Experimental Determination of Forecast Sensitivity and the Degradation of Forecasts through the Assimilation of Good Quality Data

ADRIAN SEMPLE, MICHAEL THURLOW, AND SEAN MILTON

Met Office, Exeter, United Kingdom

(Manuscript received 21 June 2011, in final form 19 December 2011)

ABSTRACT

The case of a small vigorous cyclone crossing the United Kingdom on 1 November 2009 is investigated. Met Office Global Model forecasts at the time displayed a marked change in solutions at a forecast range of 72 h, with those at longer ranges being more representative of the correct solution and those at shorter ranges only gradually migrating toward it. The strong bimodal nature of the Global Model forecasts is enough to overwhelmingly dominate the solutions from the Met Office Global Ensemble on which it is based. An investigation into the case is used as a vehicle for developing an experimental method determining the critical location of assimilated data leading to the largest impact on forecast consistency and the origins of the bimodal solutions. It allows the identification of one global positioning system radio occultation (GPSRO) and three surface observations located around the developing low that have conclusively led to the degradation in forecast skill. An assessment of these observations concludes that they are of relatively good quality and correctly assimilated. The case is suggested to be an example of forecast degradation as a result of the addition of growing errors by the data assimilation scheme.

1. Introduction

On 1 November 2009, a relatively small vigorous cyclone, approximately 650 km across, pushed rapidly across central United Kingdom, bringing severe gales and heavy rain to many areas. The cyclone initially became identifiable during 29 October as a shallow feature associated with a frontal wave in the southwestern North Atlantic, but as it approached the United Kingdom on 1 November it entered a rapidly deepening phase, whereupon it deepened by 16 hPa in 6 h. With a further deepening of 8 hPa in the next 6 h, the low achieved a central mean sea level pressure (MSLP) depth of 980 hPa as it crossed the United Kingdom.

In the runup to the event, the $T + 108$ h, $T + 96$ h, and $T + 84$ h forecasts from the Met Office Global Model (see section 2) were producing a solution of a deep cyclone with central MSLP approximately 975 hPa and a trajectory across central Scotland. At $T + 72$ h, however (initialized at 1200 UTC 29 October), there was a marked change in forecast solution with a weaker low (MSLP of 990 hPa) advancing across southwest England. The Global Model essentially persisted with this weaker low on a more southerly trajectory until one day from the event when the $T + 24$ h forecasts (initialized at 1200 UTC 31 October) produced a deeper low (981 hPa) centered on the Irish Sea at the verification time (1200 UTC 1 November 2009).

The signal from ensemble forecasting methods were mixed: the probabilities from the Met Office Global and Regional Ensemble System (MOGREPS) solution jumped with that of the Met Office Global Model, with the solution of a weaker southerly low displayed in the vast majority of the ensemble members. In contrast, the European Centre for Medium-Range Weather Forecasts (ECMWF) ensemble showed a wide spread of possible solutions.

Throughout the period in which the Global Model predicted a shallow low with a southerly trajectory, Met Office forecasters intervened to produce forecast products which represented a deeper low with a northerly trajectory over northwest England. The forecasters’ assessment at the time suggested that, given the apparent low predictability of the situation, sufficient evidence existed to maintain the more vigorous solution of the longer-range Global Model forecasts, rather than switch to the weaker solution.

This paper analyses the event in order to discover the cause behind the sudden change in model solution at the $T + 72$ h forecast range. In the process, a method was
developed by which the critical region of a forecast could be identified experimentally, so that further precise diagnostic investigation can be confidently pursued.

The specification of the Met Office Global Model system is outlined in section 2 with a discussion of the event in sections 3 and 4. The methodology developed in this study is introduced in section 5. Section 6 then uses the identification of the critical analysis area (i.e., the area that has the greatest effect on the forecast) to isolate those observations that led to the change in forecast consistency. Section 7 discusses a critical assessment of the observations that led to the analysis, whereas section 8 draws conclusions. A discussion of the importance of the results in the context of current data assimilation results is presented in section 9.

2. Met Office Global Model description

The Met Office Global Model (Cullen 1993) runs out to 6 days twice a day from the 0000 and 1200 UTC analyses and 1–2 days twice a day from the 0600 and 1800 UTC analyses. The dynamical core of the model is a semi-implicit, semi-Lagrangian, nonhydrostatic formulation (Davies et al. 2005) with the physics schemes as described by Martin et al. (2005) and references therein. The horizontal resolution at the time of the event was 0.5625° longitude by 0.375° latitude (640 × 3481 grid points; N320), which equated to 40-km grid spacing in midlatitudes (and 63 km at the equator). A total of 50 levels were used in the vertical with a model lid at 65 km. A four-dimensional variational data assimilation (4DVAR) system had been used since October 2004 (Rawlins et al. 2007).

3. Synoptic and dynamic overview

a. Synoptic perspective

On 29 October 2009, a broad longwave synoptic pattern had become established across the United States and North Atlantic. In the west, a major upper trough
resided across the central and western United States, with a downstream ridge across the Great Lakes area, extending out into the western North Atlantic. In the central North Atlantic, a major upper trough extended southward to 35°N near the Azores with a 250-hPa jet of 160 kt (1 kt = 0.5144 m s⁻¹) running along its southern flank [Fig. 1a(i)] and marking the boundary between a polar air mass of wet-bulb potential temperature 5°C to the north and a moist tropical air mass of up to 20°C to the south. At this time, a shortwave upper feature ran eastward out of Nova Scotia along the western flank of the major North Atlantic upper trough. This feature drove two shallow frontal-wave depressions within the tropical air mass into the western Atlantic, both with central surface pressures of approximately 1005 hPa [Figs. 1a(i),(ii)]. The forward low in this dual-low system, indicated by an arrow throughout Figs. 1a,b, represents the system from which the depression under study will develop over the following days.

As the major upper trough over the United States extended south, the associated downstream amplification of the synoptic pattern resulted in a rapid veering of the flow over the eastern United States, so that the shortwave upper trough entering the North Atlantic amplified in response [Figs. 1a(iii),(iv)]. With vorticity continuing to be advected into the base of the trough from upstream, the upper trough extended rapidly southward until it collapsed at 35°N, 50°W during 31 October [Figs. 1b(i),(ii)]. The continually backing flow on the forward flank of the collapsing upper trough separated the dual-low system, which had been pushing eastward: the rearmost of the surface depressions was captured by the upper vortex and became vertically aligned beneath it, whereas the forward depression moved on ahead of it into the stronger flow of the North Atlantic upper jet [Figs. 1b(i),(ii)]. This process established a strong southwesterly flow from 20° to 30°N toward the United Kingdom within which the 250-hPa jet strengthened to over 160 kt and drove the embedded preexisting forward frontal-wave depression northeastward [Fig. 1b(ii)].

The shallow surface depression remained on the warm side of the upper jet as it was driven northeastward and
as such was subject to little dynamical forcing from aloft [confirmed through fields of relative vorticity and potential vorticity (PV)], remaining as a shallow feature for much of the period. By 1 November, however, the upper trough in the western Atlantic had completed its disruption, with the southern vortex becoming slow moving at 30°N, 50°W and the resulting northern part becoming more mobile [Figs. 1b(ii),(iii)]. This transferred the axis of the upper trough eastward, where it engaged the depression as it approached the United Kingdom [Fig. 1b(iv)]. The associated strong dynamical forcing produced a late rapid deepening of the surface depression as it approached the United Kingdom, deepening the system by 40 hPa in 24 h from 0000 UTC 1 November and swinging it northeastward across northern England with a central MSLP of 980 hPa.

b. Development from the PV–theta perspective

Here we use PV on the 315-K isentropic surface as a diagnostic (Hoskins et al. 1985) with the 850-hPa wet-bulb potential temperature $\theta_w$ and MSLP to give an alternative viewpoint of this extratropical development using the PV–theta perspective of interacting upper-level (stratospheric) and low-level PV–theta anomalies.

Thursday 0000 UTC 29 October 2009 (Fig. 2a): Over northern Canada, the stratospheric reservoir of high PV (>6 PVU) air can be seen. A large-scale cyclone occupies the central Atlantic with evidence of high PV stratospheric air circulating around this system from previous filaments of PV that have been stripped off the polar vortex. Just east of Newfoundland, the low-level cyclonic feature can be seen in the MSLP field at 35°N, 50°W.
Saturday 0000 UTC 31 October 2009 (Fig. 2a): Two days later, the large-scale low in the Atlantic is slowly decaying and moving northward. In the upper troposphere/lower stratosphere, a filament of high PV air is drawn south into midlatitudes on the western flank of the low pressure system. At the same time, the small-scale low-level cyclonic feature has been advected east and north and has a tight baroclinic zone shown in the 850-hPa $\theta_w$. The European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) airmass red-green-blue (RGB) product at this time (not shown) clearly shows that the evolution of the upper-level PV anomaly is very closely tied to the evolution of the dry, descending stratospheric air mass shown in the satellite products.

Sunday 0000 and 1200 UTC 1 November 2009 (Fig. 2b): One day further on, the high PV anomaly at upper levels has been advected eastward toward the low-level cyclone, which has moved north and east and is beginning to develop a small-scale cyclone with an associated wave in the 850-hPa $\theta_w$ and very tight gradients. This alignment of the upper-level high PV anomaly (and associated tropopause fold) forcing ascent ahead of it, as well as the low-level cyclone with large amounts of warm moist air, is a classic scenario for explosive cyclogenesis. After 12 h, the surface cyclone has deepened rapidly and moved across the United Kingdom, bringing strong winds and torrential rain.

The conservation properties of PV allow the tracing of air masses back along their trajectories so that their geographical origins may be determined. Within this framework, one could conclude that the air mass engaging the cyclone as it approached the United Kingdom and associated with its rapid deepening originated within the area of Davis Strait and northeastern Canada.

A key question for the predictability of such a system is the timing and engagement of the upper-level PV and the low-level cyclone, which are essential ingredients to the explosive deepening. Errors in the evolution of either of these features (either from the model or data assimilation–observations) could potentially lead to low predictability in a deterministic sense, although we might hope that an

FIG. 3. Consistency of MSLP (hPa) forecasts for verification time 1200 UTC 1 Nov 2009 with (top left) $T + 120$ h to (bottom right) $T + 0$ h. Contours show MSLP below 1000 hPa (interval of 3 hPa). Shading indicates the magnitude of the forecast error at the forecast time (hPa).
optimally perturbed ensemble forecast would capture this event in a certain number of members.

4. Forecast consistency and ensemble solutions

Figure 3 shows a forecast consistency diagram showing all MSLP forecasts verifying at 1200 UTC 1 November across the United Kingdom. MSLP has been contoured below 1000 hPa to highlight the depression, with the shading indicating the magnitude of the difference between each forecast and the verifying analysis. The figure shows that the medium-range forecasts at $T + 108$ h to $T + 84$ h had consistently indicated a northerly track for the depression, taking a deep system across Scotland with central MSLP of around 975 hPa. The $T + 72$ h solution from the Global Model run of 1200 UTC 29 October, however, in contrast predicted that the depression would be significantly shallower with a central MSLP of 990 hPa and with a track taking it much farther south across southern England.$^1$ The next three main Global Model runs after the $T + 72$ h forecast correctly edged the solutions sporadically northward, but only to a track taking the depression across central England and maintaining a shallower system than that observed. This maintained a significant and consistent southerly component to the forecast error with an underdeveloped system. Only by the 1200 UTC 31 October $T + 24$ h forecast did the Global Model solution move more correctly into the Irish Sea, whereas the following 0000 UTC 1 November $T + 12$ h forecast failed to improve on it. At this range, the Global Model solution had a 5-hPa error with a location error of 100 km too far south.

Figure 3 shows that the $T + 72$ h forecast represents a significant shift in solutions from a system with a 90-km northerly ($-5$)-hPa error to a system with a 160-km southerly ($+10$)-hPa error. This latitudinal error corresponds to two grid points in the Global Model at forecast ranges greater than $T + 72$ h and four grid points at $T + 72$ h, with the Global Model recovering only slowly in the following forecasts with errors two to three grid points and $+6$–$10$-hPa MSLP errors.

In considering a forecast error such as this, it is important to consider whether the error is understandable from within the context of the inherent predictability of the event, such that the forecast trajectory has increased sensitivity to the initial conditions. Ensembles, in which a forecast model is run a number of times from different perturbed conditions, can be used in this context by observing the degree of spread represented within the ensemble members’ solutions, with large spread indicating a forecast with higher sensitivity. In this scenario, an isolated poor deterministic forecast is merely a consequence of low predictability.

In this case, the Met Office MOGREPS ensemble (Bowler et al. 2008) produced 7 (of 23) members with good solutions at the forecast range of $T + 84$ h (when the Global Model solution was of a deep low with a northerly track), but this dropped to 2 members at the $T + 72$ h forecast range (when the Global Model solution switched to a shallow low and a southerly track). The MOGREPS ensemble members for $T + 84$ h and $T + 72$ h are shown in Fig. 4a. The following $T + 60$ h MOGREPS run (not shown) moved in line with the Global Model solution such that the majority of the members predicted the depression to cross central England; only one member successfully represented the depression to track across northwest England. The MOGREPS ensemble, in fact, moved so drastically away from any significant cyclogenesis event at all that, although at $T + 84$ h a total of 12 (of 23) members indicated a significant depression (members 2, 8, 12, 13, 14, 15, and 18–23), the following $T + 72$ h forecast indicated only 2 members (13 and 22). Whatever factor affected the Global Model at $T + 72$ h, therefore, had an equally drastic effect on the equivalent MOGREPS ensemble, which resulted in a narrow spread in the solutions and a strong bias toward that of the Global Model.

The ECMWF ensemble (Molteni et al. 1996) at $T + 72$ h (Fig. 4b) represented a greater degree of uncertainty than MOGREPS, with a much wider spread of solutions represented for the developing low. This ensemble is therefore more suggestive of a low predictability event, in contrast to the MOGREPS ensemble, which forecasts solutions that have higher predictability but are less consistent with earlier runs.$^2$ However, this does not make so much sense in the context of the Global Model with its strong bimodal behavior before and after the critical data time of 1200 UTC 29 October (Fig. 3), because an unpredictable event is generally accompanied by more varied run-to-run solutions. Further, the magnitude of the model jump and the apparent reluctance of the Global Model in subsequent runs to move away from the solution with the southerly track, coupled with the lack of spread in the MOGREPS ensemble, is more suggestive of a significant error or a bias

---

$^1$ For comparison, the ECMWF deterministic model at $T + 72$ h represented the cyclone as a large area of low pressure to the north-northwest of Ireland and so for different reasons failed to represent the low sufficiently at this time.

$^2$ The different solutions from the two ensembles can be down to any of a number of differences, paramount of which are different driving models, the resolution at which they are ran, different sized ensembles (ECMWF being larger), different initial conditions, and the method by which the ensemble perturbations are derived.
in the Global Model solutions than in a forecast error due to low predictability. This paper therefore explores the bimodal behavior of the Met Office Global Model by assessing the \( T + 72 \) h forecast at which the discontinuity in the solutions occur.

The Met Office Global Model runs on a 6 hourly cycle: two main ("QG") forecast runs at 0000 and 1200 UTC providing forecasts to \( T + 144 \) h and two intermediate (also QG) forecast runs at 0600 and 1800 UTC running to \( T + 48 \) h. There is also a short-period \( T + 9 \) h "update" ("QU") run after each cycle to provide the best possible background field for the next main runs. This is achieved by assimilating additional observations that were not available for the main runs because of data cutoff times imposed by operational requirements. The result of this run cycle is that, although the forecast error has been identified at the QG12 29 October run, the error could in fact have entered the forecast cycle in either of the preceding runs at QG06 or QU06 29 October but was undetected because of the shorter forecast ranges involved.

To isolate the exact time of the model solution jump, the QU06 run was rerun out to \( T + 144 \) h so that the forecast of the storm track over the United Kingdom could be assessed. It was found that the forecast was in line with those preceding it, with a deep low forecast over central Scotland. This conclusively identified the forecast error to have originated in the QG12 29 October Global Model run.

5. The sensitivity grid

With the temporal origins of the forecast error accurately identified, the problem turns to identifying the geographical origins of the forecast error. This is a difficult problem since errors are not necessarily restricted to the development in which the error is primarily identified but may also be a result of error growth upstream amplifying and propagating within the flow (e.g., Persson 2000) and affecting the development of the system under study remotely. In such cases, an investigation into the depression ignoring upstream error developments would be futile, because the critical error had originated elsewhere.

One methodology of approaching this is to examine the time evolution of the error under study and to track it backward through the forecast to its first appearance as an observable error, thereby identifying any error sources feeding the error growth within the downstream depression. This may be achieved through either forecast–analysis differences (i.e., the forecast error) or forecast–forecast differences (i.e., the change in the forecast evolution relative to a previous forecast). These “error tracking” methods were employed in this case with some success and identified three potential sources: 1) developing in situ within the storm itself, 2) the central United States, and 3) northern Canada, with sources 2) and 3) both requiring the downstream propagation of errors within the flow during the forecast. The identification of these upstream sources made an efficient assessment of the initial conditions difficult as the sum total of the area to be investigated became too large.

To conclusively identify the origins of the major error source, a new methodology was devised in which the area that could possibly contain the error at the analysis time was divided into a set of \( 2 \times 2 \) grid boxes. This initial area was designed to encompass all of the error sources identified in the error-tracking analysis discussed above, with an additional area extending eastward and southward away from the storm under study to allow for the influence of observations being assimilated in the vicinity of the storm and directly affecting the analysis there. The resulting grid was termed the sensitivity grid, because it would identify the areas in which the forecast under study was particularly sensitive to the initial conditions. The hypothesis is that given a more northerly track of the depression in the 0600 UTC 29 October model run, the data analysis at some specific location in the following cycle (1200 UTC 29 October) differed sufficiently so as to produce a marked change in model solution to a much shallower depression with a more southerly track. By dividing the model area up geographically and systematically preventing new observations from being assimilated in defined areas, the location in which the greatest contribution to the model jump had occurred could be identified. The \( 2 \times 2 \) sensitivity grid resulting from this process for the 1200 UTC 29 October forecast was labeled 1–4 and is shown in Fig. 5, the area of which covering all of North America and the North Atlantic.

A series of data-denial experiments was then performed, excluding all observational data from each of the sections of the sensitivity grid in turn. A positive signal from a rerun was defined to be the case when the data was omitted from the grid box and the model solution reverted back to the more successful description of the system from the previous QU06 run: this showed that the new initial conditions in this grid box were responsible for the change in model solutions to the poorer shallower system. Conversely, a negative signal was

---

3 The QG06 run does not contribute to the following QG12 run in the Met Office system, because it is the purpose of the QU06 run to provide the background state.

4 In the Met Office system, this involves the creation of a single model rerun for which there are input files specifying criteria for each observation type. For each iteration of the grid, the relevant lines in the input files are changed to facilitate data denial in the specified area.
FIG. 4. (a) MOGREPS MSLP (hPa) ensemble solutions for (top half) $T + 84$ h and (bottom half) $T + 72$ h. The $T + 84$ h run has more successful solutions (members 2, 12, 13, 14, 15, 18, and 19) than the $T + 72$ h run (members 13 and 22). The ensemble runs also show a strong bias toward the corresponding Global Model solution (cf. Fig. 3). (b) ECMWF MSLP (hPa) ensemble solution for $T + 72$ h showing a higher degree of spread than the MOGREPS solution shown in (a).
obtained from a grid box in which nonassimilation of data in that area produced no change to the forecast under study. This showed that the forecast was insensitive to analysis changes in that location and so this area could not be responsible for the change in forecast solutions.

The grid box that then produced the biggest positive signal (i.e., in which the solution showed the move from the northerly to southerly track) was then subdivided again into another $2 \times 2$ grid and the rerun process was repeated (Fig. 5). Regularity of the grid boxes was not necessary; rather, a desire to avoid bisecting prominent dynamically active areas, which by their nature would be sensitive and thus would split the signal from the data-denial experiments. These grid boxes were labeled 4.1–4.4, thus maintaining the information from the first cycle of the sensitivity grid.

The process was repeated for a total of four cycles of the sensitivity grid with further subdivisions of the grid boxes and the labeling convention maintained through to $4.x.y.z$, with $x$, $y$, and $z$ indicating the location of the grid box in each cycle (Fig. 5). The final cycle resulted in a positive signal from an area that was sufficiently small to focus further investigation on and contained only one dynamically active feature.

This new methodology proved successful in identifying the sensitive region and the impacts on the MSLP fields at the verification time of each of the data-denial experiments are shown in Fig. 6, in which the grid box resulting in the largest positive signal is highlighted by a bold border. Results from the first set of experiments (boxes 1–4 in Fig. 5) showed that the major sensitive area for the $T + 72$ h forecast resided in the southwestern North Atlantic (box 4), because removing data from this grid box produced a solution with an excellent verification and accounted for an estimated 90% of the model error. The box to the north of this (box 2), covering the northern North Atlantic, Greenland, and northeastern Canada, also produced some signal and probably accounts for the remaining 10% of the model error.

Note that this analysis immediately allows some conclusions to be drawn about the relative importance of the error sources identified by the error-tracking results. First, the error developing in situ within the storm must be the primary error source because this is the only feature that occupies box 4. Second, it suggests that the error source from northern Canada made a small contribution as this error occupied box 2. Third, the error development in the central United States can be ruled
out because there was no signal of contribution from box 3. Overall, it shows that in this case propagation of errors from upstream is not an important issue and that the majority of the error develops locally.

The sensitivity grid was then repeated twice more on the area occupied by the major contribution in the southwestern North Atlantic to observe if the critical development area could be identified more accurately. Each application of the sensitivity grid method divided the grid box with the biggest positive contribution to the forecast further. It was possible to conclude from these experiments that most of the sensitivity to the forecast resided in a remarkably small box (30°–35°N, 50°–60°W) around the developing low in box 4.1.4.3. This area covered the western flank of the depression, which would develop and cross the United Kingdom 3 days later.

Note that the technique is applied to the grid box giving the largest signal so that smaller contributions to the forecast error may be left outside of the data-denial grid box and thus may be left to contribute toward a degradation of skill in the resulting forecast. This is observable in this case as an incremental decrease in skill when moving through the grid box sequence from box 4 to box 4.1 to box 4.1.4 and to box 4.1.4.3 and in particular in the latter stages of the process. This is useful information because it allows one to infer the extent and degree of sensitivity geographically. The results of the process in this case show that the main area of sensitivity was focused tightly around the western flank of the

![Fig. 5](image-url)
developing low within box 4.1.4.3 and then extended outwards into box 4.1.4. A further halo of minor sensitivity then extended farther northward away from the low, through the flow across Newfoundland and also down through the Davis Strait as indicated by a signal from box 2.

For the purposes of further analysis, we will primarily consider box 4.1.4 as representing the critical development area as this is the smallest box producing the strongest signal at the verification time. However, in some instances it is useful to take the analysis further and consider the smaller box 4.1.4.3 because this contains the origins of the majority of the forecast jump.

6. Identification of observations contributing to the forecast error

Observation coverage within the region of box 4.1.4 was investigated, and it was found that the following observation types were assimilated in the area (shown schematically in Fig. 7):

1) surface observations from two ships (41001, ZCIH7) and a buoy (44919): 41001 and 44919 lay to the south of the low with ZCIH7 to the north;
2) scatterometer winds from SeaWinds, with the western flank of the depression lying largely between two passes of the satellite;
3) winds derived from satellite data from National Environmental Satellite, Data, and Information Service (NESDIS) infrared (IR) and NESDIS water vapor (WV) channels;
4) three global positioning system radio occultation (GPSRO) measurements: two lay to the north and within the confines of box 4.1.4, whereas one lay farther to the south within the confines of box 4.1.4.3; and
5) Special Sensor Microwave Imager (SSM/I) coverage on both sides of box 4.1.4.

A series of data-denial experiments was carried out in order to isolate the contributions from the observation types in box 4.1.4, which contributed to the forecast jump. These experiments involved denying each observation type...
in turn in box 4.1.4 and running forecasts out to $T + 72$ h to observe the impact on the forecast at the verification time; the results of which are shown in Fig. 8. A positive signal was observed when the depression was observed to shift further north in the data-denial run relative to that in the all-data run.

Of these experiments, it was found that only surface and GPSRO observations in the area contributed to the forecast error (the “No Surfscat” run denies surface and scatterometer observations with the results being almost identical to the “No Surface” run, which proves that the impact of the scatterometer data is negligible). The model was then rerun excluding both surface and GPSRO observations combined in the area to observe the combined effects. This combined surface- and GPSRO-denied rerun (Fig. 8) produced a solution that was farther north and therefore exhibited a smaller error than each of the individual observation type of denial reruns and was also very close to the result obtained for removing all observations in box 4.1.4.

With only one of the GPSRO observations occupying box 4.1.4.3 (and therefore being more central to the critical development area), a rerun was performed in order to observe the sole impact of this GPSRO observation on the forecast. These results showed that this observation was largely responsible for the negative impact on the forecast. The combined effect of assimilating only the GPSRO observation and the three surface observations in box 4.1.4.3 is shown in Fig. 8.

The results thus showed that both a single GPSRO measurement and three surface observations in an area in the immediate vicinity of the western flank of the developing low contributed to an analysis that resulted in a poorer forecast than if neither were assimilated and that their combined effect in this location was responsible for the majority of the overall forecast error. With such a precise identification of the origins of the forecast error, it was now possible to carry out an assessment of the four observations involved.

7. Assessment of observations contributing to the forecast error

a. GPSRO observations

The assimilated GPSRO observation in box 4.1.4.3 was assessed through its refractivity and bending angle information (M. Rennie, Met Office, 2010, personal communication) with the data exhibiting a sharp transition to a narrow layer of low refractivity and low bending angle at a geopotential height of 4–5 km and an impact height of 5–6 km, respectively. The observation – background ($O - B$) errors at this height were found to be about three times the assumed observation error, and although this is considered relatively large it is not considered to be grossly

![Fig. 7. Schematic diagram illustrating the data coverage within box 4.1.4. Boundaries of box 4.1.4 and box 4.1.4.3 are indicated.](image-url)
It is not surprising therefore that the observation is within the limits of the quality control rejection criteria. GPSRO observations of this quality are therefore routinely assimilated.

The problem of assessing the quality of this observation further is difficult because it lies in a data-sparse area of the Atlantic, where observations providing similar information are relatively far away. The nearest radiosondes at the time were at 44°N, 60°W and 32°N, 65°W and these show differing degrees of dry air masses aloft. However, fields of relative humidity (RH) show that the GPSRO measurement occurs in the background field within a relatively small-scale dry feature to the northwest of the depression under study, so that other radiosondes nearby are sampling entirely different air masses and are not useful for comparison purposes.

The fact that the dry anomaly within which the GPSRO observation is situated appears in the background field does add some credibility to the observation, however, because it shows both the model and previous observations have both identified its existence. Also, GPSRO exhibits high vertical resolution so the sharp and shallow nature of the dry layer may not be unrealistic. On the other hand, GPSRO errors are known to be larger within the lower troposphere, where more than one satellite signal arrives at the receiver and interference can occur. This results in observed refractivity biases that are low relative to the model (Jensen et al. 2003), which might suggest that the observation could be indicating a layer that was drier than in reality.

On balance, the GPSRO observation is within the accuracy tolerances deemed acceptable and is observing a relatively small dry layer to the rear of the developing low that exists in the model background field. With the known characteristics of the GPSRO observing system, it appears a credible representation of the atmosphere at this location, although it cannot be conclusively corroborated.

**b. Surface observations**

It was discussed earlier that three surface observations existed in the area of box 4.1.4.3 that could have contributed to the poor forecast. The recent history of the agreement
between the observation $O$ and model background $B$ can be used as a way of assessing the quality of observations by calculating time series of $O - B$. Such time series provides $O - B$ departures over often widely varying meteorological conditions and so are an accurate way of assessing the observations' reliability.

Ship ZCIH7 lies to the north of the area and on the northwestern flank of the developing low. This ship shows a pressure bias of −1 to −2 hPa during this period but crucially does not report pressure at the analysis time of 1200 UTC 29 October so this quantity did not contribute to the analysis. The Met Office Global Model also assimilates observed surface winds and temperature. Both these quantities show relatively good agreement with the model, although the wind speed shows a small bias of 4 m s$^{-1}$.

Buoy 44919 lies within the lower southwestern area of the box. The $O - B$ time series shows that the buoy is of generally good quality with a small MSLP bias of −0.5 hPa. At 1200 UTC 29 October, $O - B$ is +0.2 hPa.

The $O - B$ time series of ship 41001 also generally indicates an observation of good quality during the period, with a MSLP $O - B$ of +0.2 hPa at 1200 UTC 29 October, while wind and temperature agreement is generally good. One period of disagreement exists for the observation, however, between 27 and 28 October in which the $O - B$ rises briefly to 4 hPa. This period makes the observation slightly questionable but, judging by the fact that agreement at other times is so good, is more suggestive that the ship passed through a deep low (for example) in which the model was poor rather than the observation.

There is thus no strong indication that the assimilated surface observations in the critical development area are in any significant error. To test this further, data-denial experiments were carried out with all possible combinations of the three surface observations in the area. It was discovered that no single observation or combination of any two observations could contribute to the forecast error but that all three surface observations were required to be present for the surface contribution to the forecast error to be generated.

8. Conclusions

A method has been described that experimentally identifies the area of an analysis that has the maximum contribution to a forecast error. The sensitivity grid method may be viewed as a way of experimentally identifying the key sensitive area of a forecast and showed that an estimated 90% of the sensitivity of the forecast was located within a relatively small area focused around the western flank of the developing low occupying 30$^\circ$–35$^\circ$N, 50$^\circ$–60$^\circ$W in the southwestern North Atlantic. In this study, this area has been termed the critical development area.

The remaining 10% of the sensitivity originated within the northerly air mass over northern Canada, as indicated by a positive signal from box 2 in the sensitivity grid experiments. This area was identified by the error-tracking techniques as one of the three possible sources of error. The area of northern Canada was also identified by the PV analysis as the origins of forecast error, although note that reliance on this diagnostic alone would have resulted in incorrect conclusions being drawn. One could infer that the PV analysis, associated as it is with the traceability of high PV air, successfully identified the role of the upper-level feature that engaged the cyclone later in its development but failed to identify the low-level sensitivity where the majority of sensitivity existed in this case.

Other analysis and forecast errors upstream of the critical development area were identified but were found not to be the dominant errors of the forecast in the verification region at the verification time. Downstream propagation of errors, which can often play an important role in forecast error development, is therefore not an important mechanism in this case.

Observations in the critical development area that have been shown to be the largest contributors to the analysis are three surface observations and one GPSRO measurement. Both observation types contribute roughly equally to the overall forecast error. It has been shown that the surface contribution only arises when all three surface observations are assimilated. A study of the characteristics of all the observations leading to the forecast error has shown that on balance their quality is relatively good and are of a typical standard for assimilated observations.

9. Discussion

a. Assimilation of good data and degradation of a forecast

The investigation discussed in this paper shows that a major component of the $T + 72$ h forecast error is the combined result of assimilating four observations that appear to be of relatively good quality. The surface contribution to this forecast error is particularly interesting because it requires all three observations to be assimilated in order for it to arise: any combination of two of the three observations has no impact on the forecast.

The assimilation of “good” observations and a resulting degradation of the forecast does not necessarily indicate an unknown problem with the data assimilation system. A model analysis is a global best fit of the observations and the background field, and it is known that
errors exist in both the observations and the model background but that it is not possible to know these errors accurately. The strategy of data assimilation is therefore to minimize on average the difference between the analysis and the true state of the atmosphere by relying on statistics of the observation and background errors. These statistics are not known accurately and are in some cases assumed.

The possibility of analysis and forecast degradation through the assimilation of observational data was illustrated by Morss and Emanuel (2002) in a simulated 3DVAR system. It was shown that, although adding observations generally improved the analysis and forecast, even with perfect observations and a perfect forecast model, there was a nonnegligible risk that adding observations could degrade the analysis and a significant risk that adding observations could degrade a forecast on a time scale of one to several days. This result illustrates that an erroneous analyzed model state is possible even with a perfect data assimilation system.

Johnson et al. (2006) examined the spatial structure of analysis increments and showed that 4DVAR preferentially generates analysis increments, which lead to growth because the evolved covariances at the end of the assimilation window contains more growing components than decaying components (Lorenc 1986). 4DVAR is therefore efficient at correcting rapidly growing errors.

The growth rate appropriate to a particular error mode is critically dependent on the specification of the background error covariances because it dictates the development of the analysis increments from the observational data. Johnson et al. showed that if the covariances are inaccurate then it is possible for the analysis increments to project incorrectly onto a growing error mode instead of a decaying error mode, leading to error growth instead of decay. In this case, 4DVAR is adding growing errors to the forecast rather than correcting them and the result is a worse forecast than if the observations had not been assimilated, even if the observations contain accurate and useful information. This random effect in the analysis process is therefore one potential source of error in determining an accurate analyzed model state.

Langland and Baker (2004) used an adjoint-based procedure for assessing the impact of observations on short-range (24 h) forecasts. They showed that the combined impact of assimilating all global observations reduced their forecast error measure (the difference between the 24-h forecast and 30-h forecast) but that a substantial fraction of observations were found to increase the error measure. This again highlights the statistical nature of the data assimilation process and the potential for negative impacts of good quality observations.

This method of observation impact assessment can also be considered as complementary to the sensitivity grid method described here, although the accuracy of the adjoint-based procedure will decrease with forecasts longer than 24 h because of the use of the adjoint.

The evidence from this study suggests the forecast error of 1 November 2009 is a result of an error in the analyzed model state that is due to an inherent random effect of the analysis process. This process occurred twice in one analysis with two independent observation types lying along the unstable baroclinic zone in the southwestern North Atlantic. There is no evidence of significant observation error or of significant errors in the model forecast formulation.

b. Impact on subsequent cycles

Given that the error had been identified with certainty as entering the forecasting system in the 1200 UTC 29 October Global Model run and that the origins of the majority of the error had been identified and could be removed from the assimilation, we considered the impact of the analysis degradation in one run on subsequent runs in the forecasting cycle. To study this, the $T + 6$ h forecast from 1200 UTC 29 October with the surface and GPSRO observations removed in the critical development area was used to generate the background state of the following cycle. The normal Global Model run sequence, with new assimilation every 6 h, was allowed to proceed through to the verification time.

The results showed (Fig. 9) that subsequent runs benefited from the improved analysis at 1200 UTC 29 October with the largest impacts being observed in the run immediately following this data time; here, a migration northward of 50 km of the low center and an increase in depth of 2 hPa was observed. The impacts in the following runs gradually decreased as the solutions converged toward their original all-data counterparts. Nevertheless, an improvement of some measure remained to even short-period forecast ranges (e.g., 2-hPa-deeper depression in the 1200 UTC 31 October $T + 36$ h forecast).

The persistence of the analysis on subsequent forecasts is an unfortunate result of the necessity to use a previous forecast in each forecast cycle and is unavoidable. The repeated use of a short-period forecast as the model background field results in the errors within the model background becoming organized and, as a consequence, rapidly growing. These project onto baroclinically unstable states where the errors and unstable states coincide. At each analysis cycle, the rapidly growing model errors are reduced through the assimilation of new data, but these grow rapidly again in the following cycle and have a potentially large impact on the model forecast skill. The introduction of an analysis
error into the forecasting system as shown here can thus last for many cycles of the model. In particular, this case illustrates that impacts of analysis changes in sensitive areas (e.g., the baroclinically unstable as in this case), especially in those which are relatively sparsely observed, can persist for many days.

c. Implications

The experimental evidence discussed in this paper strongly suggests that forecast skill has decreased as a result of the (correct) assimilation of good quality data. There is no evidence for significant errors in the observations or for significant errors in the forecast model formulation. The result is that data assimilation, although having minimized the error in the analysis, can on occasion lead to some unexpected consequences in the forecast.

Situations in which observations are selected in some way and are designed to achieve a particular response by the model may therefore be subject to unanticipated impacts. The generation of bogus observations falls specifically into this category. The use of targeted observations in which extra observations are deployed in areas of high calculated sensitivity and are designed to improve forecast skill could also lead to forecast degradation. Observation impact studies should also take the effect into account so that sufficient statistics are achieved in order to draw reliable conclusions.

Future developments in data assimilation will help alleviate this effect, particularly through the improved specification of the background error variances. Improved variances with smaller length scales and methods designed to introduce flow dependency should both allow a better representation of vertical structures and the development of fewer inappropriate structures in the model. One method of achieving this is through the use of a hybrid ensemble–VAR system that uses ensemble forecasts to provide a flow-dependent component for the static covariances of the VAR system (Hamill and Snyder 2000; Lorenc 2003). Initiatives such as this are intended to
move the assimilation process toward a more optimal treatment of observational data.

d. Summary

In conclusion, this investigation has illustrated a number of points: 1) the primary sensitive area of a system can be located within the system itself and can be relatively insensitive to larger-scale errors propagating from upstream, 2) initial low-level errors can be important in the skill of forecasts, 3) observations of probable good quality can degrade the analysis relative to the background field due to an inherent random effect of the data assimilation process which results in a significant degradation to the forecast, and 4) the impact of the error in a single forecast cycle can persist for a number of subsequent cycles.

Acknowledgments. The authors wish to thank Andrew Lorenc, Rick Rawlins, and Anders Persson for comments on earlier versions of the manuscript, and to thank the reviewers for their useful comments on the submitted manuscript. The authors also thank Mike Rennie for discussions and his help interpreting the GPSRO observations.

REFERENCES


