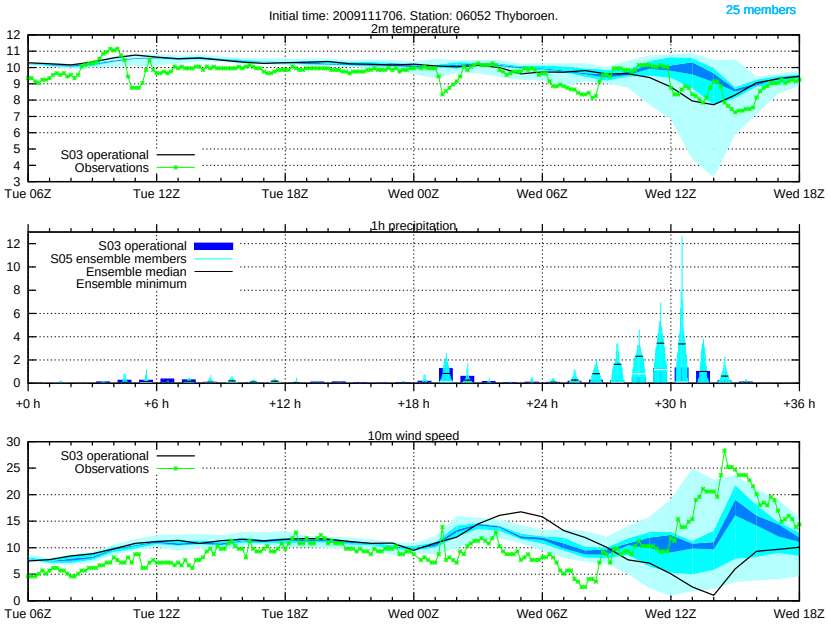
Danish Meteorological Institute
Ministry for Climate and Energy

Technical Report 09-14

A Short-Range Limited Area Ensemble Prediction System

Henrik Feddersen



Colophone

Serial title:
Technical Report 09-14

Title:
A Short-Range Limited Area Ensemble Prediction System

Subtitle:

Authors:
Henrik Feddersen

Other Contributers:

Responsible Institution:
Danish Meteorological Institute

Language:
English

Keywords:
Ensemble prediction, forecast uncertainty, HIRLAM, verification

Url:
www.dmi.dk/dmi/tr09-14.pdf

ISSN:
1399-1388

Version:

Website:
www.dmi.dk

Contents

Abstract

This report describes the experimental ensemble prediction system that has been running at DMI in real-time in its present configuration since August 2009. The ensemble prediction system is based on the HIRLAM model using the S05 setup (0.05° horizontal resolution) nested into the T15 setup (0.15° horizontal resolution). Initial condition perturbations are derived from forecast errors of the most recent S05 deterministic forecasts. Model physics is perturbed by using two differenct cloud schemes and by stochastic perturbations to the physics tendencies in the model equations ("stochastic physics"). A total of 25 ensemble members (36h forecasts) run regularly four times per day.

Verification against observations at selected Danish stations during a 90-day period (10 Aug - 7 Nov 2009) shows that DMI's ensemble prediction system compares favourably to ECMWF's ensemble prediction system. It is demonstrated that DMI's ensemble prediction system in some cases is capable of forecasting developments earlier than the operational S03 HIRLAM forecast, and it is argued that automatic precipitation point forecasts can be improved by including uncertainty estimates based on the ensemble prediction system.

As the presentation of forecasts is crucial for ensemble prediction systems, various presentation possibilities are discussed at some length in this report.

Resumé

Denne rapport beskriver det eksperimentelle ensembleprognosesystem der har kørt på DMI i sand tid i dets nuværende konfiguration siden august 2009. Ensembleprognosesystemet er baseret på HIRLAM-modellen i S05-setup'et (0,05° horisontal opløsning) indlejret i T15-setup'et (0,15° horisontal opløsning). Perturbationer af begyndelsesbetingelserne er afledt af prognosefejl fra de seneste deterministiske S05-prognoser. Modelfysikken perturberes ved at anvende to forskellige skyskemaer, samt ved at perturbere fysiktendenserne i modelligningerne ("stokastisk fysik"). I alt kører 25 medlemmer (36-timers prognoser) regelmæssigt fire gange i døgnet.

Verifikation mod observationer fra udvalgte danske stationer i en 90-dages periode (10. aug.-7. nov. 2009) viser at DMI's ensembleprognosesystem klarer sig godt i sammenligning med ECMWF's ensembleprognosesystem. Det demonstreres at DMI's ensembleprognosesystem i nogle tilfælde er i stand til at forudsige udviklinger på et tidligere tidspunkt end den operationelle S03 HIRLAM-prognose, og der argumenteres for at automatiske punktprognoser kan forbedres ved at inkludere usikkerhedsestimater baseret på ensembleprognosesystemet.

Da præsentationen af prognoserne er afgørende for ensembleprognosesystemer, diskuteres diverse præsentationsmuligheder i nogen udstrækning i denne rapport.

1. Introduction

The idea that the future state of the atmosphere could be determined by knowledge of its initial state and integration of the equations of motion dates back to the paper by Bjerkness (1904) and is still the basis of modern numerical weather prediction (NWP). With the aid of modern computers and increasingly accurate NWP models the detail and quality of weather forecasts have increased immensely since Richardson (1922) first suggested how to solve the equations numerically. These ideas are based on a deterministic view of the atmosphere and the notion that if the initial state can be determined sufficiently accurately, and the model can be made sufficiently detailed, then the future state of the atmosphere is indeed predictable. The deterministic view is also evident in, e.g., the Wikipedia definition of weather forecasting: *"Weather forecasting is the application of science and technology to predict the state of the atmosphere for a future time and a given location."*

However, the atmosphere is highly nonlinear implying that infinitesimal differences in the initial conditions may lead to a completely different forecast. Lorenz (1963) realized this as he was experimenting with a highly simplified but still nonlinear model of convective flow.

Even with the best estimates of the initial conditions and the most accurate NWP models, nonlinearities combined with unavoidable uncertainty not only in the initial conditions, but also in the formulation of the NWP models themselves, are the cause of forecast errors. The forecast errors are mostly small, but occasionally they are significant. Forecasters as well as users of weather forecasts know from experience that there is always some uncertainty associated with a forecast (Morss et al., 2008). This is reflected in the way in which the forecasts are formulated: For example, temperature and wind forecasts are presented as ranges in which the forecaster believes the true temperature and wind will fall, and rainfall may, e.g., be presented as "a risk of showers". Still, forecasts occasionally fall outside the presented uncertainty range. In the medium-range the synoptic development may diverge from the forecast, while in the short-range one of the main challenges is to predict where and when convective rainfall will develop along with the intensity of the rainfall. The "poor man's approach" to addressing forecast uncertainty is to consult alternative models and to check consistency with older forecasts.

The inherent uncertainties in the initial conditions led Epstein (1969) to propose a stochastic-dynamic approach in which the initial conditions were described in terms of probability densities. Unfortunately, the resulting equations for the evolution of the initial probability density – and even the evolution of just the first and second moments of the probability densities – are not numerically solvable in practice. Leith (1974) proposed a practical solution where an ensemble of model integrations were generated such that each member of the ensemble was started from a random perturbation of the initial condition. In this way it would be possible to produce probabilistic forecasts by simply counting the fraction of ensemble members that would predict a certain event.

Perturbation of the initial conditions is still the basis of modern ensemble forecasting, although nowadays the applied initial condition perturbations are flow dependent and favour fast growing modes, either through the use of *singular vectors* (Mureau et al., 1993; Molteni et al., 1996) or *bred vectors* (Toth and Kalnay, 1993; Tracton and Kalnay, 1993).

Thus, the perturbations are not random, and they do not sample the (unknown!) observation uncertainty. The perturbations tend to maximize the ensemble dispersion, and so the individual ensemble members will in general yield less likely but more extreme outcomes than the random approach. Nevertheless, experience has shown that ensemble forecasts tend to be underdispersive in

the sense that (too) many verifying observations fall outside the ensemble range. Though not perfect such ensemble forecasts may still provide useful information about the uncertainty of an associated deterministic forecast as well as provide forecasters with a warning of possible extreme events. For relatively small ensemble sizes extreme events (that also tend to be rare in the forecasts) would more likely be missed if a random perturbation approach was used.

The computation of initial condition perturbations implicitly assumes a perfect model, but as models in practice are not perfect, model uncertainty is, loosely speaking, included in the initial condition perturbations. That is, if the amplitudes of the initial condition perturbations are tuned to yield an ensemble spread that matches the forecast error a few days into the forecast, then they may not reflect the observation uncertainty correctly, because some of the forecast error is caused by model error. Additionally, observation error also hamper estimation of forecast error, and in general it is very difficult to separate forecast error that is due to initial condition uncertainty from forecast error that is due to model error.

Recently, there has been a growing interest in the use of *ensemble data assimilation* that relates the initial condition perturbations more to observation uncertainty. Here variants of the ensemble Kalman filter appear most popular (Houtekamer and Mitchell, 1998; Wang and Bishop, 2003), although at ECMWF tests have been made with a small ensemble of 4D-Var analyses (Palmer et al., 2007).

The above development that applies mostly to medium-range, global ensemble prediction systems has been pioneered by ECMWF and NCEP who have both been running ensemble prediction systems operationally since the early 1990ies. Short-range, limited area ensemble prediction systems do not have such a long record of operational use, primarily because of the computational costs associated with running ensemble forecasts. In addition to its own initial condition and model uncertainty a limited area model is also constrained by shortcomings in the global host model that are fed to the nested limited area model through its lateral boundaries. The lateral boundary conditions may also give rise to a practical data transfer problem if separate boundary conditions are to be transferred for each ensemble member from one computing centre to another, e.g. from ECMWF to one of its member states.

A natural first approach to short-range, limited area ensemble prediction is to just downscale each member of a global ensemble using a nested limited area model. In that way both initial condition perturbations and lateral boundaries are handled by the global ensemble prediction system. The COSMO limited area ensemble prediction system (COSMO-LEPS) is one example using this approach (Marsigli et al., 2005). Here the host ensemble is ECMWF's ensemble prediction system, and the computional costs are reduced by only downscaling selected, representative members of the ECMWF ensemble (presently 16 members). The COSMO-LEPS system is run at ECMWF and so avoids the problems of transferring huge amounts of boundary data between computational centres. The target forecast lead time for COSMO-LEPS is the early medium-range (days 2-5).

Another approach is that of the GLAMEPS project (Grand limited area modelling ensemble prediction system; Iversen et al., 2008). Here the host ensemble is a special version of the ECMWF ensemble prediction system for which initial condition perturbations are generated using singular vectors that are targeted for domains in Europe (the so-called EuroTEPS model; Frogner and Iversen, 2008) and that are only evolved for 24h, (i.e. the perturbations favour modes that has the fastest growth in the first 24h) and so are better suited for short-range forecasts than the operational ECMWF ensemble prediction system for which the singular vectors are evolved for 48h. The downscaling of the EuroTEPS ensemble is done by running the nested limited area model with its

own analysis for the unperturbed control forecast and adding the EuroTEPS perturbations to this "control analysis" and using lateral boundary conditions from the EuroTEPS ensemble. GLAMEPS has not yet reached operational status, but recent tests using a horizontal resolution of the limited area model of approximately 13 km show results that are competitive with the ECMWF operational ensemble prediction system (Feddersen and Sattler, 2009; Iversen et al., 2009).

While model uncertainty has been addressed by the use of stochastic physics (Buizza et al., 1999) in the global ECMWF ensemble prediction system, there has been more variety in the types of model perturbations that have been applied in limited area ensemble prediction systems. Both *multi-scheme* (Bright and Mullen, 2002) and *stochastic parameter* (Li et al., 2008) approaches have been applied and in some cases combined to form *multi-model* (Hou et al., 2001) ensembles. NCEP's operational short-range ensemble forecasts are based on a combined multi-model, multi-scheme ensemble prediction system (Du et al., 2009).

The experimental short-range ensemble prediction system, using the HIRLAM model, that has been established at DMI uses initial condition perturbations based on scaled lagged average forecasts (Ebisuzaki and Kalnay, 1992). Model uncertainty is addressed by running ensemble members with either STRACO (Sass, 2002) or Kain-Fritsch/Rasch-Kristjansson (Kain, 2004; Rasch and Kristjansson, 1998) convection and condensation in combination with stochastic physics. Further tests include running ensemble members with different roughness lengths and additional experiments with different types of stochastic physics and different initial condition perturbations.

Although DMI – through the HIRLAM project – participates in the development of GLAMEPS, there are several good reasons for developing in parallel a DMI ensemble prediction system (DMI-EPS):

- With the arrival of DMI's Cray XT5 supercomputer in 2008 the time seemed ripe to introduce ensemble forecasting at DMI. A similar development is seen in several other European meteorological services.

- In GLAMEPS there has been much focus on initial condition perturbations and little on model uncertainty. In DMI-EPS the main focus is on model uncertainty. Being able to experiment locally with DMI-EPS is much more efficient than having to run remotely with the GLAMEPS setup at ECMWF, and by being involved in both projects DMI (and GLAMEPS) can hopefully "get the best of both worlds."

- While the first results from GLAMEPS are promising much remains to be done before operational GLAMEPS forecasts will be available.

- DMI can afford to run a high-resolution ensemble (~ 5 km horizontal resolution) for a relatively small domain, whereas GLAMEPS must cover a pan-European domain and so cannot afford to run in as high a resolution. The high resolution is desirable for modelling convective events, an important aspect of short-range ensemble forecasting.

The configuration of the experimental DMI ensemble prediction system is further described in Section 2, verification and comparison with ECMWF's operational ensemble prediction system and DMI's operational (deterministic) forecasts are described in Section 3, and case studies and issues related to the presentation of ensemble forecasts are discussed in Section 4. The relative impact of the various perturbations to initial conditions and model physic is discussed in Section 5, and

conclusions and thoughts about the development of ensemble prediction at DMI are given in Section 6.

2. Ensemble system configuration

The model setup is similar to that used in DMI's operational setup until May 2009 (Yang et al., 2005): A small-domain ("S05") HIRLAM model is nested into a large-domain ("T15") HIRLAM model (Fig. 2.1) which is nested into the global ECMWF model. The horizontal resolution of HIRLAM is 0.05° for the inner S05 model and 0.15° for the outer T15 model. Both S05 and T15 use 40 vertical levels. A HIRLAM 3D-Var analysis is run for T15, while S05 uses the T15 analysis interpolated to the S05 grid. An additional surface analysis is used as first guess in the initialization of the S05 model. Lateral boundaries for S05 are updated every forecast hour.

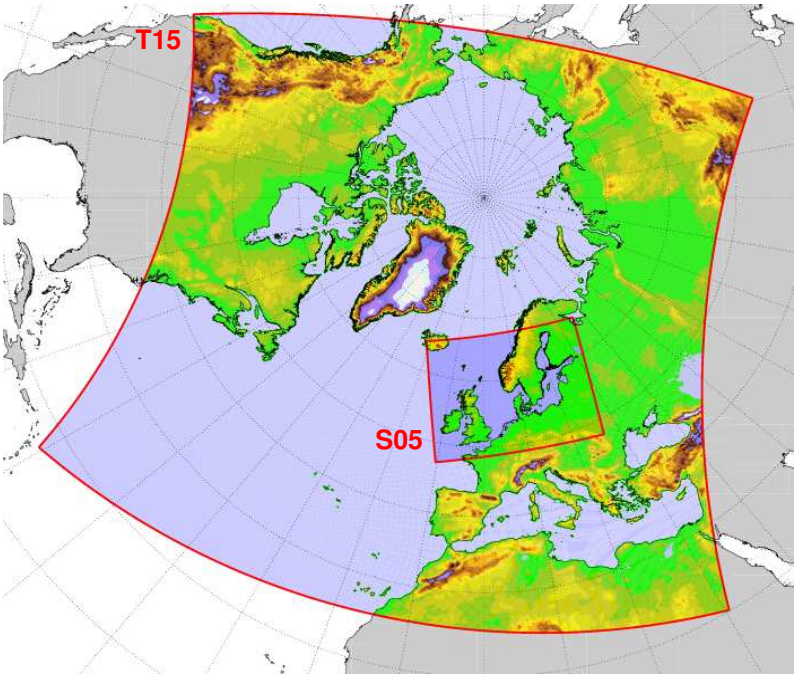


Figure 2.1*: T15 and S05 model domains.*

While the operational (deterministic) configuration used HIRLAM with certain DMI modifications, the ensemble configuration uses the reference version of the HIRLAM forecast model (version 7.2 with minor modifications). The main motivation for this is that in the reference version of HIRLAM there is an option to run with Kain-Fritsch (KF) convection and Rasch-Kristjansson (RK) condensation which is not available in DMI-HIRLAM.

The ensemble prediction system presently comprises 25 members that combine 5 different initial conditions, the two cloud schemes (STRACO and KF/RK), use of stochastic physics and perturbed roughness lengths for selected vegetation types, as illustrated in Table 2.1 and discussed in detail in sections 2.1-2.3.

2.1 Initial condition perturbations

From a research perspective the emphasis in DMI-EPS is on model uncertainty, but in order to have a useful ensemble prediction system initial condition uncertainty cannot be ignored. A simple method for perturbing the initial conditions that uses in-house data, and so does not require transfer of huge amounts of boundary data from a host ensemble prediction system, is the so-called scaled

Table 2.1: *Configuration of ensemble members 1-25. See text for details.*

Ensemble members	STRACO		KF/RK		STRACO
		Stoc.phys.		Stoc.phys	Pert.roughn.
Ini. cond. 1	1	6	11	16	21
Ini. cond. 2	2	7	12	17	22
Ini. cond. 3	3	8	13	18	23
Ini. cond. 4	4	9	14	19	24
Ini. cond. 5	5	10	15	20	25

lagged average forecasts (SLAF; Ebisuzaki and Kalnay, 1992). Here the forecast error (the difference between an old forecast and the most recent analysis) is multiplied by a scaling factor and added to or subtracted from the most recent analysis to provide a perturbed initial condition. The scaling factor controls the magnitude of the perturbation so that it becomes approximately independent of the forecast error, i.e. larger forecast errors of older forecasts are damped more than smaller forecast errors of recent forecasts. A first approximation is to assume linear forecast error growth, and consequently let the scaling factor decrease linearly with forecast age,

$$\text{initial condition} = \text{analysis} \pm \alpha_n \left(\text{forecast}_{n \text{ hours old}}(n) - \text{analysis}\right), \tag{2.1}$$

where n is the forecast age and the forecast length in hours, and α_n is the scaling factor. Presently,

$$\alpha_6 = 0.8, \quad \alpha_{12} = 0.7\,\alpha_6 \tag{2.2}$$

are used in DMI-EPS, and "analysis" and "forecast" are those of the T15 model interpolated to the S05 grid. If the linear error growth assumption had been used, α_{12} should have been half of α_6, but it was found that the forecast error of a 12h old forecast is in general less than twice that of a 6h old forecast, and so α_{12} is set to more than half of α_6. By using the 6h old and 12h old forecasts for initial condition perturbation we obtain four perturbed ensemble members in addition to the unperturbed, control forecast.

Perturbed lateral boundary conditions are generated similarly to the perturbed initial conditions (Hou et al., 2001) by simply replacing "analysis" with "forecast(t)" and "forecast(n)" with "forecast($n + t$)" in Eq. (2.1). Despite its simplicity, encouring results have been reported using the SLAF method (J.-A. Garcia Moya, personal communication).

We note that the SLAF method is really the first step towards the breeding method. In the latter the forecast error term in Eq. (2.1) is replaced by the difference between short-range forecasts from a positive and a negative perturbation. This difference is then suitably scaled and used as perturbation for the next cycle (Toth and Kalnay, 1997).

2.2 Physics perturbations

For each of the five initial conditions two runs are made: one using the STRACO scheme (Sass, 2002) and one using the Kain-Fritsch/Rasch-Kristjansson schemes (Kain, 2004; Rasch and

Kristjansson, 1998) for convection and condensation yielding a total of ten ensemble members. Each member is then run including also stochastic physics giving a total of twenty members.

The stochastic physics is of the "ECMWF-type" (Buizza et al., 1999) where the total physics tendencies for the three-dimensional model variables temperature, wind, humidity and cloud water are randomly perturbed. We can write the model equations, including the random perturbation, as

$$\dot{x}_j = A_j(\mathbf{x}, t) + P_j(\mathbf{x}, t) + r_j(\mathbf{x}, t)P_j(\mathbf{x}, t), \tag{2.3}$$

where the overdot denotes a time derivative, A_j is the dynamics, P_j the physics and r_j the random perturbation. x_j is T, u, v, q or cw, and $\mathbf{x}$ is a vector of all of the x_js.

The random perturbation is modelled by an autoregressive process, i.e.

$$r_j(t + T) = a\langle r_j(t)\rangle_D + \langle s_j\rangle_D. \tag{2.4}$$

Here T is the interval between updates of r_j, $\langle\cdot\rangle_D$ denotes a spatial average over a domain D, and s_j is a uniformly distributed random number. Values presently used in DMI-EPS are

$$T = 45 \text{ min}, \quad a = 0.9, \quad D = 53 \times 53 \text{ grid points}, \quad s_j \in U(-0.15; 0.15). \tag{2.5}$$

With a horizontal resolution of 0.05° the domain size is approximately 300×300 km^2. In order not to let the perturbation "run away" r_j is constrained to the range

$$r_j \in [-0.5; 0.5], \tag{2.6}$$

but normally the perturbations do not grow so big during a 36h forecast.

The autoregressive model ensures that the perturbations are smooth in time. Smoothness in space is obtained by smoothing the initial perturbation, although more spatial smoothing could be applied. There is no vertical variation in r_j. Figure 2.2 shows a realization of a random field at T+3h, T+18h and T+36h, while Fig. 2.3 shows time series from two neighbour domains.

With stochastic physics included one might expect an increased frequency of extreme events in the model ensemble. By inspecting the distributions of 2m temperature, 10m wind speed and 3h accumulated precipitation in the 90-day period 10 Aug - 7 Nov 2009 this does indeed appear to be the case for precipitation (data aggregated for selected Danish stations, see Section 3), while stochastic physics appears to have little impact on the model distribution of 2m temperature and 10m wind speed, see Fig. 2.4 (note the use of a logarithmic scale on the ordinate to enhance differences in the extremes). Inclusion of stochastic physics increases significantly the frequency of precipitation amounts of 12.8mm/3h or more.

The difference between the model distributions of 2m temperature and 10m wind speed is generally less than the difference between the observed distributions and either model distribution. The model has problems with small amounts of precipitation, regardless of the inclusion of stochastic physics. Cases of no precipitation are underrepresented in the model, whereas small precipitation amounts (less than 1.6mm/3h) are overrepresented in the model.

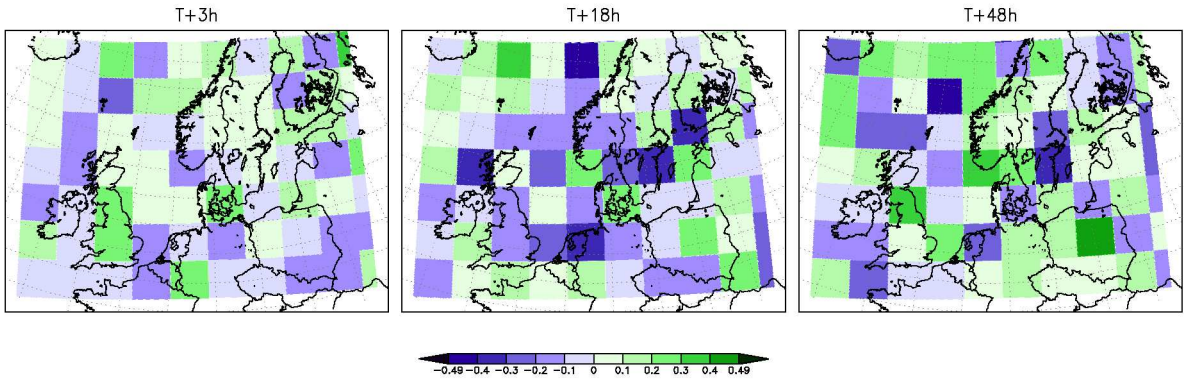


Figure 2.2: *Realization of random field used for stochastic perturbation of physics tendencies at times T+3h (left), T+18h (centre) and T+48h (right).*

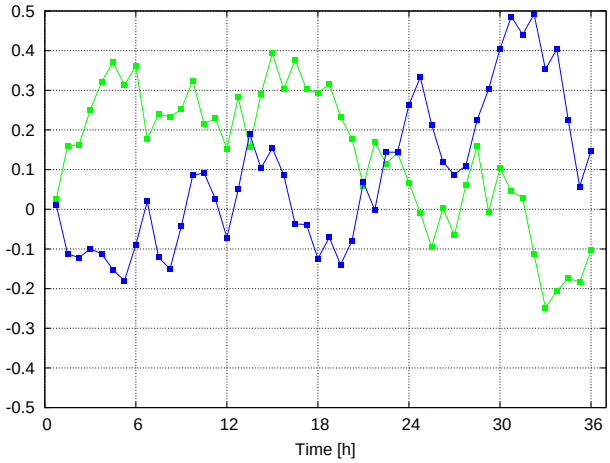


Figure 2.3: *Time series of random field at two neighbouring domains (approximately (11.5°E, 55.5°N) and (11.5°E, 53°N)).*

The model distributions are based on 30h forecasts from 0 and 12 UTC, which explains the double-hump in the 2m temperature distribution.

2.3 Experimental members

Five ensemble members have been dedicated to studying the impact of perturbing the roughness lengths for urban areas. At the start of a forecast a roughness length between 0.05 and 1.1 m is randomly chosen for each of the ensemble members 21-25. For members 1-20 the roughness length is set to 1 for urban areas. In addition, for members 21-25, stochastic perturbations are applied to the part of the physics tendencies that is due to convection and condensation, so the model perturbations for members 21-25 are more specific than the more general stochastic physics used in members 6-10 and 16-20.

2.4 Computational demands

Running ensemble forecasts is computationally demanding. But imagine a situation where a choice has to be made between increasing the horizontal resolution of a deterministic model, say by a factor 2 in both east-west and north-south direction and running an ensemble prediction system. The computational demand for running with the increased resolution would increase by approximately a factor 10 (2×2 more grid points, 2 times more time steps and a bit extra if the model is run on more

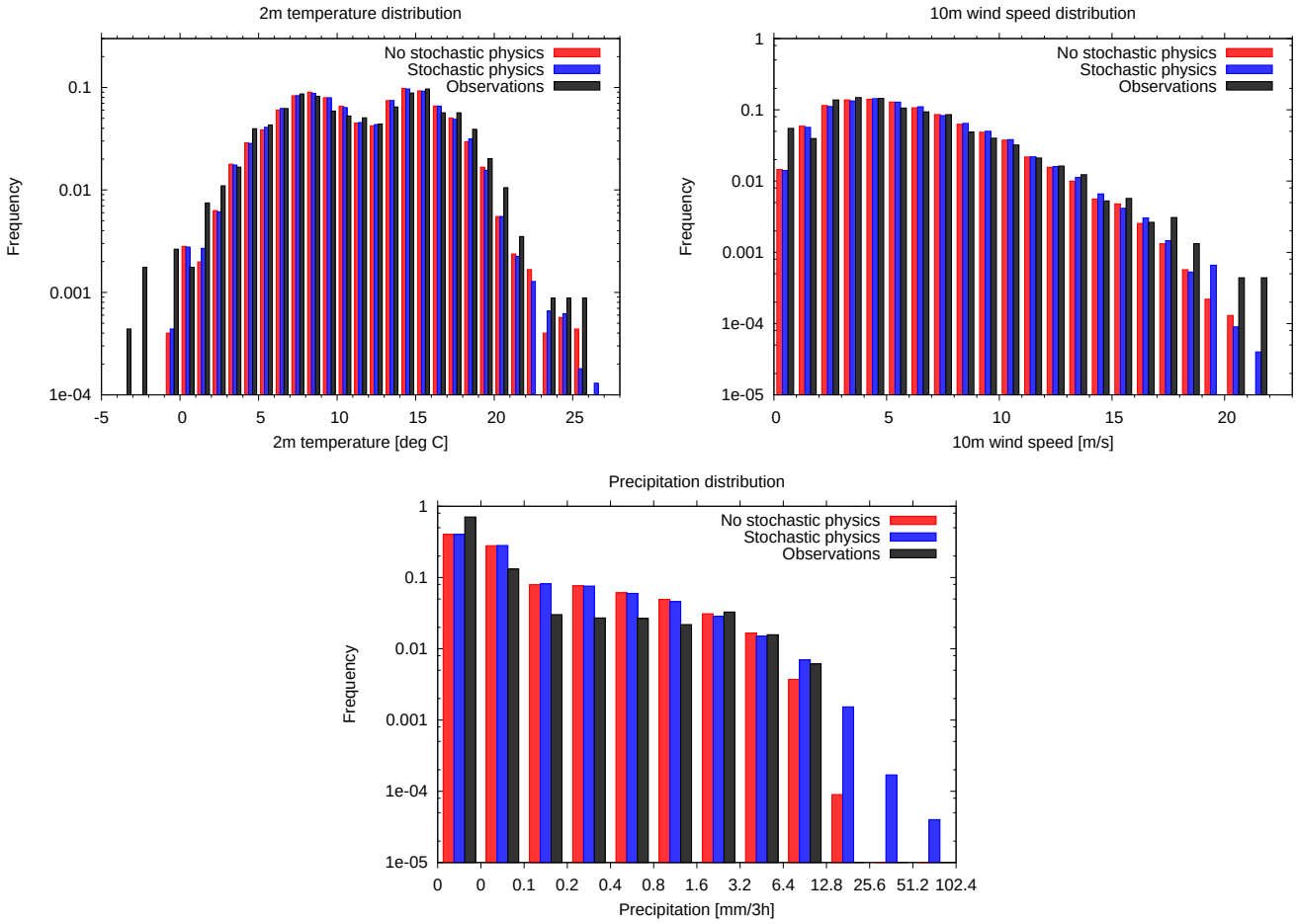


Figure 2.4: *Distribution of modelled (30h forecasts with/without stochastic physics, blue/red bars, respectively) and observed (black bars) 2m temperature (top left), 10m wind speed (top right) and 3h accumulated precipitation (bottom). Based on forecasts and observations from the 90-day period 10 Aug - 7 Nov 2009 for selected Danish stations, see Fig. 3.1.*

CPUs and the scaling is not perfect), so for the same computational costs we could run a 10-member ensemble – or even more if we are willing to sacrifice numerical "safety" as is done for DMI-EPS.

Presently, the operational S03 forecasts are run using 64 bit reals and integers, and the compilation is done using the relatively safe -O2 optimization. For DMI-EPS we use as default 32 bit reals and integers, and the compilation is done using the more aggressive -O3 optimization. The numerical uncertainty that these modifications introduce are assumed insignificant compared to the uncertainties that have been introduced by the perturbations of the initial and lateral boundary conditions and the perturbations of the model physics. The reduced accuracy and the increased optimization together reduce the runtime for a single 36h forecast by approximately 45%[1].

Each ensemble member uses a 12×13 domain decomposition and 4 compute cores for I/O or a total of 160 compute cores or 20 compute nodes on DMI's Cray XT5. A 36h forecast completes in approximately 9 min 25 sec when using the STRACO scheme and approximately 11 min 15 sec when using the KF/RK scheme. The typical wall clock time from the start of the preparation of initial and boundary condition perturbations until 25 ensemble members have completed is 40

[1]Based on a comparison between the run time of a 36h forecast run with reference-Hirlam compiled with -O3 optimization and 32 bit accuracy using the Pathscale compiler, and DMI-HIRLAM compiled with -O2 optimization and 64 bit accuracy using the PGI compiler. Run times may change for new versions of the compilers.

minutes with the present load on the operational cluster of DMI's Cray XT5 (256 compute nodes). For comparison the operational S03 model runs on 64 compute nodes and completes in approximately 38 min for a 54h forecast.

3. Ensemble forecast verification

For a deterministic forecast the forecast *quality* (or accuracy or skill) is always a measure of the "closeness" of the forecast and the verifying "truth." The fact that the verifying "truth" is always ambiguous is usually ignored, and this will also be the case in the following where observed values of temperature, wind speed and accumulated precipitation will be regarded as the truth, unless they are clearly faulty in which case the will be disregarded.

Measures of deterministic forecast quality include bias (average difference between forecast and observation), root mean square (RMS) error and various skill scores based on contingency tables, etc. See, e.g., Jollife and Stephenson (2003) for a review. A first approach for comparison of the quality of ensemble forecasts with the quality of deterministic forecasts is to simply use the ensemble mean or the ensemble median as a deterministic representative for the ensemble forecast and apply standard (deterministic) verification scores.

However, for ensemble forecasts forecast quality is not only a matter of how close the forecast is to the verifying observation, but also a matter of forecast *reliability*. The ensemble forecast should not only reflect the most likely outcome (such as the deterministic forecast), but also less likely outcomes, and it should be done in such a way that the number of members that predict a certain outcome reflects the true probability of this outcome. That is, in the long run an outcome that is predicted with a certain probability (fraction of ensemble members) should be observed with a matching frequency. This is the definition of forecast reliability. Reliability in itself does not ensure forecast quality. A forecast that samples a climatological distribution will be perfectly reliable, but has no skill beyond climatology. It has no *resolution*, i.e. the forecasts are not able to separate different outcomes.

In the following verification is done against observations at 12 Danish stations and Thorshavn at the Faroe Islands, see Fig. 3.1.

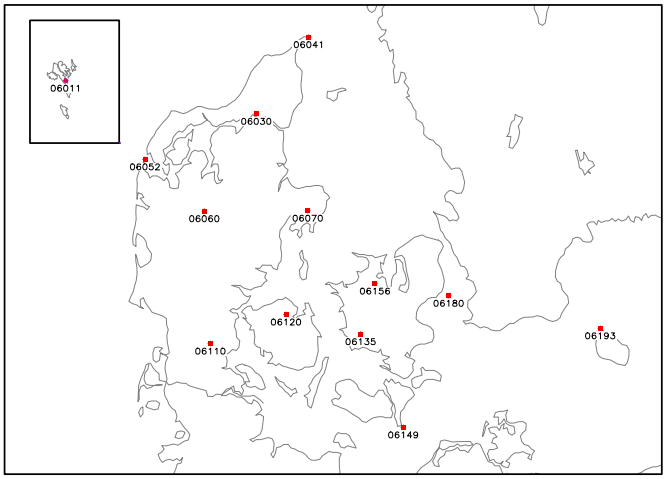


Figure 3.1*: Stations used for verification. Inset shows Faroe Islands.*

3.1 Ensemble mean scores

We may regard any component of the weather as composed of a predictable *signal* plus some unpredictable *noise*. Likewise, a deterministic forecast will be a combination of signal and noise. If we have an ensemble of *a priori* equally likely forecasts then taking the ensemble mean will filter

out some of the noise and should, consequently, verify better in the long run than any individual forecast. In practice, the ensemble members are not equally likely. The control forecast will *a priori* be likely to verify better than the perturbed forecasts. Nevertheless, taking the ensemble mean should filter out some noise and possibly verify better than the control forecast.

Figure 3.2 shows bias and RMS error for 2m temperature and 10m wind speed for the ensemble mean of the DMI ensemble compared to the control forecast (ensemble member 1: unperturbed forecast using STRACO scheme), the operational S03 forecast and the ensemble mean and control forecast of the 51 member ECMWF-EPS. The verification period here and in the following is the 90-day period 10 August - 7 November 2009 for the stations shown in Fig. 3.1.

DMI's ensemble provides the best scores both in terms of bias and RMS error for both 2m temperature and 10m wind speed. For precipitation the ECMWF ensemble median is marginally better than the DMI ensemble median. Note the ECMWF model's substantial positive bias for 10m wind speed that is not present in the Hirlam model.

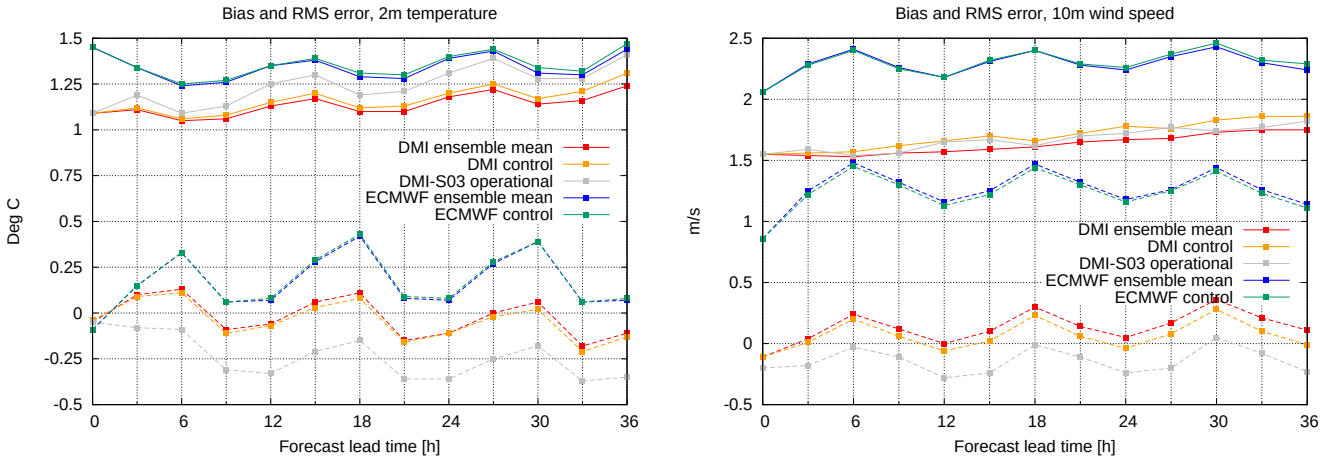


Figure 3.2*: RMS error and bias (dashed) for 2m temperature (left) and 10m wind speed (right) for DMI ensemble mean (red), DMI control (orange), DMI operational S03 (gray), ECMWF ensemble mean (blue) and ECMWF control (green).*

Precipitation data is characterized by much more scatter than 2m temperature and 10m wind speed data, and so the ensemble median is a more robust estimate of the most likely forecast than the ensemble mean. Figure 3.3 shows comparable quality of the DMI and ECMWF ensemble medians in terms of bias and RMS error for 3h accumulated precipitation. It is evident that the ensemble median scores better than the control forecast. One may speculate that the rather large RMS errors seen for the DMI control and operational forecasts are caused by the high resolution HIRLAM model generating heavy localized precipitation that is less likely to be generated in the coarser resolution ECMWF model.

3.2 Rank histograms

If all ensemble members *a priori* are equally likely, and the verifying observation is indistinguishable from the ensemble members (in a statistical sense), then the ensemble forecast is reliable. A widely used way in which to test whether this is the case is to rank the verifying observations relative to the sorted ensemble members with the lowest rank (1) assigned to observations that are less than the smallest ensemble member and the highest rank ($m + 1$ for an m-member ensemble) to observations that are greater than the largest ensemble member. For a reliable forecast a histogram of the ranks (so-called rank histogram or Talagrand histogram; Strauss

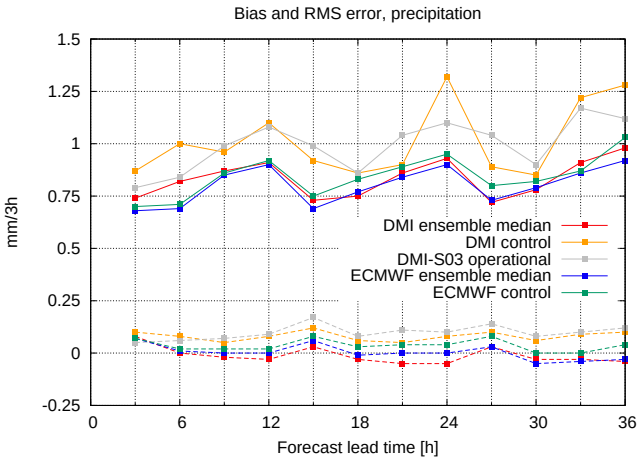


Figure 3.3: *RMS error and bias (dashed) for precipitation for DMI ensemble mean (red), DMI control (orange), DMI operational S03 (gray), ECMWF ensemble mean (blue) and ECMWF control (green).*

and Lanzinger, 1995) will be approximately flat. Note that the opposite is not necessarily true: a flat rank histogram does not guarantee forecast reliability. Rank histograms are very commonly U-shaped indicating that observations tend to fall outside the ensemble range. That is, the ensemble spread is too small to capture the expected number of observations (due to the finite ensemble size we expect a small fraction $(2/(m+1))$ of the observations to fall outside the ensemble range, even for reliable ensembles).

Figure 3.4 shows typical rank histograms for DMI-EPS and ECMWF-EPS. The U-shape is found for both 2m temperature and 10m wind speed, although for the latter the positive bias in the ECMWF model "shifts" the rank histogram towards lower ranks, so that the lowest rank is greatly over-represented. Precipitation also has a moderate over-representation of the lowest rank. This frequently happens when no precipitation is observed, and all ensemble members predict some precipitation.

A summary of the rank histograms is also shown in Fig. 3.4 in terms of the fraction of observations that are captured by the ensemble. Ideally, this capture rate should be $(m-1)/(m+1)$ for an m-member ensemble or approximately 92% for a 25-member ensemble and 96% for a 51-member ensemble.

The comparison between DMI-EPS and ECMWF-EPS is overall favourable for DMI-EPS, particularly for precipitation where the ensemble capture rate is higher for DMI-EPS for all lead times.

3.3 Spread/error relation

In an ideal ensemble prediction system the ensemble spread is an indicator of forecast uncertainty of a deterministic forecast, e.g. the control forecast or the ensemble mean (or median). Small spread indicates that the deterministic forecast is certain, and large spread that it is uncertain. Figure 3.5 shows typical scatter plots of the RMS error of the ensemble mean (median) versus the ensemble spread (ensemble spread = standard deviation of the ensemble members). Ideally, the RMS error should be small when the ensemble spread is small, but in reality this is, at best, the case only in a statistical sense. For individual ensemble forecasts the error can be quite large even if the ensemble spread is small as indicated by the error bars.

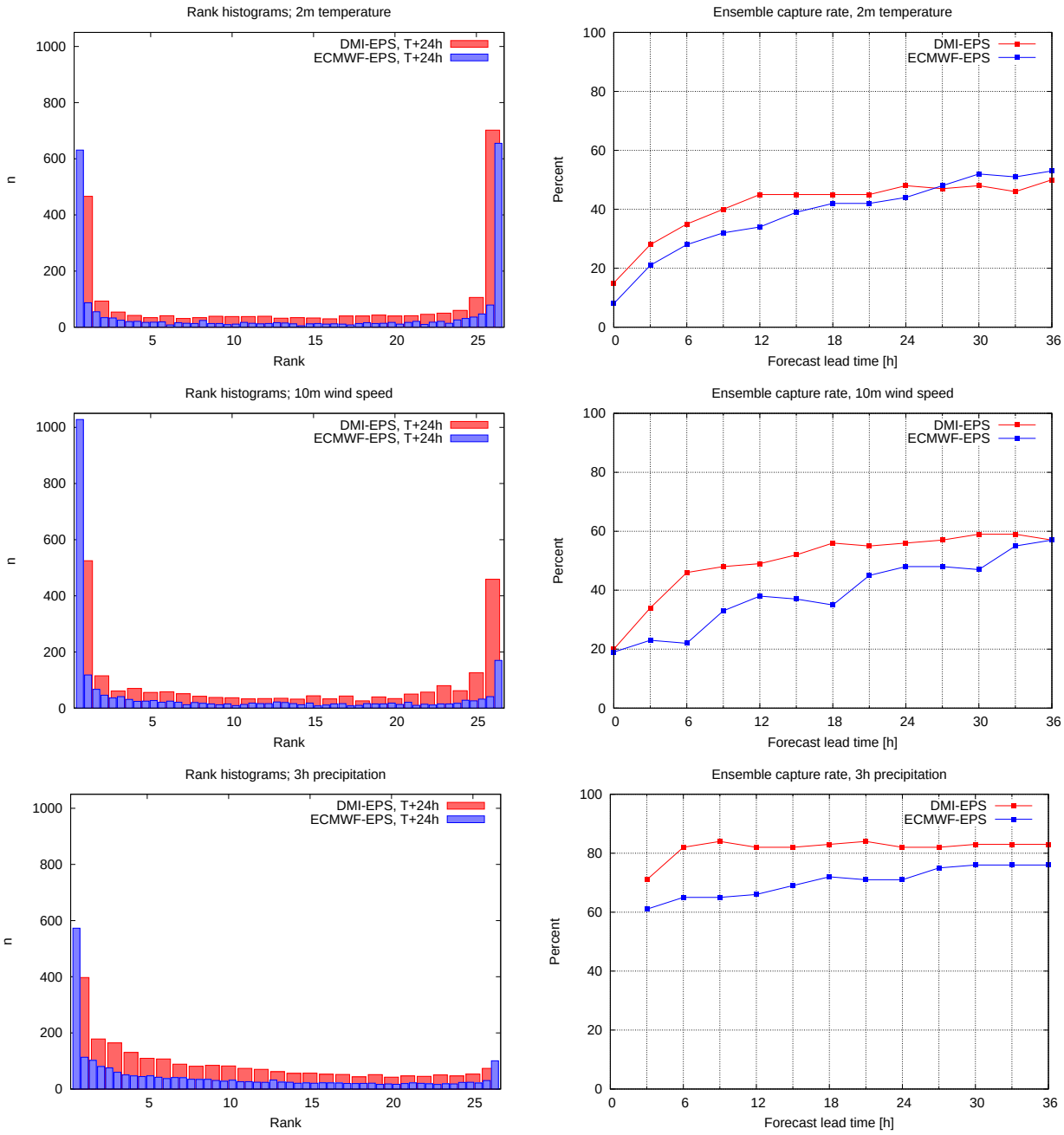


Figure 3.4: *Rank histograms (left) for 24h forecasts of 2m temperature (top), 10m wind speed (centre) and 3h accumulated precipitation (bottom); ensemble capture rates (right), i.e. the fraction of observations captured by the ensemble.*

We note that for both 2m temperature and 10m wind speed the ensemble spread is generally smaller than the RMS error of the ensemble mean, and the correlation between spread and error is small. For precipitation the relation between spread and error looks better, although for the ECMWF ensemble the error tends to exceed the spread. Low ensemble spread will often happen when there is no or little precipitation in both forecast and observations

3.4 Brier score and reliability

It is conceptually easy to verify binary events, such as “rain” or “no rain” or 2m temperature above or below 20°C. For probabilistic forecasts, including those based on ensemble prediction systems,

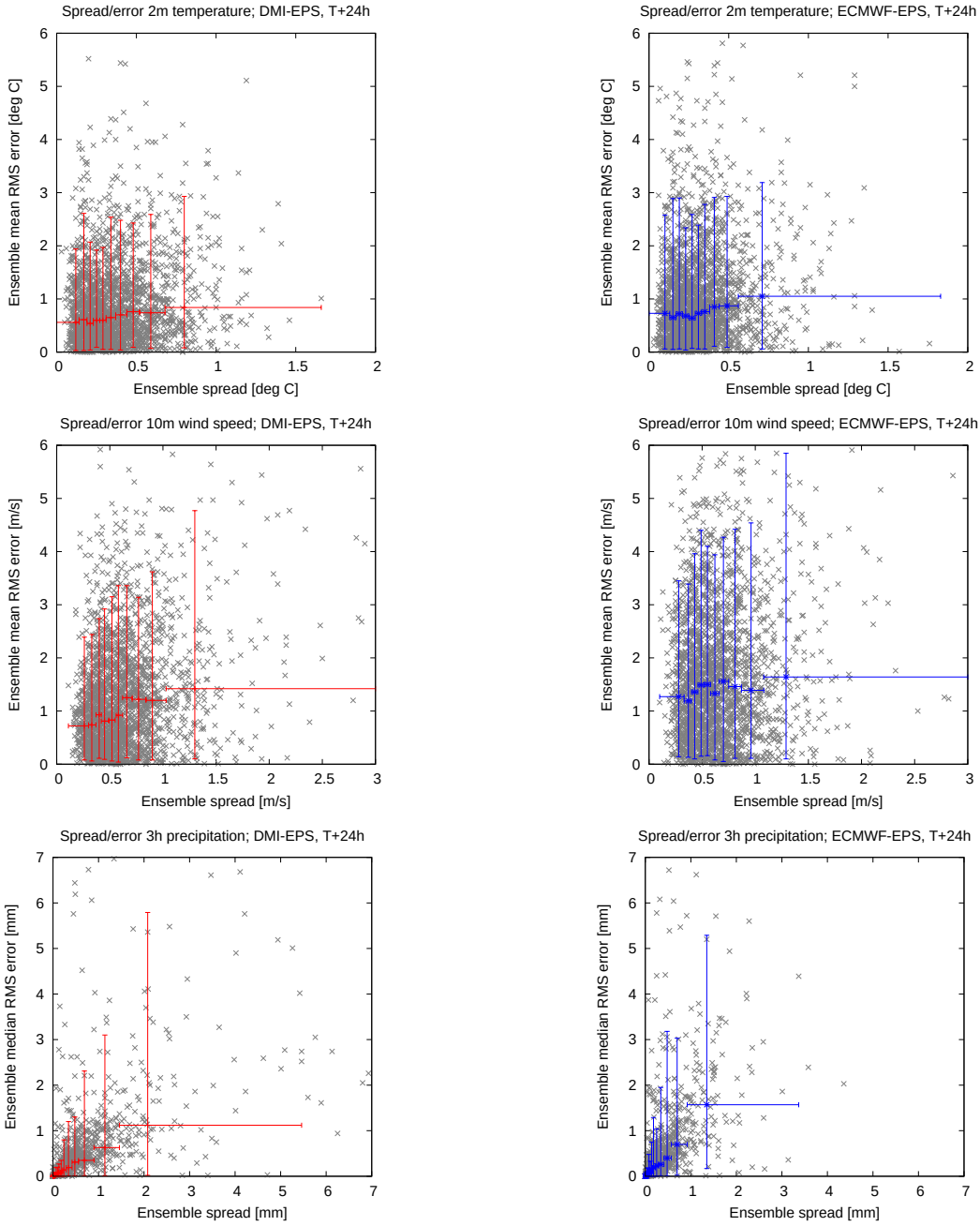


Figure 3.5: *RMS error of ensemble mean (median) versus ensemble spread for 24h forecasts for DMI-EPS (left) and ECMWF-EPS (right) and 2m temperature (top), 10m wind speed (centre) and 3h accumulated precipitation (bottom). Error bars show 90% confidence intervals for RMS error (vertical bar) for associated ensemble spread interval (horizontal bar).*

the Brier score (Brier, 1950) is among the most widely used. The Brier score is simply the mean squared probability error,

$$B = \frac{1}{n}\sum_{i=1}^{n}(p_i - o_i)^2, \tag{3.1}$$

where n is the total number of forecast/observation pairs, p_i is the forecast probability that the binary event will happen (for an ensemble forecast: the fraction of ensemble members that predict the

event), and o_i is 1 if the event actually did happen and 0 if the event did not happen. Thus, the smaller the Brier score, the better. The perfect score, 0 is obtained for a perfect, deterministic forecast for which the probability is always 1 if the event happens and 0 if the event does not happen. Note that rare events by construction will tend to score better than frequent events, so the Brier score for different events should not be compared. But a comparison between models for the same event and the same verification period shows which model is better in terms of Brier score. Figure 3.6 shows that DMI-EPS consistently scores better than ECMWF-EPS. This conclusion also holds for other thresholds (not shown).

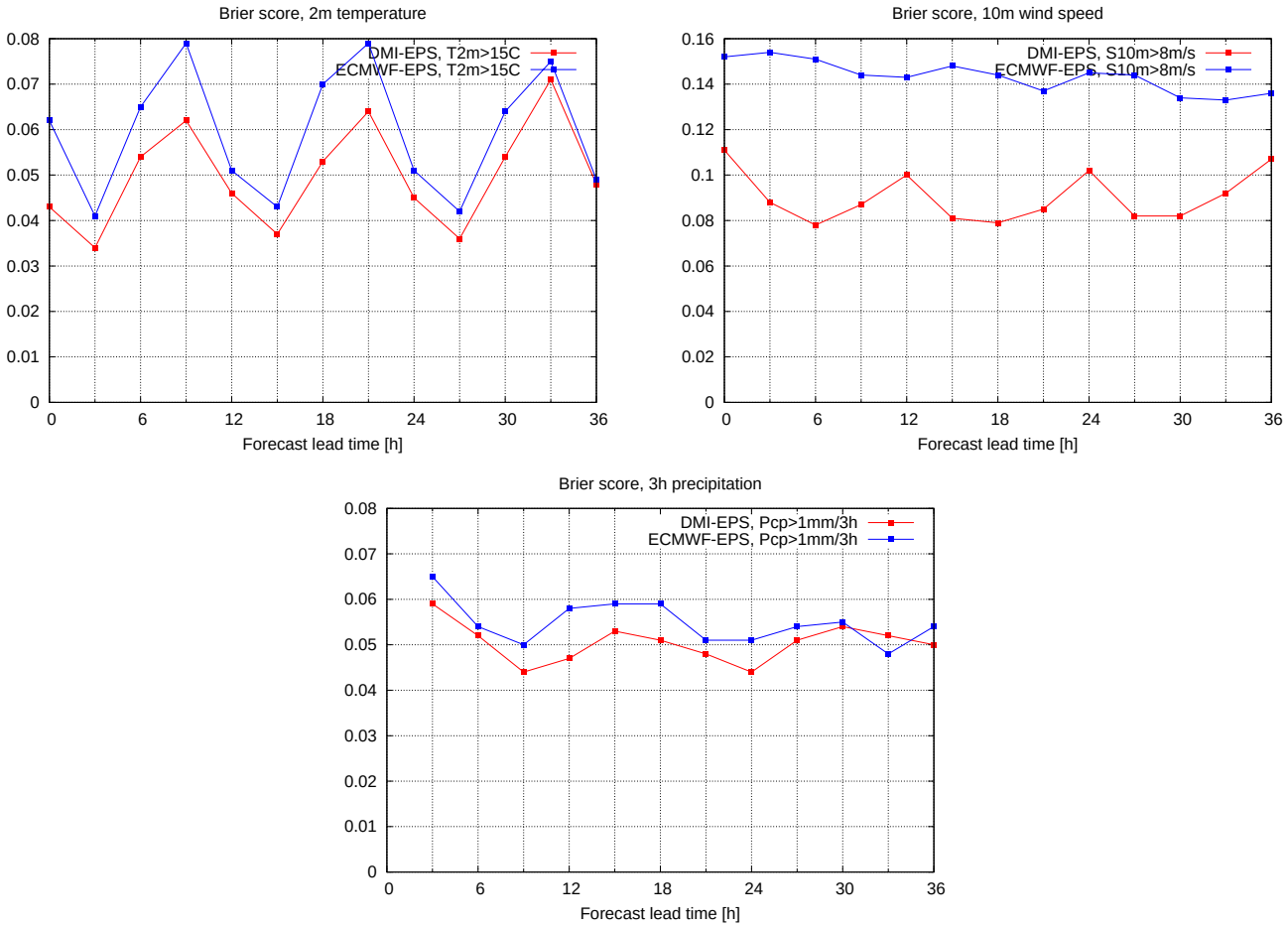


Figure 3.6*: Brier scores for 2m temperature > 15°C (top left), 10m wind speed > 8m/s (top right) and precipitation > 1mm/3h (bottom) for DMI-EPS (red) and ECMWF-EPS (blue). Lower scores are better.*

The Brier score can be decomposed into the sum of three terms: reliability, resolution and uncertainty (Murphy, 1973). Here we shall only be concerned with the reliability term which is the mean squared difference between forecast probability and observed frequency conditioned on the forecast probability of the event. Reliability is conveniently illustrated in a reliability diagram where the observed frequency of the event is plotted against the forecast probability. For a perfectly reliable forecast system the observed frequency should match the forecast probility. For example, for all forecasts that predict 70% chance of a certain event, that event should ideally happen in 70% of the verifying observations. It follows that the reliability curve for a perfectly reliable forecast lies along the diagonal in the reliability diagram. Figure 3.7 shows examples of reliability diagrams. They all suffer more or less from sampling problems, i.e. in most cases all ensemble members agree on the prediction of the event, and so there are relatively few cases where the forecast probability is not close to 0 or 1.

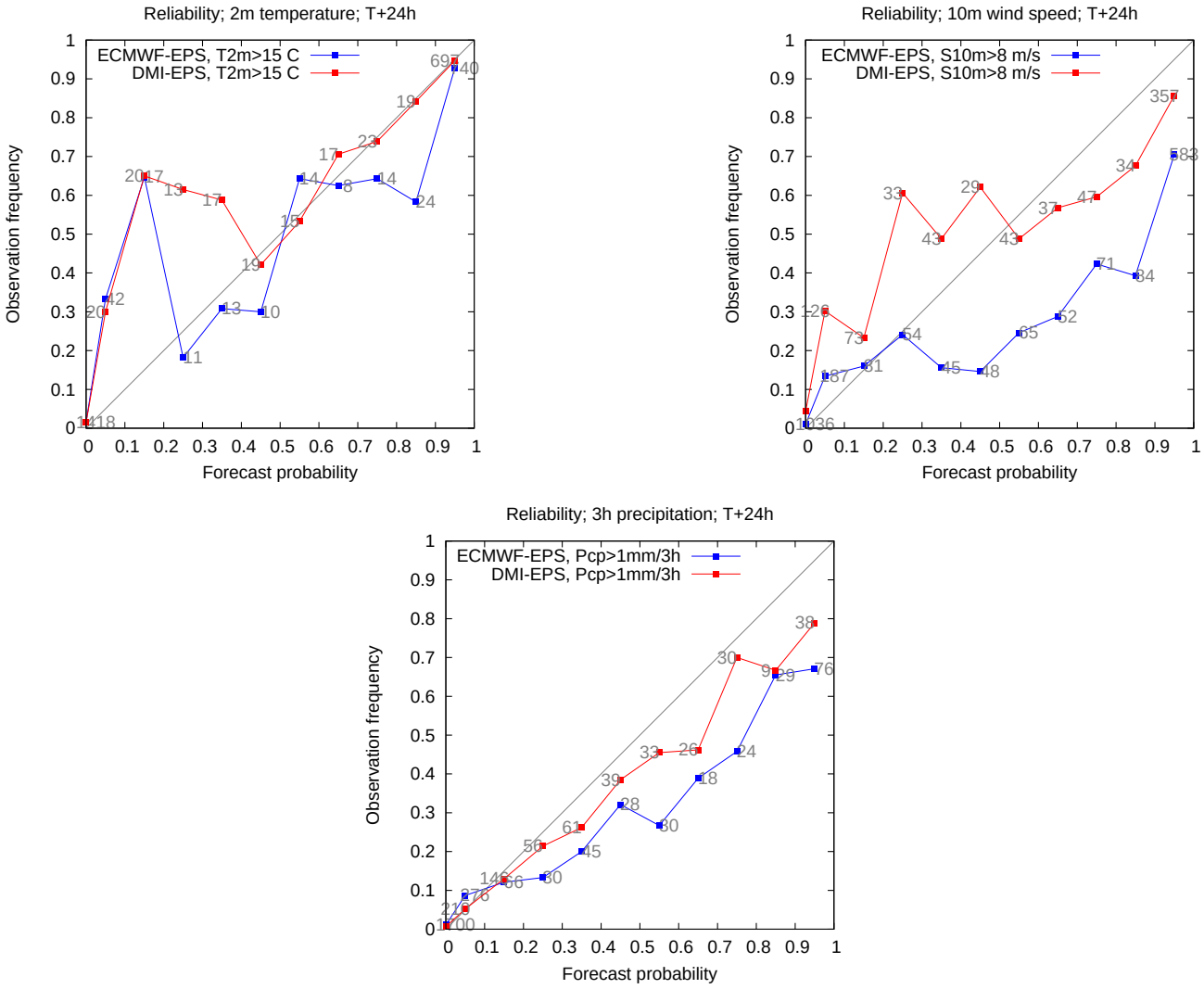


Figure 3.7: *Reliability diagrams for 24h forecasts of 2m temperature > 15°C (top left), 10m wind speed > 8m/s (top right) and precipitation > 1mm/3h (bottom) for DMI-EPS (red) and ECMWF-EPS (blue). Number of data points in each forecast probability bin is indicated on curves.*

3.5 Ranked probability score

The ranked probability score is a generalization of the Brier score that includes multiple categories (Jollife and Stephenson, 2003),

$$\mathrm{RPS} = \frac{1}{nK}\sum_{i=1}^{n}\sum_{k=1}^{K}(p_{i,k} - o_{i,k})^2 = \frac{1}{K}\sum_{k=1}^{K} B_k, \tag{3.2}$$

where $p_{i,k}$ is the forecast probability that the forecast variable $x > x_k$ where $x_1, x_2, ..., x_K$ are predefined thresholds that are ranked such that $x_1 < x_2 < \cdots < x_K$. The ranked probability score is simply the average Brier score for the K thresholds.

We use the following thresholds ($K = 5$): 0, 5, 10, 15, 20°C for 2m temperature; 3, 7, 12, 16, 20 m/s for 10m wind speed, and 0.1, 1, 5, 10, 15 mm/3h for precipitation. Figure 3.8 shows that DMI-EPS scores consistently better than ECMWF-EPS.

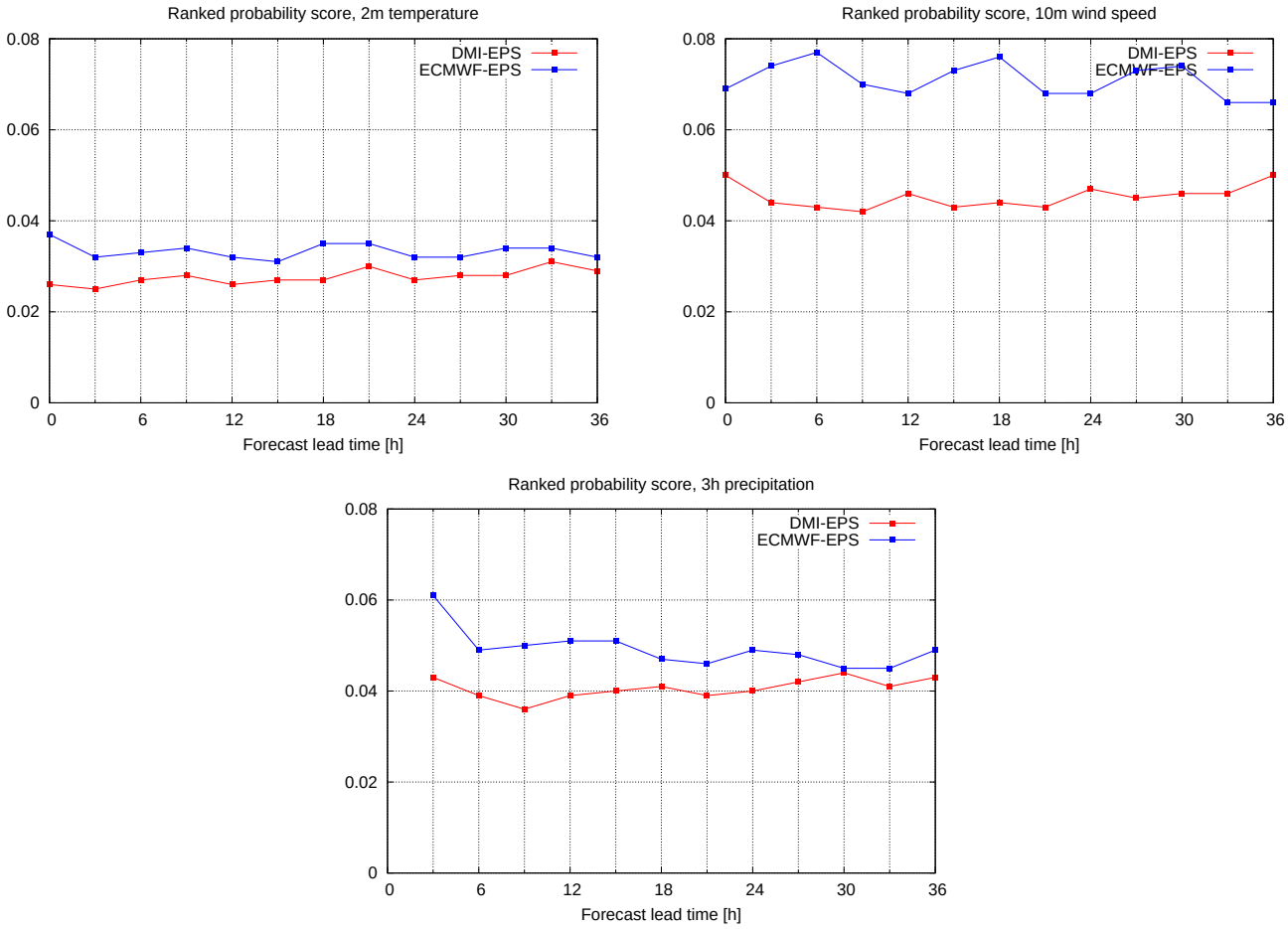


Figure 3.8: *Ranked probability scores for 2m temperature (top left), 10m wind speed (top right) and precipitation (bottom) for DMI-EPS (red) and ECMWF-EPS (blue). Lower scores are better.*

3.6 Relative operating characteristic

The relative operating characteristic (ROC) is another skill score that is based on a binary event. The so-called ROC curve is a plot of hit rate versus false alarm rate. A hit is recorded if i members of the ensemble correctly forecast the event, where $i = 1, ..., m$ for an m-member ensemble. The hit rate and false alarm rate are defined as

$$\text{Hit rate} = \frac{\text{events correctly forecast}}{\text{events occurred}} \tag{3.3}$$

and

$$\text{False alarm rate} = \frac{\text{events falsely forecast}}{\text{events non-occurred}} \tag{3.4}$$

Both hit rate and false alarm rate decrease with increasing values of i, and the ROC curve is normally extended to the limiting points (0,0) and (1,1). The closer the ROC curve is to the point (false alarm rate, hit rate)=(0,1) the better. The diagonal where hit rate equals false alarm rate marks the limit for a skillful ensemble forecast. A common measure of forecast skill is the area under the ROC curve; the closer to 1 the better. Figure 3.9 shows ROC curve examples. An attempt has been made to fit an empirical function

$$f(x;p,w) = wx^{1/p} + (1-w)(1-(1-x)^p) \tag{3.5}$$

to the data points. Provided this function gives a reasonable fit (with p positive and $0 < w < 1$), it may be useful when calculating the area under the ROC curve as the area is easily found as

$$\text{ROC area} = \int_0^1 f(x;p,w)dx = \frac{p}{p+1} \tag{3.6}$$

Alternatively, the ROC area can be calculated by simply connecting the data points by straight lines and adding the areas of each of the resulting trapezoids. In any case, when the data points do not span the whole range of false alarm rates or hit rates the use of the ROC area as a measure of forecast skill becomes ambiguous.

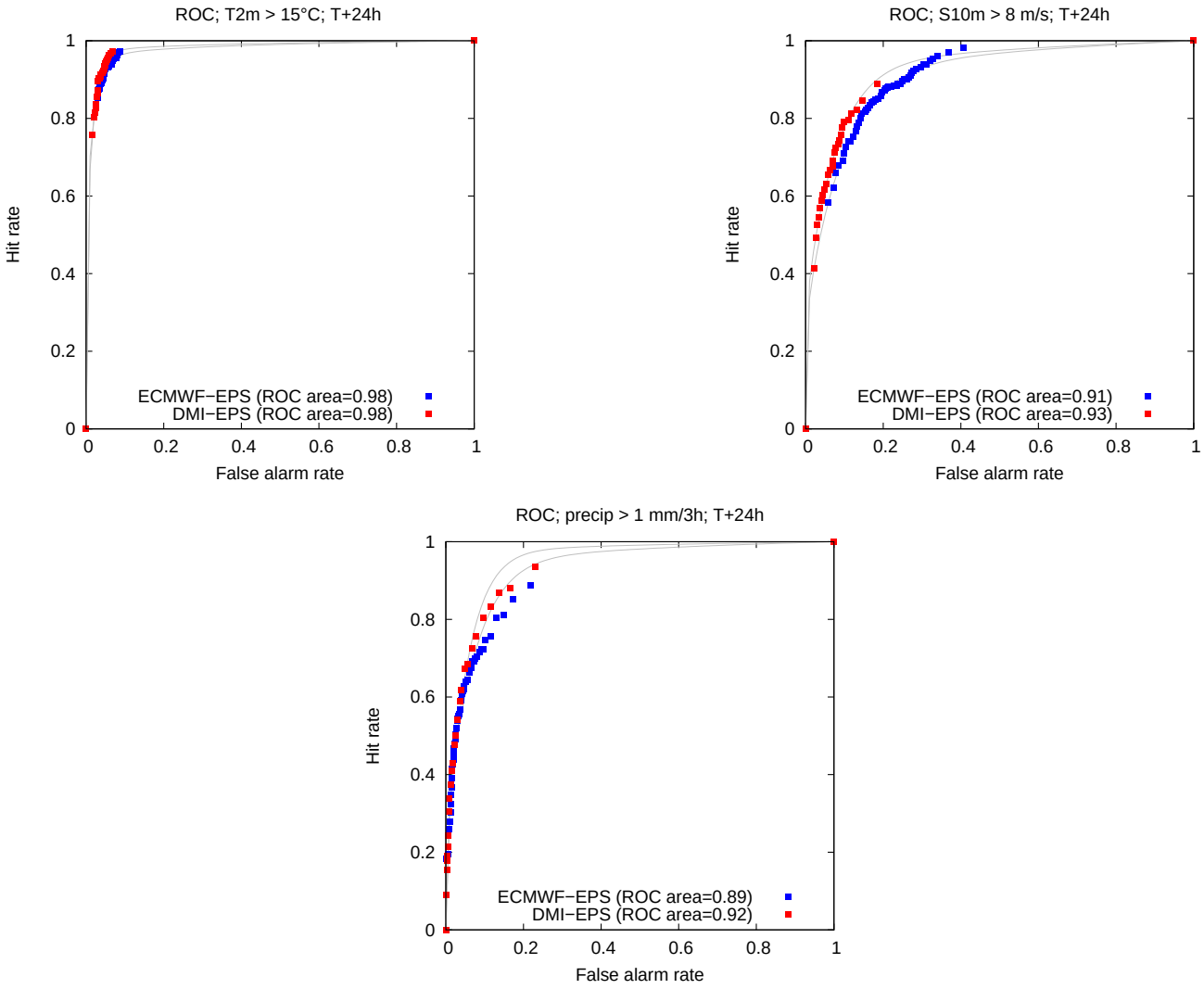


Figure 3.9: *ROC curves for 24h forecasts of 2m temperature > 15°C (top left), 10m wind speed > 8 m/s (top right) and precipitation > 1 mm/3h (bottom) for DMI-EPS (red) and ECMWF-EPS (blue). Fits to the data points are shown in gray. The calculated ROC areas are based on the fitted functions for 2m temperature and 10m wind speed and on the trapezoidal approach for precipitation (see text).*

3.7 Economic value

Imagine that a particular weather event will lead to a certain loss if it occurs, unless some precautionary (and possible costly) action is taken. A decision maker will need to decide whether to prevent the possible loss by taking action at a certain cost, or whether to accept the loss if the event occurs. With knowledge of the climatological conditions the decision maker will either always take precautionary action or will always accept the loss when the event occurs, whatever leads to the smallest expense in the long run. Whether the decision is one or the other depends on the cost/loss ratio.

When a weather forecast is also taken into account the expense can be further reduced depending on the skill of the forecast. It can be shown (see Jolliffe and Stephenson, 2003) that the expense depends on the hit rate and false alarm rate. With access to an ensemble forecast the hit rate and false alarm rate depend on how many ensemble members should forecast the event before it counts as a "hit", as discussed for the ROC curve. By picking the optimal threshold for a "hit" the expense can be further reduced.

The (relative) economic value is defined as the reduction in expense relative to the reduction that would be obtained with a perfect forecast. Thus, a forecast is skillful when the value is positive; the maximum value is 1 for a perfect forecast. Figure 3.10 shows values for optimal "hit thresholds" as a function of the cost/loss ratio for the same events that were used for the ROC curves in Fig. 3.9. The comparison between DMI-EPS and ECMWF-EPS shows some differences. For 2m temperature DMI-EPS is generally slightly more valuable than ECMWF-EPS; for 10m wind speed ECMWF-EPS is better for very low cost/loss ratios (probably related to the positive bias for the ECMWF-EPS forecasts) while DMI-EPS is better otherwise; for precipitation DMI-EPS is performing slightly better than ECMWF-EPS, except for high cost/loss ratios.

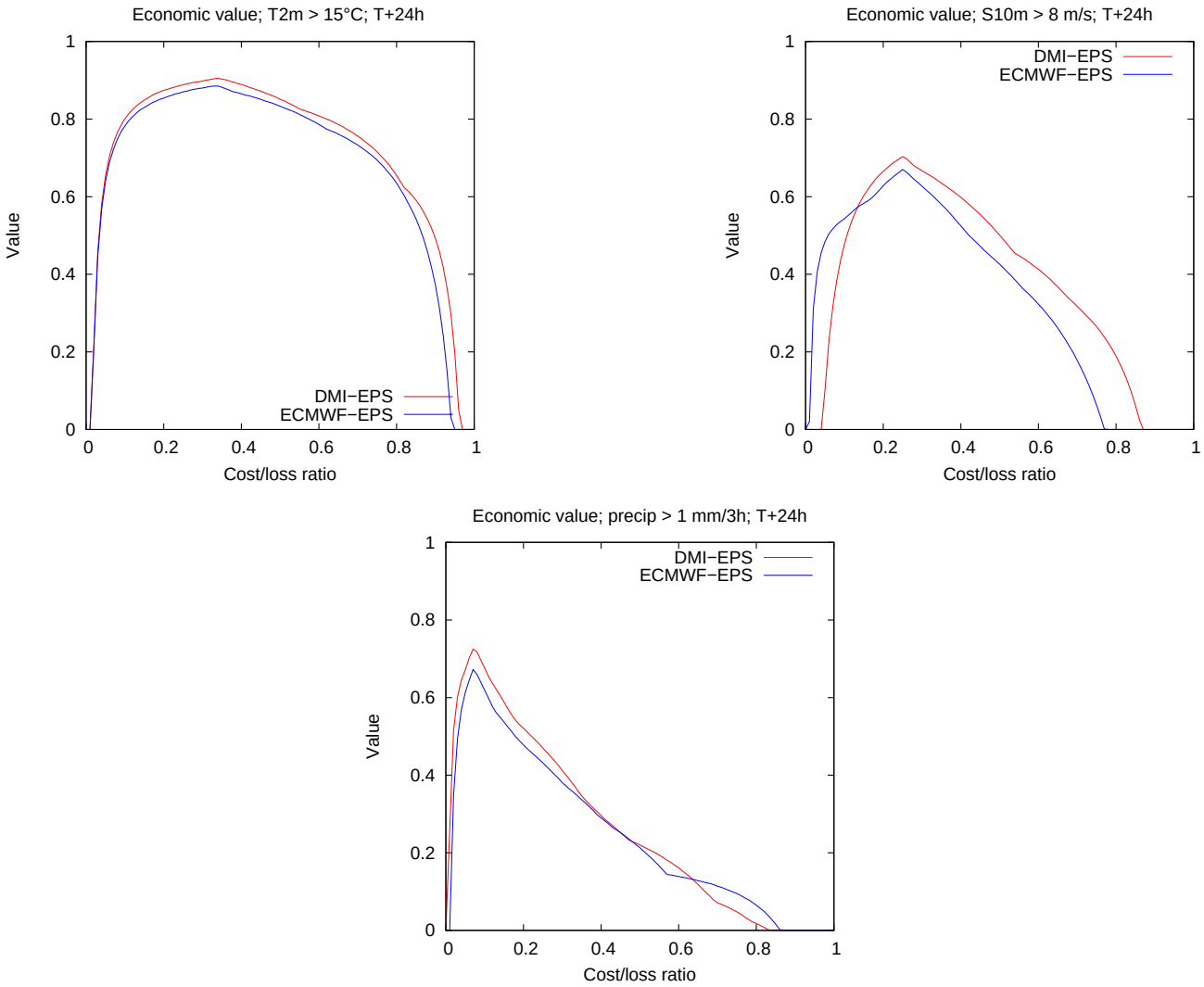


Figure 3.10*: Relative economic value for 24h forecasts of 2m temperature > 15°C (top left), 10m wind speed > 8 m/s (top right) and precipitation > 1 mm/3h (bottom) for DMI-EPS (red) and ECMWF-EPS (blue). Higher values are better.*

4. Ensemble forecast presentation: case studies

With the enormous amounts of output from ensemble prediction systems there is a challenging need to present the forecasts in a condensed form. Here we shall discuss different types of commonly used plots. The plots will be illustrated using case studies of rainfall in Denmark on 20090820 and wind speeds and mean sea level pressure from the Danish storm on 19991203. Figure 4.1 shows the operational S03 forecast from 2009082012 of precipitation accumulated from forecast hour 6 to 12 and from hour 12 to 18. The numbers on the plot show the corresponding observed precipitation. The rainfall in the north-western part of Jutland is well captured by the forecast (although the model predicts somewhat more rain than was actually observed), but the forecast fails to capture the rainfall further east, in particular we note that 12+2 mm rain was observed in Tirstrup, but less than 0.5 mm was forecast by the S03 model.

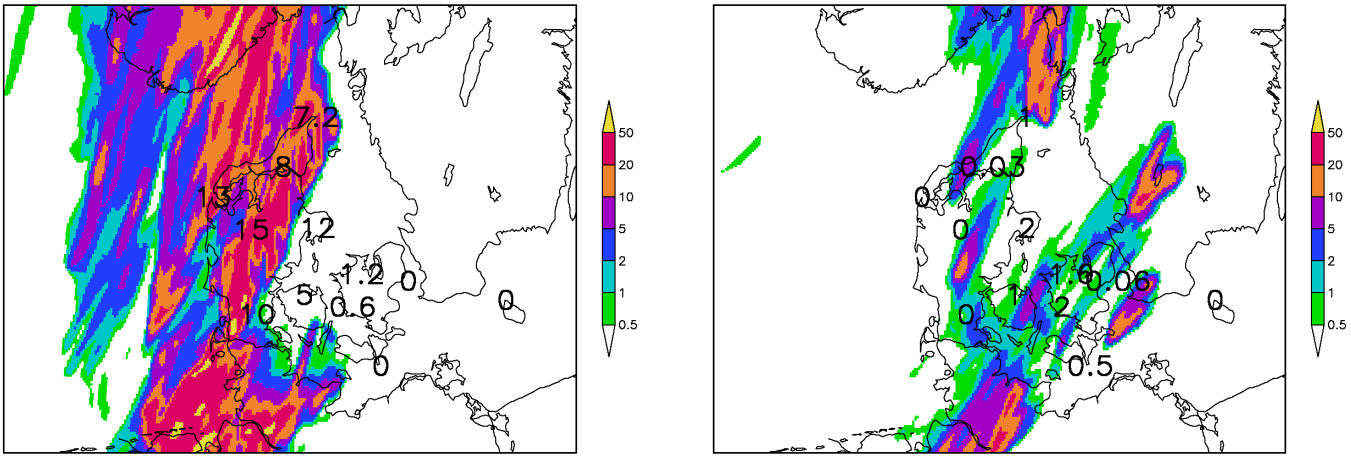


Figure 4.1: *S03 operational forecast from 2009082012: Precipitation accumulated from forecast hour 6 to 12 (left) and from hour 12 to 18 (right). Numbers show the corresponding observed precipitation in mm.*

4.1 Stamp map

The ensemble forecast's ability to capture the observed precipitation is illustrated by plotting the precipitation forecast for each member of the ensemble in a "postage stamp" map, see Figs. 4.2-4.3. The members agree on a band of rain near western Denmark between 2009082018 and 2009082100 (Fig. 4.2), although there is some uncertainty regarding how far east it will rain. Note also the spread in intensity between the members; there appears to be a tendency to areas of more intense rainfall for the members that include stochastic physics.

For the following 6h (Fig. 4.2) there is more spread in the spatial rainfall pattern.

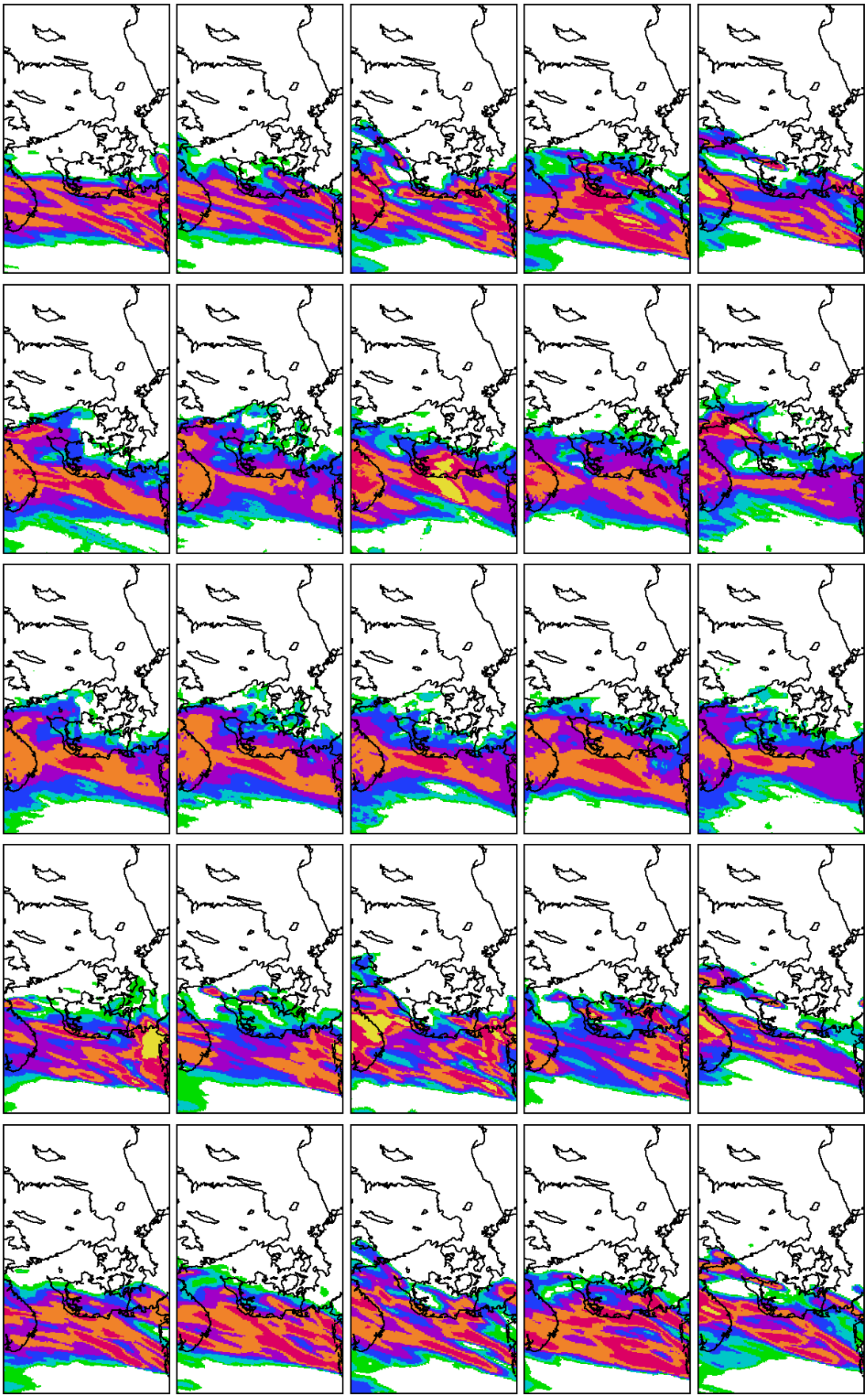


Figure 4.2: *25 ensemble members showing forecast from 2009082012: Precipitation accumulated from forecast hour 6 to 12. Each row share the same initial and lateral boundary conditions; first column uses the STRACO condensation scheme, second column uses STRACO and stochastic physics, third column uses KF/RK convection and condensation, fourth column uses KF/RK and stochastic physics, fifth column uses STRACO, stochastic physics (only applied to tendencies from STRACO) and perturbed vegetation roughness lengths. Colour scale as in Fig. 4.1.*

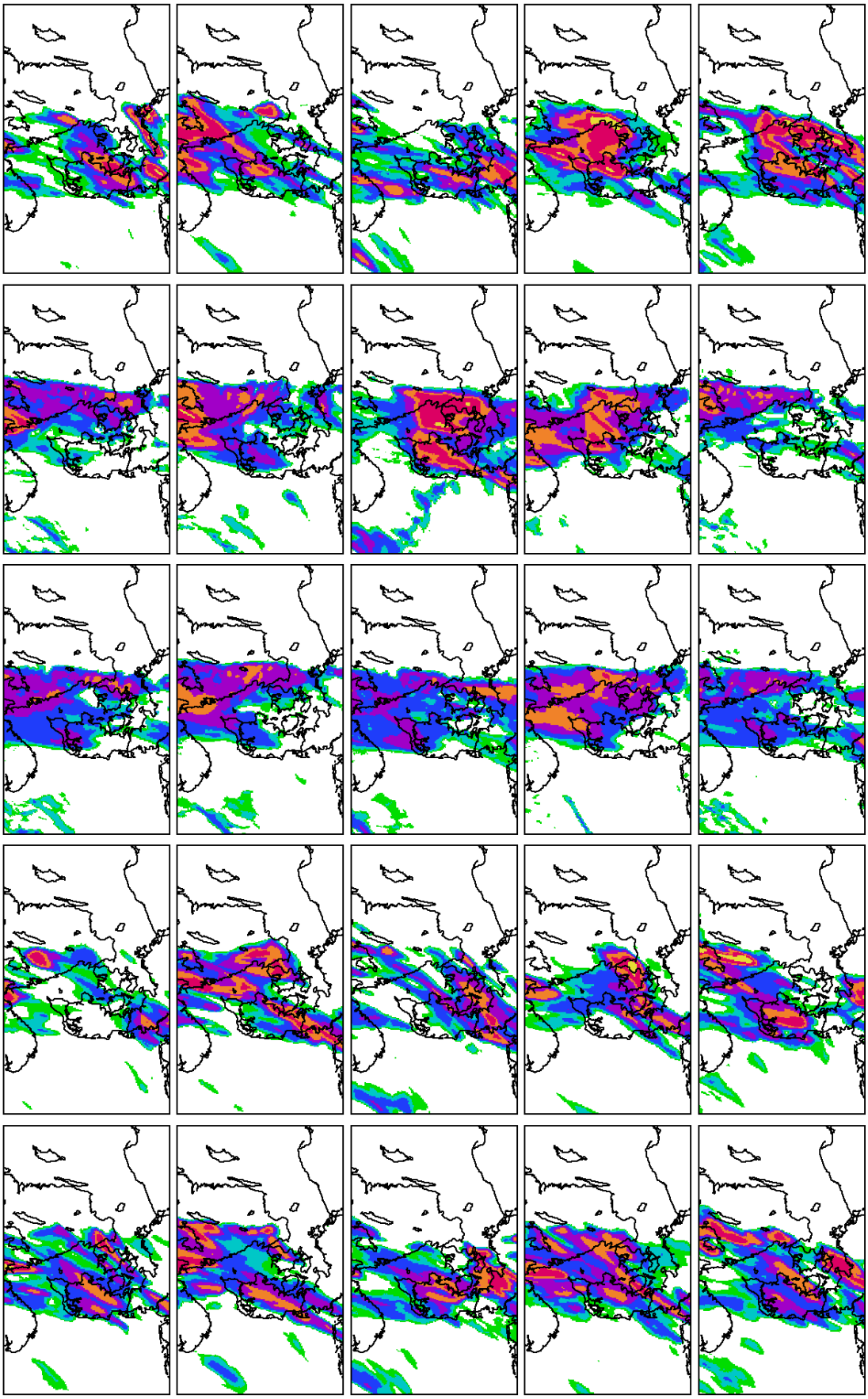


Figure 4.3: *As Fig. 4.2, but for precipitation accumulated from forecast hour 12 to 18.*

4.2 Forecast plumes and ensemble meteograms

By plotting forecasts for a single geographical point spatial information is lost, but it is straightforward to view all ensemble members on one plot and get a feeling for the ensemble spread. Figures 4.4-4.6 show forecast plumes of 2m temperature, precipitation [mm/h] and 10m wind speed. The individual ensemble members are colour-coded in the figures according to initial condition, cloud scheme and use of stochastic physics. The colour-coding gives an impression of systematic differences between the ensemble members. For example, members that use the KF/RK scheme appear to have maximum precipitation on Friday 2 UTC (Fig. 4.5), whereas there is more spread in the timing in the STRACO members. For 10m wind speed the ensemble maximum most frequently occur for members that include stochastic physics (Fig. 4.6).

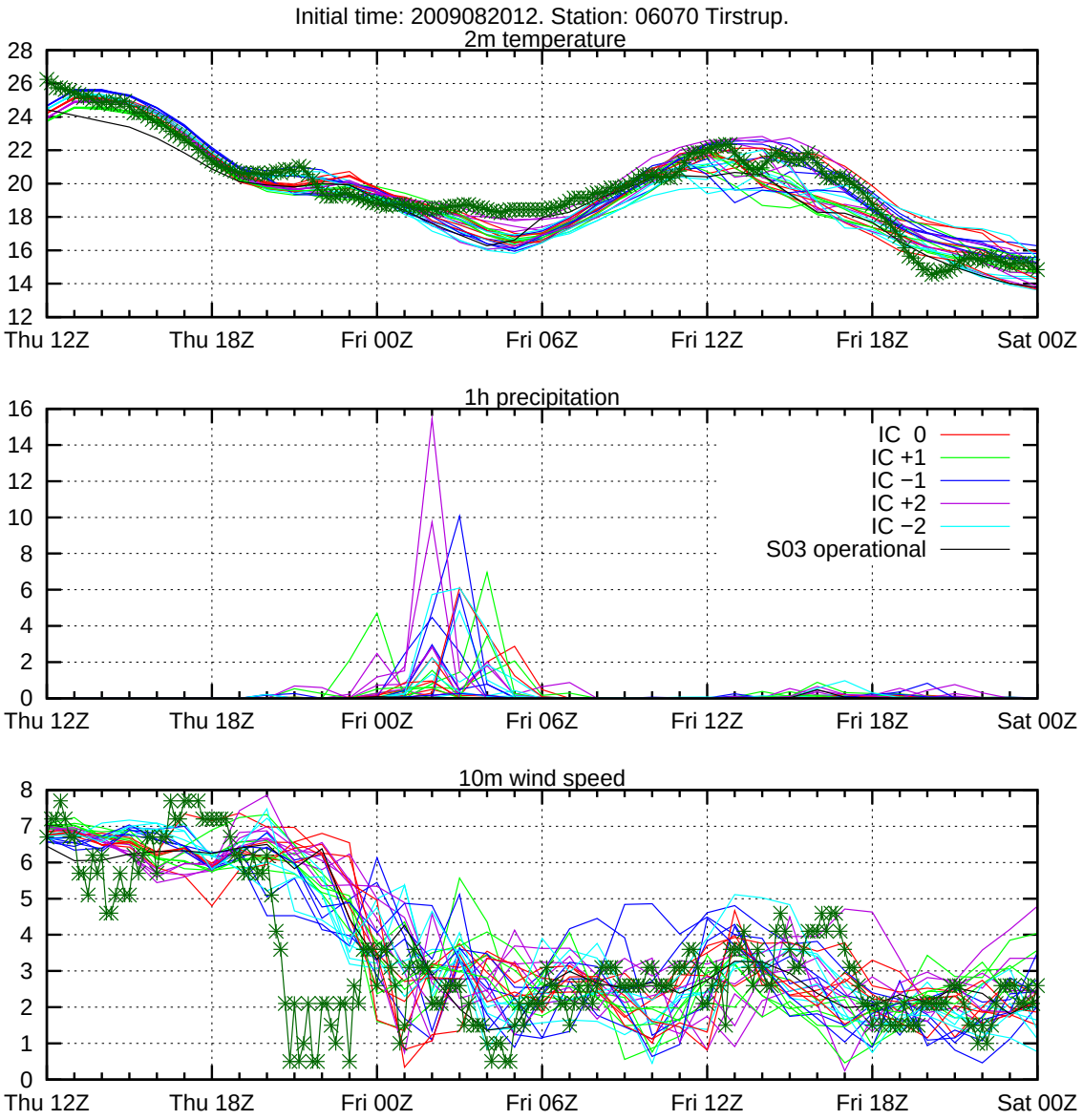


Figure 4.4: *Ensemble forecast plumes from 2009082012 for 06070 Tirstrup for 2m temperature (top), precipitation (centre) and 10m wind speed (bottom). Ensemble members are coloured according to initial condition. Dark green markes show observed 2m temperature and 10m wind speed at 10 min. intervals.*

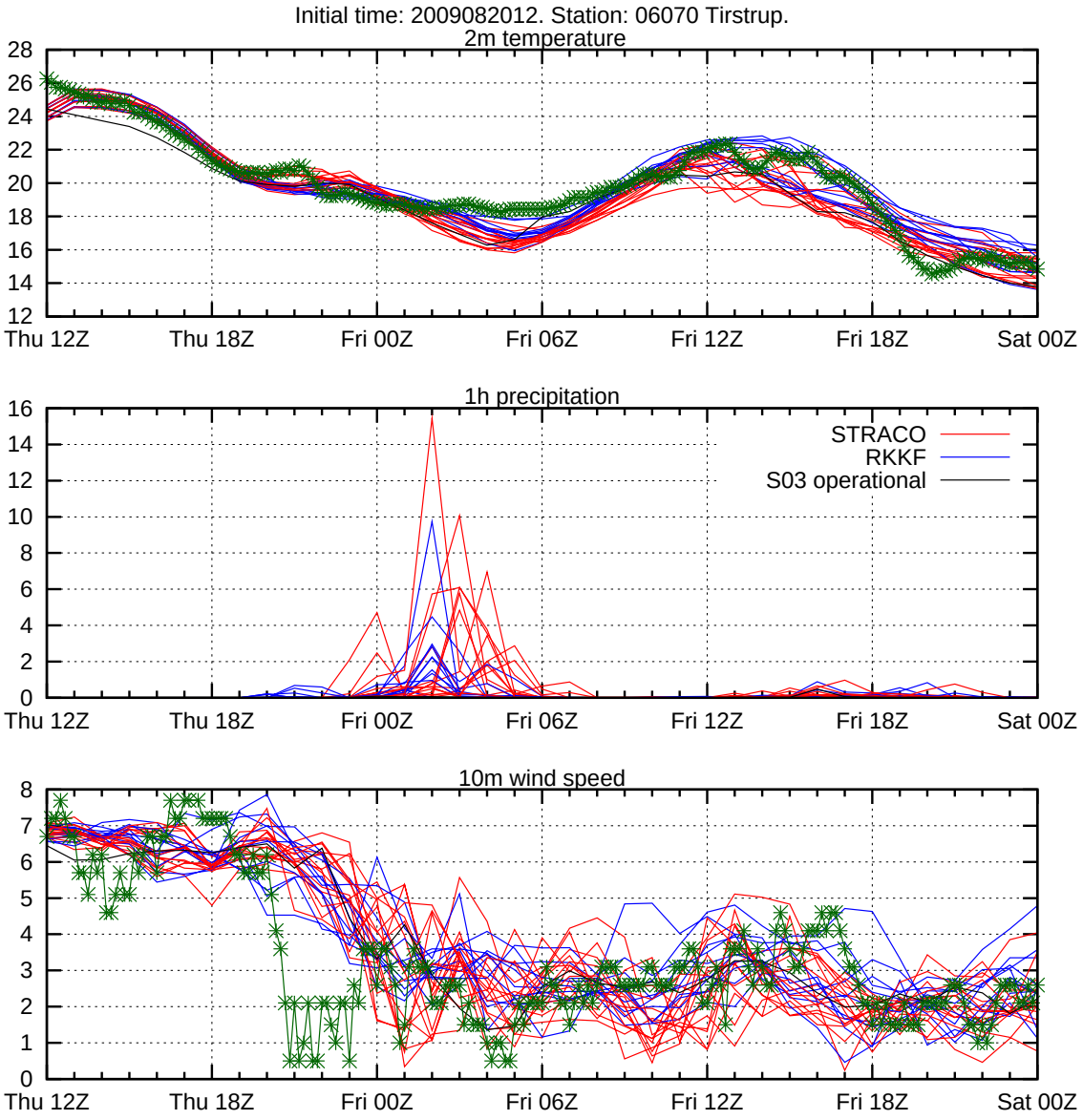


Figure 4.5*: As Fig. 4.4, but colours refer to condensation scheme.*

Note that although no ensemble member gets the timing right for the precipitation (12 mm were observed from Thursday 21 UTC until Friday 0 UTC), several members (but not the operational S03 forecast) produce precipitation a few hours later than observed.

The colour-coded ensemble members are useful for getting some insight into the ensemble, but for the average end user the many colours are probably only confusing. A more "user friendly" plot is shown in Fig. 4.7 where the colours indicate the ensemble distribution for 2m temperature and 10m wind speed, and where hourly precipitation is shown as a box-and-whisker plot. The S03 operational forecast is also shown for reference.

A couple of other plotting options for precipitation are shown in Fig. 4.8. One option shows each member plotted in a histogram for every forecast hour, where the members are sorted such that the highest values are centered and the lowest values are at the edges of the histogram. The second option shows the probability of exceedance as colours in a histogram. In both options the maximum

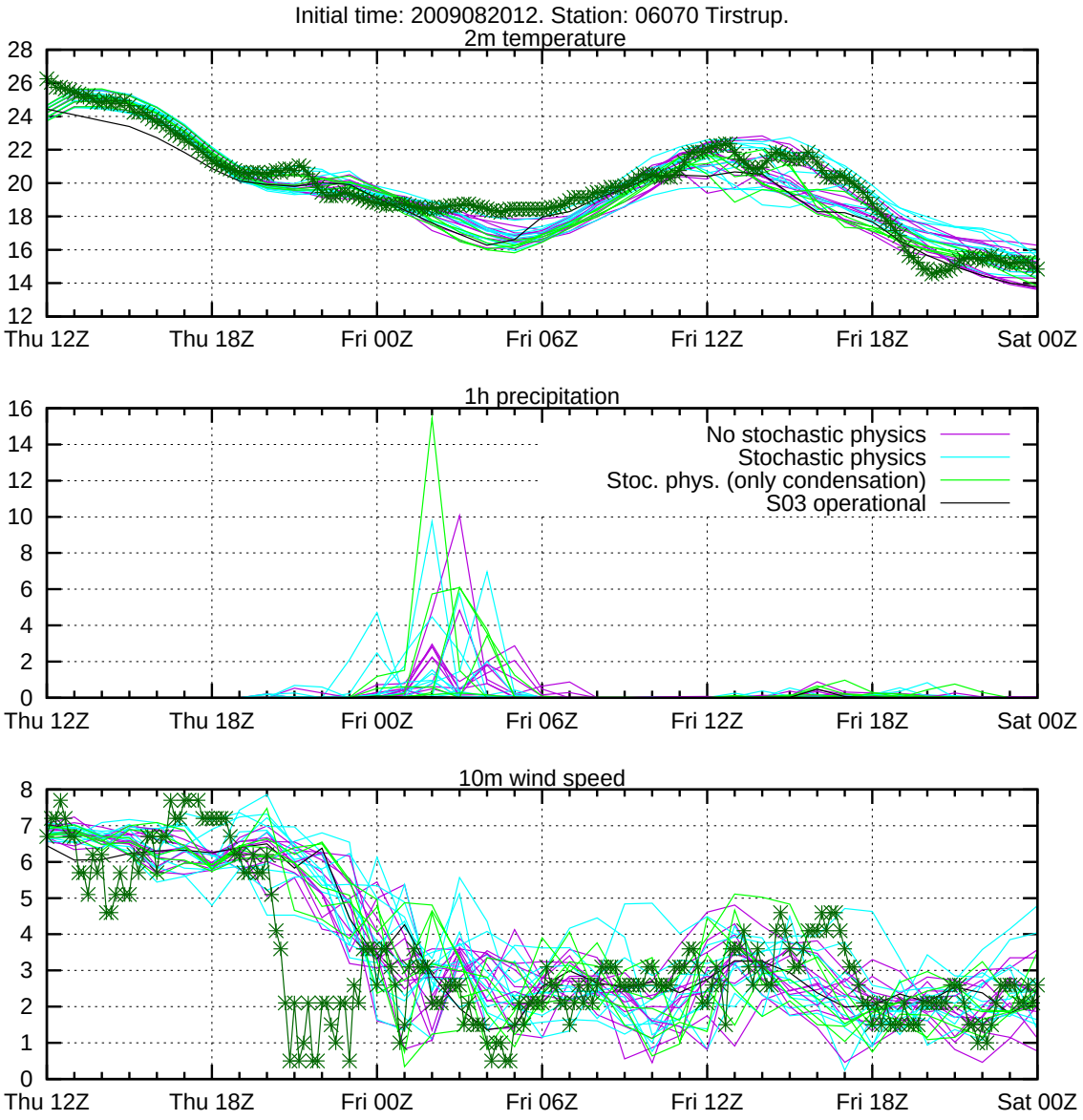


Figure 4.6*: As Fig. 4.4, but colours refer to the use of stochastic physics.*

member is excluded from the plot as this member occasionally is unrealistically high.

Figure 4.9 shows the ensemble of accumulated precipitation, coloured probability of exceedance of 3h precipitation and, in both cases, the verifying observations. Without any spatial forecast information it is evident that the ensemble forecast gives much better guidance for 06070 Tirstrup than the operational S03 forecast, even though the ensemble forecast fails to capture the precipitation between Thurday 21Z and Friday 00Z.

4.3 Spaghetti map

On 3 December 1999 Denmark was hit by “The storm of the century.” Forecasts have been rerun using the T15/S03 deterministic model configuration and the T15/S05 ensemble model configuration with lateral boundary conditions provided by the ECMWF ERA-Interim reanalyses. Figure 4.10 shows predicted and observed 10m wind speed at 06081 Blåvandshuk from around 30h prior to the

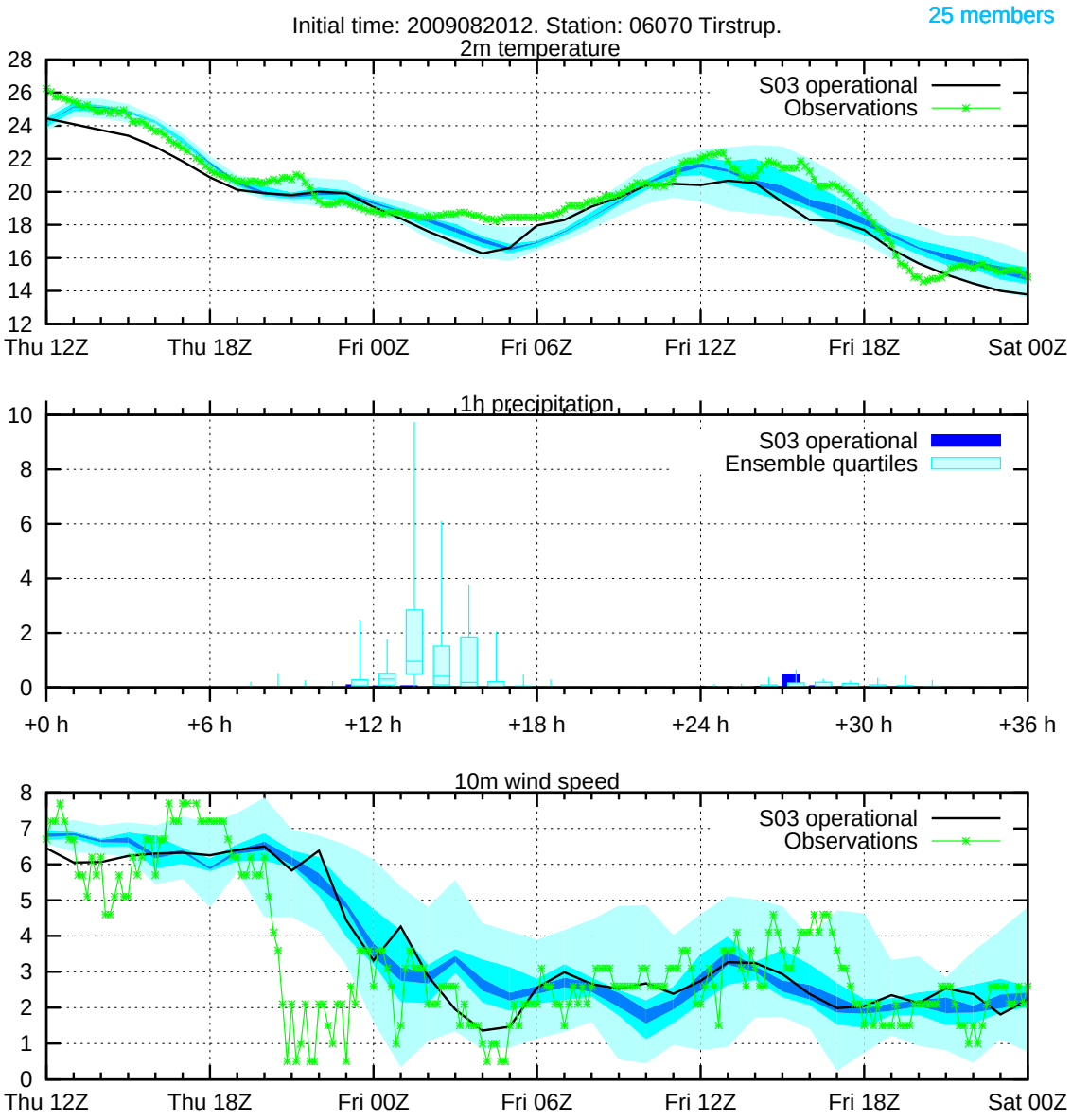


Figure 4.7: *As Fig. 4.4, but colours for 2m temperature and 10m wind speed show lowest 25%, second lowest 17%, middle 17%, second highest 17% and highest 25%. Box-and-whiskers for hourly precipitation show quartiles, median and 0.05 and 0.95 quantiles.*

wind maximum. We notice that the ensemble spread is relatively constant until the time of the maximum wind speed when the spread increases.

In order to get an impression of the spatial ensemble spread, consider a so-called spaghetti map. The spaghetti map is a plot of fixed contours of all ensemble members as in Fig. 4.11 which shows mean sea level pressure contours from two different initial times both verifying around the time of the peak of the storm. We notice a significant reduction in the “spaghetti spread” from the early to the later forecast. The lowest mean sea level pressure, 952.4 hPa was observed at Anholt at 18UTC.

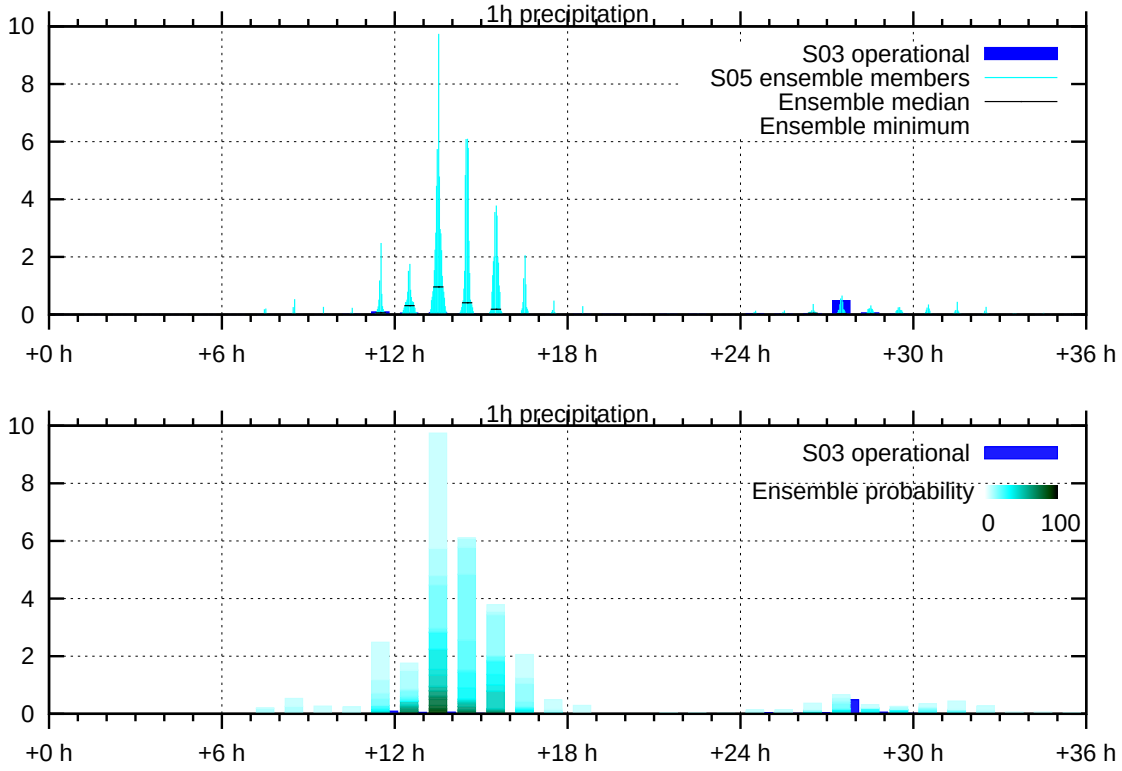


Figure 4.8*: As centre plot in Fig. 4.7, but top plot shows each ensemble member in a histogram for every forecast hour, where the members are sorted such that the highest values are centered and the lowest values are at the edges of the histogram; black horizontal bar shows median, white horizontal bar shows minimum (not visible in this plot as the minimum is 0 for all forecast lengths); bottom plot shows histogram where the colour indicates the probability of exceedance (number of members exceeding value). Both plots: operational S03 forecast is shown in blue; maximum member is excluded.*

4.4 Probability map

A probability map is simply a map of the fraction of ensemble members that exceeds a certain threshold. With a threshold of 25 m/s for 10m wind speed we see probabilities of more than 95% (Fig. 4.12) corresponding to at least 19 (of 20) members predicting more than 25 m/s where the wind field is most intense just before the storm hits land.

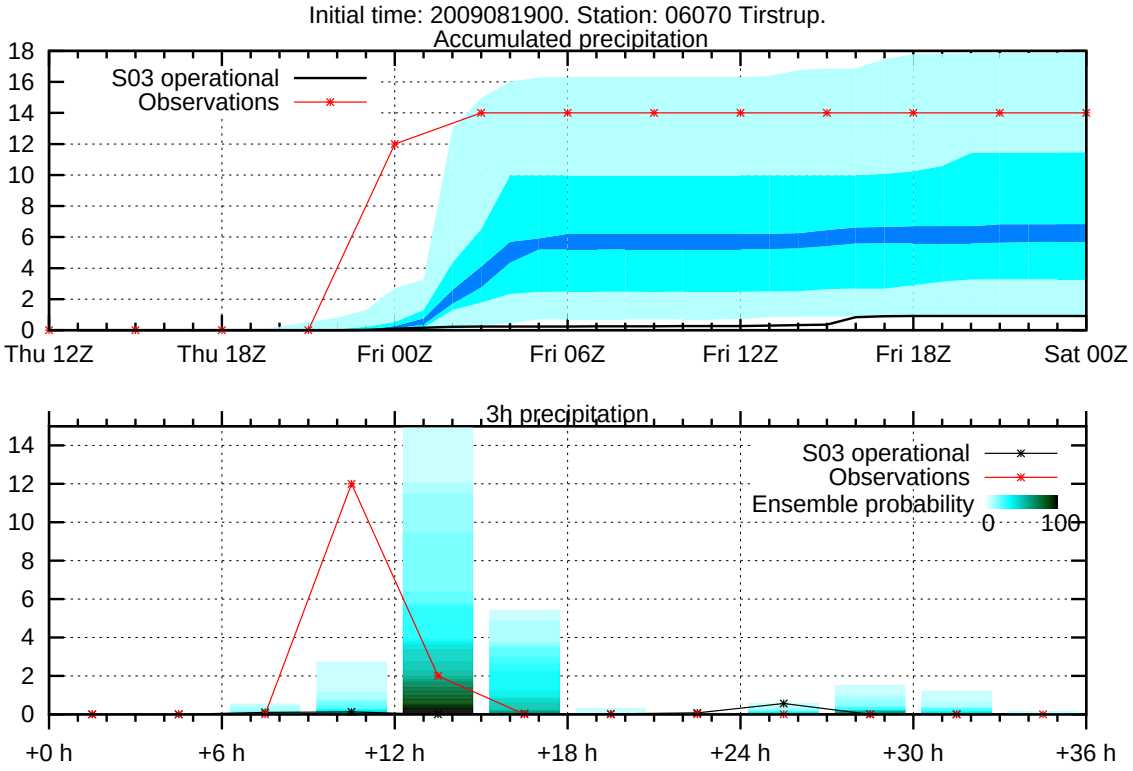


Figure 4.9: *Ensemble forecast from 2009082012 for 06070 Tirstrup for accumulated precipitation (top) and 3h precipitation (probability of exceedance; bottom). Operational S03 forecast and observed precipitation are also shown as black and red curves, respectively.*

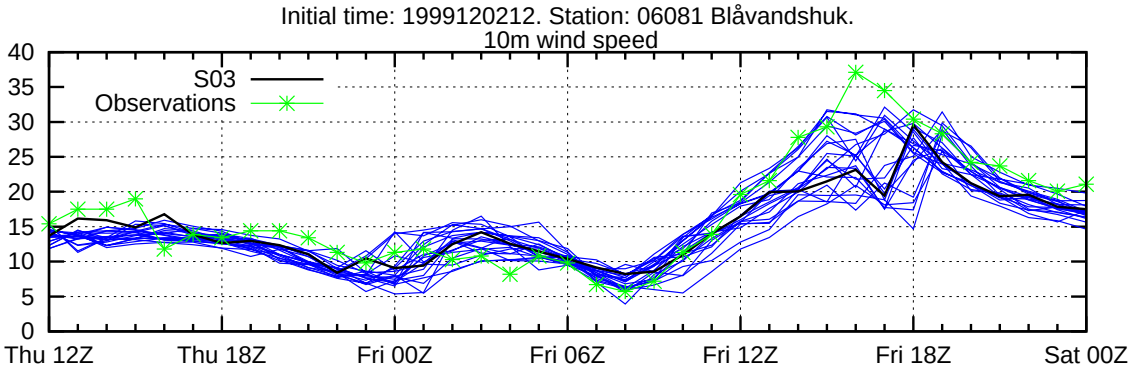


Figure 4.10: *Ensemble forecast (20 members, blue curves) from 1999120212 for 06081 Blåvandshuk for 10m wind speed. S03 forecast and observed wind speed are shown as black and green curves, respectively.*

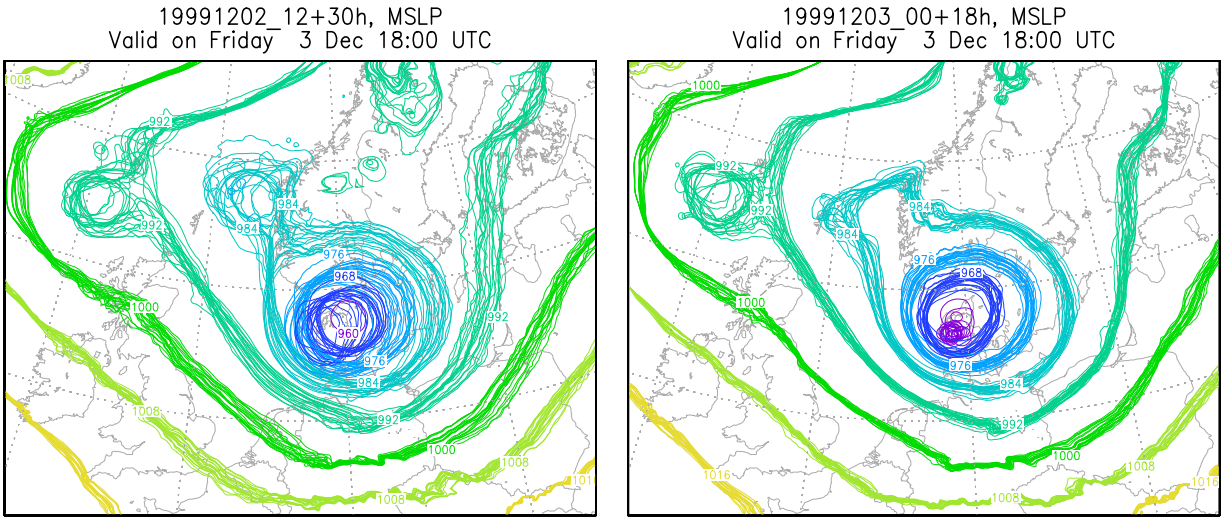


Figure 4.11: *MSLP spaghetti maps (S05 ensemble configuration) verifying on 19990318 at the peak of the storm over Denmark.*

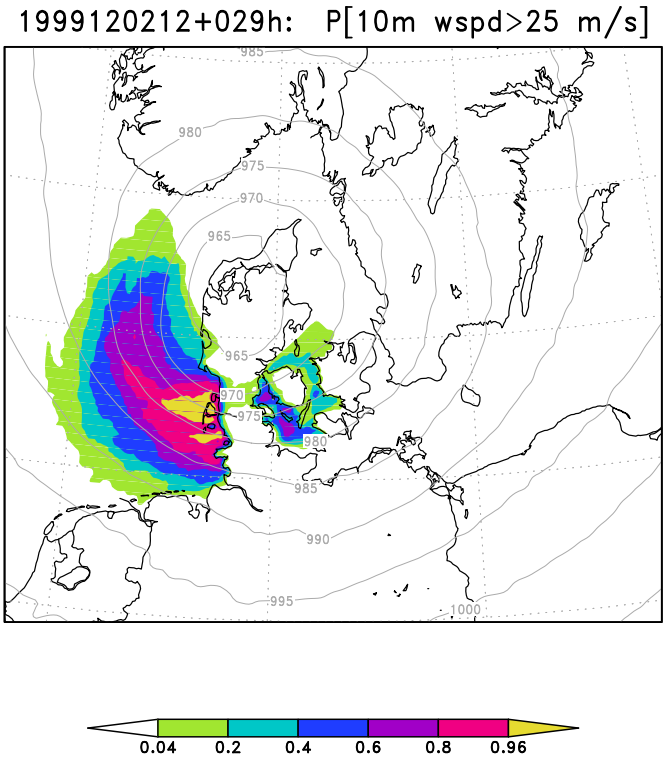


Figure 4.12: *Map showing fraction of ensemble members that predicts 10m wind speed above 25 m/s. Gray contours show MSLP for control run.*

5. Sources of ensemble spread and skill

The spread of the DMI ensembles are obtained by initial condition perturbation, multiple model configurations and stochastic physics perturbations. A natural question is, “what is more important for the performance of the ensemble prediction system?” In the following the effects of the different perturbation types are investigated by comparing verification scores for subsets of the ensemble such as “only initial condition perturbations vs. only model perturbations” and “stochastic physics vs. no stochastic physics.”

Figures 5.1-5.3 show the impact of stochastic physics. Members 1-5, 11-16 (model not including stochastic physics) are compared to members 6-10, 16-20 (model including stochastic physics). It is evident that using stochastic physics increases the ensemble spread and slightly improves the skill scores.

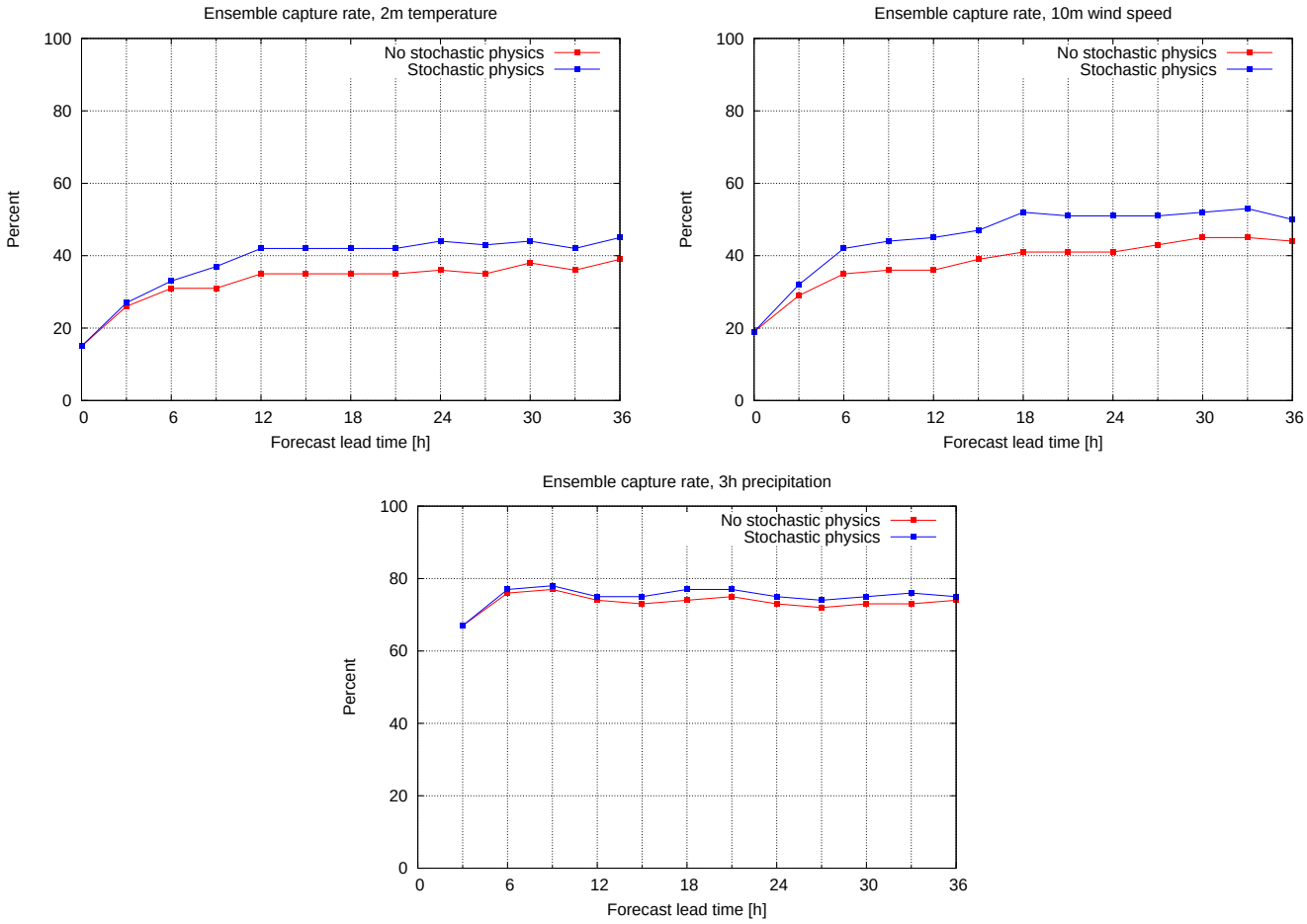


Figure 5.1: *Ensemble capture rates for forecasts of 2m temperature (top left), 10m wind speed (top right) and 3h accumulated precipitation (bottom). Red curves: model not including stochastic physics; blue curves: model including stochastic physics.*

Figures 5.4-5.6 compares spread and skill for five-member ensembles with (i) no initial perturbation (members 1, 6, 11, 16, 21), i.e. the ensemble spread is due to different model configurations, (ii) no model perturbation (members 1-5), i.e. the ensemble spread is due to different initial (and boundary) conditions, (iii) different initial conditions and stochastic physics (members 6-10), (iv) a mix of different initial conditions and model configurations (members 1, 7, 13, 19, 25) and (v) for 10m

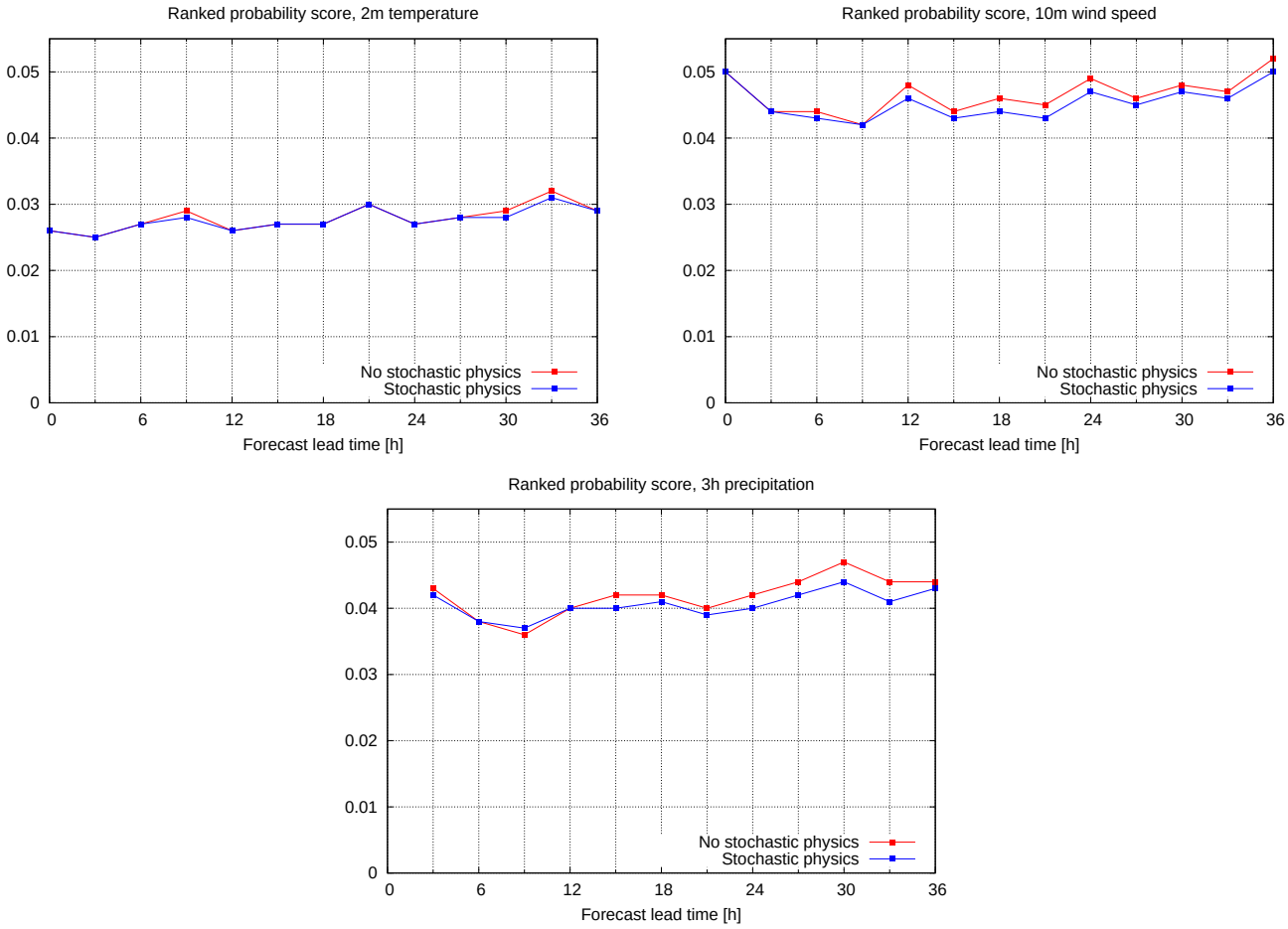


Figure 5.2*: As Fig. 5.1, but for ranked probability score (the smaller the better).*

wind speed different initial conditions, perturbed roughness lengths and stochastic physics due to convection and condensation only (members 21-25).

The mix of different initial conditions and model configurations generally yields the best forecasts, while using identical initial conditions and different model configurations generally yields the worst forecasts for 2m temperature and 10m wind speed, while the combination of different initial conditions and identical model configurations yields the worst forecasts for precipitation, i.e. preciptiation forecasts appear to be particularly sensitive to model perturbations.

The ensemble size also matters. Figures 5.7-5.9 show a positive impact on spread and skill when the enemble size is increased from 5 to 15 members (using in both cases a mix of different initial conditions and different model configurations), but the increase from 15 to 25 members only has a marginal impact. One reason for the latter might be that the 15-member ensemble is a subset of the 25-member ensemble, where the 15 members are selected so as to span the space of available initial conditions and model configurations. So, loosely speaking, the remaining members are based on perturbations that “lie between” those used for the 15-member ensemble. Thus, one should preferably use an ensemble configuration where the individual ensemble members differ both in initial condition and model configuration. While ensemble size is still important for the prediction of rare events, the ability to capture rare events will only have a marginal impact on standard verification scores such as those shown in this section.

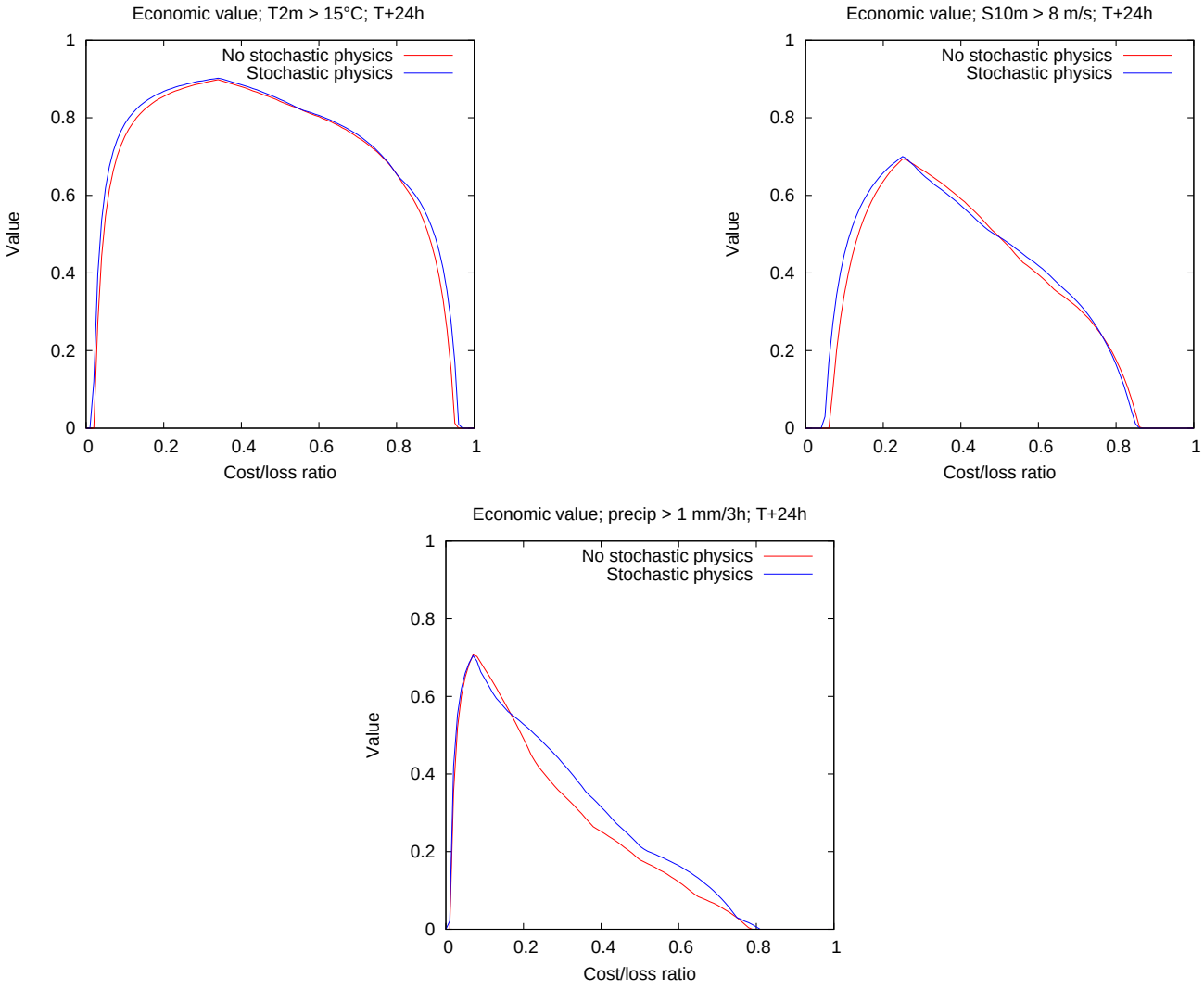


Figure 5.3*: As Fig. 5.1, but for economic value of 24h forecasts.*

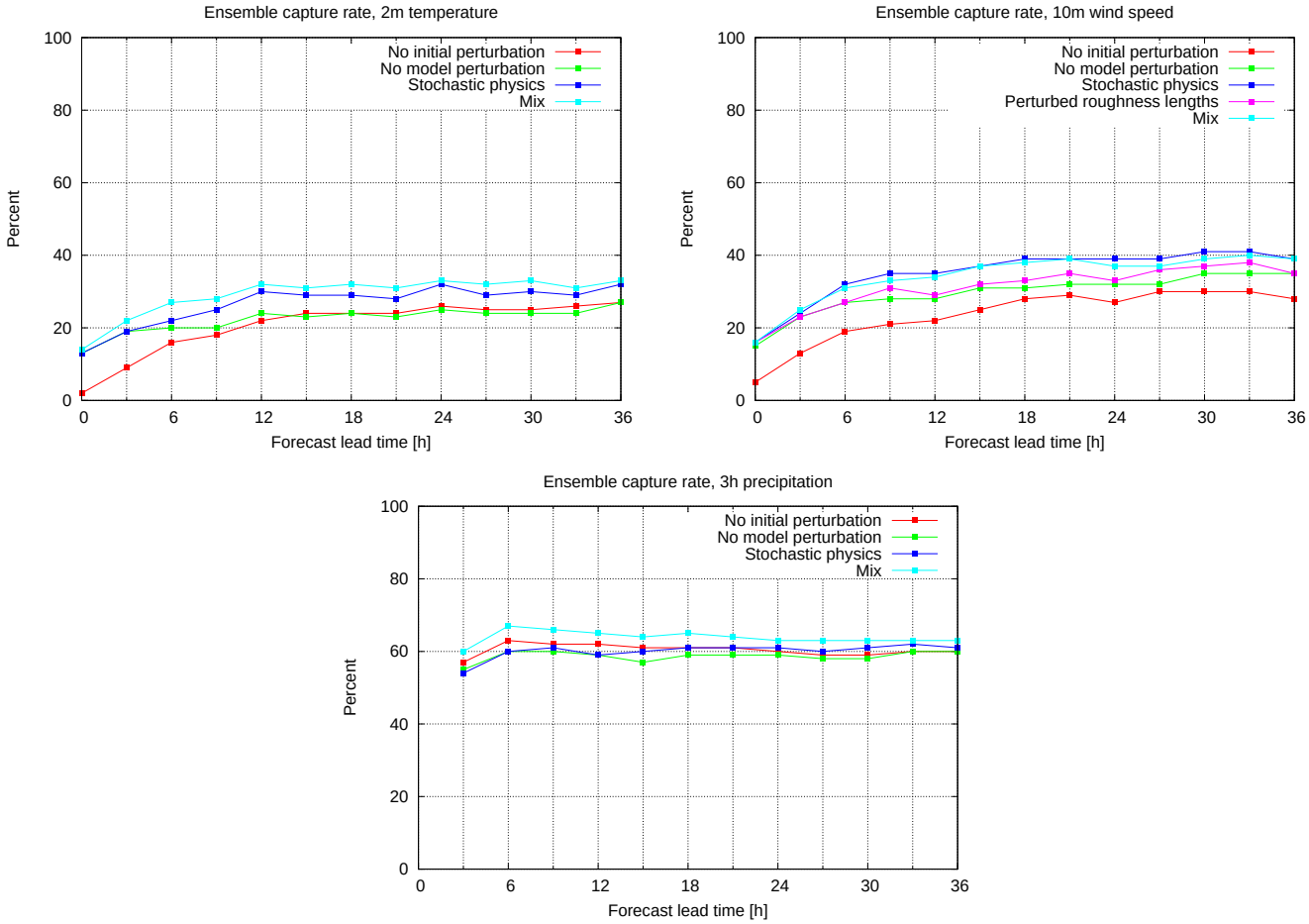


Figure 5.4: *Five-member ensemble capture rates for forecasts of 2m temperature (top left), 10m wind speed (top right) and 3h accumulated precipitation (bottom). Red curves: different model configurations, identical initial conditions; green curves: different initial conditions, identical model configurations; blue curves: different initial conditions, stochastic physics; cyan curves: different initial conditions, different model configurations; pink curve (wind speed only): different initial conditions, different roughness lengths, limited stochastic physics.*

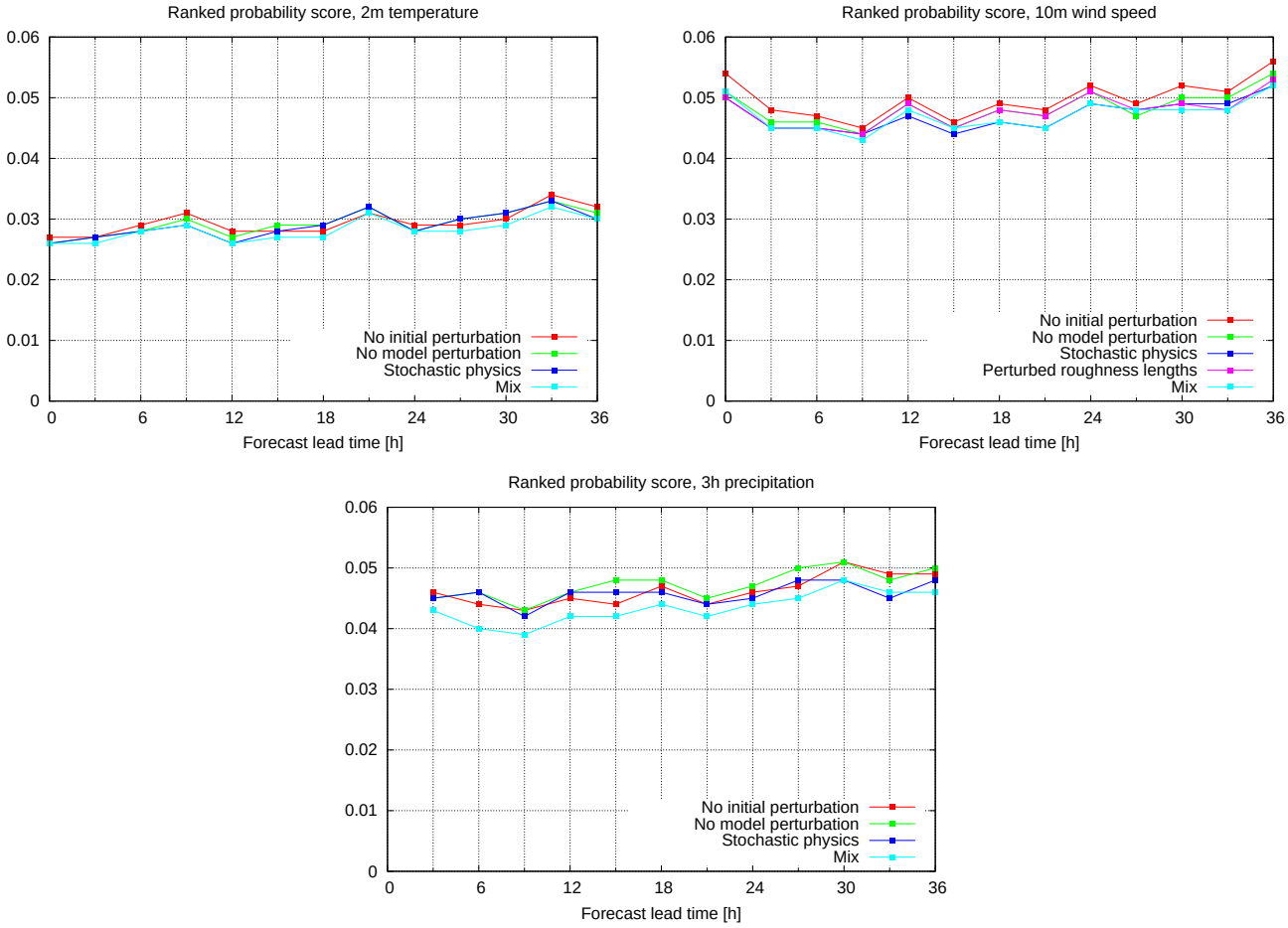


Figure 5.5*: As Fig. 5.4, but for ranked probability score (the smaller the better).*

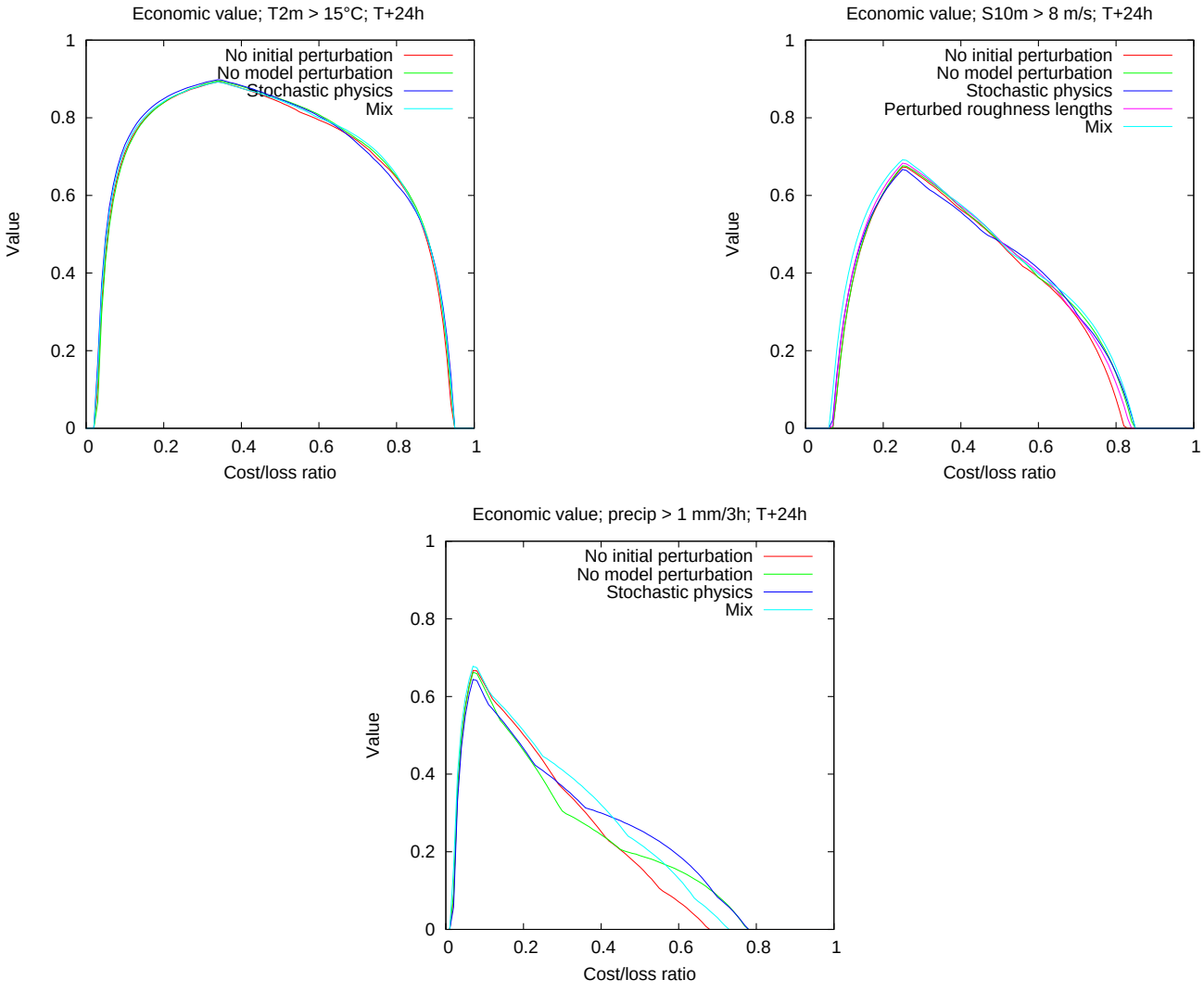


Figure 5.6*: As Fig. 5.4, but for economic value of 24h forecasts.*

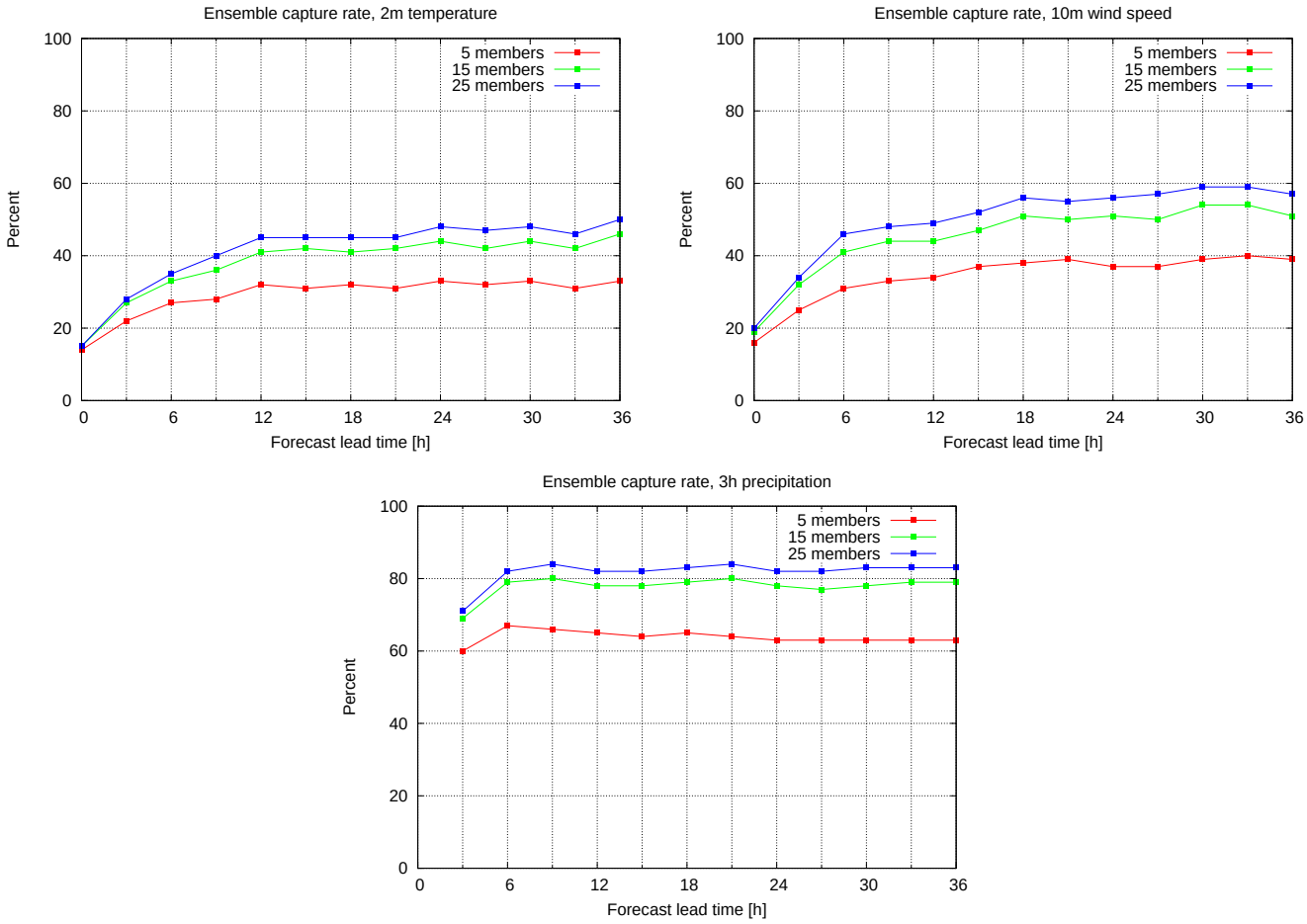


Figure 5.7: *Ensemble capture rates for forecasts of 2m temperature (top left), 10m wind speed (top right) and 3h accumulated precipitation (bottom). Red curve: 5-member ensemble; green curve: 15-member ensemble; blue curve: 25-member ensemble.*

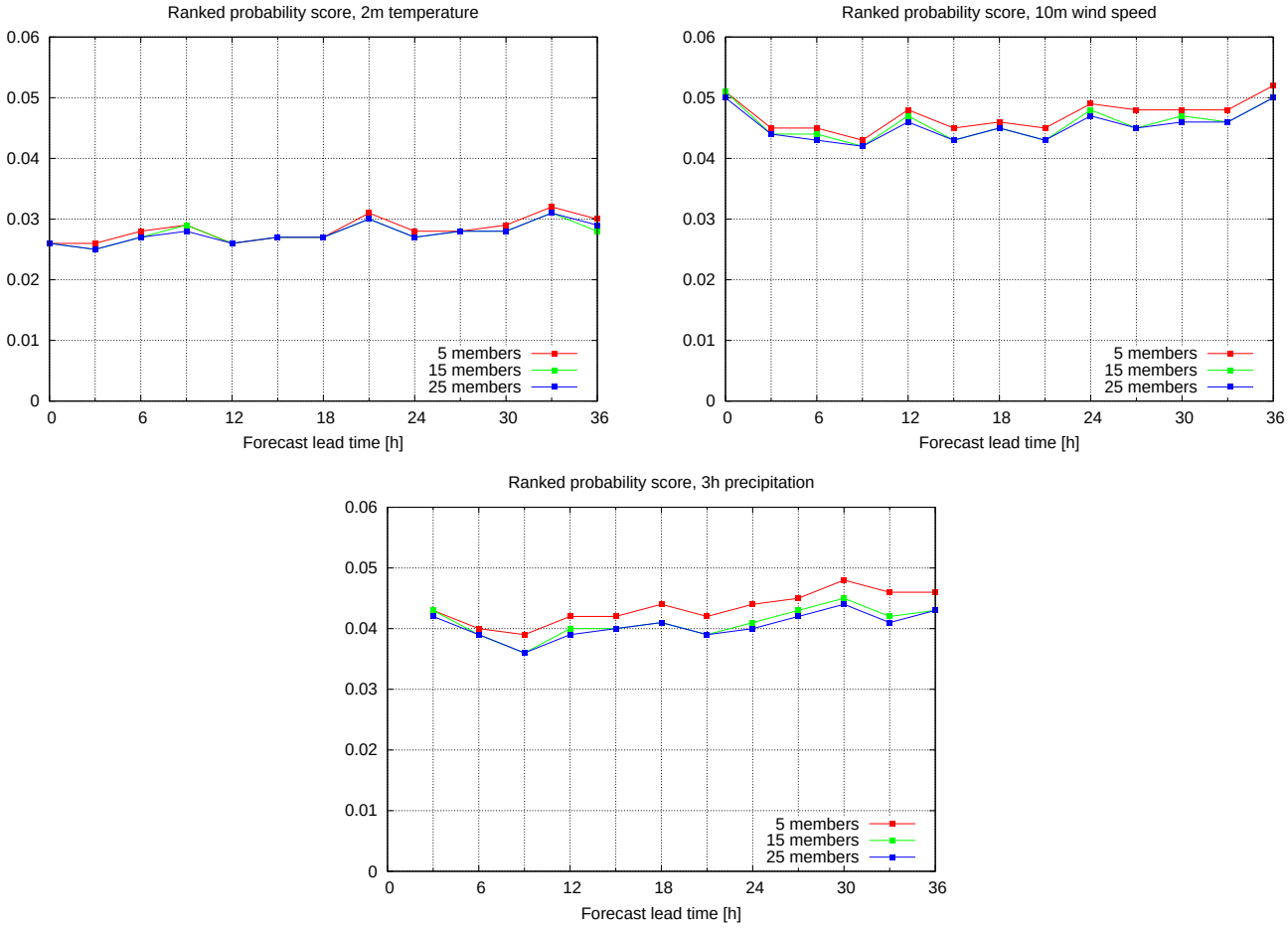


Figure 5.8*: As Fig. 5.7, but for ranked probability score (the smaller the better).*

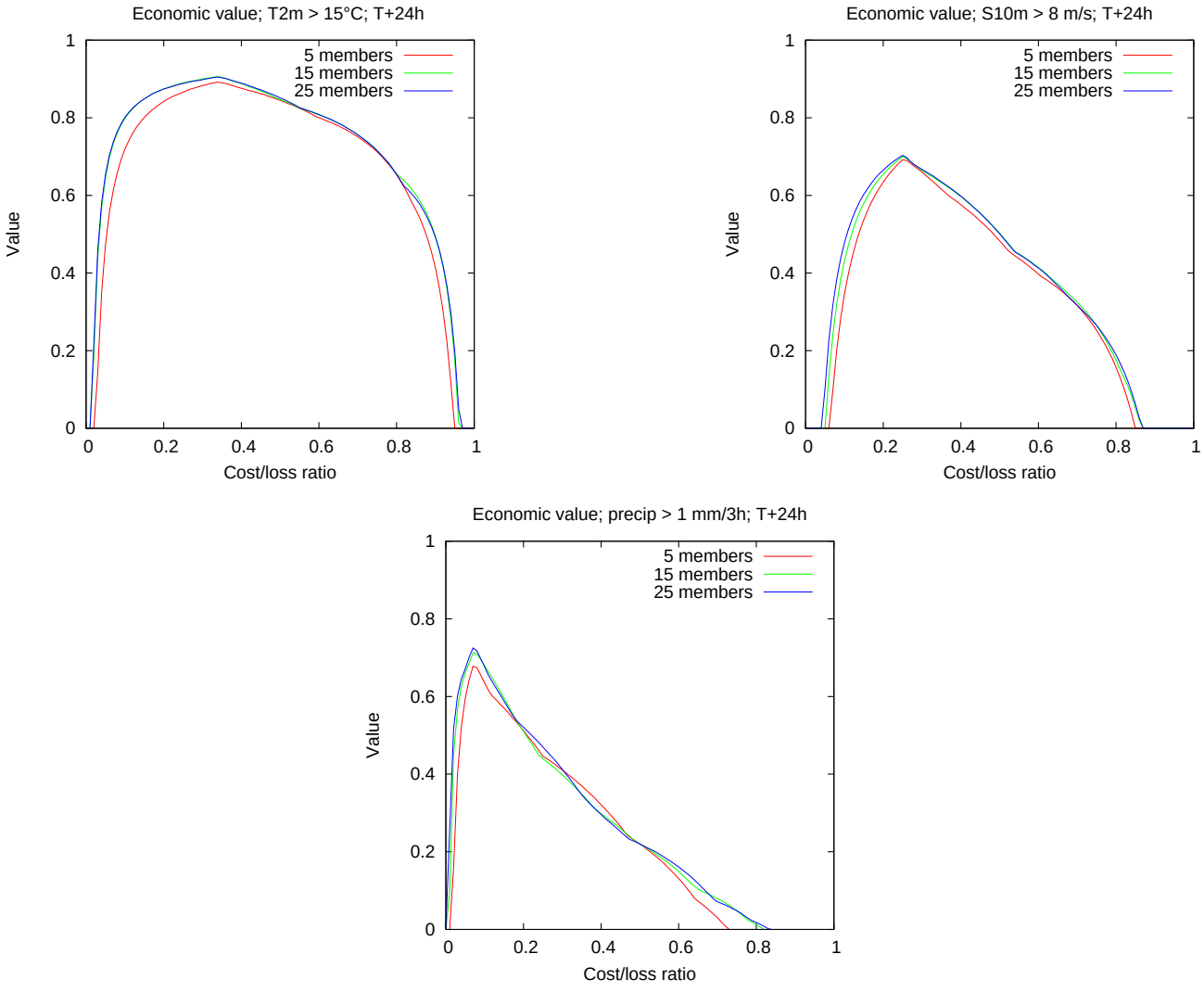


Figure 5.9*: As Fig. 5.7, but for economic value of 24h forecasts.*

6. Conclusions

Running the experimental ensemble prediction system based on the Hirlam T15/S05 configuration in real-time for several months has shown that

- it is feasible to run a reasonably sized ensemble prediction system (25 members, 36h forecasts four times per day) with the present computing facilities at DMI;
- the quality (skill) of DMI's experimental ensemble prediction system is (for the considered parameters, events, scores etc.) as good or better than ECMWF's operational ensemble prediction system for the first 36h of the forecast range;
- DMI's experimental ensemble prediction system can supplement the operational forecast in situations where the weather development is uncertain on short time scales, typically in shower conditions, as well as in situations with small-scale, fast developing weather systems and associated strong winds.

There are several aspects of the ensemble that has not yet been investigated, and there is almost certainly room for improvement. Lack of ensemble spread and the poor relation between spread and error are the two most obvious areas of potential improvement that should be kept in mind when going through the following "To Do list" for the near future:

- What is the importance of lateral boundary conditions? Rerunning forecasts combining lateral boundary conditions and initial conditions from the ensemble members may give an indication.
- Would a more advanced method for perturbing the initial conditions give more ensemble spread and better forecasts? Use of either bred vectors or singular vectors could be explored.
- Could combinations of different perturbations be enhanced as ensembles using combined perturbations (initial conditions and model physics) seem to give the best results?
- Should surface initial conditions be perturbed, or should some sort of stochastic physics be applied to the surface scheme? This might be a way to increase the ensemble spread in, e.g., 2m temperature forecasts.
- Should statistical postprocessing be applied to the ensemble forecasts in order to compensate for systematic short-comings? A simple example could be to inflate the ensembles to compensate for lack of ensemble spread.

With or without the above-mentioned possible improvements, obvious applications of the ensemble prediction system include

- automatic point forecasts (like DMI's "byvejr");
- input to storm surge and dispersion models where an ensemble forecast will enable a better risk assessment.

In addition, it is crucial that ensemble forecasts are presented in a way so that the user can easily extract the information that he/she needs. There is so much information available, that there is a substantial risk that the information signal/noise ratio becomes unacceptably low if one is not very careful with the presentation. In the present report some standard presentations have been shown (forecast plumes, probability maps, spaghetti maps, "postage stamps" etc.), but these may not be the most useful for a particular user. Other possibilities that might be more useful for another user could include maps that illustrate forecast spread or risk indices or maps based on clustering methods. Since different users are interested in different aspects of the forecasts there is a need for advice from the users regarding the best presentation of ensemble forecasts.

References

Bjerkness, V., 1904: Das Problem der Wettervorhersage, betrachtet vom Standpunkte der Mechanic und der Physik. Meteorol. Z., 21, 1-7.

Brier, G.W., 1950: Verification of forecasts expressed in terms of probability. Mon. Wea. Rev., 78, 1-3.

Bright, D.R. and S.L. Mullen, 2002: Short-Range Ensemble Forecasts of Precipitation during the Southwest Monsoon. Wea. Forecasting, 17, 1080-1100.

Buizza, R., M.J. Miller, and T.N. Palmer, 1999: Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. Quart. J. Roy. Meteor. Soc., 125, 2887-2908.

Du, J., G. DiMego, Z. Toth, D. Jovic, B. Zhou, J. Zhu, H. Chuang, J. Wang, H. Juang, E. Rogers, and Y. Lin, 2009: NCEP short-range ensemble forecast (SREF) system upgrade in 2009. 19th Conf. on Numerical Weather Prediction and 23rd Conf. on Weather Analysis and Forecasting, Omaha, Nebraska, Amer. Meteor. Soc., June 1-5, 2009 (http://www.emc.ncep.noaa.gov/mmb/SREF/SREFupgrade4NWP2009.pdf).

Ebisuzaki, W. and E. Kalnay, 1992: A modified lagged-average-forecast ensemble. WMO Research activities in atmospheric and oceanic modelling. Report 17, pp. 6.32-6.32.

Epstein, E.S., 1969: Stochastic-dynamic prediction. Tellus, 21, 739-759.

Feddersen, H. and K. Sattler, 2009: Verification of a set of GLAMEPS ensemble experiments. To appear in HIRLAM Newsletter, 55.

Frogner, I-L. and T. Iversen, 2008: Recent developments in EuroTEPS. HIRLAM Newsletter, 54, 92-95.

Hou, D., E. Kalnay, and K.K. Droegemeier, 2001: Objective Verification of the SAMEX '98 Ensemble Forecasts. Mon. Wea. Rev., 129, 73-91.

Houtekamer, P.L. and H.L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. Mon. Wea. Rev., 126, 796-811.

Iversen, T. and co-authors, 2009: Configurating GLAMEPS. To appear in HIRLAM Newsletter, 55.

Jolliffe, I.T. and D.B. Stephenson, 2003: Forecast Verification. A Practitioner's Guide in Atmospheric Science. Wiley and Sons Ltd, Chichester, 240 pp.

Kain, J.S., 2004: The Kain-Fritsch Convective Parameterization. An Update. J. Appl. Meteor., 43, 170-181.

Leith, C.E., 1974: Theoretical skill of Monte Carlo forecasts. Mon. Wea. Rev., 102, 409-418.

Li, X., M. Charron, L. Spacek and G. Candille, 2008: A Regional Ensemble Prediction System Based on Moist Targeted Singular Vectors and Stochastic Parameter Perturbations. Mon. Wea. Rev., 136, 443-462.

Lorenz, E.N., 1963: Deterministic nonperiodic flow. J. Atmos. Sci., 20, 130-141.

Marsigli, C., F. Boccanera, A. Montani, and T. Paccagnella, 2005: The COSMO-LEPS mesoscale ensemble system: validation of the methodology and verification. Nonlin. Precesses Geophys., 12, 527-536.

Molteni, F., R. Buizza, T.N. Palmer and T. Petroliagis, 1996: The new ECMWF ensemble prediction system: methodology and validation. Quart. J. Roy. Meteor. Soc., 122, 73-119.

Morss, R.E., J.L. Demuth and J.K. Lazo, 2008: Communicating uncertainty in weather forecasts: a survey of the U.S. public, Wea. Forecasting, 23, 974-991.

Mureau, R., F. Molteni and T.N. Palmer, 1993: Ensemble prediction using dynamically conditioned perturbations. Quart. J. Roy. Meteor. Soc., 119, 299-323.

Murphy, A.H., 1973: A new vector partition of the probability score. J. Appl. Met., 12, 595-600.

Palmer, T.N., R. Buizza, M. Leutbecher, R. Hagedorn, T. Jung, M. Rodwell, F. Vitart, J. Berner, E. Hagel, A. Lawrence, F. Pappenberger, Y-Y. Park, L. von Bremen and I. Gilmour, 2007: The ensemble prediction system - recent and ongoing developments. ECMWF Technical Memorandum, 540, 53 pp. (http://www.ecmwf.int/publications/library/do/references/show?id=88273).

Rasch, P.J. and J.E. Kristjansson 1998: A comparison of the CCM3 model climate using diagnosed and predicted condensate parameterizations. J. Climate 1587-1614.

Ricardson, L.F., 1922: Weather prediction by numerical process. Cambridge Univ. Press (reprint by Dover, New York, 1965).

Sass, B.H., 2002: A research version of the STRACO cloud scheme. DMI Technical Report no. 02-10 (http://www.dmi.dk/dmi/tr02-10.pdf).

Strauss, B. and A. Lanzinger, 1995: Validation of the ECMWF ensemble prediction system, in Proc. ECMWF Seminar on Predictability Vol. II, pp. 157-166 (http://www.ecmwf.int/publications/library/do/references/show?id=83450).

Toth, Z. and E. Kalnay, 1993: Ensemble forecasting at NMC: the generation of perturbations. Bull. Amer. Meteor. Soc., 74, 2317-2330.

Toth, Z. and E. Kalnay, 1997: Ensemble forecasting at NCEP: the breeding method. Mon. Wea. Rev., 125, 3297-3318.

Tracton, M.S. and E. Kalnay, 1993: Ensemble forecasting at NMC: practical aspects. Wea. Forecasting, 8, 379-398.

Wang, X. and C.H. Bishop, 2003: A comparison of breeding and ensemble transform kalman filter ensemble forecast schemes. J. Atmos. Sci., 60, 1140-1158.

Yang, X., C. Petersen, B. Amstrup, B.S. Andersen, H. Feddersen, M. Kmit, U. Korsholm, K. Lindberg, K. Mogensen, B.H. Sass, K. Sattler and N.W. Nielsen, 2005: The DMI-HIRLAM upgrade in June 2004. DMI Technical Report no. 05-09 (http://www.dmi.dk/dmi/tr05-09.pdf).

Previous reports

Previous reports from the Danish Meteorological Institute can be found on:
http://www.dmi.dk/dmi/dmi-publikationer.htm