Evaluation of Medium-Range Forecasts for Hurricane Sandy

LINUS MAGNUSSON, JEAN-RAYMOND BIDLOT, SIMON T. K. LANG, ALAN THORPE, AND NILS WEDI
European Centre for Medium-Range Weather Forecasts, Reading, United Kingdom

MUNEHIKO YAMAGUCHI
European Centre for Medium-Range Weather Forecasts, Reading, United Kingdom, and Japan Meteorological Agency, Tokyo, Japan

(Manuscript received 18 July 2013, in final form 15 January 2014)

ABSTRACT

On 30 October 2012 Hurricane Sandy made landfall on the U.S. East Coast with a devastating impact. Here the performance of the ECMWF forecasts (both high resolution and ensemble) are evaluated together with ensemble forecasts from other numerical weather prediction centers, available from The Observing System Research and Predictability Experiment (THORPEX) Interactive Grand Global Ensemble (TIGGE) archive. The sensitivity to sea surface temperature (SST) and model resolution for the ECMWF forecasts are explored. The results show that the ECMWF forecasts provided a clear indication of the landfall from 7 days in advance. Comparing ensemble forecasts from different centers, the authors find the ensemble forecasts from ECMWF to be the most consistent in the forecast of the landfall of Sandy on the New Jersey coastline. The impact of the warm SST anomaly off the U.S. East Coast is investigated by running sensitivity experiments with climatological SST instead of persisting the SST anomaly from the analysis. The results show that the SST anomaly had a small effect on Sandy’s track in the forecast, but the forecasts initialized with the warm SST anomaly feature a more intense system in terms of the depth of the cyclone, wind speeds, and precipitation. Furthermore, the role of spatial resolution is investigated by comparing four global simulations, spanning from TL159 (150 km) to TL3999 (5 km) horizontal resolution. Forecasts from 3 and 5 days before the landfall are evaluated. While all resolutions predict Sandy's landfall, at very high resolution the tropical cyclone intensity and the oceanic wave forecasts are greatly improved.

1. Introduction

On 30 October 2012 Hurricane Sandy made landfall on the New Jersey coast with a devastating impact on New York City and its surroundings. The worst problems were caused by the storm surge leading to flooding. Farther inland, the precipitation caused problems both because of the large amount and due to the fact that it fell as snow over high terrain. Before making landfall on the U.S. East Coast, the tropical cyclone had caused severe impacts in the Caribbean. A comprehensive investigation of Hurricane Sandy and its impact is given in the report from the National Hurricane Center (NHC), Miami, Florida (Blake et al. 2013).

The landfall position of Sandy was extremely unusual, with only two similar landfalls on the northern U.S. East Coast in the past hundred years: the Great New England Hurricane of 1938 and Hurricane Irene in 2011. The return period of an event like Sandy was estimated to be over 700 years in Hall and Sobe (2013). A detailed investigation of the mechanisms behind the rapid deepening of the cyclone the day before the landfall is given in Galarneau et al. (2013).

The European Centre for Medium-Range Weather Forecasts (ECMWF) has improved the average skill of its tropical cyclone predictions in the past decade (Richardson et al. 2013). However, it is the performance of weather forecasts in individual high-profile cases such as Sandy that can establish and retain the trust and confidence of the public and other forecast users, in a way disproportionate to their overall impact on average skill statistics. It is therefore of considerable interest to document the performance and to examine whether the predictive skill in such cases is sensitive to...
particular aspects of the forecasting system, albeit within the limitations of investigating a single case. Two aspects that were extensively discussed after the Sandy hurricane was the impact of model resolution on the forecasts and the impact of the warm sea surface temperature (SST) along the U.S. East Coast on the development of Sandy. The impact of satellite observations for the data assimilation on forecasts for Sandy is separately reported in McNally et al. (2013).

Here we assess the characteristics of medium-range forecast models at predicting this weather system and investigate the sensitivities to some numerical and physical factors affecting this predictive skill. Three factors are investigated: the general predictability and the difference between forecasting centers, the impact of SST, and model resolution. The focus is on the medium-range prediction of the landfall position in the eastern United States and not on the performance over the Caribbean. A short background to the case is given in section 2. In section 3, we compare the ECMWF high-resolution (HRES) and ensemble (ENS) forecasts and also forecasts from other forecasting centers from The Observing System Research and Predictability Experiment (THORPEX) Interactive Grand Global Ensemble (TIGGE) archive (Bougeault et al. 2010, available online at http://apps.ecmwf.int/datasets/). Recent comparisons of tropical cyclone forecasts from the centers participating in TIGGE are presented in Hamill et al. (2011a,b) and Yamaguchi et al. (2012). In section 4, we explore the impact of the SST and in section 5, the sensitivity to model resolution. Section 6 summarizes the main results.

2. Synoptic situation

In this section, we summarize the evolution of Hurricane Sandy. The intention is not to give a complete picture of the dynamics behind the cyclone, but to give the necessary background to the results presented in this study. A detailed investigation of the case is given in Blake et al. (2013) and Galarneau et al. (2013).

Figure 1 shows the development of the tropical cyclone in the ECMWF analysis [mean sea level pressure (MSLP)] and the precipitation during the first 6 h of the forecasts. The system formed in the Caribbean, where a closed isobar associated with the developing tropical cyclone first appeared in the analysis at 0000 UTC 23 October (Fig. 1b). One day earlier, at 0000 UTC 22 October (Fig. 1a), a convective system was present in the formation area. The storm moved northward and made landfall in Cuba and Haiti on 24 October.

After passing the Bahamas on 26 October, the storm weakened and continued northeast over the western Atlantic. The cyclone started to deepen again and on 29 October the storm began to curve to the west instead of curving toward to the east, as would be typical for the extratropical transition of a tropical cyclone (Jones et al. 2003). The evolution during the last day was influenced by the interaction with an upstream midlatitude trough and its associated cold air mass. The resulting enhanced baroclinicity led to a rapid deepening of the cyclone (Galarneau et al. 2013). At this time, the central pressure of the cyclone in the ECMWF analysis is 947 hPa (very similar to the minimum pressure of 946 hPa estimated by the NHC; Blake et al. 2013).

Figure 2 shows the analyses of 500-hPa geopotential height (Z500) and MSLP below 990 hPa from 0000 UTC 23 October to 0000 UTC 29 October (every second day). The Z500 parameter is chosen to visualize the structure of the midlatitude trough interacting with Sandy during its extratropical transition. The Z500 surface also serves as a proxy for the flow in the environment of Sandy. From 27 October and onward, Sandy started to move into the midlatitudes, downstream of the midlatitude trough located over the United States and west of a subtropical ridge. The narrow subtropical ridge is squeezed between the trough and a cutoff low to the east. During the 2 days before landfall, Sandy started to interact with the trough and moved toward the New Jersey coast. When tropical cyclones and midlatitude troughs interact, it was found in Hanley et al. (2001) that cyclones are more likely to intensify than weaken (if the horizontal scale is approximately equal), as we see during the days before landfall. Furthermore, the advection of air with low potential vorticity into the midlatitude flow by the upper-troposphere outflow of Sandy contributes to downstream ridge building east of Sandy (e.g., Atallah and Bosart 2003; Torn 2010) and excites Rossby waves (Agusti-Panareda et al. 2004; Riemer et al. 2008). During the recurvature and the extratropical transition of a tropical cyclone, the forecast is very sensitive to the phasing between the tropical cyclone and the midlatitude flow (Ritchie and Russell 2007). Here a bifurcation can occur that leads to large forecast errors (Scheck et al. 2011; Grams et al. 2013).

3. Medium-range forecasts

In this section, we investigate the skill of the predictions of Hurricane Sandy in the operational forecasts (HRES and ENS) from ECMWF, together with results from other forecasting centers available in the TIGGE archive. The horizontal resolution of ECMWF HRES is T1279 (16 km) and 91 vertical levels, while the ENS consists of 50 perturbed ensemble members plus 1 unperturbed control member (hereafter referred to as control forecast), all using T639 (32 km) resolution and
62 vertical levels. As is typical in medium-range forecasts, none of the forecasts shown here use a dynamical ocean model, but instead use persisted SST anomalies as ocean boundary conditions. At the time of Sandy, model cycle 38r1 was operational (for documentation see http://www.ecmwf.int/research/ifsdocs/CY38r1/). The other centers in the TIGGE archive used in this study are Environment Canada (CMC), the Japan Meteorological Agency (JMA), the Met Office (UKMO), and the National Centers for Environmental Prediction (NCEP). The properties of the TIGGE ensemble data available in the archive are summarized in Table 1. Please note
that the grid resolution in the TIGGE archive differs from the full model resolution, and interpolation may affect the diagnostics in this study.

Figure 3 shows the MSLP for ECMWF HRES forecasts from 0000 UTC 21 October to 0000 UTC 29 October, every 24 h. All the forecasts are valid at 0000 UTC 30 October, the time of the landfall on the New Jersey coast. The figures include the cyclone track from the analysis (red) and the forecasts (blue) together with a symbol for the cyclone center at landfall time (hourglass symbol for analyses, square for forecasts). The color of the center symbol represents the depth of the center.

The HRES forecast from 9 days before the landfall (0000 UTC 21 October) has a cyclone over the Atlantic, but on an eastward track that does not lead to landfall (Fig. 3a). The forecast from one day later (0000 UTC 22 October, Fig. 3b), predicts a landfall very close to the observed landfall, but 12 h too late. The forecast from 1200 UTC 22 October (7.5 days before landfall, not shown), has a timing error of 2 days and with a landfall too far north along the coast. From 7 days before the landfall and onward, the storm is consistently forecast to make landfall on the northern part of the U.S. East Coast, with the main error in the timing of the landfall and the path of the cyclone before the landfall. The forecasts from 25 and 26 October have the landfall point somewhat too far south, but with a good timing.

While the HRES forecast gives a reasonably consistent picture from 7 days prior to landfall, the forecaster needs to know the degree of confidence one could have in such a forecast. The ENS forecast is designed to provide an estimate of the confidence by examining the diversity among the ensemble members. Figure 4 shows cyclone tracks from the ensembles initialized at 1200 UTC 24 October, and Fig. 5 shows the ensembles initialized 12 h later (0000 UTC 25 October) from ECMWF (Fig. 5a), UKMO (Fig. 5b), NCEP (Fig. 5c), and CMC (Fig. 5d). The cyclone tracks are obtained from the tropical cyclone tracker described in Vitart et al. (1997, 2003). The results are based on all perturbed ensemble members. Please note that the ECMWF ensemble has more than twice as many ensemble members as the other centers (see Table 1). The JMA ensemble is not included as the ensemble is only run at 1200 UTC. Each line represents the cyclone track from one ensemble member. Blue lines are cyclone tracks that are counted as hits, defined as a track within a 300-km radius from the observed landfall; the brown line indicates eastward tracks (not

![Fig. 2. ECMWF analyses of Z500 (dashed, intervals of 50 m) and MSLP below 990 hPa (solid, intervals of 5 hPa).](image)

**Table 1.** Ensemble configuration for different forecasting centers, available from the TIGGE archive.

<table>
<thead>
<tr>
<th>Forecasting Center</th>
<th>Horizontal resolution</th>
<th>Vertical resolution</th>
<th>No. perturbed members</th>
<th>Grid in TIGGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>ECMWF</td>
<td>T_{639} (32 km)</td>
<td>62</td>
<td>50</td>
<td>0.5625° × 0.5625°</td>
</tr>
<tr>
<td>NCEP</td>
<td>T254 (55 km)</td>
<td>42</td>
<td>20</td>
<td>1.0° × 1.0°</td>
</tr>
<tr>
<td>UKMO</td>
<td>0.55° × 0.83° (60 km)</td>
<td>85</td>
<td>23</td>
<td>0.833° × 0.556°</td>
</tr>
<tr>
<td>CMC</td>
<td>66 km</td>
<td>74</td>
<td>20</td>
<td>1.0° × 1.0°</td>
</tr>
<tr>
<td>JMA</td>
<td>T1319 (64 km)</td>
<td>60</td>
<td>50</td>
<td>1.25° × 1.25°</td>
</tr>
</tbody>
</table>
reaching land), while gray tracks are neither of those two (usually making landfall too far north). The squares show the position of the cyclone at 0000 UTC 30 October (the time for the observed landfall) for each ensemble member with the color of the symbol indicating the depth of the cyclone. The figure also includes the observed track from NHC (black) and the landfall location on the New Jersey shore is shown with an hourglass symbol.

For the ensembles from 1200 UTC 24 October, the ECMWF has a broad spectrum of different forecast
solutions. For the tracks approaching the New Jersey coast the general feature is that most of the members predict a landfall that is too late. However, the observed track is well captured within the ensemble envelope. The UKMO ensemble has no member on an eastward track for this time, but there are a number of members with a landfall point north of Boston, Massachusetts. The JMA ensemble (not shown) has 28 members counting as hits and 11 on the eastward track (out of 50 members).

For this initial time (1200 UTC 24 October) the ensemble from NCEP has the largest percentage of members indicating Sandy’s landfall (hits), while in the CMC ensemble most members steer Sandy to the northeast. For the ensembles initialized 12 h later (0000 UTC 25 October), the opposite holds true. Here the NCEP ensemble only has five members that count as hits, while most of the CMC members are hits. The NCEP ensemble has most of the members indicating a landfall too far north. These are examples of inconsistencies in consecutive forecasts, which are undesirable and troublesome for forecasters.

To summarize the ensemble properties for all initial times, Fig. 6 shows the strike probability (Fig. 6a), the
The forecasts are verified against the estimations from the NHC. The strike probability is defined as a trajectory passing (at any time) inside a 300-km radius of the observed cyclone position at 0000 UTC 30 October. This diagnostic is therefore not sensitive to timing errors in the landfall. The other three quantities are calculated as the mean of the members that are counted as hits, in order to avoid members on other types of track. The center pressure and maximum wind speed are calculated for the time step where the track is closest to the point of landfall. A minimum of five ensemble members counted as hits are required here to be included in the plots. Therefore, the number of data points in the figure differs among the centers.

The strike probability increases as expected as the initial date of the ensemble approaches the landfall time. Around 25 October, large inconsistencies between consecutive ensembles for NCEP (black, dotted), CMC (black, solid), and to some degree UKMO (black, dashed) appear (see also Figs. 4 and 5). Overall, the forecasts from ECMWF (gray, dashed) are more consistent between initial times. For the ensembles initialized 2–3 days before
the landfall, the NCEP ensemble shows the highest strike probability. The reason for the smaller probabilities for the other centers is due to a larger ensemble spread. For example, the larger ensemble spread for ECMWF caused some members to miss landfall by more than 300 km. On average, the ECMWF ensemble spread is well tuned with regard to tropical cyclones while the NCEP ensemble has too little spread (Hamill et al. 2011a; Yamaguchi and Majumdar 2010). Although the ensemble spread cannot be evaluated for a single case, we speculate that the inconsistencies between subsequent forecasts for the NCEP ensemble may be a consequence of too little ensemble spread (too little ensemble spread leads to similar evolutions in the ensemble members).

For the error in the central pressure, all ensembles have a positive bias (too weak cyclones). For the central pressure the resolution of the forecasts in the TIGGE archive plays a role. The difference between 0.5° and 1.0° resolution for the ECMWF ensemble is found to be 1.7 hPa, averaged over the 6 initial times closest to the landfall. The worst positive pressure bias is exhibited by the UKMO ensemble (black, dashed), with a bias twice the value of ECMWF and NCEP. For the maximum wind bias, all centers underestimate the strength. Here the UKMO results are not much worse than CMC and ECMWF, even if the pressure bias is higher. The lowest wind speed bias occurs for the JMA (gray, solid) and NCEP ensembles.

Regarding the timing error, CMC has the fastest moving cyclones with too early landfalls, while ECMWF and UKMO have in general the slowest moving cyclones. For ECMWF, this has been a long-standing issue with too slow moving tropical cyclones even if it has improved during the last years (Richardson et al. 2013).

To understand the differences in the synoptic situation between good and bad ensemble members, composites of forecast errors from the members in each category have been made. The criterion for a good member is the same as above (to be counted as a hit, blue trajectories in Figs. 4 and 5) and a bad member is defined by a minimum position error of more than 600 km and a trajectory end point east of the point of the minimum error (brown trajectories in Figs. 4 and 5). By compositing the errors among each group of members, we can study common error features. We also apply
a significance test (95% level for the Student’s t test) for the structures, to see if they are significantly different from zero. We focus on ECMWF ensembles from 1200 UTC 24 October and 0000 UTC 25 October. The choice of initial times for this diagnostic is based on the last initial times when the ensemble had more than five members on the eastward track.

Figure 7 shows composites of forecast errors for members that counted as hits (left) and members that appeared on the eastward track (bad members, right). The composites are made for forecasts from 1200 UTC 24 October (top panels) and 0000 UTC 25 October (bottom panels). The error shown is for Z500 and is valid at 0000 UTC 26 October (1 and 1.5 days into the forecasts, respectively). The bifurcation of the tracks takes place between 27 and 28 October (1–2 days after the time step shown in the plots). The number of ensemble members in each category is shown in the caption of each panel.

Comparing the errors of the members that counted as hits (left) and the members ending up to the east (right), the bad members have larger negative errors connected to the weak subtropical ridge west of the cutoff low and east of the cyclone track. A negative error is associated with a weaker extension of the ridge toward northwest, steering the cyclone on an eastward path. This difference in error between the hits and misses is strongest for the forecasts initialized at 0000 UTC 25 October. The 300- and 700-hPa levels have also been investigated (not shown) and the difference is apparent for these levels, too. This result suggests that small perturbations in the strength and position of the subtropical ridge determined whether the cyclone was going to make landfall or move eastward over the Atlantic. There is no clear difference in the depth of the cyclones between the two categories. We also investigated, for longer lead times (2–4 days into the forecasts), the error structures connected to the midlatitude trough that propagated from the west. For the 0000 UTC 25 October forecast we find a weaker trough for the bad members than for the good members (already present in Fig. 7), but no similar structure for the forecast from 12 h earlier.
To investigate the sensitivities around the 25 October further, we have calculated singular vectors targeted on Sandy, for the initial time at 0000 UTC 25 October. The leading initial singular vector is the structure in the atmosphere that will have the largest impact after (in our case) 48 h for a predefined metric, based on the tangent linear assumption. Here, the metric is the total perturbation energy inside a box around the cyclone. ECMWF singular vectors targeted on tropical cyclones are operationally used to generate the perturbed initial conditions of the ensemble (Puri et al. 2001; Barkmeijer et al. 2001). The resolution of the singular vector calculations is T42. Hence, they are focused on the impact of the large-scale flow on the tropical cyclone (Peng and Reynolds 2006; Lang et al. 2012).

Figure 8 shows the vertically integrated total energy of the leading initial (top) and evolved (bottom) singular vector targeted on Hurricane Sandy. The observed track for Sandy (black line) and position at the valid time of the singular vector (black triangle) are included as a reference together with the geopotential height averaged for 700, 500, and 300 hPa from the analysis, also at the valid time for the singular vector (0000 UTC 25 and 27 October, respectively). At initial time, the most sensitive areas are found north and northeast of Sandy. The sensitivities are mainly connected to the subtropical ridge. It shows that a small perturbation in this structure will have a large impact on Sandy two days later (cf. the evolved singular vector in Fig. 8b). For the evolved singular vector, a part of the structure stretches into the midlatitudes northwest of Sandy and is associated with the outflow of the tropical cyclone that starts to interact with the midlatitude flow (Lang et al. 2012). The main part of the evolved singular vector is associated with Sandy itself, as expected for our choice of metric. The result from the singular vector analysis confirms the result from the error composites: at this stage Sandy was most sensitive to small perturbations in the subtropical ridge, in particular in the northern extent. A weaker ridge should lead Sandy on a more eastward track and increase the likelihood of not making landfall on the U.S. East Coast.

4. Impact of sea surface temperature

During the lifetime of Hurricane Sandy, the SST along the U.S. East Coast was unusually warm (Fig. 9c). In this section we investigate the impact of this anomaly on the forecast of Hurricane Sandy. The increased sea surface temperature leads to increased evaporation from the surface. A study of the water and energy budgets in tropical cyclones can be found in Trenberth and Fasullo (2007) and a sensitivity study of Hurricane Katrina to SST in Trenberth et al. (2007).

The operational setup of the ensemble prediction system in 2012 used persisted SST anomalies for the first 10 days of the forecasts. The operational ensemble will therefore be referred to as SST-Ano in this section. To test the sensitivity of the forecasts to the SST field, an experiment has been run with the SST replaced by the SST climatology (experiment SST-Clim), calculated from the past 20 years in the Interim ECMWF Re-Analysis (ERA-Interim; Dee et al. 2011). Forecasts are run from 0000 UTC 22 October to 0000 UTC 26 October every 24 h. Figures 9a and 9b show the SST field from the two forecasts initialized at 0000 UTC 24 October and valid at 0000 UTC 29 October. The figures also include the MSLP less than 990 hPa to show the position of Sandy at 0000 UTC 29 October in the control forecast. The difference between the SST’s is shown in Fig. 9c. The tropical cyclone moved over water warmer than 26°C.
until 1 day before landfall in SST-Ano, while the climatological SST is cooler (by 1°–1.5°C).

Figure 9d shows the difference between SST-Ano and SST-Clim in the upward latent heat flux from the surface, averaged over the first 5 days and all ensemble members of forecasts initialized at 0000 UTC 25 October. Positive values indicate higher upward heat fluxes in SST-Ano. Following the track of Sandy, the difference in the latent heat flux is between 50 and 100 W m$^{-2}$. The day before landfall, when the cyclone deepens rapidly, the difference reached a maximum of 150 W m$^{-2}$ (in an area where SST-Clim had a flux of 200 W m$^{-2}$). This result shows that the warmer SST led into a significantly higher energy input to the atmosphere.

Figure 10 shows the same as Fig. 6 but comparing the SST-Ano (black) and SST-Clim (gray). For the strike probability, SST-Ano shows, for four out of five initial dates, higher hit rates. For the minimum pressure, the SST-Ano has a lower central pressure for all initial dates and the mean difference is 7.6 hPa. The maximum wind
speeds are also higher in SST-Ano, with an average of 3.6 m s\(^{-1}\). The mean timing error of the landfall increases by more than 7 h for SST-Clim compared to SST-Ano. These three quantities are significantly different from zero with a 95% confidence level (Student’s \(t\) test). We have also investigated the areal coverage of high wind speeds (not shown), and we find that not only does the maximum wind speed increase but also the area covered by surface winds above 20 m s\(^{-1}\) is substantially increased. This should have affected the strength of the storm surge (not investigated in the study).

We have also compared the total accumulated precipitation from the ensemble members that are counted as hits. The precipitation is accumulated for the 24 h after the detected landfall time for each ensemble member. The forecasts are compared to the NCEP stage IV composites (Lin and Mitchell 2005) with hourly accumulations, obtained from a combination of radar and rain gauge data [Next Generation Weather Radar (NEXRAD), see online at http://www.emc.ncep.noaa.gov/mmb/ylin/pcpanl/stage4/]. The precipitation is averaged inside a box of 35°–45°N, 85°–60°W, when the NEXRAD radar dataset has coverage (Fig. 14a). The observed precipitation from NEXRAD is 16 mm (24 h)\(^{-1}\) on average inside the box. For the three forecasts initialized between 0000 UTC 24 October and 0000 UTC 26 October, the average for SST-Clim is 9.8 mm and for SST-Ano it is 13.2 mm.

5. Impact of model resolution

In this section, we investigate the impact of the horizontal resolution on the forecasts for Sandy. We compare the T\(_{1279}\) HRES forecast (which corresponds to a grid resolution of 16 km) and the T\(_{639}\) control forecast from ENS (32 km), together with a forecast using T\(_{319}\) (64 km) and T\(_{159}\) (150 km) resolution. We have also produced two forecasts (from 0000 UTC 25 and 27 October) with a very high-resolution, experimental version of the ECMWF global model (T\(_{3999}\), which corresponds to a grid resolution of 5 km). This model uses a nonhydrostatic core but still has parameterized deep convection. ECMWF’s experimentation with very high resolution is further described in Wedi et al. (2012). There are 91 vertical levels for the T\(_{1279}\) and T\(_{3999}\) simulations and 62 levels for the other simulations.
However, the distribution of levels in these two configurations is similar in the troposphere. All different resolutions were initialized from the same, T1279, analysis and interpolated to the target resolution. Notably, the higher resolutions used a correspondingly smaller time step.

Figure 11 shows the MSLP from the analysis at 0000 UTC 30 October (Fig. 11a) and single forecasts from five different model resolutions, issued 3 days earlier (0000 UTC 27 October). In the plots the cyclone tracks for the analysis (red) and forecasts (blue) are included. Here, we find an almost perfect timing of the landfall for T13999, while T1279 has the cyclone center somewhat too far east. For the lower resolutions, the cyclone is still east of the coast, indicating a slower movement in these forecasts. For the T1319 and T1159, the curve of the cyclone track to the west is not as sharp as in the analysis, while the two highest resolutions turn westward somewhat too early. Furthermore, the central pressure minimum is higher with lower resolution. The same set of experiments were initialized at 0000 UTC 25 October (not shown) with similar results as for the 27 October regarding the difference in cyclone depth and position.

Many processes are candidates for creating deeper cyclones in the higher-resolution runs. For example, we have compared the surface latent heat flux in the forecasts from 25 October and averaged it inside the box of 25°–40°N, 80°–60°W over the 5-day period leading to the landfall (for the forecasts initialized on 25 October). For these forecasts the averaged upward fluxes are for T1159–284 W m⁻², T1319–300 W m⁻², T1639–324 W m⁻², T1279–337 W m⁻², and T3999–350 W m⁻². As the flux increases with increasing resolution, more energy is transferred to the atmosphere from the ocean. This could be caused both by the greater intensity of the cyclones in the higher-resolution runs and/or the reason behind the lower central pressure.

Figure 12 shows forecasts of MSLP and significant wave height,¹ initialized at 0000 UTC 27 October and valid at 0000 UTC 30 October (same forecasts as in Fig. 11). In the figures, two buoys are marked. The time series of the observations (hourly) for the buoys together with the forecasts for 0000 UTC 27 October (3-hourly output) are plotted in Fig. 13. The verified variables are MSLP (top panels), wind speed (middle panels), and significant wave height (bottom panels). The wind speed is measured at about 5-m height and not 10 m, and has therefore been adjusted following Bidlot et al. (2002). The western buoy (hourglass symbol) is located close to New York harbor [National Data Buoy Center (NDBC) New York harbor entrance buoy, 44065] and the second buoy (diamond symbol) somewhat farther east (NDBC Long Island buoy, 44025). The center of the hurricane passed south of the two buoys.

The results for the MSLP show the differences in timing of the hurricane as seen above; with decreasing resolution the hurricane passes later. The timing for the T13999 cyclone is almost perfect, while the minimum pressure is not deep enough. This may be partly because we have used 3-hourly model output and the observations are hourly. For the T1639 and T1319 resolutions, the minimum pressure is better captured, but it could be an artifact of an error in the cyclone track (too northerly track as seen in Fig. 11). For T1319, the wind speed is temporarily lower when the pressure minimum passes, which is a sign of the closeness to the eye of the storm.

For the wind speed we also see the difference in timing. The maximum of the wind speed is well captured for all resolutions for the eastern buoy (somewhat underestimated by T1319). For the western buoy, which is located closer to New York harbor, the maximum wind speed is clearly underestimated by T1319, probably because of the resolution of the coastline (16% of the nearest grid point is considered land for this resolution). For the wave height forecasts, T13999 produces an almost perfect forecast of the peak for the western buoy, while the wave height is underestimated by all other resolutions. This difference is also apparent in the forecast maps of the wave height (Fig. 12). The better forecast for T13999 for this buoy is a combination of a better atmospheric forecast and of the higher resolution in the wave model, which allows a better representation of the bathymetry. The higher resolution of the bathymetry is most apparent close to the coasts. For the eastern buoy the results are more similar regarding the peak of the wave height, although T13999 captures the peak best.

Figure 14 shows the 24-h accumulated precipitation between 0000 UTC 30 October and 0000 UTC 31 October from forecasts initialized at 0000 UTC 27 October, compared to the U.S. radar network NEXRAD (Fig. 14a). In the figures, the MSLP valid at 1200 UTC 30 October is also plotted. Figure 11 indicates a timing difference between the experiments. The timing difference leads to different precipitation patterns; while the cyclones in T13999 and T1279 have hit land, the other resolutions have the main part of the cyclone over sea, which affects the precipitation pattern. For the highest resolution we see good agreement with the radar, due to a well-resolved orography and a good timing of the cyclone.

It may be difficult to draw conclusions regarding the role of resolution from single forecasts for longer time

¹ The ECMWF forecast system is two-way coupled to a wave model of matching horizontal resolution, but does not contain a surge model.
FIG. 11. MSLP forecasts from 0000 UTC 27 Oct to 0000 UTC 30 Oct (3-day forecasts) from different resolutions (starting from the same $T_L1279$ analysis) including the cyclone track [analysis (red) and forecast (blue)].
ranges where chaotic effects play a major role. We have therefore run a low-resolution (T1,159) ensemble for initial dates from 0000 UTC 22 October to 0000 UTC 27 October with 50 ensemble members, which we compare with the operational (T1,639) ensemble. Both the ensembles used the same initial perturbations. Figure 15 shows the same as Fig. 6 but for the operational ENS (T1,639) and the T1,159 ENS. Except for the ensemble initialized at 0000 UTC 25 October where the T1,159 ensemble shows fewer hits, the two resolutions perform in a similar way regarding the strike probability, although the T1,159 ENS has a more “jumpy” behavior. The jumpy behavior could be caused by too little ensemble spread (apparent in other experiments with this resolution, not shown). The T1,159 ensemble from 0000 UTC 25 October has a resemblance to the NCEP ensemble, where the landfall point is too far north for most of the members. Regarding the propagation speed, the result shows a slightly larger negative landfall time bias for the low-resolution ensemble. For the minimum pressure and wind speeds the lower-resolution ensemble shows weaker cyclones (as expected from the earlier results in this section). The differences in minimum pressure, maximum wind speed, and timing error are significantly different from zero with a 95% confidence level (Student’s t test).

6. Summary

In this paper we document the medium-range forecast performance for the landfall of Hurricane Sandy, affecting the New York City area on 30 October 2012, and discuss factors in the model system affecting the forecast skill and realism. We have evaluated ECMWF’s high-resolution and ensemble forecasts, together with ensemble forecasts from other forecasting centers available through the TIGGE archive. Sensitivities to sea surface temperature and the model resolution are discussed.

Hurricane Sandy made an unusual turn toward the west before making landfall. The westward movement
and rapid deepening from 29 October was most likely due to an interaction with a trough over the United States. Therefore, both the prediction of the tropical cyclone track and the U.S. trough were of importance. For the medium-range prediction, another “scene-setting” feature was the subtropical ridge that influenced the cyclone track.

Regarding the ability to predict the event, the results show that ECMWF’s operational forecasts gave an indication of what was to happen 8 days before landfall. From 7 days before (the same day the first closed MSLP isobar of the disturbance that developed into Sandy was present) the high-resolution forecast was consistent in its prediction of the landfall. The results from the ensemble forecasts allowed a significant degree of confidence to be attributed to these forecasts but also showed signs of a too slow movement of the cyclone, which led to a timing error of the landfall. The results are in line with Richardson et al. (2013), showing on average too slow propagation in ECMWF tropical cyclone forecasts, although some improvements are seen in recent years.

The TIGGE archive has been used to compare predictions from different forecasting centers. Comparing the performance, ECMWF had the highest consistency between forecasts from different initial times. We also
found that UKMO seemed to underpredict the depth of the cyclone compared to ECMWF and NCEP, and that the CMC ensemble had a larger timing error. For short-range forecasts, the NCEP ensemble showed the highest probability for a landfall within a 300-km radius of the landfall point. This higher hit rate could be due to a lower ensemble spread in the NCEP ensemble. However, to evaluate the ensemble spread, tropical cyclone tracks need to be evaluated over many cases. Comparisons between different forecasting centers for tropical cyclones can be found in Hamill et al. (2011a,b) and Yamaguchi et al. (2012). Ongoing work will evaluate the TIGGE ensembles for the recent tropical cyclone seasons, and address the question whether a systematic difference in ensemble spread is present.

In the ensemble forecasts from 5 days before the landfall and earlier, a common feature is the presence of ensemble members with an eastward path over the Atlantic, instead of making landfall on the U.S. East Coast. By using composites of the errors in the members that went on an eastward path and comparing them to members that hit the U.S. East Coast, we found a weaker
We have compared forecasts from 5 different model resolutions, spanning from 150-km (T159) to 5-km (T1399) horizontal grid spacing. All simulations were initialized from the same T1279 analysis, which was interpolated to the target resolution. In addition, for the T159 resolution we ran a 50-member ensemble, which we compared with the operational (T639) ensemble. We found that the resolution of the forecast model was not the major factor for determining whether the cyclone was going to make landfall on the U.S. East Coast, but the model resolution had, as expected, a large influence on the strength of the cyclone and consequently the surface winds and significant wave height (see below).

Our results suggest that the sharpness of the turn to the left of the cyclone on the day before landfall is dependent on the model resolution, with a sharper turn for the higher model resolutions. For the TIGGE ensembles, a common feature was landfall points too far north of the observed one. The ensembles are run with a relatively low resolution, and this might explain the error in the landfall position. One could speculate that the strength of the cyclone itself had an impact on the trajectory, by modulating the flow in the steering layer.
However, a better understanding of the mechanisms behind the left turn of the cyclone the day before landfall is beyond the scope of this paper.

For two initial dates, the wind speed, significant wave height and precipitation were compared more in-depth for the different resolutions. With the highest resolution, the extreme wave heights close to the coast were captured. This is a combined effect of an improved atmospheric forecast and a better representation of the bathymetry. Regarding the precipitation, the high-resolution runs were more strongly influenced by orographic features, which were not so well resolved in the low-resolution runs. The largest impact in New York City was due to the storm surge connected to Sandy. While a storm surge model is currently not included in the ECMWF forecasting system, our results suggest great potential in including this in the future. We have not investigated the impact of the resolution in the data assimilation system and all forecasts were initialized from the same T1279 analysis. Our results do not exclude the possibility that the resolution and the quality of the analysis played a major role for the forecast success.

While experimentation with large samples of cases would be required to confirm the general validity of the sensitivities identified here, we believe that it is important to document the forecasting system performance in such rare high-profile events. The reliability of ensemble predictions can only be evaluated by looking at multiple such rare high-profile events. The reliability of ensemble predictions can only be evaluated by looking at multiple such rare high-profile events. The reliability of ensemble predictions can only be evaluated by looking at multiple such rare high-profile events. The reliability of ensemble predictions can only be evaluated by looking at multiple such rare high-profile events. The reliability of ensemble predictions can only be evaluated by looking at multiple such rare high-profile events. The reliability of ensemble predictions can only be evaluated by looking at multiple such rare high-profile events. The reliability of ensemble predictions can only be evaluated by looking at multiple such rare high-profile events. The reliability of ensemble predictions can only be evaluated by looking at multiple such rare high-profile events. The reliability of ensemble predictions can only be evaluated by looking at multiple such rare high-profile events. The reliability of ensemble predictions can only be evaluated by looking at multiple such rare high-profile events. The reliability of ensemble predictions can only be evaluated by looking at multiple such rare high-profile events. The reliability of ensemble predictions can only be evaluated by looking at multiple such rare high-profile events. The reliability of ensemble predictions can only be evaluated by looking at multiple such rare high-profile events. The reliability of ensemble predictions can only be evaluated by looking at multiple such rare high-profile events. The reliability of ensemble predictions can only be evaluated by looking at multiple such rare high-profile events. The reliability of ensemble predictions can only be evaluated by looking at multiple such rare high-profile events. The reliability of ensemble predictions can only be evaluated by looking at multiple such rare high-profile events. The reliability of ensemble predictions can only be evaluated by looking at multiple such rare high-profile events. The reliability of ensemble predictions can only be evaluated by looking at multiple such rare high-profile events. The reliability of ensemble predictions can only be evaluated by looking at multiple such rare high-profile events. The reliability of ensemble predictions can only be evaluated by looking at multiple such rare high-profile events. The reliability of ensemble predictions can only be evaluated by looking at multiple such rare high-profile events. The reliability of ensemble predictions can only be evaluated by looking at multiple such rare high-profile events. The reliability of ensemble predictions can only be evaluated by looking at multiple such rare high-profile events. The reliability of ensemble predictions can only be evaluated by looking at multiple such rare high-profile events. The reliability of ensemble predictions can only be evaluated by looking at multiple such rare high-profile events. The reliability of ensemble predictions can only be evaluated by looking at multiple such rare high-profile events. The reliability of ensemble predictions can only be evaluated by looking at multiple such rare high-profile events. The reliability of ensemble predictions can only be evaluated by looking at multiple such rare high-profile events. The reliability of ensemble predictions can only be evaluated by looking at multiple such rare high-profile events. The reliability of ensemble predictions can only be evaluated by looking at multiple such rare high-profile events. The reliability of ensemble predictions can only be evaluated by looking at multiple such rare high-profile events. The reliability of ensemble predictions can only be evaluated by looking at multiple such rare high-profile events. The reliability of ensemble predictions can only be evaluated by looking at multiple such rare high-profile events. The reliability of ensemble predictions can only be evaluated by looking at multiple such rare high-profile events. The reliability of ensemble predictions can only be evaluated by looking at multiple such rare high-profile events. The reliability of ensemble predictions can only be evaluated by looking at multiple such rare high-profile events. The reliability of ensemble predictions can only be evaluated by looking at multiple such rare high-profile events. The reliability of ensemble predictions can only be evaluated by looking at multiple such rare high-profile events. The reliability of ensemble predictions can only be evaluated by looking at multiple such rare high-profile events. The reliability of ensemble predictions can only be evaluated by looking at multiple such rare high-profile events.


