RESEARCH ARTICLE

A well-observed polar low analysed with a regional and a global weather-prediction model

Patrick J. Stoll[1] | Teresa M. Valkonen[2] | Rune G. Graversen[1] | Gunnar Noer[3]

[1]Department of Physics and Technology, Arctic University of Norway, Tromsø, Norway

[2]Norwegian Meteorological Institute, Oslo, Norway

[3]Norwegian Meteorological Institute, Tromsø, Norway

Correspondence
P. J. Stoll, Department of Physics and Technology, The Arctic University of Norway, 9037 Tromsø, Norway.
Email: patrick.stoll@uit.no

Abstract

The capability of a regional (AROME-Arctic) and a global (ECMWF HRES) weather-prediction model are compared for simulating a well-observed polar low (PL). This PL developed on 3–4 March 2008 and was measured by dropsondes released from three flights during the IPY-THORPEX campaign. Validation against these measurements reveals that both models simulate the PL reasonably well. AROME-Arctic appears to represent the cloud structures and the high local variability more realistically. The high local variability causes standard error statistics to be similar for AROME-Arctic and ECMWF HRES. A spatial verification technique reveals that AROME-Arctic has improved skills at small scales for extreme values. However, the error growth of the forecast, especially in the location of the PL, is faster in AROME-Arctic than in ECMWF HRES. This is likely associated with larger convection-induced perturbations in the former than the latter model. Additionally, the PL development is analysed. This PL has two stages, an initial baroclinic and a convective mature stage. Sensible heat flux and condensational heat release both contribute to strengthen the initial baroclinic environment. In the mature stage, latent heat release appears to maintain the system. At least two conditions must be met for this stage to develop: (a) the sensible heat flux sufficiently destabilises the local environment around the PL, and (b) sufficient moisture is available for condensational heat release. More than half of the condensed moisture within the system originates from the surroundings. The propagation of the PL is "pulled" towards the area of strongest condensational heating. Finally, the sensitivity of the PL to the sea-surface temperature is analysed. The maximum near-surface wind speed connected to the system increases by $1-2\,\mathrm{m\cdot s^{-1}}$ per K of surface warming and a second centre develops in cases of highly increased temperature.

KEYWORDS

polar low, mesoscale cyclone, AROME-Arctic, baroclinic instability, cold-air outbreak, diabatic processes, ECMWF HRES, HARMONIE-AROME, IPY-THORPEX

1 | INTRODUCTION

Polar lows (PLs) are small but intense cyclones developing in cold air masses that flow over large water surfaces, known as cold-air outbreaks (CAOs; Rasmussen and Turner, 2003). The associated strong winds, high waves and substantial snowfall are a threat for coastal communities and maritime operations at high latitudes. The Northeast Atlantic is one of the areas with the most frequent PL occurrence (Stoll *et al.*, 2018).

PLs are mesoscale cyclones with a typical diameter of 200–600 km (Rojo *et al.*, 2015). They develop and intensify rapidly, generally within a few hours. Hence, hazardous conditions associated with PLs appear at short notice. In contrast to synoptic-scale cyclones, their lifetime rarely exceeds two days (Rojo *et al.*, 2015).

Because of their fast development and due to the sparse observation network in polar regions, the prediction of PLs is a challenge for meteorological services (Furevik *et al.*, 2015). Numerical weather prediction (NWP) models still have issues to correctly represent important details of convection and the stable atmospheric boundary layer of cold air masses (Holtslag *et al.*, 2013). These two processes are relevant during and before, respectively, the PL development. Up-to-date regional NWP models show substantial differences in their representation of convection connected to CAOs (Field *et al.*, 2017).

Detailed measurements of the development of PLs are rare. Only a handful of flight campaigns have been performed (Shapiro *et al.*, 1987; Douglas *et al.*, 1991; Douglas *et al.*, 1995; Brümmer *et al.*, 2009). In February and March 2008, in connection with the International Polar Year (IPY) of The Observing System Research and Predictability Experiment (THORPEX), several flight missions were conducted in the Northeast Atlantic (Kristjánsson *et al.*, 2011). Two PL cases and several other Arctic marine boundary-layer phenomena were observed by an aircraft. To the knowledge of the authors, the only PL observed by multiple flights was that monitored during the IPY-THORPEX campaign on 3–4 March 2008. This, commonly referred to as the THORPEX PL, is among the most investigated PLs, and is also scrutinised in this study.

Føre *et al.* (2011) described this PL based on dropsonde data obtained from the flights, satellite images and the weather-prediction model HIRLAM (High-Resolution Limited-Area Model), operational at the Norwegian Meteorological Institute (MET Norway) at that time. Føre and Nordeng (2012) use the Weather Research and Forecasting model (WRF) with 3 km horizontal grid-spacing and non-hydrostatic dynamical core to investigate the effect of surface energy fluxes and condensational heat release on the intensification of the PL. Wagner *et al.* (2011) performed WRF simulations with 2 km grid-spacing and

compared these to lidar and dropsonde measurements obtained from the flight campaigns. Innes *et al.* (2011) used the Met Office Unified Model (UM) with grid-spacing of 12, 4 and 1 km to investigate the effect of the model grid-spacing on the PL simulation. They found that the 4 km version performed considerably better than the 12 km version, while the 1 km simulation did not improve the representation of the PL with that particular model.

Føre *et al.* (2011) and Kristjánsson *et al.* (2011) suggested using the observations retrieved from the IPY-THORPEX campaign for model validation. In this study, the state-of-the-art regional weather-prediction model, AROME-Arctic (Müller *et al.*, 2017a), is validated against this dataset. AROME-Arctic (Applications of Research to Operations at MEsoscale for the European Arctic) has been used operationally at MET Norway since 2015. AROME-Arctic (AA) is the first operational model for the European Arctic with a non-hydrostatic core that permits convection. The model system from which AA originates is utilised by numerous other European meteorological services for operational weather forecasting (Bengtsson *et al.*, 2017). This model system is also currently employed for the production of the first regional reanalysis of the European Arctic. Section 2.1 gives more details on the model.

At present, AA is the main tool for forecasting PLs that develop in the Nordic Seas and offer a threat to the Norwegian coast. Due to the non-hydrostatic, convective-permitting dynamics, AA is expected to be more suited for simulating the development of PLs than previous hydrostatic models. Müller *et al.* (2017a) conclude that a PL which occurred on 8 December 2016 was represented with higher accuracy in AA than in the operational High RESolution global weather-prediction model (HRES) of the European Centre for Medium-Range Weather Forecasts (ECMWF; ECMWF, 2018). More details to ECMWF HRES are given in Section 2.2. However, in that study the models were only compared for their performance in simulating the near-surface wind speeds. The model representation of the three-dimensional dynamical structure of a PL has not yet been investigated for AA.

The capability of AA to accurately simulate the THORPEX PL is evaluated in the first part of this study, with the explicit purpose of revealing the strengths and weaknesses of the model. Furthermore, the representation of this PL by AA is compared to the performance of the weather-prediction model HRES.

In the second part of the study, the focus is on the development mechanisms of the PL in question. However, this is connected to the first, since a better understanding of the PL evolution eases the identification of the model components that need improvement to increase the forecast quality of PLs.

Multiple development mechanisms, such as baroclinic instability, shear instability, upper-level potential vorticity forcing, orographic vortex generation, convection, and diabatic processes have been recognised as being important for the intensification of PLs (Rasmussen and Turner, 2003; Terpstra *et al.*, 2015). Often, the mentioned mechanisms interact nonlinearly, implying that the role of every single component is difficult to examine (Bracegirdle, 2006).

The importance of the different mechanisms varies among PL cases, which is the major reason why no standard model for PL development has been developed. Furthermore, the importance of the different mechanisms changes during the lifetime of a PL. Some PLs were observed to develop initially in a baroclinic environment and subsequently to intensify by convective processes (e.g. Nordeng and Rasmussen, 1992). The PL investigated in this study follows such development.

Various idealised numerical simulations have been performed in order to understand the development of PLs. For example, Terpstra *et al.* (2015) applied a baroclinic channel model adapted for high-latitude conditions to demonstrate that a low-level disturbance requires a "diabatic boost" in order to amplify quickly. The occurrence of this "boost" depends on sufficient humidity and baroclinicity and weak static stability. They conceptually described the growing perturbation in the context of the Diabatic Rossby Vortex (DRV), where potential vorticity is produced below the source of latent heating.

Yanase and Niino (2005; 2007) showed in idealised experiments that the cloud structure can be associated with the dominant development mechanism. Simulations with a strong baroclinic environment lead to cyclones with comma-shaped clouds. In the absence of baroclinicity, spiral-form convective clouds develop, as seen in "hurricane-like" PLs (e.g. Nordeng and Rasmussen, 1992).

Also, NWP models have been utilised for investigation of the physical development mechanisms of PLs. Often sensitivity experiments with perturbed surface heat fluxes and condensational heat release are performed to investigate their relevance. Yanase *et al.* (2004) showed, using the Meteorological Research Institute Nonhydrological Model (MRI-NHM) with 5 km grid-spacing, that the rapid development of a PL in the Sea of Japan was mainly caused by condensational heating, whereas the surface fluxes maintained the favourable environment for the PL development.

Innes *et al.* (2011), Wagner *et al.* (2011), and Føre and Nordeng (2012) investigated the development of the THORPEX PL with sensitivity experiments. We also perform several sensitivity experiments, some of which are comparable to those in these studies. However, the earlier studies mainly examine the evolution of sea-level pressure of the PL. In this study, we analyse the PL development based on multiple relevant variables, whereby new conclusions are drawn. Additionally, we undertake new experiments to investigate the influence of the sea-surface temperature (SST) on the PL evolution. PLs develop over surfaces of open water, and the sensitivity to SST has previously been tested only in an idealised axisymmetric model (Linders *et al.*, 2011). The investigation of the effect of the SST on the PL development is of high interest for weather prediction, since it elucidates the influence of inaccurate SST fields on the forecast. In NWPs, the SST is typically set constant during the forecast. However, strong cold-air advection during which PLs occur can lead to rapidly varying SSTs (e.g. Sætra *et al.*, 2008), violating a constant SST assumption.

To summarise, the research questions posed in this study are two-fold:

1. How well does the regional NWP model AA capture the development of the well-observed THORPEX PL? How does it perform compared to the global NWP model ECMWF HRES?
2. What are the important physical mechanisms for the development of this PL? How sensitive is this PL to the sea-surface temperature?

The paper is organised as follows. In Section 2, the AA and HRES models, the observational datasets, and the applied methods are presented. Then, research questions 1 and 2 are approached in Sections 3 and 4, respectively. Finally, discussions and conclusions are provided in Section 5.

2 | DATA AND METHODS

In this study, the operational weather-forecast models, AROME-Arctic and ECMWF HRES, are compared and validated against satellite and dropsonde data from the IPY-THORPEX campaign. The models and observational datasets are introduced in the Sections 2.1 to 2.4. In Sections 2.5 to 2.7 we present the techniques applied for validation of the models in Section 3 and the comparison of the sensitivity experiments in Section 4.

2.1 | AROME-Arctic

The AROME model was developed by Météo France (Seity *et al.*, 2011), as part of the Aire Limitée Adaptation Dynamique Développement International (ALADIN) consortium. A collaboration of the ALADIN and
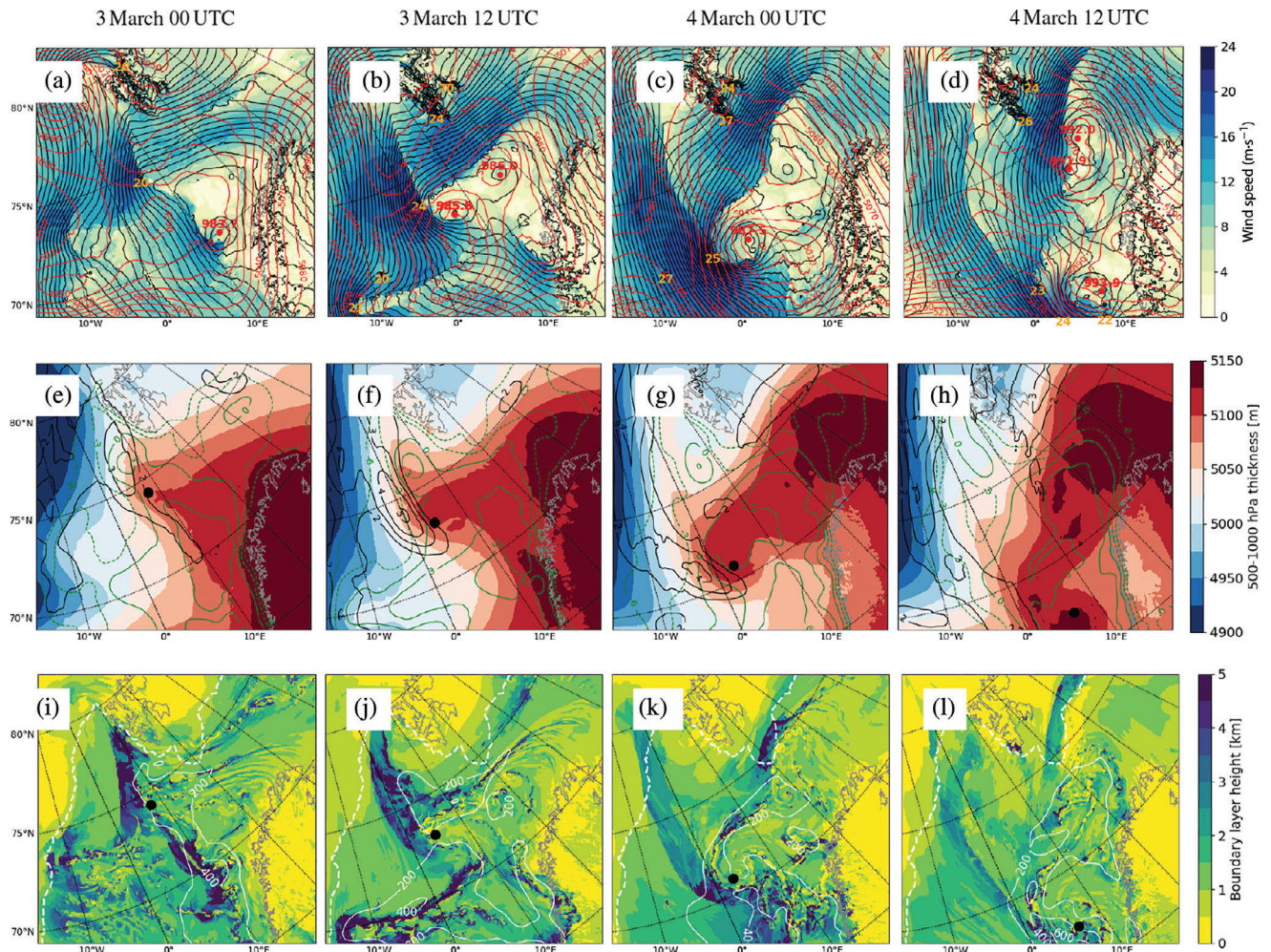
FIGURE 1 Fields from the AROME-Arctic analysis at 0000 and 1200 UTC on 3–4 March 2008. (a–d) horizontal wind speed (colour shading, m·s^{-1}), the sea-level pressure (black contours, spacing 2 hPa), and the 500 hPa geopotential height (m, red contours). Yellow and red numbers denote maximum wind speeds (m·s^{-1}) and the minimum sea-level pressure (hPa), respectively. (e–h) 500–1,000 hPa thickness (m, colour shading), as a measure of the atmospheric temperature, the baroclinicity expressed by $\nabla\theta_{850}$ (K·(100 km)$^{-1}$, black contours), and the static stability expressed by $\theta_{e,SST} - \theta_{e,500}$ (K, green contours), where positive values depict conditionally unstable conditions. (i–l) planetary boundary-layer height (km, colour shading), CAPE (J·kg^{-1}, white contours), and the location of the ice edge (white dashed line). The position of the PL centre is denoted by a black dot in (e–l)

HIRLAM consortia further adapted AROME into the HIRLAM–ALADIN Research on Mesoscale Operational NWP in Euromed (HARMONIE)-AROME model system (Bengtsson *et al.*, 2017).

The HARMONIE-AROME model system is utilised by numerous meteorological services for operational weather forecasts after adaptation for local conditions (e.g. Müller *et al.*, 2017b). MET Norway implemented a configuration of this model system, called AROME-Arctic (AA), for the European Arctic with the centre around Svalbard in November 2015 (Müller *et al.*, 2017a). For experiments in this study, version 40h1.1 of the model system is applied, which was operational for AA from 2016 until early 2019. We display only the southern half of the domain of AA (e.g. Figure 1) since the THORPEX PL evolved in that

area. The full domain is presented in Figure 1 of Müller *et al.* (2017b).

AA has a horizontal model grid-spacing of 2.5 km and 65 vertical hybrid levels, of which 32 are below 3 km. The model includes a non-hydrostatic dynamical core that permits convection. AA uses 3D-Var upper-air data assimilation of conventional and satellite observations and optimal interpolation of near-surface temperature, humidity and snow depth, both within a 3 hr cycle. Every hour, it obtains lateral and upper-boundary data from ECMWF HRES, which is presented in the next subsection. Operationally, AA retrieves data from the HRES forecast starting 12 hr earlier, because the recent HRES version is still in production. Since we reproduce an old case, we utilise the HRES forecast with the same initialisation time as the

TABLE 1 List of the performed experiments with AROME-Arctic and the main results

Name	Description	Main result
CTR = SIM-03-00	Control run; start at 0000 UTC on 3 March 2008.	PL well simulated, especially within first 24 hr of the simulation.
SIM-"day"-"hour"	Simulations starting "day" March "hour" UTC.	Realistic short-term forecasts. After 24 hr the simulations deviate.
noFLX	No turbulent heat fluxes in the domain.	The PL "consumes" the baroclinicity and decays thereafter.
noFLX-A	No turbulent fluxes in limited area (box in Figure 10e).	The local fluxes around the PL centre are most important.
2FLX	Doubled turbulent fluxes in the bulk scheme.	The convective mature stage develops considerably stronger.
noQH	No latent heat flux in the domain.	Initially comparable to CTR; no intensification in the mature stage.
noTH	No sensible heat flux in the domain.	The vortex develops weaker; the PL develops no mature stage.
noCond	No condensational heat release in the domain.	Initially reduced baroclinicity; the PL develops no mature stage.
+2/4/6 SST	Sea-surface temperature increased by 2, 4, 6 °C.	Faster and enhanced PL development; two cores develop.
−2/4/6 SST	Sea-surface temperature reduced by 2, 4, 6 °C.	Suppressed PL development after the baroclinicity is "consumed."

Note: The sensitivity experiments are started at the same time as the control run.

AA simulation. This removes differences between the AA and HRES simulations with same initialization time which originate from old boundary data of AA.

The PL developed in the morning of 3 March 2008 to the south of Svalbard and made landfall in the afternoon of 4 March 2008 in central Norway. In order to obtain accurate initial conditions for the AA simulations, a spin-up phase is started at 0000 UTC on 1 March 2008 from interpolation of the ECMWF HRES analysis. After that the model is updated 3-hourly with assimilation of observations. The main AA simulation, also referred to as the control run (CTR) and SIM-03-00, is initiated from the cycle at 0000 UTC on 3 March, just before the THORPEX PL developed, and forecasts for 48 hr until 0000 UTC on 5 March. Similar forecasts are also initiated from the consecutive cycles at 0600, 1200 and 1800 UTC on 3 and 4 March and referred to as SIM-"day"-"hour," where "day" and "hour" indicate the time of initialisation. These simulations are used for the validation of the forecast performance of the model (Section 3). In order to investigate different physical mechanisms, several sensitivity experiments are performed, beginning at the same time as CTR (Section 4). The different experiments are briefly summarised in Table 1. In experiments where the surface flux components are investigated (e.g. noTH, noQH, noFLX and 2FLX), an artificial factor was implemented into the bulk formula.

2.2 | ECMWF HRES

ECMWF HRES produces a global weather forecast for 10 days into the future. In this study, data from the model that was in operation in March 2008, is used. It is based on the ECMWF Integrated Forecast System (IFS) cycle 32r3 with a horizontal spectral resolution of T799, corresponding to a grid-spacing of about 25 km, and includes 91 vertical levels (ECMWF, 2018). The model runs twice a day, starting from 0000 and 1200 UTC. The initial state is updated by 4D-VAR data assimilation with a 12 hr window. In this study, HRES simulations from 2–4 March 2008 are compared to AA.

2.3 | Satellite data

For the qualitative validation of the PL, the model products are validated against different satellite retrievals. The U.S. National Oceanic and Atmospheric Administration (NOAA) Advanced Very High Resolution Radiometer (AVHRR) measures radiation emitted from Earth. Channel 4 retrieves infrared radiation within the spectral band of 10.3–11.3 μm, from which the emission temperature can be determined. The latter is equivalent to the cloud-top temperature in the case of cloud cover, and to the surface temperature otherwise.

QUIKSCAT (U.S. Quick Scatterometer mission carrying the SeaWinds scatterometer) is a specialised microwave radar that measures the near-surface wind vector on a swath width of 1,800 km over sea surfaces under all weather conditions (Verhoef *et al.*, 2016). The instrument measured the wind speed with a horizontal resolution of 25 km and an accuracy of 2 m·s^{-1} between June 1999 and November 2009.

2.4 | IPY-THORPEX dropsondes

The IPY-THORPEX campaign included a total of 12 flight missions between 27 February and 17 March 2008 with a total of 150 released dropsondes (Kristjánsson *et al.*, 2011). Three of the flight missions focused on the PL investigated here, with 20, 15, and 20 released dropsondes, respectively. The sondes were dropped from an altitude of about 7 km and measured the pressure, temperature, horizontal wind and relative humidity with an accuracy of 1 hPa, 0.1 K, 0.5 m·s^{-1} and 5%, respectively.

2.5 | Verification techniques

Simple error statistics – the BIAS and mean absolute error (MAE) – are calculated by comparison of the model data to the dropsondes released in the THORPEX flights. In order to exclude effects assigned to the high local variability of AA, the local average in a circle with a radius of 12.5 km, approximately the grid-spacing of HRES, is calculated for AA and presented as AA-avg. However, traditional metrics are sensitive to exact matches of observations and simulations (Ebert, 2008). Since models can have a high quality without capturing the exact location of meteorological features, spatial verification methods have been introduced for model evaluation. Different types, such as scale separation, object-oriented, field deformation and "fuzzy" verification techniques have been developed (Gilleland *et al.*, 2010).

The former three approaches are normally applied to gridded observation data, often for precipitation verification (e.g. Gilleland *et al.* (2009). This study utilises gridded observation data from satellites for infrared radiation and scatterometer wind fields. However, examples of spatial verification with these fields are rare for case-studies. Alternatively, some "fuzzy" verification techniques are commonly applied to point observations, such as the dropsondes. Fuzzy verification utilises a spatial window surrounding the location of the observation. Within this window, the data can be treated in various ways (Ebert, 2008). Here, a simple approach of Atger (2001) is applied: for a given threshold, if both the observation and at least one

grid cell within the window satisfy the threshold, a hit is obtained. Following this logic, a contingency table of hits, misses, false alarms and correct rejections can be derived which is utilised for the calculation of a skill score. Following Ebert (2008), the Hanssen and Kuipers (HK) score is calculated as:

$$
\begin{aligned}
HK &= \text{hit rate } - \text{ false alarm rate} \\
&= \frac{\text{hits}}{\text{hits} + \text{misses}} - \frac{\text{false alarm}}{\text{false alarm} + \text{correct rejection}}.
\end{aligned}
$$

A multi-event contingency table is derived by varying the threshold and the radius of the window size (scale) and displaying the result in a two-dimensional table (Ebert, 2008). Also, the equitable threat score (ETS) is applied and gives qualitatively similar results and is therefore not displayed here.

2.6 | Tracking of the polar low centre

Both for the model forecast validation (Section 3.5) and the sensitivity experiments (Section 4), the propagation of the PL is analysed. An automatic tracking procedure is applied to detect the system objectively. It consists of three steps:

- Local maxima of the filtered relative vorticity at 850 hPa are labelled as cyclone centres.
- Consecutive cyclone centres that propagated at less than 130 km·h^{-1} are merged in time to their nearest neighbour.
- The THORPEX PL is detected as the cyclone centres that propagate through the box bounded by 65°N–71°N and 5°W–10°E between hours 20 and 30 of the experiment. Satellite images reveal that the THORPEX PL was the only cyclonic system propagating through that box during that time.

Comparison of the retrieved tracks to the location of the THORPEX PL in satellite images reveals that this tracking procedure is sufficient. The detection proves insensitive to the pressure level of the vorticity, as long as the level is chosen from the lower troposphere (below 700 hPa). The maximum propagation speed may appear to be high, but was chosen because the THORPEX PL moved with a speed of up to 90 km·h^{-1} at the later stages (Wagner *et al.*, 2011), and because the centre of the PL, recognised by the applied detection algorithm, was adjusted to the location of strongest vorticity.

A Gaussian filter is applied on the relative vorticity within a radius of 100 km, cutting at one standard deviation. The size of the radius was employed after the following consideration. The smaller the filter radius, the

more individual convective cells are recognised. The larger the filter radius, the more circulation cells, including multiple PLs, are merged. In some simulations, the PL tends to split into a dual PL after more than 24 hr of forecast time (e.g. Figure 2l). For the comparison applied here, it was considered most instructive to summarise the characteristics of the PL as a single system. However, in simulations with a pronounced division of the PL centre (e.g. +6 SST in Figure 10c below), an individual investigation of the centres is insightful. The chosen filter radius of 100 km takes this into account. In cases of multiple centres within a small distance, the procedure detects an intermediate position between the centres.

2.7 | Variables in the vicinity of the polar low

After detection of the THORPEX PL, several variables are computed in order to analyse the evolution. The strength of the THORPEX PL is measured in three ways:

(a) the filtered relative vorticity at 850 hPa in the centre,

(b) the maximum wind speed at 10 m within 400 km around the centre, and

(c) the minimum sea-level pressure (SLP) within 100 km of the centre.

The location of the minimum SLP and the vorticity maximum do not coincide perfectly. In some cases, the PL does not even have a well-defined local minimum in SLP. The near-surface wind speed is influenced by the strength of both the PL and the synoptic-scale CAO. Stoll *et al.* (2018) show that the wind speed and SLP are both less effective criteria for measuring the strength of PLs than is the vorticity. However, the near-surface wind speed is likely the most relevant variable for human activities. The SLP is widely utilised as an intensity measure (e.g. Føre *et al.*, 2011), but in the present study it is demonstrated to be of little value.

The roles of the three diabatic components, the sensible and latent surface heat flux and the latent heat release by condensation, are compared. The latent heat release by condensation is deduced from the precipitation rate by using the specific latent heat for deposition, since the precipitation is almost purely in the solid phase. The mean in each of the three diabatic components within a circle of radius 300 km around the PL centre is computed in order to compare their strengths. This is necessary since the condensational heating occurs locally in convective cells, whereas surface heat fluxes are more continuous, mainly in regions of strong near-surface winds.

The gradient in the potential temperature at 850 hPa ($\nabla\theta_{850}$) is used to investigate the baroclinic development of the PL, and in the following is referred to as the baroclinicity. A Gaussian filter with 100 km radius is applied to θ_{850} prior to the calculation of the gradient in order to detect meso-α-scale baroclinic zones and to exclude temperature variations caused by small-scale convective cells. The maximum baroclinicity within a distance of 400 km of the PL centre is computed for the analysis of the evolution of the PL. Also, horizontal fields of the planetary boundary-layer height are presented. This variable is computed by AA as the lowest atmospheric level where turbulent kinetic energy is below $0.01\,\mathrm{m^2 \cdot s^{-2}}$.

Conclusions presented in the following were tested and confirmed to be insensitive to variations in the above-mentioned length-scales.

3 | MODEL VALIDATION

In this section, the capability of the weather-prediction models AA and HRES for simulating the THORPEX PL are evaluated. First, the development of the THORPEX PL in the AA simulations is described and qualitatively evaluated against satellite images (Section 3.1). Then, the representation of the PL is qualitatively (Section 3.3) and quantitatively (Section 3.4) compared between AA and HRES. Finally, the forecast qualities of the two models are compared (Section 3.5).

3.1 | Evolution of the THORPEX polar low

The evolution of the THORPEX PL is described by investigating model fields from the analysis of AA (Figure 1), and additionally by comparing the pseudo-satellite images from the analysis of AA (second column of Figure 2) to actual satellite retrievals (first column of Figure 2). The development of the THORPEX PL is also described in Føre *et al.* (2011) and Wagner *et al.* (2011). Here, a somewhat different perspective is presented by the inclusion of additional fields, such as the baroclinicity, the static stability, the planetary boundary-layer height and the convective available potential energy.

On 2 March 2008, a synoptic-scale low moved eastward across the Norwegian Sea, causing a CAO to its western side. At 0000 UTC on 3 March, the synoptic-scale low was positioned off the coast of Northern Norway (Figure 1a; 70°N, 12°E). On the western flank of the low pressure a frontal zone developed (Figure 1a; 70–78°N, 10°E). The front separated the cold air masses over the Arctic sea ice and warmer air masses over Scandinavia and developed a significant temperature gradient (Figure 1e, black contours). Along the front, the boundary layer was convective and hence reached up to 5 km altitude (Figure 1i). To the
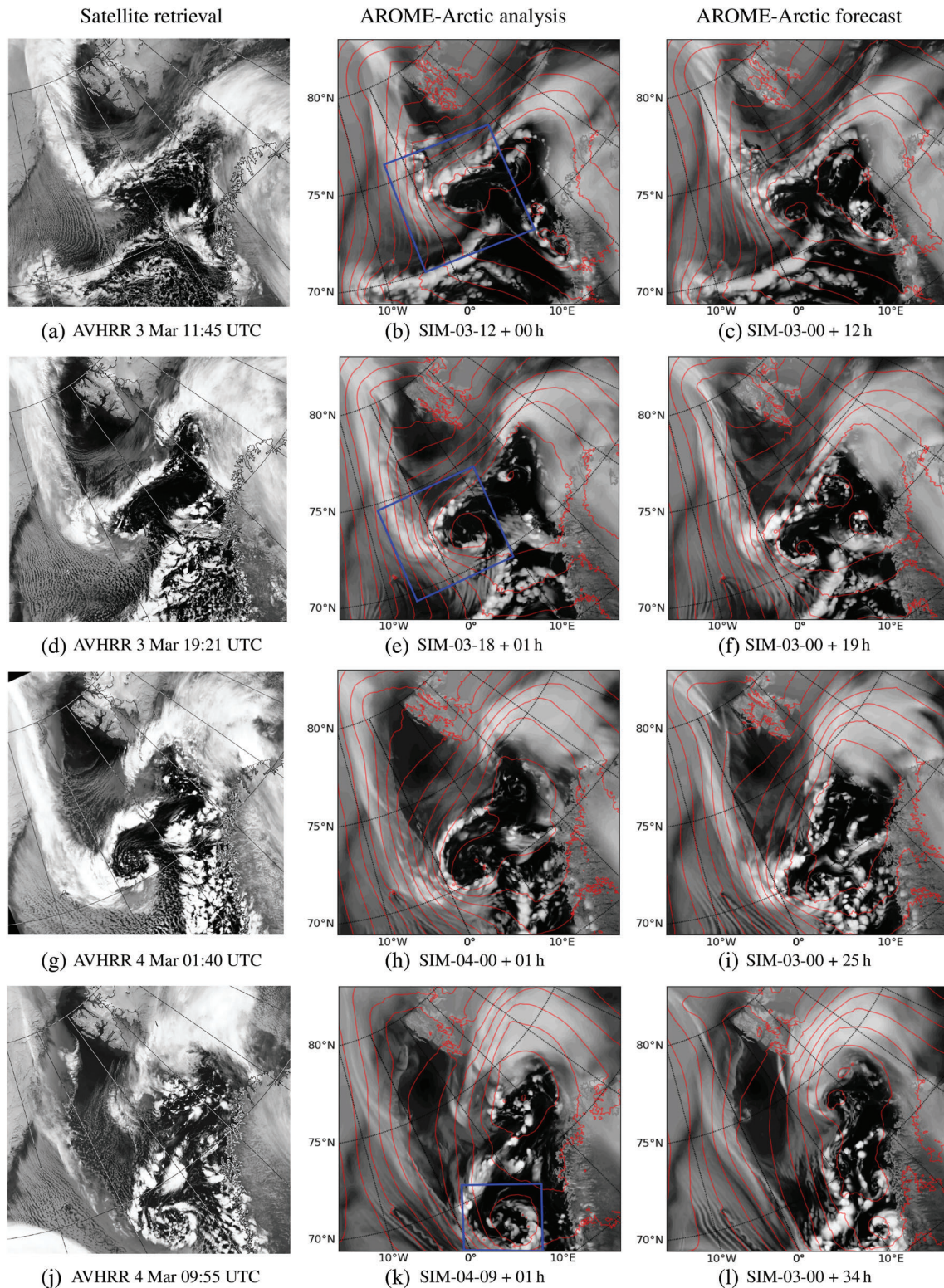
Satellite retrieval AROME-Arctic analysis AROME-Arctic forecast

(a) AVHRR 3 Mar 11:45 UTC   (b) SIM-03-12 + 00 h   (c) SIM-03-00 + 12 h

(d) AVHRR 3 Mar 19:21 UTC   (e) SIM-03-18 + 01 h   (f) SIM-03-00 + 19 h

(g) AVHRR 4 Mar 01:40 UTC   (h) SIM-04-00 + 01 h   (i) SIM-03-00 + 25 h

(j) AVHRR 4 Mar 09:55 UTC   (k) SIM-04-09 + 01 h   (l) SIM-03-00 + 34 h

FIGURE 2   (a, d, g, j) Satellite images displaying the emission temperature, equivalent to the cloud-top temperature obtained by AVHRR channel 4, and retrieved from the NERC satellite retrieving station, Dundee, UK. Pseudo-satellite images expressing the cloud-top temperature from the AROME-Arctic analysis or 1 hr forecast (b, e, h, k), and from the AROME-Arctic forecast starting at 0000 UTC on 3 March 2008 (c, f, i, l). In the latter, the lead time of the simulations is displayed in the sub-caption by "+ h." The red contours in the model fields denote the sea-level pressure with a spacing of 4 hPa. The blue boxes in (b,e,k) show the areas that are presented in Figures S1–S3, respectively

west of the front, the wind was northerly and cold, and to the east, the wind was warmer and easterly (Figure 1a,e). The front propagated westward, and at 1200 UTC on 3 March, it lay along the 2°E meridian to the west of Svalbard (Figure 1b). The satellite and pseudo-satellite images from the model (Figure 2a,b) depict a frontal cloud band. The baroclinicity of the southern part of the frontal zone, which was connected to the synoptic-scale low, decayed in intensity. In contrast, the baroclinicity of the northern part of the front amplified (Figure 1e,f), and along this frontal zone, the PL was initiated.

A secondary convergence zone formed on 3 March along 74°N to the south of Svalbard, caused by easterly winds north and southeasterly winds south of the zone (Figure 2a,b). The PL intensified around noon on 3 March in the baroclinic zone at the intersection point of the two convergence zones (Figure 2b,e). The horizontal temperature gradient increased and was maintained at approximately 5 K per 100 km (Figure 1e,f). The comma-shaped cloud structure, visible until the night of 4 March (Figure 2d,g), with the comma-head to the west of the PL center, indicate a baroclinic intensification of the PL (Yanase and Niino, 2007). The upper-level low is located to the south of the low-level center (Figure 1a,b; 72°N, 5°E). Low-level cold-air advection below the upper-level low amplifies the upper-level disturbance. The upper-level low in turn causes upper-level warm air advection above the surface low, which strengthen the low-level vortex. This is the amplification mechanism of baroclinic instability, characterised by a vertical tilt in the pressure perturbations. 1. 1 In the following, the stage of the PL until 0000 UTC on 4 March is referred to as the initial baroclinic stage.

At the end of the baroclinic stage, at 0000 UTC on 4 March, the PL formed an eye-like cloud structure (Figure 2g) with a warm core, and the baroclinicity decayed (Figure 1g). Also, the SLP and the geopotential height at 500 hPa aligned vertically, an indication of a quasi-barotropic system (Figure 1c). The highest wind speed associated with the PL occurred on the western side of the centre at the edge of the CAO. This region is referred to as the western eye-wall.

On the morning of the 4 March, the PL propagated southeastward into an area that was conditionally unstable for deep convection, indicated by $\theta_{e,\mathrm{SST}} - \theta_{e,500} > 3\,\mathrm{K}$ (Figure 1g) and by CAPE values above $400\,\mathrm{J\,kg^{-1}}$ (Figure 1k). In this environment, the PL intensified further, and strong winds of $25\,\mathrm{m\cdot s^{-1}}$ occurred in the western eye-wall (Figure 1c). The PL developed into a spiral-like system of convective clouds (Figure 2j,k). This cloud signature indicates that convective processes were of major importance for the system (Yanase and Niino, 2007). The time from 0000 UTC on 4 March, where the PL reached the highest intensity, is below referred to as the convective mature stage. Later, it will be shown that latent heat release by condensation was significant at this stage (Figure 12a).

The PL propagated further southeastward along the edge of the domain of AA from 1200 UTC on 4 March and made landfall on the coast of Norway at approximately 65°N around 1800 UTC on 4 March.

3.2 | AROME-Arctic validation against satellite images

The comparison of the satellite images with the pseudo-satellite images produced by AA, both depicted in Figure 2, reveals that the clouds are generally captured well in the AA analysis. Examples are the correct position and structure of the frontal zones and the spiral-form clouds of the PL in the mature stage. The cloud structure appears in balance with the model dynamics at the analysis time of the model. This can be seen by the lack of abrupt changes in the cloud representation within the first hours of the model simulations (not presented in detail, but indicated from a comparison of Figures 2b,c).

AA develops deep convective towers, visible as circular blobs (e.g. Figure 2e around 73°N, 3°E). In the satellite images, deep convection appears less confined and spread over larger areas (e.g. compare Figures 2g,h at 68°N, 10°E), indicating that some deep convection occurs on scales lower than the effective resolution of the model.

In the shallow CAO to the west of the frontal zone along 2°E, the model correctly simulates cloud streets (Figure 2e,h,k lower-left corner). However, the spacing between the cloud streets is about 25 km (10 grid cells, not shown), which is approximately the effective resolution of the model, and which is larger than the observed spacing of about 15 km in the satellite images (Figure 2d,g,j lower-left corner). The satellite images show that the convection in the CAO evolves into shallow convective cells during the night of 4 March (Figure 2g lower left side), whereas AA still simulates cloud streets at this time (Figure 2h). Since cloud streets are favoured over cellular convection when the vertical wind shear is large (Markowski and Richardson, 2010), AA may overestimate the vertical shear in the lowest model levels, which might be caused by an inaccurate boundary-layer parametrization.

Other interesting features captured by AA, which are not connected to the PL development, are lee vortices, visible as wave-breaking-like disturbances in the satellite image, induced by Jan Mayen, an island with a 2.2 km high mountain, located at 8°W, 71°N. These vortices can be observed in the lee of isolated terrain obstacles, when the lower atmosphere is strongly stratified, so that
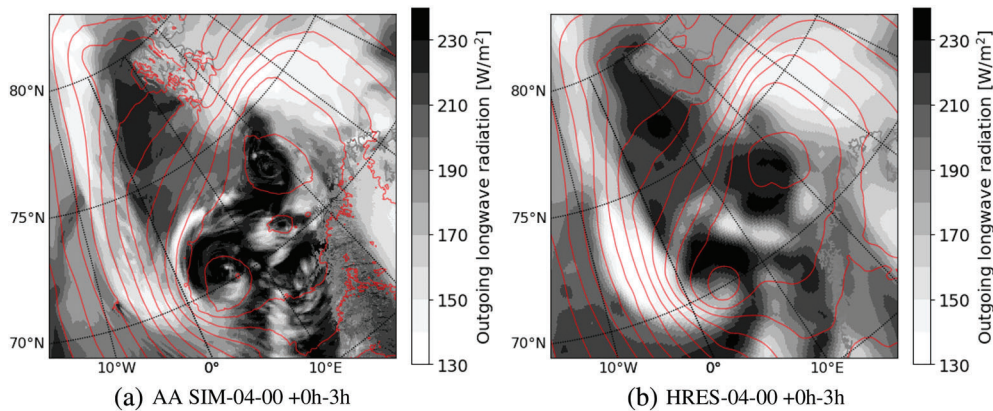
(a) AA SIM-04-00 +0h-3h   (b) HRES-04-00 +0h-3h

FIGURE 3 The outgoing long-wave radiation (shading) from the (a) AROME-Arctic and (b) HRES simulations starting at 0000 UTC on 4 March, averaged over the first 3 hr of the forecasts, to be compared with the satellite image in Figure 2g. Red contours depict the sea-level pressure (spacing 3 hPa) after 3 hr of model integration

the flow has to pass around the obstacle (section 13.3 in Markowski and Richardson, 2010). In the model, the island initialises an oscillation in the cloud street passing the mountain (Figure 2j,k). However, the effective resolution of the model appears to be insufficient to simulate the wave-breaking of the oscillation.

High clouds, connected to the jet stream, which were observed over Northern Scandinavia (see the high gradient in the geopotential height in Figure 1b), are depicted by the model, but more smoothly than observed (e.g. Figure 2a,b right side). AA has only a few model levels above 10 km altitude and is highly steered by HRES at this height. The latter has a model grid-spacing of 25 km and therefore does not resolve fine-scale structures. Locally, deviation in this high-cloud cover can lead to large differences in the local radiative balance (Valkonen *et al.*, 2020). Since the deviation in the high-cloud cover is located more than 500 km to the east of the system it has no influence on the radiative budget of the PL.

3.3 | Qualitative comparison between AROME-Arctic and ECMWF HRES

After having shown a reasonable agreement of the AA analysis to satellite images, in this and the next Section the representation of the Thorpex PL in AA and ECMWF HRESis compared, first qualitatively and then quantitatively.

3.3.1 | Cloud structure

The emission temperature displayed in Figure 2 is not a standard output parameter from ECMWF models. However, the outgoing long-wave radiation (OLR) at the top of the atmosphere can be used instead, because the emission temperature largely determines it. The OLR from the models is typically stored as the accumulation since the start

of the simulation, whereas the pseudo-satellite images, shown earlier, depict instantaneous patterns. Therefore, the OLR from AA (Figure 3a) appears more smooth than the pseudo-satellite image (Figure 2h). Since ECMWF provides the output from HRES at 3 hr intervals, the mean OLR is displayed within 3 hr period of model integration .

The comparison of the OLR between the two NWP models reveals close agreement in the representation of the comma-shaped cloud of the PL (around 72°N, 5°E), other large-scale cloud patterns (e.g. the high clouds over the Barents Sea in the upper right corner of Figure 3a,b), and areas of cloud-free conditions. However, AA better captures the shallow convection in the CAO to the west of the PL (Figure 3 lower-left corner). Also, AA resolves individual convective clouds in much more detail than HRES (e.g. Figure 3a lower edge around 10°E). These clouds that can cause a considerable amount of precipitation.

3.3.2 | Near-surface winds

Despite differences in the cloud cover, the SLP and near-surface wind fields are quite similar in the two models at near-analysis time. Especially in the initial baroclinic stage of the PL, differences in the fields are small (not shown). Also in the convective mature stage around 0600 UTC on 4 March, which is investigated in the following, the pressure field is very similar in the two models (Figure 4d,g); e.g. the synoptic-scale low at 73°N, 20°E has a comparable depths in sea-level pressure. However, the centre of the PL is 1.5 hPa deeper in AA than in HRES. Smoothing of AA to HRES resolution with a Gaussian filter of 12.5 km radius shrinks the difference in the centre pressure to 0.7 hPa. Hence, large parts of the pressure differences are attributed to the small-scale dynamics of AA.

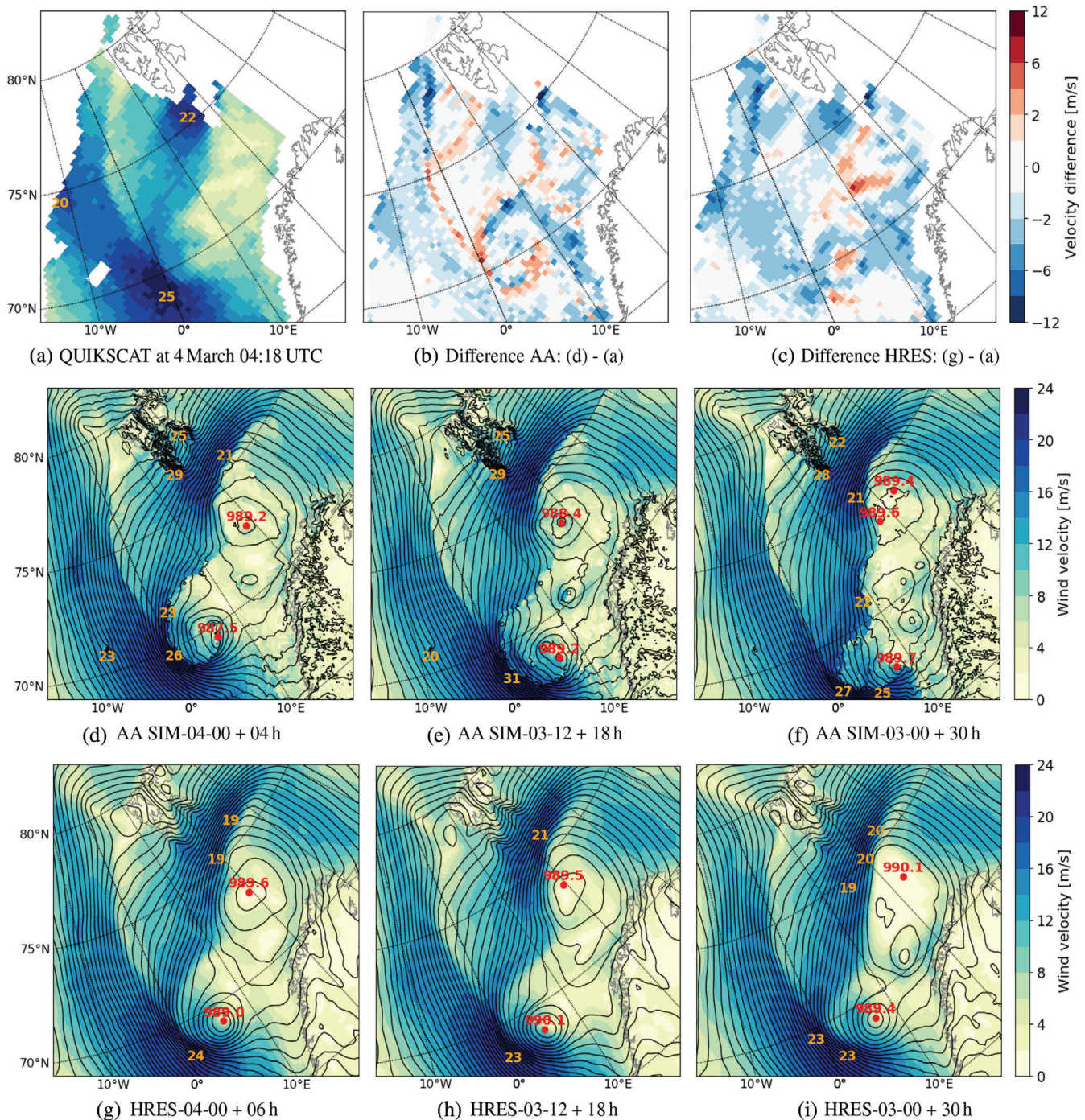AA and HRES compare well to the scatterometer wind field retrieved from QUIKSCAT (Figure 4a–c). The RMSE

FIGURE 4 (a) The 10 m wind speed (m·s⁻¹, colour shading, scale as for (d–i)) retrieved from QUIKSCAT at 0418 UTC on 4 March at the mature stage of the Thorpex PL. (d–i) The 10 m wind speed (colour shading) and sea-level pressure (contours, spacing 1 hPa) of the models AA and HRES for 0600 UTC on 4 March 2008 (but 0400 UTC for (d)). (d, g) show the near-analysis fields, (e, h) the 18 hr and (f, i) 30 hr forecasts from AA and HRES, respectively. (b, c) show the difference between the near-analysis and the QUIKSCAT satellite retrieval in (a). Red dots and numbers denote the local minima in sea-level pressure (hPa), and orange numbers the local maxima in the wind speed

of the near-surface wind fields of both models against QUIKSCAT is 1.8 m·s⁻¹. Comparable results are also found from scatterometer wind retrievals for other times (not shown).

The highest wind speed connected to this PL is observed in the western eye-wall of the PL at the edge

of the shallow CAO. The CAO is associated with the synoptic-scale situation with a low at 73°N, 20°E and both models reproduce this flow (Figure 4). The PL intensifies the flow of the CAO. Both models capture this wind intensification to the southwest of the PL. The maximum wind speed develops slightly stronger in AA (26 m·s⁻¹)

than observed by QUIKSCAT (25 m·s^{-1}), and is slightly weaker in HRES (24 m·s^{-1}). AA captures the wind speed in the shallow CAO (Figure 4b,c between 10 and 0°W) better than HRES, which underestimates the wind speed by approximately 3 m·s^{-1}. This might be attributed to improved low-level dynamics in AA, due to the increased resolution.

The largest deviation of AA from QUIKSCAT occurs at the location of the fronts, which are displaced by around 30 km (Figure 4b, red line near 0°E). The fronts are considerably sharper in AA than in QUIKSCAT. Due to its coarse resolution of 25 km, QUIKSCAT underestimates the gradient of the wind speed. Furevik *et al.* (2015) for example observe a wind gradient of 11 m·s^{-1} over a distance of 1 km in the front of a PL. Hence some of the deviations between AA and QUIKSCAT may not be associated with model deficiencies. No conclusions on wrong model dynamics of AA can be drawn from this comparison. Otherwise, the resolution of HRES appear insufficient to fully capture the flow close to the PL centre: HRES overestimates the wind speed in the almost calm centre by about 3 m·s^{-1} (Figure 4c at 70°N, 4°E) and underestimates the wind around the centre by about 4 m·s^{-1}.

In both models, large differences from QUIKSCAT are observed in the calm sector to the east of the PL (e.g. Figure 4b,c at 70°N, 11°E). The models simulate too week winds in this area which is associated with deep cellular convection (Figure 2g). As mentioned before, none of the models correctly simulate these cells (e.g. Figure 3).

3.3.3 | Qualitative comparison to dropsondes

In order to provide a more detailed validation of the models, comparison with the dropsonde data is performed. This is done qualitatively in this subsection and quantitatively in the next section. First, the qualitative analysis is performed since it highlights the challenge of model verification with the utilised observational dataset.

Examples of horizontal cross-sections of the specific humidity for both models at 850 hPa are depicted in Figure 5a,b at the time of the second THORPEX flight, at the end of the baroclinic stage of the PL. The large-scale humidity field at 850 hPa is similar for AA and HRES, but also considerable differences are recognised (Figure 5a,b). AA shows more moisture than HRES in the baroclinic zone along the 0°meridian. In this zone, the relative humidity is exceeding 90% in AA, whereas HRES rarely simulates values close to saturation.

AA simulates small-scale convective cells (e.g. Figure 5a, around 69°N, 5°E) with the relative humidity

often reaching saturation. This causes a high local variability of the humidity field. HRES reaches near saturation only in the frontal and orographic zones, but has considerably drier conditions in areas of cellular convection (Figure 5b). This arises from the advanced skills of the convection-permitting dynamics of AA.

An enlargement of the central region of the PL simulated by AA, with the observed values of the dropsondes, is presented in Figure 5c. AA shows high local variability within this region. AA and the dropsonde observations appear to have similar values, and they also appear to have a similar spatial variability of the values. Hence it is concluded that AA captures the humidity reasonably well.

In order to validate the 3D structure of the PL in more detail, vertical cross-sections AA in equivalent potential temperature, relative humidity and wind for AA are presented in Figure 6, together with dropsonde data. The cross-section goes through the main baroclinic zone during the second flight between dropsondes 5 and 9 (red line in Figure 5c). In general, AA and the dropsondes agree well on the vertical structure of the meteorological fields and on showing high local variability in the vertical direction.

AA captures the shallow CAO (west of 2°W) with low temperature and increasing humidity from the surface towards the cloud top at approximately 800 hPa (Figure 6a,b). In the baroclinic zone around 0° E, strong temperature gradients are simulated, and the observations approximately agree. The strongest winds of up to 30 m·s^{-1} are measured and simulated in this zone (dropsonde 7) at around 900 hPa (Figure 6c). From the low-level baroclinic zone, the isentropes (here surfaces of constant equivalent potential temperature) tilt towards the west with height in both model and observations. Along this tilt, frontal updraught is simulated, leading to increased relative humidity. Model and observations highlyagree in the frontal dynamics, which are causing the comma-shaped cloud.

To the east of the front, AA simulates strong convective updraughts of the order of 1 m·s^{-1} at 700 hPa between dropsondes 7 and 8, high RH of almost 100% up to 600 hPa and a conditionally unstable situation from the surface to the tropopause (450 hPa). The dropsondes largely agree with this convective behaviour.

Føre *et al.* (2011) argue that this PL is to a large degree forced by upper-level potential vorticity. They partly base their argument on a tropopause downfold, which they observe by interpolatiing dropsondes 5 to 9 of the second flight (their figure 8a). In Figure 6a, the equivalent potential temperature for the same cross-section is displayed. Dropsonde 7 reports higher temperatures than dropsondes 6 and 8 between 700 and 400 hPa. Føre *et al.* (2011) argues that this indicates the tropopause downfold. However, dropsonde 7 is located close to the warm core of the
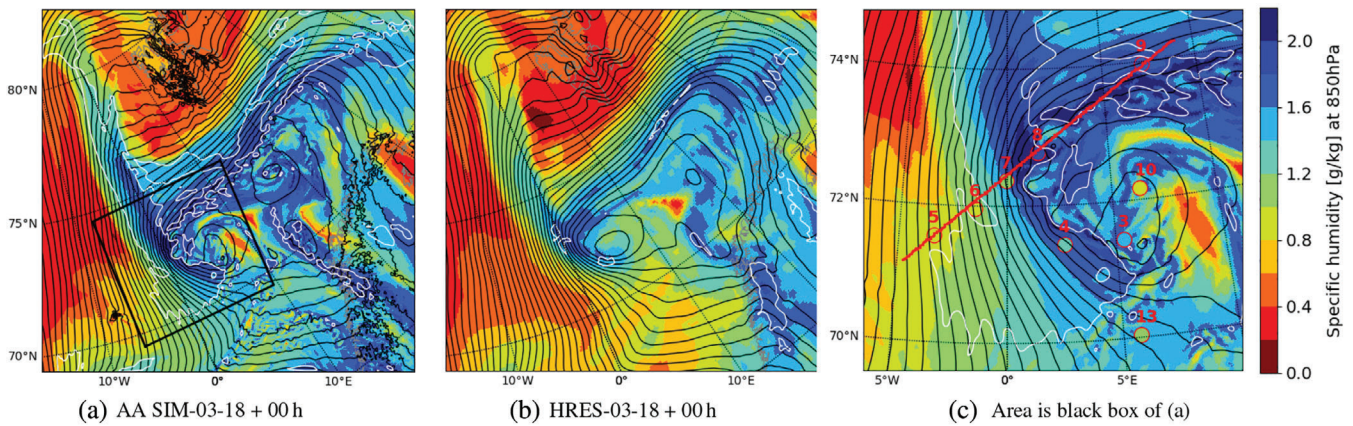
(a) AA SIM-03-18 + 00 h (b) HRES-03-18 + 00 h (c) Area is black box of (a)

FIGURE 5 The specific humidity at 850 hPa (colour shading), sea-level pressure (black contours, spacing 1 hPa) and relative humidity at 850 hPa (white contour at 90%) from (a) AROME-Arctic and (b) ECMWF HRES analysis at 1800 UTC on 3 March 2008. (c) Magnification of the area indicated by the black box in (a). In red circles the observations from the dropsondes released during the second THORPEX flight are depicted using the same colour code. Red numbers label the dropsondes, and the red line indicates the location of the cross-section presented in Figure 6
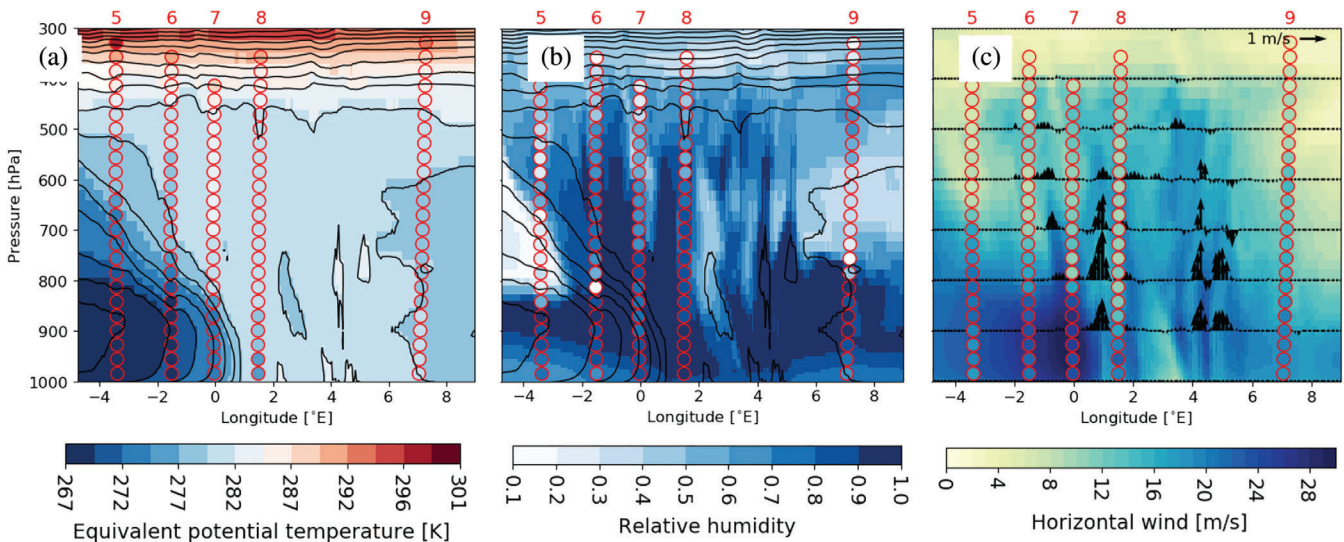

FIGURE 6 Vertical cross-section from the AA analysis at 1800 UTC on 3 March, along the line in Figure 5a during the second flight showing in colour shading (a) equivalent potential temperature), (b) relative humidity and (c) horizontal wind speed, all with data from the dropsondes in red circles. The black contours in (a,b) also depict the equivalent potential temperature with 2 K spacing. In (c) the black vertical arrows display the simulated vertical velocity. The numbers at the top label the dropsondes

PL in a convectively active region. Hence, the increased temperature for this dropsonde might be caused by adiabatic warming in the downdraught of a deep convective cell, i.e., a local tropospheric circulation not affecting the tropopause. The lidar profiles presented in figure 8 of Wagner *et al.* (2011) support this argument. An interpolation of the dropsonde data, as applied in Føre *et al.* (2011), can be misleading since it does not consider the high local variability, and the spatial extent of the interpolated values is easily exaggerated. Also, AA does not show a signal of a tropopause downfolding at the time of the

second flight but, as discussed above, the occurrence of this downfolding during the second flight is questioned here.

More horizontal cross-sections of the two models are presented in Supporting Figures S1–S4. Simulated variables (potential temperature, relative humidity and the horizontal wind velocity at 950, 850, 700 and 500 hPa) from the two models are compared to the corresponding dropsonde data of the three flights. The conclusion is qualitatively the same as for the humidity. Both models simulate the 3D structure of the PL reasonably well.

TABLE 2 Error statistics of the near-analysis time of AROME-Arctic and ECMWF HRES at different pressure levels compared to all the dropsondes released during the three flights

Level (hPa)	Model	T_{BIAS} (K)	T_{MAE} (K)	RH_{BIAS} (%)	RH_{MAE} (%)	U_{BIAS} (m·s⁻¹)	U_{MAE} (m·s⁻¹)
950	AA	0.47	0.76	–4.1	8.4	–0.55	2.76
	AA-avg	0.49	0.76	–4.4	7.9	–0.46	2.38
	HRES	–0.37	0.84	–4.7	9.0	0.11	2.23
850	AA	–0.22	0.64	7.4	14.7	–0.01	2.54
	AA-avg	–0.24	0.61	9.4	14.3	0.10	2.33
	HRES	–0.06	0.67	–7.4	15.9	–0.66	1.91
700	AA	–0.20	0.35	–1.1	19.1	–0.15	2.28
	AA-avg	–0.18	0.35	–1.2	17.0	–0.04	2.30
	HRES	–0.24	0.47	–9.1	15.5	–0.02	1.96
500	AA	–0.05	0.31	–0.9	14.5	0.14	2.03
	AA-avg	–0.05	0.32	–0.6	13.7	0.13	1.96
	HRES	–0.05	0.29	6.3	12.2	–0.18	1.97

Note: The BIAS and mean absolute error (MAE) in temperature (*T*), relative humidity (*RH*) and horizontal wind speed (*U*) are shown. AA-avg gives the values of AROME-Arctic at an approximate resolution of HRES by taking the local average within a radius of 12.5 km around the dropsonde. The model fields in these meteorological fields and the measured values by the dropsondes are depicted in Figures S1–S4.

AA shows much higher local variability than HRES in the potential temperature and relative humidity (Figures S2 and S4). High local variability was also observed in the lidar profile obtained by the aircraft passing the THORPEX low (Wagner *et al.*, 2011). Hence this variability appears realistic.

3.4 | Quantitative comparison between AROME-Arctic and ECMWF HRES

In the following, AA and ECMWF HRES are compared to the dropsondes released during the THORPEX flight campaign.

3.4.1 | Statistical scores as compared to dropsondes

In the previous sections, different fields were compared by visual inspection. Now, error statistics, such as the BIAS and mean absolute error (MAE), obtained by comparison to all the dropsondes released in the three THORPEX flights are compared for the model products (Table 2).

The MAE is in general about the same for AA at near-analysis time as for HRES. AA and HRES perform approximately equally well for the compared variables (temperature, horizontal wind speed and relative humidity) at different pressurelevels. Smoothing the AA data by applying a local average in a circle of radius

12.5 km (approximately the grid-spacing of HRES), slightly improves the MAE for AA, especially in the relative humidity and wind speed. However the skill is still similar to HRES.

The high variability of the meteorological fields makes the objective model validation challenging. For some dropsondes, model and observation are considerably different, e.g. Figure 5 dropsondes 4 and 7. For these locations, the model simulates high variability, and a small displacement creates large differences in the values. For example, the convective cells and the frontal zone of the PL are observed to be slightly displaced in AA. In classical error scores, such as the mean absolute error, the displacement of a correctly simulated feature is penalised twice: firstly since the feature is not captured at the correct location and secondly since it is simulated at a wrong location. Hence, the error statistics are weaker than if the feature had not been present in the model at all. Smoothing of AA corrects for some of this problem, as the error scores improve. However, this also degrades some of the skills of AA to simulate local extreme values. Therefore a spatial verification technique of the two models is applied and is now presented.

3.4.2 | Fuzzy verification

A "fuzzy" verification technique, which relaxes the requirement for exact collocation of observations and model simulations, is employed. Multi-event contingency

tables (Ebert, 2008) utilising the Hanssen and Kuipers (HK) score are displayed in Figure 7 for AA and HRES simulations compared to dropsonde observations from all three flights.

AA generally has highest skills on a scale of 10–20 km (Figure 7a,d), whereas HRES performs best over scales of 40–80 km (Figure 7b,e). At large scales, the models lose accuracy since the False Alarm Rate becomes as large as the Hit Rate. AA loses accuracy at smaller scales (40–80 km) than HRES (160–320 km), since the high local variability of AA generates lots of false alarms. HRES has higher skills at larger scales since the method considers displacement of features but the false alarms do not increase considerably due to a low local variability of HRES as compared to AA (e.g. compare Figure 5a,b). For the displayed fields, the skill score of AA improves slightly (around 0.1) if AA is smoothed to the HRES grid-spacing since the False Alarm Rate is reduced (not shown).

For the relative humidity, AA performs better than HRES for a scale of 10 km (Figure 7c). This indicates that AA has improved the representation of convective cells relative to HRES. Also in the wind speed, AA has very high skills at small scales ($\leq$ 20 km) and large intensities ($\geq$ 20 m·s^{-1}, Figure 7d) and is considerably advanced compared to HRES (Figure 7f). This means that AA considerably improves the capture of local extreme winds.

The fuzzy verification gives some indication of the strength and weaknesses of the models. However, it does not reveal which of the observed weather features are correctly reproduced by the models. Therefore, the qualitative validation that was previously presented is of importance.

3.4.3 | Comparison of vertical profiles

Now, the average vertical profiles of AA and HRES analysis are validated against the dropsondes released in each of the three THORPEX flights independently. The averaging is expected to correct for some of the random displacement errors between observation and models.

The average of the vertical profiles of each flight in potential temperature, relative humidity and the horizontal wind of the dropsondes and the corresponding AA and HRES grid cells containing the dropsonde is presented in Figure 8a,c,e. Figure 8b,d,f presents the BIAS and MAE of the profiles. In general, the analyses of AA and HRES agree reasonably well with the dropsonde data.

For the initial stages of the PL during the first two flights, the highest wind speeds are both observed and modelled at low levels, and the wind speed decays towards mid-levels. This wind profile is a signature of

a reverse-shear baroclinic system with a low-level jet (Terpstra *et al.*, 2016). At the mature stage (Flight 3) the wind speed and direction are almost constant in the vertical, an indication that the system is quasi-barotropic. The models capture this behaviour and do not show significant differences from the dropsondes.

For temperature, AA is warmer than the dropsondes close to the surface by 0.6–1.3 K during the first two flights. The near-surface temperature BIAS is almost as large as the MAE, meaning that the model is too warm at most dropsonde locations. This is likely attributed to overestimated surface sensible heat fluxes in the model, which might be caused by a SST warm BIAS. The strong and cold winds are cooling the sea surface, whereas in the model the SST is fixed during the simulation and only updated once a day in model cycles starting at 0000 UTC. This delayed update of the SST can cause a near-surface warm BIAS in the model. In Section 4 it is found that the PL development is quite sensitive to SST perturbations.

At approximately 850–800 hPa, this warm BIAS vanishes. Hence, on average AA is more unstable in the boundary layer than is indicated by the observations, which is also the case during the third flight (Figure 8b,d,f). Also, HRES appears to be more unstable below 900 hPa.

In terms of relative humidity, AA is on average too dry close to the surface and too moist around 800 hPa at the first two flights. This humidity profile indicates that AA overestimates shallow convection. The reduced stability in AA might explain the exaggerated convection. The near-surface dry BIAS of AA likely leads to overestimated surface latent-heat fluxes. During flight 3 both AA and HRES are considerably drier than the dropsondes through the whole troposphere, mainly around 700 hPa where the relative humidity is on average 25% lower than observed. This indicates that deep convection, different from shallow convection, is under-represented in the models. It is possible that AA would benefit from a deep convection parametrization.

3.5 | Forecast error growth

Until now, the analysis times of AA and HRES have been validated. Now short-term forecasts of the two weather-forecasting models are compared for this PL.

In the third column of Figure 2, the pseudo-satellite images of the AA simulations initiating at 0000 UTC on 3 March (SIM-03-00) are presented in order to validate the forecast quality. Also, simulations initiated at earlier and later times are compared to satellite images, but not presented here. The clouds are in general captured well by
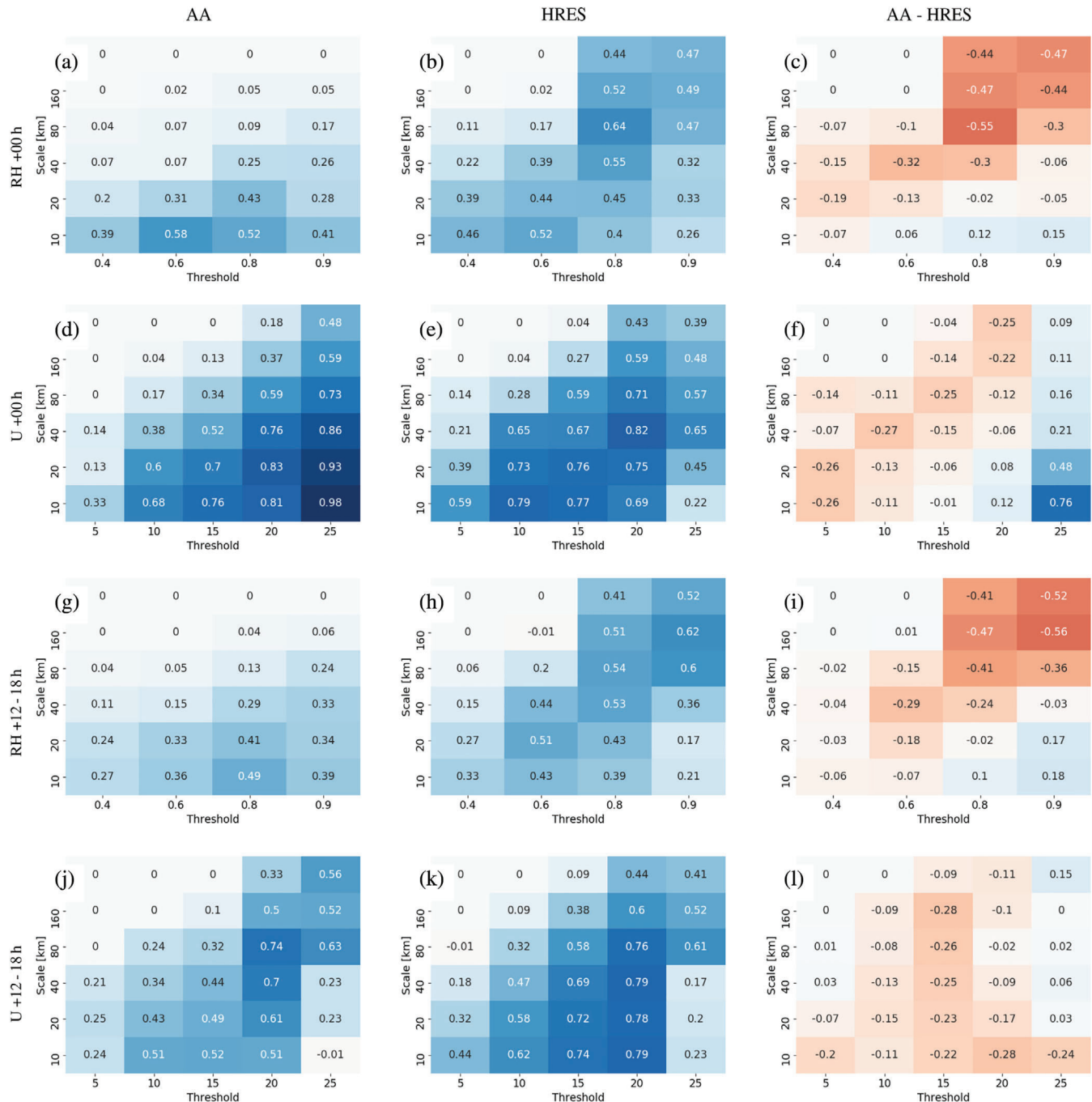
FIGURE 7 The multi-event contingency table (Atger, 2001) with the Hanssen and Kuipers (HK) score calculated for (a, d, g, j) AROME-Arctic and (b, e, h, k) HRES simulations at near-analysis time (+00 hr) and for short forecasts (+12–18 hr) compared to dropsonde observations from all three flights at pressure levels 950, 850, 700 and 500 hPa for (a, b, g, h) relative humidity (RH) and (d, e, j, k) wind speed (U). The HK score (Hit Rate minus False Alarm Rate) is 1 for a perfect model, 0 for no skill and can take negative values for higher False Alarm Rates than Hit Rates. The threshold defines the level above which an observion , or the simulated value at the location of the observation, is a hit, for RH in % and for U in m·s^{-1}. The scale defines the radius within which a hit is searched in the models. (c, f, i,l) show the differences between the HK scores of AA and HRES

AA, both in structure and placement within the first 24 hr of the simulation. For longer forecast times, the evolution of the PL starts to deviate from the observations. On the morning of 4 March, the PL is observed to be a singular spiral-like system (Figure 2j). In contrast, after 30 hr of model integration in SIM-03-00, the PL has divided into two centres (Figure 2l), a leading one (at 69°N, 7°E) and one at the intersection point of the fronts (68°N, 0°E).
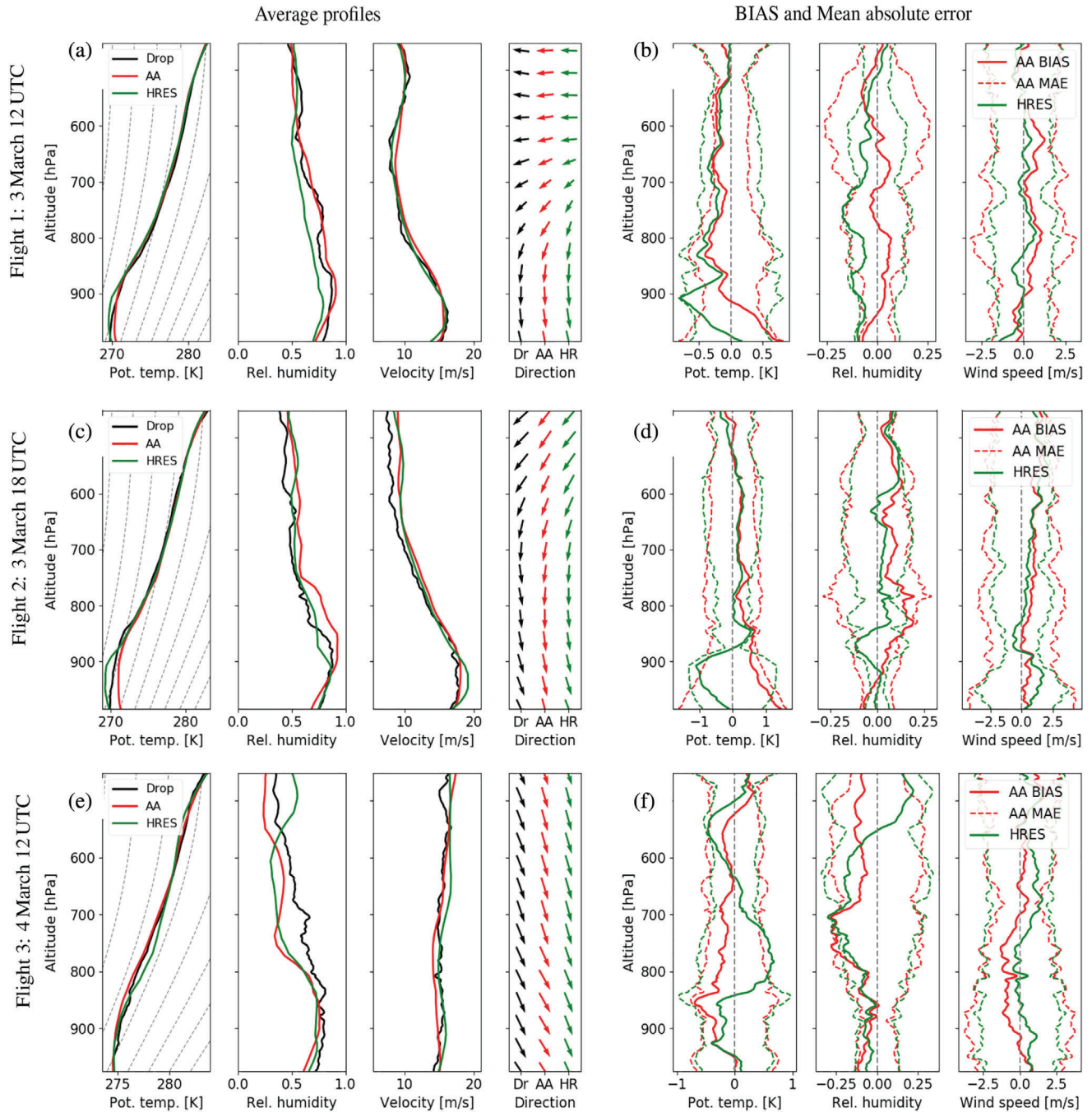
FIGURE 8 (a, c, e) Average of the potential temperature, relative humidity, horizontal wind speed and wind direction from the THORPEX dropsondes (black) released during the three flights, and from the analysis of the AA (red) and HRES (green) at grid cells containing the dropsonde at the corresponding time. The area within which the dropsondes are released are displayed by blue boxes in Figure 2. The position and measured values of the dropsondes with simulated model fields for some vertical levels are also shown in Figures S1–S4. In the potential temperature plots, moist adiabats are displayed with grey dashed lines. (b, d, f) The BIAS (solid line) and mean absolute error (MAE, dashed line) for AA (red) and HRES (green) compared to the dropsondes released during each of the flights shown in the first column. Also −MAE is added to highlight when a negative BIAS explains the total MAE

The development of two centres originates within the domain. Already after 19 hr of integration (Figure 2f), a false leading centre with a closed isobar develops around 71°N, 5°E ahead of the frontal intersection at 72°N, 0°E, where the observed PL develops.

Tracks of the PL centre for different AA simulations are presented in Figure 9a. As described in Section 2.6, the detection of the PL centre is constrained to condense the system to one if the separation of the centres is small enough. In the simulation starting 12 hr earlier, SIM-02-12,
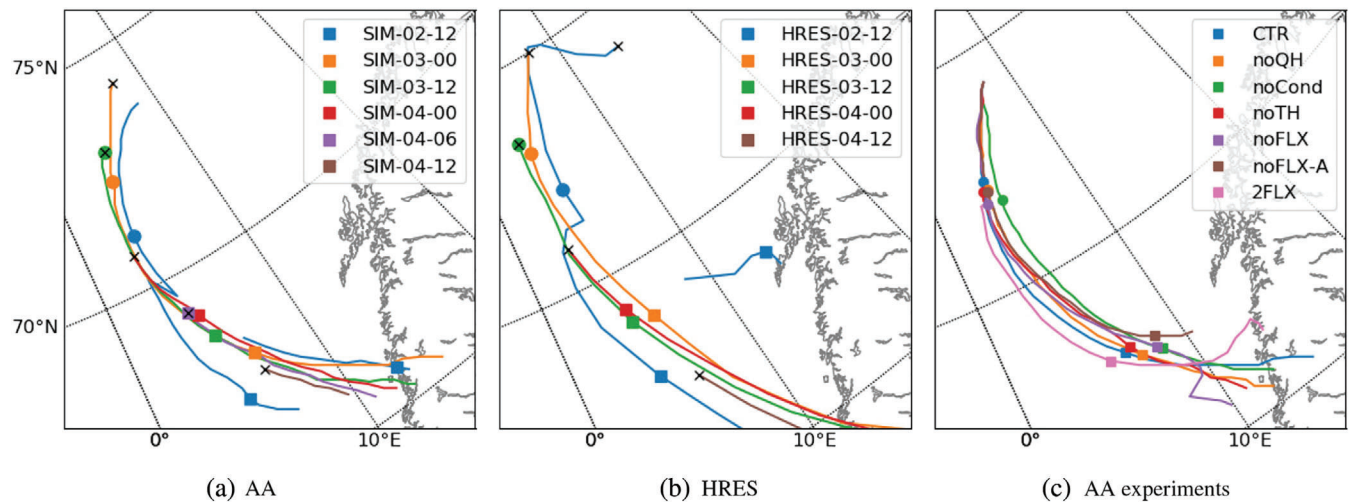
FIGURE 9 The tracks of the PL centre, defined by the local maxima in the filtered vorticity, for simulations at different initial times for (a) AROME-Arctic and (b) ECMWF HRES. Coloured circles and boxes denotes the location of each centre at 1200 UTC on 3 March and 0600 UTC on 4 March, respectively. The location at analysis times is indicated by black crosses. Note that the tracks are influenced by the boundary of the domain for the final time steps of the PL in AA. (c) depicts the tracks of the different sensitivity experiments performed with AA

the split of the PL is considerably more pronounced than in SIM-03-00, and the centre at the frontal intersection moves out of the domain (Figure 9a; the track of SIM-02-12 disappears at 66°N, 7°E since the vorticity centre leaves the domain). A split of the PL develops also in HRES-02-12 (Figure 9b) which shows that this erroneous development is an artefact across models. However, HRES-03-00, differently from SIM-03-00, correctly simulates one PL centre, and predicts its location quite accurately for the mature stage. AA generally overestimates the propagation speed of the PL in the forecasts. Already at 1200 UTC on 3 March, AA forecasts the PL considerably further to the south than observed (Figure 9a; compare orange and blue circles to the green circle with the "x", displaying the analysis time). The displacement grows until 0600 UTC on 4 March, which means that AA overestimates the propagation speed of the system. The sensitivity experiments presented in the next section reveal that suppressing condensational heat release increases the propagation speed of this PL (Figure 9c), even though it weakens the large-scale flow (Figure 10i). More discussion about this is given in Section 4.6. Hence, the faster propagation of the PL in the AA forecast might be caused by erroneous representation of convective processes, as discussed earlier.

The spatial verification of AA and HRES against dropsondes is also applied to 12–18 hr forecasts (Figure 7g–l). It reveals that AA loses some of its skill in the small scales after short forecast times. This is most pronounced in the wind speed for high values (compare Figure 7d,j), but also appears in the small scales in the humidity (compare Figure 7a,g). HRES appears not to

lose skill in the short-term forecast when compared to the analysis. At the analysis time, AA is considerably improved over HRES for high wind speed at small scales (Figure 7f). AA loses these advantages already after short forecast times of 12–18 hrs (Figure 7l). It appears that the error growth is faster for AA than for HRES. Some of this error growth is attributed to a larger displacement of the PL for short-term forecasts by AA than by HRES (compare green and red point in Figure 9a,b).

In Figure 4e,f,h,i, near-surface wind fields in the 18 and 30 hr forecasts of AA and HRES are depicted for the mature stage of the PL. The development of the PL is quite different for the AA forecasts when compared to the near-analysis. The 18 hr forecast simulates maximum wind speeds in the vicinity of the PL of up to 31 m·s^{-1} (Figure 4e), whereas 25 m·s^{-1} is observed by the QUIKSCAT instrument (Figure 4a). The 30 hr forecast experiences the separation and the overestimated propagation speed of the centre (Figure 4f). In contrast, HRES forecasts appear to differ considerably less for different lead times (Figure 4g–i). This is in accordance with Køltzow *et al.* (2019), who find that model errors grow faster for near-surface fields in high-resolution models, such as AA, than in HRES. An explanation could be given by the conceptual model of three-stage error growth suggested by Zhang *et al.* (2007): (a) convective instability causes fast error growth on small scales, which saturates within approximately 1 hr due to the complete displacement of convective cells, (b) the errors expand in space and influence the large-scale balanced flow, and (c) baroclinic

instability leads to slow growth of the balanced large-scale error component.

In AA, the model error caused by displaced convective cells is larger than for HRES, where convection is fully parametrized and therefore smoothed. Also, the qualitative validation presented earlier indicates that in AA convection is more confined than observed. Hence, larger initial perturbations are influencing the large-scale flow and eventually growing by baroclinic instability. Due to a high Coriolis parameter at high latitudes and a more shallow troposphere than in midlatitudes, this growth can be faster in PL active regions. Possibly the forecast quality of convective-permitting models could be improved by emphasising a subgrid-scale convective parametrization.

4 | SENSITIVITY EXPERIMENTS

The previous section reveals that AA performs a high-quality simulation of the PL within the first 24 hr of the model integration. Forecasts of more than 18 hr show some deviation from the observations, but the PL is still reasonably well predicted.

In this section, the development mechanisms of the PL are further investigated. For this, several sensitivity experiments are performed. The aim is to improve physical understanding of the PL development and to identify the critical components for accurate forecasts of PLs. Hence, the role of surface turbulent heat fluxes and latent heat release are investigated. Some of the forecast error in AA might be associated with a lack of updates to the SST, as discussed in the previous section. Therefore, the sensitivity of the PL to the SST is tested. All experiments are initiated at 0000 UTC on 3 March and integrated for 48 hr, until the landfall of the PL. A summary of the experiments and their main results is given in Table 1.

The wind field of the PL in the mature stage, at 0300 UTC on 4 March, in the different sensitivity experiments are displayed in Figure 10. The development of the PL in the different experiments is compared in Figure 11 by three intensity parameters, (a) the filtered vorticity, (b) the maximum wind speed, and (c) the minimum sea-level pressure, as well as by the maximum baroclinicity ($\nabla\theta_{850}$) in the vicinity of the PL centre. Additionally, the evolution of the sensible and latent surface heat flux and the latent heat release by condensation around the PL are depicted in Figure 12. Section 2.7 provides details about the computation of these parameters.

Wagner *et al.* (2011) and Føre and Nordeng (2012) mainly utilise the SLP for analysis of the PL evolution in the sensitivity experiments. However, the SLP has to be considered with caution. The SLP of the PL centre

is constant until 2100 UTC on 3 March (Figure 11a or Føre and Nordeng, 2012). Afterwards, the SLP rises even though the vorticity and wind speed are still increasing. This demonstrates that the local SLP is a misleading measure of the strength of the PL, since the SLP strongly depends on the synoptic-scale environment. The SLP of the synoptic-scale low is strongly affected in the sensitivity experiments (e.g. noFLX and 2FLX in Figure 10d,f). Hence, comparing the SLP among simulations can lead to wrong conclusions on the evolution of the PL itself. Therefore the SLP is only occasionally discussed in the following.

Føre and Nordeng (2012) perform experiments with delayed deactivation of different fluxes, which we generally consider valuable, but only when combined with a careful analysis. Because an immediate response of the SLP is lacking, they conclude that all the investigated diabatic components have a small direct effect on the PL. They also conclude, from growing SLP perturbations after long simulation times, that effects of different heat fluxes become more important at later stages of the PL. However, the SLP is a synoptic-scale field that is changing slowly and perturbations accumulate over time. For this reason, only the time derivative of the SLP difference from the control run could allow such conclusions. Other variables, like the wind speed and the vorticity that are investigated here, reveal the effects on the PL more directly. Hence due to inclusion of additional intensity measures, the analysis of the sensitivity experiments performed here is more comprehensive than the analysis in the previously mentioned studies. Additionally, the strength of the baroclinicity, turbulent fluxes and condensational heat release in the vicinity of the PL are included. In this way, the cause and effect of the diabatic components are distinguishable.

4.1 | Control run

In the control run (CTR), the vorticity of the PL increases until 0300 UTC on 4 March (27 hr into the simulation), the mature stage of the PL (Figure 10). Afterwards, the vorticity decays. The strongest winds associated with the PL, of up to 27 m·s^{-1}, are simulated between 22 and 32 hr into the simulation.

The baroclinicity ($\nabla\theta_{850}$) is high (>5K/100 km) until 1800 UTC on 3 March – called the initial baroclinic stage – and then steadily decreases in the mature stage (Figure 11a). After 0600 UTC on 4 March, the baroclinic zone is along the edge of the domain, and therefore these values are not displayed. However, a simulation with a domain further south, initialised from interpolation of the ECMWF HRES without the spin-up phase, reveals a
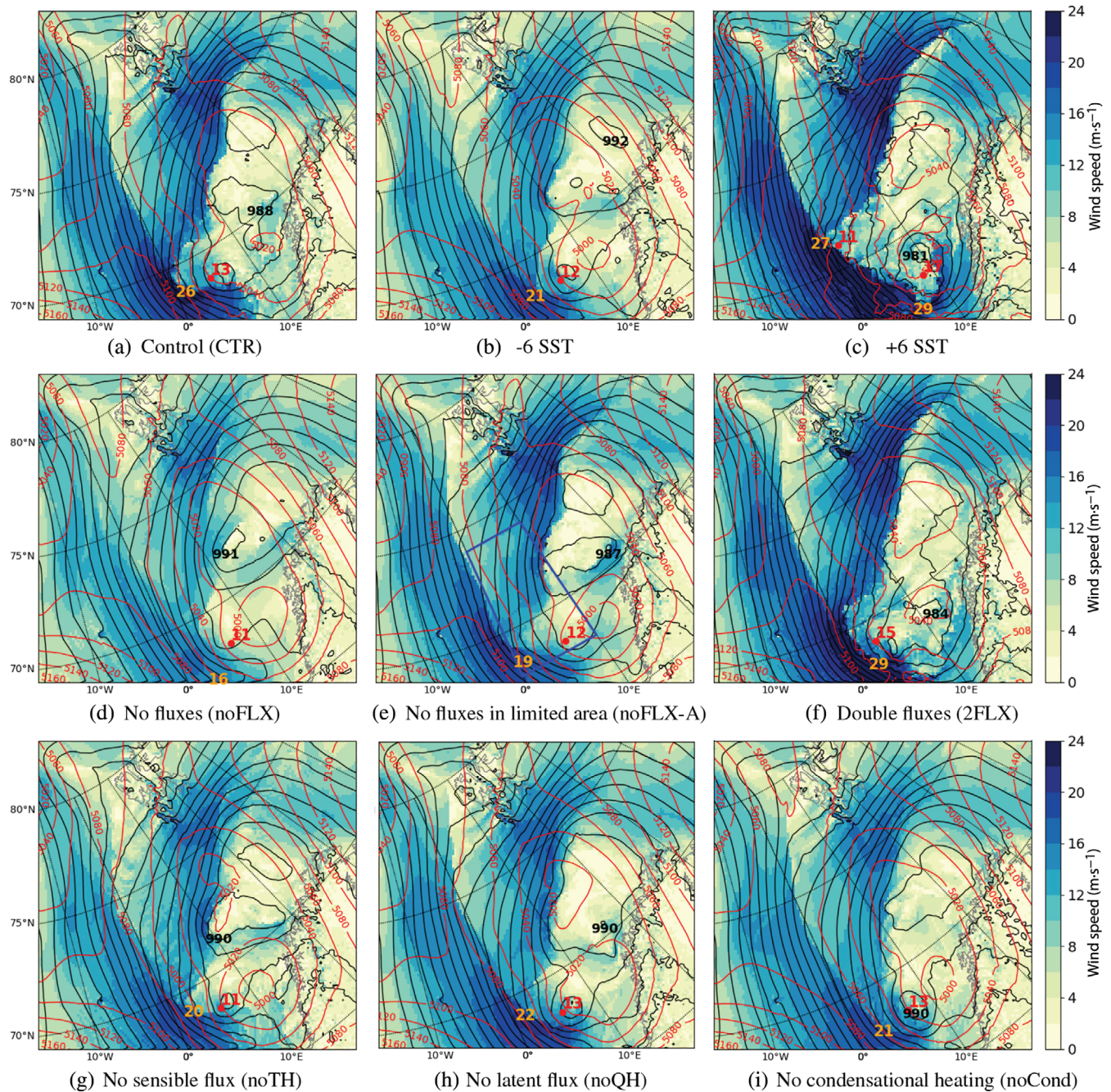
FIGURE 10 The 10 m wind speed (colour shading), sea-level pressure (black contours, spacing 2 hPa) and 500 hPa geopotential height (m, red contours) after 27 hr of model integration for different simulations starting at 0000 UTC on 3 March 2008. Red dots denote local maxima in the filtered relative vorticity at 850 hPa which defines the centres of the PL, and the red number indicates the strength in $10^{-5}s^{-1}$. The orange number depicts the maximum wind speed within 400 km of the PL centre. Black numbers show sea-level pressure minima

comparable decay of the baroclinicity in the mature stage of the PL (not shown).

Both turbulent heat fluxes around the PL increase in the baroclinic stage of the PL and eventually decrease in the mature stage after 0000 UTC on 4 March (Figure 12a). The sensible heat flux is approximately 40% higher than the latent heat flux until the middle of the mature stage of the PL at 0600 UTC on 4 March. At this stage, the air

masses around the PL are warmed considerably compared to the initial stage, so they can hold more moisture.

Interestingly, the release of latent heat by condensation is smaller than the sensible heat flux in the baroclinic stage but then it triples within 6 hr in the convective mature stage. It is also recognised that the latent heat release is higher than the surface latent heat flux by 20–30% in the baroclinic stage and more than double in theconvective
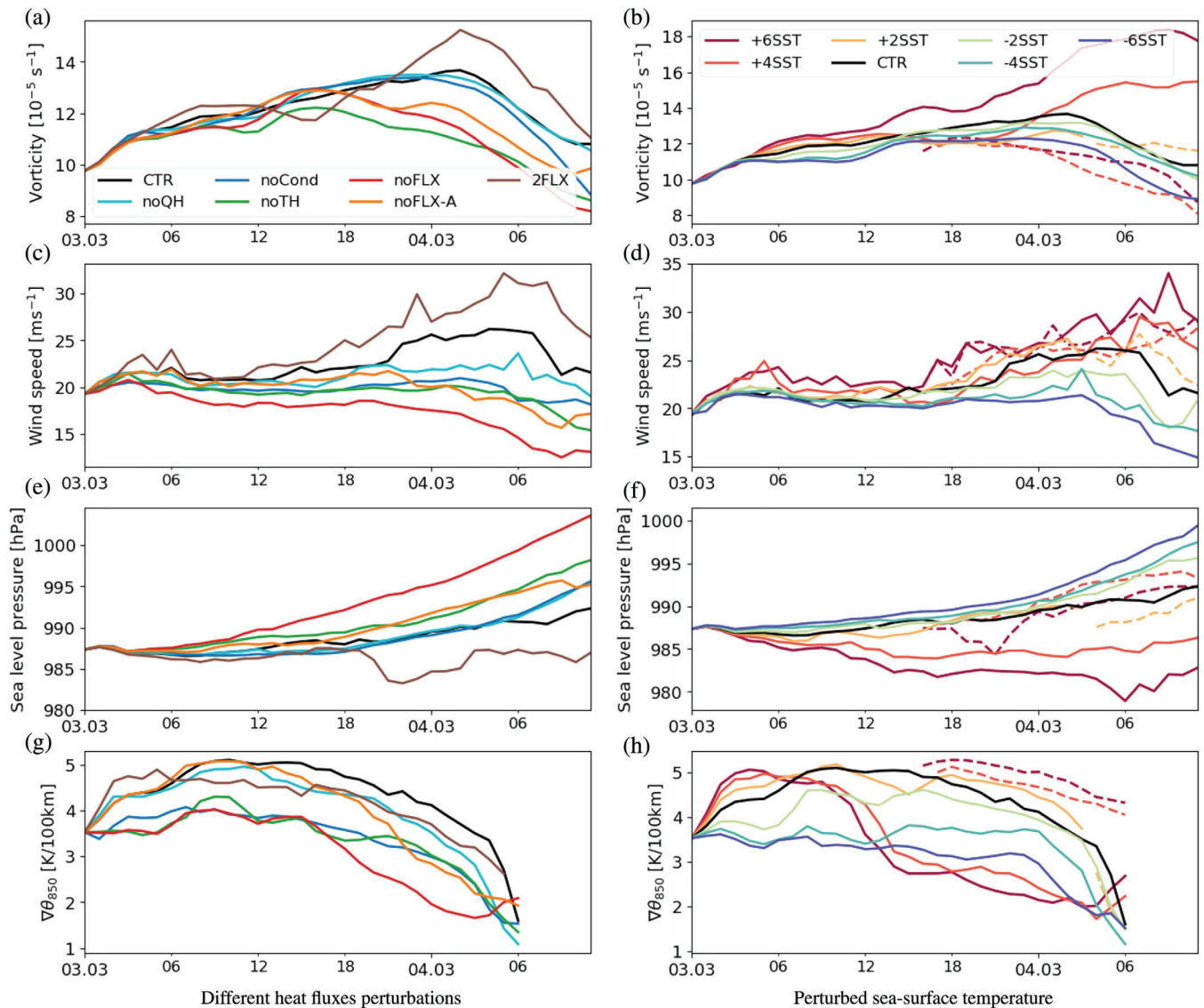
FIGURE 11 The evolution of the intensity of the PLs, as shown in Figure 10, in experiments with (a, c, e, g) perturbed fluxes, and (b, d, f, h) perturbed sea-surface temperature. The intensity is expressed as (a, b) the filtered vorticity of the centre, (c, d) the maximum wind speed within 400 km, and (e, f) the minimum sea-level pressure within 100 km distance of the vorticity centre of the PL. (g, h) show the evolution of the maximum baroclinicity ($\nabla\theta_{850}$) within 400 km distance of the PL centre. In (a, c, e, g) for +4/6SST where the PL develops two separate centres, the solid lines show the intensity of the leading centre and dashed lines of the secondary centre. Note different scales in the strength of the parameters between the two columns.

stage. This indicates that a substantial amount of the moisture is transported into the PL.

Føre and Nordeng (2012) conclude that low-level baroclinic energy conversion dominates the PL development, while other processes have a minor direct impact on the PL intensity. Here, we suggest that baroclinicity initiates and intensifies the PL, and convection maintains the PL in the mature phase of the PL from 0000 UTC on 4 March. In the following, more supporting evidence is given for this hypothesis.

4.2 | No turbulent fluxes

The role of heat fluxes from the surface is investigated in an experiment (noFLX) without both turbulent heat flux components, the sensible and the latent heat flux, over water surfaces (Figure 12d).

The maximum wind speed in the first 18 hr, measured in the western eye-wall of the PL, is somewhat weaker ($4\,\mathrm{m\cdot s^{-1}}$) than in CTR. The local wind amplification is hampered since the sharp frontal structure at the western

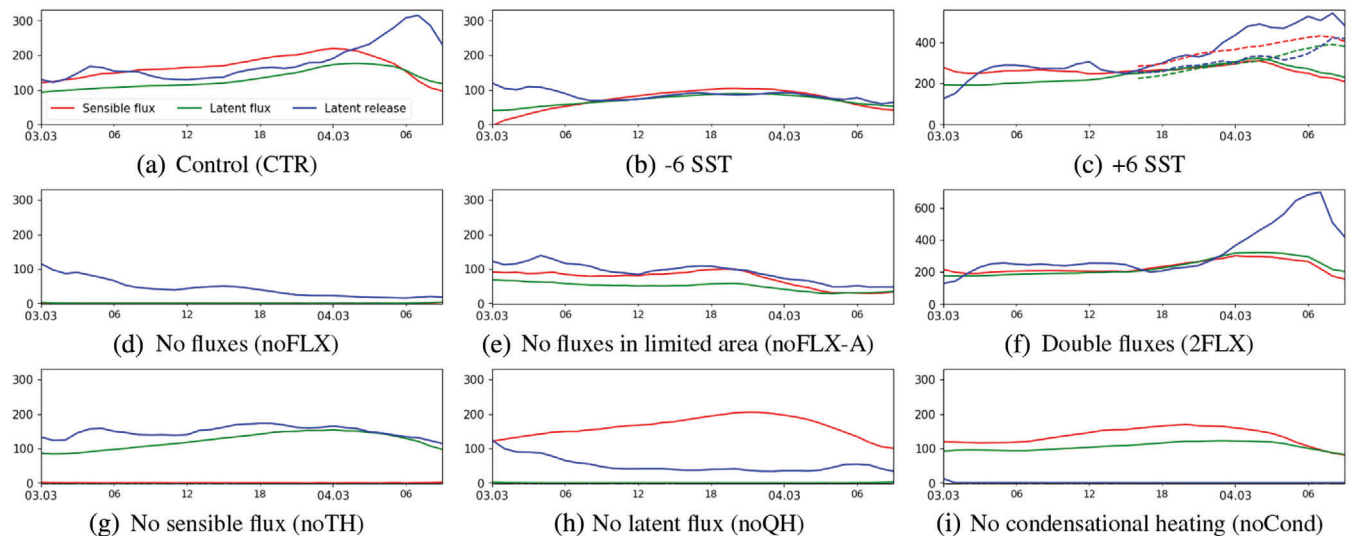**FIGURE 12** The sensible and latent surface heat fluxes and latent heat release in condensation ($W \cdot m^{-2}$) during the evolution of the PL for the different sensitivity experiments. The mean in each variable is calculated within a circle of 300 km radius around the PL centre. Note that the western eye-wall, the area of strongest surface heat fluxes, propagates along the edge of the domain after 0300 UTC on 4 March, which partly explains the reduced fluxes at that time. Note also the different vertical scales in (c) and (f). In (c) the dashed lines depict values for the secondary centre

side of the PL does not develop in the experiment without turbulent fluxes (Figure 10d). Since the baroclinicity ($\nabla\theta_{850}$) is weakened in noFLX within the first 18 hr, it is suggested that turbulent fluxes act to maintain the baroclinicity. Papritz and Spengler (2015) argue that surface sensible heat flux and latent heat release are among the main processes for the development of baroclinicity. The experiments presented later reveal that these two components are also important for the baroclinicity for this PL case.

After 18 hr of simulation, the vorticity and wind speed of the PL decay quickly, and only a moderate trough is present at 0300 UTC on 4 March (Figure 10d). Hence it is concluded that the turbulent fluxes are responsible for maintenance and further intensification of the system. These results are in accordance with Wagner *et al.* (2011) and Føre and Nordeng (2012) who performed comparable sensitivity experiments with different models without heat fluxes in simulations starting at 0600 UTC and 1200 UTC, respectively, on 3 March. They also found that the deactivation of surface heat fluxes leads to a weakening of the PL in terms of increasing SLP in the centre and prevention of convection. As here, Føre and Nordeng (2012) argue that the initial baroclinicity is "consumed" without surface heat fluxes.

4.3 | Doubled turbulent fluxes

A sensitivity experiment (2FLX) is performed where the turbulent fluxes calculated by the bulk formula are

doubled in the model simulation. This leads approximately to a doubling of the latent heat flux. In contrast, the sensible heat flux is only increased by approximately 50% because the near-surface vertical temperature gradient is reduced faster than in CTR.

In 2FLX, the PL develops similarly to CTR within the first 18 hr of the model integration, the initial baroclinic stage (Figure 11a). The baroclinicity is weaker in 2FLX because the near-surface air is heated more in the shallow CAO than on the warm and calm side of the front (not shown). Presumably, this hampers an even stronger development of the PL in the baroclinic stage in 2FLX. Interestingly, the baroclinicity develops the strongest when the turbulent fluxes are as strong as simulated in CTR. Both an increase and a decrease of the surface fluxes reduce the baroclinicity. In noFLX, the baroclinicity is "consumed," whereas in 2FLX, the baroclinicity is maintained but at a weaker level than in CTR.

From 0000 UTC on 4 March, during the convective mature stage, the intensification of the PL is strongly enhanced in 2FLX. This is indicated by an increase in the vorticity, an increase of the maximum wind speed from 27 (in CTR) to 36 m·s^{-1} (in 2FLX) and a decrease of the SLP by approximately 5 hPa until the PL encounters landfall. In this phase, the latent heat release is approximately doubled compared to CTR, leading to vortex intensification. Hence, in conclusion, the increased heat fluxes have a minor effect in the initial baroclinic stage. However, the accumulation of additional moisture leads to enhanced development in the convective stage when the latent heat is released.

4.4 | No turbulent fluxes in an area around the centre

In noFLX-A, the heat flux is turned off in a fixed area (0–10°E and 68–74°N, see blue box in Figure 10e) through which the centre is propagating within the first 27 hr with a distance of approximately 100 km to the boundary of the area10. The PL develops comparably to CTR in the initial baroclinic stage until 1800 UTC on 3 March, even though surface fluxes in the near vicinity of the PL are suppressed. In the mature stage, the omitted fluxes in the limited area prevent the PL intensifying further and developing a centre (Figure 10e). Most of the CAO, where the highest wind speed is measured, are receiving the same heat flux in noFLX-A as in CTR. For this reason, the wind strength is about the same as in CTR for the first 21 hr. Subsequently, the wind speed decays in noFLX-A, since the PL does not develop a mature stage in this experiment.

After 27 hr of model integration, at the time of highest intensity of the PL in CTR, the system leaves the area of suppressed heat fluxes (Figure 10e). However, the latent heat release does not increase in noFLX-A when the PL leaves this area, and the trough does not intensify in this experiment. In the baroclinic stage, the system appears not to accumulate enough moisture and not to develop a local statically unstable environment to further intensify convectively in the mature stage. This is in accordance to Terpstra *et al.* (2015) who conclude that interdependent thresholds in the humidity and instability are necessary for a diabatic boost of the PL development.

In conclusion, even though the PL receives some moisture from the surrounding, the local heat fluxes, particularly those leading to the accumulation of moisture and the destabilisation of the boundary layer in the baroclinic stage, are required for the development of the PL into the convective mature stage.

4.5 | Surface sensible and latent heat fluxes and latent heat release

First we give some considerations to the role of the two latent heat components. In the experiment without surface latent heat flux (noQH), moisture is still present from the initial conditions and to some extent by the boundary conditions. This moisture leads to a mean latent heat release of approximately $50\,\mathrm{W\cdot m^{-2}}$ around the PL (Figure 12h), approximately one third of the mean heat release of CTR in the baroclinic phase. In noCond the condensational heat release is completely suppressed. Consequently, the PL is weaker than in noQH, mainly as regard the wind speed, but also for the convective mature phase when it comes to vorticity. In general, it seems more meaningful to

investigate the effect of the heat release by condensation than the surface fluxes of latent heat. The former measures the consumption of "fuel," whereas the latter measures the production of the "fuel," which is not necessarily consumed.

Now the role of the different diabatic components is investigated. From the baroclinic phase, the vorticity of the PL is weakened similarly in the experiments without sensible heat flux (noTH) and with both flux components suppressed (noFLX; Figure 11a). Differently, effects on the vorticity are only recognisable later in the mature phase for suppressed condensational heat release (noCond), and negligible for suppressed latent heat flux (noQH). Hence, in the initial baroclinic stage, the sensible heat flux mainly favours the vortex intensification of this PL.

In contrast, for both noTH and noCond, the wind development is weaker than in CTR but stronger than in noFLX (Figure 11c). This means that the maximum wind speed is dependent on both sensible heat flux and latent heat release. Hence, the two diabatic components act differently on the intensity measures. Also for both noTH and noCond, the baroclinicity is weakened as in noFLX. Hence, the sensible heat flux and condensational heat release appear to be important for increasing and maintaining the baroclinicity in the initial stage of the PL (Figure 11g).

In noQH, the wind speed and baroclinicity are less influenced in the initial stage than in noCond, because the moisture present still condenses. First in the mature stage, the wind speed development is weaker in noQH than in CTR. Both in noQH and noCond the intensification of the PL in the convective mature stage, which is fuelled by latent heat release, is hampered. In noCond the latent energy is not released (by construction) and in noQH too little moisture accumulates in the baroclinic stage.

Interestingly, even though the latent heat flux is approximately the same in noTH as in CTR, the PL does not develop a convective mature stage in noTH, as the latent heat release and wind speed do not increase. This raises two suggestions: (a) the PL has to reach a certain strength before the engine of latent heat release can maintain the system in the mature stage, and (b) a destabilisation of the boundary layer by sensible heat flux is required to make latent heat release an effective intensification mechanism. Both effects may be coexisting and interacting. This suggests that all diabatic components are required to accomplish the full PL development.

These results are mainly in accordance with Føre and Nordeng (2012), but we come to opposite conclusions in two respects: (a) from theoretical considerations, the surface latent heat flux cannot be more important than the condensational heat release for the development of
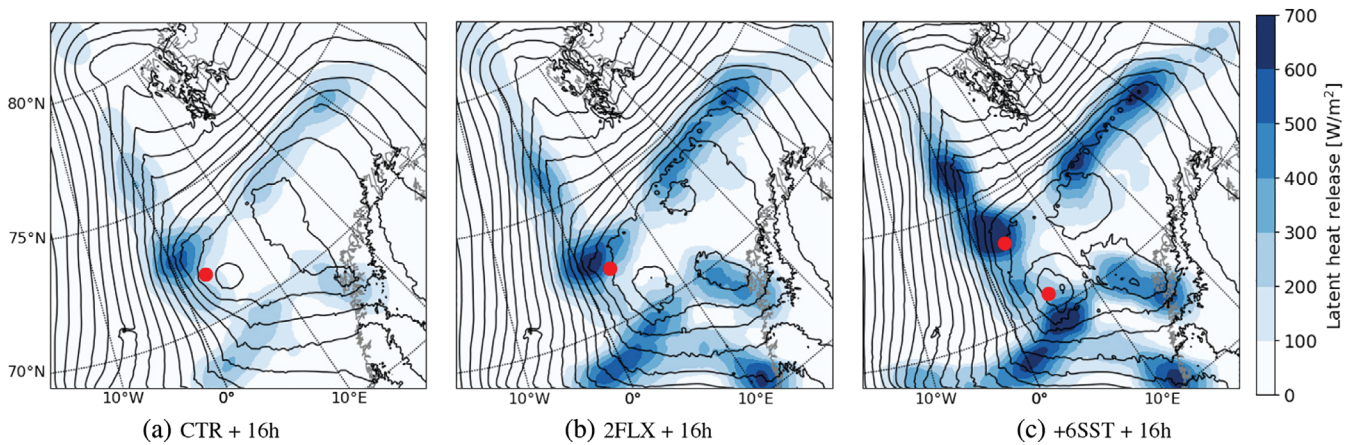
(a) CTR + 16h   (b) 2FLX + 16h   (c) +6SST + 16h

FIGURE 13 The latent heat release by condensation (colour shading) and sea-level pressure (contours, spacing 2 hPa) for (a) CTR, (b) 2FLX and (c) +6SST, all after 16 hr model integration. The red dot denotes the location of the vorticity centre of the PL. The latent heat release is derived from the precipitation rate and smoothed with a Gaussian filter of 100 km radius

dynamical systems, which is also observed for this PL, and (b) sensible heat flux is more relevant than the latent heat flux for the development of the PL in the baroclinic stage.

4.6 | Role of latent heat release on the polar low track

It is observed that the PL is "pulled" towards the area of the strongest convection. In Figure 9c the tracks of the PL in the different sensitivity experiments are displayed. At 1200 UTC on 3 March only small differences are visible in the location since the perturbations had little time to grow. However, in the experiment without condensational heat release, the PL at this time is already deflected towards the east. This deflection is explained by convection and the associated latent heat release by condensation mainly occurring on the western side of the PL (Figure 13a). The heating induces a positive potential vorticity anomaly below this heat source, hence at low levels, following the Diabatic Rossby Vortex concept (Terpstra et al., 2015). This process "pulls" the PL towards the area of the strongest latent heating and hence reduces the propagation speed of the PL. Therefore, in the experiment without condensational heating, the missing "drag" from the convection leads to a faster propagation of the PL until 0600 UTC on 4 March.

Likewise, in the experiments noFLX, noFLX-A and noQH (where latent heat release in the vicinity of the PL is dampened), the PL has also propagated faster until 0600 UTC on 4 March (Figure 9c). In contrast, the PL was slowed down in 2FLX where latent heat release is amplified (Figure 13b). This is especially recognisable since the large-scale flow is decreased in noCond and noFLX (Figure 10), which would decrease the propagation speed

of the PL, and the opposite for 2FLX. In noFLX-A, the large-scale flow is little affected, but, as for the before mentioned experiments, the PL propagates faster due to reduced local convection. The sensible heat flux, different to the latent heat release by condensation, has a negligible influence on the PL track (compare noTH to CTR).

4.7 | Perturbation of the sea-surface temperature

In order to examine the sensitivity of the PL to the sea-surface temperature (SST), experiments are performed with perturbed SST from −6 to +6 °C with 2 °C increments. This generally provides more realistic perturbations of the surface fluxes than the experiments with adapted flux components (e.g. noTH, noQH, noFLX and 2FLX). The increase (decrease) of the SST by 6 °C leads to approximately a doubling (halving) of both heat flux components in the area of the PL development (Figure 12a–c).

The initial baroclinic stage is highly influenced by the SST perturbations (Figure 11b,d,f,h). For higher (lower) SSTs, the baroclinicity is considerably stronger (weaker) than in CTR. The large-scale baroclinic environment is enhanced (weakened) by increased (decreased) SSTs, as the warm area over the sea becomes warmer (colder) by increased (decreased) heat fluxes, while the cold region, which is determined by the conditions over the Arctic sea ice, is unaffected. Interestingly, the baroclinicity does not exceed a value of approximately $5 \, \text{K} \cdot (100 \, \text{km})^{-1}$ for any experiment (Figure 11h), even though the baroclinic development is faster for increased SST, as can be seen in higher vorticity and wind speeds in Figure 11b,d for the initial stage. Likely at the threshold of $5 \, \text{K} \cdot (100 \, \text{km})^{-1}$, the baroclinicity

is produced and consumed at approximately the same rate.

The Arctic front develops considerably more strongly for increased SSTs (compare CTR to +6SST in Figure 13a,c at 69°N, 0–7°E). The initial PL centre in +6SST connects to the diabatic heating associated with the Arctic front. Since the front lies ahead of the PL, the initial centre is accelerated and "pulled" out of the baroclinic zone. Therefore the baroclinicity of the initial PL centre is low from 1200 UTC on 3 March in +6SST. Some hours later, the PL centre in +6SST intensifies further and develops into a convective system with mean values of latent heat release above 300 W·m^{-2} around the centre and of more than 500 W·m^{-2} from 0000 UTC on 4 March, approximately twice as high as in CTR (Figure 12a,c).

Also in +2FLX the Arctic front is enhanced as compared to CTR, but is considerably less than in +6SST (Figure 13). The initial PL centre does not connect the Arctic front as in +6SST, possibly since the latent heat release in the front close to the PL centre is too weak to "pull" the PL centre out of the baroclinic zone.

In +4/6SST, a second PL centre develops around 1600 UTC on 3 March at the intersection point between the baroclinic front (73–77°N, 2°E) and the convergence zone (74°N, 2–30°E in Figure 13c). It is accompanied by high baroclinicity which slowly decays, whereas latent heat release is increasing (dashed lines in Figures 11h and 12c). This also indicates a transition into a convective system for the secondary centre. However, this transition is not finalised before the PL reaches the edge of the domain. This second centre is significantly slowed down by the strong latent heat release at the intersection point behind the centre (Figures 13c and 10c).

In the mature stage, the difference in the PL intensity for various SST perturbations is very pronounced. It is estimated that 1 °C of SST increase leads to enhanced near-surface winds of 1–2 m·s^{-1} (Figure 11b). The intensity of the PL appears to increase nonlinearly with the warming of the sea surface. Different observations capture this.

- The perturbation of the vorticity and SLP nonlinearly increase with higher SSTs. The vorticity of +6SST almost doubles compared to CTR, whereas the vorticity of −6SST is only slightly lower than in CTR. The difference in SLP between CTR and −6SST is approximately 4 hPa in the mature stage of the PL, whereas the difference between CTR and +6SST is about 10 hPa. However, note that the SLP perturbations are partly caused by a deepening of the synoptic low.
- The kinetic energy increases with the square of the velocity. The wind speed increases at least linearly with

increased SST, and the area of stronger winds expands for increased SST (Figure 10k,l)

- With highly increased SSTs (+4/6SST), the PL develops a second centre (Figure 10c). The intensity of the second centre is displayed by dashed lines in Figures 11b and 12c. It reaches the same intensity as the first centre in terms of the wind speed.

The experiments performed here suggest a much higher sensitivity of the PL intensity to the SST thanwas obtained by simulation of an axisymmetric non-hydrostatic idealised model by Linders *et al.* (2011). They obtain an increase of the wind speed of about 0.6 m·s^{-1} and a decrease of the core pressure of −0.6 hPa per °C increase in SST. The increase/decrease rates obtained here are more than twice as high.

The high sensitivity of the PL development to the SST suggests that updated sea surface fields are essential for a realistic simulation of PLs. A coupled atmosphere–ocean weather prediction model could be beneficial for the forecast of PLs, especially since the strong heat fluxes can lead to a modification of the SSTs. One would expect that the CAO leads to a cooling of the ocean surface. However for one PL event, Sætra *et al.* (2008) demonstratedthat turbulent mixing of warm sub-surface currents by strong winds led to a rapid surface warming of 1–2 °C within a few hours. This SST warming is a positive feedback for further PL intensification.

5 | DISCUSSION AND CONCLUSION

In the first part of the study, the capability of the regional weather-prediction model AROME-Arctic (AA) for representing the THORPEX PL, which occurred on 3–4 March 2008 in the Norwegian Sea, is validated against observations and compared to the performance of the global model ECMWF HRES. In the second part of the study, the development mechanisms involved in this PL are investigated by sensitivity experiments with AA.

5.1 | Model validation

The comparison of the simulated cloud fields of AA with satellite images reveals the high quality of the model. AA captures the observed cloud types with a comparable structure at approximately the correct location, which is a large improvement over HRES. However, AA tends to simulate the deep and shallow convective cells more discretely than observed. The model appears to include too much

convection in the model dynamics but to underestimate subgrid-scale convection.

The near-surface wind fields of both models compare well to the scatterometer wind field from QUIKSCAT. The largest differences between AA and the satellite product occur in the zones of dislocated fronts. In this zone, the sharpness of the front in AA appears more appropriate than both QUIKSCAT and HRES.

This PL was also measured by dropsondes released from three flights during the IPY-THORPEX campaign. A one-to-one comparison of the vertical profiles from the dropsondes to the model grid cells is inappropriate, because the correct simulation of a local feature, such as a convective cell, at a wrong location, leads to a double punishment in classical error scores. Therefore, the skills in error statistics, such as the mean absolute error, are similar for AA and HRES, even though AA improves in the qualitative validation. For this reason, a "fuzzy" verification technique, which relaxes the requirement for spatial collocation of observations and simulations, is applied to the dropsonde data. This reveals that AA has higher skill than HRES for small scales (≤ 10 km) and for high intensities (e.g. for wind speeds ≥ 20 m·s^{-1}).

A few more conclusions are drawn from the comparison of AA to the dropsondes:

- The model has a near-surface warm BIAS in the initial baroclinic stage. The BIAS is likely caused by the lack of updating of the SST boundary fields in the model.
- The model is statically too unstable in the planetary boundary layer (Figure 8a,c compare the potential temperature profile to the moist adiabats), which might be caused by the near-surface warm BIAS.
- The humidity profiles indicate that shallow convection is too strong in the model. This could be induced by too weak stability of the boundary layer. Hence, an erroneous SST field appears to cause the misrepresentation of a range of other variables in the model.
- The model might overestimate the depth of deep convective cells, possibly because the implicit treatment of the model produces cells that are too confined. A deep convective parametrization scheme could relax this problem.

For predictions beyond 18 hr, AROME-Arctic deviates more from reality than the ECMWF operational model. This is in accordance with Køltzow *et al.* (2019), who also observe that error growth is faster in AA than in ECMWF HRES. Some of the faster error growth is associated with a larger dislocation of the PL in AA forecasts than in HRES forecasts. From sensitivity experiments, it is concluded that erroneous representation of convection (and

connected latent heat release) has a considerable influence on the displacement of the PL in AA.

In this study, only the deterministic forecast of the models is validated. A comparison of several ensembles of each model could give additional information about the model uncertainty. Also, a comprehensive study of more PL cases would be relevant in order to examine the forecast time upto which the high-resolution AA provides more accurate predictions than ECMWF HRES.

5.2 | Polar low development

The second focus of this study is the investigation of the development of this PL. The wind profile, cloud structure, strength of baroclinicity and heat fluxes indicate that the PL initially develops in a baroclinic zone, a remnant of an occluded synoptic-scale low. As it intensifies, the baroclinicity decreases, and the PL develops into a quasi-barotropic convective system with strong latent heat release and a warm core.

There are two general remarks on the common practice of the analysis of PLs:

- The sea-level pressure used as an intensity proxy of a PL has little relevance because it is mainly determined by the synoptic-scale environment. Most other studies that have investigated this PL have focused mainly on this parameter for the comparison of sensitivity experiments.
- The surface latent heat flux has no direct influence on the PL development. It merely creates the potential for latent heat release by condensation. The latter is of significant interest for the investigation of PLs.

Sensitivity experiments, summarised in Table 1, are performed with AA in order to study the PL in more detail. The vortex of the PL develops surprisingly similarly within the first 18 hr of the simulation in all experiments. We conclude that, in the initial baroclinic stage, the vortex development is mainly driven by the synoptic-scale environment and has limited sensitivity to different diabatic effects. However, the surface sensible heat flux and condensational heat release both contribute to enhance the baroclinicity. If these heat sources are suppressed, the PL weakens from the end of the baroclinic stage. In the initial baroclinic stage, both the sensible heat flux and the latent heat release locally intensify the near-surface wind by approximately 2 m·s^{-1}. Both diabatic contributions lead to a sharpening of the frontal zones.

In the mature stage, the baroclinicity is low, and latent heat release appears to maintain and intensify the PL, hence it is of a convective nature. At this stage, only

less than half of the consumed moisture is locally produced. The convective mature stage does not develop in the absence of sensible heat flux, latent heat flux, or latent heat release. Instead, the PL intensity decreases. Also, if the turbulent fluxes are suppressed in a limited area, through which the PL propagates in the initial baroclinic stage, the PL does not intensify further. The vortex intensifies in the mature stage when the surface fluxes or the SST are increased. Therefore, we conclude that the development of the mature stage depends on the sensible heat flux having destabilised the local environment around the PL core sufficiently, and enough moisture having accumulated for condensational heat release. It is also observed that the PL is "pulled" towards the area of the strongest convection. Following the Diabatic Rossby Vortex concept, the latent heat release associated with the convection induces a positive potential vorticity anomaly at low levels. This anomaly intensifies the PL and "pulls" the centre towards the area of latent heating. Hence, the propagation of the PL is influenced by the location of the convective area.

Sensitivity experiments with perturbed SST reveal an increased maximum in near-surface wind speed connected to the PL of 1–$2\,\mathrm{m\cdot s^{-1}}$ per K warming of the sea surface. This estimate is more than twice as high as the one provided by idealised experiments by Sætra *et al.* (2008). Further, the intensity of the PL increases nonlinearly with higher SSTs. For increased SSTs of 4 and 6 °C, a secondary PL centre develops after the first centre has propagated out of the baroclinic zone. The development of coupled atmosphere–ocean weather prediction systems with more sophisticated SSTs might considerably improve the predictions of PLs.

We conclude that baroclinicity provides the cradle for this PL, and diabatic processes in a conditionally unstable environment further intensify the system in the mature stage. The correct simulation of the latter stage appears to be more challenging for NWP models than the initial baroclinic stage.

ACKNOWLEDGEMENTS

ORCID

Patrick J. Stoll https://orcid.org/0000-0003-1120-2049

REFERENCES

Atger, F. (2001) Verification of intense precipitation forecasts from single models and ensemble prediction systems. *Nonlinear Processes in Geophysics*, 8(6), 401–417.

Bengtsson, L., Andrae, U., Aspelien, T., Batrak, Y., Calvo, J., de Rooy, W., Gleeson, E., Hansen-Sass, B., Homleid, M., Hortal, M., Ivarsson, K.-I., Lenderink, G., Niemelǎđ, S., Nielsen, K.P., Onvlee, J., Rontu, L., Samuelsson, P., Muñoz, D.S., Subias, A., Tijm, S., Toll, V., Yang, X. and Køltzow, M.Ø. (2017) The HARMONIE-AROME model configuration in the ALADIN-HIRLAM NWP system. *Monthly Weather Review*, 145(5), 1919–1935.

Bracegirdle, T.J. (2006). The role of convection in the intensification of polar lows. PhD thesis, University of Reading, UK.

Brümmer, B., Müller, G. and Noer, G. (2009) A polar low pair over the Norwegian Sea. *Monthly Weather Review*, 137(8), 2559–2575.

Douglas, M.W., Fedor, L.S. and Shapiro, M.A. (1991) Polar low structure over the northern Gulf of Alaska based on research aircraft observations. *Monthly Weather Review*, 119(1), 32–54.

Douglas, M.W., Shapiro, M.A., Fedor, L.S. and Saukkonen, L. (1995) Research aircraft observations of a polar low at the East Greenland ice edge. *Monthly Weather Review*, 123(1), 5–15.

Ebert, E.E. (2008) Fuzzy verification of high-resolution gridded forecasts: a review and proposed framework. *Meteorological Applications*, 15(1), 51–64.

ECMWF (2018) *Changes in ECMWF model.* Available at: https://www.ecmwf.int/en/forecasts/documentation-and-support/changes-ecmwf-model; accessed 19 February 2020.

Field, P.R., Brožková, R., Chen, M., Dudhia, J., Lac, C., Hara, T., Honnert, R., Olson, J., Siebesma, P., de Roode, S., Tomassini, L., Hill, A. and McTaggartâĂŘCowan, R. (2017) Exploring the convective grey zone with regional simulations of a cold air outbreak. *Quarterly Journal of the Royal Meteorological Society*, 143, 2537–2555.

Føre, I., Kristjánsson, J.E., Saetra, Ø., Breivik, Ø., Røsting, B. and Shapiro, M. (2011) The full life cycle of a polar low over the Norwegian Sea observed by three research aircraft flights. *Quarterly Journal of the Royal Meteorological Society*, 137, 1659–1673.

Føre, I. and Nordeng, T.E. (2012) A polar low observed over the Norwegian Sea on 3–4 March 2008: high-resolution numerical experiments. *Quarterly Journal of the Royal Meteorological Society*, 138, 1983–1998.

Furevik, B.R., Schyberg, H., Noer, G., Tveter, F. and Röhrs, J. (2015) ASAR and ASCAT in polar low situations. *Journal of Atmospheric and Oceanic Technology*, 32(4), 783–792.

Gilleland, E., Ahijevych, D., Brown, B.G., Casati, B. and Ebert, E.E. (2009) Intercomparison of spatial forecast verification methods. *Weather and Forecasting*, 24(5), 1416–1430.

Gilleland, E., Ahijevych, D.A., Brown, B.G. and Ebert, E.E. (2010) Verifying forecasts spatially. *Bulletin of the American Meteorological Society*, 91(10), 1365–1376.

Holtslag, A.A.M., Svensson, G., Baas, F., Basu, S., Beare, B., Beljaars, A.C.M., Bosveld, F.C., Cuxart, J., Lindvall, J., Steeneveld, G.J., Tjernstrøom, M. and Van De Wiel, B.J.H. (2013) Stable

atmospheric boundary layers and diurnal cycles: challenges for weather and climate models. *Bulletin of the American Meteorological Society*, 94(11), 1691–1706.

Innes, H.M., Kristiansen, J., Kristjánsson, J.E. and Schyberg, H. (2011) The role of horizontal resolution for polar low simulations. *Quarterly Journal of the Royal Meteorological Society*, 137, 1674–1687.

Køltzow, M., Casati, B., Bazile, E., Haiden, T. and Valkonen, T. (2019) A NWP model intercomparison of surface weather parameters in the European Arctic during the year of polar prediction Special Observing Period Northern Hemisphere 1. *Weather and Forecasting*, 43, 959–983.

Kristjánsson, J.E., Barstad, I., Aspelien, T., Føre, I., Godøy, Ø., Hov, Ø., Irvine, E., Iversen, T., Kolstad, E., Nordeng, T.E., McInnes, H., Randriamampianina, R., Reuder, J., Saetra, Ø., Shapiro, M., Spengler, T. and Ólafsson, H. (2011) The Norwegian IPY-THORPEX: polar lows and Arctic fronts during the 2008 and Andøya campaign. *Bulletin of the American Meteorological Society*, 92(11), 1443–1466.

Linders, T., Saetra, Ø. and Bracegirdle, T.J. (2011) Limited polar low sensitivity to sea-surface temperature. *Quarterly Journal of the Royal Meteorological Society*, 137, 58–69.

Markowski, P. and Richardson, Y. (2010) *Mesoscale Meteorology in Midlatitudes*. John Wiley & Sons, Chichester, UK.

Müller, M., Batrak, Y., Kristiansen, J., Køltzow, M.A.Ø., Noer, G. and Korosov, A. (2017a) Characteristics of a convective-scale weather forecasting system for the European Arctic. *Monthly Weather Review*, 145(12), 4771–4787.

Müller, M., Homleid, M., Ivarsson, K.-I., Køltzow, M.AØ., Lindskog, M., Midtbø, K.H., Andrae, U., Aspelien, T., Berggren, L., Bjørge, D., Dahlgren, P., Kristiansen, J., Randriamampianina, R., Ridal, M. and Vignes, O. (2017b) AROME-MetCoOp: a Nordic convective-scale operational weather prediction model. *Weather and Forecasting*, 32(2), 609–627.

Nordeng, T.E. and Rasmussen, E.A. (1992) A most beautiful polar low. A case study of a polar low development in the Bear Island region. *Tellus A*, 44(2), 81–99.

Papritz, L. and Spengler, T. (2015) Analysis of the slope of isentropic surfaces and its tendencies over the North Atlantic. *Quarterly Journal of the Royal Meteorological Society*, 141, 3226–3238.

Rasmussen, E.A. and Turner, J. (2003) *Polar Lows: Mesoscale Weather Systems in the Polar Regions*. Cambridge University Press, Cambridge, UK.

Rojo, M., Claud, C., Mallet, P.-E., Noer, G., Carleton, A.M. and Vicomte, M. (2015) Polar low tracks over the Nordic Seas: a 14-winter climatic analysis. *Tellus A*, 67. https://doi.org/10.3402/tellusa.v67.24660

Sætra, O., Linders, T. and Debernard, J.B. (2008) Can polar lows lead to a warming of the ocean surface?. *Tellus A*, 60(1), 141–153.

Seity, Y., Brousseau, P., Malardel, S., Hello, G., Bénard, P., Bouttier, F., Lac, C. and Masson, V. (2011) The AROME-France convective-scale operational model. *Monthly Weather Review*, 139(3), 976–991.

Shapiro, M.A., Fedor, L.S. and Hampel, T. (1987) Research aircraft measurements of a polar low over the Norwegian Sea. *Tellus A*, 39(4), 272–306.

Stoll, P.J., Graversen, R.G., Noer, G. and Hodges, K. (2018) An objective global climatology of polar lows based on reanalysis data. *Quarterly Journal of the Royal Meteorological Society*, 144(716), 2099–2117.

Terpstra, A., Michel, C. and Spengler, T. (2016) Forward and reverse shear environments during polar low genesis over the North East Atlantic. *Monthly Weather Review*, 144(4), 1341–1354.

Terpstra, A., Spengler, T. and Moore, R.W. (2015) Idealised simulations of polar low development in an Arctic moist-baroclinic environment. *Quarterly Journal of the Royal Meteorological Society*, 141, 1987–1996.

Valkonen, T.M., Stoll, P.J., Batrak, Y., Køltzow, M., Schneider, T.M., Stigter, E.E., Aashamar, O.B., Støylen, E. and Jonassen, M. (2020) Evaluation of a sub-kilometre NWP system in an Arctic fjord-valley system in winter. (unpublished manuscript)

Verhoef, A., Vogelzang, J. and Stoffelen, A. (2016). Reprocessed SeaWinds L2 winds validation report. Technical Report SAF/OSI/C-DOP2/KNMI/TEC/RP/221. KNMI, De Bilt, Netherlands.

Wagner, J.S., Gohm, A., Dörnbrack, A. and Schäfler, A. (2011) The mesoscale structure of a polar low: airborne lidar measurements and simulations. *Quarterly Journal of the Royal Meteorological Society*, 137, 1516–1531.

Yanase, W., Fu, G., Niino, H. and Kato, T. (2004) A polar low over the Japan Sea on 21 January 1997. Part II: a numerical study. *Monthly Weather Review*, 132(7), 1552–1574.

Yanase, W. and Niino, H. (2005) Effects of baroclinicity on the cloud pattern and structure of polar lows: a high-resolution numerical experiment. *Geophysical Research Letters*, 32(2). https://doi.org/10.1029/2004GL020469

Yanase, W. and Niino, H. (2007) Dependence of polar low development on baroclinicity and physical processes: an idealized high-resolution numerical experiment. *Journal of the Atmospheric Sciences*, 64(9), 3044–3067.

Zhang, F., Bei, N., Rotunno, R., Snyder, C. and Epifanio, C.C. (2007) Mesoscale predictability of moist baroclinic waves: convection-permitting experiments and multistage error growth dynamics. *Journal of the Atmospheric Sciences*, 64(10), 3579–3594.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.