Canadian Ensemble Marine Forecast System

Syd Peel * and Roop Lalbeharry

Environment Canada - Meteorological Research Division/RPN-E, Toronto, Ontario, CANADA

Corresponding author address: Syd Peel, Environment Canada - Meteorological Research Division/RPN-E, 4905 Dufferin, Toronto, Ontario, CANADA.
Syd.Peel@ec.gc.ca

1. Introduction

Wave forecasts from Environment Canada have been produced on an operational basis since 1991 (Khandekar and Lalbeharry 1996; Lalbeharry 2001). These forecasts are deterministic - a single forecast, constituting the best estimate based on the latest observations available when the integration is started, is proffered for various lead times. An ensemble forecast, by contrast, supplies a number of estimates (comprising the ensemble sample) for the same lead time, in an attempt to realistically model the uncertainty intrinsic to any meteorological or oceanographic forecast.

The ensemble mean can improve upon the deterministic forecast by eliminating higher-frequency signals which are often poorly resolved in large-scale numerical weather prediction (NWP) models. Dispersion in the ensemble sample can be used to gauge the uncertainty in the forecast, be it the ensemble mean, the ensemble control, or the corresponding higher-resolution deterministic forecast. Probabilistic forecasts can also be generated directly from the ensemble sample.

Such an ensemble of wave forecasts is now being generated daily at the Canadian Meteorological Centre by integrating the operational wave model, using the wind fields from the Canadian meteorological ensemble forecast system as forcings. From this ensemble are produced a variety of probabilistic forecasts of wave and wind properties, along with their standard deviations as indications of forecast uncertainty, and the ensemble means.

The wave model will be briefly described, along with the Canadian meteorological ensemble forecast system. The probabilistic forecasts obtained from the ensemble of wave forecasts are defined, and the verification of the ensemble wave forecasts is then framed, including a discussion of the observational data used, followed by the presentation of some preliminary results from this verification. We conclude with a discussion of these results, along with an outline of work which is ongoing/remaining.

2. The Models

The principal forcing for the wave model is the "surface" wind (in our case the 10-m wind). Supplying the wave model with the ensemble of different wind forecasts generated by the Canadian Ensemble Prediction System (EPS) (sometimes distinguished hereinafter as the "atmospheric" ensemble) yields a corresponding ensemble of wave forecasts.

a. The Canadian Ensemble Prediction System

The NWP model employed in the Canadian atmospheric EPS is the global GEM model (Côté et al. 1998a,b), integrated on 31 vertical levels with a horizontal resolution of 0.9°. A control and 20 perturbed ensemble members are obtained using an ensemble Kalman filter assimilation cycle (Houtekamer et al. 2009, 2005; Houtekamer and Mitchell 2005). In addition to perturbations of the observations initializing the integration and of the surface boundary conditions, different ensemble members are obtained from different parametrizations of sub-grid scale phenomena. The EPS thus attempts to gauge not only the uncertainty in the forecasts due to the inevitable errors in the initializing observations, but also due to imperfections in the model itself. Ensemble forecasts are produced operationally at 00 and 12 UTC, integrated out to a lead time of 16 days. The 10-m winds are among the standard

outputs of this EPS, available every 3 h out to 10 days, every 6 h for lead times exceeding 10 days.

b. The Wave Model

The WAM 4.5 wave model (WAMDI GROUP 1988; Janssen 1991) is currently used to generate the operational (deterministic) wave forecasts at the Canadian Meteorological Centre (CMC), and this wave model has been adapted to the ensemble wave forecast system. This implementation of the model allows for bottom friction but no depth refraction, diffraction, reflection, or currents. There are 25 frequency bins, spanning (logarithmically) the interval [0.042,0.41] Hz, and 24 directional bins ($\Rightarrow$ a bin width of 15°) (Lalbeharry 2001). The ensemble wave model is run in two configurations: one on a coarser grid which spans the globe, and a higher-resolution version on a window encompassing Atlantic Canada. The ensemble wave model is currently being run daily, in both of these configurations, using the output of the 00 UTC run of the atmospheric ensemble. A variety of wave properties, along with the 10-m wind speed and direction, are output for lead times out to 10 days, with a frequency of 3 h.

1) GLOBAL CONFIGURATION

The global configuration extends from 70° S to 70° N, at a resolution of 1°. The time steps for advection and for the inhomogeneous source terms in the energy balance equation are both 20 minutes.

2) ATLANTIC CANADA CONFIGURATION

The Atlantic Canada window extends from 75° W to 46° W and from 40°N to 52° N (Fig. 1), with a resolution of 0.1°. The propagation time step is 4 minutes while the source term integration time step is 12 minutes. The boundary conditions are obtained from the global ensemble run.

3. Probabilistic Forecasts

For a given lead time, the ensemble wave forecast system provides a set of forecasts for each predictand. Hence, for example, a given ensemble run supplies a set of forecasts for SWH at 24 h in the future. From this set of forecasts we can compute the probability that SWH > 2 m by tallying those ensemble members forecasting SWH's in excess of 2 m and dividing by the total number of members in the ensemble.

a. Continuous Probabilistic Models - Kernel Density Estimators

The probabilistic forecasts obtained in the manner prescribed above can be rather granular, particularly for smaller ensemble sizes. One method of reducing this granularity, as well as attempting to better estimate the tail(s) of the distributions, is to model the densities with kernel density estimators (KDEs) (Peel and Wilson 2008b; Wilks 2006; Brooks et al. 2003). Construction of a KDE is accomplished by the replacement of the step functions familiar in a typical histogram, which approximates the empirical probability density func-
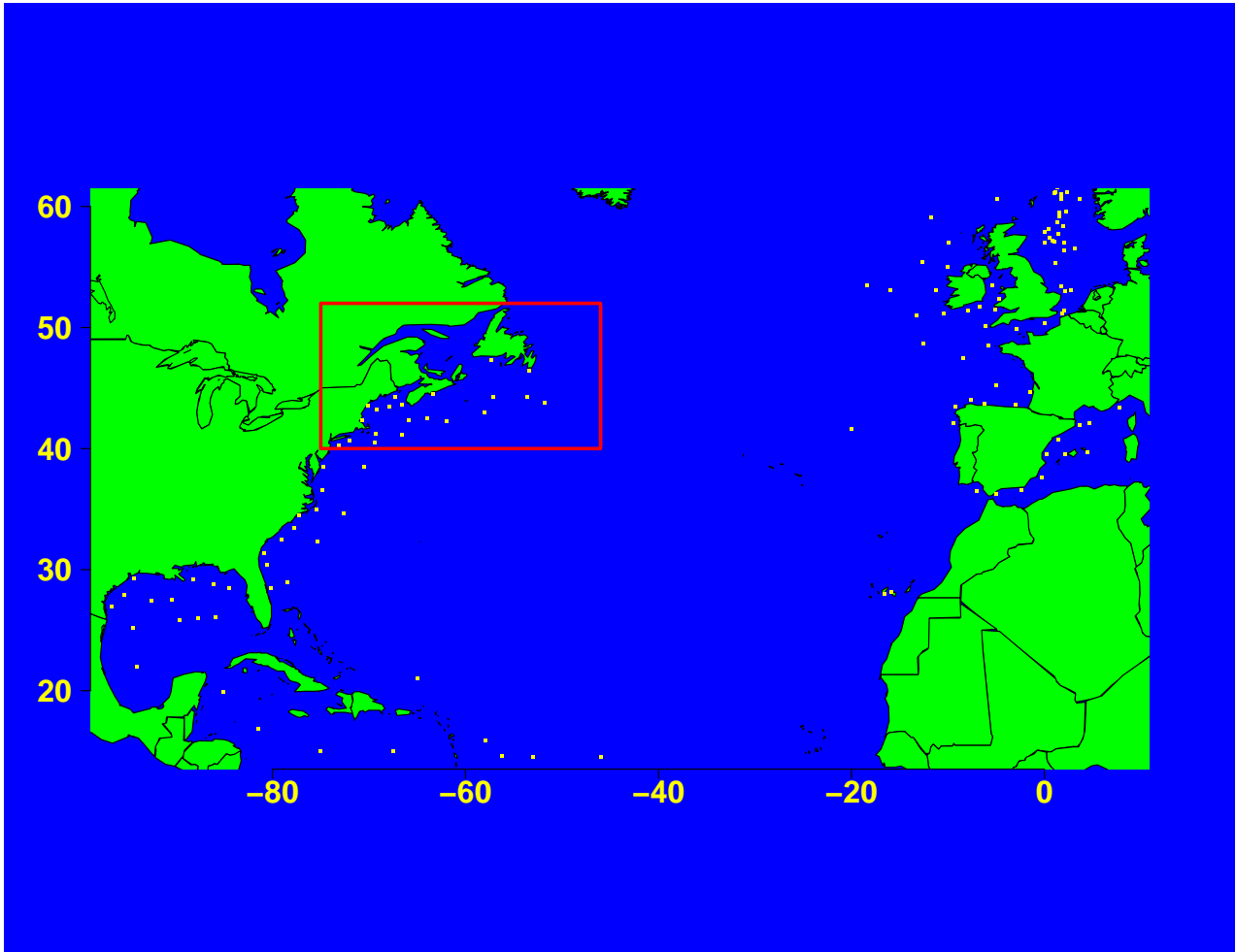
FIG. 1. Atlantic Canada nested window delineated in red (the small yellow squares indicate the positions of in-situ observation platforms).

tion (PDF), with continuous functions (termed "kernels") such as Gaussians (see Fig. 2). Effecting the requisite normalization yields a continuous PDF in place of the sum of Dirac delta functions constituting the discrete PDF obtained directly from the individual forecasts. (For a fuller treatment of the application of KDEs to the modelling of probabilistic forecasts from ensemble output please refer to Peel and Wilson (2008b) and the references therein.)

4. Verification

The preliminary results presented below were obtained from a hindcast of the ensemble waves for the period from April 1, 2008 through March 31, 2009.

a. Observational Data

While in-situ observational platforms abound (albeit mainly near littoral waters) in both the North Atlantic (see Fig. 1) and Pacific, we focussed the verification initially upon the 21 buoys in the Atlantic Canada nested window, identified in Fig. 3. The data for these platforms were extracted from the central file server (CFS) at CMC, where the raw reports
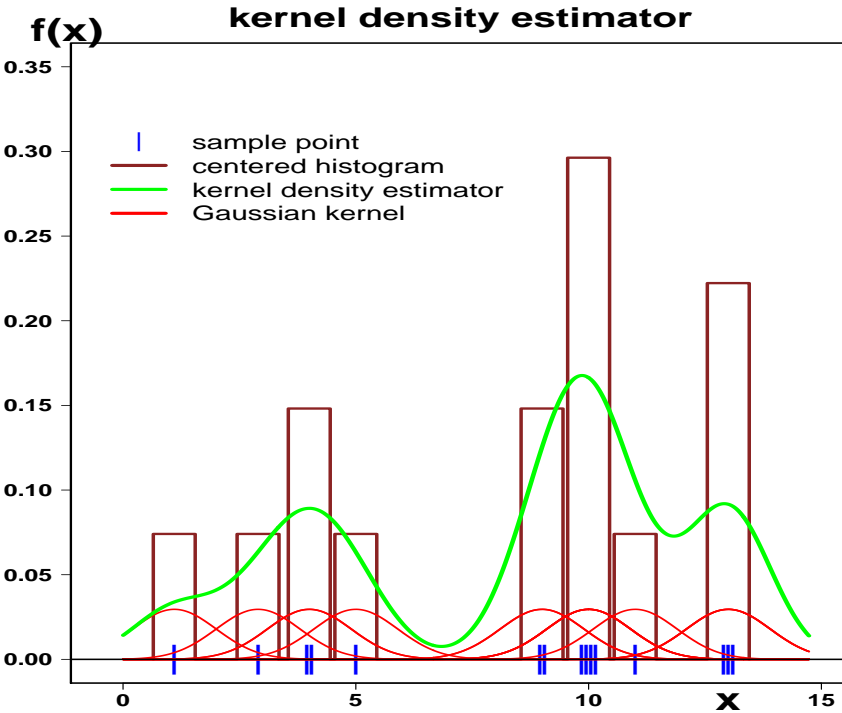
FIG. 2. Illustration of the construction of a kernel density estimator (green) of some unknown density function sampled by the data points (blue) plotted on the x-axis. (The data points have been dithered in order to reveal their multiplicities.) The kernel density estimator is obtained as the sum of the Gaussian kernels (red) centered on the data points, each kernel being weighted by the multiplicity of its corresponding data point.

transmitted to the Global Telecommunications System are archived in binary format. Since the anemometer heights at all of these buoys are $\lesssim 5$ m, their reported wind speeds were corrected to 10 m by multiplication with the adjustment factor 1.07 (Lalbeharry 2001; Smith 1988).

b. Preliminary Results

The WAM4.5 as implemented at the CMC can be configured to generate forecasts for a broad spectrum of different predictands, and a significant number of these are included in the outputs from the ensemble wave forecast system. The preliminary verification focussed on significant wave height, peak period, 10-m wind speed and direction - the results presented here are confined to the significant wave height.

Hence in Fig. 4 appear boxplots for the 0 HR forecasts of significant wave height compared against the boxplot for the buoy observations. (The top and bottom of the boxes represent the first and third quartiles, respectively, of the sample while the lines dividing the boxes indicate the sample medians. The upper and lower gates (or "whiskers") are a distance of 1.5 times the interquartile range (IQR) (represented by the height of the "box") from the third and first quartiles, respectively. Beyond the upper gate are plotted individual
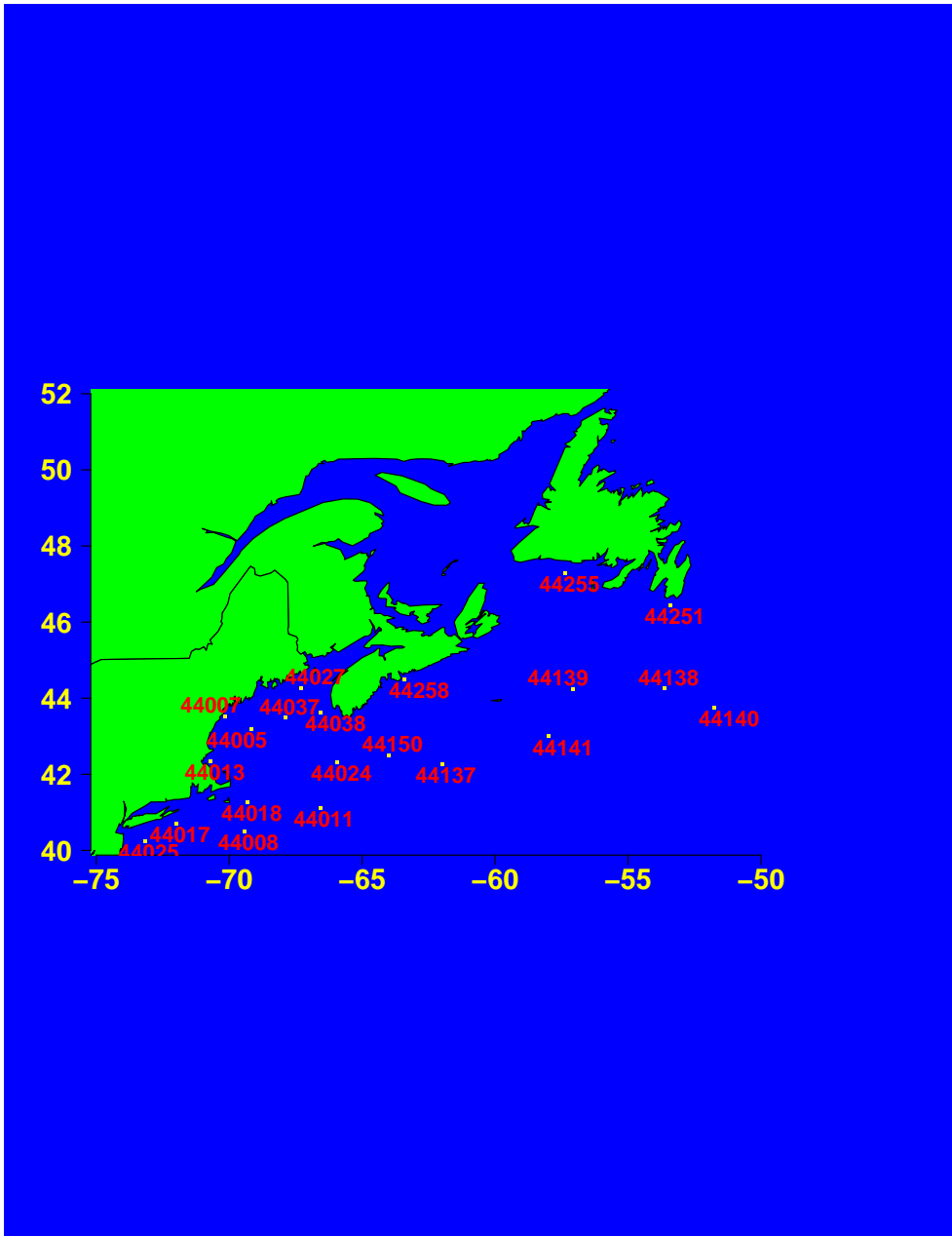
4

Fɪɢ. 3. Buoys (small yellow squares) used to verify wave forecasts on the Atlantic Canada nested window.

"outliers" from the sample (there are no "outliers" below the lower gates due to the fact that the wave height is bounded below by 0).

From these boxplots it can be seen that the control climate is less dispersive than the observed climate. The climates of the perturbed members, on the other hand, possess a dispersion (as evidenced by their IQRs, as well as the pattern of their "outliers" beyond the upper gate) comparable to that of the actual observations, but this greater dispersion appears to have been achieved at the expense of an over-forecasting bias. Moreover, the ensemble forecasts remain insufficiently dispersive, as is revealed in the rank histogram (Wilks 2006; Anderson 1996) plotted in Fig. 5.
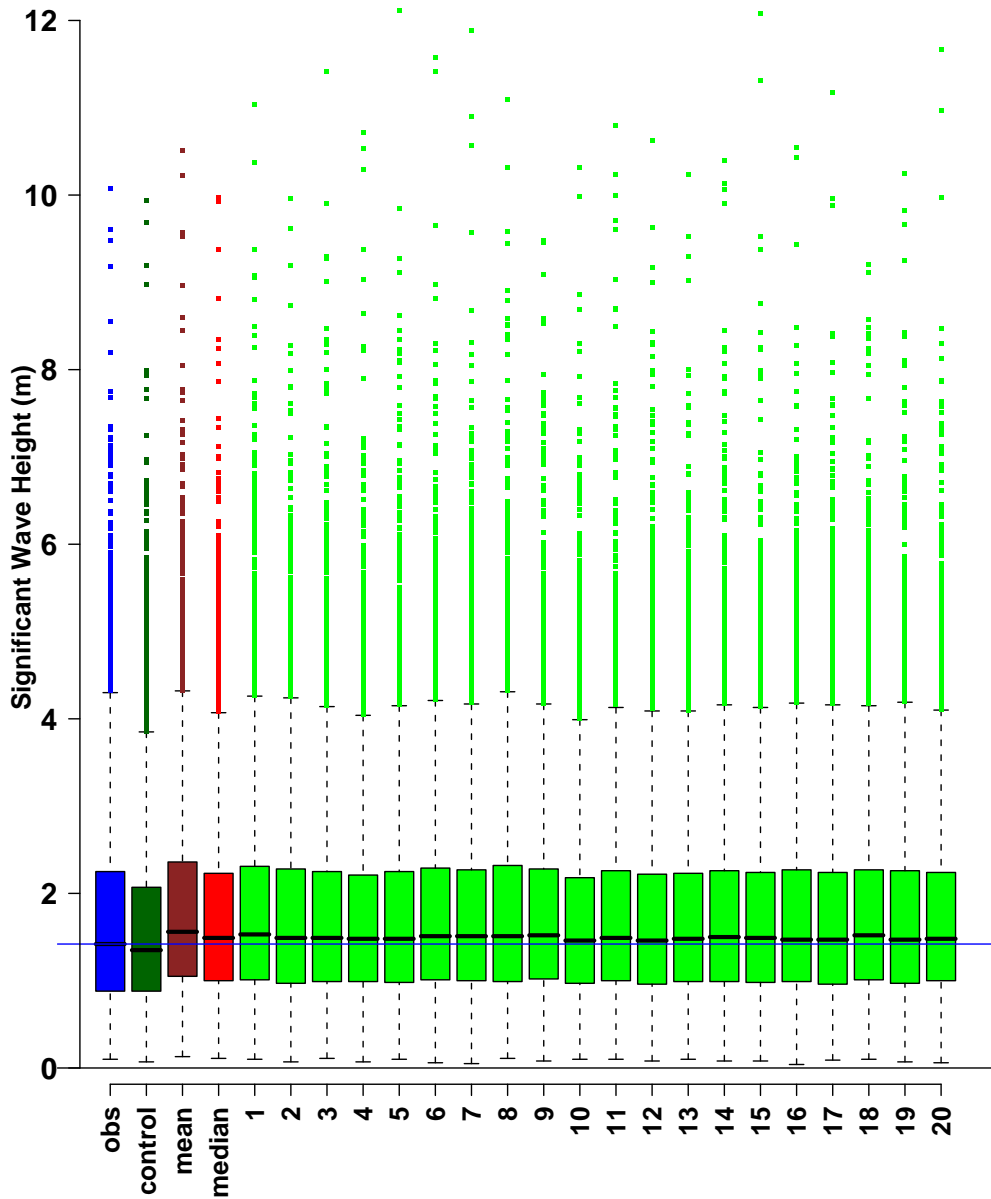
FIG. 4. Boxplots of the significant wave heights observed by the buoys (blue) and forecast by the ensemble control (dark green) and perturbations (green), along with the mean (brown) and median (red) of the ensemble perturbations.

6

This histogram is comprised of 21 bins corresponding to 20 ensemble member forecasts plus the verifying observation, the latter being ranked against the ensemble forecasts. If the observation exceeds all of the forecasts then a rank of 21 is registered, whereas if the observation is smaller than the forecasts a rank of 1 is registered, and similarly when the observation falls somewhere in between forecasts. The histogram thus depicts the distribution of the ranks of the verifying observation amongst the corresponding forecasts. For a perfectly calibrated ensemble system the rank of the verifying observation would be uniformly distributed and the rank histogram would be flat. It is clear from Fig. 5, however, that a disproportionate number of the ranks lie in the 1st and 21st bins, which means that the ensemble forecasts are underdispersive.

Replacing the typical barplots (blue) used to depict histograms with point plots (red), as illustrated in Fig. 5, facilitates the display of rank histograms for a number of lead times, as presented in Fig. 6. From this figure it can be seen that the underdispersion diminishes with increasing lead time, which is typcial for operational EPSs (Candille and Talagrand 2008; Talagrand et al. 1999), but a pronounced over-forecasting bias remains in evidence all the way out to 240 h.

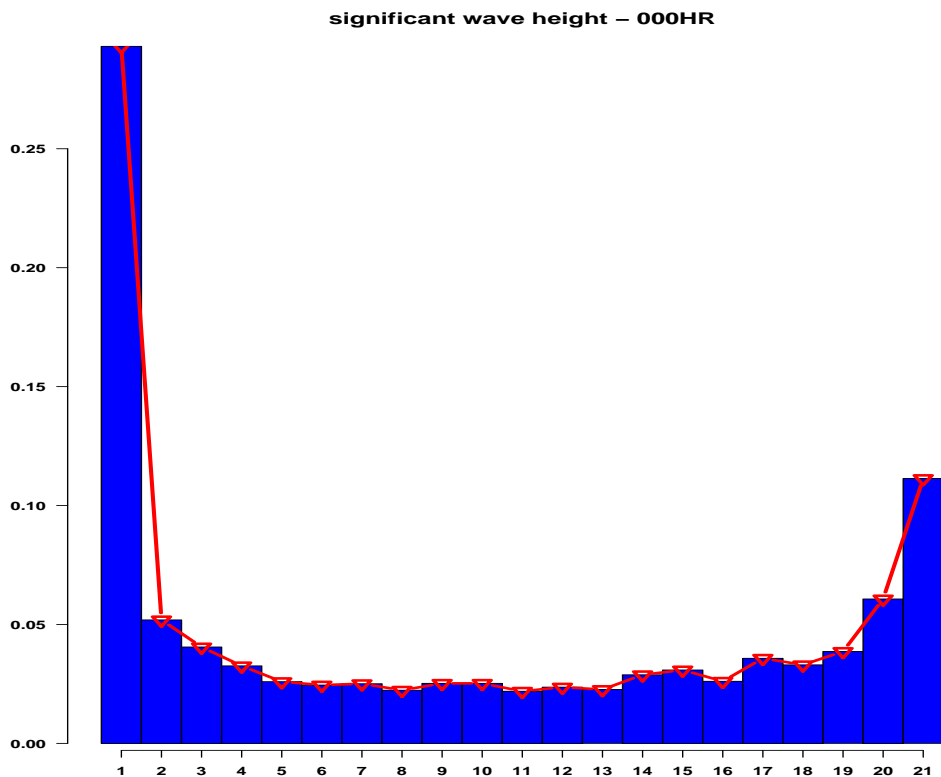**significant wave height – 000HR**

FIG. 5. Rank histogram (blue) for the 0 HR forecasts of significant wave height. The barplot typically used to render rank histograms can be replaced with a line plot (red) through the points whose abscissae are the midpoints of the histogram bins and ordinates are the corresponding normalized bin frequencies.
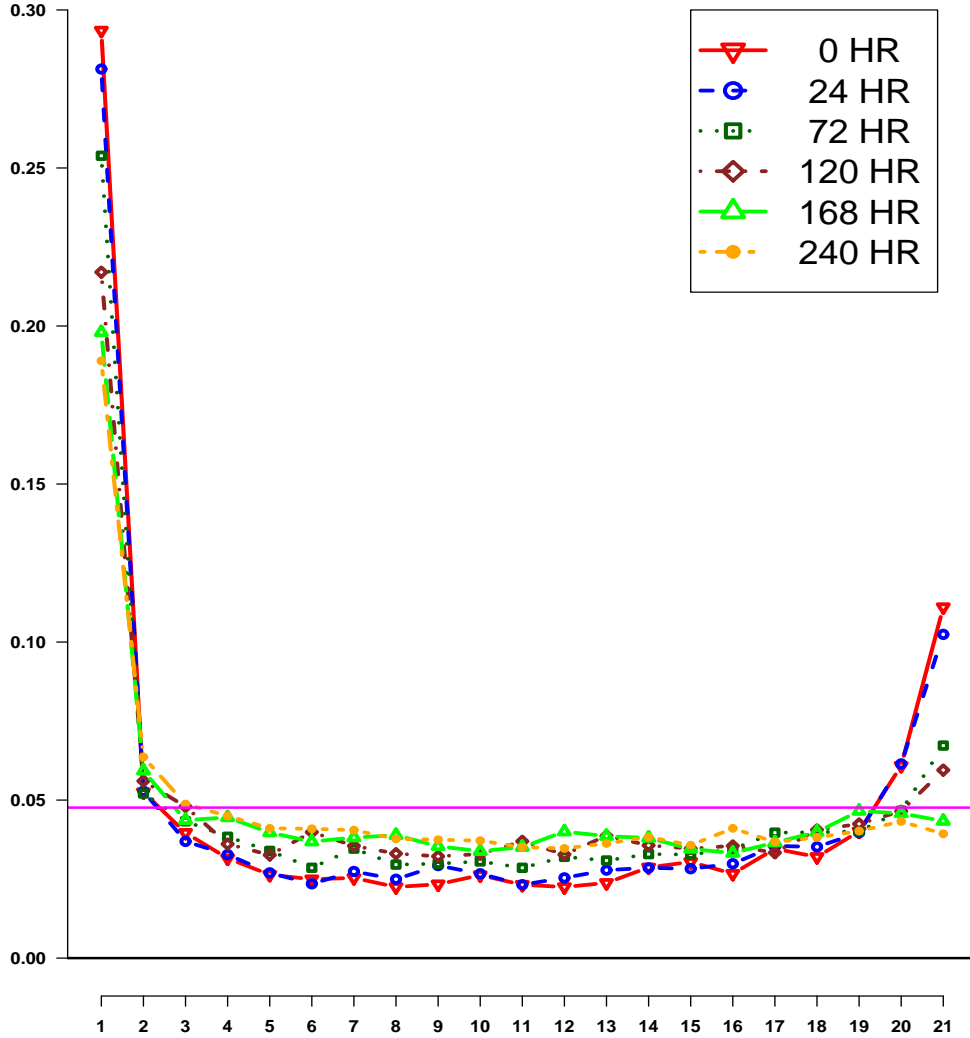
Fɪɢ. 6. Rank histograms for forecasts of significant wave height for lead times of 0 (red), 24 (blue), 72 (dark green), 120 (brown), 168 (green), and 240 (orange) h. Note the pronounced U-shape at 0 h, indicating underdispersion of the ensemble forecasts, which diminishes with increasing lead time, but still showing a significant over-forecasting bias at 240 h. The magenta horizontal line represents the uniform rank distribution of a perfectly calibrated forecast system, the verifying observation having an equal probability $(\frac{1}{21})$ of falling in each bin.

This over-forecasting bias can be seen in the reliability diagram (Wilks 2006; Murphy 1986; Peel and Wilson 2008a) plotted in Fig. 7. Forecasts of low probabilities (the first two bins) of this event (SWH > 1 m) fall above the unit-slope line (representing perfectly reliable forecasts) indicating an insufficient number of lower-probability forecasts. On the other hand forecasts in the highest three probability bins lie well below the unit-slope line, which means that these probabilities were forecast more often than they actually verified. The forecasts from the KDE probabilistic model (blue) evince greater reliability, and hence greater skill, than the raw empirical forecasts (red), particularly in the more heavily populated highest-probability bin.

This superiority of the KDE model over the raw empirical foreasts obtains for all lead times, as can be seen from the plot of their Brier Skill Scores (BSSs) vs. forecast lead time in Fig. 8. (Typical of most skill scores, perfect forecasts score a BSS of 1 while a BSS of 0 indicates no skill with respect to the reference forecast, which was sample climatology in this study. For more details on the BSS in general, and the manner in which it was applied to problems such as the one discussed here, please consult Stanski et al. (1990); Wilks (2006); Peel and Wilson (2008a); Buizza et al. (2005).) This improvement obtained with KDE smoothing is typical for EPSs exhibiting an over-forecasting bias (Peel and Wilson 2008b). As the over-forecasting bias wanes for increasing thresholds the superiority of the KDE forecasts vanishes. The Brier Skill Scores for the probabilistic forecasts at thresholds from 1-9 m, inclusive, are plotted in Figs. 9 and 10. After day 1 the skill of the forecasts exhibits the expected quasi-monotonic decay with increasing lead time, whereas skill increases with increasing threshold up to 3 m, and then starts to drop off with higher thresholds.

5. Discussion

It can be seen from Figs. 9 and 10 that skillful probabilistic forecasts of threshold exceedances for SWH can be obtained for lead times beyond 5 days for thresholds of 6 m or less, and for significantly greater lead times for thresholds less than 6 m. There is considerable noise in the signal for thresholds exceeding 6 m. These forecasts may have some skill for shorter lead times, but this skill would appear to evaporate by a lead time of four days.

Applying kernel density smoothing to the forecasts can result in improvements of almost a day's lead time, but this improvement is significant only for lower thresholds where the over-forecasting bias of the ensemble wave system appears most acute. The over-forecasting bias at all lead times dictates a systematic calibration of the forecasts, and this is currently in development.

Hindcasting is also being extended back to mid-July of 2007, when the last major change to the Canadian operational EPS occurred (Houtekamer et al. 2009). The resulting augmentation of the verification sample will help dampen the noise in the signal for the performance of forecasts at more extreme thresholds, as well as facilitating additional analyses such as seasonal stratifications of the sample in order to gauge the seasonal dependence of forecast performance.
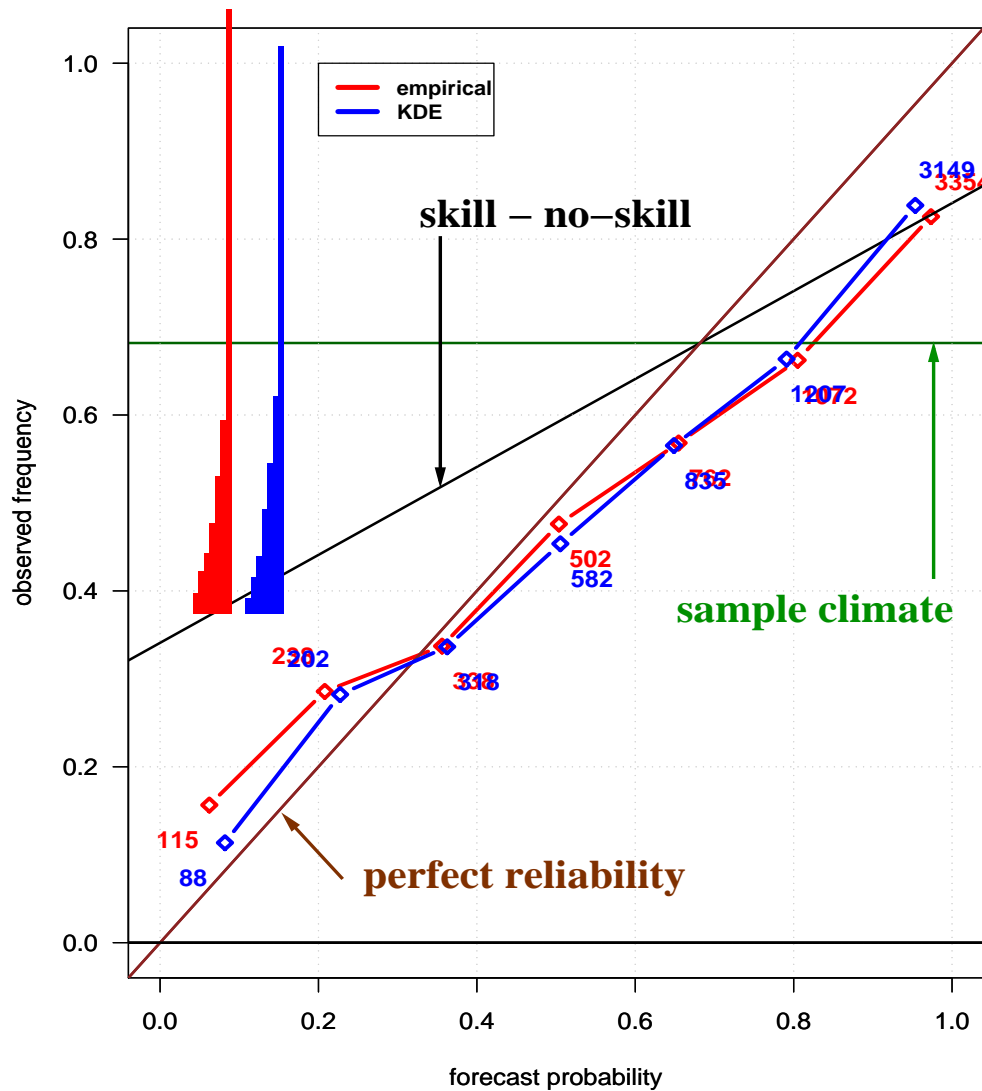
FIG. 7. Reliability diagram for 240 h forecasts of the probability that the significant wave height exceeds 1 m. The forecasts have been stratified into 7 forecast bins, the mean observed frequency for each bin plotted on the ordinate and the mean forecast probability in each bin plotted on the abscissa. Perfectly reliable forecasts would fall on the unit-slope line. The horizontal line just below 0.7 is the observed sample climate, and the line bisecting this line and the unit-slope line separates skillful forecasts from forecasts with no skill. The number of forecasts falling within each bin is plotted beside the corresponding point on the plot, and the distribution of the populations of the bins can also be gleaned from the histograms in the upper left corner. Note the effect of kernel smoothing to increase/decrease the KDE forecast probabilities (blue) with respect to the raw empirical probabilities (red) in the left-most/right-most bins, respectively. In consequence the plot points for the KDE forecasts in these two bins are pushed closer to the unit-slope line, signalling that the KDE forecasts are more reliable than the raw empirical probabilistic forecasts in these bins.
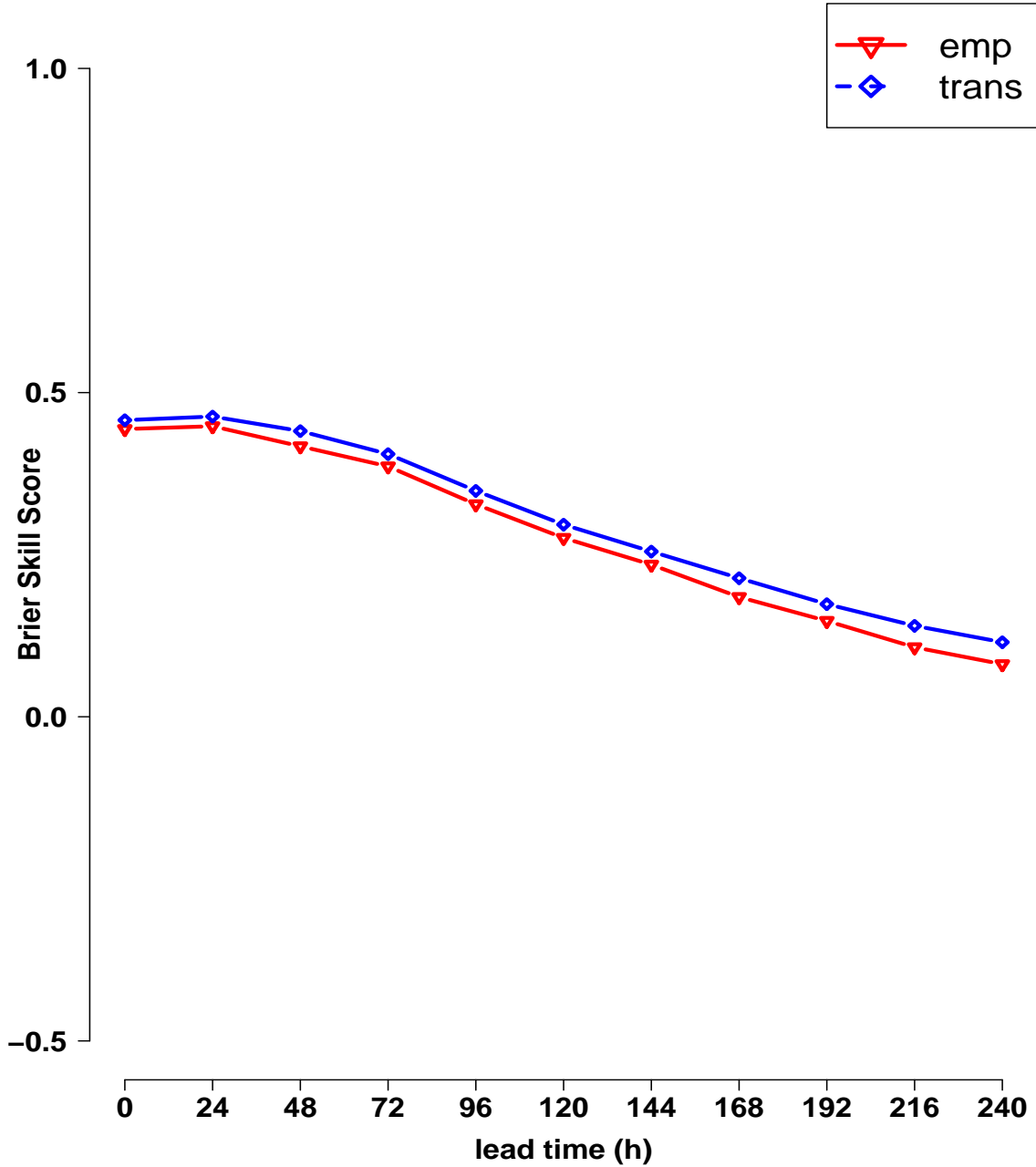
10

FIG. 8. Brier Skill Scores (BSSs) for raw empirical (red) and KDE (blue) probabilistic forecasts of SWH > 1 m, plotted as functions of lead time. Higher BSSs indicate better skill, hence the KDE forecasts are superior to those from the raw empirical model at all lead times.
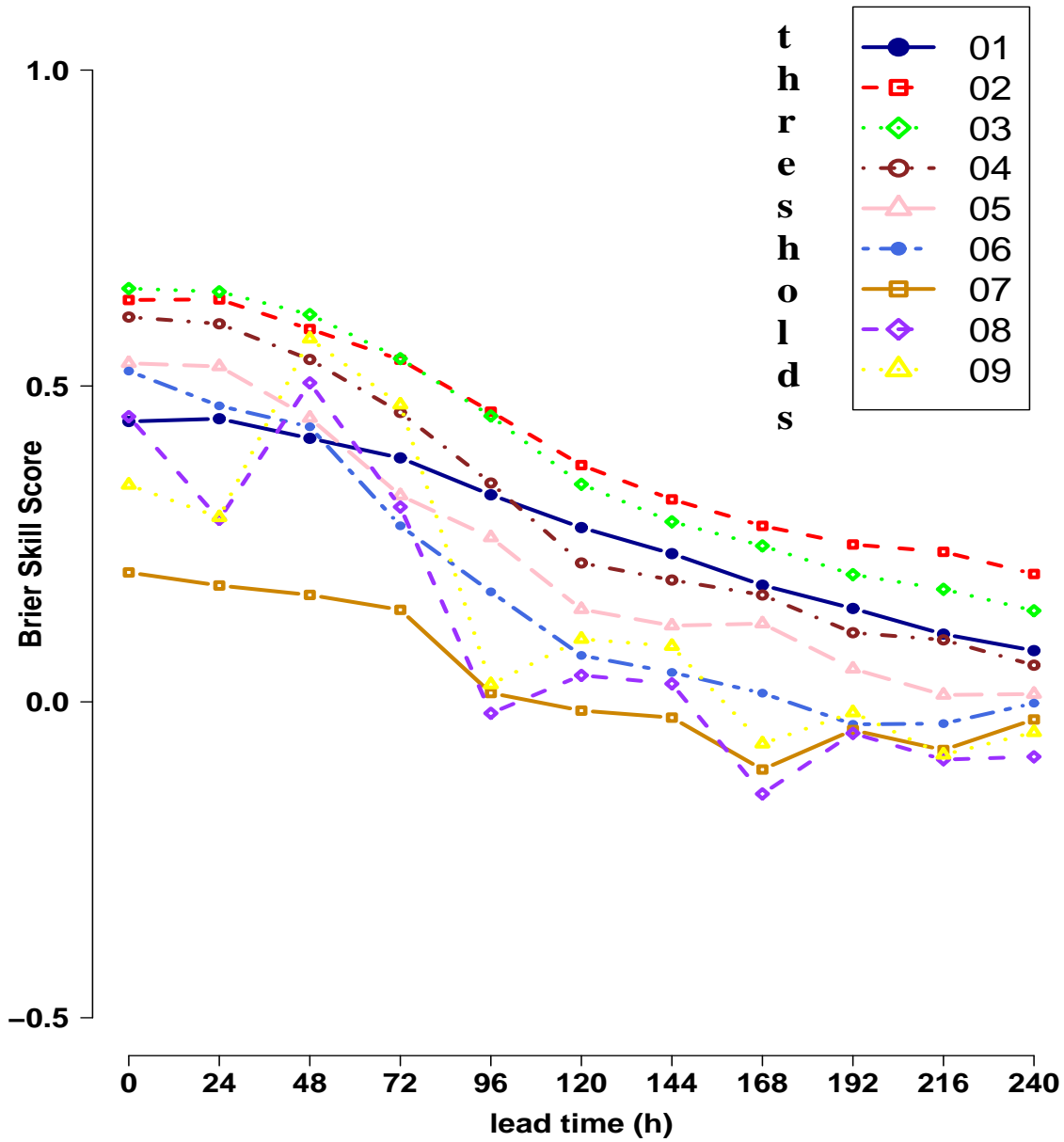
FIG. 9. BSSs for raw empirical probability forecasts of threshold exceedances by the significant wave height for thresholds from 1-9 m, as a function of lead time.
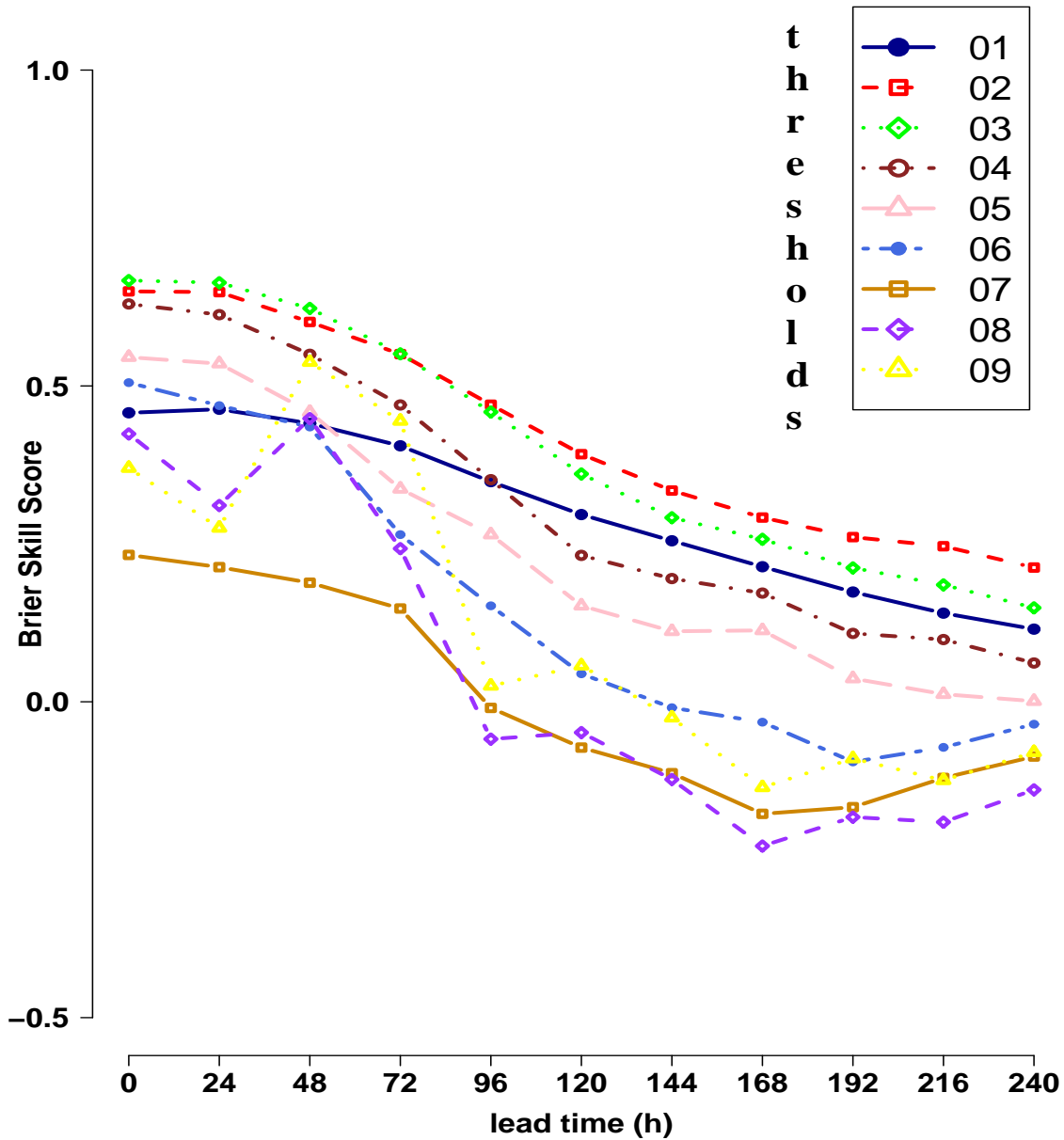
FIG. 10. BSSs for KDE probability forecasts of threshold exceedances by the significant wave height for thresholds from 1-9 m, as a function of lead time.

REFERENCES

Anderson, J. L.: 1996, A method for producing and evaluating probabilistic forecasts from ensemble model integrations. *J. Climate*, **9**, 1518–1530.

Brooks, H. E., C. A. Doswell, and M. P. Kay: 2003, Climatological estimtes of local daily tornado probability for the United States. *Wea. Forecasting*, **18**, 626–640.

Buizza, R., P. L. Houtekamer, Z. Toth, G. Pellerin, M. Wei, and Y. Zhu: 2005, A Comparison of the ECMWF, MSC, and NCEP Global Ensemble Prediction Systems. *Mon. Wea. Rev.*, **133**, 1076–1097.

Candille, G. and O. Talagrand: 2008, Evaluation of probabilistic prediction systems for a scalar variable. *Quart. J. Roy. Meteor. Soc.*, **131**, 2131–2150.

Côté, J., J.-G. Desmarais, S. Gravel, A. Méthot, A. Patoine, M. Roch, and A. Staniforth: 1998a, The Operational CMC-MRB Global Environmental Multiscale (GEM) Model. Part II: Results. *Mon. Wea. Rev.*, **126**, 1397–1418.

Côté, J., S. Gravel, A. Méthot, A. Patoine, M. Roch, and A. Staniforth: 1998b, The Operational CMC-MRB Global Environmental Multiscale (GEM) Model. Part I: Design considerations and formulation. *Mon. Wea. Rev.*, **126**, 1373–1395.

Houtekamer, P. L. and H. L. Mitchell: 2005, Ensemble Kalman Filtering. *Quart. J. Roy. Meteor. Soc.*, **131**, 3269–3289.

Houtekamer, P. L., H. L. Mitchell, and X. Deng: 2009, Model Error Representation in an Operational Ensemble Kalman Filter. *Mon. Wea. Rev.*, **137**, 2126–2143.

Houtekamer, P. L., H. L. Mitchell, G. Pellerin, M. Buehner, M. Charron, L. Spacek, and B. Hansen: 2005, Atmospheric data assimilation with an ensemble Kalman filter: Results with real observations. *Mon. Wea. Rev.*, **133**, 604–620.

Janssen, P. A. E. M.: 1991, Quasi-linear Theory of Wind-Wave Generation Applied to Wave Forecasting. *J. Phys.. Oceanogr.*, **21**, 1631–1642.

Khandekar, M. L. and R. Lalbeharry: 1996, An Evaluation of Environment Canada's Operational Ocean Wave Model Based on Moored Buoy Data. *Wea. Forecasting*, **11**, 137–152.

Lalbeharry, R.: 2001, Evaluation of the CMC Regional Wave Forecasting System Against Buoy Data. *Atmos.-Ocean*, **40**, 1–20.

Murphy, A. H.: 1986, The attributes diagram: a geometric framework for assessing the quality of probability forecasts. *Int. J. Forecasting.*, **2**, 285–293.

Peel, S. and L. J. Wilson: 2008a, A Diagnostic Verification of the Precipitation Forecasts Produced by the Canadian Ensemble Prediction System. *Wea. Forecasting*, **23**, 596–616.

— 2008b, Modeling the Distribution of Precipitation Forecasts from the Canadian Ensemble Prediction System Using Kernel Density Estimation. *Wea. Forecasting*, **23**, 575–595.

Smith, S.: 1988, Coefficients for sea surface wind stress, heat flux and wind profiles as a function of wind speed and temperature. *J. Geophys. Res.*, **93C**, 15467–15472.

Stanski, H. R., L. J. Wilson, and W. R. Burrows: 1990, *Survey of common verification methods in meteorology.*World Weather Watch Tech. Rep. 8. WMO, Geneva, Switzerland, 114 pp.

Talagrand, O., R. Vautard, and B. Strauss: 1999, Evaluation of probabilistic prediction systems. *Proceedings of workshop on predictability, October 1997. European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading, Berkshire RG2 9AX, UK*, 1–25.

WAMDI GROUP: 1988, The WAM model - A third generation ocean wave prediction model. *J. Phys.. Oceanogr.*, **18**, 1775–1810.

Wilks, D. S.: 2006, *Statistical Methods in the Atmospheric Sciences..* Elsevier Academic, second edition, 627 pp.