Ensemble-based uncertainty prediction for deterministic 2 m temperature forecasts

Master's Thesis

Faculty of Science
University of Bern

presented by
Pascal Blanc
2009

Supervisor:
PD Dr. Christoph Raible
Institute of Physics, Climate and Environmental Physics
and Oeschger Centre for Climate Change Research, Bern

Advisor:
Dr. Vanessa Stauch
MeteoSwiss, Federal Office for Meteorology and Climatology, Zurich

Abstract

The demand for local weather predictions along with uncertainty information increases. Various end-users need probabilistic forecasts to assure and optimise their activities. The energy sector uses weather forecasts to predict the production and consumption of energy (e.g., solar and wind energies). Depending on the temperature and other controlling factors, buildings have to be cooled or warmed to ensure optimal comfort. In this sector, there is unexploited potential to save energy (and ultimately, CO2 emissions) and to reduce costs based on good probabilistic forecasts.

Numerical weather prediction (NWP) models have been strongly developed in the last decade. In contrary to global NWP models, high-resolution limited-area models (LAMs) allow better simulations of small-scale weather processes which play an important role for local forecasts. However, uncertainties arising from a complex interplay of the chaotic nature of atmospheric processes, model errors, and local conditions remain. Based on global ensemble prediction systems (EPS), limitead-area EPS (LEPS) downscale the ensemble forecasts (members) to the local scale. Along with their improved formulation, LEPS are best suited to provide an estimate of local uncertainties in weather predictions. However, due to the differences between model world and reality (e.g., different orography), local forecasts are often biased. Post-processing methods are used to adapt grid point values to the local characteristics (e.g., Kalman filter).

This study aims at combining the strengths of a LAM, i.e. the best local forecast available, and a LEPS, i.e. an uncertainty information along with a forecast. For this, the relationship (SSR) between the spread of the limited area ensemble prediction system COSMO-LEPS and the skill (measured as the root mean square error [RMSE] between the prediction and the local observation) of the LAM COSMO-7 for station-based 2 m temperature forecasts is first investigated. This is somewhat unusual as it combines the output of different model systems. The resulting relationship is predominantly linear indicating that the distribution of COSMO-LEPS ensemble predictions contains precious uncertainty information that can be used to predict uncertainty in the deterministic Kalman filtered COSMO-7 forecasts (COSMO-7-KAL), supporting the new approach introduced in this study. Following influencing factors with decreasing importance for the shape of the SSR have been identified: the season, the location and the time of the day. These influences motivate the use of adaptive station-based uncertainty models. The performance of an adaptive linear model (ALM) was investigated using an optimal period of 30 days of past data (NWP model outputs and observations) and evaluated for one year of independent data. The performance of the uncertainty prediction has been increased by 66% compared to the use of the spread derived from the raw COSMO-LEPS ensemble as uncertainty information. Given that the variability of the daily cycle in RMSE is not fully simulated with the ALM, an extended ALM is introduced where the residuals from ALM are added to the uncertainty predictions. The extended ALM allows increasing the performance of the ALM by 15% compared to the ALM. On average, the uncertainty predictions using the persistence (derived from the past errors of COSMO-7-KAL only) performs the same. However, for a majority of stations (58%) the extended ALM is better than persistence, especially for low-level stations. This shows that the day to day variability of the uncertainty predictions for temperature based on ensemble information is not

very high. Nevertheless, the new approach developed in this study opens new opportunities in the field of NWP models post-processing. In particular, this model-integrating approach could be applied to other variables such as humidity or wind speed. Furthermore, if successful, it could be seen as a logical extension to any statistical bias correction of deterministic model output to infer probability distributions given a suitable PDF of the prediction errors and an appropriate ensemble prediction system.

Contents

Chapter 1

Introduction

'No forecast is complete without a forecast of the forecast skill.'

(*Tennekes et al.*, 1987)

Motivation

More and more users are interested in *local* weather predictions with uncertainty information, i.e. probabilistic forecasts. The economy is interested in probabilities to minimise costs (*Murphy*, 1977). For example, the probability that the 2 m temperature will reach or undershoot the 0°C threshold can be of interest for the agriculture. Frost can damage young plants or blooming trees and lead to economic lost. If the probability that there is frost is higher than the *'cost-loss ratio'* (costs of the preventive action(s) divided by the economic loss if no action is conducted and the event occurs), provisions should be taken. Also the transport sector (e.g., probability of frozen roads or fog at an airport) or the energy sector are dependent on the weather and need accurate forecasts to guaranty and optimise their activities. To optimise the trade and distribution of electricity (e.g., from different sources), predictions of the production are needed. The production of renewable energies are particularly dependent on the weather (e.g., solar radiation or wind strength). The needs in electricity depend on the meteorological conditions (e.g., use of air-conditioning in summer when temperature exceeds some threshold). The value of probabilistic forecasts (in particular of temperature) for optimising building climate control is being investigated (www.opticontrol.ethz.ch, MeteoSwiss is partner of this on-going project). There is potential to save energy, reduce costs and increase the comfort (e.g., natural cooling of the buildings during night in summer). Probabilistic weather

predictions are also important as inputs in hydrological models (e.g., *He et al.* [2009]). In fact, such models need probabilistic meteorological inputs like precipitation and temperature to estimate discharge probabilities. The temperature is of particular interest in mountain regions such the Alps where snow melting in spring and glaciers melting in summer are main sources of water. Precipitation (rain and snow) and discharge forecasts are needed in natural risk management (e.g., prediction of flood events). A broad scope of applications can make use of (probabilistic) forecasts provided by numerical weather prediction (NWP) models. Also forecasters need probabilistic forecasts to estimate the certainty of the current model outputs and to communicate qualitatively and quantitatively the uncertainties of weather predictions to the public (*Ban*, 2006; *Eckert and Cattani*, 2006).

NWP performance increases, uncertainties remain

NWP models have been strongly developed in the last decades and their performance constantly increases with the computational power (*Lynch*, 2008). The challenge of NWP is to mathematically describe the atmospheric phenomena and their interactions (*Holton*, 2004). Those processes occur on manifold spatial and temporal scales: large-scale (e.g., planetary waves), synoptic scale (e.g., low/high pressure systems, persistent blocking situations), meso-scale (e.g., orographic effects) and micro-scale (e.g., short-lived small-scale turbulences, single clouds, cloud microphysics). For local weather predictions, the realistic simulation of small-scale processes is of importance. Therefore, beside global circulation models, higher-resolution models resolving smaller-scale processes have been developed. They allow a better representation of regional weather conditions. Nevertheless, uncertainties remain in weather predictions (*Palmer et al.*, 2005). Due to the chaotic nature of the atmosphere, minimal deviations in the initial conditions of NWP models lead to different forecasts of the large-scale flow development (*Lorenz* [1963] and chapter 2). The initial conditions are not known with exactitude. The meteorological observational network is heterogeneously distributed around the world. Moreover, the observations are punctual and do not allow the monitoring of the current state of the atmosphere in the three dimensions. Additionally, there are measurements errors. For those reasons, the *'predictability'* of weather is limited. This is the so-called predictability of the first kind, the predictability of the second kind arising from uncertainties in the boundary conditions (*Lorenz*, 1975). There is no analytical solution for the equations that describe the

atmospheric flow, therefore numerical methods are needed. The integration of the governing equations forward in time occurs on a three-dimensional (3D) grid. Due to the limited spatial resolution of NWP models the models topography is a smoothed representation of the real orography, especially in mountain areas (e.g., non-representation of a valley / mountain ridge), ignoring local characteristics (e.g., valley winds or orographic forcing of precipitation). Also, land cover specifications need to be averaged over whole grid cells. As a result, unresolved sub-grid scale processes (e.g., sub-grid scale effects of topography and land cover) are not directly resolved and must be parametrised empirically using known grid scale parameters.

Deterministic high-resolution limited-area models

One approach to reduce the uncertainties in NWP is followed by running limited-area high-resolution models. Limited-area models are best suited to provide local forecasts due to the high spatial resolution and therefore the better representation of physical processes at the respective scales. However, they cannot be calculated globally due to computational restrictions and they need boundary conditions from global models (e.g., from the European centre for medium-range weather forecasts [ECMWF], *Persson and Grazzini* [2005]). The uncertainty about the further development of the atmospheric flow on the large scale is flow dependent, i.e., it depends on the weather situation (chapter 2). These 'large-scale' uncertainties propagate into limited-area models through the boundary conditions.

Ensemble prediction systems

Due to the sensibility of NWP models to initial conditions, *'ensemble prediction systems (EPS)'* have been developed based on Monte-Carlo experiments (*Epstein*, 1969; *Leith*, 1974; *Palmer et al.*, 2005; *Leutbecher and Palmer*, 2008). Instead of performing one single deterministic forecast, a set of predictions (so-called *'members'*) is calculated using disturbed initial conditions (chapter 2). EPS allow to estimate probabilities (e.g., probability of precipitation, probability of exceeding a certain threshold of precipitation or temperature). Ensemble predictions enable the estimation of the forecast accuracy or level of confidence for the ensemble mean [EM] which is linked to the predictability of the current flow. In other words, from the divergence of the members around the mean one can infer an error probability density function (PDF) or assess the uncertainty of the current forecast (*Ehrendorfer*, 1997). A measure

of divergence is the spread, defined here as the standard deviation of the members forecasts around the EM. Note that ensemble averaging does not lead automatically to realistic synoptic situations, especially by large spread.

In the last decade, the probabilistic approach gained importance in NWP and also limited-area EPS (LEPS) have been designed. They have a higher spatial resolution than global EPS and more complex model formulations. Therefore, they are more capable of capturing local uncertainties. An example of LEPS is COSMO-LEPS, the limited-area probabilistic model used in this study. Nested in the EPS system of the ECMWF (*Molteni et al.*, 1996; *Leutbecher and Palmer*, 2008), COSMO-LEPS is developed within the *'COSMO'* consortium (*Steppeler et al.*, 2003) and delivers operational local ensembles for Europe (*Marsigli et al.*, 2005). Figure 1.1 shows the topographies applied in the ECMWF-EPS (roughly 80 km horizontal resolution) and in the COSMO-LEPS (10 km). Better forecasts can be expected from COSMO-LEPS, especially in regions with complex terrain like the Alps.

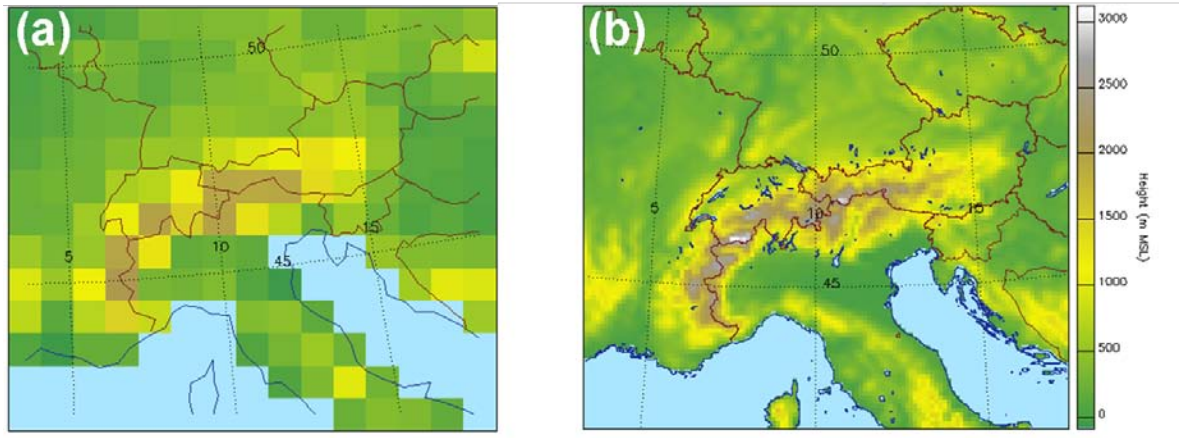

Figure 1.1: (a) ECMWF EPS ($80 \times 80km$) and (b) COSMO-LEPS ($10 \times 10km$) topographies over central Europe. Courtesy: *A. Walser, MeteoSwiss.*

Statistical post-processing increases model performance

Ångström (1814-1874), well-known for his research in spectroscopy, already recognised sources of uncertainties in weather predictions such as incomplete observations and suggested the use of statistics and physics to produce forecasts (*Liljas and Murphy*, 1994). Various statistical

postprocessing methods have been developed: Well-known are the *'model output statistics'* (MOS, *Glahn and Lowry* [1972]) and *'Kalman filtering'*, (KAL, *Persson* [1989]) amongst many others. All statistical methods are used to adapt the direct model outputs (DMO) to local characteristics using historical model outputs and observations, whereas the causes of forecast errors in the NWP models remain. They statistically remove the systematic error (*bias*) of the DMO predictions.

Similar to deterministic forecasts, the EM of an EPS can be biased. To correct for this bias, the same post-processing methods can be applied by treating the EM as one deterministic model run. The divergence of the ensemble (the spread) is often too small compared to the variability of errors of the EM (*underdispersion*). In an perfect EPS, the averaged spread (e.g., over a season) would be equal to the root mean square error (RMSE, *Palmer et al.* [2005]) of the EM. The relation between both, the spread and the RMSE of the EM, is called *'spread-skill relationship'* and is one important definition for this study. The underdispersion (RMSE>spread) in EPS is due to underdisturbed initial conditions, no consideration of inherent model uncertainties and finite ensemble size (*Buizza et al.*, 2005). To consider model inherent uncertainties, various parametrisation schemes can be used for the same subgrid scale process. There are different methods to produce initial conditions and there is no agreement about a 'best method' (*Hamill et al.*, 2002).

To tackle the underdispersion and the bias in EPS, various post-processing method exist (*ensemble-MOS methods*). They are station-based and need observations. Ensembles can also be based on predictions generated by different models ('poor man's' or multi-ensemble approach) for which similar post-processing methods are used. *'Bayesian model averaging'* (BMA, *Raftery et al.* [2005]) weights the single members based on historical data (model outputs and observations). *Ensemble dressing* is similar to BMA, with the difference that the shape of the distributions can be different for each member (*Roulston and Smith*, 2003). Also the use of *'reforecasts'* can be useful. This is done at MeteoSwiss for 2m temperature, precipitation and wind gust based on 20 years COSMO-LEPS hindcasts and observations (*Fundel et al.*, 2009). For each station and each variable, a PDF for the reforecasts and one for the observations are built. The relation between both PDF is then used to calibrate the ensemble members of future forecasts. *Hamill et al.* [2004] used reforecasts to estimate MOS coefficients for logistic regressions between the ensemble mean and the PDF of the predicted

variable (temperature or precipitation).

Jewson [2004] and *Gneiting et al.* [2005] suggested to include the spread as predictor in a linear regression framework (*non-homogeneous Gaussian regression NGR*) where the spread is used to define the standard deviation of the PDF. *Wilks* [2006] compared several ensemble-MOS post-processing methods and found ensemble dressing, logistic regression and NGR to be most promising post-processing methods in the Lorenz model. However, empirical studies are needed to confirm these results for more realistic NWP models.

Hitherto, the studies focused on the post-processing of EPS or deterministic forecasts separately. The combination of the strengths of deterministic and ensemble forecasts was only poorly investigated. Deterministic models have often higher spatial resolutions and more complete formulations. However, they are single forecasts without uncertainty information. On the other hand, EPS provide several forecasts and thus, include an indication about the uncertainty of the forecasts. Therefore, it would be intriguing to benefit from the strengths of both NWP approaches to produce improved forecasts. *Rodwell* [2005] combined deterministic and ensemble forecasts for precipitation. He 'enlarged' the ensemble by weighting the deterministic forecast more strongly than the single members of the ensemble to better predict the occurrence of precipitation, the weights decreasing with lead time. Apparently, there are no studies about such combinations for 2 m temperature. Therefore, the goal of this study is to use the spread information contained COSMO-LEPS to predict the uncertainty of deterministic COSMO-7 2 m temperature forecasts. In contrary to *Rodwell* [2005], the deterministic forecast will not be treated as a member of the ensemble. Rather, the spread of the ensemble will be used to derive a distribution around the deterministic forecast.

The thesis is organised as follows: In chapter 2, a short introduction to numerical weather prediction is given. The principles of NWP models are illustrated along with the main sources of uncertainties remaining in NWP models. The deterministic model COSMO-7 and the probabilistic model COSMO-LEPS are described. Chapter 3 evaluates the performance of 2m temperature predictions of the COSMO models. The data used and their representativeness are shortly discussed in chapter 4. In chapter 5, the spread-skill relationship is investigated. First, the spread-skill relationship between the spread of COSMO-LEPS and the RMSE of the EM is investigated. In a second step, COSMO-7-KAL is introduced and the relationship between its RMSE and the spread of COSMO-LEPS is investigated. The data analysis allows

to identify potential uncertainty models. Based on these investigations, uncertainty models are identified and evaluated in chapter 6. In the conclusion, a suitable uncertainty model setting is proposed for operationalisation at MeteoSwiss. In the outlook, further investigations are suggested.

Chapter 2

Numerical weather prediction

'One flap of a seagull's wing
may forever change
the future course of the weather...'
Lorenz, 1963

2.1 NWP model formulation

Charney published 1953 a simple set of equations capable of simulating planetary waves (birth of the global circulation models) and demonstrated the potential of NWP (*Charney*, 1953). The atmosphere can be described by the *Navier-Stokes* non-linear equations for fluid motion containing seven unknowns: pressure p, temperature T, 3 wind components (U, V, W), density ρ and water content q. These equations satisfy mass continuity, momentum and heat conservation. There is no analytical solution for these complex, partial differential equations (PDE) and the only possibility is to solve them numerically (*Lynch*, 2008). For that, a 3-dimensional grid has to be defined (Figure 2.1) and the equations are discretised. With given initial conditions, the time integration of the governing equations allows to predict the new state after a predefined time step. Grid point distance, time step and maximal velocity are important parameters to assess the stability of the integration scheme (Courant-Friedrichs-Levy condition), which will affect the forecast quality (*Ruddiman*, 2001).

The limited spatial and temporal resolutions of NWP models do not permit to describe directly sub-grid scale and short-lived physical phenomena (e.g., small-scale turbulences, single

clouds, thunderstorms). Therefore parametrisations are needed to account for such processes. Parametrisations are either statistical, physical or both together. For example the occurrence of clouds or the tendency for thunderstorms are estimated by means of known large scale parameters in the corresponding grid boxes. For instance, the instability of the air column, which can be depicted in the temperature and dew point profile, or humidity in the boundary layer and wind field play a crucial role to enable deep convection. The induced turbulences mix the components (water, aerosols, etc.) in the troposphere (e.g., through updraft, downdraft, entrainment and detrainment processes) and change the temperature profile (stabilisation). Single cells can develop into self-sustained and long-lived systems like super cells or organised mesoscale convective systems when certain conditions are fulfilled. Such events can cause severe damage (e.g., hail damage). However, most NWP models are not *'convection resolving'* due to their spatial resolution ($<$ O(1 km) is needed) and several microphysical interactions can still not be included in the physics of NWP models (e.g., hail formation), so that the model formulations remain incomplete. Current research activities go toward *'stochastic-dynamic parametrisations'* (*Palmer et al.*, 2005; *Berner et al.*, 2005; *Palmer and Hagedorn*, 2006) that account for uncertainties in the parametrisations.

2.2 Data assimilation

An immense amount of weather data is continuously collected and made available for weather centres all over the world through the *Global Telecommunication System* (GTS, Figure 2.2). These data have to be processed prior use in NWP models such as the ECMWF integrated forecast system (IFS). Synoptic, ship, buoys, aircraft, radio soundings, profilers and satellite observations are controlled and selected. The step of optimally combining observations with short-term model forecasts is called *data assimilation*. Combined with recent model outputs, *analyses* are performed delivering initial conditions, also for data-sparse region such as the southern Pacific. As an example, the ECMWF applies a variational assimilation in the space-time domain (4D-Var) that aims at matching best analyses and observations over a defined time interval by minimising a 'cost function' whereas COSMO-7 uses a latent-heat nudging scheme (see later in this chapter).

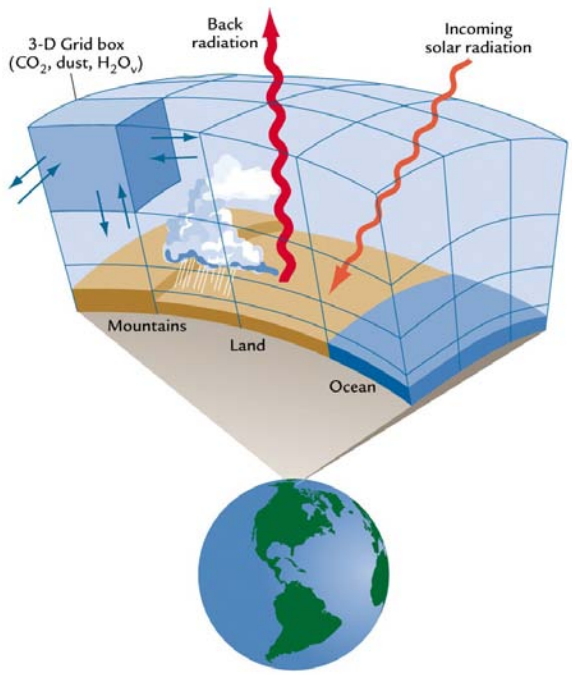

Figure 2.1: The principle of weather and climate modelling: The atmosphere is fragmented into numerous grid boxes for which the thermodynamic state and concentrations of chemical components are averaged. Topography and land type distribution are adapted to the model resolution. Sub-grid scale processes are parametrised. Source: *Ruddiman,* [2001].

2.3 Ensemble forecasting

In 1963, Edward N. Lorenz, the father of the *chaos theory*, published a paper with simulations that illustrated the sensitivity to initial conditions of a simple atmospheric flow model derived by Barry Saltzman (*Lorenz*, 1963). Minimal deviations of the initial conditions from the reality lead to various possible developments of the later state (Figure 2.3). With time, the different forecasts evolve differently depending on the state in the phase space. This divergence, or *'spread'*, can be used as indicator of *predictability*. The sensitivity of weather prediction to initial conditions (*predictability of the first kind*) underscores the importance of suitable initial conditions of the NWP models derived from a global measurements network and a data assimilation procedure.

At the time of the discovery of Lorenz, the computational power was not large enough to compute ensemble predictions, thus the meteorologists had to be satisfied with single, *deterministic* forecasts. The forecasters continued to estimate qualitatively the reliability of their

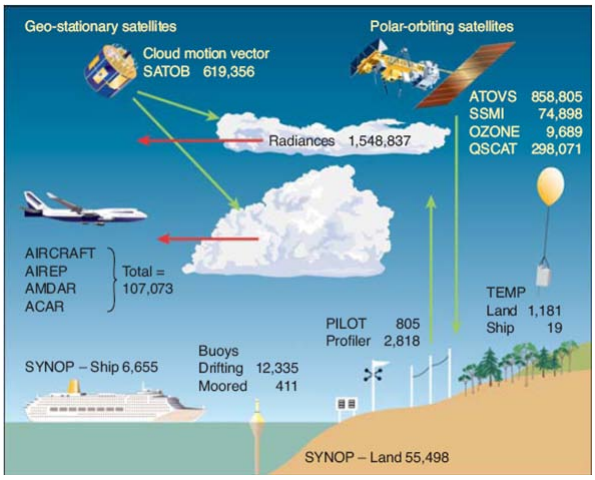

Figure 2.2: An immense amount of weather data is continuously collected and made available for weather centres all over the world through the *Global Telecommunication System (GTS)*, representing precious information for initialising NWP models. Here the observations received at ECMWF are shown for the 5 July 2004. Source: *Persson and Grazzini* [2005].

predictions by experience until operational probabilistic systems became realisable in the early 1990's (*Murphy* [1998] for an overview about early history of probabilistic forecasts). During that time span, better data assimilation (e.g., of satellite observations), improved deterministic numerical models and enhanced resolution led to large improvements in numerical weather and climate prediction. Nevertheless, due to the chaotic nature of the atmosphere, remaining uncertainties in the initial conditions and in the model formulations, the predictability of weather is still limited. Therefore, large efforts were made toward ensemble prediction systems (EPS) in the last decade. There are various proceedings to design an EPS (short review: *Hamill et al.* [2002]). The initial conditions have to be disturbed in such a way that the ensemble covers the full uncertainty range (*Ban*, 2006; *Palmer and Hagedorn*, 2006). Many different approaches to generate perturbed initial conditions have been developed but there is no agreement about which method is best (*Hamill et al.* [2002]).

2.4 Weather predictability

Predictability is flow dependent (see Figure 2.3). Applied on weather, this means that predictability is a function of the large-scale situation. For example, the development of a sus-

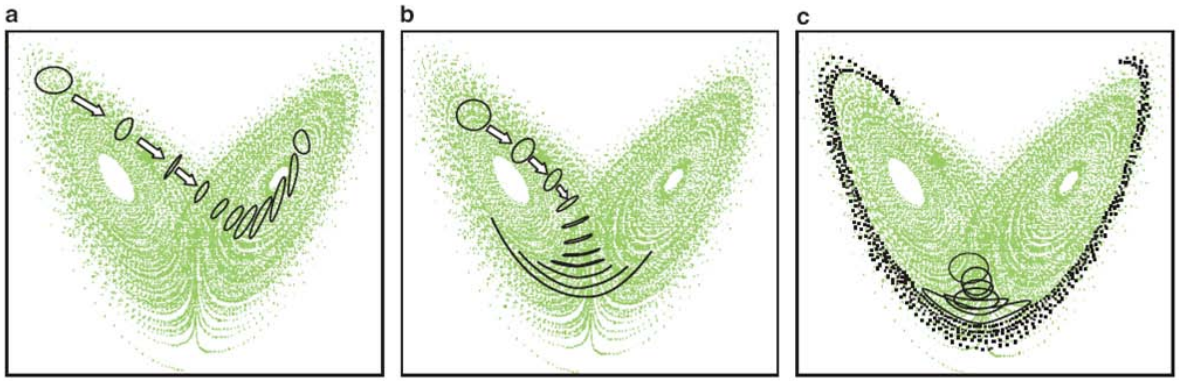

Figure 2.3: A well-known representation of the simple atmospheric model by Lorenz [1963] with a butterfly shaped phase space. In each figure, the primary circle delimits the set of initial conditions with equal uncertainty. Their temporal evolution is also plotted. Both fix points can be exemplarily interpreted as *'sunny'* or *'rainy'*. The uncertainty growth with time is flow dependent (common feature of non-linear systems such as the atmosphere): a) high predictability of the system, small ensemble spread during the whole period; b) low predictability, strong enlarging ensemble spread with time; c) flow becomes unpredictable after a very short time. Source: *Palmer et al.* [2005].

tained zonal flow with repeated cyclogenesis is 'less predictable' than a stable high pressure system (well-known example: storm 'Lothar', 26 und 27 December 1999, Western Europe). Blocking situations for example are generally quite stable and can hold two weeks or more. However, the correct forecast of their onset is difficult. Early predictions of such events have a high societal and economical relevance as they can cause long-lasting climatic extremes like the well-known 2003 summer heat wave in Western Europe (impacts on mortality, water supply, agriculture, air pollution, energy production, etc.). They are the consequence of a subtle interplay of synoptic and planetary scales (*Glisan and Lupo*, 2008). Note that the global model of the ECMWF underestimates the blocking frequency by 10% in summer and 40% in winter in the medium to extended forecast range (*Persson and Grazzini*, 2005). In the Northern Hemisphere zonal flow, storm tracks and jet streams are slightly shifted poleward (ECMWF model). As the large-scale setting from ECWMF is used to drive limited-area models (e.g., COSMO models used in this study, see next section), ECMWF model errors propagate into them. Concerning the predictability of temperature (ECMWF model), the next three days are generally well predictable. Between 3 and 5 days, there is skill in forecasting daily extremes and between 5 and 10 days in estimating daily means (*Persson and*

Grazzini, 2005).

Interestingly, the density and quality of observations where the atmospheric flow passes impacts particularly the predictability (Figure 2.4). According to the origin of the air masses (regions with abundant or sparse observations), the development is more or less predictable, respectively (*Persson and Grazzini*, 2005). Especially in 'dynamically sensitive' baroclinic regions where instabilities produce eddy kinetic energy which propagates downstream, dense observational data are necessary to assure good predictability (in terms of timing, location and intensity of the weather phenomena) downstream.

In summary, weather predictability depends on the:

- model quality (e.g., resolution, formulation);
- initial conditions (predictability of the first kind);
- boundary conditions (predictability of the second kind);
- current weather situation (predictability is flow dependent);
- location on Earth (there are 'more and less predictable' regions depending on the weather variability and local characteristics: e.g., deserts vs. Alps).

A large scope of uncertainties (from the large-scale to the local scale) are responsible for the errors in local weather predictions. However, for short-term point forecasts, the local to regional uncertainties are expected to be dominant. Regions with complex topography (e.g., the Alps) need high-resolution NWP models to simulate small-scale complex processes (fronts deformation, lee cyclogeneses, valley winds, orographic effects, snow and fog). Therefore, deterministic and probabilistic limited-area models (LAM) have been developed. The LAMs used in this study are introduced in the next subsection.

2.5 The COSMO model suite

The Consortium for Small-scale Modeling (COSMO) was formed in October 1998. Its general goal is to *'develop, improve and maintain a non-hydrostatic limited-area atmospheric model (COSMO), to be used both for operational and for research applications by the members of*

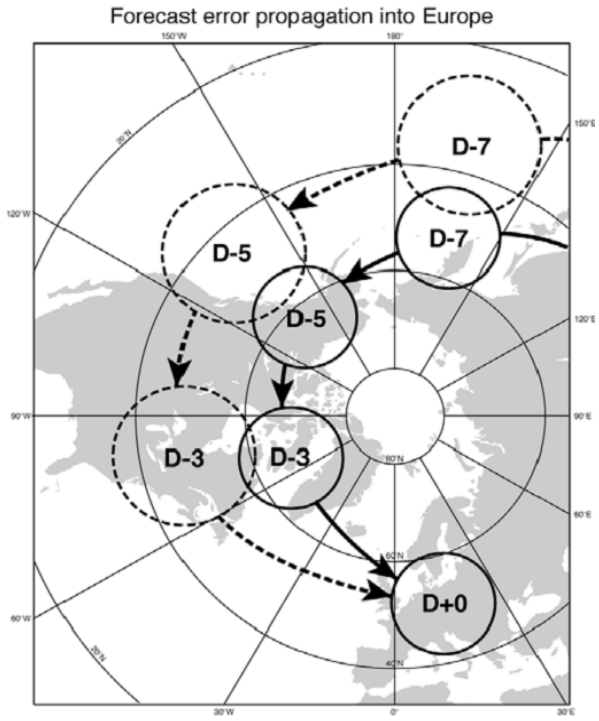

Figure 2.4: Error propagation impacts predictability. Areas (circles) in the northern hemisphere (NH) where analysis errors D-N days back in time will most affect the predictability over Europe at D+0. During summer, the main flow passes over the coarse observation network of Alaska, Northern Canada and Greenland. At wintertime, the zonal flow is predominant and crosses the dense US network (dashed circles) increasing predictability over Europe. Source: *Persson and Grazzini* [2005].

the consortium' (currently Germany, Italy, Greece, Poland and Romania and Switzerland)[1]. The approach of COSMO is to dynamically downscale the global ECMWF model to higher horizontal and vertical resolution over Europe. COSMO is designed to combine reliable large-scale predictions of the atmospheric flow from the ECMWF model with a better description of the local meteorological processes (higher resolution, more complex parametrisation schemes, see *Steppeler et al.* [2003]). COSMO is formulated in terrain-following vertical coordinates (*Schär et al.*, 2002; *Leuenberger*, 2002). The vertical grid-spacing in COSMO varies from roughly 50 m close to the surface to a few 100 m at the top of the model. Sub-grid scale

[1]http://cosmo-model.cscs.ch/, 21.08.2008. In contrary to global models, COSMO considers vertical motion in the model formulation ('non-hydrostatic'). For the COSMO model, an extensive documentation is accessible under http://www.cosmo-model.org/content/model/documentation/core/default.htm. In particular, 3 parts are available about *Dynamics and numerics, Physical parametrisation and Data assimilation.* The chapter *Postprocessing* is not disposable yet.

processes are parametrised: vertical diffusion (turbulence), cloud and precipitation formation (condensation), convection and radiation. A one-dimensional soil model is coupled to the atmospheric model.

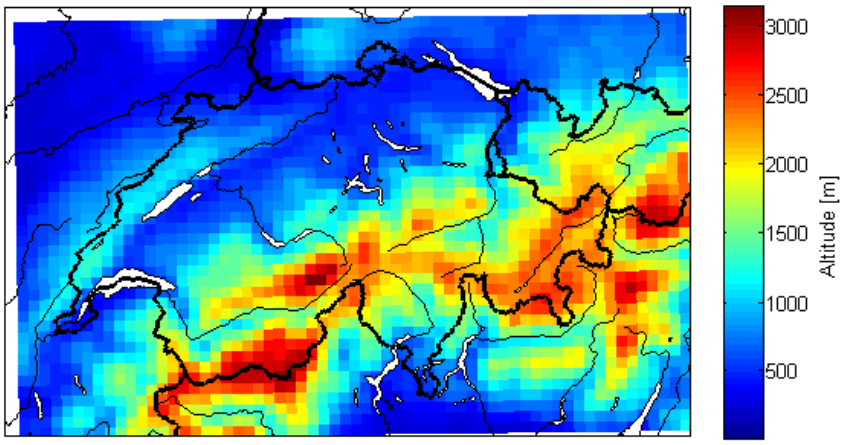

Figure 2.5: COSMO-7 topography (6.6 km resolution) over Switzerland with most of the inshore waters and the political boundaries. The colour shading corresponds to the altitude in [m]. This mask (in grey shades) is used in the background of the maps present in the thesis for highlighting the effect of topography.

Different COSMO model suites are operational at MeteoSwiss and other weather centres such as the 'Deutscher Wetterdienst' (DWD) in Germany (*DWD*, 2002). At MeteoSwiss, two deterministic models with different spatial resolution, frequency of new forecasts and prediction horizons are operational. They are computed at the Swiss national supercomputing centre (CSCS). COSMO-7 produces hourly predictions for the next 3 days on a 7×7 km grid (two daily runs at 00 and 12 UTC). In February 2008, the resolution of COSMO-7 has been reduced to 6.6 km to allow the operationalisation of the new COSMO-2 version with 2.2 km resolution. Figure 2.5 shows the 6.6 km topography of COSMO-7 over Switzerland. With COSMO-2, short-time forecasts for the next 24 hours (actualised every 3 hours) are performed for the Alpine region. With a higher resolution of the topography and a more complex description of microscale processes (e.g., turbulence, cloud formation), more precise and accurate forecasts – especially of local wind systems and precipitation events (e.g., thunderstorms in summer) – are expected. The data assimilation of COSMO-7 and COSMO-2 is based on nudging (*Leuenberger*, 2005; *Stoll*, 2005). Note that the assimilation of non-prognostic

variables is not straightforward. Diagnostic variables are variables that must be deduced from prognostic grid point values or that must be interpolated. 2 m variables are diagnosed given that the lowest level in NWP models is above 2 m. In COSMO-7 and COSMO-2 2 m temperature observations are not directly assimilated to avoid potential negative impacts in the boundary layer (*Stauffer et al.* [1991]). The 2 m temperature data is only employed to determine the stations' specific humidity content that is considered for assimilation, the 2 m temperature itself is not assimilated. Therefore, 2 m temperature model outputs and station observations can be considered as almost independent. Beside the higher resolution and the better formulation of limited-area models, the own assimilation of meso-scale observations is necessary to improve the forecasts of limited-area models compared to global models (*De Elía et al.*, 2002), especially in regions with complex terrain like the Alps (*Schraff*, 1997).

In addition, the members of the COSMO consortium have developed a limited-area ensemble prediction system (*Montani et al.*, 2003). COSMO-LEPS, issued at 12UTC each day, delivers 3-hourly forecasts for the next 5.5 days with 10 km spatial resolution. In Germany, an ensemble based on the convection resolving model COSMO-DE (roughly 3x3 km) is in development (www.dwd.de/modellierung).

2.6 The COSMO-LEPS methodology

The COSMO-LEPS model has been developed by researchers at ARPA-SMR (Italy), primarily for the prediction of extreme precipitation events (*Montani et al.*, 2003). COSMO-LEPS demonstrates a high skill in forecasting intense and localised events and is very useful for meteo-hydrological applications (*Marsigli et al.*, 2005). The boundary and initial conditions are delivered by the ECMWF-EPS. This is a downscaling approach, alternatively it is also possible to use own assimilations and initialisations (e.g., COSMO-DE ensemble). The ECMWF-EPS ensemble consists of 50 members and a control run initialised with the resolution-degraded 'best guess' initial conditions used to perform the deterministic ECMWF forecast (total: 51 members per run). Starting from probability density functions (PDF) of possible initial conditions for each grid point, 50 forecasts ('members') are generated by randomly selecting conditions from these PDFs. A 'singular vector' approach is applied to identify directions with regard to the control initial conditions that will lead to the largest error growth over a particular forecast period in order to cover the full uncertainty range as

realistic as possible (*Molteni et al.*, 1996).

To select a suitable range of boundary and initial conditions for COSMO-LEPS, the two daily EPS runs (00 and 12 UTC) are used to build a 'super-ensemble' of 102 members (Figure 2.7). These are first grouped into 16 clusters through a hierarchical analysis. Note that there is no 'best' clustering method, moreover it depends on the weather situation (*Persson and Grazzini*, 2005). The geopotential height Z, the wind components (U, V) and the specific humidity Q at three pressure levels (500, 700 and 850 hPa or roughly 5500, 3000 and 1500 m, consecutively) and at two forecast times (4^{th} and 5^{th} forecast days, 12 UTC) are used as clustering variables. The clustering and integration domains are shown in Figure 2.6. The COSMO-LEPS domain extents over central and southern Europe at a resolution of roughly 10 km (resulting in 78'948 grid points) and with 40 vertical levels (ECMWF-EPS for comparison: 80 km/40 levels, interpolation of initial and boundary conditions is needed).

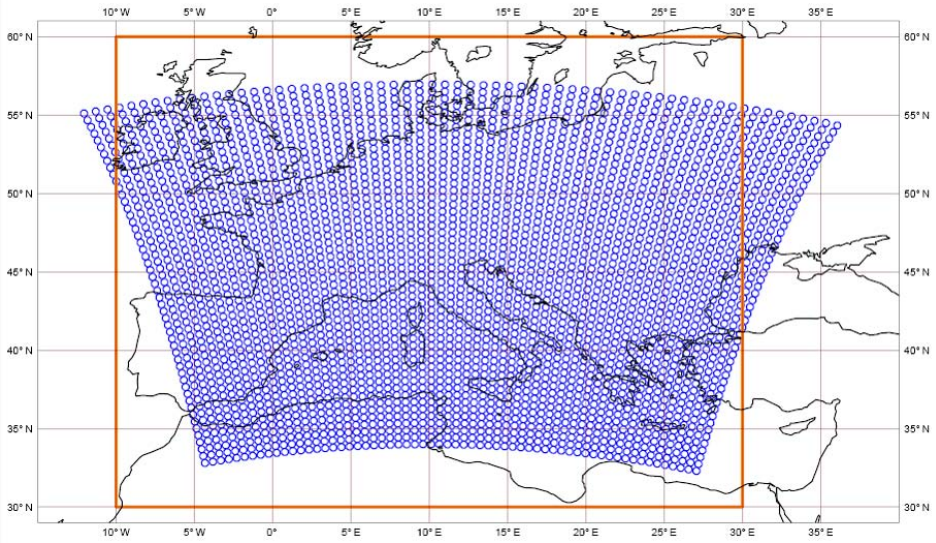

Figure 2.6: The clustering and integration domains of COSMO-LEPS. The clustering area (30°N-60°N, 10°W-30°E) is slightly larger than the integration domain (blue circles). Source: *Montani et al.*, 2004.

Within each of the 16 clusters, one representative member (RM) is selected (*Molteni et al.*, 2001).The RM has to be 'closest' to the members of its own cluster, but most 'distant' from the remaining members in terms of the clustering variables. Each of the 16 RM provides initial and boundary conditions for one simulation with COSMO-LEPS, so that a small-size high-resolution ensemble can be built. The clustering and RMs selection methods offer a

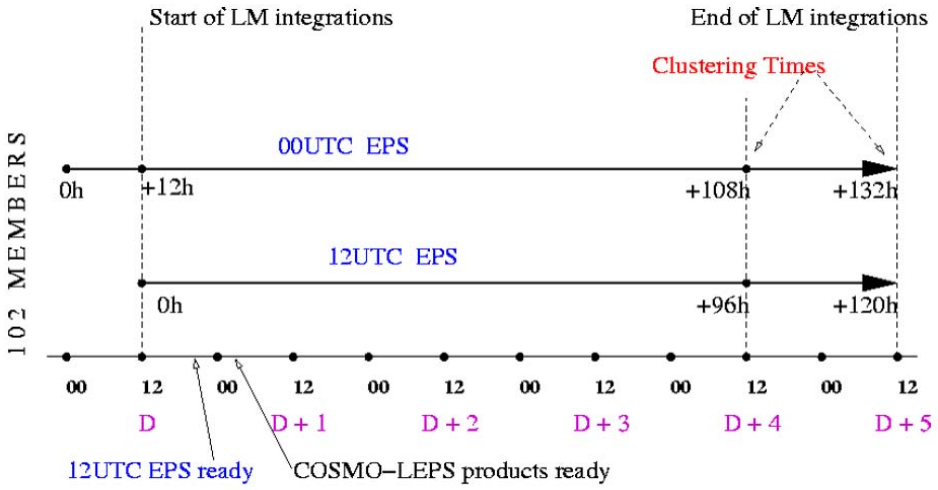

Figure 2.7: Schematic of the COSMO-LEPS methodology in the operational chain. For explanations, please refer to the text. Source: *Montani et al.*, 2004.

compromise between processor time constraints and full coverage of uncertainty. To account for model inherent uncertainties, the convection scheme (Tiedtke or Kain-Fritsch: *Tiedtke*, 1989; *Kain*, 2004) is randomly chosen.

The 12 UTC ECMWF-EPS is available in the evening (ca. 22 UTC) and COSMO-LEPS around 00 UTC. Several fields are postprocessed to get probabilistic products (e.g., precipitation's intensity probability maps, probability of the 2 m temperature being above or below a certain threshold in the last 24 h, probabilities of snow fall and wind strength, etc.) or station-based products like meteograms (precipitation, 2m temperature, surface wind, see Figure 2.8).

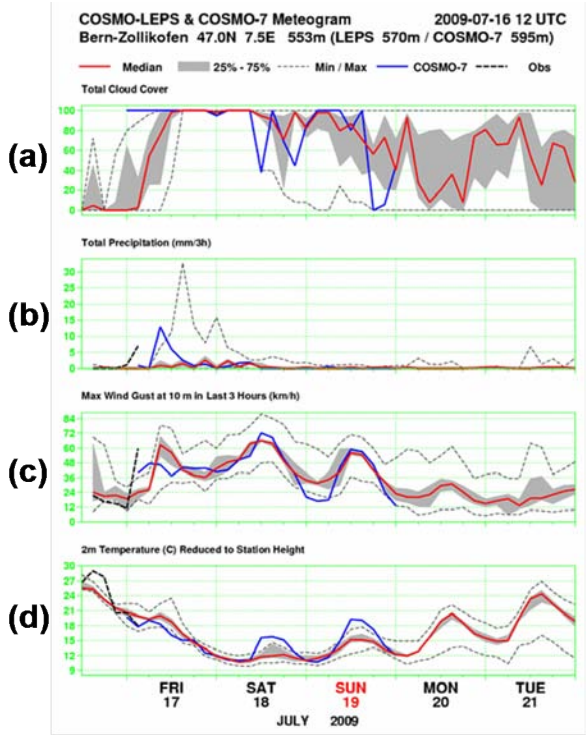

Figure 2.8: Example of a COSMO-LEPS and COSMO-7 meteogram issued 16.07.2009 12 UTC. The panels show the simulations for the next 5.5 days for the station Bern-Zollikofen: (a) total cloud cover [%], (b) precipitation in the last 3 hours [mm], (c) 10 m max. wind gusts in the last 3 hours [km/h] and (d) 2 m temperature [C]. There are 16 ensemble members. The median is plotted in red, the interquartile range (25 to 75%) in grey. The dotted lines are the minimum and maximum values. In blue the deterministic COSMO-7 run (higher spatial resolution). The spread of the COSMO-LEPS members is linked with the degree of predictability (predictability decreases with increasing lead time). Source: *MeteoSwiss*.

Chapter 3

Data and methods

3.1 Observations

Three years (January 2006 to November 2008) of quality controlled high-resolution instantaneous 10 minutes 2m temperature data ($\pm$ 0.1°C) are available for the automatic SwissMetNet stations of MeteoSwiss (Figure 3.1). A year is defined as the period December to November. For the year 2008, this means from the 1$^{\text{st}}$ of December 2007 to the 31$^{\text{th}}$ of November 2008. The seasons are defined as follows: winter (December to February, DJF), spring (March to Mai, MAM), summer (June to August, JJA), autumn (September to November, SON).

The height of 2 m is the standard of the World Meteorological Organisation's (WMO) height for measuring near surface temperature. The stations should measure the weather variables in 'undisturbed' locations. 67 stations are used in this study, they are homogeneously distributed covering regions of the Swiss Plateau, the Jura, the Alps and the Tessin. A wide range of altitudes is included from 203 m (Magadino, MAG) to 3580 m (Jungfraujoch, JUN). The maintenance of the dense Swiss observational network is expensive and time-consuming, but it is of high importance. First, these data are necessary for climate monitoring (climate warming, extreme events frequency). Second, any NWP model needs initial conditions which are delivered through the assimilation of observations. A high observations density is particularly important in regions with complex topography like the Alps (*Schraff*, 1997). Third, current observations are necessary to post-process DMO in real time (e.g., using a Kalman filter). As last example, observations are necessary in order to verify the model outputs, for instance to investigate cases when the model does not perform well or for combining rain

gauge and radar information to get high-resolution precipitation fields (*Erdin*, 2009). In this study, we neglect the uncertainties due to observation errors. We assume that they are far smaller than model forecast errors (large signal to noise ratio).

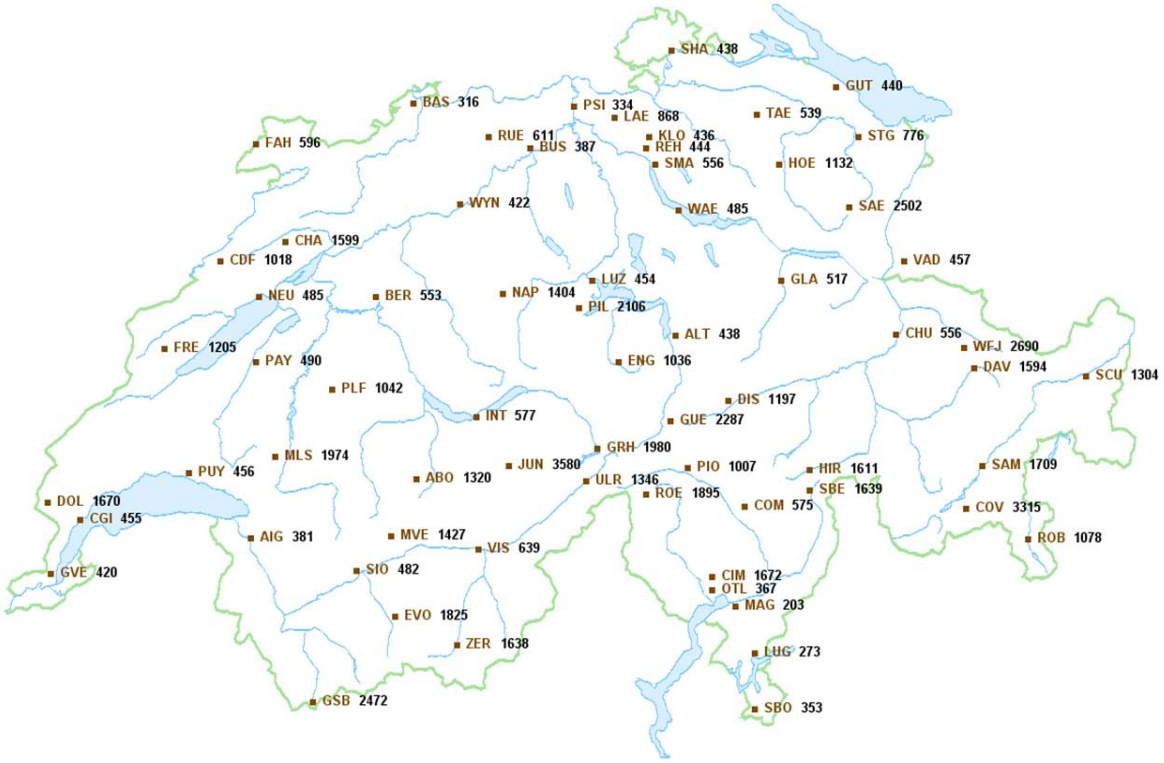

Figure 3.1: Stations of the Swiss meteorological network used in this study, the codes are given with stations height [m]. Source: *MeteoSwiss*.

3.2 COSMO model outputs

Daily at 12 UTC and for the period January 2006 to November 2008, deterministic (COSMO-7) and probabilistic (COSMO-LEPS) 2 m temperature forecasts are available for the stations used in this study. COSMO-7 provides three days of hourly forecasts, while COSMO-LEPS delivers 5.5 days of three-hourly forecasts. The COSMO-7 DMO are operationally post-processed at MeteoSwiss with an algorithm based on a Kalman filter to remove systematic errors (hereafter *COSMO-7-KAL* data). COSMO-7-KAL can be assumed bias free. The temperature is used as error predictor in the algorithm as well as the error structure of the

previous day, so that the errors emerging from systematically unresolved processes can be partly corrected. Note that the 'lead times' are defined as the time spans in hours between the forecast issuing time (here, daily at 12 UTC) and the times the forecast is valid for.

In Table 3.1, the most significant changes in COSMO-7 and COSMO-LEPS in the last three years that are relevant for the 2 m temperature forecasts are summarised. In particular, changes in boundary conditions relaxation, grid mesh, amount of vertical layers, soil model and temperature diagnostics have induced ameliorations of the forecast quality. For the stations used in this study (last section), Table 3.2 gives stations name, abbreviations, locations, stations height and the differences between stations height and grid points height.

Date	**Change**	**Impact**
01.02.2006 (00 UTC)	New high-resolution global forecasting system of ECMWF (horizontal mesh size of about 25 km).	Higher ECMWF forecast skill, (http://www.ecmwf.int/products/changes/high_resolution_2005.html), positive effects on COSMO-7 expected (COSMO-News 19).
February 2006	New COSMO-LEPS version (*Montani et al.*, 2008): Ensemble size enlargement from 10 to 16 members weighted according to the cluster populations. Random choice between two convection schemes (Tiedtke or Kain-Fritsch).	Better cover of the uncertainty range, in particular of the model inherent uncertainties arising from the parametrisation of convection.
18.12.2006 (12 UTC)	COSMO-7 assimilation cycle: new snow analysis, using observations from satellite (Meteosat Second Generation) and snow height measurements from ground stations.	More realistic 2 m temperature forecasts in winter (COSMO-News 22).
31.10.2007 (12 UTC)	New COSMO-7 version 4.0.4 (new microphysics and treatment for the lateral boundaries for snow and rain), new numerical solver: Runge-Kutta.	The new Runge-Kutta scheme is more accurate than Leap-Frog at the same efficiency. Strong positive effects during the winter period for almost all parameters (COSMO-News 25).
01.12.2007	COSMO-LEPS: Introduction of perturbations in the turbulence scheme (*Montani et al.*, 2008).	Improved consideration of uncertainties in the model formulation.
19.02.2008 (12 UTC)	COSMO-7: Grid mesh of 6.6 km instead of 7 km and new rotated grid centred over Switzerland; 60 vertical layers (instead of 45); The lowest model layer is at 10 m (instead of 30 m); New external parameters (e.g., soil type).	The 2 m temperature shows a clearly reduced negative bias over Switzerland by 0.3-0.5K (COSMO-News 26).
04.06.2008 (12 UTC)	New 2 m temperature and humidity diagnostics in COSMO-7.	More realistic diurnal cycle (COSMO-News 29).

Table 3.1: Temperature relevant COSMO (COSMO-7 and COSMO-LEPS) changes since 2006 and impacts.

n°	Code	Name	Alt. [m]	Lat. [°]	Long. [°]	LEPS	C7-07	C7-08
27	ABO	Adelboden	1320	46.5	7.5667	346.6	301.1	232.2
11	AIG	Aigle	381	46.3259	6.9208	514.9	365.1	388.2
33	ALT	Altdorf	449	46.8667	8.6333	651.1	741.1	636.2
14	BAS	Basel-Binningen	316	47.55	7.5833	-13.3	0.1	-27.8
17	BER	Bern-Zollikofen	553	46.9907	7.464	5.1	80.1	33.2
19	BUS	Buchs-Suhr	387	47.3833	8.0833	19.1	101.1	57.2
5	CDF	La Chaux-de-Fonds	1018	47.0833	6.8	18.3	64.1	59.2
7	CGI	Changins	430	46.4	6.2333	-5.3	66.1	-16.8
6	CHA	Chasseral	1599	47.1333	7.0667	-538.2	-487.9	-510.8
51	CHU	Chur-Ems	555	46.8667	9.5333	711.3	705.1	501.2
64	CIM	Cimetta	1672	46.2	8.8	-545.6	-440.9	-203.8
60	COM	Comprovasco	575	46.4667	8.9333	906.1	613.1	656.2
56	COV	Corvatsch	3315	46.4167	9.8167	-716.2	-554.9	-568.8
53	DAV	Davos	1590	46.8167	9.85	128.8	207.1	153.2
50	DIS	Disentis	1190	46.7	8.85	669.3	776.1	699.2
2	DOL	La Dole	1670	46.4333	6.1	-444.2	-416.9	-397.8
32	ENG	Engelberg	1035	46.8167	8.4167	337.6	505.1	236.2
45	EVO	Evolene-Villa	1825	46.1167	7.5167	18.3	65.1	-136.8
13	FAH	Fahy	596	47.4333	6.95	-4.7	3.1	-73.8
3	FRE	La Fretaz	1202	46.8333	6.5833	-148.5	-57.9	-70.8
35	GLA	Glarus	515	47.0333	9.0667	421.1	663.1	476.2
47	GRH	Grimsel-Hospiz	1980	46.5667	8.3333	273.6	268.1	224.2
44	GSB	Grand St.Bernard	2472	45.8667	7.1667	21.8	-66.9	-46.8
49	GUE	Guetsch	2287	46.65	8.6167	-10.4	-114.9	-0.8
25	GUT	Guettingen	440	47.6	9.2833	-22.9	-6.9	7.2
1	GVE	Geneve-Cointrin	420	46.25	6.1333	1.4	-70.9	-14.8
57	HIR	Hinterrhein	1611	46.5167	9.1833	152.5	703.1	498.2
26	HOE	Hoernli	1144	47.3667	8.95	-278.4	-247.9	-285.8
28	INT	Interlaken	580	46.6667	7.8667	430.1	528.1	451.2
48	JUN	Jungfraujoch	3580	46.55	7.9833	-830.4	-536.9	-469.8
22	KLO	Zuerich-Kloten	436	47.4833	8.5333	-10.1	17.1	33.2
20	LAE	Laegern	868	47.4833	8.4	-351.7	-380.9	-372.8
66	LUG	Lugano	273	46	8.9667	237	328.1	209.2
31	LUZ	Luzern	456	47.0333	8.3	88.8	180.1	74.2
63	MAG	Magadino	197	46.1667	8.8833	360.6	525.1	401.2
10	MLS	Moleson	1972	46.55	7.0167	-453.3	-652.9	-530.8
41	MVE	Montana	1508	46.3167	7.4833	-12.6	200.1	-25.8
29	NAP	Napf	1406	47	7.9333	-159	-397.9	-104.8
4	NEU	Neuchatel	485	47	6.95	0.5	34.1	-61.8
62	OTL	Locarno-Monti	366	46.1667	8.7833	169.2	315.1	232.2
9	PAY	Payerne	490	46.8167	6.95	18.3	42.1	1.2
30	PIL	Pilatus	2106	46.9833	8.25	-947.9	-1007.9	-876.8
65	PIO	Piotta	1007	46.5167	8.6833	617.9	835.1	717.2
12	PLF	Plaffeien-Oberschrot	1042	46.75	7.2667	80.3	-78.9	96.2

72	PSI	PSI Wuerenlingen	334	47.5333	8.2333	76.3	86.1	90.2
8	PUY	Pully	461	46.5167	6.6667	-33.4	-27.9	99.2
21	REH	Reckenholz	443	47.4333	8.5167	17.7	24.1	38.2
59	ROB	Robbia	1078	46.35	10.0667	367.7	1026.1	519.2
61	ROE	Robiei	1898	46.45	8.5167	-40.4	-52.9	32.2
15	RUE	Ruenenberg	610	47.4333	7.8833	-54.1	-11.9	-46.8
37	SAE	Saentis	2502	47.25	9.35	-1323.9	-1104.9	-1073.8
55	SAM	Samedan-St.Moritz	1705	46.5333	9.8833	478.7	529.1	562.2
58	SBE	San Bernardino	1639	46.4667	9.1833	124.5	497.1	195.2
67	SBO	Stabio	353	45.85	8.9333	19.9	46.1	-0.8
54	SCU	Scuol	1298	46.8	10.2833	735.7	735.1	727.2
16	SHA	Schaffhausen	437	47.6833	8.6167	11.9	13.1	-4.8
39	SIO	Sion	482	46.2167	7.3333	762.9	533.1	567.2
23	SMA	Zuerich SMA	556	47.3833	8.5667	-20.8	-22.9	-79.8
36	STG	St.Gallen	779	47.4333	9.4	-23.2	-65.9	-19.8
24	TAE	Taenikon	536	47.4833	8.9	-40.9	-6.9	66.2
42	ULR	Ulrichen	1345	46.5	8.3167	846	879.1	859.2
38	VAD	Vaduz	460	47.1333	9.5167	411.3	376.1	429.2
40	VIS	Visp	640	46.3	7.85	1048	782.1	866.2
34	WAE	Waedenswil	463	47.2167	8.6833	72.2	137.1	20.2
52	WFJ	Weissfluhjoch	2690	46.8333	9.8167	-450.5	-594.9	-332.8
18	WYN	Wynau	422	47.25	7.7833	77	69.1	76.2
43	ZER	Zermatt	1638	46.0333	7.75	995.6	1187.1	1088.2

Table 3.2: Table of the stations used in this study. The abbreviations, altitudes and locations are given. Under *LEPS diff.* the differences between COSMO-LEPS grid points height and stations height (grid point height minus station height) are given. The same is available for the COSMO-7 grid points prior to 19.02.2008 (*C7 diff. 07*) and afterward (*C7 diff. 08*). The differences are lower for COSMO-7 due to the higher spatial resolution. However, they are very strong correlated (R=99%). Generally, mountain stations exhibit larger height differences (R=86%).

3.3 Combination of deterministic and probabilistic forecasts

For a fictive forecast (Figure 3.2), an ensemble of 16 members with the same assumed probabilities ($\frac{1}{16}$) and one deterministic forecast are available. The NWP models forecast errors for 2 m temperature can be reasonably assumed normally distributed (chapter 6). Thus, the standard deviation (Equation 5.1) of the members (spread) is used to infer a normal PDF around the EM (black). The approach in this study is to use the spread information from

COSMO-LEPS to predict the uncertainty of the deterministic forecast COSMO-7-KAL, i.e. to predict its own error PDF. To determine the uncertainties of COSMO-7-KAL, a sample of forecast/observation pairs is needed (e.g., for a season). Determining the uncertainty of COSMO-7-KAL and the mean spread of COSMO-LEPS for each lead time, it is possible to investigate the spread-skill relationship (chapter 5) and to model it (chapter 6). The uncertainty of COSMO-7-KAL (skill) is measured using the RMSE (chapter 4). Using the identified uncertainty model, the instantaneous uncertainty of COSMO-7-KAL can be estimated using the spread of COSMO-LEPS as predictor. The use of the RMSE as predictand allows to consider possible bias in the COSMO-7-KAL forecasts for the uncertainty prediction. Seasonally, the COSMO-7-KAL forecasts are assumed bias-free.

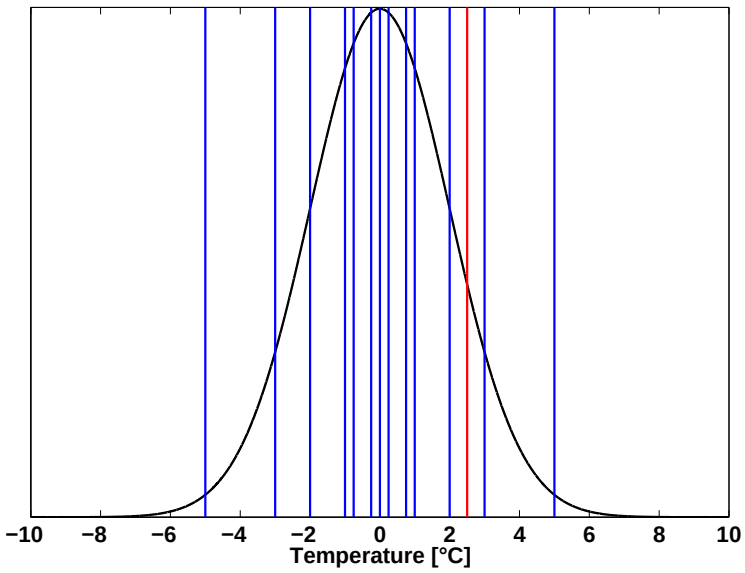

Figure 3.2: Fictive probabilistic and deterministic forecasts. The members of the ensemble are plotted in blue, the deterministic forecast is given in red. Based on the members, a normal PDF can be estimated around the EM. The deterministic run with more complex formulation has no uncertainty information. The goal of this study is to develop an uncertainty model capable of predicting the uncertainty of deterministic forecasts based on probabilistic information.

Chapter 4

Local 2 m temperature prediction in COSMO

In chapter 2 the limitations of global models in NWP were discussed, motivating the use of limited-area models for short-term weather forecasts in areas with complex orography. Due to higher spatial resolution and more complex formulation (e.g., non-hydrostatic), better forecasts can be expected from limited-area models. In this chapter, the 'local uncertainties' in COSMO-7 and COSMO-LEPS 2 m temperature predictions are discussed, that arise when comparing predictions at a model grid point with measurements at one specific location. To understand the uncertainties in local weather forecasts, the error structures must be understood. Typical errors in COSMO-7 forecasts will be presented in the second section of this chapter.

The association of a model grid point to a particular measurement site in COSMO-7 and COSMO-LEPS is facilitated by an automatic algorithm. If the nearest neighbour grid point has less than 100 m height difference to the station height, this grid point is chosen. Elsewise, the grid point with the smallest height difference in the vicinity of the stations (scanning 3×3 grid points or 2×2 grid points in 2007) is selected. The surface conditions at the grid point (e.g., land or lake) and its altitude often differ from which prevailing at the station. As the lowest atmospheric model level (COSMO-7 and COSMO-LEPS) is at 10 m, the 2m temperature is interpolated using the soil surface and the lowest model level temperatures (see first section of this chapter). Finally, the temperature is adapted to the altitude of the station considered. In doing so, a constant moist adiabatic gradient of $\frac{0.6°C}{100m}$ is assumed.

The question of the representativeness of grid point values for the observations at weather stations is an important point. Grid point values are averaged values representative for an entire grid box. On the contrary, SwissMetNet measurements are taken at specific site with standard surrounding conditions (see chapter 3). Therefore, even if the model was perfect, model forecasts at a neighbouring grid point could be quite different to local observations. Therefore, statistical post-processing is used to adapt NWP model outputs to the local conditions.

4.1 Diagnostics of 2 m temperature

Given that the lowest model layer in COSMO-7 is not located at 2 m, but at 10 m (since 19.02.2008, 30 m before), an interpolation technique is applied between the lowest atmospheric level and the ground (upper most soil model level) – the so-called 'surface layer' – to diagnose this variable. The same method is used for COSMO-LEPS (lowest model layer at 10m). The surface layer is crucial because of the interaction of the soil with the atmosphere and needs special treatments. The fluxes (essentially energy and water or water vapour) between the soil and the atmosphere must be parametrised and are modelled using a turbulent kinetic energy scheme. Surface exchange coefficients are computed using land cover and soil parameters defined for each grid cell:

- *fractional area covered by plants*;
- *leaf area index*: depends on the soil and vegetation type and shows large seasonal variations;
- *root depth*, important for plants water availability and evapotranspiration;
- *9 soil types*: each characterised by 19 parameters such as pore volume, field capacity, minimum infiltration rate, air dryness point, hydraulic diffusivity and conductivity, heat capacity and conductivity;
- *total roughness length*: or height where wind speed becomes zero, calculated on the basis of the local coverage and local orography (the local canopy term and orography term are added to get the total roughness length).

In COSMO-7 and COSMO-LEPS, the roughness height is determined by considering land use type and sub-grid scale topographical variances. The local orography part of the total roughness length is needed to encompass drag due to unresolved topography and gravity waves. A canopy height is estimated using the total roughness length. This leads generally to an overestimation of the real canopy height in mountain areas (large local orography term). Depending on the emplacement of the 2 m level (below or above canopy height), the diagnose algorithm is different (exponential or logarithmic profile, respectively). As the canopy height is overestimated in mountain regions, the exponential profile is used too often (even though the local canopy is very small: grass, glacier, etc.). Consequently, systematic errors have been identified: The 2m temperature is too closely coupled to the surface temperature, resulting in a systematically overestimated amplitude and a phase shift of the daily maximum temperature by 1-2 hours.

A new diagnostics was operationalised for COSMO-7 in May 2008 (*Buzzi*, 2008). It is based on the diagnostics from the DWD. It is tuned for typical SwissMetNet station conditions and only logarithmic profiles are used for the interpolation. The new method is significantly better for all seasons and particularly during stable conditions. The mean daily cycle is improved (shift and amplitude). On average, the negative bias in winter and the positive bias in summer are reduced. As the new diagnostics is only effective in COSMO-7, inconsistencies can be expected when combining COSMO-7 and COSMO-LEPS forecasts. However, solely the spread (and not the single temperature forecasts) of COSMO-LEPS will be used for uncertainty predictions in COSMO-7-KAL in this study. The bias of the ensemble mean has therefore no implications for the spread. When comparing the daily cycle of RMSE of COSMO-7 and COSMO-7-KAL with the daily cycle of the mean spread values of the summers 2007 (old diagnostics in COSMO-7 and COSMO-LEPS) and 2008 (new diagnostics in COSMO-7), no time shifts were observed in 2008 (not shown). Probably the spread is not sensitive to the diagnostics method. Note that on the 25.02.09 the new diagnostics was also adopted for COSMO-LEPS. Since then, COSMO-7 and COSMO-LEPS use again the same 2 m diagnostics.

4.2 COSMO performance in predicting 2 m temperature

In order to describe the performance of the COSMO temperature predictions quantitatively, several verification studies have been carried out. First, the annual verification results of the 2m forecasts of COSMO-7 and of the mean of COSMO-LEPS (COSMO-EM) for 2006-2008 are compared. Second, situations that cannot be modelled satisfactorily are discussed (focus on COSMO-7). The importance of post-processing methods to account for non-resolved and local processes is underlined. However, specific situations are highlighted where post-processing methods are not able to correct for model deficiencies.

Increased performance in 2008 compared to 2006-2007

The three-hourly forecast/observation pairs of the period 2006-2008 for all stations are used to determine the change in performance in 2008 compared to 2006-2007. As skill measure, the root mean square error (RMSE) is used:

$$\mathrm{RMSE} = \sqrt{\frac{1}{M}\sum_{m=1}^{M}(y_m - o_m)^2} \tag{4.1}$$

where y_m is the m^{th} forecast value and o_m the m^{th} observation. M is the amount of forecast/observation pairs (e.g., 366 days $\times$ 25 lead times in 2008). The RMSE is used to evaluate the average magnitude of the forecast errors (*accuracy*). The difference is squared in order to weight large deviations more strongly (*Wilks*, 1995) and thus, the RMSE is particularly useful when large errors are undesirable. The RMSE can vary between 0 and infinity. The RMSE is a negatively-oriented score, i.e., the smaller the RMSE value, the smaller are the forecast errors.

Figure 4.1 shows the relative difference of the RMSE in 2008 compared with the RMSE estimated for 2006-2007 for COSMO-7 and COSMO-EM for all stations investigated. The mean increase in skill for COSMO-7 is 18%. Only three stations exhibit a slight decrease. These results are confirmed by operational verification activities at MeteoSwiss. For COSMO-EM, the performance increased by 6% averaged over Switzerland. However, about a third of the stations (mostly low-level stations) showed less skill in 2008. Note that the EM shows a cold bias (on average for all stations and the year 2007: -0.94°C). One reason for this is a wet bias in the soil model in COSMO-LEPS (personal communication of A. Walser, MeteoSwiss).

For comparison, the overall mean bias in 2007 of COSMO-7 was 0.05°C. Assuming that the inherent predictability of the atmospheric flow in 2008 was comparable with 2006-2007, the improvements (COSMO-7 and COSMO-EM) can only be explained by model changes (see Table 3.1).

COSMO-7 behaves similar to the COSMO-LEPS members and it is known that the EM is on average better than any ensemble member (*Leith*, 1974). It explains why COSMO-7 was 2008 on average 2.5% less skilful than COSMO-EM. However, when Kalman filtering the COSMO-7 data, the best results are achieved (COSMO-7-KAL: -17% compared to COSMO-EM, Figure 4.2). In 2008, the COSMO-7-KAL forecasts were on average 18% better than the COSMO-7 DMO forecasts (Figure 4.3). Kalman filtering led to a RMSE reduction for all stations, especially for mountain stations illustrated by a strong correlation of the RMSE reduction with station height (R^2=53%). Note that R is the standard Pearson's correlation coefficient for linear relationships (*Wilks*, 1995). The values of R can vary between -1 and 1. 0 means no association, 1 perfect association with positive slope, -1 perfect correlation with negative slope. R^2 is Pearson's coefficient of determination equivalent to the explained variance of the predictand (e.g., RMSE reduction) using a certain predictor (e.g., station height). R^2 can vary between 0 (predictor does not explain the variance of the predictand at all) and 1 (100% explained variance).

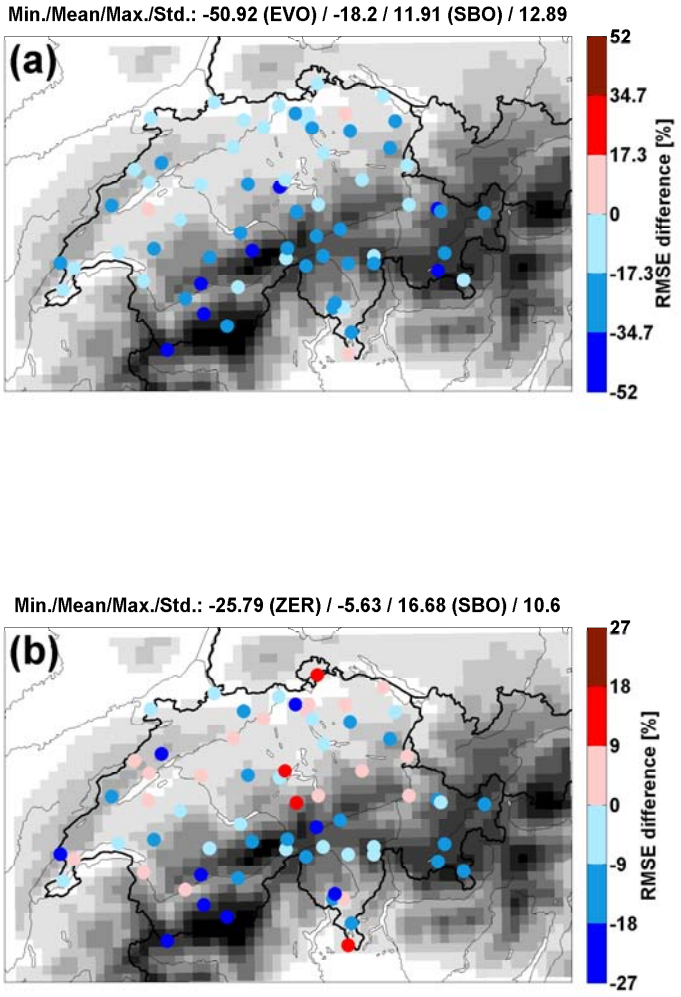

Figure 4.1: (a) COSMO-7 and (b) COSMO-EM performance in 2008 vs. 2006-2007. Mapped are the relative RMSE differences [%] between 2008 and 2006-2007. Negative values (blue) indicate skill improvement. Observe the different colorbar ranges. In this study, blue (red) always indicates improvement (degradation) of the score investigated.

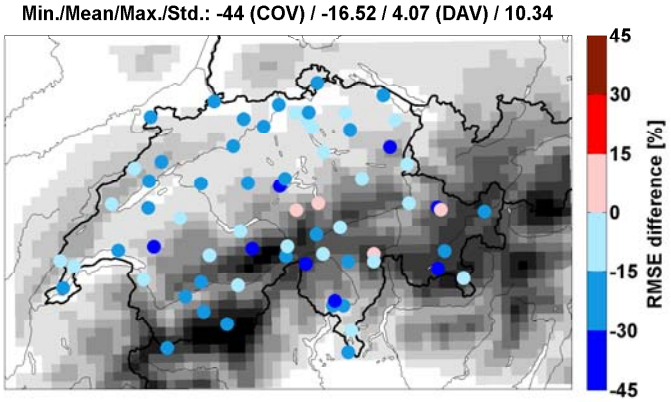

Figure 4.2: COSMO-7-KAL vs. COSMO-EM performance in 2008. COSMO-7-KAL delivers better forecasts principally due to post-processing.

For evaluating the quality of an EPS as COSMO-LEPS, the spread-skill relationship must also be considered. This will be investigated in the next chapter. Probabilistic verification methods taking into account the spread and the bias of EPS exist (e.g., *Murphy*, 1970; *Müller et al.*, 2005; *Gneiting et al.*, 2005; *Mason and Weigel*, 2009) and are needed to fully evaluate and compare EPS skills. However, they are not applied in this study. In contrast, this study focuses on the prediction of local uncertainties and how to combine the strengths of two models: COSMO-7-KAL delivering bias-free deterministic forecasts and COSMO-LEPS the uncertainty information.

In Figure 4.4 the RMSE for COSMO-7-KAL (period 2006-2007) are shown. This will be used in the verification (with independent 2008 data) as 'simplest' model for comparison. The RMSE is quite strongly correlated with the altitude of the attributed grid points (R^2=58%). This indicates that despite the strong ameliorations of the forecasts through Kalman filtering (up to -37% in RMSE values, Figure 4.3), particularly for mountain stations, the variability of the local errors remain dependent on the altitude. The performance of Kalman filtered forecasts is linked to the performance of the DMO forecasts. There are difficulties in modelling weather in complex terrain (see next section) and to expect the same performance in the Alpine region as in the Swiss Plateau, higher spatial resolution and better simulation of solar

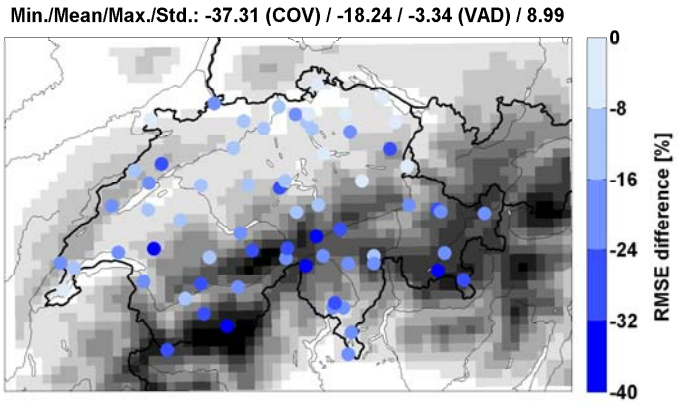

Figure 4.3: COSMO-7-KAL vs. DMO performance in 2008 (relative differences in %). Negative values indicate skill improvement.

radiation would be necessary. Precisely, this is done with the new high-resolution COSMO-2 model (chapter 2).

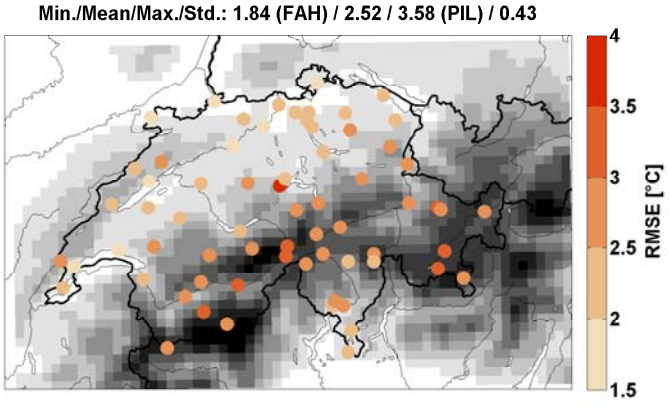

Figure 4.4: RMSE of COSMO-7-KAL in 2006-2007. All forecasts and lead times are considered. These data will be used in the verification of the developed uncertainty model (chapter 6).

Typical problems in 2 m temperature prediction

In winter, two processes are particularly difficult to simulate: the dynamics of temperature inversions and snow cover. Inversion situations are frequent during high pressure influence. Due to the low incoming solar radiation during the day and the strong outgoing long wave radiation from the surface during the night, cold air pools can develop. As examples, the stations 'La Brevine' [BRV] in the Jura (1048 m, situated in a small vale, Swiss cold temperature record of -41.8°C on the 12.01.1987) or stations in the fairly flat 'Engadine' valley (e.g., Scuol, 1304 m) record the coldest winter temperatures in Switzerland due to local characteristics. COSMO-7 and COSMO-LEPS are not able to predict such local cold air pools and Kalman filtering can help improving the forecasts (personal communication of F. Schubiger, MeteoSwiss).

Combined with low wind speed and high humidity, inversion situations are often characterised by fog formation (example in Figure 4.5). Fog acts as an absorber and backscatterer of solar radiation, leading to a radiation deficit at the ground during the day (low temperatures, no or weak daily cycle) and to a temperature increase within the inversion. Descending and adiabatically warmed air in a high pressure system (subsidence) amplifies the positive temperature anomaly above the inversion, strengthening the inversion. Such stable boundary

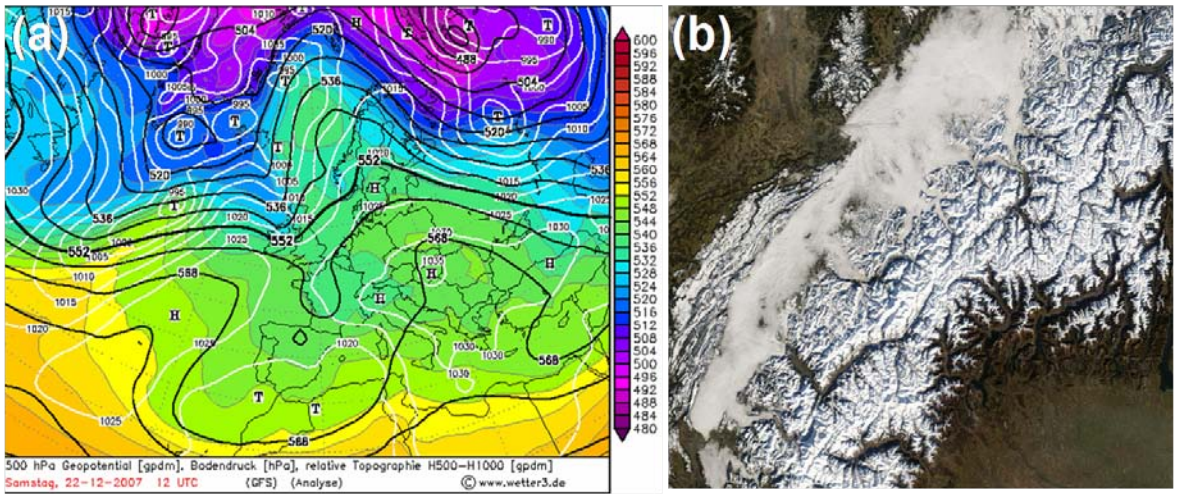

Figure 4.5: (a) Analysis from the global forecasting system (GFS, www.wetter3.de) on the 22.12.2007 12 UTC. Shown are the geopotential height at 500 hPa [gpdm, black lines], surface pressure [hPa, white lines] and thickness between 1000 and 500 hPa [gpdm, coloured surfaces]. (b) Visible satellite picture (1045 UTC, Terra/MODIS satellite: http://rapidfire.sci.gsfc.nasa.gov/realtime/2007086/) on the 20.12.2007. The stratus layer is visible over the Swiss Plateau. At the time the satellite picture was taken, a slight 'bise' (North-East wind) stream was established. Afterwards, a slight 'Foehn' (mild southerly wind) tendency prevailed.

layers are still difficult to reproduce in NWP models. In general COSMO tends to simulate too weak and too low inversions that are too rapidly dissolved (*Hess*, 2009). The parametrisation of turbulence (i.e., mixing) plays an important role to correctly predict the height and the strength of inversions. The capacity of NWP models to predict inversions with fog is crucial for the correct forecast of 2 m temperature (see examples in Figure 4.7). The amount of solar radiation at the ground affects the daily temperature cycle and its amplitude. In the case that there is an inversion and the model did not predict any inversion, the stations situated under the fog will register lower temperatures than predicted (positive forecast errors). The stations laying at the height of the inversion or slightly above would exhibit higher temperatures than predicted (negative forecast errors). Remember that the height difference between model grid point and station is considered using a constant moist adiabatic gradient (decreasing temperature with height). This is definitely a bad assumption in the particular case of inversion, given that temperature increases with height within the inversion.

To illustrate the problems of COSMO-7 forecasts during fog events, the case of the forecast on the 22.12.2007 is explained in more details. On the 22.12.2007 and during the following days, the region from the Azores up to Eastern Europe was characterised by high surface

pressure (>1015 hPa). The pressure distribution above central Europe was not very pronounced (see Figure 4.5, left). At the beginning, there was a slight Foehn tendency. Over the Swiss Plateau, the St. Galler Rhinevalley and up to the Urner Reussvalley a fog layer was present with top at 700-1000 m. The temperatures are cold and the amplitude of the daily cycle is reduced due to low irradiance (see Figure 4.7). Above the fog layer, mostly cloud free and mild temperatures are observed during the day. In the Tessin, there was high fog on the 23.12.2007 (*MeteoSwiss*, 2007). The errors are related to the height of the stations (R^2=40%, Figure 4.6). Below 1000 m positive model biases are generally observed and negative biases are observed above (Figure 4.6). This corresponds to the observation of fog top. Clearly, this case illustrates the difficulties of COSMO-7 to predict an inversion. The positive bias underneath 1000 m reveals the inexistence of the fog layer in the model, or at least a too fast dissolution of it. The negative bias above 1000 m is also an indication for a wrong prediction of the inversion.

The Kalman filtered data were even more biased (Swiss mean of 1.62°C, instead of 0.48°C with the DMO). The extension and location of the fog layer can be highly variable spatially (e.g., depending on wind strength, direction and topography) and temporally (change of

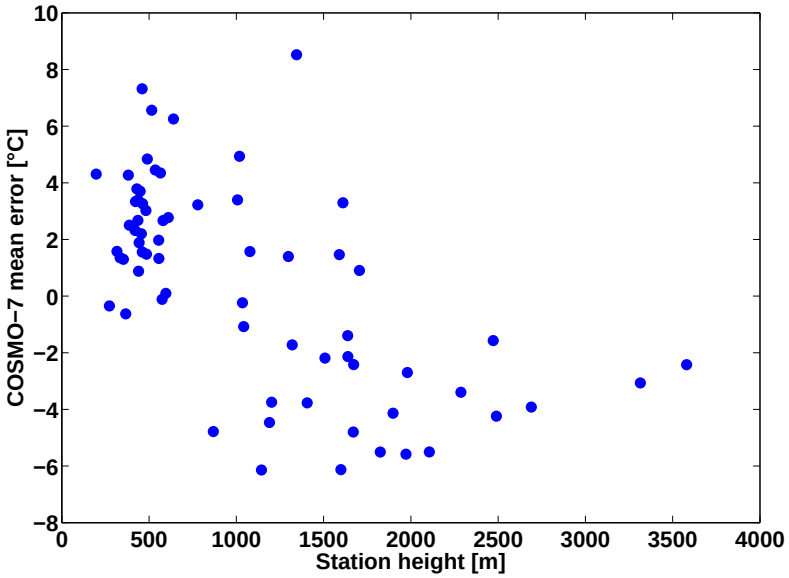

Figure 4.6: Fog event: dependence of the COSMO-7 error on the altitude. See explanations in the text.

the synoptic situation, advection of dryer air, etc.). This explains the bad performance of the Kalman filter during such situations as only the past error is used to correct the forecast. Figure 4.7 shows the meteograms for the 22.12.2007 forecasts for two stations: one underneath/in the fog (Zurich) and another above it (Hoernli). Typically, the model overestimates the amplitude of the daily cycles for stations situated below the fog layer (Figure 4.7). During the night, the large surface long-wave radiation and no or low backscattering lead to a strong surface cooling in the model. During the day, too much solar radiation and therefore too high temperatures were predicted. Above the fog layer overall too low temperatures were predicted.

Another challenge in winter, but also in spring and autumn, is the forecast of snow cover around the snow line (Figure 4.8). Due to the interpolation of the 2m temperature, the presence or absence of snow in the model world is decisive especially prior to the implementation of the new 2 m diagnostics in COSMO-7 in June 2008, when the 2m temperature was coupled too strongly to the soil temperature. The snow cover in COSMO-7 and COSMO-LEPS tends

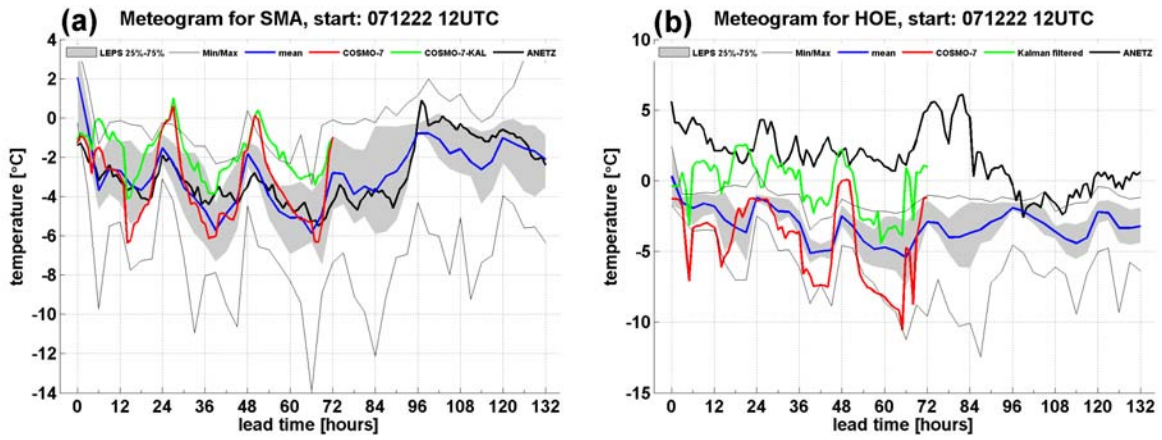

Figure 4.7: Meteograms for (a) Zurich [SMA, 556 m] and (b) Hoernli [HOE, 1144 m] during a fog event (22.12.2007 and following days). Plotted are: the interquantile range (grey patch), the min./max. of the members (thin black lines), COSMO-EM (blue), COSMO-7 (red), COSMO-7-KAL (green) and observations (bold black line). Pay attention to the different temperature scales. Zurich was probably underneath the fog layer and the station Hoernli above. Observe the different daily cycles of the model outputs and of the observations. In Zurich, the predicted night temperatures were too low: an evidence that no fog was predicted. For Hoernli, COSMO-7 predicted too low temperatures. Probably snow melting (remember, 2 m temperatures are too strongly linked to surface temperature until June 2008 in COSMO-7) exacerbated the effect of the wrong prediction of the inversion (max. of COSMO-7 forecasts around 0°C).

to melt too rapidly, leading to positive temperature biases. When snow is present, the surface temperature is 0°C or less. Without snow, the surface temperature varies depending on the radiation budget. Another factor is the difference between model grid point and station height. If the grid point is at higher elevation than the station itself (this is the case for over 60% of the stations investigated here) and if the model predicts snow for the grid point but the station is snowfree, the predicted temperature will be too low.

General properties of the 2 m temperature forecast error

The error is correlated with the magnitude of the temperature: in summer, the temperature exhibits a strong amplitude of the daily cycle due to the duration and strength of the solar radiation. Convective cloud cover and precipitation typical for a summer day are more difficult to predict than stratiform clouds and precipitation due to their local and rapid emergence. Given that cloud cover can indirectly affect temperature through diminishing irradiance and precipitation through evaporative cooling, temperature forecasts are more uncertain in summer. Also evapotranspiration plays a role: during the vegetative period green plants transpire cooling the environment (e.g., forests cooler than the surroundings). The po-

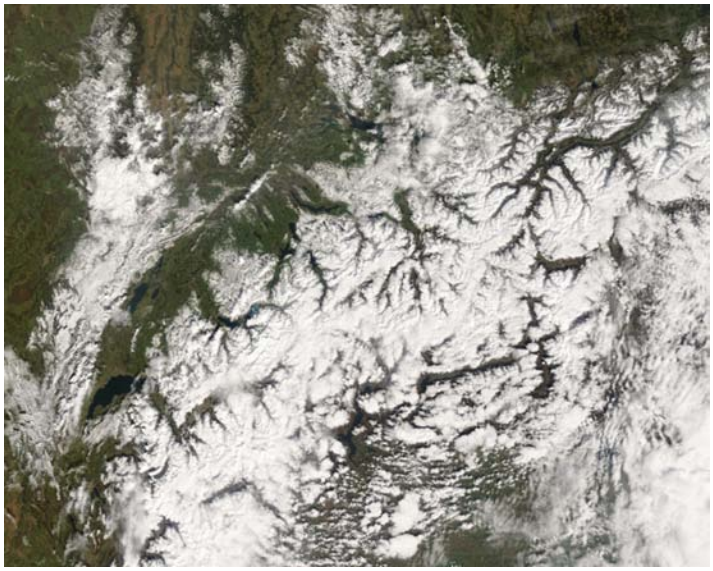

Figure 4.8: Visible satellite picture (1020 UTC, Terra/MODIS satellite: http://rapidfire.sci.gsfc.nasa.gov/realtime/2007361/) on the 27.12.2007. The snow is visible (Alps, Jura). The snow in the Swiss Plateau has already melt due to a rapid temperature increase. Snow tends to melt too rapidly in COSMO-7.

tential soil evaporation is also larger in summer (larger irradiance, higher temperatures). The actual evaporation depends on the soil moisture content. The importance of soil moisture-atmosphere interactions was investigated for the 2003 heat wave, for instance (*Fischer et al.*, 2007). Large precipitation deficits over Europe in the months prior to the heat wave led to decreasing soil moisture content limiting strongly the evaporation rates. The weak evaporative cooling allowed the temperatures to increase strongly during the heat wave. These supplementary, very local processes in summer are not fully resolved in NWP models (e.g., land cover distribution, soil moisture content) leading also to forecast errors.

Note that depending on the season, the sunrise and sunset times vary. In COSMO, the solar radiation is only actualised each hour and assumed constant in the meantime. Radiation is assumed vertical in COSMO and the slopes are not considered ('flat' terrain). Clouds over the surrounding grid cells hiding or reflecting sun radiation to the surface are not regarded. Mountains shadows (e.g., in Alpine valleys) are not taken into account. These simplifications in the model formulation lead to forecast errors of 2m temperatures that can only be partly corrected through the Kalman filter.

Summary

During winter, inversion situations are responsible for 2 m temperature forecast errors due to unsatisfactory simulations of inversions and differences between model world and reality (e.g., height difference between grid point and stations, bad constant lapse rate assumption). In autumn and spring, there are problems around the snow line. Particularly in summer, the boundary layer is well-mixed (e.g., through convection) and the constant lapse rate approximation is more realistic, but other processes linked with soil, vegetation and convection complicate 2 m temperature forecasts. Post-processing methods are helpful to adapt forecasts to the current weather situation and to local characteristics. The Kalman filter particularly improves predictions at mountain stations. However, the Kalman filter is not able to account for all forecast errors. On the one hand, COSMO-7 is not able to simulate all weather processes satisfactorily (e.g., fog, snow melt). If the differences between model world and reality change abruptly (e.g., too rapid snow melt or too rapid fog dissolution in the model compared to reality) during the 3-days forecast, the Kalman filter does not perform well. On the other hand, 'random' errors remain due to uncertainties arising from larger scales (meso-scale, syn-

optic and also large-scale). These uncertainties are called *'atmospheric uncertainties'* in this study. The strength and frequency of 'random' errors determine the uncertainty in COSMO-7-KAL forecasts. In the following chapter, the information content of the spread of the limited-area ensemble COSMO-LEPS will be investigated along with its ability to represent the uncertainties of local deterministic COSMO-7-KAL forecasts.

Chapter 5

The properties of the spread-skill relationship

5.1 Definitions and introduction

EPS are designed to represent the full uncertainty range as realistic as possible (*Ban*, 2006; *Palmer and Hagedorn*, 2006). To assess if this is the case, the spread-skill relationship (SSR) is often investigated. *Palmer et al.* [2005] showed that in a 'perfect ensemble' the mean (e.g., over a season, a year) of the spread (Equation 5.1) should be equal to the root mean square error (RMSE, see Equation 4.1 in chapter 4) of the EM over the same period. In this study, the same definitions of spread and skill are used. As the ensemble used in this study (COSMO-LEPS) only includes 16 members, the unbiased estimator of the standard deviation is used:

$$\text{spread} = \sqrt{\frac{1}{M-1}\sum_{m=1}^{M}(f_m - \overline{f})^2} \tag{5.1}$$

where M is the ensemble size, f_m is the forecast value of the m^{th} member and $\overline{f}$ represents COSMO-EM:

$$\overline{f} = \frac{1}{M}\sum_{m=1}^{M} f_m \tag{5.2}$$

Figure 5.1 shows the mean evolution in 2008 of the spread and the skill for two stations. Although Zurich (station situated on a south-west exposed slope above the lake of Zurich,

556 m) and Kloten (station situated roughly 11km northward from Zurich at a similar altitude) are geographically very close, there are obvious differences in the daily cycle (shape and amplitude) of the RMSE of COSMO-7-KAL. The RMSE of the COSMO-EM is more comparable. This is due to the fact that the spatial resolution of COSMO-LEPS is lower than that of COSMO-7 and that the COSMO-7 data are Kalman filtered, i.e. adapted to the local conditions. Also the spread is geographically less variable. In the following section, the SSR of COSMO-LEPS is investigated. In a second step, the relationship between the skill of COSMO-7-KAL and the spread of COSMO-LEPS will be characterised. Finally, both SSRs are compared quantitatively.

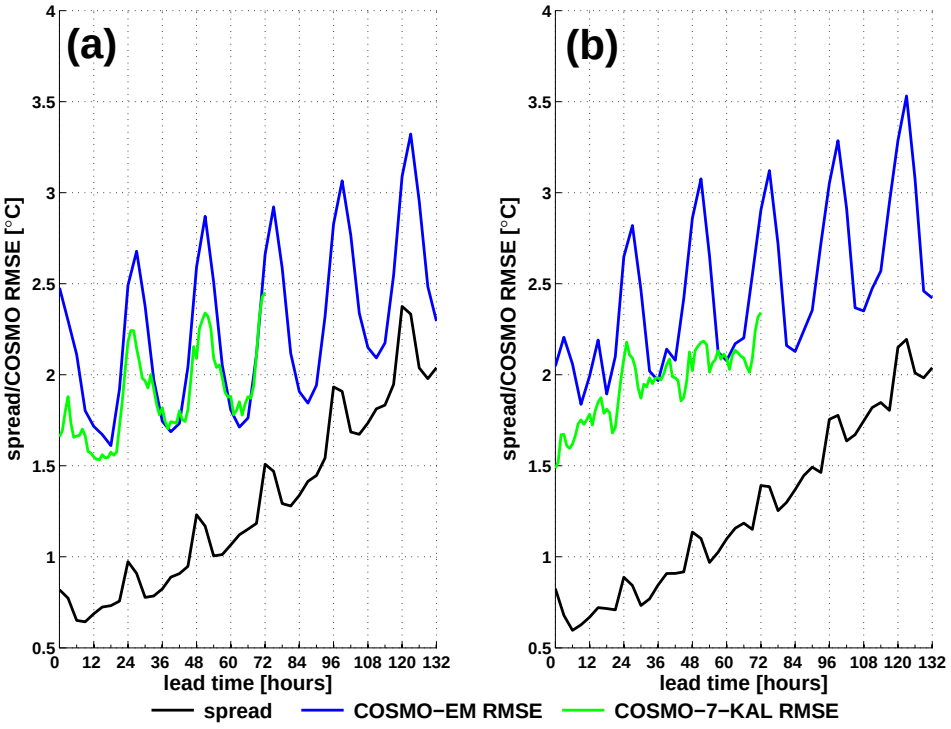

Figure 5.1: Mean spread (black) and skill (COSMO-EM [blue] and COSMO-7-KAL [green]) in 2008 for two close stations: (a) Zurich [SMA] and (b) Kloten [KLO]. The spread is too small for both stations and all lead times (*underdispersion*). The underdispersion decreases with the lead time.

5.2 The spread-skill relationship of COSMO-LEPS

To investigate the nature of the SSR for COSMO-LEPS, scatter plots are analysed. The mean spread is compared with the skill of COSMO-EM. The data of Figure5.1 for Zurich is shown as scatter plot in Figure 5.2. The plotted diagonal is the 'perfect' SSR line (spread = RMSE). All points are above this line (spread < RMSE), i.e. the spread is underdispersive for all lead times. Depending on the time of the day (12, 15, 18, 21, 00, 03, 06, 09 UTC), the points are coloured differently. The eight categories show similar slopes, but different intercepts (*'stratification of the SSR'*): i.e., the underdispersion varies during the day. When considering other representative stations, the ensemble COSMO-LEPS is mostly underdispersive (Figure 5.3). The spread is often strongly too small in the early afternoon (12 and 15 UTC), however this can not be generalised (e.g., largest underdispersion at 12 UTC for Davos, 18 UTC for Lugano). On average, the spread is too small compared to skill for all stations (Figure 5.4) indicating that it does not explain the full local variability of errors in the COSMO-7-KAL forecasts. The underdispersion tends to be larger for mountain stations (Alps, Jura). The structure of the SSR varies seasonally (example in Figure 5.5). For instance the times of the day with largest or smallest underdispersion vary depending on the season. Overall, the SSR can be assumed linear for all seasons and all stations, whereas single stations show more complex relationships for one or the other season. On average, the spread is underdispersive for all stations, i.e., it does not explain the full local variability of errors of the COSMO-7-KAL forecasts. The underdispersion tends to be larger for mountain stations (Alps, Jura).

5.3 Relationship between spread and skill of COSMO-7-KAL

The last section has shown the SSR of the COSMO-LEPS system. In this section, the novelty of this study is introduced: The SSR between the spread of the ensemble COSMO-LEPS and the skill of deterministic bias free COSMO-7 forecasts (COSMO-7-KAL) is investigated. Only 25 lead times are available (instead of 45 lead times with COSMO-LEPS) because the COSMO-7-KAL forecasts reach only up to day+3.

The similarities and differences to the SSR of COSMO-LEPS are briefly discussed, whereas similarities prevail: as for COSMO-LEPS, the SSR can be assumed linear for most stations (Figures 5.6 and 5.7) and seasons (Figure 5.9). The SSR of COSMO-7-KAL also varies

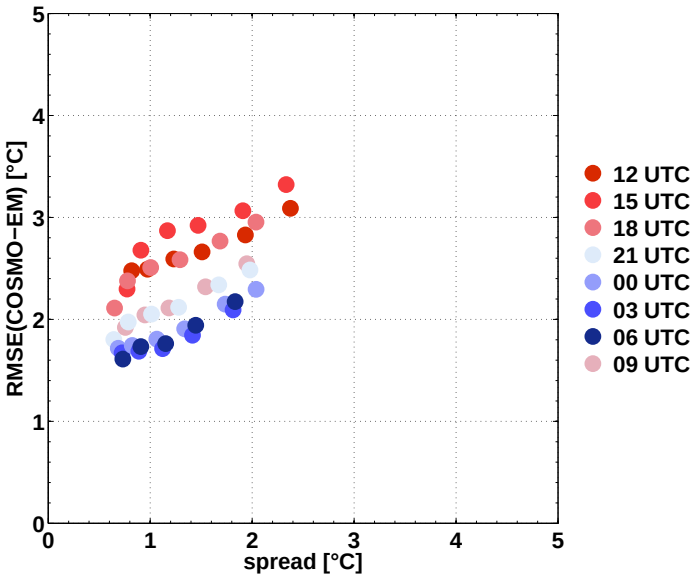

Figure 5.2: Annual COSMO-LEPS SSR for the station Zurich in 2008.

geographically (Figure 5.7) and seasonally (Figure 5.9). On average, the spread is also too small compared to the RMSE, however the mean underdispersion is reduced by 17% (Figure 5.8). The stratification of the SSR is less pronounced. Note that single stations and seasons show more complex relationships. In the next section, the SSRs using the COSMO-EM forecasts and the COSMO-7-KAL forecasts are compared quantitatively for all stations.

5.4 Seasonal comparison and discussion

In this section, the SSRs using COSMO-EM and the COSMO-7-KAL forecasts are quantitatively compared with each other. The underdispersion and the correlation of the daily cycle are investigated using two scores: one considering the overall correspondence between the spread and the RMSE, the second the correlation without considering underdispersion.

Strength of the spread-skill relationships

The differences between the RMSE (COSMO-EM and COSMO-7-KAL) and the spread values for all 3-hourly lead times until day+3 12 UTC are used to compute the root mean square difference (RMSD, Equation 5.3). This score measures the correspondence between spread

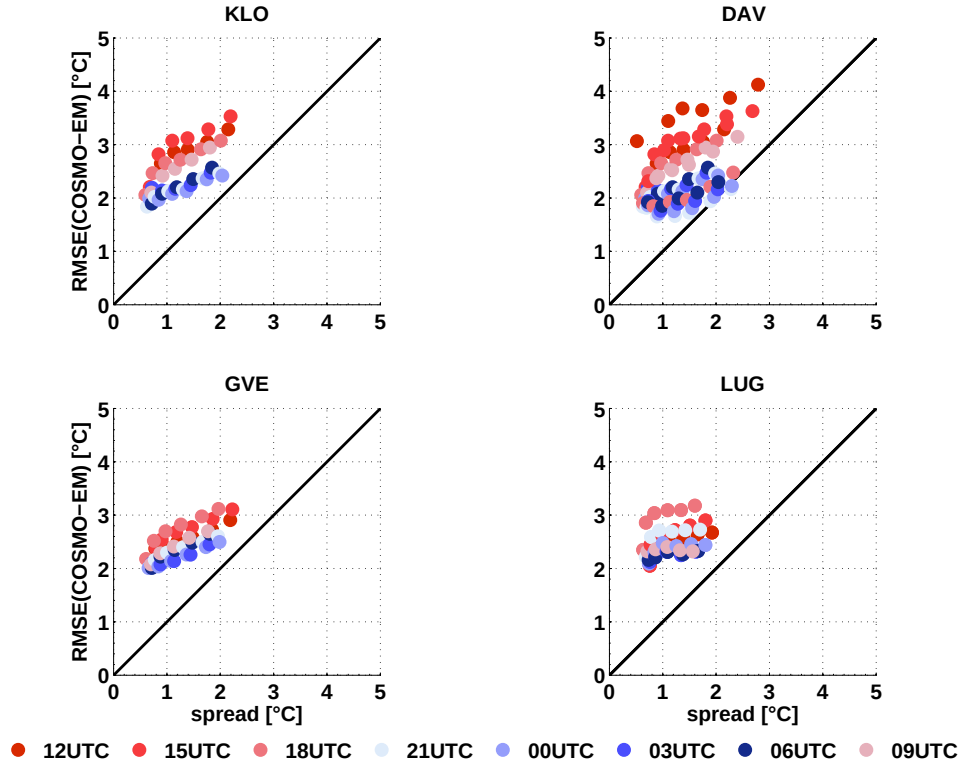

Figure 5.3: Annual COSMO-LEPS RMSE vs. spread scatterplots for four different stations situated in various topographical and meteorological environments (Kloten [KLO]: eastern Swiss Plateau; Davos [DAV]: located in a flat valley in the eastern Alps; Geneva [GEV]: western Swiss Plateau, near lake Geneva; Lugano [LUG]: Ticino (Southern part of the Alps), near lake Lugano). The diagonal is the 'perfect' SSR line. The colours represent the 8 different forecast times (12, 15, 18, 21, 00, 03, 06 and 09 UTC). Daytime lead times are plotted in reddish colours and the night-time ones in bluish colours.

and skill and is negatively-oriented: The smaller the value, the better the SSR. The RMSD can vary between infinity and 0.

$$\mathrm{RMSD} = \sqrt{\frac{1}{M}\sum_{m=1}^{M}(\mathrm{spread(m)} - \mathrm{RMSE(m)})^2} \tag{5.3}$$

The summation is applied over the M lead times. The RMSE values are for COSMO-EM or COSMO-7-KAL depending on the SSR considered.

Both the phases of the spread and skill values (e.g., shift of the daily cycle) and the underdispersion (mean distance between spread and skill values) impact the RMSD. The

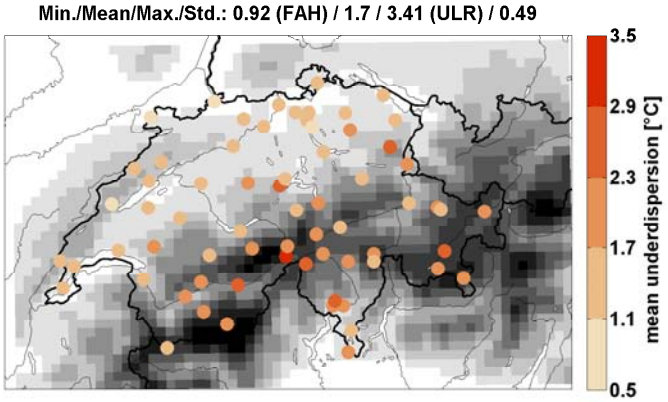

Figure 5.4: Mean underdispersion values [°C] for the SSR of COSMO-LEPS in 2008. Plotted are the differences mean skill minus mean spread (whole year, all lead times up to day+3 12 UTC). Larger values indicate larger underdispersion.

RMSDs for COSMO-LEPS in 2008 are plotted for all stations and the four seasons in Figure 5.10. The pattern is particularly pronounced in winter (2007/2008): There is a relationship between RMSD and grid point height (R^2=46%). Some middle to high altitude stations exhibit particularly large RMSDs, e.g., Cimetta (1672 m), Saentis (2490 m), Pilatus (2106 m) and Ulrichen (1345 m). The grid points attributed to these stations are too low (maximum of 1324 m height difference for the station Saentis in COSMO-LEPS, see Table 3.2 in the Appendix). This indicates that the differences between model world and reality are important for explaining the strength of the spread-skill relationship.

The RMSD values are also estimated for the SSR of COSMO-7-KAL. Instead of the absolute values, the relative differences in RMSD (RMSDD, Equation 5.4) when using COSMO-7-KAL instead of the COSMO-EM are calculated.

$$\mathrm{RMSDD}[\%] = \frac{\mathrm{RMSD}_{\text{COSMO-7-KAL}} - \mathrm{RMSD}_{\text{COSMO-EM}}}{\mathrm{RMSD}_{\text{COSMO-EM}}} \cdot 100 \qquad (5.4)$$

The RMSDs are on average 10.4% smaller for COSMO-7-KAL, i.e. the SSR is better when using COSMO-7-KAL (Figure 5.11). Only in spring, the RMSDs for COSMO-7-KAL are larger for most stations (especially mountain stations). Therefore, the correspondence between the spread and the RMSE of COSMO-7-KAL was lower than the correspondence between the

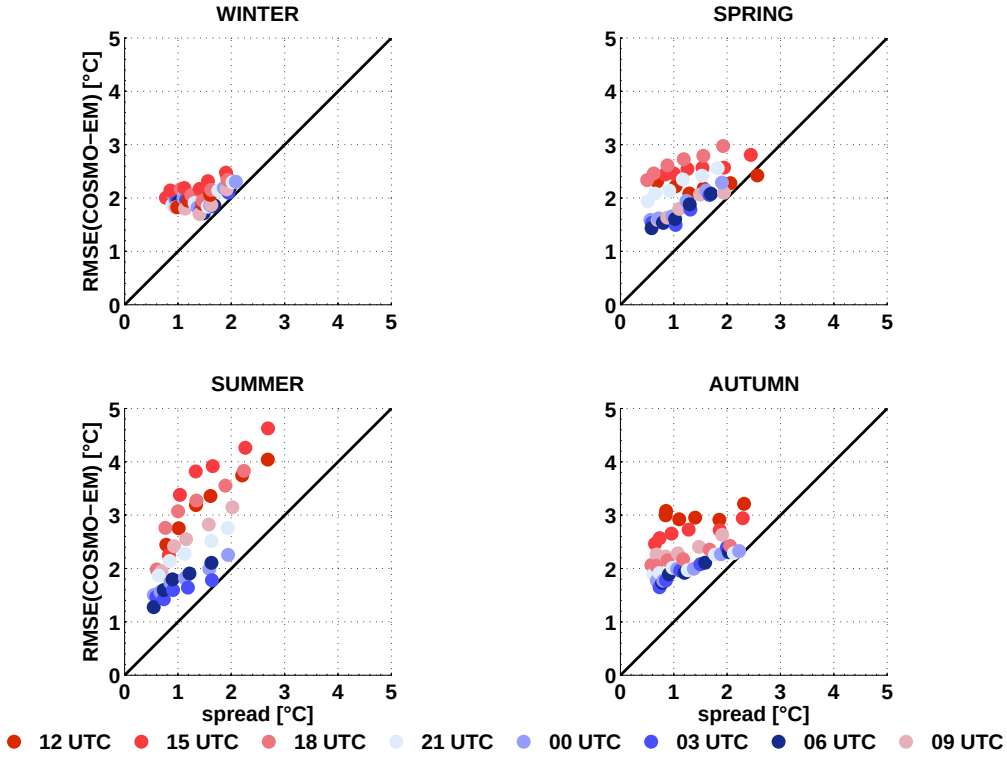

Figure 5.5: Similar to Figure 5.3 but showing seasonal scatter plots for one station (Zurich).

spread and the RMSE of the COSMO-EM. This indicates that the RMSEs of COSMO-7-KAL were larger than the RMSEs of COSMO-EM: this specific season (spring) was particularly difficult to forecast with a deterministic model. For instance, snow and fog dynamics were not satisfactorily modelled with COSMO-7 and the Kalman filter was not able to account for the resulting forecast errors (chapter 4).

Correlations between spread and skill

Important in the perspective of modelling the SSR is the correlation (or similarity) between the daily cycles of the spread and of the skill. Therefore, Pearson's correlation coefficient is used. The R values for the SSR of COSMO-LEPS are plotted in Figure 5.12. Most of the stations exhibit a positive correlation and are consistent with the expectation of increasing spread with increasing RMSE (on average, R=43.7%). Only sparse stations show negative correlations, representing a negative SSR (decreasing spread with increasing RMSE). The

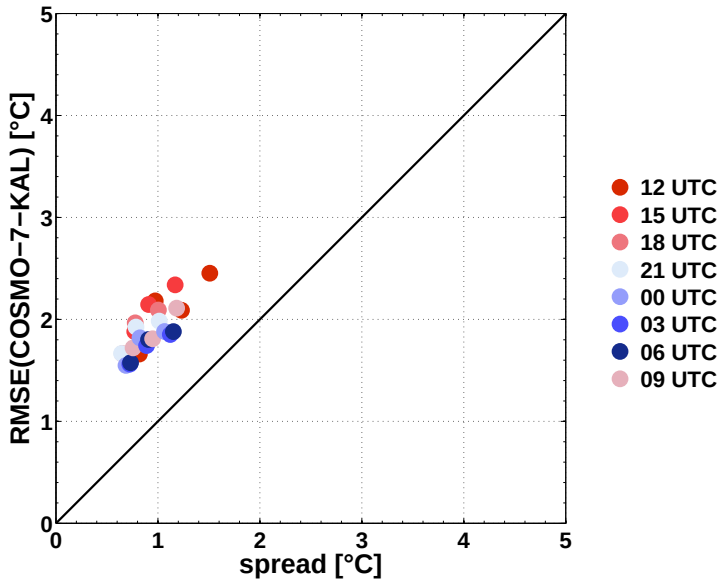

Figure 5.6: Annual COSMO-7-KAL SSR for Zurich in 2008, compare with Figure 5.2.

spread better explains the variance of the RMSE of COSMO-7-KAL than the variance of the RMSE of COSMO-EM (+10.2% in R^2 on average). The seasonal differences in R^2 (see Equation 5.5) are mapped in Figure 5.13.

$$R^2_{diff} = (R^2_{\text{COSMO-7-KAL}} - R^2_{\text{COSMO-EM}}) \tag{5.5}$$

Generally, the SSR of COSMO-7-KAL is better than the SSR of COSMO-LEPS for short-term forecasts. The underdispersion is reduced and the daily cycle of the spread corresponds better to the RMSE cycle of COSMO-7-KAL. It is not self-evident to expect from the spread of the COSMO-LEPS system to correspond better to the local variability of errors in deterministic and Kalman filtered forecasts than to the uncertainty of its own ensemble mean. The Kalman filter removes the systematic local errors and simultaneously reduces the local uncertainties. On the whole, 'random' errors depending on larger scales ('atmospheric' uncertainties) remain. The good correspondence between the spread and the variability of errors in COSMO-7-KAL indicates that the spread essentially comprises 'atmospheric uncertainties'. The COSMO-EM is not bias corrected, local errors remain that the spread cannot represent, explaining the weaker spread-skill relationship. This important conclusion justifies

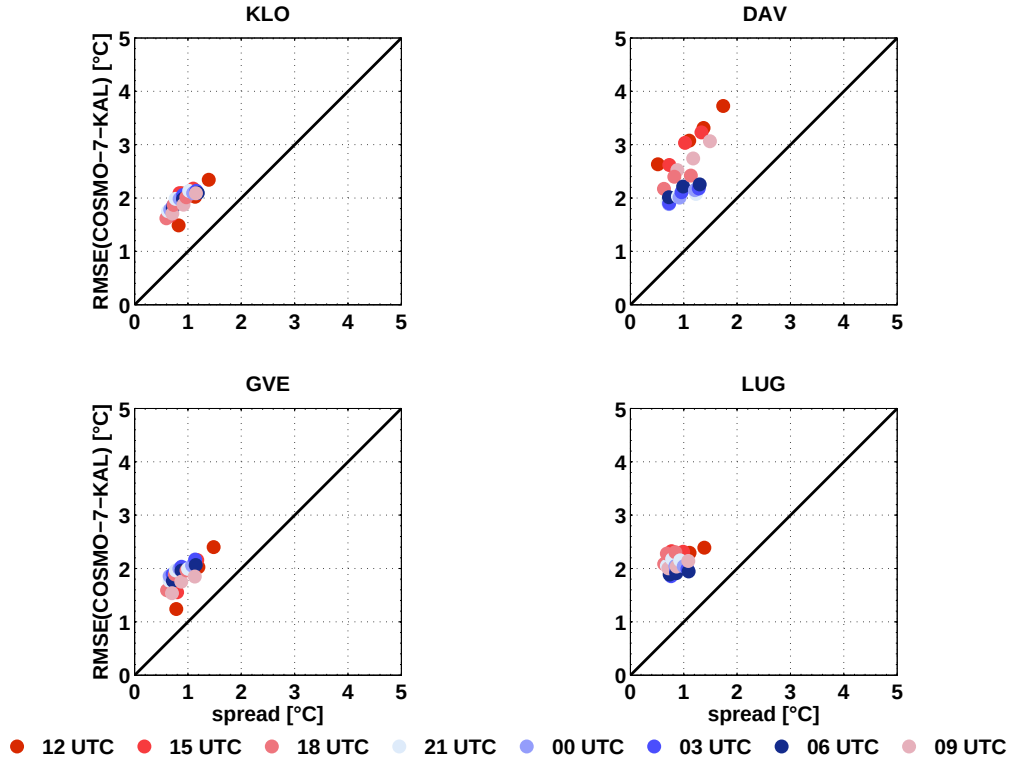

Figure 5.7: Annual COSMO-7-KAL RMSE versus spread scatter plots for four different stations situated in various topographical and meteorological environments (Figure 5.3).

the combination of COSMO-LEPS and local bias corrected COSMO-7 forecasts. The results of this chapter are encouraging in the perspective of modelling the uncertainties of the local COSMO-7-KAL predictions (next chapter).

In summary, the SSRs of COSMO-7-KAL and COSMO-LEPS depend on:

- the geographical location of the station;
- the season;
- the time of the day, whereas the dependence is much weaker for COSMO-7-KAL.

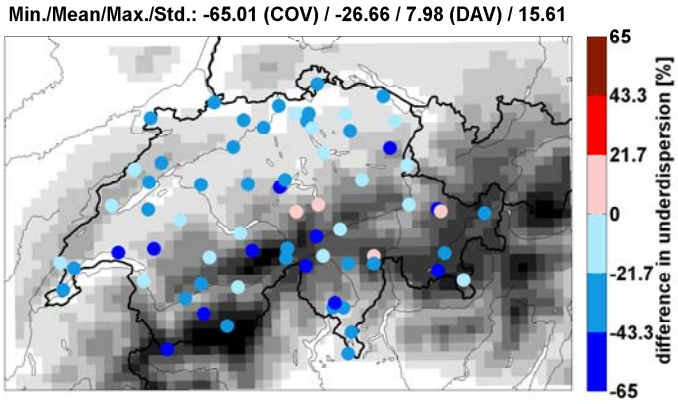

Figure 5.8: Relative change in underdispersion when using COSMO-7-KAL forecasts instead of the COSMO-EM (compared to Figure 5.4) in 2008. Only four stations show larger mean underdispersion.

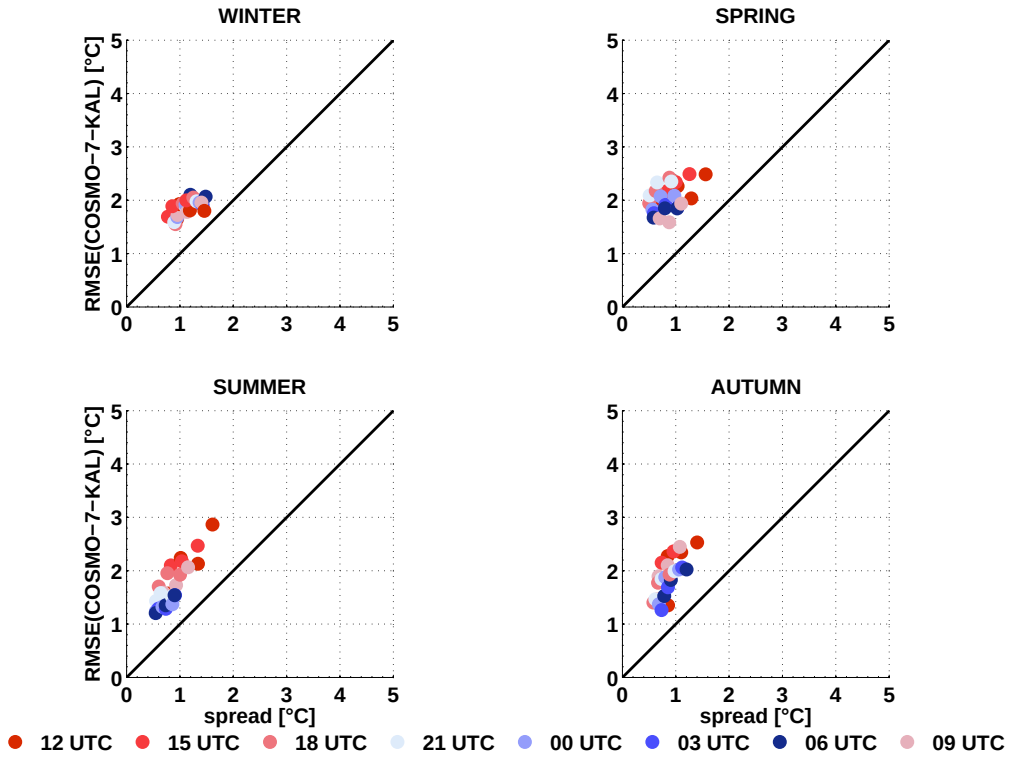

Figure 5.9: Similar to Figure 5.7 but showing the seasonal variation of the SSR for Zurich in 2008. Compare with Figure 5.5.

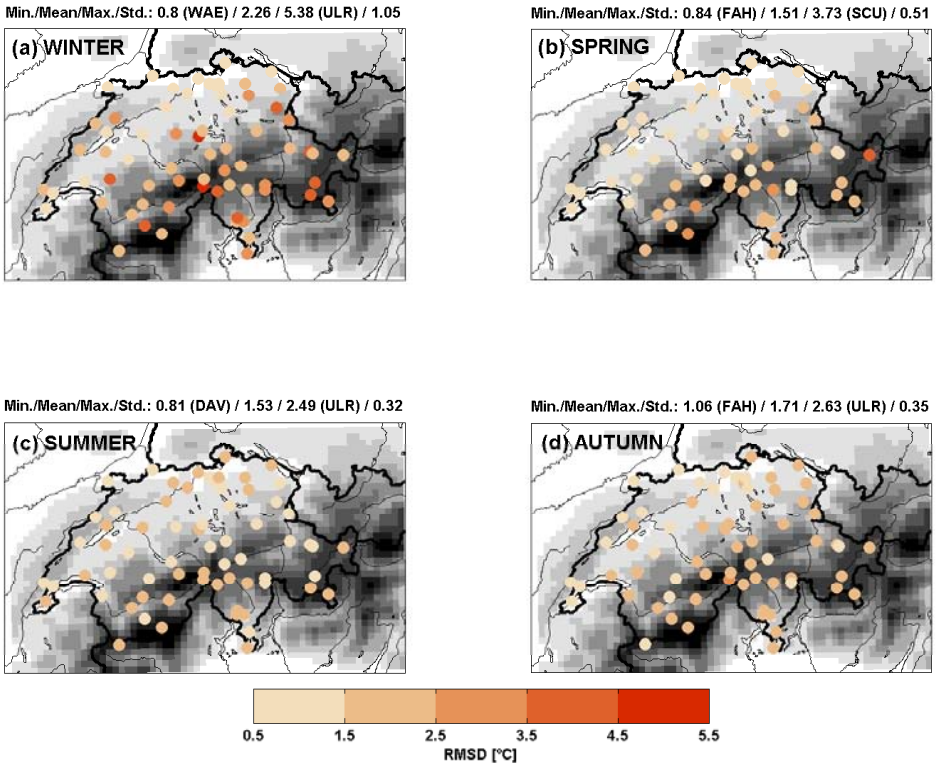

Figure 5.10: Seasonal RMSDs in 2008.

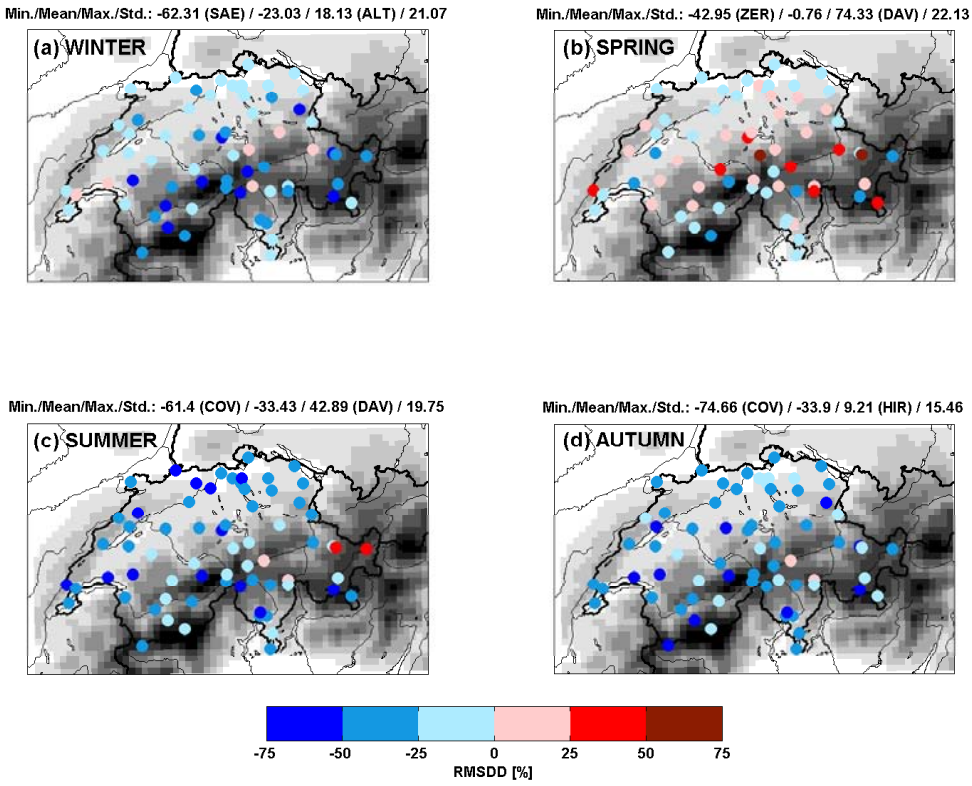

Figure 5.11: Seasonal RMSDDs in 2008. Positive values indicate smaller RMSD for the SSR of COSMO-7-KAL.

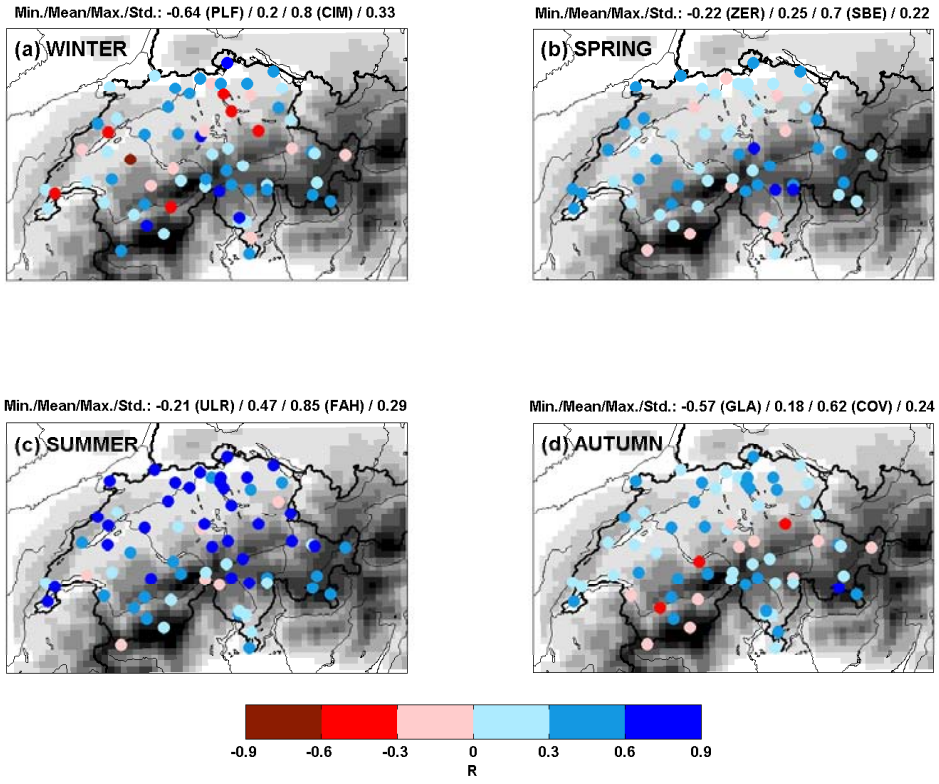

Figure 5.12: Seasonal R values in 2008 for the SSR of COSMO-LEPS.

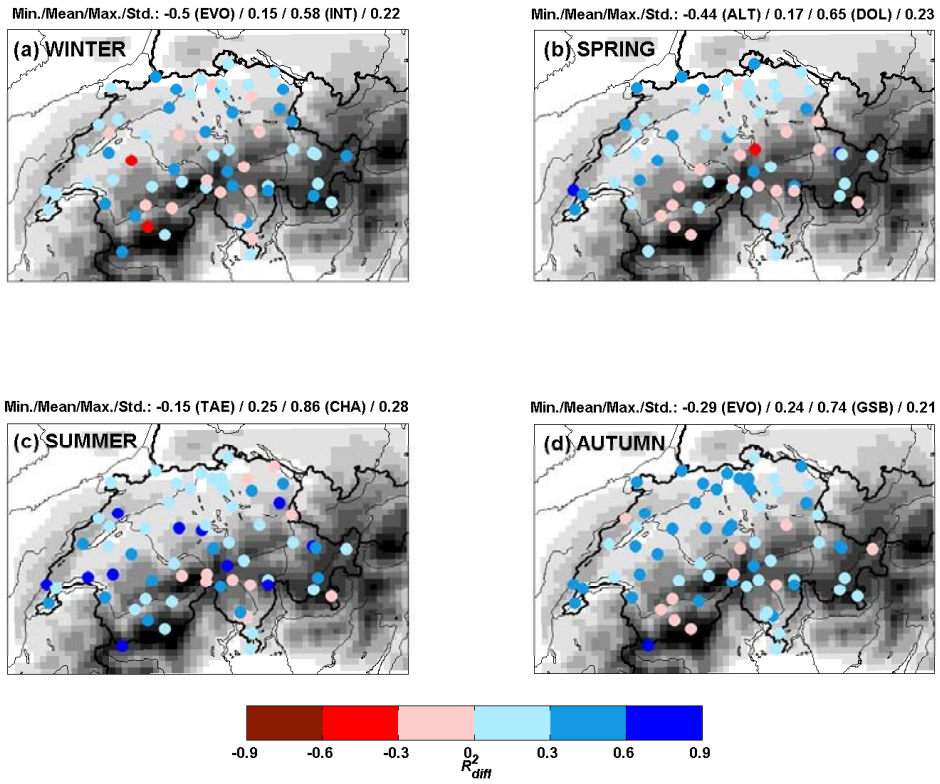

Figure 5.13: Seasonal R^2_{diff} values in 2008. Positive values correspond to the additional explained variance for the SSR of COSMO-7-KAL compared to the SSR of COSMO-LEPS.

Chapter 6

Uncertainty prediction models for COSMO-7-KAL

The investigation of the spread-skill relationships (SSR) based on the spread of COSMO-LEPS and COSMO-7-KAL forecasts shows that the use of spread information has potential to predict uncertainty in bias corrected local deterministic forecasts. The SSR can be assumed as linear. It depends on the geographical location of the station and on the season. The SSR of COSMO-7-KAL also depends on the time of the day but this dependence is less pronounced than for the SSR of COSMO-LEPS. First, the dependency on the time of day of the SSR will not be taken into account in the uncertainty model. Note that the small data basis (25 lead times for 8 times of the day) does not allow the identification of uncertainty models for each time of the day.

The geographical dependence of the SSR implies the use of *station-based* uncertainty models, whereas the seasonal variation of the SSR implies the use of *adaptive* uncertainty models. This means that model parameters are regularly (e.g., daily) adapted, or re-estimated and this for each station. Adaptive uncertainty models have the advantage of adapting rapidly in the case of NWP model changes. They are quite often in COSMO (Table 3.1 in chapter 3). For instance, changes in the consideration of inherent uncertainties in the formulation of COSMO-LEPS impact the spread (e.g., reduced underdispersion). Changes in COSMO-7 (e.g., new snow analysis, new 2 m temperature diagnostics) influence the strength and variability of errors (skill). Given that forecasts of two different NWP models are combined, both changes in the probabilistic as well as in the deterministic model can affect the nature of the

SSR. In this chapter, adaptive station-based uncertainty models with increasing complexity are identified and compared.

6.1 The adaptive linear model (ALM)

Formulation of the statistical model

The first model is the *adaptive linear model (ALM)* (Equation 6.4, Figure 6.3). Daily, the evolution of the RMSE of COSMO-7-KAL ($\overline{\text{RMSE}}_w(t)$) and of the mean spread ($\overline{\text{spread}}(t)$) in function of the lead times (t) are calculated using model outputs and observations of the last 30 days (ALM30).

To begin with, a training period of 30 days is chosen as compromise between the wish to account for the latest structure of the spread-skill relationship and the need of a minimal data sample for robust statistics. Other training periods are tested later. The ALM has two parameters: the intercept a and the slope b. They are daily estimated using the standard least square regression method (*Wilks*, 1995; *Draper and Smith*, 1998).

Outliers filtering

The 'extreme outliers' are removed to get a more robust regression. Error values below q_- (smallest negative errors) and above q_+ (largest positive errors) are not considered for the estimation of the parameters (Equations 6.1 and 6.3). This allows a more robust estimation of the model parameters. Moreover, extreme errors which can occur independently of the NWP model are filtered out. For instance, a wrong snow analysis used for assimilation and leading to biased forecasts is not representative for the performance of the model itself. For each station, an error distribution is built based on all forecast/observation pairs of the period 2006-2007 (example in Figure 6.1). The data of 2008 are not used given that the verification of the uncertainty models will be done for 2008. This ensures the independence of the data used for verification from the data used for the parametrisation of the uncertainty models. The error distributions of all Swiss stations can be assumed as Gaussian. Especially mountain stations show slightly right skewed distributions (not shown). Thus, the predicted uncertainties with the ALM can be used to derive normal distributions around the COSMO-7-KAL forecasts (chapter 3).

Figure 6.2 shows the repartition of the outliers depending on the lead times for the year 2008. The outliers from all stations are joined into the histogram. On the whole, outliers tend to be more frequent during the afternoon (12 and 15 UTC) and with increasing lead time. This is not astonishing as the forecast uncertainties are generally larger in the afternoon (chapter 5). There is no apparent link between station height and outlier frequency (not shown). The total frequency of outliers in 2008 varies between 0 and 36 depending on the station.

$$IQR = Q_{0.75} - Q_{0.25} \tag{6.1}$$

$Q_{0.25}$ and $Q_{0.75}$ are the 25 and 75% quantiles. IQR is the *interquantile range*.

$$q_- = Q_{0.25} - 3 \cdot IQR \tag{6.2}$$

$$q_+ = Q_{0.75} + 3 \cdot IQR \tag{6.3}$$

$$\widehat{\mathrm{RMSE}}_w(t) = a_w + b_w \cdot \overline{\mathrm{spread}}_w(t) \tag{6.4}$$

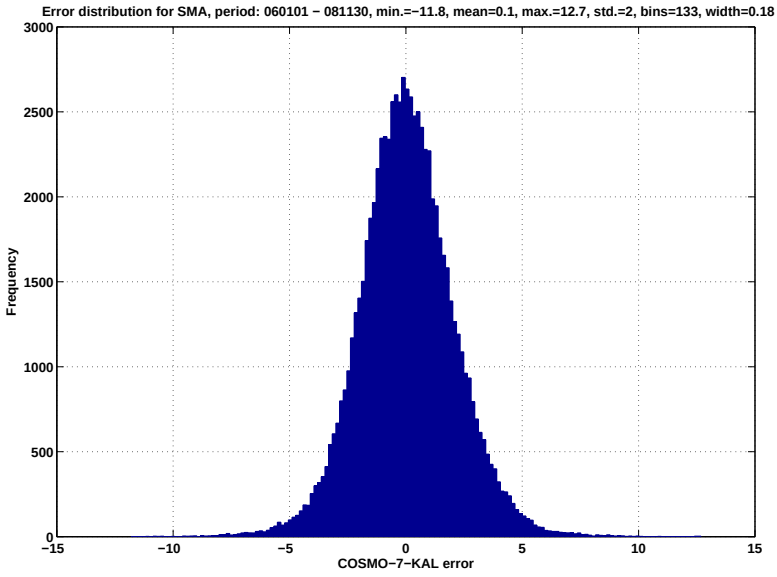

Figure 6.1: Error distribution for Zurich in 2006-2008.

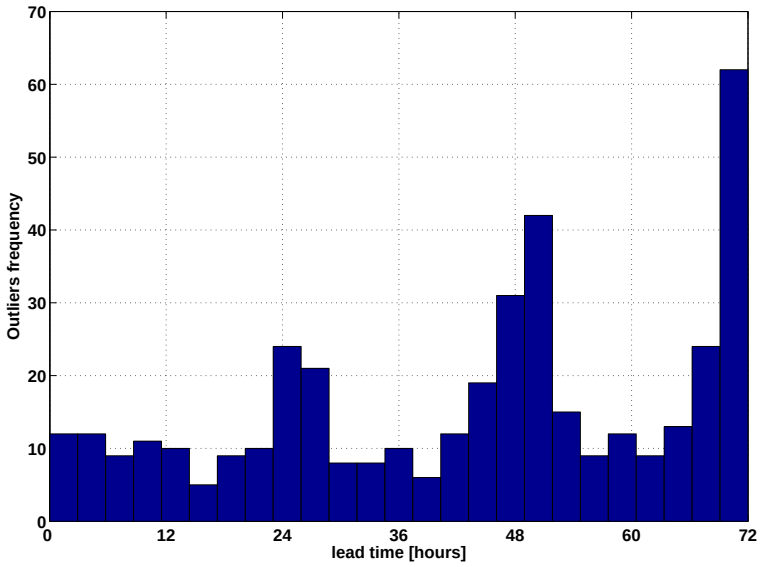

Figure 6.2: Outliers frequency in 2008 depending on the lead time. The outliers of all stations are considered.

where w is the length of the training period and t is the lead time. The parameters do not depend on the lead time.

Using this uncertainty model, simulations are performed for all stations for the year 2008. For each daily forecast of COSMO-7-KAL, the instantaneous spread(t) values of the COSMO-LEPS forecast are used to predict the uncertainties of the COSMO-7-KAL forecasts (Equation 6.5, example in Figure 6.4).

$$\sigma_{pred}(t) = a_w + b_w \cdot \text{spread}(t) \tag{6.5}$$

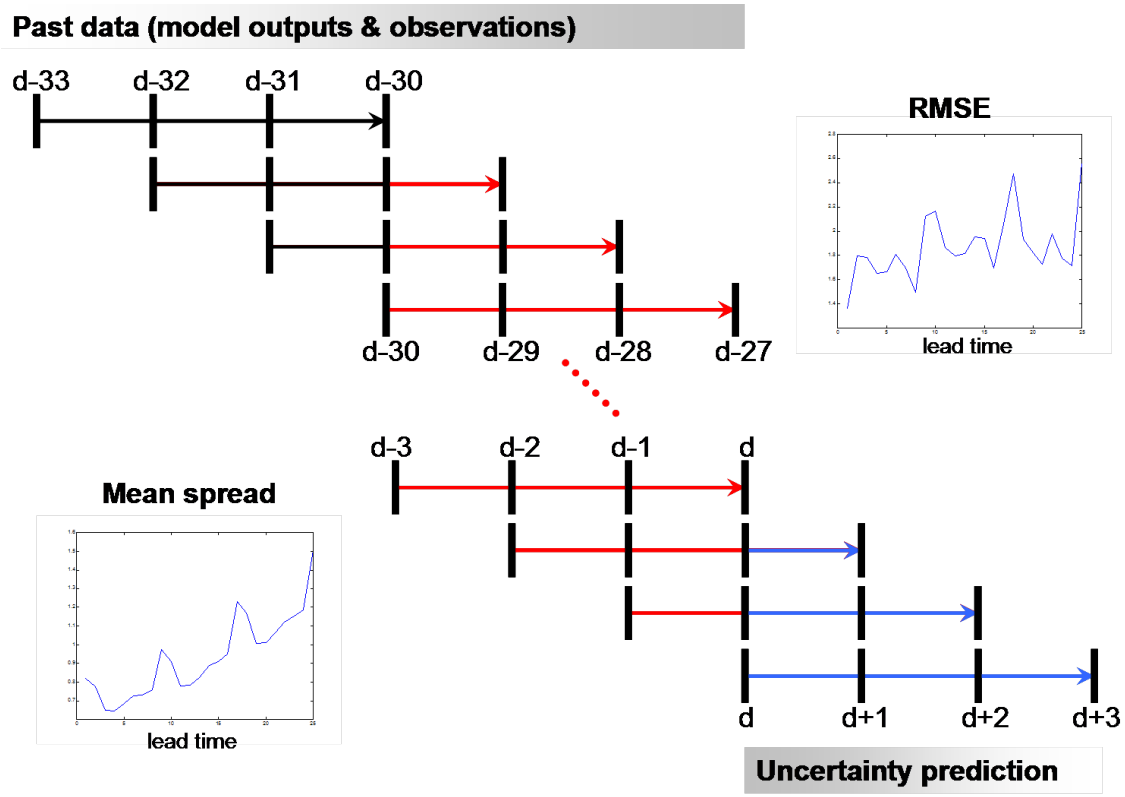

Figure 6.3: Schematic of the ALM approach. Based on the spread (COSMO-LEPS) and the RMSE (COSMO-7-KAL) of the last 30 days, a linear relationship is deduced. The model parameters are then used to predict COSMO-7-KAL uncertainty in the next three days using the spread of COSMO-LEPS as predictor.

Verification of the ALM

For each daily forecast in 2008, the RMSE of the COSMO-7-KAL forecast is calculated using Equation 4.1, where $M = 25$ lead times. For the same forecast, the 25 spread values of COSMO-LEPS are averaged. Figure 6.5 shows the DRMSE evolution without filtering of the outliers. In the same plot, the mean spread values are given. The variability of the DRMSE values is larger than the variability of the daily mean spread values. This indicates that COSMO-LEPS is not able to predict the full variability of the daily uncertainty. The large DRMSE peak on the 24.05.08, along with the increasing and decreasing in DRMSE prior and after this peak is not reproduced through COSMO-LEPS. The differences between both NWP models – COSMO-LEPS and COSMO-7 – are probably responsible for this. For instance, different snow covers in COSMO-7 and the members of COSMO-LEPS are possible. Note that the variability of DRMSE is larger for mountain stations.

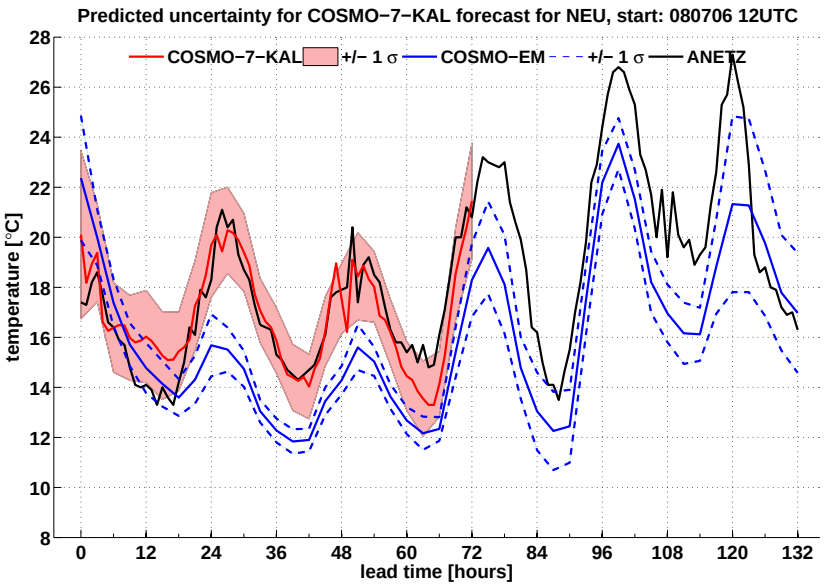

Figure 6.4: Example of an uncertainty prediction using the ALM (issuing time: 06.07.08 12 UTC) for the station Neuchatel (NEU, southerly exposed slope near lake Neuchatel). The parameters of the uncertainty model are estimated using the forecasts and observations of the last 30 days at station Neuchatel. The extreme outliers are filtered based on the error distribution of the years 2006 and 2007. The uncertainty prediction is computed using the spread of COSMO-LEPS as predictor. The spread (blue patch) is plotted around the COSMO-EM forecast (line in dark blue). When there is an overlap with the predicted uncertainty of COSMO-7-KAL (light red patch), the surface is plotted in dark red. The observations are given in black. Note that the hourly forecasts of COSMO-7-KAL and the hourly instantaneous observations are plotted, whereas the uncertainty predictions are 3-hourly (3-hourly COSMO-LEPS forecasts).

For the following investigations in this study, only outliers filtered error data are used. Figure 6.6 shows the DRMSE time series with filtered data for the same station as in Figure 6.5). Instead of the mean spread values, the daily mean predicted uncertainty values are given (DMU). For each forecast, the 25 predicted uncertainties are averaged. As the variability of the predictor is too low, the ALM also predicts too less variable DMU values. The predicted uncertainties better follow the DRMSE evolution, however there is a delay in the DMU compared to the DRMSE. An approach is proposed in the outlook to reduce the delay. Independent of the formulation of the uncertainty model, a delay is to expect as the uncertainty model is trained using historical data.

The first score which will be used for the verification and comparison of the uncertainty

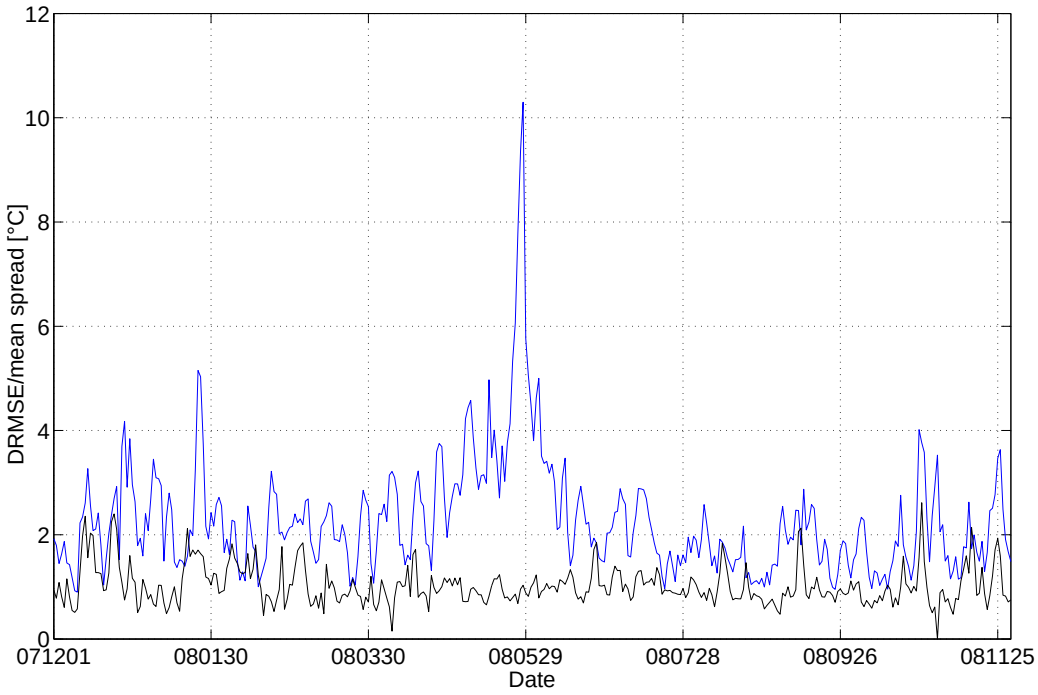

Figure 6.5: DRMSE (without filtering) and daily mean DMO spread values for Davos in 2008.

models is the root mean square difference between both DMU and DRMSE time series (Equation 6.6). The $RMSDs_{ts}$ for all stations are shown in Figure 6.7. The second score will consider the daily cycle of the predicted uncertainty (next subsection).

$$RMSD_{ts} = \sqrt{\frac{1}{M}\sum_{m=1}^{M}(\text{DMU(m)} - \text{DRMSE(m)})^2} \tag{6.6}$$

where the summation is applied over the M days (2008: 366 days).

Simulation of the daily cycle

The comparison between the mean of the daily simulated uncertainties (DMU) and the RMSE of the forecasts (DRMSE) is one aspect that must be taken into account for the verification of the uncertainty models. The second important aspect is the model ability to render correct daily cycles of the RMSE. In Figure 6.8 the monthly RMSE, DMO spread and simulated uncertainties (using the ALM30) for Davos are shown. The monthly $RMSD_{dc}$ (Equation

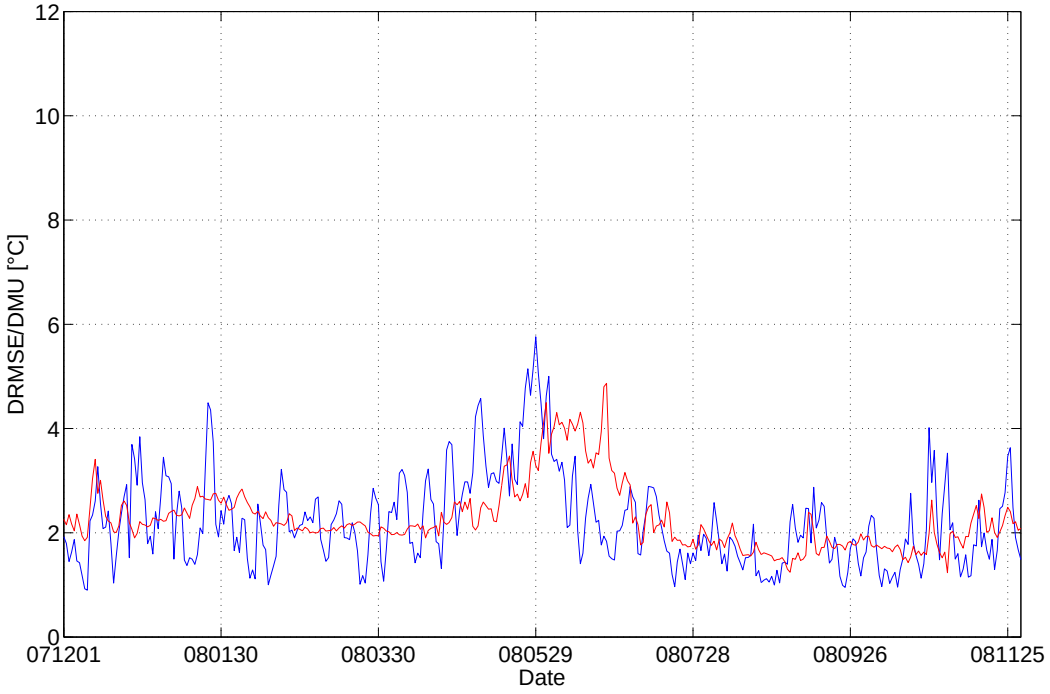

Figure 6.6: DRMSE (with filtered data) and DMU for Davos in 2008 using ALM30.

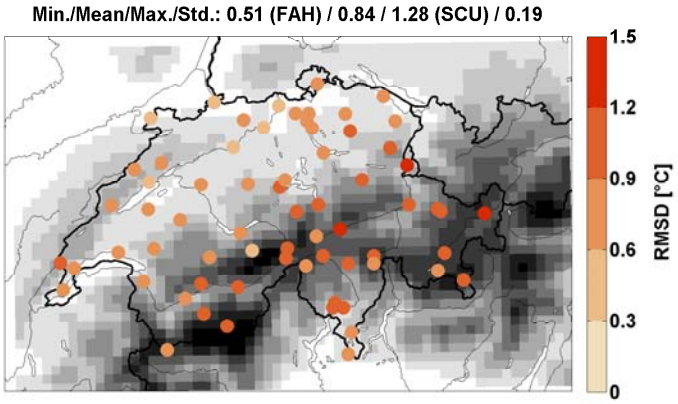

Figure 6.7: $RMSD_{ts}$ for the ALM30 in the year 2008.

6.7) between simulated uncertainties (σ_{pred}) and RMSE are averaged to get a value scoring the quality of the simulated daily cycle ($\overline{RMSD_{dc}}$). This score is plotted for all stations in Figure 6.9. Both aspects ($RMSD_{ts}$ and $\overline{RMSD_{dc}}$) will be used to determine the optimal training period after having analysed the ALM30 in more details.

$$RMSD_{dc} = \sqrt{\frac{1}{M}\sum_{m=1}^{M}\left(\sigma_{pred} - \mathrm{RMSE}(m)\right)^2} \tag{6.7}$$

where the summation is applied over the M=25 lead times.

The differences between the daily cycle of the RMSE and of the simulated uncertainty point out that the consideration of the time of the day in the uncertainty model would be beneficial (example of Davos in Figure 6.8, the difference in amplitude of the months April and May stand out). The amplitude of the daily cycle in RMSE cannot be satisfactorily simulated with the present model. To compare quantitatively the mean daily cycles of the RMSE and of the predicted uncertainties in 2008, the standard deviation of both detrended variables

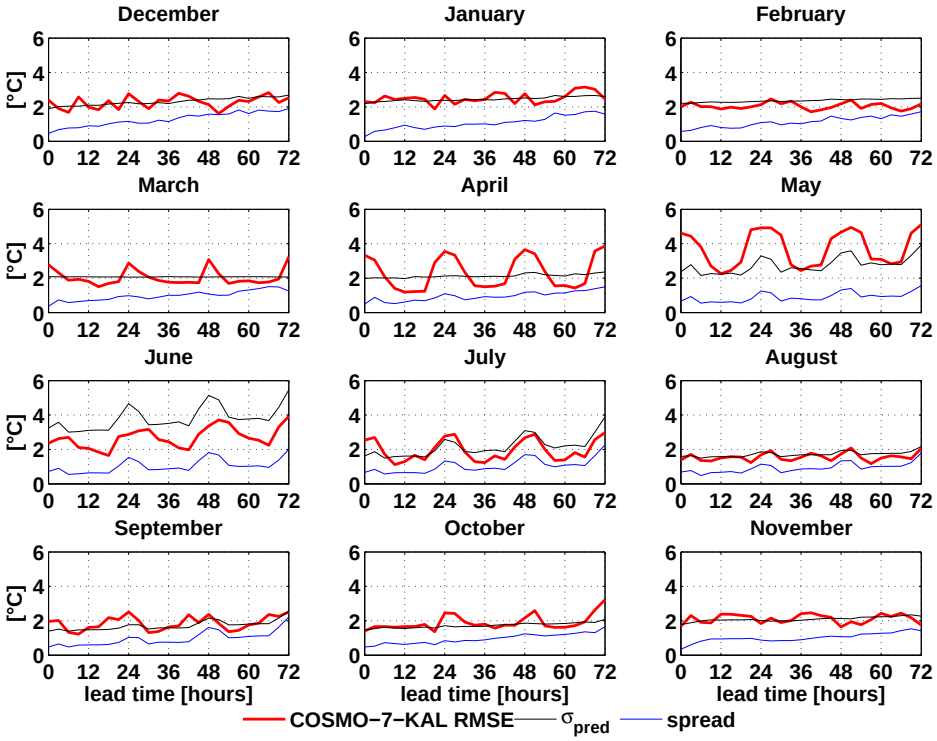

Figure 6.8: Monthly RMSE, predicted uncertainty with ALM30 and spread in function of the lead time in 2008 for Davos.

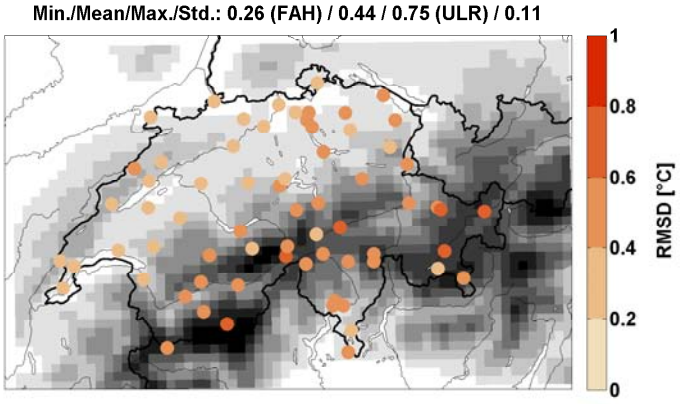

Figure 6.9: $\overline{RMSD_{dc}}$ for the ALM30 in the year 2008.

($\widetilde{\sigma_{pred}}$) and ($\widetilde{RMSE}$) are calculated. The time series are linearly detrended to eliminate the variability that is due to the trend. In doing so, the ratio only compares the variabilities in the daily cycle. The ratio between the variability of the predicted values and the variability of the RMSE (Equation 6.8) is plotted in Figure 6.10. The predicted uncertainties with the ALM30 explain on average only 40% of the variability of the daily cycle of RMSE. This motivates the consideration of the stratification of the SSR (dependency on the time of day, chapter 5) in the uncertainty model. In the next section, a pragmatic approach to extend the ALM will be proposed.

$$ratio = \frac{std(\widetilde{\sigma_{pred}})}{std(\widetilde{RMSE})} \tag{6.8}$$

Analysis of the parameters

In Figure 6.11 the parameters of the ALM30 for Zurich are shown along with the confidence intervals. The seasonal variability of the parameters at one station (standard deviation of the parameters time series in 2008) is always larger than the geographical variability of the averaged parameters (standard deviation of the mean parameter values of all stations). Especially mountain stations show large seasonal variations of the parameters compared to the mean geographical variability. Two conclusions can be drawn from this investigation: first,

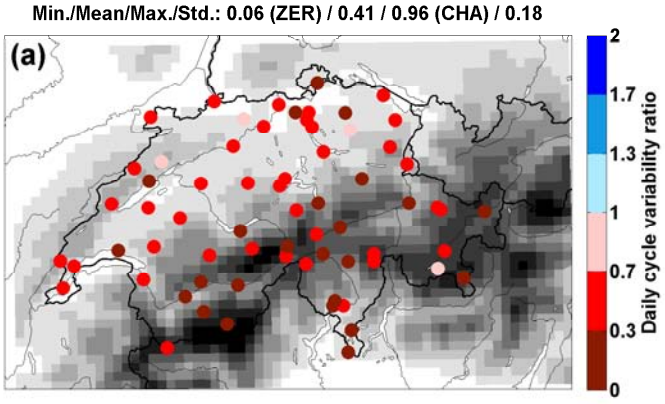

Figure 6.10: Variability ratio of the daily cycle using the ALM30.

the seasonal variability of the parameters cannot be neglected in the uncertainty models, confirming the need of adaptive models. Second, the difference in seasonal variability (e.g., Swiss Plateau vs. mountain stations) is so strong (Figure 6.12) that station-based uncertainty models are needed to take into account the geographical differences.

Sensibility to the training period

For comparing the sensibility to the training period (until now, 30 days), the uncertainty model was run with three other training periods: 15, 60 and 90 days. The two aspects, on one hand the $RMSD_{ts}$ between the DRMSE and the DMU time series and the average of the $RMSD_{dc}$ between the monthly RMSE and modelled uncertainties (daily cycle) have been investigated for the ALM in 2008. They are also calculated for the three other simulated training periods. For the ALM30, it was shown that the values of both aspects have not the same order of magnitude. While the mean $RMSD_{ts}$ was 0.84°C, the mean $RMSD_{dc}$ was 0.44°C (Figures 6.7 and 6.9). The $RMSD_{ts}$ and $RMSD_{dc}$ values for both important aspects were calculated for each station and training period. To determine the optimal training period, both aspects are weighted equally. As both values do not have the same order of magnitude, they are scaled between 0 and 1. For example, the four values (four training periods) of $RMSD_{dc}$ are taken for each station. Based on the minimum (min) and maximum

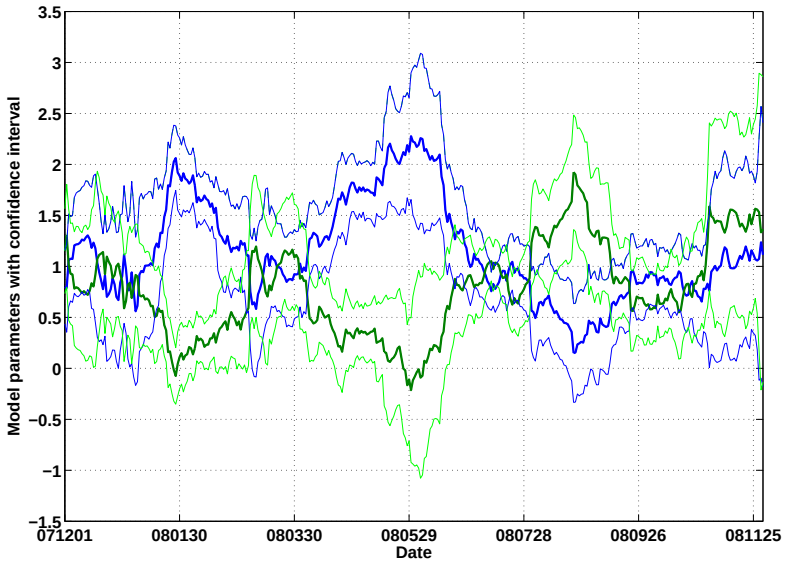

Figure 6.11: Evolution of the ALM30 parameters in 2008 for Zurich: a (blue) and b (green).

(max) of these four values, they are linearly scaled between 0 and 1 using following equation:

$$RMSD_{scaled} = \frac{1}{(max - min)} \cdot (RMSD - min) \tag{6.9}$$

where $RMSD$ can be either $RMSD_{ts}$ or $RMSD_{dc}$ depending on which values are scaled.

For each station, the four scaled $RMSD_{ts}$ values and the four scaled $RMSD_{dc}$ values are used to determine the optimal training period (example in Figure 6.13). The four absolute differences $RMSD_{ts,scaled} - RMSD_{dc,scaled}$ are calculated. The training period with the smallest difference is the optimal one. For most stations (40 stations or 60% of the 67 stations), 30 days is the optimal training period.

Residuals analysis

One assumption of linear regression is the normal distribution of the residuals (Equation 6.10), i.e. the differences between the values that are to be estimated ($\overline{\mathrm{RMSE}}_w(t)$) and the fitted values with the linear model ($\widehat{\mathrm{RMSE}}_w(t)$, Equation 6.4). If the residuals are not independent and identically-distributed (unit variance), they are not random and some systematic effects remain. In this case, the model can be improved. The distribution of all residuals for the

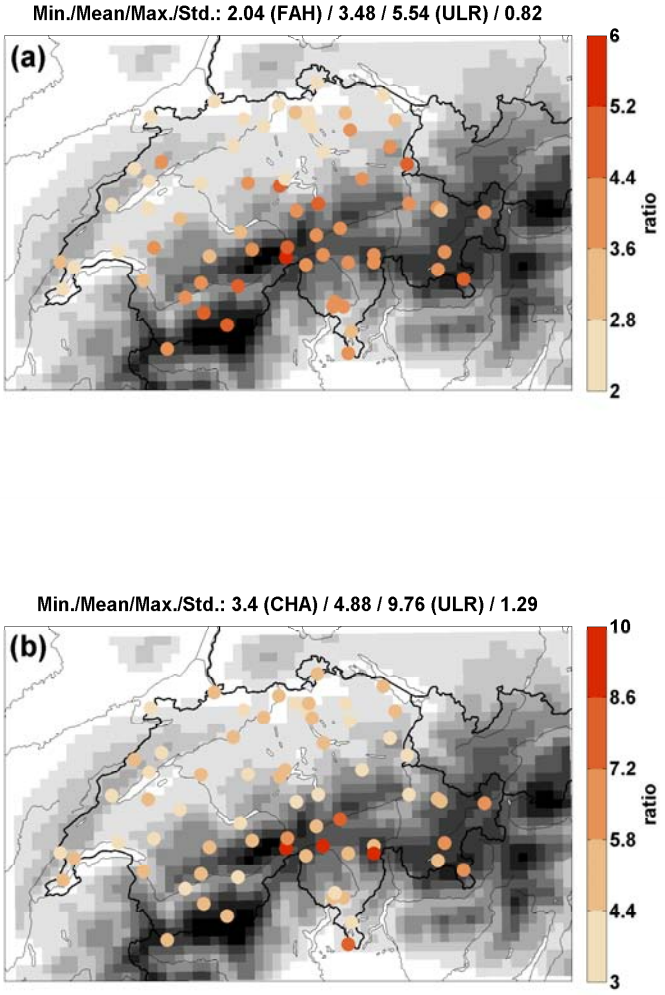

Figure 6.12: Plotted are the ratios of the seasonal variability at the stations vs. the mean geographical variability for the parameters of ALM30: (a) for a and (b) for b.

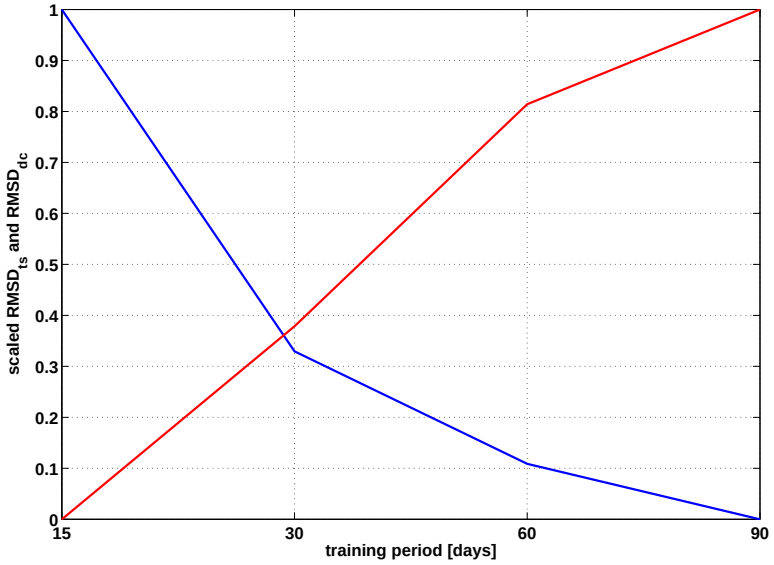

Figure 6.13: Method used to determine the optimal training period for the ALM. Example of the station Geneva. The four $RMSD_{ts,scaled}$ values are plotted in blue, the four $RMSD_{dc,scaled}$ are given in red (note that the x-axis is not linear). Theoretically, the exact x-position of the location where both lines cross could be determined. However, in this study the absolute differences between both lines are calculated for each of the four training periods, the training period with the smallest difference is assumed to be the optimal one.

station Zurich in 2008 (25×366 values) is shown in Figure 6.14. The distribution is quite normal (mean of 0, Gaussian shape), however the values in the tails are not consistent with the theoretical normal distribution: For that particular station, small negative outliers are too rare, large positive outliers are too frequent (i.e., right skewed). The tails of the error distributions are particularly sensitive to sampling (extreme values are much less frequent than values around 0). The problem in the tails of the distributions is especially demonstrative for mountain stations (example in Figure 6.15). The QQ-plots allow a more precise analyse rather than just visually considering the empirical distributions.

$$\text{residuals}(t) = \overline{\text{RMSE}}_w(t) - \widehat{\text{RMSE}}_w(t) \tag{6.10}$$

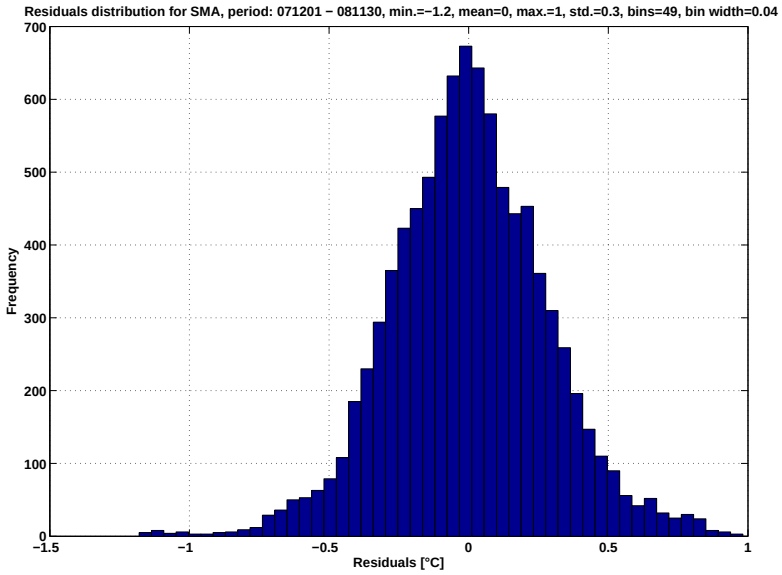

Figure 6.14: Distribution of the residuals for Zurich in 2008.

6.2 The extended ALM

In the subsection about the quality of the simulation of the daily cycle, it was shown that only a part of the variability in RMSE is explained through the ALM30. This is due to a large part to the simplification made in the formulation of the uncertainty model: the dependency of the SSR dependency on the time of the day has been neglected in the ALM30. As only 3-4 data points are available for each time of the day, it is not possible to estimate regression lines for each time of the day. The daily residuals do not always show a daily cycle (not shown) and thus, they cannot be modelled (e.g., using sine and cosine functions).

Therefore, a pragmatic approach, the extended ALM (Equation 6.11) is proposed which consists in adding the residuals (Equation 6.10) to the ALM (Equation 6.4). In doing so, the uncertainty contains a predictive part (using the spread as predictor) and a persistent part (the residuals). It is assumed that the residuals part is valid for the next forecast. There are virtually no residuals any more. The mean of the residuals is [jeweils] equal to zero. Therefore, the daily mean uncertainties (DMU) remain the same and the $RMSD_{ts}$ values do not change. However, the $RMSD_{dc}$ are different. The change in $RMSD_{dc}$ is shown in Figure 6.16. On average, the simulation (Equation 6.12) of the daily cycle of RMSE is improved for

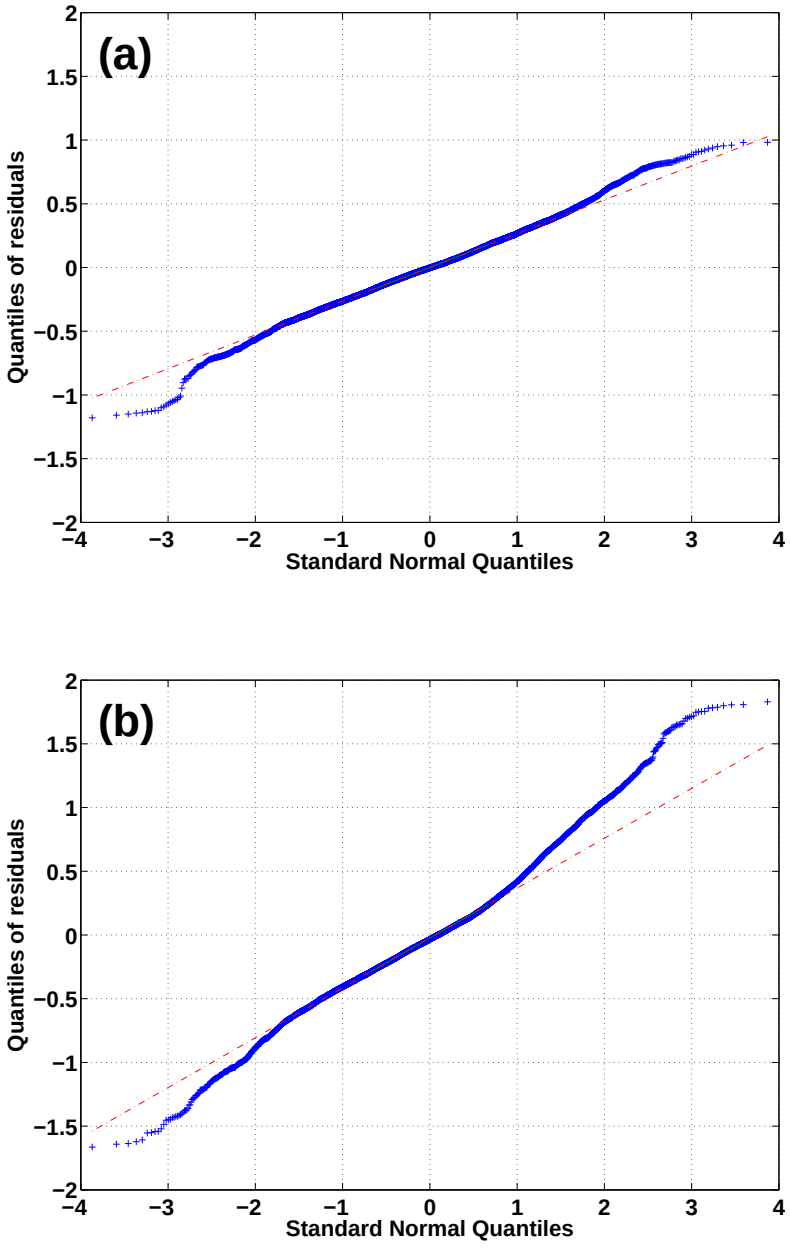

Figure 6.15: QQ-plots for two stations for the residuals of ALM30. (a) Zurich, (b) Davos. The red line represents the theoretical perfect normal distribution of the residuals. Both distributions are right skewed, but especially Davos and mountain stations in general.

all stations by 15%. Also with this extended ALM, the optimal training period remains 30 days for most stations (64%). In Table 6.1, the variability ratio (Equation 6.8) is given and plotted in Figure 6.17. The extension of the ALM30 led to a better simulation of the daily cycle of RMSE. In the last section of this chapter, the performance of the extended ALM30 is compared to other simplier models.

$$\widehat{\mathrm{RMSE}}_w^{ext}(t) = \widehat{\mathrm{RMSE}}_w(t) + \mathrm{residuals}(t) \tag{6.11}$$

And the uncertainty prediction using the instantaneous spread reads:

$$\sigma_{pred}^{ext}(t) = a_w + b_w \cdot \mathrm{spread}(t) + \mathrm{residuals}(t) \tag{6.12}$$

6.3 Comparison with simple models

Similar to Figure 5.1 in chapter 5, Figure 6.18 shows the spread and the RMSE evolution depending on the lead time in 2008. Additionally, the predicted uncertainty with the extended ALM30 is given along with the persistence and the constant RMSE value derived in chapter 4. The 'new spread' obtained with the extended ALM30 is clearly better than the spread DMO. For each uncertainty model illustrated in Figure 6.18, the $RMSD_{dc}$ is calculated using

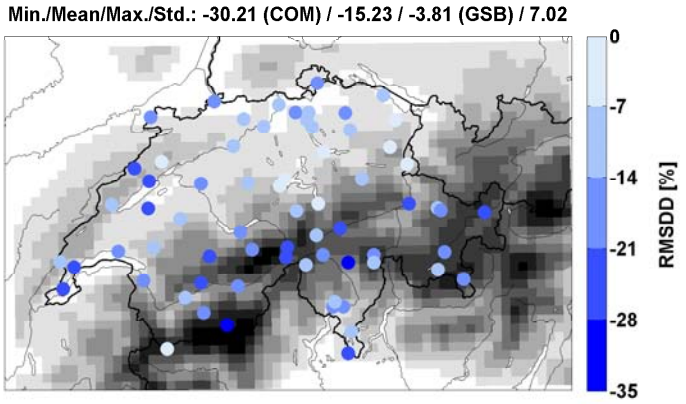

Figure 6.16: RMSDD when using the extended ALM30 instead of the simple ALM30 in 2008.

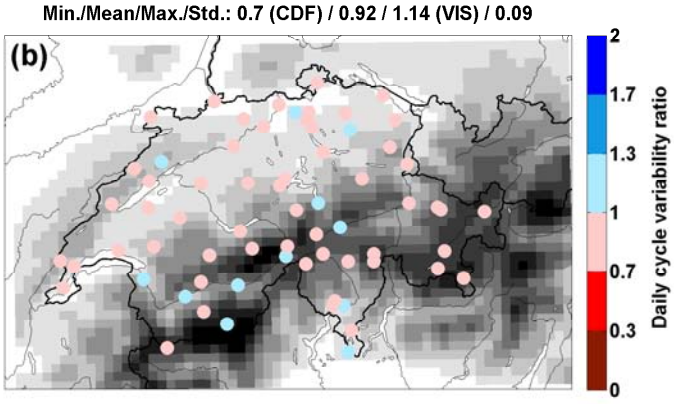

Figure 6.17: Variability ratio of the daily cycle using the extended ALM30. Compare with Figure 6.10.

Equation 6.7 where σ_{pred} is replaced by the considered model. A skill score (e.g., *Jolliffe and Stephenson* [2003]) is calculated ($RMSD_{SS}$, Equation 6.13) to obtain the relative skill when using the extended ALM30 instead of the spread DMO, the constant RMSE value or the persistence as reference (RMSD_{ref}).

$$RMSDD_{SS} = 1 - \frac{\mathrm{RMSD}_{ALM30_{ext}}}{\mathrm{RMSD}_{ref}} \tag{6.13}$$

Negative values of $RMSDD_{SS}$ indicate reduced performance with respect to the reference forecast. If $RMSDD_{SS}$ is equal to 1, the forecast is perfect. Table 6.1 shows the $RMSDD_{SS}$ compared to the spread, the constant RMSE value and the persistence for all the stations used in this study. It is not surprising that the extended ALM30 is clearly better than the spread and the constant value. Concerning the persistence: on average, persistence is as skilful as the extended ALM30. However, for a majority of stations (58%) the extended ALM is better than persistence, especially for low-level stations (Figure 6.19). This shows that the day to day variability of the uncertainty predictions for temperature based on ensemble information is not very high.

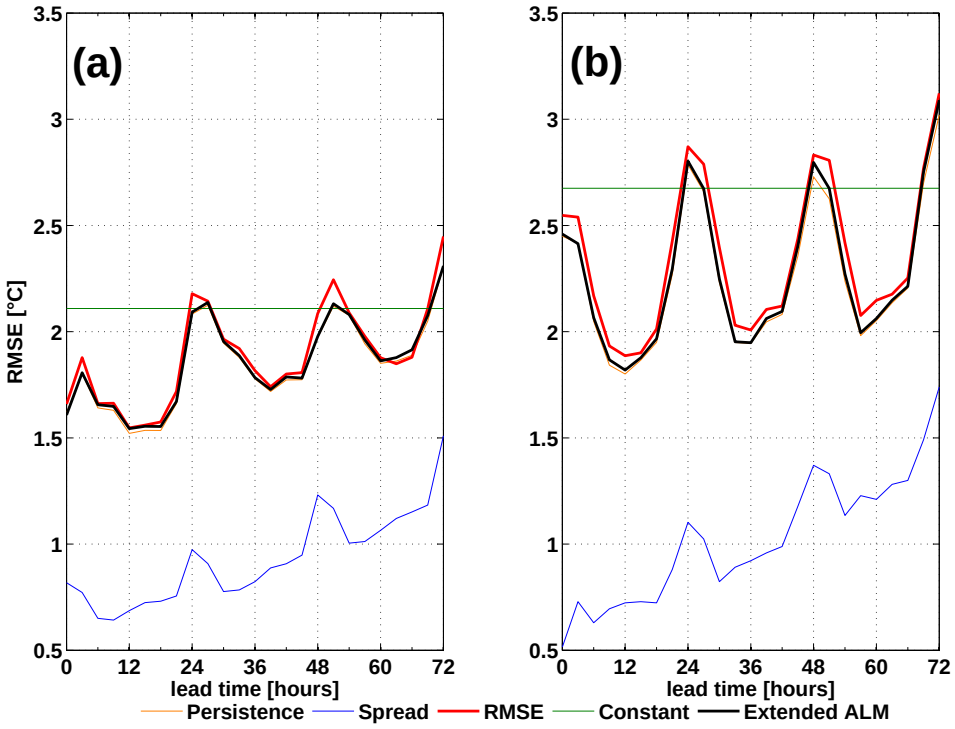

Figure 6.18: Uncertainty models comparison for (a) Zurich and (b) Davos in 2008.

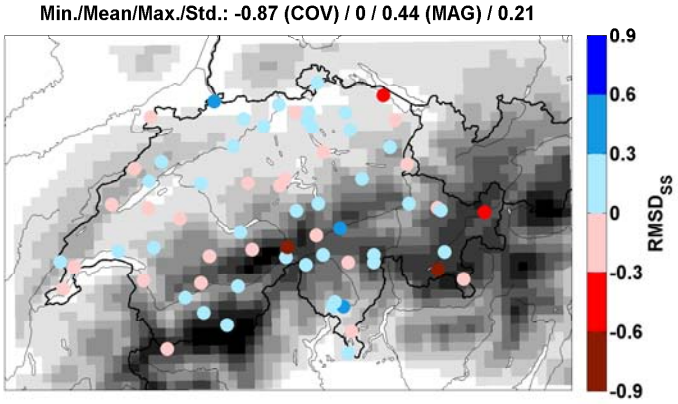

Figure 6.19: $RMSDD_{SS}$ for the extended ALM30 with respect to persistence in 2008.

Station	Variability ratio	SS wrt. spread	SS wrt. const.	SS wrt. pers.
GVE	0.95	0.97	0.89	-0.01
DOL	0.93	0.95	0.87	0.08
FRE	0.92	0.95	0.9	-0.02
NEU	0.89	0.97	0.89	0.05
CDF	0.67	0.87	0.44	-0.01
CHA	1.07	0.96	0.93	0.03
CGI	0.89	0.96	0.87	-0.2
PUY	0.96	0.97	0.9	0.11
PAY	0.92	0.96	0.87	-0.06
MLS	0.96	0.96	0.92	0.21
AIG	1.09	0.92	0.45	-0.1
PLF	0.94	0.93	0.67	-0.05
FAH	0.78	0.92	0.82	-0.13
BAS	0.97	0.98	0.94	0.35
RUE	0.96	0.96	0.89	0.12
SHA	0.81	0.93	0.72	0.05
BER	0.89	0.95	0.84	0.09
WYN	0.88	0.95	0.78	0.15
BUS	0.9	0.96	0.8	0.08
LAE	1.01	0.99	0.96	-0.19
REH	0.78	0.96	0.8	0.18
KLO	0.86	0.97	0.85	0.2
SMA	0.82	0.95	0.83	0.1
TAE	1.01	0.96	0.77	0.14
GUT	0.92	0.95	0.81	-0.35
HOE	1.04	0.97	0.93	0.01
ABO	0.89	0.95	0.8	-0.13
INT	0.97	0.95	0.71	0.1
NAP	0.91	0.97	0.94	-0.03
PIL	0.94	0.98	0.96	-0.02
LUZ	0.91	0.96	0.84	-0.02
ENG	0.91	0.95	0.66	0.07
ALT	1.18	0.97	0.54	0.18
WAE	0.8	0.95	0.81	-0.04
GLA	0.92	0.91	0.13	0.01
STG	0.86	0.91	0.3	-0.13
SAE	0.82	0.97	0.92	0.02
VAD	0.77	0.93	0.17	-0.03
SIO	0.99	0.98	0.92	0.24
VIS	1.08	0.97	0.78	0.15
MVE	0.81	0.94	0.88	-0.2
ULR	0.97	0.95	0.68	0.15
ZER	1.01	0.96	0.92	0.02
GSB	0.73	0.93	0.91	-0.16

EVO	0.84	0.96	0.94	0.05
GRH	0.89	0.95	0.91	-0.63
JUN	0.79	0.96	0.96	-0.22
GUE	0.85	0.95	0.91	-0.18
DIS	0.93	0.97	0.9	0.41
CHU	0.87	0.96	0.85	0.07
WFJ	0.69	0.94	0.94	-0.23
DAV	0.69	0.94	0.82	0.2
SCU	0.68	0.92	0.78	-0.39
SAM	0.83	0.94	0.71	0.03
COV	0.88	0.95	0.95	-0.87
HIR	0.84	0.93	0.43	0.02
SBE	0.84	0.92	0.67	0.09
ROB	0.84	0.95	0.71	-0.25
COM	0.9	0.97	0.87	-0.09
ROE	0.98	0.96	0.86	0.28
OTL	1.01	0.97	0.82	0.22
MAG	0.99	0.98	0.83	0.44
CIM	0.79	0.92	0.65	0.08
PIO	0.75	0.95	0.79	0.01
LUG	0.79	0.94	0.51	-0.07
SBO	0.96	0.96	0.75	0.17
PSI	0.85	0.96	0.93	0.07
MEAN	0.9	0.96	0.79	0.01

Table 6.1: Final scores summary. The variability ratio and the skill of the extended ALM30 with regard to the spread DMO, the constant RMSE value and the persistence are given.

Chapter 7

Conclusion and outlook

'Climate is what we expect, weather is what we get.'
probably from Mark Twain (1835-1910)

The probabilistic approach in NWP will continue to gain in importance in the next decade for the short-time to medium-range weather forecasts. The further development of ensemble prediction systems (EPS) will be central to respond to the demand for high-resolution probabilistic forecasts. For instance, the introduction of an own data assimilation and the further development of stochastic-dynamic parametrisations (*Palmer et al.*, 2005; *Berner et al.*, 2005; *Palmer and Hagedorn*, 2006) will help to better account for uncertainties in the initial conditions and the model formulation. Improvements in global EPS will indirectly improve limited-area EPS (LEPS). Simultaneously, post-processing methods (e.g., bias correction, spread calibration) will remain necessary to improve local weather predictions.

End-users are more and more interested in local uncertainty information. Also weather forecasts for the public increasingly consider uncertainties, in particular in form of exceedance probabilities for risk management, but also in form of probability distributions. For example, the probability of precipitation or probability of hail. Automatic applications are increasingly emerging on the weather market.

The present study has developed an approach that operationally predicts the expected uncertainty associated with local deterministic 2m temperature forecasts. A strong spread-skill relationship (SSR) has been found between the spread of the limited-area EPS COSMO-LEPS and COSMO-7, the deterministic limited area NWP. This relationship is mostly linear.

Therefore, given the skill could be interpreted as a measure of uncertainty of the forecast, the COSMO-LEPS spread has been identified as a suitable predictor for the COSMO-7 temperature uncertainty. The simple linear regression model only explains a part of the RMSE variability (on average 40%). Quite a large part of the uncertainty is attributed to the offset of this linear relationship indicating a considerable local uncertainty component that is independent of the uncertain atmospheric flow.

As only 3-4 data points are available for each time of the day, it is not possible to estimate regression lines for each time of the day. The daily residuals do not always show a daily cycle (not shown) and thus, they cannot be modelled (e.g., using sine and cosine functions). Therefore, an extended linear was proposed with added residuals. The extended ALM corrects the dispersion of the uncertainty model depending on the time of the day. The model already simulates much better the daily cycle of the RMSE. However, the persistence forecast is already very good (on average equal performance as the extended model) and delivers a good estimate of uncertainty if no ensemble is available.

For operational purpose, I suggest to use 00 UTC COSMO-7 outputs instead of 12 UTC as used in this study in order to use the newest COSMO-7 forecasts when COSMO-LEPS is available with a training period of 30 days.

Following the same approach for other variables like humidity, it would be possible to compute *'combined probabilities'*. For example, through multiplication of temperature and humidity probabilities, we could forecast probabilities for a 'comfort index' (e.g., how high is the probability that the temperature exceeds 25°C *and* the humidity 60%?).

The method developed in this study combines two model systems, but is very flexible and could be used to calibrate the spread of COSMO-LEPS specifically for the ensemble mean. As a result, in sympathy with the approach presented here, COSMO-LEPS DMO members would only be used to derive a probability distribution of the forecast that would be subject to calibration. In combination with an appropriate bias correction, the spread would be dressed according to past skill. Operational hourly COSMO-LEPS forecasts will be introduced in the near future. This will considerably expand the data basis (from 25 to 73 data pairs) leading to more robust daily estimations of the model parameters. Single linear regressions (5-6 values per forecast time) for each forecast time could be possible.

Acknowledgements

At this place, I would like to acknowledge all the persons without which this work could not have been realised.

Thanks to my parents, my brother, my family, and my friends for the moral and financial supports during my studies. Special thanks go to Lilly and Fritz by which I not only found a roof, but also a second family. The enriching discussions about God and the world and the encouragements during the last years were very precious to me.

Many thanks to Philippe Steiner who allowed me to write my thesis at MeteoSwiss and to Vanessa Stauch for the mentoring during the last year. When I didn't come further with my work, she was always here to help me. Thanks for the encouragements and the vitamins... I learned a lot about scientific working and writing. All the persons I became acquainted with at MeteoSwiss, especially the MO group, are acknowledged a lot for the help, the enriching discussions and the kindly working atmosphere during the last year.

Thanks a lot to Christoph Raible for the supervision of the thesis, the comments and corrections of the previous versions of this work. The quality of the work was substantially improved thank to the comments of Christoph Raible and Vanessa Stauch. Finally, I would like to thank H. Wanner for the exciting courses he taught with enthusiasm at the University of Bern and for reading my thesis.

Glossary

ALM	Adaptive linear (uncertainty) model
BMA	Bayesian model averaging
COSMO	Consortium for small-scale modeling
COSMO-7	Deterministic limited-area model with 6.6 km spatial resolution
COSMO-7-KAL	Kalman filtered COSMO-7 outputs
COSMO-EM	Ensemble mean of COSMO-LEPS
COSMO-LEPS	Limited-area ensemble prediction system with 10 km spatial resolution
CSCS	Swiss national supercomputing centre
DMO	Direct model output
ECMWF	European centre for medium-range weather forecasts
ECMWF-EPS	Ensemble prediction system of the ECMWF
EM	Ensemble mean
EPS	Ensemble prediction system
NGR	Non-homogeneous Gaussian regression
PDF	Probability density function
PDE	Partial differential equation
KAL	Kalman filter
LAM	Limited-area model
LEPS	Limited-area ensemble prediction system
NWP	Numerical weather prediction
MOS	Model output statistics
R	Pearson's correlation coefficient
R^2	Pearson's coefficient of determination
RM	Representative member
RMSD	Root mean square difference
RMSE	Root mean square error (skill measure)
spread	Standard deviation of an ensemble around its mean
SSR	Spread-skill relationship
WMO	World Meteorological Organisation
3D	three-dimensional

List of Tables

List of Figures

Bibliography

Ban, J., *Completing the forecast: characterising and communicating uncertainty for better decisions using weather and climate forecasts*, The national academies press, Washington, 2006.

Berner, J., J. Jung, P. T.N., and G. Shutts, Stochastic parametrisation of unresolved scales, *Geophysical Research Abstracts, 7*, –, 2005.

Buizza, R., P. Houtekamer, Z. Toth, G. Pellerin, and M. Wei, A comparison of the ECMWF, MSC, and NCEP global ensemble prediction systems, *Monthly Weather Review, 133*, 1076–1097, 2005.

Buzzi, M., 2008, *Challenges in Operational Numerical Weather Prediction at High Resolution in Complex Terrain*, Ph. D. thesis, Swiss Federal Institute of Technology (ETH).

Charney, J., Numerical integration of the quasi-geostrophic equations for barotropic and simple baroclinic flow, *J. Meteorol., 10(2)*, 71–99, 1953.

De Elía, R., R. Laprise, and D. Bertrand, Forecasting skill limits of nested, limited-area models: A perfect-model approach, *Monthly Weather Review, 130*, 2006–2023, 2002.

Draper, N., and H. Smith, *Applied regression analysis*, Wiley, 1998.

DWD, Die neue Modellkette des DWD I, *Promet, Tech. Rep. 27*, 2002.

Eckert, P., and D. Cattani, Construction of a combined synoptic and local confidence index, www.wmo.int/pages/prog/www/DPFS/Meetings/ET-EPS_Exeter2006/Doc4(3).doc, WMO, Meeting of the Expert Team On Ensemble Prediction Systems, Exeter, UK, 6-10 February 2006, 2006.

Ehrendorfer, M., Predicting the uncertainty of numerical weather forecasts: a review, *Meteorologische Zeitschrift, N.F.6*, 147–183, 1997.

Epstein, E., Stochastic dynamic prediction, *Tellus, 21*, 739–759, 1969.

Erdin, R., 2009, Combining rain gauge and radar measurements of a heavy precipitation event over Switzerland: Comparison of geostatistical methods and investigation of important influencing factors, Master's thesis, Swiss Federal Institute of Technology (ETH).

Fischer, E., S. Seneviratne, P. Vidale, D. Lüthi, and C. Schär, Soil moisture-atmosphere interactions during the 2003 European summer heat wave, *Journal of Climate, 20*, 5081–5099, 2007.

Fundel, F., A. Walser, M. A. Liniger, C. Frei, and C. Appenzeller, Calibrated precipitation forecasts for a limited area ensemble forecast system using reforecasts, *Monthly Weather Review,*, in press, 2009.

Glahn, H., and D. Lowry, The use of model output statistics (MOS) in objective weather forecasting, *Journal of Applied Meteorology, 11*, 1203–1211, 1972.

Glisan, J., and A. Lupo, An extreme case of atmospheric blocking over western europe, 20th Conference on Climate Variability and Change / 88th Annual Meeting of the American Meteorological Society, 2008.

Gneiting, T., A. Raftery, A. Westveld, and T. Goldman, Calibrated probabilistic forecasting using ensemble model output statistics and minimum crps estimation, *Monthly weather review, 133*, 1098–1118, 2005.

Hamill, T., J. Whitaker, and C. Snyder, Generating inital conditions for ensemble forecasts: Monte-Carlo vs. dynamic methods, organization=16th Conference on probability and statistics in the atmospheric science, 2002.

Hamill, T., J. Whitaker, and X. Wei, Ensemble reforecasting: improving medium-range forecast skill using retrospective forecasts, *Monthly Weather Review, 133*, 1155–1174, 2004.

He, Y., F. Wetterhall, H. Cloke, F. Pappenberger, M. Wilson, J. Freer, and G. McGregor, Tracking the uncertainty in flood alerts driven by grand ensemble weather predictions, *Meteorological applications,* (16), 91–101, 2009.

Hess, S., 2009, Inversion forecast with the COSMO model, Master's thesis, Swiss Federal Institute of Technology (ETH).

Holton, J., *An introduction to dynamic meteorology* (4 ed.)., Elsevier Academic Press, 2004.

Jewson, S., 2004, Probabilistic temperature forecasting: a summary of our recent research results.

Jolliffe, I. T., and D. B. Stephenson, *Forecast verification: a practitioner's guide in atmospheric science*, Wiley, Chichester, 2003.

Kain, J., The Kain-Fritsch convective parameterization: An update, *Journal of Applied Meteorology, 43*, 170–181, 2004.

Leith, C., Theoretical skill of Monte Carlo forecasts, *Monthly weather review, 115*, 409–418, 1974.

Leuenberger, D., The SLEVE coordinate in LM, *COSMO Newsletter,* (2), 105–109, 2002.

Leuenberger, D., 2005, *High-resolution Radar Rainfall Assimilation: Exploratory Studies with Latent Heat Nudging*, Ph. D. thesis, Swiss Federal Institute of Technology (ETH).

Leutbecher, M., and T. Palmer, Ensemble forecasting, *Journal of Computational Physics, 227*(7), 3515–3539, 2008.

Liljas, E., and A. Murphy, Anders Ångström and his early papers on probability forecasting and the use/value of weather forecasts, *Bulletin of the American Meteorological Society, 75*, 1227–1236, 1994.

Lorenz, E., Deterministic non-periodic flow, *Journal of Atmospheric Sciences, 20*, 130–141, 1963.

Lorenz, E., Climatic predictability, *In: The Physical Basis of Climate and Climate Modelling, WMO, Geneva, 16*, 132–136, 1975.

Lynch, P., The origins of computer weather prediction and climate modeling, *Journal of Computational Physics, 227*(7), 3431–3444, 2008.

Marsigli, C., F. Boccanera, A. Montani, and T. Paccagnella, The COSMO-LEPS mesoscale ensemble system: validation of the methodology and verification, *Nonlinear Processes in Geophysics, 12*, 527–536, 2005.

Mason, S., and A. Weigel, A generic forecast verification framework for administrative purposes, *Monthly weather review, 137*, 331–349, 2009.

MeteoSwiss, 2007, Monatliche witterungsberichte.

Molteni, F., R. Buizza, C. Marsigli, A. Montani, F. Nerozzi, and T. Paccagnella, A strategy for high-resolution ensemble prediction I: Definition of representative members and global-model experiments, *Q.J.R. Meteorol. Soc., 127*, 2069–2094, 2001.

Molteni, F., R. Buizza, T. Palmer, and T. Petroliagis, The ECMWF ensemble prediction system: Methodology and validation, *Q.J.R. Meteorol. Soc., 122*, 73–119, 1996.

Montani, A., M. Capaldo, D. Cesari, C. Marsigli, U. Modigliani, F. Nerozzi, T. Paccagnella, P. Patruno, and S. Tibaldi, Operational limited-area ensemble forecast based on the Lokal Modell, *ECMWF Newsletter,* (98), 2003.

Montani, A., C. Marsigli, and T. Paccagnella, COSMO-LEPS status report, www.wmo.ch/pages/prog/www/DPS/Meetings/ICT-DPFS_Geneva2004/INF4_Report.pdf, COSMO meeting, Milano, 2004, September.

Montani, A., C. Marsigli, and T. Paccagnella, Five years of limited-area ensemble activities at ARPA-SIM: The COSMO-LEPS system, *COSMO Newsletter,* (8), 23–26, 2008.

Müller, W., C. Appenzeller, F. Doblas-Reyes, and M. Liniger, A debiased ranked probability skill score to evaluate probabilistic ensemble forecasts with small ensemble sizes, *Journal of Climate, 18*, 1513–1523, 2005.

Murphy, A., The ranked probability score and the probability score: a comparison, *Monthly weather review, 98*(12), 917–924, 1970.

Murphy, A., The value of climatological, categorial and probabilistic forecasts in the cost-loss ratio situation, *Monthly Weather Review, 105*, 803–816, 1977.

Murphy, A., The early history of probability forecasts: Some extensions and clarifications, *Weather and Forecasting, 13*, 5–15, 1998.

Palmer, T., R. Buizza, R. Hagedorn, A. Lawrence, M. Leutbecher, and L. Smith, Ensemble prediction: A pedagogical perspective, *ECMWF Newsletter,* (106), 2005.

Palmer, T., and R. Hagedorn, *Predictability of weather and climate*, Cambridge University Press, 2006.

Palmer, T., G. Shutts, R. Hagedorn, F. Doblas-Reyes, T. Jung, and M. Leutbecher, Representing model uncertainty in weather and climate prediction, *Annu. Rev. Earth Planet Sci., 33*, 163–193, 2005.

Persson, A., Kalman filtering, a new approach to adaptive statistical interpretations of numerical meteorological forecasts, *ECMWF Newsletter,* (46), 16–20, 1989.

Persson, A., and F. Grazzini, *User Guide to ECMWF forecast products*, ECMWF, 2005.

Raftery, A., T. Gneiting, F. Balabdaoui, and M. Polakowski, Using Bayesian model averaging to calibrate forecast ensembles, *Monthly Weather Review, 133*, 1155–1174, 2005.

Rodwell, M., Comparing and combining determinstic and ensemble forecasts: How to predict rainfall occurence better, *ECMWF Newsletter,* (106), 2005.

Roulston, M., and L. Smith, Combining dynamical and statistical ensembles, *Tellus, 55A*, 16–30, 2003.

Ruddiman, W., *Earth's Climate: Past and Future*, W.H. Freeman, 2001.

Schär, C., D. Leuenberger, O. Fuhrer, D. Lüthi, and C. Girard, A new terrain-following vertical coordinate formulation for atmospheric prediction models, *Monthly Weather Review, 130*, 2459–2480, 2002.

Schraff, C., Mesoscale data assimilation and prediction of low stratus in the alpine region, *Meteorology and Atmospheric Physics, 64*, 21–50, 1997.

Stauffer, D., N. Seaman, and F. Binkowski, Use of four-dimensional data assimilation in a limited-area mesoscale model. part ii: Effects of data assimilation within the planetary boundary layer, *Monthly weather review,* (119), 734–754, 1991.

Steppeler, J., G. Doms, U. Schättler, H. Bitzer, A. Gassmann, U. Damrath, and G. Gregoric, Meso-Gamma Scale Forecasts Using the Non-hydrostatic Model LM, *Meteorology and Atmospheric Physics, 82*, 75–96, 2003.

Stoll, M., 2005, The influence of radar data assimilation on aLMo precipitation forecasts in summer 2003, Master's thesis, University of Bern.

Tennekes, H., A. Baede, and J. Opsteegh, 1987, Forecasting forecast skill.

Tiedtke, M., A comprehensive mass flux scheme for cumulus parameterization in large-scale models, *Monthly weather review, 117*, 1779–1800, 1989.

Wilks, D., *Statistical methods in atmospheric sciences* (2 ed.)., Geophysics, Academic Press, London, 1995.

Wilks, D., Comparison of ensemble-MOS methods in the Lorenz '96 setting, *Meteorological applications,* (13), 243–256, 2006.

Declaration

under Art. 28 Para. 2 RSL 05

Last, first name:	Blanc Pascal
Matriculation number:	04-105-912
Programme:	Master in Climate Sciences
Thesis title:	Ensemble-based uncertainty prediction for deterministic 2 m temperature forecasts
Thesis supervisor:	PD Dr. Christoph Raible

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person, except where due acknowledgement has been made in the text. In accordance with academic rules and ethical conduct, I have fully cited and referenced all material and results that are not original to this work. I am well aware of the fact that, on the basis of Article 36 Paragraph 1 Letter o of the University Law of 5 September 1996, the Senate is entitled to deny the title awarded on the basis of this work if proven otherwise.

Bern, 18.08.2009

Signature