UNIVERSITY OF OKLAHOMA
GRADUATE COLLEGE

INVESTIGATING ARCTIC CYCLONE-TROPOPAUSE POLAR VORTEX INTERACTIONS WITH OBSERVING SYSTEM SIMULATION EXPERIMENTS

A THESIS
SUBMITTED TO THE GRADUATE FACULTY
in partial fulfillment of the requirements for the
Degree of
MASTER OF SCIENCE

By

MATTHEW BRAY
Norman, Oklahoma
2023

INVESTIGATING ARCTIC CYCLONE-TROPOPAUSE POLAR VORTEX INTERACTIONS WITH OBSERVING SYSTEM SIMULATION EXPERIMENTS

A THESIS APPROVED FOR THE
SCHOOL OF METEOROLOGY

BY THE COMMITTEE CONSISTING OF

Dr. Steven Cavallo (Chair)

Dr. David Parsons

Dr. Xuguang Wang

Acknowledgments

I would first like to thank my advisor, Dr. Steven Cavallo for years of teaching, mentoring, and encouragement. I first started working with Steven nearly seven years ago as an undergraduate, and since then, he has guided me through several research projects, scholarship and fellowship applications, and many, many semesters of coursework. I don't know where I would be today without all of his support. Thank you also to the entire AAARG research group for providing continued feedback on this project, help with running code, and stress-relieving conversations in the office. A special thanks to Dr. Chris Riedel and Tomer Burg, who were instrumental in helping me set up and run MPAS-DART and provided plenty of analysis code. Additionally, thank you to Dr. David Parsons and Dr. Xuguang Wang for serving on my committee and providing multiple rounds of valuable feedback on this work. I'd also like to shout out all of my family and friends, both in the School of Meteorology and all around the country, for keeping me motivated with game nights, hiking trips, and late-night talks. Finally, thank you to my best friend and fiancée, Anu, for being there with me through it all and keeping me happy and calm no matter how crazy things get.

Table of Contents

List Of Tables

List Of Figures

Abstract

Tropopause polar vortices (TPVs) are upper-level cyclonic circulations that are common in high latitudes. TPVs most directly impact weather at the surface by providing baroclinic forcing for the development of Arctic cyclones (ACs), surface low pressure systems over the Arctic Ocean and surrounding landmasses. ACs, in turn, help to drive the Arctic heat and moisture budgets via transport from the midlatitudes and govern sea ice variability on short timescales via dynamic and thermodynamic influences. Although prior studies have demonstrated and examined the link between TPVs and ACs, the exact role played by the TPV and its characteristics in the development of the surface cyclone has yet to be established. In the present study, we seek to take a step forward towards closing this gap using an idealized observing system simulation experiment (OSSE). This approach involves simulating dense dropsondes over a TPV as a surface cyclone is forming and examining the changes that the additional observations produce to the TPV and resulting surface cyclone. The ECMWF Cubic Octahedral grid Nature Run (ECO1280) is taken as the truth for the OSSE (all observations are simulated from the output of this model), and the Model for Prediction Across Scales (MPAS) is coupled with the Data Assimilation Research Testbed's (DART) ensemble adjustment Kalman filter in order to run numerical experiments. It is expected that the development of the TPV and surface cyclone will be especially sensitive to moisture observations, which influence TPV strength via diabatic processes, and that proper representation of mesoscale dynamic features along the tropopause will be key to accurate forecasts.

In addition to a control in which only existing observations are simulated from ECO1280, we conduct four primary experiments assimilating additional dropsonde observations of (1) temperature, (2) humidity, (3) temperature and humidity, and (4) temperature and humidity over a broad region. All of the experiments reduce errors relative to the control throughout the atmosphere and at the surface, with Experiment 4 producing the most accurate forecast of the surface cyclone. The humidity observations lead to enhanced PV generation near the tropopause and, by the end of the forecast, a deeper surface cyclone. These improvements are more limited when temperature and humidity dropsondes are combined. Experiment 4 exhibits a much more accurate representation of the PV features surrounding the main TPV and other synoptic scale features surrounding the system (including a Rossby wave breaking into the Arctic from the midlatitudes), which may explain its better forecast performance with respect to the surface cyclone. Additionally, it is discovered that AC strength is linked to the accuracy of the modeled distance between the surface cyclone and the core of the TPV throughout its lifetime (an indicator of baroclinic growth potential). This TPV-AC distance is, in turn, controlled by a breaking midlatitude Rossby wave that was not directly observed in any of the experiments.

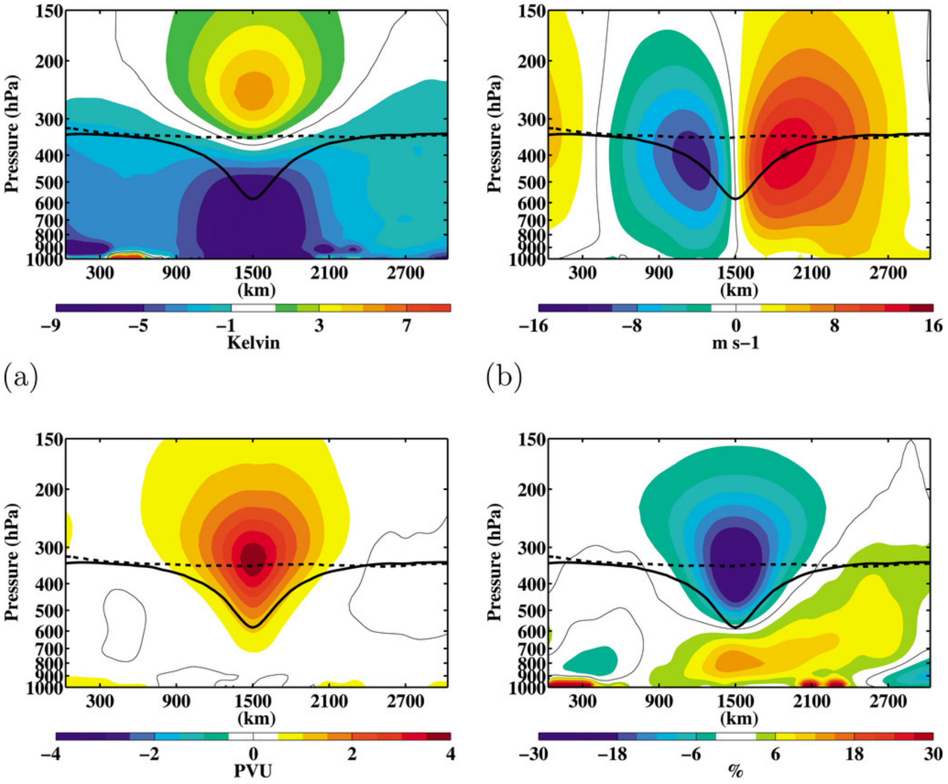
Chapter 1

Introduction and Background

1.1 Tropopause Polar Vortices

Tropopause polar vortices (TPVs) are a key dynamic feature of the upper levels of the high-latitude atmosphere. TPVs are circulations with closed material contours that are centered on the tropopause and with attendant flow extending throughout the upper troposphere and lower stratosphere (UTLS) region (Hakim, 2000; Hakim and Canavan, 2005). These vortices have radii on the order of hundreds of kilometers (sometimes over 1000 km) and exist on timescales of days to weeks (Hakim and Canavan, 2005; Cavallo and Hakim, 2009). TPVs have been observed to persist for months, especially considering the ubiquity of vortex split and merger events (Hakim and Canavan, 2005). TPVs may be cyclonic or anticyclonic depending on the direction of flow and can occur in both the Arctic and Antarctic (Cavallo and Hakim, 2009, 2010; Gordon et al., 2023); in the present study, we focus exclusively on cyclonic TPVs in the Arctic.

For conceptual purposes, TPVs are best considered as positive upper-level potential vorticity (PV) anomalies (Hakim and Canavan, 2005; Cavallo and Hakim, 2010). The general vertical structure of a TPV, including an anomalously cold troposphere and warm stratosphere, cyclonic flow throughout the atmosphere, and an enhanced moisture gradient across the tropopause can be seen in Figure 1.1. As such, the strength of a TPV can be considered as the strength of a PV anomaly, which are governed by well-established tendency equations (Pedlosky, 1992). In the upper atmosphere, diabatic forcings such as latent heating and radiation plays a dominant role in PV creation and destruction, and, in the Arctic, the radiative terms (longwave and shortwave)



(a) (b)

(c) (d)

Figure 1.1: Composite vertical cross-sections of an idealized TPV. Individual panels show anomalies from climatology of (a) temperature (K), (b) meridional wind (m s^{-1}), (c) Ertel potential vorticity (PVU), and (d) relative humidity (%). Borrowed from Cavallo and Hakim (2010).

generally predominate (Curry et al., 1996). To examine their spatial characteristics, TPVs are often viewed as (negative) potential temperature anomalies on the dynamic tropopause (defined here as the 2 PVU surface). Compared with isentropic analyses, this method has the advantage of only requiring maps at one level to see the approximate tropopause region throughout the extratropics (Morgan and Nielson-Gammon, 1998).

Several definitions have been utilized to identify TPVs in past studies, such as any PV anomaly that forms poleward of 60°latitude or stricter requirements about vortex residence time in high latitudes (Hakim and Canavan, 2005; Cavallo and Hakim, 2010). In general, regardless of the exact definition, TPVs and similar PV anomalies have been found to be common across the Arctic, especially favoring regions near high terrain or existing storm tracks (Hakim and Canavan, 2005; Kew et al., 2010; Portmann et al., 2021). TPVs have also been found to follow common pathways through and out of the Arctic (entering with midlatitudes quite frequently) and exhibit strong seasonality in their characteristics (e.g., winter TPVs tend to be stronger) and movement patterns (Hakim and Canavan, 2005; Bray and Cavallo, 2022).

TPV genesis and motion are strongly dominated by dynamic processes including advection, stretching, and deformation of the existing PV field (Cavallo and Hakim, 2012; Szapiro and Cavallo, 2018; Bray and Cavallo, 2022). On the other hand, the strengthening and weakening of the anomalies has been strongly tied to diabatic processes (Cavallo and Hakim, 2012, 2013). In particular, longwave cooling, physically related to vertical moisture gradients in the UTLS region and cloud top processes, is key to TPV intensification while latent heating from deep clouds and shortwave heating work to weaken vortices. TPVs often dissipate after being sufficiently weakened by diabatic processes (Cavallo and Hakim, 2013) or are destroyed dynamically, via shearing or reabsorption into the background flow (Portmann et al., 2021).

As positive PV anomalies, TPVs can serve as upper-level initiators of surface cyclogenesis when coupled with a surface baroclinic zone (Hoskins et al., 1985; Thorncroft et al., 1993). In particular, TPVs have been strongly associated with the development and maintenance of Arctic cyclones (ACs), which will be discussed in Section 1.2 (e.g. Simmonds and Rudeva, 2012, 2014; Tao et al., 2017b; Gray et al., 2021). Prior modeling studies have shown that TPV representation within a numerical model (as well as the representation of associated surface cyclones) is highly sensitive to the introduction or denial of new observations near the TPV, which may impact the mesoscale structure of the vortex or its interactions with surrounding features (Yamazaki et al., 2015; Johnson and Wang, 2021). Outside of the Arctic, TPVs have been linked to cold air outbreaks (Lillo et al., 2021; Biernat et al., 2021), the generation of jet streaks (Pyle et al., 2004), and other impactful weather events (Hakim et al., 1995; Hakim, 2000).

1.2 Arctic Cyclones

Arctic cyclones may be defined generally as any surface cyclone that moves north of 60°N at any point in its lifetime (Aizawa and Tanaka, 2016; Zhang et al., 2004; Gray et al., 2021), though some studies use either a stricter latitude criterion (Simmonds and Rudeva, 2014) or require longer development within the Arctic (Tanaka et al., 2012). ACs have been observed to have diameters in the thousands of kilometers (Aizawa and Tanaka, 2016; Simmonds et al., 2008). This size is much larger than Rossby deformation radius theory would suggest for circulations at such a high latitude, and the physical mechanisms behind this large size are not well understood (Emanuel and Rotunno, 1989). ACs have lifetimes on the order of days to weeks in many cases, also much longer than conventional theories would suggest (Yamagami et al., 2017; Tanaka et al., 2012; Aizawa and Tanaka, 2016; Simmonds and Rudeva, 2014).

Arctic cyclones either form locally within the Arctic or enter from the midlatitudes (Zhang et al., 2004; Sepp and Jaagus, 2011). Around half of all ACs (depending on the definition used) originate outside of the Arctic (Vessey et al., 2020; Zhang et al., 2004). During the winter, ACs often enter the Arctic along storm tracks in the North Atlantic and North Pacific Oceans (Simmonds et al., 2008) while in the summer, cyclones are relatively more likely to form in the Arctic along a band of high baroclinicity known as the Arctic frontal zone (AFZ; Crawford and Serreze, 2016). Summertime ACs generally have longer lifetimes but are weaker than their winter counterparts, mirroring trends in TPVs (Vessey et al., 2022; Zhang et al., 2004; Day and Hodges, 2018). Depending on the AC definition used, they may be either more common in the summer or occur at roughly the same frequency throughout the year (Day and Hodges, 2018; Crawford and Serreze, 2016). In addition, it is important to note that precise climatologies of AC locations and characteristics are highly sensitive to the specific tracking algorithm used (Vessey et al., 2020; Simmonds and Rudeva, 2014).

The structure of ACs throughout their lifecycle and physical controls on their development are still areas of active research. In many cases, Arctic cyclones have been observed to develop baroclinically as in the midlatitudes, especially along the AFZ (Simmonds et al., 2008; Crawford and Serreze, 2016; Tao et al., 2017a). Composites of ACs throughout their lifetimes reveal an asymmetric low-level structure and gradually lessening vertical tilt with time consistent with this midlatitude model of development (Clancy et al., 2022). The merging of different surface cyclones and linking of cyclones with TPVs then maintain these baroclinically-driven systems (Simmonds and Rudeva, 2014; Tao et al., 2017b; Yamagami et al., 2017). On the other hand, though, evidence exists to show that some ACs continue to strengthen even after assuming an equivalent barotropic and axisymmetric structure with a TPV, a process which is not yet well understood (Aizawa and Tanaka, 2016; Tao et al., 2017b). Winter ACs tend to occur in environments with stronger antecedent baroclinicity and a less defined tropopause anomaly (more closely following the midlatitudes), while the reverse is true in the summer (Vessey et al., 2022). Overall, summer Arctic cyclones often follow a two stage development cycle: initial development as a baroclinic disturbance followed by a transition to an axisymmetric structure that is strongly linked to the TPV above (Vessey et al., 2022; Yamagami et al., 2017; Tao et al., 2017b). A summary of these stages of summer AC development is provided in Figure 1.2.

Arctic cyclones are a major source of heat and moisture transport into the Arctic and thus play a crucial role in the Arctic climate system (Villamil-Otero et al., 2018; Sorteberg and Walsh, 2008). Due to a combination of this heat transport and mechanical forcings (e.g., wind and waves), ACs help to govern sea ice extent on synoptic to subseasonal timescales (Schreiber and Serreze, 2020; Blanchard-Wrigglesworth et al.,

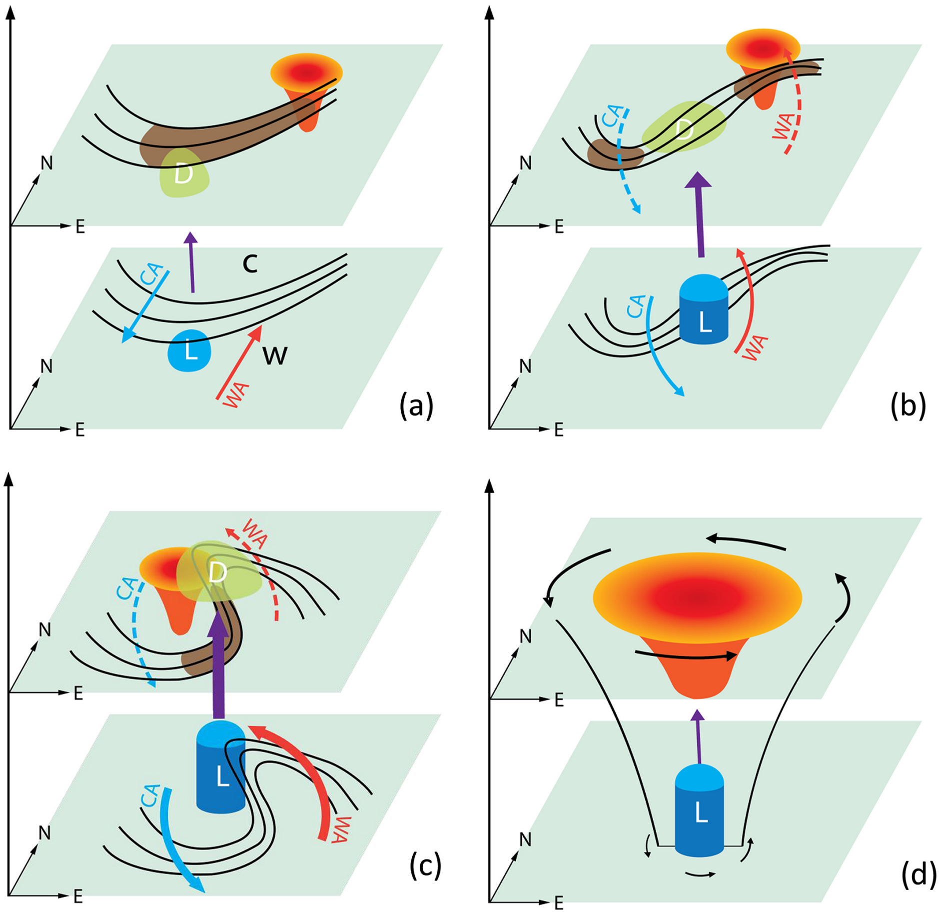


Figure 1.2: Schematic of linkages between upper and lower-level processes throughout the development of the Great Arctic Cyclone of 2012. The orange cone represents the TPV, the brown region represents and jet streak, the green region indicates divergence, and the blue cylinder represents the surface cyclone. Borrowed from Tao et al. (2017a).

2022). Links between ACs and sea ice coverage have been previously shown in individual case studies (Simmonds and Rudeva, 2012; Zhang et al., 2013) and long-term climatological studies (Simmonds and Keay, 2009; Screen et al., 2011; Wang et al., 2020). Because of their connection to sea ice, ACs also factor heavily into climate change projections. Some prior work has suggested that summertime ACs may increase in intensity in the future due to higher moisture availability and baroclinicity, which would in turn impact projections of sea ice coverage (Parker et al., 2022; Day and Hodges, 2018). The complex dynamical feedbacks between ACs and sea ice on a range of time scales makes these projections especially uncertain at the present (Valkonen et al., 2021).

1.3 Observing System Simulation Experiments

Observing system simulation experiments (OSSEs) are a type of modeling experiment used to study the impact of novel observations within a modeling and data assimilation (DA) system (Hoffman and Atlas, 2016; Masutani et al., 2013). OSSEs begin with a nature run (NR), which is a long-term free forecast produced by an accurate, high-resolution numerical model. This NR will be taken as the true state for the remainder of the OSSE, and so it is important that it accurately represents real-world climatologies (Masutani et al., 2013). From this NR, using appropriate forward operators, observations are simulated with real-world characteristics (i.e., the same variables are "observed" in the same locations as in the real world). Additionally, several sets of novel observations that are of interest to the study at hand are simulated. These observations are then provided to an experimental modeling and DA system, which should be known to accurately ingest real-world observations. A control simulation is performed in which only the simulated real-world observations are used, followed by a series of experiments in which the novel observations are provided in addition to the

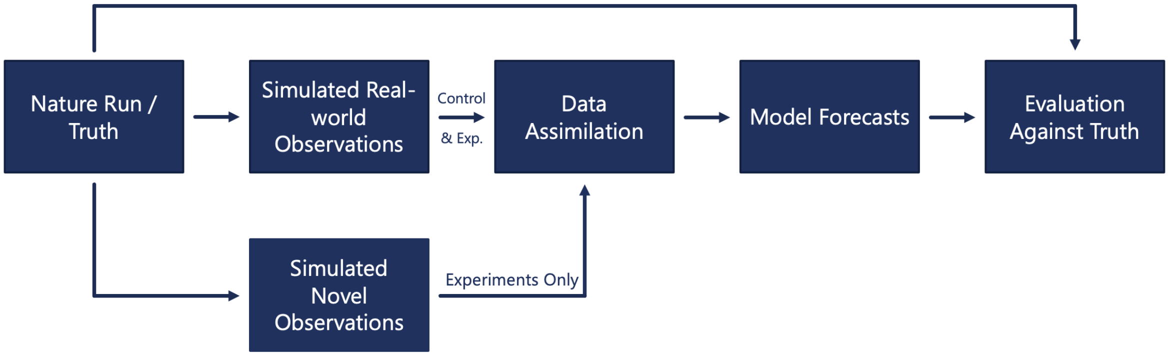


Figure 1.3: Schematic representing the basic components and workflow of an OSSE.

real-world ones. The results of these experiments can then be compared to both the control run and the NR truth; a schematic summarizing this process is provided in Figure 1.3.

OSSEs have traditionally been used to address a variety of scientific and technical goals. Chiefly, OSSEs are used to investigate the impact of new observing strategies in order to make design choices about new observation platforms (e.g., what instrumentation should be on a satellite to produce the most forecast improvement) before it is launched (Zeng et al., 2020; McCarty et al., 2021). Additionally, they have been used in operational settings to evaluate and tune modeling and DA systems (Zeng et al., 2020; Kleist and Ide, 2015). OSSEs have also proved beneficial for some less operational applications such as testing new techniques for data assimilation (e.g., Zhang et al., 2018) or testing the impact of idealized observations over data sparse or under-studied regions (e.g., English et al., 2018; Peevey et al., 2018). Operationally minded OSSEs tend to test the new observations over long periods of time and under different conditions in order to build up statistical confidence in the results; scientifically oriented studies, on the other hand, often perform "quick OSSEs" on a single case study to study observation impacts on finer scales (Zeng et al., 2020). While such experiments tend to be more computationally feasible than extensive climatological OSSEs, the limited sample size will mean that the inherent variability in real-world weather events is not captured. These quick OSSEs often use a targeted observation approach, attempting to place the simulated observations in an area that may provide special forecast benefit. The placement of these observations can be done algorithmically, perhaps via ensemble techniques (e.g., Peevey et al., 2018) or manually using physical reasoning (Majumdar, 2016). Each of these techniques has the potential to miss areas that would in fact be sensitive to error growth, and prior studies have produced conflicting results on the benefits created by targeted observations in real-world cases (Majumdar, 2016).

When conducting an OSSE, it is necessary to consider real-world sources of error and how they will (or will not) be incorporated into the OSSE framework. For example, errors between operational forecasts and the real world arise from a variety of sources (initial state errors, numerical and parameterization errors, boundary condition errors, etc.), and operational OSSEs should attempt to reproduce this level of error between the experimental forecast system and the NR (Halliwell Jr et al., 2014). In many cases, this goal is achieved by using an entirely separate modeling and DA system from the one that produced the NR to conduct the experiments (e.g., Masutani et al., 2013; Peevey et al., 2018). In other cases, though, replicating the full extent of model errors may be less necessary or adequate preexisting nature runs may be difficult to find. In such cases, OSSEs are conducted with "fraternal" (the same basic forecast system in a different configuration; e.g., Halliwell Jr et al., 2014) or "identical" (an identical forecast system to the one that produced the nature run; e.g., Zhang et al., 2018) twin setups, both of which will lead to smaller errors between experimental forecasts and the true state than would be expected in the real world.

Moreover, real-world observations include instrument, representativeness, and forward operator errors (Privé et al., 2021). Fully realistic OSSEs attempt to replicate these errors by adding noise to observations drawn from the NR based on the error characteristics of the instrument being simulated (Privé and Errico, 2013; Privé et al., 2021). The effect of including these observation errors on forecast skill has been shown to be important, but comparatively small relative to the scale of model errors (Privé and Errico, 2013). Studies which do not add errors to simulated observations are said to use "perfect" observations, the effects of which have been quantified by many previous studies (Boukabara et al., 2018; Hoffman and Atlas, 2016; Masutani et al., 2010). Generally, using perfect observations will increase forecast skill within the OSSE, easing physical interpretations at the expense of real-world applicability (McCarty et al., 2021; Boukabara et al., 2018).

1.4 Motivation and Goals

Despite the fact that extreme weather events in the Arctic are highly impactful for human communities and economic interests like shipping, fishing, and resource extraction, forecasting and observation systems for the region are severely underdeveloped (Jung et al., 2016; Melia et al., 2016). Similarly, events like Arctic cyclones can play a role in speeding up coastal erosion and ice movement, which are critical for indigenous communities in the Arctic (Nielsen et al., 2022). Especially in light of changing weather patterns and increasing human activity in the region as the climate warms, improving forecasts is of high importance (Stephenson et al., 2013). Due to their connections with winds, precipitation, and sea ice coverage around the Arctic and throughout the year, Arctic cyclones are a prime target for study (Villamil-Otero et al., 2018; Screen et al., 2011). As discussed above, the innate physical processes driving AC longevity and size are not yet well understood. Current weather models possess some medium-range predictive skill with regard to ACs (Yamagami et al., 2018). In particular, forecasts of cyclone strength are lacking, which has been shown to be related to upper level forcings like TPVs (Yamagami et al., 2018; Tao et al., 2017b).

In the present study, we will use an OSSE to investigate the effects of direct observations of a TPV and its surrounding environment on forecasts of an associated Arctic cyclone. Prior studies with real-world data have shown that observations near the TPV can produce significant forecast improvements throughout the troposphere on medium range time scales (Johnson and Wang, 2021; Yamazaki et al., 2015). The OSSE framework affords additional flexibility for testing the influence of different types of observations over specific areas to build on this previous work. In particular we will address two sets of physical questions about the TPV: 1) How does the mesoscale structure of the TPV and surrounding PV field impact cyclone development, and 2) How do the impacts of temperature versus humidity observations near the TPV differ?

We hypothesize that AC development and maintenance is sensitive not only to the general location and strength of the TPV but also to the smaller-scale shape of the TPV and surrounding PV features, like smaller satellite vortices and PV filaments. These surrounding features may help to explain AC longevity, as they could continue to induce baroclinic development even after the surface cyclone has become vertically stacked with the primary TPV. Further, we expect that although temperature observations are likely to provide more corrections to the TPV initially, observations of moisture will prove to play a larger long-term role due to the relationships between longwave cooling and PV creation. In order to best address these physical process hypotheses, we will take a relatively idealized approach to generating simulated observations. In this way, the OSSE results presented here should provide a sort of upper predictability limit for the experimental modeling system given excellent initial representation of the TPV region. Still, in addition to uncovering important physical processes, we hope that the results of these experiments will prove useful for understanding how different observation types are ingested by the DA system and planning future real-world observing campaigns.

Chapter 2

Data & Methods

2.1 OSSE Design

2.1.1 The Nature Run, Data Assimilation System, and Forecast Model

For this OSSE, we take the European Centre for Medium-Range Weather Forecasts (ECMWF) Cubic Octahedral grid Nature Run (ECO1280) as the true state (Hoffman et al., 2018). ECO1280 is a 14 month free forecast running from 30 September 2015 through 30 November 2016 and generated using the version of the ECMWF's global Integrated Forecast System (IFS) that was operational in 2016. This version of the IFS is a spectral hydrostatic model with roughly 9-km resolution in the transform grid on which physics calculations occur. The model includes 137 hybrid vertical levels, and all operationally relevant output variables are available for the entire 14 month period. Although ECO1280 is a free forecast which will not represent reality after the first couple of forecast weeks, the boundary conditions (e.g., sea surface temperatures and sea ice) are prescribed from operational analyses throughout the integration, lending ECO1280 one connection to the real world.

For this study, we utilize an experimental DA and modeling system that is entirely separate from the system that produced ECO1280 in order to introduce realistic errors between the NR and experimental forecasts. To assimilate observations simulated from the NR, we take an ensemble Kalman filter (EnKF) approach, which attempts to optimally combine observations with the background state and employs a finite ensemble of forecasts to estimate the background error covariance. Specifically, we use the ensemble adjustment Kalman filter (EAKF) developed by the Data Assimilation Research Testbed (DART) at the National Center for Atmospheric Research (NCAR; Anderson, 2001). The DART EAKF is a deterministic EnKF related to the ensemble square root filter that updates the ensemble mean and then adjusts the perturbations of the individual ensemble members in such a way that the final analysis error covariance converges to the hypothetical Kalman filter values. In order to combat filter divergence from spurious correlations within the limited ensemble, we use DART to apply Gaspari-Cohn covariance localization with the covariance set to zero past a radius of 1275 km in the horizontal and 0.4 scale heights in the vertical. Furthermore, filter divergence issues can arise from an inadequate representation of model errors within the ensemble. To address this problem, we apply prior covariance inflation in which the background ensemble perturbations are increased by a given inflation factor before assimilation. In particular, we utilize the spatially varying adaptive inflation technique described in Anderson (2009), in which the inflation factor applied at a given location evolves with each DA cycle depending on the difference between the background and observations.

To integrate the ensemble forward in time after each assimilation cycle, we use the Model for Prediction Across Scales (MPAS), also developed at NCAR (Skamarock et al., 2012). MPAS is a state-of-the-art non-hydrostatic atmospheric model which discretizes fields onto a unstructured centroidal Voronoi mesh (which allows for quasi-uniform grid cell sizes around the globe and local resolution refinement) with C-grid staggering. In particular, the MPAS configuration employed here uses a variable mesh with 15-km mesh spacing over the Arctic smoothing to 60-km mesh spacing over the rest of the globe (Fig. 2.1). MPAS uses a hybrid terrain following coordinate, and we specify 55 total vertical levels between the surface and a model top of 30 km. In general then, the resolution of our experimental system will be lower than that of the NR, as desired in an OSSE. The physical parameterizations utilized by MPAS are shown in Table

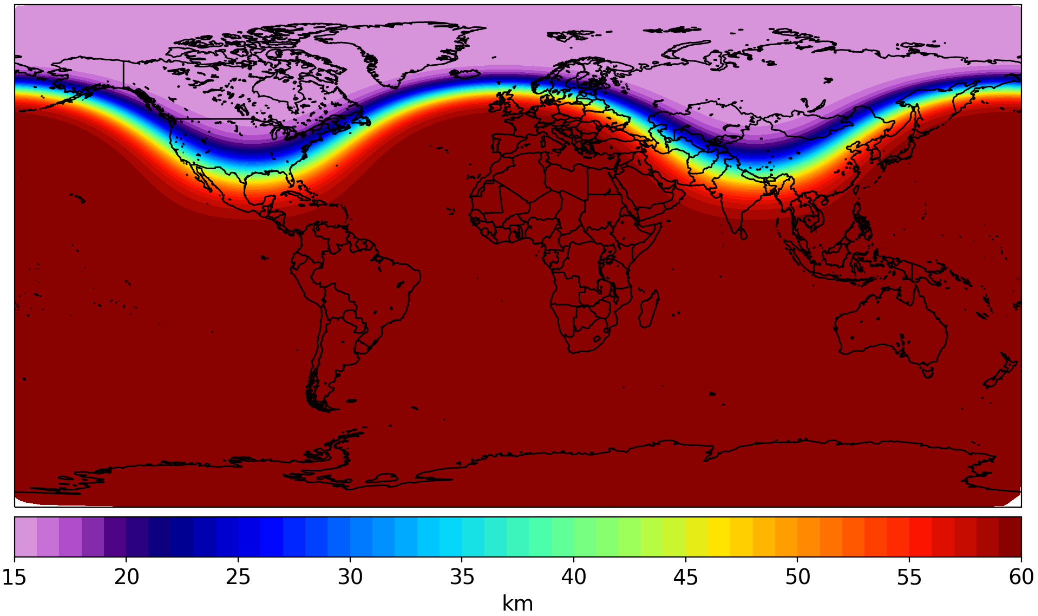


Figure 2.1: Approximate grid spacing (km) across the globe for the 60-15 km MPAS mesh used for all experiments.

2.1. These packages are the default MPAS mesoscale physics suite with the exception of the Thompson 2-moment scheme, which was chosen to more accurately represent mixed-phase processes in the Arctic. Each of these parameterizations is well-established in atmospheric research contexts. Sea surface temperatures and sea ice coverage are updated once daily at 00 UTC from National Centers for Environmental Prediction (NCEP) analyses. Note that the boundary conditions used in our experiments will thus roughly match those used in the ECO1280. When MPAS-DART cycling occurs, observations are assimilated every six hours (00, 06, 12, and 18 UTC) with an MPAS forecast in between.

2.1.2 Simulated Observations and Experiment Configuration

Conventional observations only (i.e., excluding satellite radiances and other observations with complex forward operators) are simulated from ECO1280 and assimilated in this study. In particular, PREPBUFR files, which contain the processed conventional

Process	Parameterization
Longwave Radiation	RRTMG (Iacono et al., 2008)
Shortwave Radiation	RRTMG (Iacono et al., 2008)
Boundary Layer	Yonsei University (YSU) (Hong et al., 2006)
Surface Layer	Monin-Obukhov (Janić, 2001)
Land Surface	Unified Noah LSM (Chen and Dudhia, 2001)
Microphysics	Thompson 2-moment Non-Aerosol Aware (Thompson et al., 2008)
Cumulus	New Tiedtke (Zhang et al., 2011)

Table 2.1: Physical parameterization schemes used in MPAS for all experiments.

observations used in NCEP analyses, are used as a template for simulating all existing observations (NCEP, 2008). An example of the conventional observations available in the PREPBUFR files from 16 August 2016 at 12 UTC is provided in Figure 2.2. Note that in operational settings, satellite radiances account for the vast majority of observations assimilated by modern models (though with notably more limited coverage in the Arctic). In the present study, we do not include these satellite observations to more directly examine the impact of our novel targeted observations. Observations are simulated from ECO1280 with the exact times, locations, and basic variables specified in the PREPBUFR files (which, in the case of conventional observations, is simply a matter of spatial and temporal interpolation), and observation error variances are specified exactly as in the PREPBUFR files (and by extension, in a prior version of NCEP's Global Data Assimilation System). Certain dense observations (aircraft measurements and geostationary satellite winds) are combined or "super-obed" to one observation per MPAS grid cell at specified heights to avoid known issues with correlated observations.

Additionally, any surface observations with a specified elevation that is more than 150 m away from the model value at the same point are removed.

As discussed in Section 1.3, observations in an OSSE are generally simulated with real-world error characteristics for maximum accuracy; however, because the present study aims to address scientific questions over operational ones, perfect observations with no added error are simulated from ECO1280. As a result, the impact of the observations on forecast skill will likely be greater in our experiment that would occur with similar real world observations (Hoffman and Atlas, 2016; McCarty et al., 2021). Though this method limits the direct real-world applicability of the results, it still fulfills the goals of the study and may even simplify the interpretation of results (English et al., 2018; Peevey et al., 2018). Moreover, it is important to point out that because MPAS and ECO1280 are different models run at different resolutions connected only by linearly interpolated observations, some inherent random observation is still introduced (Riishojgaard et al., 2012).

A broad overview of the OSSE workflow used in this study is found in Figure 2.3. When creating an ensemble for use with an EnKF, it is essential to have a large number of members with enough spread between them that the background error covariance is correctly approximated (Houtekamer and Zhang, 2016). To this end, we utilize an 84 member ensemble throughout the OSSE. The ensemble is initialized using real-world forecasts from the 21 members of NCEP's Global Ensemble Forecast System (GEFS). To produce 84 members, we use a time lag technique in which 24, 48, 60, and 72 hour forecasts from all 21 members valid at the same time are collected and transferred onto the MPAS grid (Kumar and Hoerling, 2000). Each of the members is advanced 12 hours from this point in order to spin up MPAS physical quantities and increase the degree of ensemble spread. Next, the ensemble is cycled via MPAS-DART with existing conventional observations simulated from ECO1280 until the ensemble contains substantial spread and is satisfactorily representative of the NR state instead

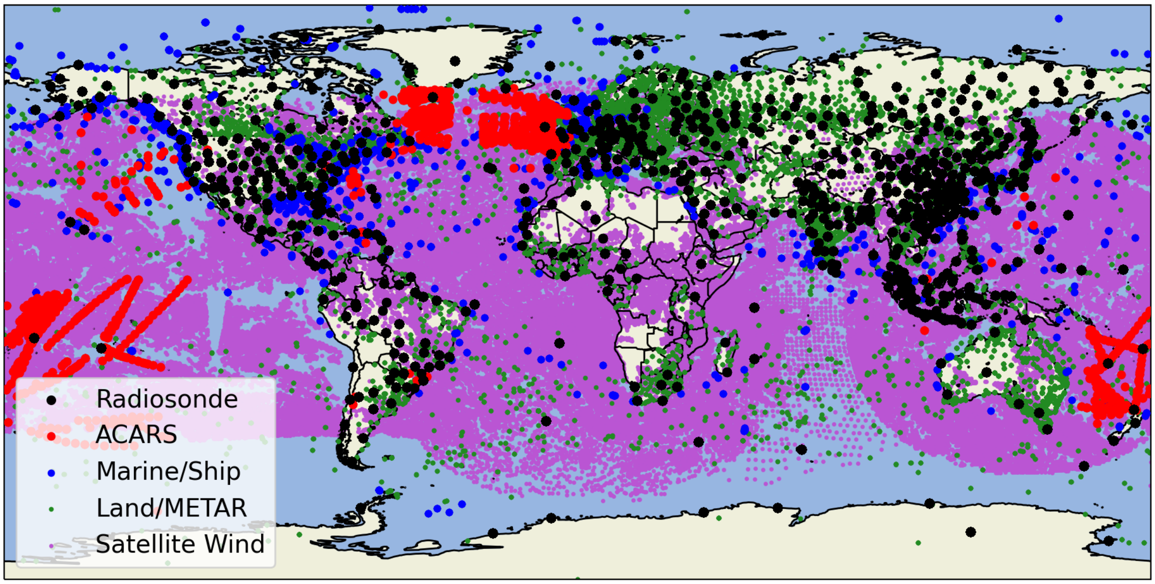


Figure 2.2: Locations of all conventional observations provided in the PREPBUFR file for 16 August 2016 at 12 UTC. Includes radiosonde launch points, aircraft data (ACARS), marine buoy and ship measurements, surface station and METAR measurements, and geostationary satellite winds.

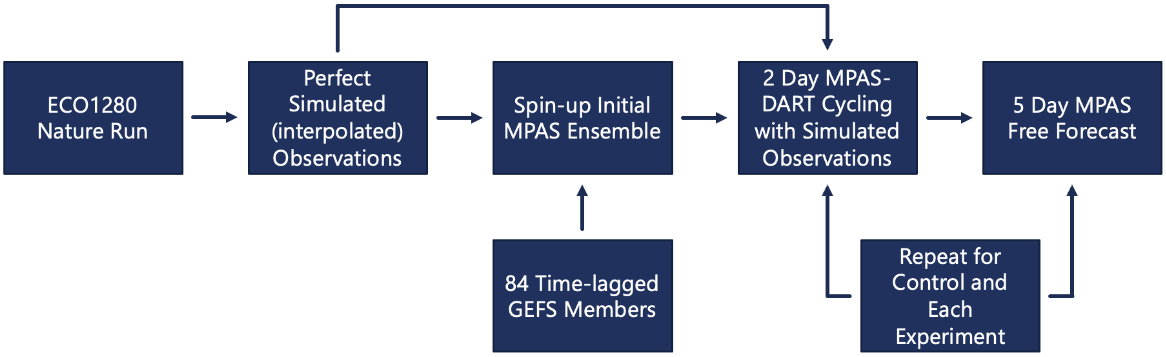


Figure 2.3: A flowchart illustrating the workflow of the OSSE as described in the text with the control and experiments as defined in Table 2.2.

of the real-world state (a total of four days of cycling; details in Section 3.3). The resulting ECO1280 ensemble is then used to initialize the individual OSSE runs, each of which consists of two days of cycling with simulated observations followed by a five day free forecast. The control run continues to cycle with only existing conventional operations for two days before the free forecast. The four main experiments cycle with existing observations plus additional simulated dropsondes (described below). A fifth sub-experiment undergoes one day of cycling with no free forecast afterwards. A summary of the individual experiments with a description of the additional simulated dropsondes included can be found in Table 2.2.

Experiment	Description
Control	Simulated conventional observations only
Exp1 (T)	Add simulated dropsondes over limited area measuring only T
Exp2 (q)	Add simulated dropsondes over limited area measuring only q
Exp3 (Tq)	Add simulated dropsondes over limited area measuring T and q
Exp4 (Tq+)	Add simulated dropsondes over large area measuring T and q
Exp 5 (Tq*) - Only run through 17 August 00Z)	Add simulated dropsondes over limited area measuring T and q with reduced q observation error

Table 2.2: Description of the observations included in each of the individual experiments conducted as part of the OSSE. The names used in the left column will be used in all future figures to distinguish results from each experiment.

Across all non-control experiments, dropsondes are simulated from ECO1280 and assimilated into the ensemble as though they are radiosonde measurements (i.e., with radiosonde observation errors, which vary with height for temperature and have a constant value of 20% relative humidity for moisture observations). All of the simulated dropsondes begin at roughly 200 hPa and continue straight to the surface, with observations taken at the majority of available ECO1280 levels to achieve around 20 hPa spacing between individual observations. As discussed earlier, this OSSE is intended to be idealized, and so the dropsonde locations do not follow a realistic flight path; instead, regularly spaced grids of dropsonde observations are centered over the ECO1280

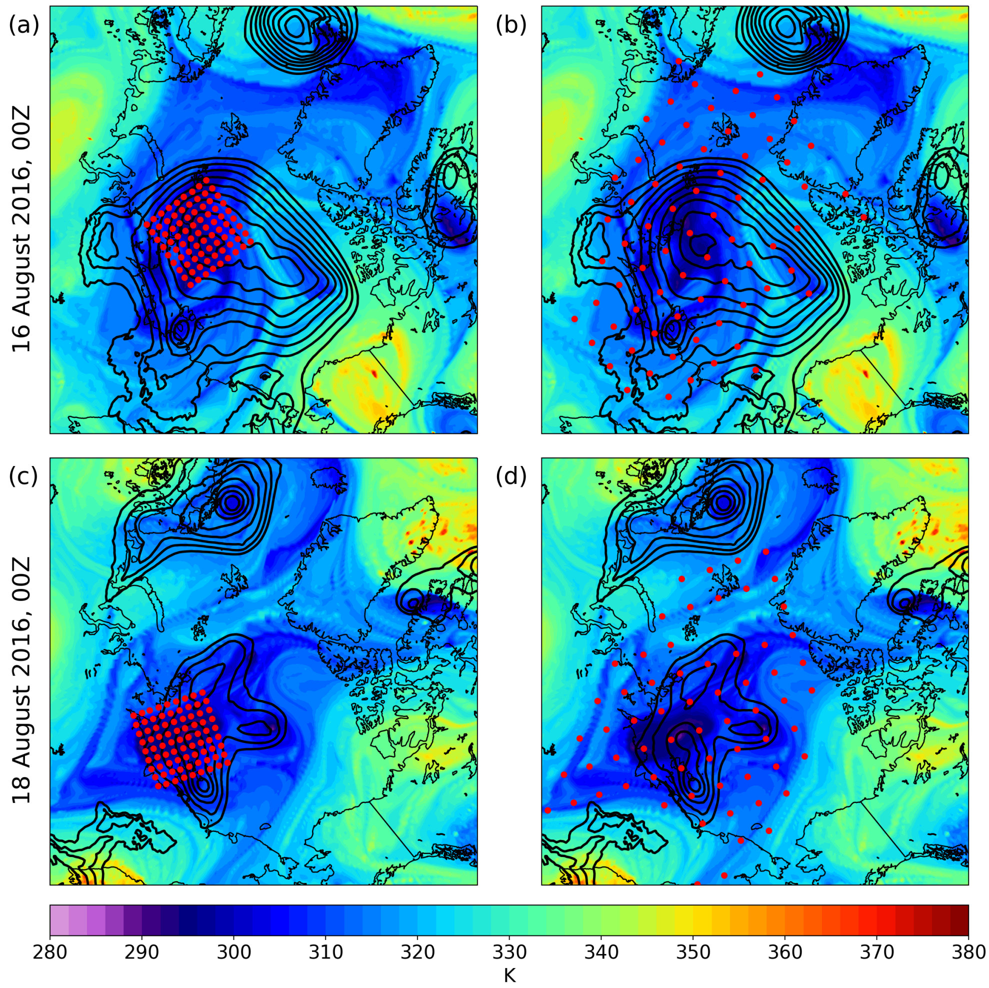


Figure 2.4: Patterns of simulated dropsondes for OSSE experiments, with locations calculated as described in the text. (a) Dropsondes simulated for Exp1 (T), Exp2 (q), Exp3 (Tq), and Exp5 (Tq*) on 16 August 2016 at 00 UTC. (b) Dropsondes simulated for Exp4 (Tq+) on 16 August 2016 at 00 UTC. (c) As in (a) but for 18 August 2016 at 00 UTC. (d) As in (b) but for 18 August 2016 at 00 UTC. Also included are the corresponding TRPT (fill; K) and MSLP (contours; hPa; every 2.5 hPa starting at 1000 hPa) fields.

TPV core at each time step for which observations are simulated (the specifics of the case will be discussed in more detail in Section 3.1). Exp1 (T), Exp2 (q), Exp3 (Tq), and Exp5 (Tq*) use a grid with 1000 km sides and 125 km spacing between individual dropsondes (Fig. 2.4a,c). Exp5 (Tq*) is simply a rerun of the first day of cycling in Exp3 (Tq) with the humidity observation error halved; the motivation for this will be discussed in Section 4.1. Exp4 (Tq+), on the other hand, uses a grid with 4000 km sides, 400 km spacing between individual dropsondes, and an additional stipulation that the dynamic tropopause potential temperature (TRPT) at the point must be less than 320 K to limit sampling to high PV features (Fig. 2.4b,d).

2.2 Verification Methods

In order to summarize impacts on forecast skill across various fields and vertical levels, we use a moist energy norm, referred to hereafter as the total moist energy error (TMEE). Following previous dry and moist energy norm definitions in Hamill et al. (2013) and (Marquet et al., 2020), we define the TMEE in this study as:

$$\mathrm{TMEE} = \left\{ \frac{1}{2} \int_A \left[\begin{array}{c} \frac{1}{4}\left({u'_{250}}^2 + {v'_{250}}^2 + \frac{c_p}{T_r}{t'_{250}}^2 + \frac{L_v^2}{c_p T_r}{q'_{250}}^2 \right) + \\ \frac{1}{4}\left({u'_{500}}^2 + {v'_{500}}^2 + \frac{c_p}{T_r}{t'_{500}}^2 + \frac{L_v^2}{c_p T_r}{q'_{500}}^2 \right) + \\ \frac{1}{4}\left({u'_{700}}^2 + {v'_{700}}^2 + \frac{c_p}{T_r}{t'_{700}}^2 + \frac{L_v^2}{c_p T_r}{q'_{700}}^2 \right) + \\ \frac{1}{4}\left({u'_{10m}}^2 + {v'_{10m}}^2 + \frac{c_p}{T_r}{t'_{2m}}^2 + \frac{L_v^2}{c_p T_r}{q'_{2m}}^2 \right) + R_d T_r \left(\frac{p'}{P_r} \right)^2 \end{array} \right] \right\}^{\frac{1}{2}}, \quad (2.1)$$

where u', v', t', and q' represent differences between the ensemble mean forecast fields and the NR truth for the zonal wind, meridional wind, temperature, and specific humidity at the given levels. Further, p' is the difference between the ensemble mean surface pressure and the NR, T_r is the reference temperature (300 K), c_p is the specific heat of dry air at constant pressure (1004 J kg^{-1} K^{-1}), L_v is the latent heat of vaporization (2.5×10^6 J kg^{-1}), R_d is the dry air gas constant (287 J kg^{-1} K^{-1}), and P_r is the reference pressure (1000 hPa). To gauge statistical significance between ensemble mean fields from different experiments, a Welch's t-test is used (Wilks, 2011). When assessing the significance of a difference between an ensemble mean and the truth, the difference is considered significant if the NR value is below the 5th percentile of ensemble values or above the 95th percentile.

To objectively track TPVs in both ECO1280 and all experimental members, we use TPVTrack, a python-based feature tracking software (Szapiro and Cavallo, 2018). TPVTrack traces TPVs by creating spatial objects based on vorticity around local minima of TRPT and searching for overlap between objects at successive time steps. Using the main TPV track calculated from ECO1280, we define a moving verification region that will be used to display information and average statistics throughout the remainder of the study (Fig. 2.5). At each analyzed time step, the verification region is centered on the ECO1280 TPV and includes all points within a 1000 km radius. This distance is roughly double the Rossby radius of deformation and is intended to capture the entirety of the TPV, AC, and immediate surrounding features at all times. The TPV core verification region is defined as all points within this verification region with a TRPT of less than 300 K. To track ACs, we use the TempestExtremes software, which identifies minima in the mean sea-level pressure (MSLP) field with closed contours and tracks them with a distance threshold (Ullrich et al., 2021).

Using these tracks, we identify best, mean, and worst performing ensemble members from all experiments using two approaches. In the first method, standardized TPV error, we take with the error between TPV location and strength in an individual ensemble member relative to ECO1280, standardize each of these errors relative to the rest of the members in that experiment, and then sum the standardized errors over the full five-day free forecast period to produce a single value of TPV forecast accuracy (more negative values represent the best members and vice versa). In the second method, TPV-AC distance, we calculate the distance between the AC and TPV in an ensemble member, calculate the error of this distance relative to the TPV-AC spacing in ECO1280, and then sum that error over the full five-day free forecast period (lower distance errors represent the best members).

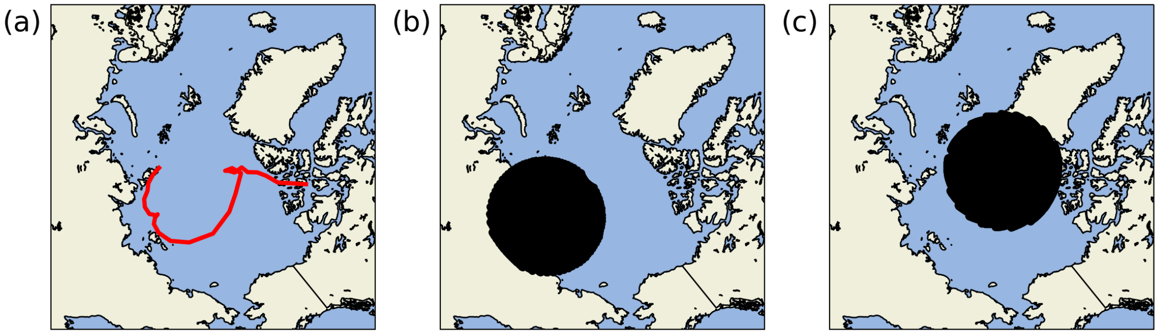


Figure 2.5: (a) The track of the main TPV of interest within the ECO1280 nature run from 16 August 2016 at 00 UTC to 23 August 2016 at 00 UTC. (b) The 1000 km radius moving verification region centered on the ECO1280 TPV track on 18 August 2016 at 00 UTC. (c) As in (b) but on 21 August 2016 at 00 UTC.

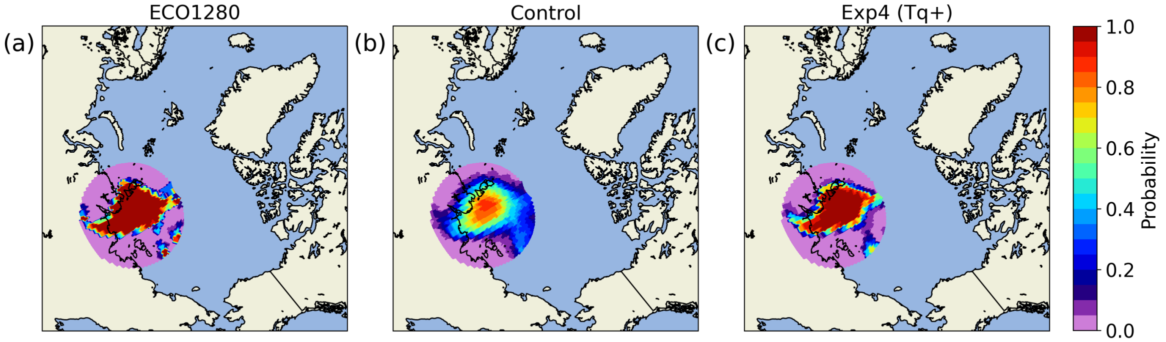


Figure 2.6: (a) Tropopause neighborhood probabilities (NP) within the moving verification region for the ECO1280 nature run on 17 August 2016 at 00 UTC. (b) Ensemble tropopause neighborhood probabilities (NEP) within the moving verification region for the control run on 17 August 2016 at 00 UTC. (c) As in (b) except for Exp4 (Tq+).

To more easily examine the different spatial structure of TPVs among experiments and compared to ECO1280, we use a neighborhood-based ensemble verification technique (Roberts and Lean, 2008; Ebert, 2009; Schwartz et al., 2010). This neighborhood approach examines TRPT within the moving verification region. The threshold TRPT is defined as the average TPV edge value in ECO1280 over the full seven day forecast period (around 300 K), as calculated by TPVTrack. For both ECO1280 and each experimental member forecast, grid points within the verification region with a TRPT less than or equal to this threshold are given a binary probability of 1, while other points are given a value of 0. A neighborhood with a radius of 50 km is then established around each grid point, and the binary probabilities of all points that fall within that radius are averaged to create a neighborhood probability (NP) value for the point. Sensitivity tests showed little qualitative change in the results for varied radius sizes between 25 and 100 km, though larger neighborhoods produce higher skill scores. For the experimental ensembles, these neighborhood probabilities are further averaged across all 84 ensemble members to create neighborhood ensemble probabilities (NEP).

The results of this process are spatial maps covering the verification region at each time step with N individual grid point neighborhood probabilities for ECO1280 ($\mathrm{NP}_{\mathrm{NR}}$) and N individual grid point neighborhood ensemble probabilities for each experiment ($\mathrm{NEP}_{\mathrm{Exp}}$). Examples of these probability maps for ECO1280, the control, and Exp4 (Tq+) are provided in Figure 2.6. These probability maps can be used to calculate the fractions skill score (Schwartz et al., 2010). First, the fractions Brier score (FBS) is calculated as:

$$\mathrm{FBS} = \frac{1}{N}\sum_{N}\left(NEP_{Exp} - NP_{NR}\right)^2. \tag{2.2}$$

The fractions skill score (FSS) is then defined as:

$$\mathrm{FSS} = 1 - \frac{FBS}{\frac{1}{N}\left(\sum_{N} NEP_{Exp}^2 + \sum_{N} NP_{NR}^2\right)}, \tag{2.3}$$

where 1 is a perfect score and 0 is a no-skill forecast. Following Ebert (2009), a skill threshold FSS above which forecasts are considered useful can be defined assuming a uniform forecast across the verifcation region as:

$$\mathrm{FSS}_{\mathrm{skill}} = 0.5 + \frac{f_{NR}}{2} \tag{2.4}$$

where f_{NR} is the fraction of the verification region less than or equal to the TRPT threshold within ECO1280. Uncertainty in the FSS is assessed using a bootstrap resampling method where 10000 random subsets of ensemble members are selected with replacement and used to form a distribution of possible FSS values.

Chapter 3

Case Overview, Nature Run Verification, and Ensemble Initialization

3.1 Case Selection and Synoptic Overview

Because we are specifically interested in studying the physical processes governing summertime, Arctic-origin ACs, we search for potential cases in ECO1280 during August 2016. A suitable case for the OSSE is identified starting around 16 August 2016 at 00 UTC; a synoptic overview of the case is provided here. The AC of interest forms along the northern coast of central Siberia on the southern flank of a large but decaying cyclone that originated in the midlatitudes (Fig. 3.1a). This existing AC is associated with a large TPV and several smaller PV anomalies, which will become key upper level forcings for our AC of interest. Over the next two days, the AC continues to develop and move along the Siberian coast, maintaining an offset of around 1000 km from the main TPV (Fig. 3.1b,c). This type of cyclogenesis along the AFZ associated with a TPV is common in real-world summers, which makes this an excellent case for the OSSE to examine (Crawford and Serreze, 2016).

Starting after 18 August at 00 UTC, the AC quickly moves off over the central Arctic and rapidly intensifies as it does so, reaching a maximum strength of near 976 hPa around 21 August 00 UTC (Fig. 3.1d-f). The primary TPV trails the cyclone by several hundred km while moving across the Arctic, but gradually the two features become vertically stacked as the cyclone peaks in strength. A tongue of higher PV air with several smaller embedded features is also visible wrapping around the AC on 21 August (Fig. 3.1f). Beyond this point, the AC and TPV take on an equivalent

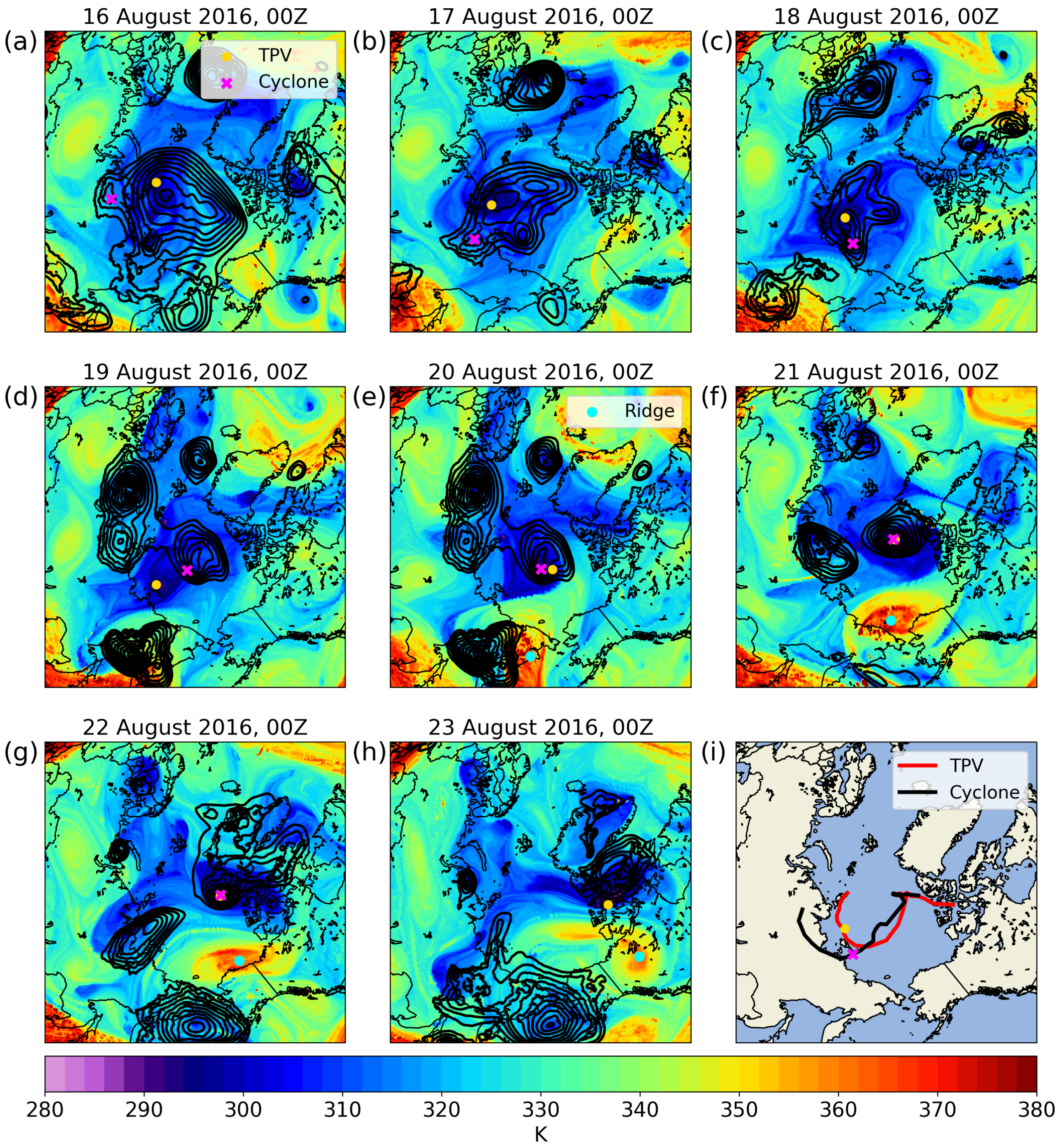


Figure 3.1: (a)-(h) Maps of TRPT (fill; K) and MSLP (contours; hPa; every 2.5 hPa starting at 1000 hPa) for ECO1280 each day at 00 UTC throughout the OSSE period. At each time, the TPV is marked with a yellow circle and the AC is marked with a pink X. The Rossby wave ridge of interest is marked with a cyan circle at relevant time steps. (i) ECO1280 primary TPV and AC tracks. The yellow dot and pink X denote the location of the TPV and AC, respectively, at the end of the cycling period (18 August 00 UTC).

Figure 3.2: A schematic displaying the simulated dates within the ECO1280 nature run during which each phase of the OSSE occurs.

barotropic structure and move into the Canadian archipelago (Fig. 3.1g,h). Other features of note include the Rossby wave ridge that builds into the Arctic and breaks as the AC intensifies and the connected region of high PV air that becomes wrapped into the AC-TPV circulation starting on 20 August.

The location of the system in this case is particularly useful for our case. In the days prior to 16 August (not shown), the preceding system moves over the observation sites of northern Europe, which should help pull the GEFS-based initial ensemble satisfactorily towards ECO1280 in this region during the four-day spin-up period, which we begin on 12 August at 00 UTC (Fig. 3.2; Fig. 2.2). From 16 August at 00 UTC when experimental cycling begins onward, the system passes very few upper-air observing locations. So, the control run should see minimal additional correction in TPV structure while the experiment runs assimilate the dropsondes. The free forecast period begins on 18 August at 00 UTC as the surface cyclone moves off the Siberian coast and begins its rapid intensification; the expectation is that the added dropsondes will allow this intensification period to be better forecasted. The forecast period ends when the system moves into the Canadian Archipelago (where tracking of the surface cyclone becomes more difficult) and begins to weaken.

3.2 Nature Run Verification

Although the ECMWF IFS a state-of-the-art model, and ECO1280 has been previously tested in the tropics and midlatitudes, the NR has yet to be verified relative to climatology specifically in the Arctic region. To address this before proceeding with the OSSE, we will examine both raw atmospheric fields and derived tracks for features of interest (TPVs and ACs) in August 2016 in ECO1280 relative to a real world climatology. To construct this climatology, we use data from 30 years worth of Augusts (1991-2020) taken from the ERA5 global reanalysis (Hersbach et al., 2020). TPV tracks for both ECO1280 and the ERA5 climatology are calculated with the same configuration of TPVTrack, and AC tracks are calculated using the same configuration of TempestExtremes.

At the 500 hPa level in the Arctic (poleward of 60°N), ECO1280 geopotential heights and temperatures fall well within the bounds of the ERA5 climatology (Fig. 3.3a,b). Additionally, ECO1280 falls on the more cyclonic end of the climatology, which is the type of real-world large-scale environment that we are interested in studying. At the 250 hPa level, while geopotential heights match the climatology, temperatures tend to be 4-5 K cooler in ECO1280 than in ERA5 (Fig. 3.3c,d). This likely indicates a high bias in tropopause heights in ECO1280 relative to ERA5 (which, notably, was produced by a slightly older version of the IFS). Because the bias is not too large and TPVs are well-represented (see below), we do not expect this bias to impact the results significantly, though our initial ensemble may be warmer than ECO1280 in the lowermost stratosphere.

Despite the slight cold bias in the lower stratosphere, the distribution of vortex minimum potential temperatures (one measure of TPV strength) for all ECO1280 TPVs closely follows ERA5 climatology (Fig. 3.4a). ECO1280 TPVs have slightly shorter lifetimes than their ERA5 counterparts, which is likely a function of the TPVTrack

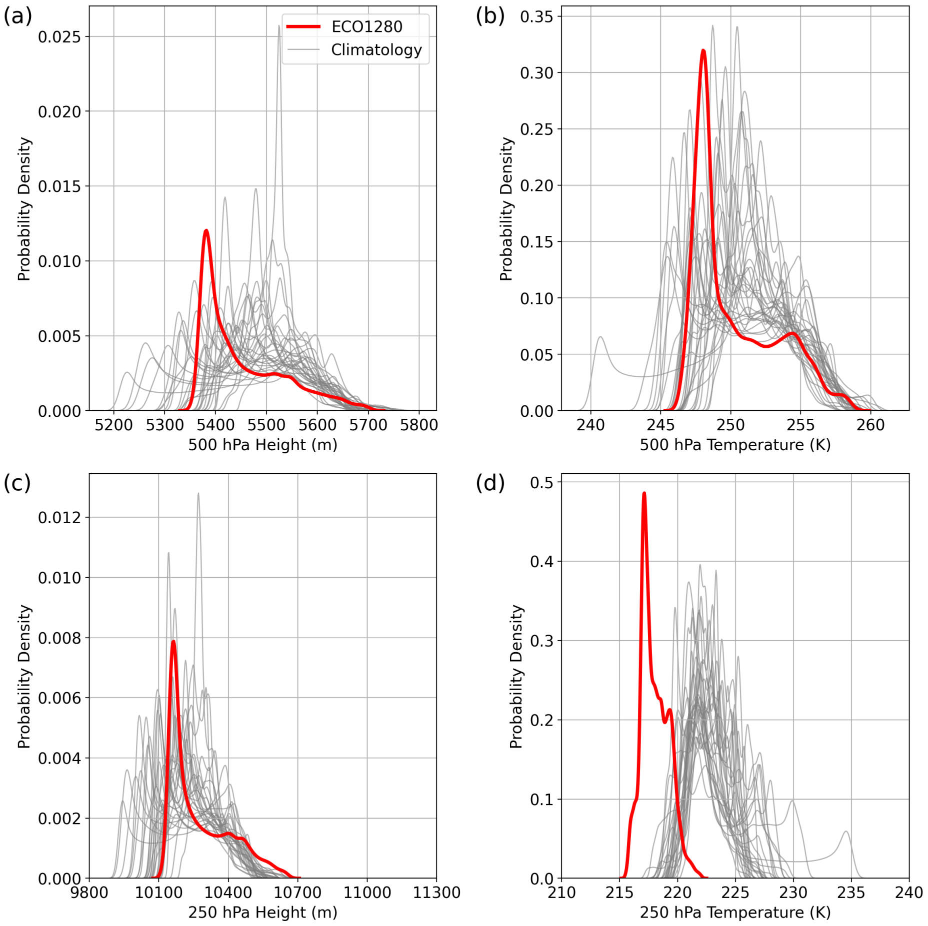


Figure 3.3: (a) Probability density function of all August 500 hPa geopotential heights (m) poleward of 60°N for ECO1280 (red) and the ERA5 climatology (gray). (b) As in (a) but for 500 hPa temperatures (K). (c) As in (a) but for 250 hPa geopotential heights (m). (d) As in (a) but for 250 hPa temperatures (K).

parameters used (Fig. 3.4b). The number of individual track points identified is similar between ECO1280 and ERA5, indicating that the difference is related to the process of stitching points together into tracks. Notably, using the same settings, TPVTrack produced slightly smaller TPV radii in ECO1280 than in ERA5, which in turn affects the rate at which track segments are combined to form longer tracks (Fig. 3.5d). The radii calculated by TPVTrack are in turn sensitive to adjustable parameters, so we do not believe that this represents any systematic error, but rather on differences in the resolution of ERA5 and ECO1280. Comparing the evolution of TPVs from August

2016 in ECO1280 and ERA5 over their lifetime, we see that the vortices inhabit roughly similar latitude bands, where most differences can be explained by different synoptic environments (Fig. 3.5a). Both minimum potential temperature and TPV amplitude (the difference in the minimum potential temperature and surrounding environment) also evolve very similarly between the two datasets (Fig. 3.4b,c). ECO1280 radii are, as mentioned, smaller, but follow a similar pattern of growth and decay (Fig. 3.4d).

Arctic cyclones in ECO1280 similarly fall within an acceptable range of ERA5 climatology. ACs in ECO1280 have minimum MSLP values on the lowest end of the ECO1280 climatology, indicating that the cyclones are slightly stronger than corresponding real-world cases in August (Fig. 3.6a); however, overall counts of cyclone points are consistent between ECO1280 and the climatology, and the differences in diagnosed intensity may be related to ECO1280's high spatial resolution. Cyclone tracks are of a similar length in ECO1280 compared to ERA5, including similar representation of the longest-lived ACs of over a week (Fig. 3.6b). Over the course of their lifetime, cyclones evolve similarly between ECO1280 and ERA5, occurring in similar latitude ranges and with comparable intensities (Fig. 3.7).

In short, ECO1280 appears to resolve Arctic features within a reasonable margin of climatology, and so we will progress forward with using it as the nature run. Certain features like the 250 hPa cool bias and stronger than average cyclones should be considered when evaluating the OSSE results.

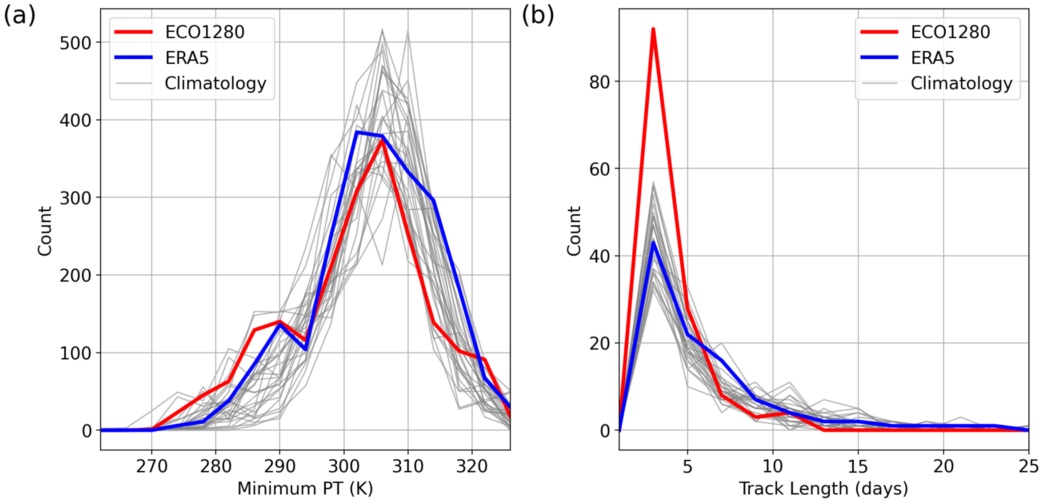


Figure 3.4: (a) Histogram of all August TPV minimum TRPT values (K) for ECO1280 (red), 2016 in ERA5 (blue), and the ERA5 climatology (gray). (b) As in (a) but for TPV lifetimes (days).

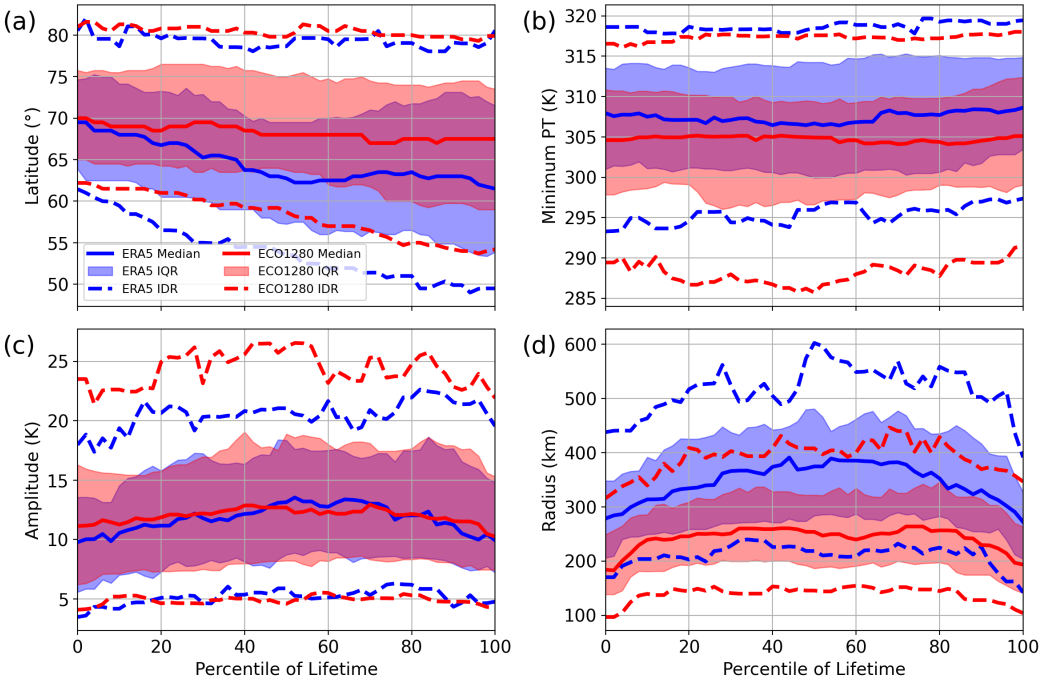


Figure 3.5: (a) TPV median (solid), interquartile range (shaded), and interdecile range (dashed) latitude (°N) as a function of percentile of total vortex lifetime in August 2016 for ECO1280 (red) and ERA5 (blue). (b) As in (a), but for TPV minimum TRPT (K). (c) As in (a), but for TPV amplitude (K). (d) As in (a), but for TPV radius (km).

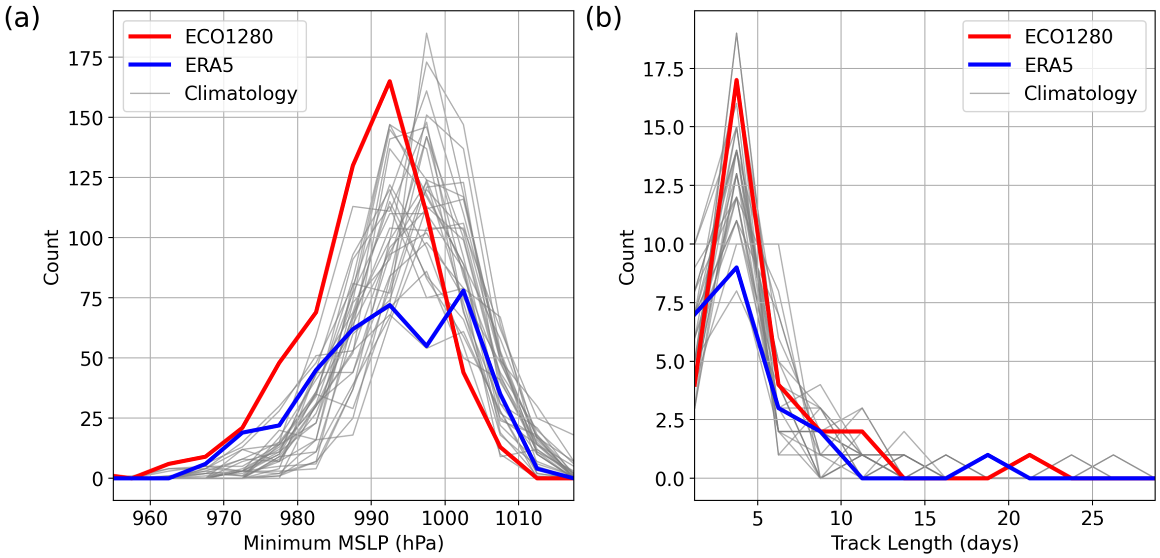


Figure 3.6: As in Figure 3.4 but for (a) AC minimum MSLP (hPa) and (b) AC lifetimes (days).

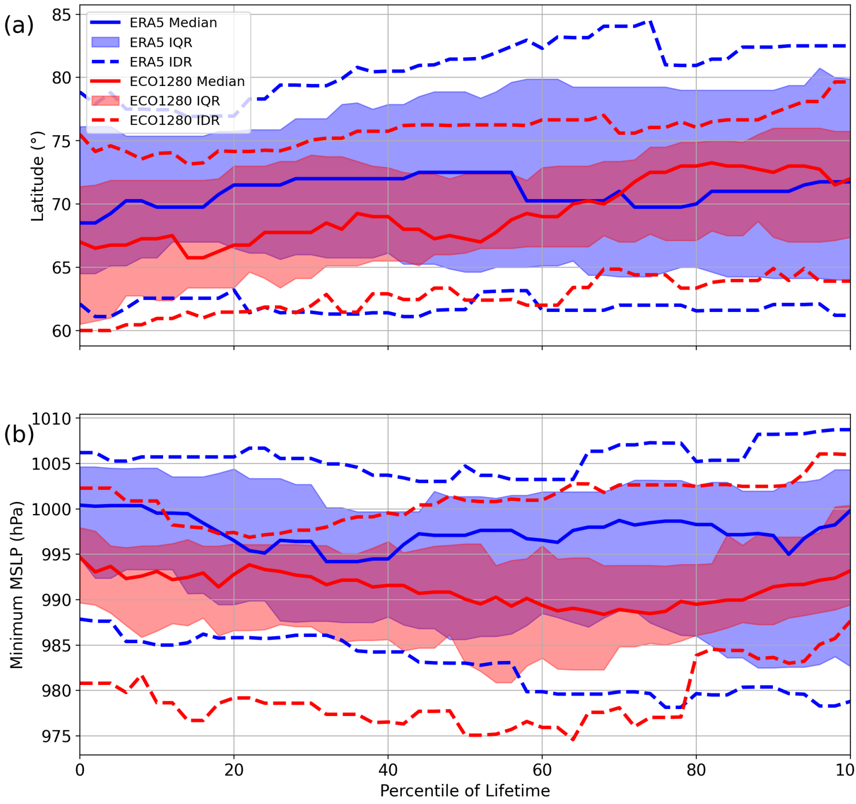


Figure 3.7: As in Figure 3.7 but for (a) AC latitude (°N) and (b) AC minimum MSLP (hPa).

3.3 Ensemble Spin-up

As discussed in Section 2.1, we initialize the 84 member ensemble from time-lagged GEFS forecasts, valid for 11 August 2016 at 12 UTC. After each of these members is advanced forward 12 hours by MPAS (enough time for MPAS physics perturbations to fully initialize), spin-up cycling with MPAS-DART begins and continues until the ensemble is deemed to be satisfactorily representative of ECO1280. This agreement between the ensemble and NR is assessed using several metrics such as root-mean-square errors (RMSEs) of forecast fields and DART-generated ensemble metrics like inflation factor, ensemble spread, and bias relative to observations. All of these metrics are evaluated after each cycling period, and experimental cycling begins once all metrics become stable in successive cycles. In the end, this process resulted in a four-day spin-up period running from 12-16 August.

First, we compare MPAS ensemble mean fields to both ECO1280 and ERA5 references; the ensemble is expected to start with lower errors relative to ERA5 and progress towards ECO1280 as more observations are assimilated. Indeed RMSE in ensemble mean 500 hPa heights across the Arctic starts at around 30 m with respect to ERA5 and 110 m with respect to ECO1280 (Fig. 3.8a). By the conclusion of the spin-up period, these values have reversed and grown level with time. A similar pattern emerges in 500 hPa temperature RMSE (Fig. 3.8c). This result is especially encouraging as it indicates that the ensemble has similar errors with respect to the nature run as an operational model has with respect to the atmospheric true state. The transition from a real-world ensemble to an ECO1280 ensemble can also be viewed via time series of Arctic-averaged fields. Ensemble mean 500 hPa heights and temperatures over the Arctic gradually transition from the ERA5 base value to the ECO1280 mean (Fig. 3.8b,d).

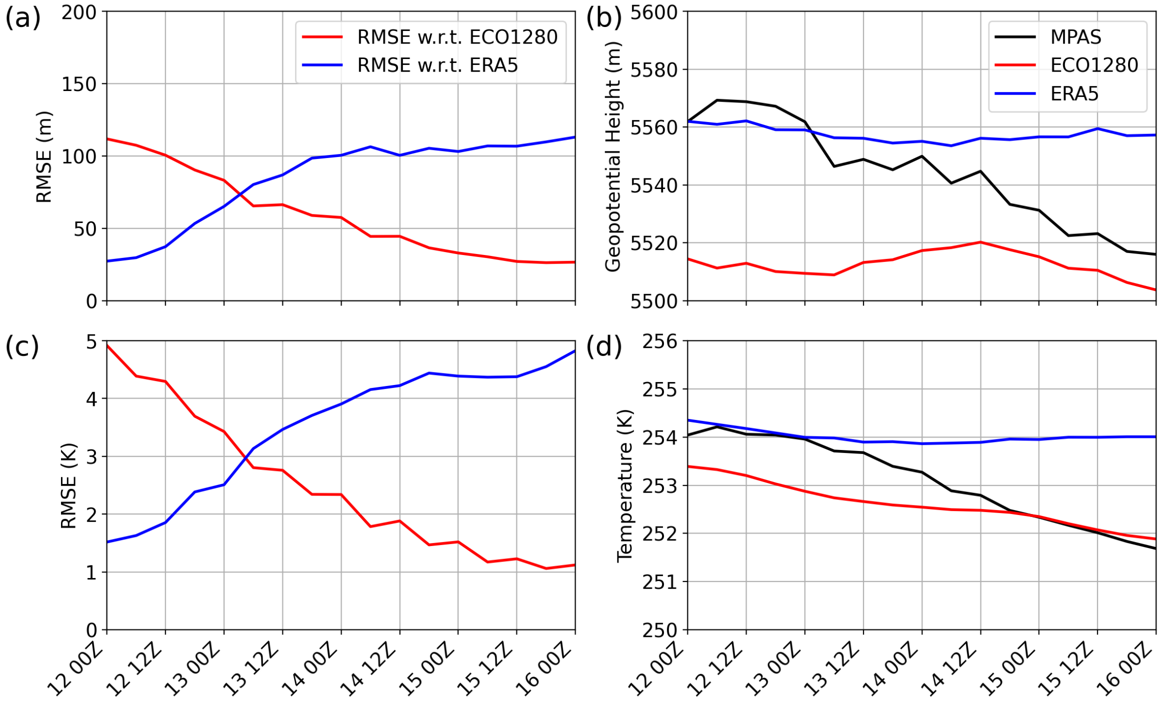


Figure 3.8: (a) RMSE of MPAS ensemble mean 500 hPa geopotential height (m) in the Arctic with respect to ECO1280 (red) and ERA5 (blue) over the spin-up period. (b) Arctic-averaged 500 hPa geopotential height (m) of the MPAS ensemble mean (black), ECO1280 (red), and ERA5 (blue) over the spin-up period. (c) As in (a) but for 500 hPa temperature (K). (d) As in (b) but for 500 hPa temperature (K).

Turning to the DART-based techniques, we first examine inflation factors. As mentioned, the configuration of DART employed here applies temporally and spatially evolving inflation factors, which change based on the innovations, background errors, and observations errors at each time step. Initially, both the mean and maximum inflation factors spike, reaching a maximum around a day into cycling (Fig. 3.9). This result is expected as innovations would be very large until the ensemble approaches the NR state. Over time, the values gradually decrease and approach generally constant values by 16 August, which indicates that the innovations and error covariances have stabilized. Global maximum inflation factors remain high throughout the spin-up period (Fig. 3.9c,d). These high values originate from surface fields over Europe, and

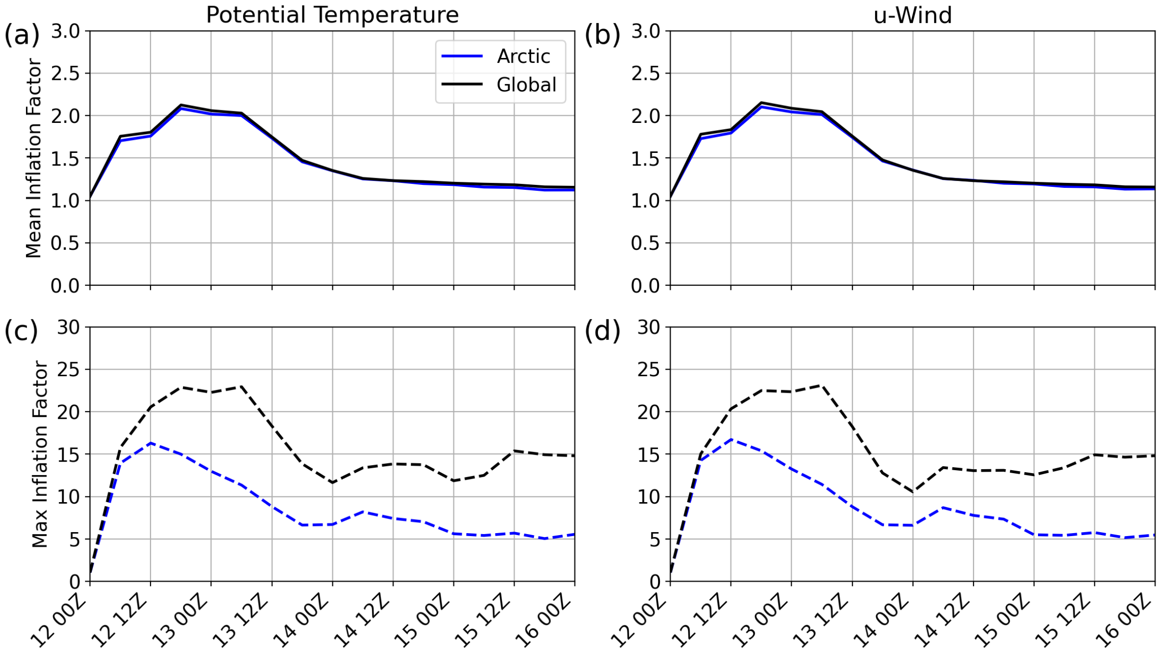


Figure 3.9: (a) Mean inflation factor for potential temperature at all levels across the Arctic (blue) and globally (black) over the spin-up period. (b) As in (a), but for zonal winds inflation factors. (c) As in (a) but for the maximum inflation factor. (d) As in (b) but for the maximum inflation factor.

we speculate that they may be related to high observation densities and differences in surface field diagnostic methods between MPAS and ECO1280.

As a final test of how well the ensemble has approached ECO1280, we examine DART observation diagnostic profiles averaged globally and over the Arctic. For brevity, we will specifically examine radiosonde temperature observations, but other observation types follow similar trends (with the exception of the humidity field which has a bias that will be discussed below). For a well-tuned ensemble, the average pre-assimilation bias should be near 0, and the DA process should reduce the magnitude of the bias even further. Although biases started out large both globally and in the Arctic (not shown), by 16 August at 00 UTC, near 0 K biases are achieved (Fig. 3.10a,c). Additionally, the total ensemble spread should be nearly equal to RMSE with respect

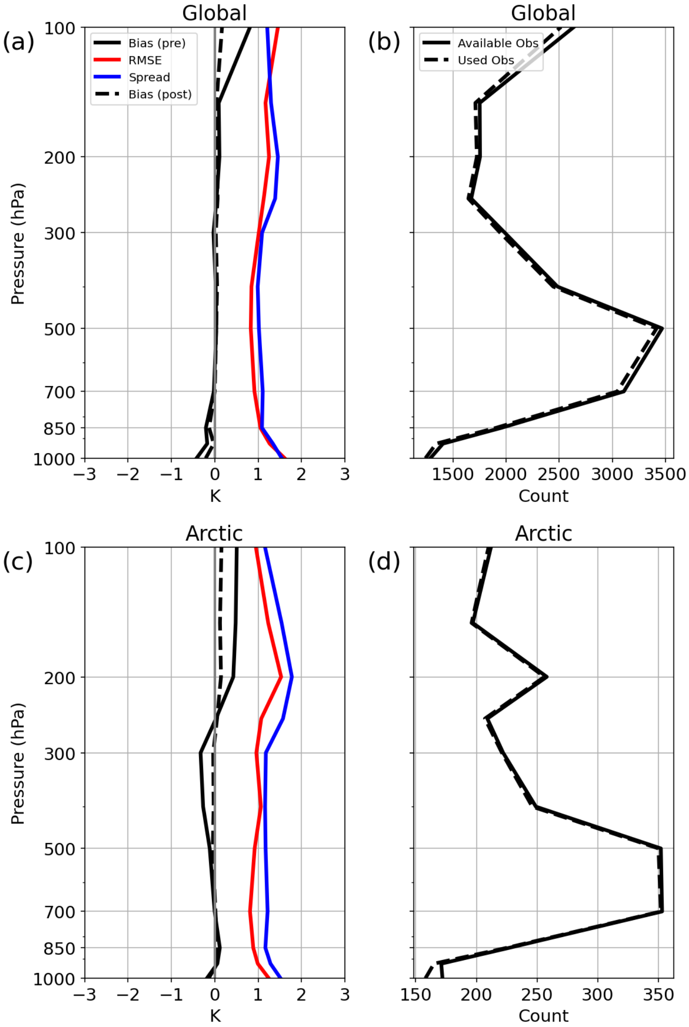


Figure 3.10: (a) Globally averaged profiles of 6-hour forecast bias (black solid), analysis bias (black dashed), ensemble total spread (blue), and RMSE (red) of the MPAS ensemble compared to radiosonde temperature observations (K) on 16 August at 00 UTC. (b) Vertical profile of global radiosonde temperature observations available (solid) and assimilated (dashed) on 16 August at 00 UTC. (c) As in (a) but averaged over the Arctic. (d) As in (b) but for the Arctic.

to observations in order to satisfy EnKF theory. At the beginning of the spin-up period, spread is low (GEFS is known to be somewhat underdispersive) and RMSE is high for all fields (not shown). At the conclusion of the spin-up period, the RMSE and total spread are nearly identical (Fig. 3.10a,c).

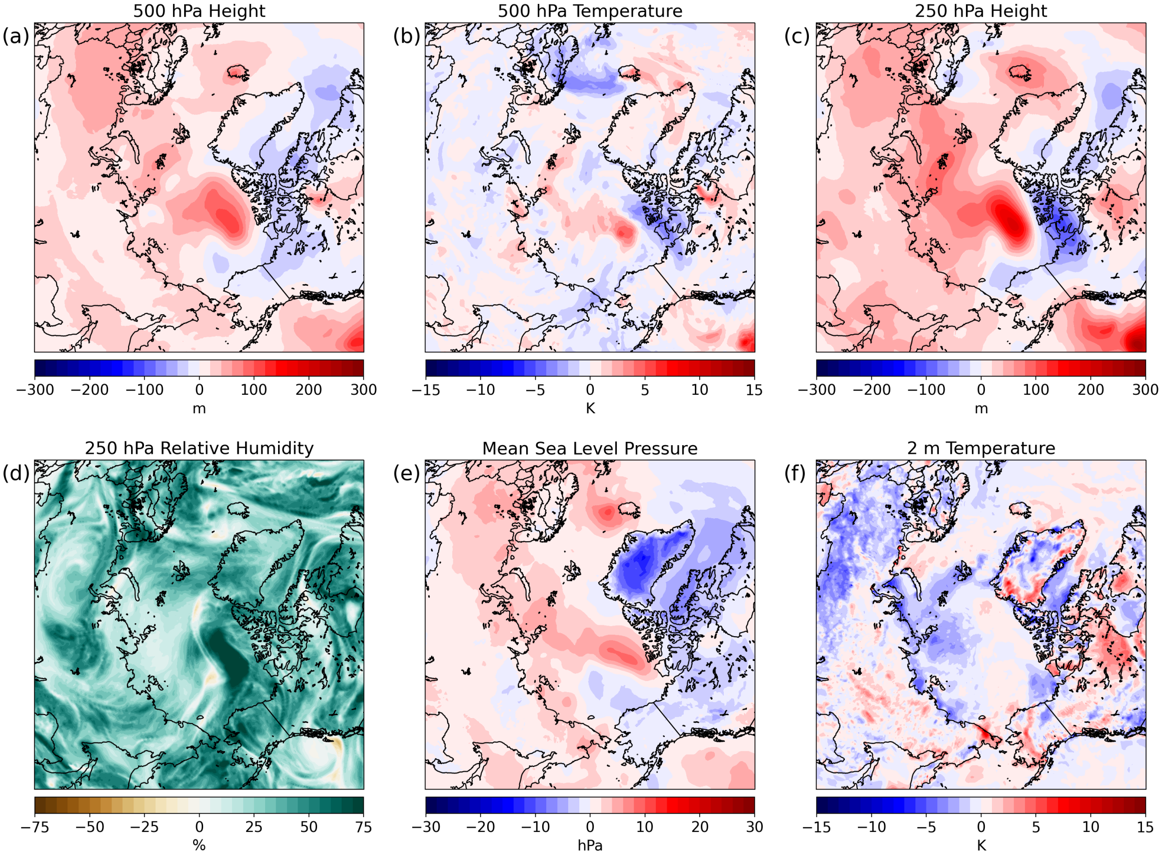


Figure 3.11: (a) Difference between MPAS ensemble mean and ECO1280 (MPAS - ECO1280) 500 hPa geopotential height (m) at 16 August 2016 00 UTC. (b) As in (a), but for 500 hPa temperature (K). (c) As in (a), but for 250 hPa geopotential height (m). (d) As in (a), but for 250 hPa relative humidity (%). (e) As in (a), but for MSLP (hPa). (f) As in (a), but for 2 meter temperature (K).

Before discussing the results from the OSSE, we will briefly discuss the errors still present in the ensemble at 00 UTC on 16 August (i.e., the starting point for each of the experiments). The most salient errors across all fields are just poleward of the Canadian Archipelago, where the MPAS ensemble is too warm, too moist, and has too high of geopotential heights (Fig. 3.11). In ECO1280, this region includes a decaying TPV-AC couplet which did not pass over any conventional observing sites during the spin-up period. During the cycling period, the system continues to weaken and passes over several Canadian sounding sites, resolving the bias. The MPAS ensemble is also too moist throughout the upper atmosphere, though especially at and above the tropopause (Fig. 3.11d). This bias arose from errors in the preprocessing software that converted GEFS forecasts to MPAS grids that were not fully rectified by the spin-up observations, despite moisture observations being available at these levels. This limited improvement may be linked to the general difficulties with assimilating upper-level humidity observations discussed in Section 4.1. Fortunately, the moisture biases are relatively low over the TPV of interest (around a 10% relative humidity error), but this bias will still be important to consider when analyzing results and applying conclusions to real-world systems. Elsewhere, biases are relatively low and within the range of errors between an operational model and the real world.

Chapter 4

Results

4.1 Experimental Performance Overview

4.1.1 Spatial Metrics and Field Differences

Before exploring specific differences between each experiment, we examine TMEE to provide a broad overview of the OSSE. Throughout most of the experimental period, TMEE in the verification region is decreased in all of the experimental means relative to the Control run (Fig. 4.1a). Experiments 1, 3, and 4 (i.e., those that assimilate temperature) see the most initial improvement while Experiment 2 improves less initially but catches up to the rest during the free forecast. This difference is expected given that humidity observations have a less direct impact than temperature on the wind fields but should have a cumulative effect on the forecasts via PV tendencies. Experiment 4 generally outperforms the other experiments as the cyclone reaches its maximum intensity, indicating that the less dense observations spread over a larger area have a more lasting effect than dense observations over the TPV core. Experiments 1 and 3 are nearly identical throughout the period, suggesting that the temperature observations dominate when combined with humidity observations. Towards the end of the forecast period, the control TMEE drops considerably compared to the experiments. This appears to result from the control briefly simulating the poleward flank of the system with high accuracy, despite errors in the preceding time steps. This effect will be discussed in more detail below.

Focusing now on the core of the TPV (i.e., within the 300 K contour on the dynamic tropopause), these trends are generally replicated (Fig. 4.1b). Experiments 1

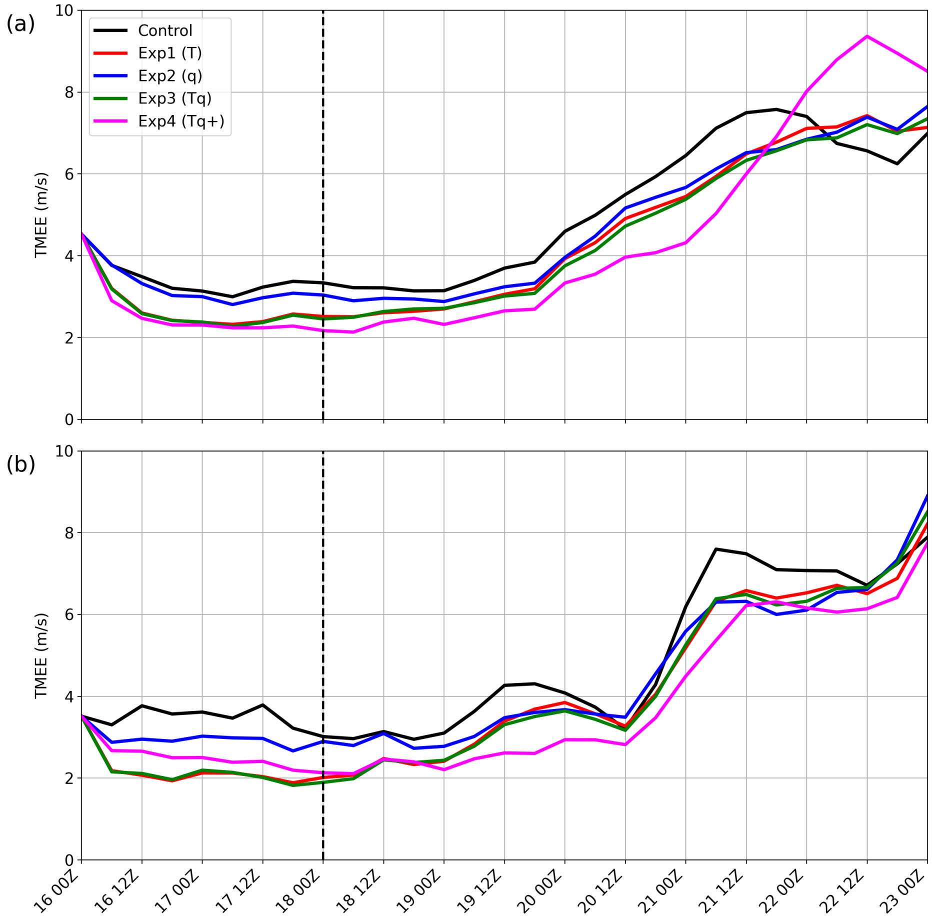


Figure 4.1: (a) Time series of TMEE (m s^{-1}) over the verification region for each experiment through the cycling and forecast periods. The dashed black line marks the start of the forecast period. (b) As in (a), but for the core verification region.

and 3 provide initial improvements but are eventually matched by Experiment 2 and outperformed by Experiment 4. The spurious improvement of the control on 22 August is less noticeable in the TPV core as well, as this improvement mostly occurred on the periphery of the system. Spatially, the largest TMEE improvements relative to the control occur over the AC-TPV core and within the ridge wrapping into the system from the south (Fig. 4.2). Although not directly observed by any of the simulated dropsondes, the progression of this region of lower-latitude air is likely tied to the strengths of the AC and TPV. It is also interesting to note the regions of forecast degradation (red shading) relative to the control over the Canadian Archipelago in

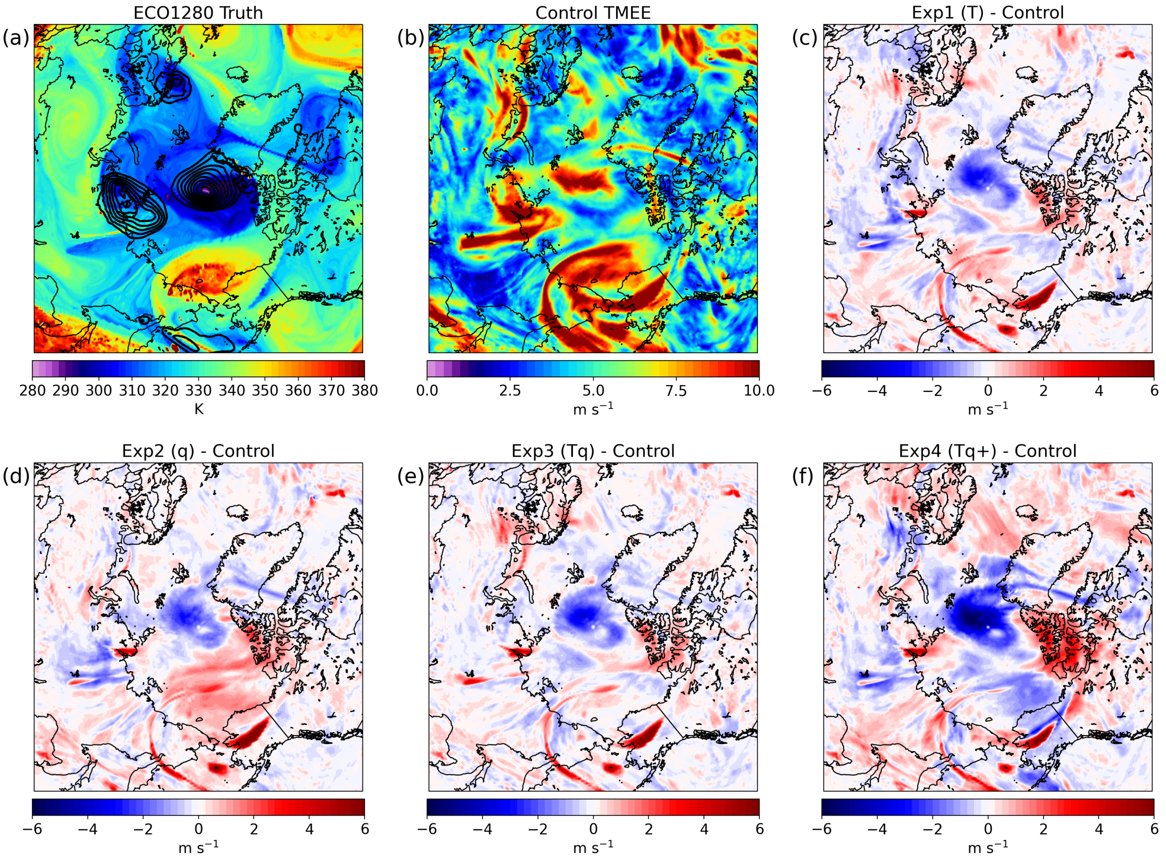


Figure 4.2: (a) ECO1280 TRPT (fill; K) and MSLP (hPa; every 2.5 hPa starting at 1000 hPa) on 21 August 00 UTC. (b) Control run TMEE (m s^{-1}) on 21 August 00 UTC. (c) Difference between Exp1 (T) and Control TMEE on 21 August 00 UTC. (d) As in (c), but for Exp2 (q). (e) As in (c), but for Exp3 (Tq). (f) As in (c), but for Exp4 (Tq+).

each experiment (Fig. 4.2c-f). These areas are along the edge of the ECO1280 TPV. The various additional dropsondes appear to have caused the vortex to expand too far into this region relative to ECO1280, while the smaller control TPV remains confined over the Arctic Ocean.

At the beginning of the forecast period, the core of the primary surface cyclone is represented similarly in the ensemble means of the experiments and the control (Fig. 4.3a-d). This result is unsurprising given that observations did not target the AC or directly observe surface pressure. Still, significant differences are present in the remote MSLP field. The decaying cyclones surrounding the main low are deeper in each of

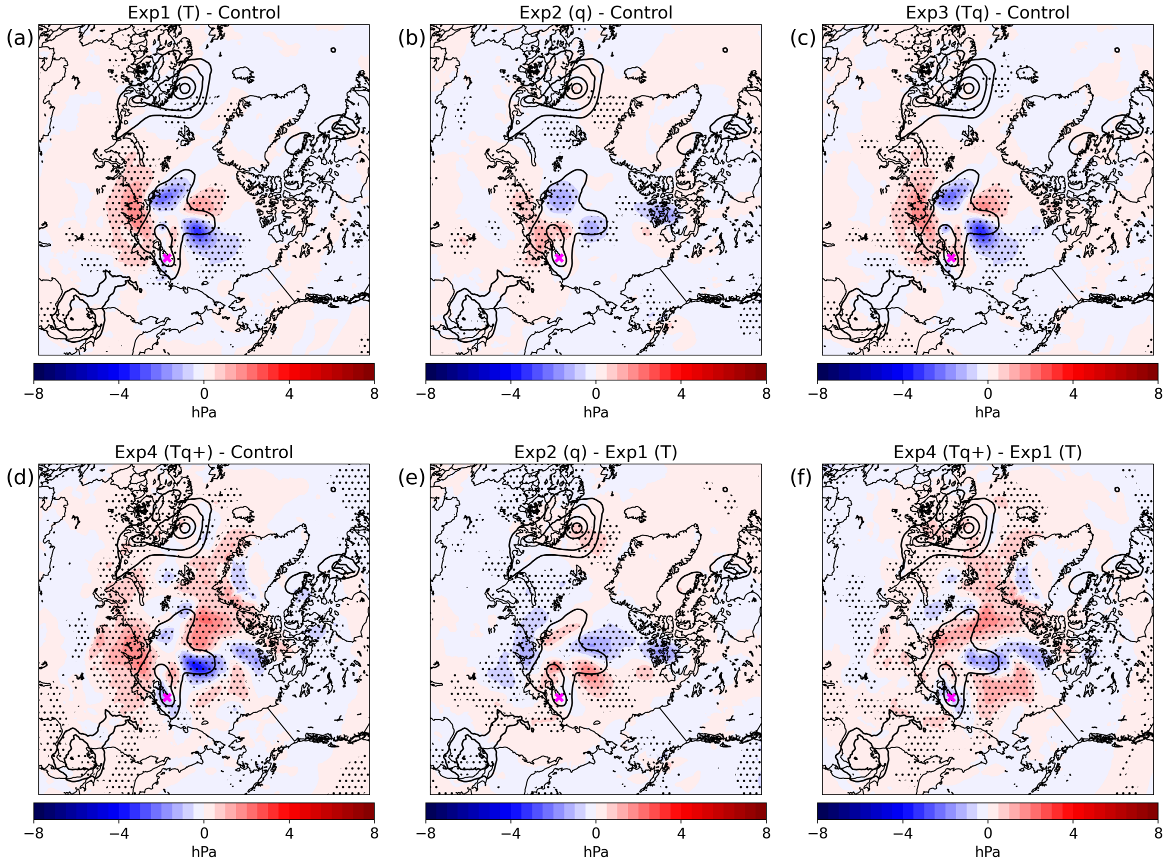


Figure 4.3: (a) Difference in Exp1 mean (T) and Control mean MSLP (fill; hPa) on 18 August 00 UTC. Stippling indicates significance at the 95% level using a Welch's t-test. ECO1280 MSLP (contours; hPa; every 5 hPa starting at 1000 hPa) is shown for reference, and the pink X represents the cyclone of interest. (b) As in (a), but for Exp2 (q) and Control. (c) As in (a), but for Exp3 (Tq) and Control. (d) As in (a), but for Exp4 (Tq+) and Control. (e) As in (a), but for Exp2 (q) and Exp1 (T). (f) As in (a), but for Exp4 (Tq+) and Exp1 (T).

the experiments than in the control, and pressure is broadly higher along the Siberian coast. Only Experiment 4 includes a main AC that is significantly deeper than in the control, along with broad changes to the surface pressure field throughout the Arctic that are not present with the dense observation experiments (Fig. 4.3d,f). Humidity observations alone generally appear to provide less correction to the initial MSLP field than temperature observations (Fig. 4.3b,e).

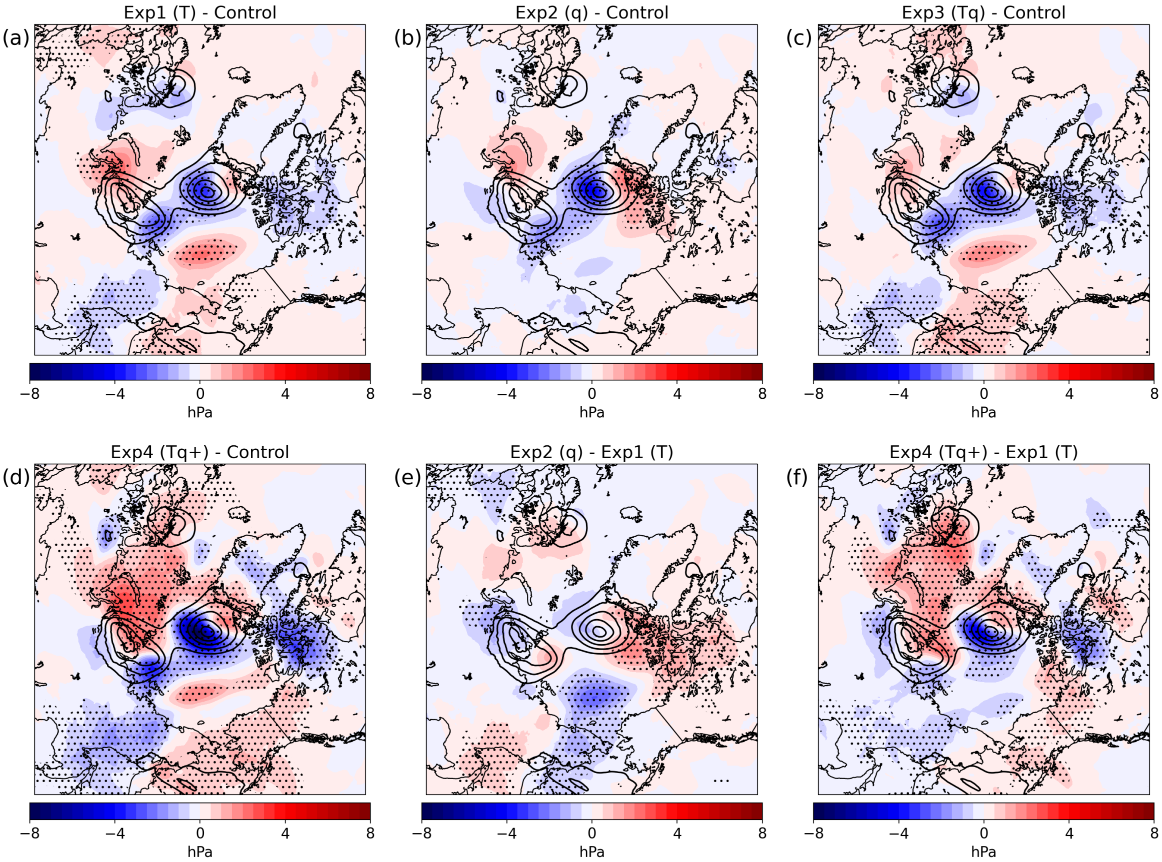


Figure 4.4: As in Figure 4.3, but on 21 August 00 UTC.

Despite small initial differences in MSLP around the AC, the experiments diverge from the control as the forecast simulation progresses. Mean differences reveal significant improvements in AC strength, location, and structure relative to the control for all experiments (Fig. 4.4a-d). The cyclone center is generally shifted away from the Canadian coast, and pressure around the cyclone core is decreased by 4 (Experiment 1) to 7 (Experiment 4) hPa, with a generally expanded region of low pressure. As seen in the TMEE analysis, Experiments 1 and 3 produce very similar results, with only slightly lower surface pressures in the latter (Fig. 4.4a,c). Experiment 2 results in comparable adjustments to the control baseline, despite the humidity observations making less of an impact on the initial MSLP field (Fig. 4.4b,e). Experiment 4 has the largest impact on the AC of interest, with significantly more deepening and a more defined spatial translation than the other experiments (Fig. 4.4d,f).

To more accurately diagnose whether the changes to the MSLP field discussed above represent improvements relative to the control, we examine differences between ensemble mean MSLP and ECO1280 (Fig. 4.5). Surface pressure is too high throughout the cyclone in the control, with especially large errors near the cyclone center (Fig. 4.5b). The magnitude of this MSLP bias is reduced in all experiments, though errors in the AC core are not completely alleviated (Fig. 4.5c-f). Note that because the additional dropsondes reduced model spread in the Arctic, the ensemble percentile method shows larger regions of significant differences in the experiments than in the control. It is also interesting to note the significant impacts on the subsequent AC along the Siberian coast, especially in Experiment 4 (Fig. 4.4d,f). This cyclone was poorly simulated in all of the ensembles, but the wide-ranging dropsondes of Experiment 4 appear to have degraded the forecast further, increasing the central cyclone pressure by around 2 hPa (Fig. 4.5f).

To examine changes in the upper atmosphere, we will use both ensemble mean TRPT (which provides a detailed view of the vortex structure) and mean 500 hPa heights (a relatively smooth field for calculating differences). At the start of the forecast period, ECO1280 exhibits a primary TPV with PV filaments extending to the north and south and two smaller vortices over the central Arctic (Fig. 4.6a). The general structure of the main TPV is captured in all of the experiments, with lower TRPT values in the experiments that assimilated temperature dropsondes (1, 3, and 4; Fig. 4.6b-f). None of the ensemble means reach the same minimum TRPT values present in ECO1280. The PV filament extending south of the main TPV (which plays a key role in its future development) is more clearly represented in each of the experimental ensemble means than in the control; however, only Experiment 4 resolves the central Arctic PV features in an ensemble mean sense.

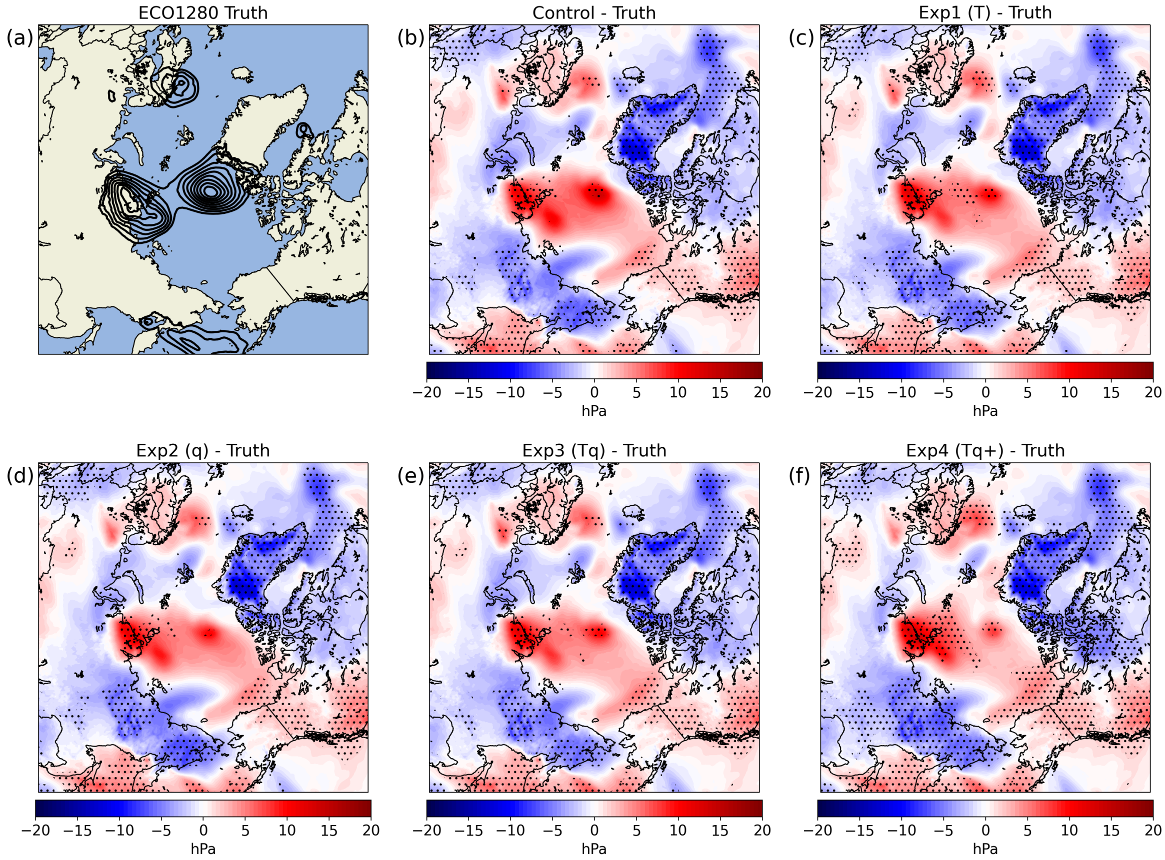


Figure 4.5: (a) ECO1280 MSLP (contours; hPa; every 2.5 hPa starting at 1000 hPa) on 21 August 00 UTC. (b) Difference in Control mean and ECO1280 MSLP (hPa) on 21 August 00 UTC. Stippling indicates significance using the ensemble percentile method described in Section 2.2 of the text. (c) As in (b), but for Exp1 (T). (d) As in (b), but for Exp2 (q). (e) As in (b), but for Exp3 (Tq). (f) As in (b), but for Exp4 (Tq+).

The 500 hPa height fields add some complexity to the analysis. The region of interest is, as would be expected, characterized by a broad low in 500 hPa heights in ECO1280 (Fig. 4.7a). Interestingly, though, control mean 500 hPa heights are slightly too low in the TPV core relative to the truth, despite a relatively weaker TPV as well (Fig. 4.7b). This disconnect seems to indicate remaining complexities in the vertical structure of the atmosphere (possibly from the preceding AC coupling event), while also exemplifying the unique, non-linear vortex characteristics of TPVs. Regardless, the dropsonde observations serve to correct these errors, significantly increasing heights

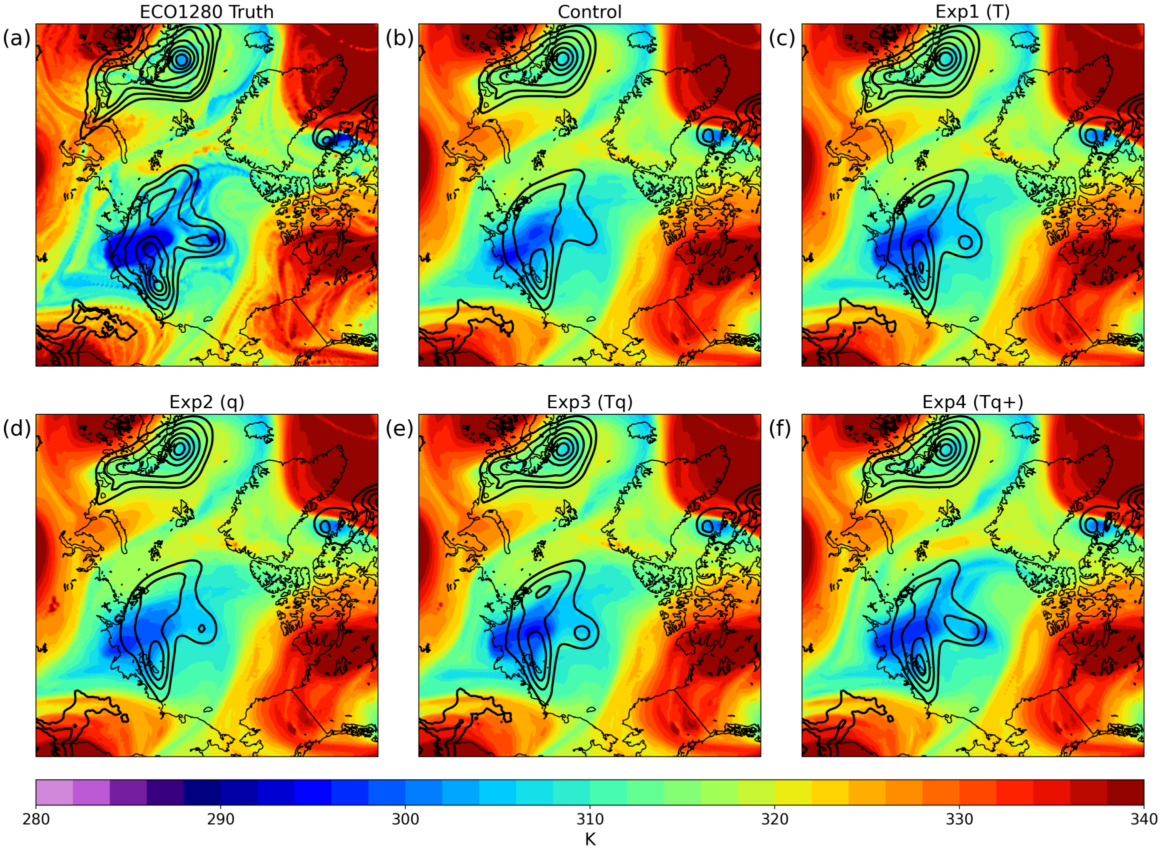


Figure 4.6: (a) ECO1280 TRPT (fill; K) and MSLP (contours; hPa; every 2.5 hPa starting at 1000 hPa) on 18 August 00 UTC. (b) Control ensemble mean TRPT (K) and MSLP (contours; hPa; every 2.5 hPa starting at 1000 hPa) on 18 August 00 UTC. (c) As in (b), but for Exp1 (T). (d) As in (b), but for Exp2 (q). (e) As in (b), but for Exp3 (Tq). (f) As in (b), but for Exp4 (Tq+).

through the TPV core with regions of lowered heights on the vortex periphery (Fig. 4.7b-f). Importantly, the trough in the south side of the vortex significantly deepens in all experiments. As in the previous fields examined, the initial changes in Experiment 2 are small relative to the other experiments, while Experiment 4 produces significant adjustments over a wider area.

As the AC reaches its maximum strength, trends at the upper and lower levels complement each other. The control mean TPV signal is weaker than in ECO1280 and offset towards the Canadian coast (Fig. 4.8a,b). In each of the experiments, the TPV is larger and matches the location of the ECO1280 TPV more closely (Fig. 4.8c-f).

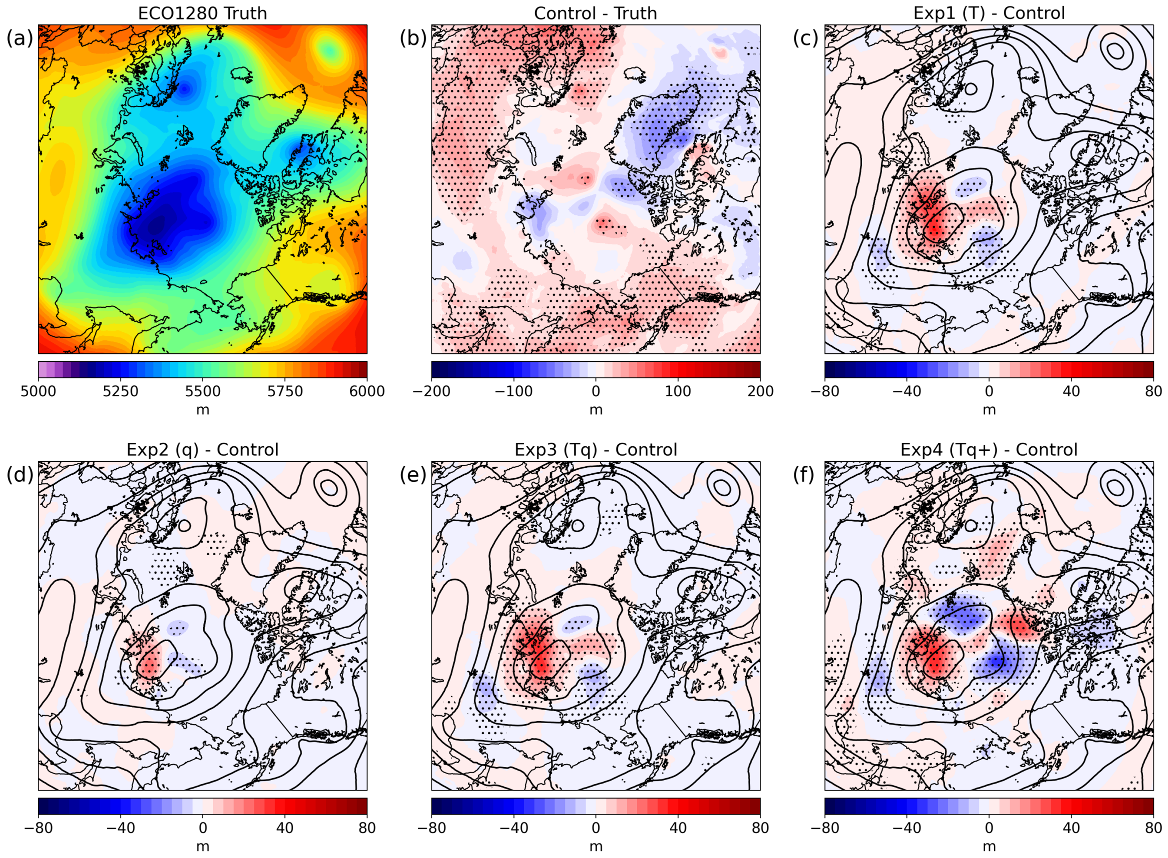


Figure 4.7: (a) ECO1280 500 hPa geopotential height (m) on 18 August 00 UTC. (b) Difference in Control mean and ECO1280 500 hPa geopotential height (m) on 21 August 00 UTC. Stippling indicates significance using the ensemble percentile method described in Section 2.2 of the text. (c) Difference in Exp1 mean (T) and Control mean 500 hPa geopotential height (fill; m) on 18 August 00 UTC. Stippling indicates significance at the 95% level using a Welch's *t*-test. ECO1280 500 hPa geopotential height (contours; m; every 100 m between 5000 and 6000 m) is shown for reference. (d) As in (c) but for Exp2 (q) and Control. (e) As in (c) but for Exp3 (Tq) and Control. (f) As in (c) but for Exp4 (Tq+) and Control.

A detailed analysis of changes in the mesoscale TPV structure is provided in Section 4.3. The ridge wrapping around the TPV over Greenland appears to have built further into the Arctic in all experiments, especially Experiment 4, which reflects broadly on the strength of the tropospheric circulation. The smaller PV features trailing the main TPV are best represented in the mean of Experiment 4, though it is important to keep in mind that these small features may not show up as clearly in ensemble means.

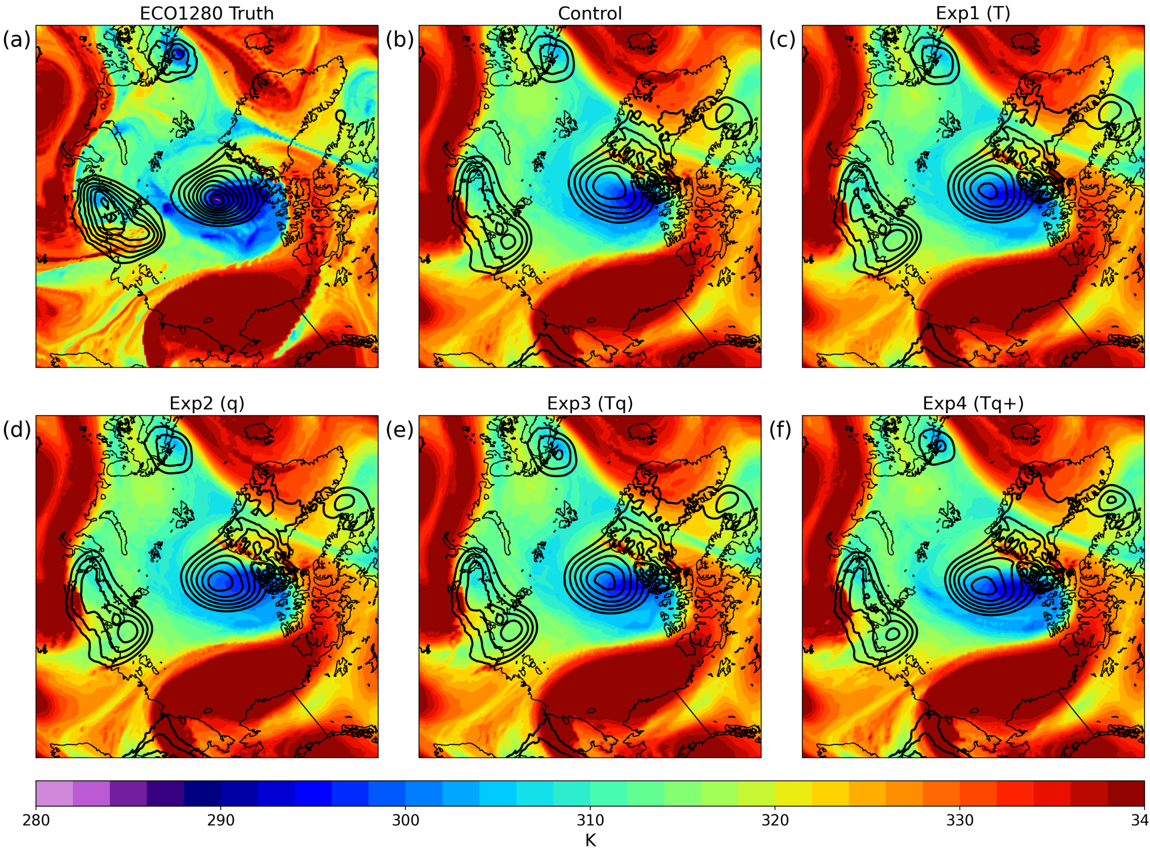


Figure 4.8: As in Figure 4.6, but on 21 August 00 UTC.

At this time step, the TRPT and 500 hPa height signals align more clearly. The ECO1280 500 hPa low has deepened significantly, with a minimum spatial minimum over the TPV core (Fig. 4.9a). Heights in the control ensemble are too high relative to ECO1280 throughout much of the Arctic and over the TPV core but too low in the region of the building ridge north of Greenland (Fig. 4.9b). The dropsonde observations generally addressed this bias, significantly lowering heights near the TPV center and raising them over the ridge (Fig. 4.9c-f). Once again, although the initial changes in Experiment 2 were minimal compared to the other experiments, biases are still significantly improved relative to the control in the core of the 500 hPa low (Fig. 4.9d). On the other hand, heights over the Canadian Archipelago are significantly lowered relative to the control, despite no preexisting error there; this trend is especially true in Experiment 4. This matches the pattern of forecast degradation seen in the TMEE fields (Fig. 4.2) and seems to represent an over-smoothing of the sharp height gradient in this region after the dropsondes are introduced.

We now examine changes in the moisture field at 250 hPa. The dry anomaly around this level (just at and above the tropopause) is key to PV creation and TPV maintenance. As discussed in Section 3.3, a positive moisture bias existed in the initial ensemble that was especially pronounced above the tropopause. Figure 4.10 indicates that this bias was far from resolved during experimental cycling, likely due to conflicting cross-correlation updates from temperature observations, which will be discussed below. With this overall bias in mind, we can still examine changes in the humidity field between experiments. The dense temperature observations of Experiment 1 lead to slightly increased humidity around the TPV core relative to the control, while the direct moisture observations in Experiment 2 produce relative drying throughout the TPV core (Fig. 4.10c,d). Experiment 3, which includes both temperature and humidity observations splits the difference between the first two experiments, with both significant drying and moistening around the TPV (Fig. 4.10e). Experiment 4 follows a similar trend around the TPV core, with larger changes outside of the main system (Fig. 4.10f). In general, Experiments 3 and 4 indicate that indirect temperature updates to the humidity field may be limiting the corrections made by direct upper-level moisture observations. This, in turn, may help explain why the initial moisture bias was not fully corrected during the spin-up and cycling periods.

Finally, we assess changes in cyclone-associated surface variables like temperature, wind speed and precipitation in order directly assess the impact of the new observations on human activities in the Arctic. As the cyclone reaches its maximum intensity, surface wind speeds surrounding the cyclone are lower than in ECO1280 across all MPAS ensembles, as would be expected from the incorrectly weak cyclones (Fig. 4.11a,b). Still, each of the experiments performs better than the control, with Experiment 4

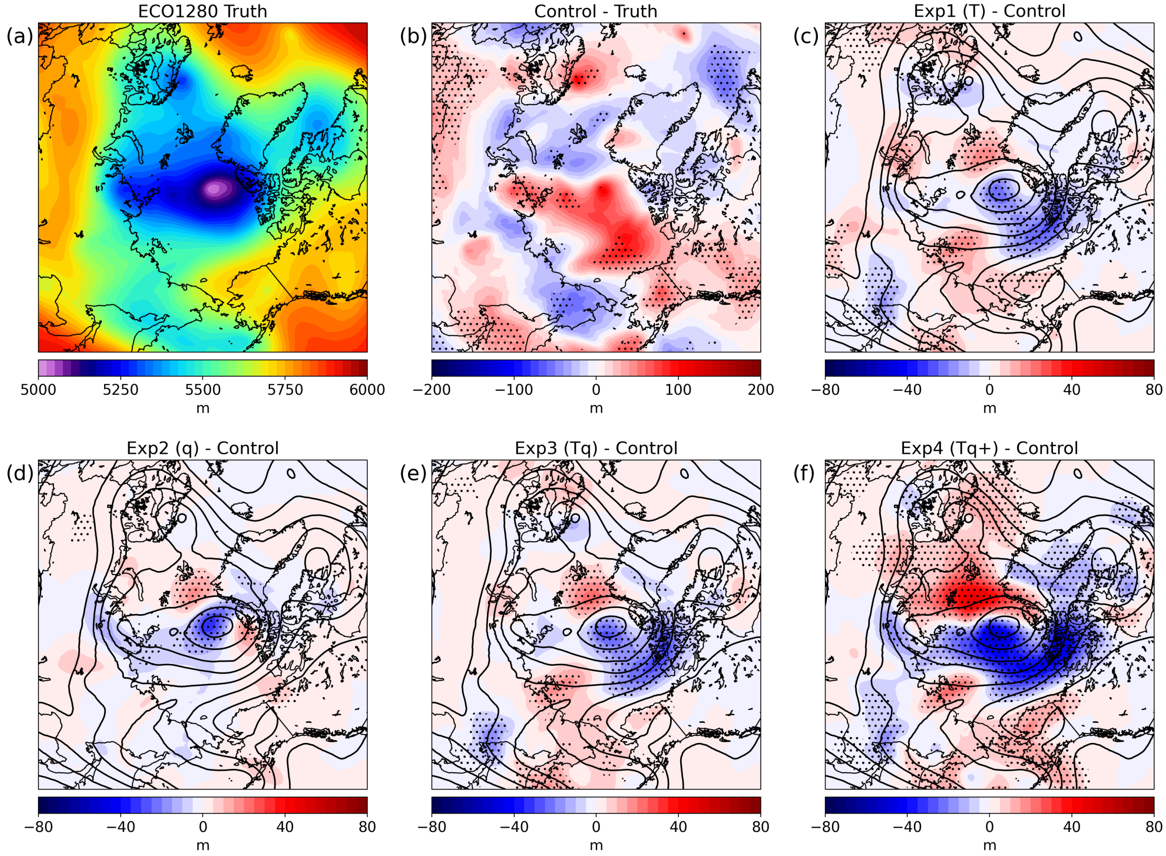


Figure 4.9: As in Figure 4.7, but on 21 August 00 UTC.

matching ECO1280 almost exactly until around 20 August 12 UTC (Fig. 4.11a). From a spatial perspective, it is clear that each of the experiments better represented the spatial extent of strong surface winds over the central Arctic, with Experiments 2 and 4 partially resolving the secondary speed maximum near Greenland (Fig. 4.12). Similarly for surface temperature, all of the ensembles exhibit smaller temperature ranges within the verification region than ECO1280 (a measure of the amount of temperature advection that has taken place), but the experiments generally outperform the control (Fig. 4.11c). This trend is predominately related to higher maximum surface temperatures near the cyclone in the dropsonde experiments than in the control, which better match ECO1280 values and are likely an indication of stronger warm air advection from the midlatitudes (Fig. 4.12). Both wind speeds and advection of warm air into

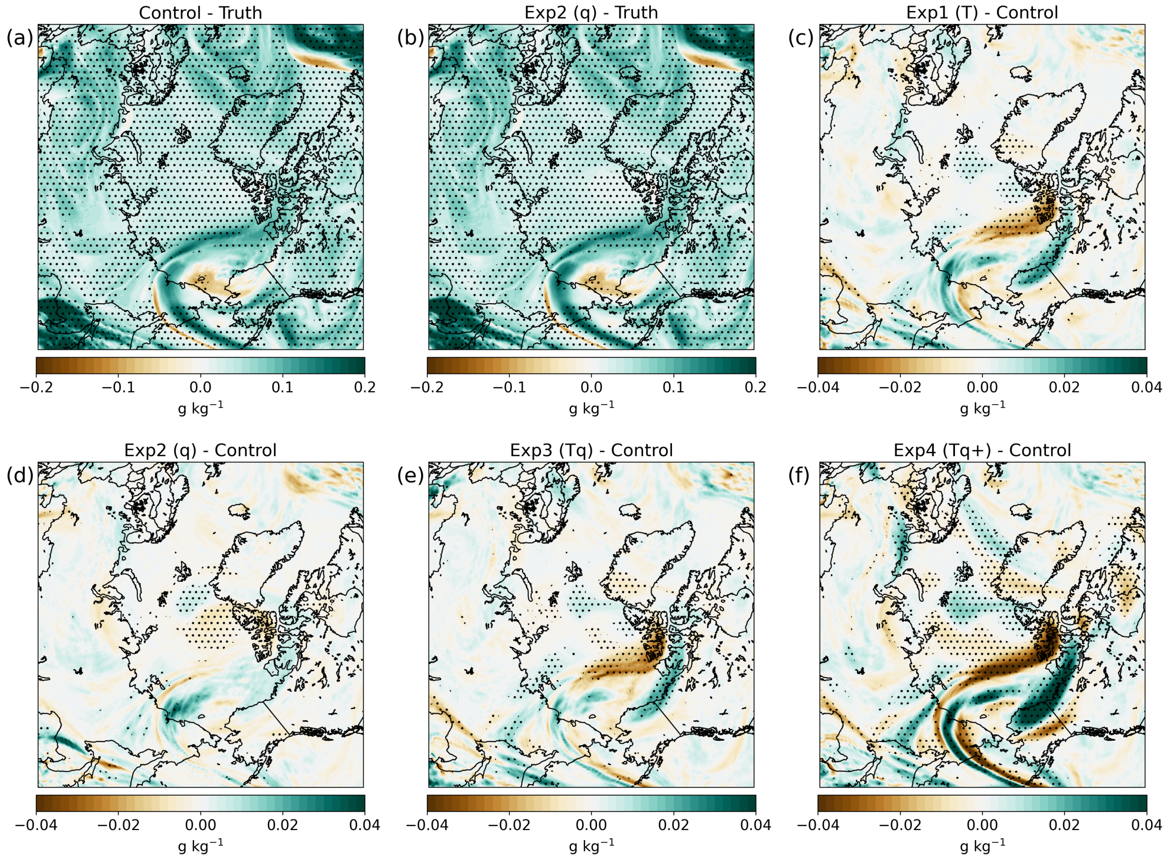


Figure 4.10: (a) Difference in Control mean and ECO1280 250 hPa specific humidity (g kg^{-1}) on 21 August 00 UTC. Stippling indicates significance using the ensemble percentile method described in Section 2.2 of the text. (b) As in (a) but for Exp2 (q). (c) Difference in Exp1 mean (T) and Control mean 250 hPa specific humidity (g kg^{-1}) on 21 August 00 UTC. Stippling indicates significance at the 95% level using a Welch's *t*-test. (d) As in (c), but for Exp2 (q) and Control. (e) As in (c), but for Exp3 (Tq) and Control. (f) As in (c), but for Exp4 (Tq+) and Control.

the Arctic are key for driving sea ice extent changes, so these results indicate that the additional dropsondes could have improved the model's forecast of sea ice impacts. Changes to the total accumulated precipitation are much more limited, with only Experiment 4 diverging from the control. Because precipitation data from ECO1280 is not available for comparison, it is difficult to determine whether this change represents and improvement or degradation (Fig. 4.11d).

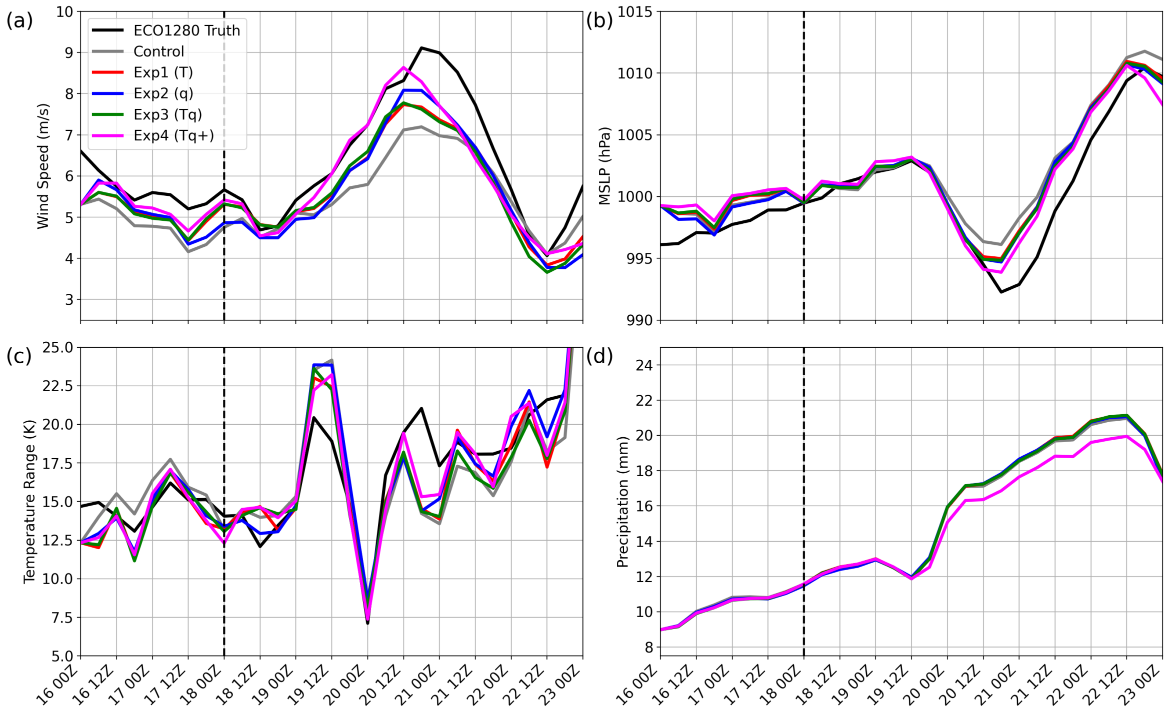


Figure 4.11: (a) Time series of average 10 meter wind speed ($\mathrm{m\,s^{-1}}$) over the verification region for each experiment and ECO1280 through the cycling and forecast periods. The dashed black line marks the start of the forecast period. (b) As in (a), but for mean sea level pressure (hPa). (c) As in (a), but for the range of 2 meter temperatures (K) present in the verification region. (d) As in (a), but for the mean total precipitation (mm).

4.1.2 Analysis Increments and Potential Vorticity Tendencies

To better understand the impact of the temperature and humidity observations, we now examine the mean analysis increments made during each DA cycling period over the TPV core. Temperature increments from the control run are small in magnitude and limited to cycling periods with a radiosonde observation in the vicinity (Fig. 4.13a). In the experiments with temperature dropsondes, initial increments are much larger, on the order of 2 K of cooling throughout the UTLS region (Fig. 4.13b,d,e). Despite much sparser observations over the TPV core, Experiment 4 produces comparable adjustments to Experiments 1 and 3. Viewing these upper-level increments spatially, a more complex picture emerges (Fig. 4.14). Initial temperature observations produce

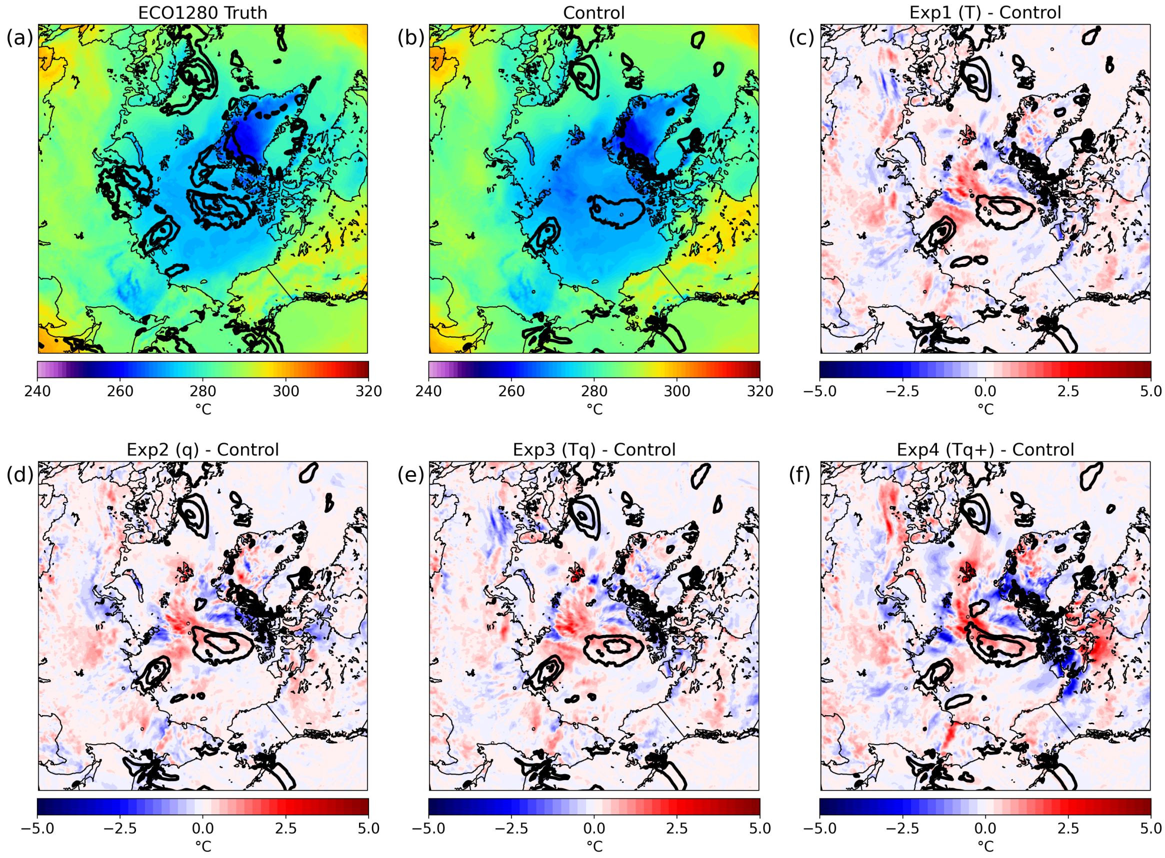


Figure 4.12: (a) ECO1280 2 meter temperature (fill; °C) and 10 meter wind speeds (contour; every 5 m s^{-1} starting at 10 m s^{-1}; m s^{-1}) for 21 August 2016 at 00 UTC. (b) As in (a), but for the control ensemble mean. (c) Difference between Exp1 (T) ensemble mean and control ensemble mean 2 meter temperature (fill; °C) and Exp1 (T) ensemble mean 10 meter wind speeds (contour; m s^{-1}) for 21 August 2016 at 00 UTC. (d) As in (c), but for Exp2 (q). (e) As in (c), but for Exp3 (Tq). (f) As in (c), but for Exp4 (Tq+).

warming increments in the lower stratosphere near the vortex core (which could be associated with strengthening the PV anomaly or shifting the vortex to the correct position) in some areas of the TPV, while cooling increments occur in others. These asymmetric adjustments to the mesoscale TPV structure are exactly what the temperature observations were intended to do. Experiment 4 in particular vastly reshapes the upper-level temperature structure of the whole Arctic, while retaining the detailed modifications over the TPV (Fig. 4.14e).

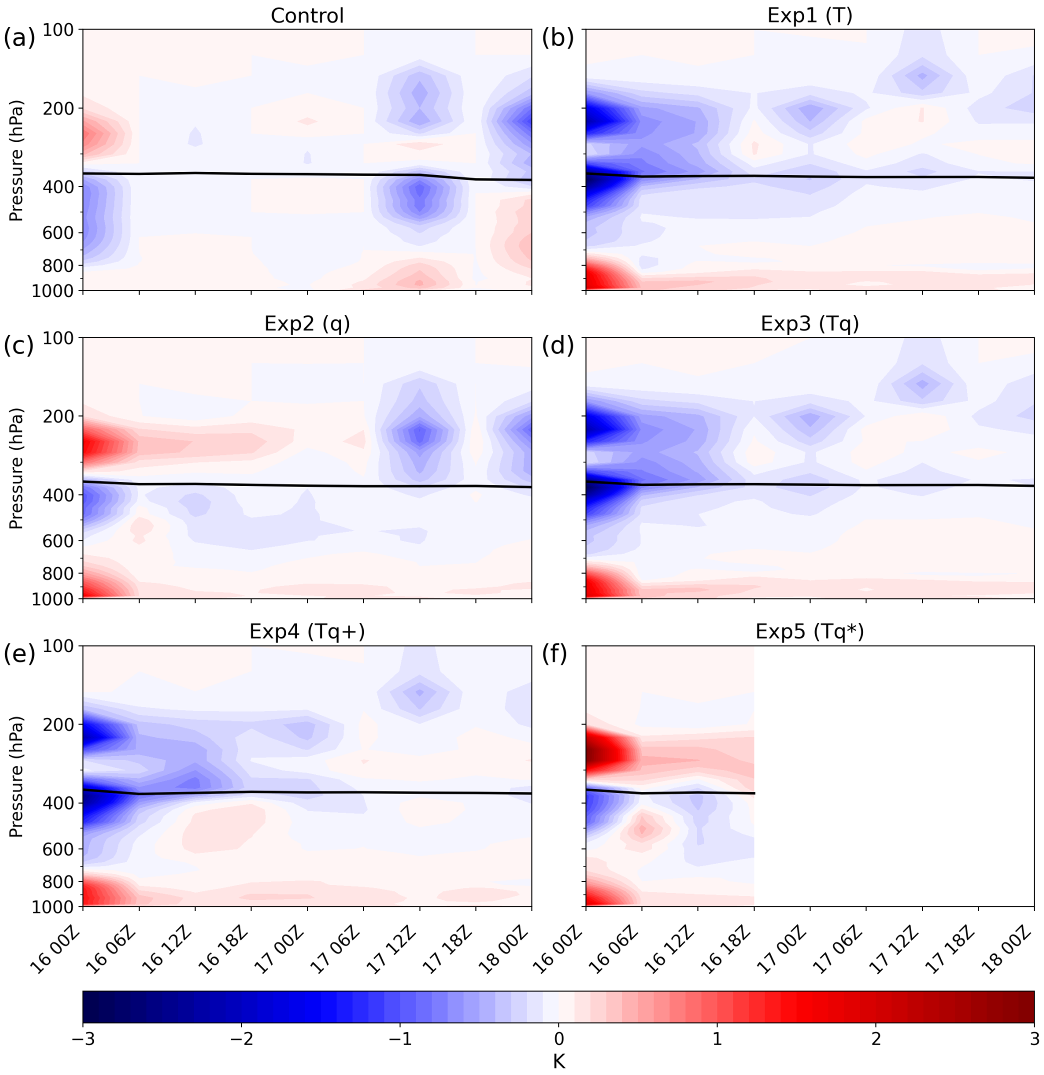


Figure 4.13: (a) Time-height plot of average potential temperature increments (K) within the core verification region in the Control cycling period. The solid black line indicates the ensemble average dynamic tropopause. (b) As in (a) but for Exp1 (T). (c) As in (a) but for Exp2 (q). (d) As in (a) but for Exp3 (Tq). (e) As in (a) but for Exp4 (Tq+). (f) As in (a) but for Exp5 (Tq*).

Humidity observations alone (i.e., Experiment 2) produce some of the same corrections via covariance with the temperature field (Fig. 4.13c). Notably, a PV anomaly strengthening dipole signature (warming above the tropopause and cooling below) emerges in the vertical composite. This dipole pattern occurs on one end of the TPV as in the temperature dropsonde experiments, while the cooling increments elsewhere

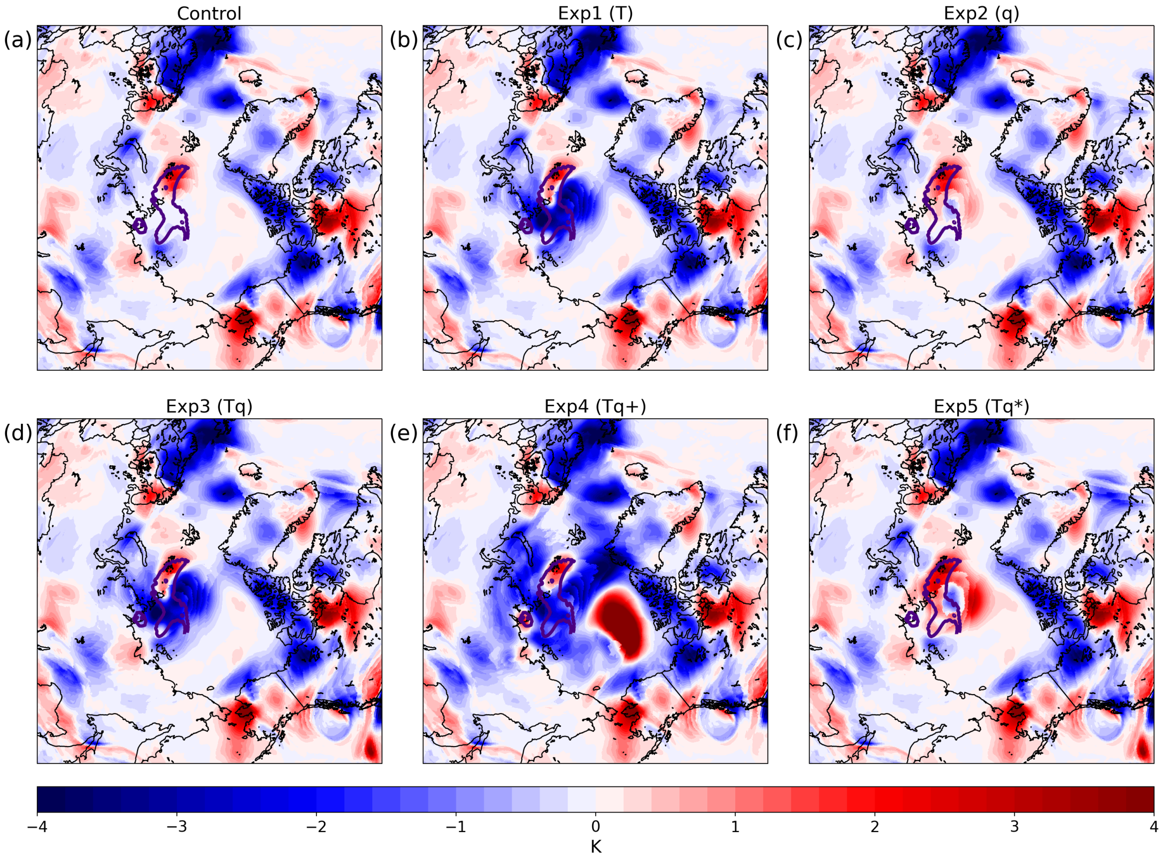


Figure 4.14: (a) Average lower stratosphere (defined as the four model levels, around 100 mb, above the tropopause) potential temperature increments (K) for the Control on 16 August 00 UTC. The purple line indicates the 300 K contour on the dynamic tropopause in ECO1280 within the moving verification region (i.e., what is included in the time-height plots of Figure 4.13). (b) As in (a), but for Exp1 (T). (c) As in (a) but for Exp2 (q). (d) As in (a) but for Exp3 (Tq). (e) As in (a) but for Exp4 (Tq+). (f) As in (a) but for Exp5 (Tq*).

in the lower stratosphere are not replicated by the humidity observations (Fig. 4.14b-e). To test the sensitivity of increments from temperature and humidity covariances to prescribed observation errors, we re-run the first day of cycling from Experiment 3 with the humidity observation errors halved (Exp5 Tq*). While Experiment 3 is very similar to Experiment 1, suggesting that the temperature observations overwhelmed the impact of the humidity observations, Experiment 5 appears to be an amplified version of Experiment 2. The same cross-tropopause temperature increment dipole (Fig. 4.13f) and broad lower stratospheric warming (Fig. 4.14f) as in Experiment 2 are present, with little evidence of the expansive cooling driven by the temperature observations. These sensitivities to observation errors suggests that there are complex correlation structures between temperature and humidity near the tropopause. This may limit the degree to which the two observation types can provide updates in tandem.

Specific humidity increments illustrate a similar disconnect between the temperature and humidity observations. Humidity observations alone act to uniformly dry the UTLS in the vicinity of the TPV, which is expected given the known moisture bias in the region (Fig. 4.15c). The bulk of these drying increments are in the same region of the TPV where the temperature observations warmed the lower stratosphere (i.e., both sets of observations are working to uniformly strengthen the TPV in this area), while limited increments occur elsewhere 4.16c). The temperature observations also produce negative humidity increments in some cases (around the tropopause initially and in the upper tropopause later in the cycling period) but actually moisten the UTLS in others (near 250 hPa initially and around the tropopause in the first day of cycling; Fig. 4.15b,d,e). As before, when temperature and humidity observations are assimilated together (Experiments 3 and 4), the humidity increments replicate the erroneous moistening of Experiment 1.

Also as in the temperature increments, reducing the humidity observation errors allows the humidity observations to make a larger impact, even when assimilated with temperature (Figs. 4.15f, 4.16f). Given the importance of vertical moisture gradients in longwave PV creation, the spurious addition of moisture into the UTLS by the temperature observations during the cycling period likely impacts the evolution of the TPV during the forecast period. PV generation will already be diminished relative to the truth in all experiments due to the starting moist bias, but only Experiments 2 and 5 produce a consistent reduction of this bias during the cycling period. This likely

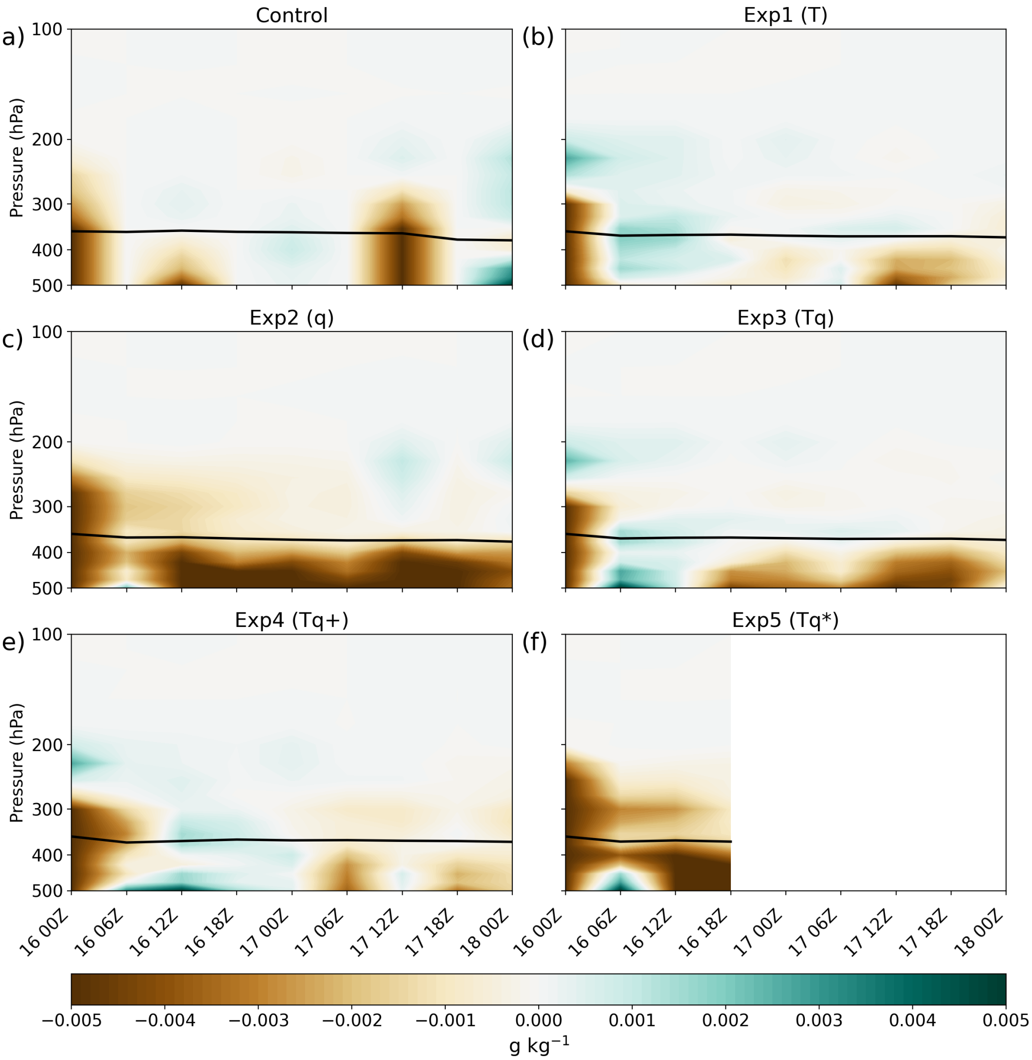


Figure 4.15: As in Figure 4.13, but for specific humidity increments (g kg^{-1}). Note that the vertical scale differs from Figure 4.13.

helps to explain why Experiment 2 starts with larger biases than the other experiments but evolves closer to them as the forecast progresses.

Although we have established that the correlation based moisture increments from temperature observations overwhelm the direct humidity observations due to their relative observation error variances, it is still unclear why the temperature and humidity increments are at odds with each other in the upper atmosphere alone. Generally, one might expect the two to be positively correlated within the ensemble, as warmer

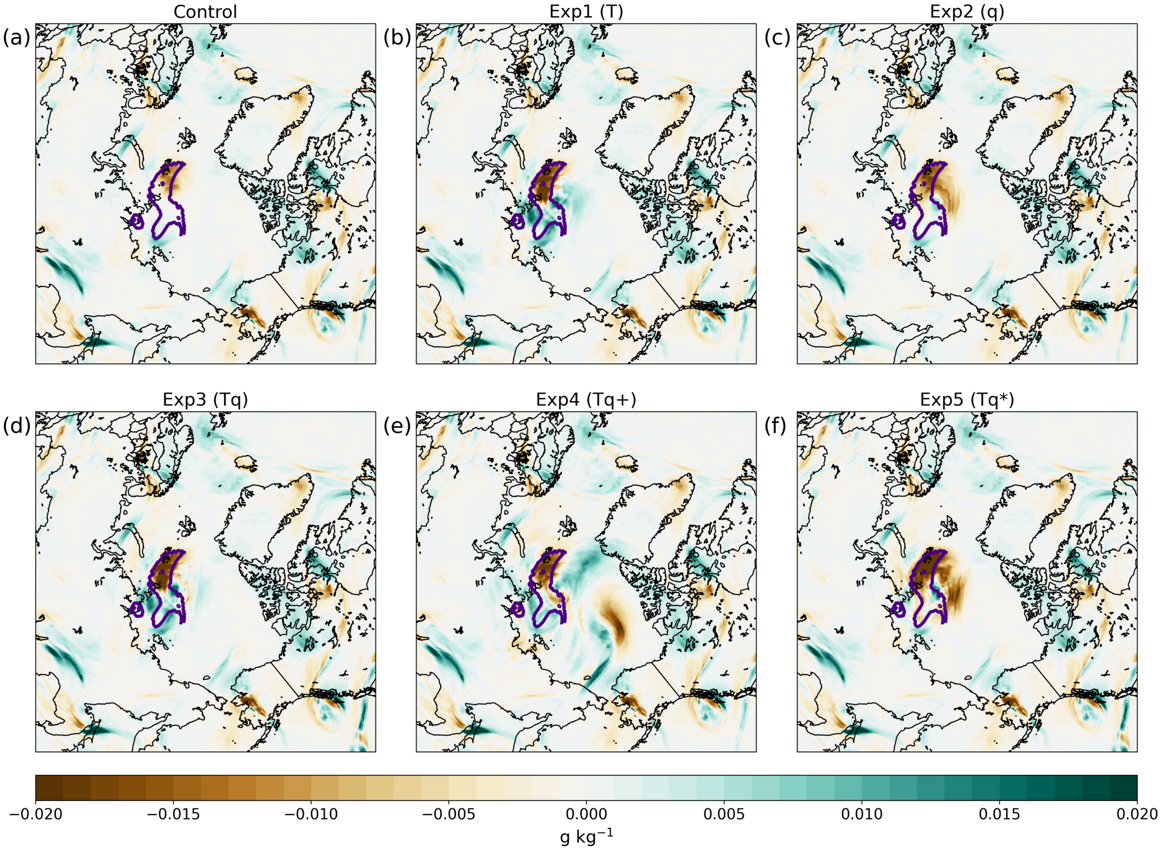


Figure 4.16: As in Figure 4.14, but for specific humidity increments (g kg^{-1}).

environments will tend to have more moisture and vice versa. In the middle and lower troposphere, this appears to be the case, with broadly positive correlations between humidity and temperature in the surrounding area (Fig. 4.17c,d). On the other hand, at 250 hPa, this correlation is reversed, so that cooling lower stratosphere increments would correspond with moistening as well (Fig. 4.17a,b). As temperatures increase and moisture decreases with height above the tropopause, this reversed correlation is reasonable aloft. Closer to the mean tropopause level where competing increments are also seen in Figures 4.13 and 4.15, general complexities within the tropopause structure of different ensemble members are a possible cause. In particular, we note that many of the cooling increments with moistening cross correlations happen along the edge of the TPV, a region that would include a sharp tropopause transition in any individual member. Overall, it seems that a better balance between prescribed temperature and humidity observation errors is needed in these regions to ensure that direct observations overshadow the correlation based updates.

To better quantify the importance of UTLS moistening or drying on PV creation or destruction, we analyze ensemble average PV tendencies from longwave radiative processes calculated by MPAS. Compared to the control run, which included some PV generation below the tropopause and mild PV destruction above throughout the experimental period, each of the temperature dropsonde experiments produce greater PV creation in the upper troposphere and greater PV destruction in the lower stratosphere (Fig. 4.18a,c,d). If the observations were purely strengthening the vortex, it may be expected that PV creation would increase throughout the UTLS. This dipole, then, is indicative of a more complicated signal than inform strengthening, instead suggesting height-varying alterations to the vertical structure of the TPV. This effect is likely combined with improvements in the initial TPV structure and increased upper troposphere PV generation provided by the TPV observations to produce better overall forecasts relative to the control, as discussed above.

Experiment 2, on the other hand, broadly reverses this trend. More PV is created in the lower stratosphere than in the control, with a mix of PV creation and destruction at and below the tropopause (Fig. 4.18b). Once again, the signal is not one of clear strengthening or weakening, but rather vertically-varied changes. Still, the pattern is notably different compared to the temperature dropsonde experiments, which highlights the impact of the moisture observations. It is also important to note that because Experiment 2 produced a great reduction of the overall moist bias, the magnitudes of the PV tendencies are also uniformly reduced. Experiment 3 exhibits modestly different tendencies in the lower stratosphere compared to Experiment 1, but as before, the impact of the humidity observations seems to have been limited. Turning to a spatial view of lower stratosphere PV tendencies (where the strongest negative

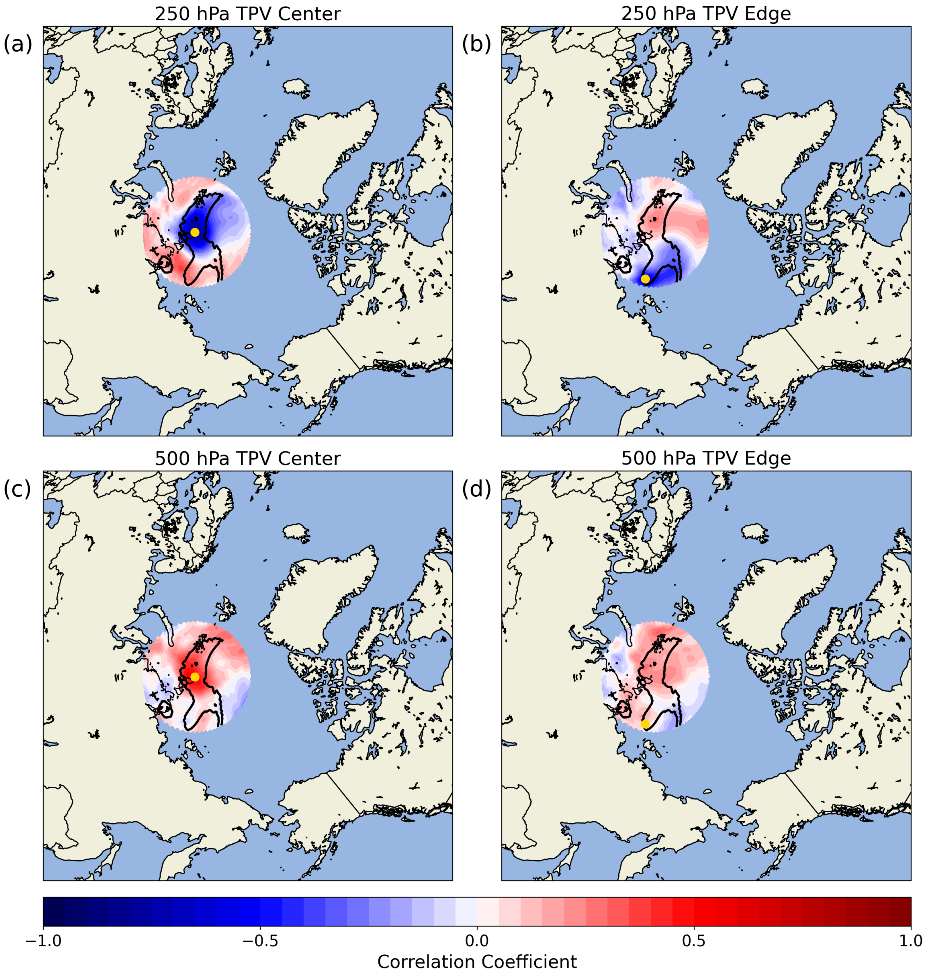


Figure 4.17: (a) 16 August 00 UTC ensemble correlation coefficient between 250 hPa specific humidity (g kg^{-1}) at the yellow dot in the TPV core and 250 hPa temperature (K) at each point in the verification region. The purple line indicates the 300 K contour on the dynamic tropopause in ECO1280. (b) As in (a), but for a point on the edge of the TPV. (c) As in (a), but for humidity and temperature at 500 hPa. (d) As in (b), but for humidity and temperature at 500 hPa.

tendencies are present in the temperature dropsonde experiments) at the beginning of the forecast period highlights some of these these differences (Fig. 4.19). Each of the temperature observation experiments include PV destruction throughout the TPV core region at this time, while Experiment 2 exhibits a mixture of limited positive and negative tendencies. The stark contrasts between tendencies in the TPV core and in

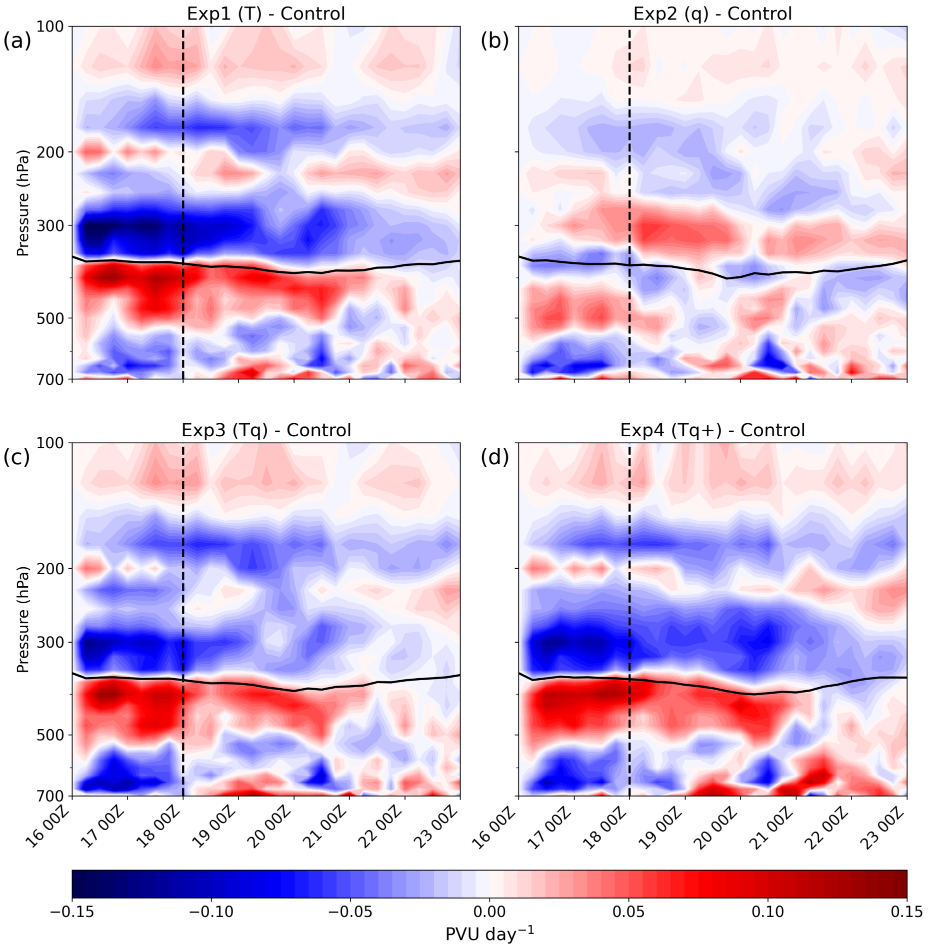


Figure 4.18: (a) Time-height plot of differences in potential vorticity tendency due to longwave radiative processes (PVU day^{-1}) between Exp1 (T) and the Control within the core verification region. The solid black line indicates the ensemble average dynamic tropopause. The dashed black line marks the start of the forecast period. (b) As in (a), but for Exp2 (q). (c) As in (a), but for Exp3 (Tq). (d) As in (a), but for Exp4 (Tq+).

the surrounding environment in each experiment also highlight the spatial intricacies of the processes at play.

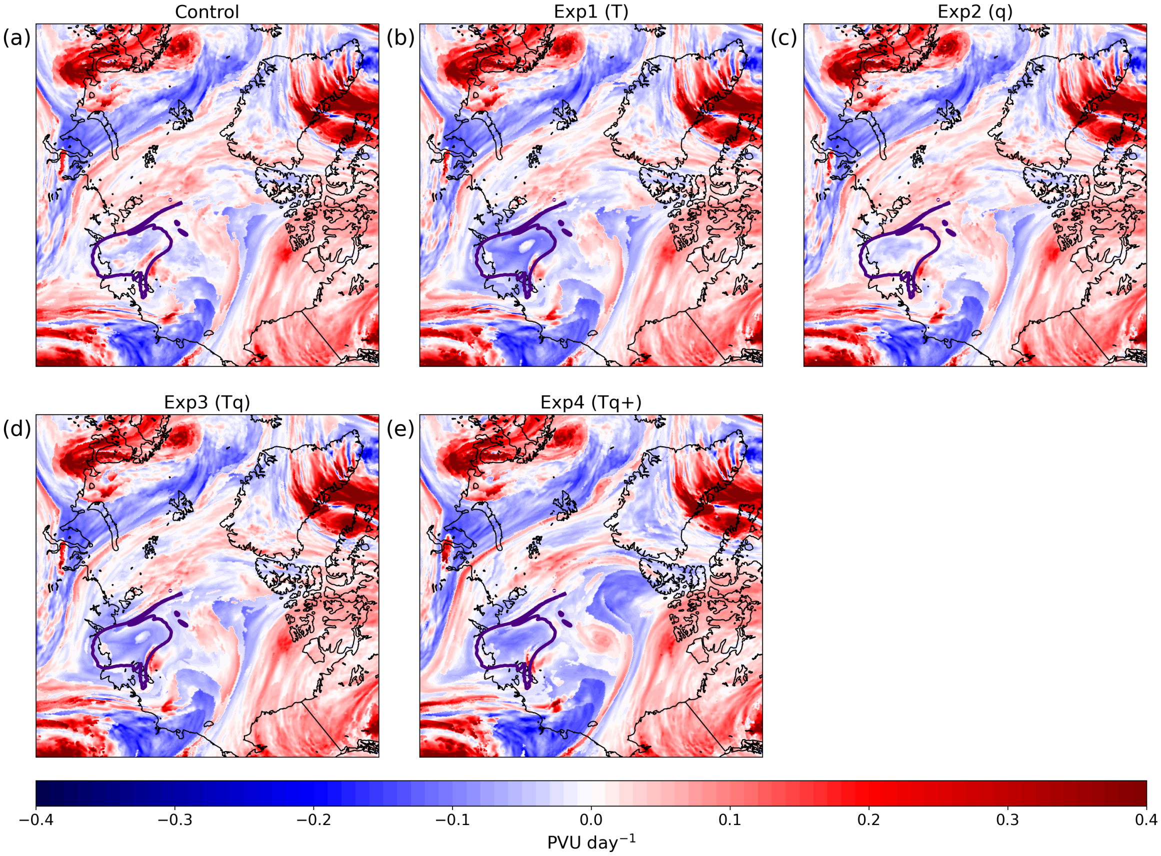


Figure 4.19: (a) Average lower stratosphere potential vorticity tendency due to longwave radiative processes (PVU day^{-1}) for the Control on 18 August 00 UTC. The purple line indicates the 300 K contour on the dynamic tropopause in ECO1280. (b) As in (a), but for Exp1 (T). (c) As in (a), but for Exp2 (q). (d) As in (a), but for Exp3 (Tq). (e) As in (a), but for Exp4 (Tq+).

4.2 TPV and Cyclone Characteristics

4.2.1 Ensemble Feature Tracks and Spatial Structures

Having analyzed the impacts of the simulated dropsondes on large-scale atmospheric fields, we now examine changes made to the evolution of the primary TPV and AC, both in the ensemble mean and in individual members. Ensemble mean TPV and AC tracks (created with TPVTrack and TempestExtremes, respectively) show relatively little variation among the experiments (Fig. 4.20). The initial TPV track is represented

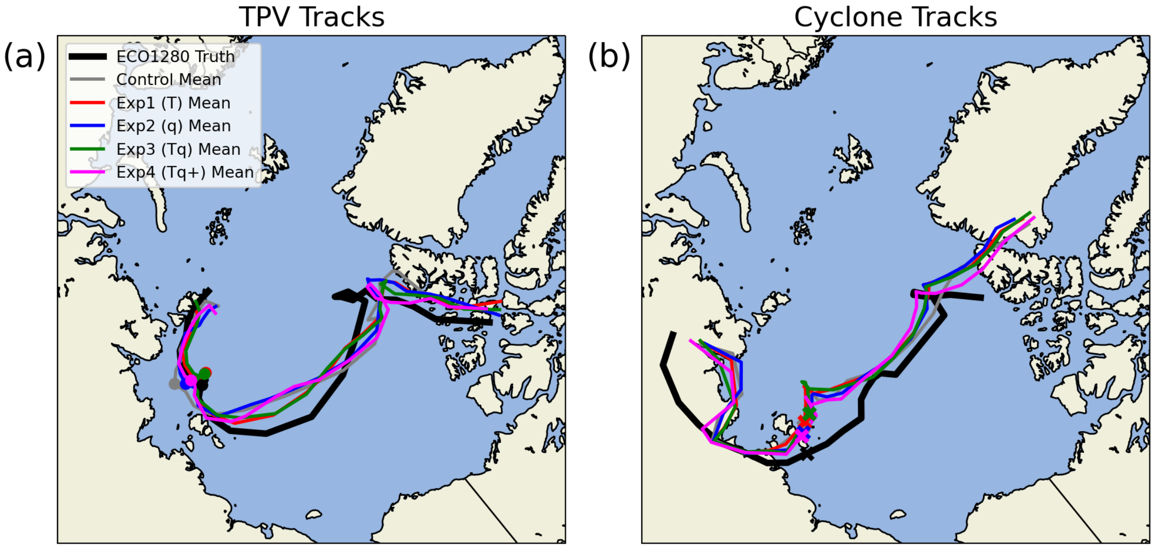


Figure 4.20: (a) The primary ECO1280 TPV track (thick black) and ensemble mean TPV tracks from each of the OSSE experiments. The position of each TPV at the end of the cycling period is represented with a dot. (b) As in (a), but for surface cyclone tracks, with the position of each AC at the end of the cycling period represented with an X.

well in all cases, with perhaps slightly more detail captured during the cycling period in the dropsonde experiments than in the control. As the system moves across the Arctic at the end of the cycling period, each ensemble mean shows a similar bias with the TPV track pushed deeper into the Arctic than in ECO1280. As a result of this initial displacement, none of the simulated TPVs correctly resolve the location of the TPV at the time of maximum cyclone strength. Experiments 2 and 4 come the closest to capturing the looped track seen in ECO1280 but still produce the feature around 250 km away. It is important to recall that these TPV tracks are point metrics following only the vortex core. Moreover, differences between the experimental ensembles and ECO1280 may arise from differences in the physical parameterizations used in each model, in addition to differences in the initial conditons. We will further explore changes in the spatial structure of the TPVs in Section 4.3. The mean AC tracks show similar trends (Fig. 4.20b). All of the forecast ensembles match ECO1280 well during the cycling period (the initial disparity is due to ensemble members in which the correct cyclone has not yet formed), but begin to diverge as they enter the Arctic, with the ensemble means pulling far to the west of the true track. In addition, the cyclones move across the Arctic relatively more slowly, with more time spent looping north of Siberia. The MPAS ensembles also diverge from ECO1280 during the later stages of the system, with the AC moving rapidly into Greenland as it decays.

Evaluating TPVTrack's diagnosed TPV characteristics over time, an interesting disconnect emerges between the point metric calculations and our qualitative assessments of the TRPT field in Section 4.1. The control experiment shows a lower core potential temperature than any of the experiments throughout the forecast period, even though the control TPV appeared visually weaker (Fig. 4.21a). On the other hand, the average TPV radius in the control is near or below that of the other experiments (Fig. 4.21b). It seems, then, that the dense dropsondes over the TPV core have a smoothing effect; although tropopause potential temperatures over a large region are decreased and the structure of the anomaly improved, the deepest mesoscale core structure has changed. This, once again, suggests that point metrics may not be sufficient for an analysis of changes to TPV structure. Errors in the location of TPVs, on the other hand, are much easier to diagnose with single point metrics. The control generally performs the worst in terms of mean TPV position, as expected from analyzing the spatial fields (Fig. 4.21c). Experiments 1 and 3 (with dense temperature observations) unsurprisingly perform best initially, with errors of less than 100 km as the TPV begins to move across the Arctic. As the cyclone reaches its maximum intensity, however, Experiments 2 and 4 become the best performers. We speculate that this shift is because of improved representation of PV tendencies and the surrounding synoptic environment, respectively.

Turning to the evolution of the surface low, two key points emerge (Fig. 4.22). First, despite briefly possessing similar minimum MSLP and MSLP amplitude (a measure of the size of the closed surface circulation) values to ECO1280 around 18 August 18

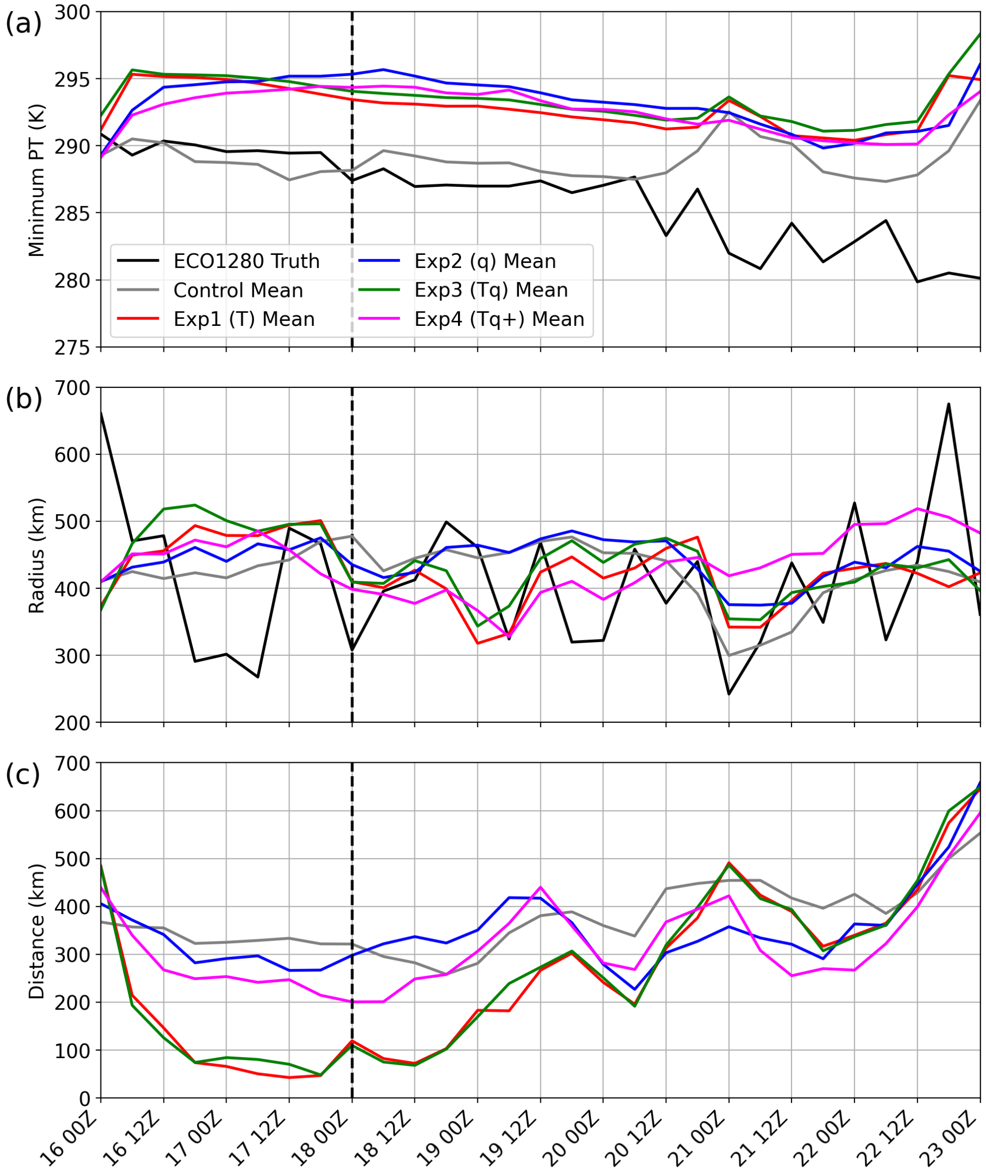


Figure 4.21: (a) Time series of ECO1280 and ensemble mean TPV minimum potential temperature (K) for each OSSE experiment. (b) As in (a), but for TPV radius (km). (c) Time series of the distance between the ensemble mean TPV and the ECO1280 TPV (km) for each OSSE experiment.

UTC, all of the MPAS ensembles diverge from the truth during the forecast and never reach the correct maximum strength. Second, the primary deepening that does occur in the experiments happens around a day earlier than the ECO1280 rapid deepening period. Starting on 18 August 18 UTC, the experimental ACs undergo a brief period of central weakening, while ECO1280 deepens slightly. Then, on 19 August at 12 UTC, the MPAS ACs begin a one day strengthening period to their eventual minimum pressures and amplitudes. In the end, Experiments 2 and 4 produce the strongest cyclones by modest margins (around 2 hPa stronger in minimum pressure and 3 hPa in amplitude relative to the control). The ACs then rapidly decay into the end of the forecast period. ECO1280, on the other hand, strengthens quickly on 20 August, reaching deeper pressures and far larger amplitudes than any of the experiments. The distance error between ECO1280 and the ensemble mean is fairly consistent throughout the forecast, with the control performing slightly worse than the dropsonde experiments (Fig. 4.22c). This early and insufficient deepening of the cyclone, along with the known errors in TPV position, suggest that the spatial relationship between the TPV and AC over time may be an important factor in its development.

An examination of the distances between the TPV and AC over time in the various experiments appears to confirm this idea. At the start of the forecast period, TPV-AC spacing is between 550 and 650 km for all ensemble means and ECO1280 (Fig. 4.23). From this point, the spacing in all MPAS ensembles decreases to between 300 and 400 km, before gradually increase as the cyclone matures and decays. In ECO1280, on the other hand, the distance decreases gradually over the course of the 3 day forecast, eventually reaching near 0 km as the cyclone matures. As discussed in Chapter 1, TPVs provide baroclinic forcing to the surface cyclone as it develops. This type of traditional baroclinic development requires that the system have a certain degree of vertical tilt. Because the systems in the MPAS ensemble became vertically stacked so quickly, the cyclones have a reduced period of optimal baroclinic growth and never become fully coupled with the TPV, lowering their ultimate strength. It is important to note, though, that Figure 4.23 may underestimate the degree of vertical coupling during the mature stage of the cyclone due to the point nature of the tracks. For

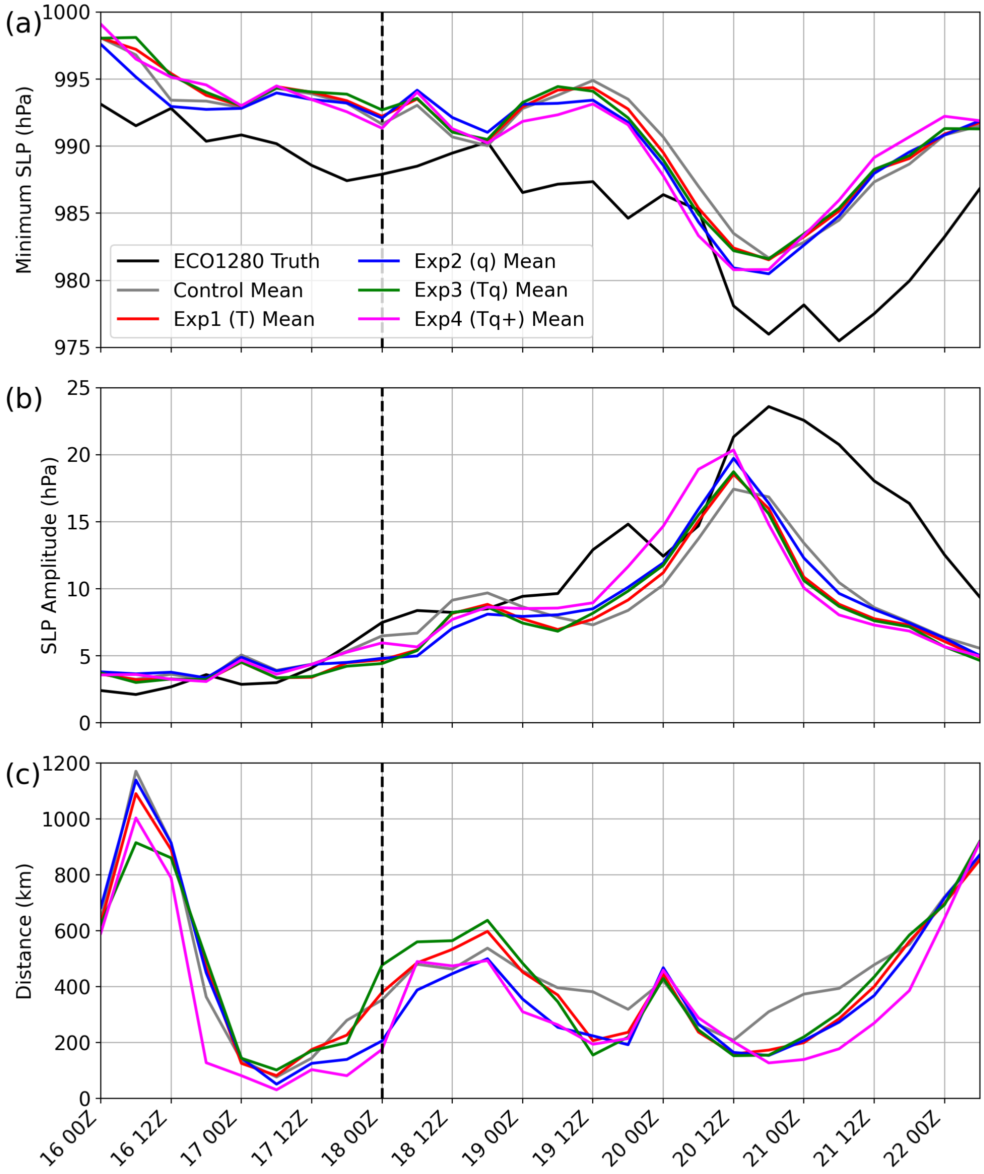


Figure 4.22: (a) Time series of ECO1280 and ensemble mean AC minimum MSLP (hPa) for each OSSE experiment. (b) As in (a), but for MSLP closed contour amplitude (hPa). (c) Time series of the distance between the ensemble mean AC and the ECO1280 AC (km) for each OSSE experiment.

example, Figure 4.8 indicates that the system is far more coupled in Experiment 1 than in the control, but this difference is not reflected in the tracks.

This early coupling between the TPV and AC is likely linked to the spatial displacement of the mean TPV tracks and slower advancement of the cyclone across the Arctic.

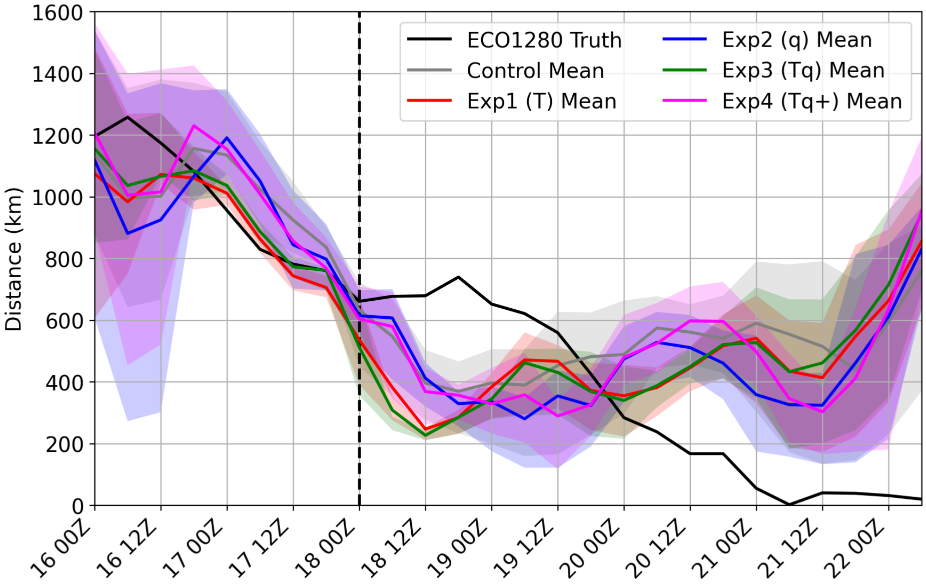


Figure 4.23: Time series of the distance between the TPV and AC in ECO1280 and the distances between ensemble mean TPVs and ACs in each of the OSSE experiments. Shading represents the interquartile range of individual ensemble members.

While local dynamic and physical processes likely play a role in these changes, examination of the TRPT field suggests that larger-scale environmental conditions drive the TPV location errors. In particular, the Rossby wave ridge breaking into the Arctic during the forecast period (which seems to be forced by a midlatitude surface cyclone) appears to progress too far poleward in each of the MPAS ensemble means (Fig. 4.24). This would, in turn, explain the displacement of the TPV tracks further into the Arctic away from the ridge and the poorly simulated horizontal spacing of the AC and TPV. This idea will be explored further using individual ensemble member forecasts in the following section.

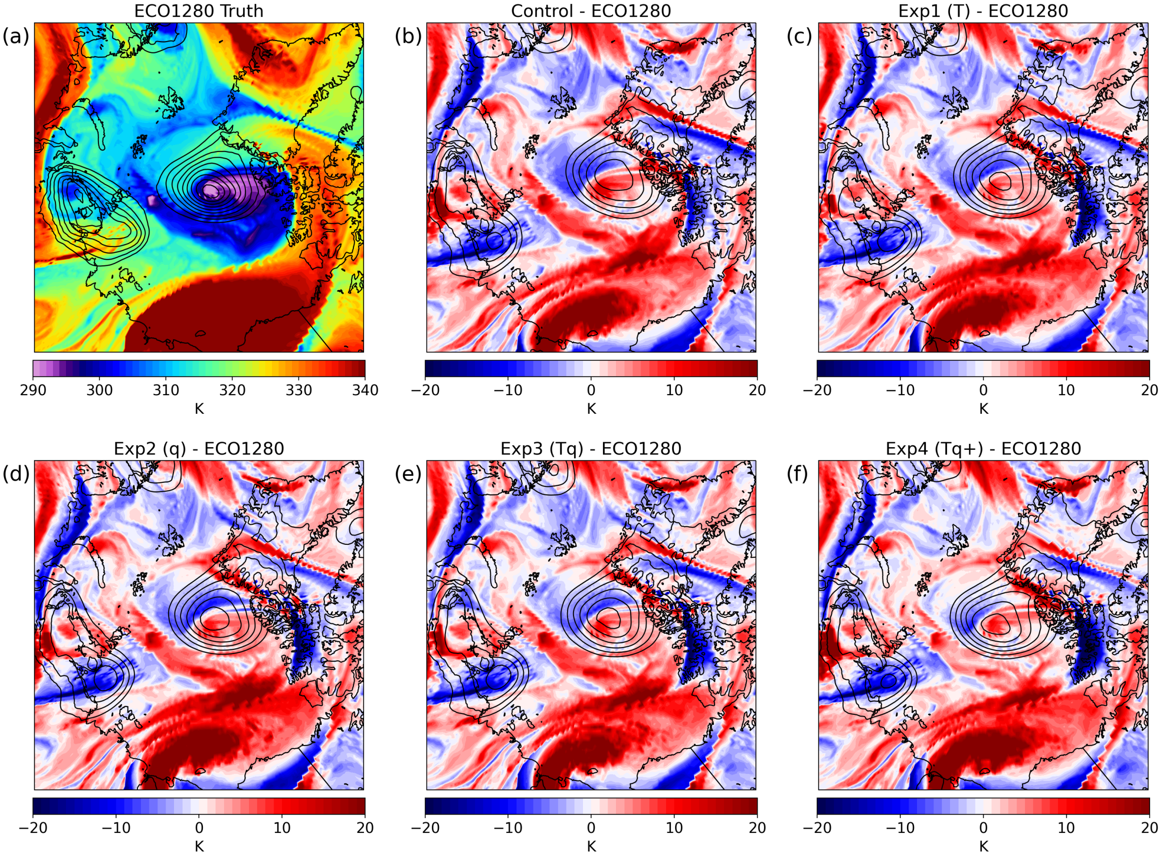


Figure 4.24: (a) ECO1280 TRPT (fill; K) and MSLP (contours; hPa; every 2.5 hPa starting at 1000 hPa) on 21 August 00 UTC. (b) Difference between the control ensemble mean TRPT and ECO1280 TRPT (fill; K) on 21 August 00 UTC. Contours represent control ensemble mean MSLP. (c) As in (b), but for Exp1 (T). (d) As in (b), but for Exp2 (q). (e) As in (b), but for Exp3 (Tq). (f) As in (b), but for Exp4 (Tq+).

4.2.2 Best, Mean, and Worst Performing Ensemble Members

Although the ensemble means examined so far provide the best estimate of the state of each experiment, certain dynamic features of interest (like the interactions between the TPV and breaking Rossby wave or impact of small scale PV features) are best studied in individual forecasts. To select individual members for study, we use two ranking methods on each ensemble (which are described in Section 2.2), targeting different physical processes of interest. These methods are TPV-AC distance (in which the best member has the best representation of the TPV-AC spacing during the forecast period) and standardized TPV error (in which the best member most accurately simulates the point metric strength and location of the ECO1280 TPV). The standardized TPV error method is weakly (though, significantly, using a Monte Carlo simulation method) correlated with AC strength (Fig. 4.25a). This weak relationship is reasonable as neither the spatial characteristics of the TPV or spatial complexities in individual members are taken into account. This method is primarily used to investigate how individual member TPVs differ from truth, and so a strong correlation to cyclone strength is not essential. The correlation between cyclone strength and the TPV-AC distance is substantially stronger, as would be anticipated from the results above. This method is best suited for examining the range of possible AC representations within the ensemble.

Interestingly, the TPVs and ACs in the best cases using the TPV-AC distance method do not necessarily seem to follow the ECO1280 tracks any more than in the ensemble mean (Fig. 4.26a,b). So, it seems in these cases that even with errors in TPV position, complimentary errors in AC position allowed the relative distance between the two to remain accurate. So, while these members may not produce an ideal forecast of the system's position relative to ECO1280, they capture the baroclinic development far more accurately. The mean and worst performing members in terms of TPV-AC

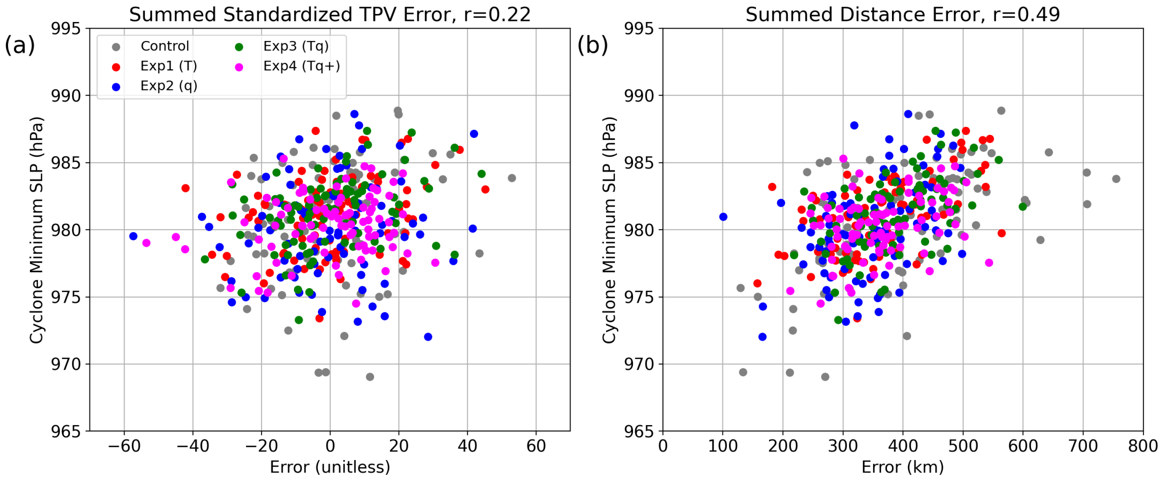


Figure 4.25: (a) Minimum MSLP reached at any point during a given member's forecast versus the standardized TPV error for that member summed over the forecast time (negative values indicate the TPV was better modeled throughout the 5 day forecast). The Pearson correlation coefficient is provided in the title. (b) As in (a) but for the TPV-AC distance error summed over the forecast time (lower values indicate the TPV-AC spacing was better modeled throughout the 5 day forecast).

distance highlight the amount of spread present in the ensemble solutions, even after two days of detailed corrections to the initial analyses (Fig. 4.26c-f). Various members appear to completely miss the TPV or AC at points in the forecast (signified by large jumps in the track maps) or follow very different development pathways from ECO1280.

Examining time series of AC development in the individual cases, we find both support for the importance of TPV-AC distance and evidence that other factors often play an important role. The best individual members capture the cyclone strengthening far better than the ensemble mean, reaching minimum pressures much nearer to the ECO1280 value (Fig. 4.27a,b). The weakening period seen in the mean on 19 August is lessened, and the cyclones deepen at a similar rate to the ECO1280. Further, although this deepening begins earlier than in ECO1280 (since the TPV track positions were not fully rectified), the lag time is reduced by 12 hours. The location errors in the best member ACs also remain low throughout the mature stage of the cyclone. The best members from Experiments 1 and 4 in particular replicate the ECO1280 AC almost

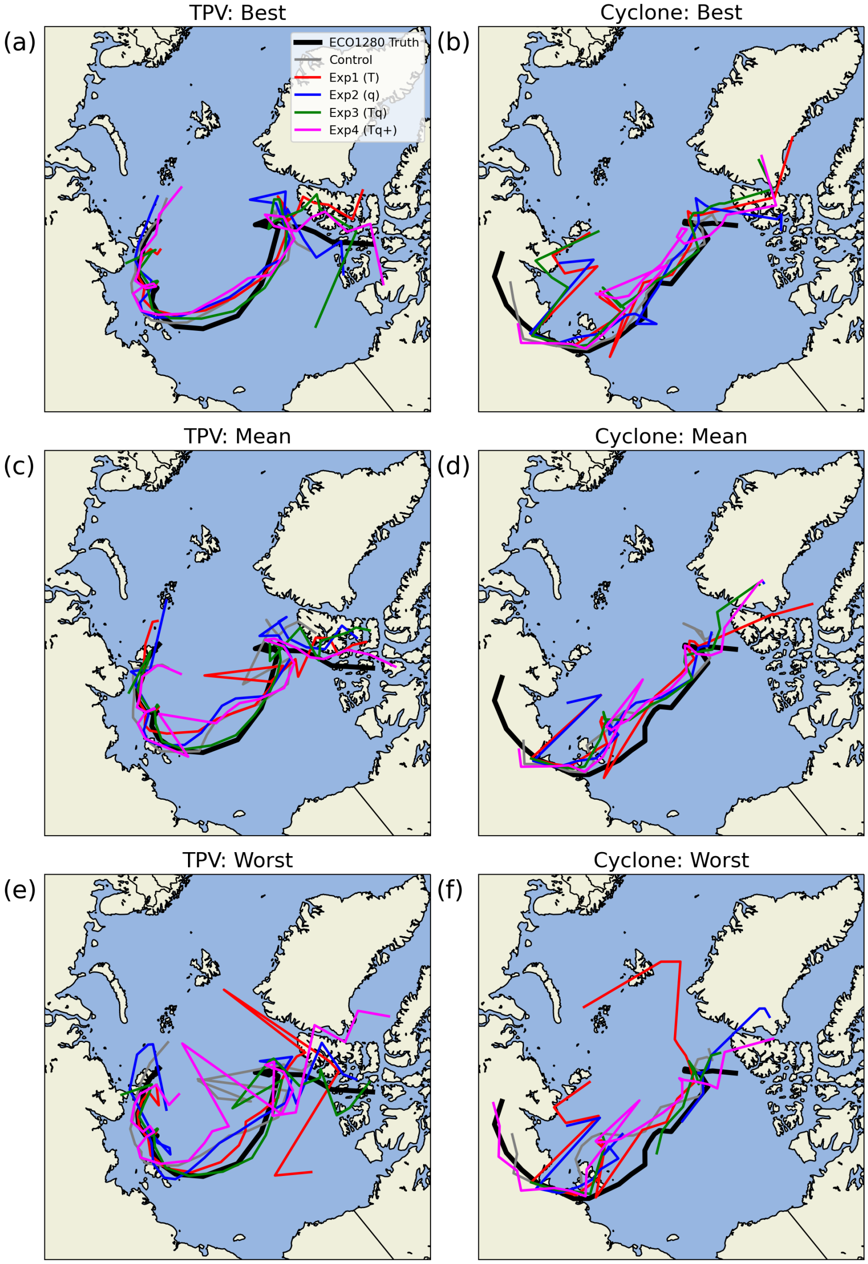


Figure 4.26: (a),(b) As in Figure 4.20, but for the best individual ensemble members using the TPV-AC distance method. (c),(d) As in (a),(b), but for the ensemble members nearest to the ensemble mean performance. (e),(f) As in (a),(b), but for the worst ensemble members.

exactly as it reaches its peak intensity. By and large, the mean and worst members follow the patterns of the ensemble mean, with ACs that develop too early and don't deepen enough (Fig. 4.27c-f). Some members (like the worst performing member of

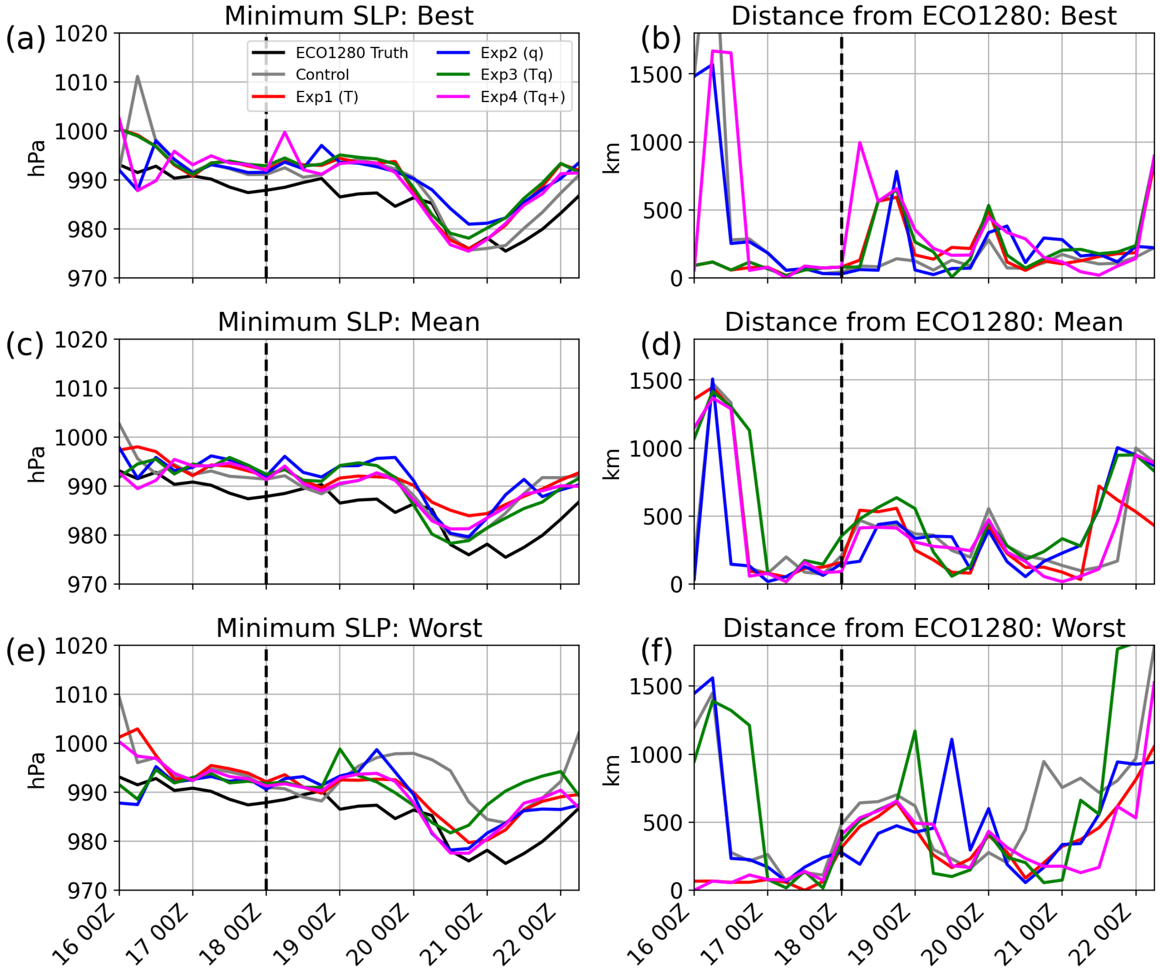


Figure 4.27: (a) Time series of AC minimum MSLP (hPa) for ECO1280 and the best individual ensemble members using the TPV-AC distance method. (b) Time series of the distance between the AC in the best ensemble member and the ECO1280 AC (km). (c),(d) As in (a),(b), but for the ensemble members nearest to the ensemble mean performance. (e),(f) As in (a),(b), but for the worst ensemble members.

Experiment 4), however, are able to fairly accurately simulate the cyclone development even with poor representation of TPV-AC spacing.

The best members in terms of TPV error, on the other hand illustrate that the strength of the upper level feature is still important, despite the lower correlation coefficients in Figure 4.25. Most of the members selected by this metric experience an initial weakening period, undergo early deepening, and have relatively high TPV-AC distance errors (Figs. 4.28 and 4.29). In spite of this, the members with the

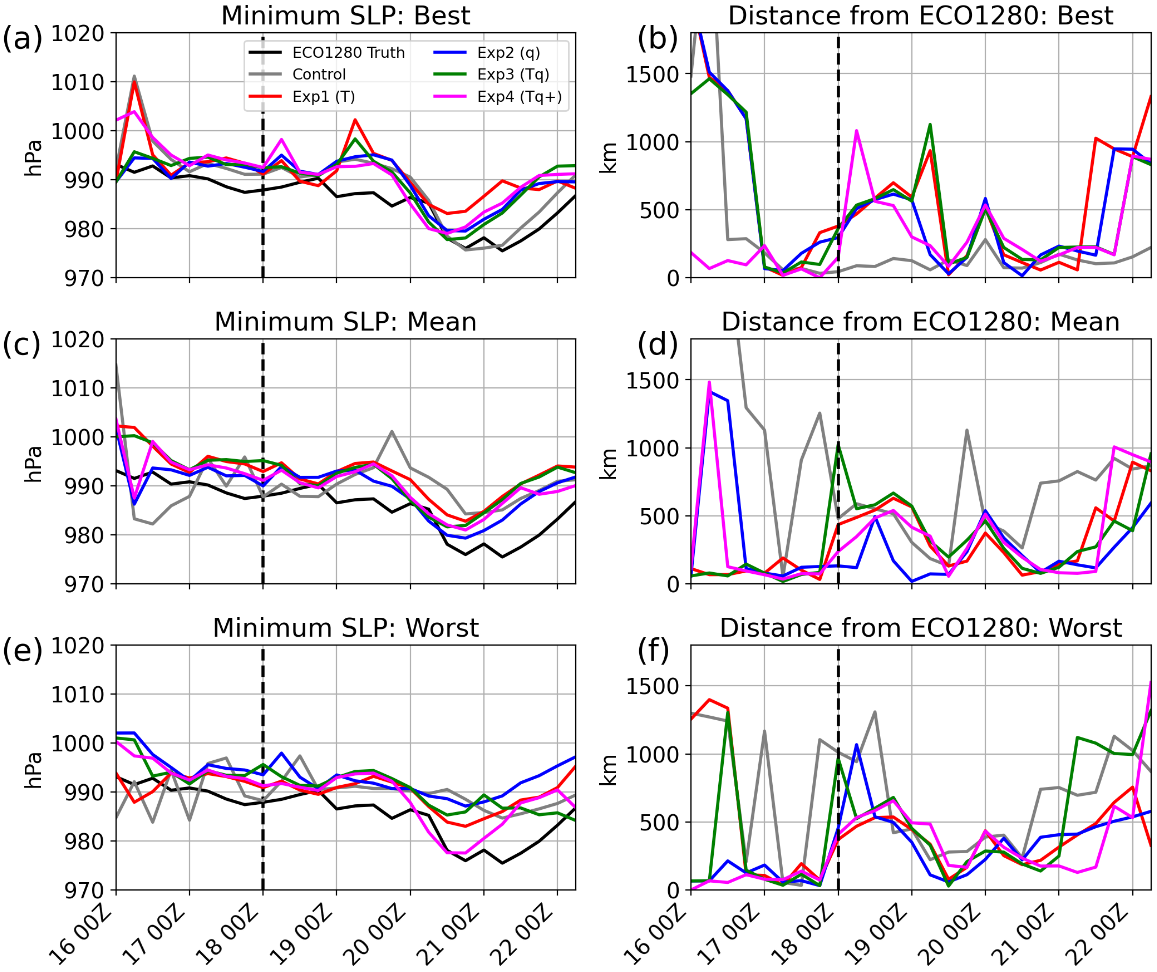


Figure 4.28: As in Figure 4.27, but for the standardized TPV error method.

lowest TPV error produce relatively strong cyclones (around 980 hPa) compared to those with average (around 983 hPa) and poor (around 987 hPa) representations of the TPV. In the end, then, it seems that the combination of a sufficiently strong TPV with proper spacing from the surface cyclone throughout the forecast is necessary to produce an optimal forecast. The TPV strength controls how much the cyclone is able to deepen, while the horizontal spacing determines when and for how long the strengthening occurs.

Examining the spatial structure of the best and worst members from the two selection methods allows us to better understand how changes in the TPV-AC distance

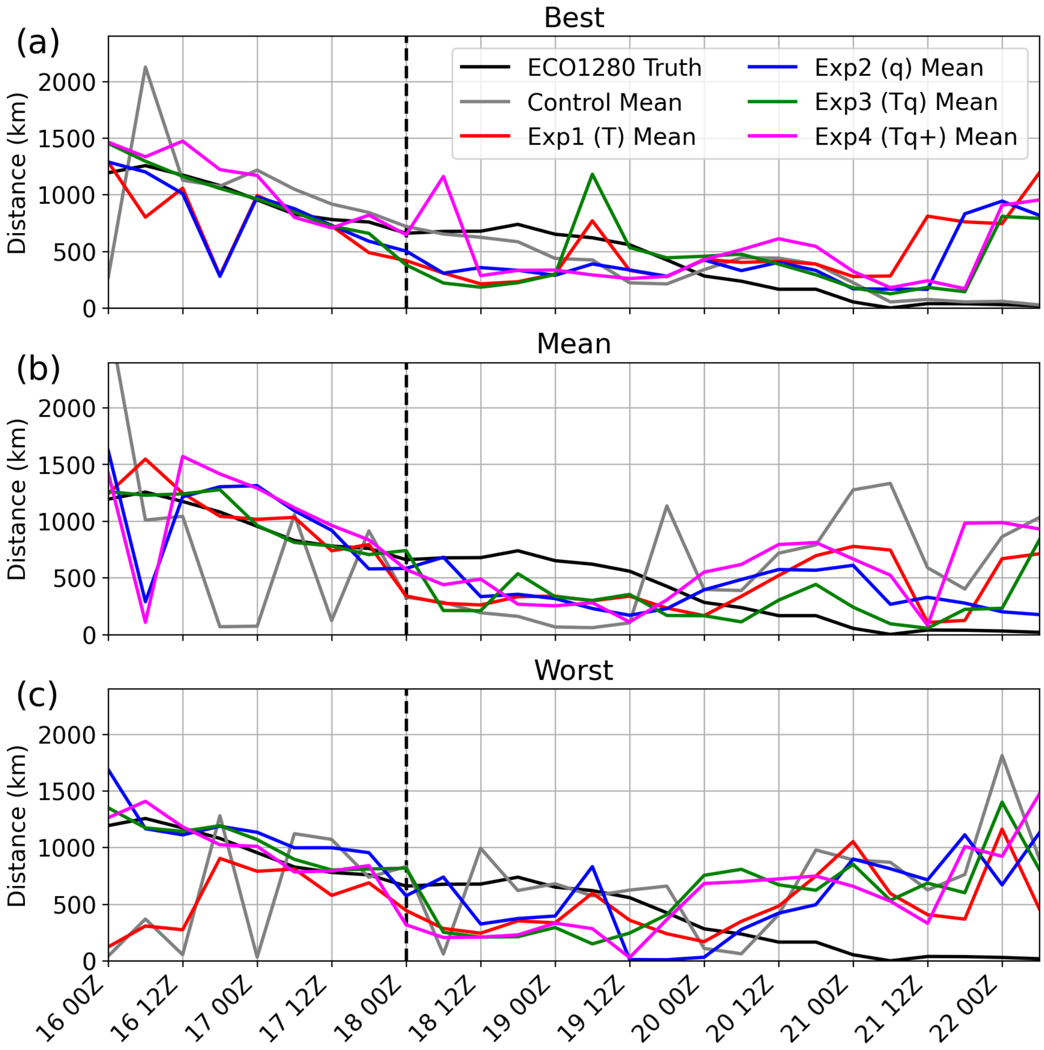


Figure 4.29: (a) Time series of the distance between the TPV and AC in ECO1280 and the distances between TPVs and ACs in the best individual ensemble members using the standardized TPV error method. (b) As in (a), but for the ensemble members nearest to the ensemble mean performance. (c) As in (a), but for the worst ensemble members.

and TPV strength play out as the cyclone reaches its maximum strength. The best performing members by the distance metric feature a wide variety of mesoscale TPV structures, but all feature vertical alignment with the surface cyclone that closely matches ECO1280 (Fig. 4.30). Moreover, in most of the high-performing cases, the streamer of high PV wrapping out of the main TPV is relatively well represented.

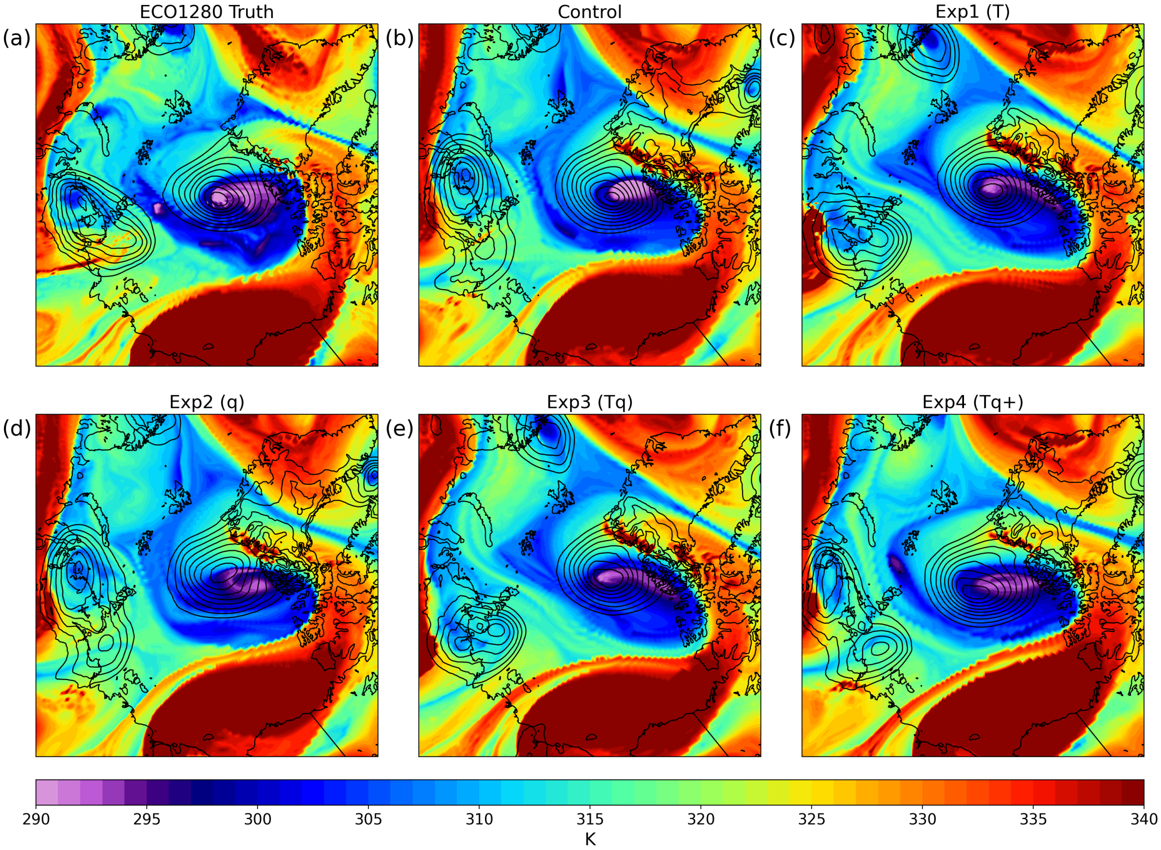


Figure 4.30: (a) ECO1280 TRPT (fill; K) and MSLP (contours; hPa; every 2.5 hPa starting at 1000 hPa) on 21 August 00 UTC. (b) As in (a), but for the best performing control ensemble member using the TPV-AC distance method. (c) As in (b), but for Exp1 (T). (d) As in (b), but for Exp2 (q). (e) As in (b), but for Exp3 (Tq). (f) As in (b), but for Exp4 (Tq+).

Because the actual location of the system was not considered in the error metric, various members include Rossby wave breaks that have progressed too far (Experiment 2) or not far enough (Experiment 3) into the Arctic. The worst members, in contrast, exhibit weaker ACs, TPV cores that are far displaced from the cyclone center, and larger errors in the synoptic pattern (Fig. 4.31). Some of the worst performing members (e.g., Experiment 3) include especially large spatial errors earlier in the forecast period, which are not included here for brevity.

The best and worst members by the TPV error method provide a somewhat cleaner spatial contrast. Despite larger variation in AC strength than in the distance method

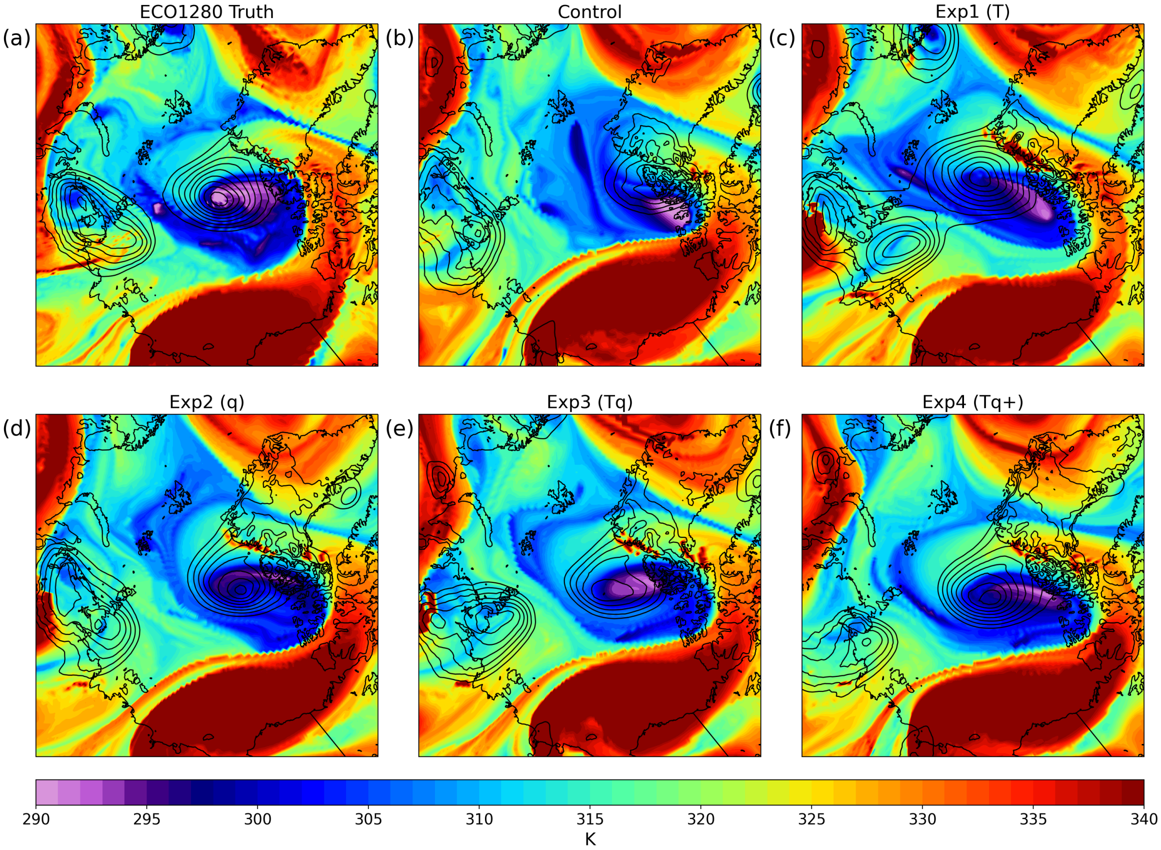


Figure 4.31: As in Figure 4.30, but for the worst performing ensemble members using the TPV-AC distance method.

(since the baroclinic growth period may have been shortened earlier in the forecast), the best performing members all closely resemble the upper-level structure of ECO1280 (Fig. 4.32). Minimum tropopause potential temperature values approaching those in the truth occur almost directly over the cyclone center, and the overall mesoscale PV structure is captured thoroughly. Conversely, the worst performing members include TPVs that are very poorly positioned with much different dynamical structures than in ECO1280 (Fig. 4.33). Interestingly, it seems that the corrections to the initial state in Experiment 4 were so extensive that none of the individual members performed that poorly, and so very small variations made the difference between the best and worst forecasts.

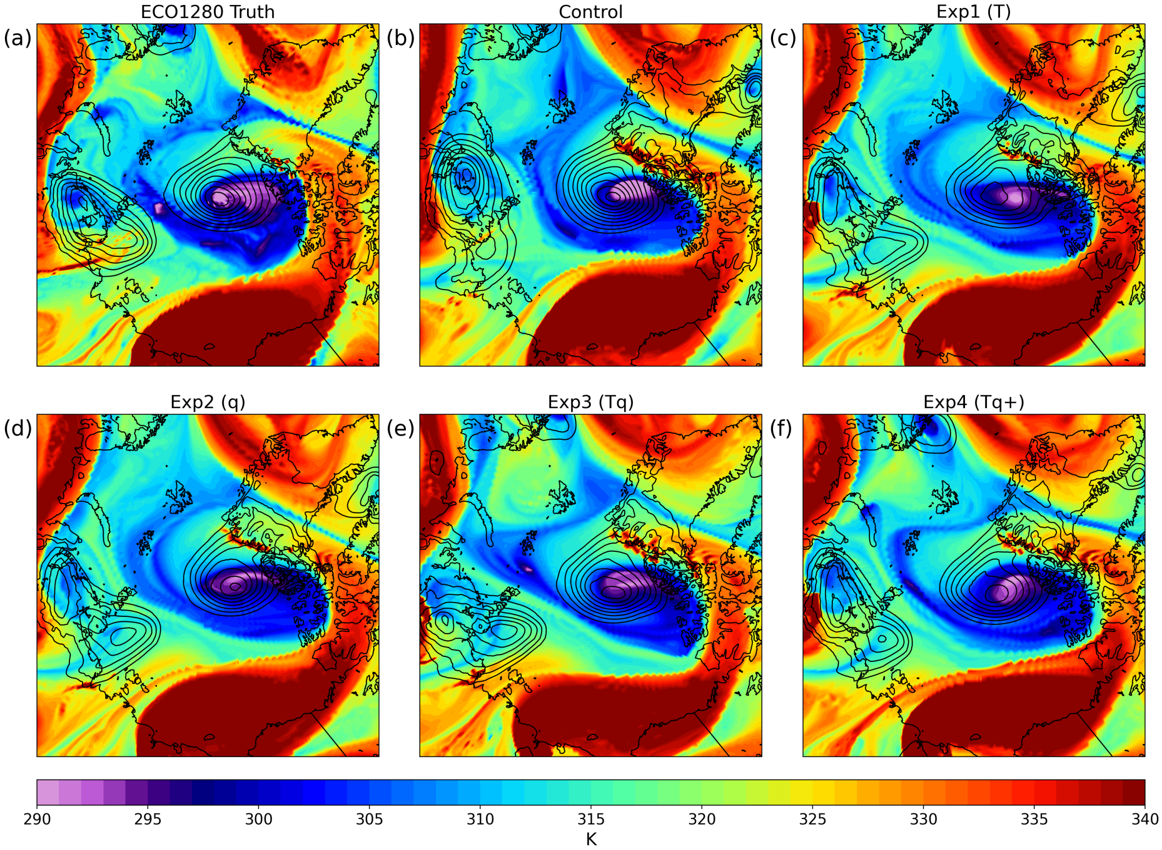


Figure 4.32: As in Figure 4.30, but for the best performing ensemble members using the standardized TPV error method.

As a whole, the members with the lowest TPV error seem to represent the breaking midlatitude ridge quite accurately, while the high error members show significant variation in ridge position. To highlight this variability, we look at spatial differences between the best and worst members for each experiment. In the control and Experiment 1, the ridge breaks further into the Arctic in the best member than in the worst member (indicated by the red areas in Figure 4.34b,c). This allows the TPVs to slide further towards the Canadian Archipelago instead of staying over the central Arctic and leads to weak, misplaced surface cyclones. On the other hand, in Experiments 2 and 3, the ridge breaks much further into the Arctic in the worst cases, matching the general trend seen in the ensemble means (Fig. 4.34d,e; blue regions). This pushes the TPV closer to Greenland than in ECO1280. Moreover, the surface cyclones in these

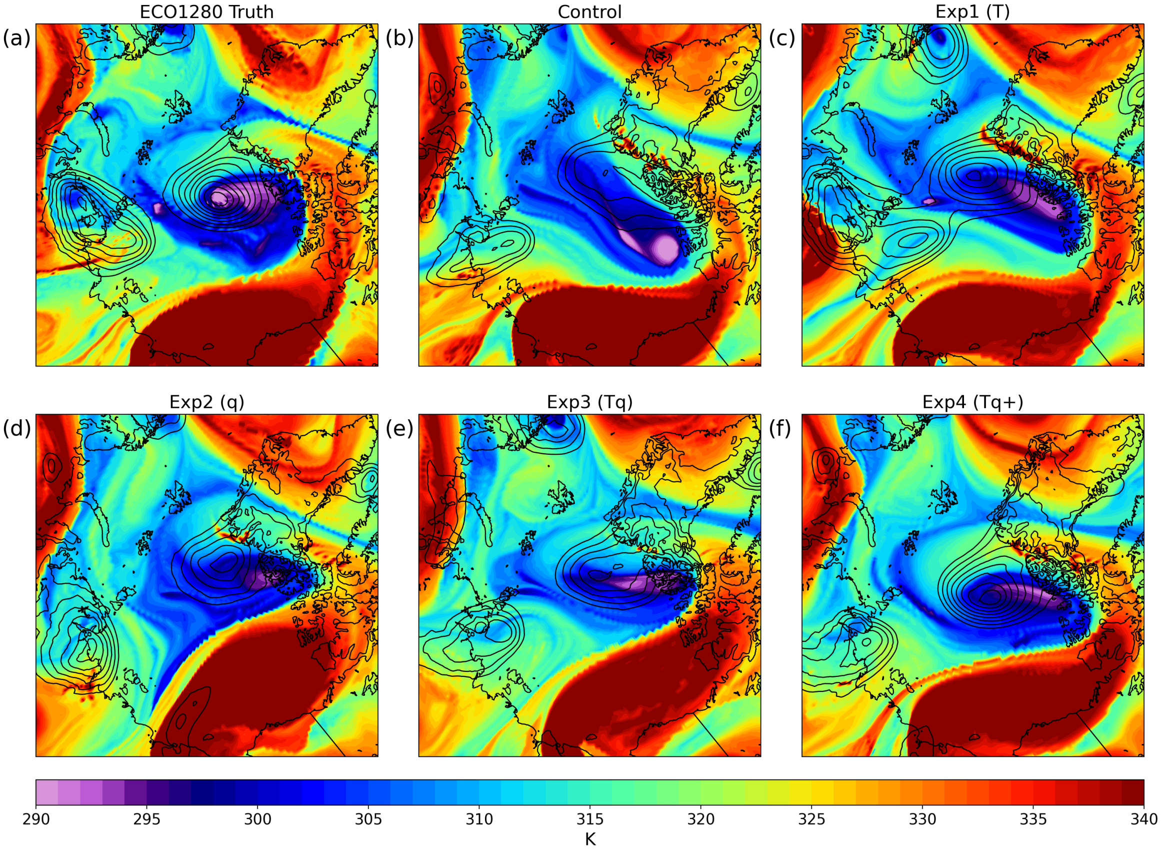


Figure 4.33: As in Figure 4.30, but for the worst performing ensemble members using the standardized TPV error method.

experiments are too weak, likely because the over-built ridge caused vertical coupling to occur to early. While these results suggest that large-scale environmental controls play a key role in driving the TPV track and concomitant AC development, the more mesoscale differences in PV structure between the best and worst members also merit further investigation.

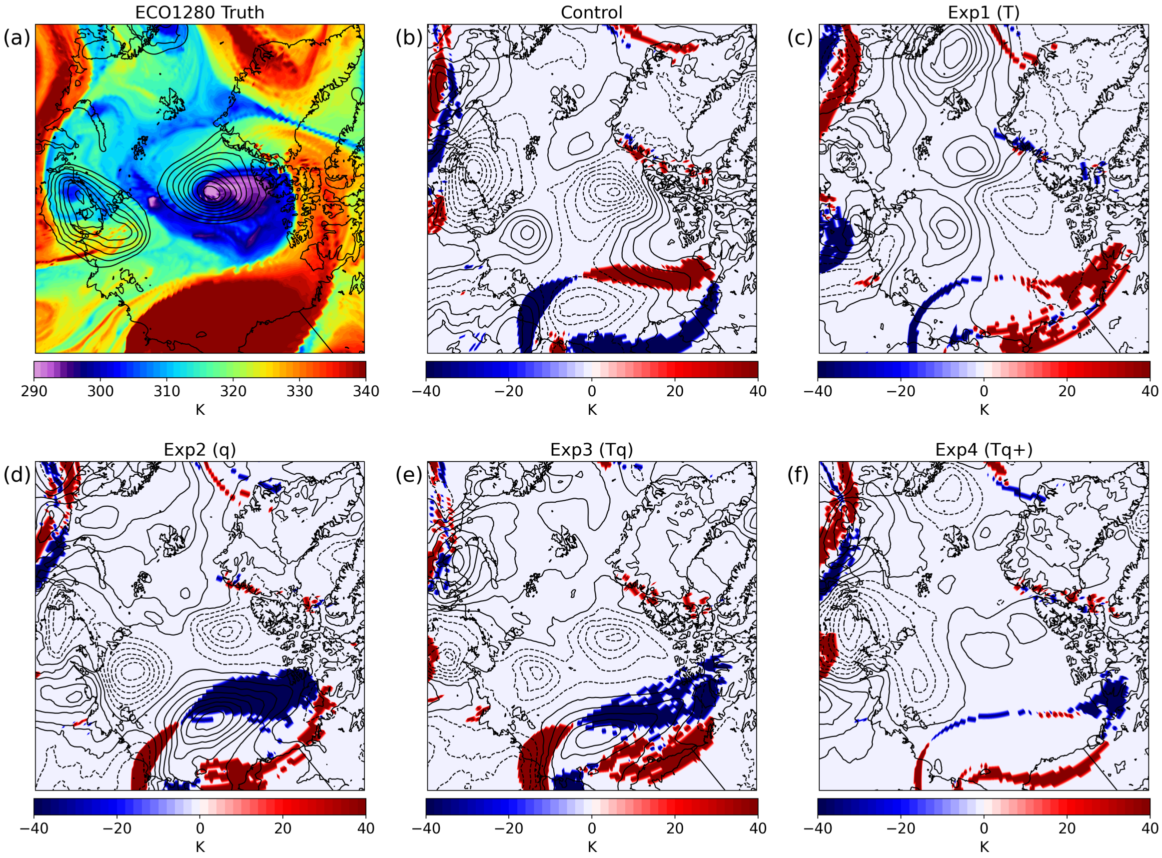


Figure 4.34: (a) ECO1280 TRPT (fill; K) and MSLP (contours; hPa; every 2.5 hPa starting at 1000 hPa) on 21 August 00 UTC. (b) Difference (best - worst) between TRPT (fill; K) and MSLP (contours; hPa; every 2.5 hPa with negative differences dashed) on 21 August 00 UTC in the best and worst individual ensemble members in the control using the standardized TPV error method. TRPT values below 340 K are set to 300 K and values above 340 K are set to 340 K in order to highlight key differences. (c) As in (b), but for the best and worst members in Exp1 (T). d) As in (b), but for Exp2 (q). (e) As in (b), but for Exp3 (Tq). (f) As in (b), but for Exp4 (Tq+).

4.3 Neighborhood Tropopause Verification

As discussed in the previous two sections, point metrics are insufficient to fully analyze variations in the TPV between experiments, and differences of tropopause potential temperature fields produce complicated spatial patterns. Therefore, it is desirable to use the threshold-based spatial neighborhood probability method described in Section 2.2 to study changes in TPV structure. From the ensemble of tropopause probability maps, we can calculate the fractions skill score (FSS), which is essentially a measure of how well the ensemble dynamic tropopause structure matches with ECO1280 in the verification region. The initial results confirm findings from the previous sections. During the cycling and early forecast periods, all dropsonde experiments have higher skill than the control (Fig. 4.35). Although Experiments 1, 3, and 4 initially lead, Experiment 2 possesses equal skill by the start of the forecast period. As the cyclone strengthens, skill scores for all ensembles begin to drop. Experiment 4 generally retains the highest skill, followed by Experiment 2, Experiments 1 and 3, and finally the control. Once again, this confirms the qualitative conclusions drawn from a visual inspection of TRPT in Section 4.1. As the cyclone decays, forecasts from the experimental ensemble members remain skillful, while the control approaches and eventually drops below the threshold for a skillful forecast.

In particular, though, we are interested in studying spatial changes in the PV field. This goal is accomplished by taking differences of ensemble probability maps from the ECO1280 probability and from other ensemble probabilities. At the beginning of the cycling period, the control ensemble is too weak and diffuse in its representation of the TPV, with lower probabilities than ECO1280 on the edges of the true vortex but higher probabilities outside of the vortex; this suggest that the TPV is not consistently represented within the ensemble (Fig. 4.36b). The temperature dropsonde experiments are a notably better match to the truth, especially in the TPV core region where

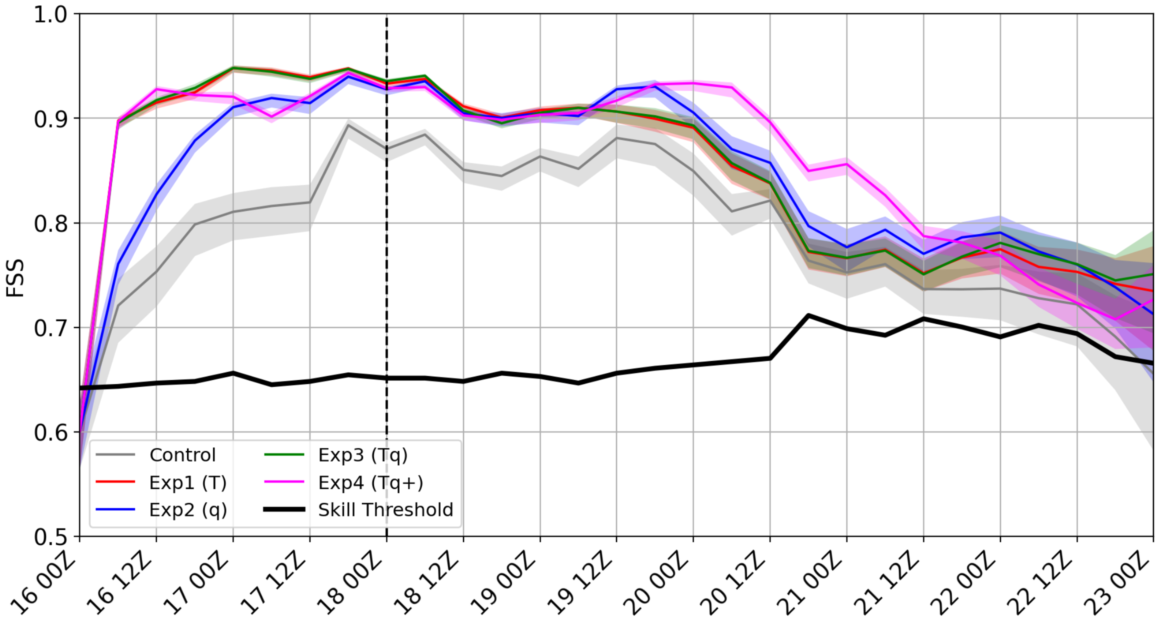


Figure 4.35: Time series of the fractions skill score, calculated using ensemble neighborhood probabilities of tropopause potential temperature as described in Section 2.2. Shading represents confidence in the skill score at the 95% level, calculated with a bootstrap resampling method. The solid black line denotes the threshold for a skillful forecast, as defined in Equation 2.4.

differences approach 0, but still contain small errors in the mesoscale features on the edges of the vortex (Fig. 4.36c,e,f). Experiment 2 lies in a middle ground between these two, with an accurately represented TPV core region, but more substantial errors on the edge of the vortex (Fig. 4.36d). In particular, all of the ensembles have lower probabilities than ECO1280 on the southeast flank of the TPV; this represents a PV filament connected to the vortex that will become important as the forecast progresses.

Comparing the experiments directly better highlights some of these differences. All of the experiments better resolve the TPV core than the control, including clear improvements along the southeast side of the vortex (Fig. 4.37b-e). Experiment 2 has lower magnitudes of error reduction than the other ensembles, but nearly identical spatial patterns, which is surprising given the somewhat indirect action of humidity observations on the PV structure. Experiments 1 and 3 exhibit nearly identical spatial

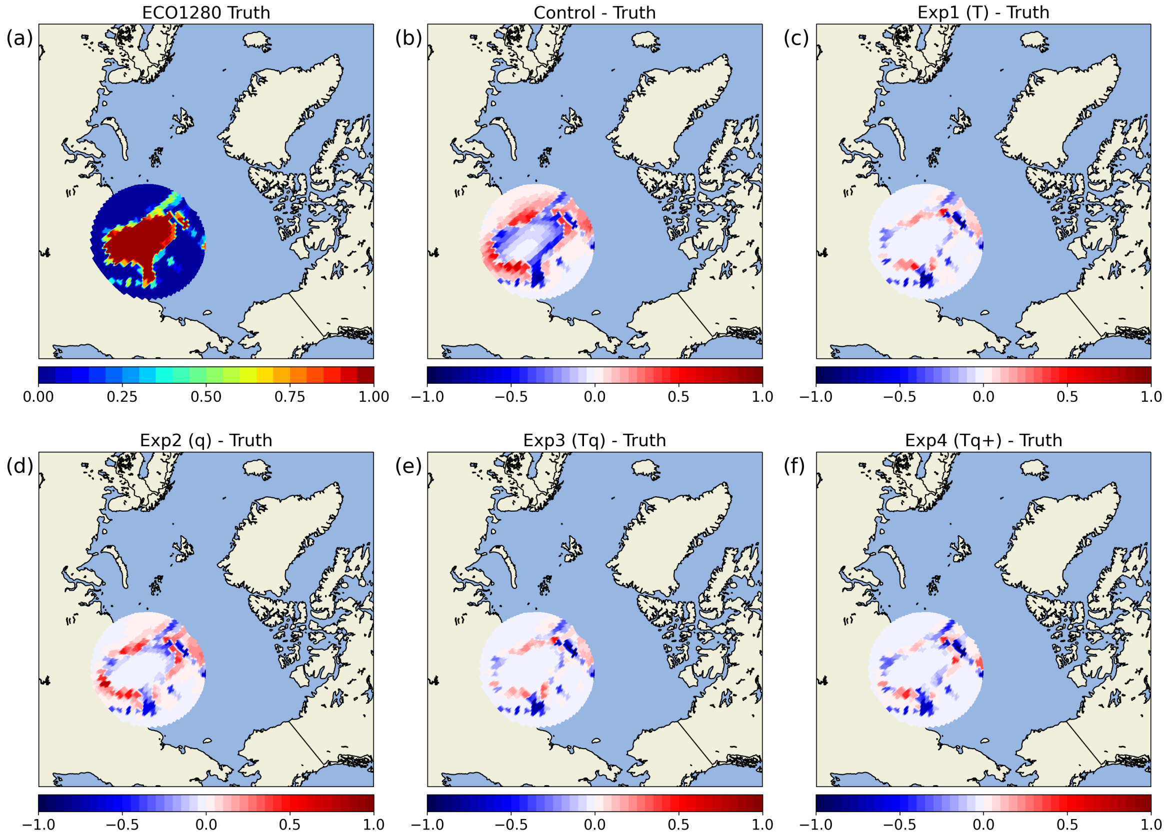


Figure 4.36: (a) ECO1280 tropopause potential temperature neighborhood probabilities on 18 August 2016 00 UTC. (b) Difference between the control ensemble neighborhood probabilities and the ECO1280 probabilities in (a). (c) As in (b), but for Exp1 (T). (d) As in (b), but for Exp2 (q). (e) As in (b), but for Exp3 (Tq). (f) As in (b), but for Exp4 (Tq+).

patterns, with only modest alterations from the humidity observations visible (Fig. 4.37g). Moreover, initial errors along the vortex edge are slightly higher in Experiment 4 than in Experiment 1 (and by extension, Experiment 3), but given the relative sparsity of dropsondes over the TPV core in the former, the difference is minimal (Fig. 4.37b-h).

As the system moves into the central Arctic, the upper level features undergo a significant reorganization (see the case study in Fig. 3.1). Although the original TPV core remains, PV filaments from the southern and western flanks of the vortex begin to wrap into the vortex and construct a new TPV core. At this point (around 19 August

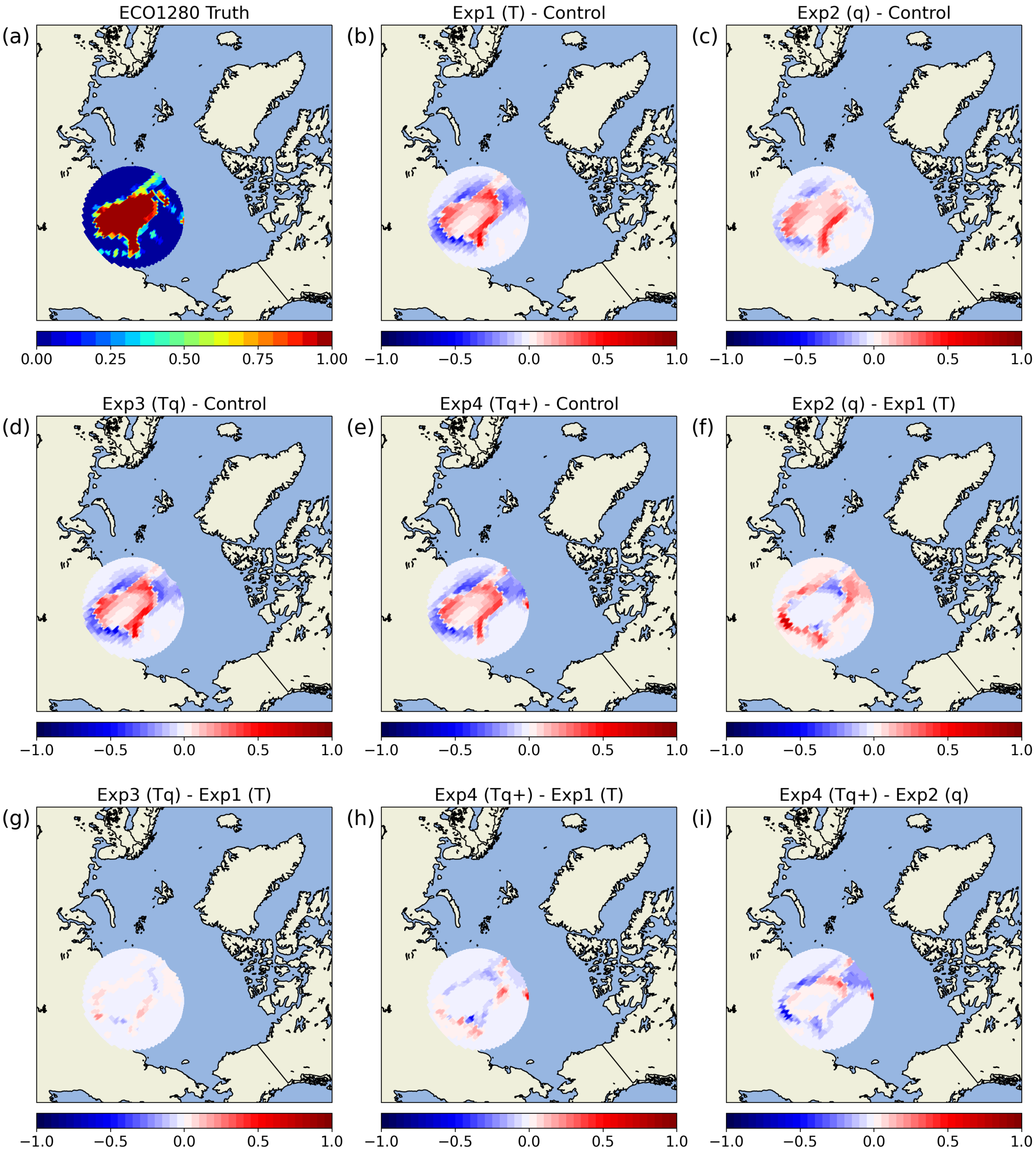


Figure 4.37: (a) ECO1280 tropopause potential temperature neighborhood probabilities on 18 August 2016 00 UTC. (b) Difference between Exp1 (T) and control ensemble neighborhood probabilities on 18 August 2016 00 UTC. (c) As in (b), but for Exp2 (q) and control. (d) As in (b), but for Exp3 (Tq) and control. (e) As in (b), but for Exp4 (Tq+) and control. (f) As in (b), but for Exp2 (q) and Exp1 (T). (g) As in (b), but for Exp3 (Tq) and Exp1 (T). (h) As in (b), but for Exp4 (Tq+) and Exp1 (T). (i) As in (b), but for Exp4 (Tq+) and Exp2 (q).

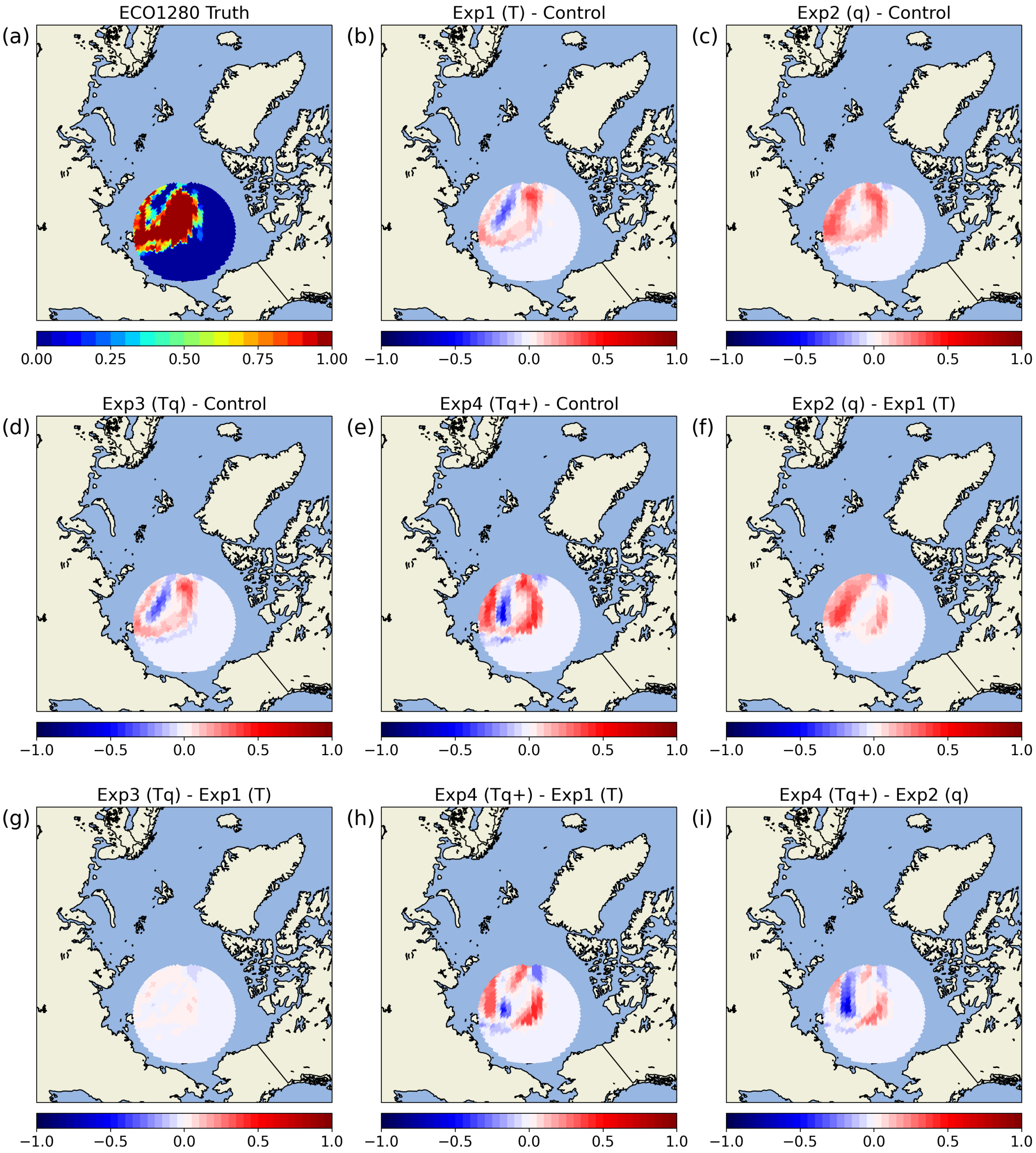


Figure 4.38: As in Figure 4.37, but for 19 August 2016 18 UTC.

18 UTC), clear differences emerge in how well the different experiments resolved these fine-scale features surrounding the TPV, including the major filament visible along the edge of the domain in Figure 4.38a. The main differences between the experiments and the control occur along the flank of the TPV where the PV filaments are joining the

main vortex (Fig. 4.38b-e). The control seems to misrepresent the detailed structure in an ensemble sense, missing the gap of low probabilities between the main vortex and the joining filament; this error is improved in all of the experiments. Compared with the other experiments, Experiment 2 does not represent the true compactness of the western filament as well, but does seem to capture the southeastern filament (on the Alaska side of the domain) more completely than Experiments 1 and 3 (Fig. 4.38f). Experiment 4 performs best once again, with more confident resolution of all of the complex structures along the vortex edge (Fig. 4.38h,i).

As the cyclone reaches its maximum strength, increased errors relative to ECO1280 have developed in all of the experimental ensembles, especially with respect to the trailing PV filament and embedded vortices (Fig. 4.39). These differences likely stem from the model resolution, which makes these very small-scale features difficult to resolve at longer lead times. The control certainly seems to fair the worst, with large errors in the position and strength of the TPV core and the surrounding environment. Experiments 1 through 3 generally reduce the magnitudes of these errors in the TPV core, but also fail to capture the main filament. Experiment 4 displays the lowest errors in the center of the TPV and reduced (though still notable) errors in along the filament. Examining the differences between the ensembles, it becomes especially clear that Experiments 1 and 3 represent improvements over the control with respect to the TPV core location and shape (Fig. 4.40b,d). The humidity observations of Experiment 2 generally produce a stronger TPV than in the control (and slightly stronger than in Experiments 1 and 3), but at the cost of shifting the vortex erroneously towards Greenland (Figs. 4.39d and 4.40c,f).

Despite three days of uncorrected free forecasts, Experiment 3 has yet to diverge significantly from Experiment 1 (Fig. 4.40g) and bears almost no resemblance to the patterns in Experiment 2 (Fig. 4.40c,d). The improvements made by the wide-ranging observations of Experiment 4 relative to the other dropsonde patterns are substantial.

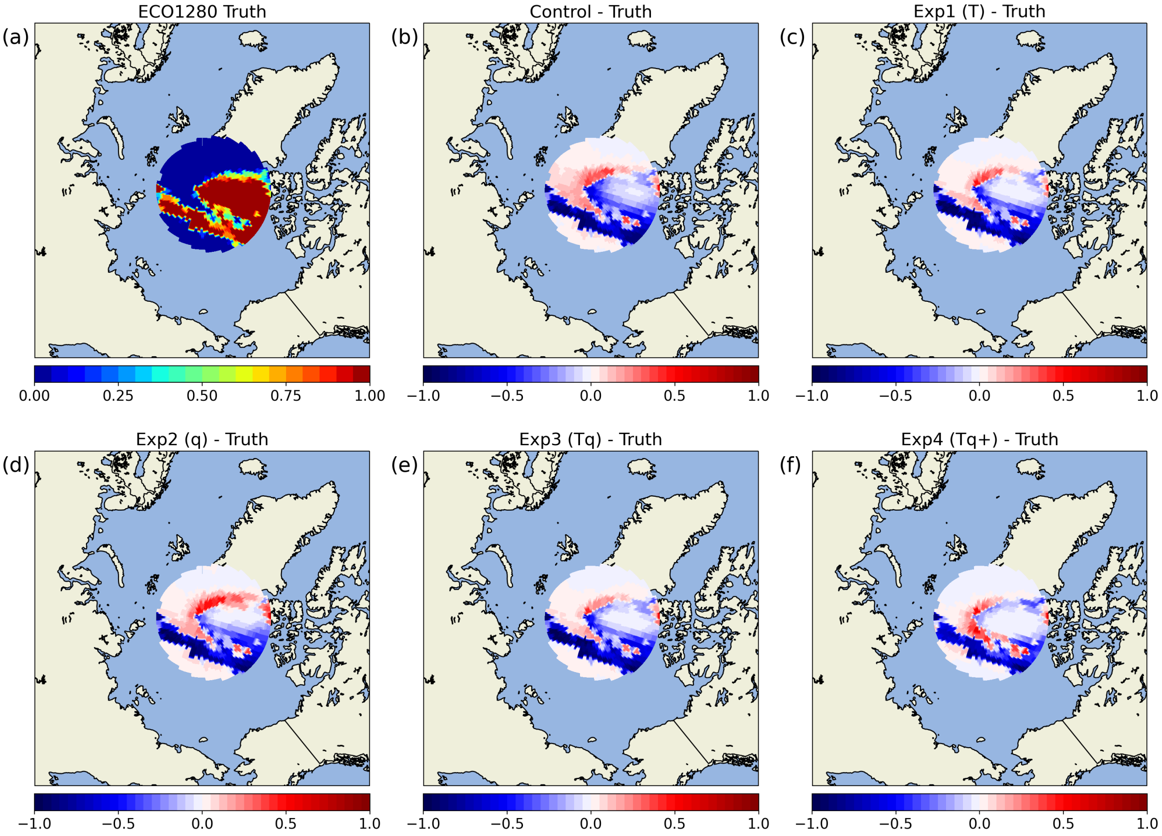


Figure 4.39: As in Figure 4.36, but for 21 August 2016 00 UTC.

The gradient on the Greenland flank of the vortex is tightened, and the core of the TPV is strengthened (Fig. 4.40e,h,i). Both of these areas originated from the merging of the original TPV core and the PV filament over Russia earlier in the free forecast period. The region in which this filament developed was only observed directly in Experiment 4. The trailing PV filament is also much more clearly resolved on 21 August in Experiment 4. This filament originated in a region away from the original TPV center, which was only observed by the more spread out dropsondes.

As the system deteriorates and begins to move over the Canadian Archipelago, the quality of the ensemble forecasts quickly decreases. As the trailing PV filament wraps into the system and the TPV couples with the surface cyclone, the ECO1280 probabilities take on more symmetric shape (Fig. 4.41a). The MPAS ensembles, on the

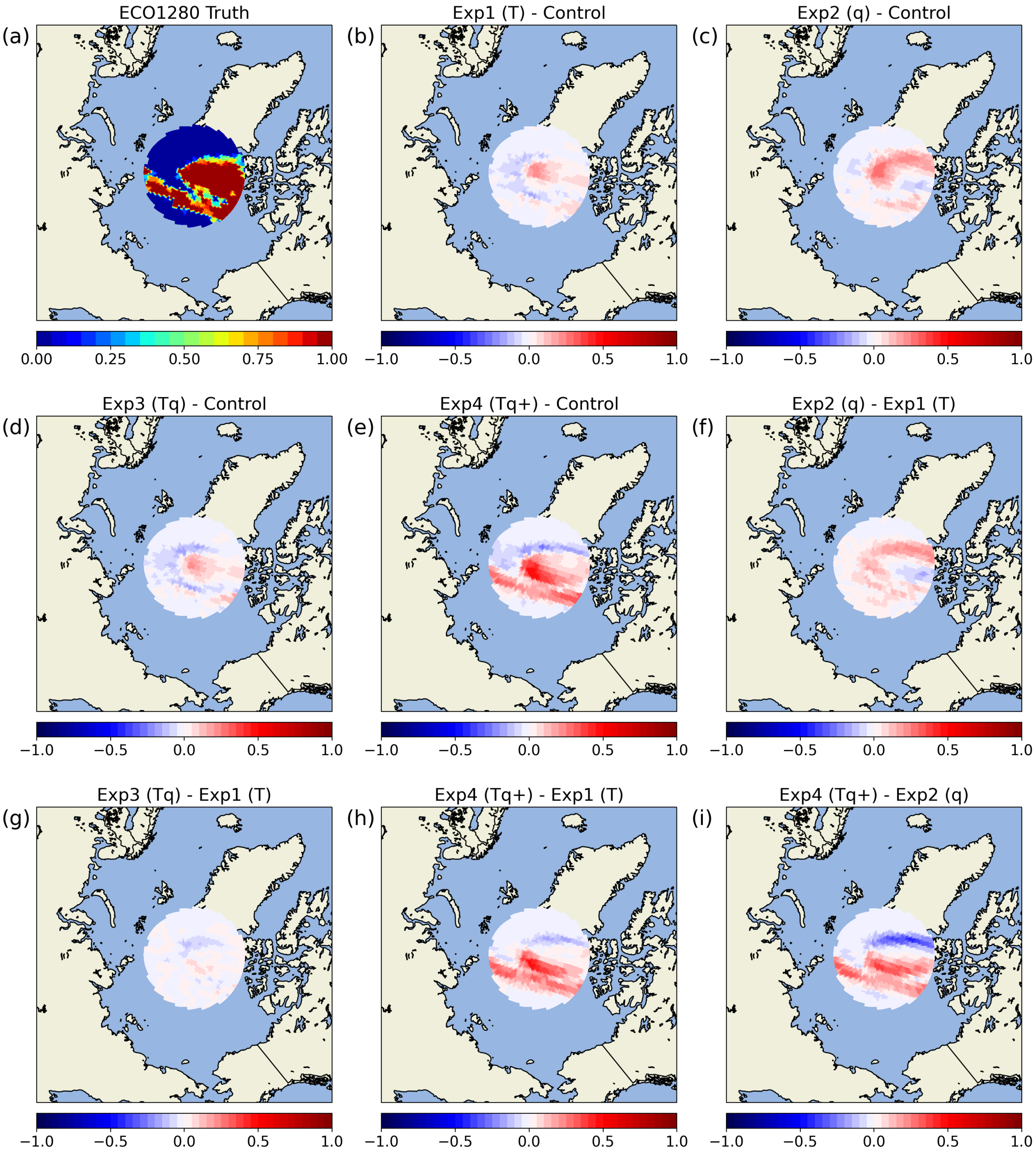


Figure 4.40: As in Figure 4.37, but for 21 August 2016 00 UTC.

other hand, retain a more elliptical shape (Fig. 4.41b-f). Although the experimental ensembles retain some additional skill relative to the control around the TPV core, they do not exhibit large improvements in the shape of the TPV. These forecast errors likely originate from a combination of errors in TPV representation and the midlatitude

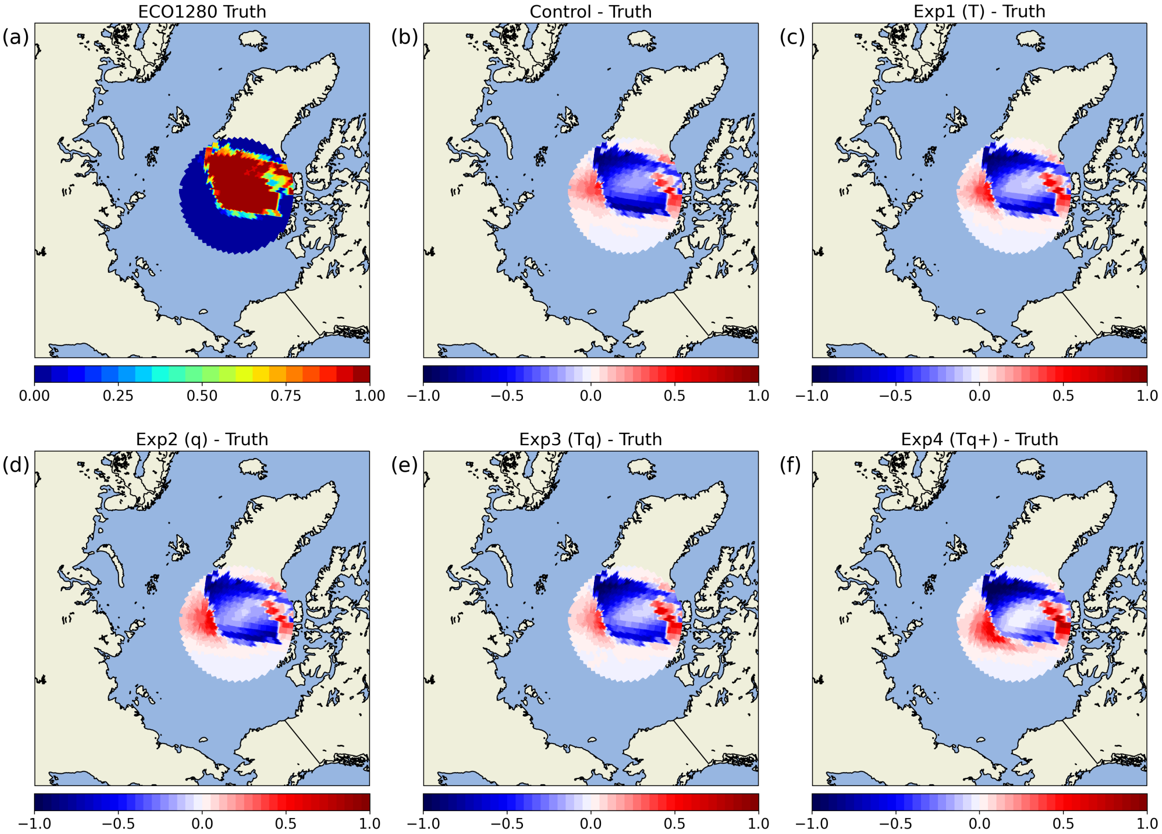


Figure 4.41: As in Figure 4.36, but for 22 August 2016 00 UTC.

ridge discussed at length in Section 4.2. The over-building of the ridge into the Arctic may have inhibited the degree to which the tropopause features could condense into a single circulation. Overall, the importance of a proper representation of mesoscale PV features seems clear. As the TPV reorganized over the Arctic, narrow PV filaments played a key role in its development, and experiments that better captured these filaments produced better forecasts. However, the inherent non-linearity of the vortex dynamics and influence of larger-scale synoptic influences makes these features difficult to forecast accurately at medium-range time scales, even with extensive corrections to the initial state. So, in addition to correcting the mesoscale vortex structure with observations in and immediately surrounding the TPV core, it is also important to sample synoptically-adjacent features that may impact the TPV's development.

Chapter 5

Summary and Conclusions

Arctic cyclones are a primary driver of high-latitude weather and sea ice changes on the synoptic scale, but large questions remain surrounding their physical drivers and medium-range predictability. ACs are known to be commonly associated with upper-level anomalies known as TPVs during their strengthening and mature stages, and in this study, we seek to investigate the connections between TPV characteristics and AC development through numerical modeling experiments. In particular, we conduct an observing system simulation experiment, in which novel observations can be simulated in any desired configuration and provided to a forecast model. In total, we conducted four primary experiments with different configurations of novel simulated dropsondes in order to test the sensitivity of a TPV-AC system to different types of corrections. In particular, we are interested in examining the importance of properly represented mesoscale tropopause PV features relative to the large-scale environment and the impact of humidity observations, which affect PV generation over time, compared to temperature observations, which more directly correct the dynamic structure of the atmosphere. We hypothesize that mesoscale corrections to the TPV structure and humidity observations will be the largest drivers of forecast improvement.

In the end, Experiment 4, which included observations of temperature and humidity spread out across the Arctic, generally produced the largest error reduction at the time of maximum cyclone strength. While experiments with many dropsondes directly over the TPV (1 and 3) corrected the initial mesoscale structure quite well, the long term impacts were more limited, suggesting that sparser observations over a large area are more beneficial than dense, local dropsondes. Experiment 2, which assimilated direct humidity observations over the TPV, produced smaller corrections at first, but eventually drew even with the other experiments during the free forecast period. The additional humidity observations significantly altered PV tendencies in the UTLS relative to the temperature-dominated experiments, which likely explains the gradual improvement of Experiment 2.

Combining the effects of temperature and humidity observations proved difficult near the tropopause, however, as the correlation-based increments provided from one observation type to the other field often opposed direct the updates. As a result, the observation error variances of each type of observation determined the direction of the increment and limited the performance of the dual observation type experiments. Although generally correcting the spatial structure, location, and track of the TPV relative to the control, the dense dropsonde observations also over-smoothed the vortex core, as evidenced by increases in TPV minimum potential temperature. At the same time, though, representation of the PV filaments surrounding the TPV and the shape of the main vortex improved in each of the temperature experiments. These mesoscale features played a key role as the TPV reorganized and strengthened later in the forecast, and so experiments that captured more initial detail near the tropopause also produced more accurate surface pressure forecasts.

Two factors that were not originally accounted for also proved to be key components of successful AC forecasts. The first is the relative distance between the centers of the TPV and the surface cyclone throughout the baroclinic development phase. On average, in the MPAS simulations, the TPV moved too directly over the AC too soon into the forecast. As a result, the optimal spacing for baroclinic development between the upper and lower level disturbances was not sustained for as long of a period as in ECO1280. Generally, ensemble members that more accurately represented the TPV-AC spacing produced deeper surface cyclones, with additional dependence on the skill of the TPV forecast.

The TPV-AC spacing, in turn, was largely governed by a Rossby wave ridge breaking into the Arctic from the northern Pacific. The wave break was driven by a midlatitude cyclone in a region that was not observed by any simulated dropsondes and acted to steer the TPV and keep it contained over the central Arctic. All of the experimental ensemble means simulated the break too far into the Arctic, meaning that the TPV was pushed deeper into the Arctic over the surface cyclone too quickly. Ensemble members that more accurately represented the location of the TPV tended to more accurately forecast the building of the ridge as well. The exact importance of the mesoscale PV structure in and around the TPV versus the influence of larger-scale features is not yet well-known, but both phenomena appear to be vital to AC development based on the results presented here. A conceptual diagram illustrating this key finding and demonstrating a proposed optimal observing pattern based on our findings is provided in Figure 5.1. In particular, we propose sparse observations of the TPV, surrounding PV features, and adjacent midlatitude features like jets and Rossby wave ridges; in this way, all regions with strong gradients of atmospheric quantities (temperature, wind, etc.) are observed, as suggested in Johnson and Wang (2021).

Although highly idealized and not necessarily representative of how real observations would impact a real-world system, this OSSE clearly demonstrates the utility of additional observations over the TPV. Given the general lack of conventional observations in the Arctic, prior studies have suggested that any additional observations (whether from satellites or new conventional observations) are likely to have a palpable positive impact on forecast skill, which was certainly demonstrated by this OSSE (Lawrence et al., 2019; Ortega et al., 2022; Jung et al., 2016). Forecast errors were significantly reduced relative to the control, and errors began to increase quickly within a day of the end of the cycling period. In particular, results from the temperature dropsonde experiments highlight the importance of detailed observations of the mesoscale

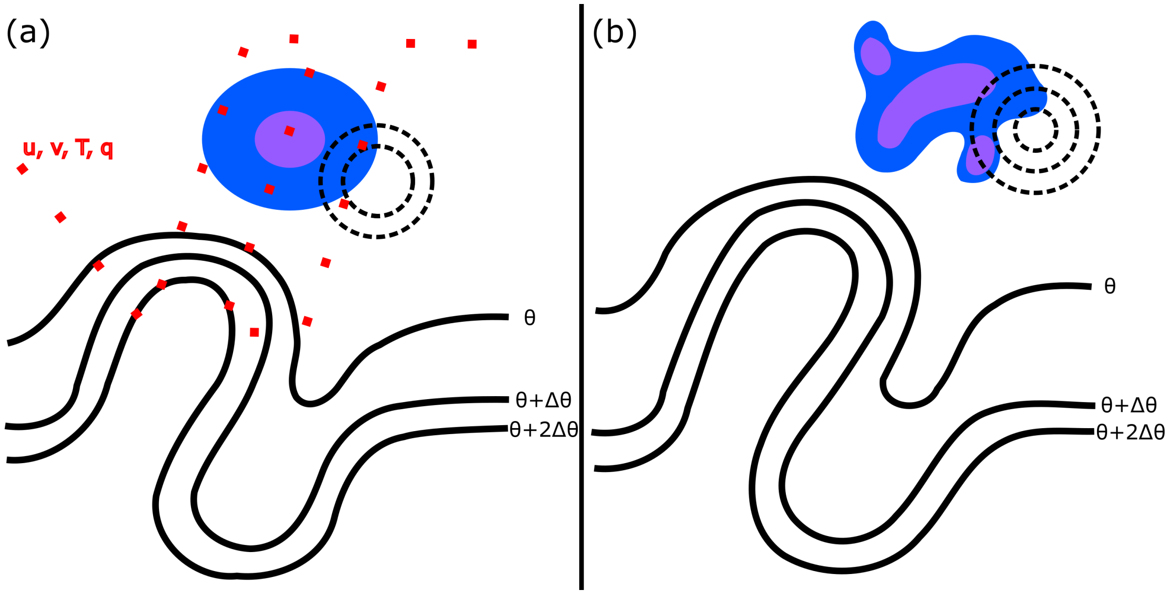


Figure 5.1: A conceptual schematic illustrating the key influences on AC track and strength. (a) An idealized AC-TPV scenario modeled prior to any additional observations being introduced. Solid black contours represent isentropes on the midlatitude dynamic tropopause, with a Rossby wave progressing into the Arctic. Blue and purple shading represents the TPV, with purple representing the deepest TPV core. The dashed black contours represent the mean sea level pressure in the Arctic (i.e., the Arctic cyclone). The red dots represent a potential observing pattern of temperature, winds, and humidity to produce the most forecast improvements base on our results. (b) As in (a), but following the assimilation of the red observations.

PV features surrounding the TPV core, which may play driving role in AC maturation. Little previous work has examined the impacts of fine-scale PV features on cyclone development specifically, but their importance was expected given the known non-linear superposition properties of PV features (Hoskins et al., 1985). These findings also parallel results from Johnson and Wang (2021) who found that upper-level temperature observations had a substantial impact on TPV-AC evolution several days into a forecast. Interestingly, though, in that study, the primary impacts were on AC tracks rather than strength, while we documented comparable effects on both track and strength.

Interestingly, the results also suggest that relatively sparse observations (though still dense enough to capture the small-scale PV features) may be preferable to avoid over-smoothing issues in the core and capture the critical features in the surrounding environment. Perhaps, then, future observation platforms like satellites, long-duration balloons, and long-distance aircraft campaigns may be most beneficial to improving forecasts. This result supports previous work on Arctic atmospheric predictability by Lawrence et al. (2019) and Køltzow et al. (2021), which found that including a large enough domain of interest and distributed observations were key to accurate Arctic forecasts. As hypothesized, humidity observations also lead to deeper surface cyclones and significantly improve forecast skill, evidently acting via physics parameterizations to strengthen the TPV. This aligns well with previous work on TPV maintenance, such as Cavallo and Hakim (2013) and directly links these PV tendency processes to surface development.

The close relationship between simulated TPV-AC distance and cyclone strength confirms the established importance of baroclinic development in Arctic cyclones (Crawford and Serreze, 2016; Tao et al., 2017a). Although the cyclone analyzed in the OSSE only exhibited an equivalent barotropic structure for around a day before moving over the Canadian Archipelago, the importance of this phase of the AC lifecycle is suggested by the experimental results. On average, the MPAS ensemble systems did not vertically align in the correct manner and decayed much more quickly than the ECO1280 cyclone. Moreover, some individual ensemble members that became vertically stacked early in the forecast period still strengthened, suggesting that physical mechanisms exist to support cyclone development even when the main TPV is centered over the cyclone, as suggested in Aizawa and Tanaka (2016) and Tao et al. (2017b).

Though not directly anticipated as a major factor when planning this OSSE, the major role played by the breaking Rossby wave follows closely from previous studies and raises key points about connections between the Arctic and midlatitudes. Rossby wave breaking (RWB) events are known to be associated with significant weather and large-scale synoptic pattern changes in the Arctic (Röthlisberger et al., 2018; Grazzini and

Vitart, 2015; Lillo and Parsons, 2017). Moreover, wave breaking events are notoriously difficult to forecast with numerical models, commonly producing medium range forecast busts (Lillo and Parsons, 2017; Gray et al., 2014; Rodwell et al., 2013). In particular, these forecasts busts are often associated with poorly observed and resolved diabatic processes in the core of the Rossby wave ridge, a potential cause of the errant forecast within our OSSE (Gray et al., 2014). Though many of these forecast impacts have been primarily documented in the midlatitudes, previous studies have shown that large-scale flow impacts tend to propagate into the Arctic (Lillo and Parsons, 2017), and Johnson and Wang (2021) showed that a nearby Rossby wave was key to the development of a TPV-AC system. In general, previous studies have shown the large-scale dynamics of the Arctic and midlatitudes to be closely connected in both directions, with midlatitude RWBs affecting the high-latitude flow and TPVs initiating new Rossby waves upon exiting the Arctic (Cohen et al., 2018; Semmler et al., 2020; Röthlisberger et al., 2018). The results presented here further cement these links and prompt new questions about the relative importance of local versus midlatitude processes in Arctic predictability.

Despite these encouraging results, there are clear avenues for future work building on this study. Fully realistic OSSEs (with more limited observation coverage and added errors) should be conducted in addition to observing system experiments (OSEs) with real data to confirm and more accurately quantify the findings. In addition, the upper atmosphere moisture bias should be considered when interpreting the results presented here, and future studies using MPAS-DART should apply corrections to remove the error. Further experiments could also be conducted using the same OSSE configuration, such as observing only the poleward breaking Rossby wave or changing the time frame in which dropsondes are simulated. Moreover, although this study briefly explored potential explanations for AC longevity, the physical mechanisms that drive AC persistence and spatial extent still need to be directly studied. This could include more idealized model experiments targeting the influence of mesoscale upper-level PV anomalies. Difficulties with assimilating temperature and moisture observations together in the vicinity of a TPV should be examined in greater detail. In particular, there is a need to understand if these issues are specific to this study or more general and whether a proper balance between prescribed observation errors variances could resolve the issue altogether. Finally, future work should attempt to more directly the characteristics of TPVs, ACs, and sea ice loss events; this could be done via fully-coupled Earth system model simulations based on this study.

References

Aizawa, T., and H. Tanaka, 2016: Axisymmetric structure of the long lasting summer arctic cyclones. *Polar Science*, **10 (3)**, 192–198, https://doi.org/10.1016/j.polar.2016.02.002.

Anderson, J., 2009: Spatially and temporally varying adaptive covariance inflation for ensemble filters. *Tellus A: Dynamic meteorology and oceanography*, **61 (1)**, 72–83, https://doi.org/10.1111/j.1600-0870.2007.00361.x.

Anderson, J. L., 2001: An ensemble adjustment kalman filter for data assimilation. *Monthly weather review*, **129 (12)**, 2884–2903, https://doi.org/10.1175/1520-0493(2001)129⟨2884:AEAKFF⟩2.0.CO;2.

Biernat, K. A., L. F. Bosart, and D. Keyser, 2021: A climatological analysis of the linkages between tropopause polar vortices, cold pools, and cold air outbreaks over the central and eastern united states. *Mon. Wea. Rev.*, **149 (1)**, 189–206, https://doi.org/10.1175/MWR-D-20-0191.1.

Blanchard-Wrigglesworth, E., M. Webster, L. Boisvert, C. Parker, and C. Horvat, 2022: Record arctic cyclone of january 2022: Characteristics, impacts, and predictability. *Journal of Geophysical Research: Atmospheres*, **127 (21)**, e2022JD037 161, https://doi.org/10.1029/2022JD037161.

Boukabara, S.-A., and Coauthors, 2018: Community global observing system simulation experiment (osse) package (cgop): perfect observations simulation validation. *Journal of Atmospheric and Oceanic Technology*, **35 (1)**, 207–226, https://doi.org/10.1175/JTECH-D-17-0077.1.

Bray, M. T., and S. M. Cavallo, 2022: Characteristics of long-track tropopause polar vortices. *Weather and Climate Dynamics*, **3 (1)**, 251–278, https://doi.org/10.5194/wcd-3-251-2022.

Cavallo, S. M., and G. J. Hakim, 2009: Potential vorticity diagnosis of a tropopause polar cyclone. *Mon. Wea. Rev.*, **137 (4)**, 1358–1371, https://doi.org/10.1175/2008MWR2670.1.

Cavallo, S. M., and G. J. Hakim, 2010: Composite structure of tropopause polar cyclones. *Mon. Wea. Rev.*, **138 (10)**, 3840–3857, https://doi.org/10.1175/2010MWR3371.1.

Cavallo, S. M., and G. J. Hakim, 2012: Radiative impact on tropopause polar vortices over the arctic. *Mon. Wea. Rev.*, **140 (5)**, 1683–1702, https://doi.org/10.1175/MWR-D-11-00182.1.

Cavallo, S. M., and G. J. Hakim, 2013: Physical mechanisms of tropopause polar vortex intensity change. *J. Atmos. Sci.*, **70 (11)**, 3359–3373, https://doi.org/10.1175/JAS-D-13-088.1.

Chen, F., and J. Dudhia, 2001: Coupling an advanced land surface–hydrology model with the penn state–ncar mm5 modeling system. part i: Model implementation and sensitivity. *Monthly weather review*, **129 (4)**, 569–585, https://doi.org/10.1175/1520-0493(2001)129⟨0569:CAALSH⟩2.0.CO;2.

Clancy, R., C. M. Bitz, E. Blanchard-Wrigglesworth, M. C. McGraw, and S. M. Cavallo, 2022: A cyclone-centered perspective on the drivers of asymmetric patterns in the atmosphere and sea ice during arctic cyclones. *Journal of Climate*, 1–47, https://doi.org/10.1175/JCLI-D-21-0093.1.

Cohen, J., K. Pfeiffer, and J. A. Francis, 2018: Warm arctic episodes linked with increased frequency of extreme winter weather in the united states. *Nature communications*, **9 (1)**, 869, https://doi.org/10.1038/s41467-018-02992-9.

Crawford, A. D., and M. C. Serreze, 2016: Does the summer arctic frontal zone influence arctic ocean cyclone activity? *Journal of Climate*, **29 (13)**, 4977–4993, https://doi.org/10.1175/JCLI-D-15-0755.1.

Curry, J. A., J. L. Schramm, W. B. Rossow, and D. Randall, 1996: Overview of arctic cloud and radiation characteristics. *Journal of Climate*, **9 (8)**, 1731–1764, https://doi.org/10.1175/1520-0442(1996)009⟨1731:OOACAR⟩2.0.CO;2.

Day, J. J., and K. I. Hodges, 2018: Growing land-sea temperature contrast and the intensification of arctic cyclones. *Geophysical Research Letters*, **45 (8)**, 3673–3681, https://doi.org/10.1029/2018GL077587.

Ebert, E. E., 2009: Neighborhood verification: A strategy for rewarding close forecasts. *Weather and Forecasting*, **24 (6)**, 1498–1510, https://doi.org/10.1175/2009WAF2222251.1.

Emanuel, K. A., and R. Rotunno, 1989: Polar lows as arctic hurricanes. *Tellus A: Dynamic Meteorology and Oceanography*, **41 (1)**, 1–17, https://doi.org/10.3402/tellusa.v41i1.11817.

English, J. M., A. C. Kren, and T. R. Peevey, 2018: Improving winter storm forecasts with observing system simulation experiments (osses). part 2: Evaluating a satellite gap with idealized and targeted dropsondes. *Earth and Space Science*, **5 (5)**, 176–196, https://doi.org/10.1002/2017EA000350-T.

Gordon, A. E., S. M. Cavallo, and A. K. Novak, 2023: Evaluating common characteristics of antarctic tropopause polar vortices. *Journal of the Atmospheric Sciences*, **80 (1)**, 337–352, https://doi.org/10.1175/JAS-D-22-0091.1.

Gray, S. L., C. Dunning, J. Methven, G. Masato, and J. M. Chagnon, 2014: Systematic model forecast error in rossby wave structure. *Geophysical Research Letters*, **41 (8)**, 2979–2987, https://doi.org/10.1002/2014GL059282.

Gray, S. L., K. I. Hodges, J. L. Vautrey, and J. Methven, 2021: The role of tropopause polar vortices in the intensification of summer arctic cyclones. *Weather and Climate Dynamics*, **2 (4)**, 1303–1324, https://doi.org/10.5194/wcd-2-1303-2021.

Grazzini, F., and F. Vitart, 2015: Atmospheric predictability and rossby wave packets. *Quarterly Journal of the Royal Meteorological Society*, **141 (692)**, 2793–2802, https://doi.org/10.1002/qj.2564.

Hakim, G. J., 2000: Climatology of coherent structures on the extratropical tropopause. *Mon. Wea. Rev.*, **128 (2)**, 385–406, https://doi.org/10.1175/1520-0493(2000)128⟨0385:COCSOT⟩2.0.CO;2.

Hakim, G. J., L. F. Bosart, and D. Keyser, 1995: The ohio valley wave-merger cyclogenesis event of 25–26 january 1978. part i: Multiscale case study. *Mon. Wea. Rev.*, **123 (9)**, 2663–2692, https://doi.org/10.1175/1520-0493(1995)123⟨2663:TOVWMC⟩2.0.CO;2.

Hakim, G. J., and A. K. Canavan, 2005: Observed cyclone-anticyclone tropopause vortex asymmetries. *J. Atmos. Sci.*, **62 (1)**, 231–239, https://doi.org/10.1175/JAS-3353.1.

Halliwell Jr, G., A. Srinivasan, V. Kourafalou, H. Yang, D. Willey, M. Le Hénaff, and R. Atlas, 2014: Rigorous evaluation of a fraternal twin ocean osse system for the open gulf of mexico. *Journal of Atmospheric and Oceanic Technology*, **31 (1)**, 105–130, https://doi.org/10.1175/JTECH-D-13-00011.1.

Hamill, T. M., F. Yang, C. Cardinali, and S. J. Majumdar, 2013: Impact of targeted winter storm reconnaissance dropwindsonde data on midlatitude numerical weather predictions. *Monthly weather review*, **141 (6)**, 2058–2065, https://doi.org/10.1175/MWR-D-12-00309.1.

Hersbach, H., and Coauthors, 2020: The era5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, **146 (730)**, 1999–2049, https://doi.org/10.1002/qj.3803.

Hoffman, R. N., and R. Atlas, 2016: Future observing system simulation experiments. *Bulletin of the American Meteorological Society*, **97 (9)**, 1601–1616, https://doi.org/10.1175/BAMS-D-15-00200.1.

Hoffman, R. N., and Coauthors, 2018: The New ECMWF Cubic Octahedral (O1280) Nature Run. *AGU Fall Meeting Abstracts*, Vol. 2018, A13C–01.

Hong, S.-Y., Y. Noh, and J. Dudhia, 2006: A new vertical diffusion package with an explicit treatment of entrainment processes. *Monthly weather review*, **134 (9)**, 2318–2341, https://doi.org/10.1175/MWR3199.1.

Hoskins, B. J., M. E. McIntyre, and A. W. Robertson, 1985: On the use and significance of isentropic potential vorticity maps. *Quart. J. Roy. Meteor. Soc.*, **111 (470)**, 877–946, https://doi.org/10.1256/smsqj.47001.

Houtekamer, P. L., and F. Zhang, 2016: Review of the ensemble kalman filter for atmospheric data assimilation. *Monthly Weather Review*, **144 (12)**, 4489–4532, https://doi.org/10.1175/MWR-D-15-0440.1.

Iacono, M. J., J. S. Delamere, E. J. Mlawer, M. W. Shephard, S. A. Clough, and W. D. Collins, 2008: Radiative forcing by long-lived greenhouse gases: Calculations with the aer radiative transfer models. *Journal of Geophysical Research: Atmospheres*, **113 (D13)**, https://doi.org/10.1029/2008JD009944.

Janić, Z. I., 2001: Nonsingular implementation of the mellor-yamada level 2.5 scheme in the ncep meso model. Tech. Rep. NCEP Office Note 437, NOAA National Centers for Environmental Prediction.

Johnson, A., and X. Wang, 2021: Observation impact study of an arctic cyclone associated with a tropopause polar vortex (tpv)-induced rossby wave initiation event. *Monthly Weather Review*, **149 (5)**, 1577–1591, https://doi.org/10.1175/MWR-D-20-0285.1.

Jung, T., and Coauthors, 2016: Advancing polar prediction capabilities on daily to seasonal time scales. *Bulletin of the American Meteorological Society*, **97 (9)**, 1631–1647, https://doi.org/10.1175/BAMS-D-14-00246.1.

Kew, S. F., M. Sprenger, and H. C. Davies, 2010: Potential vorticity anomalies of the lowermost stratosphere: A 10-yr winter climatology. *Mon. Wea. Rev.*, **138 (4)**, 1234–1249, https://doi.org/10.1175/2009MWR3193.1.

Kleist, D. T., and K. Ide, 2015: An osse-based evaluation of hybrid variational–ensemble data assimilation for the ncep gfs. part i: System description and 3d-hybrid results. *Monthly Weather Review*, **143 (2)**, 433–451, https://doi.org/10.1175/MWR-D-13-00351.1.

Køltzow, M., R. Grote, and A. Singleton, 2021: On the configuration of a regional arctic numerical weather prediction system to maximize predictive capacity. *Tellus A: Dynamic Meteorology and Oceanography*, **73 (1)**, 1–18, https://doi.org/10.1080/16000870.2021.1976093.

Kumar, A., and M. P. Hoerling, 2000: Analysis of a conceptual model of seasonal climate variability and implications for seasonal prediction. *Bulletin of the American Meteorological Society*, **81 (2)**, 255–264, https://doi.org/10.1175/1520-0477(2000)081⟨0255:AOACMO⟩2.3.CO;2.

Lawrence, H., N. Bormann, I. Sandu, J. Day, J. Farnan, and P. Bauer, 2019: Use and impact of arctic observations in the ecmwf numerical weather prediction system. *Quarterly Journal of the Royal Meteorological Society*, **145 (725)**, 3432–3454, https://doi.org/10.1002/qj.3628.

Lillo, S. P., S. M. Cavallo, D. B. Parsons, and C. Riedel, 2021: The role of a tropopause polar vortex in the generation of the january 2019 extreme arctic outbreak. *J. Atmos. Sci.*, **78 (9)**, 2801–2821, https://doi.org/10.1175/JAS-D-20-0285.1.

Lillo, S. P., and D. B. Parsons, 2017: Investigating the dynamics of error growth in ecmwf medium-range forecast busts. *Quarterly Journal of the Royal Meteorological Society*, **143 (704)**, 1211–1226, https://doi.org/doi.org/10.1002/qj.2938.

Majumdar, S. J., 2016: A review of targeted observations. *Bulletin of the American Meteorological Society*, **97 (12)**, 2287–2303, https://doi.org/10.1175/BAMS-D-14-00259.1.

Marquet, P., J.-F. Mahfouf, and D. Holdaway, 2020: Definition of the moist-air exergy norm: A comparison with existing "moist energy norms". *Monthly Weather Review*, **148 (3)**, 907–928, https://doi.org/10.1175/MWR-D-19-0081.1.

Masutani, M., and Coauthors, 2010: Observing system simulation experiments at the national centers for environmental prediction. *Journal of Geophysical Research: Atmospheres*, **115 (D7)**, https://doi.org/10.1029/2009JD012528.

Masutani, M., and Coauthors, 2013: Observing system simulation experiments; justifying new arctic observation capabilities. Tech. Rep. NCEP Office Note 473, NOAA National Centers for Environmental Prediction. https://doi.org/10.25607/OBP-815.

McCarty, W., D. Carvalho, I. Moradi, and N. C. Privé, 2021: Observing system simulation experiments investigating atmospheric motion vectors and radiances from a constellation of 4–5-μ m infrared sounders. *Journal of Atmospheric and Oceanic Technology*, **38 (2)**, 331–347, https://doi.org/10.1175/JTECH-D-20-0109.1.

Melia, N., K. Haines, and E. Hawkins, 2016: Sea ice decline and 21st century trans-arctic shipping routes. *Geophysical Research Letters*, **43 (18)**, 9720–9728, https://doi.org/10.1002/2016GL069315.

Morgan, M. C., and J. W. Nielson-Gammon, 1998: Using tropopause maps to diagnose midlatitude weather systems. *Mon. Wea. Rev.*, **126 (10)**, 2555–2579, https://doi.org/10.1175/1520-0493(1998)126⟨2555:UTMTDM⟩2.0.CO;2.

NCEP, 2008: Ncep adp global upper air and surface weather observations (prepbufr format). Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory, Boulder CO, URL https://doi.org/10.5065/Z83F-N512.

Nielsen, D. M., P. Pieper, A. Barkhordarian, P. Overduin, T. Ilyina, V. Brovkin, J. Baehr, and M. Dobrynin, 2022: Increase in arctic coastal erosion and its sensitivity to warming in the twenty-first century. *Nature Climate Change*, **12 (3)**, 263–270, https://doi.org/10.1038/s41558-022-01281-0.

Ortega, P., and Coauthors, 2022: Improving arctic weather and seasonal climate prediction: recommendations for future forecast systems evolution from the european project applicate. *Bulletin of the American Meteorological Society*, **103 (10)**, E2203–E2213, https://doi.org/10.1175/BAMS-D-22-0083.1.

Parker, C. L., P. A. Mooney, M. A. Webster, and L. N. Boisvert, 2022: The influence of recent and future climate change on spring arctic cyclones. *Nature Communications*, **13 (1)**, 6514, https://doi.org/10.1038/s41467-022-34126-7.

Pedlosky, J., 1992: *Geophysical Fluid Dynamics*, Vol. 84. 2nd ed., Springer.

Peevey, T. R., J. M. English, L. Cucurull, H. Wang, and A. C. Kren, 2018: Improving winter storm forecasts with observing system simulation experiments (osses). part i: An idealized case study of three us storms. *Monthly Weather Review*, **146 (5)**, 1341–1366, https://doi.org/10.1175/MWR-D-17-0160.1.

Portmann, R., M. Sprenger, and H. Wernli, 2021: The three-dimensional life cycles of potential vorticity cutoffs: a global and selected regional climatologies in era-interim (1979–2018). *Weather and Climate Dynamics*, **2 (2)**, 507–534, https://doi.org/10.5194/wcd-2-507-2021, URL https://wcd.copernicus.org/articles/2/507/2021/.

Privé, N. C., and R. M. Errico, 2013: The role of model and initial condition error in numerical weather forecasting investigated with an observing system simulation experiment. *Tellus A: Dynamic Meteorology and Oceanography*, **65 (1)**, 21 740, https://doi.org/10.3402/tellusa.v65i0.21740.

Privé, N. C., R. M. Errico, and W. McCarty, 2021: The importance of simulated errors in observing system simulation experiments. *Tellus A: Dynamic Meteorology and Oceanography*, **73 (1)**, 1–17, https://doi.org/10.1080/16000870.2021.1886795.

Pyle, M. E., D. Keyser, and L. F. Bosart, 2004: A diagnostic study of jet streaks: Kinematic signatures and relationship to coherent tropopause disturbances. *Mon. Wea. Rev.*, **132 (1)**, 297–319, https://doi.org/10.1175/1520-0493(2004)132⟨0297:ADSOJS⟩2.0.CO;2.

Riishojgaard, L. P., Z. Ma, M. Masutani, J. S. Woollen, G. D. Emmitt, S. A. Wood, and S. Greco, 2012: Observation system simulation experiments for a global wind observing sounder. *Geophysical research letters*, **39 (17)**, https://doi.org/doi.org/10.1029/2012GL051814.

Roberts, N. M., and H. W. Lean, 2008: Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Monthly Weather Review*, **136 (1)**, 78–97, https://doi.org/10.1175/2007MWR2123.1.

Rodwell, M. J., and Coauthors, 2013: Characteristics of occasional poor medium-range weather forecasts for europe. *Bulletin of the American Meteorological Society*, **94 (9)**, 1393–1405, https://doi.org/10.1175/BAMS-D-12-00099.1.

Röthlisberger, M., O. Martius, and H. Wernli, 2018: Northern hemisphere rossby wave initiation events on the extratropical jet—a climatological analysis. *Journal of Climate*, **31 (2)**, 743–760, https://doi.org/https://doi.org/10.1175/JCLI-D-17-0346.1.

Schreiber, E. A., and M. C. Serreze, 2020: Impacts of synoptic-scale cyclones on arctic sea-ice concentration: a systematic analysis. *Annals of Glaciology*, **61 (82)**, 139–153, https://doi.org/10.1017/aog.2020.23.

Schwartz, C. S., and Coauthors, 2010: Toward improved convection-allowing ensembles: Model physics sensitivities and optimizing probabilistic guidance with small ensemble membership. *Weather and Forecasting*, **25 (1)**, 263–280, https://doi.org/10.1175/2009WAF2222267.1.

Screen, J. A., I. Simmonds, and K. Keay, 2011: Dramatic interannual changes of perennial arctic sea ice linked to abnormal summer storm activity. *J. Geophys. Res.-Atmos.*, **116 (D15)**, https://doi.org/10.1029/2011JD015847.

Semmler, T., F. Pithan, and T. Jung, 2020: Quantifying two-way influences between the arctic and mid-latitudes through regionally increased co2 concentrations in coupled climate simulations. *Climate Dynamics*, **54 (7-8)**, 3307–3321, https://doi.org/10.1007/s00382-020-05171-z.

Sepp, M., and J. Jaagus, 2011: Changes in the activity and tracks of arctic cyclones. *Climatic Change*, **105 (3-4)**, 577–595, https://doi.org/10.1007/s10584-010-9893-7.

Simmonds, I., C. Burke, and K. Keay, 2008: Arctic climate change as manifest in cyclone behavior. *J. Climate*, **21 (22)**, 5777–5796, https://doi.org/10.1175/2008JCLI2366.1.

Simmonds, I., and K. Keay, 2009: Extraordinary september arctic sea ice reductions and their relationships with storm behavior over 1979–2008. *Geophys. Res. Lett.*, **36 (19)**, https://doi.org/10.1029/2009GL039810.

Simmonds, I., and I. Rudeva, 2012: The great arctic cyclone of august 2012. *Geophys. Res. Lett.*, **39 (23)**, https://doi.org/10.1029/2012GL054259.

Simmonds, I., and I. Rudeva, 2014: A comparison of tracking methods for extreme cyclones in the arctic basin. *Tellus*, **66 (1)**, 1–13, https://doi.org/10.3402/tellusa.v66.25252.

Skamarock, W. C., J. B. Klemp, M. G. Duda, L. D. Fowler, S.-H. Park, and T. D. Ringler, 2012: A multiscale nonhydrostatic atmospheric model using centroidal voronoi tesselations and c-grid staggering. *Monthly Weather Review*, **140 (9)**, 3090–3105, https://doi.org/10.1175/MWR-D-11-00215.1.

Sorteberg, A., and J. E. Walsh, 2008: Seasonal cyclone variability at 70 n and its impact on moisture transport into the arctic. *Tellus A: Dynamic Meteorology and Oceanography*, **60 (3)**, 570–586, https://doi.org/10.1111/j.1600-0870.2007.00314.x.

Stephenson, S. R., L. C. Smith, L. W. Brigham, and J. A. Agnew, 2013: Projected 21st-century changes to arctic marine access. *Climatic Change*, **118**, 885–899, https://doi.org/10.1007/s10584-012-0685-0,.

Szapiro, N., and S. Cavallo, 2018: Tpvtrack v1. 0: A watershed segmentation and overlap correspondence method for tracking tropopause polar vortices. *Geosci. Model Dev.*, **11 (12)**, 5173–5187, https://doi.org/10.5194/gmd-11-5173-2018.

Tanaka, H., A. Yamagami, and S. Takahashi, 2012: The structure and behavior of the arctic cyclone in summer analyzed by the jra-25/jcdas data. *Polar Science*, **6 (1)**, 55–69, https://doi.org/10.1016/j.polar.2012.03.001.

Tao, W., J. Zhang, Y. Fu, and X. Zhang, 2017a: Driving roles of tropospheric and stratospheric thermal anomalies in intensification and persistence of the arctic superstorm in 2012. *Geophysical Research Letters*, **44 (19)**, 10 017–10 025, https://doi.org/10.1002/2017GL074778.

Tao, W., J. Zhang, and X. Zhang, 2017b: The role of stratosphere vortex downward intrusion in a long-lasting late-summer arctic storm. *Quart. J. Roy. Meteor. Soc.*, **143 (705)**, 1953–1966, https://doi.org/10.1002/qj.3055.

Thompson, G., P. R. Field, R. M. Rasmussen, and W. D. Hall, 2008: Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. part ii: Implementation of a new snow parameterization. *Monthly Weather Review*, **136 (12)**, 5095–5115, https://doi.org/10.1175/2008MWR2387.1.

Thorncroft, C., B. Hoskins, and M. McIntyre, 1993: Two paradigms of baroclinic-wave life-cycle behaviour. *Quart. J. Roy. Meteor. Soc.*, **119 (509)**, 17–55, https://doi.org/10.1256/smsqj.50902.

Ullrich, P. A., C. M. Zarzycki, E. E. McClenny, M. C. Pinheiro, A. M. Stansfield, and K. A. Reed, 2021: Tempestextremes v2. 1: a community framework for feature detection, tracking, and analysis in large datasets. *Geoscientific Model Development*, **14 (8)**, 5023–5048, https://doi.org/10.5194/gmd-14-5023-2021.

Valkonen, E., J. Cassano, and E. Cassano, 2021: Arctic cyclones and their interactions with the declining sea ice: A recent climatology. *Journal of Geophysical Research: Atmospheres*, **126 (12)**, e2020JD034 366, https://doi.org/10.1029/2020JD034366.

Vessey, A. F., K. I. Hodges, L. C. Shaffrey, and J. J. Day, 2020: An inter-comparison of arctic synoptic scale storms between four global reanalysis datasets. *Climate Dynamics*, **54**, 2777–2795, https://doi.org/10.1007/s00382-020-05142-4.

Vessey, A. F., K. I. Hodges, L. C. Shaffrey, and J. J. Day, 2022: The composite development and structure of intense synoptic-scale arctic cyclones. *Weather and Climate Dynamics*, **3 (3)**, 1097–1112, https://doi.org/10.5194/wcd-3-1097-2022.

Villamil-Otero, G. A., J. Zhang, J. He, and X. Zhang, 2018: Role of extratropical cyclones in the recently observed increase in poleward moisture transport into the arctic ocean. *Advances in Atmospheric Sciences*, **35**, 85–94, https://doi.org/10.1007/s00376-017-7116-0.

Wang, Z., J. Walsh, S. Szymborski, and M. Peng, 2020: Rapid arctic sea ice loss on the synoptic time scale and related atmospheric circulation anomalies. *J. Climate*, **33 (5)**, 1597–1617, https://doi.org/10.1175/JCLI-D-19-0528.1.

Wilks, D. S., 2011: *Statistical methods in the atmospheric sciences*, Vol. 100. Academic press, https://doi.org/10.1016/C2017-0-03921-6.

Yamagami, A., M. Matsueda, and H. L. Tanaka, 2017: Extreme arctic cyclone in august 2016. *Atmospheric Science Letters*, **18 (7)**, 307–314, https://doi.org/10.1002/asl.757.

Yamagami, A., M. Matsueda, and H. L. Tanaka, 2018: Predictability of the 2012 great arctic cyclone on medium-range timescales. *Polar Science*, **15**, 13–23, https://doi.org/10.1016/j.polar.2018.01.002.

Yamazaki, A., J. Inoue, K. Dethloff, M. Maturilli, and G. König-Langlo, 2015: Impact of radiosonde observations on forecasting summertime arctic cyclone formation. *Journal of Geophysical Research: Atmospheres*, **120 (8)**, 3249–3273, https://doi.org/10.1002/2014JD022925.

Zeng, X., and Coauthors, 2020: Use of observing system simulation experiments in the united states. *Bulletin of the American Meteorological Society*, **101 (8)**, E1427–E1438, https://doi.org/10.1175/BAMS-D-19-0155.1.

Zhang, C., Y. Wang, and K. Hamilton, 2011: Improved representation of boundary layer clouds over the southeast pacific in arw-wrf using a modified tiedtke cumulus parameterization scheme. *Monthly Weather Review*, **139 (11)**, 3489–3513, https://doi.org/10.1175/MWR-D-10-05091.1.

Zhang, J., R. Lindsay, A. Schweiger, and M. Steele, 2013: The impact of an intense summer cyclone on 2012 arctic sea ice retreat. *Geophys. Res. Lett.*, **40 (4)**, 720–726, https://doi.org/10.1002/grl.50190.

Zhang, X., J. E. Walsh, J. Zhang, U. S. Bhatt, and M. Ikeda, 2004: Climatology and interannual variability of arctic cyclone activity: 1948–2002. *Journal of climate*, **17 (12)**, 2300–2317, https://doi.org/10.1175/1520-0442(2004)017⟨2300:CAIVOA⟩2.0.CO;2.

Zhang, Y.-F., C. M. Bitz, J. L. Anderson, N. Collins, J. Hendricks, T. Hoar, K. Raeder, and F. Massonnet, 2018: Insights on sea ice data assimilation from perfect model observing system simulation experiments. *Journal of Climate*, **31 (15)**, 5911–5926, https://doi.org/10.1175/JCLI-D-17-0904.1.