**BONNER METEOROLOGISCHE ABHANDLUNGEN**

Heft 65 (2014) (ISSN 0006-7156)

Herausgeber: Andreas Hense

Andreas Röpnack

# BAYESIAN MODEL VERIFICATION: PREDICTABILITY OF CONVECTIVE CONDITIONS BASED ON EPS FORECASTS AND OBSERVATIONS

# Bayesian model verification

## Predictability of convective conditions based on EPS forecasts and observations

DISSERTATION
ZUR
ERLANGUNG DES DOKTORGRADES (DR. RER. NAT.)
DER
MATHEMATISCH-NATURWISSENSCHAFTLICHEN FAKULTÄT
DER
RHEINISCHEN FRIEDRICH-WILHELMS-UNIVERSITÄT BONN

vorgelegt von
Dipl.-Meteorologe Andreas Röpnack
aus
Reutlingen

Bonn, März, 2013

Diese Arbeit ist die ungekürzte Fassung einer der Mathematisch-Naturwissenschaftlichen Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn im Jahr 2013 vorgelegten Dissertation von Andreas Röpnack aus Reutlingen.

This paper is the unabridged version of a dissertation thesis submitted by Andreas Röpnack born in Reutlingen to the Faculty of Mathematical and Natural Sciences of the Rheinische Friedrich-Wilhelms-Universität Bonn in 2013.

Anschrift des Verfassers: Address of the author:

Andreas Röpnack
Deutscher Wetterdienst
Frankfurter Straße 135
D-63067 Offenbach am Main

---

1. Gutachter: Prof. Dr. Andreas Hense, Rheinische Friedrich-Wilhelms-Universität Bonn
2. Gutachter: Prof. Dr. George Craig, Ludwig-Maximilians-Universität München

Tag der Promotion: 11.10.2013

# Contents

# Abstract

Forecasts of convective precipitation have significant uncertainties. Among the main reasons for these uncertainties are the non-linear dynamics of the atmosphere and approximations within the equations of the numerical weather prediction (NWP) models by unresolved physical processes, which have to be parametrized, and imperfect simulation of resolved physical processes. To account for the forecast uncertainties of convection permitting models, a convection permitting ensemble prediction system (EPS) based on the consortium for small-scale modeling (COSMO) model with a horizontal resolution of 2.8$km$ covering whole Germany is being developed by the Deutscher Wetterdienst (DWD). The deterministic model is named COSMO-DE. The potential of convective instability is affected by the vertical structures of temperature and humidity. These vertical profiles of the COSMO-DE-EPS and further ensembles will be investigated in this work. For verification of the vertical model profiles radiosonde observations are used. However, the observations are uncertain by themselves due to the well-known limits in observing the atmosphere.

The focus is to present a probabilistic method to verify and compare ensembles. The approach considers explicitly the observation error as well as the model uncertainty to validate multidimensional state vectors of temperature and equivalent potential temperature profiles of the COSMO-DE-EPS and of two meso-scale ensembles with horizontal resolution of 10$km$ and parametrized convection. The meso-scale ensembles are the COSMO short-range EPS (COSMO-SREPS) and the COSMO limited-area EPS (COSMO-LEPS).

The approach is based on Bayesian statistics and allows for both verification and comparison of ensembles. Both investigated variables define the dry and moist static stability of the atmosphere, and therefore they determine the necessary conditions of convection. The equivalent potential temperature contains the effect of the humidity, which cannot directly investigated, because the humidity is non-Gaussian. Since the temperature and equivalent potential temperature can be assumed to be Gaussian distributed, the Bayesian approach is solved analytically. Finally, the probabilistic approach gives an "evidence" for the ensemble under investigation in relation to a reference ensemble. This evidence is classified depending on the application either comparison or verification of ensembles.

The investigation period comprises the August 2007 for a comparison of the COSMO-DE-EPS with the COSMO-SREPS and the entire convective and orographically-induced precip-

itation study (COPS) period 2007 for a verification of the COSMO-SREPS and COSMO-LEPS against COSMO-EU analyses. It is shown that the temperature profiles modeled by the COSMO-DE-EPS have a higher evidence in view of the observations than those of the COSMO-SREPS. Furthermore, the evidence for the equivalent potential temperature is weaker due to the larger uncertainty of this variable in the model as well as in the observed state. This shows the importance of the observation uncertainty. Nevertheless, it seems that the COSMO-DE-EPS as a short range convection permitting ensemble is a suitable approach to consider the uncertainties in forecasting convection. The verification of two meso-scale ensembles COSMO-SREPS and COSMO-LEPS show a linear decrease of the probability (evidence) of the vertical temperature and equivalent potential temperature structure with increased forecast lead time. Furthermore, it is shown that the predictability of the convective conditions are up to 5 days. However, it is to consider that the typical time scale of convection is about hours, and beyond it is difficult to predict convection.

As a general result, the statistical model described in this study is appropriate to compare ensemble systems with each other. The score proposed in this work is a generalization of the Ignorance score taking additionally into account the uncertainty of the observations as well as the spatial correlation structure of the verified forecasts. The approach based on Bayesian statistics allows for a comprehensive evaluation of the forecast quality of three-dimensional samples by using just one score.

# Zusammenfassung

Vorhersagen von konvektiven Niederschlägen beinhalten erhebliche Unsicherheiten. Gründe hierfür sind die nicht-lineare Dynamik der Atmosphäre und die Näherungen in den Gleichungen der numerischen Modelle für die Wettervorhersage sowie durch nicht aufgelöste physikalische Prozesse, die parametrisiert werden müssen und weiter durch fehlerhafte Modellierung von physikalischen Prozessen. Zur Berücksichtigung der Unsicherheiten bei der Vorhersage von konvektions-erlaubenden Modellen entwickelt der Deutschen Wetterdienst (DWD) ein Ensemble-Vorhersage-System (Ensemble prediction system - EPS) für die Kurzfrist-Vorhersage, basierend auf dem COSMO-DE. Das deterministische COSMO-DE hat eine Auflösung von 2.8$km$ und umfasst ganz Deutschland. Das Potenzial der konvektiven Instabilität wird dabei durch die vertikalen Strukturen von Temperatur und Luftfeuchtigkeit bestimmt. Diese vertikalen Profile des COSMO-DE-EPS und weiterer Ensembles werden in dieser Arbeit untersucht. Für die Verifizierung der vertikalen Profile werden Radiosondenbeobachtungen verwendet. Diese Beobachtungen sind allerdings selbst auch unsicher aufgrund der bekannten Grenzen der Beobachtungen der Atmosphäre.

Das Hauptziel dieser Arbeit ist die Präsentation einer probabilistischen Methode zur Verifikation und zum Vergleich von Ensemble-Vorhersagen. Dieser Ansatz berücksichtigt explizit die Beobachtungs-Fehler sowie die Modellunsicherheit für die Validierung mehrdimensionaler Zustandsvektoren der Temperatur und der Äquivalent-potentiellen Temperaturprofile des COSMO-DE-EPS und zweier mesoskaligen Ensembles mit einer Auflösung von 10$km$ und parametrisierter Konvektion. Die mesoskaligen Ensembles sind das COSMO short-range EPS (COSMO-SREPS) und das COSMO limited-area EPS (COSMO-LEPS).

Der Ansatz basiert auf der Bayesischen-Statistik und ermöglicht sowohl eine Verifizierung als auch einen Vergleich der Ensembles. Beide untersuchten Variablen definieren die Stabilität der Atmosphäre und bestimmen die Bedingungen für Konvektion. Die Äquivalente-potentielle Temperatur enthält zudem die Wirkung der Luftfeuchtigkeit, welche nicht direkt untersucht werden kann, da die Luftfeuchtigkeit nicht Gauss-Verteilt ist. Die Temperatur und die Äquivalente-potentielle Temperatur dagegen können als Gauss-Verteilt angenommen werden, so dass der Bayesische-Ansatz analytisch gelöst werden kann. Schliesslich ergibt der probabilistische Ansatz einen "evidence" für das zu untersuchende Ensemble in Bezug auf ein Referenz-Ensemble. Dieser "evidence" wird klassifiziert, wobei die Klassifizierung dabei von der Anwendung entweder Vergleich oder Verifizierung des Ensembles abhängt.

Der Untersuchungszeitraum für den Vergleich des COSMO-DE-EPS mit dem COSMO-SREPS umfasst den August 2007 und den gesamten Zeitraum der konvektiv und orographisch induzierte Niederschlags-Studie (COPS) von 2007 für die Verifikation des COSMO-SREPS und des COSMO-LEPS gegen COSMO-EU Analysen. Es wird gezeigt, dass die Temperatur-Profile des COSMO-DE-EPS viel wahrscheinlicher sind (Sie haben einen höheren Beweis im Hinblick auf die Beobachtungen) als die des COSMO-SREPS. Ferner ist der Beweis für die Äquivalente-potentielle Temperatur schwächer aufgrund der grösseren Unsicherheiten dieser Variablen im Modell als auch in den Beobachtungen. Dies zeigt zudem die Bedeutung der Beobachtungs-Unsicherheiten. Dennoch scheint es, dass das COSMO-DE-EPS als konvektions-erlaubendes kürzestfrist Ensemble dazu geeignet ist, um die Unsicherheiten bei der Vorhersage von Konvektion zu berücksichtigen. Die Verifikation der zwei mesoskaligen Ensembles COSMO-SREPS und COSMO-LEPS zeigt eine lineare Abnahme der Wahrscheinlichkeit (evidence) der vertikalen Temperatur- und Äquivalente-potentielle Temperatur-Profile mit der Vorhersagezeit. Es wird zudem gezeigt, dass die Vorhersagbarkeit von Konvektion nach etwa 5 Tagen endet, da zu diesem Zeitpunkt die Vorhersage die gleiche Wahrscheinlichkeit wie die von Persistenz-Vorhersagen erreicht. Dabei ist zu berücksichtigen, dass die typische Zeit-Skala von Konvektion im Bereich von Stunden liegt und es darüber hinaus schwierig ist Konvektion quantitative vorherzusagen.

Es wird gezeigt, dass das statistische Modell, welches in dieser Studie beschrieben wird, geeignet ist Ensemble-Systeme miteinander zu vergleichen. Der Score, der in dieser Arbeit vorgeschlagen wird, ist eine Verallgemeinerung des Ignorance-Scores unter der zusätzlichen Berücksichtigung der Unsicherheiten der Beobachtungen sowie der räumlichen Korrelation. Der Bayes-Faktor ermöglicht dabei eine umfassende Auswertung der Prognosegüte mit einem einzigen Score.

# Chapter 1

# Introduction

Water is indispensable to life on earth. In this, precipitation is indispensable ingredient of the water cycle. On the one side, precipitation is essential for farming and consequently for food for human beings, and on the other side precipitation has also a very high potential for causing damage e.g. through flash floods. Here, the atmospheric state defines the availability of precipitation. Therefore, predictability of the state of the atmosphere in general and in particular quantitative precipitation forecasts (QPFs) are among the most important applications in numerical weather prediction (NWP). On the short-timescale up to the climate timescales, QPFs are of special economic, social and political significance.

In particular, forecasts of precipitation associated with deep convection have large uncertainties, concerning the prediction of the location and timing of the respective events (Browning et al., 2008). Figure 1.1 shows the verification results of QPFs from different meteorological weather services over several years. The forecast quality of all of them is clearly worst during the summertime with mainly convective precipitation (Ebert et al., 2003b). Weckwerth et al. (2004) underline this issue.

This is due to the fact that the instabilities of the large scale flow, e.g. expressed by positive values of convective available potential energy (CAPE), are released by random events, which themselves are triggered but not strictly determined by other flow or boundary properties such as orography, soil moisture etc. The instabilities of the atmospheric flow evolution strongly amplify small uncertainties either embedded in the large scale flow, in the boundary layer or in the surface characteristics on time scales of the order of the life time of the convective events releasing the instabilities. This limits the capability of NWP models to forecast the right diurnal cycle of precipitation (Guichard et al., 2004) and is one reason for a lack of significant improvement in QPFs during the last decades in contrast to other forecast variables (Hense et al., 2006). Further reasons for this lack of improvements are (Hense et al., 2003):

(a) *incomplete simulation of the components of the water cycle in NWP models*

(b) *gaps, non-resolved structures and errors of the initial data of NWP models*

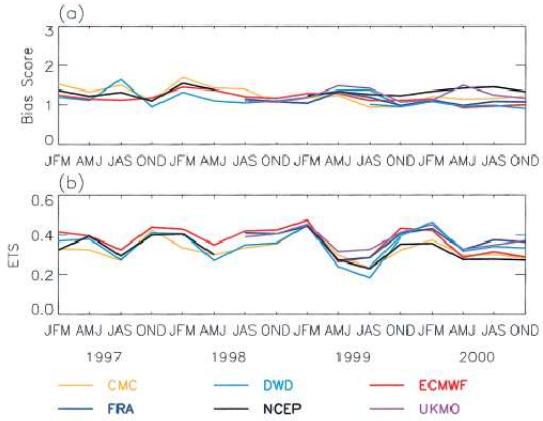



**Figure 1.1:** Time evolution of the Bias (a) and (b) of the equitable threat score (ETS) over Germany between January 1997 and June 2000 for a rain threshold of $2mmd^{-1}$ (Ebert et al., 2003b). The ETS measures the fraction of observed and forecasted events, which were correctly predicted (Ebert et al., 2010).

(c) *inadequate methods of data assimilation within NWP models*

(d) *limited predictability of the atmospheric state by deterministic NWP models*

A priority program (PP1167) "Quanitative Precipitation Forecasts" was initiated to challenge these points in a joint and coordinated effort of university institutes and other research institutions. Within the PP1167 there was a close collaboration of these university institutes and research institutions with the research and development department of the Deutsche Wetterdienst (DWD). The PP1167 started in 2004 and lasted until 2010. It was funded by the Deutsche Forschungsgemeinschaft (DFG). The main target of the PP1167 project was to improve the short and medium range QPF over central Europa. Therefore, the main assignment of tasks were (Hense and Wulfmeyer, 2008):

I *identification of physical and chemical processes responsible for the deficiencies in QPF*

II *determination and use of potentials of existing and new data and processes to improve QPF*

III *determination of the prognostic potential of NWP models by statistic-dynamic analysis with respect to QPF*

Figure 1.2 shows the topics of the PP1167. The PP1167 was subdivided into five topics. Topic (A) had the aim to investigate the atmospheric processes, which lead to precipitation. The second topic (B) aimed to improve the initial data of NWP by the use of more as well as of additional observations within different data assimilation systems. Topic (C) contained the development of new data assimilation systems and of new methods for validation of NWP models

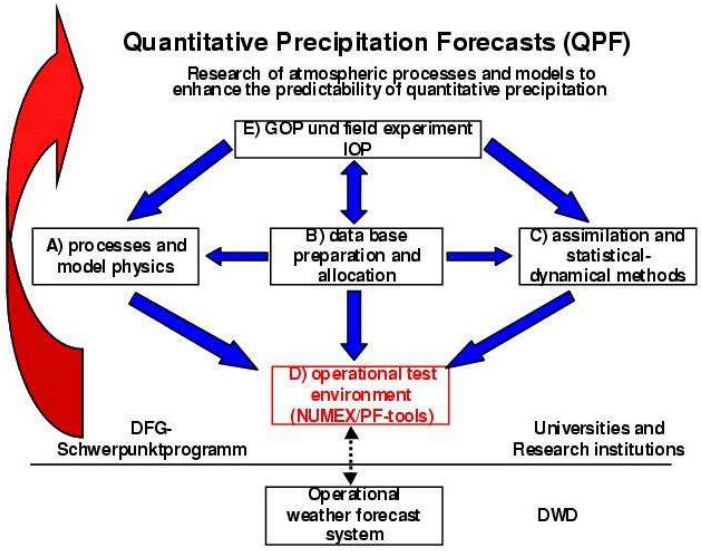



**Figure 1.2:** Structure of the priority program PP1167 "Quantitative Precipitation Forecasts", which includes the general observation period (GOP) and the convective and orographically-induced precipitation study (COPS). The COPS campaign was subdivided into several intensive observation periods (IOPs) (Hense and Wulfmeyer, 2008).

and for investigations of the predictability. One campaign is contained in topic (E). These campaign was the general observation period (GOP) including the convective and orographically-induced precipitation study (COPS). Finally, topic (D) was the connection of all topic subject areas into an operational test and evaluation environment of the DWD.

The thesis, which is presented here is a part of the PP1167 and belongs to the topic subject area (C) and has the aim to establish a new probabilistic verification method for vertical structures in temperature and moisture, which are important for convective developments. At that, the verification method offers also an access to probabilistic forecasts of ensemble prediction systems (EPS), because the approach takes up the model uncertainty from the ensemble.

Forecasts of EPS are important, because uncertainties in the initial data and the requirement to parametrize physical processes for the NWP models lead to a limited deterministic predictability. Lorenz (1963) shows the impact of small variations of the initial state to forecasts of a chaotic system. The atmosphere is such a chaotic system. Therefore, to deal with the limited deterministic predictability of processes on the small spatial scale the DWD has been developed an EPS based on the COSMO-DE, which is a convection permitting limited-area model with a horizontal resolution of 2.8*km* (Baldauf et al., 2011). The COSMO-DE was developed in the framework of the consortium for small-scale modelling (COSMO). An intermediate state of this COSMO-DE-EPS development is described in Gebhardt et al. (2011) and is investigated by the new probabilistic verification method for NWP presented in this work.

An EPS provides a sample of several deterministic realizations of the future atmospheric flow development. In COSMO-DE-EPS, this sample is obtained by perturbations of the initial and

boundary conditions and of the model physics to account for different types of uncertainties leading to the uncertainties in the forecasts. The additional information in the EPS, namely the spreading of the possible future paths in the atmospheric flow evolution reflects the forecast uncertainty in the predictions. The uncertainty can be quantified in terms of probabilities. It cannot be read off directly from the raw deterministic realizations, but requires a post processing step for deriving a probabilistic forecast based on the given forecast ensemble. Once this is done it can be shown that probability forecasts derived from an EPS are of greater benefit for decision making under uncertainty than a single deterministic forecast produced by the same model (Richardson, 2000). Similar as with deterministic forecasting, it has to be assured that the probabilistic forecasts of the post processed realizations indeed provide information of the future atmospheric state (Murphy and Winkler, 1984). This is the aim of verification of probabilistic forecasts using observations. Note that this is a problem by itself, because an EPS provides a sample of forecasts while nature provides only a single event. This will require to compare the predicted probability density of the future state of the atmosphere to the single observed state with the help of a score function (Gneiting and Raftery, 2007; Bröcker and Smith, 2007a). This score has to show the specific properties of propriety to allow an objective comparison of prediction and observations. Propriety means here that the score will always prefer a more accurate forecast (Bröcker and Smith, 2007a). The best score is obtained if and only if the predicted probability density is identical to that PDF from which the observations are drawn. Such a score is called strictly proper. The complete mathematics behind (strictly) proper scores is reviewed and presented in Gneiting and Raftery (2007).

Furthermore, the observed state is by itself uncertain due to the limited capabilities of observing the atmosphere. This again has to be taken into account to avoid that the forecasts are considered exhibiting low skill in case of verifying observations of low quality. Currently, errors of observations are an issue in research for verification. Bowler (2008) discussed the significant effect of observation errors on verification results. In particular, the effect of observation errors is not negligible when the forecast errors are small, e.g. at short lead times. This is confirmed by Candille and Talagrand (2008a), who mentioned that within forecast ranges up to two days the uncertainty of the verification results is of the same order as the uncertainty of the prediction due to the observation error. Candille and Talagrand (2008a) have introduced a method treating the observations as probability distribution. However their work is focused on binary events for given thresholds. In this work, the Bayesian statistical approach is used, which allows in a natural way to consider errors of the observations with their full probability density function (PDF), which is not threshold dependent. In the presented Bayesian approach, the uncertainty of the EPS as well as of the observation is taken into account simultaneously in a statistically consistent way without a restriction on the properties of the underlying PDFs. Using a prior probability allows to incorporate additional an unconditional prior knowledge. A further important advantage of the method is that a multivariate state of a continuous variable can be investigated.

In this study radiosonde measurements are used as verifying observations. Radiosonde observations can be considered to be of high quality, but uncertainties (sometimes called errors) are

important in our case. In this work the radiosonde observations are placed into a single column of the NWP model as a function of height. This is the same way as these observations are assimilated into the model state. Kitchen (1989) showed that this procedure is acceptable for the synoptic scale. However, for the COSMO-DE running at convection permitting scales the drifting of the radiosonde is certainly not neglectable. As a result the radiosonde observations are erroneous beyond the standard instrumental error, which has to be taken into consideration. This is done by using the observation errors of the three-dimensional variational (3dvar) data assimilation system of the DWD.

Additionally, it is often necessary to compare a specific EPS systems with another to decide about the relative quality given the same observational data set. Furthermore, in this work vertical structures are verified (realizations of multivariate random variables) to account for the dependencies (correlations) between the vertical levels. This is nearly impossible using single point information of one selected variable.

**Therefore, the key target of this work is to present a method**

- **to verify and compare ensemble predictions of atmospheric state vectors**
- **to include an uncertainty measure of the observations**
- **to allow for relative measures between different EPS systems**

All these aspects are important ingredients for verifying forecasts at resolutions, which permit convection and allow to study the predictability of convection initiation potential. For a better physical understanding of the prediction of these processes a multidimensional state vector of the forecast ensemble has to be used characterizing the vertical temperature and moisture structure. The multivariate aspect is defined by several vertical levels, which are treated simultaneously taking into account the dependencies between the levels. Other driving mechanisms like moisture convergence are not readily available from a single radiosonde ascent, but could be estimated better from a network of radiosondes. Therefore, the approach should also be capable to include several radiosondes profiles.

The probabilistic approach based on Bayesian statistic gives an "evidence" for the ensemble under investigation (i-ensemble) or rather against the reference ensemble (r-ensemble). This evidence is a continuous number ranges typically between -10 and +10, which will be classified. This classification is introduced in Chapter 4 and depends on the application either comparison or verification of ensembles. The investigated ensembles in this work are the COSMO-DE-EPS, COSMO-SREPS and COSMO-LEPS. The COSMO-SREPS and the COSMO-LEPS are ensembles based on the COSMO model, but with different driving models.

**The key questions to be answered with the help of the new method are**

Q1. *Is there a significant evidence for the new convective-permitting COSMO-DE-EPS with respect to the forecasted vertical structures when it is multivariate compared with the coarser resolved COSMO-SREPS ?*

Q2. *How does the verification of the forecasted vertical structures of COSMO-SREPS and COSMO-LEPS behave with lead time ?*

Q3. *Is there predictability at long forecasted lead times of COSMO-SREPS and COSMO-LEPS ?*

The probabilistic approach for the verification is based on the Bayesian verification method for climate change simulations by Min et al. (2004) and Min and Hense (2006). An extension of multivariate kernel dressing proposed by Schölzel and Hense (2010) is added to estimate in a more flexible way the predictive probability density from the raw ensemble samples. The new probabilistic verification method for NWP forecasts needs an estimate of the inverse covariance matrix of the internal variability of the ensemble. The standard maximum likelihood estimation of the covariance matrix often leads to non-invertible or singular matrices. This happens if the sample size of the ensemble used to estimate the covariance matrix is smaller than the dimension of the vectors, characterizing the vertical and horizontal temperature and moisture structure. In this work a method is introduced recently developed by Friedman et al. (2007) called the graphical lasso (gLasso) method, which is specifically designed to estimate non-singular covariance matrices and their inverse from small samples.

Previous work has shown that the Bayesian approach provides a convenient way to compare and verify ensemble climate simulations. Additionally, in Gneiting and Raftery (2007) it was shown that this approach leads to a proper score allowing an unbiased evaluation of the forecasts either with respect to a climatology or a different forecasting system.

The outline of the work is as follows: First, in Chapter 2 the general issue of verification and predictability is introduced. Then, the meteorological background, especially the processes of convection and convective initiation, is described in Chapter 3. In Chapter 4, the statistical fundamentals including of the Bayesian approach is introduced. The data and the methodology are described in Chapter 5. In Chapter 6, two meso-scale ensembles are compared. These two ensembles are the COSMO-DE-EPS and the COSMO-SREPS. The new COSMO-DE-EPS (DE-EPS hereafter) as a convection permitting short range ensemble is compared with the COSMO-SREPS, which uses parametrized convection. The COSMO-SREPS and COSMO-LEPS (LEPS and SREPS hereafter) is verified on the basis of analyses of temperature and equivalent potential temperature profiles as reference model in Chapter 7. Finally, the conclusions of this work and an outlook are given in Chapter 8.

# Chapter 2

# Numerical weather prediction verification and predictability

Numerical models like the NWP models are used to simulate complex physical processes to get information about the initial and the future state of physical systems. In meteorology, data assimilation combines observations with the numerical model to get the most possible accurate current state of the atmosphere. In turn, this current state is used as initial condition for NWP forecasts. Finally, for the assessment of the forecasts, verification is essential. In more details, verification is required (Ebert et al., 2010)

- to monitor the quality of NWP models,
- to improve the quality of NWP models and
- to compare the quality of different NWP models

The monitoring of the forecast quality is essential to answer the question of how precise is the forecast and whether the model developments really improve the forecasts. Furthermore, to improve the model, it is absolutely crucial to analyse the deficiencies in the model. This is the necessary very first step for improving the model. A further aspect, the comparison of NWP models is also very important and can be done with and without consideration of observations. The comparison without consideration of observations is done typically e.g. to investigate sensitivities of the model. However, for the assessment of two models, they have to be verified against observations.

Commonly, when forecasts are verified against observations, they are considered as the true state. However, they contain uncertainties. This issue is discussed in the next section and afterward the issue of verification of deterministic and probabilistic forecasts.

## 2.1 True state

It has been mentioned that the observed state $o$ is considered mostly as the true state. Thus here an overview about observation errors will be given. Ebert et al. (2010) pointed out that the "true state" of the atmosphere, which is used for verification, comes from more or less uncertain observations. In fact, all measurement instruments have a specific error range and they measure not exactly the truth. Daley (1993) divided the observation errors into natural and gross errors. The gross errors require a quality control check of the data. In the following, only the natural errors are discussed, which are

- instrument error
- and error of representativeness

Daley (1993) introduced further that these errors can be either random or systematic. Furthermore, they can be spatially or temporally correlated with each other or with the synoptic situation.

Also when analysis are used for verification instead of the raw observations, there exists always an uncertainty about the "truth state", which is in principle accessible e.g. if the analyses problem is formulated as a statistical regression problem. Until now, for verification, the errors in observational data and the errors in analysis data are mostly ignored. This proceeding seems to be justified when the errors in the observations are much smaller than the expected error in the forecast.

For reliable verification results, the forecasts and the corresponding observations are averaged over space and/or time to cancel out the effect of incorrect data, as far as possible. However, this required assumptions like independency of space and/or time if averaged over space and/or time. Thus, this proceeding can lead to problems like the masking of variations in non-homogeneous data, because of e.g. diurnal cycle, different weather regimes, station density etc. and further, standard verification methods often do not account for spatial correlations (Casati et al., 2008).

Currently, methods to account for errors in the observations are investigated e.g. in Bowler (2008) or Candille and Talagrand (2008b). Candille and Talagrand (2008a) mentioned that within forecast ranges up to two days the uncertainty of the verification results is of the same order as the uncertainty of the prediction due to the observation errors.

In this work radiosonde measurements are used as observations. The Bayesian statistics allow here explicitly to consider errors of the radiosonde observations. This is important, because in this work radiosonde data are used for verification in the same way as in the COSMO-DE where radiosonde data are used in the nudging scheme without correction for errors in position. In Chapter 5, the specifications of the radiosonde observations and of the observation errors are discussed.

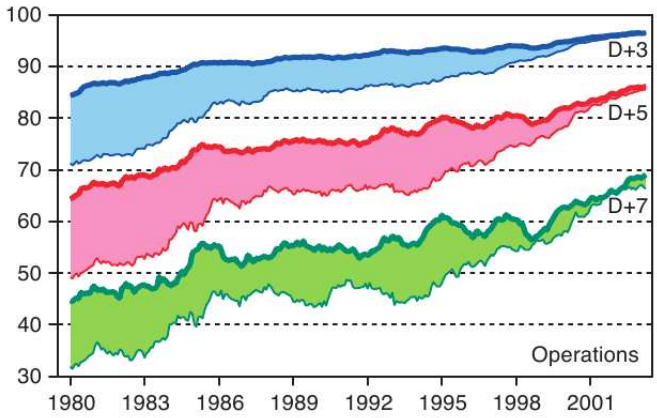



**Figure 2.1:** The historical time evolution from 1980 to 2001 of anomaly correlations of 500*hPa* height for forecasts at days D+3 (blue), D+5 (red) and D+7 (green) for ECMWF operational NWP forecasts in both hemispheres (upper/lower curve represent the northern/southern hemisphere). The area between the hemispheric anomaly correlations is shaded (Uppala et al., 2004).

## 2.2 Verification of deterministic forecasts

In general, forecast verification is the comparison of the forecasted state $f$ with the observed state $o$. Murphy and Winkler (1987) point out that this verification should be done based on the joint distribution of forecasts and observations.

The common verification of forecasts can be subdivided into

- categorical (discrete/continuous predictands)
- and continuous (only continuous predictands)

verification methods. In the following, only the issue of the continuous verification methods is discussed. Furthermore, a general overview of deterministic and probabilistic verification methods is presented. A detailed overview of the standard verification methods including the categorical verification methods can be found in Wilks (1995); Ebert (2002); Ebert et al. (2003a) and Ebert et al. (2010).

As an example of forecast quality monitoring, Fig. 2.1 shows the historical improvements of the operational deterministic NWP forecasts of the ECMWF since 1979. The main reasons for the improvement between 1980 and 2001 are the usage of advanced NWP models, but also of advanced data assimilation schemes (Uppala et al., 2004). Today, the quality of a 5 days forecast is comparable with the quality of 3 days forecast in 1980.

The Bayesian verification method, which is used in this work and the probabilistic scores with which the Bayesian verification results are compared is mathematically discussed in Chapter 4. The overview here is given for the standard verification methods and the term of predictability.

| | Range | Perfect score |
|---|---|---|
| Mean Error ($ME$) | $-\infty$ to $+\infty$ | 0 |
| Rout Mean Square Error ($RMSE$) | 0 to $+\infty$ | 0 |
| Anomaly Correlation ($AC$) | $-1$ to 1 | 1 |
| *Skill score* | $-\infty$ to 1 | 1 |

**Table 2.1:** Properties of few important scores (Ebert et al., 2010).

### Standard verification methods

For the standard verification methods the observed state is considered commonly as the true state. The model verification is done typically using standard scores like the mean error ($ME$) or the root mean square error ($RMSE$). The verification contains $N$ pairs of forecast and corresponding observation values.

$$ME = \frac{1}{N}\sum_{i=1}^{N}(f_i - o_i) \tag{2.1}$$

The $ME$ measures the systematic error (Bias) of a forecast.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(f_i - o_i)^2} \tag{2.2}$$

In contrast, the $RMSE$ respectively the $MSE$ measures the mean magnitude of the error. The value range of this scores and of further scores mentioned in the following is given by Tab. 2.1.

The anomaly correlation ($AC$) shown in Fig. 2.1 measures the spatial correspondence between a model forecast and the corresponding observations. Additionally, for the $AC$ the climatological mean $c$ at each point is subtracted. The anomaly correlation is often used to verify NWP models (Ebert et al., 2010).

$$AC = \frac{\sum_{i=1}^{N}(f_i - c_i)(o_i - c_i)}{\sqrt{\sum_{i=1}^{N}(f_i - c_i)^2}\sqrt{\sum_{i=1}^{N}(o_i - c_i)^2}} \tag{2.3}$$

At the end of this short introduction of selected verification scores, the concept of skill scores should be introduced. A skill score gives information about the benefit of a model forecast with respect to a reference forecast. For NWP models, the reference forecast is frequently the persistence or the climatology.

$$Skill\,score = \frac{score_{forecast} - score_{reference}}{score_{perfect\ forecast} - score_{reference}} \tag{2.4}$$

High resolution models like the COSMO-DE and probabilistic forecasts require advanced verification methods. This means, the standard verification methods mentioned here are only for deterministic forecasts and they are mainly for smooth forecast fields of NWP models with resolutions of about $10km$. Rossa et al. (2008) figured out this issue for QPFs. For high resolution models like the COSMO-DE ($2.8km$) the reader is referred e.g. to Wernli et al. (2007). The topic of verification of probabilistic forecasts needs methods, which can deal with an ensemble of deterministic forecasts. This will be discussed in the following. However first, the issue of predictability is discussed.

### Predictability

Predictability is a further important term. DelSole (2004) defines the term predictability as "the prediction errors have to be below those based on random selection of a realistic state". In general, four points are necessary to define adequately predictability of a deterministic model forecast:

- a forecast $f(\tau)$ at lead time $\tau$
- the observation at time $t = 0$ to define the initial error and
  the observation at time $\tau$ to define the final error
- a reference forecast $f_r(\tau)$
  to define an error level at lead time $\tau$ to be compared to the final error
- predictability is given as long as an error metric at lead time $\tau$ of the forecast
  is less than the error of the reference forecast

A common score used to measure predictability is the *MSE* or *RMSE*. The *RMSE* increases with lead time and reaches asymptotically a finite value, called saturation value. There, all predictability is lost. Figure 2.2 shows verification results of the global model GME via *RMSE*. For this verification, the temperature at $500hPa$ in different regions is verified. In general, the predictability is lost after about 7 days, because the *RMSE* has reached at this forecast lead time the quality of a climatological forecast. In case of persistence as reference forecast in place of climatology, there is also predictability up to 7 days. A saturation value can be seen only for the persistence in Fig. 2.2. Furthermore, it is shown that in the southern hemisphere the forecast quality is smaller than in the northern hemisphere. This is related to the density of the observations, which are much higher in the northern hemisphere. However, in the southern hemisphere, the usage of satellite measurements has also lead to a substantial improvement of the forecasts (Simmons and Hollingsworth, 2002).

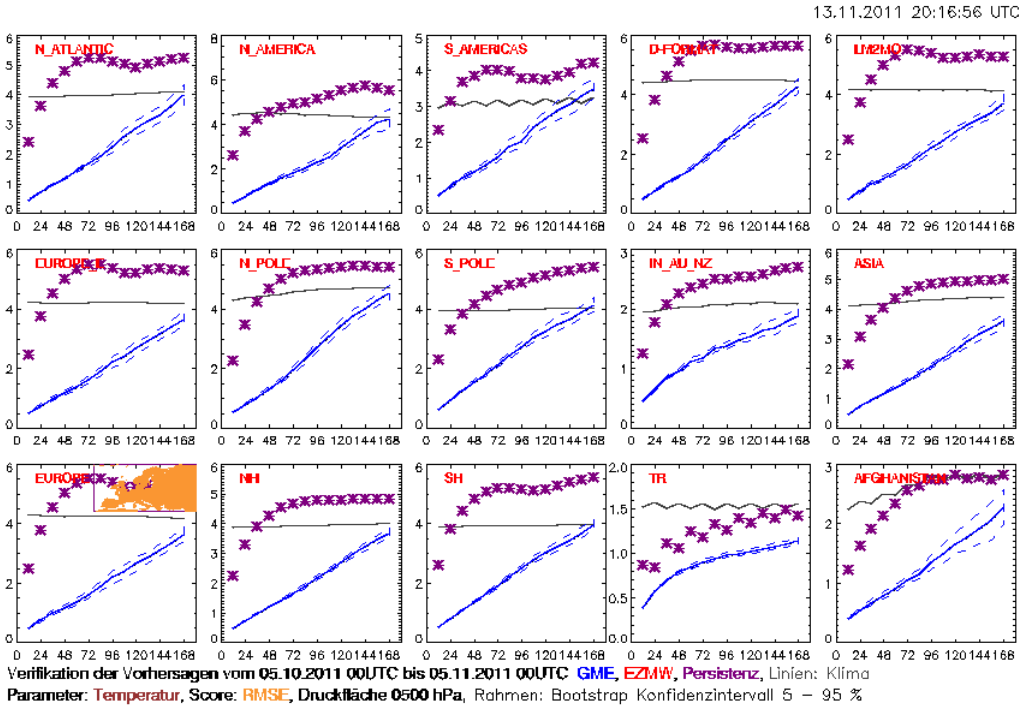



**Figure 2.2:** Example for standard verification of the GME (blue lines) for a month period in autumn 2011 using *RMSE* for different global regions at DWD. The dashed lines show the bootstrap confidence interval 5-95%. The verification of persistence (magenta crosses) and climate (black lines) are also shown. The verified variable is temperature at 500*hPa* (Damrath, 2011).

## 2.3 Verification of probabilistic forecasts

Caused by the steady increase of computing power and because of the chaotic nature of the atmosphere (Lorenz, 1984) a probabilistic point of view is more and more used today. This probabilistic forecast needs also to be verified, but with other verification methods than in the deterministic case. In this work, a probabilistic verification method to verify forecasts of vertical profiles is used to investigate the predictability of convection. This is necessary, because convection is still one of the worst predicted weather events. More about the issue of the limiting factors of predictability in NWP models relating to the simulation of convection are discussed in Chapter 3.

The general aspects about what an accurate probability forecast system has to have is specified in Ebert et al. (2010) as:

- **reliability** - agreement between forecast probability and mean observed frequency
- **sharpness** - tendency to forecast probabilities near 0 or 1
- **resolution** - ability of the forecast to resolve the set of sample events into subsets with characteristically different outcomes

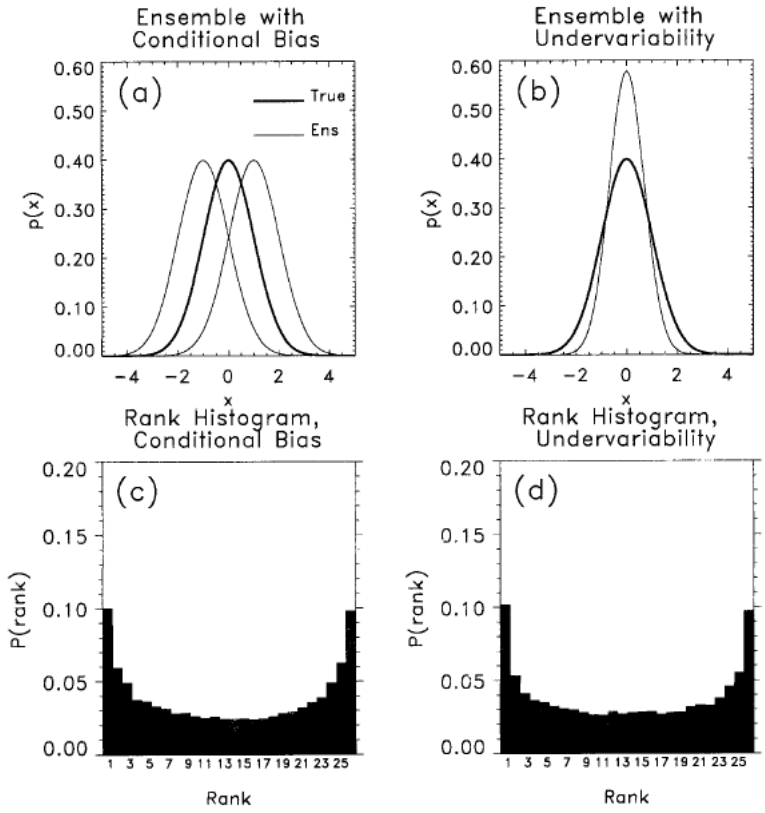



**Figure 2.3:** Illustration of different Rank histograms when the ensemble members are selected from a test probability distribution. (a) the probability distributions are two biased distributions N(-1,1) or N(1,1). (b) the distribution has a lack of variability N(0,0.69). (c) Rank histogram of a sample when ensemble members are selected from (a). (d) Rank histogram of a sample when ensemble members are selected from (b) (Hamill, 2001).

The main probabilistic verification methods are presented in the following. This probabilistic methods based mainly on scores. However, there are also purely graphical verification methods like the rank histogram, which does not explicitly need a verification score, but can be extended actually to a single number as shown in Keller and Hense (2011). The other verification is done directly by a score. This score $S$ attempts to compare the observation with the model forecast.

$$\langle S \rangle = \frac{1}{N} \sum_{i=1}^{N} S[p_i(x), o_i] \tag{2.5}$$

The verification score of probabilistic forecasts is here defined as a function $S[p(x), o]$. The ensemble forecasts is represented by a probability density function (PDF) denoted as $p(x)$ and $o$ is the observed state. A rough overview of the current status of forecast verification is given by Casati et al. (2008). For more details, the reader is referred to Gneiting and Raftery (2007); Bröcker and Smith (2007a, 2008) and Ebert et al. (2010). In the following, some scores will be further described.

### a. Verification of the raw members of an ensemble

The analysis rank histogram (ARH) or Talagrand diagram (Hamill, 2001) represents the frequency of the observation with respect to the single ensemble members sorted by class values. The histogram shows, if the ensemble has enough spread to represent the forecast uncertainty. An ensemble forecast with appropriate spread has a flat rank histogram. A "u-shape" histogram can indicate that the spread is too small (see Fig. 2.3), so that a large quantity of observations fall outside of the minimum or maximum of the ensemble forecast. Furthermore, the rank histogram shows if the ensemble has too large spread or if the ensemble contains a Bias. Figure 2.3 shows an example of a rank histogram where the ensemble spread is too small.

The ARH is based on the raw ensemble members in contrast to the probability integral transform (PIT), which based on the quantiles allowing also to compare ensembles e.g. with different number of members. The PIT is defined by the value of the predictive cumulative distribution function (CDF) at the observation point (Gneiting et al., 2008). "The PIT histogram can be interpreted in the same way as its discrete analogue, the Talagrand diagram" (Gneiting et al., 2008). Keller and Hense (2011) shows further how the PIT can be extended to a single number score. A "u-shape" histogram would correspond there to an negative score.

### b. Verification of PDFs based on the raw ensemble

The continuous rank probability score (*CRPS*) becomes just to be a very popular score. The *CRPS* in a member based formulation (Hersbach, 2000) belongs into the category verification of the raw members of an ensemble. The *CRPS* in a PDF based formulation is given by the integral of the Brier score at all possible threshold values $t$ (Gneiting et al., 2004) and belongs to the category verification of PDF based on the raw ensemble.

$$S_{crps}[p(x),o] = \int_{\infty}^{\infty} [F(t) - H(t-o)]^2 dt \tag{2.6}$$

Where $H$ is the heaviside function and $F(t) := \int_{-\infty}^{t} p(x)dx$ the cumulative distribution function (CDF). A mathematical introduction including a more detailed discussion of the *CRPS* is given in Chapter 4.

Another verification method of this type is the Ignorance score (Bröcker and Smith, 2007b). The Ignorance score does not capture the whole PDF, but rather the score verifies only one point of the ensemble PDF. The score is defined by

$$S_{ign}[p(x),o] = -log(p(o)) \tag{2.7}$$

The Ignorance score is defined as the value of the PDF at the observation value. Because of the negative sign, the logarithmic score cannot be negative. The logarithmic scoring rule is strictly proper, and smaller values of the score are better.

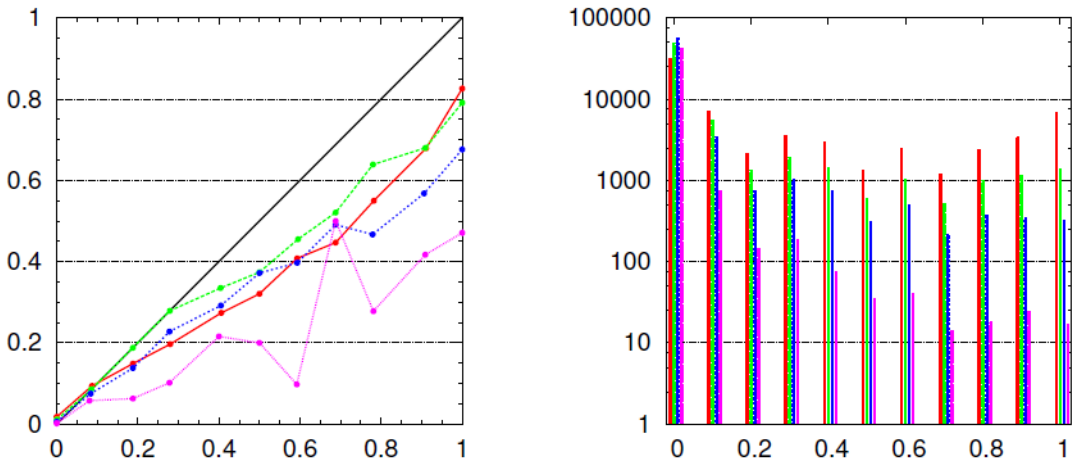



**Figure 2.4:** Reliability diagram of COSMO-LEPS 12*h* accumulated precipitation forecasts in spring 2006 on the left panel. The forecast range +18*h* for precipitation threshold 1*mm* (red solid line), 5*mm* (green), 10*mm* (blue), and 25*mm* (purple). The right panel shows the corresponding PIT histogram for the different thresholds (Walser and Liniger, 2008).

### c. Verification of the forecast of the probability of an event

Finally, the Brier score verifies the forecasted probability $P(x)$ of an event against the observed probability $P(o)$. The observed probability is defined as binary value either 1 (observed value above the threshold) or 0 (observed value under the threshold). The Brier score is defined by

$$S_{bs}[p(x),o] = (P(x) - P(o))^2 \tag{2.8}$$

The score measures the squared deviation between predicted probabilities and their outcomes, so a lower score represents higher accuracy, and the score is conceptionally similar to the mean square error.

Basically, both probabilities, which are used in the Brier score, the forecasted probability and the observed probability of a certain event can be illustrated in a Reliability diagram. The Reliability diagram plots the observed frequency against the forecast probability for a specific threshold. Figure 2.4 a) shows an example of such a Reliability diagram from Walser and Liniger (2008). The Reliability diagram shows a decreasing of reliability with an increased precipitation threshold. A perfect reliability would be the diagonal black line. Additionally, it is shown that the sample space becomes smaller for higher thresholds (Fig. 2.4 b).

## 2.4 Current area of research

Every score, if deterministic or probabilistic, investigates merely a few aspects of the type of model forecast errors. Thus, several scores have to be used to get a comprehensive quality

picture about the verified NWP model. Until now, verification scores are mostly univariate and do not consider the uncertainties of the observations. However both points are an actual issue in the development of new verification scores.

In this work, probabilistic forecasted vertical profiles are verified. A multivariate verification method based on Bayesian statistics is used including the consideration of the uncertainties of observations. The new verification score is an extension of the Ignorance score. The results of this verification method for NWP forecasts is compared with results of the Ignorance score as well as with the *CRPS*. This comparison is based on univariate results and is shown in Chapter 6. The details of the Bayesian verification method are described in Chapter 4.

# Chapter 3

# Fundamentals of convection and their forecasts

Convection is defined as vertical movement of many air parcels (plumes) due to buoyancy. For convection, the vertical structure of the atmosphere is very important, because it determines the stability of the atmosphere and therefore, the convection permitting conditions. This convection permitting conditions decide if convection can be triggered or not. The rising of the plumes are caused by buoyancy. In more detail, Emanuel (1994) defines a plume as a "buoyant jet in which the buoyancy is supplied steadily from a point source; the buoyant region is continuous". In this processes, water vapor plays a crucial role. Water vapor can reinforce the convection by release of latent heat. This physical processes are described in more detail in this chapter. Vertical temperature and humidity profiles, which determine the convection permitting conditions, are investigated in this work to study the predictability of convection by NWP models.

Generally, numerical simulation of processes in the atmosphere is a great challenge for NWP models. Since, this presupposes to consider processes on a large spectrum of spacial and temporal scales. These processes can be resolved only partly so that the unresolved processes have to be parametrized. Figure 3.1 shows the typical scales of thermo and fluid dynamic processes in the atmosphere. The spatial spectrum ranges from $10^{-3}m$ (the micro-scale, e.g. the magnitude of the dissipation of small turbulent vortexes called eddies) to $10^{7}m$ (the macro-scale, e.g. planetary waves and large cyclones). The time scale has a wide range too from about $1sec$ in the micro-scale to several years in the climatological scale.

In particular, one of the greatest challenges for NWP models is to predict convection. In this, convection belongs to the meso-scale with spatial scales between $1-10km$ and the corresponding time scale of about 1 hour. However, convection is also affected by processes on the micro and the synoptic scale (macro-scale). NWP models, which have the explicit task to give guidelines about convection are convection permitting NWP models. This NWP models simulate explicit convection without parametrizations. The guidelines relates to where the convection occurs and about the strength of the convective event. Convection permitting models like the

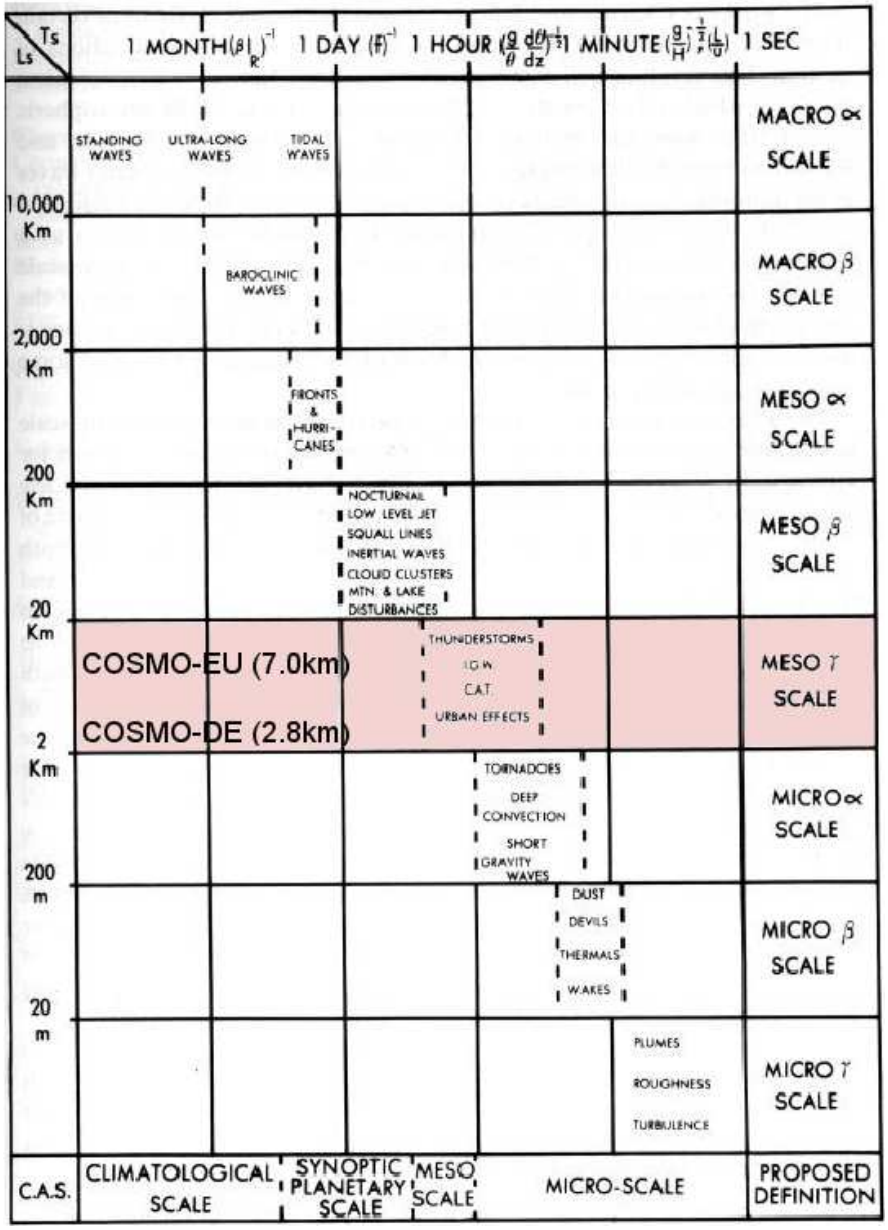



**Figure 3.1:** Scale definitions in meteorology and corresponding atmospheric processes with their characteristic scales of time and horizontal space (Orlanski, 1975).

COSMO-DE have a resolution of about 2*km* (COSMO-DE 2.8*km*) and thus belongs into the meso-scale. Figure 3.1 shows that processes of turbulence and plumes, which are also very important for the convective initiation are still unresolved with these resolutions and nevertheless, they have to be parametrized. NWP models with coarser resolutions, e.g. the COSMO-EU, belong into the meso-scale too. But with a resolution of 7*km* convection has to be parametrized completely. To summarize, it is a great challenge for NWP models to consider this huge spectrum of length and time scales shown in Fig. 3.1.

In this chapter, firstly, the general meteorological basics and following, the fundamental meteorological processes of convection and convective initiation are described. Afterwards, the simulation of convection in NWP models is explained. More precisely, the convective permitting NWP models as well as NWP models, which parametrize convection are considered. Finally, the issue of predictability of convection is discussed. The meteorological fundamentals described in this chapter are based mainly on Kraus (2004).

## 3.1 The stability of the atmosphere

The atmosphere consists mainly of nitrogen (78%) and oxygen (21%). The remaining part consists of the so-called trace gases such as argon, water vapor and carbon dioxide. Nevertheless, the water vapor in the atmosphere is extremely important for the meteorological processes in the atmosphere, because water vapor defines the general conditions for precipitation through the water cycle and its effects upon the radiation. Furthermore, water vapor is able to transport latent heat, which is very important for the energy balance of the earth-atmosphere system and for convection too. These special characteristics of water vapor are detailed discussed in the following.

### Water vapor

Water vapor is the gas phase of water. It is one state of water within the atmosphere beside the solid and liquid state of water. The proportion of water vapor content in the air can be up to 4 volume percent in the tropics. The global mean total water content is on average about $25 kgm^{-2}$, corresponding to a $25 mm$ high water column. The source of the water vapor in the atmosphere is primarily evaporation from the surface, especially from surfaces of the subtropical oceans.

The proportion of water vapor is, in contrast to many other gases, spatially and temporally highly variable and decreases rapidly with height. The ability of air to contain water vapor depends entirely on the temperature. Roughly speaking, warmer air can hold more water vapor than colder air. The Clausius-Clapeyron equation describes the exact characteristics of the saturation vapor pressure curve,

$$\frac{dE}{dT} = \frac{L}{T\Delta V} \tag{3.1}$$

where $\frac{dE}{dT}$ is the derivative of the saturation vapor pressure $E$ with respect to the temperature $T$. In the following $T$ has the dimension $K$ and $\vartheta$ gives the temperature in °C. $L$ is the latent heat of vaporization, and $\Delta V$ is the difference of specific volumes of the gas and liquid phase transition. In practical applications, the saturation vapor pressure is usually calculated by empirical functions. These empirical equation is the Magnus formula, see e.g. Gibbins (1990) or Lawrence

(2005). The Magnus formula for the saturation vapor pressure over water (temperature $\vartheta$ from $0°$C to $100°$C) is given by:

$$E = 6.1078 \cdot \exp\left(\frac{17.0809 \cdot \vartheta}{234.175 + \vartheta}\right) \tag{3.2}$$

and the Magnus formula for the saturation vapor pressure over ice (temperature $\vartheta$ from $-50°$C to $0°$C):

$$E = 6.1078 \cdot \exp\left(\frac{17.8436 \cdot \vartheta}{245.425 + \vartheta}\right) \tag{3.3}$$

As mentioned in the introduction (see Chapter 1), the water vapor plays a crucial role in the atmosphere, because water vapor has the attribute to occur in the atmosphere in all three physical forms (solid, liquid and gaseous). For the transition from one into another aggregate phase enthalpy is decreased or increased, e.g. water into water vapor consumes energy. This energy is released again as sensible energy at the opposite phase transition called latent heat. The property of the air to transport energy in form of latent heat is very important for the energy transport on earth, but also for the dynamic processes, e.g. the development of hurricanes. The latent heat of evaporation is thereby $L = 2.5 \cdot 10^6 Jkg^{-1}$ at $0°$C.

In meteorology, there are two main processes which are very fundamental related to the stability of the atmosphere namely dry and wet adiabatic processes. The ascent of an air parcel can be described either by a dry adiabatic process or by a wet adiabatic process. The wet adiabatic process is used if condensation of water vapor occurs. Both processes are described in the following.

### a. Dry adiabatic process

In case of unsaturated air, dry adiabatic processes dominate the stability of the atmosphere. The dry adiabatic process can be defined by the potential temperature $\Theta$. This temperature is a theoretical temperature of an air parcel at pressure $p$, which describes the entropy density[1] of the air parcel. The potential temperature for dry air is given by

$$\Theta = T\left(\frac{p_0}{p}\right)^{\frac{R}{c_p}} \tag{3.4}$$

where $T$ is the current absolute temperature of the parcel, $R$ the gas constant of air and $c_p$ the specific heat capacity of dry air at a constant pressure. In case of an adiabatic vertical motion, the potential temperature is constant like the specific humidity and the mixing ratio.

[1] The entropy $S$ is directly connected to $\Theta$ by $dS = c_p \ln d\Theta$ (Kraus, 2004). Thus isentropes are also isolines of equal potential temperature and entropy. Furthermore, here it is talked about the specific entropy corresponding to entropy density.

Correspondent to the potential temperature, the dry adiabatic vertical temperature gradient $\frac{\partial T}{\partial z}$ is given by

$$\left.\frac{dT}{dz}\right|_{\Theta=const} = -\Gamma_d = -\frac{g}{c_p} \tag{3.5}$$

The dry adiabatic vertical temperature gradient is denoted by $\Gamma_d$.

### b. Moist adiabatic process

In case when the air is saturated, the equivalent potential temperature $\Theta_e$ describes the entropy density of moist air. Houze (1993) explained the equivalent potential temperature as the theoretical temperature, which the parcel would have if all the water vapor condense and the complete latent heat convert into sensible heat. The equivalent potential temperature is given by

$$\Theta_e = \Theta \cdot exp\left(\frac{L\,m}{c_p T}\right) \tag{3.6}$$

This formulation of $\Theta_e$ is defined in Bolton (1980). $L$ is here the latent heat of evaporation, which was already mentioned before, and $c_p = 1004 Jkg^{-1}K^{-1}$ is the specific heat capacity of dry air. The humidity is represented by the mixing ratio $m$, for more details about $m$ and the context to the specific humidity $q_v$ see Appendix A.1. The equivalent potential temperature is conserved in a reversible moist adiabatic process meaning that the condensed water remains in the air parcel. Consequently, the vertical temperature gradient for saturated air is lower as for dry air, because of the release of the latent heat.

$$\left.\frac{dT}{dz}\right|_{\Theta_e=const} = -\Gamma_s = -\frac{g}{c_p} \cdot \frac{1+\frac{L}{R}\frac{m}{T}}{1+\frac{L}{c_p}\frac{m}{e}\frac{de}{dT}} \tag{3.7}$$

The vertical temperature gradient for saturated air ($s$) is expressed by $\Gamma_s$.

### Buoyancy

Vertical motion, more precisely, the raising of atmospheric air-masses is closely associated with the physical processes, which are responsible for precipitation. Reasons for vertical movement can be ascent at mountains, frontal cross circulations, horizontal wind convergence and convection. In case of convection, the reason for the vertical motion is buoyancy, which determines the rate of change $\frac{dw}{dt}$. In Kraus (2004), the buoyancy acceleration for an air parcel is given by

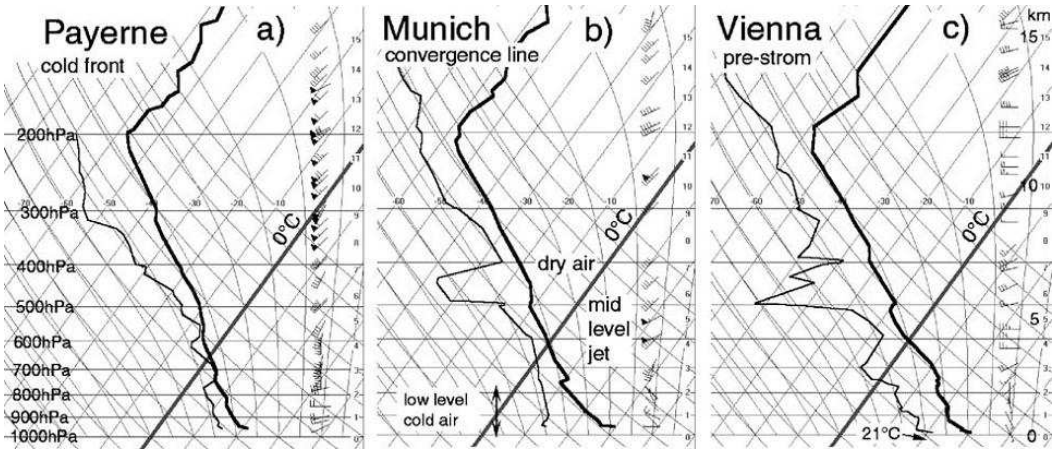



**Figure 3.2:** Vertical atmospheric structure along the cold front (a: Payerne), along the convergence line (b: Munich) and within the unstable warm sector (c: Vienna) from radiosonde soundings of 3th August 2001 at 12UTC. The corresponding modified CAPE values for pre-storm environment amounted up to $1200 Jkg^{-1}$ at location Munich (b) and eastward to Vienna (c) up to $4000 Jkg^{-1}$ (Kaltenböck, 2003).

$$\begin{aligned} \frac{dw}{dt} &= -\frac{1}{\rho}\frac{\partial p}{\partial z} - g \\ &= g\left(\frac{\rho - \rho_{parcel}}{\rho_{parcel}}\right) \\ &= g\left(\frac{T_{parcel} - T}{T}\right) \\ &= B \end{aligned} \tag{3.8}$$

These calculations contain the gas equation $\rho = \frac{p}{RT}$, the hydrostatic balance of the environment $\frac{\partial p}{\partial z} = -\rho g$ and that the pressure of the environment is equal to the pressure in the parcel $p = p_{parcel}$. The respective temperature $T_{parcel}$ is the temperature of a parcel, which ascents adiabatically, and $T$ is the temperature of the environment of the parcel. It is shown that the parcel will rise up if $T_{parcel} > T$. Therefore, the vertical velocity $w$ is strongly influenced by the environmental temperature T and furthermore, $w$ depends on the vertical structure of the environmental temperature $\frac{\partial T}{\partial z}$. However, the parcel method is a simplification of the convection occurring in the real atmosphere (Emanuel, 1994). In case of convection in the real atmosphere processes like e.g. entrainment and detrainment has to be considered too. In this sense, the environmental vertical structure of the humidity is also important. Redelsperger et al. (2002) investigated the effect of advected dry air to convection in the tropics. It was shown that extreme dry air has a clear effect on the convective development, especially on the convective cloud top height.

As main task in this work, the vertical structure of temperature and humidity is investigated. Therefore, the humidity $q_v$ will be treated by the usage of the equivalent potential temperature

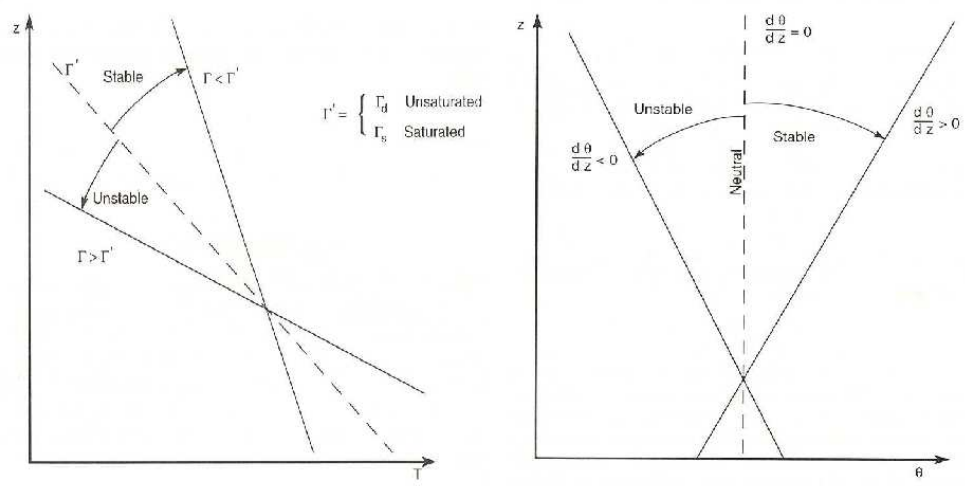



**Figure 3.3:** Illustration of stability cases of vertical temperature gradient (left panel) and of the vertical potential temperature gradient (right panel) (Sably, 1996). The stability criteria for the temperature gradient are complete, but they are not completely shown for the potential temperature gradient. The detailed stability criteria are shown in Tab. 3.1.

$\theta_e$, which allows to investigate also the humidity indirectly. The reason for this proceeding is that in this case both variables can be treated as normally distributed. More about this issue is written in Chapter 4. Figure 3.2 shows an example for vertical profiles at different weather situations. Figure 3.2 a) shows the case of a passing cold front at Payerne, (b) the passing of a convergence line at Munich and the pre-storm conditions at Vienna (c). In the area of the passing convergence line strong convection with thunderstorms was initiated. The vertical structure of the atmosphere is significant for the strength of the convection, and Fig. 3.2 shows the small scale structures especially of the dew point temperature. The crucial variables are the temperature and the humidity, which determine the potential for the convective instability.

## Stability criteria of the atmosphere

The stability of the atmosphere is determined by the atmospheric stratification of the atmosphere. This stability can be described by the vertical potential temperature gradient $\frac{\partial \Theta}{\partial z}$ in case of dry air and by the vertical equivalent potential temperature gradient $\frac{\partial \Theta_e}{\partial z}$ in case of moist air.

Figure 3.3 shows the case of stability in respect to dry and moist air. On the left hand side, the stability classification for the vertical temperature gradients is shown and the corresponding classification for the vertical potential temperature gradient on the right hand side. The classification for the vertical potential temperature applies only for dry air. The atmosphere is stable if the environmental vertical temperature gradient $\Gamma$

| | Vertical T-gradient | vertical $\Theta$ -gradient |
|---|---|---|
| Absolut stable | $\Gamma < \Gamma_s$ | $\frac{\partial \Theta_e}{dz} > 0$ |
| Wet neutral | $\Gamma = \Gamma_s$ | $\frac{\partial \Theta_e}{dz} = 0$ |
| Conditionally stable | $\Gamma_s < \Gamma < \Gamma_d$ | $\frac{\partial \Theta}{dz} > 0$ and $\frac{\partial \Theta_e}{dz} < 0$ |
| Dry neutral | $\Gamma_d = \Gamma$ | $\frac{\partial \Theta}{dz} = 0$ |
| Absolut unstable | $\Gamma_d < \Gamma$ | $\frac{\partial \Theta}{dz} < 0$ |

**Table 3.1:** Classification of the atmospheric stability through stability criteria represented by the vertical temperature gradient and the vertical potential/equivalent potential temperature gradient. $\Gamma$ without index represent the actual vertical temperature gradient. The index $d$ and $s$ stands for dry and wet adiabatic processes. The corresponding vertical $\Theta$ or rather $\Theta_e$ gradient are also shown.

$$\Gamma = -\frac{\partial T}{\partial z} \begin{cases} < & \text{(stable)} \\ = \Gamma_d & \text{(indifferent)} \\ > & \text{(unstable)} \end{cases} \tag{3.9}$$

is smaller as the dry adiabatic temperature gradient $\Gamma_d$ and unstable if $\Gamma$ is greater than $\Gamma_d$. This can be expressed also by the use of the potential temperature. The atmosphere is dry neutral, if the vertical $\Theta$ gradient is equal zero and stable (unstable) if the gradient is greater (smaller) than zero see Tab. 3.1.

Collectively, it be mentioned that an upward decrease of $\Theta_e$ is a necessary but not a sufficient condition for parcel instability (Emanuel, 1994). In case of unsaturated atmospheric air, the case of conditional stability can appear. Conditional unstable means the atmosphere is dry stable but the atmosphere becomes unstable eventually if condensation occurs. However, in case of unsaturated atmospheric air, the equivalent potential temperature has to be calculated using the temperature at the lifting condensation level $T_{LCL}$ in place of $T$ (Houze, 1993; Davies-Jones, 2009). Table 3.1 gives an overview of the stability conditions.

## Indices for potential convection

An index for the possible strength of convection is the convective available potential energy (CAPE). CAPE is thereby an index of the potential energy, which can be released from the unstable atmosphere.

$$CAPE = \int_{surface}^{EL} B \quad dz \tag{3.10}$$

CAPE is the vertical integral over the local buoyancy $B$ (see Eq. 3.8) of a parcel from the surface to the equilibrium level (EL). The level of free convection (LFC) defines the level where the air parcel rises up forced only trough his own buoyancy, and EL means the level where the buoyancy ends. CAPE is measured in $Jkg^{-1}$.

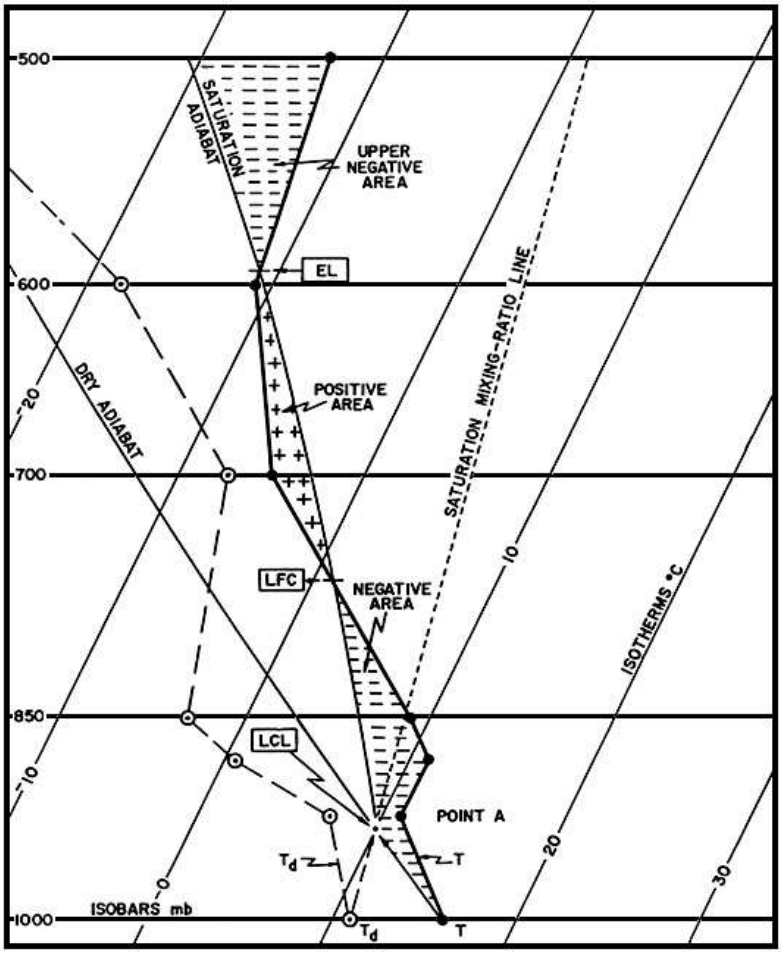



**Figure 3.4:** A schematic vertical profile of the troposphere. It is shown the level of free convection (LFC), the equilibrium level (EL) and the lifted condensation level (LCL). CAPE is represented by the positive area ('+' signes) minus the negative area ('-' signes) from the surface up to EL. CIN is represented by the negative area from the surface up to LFC, extracted from NOAA/ESRL (2011).

In the real atmosphere, there is often the case of an unstable stratification, but with a stable boundary layer. In this case, first, the unstable layer has to be overcome, e.g. by diabatic processes like heating of the ground-level trough solar radiation. An index for the stable layer is the convective inhibition (CIN).

$$CIN = \int_{surface}^{LFC} B \quad dz \tag{3.11}$$

CIN describes the energy, which is needed to overcome a stable boundary layer and it is also expressed in $Jkg^{-1}$. Figure 3.4 shows a vertical sounding of the atmosphere. CIN is marked as the negative area between the ground and the LFC and CAPE is represented as the positive area minus the negative area in Fig. 3.4. However, it should be mentioned that CAPE is a simplified theoretical estimation of the potential of convection.

So far, we have discussed the physical fundamentals about adiabatic processes and the possible stability conditions of the atmosphere. However, the circumstance of an unstable stratification of the atmosphere alone is not sufficient for initiation of convection. It is a necessary requirement. Furthermore, an explicit initiation of the convection is required. In the next section the topics of convection and convective initiation are illustrated in more detail.

## 3.2 Convection and convective initiation

In general, in physic the term convection describes vertical motion in fluids caused by differential densities. Rising plumes occur where the parcel temperature is warmer than his environment. This aspect was discussed previously in the section about buoyancy. Convection refers to the transport of properties by fluid movement. As such, it is one of the three main processes in which heat is transported: namely radiation, conduction and convection. In meteorology, convection describes the heat and mass transport by the vertical component of the flow associated with buoyancy.

Emanuel (1994) defines convection in the atmospheric science as:

> *"A class of relatively small-scale, thermally direct circulation which result from the action of gravity upon an unstable vertical distribution of mass".*

The requirements for deep convection are (Doswell, 1987; Johns and Doswell, 1992):

- *"a moist layer of sufficient depth in the low or mid-troposphere"*
- *"a steep enough lapse rate to allow for a substantial positive area (substantial amount of CAPE)"*
- *"and a lifting of a parcel from the moist layer to allow it to reach its level of free convection (LFC)"*

Doswell (1987) mentions too that an explicit trigger is necessary to initiate the process of convection. The process of triggering of convection or simple convection initiation (CI) can be done by boundary-layer forcing, upper-level forcing or initiation by previous convection (secondary generation) described in detail in Bennett et al. (2006). CI is one of the main topics in current research. In campaigns like the convective storm initiation project (CSIP) the convective initiation over the southern part of United Kingdom was investigated (Browning et al., 2008).

In general, the triggering of convection can be separated roughly into strongly forced convection (SFC), weakly forced convection (WFC) and into air-mass convection (AMC) without

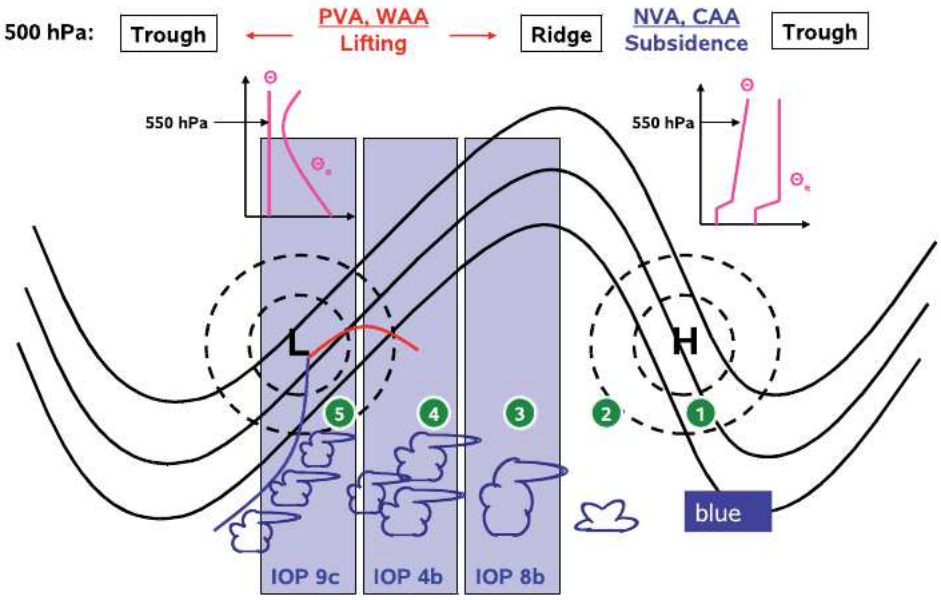



**Figure 3.5:** Convective initiation processes on the synoptic scale. The blue columns show the weather conditions during the COPS IOP 9c, 4b, and 8b (see Tab. 5.8). The black solid lines represent the 500*hPa* flow and the dashed lines the low and high surface pressure systems with warm front (red) and cold front (blue). The area with positive and negative vorticity advection is shown by (PVA, NVA) with the corresponding warm air and cold air advection (WAA, CAA). Additional typical vertical profiles of potential temperature ($\theta$) and equivalent potential temperature ($\theta_e$) downstream the trough and ridge is shown (Kottmeier et al., 2008).

forced convection. Figure 3.5 shows exemplary weather conditions in different synoptic situations from ① to ⑤. The weather conditions ranging from blue sky ① over single convective cells with ② and ③ and organised deep convection ④ to overcast sky with embedded convection ⑤ (Kottmeier et al., 2008). The three convection types will be explained in more detail in the following based on Fig. 3.5.

### Strongly Forced Convection (SFC)

This kind of convection is initiated by frontal forced events on the synoptic scale. The convection is mainly triggered by passing of fronts or convergence lines. See Fig. 3.5 area ④ and ⑤ for SFC. SFC includes organized convective events like squall lines. For the classification of the COPS-IOPs see Wulfmeyer et al. (2011).

### Weakly Forced Convection (WFC)

The convection is initiated here by non-frontal forced events at upper levels. This can be for example positive anomalies of potential vorticity associated with tropopause depressions

(Browning et al., 2008). The orography is more crucial as in the strongly forced case, and the convection is here more spatially distributed.

### Air-mass Convection (AMC)

Figure 3.5 area③ shows the synoptic weather situation for AMC. In contrast to the large-scale forced convection types, in this weather situations, the initial outbreak is triggered mainly by orography and by spatial variabilities on the small scale in

- temperature and humidity of the boundary layer
- land surface characteristics (land wetness)
- shadowing by clouds at higher levels

The occurrence of convection is here distributed on a wide area in contrast to SFC.

After Browning et al. (2008), these variabilities possibly act on scales too small to be resolved by numerical models. The importance of variations in boundary layer temperature and humidity for convective initiation is shown by Crook (1996). At synoptic scales, the area of convective initiation is most sensitive to the variability in temperature (Fabry, 2006). Keil et al. (2008) investigated the sensitivity of QPFs to variations in humidity in different layers. It was shown that an increased moistening in the boundary layer leads to an earlier initiation of convection and a higher amount of precipitation.

### Differentiation of convection-type

To distinguish the different kinds of convection from strongly forced to weakly forced convection, the convective time scale $\tau_c$ described in Keil and Craig (2011) can be used. This time scale describes the theoretical time period for the decomposition of the potential *CAPE* through $\frac{dCAPE}{dt}$. The convective time scale is defined by:

$$\tau_c = \frac{CAPE}{\frac{dCAPE}{dt}} \propto \frac{CAPE}{TOT\_PREC} \tag{3.12}$$

Where $\frac{dCAPE}{dt}$ is here estimated roughly speaking by the precipitation rate $TOT\_PREC$. In Keil and Craig (2011), $\tau_c$ of about $6h$ is defined to distinguish between forced and local/weakly forced convection:

| | |
|---|---|
| $\tau_c \geq 6h$ | Local-forced convection |
| $\tau_c < 6h$ | Forced convection |

However, Zimmer et al. (2011) shows that this choice is arbitrary. The threshold value range between 3 to 12 hours in the majority of cases, which were investigated in their study. Furthermore, it is shown that the convective time scale is a useful simplification for a numerical distinction of the predominant convection type.

### Current state of research

The last field campaign in Germany, which included the investigation of convective initiation, was the international field experiment convective and orographically-induced precipitation study (COPS) (Wulfmeyer et al., 2008). The goal of the COPS campaign in 2007 was to study the full life cycle of convective precipitation. COPS was embedded in the DFG PP1167 project with the aim to improve the orographically induced convective QPFs. Several hypotheses had been developed for COPS (Wulfmeyer et al., 2008):

- *Accurate modeling of orographically controlled convection is essential and only possible with advanced meso-scale models featuring a resolution of a few kilometers*
- *Location and timing of the initiation of convection critically depends on the structure of the humidity field in the planetary boundary layer*

The investigation of the measurement data is partly ongoing. One of the first results of the COPS campaign is from Kottmeier et al. (2008), Kalthoff et al. (2009) and Barthlott et al. (2009) the significant role of boundary layer convergence lines in relation to convective initiation. Wulfmeyer et al. (2011) underlines this result of initiation of convection by convergence lines and assess the model performance of convection-permitting models in this case as promising. The convection-permitting models are clearly more appropriate to forecast convection and they give also a better guideline for orographically influenced convection. It is further shown that the vegetation is also crucial for the sensible and latent heat flux.

In the next section the explicit simulation and predictability of convection in NWP models will be discussed.

## 3.3 Simulation of convection in NWP models

As mentioned before, upward vertical motion is the main reason for precipitation and therefore, it is very important to simulate this process correctly in NWP models. For convection, buoyancy is mainly responsible for this vertical motion. The complexity of convection and convective initiation was also explained in the previous section. In this section the numerical treatment of convection is shown. The strength of thunderstorms is also very critically dependent on the vertical velocity among others. The current state of NWP models in respect to their ability to simulate convection is shown. Therefore, the equations of the COSMO model are

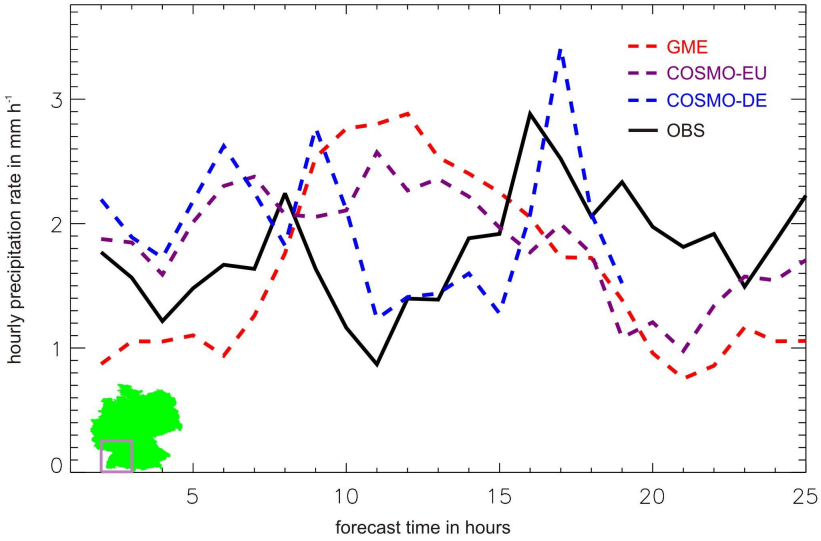



**Figure 3.6:** Mean diurnal cycle of 1 hour precipitation amounts in summer 2009. over south-western Germany. The NWP models GME (red), COSMO-EU (magenta) and COSMO-DE (blue) are compared with the observation of 25 SYNOP stations (solid black) (Baldauf et al., 2011).

shortly described in particular the equation for the vertical velocity. This description includes the approach of parametrization of convection.

### Current state of forecasts of convective precipitation

In addition to the COPS results, which were shown in the last section, verification results of the first few years since the operational launch of the convection-permitting COSMO-DE shows that NWP models still have problems with the diurnal cycle of CI and the corresponding precipitation. Baldauf et al. (2011) show in compliance to Wulfmeyer et al. (2011) that a convection-permitting NWP model is very well in the position to simulate this issue. Figure 3.6 shows the diurnal cycle of the hourly forecasted precipitation in the COPS region. The observed diurnal cycle of precipitation shows two maxima. The first maxima is shown between 7 and 8 UTC and the second between 15 and 16 UTC. It is shown that the global model GME with a resolution of $40km$ and the COSMO-EU with a resolution of $7km$ are not able to forecast this two maxima. Only the COSMO-DE is able to predict these two maxima. The reason for this is their coarser resolution and consequently the requirement to parametrize the convection. In contrast, the COSMO-DE as a convective-permitting model with a resolution of $2.8km$ is almost able to simulate the diurnal cycle, but with a small phase delay of about one hour (Baldauf et al., 2011).

As mentioned in the beginning of this chapter, for meso-scale NWP models with a grid resolutions about $\geq 10km$ the convection has to be parametrized. In case of parametrized convection, there is an unclean physical formulation when the convective cells themselves are of the order

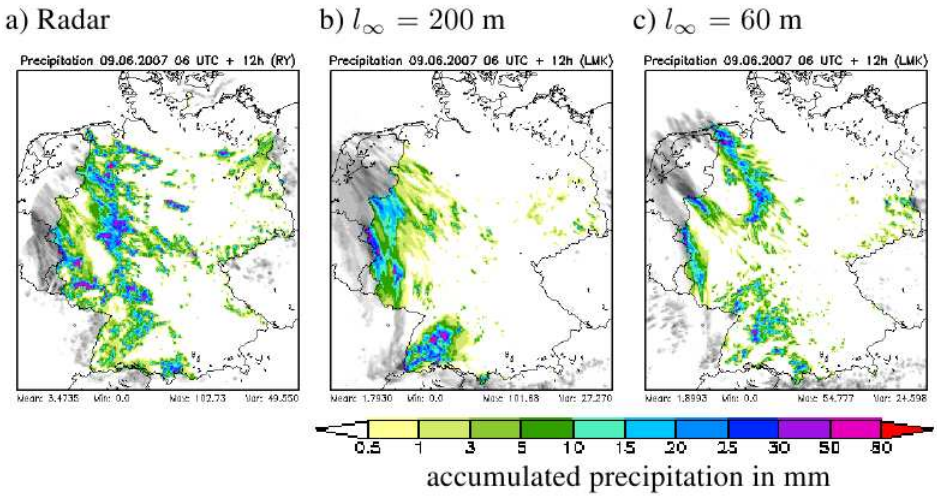



**Figure 3.7:** Accumulated surface precipitation for 9 June 2007, 6 UTC to 18 UTC. (a) shows the German radar composite, (b) 00 UTC + 18*h* COSMO-DE forecasts using the operational configuration of the PBL scheme and (c) using a reduced mixing length (Baldauf et al., 2011).

of the grid resolution and therefore partly resolvable (Kuell et al., 2007). Originally, convection schemes like Tiedtke (Tiedtke, 1989) were developed for global models with horizontal resolutions, which are too coarse for this effect. For deep convection, this problem is completely solved in a convective-permitting model where the convection is explicitly computed by the model. However, the convection has to be computed explicitly by the model. Here the triggering is one important challenge of the convective event. So, the NWP model has to simulate both, the convective-permitting conditions as well as the triggering of the convection.

The COSMO-DE has problems in forecasting moist convection especially in weakly forced situations one reason is a too stable boundary layer. This has been improved by the use of a shorter maximum turbulence mixing length $l_\infty$ (Baldauf et al., 2011). Results of this adjustment are shown in Fig. 3.7. Figure 3.7 a) shows the radar observations for the case study. The second figure (b) shows the precipitation forecast with the standard $l_\infty$ and finally, the third figure (c) shows the precipitation forecast for the reduced $l_\infty$. For $l_\infty = 60m$, the precipitation forecast is significantly improved. Particularly, the post-frontal precipitation structures are better captured.

For secondary convection the model is additionally sensitive to the cloud microphysics (Baldauf et al., 2011). Figure 3.8 shows the precipitation forecast by the usage of a more sophisticated cloud microphysic scheme (two-moment scheme). Figure 3.8 c) shows here slightly improvements for the precipitation forecast in contrast to the current operational one-moment scheme shown in figure (b).

However, in spite of all this improvements remains in the initiation of the convection as the great weakness of a deterministic NWP model. Baldauf et al. (2011) pointed out that an ensemble is needed to reasonably handle the uncertainties of the deterministic NWP model. An ensemble can capture almost all uncertainties which occur when convection is deterministi-

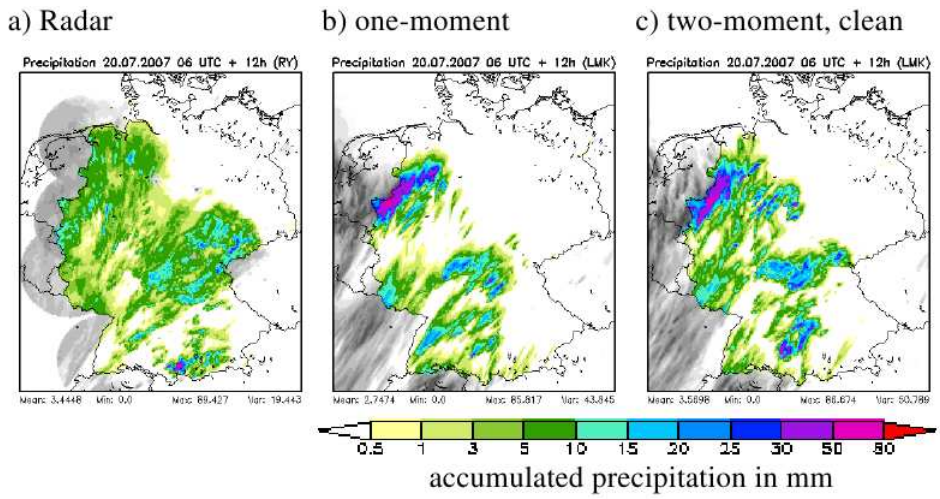



**Figure 3.8:** Accumulated surface precipitation for 20 July 2007 6 UTC to 18 UTC. (a) shows the German radar composite, (b) 00 UTC + 18*h* COSMO-DE forecasts using the one-moment scheme, and (c) the two-moment scheme (Baldauf et al., 2011).

cally simulated. More about the issue of predictability of convection is given in section 3.4. Nevertheless, the model development of the deterministic COSMO-DE is still going on. This means increased resolution, more model levels and advanced parametrization schemes especially for the turbulence in the boundary layer are indispensable to improve the deterministic NWP forecast as well as the EPS forecasts. Furthermore, both an improved deterministic NWP model and the ensemble approach complement each other.

## The governing equations of the COSMO model

The basic equations of the COSMO model and the approach of parametrization of sub-grid processes, e.g. like turbulence is introduced in the following. For a more detailed description see the COSMO documentation. For dynamics and numerics see Doms and Schättler (2002) and for the physical parametrization Doms and Förstner (2007).

The basic conservation laws of momentum, mass and heat of the COSMO model are represented by the following budget equations:

$$\rho\frac{d\mathbf{v}}{dt} = -\nabla p + \rho\mathbf{g} - 2\Omega\times(\rho\mathbf{v}) - \nabla\cdot\underline{\mathbf{t}} \tag{3.13}$$

$$\frac{d\rho}{dt} = -\rho\nabla\cdot\mathbf{v} \tag{3.14}$$

$$\rho\frac{dq_x}{dt} = -\nabla\cdot\mathbf{J}_x + I_x \tag{3.15}$$

$$\rho\frac{dE_{tot}}{dt} = -p\nabla\cdot\mathbf{v} - \nabla\cdot(\mathbf{J}_e + \mathbf{R}) + \varepsilon \tag{3.16}$$

The index $x$ is used here for

- dry air [$d$],
- water vapor [$v$],
- liquid water [$l$]
- and for water in the solid state [$f$], e.g. ice.

Eq. (3.13) describes the motion equation. The other equations are the continuity equation (3.14), the transport equation (3.15) and the energy equation (3.16) for the total specific internal energy $E_{tot}$. This equation system formulates the basic dynamic equation system for the atmospheric flow.

In the following, the focus is on the equation of the vertical motion $\frac{dw}{dt}$, which is contained in Eq. (3.13) where $\mathbf{v} = (u,v,w)^T$. More about the variables (see also the list of physical symbols, Appendix B), the other equations, the explicit formulation in rotated, terrain-following coordinates and the numerical discretization can be found in Doms and Schättler (2002). The state variables $\psi$ are separated into a base-state value $\psi_0$ and a deviation value $\psi'$. Finally, the linearized, non hydrostatic equation of the vertical velocity $w$ in the COSMO model is given by

$$\frac{dw}{dt} = -\frac{1}{\rho}\frac{\partial p'}{\partial z} + B + M_w \tag{3.17}$$

The vertical acceleration of an air parcel $\frac{dw}{dt}$ is due to three different forces: pressure gradient $\frac{\partial p'}{\partial z}$, buoyancy $B$ and subgrid effects like moist convection. The subgrid effects are denoted by $M_w$. Therefore, the buoyancy is given by

$$B = g\frac{\rho_0}{\rho}\left\{\frac{T'}{T} - \frac{T_0 p'}{T p_0} + \left(\frac{R_v}{R_d} - 1\right) q_v - q_l - q_f\right\} \tag{3.18}$$

A detailed list of symbols here are shown in Appendix B too. For a NWP model at the mesoscale, moist convection is the dominating subgrid scale process, which has to be parametrized. Thus, the subgridscale term $M_\psi$, which represent the to be parametrized processes contains the following terms for a prognostic model variable $\psi$:

$$M_\psi = M_\psi^{TD} + M_\psi^{MC} + M_\psi^{LB} + M_\psi^{CM} + M_\psi^{RD}, \tag{3.19}$$

where the individual terms have the following meaning (Doms and Schättler, 2002):

| Mesh-size | COSMO-EU $7km$ | COSMO-DE $2.8km$ | |
|---|---|---|---|
| $M_w^{TD}$ | $\neq 0$ | $\neq 0$ | 3-D TKE-based prognostic closure |
| $M_T^{TD}$ | $\neq 0$ | $\neq 0$ | |
| $M_{qv}^{TD}$ | $\neq 0$ | $\neq 0$ | |
| $M_w^{MC}$ | $= 0$ | $= 0$ | (Tiedtke, 1989) |
| $M_T^{MC}$ | $\neq 0$ | $= 0$ | |
| $M_{qv}^{MC}$ | $\neq 0$ | $= 0$ | |

**Table 3.2:** Overview of the single tendencies in the operational COSMO models after Doms and Förstner (2007). The subgrid scale terms for moist convection (MC) and for turbulent mixing (TD) are shown.

| | |
|---|---|
| $M_\psi^{TD}$ | tendency due to small scale turbulent mixing, |
| $M_\psi^{MC}$ | tendency due to subgrid scale moist convection, |
| $M_\psi^{LB}$ | lateral boundary relaxation term for one-way nesting of the model, |
| $M_\psi^{CM}$ | source term representing computational mixing, |
| $M_\psi^{RD}$ | tendency of $\psi$ due to a Rayleigh damping scheme applied within the upper boundary. |

The variables $\psi$ refers to all prognostic variables of the COSMO model. Table 3.2 shows a comparison of the COSMO-EU with the convection permitting COSMO-DE for the sub-grid scale terms for moist convection (MC) and for turbulent mixing (TD). Both are crucial for the prediction of convection. The turbulent mixing is important in the COSMO-DE as well as in the COSMO-EU. In contrast, the sub-grid scale terms for moist convection are only essential for the COSMO-EU. An overview of the sub-grid scale terms of turbulent mixing (TD) and moist convection (MC) is given in Tab. 3.2.

The parameterization of convection is explained in more details in the following. The Tiedtke scheme as well as the Kain-Fritsch scheme both will be introduced below, because they are used in the investigated ensembles. This ensembles are the COSMO-SREPS and the COSMO-LEPS. A detailed description of this ensembles can be found in Chapter 5.1.

### The parameterization of convection

The most common parametrization schemes used to simulate the sub-grid scale processes of convection in NWP models are summarized for an overview in the following. An approach developed originally for global models is the Tiedtke scheme (Tiedtke, 1989) used in the COSMO-EU. A later developed scheme is the Kain-Fritsch (Kain and Fritsch, 1990). Both schemes are used in the COSMO-SREPS as well as in the COSMO-LEPS. One of the latest schemes is the Bechtold scheme (Bechtold et al., 2001) based on the Kain-Fritsch scheme.

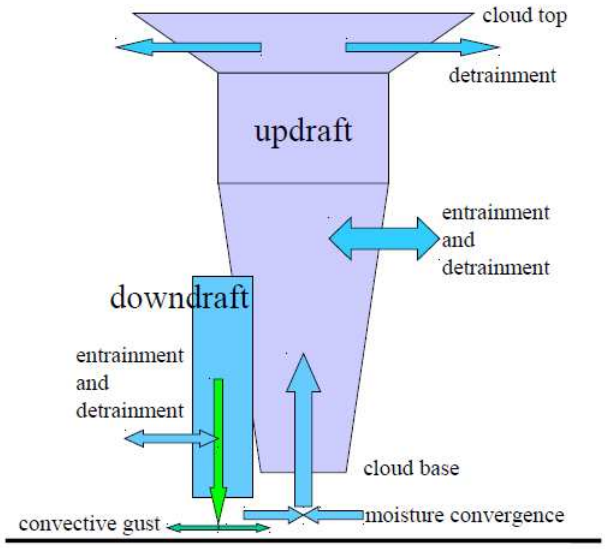



**Figure 3.9:** Schematic figure of the Tiedke scheme with up and downdraft (Majewski, 2008).

The current research includes also stochastic schemes. Plant and Craig (2008) introduce such a stochastic parametrization scheme, which based on the deterministic Kain-Fritsch scheme. In this scheme the plumes are treated in stochastic manner meaning the convective plumes are drawn from a PDF, which describes the probability of appearance of a plume of a given size within a given grid cell (Plant and Craig, 2008). However, in this section only the Tiedtke and the Kain-Fritsch scheme are introduced in more detail, because only they are used in the investigated models.

### a. The Tiedtke scheme

The convection parametrization scheme according to Tiedtke (1989) is a mass-flux approach based on the moisture convergence below the cloud base as closure assumption to represent the moist convection in the model. A schematic figure of the Tiedke scheme is shown in Fig. 3.9.

The feedback of sub-grid scale vertical fluxes of mass, heat, moisture and momentum in up- and downdrafts is calculated using a simple bulk cloud model. While the mass-flux in the updraft is determined by the moisture convergence below the cloud base, the mass-flux in the downdraft is proportional to the mass-flux in the updraft.

A necessary assumption for the Tiedtke scheme is that the horizontal area for averaging has to be large enough to contain an ensemble of cumulus clouds. However, the convective-scale eddy transport of dry static energy, moisture and momentum from cumulus updrafts, downdrafts and the cumulus-induced subsidence in the environmental air are not described in terms of contributions from the single ensemble components, but they are represented by their averaged values using an one-dimensional bulk cloud model after Yanai et al. (1973). This approximates

the net effects of an ensemble of clouds as resulting from a representative single cloud. For a more detailed description of the physics and the numerics see Tiedtke (1989); Keil (2000) and Doms and Förstner (2007). Emanuel (1994) criticise the inadequate theoretical formulation of the Tiedtke scheme.

### b. The Kain-Fritsch Scheme

The Kain-Fritsch scheme (KF) after Kain and Fritsch (1990) is another mass-flux scheme. In contrast to the Tiedtke scheme, the closure assumption of the Kain-Fritsch scheme is based on CAPE. In case of convection in the model, the KF scheme rearranges mass in a column using updraft and downdraft, and environmental mass-fluxes until at least 90% of CAPE is removed. The details of the Kain-Fritsch scheme are described in Kain and Fritsch (1990, 1993) and (Kain, 2003).

A comparison of both schemes is published in Dierer and Schubig (2008). In both schemes, a parcel is lifted level by level to test if positive buoyancy is present. In the Tiedtke scheme the convection is triggered when the parcel temperature exceeds the environment temperature by a fixed temperature threshold of $0.5K$. The temperature threshold in the KF scheme depends additionally on the large-scale vertical velocity, because large scale vertical upward motion favours convective developments.

## 3.4 Predictability of convection

In the last sections the fundamental physical processes, which are responsible for convection and the deterministic approach to simulate convection in NWP models were described. Baldauf et al. (2011) shows that convection permitting models like the COSMO-DE can provide improved forecast guidance about location, timing and strength of deep convection and improve precipitation forecasts compared to coarser NWP models that apply a parametrization of deep convection. Baldauf et al. (2011) also denote that individual convective cells are hardly predictable beyond a time scale determined by the life time of single cells of about $0.5 - 1h$.

To deal with this uncertainties of deterministic NWP models, the COSMO-DE model is being adapted to a convective-permitting ensemble named COSMO-DE-EPS at the DWD. It seems to be nearly impossible to forecast the right location of single convective cells by the approach of increasing the resolution of NWP models (Browning et al., 2008). Accordingly to Hense et al. (2003), there are three principle points which limit the predictive options of NWP models, especially for forecast of convection:

### I. Resolution of NWP models

The convective initiation is triggered by the spatial variability on the small scales in the temperature and humidity fields in the boundary layer. This is essential, especially in case of very weakly forced weather situations.

### II. Imperfect NWP models

Models like the COSMO-DE have to simulate all processes correctly, which are responsible for convection including the convective initiation. On the convective-scale a suitable parametrization of the boundary layer is absolutely crucial for the success of the forecast, especially for the explicit initiation of deep convection. But this is still an area of research and till now, the NWP models have significant weaknesses in handling the boundary layer, because large parts of the theoretical basis of boundary layer parametrization is still based on simplifications not valid at COSMO-DE resolutions.

### III. Chaotic nature of NWP models

After Lorenz (1963), it is impossible to predict exactly the state of the atmosphere beyond certain time scales, owing to the chaotic nature of the fluid dynamic equations and the imperfect knowledge about the inital state of the NWP model. This hold even if the NWP model is assumed to be perfect.

The existing observation networks have limited spatial and temporal resolution. Especially over the Oceans, which introduces uncertainty into the true initial state of the atmosphere. DelSole (2005) summarizes that classical deterministic models are perfectly predictable (no lost of predictability with lead time) if both, the initial condition and dynamical model are known perfectly, but not otherwise. Since the initial conditions and the dynamical model for NWP are not perfect, ensemble prediction systems (EPS) are used to capture the uncertainty of the forecast. An EPS involves multiple forecasts created with different model systems, different physical parametrization, or varying of initial state and for the local models varying of the boundary conditions too.

As an example of the limits of predictability of a chaotic model, the Lorenz attractor is shortly introduced in the following. For a more detailed description of the Lorenz attractor and his reference to NWP and climate models see Lorenz (1963); Buizza (2001) and Anwender (2007).

### The Lorenz attractor

The Lorenz attractor (Lorenz, 1963) is considered to be a perfect model. The key point is here the sensitivity to the initial state. So, the model is a demonstrative toy-model of the flow

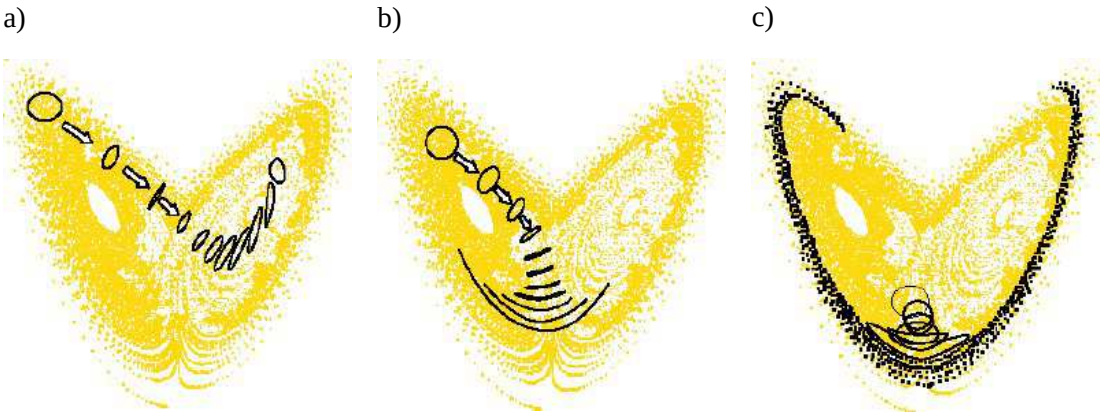



**Figure 3.10:** The two-dimensional Lorenz attractor for different initial conditions illustrated by the black cycle. The figure (a) shows a highly predictable scenario, (b) a reasonably predictable scenario and (c) a scenario where the predictability is totally lost (Buizza, 2001).

development depending on the initial state. The development in time and space of a three-dimensional non-linear system with three variables in phase space [2] $x_1, x_2, x_3$ is given by

$$\begin{aligned} \frac{dx_1}{dt} &= \sigma(x_1 - x_2) \\ \frac{dx_2}{dt} &= rx_1 - x_2 - x_1x_2 \\ \frac{dx_3}{dt} &= x_1x_2 - bx_3 \end{aligned} \tag{3.20}$$

The time development of these three variables of the model is determined by the time derivations of the variables. The parameters of the equation system $\sigma$, $r$ and $b$ are kept constant (Lorenz, 1963). However, the parameters can be changed to get a gathering of solutions. The phase space has the dimensions of the three independent variables $x_1, x_2$ and $x_3$ of Eq. (3.20).

Figure 3.10 shows two dimensions $x_1$ and $x_2$ of the phase space. In this phase space, the solutions of Eq. (3.20) are trajectories. An attractor characterizes how a dynamical system envolves over time. The black circles in Fig. 3.10 can be regarded for example as forecasts for different initial states of the atmosphere. The points in the first circle in Fig. 3.10 a), (b) and (c) are created by adding small variations to points on the attractor. Furthermore, the points in the circle can be imagined by probability density function (PDF) of the atmosphere state in the phase space. In case of the Lorenz attractor, the two regimes of the attractor can be considered as two different weather regimes (Buizza, 2001).

The first Fig. 3.10 a) shows the case of a highly predictable situation. The second picture (b) of a quite predictable situation and the third (c) for total loss of predictability. If the small perturbations are added in regions that are stable, the points are close together also even after a

[2] The phase space represents all possible instantaneous states of a system. Every trajectory in phase space is defined by a certain initial state and describes the time evolution of a single state of this system (Lorenz, 1963).

longer forecast lead time. Small perturbations lead only to small deviations within an increase of the forecast lead time. All points in Fig. 3.10 a) describe one regime only including a regime change. The case where the points are added in a less stable region is shown in Fig. 3.10 b). In this situation, in the beginning the points are close together, which is the same behavior as before in the stable region and indicates good predictability. Later, the regime change could be forecasted only with a certain probability. In the very unstable regions (Fig. 3.10 c) the predictability is given only for a very short time at the beginning. For longer lead times, the solutions evolve into completely different states.

The Lorenz attractor shows an example for limits of predictability of a simple model, which can be considered as perfect. In comparison to the Lorenz model, the NWP models are much more complex and they are far away to be perfect (especially the parametrizations), and it is impossible to know exactly the initial state of the atmosphere. This example shows that NWP models can predict the atmospheric state only with a certain probability for a certain time. After this time, the predictability is diminished as described in Chapter 2.

## 3.5 Issue of verification of forecasted vertical profiles

In this chapter, the importance of vertical temperature and humidity profiles and their impact on convective permitting conditions were shown. The coarse resolution, the imperfection of NWP models and the chaotic nature of the atmosphere limits the predictability of convective conditions.

Nevertheless, precise forecasts of the vertical structure and therefore, of the convective conditions are crucial, but not sufficient to simulate convection. The triggering of convection is necessary to initiate the convection. This issue is not part of this work. Here, only the aspect of the quality of the vertical conditions is investigated.

Two ensembles, one which explicit permits convection and the other one, which parametrizes convection will be compared in Chapter 6. This is done by the usage of three radiosonde stations to compare the forecasts of the vertical profile given the observations. The comparison aimed to investigate, which ensemble profiles are more likely and how strong is the evidence.

As discussed before in the introduction (Chapter 1) the radiosonde observations have uncertainties too. Thus, for this investigation, the uncertainties of NWP models and of observations are considered. This is possible, because the ensemble profiles are verified by a sophisticated probabilistic approach allowing in a natural way to consider the uncertainty of the observations as well as the forecast uncertainty. This Bayesian approach is introduced in the next chapter.

# Chapter 4

# Basic theory of Bayesian statistics

The Bayesian approach used in this work for comparison and verification of ensembles is described in this chapter. First, the statistical fundamentals are explained based on the frequency probability. After that, the extension to the Bayesian probability or Bayesian statistics is introduced. An more detailed introduction to statistical fundamentals for atmospheric sciences can be found e.g. in Wilks (1995).

## 4.1 Statistical Fundamentals

Statistical methods are essential for a quantitative analysis of large data sets. In general, this methods are applied to describe complex systems too e.g. in the quantum mechanics where the whole state of the system can be described only statistically. In meteorology, statistical methods are applied in a variety of ways including the evaluation of probabilistic forecasts.

In general, the probability $P$ of an event $A$ is denoted by $P(A)$ meaning in the context of frequency statistics that the number of times $n(A)$ that an event $A$ occurs in an experiment or a study normalized by the total number of cases $N$ is given by

$$P(A) = \lim_{N \to \infty} \frac{n(A)}{N} \tag{4.1}$$

Furthermore, the set $\mathcal{S}$ describes the sample space in which the event space $A$ is a subset of $\mathcal{S}$ ($A \subseteq \mathcal{S}$). The fundamental axioms for the probability theory are described by the Kolmogorov axioms (Kolmogorov, 1933):

1. $0 \leq P(A) \leq 1$
2. $P(\mathcal{S}) = 1$
3. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

The first axiom describes the probability of an event $A$ as a non-negative real number between zero and one. The second axiom shows that the probability of the sample space $\mathcal{S}$ is one. This means that there is no event outside the sample space. Finally, the third axiom describes the probability of the union of the single events $A$ and $B$. This probability is determined as the sum of the single probabilities of $A$ and $B$ minus the probability of the intersection of the single event spaces $A$ and $B$ denoted by $P(A \cap B)$. For the case that the events $A$ and $B$ are independent, the probability $P(A \cap B)$ will be the product of the single probabilities $P(A \cap B) = P(A)P(B)$. In the other case that the events are dependent, the multiplicative law of probabilities $P(A \cap B)$ is given by

$$P(A \cap B) = P(A|B)\,P(B) = P(B|A)\,P(A) \tag{4.2}$$

In which, the probability of $A$ given $B$ is called conditional probability $P(A|B)$. Furthermore, if the events $A$ and $B$ exclude each other the probability $P(A \cap B) = 0$ is zero. In the following, the statistical basics are described using continuous probability density functions (PDFs). The parameters of the PDF are estimated from the event sample.

### Continuous random variables

A continuous random variable is defined by $X = \{(x, g(x));\, x \in [a,b]\}$[1]. Then the PDF $g$ describes the theoretical distribution of the respective random variable, and the probability $P(A)$ is given by the integral of the PDF $g$ over the event space $A$.

$$P(A) = \int_A g(x)\,dx \tag{4.3}$$

The normal distribution is described in the following in more detail as a very important example of a PDF. The next function described here is the cumulative distribution function (CDF) of a PDF. The CDF describes the probability of an event $A$ to fall into the continuous interval $]-\infty, t]$. Thus the CDF $F(t)$ is the probability or frequency of occurrence of values less than or equal to $t$.

$$F(t) = \int_{-\infty}^{t} g(x)\,dx \tag{4.4}$$

The CDF is a monotone increasing (not necessarily strictly) function with a value range between 0 and 1 defined mathematically by

[1] Here, every event is associated with a real number, and the CDF is a continuously differentiable function: $g(x) = F'(x)$. Furthermore, the event space $A$ is defined as $A = \{x;\, x \in [a,b]\}$.

| | Name | Mathematical formulation | Estimator |
|---|---|---|---|
| 1st Moment | Expected value | $\mu = E[X]$ | $\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$ |
| 2nd Moment | Variance | $\sigma^2 = E[(X-\mu)^2]$ | $\sigma^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i-\mu)^2$ |

**Table 4.1:** Moments of a distributions and their corresponding Maximum-Likelihood Estimation (MLE) estimators.

$$0 \leq F(a) \leq F(b) \leq 1 \tag{4.5}$$

$$\lim_{t\to-\infty} F(t) = 0 \tag{4.6}$$

$$\lim_{t\to+\infty} F(t) = 1 \tag{4.7}$$

### Moments of distributions and further statistical indices

The expectation value of a continuous random variable $X$ is represented by $E[X]$ and the variance by $E[(X-\mu)^2]$. In statistics, a random variable can be described by one of its several moments too. The k-th moment of a random variable $X$ is denoted by

$$m_k = E[X] = \int x^k g(x)\,dx \tag{4.8}$$

and the k-th central moment as

$$\mu_k = E[(X-\mu)^k] = \int (x-m_1)^k g(x)\,dx \tag{4.9}$$

The first moment $m_1$ is the expected value, and the second central moment $\mu_2$ is the variance. Table 4.1 show an estimator for the expectation value and the variance of a given sample.

Furthermore, the quantile of a continuous distribution is defined by a point taken at a regular interval from the CDF of a random variable by

$$F(q_\tau) = \tau \tag{4.10}$$

The number $\tau$ describes the $\tau$-quantile of a random variable $X$. Finally, the median is defined as 0.5-quantile ($\tau = 0.5$).

### The normal distribution

As mentioned before, in probability theory, the normal distribution, which is also called Gaussian distribution, is one of several possible PDFs to describe sets of data. This PDF is quite

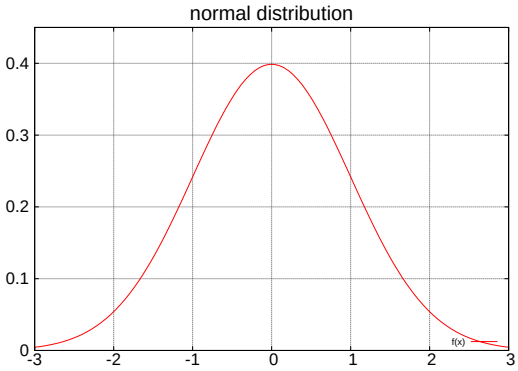



**Figure 4.1:** Normal distribution $N(\mu,\sigma^2)$ with the mean $\mu = 0.0$ and the variance $\sigma^2 = 1$.

simple and often applicable. In physics and meteorology, the normal distribution is often used as an approximation of the distribution for random variables like the averaged 2*m*-temperature in climate simulations. In statistics, a very important subject is the central limit theorem (CLT), which states that the mean of a sufficiently large set of independent random variables described by a wide variety of probability densities with a finite second moment can be assumed as normally distributed. This assumption becomes exactly true for $\lim_{N\to\infty}$. Wilks (1995) mentioned that this assumption is frequently valid already for $N \geq 30$. This illustrates the special place of the normal distribution in statistic.

In the following, first, the univariate normal distribution is described and then the multivariate normal distribution. The univariate normal distribution is defined by

$$g(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right) \tag{4.11}$$

Its described by only two parameters, the mean $\mu$ and the variance $\sigma^2$. Figure 4.1 shows a normal distribution with mean $\mu = 0$ and variance $\sigma^2 = 1$. $\mu$ and $\sigma^2$ can be estimated from the sample space in which the estimators for the mean and for the variance are defined by the maximum likelihood estimation (MLE) shown in Tab. 4.1 for the univariate case. A normal distributed random variable $X$ is denoted here as $X \sim N(\mu,\sigma)$.

### Multivariate normal distribution

The univariate normal distribution of a random variable has only one dimension. In the multivariate case, the random variable is described by a vector containing several "one dimensional" random variables, which could be additionally correlated. This means that the multivariate case treats several variables simultaneously. The multivariate random variable $X$ is here a $q$-dimensional vector of random variables $X = (x_1,\cdots,x_q)^T$ and it is called normal distributed if the density is given by

$$g(\vec{x}) = \frac{1}{\sqrt{(2\pi)^q |\Sigma|}} \exp\left(-\frac{1}{2}(\vec{x}-\vec{\mu}))^T \Sigma^{-1}(\vec{x}-\vec{\mu})\right) \tag{4.12}$$

In the multivariate case, the Gaussian distribution is determined by the mean vector $\vec{\mu}$ and the covariance matrix $\Sigma$. The covariance matrix contains the variances $\sigma^2$ of the univariate cases in the diagonal elements of the matrix. The calculation of $\Sigma$ is introduced in the following including the non-diagonal elements of the matrix, representing the correlations between the different variables of $X$. In general, the covariance matrix is defined as

$$\Sigma = E[(X-\vec{\mu})(X-\vec{\mu})^T] \tag{4.13}$$

The estimation and properties of the covariance matrix $\Sigma$ are introduced in the following section.

## 4.2 Maximum likelihood estimation (MLE)

In statistics, MLE is a method to estimate statistical parameters. The log-likelihood $L$ for a $N$ multivariate normal distributed sample $(\vec{x_1},...,\vec{x_N})$ of dimension $q$ represented by the Gaussian PDF $g(\vec{x})$ is shown in Ueno and Tsuchiya (2009). With the arithmetic average as MLE of the expectation value, $L$ can be written as

$$\begin{aligned} L &= log(\,g(\vec{x})\,) && (4.14)\\ &= -\frac{K}{2}\left[\log(2\pi) - \log(\det\Sigma^{-1}) + \mathrm{trace}(S\Sigma^{-1})\right] && (4.15)\end{aligned}$$

This is the basis for the MLE of the covariance matrix and finally, leads to the standard MLE estimator $S$ of the covariance matrix $\Sigma$

$$S = [\sigma_{ij}^2]_{q\times q} = \begin{pmatrix} \sigma_{11}^2 & \cdots & \sigma_{1q}^2 \\ \vdots & \ddots & \vdots \\ \sigma_{q1}^2 & \cdots & \sigma_{qq}^2 \end{pmatrix} \tag{4.16}$$

$$\sigma_{ij}^2 = \frac{1}{N-1}\sum_{k=1}^{N}(x_{ik}-\bar{x_i})(x_{jk}-\bar{x_j}) \tag{4.17}$$

The covariance matrix has certain properties. The covariance matrix

- is **symmetric**
  $\Leftrightarrow \Sigma = \Sigma^T$

- is **positive definite** with only positive eigenvalues $\lambda_k$
  $\Leftrightarrow \forall\lambda_k > 0 \Leftrightarrow det(\Sigma) > 0$
- has **real eigenvalues** $\lambda$ and **eigenvectors** $\hat{e}$ defined by
  $\Rightarrow \Sigma\hat{e} = \lambda\hat{e}$
  $\Rightarrow W = P^T\Sigma P = diag(\lambda_1,\cdots,\lambda_q) \quad with \quad P = [\hat{e}_1,\cdots,\hat{e}_q]$
  $\Rightarrow \Sigma = PWP^T$

To complete the topic of statistical fundamentals of multivariate random variables, the correlation between two random variables $X_i$ and $X_j$ is given by

$$\rho_{ij} = \frac{\sigma_{ij}^2}{\sqrt{\sigma_{ii}^2\,\sigma_{jj}^2}} \tag{4.18}$$

The correlation term $\rho_{ij}$ has a range of $0 \leq \rho_{ij} \leq 1$. Wilks (1995) (p.368) shows the relation between the covariance matrix $R$ and the correlation matrix $\Sigma$ as

$$R = [\rho_{ij}]_{q\times q} \tag{4.19}$$

$$R = D^{-1}\Sigma D^{-1} \tag{4.20}$$

with

$$D^{-1} = \begin{pmatrix} \frac{1}{\sigma_{11}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{\sigma_{qq}} \end{pmatrix} \tag{4.21}$$

As example, Fig. 4.2 shows a two-dimensional normal distribution with $\rho_{12} = \rho_{21} = 0$.

## 4.3 Statistics of NWP-model forecasts

After introducing the general statistical fundamentals, the statistical application to NWP models will be described now. In the following, the model forecast of the state vector of dimension $q$ of a NWP-model is denoted with $f$. The difference between the true (but unknown) state vector $f_t$ and $f$ is described by the error $\varepsilon$ of the model.

$$\varepsilon = f - f_t \tag{4.22}$$

The statistics of the model error are fully described by the PDF $g(\varepsilon)$ (Bouttier and Courtier, 1999). Here it is assumed that the expectation of $\varepsilon$ is zero, meaning there is no systematic error of the model.

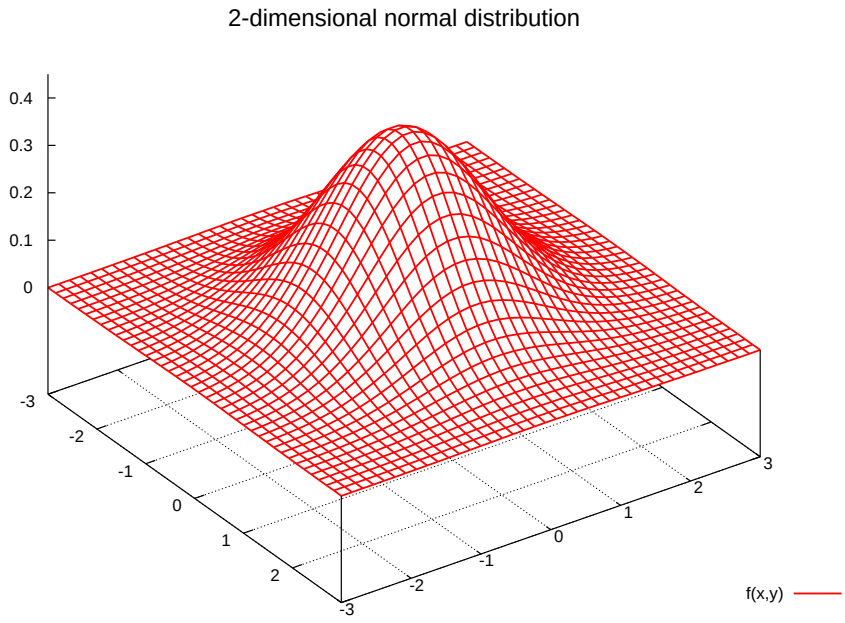



**Figure 4.2:** Two-dimensional normal distribution $N(\vec{\mu},\Sigma)$ with the means $\mu_1 = \mu_2 = 0$, the variances $\sigma^2_{1,1} = \sigma^2_{2,2} = 1$ and with the correlation coefficient $\rho_{12} = \rho_{21} = 0$.

$$E[f - f_t] = E[\varepsilon] = 0 \tag{4.23}$$

The ensemble can be understood as a Monte Carlo procedure to sample this PDF of ε. The ensemble consists of single members $f_k$ with $k = 1,..,K$ and the density $p(f)$. Following Schölzel and Hense (2010), this density can be described as the expectation of a sum of Dirac delta functions $\delta(f - f_k)$ regarding the density of the model error $g(\varepsilon)$ as

$$p(f) = E\left[\sum_{k=1}^{m} \delta(\varepsilon - (f - f_k))\right] \tag{4.24}$$

Furthermore, the expectation value of a function $h$ can be formulated as $E[h] = \int h \cdot g(\varepsilon) d\varepsilon$ for continuous random variables (Rade and Westergren, 2000). Here, the function $h$ is defined according to Eq. 4.24 as the Dirac delta function. This results finally in

$$p(f) = \int \sum_{k=1}^{m} \delta(\varepsilon - (f - f_k)) \cdot g(\varepsilon) d\varepsilon \tag{4.25}$$

$$= \frac{1}{m} \sum_{k=1}^{m} g(f - f_k) \tag{4.26}$$

showing the PDF of the ensemble forecast $p(f)$ can be formulated by a simple mixture model, which is also known as standard kernel dressing (SKD) in the univariate case see Bröcker and Smith (2008).

The standard MLE estimator of the covariance matrix was shown in the last section. Schölzel and Hense (2010) present a further method to calculate the covariance matrix without using of the arithmetic mean of the ensemble. This means, the covariance matrix can be described by the expectation value of all possible distances between all ensemble members $f_k$.

$$E[(f_k - f_{k'})(f_k - f_{k'})^T] = E[\varepsilon_k \varepsilon_k^T - \varepsilon_k \varepsilon_{k'}^T - \varepsilon_{k'} \varepsilon_k^T + \varepsilon_{k'} \varepsilon_{k'}^T] \tag{4.27}$$

$$= E[\varepsilon_k \varepsilon_k^T] + E[\varepsilon_{k'} \varepsilon_{k'}^T] \tag{4.28}$$

$$= 2E[\varepsilon_k \varepsilon_k^T] \tag{4.29}$$

$$= 2S \tag{4.30}$$

Schölzel and Hense (2010) use the assumption that the errors $\varepsilon_k$, $\varepsilon_{k'}$ are independent. The scaled difference between $f_k$ and $f_{k'}$ can be defined as $d_{kk'} = \frac{1}{\sqrt{2}}(f_k - f_{k'})$. Thus, the expectation value of $d_{kk'}$ results as the covariance matrix $S$.

$$E[d'_{kk} d_{kk'}^T] = S \tag{4.31}$$

Finally, the estimator $S$ of the error covariance matrix is given by the normalized average over all possible distances between $(f_k - f_{k'})$.

$$S = \frac{1}{2} \frac{1}{m(m-1)} \sum_{k}^{m} \sum_{k'}^{m} (f_k - f_{k'})(f_k - f_{k'})^T \tag{4.32}$$

In the following, the error covariance matrix is calculated via this way. In the next section, the extension of the classical statistic to Bayesian statistic is introduced. Bayesian statistics describe a convenient way to join all information and their inherent uncertainties (Berger, 1985). The approach on Bayesian statistics for forecast validation is used in this work, because the Bayesian statistics allows for a more extended consideration of probability than the classical statistics and an easy inclusion of the observational uncertainty. In the following the Bayesian statistics will be introduced.

## 4.4 Bayesian Statistics

In the previous section, the fundamentals of the classical frequency probability were introduced. Generally, the difference to the Bayesian probability is the different interpretation of

probability. The classical statistics interpret probability as a frequency of observed events. In contrast, the Bayesian statistics interprets this term as a "degree of plausibility" with the extended consideration on the given state of knowledge (Jaynes, 2003). More precisely, they differ mainly in the contemplation of the prior probability. The Bayesian statistic based on the Bayes theorem, which goes back to Thomas Bayes's original paper in 1763 and is given by

$$\boxed{P(H|D) = \frac{P(D|H)P(H)}{P(D)}} \tag{4.33}$$

The Bayesian approach "allows you to start with what you already believe (prior) and then see how new information changes your confidence in that belief (posterior). The Bayes theorem says simply that the probability $P$ of the hypothesis H, given the data D, is equal to the probability of the data, given the hypothesis is correct, multiplied by the probability of the hypothesis before obtaining the data, divided by the average probability of the data" (Malakoff, 1999).

In this work, the Bayesian approach is used to assess ensemble forecasts given observations. The complete forecast ensemble at lead time $\tau$ represent the hypothesis $H$. This is not applied to the full model state vector where the dimension $q$ is of orders $\mathrm{O(q)} \sim 10^8$. The aim of this Bayesian investigation is a multivariate verification of forecasts of vertical temperature and moisture profiles of several ensemble systems at a given set of radiosonde places and measurement heights. Partly this circumvents the problem of dimension reduction which will not be discussed here in detail and is referred to e.g. Jonko et al. (2009) or Hense and Römer (1995).

The multivariate structure is given by temperature and moisture at various levels. Because several ensembles are compared generated by EPS ($m_i$, $i = 1,2$) among each other or with the climatology $i = 0$ each EPS is considered as the realization of a discrete random variable $m_i = m_i(\tau)$ at forecast lead time $\tau$. Each of these ensembles is characterized by the prior probability $P(m_i)(\tau)$.

For verification the evidence of a specific ensemble $m_i(\tau)$ given the observational data $o$ is wanted to be found. This can be expressed as the conditional probability $P(m_i(\tau)|o)$, which is also called the posterior probability. The Bayes theorem relates the likelihood $l(o|m_i(\tau))$, the prior probability $P(m_i(\tau))$ and the posterior probability $P(m_i(\tau)|o)$ as:

$$P(m_i(\tau)|o) = \frac{l(o|m_i(\tau))P(m_i(\tau))}{y(o)} \tag{4.34}$$

with

$$y(o) = \sum_{j=1}^{N} l(o|m_j(\tau))P(m_j(\tau)) \tag{4.35}$$

The posterior shows the evidence of an ensemble in view of the data. The posterior evolves from the existing knowledge (the priors) and its modification through the likelihood of the observations (Min et al., 2004). The likelihood has to be further refined. For simplicity reasons

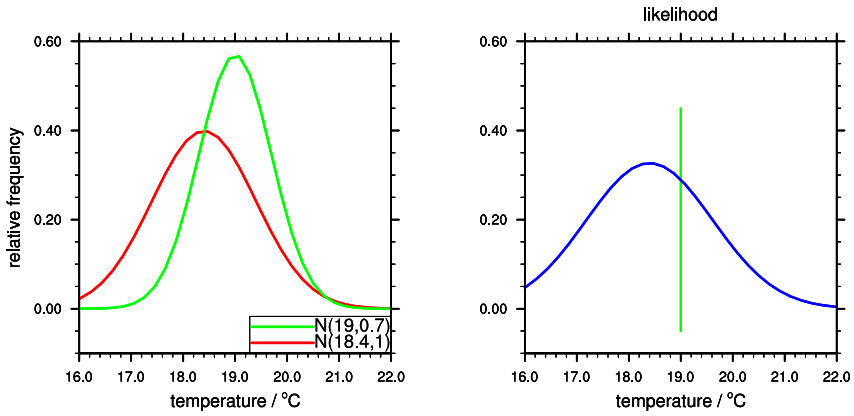



**Figure 4.3:** Univariate illustration of two Gaussian PDFs representing the model and the observation uncertainty shown by the left figure. The corresponding likelihood of this two PDFs is shown in the right figure.

the lead time $\tau$ is dropped in the following. Each ensemble $m_i$ is defined through the realizations $f_k^{(i)}$, were $k$ is the number of the respective ensemble member. Then the likelihood $l(o|m_i)$ is the integral over two PDFs. The first $p_l(o|f)$ describes the uncertainty of the observations and the second the uncertainty within the ensemble of the model $m_i$.

$$l(o|m_i) = \int p_l(o|f)p_l(f|m_i)df \tag{4.36}$$

Figure 4.3 shows a univariate example of two PDFs. In this, the likelihood is a measure of the agreement of this PDFs. The likelihood is going to zero in case when both PDFs are far apart or also when the PDFs have nearly the same mean, but one PDF has a very large standard deviation. Finally, the likelihood increases until both PDFs are nearly identical. In the following, the prior and posterior probability are discussed in more detail.

### The prior probability

The Bayesian statistic allows to combine personal believe (prior) with informations stem from data. The prior quantifies given knowledge about the forecasting system possibly in a subjective way e.g. assessed by a questionnaire among professional weather forecasters. In Jaynes (1968), there is a citation about what Laplace said to the selection of priors

> *"When the probability of a simple event is unknown, we may suppose all values between 0 and 1 as equally likely".*

In this work, an uniform prior (called Laplace prior) is used, which gives equal probability to each ensemble under investigation.

$$P(m_i) = \frac{1}{N_{eps}} \tag{4.37}$$

Furthermore, in this work always two ensembles $N_{eps} = 2$ are investigated. Additionally, the prior is systematically varied to investigate the significance of the prior to the posterior probability. A detailed discussion about the selection of prior distributions can be found e.g. in Kass and Wasserman (1996).

### The posterior probability

The posterior probability $P(m_i|o)$ can be denoted as marginal probability with respect to the forecasted state vector $f$.

$$P(m_i|o) = \int p(m_i, f|o)\, df \tag{4.38}$$

With the Bayes theorem (Eq. 4.33) the Eq. (4.38) can be written as

$$P(m_i|o) = \int p(o|m_i, f)\, p(m_i, f)\, df \cdot \frac{1}{p(o)} \tag{4.39}$$

Using the Bayes theorem once more $p(m_i, f) = p(f|m_i)p(m_i)$ the prior probability could be factored out.

$$P(m_i|o) = \int p(o|m_i, f)\, p(f|m_i)\, df \cdot \frac{P(m_i)}{p(o)} \tag{4.40}$$

Furthermore, the equation of $p(o|m_i, f) = p(o|f)$ can be simplified, because of independence with respect to the model $m_i$. The probability of observation can be expressed as $p(o) = \sum_{i=1}^{N} P(m_i) = 1$. The posterior probability is finally given as

$$P(m_i|o) = \frac{\int p_l(o|f)\, p_l(f|m_i)\, df \cdot P(m_i)}{\sum_{j=1}^{N} P(m_j|o)} \tag{4.41}$$

If for the errors in the observations an unbiased multivariate Gaussian distribution is assumed, the conditional probability $p_l(o|f)$ can be formulated as Eq. (4.42). Additionally, a multivariate Gaussian distribution $p_l(f|m_i)$ is assumed for the ensemble state shown in Eq. (4.43).

$$p_l(o|f) = \frac{1}{\sqrt{(2\pi)^q |\Sigma_o|}} \exp\left(-\frac{1}{2}(o - I(f))^T \Sigma_o^{-1} (o - I(f))\right) \tag{4.42}$$

$$p_l(f|m_i) = \frac{1}{K_i} \sum_{k=1}^{K_i} \frac{1}{\sqrt{(2\pi)^q |\Sigma_i|}} \exp\left(-\frac{1}{2}(f - f_k^{(i)})^T \Sigma_i^{-1} (f - f_k^{(i)})\right) \tag{4.43}$$

$\Sigma_o$ denotes the error covariance matrix of the observations and $\Sigma_i$ the error covariance matrix of the model $m_i$. $\Sigma_i$ is calculated by using the $K_i$ simulated vertical profiles, interpolated to the position of the radiosonde observation ($I$-operator) assumed to be linear $I(f) = I \cdot f$. For estimating the uncertainty in the ensemble the multivariate kernel dressing approach by Schölzel and Hense (2010) is used, which defines the predictive PDF for the state vector $f$ of the ensemble as a Gaussian mixture model with dressing covariance matrix $\Sigma_i$ summed over all $K_i$ realizations. The basic idea of this approach was shown in section 4.3.

Basically, the distribution of the ensemble realization is based on the SKD. Bröcker and Smith (2008) show the SKD for the univariate case based on Silverman (1986). In his work Silverman (1986) introduced additionally the usage of the so called Silverman's factor $h_S$ as a scaling factor for the raw covariance matrix. The covariance matrix is then $\Sigma_i = h_S \cdot S_{raw}$ with

$$h_S = \left( \frac{4}{K_i(q+2)} \right)^{\frac{1}{q+4}} \tag{4.44}$$

In which, $S_{raw}$ is the first approximation after Eq. (4.32) namely $S_{raw} = S$. This factor ensures that with increasing ensemble size $K_i$ the used dressing covariance matrix $\Sigma_{raw}$ becomes very small. The Silverman's factor $h_S$ is also called as the smoothing parameter. Larger dressing standard deviations by the factor leads also to a smoothing of the corresponding PDF shown by Fig. 4.5. An illustration of the SKD method and of the values of $h_S$ are shown in Fig. 4.4. In case of $h_S = 2$ a strong smoothing is shown losing the bimodal structure of the SKD PDF in Fig. 4.5 c). Figure 4.5 shows further that the likelihood is quite similar for $h_S = 1$ as well as for $h_S = 0.75$, which would be a typical value of $h_S$ for an ensemble with $K_i = 20$ and $q = 8$ (see Fig. 4.4 b). Thus in this work, a subjectively selected Silverman's factor of $h_S = 1$ is used. Nevertheless, a sensitive study to the Silverman's factor is shown in Chapter 6 to investigate the effect of $h_S$ for the comparison of two ensembles and to underline this subjective selection.

The issue of the observation uncertainty is mentioned several times before in this work mainly in the introduction (Chapter 1) and in the chapter describing the data of this work (Chapter 5). Figure 4.6 shows the sensitivity of the likelihood in view of the magnitude of the observation uncertainty represented by variation of the observation covariance $\tilde{\Sigma}_o = \gamma \cdot \Sigma_o$. It is shown that the variation of the observation error has a much stronger effect on the likelihood than the Silverman's factor, which was shown in Fig. 4.5. Due to this for the comparison of the ensembles (Chapter 6) and for the verification of the ensembles (Appendix A.3) a sensitivity study in respect to the observation uncertainty is shown in the respective chapters.

Finally, inserting Eqs. (4.42) and (4.43) into Eq.(4.36), and applying some linear algebra the integral in Eq. (4.36) can be evaluated analytically with the result

a) b)

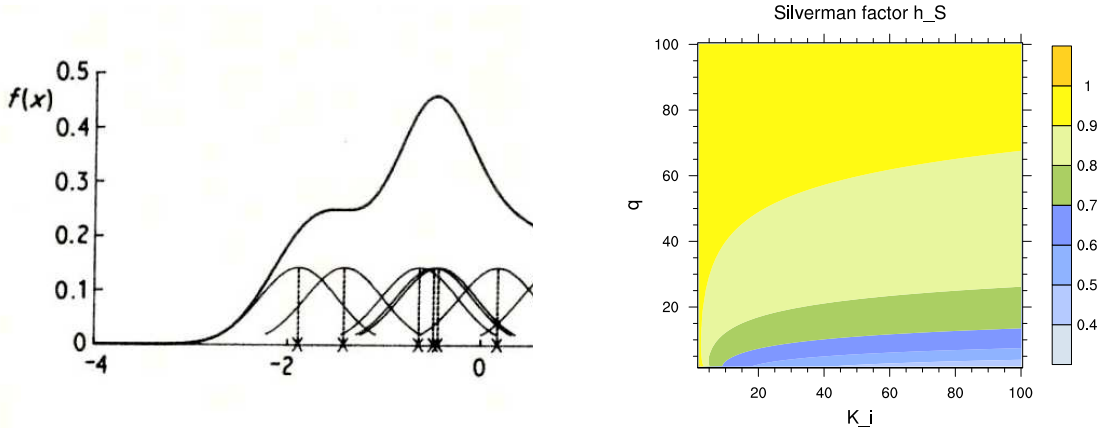



**Figure 4.4:** The left figure (a) illustrate the kernel estimation by SKD showing the individual kernels. The individual gaussian PDFs are shown with a smaller order of magnitude for a clear presentation (Silverman, 1986). Figure (b) shows the values of the Silverman's factor for different q-dimensional vectors and different ensemblesizes.

$$P(m_i|o) = \frac{1}{K_i}\sum_{k=1}^{K_i}\int \frac{1}{\sqrt{(2\pi)^{q+r}|\Sigma_i||\Sigma_o|}}\exp\left(-\frac{1}{2}(o-A_i^{-1}b_k^{(i)}))^T A_i (o-A_i^{-1}b_k^{(i)})\right) df \quad (4.45)$$
$$\cdot \exp\left(-\frac{1}{2}\Lambda_k^{(i)}\right)\frac{P(m_i)}{y(o)}$$

and the definitions

$$\begin{aligned} A_i &= \Sigma_o^{-1} + I\Sigma_i^{-1}I^T \\ b_k^{(i)} &= I\Sigma_o^{-1}o + \Sigma_i^{-1}f_k^{(i)} \\ \Lambda_k^{(i)} &= (o - I^T f_k^{(i)})^T (I^T\Sigma_i I + \Sigma_o)^{-1}(o - I^T f_k^{(i)}) \end{aligned} \quad (4.46)$$

The integral over all model realizations $f$ finally leads to the following expression:

$$\boxed{P(m_i|o) = \frac{1}{K_i}\sum_{k=1}^{K_i}\frac{\sqrt{|A_i^{-1}|}}{\sqrt{(2\pi)^q|\Sigma_i||\Sigma_o|}}\exp\left(-\frac{1}{2}\Lambda_k^{(i)}\right)\frac{P(m_i)}{r(o)}} \quad (4.47)$$

This shows that the posterior probability is a function of the Mahalanobis distances (MD) $\Lambda_k^{(i)}$ (Maesschalck et al., 2000; Mahalanobis, 1936) which describes the variance-weighted distance between the forecasted state vector $f_k^{(i)}$ and the observation $o$. The MD is invariant to nonsingular linear transformations of the state vectors $f_k^{(i)}$ and $o$ meaning that the final results are independent from the actual chosen basis (Sole and Tippett, 2007). Among other advantages this means that one can compare e.g. different variable types or variables with largely different

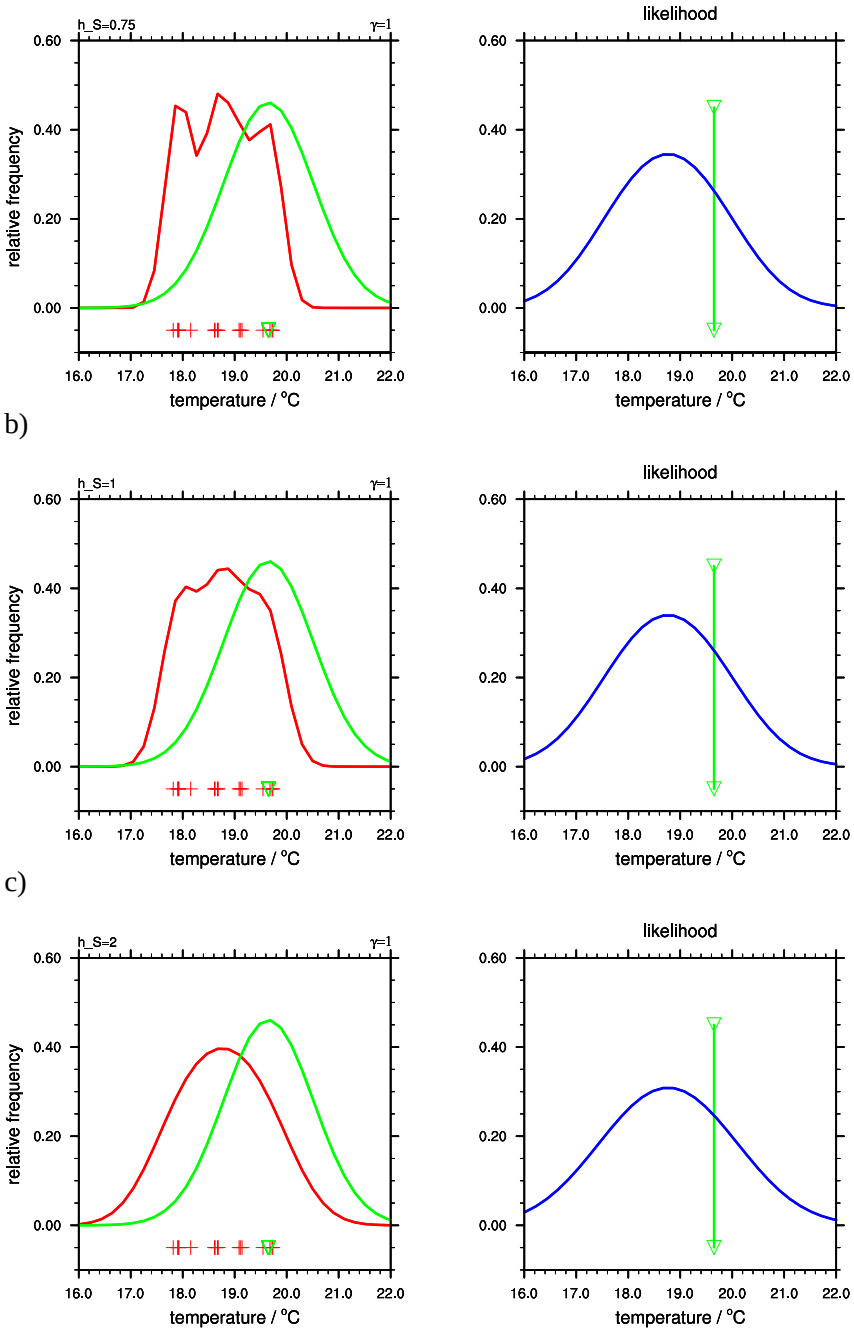



**Figure 4.5:** Univariate temperature PDF of one level (850hPa) at one gridpoint of the 6 hour SREPS forecast at 15. July 2007 (red line) and the corresponding observation PDF (green line) are shown. The raw ensemble forecasts are marked by '+' and the observation by '▼'. On the left the construction of the kernel dressing PDF is shown and on the right the likelihood function of the observation PDF and the kernel dressing PDF. The vertical green line shows the point at which the likelihood is evaluated. All this is shown for different Silverman's factors $h_S = 0.75$ (a), $h_S = 1.0$ (b) and $h_S = 2.0$ (c).

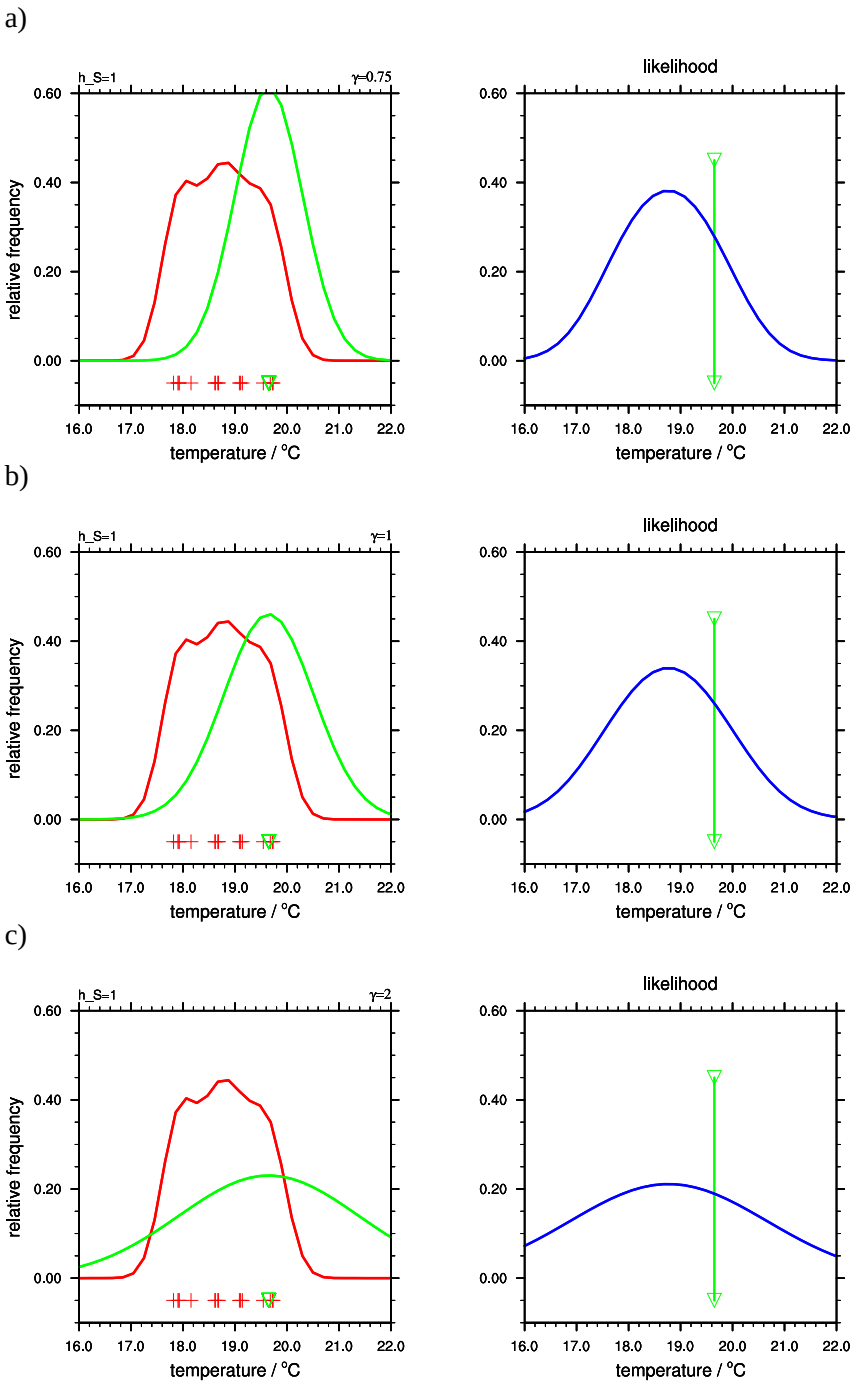



**Figure 4.6:** Univariate temperature PDF of one level (850hPa) at one gridpoint of the 6 hour SREPS forecast at 15. July 2007 (red line) and the corresponding observation PDF (green line) are shown. The raw ensemble forecasts are marked by '+' and the observation by '▼'. The figure is equal to Fig. 4.5, but here the sensitivity to the observation error is shown for different $\gamma$ factors $\gamma = 0.75$ (a), $\gamma = 1.0$ (b) and $\gamma = 2.0$ (c).

| $logB_{ij}$ | Evidence for ensemble | Validation against analysis |
|---|---|---|
| $> 5$ | Decisive for $m_i$ | n.a. |
| 2.5 - 5 | Strong for $m_i$ | n.a. |
| 1 - 2.5 | Substantial for $m_i$ | n.a. |
| $-1$ - 1 | Neutral | High level of confidence |
| $-2.5$ - $-1$ | Substantial for $m_j$ | Medium level |
| $-5$ - $-2.5$ | Strong for $m_j$ | Low level |
| $< -5$ | Decisive for $m_j$ | Very low level |

**Table 4.2:** Descriptive scales of the Bayes factor for the comparison of two ensembles after Kass and Raftery (1995). The evidence for an ensemble is used in case of comparison of two ensembles. For verification the validation column is used which is also a comparison, but against an NWP model analysis.

variability ranges as long as the errors are realizations of Gaussian distributed random variables. Additionally, the advantage of the posterior probability is the explicit inclusion of the uncertainty of the forecast ensemble and the observations. This has been not often considered in verification studies.

The ratio of the posterior from the model $m_i$ to a reference model $m_r$ can be used to compare two ensembles with each other. This ratio is called Bayes factor and is further discussed in the next section. As reference model a specific deterministic model can be defined (Chapter 7) or a specific ensemble (Chapter 6). In the deterministic case the deterministic model is considered as the mean of an artificial one member ensemble, but still including the uncertainty.

### Bayes factor

The Bayes factor characterizing the relative performance of two ensembles $m_i$ and $m_r$ or the performance of $m_i$ relative to an analysis. The Bayes factor is defined as the ratio of the posterior probabilities

$$\boxed{B_{ir} = \frac{P(m_i|o)}{P(m_r|o)}} \tag{4.48}$$

Using of the Bayes factor has the advantage that the marginal probability of the data $y(o)$ cancels out. The Bayes factor can be used to decide which ensemble is more likely with respect to the posterior probability. Gneiting and Raftery (2007) show that the logarithm of the Bayes factor $\log B_{ir}$ is proportional to the so called Ignorance score, which has the important properness feature. Therefore from now on the *log* of the Bayes factor is discussed.

In case of an analysis as reference model $r$ a *log* Bayes factor near zero means a nearly perfect forecast, because the analysis is considered as an approximation of the truth under a given uncertainty. If comparing two ensembles, a *log* Bayes factor greater than zero describes the

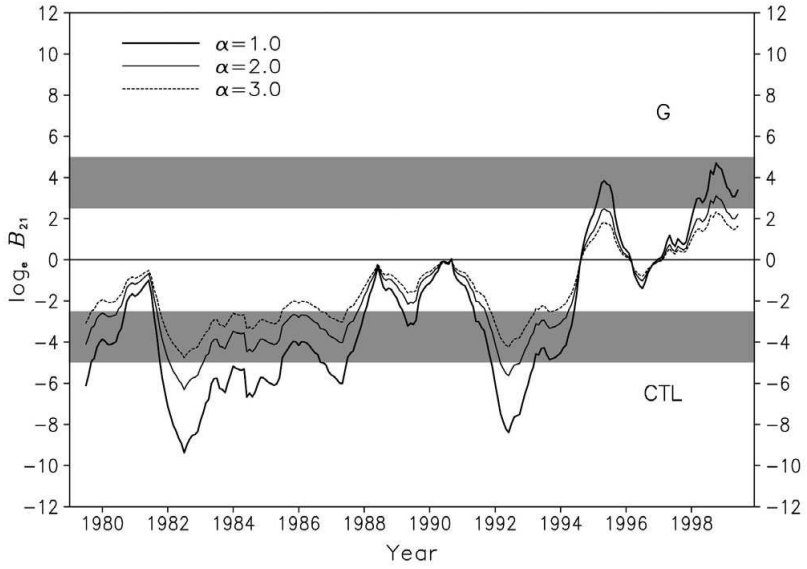



**Figure 4.7:** Time series of the Bayes factor $log\,B_{ir}$ of a control scenario (CTL) and a scenario (G) given NCEP reanalysis $2m$ and $70hPa$ temperature anomalies for the period 1979-1999. The upper (lower) shaded area indicates the interval "strong" evidence against CTL (G) scenario (Min et al., 2004).

case in which the specific model $m_i$ is more likely than the reference model $m_r$. Numbers less than zero indicate that the reference model is more likely and values around zero show that both ensemble prediction systems can not be distinguished between each other.

Table 4.2 according to Kass and Raftery (1995) introduces specific levels of evidence including a description for comparing the performance of ensemble $m_i$ vs. ensemble $m_j$ and the level of confidence of the predictive performance of the ensemble $m_i$ when compared to a verifying analysis $m_r$.

Min et al. (2004) used the Bayes factor for climate change signal analysis. In Fig. 4.7, the time series of natural logarithm of Bayes factors given the NCEP/NCAR reanalysis from 1979 to 1999 is shown. Positive values indicate evidence in favor of the greenhouse-gas forced scenario (G) while negative values show evidence in favor of the control scenario (CTL). According to Tab. 4.2, if the natural *log* of the Bayes factor exceeds 2.5 [5], the evidence against the respective alternative scenario is at least "strong" ["decisive"]. Since the late 1990s an increasing number of observations indicate partly strong evidence for the G scenario while over large parts of the record in 1980s there is substantial evidence for the CTL scenario given the data (Min et al., 2004).

## 4.5 Comparison of the Bayes factor with other probabilistic scores

The logarithm of the Bayes factor has parallels to the Ignorance score. But the Bayes factor has several advantages in contrast to the Ignorance score. These advantages are

- consideration of the observation error
- multivariate state vector
- standard kernel dressing (SDK)

The Bayesian approach includes in a natural way the uncertainty of the observations via the so called likelihood and the approach allows to investigate multidimensional state vectors of ensemble forecasts which can be treated by the standard kernel dressing.

The dissadvantage of the Bayesian approach is the limitation to normal distributed variables. The integral for the likelihood can be analytically solved only for normal distributed PDFs. But previous works has shown that the Bayesian approach provides a convenient way to compare and verify ensemble climate simulations e.g. Min et al. (2004) and Min and Hense (2006). Additionally, Gneiting and Raftery (2007) show the properness of the logarithm of the Bayes factor allowing an unbiased evaluation of the forecasts either with respect to a climatology or a different forecasting system.

In the following the Ignorance score *IGN* and the continuous ranked probability score *CRPS* are compared with the Bayes factor in more detail. Both scores were already mentioned in Chapter 2. However, here they will be discussed in more detail.

### a. Ignorance score

The Ignorance score (*IGN*) is defined by the value of a PDF at the observation point. For a normal distributed PDF $p(x) \sim N(\mu, \sigma^2)$, the *IGN* is defined by

$$IGN = -log(p(o)) \tag{4.49}$$

$$= \frac{1}{2}\log(2\pi\sigma^2) + \frac{(o-\mu)^2}{\sigma^2} \tag{4.50}$$

In general, the Ignorance score is a special case of the likelihood $l(o|m_i)$ for the case that the observation error goes to zero $\lim_{\gamma\to 0}\gamma\Sigma_o$. This used for the likelihood Eq. (4.36) gives

$$\lim_{\gamma\to 0} \quad l(o|m_i) = \int \lim_{\gamma\to 0} p_l(o|f)\, p_l(f|m_i) df \tag{4.51}$$

Using the Dirac delta function as limit for a Gaussian distribution
$\lim_{a\to 0}\left(\frac{1}{\sqrt{2\pi a}}\exp\left(-\frac{(f-o)^2}{2a}\right)\right)=\delta(f-o)$ into Eq. (4.51) leads finally to

$$\lim_{\gamma\to 0} \quad l(o|m_i)=\int \delta(f-o)\,p(f|m_i)\,df \tag{4.52}$$

$$=\int \delta(f-o)\,p(f)\,df \tag{4.53}$$

Furthermore, using of the Dirac delta function property $\int_{-\infty}^{+\infty}\delta(f-o)p(f)=p(o)$ shows the final connection of the likelihood to the Ignorance score.

$$\lim_{\gamma\to 0} \quad l(o|m_i)=p(o) \tag{4.54}$$

This shows that the Bayes factor $\log B_{ir}$ is a generalization of the Ignorance score taking additionally into account the uncertainty of the observations as well as the spatial correlation structure of the verified forecasts.

$$\lim_{\gamma\to 0} \quad (\log B_{ir})=\lim_{\gamma\to 0} \quad \log\left(\frac{P(m_i|o)}{P(m_r|o)}\right) \tag{4.55}$$

$$=\lim_{\gamma\to 0} \quad \log\left(\frac{l(m_i|o)}{l(m_r|o)}\frac{P(m_i)}{P(m_r)}\right) \tag{4.56}$$

$$=\lim_{\gamma\to 0} \quad \log\left(l(m_i|o)\right)-\log\left(l(m_r|o)\right) \tag{4.57}$$

$$=\log\left(p_{m_i}(o)\right)-\log\left(p_{m_r}(o)\right) \tag{4.58}$$

$$=IGN_r-IGN_i \tag{4.59}$$

Finally, the Bayes factor can be expressed in the case of $\lim_{\gamma\to 0}\gamma\Sigma_o$ as a summation of two Ignorance scores $IGN_i$ and $IGN_r$ with the usage of the Laplace prior, which gives all priors the same probability $P(m_i)=P(m_r)=0.5$.

### b. Continuous ranked probability score (CRPS)

The continuous ranked probability score (*CRPS*) is an extension of the Brier score. The *CRPS* is the integral of the Brier score at all possible threshold values $t$ (Gneiting et al., 2004).

$$CRPS=\int_{\infty}^{\infty}\left[F(t)-H(t-o)\right]^2 dt \tag{4.60}$$

Gneiting et al. (2004) derive the analytic solution of the integral using normal distributions as

$$CRPS = \sigma \left\{ \frac{o-\mu}{\sigma} \left[ 2\Phi\left(\frac{o-\mu}{\sigma}\right) - 1 \right] + 2\varphi\left(\frac{o-\mu}{\sigma}\right) - \frac{1}{\sqrt{\pi}} \right\} \tag{4.61}$$

Where $H$ is the Heaviside function, $\varphi\left(\frac{o-\mu}{\sigma}\right)$ the CDF and $\Phi\left(\frac{o-\mu}{\sigma}\right)$ the PDF.

The investigation results of the *CRPS* will be compared with those of the Bayes factor in Chapter 6. To compare the *CRPS* with the Bayes factor, the *CRPS* is extended using a mixture model. That means, the single CDF which is investigated in the standard *CRPS* is replaced by $K_i$ CDFs as a mixture score $CRPS_{mixture}$. $K_i$ described here the number of members of ensemble $m_i$. The extended version of the $CRPS_{mixture}$ is

$$CRPS_{mixture} = \int_{\infty}^{\infty} \left[ \sum_{k=1}^{K_i} \left( F_k(t) - H(t-o) \right) \right]^2 dt \tag{4.62}$$

Like in the standard case, the integral can be analytically solved by using of normal distributions. The final formulation of the $CRPS_{mixture}$ is

$$\begin{aligned} CRPS_{mixture} = \sigma \sum_{i=1}^{Ni} \sum_{j=1}^{Nj} \Bigg\{ & \frac{o-\mu_i}{\sigma} \cdot \Phi\left(\frac{o-\mu_i}{\sigma}\right) + \frac{o-\mu_j}{\sigma} \cdot \Phi\left(\frac{o-\mu_j}{\sigma}\right) \\ & - \frac{o}{\sigma} + \frac{\mu_i}{\sigma} \cdot F\left(\frac{\mu_j-\mu_i}{\sqrt{2}\sigma}\right) + \frac{\mu_j}{\sigma} \cdot F\left(\frac{\mu_i-\mu_j}{\sqrt{2}\sigma}\right) \\ & + \varphi\left(\frac{o-\mu_i}{\sigma}\right) + \varphi\left(\frac{o-\mu_j}{\sigma}\right) \\ & - \frac{1}{\sqrt{\pi}} \exp\left( -\frac{1}{4\sigma^2} \cdot (\mu_i-\mu_j)^2 \right) \Bigg\} \end{aligned} \tag{4.63}$$

where $\Phi$ and $\varphi$ denote the PDF and the CDF of normalized distribution. The standard *CRPS* of Eq. (4.61) is identical for the case that all $K_i$ member of an ensemble are identical.

Gneiting et al. (2004) summarize the differences and similarities of those two scores. Both, the *IGN* and *CRPS* are proper scores where smaller values are better. The key difference between both are that the *CRPS* grows linearly with the normalized error $\frac{o-\mu}{\sigma}$, whereas the *IGN* grows quadratically. On the basis of the greater robustness of the *CRPS*, the *CRPS* is to prefer. The *IGN* tends to be sensitive to events, which are outliers or extreme events. However, both scores are univariate and do not consider observation uncertainties.

### c. Skill scores of *IGN* and *CRPS*

The comparison of the Bayes factor with the *IGN* and the *CRPS* is done by usage of the skill scores (*SS*) of those scores. In the following, the *SS* of the *IGN* and the *CRPS* are denoted by

$$IGNSS = \frac{IGN_i - IGN_r}{IGN_{perf} - IGN_r} \tag{4.64}$$

$$CRPSS = \frac{CRPS_i - CRPS_r}{CRPS_{perf} - CRPS_r} \tag{4.65}$$

# Chapter 5

# Data and methodology

In this investigation, ensemble data of the demonstration of probabilistic hydrological and atmospheric simulation of flood events in the alpine region project (D-Phase) project (Arpagaus et al., 2009) are used. The D-Phase project is a forecast demonstration project (FDP) of the world weather research programme of WMO (WWRP). D-Phase is used to investigate the ability of forecasting heavy precipitation and related flooding events in the Alpine region. The domain of the D-Phase forecasts covers the whole COPS area in the south-western part of Germany and is shown in Fig. 5.1.

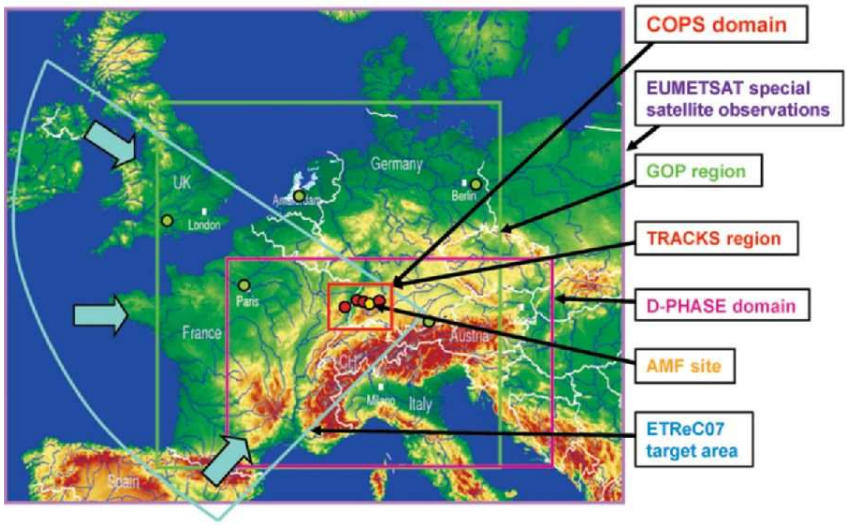



**Figure 5.1:** International collaboration within COPS during summer 2007. The green circles show the locations of major observatories, where data are stored within the GOP. The red circles indicate the COPS supersites, and the yellow circle is the ARM Mobile Facility (AMF). The blue arrows indicate the mean flow (Wulfmeyer et al., 2008).

The ensemble suite contains the SREPS and the LEPS ensembles. The SREPS is initialized by four global NWP models, while the LEPS is initialized by 16 representative members of

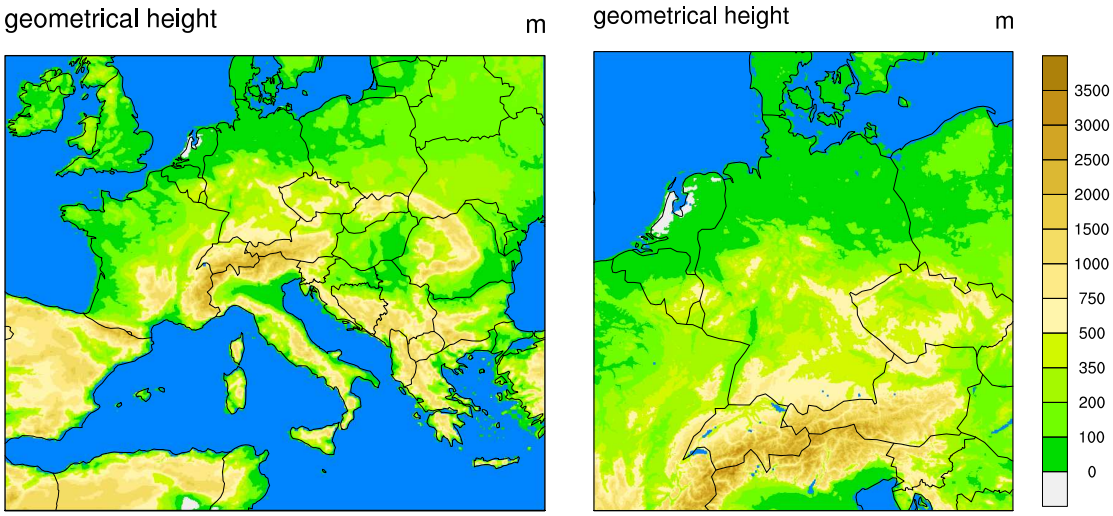



**Figure 5.2:** Model domain of COSMO-SREPS, COSMO-LEPS (left side) and COSMO-DE-EPS (right side).

the global ECMWF-EPS. Both ensembles are based on the COSMO model and contain perturbations of the model physics. The COSMO-DE-EPS (DE-EPS hereafter) provides the third regional forecast ensemble. It is development at the DWD and is a short range ensemble based on the non-hydrostatic COSMO-DE model. This model is a convection permitting limited-area model with a horizontal grid spacing of 2.8$km$ and 50 vertical model levels (Baldauf et al., 2006).

The ensemble data of the DE-EPS presented here are from runs at the DWD with an experimental version of the DE-EPS, which comprises perturbations of the initial and boundary conditions and of the model physics. The initial perturbations and the boundary data based on the courser resolved SREPS (10$km$) with parametrized convection. Table 5.1 gives an overview about the specifications of the ensemble systems SREPS, LEPS and DE-EPS. The model domains of the SREPS, LEPS and DE-EPS are shown in Fig. 5.2.

As observations radiosonde ascents of the COPS campaign (Wulfmeyer et al., 2008) are used provided by DWD. During the COPS IOPs radiosondes were released every 6 hours within the COPS area.

In the following, the three ensembles are presented in more detail as well as the observation technique by radiosondes. Finally, the procedure for the verification of vertical profiles of the ensembles is shown.

| | **COSMO-SREPS** | **COSMO-LEPS** | **COSMO-DE-EPS**<br>experimental version, not operational version |
|---|---|---|---|
| Members | 16 | 16 | 20 |
| Mesh-size | 10*km* (0.09°) | 10*km* (0.09°) | 2.8*km* (0.025°) |
| Vertical levels | 40 | 40 | 50 |
| Convection | Tiedtke/Kain-Fritsch | Tiedtke/Kain-Fritsch | explicitly resolved |
| Grid points | 258 x 306 | 258 x 306 | 461 x 421 |
| Rotated Nordpol | 40N, 170E | 40N, 170E | 40N, 170E |
| in rotated coord. | | | |
| Lower left corner | -16, -12.5 | -16, -12.5 | -5, -5 |
| Forecast range | 72h | 132h | 24h |
| Initial time | 00 UTC | 12 UTC | 00 UTC |
| Initialisation | No analysis, initialized by down-scaled forecasts | | COSMO-DE Analysis; |
| and | 4 global-models | 16 representative | initial perturbations |
| boundary | (IFS, GME, AVN, UM); | ECMWF-EPS members; | and boundary data |
| conditions | for each global model | for each of these member | based on SREPS |
| | 4 COSMO runs | one COSMO run | |
| Institutions | ARPA Emilia-Romagna | ARPA Emilia-Romagna | DWD |

**Table 5.1:** Ensemble systems of SREPS, LEPS from the ARPA-SIMC and DE-EPS from runs at DWD. The four global model are: ECMWF global (IFS), DWD global (GME), NCEP global (AVN), and UKMO global (UM). The ensembles of SREPS and LEPS contain perturbations of the model physics.

| Global Model | IC and BC ▷ 25 *km* | p1=default | p2=KF | p3=tur_len | p4=pat_len |
|---|---|---|---|---|---|
| IFS | ▷ COSMO-25 | m1 | m2 | m3 | m4 |
| GME | ▷ COSMO-25 | m5 | m6 | m7 | m8 |
| NCEP | ▷ COSMO-25 | m9 | m10 | m11 | m12 |
| UM | ▷ COSMO-25 | m13 | m14 | m15 | m16 |
| $f_{IC_{00UTC}} \Rightarrow f_{IFS_{12h-forecast}}, f_{GME_{12h-forecast}}, f_{NCEP_{12h-forecast}}, f_{UM_{12h-forecast}}$ | | | | | |

**Table 5.2:** The COSMO-SREPS consist of 16 members. Four different global models (IFS, GME, NCEP, UM) are used for the initial and the boundary conditions combined with four different physic perturbations (p1, p2, p3, p4) after Marsigli et al. (2007).

## 5.1 Ensemble Prediction Systems (EPS)

### COSMO-SREPS

The COSMO short-range ensemble prediction system (COSMO-SREPS) is a limited-area EPS. The SREPS is developed in a framework of a priority project of the COSMO consortium by ARPA-SIMC in Bologna (Marsigli et al., 2008). The ensemble has 16 members, and each of them is based on the limited-area non-hydrostatic COSMO model with a horizontal grid-spacing of 10*km* and 40 vertical levels. The model domain is shown in Fig. 5.2.

The model uncertainty is considered using a multi-analysis/multi-boundary approach (Marsigli et al., 2007). The generation of the 16 single members $m1$ to $m16$ of the SREPS is shown by Tab. 5.2. The initial (IC) and boundary condition (BC) perturbations are applied by driving the 10*km* COSMO runs with four 25*km* COSMO members of the multi-analysis/multi-boundary system of AEMet-SREPS. The AEMet-SREPS is developed by the national weather service agencia estatal de meteorologia (AEMet) in Spain. This four lower resolved COSMO runs of the AEMet-SREPS are nested finally into four different global models. This global models are the IFS (ECMWF), the GME (DWD), AVN (NCEP) and UM (UKMO).

The global models (IFS, GME, GFS, UM) are provided by the respective national weather service and the AEMet-SREPS by AEMet for this purpose. A representation of the small scale uncertainty is accomplished by applying limited-area model perturbations to the 10*km* COSMO runs. In particular, 4 different set-ups of the model physics have been adopted:

(p1) default set-up

(p2) use of the Kain-Fritsch (KF) scheme for the parametrization of the deep convection, instead of Tiedtke as in the control

(p3) tur_len (maximal turbulent length scale in $m$)
parameter equal to 1000 instead of 500 as in the control

| | | |
|---|---|---|
| Global Model (GM) | IC and BC<br>▷ clustering | p1=KF/Tiedtke, p2=tur_len, p3=pat_len |
| ECMWF EPS | ▷ RM1 .. RM16 | m1 .. m16 |
| $f_{IC_{12UTC}} \Rightarrow f_{RM1_{IFS-Analysis}}, .., f_{RM16_{IFS-Analysis}}$ | | |

**Table 5.3:** The COSMO-LEPS consist of 16 member based on the COSMO model. The perturbations are caused by the use of the global ECMWF EPS which is clustered into 16 representative member (RM) for the initial conditions and the boundary conditions combined with random choice of KF or rather Tiedtke (Montani et al., 2007; COSMO-Website, 2011).

(p4) pat_len (length scale in *m* of sub-scale surface patterns over land)
parameter equal to 10000 instead of 500 as in the control.

Finally, the combination of the 4 possible choices for the driving run with the 4 possible choices for the physics set-up leads to the 16 members ensemble (Marsigli et al., 2008).

### COSMO-LEPS

The COSMO limited-area ensemble prediction system (COSMO-LEPS) is a limited-area EPS too. Montani et al. (2003) describe the detailed generation of the ensemble. In a first step, the single ensemble members of the ECMWF-EPS are clustered into 16 groups with similar characteristics. From each of those groups a representative member (RM) is selected. Then, the RMs are used to provide the initial and boundary conditions for the single limited-area model runs of the LEPS. The LEPS consists of 16 COSMO runs with a horizontal grid-spacing of $10km$.

A special feature of the LEPS is the use of two consecutive ECMWF-EPS runs started at 00 and 12UTC for the 12UTC LEPS run. This leads to the opportunity to use a 102-member ensemble for the clustering of the IC and BC of the LEPS. In contrast, the standard ECMWF-EPS has only 51 ensemble members.

In particular, three different set-ups of the model physics have been adopted in the ensemble suite. Since December 2007, new random perturbations in each COSMO-LEPS integration are used:

(p1) random selection of the convection scheme (Kain-Fritsch or Tiedtke)

(p2) random selection of tur_len parameter

(p3) random selection of pat_len parameter.

The usage of these three possible choices for the physics set-up within the 16 COSMO runs with IC and BC from the 16 RMs leads to the 16 members ensemble.

## COSMO-DE-EPS

The COSMO-DE-EPS is developed for very short-range probabilistic forecasts. The COSMO-DE-EPS is also a limited-area EPS. Each single member of the COSMO-DE-EPS is based on the COSMO-DE. The COSMO-DE is a non-hydrostatic and convection-permitting model with a resolution of 2.8$km$. It has been developed in the framework of COSMO at the DWD (Baldauf et al., 2006).

The COSMO-DE is operational since April 2007. The model domain of the COSMO-DE (Fig. 5.2) covers entire Germany. The model has 50 vertical levels up to 30$hPa$. The cloud microphysical processes are modeled by a two-category ice scheme, which explicitly includes graupel, snow and rain. Because of its resolution, the model allows explicitly to assimilate the high-resolved radar data through the latent heat nudging (LHN) and to simulate deep convection without a parametrization scheme like KF or Tiedtke. The advantageous of the explicit simulation of deep convection was discussed in Chapter 3.

In the DE-EPS, the uncertainties of the COSMO-DE are described by perturbations of the initial state, the boundary conditions and of the model physics. The initial and boundary condition perturbations origin from the SREPS. However, this point applies only to the here used experimental DE-EPS version, which is not equal to the operational DE-EPS at the DWD. Additionally, the model physics are perturbed by changes of four namelist parameters of the COSMO-DE. The set-up of the DE-EPS is shown in Tab. 5.4 including the vertical filtering of the initial conditions of the DE-EPS (Peralta et al., 2012).

In the experimental DE-EPS version, the DE-EPS is nested into the COSMO-SREPS (10$km$) and further as mentioned before, the COSMO-SREPS into the AEMet-SREPS (25$km$) allowing to transfer the forecast uncertainty from the global scale to the short-range scale of the DE-EPS. Basically, the initial and the boundary conditions of the DE-EPS are defined by the four different global models.

The COSMO-DE-EPS is operational since May 2012. However with a slightly modified set-up constellation as described here. The COSMO-SREPS is replaced by a COSMO-7 EPS similar to the SREPS. In particular, five different set-ups of the model physic have been adopted in the DE-EPS:

(p1) entr_sc (entrainment rate of shallow convection)
parameter equal to 0.0003 instead of 0.002 as in the default

(p2) rlam_heat (scaling factor of laminar sublayers)
parameter equal to 0.1 instead of 1.0 as in the default

(p3) rlam_heat (scaling factor of laminar sublayers)
parameter equal to 10.0 instead of 1.0 as in the default

(p4) q_crit (critical value for normalized over-saturation)
parameter equal to 1.6 instead of 4.0 as in the default

| IC and BC | p1=entr_sc | p2=rlam_heat | p3=rlam_heat | p4=q_crit | p5=tur_len |
|---|---|---|---|---|---|
| SREPS IFS | m1 | m2 | m3 | m4 | m5 |
| SREPS GME | m6 | m7 | m8 | m9 | m10 |
| SREPS NCEP | m11 | m12 | m13 | m14 | m15 |
| SREPS UM | m16 | m17 | m18 | m19 | m20 |

$f_{IC_{00UTC}} = f_{DE_{Analysis}} + W(k)(f_{SREPS_{0h-forecast}} - f_{EU_{0h-forecast}})$

*Vertical low pass exponential filter:* $W(k) = \exp(-C|k/N_{ke}|^{\gamma})$, *with* $0 \leq k \leq N_{ke} = 50$

*Order of the filter:* $\gamma = 14$, *here the 5 levels closest to the surface are undisturbed*

*Constant:* $C = 73.68$

**Table 5.4:** The COSMO-DE EPS consists of 20 members. Four different members of SREPS including the four different global models are used for the initial conditions and the boundary conditions combined with four different physic perturbations (Gebhardt et al., 2011; Peralta et al., 2012).

(p5) tur_len (maximal turbulent length scale in *m*)
parameter equal to 150 instead of 500 as in the default.

The combination of this set-ups leads to the 20 members of the DE-EPS. In addition to the EPS in the next section the used observation method is described.

## 5.2 Radiosonde observations

Radiosondes are still important measurements for the initial state of NWP models and for the verification of NWP models. Recently, for limited area NWP models air plane and satellite measurements become important too. The radiosonde observations are used with additional observations in the data assimilation to provide the initial conditions for NWP models. Worldwide, there are more than 800 radiosonde launch sites per day and in Europa alone, more than 100 per day. Figure 5.3 shows the measurement network of radiosondes in Europa. The radiosonde stations are not uniformly distributed. Especially over the north sea, there are large gaps without measurement data. This is one reason for research to assimilate additional observations like satellite data into limited area NWP models.

Furthermore, the data are shared with other weather services by the global telecommunication system (GTS) through international agreements. The routine radiosonde launches occur about 45 minutes before the official observation time of 00UTC and 12UTC to provide an instantaneous profile of the atmosphere. For the initialization of NWP models and for verification of NWP models radiosonde data are still very important. Till now, satellite data are nearly almost used for global models like the GME. However limited area models like the COSMO-EU and COSMO-DE do not use them operationally till now.

Generally, a radiosonde is a measuring instrument on a weather balloon that measures various

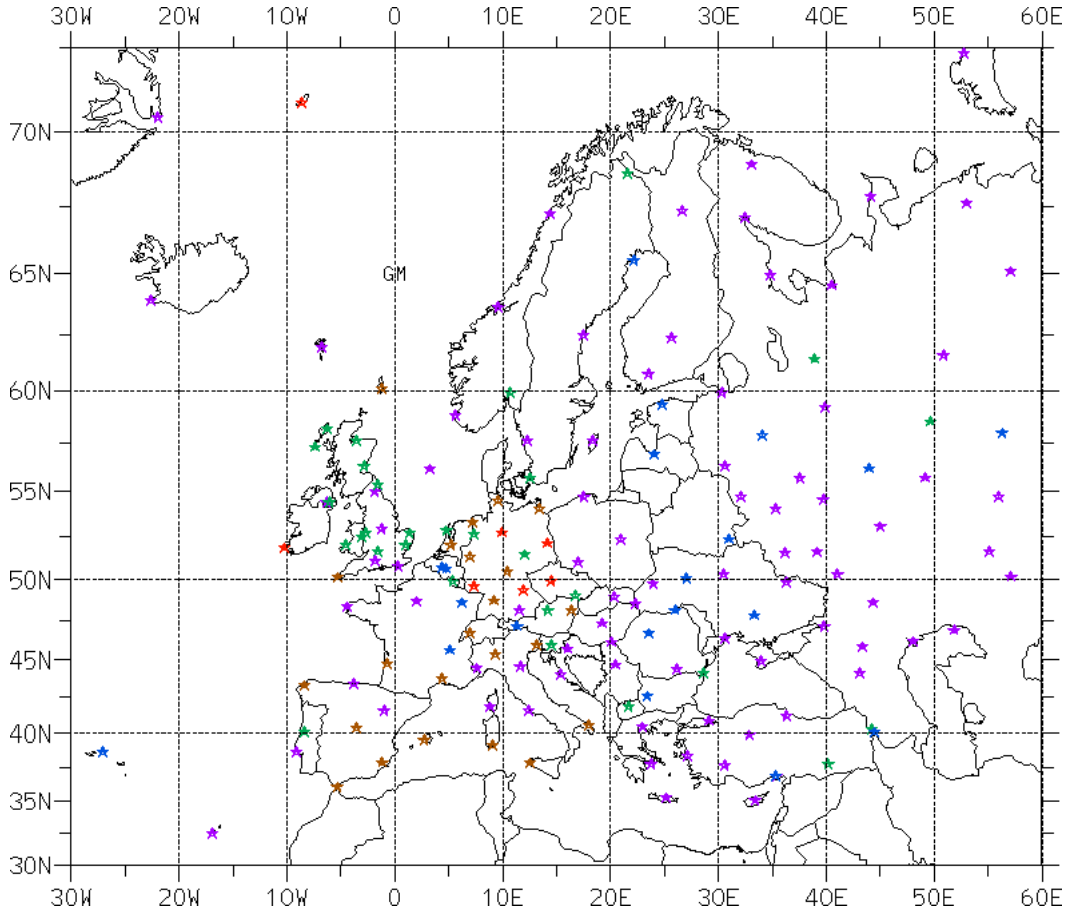



**Figure 5.3:** Locations of the radiosonde stations in Europe. The color of the marks represents the available radiosonde ascents in august 2011. The color green means 0 ascents, the colors blue 5, violet 32, yellow 63 and orange 94 ascents (Ermert, 2011).

atmospheric parameters on his ascent and transmits them to weather services like the DWD. Modern radiosondes measure and partly calculate the following variables:

- Temperature
- Pressure
- Relative humidity
- Wind speed and wind direction
- Altitude
- Geographical position (latitude/longitude)

Since July 2007, the DWD used radiosondes from Vaisala of the type RS92. These radiosondes operate with a radio frequency of 1680$MHz$ and have a measurement range from 1000$hPa$ up to

| | Temperature / K | Relative humidity / - | Pressure / *hPa* |
|---|---|---|---|
| Measurement range | $-90..+60$ | 0.00..1.00 | 1080..3 |
| Resolution | 0.1 | 0.01 | 0.1 |
| Accuracy | 0.15 | 0.02 | 0.4 |
| Measurement cycle | 1 second | | |

**Table 5.5:** Technical data of the Vaisala Radiosonde RS92-D (Vaisala, 2010).

3*hPa*. Table 5.5 shows the technical specifications of the radiosonde RS92 for the temperature, the humidity and the pressure sensor.

The data of the radiosondes are available in a special data format, the BUFR format, which are converted typically into the more usable netCDF format nowadays. The ascent of a radiosonde (called "TEMP") is stored in four parts (Part A, B, C and D). Table 5.6 shows, which data are stored in each part. The data sections A and C contain observations at the standard atmospheric pressure levels. These levels are also called the mandatory levels. The significant levels of part B and C contain only observations of levels where in the temperature or rather humidity are significant changes. In this work radiosonde data of part A are used. This has the advantage to use the observed variables on the same pressure levels on which the model data are available.

| | Range | Levels |
|---|---|---|
| Part A | SFC to 100*hPa* | Mandatory Levels |
| Part B | SFC to 100*hPa* | Significant Levels |
| Part C | 100*hPa* and higher | Mandatory Levels |
| Part D | 100*hPa* and higher | Significant Levels |

**Table 5.6:** Data section of the radiosonde obeservations.

Measurements of radiosondes are point or rather line measurements. This measurements contain atmospheric variations in time and space due to the limited sample space of the measurements (Kitchen, 1989). Additionally, it has been shown by Miloshevich et al. (2009) that there was a significant daytime Bias in the measurements, because of the effect of solar radiation. This measurements have to represent the "true state" of the atmosphere, which is discussed in detail in Chapter 2.

In this work, the observation uncertainties of radiosondes are explicitly considered. However, the previously mentioned Bias is not further treated. The standard deviation for the radiosonde data is extracted from the 3dvar data assimilation scheme of the DWD. The *RMSE* used in the 3dvar for the temperature at different heights is shown in Tab. 5.7 and based on the IFS documentation (White, 2003a). However, the *RMSE* of the 3dvar is beyond the standard instrumental error shown in Tab. 5.5, because for the data assimilation the total uncertainty of the sounding has to be considered. This includes that the sounding is not corrected for positioning errors until now. Kitchen (1989) indicates that the usage of not position corrected radiosonde data is justified for the synoptic scale. Thus, this does not apply for NWP models like the high

| | Temperature error / K | Relative humidity error / - |
|---|---|---|
| 1000*hPa* | 1.7 | 0.15 |
| 850*hPa* | 1.5 | 0.15 |
| 700*hPa* | 1.3 | 0.15 |
| 500*hPa* | 1.2 | 0.15 |
| 400*hPa* | 1.2 | 0.15 |
| 300*hPa* | 1.4 | 0.15 |
| 250*hPa* | 1.5 | 0.15 |
| 200*hPa* | 1.5 | 0.15 |
| 150*hPa* | 1.6 | 0.15 |
| 100*hPa* | 1.7 | 0.15 |

**Table 5.7:** The *RMSE* height errors at standard pressure levels of radiosondes (TEMPs) used in 3dvar. The *RMSE* is based on the IFS documentation (White, 2003a).

resulted COSMO-DE with a mash size of 2.8*km*. There are plans at the DWD to use position corrected radiosonde data in near future. Until then, this issue is treated poorly by an increased standard deviation.

For this investigation, the observations are radiosonde measurements of the DWD and of the COPS-Campaign. During the entire COPS campaign radiosondes were released every 6 hours (00,06,12,18 UTC) within the COPS area. During the COPS intensive observation periods (IOP) additional radiosondes at the German stations were started at 5,8,11,15,18 and 21 UTC. In this study the radiosonde stations Stuttgart, Idar-Oberstein and from MeteoFrance Nancy are used. The COPS radiosondes were Burnhaupt, Meistratz, Achern and Karlsruhe. All radiosonde stations are shown in Fig. 5.4. The pressure levels are 1000, 925, 850, 700, 500, 300, 250 and 200*hPa*.

## 5.3 Methodological procedure

The proceeding for the comparison and for the verification are described in this section. For both, comparison and verification, the Bayes factor is calculated mainly for three radiosonde stations. The profiles are Stuttgart $f_{stu}$, Idar-Oberstein $f_{ida}$ and Nancy $f_{nan}$ showed by the black triangle in Fig. 5.4. The three stations are treated

- arithmetically averaged to get a mean profile

$$f_{avg} = \frac{1}{3}\left(f_{stu} + f_{ida} + f_{nan}\right) \tag{5.1}$$

- and together unaveraged.

$$f_{cor} = \left[f_{stu}^{T}, f_{ida}^{T}, f_{nan}^{T},\right]^{T} \tag{5.2}$$

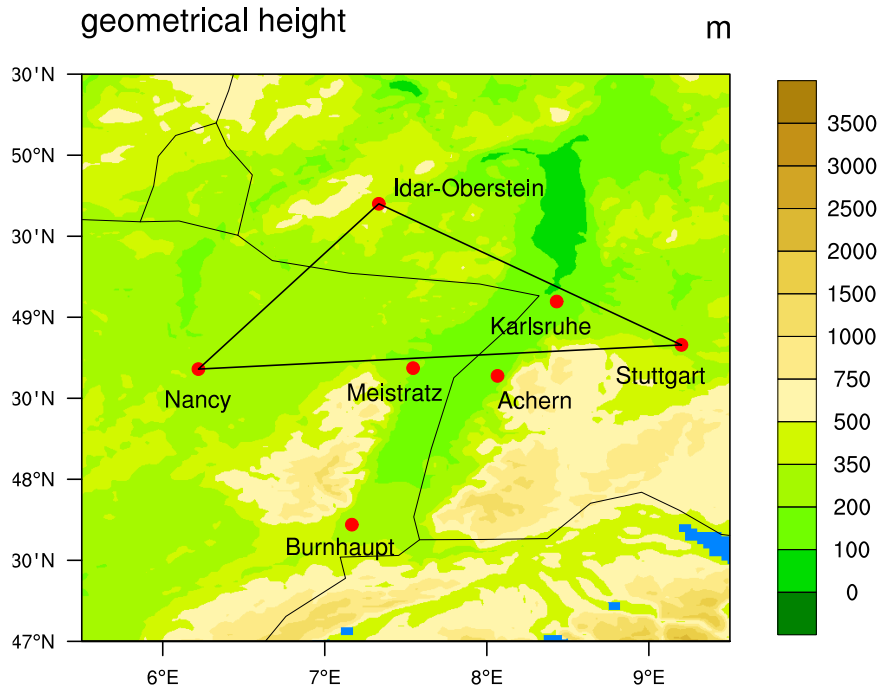



**Figure 5.4:** Radiosonde stations of the DWD (Stuttgart and Idar-Oberstein), of MeteoFrance (Nancy) and of the COPS campaign (Burnhaupt, Meistratz, Achern and Karlsruhe). The black triangle shows the main stations used in this work.

The reasons for this proceeding are on the one hand, the request for reliable results by the investigation of the arithmetically averaged profiles and on the other hand, the request for significant results as far as possible by the investigation of the unaveraged case.

The case of three stations together unaveraged in one vector considers the vertical correlations between the levels as well as the horizontal correlations between the stations. This allows among others a stricter, a more meaningful comparison. However, in the case of treating the three stations together in one vector the dimension $q$ of the vector (Eq. 5.1) makes the estimation of the covariance matrix by the standard maximum likelihood method impossible in case of a singular covariance matrix. In this case, the recently developed gLasso method by Friedman et al. (2007) is used to estimate the covariance matrix. Details of the method are presented in the **Appendix A.5**. Additionally, in **Appendix A.5** a comparison of the standard covariance matrix with the approximated gLasso-covariance matrix (Friedman et al., 2007) is shown. The comparison for the verification scenario SREPS vs. COSMO-EU analysis with three levels ($q = 3 * 3 = 9$) shows that the gLasso method provides nearly identical verification results.

In the following, the technical steps are explained, firstly. Then, the underlying scientific issue is discussed.

### Technical issue

In this work, a first step in the analysis is the conversion of the single ensemble predictions into a predictive PDF. Several methods have been described in literature e.g. Wilks and Hamill (2007) among which ensemble Gaussian kernel dressing can be found. The theoretical background was explained in Chapter 4. Essential to the method is the estimation of the dressing covariance matrix (Bröcker and Smith, 2008; Schölzel and Hense, 2010). The following steps are applied:

- This point has to be done only in case of investigation of the equivalent potential temperature $\Theta_e$. This variable allows to investigate the humidity calculated by a physical transformation described by Eq. (3.6). The basic humidity measured variables and their conversions are introduced in **Appendix A.1**.
- the first step for the temperature or the second for $\Theta_e$ is the interpolation from the surrounding grid points to the observation point. This has been done by a bi-linear interpolation described in the **Appendix A.4**.
- from a given ensemble at a fixed date and a fixed forecast lead time all possible differences between each single realization has to be calculated as a pre-whitening filter to remove approximately the true signal $f_t$ in Eq. (4.22).
- the differences are assumed to be realizations of the error scaled by a factor of $\sqrt{2}$ from which a first covariance matrix can be estimated shown by Eq. (4.31).
- the single covariance matrices for the past 5 days are calculated by this way.
- these covariance matrices $\Sigma_{i,\Delta t}$ are averaged over the past $N = 5$ days including the day, which has to investigate ($\Delta t = 0$). In doing so, $\Delta t$ describes the time distance (in days) to the investigation day $\Delta t = 0$. The daily cycle is taking into account by averaging only over the corresponding forecast lead time. The average $[\Sigma_i]$ is calculated as a weighted average:

  $$[\Sigma_i] = \frac{\sum_{\Delta t=0}^{4} w_{\Delta t} \quad \Sigma_{i,\Delta t}}{\sum_{\Delta t=0}^{4} w_{\Delta t}} \tag{5.3}$$

  The weight $w_{\Delta t}$ is given as

  $$w_{\Delta t} = N - \Delta t \tag{5.4}$$

- even this averaging is not based on a large enough sample size in case of treating three stations with eight levels ($q = 3 * 8 = 24$) to guarantee a non-singular covariance matrix from the standard maximum likelihood estimation as just mentioned. Therefore, it is used the gLasso method by Friedman et al. (2007) to estimate the covariance matrix. **Appendix A.5** shows the details.
- in the following the Bayes factor is averaged over a longer time period. If the covariance matrix has to be calculated for a day located at the edge of the time range then the period for the averaging is mirrored at this edge.

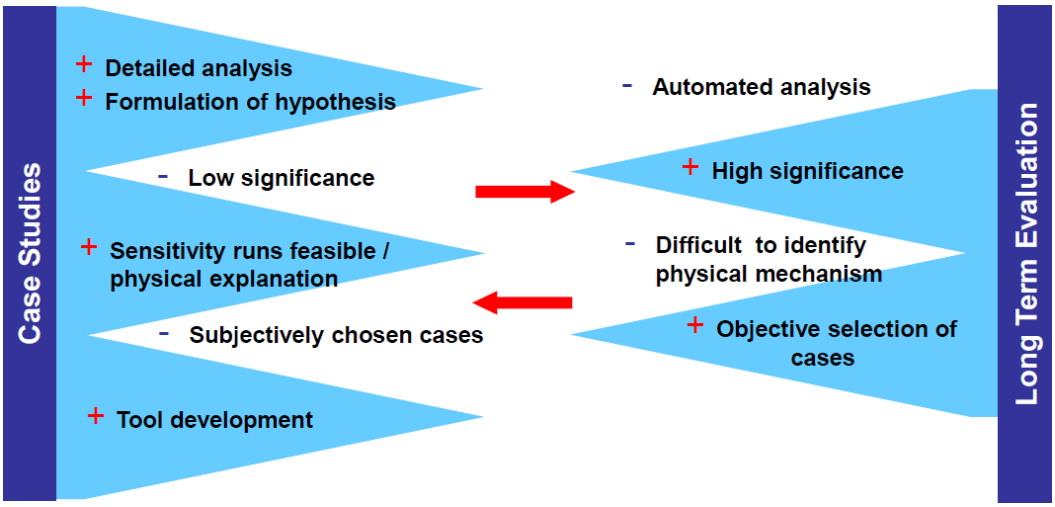



**Figure 5.5:** Comparsion of the advantages and disadvantages of case studies vs. long term evaluations (Ament, 2010).

- finally, the averaged matrix defines $\Sigma_i$ in Eq. (4.43).

- the radiosonde observations are processed as column observations at fix points at eight pressure levels:
  $(1000, 925, 850, 700, 500, 300, 250, 200 hPa)$ for the comparison DE-EPS vs. SREPS and at three pressure levels:
  $(850, 700, 500 hPa)$ for the verification SREPS and LEPS vs. COSMO-EU analysis.
  The forecasted temperature and moisture values at these pressure levels are used.

- if the surface pressure of the radiosonde station is lower than $1000 hPa$, the surface temperature of the radiosonde is extrapolated under the surface (to the $1000 hPa$ level). The approach used here is based on White (2003b) and is also used in the COSMO model when the variables are calculated on pressure levels. Details of the approach are presented in the **Appendix A.4**.

- the covariance matrix of the observations $\Sigma_o$ is assumed to be diagonal with the variances taken from the 3dvar data assimilation scheme used at the DWD.

- for the equivalent potential temperature, the covariance matrix of the observations $\Sigma_o$ has to be approximated based on the variances taken from the 3dvar. The details are shown in **Appendix A.2**.

- now, the Bayes factor $B_{ir}$ is calculated and averaged over several days. The results are given by the log $B_{ir}$ described for the comparison and for the verification by Tab. 4.2.

| IOPs | Time period | Weather type |
|---|---|---|
| IOP-1 a/b/c | 5./6./7. June | AMC / AMC / AMC |
| IOP-1 d | 8. June | AMC-SFC |
| IOP-2 | 12. June | WFC |
| IOP-3 a/b | 14./15. June | WFC / SFC |
| IOP-4 a/b | 19./20. June | AMC / SFC |
| IOP-5 a/b | 1./2. July | SFC / SFC |
| IOP-6 | 4. July | SFC |
| IOP-7 a/b | 8./9. July | SFC / SFC |
| IOP-8 a/b | 14./15. July | AMC / AMC |
| IOP-9 a/b/c | 18./19./20. July | SFC / SFC / SFC |
| IOP-10 | 23. July | SFC |
| IOP-11 a/b | 25./26. July | AMC / AMC |
| IOP-12 | 30. July | WFC |
| IOP-13 a/b | 1./2. Aug | AMC / SFC |
| IOP-14 a | 6.-7. Aug | WFC |
| IOP-14 b | 8. Aug | WFC |
| IOP-15 a/b | 12./13. Aug | AMC / WFC |
| IOP-16 | 15.-16. Aug | SFC |
| IOP-17 a | 21.-22. Aug | WFC |
| IOP-17 b | 22. Aug | WFC |
| IOP-18 a/b | 24./25. Aug | AMC / AMC |

**Table 5.8:** IOPs of the COPS campaign in 2007 with corresponding weather type classification (AMC: Air-mass convection, WFC: weakly forced convection, SFC: strongly forced convection) after Wulfmeyer et al. (2011).

## Scientific issue

The aim of this work is to investigate the predictability of convection. More precisely, the predictability of convective permitting conditions will be investigated. To do this, the scientific proceeding contains the investigation of the three radiosonde stations in a long term evaluation as well as a regime dependent comparison and verification respectively. This includes cases studies too. A schematic overview of the advantages and disadvantages of long term evaluations or rather cases studies are shown in Fig. 5.5.

- **Long-term evaluation**
  The long-term evaluation is needed to get reliable investigation results.
- **Regime-dependent verification**
  To investigate the ability of the models to predict the conditions for convection a regime-dependent comparison and verification relating to Tab. 5.8 is applied. This kind of investigation is very important to see if the results are different for different weather regimes. It

has been shown in the past that QPFs are of varying quality for different weather regimes (Keil and Craig, 2011). Table 5.8 shows the intensive observation periods (IOPs) of the COPS campaign. These IOPs are categorized into air-mass convection, weakly forced condition and strongly forced condition. This classification is used for the verification of SREPS and LEPS. For the comparison of the DE-EPS with the SREPS, a regime-selection by the convective timescale $\tau_c$ is applied (Keil and Craig, 2011).

- **Case-study vs. long-term study**
  The Bayes factor can be used to investigate explicitly single ensemble runs respectively case studies. For more meaningful results, longer time periods are mainly investigated in this work. Furthermore, a regime-dependent analysis extends the investigation results here. To complete this work, one detailed case-study is investigated (8th August 2007 in Chapter 6 and 15th July 2007 in Chapter 7) to illustrate how the Bayesian approach works and to use additional radiosonde data, which are only available for few COPS-IOPs.

  The probabilistic verification method presented here is completely new for NWP models. Because of this, both, the applicability of the Bayesian statistic as well as the predictability of convective conditions, have to be investigated by a combination of long-term, regime-dependent and case studies. The limiting factor relating to the investigated ensemble data is the availability of the ensemble data during the D-Phase project and missing radiosonde launches during the COPS period.

In the following, Chapter 6 shows the comparison of the DE-EPS with the SREPS. The application of the Bayes factor for verification is shown in Chapter 7.

# Chapter 6

# Comparison of ensemble prediction systems

The Bayesian approach allows explicitly to compare and to verify ensembles, considering the uncertainties of the observations as well as of the model uncertainty as shown in Chapter 4. This is advantageous when two ensembles have to be compared as here in this chapter, because the aim is knowledge about the significance of the result of the comparison. Particularly, in the case of a conscious investigation of a small area like the COPS region to investigate there the predictability of the convective conditions. For this investigation, it is very important to consider the uncertainties of the observations. Since, when averaging over whole Germany, the effect of uncertain observations is weaker, because the different observation errors are averaged out. This is shown in Fig. 6.1 that even though that the temperature is nearly free of a Bias (especially for vv=0*h*), the *RMSE* is quite high. This already shows, it is almost impossible to forecast exactly the convective conditions in a small domain by a deterministic model forecast.

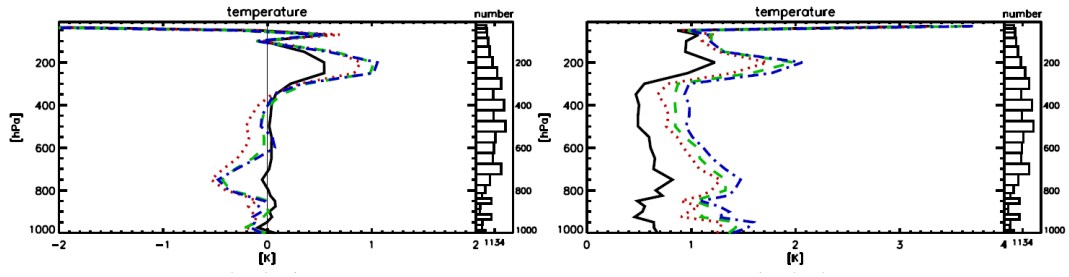



**Figure 6.1:** Standard verification of the forecasted vertical temperature profiles of vv=0,6,12,18*h* (black, red, green, blue lines) of the COSMO-DE from 00UTC runs averaged over all available radiosonde stations within the COSMO-DE domain. The thin black line represent the observation. The left panel shows the *Bias* and the right panel the *RMSE*. Additionally, the number of available radiosonde stations are shown (Pflüger, 2007).

To investigate the benefit of a highly resolved EPS with a coarser resolved EPS relating the pre-

dictability of convection here the convection permitting DE-EPS is compared with the SREPS. The aim of this comparison is to investigate the quality of the vertical profiles of the DE-EPS in comparison to the SREPS. In Chapter 3, the importance of the vertical structure of the atmosphere related to the occurrence of convection was shown. Especially, a realistic vertical structure is important for the DE-EPS, because the COSMO-DE has to simulate the convection explicitly without the usage of a convection scheme.

The Bayes factor can be evaluated separately for each day and for each forecast lead time, which is available. This is explicitly done for the COPS IOP-14b (8th August 2007). However first, for a greater representativeness the results are averaged over 21 days of August 2007. In general, the weather in this period was alternating between air-mass convection (AMC), weakly forced convection (WFC) and also strongly forced convection (SFC). The period contains the COPS IOPs 13a/b, 14a/b/c, 15a/b, 16, 17a/b and 18a/b (see Tab. 6.1). Different convection types are classified according to Keil and Craig (2011), and they are investigated by a regime dependent investigation. Additionally to the regime dependent investigation, the robustness of the results are investigated and the results are compared with those of further probabilistic scores.

## 6.1 Probabilistic comparison of COSMO-DE-EPS with the COSMO-SREPS

Regarding the comparison of the DE-EPS with the SREPS, first the temperature is compared and secondly the equivalent potential temperature. The equivalent potential temperature reflects additionally the impact of humidity.

### Comparison of temperature

Figure 6.2 a) shows that the DE-EPS is more likely than the SREPS at each forecast time. This result could be explained by the increased resolution of the DE-EPS of $2.8km$ horizontally and 50 model levels in the vertical in contrast to the $10km$ grid spacing and 40 model levels of the SREPS. This means that in the DE-EPS there are less parametrizations and more physical processes explicitly resolved by the model e.g. the convection. This is a possible cause for the more likely temperature profile. Another reason is the initialization of the DE-EPS. The DE-EPS is based on the COSMO-DE analysis including the LHN (Chapter 5) in contrast to the SREPS, which strongly relies on the four global forecast models.

The evidence for the DE-EPS in case of the calculation of the Bayes factor of three vertical profiles averaged arithmetically to get a mean profile is "strong", see Tab. 4.2. Furthermore, the evidence for the DE-EPS is even larger ("decisive") in the case of the simultaneous, joint treatment of the three profiles in one vector. In this case, the dimension of the model state vector is $q = 24$ having the advantage to get potentially clearer results. The error covariance matrix

| COPS IOPs | Basic Data | Weather type | $\tau_c$ |
|---|---|---|---|
| IOP-13 a | 1. Aug. 2007 | AMC | |
| IOP-13 b | 2. Aug. 2007 | SFC | |
| IOP-13 b | 3. Aug. 2007 | | |
| | 4. Aug. 2007 | | |
| | 5. Aug. 2007 | | |
| IOP-14 b | 8. Aug. 2007 | WFC | $\approx 0h$ |
| IOP-14 c | 9. Aug. 2007 | WFC | $\approx 0h$ |
| | 10. Aug. 2007 | | $< 6h$ |
| | 11. Aug. 2007 | | $< 6h$ |
| IOP-15 a | 12. Aug. 2007 | AMC | $\geq 6h$ |
| IOP-15 b | 13. Aug. 2007 | WFC | $\geq 6h$ |
| | 14. Aug. 2007 | | $\geq 6h$ |
| IOP-16 | 15. Aug. 2007 | SFC | $< 6h$ |
| IOP-16 | 16. Aug. 2007 | SFC | $< 6h$ |
| | 18. Aug. 2007 | | |
| | 20. Aug. 2007 | | |
| IOP-17 a | 21. Aug. 2007 | WFC | |
| IOP-17 a | 22. Aug. 2007 | | |
| | 27. Aug. 2007 | | |
| | 29. Aug. 2007 | | |
| | 30. Aug. 2007 | | |

**Table 6.1:** Basic data of the comparison DE-EPS vs. SREPS. Overall, 21 days of August 2007 are available with corresponding weather type classification (AMC: air-mass convection, WFC: weakly forced convection, SFC: strongly forced convection) after Wulfmeyer et al. (2011) and the convective time scale $\tau_c$ at afternoon from Keil and Craig (2011). The convective time scale $\tau_c$ is introduced in Chapter 3.

is estimated by the graphical lasso (gLasso) method from Friedman et al. (2007), because in this case the covariance matrix is singular in consequence of the fact that the ensemble size $K_i = 16$ (SREPS) or $= 20$ (DE-EPS) is smaller as the dimension of the model state vector $q$.

The meaning of the correlations between the levels and between the stations of this investigation is shown in Fig. 6.2 (c,d). Figure 6.2 c) shows that the vertical correlations are almost meaningless, because between the cases of consideration and non-consideration of the vertical correlations there is nearly no difference. However for the correlations between the stations the evident for the DE-EPS gets smaller if the horizontal correlations are explicitly considered shown by Fig. 6.2 d). Thus, the correlations between the stations seems to be more important than the correlations between the vertical levels. But the evident is "decisive" in all cases if the horizontal correlations are considered or not. The visualization of the corresponding correlation matrices are shown in the Appendix A.6. For the results presented in the following all correlations are considered.

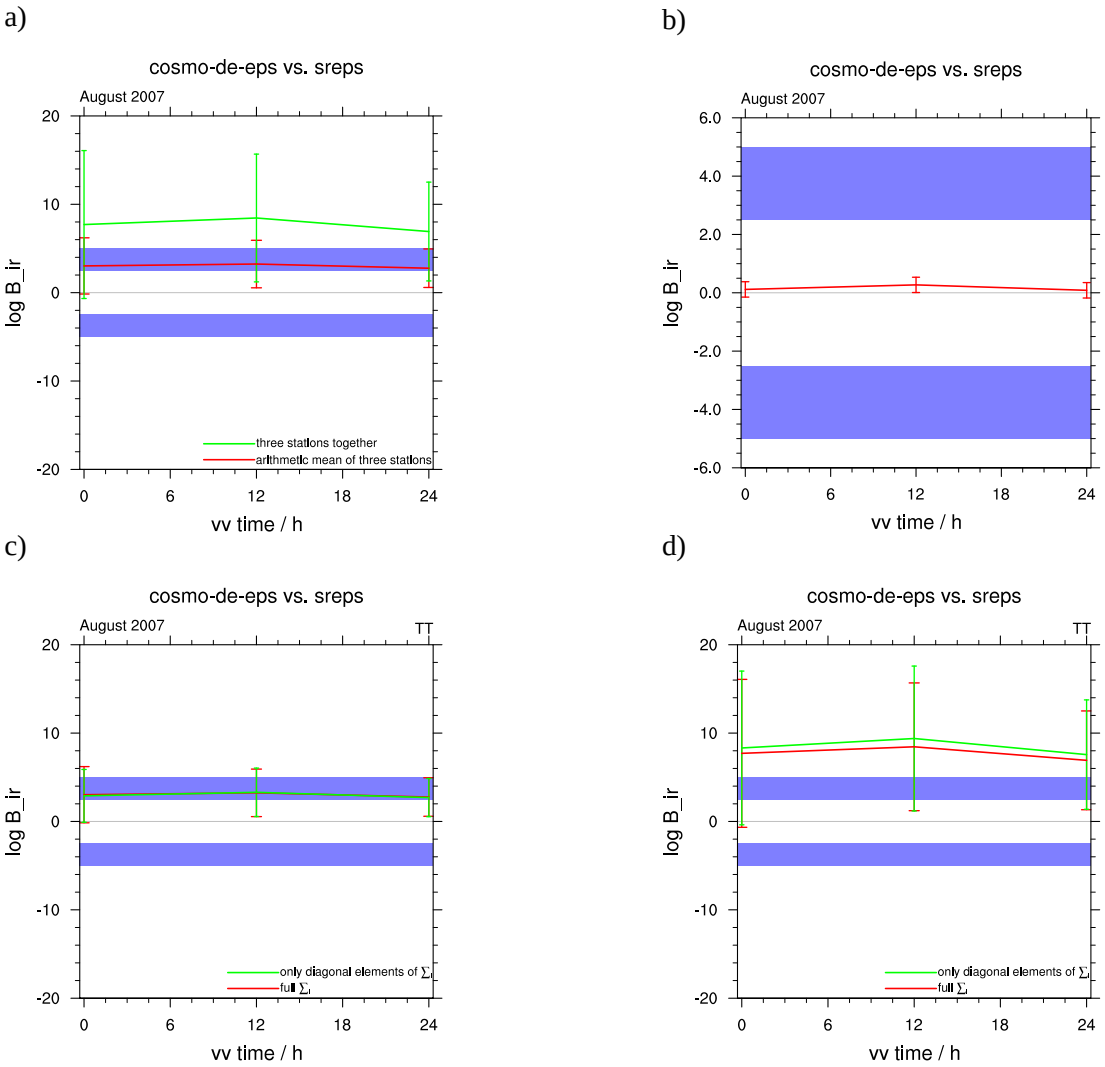



**Figure 6.2:** Time series of the Bayes factor of COSMO-DE-EPS for the temperature at forecast time (vv time) with respect to SREPS. The blue band describes the area from which the evidence for each model starts to be strong, see Tab. 4.2. Part (a) shows the multivariate case with eight vertical levels averaged over August 2007, while (b) shows the univariate (850*hPa*) case. (c) and (d) show the case of consideration and non-consideration of the correlation between the levels and stations. The mean (solid lines) and the standard deviation of August 2007 are shown.

The large standard deviation for the one month investigation period of August 2007 in Fig. 6.2 a) indicates that there are days of the investigation, which have partly an evidence substantially larger or rather smaller as the mean of the Bayes factor. This is figured out mainly by the standard deviation in the case of the joint treatment of the three profiles. Furthermore, to confirm this evidence for the DE-EPS it would be necessary to investigate a longer time period. In this chapter, the one month investigation period is investigated in more detail by a regime dependent comparison of the DE-EPS with the SREPS, which will be presented in section 6.3 to investigate the large differences in the Bayes factor for the single days.

Figure 6.2 b) shows the results for the univariate case taking as tested variable the 850*hPa*

temperature. In this case all relevant correlations between the temperature values at various levels and stations are completely lost and it is not possible to decide, which ensemble system is more likely. It clearly shows the advantage of using the multivariate approach. In the following an eof analysis is used to investigate the multivariate approach in more detail. Additionally, the sensitivity of the results to the observation error and to the Silverman's factor is shown in section 6.2. More details about the used gLasso method from Friedman et al. (2007) including a sensitivity study to the gLasso parameter ($\rho$) is shown in Appendix A.5.

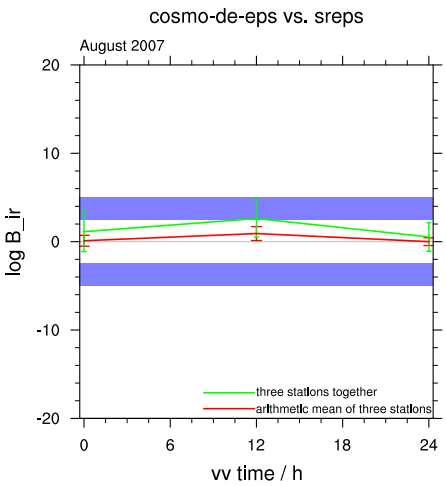
### Comparison of equivalent potential temperature

The humidity is very important for the occurrence of convection as was shown in Chapter 3. Thus, it follows the investigation of the humidity. The temperature investigated before is assumed as normally distributed. Such assumptions for the temperature are used typically in applied statistical works like e.g. Jewson and Caballero (2003). In this work, a kernel dressing approach is used additionally as shown in Chapter 4.3 to capture the partly bimodal distribution of the temperature at a single grid point. An illustration for this approach of the estimated PDF is shown in Chapter 4 (Fig. 4.5). In case of the humidity the assumption of normally distributed data is no longer valid. An alternative way is to investigate the equivalent potential temperature, because they also reflects the impact of the humidity. This way is used here with the extension of kernel dressing too. An exemplary visualization of this method is shown in the following in Fig. 6.15 for the equivalent potential temperature as well as for the temperature. A statistical treatment of the specific humidity itself would be also possible by the usage of a log-normal distribution. However, this procedure failed, due to the strong decrease of the humidity with height.

Figure 6.3 a) shows the same result as for the temperature. The DE-EPS is more likely than the SREPS in case of the equivalent potential temperature. However, in contrast to the investigation of the temperature, this applies mainly for the 12*h* forecast lead time and for the case of the simultaneous, joint treatment of the three profiles. The evidence for the DE-EPS is here "strong" and thus weaker as in the case of the temperature. For the other forecast times, the evidence is largely "neutral". The generally weaker evidence is due to the larger standard deviation of the calculated equivalent potential temperature of the observed state. Investigations to this issue including a reduction of the standard deviation of the observed state are discussed later in section 6.2 by results of a sensitivity study of the Bayes factor to the observation uncertainty. Nevertheless, it is shown that for the 12*h* forecast lead time the profile of the DE-EPS is again more likely. The reasons for that are the same as in case of the temperature. This result shows that the DE-EPS seems to have the preferable more realistic vertical profiles of temperature and equivalent potential temperature compared to the SREPS and consequently the better conditions to forecast convective events.

Figure 6.3 b) shows the univariate case in which it is again not possible to decide, which ensemble system is more likely. This is the same result as it was shown for the comparison of

a)




b)

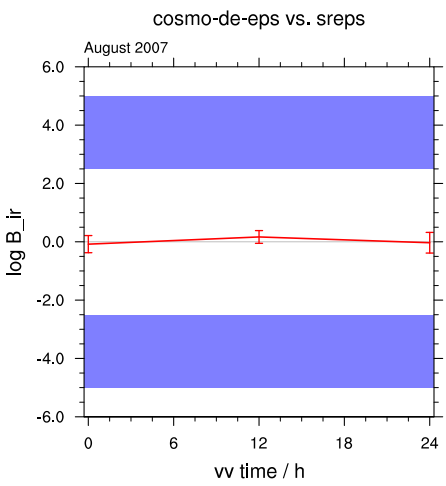



**Figure 6.3:** Time series of COSMO-DE-EPS for the equivalent potential temperature at forecast time (vv time) with respect to SREPS. The blue band describe a significant area, see 4.2. (a) shows the multivariate case with eight vertical levels, while (b) shows the univariate (850*hPa*) case. The mean (solid lines) and the standard deviation (error bars) of August 2007 are shown.

the temperature.

## Comparison of the vertical profiles

The corresponding vertical profiles of the temperature and of the equivalent potential temperature used for the Bayesian results averaged over 21 days are shown in Fig. 6.4. The vertical profiles are mean-profiles of the three radiosonde stations and also averaged over those 21 days. This illustration allows us in a first steep to look into more detail about the reasons of the Bayesian results before.

Figure 6.4 shows the mean error (*ME*) and the standard deviation of the DE-EPS and the SREPS. The reasons for the more likely vertical temperature profiles of the DE-EPS are shown by Fig. 6.4 a) and b). The DE-EPS is nearly free of a Bias whereas the SREPS has a clear positive Bias. This means, the SREPS is at all levels about 1*K* too warm. The positive bias of the SREPS is also present for the equivalent potential temperature shown in Fig. 6.4 c) and d).

The standard deviation of the mean error shows especially for the equivalent potential temperature (Fig. 6.4 b,d) clearly higher values as for the temperature (Fig. 6.4 a,c). Additionally, it is shown that the DE-EPS is not perturbed in the lowest model-levels at initial time (vv=0*h*). This is caused by the vertical low pass exponential filter introduced in Tab. 5.4.

The vertical profiles fit well to the Baysian results, but they allow only a simplified comparison of the DE-EPS with the SREPS. A regime dependent comparison of both ensembles, which will be presented later will give us the opportunity to have a more detailed view to the reasons

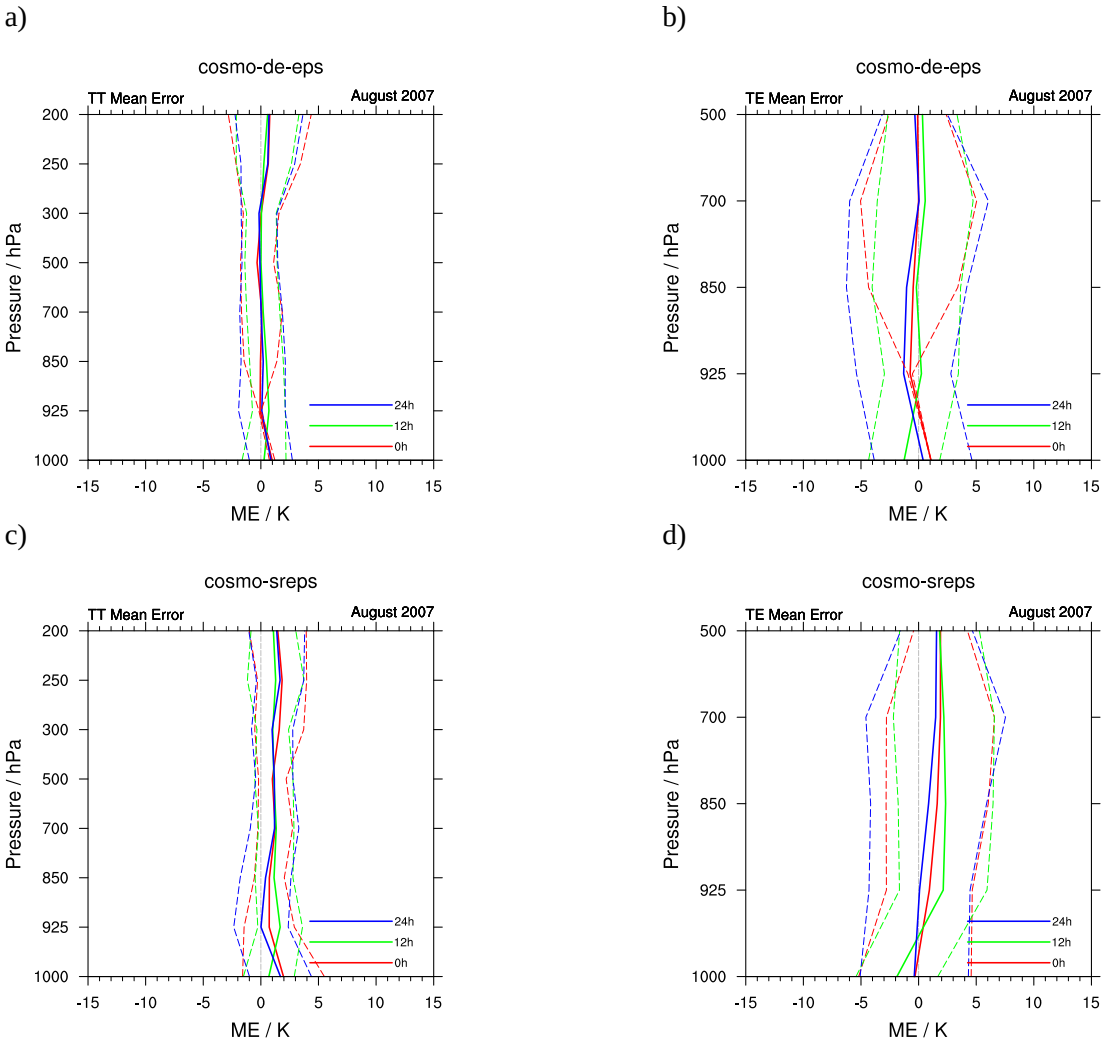



**Figure 6.4:** Vertical bias of the mean temperature of the DE-EPS (a) and of the SREPS (c) for 0*h*, 12*h* and 24*h* forecast lead time. (b) and (d) show the bias for the mean equivalent potential temperature. The dashed lines show the standard deviation plus minus the mean at the respective forecast lead times.

of the positive Bias of the SREPS.

## EOF-Analysis of the probabilistic comparison

The purpose of empirical orthogonal function (EOF) analysis is to extract important patterns from large data sets. The fundamentals are described e.g. in Wilks (1995)(p. 372) or Hannachi et al. (2007). The EOFs are obtained from the solution of the eigenvalue problem (Eq. 6.1). Where $\hat{e}$ are the eigenvectors and $\lambda$ the eigenvalues of the covariance matrix $\Sigma_i$.

$$\Sigma_i \hat{e} = \lambda \hat{e} \tag{6.1}$$

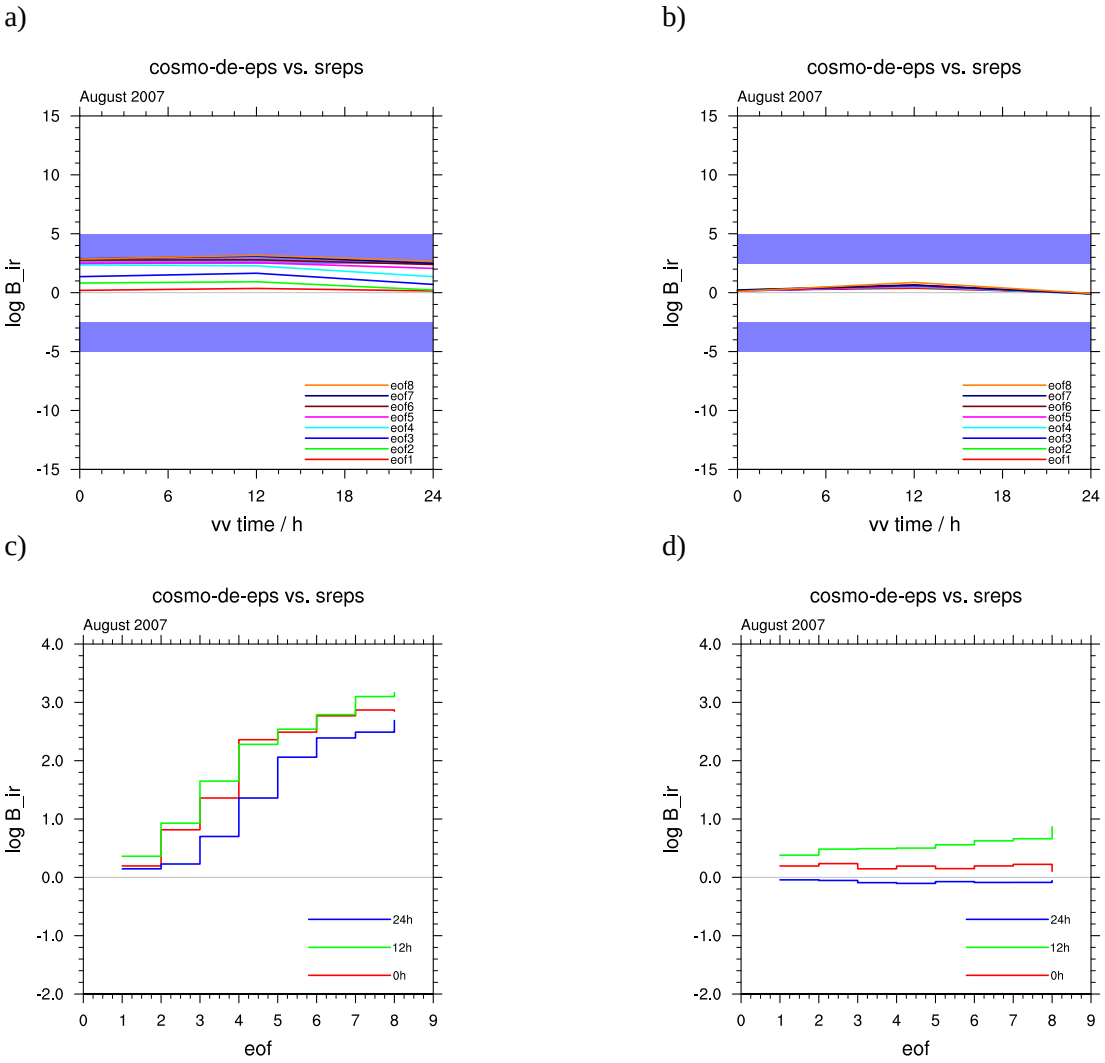



**Figure 6.5:** Empirical orthogonal function (EOF) of the Bayes factor of temperature (a,c) and (b,d) for the equivalent potential temperature. In (a,b) the term eof3 means that the first three EOFs are used (eof1-3) to present the Bayes factor as function of the lead time. Figures (c,d) show the Bayes factor for $0h$, $12h$ and $24h$ forecast lead time as function of the considered EOFs corresponding to (a,b).

The EOF analysis is used to investigate patterns like the north atlantic oscillation (NAO) or the madden julian oscillation (MJO). A detailed discussion abut the application in atmospheric science can be found e.g. in Hannachi et al. (2007). The EOF-analysis will be used here to investigate the vertical patterns, which are responsible for the greater probability of the DE-EPS.

$$\vec{x'} = \sum_{b=1}^{eof} (\vec{x'}^T \hat{e}_b)\hat{e}_b \tag{6.2}$$

Eq. (6.2) shows the vector $\vec{x'}$ described by a linear combination of EOFs. In this section the

EOF-analysis is done by systematically variation of this linear combinations from $eof = 1$ to $eof = 8$. Therefore, eof1 including only the first EOF ($eof = 1$), eof2 including the first two EOFs ($eof = 2$). Finally, eof8 including all EOFs ($eof = 8$), and the results of the Bayes factor are identical to Fig. 6.2 a) for the temperature and 6.3 a) for the equivalent potential temperature where three profiles are treated as an arithmetically averaged profile.

Figure 6.5 a) shows for the first EOF almost no difference between DE-EPS and SREPS. This means, both profiles, the profile of the DE-EPS and the profile of the SREPS, are equally probable when all vectors and matrices are expressed by the first EOF. The gain on probability for the DE-EPS results from with the higher EOFs. This issue underlines Fig. 6.5 c) where the Bayes factor is shown depending on the amount of considered EOFs. Finally, the DE-EPS becomes significantly more likely if at least five EOFs are used. This result can be connected to the higher resolution of the DE-EPS and shows that the DE-EPS is better appropriated to handle the fine vertical structures and therefore, the evolution of the temperature.

We have seen that the first EOFs are not enough to explain the difference between the DE-EPS and the SREPS. The EOF-analysis for the equivalent potential temperature (Fig. 6.5 b),d) shows no clear results like in the case for the temperature.

## 6.2 Significance of the results of the comparison

Until now a clear evidence for the DE-EPS has been shown. In this section, an investigation of the sensitivity and robustness of this result is presented. First, the sensitivity of the result with respect to the Silverman's factor and the observation uncertainty is investigated relating to the theoretical reflections in Chapter 4. Secondly, the main advantage of the Bayesian statistics is applied, which means the prior probability is varied to investigate the strength of the evidence.

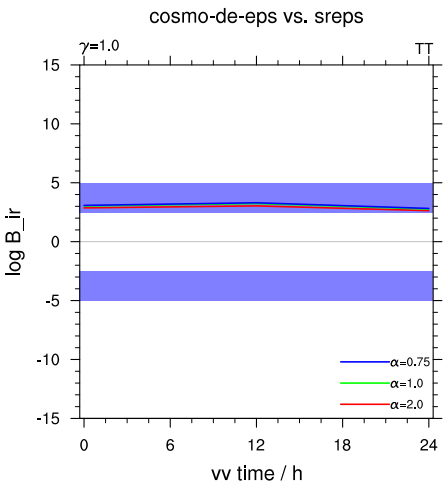
### Sensitivity to the Silverman's factor

The Silverman's factor was introduced in Chapter 4 and was named their as the smoothing parameter for PDFs. In this section the Silverman's factor is denoted as $\alpha$ defining the covariance matrix as

$$\tilde{\Sigma}_i = \alpha \cdot \Sigma_i \tag{6.3}$$

Figure 6.6 shows the sensitivity of the Bayes factor to the Silverman's factor. The sensitivity of the Bayes factor in view to the Silverman's factor is quite small. In case when the three profiles are averaged, there is almost no difference (Fig. 6.6 a). This belongs also for the case where the three profiles are put into one vector. However, for $\alpha = 2.0$ it is shown that the evident for the DE-EPS becomes weaker, but the evidence is still "decisive". Additionally, it should be mentioned that it was shown in Chapter 4 that $\alpha = 2.0$ leads to a quite strong smoothing of the

a)




b)

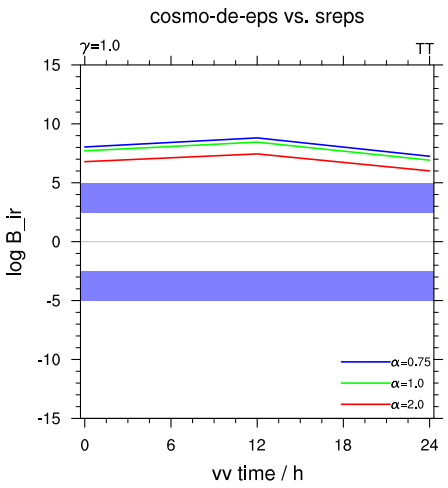



**Figure 6.6:** Time series of the Bayes factor of COSMO-DE-EPS for the temperature at forecast time (vv time) with respect to SREPS. (a) shows the case of three profile averaged and (b) shows the case of three profiles jointly together unaveraged. Different weighting factors $\alpha = 0.75, 1.0, 2.0$ for the covariance matrix are used.

PDF, which is not required in this work. All in all, this confirms the theoretical considerations from Chapter 4. The next step is the investigation of the sensitivity of the Bayes factor relating to the observation uncertainty where a much stronger influence was implied in contrast to the Silverman's factor.

### Sensitivity to the observation uncertainty

In the introduction (Chapter 1) and several times later (e.g. Chapter 5) it was mentioned that the observed state is by itself uncertain due to the limited capabilities of observing the atmosphere. In this work the radiosonde observations are placed into a single column of the NWP model as a function of height. However, for the COSMO-DE running at convection permitting scales, the drifting of the radiosonde is certainly not neglectable. As a result the radiosonde observations are erroneous beyond the standard instrumental error shown in Tab. 5.5, which has to be taken into account. This is done by using the observation errors of the 3dvar data assimilation system of the DWD. The sensitivity of the Bayes factor to the observation uncertainty is shown by multiplication of the observation error covariance matrix $\Sigma_o$ by a constant factor $\gamma$

$$\tilde{\Sigma}_o = \gamma \cdot \Sigma_o \tag{6.4}$$

The factor $\gamma$ is variated between $0.5 \leq \gamma \leq 2.0$. Values smaller than one are investigated, because of in the current IFS data assimilation scheme nowadays values smaller than those in the DWD scheme are used, which correspond approximately with $\gamma \approx 0.9$.

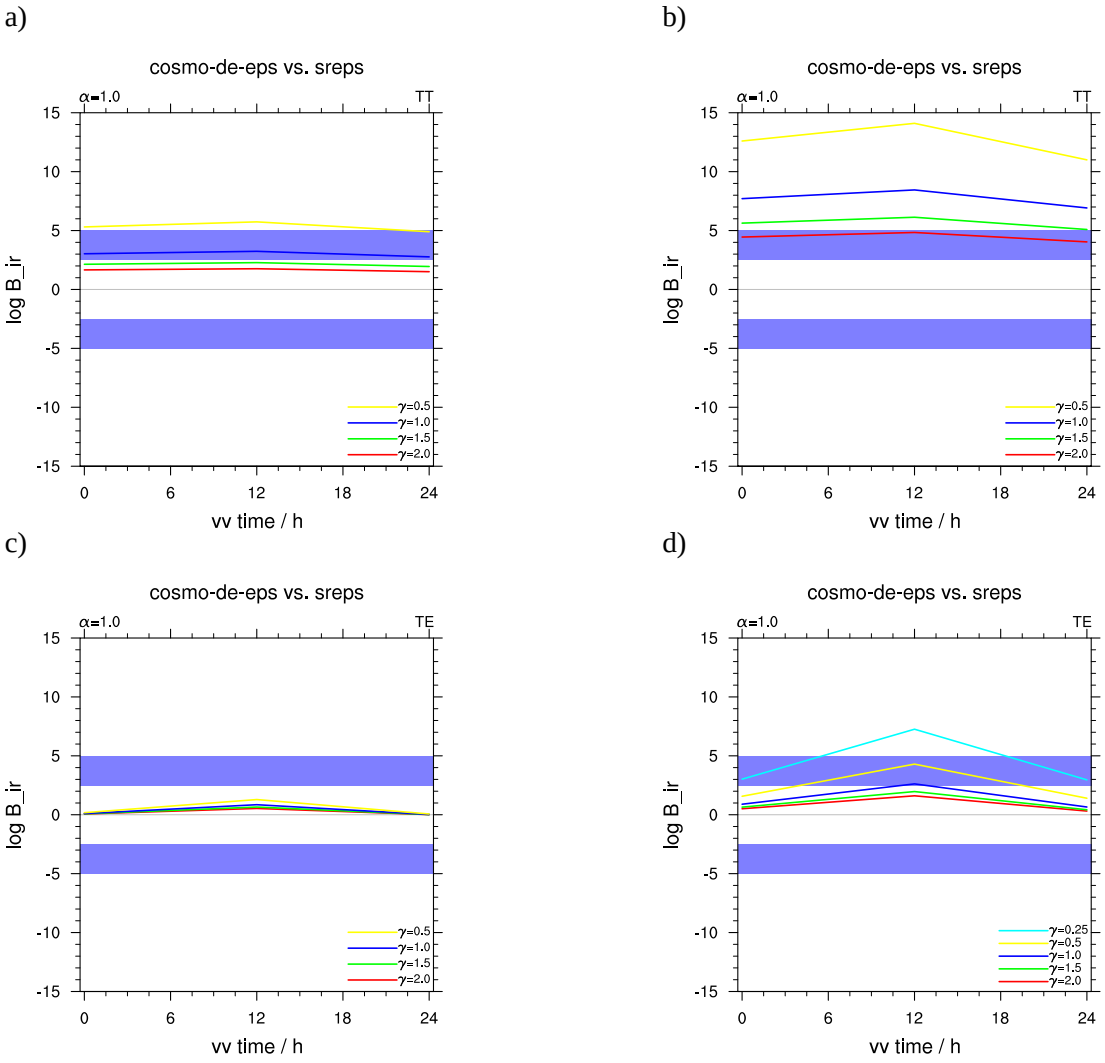



**Figure 6.7:** Time series of COSMO-DE-EPS at forecast time (vv time) with respect to SREPS. (a,b) shows the Bayes factor for the temperature and (c,b) for the the equivalent potential temperature. (a,c) shows the case of three profile averaged and (b,d) shows the case of three profiles jointly together unaveraged. Different weighting factors $\gamma = 1.0, 1.5, 2.0$ for the observational covariance matrix are used.

Figure 6.7 shows the sensitivity relating to the observation error, in each case (a,b) for the temperature and (c,d) for the equivalent potential temperature. In case of an increased observation error (increased values of $\gamma$), it is shown that the Bayes factor and therefore the evidence for the DE-EPS decrease with an increased observation error. Hence, the observations error is an index, which influence significantly the strength of the evidence. In case of a twice as large observation error, the evidence for the DE-EPS is only "substantial". This is caused by a smaller likelihood (Eq. 4.36) due to a wider PDF $p_l(o|f)$ given in Eq. (4.42). This shows that the observation error has a crucial effect on the result. But for the comparison of the DE-EPS with the SREPS the existing evidence for the DE-EPS can be further used, because the standard deviation of the radiosonde observations are realistically chosen by the values of the data

assimilation scheme.

For the equivalent potential temperature, the influence of the weighting factor $\gamma \geq 1.0$ is quite weak, because for the equivalent potential temperature the observation uncertainty is already large. Hence, an additional increase of the observation error has only weak effects in this case. However, it is shown that a decrease of the observation uncertainty has here a significant effect and lead to a "decisive" evident for the DE-EPS when $\gamma = 0.25$ [1] showing that if the observation variance is large this makes all details of the predicted PDF vanished. Only reliable enough observations allow an accurate evaluation of the ensembles.

Before as next step the sensitivity of the Bayes factor to the prior probability is discussed it should be mentioned that the sensitivity results corresponding to the Silverman's factor and to the observation uncertainty agree quit well with Chapter 4 and underline the importance of the observation uncertainty.

### Sensitivity of the Bayes factor to the prior probability

The Bayes factor for the comparison DE-EPS vs. SREPS was introduced in Chapter 4 as

$$B_{ir} = \frac{l(o|m_{\text{de-eps}})P(m_{\text{de-eps}})}{l(o|m_{\text{sreps}})P(m_{\text{sreps}})} \tag{6.5}$$

In which, the prior probabilities $P(m_{\text{de-eps}})$ and $P(m_{\text{sreps}})$ of the DE-EPS as well as of the SREPS were kept constant.

$$P(m_{\text{de-eps}}) = P(m_{\text{sreps}}) = 0.5 \tag{6.6}$$

This is an approximation for the case in which we are doubtful to what we have to belief. In this case no EPS is preferred corresponding to a uniform prior, which is also called "Laplace Prior". The Bayesian theory allows in this case to select a user-defined prior probability. But nevertheless, the data has been shown an evidence for the DE-EPS in view of the "Laplace Prior".

Furthermore, to test the robustness of this evidence for the DE-EPS, in the Bayesian statistics it is possible to vary the prior probability $P(m_i)$ for the models under investigation, here the DE-EPS and the SREPS. Because two models are compared, we get

$$P(m_{\text{de-eps}}) = 1 - P(m_{\text{sreps}}) \tag{6.7}$$

If $P(m_{\text{de-eps}})$ is larger [smaller] than $P(m_{\text{sreps}})$ then the personal belief in the DE-EPS, e.g. of a professional forecaster, is higher [lower] than for the SREPS.

[1] calculated only for the equivalent potential temperature in case of three profiles jointly together unaveraged

a) b)

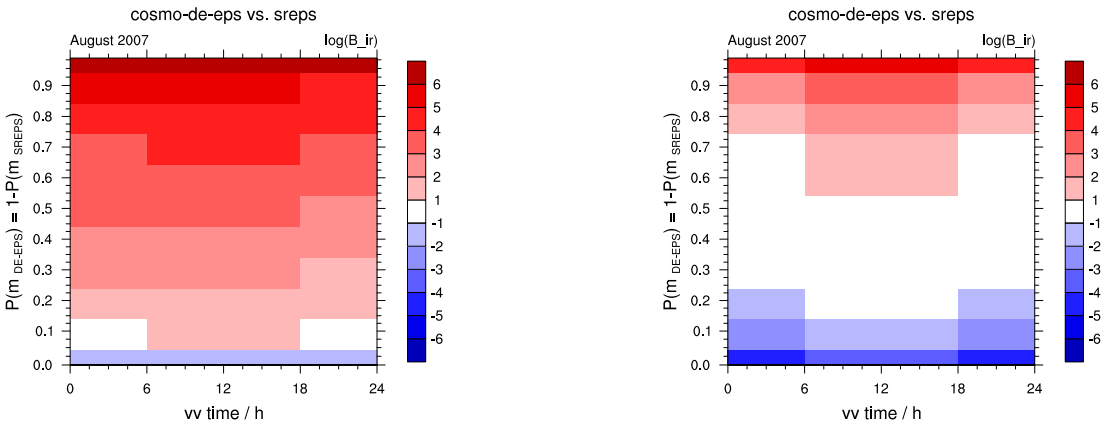



**Figure 6.8:** The distribution of the Bayes factor $logB_{ir}$ for DE-EPS ($m_{DE-EPS}$) and SREPS ($m_{SREPS}$) given the prior of $m_{DE-EPS}$ [$m_{SREPS}$] varies from 0.01 to 0.99 [from 0.99 to 0.01]. Figure a) shows the case for the temperature and (b) the case for the equivalent potential temperature.

The effect of varying the prior probabilities relating the Bayes factor and further the comparison of the DE-EPS with the SREPS for the temperature is shown in Fig. 6.8 a). The figure shows that the evidence of DE-EPS is higher than for the SREPS even if the prior probability of the DE-EPS is as small as 0.2. The previous case shown in Fig. 6.2 a) is included here for the case that both ensembles have the same prior probability $P(m_{DE-EPS}) = P(m_{SREPS}) = 0.5$.

The prior variation for the equivalent potential temperature is shown in Fig. 6.8 b). It is shown that the DE-EPS is more likely too. However, at most forecast times varying of the prior probability leads to a widely "neutral" Bayes factor. If the prior probability of the DE-EPS is quite small as 0.2, there is only a "neutral" evidence for the DE-EPS. This applies only in the case of the $12h$ forecast lead time. For the $0h$ and $24h$ forecast lead time, there is even a "substantial" evidence for the SREPS.

Summarized, it is figured out that the shown results (the evidence for the DE-EPS) are robust. Consequently, it seems the DE-EPS is appropriated to realize the vertical profiles, but mainly for the temperature. For the equivalent potential temperature, there is not such a clear evident for the DE-EPS.

### Sensitivity of the posterior probability to the prior probability

Figure 6.8 and all the previous figures have shown the Bayes factor $B_{ir}$, thus the ratio of the posterior probability of the DE-EPS to them of the SREPS. Furthermore, it is also important to have a look at the single posterior probabilities $P(m_{\text{de-eps}}|o)$ and $P(m_{\text{sreps}}|o)$. Because of that the posterior probabilities are shown in Fig. 6.9.

$$P(m_{\text{de-eps}}|o) = \frac{l(o|m_{\text{de-eps}})P(m_{\text{de-eps}})}{\sum_{j=1}^{N} l(o|m_j)P(m_j)} \tag{6.8}$$

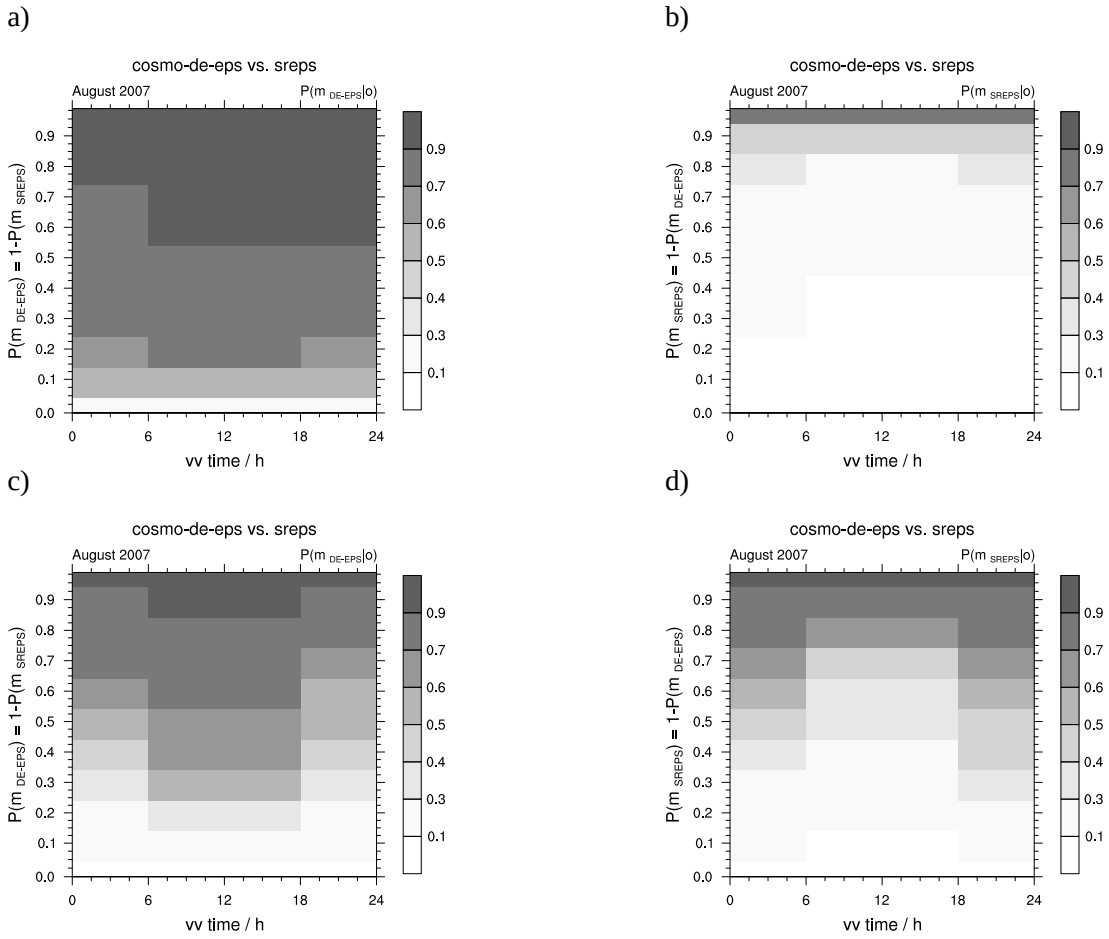



**Figure 6.9:** The distribution of the posterior probability $P(m_i|o)$ for DE-EPS ($m_{DE-EPS}$) and SREPS ($m_{SREPS}$) given the prior of $m_{DE-EPS}$ [$m_{SREPS}$] varies from 0.01 to 0.99 [from 0.99 to 0.01]. (a,b) shows $P(m_i|o)$ for the temperature and (c,d) for the equivalent potential temperature.

The figure shows a definite result of a higher posterior probability of the DE-EPS in contrast to the SREPS for this one month period. Quantified in values, the posterior probability values for the DE-EPS are largely between 0.7 and 0.9 and corresponding the posterior probability values of the SREPS are around 0.2. This regards for the temperature shown in Fig. 6.9 (a,b). Figure 6.9 (c,d) show the posterior probability for the equivalent potential temperature. The posterior probabilities for the DE-EPS are here weaker with values between 0.4 and 0.7 and therefore only just a little more likely than the SREPS.

Thus with the previous sensitivity studies relating the Silverman's factor, the observation uncertainty and the prior probability it was shown that the results for the DE-EPS are robust. In the following, the reasons for this results are investigated in more detail via a regime dependent investigation including a single case study.

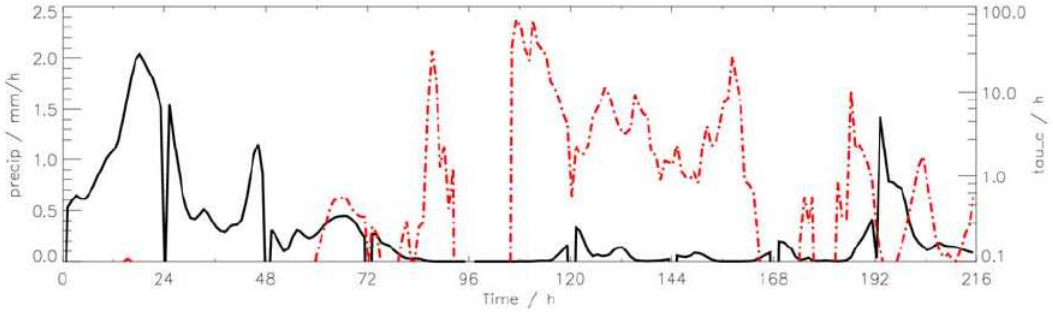



**Figure 6.10:** Time series of the ensemble mean total precipitation (solid line) and the mean convective time scale averaged over the COPS domain for the entire period from 8 to 16 August 2007 (Keil and Craig, 2011).

## 6.3 Regime-dependent comparison

In the last section, it was shown that there is an evidence for the DE-EPS both for the temperature and to a lesser degree for the equivalent potential temperature too. To learn more about the DE-EPS to forecast convection allowing conditions, a regime dependent comparison is a further option, which is applied in the following.

For this regime dependent comparison of the DE-EPS with the SREPS, the August 2007 period has to be spitted into different periods with respective prevailing weather regimes. This is done by the usage of the results of a convective adjustment time scale analysis from Keil and Craig (2011). A threshold value of convective time scale $\tau_c = 6h$ is used to distinguish between equilibrium and non-equilibrium convection (see Chapter 3). Finally, the Bayes factor is averaged over this periods taking into account the diurnal cycle.

Figure 6.10 shows the time series of the mean total precipitation in conjunction with the mean convective time scale averaged over the $360x360km^2$ COPS region. The period is split into three different episodes introduced in Tab. 6.2.

Figure 6.11 shows the Bayes factor averaged over the respective time periods specified in Tab. 6.2. The Bayes factors for the first period (8-10 August) and for the second period (12-14 August) are nearly identical. The evidence for the $12h$ temperature forecast of the DE-EPS is here "decisive" and so a little bit smaller than for the average over the whole August 2007. For the equivalent potential temperature, the evidence is "substantial" and in general, again weaker as for the temperature. Only for the third period (11, 15-16 August) the evidence is comparable to the whole month average. The regime-dependent comparison is done for the joint treatment of the three profiles to see clearer differences between the periods.

The regime-dependent comparison shows that there is a benefit of the DE-EPS for the forecast of the vertical conditions of convection. However, it is only possible to distinguish rudimental the quality of the vertical profiles in respect to the different convective weather regimes due to

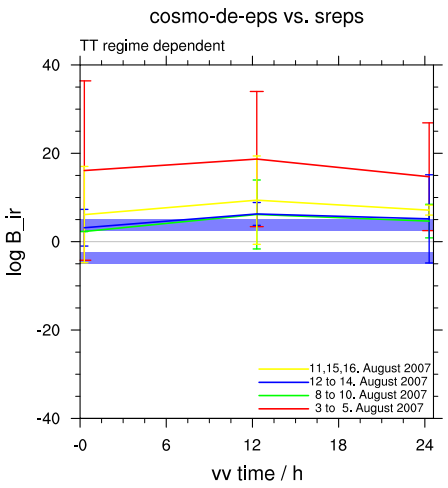
| Period | Synoptic weather situation |
|---|---|
| 8th - 10th August 2007 | [relevant forecast lead times: vv=00,12,24$h$]<br>*The weather situation is dominated by strong precipitation intensities and small convective time scales ($\tau_c < 1h$) due to a trough across central Europa leading to an easterly flow in the COPS region* (Keil and Craig, 2011). |
| 12th - 14th August 2007 | [relevant forecast lead time: vv=12$h$]<br>*The period is dominated by small mean precipitation amounts and a large mean convective time scale ($\tau_c \geq 6h$) indicating weakly foreced conditions at the synoptic scale. However, a short-term forced-frontal situation occurred in the night from 12 to 13 August indicated by a short-lived decrease in the convective time scale* (Keil and Craig, 2011). |
| 11th, 15th - 16th August 2007 | [relevant forecast lead time: vv=12$h$]<br>*The meteorological conditions were dominated by synoptic-scale disturbances crossing the COPS region and leading to a synoptically forced regime ($\tau_c < 6h$).* |
| 3rd - 5th August 2007 | [relevant forecast lead times: vv=00,12,24$h$]<br>*A trough from England is moving very fast eastwards.*<br>*It follows a marked ridge moving also very fast eastwards.*<br>*In central Europe, between the next trough in the west and the ridge in the east, extremely warm air is advected by the synoptic flow.*<br>The end of this synoptic evolution (5th August 2007, 00UTC) is shown in Fig. 6.13. |

**Table 6.2:** Time periods for the regime depended comparison. The respectively forecast lead time (vv) shows the forecast time of interest.

the small sample of cases. Nevertheless, the period (11, 15-16th August, $\tau_c < 6h$), shows that the only typical strongly forced convection period has the largest evident for the temperature of DE-EPS without the 3-5th August. This shows that the DE-EPS vertical temperature profiles here are clearly more likely as those of the SREPS anyway.

For the period from the 3-5th August, which is not included in Keil and Craig (2011), the DE-EPS is significantly more likely than the SREPS for the temperature as well as for the equivalent potential temperature. The reason for this is a misprediction of the vertical structure of the atmosphere from the SREPS. The passage of a ridge shown in Fig. 6.13 a) is predicted too early leading to a significant too warm temperature Bias at nearly all levels up to 200$hPa$ shown in Fig. 6.12 b) whereas the DE-EPS is almost Bias free (Fig. 6.12 a). Figure 6.13 b) shows the corresponding advection of warm air (above 15°C at 850$hPa$) into the COPS region from the southern part of France. This misprediction is due to the initialization of the SREPS using no analysis, but rather 12$h$ forecasts of the global models as initial state. This has to be kept in mind to interpret the clear result for the DE-EPS based on the Bayes factor.

a)




b)

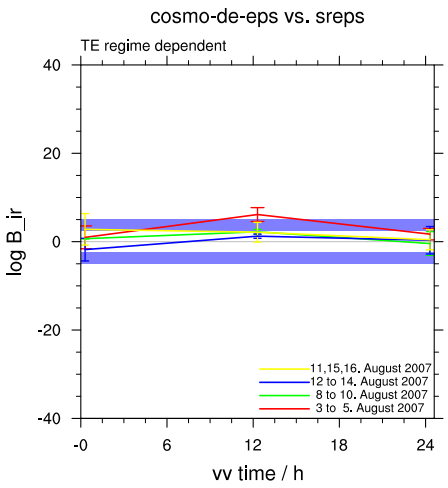
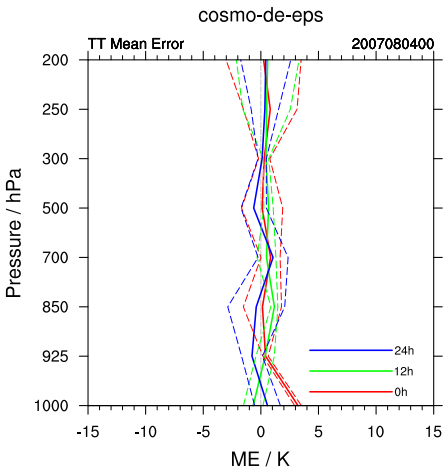



**Figure 6.11:** Time series of the log Bayes factor for three time periods selected by the convective time scale averaged over the COPS domain for four time periods. The first from 8 to 10 August, the second from 12 to 14 August and the third 11, 15 to 16 August (Keil and Craig, 2011). The fourth time period (3 to 5 August) is selected due to the large Bayes factors. Figure (a) shows the case for the temperature and (b) the case for the equivalent potential temperature. The standard deviation is shown by the error bars

a)




b)

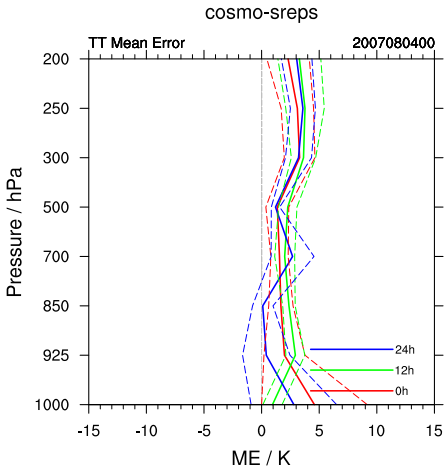



**Figure 6.12:** Vertical bias of the mean temperature of the DE-EPS (a) and of the SREPS (b) for 0*h*, 12*h* and 24*h* forecast lead time at the 4th August 2007. The dashed lines show the standard deviation plus minus the mean at the respective forecast lead times.

## Case study of single COPS IOPs

Additionally to the previous regime-dependent investigation the Bayes factor is used for a single case study. Up to now, a limiting factor for the temporal resolution was the radiosonde launch-times. The COPS campaign gives an excellent opportunity to use temporal highly resolved radiosonde ascent data. For COPS additional radiosondes were ascented at additional

a) b)

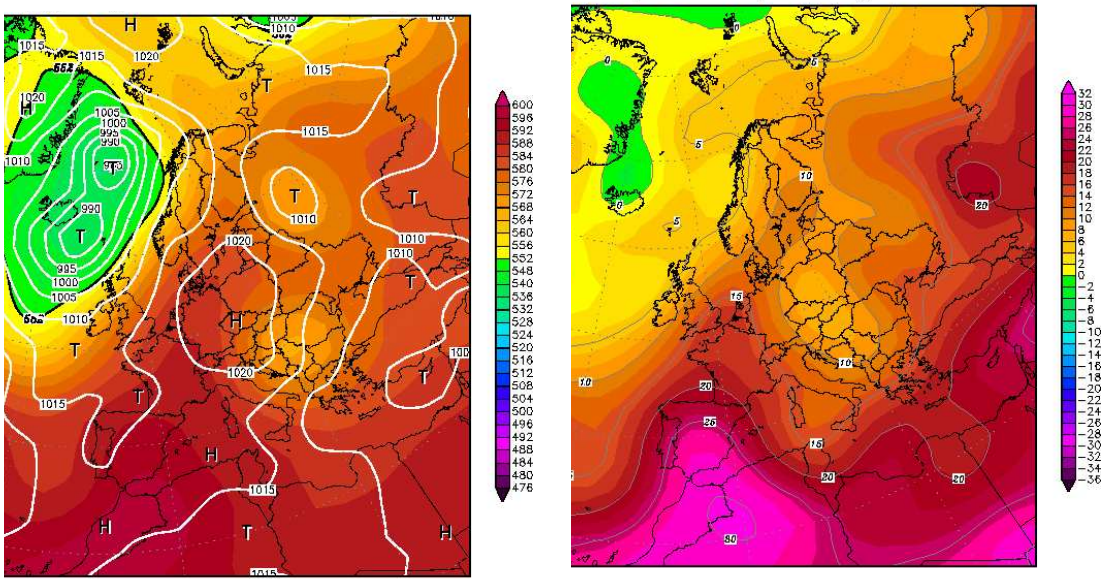


**Figure 6.13:** NCEP Reanalysis of the synoptic weather situation at the 5th August 2007 00UTC. The geopotential at 500*hPa* and the corresponding surface pressure is shown at figure (a) and the temperature is shown at the pressure-level 850*hPa* figure (b) (Wetterzentrale, 2011).

places and with a higher launching frequency (see Chapter 5).

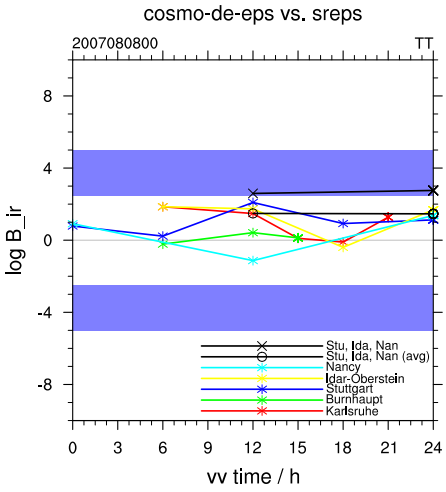
These data will be used for a comparison with higher temporal and spacious resolution inside the COPS area of the DE-EPS with the SREPS. The Bayes factor is presented separately for each COPS radiosonde station as well as for the three previous stations.

Figure 6.14 shows the comparison of the DE-EPS with the SREPS for the 8th August 2007 unaveraged as a case study. The black line shows once, the average of the three stations Stuttgart, Nancy and Idar-Oberstein and the joint treatment of the three stations. Furthermore, the stations Stuttgart, Nancy and Idar-Oberstein are also shown separately. The COPS stations in this case are Karlsruhe and Burnhaupt. In general, the COPS stations fit well with the DWD stations for the temperature (Fig. 6.14 a) as well as for the equivalent potential temperature (b). For the temperature only Nancy shows at the 12*h* forecast a "substantial" evidence for the SREPS. All other stations show a "neutral" to "strong" evidence for the DE-EPS. However, the "strong" evidence for the DE-EPS occurs only in the case where the three stations are treated jointly.

The case study shows the great variability of the quality of vertical profiles at small space. This explains the difficult task to forecast convection. The average over several profiles (temporal and spatial) are free of a Bias, e.g. see vertical profiles of the DE-EPS (Fig. 6.4). However, the single forecasted events can extremely differ from the observations.

For a complete reflexion about the case study, the univariate PDFs for the temperature and for the equivalent potential temperature at 850*hPa* are shown in Fig. 6.15. The PDF of the

a)




b)

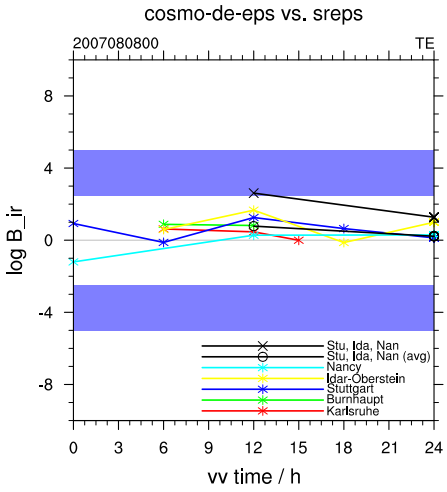



**Figure 6.14:** Time series of the log Bayes factor for one day (8th August 2007). Figure (a) shows the case for the temperature and (b) the case for the equivalent potential temperature.

temperature shows less spread for both ensembles in comparison to the observation uncertainty at vv=0*h*. The spread grows with increased forecast lead time, but both ensembles are not able to represent the observation uncertainty after 24*h* forecast time. For the equivalent potential temperature, the spread is fundamentally larger, but only after 24*h* approximately similar. It is also shown that the observation uncertainty for the calculated equivalent potential temperature is quite high, which indicates that it is difficult to get clear evidences for one ensemble relating this comparison. So further ways to handle the non-gaussian humidity has to be investigated.

In the next section, the Bayesian results are compared with other probabilistic scores. In order to assess if the results are to a certain degree comparable or not.

## 6.4 Comparison with other probabilistic scores

The comparison of the Bayes factor with other scores is done in this section to verify the previous results with further probabilistic scores. For this, the Bayes factor is compared with the *IGN* and *CRPS* (see Chapter 2 or rather 4), which are presented as skill scores. Furthermore, the comparison is done univariat and at the same points where the Bayes factor was calculated. This requires to go back to a univariate contemplation of the forecasted variables represented here by the temperature at the 850*hPa* level.

### Ignorance score

Figure 6.16 a) shows the Ignorance skill score (*IGNSS*) for the DE-EPS with the SREPS as reference ensemble and (b) shows the corresponding Bayes factor. The investigated variable

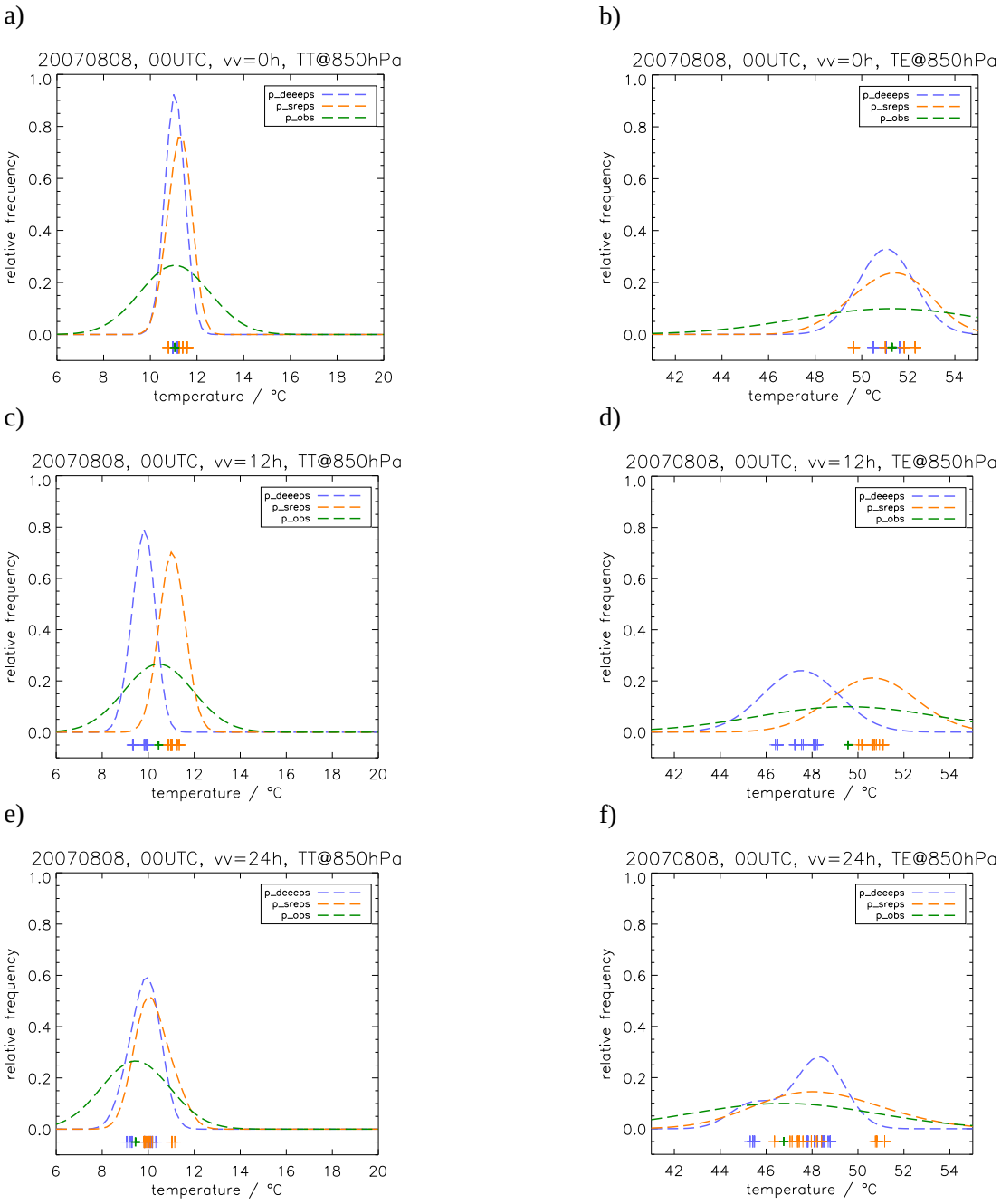



**Figure 6.15:** Univariate temperature (a,c,e) and equivalent potential temperature (b,d,f) PDFs of DE-EPS and SREPS for one level (850*hPa*) for different forecast lead time vv=0*h* (a,b), vv=12*h* (c,d) and vv=24*h* (e,f) at gridpoint Stuttgart.

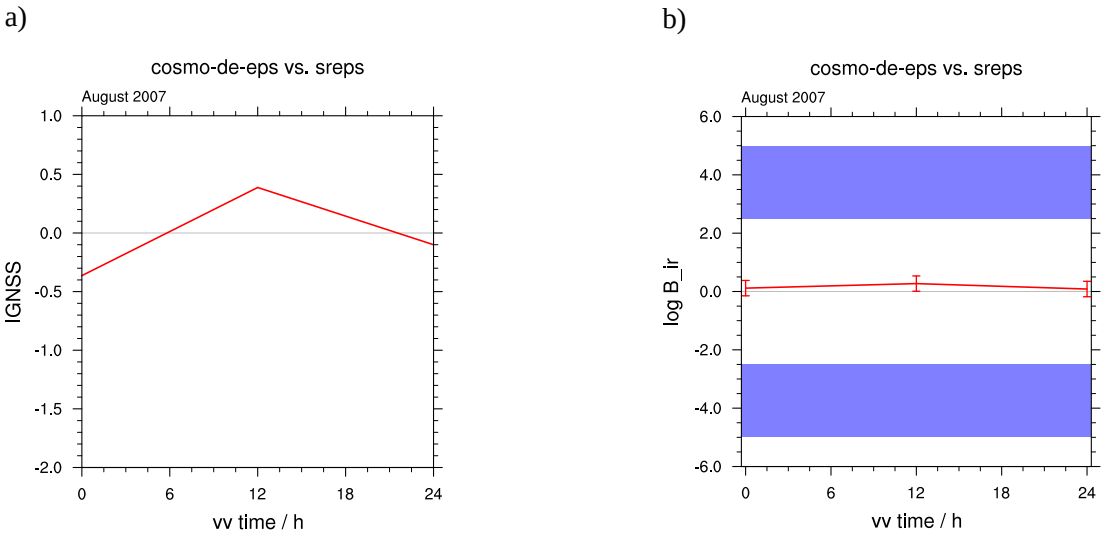



**Figure 6.16:** Time series of the Ignorance skill score (*IGNSS*) (a) and of the Bayes factor (b) for the temperature at 850*hPa*.

is the temperature at 850*hPa*. In contrast to the Bayes factor, the *IGNSS* is normed to one. The results of both scores show differences. Both have their maximum at the 12*h* forecast lead time, which means at this time the DE-EPS is most likely, but the *IGNSS* has additional negative values at the 0*h* and 24*h* forecast lead times. However, it has to be considered, the positive and the negative scales of a skill score are not comparable, because the skill score is bounded above, but not downward. Hence, this distorts the result and in addition the *IGNSS* does not consider the observation uncertainty, which is the reason for the extremely ”neutral” evidence trough the Bayes factor.

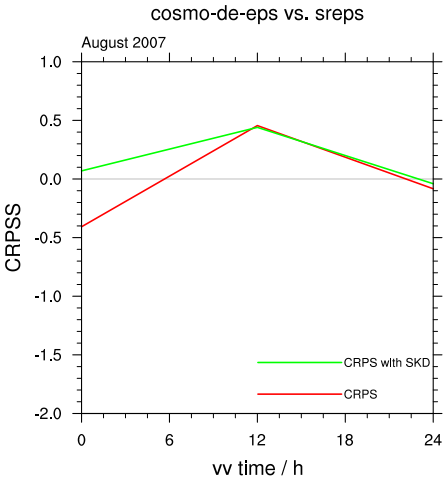
Furthermore, the Ignorance score considers only one point of the ensemble PDF (see Chapter 2) while the Bayesian approach allows to consider the full PDFs of the observation as well as of the EPS. Therefore, this comparison is not sufficient. The Bayes factor has to be compared with another probabilistic score.

### CRPS

A more sophisticated probabilistic score is the *CRPS* skill score (*CRPSS*). The comparison of the DE-EPS with the SREPS using the *CRPSS* is shown in Fig. 6.17. The *CRPSS* fits very well together with the *IGNSS*. Consequently, the *IGN*, the *CRPS* and the Bayes factor have similar characteristics. The *CRPSS* is shown for two cases. The first case, the red line, shows the standard *CRPSS* were the ensemble is represented by one normal distribution with the two parameters ensemble mean and standard deviation. In this case, the *CRPSS* agrees quite well with the *IGNSS*. The second case, the green line, represents the ensemble trough the extended SKD (see Eq. 4.63). Here, the *CRPSS* agrees better with the Bayes factor than with the *IGNSS*.

For the 0*h* and 24*h* forecast lead times, the *CRPSS* and the *IGNSS* show that it is not possi-

a)




b)

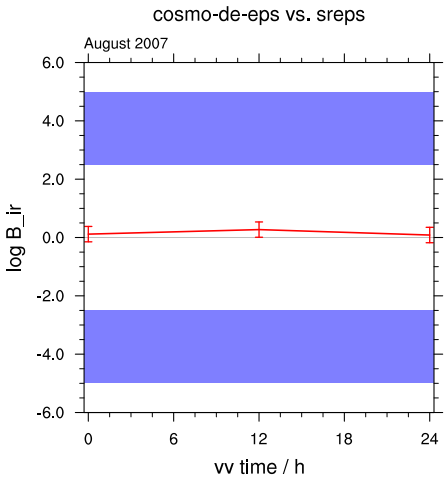



**Figure 6.17:** Time series of the Continuous ranked probability skill score (*CRPSS*). The red line described the ensemble by a normal distribution trough mean and standard deviation and the green line trough the SKD (a) and of the Bayes factor (b) for the temperature at 850*hPa*.

ble to decide, which ensemble, the DE-EPS or the SREPS, is more likely. This is the same result, which we got from the Bayes factor. Furthermore, both skill scores show for the 12*h* forecast that the DE-EPS is more likely. The Bayes factor shows the same result in principle, but the evidence is ”neutral”, because the Bayes factor considers additionally the observation uncertainty, which the others do not.

In general, all probabilistic scores including the Bayes factor show an almost ”neutral” evidence, which prefers no ensemble. In contrast, when using the eight pressure levels for the Bayes factor a "strong" to "decisive" evidence has be seen.

## 6.5 Summary and conclusion

The aim of the comparison of the DE-EPS with the SREPS in this chapter was to investigate if there is an improvement of the forecasts of the higher resolved DE-EPS to the courser SREPS relating convection permitting conditions. Therefore, a multivariate probabilistic verification method based on Bayesian statistics was used to compare two ensemble forecasts with each other relating the vertical temperature and equivalent potential temperature profiles. In which the model as well as the observation uncertainty was considered.

It was shown that the forecasted temperature profiles of the DE-EPS are much more likely than those of the SREPS even if the prior belief (e.g. of a professional forecaster) in the DE-EPS might be as low as 0.2, the posterior probability for the DE-EPS is anyhow about 0.8. Furthermore, the equivalent potential temperature was analysed to learn more about the predictability of the convective potential. The equivalent potential temperature is important for the vertical stability of the atmosphere including the impact of humidity. The DE-EPS profiles are also

more likely for the equivalent potential temperature, but it was shown that the observation uncertainty diminishes here the strength of the evidence. At the 0 hour and at the 24 hour forecast lead times, there was no evidences that one ensemble is better than the other. This result shows that the short range ensemble weather forecasts from the convection permitting COSMO-DE-EPS are a valid and useful way to quantify the uncertainty of short range weather forecasts at least for the hindcasts performed for August 2007. However, a longer period of investigations seems necessary to underline this result. This is mainly figured out by the large standard deviation, which covers over several significance levels (Tab. 4.2). Additionally, it has to be considered that until now, only profiles at three radiosonde stations within the COPS area were investigated.

In spite of the larger probability of the DE-EPS, the single case study has shown that there are large variabilities on small areas (the COPS domain) in the evidences for the probability of the vertical profiles. This shows that it is an ongoing challenge also for the development of ensembles to consider this small scale variabilities. Another important point is that in this study only the forecasts of the convection permitting conditions are investigated. The initiation of convection is another important point, which was not investigated here. This point is especially for convection permitting NWP models a great challenge, because the model has to simulate explicitly the convection initiation without parametrization. The COPS campaign has also shown that there is the preference to work on, because the vertical profiles looking promising when the NWP model uncertainty is considered.

Summarized, the issue of observation uncertainty is a very important point, because it is impossible to measure the exact truth state of the atmosphere. It was shown, reliable observation are needed to get meaningful results. Finally, we have seen that the DE-EPS is able to forecast more precisely vertical profiles, but it has also limits in the potential to forecast the exact conditions of convection.

# Chapter 7

# Verification of ensemble prediction systems

In the previous chapter two ensembles were compared. The application of verification of ensembles is presented in this chapter. However, it is again a comparison of respectively two ensembles. This means in detail that the ensemble, which has to be verified is compared with an analysis of the corresponding forecast lead time. The analysis will be treated as the "true state", but with the extension of consideration of the uncertainty of this "true state". For the Bayesian verification, the analysis has to be considered as an artificial one member ensemble.

The SREPS and the LEPS are verified for only three Levels ($850, 700, 500 hPa$) due to missing data. The temperature and the equivalent potential temperature are investigated. Furthermore, the investigation contains again a regime dependent verification and a detailed analysis of a case study where eight levels are verified.

## 7.1 Verification of SREPS and LEPS over the whole COPS period

The verification results in this section are averaged over the whole COPS period (three months, see Tab. 7.1). The Bayes factor contains three levels for this verification. Furthermore, the reference model is an analysis from COSMO-EU for the verification considered as an artificial ensemble of analysis. This means, the ensemble mean of the artificial ensemble is the actual value of the analysis. However the standard deviation has to be estimated. The analysis of the COSMO models is done by a nudging scheme (Schraff and Hess, 2003). For this nudging scheme, it is difficult to estimate the analysis uncertainty. One way would be to use an ensemble analysis e.g. from an ensemble kalman filter (EnKF). However, this is not possible in this work. Hence, the estimation of the standard deviation is done via several ways, which are compared with each other to get an understanding about the influence of the analysis uncertainty for the application of verification. This approaches are described in the following.

| COPS IOPs | Basic Data | Weather type |
|---|---|---|
| | 9. June 2007 | |
| | 10. June 2007 | |
| | 11. June 2007 | |
| | 13. June 2007 | |
| IOP-3 a | 14. June 2007 | WFC |
| IOP-3 b | 15. June 2007 | SFC |
| | 24. June 2007 | |
| | 25. June 2007 | |
| | 27. June 2007 | |
| | 28. June 2007 | |
| | 30. June 2007 | |
| | 11. July 2007 | |
| | 13. July 2007 | |
| IOP-8 b | 15. July 2007 | AMC |
| | 16. July 2007 | |
| | 17. July 2007 | |
| IOP-9 a | 18. July 2007 | SFC |
| IOP-9 b | 19. July 2007 | SFC |
| | 21. July 2007 | |
| IOP-10 | 23. July 2007 | SFC |
| | 24. July 2007 | |
| IOP-11 a | 25. July 2007 | AMC |
| IOP-13 a | 1. Aug. 2007 | AMC |
| IOP-13 b | 2. Aug. 2007 | SFC |
| | 3. Aug. 2007 | |
| | 4. Aug. 2007 | |
| | 5. Aug. 2007 | |
| IOP-14 b | 8. Aug. 2007 | WFC |
| | 9. Aug. 2007 | |
| | 10. Aug. 2007 | |
| | 11. Aug. 2007 | |
| IOP-15 a | 12. Aug. 2007 | AMC |
| IOP-15 b | 13. Aug. 2007 | WFC |
| | 14. Aug. 2007 | |
| IOP-16 | 15. Aug. 2007 | SFC |
| IOP-16 | 16. Aug. 2007 | SFC |

| | | |
|---|---|---|
| | 18. Aug. 2007 | |
| | 20. Aug. 2007 | |
| IOP-17 a | 21. Aug. 2007 | WFC |
| IOP-17 b | 22. Aug. 2007 | WFC |
| | 27. Aug. 2007 | |
| | 29. Aug. 2007 | |
| | 30. Aug. 2007 | |
| | 43 days | |

**Table 7.1:** Basic data of the verification of the SREPS and LEPS against COSMO-EU analysis. Overall, 61 days of the COPS period 2007 are available with corresponding weather type classification (AMC: Air-mass convection, WFC: weakly forced convection, SFC: strongly forced convection) after (Wulfmeyer et al., 2011).

### Standard deviation of the artificial COSMO analysis ensemble

The first approach to estimate the standard deviation of the COSMO-EU analysis is derived from the data assimilation and describes the variance of an analysis error $\sigma_a^2$ for a least-square analysis. Bouttier and Courtier (1999) show that in this case the standard deviation is determined by the sum of the inverses of the observation error variance $\sigma_o^2$ and of the background error variance $\sigma_b^2$.

$$\frac{1}{\sigma_a^2} = \frac{1}{\sigma_o^2} + \frac{1}{\sigma_b^2} \tag{7.1}$$

This formula shows that the analysis error variance is smaller than each contributory variances. In this work, it is assumed that the analysis uncertainty is close to the observation uncertainty, which is known. Furthermore, it is assumed, that the background error is equal to the observation error $\sigma_b = \sigma_o$. This approximation is applied due to the circumstance that the background error can be assumed normally to be of the same order as the observation error. Finally, the analysis uncertainty is given by

$$\sigma_a^2 = \frac{1}{2}\sigma_o^2 \tag{7.2}$$

The standard deviation of the analysis is then equal to the half of the standard deviation of the observations, and consequently smaller as the observation uncertainty. However, additional approaches have to be used, because Eq. (7.2) is only an approximation of the uncertainty for the nudging analysis of the COSMO model. Against that, Eq. (7.1) describes the real analysis error of a least-square data assimilation scheme like 3-d variational (3dvar) or rather 4-d variational

(4dvar) technique, which are based on the minimization of a so called "cost function" (Bouttier and Courtier, 1999).

The second approach in this study is the simple usage of the covariance matrix of the ensemble, which has to be verified. In this case, first, the covariance matrix of the corresponding ensemble is calculated for the initial time (vv=0*h*) and then, the covariance matrix is kept constant for the verification time range of the respective run for the reference model. Finally, the third approach uses also the covariance matrix of the corresponding ensemble. But the covariance matrix is used from the corresponding ensemble calculated for the forecast lead time (vv). Table 7.2 shows an overview of these approaches.

| | Mean value | Standard deviation |
|---|---|---|
| APPROACH 1 | $\mu_a = f_{\text{cosmo}-\text{eu}-\text{analysis}}$ | $\sigma_a^2 = \frac{1}{2}\sigma_o^2$ |
| APPROACH 2 | $\mu_a = f_{\text{cosmo}-\text{eu}-\text{analysis}}$ | $\sigma_a^2 = \sigma^2_{\text{sreps(vv=0h)}}$ |
| APPROACH 3 | $\mu_a = f_{\text{cosmo}-\text{eu}-\text{analysis}}$ | $\sigma_a^2 = \sigma^2_{\text{sreps(vv)}}$ |

**Table 7.2:** Overview about the different approaches used for the standard deviation of the artificial COSMO analysis ensemble. The overview is given for the SREPS, but belongs also for the LEPS.

In the following, the verification results of the SREPS with this three estimation variants for the standard deviation are shown.

### Verification of SREPS with different standard deviation approximations

Figure 7.1 shows the verification of the SREPS over the whole COPS period for APPROACH 1. The verification is applied to the complete 72 hour forecasts of the SREPS. At the initialization time, the Bayes factor is slightly positive. This shows that there is a small evidence against the COSMO-EU analysis at initial time and hence, a small evidence for the SREPS initial state. This result is in contrast to the fact that the SREPS has no own data assimilation and is started from dynamical downscaled forecasts of global models. However, the SREPS standard deviation is larger as the approximated standard deviation of the COSMO-EU analysis at the initial time leading to this verification result. However later, after 12 hours the evidence for the SREPS and corresponding against the COSMO-EU analysis is vanished. With advanced lead time, the evidence against the SREPS increases and reaches a Bayes factor of about $B_{ir} = -2.0$ after 72 hours forecast time. This is equivalent to a forecast quality of a "medium level of confidence" according to Tab. 4.2.

The results using the other approaches to estimate the covariance matrices (APPROACH 2 and 3) are shown in Fig. 7.2. In this alternative cases, the covariance matrix is used from the ensemble itself, which has to be verified. Figure 7.2 a) shows the case where the covariance matrix at the initial time is kept constant (vv=0*h*) and Fig. 7.2 b) where the covariance matrix is used of the corresponding forecast lead time (vv). Both approximations fit essentially well together. At the first few hours, they are consistently nearly identical. However, for APPROACH 3, the

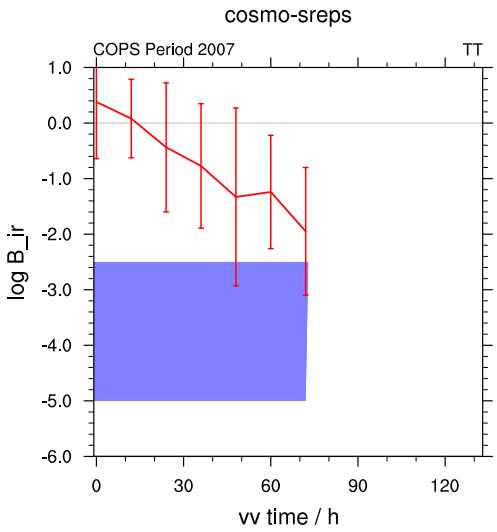



**Figure 7.1:** Time series of the Bayes factor $log\,B_{ir}$ for the verification of the SREPS with respect to the temperature at forecast time (vv time) against COSMO-EU analysis at the same time. The red line shows the Bayes factor for three stations (Stuttgart, Idar-Oberstein and Nancy). The blue band describe the area where the forecast becomes of a low level of confidence, see Tab. 4.2. The mean (solid line) and the standard deviation (error bars) of the COPS period are shown. The standard deviation of the COSMO-EU analysis is estimated after APPROACH 1.

evidence against the SREPS is only of "medium level of confidence" while in APPROACH 2, the evidence is of "low level of confidence" at the end of the forecast range. But in general, the order of magnitude of all three approaches fit together.

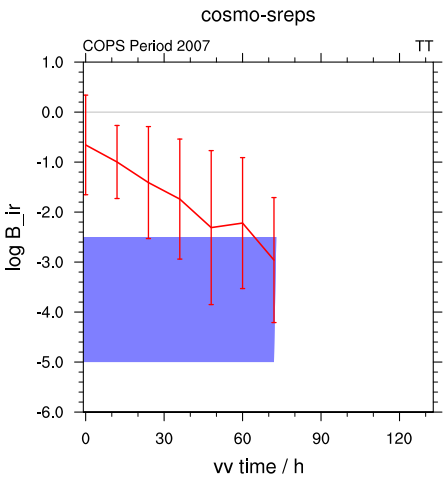
In the following, APPROACH 1 is used. This approximation allows further a comparison of the SREPS with the LEPS, because the covariance matrix of the ensemble under investigation is not used as covariance matrix for the reference ensemble. The disadvantage of this approximation could be that the approximated standard deviation of the analysis might be to small. This fact could give the SREPS and LEPS the advantage to have slightly better verification results, because they have a greater uncertainty already in the initial conditions. For the comparison of SREPS and LEPS, this disadvantage is not so important, because both ensemble are treated equally.

### Verification of COSMO-SREPS and COSMO-LEPS

In this section, firstly, verification results of the SREPS and the LEPS are presented together, so that they can be compared with each other. The covariance matrix of the analysis is approximated following APPROACH 1 of Tab. 7.2. The investigation contains again three profiles. The three stations are averaged arithmetically and in addition they are treated together in one vector unaveraged.

Figure 7.3 shows the temperature verification results of the SREPS on the left side (a,c) and

a)




b)

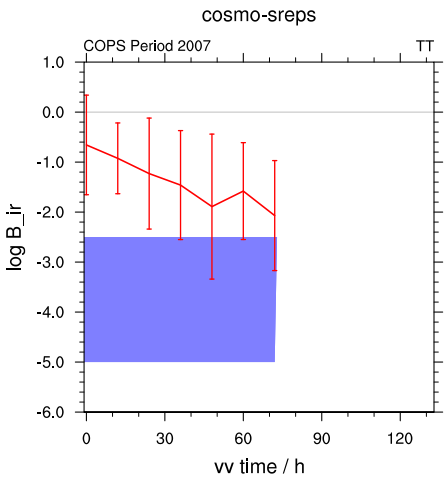



**Figure 7.2:** Time series of the Bayes factor $log\,B_{ir}$ for the verification of the SREPS with respect to the temperature at forecast time (vv time) against COSMO-EU analysis at the same time. The mean (solid lines) and the standard deviation (error bars) of the COPS period are shown. The standard deviation of the COSMO-EU analysis is estimated after APPROACH 2 (a) and APPROACH 3 (b).

of the LEPS on the right side (b,d). The verification results of the SREPS agree quite well with those of the LEPS. The LEPS allows additionally to verify a longer forecast range up to 132 hours. For both ensembles, the forecast quality is of "high to medium level" of confidence within the first 72 hours, when the three profiles are treated jointly in one vector unaveraged. After 132h, the LEPS reaches a "low level of confidence". Furthermore, the Bayes factor decreases linearly with an increased lead time. This linear decrease is in contrast to the almost constant progress of the Bayes factor for the comparison of the DE-EPS with the SREPS in Chapter 6 (e.g. Fig. 6.2), which has shown that the DE-EPS is more likely over all forecast lead times constantly. The verification of both ensembles shows that the main difference between the cases of three stations together unaveraged and three stations averaged is the strength of decrease of the forecast quality. This decrease is much stronger in case of verifying three stations together unaveraged and corresponds with the stronger evident for the DE-EPS in Chapter 6 in case where the three stations are investigated jointly too. The jointly investigation shows also in case of verification clearer results, but with the disadvantage of an increased standard deviation. The standard deviation has to be considered indicating that also when averaged over the whole COPS period there is a large variability inside the different ensemble runs. This applies mainly for the three stations together unaveraged. There the standard deviation is clearly larger as those for the three stations averaged. This indicates among others the need for a longer time period.

In Fig. 7.3 (c,d) are again the prior probabilities varied to get an idea about the robustness of the results.

$$P(m_i) = 1 - P(m_{ana}) \tag{7.3}$$

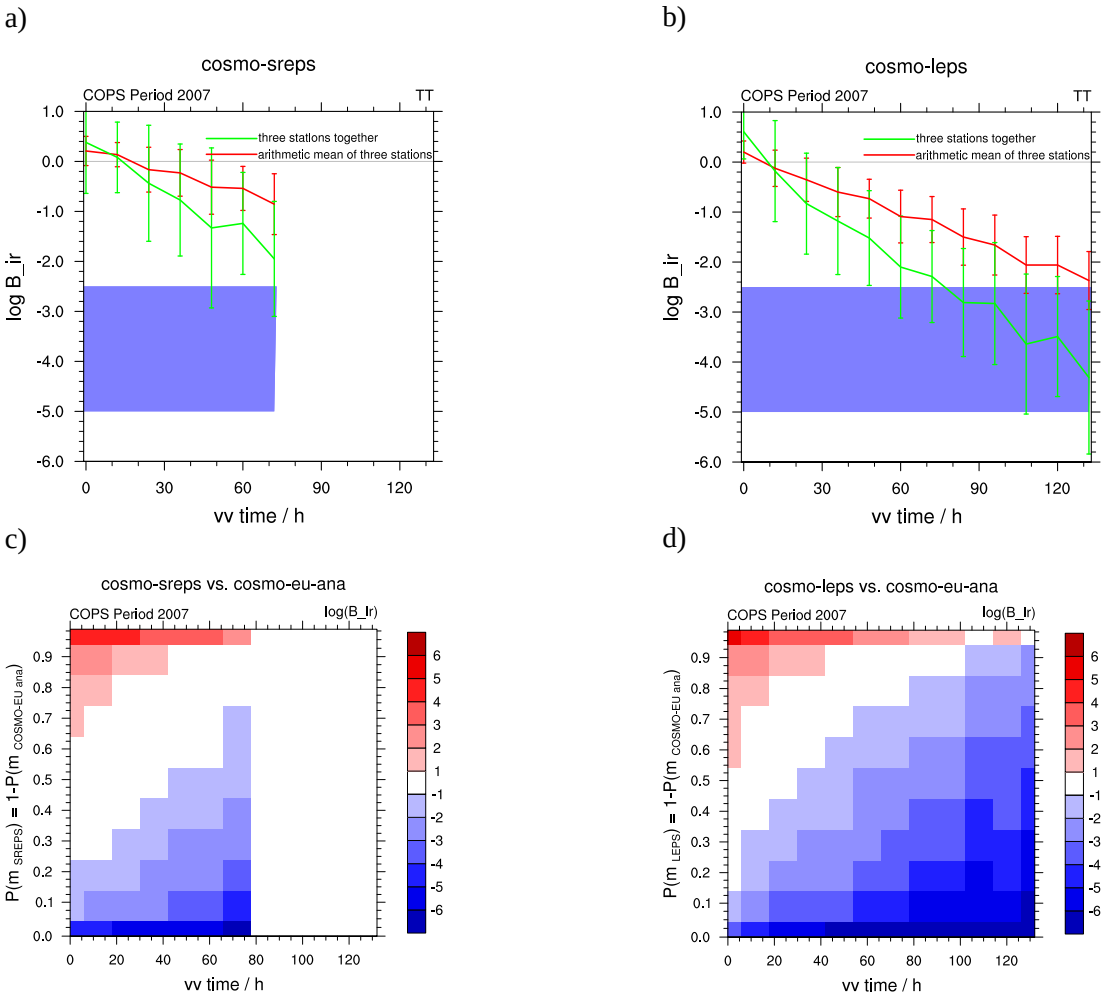



**Figure 7.3:** The distribution of the Bayes factor $log B_{ir}$ for the SREPS (a,c) and the LEPS (b,d) respectively against COSMO-EU analysis. For the SREPS, the SREPS ($m_{SREPS}$) and COSMO-EU analysis ($m_{CEUana}$) given the prior of $m_{SREPS}$ [$m_{CEUana}$] varies from 0.01 to 0.99 [from 0.99 to 0.01]. The same applies to LEPS.

Also here, both ensembles agree again very well with each other. A subjective determined prior probability e.g. for the COSMO-EU analysis of $P(m_{\text{ana}}) = 1 - P(m_{\text{sreps}}) = 0.2$ shows that the verification results are absolute robust, which means, even in the case when the the prior for the COSMO-EU analysis is to be assumed as quite small, the COSMO-EU analysis is much more likely as the SREPS forecast. The standard prior is again the Laplace prior as in Chapter 6.

$$P(m_{\text{ana}}) = P(m_{\text{sreps}}) = P(m_{\text{leps}}) = 0.5 \tag{7.4}$$

The verification of the equivalent potential temperature is shown in Fig. 7.4. The representation of the results is the same as in Fig. 7.3. The verification results for the equivalent potential temperature are quite similar to those of the temperature, which means the SREPS and LEPS fit again very well with each other. However, the decrease of evidence against the SREPS,

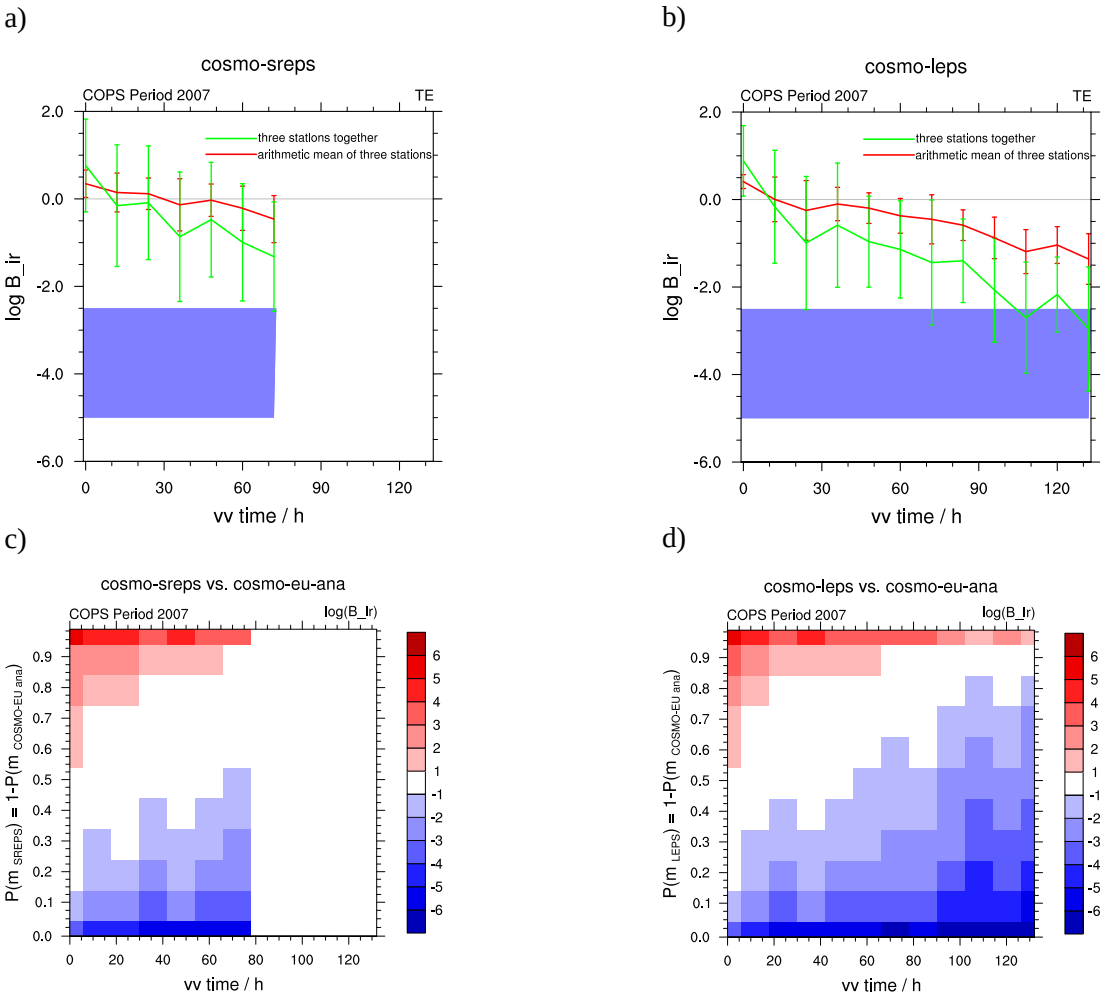



**Figure 7.4:** The distribution of the Bayes factor $log\,B_{ir}$ for the SREPS (a,c) and the LEPS (b,d) respectively against COSMO-EU analysis. For the SREPS, the SREPS ($m_{SREPS}$) and COSMO-EU analysis ($m_{CEUana}$) given the prior of $m_{SREPS}$ [$m_{CEUana}$] varies from 0.01 to 0.99 [from 0.99 to 0.01]. The same applies to LEPS.

respectively LEPS, is weaker as for the temperature. The reason is here, like for the comparison of the DE-EPS with the SREPS regarding the equivalent potential temperature in Chapter 6 that the uncertainties of the observed state for the equivalent potential temperature are larger as for the temperature. This is the reason for a weaker decrease of the forecast quality. In Appendix A.3 a sensitivity study for the uncertainties for the observed state is shown, which figures out that also for the equivalent potential temperature in case of verification the strength of the decrease of evidence depends on the amount of the observation uncertainty. This means, the uncertainty of the real state of the equivalent potential temperature in the atmosphere leading to the circumstance that there are limits of verification and therefore, it is only possible to judge the forecast quality in a particular instance if reliable observations are available. In Chapter 6 the important impact of the uncertainty of the observed state was already shown.

Furthermore, the results show that for the equivalent potential temperature it is very helpful to

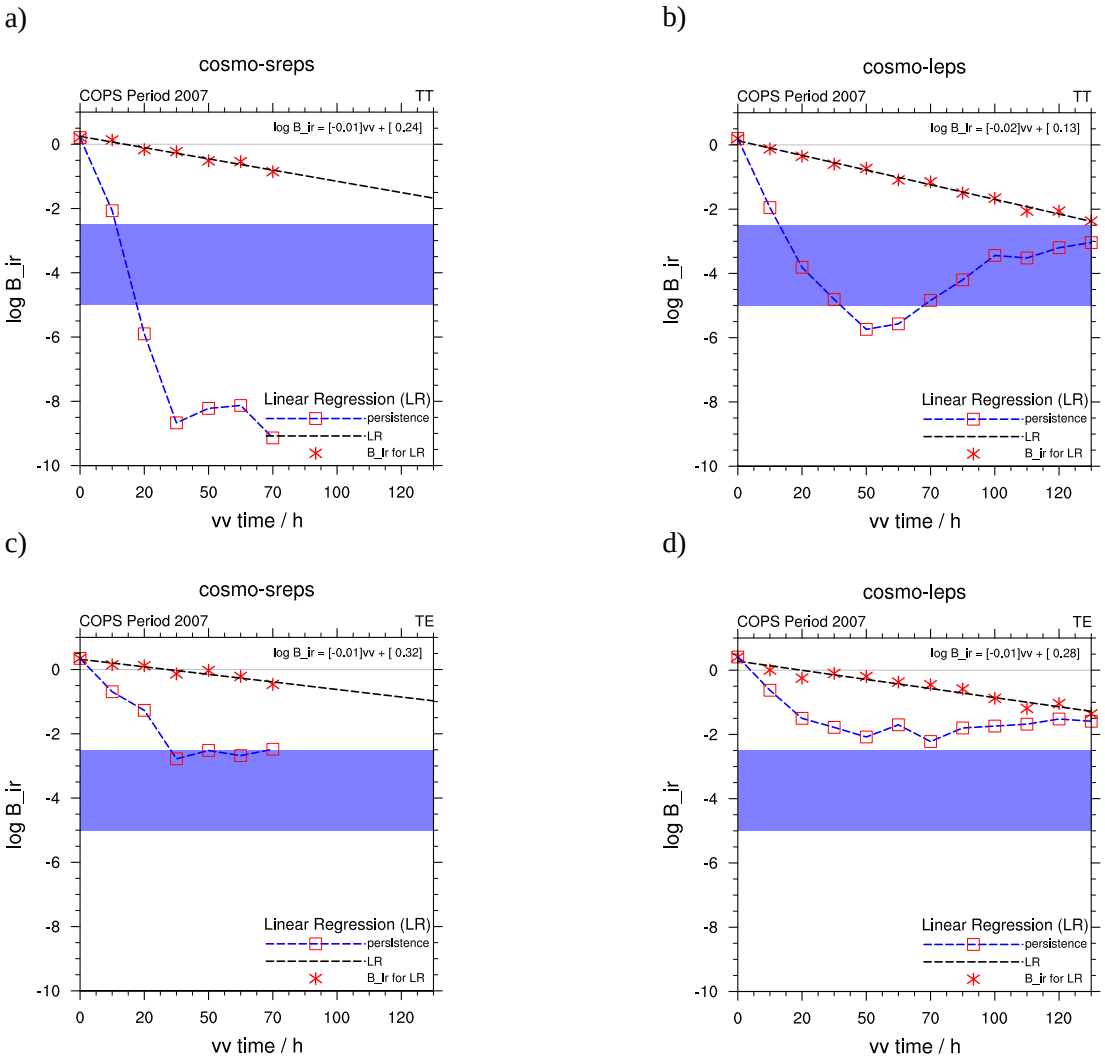



**Figure 7.5:** Linear regression of the Bayes factor for the temperature (a,b) and for the equivalent potential temperature (c,d). The Bayes factor is marked as red crosses. Additionally, the Bayes factor (red squares) is shown in case when the initial state is used as persistence with COSMO-EU analysis as reference model.

verify three profiles jointly together unaveraged, because in this case a clearer result is shown. The forecast quality is here within the first 72 hours of "high to medium level" of confidence and for the LEPS of "low level of confidence" after 132*h*.

The same behavior can be seen, when the sensitivity of the prior probability is varied shown by Fig. 7.4 (c,d).

## Predictability of convection

For the predictability investigation of convective conditions the Bayes factor for the temperature and the equivalent potential temperature of three stations averaged are shown in Fig. 7.5

including a linear regression for both ensembles and for both variables. The regression line illustrate the nearly linearly decrease of the logarithm of the Bayes factor with lead time.

In Chapter 2, it was shown that the predictability for the 500*hPa* temperature is going lost after approximate 168 hours (Fig. 2.2). This could not be investigated here due to the maximal forecast range of only 132 hours. However, at the end of the forecast range of 132 hours the regression lines of the SREPS and the LEPS show a Bayes factor about $\log B_{ir} = -2.5$ corresponding to a "medium level" of confidence, and thus the forecasts are slightly more likely than the persistence verification. The verification of the persistence forecasts of the SREPS and LEPS are done keeping the 0 hour forecasts constant while using the covariance matrix of each forecast lead time from the previous Bayes verification. The reference model is again the COSMO-EU analysis.

All in all, it is shown that there is predictability up to the forecast lead time of 132*h* or about 5 days. This correspond with the univariate predictability studies of Buizza et al. (2008) and Bougeault et al. (2010), who have shown that the predictability of the geopotential at 500*hPa* reached up to 15 days and of the temperature at 850*hPa* up to 7 days. This concerned for the geopotential of the northern hemisphere and for the temperature in the tropics of the ECMWF EPS.

## 7.2 Regime-dependent verification

The previous verification results covers temporal the whole COPS period. Additionally, a selected verification or regime dependent verification is needed to investigate if some convective regime conditions could be better forecasted as other ones. This refers to the regime dependent comparison of the DE-EPS with the SREPS in Chapter 6. There, the August 2007 period was split into three periods with different convective weather regimes. This was done by the usage of a convective adjustment time scale analysis from Keil and Craig (2011).

In this regime-dependent verification here, the selection of different convective regimes is done according the weather classification of the single COPS IOPs from Wulfmeyer et al. (2011). The convective weather regimes are air-mass convection (AMC), weakly forced convection (WFC) and strongly forced convection (SFC). For a detailed explanation of this type of convection and for the connection with the convective adjustment time scale see Chapter 3. Furthermore, for the regime dependent verification, the SREPS and the LEPS forecasts are investigated separately for the forecast ranges 0 to 24*h*, 24 to 48*h* and 48 to 72*h*. Each forecast range covers temporal a convective event of either AMC, WFC or SFC. This allows to investigate the behaviour for an increased forecast lead time and therefore the investigation of the forecast quality regarding to the different convection types.

The first regime dependent verification in Fig. 7.6 belongs to the temperature. The figure shows the verification results of the selected days with the convection types AMC, WFC and SFC. For the $0-24h$ period (Fig. 7.6 a), the SREPS forecasts of the SFC convection type are slightly

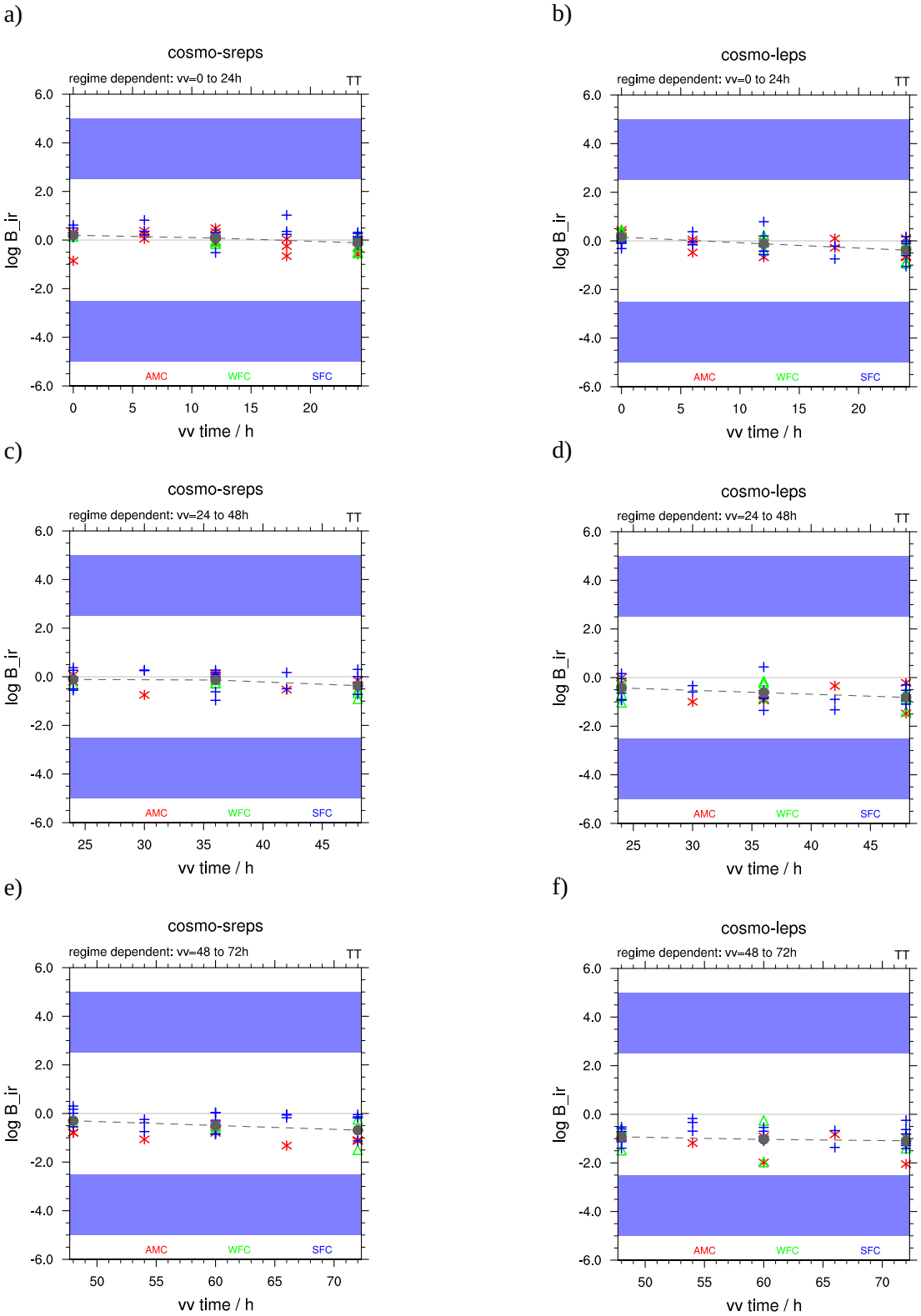



**Figure 7.6:** Regime-dependend verification of the SREPS (a,c,e) and the LEPS (b,d,f) for the temperature. (a,b) cover the $0-24h$ forecast period, (c,d) the $24-48h$ and (e,f) the $48-72h$ forecast period. The convection types are AMC, WFC and SFC following the COPS-IOP classification of Wulfmeyer et al. (2011).

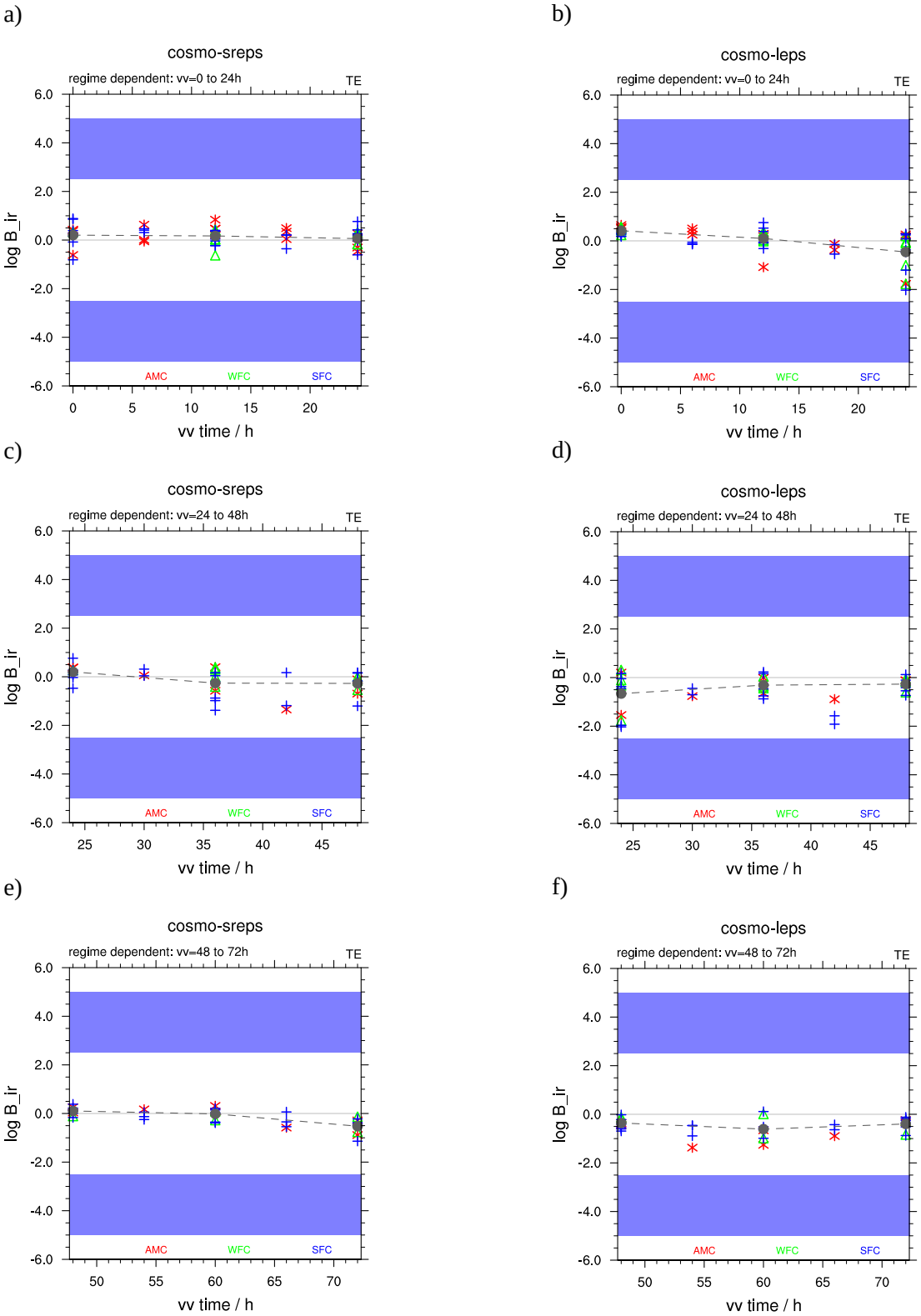



**Figure 7.7:** Regime-dependend verification of the SREPS (a,c,e) and the LEPS (b,d,f) for the equivalent potential temperature. (a,b) cover the $0-24h$ forecast period, (c,d) the $24-48h$ and (e,f) the $48-72h$ forecast period. The convection types are AMC, WFC and SFC following the COPS-IOP classification of Wulfmeyer et al. (2011).

better than those of AMC. For the LEPS forecast (Fig. 7.6 b), there is nearly no difference recognizable between the different convection types AMC, WFC and SFC. For the $24-48h$ period (Fig. c,d), only a very small difference is recognizable, which can be seen with increased forecast lead time for the $48-72h$ period clearer. Namely, it is shown that the vertical temperature profiles of SFC are slightly better predictable than those of WFC and AMC. However, the sample is too small to get a robust result. Nevertheless, the explanation would be in strongly forced weather situations the vertical stratification is better predictable as in weakly forced or non forced weather situations. Summarized, for the SREPS and for the LEPS the vertical temperature is of "high" to "medium level of confidence". But with increased forecast lead times, there are more and more verification values in the "low level of confidence" area. It is also difficult to see differences between the SREPS and the LEPS despite the fact that the LEPS is initialized 12 hours later as the SREPS.

Figure 7.7 shows the same plot, but for the equivalent potential temperature. For this forecast variable, there can be seen no differences between the three convection types. Only the LEPS shows at the $48-72h$ (Fig. e,f) period the same behaviour as before for the temperature. However, the difference between the three convection types is not so clear as in the case of the temperature. The verification classification is also very similar to those for the temperature. "high" to "medium level of confidence" for the majority part of the forecasts. But also with increased forecast lead times, there are more and more verification values in the "low level of confidence" area.

The regime-dependent analysis of the two ensembles shows that it is quite hard to select a convection type, which is better or worse forecasted respective to the vertical structure. This results from a too small sample of convective days and from the uncertainty of the observations. To get more reliable results here more levels have to be verified and more convective days have to be investigated. This investigation here can be only the beginning of a regime dependent verification, but it seems promising to get a clearer result for more levels and a larger data set. In Appendix A.7 this regime-dependent verification is also shown for the case of the jointly treatment of the three profiles unaveraged, which is here not shown due the too less number of cases for the single convective weather regimes and the large standard deviation (Fig. 7.3 and 7.4). To complete this work, for the SREPS the 15th July 2007 is investigated in more detail as a case study and demonstration of verification via the Bayes factor of a single run.

## 7.3 Verification of COSMO-SREPS - Case study 15th July 2007

Now, the Bayes factor is used to verify a single ensemble run of SREPS. As case study, the 15th July 2007, is presented, because at this day the COPS-IOP 8b promise an interesting synoptic situation as a demonstration of how the Bayesian approach works. Furthermore, this day provides the opportunity to use additional COPS radiosonde data to verify the SREPS at several different places, which are very close together. The SREPS is verified here for eight Levels $(1000, 925, 850, 700, 500, 400, 300, 250, 200 hPa)$ in case of temperature and for five Levels

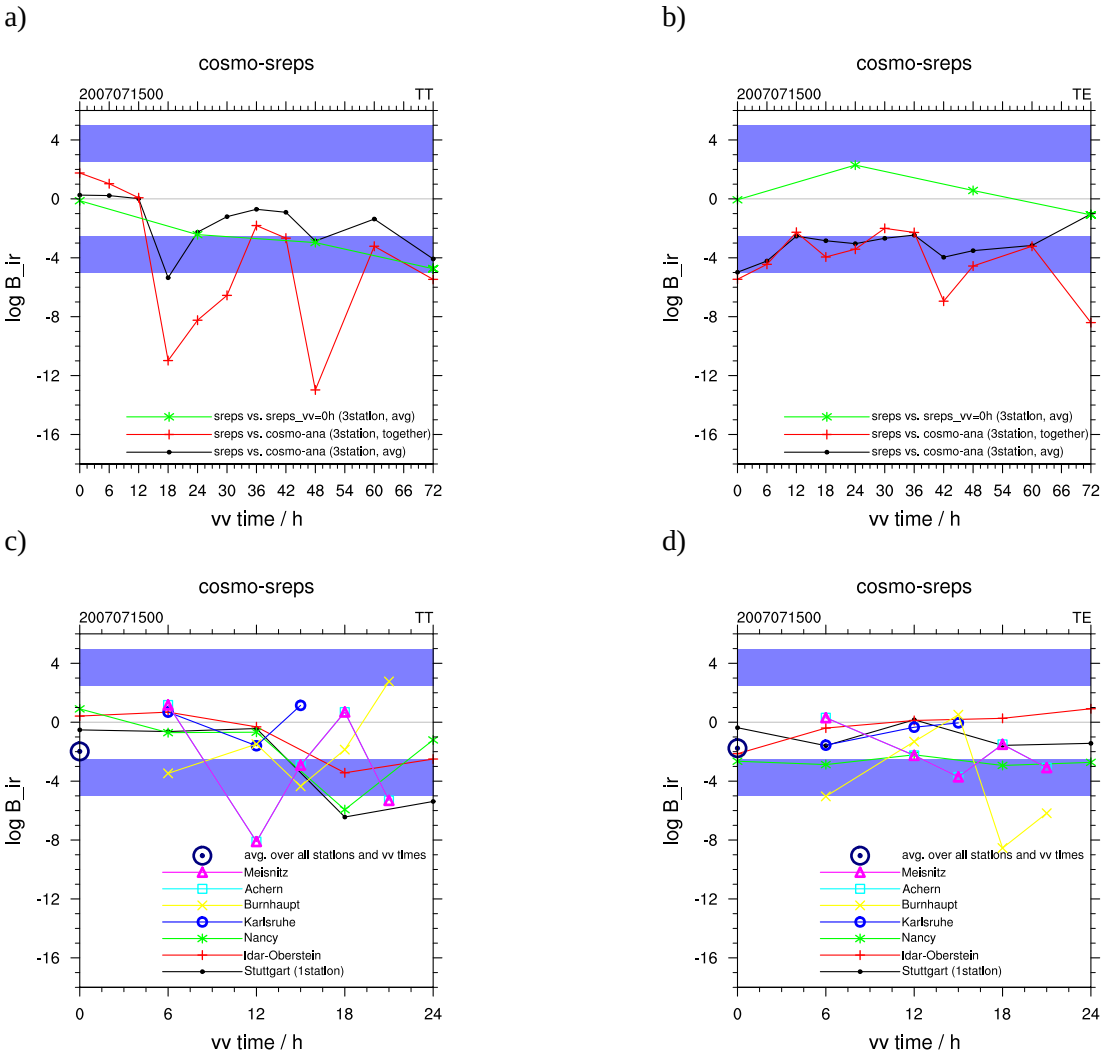



**Figure 7.8:** Time series of the Bayes factor $log\,B_{ir}$ of the temperature (a,c) and of the equivalent potential temperature (b,d) at forecast time (vv time) with respect to the COSMO-EU analyses at the same time. In (a,b) the green line shows the Bayes factor also for three stations averaged, but with the inital state of the SREPS from later runs as reference model. The red line shows the three stations (Stuttgart, Idar-Oberstein and Nancy) together unaveraged while the black line shows the three stations averaged. (c,d) shows $0-24h$ forecast period with the single verified station (Stuttgart, Idar-Oberstein and Nancy) plus the COPS stations (Meisnitz, Achern, Burnhaupt and Karlsruhe) with COSMO-EU analyses as reference model.

$(1000, 925, 850, 700, 500hPa)$ in case of the equivalent potential temperature.

In Fig. 7.8, the time series of the SREPS at 15th July 2007 for the 00UTC run is shown. Because of the verification of only one run, there is more fluctuation in the Bayes factor. Nevertheless, the decrease of the Bayes factor due to the growing forecast error with time can be seen. Figure 7.8 a) shows the verification of the temperature. At the beginning of the forecast the Bayes factor is still near zero. This shows that the model forecasts are on a "high level of confidence" in the first few hours. Later, the model skill of the SREPS is between a "medium" and a "low

| | $\sigma^2$ / $K^2$ | $\sigma^2$ / $K^2$ |
|---|---|---|
| run | 2007071600 | 2007071500 |
| vv time | 0*h* | 24*h* |
| 1000*hPa* | 9.5402 | 2.6315 |
| 925*hPa* | 4.8148 | 2.8864 |
| 850*hPa* | 1.2468 | 3.5299 |
| 700*hPa* | 2.3935 | 3.4425 |
| 500*hPa* | 0.5476 | 0.6219 |

**Table 7.3:** Variances of the SREPS for the case study of the equivalent potential temperature.

level of confidence". At the forecast lead time of 18 hours, the figure shows a strong decrease of evidence against the SREPS indicating a very unlikely forecast of the vertical temperature profile of the SREPS given the vertical profile of the COSMO-EU analysis. The comparison of the forecasted SREPS temperature profiles with the observations show a strong cold Bias in the boundary layer (see Fig. 7.9 a) being responsible for this drop in the score. Finally, the usage of the 0 hour forecast of the SREPS itself as reference model from later runs has the benefit to verify the SREPS forecast with its own initial state. But this has the disadvantage of a lower temporal resolution, because the SREPS was initialized only every 24 hours. The previous results with COSMO-EU analysis as reference model are in agreement with the results in this case.

Additionally, Fig. 7.8 c) shows the first 24 hours forecasts of the SREPS in more detail. The three stations investigated before are verified separately now plus additional COPS stations. At the beginning of the forecast, they are all close together on a "high level of confidence". But then, the single stations show partly large differences in the verification results relating that already on a small area the forecast quality of the vertical structure can be completely different. This shows the challenging task to predict convection. Even in the first hours, it seems to be hard to predict the right convective conditions. This holds also for the equivalent potential temperature shown in Fig. 7.8 d). Additionally, it is shown that the forecasts of the temperature and of the equivalent potential temperature have a "medium level of confidence" when the average over all stations and over all forecast lead times is investigated.

The verification of the equivalent potential temperature over the complete 72 hour forecast is shown by Fig. 7.8 b). In this case, two aspects have to be discussed. The first point concerns the forecast quality of "low level of confidence" of the equivalent potential temperature over the whole forecast lead time. This shows that it is important to look at several stations to get a comprehensive picture of the verification. The equivalent potential temperature is forecasted very badly in this case study. The respective vertical profiles (model minus observation) are shown in Fig. 7.9 b). For the single stations, the verification results look slightly better. At the first hours, the stations Idar-Oberstein and Nancy have already a forecast quality of "medium/low level of confidence". The second point is the behaviour of the SREPS when verified against his own initial state. In this case, the result differs completely from the verification result where

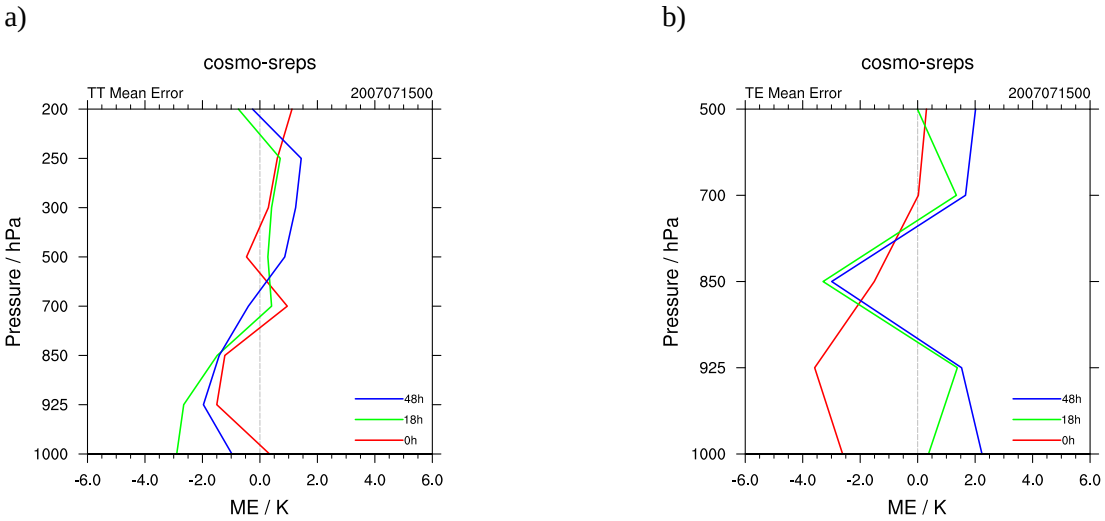



**Figure 7.9:** Vertical temperature profiles (a) and vertical equivalent potential temperature profiles (b) from SREPS at Stuttgart (model minus observation) for the forecast lead time vv=0,18,48*h*.

the COSMO-EU analysis is used as reference model. This is caused by the variance of the own initial state (vv=0*h*) of the SREPS for the 24*h* forecast lead time, which is nearly three times larger as those variance at the vv=24*h* forecast state at the 1000*hPa* pressure level. In numbers this is shown in Tab. 7.3.

## 7.4 Summary and conclusion

In this work, the second application for the Bayes factor is the verification of ensembles. For the comparison of two ensembles, it is of particular importance to consider the uncertainties of the model as well as of the observations. The aim was there to see if the evidence for an ensemble is significant or not. For verification, this points are important too, because an analysis as reference model has also an uncertainty like the observations.

The verification results of the COSMO-SREPS and COSMO-LEPS over the whole COPS period have shown that both ensembles are quite similar in relation to the forecasted convective permitting conditions. However, it is difficult to compare this both ensembles exactly, because they are initialized at different times. At the beginning of the forecast, the ensembles are not distinguishable from the COSMO-EU analysis. With an increased lead time, the Bayes factor decreases linearly and reaches at 72 hour a "medium level of confidence". In case of the LEPS, the forecast length is 132 hours. At the end of this forecast lead time, the evidence for the LEPS is going to a "low level of confidence".

The longer forecast range of the COSMO-LEPS allows to investigate the time when the predictability of the convective conditions is going lost. It was shown that there is predictability up to the end of the forecast range of 5 days.

Finally, the results show that it is possible to verify NWP ensembles with this method. This applies without restriction for the temperature. For the equivalent potential temperature, the results are also very useful, but it has been shown that the large uncertainty of the state of this variable reduces the strength of the evidence. It might be also helpful to investigate further variables like the specific humidity, which will lead to the difficulty of a non-Gaussian PDF. The large standard deviation shows again the need for a longer investigation time period. This is especially true for the jointly unaveraged treatment of three profiles.

# Chapter 8

# Conclusion and Future Works

A discussion of the main points for comparison and verification of ensembles through the Bayes factor is presented at the end of Chapter 6 and 7. In this Chapter, the main investigation results of this work are concluded and an outlook of future work will be given, regarding the possibilities of application or rather verification via the Bayesian statistic in the area of verification of NWP models.

## 8.1 Conclusion

QPFs are among the most demanding applications in NWP, because there were only a slight improvements of QPFs over the last years. Particularly, forecasts of convective precipitation have large uncertainties. However, for just in time warnings QPFs are essential due to the high potential for damage of convective precipitation.

Probabilistic forecasts of EPS are used to deal with the uncertainties of deterministic NWP model forecasts. They allow to handle with the chaotic nature of the physical processes in the atmosphere, which are responsible for convection. However, ensemble forecasts have also to be verified. There is a chain of probabilistic standard verification methods, but in case of convection forecasts, the whole vertical state is decisive and has to be verified. Hence, a verification method is needed, which allows to verify these whole profiles multivariately at small areas to avoid the effect of averaging out information of horizontal correlations. This requires a new verification technique.

The probabilistic verification approach used in this work allows to verify whole profiles considering explicitly the observation error as well as the model uncertainty. It is important to take into account that the observed state by itself is uncertain due to the limited capabilities of observing the atmosphere. The verification approach is based on the Bayesian verification method of climate change simulations from Min et al. (2004) and Min and Hense (2006). The key targets of this work were to use a probabilistic method

- **to verify and compare ensemble predictions of atmospheric state vectors**
- **to include an uncertainty measure of the observations**
- **to allow for relative measures between different EPS systems**

These aspects are important items to verifying forecasts at convection-permitting resolutions and to study the predictability of convection initiation potential. Therefore, a multidimensional state vector of the forecast ensemble has to be used characterizing at least the vertical temperature and moisture structure. The multivariate aspect was defined by several vertical levels, which are treated simultaneously taking into account the dependencies between the levels, because the vertical profile of temperature and humidity affects the potential for the convective instability. Radiosonde data are used for the verification of the vertical profiles.

The investigated ensembles in this work are the COSMO-DE-EPS, the COSMO-SREPS and the COSMO-LEPS. The main questions to answer in this work were

Q1. **Is there a significant evidence for the new convective-permitting COSMO-DE-EPS with respect to the forecasted vertical structures when it is multivariate compared with the coarser resolved COSMO-SREPS ?**

The COSMO-DE-EPS developed by the Deutscher Wetterdienst (DWD), which is based on the deterministic convection permitting COSMO-DE, was compared with the coarser resolved COSMO-SREPS. The comparison of the vertical temperature profiles has shown a significant evidence for the COSMO-DE-EPS (posterior probability of the DE-EPS about 80 %) in contrast to the COSMO-SREPS. The evidence in respect to the equivalent potential temperature is much weaker, because of the larger uncertainty of the observed state. Therefore, the DE-EPS is appropriate to handle the uncertainties of the convective permitting conditions much better than coarser models.

However, this applies to the vertical profiles. For convective initiation, it is absolutely crucial to simulate the triggering of the convective events. This issue was not investigated in this work. Furthermore, the investigation period comprises the August 2007 for this comparison. However, a much longer period of investigations seems to be necessary to underline these result.

Q2. **How does the verification of the forecasted vertical structures of COSMO-SREPS and COSMO-LEPS behave with lead time ?**

The COSMO-SREPS and the COSMO-LEPS are meso-scale ensembles with parametrized convection. The verification results over the whole COPS period for both ensembles against COSMO-EU analyses are close together. Although, the COSMO-LEPS was initialized at 12UTC, which is 12 hours later as the COSMO-SREPS. Both

ensembles show an almost linear decrease of skill with lead time. The temperature of both ensembles reached a "medium level" of confidence after 72 hour. The COSMO-LEPS shows a "low level" of confidence after 132 hours. For the equivalent potential temperature representing the humidity, the decrease of evidence with lead time is significantly weaker. This is again due to the larger uncertainty of the observed state.

The verification has shown that it is possible to verify ensembles with the great advantage to consider several levels simultaneously. This confirms, the Bayes factor allows explicit both comparison as well as verification of ensembles.

Q3. **Is there predictability at long forecasted lead times of COSMO-SREPS and COSMO-LEPS ?**

To answer this question, mainly the COSMO-LEPS is considered, because the forecast range of the LEPS is substantially longer with 132 hours as the COSMO-SREPS. It is shown that within the first 132 hours (about 5 days) of the forecast the skill (confidence level) of the verification is higher than those where a persistence forecast is verified indicating that the predictability of convective conditions reached up to about 5 days.

Furthermore, it has to be considered that the time scales of convective processes occurring in nature comprise only several hours. So, in general, it is only rudimentary possible to forecast convection over several days. However, it seems that both ensembles can give here helpful guidelines about the occurrence of convection allowing conditions including the first 5 days. The main limitation is the fact that in this study only three levels are investigated. For more meaningful results, the number of levels as well as the considered time period should be increased.

Last but not least, it should be mentioned that this study can not be considered as a full scale verification analysis of the D-Phase ensembles or the COSMO-DE-EPS. At least this task would require a much larger radiosonde network and forecasts covering much larger time periods. This work is rather a proof-of-concept or pilot study to identify demonstrative strengths and weaknesses of the presented approach, and it has been shown that the probabilistic verification method described in this study is appropriate to compare ensemble systems with each other and to verify ensembles. The Bayes factor used as a score is a generalization of the Ignorance score taking additionally into account the uncertainty of the observations as well as the spatial correlation structure of the verified forecasts. The consideration of the observation uncertainty is important, because the observations are also defective and do not represent exactly the true state of the atmosphere. The effect of the observation uncertainty is shown by the results for the equivalent potential temperature. Here the large observation uncertainty leads to a significant decrease of evidence for a respective model, because there is simply a wide area of possible atmospheric states. Till now, the most verification methods do not consider this.

The Bayes factor allows for a comprehensive evaluation of the forecast quality of three dimensional samples by using just one score. Even an extension to temporal-spatial structures

is readily possible provided that there is a method available to estimate non-singular covariance or correlation matrices of high dimensional state vectors. Future prospects are given in the following.

## 8.2 Outlook

The Bayes factor is a helpful score to investigate the quality and predictability of NWP models. The Bayes factor can be used for case studies as well as for longer time periods. This work shows only a small area of possible applications. An overview about possible future applications will be given in the following.

### Another investigation area, another weather regimes and position corrected radiosonde data

In this study the focus was on convection inside the COPS area, which is an orographically influenced area. Other areas like the northern part of Germany would be also of interest to compare the current results with results for flat terrain.

It is also thinkable to look at additional weather situations like winter storms. For the $10m$ wind gusts the vertical structure is also very important, especially again the stratification in the boundary layer. The strongest wind gusts down to the surface occur during the passing of cold fronts. For higher-resolution models well forecasted vertical profiles become extremely important.

Radiosonde data used in this work have not been position corrected, leading to a larger observation covariance matrix. The usage of the position corrected radiosonde data, which would require a smaller observation standard deviation, would lead to clearer results.

### Evaluation of longer time periods

Furthermore, a simple but important point is the investigation of longer time periods, which requires longer data sets. Therefor, extended data set with more than only three vertical levels for a long term period would be needed. Particularly, the large standard deviation has shown the need for this point. The limiting factor in this work was the lack of stored EPS data. The Bayes factor is explicitly applicable to case studies, but for meaningful results an evaluation of a longer time period would be necessary.

### Usage of additional reference models

A further aspect would be the usage of additional reference models. The Hans-Ertel centre for weather research (in German HErZ) founded by the DWD aims at developing a model climatology based on the planed COSMO-DE reanalysis (Ohlwein et al., 2011). For a model climatology a reanalysis is needed, because the standard analysis comprises different model versions over climatological timescales and therefore it is only of limited usability. The application of a model climatology as reference model would offer additional opportunities for predictability studies.

### Bayesian model averaging

The last point, which will be discussed in this outlook is the opportunity to use the Bayes factor for the so called Bayesian model averaging (BMA) described in Min and Hense (2006) and Raftery et al. (2005). The Bayes factor could be used additionally as a weight for forecast calibration e.g. to use a combination of COSMO-LEPS and COSMO-SREPS. The BMA can be defined here as a weighted average of the model-forecasted state vector $f_i$, which describes the arithmetic mean (AEM) of the corresponding model $m_i$.

$$\left[\vec{f}\right]_{BMA} = \frac{\sum_{i=1}^{N} B_{ir} f_i}{\sum_{j=1}^{N} B_{jr}} \tag{8.1}$$

If all Bayes factors would be identical to each other, the BMA would be equal to AEM for all ensemble member of all ensembles.

$$\left[\vec{f}\right]_{AEM} = \frac{1}{N} \sum_{i=1}^{N} \vec{f}_i \tag{8.2}$$

Therefore, the Bayesian statistics provide a convenient way to treat both the predictability and the forecast calibration in a single approach. Thereby, the Bayesian statistics can be also applied to deterministic models, if the deterministic model is treated as an artificial ensemble like the deterministic COSMO-EU analysis in this work.

# Appendix A

# Methodology

## A.1 Moisture variables

In the following a short overview over the moisture variables is given used in this work. A detailed description can be found in Kraus (2004). The specific humidity $q_v$ is the ratio of the densities of water vapor $\rho_v$ to air $\rho_t$ (including water vapor and dry air $\rho_d$) in a particular mass. The specific humidity ratio is expressed as a ratio of kilograms of water vapor to kilogram of total moist air $\rho_t = \rho_d + \rho_v$.

$$q_v = \frac{\rho_v}{\rho_t} \tag{A.1}$$

respectively

$$q_v = 0.622 \, \frac{e}{p - 0.378\, e} \approx 0.622 \, \frac{e}{p} \tag{A.2}$$

where $e$ denotes the partial pressure of water vapor and $p$ the total pressure of the air. The mixing ratio $m$ is the ratio of the densities of water vapor $\rho_v$ to dry air $\rho_d$. Specific humidity is related to mixing ratio (and vice versa) by:

$$m = \frac{\rho_v}{\rho_d} = 0.622 \, \frac{e}{p - e} \approx 0.622 \, \frac{e}{p} \tag{A.3}$$

Relative humidity is defined as the ratio of the partial pressure of water vapor (in a gaseous mixture of air and water vapor) to the saturated vapor pressure of water at a given temperature. In other words, relative humidity is the amount of water vapor in the air at a specific temperature compared to the maximum water vapor that the air is able to hold without condensing at that given temperature. Relative humidity is expressed as a percentage and is calculated in the following manner

$$rh = 100 \frac{e}{E} \tag{A.4}$$

Absolute humidity on a volume basis is the quantity of water in a particular volume of air, which is the density of water vapor. The dew point is the temperature to which a given parcel of humid air must be cooled, at constant barometric pressure, for water vapor to condense into water. The condensed water is called dew. The dew point is a saturation temperature. The dew point is associated with relative humidity. A high relative humidity indicates that the dew point is closer to the current air temperature. Relative humidity of 100% indicates the dew point is equal to the current temperature and the air is maximally saturated with water vapor. When the dew point remains constant and temperature increases, relative humidity will decrease.

## A.2 Equivalent potential temperature

The equivalent potential temperature was introduced in Chapter 3. Here the calculation of the standard deviation of the equivalent potential temperature will be discussed. The standard deviation of the temperature is taken from the data assimilation scheme (3dvar) of the DWD. Based on this, the standard deviation of $\Theta_e$ is approximated as follows.

The equivalent potential temperature $\Theta_e$ describes the content of energy of air masses including the humidity (mixing ratio $m$) and describes additionally the stability of the atmosphere.

$$\Theta_e = T\left(\frac{p_0}{p}\right)^{\frac{R}{c_p}} \cdot exp\left(\frac{Lm}{c_p T}\right) \tag{A.5}$$

The equation of $\Theta_e$ is here approximated in first order: $\exp(x) = 1 + x$

$$\Theta_e = T\left(\frac{p_0}{p}\right)^{\frac{R}{c_p}} + \left(\frac{p_0}{p}\right)^{\frac{R}{c_p}} \cdot \left(\frac{Lm}{c_p}\right) \tag{A.6}$$

For approximation of the standard deviation of $\Theta_e$, a formula of error propagation for functions of two variables (NIST/SEMATECH, 2010) is used $\sigma_f^2 = a^2\sigma_A^2 + b^2\sigma_B^2$ with the assumption that there are no correlations between $A$ and $B$. Additionally, the simplification $\sigma_p^2 \longrightarrow 0$ is taken. Therefore, all variables except $T$ and $m$ are treated as constant. Finally, the deviation $\sigma_{\Theta_e}^2$ can be written as

$$\sigma_{\Theta_e}^2 \approx \left(\frac{p_0}{p}\right)^{\frac{2R}{c_p}} \sigma_T^2 + \left(\frac{p_0}{p}\right)^{\frac{2R}{c_p}} \left(\frac{L}{c_p}\right)^2 \sigma_m^2 \tag{A.7}$$

$$= \left(\frac{p_0}{p}\right)^{\frac{2R}{c_p}} \left[\sigma_T^2 + \left(\frac{L}{c_p}\right)^2 \sigma_m^2\right] \tag{A.8}$$

In the data assimilation scheme of the DWD, only the standard deviation of the relative humidity is given. This requires additional the calculation of the standard deviation of the mixing ratio $m \approx 0.623\frac{E}{p} f$ . This is done by

$$\sigma_m^2 = \underbrace{\left(0.622\,\frac{E}{p}\right)^2}_{\approx 0.0001} \cdot \sigma_f^2 \tag{A.9}$$

The approximation is done because for simplification and to reduce the run time in case of calculation the Bayes factor for the equivalent potential temperature. Thus, there is finally

$$\sigma_{\Theta_e}^2 \approx \left(\frac{p_0}{p}\right)^{\frac{2R}{c_p}} \left[\sigma_T^2 + \left(\frac{L}{c_p}\right)^2 0.0001\sigma_f^2\right] \tag{A.10}$$

## A.3 The observation error statistics

In meteorology, observations are used in the data assimilation as well as for verification. However, almost entirely in the data assimilation the uncertainty of the observed state itself is considered. For the most part this does not belong for verification.

In this work, the observation error is considered explicitly based on the 3dvar from the DWD (only height dependent, see Tab. 5.7) and it is shown that this has a meaningful effect on verification results. For further investigations a systematical variation of the observation error via a constant factor $\gamma$ is done to investigate the sensitivity of the Bayes factor to the observation error.

For the observation error in the current 4dvar system of the ECMWF a slightly different approach for the observation error is used in contrast to the 3dvar system at the DWD. At ECMWF the observation error consists of

- a persistence observation error (season and geographical position dependent)
- and of a prescribed observation error (height dependent)

The final observation error is then a combination of this two above mentioned error types. The details can be found in ECMWF (2011).

A sensitivity study of the observation error is shown in Chapter 6. Additionally, a sensitivity study of the observation error in case of the verification is shown here. Figure A.1 shows the corresponding sensitivity study for the verification of the LEPS in Chapter 7 regarding the sensitivity of the LEPS in view of the observation uncertainty. Figure (a) shows the sensitivity study for the temperature and (b) for the equivalent potential temperature. It is shown that the verification strongly depends on the value of the observation uncertainty, which is variated by a constant factor $\gamma$.

The main point, which will be here discussed is that when the observation uncertainty for the equivalent potential temperature is reduced ($\gamma = 0.5$) the verification is much stronger as in the

a) b)

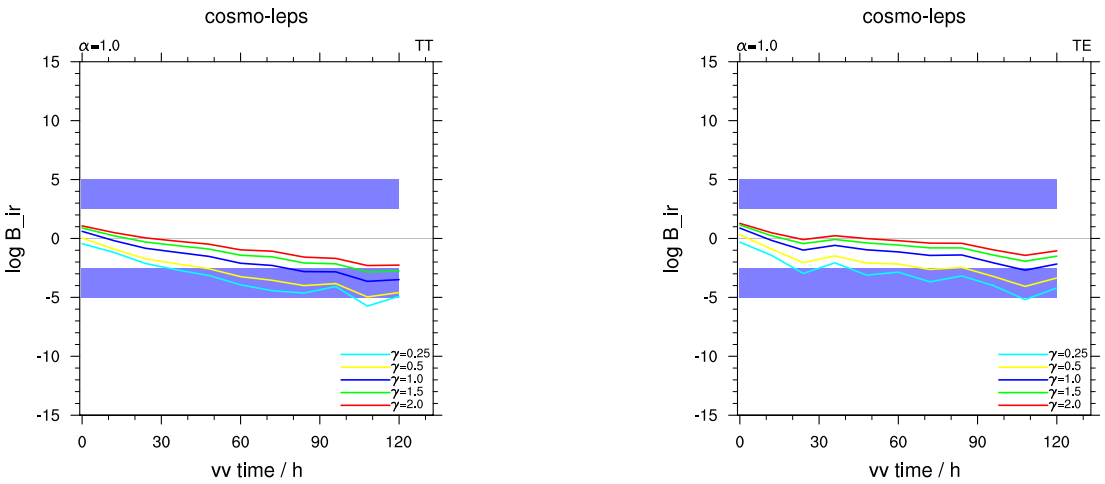



**Figure A.1:** The distribution of the Bayes factor $log\,B_{ir}$ for the LEPS (a,b) respectively against COSMO-EU analysis. Different weighting factors $\gamma = 0.25, 0.5, 1.0, 1.5, 2.0$ for the observational covariance matrix are used.

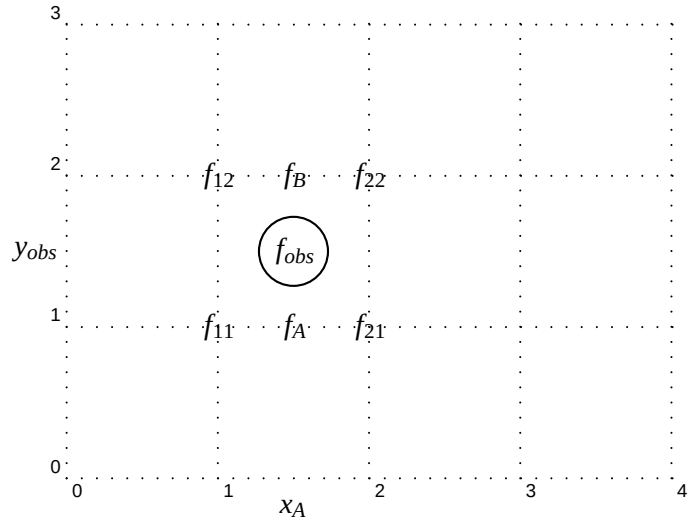
opposite case ($\gamma = 2.0$) where it is nearly impossible to decide, which one the analysis or the LEPS forecast is more likely.

## A.4 Interpolation

For the verification, the model variables calculated at grid points has to be interpolated to the observation point. This has to be done only for the horizontal, because in the vertical the model variable as well as the observation are available at the same pressure levels.

### Horizontal interpolation to the observation point

The horizontal interpolation to the observation point is done by a bilinear interpolation e.g. see White (2003b); Hackbusch et al. (2003). The method of bilinear interpolation consists of three linear interpolations.




The first two interpolations interpolate the grid points $f_{21}$ and $f_{22}$ to $f_A$ and the remaining two grid points to $f_B$. Finally, the third linear interpolation interpolates the auxiliary points $f_A$ and $f_B$ to the observation point $f_{obs}$. The equations for the bilinear interpolation are given by

$$f_A = f_{21} + \frac{f_{22} - f_{21}}{x_2 - x_1} \cdot (x_A - x_1) \tag{A.11}$$

$$f_B = f_{11} + \frac{f_{12} - f_{11}}{x_2 - x_1} \cdot (x_A - x_1) \tag{A.12}$$

$$f_{obs} = f_A + \frac{f_B - f_A}{y_2 - y_1} \cdot (y_{obs} - y_1) \tag{A.13}$$

### Vertical interpolation

The vertical interpolation from the model levels to the observation levels is not necessary, because the model data are used on pressure levels and this pressure levels correspond to the mandatory measurement levels of the radiosondes.

Only in case of the $1000hPa$ model level, here the lowest radiosonde measurement has to be vertical interpolated to the $1000hPa$ model level, but only if the lowest radiosonde measurement is above $1000hPa$. In this case the lowest radiosonde measurement is extrapolated to the $p_0 = 1000hPa$ model level. The reason for this proceeding is that there is no exact way to reconstruct the NWP model value at the lowest radiosonde observation. The extrapolation under the surface is done by the approach of White (2003b) which is also used in the COSMO model. The basic idea is

$$T_L = T_{surf} \left( 1 + y + \frac{y^2}{2} + \frac{y^3}{6} \right) \tag{A.14}$$

with:

$$y = \frac{dT}{dz} \frac{r_d}{g} \log \left( \frac{p_s}{p_0} \right) \tag{A.15}$$

The extrapolation is only used for the temperature. The relative humidity is kept constant on the value at the surface value. In this work the implementation from the COSMO source code is used.

## A.5 The gLasso method

The log-likelihood $L$ for a $K$ multivariate Normal distributed sample of dimension $q$ represented by the Gaussian PDF is shown in Eq. 4.15. This is the basis for the MLE of the covariance matrix. $S$ indicates the standard estimate of the covariance matrix $\Sigma$. The log-likelihood can be reformulated as a cost function $J$ which has to be maximized.

$$J = \log(\det \Sigma^{-1}) - \mathrm{trace}(S\Sigma^{-1}) \tag{A.16}$$

The basic starting point for the gLasso method (Friedman et al., 2007) used in this work is to estimate a sparse matrix $\Theta = \Sigma^{-1}$ which is the inverse of a covariance matrix and which maximizes the log-likelihood cost function $J$ penalized by an additional term $J_{pen}$ and already maximized partially with respect to the mean $\mu$. $S$ indicates the standard estimate of the covariance matrix as mentioned previously, $\det\Theta$ the determinant of $\Theta$ and the factor $\rho$ controls the influence of the penalizing term. The magnitude of $\rho$ ranges between 0.01 and 0.6 (in this work $\rho = 0.01$).

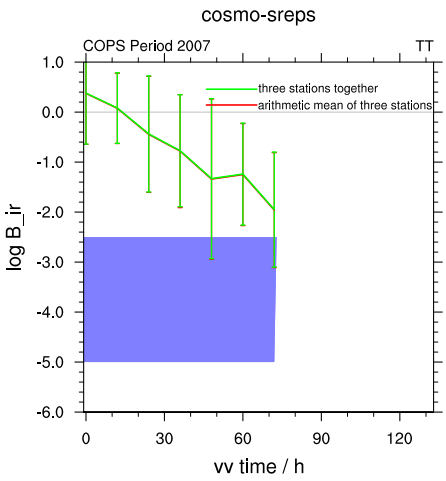
$$J = \log(\det \Theta) - \mathrm{trace}(S\Theta) + \rho J_{\mathrm{pen}} \tag{A.17}$$

The Hammersley-Clifford theorem (Wainwright and Jordan, 2008) (p.45) proves that two components $i$ and $j$ of a Gaussian distributed random vector variable (here the temperature or equivalent potential temperature values at various levels) are conditionally independent given the remaining components if the entry $\Theta_{ij}$ is zero. Therefore, it makes sense to require as much as possible entries $\Theta_{ij}$ to be zero as an additional information to estimate the covariance matrix and its inverse. Similar procedures in data assimilation applied to the covariance matrix are called localization. This can be achieved by a penalize term which sums the absolute values of the matrix entries the so called $L_1$ matrix norm of $\Theta$

$$J_{pen} = \sum_{i,j} |\Theta_{ij}| \tag{A.18}$$

The maximization of the absolute values guarantees that the extreme value of $J$ is attained at $\Theta_{ij} = 0$ (Knight and Fu, 2000). The procedure to find the extreme value of $J$ is called least absolute shrinkage and selection operator (lasso). If the vector components are viewed as nodes of a network which are linked if they are not conditionally independent and not connected if they are conditionally independent the method defines a so called graph (the joint set of nodes

a)




b)

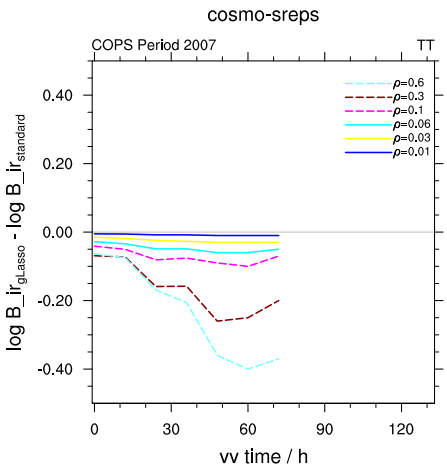



**Figure A.2:** Part (a) shows the time series of the Bayes factor $log\,B_{ir}$ of the temperature at forecast time (vv time) with respect to COSMO-EU analysis at the same time. The red line shows the Bayes factor calculated using the gLasso (Friedman et al., 2007), and the green line shows the Bayes factor calculated using the standard covariance matrix. The blue band describe a significant area, see Tab. 4.2. Part (b) shows the difference between the Bayes factor using the gLasso and the Bayes factor using the standard method for different values of ρ.

and links) or graphical model for the interactions of the vector component. Therefore, the term gLasso was coined. The interpretation of the method is that a non-singular matrix is estimated which inverse $\Theta^{-1}$ is as similar as possible to the standard sample covariance matrix $S$ and which itself has least non zero entries or the smallest necessary graph to explain the covariances among the vector components.

The algorithm of Friedman et al. (2007) is used which is available through their program package gLasso for the R programming environment. Figure A.2 a) shows a comparison of a calculation of the Bayes factor between the standard covariance matrix $\Sigma_i$ and the case where the covariance matrix is approximated by Friedman et al. (2007). The method works quite well, there is almost no difference between those two cases ($\rho = 0.01$ used in this work), see Fig. A.2 b). The difference to the standard method increases for increased values of ρ, but shows even with $\rho = 0.6$ only a maximal difference of 0.4 (Fig. A.2 b).

## A.6 The correlations

The correlation matrices for the comparison of the DE-EPS with the SREPS in Chapter 6 for three forecast lead times ($vv = 0, 12, 24h$) are shown in Fig. A.3 in this section. The correlation matrices for the August 2007 period of the comparison show only weak correlation between the single levels. Only for the $12h$ forecast lead time, larger correlation between the first two levels can be seen each for the DE-EPS and also for the SREPS. This result of low correlations between the single levels is shown before in Chapter 6 in Fig. 6.2 c).

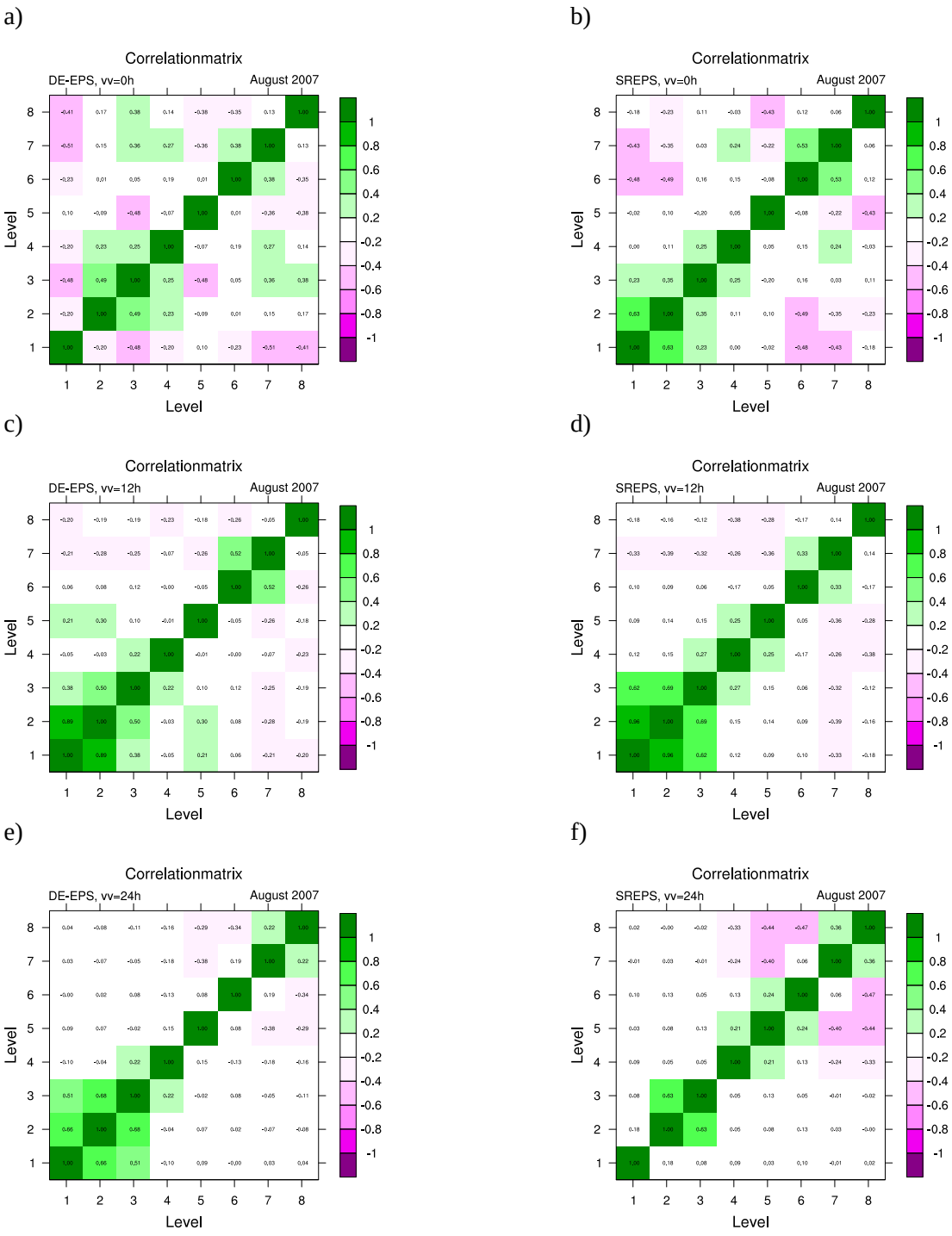



**Figure A.3:** Correlation matrices of the DE-EPS (a,c,d) and of the SREPS (b,d,f). Shown are three forecast lead times (0*h*, 12*h* and 24*h*). The correlation matrices are shown for the August 2007.

## A.7 Regime-dependent verification

A regime-dependent verification of the SREPS and LEPS is shown in Chapter 7. In addition here, the same regime-dependent verification is shown in Fig. A.4 and A.5 for the case of the jointly treatment of the three profiles unaveraged. This was not shown before, because of the large standard deviation of the Bayes factor coming from too less number of cases for the convective weather regimes. The weather regimes are air-mass convection (AMC), weakly forced convection (WFC) and strongly forced convection (SFC).

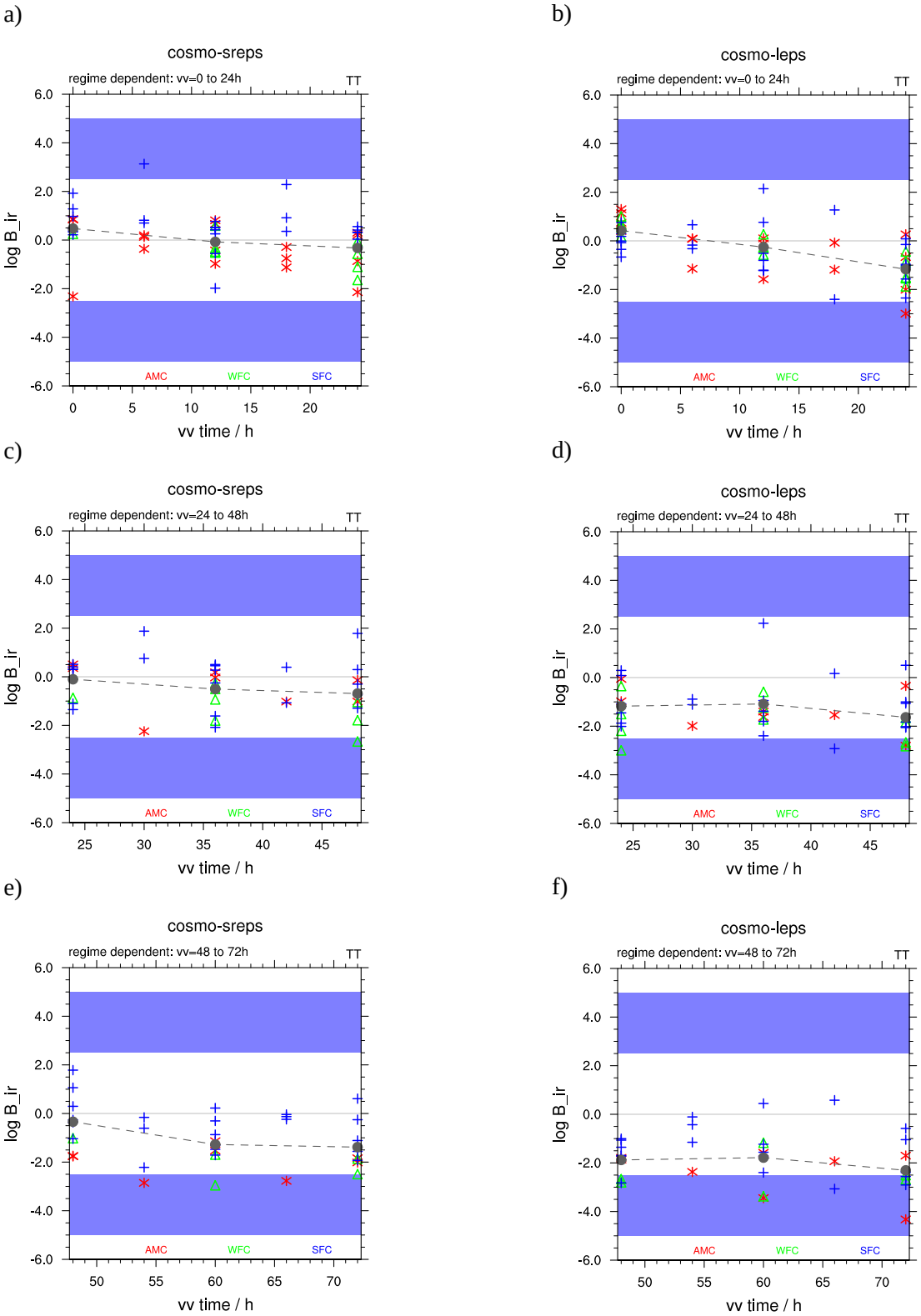



**Figure A.4:** Regime-dependend verification of the SREPS (a,c,e) and the LEPS (b,d,f) for the temperature. (a,b) cover the $0-24h$ forecast period, (c,d) the $24-48h$ and (e,f) the $48-72h$ forecast period. The convection types are AMC, WFC and SFC following the COPS-IOP classification of Wulfmeyer et al. (2011).

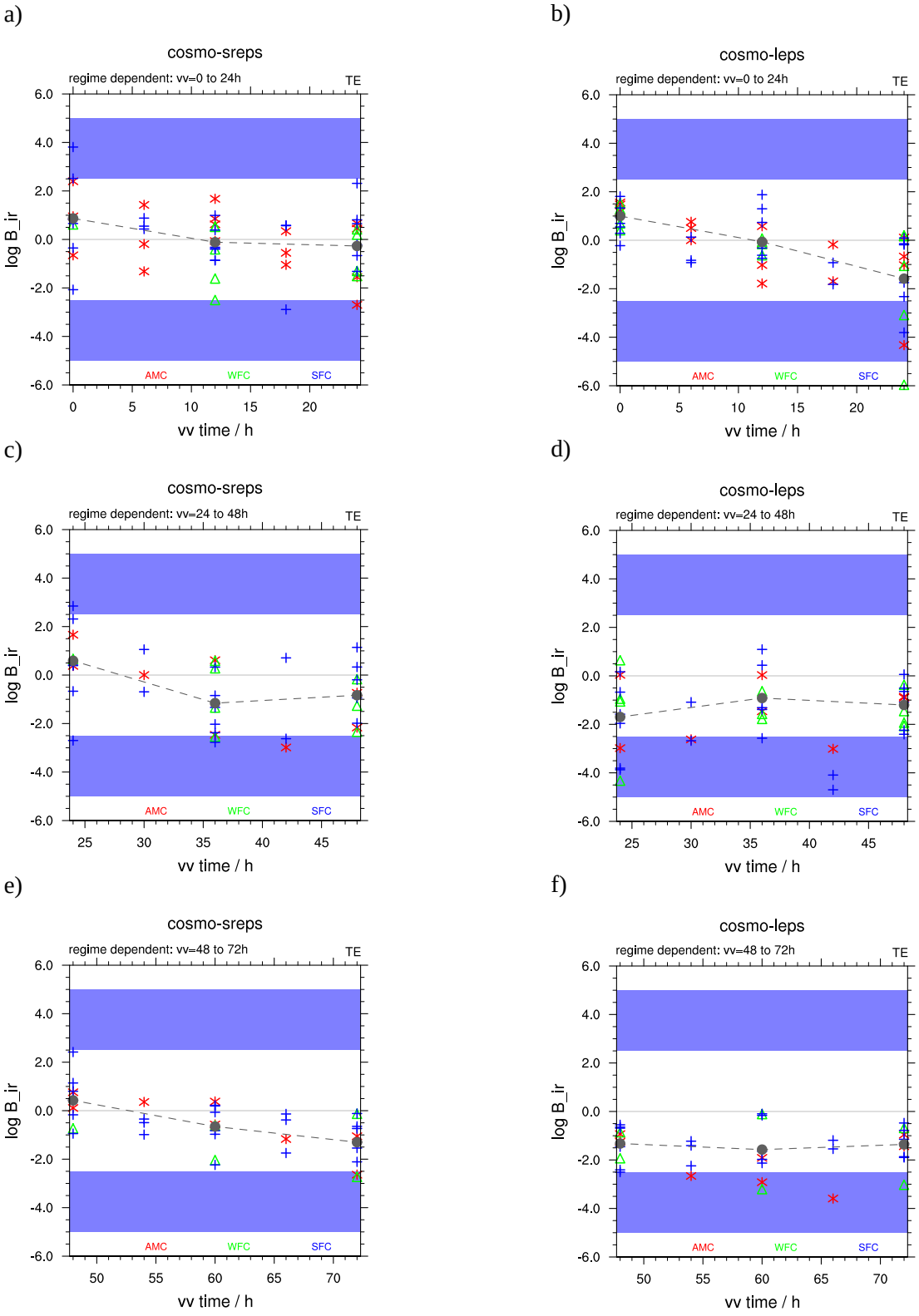



**Figure A.5:** Regime-dependend verification of the SREPS (a,c,e) and the LEPS (b,d,f) for the equivalent potential temperature. (a,b) cover the $0-24h$ forecast period, (c,d) the $24-48h$ and (e,f) the $48-72h$ forecast period. The convection types are AMC, WFC and SFC following the COPS-IOP classification of Wulfmeyer et al. (2011).

# Appendix B

# List of symbols

## Physical symbols

| Symbol | Description | Units |
|---|---|---|
| $B$ | Buoyancy | $\mathrm{kg\,m\,s^{-2}}$ |
| $c_p$ | Specific heat capacity of air | $\mathrm{J\,kg^{-1}\,K^{-1}}$ |
| $e$ | Vapour pressure over water | hPa |
| $E$ | Max. vapor pressure over water | hPa |
| $E_{tot}$ | Total specific internal energy | |
| $\varepsilon$ | Kinetic energy dissipation due to viscosity | |
| $g$ | Acceleration of gravity | $\mathrm{m\,s^{-2}}$ |
| $\Gamma$ | Temperature gradient | $\mathrm{K(100m)^{-1}}$ |
| $\Gamma_s$ | Wet-adiabatic Temperature gradient | $\mathrm{K(100m)^{-1}}$ |
| $\Gamma_d$ | Dry-adiabatic Temperature gradient | $\mathrm{K(100m)^{-1}}$ |
| $I$ | Term of sources/sinks | |
| $\mathbf{J}$ | Diffusion flux | |
| $k_b$ | Boltzmann constant | $\mathrm{J\,K^{-1}}$ |
| $L$ | Latent heat of evaporation | $\mathrm{J\,kg^{-1}}$ |
| $\Omega$ | Constant angular velocity of earth rotation | |
| $p$ | Pressure | hPa |
| $q_f$ | Specific humidity of water in the solid state | $\mathrm{kg\,kg^{-1}}$ |
| $q_l$ | Specific humidity of liquid water | $\mathrm{kg\,kg^{-1}}$ |
| $q_v$ | Specific humidity of water vapor | $\mathrm{kg\,kg^{-1}}$ |
| $m$ | Mixing ratio | $\mathrm{kg\,kg^{-1}}$ |
| $R$ | Universal gas constant | $\mathrm{J\,mol^{-1}\,K^{-1}}$ |
| $\mathbf{R}$ | flux density of solar and thermal radiation | |
| $R_a$ | Gas constant for air | $\mathrm{J\,mol^{-1}\,K^{-1}}$ |
| $R_w$ | Gas constant for water | $\mathrm{J\,mol^{-1}\,K^{-1}}$ |
| $rh$ | Relative humidity | - |

| | | |
|---|---|---|
| $\rho$ | Density | $\mathrm{kg\,m^{-3}}$ |
| $T$ | Air temperature | K |
| $t$ | Time | s |
| $\underline{\mathbf{t}}$ | Stress tensor due to viscosity | |
| $TOT\_PREC$ | Total precipitation | $\mathrm{kg\,m^{-2}}$ |
| $\Theta$ | Potential temperature | K |
| $\Theta_e$ | Equivalent potential temperature | K |
| $\tau_c$ | Convective timescale | h |
| $vv$ | Forecast lead time | h |
| $\mathbf{v} = (\mathbf{u}, \mathbf{v}, \mathbf{w})^{\mathbf{T}}$ | Wind vector | $\mathrm{m\,s^{-1}}$ |
| $z$ | Height | m |

## Statistical symbols

| Symbol | Description | Units |
|---|---|---|
| $B_{ir}$ | Bayes factor | - |
| $\varepsilon$ | Model uncertainty | |
| $E[X] = \mu$ | Expected value | |
| $E[(X-\mu)^2] = \sigma^2$ | Variance | |
| $f_i^k$ | Forecast state vector of ensemble $m_i$, member $k$ | |
| $f_t$ | True state vector | |
| $h_S$ | Silverman's factor | - |
| $I(f)$ | $I$-Operator | - |
| $K_i$ | Number of ensemble members | - |
| $l(o\|m_i)$ | Likelihood | - |
| $m_{i,ref}$ | Ensemble system $i$, $reference$ | - |
| $o$ | Observation vector | |
| $P(m_i)$ | Prior probability | - |
| $P(m_i\|o)$ | Posterior probability | - |
| $q$ | Dimension of $f_i^k$ and $o$ | - |
| $S$ | Estimator of $\Sigma_i$ | |
| $\Sigma_i$ | Error covariance matrix of the model $m_i$ | |
| $\Sigma_o$ | Error covariance matrix of the observations | |

# Appendix C

# List of abbreviations

| Abbreviation | Description |
|---|---|
| AC | Anomaly Correlation |
| AMC | Air-mass Convection |
| ARPA-SIMC | Agenzia Regionale Prevenzione e Ambiente dell Emilia-Romagna Servizio Idro-Meteo-Clima |
| BC | Boundary Conditions |
| CAPE | Convective Available Potential Energy |
| CDF | Cumulative Distribution Function |
| CI | Convective Initiation |
| CIN | Convective Inhibition |
| CMC | Canadian Weather Service |
| COPS | Convective and Orographically-induced Precipitation Study |
| COSMO | Consortium for Small-scale Modeling |
| CRPS | Continuous Rank Probability Score |
| D-Phase | Demonstration of Probabilistic Hydrological and Atmospheric Simulation of flood Events in the Alpine region |
| DE-EPS | COSMO-DE Ensemble Prediction System |
| DWD | Deutscher Wetterdienst (German Weather Service) |
| ECMWF | European Centre for Medium-Range Weather Forecasts |
| EL | Equilibrium level |
| EPS | Ensemble Prediction System |
| FDP | Forecast Demonstration Project |
| FRA | Meteo France |
| GFS | Global Forecast System |
| GME | Global Model at DWD |
| GOP | General Observation Period |
| IC | Initial Conditions |
| IFS | Integrated Forecast System |

| | |
|---|---|
| IGN | Ignorance Score |
| IOP | Intensive Observation Period |
| LEPS | COSMO Limited-Area Ensemble Prediction System |
| LFC | Level of free convection |
| ME | Mean Error |
| NCEP | National Centers for Environmental Prediction |
| NWP | Numerical Weather Prediction |
| PDF | Probability Density Function |
| QPF | Quantitative Precipitation Forecast |
| RMSE | Root Mean Squared Error |
| SKD | Standard Kernel Dressing |
| SFC | Strongly Forced Convection |
| SREPS | COSMO Short-Range Ensemble Prediction System |
| UK | UK Met Office |
| UKMO | see UK |
| UM | Unified Model |
| WFC | Weakly Forced Convection |
| WMO | World Meteorological Organization |
| WWRP | World Weather Research Programme |

# Bibliography


Ament, F., 2010: Quest 3 phases from 2004 to 2010. Available online at `http://gop.meteo.uni-koeln.de/ag\_crewell/`; visited on December 7th 2011.

Anwender, D., 2007: *Extratropical Transition in the Ensemble Prediction System of the ECMWF: case studies and experiments*. Dissertation der Fakultät für Physik an der Universität Karlsruhe, 184 pp.

Arpagaus, M., M. W. Rotach, P. Ambrosetti, F. Ament, C. Appenzeller, H.-S. Bauer, A. Behrendt, F. Bouttier, A. Buzzi, M. Corazza, S. Davolio, M. Denhard, M. Dorninger, L. Fontannaz, J. Frick, F. Fundel, U. Germann, T. Gorgas, G. Grossi, C. Hegg, A. Hering, S. Jaun, C. Keil, M. A. Liniger, C. Marsigli, R. McTaggart-Cowan, A. Montani, K. Mylne, L. Panziera, R. Ranzi, E. Richard, A. Rossa, D. Santos-Munoz, C. Schaer, Y. Seity, M. Staudinger, M. Stoll, S. Vogt, H. Volkert, A. Walser, Y. Wang, J. Werhahn, V. Wulfmeyer, C. Wunram, and M. Zappa, 2009: MAP D-Phase: Demonstrating forecast capabilities for flood events in the Alpine region. **78**, 75 pp.

Baldauf, M., A. Seifert, J. Förstner, D. Majewski, M. Raschendorfer, and T. Reinhardt, 2011: Operational convective-scale numerical weather prediction with the COSMO model. *Mon. Wea. Rev.*, **139**, 3887–3905.

Baldauf, M., K. Stephan, S. Klink, C. Schraff, A. Seifert, J. Förstner, T. Reinhardt, and C. J. Lenz, 2006: The new very short range forecast model LMK for the convection-resolving scale. *Second THORPEX International Science Symposium*, **Volume of extended abstracts, Part B**, 148–149.

Barthlott, C., J. W. Schipper, N. Kalthoff, B. Adler, C. Kottmeier, A. Blyth, and S. Mobbs, 2009: Model representation of boundary-layer convergence triggering deep convection over complex terrain: a case study from COPS. *Atmos. Res.*, **95**, 172–185.

Bechtold, P., E. Bazile, F. Guichard, P. Mascart, and E. Richard, 2001: A mass-flux convection scheme for regional and global models. *Quart. J. Roy. Met. Soc.*, **127**, 869–886.

Bennett, L. J., K. A. Browning, A. M. Blyth, D. J. Parker, and P. A. Clark, 2006: A review of the initiation of precipitating convection in the United Kingdom. *Quart. J. Roy. Met. Soc.*, **132**, 1001–1020.

Berger, J. O., 1985: *Statistical decision theory and Bayesian analysis*. 2nd ed. Springer Series in Statistics, Springer, New York, 618 pp.

Bolton, D., 1980: The computation of equivalent potential temperature. *Mon. Wea. Rev.*, **108**, 1046–1053.

Bougeault, P., Z. Toth, C. Bishop, B. Brown, D. Burridge, D. Chen, E. Ebert, M. Fuentes, T. Hamill, K. Mylne, J. Nicolau, T. Paccagnella, Y.-Y. Park, D. Parsons, B. Raoult, D. Schuster, P. S. Dias, R. Swinbank, Y. Takeuchi, W. Tennant, L. Wilson, and S. Worley, 2010: The THORPEX interactive grand global ensemble. *Bull. Am. Met. Soc.*, **91**, 1059–1072.

Bouttier, F. and P. Courtier, 1999: Data assimilation concepts and methods. *ECMWF - Lecture Notes NWP Course*, 58 pp.

Bowler, N. E., 2008: Accounting for the effect of observation errors on verification of mogreps. *J. Appl. Met.*, **15**, 199–205.

Bröcker, J. and L. Smith, 2007a: Scoring probabilistic forecasts: the importance of being proper. *Weather and Forecasting*, **22**, 382–388.

— 2007b: Scoring probabilitic forecasts: The impact of being proper. *Weather and Forecasting*, **22**, 382–388.

Bröcker, J. and L. A. Smith, 2008: From ensemble forecast to predictive distribution function. *Tellus*, **60**, 663–678.

Browning, K. A., A. M. Blyth, and P. A. Clark, 2008: The convective storm initiation project. *Bull. Am. Met. Soc.*, **88**, 1939–1955.

Buizza, R., 2001: Chaos and weather prediction. Available online at `http://www.ecmwf.int/staff/roberto\_buizza/Work/Lectures/Confs\_WSs\_Sems/`; visited on September 8th 2011.

Buizza, R., Y.-Y. Park, M. Leutbecher, and F. Pappenberger, 2008: Predictability studies using TIGGE data. *ECMWF Newsletter*, **No. 116**, 16–20.

Candille, G. and O. Talagrand, 2008a: Impact of observational error on the validation of ensemble prediction systems. *Quart. J. Roy. Met. Soc.*, **134**, 959–971.

— 2008b: Impact of observational error on the validation of ensemble prediction systems. *Quart. J. Roy. Met. Soc.*, **134**, 509–521.

Casati, B., L. J. Wilson, D. B. Stephenson, P. Nurmi, A. Ghelli, M. Pocernich, U. Damrath, E. E. Ebert, B. G. Brown, and S. Mason, 2008: Forecast verification: current status and future directions. *Meteorological Aplications*, **15**, 3–18.

COSMO-Website, 2011: Website of consortium for small-scale modeling (COSMO). URL `http://www.cosmo-model.org`

Crook, N. A., 1996: Sensitivity of moist convection forced by boundary layer processes to low-level thermodynamic fields. *Bull. Am. Met. Soc.*, **8**, 1767–1785.

Daley, R., 1993: *Atmospheric Data Analysis*. University Press, Cambridge, 457 pp.

Damrath, U., 2011: Standard-verification of GME at the Deutsche Wetterdienst (DWD). Internal note.

Davies-Jones, R., 2009: On formulas for equivalent potential temperature. *Mon. Wea. Rev.*, **137**, 3137–3148.

DelSole, T., 2004: Predictability and information theory. Part I: Measures of predictability. *J. Atmos. Sci.*, **61**, 2425–2440.

— 2005: Predictability and information theory. Part II: Imperfect forecasts. *J. Atmos. Sci.*, **62**, 3368–3381.

Dierer, S. and F. Schubig, 2008: Evaluation of the Kain-Fritsch/Bechtold convection scheme. *COSMO Newsletter*, **No. 7**, 37–42.

Doms, G. and J. Förstner, 2007: A description of the nonhydrostatic regional model lm. Part II: Physical parameterization. *Deutscher Wetterdienst (DWD), Offenbach*, 132 pp.

Doms, G. and U. Schättler, 2002: A description of the nonhydrostatic regional model lm. Part I: Dynamics and numerics. *Deutscher Wetterdienst (DWD), Offenbach*, 134 pp.

Doswell, C. A., 1987: The distinction between large-scale and mesoscale contribution to severe convection: A case study example. *Weather and Forecasting*, **2**, 3–16.

Ebert, B., 2002: Verifying satellite precipitation estimates for weather and hydrological applications. *IPWG International Precipitation Working Group*, **1st Workshop Proceedings**, 8 pp.

Ebert, B., F. Atger, M. Baldwin, K. Brill, H. Brooks, B. Brown, B. Casati, U. Damrath, C. Ferro, A. Ghelli, M. Göber, J. Jenkner, I. Jolliffe, C. Keil, T.-Y. Koh, P. Nurmi, P. Roebber, D. Stephenson, C. Wilson, and L. Wilson, 2010: Forecast verification: Issues, methods and FAQ. Website, visited on October 15th 2010.
URL `http://www.cawcr.gov.au/projects/verification/`

Ebert, B., U. Damrath, W. Wergen, and M. Baldwin, 2003a: Supplement to the WGNE assessment of short-term Quantitative Precipitation Forecasts. *Bull. Am. Met. Soc.*, **84**, 492–493.

— 2003b: The WGNE assessment of short-term Quantitative Precipitation Forecasts. *Bull. Am. Met. Soc.*, **84**, 481–492.

ECMWF, 2011: IFS documentation - CY37r2 Part II: Data assimilation. *European Centre for Medium-Range Weather Forecasts (ECMWF), Reading*, 174 pp.

Emanuel, K. A., 1994: *Atmospheric Convection*. Oxford University Press, New York, 580 pp.

Ermert, V., 2011: Radiosondenaufstiege in Europa. Available online at http://www.uni-koeln.de/math-nat-fak/geomet/meteo/; visited on September 10th 2011.

Fabry, F., 2006: The spatial variability of moisture in the boundary layer and its effect on convective initiation: Project-long characterization. *Mon. Wea. Rev.*, **134**, 79–91.

Friedman, J., T. Hastie, and R. Tibshirani, 2007: Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, **9**, 432–441.

Gebhardt, C., S. E. Theis, M. Paulat, and Z. B. Bouallegue, 2011: Uncertainties in COSMO-DE precipitation forecasts introduced by model pertubations and variation of lateral boundaries. *Atmos. Res.*, **100**, 168–177.

Gibbins, C. J., 1990: A survey and comparison of relationships for the determination of the saturation vapour pressure over plane surfaces of pure water and of pure ice. *Ann. Geophys.*, **8**, 859–885.

Gneiting, T., A. Raftery, A. Westveld, and T. Goldman, 2004: Calibrated probabilistic forecasting using ensemble output statistics and minimum crps estimation. *Mon. Wea. Rev.*, **133**, 1098–1118.

Gneiting, T. and E. A. Raftery, 2007: Strictly proper scoring rules, Prediction and estimation. *J. Americ. Statist. Assoc.*, **102**, 359–378.

Gneiting, T., L. I. Stanberry, E. P. Grimit, L. Held, and N. A. Johnson, 2008: Assessing probabilistic forecasts of multivariate quantities, with an application to ensemble predictions of surface winds. *Technical Report, Department of Statistics, University of Washington*, **537**, 26 pp.

Guichard, F., J. C. Petch, and J.-L. Redelsperger, 2004: Modelling the diurnal cycle of deep precipitation convection over land with cloud-resolving models and single-column models. *Quart. J. Roy. Met. Soc.*, **130**, 3139–3172.

Hackbusch, W., H. R. Schwarz, E. Zeidler, and I. N. Bronstein, 2003: *Teubner - Taschenbuch der Mathematik: TEIL I*. Teubner Verlag / GMV Fachverlag, Wiesbaden, 830 pp.

Hamill, T. M., 2001: Interpretation of rank histograms for verifying ensemble forecasts. *Mon. Wea. Rev.*, **129**, 550–560.

Hannachi, A., I. T. Jolliffe, and D. B. Stephenson, 2007: Empirical orthogonal functions and related techniques in atmospheric science: A review. *Int. J. Climatol*, **27**, 1119–1152.

Hense, A., G. Adrian, C. Kottmeier, C. Simmer, and V. Wulfmeyer, 2003: Das Schwerpunktsprogramm SPP 1167 der Deutschen Forschungsgemeinschaft "Quantitative Niederschlagsvorhersage". *Proposal of the DFG - SPP1167*, 8 pp.

— 2006: The german priority Program SPP1167 PQP Quantitative Precipitation Forecast: an overview - 2nd International Symposium on Quantitative Precipitation Forecasting (QPF) and Hydrology, Boulder, CO USA, June 4-8.

Hense, A. and U. Römer, 1995: Statistical analysis of tropical climate anomaly simulations. *Climate Dynamics*, **11**, 178–192.

Hense, A. and V. Wulfmeyer, 2008: The german priority programm SPP1167 "Quantitative Precipitation Forecast". *Meteorol. Zeitschr.*, **17**, 703–705.

Hersbach, H., 2000: Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather and Forecasting*, **15**, 559–570.

Houze, R. A., 1993: *Cloud Dynamics*. Academic Press, San Diego, 573 pp.

Jaynes, E. T., 1968: Prior probabilities. *System Scienes and Cybernetics*, **4**, 227–241.

— 2003: *Probability Theory*. University Press, Cambridge, 758 pp.

Jewson, S. and R. Caballero, 2003: The use of weather forecasts in the pricing of weather derivatives. *Meteorol. Appl.*, **00**, 1–13.

Johns, R. H. and C. A. Doswell, 1992: Severe local storms forecasting. *Weather and Forecasting*, **7**, 588–612.

Jonko, A. K., A. Hense, and J. J. Feddema, 2009: Effect of landcover change on the tropical circulation in a GCM. *Climate Dynamics*, doi:10.1007/s00382–009–0684–7.

Kain, J. S., 2003: The Kain-Fritsch convection parameterization: An update. *J. Appl. Met.*, **43**, 170–181.

Kain, J. S. and J. M. Fritsch, 1990: A one-dimensional entraining/detraining plume model and its application in convection parameterization. *J. Atmos. Sci.*, **47**, 2784–2802.

— 1993: Convective parameterization for mesoscale models: The Kain-Fritsch scheme. The representation of cumulus convection in numerical models. *Meteor. Monogr.*, **24**, 165–170.

Kaltenböck, R., 2003: The outbreak of severe storms along convergence lines northeast of the Alps. Case study of the 3 August 2001 mesoscale convective system with a pronounced bow echo. *Atmos. Res.*, **70**, 55–75.

Kalthoff, N., B. Adler, C. Barthlott, U. Corsmeier, S. Mobbs, S. Crewell, K. Träumner, C. Kottmeier, A. Wieser, V. Smithd, and P. D. Girolamo, 2009: The impact of convergence zones on the initiation of deep convection: a case study from COPS. *Atmos. Res.*, **93**, 680–694.

Kass, R. E. and A. E. Raftery, 1995: Bayes factors. *J. Americ. Statist. Assoc.*, **90**, 773–795.

Kass, R. E. and L. Wasserman, 1996: The selection of prior distributions by formal rules. *J. Americ. Statist. Assoc.*, **91**, 1343–1370.

Keil, C., 2000: *Numerische Simulation von Starkniederschlagsereignissen mit mesoskaligen Wettervorhersagemodellen*. Dissertation der Fakultät für Physik der Ludwig-Maximilians Universität München, 123 pp.

Keil, C. and G. Craig, 2011: Regime-dependent forecast of convective precipitation. *Meteorol. Zeitschr.*, **20**, 145–151.

Keil, C., A. Röpnack, G. Craig, and U. Schumann, 2008: Sensitivity of quantitative precipitation forecast to height dependent changes in humidity. *Geophys. Res. Lett.*, **35**, 1–5.

Keller, J. D. and A. Hense, 2011: A new non-gaussian evaluation method for ensemble forecasts based on analysis rank histograms. *Meteorol. Zeitschr.*, **20**, 107–117.

Kitchen, M., 1989: Representativeness errors for radiosonde observations. *Quart. J. Roy. Met. Soc.*, **115**, 673–700.

Knight, K. and W. Fu, 2000: Asymptotics for lasso-type estimators. *The Annals of Statistics*, **28**, 1356–1378.

Kolmogorov, A. N., 1933: *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Springer, Berlin, 63 pp.

Kottmeier, C., N. Kalthoff, C. Barthlott, U. Corsmeier, J. van Baelen, A. Behrendt, R. Behrendt, A. Blyth, R. Coulter, S. Crewell, P. di Girolamo, M. Dorninger, C. Flamant, T. Foken, M. Hagen, C. Hauck, H. Höller, H. Konow, M. Kunz, H. Mahlke, S. Mobbs, E. Richard, R. Steinacker, T. Weckwerth, A. Wieser, and V. Wulfmeyer, 2008: Mechanisms initiating deep convection over complex terrain during COPS. *Meteorol. Zeitschr.*, **17**, 931–948.

Kraus, H., 2004: *Die Atmosphäre der Erde*. Springer-Verlag, Berlin Heidelberg, 422 pp.

Kuell, V., A. Gassmann, and A. Bott, 2007: Towards a new hybrid cumulus parametrization scheme for use in non-hydrostatic weather prediction models. *Quart. J. Roy. Met. Soc.*, **133**, 479–490.

Lawrence, M. G., 2005: The relationship between relative humidity and the dewpoint temperature in moist air. a simple conversion and applications. *Bull. Am. Met. Soc.*, **86**, 225–233.

Lorenz, E. N., 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.*, **20**, 130–141.

— 1984: Irregularity: a fundamental property of the atmosphere. *Tellus*, **36**.

Maesschalck, R. D., D. Jouan-Rimbaud, and D. L. Massart, 2000: The Mahalanobis distance. *Chemometrics and Intelligent Laboratory Systems*, **50**, 1–18.

Mahalanobis, P. C., 1936: On the generalised distance in statistics. *Proceedings of the National Institute of Science of India*, **12**, 49–55.

Majewski, D., 2008: Further development of HRM. Available online at `http://www.met.gov.om/`; visited on May 2nd 2012.

Malakoff, D., 1999: Statistics: A brief guide to Bayes Theorem. *Science*, **286**, 1461.

Marsigli, C., A. Montani, and T. Paccagnella, 2007: Ensemble activities at ARPA-SIM: the COSMO-LEPS and COSMO-SREPS systems. *29th International Conference on Alpine Meteorology*, **Extended abstracts**, available online at `http://www.cnrm.meteo.fr/icam2007/ICAM2007/extended/manuscript\_118.pdf`; visited on March 1st 2013.

— 2008: The COSMO-SREPS ensemble for short-range: system analysis and verification on the MAP-DPhase DOP. *Proceedings of the Joint MAP-DPhase Scientific Meeting-COST 731 mid-term seminar, 19-22 May 2008, Bologna, Italy*, 9–14.

Miloshevich, L. M., H. Vomel, D. N. Whiteman, and T. Leblanc, 2009: Accuracy assessment and correction of Vaisala RS92 radiosonde water vapor measurements. *J. Geophys. Res.*, **114**, 23 pp.

Min, S.-K. and A. Hense, 2006: A Bayesian approach to climate model evaluation and multi-model averaging with an application to global mean surface temperatures from IPCC AR4 coupled climate models. *Geophys. Res. Lett.*, **33**.

Min, S.-K., A. Hense, H. Paeth, and W.-T. Kwon, 2004: A Bayesian decision method for climate change signal analysis. *Meteorol. Zeitschr.*, **13**, 421–436.

Montani, A., M. Capaldo, D. Cesari, C. Marsigli, U. Modigliani, F. Nerozzi, T. Paccagnella, P. Patruno, and S.Tibaldi, 2003: Operational limited-area ensemble forecasts based on the Lokal Modell. *ECMWF Newsletter*, **No. 98**, 2–7.

Montani, A., C. Marsigli, and T. Paccagnella, 2007: Five years of limited-area ensemble activities at ARPA-SIM: the COSMO-LEPS system. *11th Workshop on Meteorological Operational Systems, Reading, 12-16 November 2007*.

Murphy, A. H. and R. L. Winkler, 1984: Probability forecasting in meteorology. *J. Americ. Statist. Assoc.*, **79**, 489–500.

— 1987: A general framework for forecast verification. *Mon. Wea. Rev.*, **115**, 1330–1338.

NIST/SEMATECH, 2010: e-handbook of statistical methods. Available online at `http://www.itl.nist.gov/div898/handbook/`; visited on November 8th 2011.

NOAA/ESRL, 2011: Schematic vertical profile of the troposphere. Available online at `http://laps.noaa.gov/afwa/eds\_stuff/`; visited on September 9th 2011.

Ohlwein, C., P. Friedrich, A. Hense, A. Kapala, and S. Crewell, 2011: Retrospective analysis of regional climate. Available online at `http://www.herz-tb4.uni-bonn.de/`; visited on October 31th 2011.

Orlanski, I., 1975: A rational subdivision of scales for atmospheric processes. *Bull. Am. Met. Soc.*, **56**, 527–530.

Peralta, C., Z. B. Bouallegue, S. E. Theis, C. Gebhardt, and M. Buchhold, 2012: Accounting for initial condition uncertainties in COSMO-DE-EPS. *J. Geophys. Res.*, **117**, 13 pp.

Pflüger, U., 2007: Standard-temp-verification of COSMO-DE at the Deutsche Wetterdienst (DWD). Internal note.

Plant, R. S. and G. C. Craig, 2008: A stochastic parameterization for deep convection based on equilibrium statistics. *J. Atmos. Sci.*, **65**, 87–105.

Rade, L. and B. Westergren, 2000: *Mathematische Formeln*. Springer, Berlin Heidelberg, 551 pp.

Raftery, A. E., T. Gneiting, F. Balabdaoui, and M. Polakowski, 2005: Using Bayesian model averaging to calibrate forecast ensembles. *Mon. Wea. Rev.*, **133**, 1155–1174.

Redelsperger, J. L., D. B. Parsons, and F. Guichard, 2002: Recovery processes and factors limiting cloud-top height following the arrival of a dry intrusion observed during toga coare. *J. Atmos. Sci.*, **59**, 2438–2457.

Richardson, D. S., 2000: Skill and relative economic value of the ECMWF ensemble prediction system. *Quart. J. Roy. Met. Soc.*, **126**, 649–667.

Rossa, A., P. Nurmi, and E. Ebert, 2008: Overview of methods for the verification of quantitative precipitation forecasts. *Precipitation: Advances in Measurement, Estimation and Prediction*, 418–450.

Sably, M. L., 1996: *Fundamentals of Atmospheric Physics*. Academic Press Elsevier, San Diego, CA, 627 pp.

Schölzel, C. and A. Hense, 2010: Probabilistic assessment of regional climate change in Southwest Germany by ensemble dressing. *Climate Dynamics*, doi:10.1007/s00382–010–0815–1.

Schraff, C. and R. Hess, 2003: A description of the nonhydrostatic regional model lm. Part III: Data assimilation. *Deutscher Wetterdienst (DWD), Offenbach*, 79 pp.

Silverman, B. W., 1986: *Density Estimation for Statisticas and Data Analysis*. Monographs on Statistics and Applied Probability 26, Chapman and Hall/CRC, London/Boca Raton, 175 pp.

Simmons, A. J. and A. Hollingsworth, 2002: Some aspects of the improvement in skill of numerical weather prediction. *Quart. J. Roy. Met. Soc.*, **128**, 647–677.

Sole, T. D. and M. K. Tippett, 2007: Predictability: Recent insights from information theory. *Rev. Geophys.*, **45**, doi:10.1029/2006RG000202.

Tiedtke, M., 1989: A comprehensive mass flux scheme for cumulus parameterization in large-scale models. *Mon. Wea. Rev.*, **117**, 1779–1800.

Ueno, G. and T. Tsuchiya, 2009: Covariance regularization in reverse space. *Quart. J. Roy. Met. Soc.*, **135**, 1133–1156.

Uppala, S., P. Kallberg, A. Hernandez, S. Saarinen, M. Fiorino, X. Li, K. Onogi, N. Sokka, U. Andrae, and V. D. C. Bechtold, 2004: ERA-40: ECMWF 45-year reanalysis of the global atmosphere and surface conditions 1957-2002. *ECMWF Newsletter*, **No. 101**, 2–21.

Vaisala, 2010: Vaisala radiosonde RS92-D. Available online at http://www.vaisala.com; visited on November 27th 2011.

Wainwright, M. J. and M. I. Jordan, 2008: Graphical models, exponential families, and variational inference. *Found. Trends Mach. Learn.*, **1**, 1–305, doi:10.1561/2200000001.
URL http://dl.acm.org/citation.cfm?id=1498840.1498841

Walser, A. and M. A. Liniger, 2008: Probabilistic verifcation of COSMO-LEPS forecasts with SYNOP stations. *COSMO Newsletter*, **No. 7**, 49–56.

Weckwerth, T. M., D. B. Parsons, S. E. Koch, J. A. Moore, M. A. Lemone, B. B. Demoz, C. Flamant, B. Geerts, J. Wang, and W. F. Feltz, 2004: An overview of the international h2o project (ihop_2002) and some prelininary highlights. *Bull. Am. Met. Soc.*, **85**, 253–277.

Wernli, H., M. Paulat, M. Hagen, and C. Frei, 2007: SAL - a novel quality measure for the verification of quantitative precipitation forcasts. *Mon. Wea. Rev.*, **136**, 4470–4487.

Wetterzentrale, 2011: Website of weather forecasts.
URL http://www.wetterzentrale.de

White, P. W., 2003a: IFS documentation - CY25r1 Part I: Observation processing. *European Centre for Medium-Range Weather Forecasts (ECMWF), Reading*, 160 pp.

— 2003b: IFS documentation - CY25r1 Part VI: Technical and computational procedures. *European Centre for Medium-Range Weather Forecasts (ECMWF), Reading*, 165 pp.

Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Science*. International Geophysics Series, Academic Press, San Diego, 467 pp.

Wilks, D. S. and T. M. Hamill, 2007: Comparison of Ensemble-MOS Methods using GFS Reforecasts. *Mon. Wea. Rev.*, **135**, 2379–2390.

Wulfmeyer, V., A. Behrendt, and H.-S. Bauer, 2008: The Convective and Orographically-induced Precipitation study. *Bull. Am. Met. Soc.*, **89**, 1477–1486.

Wulfmeyer, V., A. Behrendt, C. Kottmeier, U. Corsmeier, C. Barthlott, G. C. Craig, M. Hagen, D. Althausen, F. Aoshima, M. Arpagaus, H.-S. Bauer, L. Bennett, A. Blyth, C. Brandau, C. Champollion, S. Crewell, G. Dick, P. D. Girolamo, M. Dorninger, Y. Dufournet, R. Eigenmann, R. Engelmann, C. Flamant, T. Foken, T. Gorgas, M. Grzeschik, J. Handwerker, C. Hauck, H. Höller, W. Junkermann, N. Kalthoff, C. Kiemle, S. Klink, M. König, L. Krauss, C. N. Long, F. Madonna, S. Mobbs, B. Neininger, S. Pal, G. Peters, G. Pigeon, E. Richard, M. W. Rotach, H. Russchenberg, T. Schwitalla, V. Smith, R. Steinacker, J. Trentmann, D. D. Turner, J. van Baelen, S. Vogt, H. Volkert, T. Weckwerth, H. Wernli, A. Wieser, and M. Wirth, 2011: The Convective and Orographically-induced Precipitation study (COPS): the scientific strategy, the field phase, and research highlights. *Quart. J. Roy. Met. Soc.*, **137**, 3–30.

Yanai, M. S., S. Esbensen, and J. Chu, 1973: Determination of bulk properties of tropical cloud clusters from large-scale heat and moisture budgets. *J. Atmos. Sci.*, **30**, 611–627.

Zimmer, M., G. C. Craig, C. Keil, and H. Wernli, 2011: Classification of precipitation events with a convective response timescale and their forecasting characteristics. *Geophys. Res. Lett.*, **38**, doi:10.1029/2010GL046199.

# BONNER METEOROLOGISCHE ABHANDLUNGEN

Herausgegeben vom Meteorologischen Institut der Universität Bonn durch Prof. Dr. H. FLOHN (Hefte 1-25), Prof. Dr. M. HANTEL (Hefte 26-35), Prof. Dr. H.-D. SCHILLING (Hefte 36-39), Prof. Dr. H. KRAUS (Hefte 40-49), ab Heft 50 durch Prof. Dr. A. HENSE.

Heft 1-39: siehe `http://www2.meteo.uni-bonn.de/bibliothek/bma.html`

Heft 40: ***Hermann Flohn***: Meteorologie im Übergang Erfahrungen und Erinnerungen (1931-1991). 1992, 81 S. + XII. € 23

Heft 41: ***Adnan Alkhalaf and Helmut Kraus***: Energy Balance Equivalents to the Köppen-Geiger Climatic Regions. 1993, 69 S. + IX. € 19

Heft 42: ***Axel Gabriel***: Analyse stark nichtlinearer Dynamik am Beispiel einer rei- bungs-freien 2D-Bodenkaltfront. 1993, 127 S. + XIV. € 30

Heft 43: ***Annette Münzenberg-St.Denis***: Quasilineare Instabilitätsanalyse und ihre Anwendung auf die Strukturaufklärung von Mesozyklonen im östlichen Weddellmeergebiet. 1994, 131 S. + XIII. € 33

Heft 44: ***Hermann Mächel***: Variabilität der Aktionszentren der bodennahen Zirkulation über dem Atlantik im Zeitraum 1881-1989. 1995, 188 S. + XX. € 48

Heft 45: ***Günther Heinemann***: Polare Mesozyklonen. 1995, 157 S. + XVI. € 46

Heft 46: ***Joachim KlaSSen:*** Wechselwirkung der Klima-Subsysteme Atmosphäre, Meereis und Ozean im Bereich einer Weddellmeer-Polynia. 1996, 146 S. + XVI. € 43

Heft 47: ***Kai Born***: Seewindzirkulationen: Numerische Simulationen der Seewind- front. 1996, 170 S. + XVI. € 48

Heft 48: ***Michael Lambrecht***: Numerische Untersuchungen zur tropischen 30-60-tägigen Oszillation mit einem konzeptionellen Modell. 1996, 48 S. + XII. € 15

Heft 49: ***Cäcilia Ewenz***: Seewindfronten in Australien: flugzeuggestützte Messungen und Modellergebnisse. 1999, 93 S. + X. € 30

Heft 50: ***Petra Friederichs***: Interannuelle und dekadische Variabilität der atmosphärischen Zirkulation in gekoppelten und SST-getriebenen GCM-Experimenten. 2000, 133 S. + VIII. € 25

Heft 51: ***Heiko Paeth***: Anthropogene Klimaänderungen auf der Nordhemisphäre und die Rolle der Nordatlantik-Oszillation. 2000, 168 S.+ XVIII. € 28

Heft 52: ***Hildegard Steinhorst***: Statistisch-dynamische Verbundsanalyse von zeitlich und räumlich hoch aufgelösten Niederschlagsmustern: eine Untersuchung am Beispiel der Gebiete von Köln und Bonn. 2000, 146 S. + XIV. € 25

Heft 53: ***Thomas Klein***: Katabatic winds over Greenland and Antartica and their interaction with mesoscale and synoptic-scale weather systems: three-dimensional numerical models. 2000, 146 S. + XIV. € 25

Heft 54: ***Clemens Drüe***: Experimentelle Untersuchung arktischer Grenzschichtfronten an der Meereisgrenze in der Davis-StraSSe. 2001, 165 S. + VIII. €

Heft 55: ***Gisela Seuffert***: Two approaches to improve the simulation of near surface processes in numerical weather prediction models. 2001, 128 S. + VI. € 25

Heft 56: ***Jochen Stuck***: Die simulierte axiale atmosphärische Drehimpulsbilanz des ECHAM3-T21 GCM. 2002, 202 S. + VII. € 30

Heft 57: ***Günther Haase***: A physical initialization algorithm for non-hydrostatic weather prediction models using radar derived rain rates. 2002, 106S. + IV. € 25

Heft 58: ***Judith Berner***: Detection and Stochastic Modeling of Nonlinear Signatures in the Geopotential Height Field of an Atmospheric General Circulation Model. 2003, 157 S. + VIII. € 28

Heft 59: ***Bernd Maurer***: Messungen in der atmosphärischen Grenzschicht und Validation eines mesoskaligen Atmosphärenmodells über heterogenen Landoberflächen. 2003, 182 S. + IX. € 30

Heft 60: ***Christoph Gebhardt***: Variational reconstruction of Quaternary temperature fields using mixture models as botanical – climatological transfer functions. 2003, 204 S. + VIII. € 30

Heft 61: ***Heiko Paeth***: The climate of tropical and northern Africa – A statistical-dynamical analysis of the key factors in climate variability and the role of human activity in future climate change. 2005, 316 S. + XVI. € 15

Heft 62: ***Christian Schölzel***: Palaeoenvironmental transfer functions in a Bayesian framework with application to Holocene climate variability in the Near East. 2006, 104 S. + VI. € 15

Heft 63: ***Susanne Bachner***: Daily precipitation characteristics simulated by a regional climate model, including their sensitivity to model physics, 2008, 161 S. € 15

Heft 64: ***Michael Weniger***: Stochastic parameterization: a rigorous approach to stochastic three-dimensional primitive equations, 2014, 148 S. + XV. open access

Heft 65: ***Andreas Röpnack***: Bayesian model verification: predictability of convective conditions based on EPS forecasts and observations, 2014, 152 S. open access[1]

[1] Available at `http://hss.ulb.uni-bonn.de/fakultaet/math-nat/`

METEOROLOGISCHES INSTITUT
MATHEMATISCH NATURWISSENSCHAFTLICHE FAKULTÄT
UNIVERSITÄT BONN

Meteorologisches Institut
Universität Bonn