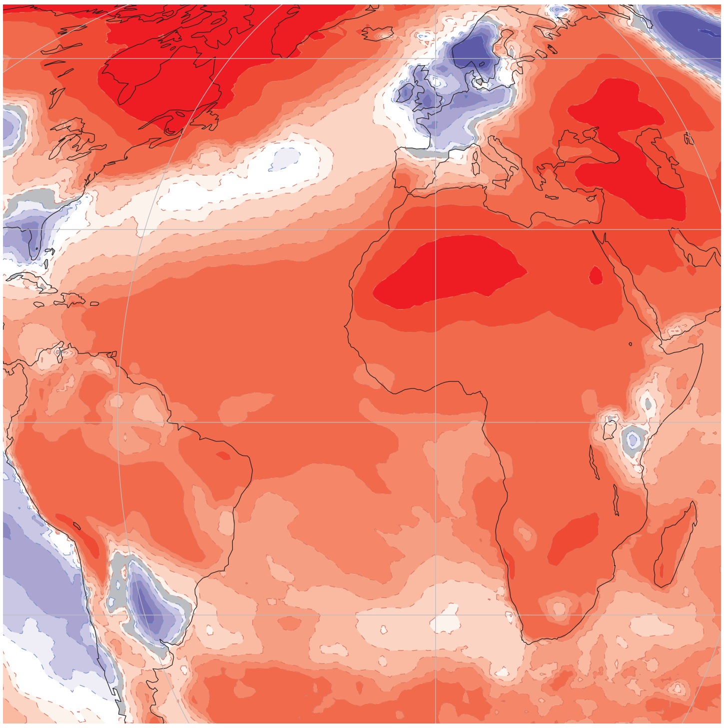
# ECMWF Feature article

from Newsletter Number 139 – Spring 2014

METEOROLOGY

# Flow-dependent verification of the ECMWF ensemble over the Euro-Atlantic sector



www.ecmwf.int/en/about/news-centre/media-resources

doi:10.21957/j1ipst9o

*This article appeared in the Meteorology section of ECMWF Newsletter No. 139 – Spring 2014, pp. 34–38.*

# Flow-dependent verification of the ECMWF ensemble over the Euro-Atlantic sector

Laura Ferranti, Susanna Corti, Martin Janousek

Because the atmospheric circulation is chaotic and its evolution is sensitive to the initial state, the skill of numerical weather predictions is flow dependent. This means that it is easier to make skilful predictions starting from some flow configurations than from others. For simplified models, chaos theory can provide the 'intrinsic' predictability level of atmospheric variations, but in operational practice, estimates of predictability are made from forecast ensembles.

ECMWF runs an ensemble of 50 independent forecasts (with perturbed initial conditions and model physics) to estimate forecast uncertainty, such that the spread amongst ensemble members gives an estimate of predictability. On some days, the spread will be small, implying that the atmosphere is very predictable. On other days, the ensemble of forecasts will diverge considerably, indicating that the atmosphere is less predictable. Identifying which circulation patterns lead to more predictable states than others (i.e. forecasting the forecast skill) is relevant for interpreting the forecast.

This study aims to assess the relative skill of medium-range weather forecasts depending on which flow pattern is in place over the North Atlantic when the forecast is initiated. A key aspect in the evaluation of flow-dependent predictability is that a defined flow circulation pattern must occur with sufficient frequency that statistics of ensemble forecast spread can be gathered. For this reason we use the concept of weather regimes to classify a small number of flow patterns. Consequently, the intra-seasonal variability of the North-Atlantic atmospheric circulation is described as transitions between a small number of recurrent and quasi-stationary states called weather regimes.

Weather regimes are generally computed by applying clustering algorithms on a circulation variable (such as the geopotential height at 500 hPa). The study of the frequency of occurrence and/or persistence of weather regimes provides a framework for the analysis of the complex atmospheric dynamics. This description assumes that there are preferred regions in the phase space (the space in which all possible states of a system are represented) where atmospheric trajectories tend to reside for extended periods. This study uses the four Euro-Atlantic climatological regimes (Figure1) that explain a large portion of the low-frequency variability in this geographic area. These regimes are:

- Positive North Atlantic Oscillation (NAO+)
- Negative North Atlantic Oscillation (NAO-)
- Scandinavian Blocking (BL)
- Atlantic Ridge (AR)

See Box A for further information.

**AFFILIATIONS**
**Laura Ferranti, Martin Janousek:** ECMWF, Reading, UK
**Susanna Corti:** Istituto di Scienze dell'Atmosfera e del Clima (ISAC), CNR, Bologna, Italy

 doi:10.21957/j1ipst9o

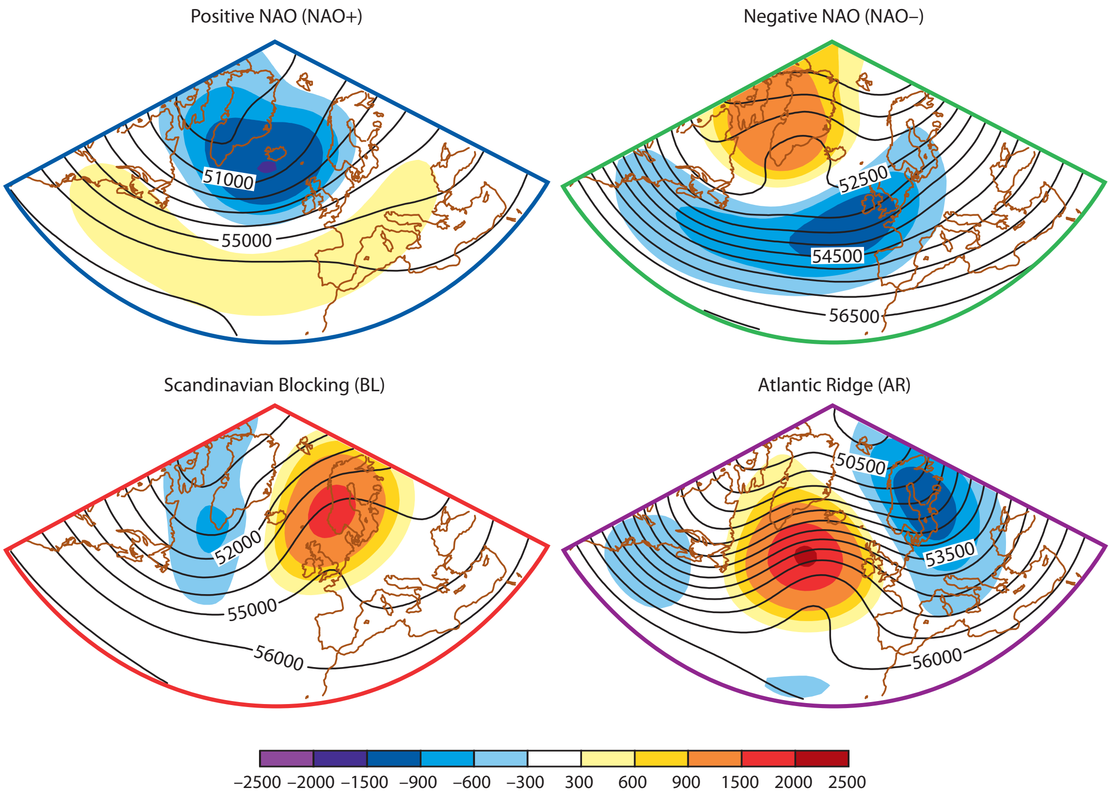



**Figure 1** Geographical patterns of the four Euro-Atlantic climatological regimes (both anomalies and full fields) for the October to April cold season. The geopotential anomalies (colour shading) and geopotential (contours) at 500 hPa in units of $m^2s^{-2}$ are derived from ECMWF's reanalysis data.

## Data and methods

**A**

The present analysis uses the ECMWF operational ensemble forecast (ENS) (*Leutbecher & Palmer,* 2008) and the ECMWF operational analyses of daily geopotential height at 500 hPa. The data used covers five cold seasons from October 2007 to April 2012. The ENS, based on 51 members (1 unperturbed and 50 starting from slightly perturbed initial conditions), has been designed to simulate initial and, through the application of stochastic physics, model uncertainties. At present, ENS runs with approximately 32-km horizontal resolution up to forecast day 10, and 64 km thereafter. Since the ECMWF forecasting system is regularly upgraded, the evaluation is confined to the five most recent winters. This is a compromise between reducing discontinuities associated with the impact of model changes in the forecast data and retaining a sufficient amount of cases.

The climatological regimes used in this study have been computed by using the 'k-means' clustering algorithm on daily anomalies of 500 hPa geopotential height taken from ECMWF reanalysis over the domain (80°W–40°E, 30°–90°N) for the 29 cold seasons (October to April) 1980–2008. The patterns obtained correspond to the four well-known clusters described by many authors (e.g. Cassou, 2008). There are the two patterns describing the opposite phases of the North Atlantic Oscillation (NAO+, NAO-), the Scandinavian Blocking (BL) and the Atlantic Ridge (AR) (Figure 1). It is interesting to note that the two phases of the NAO together with the AR regimes describe the three preferred North Atlantic jet stream locations (*Woollings et al.,* 2010), namely, NAO-, NAO+ and AR correspond to southern, central and northern jet-states respectively.

The four regimes are used in the ECMWF medium-range clustering products (*Ferranti & Corti,* 2011) to provide additional information about the ENS in terms of large-scale circulation and to allow an objective verification of the regime transitions. A pattern-matching algorithm is used to assign each individual forecast member to the closest climatological weather regime (in terms of the root mean square difference). To account for the seasonal evolution (in the classification), the patterns and amplitudes of the climatological regimes are adjusted month by month.

## Which flow regime leads to less or more skilful predictions?

Changes between the four weather regimes shown in Figure 1 are used to describe the low-frequency component of the atmospheric variability. In this simplified representation, where only four possible flow configurations are considered, we assess which circulation regime leads to more or less accurate predictions over the Euro-Atlantic sector. All forecasts are stratified according to the regime in the initial conditions. For example, all the forecasts initiated with a dominant zonal flow over the Atlantic are grouped in the category of forecast initiated with a NAO+ regime.

The next step is to consider the anomaly correlation of the ensemble means forecast for the four categories as a measure of deterministic skill (Figure 2). Between day 9 and day 13 the forecasts initiated in a Scandinavian Blocking or Atlantic Ridge flow-type show a larger drop in skill than the forecasts initiated in NAO- or NAO+. By day 15, forecasts initiated in a blocking regime have the lowest anomaly correlations. Forecasts initiated in NAO- are the most skilful beyond 10 days.

The Root Mean Square Error (RMSE) of the ensemble mean normalized by the standard deviation of the analysis (not shown) provides equivalent results. Probabilistic scores are also consistent, although the differences in skill levels between the four categories appear to be less significant.

Several studies show that instability processes of the large-scale flow play major roles in the development of blocking anomalies and in the growth of errors during blocking transitions. The fact that in the late medium range, forecasts initiated during a blocking regime are generally less skilful suggests that further progress is needed to understand the processes that maintain the blocking circulation.

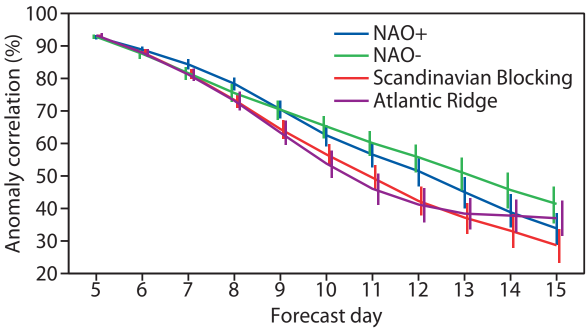



**Figure 2** Anomaly correlation of the ensemble for the four forecast categories as a function of forecast range for Europe for five cold seasons (October–March 2007/08 to October–March 2011/12). The bars, based on 1,000 subsamples generated with the bootstrap method, indicate the 95% confidence intervals.

## Regime transitions

The model's ability to correctly reproduce regime transitions and regime persistence is assessed by stratifying the forecasts according to both their initial conditions and their accuracy at day 10. All ensembles of forecasts, initiated with a given regime, are grouped into the same category within which we distinguish two additional groups: the good and poor forecasts. We define as poor (good) forecasts those with a RMSE of the ensemble mean being in the upper (lower) fifth of the whole RMSE distribution computed over the European domain (12.5°W–42.5°E, 35.0°N–75.0°N) at day 10. For each group and each category we compute composite maps of anomalies of 500 hPa geopotential height.

Figure 3 shows the composites of the anomalies for the poor forecasts initiated in the NAO+ regime: at initial time (Figure 3a) and after 10 days (Figure 3b), with the composites of the verification anomalies (Figure 3c). Over the Euro-Atlantic sector the model composite at day 10 exhibits a similar anomaly pattern to that of the initial conditions, indicating that in both cases the large-scale flow is characterized by enhanced westerlies across the Atlantic. On the other hand, the verifying composite, with a high anomaly over the Scandinavian Peninsula, exhibits the typical blocking circulation pattern. Such a high level of spatial coherence in the observed anomaly patterns of the composites after 10 days and their similarity to the Scandinavian Blocking regime structure is remarkable and indicates that most of the poor forecasts are missing the same observed regime transition. The composite for the poor forecasts clearly suggests that the model failed to make a transition from a strong zonal flow to a blocking pattern, instead favouring the persistence of the zonal circulation.

It is interesting to note that the change from NAO+ (zonal flow) to a blocked flow is one of the preferred observed transitions documented by *Vautard* (1990). Table 1 shows the population of the four climatological regimes (as a percentage) at different time ranges for the good and poor forecasts initiated in NAO+. The numbers in black indicate the forecast values and in red the verification values (if different). Looking at the

 doi:10.21957/j1ipst9o

poor forecasts in Table 1, it can be seen that 40% of the observed cases developed into a blocking type of flow by day 5 and by day 10 those cases increased to 51%. In the forecast the number of transitions to a blocking regime at day 10 are underestimated (42% versus 51%) and the persistence of the prevalent zonal flow is over-estimated (37% versus 21%).

The composite anomalies (Figure 3a) at initial time show a coherent structure over the Pacific sector reminiscent of the negative phase of the Pacific North Atlantic circulation pattern (PNA). This is consistent with analysis from *Corti & Palmer* (1997) which showed that the largest NAO sensitivity to small initial perturbation, and therefore loss of predictability, is associated with a negative phase of the PNA.

The composite anomalies associated with the poor forecasts documented by *Rodwell et al*. (2013) are very similar to the ones represented in Figure 3c. However, the flow conditions preceding the poor forecast events in their study bear no similarity with those depicted in Figure 3a. This inconsistency could be due to us looking at different forecast ranges (10 days versus 6 days) and the poor forecasts in their study occurring in a different season (late spring).

For the good forecasts initiated in NAO+, Table 1 shows that these are characterized by 35% of cases during which the zonal flow persisted, 28% of transitions to blocking and 21% of transitions to an NAO- regime. The forecast, for these selected good cases, was able to represent the correct percentage of transitions to blocking as well as to the other flow patterns. As opposed to the poor forecasts, the good forecast composites at the initial condition (not shown) do not present a definite coherent structure over the Pacific area, perhaps suggesting a reduced sensitivity to initial perturbations and in turn an increased predictability.

The composites for the ‘poor’ and ‘good’ forecasts initiated in the other three regimes are not shown here for the sake of brevity. However, we can point out that:

- Poor forecasts initiated in NAO- underestimate the transitions to the blocking regime; the good forecasts are mainly dominated by the cases with persistence of NAO-.
- Poor forecasts initiated in blocking are characterized by the model failure to maintain the blocking regime and favouring instead transitions to the AR and zonal regimes. The poor forecasts initiated in blocking show the largest errors compared with the poor forecasts initiated in any other regime.

Overall the main forecast deficiency, in terms of flow regimes, is in reproducing transitions to blocking and in maintaining the blocking circulation.

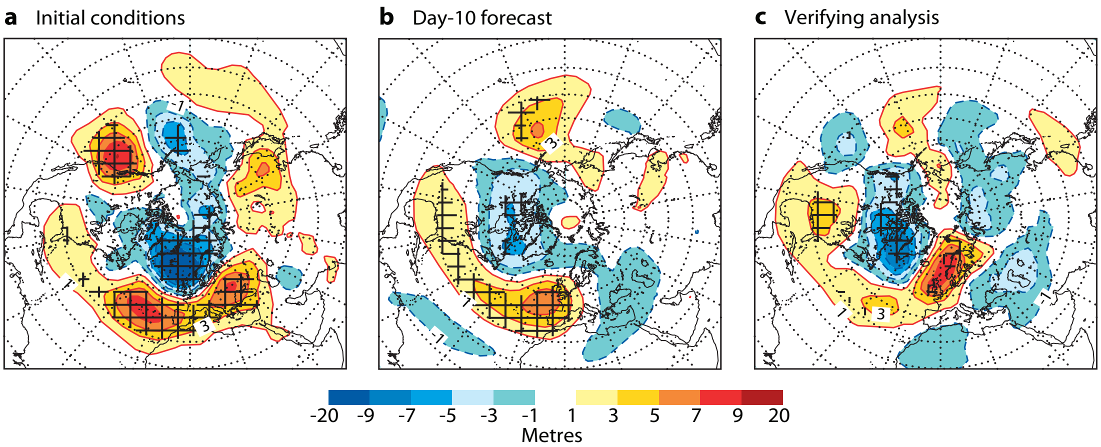



**Figure 3** Anomaly composites of 500 hPa geopotential height for the poor forecasts initiated during NAO+ for (a) the initial conditions, (b) the forecasts at day 10 and (c) the corresponding verifying analysis. Hatched shading indicates statistical significance at the 10% level.

| | Day 0 | Day 1 | Day 5 | Day 7 | Day 10 |
|---|---|---|---|---|---|
| **Forecasts with large RMSE at day 10 (poor forecasts)** | | | | | |
| NAO+ | 100 | 81 | 56, 44 | 54, 40 | 37, 21 |
| BL | 0 | 8 | 28, 40 | 35, 53 | 42, 51 |
| NAO- | 0 | 2 | 0 | 2 | 2, 5 |
| AR | 0 | 9 | 16 | 9, 5 | 19, 23 |
| **Forecast with small RMSE at day 10 (good forecasts)** | | | | | |
| NAO+ | 100 | 65 | 40, 35 | 28, 33 | 37, 35 |
| BL | 0 | 24 | 30, 33 | 30 | 28 |
| NAO- | 0 | 2 | 19, 23 | 28 | 23, 21 |
| AR | 0 | 9 | 11, 9 | 14, 9 | 12, 16 |

**Table 1** shows the population in percentage of the four climatological regimes at different time ranges for the good and poor forecasts initiated in NAO+. The numbers in black indicate the forecast values and in red the verification values if they are different.

## Relationship between spread and error

It is possible that some of the forecast failures in capturing the flow transitions from one circulation regime to another are a consequence of an intrinsic low predictability of such events. This can be addressed by considering the variations of spread of the ensemble forecasts in different flow configurations. Consequently, we investigate whether there is a relation between flow changes associated with large forecast errors (such as transition to/from a blocking regime) and large uncertainties measured in terms of spread of the ensemble forecast.

By incorporating uncertainties associated with initial conditions and model formulation into the forecast process, an ensemble of forecasts automatically takes account of flow dependence. For an ideal ensemble that accurately accounts for all sources of forecast uncertainty, the verifying truth should be statistically indistinguishable from the members of the forecast ensemble. Consequently, the spread of such an ideal forecast ensemble should provide an estimate of the forecast uncertainty: cases with large (small) ensemble spread should be associated with large (small) forecast uncertainty. The probability of specific weather events could be reliably specified from such ideal uncertainty forecasts, allowing forecasters and other users to determine the associated risk. Operational forecast ensembles are naturally imperfect and they may require statistical post-processing to generate calibrated probability forecasts for users. Nevertheless, it is interesting to look at the raw ensemble data to assess the ability to capture some fraction of the true forecast uncertainty.

We first show that for the operational forecasts covering the cold seasons 2007–2012 the spread is a good indicator of the expected forecast error. Figure 4 shows a scatterplot of RMSE versus ensemble spread at day 10 for all the forecasts. The ensemble spread distribution is binned into ten equally-populated categories, and the RMSE is averaged over each bin. After this bin averaging, properly tuned spread and error measures should then equate (ignoring observation error), and a perfect ensemble forecast should therefore produce points lying along a 45° line. Indeed Figure 4 shows that the ECMWF ensemble exhibits a good spread-error relationship.

Then, by considering the ensemble spread distribution for all the forecasts initiated in each of the four regimes (Figure 5), we evaluate whether the variability in the ensemble spread exhibits any flow dependency. The spread distribution for the forecasts initiated in NAO- has significantly the smallest mean value according to the Kolmogorov Smirnov test ($p<0.001$). This is consistent with the fact that the NAO- is the regime leading to the most skilful predictions at day 10. On the other hand, the spread distributions for the forecasts initiated in the other regimes are not significantly different from each other. It follows that, for the sample considered, the flow dependency of the ensemble spread is evident only for the forecasts initiated in NAO-.

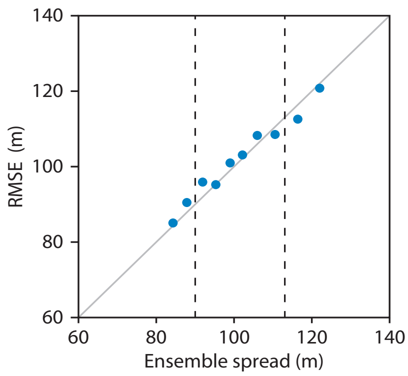



**Figure 4** Scatterplot of RMSE versus the spread for day 10 forecasts. The vertical lines in the scatterplot represent the upper and lower fifth values of the ensemble spread distribution.


doi:10.21957/j1ipst9o

## Summary and outlook

In this study weather regimes have been used to describe the low-frequency atmospheric variability in the Europe-Atlantic area, focusing on the prediction of regime transitions in the late medium range (around day 10) in winter. The regimes leading to either more or less skilful forecasts have been identified.

Overall the model performance, measured in terms of anomaly correlation coefficient, is reasonably good (i.e. correlation greater than 0.6): up to day 9 for predictions initiated in Scandinavian Blocking and Atlantic Ridge regimes, and up to day 10.5 for predictions initiated in either phase of the NAO.

The skills of the forecasts initiated in the NAO+ and NAO- regimes are comparable up to days 10–13. Poor forecasts fail to predict transitions from a strong zonal flow to a blocking pattern, favouring instead the persistence of the zonal circulation. The initial conditions leading to such poor forecasts show a coherent structure over the Pacific reminiscent of the negative phase of the PNA.

Blocking is the regime associated with the least accurate forecasts. Poor forecasts tend to underestimate the persistence of blocking, while overestimating the maintenance of and transitions to zonal flow (NAO+). Consistent with several previous studies, our results show that transition to blocking is also difficult to predict. The least skilful forecasts are mainly associated with unpredicted onset of blocking. It is found that the forecasting of blocking onset is particularly difficult when, at initial time, the westerly jet across the Atlantic is in its southern (NAO-) or northern location (Atlantic Ridge). The Atlantic Ridge is the other regime that leads to lower forecast accuracy. Most of the poor forecasts initiated in the Atlantic Ridge regime missed the transitions to blocking and tended instead to persist in the same regime. Consistent with our results, *Frame et al.* (2011) showed that the ensemble predictions are less skilful when the initial conditions have the jet shifted to the north.

At forecast day 10 the ensemble spread over Europe is a useful indicator of the forecast error. The spread of forecasts initiated in the NAO- regime is significantly smaller than for forecasts initiated in the other regimes. This is consistent with their higher skill.

According to the last five years of forecast data, NAO- is the circulation regime that leads to the most skilful forecasts. Consistent with this, the ensemble spread is generally small for the forecast initiated in NAO- indicating a relatively high level of inherent predictability. Generalizing the present results only on the basis of five cold seasons might be difficult. For example, in Europe, the winter of 2009/2010 was unusually cold and coincided with an exceptionally long occurrence of NAO- events persisting for about two weeks in December 2009 and February 2010. However results from a recent study, looking at a longer dataset from NCEP reforecasts and TIGGE (THORPEX Interactive Grand Global Ensemble) data, provide supporting evidence.

Since this flow-dependent predictability analysis is based on Euro-Atlantic weather regimes, it does not directly provide information on a global scale although to obtain good regime predictions at the medium range a global model is needed. It is also worth noting that there is some level of arbitrariness in considering a specific number of flow patterns. The choice of four weather regimes is a compromise: the aim was to explain the maximum portion of the low-frequency variability in the region whilst using as small a number as possible to increase the representativeness of each regime.

The present study documents the existence of flow dependency in the model's performance in the late medium range. This constitutes the basis for further research into the dynamical and physical processes that initiate regime transitions or favour the maintenance of a specific flow pattern. The ultimate goal is to establish which aspects of the forecasting system should be improved in order to obtain more accurate and reliable predictions at this time range.

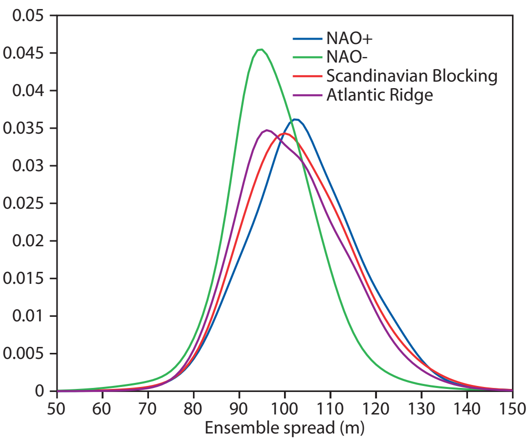



**Figure 5** Ensemble spread distribution at day 10 for forecasts initiated in NAO+, NAO-, Scandinavian Blocking and Atlantic Ridge regimes. The NAO- spread distribution is significantly ($p<0.001$) different from the other spread distributions according to the Kolmogorov Smirnov test.

doi:10.21957/j1ipst9o

## Further reading

**Cassou, C.,** 2008: Intraseasonal interaction between the Madden-Julian oscillation and the North Atlantic Oscillation. *Nature,* **455,** 523–527.

**Corti, S.** & **T.N. Palmer,** 1997: Sensitivity analysis of atmospheric low-frequency variability. *Q. J. R. Meteorol. Soc.,* **123,** 2425–2447.

**Ferranti, L.** & **S. Corti,** 2011: New clustering products. ECMWF *Newsletter No. 127,* 6–11.

**Frame, T.H.A., J. Methven, S.L. Gray** & **M.H.P. Ambaum,** 2013: Flow-dependent predictability of the North Atlantic jet. *Geophys. Res. Letters,* **40,** 2411–2416.

**Leutbecher, M.** & **T.N. Palmer,** 2008: Ensemble forecasting. *J. Comput. Phys.,* **227,** 3515–3539.

**Rodwell, M. J.** & **co-authors,** 2013: Characteristics of occasional poor medium-range forecasts for Europe. *Bull. Am. Meteorol. Soc.,* **94,** 1393–1405.

**Vautard**, **R.,** 1990: Multiple weather regimes over the North Atlantic: analysis of precursors and successors. *Mon. Weather Rev.,* **118,** 2056–2080.

**Woollings, T., A. Hannachi** & **B.J. Hoskins,** 2010: Variability of the North Atlantic eddy-driven jet stream. *Q. J. R. Meteorol. Soc.,* **136,** 856–868.



 doi:10.21957/j1ipst9o