

The impact of targeted observations made during the Greenland Flow Distortion Experiment

E. A. Irvine,^{a*} S. L. Gray,^a J. Methven,^a I. A. Renfrew,^b K. Bovis^{c†} and R. Swinbank^{c†}

^aUniversity of Reading, UK

^bUniversity of East Anglia, Norwich, UK

^cMet Office, Exeter, UK

ABSTRACT: The impact of targeted sonde observations on the 1–3 day forecasts for northern Europe is evaluated using the Met Office four-dimensional variational data assimilation scheme and a 24 km gridlength limited-area version of the Unified Model (MetUM). The targeted observations were carried out during February and March 2007 as part of the Greenland Flow Distortion Experiment, using a research aircraft based in Iceland. Sensitive area predictions using either total energy singular vectors or an ensemble transform Kalman filter were used to predict where additional observations should be made to reduce errors in the initial conditions of forecasts for northern Europe. Targeted sonde data was assimilated operationally into the MetUM. Hindcasts show that the impact of the sondes was mixed. Only two out of the five cases showed clear forecast improvement; the maximum forecast improvement seen over the verifying region was approximately 5% of the forecast error 24 hours into the forecast. These two cases are presented in more detail: in the first the improvement propagates into the verification region with a developing polar low; and in the second the improvement is associated with an upper-level trough. The impact of cycling targeted data in the background of the forecast (including the memory of previous targeted observations) is investigated. This is shown to cause a greater forecast impact, but does not necessarily lead to a greater forecast improvement. Finally, the robustness of the results is assessed using a small ensemble of forecasts. Copyright © 2009 Royal Meteorological Society and Crown Copyright.

KEY WORDS predictability; sensitive area prediction; singular vectors; ensemble transform Kalman filter

Received 28 November 2008; Revised 30 January 2009; Accepted 11 February 2009

1. Introduction

The aim of targeted observing is to improve the short-term forecast for a specified region and time by the assimilation of additional observational data into operational weather forecasts. For a limited-area model, there are three potential sources of forecast error: model deficiencies, errors propagating in from lateral boundary conditions and errors in the initial conditions. The aim of targeted observing is to reduce the initial condition error in a sensitive region, where errors have the potential to grow (or are large at the targeting time) and therefore degrade the forecast downstream of the target region a short time later. Targeted observing should therefore reduce forecast ‘busts’ which occur as a result of the growth of initial condition error. Field programmes which have trialled targeted observing have taken these observations in oceanic regions, away from the radiosonde network, with the rationale that initial condition errors are most likely

to occur in regions where there are fewer observations available to constrain the analysis.

Targeted observations have the potential to be used as part of an adaptive observing system, where additional observations are targeted according to the synoptic situation. This was trialled during 2008 under the Eurorisk-PREVIEW (PREvention, Information and Early Warning) programme (<http://www.preview-risk.com>). Investigating the benefit of targeted observations is also one of the aims of The Observing System Research and Predictability Experiment (THORPEX), a World Meteorological Organisation programme (<http://www.wmo.int/thorpex>). This paper aims to contribute to the debate on the effectiveness of targeted observing, by analysing the impact of targeted observations made during the field campaign of the Greenland Flow Distortion experiment (GFDex) in February–March 2007 (Renfrew *et al.*, 2008).

There have been a number of trials of targeted observations during field campaigns; Langland (2005) provides an overview of these campaigns. These have had mixed success in their attempts to improve forecasts. The Fronts and Atlantic Storm Track Experiment (FASTEX) in 1997 (Bergot, 1999; Szunyogh *et al.*, 1999) was the first field campaign to objectively target sensitive areas to improve the forecasts of landfalling systems downstream of the

*Correspondence to: E. A. Irvine, Department of Meteorology, University of Reading, PO Box 243, Reading, Berkshire RG6 6BB, UK. E-mail: e.a.irvine@reading.ac.uk

†The contributions of K. Bovis and R. Swinbank were written in the course of their employment at the Met Office, UK, and are published with the permission of the Controller of HMSO and the Queen’s Printer for Scotland.

target area (the North Atlantic). An average reduction in the two-day forecast error of 15% was seen during the period of the field campaign (Montani *et al.*, 1999). This large forecast improvement seen during FASTEX encouraged further trials of targeted observations, in the North Pacific Experiment (NORPEX) in 1998 (Langland *et al.*, 1999; Liu and Zou, 2001) and the Winter Storm Reconnaissance project (WSR) (Szunyogh *et al.*, 2000, 2002) which has run every year since 1999. In both of these field campaigns, a reduction in the two-day forecast error of 10–20% was achieved as a result of assimilating targeted sonde data. These early studies used global forecast models with gridlengths of 100 km and greater – now considered to be relatively coarse resolution (the possible exception to this is the study of Montani *et al.* (1999), who used a spectral model at T213 truncation, which is approximately 60 km resolution), and used data assimilation schemes that did not take into account the exact time at which the observations were made (i.e. were not four-dimensional schemes). More recently, the Atlantic THORPEX Regional Campaign (ATReC) showed a mixture of small improvement and degradation due to targeted observing, using four-dimensional schemes to assimilate the targeted data (Petersen and Thorpe, 2007; Fourrié *et al.*, 2006). The reduced impact of the targeted data during ATReC, compared to earlier targeting studies, could be due to improved data assimilation and model formulation, along with an increased number of routine observations, which together have limited the scope for targeting. This does however highlight the need for a study of targeted observations using an up-to-date model of higher resolution than has been used previously for this. This study addresses this issue by using a 24 km resolution regional model to run the hindcast studies.

Field campaign trials of targeted observations are limited by their small number of cases, which make it difficult to draw general conclusions about the success of the targeting and assess whether the results are statistically significant. Buizza *et al.* (2007) (also Cardinali *et al.*, 2007; Kelly *et al.*, 2007) addressed these issues by running large numbers of hindcast experiments for cases where the extra observations were placed over the Atlantic or Pacific Oceans and were either targeted in singular vector (SV) sensitive regions, or random regions. They found that the impact of targeted observations in SV sensitive regions was larger in winter than summer, and larger for the Pacific than the Atlantic region. The routine observing network in the target region was also found to greatly influence the forecast impact from targeted observations, with an average forecast error reduction of 19.1% if the observations were targeted in data-sparse regions reducing to only 2% if the observing network in the targeting region could not be considered sparse. They also concluded that targeted observations in sensitive regions were more valuable than in random regions. As our study only comprises five targeting cases, general conclusions about the success of targeting cannot be drawn. However it does attempt to quantify the significance of the

improvement that is seen in two cases by using ensemble methods. This method may be useful to try and draw useful conclusions about the success or failure of targeted observations from future field campaigns.

Targeting studies to date have largely focused on the use of dropsonde observations as targeted observations. There are many advantages to using direct measurements of the atmosphere such as from dropsondes as opposed to satellite radiances, although the disadvantage is primarily the expense. Dropsonde data is of higher resolution in the vertical than satellite data and has smaller errors. The current observing network lacks high-resolution profile data in data-sparse regions such as the North Atlantic. Radiosonde stations provide high-resolution profile data but, with the exception of a few ship-released radiosondes, these data are generally only available over populated areas of land. Satellite observations might therefore be considered ideal to fill this gap, however satellites measuring infrared radiation cannot see through cloud. As targeted observations are typically taken in baroclinically active regions, which are often cloudy (McNally, 2002), the satellites may not be able to sample shallow sensitive regions indicated by SVs to be located within the cloud layer (Lawrence *et al.*, 2009). Microwave sensors are less sensitive to the presence of cloud, however Dando (2007) found only small forecast improvements by assimilating targeted Advanced TIROS Operational Vertical Sounder (ATOVS) observations. This study therefore uses only *in situ* dropsonde data as targeted data.

The GFDex was a three-week field campaign which used the Facility for Atmospheric Airborne Measurement (FAAM) BAe 146 aircraft based at Keflavik, Iceland from 19 February to 12 March 2007. The GFDex aimed to study high-wind-speed events such as tip-jets off Cape Farewell and barrier winds along the southeastern coast of Greenland, caused by the interaction of the synoptic-scale flow with Greenland, as well as the effect this has on downstream predictability. (A complete description of the field campaign and the different observational aims is given in Renfrew *et al.*, 2008.) The GFDex carried out targeted observing flights in the region around Iceland and southern Greenland using the FAAM aircraft to release dropsondes into sensitive regions. Here we present an analysis of the impact of these targeted observations. The set-up of the targeting experiments and the details of the hindcasts performed to analyse the results are given in sections 2 and 3. An overview of the impact seen in each of the five targeting cases is given in section 4. Two cases where the targeting was successful are analysed in more detail in sections 5 and 6. As sondes were released for several targeting flights over only two and a half weeks, this offers us the chance to study the impact of memory of previous targeted observations on the forecast, examined in section 7. Ensemble methods are used to quantify the significance of the forecast improvement in section 8, followed by discussion and conclusion in section 9.

2. Set-up of the GFDex targeting experiments

This section describes the set-up of the targeting experiments during GFDex: the sensitive area predictions, the timelines involved in the targeting and finally an overview of the targeting flights made.

2.1. Sensitive area predictions

An important part of the observation targeting process is defining the *target region*, where the additional observations will be made. Sensitive area prediction (SAP) plots are used to determine this. They predict where the forecast for a specified *verification region* will be most sensitive to errors in the initial conditions for a specific *optimisation time*, i.e. the time interval between performing the targeting and verifying the forecast (Figure 1). Here the word sensitive is used to mean that a large forecast error may result in the verification region at the verification time as a result of an initial error elsewhere in the forecast domain. To identify this sensitive region, two objective methods were used during the GFDex: total energy singular vectors (TESV; Buizza and Montani, 1999) and the ensemble transform Kalman filter (ETKF; Bishop *et al.*, 2001).

The sensitivity calculations were performed for two fixed verification regions. These were approximately square regions of side 2000 km. One region was positioned over northwest Europe centred over the UK and the other over Scandinavia (Figure 2). The verification regions were placed downstream of Greenland, in areas to which flow perturbations from the target region might be expected to move within the specified optimisation times. The regions were similar in size to those used by previous targeting experiments such as FASTEX (Montani *et al.*, 1999) and ATReC (Petersen and Thorpe, 2007).

The TESV-SAPs were generated by ECMWF, at T42 resolution, using SVs initialised from their deterministic forecast model. Total energy SVs are used to identify the fastest-growing perturbations over the optimisation time that would impact the verification region. The growth of the SVs was measured using a total energy norm. This is an appropriate metric to use for targeting purposes as it provides a good approximation to the analysis-error covariance matrix used by data assimilation schemes (Palmer *et al.*, 1998). The SAP plots are generated by vertically averaging the ten leading SVs, which is sufficient to gain a good representation of the target region (Buizza and Montani, 1999).

The ETKF-SAPs were provided by the Met Office, at a resolution of 2.5° , using output from the 23-member

15-day global component of the Met Office Global and Regional Ensemble Prediction System (MOGREPS; Bowler *et al.*, 2008). The ETKF takes the MOGREPS perturbations (ensemble member minus the ensemble mean) and transforms them to obtain an estimate of the analysis-error covariance within the verification region from assimilating routine observations only, or routine observations plus targeted observations. The ETKF assumes a routine observational network of radiosondes and some satellite temperature measurements, and uses a set of nine test soundings to represent targeted observations. The reduction in forecast-error variance resulting from assimilating the test set of targeted observations is calculated for different locations of the test soundings, using total energy as a metric. The SAP plot displays the optimal location for targeting observations in order to get the maximum forecast-error reduction within the verification region. Plots of non-normalised signal variance, roughly equivalent to the reduction in forecast-error variance from assimilating real targeted observations (Majumdar *et al.*, 2001), were also provided by the Met Office during the campaign. This allowed the magnitude of forecast sensitivity to be compared from one day to the next; this information was not provided by the TESV sensitivity calculations.

There are differences in the types of region identified by ETKF- and TESV-SAPs. The reader is referred to Majumdar *et al.* (2002) for a detailed discussion of the reasons for this. SVs identify regions where a small initial error will grow rapidly over the forecast time, thus SVs measure the rate of error growth. By contrast, the ETKF identifies regions that result in a large error at verification time, thus the ETKF identifies both regions where an initial error has grown rapidly to be large at verification time and regions where the initial error was large but did not grow. SVs computed with the total energy norm often identify sensitive regions over data-rich regions, as the position of the routine observing system is not taken into account during their calculation. As the ETKF takes into account the position of the radiosonde network, it tends to identify only regions away from the routine observing network, however it may also identify spurious sensitive regions due to a lack of covariance localisation (Majumdar *et al.*, 2002).

Little overlap between the ETKF- and TESV-SAPs was observed during the campaign. This was also the case during the ATReC campaign, where the ETKF and SV sensitive areas overlapped by more than half in only 46% of cases (Leutbecher *et al.*, 2004). There was no overlap in the regions of maximum sensitivity for the targeted

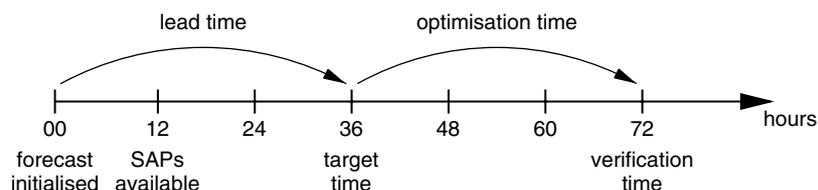


Figure 1. An illustration of the timeline involved in making targeted observations. In this example the lead time is 36 h and the optimisation time for the sensitive area predictions is also 36 h.

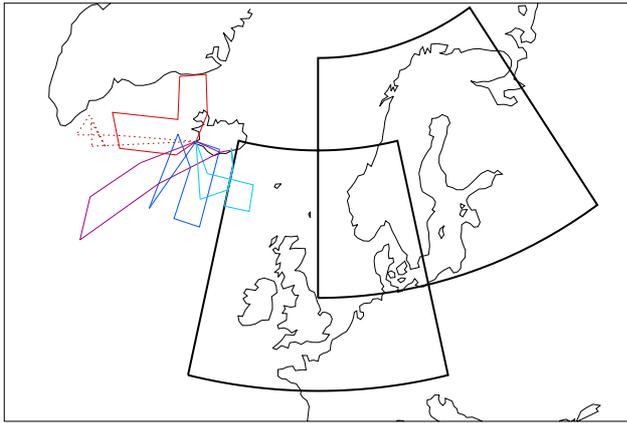


Figure 2. Map of the study region, showing the flight tracks of the targeting (solid lines) and null (dotted line) flights during the GFDEX. The flights started from Iceland. The two verification regions, one centred over the UK (referred to in the text as northwest Europe) and one over Scandinavia, are marked by bold lines. This figure is available in colour online at www.interscience.wiley.com/journal/qj

cases. This meant that flights were tasked using either TESV- or ETKF-SAPs, but not both. It was noted that the TESV sensitive regions tended to be located further north and on the eastern side of low pressure systems, whereas the ETKF-SAPs were located further south and towards the centre of lows. It is beyond the scope of this paper to analyse the reasons for these differences in locations; further analysis would be required to determine whether the differences are systematic and their cause.

2.2. Timelines involved in targeted observations

Targeted observations were made over a period of approximately 4 h, centred on 1200 UTC (hereafter referred to as the *target* time; Figure 1). A constraint on operational targeted observing is the need for a time-lag between running the sensitivity calculations and making the targeted observations. This so-called *lead time* allows the resources to be mobilised and flight plans to be drawn up. During the GFDEX, air traffic control restrictions meant that flight plans had to be filed 24 h before take-off. Thus SAPs with a 36 h lead time were used to plan the targeted observation flights. SAPs were also received with lead times of 48 and 60 h, which were used to make preliminary flight plans. SAPs with a 24 h lead time were used as a final check on the location of the sensitive region. It was observed during the campaign that SAPs with a 24 h lead time were consistent with the SAPs with a 36 h lead time. This agrees with the findings of Petersen *et al.* (2007) that there is good spatial coherence between SAP plots generated 12 h apart. SAPs were generated for three optimisation times: 24, 36 and 48 h. A range of optimisation times was necessary to allow the signal from the dropsondes to move from the target region into the verification region. This distance varied from case to case, as the target area ranged from the Denmark Strait to the southeast of Iceland. The speed with which the impact from targeted observations propagates downstream was shown in Szunyogh *et al.* (2000) to be 30° longitude

per day, suggesting that downstream development may be important, as the speed is consistent with the group velocity of a dispersive Rossby wave in the atmosphere (Persson, 1999, and references therein).

The SAPs were generated from forecasts run at 0000 UTC and 1200 UTC daily. The SAP plots were available for us to analyse 7–8 h (ECMWF SAPs) and 10–11 h (Met Office SAPs) after this. This means that the total time from forecast initialisation to verifying the forecast containing the targeted observations varied from 48 to 108 h. This tests the limits of the assumptions of linearity made by both ETKF and SV methods, as a growing perturbation is likely to be in a nonlinear regime after 48 h (Gilmour *et al.*, 2001). However during the field campaign it was observed that there was little variation in the location of the sensitive regions identified by either the ETKF or SV methods as the total time decreased.

2.3. Overview of GFDEX targeting flights

Targeted observing was carried out on four out of twelve flights during the field campaign; three of these flights were into TESV-predicted sensitive regions and one into an ETKF-predicted sensitive region. Targeting flights were made nearly every time a sensitive region was predicted to be in range of the aircraft. This is a different strategy to that which would be used for operational targeting as, due to the cost of targeting, it could only be justified to reduce the uncertainty in forecasts which indicate the possibility of a high-impact weather event. After the campaign, a null case was chosen from the other flights. This was a flight into a Greenland lee cyclone, into an area which neither the ETKF- nor TESV-SAPs predicted to be sensitive to initial condition errors. The assimilation of additional data here should not significantly impact the forecast within the verification regions. Details of the flights are given in Table I, and the flight tracks are shown in Figure 2.

Global Positioning System (GPS) dropsondes (Vaisala RS92) released from the aircraft were used to make the observations. These measure vertical profiles of height, pressure, air temperature, humidity and wind, transmitting the data back to the aircraft. The data were sent onto the global telecommunications system (GTS) via a satellite communications device during the flight. The data were therefore available for meteorological centres to use in the operational 1200 UTC forecasts. Data from all GFDEX sondes (targeted and non-targeted) were assimilated into operational forecasts.

Each targeted observing flight during the GFDEX released 7–11 targeted sondes (out of 11–18 total sondes per flight; Table I), covering an area of approximately $1 \times 10^5 \text{ km}^2$. For the null case, five of the 16 sondes that were launched during the flight are treated as targeted sondes. Note that some flights released additional sondes, some inside and some outside the predicted sensitive regions, for additional observational aims. The treatment of these non-targeted sondes is discussed in section 3.2. The spatial coverage of the target area and number of dropsondes released within it have been shown to affect

Table I. Specification of targeting and null case studies.

Date	Target region	SAP type	Sensitivity for NW Europe for optimisation times			Sensitivity for Scandinavia for optimisation times			Targeted/ total sondes deployed
			24 h	36 h	48 h	24 h	36 h	48 h	
24 Feb 07	S of Iceland	TESV	Y	Y	Y	N	N	Y	11/11
26 Feb 07	SE of Iceland	TESV	Y	Y	Y	Y	Y	Y	7/14
01 Mar 07	Denmark Strait	TESV	N	Y	Y	Y	Y	Y	9/18
03 Mar 07	SW of Iceland	NULL	N	N	N	N	N	N	5/16
10 Mar 07	SW of Iceland	ETKF	Y	Y	Y	Y	Y	Y	9/15

the success of the targeting. Idealised modelling studies suggest that releasing between 10 and 40 dropsondes within an area of $3 \times 10^6 \text{ km}^2$ should be sufficient to improve the 2 d forecast over Europe (Leutbecher *et al.*, 2002). Buizza and Montani (1999) showed that insufficient sampling of a target area can result in only 65% of the possible theoretical forecast improvement being realised. However, sampling the large area indicated necessary by the modelling studies with one aircraft like the BAE 146 used in the GFDex is logistically impossible. This is due to the limited range of the plane and the need to release all the sondes within a 4.5 h time window so that they do not miss the cut-off time for the operational Met Office global forecast model (1345 UTC for the 1200 UTC forecast).

The spacing of the dropsondes can also be important to the success of the targeting. Numerical modelling studies indicate that the spacing of the dropsondes should be 1–2 times the horizontal correlation length-scale of the variables, such as wind and temperature, being assimilated (Leutbecher *et al.*, 2002). As the analysis increments are spread out according to the error correlation specification in the data assimilation system, it is evident that dropsondes placed closer than a correlation length will lead to increments which overlap. The horizontal correlation length-scale for temperature for example varies with height, from around 280 km at the height the dropsondes were released to approximately 200 km near the ground (Ingleby, 2001). Taking these results into account, and allowing for the range of the plane, the spacing between targeted dropsondes in the GFDex was always greater than 180 km and for most flights was around 220 km.

3. Design of hindcast studies

This section describes the model used to run the hindcasts, and the different hindcast runs performed with the model.

3.1. Description of model

The Met Office Unified Model (MetUM) version 6.1, operational at the time of the field campaign, was used to perform the hindcast studies. The model dynamics are

non-hydrostatic. It was run as a limited-area model, over the operational North Atlantic European (NAE) domain, with 24 km grid spacing and 38 vertical levels. The model was run at half the horizontal resolution of the operational NAE forecasts, due to the large computational cost of running the hindcasts. The justification for running a limited-area model, rather than a global model, was that both the target and verification regions were inside the limited-area domain, although near the northern edge, and this was sufficient to study the impact of sonde data on the forecasts for the verification regions. Using a limited-area model also allowed the hindcasts to be run at higher resolution. Lateral boundary conditions were provided by the operational global control run of MOGREPS. The MetUM uses a four-dimensional variational (4D-Var) data assimilation scheme with a 6 h window for the assimilation of observations (Rawlins *et al.*, 2007). The hindcast studies used the same assimilation window as the operational NAE forecast cycle, which for the 1200 UTC forecast was 0900 UTC to 1505 UTC, but a 48 km gridlength for the 4D-Var scheme (half the operational resolution).

Before observations are passed to the data assimilation scheme, they are first subject to a quality control process, based on Bayesian methods, to remove ‘bad’ observations. For dropsonde observations, which are treated like radiosonde observations, this means checking the profile for consistency and performing ‘buddy checks’ (Ingleby and Lorenc, 1998). The buddy checks compare the difference between two like observations (e.g. radiosonde–dropsonde), located within 150 km of each other, against the background and determine whether it is the background or the observation that is in error. With the exception of some wind data from one malfunctioning sonde, the GFDex observations were accepted by the quality control system and passed to the data assimilation scheme; only 0.2% of GFDex observations were rejected.

3.2. Description of the hindcast runs

To assess the impact of the targeted dropsondes on the forecast, the complete forecast cycle of four forecasts per day was re-run for the period 19 February to 13 March 2007, with different combinations of observations excluded from the forecast. Four separate forecast cycles

Table II. Definition of the hindcast runs.

Run	Routine observations	Non-targeted GFDEX sondes	Targeted GFDEX sondes	Continuous assimilation
CONTROL	Y	N	N	Y
TNOMEM	Y	N	Y	N
TMEM	Y	N	Y	Y
ALL DATA	Y	Y	Y	Y

were run: labelled CONTROL, TNOMEM, TMEM and ALL DATA (Table II).

The CONTROL run assimilated only routine observations. Here ‘routine observations’ is taken to mean the non-adaptive component of the observing network, such as radiosonde, aircraft meteorological data relay (AMDAR), satellite and surface station observations. The CONTROL forecast run therefore produces a set of forecasts which have not been influenced by the additional observations made during the GFDEX. Note that during the GFDEX there were some additional radiosonde launches from land stations in Greenland and Iceland at 0600 and 1800 UTC and from three ships when they were within a pre-defined area around Iceland and southern Greenland (Renfrew *et al.*, 2008). These have been classified as routine observations for this study, as the additional land launches were outside the observation window for the data assimilation scheme for the 1200 UTC forecast and the ship releases were not within sensitive areas. Two other runs assimilated targeted observations along with routine observations. The TNOMEM hindcasts took the background from the CONTROL forecast and assimilated targeted observations for the 1200 UTC forecast on that *one particular targeting day*. This was done separately for each case-study, so that the background for each TNOMEM forecast contains only routine observations. The set-up of the hindcasts means that any differences between the TNOMEM forecast and the CONTROL forecast are due to the targeted observations made that day. In contrast with the TNOMEM run, the TMEM run assimilated targeted observations *every time they were available*, so that the targeted data was cycled in the forecast. This allows the effect of memory of previous observations in the forecast to be investigated. Therefore any differences between the TMEM and TNOMEM forecasts are due to differences in the background of the two forecasts. As the model was run with a 24 km resolution instead of the 12 km used operationally, a run with the same observations as the operational run was also produced. This included routine observations and all the GFDEX sondes, including non-targeted sonde data. This is the ALL DATA run, and was used to produce the analyses needed to verify the forecasts. The difference between the TMEM and ALL DATA runs is that the TMEM run does not include non-targeted sonde data.

The data assimilation was run at the Met Office, using the operational MetUM version 6.1. The forecasts were then run from the output produced by the data assimilation runs on the UK’s high-performance computing

facility (HPCx), available via the National Centre for Atmospheric Science (NCAS), using the release version of the MetUM version 6.1. Each forecast was run for 63 h, starting at 0900 UTC on the targeting day and outputting data every 6 h.

4. Impact of the targeted observations

The targeted sonde data are taken in regions where there are few high-density profile measurements to constrain the background field. This may result in larger initial-condition errors in this region, which may then propagate into the verification region. In order to correct these errors, if such errors exist, the sonde data must be different from the background field and its inclusion in the forecast should lead to an adjustment of the background field towards the sonde observations. Figure 3(a)–(e) shows the *difference* between the v -component (north–south) of the wind measured by the sonde and the CONTROL model v (at the location where the sonde was released), defined as

$$difference = | CONTROL - SONDE | . \quad (1)$$

The CONTROL model data have been interpolated to the position at which the dropsonde was released, and both model and dropsonde data have been interpolated onto common pressure levels for ease of comparison. Here, hourly model output was used, taking the model output closest to the time of the dropsonde release, so that the background field should be representative of the true atmospheric state at the time the sonde data were collected. It is seen that the *difference* is larger for the last three cases, and in particular the last case, where the v -wind *difference* is as high as 15 m s^{-1} (Figure 3(e)), and the upper-level temperature difference is as high as 6°C (not shown). It is likely that a larger *difference* will have a larger impact on the forecast.

The model *adjustment* towards the sonde data, shown in Figure 3(f)–(j), is defined as

$$adjustment = | CONTROL - SONDE | - | TNOMEM - SONDE | . \quad (2)$$

Positive values of *adjustment* mean that the model profile from the forecast containing the targeted observations (TNOMEM) is closer to the sonde observations than the profile from the CONTROL forecast which does

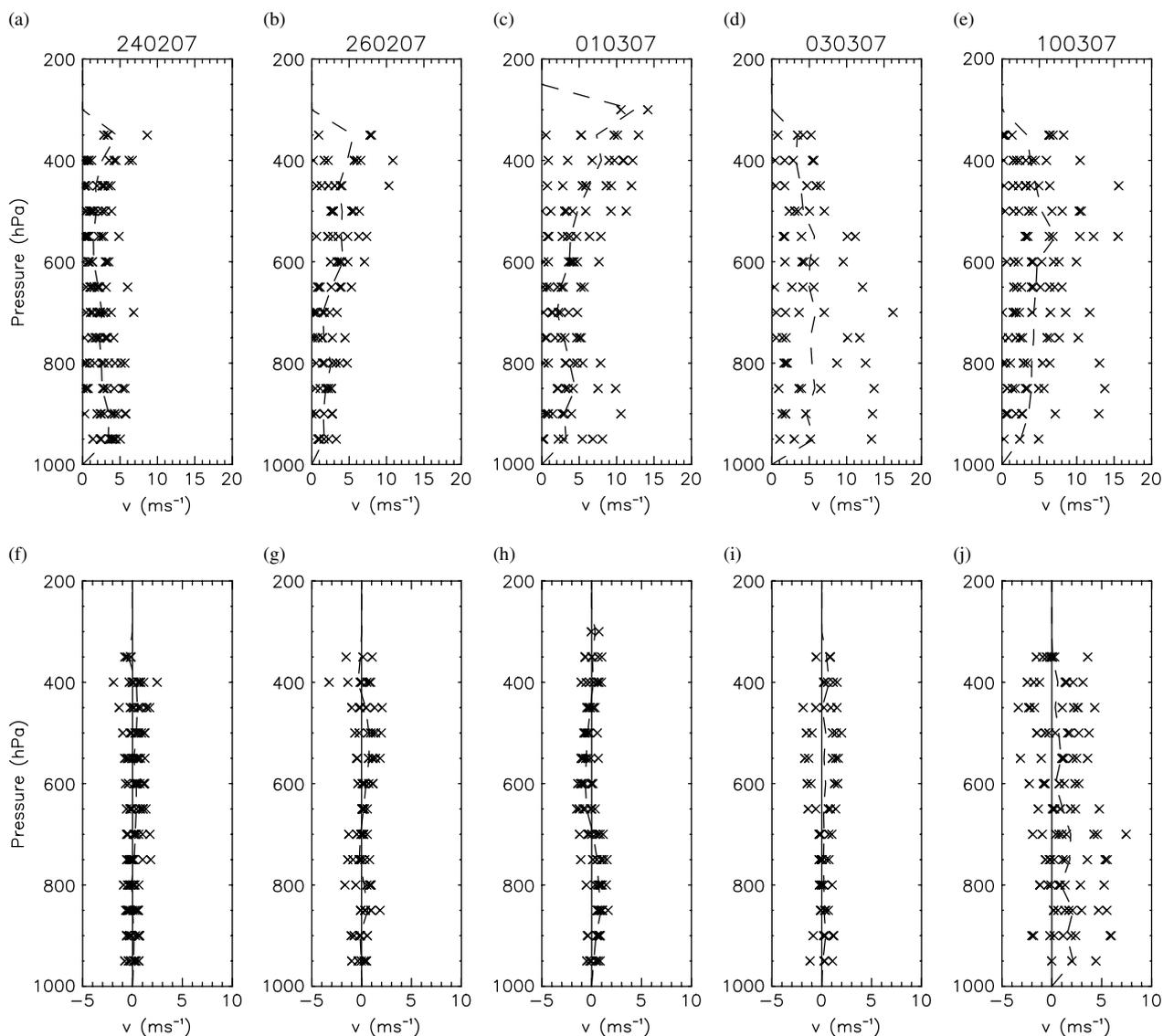


Figure 3. Sonde–model comparison for v -component of the wind for each of the five cases; dates (ddmmy) are given above the top panels. (a) to (e) show the *difference* (Eq. (1)) between the CONTROL forecast and sondes for each case. (f) to (j) show the *adjustment* (Eq. (2)); this is positive if the forecast containing the dropsonde information is closer to the dropsonde data than the CONTROL forecast and negative if it is further away. Each cross represents data from one sonde at that pressure level, and the dashed line represents the average value. Note that the altitude of the aircraft varied both within and between flights therefore sondes were released from different altitudes.

not include targeted observations. It implies that the data assimilation scheme has made use of the data to modify the background field, and negative or zero values of *adjustment* imply that it has not. It is seen that for most cases the *adjustment* is near zero, although with a slight positive bias, which for the 01 and 03 March cases where the *difference* is large may indicate that the data assimilation system has been unable to use the information contained in the targeted observations. For the 24 February and 26 February cases, the average *difference* is below 5 m s^{-1} and the *adjustment* is correspondingly small. This suggests that the targeted sondes will have little impact on the forecast in these cases. The 10 March case shows a relatively large *difference* and large *adjustment* with an obvious positive bias.

To quantify whether the dropsonde observations had a positive or negative impact on the forecast for the

verification regions, the forecasts must be verified against some measure of the true state of the atmosphere at verification time. Model analyses from the ALL DATA forecast run were used as the best estimate of the ‘truth’. Following most previous targeting studies, model analyses were used instead of observations due to the poor and irregular spatial coverage of profile observations, such as radiosonde data, available to verify the forecasts. The forecasts for one case were also verified against operational ECMWF analyses, interpolated to the same grid and resolution, as a comparison. It was found that the improvement was similar in magnitude and location to the improvement seen when MetUM analyses were used to verify the forecasts.

The results are presented here in terms of total energy. Total energy was chosen to measure the forecast improvement, because it is used as a metric in calculating the

TESV- and ETKF-SAPs. The total energy, TE , of the deviation of a forecast from the analysis can be calculated as

$$TE_{F-A} = \frac{1}{2} (u_{F-A}^2 + v_{F-A}^2) + \frac{1}{2} \frac{c_p}{T_{ref}} (T_{F-A}^2), \quad (3)$$

where u is the east–west oriented wind component, T is the temperature, and the subscript F–A indicates a forecast minus analysis difference. Here the reference temperature $T_{ref}=300$ K and c_p is the specific heat capacity. The quantity TE_{F-A} was calculated at 850, 500 and 250 hPa levels, and the results summed. These levels were chosen to be consistent with the levels used in calculating the ETKF-SAPs. An impact metric I is then defined, in terms of total energy, such that $I > 0$ implies that the targeted observations have improved the forecast, but $I < 0$ implies that they have degraded it:

$$I = TE_{C-A} - TE_{T-A}, \quad (4)$$

where C is the CONTROL forecast, A is the verifying analysis (from the ALL DATA hindcast run, section 3.2) and T is the forecast containing targeted observations, which for the results in this section is TNOMEM.

Figure 4 shows the impact as a result of assimilating the targeted data, using Eq. (4) averaged over the verification regions. For all cases, and in both verification regions, the average impact of targeted data on the forecasts is small (Figure 4(a) and (c)), which may be related to deficiencies in the ability of the data assimilation to make use of the targeted data, as previously discussed. There is no improvement to the forecast in the northwest Europe verification region (Figure 4(a)) at the forecast times that the SAPs indicated sensitivity for (24–48 h) for either the 24 and 26 February or 10 March cases. The 01 March and 03 March cases show some improvement at longer forecast times, but on inspection of the individual forecasts it is not possible to conclusively attribute this to the targeted observations. For the Scandinavian verification region, Figure 4(c), there is clear improvement ($I > 0$) in the 01 and 10 March cases at the 24 h forecast time. The forecast improvement for these cases grows from near zero at target time to a peak at around 24 h forecast time after which I decreases, so that the forecast is degraded after approximately 30 h forecast time. Upon inspection of the spatial variation in impact, it is only for these two cases that it is possible to track the impact from the target region to the verification region. The mechanisms for this forecast impact for these cases are examined in detail in sections 5 and 6. Small forecast impact is seen for the other two targeted cases and the null case for this verification region. Recall that, for 24 and 26 February, the *difference* was smaller than for the other cases, suggesting that there were only very small errors in the initial conditions in the target region (i.e. the forecast background well represented the state of the atmosphere). It is therefore unsurprising that there is only small forecast impact for these cases.

To put into perspective the magnitude of the impact of the targeted observations, the magnitude of the forecast error (defined as Eq. (5)) within the verification regions, which the targeted observations aim to reduce, is examined.

$$error = TE_{C-A}. \quad (5)$$

Figure 4(b) and (d) shows that the forecast error increases with time, to 30–40 m^2s^{-2} after 24 h optimisation time for the 01 March and 10 March cases, which is an order of magnitude larger than the modulus of the forecast impact (Figure 4(a) and (c)) at that time. The forecast error is non-zero at target time because the CONTROL forecast is verified against analyses from the ALL DATA run. Therefore there are differences in the number of observations both assimilated and in the background fields. The operational NAE forecasts generally verified well during this time and the forecast error, and therefore the potential to improve the forecast, was small. The maximum reduction in forecast error due to the targeted observations was 5% at 24 h forecast time for the 01 March case, and the forecast degradation reached around 3% in several cases.

5. Targeting case-study: 01 March 2007

In this section the impact of targeted observations for the 01 March case is examined in greater detail. An overview of the case is given, including the synoptic situation at target time. The initial increments to the background field are then examined from a dynamical perspective, to understand how the targeted observations modified the background field, and how the impact of the targeted observations propagated through the forecast.

5.1. Overview

At targeting time on 01 March, the surface pressure chart showed a small Greenland lee cyclone (labelled ‘G’ on Figure 5(a)) to the southwest of Iceland, and weak barrier flow through the Denmark Strait. The barrier flow was forecast to strengthen as a low pressure system, southeast of Cape Farewell at this time, moved north towards Iceland and deepened (Petersen *et al.*, 2009, provide further details). At upper levels there was westerly flow on the south side of a trough positioned over northeast Greenland and extending out into the Greenland Sea (not shown). This, with an associated upper-level potential vorticity (PV) anomaly, was providing upper-level forcing for polar low development (labelled ‘P’ on Figure 5(a)) in the Greenland Sea, within the cold-air outbreak that extended as far south as the UK.

The TESV-SAPs indicated the sensitive region for short-range forecasts over northwest Europe and Scandinavia to be west of Iceland, over the Denmark Strait and Irminger Sea (Figure 5(a)). This corresponds with

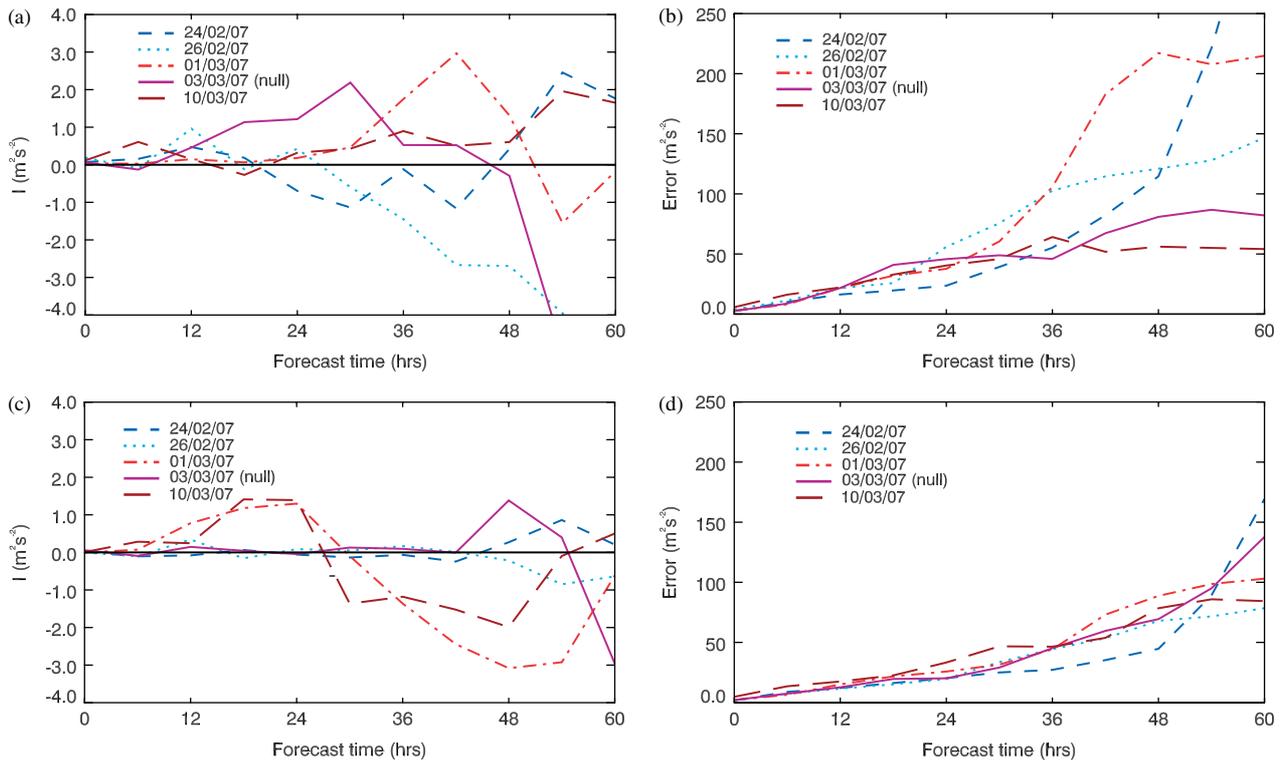


Figure 4. (a, c) show the Impact, I (Eq. (4)), of the targeted sondes for the TNOMEM forecast, and (b, d) the forecast error (of the CONTROL forecast) during the GFDex. (a, b) show the results for the northwest Europe verification region, and (c, d) for the Scandinavian verification region. Note the difference in scale used for the y-axis for I and $error$.

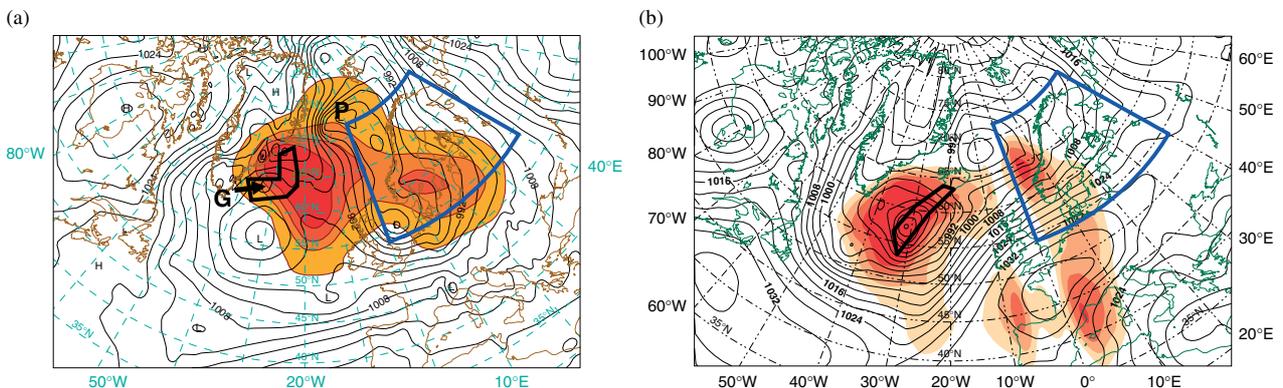


Figure 5. (a) The ECMWF TESV-SAP used to plan the 01 March flight and (b) the Met Office ETKF-SAP used to plan the 10 March flight. The SAPs show the locations of the sensitive regions, where the addition of sonde data could improve the forecast for the Scandinavian verification region (36 h lead time, 24 h optimisation time). The colour shading indicates the sensitive regions, with the most sensitive $1, 2, 4$ and 8×10^6 km² of the forecast domain shaded. The darkest shading indicates the region of maximum sensitivity. The black contours are the 36 h forecast of mean sea-level pressure valid at targeting time. The flight tracks and Scandinavian verification regions are outlined. In (a), P indicates the position of the developing polar low at targeting time, and G the location of the Greenland lee cyclone.

the location of the weak barrier winds and lee cyclone. By contrast, the ETKF-SAPs showed little forecast sensitivity to additional observations in this area, with signal variances of only $0.2 \text{ m}^2 \text{ s}^{-2}$ for both verification regions for SAPs with a 24 h lead time. This was below the average seen during the field campaign, and so the ETKF-SAPs were not used for flight planning for this case. The flight released nine dropsondes for targeting, equally spaced along the flight track. Of these nine sondes, one sonde failed completely and therefore data from only eight sondes were transmitted onto the GTS. An additional nine sondes were released during the flight

to study the barrier wind and lee cyclogenesis. These additional sonde data were assimilated into the ALL DATA run, along with the targeted sonde data.

5.2. Use of targeted sonde data by the assimilation scheme

The 4D-Var assimilation scheme used to assimilate the sonde data seeks to find an increment to the background state that best minimises a cost function over a

6 h period. Increments are then applied to model-grid-orientated wind components (u_m and v_m), potential temperature, Exner pressure, total water and ozone, at the start of the forecast, 0900 UTC. The analysis increment therefore indicates how much ‘new’ information was given to the data assimilation scheme by the targeted data. In this context, ‘new’ information may be information which is not already contained in the background field, or information that is not provided by routine observations used to constrain the background field. The size of the analysis increment is dependent on the magnitude of the difference between the background field and observations.

The difference between the CONTROL and TNOMEM analysis increments (Figure 6) shows the initial impact of the targeted sonde data. The background used in both forecasts is the same, and the only differences are that the TNOMEM forecast has assimilated targeted sonde data in addition to the routine observations assimilated by the CONTROL forecast. Therefore any differences between the analysis increments for these two forecasts identify where the targeted sonde data are being used to constrain the background field. The impact of the targeted data is localised around the region where the sondes were released. There is a cyclonic wind increment in the southern portion of the flight track, in the region of the small lee cyclone. The effect of the targeted observations on the background at this location is to strengthen this lee cyclone. This is consistent with the pressure increment, which is negative over the lee cyclone indicating that the cyclone is too weak in the CONTROL forecast.

The vertical structure of the TNOMEM-CONTROL analysis increments and CONTROL forecast background fields were examined along an east–west section at 68°N. The background fields (Figures 7(a) and (c)) show an approximately zonally oriented jet at 350 hPa near the coast of Greenland and a weak barrier flow at 950 hPa. The size of the analysis increments (Figures 7(b) and (d)) is small, relative to the background winds. The TNOMEM increment is strengthening the u -component of the flow and enhancing the southward component of the low-level flow near the coast of Greenland. The v -component of the flow is weak, and thus small analysis increments have a larger impact on the background fields.

There is no strong evidence of upstream vertical tilting in the increments, which would be an indication of a structure conducive to baroclinic growth. The lack of tilt may therefore indicate a barotropic structure, as obtaining a tilted increment would require that the observation minus first-guess field tilt with height. This pattern of increments also reflects the assumed background-error covariance structures used in the data assimilation scheme, which vary little with height.

5.3. The impact of targeted sonde data on the forecasts

It was seen in section 4 that the impact of sonde data on the forecast of the vertically summed total energy was small when averaged over the verification region. Figure 4(c) showed that for the 01 March case, a small improvement was seen in the first 30 h. To be able to

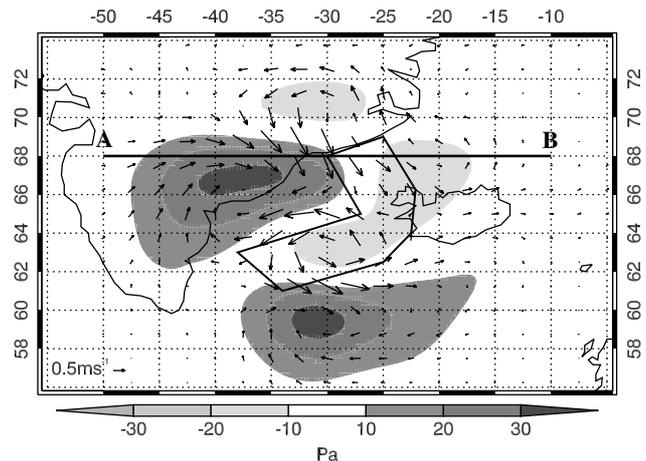


Figure 6. TNOMEM minus CONTROL analysis increment for the 01 March case: pressure (shaded) and wind strength and direction (arrows). The wind differences are a pressure-weighted vertical average over all 38 model levels, and the pressure increment difference is shown at model level 16 (≈ 500 hPa). The flight track is overlaid. The position of the cross-section displayed in Figure 7 is shown by the bold line connecting points A and B.

say that the improvement was due to the influence of the targeted sonde data, it is necessary to track the growing improvement from the target region at target time to the verification region at verification time.

Figure 8 shows the TNOMEM forecast impact during the first 36 h of the forecast, calculated using Eq. (4). The forecast impact can be split into two components: a component associated with the lee cyclone which grows *in situ*, and a second component associated with an upper-level trough and jet streak and which is advected into the verification region with a developing polar low, initially located at 72°N, 13°W. Small forecast improvement is seen in the target region at target time (Figure 8(a)). Thus the targeting has succeeded in reducing the initial condition error in the target region. Most of the initial improvement is in the southern part of the target region and is associated with the lee cyclone. This impact grows after 12 h but does not propagate into either verification region. Instead it improves the forecast of a cyclone which, initially located south of the lee cyclone (Figure 5(a)), moves into the lee of Greenland and deepens rapidly. Figure 4(a) showed no impact in the northwest Europe verification region until 36 h into the forecast. It is seen in Figure 8 that no impact moves from the target region into the northwest Europe verification region, which explains the lack of impact initially. There is some impact at 36 h to a baroclinic wave as it moves eastwards to the south of the UK. However this impact does not appear to be linked to the target region, and therefore the impact seen may be due to random noise in the initial field and cannot be attributed to the targeted observations.

At target time there is also small improvement on the eastern coast of Greenland (Figure 8(a)). This propagates east into the verification region with the developing polar low, with the improvement mostly confined to the warm sector of the polar low (Figure 8(b)). The improvement

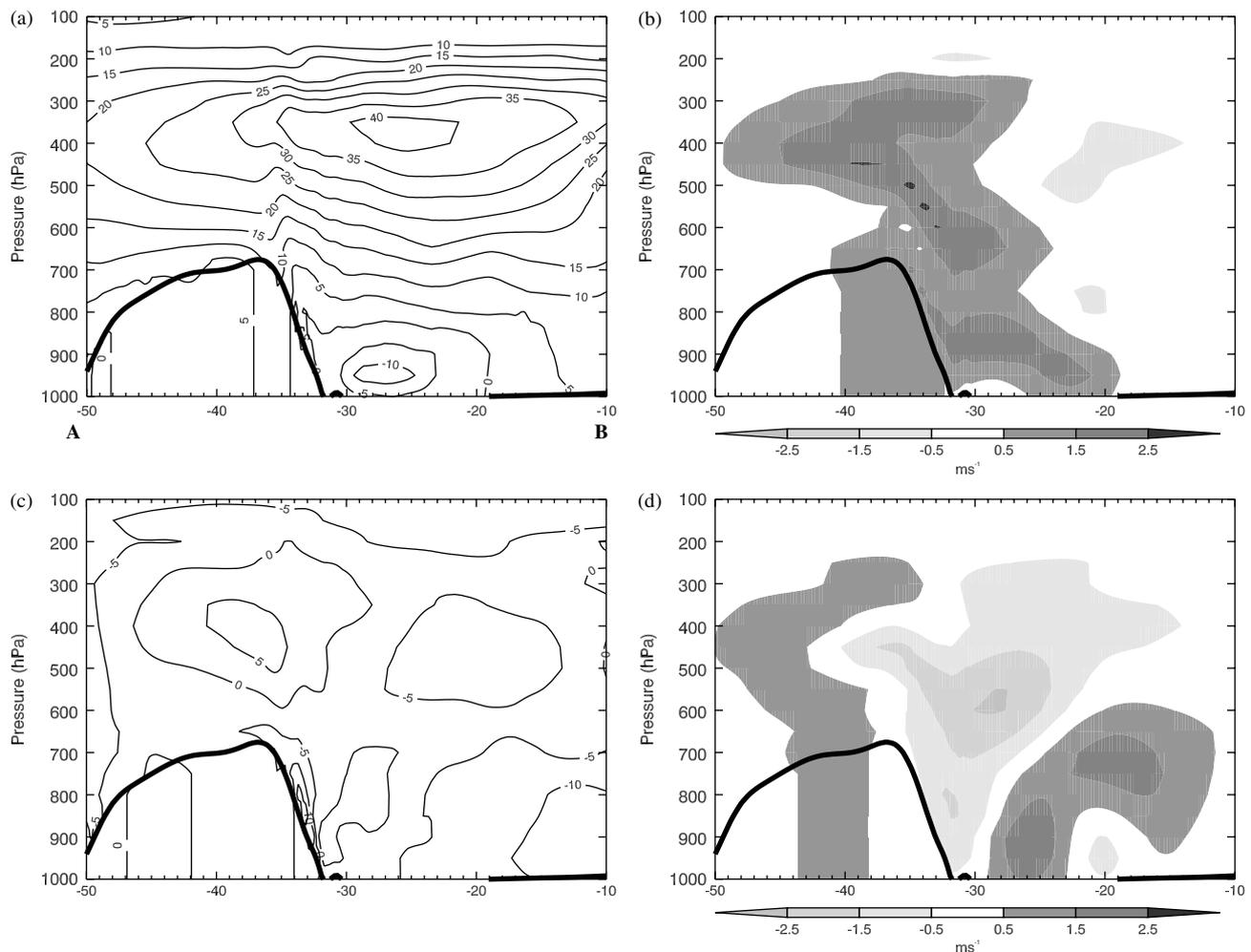


Figure 7. Cross-sections through the Denmark Strait at 68°N from 50°W to 10°W for the 01 March case. (a, c) show the background field used in the data assimilation system for (a) the u -component of wind and (c) the v -component of wind. (b, d) show the TNOMEM minus CONTROL analysis increments of (b) u -component of wind and (d) v -component of wind. The bold black line is the surface pressure, used to show the orography; the cross-section cuts through Greenland on the left. The position of A and B relative to the flight track is shown on Figure 6.

is largest 24 h after targeting (Figure 8(c)), with values greater than $30 \text{ m}^2 \text{ s}^{-2}$. The impact then moves out through the northern boundary of the verification region with the decaying polar low by 36 h after targeting (Figure 8(d)), after which time the forecast is degraded.

The targeted sondes sampled the southern part of the upper-level trough and its associated upper-level PV anomaly, and the impact was advected at upper levels by the PV anomaly, which moved with the polar low into the verification region, and caused some forecast improvement. This can be clearly seen by examining a vertical cross-section along 68°N through the forecast domain to see how the impact of the targeted data moves from the target region to the verification region. At target time the difference between the TNOMEM and CONTROL forecast is located mainly at lower levels at around 750 hPa on the eastern side of Greenland (Figure 9(a)). This propagates upwards and to the east so that, after 12 h optimisation time, the maximum difference is located at the dynamical tropopause and associated with the PV anomaly (tropopause fold) that is forcing the development of the polar low (Figure 9(c)). The

difference at this time is also associated with a warm anomaly at the surface, particularly to the north of the system where the tropopause fold is shallower and there is a strong zonal gradient in wind speed from the calmer core of the polar low to the strong low-level northerly jet to its west (not shown). The difference is initially located on the western side of the PV anomaly, and moves around the southern edge of the PV anomaly into the verification region (Figures 9(e) and (f)) as the anomaly moves eastwards. The inclusion of targeted sonde data in the forecast caused the eastern edge of the PV anomaly to be shifted west, by up to 200 km (Figure 9(e)). This caused the cyclonic circulation associated with the PV anomaly to be shifted to the west, so that when the CONTROL and TARGETED forecasts are differenced, a dipole structure is seen in the v -wind field. This is illustrated schematically in Figure 10. The mechanism that caused the impact in the Scandinavian verification region is therefore the modification of the position of a PV anomaly associated with an upper-level trough, in a region of baroclinic instability. The interaction of this upper-level trough with an occluded front led to the

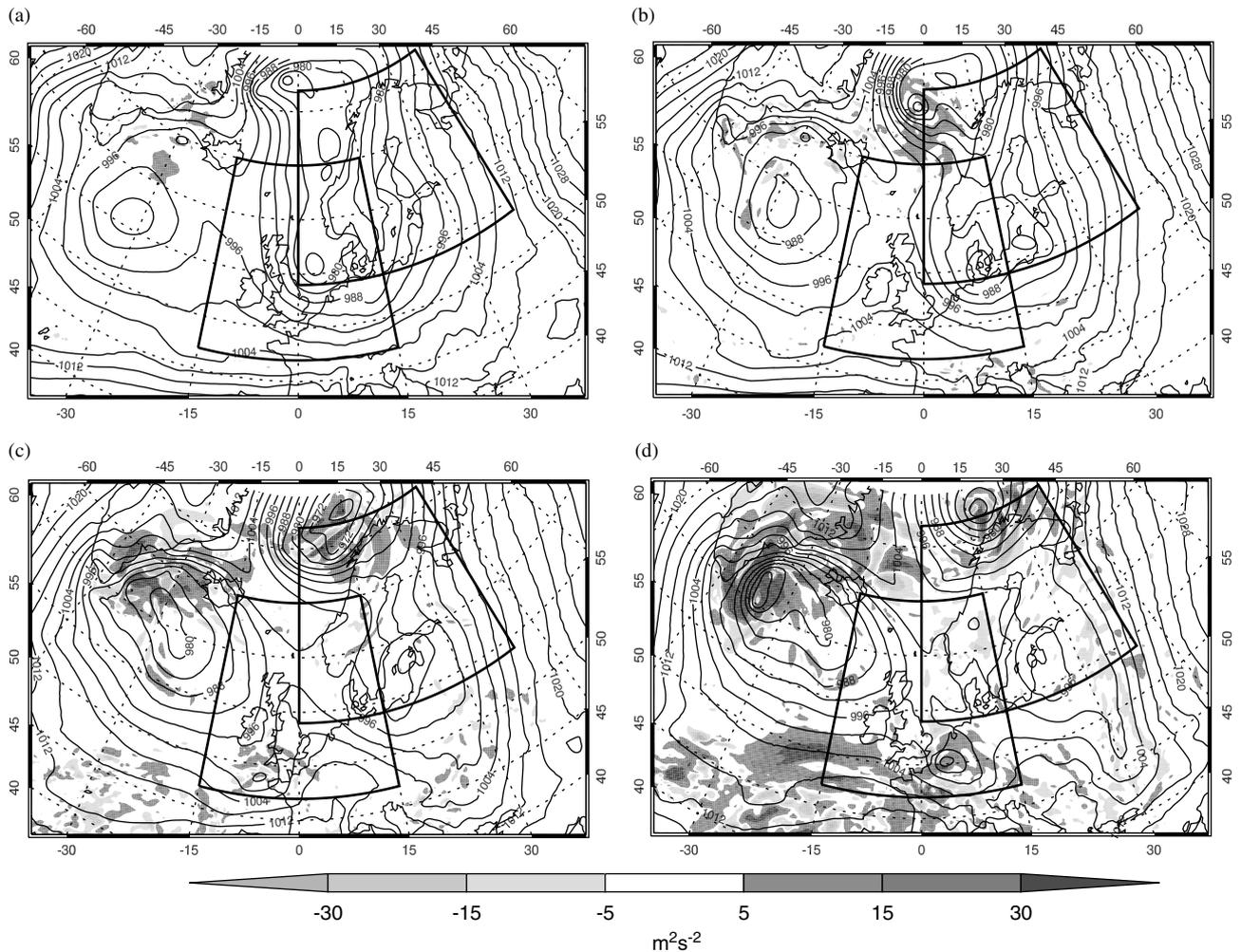


Figure 8. Impact, I (Eq. (4)), of the targeted sondes for the TNOMEM forecast of the 01 March case at (a) 0, (b) 12, (c) 24 and (d) 36 h optimisation times, shown by shading. The CONTROL forecast of mean sea-level pressure is contoured, and the verification regions are outlined.

development of a polar low, and the impact moved east with this system into the verification region.

6. Targeting case-study: 10 March 2007

A brief analysis of the 10 March case is provided here to contrast with the 01 March case. Forecast improvement of similar magnitude to the 01 March case was seen in the first 30 h in the Scandinavian verification region (Figure 4(c)), but the mechanism for the impact appears to be different.

6.1. Overview

At target time a synoptic-scale cyclone was located south-west of Iceland (Figure 5(b)), and was moving rapidly towards Iceland as it developed. During flight planning, it was noted that there was considerable uncertainty in the forecasts from different forecast centres in the location and strength of the cyclone at targeting time. This made planning a flight track difficult and resulted in the aircraft flying directly over the centre of the cyclone rather than to the west of the centre as planned. The cyclone

was located on the poleward side of the upper-level jet stream at targeting time, and although the centre of the cyclone was forecast to move over Iceland, the trailing fronts were forecast to move over the UK, through the northwest European verification region. An overview of this case is also given in Renfrew *et al.* (2008).

The ETKF-SAPs with a 36 h lead time identified the centre of the low pressure system to be most sensitive for the forecasts for Scandinavia and northwest Europe (Figure 5(b)), and these were used for flight planning. The TESV-SAPs (not shown) identified both the eastern edge of the low pressure system and a region to the far south, in the mid-Atlantic. The aircraft was not able to sample both sensitive regions; the ETKF-predicted sensitive region was chosen as this was the main sensitive region for both verification regions.

6.2. The impact of targeted sonde data on the forecasts

The impact calculated using Eq. (4) for the 10 March case is shown in Figure 11. The impact to the 1200 UTC analysis is mostly positive (Figure 11(a)), and confined to the central and northern areas of the cyclone (note

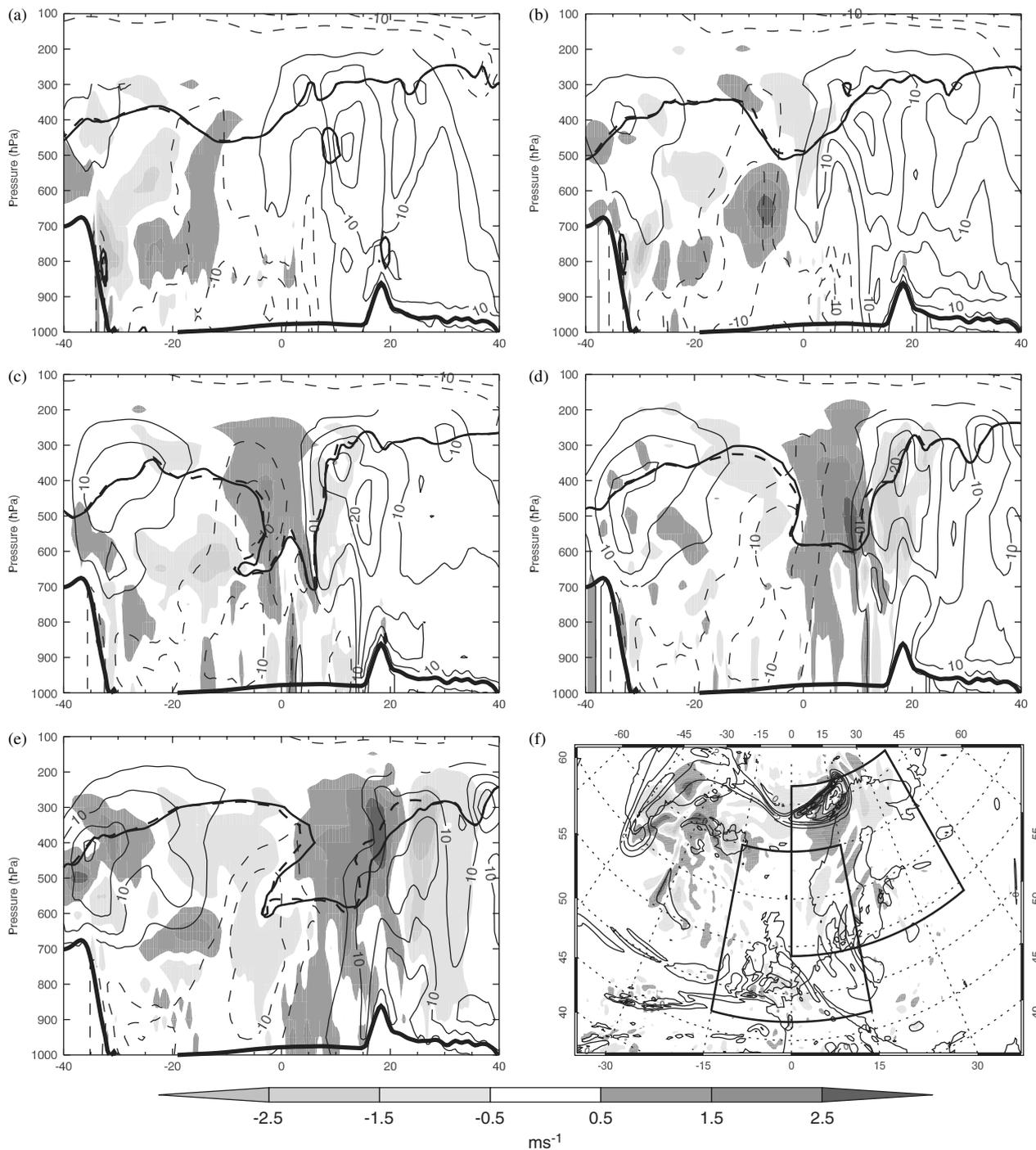


Figure 9. (a) to (e) show the TNOMEM minus CONTROL forecast (shading) of the v -component of wind at 68°N from -40°E to $+40^{\circ}\text{E}$ at (a) 0, (b) 6, (c) 12, (d) 18 and (e) 24h optimisation time for the 01 March case. The CONTROL forecast of v is contoured at 10 m s^{-1} intervals with thin solid (southerly) and dashed (northerly) lines. The dynamical tropopause is shown using the $\text{PV}=2\text{PVU}$ contour, where the CONTROL forecast is the bold solid line and the TNOMEM forecast is the bold dashed line. The surface pressure (thick black line) is used to indicate the orography, showing Greenland on the left of the cross-section and Norway on the right. (f) shows the TNOMEM minus CONTROL forecast of v at 500 hPa at 24h optimisation time, using the same scale as (a) to (e). The PV field at 500 hPa is contoured with a contour interval of 1PVU.

that the cyclone was at this stage moving slowly in a northerly direction). The cyclone stalls over Iceland as it begins to decay, and much of the impact from the targeted sondes remains with the cyclone at the later forecast times (Figure 11(c) and (d)). After the 12h forecast, the impact also begins to spread out along the fronts associated with the cyclone (which lie along

kinks in the mean sea level pressure fields). The impact is greatest at upper levels (not shown) and appears to propagate downstream along the jet stream in a manner similar to a jet streak. The impact amplifies rapidly in the downstream trough and leaves the domain to the south. This case is discussed in more detail in the following section.

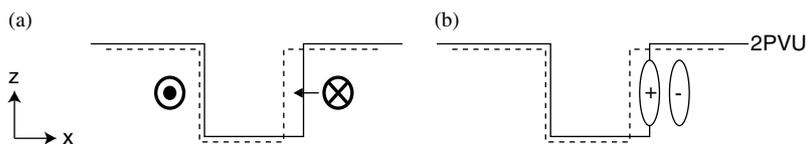


Figure 10. Schematic diagram to show how targeted sondes modified an upper-level PV anomaly and associated wind field. (a) shows the cyclonic circulation associated with an upper-level anomaly to the dynamic tropopause in the CONTROL forecast (solid line). If the eastern edge of the anomaly is moved to the west in the TNOMEM forecast (dashed line) then the associated circulation is also moved. The resulting difference in v -wind field has a dipole structure, as shown in (b).

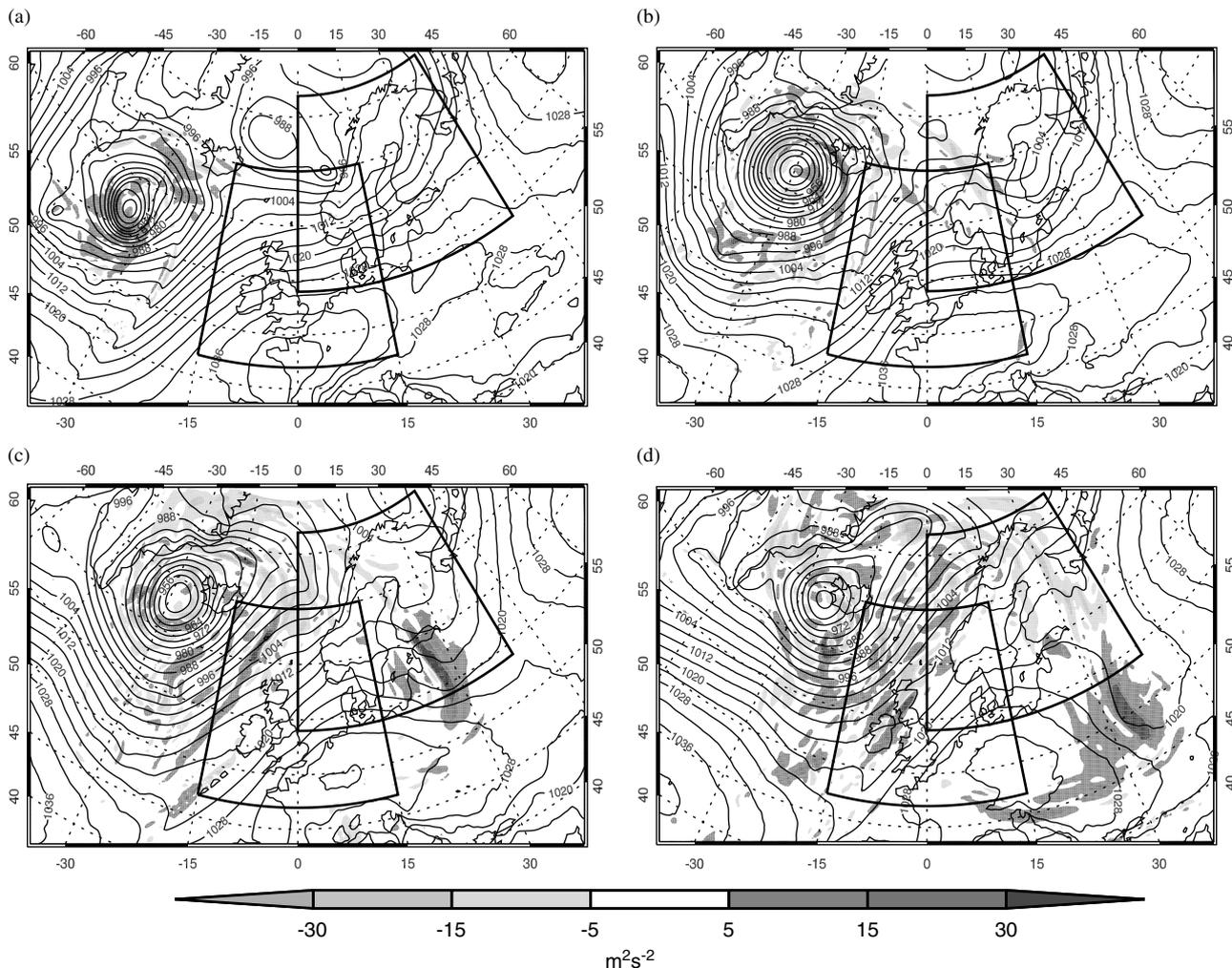


Figure 11. Impact, I (Eq. (4)), of the targeted sondes for the TNOMEM forecast of the 10 March case at (a) 0, (b) 12, (c) 24 and (d) 36 h optimisation times, shown by shading. The CONTROL forecast of mean sea-level pressure is contoured, and the verification regions are outlined.

7. The effect of cycling observations in the forecast background

The results presented thus far have focussed on the TNOMEM targeted run. This is because it provides a clean comparison with the CONTROL run, as the only differences between the two forecast runs are the addition of targeted data to the TNOMEM forecast for a particular targeting case. This also represents a situation where targeting is done infrequently, i.e. less than once a week. If targeting were done more frequently, for example if a similar area was targeted twice within a week, then perturbations from the previous targeting may

still be impacting the forecast. As a 6 h forecast from the 1200 UTC run is used as the background to the 1800 UTC forecast and so on, targeted observations can impact subsequent forecasts through the effect of memory in the forecast system.

The impact of assimilating the targeted data in the TMEM forecasts is shown in Figure 12. (Figures 12(a),(b) are equivalent to Figures 4(a),(c), but are for the TMEM forecast instead of the TNOMEM forecast.) By comparison with Figure 4, it is seen that with memory of previous targeted observations in the forecast, the magnitude of the forecast impact is increased at some forecast times, meaning that the TMEM forecasts

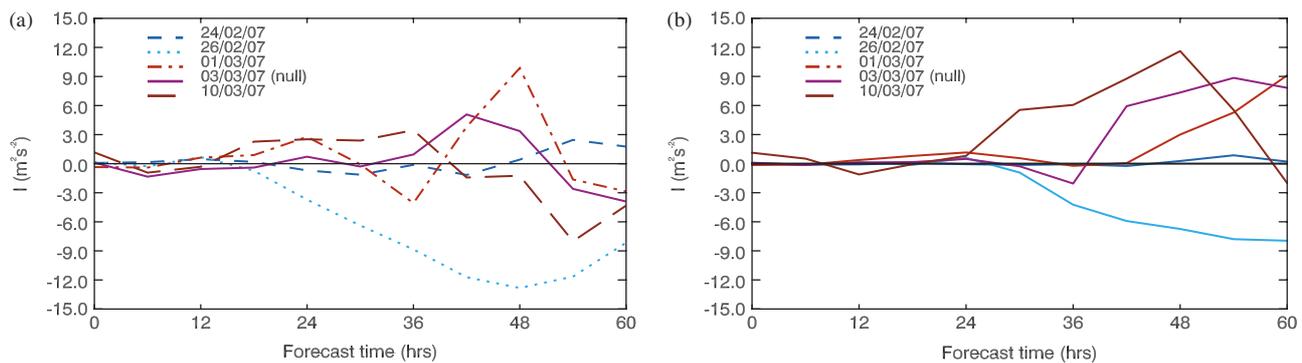


Figure 12. Impact, I (Eq. (4)), of the targeted sondes for the TMEM forecast for (a) northwest Europe and (b) Scandinavian verification regions. Note the difference in scale between this figure and Figure 4(a) and (c). This figure is available in colour online at www.interscience.wiley.com/journal/qj

are more different from the CONTROL forecast. The exception to this is the 24 February case, the first targeting experiment of the GFDex, where the TMEM and TNOMEM forecasts are, by definition, identical. It can be seen that for the 26 February case, the level of forecast degradation is increased, whereas for the final three cases the level of forecast improvement is increased in the Scandinavian verification region. It is not possible to conclude from this that the direction of the change is due to the memory of targeted observations; it is just as likely that the effect of cycling targeted observations is to introduce random perturbations into the background field, some of which grow.

Given that the hindcasts were run using a limited-area domain, it is perhaps surprising that there is such a large difference between the TNOMEM and TMEM forecast impacts for the 10 March case. As the 10 March case occurs seven days after the previous targeting case, it might be expected that any perturbations introduced into the forecast background from assimilating previous targeted observations might have moved out of the boundary of the limited-area domain, so that the TMEM and TNOMEM forecasts would show similar impact. Figure 13 shows the impact to the TMEM forecast calculated using Eq. (4) in the first 36 h of the forecast for the 10 March case. By comparison with Figure 11 (note the difference in scale for the impact), it is seen that the magnitude of the impact is increased. At initial time, the forecast impact is not confined to the area in which the sondes were released, but are spread along the fronts associated with the cyclone. This additional initial impact is due to differences between the TNOMEM and TMEM background fields (the TNOMEM and CONTROL background fields are the same). This shows that, although the GFDex dropsonde observations had a small impact on the forecast, the perturbations from targeting seem to have persisted in the forecast background for 7 d after targeting.

The inclusion of earlier sonde data in the background of the forecast does not necessarily lead to greater forecast improvement. Rather the trajectories of the CONTROL forecast and targeted forecasts differ more as the optimisation time increases when additional observational data

are included in the background of the forecast as well as being assimilated into the forecast. It can be concluded from these results that differences in the background field may sometimes have a greater impact on the forecast than the analysis increment due to the assimilation of sondes. Cardinali *et al.* (2004) showed that up to 85% of forecast information is provided by the background field rather than observations within the 4D-Var time window. That perturbations still impact forecasts up to a week later is perhaps surprising given that the perturbations are small, localised and are inserted into a limited-area domain. The increased impact of targeted observations seen from the TMEM hindcast runs also has implications for studies of targeted observing. It implies that studies that only investigate the impact of targeted observations by running a 'TMEM' forecast and compare it to a control cycle containing no targeted observations may be overestimating the direct impact of the targeted observations. This is particularly true for field campaigns, where targeting may be done many times over a short time span of a few weeks for example.

8. Robustness of the forecast improvement

It has been shown in previous sections that some improvement was seen at specific optimisation times as a result of assimilating GFDex sondes into the forecast. In this section the robustness of this improvement is inferred by seeing how sensitive the results are to the background field and analysis increment used to make the forecast.

An ensemble of five targeted forecasts was created (Table III). In this case the CONTROL forecast is not counted as an ensemble member, as this is what the perturbed forecasts are compared against. Three of the ensemble members are the ALL DATA, TMEM and TNOMEM forecasts (labelled as \times in Table III; recall that the TNOMEM forecast uses the background from the CONTROL forecast), detailed in section 3.2. Two additional ensemble members (labelled as \diamond in Table III) were created by adding the analysis increments from the TMEM and ALL DATA hindcast runs to the CONTROL background.

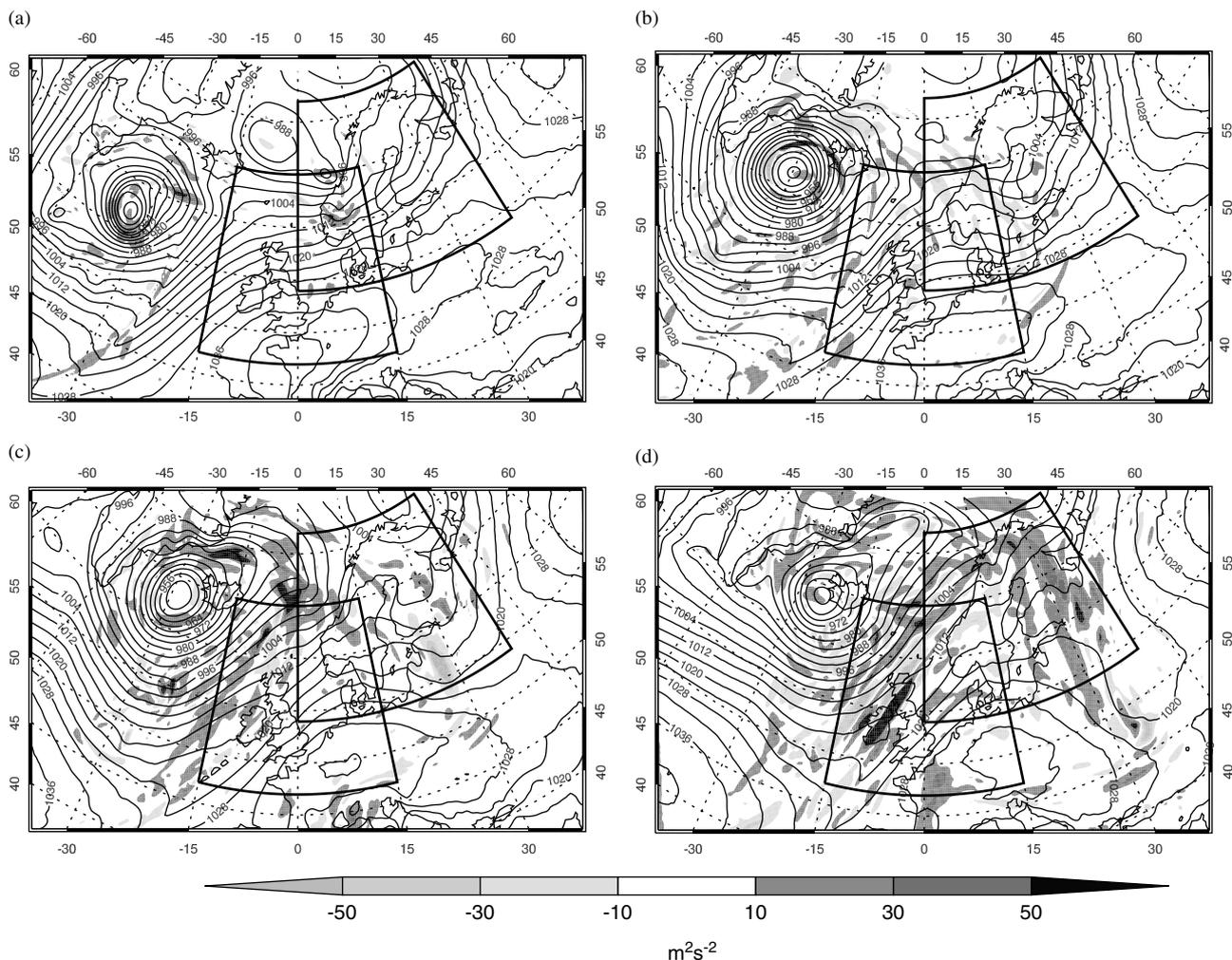


Figure 13. Impact, I (Eq. (4)), of the targeted sondes for the TMEM forecast of the 10 March case at (a) 0, (b) 12, (c) 24 and (d) 36 h optimisation times, shown by shading. The CONTROL forecast of mean sea-level pressure is contoured, and the verification regions are outlined. Note the difference in the I scale between this figure and Figure 11.

Table III. Definition of ensemble forecast members.

Forecast	Analysis increment		
	TNOMEM	TMEM	ALL DATA
background			
CONTROL	×	◇	◇
TMEM		×	
ALL DATA			×

The original simulations (section 3.2) are denoted ×. These symbols are used to identify ensemble members in Figure 14.

Figure 14(a) shows the impact for each ensemble member, calculated using Eq. (4) averaged over the Scandinavian verification region, for the 01 March case. The spread of the plume is small initially, with all members showing the robust improvement in the short-range forecast for Scandinavia that was previously identified in the TNOMEM and TMEM forecasts. The spread of the plume exceeds the absolute value of the mean from 24 h onwards, indicating that any improvement or degradation here is not robust, and due to differences in the

background field rather than being caused by the targeted data. For the 10 March case (Figure 14(b)), not all members show improvement in the forecast between 18 and 24 h after targeting, which implies that the impact seen in Figure 4(c) is not positive definite.

The aim of the ensembles shown in Figure 14 is to estimate the ratio of signal to noise associated with the value of impact. For the ensembles shown in Figure 14, the ratio of standard deviation to ensemble mean at 24 h forecast time is 0.3 for the 01 March case and 0.5 for the 10 March case. Although a small ensemble, its normalised spread is similar to the 23-member operational MOGREPS ensemble for these cases, which has a ratio of standard deviation to mean of 0.4 for the 01 March case and 0.6 for the 10 March case. It can be seen from Figure 14 that the average magnitude of impact is small. This is due to the small size of the perturbations associated with the targeted observations, which are roughly one-tenth of the size of those used to construct the MOGREPS ensemble. By design, the amplitude of a MOGREPS initial perturbation (not shown) is comparable to analysis error and its growth comparable to the growth

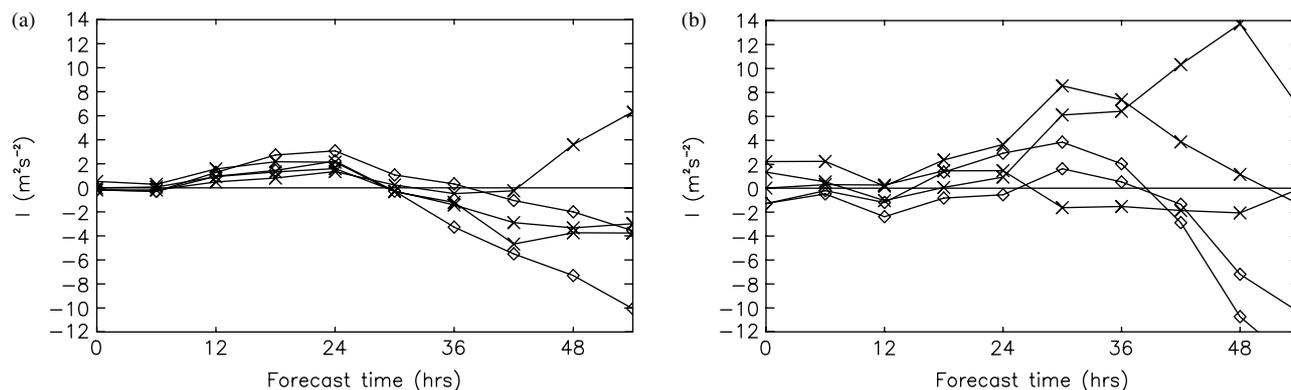


Figure 14. Ensemble plume of the impact, I (Eq. (4)), averaged over the Scandinavian verification region for (a) 01 March and (b) 10 March cases. The symbols used to mark the different members are defined in Table III.

of the forecast error over the 48 h ETKF window. The forecast error for the GFDex cases (Figures 4(b) and (d)) is also comparable in magnitude to the MOGREPS ensemble perturbations. However since the MOGREPS initial perturbations are created under assumption of linearised dynamics, which is valid at least for a 24 h window (Gilmour *et al.*, 2001), we know that if the initial perturbations were to be rescaled by a constant factor (say 0.1) and rerun, the spread/mean would be unchanged. Therefore, assuming that the perturbations associated with targeted observations evolve dynamically in a similar manner to perturbations obtained by the MOGREPS ETKF system, the spread of impact for this small ensemble gives a reasonable indication of the uncertainty in the value of impact.

9. Conclusions

Targeted observations were made during the GFDex, in the region around Iceland and southern Greenland, with the aim of improving the 24–48 h forecast over the UK and Scandinavia. The additional observational data were assimilated operationally into the Met Office forecasts. The impact of these targeted observations has been assessed by running hindcasts using a limited-area model with a 4D-Var data assimilation scheme. There were a total of four targeted observing flights into TESV or ETKF predicted sensitive regions. One additional ‘null’ flight was also included in this assessment.

For two of the targeted cases, 24 and 26 February, the difference between the background and sonde data was small, therefore the dropsonde data gave little new information to the data assimilation system, resulting in little forecast impact. For the 01 and 10 March targeting cases, the targeted dropsonde data gave new information to the forecast system and these data were, in general, used by the assimilation scheme to correct errors in the background field in the vicinity of the sondes. This resulted in a small improvement in the analysis at the target time in the target region. This improvement did not always propagate from the target region into the verification region; some improvement appeared to grow *in situ*. However, the resulting forecast improvement was

small, with a maximum forecast improvement of 5% and degradation of 3%.

The mechanisms for the forecast impact for the 01 and 10 March cases were analysed in more detail. For the 01 March case, the effect of the targeted sondes on the analysis increment was a deepening and spin-up of a low-level cyclone. It has been shown that the subsequent modification of the position of a tropopause fold and its associated wind field, which was forcing the development of a polar low, caused an improvement to the forecast over Scandinavia. For the 10 March case the impact was also mostly seen at upper levels, in the downstream development of a trough and cut-off low in the Scandinavian verification region.

The magnitude of the dropsonde signal and the impact on the forecasts was larger when previous targeted observations were included in the background state. It was shown that for one case, 10 March 2007, this caused a larger forecast improvement (for the verification region) than assimilating targeted data for that day alone. It is likely that the impact of cycling observations in the forecast background is to introduce random perturbations to the background field, some of which grow, which may improve or degrade the forecast. The implication of this result for operational targeting is that previous targeting experiments may affect the success of subsequent targeting experiments on following days.

To determine the robustness of our impact studies, a small ensemble of forecasts was created, by varying the background and analysis increment used to run each forecast. Using this method, the improvement to the Scandinavian verification region in the first 24 h of the forecast for the 01 March case was shown to be a robust result. Using the same method, it was shown that the forecast improvement seen for the 10 March case was not robust.

Overall, the impact of targeted observations seen in GFDex is both small and mixed, though generally positive or neutral for forecasts of lengths equivalent to the shortest optimisation time considered for the sensitive area predictions. This is consistent with other recent investigations of targeting over the north Atlantic. Further work is currently being carried out to investigate

whether the small impact seen in this study reflects how the additional observations are treated by the data assimilation scheme, and thus whether there is potential to increase the impact of targeted observations.

Acknowledgements

We would like to thank Philippe Bougeault, Martin Leutbecher and David Richardson (ECMWF) and Alan Thorpe (NERC) for early support and encouragement. Andrew Lawrence at ECMWF provided the TESV SAPs. The authors wish to thank Guðrún Nína Petersen for valuable discussions throughout the course of the study. We gratefully acknowledge everyone at FAAM, Avalon and Directflight for the invaluable support they provided during the field campaign. We thank the Met Office for making the MetUM available and the National Centre for Atmospheric Science Computational Modelling Support for providing computing and technical support. The study was supported by a Natural Environment Research Council (NERC) grant (NE/C003365/1).

References

- Bergot T. 1999. Adaptive observations during FASTEX: A systematic survey of upstream flights. *Q. J. R. Meteorol. Soc.* **125**: 3271–3298.
- Bishop CH, Etherton BJ, Majumdar SJ. 2001. Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Mon. Weather Rev.* **129**: 420–436.
- Bowler NE, Arribas A, Mylne KR, Robertson KB, Beare SE. 2008. The MOGREPS short-range ensemble prediction system. *Q. J. R. Meteorol. Soc.* **134**: 703–722.
- Buizza R, Cardinali C, Kelly G, Thépaut J-N. 2007. The value of targeted observations. II: The value of observations taken in singular-vector-based target areas. *Q. J. R. Meteorol. Soc.* **133**: 1817–1832.
- Buizza R, Montani A. 1999. Targeting observations using singular vectors. *J. Atmos. Sci.* **56**: 2965–2985.
- Cardinali C, Buizza R, Kelly G, Shapiro M, Thépaut J-N. 2007. The value of targeted observations. III: Influence of weather regimes on targeting. *Q. J. R. Meteorol. Soc.* **133**: 1833–1842.
- Cardinali C, Pezzulli S, Andersson E. 2004. Influence-matrix diagnostic of a data assimilation system. *Q. J. R. Meteorol. Soc.* **130**: 2767–2786.
- Dando ML. 2007. The impact of targeted satellite observations on numerical weather prediction. *Q. J. R. Meteorol. Soc.* **133**: 1945–1957.
- Fourrié N, Marchal D, Rabier F, Chapnik B, Desroziers G. 2006. Impact study of the 2003 North Atlantic THORPEX Regional Campaign. *Q. J. R. Meteorol. Soc.* **132**: 275–295.
- Gilmour I, Smith LA, Buizza R. 2001. Linear regime duration: Is 24 hours a long time in synoptic weather forecasting? *J. Atmos. Sci.* **58**: 3525–3539.
- Ingleby NB. 2001. The statistical structure of forecast errors and its representation in the Met. Office global 3D variational data assimilation scheme. *Q. J. R. Meteorol. Soc.* **127**: 209–231.
- Ingleby NB, Lorenc AC. 1998. 'Generic quality control'. OPS Scientific Documentation Paper No. 2. Met Office: Exeter, UK.
- Kelly G, Thépaut J-N, Buizza R, Cardinali C. 2007. The value of targeted observations – I: Data denial experiments for the Atlantic and the Pacific. *Q. J. R. Meteorol. Soc.* **133**: 1803–1815.
- Langland RH. 2005. Issues in targeted observing. *Q. J. R. Meteorol. Soc.* **131**: 3409–3425.
- Langland RH, Toth Z, Gelaro R, Szunyogh I, Shapiro MA, Majumdar SJ, Morss RE, Rohaly GD, Velden C, Bond N, Bishop CH. 1999. The North Pacific Experiment (NORPEX-98): Targeted observations for improved North American weather forecasts. *Bull. Am. Meteorol. Soc.* **80**: 1363–1384.
- Lawrence AR, Leutbecher M, Palmer TN. 2009. The characteristics of Hessian singular vectors using an advanced data assimilation scheme. *Q. J. R. Meteorol. Soc.* **135**: 1117–1132.
- Leutbecher AR, Barkmeijer J, Palmer TN, Thorpe AJ. 2002. Potential improvements to forecasts of two severe storms using targeted observations. *Q. J. R. Meteorol. Soc.* **128**: 1641–1670.
- Leutbecher MA, Doerenbecher A, Grazzini F, Cardinali C. 2004. 'Planning of adaptive observations during the Atlantic THORPEX regional campaign 2003'. *ECMWF Newsletter* **102**: 16–25.
- Liu H, Zou X. 2001. The impact of NORPEX targeted dropsondes on the analysis and 2–3-day forecasts of a landfalling Pacific winter storm using NCEP 3D-Var and 4D-Var systems. *Mon. Weather Rev.* **129**: 1987–2004.
- Majumdar SJ, Bishop CH, Etherton BJ, Szunyogh I, Toth Z. 2001. Can an ensemble transform Kalman filter predict the reduction in forecast error variance produced by targeted observations? *Q. J. R. Meteorol. Soc.* **127**: 2803–2820.
- Majumdar SJ, Bishop CH, Etherton BJ. 2002. Adaptive sampling with the ensemble transform Kalman filter. Part II: Field program implementation. *Mon. Weather Rev.* **130**: 1356–1369.
- McNally AP. 2002. A note on the occurrence of cloud in meteorologically sensitive areas and the implications for advanced infrared sounders. *Q. J. R. Meteorol. Soc.* **128**: 2551–2556.
- Montani A, Thorpe AJ, Buizza R, Undén P. 1999. Forecast skill of the ECMWF model using targeted observations during FASTEX. *Q. J. R. Meteorol. Soc.* **125**: 3219–3240.
- Palmer TN, Gelaro R, Barkmeijer J, Buizza R. 1998. Singular vectors, metrics and adaptive observations. *J. Atmos. Sci.* **55**: 633–653.
- Persson A. 1999. 'Synoptic-dynamic diagnosis of medium range weather forecast systems'. In *Proceedings of Seminar on Diagnosis of Models and Data Assimilation Systems*. ECMWF: Reading, UK pp 123–137.
- Petersen GN, Majumdar SJ, Thorpe AJ. 2007. The properties of sensitive area predictions based on the ensemble transform Kalman filter (ETKF). *Q. J. R. Meteorol. Soc.* **133**: 697–710.
- Petersen GN, Renfrew IA, Moore GWK. 2009. An overview of barrier winds off southeastern Greenland during GFDEX. *Q. J. R. Meteorol. Soc.* In press.
- Petersen GN, Thorpe, AJ. 2007. The impact on weather forecasts of targeted observations during A-TReC. *Q. J. R. Meteorol. Soc.* **133**: 417–431.
- Rawlins F, Ballard SP, Bovis KJ, Clayton AM, Li D, Inverarity GW, Lorenc AC, Payne TJ. 2007. The Met Office global four-dimensional variational data assimilation scheme. *Q. J. R. Meteorol. Soc.* **133**: 347–362.
- Renfrew IA, Moore GWK, Kristjánsson JE, Ólafsson H, Gray SL, Petersen GN, Bovis K, Brown P, Føre I, Haine T, Hay C, Irvine EA, Ohigashi T, Outten S, Pickart RS, Shapiro M, Sproson D, Swinbank R, Woolley A, Zhang S. 2008. The Greenland Flow Distortion Experiment. *Bull. Am. Meteorol. Soc.* **89**: 1307–1324.
- Szunyogh I, Toth Z, Emanuel KA, Bishop CH, Snyder C, Morss RE, Woollen J, Marchok T. 1999. Ensemble-based targeting experiments during FASTEX: The effect of dropsonde data from the Lear jet. *Q. J. R. Meteorol. Soc.* **125**: 3189–3217.
- Szunyogh I, Toth Z, Morss RE, Majumdar SJ, Etherton BJ, Bishop CH. 2000. The effect of targeted dropsonde observations during the 1999 Winter Storm Reconnaissance Program. *Mon. Weather Rev.* **128**: 3520–3537.
- Szunyogh I, Toth Z, Zimin AV, Majumdar SJ, Persson A. 2002. Propagation of the effect of targeted observations: The 2000 Winter Storm Reconnaissance Program. *Mon. Weather Rev.* **130**: 3520–3537.