



universität
wien

DIPLOMARBEIT

Titel der Diplomarbeit

Evaluation of the added value of LAMEPS in contrast
to global EPS

Verfasserin

Karin Schmeisser

angestrebter akademischer Grad

Magistra der Naturwissenschaften (Mag. rer. nat.)

Wien, März 2012

Studienkennzahl: A 415

Studienrichtung: Meteorologie

Betreuer: Ao. Univ. Prof. Mag. Dr. Leopold Haimberger

Contents

1	Abstract	4
2	Zusammenfassung	5
3	Introduction and Motivation	6
4	Numerical weather prediction models (NWP)	9
4.1	History of NWP development	9
4.2	The ECMWF global atmospheric model	11
4.3	The limited area model ALADIN-AUSTRIA	13
5	EPS	16
5.1	Basic Principles and Sources of uncertainty	16
5.2	A comparison of global ensemble prediction systems	18
5.3	The ECMWF ensemble prediction system	19
5.3.1	Model characteristics	19
5.3.2	Singular vector method	20
5.4	The Central European limited area forecasting system: ALADIN - LAEF	23
5.4.1	Model characteristics	23
5.4.2	The breeding method	23
5.4.3	The blending mechanism	25
5.4.4	Physical parameterization - Multi physics	26
5.4.5	Generation of surface level IC perturbations	28
5.4.6	Generation of upper level IC perturbations	30
5.4.7	Generation of the 16 ALADIN-LAEF members	31
5.4.8	Example of the blending method	32
6	Data	36
7	Verification	37
7.1	The purpose of Forecast Verification	37
7.2	Scalar Attributes of a Forecast	40
7.3	Binary Events	42
7.4	Probability Considerations	43
7.5	Verification of Ensemble Forecasts	45
7.5.1	Reliability and Resolution	45
7.5.2	Bias, RMSE and Ensemble Spread	46
7.5.3	The Brier Score and Brier Skill Score	47
7.5.4	Ranked Probability Score and Ranked Probability Skill Score . . .	49
7.5.5	Continuous Ranked Probability Score and Continuous Ranked Probability Skill Score	50
7.5.6	The ROC diagram	50
7.5.7	Talagrand diagram	51
7.5.8	Statistical Significance	52

8	Results	55
8.1	Verification of surface fields	55
8.1.1	Mean Sea Level Pressure	56
8.1.2	Temperature Anomaly	64
8.1.3	Wind Speed	70
8.1.4	Total Precipitation	76
8.1.5	Short Summary	85
8.2	Verification of upper level fields	86
8.2.1	Geopotential in 500 hPa-level	87
8.2.2	Geopotential in 850 hPa-level	90
8.2.3	Temperature Anomaly in 500 hPa	93
8.2.4	Temperature Anomaly in 850 hPa	98
8.2.5	Wind speed in 500 hPa	100
8.2.6	Wind speed in 850 hPa	104
8.2.7	Relative Humditiy in 500 hPa	107
8.2.8	Relative Humidity in 850 hPa	111
8.2.9	Short Summary	117
9	Statistical Post-Processing of EPS forecasts	119
9.1	Calibration	119
9.2	Previous investigations to calibrate ALADIN-LAEF	122
10	Conclusion	125
11	Appendix	129
12	Danksagung	173
13	Eidesstattliche Erklärung	175

1 Abstract

The limited area ensemble prediction system ALADIN-LAEF (Aire Limitée Adaption Dynamique Développement International - Limited Area Forecasting) has been developed in an effort to better predict high impact weather. Since it is operated on a higher horizontal resolution than its global counterpart the ECMWF-EPS (European Centre for Medium Range Weather Forecasts-Ensemble Prediction System), effects through local topography can be better resolved. In the first version of ALADIN-LAEF, initial perturbations were generated by dynamically downscaling the initial perturbations provided by the ECMWF-EPS. After a period of research a newer version was launched in February 2009. This version contained a method called BBSM (breeding blending surface multi-physics) which was based on a mechanism called "breeding". Breeding a numerical model means to initialize a model "pre-run". The method is based on the notion that during the integration of the numerical model dominant errors evolve, whereas other disappear. Therefore the resulting field accounts for involved uncertainties. ALADIN-LAEF and the ECMWF-EPS differ not only in regard to the generation of the initial perturbations, but also in regard to their physical configurations. In the ECMWF-EPS the physical settings are based on stochastic physics, whereas in ALADIN-LAEF each ensemble members is linked to different physical parameterization schemes. In an effort to filter the added value of the LAMEPS ALADIN-LAEF in contrast to the global ECMWF-EPS, a verification has been performed. Therefore the forecasts of a 2 months period have been applied (20. June - 20. August 2007). The performance of both models has been assessed using multiple verification scores such as Bias, Brier Score and CRPS. Results of the verification of the surface field show that ALADIN-LAEF generated superior wind speed forecasts as well as precipitation forecasts. In contrast the ECMWF-EPS turned out to produce better mean sea level pressure forecasts as well as 2 m temperature forecasts. However it should be noted that the temperature forecasts generated by ALADIN-LAEF can easily be improved by statistical post-processing. Results of the verification of the upper level field reveal a superior performance of the ECMWF-EPS in regard to the prediction of wind speed, relative humidity as well as the 500 hPa geopotential. Only the 500 hPa temperature forecasts of ALADIN-LAEF are better than those of the ECMWF-EPS. This result should be linked to the notable systematic error in the temperature forecasts of the ECMWF-EPS. Since systematic errors have a remarkable impact on the quality of the forecasts, calibration methods should be considered. In a first attempt to correct the systematic error in the 2 m temperature forecasts of ALADIN-LAEF, a calibration method has been successfully tested at ZAMG. In conclusion it has been found that the ability of ALADIN-LAEF to improve the ensemble forecasts generated by the ECMWF-EPS strongly depends on the selected parameter.

2 Zusammenfassung

Um eine genauere Vorhersage von signifikanten Wetterereignissen zu ermöglichen, wurde das regionale Ensemblevorhersagesystem ALADIN-LAEF (Aire Limitée Adaption Dynamique Développement International - Limited Area Forecasting) entwickelt. Dieses kann aufgrund seiner deutlich höheren Auflösung im Vergleich zu globalen Ensemblevorhersagesystemen, lokale Gegebenheiten wie Gebirge besser berücksichtigen. Aus diesem Grund nimmt man an, dass die Qualität der Prognosen klein skaliger Wetterphänomene durch den Einsatz von regionalen Ensemblevorhersagesystemen steigt. Um diese Annahme zu überprüfen, wurden die Prognosen von ALADIN-LAEF und dem globalen ECMWF-EPS (European Centre for Medium Range Weather Forecasts - Ensemble Prediction System) verifiziert und anschließend verglichen. Ziel der Verifikation war es, potentielle Vorteile des regionalen Ensemblevorhersagesystems ALADIN-LAEF herauszufiltern und Nachteile zu lokalisieren. Für die Verifikation wurden sämtliche Prognosen im Zeitraum vom 20. Juni 2007 bis 20. August 2007 überprüft. Zur Beurteilung der Prognosequalität wurde eine breite Auswahl an VerifikationsmaSSen berechnet. Dazu zählen beispielsweise Bias, Brier Score oder auch CRPS. Der Vergleich der Ergebnisse der Bodenfelder zeigt, dass ALADIN-LAEF bessere Wind- und Niederschlagsprognosen erzeugt. Im Gegensatz dazu produziert das ECMWF-EPS deutlich bessere Temperatur- und Bodendruckprognosen. Dabei sollte erwähnt werden, dass vor allem die Qualität der Temperaturprognosen von ALADIN-LAEF durch Kalibration deutlich erhöht werden könnte. Die Verifikation der Höhenfelder zeigt, dass das ECMWF-EPS deutlich bessere Wind-, Feuchte- und 500 hPa Geopotential-Prognosen erzeugt. Dagegen produziert ALADIN-LAEF bessere Temperaturprognosen für das 500 hPa-Feld. Dabei muss erwähnt werden, dass die 500 hPa Temperaturprognosen des ECMWF-EPS einen deutlichen Bias aufweisen. Da systematische Fehler einen beträchtlichen Einfluss auf die Qualität der Prognosen haben, sind Korrekturmetho-den von besonderem Interesse. Kalibrationsmethoden für die 2 m Temperaturprognosen wurden an der ZAMG bereits erfolgreich getestet. Fazit dieser Untersuchung ist, dass der Mehrwert von ALADIN-LAEF im Vergleich zum ECMWF-EPS stark von dem betrachteten Parameter abhängt.

3 Introduction and Motivation

"Ensemble forecasting has proved to be a successful way of dealing with the inherent uncertainty of weather and climate forecasts." (Buizza et al., 2005)

Since the atmosphere is a highly non-linear system, even small differences in the initial conditions can cause striking differences in the evolution. Although the development of sophisticated numerical weather prediction models was quite successful during the past decades, there are still sources of uncertainty. Beyond others the main components of uncertainty are the initial conditions and model errors. The first operational ensemble forecasting systems were launched in December 1992, both ECMWF-EPS (European Centre for Medium Range Forecasts-Ensemble Prediction System, UK) and NCEP-EPS (National Centers for Environmental Prediction, USA). In principle an ensemble with several members to predict different possible evolutions of the atmosphere offers a better chance to "catch" the real development than 1 deterministic model. The main issue for running an ensemble is the generation of appropriate initial perturbations as well as the consideration of model errors. This is a very challenging task since the atmosphere is a highly non-linear system with features at distinct scales such as upper level flow or convection. Generally a global EPS runs at a coarser horizontal resolution than the deterministic forecast ranging from 30 to 50 km. Therefore it cannot resolve local weather events like orographically induced high precipitation. Since large parts of Austria are located in the Alps, limited area models that can better represent these phenomena are of special interest. In many cases the extreme weather events are concentrated on small areas and they can cause enormous damage, e.g. heavy precipitation that regularly causes floods in Austria, or hail. Global EPS like ECMWF-EPS or NCEP-EPS have been designed for forecast lead times ranging between three and ten days. Thinking of severe small scale weather phenomena, there is also significant forecast uncertainty on shorter time and length scales. In an effort to meet those demands a Limited Area Model EPS denoted ALADIN-LAEF (Airé Limitée Dynamique Développement International- Limited Area Ensemble Forecasting) was developed and became preoperational at ZAMG (Zentralanstalt für Meteorologie und Geodynamik) in 2007. After a period of research a newer version was launched in February 2009. This survey concentrates on the comparison of two distinct EPS systems, a regional one (limited area ensemble - ALADIN-LAEF) and a global one (ECMWF-EPS).

Preceding studies on this topic have mostly concentrated on prominent weather events and only a few were performed over a longer period of time. Here is a brief review of the development of limited area ensemble prediction systems at various meteorological services:

Encouraged by the success of global medium range ensemble systems, pioneer work concerning LAMEPS was done at NCEP in the early 90's (Toth and Kalnay, 1997). Therefore the so called "breeding method" which had been previously developed for the global system was adapted to a regional ensemble prediction system. This breeding method basically consists of a "pre run" of the NWP- model, whereas errors due to the data assimilation scheme as well as due to the model uncertainties could evolve, producing initial perturbations. For the NCEP SREF (Short Range Ensemble Forecasting) the lateral boundary conditions were provided by the global ensemble system (Trac-

ton et al., 1998). NCEP SREF has been implemented operationally in May 2001, being the first real-time, operational regional ensemble prediction system in the world. In an effort to better predict heavy rainfall events, the Italian hydro meteorological Service (ARPA-SIM) in Cooperation with COSMO (Consortium for Small Scale Modeling) established a limited area ensemble prediction system, the COSMO-LEPS. COSMO is a consortium of Germany, Italy, Switzerland, Greece and Poland which aims to develop, improve and maintain the non-hydrostatic limited area model, formerly called Lokal Modell (Marsigli et al., 2005). The COSMO-LEPS contained a special algorithm to select certain members out of a global ensemble system, whereas for the global model the 51-member ECMWF-EPS has been used. The selected ensemble members (called Representative Members) were finally downscaled and represented the initial perturbations that could be implemented into the COSMO-LEPS. In an effort to assess the model performance a 3 months period (September-November 2003) has been verified. Therefore the forecast precipitation accumulated over 24 hours was compared to the observed data. A comparison in terms of average precipitation values showed that the ECMWF-EPS (51 members) performed better. In contrast when the comparison was carried out in terms of maximum values, the COSMO-LEPS revealed better results. In consequence addressing the regional model will be beneficial in case of intense and localized events. Next to comparing the regular COSMO-LEPS setting to the ECMWF-EPS, different configurations have been tested. As it turned out doubling the Representative Members (from 5 to 10) produced the greatest improvement of skill. Moreover the implementation of time lagged EPS systems seemed to be beneficial. Finally these results lead to a modification of the COSMO-LEPS at the beginning of June 2004: "The super-ensemble has been built using the 2 most recent EPS and the number of clusters has been fixed to 10, nesting Local Model on each of the so selected 10 Representative Members"(Marsigli et al., 2005).

Another investigation concerning LAMEPS was done by the Croatian meteorological and hydrological Service (CHMS) in 2004. A limited area EPS was generated by dynamically downscaling the global ECMWF-EPS (T_L255) members, adapting them to the 12.2km Croatian ALADIN -version (Brankovic et al., 2008). In contrast to COSMO-EPS there was no selection of representative members and in consequence no loss of information. Because of a relatively large computational demand using all 50 ECMWF-EPS members, such an approach might only be feasible for case studies. For the selection of information an identical clustering method was applied to the ECMWF members as well as to the downscaled ensemble members. Apart from testing the model performance in contrast to the ECMWF-EPS, the clustering methods were being verified. The intention was to find out how clustering could affect the results. In an effort to evaluate the model performance 4 special cases of Severe Weather (period: summer2003-autumn2004) were investigated. Results indicated that the identical clustering algorithm may yield differing results when applied to either global or to downscaled ensembles. The rationale might be the fact that a downscaled, higher-resolution ensemble resolves more explicitly small-scale features, in particular those strongly influenced by orographic forcing. This conclusion has important implications in limited area ensemble prediction, since it implies that downscaling may affect the interpretation or relevance of global ensemble forecasts. In consequence it might not always be feasible to choose a selection of representative global lower-resolution ensemble members and scale them down since representative members of the global ensemble are not necessarily representative members of the limited area ensemble.

At the UK Met Office a short-range ensemble prediction system called MOGREPS (Met Office Global and Regional Ensemble Prediction System) has been developed. It con-

sists of a global EPS and a regional EPS that can be referred to as NAE (North Atlantic Europe). The global ensemble provides the boundary conditions and the initial perturbations for the regional ensemble (Bowler et al., 2008). In contrast to other systems, perturbations to the initial conditions are being calculated using the ensemble transform Kalman filter (ETKF). Basically the Kalman filter provides an optimal estimate of the future state of a system given certain conditions. Therefore the estimate of the true state is updated by forecasts and observations at each time step. This method is then applied for all 24 members after downscaling the members of the global model. MORGREPS has been run in the Met Office operational suite since August 2005. Results from the verification period from March- May 2006 reveal slightly higher Brier Skill Scores for the regional ensemble system in contrast to the global ensemble and the ECMWF-EPS concerning wind speed and surface temperature.

On account of the increasing importance and success of LAMEPS the Central European Limited Area Ensemble Forecasting system ALADIN-LAEF was developed at ZAMG (ZentralAnstalt für Meteorologie und Geodynamik). This project was realized in the frame of the international cooperation LACE (Limited Area modeling Central Europe) and the first version of LAEF became preoperational in March 2007. The ensemble system is based on the NWP model ALADIN (Aire Limitée Adaption dynamique Développement InterNational) that was developed by 13 European national weather services. The first version of LAEF (DOWN) is characterized by dynamical downscaling the global EPS from ECMWF to the ALADIN-LAEF domain. After a period of testing and research a newer version was launched in February 2009 which contained a different technique. It is called BBSM (Breeding- Blending Surface perturbation-Multiphysics). "The idea is using the ALADIN blending technique, a digital filter and spectral analysis method, to combine the large-scale uncertainty generated by singular vectors of the ECMWF model with the small scale uncertainty generated by breeding with ALADIN model" (Wang et al., 2011). The set of techniques and model characteristics will be described in detail in chapter 5. It should be noted that both versions of ALADIN-LAEF (BBSM and DOWN) were compared using a verification period of 2 months (20.June- 20.August 2007). Comprehensive EPS verification scores considering statistical reliability, resolution and discrimination have been calculated. The performance of BBSM measured by the quotient spread/RMSE showed clear predominance compared to DOWN concerning upper air weather variables. There were also remarkable benefits achieved for the surface weather variables in respect to skill and resolution. However for surface weather variables, such as 2m temperature and 10m wind forecast, there remains under-dispersion. It is assumed this might result from the lack of quantification of uncertainties related to the model surface physics. As a next step the performance of ALADIN-LAEF in contrast to a global ensemble system had to be investigated. As mentioned above, LAEF was developed in an effort to better predict small scale weather phenomena such as flash floods in contrast to global ensemble prediction models. Therefore a verification project was started comparing ALADIN-LAEF (BBSM) to the global ECMWF-EPS. The main purpose of this investigation is to filter and locate the added value from LAMEPS ALADIN-LAEF in contrast to ECMWF-EPS. This would give additional justification to further develop ALADIN-LAEF. In regard to an objective investigation, it is important to know about the characteristics of a numerical model and the corresponding EPS. For this reason I will start with an introduction about general model characteristics and developments.

4 Numerical weather prediction models (NWP)

4.1 History of NWP development

Numerical weather prediction models form the fundament of today's weather prediction systems. They are based on a long history of innovative developments, starting in 1904 as a result of the thesis of the Norwegian scientist Vilhelm Bjerknes. He suggested that the weather could be predicted mathematically by a set of hydrodynamic and thermodynamic equations. This thesis was further developed and tested by Lewis F. Richardson who employed scientists to do hand-calculations. Lynch (2006) described Richardson's famous experiment as follows: "Richardson constructed a systematic mathematical method for predicting the weather and demonstrated its application by carrying out trial forecasts. His dream was to speed up the calculations but back then the technical resources were not available. History has shown that his innovative ideas were fundamentally sound: the methodology proposed by him is essentially the same used in practical weather forecasting today. However, the method devised by Richardson was utterly impractical at the time of his publications and the results of his trial forecasts appeared to be little short of outlandish. As a result, his ideas were eclipsed for decades, though there remained his dream." Richardson himself wrote about this dream:

"Perhaps some day in the dim future it will be possible to advance the computations faster than the weather advances and at a cost less than the saving to mankind due to the information gained. But that is a dream"

(Weather Prediction by Numerical Process, Lewis Fry Richardson, 1922)

According to Peter Lynch, an Irish Professor of Meteorology, who collected all the information about the history of NWP-models, the true significance of Richardson's work is evident (Lynch, 2006). With the technical evolution during the 20th century and the invention of computers as well as the expansion of meteorological observing stations, a rapid progress was recorded. In 1950 Charney managed the first successful numerical integration of the barotropic vorticity equation. On the basis of the so called "primitive equations" further developments on the numerical models were done, though it soon turned out that the nature of the problem was much more complicated than envisaged.

Encouraged by the success of the research in the United States and experience with short range and climatological simulations the ECMWF was established in 1971. In October 1973 the ECMWF convention was signed by nineteen European states including Austria. During the last years the community has been enlarged. At present (May 2010) the community includes the following member states: Belgium, Denmark, Germany, Greece, Spain, France, Ireland, Italy, Luxembourg, the Netherlands, Norway, Austria, Portugal, Switzerland, Finland, Sweden, Turkey, United Kingdom. Beyond that co-operation agreements with Czech Republic, Montenegro, Estonia, Croatia, Iceland, Latvia, Lithuania, Hungary, Morocco, Romania, Serbia, Slovenia and Slovakia

have been assigned.

The fundamental ambitions of the ECMWF are formulated as follows:

- To develop dynamic models of the atmosphere with a view to preparing medium-range weather forecasts by means of numerical methods
- To prepare, on regular basis, the data necessary for the production of medium-range weather forecasts
- To carry out scientific and technical research directed towards the improvement of these forecasts
- To collect and store appropriate meteorological data
- To make available to the meteorological offices of the Member States, in the most appropriate form, the results of the studies and research provided for in the first and third objectives above and the data referred to in the second and fourth objectives
- To make available sufficient proportion of its computing capacity to the meteorological offices of the member States for their research, priority being given to the field of numerical forecasting. The allocation of the properties would be determined by Council
- To assist in implementing the programs of the World Meteorological Organization
- To assist in advanced training for the scientific staff of the meteorological offices of the Member states in the field of numerical weather forecasting. (Persson and Grazzini, 2005)

The first operational forecast was launched on 1 August 1979. Since then the model underwent a broad variety of changes and further increments. Basically it consists of five components: a general circulation model, an ocean wave model, a data assimilation scheme and since 1992 an ensemble forecast system. Moreover a seasonal forecasting system was established 1998 and 4 years later a monthly forecasting system was installed. For the purpose of this investigation, the characteristics of the models operational in 2007 - the verification period - should be described precisely, since the ECMWF NWP model has formed the basis for all further model developments in Europe and since it supports the ECMWF-EPS which incorporates lots of its facilities.

4.2 The ECMWF global atmospheric model

The ECMWF global atmospheric model is based on a general circulation model and includes a dynamical component, a physical component and a coupled ocean wave component. The version which was operational in the period of investigation is denoted T_L799L91 (Persson and Grazzini, 2005). The model formulation for wind and temperature is based on a spherical harmonic representation which is triangularly truncated at total wave number 799 corresponding to a grid length of about 25 km. For the description of additional variables and model physics a reduced Gaussian grid is used. That allows a constant east-west separation between model points and prevents numerical instabilities which could arise due to rapidly decreasing distances near the poles. For the vertical representation the atmosphere is divided into 91 vertical layers up to 0.01 hPa (about 80 km), whereas the vertical resolution is finest in the planetary boundary (see fig. 4.1). As a result there are as many levels in the lowest 1.5km of the model atmosphere as in the highest 45km.

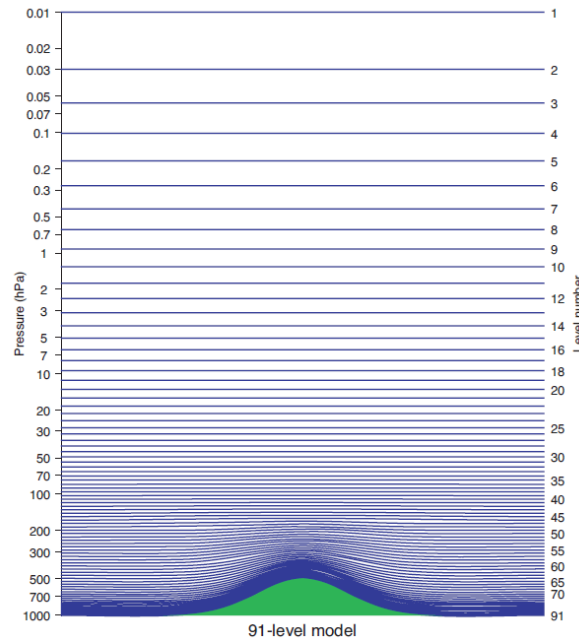


Figure 4.1: Distribution of the vertical levels in the TL799L91 model (Persson and Grazzini, 2005)

The ECMWF produces routine global analyses for the four main synoptic hours 00, 06, 12 and 18 UTC and global 10-day forecasts based on the 00 and 12 UTC data. The model framework can be summarized by six basic physical equations: the gas law, the hydrostatic equation, the equation of continuity, the equation of motion, the thermodynamic equation and the conservation of moisture. For the numerical formulation a semi-Lagrangian numerical scheme is being used. In principle the Lagrangian approach is given by: $du/dt = 0$. This means the value of a particle will not change in time, only the position of each particle has to be located and as a further consequence the trajectories. At every time step the grid-points of the numerical mesh are representing the arrival points of backward trajectories. During the transport the particle encounters various physical and dynamical forcing. Finally all prognostic variables are found through interpolation. In contrast to the Eulerian framework the semi-

Lagrangian scheme allows the use of large time steps without limiting the stability. For the numerical integration of the NWP model a temporal resolution of 12 minutes is being used. To account for the land topography, a model-topography is being employed. This model-topography is based on a data set with 1 km resolution that contains the mean elevation over the mean sea level. In addition a land-sea mask that contains the relative percentage of land and water ranging from values between 0 (100% sea) to 1 (100% land) is being considered. Physical processes like radiation, turbulence, friction and the formation of clouds follow physical equations, though due to their small scales they have to be described in a statistical way as parameterization processes (see fig. 4.2).

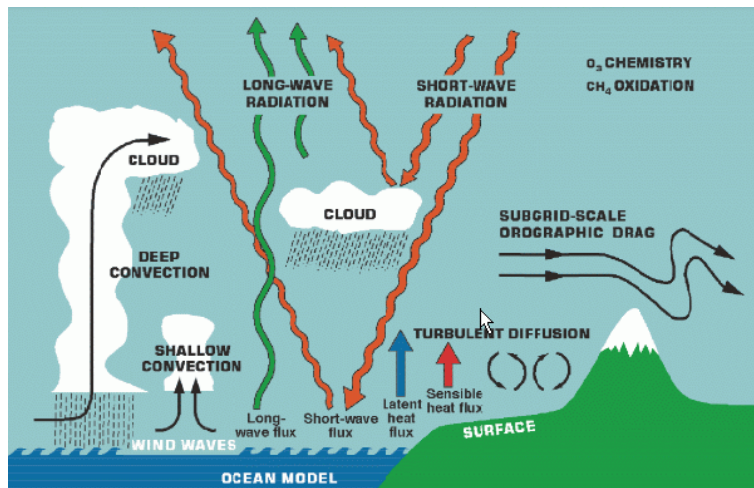


Figure 4.2: Main physics represented in the ECMWF model

Source: Persson and Grazzini, User guide to ECMWF forecast products, 2005

To account for radiation processes the spectrum is divided into a long wave part (thermal) and a short wave part (solar), whereas the thermal part includes 16 bands and the solar part 6 bands. These radiation processes are mainly affected by pressure, temperature, moisture, cloud cover and cloud water content. The radiation scheme is only called every 3 hours to save computational costs. In addition, there is also a convection scheme computing cloud production as well as vertical transports of moisture and momentum. Moreover it is able to distinguish deep, shallow and mid-level convection. "Although the physics computations are performed only in the vertical, the complexity of processes and feedback mechanisms between various processes makes the computations complex and expensive. While the dynamics as such only occupy 23% of the computational time, the physical processes (including radiation) account for 36% and the ocean wave model alone 10%. The remaining 31% are spent on communications, numerical transformations and spectral computations" (Persson and Grazzini, 2005).

The analysis of the atmospheric conditions is based on a data assimilation scheme. In consequence the quality of the estimate of the atmospheric conditions depends on the skill of the assimilation scheme. In a typical 12-hour period, there is a total of 75 million pieces of data available, whereas 98% result from satellite. Generally there are two types of measurements: in situ observations and remote sensing. Apparently various types of data have different characteristics in terms of geographical coverage, vertical structure and temporal distribution. Conventional data result from SYNOP codes, radiosonde measurements, wind profilers and similar surface instruments. During the last decades

there has been a significant increase in the quantity, quality and diversity of satellite observations (see fig. 4.3). Although satellite data cannot replace conventional observations completely, they provide several clear advantages such as high temporal and spatial coverage. This broad spatial coverage of satellite data ensures that the volatile small amplitude-large scale errors over the oceans are corrected for, something which isolated measurements would have difficulties to do. Consequently there is a strong benefit from satellite data in the ECMWF, especially over the Southern Hemisphere where there is lack of conventional data. However there are limitations of the use of satellite data since the measurements are bounded to certain criteria such as specific wavelengths. Before being implemented in a numerical model, quality controls as well as redundancy checks have to be done. The quality control includes a thinning procedure, where data is removed that is either redundant or has highly correlated errors. At the ECMWF a sophisticated assimilation scheme has been developed and a consequent monitoring of the data platforms is one of its main characteristics. Platforms which are found to report biased or erratic observations are put on a so called blacklist. During the quality control the number of observations is reduced significantly, whereas the satellite data is mainly affected. In particular only around 5% of the radiance observations are being used. However this small fraction in absolute terms is still ten times the total amount of all other types of observations.

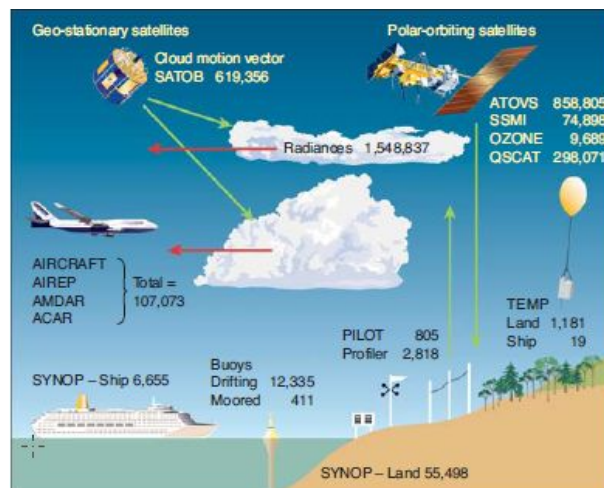


Figure 4.3: Summary of observations received at the ECMWF, 5 July 2004

Source: Persson and Grazzini, User guide to ECMWF forecast products, 2005

4.3 The limited area model ALADIN-AUSTRIA

ALADIN-AUSTRIA is a limited area numerical weather prediction model that was operationally implemented at ZAMG 2003, producing forecasts up to 48 hours (Wang et al., 2006). It has been developed on the basis of the limited area NWP-model ALADIN (Aire Limitée Adaption dynamique Développement International) which was initialized by Meteo France. Within an international project through RC LACE (Regional Cooperation for Limited Area Modeling in Central Europe) 13 European national weather services including ZAMG joined the further development of ALADIN. With the technical development at ZAMG (high performance computer SIG-3400) it became possible to run an own limited area model operationally. As a result ALADIN-VIENNA was started on 22 September 1998, generating output twice per day. The

model domain was nested in ALADIN-LACE (including Central Europe) and covered Austria and parts of Central Europe (see fig. 4.5). The boundary conditions were provided by the global model ARPEGE/IFS (Action de Recherche Petite Echelle Grande Echelle/ Integrated Forecasting System, Météo France) in a 3-hourly interval. For the vertical discretization pressure-based hybrid coordinates were being implemented. In respect to the horizontal representation the model was formulated in a spectral way which is a consequence of the general agreement among the ALADIN international co-operation. At the initialization of the project ALADIN it was decided that ALADIN and modified versions should be developed as the LAM counterpart of the global spectral model ARPEGE/IFS in order to share the development and progress. It should be noted that spectral discretization can cause problems for limited area models because the periodicity is missing and has to be created artificially. With the introduction of the newer version ALADIN-AUSTRIA there were several refinements done. The vertical resolution was increased from 37 levels (ALADIN LACE as well as ALADIN VIENNA) to 45 levels, whereas most additional levels were added in the lower atmosphere. In ALADIN-AUSTRIA the horizontal resolution remained 9.6 km like in ALADIN-VIENNA, because experience at ZAMG revealed that increasing resolution does not necessarily improve the forecast. According to the results it appeared better to stick to the Lindzen and Fox-Rabinovitz consistency criterion. This criterion suggests that there must be a reasonable proportion between horizontal and vertical resolution. In regard to the numerical formulation the ALADIN-AUSTRIA is built up on the following equations: momentum-, hydrostatic -, continuity, thermodynamic and moisture equations. For the time-integration a semi-Lagrangian semi-implicit (SLSI) scheme is being employed that maintains stability through a semi-implicit correction. With the use of this correction term it is possible to use larger time steps than the Courant-Levy-criteria would permit. As a consequence of its spectral formulation the model domain has to be split into 3 different regions: zone C, zone I and zone E. Zone C is the area of meteorological interest in the centre of the model domain. Zone I denotes the region where the boundaries are adapted from ARPEGE. The outermost region is zone E which exists for computational reasons, e.g. interpolation processes to create an overall periodical field. In consequence noises that develop at the boundary conditions due to poor conformance between the global and the regional model have to be suppressed when entering region C. This is done by using a Digital Filter Initialization (DFI). With this technique the high frequency noise can be damped. For ALADIN-AUSTRIA the physical parameterizations are constructed in a way that each model column is treated independently. There, exchanges are only possible through dynamics or horizontal diffusion. One part of this parameterization is the radiation scheme that has one spectral interval in the solar range and one in the thermal range. It is able to distinguish liquid and ice phase effects in clouds as well as all active gases. To account for the turbulent transports of momentum, heat and water vapour a first-order turbulence closure from Louis (1979) is employed. For physical parameterization there is also a cloudiness scheme included, whereas cloudiness and cloud water content are determined simply as a function of humidity and temperature. Besides several other parameterizations there is also the deep convection scheme that is based on a mass-flux-type-scheme of Bougeault (1985), though it has been refined concerning several aspects. Differences in contrast to the ECMWF global model are not only found concerning resolution and topography, but also concerning physical parameterizations such as precipitation processes or the treatment of soil.

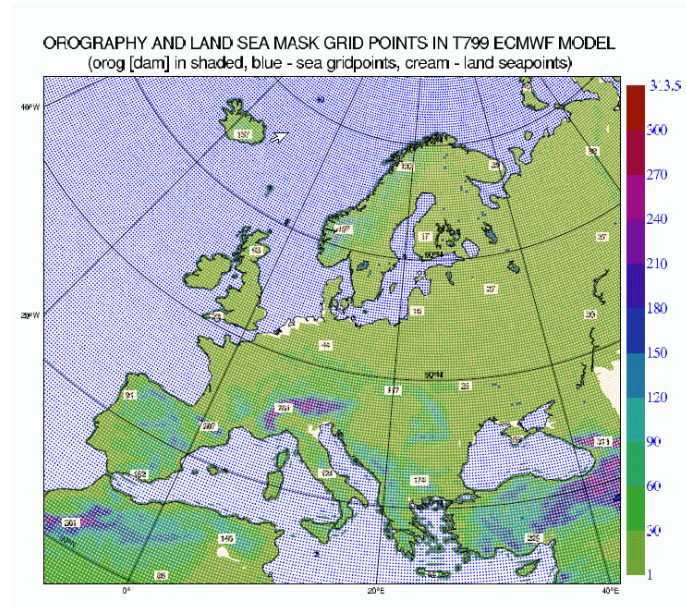


Figure 4.4: Orography and land-sea mask of the ECMWF general circulation model TL799L91; Source: Persson and Grazzini, User guide to ECMWF forecast products, 2005

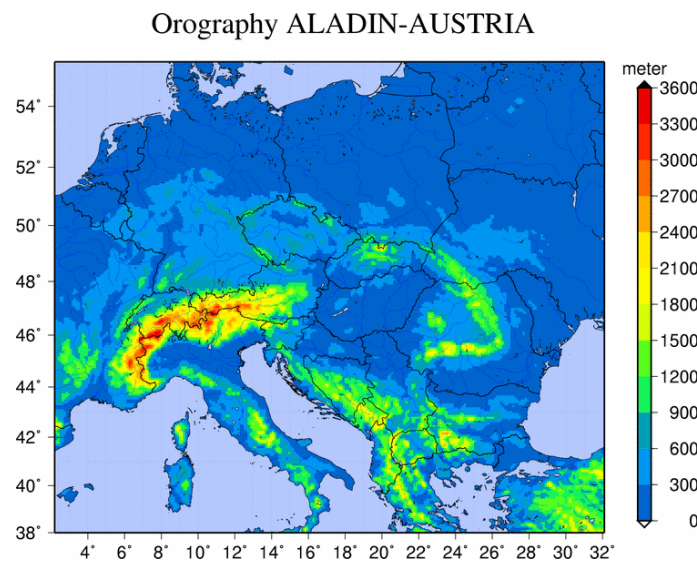


Figure 4.5: Domain and topography of ALADIN-AUSTRIA
Source: Wang et al., The operational Limited Area Modeling system at ZAMG: ALADIN-AUSTRIA, 2006

5 EPS

5.1 Basic Principles and Sources of uncertainty

The numerical formulation of the partial differential equations that determine the atmospheric dynamics is based upon a broad range of approximations. Beyond that the atmospheric conditions implemented in a numerical model can only be estimated within certain accuracy. Generally there exist four sources of uncertainty that have to be taken into consideration:

1. Uncertainties due to observational errors and data assimilation method, e.g. incomplete data coverage, measurement errors
2. Uncertainties due to errors in boundary conditions, e.g. as a consequence of coupling the LAMEPS with a global counterpart
3. Uncertainties resulting from imperfect description of surface conditions and physical parameterizations, e.g. the effect of unresolved scales
4. Uncertainties as a consequence of the approximations in the model formulation

Since the atmosphere can never be completely observed, either in terms of spatial coverage or accuracy of measurements, a fluid-dynamical weather prediction model will always begin calculating forecasts from a state at least slightly different from that of the real atmosphere. On this account the correct estimation of the level of uncertainty is the main challenge when dealing with numerical models. Consider the fact that the initial state is uncertain but can be described by a probability density function $p(x)$. According to Leutbecher and Palmer (Leutbecher and Palmer, 2008) the evolution of the p.d.f. $p(x)$ under the dynamics of the forecast model can be described by Liouville's equation. This is a linear partial differential equation defined on a subset of \mathbb{R}^n . Its numerical solution is impractical even for dimensions N as low as 100. Yet, current numerical weather prediction models have phase spaces of dimension N of 10^6 - 10^8 . Therefore computational problems make the use of these equations unfeasible for numerical weather prediction. The only feasible technique to obtain estimates of the nonlinearly evolved p.d.f. are Monte-Carlo techniques that sample the p.d.f. at initial time and evolve the sampled initial states with the forecast model - or a perturbed version of the forecast model to account for model uncertainty. The sample is usually referred to as ensemble and individual elements as ensemble members. This approach was introduced already in the 1960's by Lorenz. In 1969, Epstein presented a theoretical approach which he denoted stochastic dynamic prediction, whereas atmospheric motions were governed by physical laws and the evolution of the atmosphere was considered to be deterministic (Epstein, 1969). The uncertainty about the initial state of the atmosphere, from which a simplified NWP model was initialized, was conceived of as a probability distribution in the phase space. In Epstein's concept, a stochastic dynamic forecast moved the probability distribution of the initial state through the phase space as the forecast advanced in time, according to the laws of fluid dynamics represented in the NWP model equations. As a consequence, the shape of the initial distribution

was stretched and distorted as the forecast advanced. It also became more dispersed at longer forecast projections, reflecting the increased uncertainty of forecasts further into the future (see fig. 5.1). Equations describing the evolution of the initial-condition probability distribution could be derived through introduction of a continuity, or conservation equation for probability (Ehrendorfer, 1994). However, the dimensionality of phase spaces for problems of practical forecasting was too large to allow direct solution of these equations.

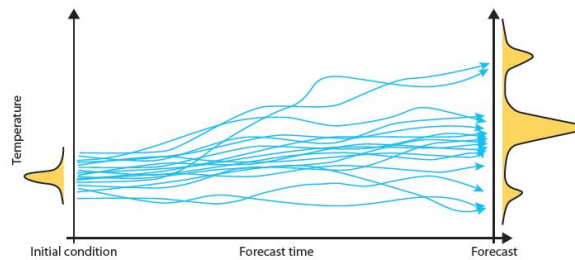


Figure 5.1: The basic principle of ensemble-based probabilistic forecasting

Source: <http://www.ecmwf.int/about/information-leaflets/EPS.pdf/status>
18.6.2010

Therefore an alternative approach had to be found. In 1974 Leith proposed a practical solution to these stochastic dynamic equations, using Monte-Carlo techniques (Wilks, 2006). The basic idea is to collect a finite sample from the probability distribution describing the uncertainty of the initial state of the atmosphere. These few members are picked randomly and located around the mean estimated atmospheric state in the phase space. Collectively, these points are called ensemble of the initial conditions, and each represents a plausible initial state of the atmosphere consistent with the uncertainties in observation and analysis. Rather than explicitly predicting the movement of the entire initial-state probability distribution through phase space, the movement is approximated by the collective trajectories of the ensemble of sampled initial points. In other words each point in the initial ensemble provides the initial conditions for a separate run of the NWP model. At this initial time, all the ensemble members are very similar to each other. After the integration the distribution of the ensemble shows the range of possible prospective atmospheric conditions, reflecting the involved uncertainties (see fig. 5.2).

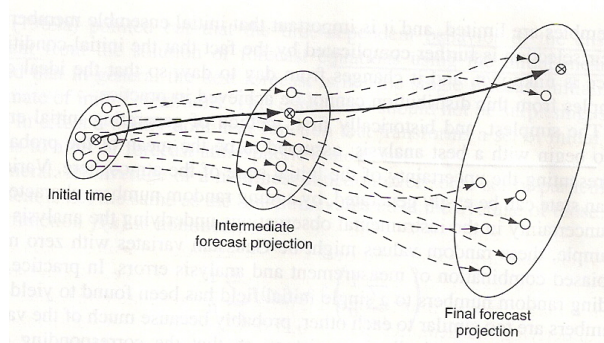


Figure 5.2: Schematic Illustration of concepts in ensemble forecasting

Source: Wilks, Statistical methods in the Atmospheric Sciences, 2006

Although this range of uncertainty is very valuable for a correct estimation of the future conditions, the information finally has to be condensed. A very simple approach is to calculate the ensemble mean, whereas elements of disagreements are being filtered. In 1969 Epstein pointed out that the time-dependent behavior of the ensemble mean is different to the solution of a single forecast initialized with the best estimate of the initial state. However condensing information by calculating the ensemble mean will only be valuable until there is a change in long-wave pattern (Palmer, 1993). If a regime changes, it is very likely that the ensemble members develop completely different trajectories and the ensemble average will no longer present the best approximation.

Although information finally has to be filtered, the striking characteristic of ensemble prediction systems is the capacity to yield information about the magnitude and nature of uncertainty in a forecast. Looking at an ensemble prediction system it seems intuitively clear, that an ensemble mean will be close to the eventual atmospheric state if the dispersion of the ensemble is small. In contrast the atmospheric state appears to be very uncertain, if the individual ensemble members are widely distributed. These methods are based on intuition, but for the correct use of this level of uncertainty concrete methods have to be applied. In meteorological offices, the forecasters consider a range of uncertainty by using different sources of information. A probably more accurate method would be to forecast "forecast skill" in terms of ensemble spread and accuracy (Ehrendorfer and Tribbia, 1997). This method would lead a better perception of the forecast. If the magnitude of the ensemble spread (standard deviation) corresponds to the magnitude of accuracy of the ensemble mean (root mean square error), the model is considered to account for the involved uncertainties. It should be noted that the level of uncertainty can vary from situation to situation. Therefore a large number of ensemble members can be very useful to get a reliable approximation. Since the computer time is a limiting factor at operational forecast centres, the appropriate selection of ensemble members and spatial resolution has to be balanced. For this reason the number of ensemble members is limited. Consequently the main challenge in ensemble prediction systems is the generation of adequate initial perturbations within a reasonable time range.

According to Wilks (2006) the simplest and historically first method of generating initial ensemble members is to begin with a best analysis, assumed to be the mean of the probability distribution representing the uncertainty of the initial state in the atmosphere. Variations around this mean state can be easily generated, by adding random numbers characteristic of the errors or uncertainty in the instrumental observations underlying the analysis. For example, these random values might be Gaussian variates with zero mean, implying an unbiased combination of measurement and analysis errors. In practice, however, simply adding random numbers to a single initial field has been found to yield ensembles whose members are too similar to each other, probably because much of the variation introduced in this way was dynamically inconsistent, so that the corresponding energy is quickly dissipated in the model (Palmer et al., 1990). For this reason the variability of the resulting forecast ensemble underestimates the uncertainty in the forecast. Based on this notion several methods to generate appropriate initial conditions have been developed at different national weather prediction centres.

5.2 A comparison of global ensemble prediction systems

This section gives a brief comparison of the different global ensemble prediction systems developed at the 3 different meteorological services namely The European Centre

for Medium-Range Weather Forecasts (ECMWF), the Meteorological Service of Canada (MSC) and the National Centers for Environmental Prediction (NCEP). At the NCEP the so called breeding method has been developed. This method, discussed later in more detail, is based on the notion that fast-growing perturbations develop naturally in a data assimilation cycle as well as in the model integration process. So at the end of the integration the results will include the dominating errors and therefore represent a perturbation to the real state of the atmosphere. The main advantage about breeding is its rather easy application, since there are hardly any additional computations necessary.

The method used at the ECMWF is completely different and requires more computational effort. The intention is not only to consider error growth like it is done in the breeding method, but to identify the directions of maximum error growth and use them for generating perturbations. Therefore a singular vector approach is being used, which is able to mark the directions of maximum error growth over a fixed time interval. If the perturbations point along these directions, the maximum deviations to the control run can be obtained. A detailed description of this method will be given in next section. At MSC a completely different method has been developed for the generation of perturbations. The technique bases on the selection of randomly perturbed observations (Houtekamer et al., 1996). The initial perturbations are generated by assimilating randomly perturbed observations, using different model versions in a number of independent assimilation cycles. Since the analysis and forecast process is repeated several times with different random input, the perturbed-observation method (PO) is a classic example of the Monte Carlo approach. Finally the results are thought to account for uncertainties involved in the assimilation processes and model integration.

In 1995 Houtekamer and Derome compared these different strategies to generate initial perturbations using a simplified model (3level- quasigeostrophic). Additionally they created an artificial observational data set. Although the 3 strategies for generating appropriate perturbations differ considerably, the results turned out to be quite comparable. However it should be noted that the simplification of the model and the data net could have excluded important aspects. A further investigation of Buizza, Houtekamer and Toth (Buizza et al., 2005) revealed that the performance of the EPS strongly depends on the quality of the data assimilation system used to create unperturbed (best) initial conditions and the numerical models involved. For all three global systems, the spread of the ensemble forecasts turned out to be insufficient to capture reality, suggesting that none of them is able to simulate all sources of forecast uncertainty.

5.3 The ECMWF ensemble prediction system

5.3.1 Model characteristics

The first version of the ECMWF ensemble prediction system was introduced in December 1992 and contained 33 members in a T63L91 model configuration, where one control run and 32 perturbed members simulated the error growth during the first 48 hours forecast range (Buizza et al., 2005). Since the number of ensemble members and the choice of resolution depend on the computational capacity, the model resolution was upgraded in 1996 and the number of ensemble members extended to 50 thanks to improved technical resources. This newer version was denoted T_L159L31. In March 1998 a scheme accounting for initial uncertainties due to perturbations that had grown during the 48h prior to the starting time was introduced (evolved singular vectors, Barkmeijer et al. 1999). Another model upgrade came in October 1998. Then a scheme

which could simulate uncertainties due to random model error in the parameterized physical processes was added. Within the next few years the ensemble prediction system underwent further revisions. In October 1999 the horizontal resolution was increased and in November 2000 the number of vertical levels extended. The resulting version was denoted T_L255L40. In January 2002 the model was adapted and tropical initial perturbations were being implemented. (Barkmeijer et al., 2001). Based on further model revisions, regional characteristics due to their geographical location could be considered.

The model version valid for the period of investigation (2007) is denoted T_L399L62 (corresponding deterministic model: T_L799L91). The horizontal representation of 399 linear spectral components approximately corresponds to a horizontal resolution of 50 km. For the vertical representation, the atmosphere is divided into 62 layers between the surface and the 5hPa level. The key characteristic of the ensemble prediction system is the generation of perturbations. For the ECMWF-EPS the singular vector approach is being used. A detailed description of this technique will be given in the next section. For the numerical integration a semi-Lagrangian scheme has been implemented and the corresponding integration time step is 1800 seconds. The parameterization of sub grid-scale processes is based on stochastic physics. Physical processes on sub grid scale are represented as relatively simple functions of the explicitly resolved variables. Since these functions cannot fully capture the processes, one way to account for involved uncertainty is to perturb the functions randomly. Ensemble forecasts are generated twice per day, at 00 UTC and 12 UTC, producing forecasts valid for a time range of 10 days.

5.3.2 Singular vector method

To account for the "worst possible scenario" a method has been developed at the ECMWF that generates perturbations with maximum deviation to the control run. This method is based on a singular vector method and the main focus is to maximize the involved energy within a time range of 48 hours. The advantage of using singular vectors is that if the forecast error evolves linearly and the proper initial time norm is used, the resulting ensemble captures the largest amount of forecast error covariance at optimization time (Ehrendorfer and Tribbia, 1997). Therefore the involved perturbations have been designed to point along the directions of maximum error growth. Their growth rate can be defined by a metric called total energy norm and is limited by a time range that is usually referred to as optimization time interval. For the ECMWF an optimization interval of 48 hours has been implemented. According to Buizza (Predictability of Weather and Climate, 2006) the generation of perturbations can be described in the following way: Let X denote the state vector of the deterministic prediction system, whose simplified evolution equations can be formally written as:

$$\frac{\partial X}{\partial t} = A(X) \quad (5.1)$$

As a result its evolution can be obtained by integrating from an initial time to a final time. Additionally there will as well be an evolution of the related perturbations x , showing differences to the main trajectory. Their evolution can be described in a first approximation by the linearized model equations:

$$\frac{\partial x}{\partial t} = A_1 x \quad (5.2)$$

whereas A_1 is the tangent operator computed along the main trajectory $X(t)$. In consequence the perturbation trajectory $x(t)$ from the initial to the final time can be described

by:

$$x(t) = L(t, t_0)x(t_0) \quad (5.3)$$

The operator $L(t, t_0)$ can be referred to as tangent-linear propagator. In an effort to find the maximum deviation to the main trajectory, the direction of maximum error growth has to be found. Therefore the tangent-linear operator is factorized using the singular value decomposition:

$$x(t) = U\Sigma V^T x(t_0) \quad (5.4)$$

U denotes the matrix which contains the orthonormal eigenvectors of AA^T in its columns. Σ is a diagonal matrix which contains the singular values of A and V is the matrix which contains the orthonormal eigenvectors of A^T in its columns. Choosing the right singular vector $v_i(t)$ corresponding to the largest singular value we get the direction of maximum error growth. This vector is finally implemented for $x(t_0)$. It should be noted that the perturbations pointing along different axes of the system show different evolutions. Only those perturbations with singular values exceeding the critical value 1 that point along the axis of the related singular vector experience maximum growth. For the correct estimation of maximum error growth the energy norm is being used and as a result the equation for the final perturbations is slightly modified:

$$\|x(t)\| = x^T(t)C^T C x(t) = x^T(t_0)L^T C^T C L x(t_0) \quad (5.5)$$

The singular decomposition of the involved operators show, that the final form is linked to an eigenvalue problem. The square roots of these eigenvalues represented in the diagonal of the matrix Σ , are called the singular values. As already mentioned the singular vectors corresponding to the largest singular values mark the directions of maximum growth. As a result at optimization time t , the i^{th} singular vector evolves into:

$$v_i(t) = L(t, t_0)v_i(t_0) \quad (5.6)$$

a vector with total norm equal to

$$\|v_i(t)\| = \sigma_i \quad (5.7)$$

Finally any perturbation can be written as linear combination of the singular vectors $v_i(t)$ and as a consequence

$$\max \frac{\|x(t)\|}{\|x(t_0)\|} = \sigma_i \quad (5.8)$$

which implies that maximum growth as measured by the norm is associated with the dominant singular vectors. The descriptions given above describe the main idea to construct appropriate perturbations, but since the Matrix of the tangent-linear Operator is never fully available, the problem has to be solved iteratively. For the technical implementation at the ECMWF the Lanczos algorithm is being used.

Based on this approach an ensemble prediction system can be created by adding perturbations to the analysis. Formally each member e_j can be written as time integration (Palmer and Hagedorn, 2006):

$$e_j(d, t) = e_j(d, 0) + \int_{t_0}^t (A(e_j, t) + P_j'(e_j, t)) dt \quad (5.9)$$

of the perturbed model equations:

$$\frac{de_j}{dt} = A(e_j, t) + P_j'(e_j, t) \quad (5.10)$$

Here j denotes the individual members, d the initial day and t the time of forecast. A identifies the analysis at each time step and P' denotes the contribution generated by a parameterized physical processes (stochastic perturbation). For each grid point

$$r = (\lambda, \phi, \sigma) \quad (5.11)$$

identified by its latitude, longitude and vertical hybrid coordinate, the perturbed parameterized tendency of each state vector component is defined as

$$P'_j(e_j, t) = [1 + \langle s_j(\lambda, \phi, t) \rangle_{D,T}] P(e_j, t) \quad (5.12)$$

There P is the unperturbed diabatic tendency and $\langle \dots \rangle_{D,T}$ indicates that the same random number s_j has been used for all grid points inside a $D \times D$ degree box and over T time steps (Buizza et al., 1999). The introduction of space and time coherence in the stochastic perturbations was based on the assumption that organized systems have some intrinsic space and time scales that may span more than one model time step and more than one model grid point. Making the stochastic uncertainty proportional to the tendency was based on the concept that organization (away from the notion of a quasi-equilibrium ensemble of sub grid processes) is likely to be stronger, as the parameterized contribution becomes stronger. For starting the forecast (see eq.5.9) the initial members $e_j(d,0)$ of the EPS are generated as follows:

$$e_j(d, 0) = e_0(d, 0) + \delta e_j(d, 0) \quad (5.13)$$

e_0 denotes the unperturbed initial condition for the forecast day d and forecast time 0 . where each initial perturbation is generated using singular vectors computed to maximize the total energy norm over a 48-hour time interval and scaled to have an amplitude comparable to analysis error estimates. Singular vectors are usually located in regions of strong barotropic and baroclinic activity: at initial time, they have most of their energy confined in the small scale and are confined vertically in the lower troposphere. During the optimization time interval, they change shape and grow in scale, and vertically propagate upward. This propagation is linked to the conversion of initial-time potential energy into final-time kinetic energy (Buizza and Palmer, 1995).

5.4 The Central European limited area forecasting system: ALADIN - LAEF

5.4.1 Model characteristics

ALADIN-LAEF (Aire Limitée Adaption Dynamique Développement International - Limited Area Forecasting) is a limited area ensemble prediction system that was developed at ZAMG to better predict high impact weather. Therefore the application of the high resolution limited area model ALADIN is one of the key elements in this system. The horizontal spectral resolution corresponds to approximately 18 km and the vertical representation is based on 37 levels. The model domain includes Europe, parts of Northern Africa, Western Russia as well as parts of the Atlantic Ocean (see fig. 5.3). ALADIN-LAEF includes 16 ensemble members and generates forecasts up to a forecast lead time of 60 hours. In the first version of ALADIN-LAEF, the initial condition (IC) perturbations and the lateral boundary conditions were generated by dynamically downscaling the first 16 ECMWF-EPS members. Model uncertainties and surface uncertainties were not taken into account in this version (Wang et al., 2011). In an effort to better incorporate small scale perturbations, a new approach was designed. It combined the large scale perturbations provided by the ECMWF singular vectors and the small scale perturbations generated through a breeding method. The resulting version was denoted BBSM (Breeding-Blending-Surface perturbation-Multiphysics) and is based on 3 key characteristics:

- a blending technique for generating initial condition perturbations
- a multi-physics scheme
- a technique to construct suitable initial surface conditions

In ALADIN-LAEF the generation of IC perturbation includes **2 independent mechanisms**:

1. the generation of IC upper level perturbations
2. the generation of IC surface perturbations which is called NCSB (Non-Cycling Surface Breeding)

In NCSB different atmospheric conditions are combined with a joint surface analysis. When these different settings are implemented into the model and the model is integrated forwards, different surface conditions are obtained. These surface conditions account for a range of uncertainties due to imperfect surface observations. However the generation of appropriate IC perturbations for both the surface field and the upper level field is based on the breeding method.

5.4.2 The breeding method

The breeding method was first implemented at NCEP in 1992 for generating perturbations for the global EPS. The major advantage about breeding is the fact that is simpler and at less computational costs than the calculation of singular vectors. According to Toth and Kalnay (Toth and Kalnay, 1997) the breeding method simulates the development of growing errors in the analysis cycle. A difference field between two nonlinear forecasts is carried forward and scaled down at regular intervals upon the evolving

LAEF Domain & Topography

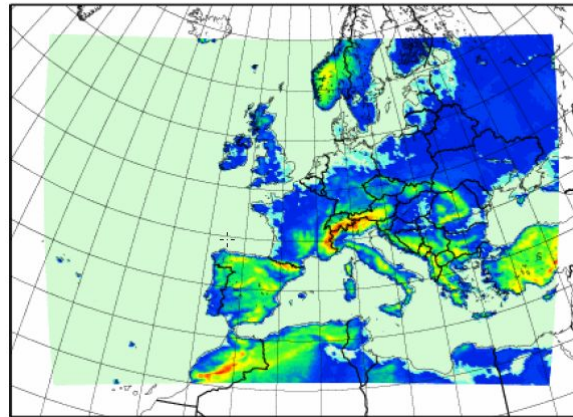


Figure 5.3: ALADIN-LAEF-domain, source: ZAMG

atmospheric fields. By construction, the bred vectors are superpositions of the leading local (time-dependent) Lyapunov vectors (LLVs) of the atmosphere. An important property is that all random perturbations take over the structure of the leading LLVs after a transient period, which for large-scale atmospheric processes is about 3 days. The importance of this property of LLVs in meteorology was first recognized by Lorenz (Lorenz, 1965). Because of nonlinear interactions and the existence of many regional features, different breeding cycles do not converge to a single leading LLV, but rather span the subspace of the fastest growing perturbations that can occur at the chosen level of perturbations amplitudes. When several independent breeding cycles are performed, the phases and amplitudes of individual leading LLVs are random, which ensures quasi-orthogonality among the global bred vectors from independent breeding cycles.

The general breeding method can be divided into **5 steps**:

1. starting with the introduction of an arbitrary perturbation on to control analysis, this should be done only once
2. integrating the model forward starting from the control analysis and the perturbed initial conditions
3. building the difference between the perturbed forecast and the control analysis at a fixed time interval (see fig. 5.4)
4. scaling down the forecast difference in amplitude to the size of the perturbation
5. adding/subtracting the rescaled difference to the new control analysis
6. the steps 2-4 are then repeated and perturbations being bred that grow along the forecast trajectory

In ALADIN-LAEF the breeding method starts with the implementation of a selected number of pairs of perturbations. These positive and negative perturbations are placed around the control analysis. Then the model is integrated forward, whereas the boundary conditions are provided by the ECMWF-ensemble prediction system. After the

breeding process the perturbations have to be rescaled in order to represent a physically relevant size. The rescaling constant S depends on the difference between the involved perturbations and can also be directed by a tuning constant C . The rescaling constant S is defined by the tuning constant C and the average difference between a set of forecast pairs ΔP .

$$S = \frac{C}{\Delta P} \quad (5.14)$$

Consider the temperature forecasts for the 850hPa-level. At each model grid point, the difference between the positive and negative short-term temperature forecasts is calculated (Wang et al., 2011). The resulting average difference can finally be modified by the tuning constant.

$$\Delta P = \sqrt{\sum_1^N \frac{T(850)^P - T(850)^n}{N}} \quad (5.15)$$

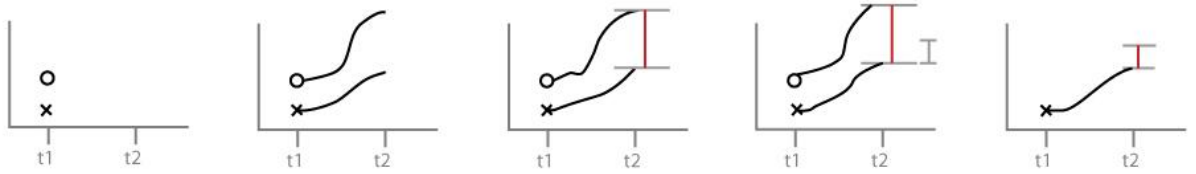


Figure 5.4: Visualization of a simplified breeding mechanism, source: Karin Schmeisser

5.4.3 The blending mechanism

The rationale behind blending is the fact that the singular vector method has shown advantages in medium-time range forecasts, whereas the breeding method has revealed better results for small scale and short-time range forecasts (Buizza et al., 2005). Basically breeding attempts to give the best estimate of the actual errors in the initial analysis based on past information of the flow, whereas the singular vectors of the global model contain future information of possible forecast error (Toth and Kalnay, 1997). In an effort to **include both advantages** a blending technique was developed by Brozkova et al. (2001). The principal part of this method is the implementation of a digital filter that enables the filtering of selected frequencies (small scale noise). Therefore a standard Dolph-Chebyshev digital filter is being used. A detailed description of digital filter technique for NWP can be found in Lynch and Huang (1992). In regard to ALADIN-LAEF this filter has been applied on both the ECMWF singular vectors and the breeding vectors on the original ALADIN grid but at a lower spectral resolution. The difference between those filtered files is then transferred to the ALADIN original spectral resolution and finally added to the ALADIN breeding vector. A symbolic equation representing the blending mechanism is given in Radmila et al. (2006):

$$ICblend(m) = Abred(m) + ((ASV(m)^{DF})_{LOW} - (Abred(m)^{DF})_{LOW})_{HIGH} \quad (5.16)$$

m represents the individual ensemble members and $ICblend$ denotes the IC perturbations after blending. ASV stands for the perturbed analysis generated by ECMWF singular vectors and $Abred$ represents the breeding vectors. LOW indicates the lower spectral resolution (cut-off truncation) which is being used in the digital filter technique. The index $HIGH$ refers to the ALADIN original resolution.

5.4.4 Physical parameterization - Multi physics

The numerical formulation of the partial differential equations that govern the dynamics of the atmosphere implies a range of approximations. Some of these approximations are due to the sub-grid scale characteristic of physical processes such as soil or radiation processes. Popular methods for representing those model uncertainties are stochastic physics where physical tendencies are randomly perturbed or different parameterizations for the individual ensemble members. In ALADIN-LAEF different physics configurations as well as different variations of certain parameterizations are being implemented.

Here is an overview the physical parameterizations used in ALADIN-LAEF:

The left column represents the 16 individual perturbed ensemble members, a control member M0 and the deterministic forecast M99 (see left column in fig. 5.5). In the second vertical column the different configurations can be seen. These configurations include different types of parameterization schemes which address cloud physics, deep convection or radiation.

mem #	configuration	Cloud-physics	deep convection	radiation	turbulent transport	shallow convection	mixing length & entrainment rate
M 1	ALADIN-25	Kessler	BGMC	RG	Louis81	JFG03	Setting_0
M 2	ALADIN-25	Kessler	BGCP	RG	Louis81	JFG03	Setting_1
M 3	HARMONIE	Sunquist	STRACO	Savijarvi90	CBR+S90	JFG03	---
M 4	ALARO+3MT	Alaro	3MT	JFG05	JFG06	JFG03	---
M 5	ALADIN-32	Lopez	BGMC	ECMWF	Louis81	KFB	Setting_0
M 6	ALADIN-32	Lopez	BGCP	ECMWF	Louis81	KFB	Setting_1
M 7	ALARO	Alaro	BG_MCON	JFG05	JFG06	JFG03	---
M 8	ALARO	Alaro	BG_MCON	JFG05	JFG06	JFG03	---
M 9	ALADIN-32	Lopez	BG_MCON	ECMWF	CBR+B81	KFB	Setting_0
M 10	ALADIN-32	Lopez	BG_CAPE	ECMWF	CBR+B81	KFB	Setting_1
M 11	ALADIN-32	Lopez	BG_MCON	ECMWF	CBR+S90	KFB	Setting_0
M 12	ALADIN-32	Lopez	BG_CAPE	ECMWF	CBR+S90	KFB	Setting_1
M 13	ALADIN-32	Lopez	BG_MCON	ECMWF	CBR+S90	JFG03	Setting_0
M 14	ALADIN-32	Lopez	BG_CAPE	ECMWF	CBR+S90	JFG03	Setting_1
M 15	ALARO+3MT	Alaro+XR	3MT	JFG05	JFG06	JFG03	---
M 16	ALARO+3MT	Alaro+XR1	3MT	JFG05	JFG06	JFG03	---
M 0	ALARO	Alaro	BG_MCON	JFG05	JFG06	JFG03	---
M 99	ALADIN-32	Lopez	BG_MCON	ECMWF	Louis81	KFB	Setting_0

Figure 5.5: The physical settings used in ALADIN-LAEF

source: Wang et al., The Central European Limited Area Ensemble Forecasting System ALADIN-LAEF, 2011

The following brief introduction is based on the general information given in "The central European limited area forecasting system: ALADIN-LAEF" (Wang et al., 2011). ALADIN-LAEF includes the following **5 main physical configurations**: ALADIN-25, HARMONIE, ALARO+3MT, ALADIN-32, and ALARO. Note that there exist variations of each main setting.

1. The physical setting **ALADIN-25** has been attached to the 1st and 2nd member of ALADIN-LAEF (see 1. and 2. row in table 5.5). ALADIN-25 represents the basic physical settings of ALADIN. Although for both members the main features are the same, there exist variations such as different settings for mixing length. Therefore the two ensemble members have a different fine tuning. ALADIN-25 includes a diagnostic scheme for the representation of cloudiness and large-scale

precipitation. The resolved cloudiness and cloud water content are determined as a function of humidity and temperature. The precipitation flux is computed from the condensation rates with the assumption that any supersaturation is converted to precipitation instantaneously. Evaporation, melting and freezing of precipitation are taken into account by applying a revised Kessler (1969) scheme. Moreover there is a deep convection scheme included. It has been based on the mass-flux-type scheme of Bougeault (1985) and contains several refinements and modifications that can be found in Geleyn (2003). Apart from that a radiation scheme developed by Geleyn and Ritter (1992) has been implemented. The computation of turbulent fluxes of heat, water vapour and momentum have been linked to a scheme of Louis et al. (1981), whereas the parameterization of shallow convection mainly follows Geleyn (1987). Moreover for the calculation of mountain drag, the linear gravity wave drag contribution is based on the ideas of Boer et al. (1984). The form drag contribution follows Lott and Miller (1997). Some other effects of unresolved features due to topography have been taken into account by Geleyn (2003). Beyond that for the consideration of soil processes a scheme developed by Noilham and Planton (1989) has been used.

2. The physical configuration **HARMONIE** (HIRLAM ALADIN Regional/Meso-scale Operational NWP In Europe) has been attached to the 3rd ALADIN-LAEF member (see 3.row in table 5.5). This setting enables a combination of ALADIN dynamics with HIRLAM (High Resolution Limited Area Model) physics. It is characterized by a radiation scheme (Savijarvi), a vertical diffusion scheme (CBR) and a convection and condensation scheme (STRACO).
3. The 4th, 15th and 16th member of ALADIN-LAEF have been linked to the **ALARO+3MT** setting (see 4.row in table 5.5). 3MT refers to a new parameterization scheme (Modular Multiscale Microphysic and Transport) which has been developed for handling the transition from the parameterization of sub-grid convection at coarse horizontal resolution to explicit computation of convection at cloud resolving resolutions for horizontal model resolutions between 7km and 2km (Gerard 2007; Gerard and Geleyn 2005; Priou et al. 2007).
4. In contrast the configuration **ALADIN-25** has been attached to several ensemble members: 5, 6, 9-14. (see for example 5.row in table 5.5). This configuration is based on the operational setup at Météo-France, but differs in regard to the large-scale cloudiness scheme and the precipitation scheme. Further information can be found in Lopez et al. (2002) and Geleyn (2003). It should be noted that there exist several variations for each setting such as ALARO+3MT or ALADIN-32.
5. The 7th and 8th member of ALADIN-LAEF are linked to the physical configuration **ALARO** (see 7. and 8. row in table 5.5). This physical setting differs from the basic ALADIN-25 configuration concerning large-scale precipitation, turbulent transport and radiation. In contrast to ALADIN-25 a prognostic parameterization type for resolved cloudiness and precipitation is implemented. Moreover a statistical approach is applied for the sedimentation of particles, where instead of fixed fall velocities, probabilities are used to describe the downward transfer of

precipitating species. Further information can be found in Geleyn et al. (2008). Beyond that the radiation scheme also differs from the basic setting. Therefore several modifications have been introduced. Among others a new saturation cloud model is being used, which contains a specific dependency of the cloud optical properties on cloud water and cloud ice. Besides this physical setup is also used in the operational ALADIN-AUSTRIA.

Finally it should be noted that dealing with physical parameterizations is very complex and demands lots of expertise and practice. Therefore even changing single parameters can cause completely different results.

5.4.5 Generation of surface level IC perturbations

Since the atmospheric surface conditions are never known completely due to incomplete spatial coverage as well as errors in measurements, methods to account for these uncertainties have to be found. Therefore an ensemble-based approach is considered to be suitable. Perturbing initial surface conditions, such as soil moisture, should have beneficial impact on the skill of short-range probabilistic forecasts of surface weather parameters (Sutton et al., 2006). Based on this notion a special strategy to generate perturbations to the surface variables has been developed. It is called **NCSB (Non-Cycling Surface Breeding)** and as the name already indicates, it includes another breeding variant.

For the generation of the appropriate surface perturbations, the first 16 ECMWF-EPS members are being dynamically downscaled and each member is linked to the same initial surface analysis. This surface analysis is derived from the global model ARPEGE the global counterpart to ALADIN, since there is no data assimilation done in the limited area model ALADIN. Beyond that ALADIN and ARPEGE show similar facilities concerning physical parameterization. Therefore the ARPEGE surface analysis is considered to be more compatible to ALADIN-LAEF than the ECMWF surface analysis. In the beginning each ensemble member contains its individual upper level condition and a joint surface condition (see fig. 5.6). After the breeding process the situation has considerably changed. The interaction between the individual upper level condition and the related surface condition as well as different physical settings lead to different surface conditions. These surface conditions are then blended with the ECMWF surface analysis. Finally a set of appropriate surface conditions is obtained that can be implemented into the model.

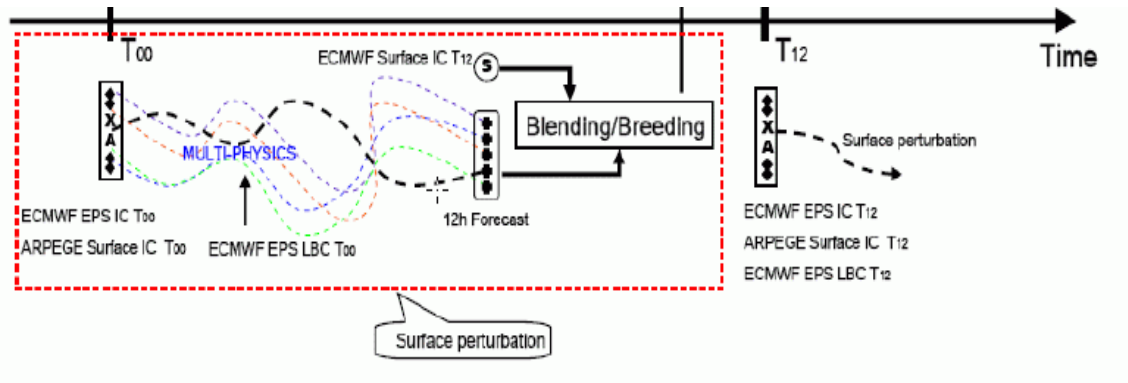


Figure 5.6: ALADIN-LAEF - Generation of IC surface perturbations,
source: Wang et al. 2011

5.4.6 Generation of upper level IC perturbations

The rationale behind the generation of upper level IC perturbations is the same as for the surface IC perturbations. Atmospheric conditions are never known completely and the mission of ensemble prediction systems is to identify the range of possibilities. In ALADIN-LAEF the generation of appropriate upper level perturbations includes two mechanisms: the breeding process and the blending mechanism. For the initialization of the process the most recent ensemble forecast is implemented (see fig. 5.7). These forecasts serve as an initial condition, whereas each ensemble member is linked to an individual physical setting. For starting the model, time lagged ECMWF-EPS forecasts serves as boundary conditions. After the integration process, the resulting ensemble forecasts (breeding vectors) are blended with the ECMWF-EPS forecasts (singular vectors). This method, as already mentioned, is thought to account for both the short time range and medium time range uncertainties (see: the blending mechanism). Finally a set of perturbations is obtained that can be implemented into the model.

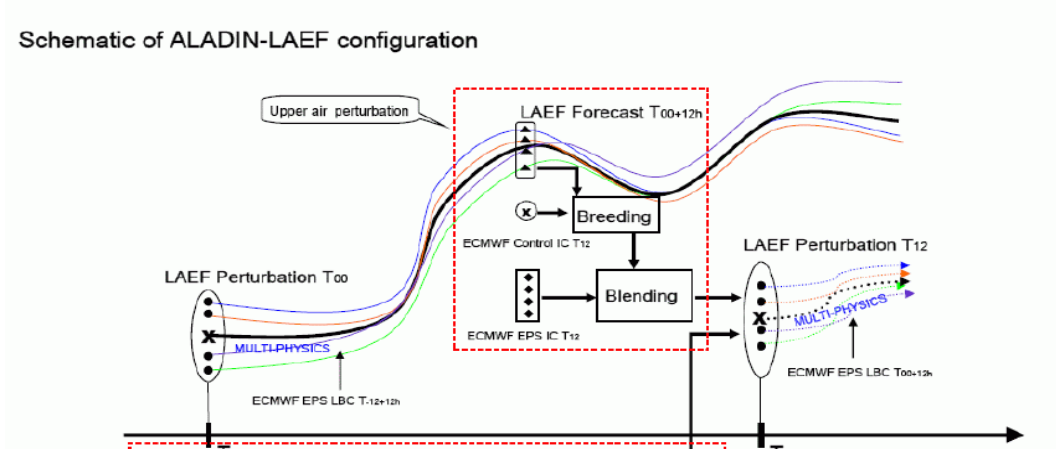


Figure 5.7: ALADIN-LAEF - Generation of IC upper level perturbations,
Source: Wang et al. 2011

5.4.7 Generation of the 16 ALADIN-LAEF members

For the generation of the 16 ALADIN-LAEF members 16 upper level perturbations are combined with 16 surface perturbations. For both sets of perturbations the ECMWF-EPS serves as an useful reference. After modifying the initial sets of perturbations (breeding & blending) the model can be started.

The preparations for the model run initialized at 12 UTC have to be started at 00 UTC. The lateral boundary conditions are provided by a 12h time lagged ECMWF-EPS forecast, since the time lagged forecasts permits an earlier model output which could not be provided using an current forecast. As already discussed, the generation of the initial surface perturbations is performed by adding 16 members from the ECMWF-EPS to a surface analysis (ARPEGE). During breeding process the surface conditions for each member will change due to interaction with the individual upper level members and their multi-physics. As a result a range of initial surface conditions is obtained. These surface conditions are blended with the ECMWF- initial surface conditions for the 12UTC model run- and can afterwards be implemented into the model. The mechanism for the free atmosphere perturbations is quite similar. As discussed before the first model start is initialized with an previously generated forecast. These perturbations are implemented in the breeding cycle, whereas the boundary conditions are again provided from the ECMWF-EPS. After the breeding cycle, the resulting perturbations are blended with the ECMWF-EPS members using a digital filter, to exclude high frequencies. Finally the combination of both results reveal requested initial conditions. The whole process is illustrated in the following imagery (fig. 5.8):

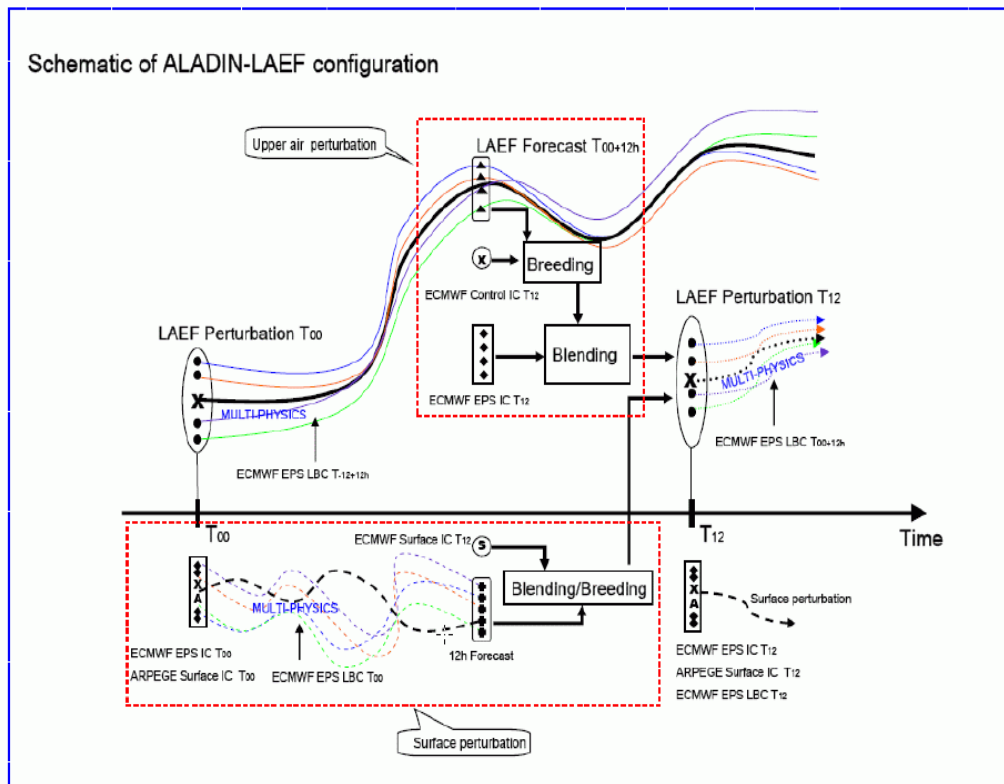


Figure 5.8: ALADIN-LAEF - Generation of Perturbations, source: Wang et al. 2011

5.4.8 Example of the blending method

The following example demonstrates the effects of the blending mechanism (see fig.5.9). Therefore 3 different fields of perturbations have been generated. For this purpose the perturbations for 1 ensemble member (here indicated as m01) on the 10th of August 2007 have been calculated. The upper graphic in picture 5.9 shows the perturbations after the breeding mechanism. Apparently the high resolution of the limited area model leads to a fine structured field of perturbations. The picture in the center represents the field of perturbations generated by singular vectors. This field is remarkably smoother and does not contain any fine structures. The bright colours (closer to white shades) indicate that the perturbations are less distinctive than those resulting of the breeding method. Several differences between the two upper pictures can be detected. Examples:

- The biggest differences between both methods appear in the Northern latitudes. For example, over Norway the breeding produces positive temperature anomalies, whereas the singular vectors generate slight negative temperature anomalies.
- Over Central Europe, e.g. Austria, again the breeding produces pronounced positive temperature anomalies, whereas the singular vectors generate extremely weak temperature anomalies.
- In the Western Parts of the Mediterranean Sea, the breeding generates positive temperature anomalies, whereas the singular vectors generate negative temperature anomalies. In the Eastern Parts of the Mediterranean Sea the opposite structures appears.
- Over the Atlantic Ocean, close to the coast of Portugal, the singular vectors produce a large area with negative temperature anomalies. In contrast the breeding mechanism generates a more distinct field, which also include a stripe of positive temperature anomalies.

The blending of these two different fields of perturbations finally produces the field below (see fig. 5.9 bottom). Apparently this field contains fine structures resulting from the breeding method. The closer analysis shows that the large scale characteristics result from the singular vector perturbations. For example in the Western Parts of the Mediterranean Sea appear negative temperature anomalies. The resulting field resembles the field generated by the singular vectors. In contrast the field generated by the breeding method contains positive temperature anomalies in this region. Beyond that similar examples can be found in different regions. This result proves what could have been expected intuitively: Large scale structures are determined by the singular vectors, whereas fine structures result from breeding the limited area model. Another interesting aspect is the impact of those fields of perturbations. For this purpose the temperature fields corresponding to the individual fields of perturbations have been calculated. In the picture below (see fig. 5.10) the upper graphic shows the temperature field which results from the breeding perturbations. The graphic in the center shows the temperature field which results from the singular vector perturbations. Again the graphic on the bottom represents the result of the blended perturbations. At first sight, the three graphics look very similar too each other. This seems a bit surprising, since the field of perturbations differ remarkably. However the closer analysis reveals a lot of differences:

- In the Atlantic Ocean, West of Ireland, the temperature field resulting from the breeding method reveals fine structures. Moreover the negative temperature anomalies are larger than those resulting from the singular vector perturbations.
- Over the Atlantic Ocean, West of Portugal, appears a pronounced stripe of positive temperature anomalies resulting from the breeding method.
- Over Western Russia, the temperature field resulting from the breeding method shows larger positive temperature anomalies than those resulting from the singular vector perturbations.

Since the formation of frontal systems which affect Europe mainly takes place over the Atlantic Ocean, it is important to know about the initial field. Differences between initial fields can lead to completely different developments!

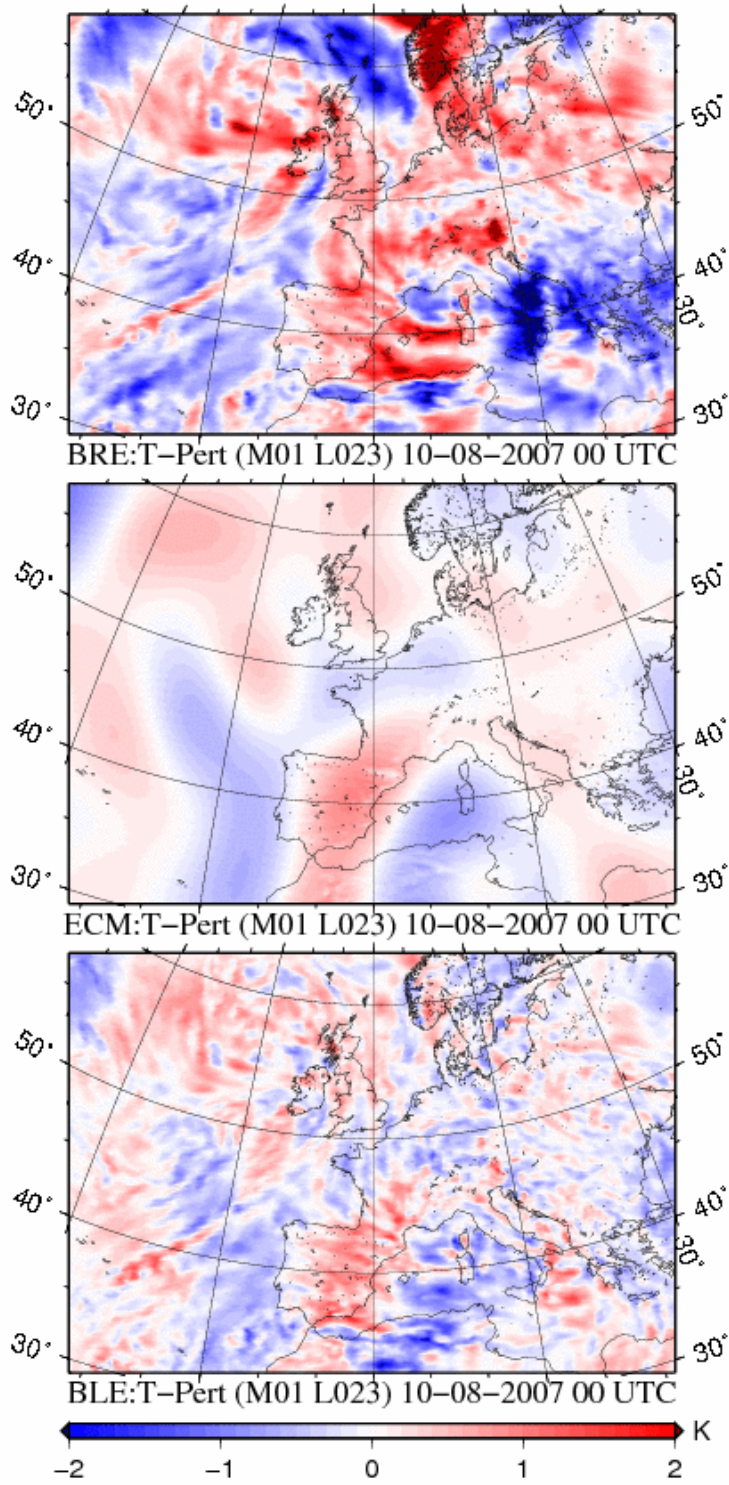


Figure 5.9: Blending Example - Perturbation
author: Martin Bellus, Project ZAMG-SHMU

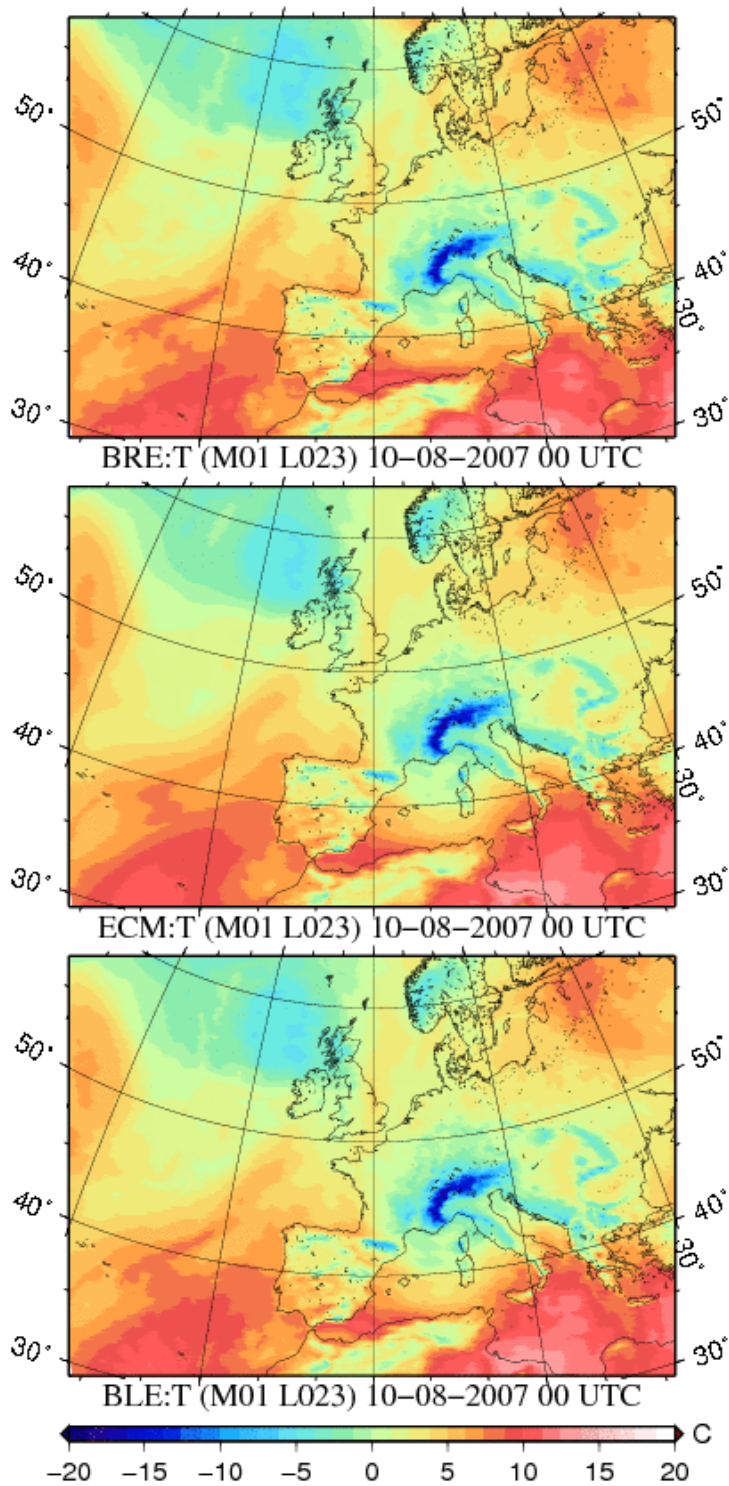


Figure 5.10: Blending Example - Resulting Temperature Field
author: Martin Bellus, Project ZAMG-SHMU

6 Data

The data involved in this investigation include the operational ECMWF-EPS data as well as the ALADIN-LAEF data ranging from 20.6.2007-20.8.2007. The reference data that serves for the skill scores in the verification process, is provided by the deterministic model ALADIN-AUSTRIA. Since ensemble prediction systems produce a large amount of data, the storage of EPS-data demands substantial computational resources, disk space and lots of effort. Therefore only a special selection of ALADIN-LAEF and ALADIN-AUSTRIA data has been stored at ZAMG. In order to adequately deal with the large data amounts produced every single day, the ECMWF has established a data archive that is called MARS (Meteorological Archival and Retrieval System). It is the main repository of meteorological data at ECMWF and contains terabytes of operational and research data as well as data from special projects. The data for Meteorological fields is available in GRIB (GRID in binary) format and BUFR (Binary Universal Form Representation) format for meteorological observations. Since its start in 1985, it has both grown in size and diversity. In order to cope with this growing archive and the ever changing requirements, a project was set up in 1996, called the DHS project (Data Handling System). In consequence MARS was totally redesigned to run on a UNIX platform. MARS data are freely available to registered users in the member states and co-operating States. There is no public access to MARS. For research and commercial use, data can be obtained through special data services. For this investigation the ECMWF-EPS data was retrieved from the MARS archive and then subjected to an interpolation process. The variables implemented in the verification procedure are the mean sea level pressure, the 2m temperature, the 10m wind, total precipitation/12h, relative humidity (500 hPa and 850hPa), Temperature (500hPa and 850 hPa) and Geopotential (500 hPa and 850 hPa). The complete set of data comprises 110 GB.

7 Verification

7.1 The purpose of Forecast Verification

Forecast verification is the main method to assess the quality of forecasts. Therefore the relationship between a matched set of forecasts and observations is closely analyzed. Verification activities are useful only if they lead to some further decisions. The decision will either generate changes in the product or in the way the forecasts are made, or it might be a "do nothing" decision which confirms that the product or service is satisfactory (Stanski et al., 1989). The first attempts to address the quality of forecasts were done in 1884, when Sergeant Finley of the US Army signal Corps tried to analyze the value of tornado forecasts (Finley, 1884). His data set only contained the information whether a tornado had or had not occurred when being predicted. Therefore he constructed a matrix that included four entries (see table 7.1). On the x-axis the observations (yes or no) were listed and on the y-axis the forecasts (yes or no). Using this list he was able to calculate several quality-defining scores. Although this was a fairly simple approach, it still serves as the basis for forecast verification. During the first half of the 20th century hardly any attempts to improve the verification techniques were done (Muller, 1944). Progressive advancements in technology enabled the construction of better numerical weather prediction models. In consequence a large expansion of weather forecast products occurred from the 1950s onwards. According to Jolliffe and Stephenson (Jolliffe and Stephenson, 2003) there was little change in verification practice until the 1980s. With the introduction of methods from the signal detection Theory (Mason, 1980) and the development of a general framework for forecast verification (Murphy and Winkler, 1987), two new concepts provided the theoretical basis for understanding the skill of binary forecasts:

Murphy and Winkler distinguished between verification measures, performance measures and scoring rules. A verification measure is any function of the forecasts or observations, but does not concern their correspondence. For example the base rate (probability of observed events) can be regarded as verification measure. In contrast performance measures address the relationship between forecasts and the related observations, either on individual or collective basis. Prominent examples are the hit rate or the false alarm rate. The scoring rule is a performance measure which is defined for each individual pair of forecasts and observations. For example the squared error can be regarded as scoring rule, but not the hit rate since it requires a number of estimates. Considering forecast quality there is an additional characteristic that should be aspired which is called "consistency". A forecast is considered to be consistent if it corresponds with the forecaster's best judgment. Some scoring rules lead to inconsistent predictions. For example, issuing forecasts that are closer to the climatologically average may reveal an overall better performance, but has nothing to do with the forecasters' best judgment. Therefore a proper scoring rule is defined in such a way that forecasters are rewarded with the best expected scores if their forecasts correspond with their judgments. A prominent example is the Brier Score. Since forecaster's judgments contain an element of uncertainty, this concept is applicable only to probabilistic forecasts. The probabilistic approach is the only method which includes all available

information. However problems due to complex dimensionality can arise. In 1979 Mason recommended an approach to deal with this dimensionality problem. With the implementation of a decision threshold, forecasts could be divided into several classes and then finally be evaluated (Mason, 1979). As a further consequence of the constant growth in forecast products, the World Meteorological Organization (WMO) published a 114-page technical report which gave a comprehensive survey of forecast verification methods in use in the late 1980s (Stanski et al., 1989).

According to Stanski verification activities in meteorology are directed by two main types of goals:

1. **Administrative:** Initially the Meteorological Service of Canada (since 1871) used verification to justify to Parliament the provision of a forecast service. In addition to justify the cost of a weather service, administrative activities today also includes the purchase of major equipment such as larger computers, the determination of when or whether to replace a forecast product with a new one, and many other decisions on the optimum employment of human and equipment resources in a weather service. Administrative verification is done on a continuing basis in an effort to monitor the overall quality of forecasts and to track changes in their quality over periods of time.
2. **Scientific:** The scientific interest in forecast verification concentrates on the identification of strengths and weaknesses of forecast products. If this is done in sufficient detail, it becomes possible to improve the forecasts. That is, to provide information to direct research and development. Apart from these two main purposes, forecast verification is also a subject for economic decisions. These decisions are highly user specific and must be evaluated on a case-by-case basis (Katz and Murphy, 1997).

All verification processes start with a matched (spatially and temporally) set of observations and forecasts. For spatial matching, observation data must either be transformed to the grid of the forecast data, or the forecast data must be interpolated in space to observation points. The former often leads to the criticism that the analyzed observations are not as representative of the truth as they were before the interpolation. This criticism becomes especially important if the analysis is an output trial field of the model which is being verified. In that case, there is potential bias in the results in favour of the model. Once the matching has been done, the next steps in processing the datasets depend on the decisions that have to be made. These decisions address the purpose of the verification (scientific or administrative). For the verification of numerical models the administrative verification answers questions about trends in model accuracy and skill. Beyond that there is often a tendency to reduce the results to a few numerical values through the use of summary scores. In contrast to the administrative purposes the scientific verification involves answering questions about spatial/and or temporal variations in the performance of the models. These can provide information that can be fed back to model developers to improve the model. For the calculation of verification scores it should be noted that there exists a range of distinct predictand types which have to be treated differently. For continuous predictands a specific value or range of values is being forecasted, such as temperature or wind. In contrast categorical predictands represent those forecasts for which the occurrence is considered to be mutually exclusive and exhaustive like for example the occurrence of measurable precipitation (either it rains or it does not). Probability forecasts are viewed as more general categorical forecasts. They apply to categorized predictands, but each category

is assigned a probability of occurrence. These probabilities must add up to 1 over all the categories of that predictand. Conversely, a categorical forecast can be viewed as a restrictive probability forecast where only the probabilities 0% and 100% are allowed for all categories. Categorical forecasts imply certainty that the chosen category will occur. Once the purpose of the verification as well as the involved predictand types have been chosen, there is a range of verification scores that can be calculated. Here is a brief sequence of this process (see fig. 7.1):

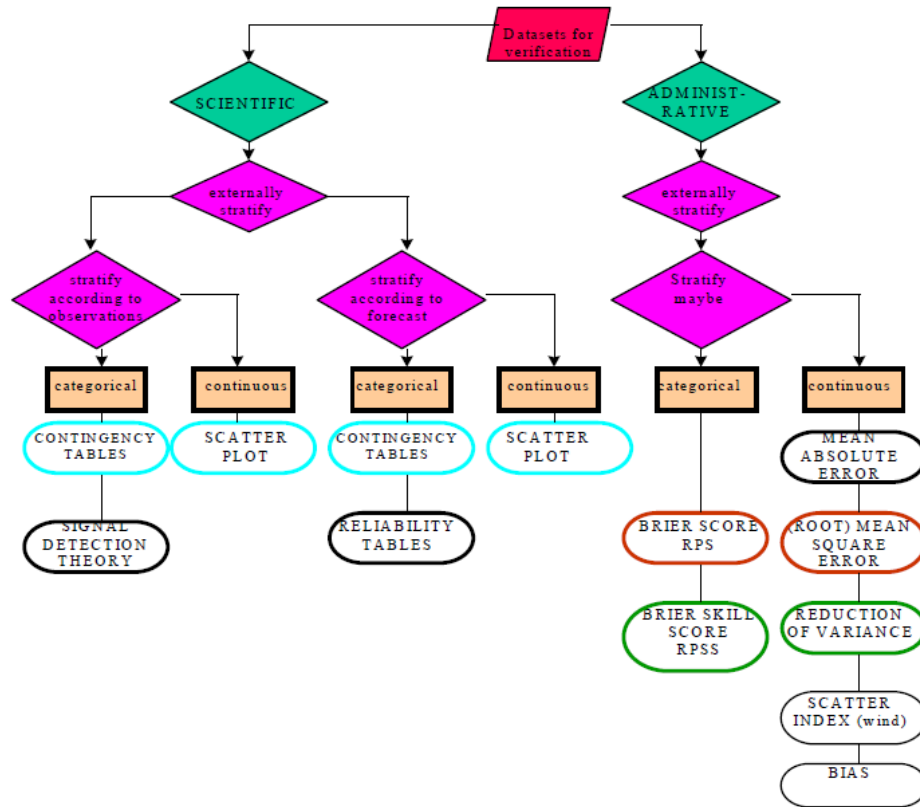


Figure 7.1: A basic model for the verification of weather elements
 source: Stanski et al., 1989, Survey of common verification methods in meteorology

7.2 Scalar Attributes of a Forecast

The quality of forecasts is often described using attributes such as skill, reliability or resolution. Although these parameters provide information about certain characteristics of the forecasts, they can not display the whole information. Even in the simplest case of a 2 dimensional matrix (e.g. table 7.1), 3 verification measures are required to specify the forecast performance. This dimensionality problem leads to difficulties in understanding and the comparison of verification statistics (Wilks, 2006). These problems can be further intensified in verification situations, where the sample sizes turns out to be too small to produce appropriate verification scores. Therefore it has become a common strategy to summarize forecast performance using one or several scalar verification scores. It should be noted that no single verification measure provides complete information about the quality of a product. What is most often done then, is to choose a suitable set of parameters describing the performance of a forecast model. The following summary of the most important forecast attributes is based on the definitions given in Stanski et al. (1989) as well as in Wilks (2006). Although the overall performance of ensemble prediction systems can be measured using these attributes, an additional selection of verification scores can give further information.

- **Accuracy** refers to the average correspondence between individual forecasts and the true weather represented by observations. As a result the forecast error is defined by the difference between a forecast value and the observed value. The smaller this difference is, the smaller the error and the greater the accuracy. A prominent measure of accuracy is the mean absolute error (MAE). However the difficult aspect about accuracy is communicating the scale of accuracy, i.e. the boundaries or limits of acceptability. A statistical forecast for freezing rain may be considered accurate for statistical purposes, but not accurate enough to be of benefit to an operational forecaster. Accuracy measures are sometimes considered to be divisible into measures of other component attributes of forecasts (reliability and resolution) and/or observations (uncertainty). For example the Brier Score can be divided into those components.
- The **skill** of a forecast defines the relative accuracy of a forecast in contrast to a reference/standard forecast. Common standards, which are considered to have no skill (i.e. the standard forecast can be generated from the observations alone), are climatology, persistence and chance. Skill scores provide a means to account for variations in accuracy which has nothing to do with the forecaster's ability to forecast.
- The **Bias**, or unconditional bias, or systematic bias, measures the correspondence between the average forecast and the average observed value of predictand. It is simply the ratio of the number of yes forecasts to the number of yes observations. Unbiased forecasts exhibit a Bias of 1, indicating that the event was forecast the same number of times that was observed. This concept is different from accuracy, which measures the average correspondence between individual pairs of forecasts and observations. Temperature forecasts that are consistently too warm or precipitation forecasts that are consistently too wet, both exhibit a bias. Therefore forecasts can be improved by feeding back bias information to the model developers.
- **Reliability**, or calibration, or conditional bias, pertains to the relationship of a forecast to the average observation, for specific values of (i.e., conditional on) the

forecast. Reliability can be represented in a reliability curve and will be discussed later in detail.

- **Resolution** is the ability of forecasting systems to discriminate among different events, producing distinct probability density functions. For example, if the distribution of observed temperatures in case of a -10°C is forecast is different to the distribution in case of a -5°C forecast, the forecasting system is said to have resolution. It should be noted that it does not matter that the label is correct. The mean observed temperature for forecasts of -5°C could be 0 and the mean observed temperature for the -10°C forecasts could be -15°C . All that is important is the sampling into characteristically different components. Since resolution depends on the ability to discern different events, resolution can be tied to the overall experience of the forecaster. When assessing the quality of forecasts, climatologically forecasts often serve as a reference forecast showing high reliability and low resolution. In contrast persistence is considered to provide high reliability and high resolution.
- **Sharpness**, or refinement, is an attribute of the forecasts alone, without regard to their corresponding observations. Forecasts that rarely deviate much from the climatologically value of the predictand exhibit low sharpness. By contrast, forecasts that are frequently much different from the climatologically value of the predictand are sharp. For probability forecasts, sharpness is the tendency toward forecasting 0% and 100% probability. Sharp forecasts will be accurate only if they also exhibit good reliability, or calibration: anyone can produce sharp forecasts, but the difficult task is to ensure that these forecasts correspond well to the subsequent observations. The essential difference between resolution and sharpness is that the former depends on both observations and forecasts while the latter depends only on forecasts. It is possible to have a forecast which is sharp but has no resolution. For example, a probability forecast of rain/no rain is sharp by tending to forecast only 100% and 0%, but has resolution only if the forecast frequency of rain given that rain occurred, is significantly different from the frequency of rain when rain did not occur. Hence resolution implies sharpness.
- **Uncertainty** is the variance of the observations in the verification example and does not depend on forecasts in any way. It is the variation in the uncertainty between datasets that make it hazardous to compare verification statistics that are sensitive to uncertainty. For example, variations in the uncertainty of observations make it inappropriate to compare temperature forecasts from one region to another without compensation for this factor.
- **Value Added:** For the user, a measure of value (utility and economic worth) is often more important than a measure of skill. For the calculation of the added value a baseline (reference) is employed. This baseline results from an unskilful forecast (e.g. climatology or persistence). For the development of weather forecast models the economic approach is an important aspect since it can justify further developments. Therefore the subject economic value and skill will be discussed in the next sections.

There is one further aspect that has to be considered when assessing the forecast quality, this is the **statistical significance**. Having calculated the skill of the forecast in relation to a particular baseline, there remains one question. Could the improvement of a forecast have arisen by chance? Therefore statistical tests namely hypothesis testing

and confidence intervals are needed to verify the result. These tests presume independence of observations. As a matter of fact the data of atmospheric conditions often contains temporal and spatial correlations and independence can not be assumed. In consequence the usual procedures have to be adapted to enable statistical testing. For the purpose of this investigation statistical tests have not been done due to the limited data. Note that ensemble verification includes a huge amount of data, but it is restricted in the sense that the data is subdivided into many parts. For common ensemble verification procedures a 2 months data period is applied which has also been done for this investigation. However the data included is not enough for sophisticated statistical tests.

7.3 Binary Events

Verification methods base on the attempts of Sergeant Finley who investigated the occurrence of tornadoes in the United States. These forecasts relate to a simple set of binary (dichotomous) events and are sometimes called yes/no forecast. Therefore a 2x2 contingency table (see table 7.1) was designed to display the possible outcomes. With the application of this table, several scores could be computed such as the hit rate, the false alarm rate or Proportion Correct.

Contingency table Forecasts	Observations		Total
	YES	NO	
YES	a	b	a+b
NO	c	d	c+d

Table 7.1: Contingency table

On the basis of this contingency table new scores were developed such as the Heidke Skill Score (Heidke 1926) or the critical success index (Donaldson et al, 1975). Here is a brief summary of the most common verification scores for binary forecasts:

- Hit Rate: $h=a/(a+c)$
- False Alarm Rate: $F=b/(b+d)$
- Proportion Correct: $PC=(a+d)/n$
- Heidke Skill Score: $HSS= (PC-E)/1-E$, whereas E denotes the proportion of forecasts that would have been correct if forecasts and observations were independent and assuming the same proportion of forecasts of occurrence to non-occurrence.
- Critical Success index: $CSI=a/(a+b+c)$

These scores are very convenient for the verification of binary forecasts, but for the evaluation of probabilistic forecasts a better fitting set of scores had to be developed. The probabilistic approach is the only method which includes all available information, but as already mentioned before problems due to complex dimensionality can arise.

7.4 Probability Considerations

On fundamental level, forecast verification involves the investigation of the properties of the joint probability distribution of forecasts and observations (Murphy and Winkler, 1987). This joint probability distribution will reflect their association and reveal deficiencies. Although this probabilistic approach includes the whole information, dealing with probability forecasts is somewhat more subtle than the verification of nonprobabilistic forecasts. The difficulty arises because for probability forecasts a broad range of values can be issued, whereas apart from the two extremes (0% and 100%) neither turns out to be completely right or wrong. Each of this probability forecasts has to be linked to the corresponding observation and finally the data involved can be enormous. Although the theoretical basis may sometimes be complex, it is important to know about the principles and how to switch between the binary and the probabilistic approach. Starting with some statistical background, we consider the observations of a categorical variable. If these observations are made over a sufficiently long period of time, the relative frequency of the event will tend to some limiting value, which is the probability of that event. Consider a random variable X that is linked to a numerical value for a mutually exclusive event. For example, $X=1$ if a thunderstorm occurs and $X=0$ if there was no thunderstorm. A particular value of the random variable X is denoted by x . The probability density function $p(x)$ of a discrete variable associates a probability with each of the possible values that can be taken by X . For example, the probability function for a thunderstorm $p(1)$ might be 0.22, whereas the probability that it does not occur is $p(0)=0.88$. What is important is that the sum of $p(x)$ over all possible values of x must be 1 by definition.

In contrast for continuous random variables the probability linked to any particular exact value is zero and positive probabilities can only be assigned to a range of values of X (Jolliffe and Stephenson, 2003). The probability density function $f(x)$ for a continuous variable has the following properties:

$$f(x) \geq 0 \quad (7.1)$$

$$\int_a^b f(x) dx = P(a \leq X \leq b) \quad (7.2)$$

where P denotes the probability that X lies in the interval from a to b

$$\int_{-\infty}^{\infty} f(x) dx = 1 \quad (7.3)$$

The joint probability density function reflects the probability that the forecast x shows a particular value when being given a particular observation value. For example a mutually exclusive event like the occurrence of a thunderstorm then shows four different probability values ($p(1,1)$, $p(1,0)$, $p(0,1)$ and $p(0,0)$) which must add up to unity. In case of continuous variables, the joint probability density function $f(x,o)$ is a function with the following properties:

$$f(x, o) \geq 0 \quad (7.4)$$

$$\int_a^b \int_c^d f(x, o) dx do = P(a \leq X \leq b \text{ and } c \leq O \leq d) \quad (7.5)$$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, o) dx do = 1 \quad (7.6)$$

The joint probability density function of forecasts and observations can be very useful, because a comprehensive appreciation of forecast quality can be achieved. Consider a forecast y_i , which can take any of the I values y_1, y_2, \dots, y_i and the corresponding observation as o_j , which can take any of the J values o_1, o_2, \dots, o_j . Then the joint probability density function of the forecasts and observations is denoted:

$$p(y_i, o_j) = \text{Probability}\{y_i, o_j\} = \text{Probability}\{y_i \cap o_j\} \quad i = 1, \dots, I \quad j = 1, \dots, J \quad (7.7)$$

This is a discrete bivariate probability density function, associating a probability with each of the $I \times J$ possible combinations of forecast and observation. Even in the simplest cases, for which $I=J=2$, this joint distribution can be difficult to use directly. Therefore Murphy and Winkler published a general framework for forecast verification based on the joint distribution of forecasts and observation. Their concept was contrary to the traditional measures-oriented approaches which often involved information loss. This alternative approach is known as distribution-orientated verification or diagnostic verification. Murphy and Winkler developed two different factorizations to build the joint distribution of forecasts and observations. With the use of the conditional probability these two different factorizations can be calculated, namely the calibration-refinement factorization and the likelihood-base rate factorization. The calibration-refinement factorization is defined by the product of the conditional probability for all the possible outcomes o_j , given one forecast y_i and the relative frequency of the forecasts (marginal distribution):

$$p(y_i, o_j) = p(o_j|y_i) p(y_i) \quad i = 1, \dots, I \quad j = 1, \dots, J \quad (7.8)$$

The term refinement refers to the dispersion of a distribution $p(y_i)$ (Wilks, 2006). Therefore a refinement distribution with a large spread implies refined forecasts, where different forecasts are issued relatively frequent, and so have the potential to discern a broad range of conditions. Conversely, if most of the forecasts y_i are the same or very similar, $p(y_i)$ is narrow, which indicates a lack of refinement. This attribute of forecast refinement often is referred to as sharpness. A prominent example to display the information based on the calibration-refinement approach is the reliability diagram. A reliability diagram consist of plots of both the calibration function and the refinement distribution, and so is a full graphical representation of the joint distribution of the forecasts and observations, through its calibration-refinement factorization.

The second factorization of the joint distribution of forecasts and observations is the likelihood-base rate factorization:

$$p(y_i, o_j) = p(y_i|o_j)p(o_j) \quad i = 1, \dots, I \quad j = 1, \dots, J \quad (7.9)$$

This factorization consists of the conditional distribution $p(y_i | o_j)$ and the relative frequency of observations $p(o_j)$. The conditional distribution expresses the likelihoods that each of the forecast values y_i would have been issued in advance of each of the observed weather events. "Although this concept may seem logically reversed, but it can reveal useful information about the nature of forecast performance" (Wilks, 2006). The relative frequency of observations is often denoted sample climatology.

Although the two factorizations of the joint distribution of forecasts and observations organize the verification information conceptually, neither reduces the dimensionality (Murphy, 1991), or degrees of freedom, of the verification problem. In fact all probabilities of the joint distribution which add to 1 must be included to express the information completely. Because of the finite nature of the verification data, explicitly

using the concept of the joint distribution requires that a parametric distribution has to be fit. The evaluation of the forecast quality is then based on the parameters of the distribution. For example, fitting a bivariate normal distribution to forecast and observations of a continuous variable would lead to five parameters: the means and variances between the forecasts and observations, and their correlation (Katz, 1982). Parametric distributions and other statistical models occasionally are assumed for the joint distribution of forecasts and observations or their factorizations, but it is far more common that scalar performance and skill measures, computed using individual forecast/observation pairs, are used for forecast verification. In practice, multiple sets of binary forecasts are often produced by varying a decision threshold over a range of control values and afterwards these sets of forecasts need to be evaluated together.

7.5 Verification of Ensemble Forecasts

7.5.1 Reliability and Resolution

The main advantage of ensemble prediction systems is their ability to produce probabilistic forecasts. Probabilistic forecasts assign values between 0 and 1 to possible future states and consequently rate the level of uncertainty. However the forecasts generated by the ensemble prediction system can only be useful, if the forecasting system satisfies several quality characteristics. There are two main forecast attributes which define a forecasting system: reliability and resolution. A forecasting system is only reliable if the probability density function of the forecasts matches to the probability density function of the observations. Note that the systematic prediction of the climatological-based values will be perfectly reliable, but there will be no discrimination among different events. In consequence another forecast characteristic is required to ensure forecast quality. The ability to differentiate different events by applying distinct forecast probabilities is called resolution. The main difference between these two attributes is the fact that resolution cannot be improved by post-processing processes whereas there is the chance to improve reliability by statistical calibration. To provide good resolution a forecast system must be able to a priori separate cases into groups with as different future outcomes as possible, so each forecast group is associated with a distinct distribution of verifying observations. In consequence a probabilistic forecast system generating perfectly reliable forecasts at maximum resolution would constitute a perfect deterministic forecast system. Reliability can be displayed using the so called reliability curve, whereas the x-axis denotes the forecast probability and the y-axis the observed probability (see fig.7.2). A perfectly reliable forecasting system would reveal a line tilted to 45 degrees. If the resulting line exceeds this reference line, the system is said to be "underforecasting" and vice versa. Note that deviations from the diagonal are not necessarily indicating true deviations from reliability but can also be due to sampling variations (Jolliffe and Stephenson, 2003). Bootstrap methods (Efron and Tibshirani, 1993) could easily be developed to quantify the sampling uncertainty in these estimates of reliability. Additionally to the reliability curve a sharpness diagram should be given. This diagram shows the frequencies for the forecast probabilities and can therefore discriminate climatology-based predictions from prediction with high resolution (see fig. 7.2).

In an effort to assess the forecast quality, different scores which incorporate attributes like reliability and resolution have been developed. The development of these scores bases on the following notion: A probabilistic forecast can be transformed into a set of discrete events by choosing a decision threshold, whereas in contrast a set of discrete

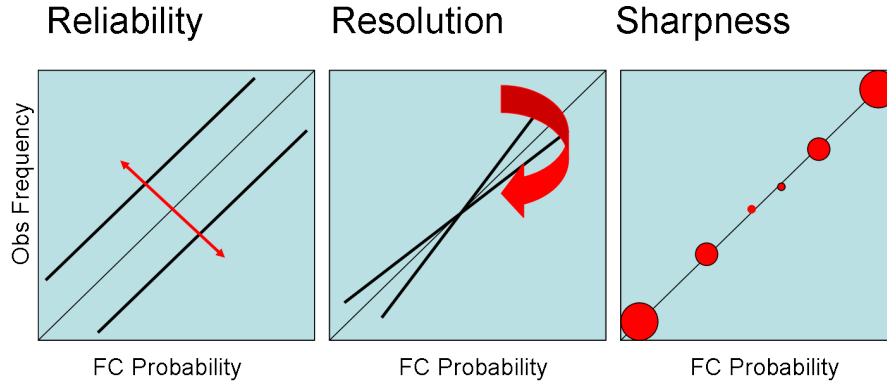


Figure 7.2: Reliability-Resolution-Sharpness
author: Prof. Manfred Dörninger, University Vienna

variables or categories can be considered as a probabilistic forecast.

7.5.2 Bias, RMSE and Ensemble Spread

The Bias which is also called systematic error, measures the average difference between an individual forecast value and the related observation at the same time and model grid point. Consequently forecasts which are continuously deviating from the corresponding observations exhibit bias. For example, temperature forecasts that are constantly too warm are considered to be biased. However there are possibilities to correct these systematic errors. The bias can be calculated in the following way (Jolliffe and Stephenson, 2003):

$$Bias = \frac{1}{N} \sum_{i=1}^N (f_i - o_i) \quad i = time \quad (7.10)$$

Here f_i denotes the forecasts and o_i denotes the corresponding observations at each forecast time step i . N refers to total number of forecast time steps.

Another measure of forecast performance is the mean squared error (MSE) which includes the quadratic difference between forecasts and observations. Therefore it is more sensitive to outliers than the bias. The MSE can be calculated according to the following equation:

$$MSE = \frac{1}{N} \sum_{i=1}^N (f_i - o_i)^2 \quad (7.11)$$

In an effort to improve model performance it appears rational to penalize larger errors. Apart from that a measure providing the same units like the observed values can be very convenient. Therefore instead of the MSE, the root mean squared error (RMSE) is being used. The RMSE meets both requirements and can be calculated in the following way:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (f_i - o_i)^2} \quad (7.12)$$

In regard to ensemble verification the RMSE also serves for different considerations. The main purpose of an ensemble prediction system is to capture the range of different developments of the atmospheric conditions. Therefore the ensemble spread should

match the estimated error size. Consequently a perfect ensemble forecasting system would reveal the quotient Spread/RMSE=1.

7.5.3 The Brier Score and Brier Skill Score

In 1950 Glenn W. Brier published a quadratic scoring measure addressing the forecast quality (Brier, 1950). In contrast to many other verification scores this measure bases on the probabilistic approach. The equation may look similar to the formula describing the mean square error (see eq. 7.11), but it differs through the implementation of probabilities. The score averages the squared differences between pairs of forecast probabilities and subsequent binary observations and is therefore negatively orientated. In a perfect forecast the Brier Score would be zero. It should be mentioned that in the original formula from 1950 the observations could only take two values (0 or 1) depending on whether the event occurred or not. The calculation of the Brier Score which has been commonly used in the last decades can be described in the following way:

$$Brier\ Score = \frac{1}{n} \sum_{k=1}^n (y_k - o_k)^2 \quad (7.13)$$

Here n refers to the total number of pairs of forecasts y_k and observations o_k . Since the Brier Score is quadratic, it can be usefully decomposed into three individual parts due to the binomial law. In 1973 Murphy introduced a decomposition of the Brier Score containing a reliability term a resolution term and an uncertainty term.

$$Brier\ Score = \frac{1}{n} \sum_{i=1}^I N_i (y_i - \bar{o}_i)^2 - \frac{1}{n} \sum_{i=1}^I N_i (\bar{o}_i - \bar{o})^2 + \bar{o} (1 - \bar{o}) \quad (7.14)$$

In this equation the first term refers to "reliability", the second term to "resolution" and the last term to "uncertainty".

The following derivation is based on the information given in (Wilks, 2006):

Consider a set of verification data which contains forecasts y_i with discrete numbers within an interval I . Then the total number of forecast-event pairs is:

$$n = \sum_{i=1}^I N_i \quad (7.15)$$

The marginal distribution of the forecasts- the refinement- in the calibration refinement factorization consists simply of the relative frequencies:

$$p(y_i) = N_i/n \quad (7.16)$$

For each of the subsamples delineated by the I allowable forecast values there is a relative frequency of occurrence of the forecast event. Since the observed event is dichotomous, a single conditional relative frequency defines the conditional distribution of observations given each forecast y_i . Then the conditional average observation can be calculated in the following way:

$$\bar{o}_i = p(o_1|y_i) = \frac{1}{N_i} \sum_{k=1}^{N_i} o_k \quad (7.17)$$

where $o_k = 1$ if the event occurs for the k^{th} forecast-event pair, $o_k = 0$ if it does not and the summation conducted only for those values of k when the forecast y_i was issued. Moreover the overall (unconditional) relative frequency, or sample climatology, of the observations is given by:

$$\bar{o} = \frac{1}{n} \sum_{i=1}^n o_k = \frac{1}{n} \sum_{i=1}^I N_i \bar{o}_i \quad (7.18)$$

On the basis of these different components, the Brier score can be expressed in terms of reliability, resolution and uncertainty (see eq. 7.14). Since a good forecasting system reveals a small Brier score, the reliability term in this decomposition has to be small, whereas the resolution term should be rather large. The uncertainty term is only linked to the sample climatology and therefore cannot be influenced by the forecasting system directly. In contrast to other verification measures, the Brier Score is strictly proper which means that the score can not be influenced by hedging or gaming. The latter implies forecasting something other than forecaster's true beliefs about future weather events. Consequently the Brier Score is often preferred when evaluating forecasts. Since the performance of a forecast model is often evaluated in comparison to another model to justify costs or further developments, an additional score is required. For this purpose the so called "skill scores" have been invented. These scores determine the forecast quality in reference to a baseline. The general definition of the skill score is the following one:

$$\text{skill score} = \frac{\text{score} - \text{score}(\text{reference})}{\text{score}_{\text{perfect}} - \text{score}(\text{reference})} \quad (7.19)$$

Consequently the skill score for a perfect forecast would be 1 and zero if the forecast was identical to the reference forecast. The choice of the reference forecast depends on several aspects such as the temporal scale and availability of data. For example, persistence is suitable for short-range forecasts, whereas climatology may be more appropriate for longer-range forecasts.

For this investigation a few additional considerations have to be made: For the calculation of the Brier Skill Score (BSS) of the surface fields, the SYNOP-based observations have been used as a reference. Apparently this set of data is only available for surface fields. Therefore a different approach has to be used for the upper level fields. For this purpose the analysis generated by a data assimilation scheme of a numerical model can be used. For the calculation of the BSS of the upper level fields, the ECMWF-analysis serves as a reference. There might arise the question why the ECMWF-analysis has been used and not the ARPEGE-analysis. The main reason is that the data assimilation scheme used at the ECMWF has proved to be superior to other assimilation schemes (Buizza et al., 2005). However it should be noted that using the analysis from the corresponding assimilation scheme can have an advantage measuring the performance. Therefore the ECMWF-EPS could benefit from this setting. This fact should be considered when addressing the model performance. The Brier Skill Score can be calculated as follows:

$$BSS = \frac{BS - BS(\text{reference})}{BS_{\text{perfect}} - BS(\text{reference})} \quad (7.20)$$

For the surface fields the 2 different reference scores have been calculated:

$$BS(ANA_{\text{surface}}) = BS + (\text{observation}(t - 6h) - \text{observation}(t)) \quad (7.21)$$

$$BS(REF_{\text{surface}}) = BS + (ALADIN - AUSTRIA_{\text{forecasts}}(t) - \text{observation}(t)) \quad (7.22)$$

For the upper level fields the 2 different reference scores have been calculated:

$$BS(AN A_{\text{upper level}}) = BS + (ECMWF_{\text{analysis}}(t - 6h) - ECMWF_{\text{analysis}}(t)) \quad (7.23)$$

$$BS(REF_{\text{upper level}}) = BS + (ECMWF_{\text{forecast}}(t) - ECMWF_{\text{analysis}}(t)) \quad (7.24)$$

7.5.4 Ranked Probability Score and Ranked Probability Skill Score

The Ranked Probability Score (RPS) can be considered as the extension of the Brier Score and is able to account for multiple events. The RPS is sensitive to distance and penalizes forecasts increasingly the more they deviate from the corresponding observations. Therefore the squared distances are computed for the cumulative probabilities of the forecast and observation vectors. The Ranked Probability Score is given by:

$$RPS = \sum_{m=1}^J (Y_m - O_m)^2 \quad (7.25)$$

Here J denotes the number of event categories and therefore also the number of probabilities included in each forecast (see example below). The RPS ranges from zero to one and is negatively orientated (the lower the value the better the forecast). In a perfect forecast system, every single vector would correspond to the observation vectors and so would their accumulations. Therefore the perfect Ranked Probability Score is zero. The cumulative forecasts and observations, denoted Y_m and O_m , are defined as functions of the components of the forecast vector y_j and observation vector o_j (Wilks, 2006).

$$Y_m = \sum_{j=1}^m y_j \quad m = 1, \dots, J \quad (7.26)$$

$$O_m = \sum_{j=1}^m o_j \quad m = 1, \dots, J \quad (7.27)$$

In consequence the RPS can be defined by:

$$RPS = \sum_{m=1}^J \left[\left(\sum_{j=1}^m y_j \right) - \left(\sum_{j=1}^m o_j \right) \right]^2 \quad (7.28)$$

Consider one single event with 3 possible outcomes such as no precipitation, light precipitation and heavy precipitation. Then each of these outcomes will be assigned a particular probability and their sum must add to one since they refer to one event. For example we assume the probability for no precipitation is 0.2, the probability for light precipitation is 0.5 and the probability for heavy precipitation is 0.3. Then the related cumulative forecast vector for this event would be: $Y_1=y_1=0.2$; $Y_2=y_1+y_2=0.7$; $Y_3=y_1+y_2+y_3=1$; Accordingly the observation has to be divided into three categories, whereas the category with the actual outcome will be assigned 1. For example, if there occurred heavy rain, the resulting cumulative observation vectors would be: $O_1=o_1=0$; $O_2=o_1+o_2=0$; $O_3=o_1+o_2+o_3=1$. As a next step the squared difference between the accumulated vectors will be calculated and then the difference will be stored. Finally the

sum of the stored differences will be calculated, reflecting the overall distance to the observations. For the special case of only two event categories, the Ranked Probability Score reduces to the Brier Score. In order to compare the model performance to reference system, the Ranked Probability Skill Score (RPSS) can be calculated. A perfect forecasting system would reveal a RPSS of 1. The RPSS is given by:

$$RPSS = \frac{RPS - RPS_{\text{ref}}}{0 - RPS_{\text{ref}}} = 1 - \frac{RPS}{RPS_{\text{ref}}} \quad (7.29)$$

7.5.5 Continuous Ranked Probability Score and Continuous Ranked Probability Skill Score

In contrast to the Ranked Probability Score which is able to account for a range of events and thresholds, the Continuous Ranked Probability Score (CRPS) can be defined for all possible thresholds. According to Wilks (2006) it is usually logistical difficult to provide a full continuous probability density function (PDF) or cumulative distribution function (CDF) for a probability forecast for a continuous predictand y , unless a conventional parametric form is assumed. In that case a particular forecast PDF or CDF can be summarized with a few specific values for the distribution parameters. Regardless of how a forecast probability distribution is expressed, it is able to include the whole range of information and rewards forecasts located near the observed value. The Continuous Ranked Probability Score can be calculated as follows (Hersbach, 2000):

$$CRPS = \int_{-\infty}^{\infty} (F(y) - F_0(y))^2 dy \quad (7.30)$$

The cumulative distribution $F(y)$ is defined by:

$$F(y) = Pr(X \leq x) = \int_{X \leq x} f(x) dx \quad (7.31)$$

$F_0(y)$ is a cumulative-probability step function that jumps from 0 to 1 at the point where the forecast variable equals the observation. Beyond that for the evaluation of the model performance in contrast to a reference forecasting system, the continuous ranked probability skill score (CRPSS) can be calculated using the following equation:

$$CRPSS = \frac{CRPS - CRPS_{\text{reference}}}{CRPS_{\text{perfect}} - CRPS_{\text{reference}}} \quad (7.32)$$

7.5.6 The ROC diagram

The Relative (or Receiver) Operating Characteristic (ROC) is a measure for the ability of a system to discriminate among different events. The method was developed for purposes in electrical engineering and has also been used widely in psychology. In 1982 Mason introduced the concept for meteorological applications. Basically the forecasting system has to choose among two alternatives: whether the event is expected to occur or not. This decision has to be based on a selected threshold which varies from situation to situation. For each selected threshold a contingency table can be drawn. This table finally provides the hit rate and the false alarm rate. The ROC diagram can be constructed by implementing the hit rate for the y-axis and the false alarm rate for the x-axis (see fig. 7.3). Point zero (0,0) in that diagram corresponds to never forecasting

the event and point 1 (1,1) to always forecasting the event. Consider the ROC diagram for perfect forecasts, which use only two probabilities: 0% and 100%. For such forecasts there is only one probability threshold from which to calculate a 2x2 contingency table (Wilks, 2006). That table for perfect forecasts exhibits false alarm rate $F = 0.0$ and hit rate $H = 1.0$, so its ROC curve consists of two line segments coincident with the left boundary and the upper boundary of the ROC diagram. At the other extreme of forecast performance, random forecasts consistent with sample climatological probabilities will exhibit the same amount of hits as false alarm rates regardless of how many or how few different probabilities are used. Consequently their ROC curve will consist of the 45° diagonal connecting the points (0,0) and (1,1). As a result forecasting systems will reveal ROC curves between these two boundaries. Comparing different forecasting systems, it can be very useful to use scalar measures instead of graphical figures. For this purpose the area under the ROC curve can be used, whereas in a perfect forecasting system the area would be 1.

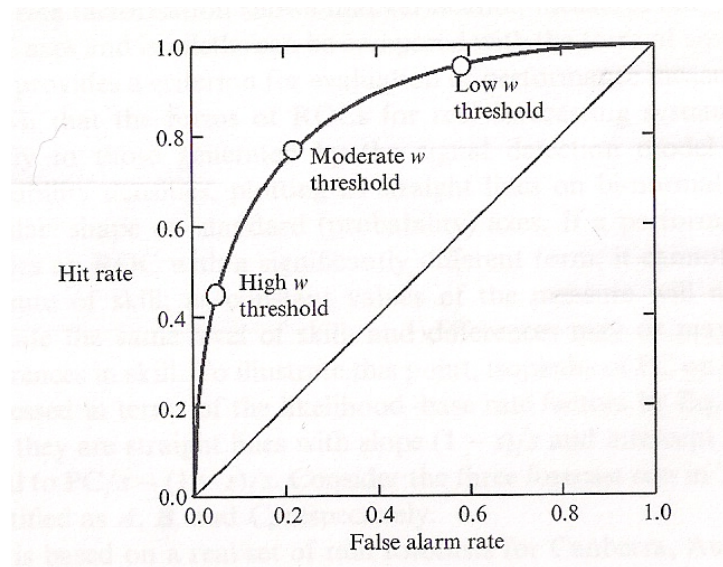


Figure 7.3: ROC; source: Jolliffe and Stephenson, 2003

7.5.7 Talagrand diagram

When assessing the quality of a forecasting system, it is important to check the consistency of the forecasts. Consistency of an ensemble means that the future atmospheric state behaves like a random draw from the same distribution that produced the ensemble (Anderson, 1997). That is, the number of the forecasts has to match the frequency of occurrence. Therefore a rank histogram, which is often called Talagrand diagram can be constructed (Talagrand et al., 1997). The x-axis of this diagram contains the ensemble members and the y-axis the quotient frequency/expected frequency (see fig. 8.17). For every observation point and date all the members are ordered according to their forecast values, from the smallest to biggest. If the consistency condition has been met, the diagram reveals a set of uniform bars, reflecting equiprobability of the observations with their ensemble distributions, except for departures that are small enough to be attributable to sampling variations (Wilks, 2006). Departures from the ideal uniformity can be used to diagnose aggregate deficiencies of the ensembles (Hamill, 2001).

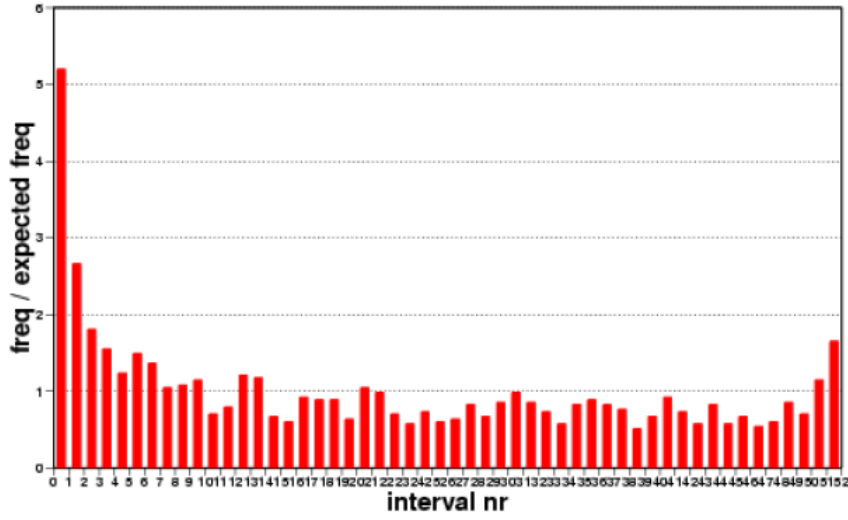


Figure 7.4: Talagrand diagram; source: ECMWF user guide, 2006

In order to meet the consistency criteria ensemble forecasts need an appropriate degree of ensemble dispersion. Moreover the consistency criteria can be affected by unconditional biases, which can also be detected in the Talagrand diagram. If an ensemble is biased, the rank histogram shows a concentration of high ranks on either the left or the right side. Moreover underdispersion will reveal a U-shaped diagram, whereas overdispersion tends to a high concentration of ranks in the center.

7.5.8 Statistical Significance

Since numerical models include a range of assumptions, results of the verification are expected to show deviations from the perfect score. Apparently the magnitude of the deviations varies for different parameters and forecast lead times. In an effort to filter significant deviations from others, the standard deviation of the scores has to be calculated. For some scores such as bias, results can be considered to follow a standard normal distribution. It should be noted that this is a special case of the Gaussian normal distribution. The density function of the Gaussian normal distribution can be written as:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad -\infty < u < \infty \quad (7.33)$$

In contrast a normal distribution with the parameters $\mu = 0$ and $\sigma = 1$ represents the standard normal distribution. The density function of the standard normal distribution can be written as:

$$\phi(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} \quad -\infty < u < \infty \quad (7.34)$$

This density function has the following main characteristics:

- $\phi(u)$ is symmetric
- the maximum of the function is located at $x = \mu$
- $\phi(u)$ is standardized, e.g. the area under the curve sums up to 1.

Thus, the area under the curve serves as a probabilistic measure. Consider a normal distributed random variable X . This variable can be easily transformed into a standard normal distributed random variable according to a linear transformation:

$$U = \frac{X - \mu}{\sigma} \quad (7.35)$$

The calculation whether a normal distributed random variable X is located within a specific region (see fig. 7.5) works as follows:

$$Pr\{X \leq x\} = F(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_x^{-\infty} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt \quad (7.36)$$

Here $Pr\{X \leq x\}$ denotes the probability that X is located in a region with the upper limit x . The same method can be applied to ensure a variable is located within a specific interval (see fig.7.6).

$$Pr\{\mu - k\sigma \leq x \leq \mu + k\sigma\} = F(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{\mu+k\sigma}^{\mu-k\sigma} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt \quad (7.37)$$

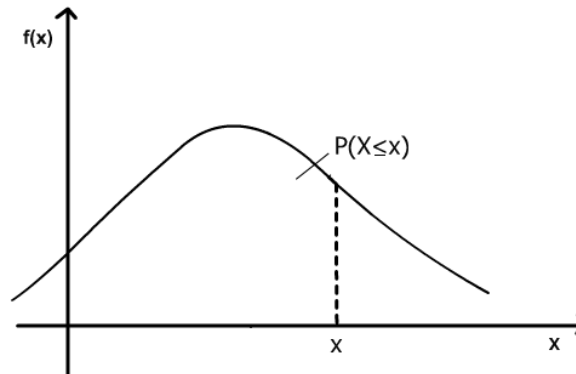


Figure 7.5: cumulative distribution function, author: Karin Schmeisser

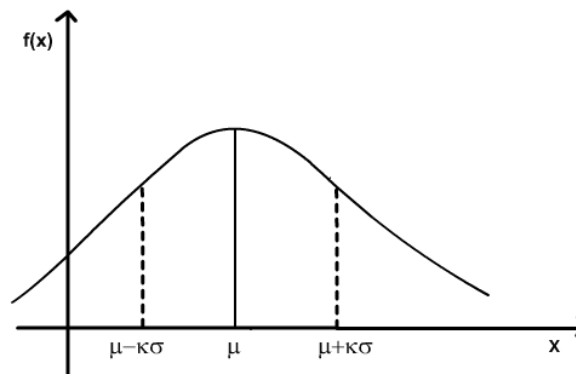


Figure 7.6: cumulative distribution function, author: Karin Schmeisser

Since the normal distribution follows a bell shaped curve and the area under this curves represents the probability that a selected variable is located within a specific region, it seems to be beneficial to stick to the mean of the distribution. Thus, the interval should be centered around the mean of the distribution to gain the maximum area under the curve. Given a set of normal distributed random variables, about 95,4% are located within the interval of $(\mu - 2\sigma, \mu + 2\sigma)$ (Papula L., 2007). According to Papula (2007) this derivation works as follows:

$$P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) = P(-2 \leq U \leq 2) = 2\phi(2) - 1 = 2 \cdot 0.9772 - 1 = 0.9544 \quad (7.38)$$

In consequence results of the verification are considered to deviate significantly, if they are located outside the interval $(\mu - 2\sigma, \mu + 2\sigma)$.

When aspiring a strictly proper statistical comparison some further aspects should be noted: The comparison of two scores involves two different probability density functions, whereas each of them is based on an estimate of the ensemble mean. Therefore both scores contain a source of uncertainty which needs to be considered when calculating statistical significant deviations. Beyond that the estimate of the standard deviations causes similar problems. For this reason an assessment of a combined uncertainty needs to be included. Based on this notion, results can only be considered to differ significantly, if there is large difference between the two intervals $(\mu - 2\sigma, \mu + 2\sigma)$. For a strictly proper statistical comparison the application of the t-distribution should be considered.

8 Results

8.1 Verification of surface fields

The quality assessment of forecasts for the surface fields has been based on the prediction of 4 important parameters, namely mean sea level pressure, temperature anomaly, wind speed and total precipitation. Atmospheric processes such as the formation of cyclones are directly linked to the location of pressure fields. If the surface pressure field differs from the upper level pressure field, further evolution of frontal systems can be expected. Therefore pressure fields such as the mean sea level pressure field are of great interest. Based on the importance of pressure fields, both ensemble prediction systems will be first evaluated in regard to the quality of their pressure forecasts. For this purpose the performance of the ECMWF-EPS based on 16 members will be compared to the performance of ALADIN-LAEF.

Since the prevailing temperatures have a major impact on people's way of living, the correct prediction of temperature can be essential. For instance agricultural work particularly depends on correct forecasts of droughts or temperatures below freezing level. However the correct prediction of temperature can be very challenging, since the temperature depends on several other parameters such as pressure, predominant winds and effects of local topography. Since the temperature is an extremely variable parameter, the verification has been linked to climatology. For each model grid point a climatological value based on the ERA40-data has been calculated and serves as a reference. Consider the verification of forecasts exceeding 30°C. Apparently in some regions this value will often occur, whereas in others temperatures will hardly ever reach this value, e.g mountainous areas. Therefore the results of the verification would be influenced by local effects. For this reason it is beneficial to detect deviations from local climatology. Therefore the verification of temperature forecasts has been linked to temperature anomalies.

In many disciplines the correct prediction of wind speed can be essential. For example in the construction industry certain operations have to be modified when strong winds are expected. Therefore the wind speed forecasts expected in a height of 10 m have been evaluated.

The formation of precipitation is a very complex process and depends on several components such as humidity, temperature and wind (updraft, downdraft). Since the occurrence of precipitation is often restricted to sub grid-scale, precipitation forming processes have to be parameterized. This fact already demonstrates that there is a huge range of uncertainty involved in the prediction of precipitation. However the correct prediction of precipitation can be crucial for outdoor-events or road constructions. Meteorological offices are frequently consulted for rain forecasts. Therefore it is important to be aware of the the quality of rain forecasts as well as its drawbacks. Beyond that it should be noted that both EPS differ significantly in regard to the modeling of precipitation. This fact would already suggest different results for both models.

For the verification of these 4 parameters, the surface observations have been used as a reference. Each parameter has been addressed individually and analyzed for selected thresholds and forecast lead times. In the consequent subchapters a brief selection of

the most striking results will be shown. Finally a comprehensive overview of the model performance will be given. Detailed results can be extracted from the attached data sheet in the appendix.

8.1.1 Mean Sea Level Pressure

As a first measure of model performance, systematic errors (=bias) should be identified. In the verification diagram the bias of the ECMWF-EPS is negatively orientated and significantly deviates from the perfect score which is zero (see fig. 8.1). In contrast the bias of ALADIN-LAEF is positively orientated and also deviates remarkably from the perfect score. Moreover both models forecasts show diurnal variations, why is a typical characteristic of numerical models. The average performance over all forecast lead times reveals that the ECMWF-EPS (average bias = -0.26160) is slightly closer to the perfect score of zero than ALADIN-LAEF (average bias = 0.39525). In an effort to demonstrate statistical significant deviations between both models, the standard deviation σ has been calculated. Therefore the bias scores of the ensemble mean have been used. Statistical significant deviations are considered to be located outside the interval of $\pm\sigma$. Results show a clear overlap between both models for shorter forecast lead times. Thus, no significant deviations can be detected. However the ECMWF-EPS reveals a remarkable larger dispersion. For longer forecast lead times, the error bars show small regions with no overlap. Since this differences are extremely small, it can be said that the comparison of both models shows no statistical significant differences. In contrast to the bias, the root mean square error (RMSE) penalizes outliers more vigorously. Results reveal that both models differ only slightly in regard to RMSE. During shorter forecast lead times (up to 18h) the performance of ALADIN-LAEF is a little better than the performance of the ECMWF-EPS. This result is quite encouraging, since ALADIN-LAEF has been particularly designed for short forecast lead times. For larger forecast lead times the ECMWF-EPS scores better in terms of RMSE. Moreover the performance of both models show hardly any differences in regard to ensemble spread. Since one of the main objectives of ensemble prediction systems is to account for the involved uncertainties, the ensemble spread should match the RMSE. In terms of RMSE/spread the average performance of ALADIN-LAEF is better than the average performance of the ECMWF-EPS.

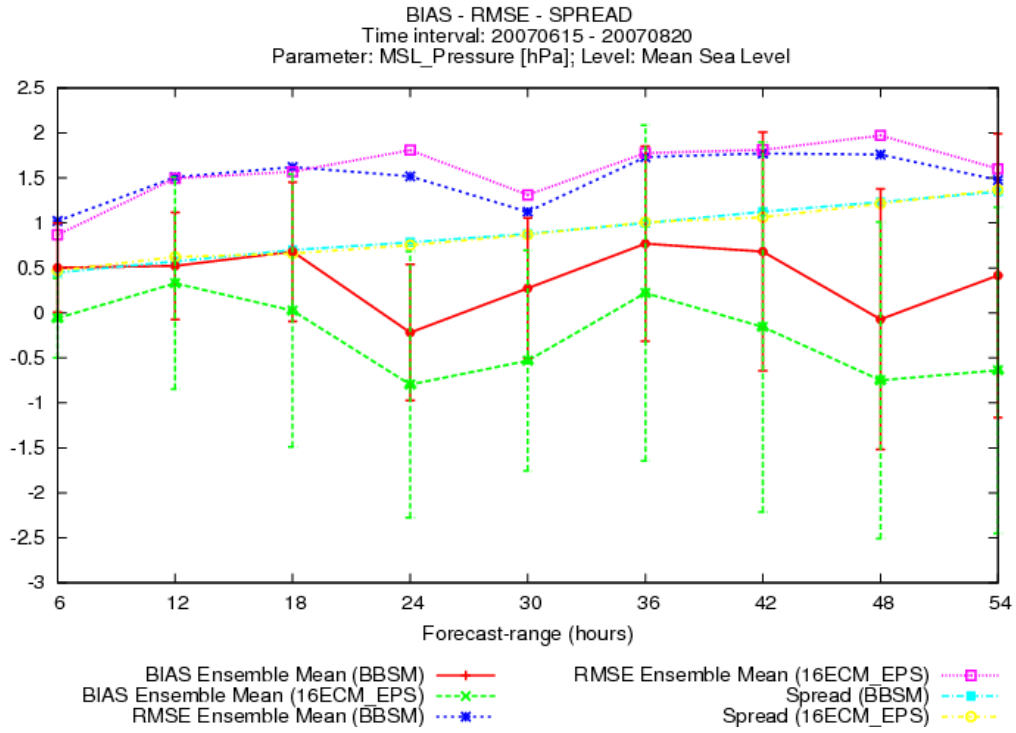


Figure 8.1: mean sea level pressure - BIAS-RMSE-SPREAD

Another quality defining measure is the Brier Score. For the verification of the mean sea level pressure forecasts, three different thresholds (1005 hPa, 1010 hPa and 1015 hPa) have been applied. Generally there are hardly any differences detectable between the performance of ALADIN-LAEF and the ECMWF-EPS. Both models score best for the pressure threshold of 1005 hPa (see fig. 8.2). This overall similar performance indicates that the forecast probabilities, generated at each model grid point, are quite similar. Thus, the ensemble spread is very similar, as has been already shown above.

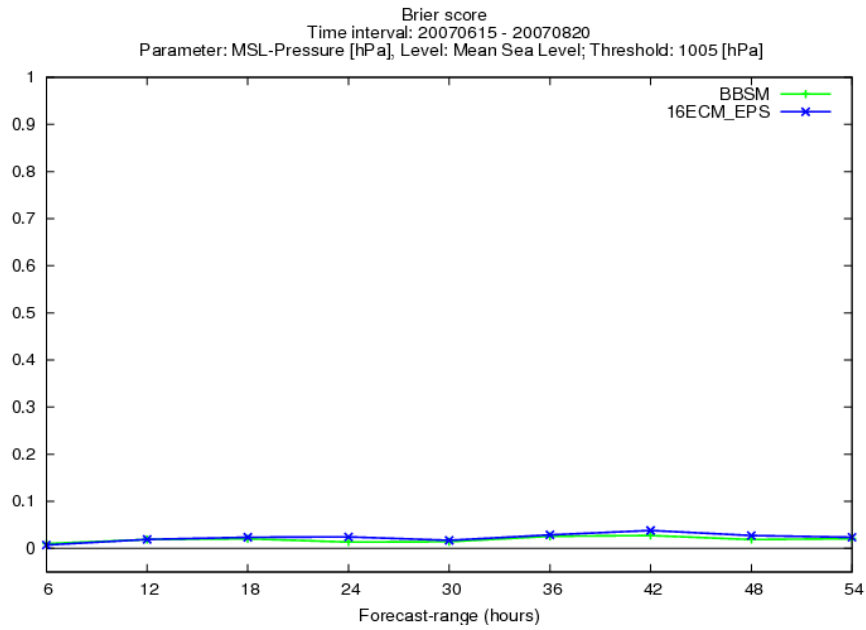


Figure 8.2: mean sea level pressure - Brier Score, threshold 1005 hPa

Since the quality of a numerical model is often addressed in comparison to another (former) model, verification scores such as skill scores which highlight these differences are extremely valuable. For this purpose the Brier Skill Score has been calculated. Two references have been used: The first reference includes preceding observations (see chapter 5.5.3). The second reference is based on the forecasts of the deterministic model ALADIN. Results show that for both models and both references the forecasts of the 1005 hPa-field were close to the perfect score which is 1 (see fig. 8.4). In contrast the forecasts for the 1015 hPa-field were inferior. The Brier Skill Score for both models ranged between 0.7 and 0.9 and showed diurnal variations.

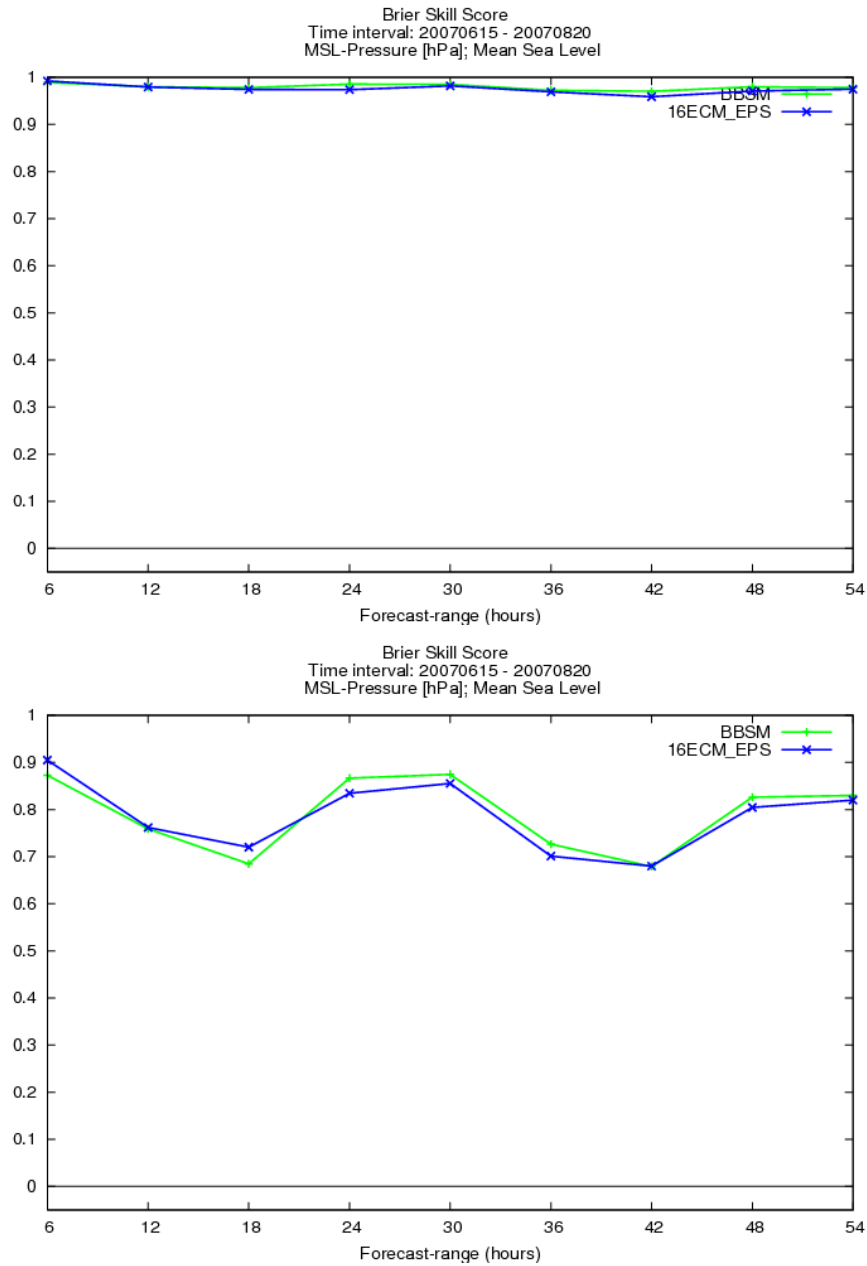


Figure 8.3: mean sea level pressure - Brier Skill Score, threshold 1005 hPa (top) and 1015 hPa (below), reference: Analysis

Beyond that it can be interesting to analyze the impact of the different references. Therefore the results based on the two different references (preceding observations & ALADIN forecasts) have been compared. As it turned out the Brier Skill Score differs most for the threshold of 1015 hPa. The performance of both models is clearly better, when being compared to preceding observations. When the quality of the forecasts is compared to the ALADIN deterministic models, the Brier Skill Scores shows significantly smaller values.

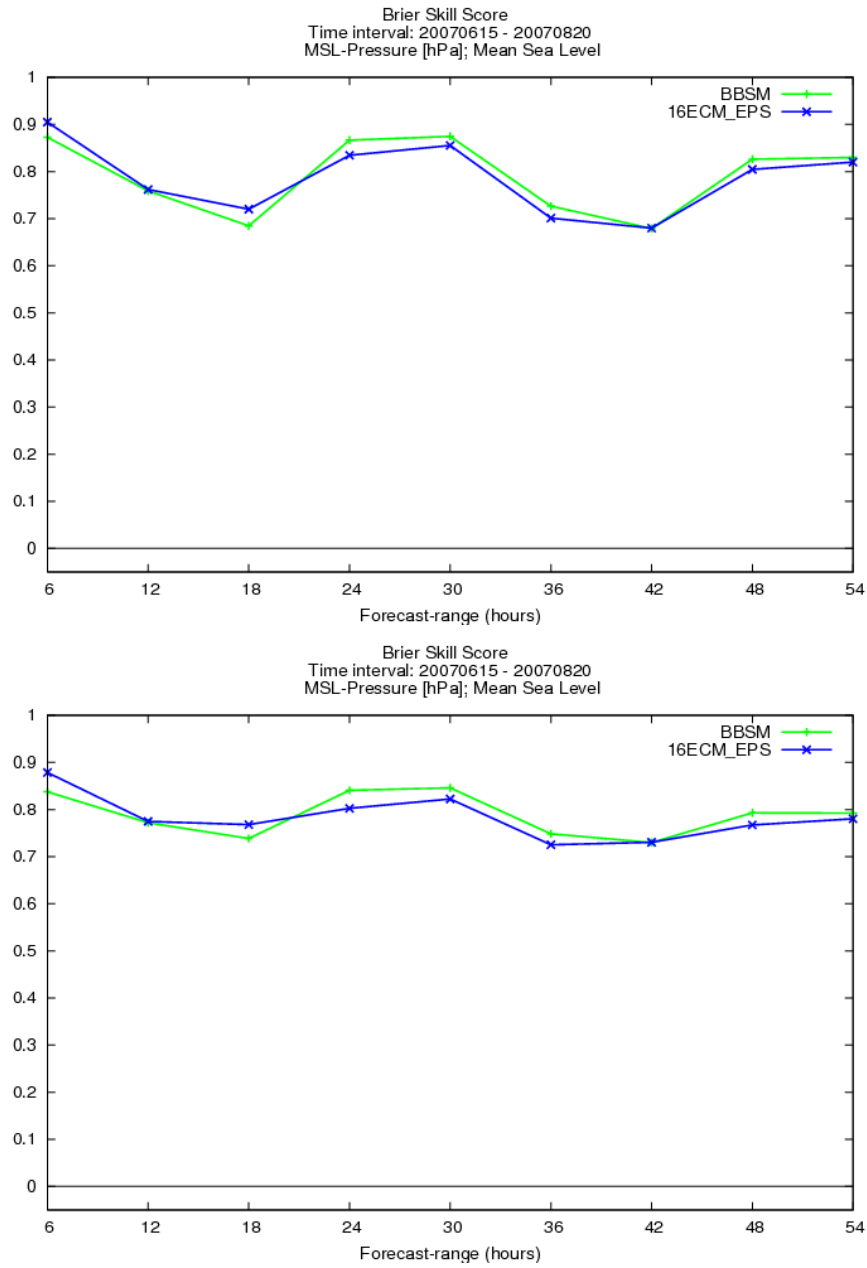


Figure 8.4: Comparison of Brier Skill Score for two different references:
ALADIN analysis (top) and ALADIN forecasts (below),
mean sea level pressure, threshold 1015hPa

Since the Brier Skill Score can only consider one specific threshold, another measure including a set of different thresholds can give a more comprehensive overview. For this purpose Ranked Probability Score (RPS) has been calculated (see fig. 8.5). Both models score very similarly and are close to the perfect score of zero.

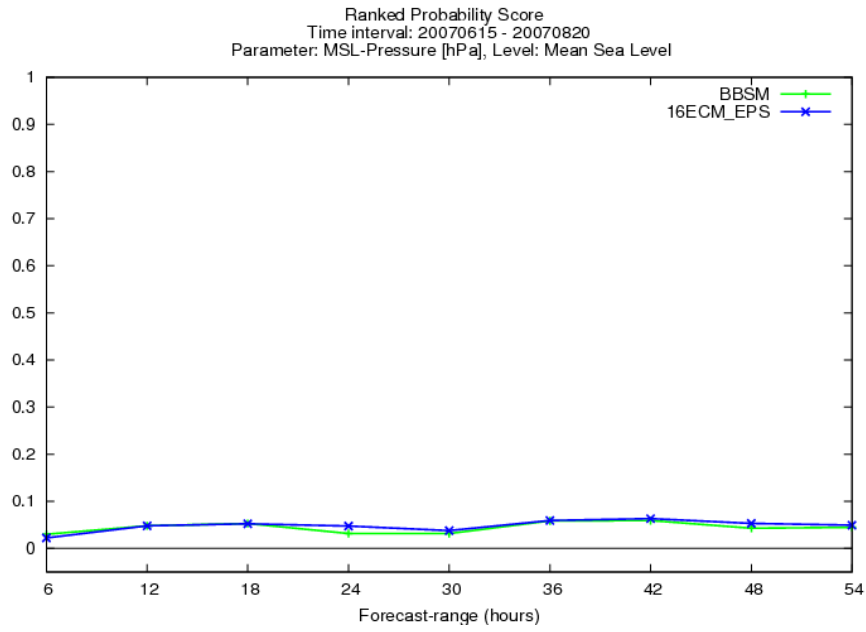


Figure 8.5: mean sea level pressure - Ranked Probability Score

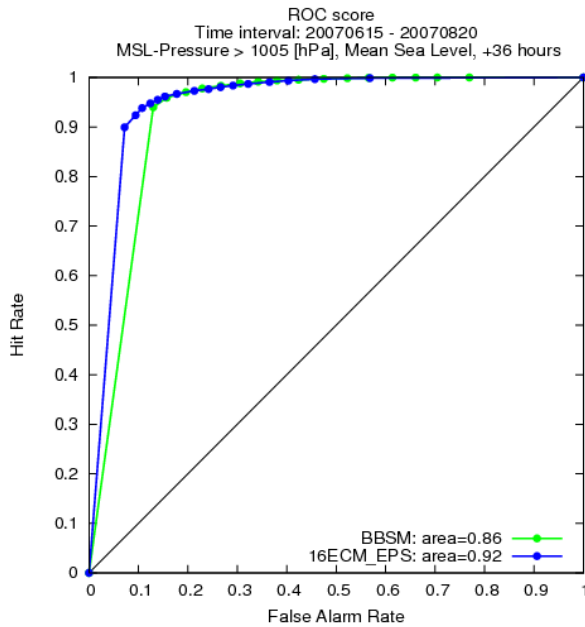


Figure 8.6: mean sea level pressure - ROC, 36h-forecast - threshold: 1005hPa

Apart from measuring the quality of forecasts in terms of accuracy (bias, RMSE), the forecasts should also be analyzed in regard to reliability. For this verification the thresholds: 1005 hPa, 1010 hPa and 1015 hPa as well as forecast lead times ranging from 6 to 54 hours have been applied. Since perfect reliable forecasts are located on the diagonal of the reliability diagram, their averaged value would be 0.5. In an effort to enable an objective comparison, the average reliability has been calculated. The comparison of both models reveals an overall very similar performance. However the quality of the forecasts varies for different forecast lead times. For forecast lead times of +6 hours

and all three thresholds the ECMWF-EPS clearly dominates ALADIN-LAEF. This result should be linked to extremely small bias of the ECMWF-EPS at that particular forecast lead time. For longer lead times the differences between both models get smaller. For forecast lead times ranging from 24 to 30 hours, ALADIN-LAEF performs better (see fig. 8.7). For these forecast lead times ALADIN-LAEF exhibits a clearly smaller bias than the ECMWF-EPS. The analysis of all forecast lead times and different thresholds reveals a slightly better performance of ALADIN-LAEF. ALADIN-LAEF scores better in 16 of 27 different settings.

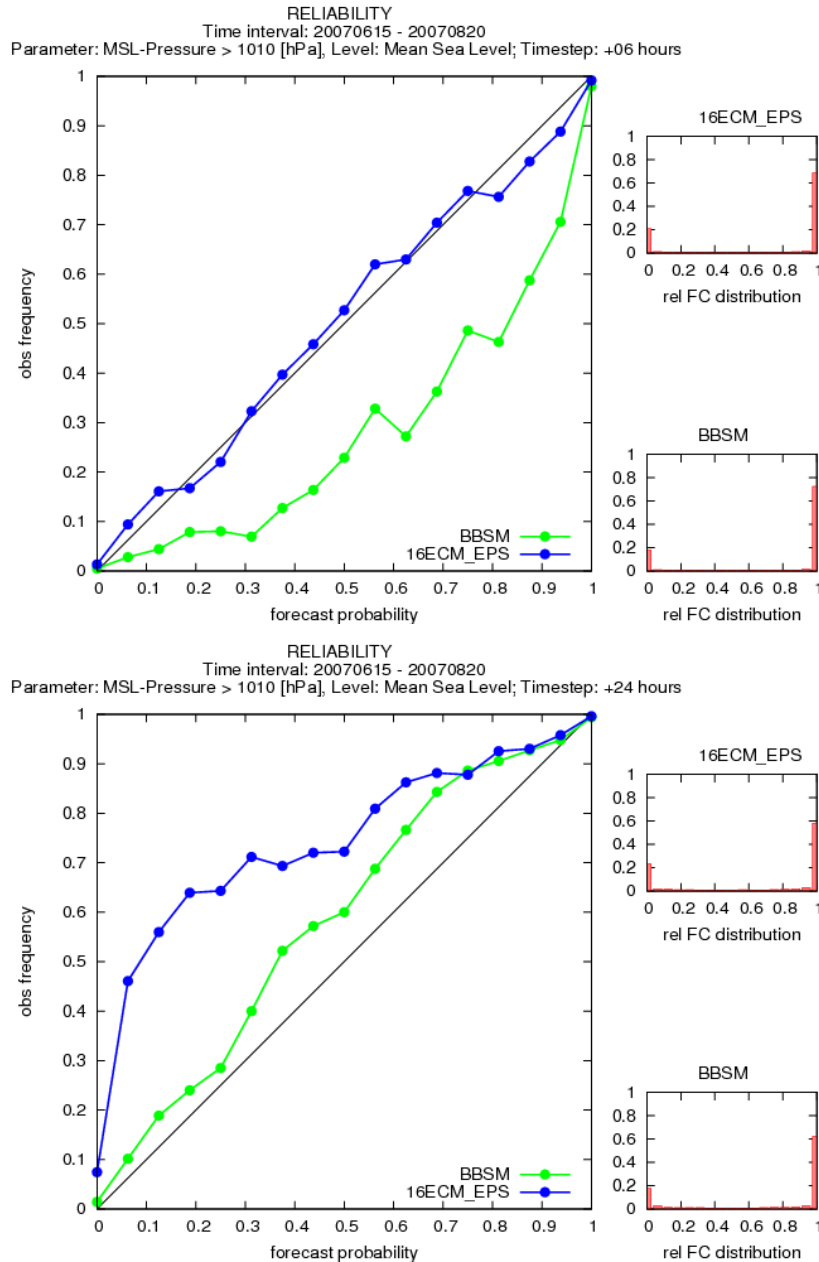


Figure 8.7: mean sea level pressure - Reliability diagram, 6h- and 24h- forecasts, threshold: 1010hPa

The quality of ensemble prediction systems also depends on their ability to discriminate among different events. This ability can be assessed using the so called ROC-curve,

which compares hits (predicted events that did occur) with false alarms (predicted events that did not occur). For the verification of the mean sea level pressure field three different thresholds (1005 hPa, 1010 hPa and 1015 hPa) have been used as well as forecast lead times ranging from 6 to 54 hours. For the forecast lead time of 6 hours and all three thresholds, the ECMWF-EPS scores better. The ROC-areas calculated for the other forecast lead times are quite similar apart from one special result: The biggest difference in model performance is detected for the forecast lead time of 36 hours and the pressure threshold of 1005 hPa (see fig. 8.6). For this setting, the ECMWF-EPS scores notably better than ALADIN-LAEF. This fact can easily be assessed comparing the two areas under the ROC-curves, whereas the area for the ECMWF-EPS is larger.

In summary the verification of the mean sea level forecasts reveals only small differences between both EPS. The differences are mainly present for the forecast lead time of 6 hours, whereas the ECMWF-EPS there scores better. Although the differences between both models are relatively small, the comprehensive analysis of 80 different configurations reveals a better performance of the ECMWF-EPS. In 47 out of 80 cases the ECMWF-EPS scores better than ALADIN-LAEF (see table 8.1 in short summary). The reason might be a better incorporation of boundary conditions for the shortest forecast lead time of 6 hours. However it should be noted that this detailed comparison has one considerable drawback: The impact of the different scores to the final result is not balanced. Thus, the bias scores only once, whereas for example reliability scores 9 times (9 different settings). In an effort to generate a fair comparison, 16 major scores have been calculated (see table 8.1 in short summary). The calculation of each major score is based on the ensemble mean averaged over all thresholds. Thus, both bias and reliability get the same impact (1/16) to the final result. The comparison of these 16 major scores shows a better performance of ALDIN-LAEF. ALADIN-LAEF scores better in 9 of 16 cases. However the differences between both models are rather small. It should be further noted that only in case of the mean sea level pressure, results based on 16 major scores draw a different picture than results based on individual thresholds.

8.1.2 Temperature Anomaly

Results of the verification show that the ECMWF-EPS scores better in terms of bias (see fig. 8.8). The temperature forecasts of the ECMWF-EPS contain hardly any remarkable bias. In contrast ALADIN-LAEF constantly predicts too low temperatures (average bias: -0.99255). The constantly negative bias results from the surface set up in ALADIN-LAEF. This surface set up includes the ARPEGE surface analysis and introduces a negative bias. Experiments have shown that with the implementation of the ECMWF surface analysis an even larger negative bias would be introduced. Therefore it seems to be better to stick to the ARPEGE surface analysis. As the calculation of the standard deviation shows, the differences between both models are significant. Thus, the error bars containing the confidence interval of $\pm\sigma$, differ remarkably. Moreover the verification of the temperature forecasts shows that the ECMWF-EPS scores better in terms of RMSE for all forecast lead times. Beyond that the ECMWF-EPS is superior in terms of RMSE/spread.

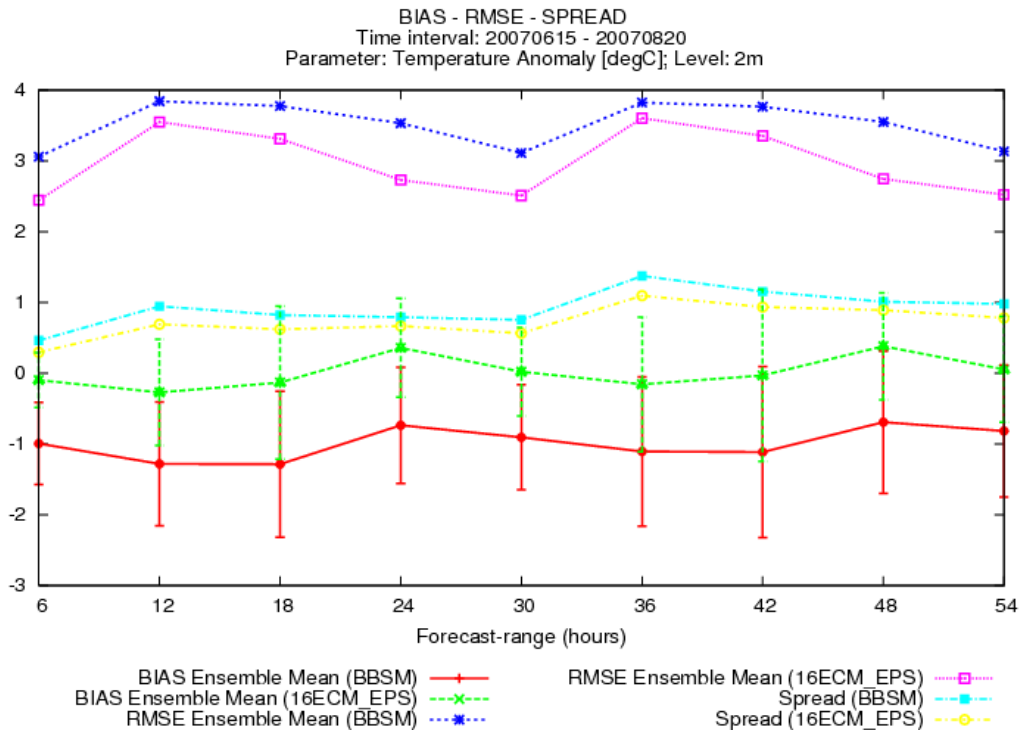


Figure 8.8: temperature anomaly - Bias-RMSE-Spread

When the performance of the models is analyzed in terms of Brier Score, there are hardly any differences detectable (see fig. 8.9 and fig. 8.10). The ECMWF-EPS scores slightly better for negative temperature anomalies ($< -2^{\circ}\text{C}$) and ALADIN-LAEF for positive temperature anomalies ($> 2^{\circ}\text{C}$). In consequence the Ranked Probability Score reveals an almost similar result for both models which is close to 0.1 (see fig. 8.12). Since the ECMWF-EPS reveals better results in terms of the Continuous Ranked Probability Score, both models obviously differ when a broader range of different thresholds is applied (see fig. 8.14). When the model performance is evaluated in regard to reliability (see fig. 8.15 and fig. 8.16), the ECMWF-EPS generally scores better for negative temperature anomalies ($< -2^{\circ}\text{C}$). In contrast ALADIN-LAEF shows a better performance for most of the positive temperature anomalies ($> 2^{\circ}\text{C}$). This demonstrates again, what the

results of the Brier Scores have already shown. The comparison of all reliability diagram reveals a better performance of the ECMWF-EPS. However it should be noted that this result is linked to the remarkably smaller bias of the ECMWF-EPS! Although the differences are relatively small (range 0.1), the ECMWF-EPS outperforms ALADIN-LAEF in 25 of 27 cases. Next to reliability, the ability to discriminate among different events is an important attribute. Therefore the Relative Operating Characteristic serves as a measure. When the performance of both models is assessed in terms of ROC, the ECMWF-EPS scores better in 16 out of 27 cases. Finally the quality of the models can be analyzed using the Talagrand diagram. For both models this diagram reveals under dispersion (U-shaped form). This means that the ensemble members tend to be much like each other. For all forecast lead times ALADIN-LAEF produces a Talagrand diagram which is closer to the ideal rank distribution (see fig. 8.17). However it should be noted that ALADIN-LAEF is negatively biased, which can be detected in the Talagrand diagram as well (left column).

The final comprehensive analysis of all scores yields a better performance of the ECMWF-EPS. The ECMWF-EPS scores better in 55 of 82 different settings (see table 8.1). This superiority not only addresses accuracy (bias, RMSE, etc.), but also reliability. The comparison of 16 major scores also highlights the better performance of the ECMWF-EPS (see table 8.1).

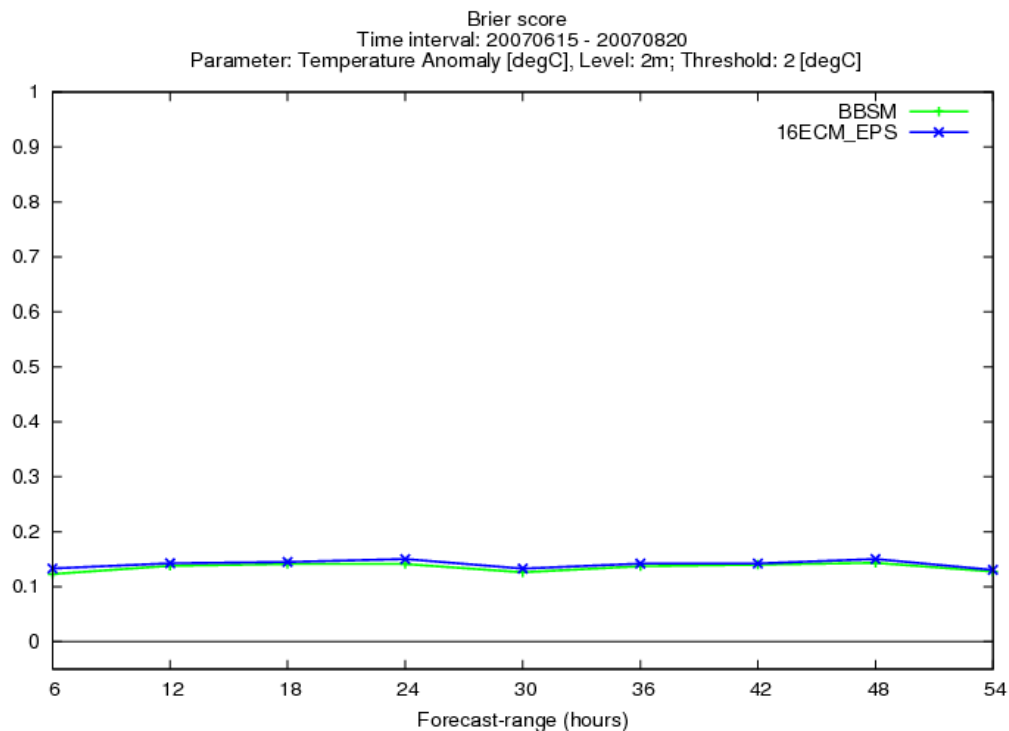


Figure 8.9: temperature anomaly - Brier Score, threshold: $>2^{\circ}\text{C}$

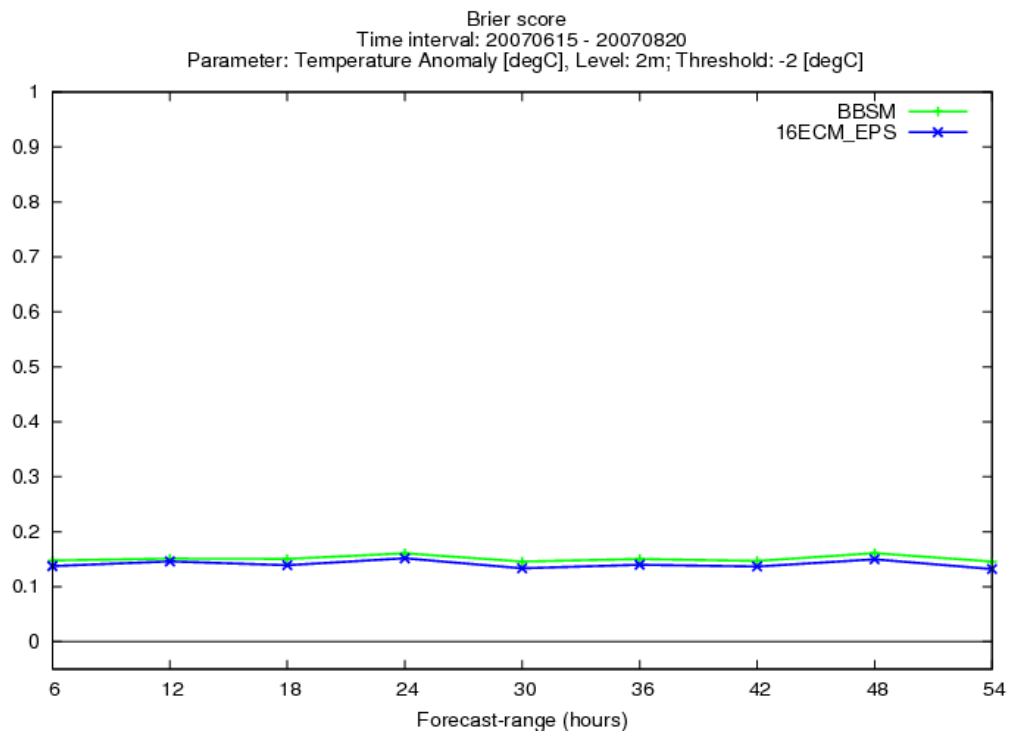


Figure 8.10: temperature anomaly - Brier Score, thresholds: $<-2^{\circ}\text{C}$

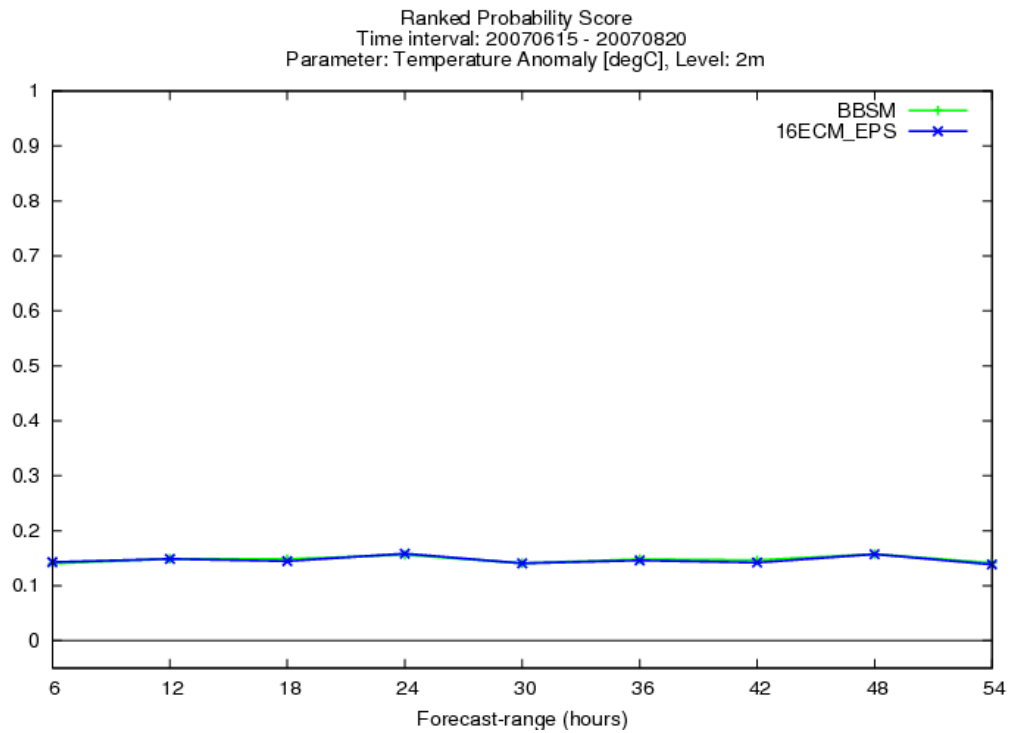


Figure 8.11: temperature anomaly - Ranked Probability Score

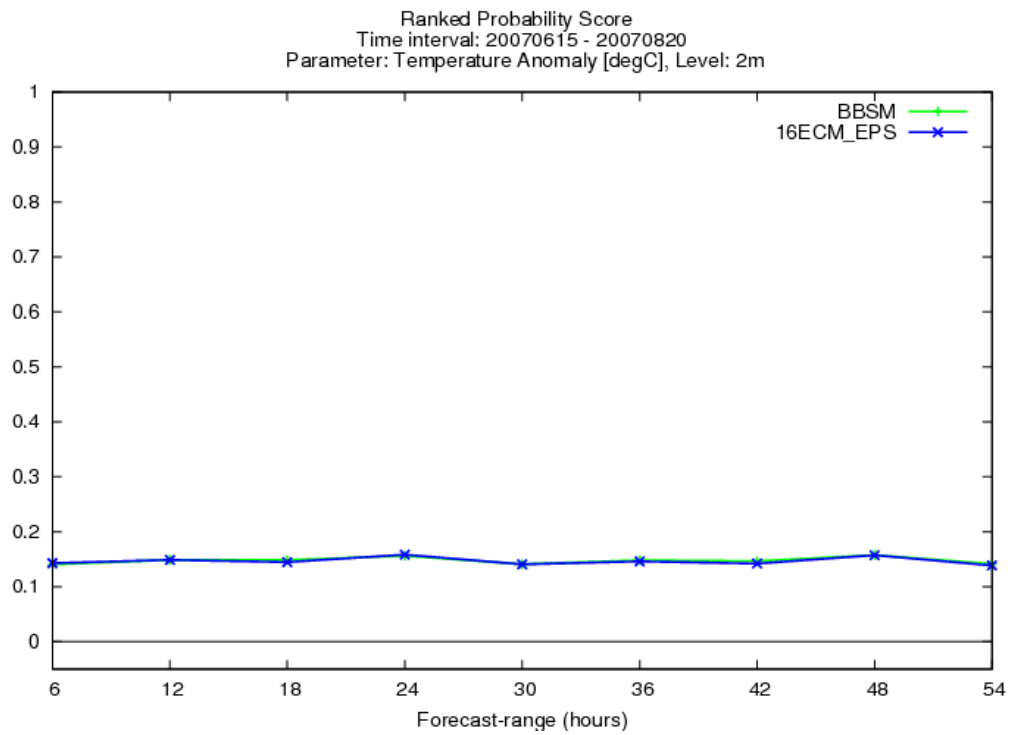


Figure 8.12: temperature anomaly - Ranked Probability Score

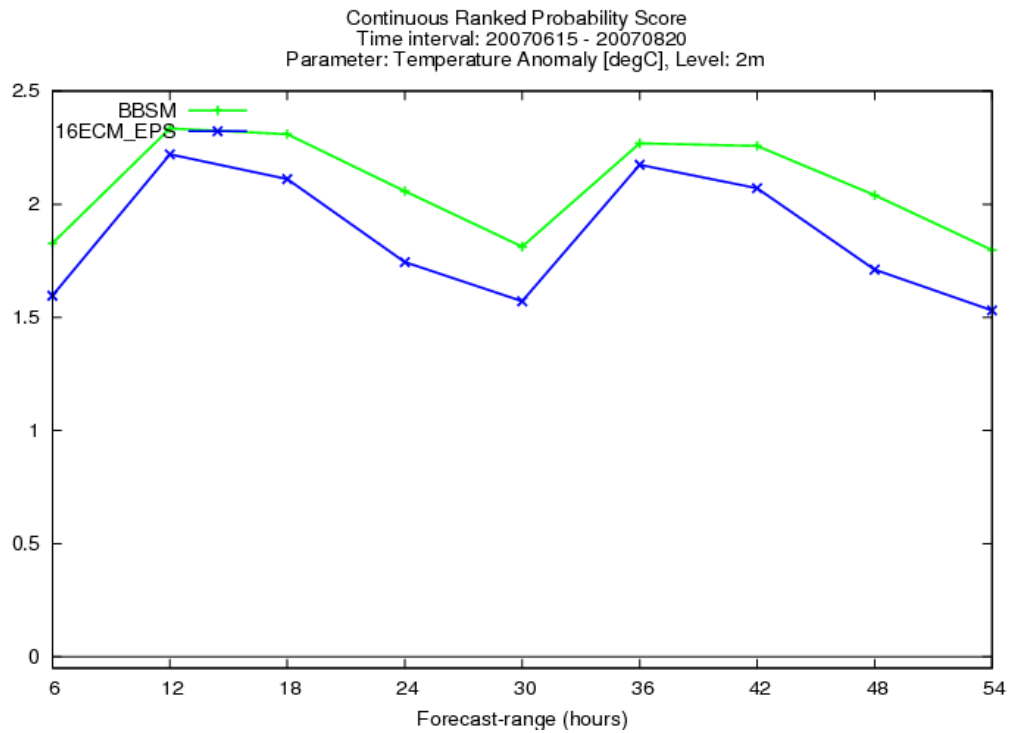


Figure 8.13: temperature anomaly - Continuous Ranked Probability Score

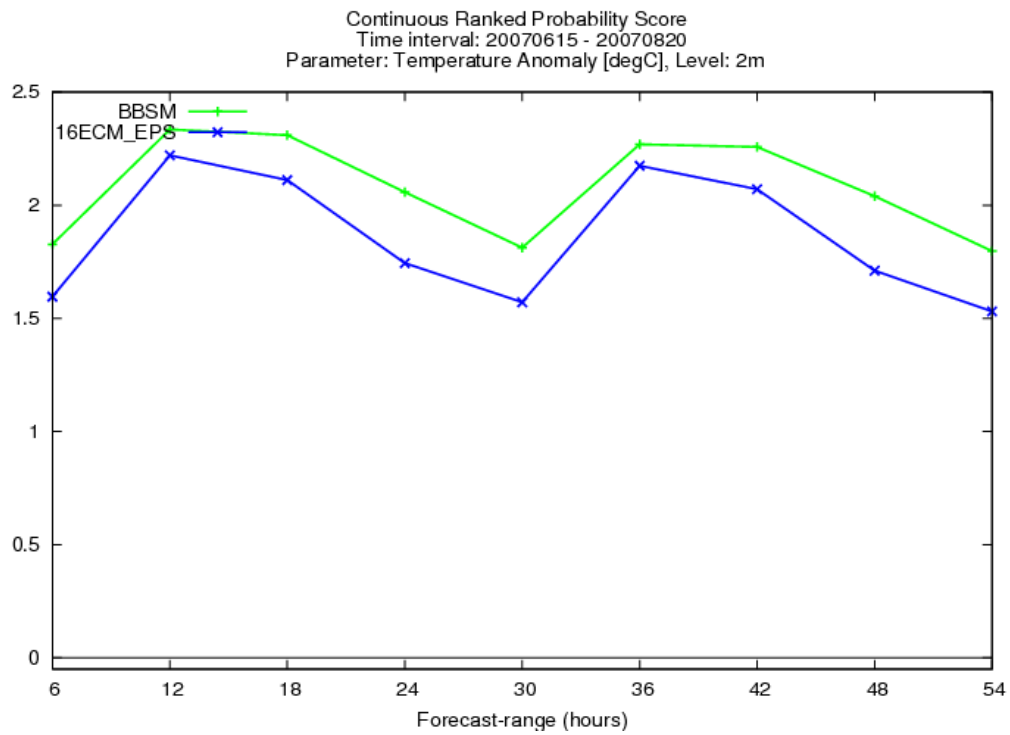


Figure 8.14: temperature anomaly - Continuous Ranked Probability Score

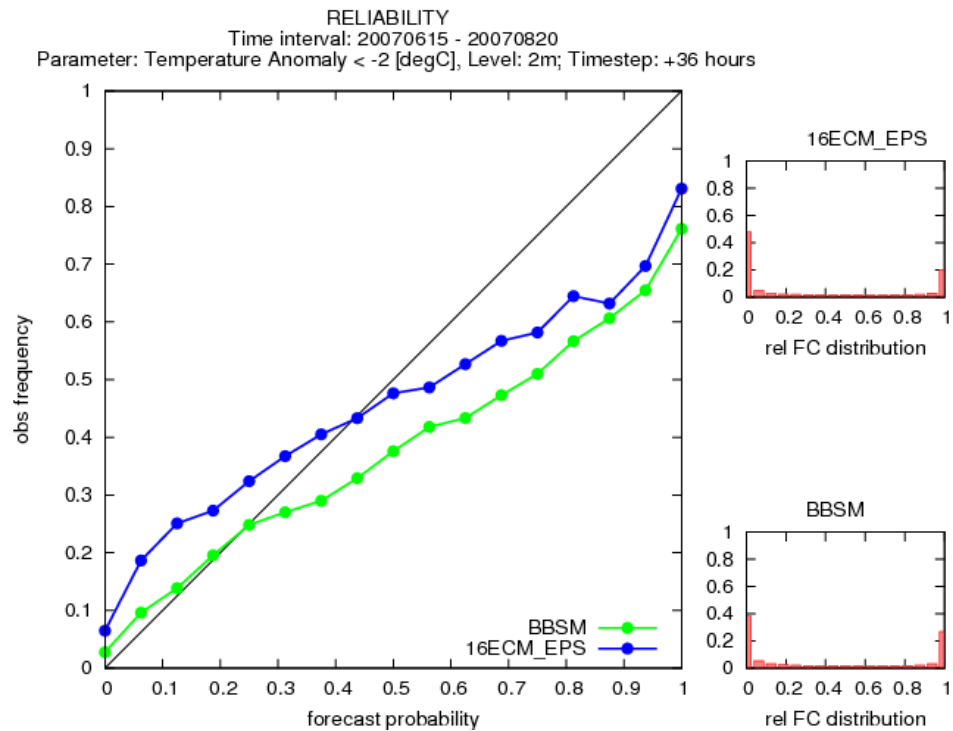


Figure 8.15: temperature anomaly - Reliability diagram
forecast lead time:36h, negative temperature anomaly

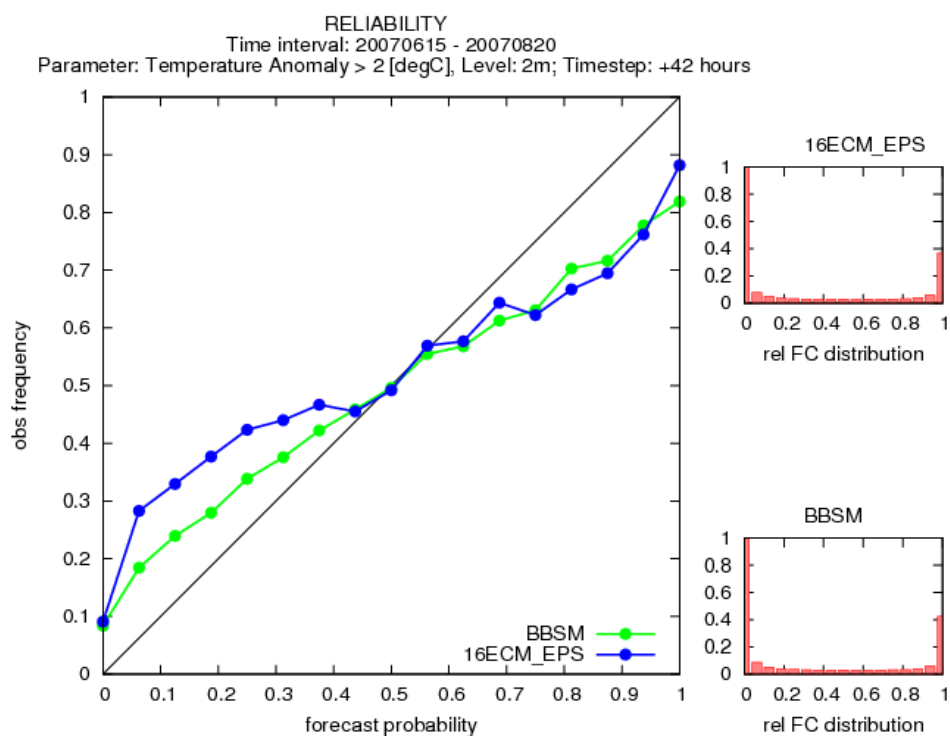


Figure 8.16: temperature anomaly - Reliability diagram
forecast lead time: 42h, positive temperature anomaly

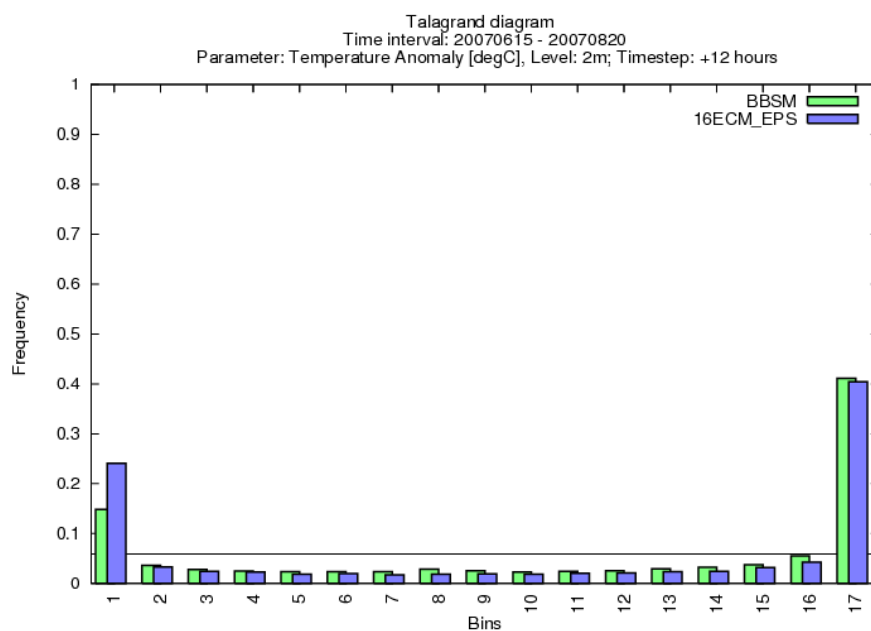


Figure 8.17: temperature anomaly - Talagrand diagram, forecast lead time: 12 hours

8.1.3 Wind Speed

Results show that ALADIN-LAEF scores better in terms of bias and RMSE (see fig. 8.18). The average bias of ALADIN-LAEF is -0.18288, whereas the average bias of the ECMWF-EPS is 0.56527. The calculation of the standard deviation based on the bias data, reveals statistical significant differences between both models. The error bars containing the interval of $\pm\sigma$ show only a few areas with overlap. Like in the case of the mean sea level surface pressure, for the surface wind speed the ECMWF-EPS shows a remarkable larger statistical spread. Since the differences between both models are statistical significant, it can be said that that ALADIN-LAEF scores definitely better in terms of bias. In contrast the differences between both models in terms of RMSE and ensemble spread are notably smaller. However ALADIN-LAEF scores better in terms of RMSE/spread.

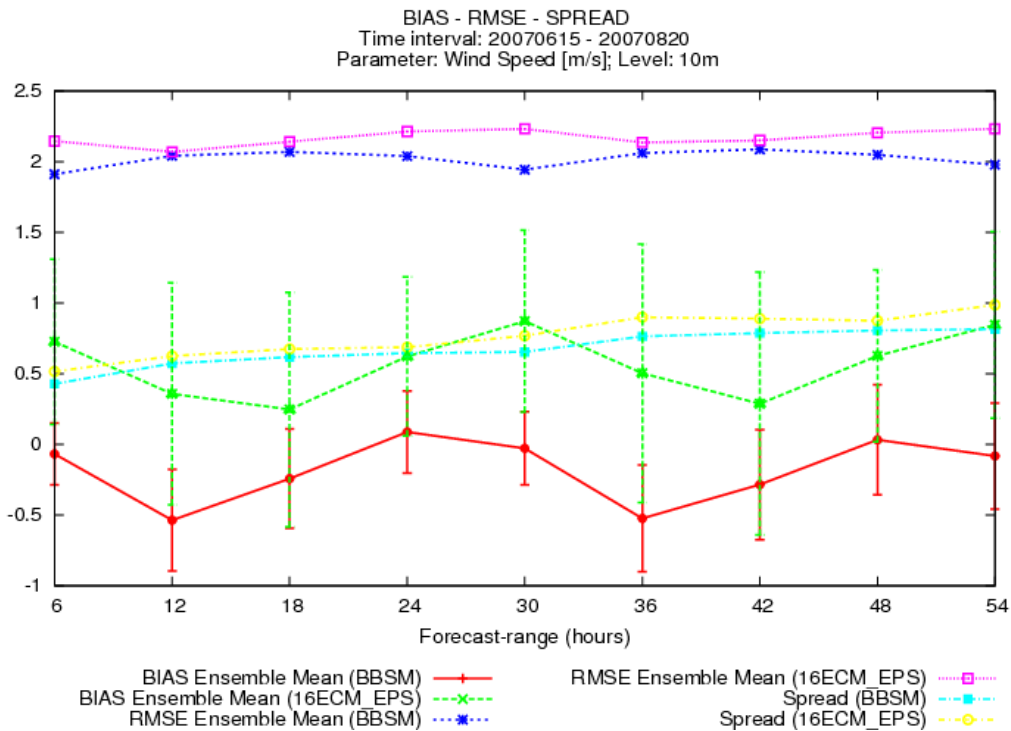


Figure 8.18: wind speed - BIAS-RMSE-Spread

The further assessment has been done using the Brier Score. Therefore 4 different thresholds, namely 1 m/s, 2 m/s, 4 m/s and 10 m/s have been evaluated. For the smallest threshold there are hardly any differences detectable between both EPS. For the threshold of 2 m/s ALADIN-LAEF scores slightly better. Moreover ALADIN-LAEF is clearly superior in regard to the prediction of wind speeds > 4 m/s and wind speed > 6 m/s (see fig. 8.20 top). Since one of the main reasons to develop ALADIN-LAEF, was to better predict high wind speeds, this result is quite encouraging. In addition to the Brier Scores the Brier Skill Scores have been computed. Again for the smallest threshold of 1 m/s there are hardly any differences detectable. For the larger thresholds ALADIN-LAEF scores clearly better than the ECMWF-EPS (see fig. 8.20 below, wind speed > 4 m/s).

Since ensemble prediction systems have been designed to account for a range of uncertainties, the ensemble spread should cover all possible outcomes. Weather forecasts

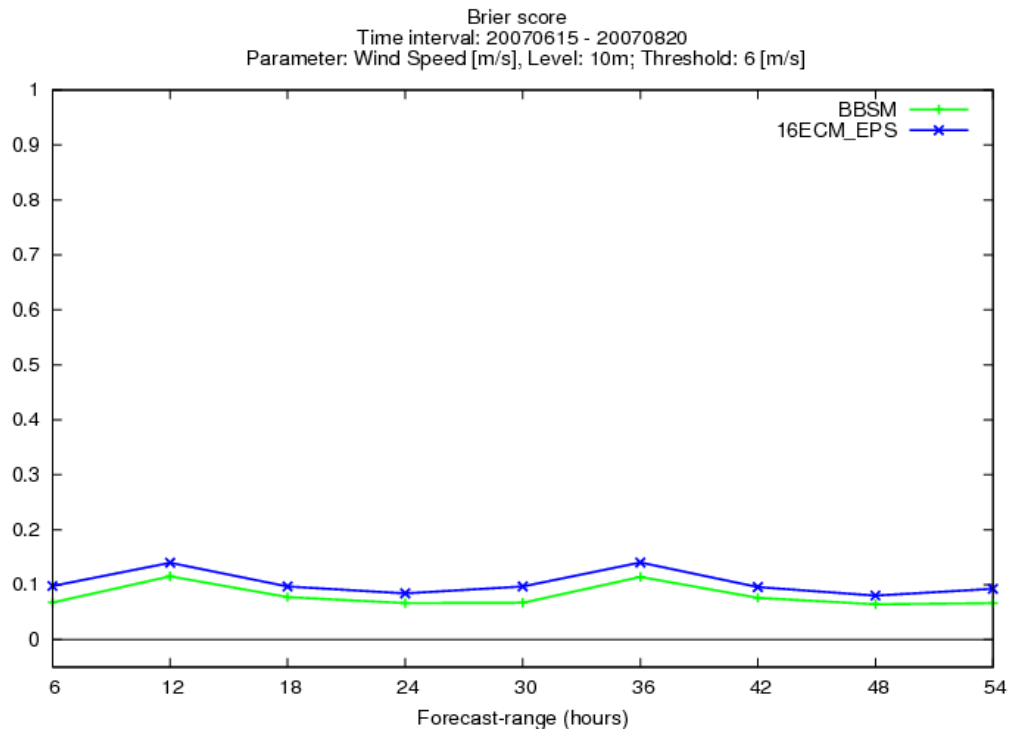


Figure 8.19: wind speed - Brier Score, threshold: 6 m/s

which are not located within this spread, are considered as outliers. Apparently the number of outliers should be as small as possible, but larger than 2/17 (see chapter 5). The calculation of the outliers reveals that ALADIN-LAEF shows a smaller number of outliers than the ECMWF-EPS (see fig. 8.21).

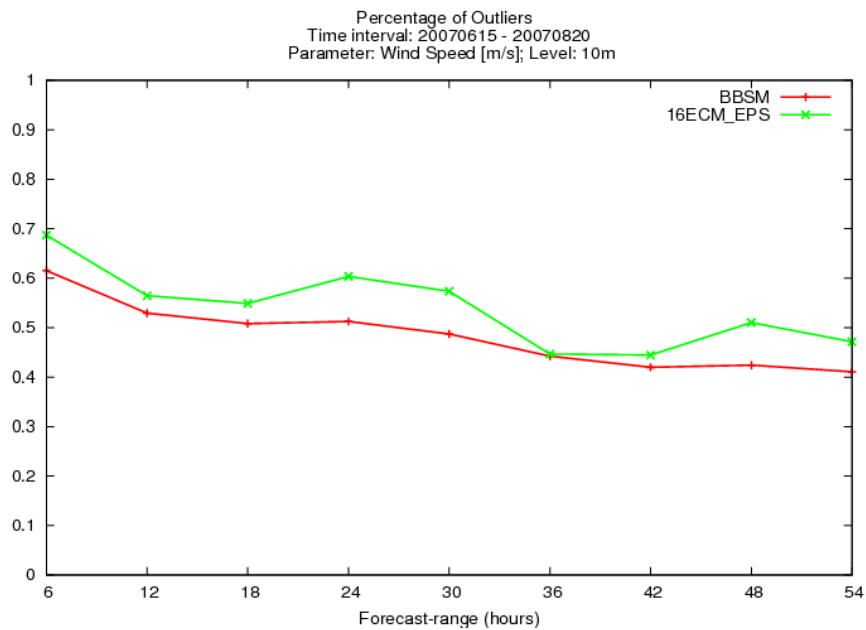


Figure 8.21: wind speed - percentage of outliers

The quality of wind speed forecasts has been further assessed in regard to reliability.

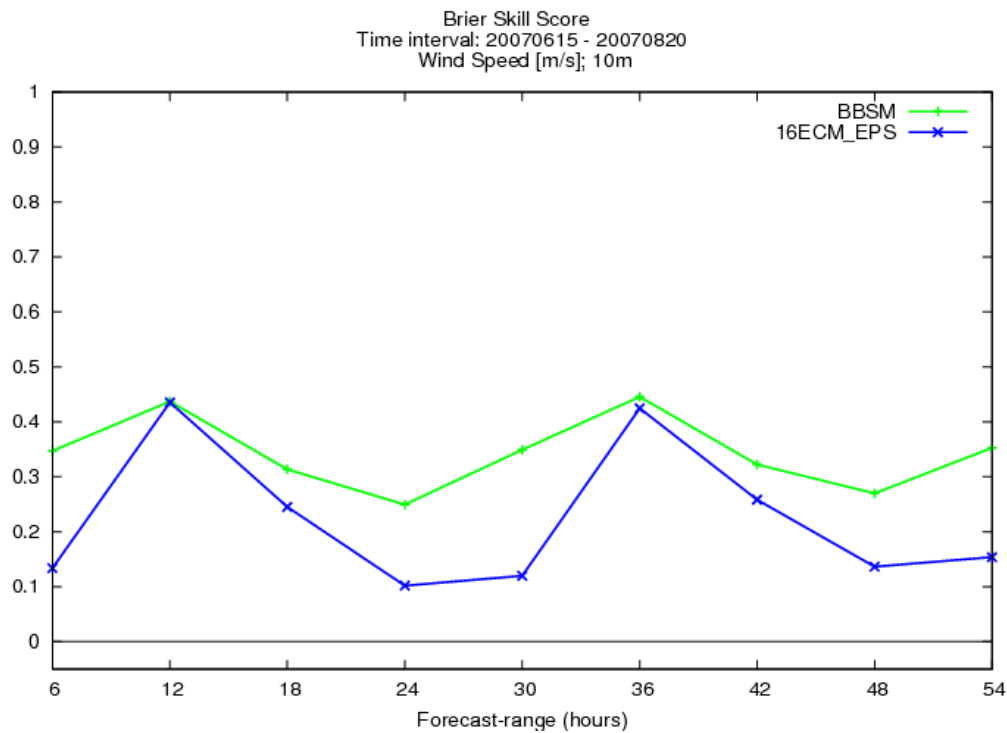


Figure 8.20: wind speed - Brier Skill Score, reference:analysis

Results for small wind speeds (>1 m/s) and all forecast lead times reveal poor resolution, indicated by a nearly horizontal curve in the centre (see fig. 8.22). This result cannot be improved by bias correction, since the bias is rather small for short forecast lead times. For example: ALADIN-LAEF, fc.time, 6h, bias= -0,06760. Apart from that there is a general tendency to underforecast lower probabilities and overforecast higher probabilities. For example the forecast probability of 0.1 for wind speed >1 m/s is linked to a observed frequency of 0.3 (see fig. 8.23). Thus, in that case the event was more often observed than being predicted. It should be mentioned that the result for the forecast lead time of 12 hours, is linked to large bias. If the bias was corrected, the resulting curve would be shifted towards the perfect line and cut it somewhere around the center. However it should be mentioned that for all forecast lead times, reliability increases with increasing wind speed (see for example fig. 8.23). This is indicated by a slight tilting of the curve towards the perfect line tilted to 45 degrees.

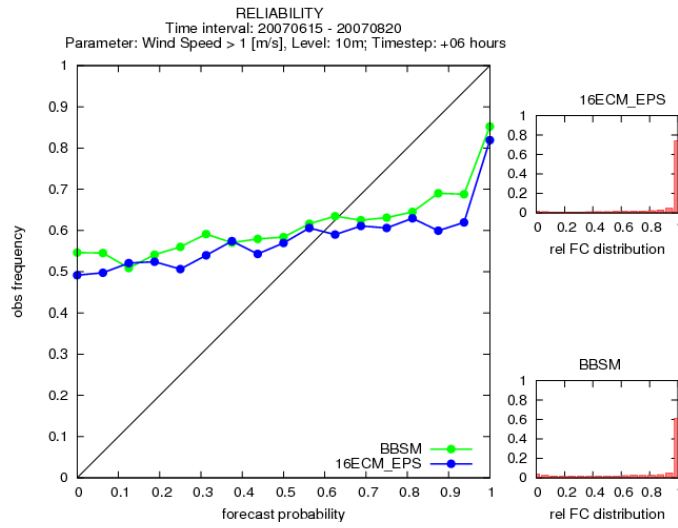


Figure 8.22: wind speed - Reliability, Poor Resolution indicated by nearly horizontal curve

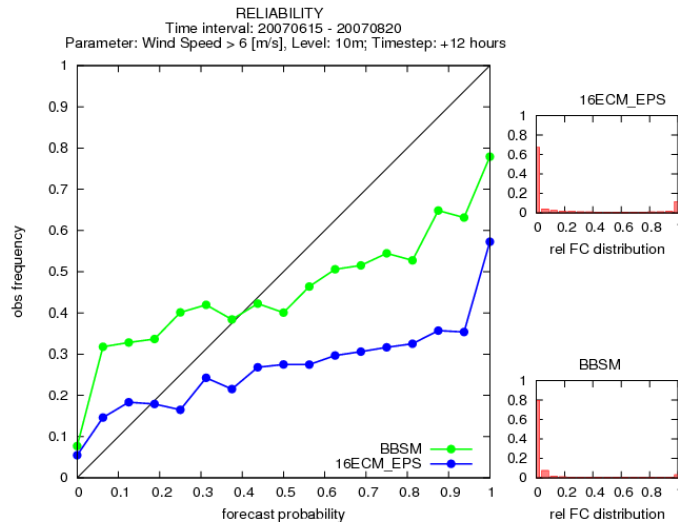


Figure 8.23: wind speed - Reliability

The comparison of all reliability diagrams shows a clear dominance of ALADIN-LAEF. ALADIN-LAEF scores better in 13 of 23 different settings. However this result is clearly an effect of the smaller bias. The reason might be better incorporation of local characteristics such as topography. It should be considered that for this verification the ECMWF-EPS was dynamically downscaled. Thus, unlike the regional model ALADIN-LAEF, the ECMWF-EPS can not account for specific local topography. Although ALADIN-LAEF reveals high reliability in regard to the prediction of wind speed, there are some further aspects to be considered. The calculation of the ROC-curves demonstrates again that in most of the cases ALADIN-LAEF scores better. However this notion is not true for high wind speed forecasts (6 m/s). For almost all forecast lead times and the threshold 6 m/s the ECMWF-EPS scores better. Wind speeds of that category are often linked to large scale events which might be better incorporated by a global system. It might have been expected that the ECMWF-EPS

scores better for larger forecast lead times and high wind speed, but it also does for shorter forecast lead times (see fig. 8.24). This is an important aspect, since correct prediction of high wind speeds within the next 12 hours can be essential.

Finally the rank distribution of both models has been analyzed. For both EPS the ensemble members tend to be too similar and underdispersion can be observed. However for all forecast lead times, ALADIN-LAEF shows a better rank distribution than the ECMWF-EPS.

The comparison of all scores and selected thresholds demonstrates a clear dominance of ALADIN-LAEF. ALADIN-LAEF scores better in 71 of 106 different configurations (see table 8.1 in short summary). The same result is obtained when comparing the 16 major scores. In 12 of 16 cases ALADIN-LAEF outperforms the ECMWF-EPS. Nevertheless some restrictions should be noted. Although ALADIN-LAEF turns out to be generally very reliable, drawbacks in regard to the prediction of high wind speed should be considered.

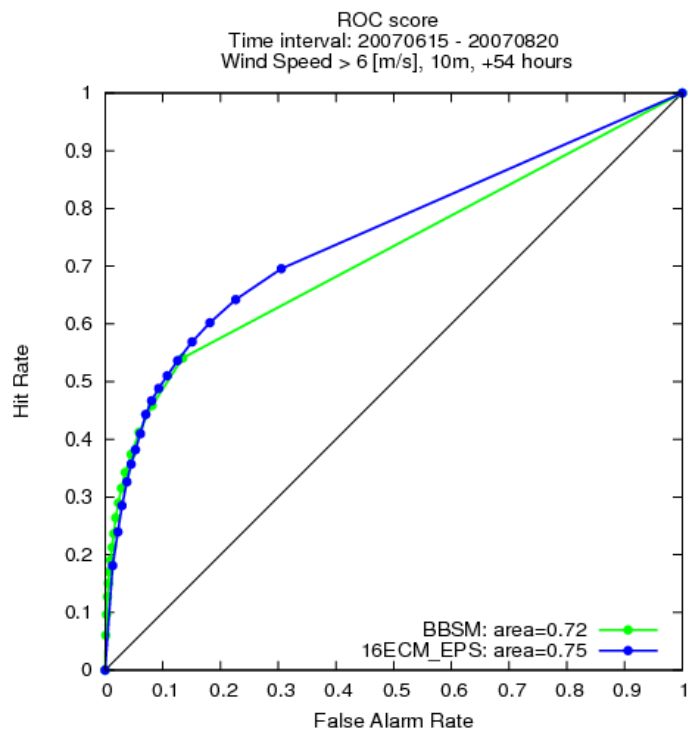
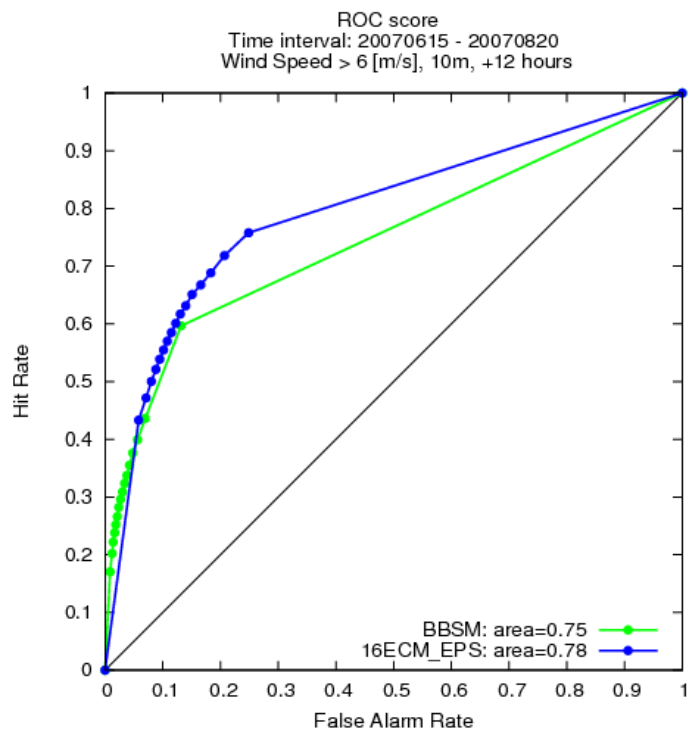


Figure 8.24: wind speed - Relative Operating Characteristic

8.1.4 Total Precipitation

As a first measure of model performance, systematic errors (bias) have been identified. For almost all forecast lead times the ECMWF-EPS shows a better performance than ALADIN-LAEF (see fig. 8.25). The average bias of the ECMWF-EPS is 0.16210, whereas the average bias of ALADIN-LAEF is 0.39149. However it should be mentioned that the differences between both models are not statistically significant. As the error bars demonstrate, the intervals containing $\pm\sigma$ overlap for all forecast lead times. For the forecast lead time of 42 hours, there appears a small region with no overlap, but it is extremely small in comparison to the complete interval. Therefore differences between both models should be considered as small and statistically not significant. Measuring the performance in terms of RMSE, both models show very similar results which are ranging between 4 and 4.5 mm/12h. Results based on the ratio RMSE/spread show that ALADIN-LAEF is closer to 1 than the ECMWF-EPS.

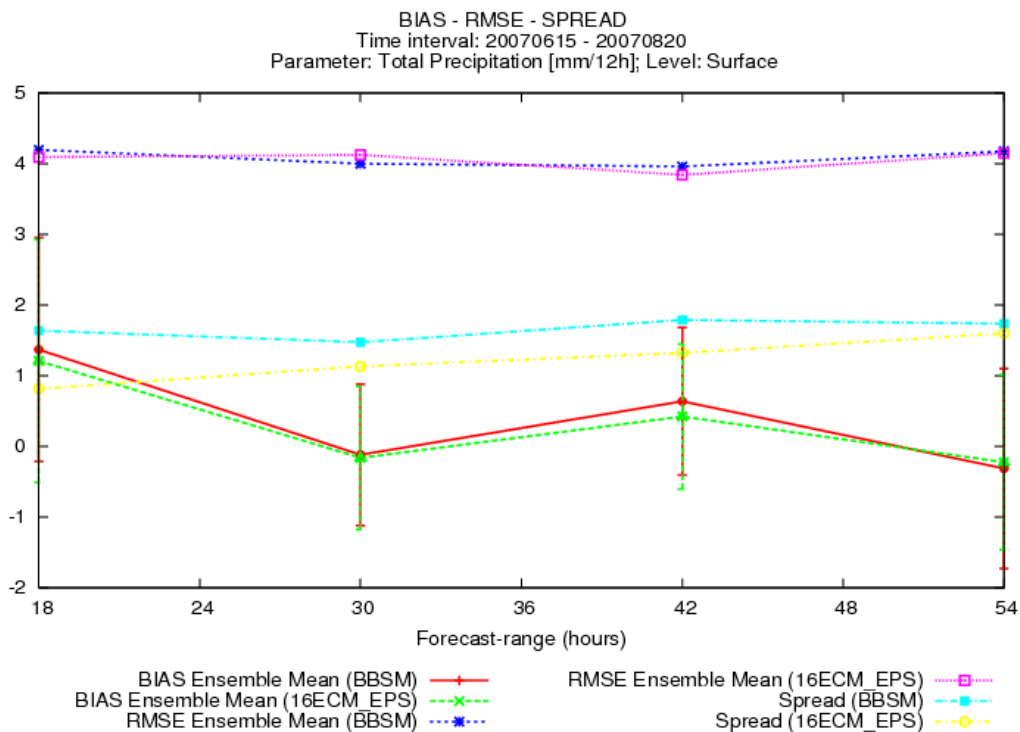


Figure 8.25: total precipitation - BIAS-SPREAD-RMSE

When the performance of both models is assessed in terms of Brier Score, the smallest threshold values (1 mm) reveals the greatest discrepancies (see fig. 8.26). For this threshold value, ALADIN-LAEF scores better for shorter forecast lead times (18h-30h), whereas the ECMWF-EPS is superior for larger forecast lead times. For larger thresholds (2 mm, 5 mm and 10 mm) there are only very small differences detectable. For the largest threshold of 25 mm, both models show almost similar performance (see fig. 8.27).

Results based on the Ranked Probability Score and the Continuous Ranked Probability Score reveal again that the largest differences between the two models are present for smaller forecast lead times (18h-36h). ALADIN-LAEF scores better for these forecast lead times. The reason might be again the increased horizontal resolution that enables a better allocation of rainfall. For larger forecast lead times, the differences between

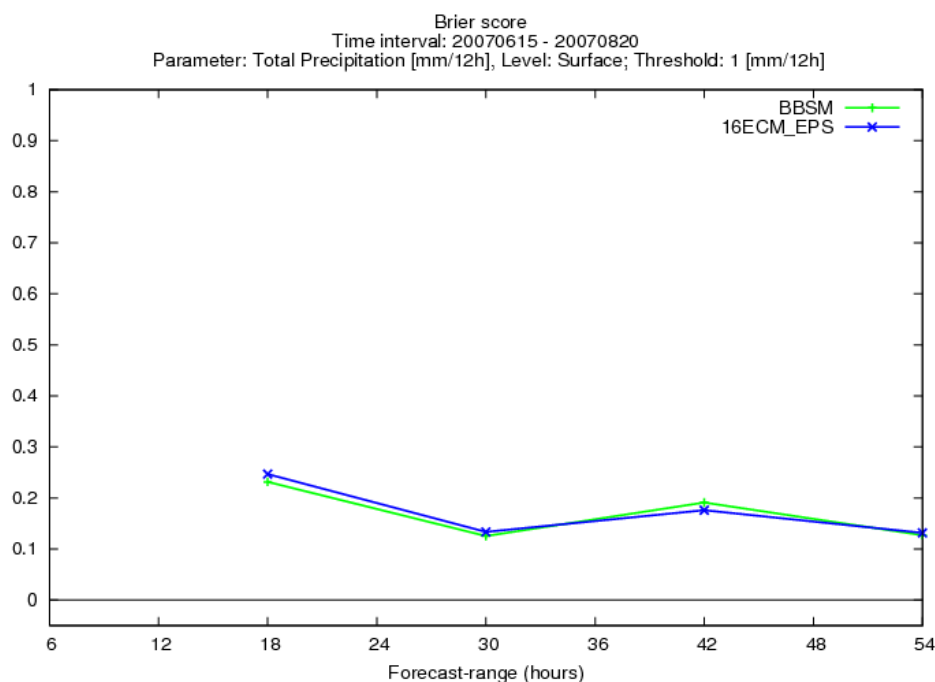


Figure 8.26: total precipitation - Brier Score: threshold 1 mm

the two models get smaller (see fig. 8.28). The assessment of the performance based on the percentage of outliers shows that ALADIN-LAEF is superior. Moreover results point out again that the uncertainty for shorter forecast lead times is larger. Hence, both models have more outliers for short forecast lead times (see fig. 8.29).

The fact that both EPS show a higher level of uncertainty for shorter forecast lead times can be also seen in the reliability diagrams. For example for the forecast lead time of 18 hours, both models are positively biased. The corresponding reliability diagram differs remarkably from the perfect curve and shows a strong tendency to overforecast high wind speeds (see fig. 8.30). Results show that reliability increases with increasing forecast lead times (see fig. 8.31). This finding might seem surprising, since short time forecasts are usually considered to be more reliable. The reason behind that is the ensemble spread which is too small for short forecast lead times. This can be also seen in the outliers diagram (fig. 8.29). For events where hardly any heavy rainfall is expected (probability ranging from 0.0 to 0.2) the reliability is very good. For heavy rainfall (threshold 10 mm/12h) there is a strong tendency to overforecast that event. This is an often observed characteristic of NWP models. For example in ALADIN-LAEF the predicted probability of heavy rain corresponds to 0.9, whereas the observed frequency is only about 0.6. However it should be noted that the number of heavy rainfall events (> 10mm) for the investigated period was rather small (see climatological data in fig.8.32, fig.8.33 and fig. 8.34).

Since the ability to discriminate among heavy and light rainfall events can be essential, both models have been evaluated in regard to ROC. Results reveal that ALADIN-LAEF scores better in 12 of 20 different settings. Again the largest differences between both models are present for shorter forecast lead times, e.g. 18 hours (see fig.8.35 and fig.8.36). This result is also reflected in the Talagrand diagram, where shorter forecast lead times show underdispersion (see fig.8.37).

The comparison of all scores and selected thresholds reveals a clearly better perfor-

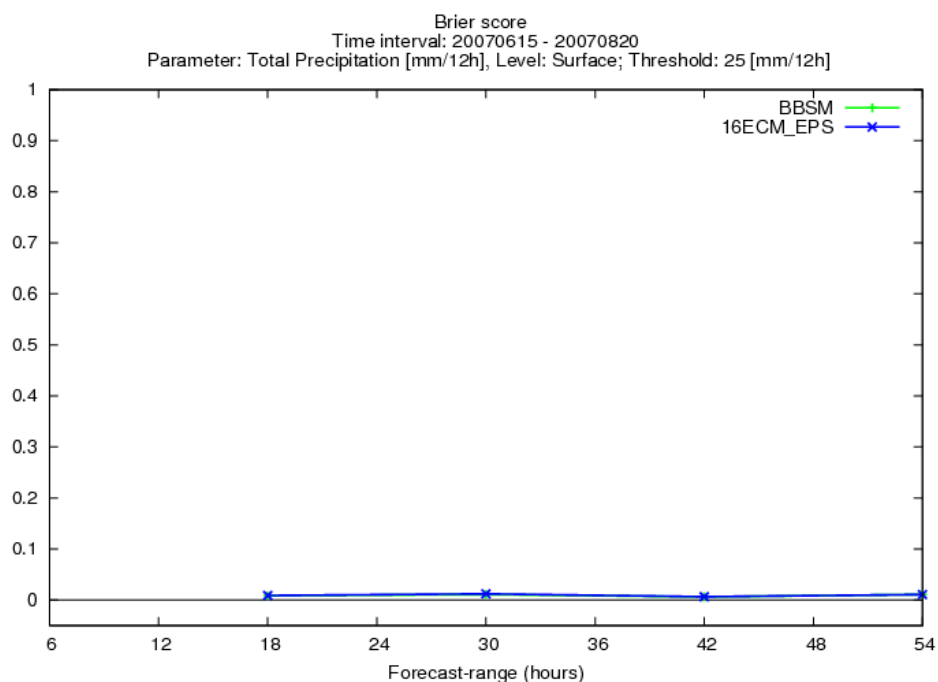


Figure 8.27: total precipitation - Brier Score: threshold 25 mm

mance of ALADIN-LAEF. In 54 of 69 different settings the model is superior (see table 8.1). When the comparison is based on 16 major scores, ALADIN-LAEF performs better in 13 of 16 different cases. The better performance is particularly linked to shorter forecast lead times. This might be an effect of the increased horizontal resolution of ALADIN-LAEF. Since ALADIN-LAEF has been particularly designed to better predict small scale weather events such as orographically induced heavy rainfall, these results appear quite encouraging.

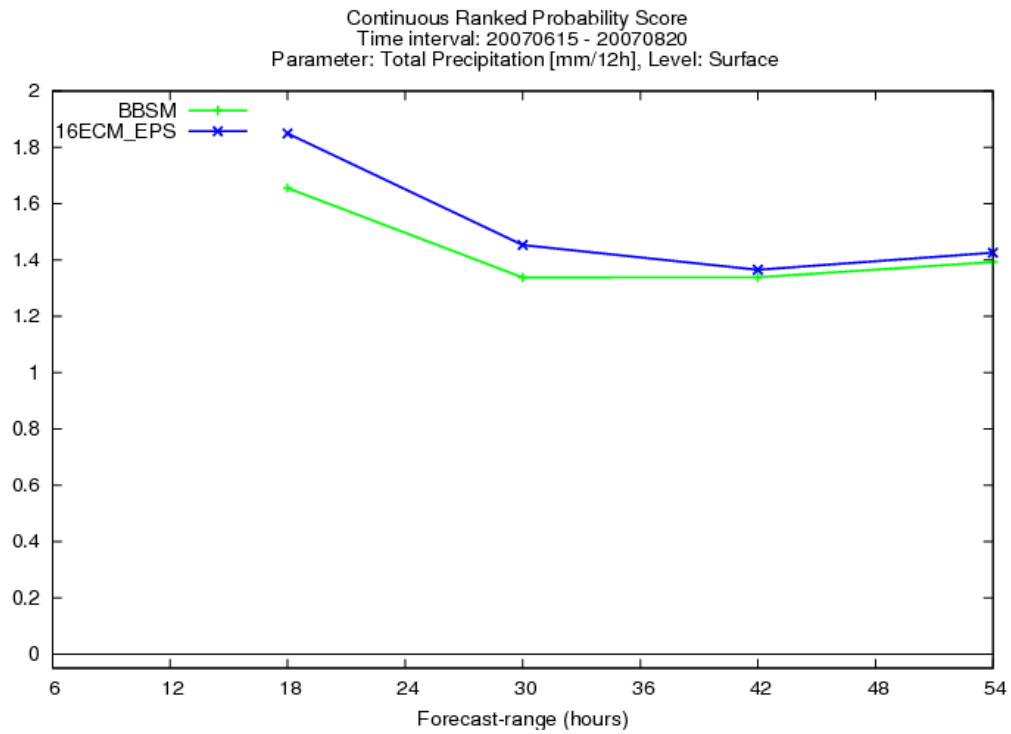


Figure 8.28: total precipitation - Continuous Ranked Probability Score

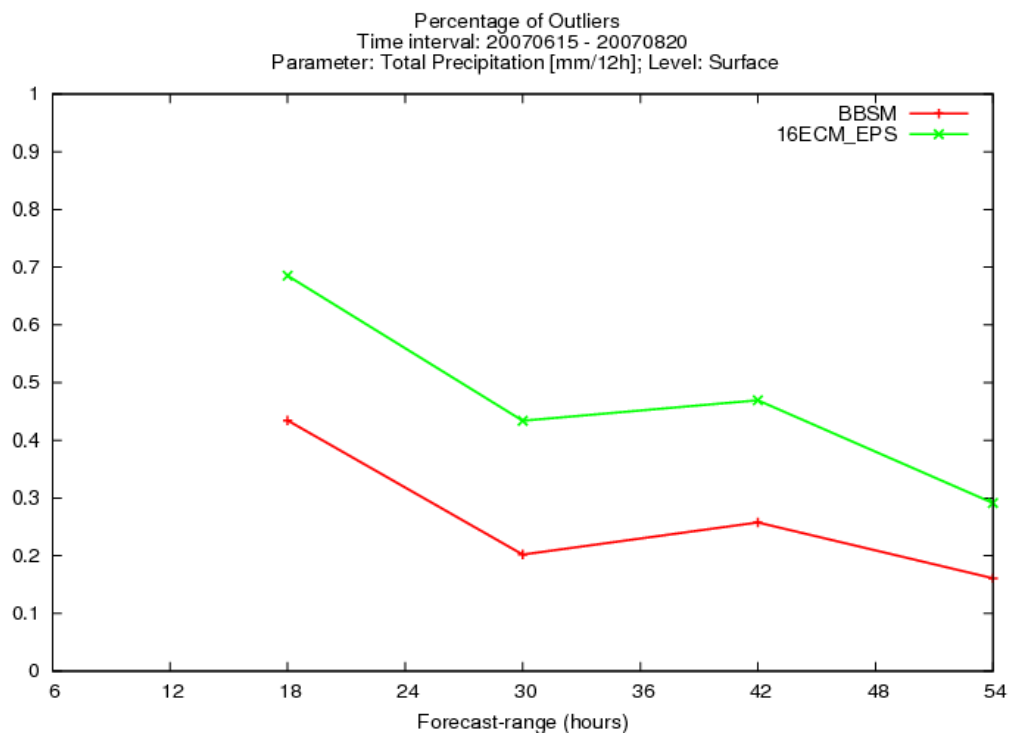


Figure 8.29: total precipitation - Percentage of Outliers

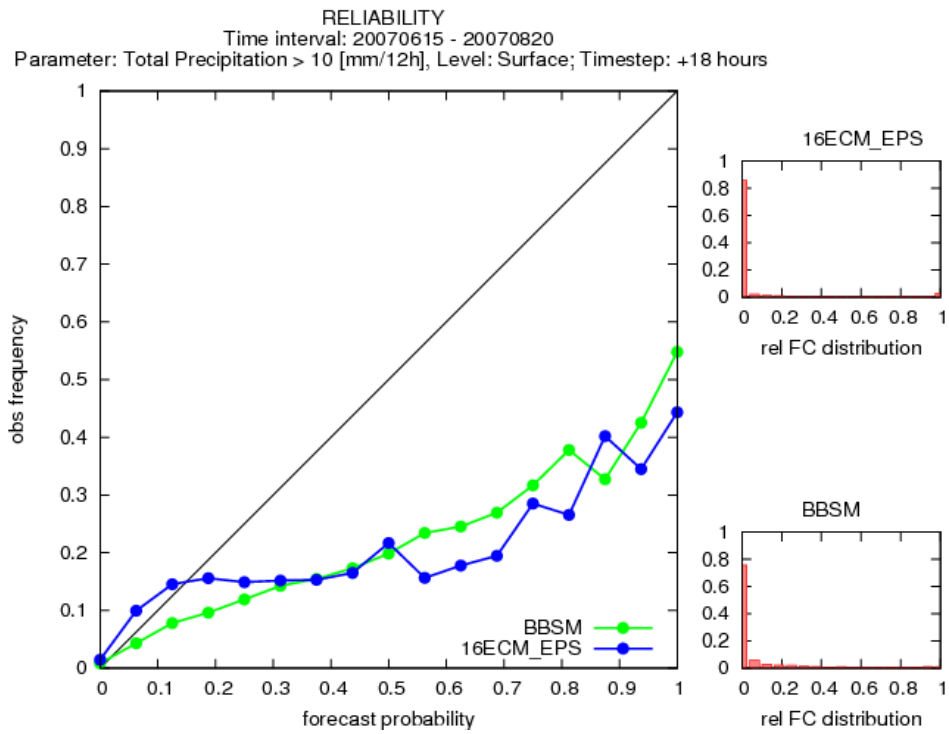


Figure 8.30: total precipitation - Reliability diagram: fc.time 18h, thr. 10 mm

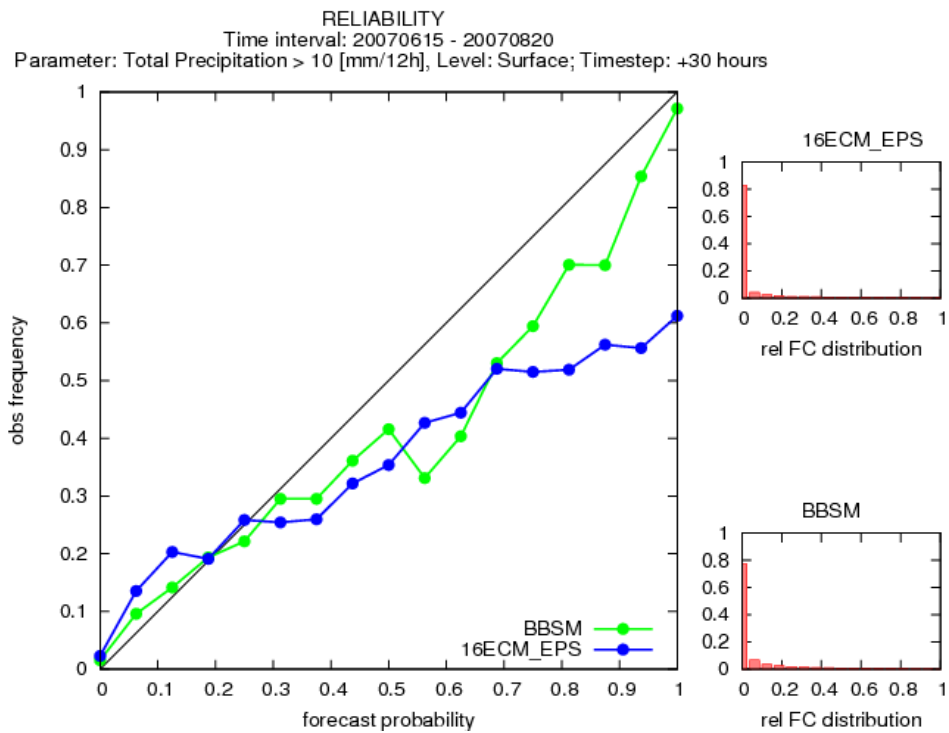


Figure 8.31: total precipitation - Reliability diagram: fc.time 30h, thr. 10 mm

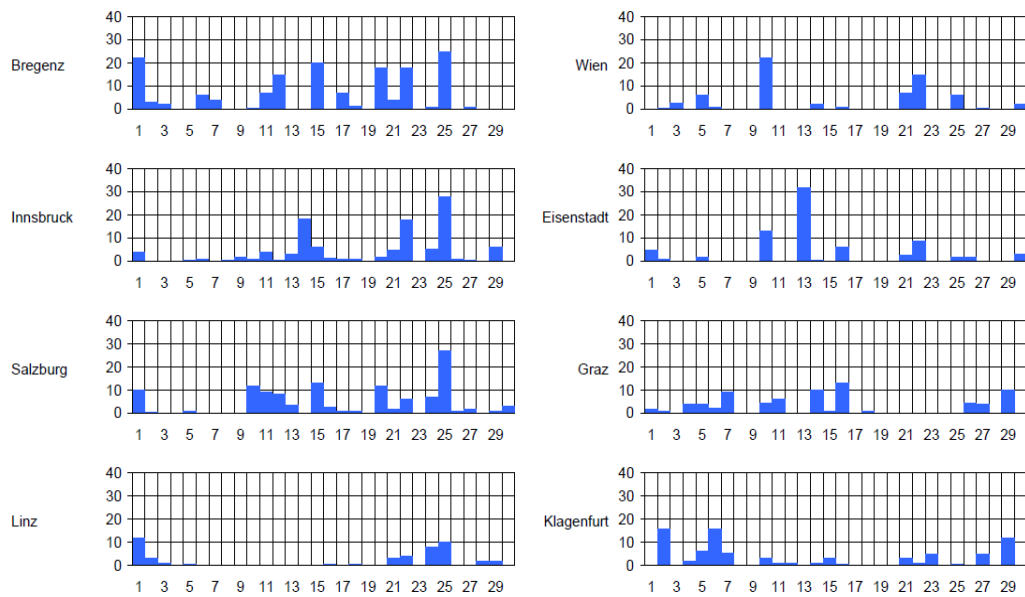


Figure 8.32: total precipitation - Total Precipitation for selected stations
June 2007, horizontal axis: days, vertical axis: mm/24h

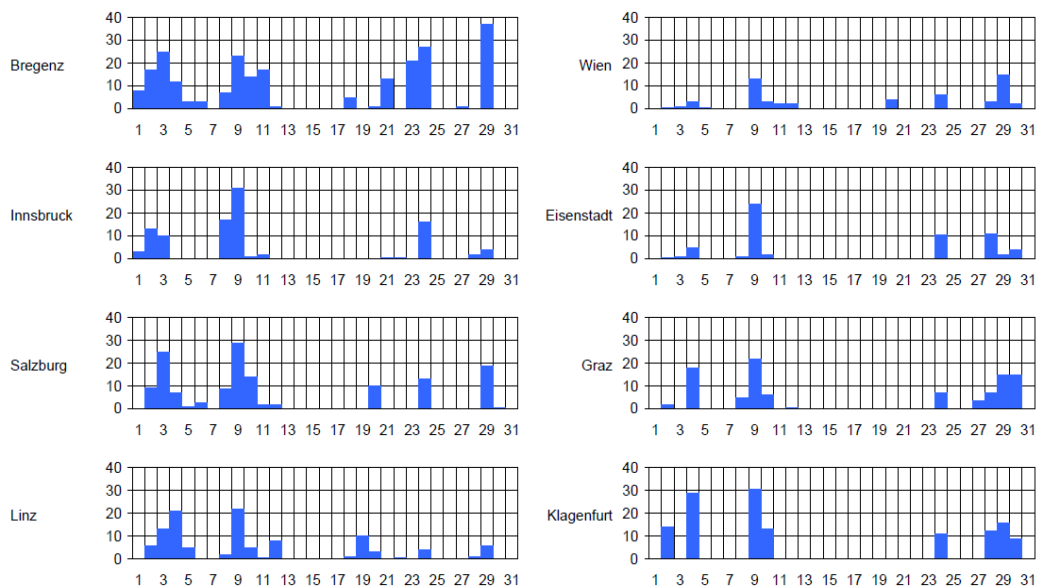


Figure 8.33: total precipitation - Total Precipitation for selected stations
July 2007, horizontal axis: days, vertical axis: mm/24h

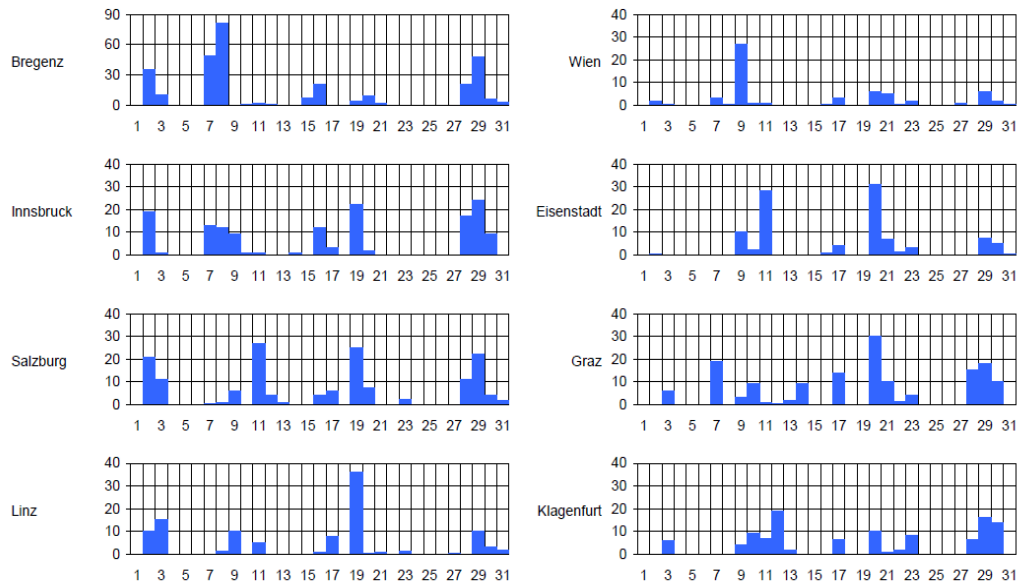


Figure 8.34: total precipitation - Total Precipitation for selected stations
August 2007, horizontal axis: days, vertical axis: mm/24h

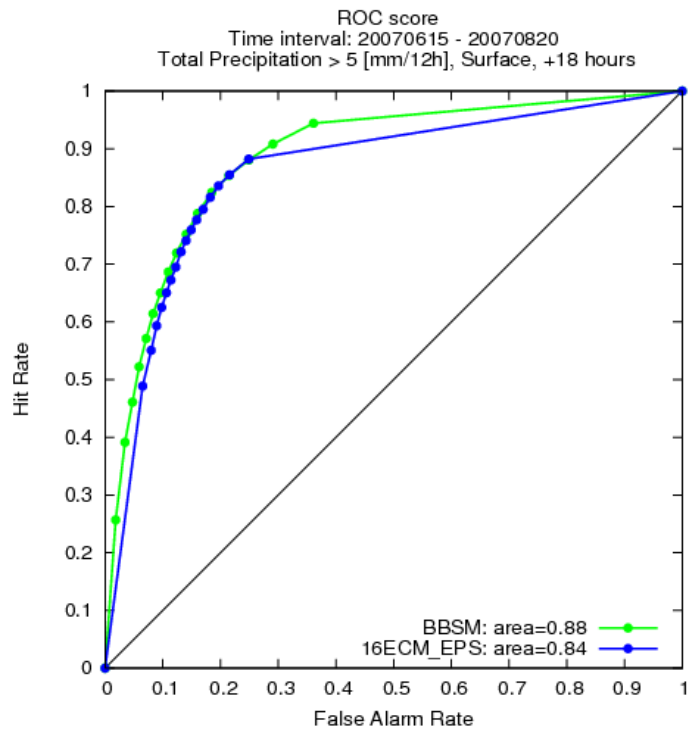


Figure 8.35: total precipitation - ROC: fc.time 18h, thr. 5 mm

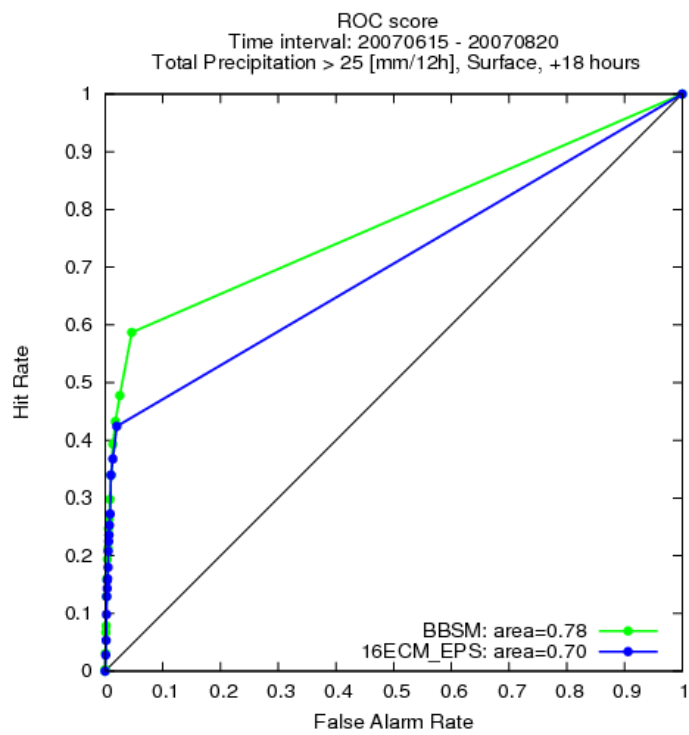


Figure 8.36: total precipitation - ROC: fc.time 18h, thr. 25 mm

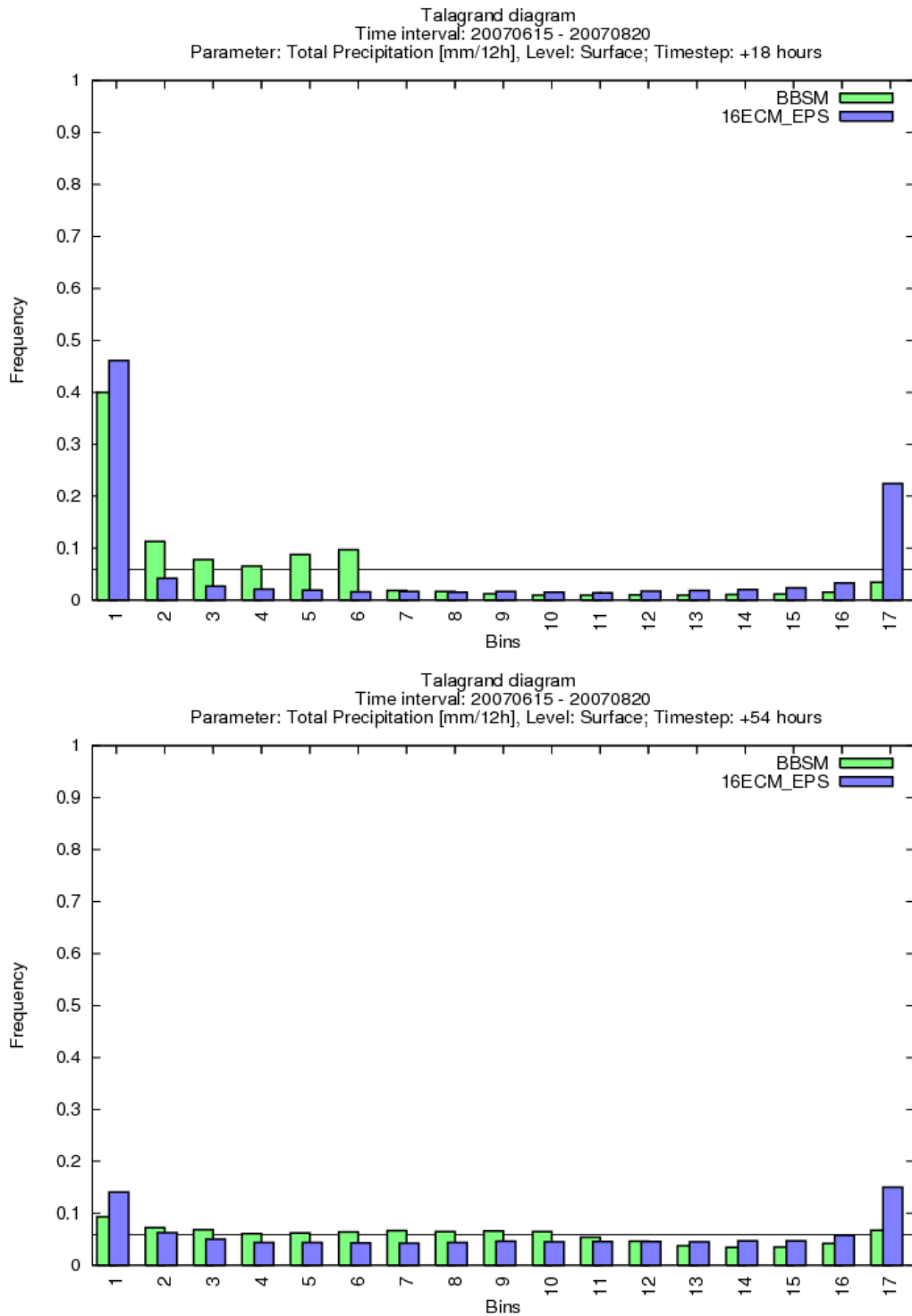


Figure 8.37: total precipitation - Talagrand diagram: 18h and 54h forecasts

8.1.5 Short Summary

The final comprehensive comparison of ALADIN-LAEF and the ECMWF-EPS has been based on all different configurations. ALADIN-LAEF has proved to be superior in regard to the prediction of wind speed and precipitation. In contrast the ECMWF-EPS turned out to be predominant for pressure and temperature forecasts (see table 8.1). Nevertheless this general performance should not be used as a single directive, since results differ for selected thresholds and forecast lead times. The chapter Appendix contains a detailed list of the results as well as an explanation of this statistic.

Results based on 16 major scores		
Parameter	ALADIN-LAEF	ECMWF-EPS
Mean Sea Level Pressure	9	7
Temperature Anomaly	2	14
Wind Speed	12	4
Total Precipitation	13	3
Total	36	28
Results based on individual thresholds		
Parameter	ALADIN-LAEF	ECMWF-EPS
Mean Sea Level Pressure	33	47
Temperature Anomaly	27	55
Wind Speed	71	35
Total Precipitation	54	15
Total	185	152

Table 8.1: Final Results - Surface Parameters

8.2 Verification of upper level fields

Large scale atmospheric motions govern the formation of weather systems such as troughs and cold fronts. The driving force behind these large scale motions is the pursuit to keep balance. If the atmospheric conditions are unbalanced, further development can be expected. For example if the upper level pressure field differs from the lower level pressure field, the formation of weather regimes can be observed. Apparently weather forecasts particularly depend on the quality of these large scale parameters. For the verification of the upper level field 4 selected parameters have been analyzed. These parameters are geopotential, temperature anomaly, wind speed and relative humidity. For each parameter the forecasts for the 500 hPa-level and the 850hPa-level have been evaluated. For the verification the ECMWF analysis serves as a reference.

The large scale flow is governed by the geopotential in 500 hPa. This flow determines the formation of upper level troughs, cut off lows and upper level ridges. These systems have a major impact on the formation of frontal systems, whereas the position of the upper level field in relation to the surface field is crucial. Based on the importance of the large scale flow, the 500 hPa geopotential forecasts of both ensemble prediction systems have been evaluated. Since the large scale flow directs the transport of air masses, it has an crucial impact on the development of cold air outbreaks or episodes of heat. Therefore the temperature forecasts for the 500 hPa have been evaluated. Based on the wind speed forecasts for the 500hPa level, the further movement of the jet stream can be predicted. The jet stream has a crucial impact on the development of frontal systems and can lead to frontal intensification as well as frontal decay. Therefore it is important to monitor the quality of the wind speed forecasts. Beyond that humidity forecasts have been verified. In an effort to keep consistent with the levels of the other parameters, the humidity profile in 500 hPa has been analyzed. However it should be noted that 500 hPa is rather high for humidity analysis and these humidity predictions can only be used for certain applications. For example the humidity profile in 500 hPa gives a good indication where the jet stream can be found. Inside the jet stream a rotating motion takes place, whereas the sinking motion at the rear edge leads to a drying process. This can be seen in the humidity field. As already mentioned this information is very valuable, since the jet stream has the power to intensify frontal systems. The geopotential forecasts for the 850 hPa level form the basis for several different applications. For example the retrieval of the snow line depends on the geopotential in 850 hPa and the corresponding equivalent potential temperature. Moreover the geopotential in 850 hPa is particularly useful in mountainous areas. For example forecasts of phenomena such as foehn rely on that information. Temperature forecasts for the 850 hPa level are particularly helpful for stability analysis. The calculation of many stability indices such as the Showalter index involves the temperatures in the 850 hPa level. Based on these temperatures further assessments in regard to convection can be done. These examples indicate that the quality of temperature forecasts for the 850 Pa level is important. Wind speed forecasts for the 850 hPa level are essential in mountainous areas. The wind speed information serves as basis to predict phenomena such as foehn. Moreover the correct prediction of high wind speed in mountainous areas is essential for hiking or climbing. Since large parts of Austria are located in mountainous regions, the wind speed forecasts for the 850 hPa level are of great interest. One important aspect of weather forecasts is the correct prediction of precipitation. Therefore it is not only important to locate frontal systems, but also to assess the corresponding amount of precipitation. For this purpose humidity information in 850hPa is needed.

Based on that information further assumptions can be taken. However it should be noted, that the correct prediction of precipitation is extremely challenging, since it involves a broad range of uncertainty. For example rainfall forecasts rely on the correct assessment of wind speed as well as frontal movements. In the following subchapters each of the briefly described parameter will be addressed individually for selected thresholds and forecast lead times. Finally a comprehensive overview will highlight general tendencies and the most important features.

8.2.1 Geopotential in 500 hPa-level

Results of the verification indicate that the ECMWF-EPS scores better in terms of bias and RMSE (see fig.8.38). The average bias of the ECMWF-EPS corresponds to 15.01890, whereas the average bias of ALADIN-LAEF is 29.41804. These values might appear surprisingly high compared to the results of other parameters. However it should be considered that the involved Geopotential corresponds to values with a magnitude of 1000 (10 geopotential decameter). When the performance of both models is assessed in terms of RMSE/spread, ALADIN-LAEF reveals a better performance.

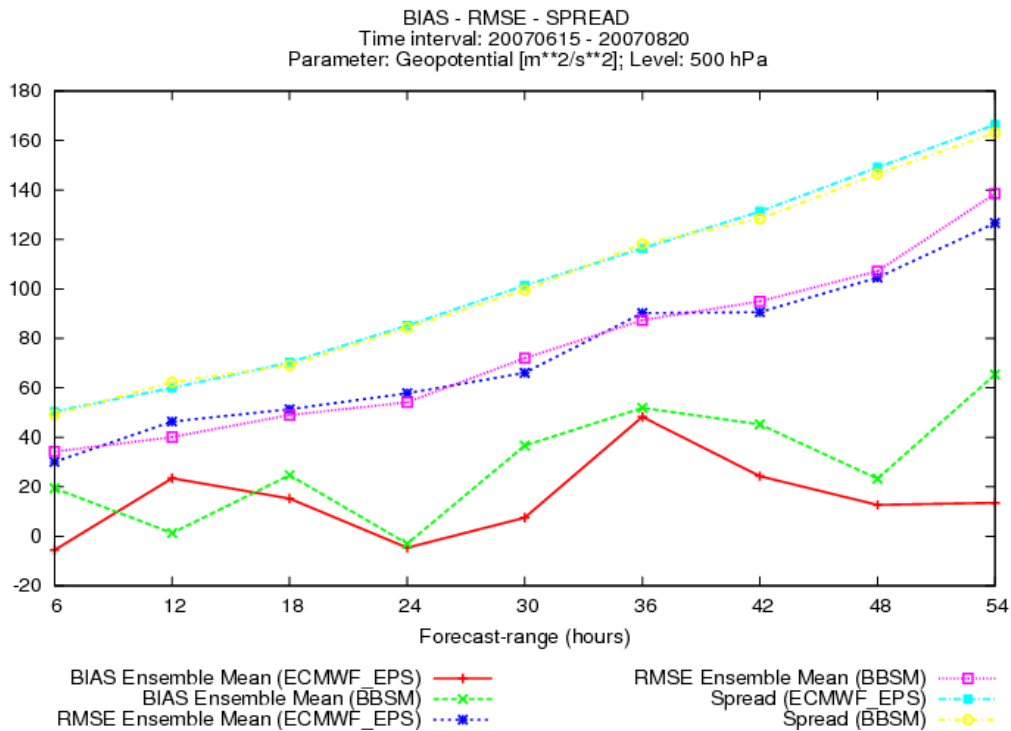


Figure 8.38: Geopotential in 500hPa - Bias, RMSE and Ensemble Spread

Results based on the calculation of outliers, show a better performance of the ECMWF-EPS (see fig. 8.39). The average percentage of outliers (here given as value between zero and 1) of the ECMWF-EPS is 0.02182. Thus, about 2 percent of the forecasts generated by the ECMWF-EPS are located outside of the ensemble spread. In contrast about 4.8 percent of the forecasts generated by ALADIN-LAEF are located outside of the ensemble spread. This result can be linked to the larger bias of the ALADIN-LAEF forecasts.

For the sake of completeness it should be noted that no reliability diagrams and ROC-areas have been calculated for the geopotential. This bases on the fact, that the calculation of these scores requires an appropriate threshold. Therefore a climatologi-

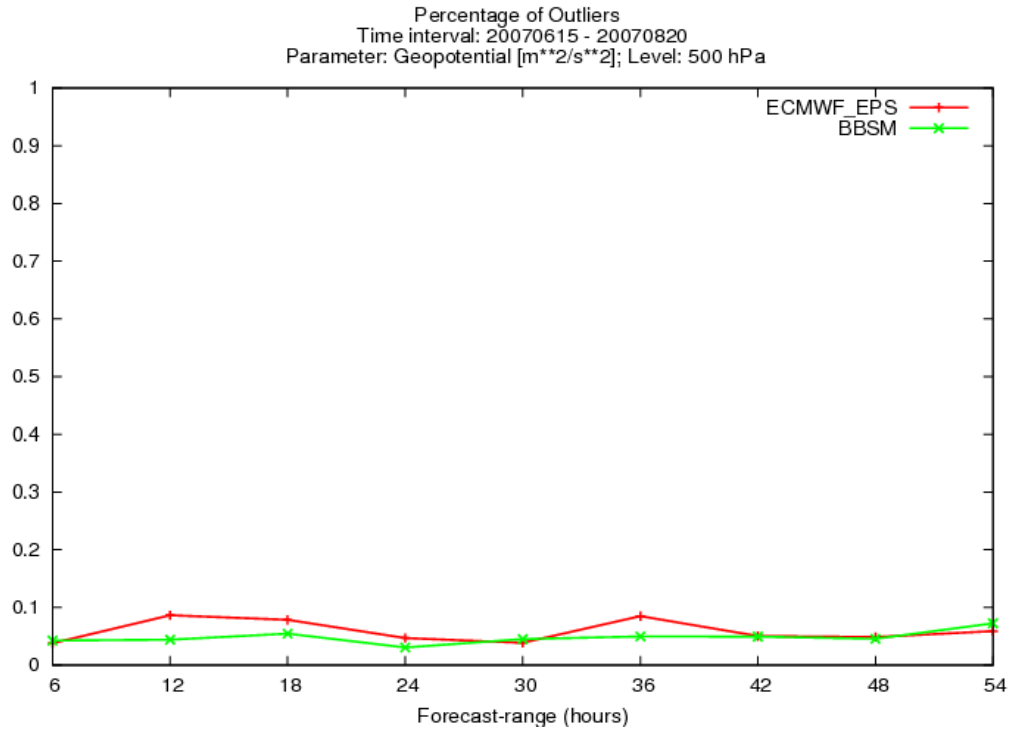


Figure 8.39: Geopotential in 500hPa - Outliers

cal analysis would be needed to set up an appropriate threshold. In contrast to all the other parameters, no climatology analysis was available. The calculation of the rank distribution shows overdispersion for both models and all forecast lead times. This results stands in contrast to most of the other parameters (for example temperature anomaly), where underdispersion can be observed. In the Talagrand diagram of the 500hPa geopotential high concentrations of ranks in the center can be found (see fig. 8.40). This fact indicates that the ensemble spread is too large. This notion is also confirmed by the small number of outliers of both models. Although the number of outliers should be as small as possible, it should not fall below 2/17. The comparison of all forecast lead times shows that ALADIN-LAEF produces better rank distributions. The best result turns up for the forecast lead time of 18 hours (ALADIN-LAEF), whereas the worst rank distribution corresponds to a forecast lead time of 6 hours (ECMWF-EPS) (see fig. 8.40).

The final comparison of all scores reveals an overall better performance of the ECMWF-EPS. ECMWF-EPS scores better in 6 of 8 different settings (see table 8.2). It should be noted that no reliability- and ROC-data were available.

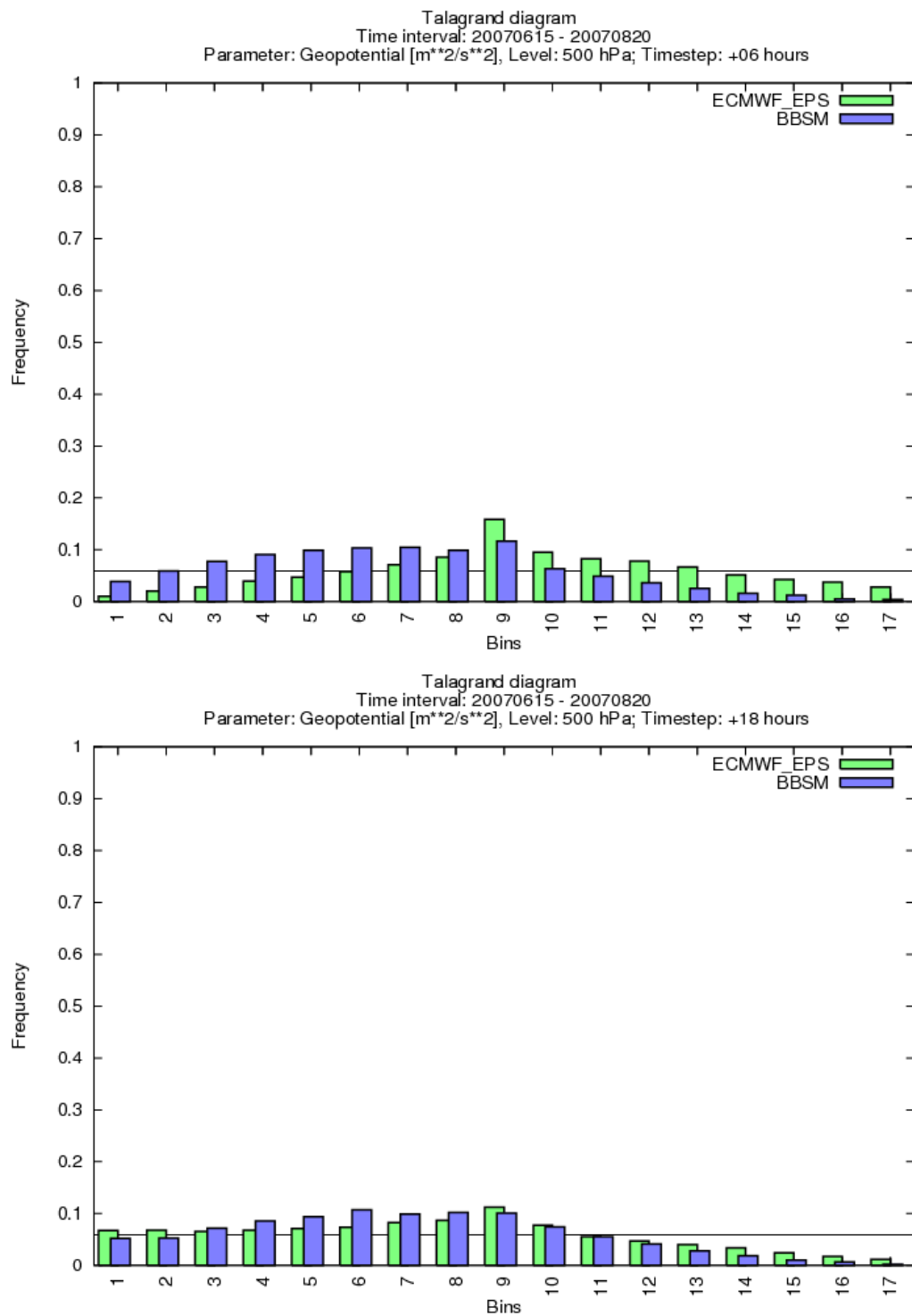


Figure 8.40: Geopotential in 500hPa
Talagrand diagram, forecast lead times: 6h and 18h

8.2.2 Geopotential in 850 hPa-level

As a first measure of model performance the bias has been calculated. As results demonstrates, both EPS differ remarkably in terms of bias. The bias of ALADIN-LAEF shows a slight oscillation around the perfect score which is zero (see fig. 8.41). In contrast the ECMWF-EPS constantly predicts too low values for the geopotential in 850hPa. Thus, the bias is negatively orientated and beyond that also remarkably larger than the bias of ALADIN-LAEF. The average bias of ALADIN LAEF is -2.28350, whereas the average bias of the ECMWF-EPS is -28.26636. In the bias diagram a maximum deviation of about $-50 \text{ m}^2/\text{s}^2$ can be detected. ALADIN-LAEF also scores better in terms of RMSE. In opposition to bias and RMSE, both models hardly differ in terms of spread. However the ratio RMSE/spread of the ECMWF-EPS is notably closer to the perfect value of 1 than that of ALADIN-LAEF.

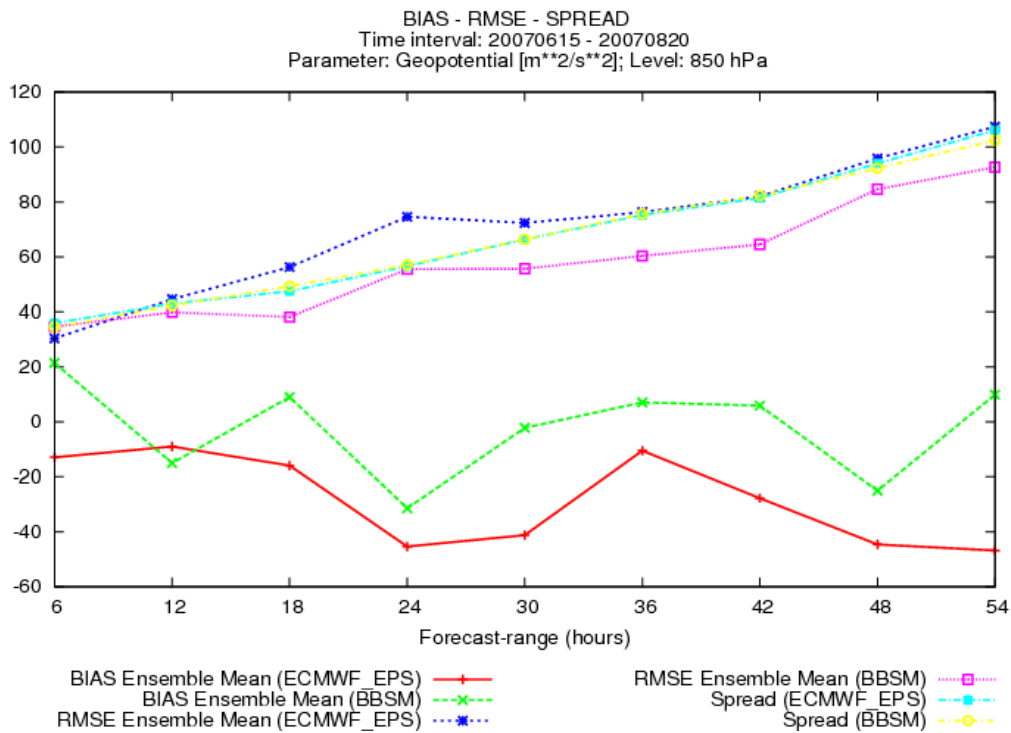


Figure 8.41: Geopotential in 850hPa - Bias, RMSE and ensemble spread

Results for the Continuous Ranked Probability Score reveal a superior performance of ALADIN-LAEF. ALADIN-LAEF scores better for forecast lead times ranging from 12 to 54 hours, whereas the ECMWF-EPS scores a little better for a forecast lead time of 6 hours (see fig. 8.42).

When the Talagrand diagrams are compared for all forecast lead times, the ECMWF-EPS performs better for shorter forecast lead times, whereas ALADIN-LAEF performs better for longer forecast lead times (compare fig. 8.43 and fig. 8.44). The final comparison of all scores indicates a better performance of ALADIN-LAEF. ALADIN-LAEF performs better in 7 of 8 different configurations (see table 8.2). It should be noted that no reliability- and ROC-data were available.

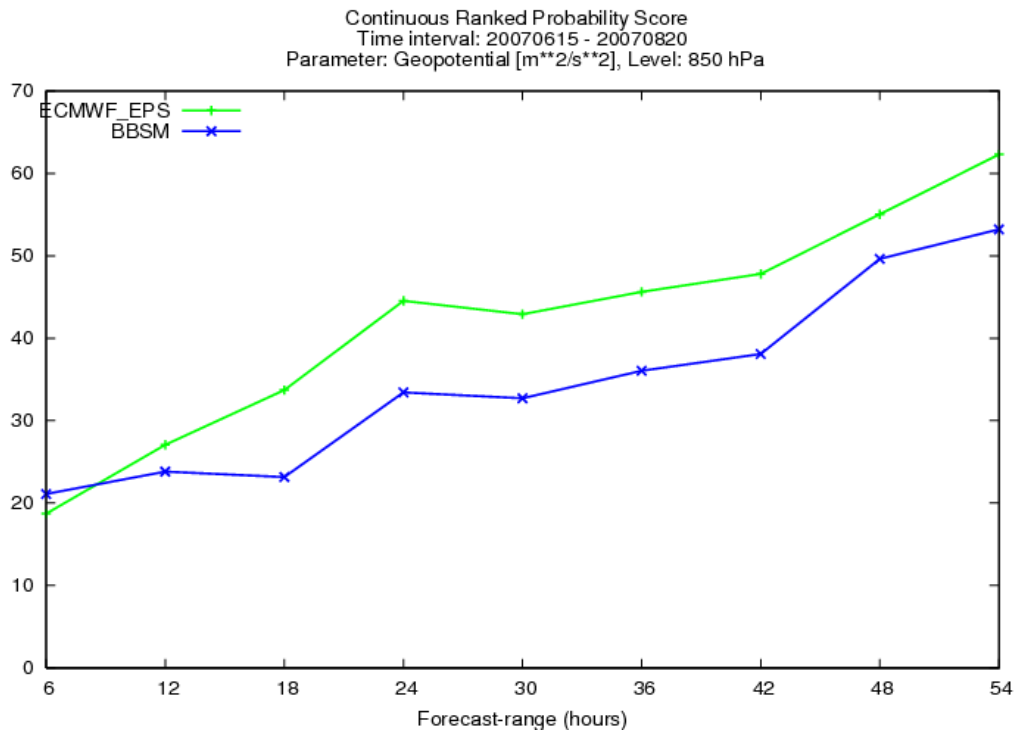


Figure 8.42: Geopotential in 850hPa - Continuous Ranked Probability Score

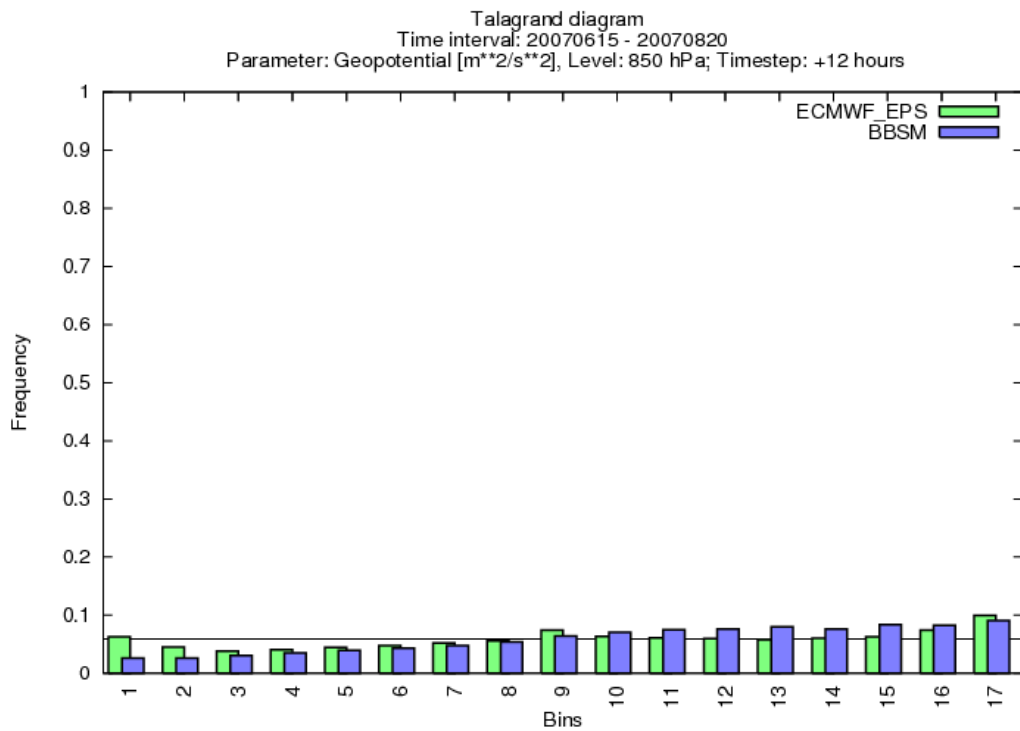


Figure 8.43: Geopotential in 850hPa - Talagrand diagram - 12h forecasts

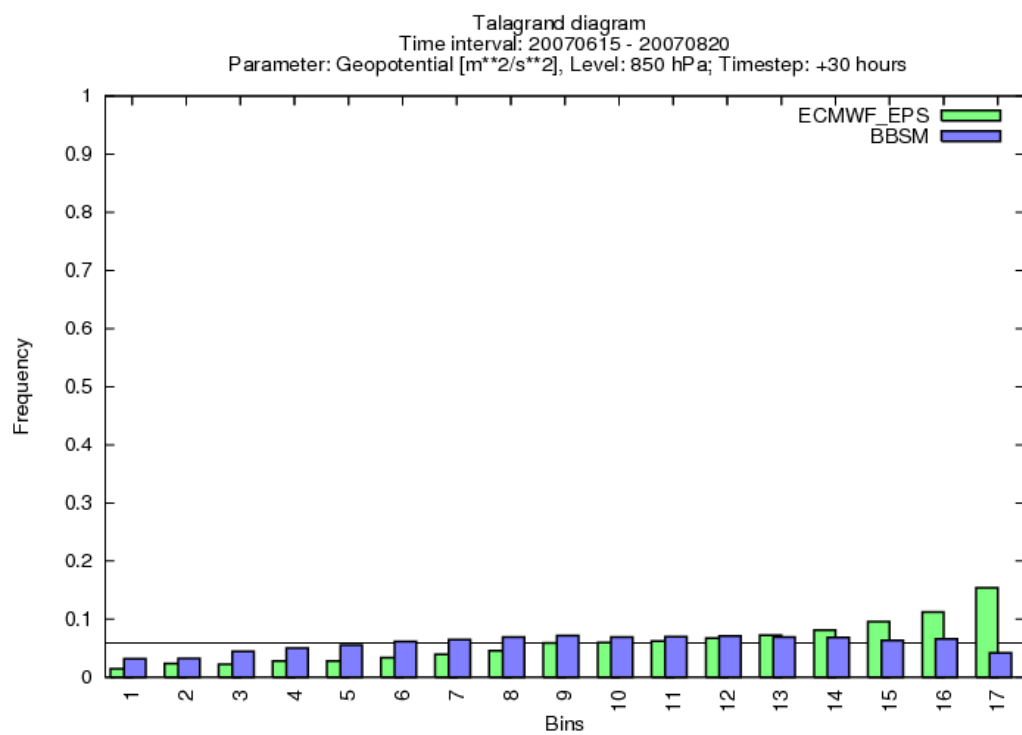


Figure 8.44: Geopotential in 850hPa - Talagrand diagram - 30h forecasts

8.2.3 Temperature Anomaly in 500 hPa

The evaluation of the temperature forecasts for the 500hPa level reveals, that ALADIN-LAEF exhibits a clearly smaller bias than the ECMWF-EPS for all forecast lead times. Results show that the average bias of ALADIN-LAEF is 0.14022, whereas the average bias of the ECMWF-EPS is 0.22305. Although ALADIN-LAEF also scores better in terms of RMSE, the ratio RMSE/spread is inferior to that of the ECMWF-EPS (see fig. 8.45).

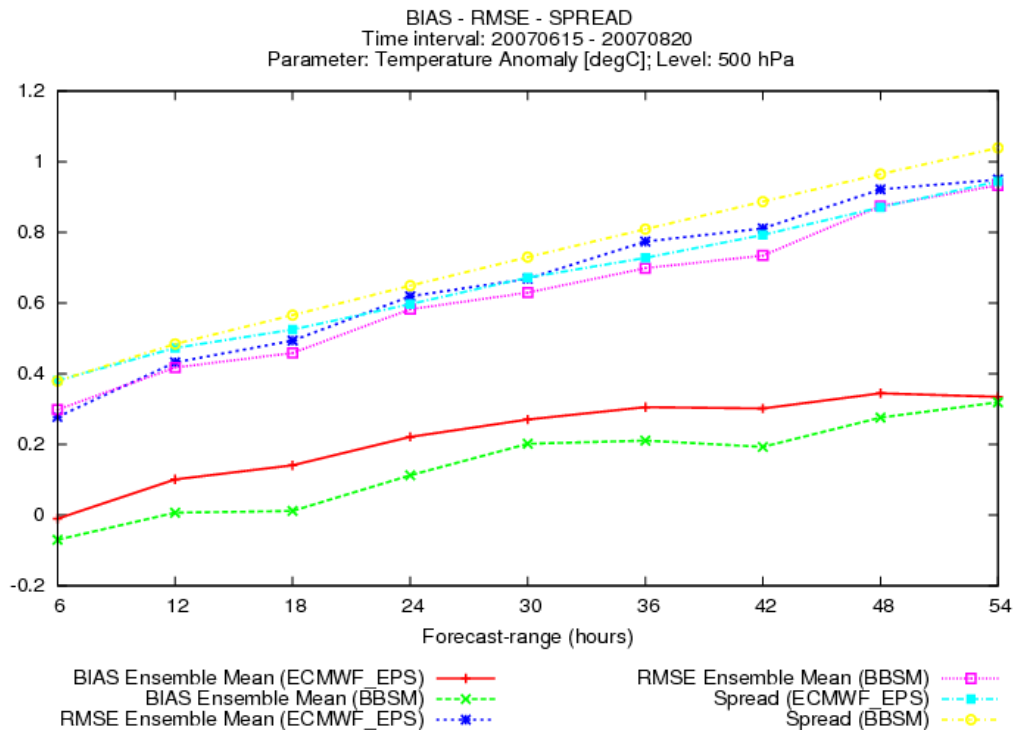


Figure 8.45: Temperature Anomaly in 500hPa - Bias, RMSE and Ensemble Spread

When the performance of both models is compared in terms of Brier Score, only slight differences can be detected, whereas the average performance of ALADIN-LAEF is a little better (range of differences 0.01). The computation of the Continuous Ranked Probability Score also reveals a better performance of ALADIN-LAEF (see fig. 8.46). Beyond that results based on the Continuous Ranked Probability Skill Score reveal a better performance of ALADIN-LAEF (see fig. 8.47). As it turns out, for both references (ECMWF analysis as well as ECMWF forecasts) ALADIN-LAEF scores slightly better than the ECMWF-EPS (range 0.01). It should be further noted, that both model score better when they are compared to the ECMWF analysis.

The calculation of the outliers also shows a better performance of ALADIN-LAEF (see fig. 8.48). Thus, more observations are located within the ensemble spread. This result is not surprising, since ALADIN-LAEF also produces a remarkably larger ensemble spread.

Results based on the calculation of reliability curves, show that both models perform better for shorter forecast lead times (6-18h). Although both models exhibit a tendency to overforecast lower probabilities and underforecast larger probabilities (see fig. 8.49), the average reliability is very good. Forecasts for larger forecast times show a high tendency to underforecast positive temperature anomalies and overforecast negative temperature anomalies. This tendency can be detected for all forecast lead times rang-

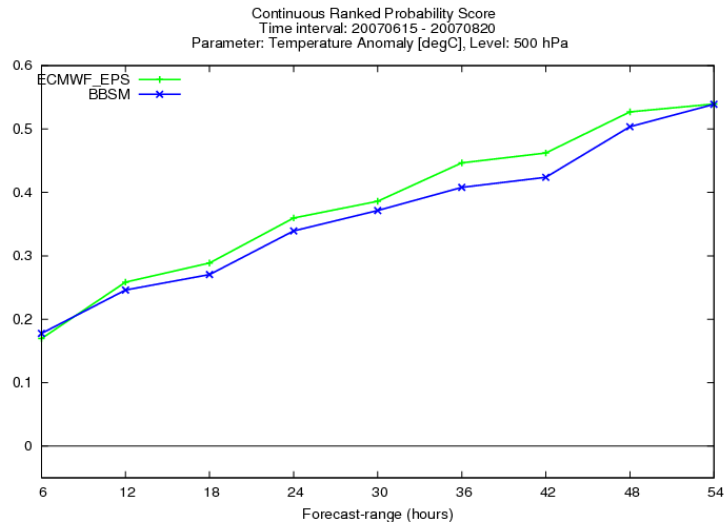


Figure 8.46: Temperature Anomaly 500hPa - Continuous Ranked Probability Score

ing from 24 to 54 hours (see fig. 8.50). This result is linked to the increasing bias with increasing forecast lead time. The comparison of all reliability diagrams reveals a better performance of ALADIN-LAEF in 19 of 27 different settings. Again it should be said that this is an effect of the overall smaller bias of ALADIN-LAEF.

The ROC diagrams show again that both models score best for shorter forecast lead times. The best results for both models are obtained for the forecast lead time of 6 hours. This means, that forecasts for short forecast lead times are not only very reliable, but also contain a high hit rate. This hit rate drops a little for longer forecast lead times (see fig. 8.51).

The final comparison of all scores reveals that ALADIN-LAEF performs better in 61 of 82 cases (see table 8.2 in short summary). The comparison of 16 major scores leads to the same result. Nevertheless some limitations should be considered. The ECMWF-EPS is more reliable for positive temperature anomalies and has a higher hit rate for short forecast lead times.

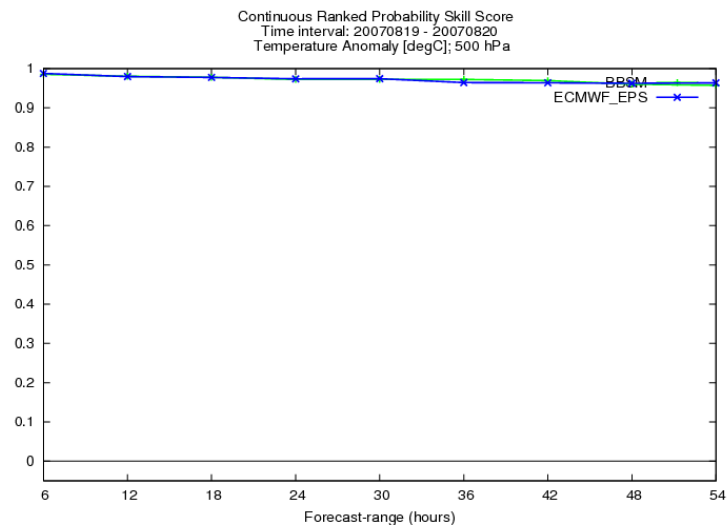


Figure 8.47: Temperature Anomaly 500hPa - Continuous Ranked Probability Skill Score - Reference: Analysis

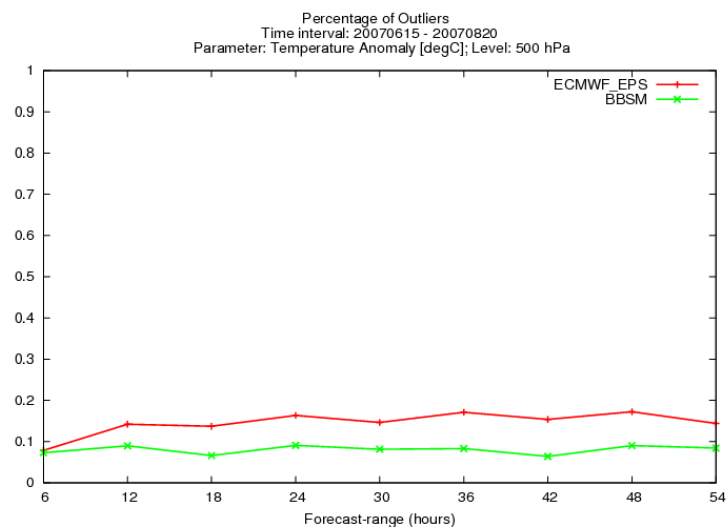


Figure 8.48: Temperature Anomaly 500hPa - Outliers

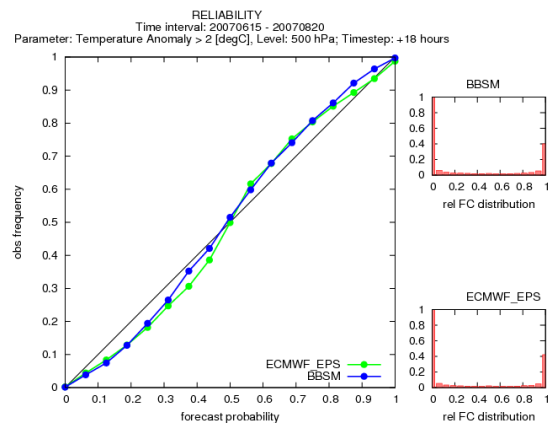


Figure 8.49: Temperature Anomaly 500hPa - Reliability

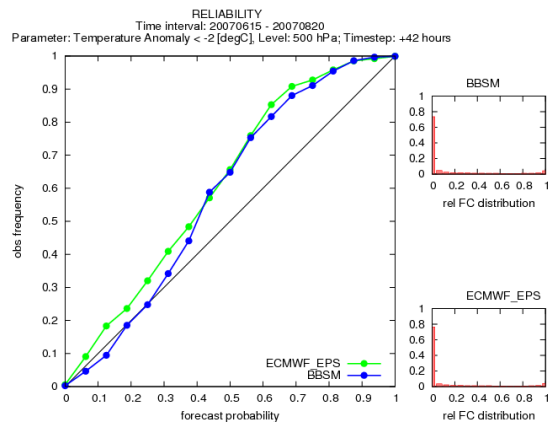


Figure 8.50: Temperature Anomaly 500hPa - Reliability

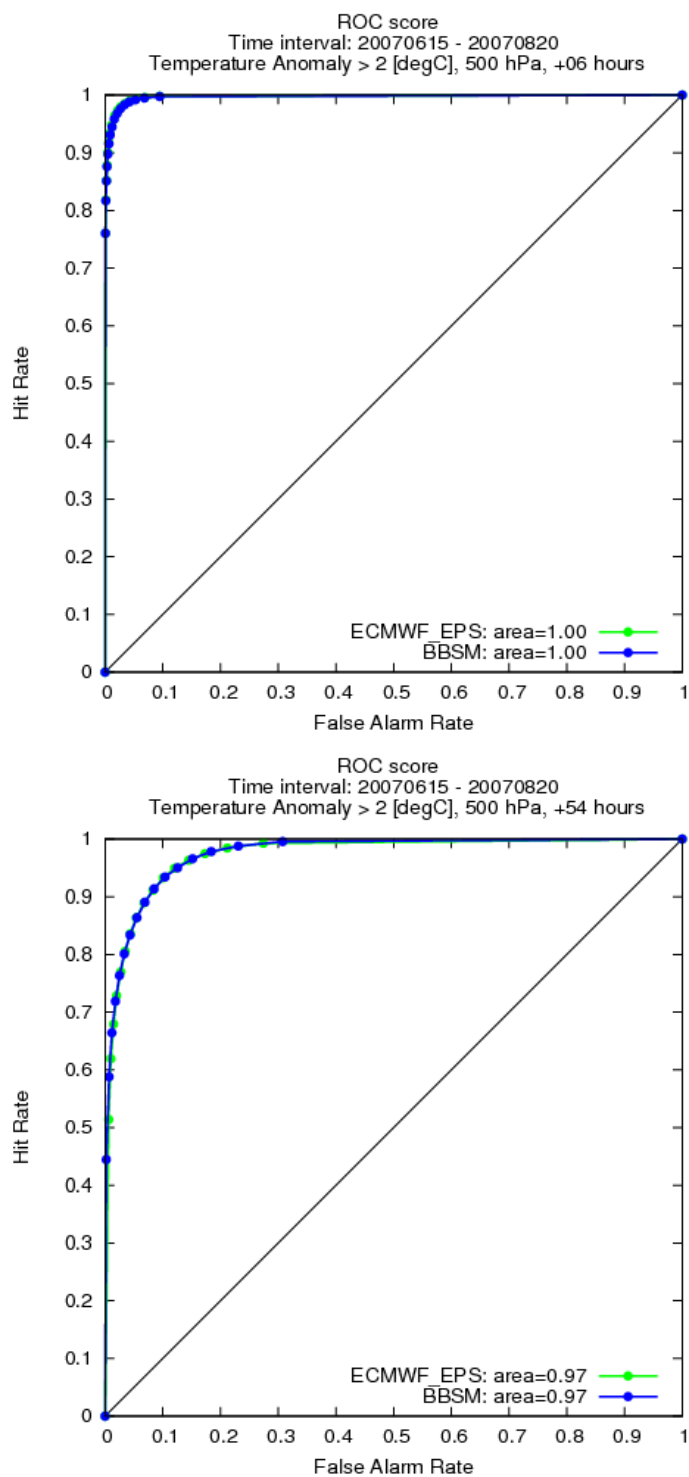


Figure 8.51: Temperature Anomaly 500hPa - Relative Operating Characteristic

8.2.4 Temperature Anomaly in 850 hPa

Results of the verification show that both models tend to predict too high temperatures. Hence, the forecasts exhibit a positive bias. For shorter forecast lead times up to 30 hours, ALADIN-LAEF shows a smaller bias than the ECMWF-EPS (see fig. 8.52). In contrast the ECMWF-EPS scores better for longer forecast lead times. When the average bias is calculated, ALADIN-LAEF turns out to be superior. However ALADIN-LAEF produces a larger average RMSE than the ECMWF-EPS. This result indicates that ALADIN-LAEF must produce more outliers than the ECMWF-EPS. The calculation of the outliers proves that this notion is correct (see fig. 8.53). When the performance is evaluated in terms of RMSE/spread, ALADIN-LAEF scores better. The average ratio RMSE/spread of ALADIN-LAEF is 1.08116, whereas the average ratio RMSE/spread of the ECMWF-EPS is 1.18198.

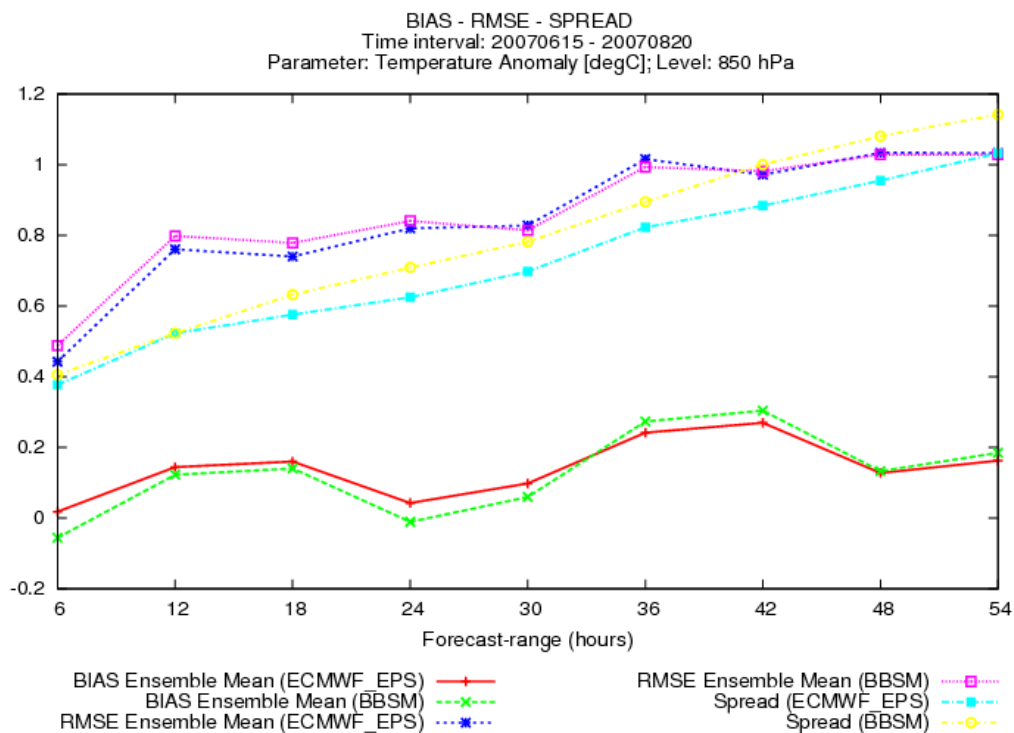


Figure 8.52: Temperature Anomaly in 850hPa - Bias, RMSE and Ensemble Spread

Results based on the calculation of reliability reveal a clearly better performance of the ECMWF-EPS. The ECMWF-EPS scores better in 19 of 27 different settings. This better performance is mainly linked to positive temperature anomalies (see fig. 8.54). The other 8 cases, in which ALADIN-LAEF performs better are linked to negative temperature anomalies and short forecast lead times (for example see fig. 8.54). As the results demonstrate (for example fig. 8.54), both models issue very reliable forecasts. Thus, the curves are close to the perfect line. This is an effect of the small bias, especially for short forecast lead times.

When the performance of the models is evaluated in terms of ROC, the ECMWF-EPS scores better in 18 of 27 different settings. Again the cases where ALADIN-LAEF turns out to be superior, are those with negative temperature anomaly and short forecast lead times. Moreover the Talagrand diagrams show that the ECMWF-EPS is superior in regard to the prediction of 850 hPa temperature forecasts.

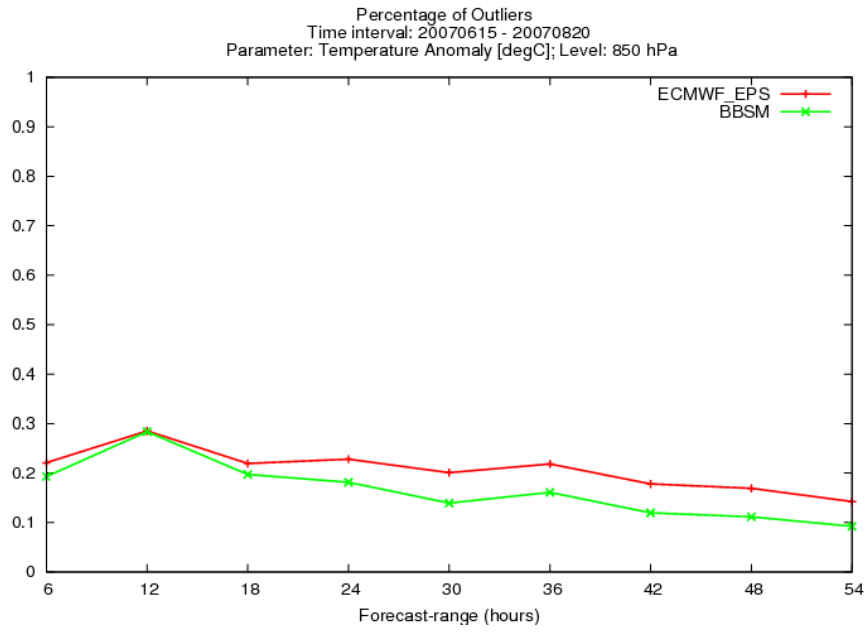


Figure 8.53: Temperature Anomaly in 850hPa - Outliers

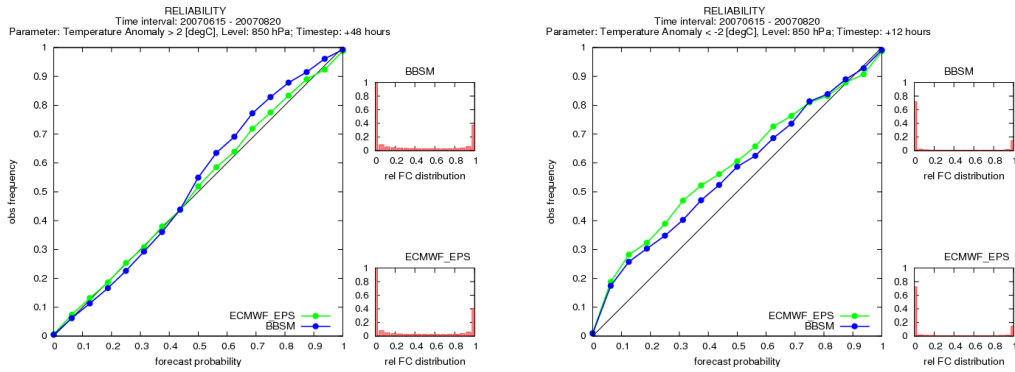


Figure 8.54: Temperature Anomaly in 850hPa - Reliability of forecasts, 12h and 48h

The final comprehensive assessment of all scores, thresholds and forecast lead times points out that the ECMWF-EPS scores better in 53 of 82 cases (see table 8.2 in short summary). The comparison of 16 major scores also reveals the superior performance of the ECMWF-EPS.

8.2.5 Wind speed in 500 hPa

Results based on bias and RMSE show an average better performance of the ECMWF-EPS (see fig. 8.55). The average bias of the ECMWF-EPS corresponds to 0.01738, whereas the average bias of ALADIN-LAEF is -0.18058. Thus, the wind speed forecasts of ALADIN-LAEF are constantly too low. When the performance of both models is compared in terms of RMSE/spread the ECMWF-EPS scores also better. However it should be noted that the differences in terms of RMSE/spread are extremely small (1.00078 versus 0.97929).

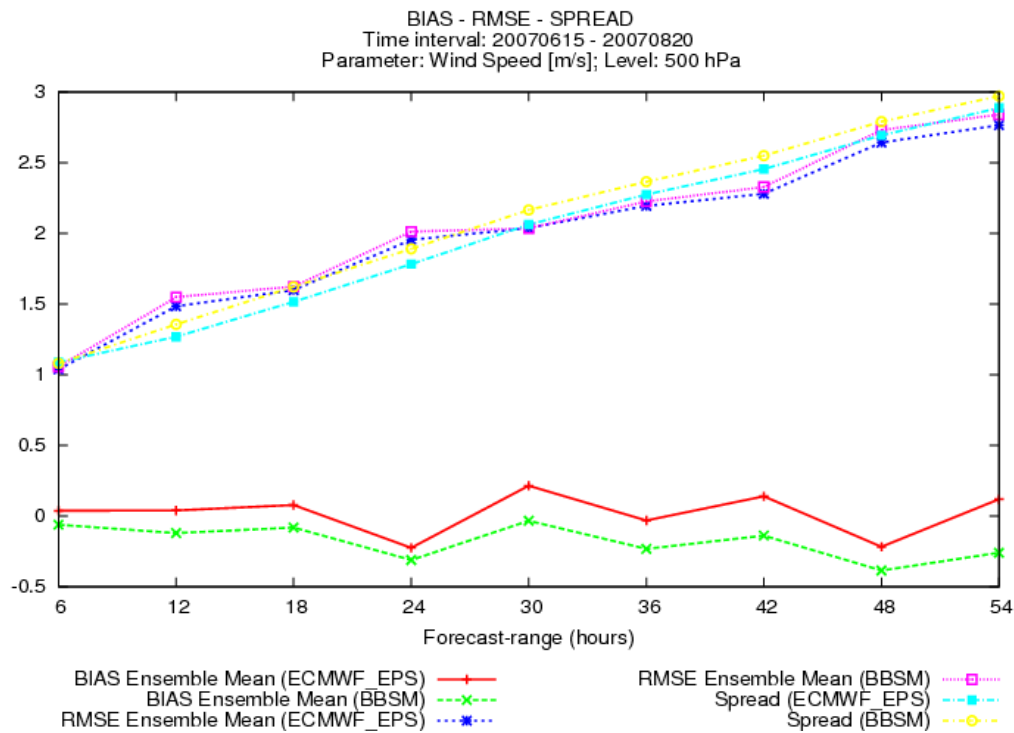


Figure 8.55: Wind Speed in 500hPa - Bias, RMSE and ensemble spread

Results based on the calculation of the Brier Score and Ranked Probability Score show hardly any differences between both models. The calculation of the Continuous Ranked Probability Score considers a broader range of threshold values. Results show that in terms of CRPS the ECMWF-EPS performs better (see fig. 8.56). The calculation of the percentage of outliers reveals a better performance of ALADIN-LAEF for all forecast lead times (see fig. 8.57). The reason behind this result might be linked to the larger ensemble spread generated by ALADIN-LAEF. However both ensemble prediction systems reveal good reliability for almost all forecast lead times and thresholds (see fig. 8.58). The best results turn up for the forecast lead time of 6 hours. However both model exhibit a general tendency to underforecast low wind speeds for forecast lead times ranging from 12 to 54 hours (see fig. 8.59). In terms of reliability, the ECMWF-EPS is clearly superior to ALADIN-LAEF. It performs better in 29 of 36 different settings. This result is based on the smaller bias of the ECMWF-EPS. Results based on the ROC diagram demonstrate again, the both models score best for the forecast lead time of 6 hours regardless of the selected threshold (see fig. 8.60). For larger forecast lead times and lower wind speeds (threshold: 2 m/s and 4 m/s), the hit rate decreases slightly (see fig. 8.61). Although it turns out that the ECWMF-EPS performs better for

almost all scores and forecast lead times, the Talagrand diagram draws a completely different picture. Results based on the Talagrand diagram show that ALADIN-LAEF scores better in 8 of 9 different settings. This output could be linked to the notably larger spread generated by ALADIN-LAEF. The final comparison of all scores and different settings reveals a superior performance of the ECMWF-EPS. The ECMWF-EPS performs better in 87 of 103 different settings (see table 8.2 in short summary). This result is also confirmed by the comparison of 16 major scores.

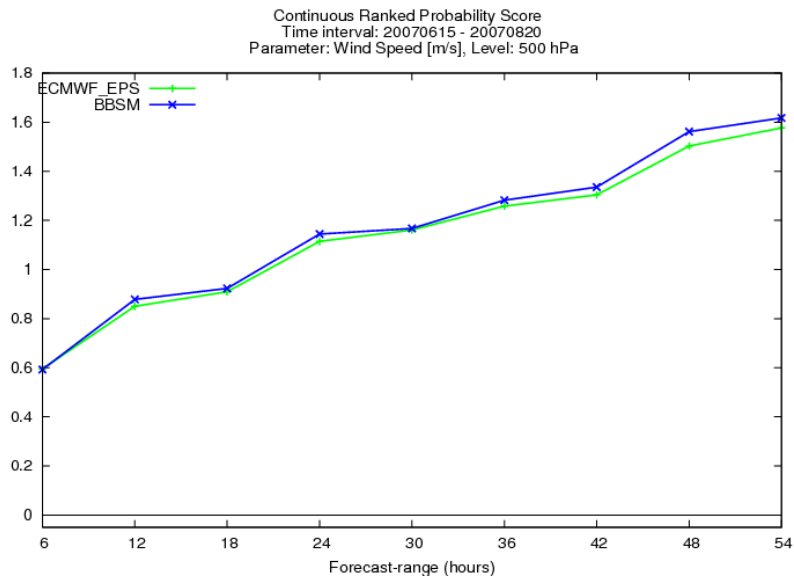


Figure 8.56: Wind Speed in 500hPa - Continuous Ranked Probability Score

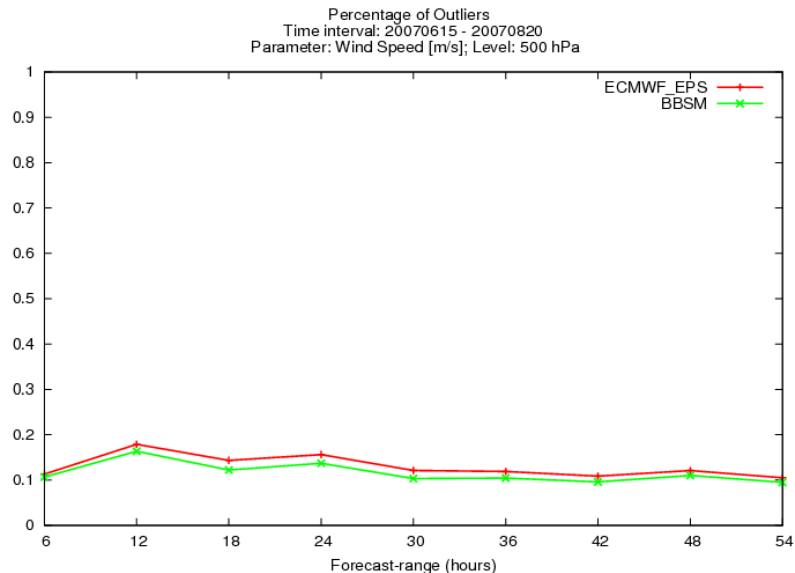


Figure 8.57: Wind Speed in 500hPa - Outliers

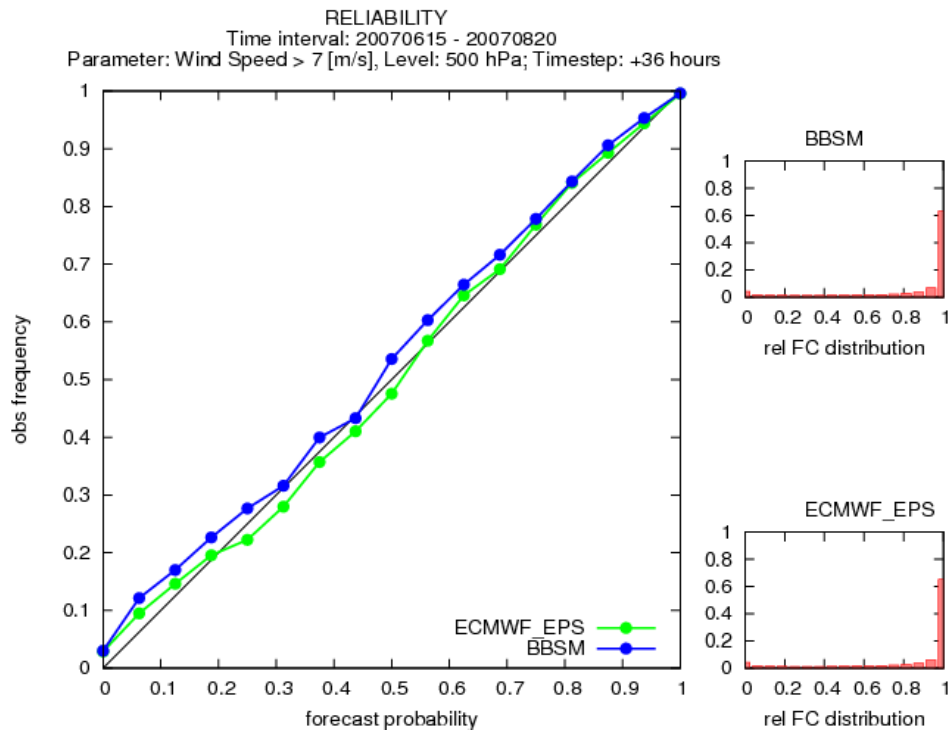


Figure 8.58: Wind Speed in 500hPa - Reliability

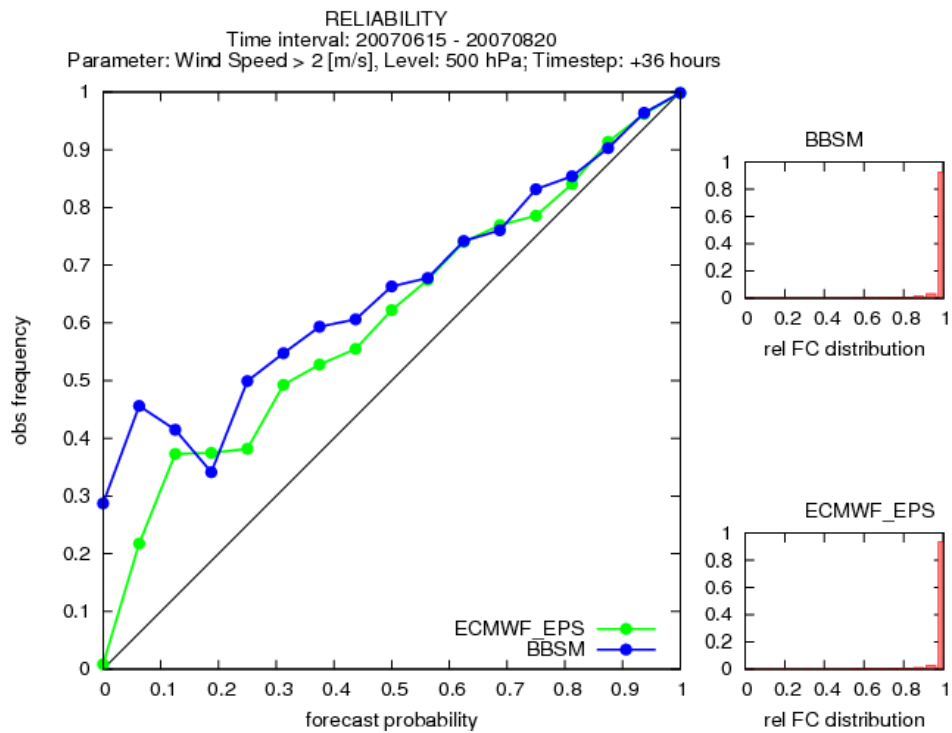


Figure 8.59: Wind Speed in 500hPa - Reliability

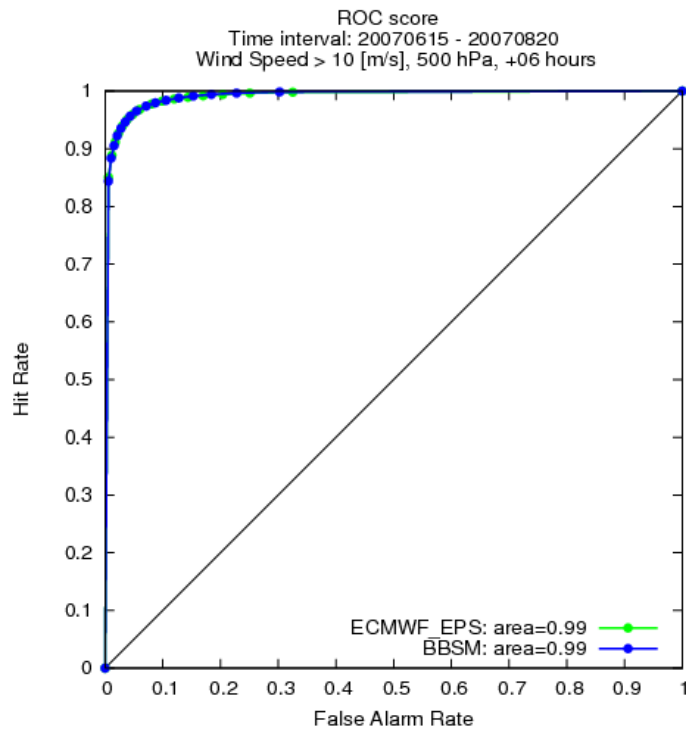


Figure 8.60: Wind Speed in 500hPa - Relative Operating Characteristic

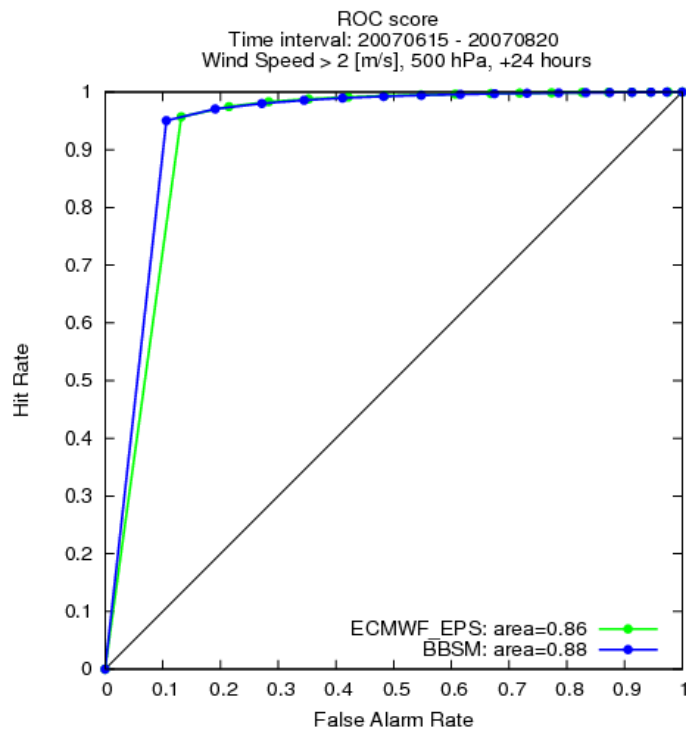


Figure 8.61: Wind Speed in 500hPa - Relative Operating Characteristic

8.2.6 Wind speed in 850 hPa

Results of the verification show that the forecasts generated by ALADIN-LAEF are only slightly biased (average bias: 0.02843). In contrast the ECMWF-EPS forecasts contain a remarkably larger bias (average bias: 0.32763). When the performance is measured in terms of RMSE, the ECMWF-EPS scores better for forecast lead times up to 24 hours (see fig. 8.62). For larger forecast lead times ALADIN-LAEF performs better. However it should be noted that the involved differences are really small. Measuring the performance in terms of RMSE/spread ALADIN-LAEF scores clearly better.

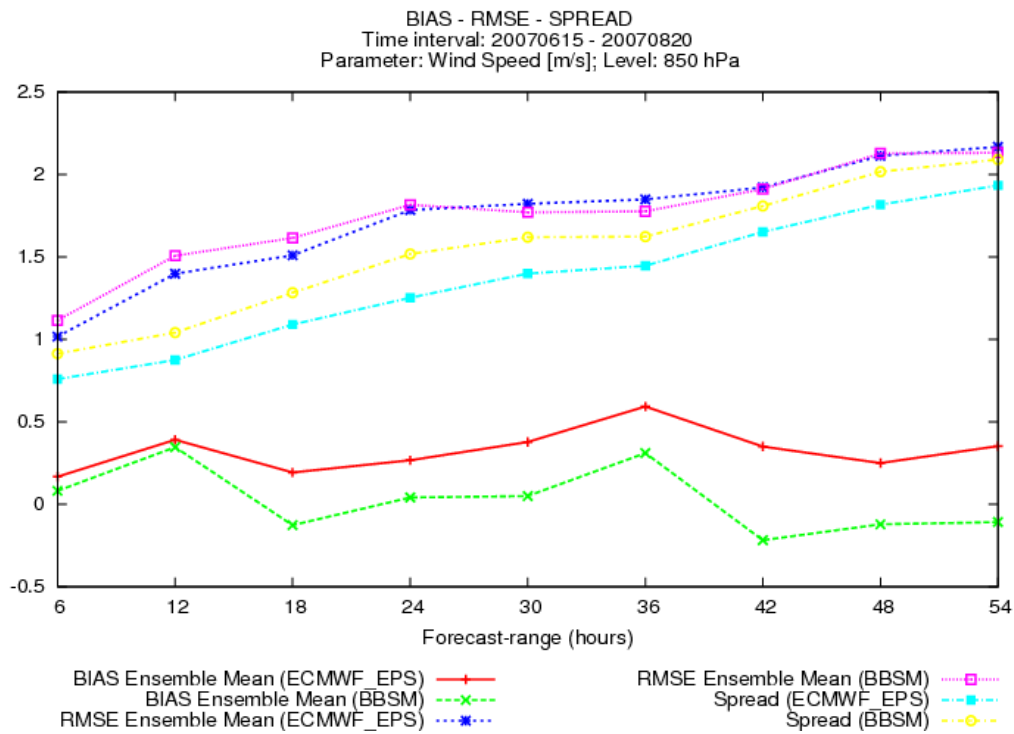


Figure 8.62: Wind Speed in 850hPa - Bias, RMSE and ensemble spread

When the performance is analyzed in regard to reliability, results show that the ECMWF-EPS generates better forecasts for low wind speeds (see fig.8.2.6). The observation of all reliability diagram reveals a general tendency to underforecast low wind speeds. For the threshold value of 2 m/s, the ECMWF-EPS scores better for all forecast lead times. In contrast ALADIN-LAEF produces more reliable forecasts for higher wind speeds (threshold values 7 m/s and 10 m/s - see fig. 8.2.6). The better performance of ALADIN-LAEF can be linked to the smaller bias. The comparison of all ROC areas shows that the ECMWF-EPS scores better for all forecast lead times and selected thresholds. This means that the ECMWF-EPS produces more hits than ALADIN-LAEF. In contrast results based on the calculation of the rank distribution, demonstrate a better performance of ALADIN-LAEF. Since both models produce rank distributions with slight underdispersion, the ensemble member tend to be too similar but different from the verification. The best results can be observed for the forecast lead time of 48 hours (see ALADIN-LAEF in fig. 8.2.6). The final comparison of all scores and selected thresholds indicates a better performance of the ECMWF-EPS. The ECMWF-EPS scores better in 71 of 103 different settings (see table 8.2 in short summary). This finding is also confirmed by the comparison of 16 major scores.

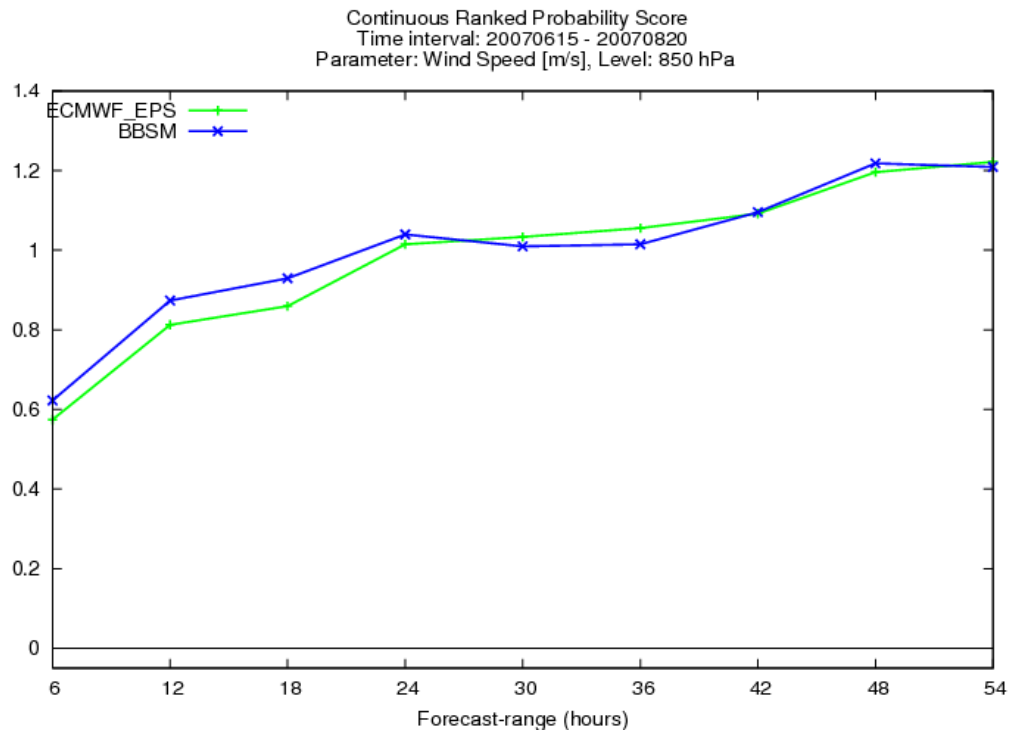


Figure 8.63: Wind Speed in 850hPa - Continuous Ranked Probability Score

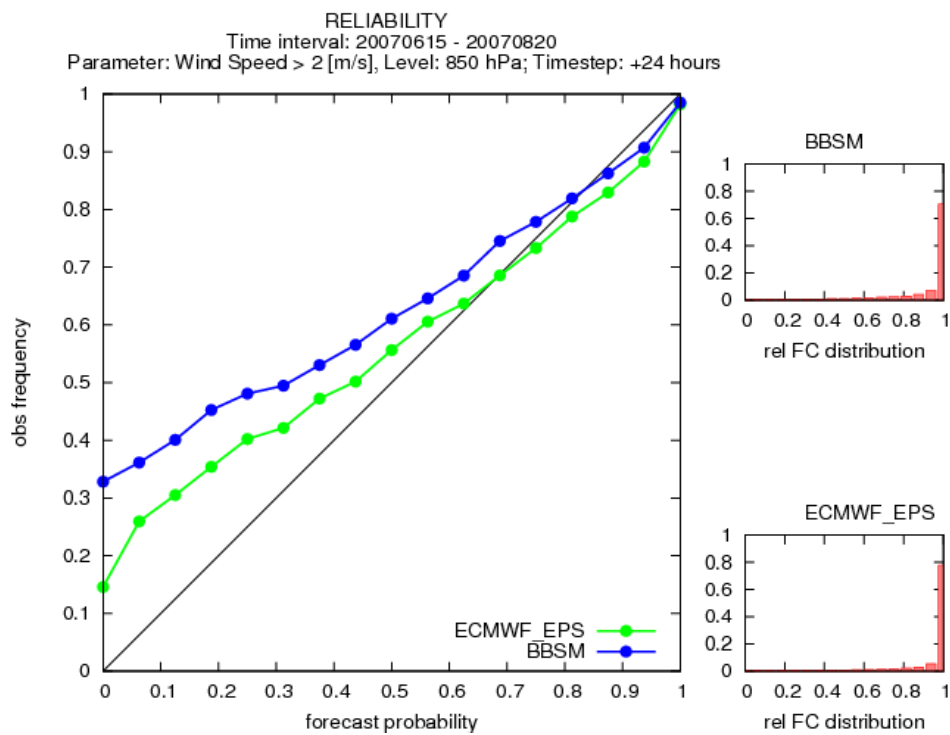


Figure 8.64: Wind Speed in 850hPa - Reliability, forecast lead time: 24h, thr. 2 m/s

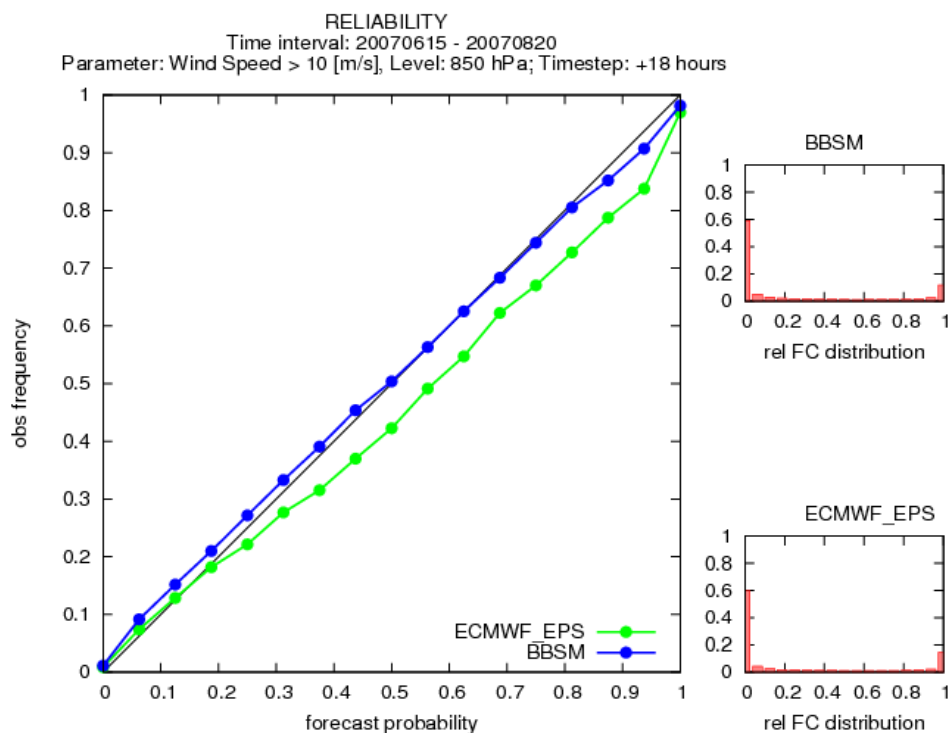


Figure 8.65: Wind Speed in 850hPa - Reliability
forecast lead times: 18h and threshold 10 m/s

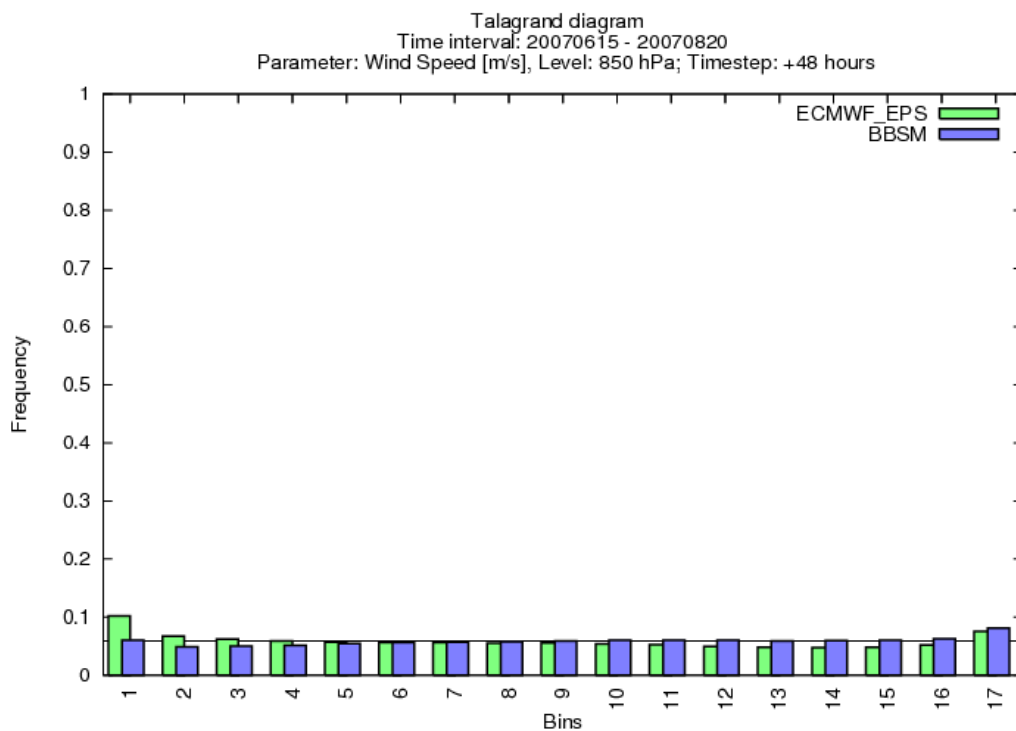


Figure 8.66: Wind Speed in 850hPa - Talagrand diagram, forecast lead time: 48h

8.2.7 Relative Humidity in 500hPa

The forecasts generated by ALADIN-LAEF reveal a smaller bias (average bias: -0.82311) than those of the ECMWF-EPS (average bias: -1.63412). In terms of RMSE, the ECMWF-EPS performs slightly better for shorter forecast lead times up to 24 hours (see fig. 8.67). For longer forecasts lead times ALADIN-LAEF reveals a better performance. Nevertheless the comparison of the average RMSE indicates a better overall performance of the ECMWF-EPS. However if both models are compared in terms of RMSE/spread, ALADIN-LAEF is superior.

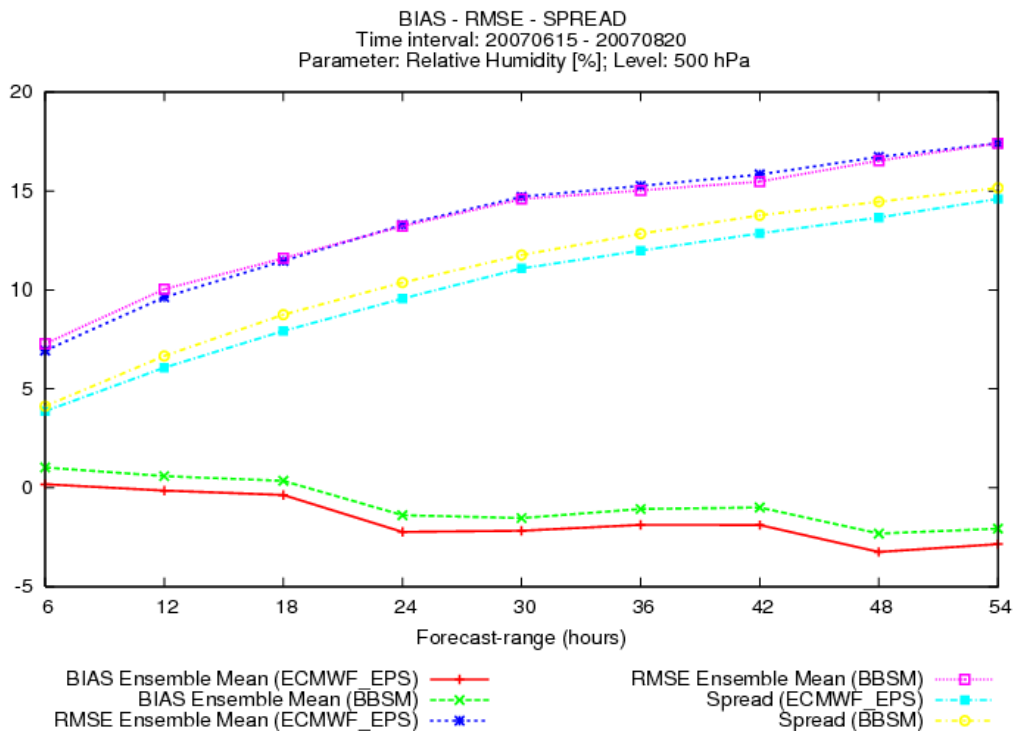


Figure 8.67: Relative Humidity in 500hPa - Bias, RMSE and ensemble spread

Like the results for the RMSE, the results for the Continuous Ranked Probability Score show only small differences between both EPS. The ECMWF-EPS turns out to be superior for shorter forecast lead times (up to 24 hours), whereas ALADIN-LAEF scores better for larger forecast lead times (see fig. 8.68).

The calculation of the percentage of outliers demonstrates that ALADIN-LAEF scores better for almost all forecast lead times (see fig. 8.69). Only for the short forecasts up to 12 hours the ECMWF-EPS seems to be slightly superior. Also in regard to reliability ALADIN-LAEF turns out to be clearly predominant (see fig. 8.71). ALADIN-LAEF shows a higher reliability than the ECMWF-EPS in 20 of 27 different settings. This result is an effect of the smaller bias of ALADIN-LAEF for larger forecast lead times (24h-54h). Only for short forecast lead times the ECMWF-EPS performs better (see fig. 8.70). The comparison of the calculated ROC-areas for both models reveals a clearly superior performance of the ECMWF-EPS. The ECMWF-EPS turns out to be superior for all forecast lead times. Moreover the comparison of the rank distribution shows that the ECMWF-EPS performs better. The final comparison of all scores and thresholds shows that the ECMWF-EPS performs better in 58 of 82 different cases (see table 8.2 in short summary). When the comparison is based on 16 major scores, the same results is

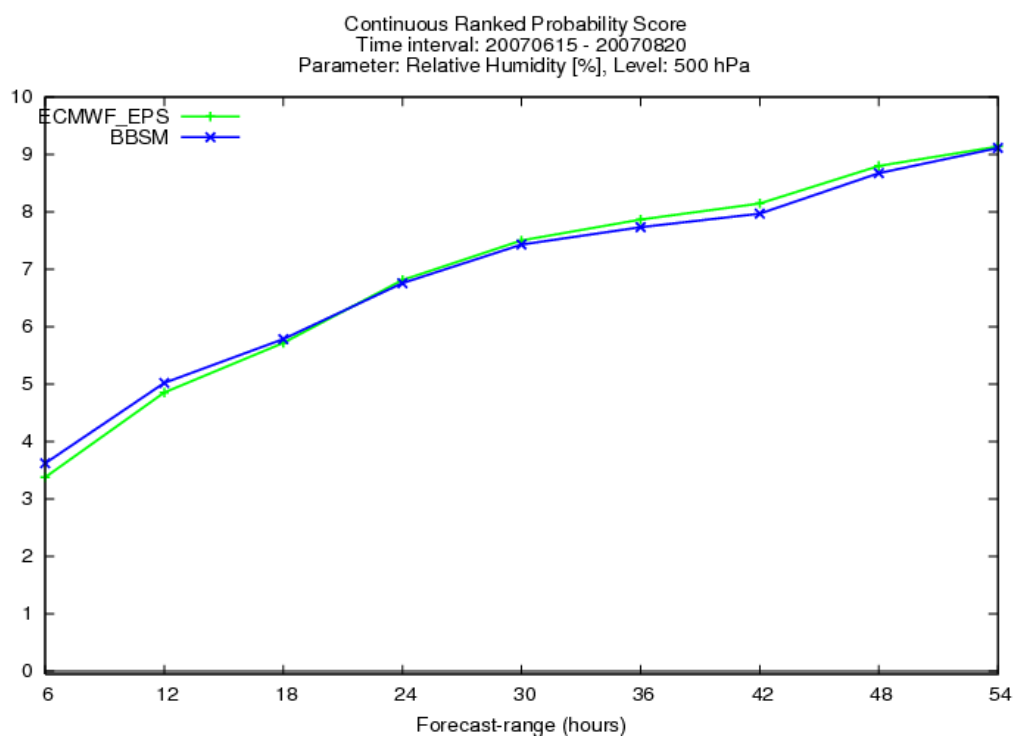


Figure 8.68: Relative Humidity in 500hPa - Continuous Ranked Probability Score

obtained.

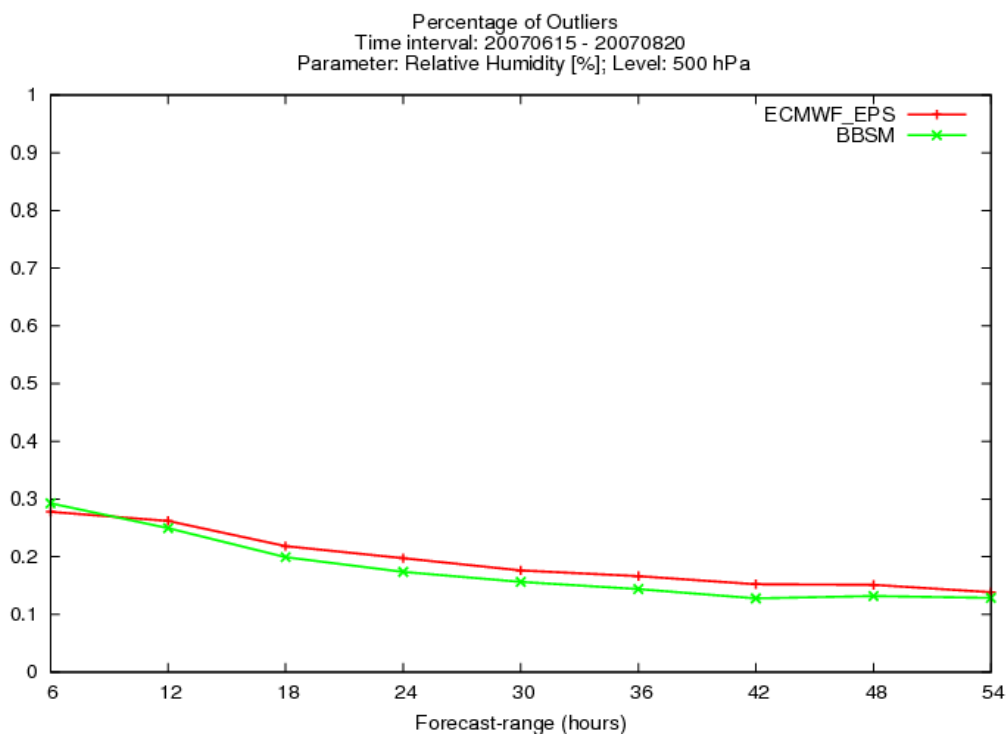


Figure 8.69: Relative Humidity in 500hPa - Outliers

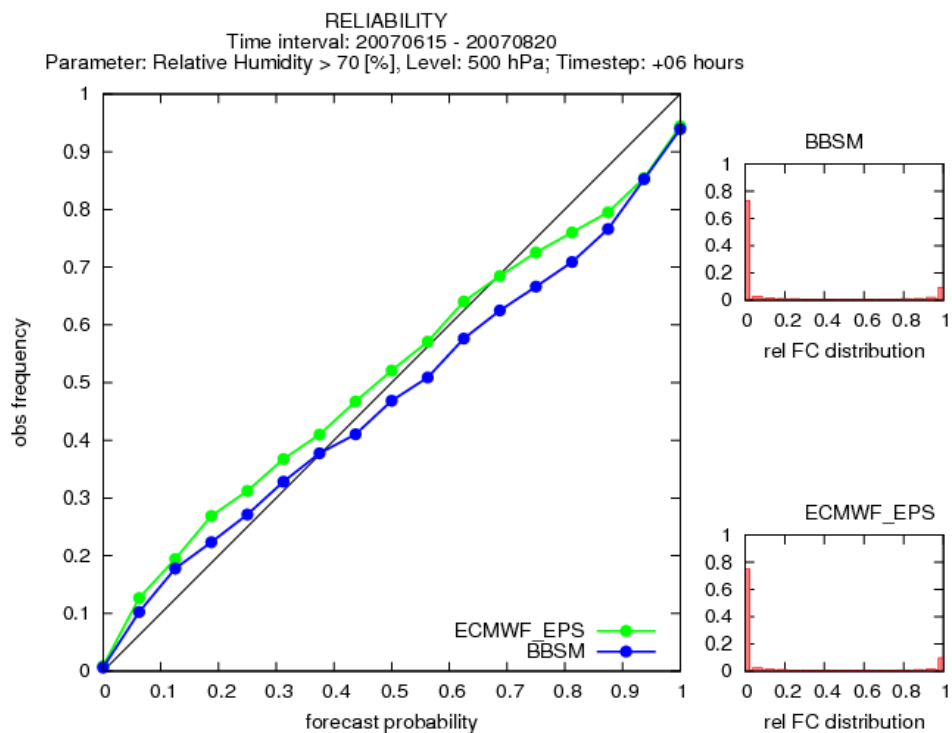


Figure 8.70: Relative Humidity in 500hPa - Reliability
forecast lead time: 6h, threshold 70%

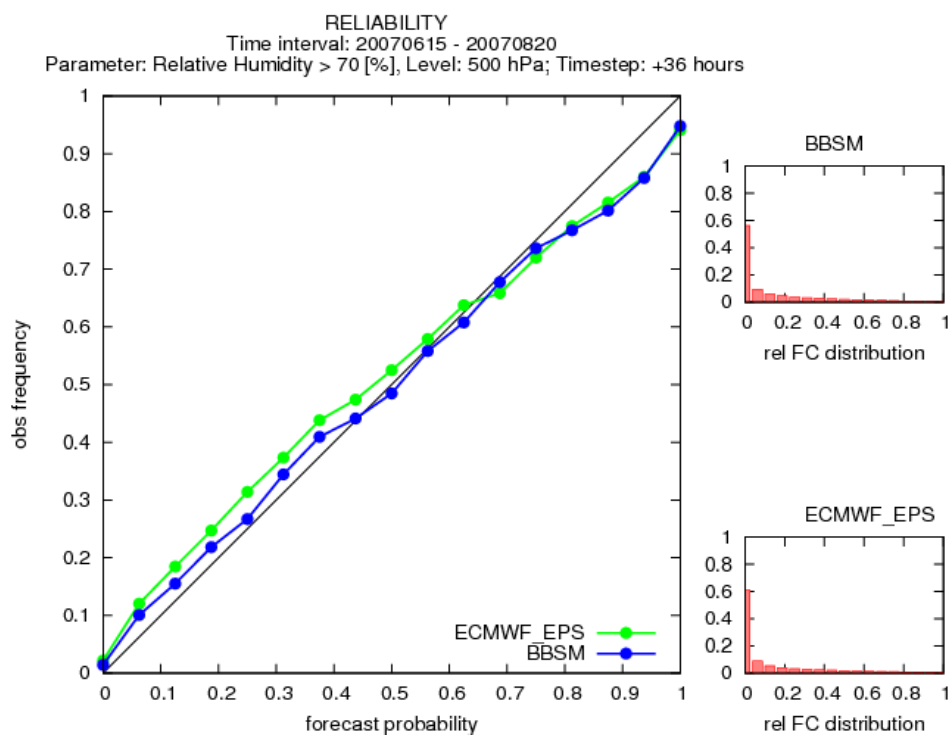


Figure 8.71: Relative Humidity in 500hPa - Reliability
forecast lead time: 36h, threshold 70%

8.2.8 Relative Humidity in 850 hPa

Results show that the forecasts generated by ALADIN-LAEF contain a larger bias than the forecasts of the ECMWF-EPS (see fig. 8.72). Moreover ALADIN-LAEF reveals a larger RMSE than the ECMWF-EPS. However since ALADIN-LAEF produces a notably larger spread than the ECMWF-EPS, it performs better in terms of RMSE/spread.

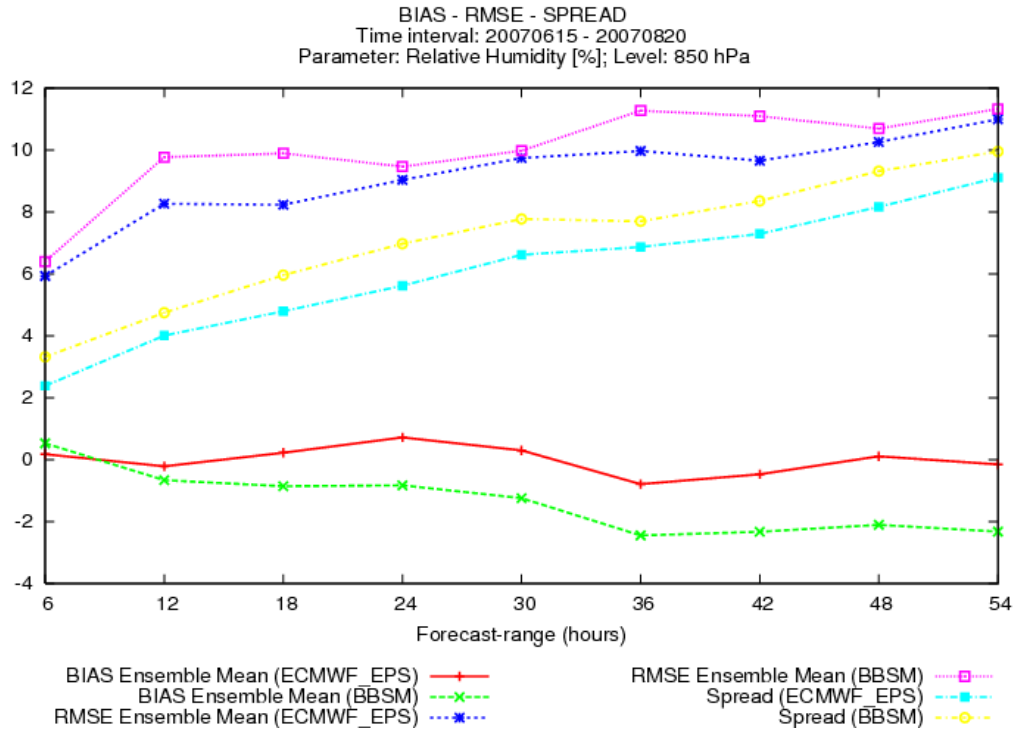


Figure 8.72: Relative Humidity in 850hPa - Bias, RMSE and ensemble spread

Results based on the Brier Scores show remarkable differences between both models. For the smallest selected threshold of 40% relative humidity, the differences are relatively small (see fig. 8.73). For this threshold the ECMWF-EPS performs better. The average Brier Score of the ECMWF-EPS corresponds to 0.05135, whereas the average Brier Score of ALADIN-LAEF is 0.05793. For the threshold 55% the ECMWF-EPS performs clearly better (see fig. 8.74). The largest differences between both models turn up for the largest threshold of 70% relative humidity. For this threshold the ECMWF-EPS again performs better. However it should be noted that this superior performance is linked to certain forecast lead times (see fig. 8.75). The ECMWF-EPS scores better for the forecast lead times of 12, 18, 36 and 42 hours. For the other forecast lead times, the differences between both models are relatively small. The better performance of the ECMWF-EPS is also demonstrated by the Continuous Ranked Probability Score. This indicates that regardless of the selected threshold, the ECMWF-EPS scores better (see fig. 8.76). The evaluation of the model performance in regard to reliability shows again a better overall performance of the ECMWF-EPS. The ECMWF-EPS scores better in 23 of 27 different cases. This result is linked to the remarkably smaller bias of the ECMWF-EPS. However it should be noted that the differences between both models in terms of reliability are relatively small. The comparison of all thresholds and forecast lead times reveals that both models tend to overforecast low relative humidity, whereas they underforecast high relative humidity (see fig. 8.78 and fig. 8.79).

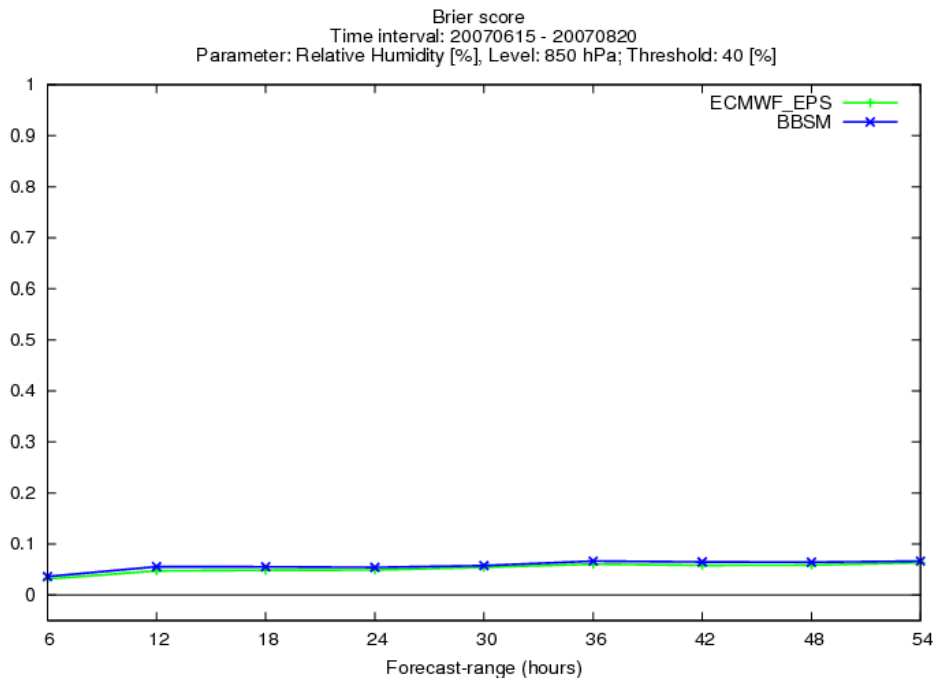


Figure 8.73: Relative Humidity in 850hPa - Brier Score, threshold 40%

The comparison of the ROC-areas for all forecasts lead times and selected threshold demonstrates a superior performance of the ECMWF-EPS. For all settings the ECMWF-EPS turns out to be superior. The best results are obtained for the threshold 55% which indicates that the prediction of moderate humidity might be easier. In contrast the prediction of low relative humidity turns out to be more challenging. The worst result is obtained for ALADIN-LAEF, a forecast lead time of 12 hours and the threshold 40% (see fig. 8.80). For this threshold the ECMWF-EPS scores better, but also differs notably from the perfect score 1. Results based on the Talagrand diagram show only small differences. However the ECMWF-EPS outperforms ALADIN-LAEF in 6 of 9 different settings. Both models score best for the forecast lead time of 48 hours (see fig. 8.81).

The general comparison of all scores and different settings points out a clearly dominant performance of the EMCWF-EPS. The ECMWF-EPS performs better in 72 of 82 different cases (see table 8.2). This result might be based on the fact, that the transport of moist air can be better incorporated in a global model. The development of major frontal systems which affect the European continent, mainly takes place in the Atlantic Ocean. Therefore the transport of moist air masses might be better considered when operating on a larger domain like that of the ECMWF-EPS.

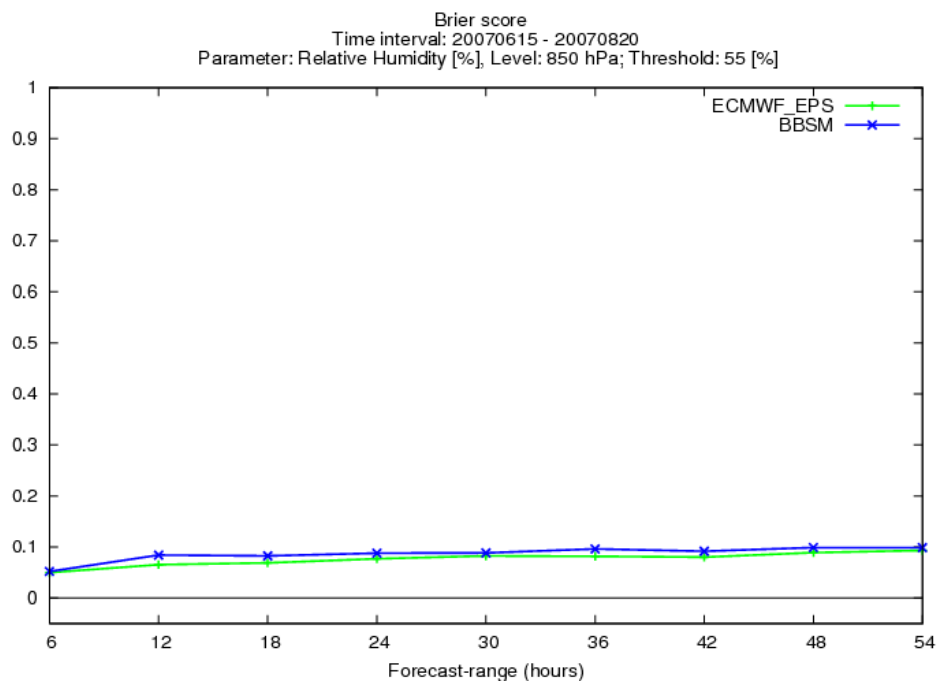


Figure 8.74: Relative Humidity in 850hPa - Brier Score, threshold 55%

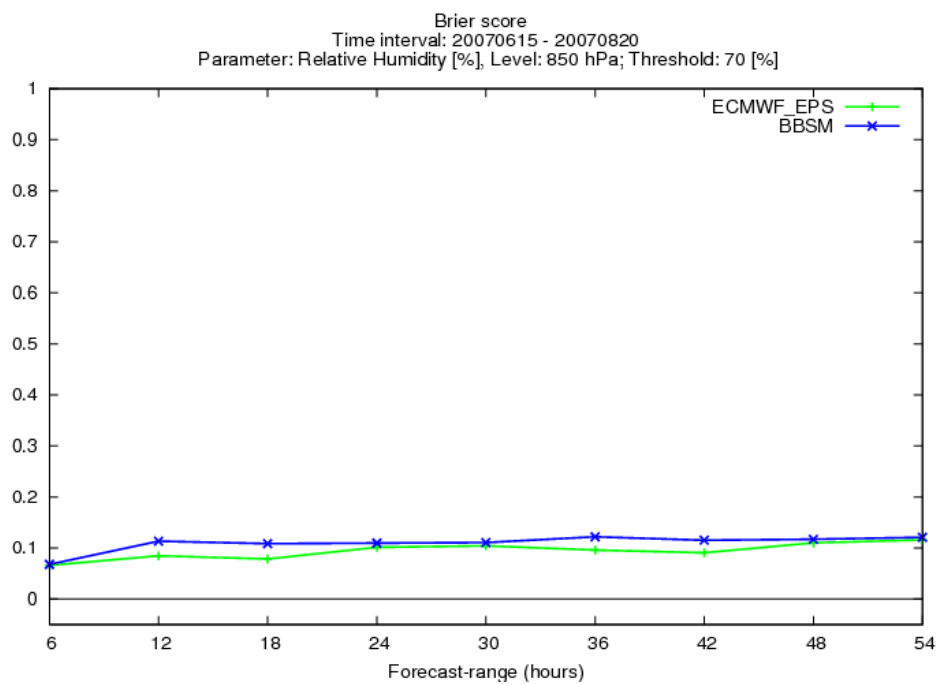


Figure 8.75: Relative Humidity in 850hPa - Brier Score, threshold 70%

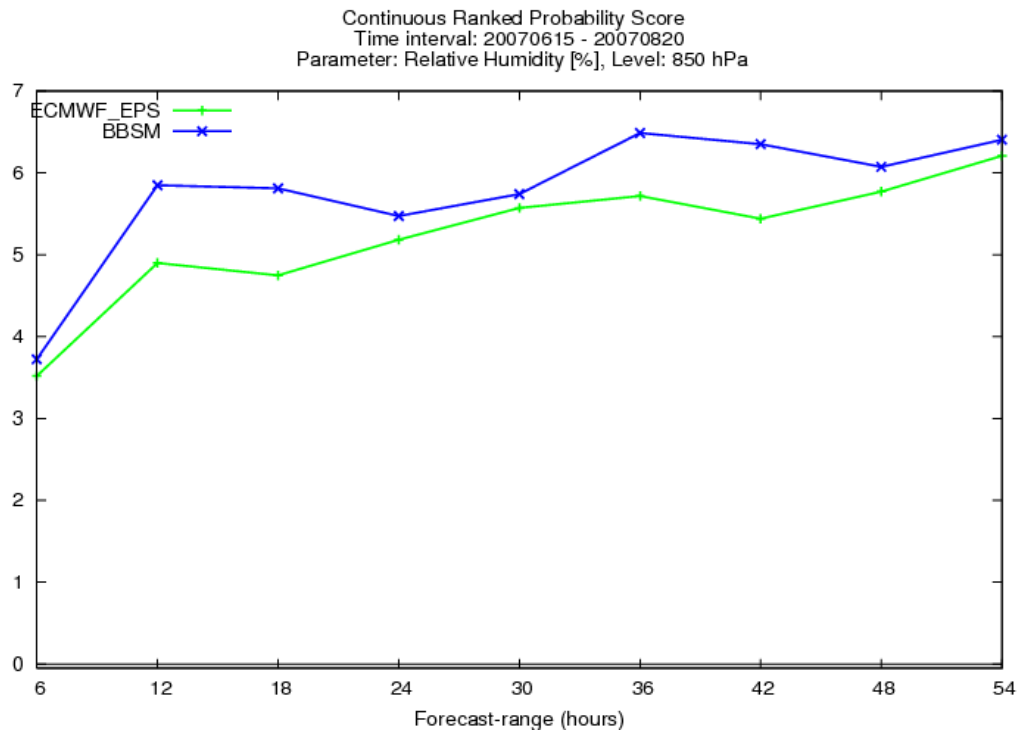


Figure 8.76: Relative Humidity in 850hPa - Continuous Ranked Probability Score

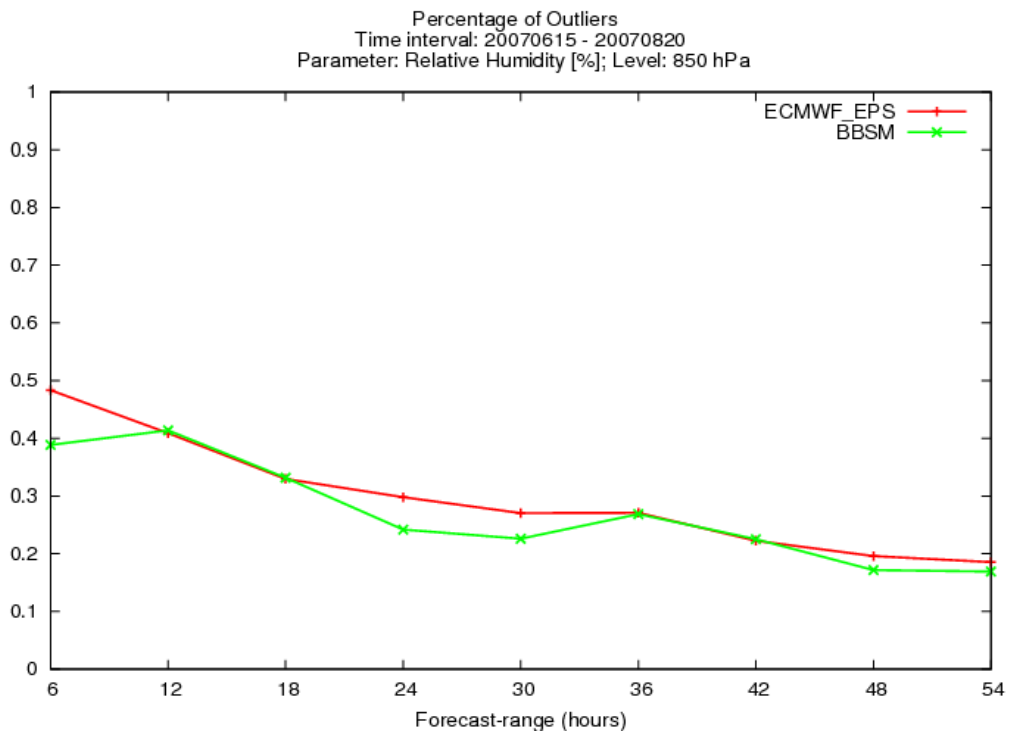


Figure 8.77: Relative Humidity in 850hPa - Percentage of Outliers

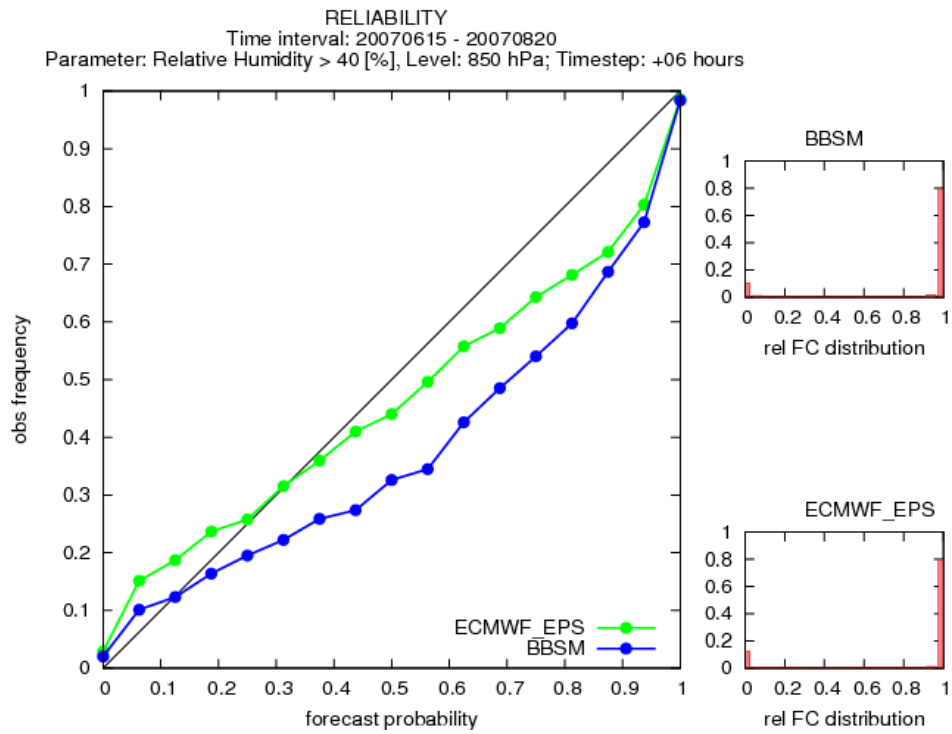


Figure 8.78: Relative Humidity in 850hPa - Reliability, 40% relative humidity

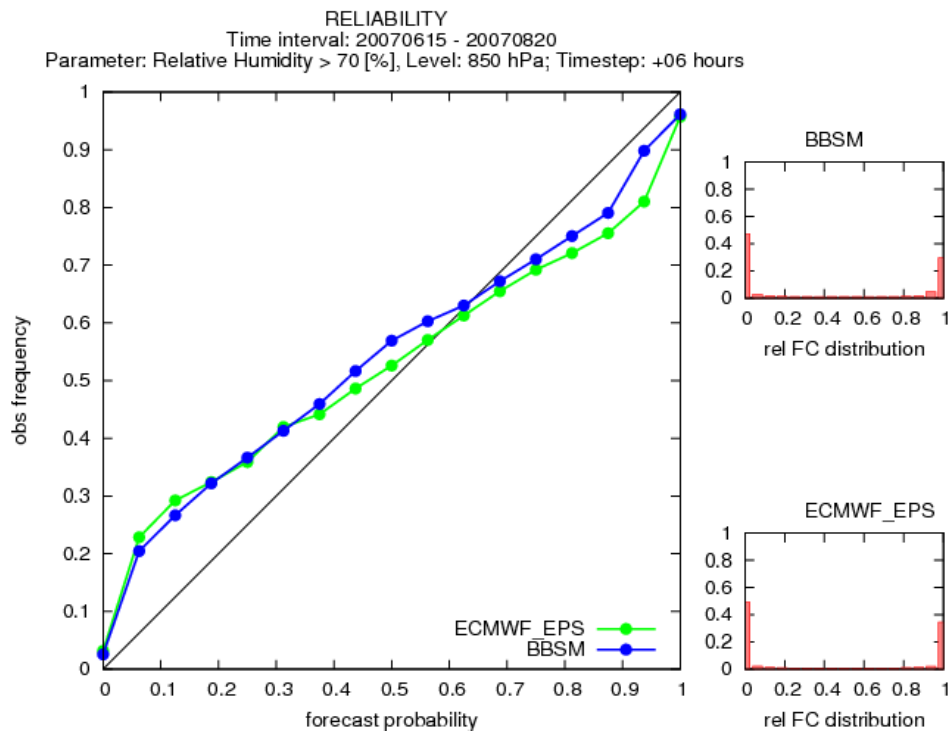


Figure 8.79: Relative Humidity in 850hPa - Reliability, 70% relative humidity

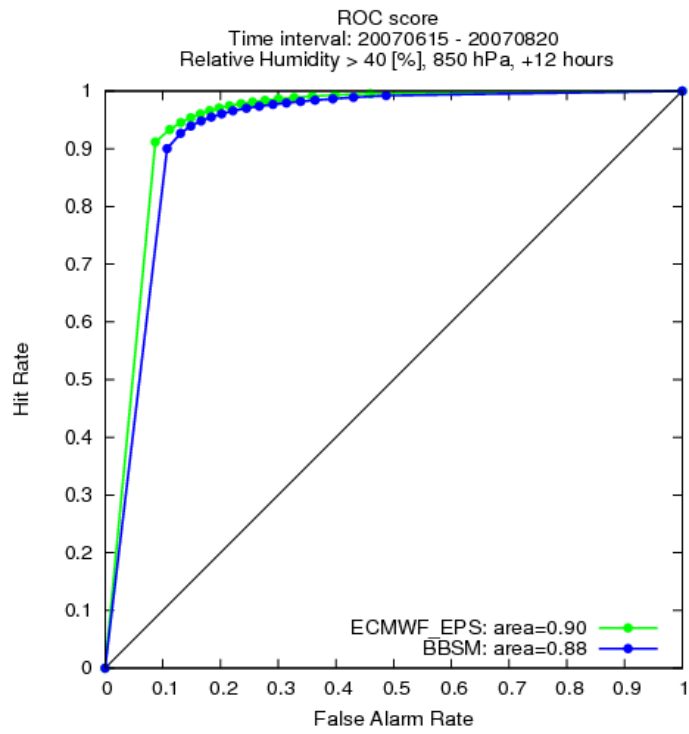


Figure 8.80: Relative Humidity in 850hPa - ROC

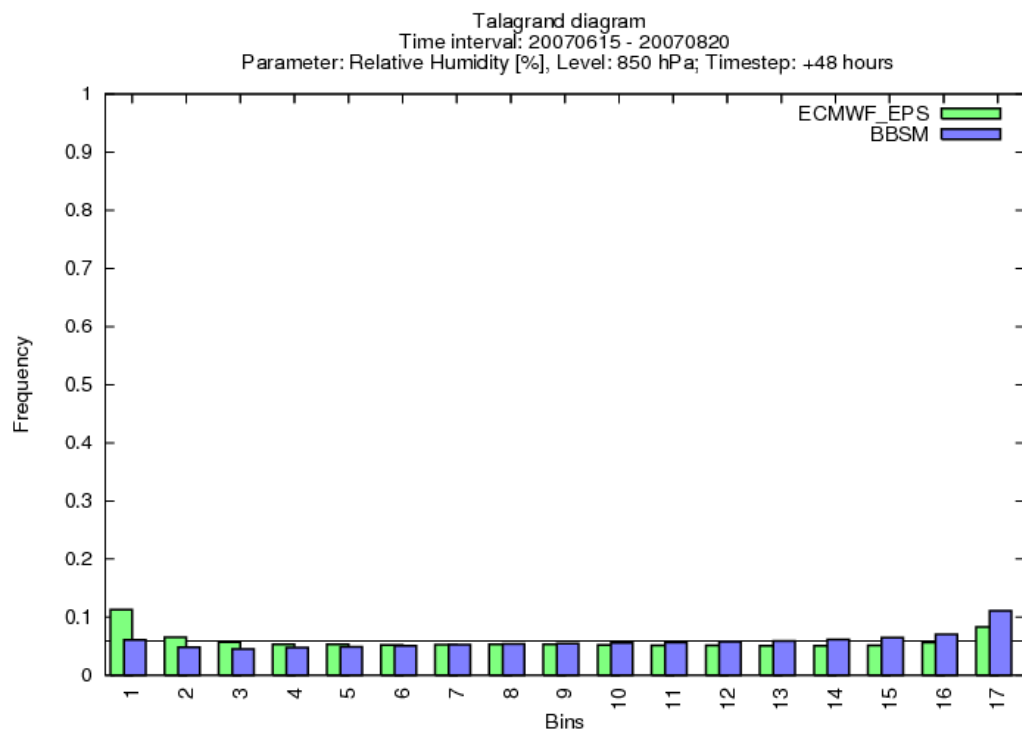


Figure 8.81: Relative Humidity in 850hPa - Talagrand diagram

8.2.9 Short Summary

The comprehensive evaluation of all upper level parameters reveals a better overall performance of the ECMWF-EPS. The ECMWF-EPS performs better in 370 of 550 different settings (see table 8.2). Although the differences between both models are often very small, the ECMWF-EPS performs better for almost all parameters. The dominance of the ECMWF-EPS is mainly linked to forecasts for the 500hPa-level. For example the 500hPa wind speed forecasts of the ECMWF-EPS are clearly superior. Moreover the ECMWF-EPS produces better forecasts for the humidity field in 500 hPa. In contrast ALADIN-LAEF generates better temperature forecasts for the 500 hPa level. This result seems a bit surprising. Generally the large scale transport of air masses can be better predicted by a global model. Therefore the ECMWF-EPS is expected to score better for all parameters in the 500hPa level. However this is not true for the temperature anomaly. The rationale behind that is the remarkably larger bias of the ECMWF-EPS temperature forecasts for the 500 hPa level. The comparison of the forecasts for the 850 hPa field, reveals a better performance of the ECMWF-EPS. The ECMWF-EPS produces better forecasts for the humidity field, the wind speed and the temperature anomaly. It should be noted that these results are based on the average performance. Thus, the general performance should not be based as a single directive, since results differ for selected thresholds and forecast lead times. For example ALADIN-LAEF is more reliable for high wind speed forecasts (7 m/s and 10 m/s). Moreover the temperature forecasts of ALADIN-LAEF are less biased than those of the ECMWF-EPS. These examples show that the average performance only gives a hint to the quality of a numerical model, but it cannot account for specific characteristics. However it should be stressed again, that for the upper level the ECMWF-EPS forecasts score clearly better. The chapter Appendix contains a detailed list of the results as well as an explanation of this statistic.

Results based on 16 major scores		
Parameter	ALADIN-LAEF	ECMWF-EPS
Geopotential in 500hPa	2	6
Geopotential in 850hPa	7	1
Temperature Anomaly in 500hPa	13	3
Temperature Anomaly in 850hPa	5	11
Wind Speed in 500hPa	1	15
Wind Speed in 850hPa	4	12
Relative Humidity in 500hPa	3	13
Relative Humidity in 850hPa	1	15
total	36	76
Results based on detailed thresholds		
Parameter	ALADIN-LAEF	ECMWF-EPS
Geopotential in 500hPa	2	6
Geopotential in 850hPa	7	1
Temperature Anomaly in 500hPa	61	21
Temperature Anomaly in 850hPa	29	53
Wind Speed in 500hPa	16	87
Wind Speed in 850hPa	32	71
Relative Humidity in 500hPa	24	58
Relative Humidity in 850hPa	9	73
total	180	370

Table 8.2: Final Results - Upper Level Parameters

9 Statistical Post-Processing of EPS forecasts

9.1 Calibration

Numerical weather prediction models include a range of parameterizations, since some atmospheric processes take place on a horizontal scale which cannot be resolved by the numerical model. As a result the generated forecasts are frequently affected by systematic errors. It is considered that the correction of these errors can significantly improve the quality of the forecasts. For this purpose a calibration of the forecasts can be performed. The goal of calibration is to correct for such known model deficiencies and construct predictions with statistical properties similar to the observations (Hagedorn, 2010). There exist multiple methods how to post-process the forecasts. However all of them require a record of prediction-observation pairs, whereas the length of the data record plays a crucial role. Here is a short list of calibration methods which can be applied to the forecasts:

- Bias correction
- Multiple implementation of deterministic Model Output Statistics (MOS)
- Ensemble dressing
- Bayesian model averaging
- Non-homogeneous Gaussian regression

The simplest way of eliminating systematic errors is achieved by removing first-order bias. As a simple first order calibration a bias correction can be applied (Hagedorn, 2010).

$$c = \frac{1}{N} \sum_{i=1}^N \bar{e}_i - \sum_{i=1}^N o_i \quad (9.1)$$

Here the first term contains the ensemble mean for the i^{th} forecast and the second term the value of the i^{th} observation. The factor c is applied to each ensemble member, i.e. spread is not affected. Hamill et al. (2008) found the bias correction contributing to a large extent to full calibration.

From the outset of ensemble forecasting (Leith, 1974) it was anticipated that the use of finite ensembles would yield errors in the forecast ensemble mean that could be statistically corrected using a database of previous errors - essentially a MOS post-processing for the ensemble mean. However this approach should be applied carefully since for longer lead-times the MOS tends to correct towards climatology (Hagedorn, 2010). Hamill et al. (2004) report very effective MOS post-processing of the ensemble mean for both surface temperature anomalies and accumulated precipitation, at 6-10 day and 8-14 day lead times.

Another method to post-process forecasts is to produce an ensemble dressing. Therefore a probability distribution around either the ensemble mean or around each of the ensemble members is defined. There exists a number to find appropriate dressing kernels ("best-member dressing", "error dressing", etc.) The distributions that are superimposed are derived from historical error statistics of the ensemble prediction system. Atger (1999) used Gaussian distributions around the ensemble mean for the 500mb height, with standard deviations proportional to the forecast ensemble standard deviation. Since it appears beneficial to dress individual ensemble members rather than the ensemble mean, the procedure yields state-dependent uncertainty information even if the spread of the added error distribution is not conditional on the ensemble spread (Wilks, 2006). Bremnes (2004) forecasts probability distributions for precipitation are based on a two-stage ensemble MOS procedure that uses selected quantiles for the forecast ensemble precipitation distribution as predictors. According to Wilks (2006) first, the probability of nonzero precipitation is forecasted using a probit regression, which is similar to logistic regression, but using the cumulative distribution function (CDF) of the standard Gaussian distribution to constrain the linear function of the predictors to the unit interval. Here the CDF is denoted $F(x)$:

$$F(x) = Pr \{X \leq x\} = \int_{X \leq x} f(x) dx \quad (9.2)$$

$$f(x) = \phi(b_0 + b_1x_1 + b_2x_2 + b_3x_3) \quad (9.3)$$

The three predictors (x_1, x_2, x_3) represent the ensemble minimum, the ensemble median and the ensemble maximum. Second, conditional on the occurrence of nonzero precipitation, the 5th, 25th, 50th, 75th and 95th percentiles of the precipitation amount distributions are specified with separate regression equations, which each use the two ensemble quartiles as predictors. The final post-processed precipitation probabilities then are obtained through multiplicative law of probability:

$$Pr \{E_1 \cap E_2\} = P \{E_1|E_2\} Pr \{E_2\} \quad (9.4)$$

Here E_1 denotes the event when nonzero precipitation occurs and E_2 denotes the event when rainfall occurs. E_2 is defined by a combination of the forecast percentiles produced by the second regression step.

The Bayesian model averaging is closely linked to ensemble dressing. Like for the ensemble dressing method a probability distribution is applied. However the dressing kernels do not need to be the same for all ensemble members. Moreover different ensemble members receive different weights. The estimation of the weights is performed via maximum-likelihood methods (Hagedorn, 2010).

Another prominent calibration method is the non-homogeneous Gaussian regression. The non-homogeneous Gaussian regression yields probabilistic forecasts that take the form of Gaussian predictive PDFs, whereas the predictive mean is a bias-corrected weighted average of the ensemble member forecasts. The calibration technique is based on multiple linear regression and takes into account the spread/skill relationship. The following derivation of the non-homogeneous Gaussian regression is based on Gneiting et al. (2004):

Suppose that X_1, \dots, X_m denote an ensemble of individually distinguishable forecasts for an univariate weather event. A multiple linear regression equation for Y in terms of the ensemble member forecasts can be written as:

$$Y = a + b_1X_1 + \dots + b_mX_m + \epsilon \quad (9.5)$$

Here a and b_1, \dots, b_m are regression coefficients, whereas ϵ is an error term that averages to zero. Regression approaches of this type have been shown to improve the deterministic-style forecast accuracy of synoptic weather and seasonal climate ensembles (Krishnamurti et al. 1999). Standard regression theory suggests a straightforward way of constructing predictive PDFs and CDFs from a regression equation, by taking them to be Gaussian with predictive mean equal to the regression estimate and predictive variance mean equal to the mean squared prediction error for the training data. This approach corrects for model biases and takes account of dispersion errors. However, the resulting assessment of uncertainty is static, in that the predictive variance is independent of the ensemble spread, thereby negating the spread-skill relationship (Whitaker and Lough 1998). In an effort to account for the relation between model error and ensemble spread, the following linear function is used:

$$Var(\epsilon) = c + dS^2 \quad (9.6)$$

Here S^2 is the ensemble variance, and c and d are nonnegative coefficients. Combining both equations yields the Gaussian predictive distribution, whose mean derives from the regression equation and forms a bias-corrected weighted average of the ensemble member forecasts, and whose variance depends linearly on the ensemble variance:

$$N(a + b_1X_1 + \dots + b_mY_m, c + dS^2) \quad (9.7)$$

For estimating the regression coefficients, the method of minimum continuous ranked probability score estimation can be used. This technique finds the coefficient values that optimize the CRPS for the training data. The CRPS can be considered as a robust score and represents a general version of the MAE (Gneiting, 2004). The CRPS is defined as the integral of the Brier Score for all possible threshold values t of the continuous predictand (Hersbach 2000). Specifically, if F is the predictive CDF and y is the verifying observation, the continuous ranked probability score is defined as:

$$CRPS(F, y) = \int_{-\infty}^{\infty} [F(t) - H(t - y)]^2 dt \quad (9.8)$$

Here $H(t - y)$ denotes the Heaviside function and takes the value 0 when $t < y$ and the value 1 otherwise. However, when F is the CDF of a normal distribution with mean μ and variance σ^2 , repeated partial integration shows that:

$$CRPS[N(\mu, \sigma^2), y] = \sigma \left\{ \frac{y - \mu}{\sigma} \left[2\Phi\left(\frac{y - \mu}{\sigma}\right) - 1 \right] + 2\phi\left(\frac{y - \mu}{\sigma}\right) - \frac{1}{\sqrt{\pi}} \right\} \quad (9.9)$$

Here $\phi\left(\frac{y - \mu}{\sigma}\right)$ denotes the PDF and $\Phi\left(\frac{y - \mu}{\sigma}\right)$ denotes the CDF of the normal distribution with mean 0 and variance 1 evaluated at the normalized prediction error $\frac{y - \mu}{\sigma}$. In an effort to find the minimum CRPS we write the CRPS as an analytic function of the regression coefficients:

$$\Gamma(a; b_1; \dots; b_m; c; d) = \frac{1}{k} \sum_{i=1}^k (c + dS_i^2)^{1/2} \left\{ Z_i [2\Phi(Z_i) - 1] + 2\phi(Z_i) - \frac{1}{\sqrt{\pi}} \right\} \quad (9.10)$$

where $Z_i = \frac{Y_i - (a + b_1X_{i1} + \dots + b_mX_{im})}{(c + dS_i^2)^{1/2}}$ is a standardized forecast error, and where ϕ denotes the PDF and Φ denotes the CDF of a normal distribution with mean 0 and variance 1. According to Gneiting (2004) the coefficient values that minimize the CRPS numerically can be found using the Broyden-Fletcher-Goldfarb-Shanno algorithm (Press

et al. 1992). The application of this method shows encouraging results. According to Gneiting (2004) the technique was applied to 48-h sea level pressure mesoscale ensemble forecasts as well as surface temperature forecasts of the University of Washington. The method was tested for the North American Pacific region for a period in spring 2000. When the uncalibrated forecasts were compared to the bias-corrected ensemble, the calibrated forecasts of sea level pressure had root-mean-square error 9% less and mean absolute error 7% less. The predictive PDFs were sharp, and much better calibrated than the raw ensemble or the bias-corrected ensemble (Gneiting et al., 2004).

9.2 Previous investigations to calibrate ALADIN-LAEF

Since ALADIN-LAEF has been put into preoperational phase, it has been subject to several investigations. Like most of the ensemble prediction systems, ALADIN-LAEF reveals systematic deficiencies which can be linked to physical parameterization. As a consequence ALADIN-LAEF produces forecasts which constantly deviate from the observations. Results of preceding investigations show that the magnitude of the deviations varies remarkably for different parameters and forecast lead times. For some parameters such as the surface wind speed, the deviations are relatively small (range: ± 0.5 m/s). For others like the temperature forecasts, the systematic error has a notable impact on the forecasts. For these parameters it seems beneficial to remove the bias, before the forecasts are issued. As already mentioned in the previous section, several different methods to correct systematic deficiencies can be applied. In a preceding investigation at ZAMG, the calibration of temperature forecasts was tested. The main objective was to filter the potential added value of calibrated forecasts. For this purpose Kann et al. (2009) used a non-homogeneous Gaussian regression (NGR) method. This method is based on a multiple linear regression and corrects the first and second moment (mean bias and dispersion) for Gaussian-distributed continuous variables. Fitting the regression coefficients, a minimum continuous ranked probability score (CRPS) estimation has been chosen instead of the more traditional maximum likelihood technique (Kann et al., 2009). The CRPS is able to summarize the overall behavior of the ensemble compared to the observations. A detailed description of this technique can be found in the previous section. For the calculation of the CRPS the high-resolution analysis provided by INCA (Integrated Nowcasting through Comprehensive Analysis) serves as a reference. INCA incorporates observations as well as the ALADIN-AUSTRIA first guess field and interpolates it onto a 1km \times 1km data grid. In an effort to account for different seasonal characteristics such as convection or stationary frontal systems, 2 different periods were tested: December 2007 as well as July 2008. For example, applying the calibration to the 1st December 2007 00UTC, the training period consisted of the previous 30 days for a +6h projection (Kann, 2009). Results based on the data set for December 2007 show that the calibration removes the bias to a great extent. About half of the improvement achieved by full calibration can be attributed to the correction of the mean bias. According to Kann et al. (2009) the RMSE is reduced from about 3 to 2.4 K (see fig.9.1). Moreover the standard deviation of the ensemble members (from ensemble mean) is increased from 0.1-0.7 to 1.6 K. Thus, the calibrated ensemble seems to reflect the model uncertainty much better than the direct model output of the ensemble (spread-skill). Moreover results showed that the rank histograms became flatter and the percentage of outliers decreased. Beyond that the CRPS of the calibrated ensemble was reduced by 35%. Since it is considered that the length of the data set could have an impact on the quality of the calibration, different training sets

have been tested. Results show that in the case of the 2-m temperature, the length of the data has only marginal impact on calibration success. Different sliding periods of 30, 50 and 100 days were used and their results compared, but the differences between 30 and 50 days was negligible (Kann et al., 2009). Beyond that different tunings of the shape of the Gaussian PDF were tested. Therefore different rescaling factors have been applied, whereas rescaling the predictive PDF to half of the RMSE turned out to be most beneficial. Results show that rescaling the predictive PDF in order to obtain sharp forecasts is especially important for short range forecasts. When the same method was applied to the global ensemble (ECMWF-EPS) the improvement was of similar magnitude. However the calibrated 2 m temperature forecasts of ALADIN-LAEF still remained better than the 2 m temperature forecasts from the calibrated ECMWF-EPS (Kann, 2009). This result suggests that statistical downscaling of EPS cannot replace dynamical downscaling. Since the NGR method cannot account for temporal variations of the model error, another refined technique has been tested as well. This method called NGR TD (Non Homogeneous Gaussian Regression Time Decayed) incorporates a time-decaying averaging method. Instead of averaging over the entire period, the CRPS is adapted by a weighted mean of the prior error estimate and an update. Results show that a further improvement of the forecasts of about 5% can be obtained.

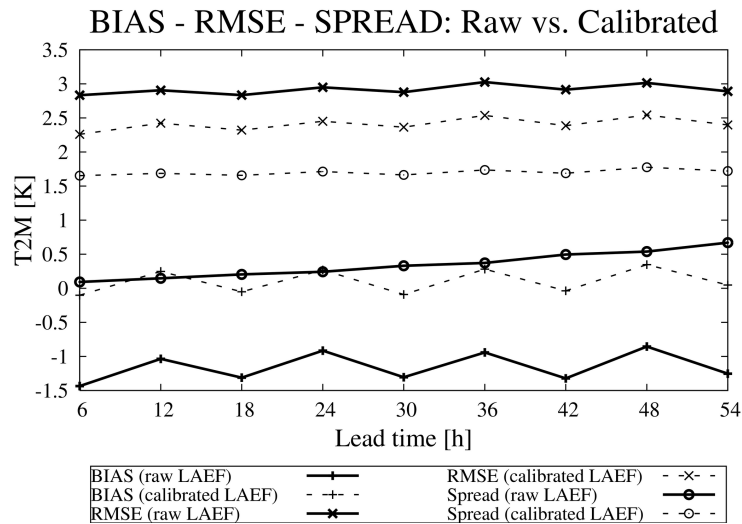


Figure 9.1: BIAS, RMSE and ensemble spread of the raw and calibrated ALADIN-LAEF system, source: Kann, 2009

Since precipitation forecasts generated by ALADIN-LAEF also reveal systematic errors, investigations in regard to effective calibration have been performed at ZAMG. Since the predictive PDF of precipitation does not reflect Gaussian shape, another calibration method than the NGR had to be found. Therefore Kalin et al. (2008) calibrated rainfall forecasts using logistic regression. The same method had already been tested at the ECMWF-EPS in 2007 (Hamill et. al) and proved to be successful. It should be noted that the calibration of precipitation forecasts is rather complicated due to the complex temporal and spatial distribution of rainfall. For the verification 6h precipitation forecasts were applied. For this purpose a 3 months data set (June - September 2007) was used. Results show significant improvement for lower thresholds (0.1 mm and 0.3 mm) (Kalin, 2009). However for larger thresholds (≥ 3 mm) no improvement could be achieved. Since the training period contained only a few heavy precipitation

events, a larger data set is considered to improve the skill of calibrated forecasts. For this purpose in 2009 another verification was performed. This time a notably larger data set was applied (almost 2 years of training data). Results showed a significant increase of skill of calibrated forecast compared to the raw ones. Moreover the positive impact of calibration was shifted to the higher thresholds. Still, the calibration failed for very high thresholds, even when the duration of precipitation event is reduced, e.g. to 3 hours. Since the quality and amount of data seems to be the critical part, further effort have to be made (Kalin, 2009).

Next to temperature- and precipitation forecasts, the calibration of other parameters such as geopotential or relative humidity should be analyzed more closely. As the following table (table9.1) demonstrates, the correction of the mean bias could notably improve the forecasts. Therefore this aspect will be subject to further investigations.

Parameter	Experiment	Bias (average)
geopotential 850 hPa	ECMWF-EPS	-28,26636
geopotential 850 hPa	BBSM	-2,28350
relative humidity 500 hPa	ECMWF-EPS	-1,63412
relative humidity 850 hPa	BBSM	-1,36375
sfc. temperature anomly	BBSM	-0,99255
relative humidity 500 hPa	BBSM	-0,82311
msl. pressure	ECMWF-EPS	-0,26160
sfc. wind speed	BBSM	-0,18288
wind speed 500 hPa	BBSM	-0,18058
relative humidity 850 hPa	ECMWF-EPS	-0,00590
sfc. temperature anomaly	ECMWF-EPS	0,01408
wind speed 500 hPa	ECMWF-EPS	0,01738
wind speed 850 hPa	BBSM	0,02843
temp anom. 850 hPa	BBSM	0,12784
temp anom. 850 hPa	ECMWF-EPS	0,13973
temp anom. 500 hPa	BBSM	0,14022
total precip.	ECMWF-EPS	0,16210
temp anom. 500 hPa	ECMWF-EPS	0,22305
wind speed 850 hPa	ECMWF-EPS	0,32763
total precip.	BBSM	0,39149
msl. pressure	BBSM	0,39525
sfc. wind speed	ECMWF-EPS	0,56527
geopotential 500 hPa	ECMWF-EPS	15,01890
geopotential 500 hPa	BBSM	29,41804

Table 9.1: bias table

10 Conclusion

Limited area ensemble prediction systems like ALADIN-LAEF have been developed to better predict high impact weather. Since they are operated on a smaller scale than global EPS, they are considered to better resolve phenomena such as high wind speeds or orographically induced rainfall. In an effort to filter the added value of ALADIN-LAEF in contrast to the global EPS of the ECMWF, a verification was performed over a 2 months period. ALADIN-LAEF differs from the global ECMWF-EPS in regard to the generation of initial perturbations as well as physical parameterization. The ECMWF-EPS applies a singular vector approach to generate initial perturbations. This method is based on an eigenvalue approach which determines the fastest growing perturbations within a selected time range. In contrast the generation of initial perturbation in ALADIN-LAEF is based on a method called BBSM (Breeding-Blending-Surface perturbation-Multiphysics). This method combines so called "breeding vectors" with singular vectors provided by the ECMWF-EPS. The breeding vectors are generated during a model "pre-run". The rationale for performing a model "pre-run" is the notion, that dominating errors evolve during the integration process, whereas others disappear. Breeding has proved to be beneficial during shorter forecast lead times (Buizza et al. 2005), whereas the singular vector method is considered to be superior for longer forecast lead times. Thus, the BBSM approach attempts to include both advantages. Apart from the generation of initial perturbations, ALADIN-LAEF and the ECMWF-EPS also differ in regard to physical parameterization. In the ECMWF-EPS physical configurations are based on stochastic physics, whereas in ALADIN-LAEF each ensemble member is linked to an individual physical parameterization scheme. Another main difference between both models is the horizontal resolution. ALADIN-LAEF operates on 18 km horizontal resolution, whereas the ECMWF-EPS was run on 50 km horizontal resolution in 2007. Beyond that both models differ in regard to the number of their ensemble members. ALADIN-LAEF consists of 16 different members, whereas the ECMWF-EPS contains 50 members. It should be considered that both increased horizontal resolution as well as an increased number of ensemble members occupy more computational time. Since operational models should be run within a reasonable time range, models settings are based on a good compromise.

In an effort to ensure a fair comparison, the first 16 members of the ECMWF-EPS were compared to the 16 members of ALADIN-LAEF. The performance of both models was compared using the following verification scores: Bias, RMSE, ensemble spread, Brier Score, Brier Skill Score, Ranked Probability Score, Ranked Probability Skill Score, Continuous Ranked Probability Score and Continuous Ranked Probability Skill Score. Moreover the quality was addressed in regard to reliability, hit rates (ROC) and rank distribution (Talagrand diagram). The verification of the surface fields was performed at observation locations. Forecast values were interpolated to the observation site for smoothly varying fields, such as 2 m temperature, 10 m wind speed and surface pressure. For precipitation which has strong spatial gradients, the observations were matched to the nearest grid point. For the verification of the upper level fields, the ECMWF analysis was used as a reference. Both analysis and forecasts were interpolated to a common regular 0.15×0.15 degree latitude/longitude grid. The verification

was performed for a two months period from 20th June to 20th August. The application of a 2-months period is fairly common for the verification of ensemble forecasts, since ensemble prediction systems produce remarkably large data sets.

Results of the verification of the surface fields reveal an overall better performance of ALADIN-LAEF. On average ALADIN-LAEF produces better mean sea level forecasts, wind speed forecasts as well as precipitation forecasts (see table.10.1). However it should be noted that the quality of the forecasts varies for different parameters and forecast lead times. In an effort to enable an objective comparison between both models, 16 main verification scores for the ensemble mean have been calculated. These scores are averaged values of the ensemble mean over all forecast lead times and thresholds. Therefore they only represent the mean performance and cannot account for specific characteristics. The comparison of the mean sea level pressure forecasts reveals a slightly better performance of ALADIN-LAEF. ALADIN-LAEF performs better in 9 of 16 scores. In contrast ALADIN-LAEF is clearly superior in regard to surface wind speed forecasts. ALADIN-LAEF scores better in 12 of 16 cases. The forecasts generated by ALADIN-LAEF show a remarkably smaller bias as well as better forecast probabilities (for example Brier Score). This result might be linked to a better incorporation of local characteristics such as topography. The verification of the rainfall forecasts also reveals a clearly better performance of ALADIN-LAEF. The comparison of 16 major scores indicates a better performance in 13 of 16 different cases. In ALADIN-LAEF the ratio RMSE/ensemble spread is notably closer to the perfect score of 1. Thus, the forecasts better account for the involved uncertainty. In contrast the 2 m- temperature forecasts in ALADIN-LAEF are inferior to those of the ECMWF-EPS. ALADIN-LAEF is significantly negatively biased, whereas the ECMWF-EPS shows hardly any bias. As a result the forecasts generated by ALADIN-LAEF are also less reliable. The comparison of 16 major scores reveals a better performance of the ECMWF-EPS in 14 cases. The rationale behind the inferior performance of ALADIN-LAEF in regard to temperature forecasts is a known problem and linked to the inconsistencies in the surface parameterization in ALADIN-LAEF.

In contrast to the verification of the surface field, the verification of the upper level parameters draws a completely different picture. For almost all parameters the ECMWF-EPS performs better (see table.10.1). Results show that the ECMWF-EPS produces clearly superior wind speed forecasts as well as humidity forecasts. Especially the 500hPa level wind speed forecasts of the ECMWF-EPS are of high quality: extremely low bias as well as a good matching of RMSE and ensemble spread. Thus, the comparison of both models reveals a clear dominant performance of the ECMWF-EPS. The ECMWF-EPS scores better in 15 of 16 different cases! The verification of the wind speed forecast for the 850 hPa field, also demonstrates a better overall performance of the ECMWF-EPS. Despite this overall better performance it should be noted that ALADIN-LAEF exhibits a notably smaller bias. The comparison of the humidity forecasts also reveals a better performance of the ECMWF-EPS. Although the ALADIN-LAEF forecasts are generally more accurate (smaller bias), it produces notably more outliers. In consequence the forecast probability (Brier Score) is better in the ECMWF-EPS. Therefore the ECMWF-EPS scores better in 13 of 16 different cases. Beyond that the ECMWF-EPS produces clearly superior humidity forecasts for the 850 hPa level. The verification of the geopotential forecasts reveals opposed results. The ECMWF-EPS generates better forecasts for the 500hPa level, whereas ALADIN-LAEF produces better 850 hPa level forecasts. Both comparisons reveal a clear winner (see table.10.1). Like the comparison of the geopotential forecasts, the comparison of the temperature forecasts reveals opposed results. Thus, ALADIN-LAEF produces better 500hPa temperature forecasts,

whereas the ECMWF-EPS produces better 850 hPa forecasts. These results might seem a bit surprising, since ALADIN-LAEF could be expected to generate better medium level (850 hPa) forecasts. Although the temperature forecasts of ALADIN-LAEF seem to be generally more accurate (see bias), the forecast probability (see Brier Score) resulting from the complete set of members, is better in ALADIN-LAEF. Thus, the ECMWF-EPS performs better in 11 of 16 different cases. In contrast ALADIN-LAEF produces better temperature forecasts for the 500hPa level. This result should be linked to the notable larger bias of the ECMWF-EPS that affects the forecasts.

Preceding investigations on how to post-process the forecasts have shown promising results. For example the correction of the mean bias in the temperature forecasts increased the forecast quality significantly. The verification of the calibrated forecasts showed that the CRPS could be reduced by 35% (Kann, 2009). The calibration of precipitation forecasts also proved that post-processing can notably increase the skill of forecasts (Kalin, 2009). Although the main objective in numerical models is to reflect the true conditions as best as possible, statistical post-processing represents a final option in optimizing forecasts. After a long period of development and preoperational phase, ALADIN-LAEF will be put into operational mode presumably in April 2011. Based on the increasing importance of limited area ensemble prediction system, additional European countries such as Turkey will step in. Thus, the domain is going to be enlarged (see fig.10.1). In this graphic the green border represents the current domain, and the blue border the new one. Further plans for the future are to increase the horizontal resolution from 18 km to 11 km. Beyond that physical parameterizations will be adapted. Since stochastic physics have proved to be very successful, the multi-physics scheme in ALADIN-LAEF will be revised and stochastic physics will be adapted. At present stochastic physics are tested for the surface fields at ZAMG. Results of the verification will direct further steps in the development.

Parameter	ALADIN-LAEF	ECMWF-EPS
mean sea level pressure	9	7
sfc. temperature anomaly	2	14
sfc. wind speed	12	4
total precipitation	13	3
surface total	36	28
Geopotential in 500hPa	2	6
Geopotential in 850hPa	7	1
Temperature Anomaly in 500hPa	13	3
Temperature Anomaly in 850hPa	5	11
Wind Speed in 500hPa	1	15
Wind Speed in 850hPa	4	12
Relative Humidity in 500hPa	3	13
Relative Humidity in 850hPa	1	15
upper level total	36	76
Total	72	104

Table 10.1: final results based on 16 major scores

ALADIN-LAEF (old:G, new:B) vs GLAMEPS (R)

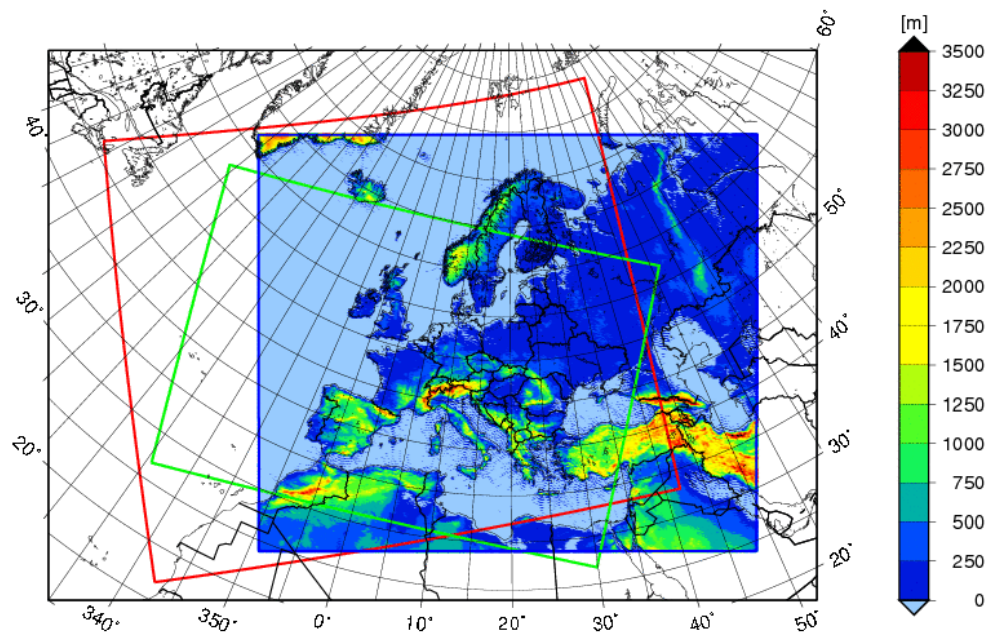


Figure 10.1: Future domain of ALADIN-LAEF

11 Appendix

Since the involved data set is rather large (110 GB), only a well chosen selection can be presented here. However the complete data set is available on the attached data-CD. In an effort to highlight the general performance of both EPS, the average performance has been calculated. In a first step the performance of the ensemble mean has been analyzed for each parameter. Therefore several scores such as bias, Brier Score etc. were calculated for several forecast lead times and selected thresholds. In a second step, the performance averaged over all forecast lead times and selected thresholds has been computed. This average performance enables a general comparison of both EPS. Apparently this average performance is not able to consider particular characteristics linked to specific forecast lead times and thresholds. For this purpose the performance of both EPS has been analyzed closely in the chapter 8 ("Results"). The table here works as follows: The first box called "Final Results- scores averaged over all thresholds" enables a fair comparison between both models. Thus, each scores has the same impact (1/16) on the final result. The second box called "Detailed Results- scores for individual thresholds" enables a closer assessment of specific characteristics. In the following table the detailed comparison is represented. For each parameter the scores are presented in alphabetical order.

- The first column contains the different scores.
- The second column represents the involved ensemble prediction system. BBSM corresponds to ALADIN-LAEF (as explained in the chapter 5).
- The third column shows the largest involved forecast lead times. For example "fc.time 54" means that the performance has been averaged over the forecast lead times ranging from 6 hours (start point) to 54 hours.
- The 4th column shows if a particular threshold has been applied.
- The 5th column shows the **averaged value for the selected score**. For example in the 1.row, the 5.column contains the bias of the ensemble mean, averaged over all forecast lead times.
- The 6th and 7th column have been attached in order to compare the general performance of both EPS. That ensemble prediction system, which is **closer to the perfect score is then linked to 1**, whereas the other EPS is linked to zero. Finally the results for the 6th and 7th column are added up for each parameter and that EPS, which has achieved the higher number of "victories", is considered to perform better.

FINAL RESULTS - scores averaged over all thresholds		
Surface - 16 scores	BBSM	ECMWF-EPS
msl. Pressure	9	7
temperature anomaly	2	14
wind speed	12	4
total precipitation	13	3
	36	28
Upper level - 16 scores	BBSM	ECMWF-EPS
temperature anom. 500hPa	13	3
temperature anom. 850hPa	5	11
geopotential 500hPa	2	6
geopotential 850hPa	7	1
wind speed 500hPa	1	15
wind speed 850hPa	4	12
relative humidity 500hPa	3	13
relative humidity 850hPa	1	15
	36	76

Detailed Results - scores for individual thresholds		
Surface	BBSM	ECMWF-EPS
msl. pressure	33	47
temperature anomaly	27	55
wind speed	71	35
total precipitation	54	15
	185	152
Upper Level	BBSM	ECMWF-EPS
Geopotential 500	2	6
Geopotential 850	7	1
Temp. Anomaly 500	61	21
Temp. Anomaly 850	29	53
Wind Speed 500	16	87
Wind Speed 850	32	71
Rel. Humidity 500	24	58
Rel. Humidity 850	9	73
	180	370

Parameter: Mean Sea Level Pressure						
Score	Experiment	fc.time	thr	Average (thr, time)	BBSM	ECMWF-EPS
BIAS	ECMWF_EPS	54	No	-0,26160		
BIAS	BBSM	54	No	0,39525	0	1
BRIERSCOREEPS	BBSM	54	1005	0,01885		
BRIERSCOREEPS	ECMWF_EPS	54	1005	0,02293	1	0
BRIERSCOREEPS	BBSM	54	1010	0,05529		
BRIERSCOREEPS	ECMWF_EPS	54	1010	0,05907	1	0
BRIERSCOREEPS	BBSM	54	1015	0,05912		
BRIERSCOREEPS	ECMWF_EPS	54	1015	0,05962	1	0
BRIERSSANA	BBSM	54	1015	0,79103		
BRIERSSANA	ECMWF_EPS	54	1015	0,79114	1	0
BRIERSSANA	ECMWF_EPS	54	1010	0,91512		
BRIERSSANA	BBSM	54	1010	0,92017	0	1
BRIERSSANA	ECMWF_EPS	54	1005	0,97541		
BRIERSSANA	BBSM	54	1005	0,97978	0	1
BRIERSSREF	ECMWF_EPS	54	1005	0,01988		

BRIERSSREF	BBSM	54	1005	0,18107	0	1
BRIERSSREF	ECMWF_EPS	54	1010	0,21035		
BRIERSSREF	BBSM	54	1010	0,27439	0	1
BRIERSSREF	BBSM	54	1015	0,32410		
BRIERSSREF	ECMWF_EPS	54	1015	0,32410	0	0
CRPSEPS	BBSM	54	No	0,79597		
CRPSEPS	ECMWF_EPS	54	No	0,86648	1	0
CRPSREF	BBSM	54	No	1,12457		
CRPSREF	ECMWF_EPS	54	No	1,12457	0	0
CRPSSANA	ECMWF_EPS	54	No	0,99915		
CRPSSANA	BBSM	54	No	0,99921	0	0
CRPSSREF	ECMWF_EPS	54	No	0,21994		
CRPSSREF	BBSM	54	No	0,28039	0	1
OUTLIERS	ECMWF_EPS	54	No	0,20862		
OUTLIERS	BBSM	54	No	0,24707	0	1
RELIABILITY	BBSM	6	1005	0,33493		
RELIABILITY	ECMWF_EPS	6	1005	0,51219	0	1
RELIABILITY	BBSM	12	1005	0,46160		
RELIABILITY	ECMWF_EPS	12	1005	0,58050	1	0
RELIABILITY	BBSM	18	1005	0,46432		
RELIABILITY	ECMWF_EPS	18	1005	0,69792	1	0
RELIABILITY	BBSM	24	1005	0,56157		
RELIABILITY	ECMWF_EPS	24	1005	0,78595	1	0
RELIABILITY	BBSM	30	1005	0,37542		
RELIABILITY	ECMWF_EPS	30	1005	0,70572	1	0
RELIABILITY	BBSM	36	1005	0,41171		
RELIABILITY	ECMWF_EPS	36	1005	0,66237	1	0
RELIABILITY	BBSM	42	1005	0,41101		
RELIABILITY	ECMWF_EPS	42	1005	0,68590	1	0
RELIABILITY	BBSM	48	1005	0,42187		
RELIABILITY	ECMWF_EPS	48	1005	0,68285	1	0
RELIABILITY	BBSM	54	1005	0,34123		
RELIABILITY	ECMWF_EPS	54	1005	0,64947	0	1
RELIABILITY	BBSM	6	1010	0,29479		
RELIABILITY	ECMWF_EPS	6	1010	0,49905	0	1
RELIABILITY	BBSM	12	1010	0,42407		
RELIABILITY	ECMWF_EPS	12	1010	0,47822	0	1
RELIABILITY	BBSM	18	1010	0,36300		
RELIABILITY	ECMWF_EPS	18	1010	0,55973	0	1
RELIABILITY	BBSM	24	1010	0,58125		
RELIABILITY	ECMWF_EPS	24	1010	0,74917	1	0
RELIABILITY	BBSM	30	1010	0,41050		
RELIABILITY	ECMWF_EPS	30	1010	0,66872	1	0
RELIABILITY	BBSM	36	1010	0,37410		
RELIABILITY	ECMWF_EPS	36	1010	0,51928	0	1
RELIABILITY	BBSM	42	1010	0,39013		
RELIABILITY	ECMWF_EPS	42	1010	0,57246	1	0
RELIABILITY	BBSM	48	1010	0,48471		
RELIABILITY	ECMWF_EPS	48	1010	0,67687	1	0
RELIABILITY	BBSM	54	1010	0,38584		
RELIABILITY	ECMWF_EPS	54	1010	0,63651	1	0
RELIABILITY	BBSM	6	1015	0,28198		
RELIABILITY	ECMWF_EPS	6	1015	0,51583	0	1

RELIABILITY	BBSM	12	1015	0,32529		
RELIABILITY	ECMWF_EPS	12	1015	0,34648	0	1
RELIABILITY	BBSM	18	1015	0,30320		
RELIABILITY	ECMWF_EPS	18	1015	0,38463	0	1
RELIABILITY	BBSM	24	1015	0,54705		
RELIABILITY	ECMWF_EPS	24	1015	0,64009	1	0
RELIABILITY	BBSM	30	1015	0,42451		
RELIABILITY	ECMWF_EPS	30	1015	0,58854	1	0
RELIABILITY	BBSM	36	1015	0,32672		
RELIABILITY	ECMWF_EPS	36	1015	0,37617	0	1
RELIABILITY	BBSM	42	1015	0,35280		
RELIABILITY	ECMWF_EPS	42	1015	0,43203	0	1
RELIABILITY	BBSM	48	1015	0,51970		
RELIABILITY	ECMWF_EPS	48	1015	0,59429	1	0
RELIABILITY	BBSM	54	1015	0,43656		
RELIABILITY	ECMWF_EPS	54	1015	0,58309	1	0
RMSE	BBSM	54	No	1,50404		
RMSE	ECMWF_EPS	54	No	1,58043	1	0
ROC_AREA	BBSM	6	1005	0,94903		
ROC_AREA	ECMWF_EPS	6	1005	0,98406	0	1
ROC_AREA	BBSM	6	1010	0,94078		
ROC_AREA	ECMWF_EPS	6	1010	0,97968	0	1
ROC_AREA	BBSM	6	1015	0,96776		
ROC_AREA	ECMWF_EPS	6	1015	0,98302	0	1
ROC_AREA	BBSM	12	1005	0,89959		
ROC_AREA	ECMWF_EPS	12	1005	0,93256	0	1
ROC_AREA	BBSM	12	1010	0,90708		
ROC_AREA	ECMWF_EPS	12	1010	0,93178	0	1
ROC_AREA	BBSM	12	1015	0,94840		
ROC_AREA	ECMWF_EPS	12	1015	0,95951	0	1
ROC_AREA	BBSM	18	1005	0,90361		
ROC_AREA	ECMWF_EPS	18	1005	0,95208	0	1
ROC_AREA	BBSM	18	1010	0,90835		
ROC_AREA	ECMWF_EPS	18	1010	0,93733	0	1
ROC_AREA	BBSM	18	1015	0,94737		
ROC_AREA	ECMWF_EPS	18	1015	0,95520	0	1
ROC_AREA	BBSM	24	1005	0,96694		
ROC_AREA	ECMWF_EPS	24	1005	0,98191	0	1
ROC_AREA	BBSM	24	1010	0,97674		
ROC_AREA	ECMWF_EPS	24	1010	0,97688	0	1
ROC_AREA	BBSM	24	1015	0,97378		
ROC_AREA	ECMWF_EPS	24	1015	0,96930	1	0
ROC_AREA	BBSM	30	1005	0,95235		
ROC_AREA	ECMWF_EPS	30	1005	0,98494	0	1
ROC_AREA	BBSM	30	1010	0,97061		
ROC_AREA	ECMWF_EPS	30	1010	0,98185	0	1
ROC_AREA	BBSM	30	1015	0,97951		
ROC_AREA	ECMWF_EPS	30	1015	0,97696	1	0
ROC_AREA	BBSM	36	1005	0,86386		
ROC_AREA	ECMWF_EPS	36	1005	0,94865	0	1
ROC_AREA	BBSM	36	1010	0,90250		
ROC_AREA	ECMWF_EPS	36	1010	0,94447	0	1
ROC_AREA	BBSM	36	1015	0,95107		

ROC_AREA	ECMWF_EPS	36	1015	0,95089	1	0
ROC_AREA	BBSM	42	1005	0,88582		
ROC_AREA	ECMWF_EPS	42	1005	0,92915	0	1
ROC_AREA	BBSM	42	1010	0,91832		
ROC_AREA	ECMWF_EPS	42	1010	0,94526	0	1
ROC_AREA	BBSM	42	1015	0,95045		
ROC_AREA	ECMWF_EPS	42	1015	0,94978	1	0
ROC_AREA	BBSM	48	1005	0,95758		
ROC_AREA	ECMWF_EPS	48	1005	0,97755	0	1
ROC_AREA	BBSM	48	1010	0,96808		
ROC_AREA	ECMWF_EPS	48	1010	0,97254	0	1
ROC_AREA	BBSM	48	1015	0,96541		
ROC_AREA	ECMWF_EPS	48	1015	0,96762	0	1
ROC_AREA	BBSM	54	1005	0,94314		
ROC_AREA	ECMWF_EPS	54	1005	0,97669	0	1
ROC_AREA	BBSM	54	1010	0,96118		
ROC_AREA	ECMWF_EPS	54	1010	0,97205	0	1
ROC_AREA	BBSM	54	1015	0,97043		
ROC_AREA	ECMWF_EPS	54	1015	0,97145	0	1
RPSEPS	BBSM	54	No	0,04442		
RPSEPS	ECMWF_EPS	54	No	0,04720	1	0
RPSSANA	ECMWF_EPS	54	No	0,92624		
RPSSANA	BBSM	54	No	0,93041	0	1
RPSSREF	ECMWF_EPS	54	No	0,24256		
RPSSREF	BBSM	54	No	0,28654	1	0
SPREAD	ECMWF_EPS	54	No	0,89237		
SPREAD	BBSM	54	No	0,89853	1	0
TALAGRAND	BBSM	6	No	0,33261		
TALAGRAND	ECMWF_EPS	6	No	0,07916	1	0
TALAGRAND	BBSM	12	No	0,28283		
TALAGRAND	ECMWF_EPS	12	No	0,18628	0	1
TALAGRAND	BBSM	18	No	0,28661		
TALAGRAND	ECMWF_EPS	18	No	0,16527	0	1
TALAGRAND	BBSM	24	No	0,07702		
TALAGRAND	ECMWF_EPS	24	No	0,04100	1	0
TALAGRAND	BBSM	30	No	0,11754		
TALAGRAND	ECMWF_EPS	30	No	0,03551	0	1
TALAGRAND	BBSM	36	No	0,23166		
TALAGRAND	ECMWF_EPS	36	No	0,13324	0	1
TALAGRAND	BBSM	42	No	0,19492		
TALAGRAND	ECMWF_EPS	42	No	0,09749	0	1
TALAGRAND	BBSM	48	No	0,06398		
TALAGRAND	ECMWF_EPS	48	No	0,02106	1	0
TALAGRAND	BBSM	54	No	0,09491		
TALAGRAND	ECMWF_EPS	54	No	0,01493	1	0

33

47

Parameter: Temperature Anomaly						
Score	Experiment	fc.time	thr	Average (thr, time)	BBSM	ECMWF-EPS
BIAS	BBSM	54	No	-0,99255		
BIAS	ECMWF_EPS	54	No	0,01408	0	1
BRIERSCOREEPS	BBSM	54	2	0,13533		
BRIERSCOREEPS	ECMWF_EPS	54	2	0,13967	1	0
BRIERSCOREEPS	ECMWF_EPS	54	-2	0,13984		

BRIERSCOREEPS	BBSM	54	-2	0,15113	0	1
BRIERSCOREEPS	BBSM	54	0	0,15649		
BRIERSCOREEPS	ECMWF_EPS	54	0	0,15688	1	0
BRIERSSANA	ECMWF_EPS	54	2	0,60180		
BRIERSSANA	BBSM	54	2	0,61448	0	1
BRIERSSANA	ECMWF_EPS	54	0	0,70807		
BRIERSSANA	BBSM	54	0	0,70873	0	1
BRIERSSANA	BBSM	54	-2	0,78748		
BRIERSSANA	ECMWF_EPS	54	-2	0,80328	1	0
BRIERSSREF	BBSM	54	-2	0,10572		
BRIERSSREF	ECMWF_EPS	54	-2	0,17255	1	0
BRIERSSREF	ECMWF_EPS	54	0	0,13075		
BRIERSSREF	BBSM	54	0	0,13321	0	1
BRIERSSREF	ECMWF_EPS	54	2	0,13501		
BRIERSSREF	BBSM	54	2	0,16272	0	1
CRPSEPS	ECMWF_EPS	54	No	1,84631		
CRPSEPS	BBSM	54	No	2,07846	0	1
CRPSREF	BBSM	54	No	2,38244		
CRPSREF	ECMWF_EPS	54	No	2,38244	0	0
CRPSSANA	BBSM	54	No	0,88421		
CRPSSANA	ECMWF_EPS	54	No	0,89789	0	1
CRPSSREF	BBSM	54	No	0,12504		
CRPSSREF	ECMWF_EPS	54	No	0,22498	0	1
OUTLIERS	ECMWF_EPS	54	No	0,52205		
OUTLIERS	BBSM	54	No	0,53133	0	1
RELIABILITY	BBSM	6	-2	0,34230		
RELIABILITY	ECMWF_EPS	6	-2	0,43326	0	1
RELIABILITY	ECMWF_EPS	6	0	0,54068		
RELIABILITY	BBSM	6	0	0,60265	0	1
RELIABILITY	BBSM	6	2	0,36087		
RELIABILITY	ECMWF_EPS	6	2	0,46787	0	1
RELIABILITY	BBSM	12	-2	0,38001		
RELIABILITY	ECMWF_EPS	12	-2	0,45586	0	1
RELIABILITY	ECMWF_EPS	12	0	0,58575		
RELIABILITY	BBSM	12	0	0,59753	0	1
RELIABILITY	BBSM	12	2	0,43570		
RELIABILITY	ECMWF_EPS	12	2	0,46644	0	1
RELIABILITY	BBSM	18	-2	0,44476		
RELIABILITY	ECMWF_EPS	18	-2	0,50916	0	1
RELIABILITY	ECMWF_EPS	18	0	0,51821		
RELIABILITY	BBSM	18	0	0,58267	0	1
RELIABILITY	BBSM	18	2	0,44434		
RELIABILITY	ECMWF_EPS	18	2	0,50510	0	1
RELIABILITY	BBSM	24	-2	0,42481		
RELIABILITY	ECMWF_EPS	24	-2	0,50160	0	1
RELIABILITY	ECMWF_EPS	24	0	0,45740		
RELIABILITY	BBSM	24	0	0,50753	1	0
RELIABILITY	BBSM	24	2	0,41957		
RELIABILITY	ECMWF_EPS	24	2	0,45608	0	1
RELIABILITY	BBSM	30	-2	0,39397		
RELIABILITY	ECMWF_EPS	30	-2	0,46273	0	1
RELIABILITY	ECMWF_EPS	30	0	0,52608		
RELIABILITY	BBSM	30	0	0,56255	0	1

RELIABILITY	BBSM	30	2	0,39615		
RELIABILITY	ECMWF_EPS	30	2	0,47631	0	1
RELIABILITY	BBSM	36	-2	0,38750		
RELIABILITY	ECMWF_EPS	36	-2	0,47529	0	1
RELIABILITY	ECMWF_EPS	36	0	0,56429		
RELIABILITY	BBSM	36	0	0,57274	0	1
RELIABILITY	BBSM	36	2	0,46696		
RELIABILITY	ECMWF_EPS	36	2	0,48294	0	1
RELIABILITY	BBSM	42	-2	0,45952		
RELIABILITY	ECMWF_EPS	42	-2	0,52516	0	1
RELIABILITY	ECMWF_EPS	42	0	0,49671		
RELIABILITY	BBSM	42	0	0,56190	0	1
RELIABILITY	BBSM	42	2	0,45501		
RELIABILITY	ECMWF_EPS	42	2	0,52192	0	1
RELIABILITY	BBSM	48	-2	0,42595		
RELIABILITY	ECMWF_EPS	48	-2	0,48985	0	1
RELIABILITY	ECMWF_EPS	48	0	0,45567		
RELIABILITY	BBSM	48	0	0,50524	1	0
RELIABILITY	BBSM	48	2	0,43192		
RELIABILITY	ECMWF_EPS	48	2	0,45927	0	1
RELIABILITY	BBSM	54	-2	0,38459		
RELIABILITY	ECMWF_EPS	54	-2	0,46031	0	1
RELIABILITY	ECMWF_EPS	54	0	0,51149		
RELIABILITY	BBSM	54	0	0,53945	0	1
RELIABILITY	BBSM	54	2	0,43125		
RELIABILITY	ECMWF_EPS	54	2	0,46484	0	1
RMSE	BBSM	54	No	2,93140		
RMSE	ECMWF_EPS	54	No	2,97521	0	1
RPSANA	BBSM	54	No	0,53410		
RPSANA	ECMWF_EPS	54	No	0,53410	0	0
RPSEPS	ECMWF_EPS	54	No	0,14546		
RPSEPS	BBSM	54	No	0,14765	0	1
RPSREF	BBSM	54	No	0,17082		
RPSREF	ECMWF_EPS	54	No	0,17082	0	0
RPSSANA	BBSM	54	No	0,72351		
RPSSANA	ECMWF_EPS	54	No	0,72763	0	1
RPSSREF	BBSM	54	No	0,13353		
RPSSREF	ECMWF_EPS	54	No	0,14601	0	1
SPREAD	ECMWF_EPS	54	No	0,72694		
SPREAD	BBSM	54	No	0,92181	1	0
TALAGRAND	BBSM	6	No	0,21708		
TALAGRAND	ECMWF_EPS	6	No	0,31259	1	0
TALAGRAND	BBSM	12	No	0,14807		
TALAGRAND	ECMWF_EPS	12	No	0,21147	1	0
TALAGRAND	BBSM	18	No	0,15799		
TALAGRAND	ECMWF_EPS	18	No	0,23512	1	0
TALAGRAND	BBSM	24	No	0,24678		
TALAGRAND	ECMWF_EPS	24	No	0,32891	1	0
TALAGRAND	BBSM	30	No	0,16760		
TALAGRAND	ECMWF_EPS	30	No	0,24552	1	0
TALAGRAND	BBSM	36	No	0,11321		
TALAGRAND	ECMWF_EPS	36	No	0,15773	1	0
TALAGRAND	BBSM	42	No	0,12954		

TALAGRAND	ECMWF_EPS	42	No	0,18116	1	0
TALAGRAND	BBSM	48	No	0,20137		
TALAGRAND	ECMWF_EPS	48	No	0,26458	1	0
TALAGRAND	BBSM	54	No	0,14295		
TALAGRAND	ECMWF_EPS	54	No	0,19174	1	0
ROC_AREA	BBSM	6	-2	0,82824		
ROC_AREA	ECMWF_EPS	6	-2	0,81932	1	0
ROC_AREA	BBSM	6	0	0,85443		
ROC_AREA	ECMWF_EPS	6	0	0,82920	1	0
ROC_AREA	BBSM	6	2	0,83782		
ROC_AREA	ECMWF_EPS	6	2	0,83467	1	0
ROC_AREA	BBSM	12	-2	0,83994		
ROC_AREA	ECMWF_EPS	12	-2	0,85056	0	1
ROC_AREA	BBSM	12	0	0,86113		
ROC_AREA	ECMWF_EPS	12	0	0,85392	1	0
ROC_AREA	BBSM	12	2	0,84101		
ROC_AREA	ECMWF_EPS	12	2	0,85271	0	1
ROC_AREA	BBSM	18	-2	0,83669		
ROC_AREA	ECMWF_EPS	18	-2	0,85653	0	1
ROC_AREA	BBSM	18	0	0,86182		
ROC_AREA	ECMWF_EPS	18	0	0,85249	1	0
ROC_AREA	BBSM	18	2	0,83746		
ROC_AREA	ECMWF_EPS	18	2	0,85617	0	1
ROC_AREA	BBSM	24	-2	0,81695		
ROC_AREA	ECMWF_EPS	24	-2	0,81601	1	0
ROC_AREA	BBSM	24	0	0,83193		
ROC_AREA	ECMWF_EPS	24	0	0,81578	1	0
ROC_AREA	BBSM	24	2	0,82707		
ROC_AREA	ECMWF_EPS	24	2	0,83660	0	1
ROC_AREA	BBSM	30	-2	0,84462		
ROC_AREA	ECMWF_EPS	30	-2	0,84394	1	0
ROC_AREA	BBSM	30	0	0,86260		
ROC_AREA	ECMWF_EPS	30	0	0,84937	1	0
ROC_AREA	BBSM	30	2	0,84854		
ROC_AREA	ECMWF_EPS	30	2	0,85391	0	1
ROC_AREA	BBSM	36	-2	0,85139		
ROC_AREA	ECMWF_EPS	36	-2	0,86744	0	1
ROC_AREA	BBSM	36	0	0,86313		
ROC_AREA	ECMWF_EPS	36	0	0,86644	0	1
ROC_AREA	BBSM	36	2	0,84842		
ROC_AREA	ECMWF_EPS	36	2	0,86483	0	1
ROC_AREA	BBSM	42	-2	0,84870		
ROC_AREA	ECMWF_EPS	42	-2	0,86907	0	1
ROC_AREA	BBSM	42	0	0,86609		
ROC_AREA	ECMWF_EPS	42	0	0,86811	0	1
ROC_AREA	BBSM	42	2	0,84608		
ROC_AREA	ECMWF_EPS	42	2	0,86760	0	1
ROC_AREA	BBSM	48	-2	0,82237		
ROC_AREA	ECMWF_EPS	48	-2	0,82569	0	1
ROC_AREA	BBSM	48	0	0,83261		
ROC_AREA	ECMWF_EPS	48	0	0,82726	1	0
ROC_AREA	BBSM	48	2	0,82959		
ROC_AREA	ECMWF_EPS	48	2	0,84376	0	1

ROC_AREA	BBSM	54	-2	0,85052		
ROC_AREA	ECMWF_EPS	54	-2	0,85228	0	1
ROC_AREA	BBSM	54	0	0,86435		
ROC_AREA	ECMWF_EPS	54	0	0,86041	1	0
ROC_AREA	BBSM	54	2	0,85279		
ROC_AREA	ECMWF_EPS	54	2	0,86347	0	1
				82	27	55
Parameter: Wind Speed						
Score	Experiment	fc.time	thr	Average (thr, time)	BBSM	ECMWF-EPS
BIAS	BBSM	54	No	-0,18288		
BIAS	ECMWF_EPS	54	No	0,56527	1	0
BRIERSCOREANA	BBSM	54	6	0,09987		
BRIERSCOREANA	ECMWF_EPS	54	6	0,09987	0	0
BRIERSCOREANA	ECMWF_EPS	54	4	0,26612		
BRIERSCOREANA	BBSM	54	4	0,26676	0	1
BRIERSCOREANA	BBSM	54	2	0,58965		
BRIERSCOREANA	ECMWF_EPS	54	2	0,58965	0	0
BRIERSCOREANA	BBSM	54	1	0,81406		
BRIERSCOREANA	ECMWF_EPS	54	1	0,81406	0	0
BRIERSCOREEPS	BBSM	54	6	0,07949		
BRIERSCOREEPS	ECMWF_EPS	54	6	0,10191	0	1
BRIERSCOREEPS	BBSM	54	1	0,16697		
BRIERSCOREEPS	ECMWF_EPS	54	1	0,16710	1	0
BRIERSCOREEPS	BBSM	54	4	0,17112		
BRIERSCOREEPS	ECMWF_EPS	54	4	0,19608	1	0
BRIERSCOREEPS	BBSM	54	2	0,23364		
BRIERSCOREEPS	ECMWF_EPS	54	2	0,25264	1	0
BRIERSCOREREF	ECMWF_EPS	54	6	0,09491		
BRIERSCOREREF	BBSM	54	6	0,21252	0	1
BRIERSCOREREF	BBSM	54	4	0,21350		
BRIERSCOREREF	ECMWF_EPS	54	4	0,21350	0	0
BRIERSCOREREF	BBSM	54	1	0,23487		
BRIERSCOREREF	ECMWF_EPS	54	1	0,23487	0	0
BRIERSCOREREF	BBSM	54	2	0,30680		
BRIERSCOREREF	ECMWF_EPS	54	2	0,30680	0	0
BRIERSSANA	ECMWF_EPS	54	6	-0,04275		
BRIERSSANA	BBSM	54	6	0,19492	1	0
BRIERSSANA	ECMWF_EPS	54	4	0,23057		
BRIERSSANA	BBSM	54	4	0,34292	1	0
BRIERSSANA	ECMWF_EPS	54	2	0,54863		
BRIERSSANA	BBSM	54	2	0,58948	1	0
BRIERSSANA	ECMWF_EPS	54	1	0,78676		
BRIERSSANA	BBSM	54	1	0,78804	1	0
BRIERSSREF	ECMWF_EPS	54	2	0,17434		
BRIERSSREF	BBSM	54	2	0,23643	1	0
BRIERSSREF	ECMWF_EPS	54	4	0,06667		
BRIERSSREF	BBSM	54	4	0,19958	1	0
BRIERSSREF	ECMWF_EPS	54	6	-0,08440		
BRIERSSREF	BBSM	54	6	0,16231	1	0
BRIERSSREF	BBSM	54	1	0,29032		
BRIERSSREF	ECMWF_EPS	54	1	0,30569	0	1
CRPSEPS	BBSM	54	No	1,14423		
CRPSEPS	ECMWF_EPS	54	No	1,31039	1	0

CRPSREF	ECMWF_EPS	54	No	1,43073		
CRPSREF	BBSM	54	No	1,44133	0	1
CRPSSANA	ECMWF_EPS	54	No	0,53818		
CRPSSANA	BBSM	54	No	0,60051	0	1
CRPSSREF	ECMWF_EPS	54	No	0,08673		
CRPSSREF	BBSM	54	No	0,20530	0	1
OUTLIERS	ECMWF_EPS	54	No	0,44889		
OUTLIERS	BBSM	54	No	0,48333	0	1
RELIABILITY	ECMWF_EPS	6	1	0,57052		
RELIABILITY	BBSM	6	1	0,61237	0	1
RELIABILITY	ECMWF_EPS	12	1	0,85476		
RELIABILITY	BBSM	12	1	0,86944	0	1
RELIABILITY	ECMWF_EPS	18	1	0,73459		
RELIABILITY	BBSM	18	1	0,74177	0	1
RELIABILITY	ECMWF_EPS	24	1	0,57978		
RELIABILITY	BBSM	24	1	0,61269	0	1
RELIABILITY	ECMWF_EPS	30	1	0,56355		
RELIABILITY	BBSM	30	1	0,61602	0	1
RELIABILITY	ECMWF_EPS	36	1	0,81662		
RELIABILITY	BBSM	36	1	0,87685	0	1
RELIABILITY	ECMWF_EPS	42	1	0,74506		
RELIABILITY	BBSM	42	1	0,74808	0	1
RELIABILITY	ECMWF_EPS	48	1	0,57371		
RELIABILITY	BBSM	48	1	0,61792	0	1
RELIABILITY	ECMWF_EPS	54	1	0,55276		
RELIABILITY	BBSM	54	1	0,62503	0	1
RELIABILITY	ECMWF_EPS	6	2	0,33322		
RELIABILITY	BBSM	6	2	0,44654	1	0
RELIABILITY	ECMWF_EPS	12	2	0,62574		
RELIABILITY	BBSM	12	2	0,66826	0	1
RELIABILITY	ECMWF_EPS	18	2	0,53148		
RELIABILITY	BBSM	18	2	0,55678	0	1
RELIABILITY	ECMWF_EPS	24	2	0,37076		
RELIABILITY	BBSM	24	2	0,43519	1	0
RELIABILITY	ECMWF_EPS	30	2	0,34056		
RELIABILITY	BBSM	30	2	0,45542	1	0
RELIABILITY	ECMWF_EPS	36	2	0,63003		
RELIABILITY	BBSM	36	2	0,67325	0	1
RELIABILITY	ECMWF_EPS	42	2	0,53291		
RELIABILITY	BBSM	42	2	0,56624	0	1
RELIABILITY	ECMWF_EPS	48	2	0,38218		
RELIABILITY	BBSM	48	2	0,45582	1	0
RELIABILITY	ECMWF_EPS	54	2	0,35279		
RELIABILITY	BBSM	54	2	0,47977	1	0
RELIABILITY	ECMWF_EPS	6	4	0,23787		
RELIABILITY	BBSM	6	4	0,40785	1	0
RELIABILITY	ECMWF_EPS	12	4	0,38750		
RELIABILITY	BBSM	12	4	0,54478	1	0
RELIABILITY	ECMWF_EPS	18	4	0,32800		
RELIABILITY	BBSM	18	4	0,41398	1	0
RELIABILITY	ECMWF_EPS	24	4	0,25272		
RELIABILITY	BBSM	24	4	0,33929	1	0
RELIABILITY	ECMWF_EPS	30	4	0,24072		

RELIABILITY	BBSM	30	4	0,40746	1	0
RELIABILITY	ECMWF_EPS	36	4	0,36339		
RELIABILITY	BBSM	36	4	0,54127	1	0
RELIABILITY	ECMWF_EPS	42	4	0,32879		
RELIABILITY	BBSM	42	4	0,43477	1	0
RELIABILITY	ECMWF_EPS	48	4	0,26629		
RELIABILITY	BBSM	48	4	0,36706	1	0
RELIABILITY	ECMWF_EPS	54	4	0,26207		
RELIABILITY	BBSM	54	4	0,43439	1	0
RELIABILITY	ECMWF_EPS	6	6	0,17759		
RELIABILITY	BBSM	6	6	0,34962	1	0
RELIABILITY	ECMWF_EPS	12	6	0,26446		
RELIABILITY	BBSM	12	6	0,45322	1	0
RELIABILITY	ECMWF_EPS	18	6	0,22218		
RELIABILITY	BBSM	18	6	0,35273	1	0
RELIABILITY	ECMWF_EPS	24	6	0,18738		
RELIABILITY	BBSM	24	6	0,30738	1	0
RELIABILITY	ECMWF_EPS	30	6	0,17904		
RELIABILITY	BBSM	30	6	0,35631	1	0
RELIABILITY	ECMWF_EPS	36	6	0,25238		
RELIABILITY	BBSM	36	6	0,46299	1	0
RELIABILITY	ECMWF_EPS	42	6	0,22994		
RELIABILITY	BBSM	42	6	0,38440	1	0
RELIABILITY	ECMWF_EPS	48	6	0,21719		
RELIABILITY	BBSM	48	6	0,34648	1	0
RELIABILITY	ECMWF_EPS	54	6	0,19972		
RELIABILITY	BBSM	54	6	0,39083	1	0
RMSE	BBSM	54	No	2,02026		
RMSE	ECMWF_EPS	54	No	2,16984	1	0
RPSEPS	BBSM	54	No	0,16281		
RPSEPS	ECMWF_EPS	54	No	0,17944	1	0
RPSSANA	ECMWF_EPS	54	No	0,58025		
RPSSANA	BBSM	54	No	0,62288	1	0
RPSSREF	ECMWF_EPS	54	No	0,15421		
RPSSREF	BBSM	54	No	0,23248	1	0
SPREAD	BBSM	54	No	0,67688		
SPREAD	ECMWF_EPS	54	No	0,76929	0	1
ROC_AREA	BBSM	6	1	0,54161		
ROC_AREA	ECMWF_EPS	6	1	0,46484	1	0
ROC_AREA	BBSM	6	2	0,74284		
ROC_AREA	ECMWF_EPS	6	2	0,67393	1	0
ROC_AREA	BBSM	6	4	0,77724		
ROC_AREA	ECMWF_EPS	6	4	0,77057	1	0
ROC_AREA	BBSM	6	6	0,73669		
ROC_AREA	ECMWF_EPS	6	6	0,75859	0	1
ROC_AREA	BBSM	12	1	0,42758		
ROC_AREA	ECMWF_EPS	12	1	0,34907	1	0
ROC_AREA	BBSM	12	2	0,65198		
ROC_AREA	ECMWF_EPS	12	2	0,59135	1	0
ROC_AREA	BBSM	12	4	0,75925		
ROC_AREA	ECMWF_EPS	12	4	0,75596	1	0
ROC_AREA	BBSM	12	6	0,74558		
ROC_AREA	ECMWF_EPS	12	6	0,79313	0	1

ROC_AREA	BBSM	18	1	0,53146		
ROC_AREA	ECMWF_EPS	18	1	0,46026	1	0
ROC_AREA	BBSM	18	2	0,68296		
ROC_AREA	ECMWF_EPS	18	2	0,67055	1	0
ROC_AREA	BBSM	18	4	0,73387		
ROC_AREA	ECMWF_EPS	18	4	0,74475	0	1
ROC_AREA	BBSM	18	6	0,72217		
ROC_AREA	ECMWF_EPS	18	6	0,74919	0	1
ROC_AREA	BBSM	24	1	0,55120		
ROC_AREA	ECMWF_EPS	24	1	0,43031	1	0
ROC_AREA	BBSM	24	2	0,70762		
ROC_AREA	ECMWF_EPS	24	2	0,66732	1	0
ROC_AREA	BBSM	24	4	0,74348		
ROC_AREA	ECMWF_EPS	24	4	0,74800	0	1
ROC_AREA	BBSM	24	6	0,70836		
ROC_AREA	ECMWF_EPS	24	6	0,72503	0	1
ROC_AREA	BBSM	30	1	0,59935		
ROC_AREA	ECMWF_EPS	30	1	0,52890	1	0
ROC_AREA	BBSM	30	2	0,76017		
ROC_AREA	ECMWF_EPS	30	2	0,70118	1	0
ROC_AREA	BBSM	30	4	0,77185		
ROC_AREA	ECMWF_EPS	30	4	0,77617	0	1
ROC_AREA	BBSM	30	6	0,72705		
ROC_AREA	ECMWF_EPS	30	6	0,76362	0	1
ROC_AREA	BBSM	36	1	0,47134		
ROC_AREA	ECMWF_EPS	36	1	0,43893	1	0
ROC_AREA	BBSM	36	2	0,66931		
ROC_AREA	ECMWF_EPS	36	2	0,62417	1	0
ROC_AREA	BBSM	36	4	0,75342		
ROC_AREA	ECMWF_EPS	36	4	0,75618	0	1
ROC_AREA	BBSM	36	6	0,73602		
ROC_AREA	ECMWF_EPS	36	6	0,79828	0	1
ROC_AREA	BBSM	42	1	0,57453		
ROC_AREA	ECMWF_EPS	42	1	0,51819	1	0
ROC_AREA	BBSM	42	2	0,69485		
ROC_AREA	ECMWF_EPS	42	2	0,68416	1	0
ROC_AREA	BBSM	42	4	0,72755		
ROC_AREA	ECMWF_EPS	42	4	0,74327	0	1
ROC_AREA	BBSM	42	6	0,71475		
ROC_AREA	ECMWF_EPS	42	6	0,74871	0	1
ROC_AREA	BBSM	48	1	0,59427		
ROC_AREA	ECMWF_EPS	48	1	0,51087	1	0
ROC_AREA	BBSM	48	2	0,71486		
ROC_AREA	ECMWF_EPS	48	2	0,69175	1	0
ROC_AREA	BBSM	48	4	0,73556		
ROC_AREA	ECMWF_EPS	48	4	0,74362	1	0
ROC_AREA	BBSM	48	6	0,70139		
ROC_AREA	ECMWF_EPS	48	6	0,72528	1	0
ROC_AREA	BBSM	54	1	0,63106		
ROC_AREA	ECMWF_EPS	54	1	0,60034	1	0
ROC_AREA	BBSM	54	2	0,75918		
ROC_AREA	ECMWF_EPS	54	2	0,73468	1	0
ROC_AREA	BBSM	54	4	0,76308		

ROC_AREA	ECMWF_EPS	54	4	0,77411	0	1
ROC_AREA	BBSM	54	6	0,72046		
ROC_AREA	ECMWF_EPS	54	6	0,75556	1	0
TALAGRAND	BBSM	6	No	0,36416		
TALAGRAND	ECMWF_EPS	6	No	0,49439	1	0
TALAGRAND	BBSM	12	No	0,19971		
TALAGRAND	ECMWF_EPS	12	No	0,31786	1	0
TALAGRAND	BBSM	18	No	0,24102		
TALAGRAND	ECMWF_EPS	18	No	0,29707	1	0
TALAGRAND	BBSM	24	No	0,32098		
TALAGRAND	ECMWF_EPS	24	No	0,40796	1	0
TALAGRAND	BBSM	30	No	0,28490		
TALAGRAND	ECMWF_EPS	30	No	0,40502	1	0
TALAGRAND	BBSM	36	No	0,15161		
TALAGRAND	ECMWF_EPS	36	No	0,24049	1	0
TALAGRAND	BBSM	42	No	0,18309		
TALAGRAND	ECMWF_EPS	42	No	0,22281	1	0
TALAGRAND	BBSM	48	No	0,25319		
TALAGRAND	ECMWF_EPS	48	No	0,31906	1	0
TALAGRAND	BBSM	54	No	0,23145		
TALAGRAND	ECMWF_EPS	54	No	0,30144	1	0
				106	71	35

Parameter: Total Precipitation

Score	Experiment	fc.time	thr	Average (thr, time)	BBSM	ECMWF-EPS
BIAS	ECMWF_EPS	54	No	0,16210		
BIAS	BBSM	54	No	0,39149	0	1
BRIERSCOREEPS	BBSM	54	25	0,00915		
BRIERSCOREEPS	ECMWF_EPS	54	25	0,00942	1	0
BRIERSCOREEPS	BBSM	54	10	0,04488		
BRIERSCOREEPS	ECMWF_EPS	54	10	0,04705	1	0
BRIERSCOREEPS	BBSM	54	5	0,08200		
BRIERSCOREEPS	ECMWF_EPS	54	5	0,09119	1	0
BRIERSCOREEPS	BBSM	54	2	0,13390		
BRIERSCOREEPS	ECMWF_EPS	54	2	0,14121	1	0
BRIERSCOREEPS	BBSM	54	1	0,16886		
BRIERSCOREEPS	ECMWF_EPS	54	1	0,17030	1	0
BRIERSSANA	ECMWF_EPS	54	25	-0,06093		
BRIERSSANA	BBSM	54	25	-0,02974	1	0
BRIERSSANA	ECMWF_EPS	54	10	0,09782		
BRIERSSANA	BBSM	54	10	0,13967	1	0
BRIERSSANA	ECMWF_EPS	54	1	0,23236		
BRIERSSANA	BBSM	54	1	0,23792	0	0
BRIERSSANA	ECMWF_EPS	54	2	0,20877		
BRIERSSANA	BBSM	54	2	0,24953	1	0
BRIERSSANA	ECMWF_EPS	54	5	0,14472		
BRIERSSANA	BBSM	54	5	0,23386	1	0
BRIERSSREF	ECMWF_EPS	54	1	0,26106		
BRIERSSREF	BBSM	54	1	0,26800	1	0
BRIERSSREF	ECMWF_EPS	54	2	0,29030		
BRIERSSREF	BBSM	54	2	0,32621	1	0
BRIERSSREF	ECMWF_EPS	54	5	0,32823		
BRIERSSREF	BBSM	54	5	0,38969	1	0

BRIERSSREF	ECMWF_EPS	54	10	0,37149		
BRIERSSREF	BBSM	54	10	0,40020	1	0
BRIERSSREF	ECMWF_EPS	54	25	0,42558		
BRIERSSREF	BBSM	54	25	0,44237	1	0
CRPSEPS	BBSM	54	No	1,43122		
CRPSEPS	ECMWF_EPS	54	No	1,50026	1	0
CRPSSANA	ECMWF_EPS	54	No	0,19848		
CRPSSANA	BBSM	54	No	0,23633	1	0
CRPSSREF	ECMWF_EPS	54	No	0,33600		
CRPSSREF	BBSM	54	No	0,36473	1	0
OUTLIERS	BBSM	54	No	0,26375		
OUTLIERS	ECMWF_EPS	54	No	0,38241	1	0
RELIABILITY	ECMWF_EPS	18	1	0,14663		
RELIABILITY	BBSM	18	1	0,16263	1	0
RELIABILITY	ECMWF_EPS	30	1	0,35997		
RELIABILITY	BBSM	30	1	0,37686	1	0
RELIABILITY	BBSM	42	1	0,22411		
RELIABILITY	ECMWF_EPS	42	1	0,23887	0	1
RELIABILITY	ECMWF_EPS	54	1	0,38729		
RELIABILITY	BBSM	54	1	0,41588	1	0
RELIABILITY	ECMWF_EPS	18	2	0,17806		
RELIABILITY	BBSM	18	2	0,19011	1	0
RELIABILITY	ECMWF_EPS	30	2	0,38261		
RELIABILITY	BBSM	30	2	0,42731	1	0
RELIABILITY	BBSM	42	2	0,26337		
RELIABILITY	ECMWF_EPS	42	2	0,27020	0	1
RELIABILITY	ECMWF_EPS	54	2	0,40602		
RELIABILITY	BBSM	54	2	0,44338	1	0
RELIABILITY	ECMWF_EPS	18	5	0,20542		
RELIABILITY	BBSM	18	5	0,22825	1	0
RELIABILITY	ECMWF_EPS	30	5	0,39776		
RELIABILITY	BBSM	30	5	0,42758	1	0
RELIABILITY	ECMWF_EPS	42	5	0,29923		
RELIABILITY	BBSM	42	5	0,32135	1	0
RELIABILITY	ECMWF_EPS	54	5	0,42016		
RELIABILITY	BBSM	54	5	0,46659	1	0
RELIABILITY	ECMWF_EPS	18	10	0,20775		
RELIABILITY	BBSM	18	10	0,22105	1	0
RELIABILITY	ECMWF_EPS	30	10	0,37299		
RELIABILITY	BBSM	30	10	0,41886	1	0
RELIABILITY	ECMWF_EPS	42	10	0,32431		
RELIABILITY	BBSM	42	10	0,33481	1	0
RELIABILITY	ECMWF_EPS	54	10	0,41071		
RELIABILITY	BBSM	54	10	0,46541	1	0
RELIABILITY	ECMWF_EPS	18	25	0,16301		
RELIABILITY	BBSM	18	25	0,19903	1	0
RELIABILITY	ECMWF_EPS	30	25	0,13181		
RELIABILITY	BBSM	30	25	0,22657	1	0
RELIABILITY	ECMWF_EPS	42	25	0,12745		
RELIABILITY	BBSM	42	25	0,22350	1	0
RELIABILITY	ECMWF_EPS	54	25	0,15506		
RELIABILITY	BBSM	54	25	0,15644	1	0
RMSE	ECMWF_EPS	54	No	4,05285		

RMSE	BBSM	54	No	4,08245	0	1
ROC_AREA	BBSM	18	1	0,83868		
ROC_AREA	ECMWF_EPS	18	1	0,78582	1	0
ROC_AREA	BBSM	18	2	0,86741		
ROC_AREA	ECMWF_EPS	18	2	0,82644	1	0
ROC_AREA	BBSM	18	5	0,88309		
ROC_AREA	ECMWF_EPS	18	5	0,85970	1	0
ROC_AREA	BBSM	18	10	0,86910		
ROC_AREA	ECMWF_EPS	18	10	0,84039	1	0
ROC_AREA	BBSM	18	25	0,77645		
ROC_AREA	ECMWF_EPS	18	25	0,73671	1	0
ROC_AREA	BBSM	30	1	0,87738		
ROC_AREA	ECMWF_EPS	30	1	0,87808	0	1
ROC_AREA	BBSM	30	2	0,87982		
ROC_AREA	ECMWF_EPS	30	2	0,87758	1	0
ROC_AREA	BBSM	30	5	0,87188		
ROC_AREA	ECMWF_EPS	30	5	0,86041	1	0
ROC_AREA	BBSM	30	10	0,85245		
ROC_AREA	ECMWF_EPS	30	10	0,83187	1	0
ROC_AREA	BBSM	30	25	0,75495		
ROC_AREA	ECMWF_EPS	30	25	0,72492	1	0
ROC_AREA	BBSM	42	1	0,85983		
ROC_AREA	ECMWF_EPS	42	1	0,87051	0	1
ROC_AREA	BBSM	42	2	0,86946		
ROC_AREA	ECMWF_EPS	42	2	0,87635	0	1
ROC_AREA	BBSM	42	5	0,87031		
ROC_AREA	ECMWF_EPS	42	5	0,87209	0	1
ROC_AREA	BBSM	42	10	0,85191		
ROC_AREA	ECMWF_EPS	42	10	0,83450	1	0
ROC_AREA	BBSM	42	25	0,76137		
ROC_AREA	ECMWF_EPS	42	25	0,73128	1	0
ROC_AREA	BBSM	54	1	0,86028		
ROC_AREA	ECMWF_EPS	54	1	0,86886	0	1
ROC_AREA	BBSM	54	2	0,86221		
ROC_AREA	ECMWF_EPS	54	2	0,86515	0	1
ROC_AREA	BBSM	54	5	0,85728		
ROC_AREA	ECMWF_EPS	54	5	0,85402	1	0
ROC_AREA	BBSM	54	10	0,83332		
ROC_AREA	ECMWF_EPS	54	10	0,83968	0	1
ROC_AREA	BBSM	54	25	0,70071		
ROC_AREA	ECMWF_EPS	54	25	0,73535	0	1
RPSEPS	BBSM	54	No	0,08776		
RPSEPS	ECMWF_EPS	54	No	0,09184	0	1
RPSSANA	ECMWF_EPS	54	No	0,19067		
RPSSANA	BBSM	54	No	0,22651	1	0
RPSSREF	ECMWF_EPS	54	No	0,30071		
RPSSREF	BBSM	54	No	0,33042	1	0
SPREAD	ECMWF_EPS	54	No	1,21742		
SPREAD	BBSM	54	No	1,65820	1	0
TALAGRAN	BBSM	18	No	0,40000		
TALAGRAN	ECMWF_EPS	18	No	0,42252	1	0
TALAGRAN	BBSM	30	No	0,12755		
TALAGRAN	ECMWF_EPS	30	No	0,16512	1	0

TALAGRAND	BBSM	42	No	0,21757		
TALAGRAND	ECMWF_EPS	42	No	0,27239	0	1
TALAGRAND	BBSM	54	No	0,09343		
TALAGRAND	ECMWF_EPS	54	No	0,09446	1	0

68 54 14

Parameter: Temperature Anomaly in 500hPa level						
Score	Experiment	fc.time	thr	Average (thr, time)	BBSM	ECMWF-EPS
BIAS	BBSM	54	No	0,14022		
BIAS	ECMWF_EPS	54	No	0,22305	1	0
BRIERSCOREEPS	BBSM	54	-2	0,02433		
BRIERSCOREEPS	ECMWF_EPS	54	-2	0,02508	1	0
BRIERSCOREEPS	BBSM	54	0	0,03975		
BRIERSCOREEPS	ECMWF_EPS	54	0	0,04007	1	0
BRIERSCOREEPS	BBSM	54	2	0,05230		
BRIERSCOREEPS	ECMWF_EPS	54	2	0,05341	1	0
BRIERSSANA	ECMWF_EPS	54	-2	0,82384		
BRIERSSANA	BBSM	54	-2	0,82903	1	0
BRIERSSANA	ECMWF_EPS	54	0	0,86101		
BRIERSSANA	BBSM	54	0	0,86211	1	0
BRIERSSANA	ECMWF_EPS	54	2	0,90301		
BRIERSSANA	BBSM	54	2	0,90499	1	0
BRIERSSREF	ECMWF_EPS	54	-2	0,36873		
BRIERSSREF	ECMWF_EPS	54	0	0,38017	0	1
BRIERSSREF	BBSM	54	0	0,38567		
BRIERSSREF	BBSM	54	-2	0,38605	1	0
BRIERSSREF	ECMWF_EPS	54	2	0,39619		
BRIERSSREF	BBSM	54	2	0,40861	1	0
CRPSEPS	BBSM	54	No	0,36441		
CRPSEPS	ECMWF_EPS	54	No	0,36956	1	0
CRPSSANA	ECMWF_EPS	54	No	0,96963		
CRPSSANA	BBSM	54	No	0,97006	1	0
CRPSSREF	ECMWF_EPS	54	No	0,39319		
CRPSSREF	BBSM	54	No	0,40147	1	0
OUTLIERS	ECMWF_EPS	54	No	0,07934		
OUTLIERS	BBSM	54	No	0,08032	0	1
RELIABILITY	BBSM	6	-2	0,47277		
RELIABILITY	ECMWF_EPS	6	-2	0,54103	1	0
RELIABILITY	BBSM	12	-2	0,53960		
RELIABILITY	ECMWF_EPS	12	-2	0,59608	1	0
RELIABILITY	BBSM	18	-2	0,54096		
RELIABILITY	ECMWF_EPS	18	-2	0,62147	1	0
RELIABILITY	BBSM	24	-2	0,53806		
RELIABILITY	ECMWF_EPS	24	-2	0,60125	1	0
RELIABILITY	BBSM	30	-2	0,56599		
RELIABILITY	ECMWF_EPS	30	-2	0,61846	1	0
RELIABILITY	BBSM	36	-2	0,57286		
RELIABILITY	ECMWF_EPS	36	-2	0,61258	1	0
RELIABILITY	BBSM	42	-2	0,58192		
RELIABILITY	ECMWF_EPS	42	-2	0,62076	1	0
RELIABILITY	BBSM	48	-2	0,58222		
RELIABILITY	ECMWF_EPS	48	-2	0,60623	1	0
RELIABILITY	BBSM	54	-2	0,59439		
RELIABILITY	ECMWF_EPS	54	-2	0,61652	1	0

RELIABILITY	ECMWF_EPS	6	0	0,50924		
RELIABILITY	BBSM	6	0	0,54611	0	1
RELIABILITY	ECMWF_EPS	12	0	0,42968		
RELIABILITY	BBSM	12	0	0,47667	1	0
RELIABILITY	ECMWF_EPS	18	0	0,41503		
RELIABILITY	BBSM	18	0	0,48164	1	0
RELIABILITY	ECMWF_EPS	24	0	0,40337		
RELIABILITY	BBSM	24	0	0,45304	1	0
RELIABILITY	ECMWF_EPS	30	0	0,37949		
RELIABILITY	BBSM	30	0	0,41173	1	0
RELIABILITY	ECMWF_EPS	36	0	0,38085		
RELIABILITY	BBSM	36	0	0,41524	1	0
RELIABILITY	ECMWF_EPS	42	0	0,38639		
RELIABILITY	BBSM	42	0	0,43057	1	0
RELIABILITY	ECMWF_EPS	48	0	0,39856		
RELIABILITY	BBSM	48	0	0,42595	1	0
RELIABILITY	ECMWF_EPS	54	0	0,38987		
RELIABILITY	BBSM	54	0	0,40538	1	0
RELIABILITY	ECMWF_EPS	6	2	0,52968		
RELIABILITY	BBSM	6	2	0,53220	0	1
RELIABILITY	ECMWF_EPS	12	2	0,50365		
RELIABILITY	BBSM	12	2	0,51299	0	1
RELIABILITY	ECMWF_EPS	18	2	0,49449		
RELIABILITY	BBSM	18	2	0,50348	0	1
RELIABILITY	ECMWF_EPS	24	2	0,46000		
RELIABILITY	BBSM	24	2	0,46723	0	1
RELIABILITY	BBSM	30	2	0,46440		
RELIABILITY	ECMWF_EPS	30	2	0,46555	1	0
RELIABILITY	ECMWF_EPS	36	2	0,45998		
RELIABILITY	BBSM	36	2	0,46735	0	1
RELIABILITY	ECMWF_EPS	42	2	0,46626		
RELIABILITY	BBSM	42	2	0,47619	0	1
RELIABILITY	ECMWF_EPS	48	2	0,45287		
RELIABILITY	BBSM	48	2	0,45997	0	1
RELIABILITY	BBSM	54	2	0,46520		
RELIABILITY	ECMWF_EPS	54	2	0,46662	1	0
RMSE	BBSM	54	No	0,62535		
RMSE	ECMWF_EPS	54	No	0,65321	1	0
RPSEPS	BBSM	54	No	0,03879		
RPSEPS	ECMWF_EPS	54	No	0,03952	1	0
RPSSANA	ECMWF_EPS	54	No	0,87917		
RPSSANA	BBSM	54	No	0,88137	1	0
RPSSREF	ECMWF_EPS	54	No	0,38597		
RPSSREF	BBSM	54	No	0,39718	1	0
SPREAD	ECMWF_EPS	54	No	0,66011		
SPREAD	BBSM	54	No	0,72311	1	0
ROC_AREA	BBSM	6	-2	0,99792		
ROC_AREA	ECMWF_EPS	6	-2	0,99811	0	1
ROC_AREA	BBSM	6	0	0,99500		
ROC_AREA	ECMWF_EPS	6	0	0,99597	0	1
ROC_AREA	BBSM	6	2	0,99599		
ROC_AREA	ECMWF_EPS	6	2	0,99685	0	1
ROC_AREA	BBSM	12	-2	0,99309		

ROC_AREA	ECMWF_EPS	12	-2	0,99323	0	1
ROC_AREA	BBSM	12	0	0,98916		
ROC_AREA	ECMWF_EPS	12	0	0,98973	0	1
ROC_AREA	BBSM	12	2	0,99275		
ROC_AREA	ECMWF_EPS	12	2	0,99234	1	0
ROC_AREA	BBSM	18	-2	0,99437		
ROC_AREA	ECMWF_EPS	18	-2	0,99225	1	0
ROC_AREA	BBSM	18	0	0,99119		
ROC_AREA	ECMWF_EPS	18	0	0,98676	1	0
ROC_AREA	BBSM	18	2	0,99270		
ROC_AREA	ECMWF_EPS	18	2	0,99106	1	0
ROC_AREA	BBSM	24	-2	0,99103		
ROC_AREA	ECMWF_EPS	24	-2	0,99134	0	1
ROC_AREA	BBSM	24	0	0,98105		
ROC_AREA	ECMWF_EPS	24	0	0,97847	1	0
ROC_AREA	BBSM	24	2	0,98772		
ROC_AREA	ECMWF_EPS	24	2	0,98673	1	0
ROC_AREA	BBSM	30	-2	0,99137		
ROC_AREA	ECMWF_EPS	30	-2	0,98864	1	0
ROC_AREA	BBSM	30	0	0,98129		
ROC_AREA	ECMWF_EPS	30	0	0,97715	1	0
ROC_AREA	BBSM	30	2	0,98664		
ROC_AREA	ECMWF_EPS	30	2	0,98596	1	0
ROC_AREA	BBSM	36	-2	0,98617		
ROC_AREA	ECMWF_EPS	36	-2	0,98163	1	0
ROC_AREA	BBSM	36	0	0,97753		
ROC_AREA	ECMWF_EPS	36	0	0,97201	1	0
ROC_AREA	BBSM	36	2	0,98342		
ROC_AREA	ECMWF_EPS	36	2	0,97965	1	0
ROC_AREA	BBSM	42	-2	0,98329		
ROC_AREA	ECMWF_EPS	42	-2	0,98205	1	0
ROC_AREA	BBSM	42	0	0,98389		
ROC_AREA	ECMWF_EPS	42	0	0,97660	1	0
ROC_AREA	BBSM	42	2	0,98276		
ROC_AREA	ECMWF_EPS	42	2	0,98012	1	0
ROC_AREA	BBSM	48	-2	0,98041		
ROC_AREA	ECMWF_EPS	48	-2	0,97940	1	0
ROC_AREA	BBSM	48	0	0,96838		
ROC_AREA	ECMWF_EPS	48	0	0,96670	1	0
ROC_AREA	BBSM	48	2	0,97627		
ROC_AREA	ECMWF_EPS	48	2	0,97437	1	0
ROC_AREA	BBSM	54	-2	0,97723		
ROC_AREA	ECMWF_EPS	54	-2	0,98158	0	1
ROC_AREA	BBSM	54	0	0,96526		
ROC_AREA	ECMWF_EPS	54	0	0,96828	0	1
ROC_AREA	BBSM	54	2	0,97373		
ROC_AREA	ECMWF_EPS	54	2	0,97545	0	1
TALAGRAND	BBSM	6	No	0,02012		
TALAGRAND	ECMWF_EPS	6	No	0,01858	0	1
TALAGRAND	BBSM	12	No	0,04970		
TALAGRAND	ECMWF_EPS	12	No	0,06443	1	0
TALAGRAND	BBSM	18	No	0,03385		
TALAGRAND	ECMWF_EPS	18	No	0,06784	1	0

TALAGRAND	BBSM	24	No	0,06237		
TALAGRAND	ECMWF_EPS	24	No	0,08234	1	0
TALAGRAND	BBSM	30	No	0,06424		
TALAGRAND	ECMWF_EPS	30	No	0,07522	1	0
TALAGRAND	BBSM	36	No	0,06265		
TALAGRAND	ECMWF_EPS	36	No	0,09207	1	0
TALAGRAND	BBSM	42	No	0,04524		
TALAGRAND	ECMWF_EPS	42	No	0,07354	1	0
TALAGRAND	BBSM	48	No	0,07008		
TALAGRAND	ECMWF_EPS	48	No	0,08106	1	0
TALAGRAND	BBSM	54	No	0,06726		
TALAGRAND	ECMWF_EPS	54	No	0,06305	0	1

61

21

Parameter: Temperature Anomaly in 850hPa level						
Score	Experiment	fc.time	thr	Average (thr, time)	BBSM	ECMWF-EPS
BIAS	BBSM	54	No	0,12784		
BIAS	ECMWF_EPS	54	No	0,13973	1	0
BRIERSCOREEPS	BBSM	54	-2	0,03416		
BRIERSCOREEPS	ECMWF_EPS	54	-2	0,03489	1	0
BRIERSCOREEPS	BBSM	54	0	0,04207		
BRIERSCOREEPS	ECMWF_EPS	54	0	0,04301	1	0
BRIERSCOREEPS	ECMWF_EPS	54	2	0,04142		
BRIERSCOREEPS	BBSM	54	2	0,04295	1	0
BRIERSSANA	BBSM	54	2	0,89712		
BRIERSSANA	ECMWF_EPS	54	2	0,90078	0	1
BRIERSSANA	ECMWF_EPS	54	0	0,92872		
BRIERSSANA	BBSM	54	0	0,93030	1	0
BRIERSSANA	ECMWF_EPS	54	-2	0,95455		
BRIERSSANA	BBSM	54	-2	0,95548	1	0
BRIERSSREF	ECMWF_EPS	54	-2	0,39094		
BRIERSSREF	BBSM	54	-2	0,40373	1	0
BRIERSSREF	ECMWF_EPS	54	0	0,38392		
BRIERSSREF	BBSM	54	0	0,39561	1	0
BRIERSSREF	BBSM	54	2	0,40093		
BRIERSSREF	ECMWF_EPS	54	2	0,42560	0	1
CRPSEPS	ECMWF_EPS	54	No	0,47458		
CRPSEPS	BBSM	54	No	0,49379	0	1
CRPSSANA	BBSM	54	No	0,96319		
CRPSSANA	ECMWF_EPS	54	No	0,96462	0	1
CRPSSREF	BBSM	54	No	0,38730		
CRPSSREF	ECMWF_EPS	54	No	0,41275	0	1
OUTLIERS	ECMWF_EPS	54	No	0,12407		
OUTLIERS	BBSM	54	No	0,16424	0	1
RELIABILITY	ECMWF_EPS	6	-2	0,49621		
RELIABILITY	BBSM	6	-2	0,49789	1	0
RELIABILITY	BBSM	12	-2	0,56378		
RELIABILITY	ECMWF_EPS	12	-2	0,59037	1	0
RELIABILITY	BBSM	18	-2	0,56798		
RELIABILITY	ECMWF_EPS	18	-2	0,58592	1	0
RELIABILITY	ECMWF_EPS	24	-2	0,53457		
RELIABILITY	BBSM	24	-2	0,53476	0	1
RELIABILITY	ECMWF_EPS	30	-2	0,52441		
RELIABILITY	BBSM	30	-2	0,55335	0	1

RELIABILITY	ECMWF_EPS	36	-2	0,59179		
RELIABILITY	BBSM	36	-2	0,59524	0	1
RELIABILITY	ECMWF_EPS	42	-2	0,60009		
RELIABILITY	BBSM	42	-2	0,60100	0	1
RELIABILITY	ECMWF_EPS	48	-2	0,55980		
RELIABILITY	BBSM	48	-2	0,57056	0	1
RELIABILITY	ECMWF_EPS	54	-2	0,55580		
RELIABILITY	BBSM	54	-2	0,58583	0	1
RELIABILITY	ECMWF_EPS	6	0	0,50825		
RELIABILITY	BBSM	6	0	0,55248	0	1
RELIABILITY	ECMWF_EPS	12	0	0,41103		
RELIABILITY	BBSM	12	0	0,44578	1	0
RELIABILITY	ECMWF_EPS	18	0	0,40929		
RELIABILITY	BBSM	18	0	0,44640	1	0
RELIABILITY	ECMWF_EPS	24	0	0,48990		
RELIABILITY	BBSM	24	0	0,51394	1	0
RELIABILITY	ECMWF_EPS	30	0	0,48602		
RELIABILITY	BBSM	30	0	0,48655	1	0
RELIABILITY	BBSM	36	0	0,42467		
RELIABILITY	ECMWF_EPS	36	0	0,42989	0	1
RELIABILITY	ECMWF_EPS	42	0	0,42400		
RELIABILITY	BBSM	42	0	0,42759	1	0
RELIABILITY	BBSM	48	0	0,47366		
RELIABILITY	ECMWF_EPS	48	0	0,47752	0	1
RELIABILITY	BBSM	54	0	0,46437		
RELIABILITY	ECMWF_EPS	54	0	0,47713	0	1
RELIABILITY	ECMWF_EPS	6	2	0,49569		
RELIABILITY	BBSM	6	2	0,52222	0	1
RELIABILITY	ECMWF_EPS	12	2	0,50779		
RELIABILITY	BBSM	12	2	0,51134	0	1
RELIABILITY	ECMWF_EPS	18	2	0,50185		
RELIABILITY	BBSM	18	2	0,50490	0	1
RELIABILITY	ECMWF_EPS	24	2	0,50508		
RELIABILITY	BBSM	24	2	0,52421	0	1
RELIABILITY	ECMWF_EPS	30	2	0,48722		
RELIABILITY	BBSM	30	2	0,52417	0	1
RELIABILITY	ECMWF_EPS	36	2	0,50063		
RELIABILITY	BBSM	36	2	0,50504	0	1
RELIABILITY	ECMWF_EPS	42	2	0,49954		
RELIABILITY	BBSM	42	2	0,50510	0	1
RELIABILITY	ECMWF_EPS	48	2	0,51053		
RELIABILITY	BBSM	48	2	0,52272	0	1
RELIABILITY	ECMWF_EPS	54	2	0,50442		
RELIABILITY	BBSM	54	2	0,52801	0	1
RMSE	ECMWF_EPS	54	No	0,84380		
RMSE	BBSM	54	No	0,86145	0	1
RPSEPS	BBSM	54	No	0,03973		
RPSEPS	ECMWF_EPS	54	No	0,03977	1	0
RPSSANA	ECMWF_EPS	54	No	0,93328		
RPSSANA	BBSM	54	No	0,93336	1	0
RPSSREF	BBSM	54	No	0,39991		
RPSSREF	ECMWF_EPS	54	No	0,40121	0	1
SPREAD	ECMWF_EPS	54	No	0,71389		

SPREAD	BBSM	54	No	0,79678	1	0
ROC_AREA	BBSM	6	-2	0,99403		
ROC_AREA	ECMWF_EPS	6	-2	0,99347	1	0
ROC_AREA	BBSM	12	-2	0,97873		
ROC_AREA	ECMWF_EPS	12	-2	0,97866	1	0
ROC_AREA	BBSM	18	-2	0,98292		
ROC_AREA	ECMWF_EPS	18	-2	0,98283	1	0
ROC_AREA	BBSM	24	-2	0,98420		
ROC_AREA	ECMWF_EPS	24	-2	0,98553	0	1
ROC_AREA	BBSM	30	-2	0,98691		
ROC_AREA	ECMWF_EPS	30	-2	0,98649	1	0
ROC_AREA	BBSM	36	-2	0,97951		
ROC_AREA	ECMWF_EPS	36	-2	0,97778	1	0
ROC_AREA	BBSM	42	-2	0,98118		
ROC_AREA	ECMWF_EPS	42	-2	0,98134	0	1
ROC_AREA	BBSM	48	-2	0,98182		
ROC_AREA	ECMWF_EPS	48	-2	0,98215	0	1
ROC_AREA	BBSM	54	-2	0,98210		
ROC_AREA	ECMWF_EPS	54	-2	0,98258	0	1
ROC_AREA	BBSM	6	0	0,99165		
ROC_AREA	ECMWF_EPS	6	0	0,99152	1	0
ROC_AREA	BBSM	12	0	0,97274		
ROC_AREA	ECMWF_EPS	12	0	0,97844	0	1
ROC_AREA	BBSM	18	0	0,97957		
ROC_AREA	ECMWF_EPS	18	0	0,98288	0	1
ROC_AREA	BBSM	24	0	0,98101		
ROC_AREA	ECMWF_EPS	24	0	0,98076	1	0
ROC_AREA	BBSM	30	0	0,98568		
ROC_AREA	ECMWF_EPS	30	0	0,98305	1	0
ROC_AREA	BBSM	36	0	0,97324		
ROC_AREA	ECMWF_EPS	36	0	0,97439	0	1
ROC_AREA	BBSM	42	0	0,97787		
ROC_AREA	ECMWF_EPS	42	0	0,97824	0	1
ROC_AREA	BBSM	48	0	0,97882		
ROC_AREA	ECMWF_EPS	48	0	0,97574	1	0
ROC_AREA	BBSM	54	0	0,97962		
ROC_AREA	ECMWF_EPS	54	0	0,97841	0	1
ROC_AREA	BBSM	6	2	0,99186		
ROC_AREA	ECMWF_EPS	6	2	0,99319	0	1
ROC_AREA	BBSM	12	2	0,97872		
ROC_AREA	ECMWF_EPS	12	2	0,98309	0	1
ROC_AREA	BBSM	18	2	0,98343		
ROC_AREA	ECMWF_EPS	18	2	0,98606	0	1
ROC_AREA	BBSM	24	2	0,98334		
ROC_AREA	ECMWF_EPS	24	2	0,98635	0	1
ROC_AREA	BBSM	30	2	0,98539		
ROC_AREA	ECMWF_EPS	30	2	0,98686	0	1
ROC_AREA	BBSM	36	2	0,97889		
ROC_AREA	ECMWF_EPS	36	2	0,97989	0	1
ROC_AREA	BBSM	42	2	0,98137		
ROC_AREA	ECMWF_EPS	42	2	0,98257	0	1
ROC_AREA	BBSM	48	2	0,98172		
ROC_AREA	ECMWF_EPS	48	2	0,98262	0	1

ROC_AREA	BBSM	54	2	0,98244		
ROC_AREA	ECMWF_EPS	54	2	0,98247	0	1
TALAGRAND	BBSM	6	No	0,06886		
TALAGRAND	ECMWF_EPS	6	No	0,08082	1	0
TALAGRAND	BBSM	12	No	0,17011		
TALAGRAND	ECMWF_EPS	12	No	0,13470	0	1
TALAGRAND	BBSM	18	No	0,12024		
TALAGRAND	ECMWF_EPS	18	No	0,09609	0	1
TALAGRAND	BBSM	24	No	0,07912		
TALAGRAND	ECMWF_EPS	24	No	0,07194	0	1
TALAGRAND	BBSM	30	No	0,06927		
TALAGRAND	ECMWF_EPS	30	No	0,06499	0	1
TALAGRAND	BBSM	36	No	0,11215		
TALAGRAND	ECMWF_EPS	36	No	0,09236	0	1
TALAGRAND	BBSM	42	No	0,08577		
TALAGRAND	ECMWF_EPS	42	No	0,07564	0	1
TALAGRAND	BBSM	48	No	0,06177		
TALAGRAND	ECMWF_EPS	48	No	0,05514	0	1
TALAGRAND	BBSM	54	No	0,05572		
TALAGRAND	ECMWF_EPS	54	No	0,04457	0	1

29

53

Parameter: Geopotential in 500hPa level						
Score	Experiment	fc.time	thr	Average (thr, time)	BBSM	ECMWF-EPS
BIAS	BBSM	54	No	29,41804		
BIAS	ECMWF_EPS	54	No	15,01890	0	1
CRPSEPS	ECMWF_EPS	54	No	44,24333		
CRPSEPS	BBSM	54	No	46,19936	0	1
CRPSSANA	BBSM	54	No	0,99918		
CRPSSANA	ECMWF_EPS	54	No	0,99922	0	1
CRPSSREF	BBSM	54	No	0,29671		
CRPSSREF	ECMWF_EPS	54	No	0,31916	0	1
OUTLIERS	ECMWF_EPS	54	No	0,02182		
OUTLIERS	BBSM	54	No	0,04827	0	1
RMSE	ECMWF_EPS	54	No	73,45551		
RMSE	BBSM	54	No	75,29250	0	1
SPREAD	ECMWF_EPS	54	No	102,25256		
SPREAD	BBSM	54	No	102,30833	1	0
TALAGRAND	ECMWF_EPS	all	No	0,01826		
TALAGRAND	BBSM	all	No	0,03949	1	0

2

6

Parameter: Geopotential in 850hPa level						
Score	Experiment	fc.time	thr	Average (thr, time)	BBSM	ECMWF-EPS
BIAS	ECMWF_EPS	54	No	-28,26636		
BIAS	BBSM	54	No	-2,28350	1	0
CRPSEPS	BBSM	54	No	34,57075		
CRPSEPS	ECMWF_EPS	54	No	40,80263	1	0
CRPSSANA	ECMWF_EPS	54	No	0,99721		
CRPSSANA	BBSM	54	No	0,99764	1	0
CRPSSREF	ECMWF_EPS	54	No	0,26221		
CRPSSREF	BBSM	54	No	0,36633	1	0
OUTLIERS	ECMWF_EPS	54	No	0,07819		
OUTLIERS	BBSM	54	No	0,09341	0	1
RMSE	BBSM	54	No	58,45002		

RMSE	ECMWF_EPS	54	No	70,77056	1	0
SPREAD	ECMWF_EPS	54	No	66,76489		
SPREAD	BBSM	54	No	66,91133	1	0
TALAGRAND	ECMWF_EPS	54	No	0,01984		
TALAGRAND	BBSM	54	No	0,04465	1	0

7 1

Parameter: Wind Speed in 500hPa level						
Score	Experiment	fc.time	thr	Average (thr, time)	BBSM	ECMWF-EPS
BIAS	BBSM	54	No	-0,18058		
BIAS	ECMWF_EPS	54	No	0,01738	0	1
BRIERSCOREEPS	ECMWF_EPS	54	2	0,00770		
BRIERSCOREEPS	BBSM	54	2	0,00858	0	1
BRIERSCOREEPS	ECMWF_EPS	54	4	0,02043		
BRIERSCOREEPS	BBSM	54	4	0,02179	0	1
BRIERSCOREEPS	ECMWF_EPS	54	7	0,04214		
BRIERSCOREEPS	BBSM	54	7	0,04459	0	1
BRIERSCOREEPS	ECMWF_EPS	54	10	0,05763		
BRIERSCOREEPS	BBSM	54	10	0,06068	0	1
BRIERSCOREREF	BBSM	54	2	0,01502		
BRIERSCOREREF	ECMWF_EPS	54	2	0,01514		
BRIERSSANA	BBSM	54	10	0,91199		
BRIERSSANA	ECMWF_EPS	54	10	0,91641	0	1
BRIERSSANA	BBSM	54	7	0,94722		
BRIERSSANA	ECMWF_EPS	54	7	0,95012	0	1
BRIERSSANA	BBSM	54	4	0,97711		
BRIERSSANA	ECMWF_EPS	54	4	0,97854	0	1
BRIERSSANA	BBSM	54	2	0,99171		
BRIERSSANA	ECMWF_EPS	54	2	0,99222	0	1
BRIERSSREF	BBSM	54	2	0,45639		
BRIERSSREF	ECMWF_EPS	54	2	0,48918	0	1
BRIERSSREF	BBSM	54	4	0,44078		
BRIERSSREF	ECMWF_EPS	54	4	0,47450	0	1
BRIERSSREF	BBSM	54	7	0,44062		
BRIERSSREF	ECMWF_EPS	54	7	0,47065	0	1
BRIERSSREF	BBSM	54	10	0,43299		
BRIERSSREF	ECMWF_EPS	54	10	0,45958	0	1
CRPSEPS	ECMWF_EPS	54	No	1,10512		
CRPSEPS	BBSM	54	No	1,16693	0	1
CRPSSANA	BBSM	54	No	0,91735		
CRPSSANA	ECMWF_EPS	54	No	0,92172	0	1
CRPSSREF	BBSM	54	No	0,42364		
CRPSSREF	ECMWF_EPS	54	No	0,45307	0	1
OUTLIERS	ECMWF_EPS	54	No	0,06205		
OUTLIERS	BBSM	54	No	0,11537	0	1
RELIABILITY	ECMWF_EPS	6	2	0,52240		
RELIABILITY	BBSM	6	2	0,52827	0	1
RELIABILITY	ECMWF_EPS	12	2	0,59458		
RELIABILITY	BBSM	12	2	0,63210	0	1
RELIABILITY	ECMWF_EPS	18	2	0,56264		
RELIABILITY	BBSM	18	2	0,56398	0	1
RELIABILITY	ECMWF_EPS	24	2	0,57945		
RELIABILITY	BBSM	24	2	0,61938	0	1
RELIABILITY	ECMWF_EPS	30	2	0,56532		

RELIABILITY	BBSM	30	2	0,63224	0	1
RELIABILITY	ECMWF_EPS	36	2	0,58019		
RELIABILITY	BBSM	36	2	0,65531	0	1
RELIABILITY	BBSM	42	2	0,52803		
RELIABILITY	ECMWF_EPS	42	2	0,59191	1	0
RELIABILITY	ECMWF_EPS	48	2	0,51749		
RELIABILITY	BBSM	48	2	0,62523	0	1
RELIABILITY	ECMWF_EPS	54	2	0,48876		
RELIABILITY	BBSM	54	2	0,62719	0	1
RELIABILITY	ECMWF_EPS	6	4	0,49091		
RELIABILITY	BBSM	6	4	0,50417	1	0
RELIABILITY	ECMWF_EPS	12	4	0,51581		
RELIABILITY	BBSM	12	4	0,54420	0	1
RELIABILITY	ECMWF_EPS	18	4	0,51488		
RELIABILITY	BBSM	18	4	0,53560	0	1
RELIABILITY	ECMWF_EPS	24	4	0,54052		
RELIABILITY	BBSM	24	4	0,56754	0	1
RELIABILITY	ECMWF_EPS	30	4	0,50956		
RELIABILITY	BBSM	30	4	0,54007	0	1
RELIABILITY	ECMWF_EPS	36	4	0,50456		
RELIABILITY	BBSM	36	4	0,53677	0	1
RELIABILITY	ECMWF_EPS	42	4	0,48260		
RELIABILITY	BBSM	42	4	0,52635	0	1
RELIABILITY	ECMWF_EPS	48	4	0,52046		
RELIABILITY	BBSM	48	4	0,56921	0	1
RELIABILITY	ECMWF_EPS	54	4	0,50195		
RELIABILITY	BBSM	54	4	0,55378	0	1
RELIABILITY	ECMWF_EPS	6	7	0,49522		
RELIABILITY	BBSM	6	7	0,51892	0	1
RELIABILITY	ECMWF_EPS	12	7	0,46561		
RELIABILITY	BBSM	12	7	0,50495	1	0
RELIABILITY	ECMWF_EPS	18	7	0,48498		
RELIABILITY	BBSM	18	7	0,51760	0	1
RELIABILITY	ECMWF_EPS	24	7	0,54046		
RELIABILITY	BBSM	24	7	0,56431	0	1
RELIABILITY	ECMWF_EPS	30	7	0,47306		
RELIABILITY	BBSM	30	7	0,51554	1	0
RELIABILITY	ECMWF_EPS	36	7	0,49860		
RELIABILITY	BBSM	36	7	0,52770	0	1
RELIABILITY	ECMWF_EPS	42	7	0,48809		
RELIABILITY	BBSM	42	7	0,52682	0	1
RELIABILITY	ECMWF_EPS	48	7	0,53095		
RELIABILITY	BBSM	48	7	0,56242	0	1
RELIABILITY	ECMWF_EPS	54	7	0,49252		
RELIABILITY	BBSM	54	7	0,54260	0	1
RELIABILITY	ECMWF_EPS	6	10	0,48396		
RELIABILITY	BBSM	6	10	0,51237	1	0
RELIABILITY	ECMWF_EPS	12	10	0,47131		
RELIABILITY	BBSM	12	10	0,50830	1	0
RELIABILITY	ECMWF_EPS	18	10	0,49196		
RELIABILITY	BBSM	18	10	0,52779	0	1
RELIABILITY	ECMWF_EPS	24	10	0,54851		
RELIABILITY	BBSM	24	10	0,55859	0	1

RELIABILITY	ECMWF_EPS	30	10	0,47326		
RELIABILITY	BBSM	30	10	0,50785	1	0
RELIABILITY	ECMWF_EPS	36	10	0,51183		
RELIABILITY	BBSM	36	10	0,53198	0	1
RELIABILITY	ECMWF_EPS	42	10	0,50325		
RELIABILITY	BBSM	42	10	0,53705	0	1
RELIABILITY	ECMWF_EPS	48	10	0,53750		
RELIABILITY	BBSM	48	10	0,55913	0	1
RELIABILITY	ECMWF_EPS	54	10	0,48616		
RELIABILITY	BBSM	54	10	0,52662	0	1
RMSE	ECMWF_EPS	54	No	1,97843		
RMSE	BBSM	54	No	2,04498	0	1
ROC_AREA	BBSM	6	2	0,96325		
ROC_AREA	ECMWF_EPS	6	2	0,97720	0	1
ROC_AREA	BBSM	6	4	0,98143		
ROC_AREA	ECMWF_EPS	6	4	0,98924	0	1
ROC_AREA	BBSM	6	7	0,98812		
ROC_AREA	ECMWF_EPS	6	7	0,99071	0	1
ROC_AREA	BBSM	6	10	0,98794		
ROC_AREA	ECMWF_EPS	6	10	0,99045	0	1
ROC_AREA	BBSM	12	2	0,90102		
ROC_AREA	ECMWF_EPS	12	2	0,93107	0	1
ROC_AREA	BBSM	12	4	0,94664		
ROC_AREA	ECMWF_EPS	12	4	0,95861	0	1
ROC_AREA	BBSM	12	7	0,97135		
ROC_AREA	ECMWF_EPS	12	7	0,97713	0	1
ROC_AREA	BBSM	12	10	0,97829		
ROC_AREA	ECMWF_EPS	12	10	0,98177	0	1
ROC_AREA	BBSM	18	2	0,93247		
ROC_AREA	ECMWF_EPS	18	2	0,95065	0	1
ROC_AREA	BBSM	18	4	0,94885		
ROC_AREA	ECMWF_EPS	18	4	0,96729	0	1
ROC_AREA	BBSM	18	7	0,97502		
ROC_AREA	ECMWF_EPS	18	7	0,97897	0	1
ROC_AREA	BBSM	18	10	0,97915		
ROC_AREA	ECMWF_EPS	18	10	0,97946	0	1
ROC_AREA	BBSM	24	2	0,88459		
ROC_AREA	ECMWF_EPS	24	2	0,91286	0	1
ROC_AREA	BBSM	24	4	0,94060		
ROC_AREA	ECMWF_EPS	24	4	0,95643	0	1
ROC_AREA	BBSM	24	7	0,96446		
ROC_AREA	ECMWF_EPS	24	7	0,97321	0	1
ROC_AREA	BBSM	24	10	0,96409		
ROC_AREA	ECMWF_EPS	24	10	0,96979	0	1
ROC_AREA	BBSM	30	2	0,89071		
ROC_AREA	ECMWF_EPS	30	2	0,90736	0	1
ROC_AREA	BBSM	30	4	0,94971		
ROC_AREA	ECMWF_EPS	30	4	0,96293	0	1
ROC_AREA	BBSM	30	7	0,95863		
ROC_AREA	ECMWF_EPS	30	7	0,96670	0	1
ROC_AREA	BBSM	30	10	0,96519		
ROC_AREA	ECMWF_EPS	30	10	0,96944	0	1
ROC_AREA	BBSM	36	2	0,85234		

ROC_AREA	ECMWF_EPS	36	2	0,92041	0	1
ROC_AREA	BBSM	36	4	0,93646		
ROC_AREA	ECMWF_EPS	36	4	0,95235	0	1
ROC_AREA	BBSM	36	7	0,95613		
ROC_AREA	ECMWF_EPS	36	7	0,96643	0	1
ROC_AREA	BBSM	36	10	0,96468		
ROC_AREA	ECMWF_EPS	36	10	0,96976	0	1
ROC_AREA	BBSM	42	2	0,82451		
ROC_AREA	ECMWF_EPS	42	2	0,90927	0	1
ROC_AREA	BBSM	42	4	0,93253		
ROC_AREA	ECMWF_EPS	42	4	0,95597	0	1
ROC_AREA	BBSM	42	7	0,95948		
ROC_AREA	ECMWF_EPS	42	7	0,96572	0	1
ROC_AREA	BBSM	42	10	0,95837		
ROC_AREA	ECMWF_EPS	42	10	0,96470	0	1
ROC_AREA	BBSM	48	2	0,83634		
ROC_AREA	ECMWF_EPS	48	2	0,90628	0	1
ROC_AREA	BBSM	48	4	0,93498		
ROC_AREA	ECMWF_EPS	48	4	0,95129	0	1
ROC_AREA	BBSM	48	7	0,94485		
ROC_AREA	ECMWF_EPS	48	7	0,95485	0	1
ROC_AREA	BBSM	48	10	0,94331		
ROC_AREA	ECMWF_EPS	48	10	0,95284	0	1
ROC_AREA	BBSM	54	2	0,87276		
ROC_AREA	ECMWF_EPS	54	2	0,92219	0	1
ROC_AREA	BBSM	54	4	0,94295		
ROC_AREA	ECMWF_EPS	54	4	0,95972	0	1
ROC_AREA	BBSM	54	7	0,96171		
ROC_AREA	ECMWF_EPS	54	7	0,96912	0	1
ROC_AREA	BBSM	54	10	0,96477		
ROC_AREA	ECMWF_EPS	54	10	0,96989	0	1
RPSEPS	ECMWF_EPS	54	No	0,03197		
RPSEPS	BBSM	54	No	0,03382	0	1
RPSSANA	BBSM	54	No	0,96108		
RPSSANA	ECMWF_EPS	54	No	0,96320	0	1
RPSSREF	BBSM	54	No	0,43817		
RPSSREF	ECMWF_EPS	54	No	0,46753	0	1
SPREAD	ECMWF_EPS	54	No	1,97689		
SPREAD	BBSM	54	No	2,08822	1	0
TALAGRAND	BBSM	6	No	0,04845		
TALAGRAND	ECMWF_EPS	6	No	0,02722	1	0
TALAGRAND	BBSM	12	No	0,06958		
TALAGRAND	ECMWF_EPS	12	No	0,05012	0	1
TALAGRAND	BBSM	18	No	0,05414		
TALAGRAND	ECMWF_EPS	18	No	0,03584	1	0
TALAGRAND	BBSM	24	No	0,04919		
TALAGRAND	ECMWF_EPS	24	No	0,03037	1	0
TALAGRAND	BBSM	30	No	0,04911		
TALAGRAND	ECMWF_EPS	30	No	0,03165	1	0
TALAGRAND	BBSM	36	No	0,03900		
TALAGRAND	ECMWF_EPS	36	No	0,02216	1	0
TALAGRAND	BBSM	42	No	0,04025		
TALAGRAND	ECMWF_EPS	42	No	0,02354	1	0

TALAGRAND	BBSM	48	No	0,04149		
TALAGRAND	ECMWF_EPS	48	No	0,01934	1	0
TALAGRAND	ECMWF_EPS	54	No	0,02897		
TALAGRAND	BBSM	54	No	0,04804	1	0

16

87

Parameter: Wind Speed in 850hPa level						
Score	Experiment	fc.time	thr	Average (thr, time)	BBSM	ECMWF-EPS
BIAS	BBSM	54	No	0,02843		
BIAS	ECMWF_EPS	54	No	0,32763	1	0
BRIERSCOREEPS	ECMWF_EPS	54	2	0,05362		
BRIERSCOREEPS	BBSM	54	2	0,06053	0	1
BRIERSCOREEPS	ECMWF_EPS	54	4	0,09073		
BRIERSCOREEPS	BBSM	54	4	0,09563	0	1
BRIERSCOREEPS	ECMWF_EPS	54	7	0,08763		
BRIERSCOREEPS	BBSM	54	7	0,08840	0	1
BRIERSCOREEPS	ECMWF_EPS	54	10	0,06047		
BRIERSCOREEPS	BBSM	54	10	0,06205	0	1
BRIERSSANA	BBSM	54	10	0,72851		
BRIERSSANA	ECMWF_EPS	54	10	0,73533	0	1
BRIERSSANA	BBSM	54	7	0,79885		
BRIERSSANA	ECMWF_EPS	54	7	0,80039	0	1
BRIERSSANA	BBSM	54	4	0,86804		
BRIERSSANA	ECMWF_EPS	54	4	0,87475	0	1
BRIERSSANA	BBSM	54	2	0,93356		
BRIERSSANA	ECMWF_EPS	54	2	0,94114	0	1
BRIERSSREF	BBSM	54	2	0,45045		
BRIERSSREF	ECMWF_EPS	54	2	0,51388	0	1
BRIERSSREF	BBSM	54	4	0,42705		
BRIERSSREF	ECMWF_EPS	54	4	0,45799	0	1
BRIERSSREF	BBSM	54	7	0,42155		
BRIERSSREF	ECMWF_EPS	54	7	0,42699	0	1
BRIERSSREF	BBSM	54	10	0,41057		
BRIERSSREF	ECMWF_EPS	54	10	0,42735	0	1
CRPSEPS	ECMWF_EPS	54	No	0,96011		
CRPSEPS	BBSM	54	No	1,00146	0	1
CRPSSANA	BBSM	54	No	0,85914		
CRPSSANA	ECMWF_EPS	54	No	0,86491	0	1
CRPSSREF	BBSM	54	No	0,42262		
CRPSSREF	ECMWF_EPS	54	No	0,44797	0	1
OUTLIERS	ECMWF_EPS	54	No	0,13079		
OUTLIERS	BBSM	54	No	0,17339	0	1
RELIABILITY	ECMWF_EPS	6	2	0,48565		
RELIABILITY	BBSM	6	2	0,56369	0	1
RELIABILITY	ECMWF_EPS	12	2	0,51497		
RELIABILITY	BBSM	12	2	0,57374	0	1
RELIABILITY	ECMWF_EPS	18	2	0,54614		
RELIABILITY	BBSM	18	2	0,63655	0	1
RELIABILITY	ECMWF_EPS	24	2	0,55261		
RELIABILITY	BBSM	24	2	0,62675	0	1
RELIABILITY	ECMWF_EPS	30	2	0,54188		
RELIABILITY	BBSM	30	2	0,61274	0	1
RELIABILITY	ECMWF_EPS	36	2	0,52869		
RELIABILITY	BBSM	36	2	0,58662	0	1

RELIABILITY	ECMWF_EPS	42	2	0,55280		
RELIABILITY	BBSM	42	2	0,64418	0	1
RELIABILITY	ECMWF_EPS	48	2	0,55000		
RELIABILITY	BBSM	48	2	0,63195	0	1
RELIABILITY	ECMWF_EPS	54	2	0,53189		
RELIABILITY	BBSM	54	2	0,62468	0	1
RELIABILITY	ECMWF_EPS	6	4	0,45358		
RELIABILITY	BBSM	6	4	0,51160	1	0
RELIABILITY	ECMWF_EPS	12	4	0,41808		
RELIABILITY	BBSM	12	4	0,46432	1	0
RELIABILITY	ECMWF_EPS	18	4	0,47248		
RELIABILITY	BBSM	18	4	0,55157	0	1
RELIABILITY	ECMWF_EPS	24	4	0,47884		
RELIABILITY	BBSM	24	4	0,52870	0	1
RELIABILITY	ECMWF_EPS	30	4	0,45234		
RELIABILITY	BBSM	30	4	0,51748	1	0
RELIABILITY	ECMWF_EPS	36	4	0,40867		
RELIABILITY	BBSM	36	4	0,47289	1	0
RELIABILITY	ECMWF_EPS	42	4	0,46000		
RELIABILITY	BBSM	42	4	0,55187	0	1
RELIABILITY	ECMWF_EPS	48	4	0,49414		
RELIABILITY	BBSM	48	4	0,54874	0	1
RELIABILITY	ECMWF_EPS	54	4	0,47446		
RELIABILITY	BBSM	54	4	0,54091	0	1
RELIABILITY	ECMWF_EPS	6	7	0,43747		
RELIABILITY	BBSM	6	7	0,47499	1	0
RELIABILITY	ECMWF_EPS	12	7	0,39311		
RELIABILITY	BBSM	12	7	0,42296	1	0
RELIABILITY	ECMWF_EPS	18	7	0,44198		
RELIABILITY	BBSM	18	7	0,51505	1	0
RELIABILITY	ECMWF_EPS	24	7	0,43963		
RELIABILITY	BBSM	24	7	0,48239	1	0
RELIABILITY	ECMWF_EPS	30	7	0,41364		
RELIABILITY	BBSM	30	7	0,48660	1	0
RELIABILITY	ECMWF_EPS	36	7	0,37511		
RELIABILITY	BBSM	36	7	0,44429	1	0
RELIABILITY	ECMWF_EPS	42	7	0,42641		
RELIABILITY	BBSM	42	7	0,52373	1	0
RELIABILITY	ECMWF_EPS	48	7	0,44950		
RELIABILITY	BBSM	48	7	0,50517	1	0
RELIABILITY	ECMWF_EPS	54	7	0,44019		
RELIABILITY	BBSM	54	7	0,50861	1	0
RELIABILITY	ECMWF_EPS	6	10	0,44443		
RELIABILITY	BBSM	6	10	0,44815	1	0
RELIABILITY	ECMWF_EPS	12	10	0,37655		
RELIABILITY	BBSM	12	10	0,38839	1	0
RELIABILITY	ECMWF_EPS	18	10	0,44986		
RELIABILITY	BBSM	18	10	0,50471	1	0
RELIABILITY	ECMWF_EPS	24	10	0,42374		
RELIABILITY	BBSM	24	10	0,46146	1	0
RELIABILITY	ECMWF_EPS	30	10	0,41289		
RELIABILITY	BBSM	30	10	0,46762	1	0
RELIABILITY	ECMWF_EPS	36	10	0,38242		

RELIABILITY	BBSM	36	10	0,42980	1	0
RELIABILITY	ECMWF_EPS	42	10	0,43142		
RELIABILITY	BBSM	42	10	0,51668	1	0
RELIABILITY	ECMWF_EPS	48	10	0,43749		
RELIABILITY	BBSM	48	10	0,48999	1	0
RELIABILITY	ECMWF_EPS	54	10	0,43594		
RELIABILITY	BBSM	54	10	0,49304	1	0
RMSE	ECMWF_EPS	54	No	1,71619		
RMSE	BBSM	54	No	1,75258	0	1
ROC_AREA	BBSM	6	2	0,90723		
ROC_AREA	ECMWF_EPS	6	2	0,92908	0	1
ROC_AREA	BBSM	6	4	0,96151		
ROC_AREA	ECMWF_EPS	6	4	0,96181	0	1
ROC_AREA	BBSM	6	7	0,97674		
ROC_AREA	ECMWF_EPS	6	7	0,97789	0	1
ROC_AREA	BBSM	6	10	0,98154		
ROC_AREA	ECMWF_EPS	6	10	0,98504	0	1
ROC_AREA	BBSM	12	2	0,81464		
ROC_AREA	ECMWF_EPS	12	2	0,86358	0	1
ROC_AREA	BBSM	12	4	0,91588		
ROC_AREA	ECMWF_EPS	12	4	0,92675	0	1
ROC_AREA	BBSM	12	7	0,95525		
ROC_AREA	ECMWF_EPS	12	7	0,96070	0	1
ROC_AREA	BBSM	12	10	0,96544		
ROC_AREA	ECMWF_EPS	12	10	0,97272	0	1
ROC_AREA	BBSM	18	2	0,84841		
ROC_AREA	ECMWF_EPS	18	2	0,87066	0	1
ROC_AREA	BBSM	18	4	0,92108		
ROC_AREA	ECMWF_EPS	18	4	0,92455	0	1
ROC_AREA	BBSM	18	7	0,94871		
ROC_AREA	ECMWF_EPS	18	7	0,95865	0	1
ROC_AREA	BBSM	18	10	0,96223		
ROC_AREA	ECMWF_EPS	18	10	0,97199	0	1
ROC_AREA	BBSM	24	2	0,82481		
ROC_AREA	ECMWF_EPS	24	2	0,85577	0	1
ROC_AREA	BBSM	24	4	0,90750		
ROC_AREA	ECMWF_EPS	24	4	0,91302	0	1
ROC_AREA	BBSM	24	7	0,93948		
ROC_AREA	ECMWF_EPS	24	7	0,94339	0	1
ROC_AREA	BBSM	24	10	0,95485		
ROC_AREA	ECMWF_EPS	24	10	0,96087	0	1
ROC_AREA	BBSM	30	2	0,85277		
ROC_AREA	ECMWF_EPS	30	2	0,86003	0	1
ROC_AREA	BBSM	30	4	0,91766		
ROC_AREA	ECMWF_EPS	30	4	0,91360	1	0
ROC_AREA	BBSM	30	7	0,94521		
ROC_AREA	ECMWF_EPS	30	7	0,94470	1	0
ROC_AREA	BBSM	30	10	0,95628		
ROC_AREA	ECMWF_EPS	30	10	0,96230	0	1
ROC_AREA	BBSM	36	2	0,80829		
ROC_AREA	ECMWF_EPS	36	2	0,84336	0	1
ROC_AREA	BBSM	36	4	0,90492		
ROC_AREA	ECMWF_EPS	36	4	0,90810	0	1

ROC_AREA	BBSM	36	7	0,94330		
ROC_AREA	ECMWF_EPS	36	7	0,94553	0	1
ROC_AREA	BBSM	36	10	0,95629		
ROC_AREA	ECMWF_EPS	36	10	0,96162	0	1
ROC_AREA	BBSM	42	2	0,84757		
ROC_AREA	ECMWF_EPS	42	2	0,86049	0	1
ROC_AREA	BBSM	42	4	0,91199		
ROC_AREA	ECMWF_EPS	42	4	0,91249	0	1
ROC_AREA	BBSM	42	7	0,93431		
ROC_AREA	ECMWF_EPS	42	7	0,93962	0	1
ROC_AREA	BBSM	42	10	0,94680		
ROC_AREA	ECMWF_EPS	42	10	0,95686	0	1
ROC_AREA	BBSM	48	2	0,82361		
ROC_AREA	ECMWF_EPS	48	2	0,86179	0	1
ROC_AREA	BBSM	48	4	0,89433		
ROC_AREA	ECMWF_EPS	48	4	0,90449	0	1
ROC_AREA	BBSM	48	7	0,92054		
ROC_AREA	ECMWF_EPS	48	7	0,92975	0	1
ROC_AREA	BBSM	48	10	0,93748		
ROC_AREA	ECMWF_EPS	48	10	0,94516	0	1
ROC_AREA	BBSM	54	2	0,84125		
ROC_AREA	ECMWF_EPS	54	2	0,86789	0	1
ROC_AREA	BBSM	54	4	0,91506		
ROC_AREA	ECMWF_EPS	54	4	0,91836	0	1
ROC_AREA	BBSM	54	7	0,94296		
ROC_AREA	ECMWF_EPS	54	7	0,94756	0	1
ROC_AREA	BBSM	54	10	0,95533		
ROC_AREA	ECMWF_EPS	54	10	0,96246	0	1
RPSEPS	ECMWF_EPS	54	No	0,07311		
RPSEPS	BBSM	54	No	0,07665	0	1
RPSSANA	BBSM	54	No	0,86695		
RPSSANA	ECMWF_EPS	54	No	0,87305	0	1
RPSSREF	BBSM	54	No	0,42682		
RPSSREF	ECMWF_EPS	54	No	0,45465	0	1
SPREAD	ECMWF_EPS	54	No	1,34822		
SPREAD	BBSM	54	No	1,54611	1	0
TALAGRAND	BBSM	6	No	0,09740		
TALAGRAND	ECMWF_EPS	6	No	0,09222	0	1
TALAGRAND	BBSM	12	No	0,16985		
TALAGRAND	ECMWF_EPS	12	No	0,15065	0	1
TALAGRAND	BBSM	18	No	0,07854		
TALAGRAND	ECMWF_EPS	18	No	0,08911	1	0
TALAGRAND	BBSM	24	No	0,08828		
TALAGRAND	ECMWF_EPS	24	No	0,09853	1	0
TALAGRAND	BBSM	30	No	0,07376		
TALAGRAND	ECMWF_EPS	30	No	0,08685	1	0
TALAGRAND	BBSM	36	No	0,09737		
TALAGRAND	ECMWF_EPS	36	No	0,10037	1	0
TALAGRAND	BBSM	42	No	0,05177		
TALAGRAND	ECMWF_EPS	42	No	0,06297	1	0
TALAGRAND	BBSM	48	No	0,06053		
TALAGRAND	ECMWF_EPS	48	No	0,05366	0	1
TALAGRAND	BBSM	54	No	0,08551		

TALAGRAND	ECMWF_EPS	54	No	0,08689	1	0
					32	71
Parameter: Relative Humidity in 500hPa level						
Score	Experiment	fc.time	thr	Average (thr, time)	BBSM	ECMWF-EPS
BIAS	ECMWF_EPS	54	No	-1,63412		
BIAS	BBSM	54	No	-0,82311	1	0
BRIERSCOREEPS	ECMWF_EPS	54	70	0,07328		
BRIERSCOREEPS	BBSM	54	70	0,07514	0	1
BRIERSCOREEPS	ECMWF_EPS	54	55	0,09241		
BRIERSCOREEPS	BBSM	54	55	0,09427	0	1
BRIERSCOREEPS	ECMWF_EPS	54	40	0,09832		
BRIERSCOREEPS	BBSM	54	40	0,10022	0	1
BRIERSSANA	BBSM	54	70	0,52941		
BRIERSSANA	ECMWF_EPS	54	70	0,54106	0	1
BRIERSSANA	BBSM	54	55	0,64598		
BRIERSSANA	ECMWF_EPS	54	55	0,65302	0	1
BRIERSSANA	BBSM	54	40	0,75270		
BRIERSSANA	ECMWF_EPS	54	40	0,75735	0	1
BRIERSSREF	BBSM	54	55	0,45068		
BRIERSSREF	BBSM	54	40	0,45130	0	1
BRIERSSREF	ECMWF_EPS	54	55	0,46239		
BRIERSSREF	BBSM	54	70	0,46290	0	1
BRIERSSREF	ECMWF_EPS	54	40	0,46290		
BRIERSSREF	ECMWF_EPS	54	70	0,47669	0	1
CRPSEPS	ECMWF_EPS	54	No	6,71116		
CRPSEPS	BBSM	54	No	6,90097	0	1
CRPSSANA	BBSM	54	No	0,81990		
CRPSSANA	ECMWF_EPS	54	No	0,82485	0	1
CRPSSREF	BBSM	54	No	0,45932		
CRPSSREF	ECMWF_EPS	54	No	0,47543	0	1
OUTLIERS	ECMWF_EPS	54	No	0,10702		
OUTLIERS	BBSM	54	No	0,17818	0	1
RELIABILITY	BBSM	6	40	0,44031		
RELIABILITY	ECMWF_EPS	6	40	0,48414	0	1
RELIABILITY	BBSM	12	40	0,48982		
RELIABILITY	ECMWF_EPS	12	40	0,50343	0	1
RELIABILITY	BBSM	18	40	0,50548		
RELIABILITY	ECMWF_EPS	18	40	0,51452	1	0
RELIABILITY	BBSM	24	40	0,55011		
RELIABILITY	ECMWF_EPS	24	40	0,56801	1	0
RELIABILITY	BBSM	30	40	0,54395		
RELIABILITY	ECMWF_EPS	30	40	0,56430	1	0
RELIABILITY	BBSM	36	40	0,53073		
RELIABILITY	ECMWF_EPS	36	40	0,55075	1	0
RELIABILITY	BBSM	42	40	0,52771		
RELIABILITY	ECMWF_EPS	42	40	0,54532	1	0
RELIABILITY	BBSM	48	40	0,55373		
RELIABILITY	ECMWF_EPS	48	40	0,57696	1	0
RELIABILITY	BBSM	54	40	0,54299		
RELIABILITY	ECMWF_EPS	54	40	0,56632	1	0
RELIABILITY	BBSM	6	55	0,46368		
RELIABILITY	ECMWF_EPS	6	55	0,50595	0	1
RELIABILITY	BBSM	12	55	0,48725		

RELIABILITY	ECMWF_EPS	12	55	0,51272	0	1
RELIABILITY	BBSM	18	55	0,48989		
RELIABILITY	ECMWF_EPS	18	55	0,51325	1	0
RELIABILITY	BBSM	24	55	0,53944		
RELIABILITY	ECMWF_EPS	24	55	0,56776	1	0
RELIABILITY	BBSM	30	55	0,53302		
RELIABILITY	ECMWF_EPS	30	55	0,55409	1	0
RELIABILITY	BBSM	36	55	0,51971		
RELIABILITY	ECMWF_EPS	36	55	0,54669	1	0
RELIABILITY	BBSM	42	55	0,51498		
RELIABILITY	ECMWF_EPS	42	55	0,53864	1	0
RELIABILITY	BBSM	48	55	0,54210		
RELIABILITY	ECMWF_EPS	48	55	0,57287	1	0
RELIABILITY	BBSM	54	55	0,53100		
RELIABILITY	ECMWF_EPS	54	55	0,56082	1	0
RELIABILITY	BBSM	6	70	0,47117		
RELIABILITY	ECMWF_EPS	6	70	0,51195	0	1
RELIABILITY	BBSM	12	70	0,46572		
RELIABILITY	ECMWF_EPS	12	70	0,50955	0	1
RELIABILITY	BBSM	18	70	0,46447		
RELIABILITY	ECMWF_EPS	18	70	0,50350	0	1
RELIABILITY	BBSM	24	70	0,50672		
RELIABILITY	ECMWF_EPS	24	70	0,54380	1	0
RELIABILITY	BBSM	30	70	0,50527		
RELIABILITY	ECMWF_EPS	30	70	0,52850	1	0
RELIABILITY	BBSM	36	70	0,49352		
RELIABILITY	ECMWF_EPS	36	70	0,52497	1	0
RELIABILITY	BBSM	42	70	0,50172		
RELIABILITY	ECMWF_EPS	42	70	0,53055	1	0
RELIABILITY	BBSM	48	70	0,51011		
RELIABILITY	ECMWF_EPS	48	70	0,55198	1	0
RELIABILITY	BBSM	54	70	0,50585		
RELIABILITY	ECMWF_EPS	54	70	0,54360	1	0
RMSE	ECMWF_EPS	54	No	13,29561		
RMSE	BBSM	54	No	13,46354	0	1
ROC_AREA	BBSM	6	40	0,97500		
ROC_AREA	ECMWF_EPS	6	40	0,98168	0	1
ROC_AREA	BBSM	6	55	0,97359		
ROC_AREA	ECMWF_EPS	6	55	0,97789	0	1
ROC_AREA	BBSM	6	70	0,97233		
ROC_AREA	ECMWF_EPS	6	70	0,97693	0	1
ROC_AREA	BBSM	12	40	0,96123		
ROC_AREA	ECMWF_EPS	12	40	0,96615	0	1
ROC_AREA	BBSM	12	55	0,95326		
ROC_AREA	ECMWF_EPS	12	55	0,96030	0	1
ROC_AREA	BBSM	12	70	0,95293		
ROC_AREA	ECMWF_EPS	12	70	0,95910	0	1
ROC_AREA	BBSM	18	40	0,95046		
ROC_AREA	ECMWF_EPS	18	40	0,95441	0	1
ROC_AREA	BBSM	18	55	0,94137		
ROC_AREA	ECMWF_EPS	18	55	0,94659	0	1
ROC_AREA	BBSM	18	70	0,94293		
ROC_AREA	ECMWF_EPS	18	70	0,94739	0	1

ROC_AREA	BBSM	24	40	0,93613		
ROC_AREA	ECMWF_EPS	24	40	0,94075	0	1
ROC_AREA	BBSM	24	55	0,92690		
ROC_AREA	ECMWF_EPS	24	55	0,92781	0	1
ROC_AREA	BBSM	24	70	0,92754		
ROC_AREA	ECMWF_EPS	24	70	0,92912	0	1
ROC_AREA	BBSM	30	40	0,92771		
ROC_AREA	ECMWF_EPS	30	40	0,93204	0	1
ROC_AREA	BBSM	30	55	0,91797		
ROC_AREA	ECMWF_EPS	30	55	0,92001	0	1
ROC_AREA	BBSM	30	70	0,91760		
ROC_AREA	ECMWF_EPS	30	70	0,92099	0	1
ROC_AREA	BBSM	36	40	0,91535		
ROC_AREA	ECMWF_EPS	36	40	0,91943	0	1
ROC_AREA	BBSM	36	55	0,90436		
ROC_AREA	ECMWF_EPS	36	55	0,91094	0	1
ROC_AREA	BBSM	36	70	0,90408		
ROC_AREA	ECMWF_EPS	36	70	0,91067	0	1
ROC_AREA	BBSM	42	40	0,90593		
ROC_AREA	ECMWF_EPS	42	40	0,91074	0	1
ROC_AREA	BBSM	42	55	0,89651		
ROC_AREA	ECMWF_EPS	42	55	0,90249	0	1
ROC_AREA	BBSM	42	70	0,89921		
ROC_AREA	ECMWF_EPS	42	70	0,90478	0	1
ROC_AREA	BBSM	48	40	0,89650		
ROC_AREA	ECMWF_EPS	48	40	0,90345	0	1
ROC_AREA	BBSM	48	55	0,88516		
ROC_AREA	ECMWF_EPS	48	55	0,89375	0	1
ROC_AREA	BBSM	48	70	0,88583		
ROC_AREA	ECMWF_EPS	48	70	0,89479	0	1
ROC_AREA	BBSM	54	40	0,88890		
ROC_AREA	ECMWF_EPS	54	40	0,89857	0	1
ROC_AREA	BBSM	54	55	0,87864		
ROC_AREA	ECMWF_EPS	54	55	0,88932	0	1
ROC_AREA	BBSM	54	70	0,87878		
ROC_AREA	ECMWF_EPS	54	70	0,89039	0	1
RPSEPS	ECMWF_EPS	54	No	0,08800		
RPSEPS	BBSM	54	No	0,08988	0	1
RPSSANA	BBSM	54	No	0,67565		
RPSSANA	ECMWF_EPS	54	No	0,68240	0	1
RPSSREF	BBSM	54	No	0,45430		
RPSSREF	ECMWF_EPS	54	No	0,46655	0	1
SPREAD	ECMWF_EPS	54	No	10,19411		
SPREAD	BBSM	54	No	10,87700	1	0
TALAGRAND	BBSM	6	No	0,20140		
TALAGRAND	ECMWF_EPS	6	No	0,10679	0	1
TALAGRAND	BBSM	12	No	0,15595		
TALAGRAND	ECMWF_EPS	12	No	0,09087	0	1
TALAGRAND	BBSM	18	No	0,11828		
TALAGRAND	ECMWF_EPS	18	No	0,06536	0	1
TALAGRAND	BBSM	24	No	0,08110		
TALAGRAND	ECMWF_EPS	24	No	0,03890	0	1
TALAGRAND	BBSM	30	No	0,06940		

TALAGRAND	ECMWF_EPS	30	No	0,03370	0	1
TALAGRAND	BBSM	36	No	0,07203		
TALAGRAND	ECMWF_EPS	36	No	0,03718	0	1
TALAGRAND	BBSM	42	No	0,06279		
TALAGRAND	ECMWF_EPS	42	No	0,03164	1	0
TALAGRAND	BBSM	48	No	0,05452		
TALAGRAND	ECMWF_EPS	48	No	0,02434	1	0
TALAGRAND	ECMWF_EPS	54	No	0,05035		
TALAGRAND	BBSM	54	No	0,09650	0	1

24

58

Parameter: Relative Humidity in 850hPa level						
Score	Experiment	fc.time	thr	Average (thr, time)	BBSM	ECMWF-EPS
BIAS	BBSM	54	No	-1,36375		
BIAS	ECMWF_EPS	54	No	-0,00590	0	1
BRIERSCOREEPS	ECMWF_EPS	54	40	0,05135		
BRIERSCOREEPS	BBSM	54	40	0,05793	0	1
BRIERSCOREEPS	ECMWF_EPS	54	55	0,07492		
BRIERSCOREEPS	BBSM	54	55	0,08677	0	1
BRIERSCOREEPS	ECMWF_EPS	54	70	0,09218		
BRIERSCOREEPS	BBSM	54	70	0,10947	0	1
BRIERSSANA	BBSM	54	70	0,74700		
BRIERSSANA	ECMWF_EPS	54	70	0,78658	0	1
BRIERSSANA	BBSM	54	55	0,86566		
BRIERSSANA	ECMWF_EPS	54	55	0,88403	0	1
BRIERSSANA	BBSM	54	40	0,92930		
BRIERSSANA	ECMWF_EPS	54	40	0,93733	0	1
BRIERSSREF	BBSM	54	40	0,43281		
BRIERSSREF	ECMWF_EPS	54	40	0,49796	0	1
BRIERSSREF	BBSM	54	55	0,43635		
BRIERSSREF	ECMWF_EPS	54	55	0,51406	0	1
BRIERSSREF	BBSM	54	70	0,43340		
BRIERSSREF	ECMWF_EPS	54	70	0,52153	0	1
CRPSEPS	ECMWF_EPS	54	No	5,11382		
CRPSEPS	BBSM	54	No	5,76833	0	1
CRPSSANA	BBSM	54	No	0,90865		
CRPSSANA	ECMWF_EPS	54	No	0,91901	0	1
CRPSSREF	BBSM	54	No	0,41920		
CRPSSREF	ECMWF_EPS	54	No	0,48465	0	1
OUTLIERS	ECMWF_EPS	54	No	0,19851		
OUTLIERS	BBSM	54	No	0,27076	0	1
RELIABILITY	BBSM	6	40	0,38358		
RELIABILITY	ECMWF_EPS	6	40	0,45855	0	1
RELIABILITY	BBSM	6	55	0,48359		
RELIABILITY	ECMWF_EPS	6	55	0,49845	0	1
RELIABILITY	ECMWF_EPS	6	70	0,52553		
RELIABILITY	BBSM	6	70	0,53881	0	1
RELIABILITY	ECMWF_EPS	12	40	0,46389		
RELIABILITY	BBSM	12	40	0,47100	1	0
RELIABILITY	ECMWF_EPS	12	55	0,48839		
RELIABILITY	BBSM	12	55	0,54593	0	1
RELIABILITY	ECMWF_EPS	12	70	0,52249		
RELIABILITY	BBSM	12	70	0,60783	0	1
RELIABILITY	ECMWF_EPS	18	40	0,45440		

RELIABILITY	BBSM	18	40	0,49986	1	0
RELIABILITY	ECMWF_EPS	18	55	0,47560		
RELIABILITY	BBSM	18	55	0,54632	0	1
RELIABILITY	ECMWF_EPS	18	70	0,51721		
RELIABILITY	BBSM	18	70	0,58432	0	1
RELIABILITY	ECMWF_EPS	24	40	0,51780		
RELIABILITY	BBSM	24	40	0,53822	0	1
RELIABILITY	ECMWF_EPS	24	55	0,52714		
RELIABILITY	BBSM	24	55	0,57710	0	1
RELIABILITY	ECMWF_EPS	24	70	0,49899		
RELIABILITY	BBSM	24	70	0,56053	0	1
RELIABILITY	BBSM	30	40	0,51825		
RELIABILITY	ECMWF_EPS	30	40	0,52407	1	0
RELIABILITY	ECMWF_EPS	30	55	0,52758		
RELIABILITY	BBSM	30	55	0,57154	1	0
RELIABILITY	ECMWF_EPS	30	70	0,51163		
RELIABILITY	BBSM	30	70	0,57032	1	0
RELIABILITY	ECMWF_EPS	36	40	0,49570		
RELIABILITY	BBSM	36	40	0,53192	0	1
RELIABILITY	ECMWF_EPS	36	55	0,52319		
RELIABILITY	BBSM	36	55	0,58716	0	1
RELIABILITY	ECMWF_EPS	36	70	0,54350		
RELIABILITY	BBSM	36	70	0,62443	0	1
RELIABILITY	ECMWF_EPS	42	40	0,48740		
RELIABILITY	BBSM	42	40	0,54296	0	1
RELIABILITY	ECMWF_EPS	42	55	0,51924		
RELIABILITY	BBSM	42	55	0,58033	0	1
RELIABILITY	ECMWF_EPS	42	70	0,53456		
RELIABILITY	BBSM	42	70	0,60287	0	1
RELIABILITY	ECMWF_EPS	48	40	0,53309		
RELIABILITY	BBSM	48	40	0,56315	0	1
RELIABILITY	ECMWF_EPS	48	55	0,54016		
RELIABILITY	BBSM	48	55	0,59557	0	1
RELIABILITY	ECMWF_EPS	48	70	0,51020		
RELIABILITY	BBSM	48	70	0,57853	0	1
RELIABILITY	ECMWF_EPS	54	40	0,51813		
RELIABILITY	BBSM	54	40	0,53699	0	1
RELIABILITY	ECMWF_EPS	54	55	0,53116		
RELIABILITY	BBSM	54	55	0,58402	0	1
RELIABILITY	ECMWF_EPS	54	70	0,51935		
RELIABILITY	BBSM	54	70	0,58679	0	1
RMSE	ECMWF_EPS	54	No	9,03997		
RMSE	BBSM	54	No	9,98986	0	1
ROC_AREA	BBSM	6	40	0,91954		
ROC_AREA	ECMWF_EPS	6	40	0,94207	0	1
ROC_AREA	BBSM	6	55	0,94863		
ROC_AREA	ECMWF_EPS	6	55	0,95482	0	1
ROC_AREA	BBSM	6	70	0,94019		
ROC_AREA	ECMWF_EPS	6	70	0,95048	0	1
ROC_AREA	BBSM	12	40	0,87924		
ROC_AREA	ECMWF_EPS	12	40	0,92066	0	1
ROC_AREA	BBSM	12	55	0,90469		
ROC_AREA	ECMWF_EPS	12	55	0,94637	0	1

ROC_AREA	BBSM	12	70	0,89898		
ROC_AREA	ECMWF_EPS	12	70	0,93695	0	1
ROC_AREA	BBSM	18	40	0,91666		
ROC_AREA	ECMWF_EPS	18	40	0,93954	0	1
ROC_AREA	BBSM	18	55	0,92365		
ROC_AREA	ECMWF_EPS	18	55	0,95324	0	1
ROC_AREA	BBSM	18	70	0,91983		
ROC_AREA	ECMWF_EPS	18	70	0,94895	0	1
ROC_AREA	BBSM	24	40	0,93193		
ROC_AREA	ECMWF_EPS	24	40	0,94615	0	1
ROC_AREA	BBSM	24	55	0,92989		
ROC_AREA	ECMWF_EPS	24	55	0,94419	0	1
ROC_AREA	BBSM	24	70	0,92783		
ROC_AREA	ECMWF_EPS	24	70	0,94092	0	1
ROC_AREA	BBSM	30	40	0,92323		
ROC_AREA	ECMWF_EPS	30	40	0,93726	0	1
ROC_AREA	BBSM	30	55	0,93439		
ROC_AREA	ECMWF_EPS	30	55	0,94144	0	1
ROC_AREA	BBSM	30	70	0,92693		
ROC_AREA	ECMWF_EPS	30	70	0,93688	0	1
ROC_AREA	BBSM	36	40	0,90442		
ROC_AREA	ECMWF_EPS	36	40	0,93019	0	1
ROC_AREA	BBSM	36	55	0,91917		
ROC_AREA	ECMWF_EPS	36	55	0,94674	0	1
ROC_AREA	BBSM	36	70	0,91297		
ROC_AREA	ECMWF_EPS	36	70	0,93940	0	1
ROC_AREA	BBSM	42	40	0,92712		
ROC_AREA	ECMWF_EPS	42	40	0,94280	0	1
ROC_AREA	BBSM	42	55	0,93125		
ROC_AREA	ECMWF_EPS	42	55	0,95174	0	1
ROC_AREA	BBSM	42	70	0,92612		
ROC_AREA	ECMWF_EPS	42	70	0,94764	0	1
ROC_AREA	BBSM	48	40	0,93162		
ROC_AREA	ECMWF_EPS	48	40	0,94108	0	1
ROC_AREA	BBSM	48	55	0,93021		
ROC_AREA	ECMWF_EPS	48	55	0,93991	0	1
ROC_AREA	BBSM	48	70	0,92611		
ROC_AREA	ECMWF_EPS	48	70	0,93547	0	1
ROC_AREA	BBSM	54	40	0,92488		
ROC_AREA	ECMWF_EPS	54	40	0,93447	0	1
ROC_AREA	BBSM	54	55	0,93038		
ROC_AREA	ECMWF_EPS	54	55	0,93512	0	1
ROC_AREA	BBSM	54	70	0,92322		
ROC_AREA	ECMWF_EPS	54	70	0,92989	0	1
RPSEPS	ECMWF_EPS	54	No	0,07282		
RPSEPS	BBSM	54	No	0,08472	0	1
RPSSANA	BBSM	54	No	0,86610		
RPSSANA	ECMWF_EPS	54	No	0,88492	0	1
RPSSREF	BBSM	54	No	0,43449		
RPSSREF	ECMWF_EPS	54	No	0,51413	0	1
SPREAD	ECMWF_EPS	54	No	6,06000		
SPREAD	BBSM	54	No	7,12322	1	0
TALAGRAND	BBSM	6	No	0,22129		

TALAGRAND	ECMWF_EPS	6	No	0,21683	0	1
TALAGRAND	BBSM	12	No	0,19707		
TALAGRAND	ECMWF_EPS	12	No	0,15602	0	1
TALAGRAND	BBSM	18	No	0,15868		
TALAGRAND	ECMWF_EPS	18	No	0,12770	0	1
TALAGRAND	BBSM	24	No	0,10926		
TALAGRAND	ECMWF_EPS	24	No	0,12189	1	0
TALAGRAND	BBSM	30	No	0,09843		
TALAGRAND	ECMWF_EPS	30	No	0,10577	1	0
TALAGRAND	BBSM	36	No	0,09989		
TALAGRAND	ECMWF_EPS	36	No	0,08143	0	1
TALAGRAND	BBSM	42	No	0,08740		
TALAGRAND	ECMWF_EPS	42	No	0,06688	0	1
TALAGRAND	BBSM	48	No	0,06072		
TALAGRAND	ECMWF_EPS	48	No	0,06396	1	0
TALAGRAND	ECMWF_EPS	54	No	0,11088		
TALAGRAND	BBSM	54	No	0,12152	0	1
					9	73

Bibliography

- Anderson, J. (1997). The impact of dynamical constraints on the selection of initial conditions for ensemble predictions: low-order perfect model results. *MONTHLY WEATHER REVIEW*, (125):2969–2983.
- Barkmeijer, J., Buizza, R., Palmer, T., Puri, K., and Mahfouf, J.-F. (2001). Tropical singular vectors computed with linearized diabatic physics. *QUARTERLY JOURNAL OF THE ROYAL METEOROLOGICAL SOCIETY*, (127):685–708.
- Bowler, N. E., Arribas, A., Mylne, K. R., Robertson, K. B., and Beare, S. E. (2008). The MOGREPS short-range ensemble prediction system. *QUARTERLY JOURNAL OF THE ROYAL METEOROLOGICAL SOCIETY*, (134):703–722.
- Brankovic, C., Matjacic, B., Ivatek-Sahdan, S., and Buizza, R. (2008). Downscaling of ECMWF Ensemble Forecasts for Cases of Severe Weather: Ensemble Statistics and Cluster Analysis. *MONTHLY WEATHER REVIEW*, (136):3323–3342.
- Brier, G. (1950). Verification of forecasts expressed in terms of probability. *MONTHLY WEATHER REVIEW*, (78):1–3.
- Buizza, R., Houtekamer, P., Toth, Z., Pellerin, G., Wei, M., and Zhu, Y. (2005). A comparison of the ECMWF, MSC and NCEP Global Ensemble Prediction System. *MONTHLY WEATHER REVIEW*, (133):1076–1097.
- Buizza, R., Miller, M., and Palmer, T. (1999). Stochastic simulation of model uncertainties. *QUARTERLY JOURNAL OF THE ROYAL METEOROLOGICAL SOCIETY*, (125):2887–2908.
- Buizza, R. and Palmer, T. (1995). The singular-vector structure of the atmospheric global circulation. *Journal of the Atmospheric Sciences*, (52):1434–1456.
- Ehrendorfer, M. (1994). The Liouville equation and its potential usefulness for the prediction of forecast skill. *MONTHLY WEATHER REVIEW*, (122):703–713.
- Ehrendorfer, M. and Tribbia, J. (1997). Optimal prediction of forecast error covariances through singular vectors. *Journal of the Atmospheric Sciences*, (54):286–313.
- Epstein, E. S. (1969). Stochastic dynamic prediction. *Tellus*, (21):739–759.
- Finley, J. (1884). Tornado predictions. *American Meteorological Journal*, pages 85–88.
- Gneiting, T. ., Raftery, A., Westveld, A., and Goldman, T. (2004). Calibrated Probabilistic Forecasting Using Ensemble Model Output Statistics and Minimum CRPS Estimation. *MONTHLY WEATHER REVIEW*, (133):1098–1118.
- Hagedorn, R. (2010). *Post-Processing of EPS Forecasts: Training Course 2010*: http://www.ecmwf.int/newsevents/training/meteorological_presentations/MET_PR.html.
- Hamill, T. (2001). Interpretation of rank histograms for verifying ensemble forecasts. *MONTHLY WEATHER REVIEW*, (129):550–560.

- Hersbach, H. (2000). Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather and Forecasting*, (15):559–570.
- Houtekamer, P., Lefaivre, L., and Derome, J. (1996). The RPN ensemble prediction system. *Proc. ECMWF Seminar on Predictability- ECMWF*, pages 121–246.
- Jolliffe, I. T. and Stephenson, D. B. (2003). *Forecast verification: A practitioner's guide in atmospheric science*. Wiley and J. Wiley.
- Kann, A., Wittmann, C., Wang, Y., and Ma, X. (2009). Calibrating 2m-Temperature of Limited Area Ensemble Forecasts Using High-Resolution Analysis. *American Meteorological Society*.
- Katz, R. (1982). Statistical evaluation of climate experiments with general circulation models: a parametric time series modeling approach. *Journal of the Atmospheric Sciences*, (39):1446–1455.
- Katz, R. and Murphy, A. (1997). *Forecast value: prototype decision-making models*. In: *Economic Value of Weather and Climate Forecasts* (eds. R.W. Katz and A.H. Murphy): 183–217. Cambridge Univ. Press, Cambridge.
- Leith, C. (1974). Theoretical skill of Monte-Carlo forecasts. *MONTHLY WEATHER REVIEW*, pages 409–418.
- Leutbecher, M. and Palmer, T. (2008). Ensemble Forecasting. *Journal of Computational Physics*, (227):3515–3539.
- Lorenz, E. (1965). A study of the predictability of a 28-variable atmospheric model. *Tellus*, (17):321–333.
- Lynch, P. (2006). *The emergence of numerical weather prediction: Richardson's dream*. Cambridge University Press and Cambridge Univ. Press, Cambridge.
- Marsigli, C., Boccanera, F., Montani, A., and Paccagnella, T. (2005). The COSMO-LEPS mesoscale ensemble system: validation of the. *Nonlinear Processes in Geophysics*, (12):527–536.
- Mason, I. (1979). On reducing probability forecasts to yes/no forecasts. *MONTHLY WEATHER REVIEW*, (107):207–211.
- Mason, I. (1980). Decision-theoretic evaluation of probabilistic predictions. In: *WMO Symposium on Probabilistic and Statistical Methods in Weather Forecasting*, Nice, 8–12 September. pages 219–228.
- Muller, R. (1944). Verification of short-range weather forecasts (a survey of the literature). *Bulletin of the American Meteorological Society*, (25):18–27, 47–53, 88–95.
- Murphy, A. (1991). Forecast verification: its complexity and dimensionality. *MONTHLY WEATHER REVIEW*, (119):1590–1601.
- Murphy, A. and Winkler, R. (1987). A general framework for forecast verification. *MONTHLY WEATHER REVIEW*, (115):1330–1338.
- Palmer, T. (1993). Extended-range atmospheric prediction and the Lorenz model. *Bulletin of the American Meteorological Society*, (74):49–65.

- Palmer, T. and Hagedorn, R. (2006). *Predictability of weather and climate*. Cambridge University Press, Cambridge.
- Palmer, T., Mureau, R., and Molteni, F. (1990). The Monte Carlo forecast. *Weather*, pages 198–207.
- Papula L. (2007). *Mathematik für Ingenieure und Naturwissenschaftler, Band 3*. Vieweg & Sohn Verlag, Wiesbaden, 11 edition.
- Persson, A. and Grazzini, F. (2005). *User guide to ECMWF forecast products*: <http://www.ecmwf.int/products/forecasts/guide/>.
- Stanski, H. R., Wilson, L. J., and Burrows, W. R. (1989). *Survey of common verification methods in meteorology: WMO World Weather Watch Technical Report No.8, WMO/TD No. 358*: http://www.cawcr.gov.au/projects/verification/Stanski_et_al/Stanski_et_al.html.
- Sutton, C., Hamill, T., and Warner, T. (2006). Will perturbing soil moisture improve warm-season ensemble forecast? *MONTHLY WEATHER REVIEW*, (134):3174–3189.
- Talagrand, O., Vautard, R., and Strauss, B. (1997). Evaluation of probabilistic prediction systems: Proceedings. *ECMWF Workshop on Predictability*, pages 1–25.
- Toth, Z. and Kalnay, E. (1997). Ensemble Forecasting at NCEP: The breeding method. *MONTHLY WEATHER REVIEW*, (125):3297–3318.
- Tracton, M., Du, J., Toth, Z., and Juang, H. (1998). Short-range ensemble forecasting (SREF) at NCEP/EMC: 12th conference on Numerical Weather Prediction, Phoenix. *American Meteorological Society*, pages 269–272.
- Wang, Y., Bellus, M., Wittmann, C., Steinheimer, M., Ivatek-Sahdan, S., Kann, A., Tian, W., Ma, X., Tascu, S., and Bazile, E. (2011). The Central European limited area ensemble forecasting system: ALADIN-LAEF. *QUARTERLY JOURNAL OF THE ROYAL METEOROLOGICAL SOCIETY*, (137):483–502.
- Wang, Y., Haiden, T., and Kann, A. (2006). The operational Limited Area Modelling system at ZAMG: ALADIN-Austria. *Österreichische Beiträge zu Meteorologie und Geophysik*, (37).
- Wilks, D. (2006). *Statistical Methods in the Atmospheric Sciences*, volume 91 of *International geophysics series*. Academic Press, Amsterdam.

List of Figures

4.1	Distribution of the vertical levels in the TL799L91 model (Persson and Grazzini, 2005)	11
4.2	Main physics represented in the ECMWF model Source: Persson and Grazzini, User guide to ECMWF forecast products, 2005	12
4.3	Summary of observations received at the ECMWF, 5 July 2004 Source: Persson and Grazzini, User guide to ECMWF forecast products, 2005	13
4.4	Orography and land-sea mask of the ECMWF general circulation model TL799L91; Source: Persson and Grazzini, User guide to ECMWF forecast products, 2005	15
4.5	Domain and topography of ALADIN-AUSTRIA Source: Wang et al., The operational Limited Area Modeling system at ZAMG: ALADIN-AUSTRIA, 2006	15
5.1	The basic principle of ensemble-based probabilistic forecasting Source: http://www.ecmwf.int/about/information-leaflets/EPS.pdf /status 18.6.2010	17
5.2	Schematic Illustration of concepts in ensemble forecasting Source: Wilks, Statistical methods in the Atmospheric Sciences, 2006 . . .	17
5.3	ALADIN-LAEF-domain, source: ZAMG	24
5.4	Visualization of a simplified breeding mechanism, source: Karin Schmeisser	25
5.5	The physical settings used in ALADIN-LAEF source: Wang et al., The Central European Limited Area Ensemble Forecasting System ALADIN-LAEF, 2011	26
5.6	ALADIN-LAEF - Generation of IC surface perturbations, source: Wang et al. 2011	29
5.7	ALADIN-LAEF - Generation of IC upper level perturbations, Source: Wang et al. 2011	30
5.8	ALADIN-LAEF - Generation of Perturbations, source: Wang et al. 2011 .	31
5.9	Blending Example - Perturbation author: Martin Bellus, Project ZAMG-SHMU	34
5.10	Blending Example - Resulting Temperature Field author: Martin Bellus, Project ZAMG-SHMU	35
7.1	A basic model for the verification of weather elements source: Stanski et al., 1989, Survey of common verification methods in meteorology	39
7.2	Reliability-Resolution-Sharpness author: Prof. Manfred Dorninger, University Vienna	46
7.3	ROC; source: Jolliffe and Stephenson, 2003	51
7.4	Talagrand diagram; source: ECMWF user guide, 2006	52
7.5	cumulative distribution function, author: Karin Schmeisser	53

7.6	cumulative distribution function, author: Karin Schmeisser	53
8.1	mean sea level pressure - BIAS-RMSE-SPREAD	57
8.2	mean sea level pressure - Brier Score, threshold 1005 hPa	58
8.3	mean sea level pressure - Brier Skill Score, threshold 1005 hPa (top) and 1015 hPa (below), reference: Analysis . . .	59
8.4	Comparison of Brier Skill Score for two different references: ALADIN analysis (top) and ALADIN forecasts (below), mean sea level pressure, threshold 1015hPa	60
8.5	mean sea level pressure - Ranked Probability Score	61
8.6	mean sea level pressure - ROC, 36h-forecast - threshold: 1005hPa	61
8.7	mean sea level pressure - Reliability diagram, 6h- and 24h- forecasts, threshold: 1010hPa	62
8.8	temperature anomaly - Bias-RMSE-Spread	64
8.9	temperature anomaly - Brier Score, threshold: >2°C	65
8.10	temperature anomaly - Brier Score, thresholds: <-2°C	66
8.11	temperature anomaly - Ranked Probability Score	66
8.12	temperature anomaly - Ranked Probability Score	67
8.13	temperature anomaly - Continuous Ranked Probability Score	67
8.14	temperature anomaly - Continuous Ranked Probability Score	68
8.15	temperature anomaly - Reliability diagram forecast lead time:36h, negative temperature anomaly	68
8.16	temperature anomaly - Reliability diagram forecast lead time: 42h, positive temperature anomaly	69
8.17	temperature anomaly - Talagrand diagram, forecast lead time: 12 hours	69
8.18	wind speed - BIAS-RMSE-Spread	70
8.19	wind speed - Brier Score, threshold: 6 m/s	71
8.21	wind speed - percentage of outliers	71
8.20	wind speed - Brier Skill Score, reference:analysis	72
8.22	wind speed - Reliability, Poor Resolution indicated by nearly horizontal curve	73
8.23	wind speed - Reliability	73
8.24	wind speed - Relative Operating Characteristic	75
8.25	total precipitation - BIAS-SPREAD-RMSE	76
8.26	total precipitation - Brier Score: threshold 1 mm	77
8.27	total precipitation - Brier Score: threshold 25 mm	78
8.28	total precipitation - Continuous Ranked Probability Score	79
8.29	total precipitation - Percentage of Outliers	79
8.30	total precipitation - Reliability diagram: fc.time 18h, thr. 10 mm	80
8.31	total precipitation - Reliability diagram: fc.time 30h, thr. 10 mm	80
8.32	total precipitation - Total Precipitation for selected stations June 2007, horizontal axis: days, vertical axis: mm/24h	81
8.33	total precipitation - Total Precipitation for selected stations July 2007, horizontal axis: days, vertical axis: mm/24h	81
8.34	total precipitation - Total Precipitation for selected stations August 2007, horizontal axis: days, vertical axis: mm/24h	82
8.35	total precipitation - ROC: fc.time 18h, thr. 5 mm	82
8.36	total precipitation - ROC: fc.time 18h, thr. 25 mm	83
8.37	total precipitation - Talagrand diagram: 18h and 54h forecasts	84
8.38	Geopotential in 500hPa - Bias, RMSE and Ensemble Spread	87

8.39	Geopotential in 500hPa - Outliers	88
8.40	Geopotential in 500hPa Talagrand diagram, forecast lead times: 6h and 18h	89
8.41	Geopotential in 850hPa - Bias, RMSE and ensemble spread	90
8.42	Geopotential in 850hPa - Continuous Ranked Probability Score	91
8.43	Geopotential in 850hPa - Talagrand diagram - 12h forecasts	91
8.44	Geopotential in 850hPa - Talagrand diagram - 30h forecasts	92
8.45	Temperature Anomaly in 500hPa - Bias, RMSE and Ensemble Spread . .	93
8.46	Temperature Anomaly 500hPa - Continuous Ranked Probability Score .	94
8.47	Temperature Anomaly 500hPa - Continuous Ranked Probability Skill Score - Reference: Analysis	95
8.48	Temperature Anomaly 500hPa - Outliers	95
8.49	Temperature Anomaly 500hPa - Reliability	95
8.50	Temperature Anomaly 500hPa - Reliability	96
8.51	Temperature Anomaly 500hPa - Relative Operating Characteristic	97
8.52	Temperature Anomaly in 850hPa - Bias, RMSE and Ensemble Spread . .	98
8.53	Temperature Anomaly in 850hPa - Outliers	99
8.54	Temperature Anomaly in 850hPa - Reliability of forecasts, 12h and 48h .	99
8.55	Wind Speed in 500hPa - Bias, RMSE and ensemble spread	100
8.56	Wind Speed in 500hPa - Continuous Ranked Probability Score	101
8.57	Wind Speed in 500hPa - Outliers	101
8.58	Wind Speed in 500hPa - Reliability	102
8.59	Wind Speed in 500hPa - Reliability	102
8.60	Wind Speed in 500hPa - Relative Operating Characteristic	103
8.61	Wind Speed in 500hPa - Relative Operating Characteristic	103
8.62	Wind Speed in 850hPa - Bias, RMSE and ensemble spread	104
8.63	Wind Speed in 850hPa - Continuous Ranked Probability Score	105
8.64	Wind Speed in 850hPa - Reliability, forecast lead time: 24h, thr. 2 m/s . .	105
8.65	Wind Speed in 850hPa - Reliability forecast lead times: 18h and threshold 10 m/s	106
8.66	Wind Speed in 850hPa - Talagrand diagram, forecast lead time: 48h . . .	106
8.67	Relative Humidity in 500hPa - Bias, RMSE and ensemble spread	107
8.68	Relative Humidity in 500hPa - Continuous Ranked Probability Score . .	108
8.69	Relative Humidity in 500hPa - Outliers	109
8.70	Relative Humidity in 500hPa - Reliability forecast lead time: 6h, threshold 70%	109
8.71	Relative Humidity in 500hPa - Reliability forecast lead time: 36h, threshold 70%	110
8.72	Relative Humidity in 850hPa - Bias, RMSE and ensemble spread	111
8.73	Relative Humidity in 850hPa - Brier Score, threshold 40%	112
8.74	Relative Humidity in 850hPa - Brier Score, threshold 55%	113
8.75	Relative Humidity in 850hPa - Brier Score, threshold 70%	113
8.76	Relative Humidity in 850hPa - Continuous Ranked Probability Score . .	114
8.77	Relative Humidity in 850hPa - Percentage of Outliers	114
8.78	Relative Humidity in 850hPa - Reliability, 40% relative humidity	115
8.79	Relative Humidity in 850hPa - Reliability, 70% relative humidity	115
8.80	Relative Humidity in 850hPa - ROC	116
8.81	Relative Humidity in 850hPa - Talagrand diagram	116

9.1	BIAS, RMSE and ensemble spread of the raw and calibrated ALADIN-LAEF system, source: Kann, 2009	123
10.1	Future domain of ALADIN-LAEF	128

12 Danksagung

Ich habe bei der Erstellung dieser Diplomarbeit sehr viel gelernt und möchte mich an dieser Stelle bei vielen Menschen bedanken, die mich in den schwierigen Phasen unterstützt haben. Ich bedanke mich recht herzlich bei Herrn Prof. Haimberger für die hilfreichen Kommentare und Vorschläge. Weiters möchte ich mich ganz herzlich bei Herrn Dr. Wang bedanken, der es mir ermöglicht hat, an der Zentralanstalt für Meteorologie und Geodynamik (ZAMG) eine Diplomarbeit zu schreiben und zeitgleich erste Berufserfahrungen zu sammeln. Ein ganz besonderer Dank gilt meinem Betreuer Florian Weidle, der sehr viel Zeit aufgebracht hat um mich mit den Programmen vertraut zu machen und alle meine Fragen zu beantworten. Vielen vielen Dank! Zuletzt möchte ich mich noch bei meiner Familie und Mathias bedanken, die mich immer wieder aufgemuntert haben und eine große Unterstützung waren.

Lebenslauf

Persönliche Angaben

Name:	Karin Schmeisser
Geburtstag/Geburtsort	06.10.1985 in Kirchdorf an der Krems
Staatsangehörigkeit:	Österreich
Familienstand:	Ledig

Ausbildung

Seit 2004	Meteorologie und Geophysikstudium an der Hauptuniversität Wien
2000 – 2004	Adalbert Stifter Oberstufenrealgymnasium in Linz Schwerpunkt Musik

Praktika und Berufserfahrung

seit April 2010	ZAMG Kundenservice (Satellitenmeteorologie)
2009 August	ZAMG (Freier Dienstvertrag) Projekt EUMeTrain
2009 Juli	Austro Control Wien (Praktikum) Projekt Hageldatenbank sowie Fallstudien
2008 (1 Monat)	Austro Control Linz (Praktikum) Verifikation des TAF-guidance der metalliance
2007 (6 Wochen)	Labor-Kunststofffirma AGRU
2006 (1 Monat)	Labor-Kunststofffirma AGRU

13 Eidesstattliche Erklärung

Ich erkläre hiermit an Eides statt, dass ich die vorliegende Arbeit selbständig und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht. Die Arbeit wurde bisher in gleicher oder ähnlicher Form keiner anderen Prüfungsbehörde vorgelegt und auch noch nicht veröffentlicht.

Wien, März 2012