Norwegian Meteorological Institute
met.no
met.no report

no. 6/2005
HIRLAM, impact study

Results from the Winter 2003/2004 HIRLAM Analysis Impact Study

Ole Vignes, Frank Thomas Tveter and Harald Schyberg
Norwegian Meteorological Institute
Oslo, Norway

Norwegian Meteorological Institute
met.no

report

Title	Date
Results from the Winter 2003/2004 HIRLAM Analysis Impact Study	26th January 2005
Section	**Report no.**
Section for Remote Sensing	no. 6
Author(s)	**Classification**
Ole Vignes, Frank Thomas Tveter and Harald Schyberg	● Free ○ Restricted
	ISSN 1503-8025
	e-ISSN 1503-8025
Client(s)	**Client's reference**
NA	NA

Abstract

This report presents results from 8 HIRLAM impact experiments for the 4 month period from November 2003 until February 2004. The purpose of the experiments was to study how the current operational HIRLAM 20km assimilation system could be improved. It turns out that initializing the HIRLAM reruns using the ECMWF analysis and assimilating QuikScat and ATOVS observations improved the verification of the resulting HIRLAM forecasts over the specified 4 month verification period.

Keywords

HIRLAM, impact study

Disiplinary signature	Responsible signature

Postal address P.O Box 43 Blindern N-0313 OSLO Norway

Office Niels Henrik Abels vei 40

Telephone +47 2296 3000

Telefax +47 2296 3050

e-mail: met.inst@met.no **internet:** met.no

Bank account 7694 05 00601

Swift code DNBANOKK

1 Introduction

The Norwegian Meteorological Institute (met.no) currently has three different HIRLAM models in daily production. Of these, only the largest and coarsest model at 0.2° (≈ 20 km) resolution has its own assimilation system for observations. This model has a 6-hourly assimilation cycle, and it produces 4 forecasts of length 60 hours per day. In addition, there are reruns of the 00Z and 12Z forecasts (before the 06Z and 18Z runs) in order to take advantage of more observations and more recent lateral boundaries. The lateral boundaries are frames produced by the ECMWF global model (the LBC project). The forecast model in current operational use is version 6.2.0 of the HIRLAM reference. All the experiments reported here used this same version.

The HIRLAM models are considered to be important tools for short range forecasting at the institute, and they supplement the global forecasts available twice daily from ECMWF. These models complement the global model predictions available at the Norwegian Meteorological Institute in two ways, firstly by providing higher resolution information, and secondly by earlier availability of forecasts, partly obtained by a shorter observation cutoff than the global models.

In general, the quality of a limited-area model in an operational setup is determined by

1. quality of initial state
2. quality of the forcing on the lateral boundaries
3. quality of lower boundary (surface) forcing
4. accuracy of the forecast model equations and algorithms.

In this study we are mainly concerned about the strategy to improve the first item above, the initial state. In limited-area models the impact of observations improving the initial state estimate is a function of the domain size, and is largest for large domains. The relative influence of the information input from the initial state becomes less, relative to the forcing from the lateral boundaries, as forecast time increases. Thus the impact of introducing new observations is larger in global models, and for limited area models it decreases with decreasing domain size as well as with increasing forecast range.

One of the goals of this study is to assess the impact of various observation systems to decide whether we can run with a shorter observation cutoff than today and still have a comparable quality of the initial state. Present cutoff-times for the analysis at met.no are 2h 15 min for 00Z the 12Z runs and 4h 30 min for the 06 and 18. Due to its large impact in the past, the designing factor for the analysis cutoff has been the arrival of radiosondes. Eerola (2003) did an investigation of arrival times of radiosondes at the Finnish Meteorological Institute, which probably is representative also for the arrival times at met.no. This showed that after 2 hrs, 88% of the first parts of the radiosonde messages had arrived, and 71% of the sondes had arrived in complete form. Thus the present cutoff-time at met.no of 2h 15 min for the main termins is necessary to ensure that a significant portion of the radiosondes are available.

Recently more satellite data have become available early. For instance, AMSU-A observations are available over the North-Atlantic within 30 minutes after observation through the EUMETSAT ATOVS retransmission service (EARS). It is of interest to see whether the situation where radiosondes are a necessary factor to obtain high quality of the initial state, has changed.

Several impact studies of various observation systems have been performed in limited-area models addressing the impacts of various observations, see for instance Amstrup and Mogensen

(2004) and several papers in Böttger et al (2004). It is also of interest to see if the impact of the various observation systems in the operational system of the Norwegian Meteorological Institute shows any significantly differences from that found in other limited-area systems. If such differences are found, it is of interest to locate the reasons for this.

We also study whether there are improvements by use of ECMWF analysis in reanalysis cycles or by increasing the analysis update frequency to allow better use of the observations. The present 3D-Var system uses the FGAT approximation (First Guess at Appropriate Time), and this approximation should becomes less severe with a 3 hours cycle instead of a 6 hours cycle as used currently.

To assess these topics, we assess how the HIRLAM 20km analysis and forecast quality would be affected by

- not assimilating QuikScat observations
- not assimilating ATOVS observations
- initializing reruns using an ECMWF analysis
- using 3-hourly assimilation cycles
- using newly generated background error structure functions
- removal of radiosondes in presence of remotely sensed data

In order to study each of these changes, a "reference" and a number of "experiments" were designed. The reference was originally expected to score fairly well in the verification. Each experiment had (only) one change compared to the reference. In addition to the changes listed above, an experiment with no observations at all was also designed. This was done in order to verify the importance of an assimilation system in a limited-area model with constant feeding of information through the lateral boundaries.

For historical reasons, HIRLAM experiments are usually given a three letter name. Table 1 gives an overview of the characteristics of the various HIRLAM experiments.

Name	ECMWF analysis	QuikScat	ATOVS	3-hourly cycle	New str.fun.	TEMP	Other conv.
REF		X	X	X		X	X
ECA	X	X	X	X		X	X
NQS			X	X		X	X
NAT		X		X		X	X
6HA		X	X			X	X
NSF		X	X	X	X	X	X
NTP		X	X	X			X
NOB				X			

Table 1: Overview of experiment characteristics

2 Results

The reference and the experiments were run on a verification period of 4 months, from 5th of November 2003 until 2nd of March 2004. For verification, the forecasts originating from the

midnight cycle are used. The forecasts are verified at the main synoptic hours (0Z, 6Z, 12Z, 18Z). Only EWGLAM observations are used in the verification. The RMS between the reference observations and the model forecasts is calculated. If the experiment has a significantly smaller RMS than the reference, the experiment is considered to be better than the reference.

For each experiment we present verification of surface parameters against SYNOP stations as a function of forecast range up to 48 hours. We also present vertical profiles of verification against EWGLAM radiosondes. These curves compare experiment and reference for +6, +24 and +48 hours. In the figures, "m-o" indicates "model" minus "reference observations".

Furthermore, we present time-series of daily verifications for the experiment period for the surface parameters, to assess whether the difference in verification from the reference comes from single episodes or has a more constant nature.

Finally, in order to study systematic geographic effects of the changes, the geographic mean difference between the experiment and the reference parameters is calculated. These plots are normalized to show the relative contribution of the mean difference to the RMS. For instance, if the mean difference contributes to 50% of the RMS, the normalized value is 0.5. If the normalized value is less than 0.5, the standard deviation dominates the RMS, and if it is greater than 0.5, the mean difference dominates the RMS. The appendix gives more details.

2.1 The Reference

The reference run (REF) in this experiment series is neither the same as the current HIRLAM reference nor the current operational *met.no* HIRLAM version. The reference used here has 3-hourly forecast cycles (same as the HIRLAM reference). But unlike the HIRLAM reference and the operational *met.no* version, ATOVS (AMSU-A) and QuikScat scatterometer observations are used in addition to the conventional observations used by all. So the reference here was designed to be close to an eventual future operational HIRLAM system for *met.no*. In the following, all other experiments are compared to this reference.

2.2 The ECA experiment

In the ECA experiment we have tried to see if we can extract information from the EC analysis, which are known to be of relatively high quality. To do this, it is desirable to retain high resolution information from the HIRLAM forecast, but also to ensure an initial state close to balance. An approach already implemented operationally at the Danish Meteorological Institute has been followed here.

Thus, the ECA experiment is the only experiment where the forecast following the assimilation is slightly different from the REF run. The reason for this is that the ECMWF analysis is blended with the HIRLAM first guess via an *incremental* digital filter, like this:

$$HL_{in} = HL_{fg} + DFI(EC_{an}) - DFI(HL_{fg}), \tag{1}$$

where HL_{fg} is the HIRLAM first guess (a +3h forecast), EC_{an} is the LBC ECMWF analysis (at 0.5° resolution), and HL_{in} is the resulting initialized analysis. In the other experiments, an ordinary digital filter is used, like this:

$$HL_{in} = DFI(HL_{an}),$$

where HL_{an} is the HIRLAM 20km analysis. The blending described above was performed twice daily, at the 00Z and 12Z runs. At 06Z and 18Z equation 1 applies, if EC_{an} is substituted by HL_{an}.

With hindsight the blending strategy should have been chosen differently, to more closely resemble what can be achieved operationally. It would have been better to do it at e.g. 06Z and 18Z, not at the same cycle that is later verified.

Figure 1 shows the RMS for MSLP, T2M and FF10 as a function of forecast length for REF and ECA. We observe that the ECA experiment has a significantly better RMS for all parameters. Figure 2 show the vertical distribution of the mean squared error for Z, T and Rh in the +6h, +24h and +48h forecasts. Again we observe that the ECA experiment has a better score than the reference. Figures 3 to 5 show the time series of the daily contribution to the integrated RMS for MSLP, T2m and FF10. We observe that the MSLP parameter in the ECA experiment has a better daily RMS than the reference for almost the entire verification period. The other parameters also show the same tendency. Figures 6 to 13 show the geographical distribution of the mean difference between ECA and REF for T2m, Z, Rh and the 10m wind. There are significant systematic differences between land and sea for the surface temperature, the ECA experiment is warmer over land and polar ice and colder over the sea. There are also large systematic differences in temperature and geopotential higher up in the atmosphere. We observe systematic differences in relative humidity in the atmosphere over Russia. The 10m wind vector difference is large over Greenland, and here the ECA wind speed is larger than the reference. We also observe larger winds over Russia, and there appears to be a vortex in the lower right corner of the grid (by the Red Sea, with high wind vector differences that are normal to the average wind direction).

The improved score by using EC analysis in this way, must be due to the EC analysis having higher quality than the HIRLAM analysis. This is not so surprising, since EC has a more mature assimilation system (in addition differences in the forecast model). For instance, the 4D-Var scheme probably makes better use of the observations than a 3D-Var scheme, the EC analysis uses more types of satellite observations than HIRLAM 3D-Var and can also get benefit from observations outside the LAM domain.

2.3 The NQS experiment

In this experiment, the QuikScat scatterometer observations were omitted, while the reference assimilates QuikScat following a method described in Tveter (Tveter 2002).

Figure 14 shows the RMS for MSLP, T2M and FF10 as a function of forecast length for REF and NQS. We observe that the QuikScat observations have a small positive effect on the forecasts. Figure 15 show the vertical distribution of the mean squared error for Z, T and Rh. The QuikScat observations have a neutral to positive effect on the geopotential. Figures 16 to 18 show the time series of the daily contribution to the integrated RMS for MSLP, T2m and FF10. The improvement in MSLP originates from a period from 11th of December 2003 until the 19th of February 2004. The same trend is detectable in T2m and FF10.

2.4 The NAT experiment

The operational HIRLAM runs assimilates AMSU-A observations over the ocean (Schyberg et al, 2003). The reference run contains observations from the NOAA-15 and 16 satellites received from several reception stations through the EARS service of EUMETSAT. In this experiment, all these ATOVS radiance observations were omitted.

Figure 19 shows the RMS for MSLP, T2M and FF10 as a function of forecast length for REF and NAT. We observe that the ATOVS observations have a small positive effect on the forecasts. Figure 20 show the vertical distribution of the mean squared error for Z, T and Rh.

The ATOVS observations have a neutral to positive effect on the +48h forecasts in the three parameters. Figures 21 to 23 show the time series of the daily contribution to the integrated RMS for MSLP, T2m and FF10. We observe that the positive impact is accumulated over the whole verification period.

2.5 The 6HA experiment

In this experiment, a 6-hourly assimilation cycle was used as opposed to the 3-hourly cycle used in the reference.

Figure 24 shows the RMS for MSLP, T2M and FF10 as a function of forecast length for REF and 6HA. Using a 3-hourly cycle gives a significantly worse score in MSLP compared to using a 6-hourly cycle. The 3-hourly cycle scores equal to or worse than the 6-hourly cycle in T2m and FF10. Figure 25 show the vertical distribution of the mean squared error for Z, T and Rh. The 6-hourly cycle scores significantly better than the 3-hourly cycle in temperature and geopotential. Figures 26 to 28 show the time series of the daily contribution to the integrated RMS for MSLP, T2m and FF10. The positive impact of moving back to a 6-hourly cycle is distributed evenly over the verification period, although FF10 seems to have a shorter period if negative impact around the last half of December. Figures 29 to 32 show the geographical distribution of the mean difference between 6HA and REF for T2m, Z and 10m wind. There seems to be a land/ocean effect in the geographic distribution of the T2m mean difference. There are also large differences in the geopotential. The largest 10m wind mean differences seem to be associated with the boundary.

It could seem surprising that the 6-hourly cycle actually has a better score than the 3-hourly cycle, when the latter has a potential to make better use of the non-synoptic observations. We have already seen in the above experiments that we get benefits from QuikScat and ATOVS observations, and there are also non-synoptic aircraft data in the system. The most probable reason for this is that the relative weighting of first guess and observations used in the assimilation has not been adjusted, even if the first guess is now a 3-hour forecast rather than a 6-hour forecast. The assumed background error covariance matrix used here, is thus probably less optimal for 3-hourly cycling than 6-hourly. For seeing improved results of the 3-hourly cycle, a retuning of the background error statistics would probably be necessary.

2.6 The NSF experiment

In this experiment, newly generated statistics for the background error model were applied. These structure functions were generated from RCR[1] data by the NMC method (84 pairs of differences between +24h and +48h forecasts spread over all seasons). The RCR data is at 0.2° resolution and 40 levels, which is the same as in our runs. This allowed the analysis to be done with high resolution increments (similar to forecast resolution). In the reference run (and all the other experiments), the Hirlam reference statistics generated from SMHI data at 0.4° horizontal resolution and 31 vertical levels were used. In these runs, the reference statistics were interpolated to 40 levels, and analysis increments were at half the forecast resolution (0.4°), i.e., the resolution of the reference SMHI data. All runs have used the so-called analytical balance error model, statistical balance has not been tried in this series of experiments.

Figure 33 shows the RMS for MSLP, T2M and FF10 as a function of forecast length for REF and NSF. The introduction of the new structure functions has a negative effect on the surface

[1]Regular Cycles with the (Hirlam) Reference, run at FMI.

parameters. Figure 34 show the vertical distribution of the mean squared error for Z, T and Rh. The new structure functions also have a negative effect on these three parameters. Figures 35 to 37 show the time series of the daily contribution to the integrated RMS for MSLP, T2m and FF10. The contribution to the negative effects on MSLP is variable throughout the verification period, and it is dominated by relatively few events. The negative contributions to the FF10 RMS are more evenly distributed. Figures 38 to 41 show the geographical distribution of the mean difference between NSF and REF for T2m, Z and 10m wind. There appears to be some coastal phenomenon that creates a T2m mean difference between the NSF experiment and the reference. There is also an anomaly in the lower right corner of the grid, which is visible in several of the plots.

Based on these results, it does not seem beneficial to replace the "old" structure functions used in the present reference.

2.7 The NTP experiment

In this experiment, no radiosondes (TEMP) were assimilated.

Figure 42 shows the RMS for MSLP, T2M and FF10 as a function of forecast length for REF and NTP. Removing the radiosonde observations from the assimilation has a clear negative effect on the resulting forecasts. Figure 43 shows the vertical distribution of the mean squared error for Z, T and Rh. Radiosonde observations are important for the quality of these three parameters also. Figures 44 to 46 show the time series of the daily contribution to the integrated RMS for MSLP, T2m and FF10. Negative contributions are accumulated during the entire verification period. Figures 47 to 54 show the geographical distribution of the mean difference between NTP and REF for T2m, T, Z, Rh and 10m wind. There is some sort of coastal phenomenon in the T2m mean difference, and we also observe an anomaly in the lower right corner of the grid. There are clear systematic mean differences in temperature and humidity, and there is a strong anomaly in the wind field in the lower right corner of the grid.

The degradation by not using radiosondes is quite significant, and much larger than the impact of ATOVS or QuikScat data. This shows that radiosondes are still - in the presence of new satellite data - an essential ingredient in the forecast quality in the HIRLAM system run here.

2.8 The NOB experiment

In this experiment, no conventional (TEMP and SYNOP), ATOVS, QuikScat or any other observations were assimilated. Thus, the only external source of information forcing the simulation to the truth comes from the lateral boundary data. As mentioned above, the impact of these data increases with decreasing domain. This experiment throws light on how much the initial state really contributes to the forecast quality with the domain size used here.

Figure 55 shows the RMS for MSLP, T2M and FF10 as a function of forecast length for REF and NOB. Removing all observations has a dramatic negative effect on the verification. Figure 56 show the vertical distribution of the mean squared error for Z, T and Rh. Observations are clearly important for vertical parameters also. Note the negative mean difference in the high level temperatures. Figures 57 to 59 show the time series of the daily contribution to the integrated RMS for MSLP, T2m and FF10. The NOB experiment scores worse than the reference on every single day in the verification period. Figures 60 to 67 show the geographical distribution of the mean difference between NOB and REF for T2m, T, Z, Rh and 10m wind. There appears to be some coastal phenomenon at work in the T2m mean difference, and we

observe an anomaly in the lower right corner of the grid. The upper atmosphere in the NOB experiment is cold in the northern Atlantic. We observe systematically dry air over Russia. There are lower 10m winds at the boundaries, and higher winds in central parts of the model area. The anomaly in the 10m wind in the lower right corner is also visible in this experiment.

Excluding all observations gives a surprisingly large negative effect on the forecasts. One would expect that updated fields that propagate inward from the boundaries would rapidly replace the air masses in the model area, and ensure a minimum of updated information in the model area. However, the absence of observations in the model area seems to allow the creation of a self sustainable blocking condition that either repels the boundary air masses or somehow degrades their information so that the blocking condition persists and the resulting forecasts are poor.

3 Summary and conclusions

We have demonstrated that improving the initial state by data assimilation is essential to the forecast quality in the met.no operational setup. The NOB experiment showed a surprisingly large negative impact by dropping data assimilation altogether. This shows that the continuous feeding of information through the lateral boundaries can not bring the model toward a reasonable state, even for the longest forecast ranges studied here. The model domain must probably be significantly smaller than now for data assimilation not to be of large importance.

The impact experiments with ATOVS and scatterometer data show that assimilating each of these types of satellite data gives a slight positive impact.

Even in the presence of these satellite data, radiosondes is still a main contributor to the forecast quality. The gain in quality from satellite data with new fast delivery systems such as EARS, can not outweigh the loss from receiving fewer radiosondes. This means that a shortening of the analysis cutoff-time will be detrimental to the forecast quality, and we should probably keep an analysis cutoff time of the same magnitude as today for the 00 and 12Z runs (2h 15 min).

With the present tuning of the assimilation system, it turned out to be detrimental to replace the 6-hourly cycles with 3-hourly cycles. A retuning is probably necessary to take benefit from more frequent analysis.

In the ECA experiment we have shown that there is additional information content in the ECMWF analysis fields, not found in the HIRLAM 3D-Var analysis, which can be exploited in reanalysis cycles with the digital filter approach demonstrated here.

References

Amstrup, Bjarne and Kristian S. Mogensen, 2004: Observing system experiments with the DMI HIRLAM 3D-Var data assimilation system in a winter and summer period in 2002. DMI Technical report, available from Danish Meteorological Institute.

Böttger, Horst, Paul Menzel and Jean Pailleux (ed.), 2004: Proceedings of the third WMO workshop on the impact of various observing systems on numerical weather prediction, Alpbach, Austria 9-12 March 2004. WMO publication WMO/TD No 1228.

Eerola Kalle, 2003: Statistics of the arrival times of conventional observations for HIRLAM at FMI. In: HIRLAM Newsletter no 44, November 2003. Available from HIRLAM-5 Project c/o Per Undén, SMHI, S-60176 Norrköping, Sweden.

Schyberg, Harald, Tomas Landelius, Sigurdur Thorsteinsson, Frank Thomas Tveter, Ole Vignes and others, 2003: Assimilation of ATOVS data in the HIRLAM 3D-Var system. HIRLAM Technical Report No. 60, April 2003. Available from HIRLAM-5 Project c/o Per Undén, SMHI, S-60176 Norrköping, Sweden.

Tveter, Frank, 2002: Assimilation of QuikScat in HIRLAM 3D-var. Met.no research note no 84. Available from The Norwegian Meteorological Institute.

Appendix: Methodological comments

MSE verification

The basis for our Mean Squared Error (MSE) verification is the penalty function $\mathrm{MSE}[t] = E[(\mathbf{X}[t] - \mathbf{X}_t[t])^2]$ where $\mathbf{X}[t]$ is a $+t$ hour forecast for the state of the atmosphere and $\mathbf{X}_t[t]$ is the true state of the atmosphere at the forecast time. However, the true state of the atmosphere is unknown. We use independent reference observations to estimate

$$\mathrm{RMS}[t] \quad = \quad \sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_{r,i} - \mathbf{H}_{r,i}\mathbf{X}[t])^2} \tag{2}$$

where m is the number of independent reference observations, $\mathbf{H}_{r,i}$ is the forward operator for reference observation i, $\mathbf{X}[t]$ is a $+t$ hour forecast for the state of the atmosphere which is valid for the observation time. We assume further that the reference observations contain the same amount of information about each model state variable. It is then reasonable to expect that experiments with a small RMS also will have a small MSE.

Only forecasts that start at $00Z$ hours are used in the calculation of the RMS, and only reference observations at $00Z$, $06Z$, $12Z$ and $18Z$ are used in the verification. Only one parameter can be verified at a time (say the MSLP). The configuration that scores lowest in the MSE verification (using the same independent information), is considered to be better. In the figures, "m-o" indicates "model" minus "reference observations". Also, the vertical error figures show the +6, +24 and +48.

The atmospheric governing equations have chaotic properties that become visible if one applies a neutral change to the forecasting system, for instance by assimilating observations that do not improve nor degrade the forecast quality. In this case there will be changes from day to day, sometimes the original system scores better and sometimes it scores worse compared to the modified system. The reason for this is that the governing equations cause insignificant perturbations to grow rapidly, so that two equally probable states of the atmosphere may evolve into two very different weather situations, where one will score better than the other. We expect that our two systems will have comparable scores if one averages over a long enough period (we have then implicitly assumed that the "tails" of the probability distribution for the random perturbations are small). Much of the motivation for choosing a rather long verification period (4 months) was to increase the chances of having significant verification results.

The significance of any MSE comparison (using the null-hypothesis that two systems have equal quality) is difficult to determine since we do not know how much independent information we are using in our verification statistics. We may raise the question if the MSE verification always converges (as long as we verify over a long enough time-period). This is an interesting theoretical question, but we will not develop this discussion any further here.

We may wish to acquire an impression of how the MSE verification score over the 4 months accumulate. A fruitful approach is then to look at the contribution to the "area under the daily $\mathrm{RMS}[t]$ curve",

$$\phi = \sqrt{\sum_{t=6}^{48}(\mathrm{RMS}[t])^2}.$$

If we plot ϕ as a function of the day in the verification period for both the reference and our experiment, we can identify how any difference in final verification score was accumulated. If

there are just a few events with a large difference we may argue that the difference in verification score is less significant since if, say, the verification had stopped short of one of these events, the MSE verification score would have been very different.

Note that for the first couple of days in the period, there may only be short forecasts available, and the daily RMS will then tend to be lower. The opposite may be true for the last couple of days in the period, where only long forecasts may be available, and the daily RMS will tend to be higher.

RMS comparison

It may be interesting to examine the geographical distribution of the difference between an experiment and the reference for a particular parameter. Let us define

$$\mathrm{RMS}_{ij}^2 = \mu_{ij}^2 + \mathrm{Std}_{ij}^2$$

where $\mu_{ij} = \mathrm{E}[e_{ij} - r_{ij}]$, $\mathrm{Std}_{ij} = \sqrt{\mathrm{E}[(e_{ij} - r_{ij} - \mu_{ij})^2]}$ and e_{ij} is the experiment parameter value at grid location ij and r_{ij} is the reference parameter value at the same grid location. Note that $\mathrm{RMS}_{ij}^2 = \mathrm{E}[(e_{ij} - r_{ij})^2]$. The function

$$\gamma = \frac{\mu_{ij}|\mu_{ij}|}{\mathrm{RMS}_{ij}^2} \tag{3}$$

will give $\sim \pm 1$ in areas where the main contribution to the RMS is a mean difference (μ_{ij}) while it is close to zero in areas where the RMS is not dominated by a mean difference. We note that $\gamma \sim 0.5$ in areas where the mean difference, μ_{ij}, gives a contribution to RMS_{ij} that is comparable to the contribution from the standard deviation, Std_{ij}. When we analyse the wind vector difference, we use the definition

$$|\gamma| = \frac{\boldsymbol{\mu}_{ij}\boldsymbol{\mu}_{ij}}{\mathrm{RMS}_{ij}^2} \tag{4}$$

and the sign of γ is positive when the average wind speed of the experiment is larger than the reference (negative otherwise).

List of Figures

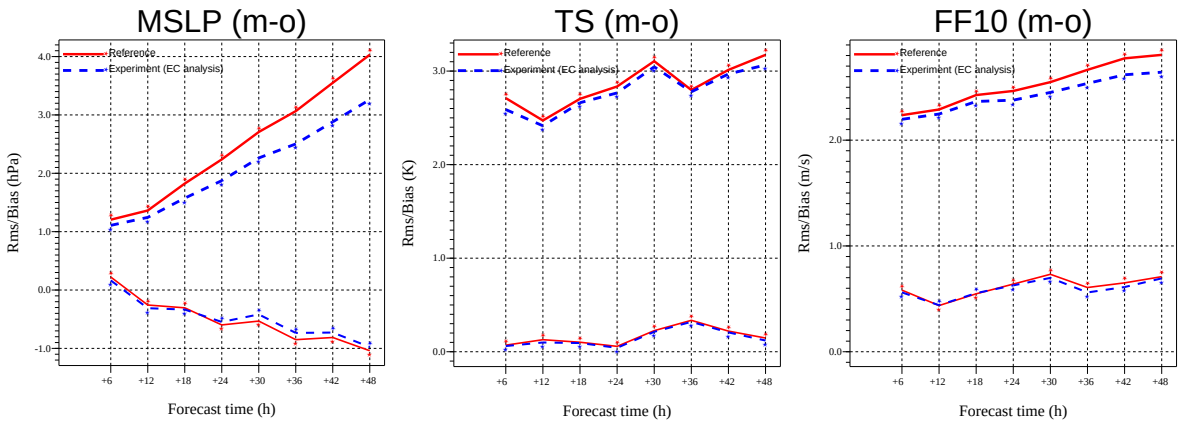

Figure 1: MSLP, T2m and FF10 error as a function of forecast length, in the ECA experiment.

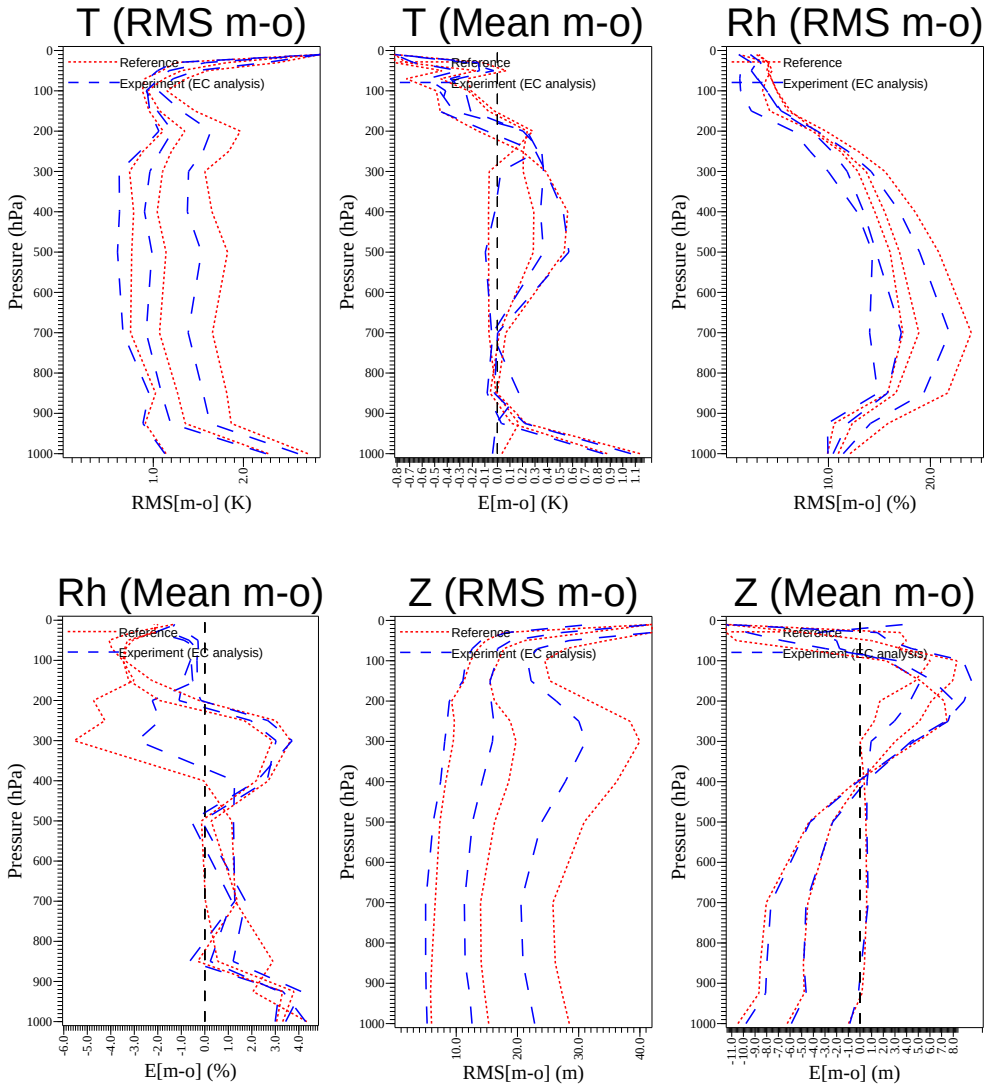

Figure 2: Vertical distribution of the temperature, relative humidity and geopotential error in the ECA experiment.

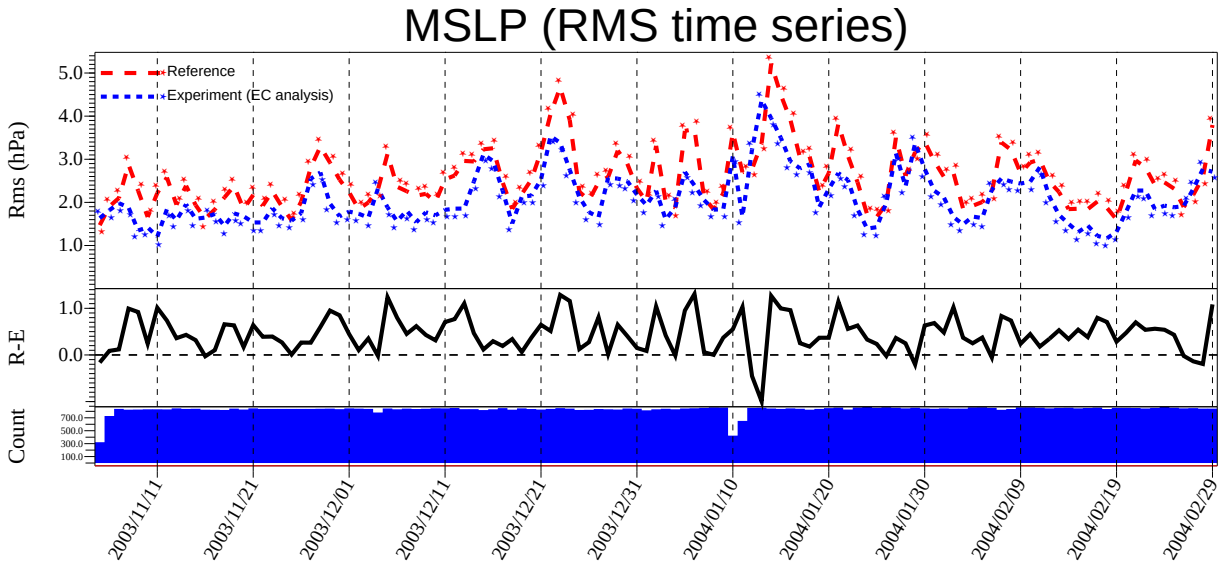

Figure 3: Time series of the daily contribution to the MSLP error in the ECA experiment.

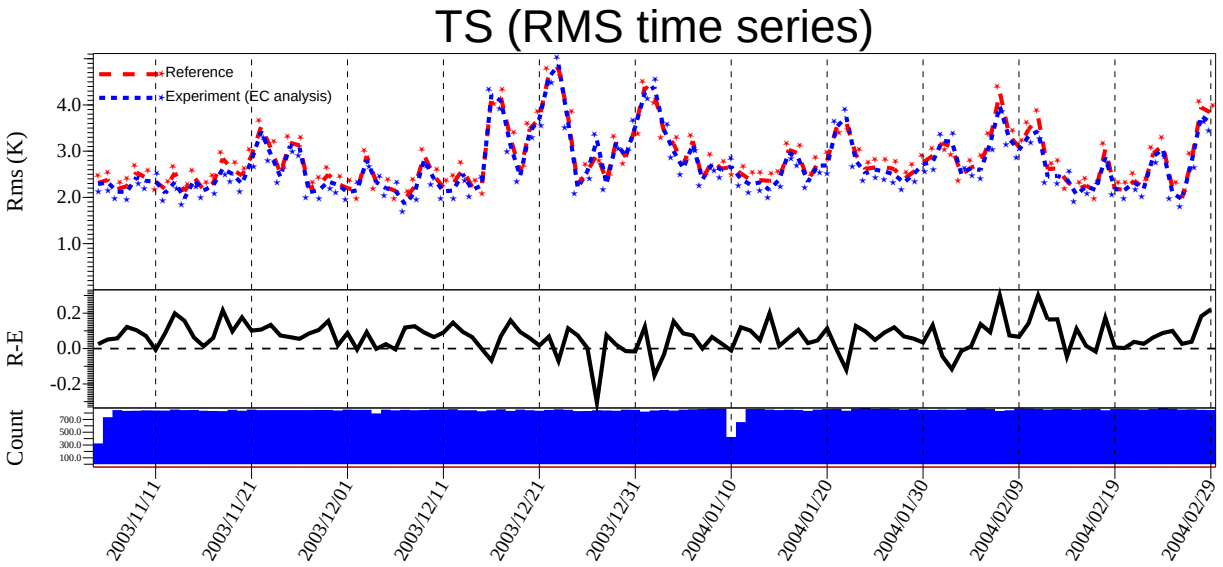

Figure 4: Time series of the daily contribution to the T2m error in the ECA experiment.

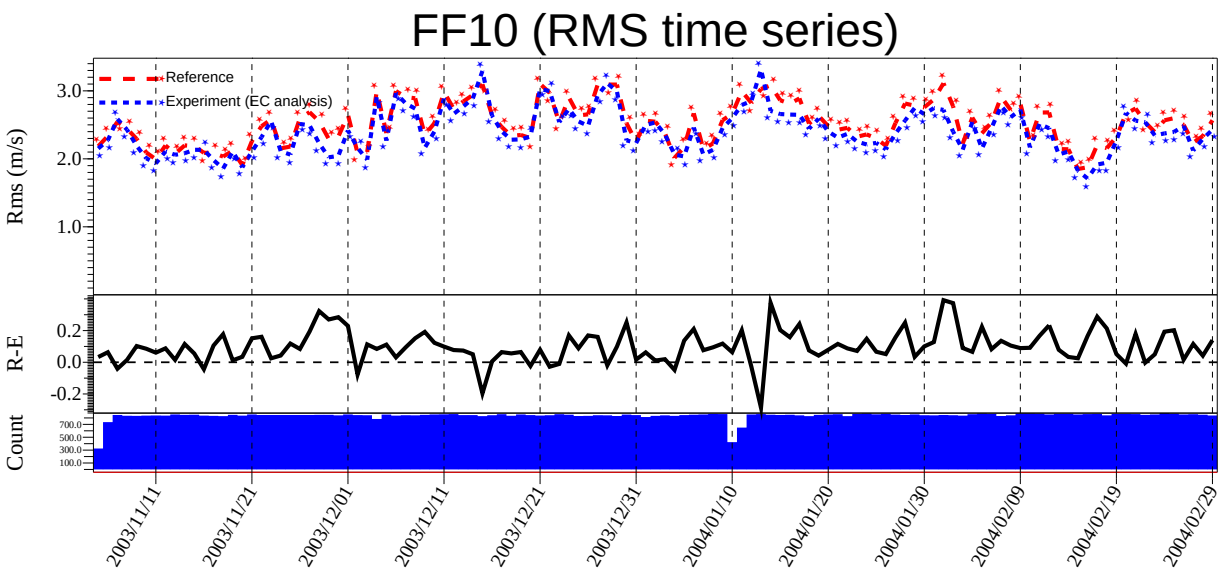

Figure 5: Time series of the daily contribution to the FF10 error in the ECA experiment.

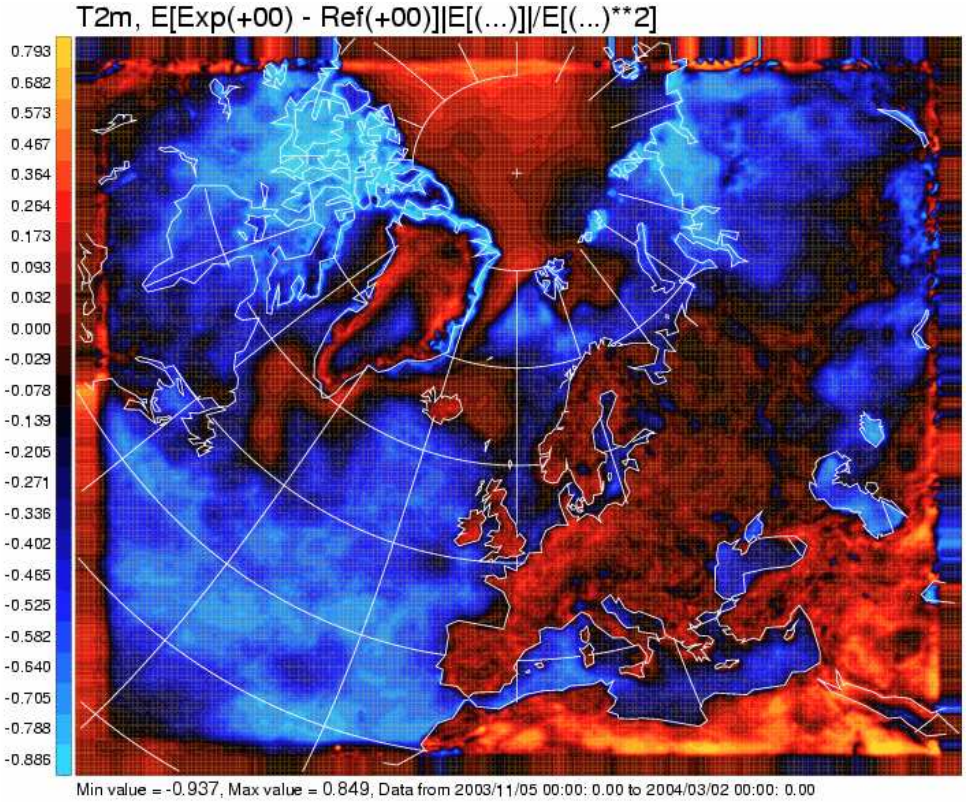

Figure 6: Geographical distribution of the T2m +0h mean difference between ECA and REF.

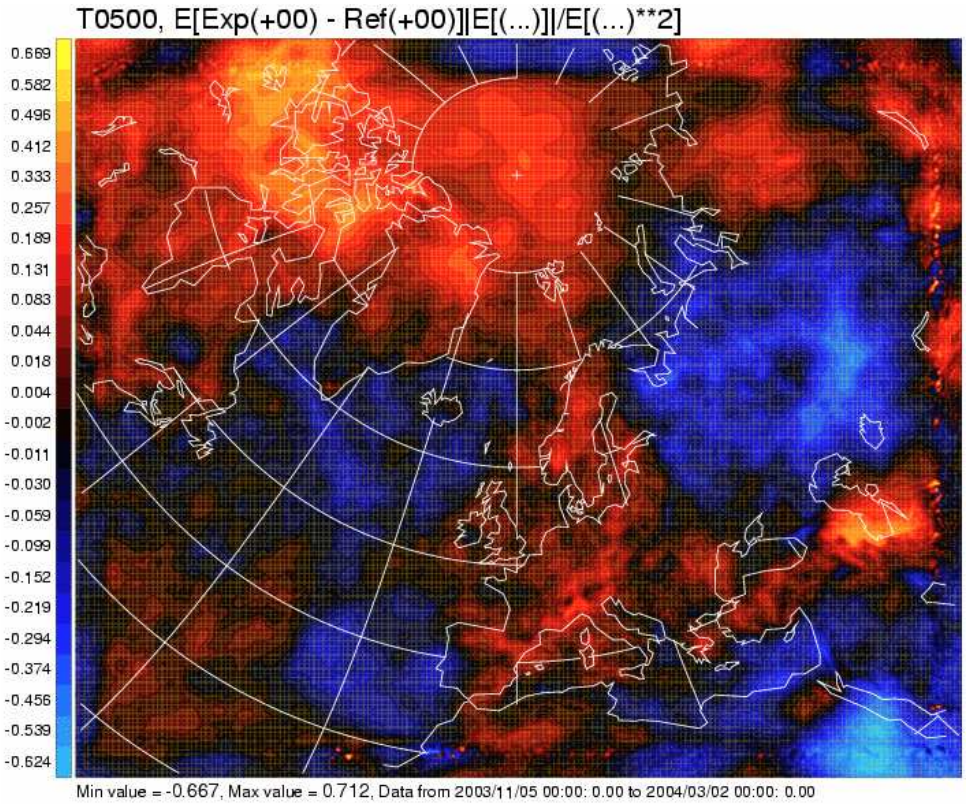

Figure 7: Geographical distribution of the T500 +0h mean difference between ECA and REF.

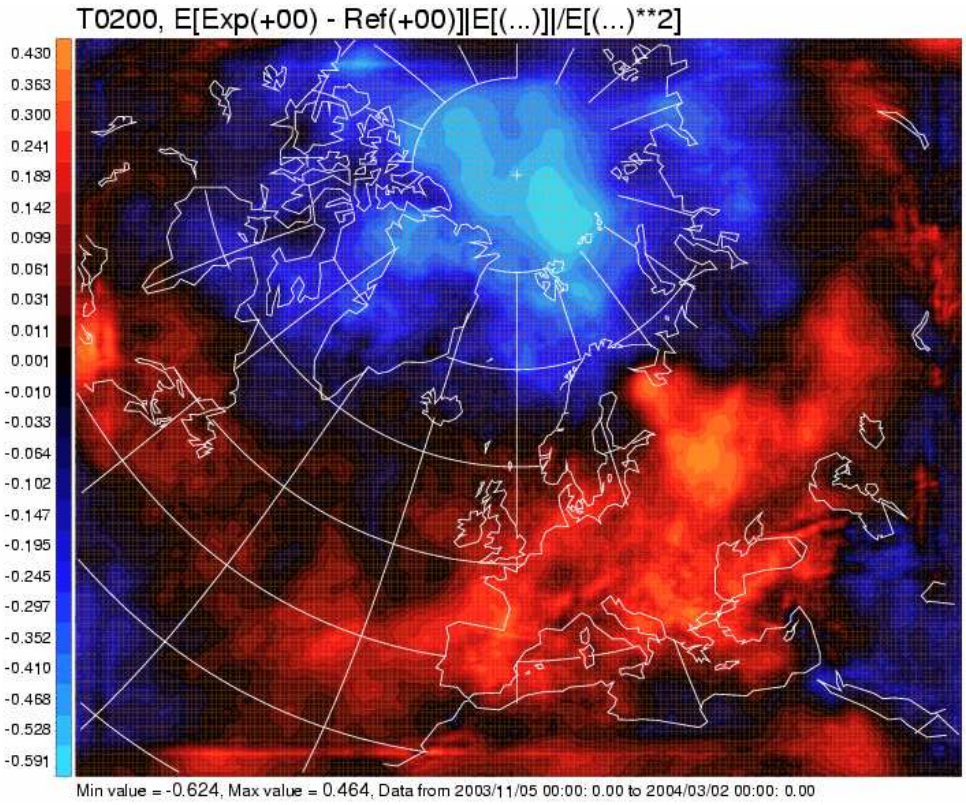

Figure 8: Geographical distribution of the T200 +0h mean difference between ECA and REF.

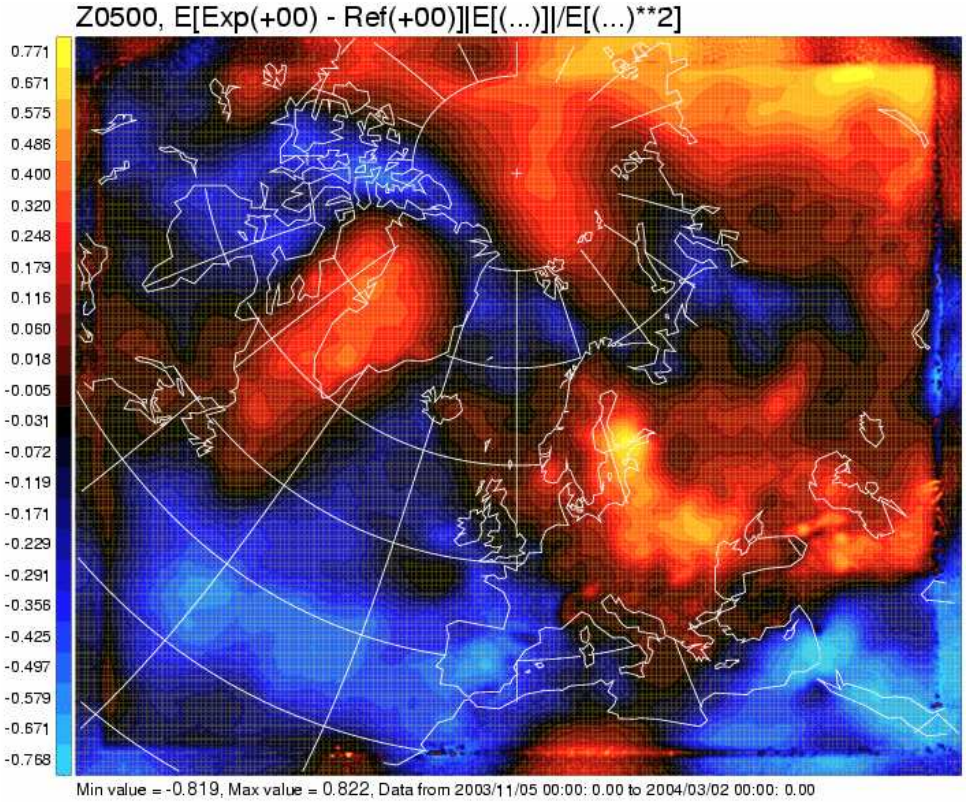

Figure 9: Geographical distribution of the Z500 +0h mean difference between ECA and REF.

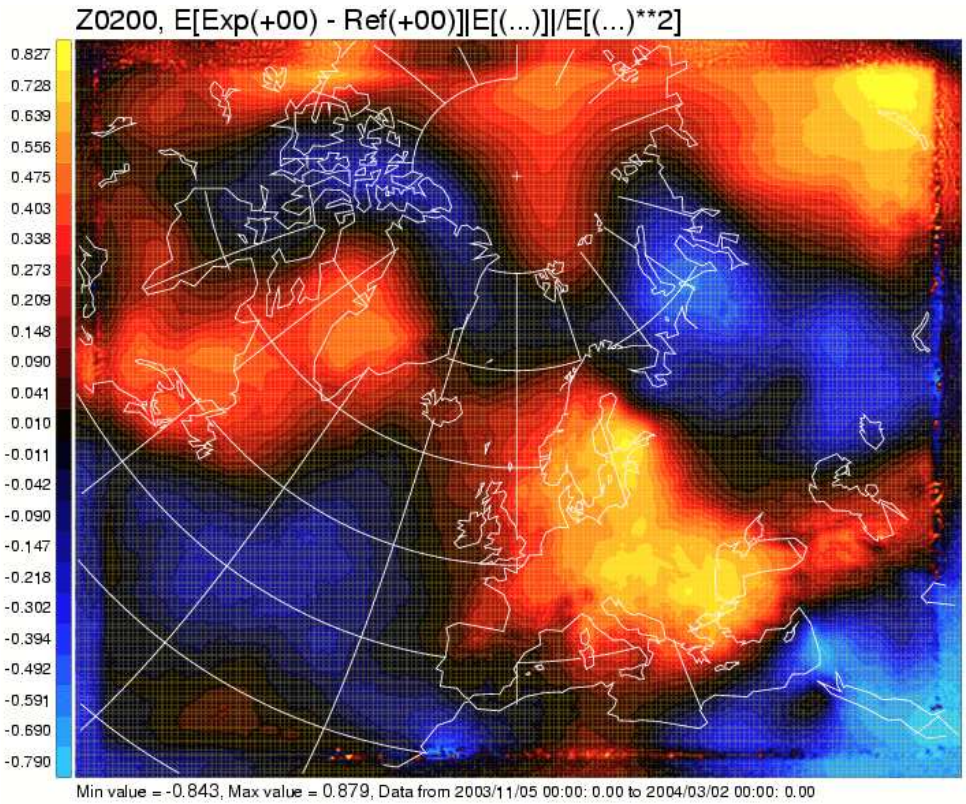

Figure 10: Geographical distribution of the Z200 +0h mean difference between ECA and REF.

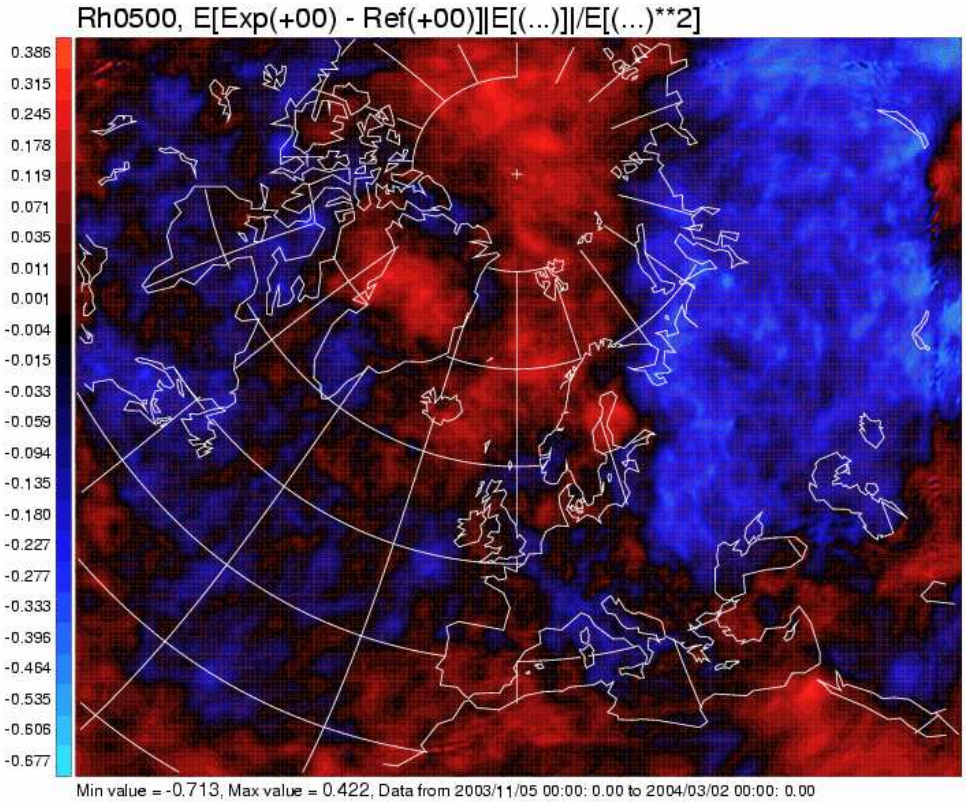

Figure 11: Geographical distribution of the Rh0500 +0h mean difference between ECA and REF.

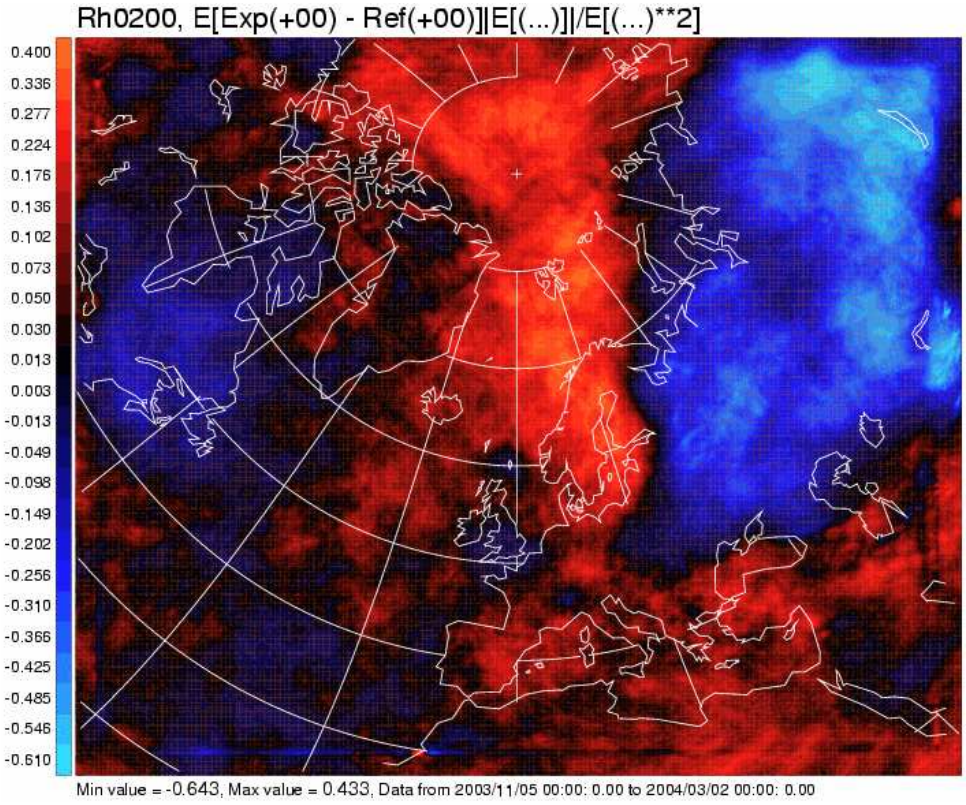

Figure 12: Geographical distribution of the Rh200 +0h mean difference between ECA and REF.

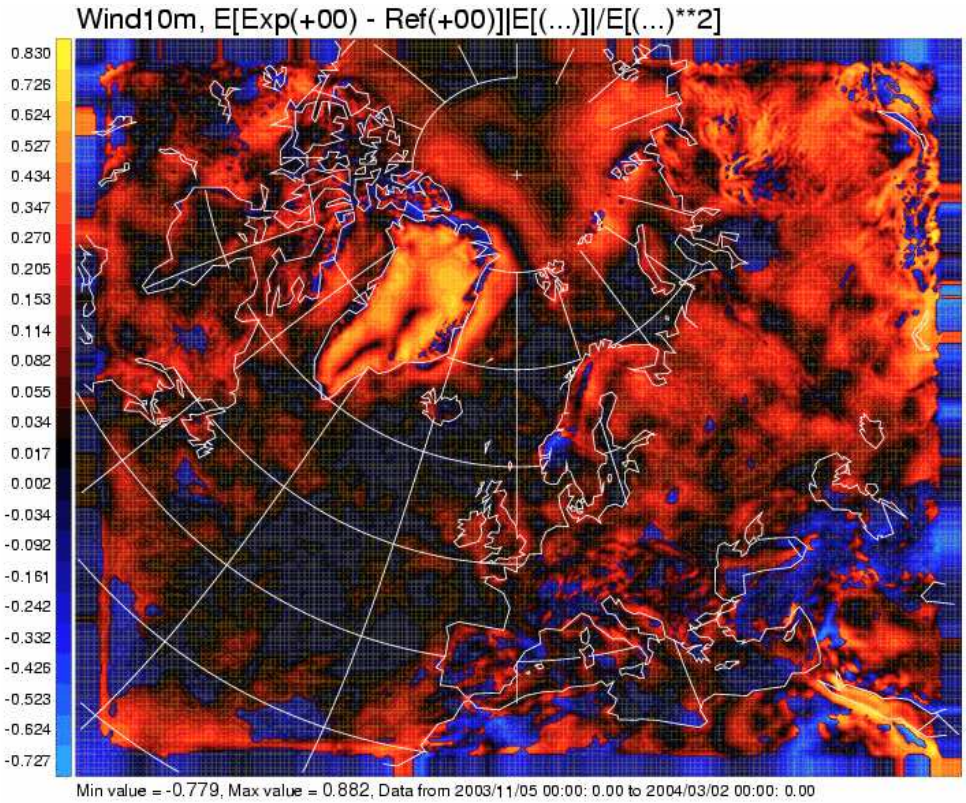

Figure 13: Geographical distribution of the wind speed +0h mean difference between ECA and REF.

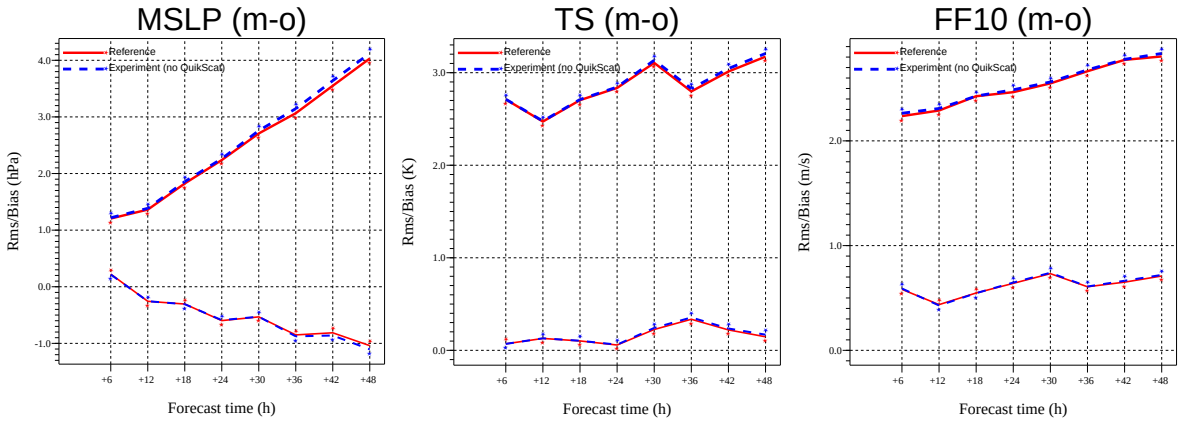

Figure 14: MSLP, T2m and FF10 error as a function of forecast length, in the NQS experiment.

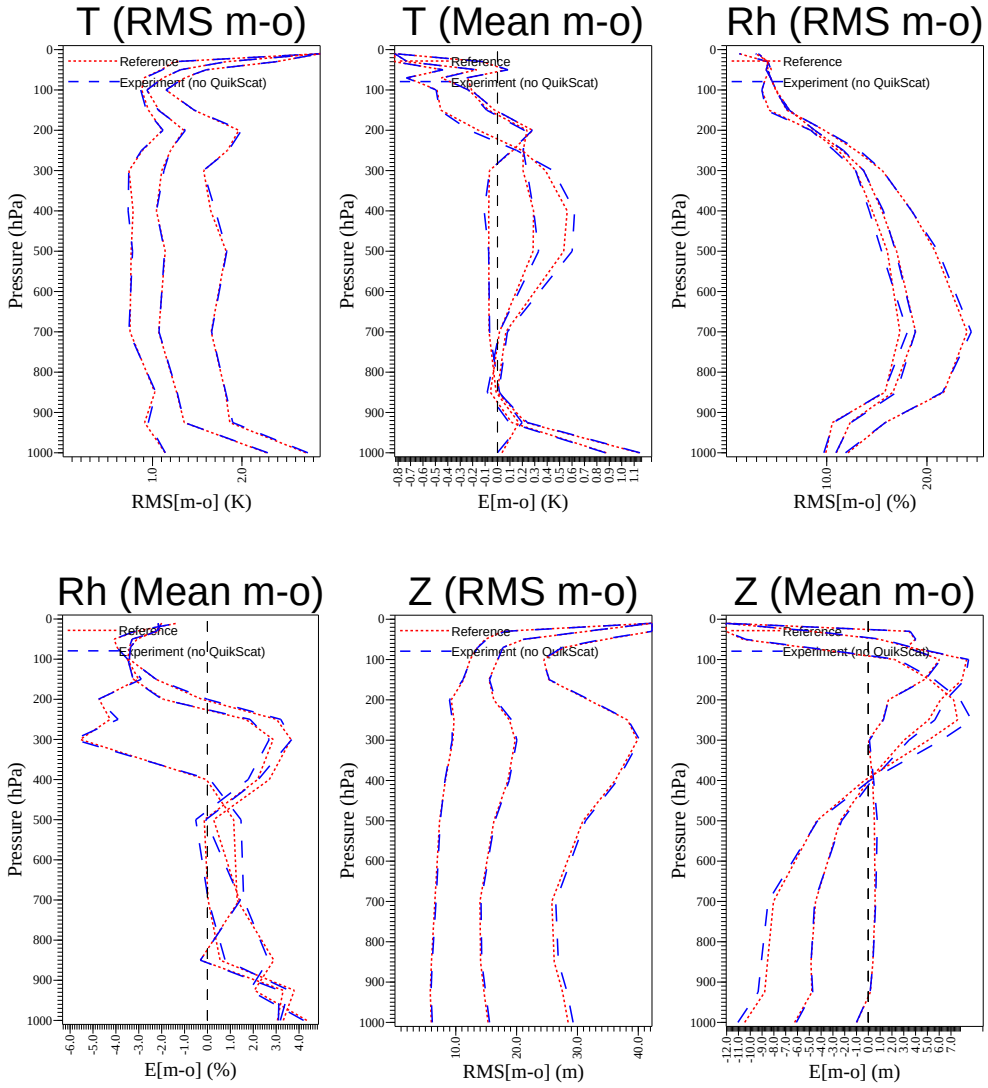

Figure 15: Vertical distribution of the temperature, relative humidity and geopotential error.

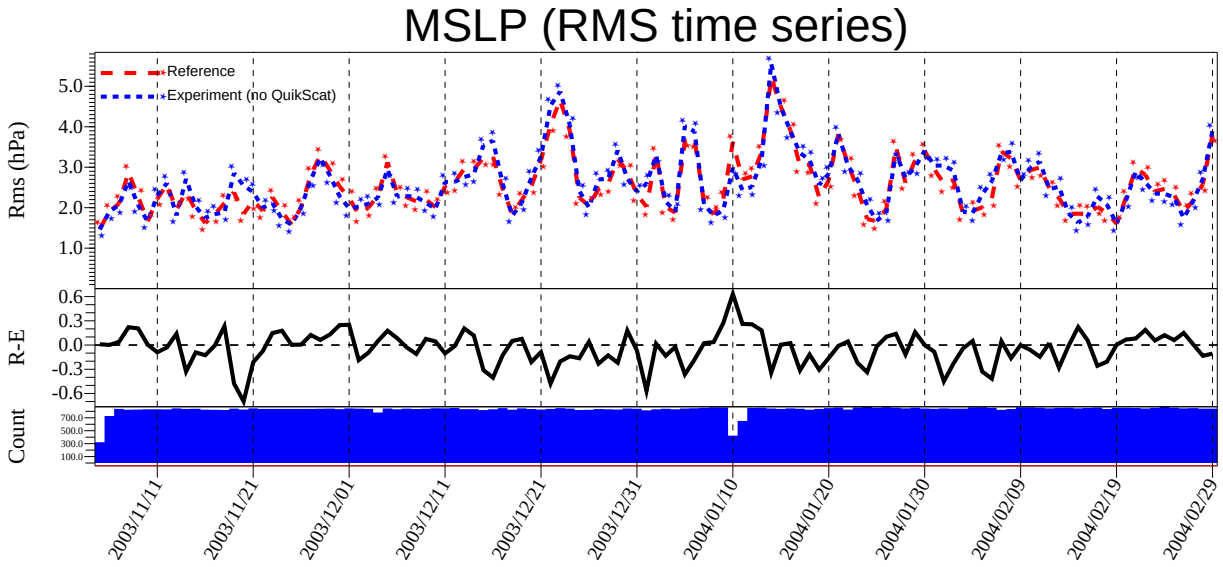

Figure 16: Time series of the daily contribution to the MSLP error in the NQS experiment.

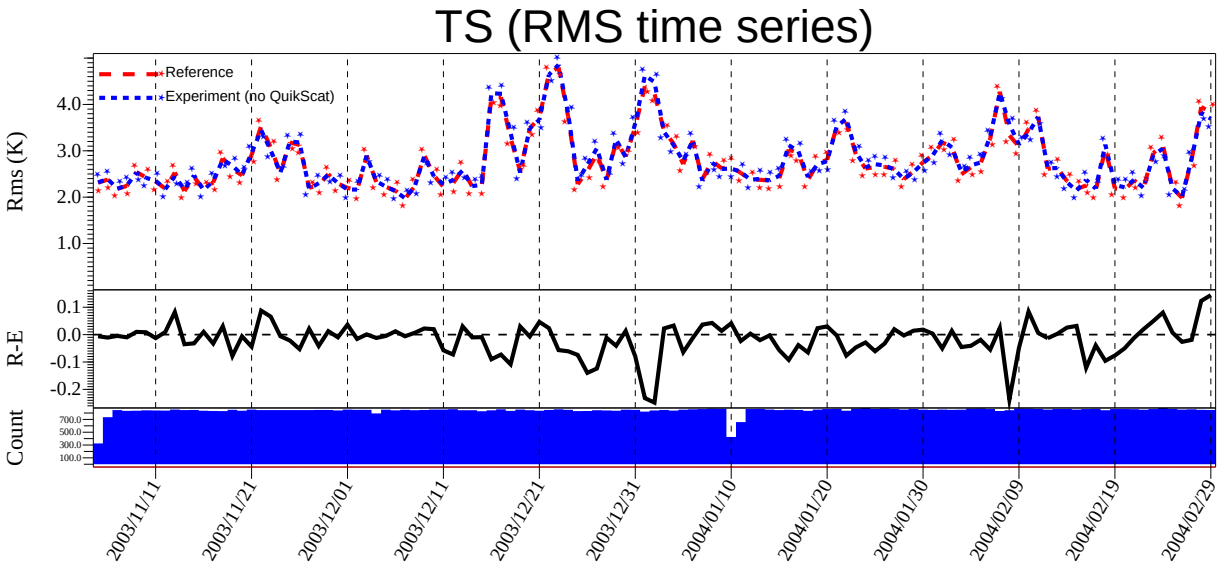

Figure 17: Time series of the daily contribution to the T2m error in the NQS experiment.

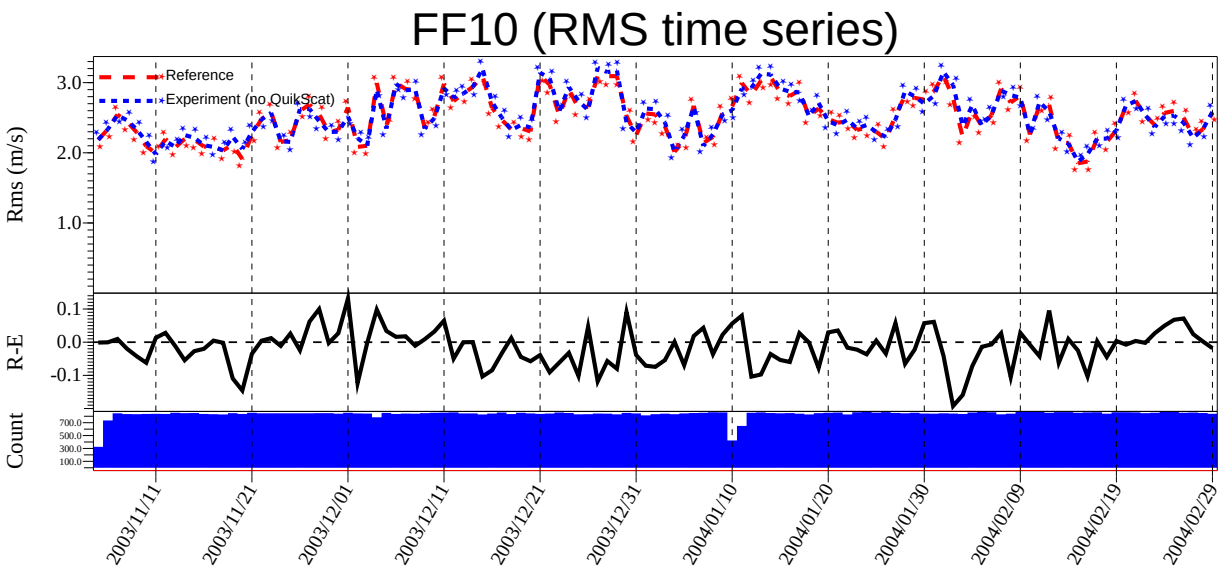

Figure 18: Time series of the daily contribution to the FF10 error in the NQS experiment.

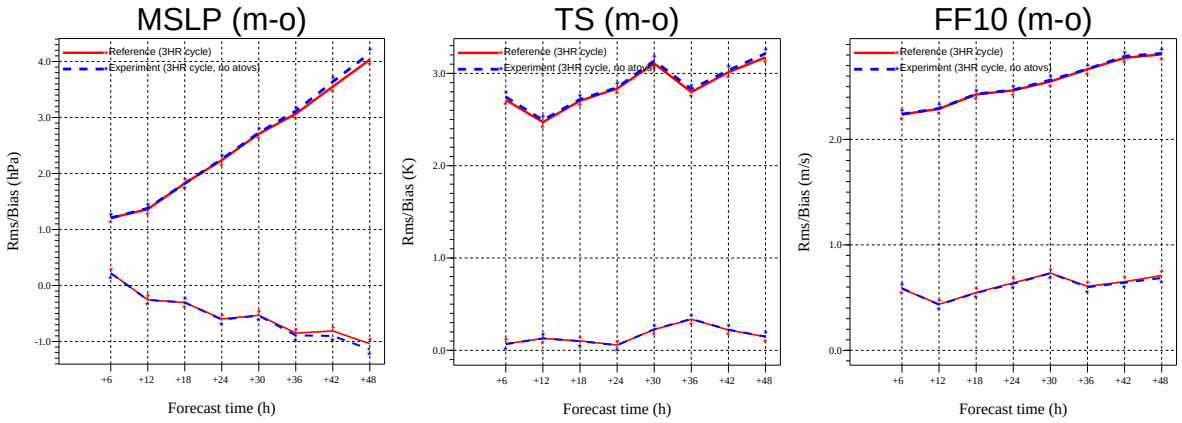

Figure 19: MSLP, T2m and FF10 error as a function of forecast length, in the NAT experiment.

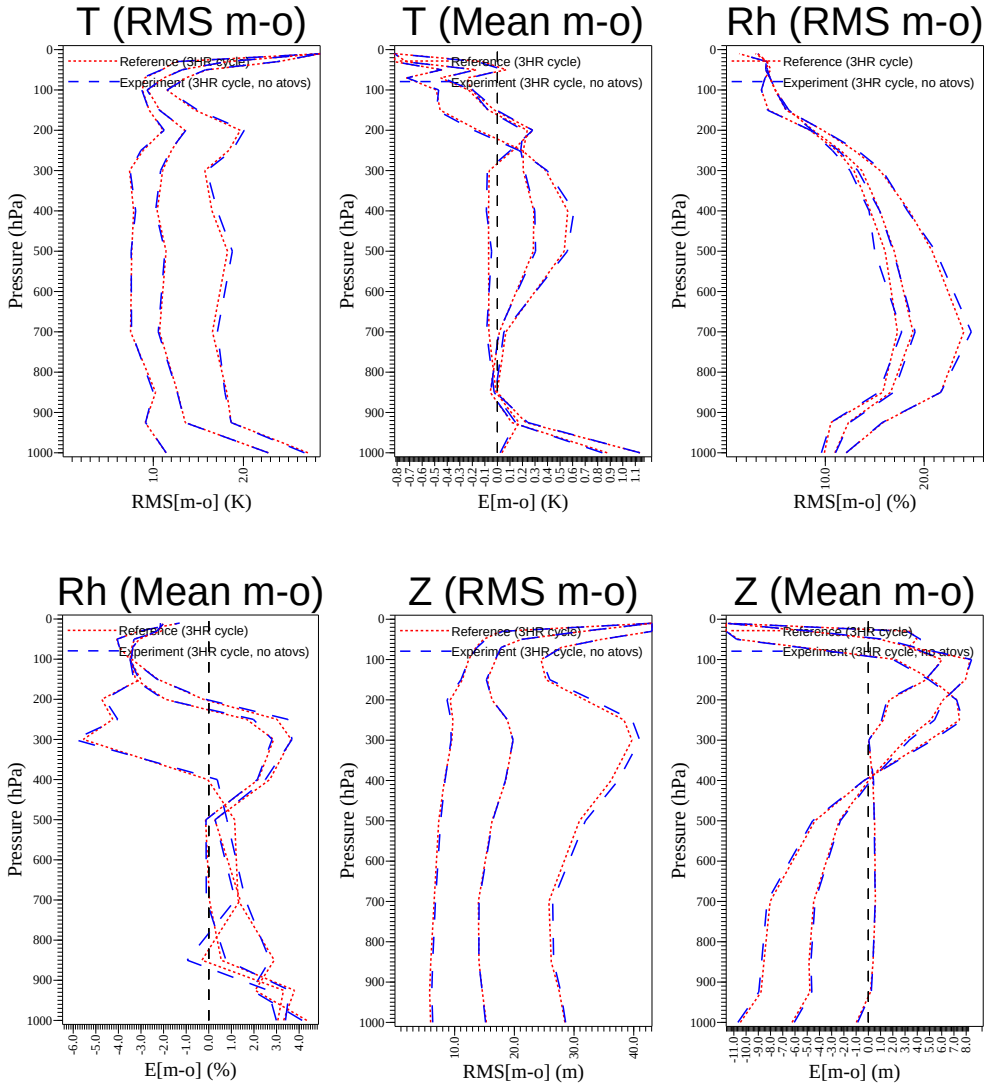

Figure 20: Vertical distribution of the temperature, relative humidity and geopotential error.

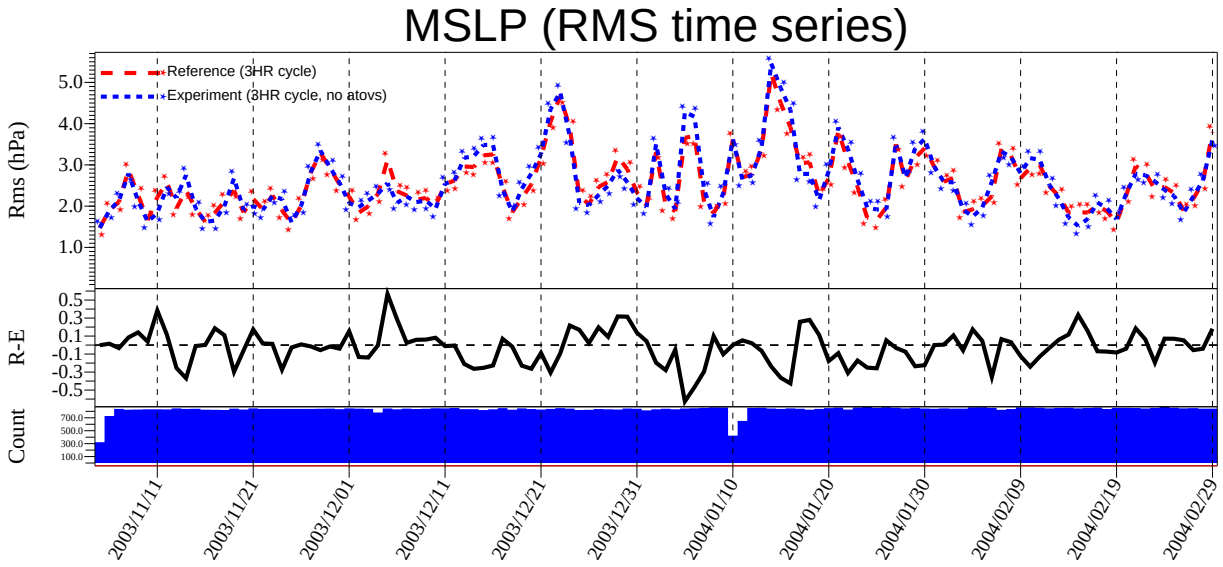

Figure 21: Time series of the daily contribution to the MSLP error in the NAT experiment.

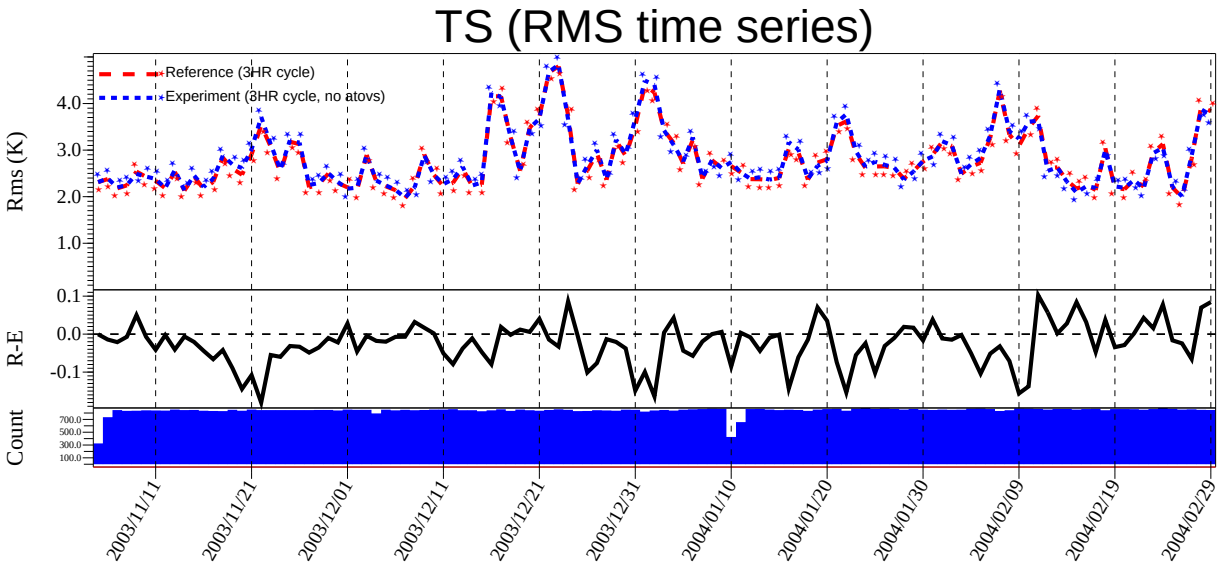

Figure 22: Time series of the daily contribution to the T2m error in the NAT experiment.

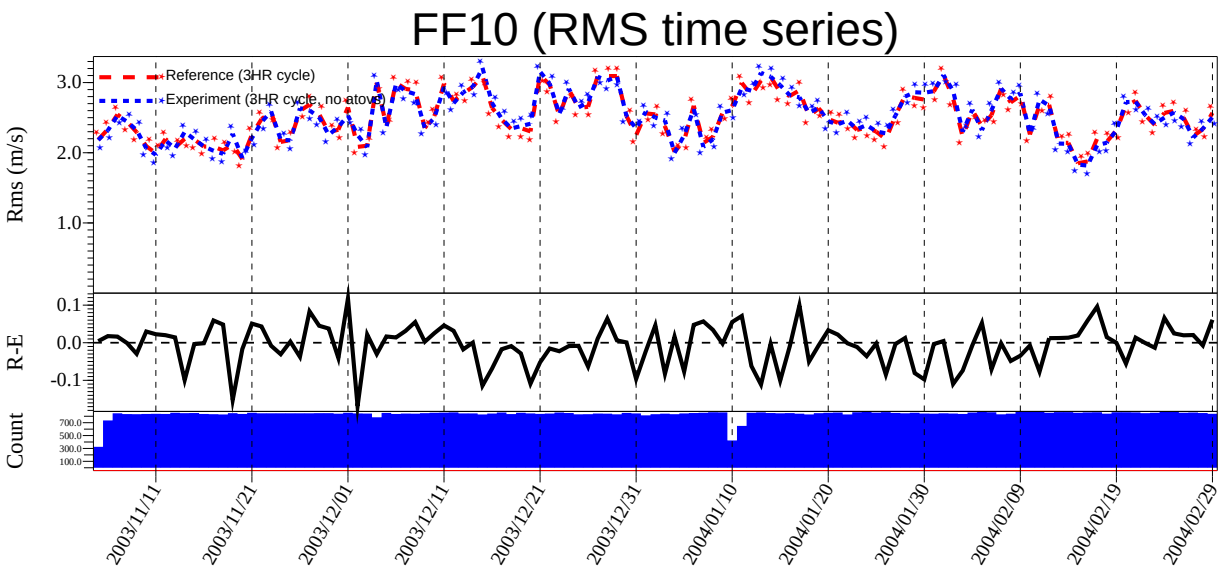

Figure 23: Time series of the daily contribution to the FF10 error in the NAT experiment.

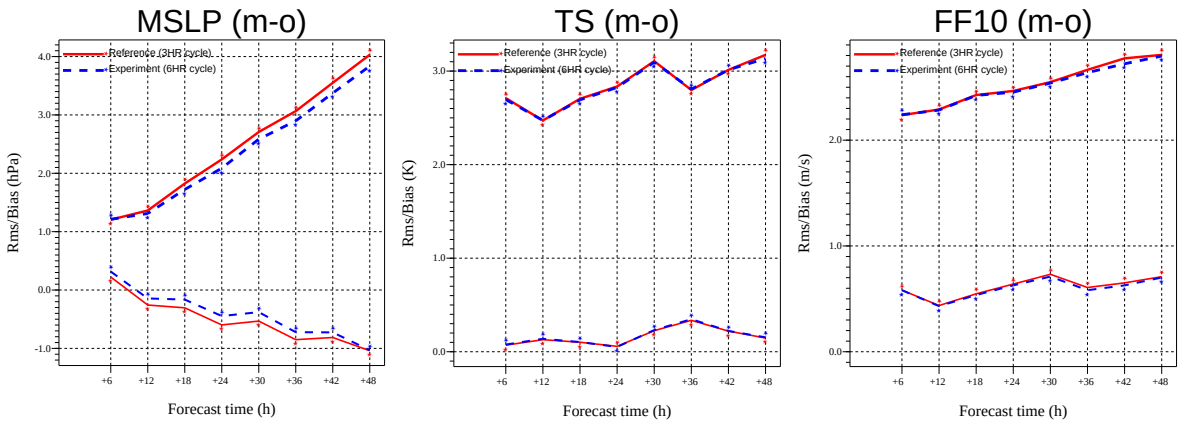

Figure 24: MSLP, T2m and FF10 error as a function of forecast length, in the 6HA experiment.

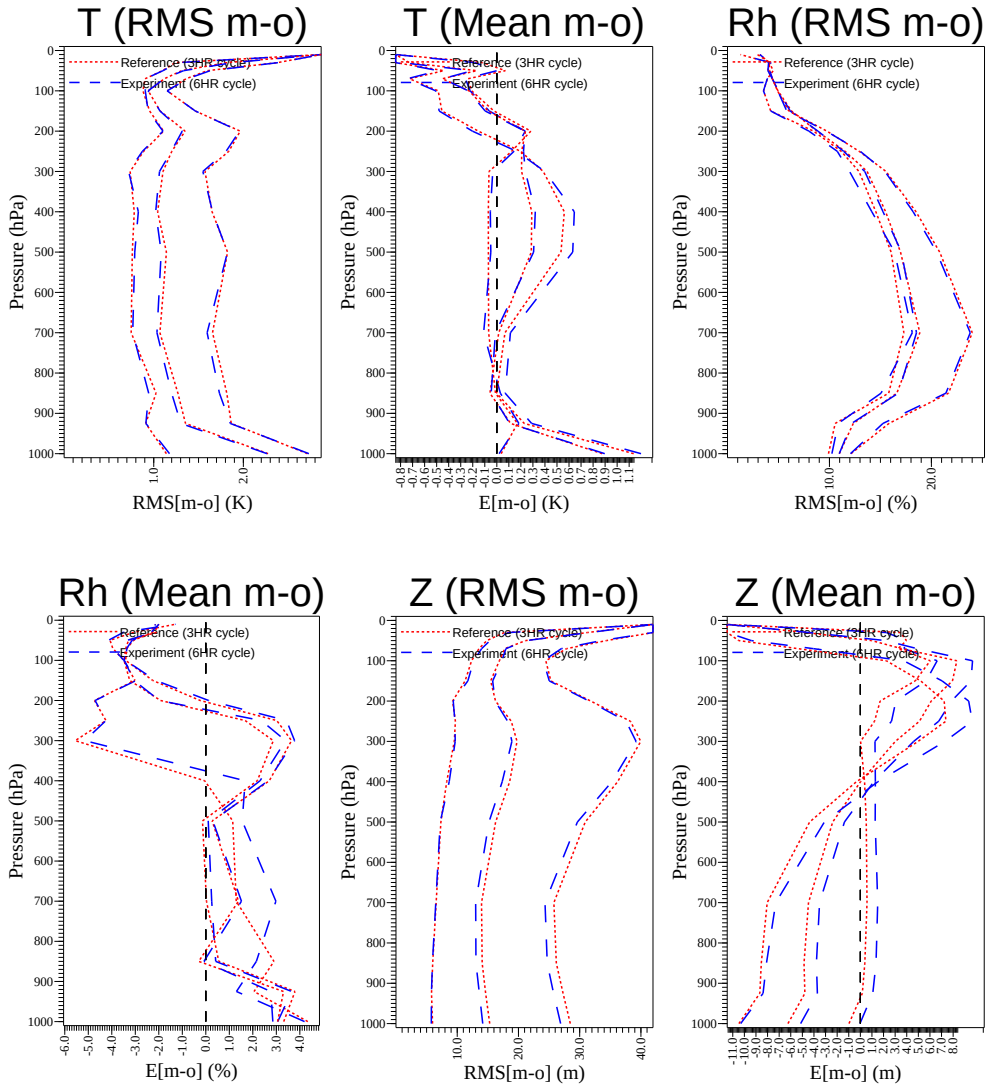

Figure 25: Vertical distribution of the temperature, relative humidity and geopotential error in the 6HA experiment.

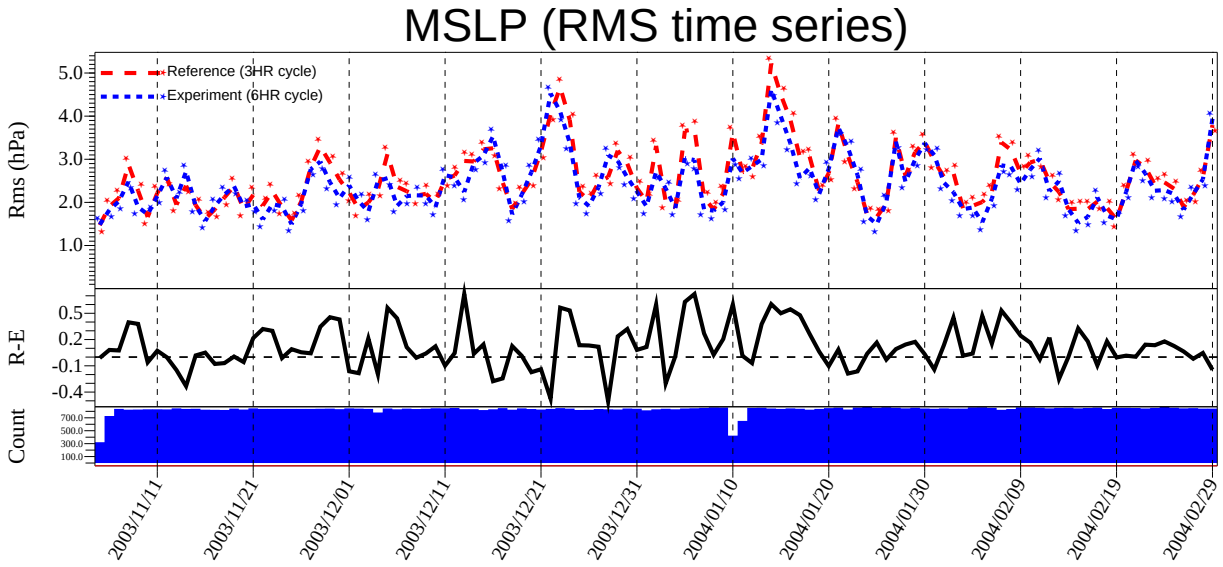

Figure 26: Time series of the daily contribution to the MSLP error in the 6HA experiment.

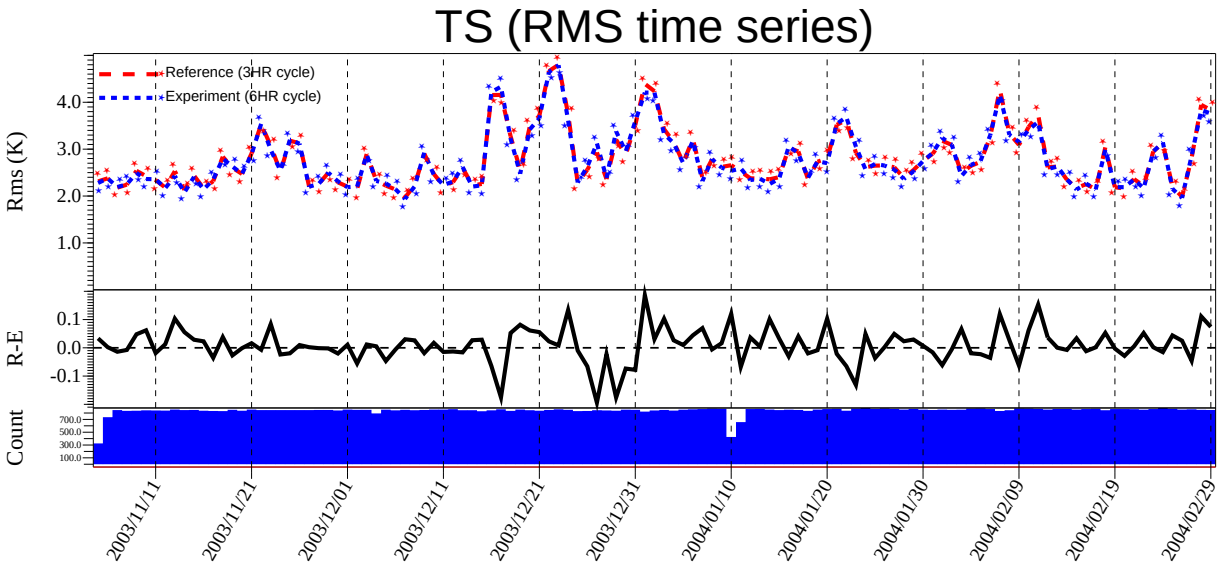

Figure 27: Time series of the daily contribution to the T2m error in the 6HA experiment.

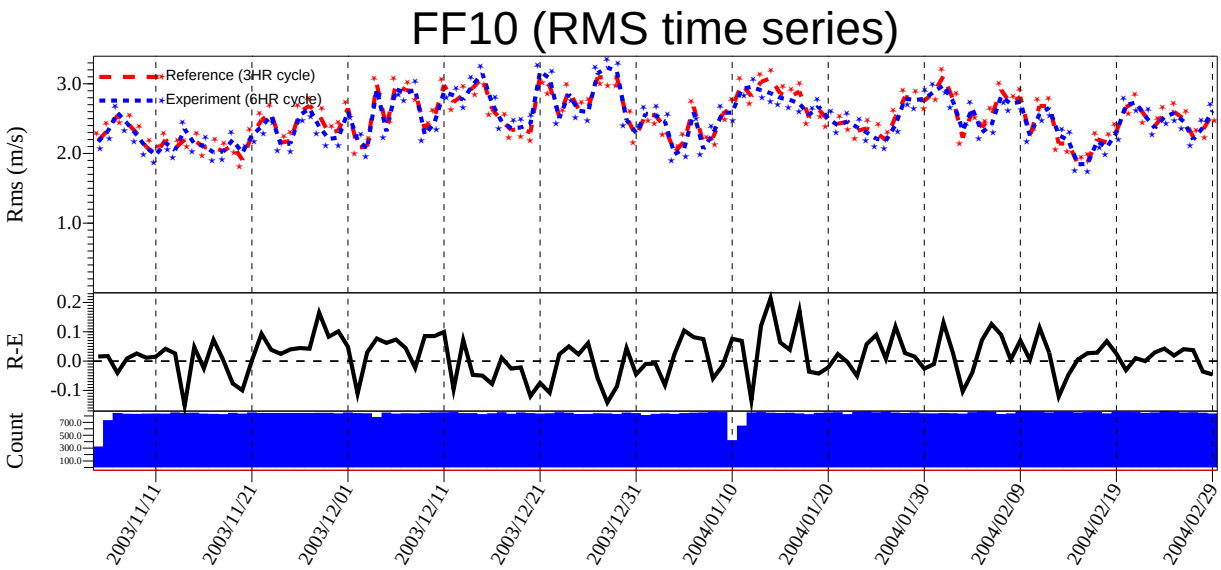

Figure 28: Time series of the daily contribution to the FF10 error in the 6HA experiment.

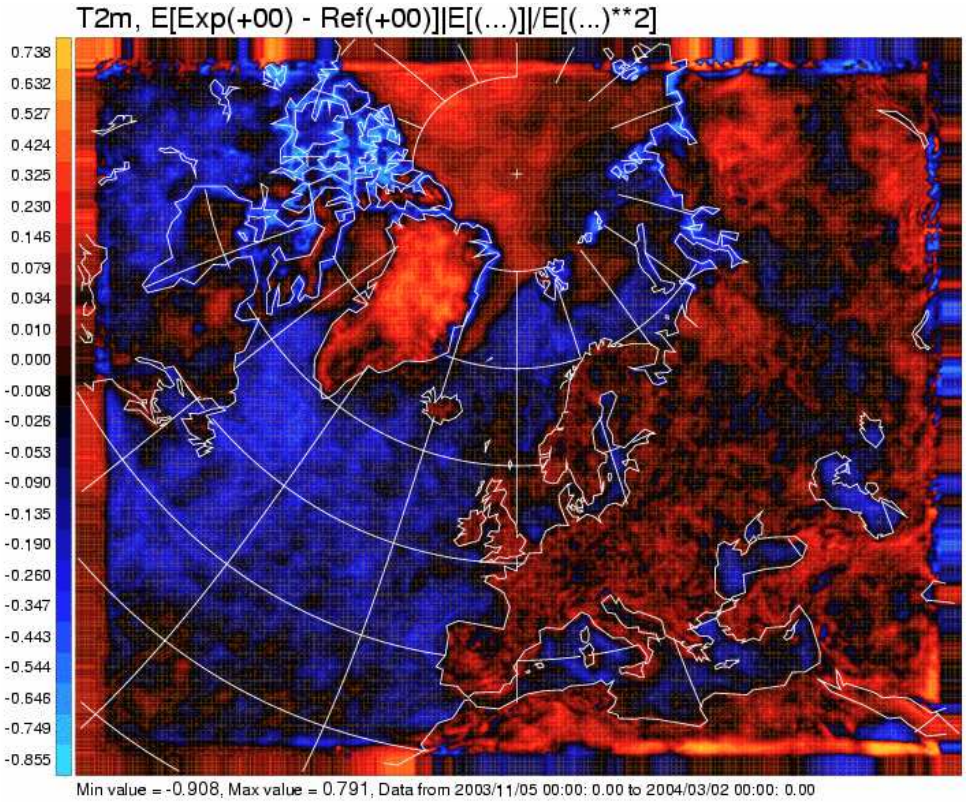

Figure 29: Geographical distribution of the T2m +0h mean difference between 6HA and REF.

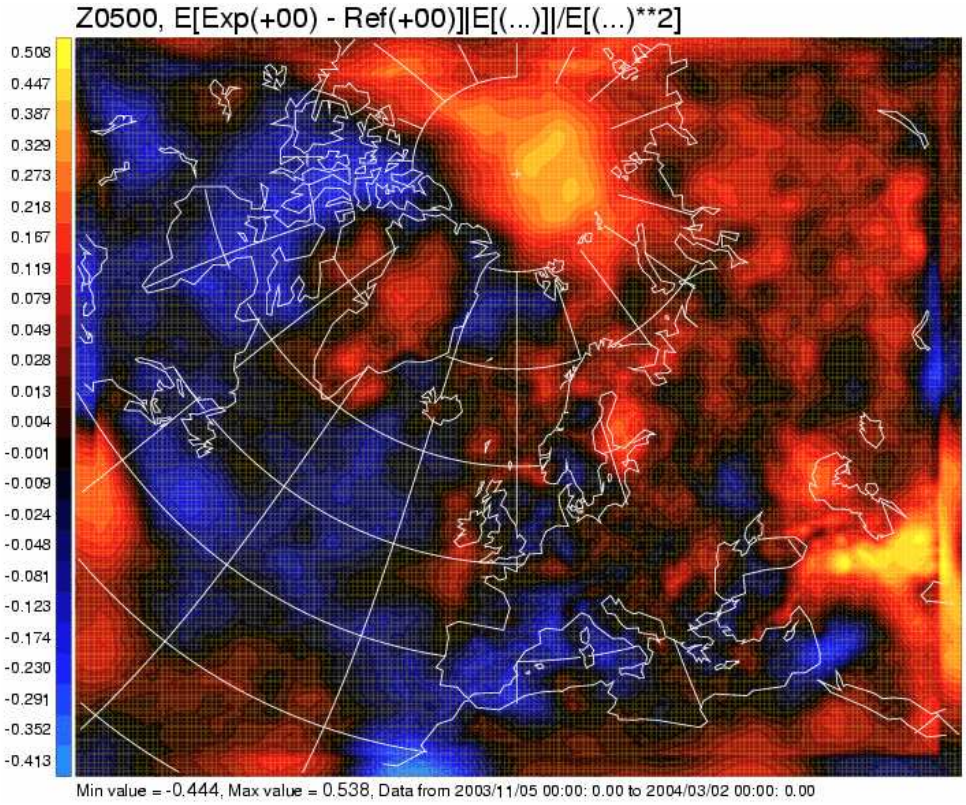

Figure 30: Geographical distribution of the Z500 +0h mean difference between 6HA and REF.

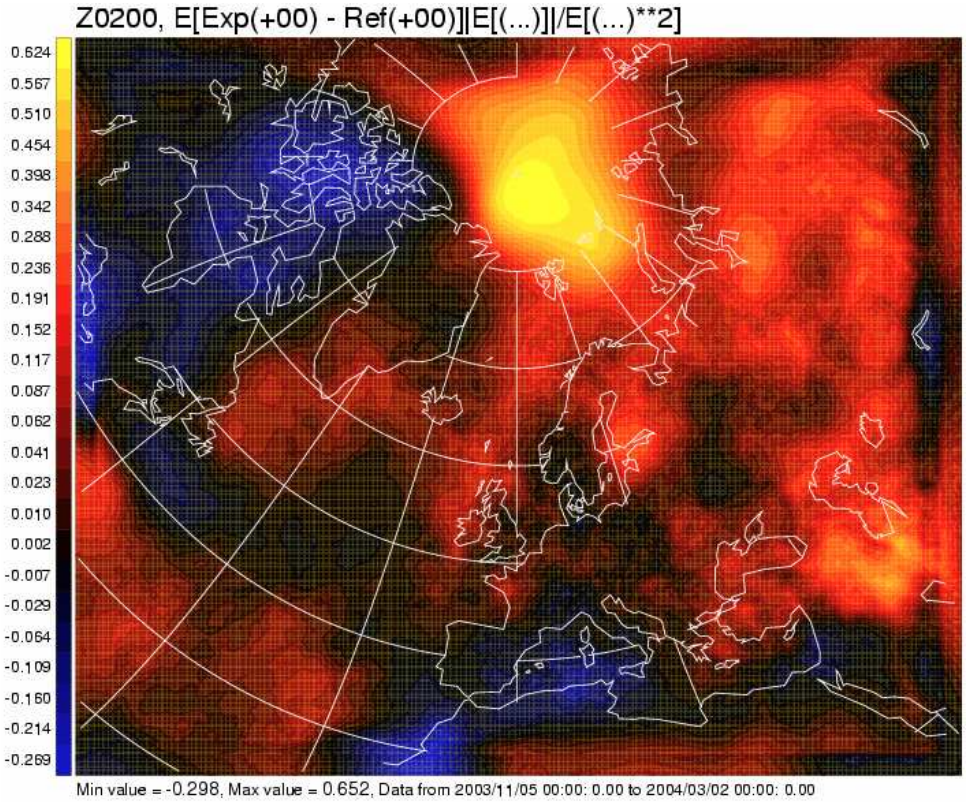

Figure 31: Geographical distribution of the Z200 +0h mean difference between 6HA and REF.

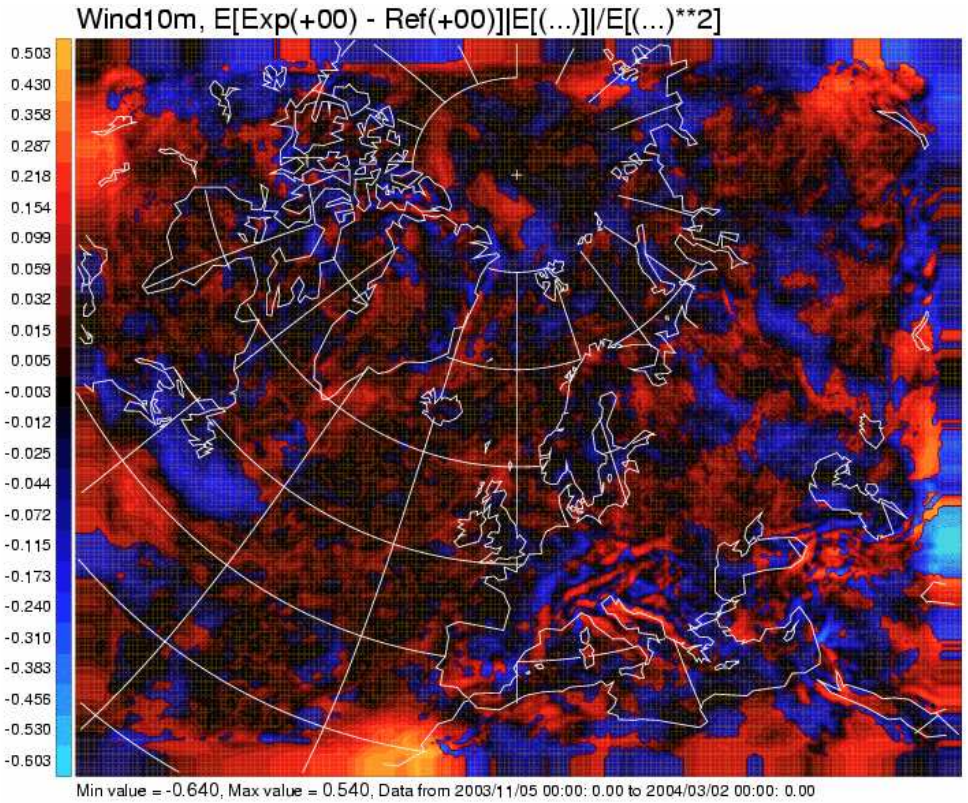

Figure 32: Geographical distribution of the wind speed +0h mean difference between 6HA and REF.

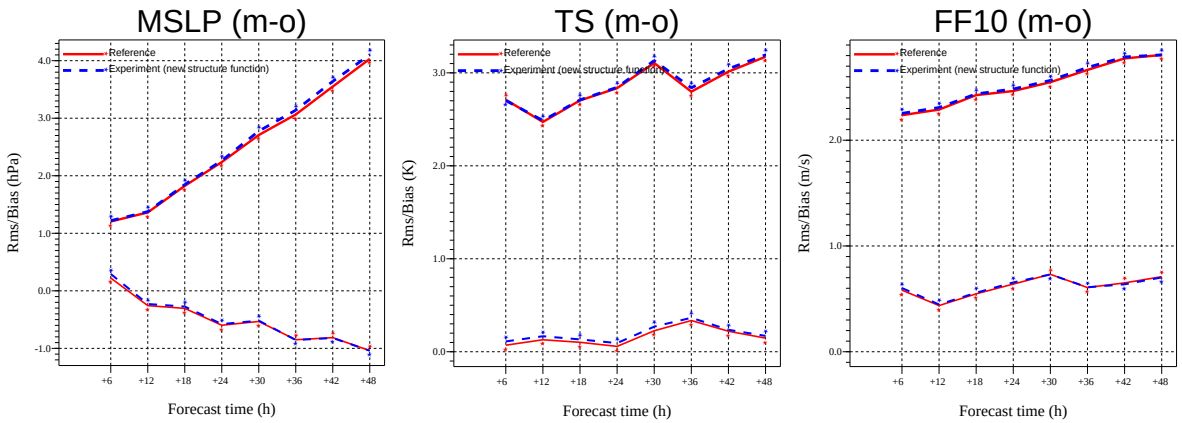

Figure 33: MSLP, T2m and FF10 error as a function of forecast length, in the NSF experiment.

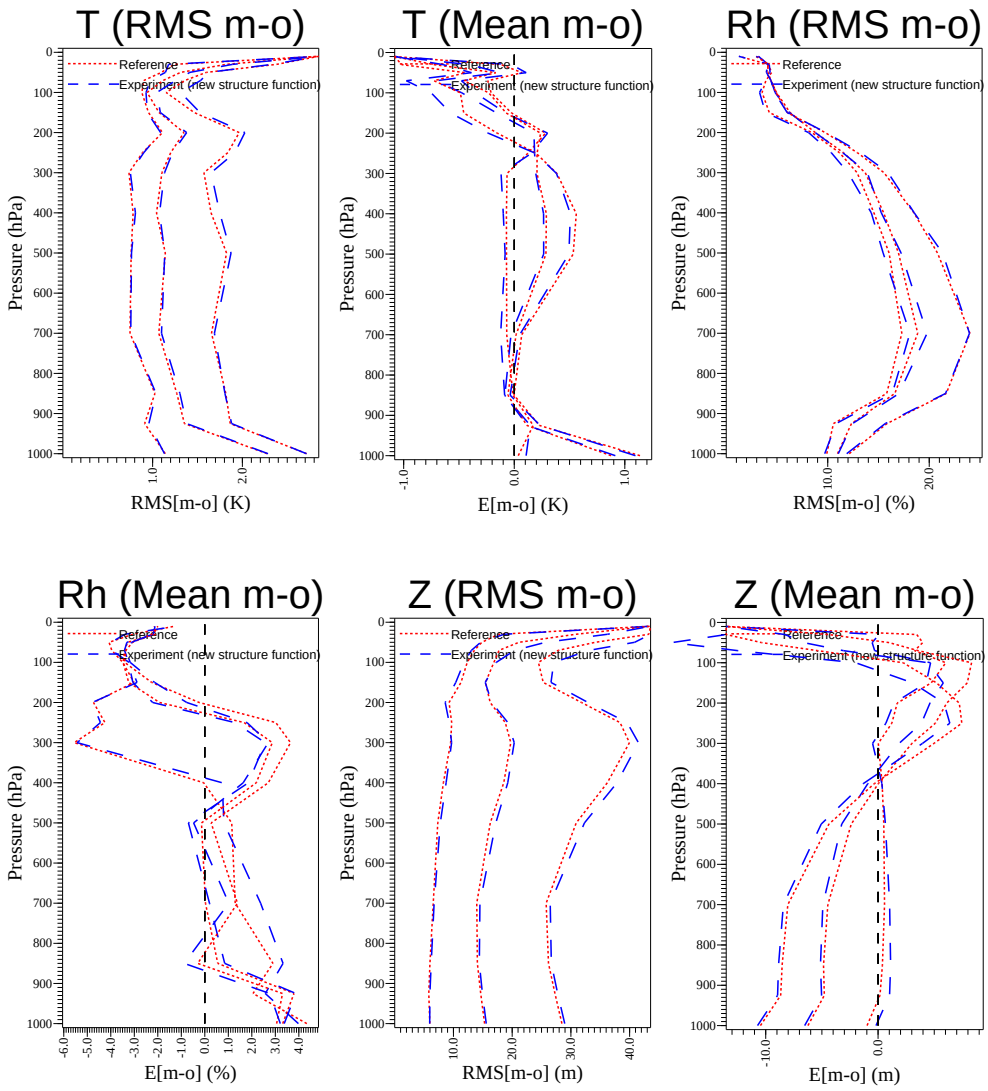

Figure 34: Vertical distribution of the temperature, relative humidity and geopotential error in the NSF experiment.

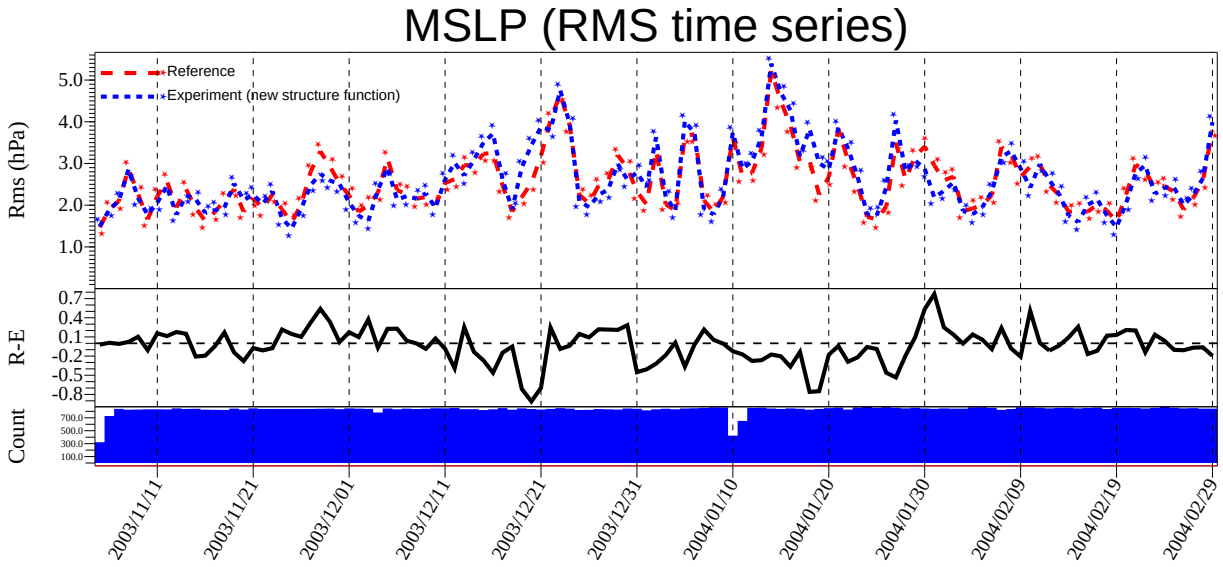

Figure 35: Time series of the daily contribution to the MSLP error in the NSF experiment.

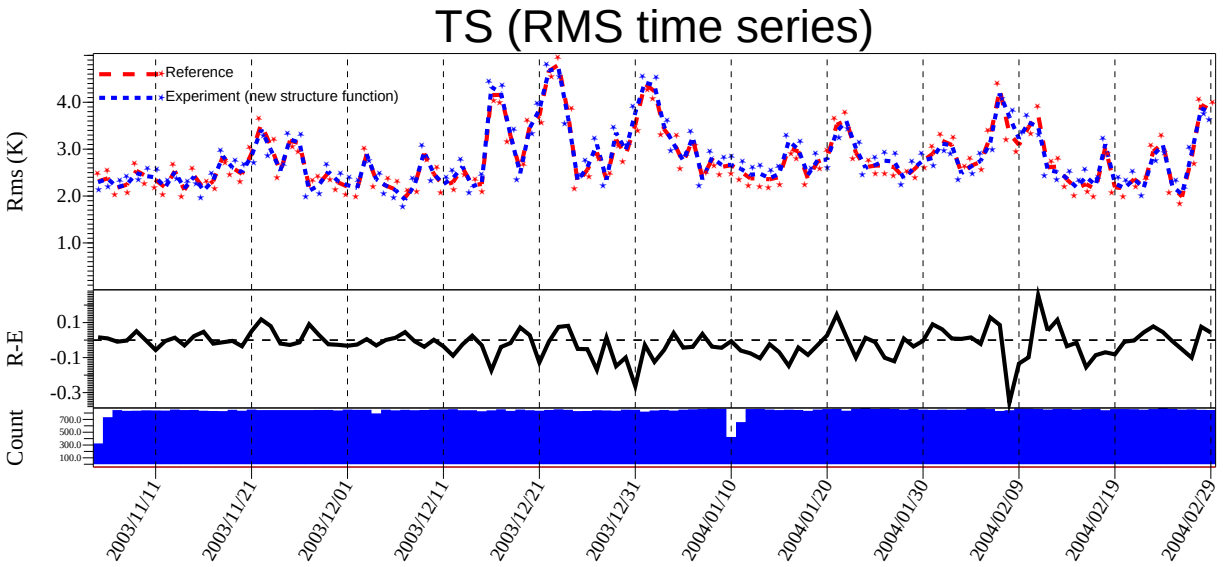

Figure 36: Time series of the daily contribution to the T2m error in the NSF experiment.

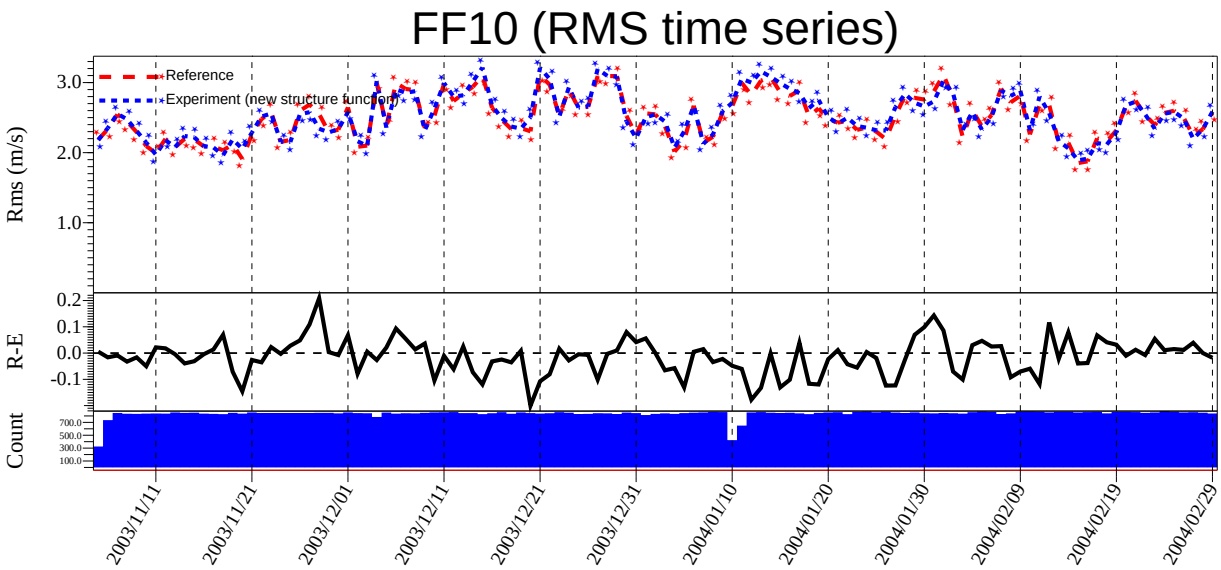

Figure 37: Time series of the daily contribution to the FF10 error in the NSF experiment.

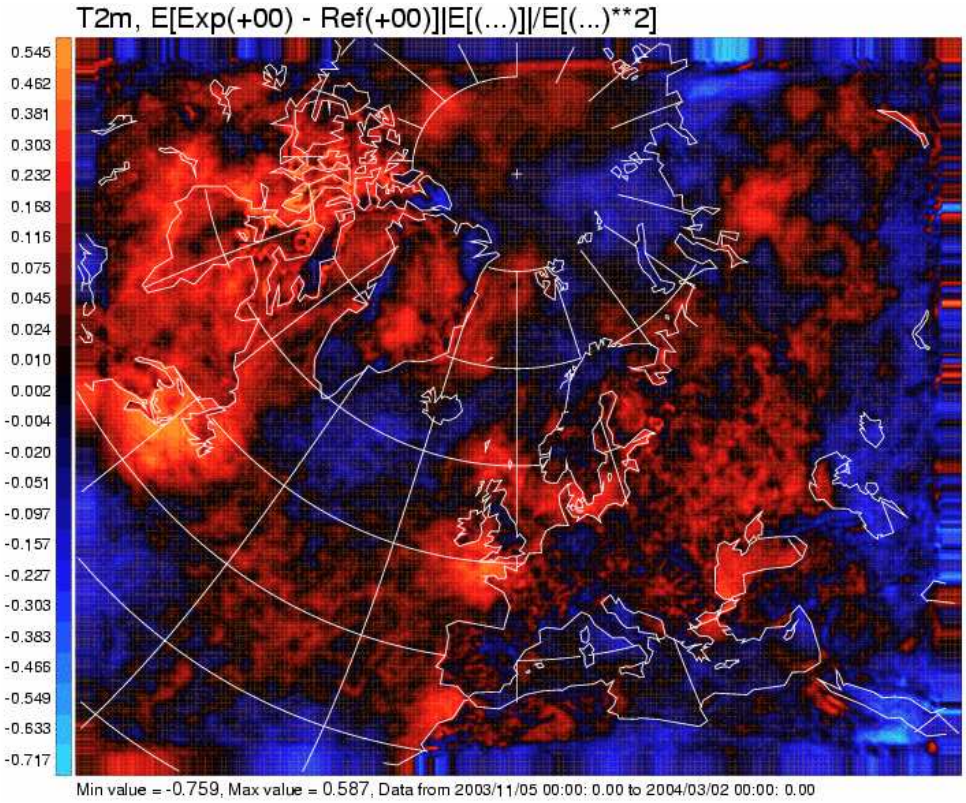

Figure 38: Geographical distribution of the T2m +0h mean difference between NSF and REF.

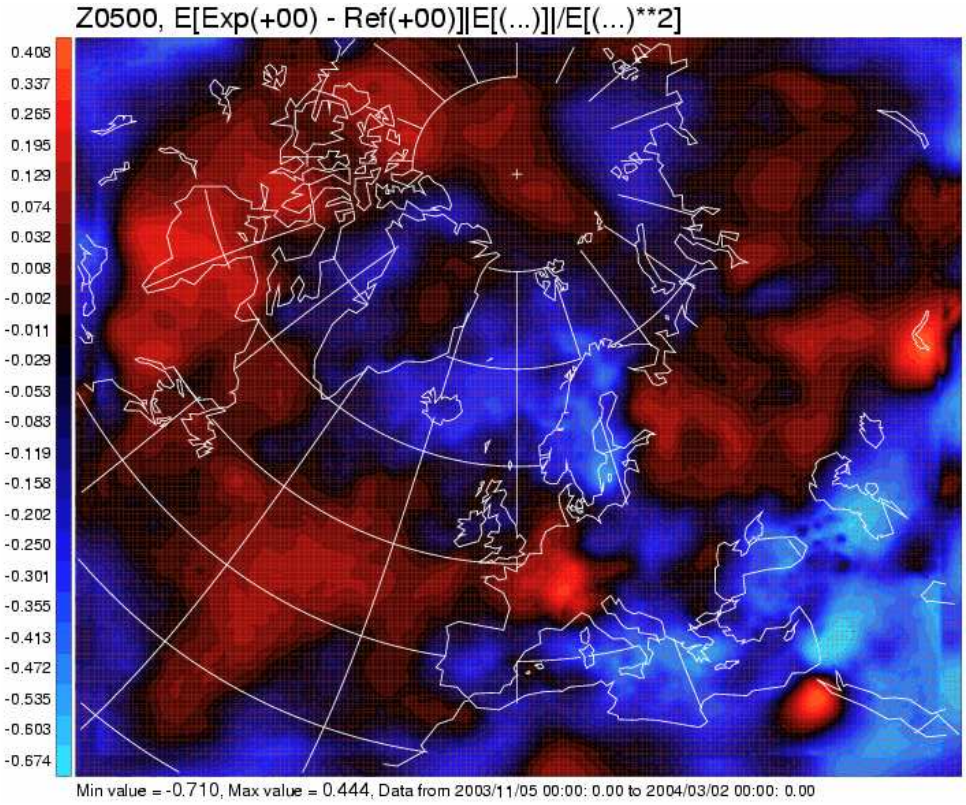

Figure 39: Geographical distribution of the Z500 +0h mean difference between NSF and REF.

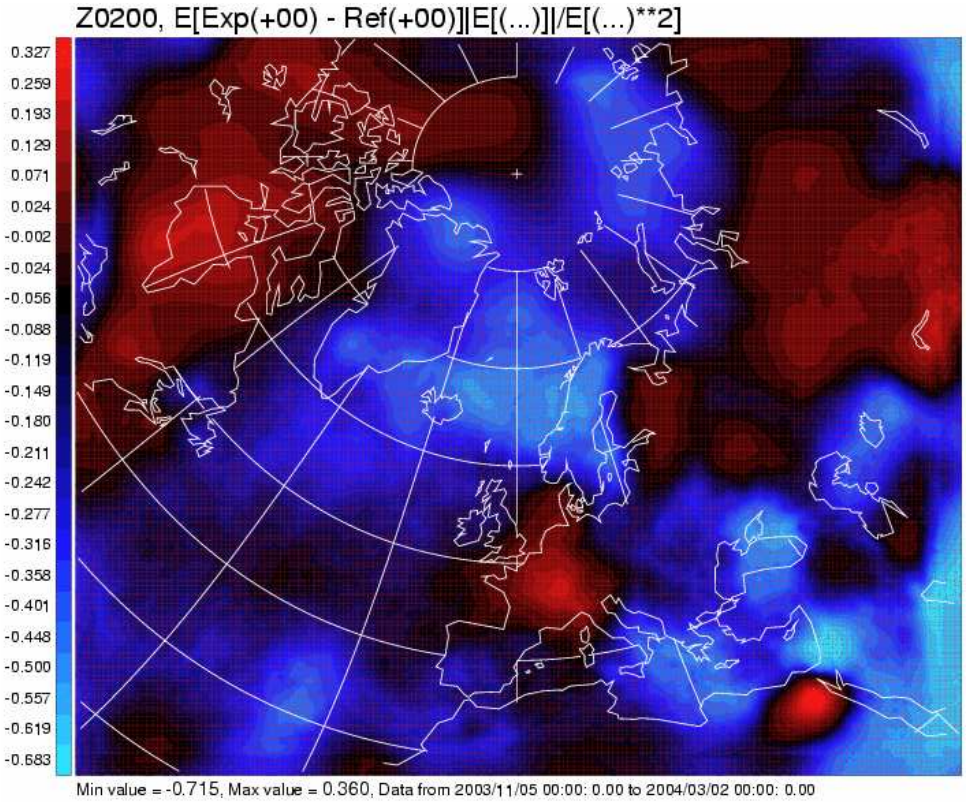

Figure 40: Geographical distribution of the Z200 +0h mean difference between NSF and REF.

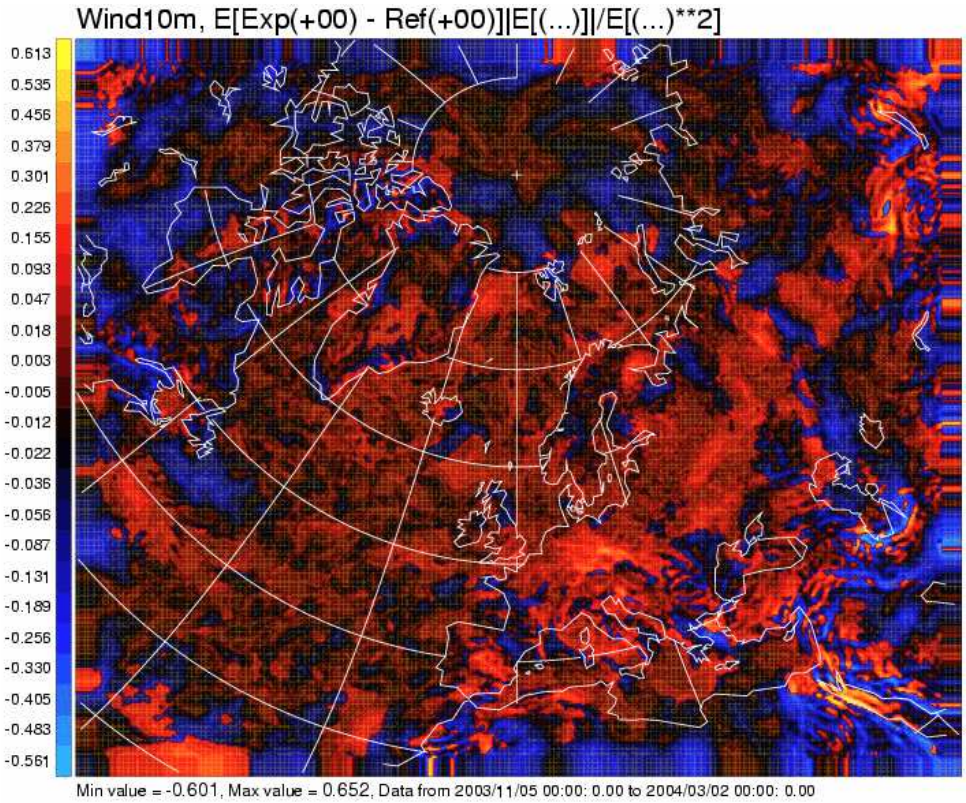

Figure 41: Geographical distribution of the wind speed +0h mean difference between NSF and REF.

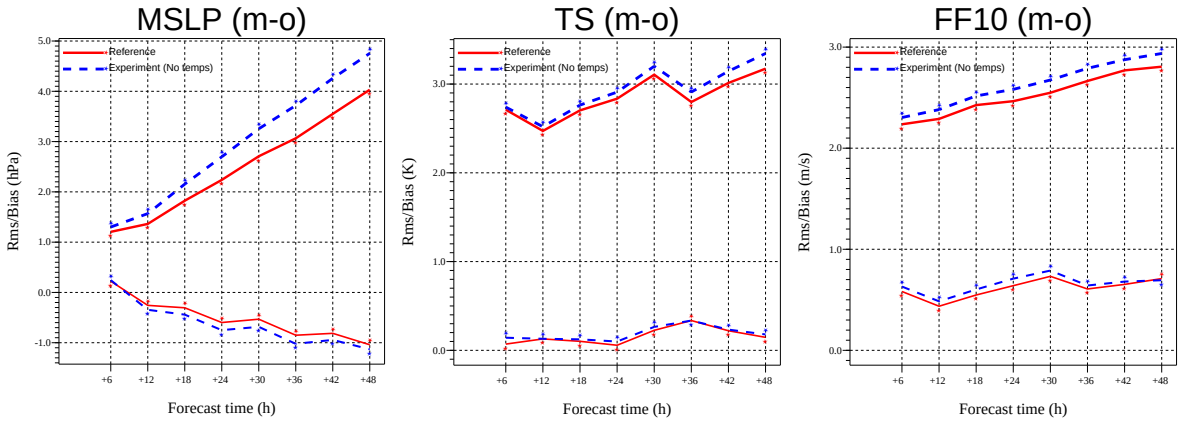

Figure 42: MSLP, T2m and FF10 error as a function of forecast length, in the NTP experiment.

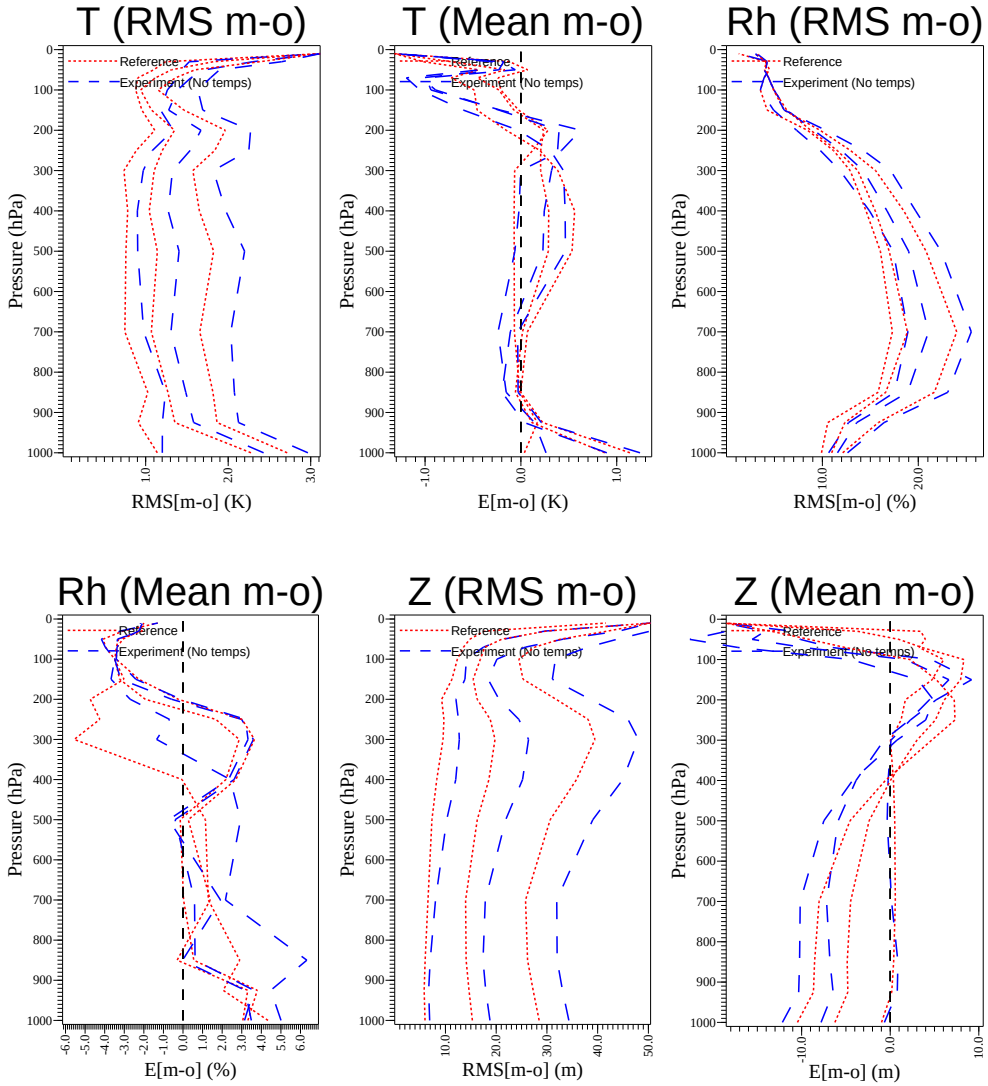

Figure 43: Vertical distribution of the temperature, relative humidity and geopotential error.

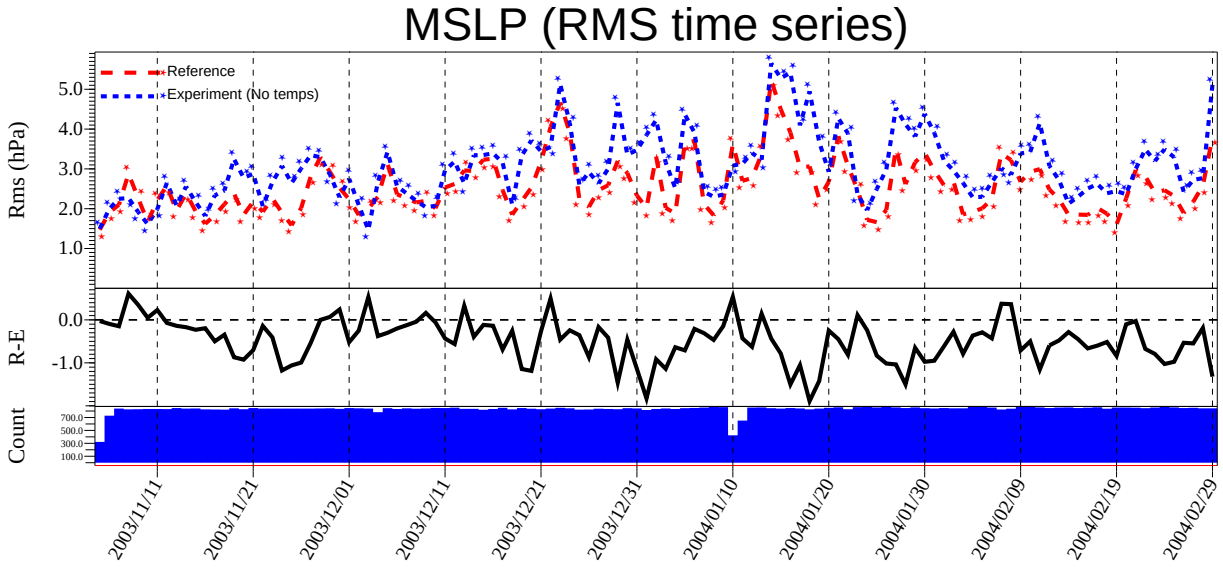

Figure 44: Time series of the daily contribution to the MSLP error in the NTP experiment.

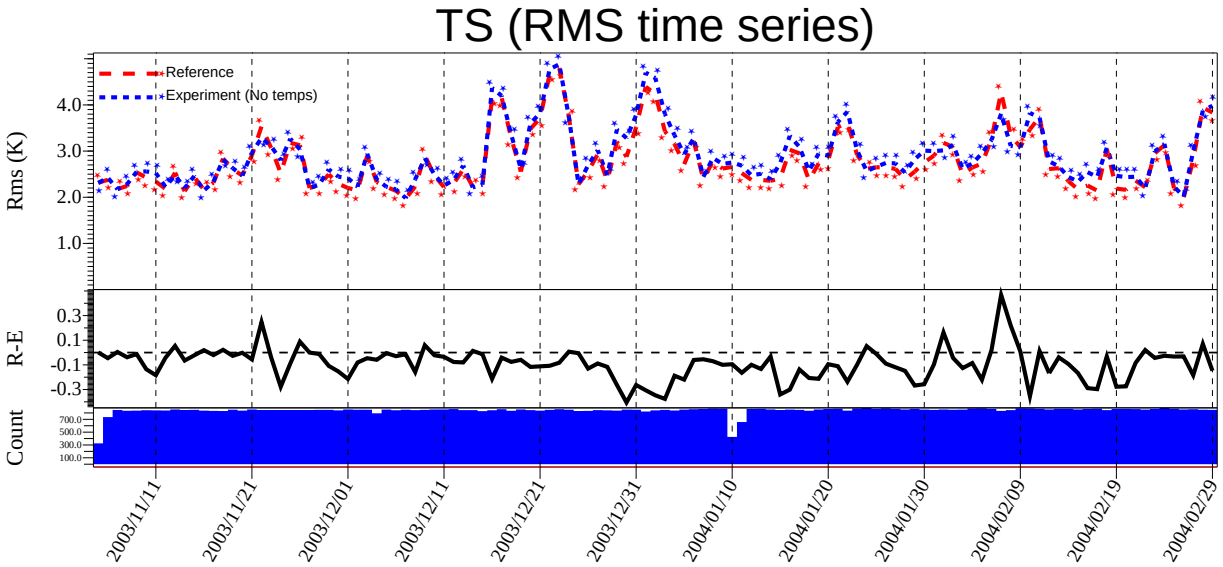

Figure 45: Time series of the daily contribution to the T2m error in the NTP experiment.

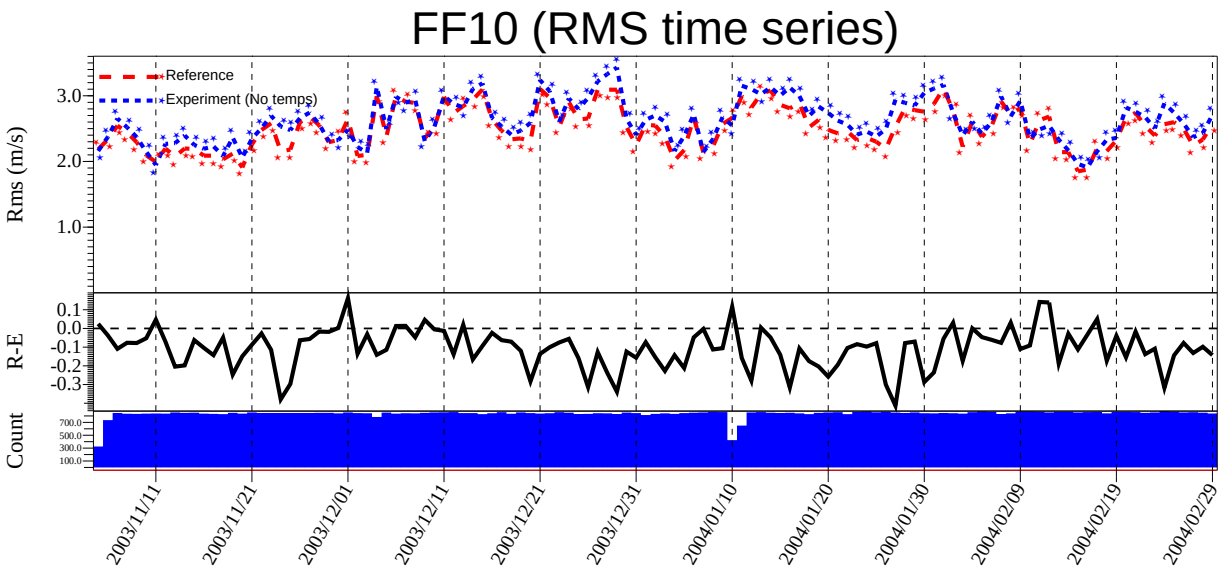

Figure 46: Time series of the daily contribution to the FF10 error in the NTP experiment.

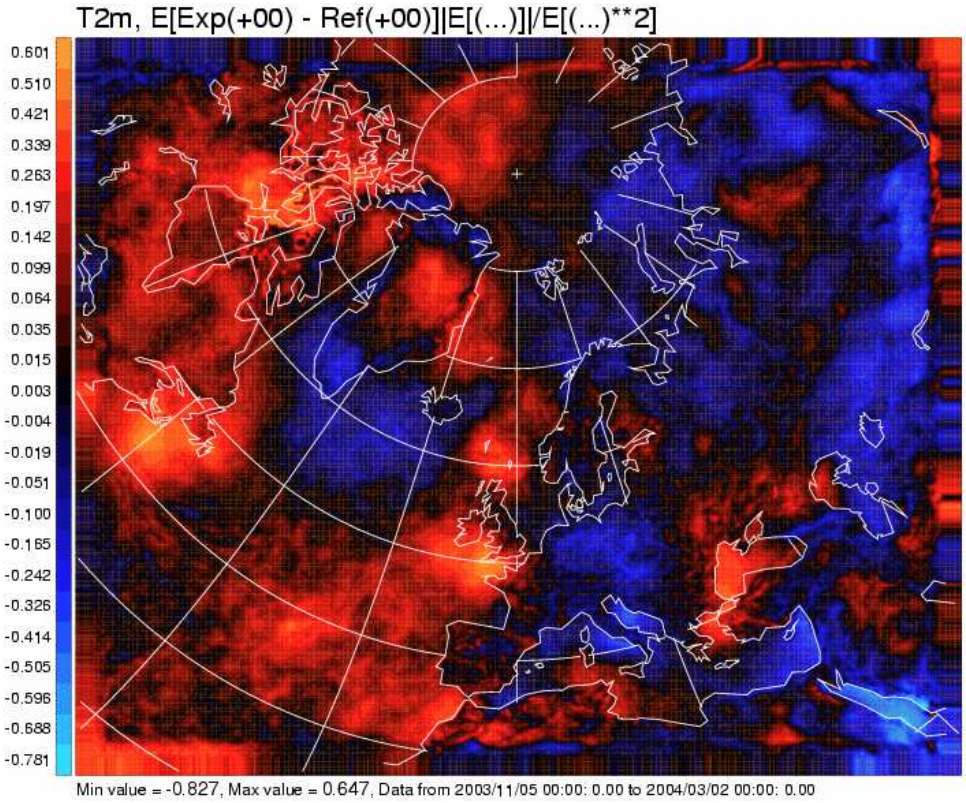

Figure 47: Geographical distribution of the T2m +0h mean difference between NTP and REF.

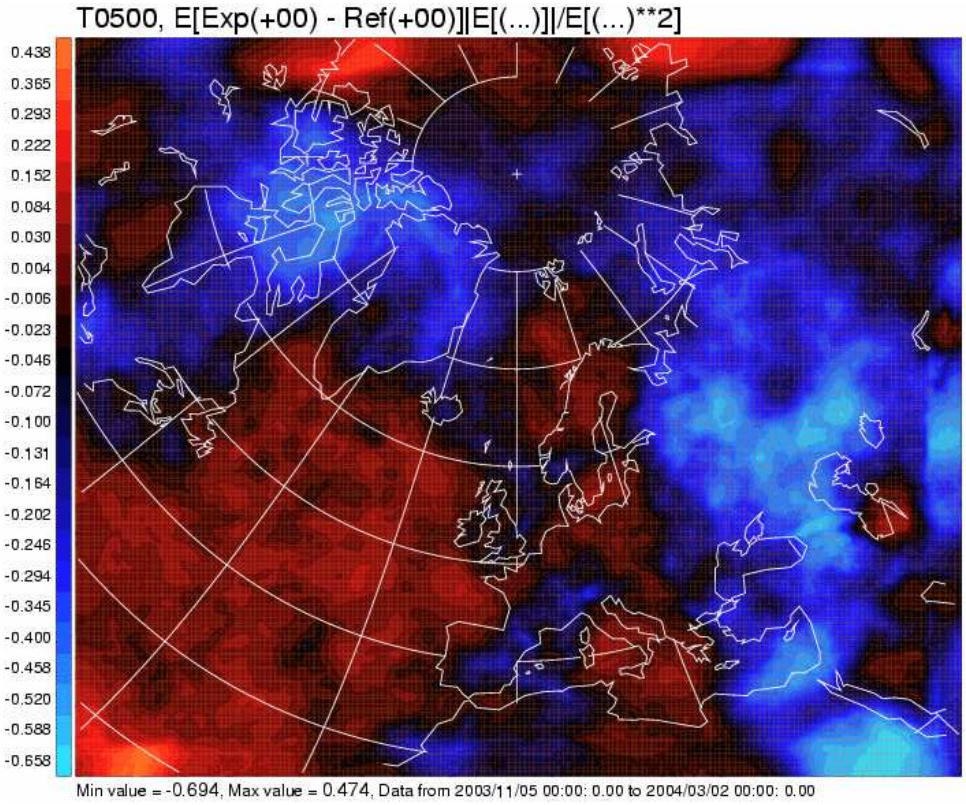

Figure 48: Geographical distribution of the T500 +0h mean difference between NTP and REF.

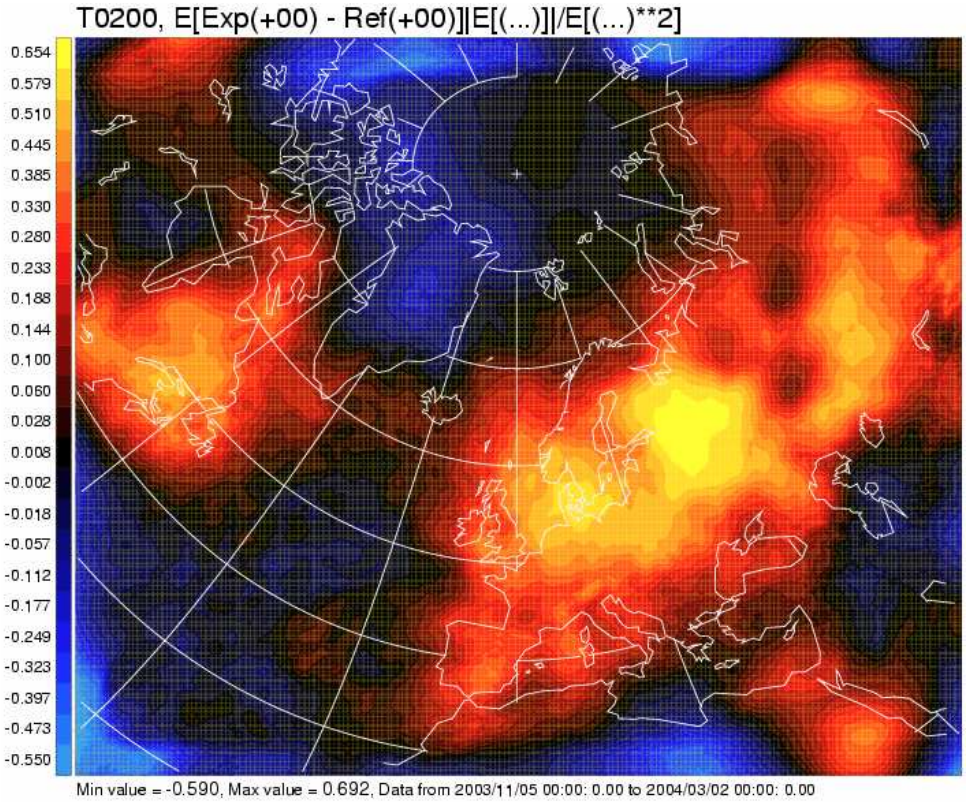

Figure 49: Geographical distribution of the T200 +0h mean difference between NTP and REF.

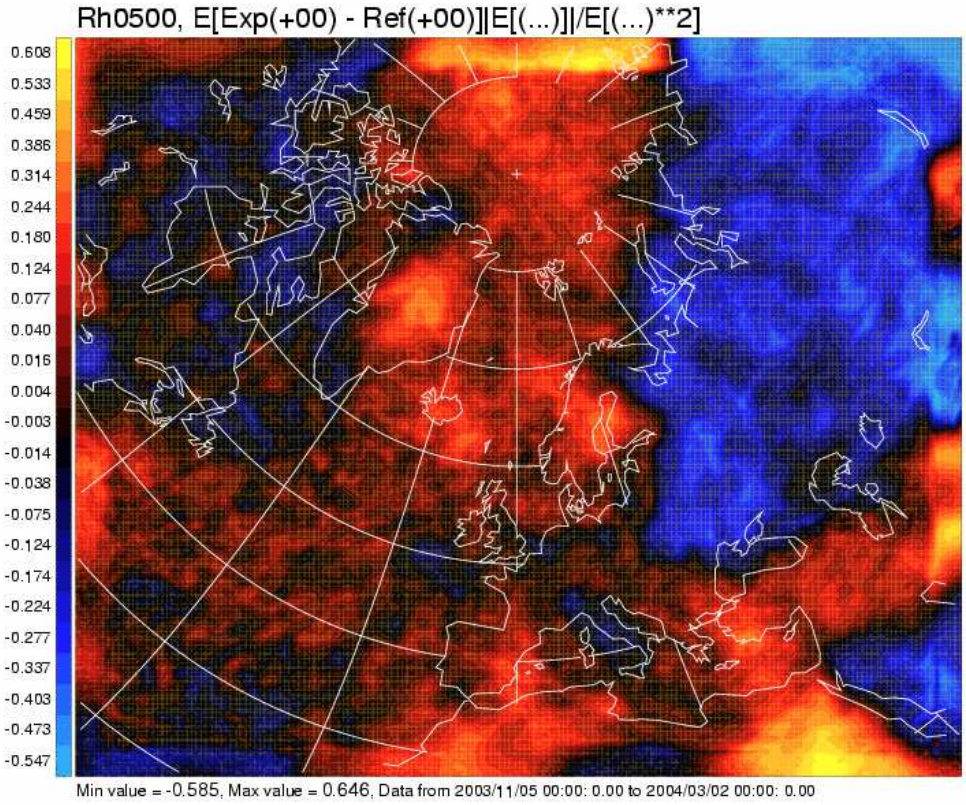

Figure 50: Geographical distribution of the Rh500 +0h mean difference between NTP and REF.

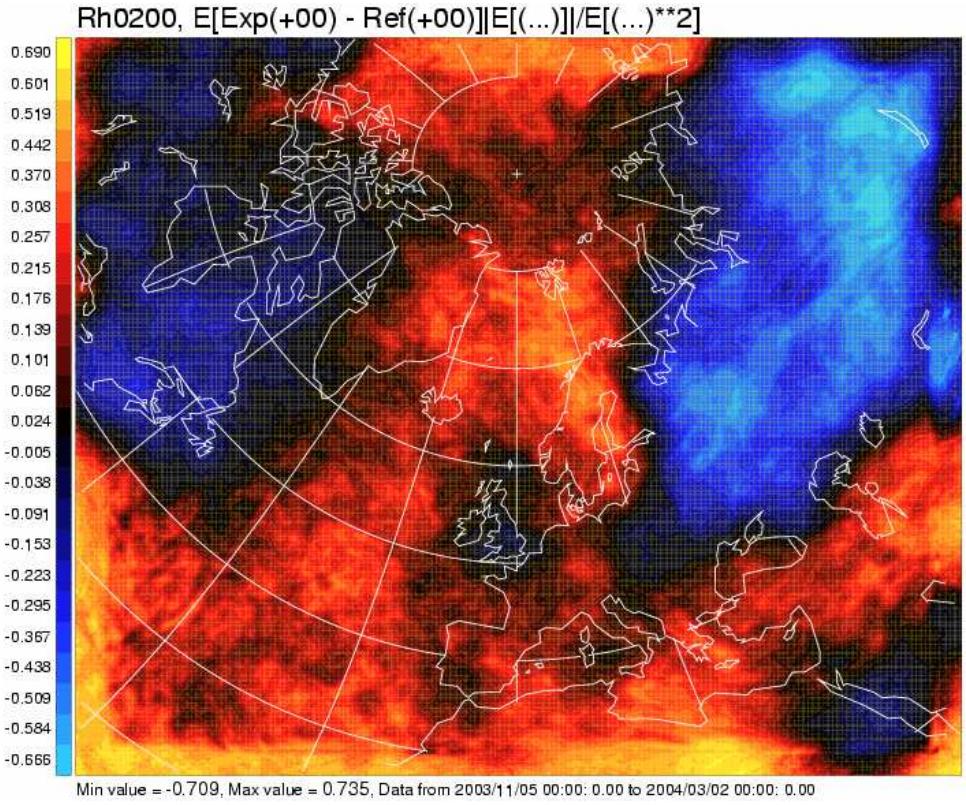

Figure 51: Geographical distribution of the Rh200 +0h mean difference between NTP and REF.

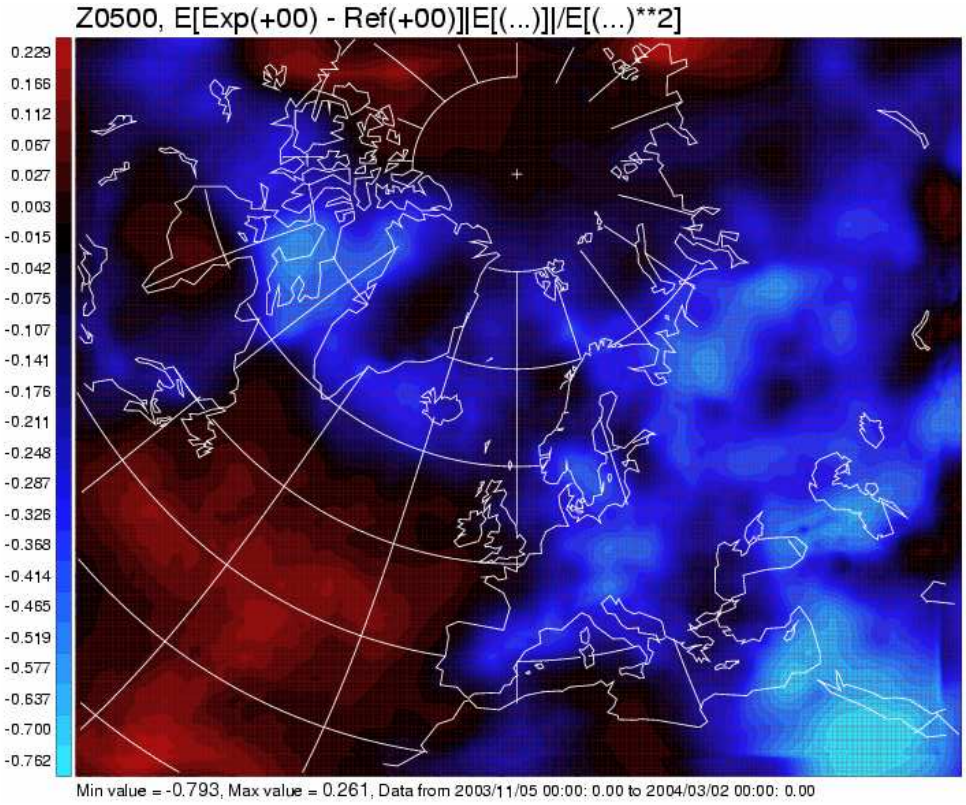

Figure 52: Geographical distribution of the Z500 +0h mean difference between NTP and REF.

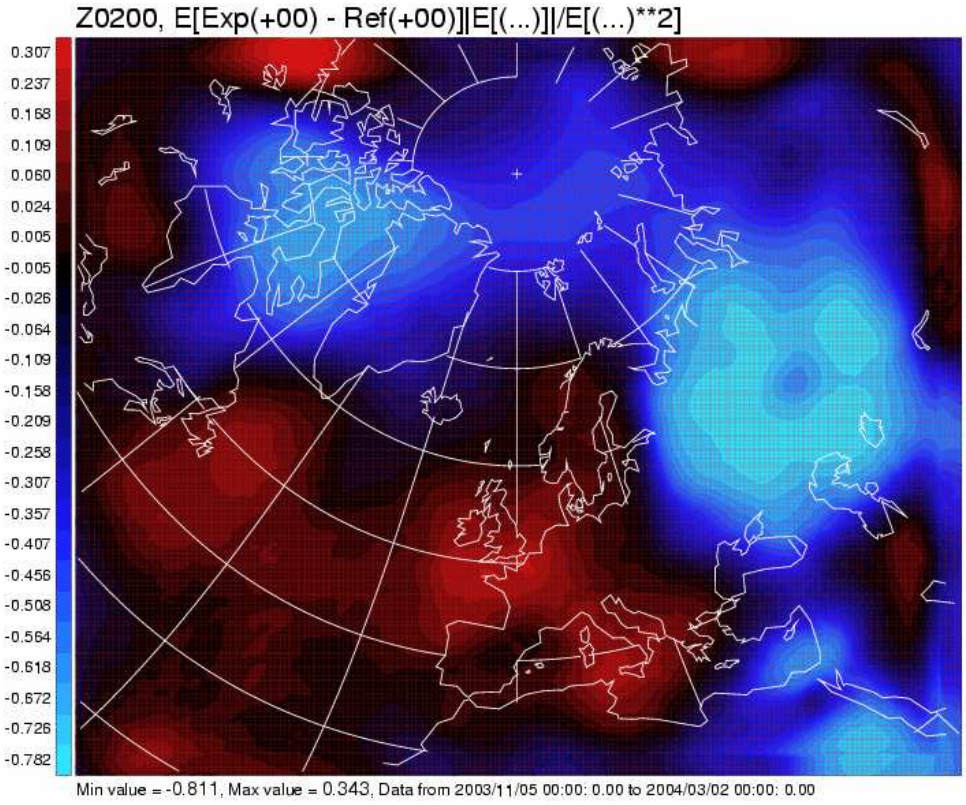

Figure 53: Geographical distribution of the Z200 +0h mean difference between NTP and REF.

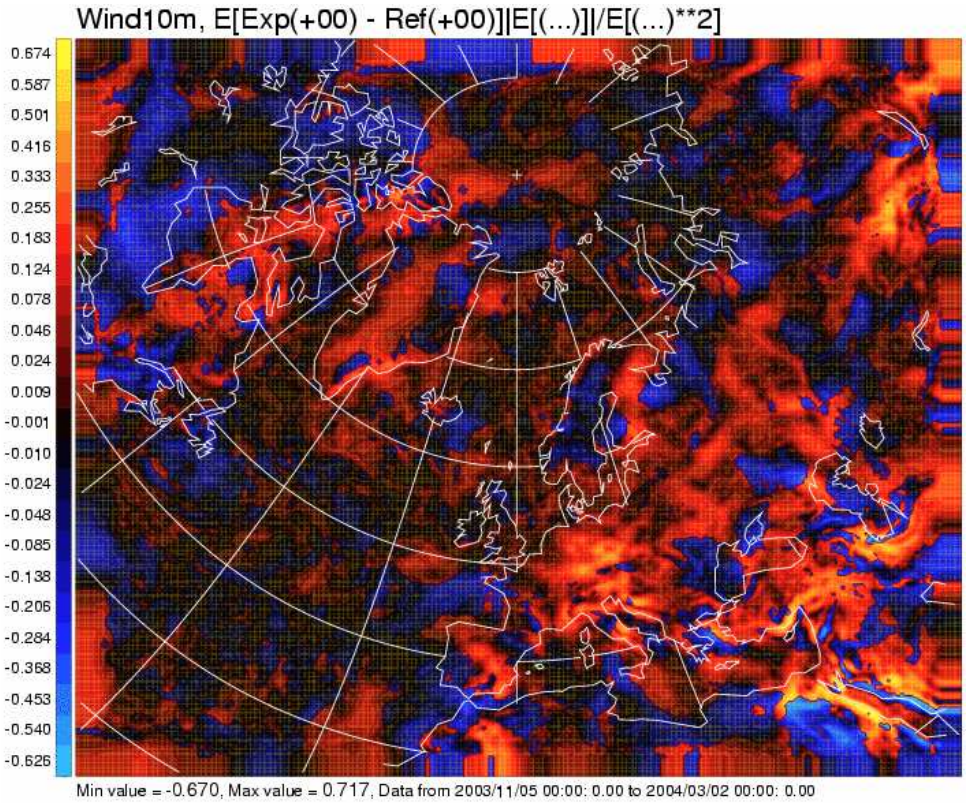

Figure 54: Geographical distribution of the wind speed +0h mean difference between NTP and REF.

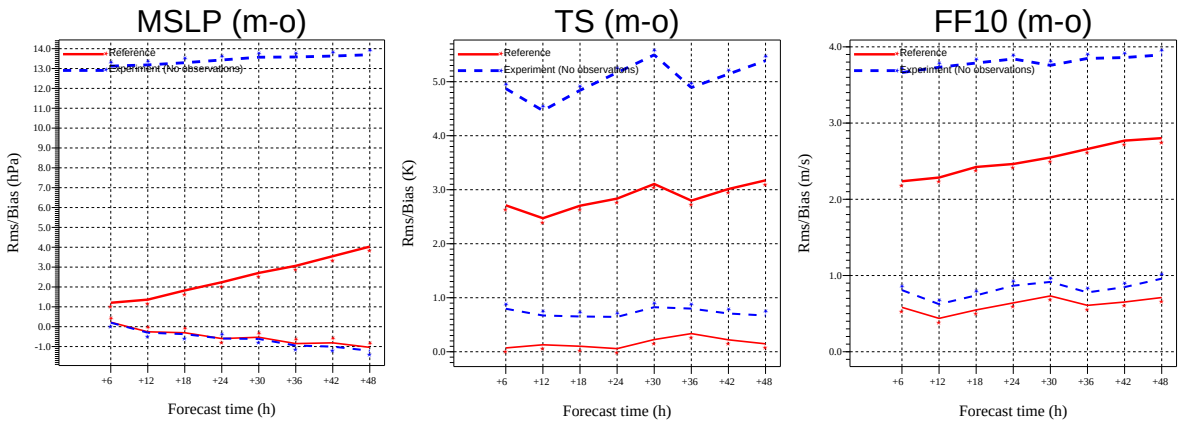

Figure 55: MSLP, T2m and FF10 error as a function of forecast length, in the NOB experiment.

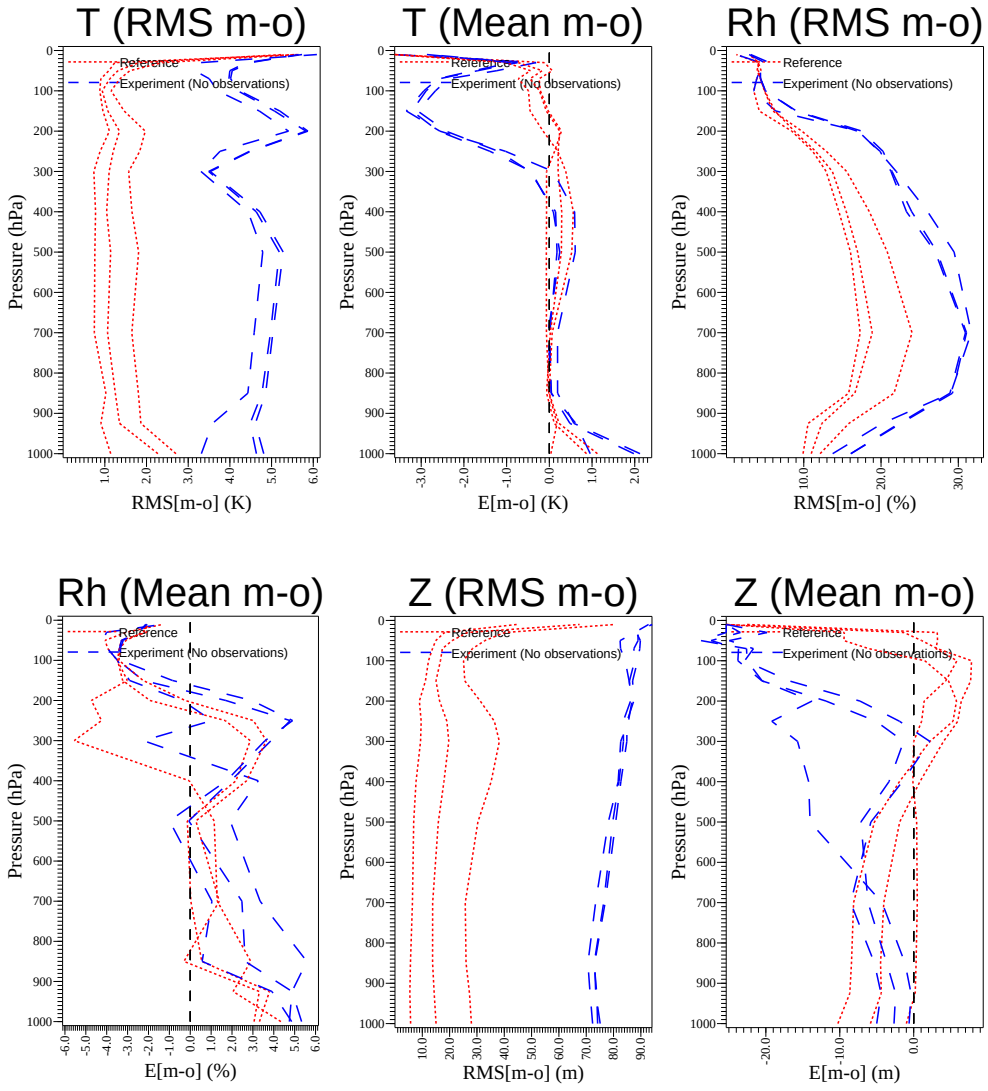

Figure 56: Vertical distribution of the temperature, relative humidity and geopotential error in the NOB experiment.

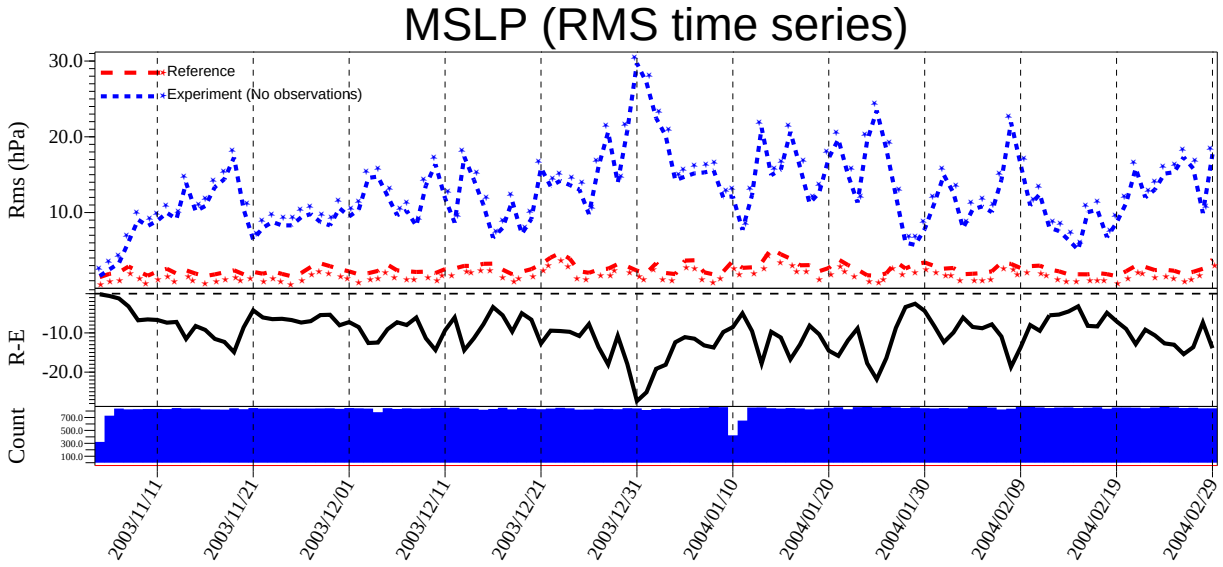

Figure 57: Time series of the daily contribution to the MSLP error in the NOB experiment.

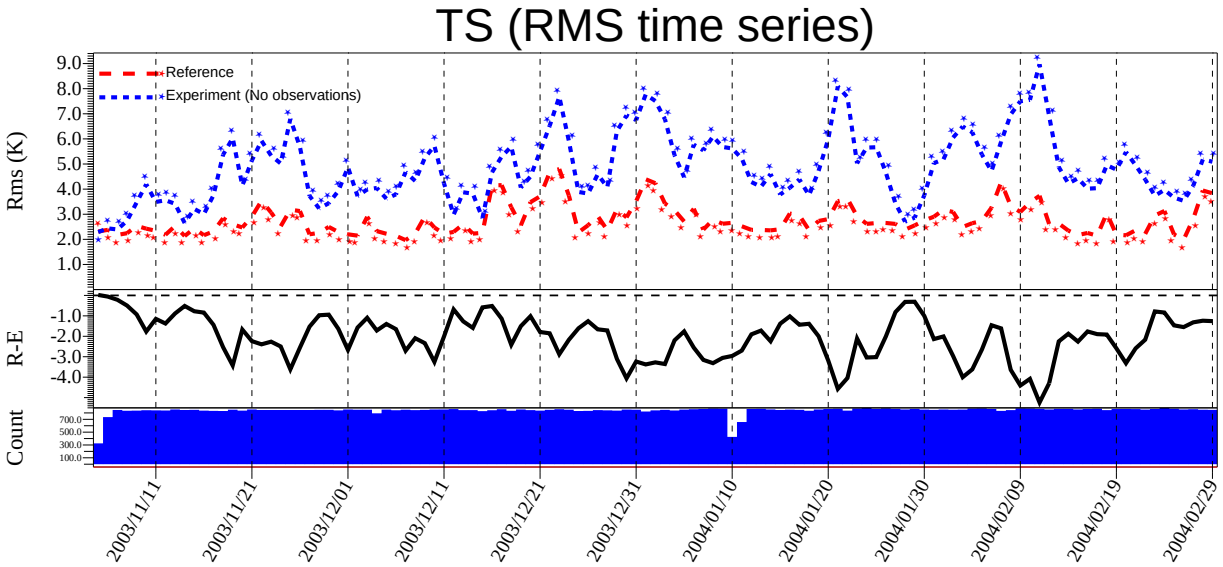

Figure 58: Time series of the daily contribution to the T2m error in the NOB experiment.

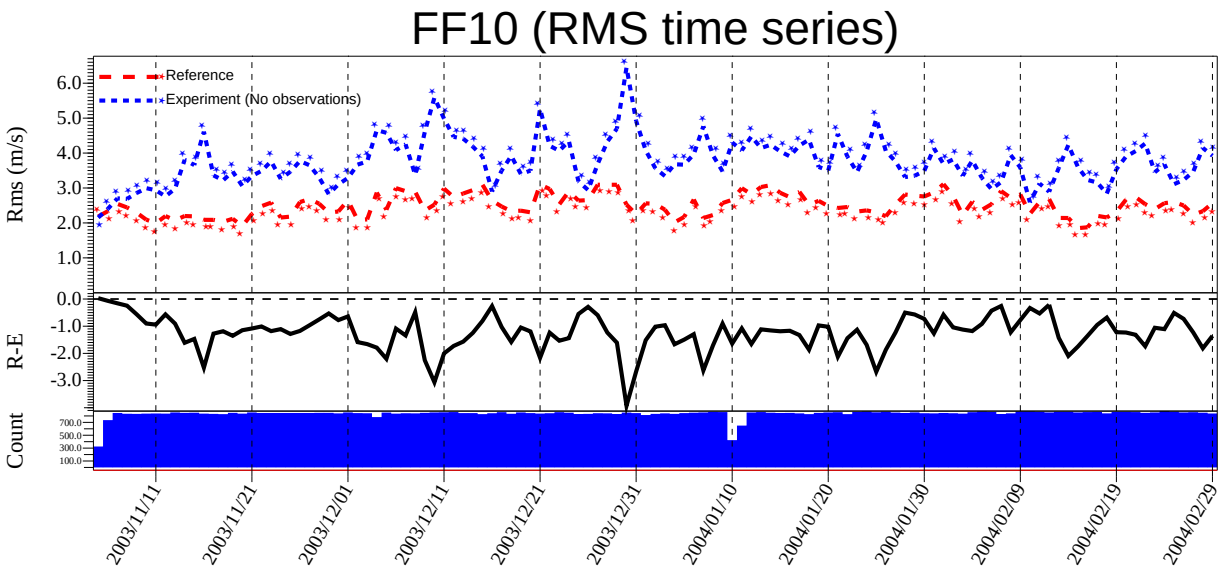

Figure 59: Time series of the daily contribution to the FF10 error in the NOB experiment.

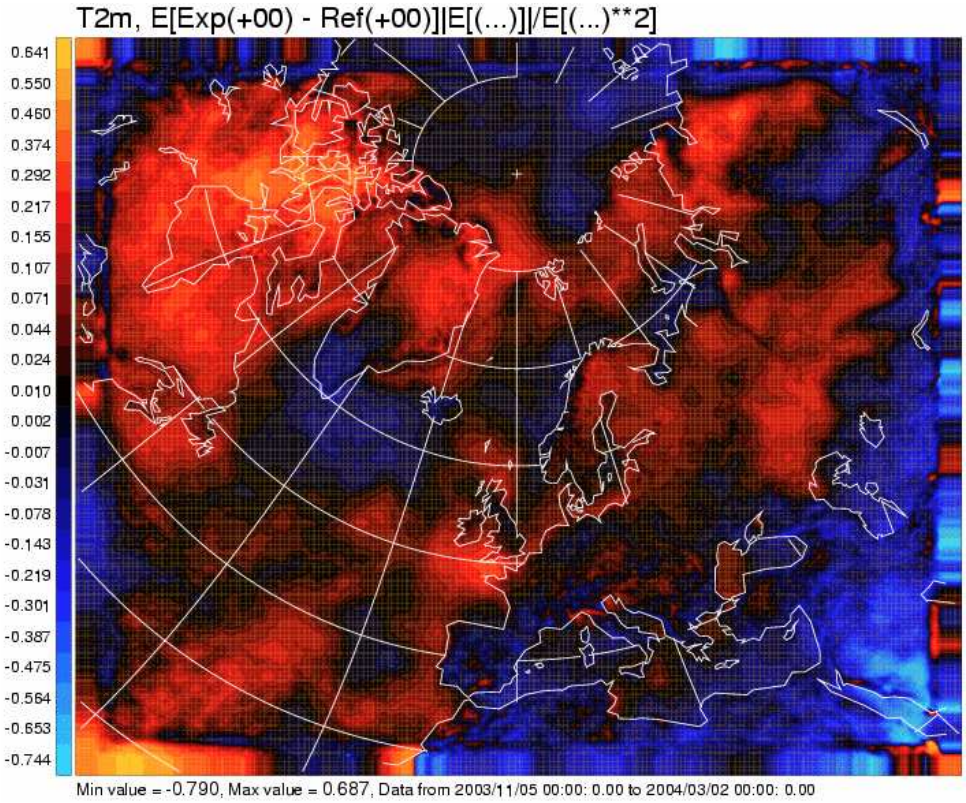

Figure 60: Geographical distribution of the T2m +0h mean difference between NOB and REF trial.

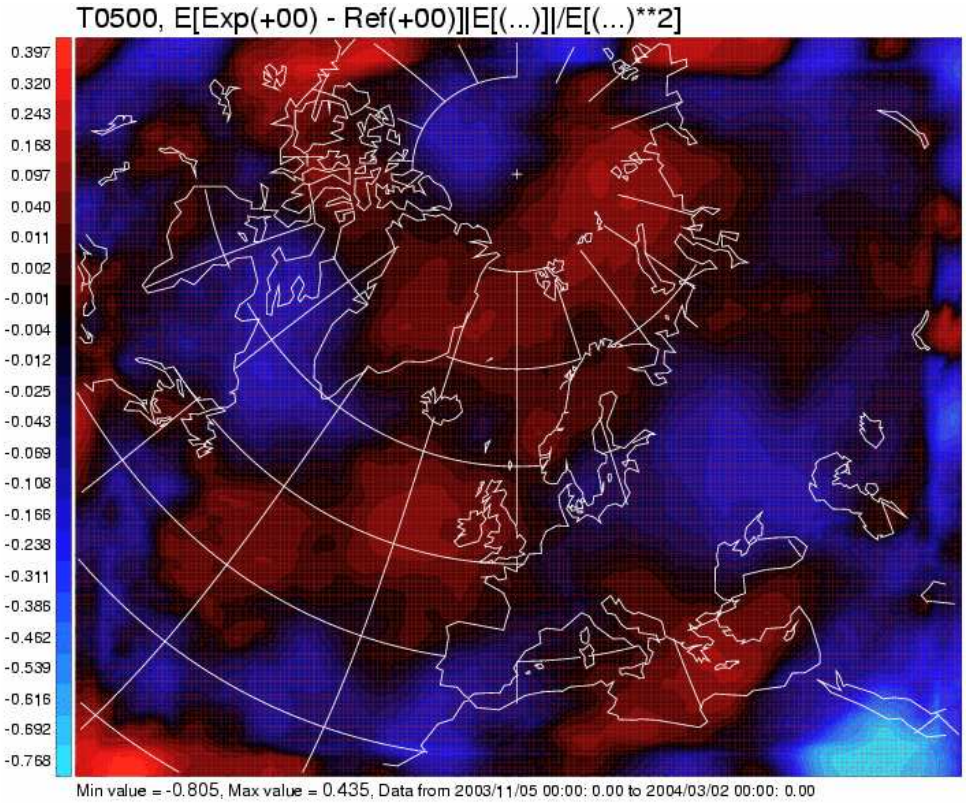

Figure 61: Geographical distribution of the T500 +0h mean difference between NOB and REF.

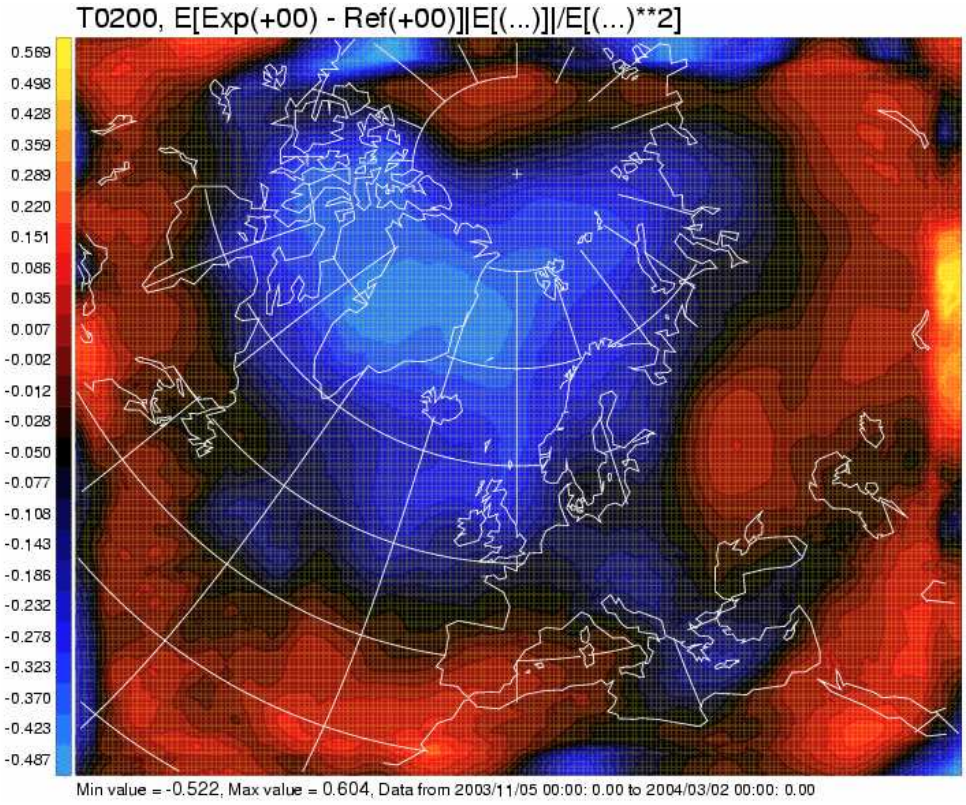

Figure 62: Geographical distribution of the T200 +0h mean difference between NOB and REF.

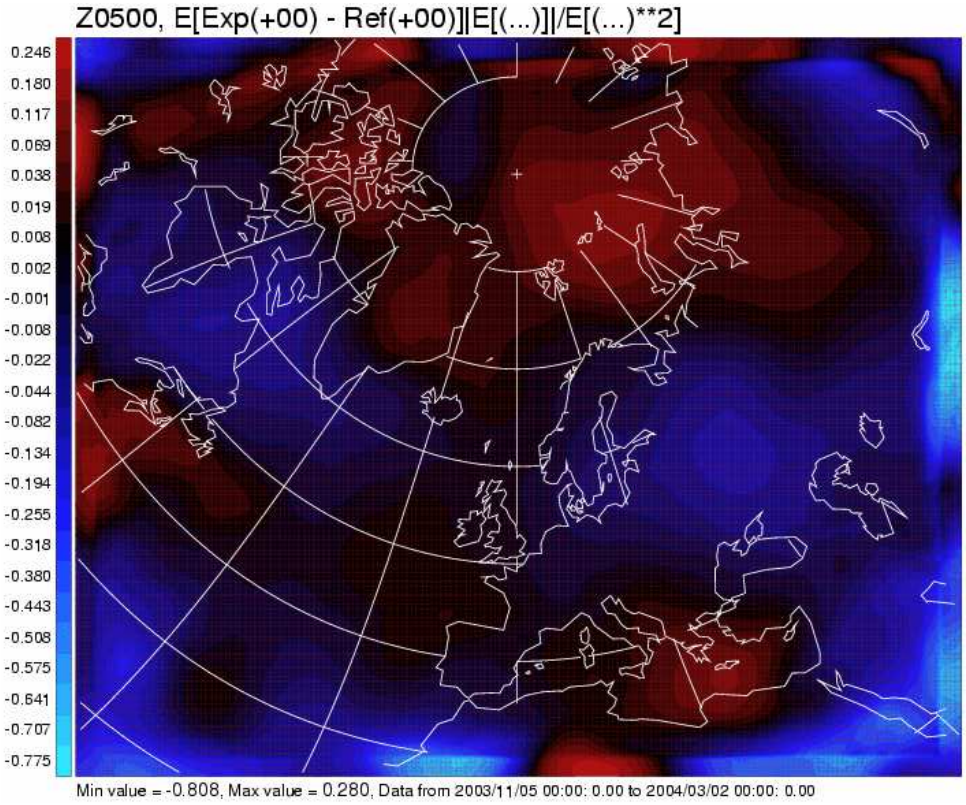

Figure 63: Geographical distribution of the Z500 +0h mean difference between NOB and REF.

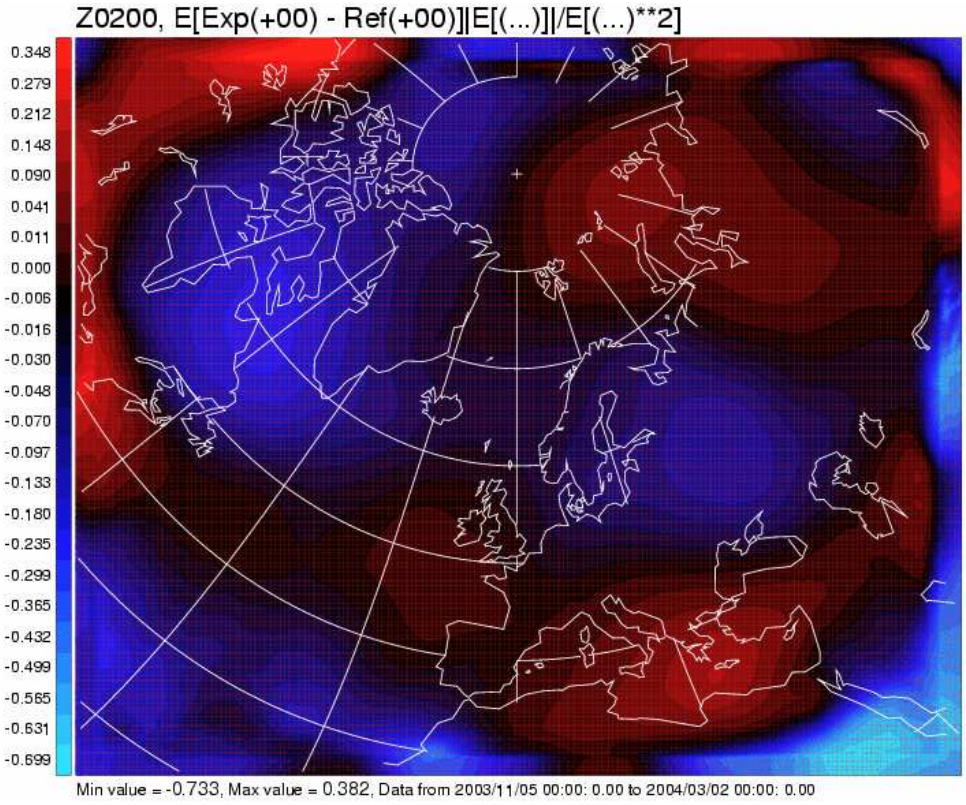

Figure 64: Geographical distribution of the Z200 +0h mean difference between NOB and REF.

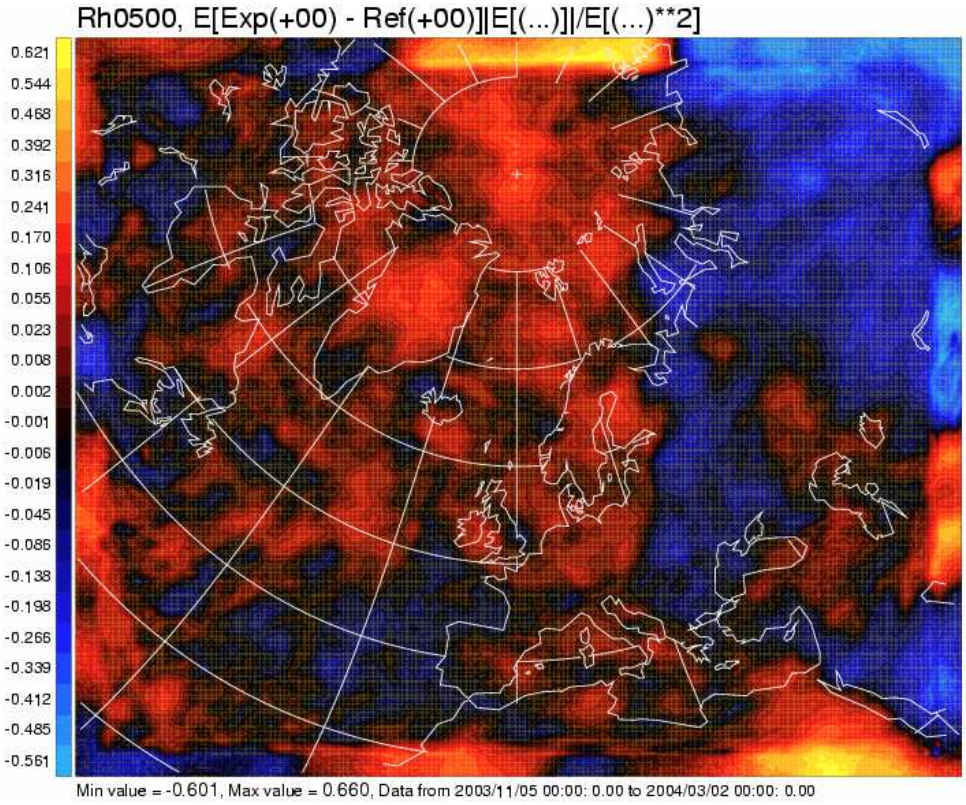

Figure 65: Geographical distribution of the Rh500 +0h mean difference between NOB and REF.

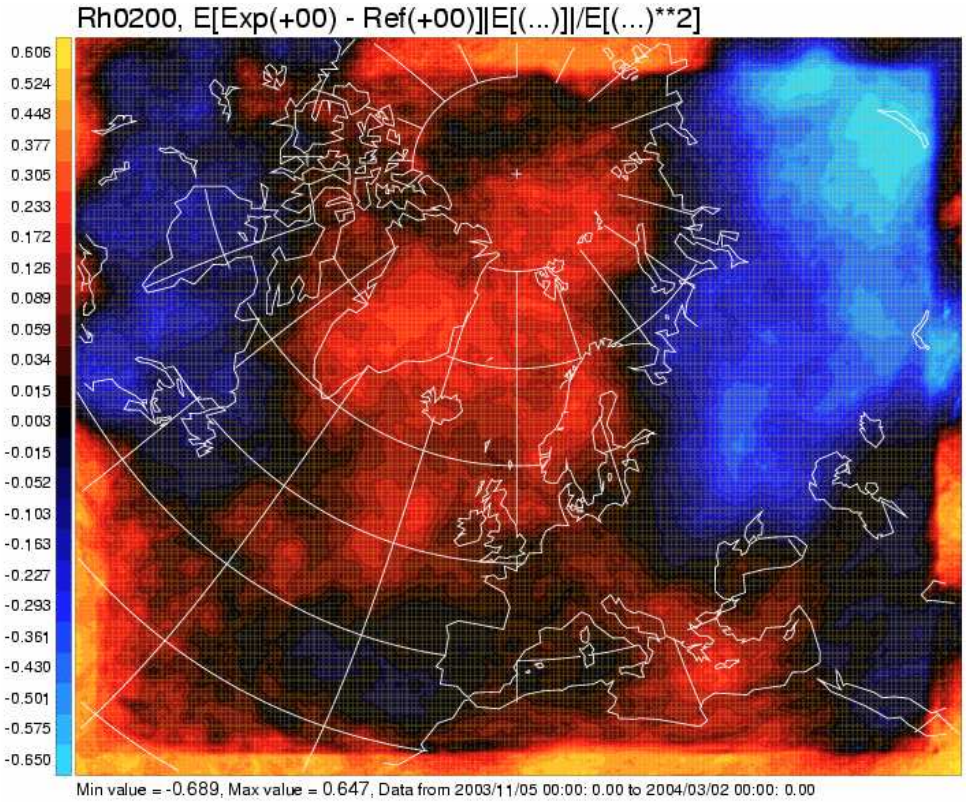

Figure 66: Geographical distribution of the Rh200 +0h mean difference between NOB and REF.

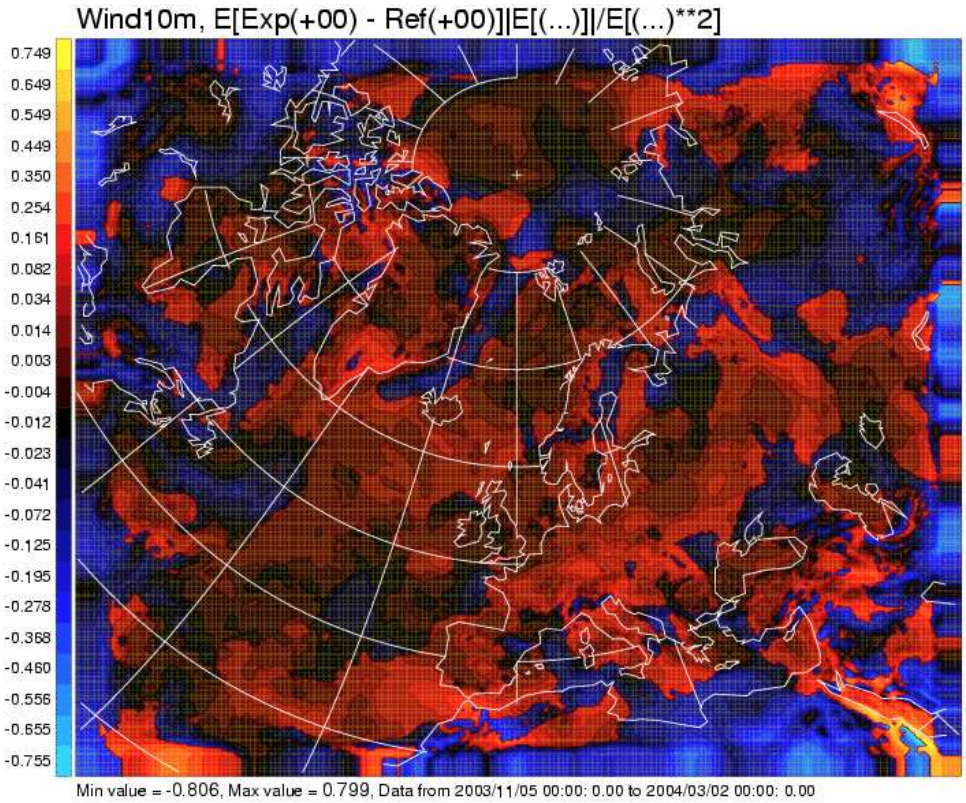

Figure 67: Geographical distribution of the wind speed +0h mean difference between NOB and REF.