

Can we trust climate models to realistically represent severe European windstorms?

Tomasz M. Trzeciak¹ · Peter Knippertz² · Jennifer S. R. Pirret¹ · Keith D. Williams³

Received: 2 February 2015 / Accepted: 19 July 2015 / Published online: 20 August 2015
© The Author(s) 2015. This article is published with open access at Springerlink.com

Abstract Cyclonic windstorms are one of the most important natural hazards for Europe, but robust climate projections of the position and the strength of the North Atlantic storm track are not yet possible, bearing significant risks to European societies and the (re)insurance industry. Previous studies addressing the problem of climate model uncertainty through statistical comparisons of simulations of the current climate with (re-)analysis data show large disagreement between different climate models, different ensemble members of the same model and observed climatologies of intense cyclones. One weakness of such evaluations lies in the difficulty to separate influences of the climate model's basic state from the influence of fast processes on the development of the most intense storms, which could create compensating effects and therefore suggest higher reliability than there really is. This work aims to shed new light into this problem through a cost-effective “seamless” approach of hindcasting 20 historical severe storms with the two global climate models, ECHAM6 and GA4 configuration of the Met Office Unified Model, run in a numerical weather prediction mode using different lead

times, and horizontal and vertical resolutions. These runs are then compared to re-analysis data. The main conclusions from this work are: (a) objectively identified cyclone tracks are represented satisfactorily by most hindcasts; (b) sensitivity to vertical resolution is low; (c) cyclone depth is systematically under-predicted for a coarse resolution of T63 by both climate models; (d) no systematic bias is found for the higher resolution of T127 out to about three days, demonstrating that climate models are in fact able to represent the complex dynamics of explosively deepening cyclones well, if given the correct initial conditions; (e) an analysis using a recently developed diagnostic tool based on the surface pressure tendency equation points to too weak diabatic processes, mainly latent heating, as the main source for the under-prediction in the coarse-resolution runs. Finally, an interesting implication of these results is that the too low number of deep cyclones in many free-running climate simulations may therefore be related to an insufficient number of storm-prone initial conditions. This question will be addressed in future work.

Keywords Windstorms · Seamless approach · Climate models · Diabatic processes · Horizontal resolution

✉ Tomasz M. Trzeciak
t.m.trzeciak@gmail.com

Peter Knippertz
peter.knippertz@kit.edu

Jennifer S. R. Pirret
ee09jso@leeds.ac.uk

Keith D. Williams
keith.williams@metoffice.gov.uk

¹ University of Leeds, Leeds LS2 9JT, UK

² Karlsruhe Institute of Technology, 76128 Karlsruhe, Germany

³ Met Office, FitzRoy Road, Exeter EX1 3PB, UK

1 Introduction

Windstorms associated with intense wintertime cyclones from the North Atlantic Ocean are among the most frequent and most devastating natural hazards affecting Europe (Munich 2009; Schwierz et al. 2010). Several storms during the last decades, such as the Great October Storm (October 1987), Daria (January 1990), Lothar (December 1999), Jeannette (October 2002), Kyrill (January 2007) and Klaus (January 2009) to name just a few, caused fatalities

and billions of euros in economic losses. Changes in the track, frequency and intensity of such events in the decades to come are a potential threat to European societies (Leckebusch et al. 2007; Pinto et al. 2007a). Reliable projections of the changes in the related meteorological extremes are therefore of paramount importance.

Despite all recent progress in climate change research for variables like near-surface temperature, confidence in projections of storm tracks, cyclone intensity and extreme winds in the northern hemisphere remains low according to the 5th Assessment Report of the Intergovernmental Panel on Climate Change (IPCC; Christensen et al. 2013) and the UK Climate Projections (UKCP 2009). This low confidence is related to concerns over the skill of many models in realistically representing large-scale features such as the stratosphere (Scaife et al. 2012; Manzini et al. 2014) and ocean circulation (Woollings et al. 2012) as well as the dynamics of individual storm systems, for example due to insufficient horizontal resolution (Willison et al. 2013). Consequently, the consensus between different models on a climate change signal remains rather low (Harvey et al. 2012; Zappa et al. 2013b).

With respect to reliable projections of climate change impacts into the future, a reproduction of the climatology of extreme extratropical cyclones is generally regarded as an important requisite. Often, however, the deviations between model and observations for the current climate are so large that future projections rely solely on changes with respect to the model's own climate, e.g. with respect to percentiles in wind speed or cyclone occurrence frequencies. Particularly for intense events, the inter-model variability is considerable (Lambert and Fyfe 2006). Some models (or model versions) show a general under-representation of cyclones with core pressures below 970 hPa over the North Atlantic/Europe, i.e. the ones that are of greatest importance with respect to impacts (Knippertz et al. 2000; Pinto et al. 2006, 2007b, 2009a). This pertains not only to the frequency of such systems, but also to deepening rates and minimum core pressures, leading to a reduction in frequency as high as 50 % in subregions (Pinto et al. 2006). Other studies, however, report a good correspondence (Bengtsson et al. 2006) or even a higher number of intense systems in the model's control climate than in analysis data (Bengtsson et al. 2009). Such biases also occur in operational weather forecasts as documented by Froude (2009) for the ensemble prediction system (EPS) of the European Centre for Medium-Range Weather Forecasts (ECMWF) over the North Atlantic. A fundamental understanding of the exact reasons for such deviations is of great importance for assessing sources of uncertainty in climate projections.

Generally speaking, the three main sources of uncertainties in climate projections are related to the evolution of emissions, internal climate variability and errors in climate

models (Hawkins and Sutton 2009). While the first two can be addressed through consideration of scenarios and long-term climate ensembles, the last is a more fundamental problem, generally associated with the model's coarse horizontal and vertical resolutions or deficiencies in model physics and the dynamical core. Model errors also affect simulations of past atmospheric states used to create event sets for the insurance industry. The most common approach to assess model error is to conduct multi-model ensemble simulations for the recent climate, using the ensemble spread and deviations from observational datasets to provide a measure of uncertainty. This becomes rather problematic, if models have similar biases (e.g. Pennell and Reichler 2011) or if different errors compensate each other. For example, a general large-scale negative bias in surface pressure, related, say, to processes in the ocean or stratosphere, could mask a systematic underprediction of storm deepening due to an insufficient representation of the fast dynamical processes associated with cyclone intensification. A clear separation of the relative influences of different types of errors is difficult or impossible in the statistical evaluations of long-term climate simulations cited above and therefore alternative approaches are needed.

In the following we will summarise some recent work on model deficiencies related to cyclone activity. Model resolution is most commonly employed to explain differences between global models and (re-)analyses in the frequency of intense cyclones (Knippertz et al. 2000; Pinto et al. 2006; Leckebusch et al. 2008; Zappa et al. 2013a). Jung et al. (2006) distinguish dynamical effects, related to the supposedly worse representation of crucial physical processes and the reduced height of orographic barriers such as Greenland and the Alps in coarser-resolution models, and truncation effects to do with changing the chances of detecting small-scale, short-lived cyclones with automatic tracking routines. They find that the effect of dynamics, physics and orography dominates over the truncation effect for intense cyclones, whereas the truncation effect dominates for shallow cyclones. Willison et al. (2013) argue that coarser-resolution models are not capable to reproduce aspects such as the diabatic generation of potential vorticity along fronts, leading to feedbacks on the evolution of individual cyclones (see also Madonna et al. (2014) for effects of Warm Conveyor Belts) and of the entire storm track and its energetics. Pinto et al. (2007b, 2009a) compare intense winter cyclones over the North Atlantic in ECHAM5 and in NCEP re-analysis with the same horizontal resolution and find fewer and weaker systems in the model despite a similar mean jet speed. This raises the question to what degree increasing horizontal resolution alone can improve the representation of deep cyclones. In addition, there are even differences in storm intensity for different analysis products, which is again to some degree related to horizontal

resolution but also to the data assimilation system (Hodges et al. 2003; Löptien et al. 2008; Bengtsson et al. 2009). This illustrates the problem of which “truth” the control simulations of climate models should be compared to.

Several authors find systematic biases and large inter-model differences in number and intensity of North Atlantic cyclones with storm tracks tending to be either too zonal, displaced southward or not extending far enough into the European continent (Pinto et al. 2006; Greeves et al. 2007; Ulbrich et al. 2008; Zappa et al. 2013a). Greeves et al. (2007) state that this result depends on the dynamical core and the horizontal resolution. At lower resolution, a semi-Lagrangian core is less able to produce small-scale eddies, which results in weaker eddy kinetic energy, and in weaker and fewer cyclonic features, while an Eulerian core reacts to reduced resolution with changes in the storm track location with intensities less affected. Other potential sources of error include parameterisations of convection, surface energy fluxes, boundary layer turbulence etc. It is therefore unclear to what degree the lack (abundance) of intense cyclones in some models is mainly the product of too rare (frequent) occurrences of storm-prone circulation patterns or of problems spinning up intense storms due to deficiencies in model physics. Leckebusch et al. (2008) find indications that the four large-scale circulation clusters that are associated with most intense storms in observations are systematically under-represented in different model simulations of the current climate. The differences in basic states between different global climate models are so large that the driving model is often the dominant influence on simulations with regional climate models, limiting their usefulness to improve the reliability of climate projections (Leckebusch et al. 2006; Donat et al. 2011).

A rather technical aspect, which nonetheless adds to the difficulty in bringing in line results from various studies, are differences in the objective methods used to identify winter storms. Common approaches use mean sea-level pressure (MSLP), vorticity or wind measures at different levels, the usage of which can in fact affect resulting trends, again with a certain influence of the available horizontal resolution (Pinto et al. 2005; Ulbrich et al. 2009). The use of near-surface wind speed is particularly difficult due to the strong influence of the model’s representation of the boundary layer (Leckebusch et al. 2006; Rockel and Woth 2007). However, results for the most intense storm systems are more robust than for total numbers (Neu et al. 2013).

In recent years many operational and research centres around the world have recognised that the scientific and technical challenges of short-term numerical weather prediction (NWP), seasonal forecasting and climate projections are in many ways interrelated. Systematic errors in climate models often resemble those in NWP models (e.g. Williams and Brooks 2008). Froude (2009) for example

found a larger error in the predicted amplitudes of intense storms than of weaker storms in the operational ECMWF EPS system consistent with many climate models (Lambert and Fyfe 2006). This general perception led to the concept of “seamless prediction”, which is based on the idea of using only one modelling system in different configurations to predict the state of the atmosphere over a large range of spatial and temporal scales (e.g. Rodwell and Palmer 2007). Possible applications of this idea include a calibration of climate projections using the reliability of probabilistic seasonal forecasts (Palmer et al. 2008). A prime example of such a seamless strategy is the Unified Model (UM) of the UK Met Office that comprises model versions used for high-resolution short-term NWP as well as climate and Earth system simulations.

To the best of our knowledge, this study is the first to apply the novel idea of a seamless weather–climate prediction to the specific question of evaluating model simulations of severe storms affecting Europe. The main objective of this work is to test whether state-of-the-art climate models such as ECHAM6 and the UM are capable of realistically reproducing the dynamical processes involved in the usually rapid deepening of historical European windstorms as compared to re-analyses. The influence of different horizontal and vertical resolutions on the evolution of the cyclones’ core pressure will be tested for different lead times using an objective tracking method. Physical reasons for differences in cyclone deepening will be assessed using a recently developed diagnostic tool based on the surface pressure tendency equation (PTE). The results allow for the first time a clear separation into fast and slow model errors to guide future model development. The paper is structured as follows: Sect. 2 provides details on the methodology including information on the selected storms, observational data, the employed models and numerical experiments as well as the tracking and PTE methods. Results of the simulations and their interpretation are given in Sect. 3 followed by discussion and conclusions in Sect. 4.

2 Methodology

The following subsection will explain the method of storm selection and provide some background information on the selected storms. Section 2.2 gives a description of the two climate models employed together with a discussion of the conducted hindcast simulations. The two last subsections then provide information on the cyclone tracking (Sect. 2.3) and the pressure tendency diagnostic used to analyse physical reasons for differences in model performance (Sect. 2.4). A key dataset used for various aspects throughout this paper is the 6-hourly ERA-Interim reanalysis (Dee et al. 2011).

2.1 Storm selection

For this study, 20 winter (October–March) storms from the period 1990–2010 were selected based on their potential wind damage. The selection was made using the Storm Severity Index (SSI) developed by Leckebusch et al. (2008), which is based on the cube of the wind speed above the local 98th percentile of the wind climatology. Other meteorological indices are available, such as that proposed by Lamb and Frydendahl (1991). The Lamb Index uses the greatest observed wind speed cubed, multiplied by the area and duration of the storm; a similar index was used in the recent work of Roberts et al. (2014). Both indices endeavour to use meteorological or physical properties of the storm as a proxy for the economic damage a storm inflicts. The Lamb Index was not employed in the current work because in the SSI, use of the 98th percentile allows for adaptation; an area with a higher 98th percentile of the wind climatology will likely be more prepared for strong winds and so less damage will be inflicted. Using SSI means that more Mediterranean cyclones appear in a list of severe wind events than if the Lamb Index were used (Roberts et al. 2014), but this was easily overcome in the current work by manually excluding such systems.

Using ERA-Interim data, the SSI was calculated for each grid point, summed over the region 40–60°N and 10°W–20°E, which includes central and northern Europe and parts of the adjacent Atlantic Ocean and Mediterranean Sea, and ranked by the total value. Out of the top 75 values found, 29 were within 24 h of an even higher value associated with the same meteorological system and were therefore excluded from further investigation. 26 values were associated with weather phenomena other than Atlantic cyclones and also excluded: Mediterranean cyclones (13), polar lows (3), large pressure gradients at the fringe of strong high-pressure systems (9) and orographic effects (1). These were identified through a subjective analysis of weather charts and horizontal distributions of SSI (not shown). The 20 storms selected are listed in Table 1.

The resulting list features many famous storms that have caused numerous fatalities and injuries as well as substantial damage and insurance losses. It includes the destructive storm series of January–March 1990 (Daria, Vivian, Wiebke) and December 1999 (Anatol, Lothar, Martin). The synoptic evolution, dynamics and impacts of many of these storms have been analysed in detail by various authors: e.g. Vivian (Goyette et al. 2001), Lothar (Wernli et al. 2002; Rivière et al. 2010), Kyrill (Fink et al. 2009), Klaus (Liberato et al. 2011; Rivière et al. 2014) and Xynthia (Rivière et al. 2012; Ludwig et al. 2014). They belong to a class of storms that is essential for climate models to reproduce satisfactorily in order to allow reliable estimates of future storm risk.

Table 1 Details of the selected storms

Name	Date and time	Min P (hPa)	Duration (days)
Agnes	1993-01-24 00:00	966.7	4.25
Anatol	1999-12-03 18:00	956.1	4.50
Daria	1990-01-25 18:00	949.1	6.50
Elke	2000-12-08 06:00	972.4	7.25
Emma	2008-02-29 18:00	959.7	8.00
Gero	2005-01-12 00:00	947.8	6.75
Hanno	2007-01-14 00:00	970.4	6.00
Jeanette	2002-10-27 12:00	975.1	7.00
Jennifer	2002-01-27 12:00	953.3	7.50
Klaus	2009-01-24 00:00	966.0	2.50
Kyrill	2007-01-19 06:00	961.4	7.00
Lara	1999-02-05 00:00	949.3	2.00
Lothar	1999-12-26 12:00	976.0	5.50
Martin	1999-12-27 18:00	968.1	3.75
Rebekka	2000-11-06 06:00	965.8	10.00
Silke	1998-12-26 12:00	949.2	6.50
Udine	1991-01-05 06:00	948.0	6.25
Vivian	1990-02-27 12:00	941.0	6.50
Wiebke	1990-03-01 06:00	971.5	8.00
Xynthia	2010-02-28 00:00	968.3	8.50

Included are their names in alphabetical order as given by the Free University of Berlin (for more information, see <http://www.met.fu-berlin.de/adopt-a-vortex/>), the time and value of their minimum pressure as well as the track length in days, all as identified from objective tracking (see Sect. 2.3)

2.2 Climate models and their configuration

The experimental set-up used in this study builds on the Transpose-AMIP experimental methodology, in which climate models are run in NWP mode (Phillips et al. 2004; Boyle et al. 2005). The details of the methodology for the latest Transpose-AMIP II experiment are described in Williams et al. (2013). In short, the models are configured as for the AMIP experiment, but only a few days long hind-cast runs are performed instead of a long, multi-year climate run. The model state variables are initiated from re-analysis data giving a well defined initial state and a reference data to evaluate the model performance against. Non-atmospheric state variables that spin up relatively slowly, e.g. land surface and aerosols, are initiated from a climatology derived from a model's AMIP run or through the nudging method of Boyle et al. (2005). Variables that spin up quickly, e.g. cloud fraction and precipitation, are initiated from zero or again using the nudging method of Boyle et al. (2005). For the present study the spin-up of the land surface, which can take several years (Yang et al. 1995), is not so crucial, since the storm development takes places prevalently over the Atlantic Ocean, and therefore

the option of climatological initialisation was chosen. Aerosols are expected not to constitute a major driver in the storm development and thus aerosol concentrations are also initiated from climatological values, while fast changing variables such as cloud fraction and precipitation that typically spin up within several hours of the forecast are initiated from zero. For the initialisation of the atmospheric state (temperature, water vapour, divergence and vorticity or wind velocities, surface pressure) ERA-Interim re-analysis data from the ECMWF are used. Sea surface temperature and ice cover are also taken from ERA-Interim and are kept fixed for the duration of the hindcast. It should be noted that the native spatial resolution of ERA-Interim is T255 (ca. 80 km), such that the hindcasts may to some degree benefit from the higher resolution influence on the initial conditions, which would not be the case in longer climate simulations.

Two models are used here: ECHAM6 (Max Planck Institute for Meteorology (MPI), Germany), which is the version used for the Coupled Model Intercomparison Project Phase 5 (CMIP5) round of experiments, and UM-GA4 (Met Office, UK; Walters et al. 2014). A larger selection of models would be desirable for better inter-comparison, however technical challenges of setting up and running the models made this choice impractical. The models were run for the selected windstorm events listed in Table 1 at lead times (taken with respect to the time of the cyclone core pressure minimum) that ranged from 18 to 96 h. ECHAM6 was configured with T63 spectral truncation, which corresponds to about 1.9° or 200 km horizontal resolution and 47 levels for the vertical hybrid coordinate with pressure at the top of the atmosphere taken as $p_{\text{top}} = 0$. Additionally, a two times higher horizontal and vertical resolution of T127 (about 0.94° or 100 km grid spacing) and 95 levels, respectively, was used as well. UM was run at N96 horizontal resolution and 85 vertical levels with the lid at 85 km above the mean sea level. The N96 configuration of UM has the same number of grid points (grid spacing) in the longitudinal direction as T63 of ECHAM6, but one and a half times more in the latitudinal direction, thus this resolution is effectively in between T63 and T127 of ECHAM6. In the vertical both models are stratosphere resolving.

2.3 Cyclone tracking method

Objective identification and tracking of extratropical cyclones is a challenging problem. There is no widely accepted scientific definition of what an extratropical cyclone is and as a result various practical methods have been proposed to date, typically relying on tracking extrema in a relevant meteorological variable such as MSLP or low-level vorticity. A number of existing methods have recently been inter-compared in the Intercomparison

of mid latitude storm diagnostics (IMILAST) project (Neu et al. 2013).

Identification and tracking of the severe windstorms is all the more challenging due to their often fast propagation and greater tendency to develop secondary extrema (Hanley and Caballero 2011). This prompted us to develop our own methodology with the main goals being: applicability to data at native resolutions, and thus avoidance of potential bias from input preprocessing such as truncation to lower resolution, and stability of output tracks for (not excessively large) changes of the algorithm's tuning parameters.

In our algorithm, cyclones are identified similarly to Murray and Simmonds (1991) and Pinto et al. (2005) in that they are taken as the minimum of MSLP or, if not found, the minimum of its gradient in the vicinity of a vorticity maximum (within 5° latitude search radius). Each vorticity maximum is associated with at most one such minimum in our method. The minima and maxima are located on a 0.25° grid, which is higher than native resolution of the input data and to which all variables are first interpolated using cubic splines. The use of the pressure gradient criterion is necessary, since not all midlatitude cyclonic systems are associated with a closed contour pressure depression, i.e. a proper minimum in MSLP, and inclusion of open pressure depressions is required.¹ The original method of Murray and Simmonds (1991) uses geostrophic vorticity calculated as the Laplacian of MSLP and smoothed out within a prescribed radius. While this is also a possibility in our method, using 850 hPa vorticity truncated at T63 resolution proved to work more reliably, in particular over mountainous regions, and to have a more uniform range of values in the meridional direction, thus making the choice of the threshold value for vorticity less critical (set to $v_{\text{th}} = 10^{-5} \text{ s}^{-1}$ for this study). The main role of using fixed resolution for the vorticity (the pressure variable is taken at native resolution) is to restrict the number and location of admitted open depression systems. In that regard we found T63 resolution to be close to optimal: a lower resolution of T42, which is for example used in the method of Hodges (1994), tended to eliminate some valid pressure minima from the consideration, while higher resolution would admit too many spurious open depression systems and orography originated pressure minima. The procedure of cyclone identification detailed above combined with track pathway restriction and optimisation described later in this section worked sufficiently reliably that no special treatment of elevated terrain was necessary, such as

¹ As an alternative to pressure gradient we also tried using filtered pressure variable with the coarsest scales (corresponding to the first 5 or so wavenumbers) being removed, but this tended to shift the position of minima, especially for cyclones embedded within large scale pressure gradients, and thus we abandoned this approach.

orography filtering (Murray and Simmonds 1991; Pinto et al. 2005) or variables on terrain-following coordinates (Grise et al. 2013) found in other, similar methods.

Next, the identified cyclone centres are connected across subsequent time steps to form tracks, using the following multi-step procedure. First, for each two successive time steps pairs of the cyclone centres are formed if the cross-correlation coefficient S_{corr} of MSLP for the 10° longitude by 5° latitude boxes around cyclone centres is non-negative, and if the distance between the centres (taken as the great circle arc angle) is less than the threshold d_{th} calculated as:

$$d_{\text{th}} = \frac{d_1}{1 + \left(\frac{d_1}{d_2} - 1\right) \frac{v_{\text{th}}}{\min(v_i, v_{i+1})}} \quad (1)$$

where $d_1 = 11^\circ$ and $d_2 = 3^\circ$ are the upper and lower limits for the distance threshold, respectively, and v_i denotes the vorticity value at the identified cyclone centre. The effect of this formula is to have a more restrictive threshold for weaker cyclones than for stronger ones, which are more likely to move more rapidly. The cross-correlation coefficient, calculated as the covariance of two variables divided by the product of their individual standard deviations, provides a measure of similarity in pressure field around identified cyclone centres between two consecutive time steps, with larger values giving higher likelihood that a given pair of cyclone centres belong to the same cyclonic system propagating in space and time.

Next, the pairs formed in the previous step are joined on the shared points to form three-point track fragments (hereafter simply called triplets) under a smoothness constraint $S_{\text{smooth}} < 1$, with the smoothness score S_{smooth} calculated as:

$$S_{\text{smooth}} = \frac{|\mathbf{P}_i - \mathbf{P}_m|}{\max(|\mathbf{P}_{i-1} - \mathbf{P}_{i+1}|/2, d_{\text{st}})} + \frac{|p_{i-1} - 2p_i + p_{i+1}|}{p_{\text{sc}}} \quad (2)$$

where in the first term $|\mathbf{P}_i - \mathbf{P}_m|$ denotes the great circle arc distance between the points \mathbf{P}_i and \mathbf{P}_m , \mathbf{P}_m is the mid point on the great circle arc between the triplet end points \mathbf{P}_{i-1} and \mathbf{P}_{i+1} , which is expected to be close to the point \mathbf{P}_i for uniformly moving cyclones, and $d_{\text{st}} = 3^\circ$ is a cut-off distance at which smoothness is taken in absolute rather than relative terms (this avoids over-restricting near stationary cyclones that may wander randomly within the cut-off distance and exceed the relative smoothness criterion). Hodges (1995) proposed a different smoothness criterion by generalising the approach of Salari and Sethi (1990) to motion on the sphere. Their criterion is a weighted sum of two terms separately penalising changes in speed and direction, which provides additional tuning flexibility. However, we have not found this to be required in practice, while their approach does not properly cater for slow moving cyclones. The second term in our smoothness criterion

gives a restriction for smoothness of pressure p_i along the three consecutive points normalised by a scaling constant of $p_{\text{sc}} = 8 \text{ kPa}$.

In addition to the smoothness criterion, we retain at most n_{br} best (lowest score) triplets that share the same middle point. In this way we limit the number of possible branching choices (i.e., the number of different pathways tracking through a given point) to the most likely ones. The initial tracks are then formed by linking overlapping triplets, the smoothest ones first, and after that the tracks are optimised by swapping parts to obtain the lowest total score as explained later in this section. With $n_{\text{br}} = 1$ the initial tracks are also the final ones, since no branching is allowed (each cyclonic feature point is assigned at most one triplet passing through it) and tracks are either extended with an overlapping triplet or terminated. This is not always satisfactory, in particular if one is interested in forming the longest possible tracks, since extratropical cyclones often develop multiple pressure minima and vorticity maxima, which may cause premature track termination by taking the branch that is locally the smoothest one, but ultimately not propagating further. Hanley and Caballero (2011) studied this problem of multiple centres and found that they are much more prevalent in intense storms, which are the kind of storms that also have the highest practical importance and are the focus of this study. Through subjective visual analysis, we found that the majority of benefits from allowing track branching is realised already for $n_{\text{br}} = 2$, with only minor improvements for some of the tracks for $n_{\text{br}} = 3$, which is the value used for this study. Larger values of n_{br} were not beneficial due to increased risk of including some unrelated or misidentified centres. Considering only a limited number of locally smoothest connections tracking through a given point helps to make the right choice between extending or terminating the track, which makes the method more robust and less sensitive to the cyclone identification procedure.

For the track optimisation step, we first implemented the method of Salari and Sethi (1990), which was also used by Hodges (1994). The method works by exchanging a pair of points between tracks that provides the biggest improvement to some target measure of goodness, such as better overall smoothness of tracks. This point exchange progresses along the tracks from start to finish and then is repeated from the beginning until no more possible exchanges are possible. The disadvantage of this method is that it is sensitive to the initial set of tracks being used.² Salari and Sethi (1990) addressed this issue by running the

² Optimizing path connectivity in some global sense is computationally intractable for larger problem sizes due to combinatorial explosion of the number of possible paths. Therefore, most methods work by iteratively refining some initial guess in the local sense and thus do not guarantee finding the global optimum, making them dependent on the quality of the initial guess.

algorithm alternately in forward and backward direction until convergence. Unfortunately, with this modification the algorithm is no longer guaranteed to terminate (Veenman et al. 1998) and in our testing this lack of convergence occurred frequently enough that we decided to take a different approach. Rather than exchanging individual points between tracks, we split them at a given time step into two parts and exchange these parts instead, choosing the exchange that offers maximal improvement to the objective function being optimised. This has the advantage that the calculated improvement is independent from the direction in which the exchanges proceed, thus the final, converged set of tracks will be invariant to that direction. For the objective function to minimise we tested using the total sum of the smoothness scores S_{smooth} and of cross-correlation scores S_{corr} (see above) taken along all tracks. Both of these choices performed well, but their combination (through arithmetic mean) performed better still in our subjective visual assessment (although not markedly so), thus this was the final choice used for this study.

The method described above proved to perform very well and it reliably produced very smooth tracks with hardly any visible artefacts. We tested it also using 12-hourly data, which only required more relaxed distance parameters $d_1 = 18^\circ$ and $d_2 = 5^\circ$ (due to longer time step) to produce tracks of comparable quality to the 6-hourly data. One aspect remained problematic for our intended purpose, however. For some cyclones with multiple centres (e.g. Kyrill) the method produces two shorter tracks rather than a longer, single one. This is not a mistake per se and, e.g. Fink et al. (2009), Fink et al. (2012) also treats the case of storm Kyrill as two cyclones, but this makes inter-comparison of individual tracks more cumbersome. For this reason we modified the step of exchanging the track parts by calculating the gain in the objective function due to exchange with an added, small bias term that favours formation of longer tracks, even if somewhat less optimal with respect to the chosen objective function. The bias term was calculated as an increase in the difference between the tracks' lengths after the exchange (taken as the number of time steps) times some small constant (0.02 in this study). We should note here that in rare cases this modification may prevent the track optimisation step from converging, especially if the scaling constant in the bias term is too large. However, in practice this has not proved to be a problem in our experience.

As a matter of practical convenience we also added the possibility of running our tracking method with a reference track. This does not change the tracks being produced for the given input data, but only the sorting order of the output, with the tracks best matching the reference being enumerated first. The distance between the tracks is calculated as the sum of distances between corresponding points

within the overlap time window of the two tracks or the maximum distance of $d_1 = 11^\circ$ outside of this window.

2.4 Pressure tendency equation diagnostic

An interesting question in the context of this study is whether differences in model performance for the different configurations tested here can be assigned to certain physical processes involved in cyclone deepening. As the simulations are started from analysis data, we assume that the large-scale environment is sufficiently constrained and that differences are predominantly related to dynamical processes in the immediate vicinity of the centre of the deepening cyclone. Generally speaking the main contributors to the deepening of cyclones are baroclinic conversions (transport of warm air upwards and polewards, and transport of cold air equatorwards and downwards) and diabatic processes (latent heating, radiation, surface fluxes).

Recently, Fink et al. (2012) proposed a new tool to automatically diagnose different contributions on the basis of the surface pressure tendency equation (PTE). In this approach the PTE is re-formulated from the classical mass-based version using virtual temperature as the main variable (see Fink et al. 2012 for details). Essentially, the tendency of surface pressure (hereafter DP) then equals to the vertical integral of the time change in virtual temperature from the surface to an upper boundary (ITT). Column warming (cooling) is associated with surface pressure fall (rise). In order to close the equation correction terms for mass loss (gain) through precipitation (evaporation; EP) and for geopotential tendencies at the upper boundary ($D\phi$) need to be taken into account, which are usually small compared to ITT (see discussion in Knippertz et al. 2009). Finally the ITT term can be split into contributions from horizontal temperature advection ($TADV$), vertical motion (VTM) and diabatic heating ($DIAB$). The method then uses an objective tracking algorithm (see Sect. 2.3) that identifies the position of a given cyclone every 6 h. For each time step a 3 by 3 degree box reaching from the surface to 100 hPa is centred on the cyclone position and the PTE is evaluated for this box using ERA-Interim or climate model data for the 6 h preceding the arrival of the cyclone centre. Diabatic processes are thereby calculated as a residual as in Fink et al. (2012). The method therefore identifies the processes that take place downstream of the cyclone centre to create the surface pressure fall during the approach of the system in a mixed Lagrangian–Eulerian framework. Fink et al. (2012) already applied this method to some of the storms under consideration here showing that (a) ITT clearly dominates all other terms during the main deepening phase, (b) $TADV$ and $DIAB$ both contribute to deepening while VMT causes pressure rise through adiabatic cooling, (c) relative contributions from $DIAB$

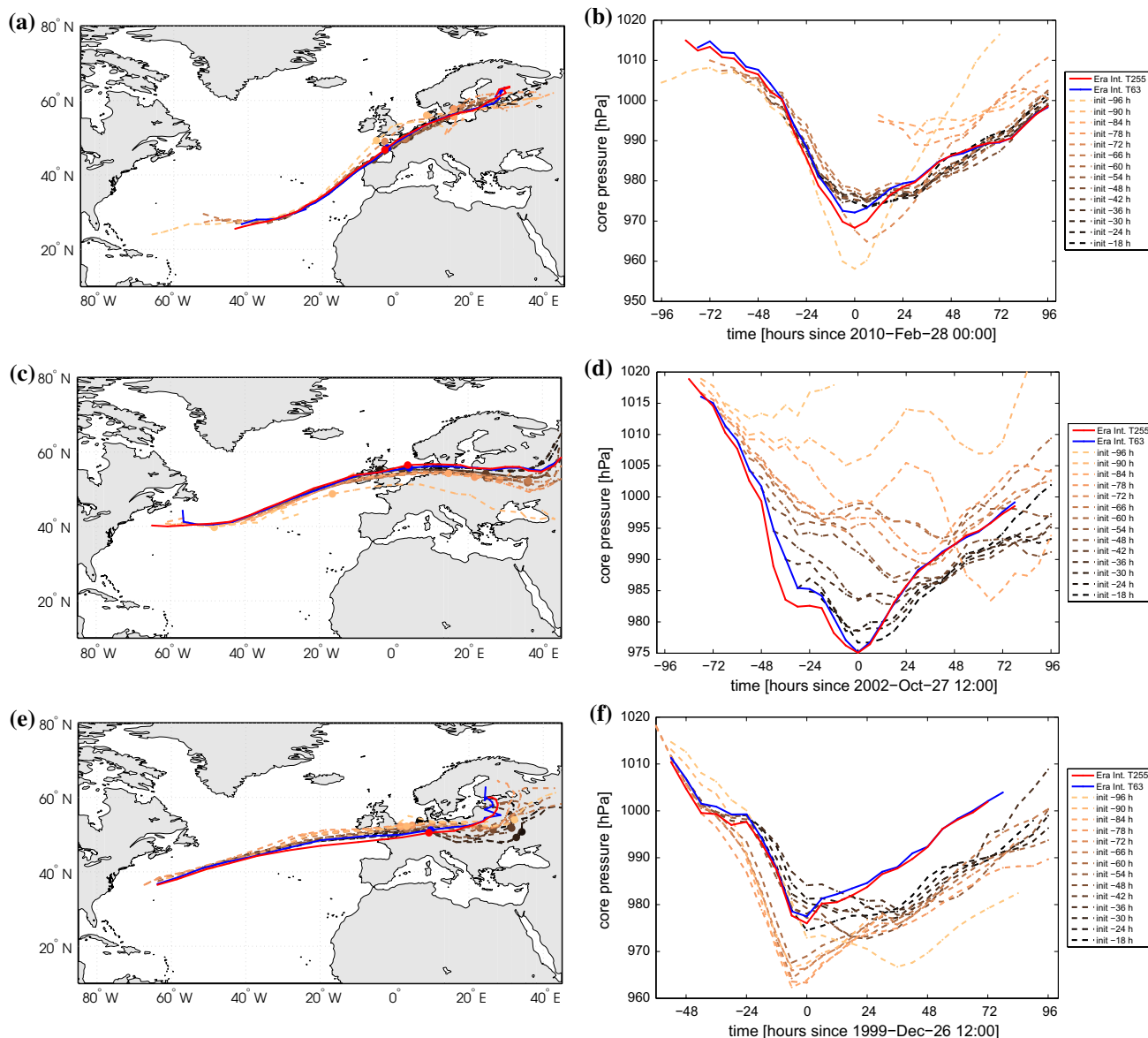


Fig. 1 Examples of storm development in ECHAM6 T63L47 simulations: Xynthia (*top*), Jeanette (*middle*) and Lothar (*bottom*). Left panels show tracks, while right panels show the time evolution of core pressure relative to the time of deepest pressure in the analysis data (00 UTC 28 February 2010 for Xynthia, 12 UTC 27 October

2002 for Jeanette, 12 UTC 26 December 1999 for Lothar). *Red (blue) lines* represent reference tracks in ERA-Interim data in T255 (T63) horizontal resolution. The *dashed lines* represent 14 hindcasts with lead times from 18 (*beige*) to 96 h (*black*) relative to the time of deepest pressure in the analysis data

vary strongly from storm to storm with Xynthia, Lothar and Klaus showing largest values.

3 Results

3.1 Results for ECHAM6 at T63 horizontal resolution

In this subsection we will briefly discuss some of the simulation results using the ECHAM6 model at T63L47 resolution. A spectral resolution of T63, which corresponds to

about 1.9° or 200 km grid-spacing, is still a fairly typical resolution for climate or Earth system models and was used for ECHAM6 in the most recent Assessment Report of the Intergovernmental Panel on Climate Change (Flato et al. 2013).

Figure 1 shows the tracks and core pressure evolution of three example cyclones (Xynthia in February–March 2010, Jeanette in October 2002 and Lothar in December 1999) as tracked with the objective algorithm described in Sect. 2.3. Results from 14 hindcast simulations started every 6 h are shown together with tracks based on the

ERA-Interim reanalysis data. The hindcasts were started from 96 h (4 days) to 18 h before the minimum core pressure in the reanalysis. The reference tracks are determined from data in the original resolution of T255 (about 0.47° or 50 km grid spacing) and from the same data coarse-grained to T63 in order to match the climate model data. This was done to demonstrate the truncation effect discussed by Jung et al. (2006). For Xynthia and Jeanette, the coarse-graining has very little effect on the track (Fig. 1a, c), while Lothar, which was a relatively small storm embedded within large scale pressure gradient, is shifted northward in the T63 data (Fig. 1e). As expected, truncation effects on core pressure are more marked. For Xynthia, there is a slight systematic offset throughout the deepening phase, reaching about 3 hPa at the time of minimum pressure, with better agreement at later stages (Fig. 1b), possibly due to a larger size of the mature system. Jeanette shows similar differences during the deepening phase, but then good agreement from the minimum onwards (Fig. 1d). For Lothar, truncation errors are somewhat smaller and occur at different stages of the development (Fig. 1f).

Given the coarse resolution of the climate model, the hindcasts reproduce the tracks of all three cyclones surprisingly well. Xynthia is characterised by an extremely large meridional displacement from the subtropical Atlantic to Finland or even northwestern Russia (Fig. 1a). ECHAM6 reproduces this unusual track well apart from a slight northward excursion into England and the North Sea in some of the earlier runs. It is also remarkable that some of the tracks in the hindcasts are even longer than in the reanalysis (both earlier start and later decay). The core pressure evolution shown in Fig. 1b also matches well, with the majority of simulations following the reanalysis within just a few hPa. Notable exceptions are two of the earlier runs with core pressures even deeper than observed and three runs that fail to spin up Xynthia altogether, only generating a much weaker low during the late stages of development. Such a behaviour is to be expected in some cases given the large sensitivity to initial conditions in highly baroclinic environments.

Results for Jeanette show a very different picture. Agreement with respect to track is very good for the large majority of runs (Fig. 1c). One very early run shows a significant southward shift already over the central North Atlantic, while some others show slight southward shifts in the very final stages only. With regard to core pressure, the ECHAM6 hindcasts show a systematic and large underestimation of the development (Fig. 1d). While the earliest run ends after a few days, the following simulations show a much later intensification of Jeanette after an intermediate period of weakening. After that there is a systematic improvement with lead time starting from a positive bias of about 20 hPa around 3-days lead time to a satisfactory

reproduction for the -18 h run, started when the cyclone was already fully developed. Many of the later runs show a delay in the filling of Jeanette. This behaviour suggests that important physical processes in the development of the storm are not well reproduced by the coarse resolution model. The work by Willison et al. (2013) suggests that the lack of diabatic contributions along insufficiently resolved frontal rain bands is a likely cause. This may be particularly important for Jeanette, as it occurred in October when baroclinicity is usually still relatively low, but sea surface temperatures are high. This question will be further addressed in Sect. 3.4.

Finally, Lothar shows a tendency of hindcast tracks to be shifted northwards in agreement with the coarse-grained reanalysis track (Fig. 1e). There is also some disagreement in the late stages, when the reanalysis tracks curve northwards towards Finland, while several hindcast tracks continue on into Russia. The core pressure evolution shows a clear clustering into two stereotypical behaviours (Fig. 1f). Most later runs reproduce the deepening phase very well, but then show a clear delay in the weakening with an offset of about a day. The earlier runs show an even steeper deepening and reach a core pressure on the order of 965 hPa, about 10 hPa deeper than the reanalysis, followed by a delayed filling similar to the later runs. This is quite a surprising outcome given the already extremely destructive nature of the real Lothar. This behaviour is, however, consistent with the ECMWF operational ensemble prediction at the time, which contained several members much deeper than the analysed storm (see Fig. 7.26 in Wilks 2011). One important caveat in this context is that Lothar was an explosively developing, but very small storm, which is not very well captured in reanalysis data, particularly of course the part over the Atlantic, where observations are sparse (T. Hewson, ECMWF, personal communication, 2014).

In summary, these results show that even at a coarse resolution of T63, climate models are capable of realistically simulating the tracks of severe cyclones, when given the analysed initial conditions. Such a behaviour was found for most of the 20 storms selected for this study (not shown). The evolution of the core pressure shows large case-to-case differences as already documented in Fig. 1. In the following subsection we will investigate the influence of horizontal and vertical resolutions.

3.2 Sensitivity to horizontal resolution

As explained in Sect. 2.2, the runs with a resolution of T63 discussed in Sect. 3.1 were repeated consistently for a finer resolution of T127 (about 0.94° or 100 km grid spacing). This model configuration was available only together with increased vertical resolution of 95 levels. A comparison between results with T63L95 and T63L47 configuration,

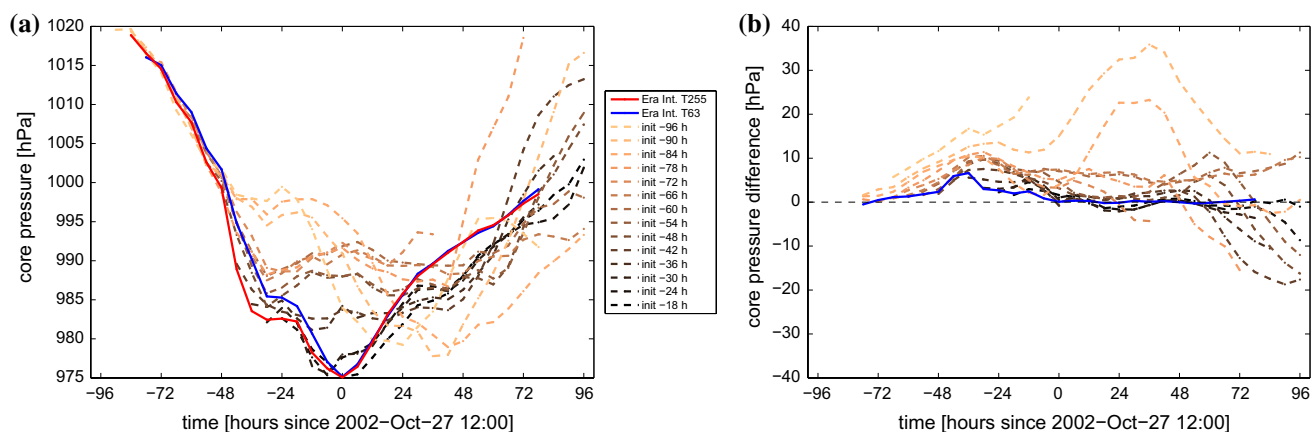


Fig. 2 Influence of resolution on the simulation of storm Jeanette. **a** Time evolution of core pressure relative to the time of deepest pressure in the analysis data (12 UTC 27 October 2002) as in Fig. 1d but for ECHAM6 in T127L95 resolution. **b** Differences between the

T63L47 and T127L95 simulations (Fig. 1d minus Fig. 2a). *Blue solid line* shows the truncation error for coarse-grained Era-Interim (T63 minus T255)

however, did not show any appreciable differences (not shown), strongly suggesting that the increase in vertical resolution has only a minor impact on the results. As an example of the impact of increased horizontal resolution, Fig. 2a shows the core pressure evolution for Jeanette, which should be compared to the corresponding plot in Fig. 1d. To facilitate this comparison, Fig. 2b shows the track-to-track difference in pressure for each hindcast using the same colour coding. It is clear from these plots that particularly the early stages of the development are much improved. After that, the earlier runs still show a delay of the intensification as in Fig. 1d, but at a deeper and therefore more realistic pressure level. Later runs, started after the cyclone has already been initiated, show only a mild improvement. This suggests that, in this case at least, the representation of key physical processes during the early stages of development are particularly sensitive to model resolution. These results also demonstrate that the dynamical effect (magnitude of differences in Fig. 2b) is much larger in this case than the truncation effect (differences between the high- and low-resolution analysis) of a few hPa.

Figure 3 summarises the influence of horizontal resolution statistically for all 20 selected cyclones. As a common frame of reference the time of deepest core pressure found in the reanalysis (T255) is used, similar to Figs. 1 and 2. Differences between ECHAM6 hindcasts and ERA-Interim reanalysis (T255) are depicted as box-and-whisker diagrams (see caption for a detailed explanation). The top row of Fig. 3 shows results for core pressure for the two horizontal resolutions T63 and T127, already exemplarily compared in Fig. 2. The values compared here are the minimum core pressure in the reanalysis and the minimum core pressure of the cyclone simulated by the climate model with a certain lead time relative to the deepest

pressures in the reanalysis, given on the x-axis. Of course the cyclone in the hindcast can be shifted in space and may have its minimum of core pressure at a time different from the reanalysis. These differences will be discussed in the following panels c–f. Also, not all hindcasts actually produce a track comparable to the reanalysis (see discussion of track matching in Sect. 2.3). Therefore the box-and-whisker diagrams in Fig. 3 do not contain all 20 members, particularly not those for long lead times (numbers in brackets).

As suggested by the case example of Jeanette discussed in Sect. 3.1, Fig. 3a shows a significant systematic underrepresentation of cyclone strength measured in terms of core pressure in the T63 simulations. The median bias is about 3 hPa for 18 h lead time (typical order of magnitude of the truncation error) and then gradually increases to more than 10 hPa for 66–78 h, after which it slightly drops again. For these longest lead times the range of biases is on the order of 50 hPa, making a meaningful comparison between the two datasets somewhat questionable. The 25th percentile is negative only for 18 h lead time, demonstrating a positive bias for the large majority of cases, although the whiskers show that for all lead times there are runs exhibiting deeper developments than analysed (as in the example of Lothar in Fig. 1f). The corresponding analysis for T127 shows a dramatically different picture (Fig. 3b). The median bias is very close to zero up to 48 h lead time with an almost symmetric distributions. Results are still surprisingly good for the 54–66 h range. After that, biases become more clearly positive and the range increases, but results are still much better than for the coarser resolution (Fig. 3a).

Panels c and d in Fig. 3 show the corresponding analysis for the time of the deepest core pressure. For T63, the

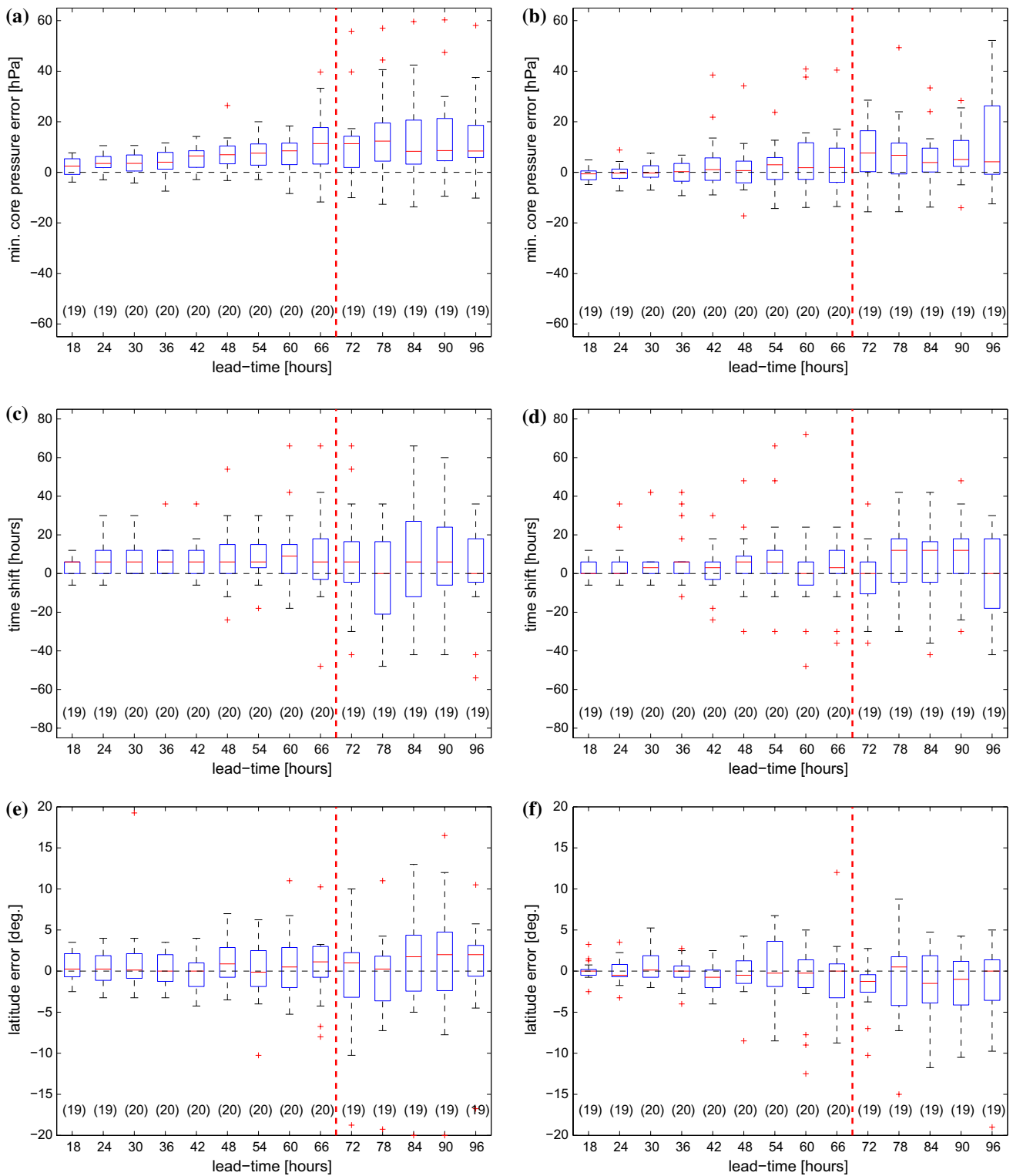


Fig. 3 Statistical analysis of the influence of horizontal resolution on all 20 storms. Differences between ECHAM6 simulations and ERA-Interim (T255) for the time of the deepest pressure in the analysis data are shown as *box-and-whisker plots* (red line median, blue boxes interquartile range, *whiskers* extrema, *pluses* outliers) against hindcast lead time in hours. The number of storms matched and used for

analysis is given in parentheses along the x-axis. Shown resolutions are T63L47 (*left panels*) and T127T95 (*right panels*) for the three parameters core pressure (hPa; *top*), time (hours; *middle*) and latitude (degrees; *bottom*). The *dashed red lines* mark 3 days lead time, when hindcast typically start to degrade

median bias is almost constantly 6 h, meaning that the cyclones are one tracking time step too slow on average, with few occurrences of 0 and 12 h (Fig. 3c). The interquartile range comprises mostly positive values or at least shows positive skew. As for pressure (Fig. 3a), the range increases strongly at 78 h, making interpreting the long lead times somewhat questionable with time differences of the order of 2 days as in some of the runs for Jeanette shown in Fig. 1d. The corresponding values for T127 (Fig. 3d) are slightly better out to 72 h lead time with more medians falling on the zero line. There is also a general tendency to more positive values as in T63. While the interquartile range and whiskers tend to be smaller than for T63, a few large outliers occur, the cause of which is unclear. Overall, these results show a somewhat delayed development of the severe cyclones in general (as illustrated in Figs. 1d and 2a for Jeanette), broadly consistent with the findings of Froude et al. (2007) of too slow propagation of cyclones in operational forecasts. One interesting implication of this is that on average the model cyclones spend more time over water, where frictional convergence is reduced, which should offset the tendency for weaker storms at lower resolution to some extent. This means that the underestimation of core pressure due to dynamical processes alone may be even larger than shown in Fig. 3a. A quantitative analysis of this effect, however, is beyond the scope of this paper.

The bottom panels of Fig. 3 show differences in latitudinal position between ECHAM6 hindcasts and ERA-Interim, again for T63 (Fig. 3e) and T127 (Fig. 3f). Biases are generally small (few degrees) and mostly symmetric around the zero line for both resolutions up to about 3 days lead time. There is a very slight tendency for the T63 cyclones to be on a northward shifted track (as for Lothar in Fig. 1e), while T127 tracks are a little further south. Overall the differences between the two resolutions are not very large, with a slightly smaller spread for the higher resolution, as already found for time (Fig. 3d). The latitudinal position bias is moderately correlated in negative direction with the core pressure bias (i.e., southward/northward shift correlates with weaker/stronger cyclone intensity), with similar correlation coefficients of -0.53 and -0.52 for T63 and T127, respectively.

In summary, these results clearly demonstrate that even a still relative coarse resolution of T127 is sufficient to reproduce the most severe cyclones affecting Europe, if the model is provided with realistic initial conditions, while T63 clearly misses out on some essential physical processes involved in the deepening, not so much in the propagation of the cyclones. This result is in general agreement with the work by Jung et al. (2006) and Willison et al. (2013) as discussed in Sect. 1.

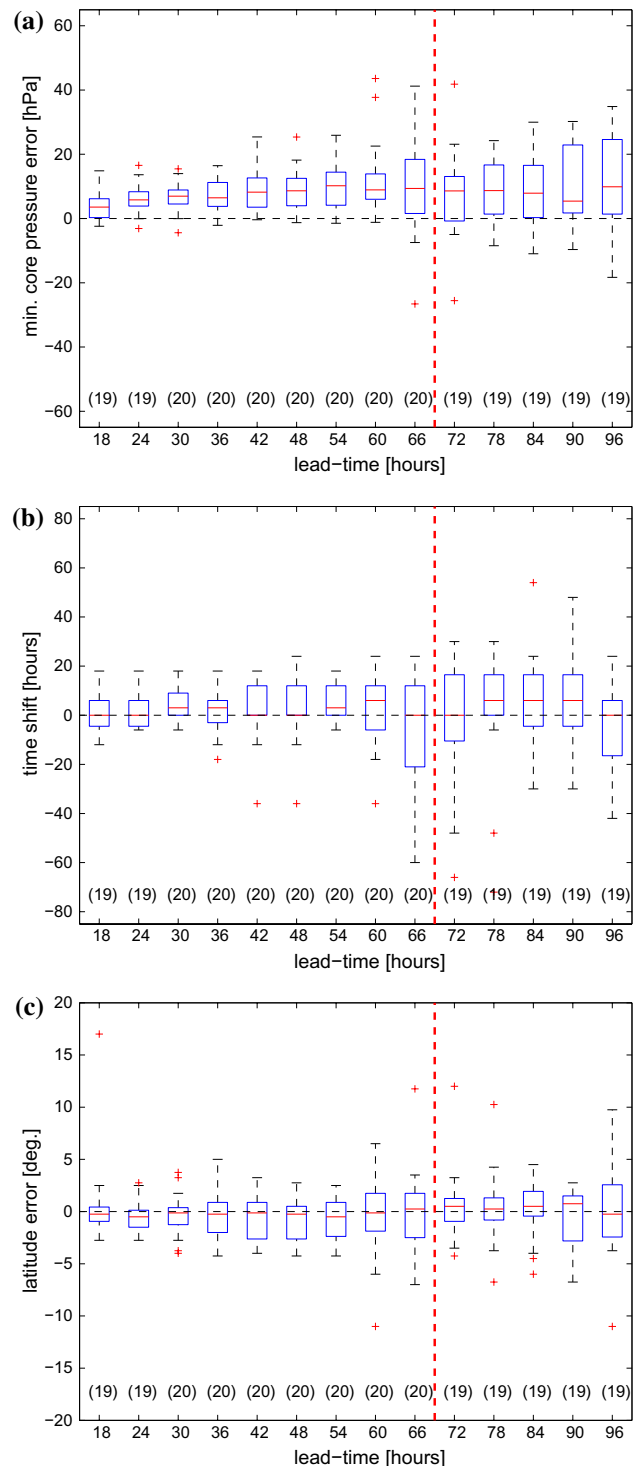


Fig. 4 As in Fig. 3a, c and e but for UM N96L85

3.3 Comparison between ECHAM6 and UM

An important question is how model dependent the results in the previous section are. To test this, identical hindcasts

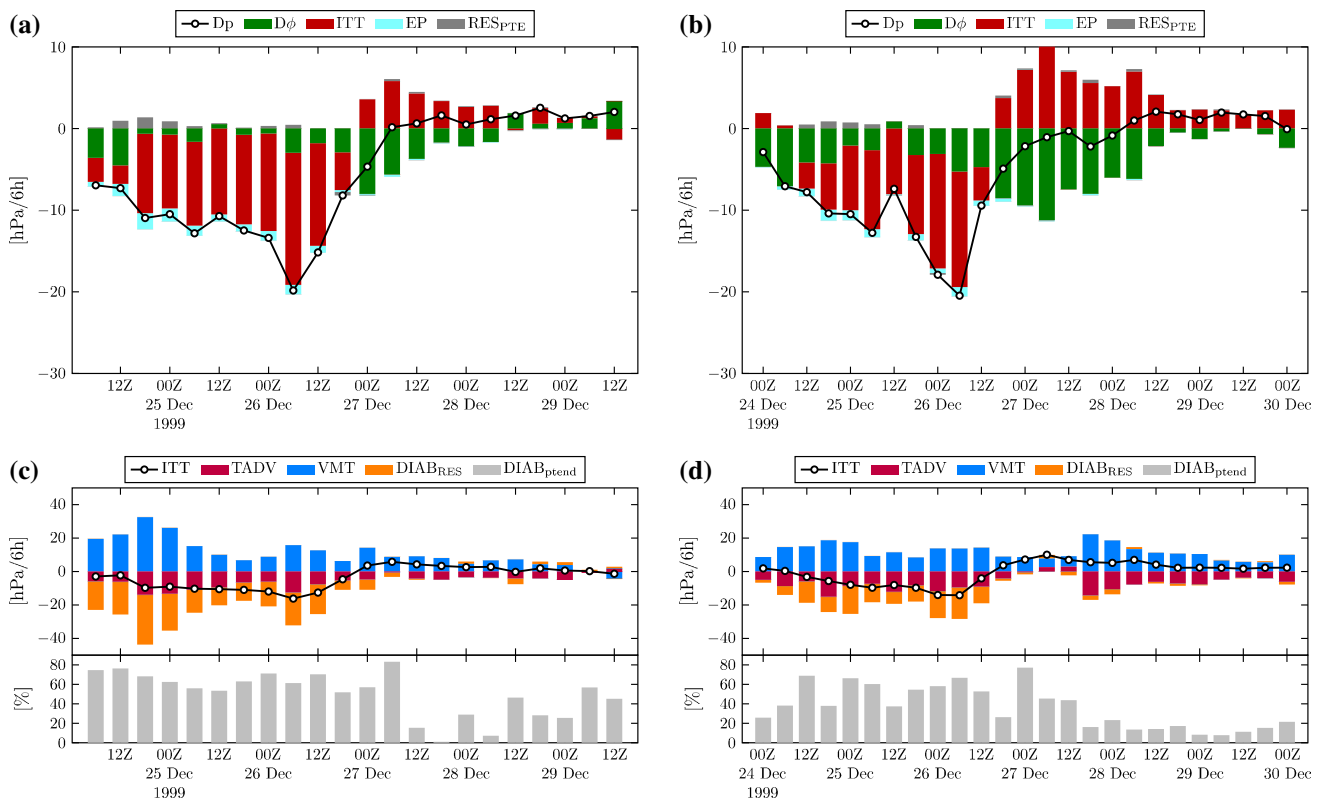


Fig. 5 PTE analysis for Lothar. *Left panels* show the time evolution of PTE terms from 06 UTC 24 December to 12 UTC 29 December 1999 for ERA-Interim reanalysis and *right panels* the corresponding results for the ECHAM6 T63L45 hindcast started at 12 UTC 23 December 1999 (corresponds to 72h in Fig. 1, bottom row). *Top row* shows the overall budget, while the *bottom row* shows the split

of the total column warming (*ITT*) into horizontal advection (*TADV*), diabatic heating (*DIAB_{RES}*) and effects of vertical motion (*VMT*). For a detailed explanation of the meanings of the different terms, see Sect. 2.4. The *dashed red lines* mark 3 days lead time, when hindcast typically start to degrade

for the 20 cyclones were generated using the UM in a horizontal resolution of N96 with 85 levels and statistics analogous to Fig. 3 were produced (Fig. 4). The three panels of Fig. 4 can be directly compared with Fig. 3a, c and e. Despite the different model architecture, physics and somewhat higher horizontal resolution, many parallels are found. Similar to ECHAM6 (Fig. 3a), the UM (Fig. 4a) shows a significant positive bias in core pressure,³ which is even stronger during the first day and then increases only slowly after that, giving overall larger median and interquartile values out to 60 h lead time, after which the differences are less clear with a slightly larger spread in ECHAM6. The results for timing of deepest pressure are also positively skewed but slightly less so

than in ECHAM6 (cf. Figs. 3c, 4b). Again the spread in these values is slightly smaller for the UM. The match with latitudinal position is also slightly better for the UM with medians closer to zero, more symmetric distributions and a smaller spread. The correlation of latitudinal position with core pressure bias is -0.45 , thus again negative but slightly weaker than for ECHAM6. Overall these results suggest that the general problem of coarse-resolution climate models to realistically reproduce key processes of cyclone deepening is not model-dependent, but inherent in insufficiently resolved dynamical features such as fronts.

3.4 Pressure tendency equation analysis

An important question remaining unanswered from the previous analysis is that of the physical reasons for the underestimation of cyclone deepening in the T63 (and also UM N96) simulations. As already outlined in Sect. 2.4, this question will be addressed here with the aid of the PTE diagnostic proposed by Fink et al. (2012). Figure 5 shows

³ The negative bias in storm strength in the UM has been considerably improved in GA6 compared to GA4 through the introduction of ENDGame dynamical core, which improves the accuracy of the departure point calculation in the semi-Lagrangian dynamical core (C. Sanchez, Met Office, personal communication, 2014).

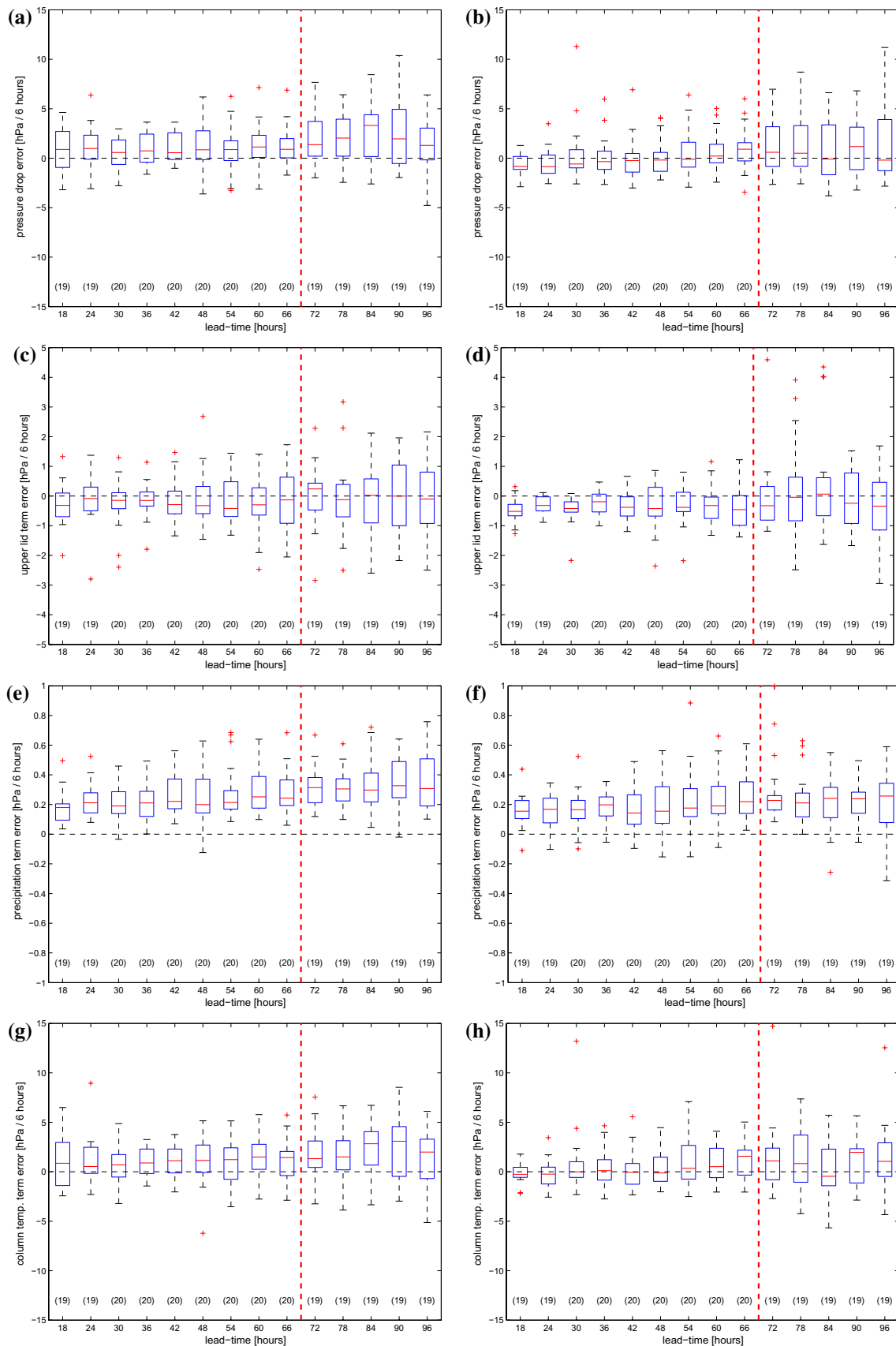
one detailed example of such an analysis for the storm Lothar (already discussed in Fig. 1). Left panels show results based on ERA-Interim (T255), left panels those from a selected ECHAM6 hindcast (the one started at 12 UTC 23 December 1999, i.e. 72 h before the deepest core pressure). The overall budget for ERA-Interim shows the intense and prolonged deepening and mature phases with negative Dp values until 00 UTC 27 December (Fig. 5a). As expected and already discussed in Fink et al. (2012), this phase is dominated by intense warming of the atmospheric column (up to 100 hPa) as shown by the negative ITT terms. There are small contributions to the pressure fall from mass removal through precipitation (EP) and geopotential tendencies at the 100 hPa level ($D\phi$). It is interesting to note that there is also a small residual (RES_{PTE}) during times with large EP contributions, pointing to some problems with precipitation in the reanalysis data. The splitting of the ITT term in different components (Fig. 5c) shows that the deepening phase is characterised by very large diabatic contributions ($DIAB_{RES}$) that make up on the order of 70 % of all negative terms (i.e. the sum of $TADV$ and $DIAB_{RES}$). This quantity is termed $DIAB_{ptend}$ and shown by grey bars in the bottom panels in Fig. 5. Large parts of this are compensated by cooling through vertical motions (VMT), but overall an almost constant warming (ITT) is observed from 18 UTC 24 December to 12 UTC 26 December. The large contribution from diabatic heating is consistent with the idea of a diabatic Rossby wave, discussed specifically for Lothar in Wernli et al. (2002). The main point there is that latent heating creates a potential vorticity anomaly in the lower troposphere that later interacts with an upper-level jet stream to create an explosive baroclinic development. Finally, the late stages are characterised by unusual column cooling and stronger movements of the upper integration boundary (Fig. 5a). The former is mostly related effects of vertical motion not being compensated by diabatic heating or horizontal temperature advection (Fig. 5c). This was a period when the storm was almost stationary over Russia and slowly decaying, making the PTE analysis somewhat less meaningful (see also Fink et al. 2012).

Figure 5b, d show the corresponding analysis for ECHAM6. The broad patterns of evolution are similar to ERA-Interim, but a number of interesting details differ quite significantly. The most striking difference is the much larger contributions of geopotential tendencies at the upper integration boundary (Fig. 5b), particularly during the phase immediately after the deepest core pressure (i.e. 18 UTC 26 December to 06 UTC 28 December). These changes are mostly compensated by smaller negative ITT terms in early stages and larger positive ITT terms in later stages, leading to an overall similar evolution of the pressure tendency term Dp , which however stays negative for longer in the ECHAM6 simulation. Contributions from the

Fig. 6 Statistical analysis of the influence of horizontal resolution on the main PTE budget for all 20 storms. The depiction is analogous to Fig. 3 with results for T63L47 in the left panels and T127T95 in the right panels. Displayed parameters are pressure tendency (Dp , a, b), tendency at the upper integration boundary ($D\phi$, c, d), precipitation/evaporation term (EP , e, f) and column temperature tendency term (ITT , g, h). For a detailed explanation of the meanings of the different terms, see Sect. 2.4. The dashed red lines mark 3 days lead time, when hindcast typically start to degrade

EP term are comparable and the residual RES_{PTE} is also similar. As a consequence of the changes to the ITT term, diabatic contributions ($DIAB_{RES}$) and those from vertical motions (VMT) are both reduced in magnitude during early stages, while $TADV$ is less affected (Fig. 5d). This implies a lesser relative role of diabatic processes as shown by smaller $DIAB_{ptend}$ values. At later stages, there are much stronger contributions from vertical motions (VMT) in ECHAM6, although also some stronger compensation from $TADV$. This comparison demonstrates that even in the case of a similar track and core pressure evolution, different physical processes can dominate the deepening. The strong reduction of diabatic processes in the coarse-resolution model is consistent with the idea of insufficiently resolving frontal precipitation zones (e.g. Willison et al. 2013). One may have expected that such a deficit would lead to an overall weaker cyclone development as demonstrated in many case studies, and consistent with the results shown in Figs. 2, 3 and 4. The fact that in this case the lack of diabatic contributions is compensated by larger $D\phi$ values is remarkable and point to a stronger upper-level control of the cyclogenesis, which can be sufficiently resolved by the coarse-resolution model. This idea is consistent with the slight northward shift in the track shown in Fig. 1f. It almost appears as if the situation was so prone to the development of an intense storm that different physical pathways exist that can lead to similar cyclonic developments and that the choice of pathways is resolution-dependent.

An important question now is whether this behaviour is representative of the ensemble of 20 storms studied in this paper. In order to find this out, box-and-whisker plots for the PTE terms Dp , $D\phi$, EP and ITT (Fig. 6), and $TADV$, $DIAB_{RES}$ and VMT were created (Fig. 7), again for ECHAM6 runs with resolutions T63 (left) and T127 (right). Before the statistical analysis, the PTE terms are averaged over the deepening period (negative Dp) in the hindcasts and the corresponding time in the ERA-Interim. As expected from Figs. 3 and 4, the Dp term shows a consistent positive bias in the median for all lead times (Fig. 6a). Somewhat surprisingly this bias remains quite constant around 1 hPa per 6 h for lead times from 18 to 66 h and then increases abruptly accompanied by a larger spread. A marked change in median and spread around this lead time has already been observed for earlier plots. A possible explanation for this behaviour is that Dp is not



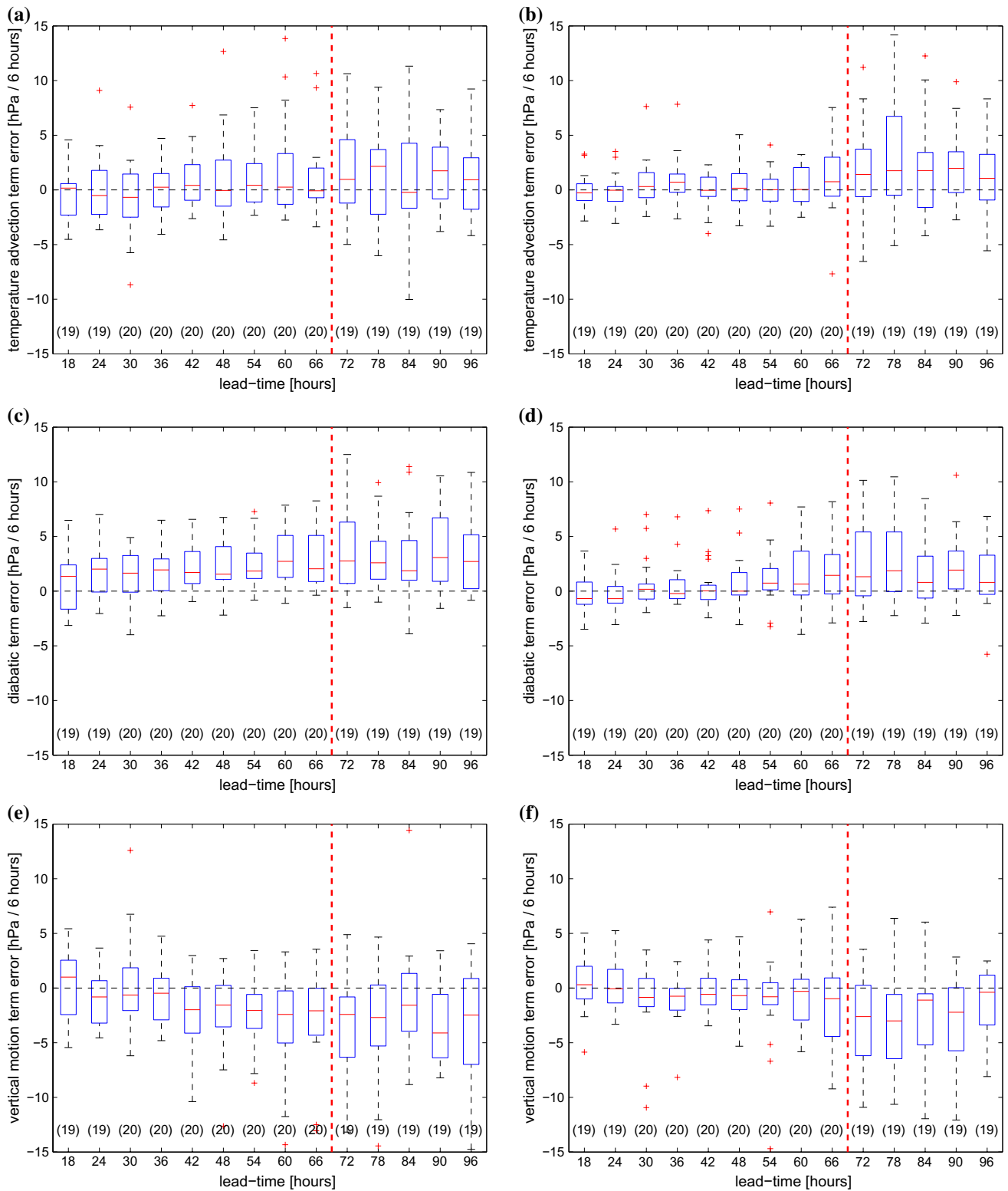


Fig. 7 As Fig. 6 but showing the split of the total column warming (*ITT*, Fig. 6g, h) into horizontal advection (*TADV*, **a**, **b**), diabatic heating (*DIAB_{RES}*, **c**, **d**) and effects of vertical motion (*VMT*, **e**, **f**).

For a detailed explanation of the meanings of the different terms, see Sect. 2.4. The dashed red lines mark 3 days lead time, when hindcast typically start to degrade

a direct measure of the absolute depth of a cyclone but of the pressure fall at a given location during the time when the cyclone approaches. We have already seen in Fig. 3a that the absolute core pressure bias increases with hindcast lead time. The fact that Dp does not increase in the same way suggests that the pressure gradient to the east of the cyclone centre is affected less, as this plays a big role for the Eulerian time change measured with Dp , which is also not affected by a linear increase in pressure bias with lead time. Going to higher resolution results in a general shift to smaller pressure biases, as expected, but surprisingly shorter lead times now even show a negative bias (Fig. 6b). As this is not found for absolute core pressure (Fig. 3b), a possible explanation is a difference in the morphology of the cyclone (size, distribution of pressure gradients) between ERA-Interim and ECHAM6 as discussed above or possibly effects to do with the spin-up of clouds.

Contributions to deepening from geopotential height tendencies at the upper integration boundary ($D\phi$) show negative medians, typically of about 0.5 hPa per 6 h and less, for all lead times less than 3 days for both resolutions (Fig. 6c, d). After that the values become more varied and the spread increases. Differences between the two resolutions are small with T127 showing a slightly smaller spread and overall more negative values. It remains an open question why ECHAM6 shows such a systematically different behaviour in 100-hPa geopotential height tendencies compared to ERA-Interim. Vertical resolution can be ruled out as a dominating effect, as the T127 simulations have 95 levels and are therefore better resolved than ERA-Interim with 60. It appears more likely that the different model physics create differences in the way that cyclone deepening is represented dynamically. The median of the differences in the EP term is smaller than the other terms and clearly positive for all lead times and both resolutions (around 0.2 hPa per 6 h; Fig. 6e, f). This implies less precipitation (and/or more evaporation) in ECHAM6 than in ERA-Interim in the area around the cyclone centre. Differences between lead times are small and not systematic. Values for T127 are generally a little smaller, but the overall pattern is strikingly similar. Again this suggests that these deviations are resulting from fundamental differences in the physics of the ECMWF and ECHAM6 models. It should be pointed out here that precipitation and evaporation are purely model-generated in ERA-Interim and therefore only indirectly affected by the assimilation of observational data in contrast to, for example, temperature, humidity or winds.

The last term needed to close the overall budget is ITT (Fig. 6g, h). As the usually dominating term, many similarities to the evolution of Dp (Fig. 6a, b) are found. For the T63, all medians are positive, increase slightly with lead time accompanied by an increase in spread for the longest lead times (Fig. 6g). The similarities to Dp are

to a large extent due to a cancellation of negative differences from $D\phi$ (Fig. 6c) and positive differences from EP (Fig. 6e). For T127 (Fig. 6h), simulations agree much better with ERA-Interim with medians very close to zero for the first two simulation days and mostly positive values and a larger spread afterwards. For the first 48 h, this good agreement is, at least mathematically, the consequence of the negative $D\phi$ value (Fig. 6d) overcompensating positive EP values (Fig. 6f) to create the negative Dp values in Fig. 6b. So effectively, the largest difference between the two resolutions is a lack of column warming in the coarse resolution, ultimately leading to too small pressure tendencies down-track of the cyclone centre. The reasons for these differences in ITT terms will be looked at next.

Figure 7 shows the split-up of the ITT term into its three components $TADV$, $DIAB_{RES}$ and VMT . The contribution of horizontal temperature advection appears to be well represented by both resolutions out to about 66 h lead time, after which it becomes positive and spread increases (Fig. 7a, b). As in many such plots before, the spread is smaller for T127, but differences in the median are rather small, indicating that both resolutions are capable of realistically handling this part of the cyclone deepening process. This is not surprising as it mostly depends on the large-scale baroclinicity and the cyclone-scale horizontal winds. In contrast, $DIAB_{RES}$ is highly resolution-dependent. For T63, all medians are positive with large values around 2 hPa per 6 h during the first three days and comparably large spread, which then increases even further for the longest lead times (Fig. 7c). Simulations at T127 (Fig. 7d) in contrast start with slightly negative medians, stay close to zero for 30–48 h lead times and then increase to positive values, however still smaller than for T63. The spread is significantly smaller than for T63, particularly for the shorter lead times. These results are consistent with more precipitation in T127 (see Fig. 6e, f) and strongly suggest that this the most important difference between the two resolutions, as for example proposed in a longer-term modelling experiment by Willison et al. (2013). Finally, Fig. 7e, f show differences between the two sets of ECHAM6 experiments and ERA-Interim for the VMT . For T63, this term is initially close to zero, but then increasingly becomes negative, reaching values of 2–3 hPa per 6 h for longer lead times. This behaviour compensates some of the positive values in $TADV$ (Fig. 7a, longer lead times only) and $DIAB_{RES}$ (Fig. 7c, all lead times) to create overall smaller errors in ITT (Fig. 6g). This is physically plausible as both positive horizontal temperature advection and latent heat release are connected to upward vertical motion, which creates adiabatic cooling. T127 shows a similar behaviour for lead times from 66 h onwards, but for the shorter lead times medians are close to zero or

slightly negative with small spreads, in agreement with the good results for *TADV* (Fig. 7b), *DIAB_{RES}* (Fig. 7d) and *ITT* (Fig. 6g).

4 Discussion and conclusions

To the best of our knowledge, this study is the first to apply the seamless approach of combining aspects of weather and climate modelling to the specific problem of evaluating model simulations of severe North Atlantic windstorms that are an important natural hazard for Europe. Currently, there is relatively little consensus between different climate models with regard to the frequency and intensity of such storms and the confidence in any climate change signal is low (Christensen et al. 2013). This poses a significant risk to European societies and the (re-)insurance business. Widely used statistical analyses of climate model output make it difficult to disentangle different sources of model errors such as those related to slow, large-scale processes like in the stratosphere and oceans and those associated with the actual fast dynamics of explosively deepening cyclones. The seamless approach followed in this study provides an alternative that alleviates this difficulty by performing short, NWP-type hindcasts of the most destructive storms, which are initialised from a well defined atmospheric state and then compared with the best approximation of the actual evolution of the atmosphere as provided by reanalysis products, allowing a clear separation between fast and slow processes. 20 historical storms during the last quarter of a century were objectively selected for this study. In total over 1500 hindcasts were conducted using the state-of-the-art climate models ECHAM6 and UM-GA4 with differing lead times and horizontal and vertical resolutions. Cyclones were then objectively tracked and position and core pressures compared to the corresponding tracks from ERA-Interim data. Physical reasons for differences between the simulations were analysed with a PTE diagnostic tool developed by Fink et al. (2012).

The main conclusions from this study are:

- A. Despite the considerable biases found in free-running simulations (e.g. Zappa et al. 2013a), the two climate models used here are generally capable of reproducing the selected intense cyclones when given the analysed initial conditions.
- B. Particularly the cyclone tracks are generally well reproduced up to several days lead time.
- C. However, the deepening is systematically too weak in the coarser resolution runs (T63 for ECHAM6 and N96 for UM-GA4, corresponding to about 200 and 140 km, respectively), reaching biases of 5 hPa after just a couple of days simulation time. This result is consistent

with the statistical analysis by Zappa et al. (2013a), strongly suggesting that this resolution is too coarse for storm studies and should not be used for this purpose in the future.

- D. Sensitivity to vertical resolution is small, suggesting a minor influence on the storm evolution.
- E. According to the PTE analysis, the main source of error at coarse resolution is due to the representation of diabatic processes (mainly latent heating). This result is in agreement with experiments by Willison et al. (2013) with a regional climate model, suggesting that key dynamical features like the diabating production of potential vorticity along fronts are insufficiently represented. This model deficiency may become even more problematic in future climate projections, as diabatic processes have been suggested to become more important for cyclone deepening in a warmer climate (e.g., Colle et al. 2013; Marciano et al. 2014).
- F. A doubling of the resolution to T127 (corresponds to about 100 km) appears to be already sufficient to remove most of these biases, at least up to 3 days lead time.

It would be interesting to compare the modelling results presented here with output from state-of-the-art NWP models to gauge the potential improvement from even higher resolution and better model physics. At the same time, comparisons to ensemble predictions would give a measure for the sensitivity to initial conditions. Other remaining questions are sensitivities of the results to the tracking method (see Neu et al. 2013) and the influence of resolution on intense open-wave systems that are particularly difficult to track. The good performance of the T127 simulations in combination with the often observed underprediction of intense cyclones in free-running climate models raises the question of a possible lack of “storm-prone situations” in those simulations, i.e. large-scale conditions over the North Atlantic known to favour the occurrence of explosively deepening storms such as certain phases of the North Atlantic Oscillation (Pinto et al. 2009a, 2014; Vitolo et al. 2009; Hanley and Caballero 2012). Another possible explanation is that the model runs benefit to some degree from the finer resolution inherent in the initial conditions provided by ERA-Interim re-analyses produced at T255. However, Jung et al. (2006) show that numbers of cyclones with core pressures less than 980 hPa are still underestimated in free-running seasonal integrations at T255. The work of Willison et al. (2013) suggests that deficiencies in the representation of individual storms can feedback onto the storm track as a whole and therefore possibly on conditions for subsequent storm evolutions. This is a question of great importance for future model improvements that will be addressed in future work.

Acknowledgments The authors would like to thank the AXA Research Fund for financing the SEAMSEW (A Seamless Approach to Assessing Model Uncertainties in Climate Projections of Severe European Windstorms) project at the University of Leeds; Sebastian Rast (Max-Planck Institute) and Daniel Klocke (German Weather Service) for help with setting up ECHAM6; the N8 High Performance Computing Research Partnership for computing facilities and support. We acknowledge use of the MONSooN system, a collaborative facility supplied under the Joint Weather and Climate Research Programme, which is a strategic partnership between the Met Office and the Natural Environment Research Council. K.D. Williams was supported by the Joint UK DECC/Defra Met Office Hadley Centre Climate Programme (GA01101). We also thank two anonymous reviewers for their helpful comments that contributed to improve this manuscript.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Bengtsson L, Hodges KI, Roeckner E (2006) Storm tracks and climate change. *J Clim* 19:3518–3543
- Bengtsson L, Hodges KI, Keenlyside N (2009) Will extratropical storms intensify in a warmer climate? *J Clim* 22:2276–2301
- Boyle JS, Williamson D, Cederwall R, Fiorino M, Hnilo J, Olson J, Phillips T, Potter G, Xie S (2005) Diagnosis of Community Atmospheric Model 2 (CAM2) in numerical weather forecast configuration at atmospheric radiation measurement sites. *J Geophys Res* 110(D15):D15S15. doi:10.1029/2004JD005042
- Christensen JH, Kumar KK, Aldrian E, An SI, Cavalcanti IFA, de Castro M, Dong W, Goswami P, Hall A, Kanyanga JK, Kitoh A, Kossin J, Lau NC, Renwick J, Stephenson DB, Xie SP, Zhou T (2013) Climate phenomena and their relevance for future regional climate change. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. In: Stocker TF, Qin D, Plattner GK, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM (eds) *Climate change 2013: the physical science basis*. Cambridge University Press, Cambridge
- Colle BA, Zhang Z, Lombardo KA, Chang E, Liu P, Zhang M (2013) Historical evaluation and future prediction of Eastern North American and Western Atlantic extratropical cyclones in the CMIP5 models during the cool season. *J Clim* 26(18):6882–6903. doi:10.1175/JCLI-D-12-00498.1
- Dee DP, Uppala SM, Simmons AJ, Berrisford P, Poli P, Kobayashi S, Andrae U, Balmaseda MA, Balsamo G, Bauer P, Bechtold P, Beljaars ACM, van de Berg L, Bidlot J, Bormann N, Delsol C, Dragani R, Fuentes M, Geer AJ, Haimberger L, Healy SB, Hersbach H, Hólm EV, Isaksen I, Kållberg P, Köhler M, Matricardi M, McNally AP, Monge-Sanz BM, Morcrette JJ, Park BK, Peubey C, de Rosnay P, Tavolato C, Thépaut JN, Vitart F (2011) The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q J R Meteorol Soc* 137(656):553–597
- Donat MG, Leckebusch GC, Wild S, Ulbrich U (2011) Future changes in European winter storm losses and extreme wind speeds inferred from GCM and RCM multi-model simulations. *Nat Hazards Earth Syst Sci* 11(5):1351–1370. doi:10.5194/nhess-11-1351-2011, <http://www.nat-hazards-earth-syst-sci.net/11/1351/2011/>
- Fink AH, Bruecher T, Ermert V, Krueger A, Pinto JG (2009) The European storm Kyrill in January 2007: synoptic evolution, meteorological impacts and some considerations with respect to climate change. *Nat Hazards Earth Syst Sci* 9(2):405–423
- Fink AH, Pohle S, Pinto JG, Knippertz P (2012) Diagnosing the influence of diabatic processes on the explosive deepening of extratropical cyclones. *Geophys Res Lett* 39(7):L07–803. doi:10.1029/2012GL051025
- Flato G, Marotzke J, Abiodun B, Braconnot P, Chou S, Collins W, Cox P, Driouech F, Emori S, Eyring V, Forest C, Gleckler P, Guilyardi E, Jakob C, Kattsov V, Reason C, Rummukainen M (2013) Evaluation of climate models. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. In: Stocker TF, Qin D, Plattner GK, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM (eds) *Climate change 2013: the physical science basis*. Cambridge University Press, Cambridge
- Froude LSR (2009) Regional differences in the prediction of extratropical cyclones by the ECMWF Ensemble Prediction System. *Mon Weather Rev* 137:893–911
- Froude LSR, Bengtsson L, Hodges KI (2007) The prediction of extratropical storm tracks by the ECMWF and NCEP Ensemble Prediction Systems. *Mon Weather Rev* 135(7):2545–2567. doi:10.1175/MWR3422.1
- Goyette S, Beniston M, Caya D, Laprise R, Jungo P (2001) Numerical investigation of an extreme storm with the Canadian Regional Climate Model: the case study of windstorm Vivian, Switzerland, February 27 1990. *Clim Dyn* 18(1–2):145–168. doi:10.1007/s003820100166
- Greeves CZ, Pope VD, Stratton RA, Martin GM (2007) Representation of Northern Hemisphere winter storm tracks in climate models. *Clim Dyn* 28:683–702
- Grise KM, Son SW, Gyakum JR (2013) Intraseasonal and interannual variability in north american storm tracks and its relationship to equatorial pacific variability. *Mon Weather Rev* 141(10):3610–3625. doi:10.1175/MWR-D-12-00322.1
- Hanley J, Caballero R (2011) Objective identification and tracking of multicentre cyclones in the ERA-Interim reanalysis dataset. *Q J R Meteorol Soc*. doi:10.1002/qj.948
- Hanley J, Caballero R (2012) The role of large-scale atmospheric flow and Rossby Wave Breaking in the evolution of extreme windstorms over Europe. *Geophys Res Lett* 39:L21–708
- Harvey BJ, Shaffrey LC, Woollings TJ, Zappa G, Hodges KI (2012) How large are projected 21st century storm track changes? *Geophys Res Lett*. doi:10.1029/2012GL052873
- Hawkins E, Sutton R (2009) The potential to narrow uncertainty in regional climate predictions. *Bull Am Meteorol Soc* 90(8):1095–1107. doi:10.1175/2009BAMS2607.1
- Hodges KI (1994) A general method for tracking analysis and its application to meteorological data. *Mon Weather Rev* 122(11):2573–2586. doi:10.1175/1520-0493(1994)122<2573:AGMFTA>2.0.CO;2
- Hodges KI (1995) Feature tracking on the unit sphere. *Mon Weather Rev* 123(12):3458–3465. doi:10.1175/1520-0493(1995)123<3458:FTOTUS>2.0.CO;2
- Hodges KI, Hoskins BJ, Boyle J, Thorncroft C (2003) A comparison of recent reanalysis datasets using objective feature tracking: storm tracking and tropical easterly waves. *Mon Weather Rev* 131:2012–2037
- Jung T, Gulev SK, Rudeva I, Soloviev V (2006) Sensitivity of extratropical cyclone characteristics to horizontal resolution in the ECMWF model. *Q J R Meteorol Soc* 132(619):1839–1857. doi:10.1256/qj.05.212
- Knippertz P, Ulbrich U, Speth P (2000) Changing cyclone and surface wind speeds over the North Atlantic and Europe in a transient GHG experiment. *Clim Res* 15:109–120
- Knippertz P, Fink AH, Pohle S (2009) Comments on “dry-season precipitation in tropical West Africa and its relation to forcing from

- the extratropics" reply. *Mon Weather Rev* 137(9):3151–3157. doi:[10.1175/2009MWR3006.1](https://doi.org/10.1175/2009MWR3006.1), <http://journals.ametsoc.org/doi/pdf/10.1175/2009MWR3006.1>
- Lamb H, Frydendahl K (1991) *Historic storms of the North Sea*. Cambridge University Press, British Isles
- Lambert SJ, Fyfe JC (2006) Changes in winter cyclone frequencies and strengths simulated in enhanced greenhouse warming experiments: results from the models participating in the IPCC diagnostic exercise. *Clim Dyn* 26:713–728
- Leckebusch G, Renggli D, Ulbrich (2008) Development and application of an objective storm severity measure for the northeast Atlantic region. *Meteorol Z* 17(5):575–587
- Leckebusch GC, Koffi B, Ulbrich U, Pinto JG (2006) Analysis of frequency and intensity of European winter storms from a multi-model perspective, at synoptic and regional scales. *Clim Res* 31:59–74
- Leckebusch GC, Ulbrich U, Fröhlich L, Pinto JG (2007) Property loss potentials for European midlatitude storms in a changing climate. *Geophys Res Lett* 34(L05):703
- Liberato MLR, Pintel JG, Trigo IF, Trigo RM (2011) Klaus—an exceptional winter storm over northern Iberia and southern France. *Weather* 66:330–334
- Löptien U, Zolina O, Gulev S, Latif M, Soloviev V (2008) Cyclone life cycle characteristics over the Northern Hemisphere in coupled GCMs. *Clim Dyn* 31(5):507–532. doi:[10.1007/s00382-007-0355-5](https://doi.org/10.1007/s00382-007-0355-5)
- Ludwig P, Pinto JG, Meyers M, Gray SL (2014) The role of anomalous SST and surface fluxes over the southeastern North Atlantic in the explosive development of windstorm Xynthia. *Q J R Meteorol Soc* 140(682):1729–1741. doi:[10.1002/qj.2253](https://doi.org/10.1002/qj.2253)
- Madonna E, Boettcher M, Grams CM, Joos H, Martius O, Wernli H (2014) Verification of North Atlantic warm conveyor belt outflows in ECMWF forecasts. *Q J R Meteorol Soc*. doi:[10.1002/qj.2442](https://doi.org/10.1002/qj.2442)
- Manzini E, Karpechko AY, Anstey J, Baldwin MP, Black RX, Cagnazzo C, Calvo N, Charlton-Perez A, Christiansen B, Davini P, Gerber E, Giorgetta M, Gray L, Hardiman SC, Lee YY, Marsh DR, McDaniel BA, Purich A, Scaife AA, Shindell D, Son SW, Watanabe S, Zappa G (2014) Northern winter climate change: assessment of uncertainty in CMIP5 projections related to stratosphere–troposphere coupling. *J Geophys Res Atmos* 119(13):7979–7998. doi:[10.1002/2013JD021403](https://doi.org/10.1002/2013JD021403)
- Marciano CG, Lackmann GM, Robinson WA (2014) Changes in U.S. East Coast cyclone dynamics with climate change. *J Clim* 28(2):468–484. doi:[10.1175/JCLI-D-14-00418.1](https://doi.org/10.1175/JCLI-D-14-00418.1)
- Munich Re (2009) World map of natural hazards. available on request from Munich Re Touch
- Murray RJ, Simmonds I (1991) A numerical scheme for tracking cyclone centres from digital data, part I: development and operation of the scheme. *Aust Meteorol Mag* 39:155–166
- Neu U, Akperov MG, Bellenbaum N, Benestad R, Blender R, Caballero R, Cocozza A, Dacre HF, Feng Y, Fraedrich K, Grieger J, Gulev S, Hanley J, Hewson T, Inatsu M, Keay K, Kew SF, Kindem I, Leckebusch GC, Liberato MLR, Lionello P, Mokhov II, Pinto JG, Raible CC, Reale M, Rudeva I, Schuster M, Simmonds I, Sinclair M, Sprenger M, Tilinina ND, Trigo IF, Ulbrich S, Ulbrich U, Wang XL, Wernli H (2013) IMILAST: a community effort to intercompare extratropical cyclone detection and tracking algorithms. *Bull Am Meteorol Soc* 94(4):529–547. doi:[10.1175/BAMS-D-11-00154.1](https://doi.org/10.1175/BAMS-D-11-00154.1)
- Palmer TN, Doblas-Reyes FJ, Weisheimer A, Rodwell MJ (2008) Toward seamless prediction: calibration of climate change projections using seasonal forecasts. *Bull Am Meteorol Soc* 89:459–470
- Pennell C, Reichler T (2011) On the effective number of climate models. *J Clim* 24(9):2358–2367. doi:[10.1175/2010JCLI3814.1](https://doi.org/10.1175/2010JCLI3814.1)
- Phillips TJ, Potter GL, Williamson DL, Cederwall RT, Boyle JS, Fiorino M, Hnilo JJ, Olson JG, Xie S, Yio JJ (2004) Evaluating parameterizations in general circulation models: climate simulation meets weather prediction. *Bull Am Meteorol Soc* 85(12):1903–1915. doi:[10.1175/BAMS-85-12-1903](https://doi.org/10.1175/BAMS-85-12-1903), <http://journals.ametsoc.org/doi/abs/10.1175/BAMS-85-12-1903>
- Pinto JG, Spanghel T, Ulbrich U, Speth P (2005) Sensitivities of a cyclone detection and tracking algorithm: individual tracks and climatology. *Meteorologische Zeitschrift* 14:823–838. doi:[10.1127/0941-2948/2005/0068](https://doi.org/10.1127/0941-2948/2005/0068), <http://www.ingentaconnect.com/content/schweiz/mz/2005/00000014/00000006/art00015>
- Pinto JG, Spanghel T, Ulbrich U, Speth P (2006) Assessment of winter cyclone activity in a transient ECHAM4-OPYC3 GHG experiment. *Meteorol Z* 15:279–291
- Pinto JG, Fröhlich E, Leckebusch G, Ulbrich U (2007a) Changing European storm loss potentials under modified climate conditions according to ensemble simulations of the ECAHM5/MPI-OM1 GCM. *Nat Hazards Earth Syst Sci* 7:165–175
- Pinto JG, Ulbrich U, Leckebusch GC, Spanghel T, Meyers M, Zacharias S (2007b) Changes in storm track and cyclone activity in three SRES ensemble experiments with the ECHAM5/MPI-OM1 GCM. *Clim Dyn* 29:195–210
- Pinto JG, Zacharias S, Fink AH, Leckebusch GC, Ulbrich U (2009) Factors contributing to the development of extreme North Atlantic cyclones and their relationship with the NAO. *Clim Dyn* 32:711–737
- Pinto JG, Gmara I, Masato G, Dacre HF, Woollings T, Caballero R (2014) Large-scale dynamics associated with clustering of extratropical cyclones affecting Western Europe. *J Geophys Res Atmos*. doi:[10.1002/2014JD022305](https://doi.org/10.1002/2014JD022305)
- Rivière G, Arbogast P, Maynard K, Joly A (2010) The essential ingredients leading to the explosive growth stage of the European wind storm Lothar of Christmas 1999. *Q J R Meteorol Soc* 136(648):638–652. doi:[10.1002/qj.585](https://doi.org/10.1002/qj.585)
- Rivière G, Arbogast P, Lapeyre G, Maynard K (2012) A potential vorticity perspective on the motion of a mid-latitude winter storm. *Geophys Res Lett*. doi:[10.1029/2012GL052440](https://doi.org/10.1029/2012GL052440)
- Rivière G, Arbogast P, Joly A (2014) Eddy kinetic energy redistribution within windstorms Klaus and Friedhelm. *Q J R Meteorol Soc*. doi:[10.1002/qj.2412](https://doi.org/10.1002/qj.2412)
- Roberts JF, Champion AJ, Dawkins LC, Hodges KI, Shaffrey LC, Stephenson DB, Stringer MA, Thornton HE, Youngman BD (2014) The XWS open access catalogue of extreme European windstorms from 1979 to 2012. *Nat Hazards Earth Syst Sci* 14(9):2487–2501. doi:[10.5194/nhess-14-2487-2014](https://doi.org/10.5194/nhess-14-2487-2014), <http://www.nat-hazards-earth-syst-sci.net/14/2487/2014/>
- Rockel B, Woth K (2007) Extremes of near-surface wind speed over Europe and their future changes as estimated from an ensemble of RCM simulations. *Clim Change* 81:267–280
- Rodwell MJ, Palmer TN (2007) Using numerical weather prediction to assess climate models. *Q J R Meteorol Soc* 133:129–146. <http://onlinelibrary.wiley.com/doi/10.1002/qj.23/abstract>
- Salari V, Sethi IK (1990) Feature point correspondence in the presence of occlusion. *IEEE Trans Pattern Anal Mach Intell* 12(1):87–91. doi:[10.1109/34.41387](https://doi.org/10.1109/34.41387)
- Scaife A, Spanghel T, Fereday D, Cubasch U, Langematz U, Akiyoshi H, Bekki S, Braesicke P, Butchart N, Chipperfield M, Gettelman A, Hardiman S, Michou M, Rozanov E, Shepherd T (2012) Climate change projections and stratosphere–troposphere interaction. *Clim Dyn* 38(9–10):2089–2097. doi:[10.1007/s00382-011-1080-7](https://doi.org/10.1007/s00382-011-1080-7)
- Schwierz C, Köllner-Heck P, Zenklusen Mutter E, Bresch DN, Vidale PL, Wild M, Schär C (2010) Modelling European winter wind storm losses in current and future climate. *Clim Change* 101(3–4):485–514. doi:[10.1007/s10584-009-9712-1](https://doi.org/10.1007/s10584-009-9712-1)
- UKCP (2009) <http://ukclimateprojections.defra.gov.uk>

- Ulbrich U, Pinto JG, Kupfer H, Leckebusch GC, Spanghel T, Reyers M (2008) Changing Northern Hemisphere storm tracks in an ensemble of IPCC climate change simulations. *J Clim* 21:1669–1679
- Ulbrich U, Leckebusch GC, Pinto JG (2009) Extra-tropical cyclones in the present and future climate: a review. *Theor Appl Climatol* 96(1–2):117–131
- Veenman CJ, Hendriks EA, Reinders MJT (1998) A fast and robust point tracking algorithm. In: *IEEE international conference on image processing*, vol III, pp 653–657
- Vitolo R, Stephenson DB, Cook IM, Mitchell-Wallace K (2009) Serial clustering of intense European storms. *Meteorol Z* 18:411–424
- Walters DN, Williams KD, Boutle IA, Bushell AC, Edwards JM, Field PR, Lock AP, Morcrette CJ, Stratton RA, Wilkinson JM, Willett MR, Bellouin N, Bodas-Salcedo A, Brooks ME, Copsey D, Earnshaw PD, Hardiman SC, Harris CM, Levine RC, MacLachlan C, Manners JC, Martin GM, Milton SF, Palmer MD, Roberts MJ, Rodríguez JM, Tennant WJ, Vidale PL (2014) The Met Office Unified Model Global Atmosphere 4.0 and JULES Global Land 4.0 configurations. *Geosci Model Devel* 7(1):361–386. doi:10.5194/gmd-7-361-2014, <http://www.geosci-model-dev.net/7/361/2014/>
- Wernli H, Dirren S, Liniger MA, Zillig M (2002) Dynamical aspects of the winter storm ‘Lothar’ (24–26 December 1999). *Q J R Meteorol Soc* 128:405–429
- Wilks D (2011) *Statistical methods in the atmospheric sciences*. Academic Press, London
- Williams KD, Brooks ME (2008) Initial tendencies of cloud regimes in the Met Office Unified Model. *J Clim* 21(4):833–840. doi:10.1175/2007JCLI1900.1
- Williams KD, Bodas-Salcedo A, Déqué M, Fermepein S, Medeiros B, Watanabe M, Jakob C, Klein SA, Senior CA, Williamson DL (2013) The Transpose-AMIP II experiment and its application to the understanding of Southern Ocean cloud biases in climate models. *J Clim* 26(10):3258–3274. doi:10.1175/JCLI-D-12-00429.1
- Willison J, Robinson WA, Lackmann GM (2013) The importance of resolving mesoscale latent heating in the North Atlantic storm track. *J Atmos Sci* 70(7):2234–2250. doi:10.1175/JAS-D-12-0226.1
- Woollings T, Gregory JM, Pinto JG, Reyers M, Brayshaw DJ (2012) Response of the North Atlantic storm track to climate change shaped by ocean–atmosphere coupling. *Nat Geosci* 5(5):313–317. doi:10.1038/ngeo1438
- Yang ZL, Dickinson RE, Henderson-Sellers A, Pitman AJ (1995) Preliminary study of spin-up processes in land surface models with the first stage data of Project for Intercomparison of Land Surface Parameterization Schemes Phase 1(a). *J Geophys Res Atmos* 100(D8):16553–16578. doi:10.1029/95JD01076
- Zappa G, Shaffrey L, Hodges K (2013a) The ability of CMIP5 models to simulate North Atlantic extratropical cyclones. *J Clim* 26:5379–5396
- Zappa G, Shaffrey LC, Hodges KI, Sansom PG, Stephenson DB (2013b) A multimodel assessment of future projections of North Atlantic and European extratropical cyclones in the CMIP5 climate models. *J Clim* 26(16):5846–5862. doi:10.1175/JCLI-D-12-00573.1