Tellus (2008), 60A, 249–260
Printed in Singapore. All rights reserved

TELLUS

A cycled sensitivity observing system experiment on simulated Doppler wind lidar data during the 1999 Christmas storm 'Martin'

By GERT-JAN MARSEILLE*, AD STOFFELEN and JAN BARKMEIJER, *KNMI, Wilhelminalaan 10, De Bilt, The Netherlands*

(Manuscript received 15 January 2007; in final form 15 October 2007)

ABSTRACT

In a companion paper in this issue Sensitivity Observing System Experiment (SOSE) has been introduced as a new method to assess the potential added value of future observing systems for Numerical Weather Prediction (NWP). There, SOSE was introduced as a single cycle experiment, meaning that additional synthetic observations, to extend the existing global observing system (GOS), are applied in only one assimilation cycle. In this study SOSE has been extended to run over three subsequent days to enable impact assessment of additional prospective observations over a prolonged period prior to an event. This is achieved by a cycled implementation of the SOSE method where analysis adaptations from previous cycles evolve progressively in subsequent cycles. This implementation of a cycled SOSE results in a sequence of pseudo-true atmospheric states that are subsequently used for the simulation of prospective extensions of the existing GOS. A cycled SOSE has two attractive properties as compared to the single-cycle implementation (i) the resulting pseudo-true atmospheric state at the end of the cycling period, that is, at forecast initial time, provides a better forecast and (ii) the cycling implementation makes the SOSE method more suitable for absolute impact assessment of continuously operated observing systems such as from polar satellite platforms. The NWP case investigated concerns the 1999 Christmas storm 'Martin' that caused much havoc in Western Europe. We show that additional observations from a spaceborne Doppler wind lidar over a 3-d period would have improved the 2-d forecast of 'Martin' substantially. This is substantiated by a 50-member ensemble forecast run.

1. Introduction

Meteorological data assimilation systems are usually cycled over subsequent time windows of several hours length. Information from observations in one cycle is propagated forward in time to next cycles and contribute to the following analyses in a constructive manner. The full impact of a new observing system can therefore only be assessed in a cycled experiment. Another time aspect lies in the local sampling of for example polar satellite instruments which only sample twice daily at any longitude (away from the poles). To assess the impact of prospective polar satellite observing systems in for example particular cases of extreme or dynamic weather one would therefore need to consider an extended time window over several cycles for experimentation. In this manuscript we report on the effect of data assimilation system cycling for a Sensitivity Observing System Experiment (SOSE). The case under investigation is the storm called 'Martin' that hit Western Europe over Christmas in 1999 and caused large social and economic losses.

SOSE is a new method in Numerical Weather Prediction (NWP) to assess the impact of prospective extensions of the global observing system (GOS) and is presented in a companion paper in the context of other experiments for NWP impact assessment, see Marseille et al. (2007a), hereafter MSBa. In a SOSE a reference analysis is modified synthetically in a so-called sensitivity computation such that the adapted analysis still fits with the observations of all existing observing systems, but at the same time does improve the subsequent NWP 2-d forecast. In MSBa it is shown that the forecast improvements that are constrained on the 2-d forecast, actually are maintained over the full useful forecast range (7 d). This suggests that improvements brought by a single cycle SOSE beneficially affect the weather model trajectory over a prolonged period, and in fact may be reinforced by the next single cycle SOSE computation in the data assimilation process. One would thus expect that cycling results in larger 2-d forecast improvement. Moreover, in MSBa the spatial properties of the structure of these unobserved analysis modifications were verified to be similar to those of the control analysis

*Corresponding author.
e-mail: Gert-Jan.Marseille@knmi.nl
DOI: 10.1111/j.1600-0870.2007.00290.x

error, and the modifications are therefore considered to be realistic. The adapted atmospheric state is denoted pseudo-truth and is used to simulate the new observing system(s). A consistent GOS with extended capability is thus established, and simulated and existing observations are ready to be used simultaneously in the SOSE. The better the prospective observing system is capable of observing the analysis modifications as contained in the pseudo-truth, the more beneficial impact it has and the larger the NWP forecast improvement will be.

SOSE was used in a second companion paper of Marseille et al. (2007b), hereafter MSBb, to test prospective Doppler wind lidar (DWL) scenarios on their capability to prevent large forecast failures by measuring rapidly growing structures not captured by the existing GOS. The SOSEs were conducted in a single cycle mode, where synthetic observations are added in one data assimilation time window only. As such, the observation impact is not fully representative for the expected impact of the observing system in operational NWP. For an absolute impact assessment of prospective observing systems, the SOSE method needs to be applied in so-called cycling mode to take full advantage of nowadays data assimilation systems to propagate observational information forward in time and as such to exploit observations of the new observing system in the time prior to the event.

The objective of this paper is to study whether the SOSE method can be extended to generate a pseudo-true trajectory, extending over a prolonged period prior to the event, rather than a single-cycle pseudo-true atmospheric state (MSBa). This may be achieved by a sequence of pseudo-true atmospheric states as discussed in Section 2. The hypothesis is that sensitivity structures from previous cycles progressively evolve, giving better forecast initial states in later cycles. An additional sensitivity computation, initiated with a better forecast initial state, further improves the forecast initial state potentially giving a better pseudo-true atmospheric state then in single cycle mode. The experimental setup is described in Section 2 and the hypothesis is verified in Section 3 for a case study including the Christmas storm 'Martin' over Western Europe in December 1999. It is shown that the sequence of pseudo-true atmospheric states resulting from the cycling implementation fulfils the pseudo-truth requirements as defined in MSBa: the pseudo-truth (i) is realistic, mimicking real analysis errors, (ii) is compatible with existing observations and (iii) improves the 2-d forecast. In addition, the forecast resulting from the cycled implementation is better than from the single cycle implementation, an indication of the validity of the hypothesis for this particular case.

For the obtained trajectory of (pseudo-true) atmospheric states the tandem-Aeolus scenario described in MSBb is simulated to generate profiles of synthetic horizontal line-of-sight (HLOS) wind components that are subsequently used together with existing observations in a SOSE impact experiment as explained in Section 3. In Section 4, we zoom in on the second Christmas storm 'Martin' making landfall in Brittany (Fr) on 27 December 18 UTC and causing much damage over France and Germany in the following 24 h. It is shown that additional observations from a spaceborne DWL would have improved the forecast of 'Martin' substantially. The results are summarized in Section 5 followed by conclusions and recommendations.

2. Setup of a cycled SOSE

A proper SOSE experiment requires a sensitivity computation, for the definition of the pseudo-truth, with the same model that produced the forecast and verifying analysis, see MSBa. Therefore a rerun of the case with a recent model version is performed as a first step. The case rerun is in the remainder also denoted control experiment or CONTROL. The rerun extends over a period including a 2-d lead period, the cycling period and the forecast range, see Fig. 1. The atmospheric state is initiated at t_0, in our implementation with model fields from the ERA-40 re-analysis experiment Uppala et al. (2005). After the 2-d lead-time possible spin-up problems due to biases between the different model versions used for ERA-40 and the SOSE experiment will be eliminated. The cycling period is composed of consecutive assimilation windows of typically 6 or 12 h. In each window, a SOSE analysis experiment is performed, as described in MSBa and shortly summarized below. Part of the SOSE analysis is a background sensitivity experiment that includes a 48-h forecast initiated with the background state at the beginning of the assimilation window, that is, the state at the end of the previous window in a 4D-Var context. For the first window in the cycling period the background state is extracted from the rerun. The forecast is verified against the analysed state from the rerun 48 h later. Key background errors are obtained from a sensitivity computation using the forecast error as input and the background error covariance matrix in the initial time norm, see MSBa. The adapted background is used in the analysis with existing observations to produce the SOSE analysis, also denoted pseudo-truth. This process is repeated until the end of the cycling period.

This cycling procedure propagates sensitivity structures from previous cycles forward in time. It is expected that (part) of these structures are maintained and evolve in subsequent cycles, in particular over data sparse areas. In addition, it is known that sensitivity structures not only improve the 48-h forecast but also forecasts beyond, see for example, Isaksen et al. (2005) and MSBa. This property of sensitivity structures suggests a progressive evolution of sensitivity structures during cycling providing better forecasts at the end of the cycling period than the single cycle mode forecasts. The result of the cycling procedure is a sequence of pseudo-true atmospheric states over a prolonged period prior to the event that are subsequently used for the simulation of the new observing system(s). The cycled SOSE procedure is applied to the Christmas 1999 period in the next section.

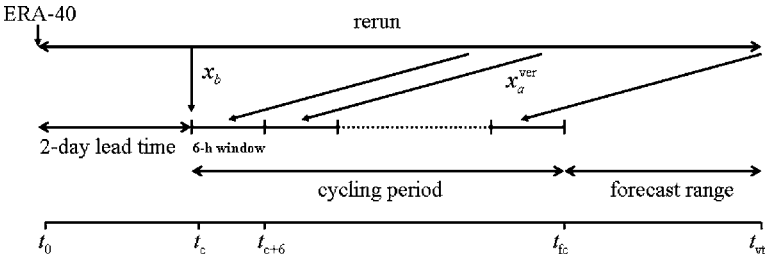

Fig. 1. Timescale of a typical cycled SOSE experiment. t_0 and t_c denote the start of the rerun and cycling period, respectively. The rerun is initiated with fields from the ERA-40 experiment. Cycling is initiated with the background fields, x_b, from the rerun. t_{fc} and t_{vt} denote the forecast initial and verification time. In each 6-h window in the cycling period a background sensitivity experiment and SOSE analysis are conducted as displayed in Fig. 1 of MSBa, using the verifying analysis, x_a^{ver}, from the rerun as input.

3. Cycled SOSE case study

The procedure described in the previous section to generate a pseudo-true trajectory over a prolonged period through a sequence of pseudo-true atmospheric states is applied to the December 1999 Christmas period. This period was characterized by two storms named 'Lothar' and 'Martin' causing much havoc in Western Europe, see for example, Buizza and Hollingsworth (2002). These storms are interesting for case studies because almost all operational NWP models failed to forecast these storms even on the short (48-h) term. In this section, we focus on the 2-d forecast of 500 hPa geopotential height (Z500) for 28 December 12 UTC, for which the operational model (OPER) is in the top 3 of worst ECMWF forecasts over Europe for the period 1998–2004, see MSBb. The forecast error is displayed in Fig. 2a and quantified in the second row of Table 1. In the next section we zoom in on the second Christmas storm 'Martin'.

According to the timetable in Fig. 1 a re-run is conducted over a 7-d period. The model version operational at ECMWF from October 2003 until March 2004 (26r3) was used for this purpose. The rerun starts at 21 December 12 UTC (t_0) initiated with model fields from the ERA-40 re-analysis experiment and finishes at forecast verification time (t_{vt}) 28 December 1999 12 UTC. Figure 2b shows that the 2-d rerun (CONTROL) forecast is slightly better than 1999 operations (OPER), see also Table 1. The cycling period extends over 84 h and is composed of 14 6-h assimilation windows, the first centred at 23 December 06 UTC (t_c) and the last at 26 December 12 UTC. The objective is to improve the 2-d forecast verifying on 28 December 12 UTC (t_{vt}) by adapting the forecast initial state (control analysis) on 26 December 12 UTC (t_{fc}). Two experiments are conducted for this purpose.

(1) A single-cycle SOSE experiment as in MSBb. This is in fact a special case of the cycling experiment, that is, using a single cycle window with $t_c = t_{fc}$. Figure 3a and b show the temperature and wind analysis adaptations, defining the pseudo-truth (control analysis plus adaptation). The pseudo-truth is a better forecast initial state, reducing the Z500 forecast error over Europe to 34.03 m as compared to 44.66 m for the control forecast or by 24%, see Table 1 and Fig. 2c.

(2) A cycled SOSE experiment. The resulting temperature and wind analysis adaptations at forecast initial time (t_{fc}), that is, the difference of the pseudo-truth and control analysis, are displayed in Fig. 3c and d. Clearly, the amplitude of the adaptations is much larger than for a single-cycle experiment by a factor of 4 on average up to a factor of 10 locally, see also Fig. 4. Also, most of the adaptation is over the data sparse oceans as expected, because of generally better analyses over the continents. The resulting forecast initiated with the pseudo-truth on 26 December 12 UTC is better than the forecast initiated with the pseudo-truth from the single-cycle experiment, see Fig. 2c and d. In cycling mode the Z500 forecast error over Europe is reduced to 26.49 m or by 41% with respect to the control forecast, see Table 1.
Another way of showing the growth of the amplitude of the analysis adaptations during cycling is in Fig. 5. The energy of the adaptation in the first cycle, Fig. 5a, is similar to the amplitude of the adaptation in a single-cycle experiment. The energy increases during cycling with a maximum reached after 72 h (not shown). This anticipated saturation of sensitivity correction impact is important and demonstrates the inherent ability of the data assimilation system to follow the observations closely and prevent excessive deviations of the modelled atmosphere from the true atmospheric state. As a consequence, the analysis adaptations generated in cycling mode are well within the analysis error variance statistics, for example, compare to the energy of analysis errors in fig. 10 of MSBa.

3.1. Cycled SOSE verification

For a proper SOSE experiment, the data usage and observation statistics of existing observing systems should be similar to the CONTROL experiment. A large deviation would indicate either inconsistency between the pseudo-true atmospheric state and existing observations resulting, for example, in a larger rejection of observations in the SOSE experiment and/or larger deviations of observations from the background. Table 2 shows that the data usage of aircraft and radiosonde wind observations is

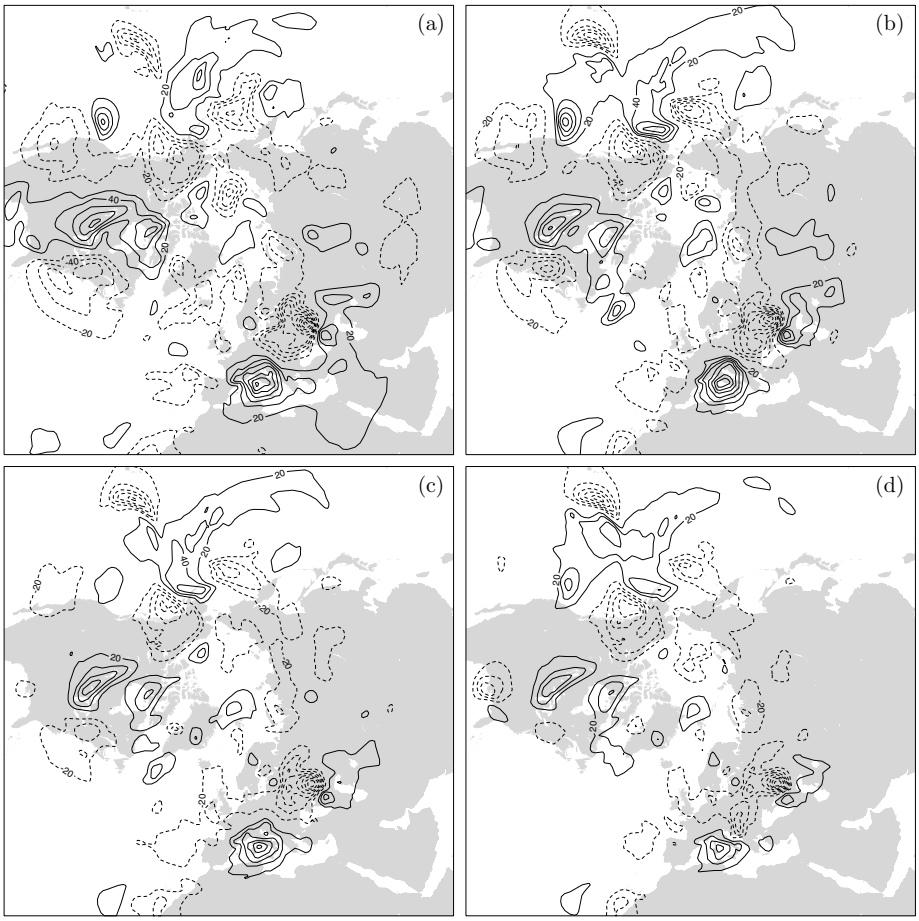

Fig. 2. ECMWF 2-d forecast error of 500 hPa geopotential height, verifying at 28 December 1999 12 UTC for the forecast initiated with the analysis at 26 December 12 UTC (t_{fc}) from (a) the 1999 operational model, (b) the rerun (control) using the 2004 operational model (version 26r3), (c) the single-cycle experiment pseudo-truth and (d) the 84-h cycling pseudo-truth. The contour interval is 20 m. Solid/dashed contour lines are for positive/negative values.

very similar for the SOSE and CONTROL experiment. For radiosondes the data usage is slightly larger (0.3%) in the SOSE experiment. The usage in both experiments for other observation types is also very similar (not shown). The observation minus background (o-b) and observation minus analysis (o-a) statistics for aircraft wind observations in Fig. 6 and the numerical values in Table 3 also show very similar values for the used observations, with slightly smaller biases and larger root-mean-square errors for aircraft and radiosonde winds in the SOSE experiment. Observation statistics of other observation types are also very similar in both experiments. It is concluded that the data usage and statistics of existing observations in the CONTROL and cycled SOSE experiments are very similar. This implies that differences between the control and SOSE impact experiments may be fully attributed to the additional new observing system(s). The next section demonstrates the added value of DWL on the Z500 forecast.

3.2. Cycled SOSE DWL impact experiment

The capability of future spaceborne DWLs to extend the GOS progressively and improve forecasts is assessed for the tandem-Aeolus scenario, also denoted T_2A, as described in MSBb. The tandem-Aeolus scenario is a constellation of two Aeolus satellites in the same orbit but separated by 180°. Aeolus is a polar orbiting satellite carrying a DWL to give a global coverage of profiles of single HLOS wind components, see Stoffelen et al. (2005). The tandem-Aeolus scenario has double the coverage of a single Aeolus as described in MSBb. HLOS wind observations are simulated as explained in section 4 of Marseille et al. (2007b), that is, using the SOSE pseudo-truth as true atmospheric state and including observation errors that take into account instrumental characteristics and atmosphere optical properties. Three DWL impact experiments have been conducted, denoted EXP1, EXP2 and EXP3:

Table 1. 2-d forecast error of 500 hPa geopotential height (m) for global regions (N/S/W/E): N.Hemis (90/30/−180/180), Europe (75/35/−12.5/42.5) and N.America (60/25/−120/−75). Verification time is 28 December 1999 12 UTC. OPER corresponds to the 1999 ECMWF operational model, CONTROL to the rerun with the 2004 model version (26r3). Both are based on observations from the existing GOS only. The pseudo-truth experiments use the modified analysis (pseudo-truth) at t_{fc} as forecast initial state, the DWL-T_2A experiments use observations from the existing GOS and additional synthetic DWL observation from the tandem-Aeolus scenario, according to EXP1, EXP2 and EXP3 (see text)

Z500 RMSE (m)	N.Hemis	Europe	N.America
OPER	31.20	47.09	38.00
CONTROL	29.78	44.66	32.46
Pseudo-truth; single-cycle	23.61	34.03	23.56
Pseudo-truth; cycling	22.24	26.49	21.64
DWL-T_2A; EXP1	28.50	44.58	29.27
DWL-T_2A; EXP2	26.36	42.86	28.61
DWL-T_2A; EXP3	24.07	34.20	26.39

(1) *EXP1.* A single-cycle SOSE experiment, using existing observations and 6 h of synthetic DWL observations with the locations of the satellite tracks selected to give maximum coverage over the Northern Hemisphere oceans, as in MSBb, see Fig. 7.

(2) *EXP2.* A cycled SOSE experiment using existing observations and synthetic DWL observations, the latter only added in the last two cycles (12 h) of the cycling period. Then, half of the tracks are over the data dense continents giving minimal added value for NWP. The DWL coverage over the oceans is thus similar to the single-cycle experiment, but sampling a different pseudo-truth, according to Fig. 3b and d.

(3) *EXP3.* A cycled SOSE experiment using existing observations and 84 h synthetic DWL observations covering the complete cycling period.

Realistic polar orbiting scenarios were used in all three experiments, with no preference to target specific (data sparse) areas. Figure 7 shows the impact, Δ_{an}, of DWL on the 26 December 12 UTC analysis for the three experiments. Here, impact is defined

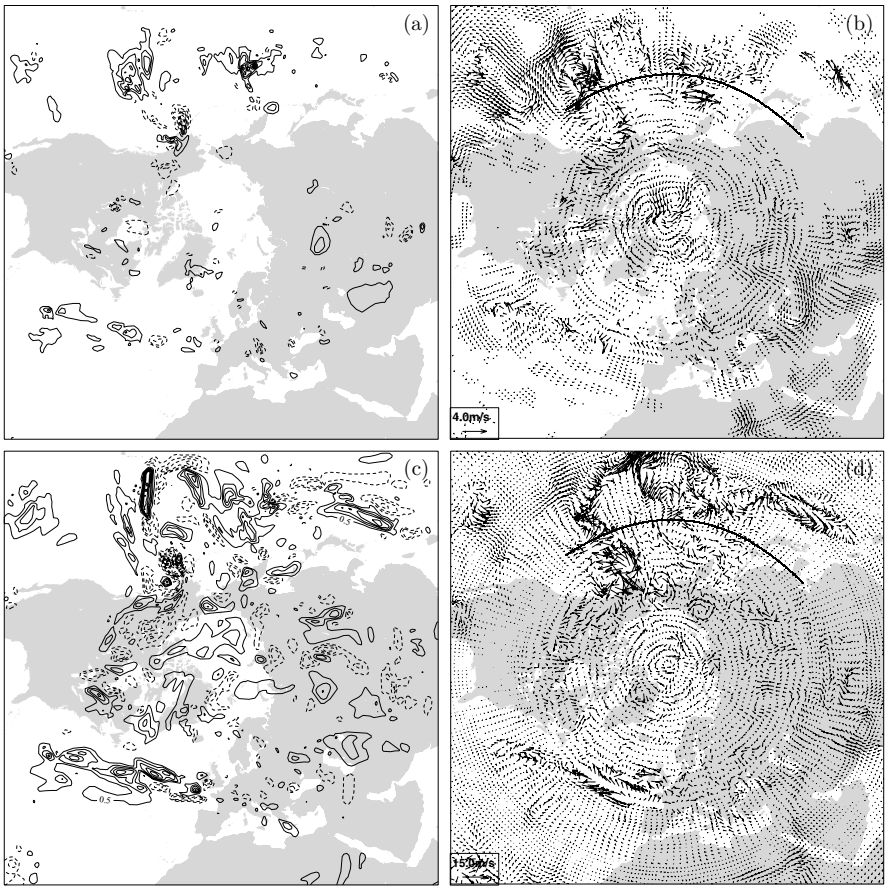

Fig. 3. 500 hPa SOSE analysis adaptations at forecast initial time (t_{fc}) 26 December 1999 12 UTC from a single-cycle (top row) and 84-h cycling (bottom row) experiment. Left panels show temperature with a contour line interval of (a) 0.25 K and (c) 0.5 K . Solid/dashed contour lines are for positive/negative values. The right panels show wind (ms^{-1}). Note that the wind arrow unit velocity (at the bottom left in the panels) is about a factor of 4 larger in (d) as compared to (b). The solid thick lines over the Pacific near 40°N show the location of the vertical cross-section displayed in Fig. 4.

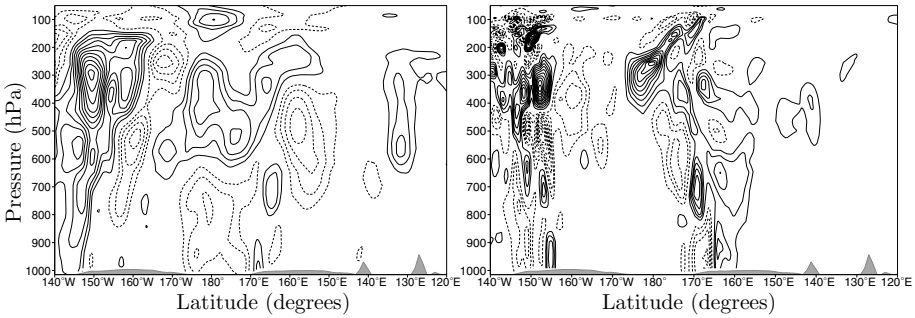

Fig. 4. Cross-section of zonal wind component analysis adaptation over the Northern Pacific at 40°N, see Fig. 3, for 26 December 1999 12 UTC based on a single-cycle (left-hand panel) and 84-h cycling (right-hand panel) SOSE experiment. The contour interval in the left/right-hand panel is 0.1/1 ms^{-1}. Solid/dashed contour lines are for positive/negative values.

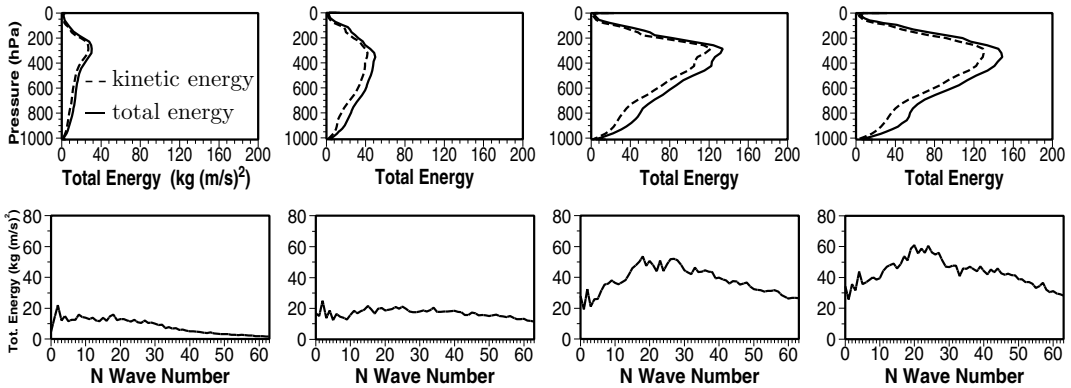

Fig. 5. Energy (kg m^2 s^{-2}) evolution of SOSE analysis adaptations (top row) and the horizontal total energy spectrum (bottom row) over the 84-h cycling period. Panels from left to right correspond to time instants 23 December 6 UTC, 24 December 12 UTC, 25 December 12 UTC and 26 December 12 UTC. Dashed and solid lines in the top row correspond to kinetic and total energy, respectively.

Table 2. Global data usage for AIREP (aircraft) and RAOB (radiosonde) zonal wind components in the control (CTL) and cycled SOSE (CYC) experiment in the 6-h window centred at 26 December 1999 12 UTC, that is, corresponding to the last window (at t_{fc}) of the cycling period. The values correspond to the total/used/rejected/blacklisted number of observations. A fraction of the rejected observation is related to variational quality control (varQC) during assimilation

# Observations		TOTAL	USED	varQC	REJ	BLK
AIREP-U	CTL	10751	8539	43	2211	1
	CYC	10751	8537	40	2213	1
RAOB-U	CTL	17882	15890	152	967	1025
	CYC	17882	15945	164	912	1025

as the difference between the absolute values of the analysis errors for the CONTROL and DWL experiments:

$$\Delta_{\text{an}} = \left|x_a^c - x_t^p\right| - \left|x_a^d - x_t^p\right|. \quad (1)$$

Here, x_a^c and x_a^d denote the analysis for the CONTROL and DWL experiment, respectively, and analysis error is defined with respect to the pseudo-truth, x_t^p. Positive/negative values indicate positive/negative DWL impact. It is clear from the amplitude of the analysis adaptations that larger impact values are expected from the cycling experiment. This is confirmed in Fig. 7 and Table 4. In addition, by comparing Fig. 7c and e and the last two columns in Table 4 it follows that DWL observations from previous cycles add progressively to the quality of backgrounds and analyses in subsequent cycles, resulting in a better forecast initial state than when using DWL observations only close to the forecast initial time. Also the forecast over Europe is substantially improved when using additional DWL over a prolonged period prior to the forecast initial time, see Fig. 7b,d and f and the last three columns of Table 1. The Z500 forecast error over Europe goes down from 44.66 m (CONTROL), 44.58 m (EXP1), 42.86 m (EXP2) to 34.20 m (EXP3) or by 0.2, 4.0 and 23.4% for the DWL experiments, respectively, with respect to the control forecast.

Following the conclusion from Section 3.1 we conclude that the improved quality of analyses and forecasts in the DWL experiments may be attributed to the additional DWL observations. Moreover, it is clear that the cycled DWL impact is a much larger fraction of the maximum achievable impact than the single-cycle

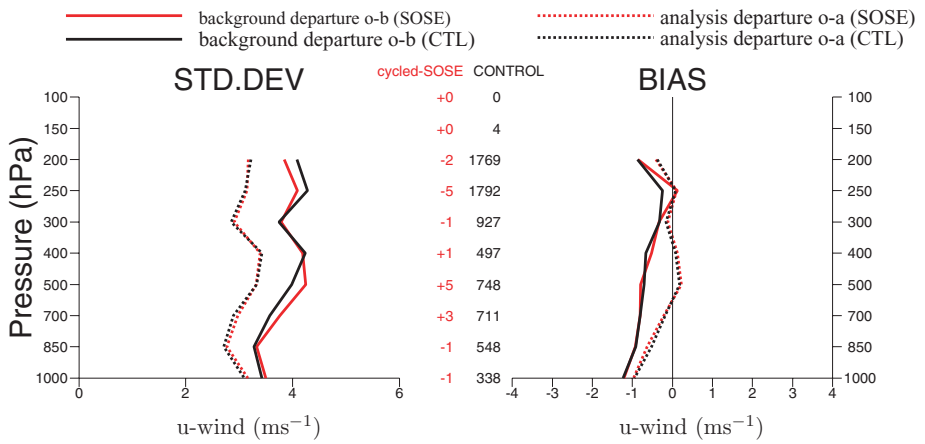

Fig. 6. Observation statistics of the zonal component of AIREP winds (ms^{-1}) for the Northern Hemisphere extra tropics above 30 N, valid in the last window of the cycling period on 26 December 1999 12 UTC. Standard deviation left and bias right. Black/red corresponds to the CONTROL (rerun)/cycled SOSE experiment. The black numbers between the two figures denote the total used observations in the analysis by the CONTROL experiment. The numbers in red the additional number of used observations in the cycled SOSE experiment.

Table 3. Statistics of used observations, see Table 2, over the Northern Hemisphere extra tropics above 30 N. Bias, root-mean-square (RMS) error and error standard deviation (STD) of observations minus background (o-b) and observations minus analysis (o-a), in ms^{-1}. The statistics include all vertical levels

(o-b)/(o-a) (ms^{-1})		BIAS	RMS	STD
AIREP-U	CTL	−0.54/−0.14	3.9/3.0	3.9/3.0
	CYC	−0.47/−0.15	3.9/3.0	3.8/3.0
RAOB-U	CTL	0.028/0.011	3.5/2.6	3.5/2.6
	CYC	0.026/0.004	3.6/2.6	3.6/2.6

DWL impact is. Continuous and homogeneous sampling of the atmosphere in the days prior to an extreme event thus appears effective in reducing the forecast error.

4. Christmas 1999 storm 'Martin'

In this section we zoom in on the second Christmas storm 'Martin' that made landfall in Brittany (Fr) on 27 December 18 UTC and causing much damage over France and Germany in the following 24 h. The verifying analysis from the rerun (CONTROL) experiment using the 2004 ECMWF operational model, see Fig. 8a, is very close to the 1999 operational analysis (not shown) with large wind speeds up to 10 Bft in the Gulf of Biscay. The 48-h control forecast from 25 December 18 UTC fails to predict the storm, see Fig 8b similar to the 1999 operational model (not shown). The mean sea level pressure (MSLP) RMS forecast error over France, (N/S/W/E) = (50/42/−20/10), relative to the verifying analysis in Fig. 8a is 9.6 hPa. The forecast initiated with the 25 December 18 UTC pseudo-truth from the cycled SOSE experiment represents the best achievable forecast in a cycled SOSE experiment. This forecast is displayed in Fig. 8c and does predict a storm, although the low is not deep enough and its location is slightly wrong. The MSLP forecast error is 4.6 hPa. Figure 8d shows that the additional 60 h of DWL observations over the period 23 December 06 UTC until 26 December 12 UTC from a tandem-Aeolus scenario improve the forecast substantially. The remaining MSLP forecast error is 6.5 hPa. In other words the additional DWL observations have reduced the forecast error by 62% relative to the maximum reduction (from the pseudo-truth forecast) that can be achieved by additional observations in SOSE. In addition, the position error metric, that is defined as the distance of the position of the center of the low relative to the position of the low in the pseudo-truth forecast, shows a reduction by 63% from 433 km (CONTROL) to 158 km (DWL). The intensity error has been reduced by 36%, from 994.99 hPa to 987.80 hPa relative to 974.87 hPa. Although these results clearly demonstrate the added value of a DWL, the deterministic DWL forecast in Fig. 8d still does not predict a severe storm. However, a deterministic forecast gives no complete answer on the probability of occurrence of a severe storm. This may be achieved by an ensemble experiment as discussed in the next section.

4.1. *'Martin' ensemble experiment*

Nowadays there is little doubt about the benefit one gains in weather forecasting by running ensembles. Many NWP centres have now developed approaches to search for growing structures in the initial state or shortcomings in the model formulation and to use this knowledge in designing ensembles. See Buizza et al. (2005) for a recent comparison of the performance of various operational ensemble approaches.

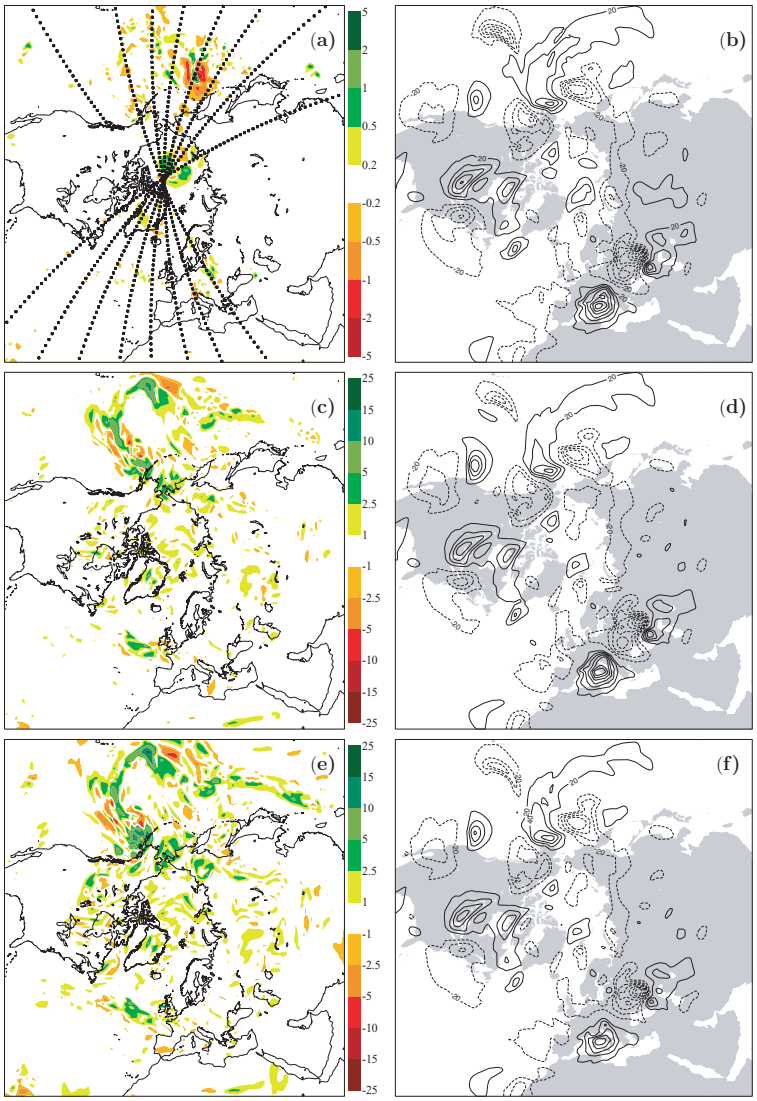

Fig. 7. DWL impact, eq. (1), on the wind analysis (ms^{-1}) of 26 December 12 UTC (left-hand column) and 500 hPa geopotential height (Z500) 2-d forecast error (m) at verification time 28 December 12 UTC (right-hand column) for (a, b) EXP1, (c, d) EXP2 and (e, f) EXP3 (see text). The dots in (a) denote the DWL observation locations in the single-cycle experiment. Note the different colour scales in panels (a) and (c, e). The Z500 contour interval is 20 m. The results are quantified in Tables 1 and 4.

Table 4. DWL tandem-Aeolus (T_2A) scenario impact, eq. (1), on the 26 December 12 UTC 500 hPa wind analysis for the three DWL experiments. The global regions (N/S/W/E) are defined through: NHem (90/30/−180/180), Eur (75/35/−12.5/42.5), NAtl (75/20/−75/−5), and NAme (60/25/−120/−75), NPac (75/20/140/−120), NAsia (80/40/20/180), NPole (90/75/−180/180)

Δ_{an} (ms^{-1})	NHem	Eur	NAtl	NAme	NPac	NAsia	NPole
T_2A; EXP1	−0.01	−0.03	0.00	0.00	−0.03	−0.05	0.11
T_2A; EXP2	0.36	−0.01	0.17	0.14	0.68	0.22	0.53
T_2A; EXP3	0.54	0.03	0.25	0.13	1.02	0.37	0.72

In this study we have employed a recent version of the ECMWF Ensemble Prediction System. Using the initial singular vectors of 25 December 18 UTC and the evolved singular vectors of 23 December 18 UTC, both at T42L60 resolution (∼500 km horizontal resolution and 60 vertical levels), we have applied the so-called Gaussian sampling technique of Ehrendorfer and Beck (2003), which was recently introduced as method to create initial time perturbations for the EPS. Here, we sample from 25 singular vectors to produce 50 initial state perturbations. The ensemble comprising 50 perturbed members and the unperturbed control forecast was run with resolution T319L60 (∼60 km horizontal resolution). Since our main interest is to study the impact of different initial conditions, we have not applied the stochastic physics perturbations during the model run.

Three ensemble experiments were conducted, based on analyses from the control (no DWL), the DWL and the cycled SOSE pseudo-truth. The latter is used for reference to indicate the maximum achievable result for an experiment with additional observations to the GOS. A storm is identified if either the maximum wind speed exceeds 10 Bft (24.5 ms^{-1}) or the minimum mean sea level pressure is less than 980 hPa over the verification

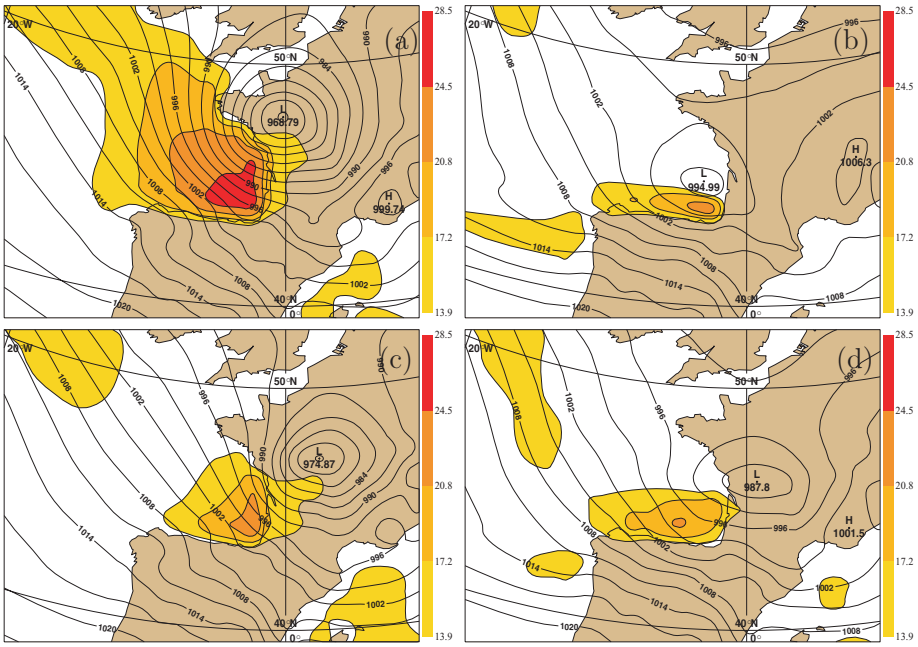

Fig. 8. Christmas 1999 storm 'Martin'. ECMWF surface pressure (hPa) and wind speed (ms^{-1}) at verification time 27 December 1999 18UTC for (a) the control verifying analysis, (b) the control 48-h forecast, (c) the pseudo-truth 48-h forecast and (d) the DWL (tandem-Aeolus scenario) 48-h forecast. Solid lines denote the mean sea level pressure with 3 hPa contour interval. Shaded areas denote the surface wind speed, corresponding to the range 7–10 Bft.

Table 5. 50-Member Ensemble Prediction System (EPS) experiment. Number of members (out of 50) including a storm in the verification area over part of the Atlantic, Gulf of Biscay and France (N/S/W/E) = (55/40/−20/10). Forecast initial time is 25 December 1999, 18 UTC. The second/third column shows the number of storm members (see text for the definition of a storm) for the 48 and 54-h forecast verifying at 27 Dec. 18 UTC and 28 Dec. 00 UTC, respectively. Analyses for the ensemble experiment are extract from (i) the control experiment, (ii) the DWL EXP3 experiment based on 60 h of HLOS wind observations according to a tandem-Aeolus scenario in addition to the GOS and (iii) the cycled SOSE pseudo-truth

# Storm members	48-h FC	54-h-FC
CONTROL	5	5
DWL-T_2A; EXP3	11	15
pseudo-truth; cycling	18	38

area which includes part of the Atlantic, the Gulf of Biscay and the French mainland. The wind speed criterion is extracted from the public warning system used by most meteorological offices. Table 5 shows that for the control experiment 5 out of 50 members include a storm (verification time 27 December 18 UTC), while for the DWL experiment the number of storm members increases to 11, that is, a more than doubling of the probability (from 10 to 22%) of forecasting a severe storm in case of additional DWL observations. For the 54-h forecast, verifying at 28 December 00 UTC, that is, when the storm has moved further into France, the impact of additional DWL is even more clear with a 30% probability of a severe storm against 10% for the control (no DWL) experiment.

Another measure to indicate the severity of a storm is by wind gust that is defined by the World Meteorological Organization (WMO) as the maximum of the wind averaged over 3 second intervals. Figure 9a shows the verifying analysis at 28 December 1999 00 UTC and corresponding 54-h forecasts from the unperturbed analyses (ensemble control forecast) of the three ensemble experiments. The verifying analysis shows wind gusts up to 12 Bft over the French coast and the Gulf of Biscay and up to 10 Bft in the Mediterranean. These large wind speeds are absent in the control forecast of the control (no DWL) experiment in see Fig. 9b, and moreover the low is not deep enough and out of position. Also the control forecast of the DWL experiment in Fig. 9d shows no extreme wind gusts, but the position of the low pressure system is much closer to the reference forecast from the pseudo-truth in Fig. 9c. In addition some members of the ensemble do forecast a severe storm with wind gusts up to 12 Bft, see Fig. 10 for a typical example.

The probability maps in Fig. 11 show the probability of exceeding a wind gust threshold for the three ensemble

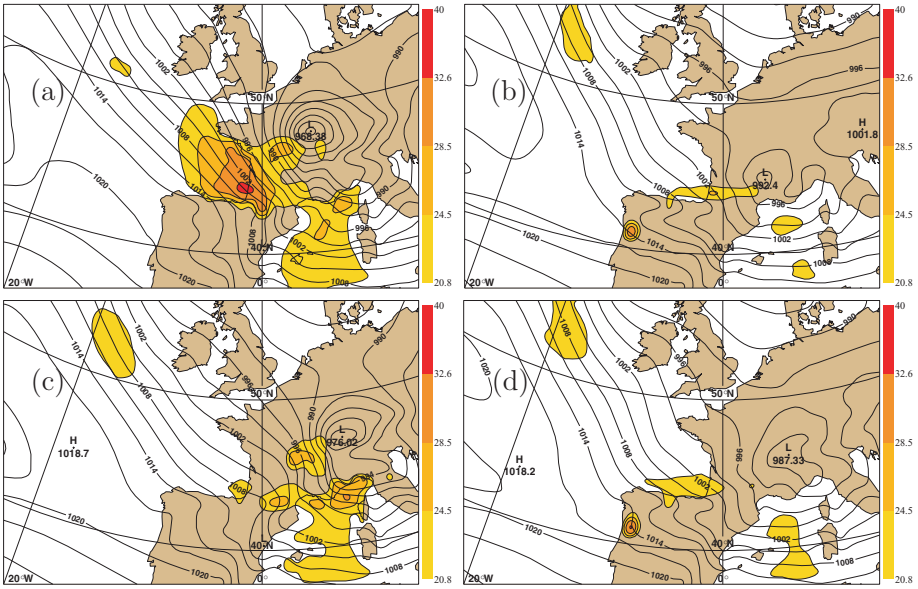

Fig. 9. Mean sea level pressure (hPa) and surface wind gust (ms^{-1}) for (a) the ECMWF verifying analysis at 28 December 1999 00 UTC, (b) the 54-h EPS control forecast for the control (no DWL) experiment, (c) the 54-h EPS control forecast for the cycled SOSE experiment and (d) the 54-h EPS control forecast for the DWL experiment. Solid lines denote the mean sea level pressure with 3 hPa contour interval. Shaded areas denote the surface wind gust, corresponding to the range 9–12 Bft.

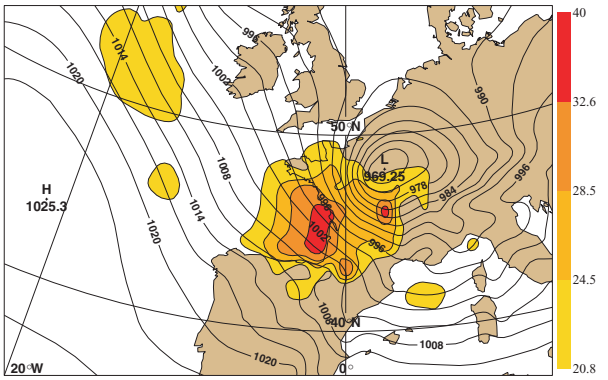

Fig. 10. Typical storm member from the DWL ensemble showing wind gusts up to 12 Bft in the 54-h forecast verifying at 28 December 00UTC. Solid lines denote the mean sea level pressure with 3 hPa contour interval. Shaded areas denote the surface wind gust (ms^{-1}), corresponding to the range 9–12 Bft.

experiments. These maps show a larger probability for the occurrence of excessive winds in the DWL experiment than in the control (no DWL) experiment in addition to an improved location of the forecasted severe weather. These results indicate that the GOS would have benefited from additional spaceborne DWL observations in the Christmas 1999 period to better forecast the damaging storm 'Martin' over Western Europe.

5. Summary, conclusions and discussion

A cycled SOSE is run over three subsequent days and the effects over multiple data assimilation cycles are compared to the effects in a SOSE over a single assimilation cycle with a 6-h time window. To this end, the SOSE method is extended in this paper to enable NWP impact assessment of prospective observations over a prolonged period of several days. It has been demonstrated that in a cycled SOSE a substantial part of the pseudo-truth analysis adaptations are maintained by the model, not rejected by existing observations (including all available satellite systems) and propagated progressively in subsequent assimilation cycles. This is another indication of the realism of the spatial structures of these adaptations. Unrealistic structures would be destroyed by either the model (Caron et al., 2007) or observations in the analysis (Isaksen et al., 2005). A cycled SOSE experiment with unrealistic structures would therefore not result in a stronger signal. However, the stronger signal at the end of the cycling period demonstrates that the generated structures constructively interfere from one cycle to the next while remaining compatible with all existing observations.

The resulting sequence of pseudo-true atmospheric states are used for the simulation of prospective extensions of the existing global observing and system (GOS). This opens the way for absolute impact assessment of future observing systems through SOSE.

The cycled SOSE has been applied to one case to give an indication of the added value of future spaceborne DWL

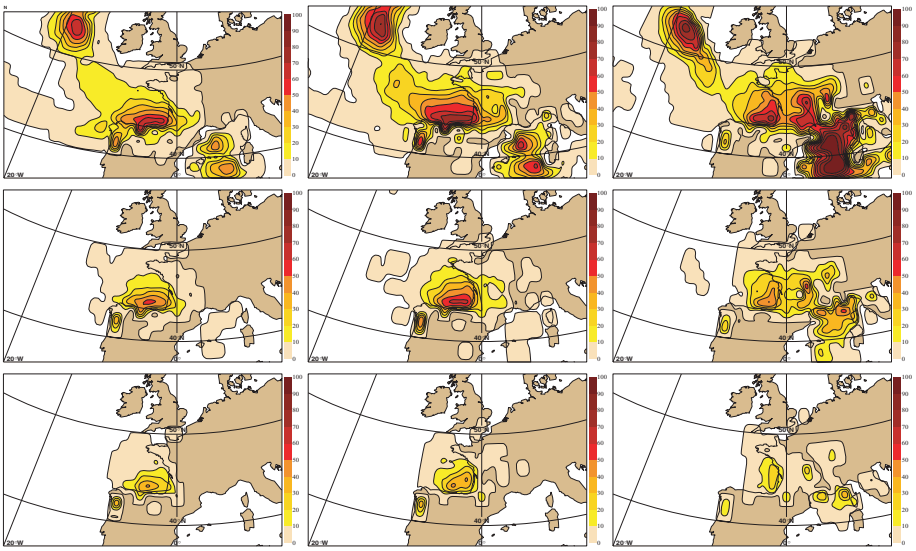

Fig. 11. Probability maps (%) for wind gust exceeding 9 Bft (top row), 10 Bft (middle row) and 11 Bft (bottom row) based on 54-h forecasts, verifying at 28 December 1999 00 UTC, from ensemble experiments using (i) no DWL observations (left-hand column), (ii) DWL observations (middle column) and (iii) the cycled SOSE pseudo-truth.

for NWP. The case investigated concerns the 1999 Christmas storm ‘Martin’ that caused much havoc in Western Europe. We apply the cycled SOSE to the Christmas 1999 period and show that additional observations from a spaceborne Doppler wind lidar over a 3-d period would have improved the 2-d forecast of the second Christmas storm ‘Martin’ substantially. This is further substantiated in a 50-member ensemble forecast run.

A cycled SOSE has two attractive properties as compared to the single-cycle implementation: (i) the resulting pseudo-true atmospheric state at the end of the cycling period, that is, at forecast initial time, provides a better forecast and (ii) the cycled implementation makes the SOSE method more suitable for absolute impact assessment of continuously operated observing systems such as from polar satellite platforms. Point (i) is of great interest since it indicates that the initial state may be able to explain a larger part of the forecast error (∼45% for ‘Martin’) than in single cycle sensitivity experiments (∼25% on average). Effects of model error, including the natural unpredictability (butterfly effect), thus appear less dominant, whereas about half of the forecast error remains unexplained still. It therefore provides more scope for an observing system to improve the analyses and their subsequent forecasts.

On the other hand the analysis modifications in the cycled SOSE are an order of magnitude larger than in a single-cycle SOSE, as these changes expectedly grow towards the size of the estimated analysis error due to the cycling over a few days. At the same time the spatial structure of the adaptations remains similar to the expected analysis error structure. As such, compared to the single-cycle ones, the cycled SOSE adaptations for the pseudo-truth appear realistic but are less effective per unit variance to reduce the forecast error, as is now more in line with real analysis errors.

Point (ii) concerns the fact that observing systems not only contribute to the quality of the analysis by their observations in the current cycle time window, but also by the observations performed in the days before, which information is propagated by the cycled data assimilation system. In a cycled SOSE both contributions of a new observing system are taken into account, whereas in a single-cycle SOSE only the instantaneous contribution is counted and not the longer term one. In a single-cycle SOSE the relative contribution of different prospective observing systems to NWP may be evaluated, but more absolute value will be obtained in a cycled SOSE. Point (i) results in larger NWP impact of any observing system, as does point (ii). On the other hand, from point (ii) we infer that a single observation has both instantaneous and long-term value and thus in a cycled SOSE the optimal sampling strategy may appear somewhat different. This consideration is also relevant for observation targeting strategies (THORPEX). We find for ‘Martin’ that a tandem Aeolus already constitutes ∼60% of the maximum achievable 2-d forecast impact in the cycled SOSE, whereas for the single-cycle SOSE this was only ∼8% of the (smaller) maximum achievable impact over all 38 cases run (MSBb).

The points above raise the issue of SOSE calibration both for absolute and relative NWP impact assessment. Calibration is an important aspect of, for example, OSSE (observing system simulation experiments) to verify the experimental setup and the realism of the results. Calibration is often associated to impact verification of simulated observations in the experiment with

the corresponding real observations in operational NWP. This is only part of the calibration. It is at least as important to verify the use of observations in the experiment with operations through statistics of observations minus background (o-b), observations minus analysis (o-a) and observation rejection. For the existing observations no substantial discrepancies were found in the data usage statistics of the CONTROL and SOSE experiments. Although applied to a single case, this result is a strong indication of a valid setup of the cycled SOSE experiment: the new observing system (DWL) is capable to resolve (part of) the additional variance brought into the system, through the analysis modifications that define the pseudo-truth, without disturbing the existing observations.

On the other hand, the basic back-predictability of the 2-d forecast errors to the SOSE initial state adaptations is low and there is no evidence that the sequence of pseudo-true atmospheric states in a cycled SOSE is any closer to the true atmospheric state than the sequence of control analyses (the truth is inherently unknown since unobserved). Compared to the single-cycle SOSE where the adaptations mainly replace dynamic analysis errors, as described in Marseille et al. (2007a), in a cycled SOSE, due to the cycling, also part of the stochastic analysis error is replaced, particularly in data sparse regions. For cycled SOSE, it is also expected that observation impact will not dramatically depend on the data assimilation system. It remains of interest to test the cycled SOSE in similar ways as proposed for the single-cycle SOSE, across data assimilation systems or in a weak constraint approach (Marseille et al., 2007a).

A quantitive assessment to project simulated impact to expected impact in operations further requires impact experiments, for example, to verify the impact of synthetic observations, from existing observing systems, in SOSE with the impact of the corresponding real observations in operational NWP. This could be done for example by excluding all wind observations from the analysis in a control experiment, construct the SOSE pseudo-truth over this control experiment and simulate all excluded wind observations by using this pseudo-truth. Subsequently, both a SOSE and an OSE (observing system experiment) should be run. Ideally the impact of the simulated observations in the SOSE should then be similar to the NWP impact of the real observations in the OSE. The impact should be assessed both on the analyses and the forecasts. Alternatively the Atlantic THORPEX Regional Campaign (A-TReC) period (Weissmann and Cardinali, 2007) may be used for calibration. During this measurement campaign in autumn 2003 additional observations from dropsondes and airborne DWL were obtained in targeted areas over the Atlantic near Iceland and Greenland. ECMWF demonstrated positive impact in an OSE (Weissmann and Cardinali, 2007). Although the period and target area are limited a SOSE over this period may provide an indication on calibration issues.

Fisher (2004) presented a long-loop 4D-Var method that in principle should construct improved analysed NWP model trajectories by considering both past and future observations in a long 4D-Var time window. Such method needs the development of a model tendency error penalty term in 4D-Var, but would be useful for generating a trajectory of pseudo-truth. However, the development of such long-loop 4D-Var and running it requires both excessive human and computational resources. Meanwhile the SOSE method may be further elaborated as an affordable alternative.

6. Acknowledgments

The authors thank the members of the ESA Aeolus Mission Advisory Group for stimulating discussions that have led to the work as described in this paper. Also, our ECMWF colleagues are acknowledged for their support and advice.

References

Buizza, R. and Hollingsworth, A. 2002. Storm prediction over Europe using the ECMWF Ensemble Prediction System. *Meteorol. Appl.* **9**, 289–305.

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

Caron, J. F., Yau, M. K., Laroche, S. and Zwack, P. 2007. The characteristics of key analysis errors. Part I: dynamical balance and comparison with observations. *Mon. Wea. Rev.* **135**, 249–266.

Ehrendorfer, M. and Beck, A. 2003. Singular vector-based multivariate normal sampling in ensemble prediction. *ECMWF Technical Memorandum* **416**.

Fisher, M. 2004. On the equivalence between the Kalman smoother and long-window, weak-constraint 4D-Var. In: *Proceedings of the Sixth Workshop on Adjoint Applications in Dynamic Meteorology, Maratea (It.), 2004*, 45–64.

Isaksen, L., Fisher, M., Andersson, E. and Barkmeijer, J. 2005. The structure and realism of sensitivity perturbations and their interpretation as 'Key Analysis Errors'. *Q. J. R. Meteorol. Soc.* **131**, 3053–3078.

Marseille, G. J., Stoffelen, A. and Barkmeijer, J. 2007a. Sensitivity Observing System Experiment (SOSE)–a new effective nwp-based tool in designing the Global Observing System. *Tellus* **60A**, doi: 10.1111/j.1600-0870.2007.00288.x

Marseille, G. J., Stoffelen, A. and Barkmeijer, J. 2007b. Impact assessment of prospective spaceborne Doppler wind lidar observation scenarios. *Tellus* **60A**, doi: 10.1111/j.1600-0870.2007.00289.x

Stoffelen, A., Pailleux, J., Källén, E., Vaughan, J. M., Isaksen, L., and co-authors. 2005. The Atmospheric dynamics mission for global wind measurement. *Bull. Am. Meteorol. Soc.* **86**, 73–87.

Uppala, S. M., Kållberg, P. W., Simmons, A. J., Andrae, U., da Costa Bechtold, V., and co-authors. 2005. The ERA-40 re-analysis. *Q. J. R. Meteorol. Soc.* **131**, 2961–3012.

Weissmann, M. and Cardinali, C. 2007. Impact of airborne Doppler lidar observations on ECMWF forecasts. *Q. J. R. Meteorol. Soc.* **133**, 107–116.