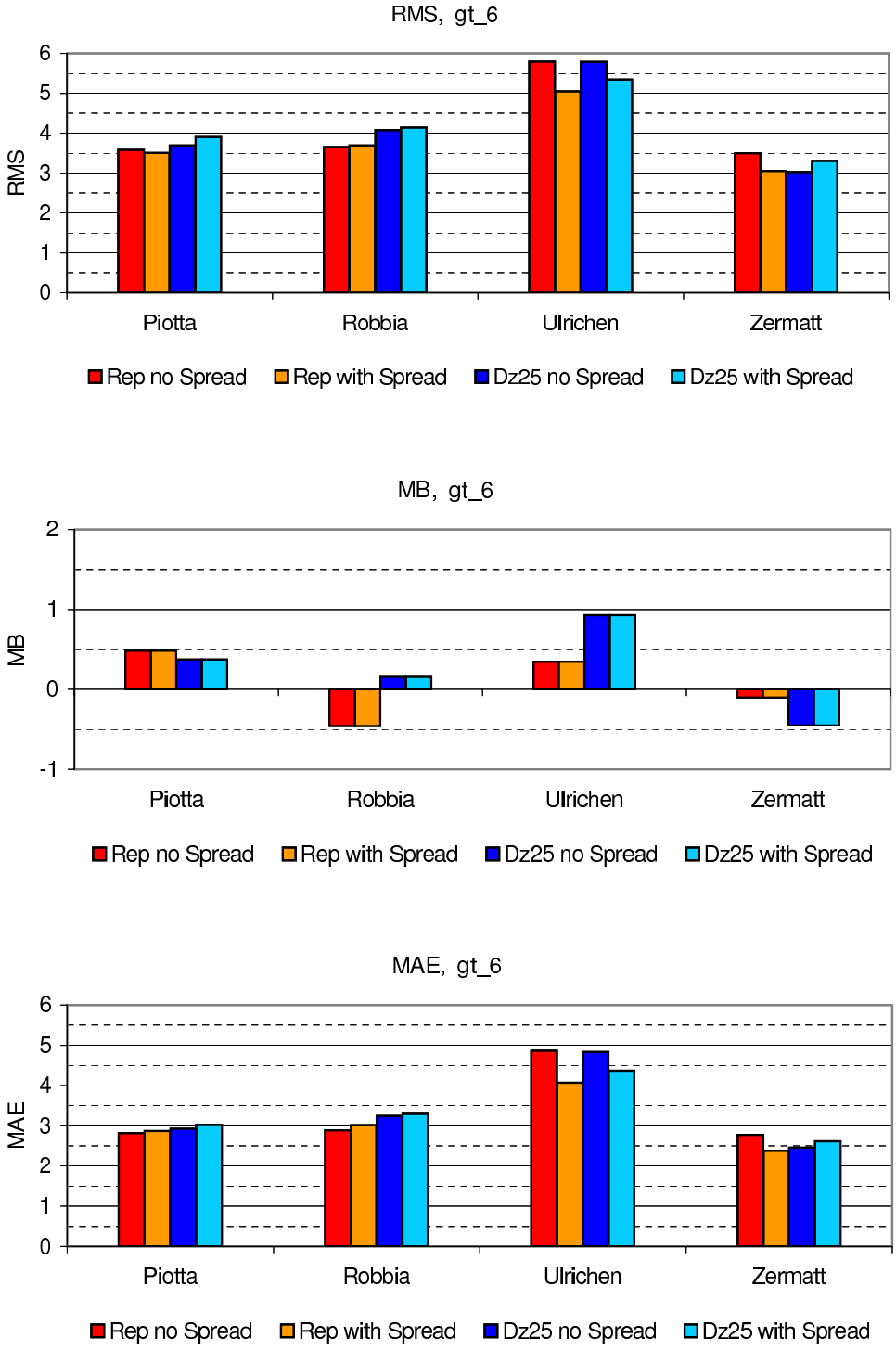


Figure 5.29: The impacts of the spreading procedure. The three different statistical quantities (RMS, MB, and MAE) are shown for each station, data source and the two approaches (with/without spreading).

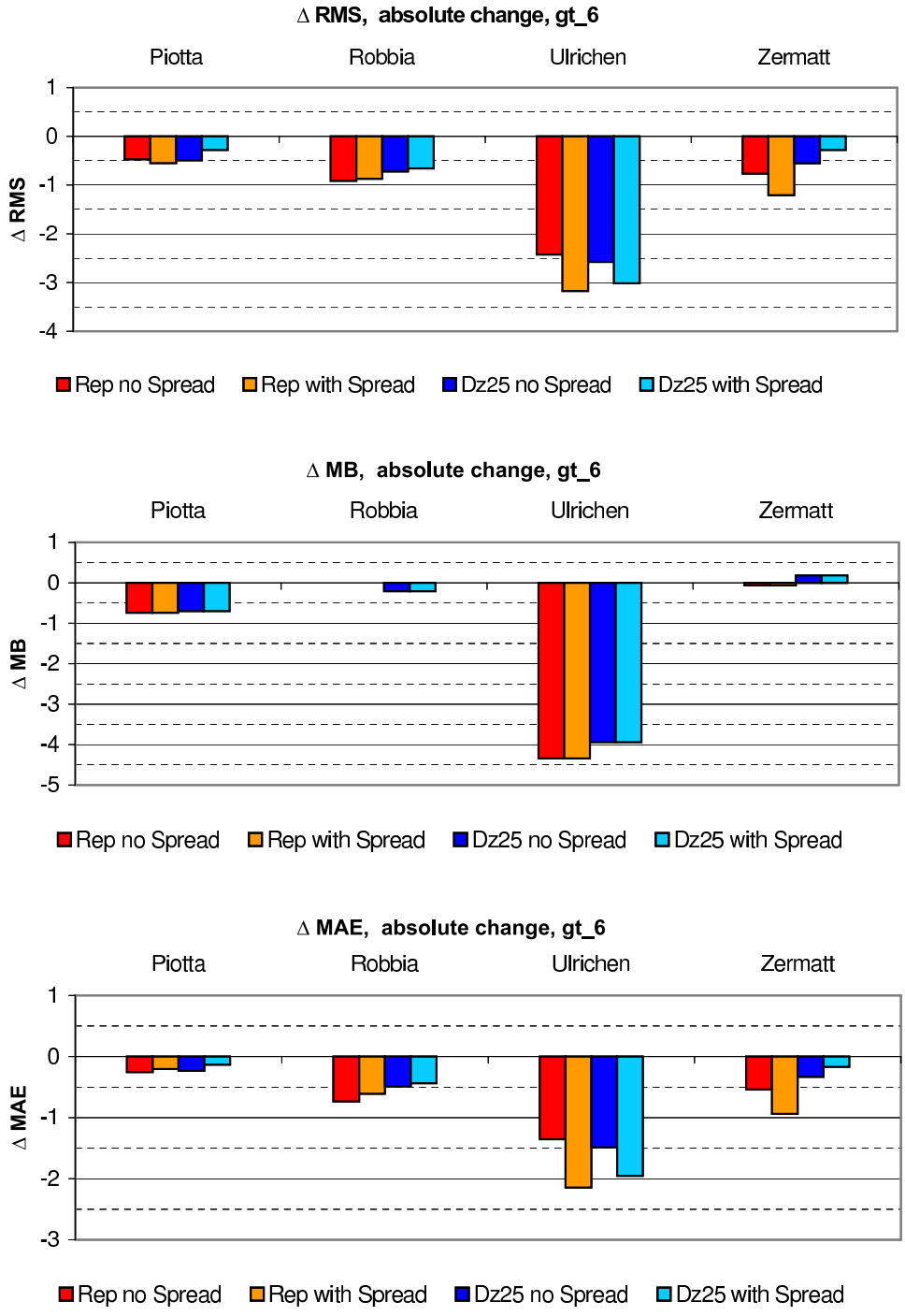


Figure 5.30: The impacts of the spreading procedure. The changes in the "absolute values" of the different statistical quantities are shown.
(example: $\Delta RMS = |RMS_{modif.}| - |RMS_{only\ alt.}|$, the subscript *only alt.* refers to only altitude corrected modelled data and *modif.* to the modificated model data (correction with diurnal cycle and where indicated with the spreading procedure)).
Negative values indicate an improvement through the modifications compared with the only altitude corrected data. "rep" gp: the bars in red and orange, "min_dz25" gp: the bars in blue colours.

On the other hand, although the ranges of the two datasets matched much better after applying the spreading procedure, the deviations between the two datasets (shown in the lower panels) did not clearly become smaller in Fig. 5.32. The reason for this behaviour were wrongly determined initial temperatures (00-UTC values) by adding the offset described on p. 77.

In the Figures 5.33 and 5.34 on p. 86/87, the scatter plots for Ulrichen (class: "greater than 6") are presented.
These Figures show the hourly 2m-temperatures of days belonging to the "greater than 6"-class. Fig. 5.33 shows in the upper panel the unmodified (only altitude corrected) modelled temperatures, while in the lower panel the model data was corrected with the mean observed deviations (model↔observation).
In Fig. 5.34 the same is shown, but this time the modelled 2m-temperatures in the lower panel were additionally modified with the spreading procedure.
We want to point out the rotation of the whole data cloud toward the 45°-line. This rotation was more pronounced in case the spreading procedure was performed. We think this is of importance because the correlation coefficient is a measurement for the similarity of the behaviour of two datasets, but a slope close to "1" on the other hand told us, that our correction of the modelled temperatures had effectively a positive impact in conforming the modelled values to the observed ones.
On the other side, the drawback of this procedure was a broader distribution of the data points, i.e. the cloud of data points became diffuser.

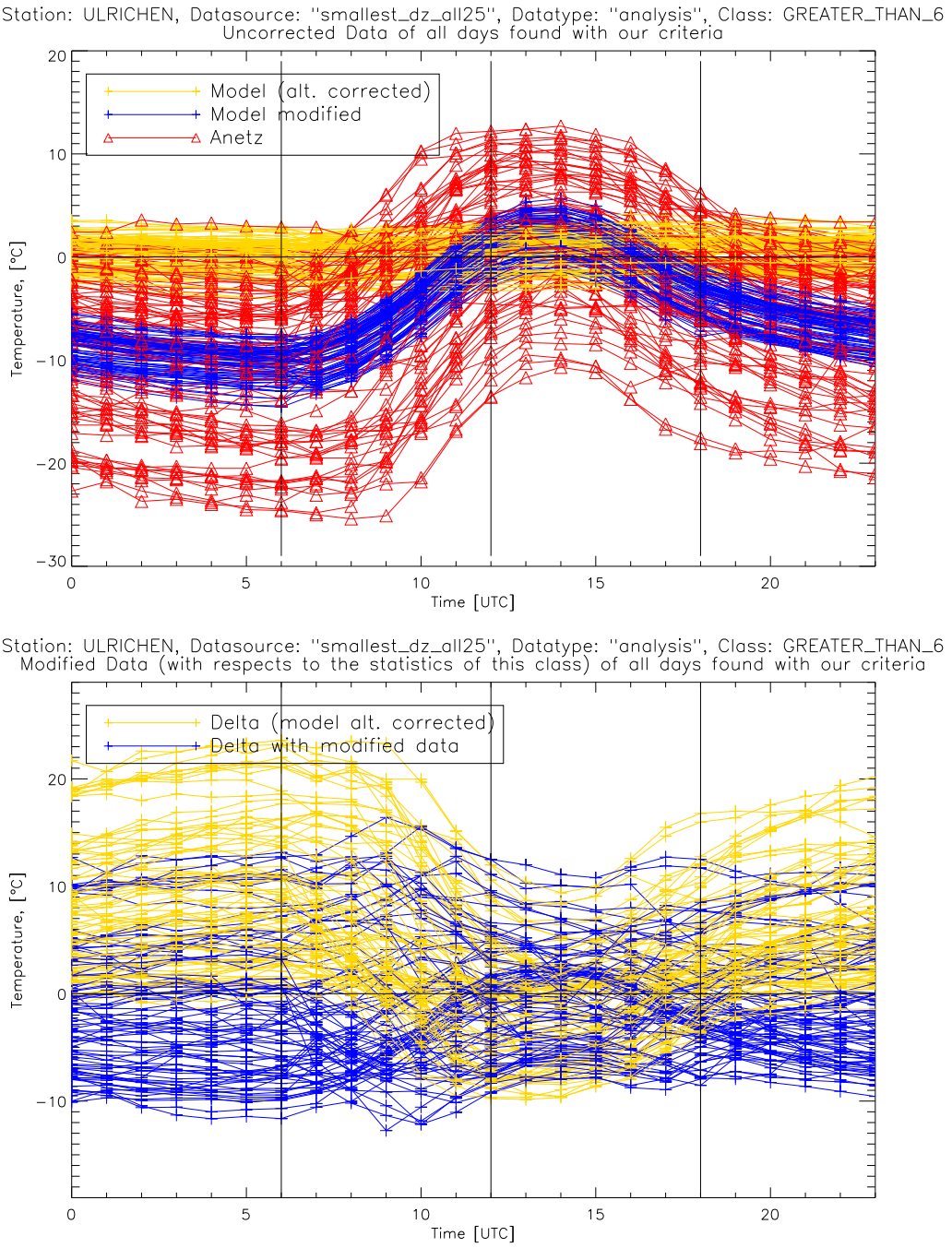


Figure 5.31: Daily cycles of Ulrichen, main class: "greater than 6", without spreading procedure.
Yellow: altitude corrected model data without any further modifications.
Blue: fully corrected model data (i.e. the mean daily cycle of this class is added to the altitude corrected model data).
Red: observed values of the surface station.
Lower panel: the differences between modelled and observed data (model - obs.).
Without spreading

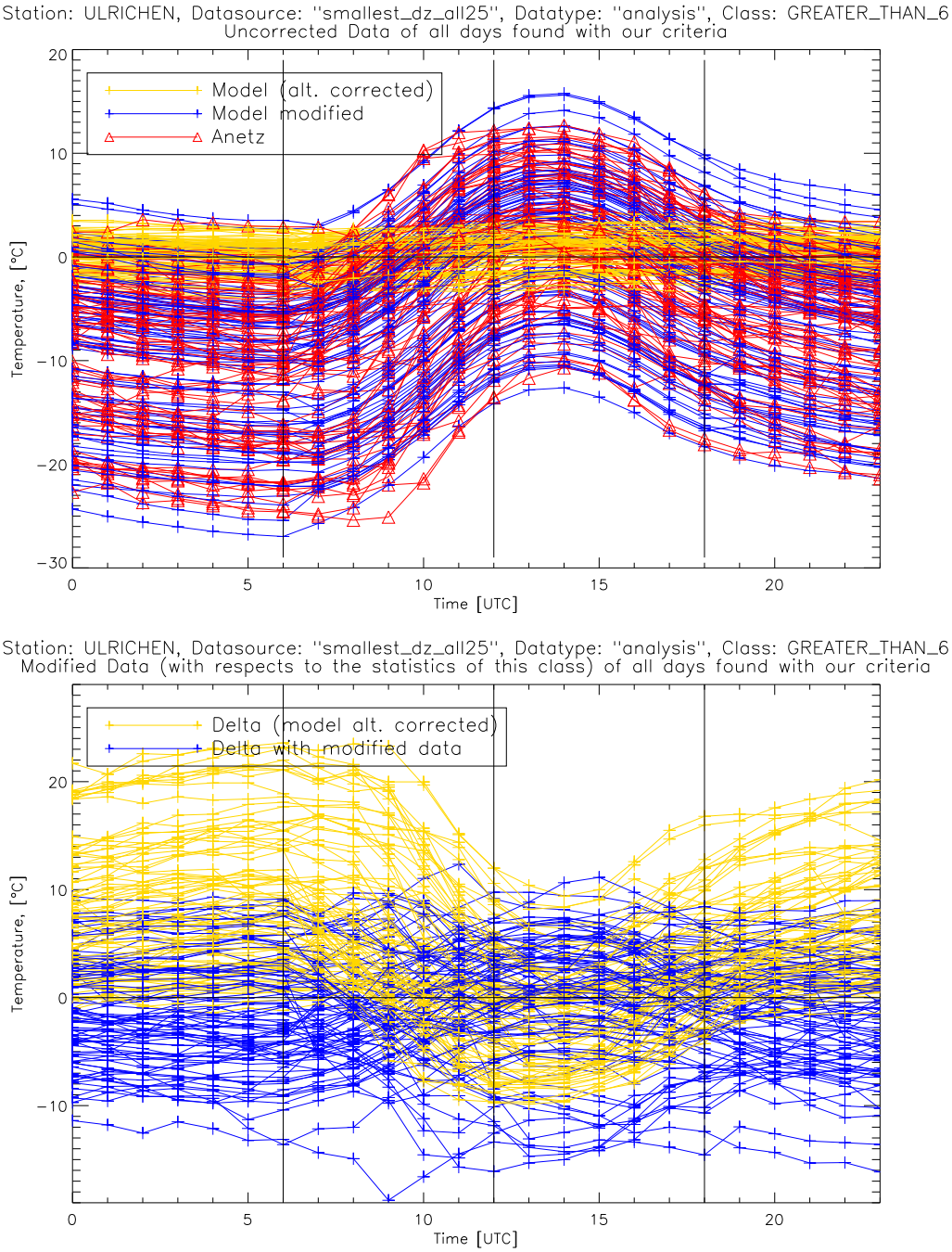


Figure 5.32: Daily cycles of Ulrichen, main class: "greater than 6", with spreading. Additionally to the modification with a mean diurnal cycle, the spreading procedure was applied. For further remarks we refer to Fig. 5.31.
With spreading

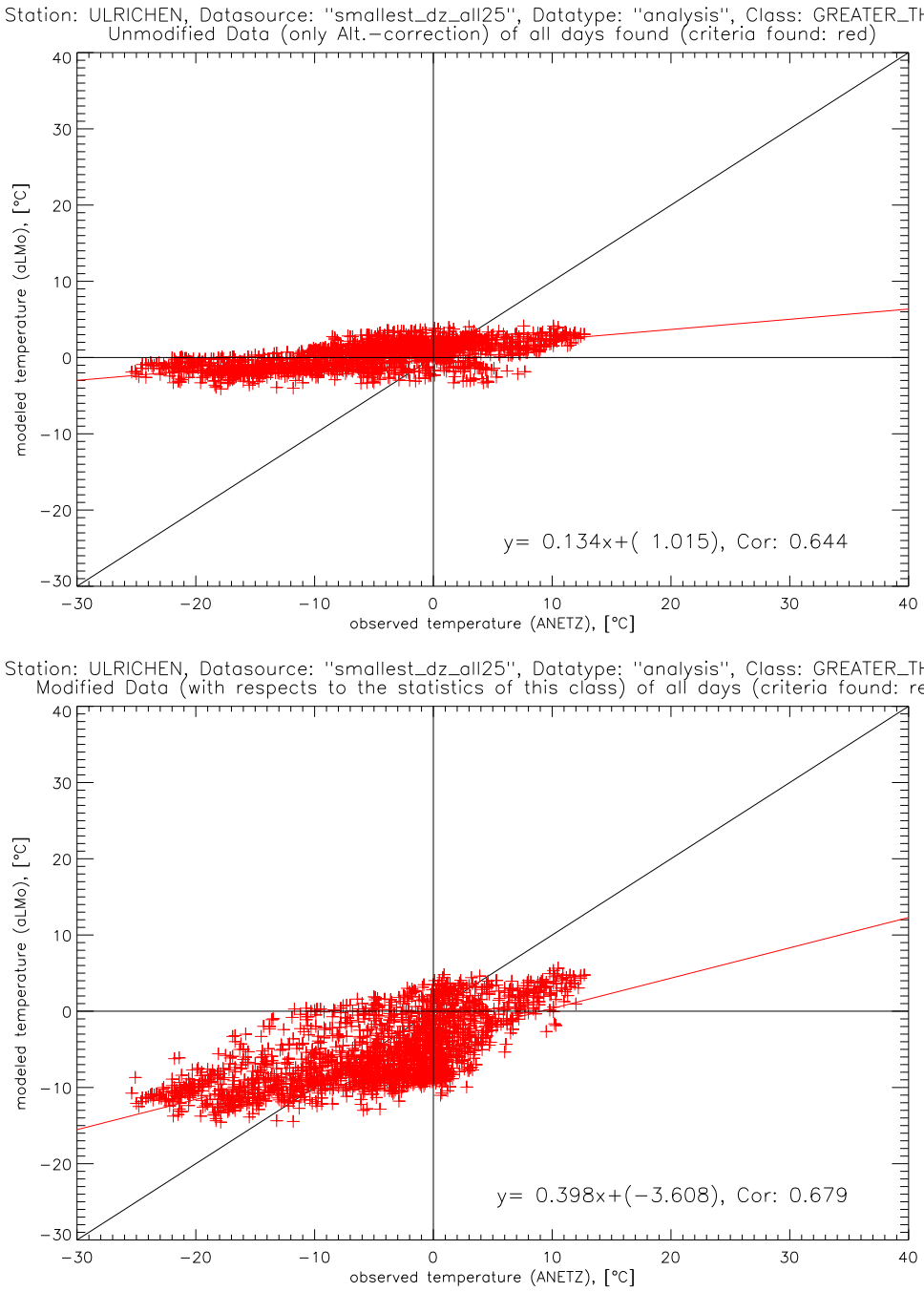


Figure 5.33: Comparison between modelled and observed 2m-temperatures at Ulrichen on "greater than 6"-days.
Here, only the data of these days are shown, which actually belong to the specific class.
Upper panel: only the altitude corrected model data.
Lower panel: modified model data, i.e. the mean deviations (model-observation) of the "greater than 6"-class were added.
No Spreading

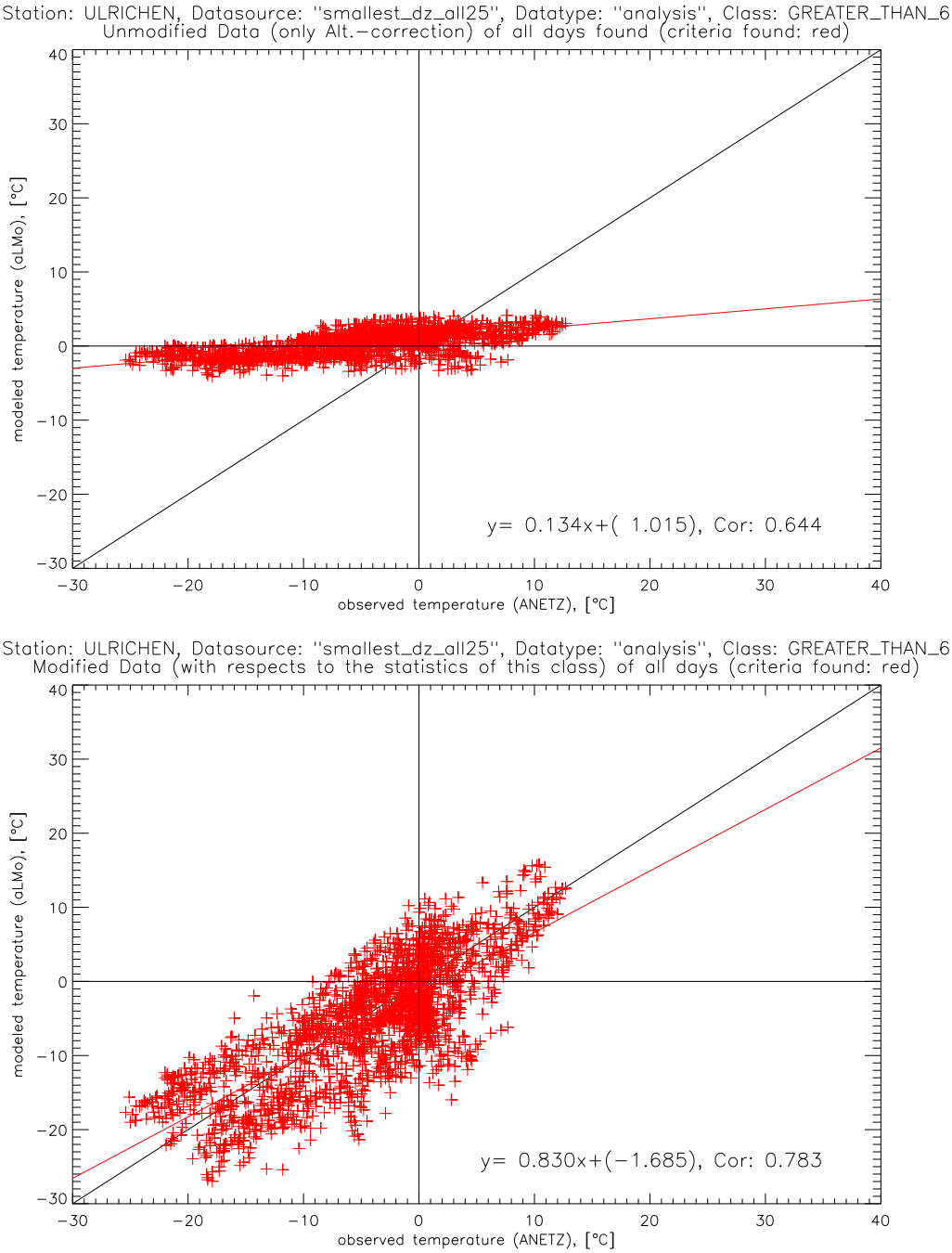


Figure 5.34: Comparison between modelled and observed 2m-temperatures at Ulrichen on "greater than 6"-days.
This time, the model data shown in the lower panel was additionally corrected by applying the spreading procedure.
For further remarks we refer to Fig. 5.33. With Spreading

5.9 Test on an Independent Dataset

In this section, the impact of the suggested corrections on the stations of the control group (Comprovasco, Engelberg, Samedan, and Scuol) are presented. In the course of this final analysis explicitly no piece of information from the specific observations at these locations was incorporated in correcting the modelled data.
Moreover, the influence of the previously introduced spreading procedure is shown. Further, we present the results of using different mean daily cycles for the correction of the modelled data. The aim of carrying out this final analysis using different types of diurnal-cycles was to determine the sensitivity of the statistical quantities (RMS, MB, and MAE) to varying sources of mean diurnal cycles (cf. Methods, p. 33 and Table 4.2).
Because we avoided the use of observational data in the correction procedure of the specific station (control group) we used data, i.e. diurnal cycles of the development group in order to correct the modelled 2m-temperatures. Beside a mean daily cycle of all four stations (Piotta, Robbia, Ulrichen, and Zermatt) we calculated mean values of Piotta, Robbia, and Zermatt ("Mean_PioRobZer"-cycle). This combination seemed to be appropriate regarding Figures 5.16-5.18 on p. 55-57. There (especially at the "gt_6"-class) we found on the lower panel, that Piotta, Robbia, and Zermatt possessed rather similar diurnal cycles of the differences (model-observation), used in the correction procedure while Ulrichen behaved differently.

Impacts of the different Diurnal Cycles and the Spreading Procedure

The following comments describe the impact of spreading procedure and determine which of the different daily cycles (used for the correction) was most appropriate for the specific station. For more detailed information we refer to the Tables B.1-B.4 on p. 133 et seq. (Appendix). Note, this time the full-year dataset was used to derive the RMS, MB, and MAE and not only these days, belonging to a specific class. While the impacts (positive, negative) of the applied corrections were of course not affected by this larger dataset, the results were less pronounced (see below) due to incorporating days when no correction at all took place.

Evaluation of the different Diurnal Cycles

For each station the most appropriate type of diurnal cycle was found in the Tables B.1-B.4 (Appendix).

Comprovasco: At both data sources ("rep", "min_dz25") the Mean_PioRobZer diurnal cycle was the most appropriate one and led to better statistical quantities than in case of only the altitude correction was performed.

Engelberg: While at the "rep" gp the Mean_all4 daily cycle led to the most positive results, at the "min_dz25" gp the Mean_PioRobZer diurnal cycle was clearly most appropriate. Note that the applied corrections led generally to worse results than in case of only an altitude correction of the model data.

Samedan: While at the "rep" gp the Mean_all4 daily cycle led to positive results, the typical diurnal cycle of Ulrichen (Ulr) was clearly most appropriate at the "min_dz25" gp. On both data sources an improvement of the statistical values through the applied correction could be found.

Scuol: On both data sources, the Mean_all4 gp led to the most positive results.

Comments on the figures, showing the RMS, MB, and MAE, using different diurnal cycles for the corrections

In the Figs. 5.35-5.37, the statistical quantities (RMS, MB, and MAE) for each station, data source, and correction procedure (with/without spreading) at "greater than 6"-days are shown. The differences between the Figures (5.35-5.37) are the kinds of diurnal-cycles that were used to correct the modelled 2m-temperatures (cf. Methods, Table 4.2 on p. 34). The three figures show rather similar results, although they are not equal in their magnitudes.

Data Sources

No matter which diurnal cycle was applied to correct the modelled 2m-temperatures (Mean_all4, Mean_PioRobZer or Ulr), at Engelberg, Samedan, and Scuol the use of the "min_dz25" gp led to better RMSs and MAEs. The RMSs improved up to 1.1 K, (Samedan, "Ulr") and the MAEs up to 0.9 K (Samedan, "Ulr").

The impacts on the MBs were not clear but in case the diurnal cycle of "Ulr" was chosen to correct the model data, better MBs were found at these three stations at the "'min_dz25" gp.

At Comprovasco, the differences were generally smaller and indicated a worsening in case of the "min_dz25" gp was used. The RMS increased maximal +0.46 K (Mean_PioRobZer), and the MAE +0.28 K (both values became worse).

The magnitudes of the differences depended on the station and the type of diurnal cycle used for the correction.

To summarise, on three out of four stations of the control group, the chosen statistical measures indicated an improvement in case the "min_dz25" gp was chosen as data source.

Spreading Procedure

At Samedan, the spreading had a positive impact and let the RMS decrease by maximally -0.67 K ("Ulr") and the MAE by maximally -0.32 K ("Ulr"). The positive impacts at this station were more pronounced at the "rep" gp.

At Comprovasco and Scuol, the spreading procedure had a negative impact, no matter which data source and type of diurnal cycle was applied for the correction (Mean_all4, Mean_PioRobZer, Ulr). The RMSs increased by maximal +0.36 K (Scuol, Mean_all4), the MAEs by maximal +0.33 K (Scuol, Mean_PioRobZer).

At Engelberg, the spreading procedure had no strong influence. The only marginal changes were in the order of ±0.05 K (RMS) and ±0.07 K (MAE). Note, that the spreading procedure had no impact on the Mean Bias (MBs).

To summarise, the spreading procedure was not in general successful in improving the chosen statistical measures at the four stations of the control group. We can state, that this approach, which was more promising at the development group had not the same positive impact when transfered to other locations.

Comments on the figures, showing the changes in the statistical quantities, in case of applying different types of diurnal cycles for the corrections

The Figures 5.38-5.40 present the effectivity of the different corrections at all four stations, data sources and types of diurnal cycles chosen for the modification of the modelled 2m-temperatures. On these figures, the change of the three statistical quantities between only altitude corrected and the additionally modified data can be seen. The absolute changes were calculated by subtracting the absolute errors of the altitude corrected data from the ones of the modified data (added diurnal cycle and spreading, respectively). Note that positive values indicate a worsening, negative ones an improvement of the statistical quantity through the applied corrections. Though, the larger a "negative" bar, the greater the positive impact of the correction on the considered statistical quantity.

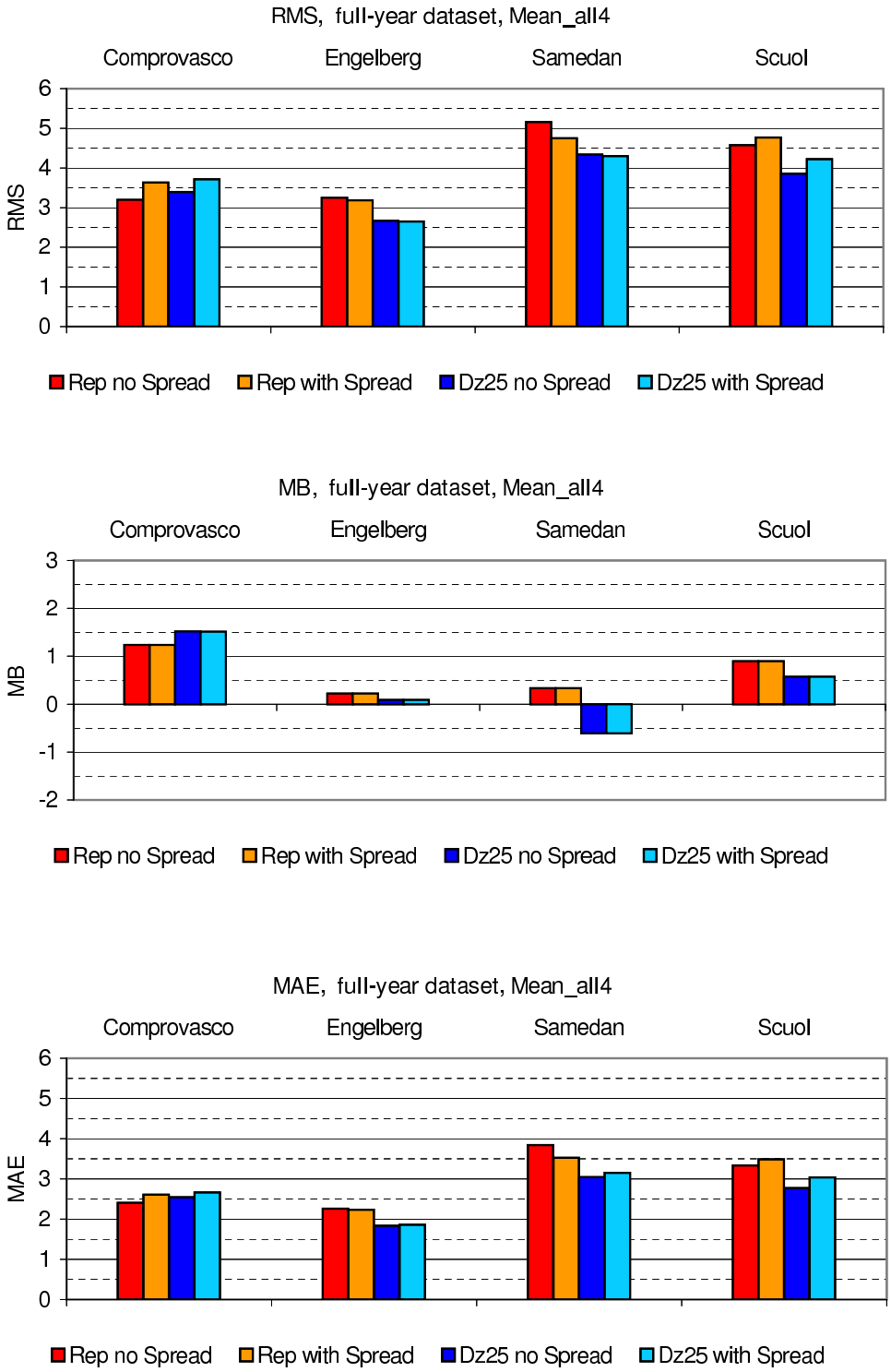


Figure 5.35: The statistical quantities (RMS, MB, and MAE) at all four stations of the control group. The model data was corrected with a mean diurnal cycle of all four stations of the development group ("Mean all 4").

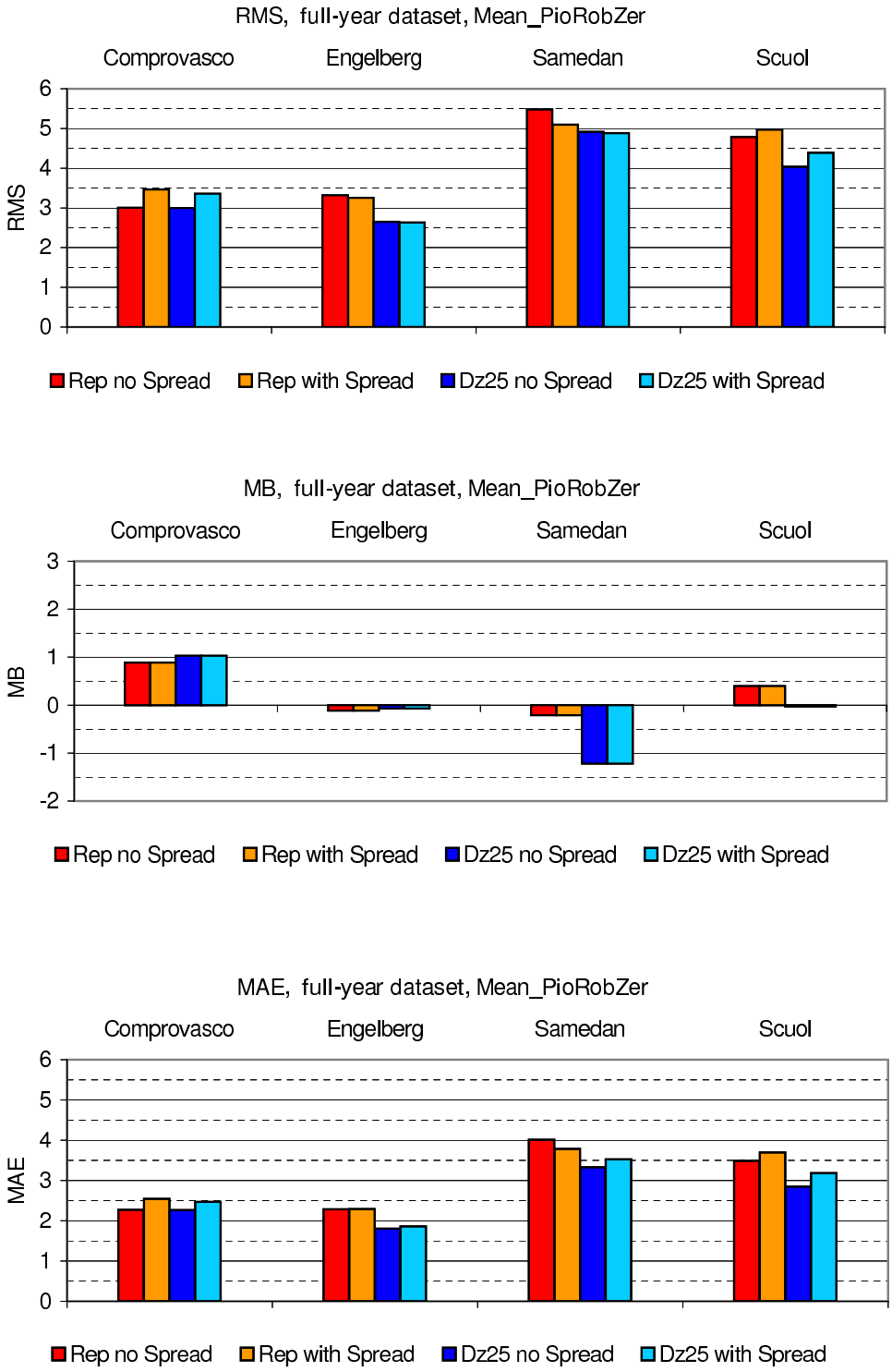


Figure 5.36: The statistical quantities (RMS, MB, and MAE) at all four stations of the control group. The model data was corrected with a mean diurnal cycle of Piotta, Robbia, and Zermatt ("Mean PioRobZer").

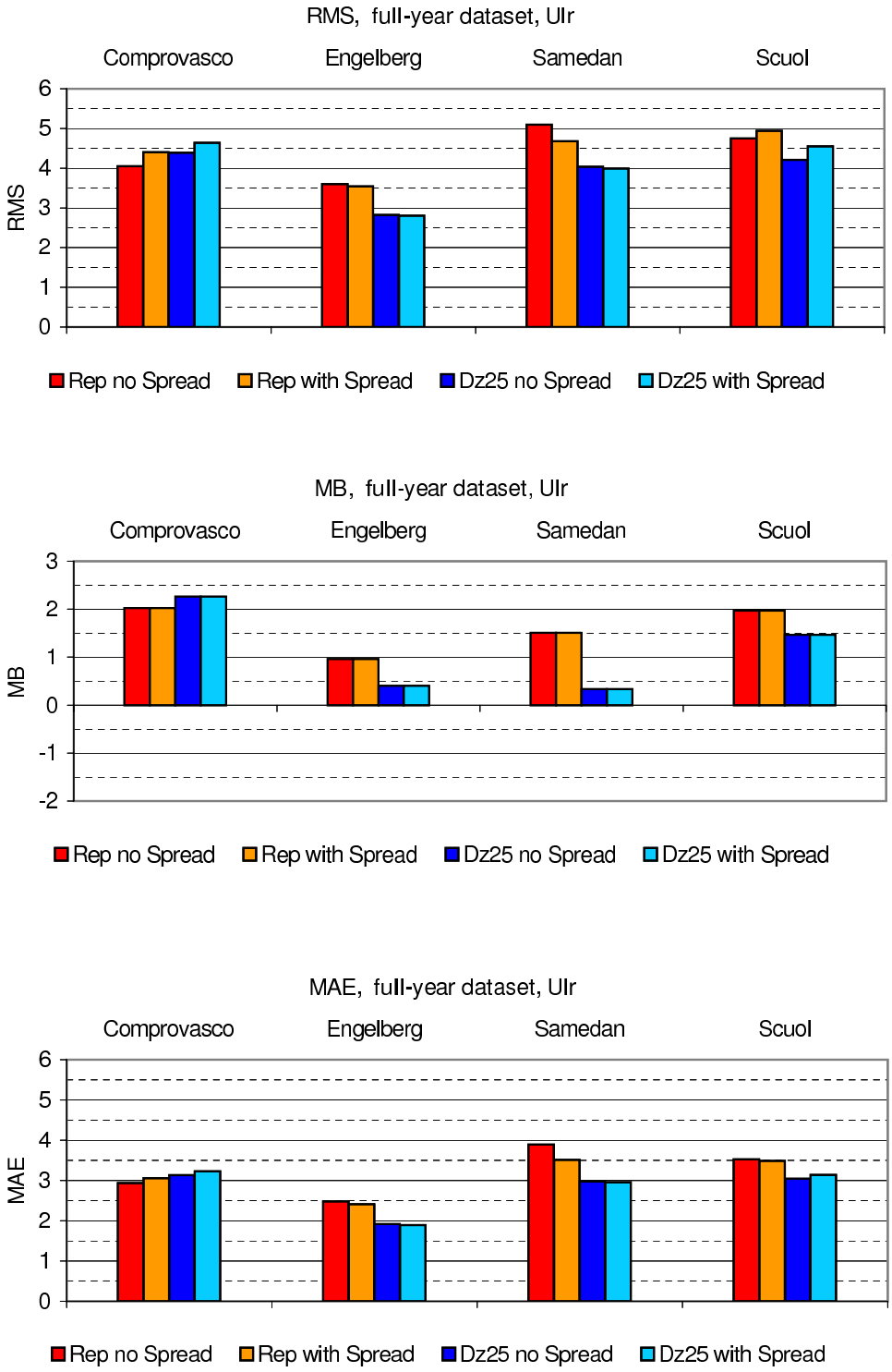


Figure 5.37: The absolute errors of all three statistical quantities (RMS, MB, and MAE) at all four stations of the control group. The model data was corrected with the typical diurnal cycle of Ulrichen ("Ulr").

At Samedan, the impacts of the correction on the RMSs and MAEs were positive, no matter which diurnal cycle was chosen to correct the model data. Most effective was the correction in case the typical daily cycle of Ulrichen ("Ulr") was used. Note that Ulrichen and Samedan were already a pair in Table 3.1 on p. 16 (Data). Reductions of the RMS > 1 K and of the MAE > 0.5 K were reached compared with the impacts in case the model data was only altitude-corrected.
At the other three stations (Comprovasco, Engelberg, and Scuol) the "Ulr" diurnal cycle was not appropriate and led to a general worsening of the RMSs, MBs, and MAEs. At Comprovasco and Scuol, the results were ambiguous. We found, that the "Mean_PioRobZer" daily cycle was most appropriate, i.e. led to the smallest worsenings. The impacts were generally small ($<\sim 0.3$ K).
At Engelberg, no matter which type of diurnal cycle was chosen, the applied corrections had negative impacts on the selected statistical measures. These negative influences were most pronounced in case of the "Ulr" daily cycle was used. Applying the other two diurnal cycles (Mean_all4, Mean_PioRobZer), only marginal changes (generally worsenings) smaller than 0.1 K were found.
In summary, for one station (Samedan), an appropriate diurnal cycle ("Ulr") could be assigned, which led to improved statistical quantities. At the other three stations, no appropriate daily cycle in order to correct the modelled 2m-temperatures was found. The results were ambiguous and not satisfying.
We can state, it is difficult to assign an appropriate diurnal cycle for the correction of the modelled data to a certain location. This approach only led at one station out of four to clearly improved results.

Impact of the Spreading Procedure at Scuol and Samedan

The Figures 5.41 and 5.42 on p. 99/100 present the results of this analysis at Scuol.
In both figures, a mean daily cycle of all four stations of the development group (Piotta, Robbia, Ulrichen, and Zermatt) was used to correct the modelled data, i.e. "Mean_all4". In Fig. 5.41, no spreading procedure was applied, though only the mean daily cycle was added to the altitude corrected modelled data. In Fig. 5.42, additionally to the modification through the mean daily cycle, the spreading procedure was performed.
The Figures 5.43 and 5.44 on p. 101/102 present the same kind of results as described above but for Samedan. This time the typical daily cycle of Ulrichen ("Ulr") was used for the correction of the modelled data.

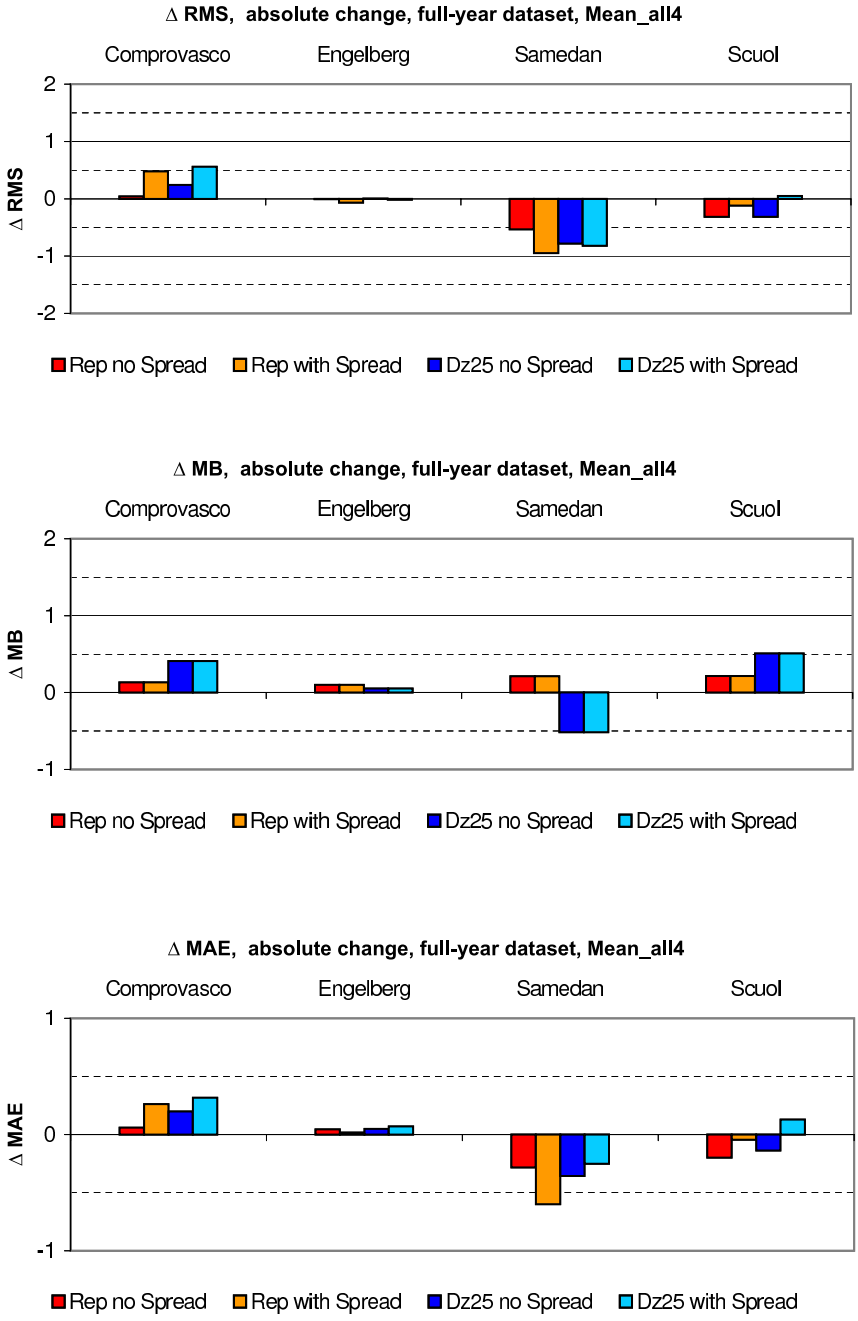


Figure 5.38: The differences in the absolute values of all three statistical quantities (RMS, MB, and MAE), before and after the applied modifications
(example: $\Delta RMS = |RMS_{modif.}| - |RMS_{only\ alt.}|$, the subscript *only alt.* refers to only altitude corrected modelled data and *modif.* to the modificated model data (correction with diurnal cycle and where indicated with the spreading procedure)).
Negative values indicate an improvement through the modifications compared with the only altitude corrected data. Note the model data was modified with a mean diurnal cycle of all four stations of the development group.
Mean all 4

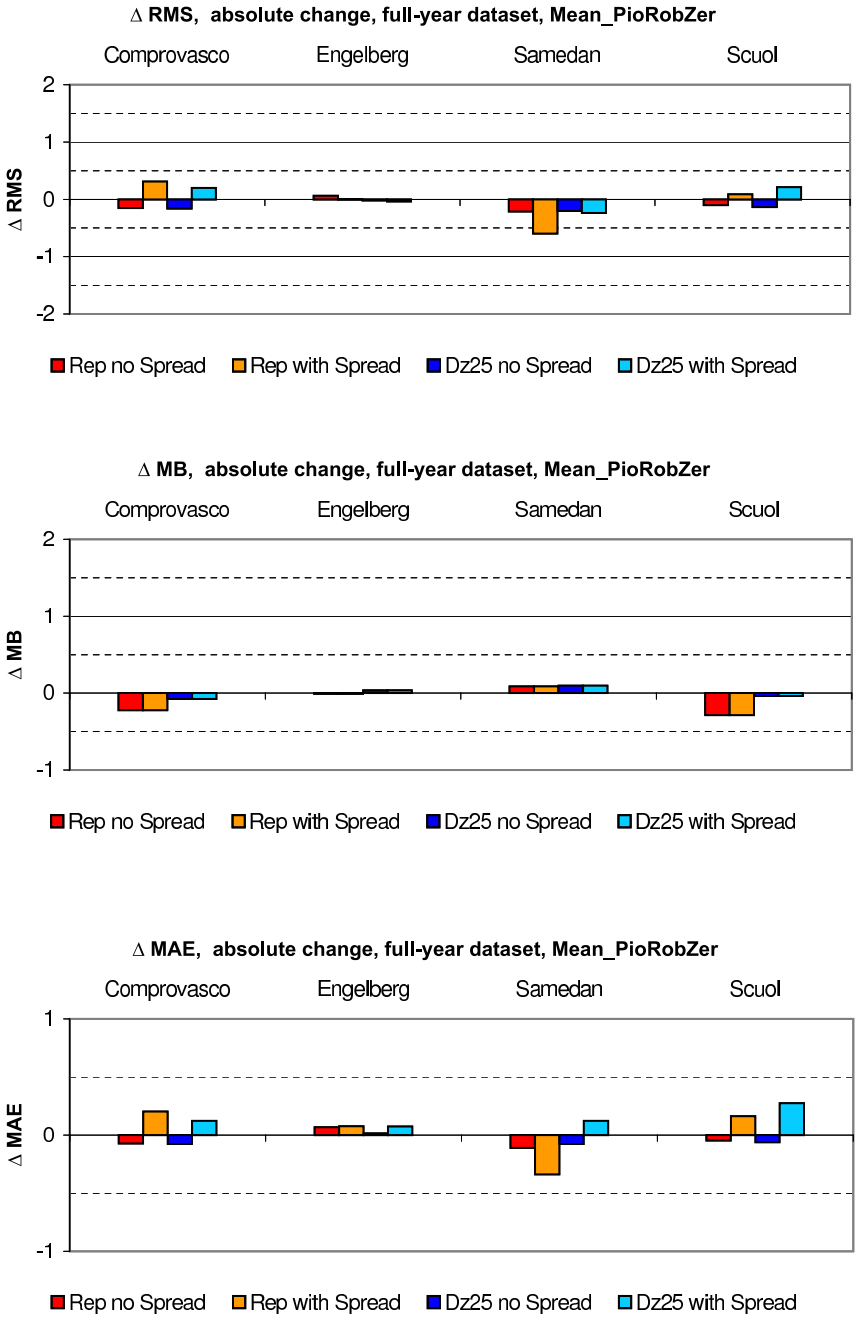


Figure 5.39: The differences in the absolute values of all three statistical quantities (RMS, MB, and MAE), before and after the applied modifications (cf. example in the remarks on Fig. 5.38). Negative values indicate an improvement through the modifications compared with the only altitude corrected data. Note, the model data was corrected with a mean diurnal cycle of Piotta, Robbia, and Zermatt.
Mean PioRobZer

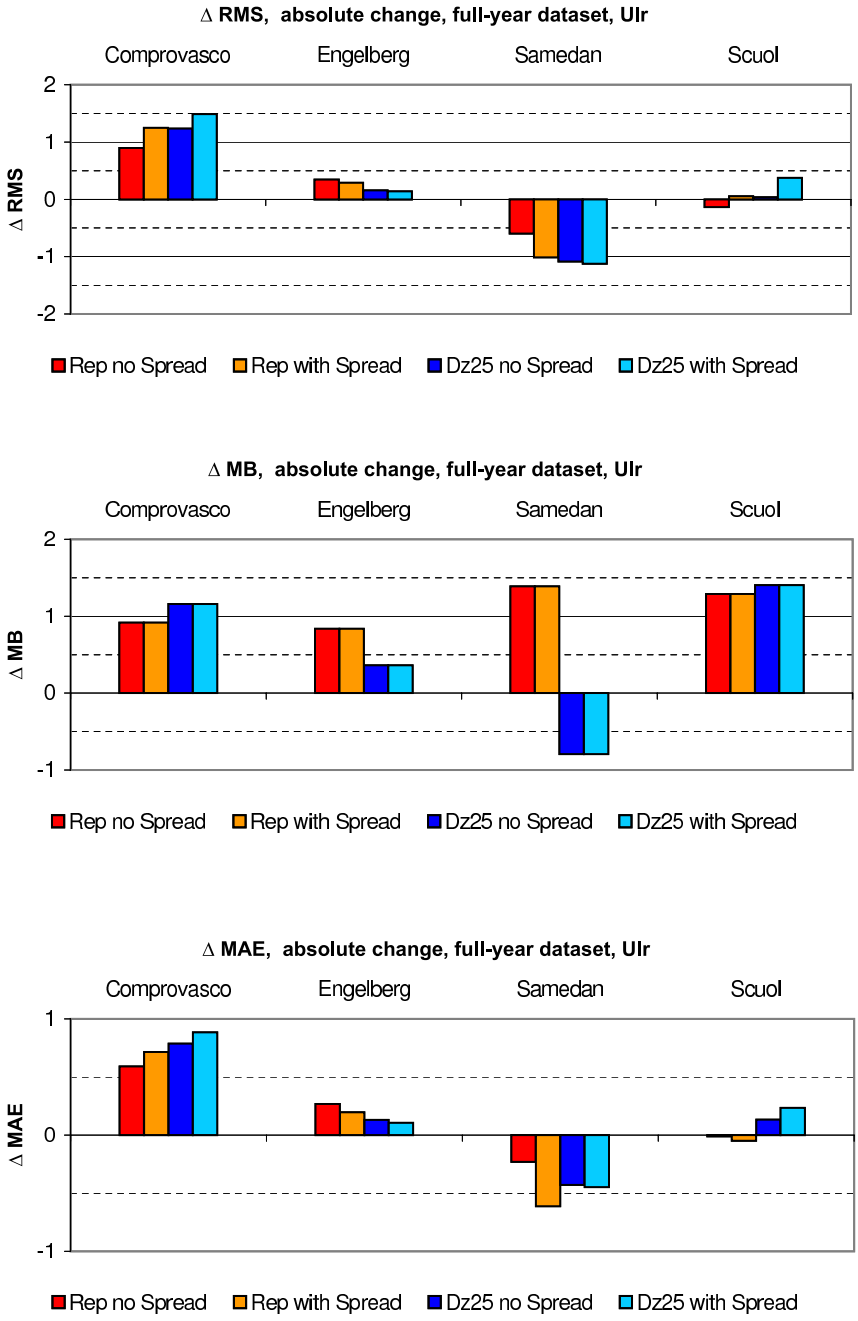


Figure 5.40: The changes in the absolute values of all three statistical quantities (RMS, MB, and MAE), before and after the applied modifications (cf. example in the remarks on Fig. 5.38). Negative values indicate an improvement through the modifications compared with the only altitude corrected data.
Note, the model data was corrected with the typical daily cycle of Ulrichen.
Ulr

For an overall estimation of the impact of the spreading procedure (quantitative) on each of the four stations of the control group we refer to the Tables B.1-B.4 on p. 133-136 (Appendix) with the statistical values of both approaches, i.e. with and without the spreading procedure, respectively.

Correlation Coefficients and the Rotation of the Data Clouds

In Table 5.8 the correlation coefficients of the four stations of the control group are shown. For the stations of Comprovasco, Engelberg, and Scuol the results are shown in case of using the the "Mean_all4" daily cycle in order to correct the model data. At Samedan the most appropriate daily cycle for the correction was the one at Ulrichen ("Ulr"). Therefore, the results are presented in case the "Ulr" diurnal cycle was chosen for the correction at this station.
We found higher correlation coefficients at Samedan and Scuol in case of the model data was modified through a mean diurnal cycle and where indicated by using the spreading procedure. At Engelberg the impacts of the correction were marginal (especially at the min_dz25" gp). At Comprovasco the spreading procedure was not successful at the "rep" gp.

The rotation of the data cloud towards the 1:1-line, described on p. 83 could also be seen at all four stations of the control group (cf. Table 5.9). At Engelberg, Samedan and Scuol, the spreading procedure supported this rotation towards the 1:1-line. Only at Comprovasco this transformation was too strong and the slope became steeper then 45°.

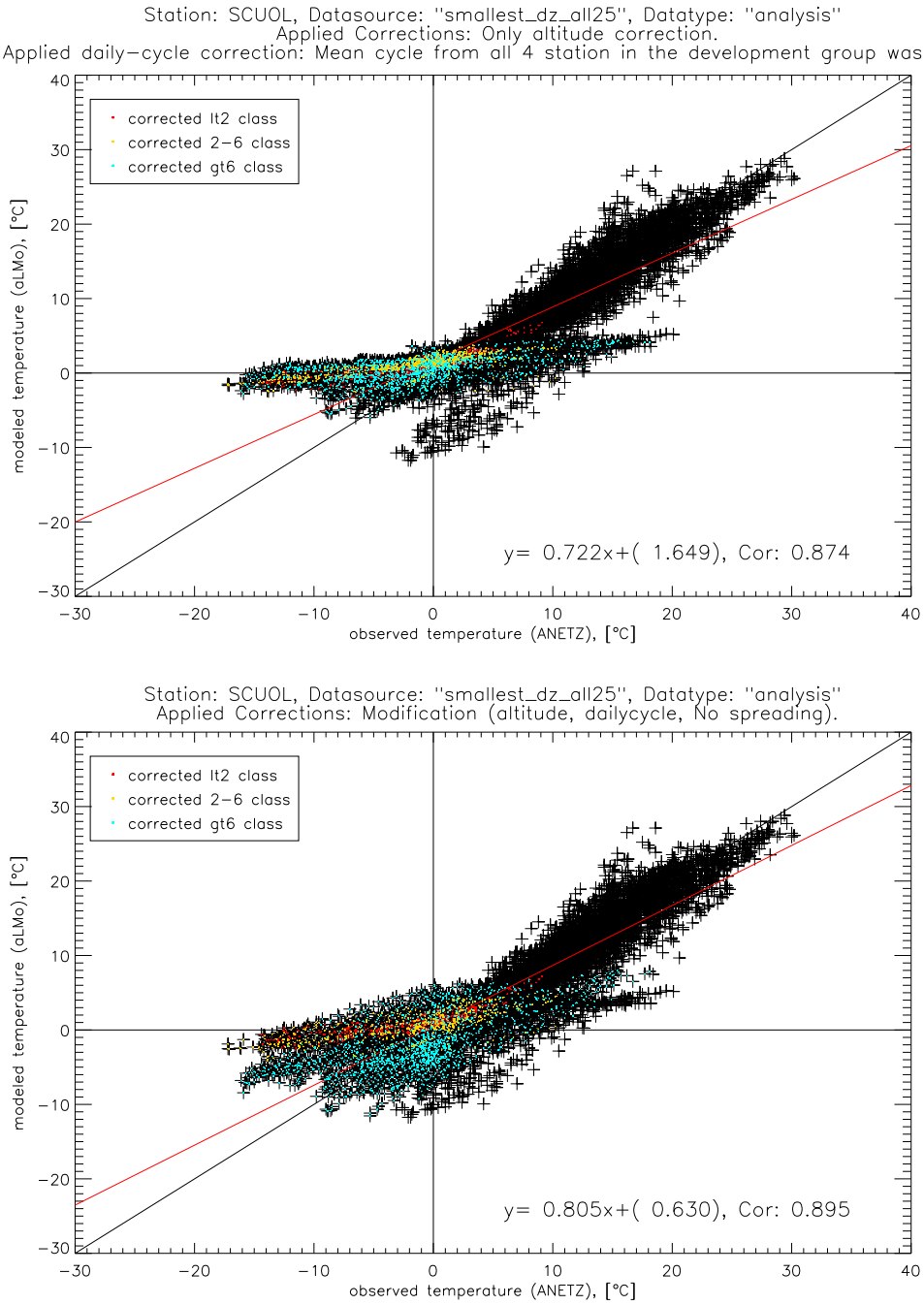


Figure 5.41: The full-year dataset (2m-temperatures) of Scuol.
The points in red, yellow, and blue indicate the data of the three different classes (cf. legend). While in the lower panel the data of the three classes were corrected (diurnal cycle and where indicated with the spreading procedure), the data points in the upper panel show where the original data (only altitude corrected) was situated.
Upper panel: only the altitude correction of the model data
Lower panel: model data of the three different classes were modified with a mean diurnal cycle of all four stations of the development group ("Mean all 4").
No Spreading

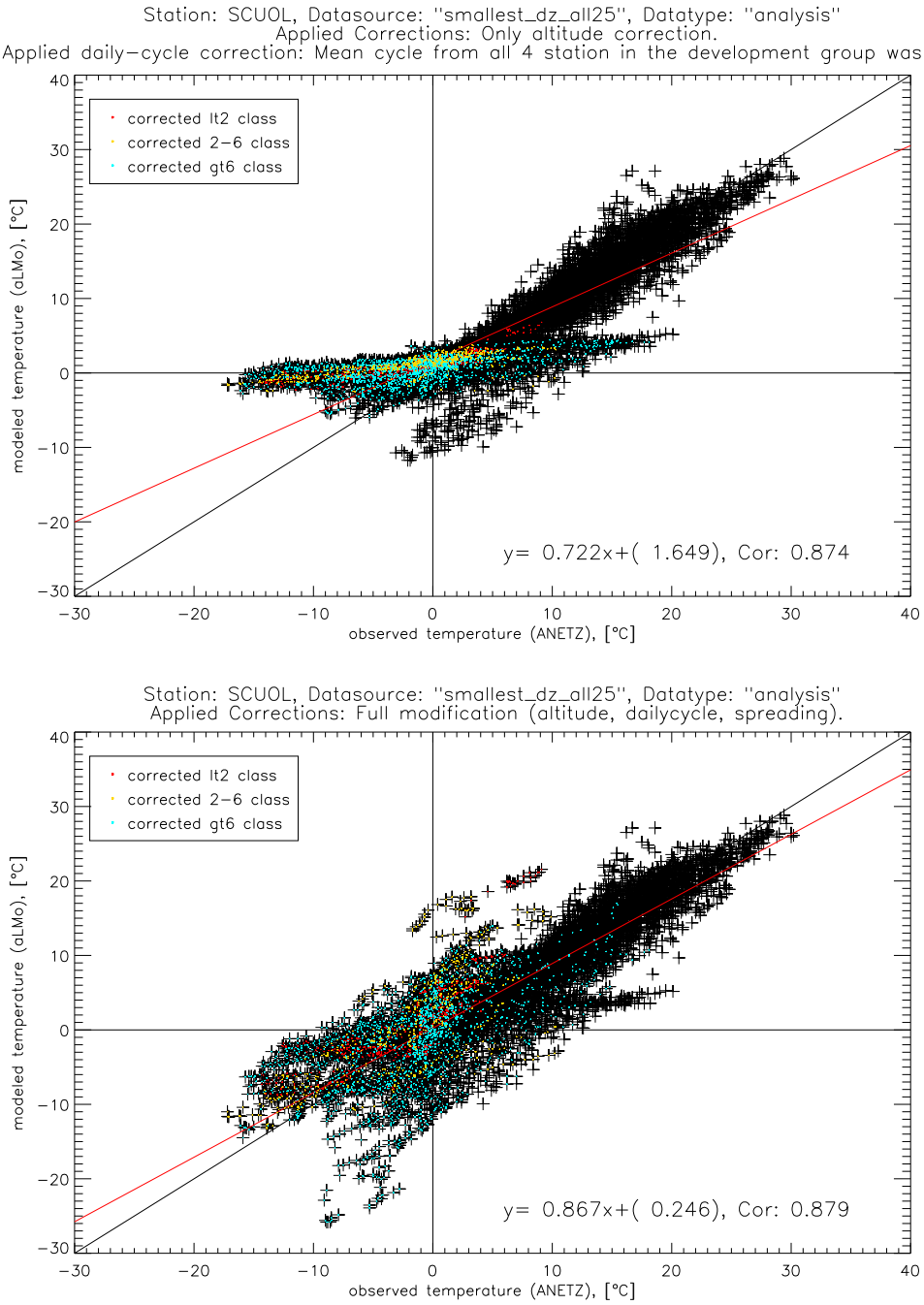


Figure 5.42: The full-year dataset (2m-temperatures) of Scuol.
The mean diurnal cycle and the spreading procedure were used for the correction of the model data (lower panel).
For further remarks we refer to Fig. 5.41.
With Spreading

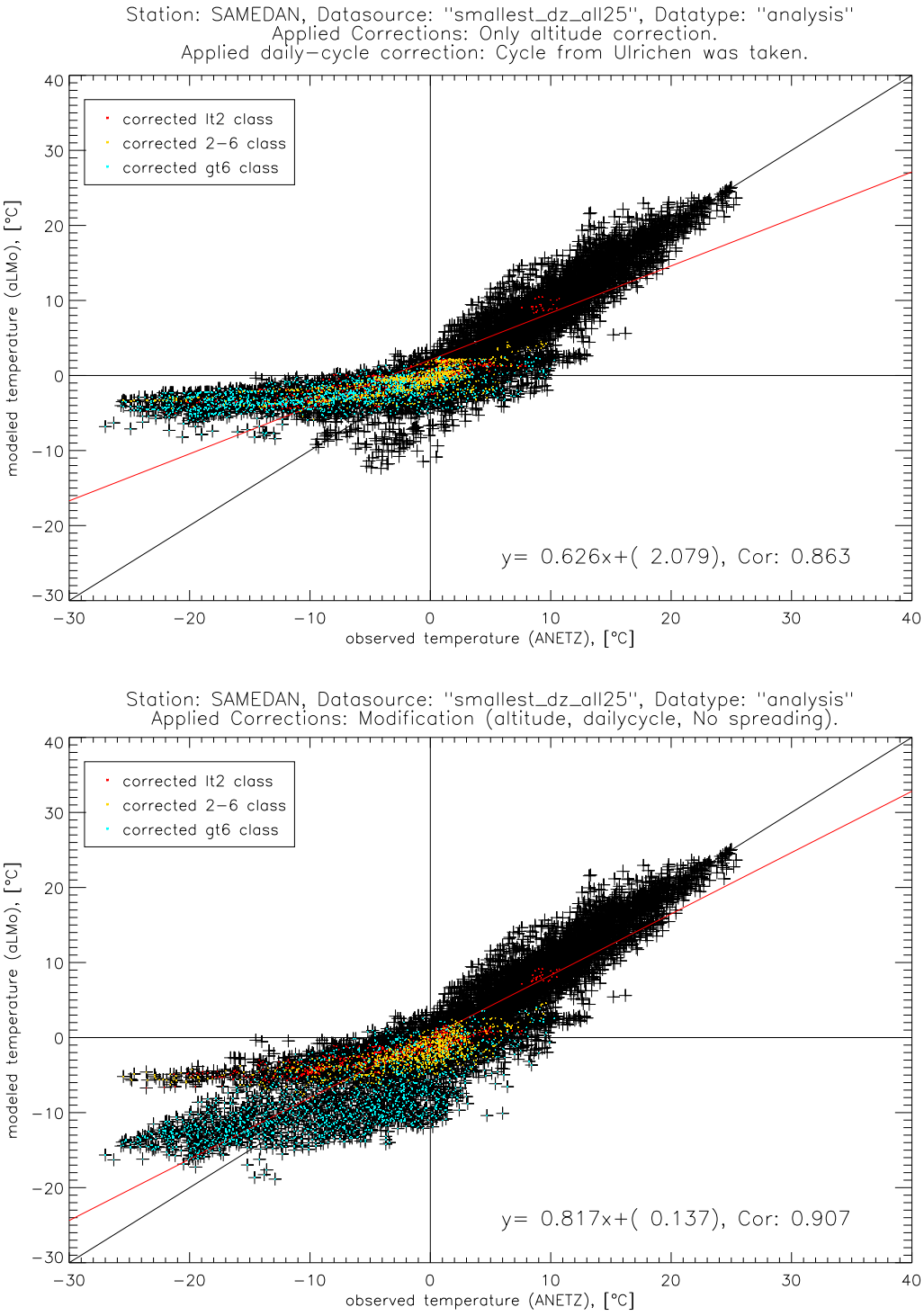


Figure 5.43: The full-year dataset (2m-temperatures) of Samedan.
The points in red, yellow, and blue indicate the data of the three different classes (cf. legend). While in the lower panel the data of the three classes were corrected (diurnal cycle and where indicated with the spreading procedure), the data points in the upper panel show where the original data (only altitude corrected) was situated.
Upper panel: only the altitude correction of the model data
Lower panel: model data of the three different classes were modified with the typical diurnal cycle of Ulrichen ("Ulr")
No Spreading

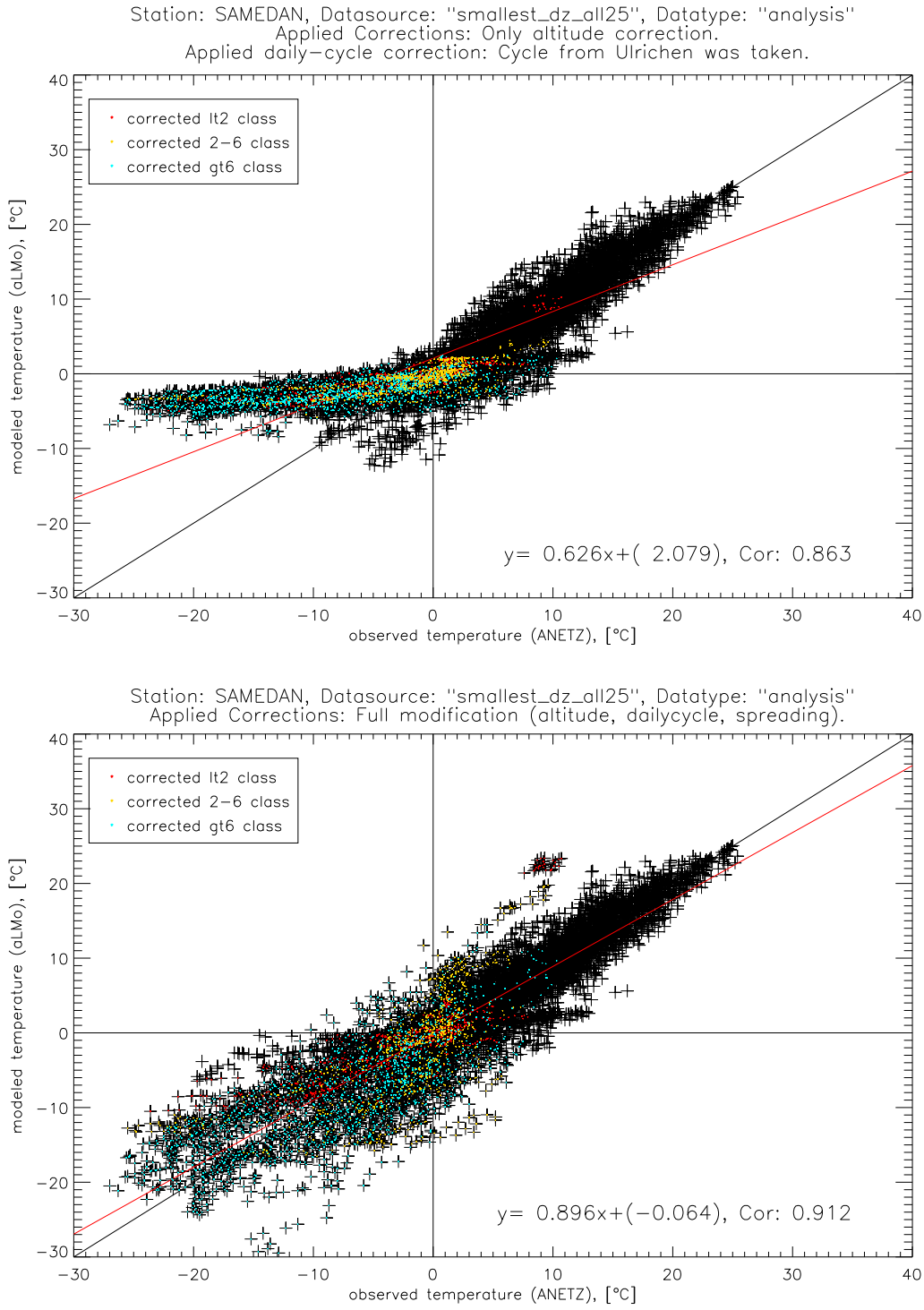


Figure 5.44: The full-year dataset (2m-temperatures) of Samedan.
The mean diurnal cycle and the spreading procedure were used for the correction of the model data (lower panel).
For further remarks we refer to Fig. 5.43.
With Spreading

Table 5.8: The correlation coefficients of the corrected modelled data (for each station of the control group) and the observed data are shown. Note that the full-year datasets were used for calculating the coefficients and not only these days that were actually corrected and therefore belonged to one of the three classes (lt_2, gt_6, and 2-6).
The modelled data of Comprovasco, Engelberg, and Scuol were corrected using the "Mean_all4" diurnal cycle, the data of Samedan by using the "Ulr" daily cycle.
In **bold** letters, where an improvement through the spreading resulted.
In *italic* letters, when the spreading led to even worse results than in case only the altitude correction was performed.
The abbreviation "only alt." refers to the results in case of the modelled data was only altitude corrected.

	Representative			Smallest_dz_all25		
Station	only alt.	no spread	with spread	only alt.	no spread	with spread
Comprovasco	0.919	0.925	*0.913*	0.919	0.925	0.920
Engelberg	0.915	0.915	**0.921**	0.944	0.943	**0.944**
Samedan	0.814	0.860	**0.886**	0.863	0.907	**0.912**
Scuol	0.827	0.852	0.841	0.874	0.895	0.879

Table 5.9: The slopes of the linear regressions (modelled data↔observed data) are shown. Note, that the full-year datasets were used for calculating the regression and not only these days that were actually corrected and therefore belonged to one of the three classes (lt_2, gt_6, and 2-6).
The modelled data of Comprovasco, Engelberg, and Scuol were corrected using the "Mean_all4" diurnal cycle, the data of Samedan by using the "Ulr" daily cycle.
In **bold** letters, when the "additional" rotation through the spreading led to slopes closer to "1".
In *italic* letters, when the "additional" rotation through the spreading led to slopes more distant to "1" than in case only the altitude correction was performed.
The abbreviation "only alt." refers to the results in case of the modelled data was only altitude corrected.

	Representative			Smallest_dz_all25		
Station	only alt.	no spread	with spread	only alt.	no spread	with spread
Comprovasco	0.939	0.987	*1.046*	0.939	*1.018*	*1.081*
Engelberg	0.805	0.834	**0.909**	0.851	0.870	**0.916**
Samedan	0.552	0.716	**0.817**	0.626	0.817	**0.896**
Scuol	0.637	0.700	**0.780**	0.722	0.805	**0.867**

5.10 Analysis of the Dataset 2004

Because of the programming mistake during the analysed time period 2002/03 (cf. Introduction, p. 3), the output fields of the near-surface temperatures, especially during the winter were not representative for other years. Therefore, we decided to analyse another year of data, namely from 1 January to 31 December 2004. Due to time constraints, only a "reduced" analysis could be performed.

The Table 5.10 on p. 105 presents the correlation coefficients of the modelled and observed 2m-temperatures. Note that the temperature data of the model was corrected regarding the difference in altitude between surface station and model grid point (climatological gradient).
We found at each station of the development group, no matter which grid point was chosen as data source a better correspondence between modelled and observed values in 2004 than during the time period 2002/03.

In Figures 5.45 and 5.46 on p. 106/107, the full-year temperature data at Ulrichen and Robbia of 2002/03 and 2004, respectively, are shown. At none of the four stations of the development group such striking horizontal structures at the "cold end", previously referred to as "tail"-region, could be found.
At Ulrichen, at both data sources, still a slightly positive curvature was seen in this part of the dataset, where the observed temperatures lay below 0°C (cf. Fig. 5.45 on p. 106).
Our conclusion of the missing tail-structures in winter 2002/03 was that the programming mistake was successfully removed after the winter 02/03.

Although not as much pronounced as during the time period 2002/03, in 2004 were still horizontally shaped data clouds ("nose"-regions) at the "warm end" visible. These data clusters were found at each station of the development group with differing distinctiveness.
We analysed the number of days found in the "Nose"-regions of all four stations and both data sources. The results are shown in Table 5.11 on p. 110. Note, that the threshold-temperatures, defining the "Nose"-regions, were manually determined.
The number of days, belonging to the "nose"-region tended to be smaller during the time period of 2004.

Table 5.10: A comparison of the correlation coefficients of modelled and observed 2m-temperatures during 2002/03 and 2004. Note, the modelled data was altitude-corrected.

	Representative		**Smallest_dz_all25**	
	2002/03	2004	2002/03	2004
Piotta	0.912	0.948	0.923	0.948
Robbia	0.858	0.916	0.912	0.920
Ulrichen	0.855	0.923	0.855	0.921
Zermatt	0.894	0.924	0.905	0.926

We were interested how the diurnal cycles of such "Nose"-days look like. Some examples are presented in the Figures 5.47 and 5.48 on p. 108/109. During these days, we found almost constant modelled temperatures between $\sim$ 9-16 UTC. Obviously, the warming process before noon, due to solar insolation was abruptly stopped as soon as a certain temperature was reached. Performing such a comparison between modelled and observed temperatures considering raw model data, i.e. without an altitude correction, we found that this stop of a further warming and therefore a suppressed continuous daily cycle occurred at temperatures around 0°C. At the same time, the snow water content started to decrease. This phenomena is shown in Fig. 5.49 on p. 110, where the situation can be seen at Ulrichen on a day in May 2004 at the "representative" gp.
To summarise, these structures occur mostly during a time period between April and May 2004. As soon as the temperature in the model became positive, the snow melting began and no further increase of the 2m-temperature was possible.

Despite of the correction of the programming bug, suppressing a diurnal cycle in the near-surface temperatures during periods where the soil was covered with snow in the winter 2002/03, still wrong estimations of the model leading to too weak daily cycles in the 2m-temperatures could be found in 2004. The reasons, seasons, and the temperature-"regions" where such a pattern was found in 2004 were different than during the winter 2002/03.

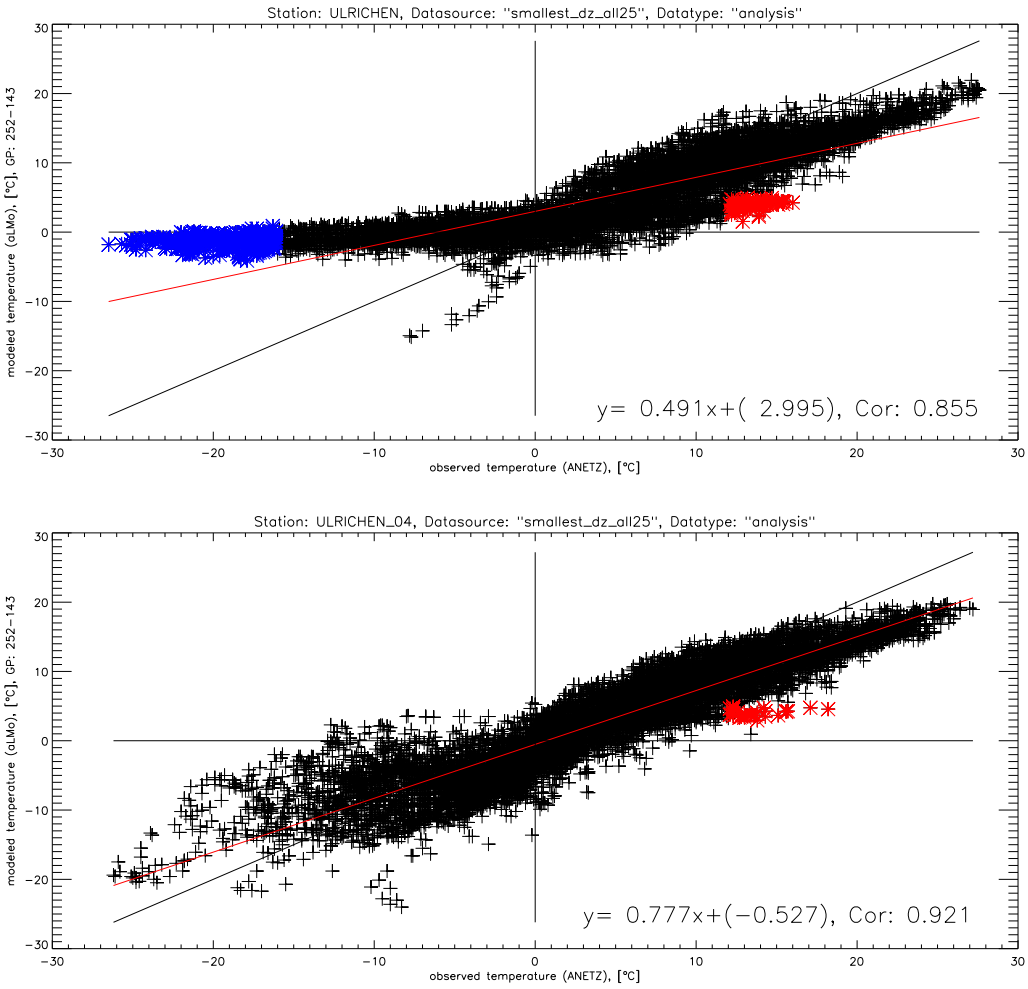


Figure 5.45: The two full-year datasets (2m-temperatures) of Ulrichen in 2002/03 (upper) and 2004 (lower panel), respectively.

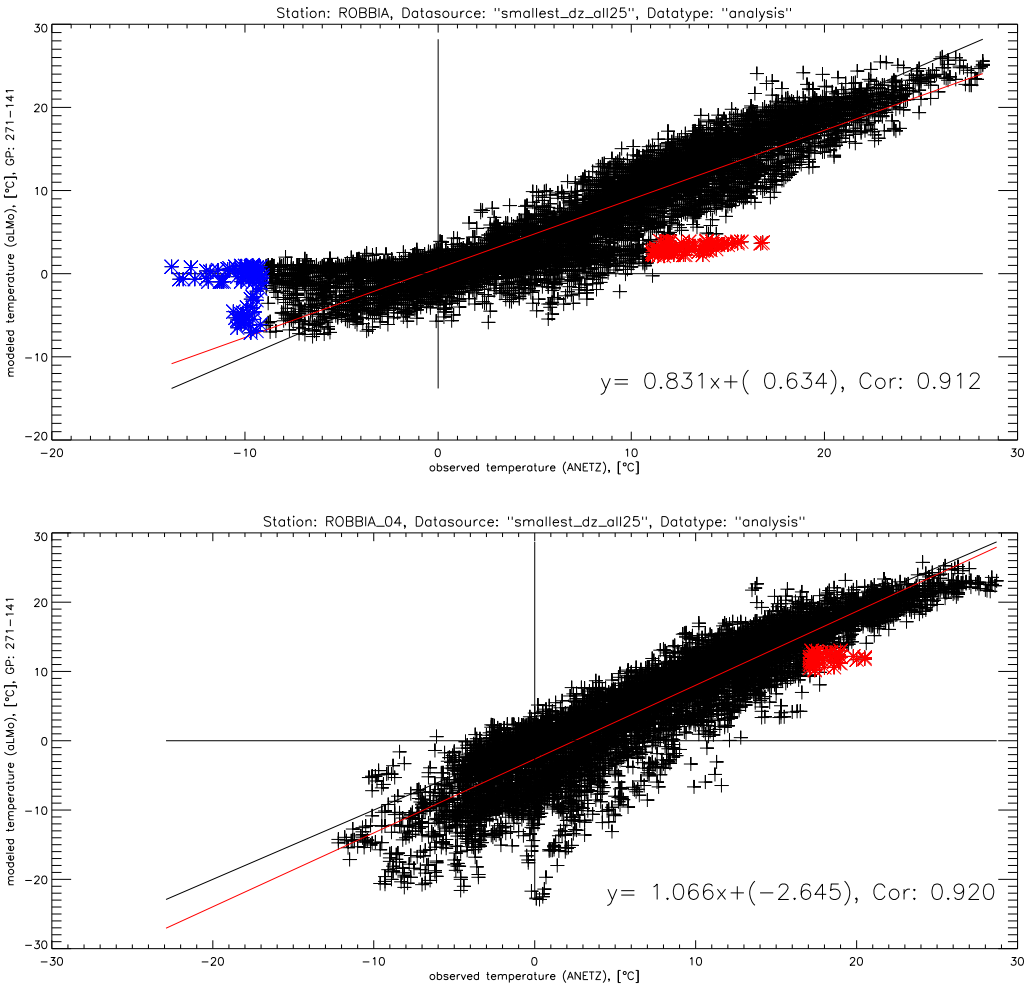


Figure 5.46: The two full-year datasets (2m-temperatures) of Robbia in 2002/03 (upper) and 2004 (lower panel), respectively.

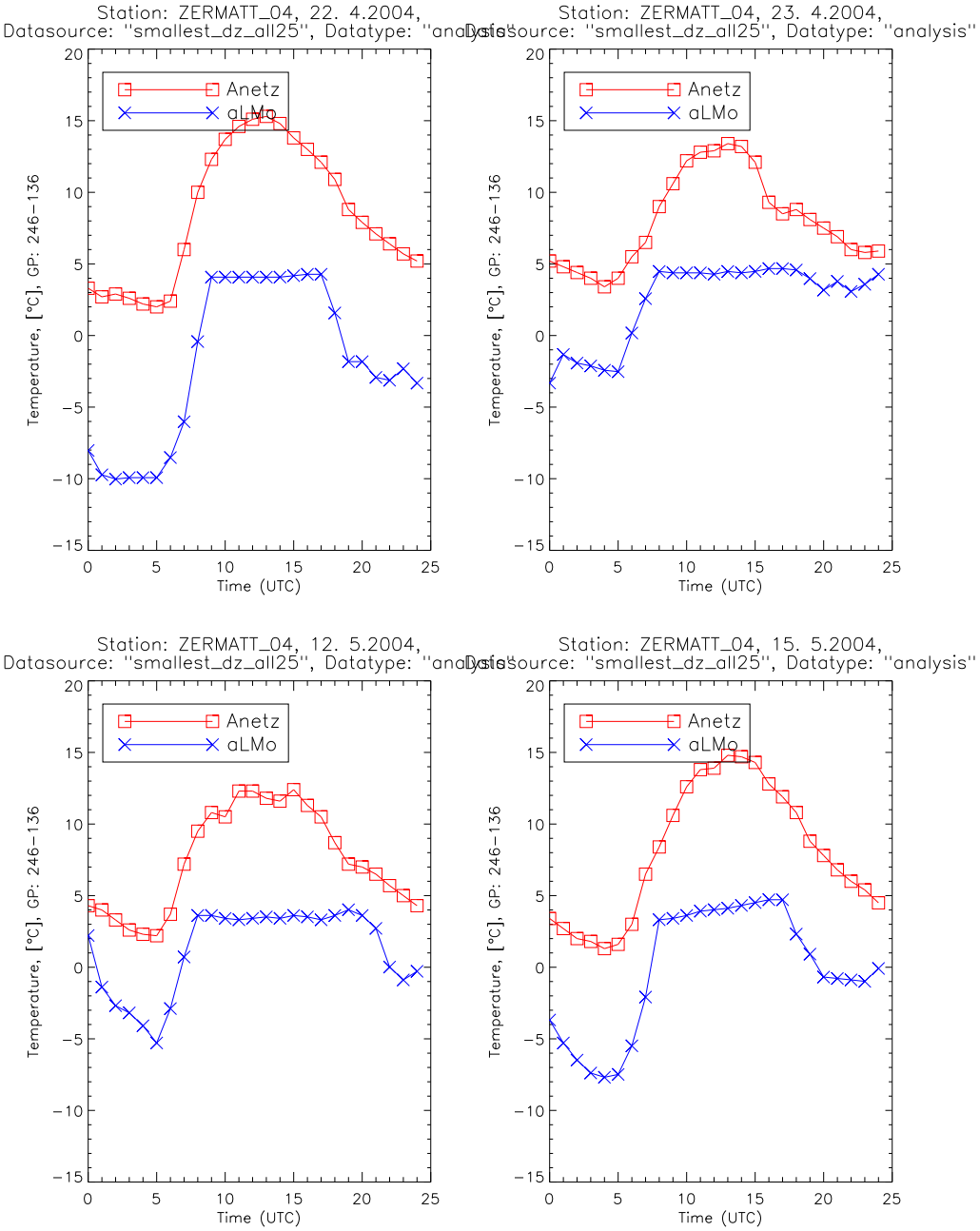


Figure 5.47: The diurnal cycles (2m-temperatures) of four days in the "Nose"-region at Zermatt in 2004.
Red: observed temperatures, blue: altitude corrected modelled values.

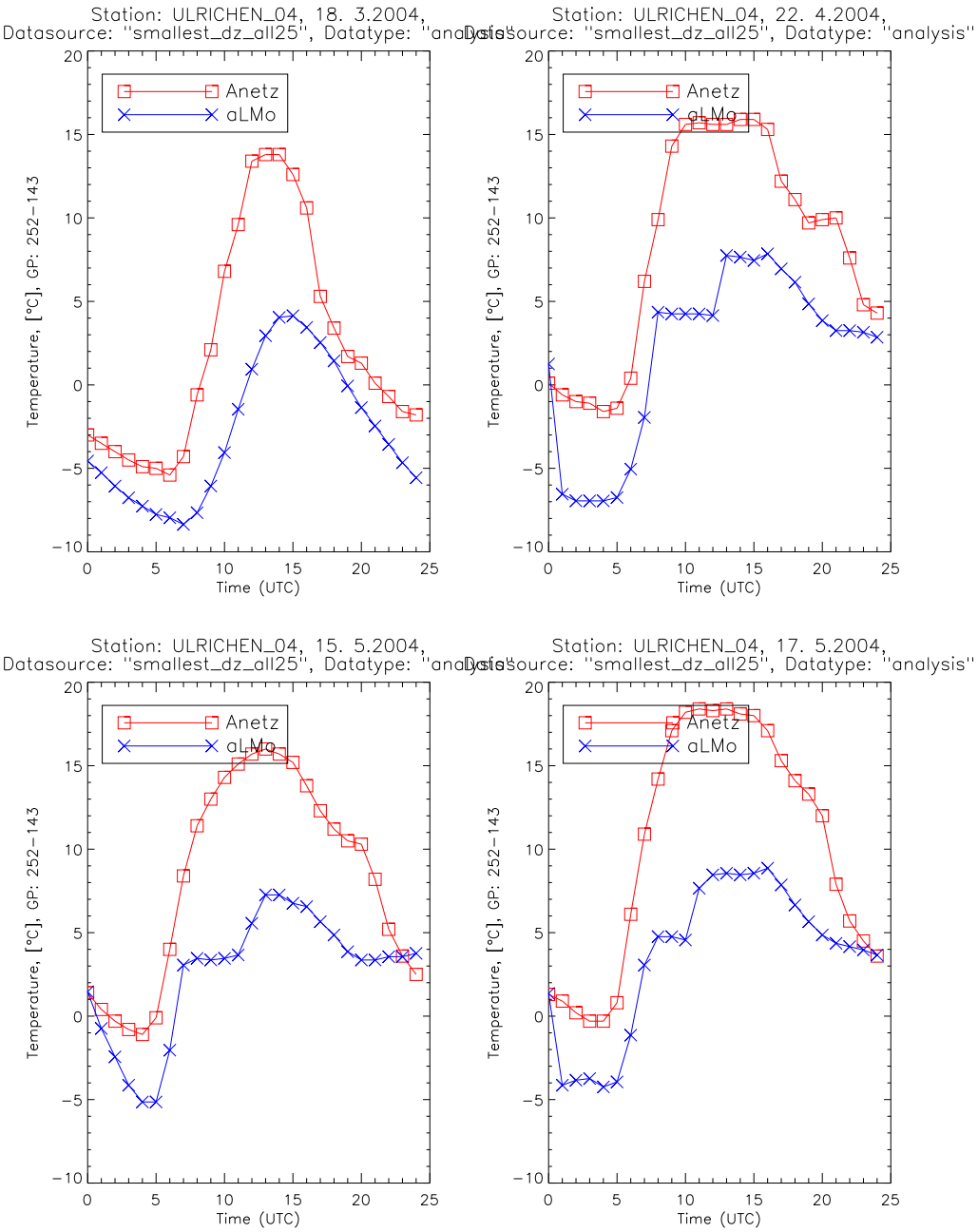


Figure 5.48: The diurnal cycles (2m-temperatures) of four days in the "Nose"-region at Ulrichen in 2004.
Red: observed temperatures, blue:altitude corrected modelled values.

Table 5.11: A comparison of the number of days found in the "Nose"-regions in the full-year datasets of 2002/03 and 2004. For more information about the identification of "Nose"-days we refer to p. 30 (Methods) and to p. 40 (Results).
Bold, where a smaller number was found in 2004 than during the time period 2002/03.

	Representative		**Smallest_dz_all25**	
	2002/03	2004	2002/03	2004
Piotta	17	18	32	**10**
Robbia	22	**13**	17	**16**
Ulrichen	13	**8**	18	**13**
Zermatt	13	**7**	24	24

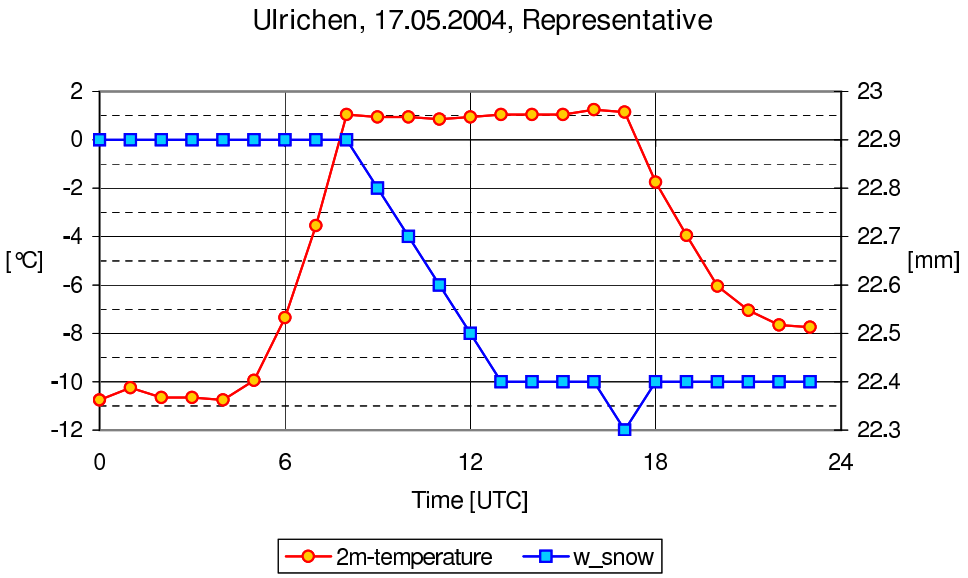


Figure 5.49: 2m-temperatures (red) and the snow water content, w_snow (blue) at Ulrichen (17 May 2004). Note that 1 mm snow water content is equivalent to a thickness of the snow layer of approximately 10 mm. This example shows the raw modelled temperature (without any altitude correction) at the "representative" gp.

Chapter 6

Conclusions and Outlook

Conclusions

In this study temperatures on the 2m-level generated by the operational NWP-model at MeteoSwiss were compared with observations taken by the automatic measurement net of Switzerland. Stations in particularly complex topography were chosen.
The major interest of this diploma thesis was to find a relation between the model-output at a certain grid point and the measurements taken at a neighbouring surface station. Therefore, the performance of the model during specific days, when large deviations between the modelled and observed temperatures occurred was analysed. An important part of this work was the development of a correction algorithm, modifying the modelled data on such critical days. In the beginning of this study we built two groups, each one containing four surface stations. While the data of the stations of the first group were used for developing the different algorithms, criteria, and threshold-values, these were finally tested on the stations of the second (control-) group (independent dataset).

In 2002/03 a bug in the model led to erroneous estimates of the 2m-temperature in case the surface was snow-covered.
Unfortunately, due to a communication error, the period for the data analysis was selected in such a way that this bug was still active. Therefore, the results as presented here are hence to be taken as a 'principle study' and the specific results are not representative for other years, where this programming error was not 'active'.

We found great deviations between the modelled and observed 2m-temperatures during wintertime 2002/03 which had their source in the previously mentioned model bug. These differences were mainly caused by a missing diurnal cycle in the modelled temperatures, while at the same time the surface stations recorded much stronger daily variations. These days in question most often occurred during a period of time between November until March. That corresponded well with the programming mistake in the model's operational source code in parameterising the surface temperatures (temperatures at the boundary between the surface and the atmosphere) whenever the surface was snow-covered (cf. Appendix, p. 137).
To avoid such erroneous estimations of the model, we established a set of criteria that was able to identify most critical days in advance only using the output variables of the model without considering the actual measurements at the surface station.
Adding an observed mean diurnal cycle to the model data on such days, we were able to improve in most cases the correspondence between modelled and observed data.
At Ulrichen, we found large errors in the statistical measures on days identified as most problematic ("gt_6"-class), i.e. a RMS $\sim$ 8.3 K, a MB $\sim$ -4.6 K, and an MAE of around 6.3 K. In the course of this analysis we were able to improve the RMS and MAE by more than 2,5 K, the MB even by 4 K.
At the other three stations of the development group (Piotta, Robbia, and Zermatt) the initially found errors were smaller, i.e. RMS: 3.6-4.8 K, MB: 0.2-1.2 K, and MAE : 2.8-3.7 K. These RMSs and MAEs were reduced by $\sim$ 0.5 K. Only small improvements were reached in the MBs.

We analysed the impact of choosing different grid points (gp) in the comparison with the observed data. For this purpose the differences between the modelled and observed daily minimum and maximum temperature, respectively, were considered. Most common for the verification procedures at MeteoSwiss is the so-called "representative" gp, which is usually the (horizontally) nearest one to the station. However, in case of a difference in altitude between gp and station exceeding 100 m, the gp of the nearest four with the least difference in altitude is believed to be the representative one.
This analysis revealed, that beside the commonly used "representative" gp, the one of the nearest 25 gps to a station showing the least difference in altitude ("min_dz25") led to comparable and often better results.
Considering the horizontal distance (gp $\rightarrow$ station), which was usually much larger in case of the "min_dz25" gp (up to 21 km, cf. Table 3.1 on p. 16), these results are even more interesting.

Further, in order to compare the modelled 2m-temperatures directly with observed values, taken at a surface station, the model data was altitude-corrected. We considered three different lapse rates: dry- and wet-adiabatic, and a variable climatological gradient (monthly mean values, cf. Data, p. 24).

At each station of the development group the correlation coefficients between modelled and observed values improved, in case of using the variable climatological instead of a constant gradient (dry, wet).

The improvements in case the variable climatological gradients instead of a constant lapse rate (dry-/wet-adiabatic) were used, depended on the data source (representative: "rep" gp, smallest_dz_all25: "min_dz25" gp).

At the "rep" gp, the correlation coefficients of the four stations of the development group slightly improved by $\sim$ 0.020, the ones at the "min_dz25" gp by $\sim$ 0.006 (cf. Table 5.1 on p. 40).

Additionally, a spreading procedure was introduced that took the much broader distribution of the observed diurnal cycles, compared with a narrow band of the modelled daily-cycles into account.

The impact of this approach was ambiguous and must be considered for each station and grid point separately.

Outlook

Based on the finding of this study, we suggest the following key aspects for future works:

- *Performing the same Analysis for another Year of Data:* Due to the previously explained programming error (cf. Introduction, p. 3 and Appendix, p. 137), the analysed winter 2002/03 can not be considered as generally representative. Therefore, a similar study to the one at hand should be performed for another full year of data, when no programming bug influences the mean temperatures and diurnal cycles during the snow season.
  Such an analysis will not only reveal that the error in the LM source code was effectively corrected, but also enables to concentrate on other interesting patterns (e.g. topographic shading), that could be found at alpine locations.
  One of these patterns, that should be analysed, is the particularly too strong nocturnal cooling in the model especially pronounced between December and February[1].

- *Diurnal Cycles used for the Corrections:* In this study, different mean diurnal cycles were used for the correction of the modelled data. The use of these variously composed daily cycles led to different results of our analysis. In case that such great deviations and missing diurnal cycles can also be found during other winters (without the programming error), we believe, that this technique should be carried on and be applied on a much larger set of alpine surface stations. There are other criteria, as for instance the elevation of the surface stations or geographic regions that can be considered, in order to gain other types of diurnal cycles.

- *Spreading Procedure:* In the course of this study we found, that all as problematic identified (gt_6-) days showed modelled diurnal cycles, that were not only almost constant, but also situated inside a very narrow bandwidth. The observed data, on the other hand, presented a much broader distribution of the recorded daily cycles. These results led us to the application of the spreading procedure.
  We found moderate to very high correlations between the 0 UTC temperatures of the model and the ones of the observation.
  In case that this pattern can also be found during other winters (without the programming error), we believe that a further development of this approach could lead to improved results. Instead of only using the temperature deviations at 0 UTC, more or other variables could be searched in order to find an appropriate correction for this difference in distribution.

[1] cf. the regular verification of "aLMo" done by F. Schubiger, Intranet, MeteoSwiss

- *Local Shading in Alpine Valleys.* It is a fact, that the orography of even high-resolution NWP-models (e.g. "aLMo") are still not able to represent a complex topography accurately enough to describe small scale phenomena, found in mountainous environments. Therefore, local characteristics as radiation balances or the times of the local sunrise and sunset are still inaccurately displayed by the model, which influence at the same time strongly the temperatures found inside a valley. Due to this simplified and too inaccurate representation of the real topography by in the model, time lags between the local sunrise/sunset in the model's orography and the actual experienced ones in the real topography occur. To account for these differences, we think it would be appropriate to find a connection between the time lags in the local sunrise and sunsets, respectively, and the temperatures. We propose a semi-statistical approach, including a regression combined with the use of a high-resolution digital elevation model.

  Even though the spatial resolution of NWP-models will become finer in the near future, statistical approaches in enhancing the quality of the model output show already today promising results in case-studies, where mesoscale models were operated on a very fine spatial resolution (cf. Hart et al. (2004), MM5, 4 km).

  Efforts in improving the ways, how the model treats radiation budgets in a complex terrain were made in Mueller and Scherer (2004) and Colette et al. (2003) (in Mueller and Scherer (2004)). But still, even if local variables (e.g. radiation) are treated more detailed, means over the full area of a grid cell (in aLMo: 49 km^2) have to be built for the implementation in the model.

- *Vertical Gradient in Temperature:* In the course of any comparison between modelled and observed values, a modification of one of the datasets has to be performed in order to account for the different location of the data in space and/or time. The concept of this modification depends on the variable of question. In case of the 2m-temperature, a correction of the modelled data, considering the vertical difference in altitude between the model gp and the surface station was performed. An important aspect of this approach is to choose a close to real vertical temperature gradient. Beside the three different lapse rates we considered in this study, (dry-, wet-adiabatic, and climatological) one can apply as well the data of a sounding (twice a day) or even calculate an appropriate lapse rate from the model data at the specific location.

According to the good correspondence between the modelled and observed temperatures (daily minimal and maximal temperatures) in case the grid point was used with the least difference in altitude of all a station surrounding 25 gps, we believe an effort in determine more accurate lapse rates would improve the results.

Chapter 7

Acknowledgements

I am most grateful to the efforts and support of Mathias Rotach. He gave me the chance to write this extraordinary interesting diploma thesis at the MeteoSchweiz. I am much obliged for the many hours he invested to discuss our strategies and approaches to the various tasks. I never felt dismissed but rather encouraged in my own ideas and suggestions.

I also appreciated to work next to Matteo Buzzi, the PhD student who currently is on his way to bring aLMo on a higher level in terms of spatial resolution. He gave me many recommendations and tips how to handle Unix and IDL, which was a great help. It was always a pleasure. Thanks.

Furthermore, I am grateful for the numerous discussions with Francis Schubiger, who always found some time for explanations in model-verification and general modelling belongings.

Many thanks go to all other members of the MO-Group of the MeteoSwiss and especially to Philippe Steiner for the nice talks in the tramway, to Jean Marie Bettems and Guy de Morsier for the discussions about data assimilation and soil processes and to Marco Arpagaus for the explanations in modelling and math support. Moreover, I would like to thank Pirmin Kaufmann and Emanuele Zala for the support in IDL problems and many other things.

I also appreciated very much the valuable work by Simon Scherrer and Heike Kunz in providing the necessary data to perform the altitude correction with climatologic gradients. Similarly, many thanks go to the staff from the "data warehouse" providing the data of the surface stations, used in this work.

In addition, I would like to thank Andreas Fischer who proofread my diploma thesis and supported me in several language concerns.

Many thanks go as well to the whole staff at MeteoSchweiz, who made this work possible. Namely the IT for the continuously enlargement of the quota and the forecasters on duty for the numerous interesting discussions should be mentioned here especially.

Vielen Dank auch an meine Eltern, welche mich waehrend des ganzen Studiums sehr unterstuetzt haben und immer hinter mir gestanden sind. Ohne euch waere diese Arbeit in dieser Form nicht moeglich gewesen.
Merci vielmal!

Chapter 8

Nomenclature

aLMo	Alpines Model
ANETZ	Automatisches Messnetz der Schweiz
ARPS	Advanced Regional Prediction System
asl	above sea level
CAPE	Convective Available Potential Energy
COSMO	Consortium for Small-scale Modelling
ECMWF	European Center for Medium-Range Weather Forecasts
GME	Global-Modell (DWD)
gp	grid point
gt_6	greater than 6 (class definition)
hPa	pressure unit in hundreds of Pascal (N/m^2)
IFS	Integrated Forecast System
K	temperature unit: "Kelvin"
LM	Lokal-Modell
min_dz25	gp of the surrounding 25 of a certain station that shows the smallest difference in altitude between model gp and the surface station
MOS	Model Output Statistic
NCCR	National Centres of Competence in Research

NN	Neural Network
NWP	Numerical Weather Prediction
rep	representative gp
SST	Sea Surface Temperature
SVF	Sky View afctor
WMO	World Meteorological Organisation

Bibliography

Bundesamt fuer Landestopographie, 2001: Formeln und Konstanten fuer die Berechnung der Schweizerischen schiefachsigen Zylinderprojektion und der Transformation zwischen Koordinatensystemen. Eidgenoessische Vermessungsdirektion, Tech. rep.

Colette, A., F. K. Chow, and R. L. Street, 2003: A numerical study of inversion-layer breakup and the effects of topographic shading in idealized valleys. *Journal of Applied Meteorology*, **42**, 1255–1272.

De Morsier, G., 2001: Das Lokal-Modell (LM) der Wetterdienste von Deutschland, Griechenland, Italien und der Schweiz. *Oesterreichische Beitraege zu Meteorologie und Geophysik, ISSN Heft Nr. 27*.

De Morsier, G., 2004: *Newsletter No.4*, chap. 4 Operational Applications, MeteoSwiss (Zuerich), pp. 30–34, Consortium for Small-Scale Modelling.

Deng, X., and R. Stull, 2004: A mesoscale analysis method for surface potential temperature in mountainous and coastal terrain, revised manuscript to Monthly Weather Review.

Doms, G., and U. Schaettler, Eds., 2002: *Newsletter No.2*, chap. 3 Model System Overview, pp. 9–23, Consortium for Small-Scale Modelling.

Doms, G., and U. Schaettler, Eds., 2004: *Newsletter No.4*, chap. 3 Model System Overview, pp. 9–23, Consortium for Small-Scale Modelling.

Flemming, J., et al., 2004: E.C.M.W.F's data and products for GMES, HALO report.

Hart, K. H., W. J. Steenburgh, D. J. Onton, and A. J. Siffert, 2004: An evaluation of mesoscale-model-based model output statistics (MOS) during the 2002 Olympic and Paralympic Winter Games. *Weather and Forecasting*, **19**, 200–218.

Kalnay, E., 2003: *Atmospheric Modeling, Data Assimilation and Predictability*, chap. Appendix C, pp. 276–279, Press Syndicate of the University of Cambridge.

Kendall, M., and W. Buckland, 1971: *A Dictionary of Statistical Terms*. 3rd ed., Longman Group Limited.

Kretzschmar, R., P. Eckert, and D. Cattani, 2004: Neural network classifiers for local wind prediction. *Journal of Applied Meteorology*, **43**, 727–738.

Liljequist, G., and K. Cehak, 1984: *Allgemeine Meteorologie*. 3. Auflage, Vieweg.

Linder, W., 2005: High resolution temperature interpolation for the Swiss Alps. *submitted to Meteorologische Zeitschrift*.

Marzban, C., 2003: Neural networks for post-processing model output: ARPS. *Monthly Weather Review*, **131**(6), 1103–1111.

Matzinger, N., M. Andretta, E. van Gorsel, R. Vogt, A. Ohmura, and M. W. Rotach, 2003: Surface radiation budget in an alpine valley. *Quarterly Journal of the Royal Meteorological Society*, **129**(588), 877–895.

Mueller, M. D., and D. Scherer, 2004: A grid and subgrid scale radiation parameterization of topographic effects for mesoscale weather forecast models, Institute of Meteorology Climatology and Remote Sensing, University of Basel.

Whiteman, C. D., 1990: Observations of thermally developed wind systems in mountainous terrain. *Atmospheric Processes over Complex Terrain, Meteorological Monographs*, **23**(45), 5–42. American Meteorological Society.

Zala, E., 2002: *Newsletter No.2*, chap. 4 Operational Applications, MeteoSwiss (Zuerich), pp. 29–33, Consortium for Small-Scale Modelling.

Appendix A

Analysis of the Criteria

In this appendix, the statistical measures Root Mean Square (RMS), Mean Bias (MB), and Mean Absolute Error (MAE) of observed and modelled 2m-temperatures at the selected stations of the development group are shown. The modelled data was corrected, using the procedures outlines on p. 70/76 (Results) and p. 32 (Methods), i.e. corrections through adding a diurnal cycle and a spreading procedure.

The data, considered for this analyses consisted of all days of the specific class (lt_2, 2-6, and gt_6), i.e. only data, that was actually corrected was taken into account to compute the statistical measures, seen in the Tables A.1-A.8.

The first column in these Tables specifies the statistical quantity (RMS, MB, or MAE). In the second column (grey), the results in case of only the altitude correction was carried out are shown while in the third one the data was additionally modified with a typical diurnal cycle of the specific station and, where indicated with the spreading procedure.

The green (red) columns at the end indicate the improvements (worsenings) through the additional corrections (daily cycle, spreading) compared with the only altitude corrected data. In the first column of these two the difference between the absolute errors are shown, the second one refers to the relative change between these two approaches. A negative algebraic sign indicates an improvement through the corrections, a positive one a worsening.

While in the Tables A.1 and A.2 the results without the application of the spreading procedure at the "rep" gp are shown, the Tables A.3 and A.4 present the results with the spreading procedure at the data source. The Tables A.5 and A.6 present the results without the application of the spreading procedure at the "min_dz25" gp. The Tables A.7 and A.8 on the other hand show the results at the same data source with the spreading procedure.

Table A.1: Statistics at the "representative" gp at Piotta and Robbia. The individual daily-cycles were used for the correction, without spreading.
Piotta and Robbia, "rep" gp, no spreading

Piotta	Class: ' lower_than_2 '			n = 552
	Only altitude correction	Full modification	Improvement absolute	Improvement relative
RMS	2.614	2.401	-0.212	-8.1%
MB	1.250	0.829	-0.420	-33.6%
MAE	2.073	1.874	-0.199	-9.6%

	Class: ' 2.0-6.0 '			n = 696
	Only altitude correction	Full modification	Improvement absolute	Improvement relative
RMS	2.910	2.916	0.005	0.2%
MB	0.326	-0.741	0.415	127.4%
MAE	2.273	2.147	-0.126	-5.5%

	Class: ' greater_than_6 '			n = 1752
	Only altitude correction	Full modification	Improvement absolute	Improvement relative
RMS	4.051	3.580	-0.471	-11.6%
MB	1.223	0.484	-0.739	-60.4%
MAE	3.072	2.819	-0.254	-8.3%

Robbia	Class: ' lower_than_2 '			n = 528
	Only altitude correction	Full modification	Improvement absolute	Improvement relative
RMS	2.839	2.888	0.049	1.7%
MB	0.686	-0.907	0.221	32.2%
MAE	2.356	2.139	-0.217	-9.2%

	Class: ' 2.0-6.0 '			n = 1224
	Only altitude correction	Full modification	Improvement absolute	Improvement relative
RMS	5.060	3.818	-1.241	-24.5%
MB	3.191	1.498	-1.693	-53.1%
MAE	4.011	3.067	-0.944	-23.5%

	Class: ' greater_than_6 '			n = 1776
	Only altitude correction	Full modification	Improvement absolute	Improvement relative
RMS	4.564	3.652	-0.912	-20.0%
MB	0.459	-0.457	-0.002	-0.4%
MAE	3.623	2.888	-0.735	-20.3%

Table A.2: Statistics at the "representative" gp at Ulrichen and Zermatt. The individual daily-cycles were used for the correction, without spreading.
Ulrichen and Zermatt, "rep" gp, no spreading

Ulrichen	Class: ' lower_than_2 '			n = 552
	Only altitude correction	Full modification	Improvement absolute	Improvement relative
RMS	2.527	2.332	-0.195	-7.7%
MB	-0.725	0.165	-0.560	-77.2%
MAE	1.885	1.765	-0.120	-6.4%

	Class: ' 2.0-6.0 '			n = 984
	Only altitude correction	Full modification	Improvement absolute	Improvement relative
RMS	6.025	5.682	-0.343	-5.7%
MB	-1.017	0.034	-0.983	-96.6%
MAE	3.953	4.044	0.092	2.3%

	Class: ' greater_than_6 '			n = 2256
	Only altitude correction	Full modification	Improvement absolute	Improvement relative
RMS	8.219	5.794	-2.426	-29.5%
MB	-4.684	0.343	-4.341	-92.7%
MAE	6.216	4.862	-1.354	-21.8%

Zermatt	Class: ' lower_than_2 '			n = 504
	Only altitude correction	Full modification	Improvement absolute	Improvement relative
RMS	2.490	2.504	0.014	0.6%
MB	-0.033	-0.541	0.509	1560.4%
MAE	1.897	1.835	-0.062	-3.3%

	Class: ' 2.0-6.0 '			n = 792
	Only altitude correction	Full modification	Improvement absolute	Improvement relative
RMS	3.235	3.024	-0.211	-6.5%
MB	-0.138	0.121	-0.017	-12.2%
MAE	2.374	2.346	-0.028	-1.2%

	Class: ' greater_than_6 '			n = 2376
	Only altitude correction	Full modification	Improvement absolute	Improvement relative
RMS	4.265	3.500	-0.765	-17.9%
MB	0.162	-0.103	-0.060	-36.8%
MAE	3.316	2.775	-0.541	-16.3%

Table A.3: Statistics at the "representative" gp at Piotta and Robbia. The individual daily-cycles and the spreading procedure were used for the correction.
Piotta and Robbia, "rep" gp, with spreading

Piotta	Class: ' lower_than_2 '			n = 552
	Only altitude correction	Full modification	Improvement	
			absolute	relative
RMS	2.614	2.504	-0.110	-4.2%
MB	1.250	0.829	-0.420	-33.6%
MAE	2.073	1.976	-0.097	-4.7%

	Class: ' 2.0-6.0 '			n = 696
	Only altitude correction	Full modification	Improvement	
			absolute	relative
RMS	2.910	2.946	0.036	1.2%
MB	0.326	-0.741	0.415	127.4%
MAE	2.273	2.341	0.069	3.0%

	Class: ' greater_than_6 '			n = 1752
	Only altitude correction	Full modification	Improvement	
			absolute	relative
RMS	4.051	3.502	-0.549	-13.5%
MB	1.223	0.484	-0.739	-60.4%
MAE	3.072	2.868	-0.204	-6.6%

Robbia	Class: ' lower_than_2 '			n = 528
	Only altitude correction	Full modification	Improvement	
			absolute	relative
RMS	2.839	2.204	-0.635	-22.4%
MB	0.686	-0.907	0.221	32.3%
MAE	2.356	1.709	-0.647	-27.5%

	Class: ' 2.0-6.0 '			n = 1224
	Only altitude correction	Full modification	Improvement	
			absolute	relative
RMS	5.060	3.560	-1.499	-29.6%
MB	3.191	1.498	-1.693	-53.1%
MAE	4.011	2.812	-1.199	-29.9%

	Class: ' greater_than_6 '			n = 1776
	Only altitude correction	Full modification	Improvement	
			absolute	relative
RMS	4.564	3.693	-0.870	-19.1%
MB	0.459	-0.457	-0.002	-0.3%
MAE	3.623	3.013	-0.610	-16.8%

Table A.4: Statistics at the "representative" gp at Ulrichen and Zermatt. The individual daily-cycle and the spreading procedure were used for the correction.
Ulrichen and Zermatt, "rep" gp, with spreading

Ulrichen	Class: ' lower_than_2 '			n = 552
	Only altitude correction	Full modification	Improvement	
			absolute	relative
RMS	2.527	2.283	-0.244	-9.7%
MB	-0.725	-0.165	-0.560	-77.2%
MAE	1.885	1.792	-0.092	-4.9%

	Class: ' 2.0-6.0 '			n = 984
	Only altitude correction	Full modification	Improvement	
			absolute	relative
RMS	6.025	5.253	-0.773	-12.8%
MB	-1.017	0.034	-0.983	-96.6%
MAE	3.953	4.194	0.242	6.1%

	Class: ' greater_than_6 '			n = 2256
	Only altitude correction	Full modification	Improvement	
			absolute	relative
RMS	8.219	5.045	-3.174	-38.6%
MB	-4.684	0.343	-4.341	-92.7%
MAE	6.216	4.071	-2.145	-34.5%

Zermatt	Class: ' lower_than_2 '			n = 504
	Only altitude correction	Full modification	Improvement	
			absolute	relative
RMS	2.490	2.508	0.018	0.7%
MB	-0.033	-0.541	0.509	1565.5%
MAE	1.897	1.990	0.093	4.9%

	Class: ' 2.0-6.0 '			n = 792
	Only altitude correction	Full modification	Improvement	
			absolute	relative
RMS	3.235	2.606	-0.630	-19.5%
MB	-0.138	0.121	-0.017	-12.1%
MAE	2.374	2.005	-0.369	-15.5%

	Class: ' greater_than_6 '			n = 2376
	Only altitude correction	Full modification	Improvement	
			absolute	relative
RMS	4.265	3.057	-1.208	-28.3%
MB	0.162	-0.103	-0.060	-36.8%
MAE	3.316	2.378	-0.938	-28.3%

Table A.5: Statistics at the "min_dz25" gp at Piotta and Robbia. The individual daily-cycles were used for the correction, without spreading.
Piotta and Robbia, "min_dz25" gp, no spreading

Piotta	Class: ' lower_than_2 '			n = 600
	Only altitude correction	Full modification	Improvement absolute	Improvement relative
RMS	2.969	2.937	-0.032	-1.1%
MB	-0.017	0.191	0.173	1008.8%
MAE	2.364	2.335	-0.029	-1.2%

	Class: ' 2.0-6.0 '			n = 624
	Only altitude correction	Full modification	Improvement absolute	Improvement relative
RMS	3.525	3.478	-0.047	-1.3%
MB	0.132	-0.360	0.228	173.2%
MAE	2.685	2.645	-0.041	-1.5%

	Class: ' greater_than_6 '			n = 1512
	Only altitude correction	Full modification	Improvement absolute	Improvement relative
RMS	4.184	3.693	-0.492	-11.8%
MB	1.069	0.372	-0.697	-65.2%
MAE	3.159	2.926	-0.233	-7.4%

Robbia	Class: ' lower_than_2 '			n = 408
	Only altitude correction	Full modification	Improvement absolute	Improvement relative
RMS	3.594	3.950	0.356	9.9%
MB	-1.124	-1.991	0.867	77.2%
MAE	2.736	2.874	0.139	5.1%

	Class: ' 2.0-6.0 '			n = 456
	Only altitude correction	Full modification	Improvement absolute	Improvement relative
RMS	2.663	2.630	-0.032	-1.2%
MB	-0.054	0.761	0.707	1299.7%
MAE	2.094	2.169	0.075	3.6%

	Class: ' greater_than_6 '			n = 1080
	Only altitude correction	Full modification	Improvement absolute	Improvement relative
RMS	4.800	4.079	-0.721	-15.0%
MB	-0.362	0.155	-0.207	-57.2%
MAE	3.737	3.247	-0.490	-13.1%

Table A.6: Statistics at the "min_dz25" gp at Ulrichen and Zermatt. The individual daily-cycles were used for the correction, without spreading.
Ulrichen and Zermatt, "min_dz25" gp, no spreading

Ulrichen	Class: ' lower_than_2 '			n = 696
	Only altitude correction	Full modification	Improvement absolute	Improvement relative
RMS	5.825	5.381	-0.444	-7.6%
MB	-2.524	-1.446	-1.078	-42.7%
MAE	3.730	3.512	-0.218	-5.8%

	Class: ' 2.0-6.0 '			n = 1080
	Only altitude correction	Full modification	Improvement absolute	Improvement relative
RMS	6.909	6.184	-0.725	-10.5%
MB	-2.841	-1.290	-1.551	-54.6%
MAE	4.813	4.429	-0.384	-8.0%

	Class: ' greater_than_6 '			n = 1992
	Only altitude correction	Full modification	Improvement absolute	Improvement relative
RMS	8.362	5.785	-2.577	-30.8%
MB	-4.868	0.929	-3.940	-80.9%
MAE	6.321	4.835	-1.486	-23.5%

Zermatt	Class: ' lower_than_2 '			n = 264
	Only altitude correction	Full modification	Improvement absolute	Improvement relative
RMS	3.149	2.824	-0.324	-10.3%
MB	1.188	0.333	-0.855	-72.0%
MAE	2.521	2.294	-0.227	-9.0%

	Class: ' 2.0-6.0 '			n = 456
	Only altitude correction	Full modification	Improvement absolute	Improvement relative
RMS	3.608	3.474	-0.134	-3.7%
MB	-0.545	-0.416	-0.130	-23.8%
MAE	2.748	2.705	-0.044	-1.6%

	Class: ' greater_than_6 '			n = 1848
	Only altitude correction	Full modification	Improvement absolute	Improvement relative
RMS	3.584	3.032	-0.553	-15.4%
MB	0.271	-0.452	0.181	66.9%
MAE	2.785	2.454	-0.331	-11.9%

Table A.7: Statistics at the "min_dz25" gp Piotta and Robbia. The individual daily-cycles and the spreading procedure were used for the correction.
Piotta and Robbia, "min_dz25" gp, with spreading

Piotta	Class: ' lower_than_2 '			n = 600
	Only altitude correction	Full modification	Improvement absolute	Improvement relative
RMS	2.969	2.748	-0.221	-7.4%
MB	-0.017	-0.191	0.173	1014.0%
MAE	2.364	2.309	-0.055	-2.3%

	Class: ' 2.0-6.0 '			n = 624
	Only altitude correction	Full modification	Improvement absolute	Improvement relative
RMS	3.525	3.633	0.109	3.1%
MB	0.132	0.360	0.228	173.2%
MAE	2.685	2.969	0.284	10.6%

	Class: ' greater_than_6 '			n = 1512
	Only altitude correction	Full modification	Improvement absolute	Improvement relative
RMS	4.184	3.907	-0.278	-6.6%
MB	1.069	0.372	-0.697	-65.2%
MAE	3.159	3.025	-0.133	-4.2%

Robbia	Class: ' lower_than_2 '			n = 408
	Only altitude correction	Full modification	Improvement absolute	Improvement relative
RMS	3.594	3.741	0.147	4.1%
MB	-1.124	-1.991	0.867	77.2%
MAE	2.736	3.078	0.342	12.5%

	Class: ' 2.0-6.0 '			n = 456
	Only altitude correction	Full modification	Improvement absolute	Improvement relative
RMS	2.663	2.802	0.139	5.2%
MB	-0.054	0.761	0.707	1301.3%
MAE	2.094	2.290	0.196	9.3%

	Class: ' greater_than_6 '			n = 1080
	Only altitude correction	Full modification	Improvement absolute	Improvement relative
RMS	4.800	4.142	-0.657	-13.7%
MB	-0.362	0.155	-0.207	-57.2%
MAE	3.737	3.299	-0.438	-11.7%

Table A.8: Statistics at the "min_dz25" gp Ulrichen and Zermatt. The individual daily-cycles and the spreading procedure were used for the correction.
Ulrichen and Zermatt, "min_dz25" gp, with spreading

Ulrichen	Class: ' lower_than_2 '			n = 696
	Only altitude correction	Full modification	Improvement absolute	Improvement relative
RMS	5.825	4.818	-1.006	-17.3%
MB	-2.524	-1.446	-1.078	-42.7%
MAE	3.730	3.319	-0.411	-11.0%

	Class: ' 2.0-6.0 '			n = 1080
	Only altitude correction	Full modification	Improvement absolute	Improvement relative
RMS	6.909	6.062	-0.847	-12.3%
MB	-2.841	-1.290	-1.551	-54.6%
MAE	4.813	4.704	-0.109	-2.3%

	Class: ' greater_than_6 '			n = 1992
	Only altitude correction	Full modification	Improvement absolute	Improvement relative
RMS	8.362	5.345	-3.017	-36.1%
MB	-4.868	0.929	-3.940	-80.9%
MAE	6.321	4.368	-1.953	-30.9%

Zermatt	Class: ' lower_than_2 '			n = 264
	Only altitude correction	Full modification	Improvement absolute	Improvement relative
RMS	3.149	2.894	-0.254	-8.1%
MB	1.188	0.333	-0.855	-72.0%
MAE	2.521	2.247	-0.274	-10.9%

	Class: ' 2.0-6.0 '			n = 456
	Only altitude correction	Full modification	Improvement absolute	Improvement relative
RMS	3.608	3.852	0.244	6.8%
MB	-0.545	-0.416	-0.130	-23.8%
MAE	2.748	2.929	0.181	6.6%

	Class: ' greater_than_6 '			n = 1848
	Only altitude correction	Full modification	Improvement absolute	Improvement relative
RMS	3.584	3.305	-0.279	-7.8%
MB	0.271	-0.452	0.181	66.9%
MAE	2.785	2.616	-0.168	-6.0%

Appendix B

Analysis on an Independent Dataset

In this appendix, the statistical measures Root Mean Square (RMS), Mean Bias (MB), and Mean Absolute Error (MAE) of observed and modelled 2m-temperatures at the selected stations of the control group are shown. The modelled 2m-temperatures were corrected, using three different diurnal cycles (Mean_all4, Mean_PioRobZer, and Ulr). For more information about this approach we refer to Methods, p. 33 and Results, p. 88.

The data, considered for calculating the different statistical measures, consisted of the full-year dataset 2002/03. Note that for this analysis not only these days were considered that were actually corrected ("lt_2"-, "2-6"-, and "gt_6"-days).

Tables B.1 and B.2 present the results of this analysis at the four stations of the control group (Comprovasco, Engelberg, Samedan, and Scuol) at the "representative" gp.

The first column specifies the different statistical quantities, i.e. RMS, MB, and MAE. The second column (grey) shows the results in case of only the altitude correction was performed for modifying the model data. The next three columns (green, yellow, and red) present the results in case the data was additionally modified with different diurnal cycles and where indicated with the spreading procedure. To visualise which diurnal cycle led to best results, the cells are marked with different colours. Green indicates the most appropriate daily-cycle, yellow the intermediate one and red the worst diurnal cycle for a specific statistical quantity. Note, in case of the correction even with the most appropriate diurnal cycle also led to a worsening, we used a dark green.

The Tables B.3 and B.4 show the same kind of results in case the "min_dz25" gp was used.

Table B.1: The statistical results of the final analysis at the control group. The data of the "representative" gp was used and the correction included the different types of diurnal cycles while no spreading procedure was applied.
Colours:
grey: only altitude correction
green: best correction type (dark green: best correction, but still worse than "only alt. correction")
yellow: intermediate quality of the correction
red: worst correction type
Representative, no spreading

Comprovasco n = 8160		Only altitude correction	Full correction		
			Different types of corrected daily-cycle		
			Mean_all4	Mean_PioRobZer	Ulr
RootMeanSquare	RMS	3.154	3.193	3.003	4.049
Mean Bias	MB	1.105	1.238	0.884	2.022
Mean Absolute Error	MAE	2.345	2.404	2.273	2.937

Engelberg n = 8160		Only altitude correction	Full correction		
			Different types of corrected daily-cycle		
			Mean_all4	Mean_PioRobZer	Ulr
RootMeanSquare	RMS	3.254	3.249	3.318	3.601
Mean Bias	MB	0.121	0.220	-0.112	0.958
Mean Absolute Error	MAE	2.216	2.260	2.282	2.483

Samedan n = 8160		Only altitude correction	Full correction		
			Different types of corrected daily-cycle		
			Mean_all4	Mean_PioRobZer	Ulr
RootMeanSquare	RMS	5.694	5.159	5.694	5.097
Mean Bias	MB	0.121	0.332	-0.209	1.508
Mean Absolute Error	MAE	4.123	3.840	4.013	3.894

Scuol n = 8160		Only altitude correction	Full correction		
			Different types of corrected daily-cycle		
			Mean_all4	Mean_PioRobZer	Ulr
RootMeanSquare	RMS	4.886	4.573	4.785	4.749
Mean Bias	MB	0.681	0.896	0.395	1.970
Mean Absolute Error	MAE	3.534	3.337	3.488	3.524

Table B.2: The data of the "representative" gp was used and the correction included the different types of diurnal cycles and the spreading procedure.
For further remarks we refer to Table. B.1
Representative, with spreading

Comprovasco		Only altitude corrected	Full correction		
n = 8160			Different types of corrected daily-cycle		
			Mean_all4	Mean_PioRobZer	Ulr
RootMeanSquare	RMS	3.154	3.630	3.464	4.402
Mean Bias	MB	1.105	1.238	0.884	2.022
Mean Absolute Error	MAE	2.345	2.606	2.547	3.060

Engelberg		Only altitude corrected	Full correction		
n = 8160			Different types of corrected daily-cycle		
			Mean_all4	Mean_PioRobZer	Ulr
RootMeanSquare	RMS	3.254	3.186	3.257	3.544
Mean Bias	MB	0.121	0.220	-0.112	0.958
Mean Absolute Error	MAE	2.216	2.233	2.292	2.411

Samedan		Only altitude corrected	Full correction		
n = 8160			Different types of corrected daily-cycle		
			Mean_all4	Mean_PioRobZer	Ulr
RootMeanSquare	RMS	5.694	4.748	5.096	4.681
Mean Bias	MB	0.121	0.332	-0.209	1.508
Mean Absolute Error	MAE	4.123	3.524	3.785	3.512

Scuol		Only altitude corrected	Full correction		
n = 8160			Different types of corrected daily-cycle		
			Mean_all4	Mean_PioRobZer	Ulr
RootMeanSquare	RMS	4.886	4.770	4.974	4.939
Mean Bias	MB	0.681	0.896	0.395	1.970
Mean Absolute Error	MAE	3.534	3.490	3.696	3.484

Table B.3: The data of the "smallest_dz_all25" gp was used and the correction included the different types of diurnal cycles without the spreading procedure.
For further remarks we refer to Table. B.1
Smallest_dz_all25, no spreading

Comprovasco		Only altitude correction	Full correction		
n = 8160			Different types of corrected daily-cycle		
			Mean_all4	Mean_PioRobZer	Ulr
RootMeanSquare	RMS	3.154	3.394	2.995	4.388
Mean Bias	MB	1.105	1.515	1.029	2.263
Mean Absolute Error	MAE	2.345	2.544	2.268	3.131

Engelberg		Only altitude correction	Full correction		
n = 8160			Different types of corrected daily-cycle		
			Mean_all4	Mean_PioRobZer	Ulr
RootMeanSquare	RMS	2.666	2.668	2.648	2.821
Mean Bias	MB	-0.038	0.089	-0.072	0.400
Mean Absolute Error	MAE	1.788	1.836	1.804	1.918

Samedan		Only altitude correction	Full correction		
n = 8160			Different types of corrected daily-cycle		
			Mean_all4	Mean_PioRobZer	Ulr
RootMeanSquare	RMS	5.118	4.337	4.916	4.035
Mean Bias	MB	-1.122	-0.606	-1.218	0.330
Mean Absolute Error	MAE	3.403	3.047	3.326	2.974

Scuol		Only altitude correction	Full correction		
n = 8160			Different types of corrected daily-cycle		
			Mean_all4	Mean_PioRobZer	Ulr
RootMeanSquare	RMS	4.173	3.858	4.037	4.211
Mean Bias	MB	0.065	0.573	-0.027	1.467
Mean Absolute Error	MAE	2.909	2.772	2.848	3.042

Table B.4: The data of the "smallest_dz_all25" gp was used and the correction included the different types of diurnal cycles and the spreading procedure.
For further remarks we refer to Table. B.1.
Smallest_dz_all25, with spreading

Comprovasco n = 8160		Only altitude correction	Full correction		
			Different types of corrected daily-cycle		
			Mean_all4	Mean_PioRobZer	Ulr
RootMeanSquare	RMS	3.154	3.714	3.353	4.640
Mean Bias	MB	1.105	1.515	1.029	2.263
Mean Absolute Error	MAE	2.345	2.662	2.467	3.228

Engelberg n = 8160		Only altitude correction	Full correction		
			Different types of corrected daily-cycle		
			Mean_all4	Mean_PioRobZer	Ulr
RootMeanSquare	RMS	2.666	2.649	2.630	2.803
Mean Bias	MB	-0.038	0.089	-0.072	0.400
Mean Absolute Error	MAE	1.788	1.858	1.862	1.895

Samedan n = 8160		Only altitude correction	Full correction		
			Different types of corrected daily-cycle		
			Mean_all4	Mean_PioRobZer	Ulr
RootMeanSquare	RMS	5.118	4.297	4.881	3.993
Mean Bias	MB	-1.122	-0.606	-1.218	0.330
Mean Absolute Error	MAE	3.403	3.152	3.524	-2.921

Scuol n = 8160		Only altitude correction	Full correction		
			Different types of corrected daily-cycle		
			Mean_all4	Mean_PioRobZer	Ulr
RootMeanSquare	RMS	4.173	4.223	4.386	4.547
Mean Bias	MB	0.065	0.573	-0.027	1.467
Mean Absolute Error	MAE	2.909	3.038	3.182	3.142

Appendix C

Programming Error in LM-Code 2002/03

In this chapter we describe a programming bug in the LM source code (lm 2.17), that was unfortunately in the operational "aLMo" version between June 2002 and April 2003. In earlier versions, the calculation of the partial snow coverage was not identical in the parts "soil" and "radiation" with impacts on the variable t_g (cf. below, Formulas) and the albedo. Note that the temperature t_g directly influences the temperature on the 2 m-level. In the course of this unification, a programming bug in the subroutine `tgcom` was implemented, that caused wrong t_g (and therefore 2m-temperatures).
A correct version of `tgcom` was implemented in the operational aLMo (lm 2.18) at 10th. April 2003.

Formulas

The following set of equations represents the different versions that were used to calculate the temperature t_g.
In eq. C.1, the wrong version, used in lm 2.17 can be seen. Note the multiplication operator between the constant factor ($rhde$) and the snow water content (w_snow). Eq. C.2 shows the corrected version of eq. C.1, where the multiplication between w_snow and $rhde$ was replaced by a division. The actually implemented correct formula C.3 in `tgcom` (lm 2.18) is a linear approximation (Taylor, 1st. order) of eq. C.2 (in case of $w\_snow \leq rhde$).

t_g : temperature at the boundary surface - atmosphere

t_s : temperature at the ground surface

t_snow : temperature at the snow surface

$$t\_g = t\_snow + e^{-rhde \cdot w\_snow} \cdot (t\_s - t\_snow) \tag{C.1}$$

$$t\_g = t\_snow + e^{-w\_snow/rhde} \cdot (t\_s - t\_snow) \tag{C.2}$$

$$t\_g = t\_snow + (1 - \min(1, w\_snow/rhde)) \cdot (t\_s - t\_snow) \tag{C.3}$$

Consequences

We want to point out, that performing a multiplication of a constant factor (i.e. *rhde*) by the snow water content was correctly used in a former version (1998) where an exponential term $e^{-0.2 \cdot w\_snow}$ was applied and *w_snow* was expressed in [kg/m^2]. However, the multiplication of *w_snow* with *rhde* (0.015), instead of a division, led to values of *t_g* similar to the ones of *t_s* in case of more than 15 cm snow depth (i.e. *w_snow*$>$ 0.015). This programming bug had no influence in case of $w\_snow = 0$.

This was especially problematic during time periods where the soil was covered with snow and therefore was isolated from the atmospheric conditions above the snow layer. The consequences of this programming bug were not only too warm values of *t_g* (insulation-effect) but also suppressed diurnal cycles of the same variable.