# Technical Memo

**ECMWF**
European Centre for Medium-Range
Weather Forecasts

# 888

# Tropical cyclone activities at ECMWF

Linus Magnusson, Sharan Majumdar*, Rebecca Emerton, David Richardson, Magdalena Alonso-Balmaseda, Calum Baugh, Peter Bechtold, Jean Bidlot, Antonino Bonanni, Massimo Bonavita, Niels Bormann, Andy Brown, Phil Browne, Hilda Carr, Mohamed Dahoui, Giovanna De Chiara, Michail Diamantakis, David Duncan, Steve English, Richard Forbes, Alan Geer, Thomas Haiden, Sean Healy, Tim Hewson, Bruce Ingleby, Martin Janousek, Christian Kuehnlein, Simon Lang, Sarah-Jane Lock, Tony McNally, Kristian Mogensen, Florian Pappenberger, Inna Polichtchouk, Fernando Prates, Christel Prudhomme, Florence Rabier, Patricia de Rosnay, Tiago Quintino, Mike Rennie, Helen Titley**, Filip Vana, Frederic Vitart, Francis Warrick, Nils Wedi, Ervin Zsoter

* University of Miami
** Met Office and University of Reading

**October 2021**

Series: ECMWF Technical Memoranda

A full list of ECMWF Publications can be found on our web site under:

*http://www.ecmwf.int/publications/*

Contact: library@ecmwf.int

# Contents

## Abstract

ECMWF has a wide range of **users** of forecasts of tropical cyclones (TCs). Several member states have territories that are frequently hit by TCs, and Météo-France is the responsible Regional Specialized Meteorological Centre (RSMC) for the southern Indian Ocean. ECMWF forecasts of TCs are also routinely used by WMO member states. Additionally, ECMWF recently opened the operational charts to the public. Many commercial customers have activities related to TCs.

Traditionally, forecasters and users have relied on global models including ECMWF to predict the **position** (or "track") of the centre of the TC out to 5 days. Nowadays, users are increasingly demanding a wider range of products related to TCs, and NWP centres are working to meet their needs. These needs include **seasonal forecasts of TC activity**, sub-seasonal forecasts of the potential for **TC genesis** (formation), accurate medium-range forecasts of the **intensity and structure** of each TC, and **downstream influences** of TCs on extratropical weather (including over Europe). The interest in forecasting **impacts** of TCs has also increased. At ECMWF, the Copernicus Emergency Management Service (CEMS) run global **flood forecasts**. Given that an important fraction of flood events in the tropics are related to TCs, accurate forecasts of the magnitude, structure, and duration of rainfall from TCs are necessary. Several external institutes also use ECMWF forecasts to model **storm surges** induced by TCs; storm surges have historically been the largest cause of death from TCs.

Within this perspective of ever-increasing user needs, this report has been developed to document progress and challenges in the ECMWF forecasting system of special relevance to TCs. These included observations (Section 2), tracking of TCs (Section 3), verification (Section 4), predictions (Section 5), data assimilation (Section 6), modelling (Section 7), forecast products (Section 8), and impact forecasting applications (Section 9). Of note, especially in Sections 6 and 7, are new results from a series of special experiments conducted over a common 37-day period in August-September 2020 that explored potential future avenues in data assimilation and modelling. Section 10.1 summarizes key progress and challenges in each of the above areas, and Section 10.2 provides a larger-scale vision for future improvement. As this report was spawned by a 1-year visit by Prof. Sharan Majumdar from University of Miami, and due to the unusually high activity in the Atlantic basin during 2020, a special attention is paid to the performances in the Atlantic. The breadth of this report has resulted in a large amount of material. To improve its readability, the report is written in a way that each section can be treated as a stand-alone article on a specific topic, summarizing progress, challenges, and future directions.

## 1.     Introduction

### 1.1.    Background

Tropical cyclones (TCs) are deadly extreme events. They often bring extreme winds as they make landfall, leading to storm surges along the coastlines together with direct damage from the winds. They can also produce extreme rainfall, which can lead to coastal and inland flooding, especially if the TC moves slowly after landfall as in the cases of Harvey (2017; Blake and Zelinsky, 2018) and Idai (2019; Emerton et al., 2020). If a TC curves into the mid-latitudes, it may undergo extratropical

transition or be absorbed by an extratropical westerlies or a low-pressure system, which can lead to windstorms over western Europe or western North America, or downstream development of extratropical events with low predictability (Keller et al., 2019).

From an ECMWF perspective, accurate predictions of TCs are important for several reasons. Several of our Member States have overseas territories that are regularly affected by TCs and/or are responsible for the forecasting and warning of TC impacts. For example, during 2017, Hurricane Irma severely affected several of our Member States' territories on the Leeward Islands. In the southern Indian Ocean, Météo-France is the responsible Regional Specialized Meteorological Centre (RSMC). Furthermore, as mentioned above, TCs that recurve into the extratropics can affect Europe. Occasionally, former TCs can impact western Europe in their post-tropical stage (e.g., Leslie 2018, Lorenzo 2019). The forecasting and timely warning of events such as these is vital.

ECMWF forecasts are also widely used outside the European continent by WMO member states, commercial actors (providing forecasts to media) and private users via the recently gained access to the products on the ECMWF website or from other web pages that show ECMWF forecasts. Around the globe, it is also important to issue early warnings for events that will cause a demand for humanitarian aid. In recent years, the concept of forecast-based financing has been introduced. For example, the Red Cross aims to trigger humanitarian action ahead of an event, based on forecast information (IFRC, 2021; Coughlan de Perez et al., 2015).

ECMWF's global reputation is partly built on the performance for high-profile TCs such as Sandy (2012) and Irma (2017). TCs generate extensive media coverage before, during, and after the event. To respect the requirement not to comment on events as they happen so as not to distract affected populations from the local authoritative voice, ECMWF abstains as much as possible from commenting before and during the event. Even without commenting or issuing any statements, ECMWF is a part of conversations on social media and the forecasts are used on most TV channels in the United States. After the event has happened, and by capitalising on the existing momentum through accepting scientific interviews, TCs have proven to be an excellent way to improve awareness of ECMWF. Such interviews allow ECMWF to speak more broadly about the model and the science that sustains it, rather than focusing solely on TCs.

The primary metric upon which ECMWF's reputation in TC forecasting has been built is the **position** (or track) of the centre of the TC. To monitor progress from year to year, ECMWF has adopted the TC position error for the 3-day operational high-resolution (HRES) forecast as one of the supplementary headline scores. That score for 3-day and 5-day forecasts, together with verification of TC **central pressure ("Pmin")** mean absolute error and mean error (or bias) for 3-day forecasts is shown in Figure 1. The corresponding scores are also shown for the ensemble mean, derived from the track of each ensemble member. To account for the interannual variability of the sample size and overall forecast difficulty, the results for ERA5 forecasts are included as well. The position error at day 3 and 5 in HRES and ensemble mean has steadily improved between 2006-2015 (Figure 1(a)). Comparing with ERA5, which is based on a non-evolving forecasting system, the improvements are incremental with contributions from many upgrades of the operational forecasting system. In 2020, the position error was highest in the time-series for ERA5 for 5-day forecasts, indicating that it was the least predictable year, and this explains the relatively high position errors in HRES. Over the

years, the day 3 position error of the ensemble mean has approached the HRES position error. In contrast, the corresponding day 5 position error of the ensemble mean has been consistently lower than HRES, as a result of the filtering during unpredictable situations by the ensemble. For the Pmin mean absolute error (Figure 1(b)) and mean error (Figure 1(c)), discrete improvements in both quantities for Pmin are evident in 2010 and 2016, when the model resolution was increased. The ensemble mean has a lower mean absolute error than HRES, although the weak bias (Pmin too high) is larger due to its reduced resolution.
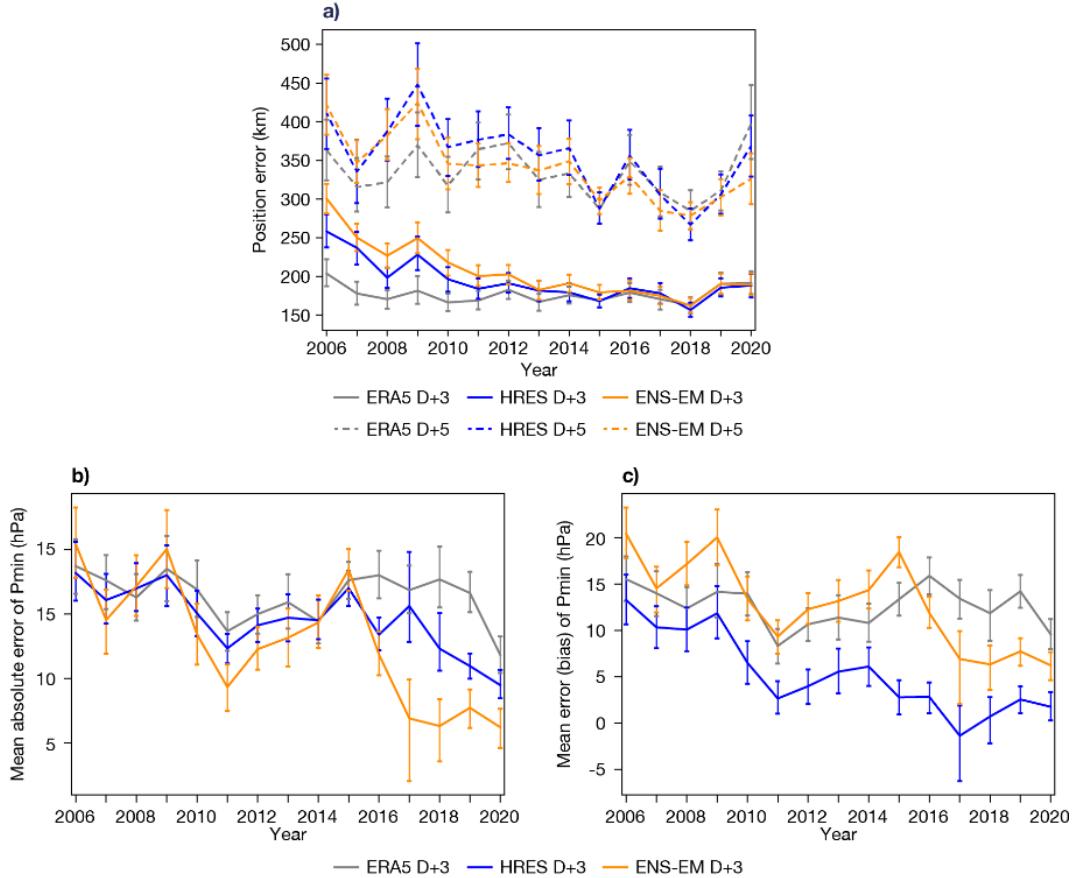


Figure 1: (a) 3-day (solid) and 5-day (dashed) forecast errors of TC position forecasts; (b) 3-day mean absolute error of TC central pressure (Pmin); (c) As for (b), for the mean error. Blue: HRES; Orange: ensemble mean; Grey: ERA5. The data is aggregated from 1 December the year before to 1 December of the labelled year.

It is a common practice for TC forecasts from different providers to be compared against each other. For example, Heming et al. (2019) summarised the verification of TC position in a report provided to the 2018 International Workshop on Tropical Cyclones. TC position forecasts from deterministic operational forecasts have been verified by the Japan Meteorological Agency (JMA) under an inter-comparison project of the Working Group on Numerical Experimentation (WGNE) since 1991 (Yamaguchi et al., 2017). The evaluation is based on meteorological fields provided to JMA who run the tracker, and an updated plot for the north-western Pacific basin is presented in Figure 2a. While there is a large year-to-year variability in the ordering of the models, ECMWF has always been

among the best if not the best. However, the spread among the group of models has decreased in recent years. For other basins, the sample is usually too small to draw conclusions for single years, but for the south-western Indian Ocean ECMWF has not been the superior model in recent years (not shown).  The NOAA National Hurricane Center (NHC) in the United States routinely verifies the models that are used in their operational activities, as shown in Figure 2(b) for 3-day position forecasts in the Atlantic basin provided by ECMWF, Met Office, and NOAA (GFS and two regional models). For 2020, the position error of the three global models was very similar. Between 2010-2020, ECMWF had the lowest errors for 7 out of 11 years, UKMO 2 years and GFS 2 years. This indicates an interannual variability in the ranking even though ECMWF has the lowest errors averaged over the full period. However, there is also a large case-to-case variability for the ranking of the models, as illustrated in e.g Tang et al. (2021). Therefore, multi-model ensembles have been proven to be superior than a single (ensemble) forecast (Titley et al., 2020).

Since the 1990s, the use of **ensemble forecasts** has increased over a range of applications, including TCs. To better understand the use of these forecasts, Titley et al. (2019) conducted a survey about the use of ensemble forecasts in TC forecasting. The study showed that nearly all operational centres use ensemble forecasts, particularly for TC track and genesis forecasting. However, the information from the dynamical ensemble is to a lesser degree communicated in official graphical products and uncertainty information. While it was noted that the quality of the track forecasts is on a sufficient level, the quality of the intensity forecasts in operational ensembles is still poor.
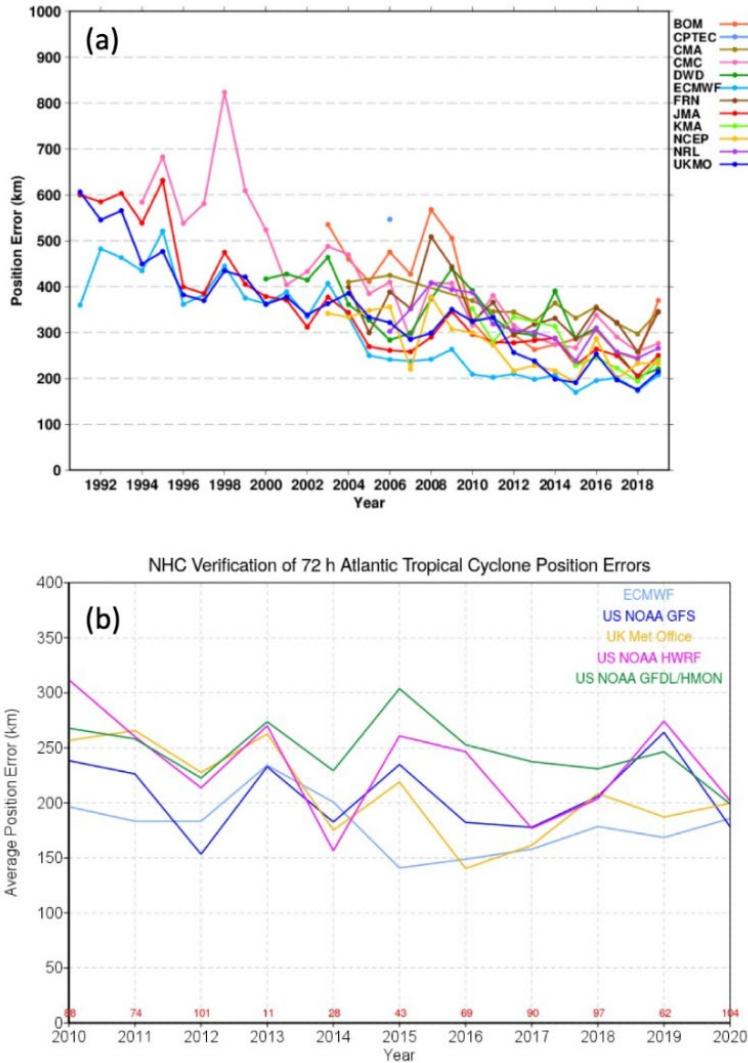
*Figure 2: 3-day position error for different NWP centres for (a) the north-western Pacific basin provided by JMA via WGNE and (b) for the Atlantic basin provided by NHC. Panel (a) is taken with permission from the presentation on 35th WGNE meeting (http://wgne.meteoinfo.ru/wp-content/uploads/2020/10/WGNE35_Ujiie_tcverif.pdf).*


## 1.2.    Purpose and Structure of this Report

The goal of this report is to provide insights into the scientific and operational status of ECMWF's modelling and predictions of TCs, at the time during which the 47r1 cycle was operational. This includes understanding ECMWF's progress in TC simulation and prediction in comparison with other centres, to help provide directions on how ECMWF can further improve its TC predictions.

This report contains ten sections, comprising all parts of the forecast production chain relevant to TCs where ECMWF is involved. Each section can be treated as a stand-alone article on a specific topic, summarizing progress, challenges, and future directions.  The report starts with background information, first outlining available **observations** relevant for assimilation and evaluation of TCs

(Section 2), then documenting how **TC tracking** is done at ECMWF (Section 3) to create operational products and perform **verification** (Section 4). The subsequent three sections describe new scientific results. Section 5 discusses various **predictability** aspects of the ECMWF forecasting system during the life cycle of the TC on different timescales. In Section 6, aspects of **assimilation of observations** are presented, including new results from a range of assimilation experiments that were conducted at ECMWF for a 37-day period in August-September 2020. In Section 7, **modelling activities** at ECMWF are described, assisted by modelling experiments conducted during the same 37-day period. The next two sections give examples of forecast and product usage and delivery. In Section 8, current and planned **forecast products** are presented, and progress and goals for applications in terms of **hazard and impact forecasts** are described in depth in Section 9. Finally, concluding remarks and priorities for future improvement are discussed in Section 10.

Throughout the report, most of the illustrated examples are provided from 2020. Due to the Atlantic basin having a record-breaking number of TCs (with the Pacific having relatively low activity), attention is paid more to Atlantic TCs. Reference materials in the Appendices include a list of abbreviations (Appendix A), the TCs during the experiment period in 2020 (Appendix B), and a summary of verification scores of wind and temperature in the tropics (Appendix C). Much of the work presented in this report was spawned by a one-year visit by Prof. Sharan Majumdar (University of Miami) between July 2020–July 2021.

## 2.     Observations of Tropical Cyclones

### 2.1.     Introduction

This section gives an overview of the different kinds of observations that provide information about the current conditions and can help to determine the future progression of TCs. The observations are important in our forecast process both for data assimilation and for post-event evaluation, although not all available observations are used for both purposes. For a more comprehensive report on the use of observations in verification, see WMO (2013). Table 1 lists the observation types that are used regularly at ECMWF for assimilation, and for diagnostics and evaluation of conditions in TCs and their vicinity. In Sections 4-5, the observation types for evaluation are presented, and in Section 6 the impact of assimilating selected observation types on subsequent TC analyses and forecasts is described.

*Table 1: Use of TC-related observations at ECMWF.*

|  |  | Assimilation | Diagnostics and Evaluation of TC |
|---|---|---|---|
| In-situ | Land/Ship/Buoy | X | X |
|  | Radiosondes | X |  |
|  | Aircraft | X |  |

| Satellite | Microwave (MW) | X | |
|---|---|---|---|
| | Infrared (IR) | X | X |
| | GPS Radio Occultation | X | |
| | Atmospheric Motion Vectors (AMV) | X | |
| | Aeolus | X | |
| | Scatterometer | X | X |
| | Synthetic Aperture Radar (SAR) | | |
| | SMOS/SMAP | | |
| Aircraft | Dropsondes | X | X |
| | In-flight data | | |
| | Doppler Radar | | |
| | Stepped Frequency Microwave Radiometer (SFMR) | | |
| Cyclone estimate | Best Track | | X |
| Sea-surface temperature | OSTIA | X | X |

## 2.2. In-situ observations

TCs are very difficult to directly observe with conventional observations, due to their extreme conditions and since they often appear in remote locations. The availability of SYNOP observations of TCs from land stations, ships, and buoys are very sparse, due to the low probability of a TC passing over a permanent location and that ships avoid the vicinity of the TC. And if a TC passes close to a station, the risk is very high for the station or buoy to be damaged.

It is even more rare that a TC can be directly observed by a regular radiosonde or by commercial aircraft. However, the assimilation of data from both platforms is hypothesized to improve the predictions of the large-scale flow that determines the propagation of the TCs. It is therefore common practice in the United States (NOAA / National Weather Service) to launch special radiosondes at 06 UTC and 18 UTC ahead of a potentially landfalling TC.

## 2.3. Temperature and humidity from satellites

Infrared observations (including near infrared and visible) are provided by sensors onboard geostationary (GEO) and low-Earth orbit (LEO) satellites and are critically important for monitoring the current state of a TC and predicting its future evolution. High resolution imagery from LEO and GEO sensors provides the most spatially detailed view of the storm, primarily showing the cloud tops, with some GEO sensors (such as the Advanced Baseline Imager on the GOES satellites) capable of refreshing this view every five minutes. The imagery is used to first identify precursor disturbances that may develop into a TC, and then the TC itself by Regional Specialized Meteorological Centres (RSMCs) and Tropical Cyclone Warning Centres (TCWCs). It is also used to estimate the central pressure and maximum wind speed via the Dvorak technique (see Section 2.7). In clear air, the infrared radiances inform on the 3D atmospheric temperature field, mid and upper-level humidity and the ocean surface temperature.

Microwave observations are relatively insensitive to attenuation from clouds and can thus provide information on atmospheric temperature, humidity, and hydrometeors at levels throughout the troposphere. The large number of microwave sensors in LEO provides high spatiotemporal coverage. Microwave observations can thus constrain the steering flow of TCs while also providing unique information below cloud tops within the TC itself. In current ECMWF operational assimilation, microwave imager and humidity sounding channels are utilised in "all-sky" conditions (i.e., including regions with clouds and precipitation); these form a crucial constraint on the analysis of columnar water vapour and hydrometeors due to sensitivity to low- and mid-level moisture. This can also benefit wind forecasts through the tracer effect of humidity (Geer et al., 2018). Temperature sounding channels (e.g., AMSU-A) are assimilated in clear-sky conditions currently, meaning that temperature profile information is available in clear and lightly cloudy areas, but not within thicker clouds such as in deep convection. All-sky AMSU-A assimilation will be included in the 47r3 IFS cycle, so that temperature profile information will be assimilated even within TCs, and the impact of this change is tested in Section 6.

GNSS radio occultation (GNSS-RO) measurements provide information about temperature and humidity in the atmosphere. In 2020, there was a large increase in the number of GNSS-RO available for operational assimilation, with the active assimilation of COSMIC-2 data from 25 March 2020, and the use of Spire data from 13 May until 30 September 2020. GNSS-RO can be considered all-sky capable. The measurements have a limb geometry, and provide high vertical resolution information, but have poorer horizontal resolution (~100's km). They are assimilated with a two-dimensional observation operator, using NWP information in an "occultation plane", to mitigate errors caused by horizontal gradients in the atmosphere. The information content of GNSS-RO measurements is largest for upper-tropospheric and stratospheric temperatures. However, the inclusion of >4000 COSMIC-2 measurements per day, in the latitude band between ±40 degrees, had a clear positive impact on tropospheric temperature, humidity and wind forecasts in the tropics (Ruston and Healy, 2021). The positive impact on tropical tropospheric humidity was difficult to demonstrate prior to COSMIC-2, but we now see a clear signal in the departure statistics of almost all observations sensitive to tropospheric humidity. Therefore, some impact of GNSS-RO on the steering flow of TCs could be anticipated. In Section 6 we evaluate the impact of from COSMIC-2 on TCs.

## 2.4.    Winds from satellites

Wind estimates can be derived from satellite observations in the free atmosphere and surface in different ways. For the free atmosphere, Atmospheric Motion Vectors (AMVs) are derived from cloud motion, or water vapour features, through a series of satellite images. In the tropics, AMVs account for most wind observations in the upper-level troposphere, particularly over the ocean. However, the impact of AMVs on TC forecasts at ECMWF is limited using 200x200 km thinning boxes (to mitigate spatially correlated errors), and uncertainties exist in attributing representative heights for AMVs. Restrictions on the assimilation of AMVs at high pressure levels further decrease the number of AMVs assimilated over TCs due to their very high cloud tops (not using water vapour winds with pressure below 200 hPa for latitudes above 25°N, or GOES winds anywhere with pressure < 150 hPa). In some cases, AMVs are not available at all as there is not enough contrast over the central dense overcast region of the TC for AMVs to be derived. These limitations affect the ability of assimilated AMVs to capture the curvature of the wind field and their overall impact. Nevertheless, previous research has shown some benefits in TC verification scores by assimilating AMVs.

Since 9 January 2020, horizontal line-of-sight winds from ESA's Earth Explorer satellite Aeolus have been operationally assimilated at ECMWF. Aeolus is the first Doppler wind lidar instrument in space. It produces two main types of wind information: Mie-cloudy measuring winds from the top of optically thick clouds, through semi-transparent cirrus and from aerosols with sufficiently large backscatter; and Rayleigh-clear winds which are based on backscatter from air molecules in clear air. Aeolus measures only a component of the vector wind along the line-of-sight of the lidar and we assimilate the horizontal line-of-sight wind component (see Figure 3 for example). The bulk of the coverage comes from Rayleigh-clear winds; however, Mie-cloudy winds are much less noisy than the Rayleigh-clear. There are some residual biases for the Mie-cloudy winds as a function of wind speed and an apparent temperature dependence to the Rayleigh-clear bias (more negative bias near surface) which could affect the impact.

Aeolus cannot measure below optically thick clouds and hence cannot see directly into a TC. It does however provide good quality Mie wind information in the outflow cirrus of TCs when a direct hit is made. Aeolus is probably more suited to capturing the winds around a TC, such as the broad circulation and vertical wind shear. One can expect these wind measurements to fill gaps in the observing system in the tropics, where vertical wind profile information is lacking, and potentially improve forecasts of tropical cyclogenesis, track, and the structure. Since Aeolus is a demonstration mission, it has not been available all the time in the latter half of 2020 and into 2021, due to various instrument test procedures. An ESA project started in July 2021 on the assessment of Aeolus winds on extreme events like TCs and extra-tropical storms.

For surface winds, there are several products available that derive winds from satellite instruments. Scatterometers measure the ocean surface roughness which is related to the local wind conditions. At ECMWF scatterometer winds have been used since 1995. Currently, the vector winds from ASCAT onboard MetOp-A/B/C and HY-2B (since June 2021) are operationally assimilated. However, the vectors have ambiguities, which are solved using information from the background winds, and this could lead to wrong selections if, for example, a TC is misplaced in the background fields (see Section 6.6). Another problem related to the use of scatterometer winds for TCs is the saturation of

the signal around 30 m/s (for the C-band instruments). Ku-band observations, like HY-2B, are less reliable at high winds and they are assimilated up to 25 m/s. Also, since Ku-band measurements are affected by the rain, the observations are filtered out around the centre of TCs.

There are also <u>other satellite channels</u> that provide wind information that are not currently used at ECMWF for wind assimilation. C-band Synthetic Aperture Radar (SAR) is the only instrument providing all-weather wind measurements at very high resolution. Thanks to its high sensitivity it can provide wind speeds up to 75 m/s in major hurricanes (Mouche et al., 2019). As for scatterometers, a wind direction ambiguity is removed by taking the closest direction to the ECMWF model winds. Due to other applications of the instrument, it is not possible to have a continuous sampling of the surface winds at such high resolution. Within the current CYMS (Cyclone Monitoring Service based on Sentinel-1) project, using TC track forecasts available from ECMWF, ESA plans the most suitable Sentinel-1 acquisition on the expected location of the hurricane's eye over the next five days. Currently the TC products from Sentinel 1 A/B and Radarsat-2 are available in near-real-time with a 3-km horizontal resolution. Thanks to the high resolution, SAR images demonstrate great potential for capturing a well-defined TC eye. A first assessment study of high-resolution Sentinel-1 data in an NWP assimilation, performed by Météo-France at the Regional Specialized Meteorological Center La Réunion (Duong et al., 2021) showed that the wind observations are biased at low (20–25°) or high (40–45°) incidence angles with sometimes large differences in maximum winds between SAR and Best Track. Their impact experiments, run with the LAM AROME OI, on a couple of cases showed that the assimilation of high-resolution Sentinel-1 wind data leads to a better TC positioning in the analysis and an improved representation of the outer vortex structure. Analysis increments were not confined to surface wind fields but are also visible in the upper layers up to 400 hPa in temperature, humidity and wind.

L-band passive sensors such as SMOS and SMAP provide all-weather wind information (speed only) near the ocean surface. At this frequency, there is no sensitivity saturating at moderately high wind speed, as occurs for scatterometer and higher frequency passive observations (De Chiara and English, 2016). SMOS wind retrievals have reduced sensitivity at low–moderate wind speeds. Cotton at al. (2018) showed that above 15 m s$^{-1}$, SMOS winds tend to be stronger than the Met Office model winds; it also reported the quality of the retrieved winds to be reduced in the presence of sea ice, radio-frequency interference (RFI) contamination and strong river plumes. The L-band passive sensors are limited by their coarse spatial resolution (40-50 km) but, in general, they do provide meaningful information on the wind radii of gale (34 kt), storm (50 kt), and hurricane (64 kt) force winds (Reul et al., 2017). These products could be considered in the future for the validation of the ECMWF tracker wind radii outputs (see Section 4). Assimilation experiments run at the Met Office showed that when the storm radius is small, SMOS is unable to resolve the eye structure present in the model. When the storm radius is much larger, SMOS can resolve the eye structure, and analysed and short-range forecast central pressures are closer to the Best Track when SMOS observations are assimilated (Cotton et al., 2018).

There are other ongoing satellite missions targeting surface wind near TCs. The Cyclone Global Navigation Satellite System (CYGNSS, Ruf et al., 2016), a constellation of eight micro-satellites launched in 2016 in LEO, is one such example. The radar receiver on each microsatellite measures GPS signals reflected from the ocean surface, and the wind speed is derived from these signals. Work

is still ongoing to determine if CYGNSS has significant value for data assimilation around TCs. Given that CYGNSS currently has a latency of around three days, the potential value of CYGNSS to ECMWF may lie in the evaluation of analyses and forecasts of surface wind fields in TCs.

## 2.5.    Sea surface temperature, ocean waves and sub-surface ocean

ECMWF currently deploys separate data assimilations for the atmosphere, for the ocean waves and for the subsurface oceans. In this section we outline the observation information that is used in these different components and that also can be applied for evaluation.

The main source of information of sea-surface temperature (SST) comes from the <u>Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA)</u> system, which generates global, daily, gap-filled foundation SST fields from satellite data and in situ observations (Good et al., 2020). Although the system operates in near-real time, OSTIA is a daily product which results in ~1-day delay between the satellite observations used in the SST and the data assimilation. However, during cloudy conditions the effective delay could be longer, As discussed in Section 6, this could have a significant impact on the SST analysis in the wake of TCs.

<u>Significant wave height</u> observations from nadir looking space-borne altimeters are used by the ocean wave data assimilation. Most altimeters have a small footprint, therefore limiting the likelihood of a direct observation of TC's waves, yet their all-weather capabilities do offer insight into waves in a TC when there is a direct overpass (Magnusson and Bidlot, 2020). Nevertheless, waves propagate away from storm centres as swell and subsequent observations are very useful in correcting swell error. Novel and future altimeters have wider swath capabilities, enhancing their potential impact. Not assimilated yet at ECMWF, ocean wave data from SAR images potentially provide some more information on the wave spectrum, but they are also limited in their coverage. This relates to the discussion about surface winds from SAR discussed above, as the satellite cannot measure both the waves and the winds at the same time.

<u>Moored buoys and platforms</u>, usually deployed in near coastal areas, provide wave information. The wave data are currently used only for model validation and development due to their geographical distribution which has been shown to limit the impact such data can have at a global scale. On occasions, TC's conditions are observed at those locations and severe cases usually lead to the loss of the assets.

<u>Drifting buoys and floats</u> can now be equipped with ocean wave sensors. There is an increasing number of such devices in all ocean basins, some already on the GTS, others in private company's hands. The full potential of these data for both model diagnostics and assimilation has not yet been explored. Some deployments ahead of TCs have shown their potential for wave forecast evaluation, such as for TC Teddy in 2020 (see Section 7).

Subsurface information about <u>ocean temperature and salinity</u> is provided by four main sources: profiling floats, moored buoys, ship based XBTS/CTDS, and marine mammals. The latter two provide little information in the vicinity of TCs given that ships will avoid such situations and mammal borne observations tend to be from seals in high latitude regions. Moored buoys provide high frequency observations of temperature and salinity with low latency given their constant surface presence. Their coverage is higher in the tropical Pacific and Indian oceans compared with the Atlantic basin. The

backbone of ocean *in situ* observations are provided by Argo profiling floats. Argo profilers typically provide one profile in a 10-day period (to save battery life and provide timeseries for climate monitoring purposes). There are more experimental observation platforms available targeting TCs such as available surface drifter observations of ocean temperature, ocean gliders, Uncrewed Surface Vehicles, and Alamo floats. These are currently retained for verification purposes. Alamo floats are Argo-like floats that can be dropped from aircraft ahead of TCs (Sanabia and Jayne, 2020). In an ongoing collaboration with the US Navy and Woods Hole Oceanographic Institution, Alamo floats are used to evaluate the response in the ocean model behind TCs.

## 2.6.    Aircraft reconnaissance and surveillance

Manned reconnaissance and surveillance flights into and around TCs are regularly performed in the Atlantic (and occasionally the Eastern North Pacific) by NOAA (Gulfstream G-IV; P-3) and the United States Air Force (USAF; C-130). The P-3 and C-130 flights provide a variety of observations to fix the location and intensity of the TC ("reconnaissance"), investigate the inner-core structure (using instrumentation such as airborne Doppler radars), and provide input to some NWP systems. The missions are undertaken with different constraints. The G-IV samples the synoptic environment of the TC from high altitude ("synoptic surveillance"), but cannot cross over the TC due to the risk of severe turbulence. On the other hand, the P-3 and C-130 are able to fly in the core of the TC but are usually flying at a much lower altitude (700 hPa) to avoid icing.

Dropsondes are usually launched during the flights. The data are provided on the GTS and are assimilated in the ECMWF forecasts. The assimilation has occasionally been problematic as discussed in Bonavita et al. (2017). One source of the problem was identified to be the horizontal drift of the sonde during the descent. Since that report, some of the dropsondes have started to include position information for each observation, and ECMWF has started to assimilate this information (see Section 6).

Recent improvements in the dropsonde technology also allow a higher temporal frequency of the dropsondes. For example, this was tested during the NASA/NOAA SHOUT campaign in 2015-16 with the unmanned Global Hawk aircraft. Another advantage with this observation platform was a very long operational range making it possible to sample systems in an early stage over the tropical Atlantic and the ability to cross the centre of the TC as it operated at very high altitude (~20 km). However, financial and logistical considerations have shelved the Global Hawk in recent years.  On the other hand, small, low-altitude Unmanned Aircraft Systems (Cione et al., 2021) are expected to be significantly more cost-effective and could be used in the future for assimilation and evaluation.

Other products from the manned NOAA and USAF aircraft include measurements of precipitation and winds from airborne Doppler radar in the core of the TC, and surface wind speeds inferred from the Stepped Frequency Microwave Radiometer (SFMR). The aircraft also records winds and temperature at the flight level, in a similar way as commercial flights. The flight-level data are operationally assimilated by NOAA (Christophersen et al., 2021). As assimilation procedure exists for commercial flight data, first exploration of the observations from reconnaissance flights are under way at ECMWF.

ECMWF ensemble data are being used in a project that aims to develop products to identify sensitive regions for TC track forecasts, which could be targeted for aircraft surveillance and additional radiosonde observations. This method was first demonstrated, after a period of testing, by NOAA NHC forecasters using ECMWF forecasts during 2017 and 2018. Led by Ryan Torn at the University of Albany, this technique can be applied to metrics such as the TC latitude or longitude, or going beyond the TC track, could potentially be applied to intensity, wind field or precipitation. It allows ECMWF forecast data to inform on optimum aircraft surveillance flight paths and/or rawinsonde drop locations, to provide additional observation data for data assimilation.

## 2.7.    "Best Track" estimates of position and intensity

At least four times a day, each of the RSMC/TCWCs produce estimates of the position and intensity for all present TCs in their basin. These observations are often referred to as Best Track, but the user needs to be aware of the ambiguity of this term (see below). In this report, we will use the term in a wide sense of the estimates from RSMC/TCWCs. Most of the information in this sub-section is based on WMO (2013).

The Best Track is a subjective human assessment of the TC centre location, intensity, and structure, using all observations available at the time of the analysis. As aircraft missions are generally only present in the Atlantic, the estimates are often only based on different satellite products. A common tool is the Dvorak technique (Dvorak, 1984) where the analyst identifies patterns in cloud features in satellite visible and enhanced IR imagery, and associates them with an intensity (T) number (Velden et al. 2006). From this, look-up tables are available to determine the minimum central pressure (Pmin) and maximum wind (Vmax). As this technique involves a human judgement, uncertainties naturally arise. To minimise this effect, the Advanced Dvorak Technique (Olander and Velden, 2007) is an automated technique to classify cloud patterns and apply the Dvorak rules.

A Best Track record typically consists of centre longitude and latitude, maximum surface wind speeds, central pressures, and may also include the radius of maximum wind, and the maximum radius to 34-, 50-, and 64-kt winds in each quadrant (often referred to as wind radii). The uncertainties in these estimates have been evaluated by Torn and Snyder (2012) and Landsea and Franklin (2013). Torn and Snyder (2012) found a strong case dependency with the position uncertainty being lower for strong TCs while the other way around for intensity (similar to the analysis errors discussed in Section 5.3). They also found a dependence on the presence of reconnaissance flights. The average uncertainties were found to be around 50 km for position, 8 kt for Vmax and 4hPa for Pmin. Landsea and Franklin (2013) finds similar results but it is worth pointing out the difference in position error for tropical storms without reconnaissance flights (63 km) and major hurricanes with flights (15 km).

Once the Best Track estimates are ready, the information is distributed on the GTS network. This product is referred to as working/operational Best Track. The data are also collected in real-time in "TCVitals" files. Note that in some basins, several institutes can issue estimates for the same TCs. After each season, the TCs are re-evaluated, and the estimates can be modified before the final Best Track is completed. The International Best Track Archive for Climate Stewardship (IBTrACS) combines track and intensity estimates from several RSMCs and other agencies to provide a central repository of both working and final Best Track (Knapp et al., 2010).

At ECMWF, the product that is internally referred to as Best Track is delivered from the Met Office and is based on real-time bulletins from RSMC/TCWCs. The current version includes information about position, intensity in terms of central pressure and category. This is used for triggering singular-vector calculations (see Section 6), product generation and for verification purposes. Due to the missing wind information, a different source for the operational verification may be considered for use in the future (see Section 4).

## 2.8. Summary and discussion

In this section, we have described various types of observations relevant to TCs, for assimilation and evaluation.

Given that TCs spend most or all their lifetimes over the ocean, there is a strong reliance on satellite data (radiances; retrieved profiles of humidity and temperature; wind retrievals). Data from new satellite platforms and instruments continue to be investigated. One example is SAR imagery, from which either surface wind or wave height data can be extracted and tested for both evaluation and assimilation. One current problem with the product is the limited availability, as the instrument needs to be dedicated to one parameter at the time.

Reconnaissance and surveillance aircraft for TCs, which normally only fly in the Atlantic basin, provide a range of observation types. Currently, ECMWF is mainly using dropsondes for assimilation and occasionally for evaluation. Recently, NOAA has implemented a data assimilation scheme in its operational limited-area Hurricane Weather Research and Forecasting (HWRF) model to assimilate aircraft data in the inner-core of TCs (Zawislak et al., 2021). For ECMWF, a feasible test would be to assimilate the flight level data, if the data can be acquired in real-time. For evaluation, a more regular use of SFMR horizontal surface wind profiles can be explored to diagnose the size of the TCs.

Operational estimates of TC position and intensity, referred to as Best Track, are regularly used for evaluation (see following sections) and is tested in Section 6 for assimilation. The estimates are based on all available observations, such as infrared and visible satellite images via the Dvorak technique, satellite surface winds and reconnaissance flights. However, the practices and availability of observations differ between different parts of the world, which leads to inconsistencies. Additionally, there are several databases that contain the data. From the ECMWF side, we need to discuss the most feasible source for the forecast production and the evaluation, to fulfil future requirements for verification.

## 3. Tracking tropical cyclones

### 3.1. Introduction

In this section, ECMWF's cyclone tracking algorithms are described. For TCs, ECMWF uses the same tracker for all timescales based on Vitart et al. (1997) and van der Grijn et al. (2005), described in Vitart et al. (2012). For extra-tropical cyclones, the tracker described in Hewson and Titley (2010) is used.  As there is an overlap between the two trackers during extra-tropical transitions, both trackers are described in Sections 3.2 and 3.3 respectively. Next, future possibilities to utilise machine learning in TC tracking are outlined in Section 3.4, followed by a short discussion in Section 3.5.

## 3.2. ECMWF operational tropical cyclone tracker

The main three steps of ECMWF's TC tracker are illustrated in Figure 3. In Step 1, the warm-core TCs are identified using 6-hourly fields of vorticity, temperature, and mean sea level pressure (MSLP) fields at a low resolution ($T_L159$, about 100 km). The use of low-resolution data is motivated by a reduction of the computing cost and the need to filter small-scale vorticity fields.
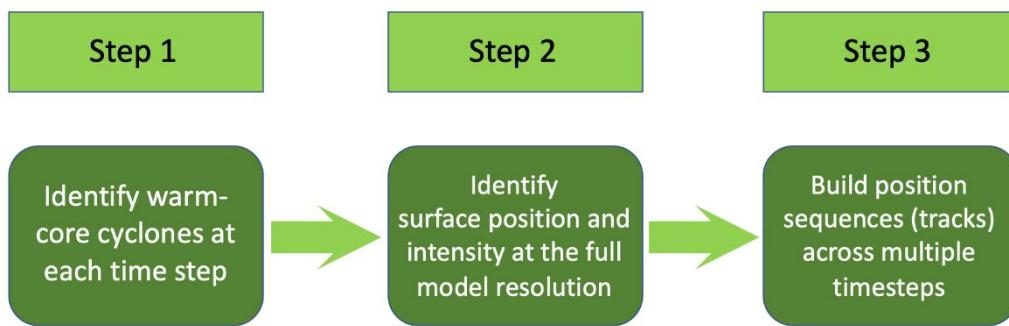


*Figure 3: The main three steps of the ECMWF TC tracker.*

Following Vitart et al. (1997), a series of criteria illustrated in Figure 4 then need to be met to satisfy Step 1. Although Criterion 3 on the maximum temperature anomaly and Criterion 4 on the maximum 200-1000 hPa thickness seem redundant as they both characterize the presence of a warm core, the use of both criteria is helpful to eliminate some extratropical cyclones that accidentally satisfy one of these criteria. To be defined as a TC, criteria 1, 2a, 3a, and 4a must be verified at all time steps, while criteria 3b and 4b need to be verified at least once in the lifetime of a TC (see step 3 in Figure 3).
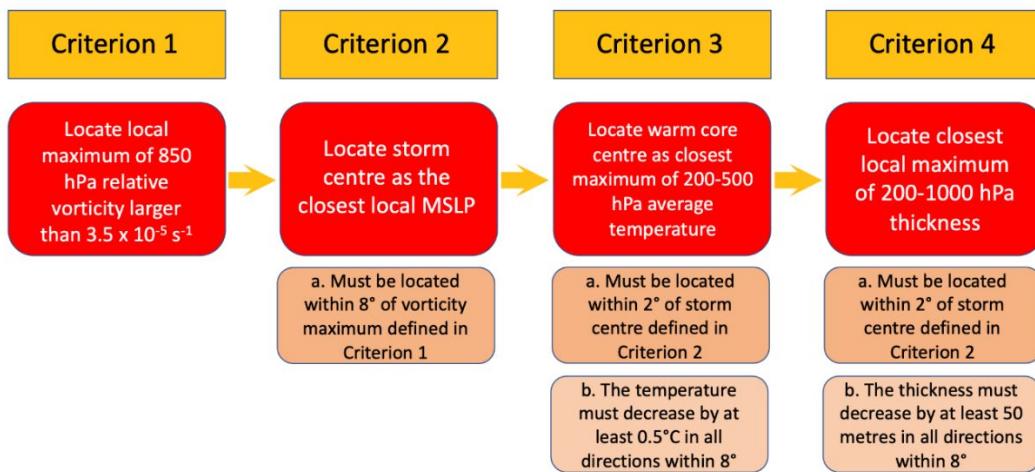


*Figure 4: The criteria that need to be met to satisfy the identification of a warm core cyclone (Step 1 in Figure 3).*

In Step 2 of Figure 3, the warm core TCs identified in Step 1 are detected using the full model resolution to retrieve the most accurate position and intensity. Since Cycle 47r1, the maximum radius to 34, 50 and 64 knots are also calculated for the four different quadrants (often referred to as wind radii), using a module provided from the Vortex Tracker package (Biswas et al., 2018) developed at the Geophysical Fluid Dynamics Laboratory (GFDL).

Step 3 of Figure 3 builds 6-hourly position sequences (tracks) from the warm core TCs identified in Steps 1 and 2. At each time step, the next position of the storm is calculated using a first guess position based on the steering wind between 200-850 hPa. The warm core TC identified in Steps 1 and 2 that is the closest to the first guess and within 350 km is considered as belonging to the same track. The difference in position between the first guess and the chosen position will be used in a correction term in the following step. A TC may be allowed to "disappear" for a maximum period of 24 hours (a TC may weaken for a short period of time when crossing an island for instance). All the criteria of Figure 4 defined in Step 1 need to be met a least once during the TC lifetime to be included in the tracker database. This means that the track contains both the early part of the life of a TC and the later parts (extra-tropical transition; after landfall). The TC also needs to be present for a minimum of two time-steps and to be located over a sea point at least once or being present in the initial conditions. Finally, the resulting track points are filtered into each of the TC basins. The tracker only searches for TCs over the basins recognised by WMO.

The final tracks are used for ECMWF forecast products and verification. They are also disseminated in BUFR format via FTP, on GTS and to the TIGGE-XML archive.

## 3.3.    Extratropical cyclone tracking in Cyclone DataBase

For extra-tropical cyclone tracking, ECMWF uses the package described in Hewson and Titley (2010) to deliver operational products. This tracker is also able to pick up tropical cyclones that approach the extra-tropics. Shorthand for this facility and its output is "CDB" - i.e. the "Cyclone Data Base". The CDB is currently developed and maintained in collaboration with the Met Office.

During identification the CDB first uses objectively-defined lower tropospheric fronts, based on wet-bulb potential temperature, and then ostensibly looks for local maxima, on those fronts, in the vorticity of the cross-front geostrophic wind, which help define front-related cyclonic features of various types. These are complemented by a set of more simply identified low pressure centres. Together these all form the full "cyclone set" for a given time, albeit with close/co-location avoided by iterative combination, using an all-feature separation matrix, of cyclone pairs that are closer together in space than a pre-defined threshold. The second step of association - i.e. joining the dots, between consecutive times, employs the "half-time tracking" technique, whereby features from timesteps n and n+1 are respectively, and nominally, moved forwards and backwards in time, by half of the timestep interval, using an upper-level steering wind, and previous movement (for continuity) if available, in order to assess the credible and the most probable feature associations. In addition to separation at half-time, the CDB also uses cyclone type change probability (related to training data and a cyclone life-cycle conceptual model), and thickness changes in the algorithm to decide which feature moves where, in an iterative process.

The CDB uses for input standard gridded pressure level data (u, v, T, q) vertically interpolated to a height 1km above the model topography, with an upscaled grid length of ~50km used to retain a synoptic scale focus befitting cyclones. Many diagnostics are computed from the input fields before graphics-based post-processing of these pinpoints the fronts and cyclonic features. This graphical post-processing approach facilitates off-grid cyclone positions. Many thresholds are inevitably employed; these were refined over 2 years of working with real time data with forecasters.

Operationally at ECMWF CDB runs for a domain focusing over the north Atlantic and Europe. The low centre identification algorithm serves to successfully identify the TCs in tropical regions, and during extra-tropical transition (ET) these features are ordinarily seen to convert into the frontal wave cyclone type, as they ingest a semi-linear (i.e. non-axisymmetric) lower tropospheric thermal gradient.

The results from CDB processing are currently used in graphical products but the tracks are not distributed in an alphanumeric format.

### 3.4.   Machine Learning approaches to tropical cyclone detection

In recent years, the problem of identifying TCs (described as Step 1 above) has become a popular application for machine learning solutions. In 2020, ECMWF has started a collaboration with NOAA and NVIDIA for exploring machine learning solutions for this problem. NOAA has initially provided a pre-developed machine learning tool (described in Kumler-Bonfanti et al., 2020) based on Convolutional Neural Networks, and within this collaboration ECMWF are optimizing selected parts of the algorithm. A similar application of machine learning was also explored in the ECMWF Summer of Weather Code 2020.

This project aims at exploring scientific challenges of TC tracking through machine learning, and more general technical aspects related to deploying a machine learning prediction system in operations (e.g., system requirements, reliability, maintainability, etc.). It aims to facilitate the integration of pre-trained machine learning models in high-performance computing environments for inference purposes, inside the NWP models and its post-processing workflows.

The algorithm under development can be used to detect the presence and location of TCs from fields of selected meteorological parameters. The network is currently trained using input fields from ERA5 reanalysis and ground-truth output fields, made by labelling the TC's areas from positional data, as provided by Best Track from the IBTrACS Database (see Section 2). During training the weights of the network are updated to minimize a pre-defined prediction error.

A Neural Network is used to scan multiple global fields and produce a single-channel output field where the TC areas are numerically labelled. This type of algorithm falls into the more general Machine Learning class of Image segmentation, where an image is labelled on a per-pixel basis according to pre-defined categories.

In this works there are many parameters to explore. The network used for this work has a U-NET architecture. The algorithm is currently trained on precipitable water content to mimic satellite images, but the variables in the training data set can be extended (e.g., to the same set as in the operational tracker described above) to include surface pressure, wind speeds, etc. The detection rate vs. computational performance can also be explored by using higher resolution in the input field such

as from ECMWF HRES analyses, or by applying cropping of the input data during the training phase. However, pairing analysis data (ERA5 or HRES) with Best Track can introduce inconsistencies if large errors are present for the TC position in the analysis or Best Track (see Section 5 for analysis errors). The network configuration itself also includes several settings to be further explored.

The future plans include to use this work as a basis to further understand processes behind tropical cyclone genesis and shortcomings in the ECMWF forecast model. This work will be undertaken in the upcoming EU- funded project H2020-CLINT.

## 3.5.    Discussion

In this section we present details about tracking algorithms related to tropical cyclones that are used at ECMWF. We also present the first steps to explore machine learning for this application.

We are aware of the existence of several other tracking software packages, both in ECMWF Member States and elsewhere. One open question here is if it is possible to obtain future synergies in the development and maintenance of the software. ECMWF aims to make the tracking code open to the community in the future.

All TC trackers possess challenges, which are partially due to the difficulty for a model to accurately represent TCs and their vital statistics, and because of the complexity of TC structures in nature. A calibrated tracking algorithm may need updating as models are upgraded, and TCs in nature may not follow straightforward physical relationships that allow for easy calibration. This is especially true during the early stage, when the centre is not well-defined, and there may be significant vertical variation in the central location. The latter also holds for strongly sheared TCs. In contrast, for mature hurricanes, the central position is expected to be tracked well. Although the central pressure (Pmin) is a stable quantity to track, the maximum wind speed (Vmax) and radius of maximum wind are difficult, since they are dominated by small scales and can shift in space and time. Another major challenge in tracking is in distinguishing between tropical, subtropical, and extratropical cyclones. Accordingly, especially for these types of storms as they enter the extratropics, forecast verification statistics may be affected depending on the type of tracker used.

Although the use of machine learning for meteorological applications is still in its infancy, a particularly promising area for use is TC tracking. This is partly due to the presence of a well-defined observation dataset (Best Track, Section 2) to be used as labels in supervised training of a model. While the results are not yet operationally viable, the technique has potential to better mimic the TC properties in the Best Track than a manually calibrated tracking algorithm.

# 4.    Verification Methods for Tropical Cyclones

## 4.1.    Introduction

In the section, we discuss verification practices for TCs that are applied at ECMWF and elsewhere and highlight ongoing developments of the verification at ECMWF. Verification methods for TCs were comprehensively outlined in WMO (2013). The methodologies discussed here will be applied in the subsequent sections about predictability and forecast experiment evaluation.

Figure 5 shows the surface wind structure in Hurricane Laura (2020). When it comes to quantifying their predictive skill, TCs possess an advantage over other weather systems since they are self-contained features with distinct, measurable characteristics, a few of which are tabulated in the Best Track estimates. Such characteristics (highlighted in bold) include the mere **existence of a TC** (to measure TC activity), the latitudinal and longitudinal **position of the centre** of the TC (often referred to as the "track"), and various measures of the TC surface wind structure. The most common measure of the structure is the "intensity", which is usually represented by either the **minimum central pressure** ("Pmin") or the **maximum sustained surface wind speed** anywhere within the TC ("Vmax"). The distance between the points for Pmin and Vmax is referred to as the **radius of maximum wind.** All these metrics are routinely produced in the ECMWF TC tracker.
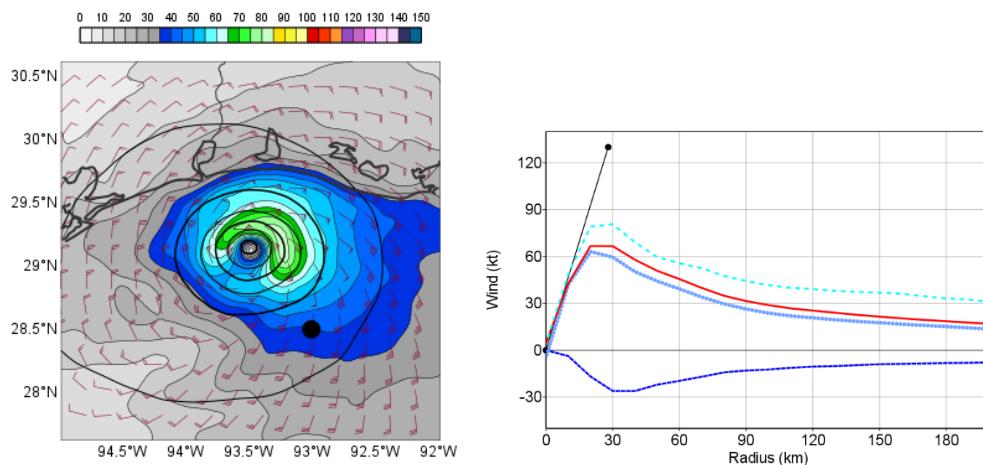


*Figure 5: Visualisation of the surface wind structure in Hurricane Laura, for a 60 h HRES forecast initialized on 1200 UTC 24 August 2020. Left: MSLP (black contours), 10-metre winds (wind barbs, in kt), 10-metre wind speed (shading, in kt). Black dot: NHC Best Track position. Right: Radial mean of total (red), tangential (light blue, dotted) and radial (blue) 10-metre wind speed, and maximum wind speed at each radius (cyan, dashed). Black dot: Best Track value of Vmax and the radius to Vmax.*

Since 2020, the **radial extent of specified surface wind speeds** (34 kt or 17 m/s, 50 kt or 26 m/s, 64 kt or 33 m/s) in the northeast, southeast, southwest, and northwest quadrants have also been produced in the tracker output and an example is given in Figure 6 for a 60 h forecast of TC Laura. Examples of strong hurricanes (e.g., HRES) to less intense hurricanes with an asymmetric wind distribution are illustrated. The tropical storm force winds in each of the illustrated forecast members extend further to the east than to the west, raising the possibility of broader wind damage and storm surges on the eastern side of the hurricane. The HRES and one of the ensemble members produced accurate 60 h forecasts of the position and wind distribution compared with the NHC Best Track (Figure 6, right panel), whereas other members were weaker or even not tracked.
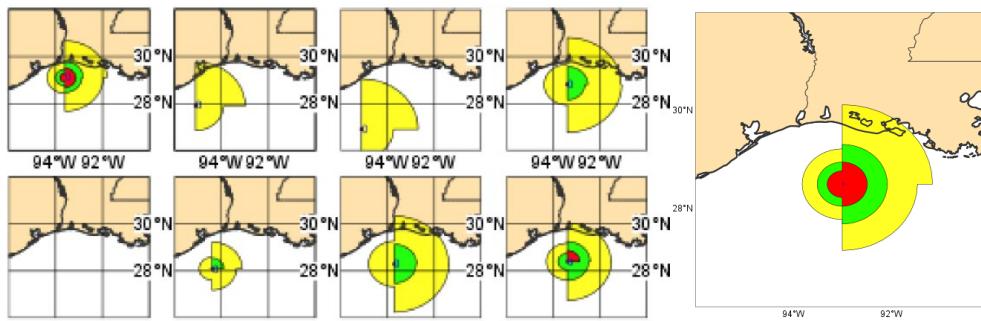
*Figure 6: HRES (top left), 47r1-ENS-CF (second from left), and 6 of the 50 ensemble members' predictions of the radii of (yellow) 34 kt, (green) 50 kt, and (red) 64 kt surface winds, for the 60 h forecast of Hurricane Laura initialized at 24 August 12UTC 2020. Right: NHC Best Track of position and surface wind radii at 27 August 00UTC 2020.*

To characterise the movement of the TC, the **propagation speed** can be calculated using the current and previous position (usually 6 hours earlier). Forecast track errors can also be decomposed into an **along-track part and a cross-track part** (see Figure 3 in WMO13). Other relevant aspects of the position are to correctly predict **genesis** (both location and timing) and to capture risk for **landfall**.

The key verification measures of TCs include both the predictive skill (via measures such as the Mean Absolute Error or MAE; Wilks, 2006) and bias (via the Mean Error or ME). For many TC characteristics, a probabilistic framework for prediction and verification is desirable.

The predictive skill of these TC metrics, and potentially others, depends not only on the intrinsic predictability of the Earth system, but the ability of the full NWP system to accurately simulate the processes on multiple scales. Traditionally, the focus at ECMWF (and other global NWP centres) has been on track prediction, which is largely (but not entirely) driven by synoptic-scale processes, which are well represented in the ECMWF data assimilation and modelling system. More recently, with the model grid spacing improving to less than 10 km, the model can capture mesoscale features such as the TC eyewall and rainbands. Accordingly, attention is now also being focused on structural characteristics, especially intensity. The intensity is defined in two ways by the TC community: Vmax is used by the NHC and other agencies, while Pmin is commonly used by Met Office and ECMWF, for example. It is generally viewed that Pmin is a more stable and robust metric by which to evaluate TC intensity in global models (Davis, 2018). The Vmax metric, while more volatile than Pmin in its predictability and more difficult to observe, was designed half a century ago with the worst-case scenarios for housing damage in mind. Vmax is largely dependent on convective-scale (O(1 km)) processes that are not resolved in the global models, and it is therefore expected to be regularly underestimated. Another difficulty with wind speed is that different reporting centres use different criteria to estimate the maximum wind speed. While the WMO recommends a 10-minute averaging period, in some centres the Vmax is measured over a 1-minute interval. This inconsistency is not taken into consideration in the verification of Vmax or wind radii in this report. The inclusion of both Pmin and Vmax in Best Track data allows for the evaluation of numerical predictions of both metrics, and they are also considered by other centres for use in assimilation and initialization. In this paper, "intensity" can refer to both Pmin and Vmax.

## 4.2. Operational verification and recent developments at ECMWF

Based on the verification against the Best Track dataset described in Section 2, statistics on position and Pmin errors are routinely produced at ECMWF. It is common practice to only include TCs that are present at the initial time of the forecast, which excludes forecasts of genesis. When comparing two (or more) models, homogeneous samples are based on the same pairs of forecast events shared by the different models.

As discussed in the Introduction, ECMWF has adopted the 3-day HRES forecast position error as one of its headline scores (see Figure 1). That score is included in ECMWF annual verification report (e.g., Haiden et al., 2021). The report also contains the RMSE of position errors at day 5, mean error and mean errors (bias) of propagation speed and intensity (in terms of central pressure) of HRES and the ENS control forecast. The metrics are also compared to ERA5 forecasts as a reference. For the ensemble, results of RMSE of the position error for ensemble mean and the ensemble standard deviation, as well as probabilistic verification of ensemble track forecasts are shown. Finally, the report includes results from basin-wide TC activity in the seasonal forecast.

Recently, ECMWF began an inter-comparison of the TC forecast performance between ECMWF and other centres. The track forecasts used are from the TIGGE-TC exchange, which provides tracks for both deterministic and ensemble models. All TCs that exist at the time of the forecast initialisation are included.
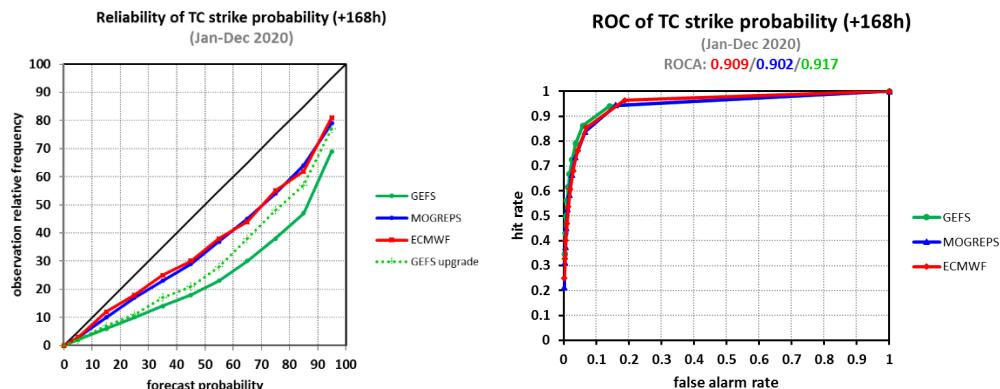


*Figure 7: Probabilistic verification of ensemble TC forecasts at day 7 for January-December 2020; ENS (red), MOGREPS-G (blue) and GEFS (green). The dashed line (green) includes only the upgraded GEFS from 23 September 2020. Left: reliability diagram; Right: standard ROC diagram.*

Evaluations of the 7-day strike probabilities are presented in Figure 7 for the ECMWF ENS, Met Office MOGREPS-G and NCEP GEFS ensemble forecast systems. This product is defined as the probability that the central position of an existing TC will pass within 120 km during the next 168 hours. All models are found to be overconfident, as shown by the curves below the diagonal line (0,0) to (1,1) in the reliability diagram (left panel) with the GEFS showing the lowest reliability. A significant model upgrade was announced by NCEP on 23 September 2020; the number of perturbed members increased from 21 to 31 and the resolution increased from 33 to 25 km. Preliminary results indicate an improvement of the NCEP GEFS reliability curve based on a smaller sample (from end of

September until December 2020). Using the same ensemble models, Titley et al. (2019) obtained better reliability curves based on a two-year period. Their results considered TCs with Best Track intensities greater than 34 knots (thus excluding tropical depressions) whereas the strike probabilities included times up to the last time that a matching observation of tropical storm intensity or higher was available, thus reducing the false alarms.

In contrast to the reliability diagrams, the Relative Operating characteristic (ROC) curves are almost identical between the three ensembles despite the differences of hit rates for larger probability thresholds for very small false alarm rates. GEFS has the best (largest) area under ROC curve amongst the three ensemble systems (0.917 against 0.909 and 0.902 of ENS and MOGREPS-G respectively).

When comparing the performance of TC forecasts from different centres, one faces several challenges. One is whether to use tracks calculated at each centre or to obtain the meteorological fields and apply the same tracker. The advantage of the former is that it reduces the amount of data to transfer but has the disadvantage of the properties of the different trackers affecting the forecasts (see Horn et al., 2014 for comparison of different tracking schemes). However, if the aim is to score the final product this is not a problem. One difficulty to obtain the meteorological fields is to make sure that they are on a resolution as close as possible to the native model resolution. In the TIGGE archive for example, the resolution is reduced for most centres, which affect the intensity of the TCs. To simplify the exchange of TC track data, there is a dedicated TIGGE archive (TIGGE-TC) for ensemble forecasts of TC tracks (NCAR, 2021), based on the trackers from each centre.

ECMWF is also exploring intensity verification based on wind speed, using Best Track files from other sources than those routinely used at the Centre (see Section 2). ECMWF occasionally produces statistics of Vmax verification, wind-pressure relations (see Section 5) and lately verification of wind radii as illustrated in Figure 8.
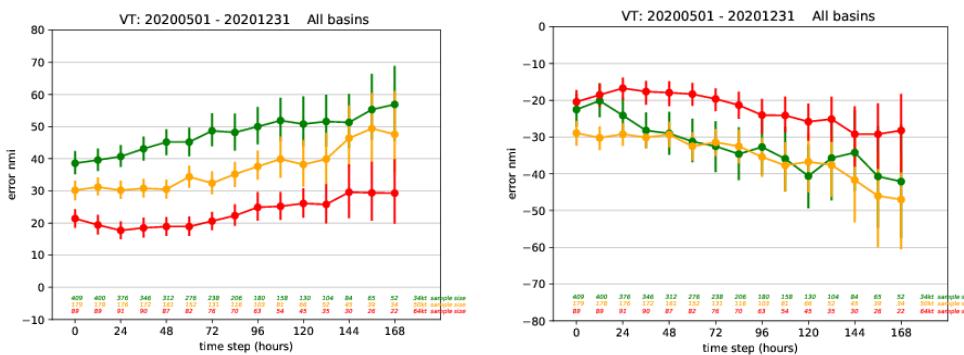


*Figure 8: Mean absolute error (left) and mean error (right) for wind radii of 34- (green), 50- (yellow), and 64-knot (red) thresholds, for HRES operational forecasts between 1 May – 31 December 2020. Units in nautical miles (1 n mi = 1.852 km). Vertical bars represent 95% confidence intervals.*

The verification of wind radii forecasts is difficult due to the scarcity of surface wind observations, which are critical for obtaining accurate wind structures of storms (Cangialosi and Landsea, 2016).

Because of this, TC forecast centres are still reluctant to publicly release verification on a regular basis. Figure 8 illustrates an example of wind radii verification following the implementation in operations of wind radii product in July 2020. A systematic negative bias is seen for the 34-, 50- and 64-knot thresholds up to 168 h. Overall, the HRES under-estimates the TC size in terms of the wind speed. This could partly be related to an underestimation of the wind-pressure relation discussed in Section 5. While the 34-knot wind radii observations are reported in all basins (south and north hemispheres), the 50- and 64-knot wind radii are only limited to Eastern Pacific and North Atlantic basins. For each wind speed threshold, the mean absolute radii error increases with the forecast lead time. Errors are slightly larger in the ENS control forecast than the HRES suggesting that model resolution has an impact on the TC size (not shown). The amplitude of the error of the 64-knot wind threshold is smaller comparatively with the 50- and 34-knot thresholds. This is expected as 34-knot wind radii is remarkably higher than the other two thresholds (Cangialosi and Landsea, 2016). A future development of this verification would therefore be to scale the error with the radius.

## 4.3.     Future aspects of verification at ECMWF

The previous two sub-sections have summarized the progress in verification of TC position and surface wind, with the increasing ability to simulate TC structure enabling new metrics such as Vmax and the wind radii. Forecasting challenges related to TC structure will be further discussed in Section 5.

One future goal is to evaluate other variables that convey the impact, such as rainfall in connection with TCs. However, it is problematic to find reliable ground measurements from rain gauges and radar, due to the harsh environment. During a visit to ECMWF in 2016-2017 by Tung Nguyen, sponsored by WMO, the skill for rainfall during TC landfalls in Vietnam was evaluated over a ten-year period in ECMWF forecasts against high-density rainfall observations provided by the Vietnamese Meteorological Service. The skill was found to be on a higher level compared to other type of events. Another option for verification is to use satellite derived rainfall from the GPM dataset (Skofronick-Jackson et al, 2017; Huffman et al., 2014) along the TC tracks (Omranian et al., 2018). An example for such comparison is shown in Section 9 (64) for TC Idai (2019). As further discussed in Section 9, accurate representations of precipitation structure and duration are also necessary for flood forecasting and hazard mitigation.

The surface and 3-dimensional wind structure and precipitation distribution are important for capturing rapid intensification, for example. The TC structure is also important on its own, to provide improved forecasts and warnings of hazards associated with TCs. Examples include the broader wind field outside Vmax (which is only at one point), which contributes to storm surges and large-scale wind or wave damage as discussed above.  However, many of the structural aspects of the hurricane are poorly observed and predicted, such as the fluxes at the air-sea interface, the structure and processes within the boundary layer, the organization and consequences of updrafts and downdrafts, and eyewall processes such as eyewall replacement cycles and intense mesovortices inside the eyewall. Some quantities can be derived such as outflow from atmospheric motion vectors and rain-structure from radar (airborne or ground-based). But it is not straightforward to set up regular verification of these quantities, and is expected to be time-consuming. It is therefore an area where ECMWF can seek collaboration for the wider community.

Another future goal is to verify variables that characterise the vertical structure of TCs. While these variables do not directly convey the impact, they are useful to evaluate the fidelity of the model analyses and forecasts. Characteristics include the vertical change in wind structure above the boundary layer, the warm core, and the upper-tropospheric outflow. However, it is difficult to verify these quantities routinely, given the limited availability and quality of observations in TCs.

As will be demonstrated in Section 5, the characteristics of the forecast errors and biases depend on the characteristics of the TCs themselves, such as the size or intensity. Hence, there is an ongoing need to perform conditional verification to understand the regional differences in errors and how they differ during different phases of the TC life cycle. In the past, this has helped to understand the impact on ocean coupling.

In operational ensemble verification of field variables, observation uncertainty is taking into account for the spread-error relation (metric discussed in Section 5). For TC verification, ECMWF could in the future explore to take the Best Track estimated error into account for this type of verification.

One uncertainty in the intensity verification is the inconsistency in the Best Track estimates due to different practices in different basins and availability of observations. Since 2016, ECMWF operationally produce simulated satellite images based on infrared channels on Meteosat (see Figure 54). This gives the opportunity to track TCs and estimate intensity in the model with the Advanced Dvorak Technique (Olander and Velden, 2007) in a consistent way as done with real satellite images. This was documented in Magnusson et al. (2017). The result was successful for most of the cases. The main shortcoming that was identified was in situations when cirrus clouds obscure the eye of the cyclones.

Direct evaluation based on IR-images as shown in Section 7 should also be considered in the region around tropical cyclones. Such targeted evaluation could also be applied based on data assimilation output for other assimilated observations. We could also consider making direct evaluations of non-assimilated observations discussed in Section 2, such as SFMR surface winds from reconnaissance flights.

ECMWF does not currently perform a regular verification of TC activity/genesis for medium-range forecasts. However, it is evaluated for extended-range forecasts. In Yamaguchi et al. (2015), the TC activity of ECMWF forecasts was compared with other forecasts in the TIGGE archive. One difficulty here is the use of specific wind speed threshold to detect cyclones, which makes the number of the TCs sensitive to biases in maximum wind speed. The solution applied in Yamaguchi et al. (2015) was to find the intensity threshold that maximised the skill for each system. As similar method was applied in Bergman et al. (2019) for seasonal forecasts. The results in Yamaguchi et. (2015), performed during a sabbatical at ECMWF by Munehiko Yamaguchi (JMA), showed an advantage for ECMWF forecasts in most of the basins. However, as this study is based on forecasts that are about ten years old, it is time to redo this type of verification. The probability of TC genesis in the ENS forecasts is discussed in Section 5.

Focusing on seasonal timescales, Bergman et al. (2019) made verification of landfall risks in ECMWF forecasts. Such verification could also be adapted to shorter timescales such as extended-range.

# 5. Forecast challenges at different stages of the TC life cycle

## 5.1. Introduction

As for all weather systems, predictions of TCs pose challenges on multiple scales, spanning from correctly initialising the system on the relevant scales, to accurately modelling the processes on multiple scales, to accurately representing the climatology of their occurrence. In this section, the characteristics of predictions and their errors on different timescales are presented. To begin, the forecast and analysis errors based on ECMWF's standard overall verification are described in Sections 5.2 and 5.3 respectively. Following this, the characteristics of TC position (track) and intensity errors are provided in Section 5.4 and 5.5 respectively. Prediction of genesis, marking the beginning of the TC's lifetime, is addressed in Section 5.6. Extratropical transition and decay, marking the end of the TC's lifetime, is described in Section 5.7. Extended-range and seasonal predictions of TC activity are summarized in Section 5.8, followed by a summary of the key results in Section 5.9 and main challenges in Section 5.10.

## 5.2. Effect of model resolution on forecast errors

To give an overview of the errors in the operational forecast with different resolutions, Figure 9 shows the mean absolute error (MAE; Wilks, 2006) for position, Pmin, and Vmax, the mean error (or bias, ME) for Pmin and Vmax, and the wind-pressure relation for the following three sets of forecasts during July-November 2020:

    (i)    Operational HRES for Cycle 47r1;

    (ii)   Ensemble control for Cycle 47r1 (ENS-CF-47r1); and

    (iii)  Pre-operational control for Cycle 47r2 (ENS-CF-47r2).

During this period, the horizontal resolution for HRES was 9 km and ensemble control forecast (CF) 18 km. In Cycle 47r2 that became operational in May 2021, the HRES forecasts performed very similarly to the previously operational Cy47r1 HRES forecasts (not shown). In Cycle 47r2, the vertical resolution of the ensemble was increased from 91 to 137 levels to make it the same as HRES (Lang et al., 2021). The cycle also included a change to single precision that did not have any significant impact on TC performance.

The position errors in HRES, ENS-CF-47r1, and ENS-CF-47r2 are very similar up to day 3 (Figure 9(e)). Beyond day 3, the ENS-CF-47r1 errors become larger. On the other hand, the position error for ENS-CF-47r2 becomes closer to HRES. However, all differences are within the 95% (unpaired) confidence interval. One can note that the confidence interval is large for long lead times as the sample size decreases, and TCs are more likely to be interacting with extratropical systems where very large errors can occur.

For Pmin, the HRES ME (bias) remains at about +2 hPa during days 0-3, but then decreases and becomes negative after day 4 (Figure 9(b)). The change in ME is likely related to the stage of the TC lifecycle, as the sample in the verification of longer lead-times contains a higher fraction of TCs that were mature at the initial time (Rodwell et al., 2017, Figure 1). ENS-CF-47r1 develops a higher positive ME after the initialisation, given a more limited ability to intensify TCs due to the lower resolution. The difference in bias is reflected in the lower MAE for HRES than the ENS-CF-47r1

(Figure 9(a)). Stronger TCs are evident in ENS-CF-47r2. We believe that this is linked to a larger fraction of large-scale (resolved) precipitation with the higher vertical resolution, which helps to build up the vertical circulation in the TC (see discussion around explicit deep convection in Section 7). Finally, as will be discussed further down in this section, strong biases in HRES exist for Pmin when the initial TC is below hurricane strength, even if the average bias is small.

For Vmax, all configurations underestimate the maximum wind speed (Figure 9(d)). The relation between Pmin and Vmax, visualised as a wind-pressure plot in Figure 9(f), aggregated for 24-120 h forecasts, does not show any large differences between different lead-times. The lines are based on a quantile-quantile mapping between the two quantities. As expected, TCs with a lower Pmin possess a stronger Vmax. All configurations underestimate the winds for a given Pmin, and the underestimation is worse for both ENS-CF-47r1 and ENS-CF-47r2 compared with HRES. This indicates that increased horizontal resolution improves the relationship (see also Section 7).
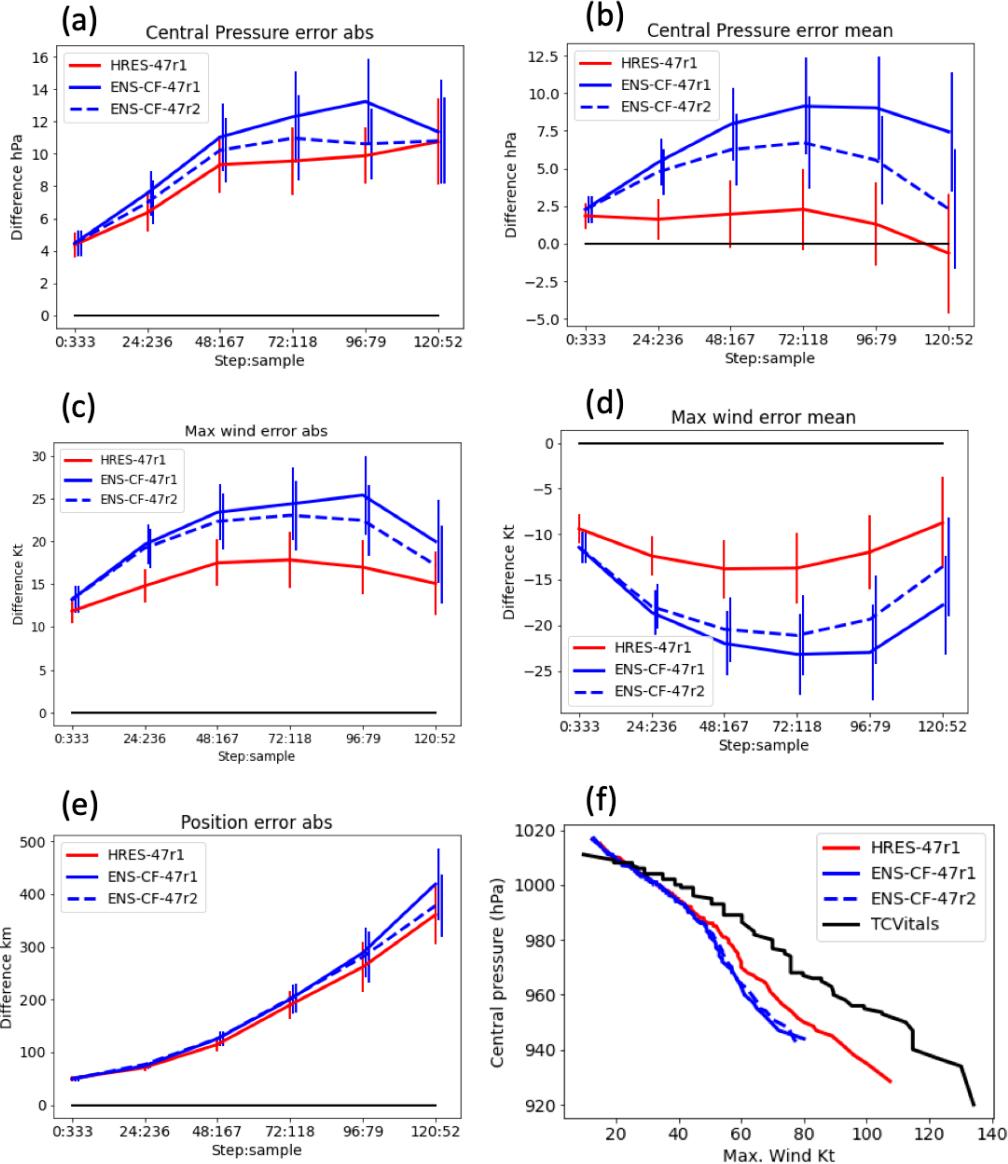
*Figure 9: Position and intensity evaluations for (red) HRES, (solid blue) ENS Control, and (dashed blue) pre-Operational 47r2 ENS Control: (a) Mean Absolute Error and (b) Mean Error for Pmin. (c) Mean Absolute Error and (d) Mean Error for Vmax. (e) Mean Absolute Error for position. (f) Wind-pressure relation for 24–120-hour forecasts.*

## 5.3. Analysis errors of tropical cyclones

Initializing an accurate physical structure of the TC in the data assimilation is a challenging task, due to the lack of observations within the TC to constrain the system and the limited resolution of the data assimilation. In this section, the analysis errors are presented. A more in-depth discussion of the data assimilation challenges will be provided in Section 6.

In Figure 10, the position error of the ECMWF analysis is illustrated for all Atlantic TCs after 1 July 2020 (cycle 47r1).  The position is produced by the ECMWF tracker, and the verifying value is the Best Track from NHC. We note that there are uncertainties in the Best Track, as documented in Landsea and Franklin (2013), although we expect that these uncertainties are now lower in the 2020s due to more advanced satellite imagery. As an example, the 2013 NHC estimate of position error uncertainty for tropical storms (Vmax < 33 m/s) was 53 km if no aircraft reconnaissance data were available. For major hurricanes (Vmax > 49 m/s), the corresponding estimate was 26 km, given the relative ease in identifying the position of the eye. Overall, the majority of the 226 ECMWF analysis errors of TC position is 30 km or less. However, a small fraction of these errors is large (> 60 km), and most of these cases are within the subset of weak TCs including depressions (pink bars). On the other hand, no such large analysis errors of position are evident for hurricanes (light orange bars).
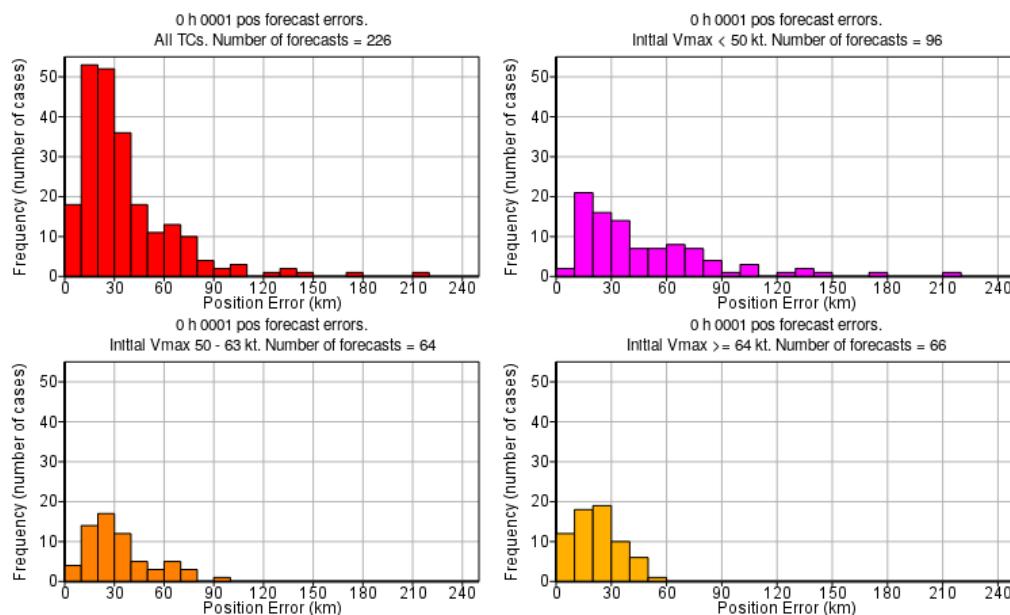


*Figure 10: ECMWF analysis position error as a function of initial Vmax (Best Track value).*

The corresponding ECMWF analysis error estimates for Pmin are shown in Figure 11. While over 100 of the 226 cases show a very small absolute analysis error (<2.5 hPa), there is approximately an equal number of cases where the analysis value of Pmin is too weak by at least 2.5 hPa. Several of these cases, especially the small number of extreme ones, correspond to hurricanes that are far too weak in the initial conditions (up to 40 hPa). Although not shown here, the majority of the analysis values of Vmax are at least 5 m/s weaker than the corresponding NHC Best Track values.

To conclude, the largest initial position errors usually occur in weak TCs while the largest initial intensity errors usually occur in strong TCs.
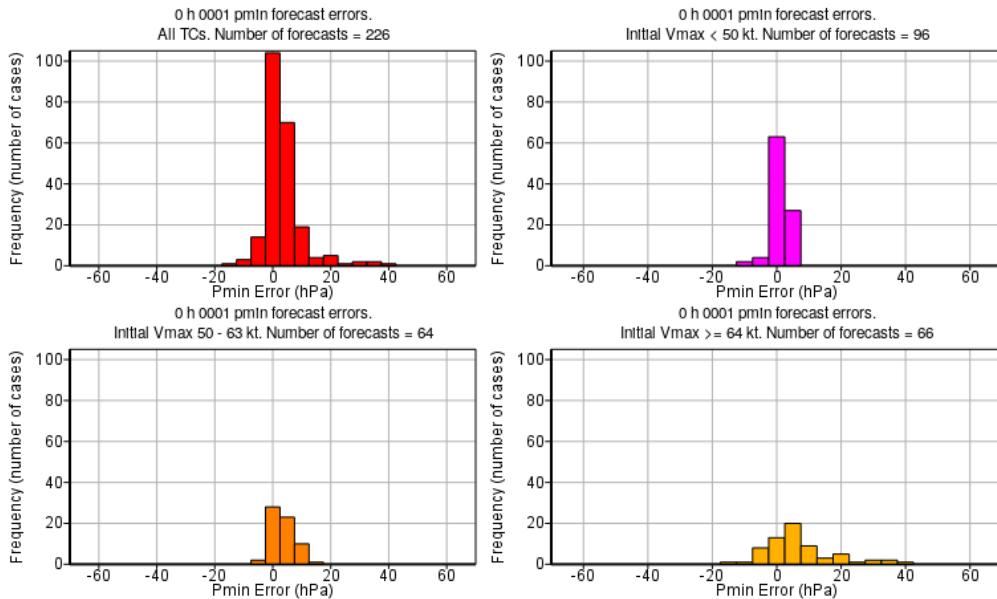
*Figure 11: ECMWF analysis Pmin error as a function of initial Vmax (Best Track value).*

To put these results in perspective, these values are compared against the US Navy's operational COAMPS-TC® system (Doyle et al. 2014) which uses a TC vortex-following grid with an innermost grid spacing of 4 km. Unlike ECMWF, COAMPS-TC employs the insertion of a synthetic, relocated and balanced Rankine vortex to initialize all numbered TCs. The value of Vmax is constructed to match the operational RSMC estimate, and the radial wind profile is adjusted to fit the RSMC estimate of the radius of maximum winds and radius of 17 m/s winds (Komaromi et al., 2021). Hence, by design, the initial vortex in COAMPS-TC is forced to contain much lower initial condition errors of position and intensity. The COAMPS-TC analysis errors of Pmin (evaluated versus the NHC Best Track) are shown in Figure 12. A comparison of this figure with Figure 11 shows a higher fraction of COAMPS-TC Pmin analysis errors nearer to zero. The contrasts between the COAMPS-TC and ECMWF analysis errors are more distinct in the position and especially the Vmax errors.
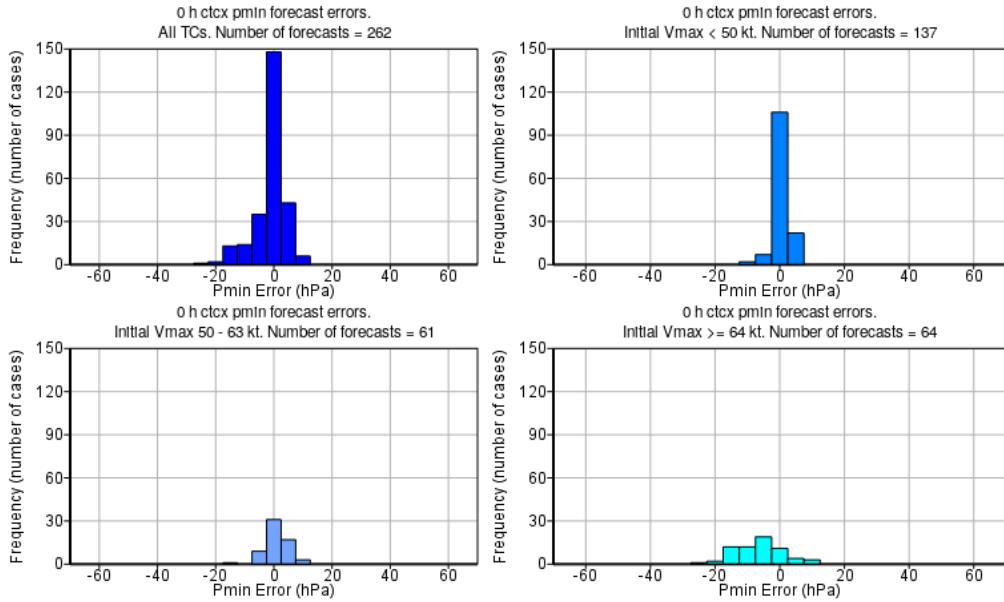
*Figure 12: Same as Figure 11 but for COAMPS-TC.*

While a homogeneous sample during the 2020 Atlantic hurricane season was used in this comparison between ECMWF and COAMPS-TC, the number of cases, even at 0 h, was substantially less in ECMWF (226) than in COAMPS-TC (262). The reason for this is that weak TCs were sometimes not trackable in the ECMWF analysis fields. In COAMPS-TC, the full sample of TCs exists by design of the synthetic vortex initialization.

## 5.4. Position (track) predictions

Arguably the most important aspect to predict for TCs in the medium range is its future position (track). As seen in Section 1, the skill of ECMWF TC position forecasts has improved over the past decades, but other centres have also improved at a similar pace. However, there exists a year-to-year variability in the errors. For example, as illustrated in Figure 1, the average 5-day HRES position errors in 2020 were larger than those of the previous five years for HRES. There are still several challenges in track prediction, some of which are described and illustrated below. In Magnusson et al. (2019b), the diagnostics of several challenging cases were discussed. Examples of tools used in such investigations include ensemble sensitivity (Torn et al., 2018), adjoint sensitivities (Doyle et al., 2012), steering flow diagnostics (Tang et al., 2021) and relaxation experiments (Magnusson et al., 2019b). Difficult cases to predict are often related to bifurcation points in the steering flow (Riemer and Jones, 2012), where a small error in the position of the TC can make a difference between a TC recurving into the mid-latitude flow, or instead continuing its westward propagation.

In Figure 10 we found that the analysis error and uncertainty for position is larger for weak TCs compared with stronger ones. From this finding, one can ask how the initial intensity of the TC affects the position (track) forecast. To answer this question, Figure 13 shows the position error for Atlantic TCs between 1 July – 18 November 2020 (Cy47r1), divided according to their initial intensity. It is evident that the average position error is over 70% larger for initially weaker TCs (magenta line) than

those for initially mature hurricanes (brown line) in the short range (24-48h), and this large discrepancy continues out to 84 h. This is an expected result, given that there is larger initial uncertainty in the structure of the weaker TCs and their environmental interactions.
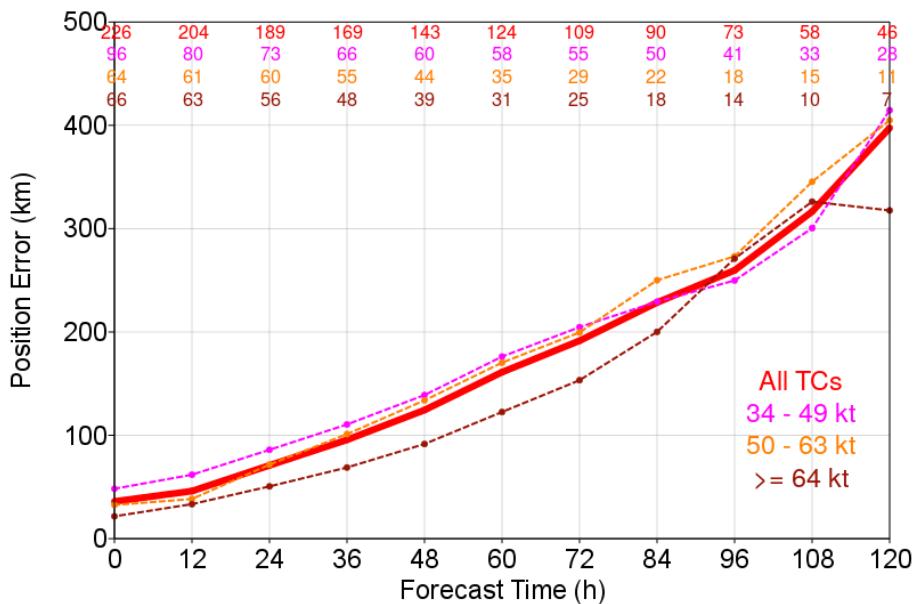


*Figure 13: MAE of operational HRES forecasts for position, for all Atlantic basin TCs between 1 July – 18 November 2020 (AL05 Edouard – AL31 Iota), stratified by Vmax at the initial time. Red: all cases. Magenta: only TCs with Vmax between 34-49 kt (weak tropical storms). Orange: only TCs with intensity 50-63 kt (strong tropical storms). Brown: only TCs with Vmax> 64 kt (hurricane strength). The sample sizes of each subset are listed by their respective colour.*

The key tool for predicting TC position in the medium range is ensemble forecasts, and we next focus the diagnostics on this system. A common metric in ensemble-based prediction and evaluation is the ensemble mean. For TC position, it refers to the Euclidean mean of the positions from each ensemble member (in contrast to other metrics where the ensemble mean refers to a field average of all members). As a start, the error of the ensemble mean and standard deviation (spread hereafter) of TC position are displayed in Figure 14 for all TCs in 2020 in all basins. In a well-tuned ensemble, these two quantities should coincide over many cases. On average, we find that the ensemble spread is lower than the average error of the ensemble mean. Figure 15 shows a scatterplot of ensemble spread and ensemble mean error for 3-day and 5-day forecasts from the Atlantic cases in 2020. The plot also includes the running mean based on 20 ensemble spread points sorted from low to high. For both lead-times, the running-mean line is above the 1-1 line, indicating a larger error than spread. However, it is also evident that the ensemble has some discriminatory skill to predict cases with larger error from the ensemble spread. For the 3-day lead-time, the most problematic forecasts were for Laura, Eta and Zeta. In the case of Laura, the forecast issued on 24 August 00UTC showed very large ensemble spread, while the forecast from 25 August 00UTC showed very large ensemble mean error, and this case will be discussed further down in this section.
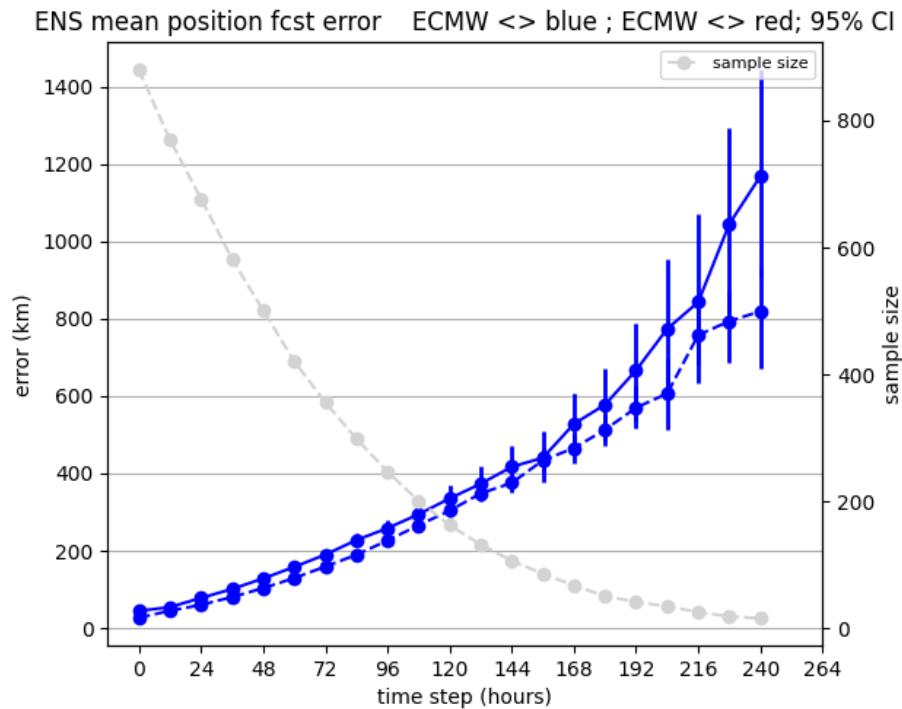
*Figure 14: Mean position error of ensemble mean (solid) and ensemble spread (dashed) for cases during 2020 for all basins. Units in km. Vertical bars represent 95% CI.*
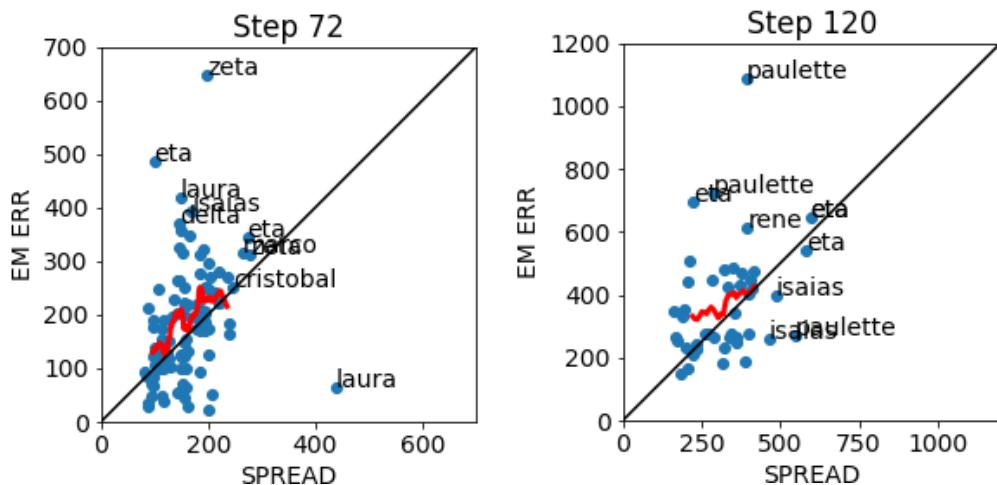


*Figure 15: Scatterplot of 3-day (left) and 5-day (right) ensemble spread (x-axis) and ensemble mean position error (y-axis) in km for cyclones in the Atlantic during the 2020 season. A running mean based on 20 ensemble spread points are included as red line. The 1-1 line is included in black. The most extreme cases are annotated.*

The four forecasts with the largest 5-day ensemble mean position errors in 2020 (each exceeding 600 km) are illustrated in Figure 16. In each case, the TC was gaining latitude during the forecast window as it moved towards the extra-tropics. The ensemble spread in each of these cases (dashed green line)

was sometimes substantially lower than the corresponding error of the ensemble mean (solid green line), consistent with the underprediction of the ensemble mean error shown in Figure 14. As is evident from a visual inspection of the ensemble forecast tracks (grey lines), the verification at five days lies either on the periphery of the ensemble, or outside the entire ensemble. Further investigation is required on the challenges of ensemble forecasts to encapsulate the actual TC position in a disproportionately large number of cases. These challenges include ensemble initialisation and model perturbation techniques, which will be discussed in Sections 6.9 and 7.4 respectively.
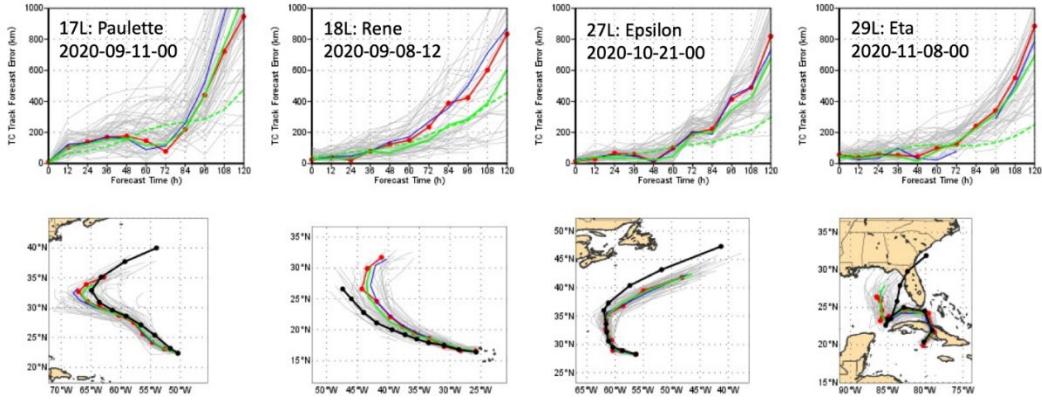


*Figure 16: Operational HRES (red), ENS-CF-47r1 (blue), ensemble members (grey), ensemble mean (solid green) forecast errors (top) and forecasts (bottom) out to 5 days, evaluated against the NHC Best Track (black) for four TC cases. These cases possessed the largest 5-day ensemble mean position errors of all Atlantic TC forecasts in 2020. The ensemble spread is represented by the dashed green line in the top panels.*

To further illustrate the forecasting issues with TC Laura, Figure 17 shows the longitude for the crossing of 30°N in each ensemble member (red) and HRES (blue), for forecasts initialised on different dates. The crossing of 30°N was chosen to be the latitude of the coast of Texas/Louisiana where the cyclone made landfall on 27 August. The forecasts issued from 24 August 12 UTC to 25 August 12 UTC had almost all ensemble members to the west of the actual crossing. For earlier forecasts, a large ensemble spread was present, but we can also see consecutive forecasts in which a majority of the ensemble members "flip-flop" between the east and west of the actual crossing, which is undesirable from a forecaster's perspective.
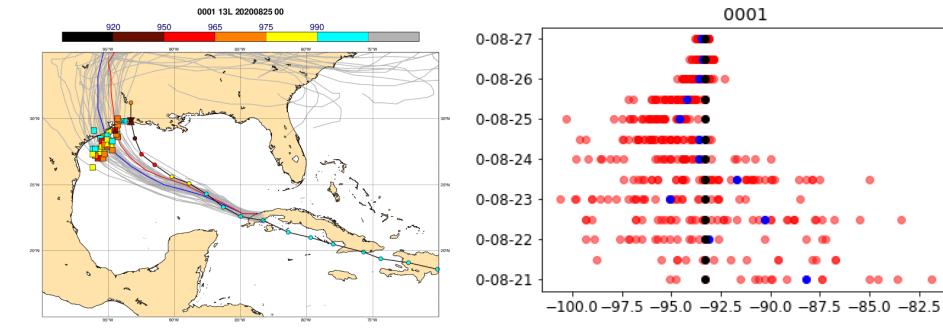
*Figure 17: Left: Forecast for TC Laura from 25 August 00UTC. HRES (red), ENS Control (blue), ensemble members (grey) and BestTrack (black). Symbols indicate position and intensity on 27 August 00UTC. Right: Longitude of forecasts passing the latitude of 30N for HRES (blue) and ensemble members (red) for different initial times. Best Track longitude marked with black.*

We can summarise the inconsistency (jumpiness) in a sequence of ensemble forecasts using the divergence index (DI) of Richardson et al. (2020), shown in Figure 18. Applying this to the cross-track errors for the 2020 Atlantic TCs shows that Laura stands out as the most inconsistent of all these cases. This is in agreement with subjective feedback from forecasters who were trying to assess the areas most at risk along the US Gulf coast. The inconsistency was especially large around the time of landfall (verification times between 26 August 12UTC and 27 August 12UTC).
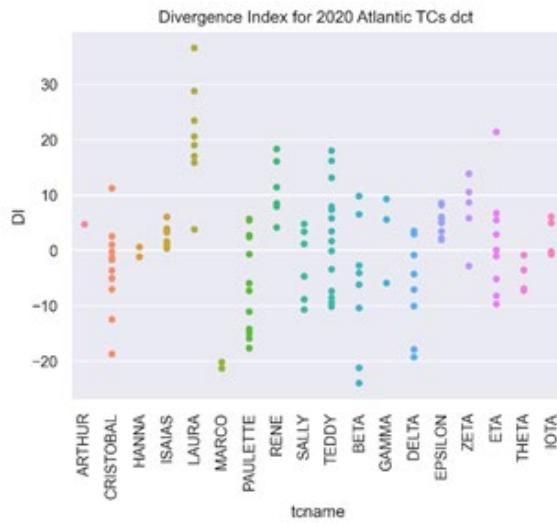


*Figure 18: Inconsistency of ECMWF ensemble track forecasts (2020 Atlantic; all cases with at least 20 members for at least T+12,…T+60). Divergence index (DI) – large positive numbers indicate inconsistent cases (large negative numbers indicate high consistency).*

While one can argue that the cross-track part of the error is most important to determine the position of the landfall, the along-track error can affect the timing of the landfall and hence the interaction between the storm surge and tides. Both can also be critical in determining interaction with extra-tropical features, such as an upper trough. Whether or not the TC ends up in the mid-latitudes, ahead

of the trough (where it may well be reinvigorated) or stays in the tropics can depend on whether it moves fast enough to catch the upper trough 'bus'. Some of the largest position errors can occur in such circumstances. Hurricane Leslie, which caused widespread damage when it hit Portugal as a re-intensified feature in October 2018, was one such case.

Figure 19 shows the distribution of along-track and cross-track errors for all 2-day forecasts of Atlantic TCs in 2020. For the cross-track biases (left panel), the westward propagating TCs (left panel) exhibit a slight overall northward (right-of-track) bias. For the subset of weak westward propagating TCs, this drift to the north is often evident, as illustrated in Figure 20. Most or all of the ensemble members (grey lines) are north of the actual track (black line), and this challenge warrants further investigation. On the other hand, TCs that propagated to the north, which often were strong TCs, in many cases showed a west (left-of-track) bias (Figure 19, right panel). For these northward moving storms, a strong negative (slow) along-track bias is also present.
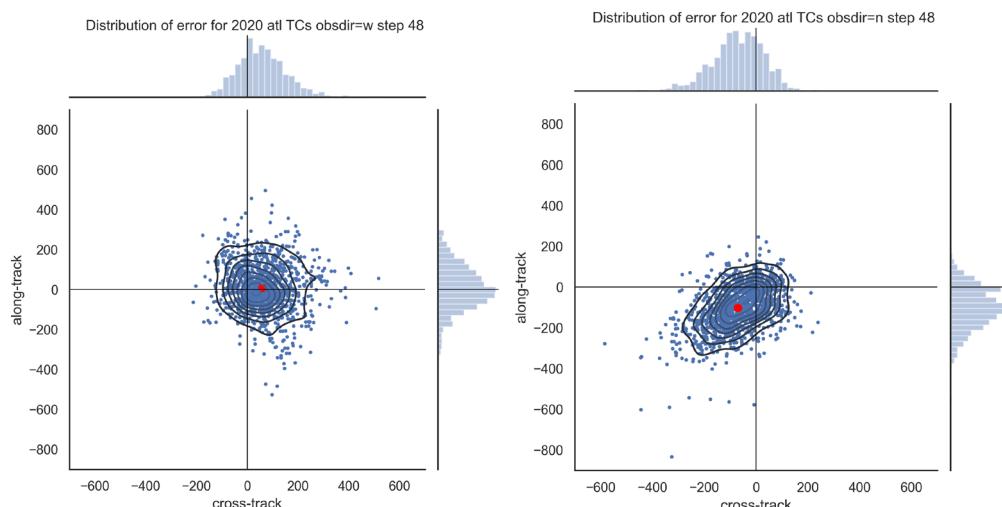


*Figure 19: Cross-track and along track errors for westward moving (left) and northward moving (right) TCs (2020 Atlantic). T+48. The red circle shows the distribution mean. The sub setting of the cases is based on a movement direction +/- 20 degrees from west and north respectively. Positive cross-track bias indicates right-hand bias.*
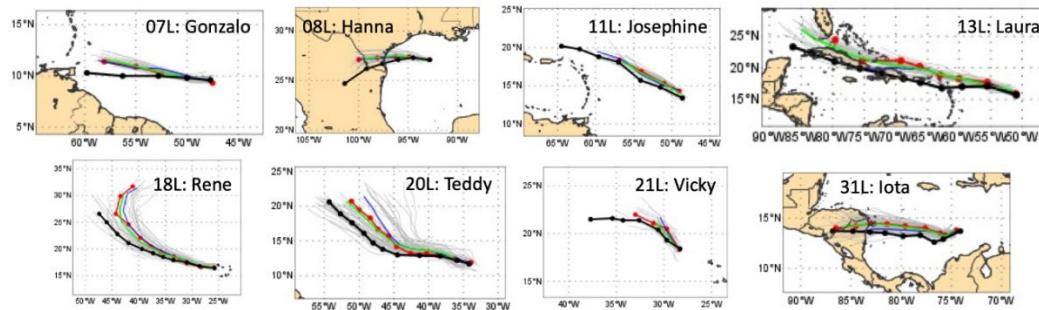
*Figure 20: Examples of 2020 forecasts in which weak, westward-moving Atlantic TCs were predicted to the right (north) of the actual track. Black: NHC Best Track. Red: Operational HRES (47r1). Blue: ENS-CF-47r1. Grey: Ensemble members. Green: Ensemble mean. Dots: 12-hourly locations.*

The along-track bias for short forecast ranges (48-hour) agrees with the propagation speed bias diagnostic that is routinely produced in the ECMWF verification report (Figure 32 in Haiden et al., 2021), where we find a long-term negative bias on the order of 1 km/h. This also agrees with other studies as Leonardo and Colle (2020), who used ECMWF ensemble forecasts from the TIGGE archive. They found the largest contribution to the bias from cases approaching the mid-latitudes. However, in their data a (smaller) bias was also present in cases that dissipated in the tropics. One could speculate that the stronger signal for the high-latitude cases is caused by the acceleration of the TC when interacting with the mid-latitude flow, which stretches the differences. In Chen et al. (2019), the GFDL-FV3 model and IFS with the same (ECMWF) initial conditions were compared. FV3 showed much less of a slow bias than IFS and had lower track errors for longer lead times.

In 2018 scientists at ECMWF made a range of sensitivity experiments to target the slow propagation bias. Even if some results indicated a sensitivity to the model time-step, the results were mainly inconclusive. This topic will be further explored in Section 7.

## 5.5.    Intensity predictions

In contrast to track forecasts, operational forecasts of intensity did not improve at the same pace in the 1990s-early 2000s (e.g., Fig. 6 of Rappaport et al., 2009). In response to this concern, together with multiple hurricanes that made landfall in the United States in 2004 and 2005, NOAA initiated the Hurricane Forecast Improvement Program in 2009 (Gall et al., 2013). In Titley et al. (2019), intensity forecasts from ensembles were highlighted by users as an area where they want to see progress. The approach at many weather services has been to use regional NWP systems with TC-following nested grids with grid spacings of 4 km or less such as AROME Overseas, HWRF and COAMPS-TC, to resolve the convective-scale processes that produce high wind speeds in the eyewall. These regional systems have also been employed over the past decade in hundreds of research articles that shed new light on TC structure and intensity evolution, and their predictability. As is now evident in regional systems such as HWRF and COAMPS-TC, predictions of Vmax have improved substantially during the late 2010s (Cangialosi et al., 2020). It is worth noting that these models have their own vortex initialization schemes, with HWRF recently introducing regional data assimilation of inner-core data including aircraft data (Zawislak et al., 2021; Christophersen et al., 2021).

With the horizontal grid spacing of the ECMWF HRES now smaller than 10 km, it is finally within reach to accurately predict the intensity of larger TCs with eyewalls that can be resolved in the model, and an improvement in Pmin errors has been achieved in the past decade (Figure 1).

The evaluation of Pmin and Vmax analyses and forecasts for all Atlantic TCs between 1 July – 18 November 2020 reveals some important although unsurprising results for consideration. First, the MAE in intensity analyses and forecasts vary based on the initial intensity. Especially for forecasts of three days or less, the initially weaker TCs have lower forecast errors (Figure 21(a)). This is largely due to the presence of a high proportion of weaker TCs that do not intensify significantly, thereby producing several low-error cases. On the other hand, for TCs of hurricane intensity, the initial MAE of Pmin is 7.5 hPa (Figure 21(a)), indicating limitations in the data assimilation system to deepen strong TCs to their estimated values in the Best Track. This MAE increases to approximately 13 hPa for 36-48 h forecasts of TCs of hurricane strength. Similar conclusions are drawn for Vmax forecasts (Figure 21(b)). For weaker TCs, the initial Vmax errors are small, and they amplify gradually out to five days. For TCs of hurricane strength, the average initial Vmax errors are very large, exceeding 25 kt. The Vmax error decreases with longer lead times, partly because of the storms becoming weaker and partly as a compensation of errors with the developing Pmin bias.

The corresponding biases in the intensity analyses and forecasts provide further insights. For initially weak TCs, the Pmin forecasts are consistently too weak (Figure 21(c)). In this sample, several cases exist in which the forecast erroneously does not intensify an initially weak TC. For TCs of hurricane strength, the forecasts beyond one day yield overly strong TCs. An examination of the individual cases reveals that this strong bias reflects a generally limited ability of the forecasts to increase the Pmin during the weakening phase of the TC. The biases in Vmax demonstrate a systematic underestimation of the maximum surface wind at all intensities (Figure 21(d)). The bias at longer lead times is lowest for initially strong cyclones, which should be related to the too low Pmin in these cases.

To summarise, the stratification by initial intensity reveals several challenges in analysing and predicting TCs, with contrasting challenges for weak versus strong TCs.
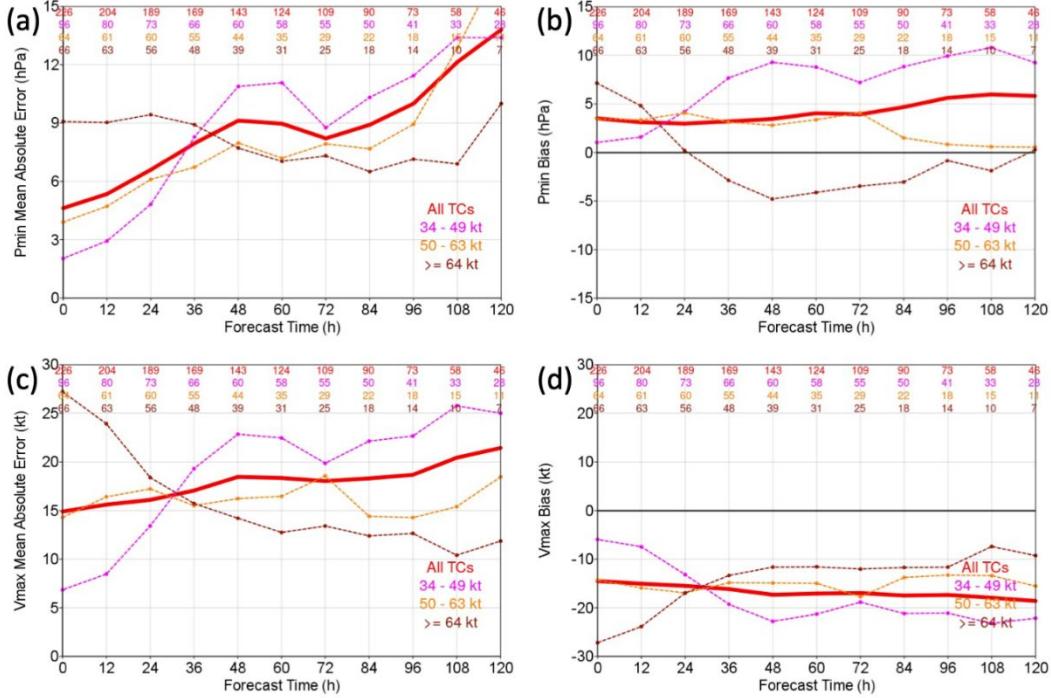
*Figure 21: MAE of operational HRES forecasts for (a) Pmin and (c) Vmax, and respective mean error or bias (average of forecast value – verification value) for (b) Pmin and (d) Vmax, for all Atlantic TCs between 1 July – 18 November 2020, stratified by Vmax at the initial time. Red: all cases. Magenta: only TCs with Vmax between 34-49 kt (weak tropical storms). Orange: only TCs with intensity 50-63 kt (strong tropical storms). Brown: only TCs with Vmax> 64 kt (hurricane strength). The sample sizes of each subset are listed by their respective colour.*

TC intensification and weakening processes need to be captured more accurately in the NWP system. The intensity change depends on a variety of processes on multiple scales, which may act non-linearly to positively reinforce each other or compete against each other. First, the coupling with the ocean is important to ensure that the fluxes of temperature and moisture at the ocean surface are appropriately represented. Further details of ocean coupling are provided in Section 7. Second, the environmental wind field, commonly represented by the 200-850 hPa vertical wind shear, is known to be influential on intensity change. However, although the general assumption is that moderate to strong wind shear (5-15 m/s) is usually detrimental to intensification, there are situations in which TCs can intensify, even rapidly, in these conditions. Third, the presence of low-humidity air is also usually assumed to be detrimental to intensification. Other environmental factors include interactions with subtropical or extratropical jets, which may control the outflow and thereby the intensity change, and interactions with land which modifies the supply of thermal energy and moisture through the TC surface layer and boundary layer.

Additionally, and of equal importance, are inner-core processes on the mesoscale and convective scales, which are governed in part by the environmental interactions. These inner-core processes include the organization of initially disorganized convection, rainband formation, and eyewall formation leading to intensification. The reverse of these processes can lead to TC weakening.

An especially challenging and societally important aspect of intensity change is Rapid Intensification (RI), in which Vmax increases by at least 30 kt (15 m/s) during a 24-hour period.  The statistical and dynamical (global and regional) models to date have not exhibited high skill in predicting RI (Cangialosi et al., 2020), and it remains a very active research topic in the field of tropical meteorology. It is very difficult to discriminate between RI and steady intensification (intensification at a rate of < 15 m/s in 24 hours).  RI can often begin early in the life cycle of a TC, even before it is a named tropical storm, which presents an additional challenge in the accurate initialization of a weak, disorganized TC.  The abrupt axisymmetrization of convection and corresponding sharp increase in vorticity remains one of the toughest analysis and forecasting challenges in all NWP, even in regional models with a 1 km inner grid. While the environmental interactions are important, it is the precise nature of the O(1 km) inner-core processes, including eyewall formation and replacement, that needs to be captured physically. Nevertheless, some recent forecast cases (COAMPS-TC for Hurricanes Eta and Iota in 2020) demonstrated an ability to capture the RI process, albeit imperfectly, suggesting that there is potential for improvement in future global NWP systems when reaching 4-5 km resolution.  A probabilistic approach to RI, and intensity change in general, would likely be necessary.

Figure 22 shows the Pmin change rate over six hours (48-42h) in BestTrack and HRES for 2020 cases in all basins. The slope of the quantile-quantile matching line shows that the model is both too slow for the intensification and too slow for the weakening, compared with the Best Track. This result is consistent with the discussion above based on the conditional bias for initially weak storms that develop a weak intensity bias, while initially strong storms develop a strong intensity bias. We can also see three cases that stand out for missed intensifications: Goni (NW Pac), Iota (Atl), Eta(Atl). These three were small TCs that underwent a very RI that was clearly missed in HRES. But as mentioned above, the RI was captured by the COAMPS-TC regional hurricane model, illustrating the future prospects for predicting RI.
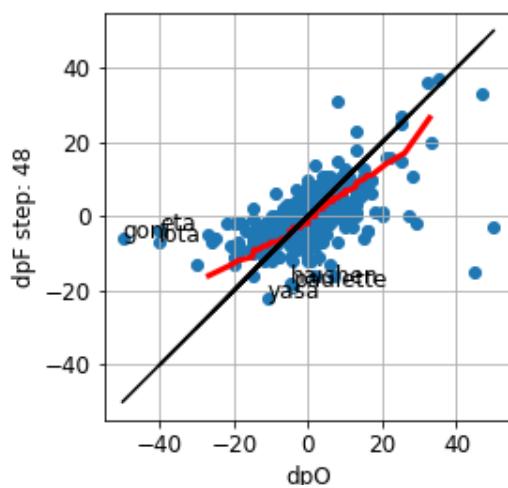


*Figure 22: Intensification rate (hPa/6h) in BestTrack (x-axis) and 48-42 hour forecasts (y-axis) from 2020. Red line shows quantile-quantile matching between predicted and BestTrack.*

## 5.6.    Genesis

Tropical cyclogenesis (frequently referred to as "genesis") refers to the instant at which a TC forms. In its basic state, the tropical atmosphere is benign. For example, in the tropical Atlantic, the relative humidity above the boundary layer is typically 60-70%, and weak subsidence serves to impede the development of new clouds. Hence, the middle troposphere needs to be moistened by repeated convection for a TC to develop. In addition to this enhanced moisture, it is conventionally thought that low vertical wind shear and high thermodynamic instability (aided by a warm ocean) are favourable environmental conditions.

We first evaluate the capability of the model to simulate the frequency of genesis occurrences. However, this is dependent on the wind threshold set to count as a genesis event. Given the low Vmax bias discussed in Section 5.2, we expect the model to underpredict the number of cases if we apply the same threshold for a tropical storm (17 m/s) as in the Best Track. We illustrate the effect of this Vmax threshold on the number of forecast TCs in Figure 23 . In general, the total number of TCs (including tropical depressions and higher) is overpredicted in the model compared with the observed number, while the number of tropical storms is underpredicted, consistent with the general low bias in Vmax. The number of TCs is increased in cycle 47r2, as expected from the increased intensities shown in Section 5.2. This cycle increased the vertical resolution in the ENS to be the same as that of the HRES, and the increase in the number of TCs is consistent with the increased intensity seen in Figure 9.

We can evaluate the impacts of the wind speed threshold and the cycle change using the performance diagram (Halperin et al., 2013). Figure 24 shows for five-day forecasts that adjusting the maximum wind threshold to reduce overall bias is a trade-off between number of hits and false alarms (success ratio is equal to one-false alarm ratio), while the overall skill measured by the threat score (curved lines – closer to top-right corner indicates higher skill). Cycle 47r2 improves the skill in forecasting the genesis of tropical storms, especially at the longer forecast ranges.
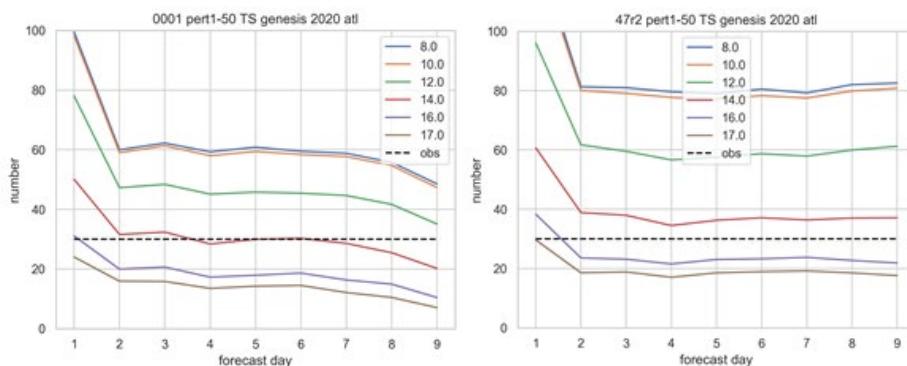


*Figure 23: Genesis of TCs in the ENS. Average number of TCs that develop in the perturbed ENS members on forecast days 1 to 9 in the Atlantic basin in 2020 (10 May – 30 November). Left: operational ENS (47r1), right: pre-operational cycle 47r2. Coloured lines show the average number of TCs defined using different thresholds of maximum wind speed from 8 m/s (threshold for tropical depression in operational tracker) to 17 m/s (threshold for tropical storm). Black dashed line shows the observed number of tropical storms (30 cases; 31 TCs altogether including tropical depression).*
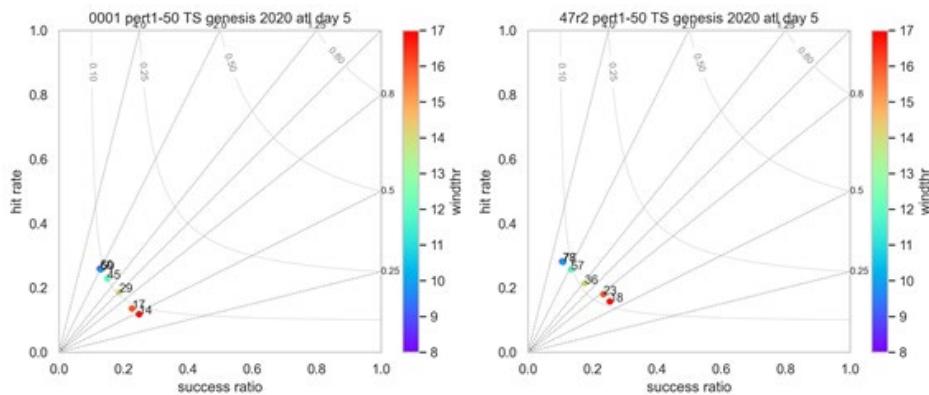
*Figure 24: Performance diagram for genesis of tropical storms in the ENS at forecast day 5 the Atlantic basin in 2020 (10 May – 30 November) for operational ENS (left) and pre-operational cycle 47r2 (right). Different points are for TCs defined using different thresholds of maximum wind speed from 8m/s (blue) to 17 m/s (red); average number of TCs shown for each case. Dashed straight lines show bias; curved lines show threat score.*

For genesis to occur, a precursor disturbance is necessary. These disturbances vary widely among different basins, and even in each individual basin. In the Atlantic basin, the majority of hurricanes initially developed from African Easterly Waves (AEWs). Other precursor disturbances include tropical waves of non-African origin, wave-ITCZ mergers, low pressure systems that did not originate from waves, broad gyres of low pressure over Central and South America, and extratropical baroclinic zones and even cyclones that drift into lower latitudes. In the north-western Pacific basin, TCs can develop from precursors such as a large monsoon trough, or easterly wave disturbances. In some instances, the wave structure and activity can be explained by shallow-wave theory (Kelvin waves, inertia-gravity waves and equational Rossby waves). Genesis can also be enhanced or hindered by large-scale oscillations, depending on their amplitudes and phases, and how these modulate the aforementioned environmental conditions such as wind shear and moisture. Such oscillations include El Niño - Southern Oscillation (ENSO), the Madden-Julian Oscillation (MJO), and Convectively Coupled Kelvin Waves. These and other oscillations may be concurrent, and their combined influence on genesis may be nonlinear, adding to the complications of predicting genesis on timescales of weeks to seasons. For example, in the south-west Indian Ocean, where Météo-France is the responsible RSMC, TCs most frequently form during phases 2 – 4 of the MJO, when MJO convection is active over the Indian Ocean. Landfall in south-east Africa has also been shown to be more common during La Niña, the cool phase of ENSO (Vitart et al., 2003). The potential for introducing diagnostics for tropical waves has recently been discussed with Météo-France, who demonstrated the utility of their tropical wave tracking software, and such diagnostics could provide important insights regarding TCs and genesis. Tropical wave diagnostics are also a part of the German Waves2Weather project. ECMWF aims to explore the utility of tropical wave diagnostics further in collaboration with these groups and in the project H2020-CLINT.

How far in advance are we able to predict genesis of tropical cyclones? It is a difficult question to answer as the variability is very large from case to case. For example, for Marco (Figure 25(a)) and Teddy (Figure 25(b)), the probability steadily grew from near 0% to near 100% as the forecast time

was shortened from 240 h to 0 h. On the other hand, for TCs such as Nana (Figure 25(c)), the corresponding probabilities were very small even three days prior to the TC being named. Additionally, isolated cases, especially pre-Laura (Figure 25(d)), exhibited "jumpiness" in the probabilities, even one-three days prior to genesis. Given the reliance on ensemble-based probabilities in products such as the NHC's Tropical Weather Outlook, and the rapid development or landfall (or both) in some TCs shortly after genesis, the predictive skill and predictability of genesis (and subsequent development) are important challenges to address.
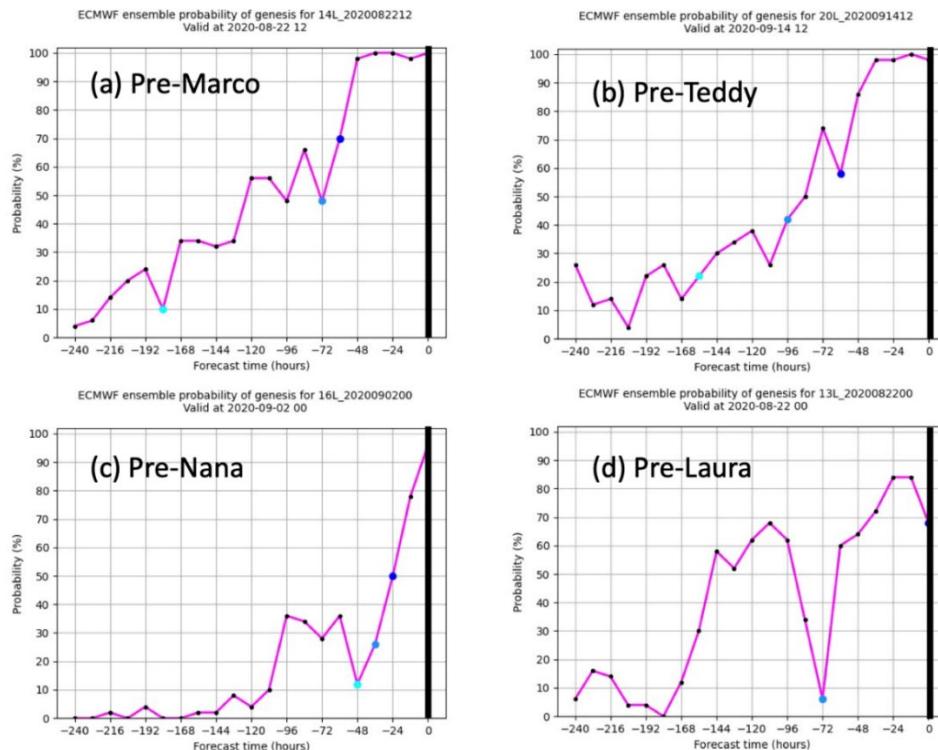


*Figure 25: ECMWF ensemble-based probability of the existence of a TC within 500 km of the actual location of the named TC, at the fixed verification time that it became a named Tropical Storm. The x-axis shows the time (in hours) between the initial time of the ensemble and the fixed verification time.*

Through a review of NHC Tropical Cyclone Reports and investigations of ERA5 reanalysis fields together with ECMWF forecasts, multiple pathways to genesis are evident, and these influence the predictability and ensemble forecast probabilities. The most straightforward cases in 2020 (Isaias, Marco, Rene, Teddy) were isolated AEWs with an amplitude stronger than most waves (measured, for example, by the 700 hPa relative vorticity averaged within a disk of radius 500 km). These robust waves developed in favourable environmental conditions, and we consider them to be the types of disturbances with the highest predictability. In other cases, such as Laura and Paulette, AEWs interacted with each other and/or low-pressure systems, with the probability of genesis being influenced strongly by the characteristics of these complex interactions. The ability of the analysis scheme to accurately represent these interactions is important in providing an accurate forecast. However, this is not straightforward, given the lack of observations in cloudy, mid-lower tropospheric

regions where the areas of weak vorticity are potentially beginning to amalgamate.  Another substantial challenge in the theory and prediction of genesis is the initiation of convection, and how it organizes around a wave or broad region of low pressure. This challenge may be particularly important to address in the cases of very weak waves and disturbances, such as those that led to Gonzalo, Hanna, Josephine, Nana, Sally, Gamma, and Delta in 2020. In each of these cases, the NHC predicted 20% or lower chance of genesis five days before the TC was named. In many of these cases, the NHC probability was still low (20% or less) even just two days before.  The ECMWF ensemble also predicted similarly low probabilities at five-day lead times. In several instances, the ECMWF ensemble probability rose sharply between three days and two days lead time.  We suggest that accurate probabilistic predictions from weak waves or disturbances are the primary challenge for genesis prediction, to reduce the number of "misses" just two-three days out.

It is also important to account for non-developing disturbances in evaluating genesis probabilities. Using a wave tracker (developed by Quinton Lawton, PhD student at the University of Miami, based on Brammer and Thorncroft 2015 and Elless and Torn, 2018), many non-developing waves in the Atlantic basin were objectively identified.  Some of these disturbances were weak waves in an unfavourable environment for genesis, whereas others were stronger waves that seemed more likely (but not certain) to develop. The ECMWF ensemble provided what we subjectively think are reasonable probabilities of development for these disturbances that ultimately did not develop. For the weak waves, these forecast probabilities were less than 20% for all lead times. For the stronger waves, the probabilities rose to 30-60% in some instances, but there were no cases in 2020 where there were obvious "false alarms". The main challenge in genesis prediction is to discriminate between weaker waves that develop into TCs, and weaker waves that do not.

## 5.7.    Extratropical transitions and tropical cyclone decay

At the end of the lifecycle, some TCs curve towards the extratropics and start to interact with the waveguide in the mid-latitudes. During the extratropical transition, the cyclone becomes asymmetric with a frontal structure and the core changes from warm to cold. For a fundamental overview of the processes, see Jones et al (2003). Extra-tropical transitions can cause substantial impact in the mid-latitudes, both if the cyclones that directly (Evans et al.,2017; Baker et al., 2021) hit in a sub-tropical stage or soon after extra-tropical transition (e.g., Sandy, 2012; Leslie, 2018; Lorenzo, 2019) or indirectly (Keller et al., 2019) as the extratropical transitions can lead to downstream development (e.g., after TC Karl, 2016; Schäfler et al., 2018). Even if the cyclones do not hit land, the ocean waves can propagate long distances and hit the coasts of Europe.

Whether or not a TC will approach the extra-tropics is determined by the steering flow. If a TC is close to a bifurcation point in the flow (Riemer and Jones, 2013), very large track forecast uncertainties and track errors can occur. It is therefore critical to correctly predict bifurcations in the steering flow and the TC track towards these points. An example of such sensitivity is discussed in Magnusson et al. (2014) for TC Sandy (2012) and in Magnusson et al. (2019b) for TC Joaquin (2015), where small changes in the sub-tropical ridge caused very large differences in the future track of these TCs.

A related uncertainty is the phasing with the mid-latitude wave guide, where an upstream trough favours a northward propagation into the extratropics. Correctly predicting the mid-latitude wave guide is crucial to capture the extratropical transitions. Such a sensitivity was highlighted in McNally et al. (2014) where they found that satellite data over the northern Pacific influenced the predictions of the landfall of TC Sandy. It is difficult to determine if the forecast error in extratropical transitions is due to incorrect predictions of the TC or of the mid-latitude wave guide, as the latter can affect the track of the TC. More studies are needed to compare situations with an active waveguide with and without extratropical transitions to investigate the impact on the forecast skill from ET on the mid-latitudes. During the extra-tropical transition, the propagation speed of the cyclone can increase a lot, which also can lead to very large position errors. Large uncertainties also lie in the possibility for the cyclone to re-intensify when entering the extra-tropical stage.

The propagation speed of the TC in the tropics determines the phasing with bifurcation points and also the mid-latitude flow, while the propagation speed after the curving towards the extratropics determines the phase-lock with the movement of the mid-latitude waveguide. Both these aspects can cause substantial errors in the predictions of the extratropical transitions. As discussed above, the ECMWF forecasts have a negative bias in the propagation speed. Such bias is present both in cases that did and did not undergo extratropical transitions but is larger in the transition cases (Leonardo and Colle, 2020). However, further diagnostics are needed to understand the impact from this bias on the mid-latitude skill.

The divergent outflow from the TC can also modify the large-scale flow and contribute to the outcome of the extra-tropical transition (e.g., Agusti-Panareda et al., 2004; Keller et al., 2019). The modification of the potential vorticity in the outflow is governed partially by condensational warming in the TC. Therefore, both the strength of the secondary circulation (connected to intensity) in the TC and the precipitation rate can impact the transition. As diabatic processes are parametrized in the model this is a potential source for forecast errors. Leonardo and Colle (2020) found a sensitivity to the rain rate in TCs bound for extra-tropical transition to the propagation speed of the TCs.

It has been suggested that these transitions decrease the medium-range predictability over, for example, Europe (Keller et al., 2019). Lillo and Parsons (2017) investigated the climatology of forecast bust cases over Europe in the ERA-Interim forecasts and found that the busts are most frequent in September and October, which coincides with the most active period for tropical cyclones in the Atlantic. In a recent (unpublished) update of the annual forecast bust frequency, ECMWF and other NWP centres saw an anomalous high number of busts in the summer-autumn 2017. But even if the worst period of low skill coincided with TC Harvey, the forecast error was tracked to a tropical depression east of Florida at the same time, which shows that it is not straightforward to link the low skill to the most severe hurricanes. In recent years the evaluation at ECMWF has shown that the medium-range performance relative to other centres is worst during the autumn period, but the relation to tropical cyclones is still to be understood.

While the TCs that undergo extra-tropical transitions may create substantial impacts downstream over Europe, the majority of TCs do not undergo extra-tropical transition. As was especially evident in 2020, several TCs can make landfall in the deep tropics or subtropics, spinning down quickly into a remnant low pressure system that can bring substantial flooding rainfall for several days. Other TCs

weaken as they encounter high vertical wind shear or substantial low-humidity air, which may occur in the tropics and especially the extratropics. As a TC moves into the extratropics, it also encounters much colder waters, removing the supply of thermal energy and moisture from the ocean that is necessary to maintain the TC.  As is implicit within the results of Figure 21, the HRES forecast often does not weaken strong TCs as fast as the actual weakening rate, which contributes to large errors in the intensity forecasts.  Overall, intensity forecast errors are dominated by difficulties in strengthening initially weak TCs (and some stronger TCs that intensify explosively), and difficulties in weakening initially strong TCs.

## 5.8.     Extended-range and seasonal predictions

This sub-section will discuss aspects of predicting TCs on extended (sub-seasonal) and seasonal ranges. Predicting genesis of TCs is an essential part of successful extended-range forecasts, and the subject was covered in Section 5.6.

The World Weather Research programme (WWRP) / World Climate Research Programme (WCRP) Sub-seasonal to Seasonal Prediction project (S2S) was established in 2013. Its main goals are to improve skill and our understanding of sub-seasonal to seasonal predictability and promote the uptake of S2S forecasts by the application community. A main delivery of this project has been the creation of a multi-model database containing 3-week behind real-time forecasts and re-forecasts from 11 operational centres. The S2S database represents a unique opportunity to investigate the skill of S2S models to predict TC activity up to week 6. Lee et al. (2018) produced an inter-comparison of the skill of the S2S database models to predict weekly TC genesis probabilities. Results show that the ECMWF model has the largest skill over the Atlantic, western North Pacific, eastern North Pacific and South Pacific, compared with the other models. The ECMWF extended-range forecasts display significant skill up to week 5 over the North Atlantic and western North Pacific and week 2 over the eastern North Pacific and South Pacific. Over the other basins, the skill is limited to week 1.

Important sources of extended-range predictability include ENSO and SST, and the MJO discussed in Maloney and Hartmann (2000). Vitart (2009) showed that this modulation of TCs by the MJO is well simulated in the ECMWF extended-range forecasts, except over the eastern Pacific and the Atlantic where the MJO teleconnections are too weak. Lee et al. (2018) showed that all the S2S models reproduce well this modulation, and there is a clear relationship between the ability of S2S models to represent the MJO and its impact on the TC activity and their skill in predicting extended-range TC activity. The model TC climatology also influences the performance in sub-seasonal prediction. S2S models are generally more skilful at predicting the probability of TC occurrence during the favourable phases of the MJO (Lee et al., 2020).

Lee et al. (2020) explored the impact of three different calibration methods to remove the mean TC genesis and occurrence biases, as well as the impact of a linear regression technique (van den Dool, 2017) on the probabilistic forecast skill scores. The linear regression method performed the best.

In recent years, and especially in 2020, the seasonal forecasts from ECMWF (SEAS5) of TC activity clearly missed the anomalies in the Atlantic TC activity, as seen in Figure 26. To follow up this issue, an internal working group met three times during the winter 2020/2021. The paragraphs below will therefore have a focus on seasonal forecasts for the Atlantic.
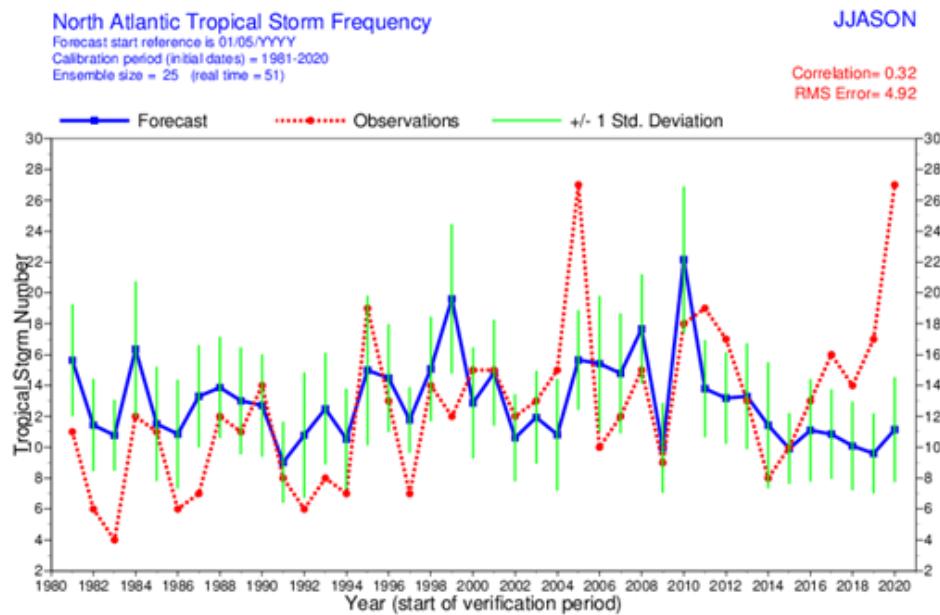
*Figure 26: TC activity in seasonal forecasts from SEAS5 (blue) and observations (red) before introducing the revised calibration.*

For longer lead times, the challenge is to identify drivers of predictability and the modulation of (basin-wide) anomalies in TC activity. Figure 27 shows the correlation between SST indices and TC frequency in the Atlantic, both based on observations and in SEAS5. While the SEAS5 has a comparable correlation from Nino3.4 with the observations, they are clearly missing the correlation with the local SST in the main development region (MDR) of TCs in the Atlantic, and also underestimate the link from Nino1.2 (easternmost part of the Pacific). Another important factor for the TC activity in the Atlantic is the vertical wind shear (difference of horizontal wind between 200 and 850 hPa) in the MDR, where large wind shear decreases the activity. The wind shear is partly modulated by the ENSO state as El Niño tends to increase the shear (Gray, 1979).
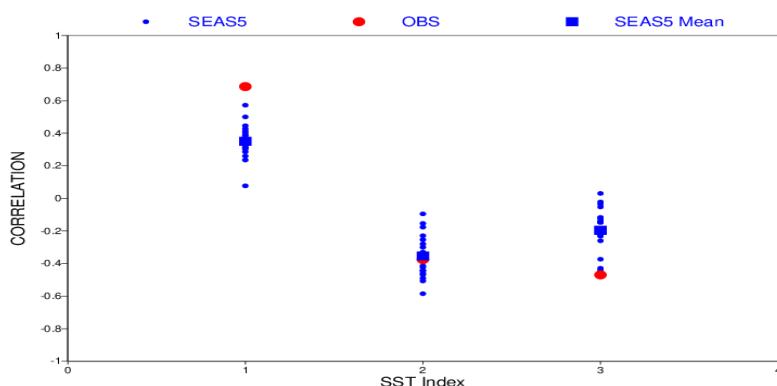


*Figure 27: Correlation between Atlantic TC activity and area averaged SST in the Atlantic main development region (left), Nino3.4 (middle) and Nino1.2 (right).*

In a recent investigation at ECMWF, the too weak link to the local SST was pointed out as one candidate, as 2020 saw a strong positive anomaly in the local SST. A positive trend in wind shear over the MDR was also found in SEAS5. This positive trend, which is not present in ERA5 reanalysis, could help explain the underprediction of TCs in the Atlantic in recent years. The underlying mechanism for this erroneous trend is under investigation.

Figure 28 highlights another problem for TCs in the Atlantic: the model climatology of TCs. The seasonal forecasts produce too many TCs in the Main Development Region and far too few in the Gulf in Mexico. This could have been an additional factor for the bad performance in 2020 as we saw a high activity in the Gulf of Mexico in the observations. Ultimately, the seasonal forecasts should be able to predict seasonal anomalies in landfalling TCs. This capability was investigated in Bergman et al. (2019) for ECMWF System 4, and the main result was that the skill of predicting landfalls was lower than the basin-wide activity. Marginal skill was found for landfall on the Atlantic coast, while the forecasts were not skilful for landfalls around the Gulf of Mexico.
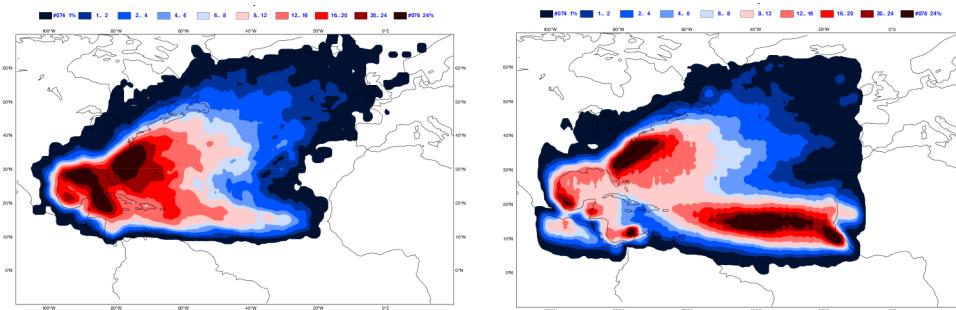


*Figure 28: Climatology of tropical storm density from observations (left) and Seas5 for month 2-6 from May start dates.*

To summarize the challenges for the seasonal forecasts, the important aspects to understand are the weak link between the local SST and TC activity, the suspicious trend in vertical wind shear, and the biased geographical distribution of TC activity. To temporarily minimise the effect of the erroneous trend the calibration period for TC activity in the seasonal forecast products was changed in 2021 from a fixed period (1993-2016) to the previous ten years.

## 5.9. Summary of challenges

The main findings from this section for all facets of predictive skill, related both to the intrinsic predictability of the atmosphere and the ECMWF system, are as follows:

**Analysis errors:**

- Initial errors and uncertainties of TC position are largest for weak TCs, where the centre is diffuse and difficult to determine from satellite observations. This applies to both ECMWF analyses and forecasters' Best Track estimates of position.

- Initial errors and uncertainties of TC intensity are largest for strong TCs, as the data assimilation struggles to replicate the convective-scale and mesoscale processes central to intense TCs. Further discussion of data assimilation in TCs will be provided in Section 6.

**Position (track) predictions:**

- Forecast errors of position are substantially larger for initially weak TCs than for strong TCs on average, especially in the short range.

- ECMWF forecasts have a long-standing systematic error with the propagation speed being too slow. There are indications that the main contribution to the error is during the propagation into the mid-latitudes, which could impact the extra-tropical transitions. This bias will be further discussed in Section 7.

- ECMWF forecasts of westward-moving TCs, especially weak TCs, often drift anomalously northward.

- ECMWF forecasts of TCs with a more northward component often drift anomalously westward.

- There can be run-to-run inconsistency ("jumpiness") in HRES and ensemble predictions.

**Pmin predictions:**

- Forecast errors are generally lowest for initially weak TCs, although there is a positive bias in Pmin forecasts which indicates that intensification rate is being under-predicted (too slow).

- Short-range forecast errors are generally highest for initially strong TCs, with a large positive bias. Later on, this bias for the initially strong cases changes sign as the weakening rate is under-predicted (too slow).

- Intensification in larger-scale TCs is captured better than intensification in smaller TCs.

**Vmax predictions:**

- There is a strong negative bias of Vmax, especially for initially strong TCs in the short-range. This is expected, as the definition of Vmax requires the simulation of convective-scale processes (O(1 km)).

- The under-prediction of the wind speed also affects the wind radii prediction (see Section 4).

**Wind-pressure (Vmax vs Pmin) predictions:**

- The wind-pressure relation is underestimated (see Vmax point above), and the underestimation is larger in the 18 km resolution than in 9 km resolution. The relation will be further discussed in Section 7.

**Genesis predictions:**

- There is a large variability in probabilistic predictions of genesis, and hence predictability, due to the variety of genesis mechanisms.

- For stronger, solitary African Easterly Waves, the pathway to genesis, and therefore the probabilistic predictions, are more straightforward and evident even over a week before genesis.

- For weaker African Easterly Waves, or wave/ITCZ/low pressure interactions, the pathway to genesis is more complex, and the probabilities of genesis are often very low less than three days prior to genesis. It is usually not straightforward to discriminate between developers and non-developers for these systems until one-two days prior to genesis.

- For non-developers (African Easterly Waves that do not develop into TCs), the probabilities of genesis do not appear to be unrealistically high.

- The precise timing of genesis is a challenge for the operational forecaster, and for the model, given the weak and disorganized structure of the TC at the time of genesis.

   **Extratropical transition predictions:**

- Extratropical transitions are known to insert uncertainties in the mid-latitude flow. However, it is not fully understood whether the TC is the main source of the uncertainty, or if it acts as an amplifier of existing uncertainties from the upstream mid-latitude waveguide, or both.

- The effect on the slow propagation speed bias on extratropical transitions needs to be further understood.

- The effect from the intensity prediction to TC outflow and further on the extratropical transition needs to be explored.

   **Decay predictions:**

- Decay is often under-predicted, due to the model maintaining too robust a TC as it moves towards the extratropics and encounters cooler water, stronger wind shear, and drier air.

   **Extended-range predictions of TC activity:**

- Large biases in the model climatology of TCs were identified, especially between sub-basins in the Atlantic where the model produced too many TCs west of the coast of Africa and too few in the Gulf of Mexico.

- For the activity in the Atlantic, the model underestimated the link from the SST in the western tropical Pacific (related to ENSO) and the link from local SST in the Atlantic.

- The erroneous trend in SEAS5 wind shear over the northern tropical Atlantic need to be further understood, for example in the relation to global warming patterns in the model and reanalysis.

This section summarizes ECMWF's key challenges in TC prediction, for several aspects of TCs including their formation, motion, intensity, decay, and overall activity. These challenges arise in part due to the breakdown of intrinsic predictability of relevant atmospheric processes on the relevant timescales, and to limitations in the data assimilation (including observational availability) and modelling. In Sections 6 and 7, progress and challenges in ECMWF's data assimilation and modelling respectively are presented in detail, in the context of TC analysis and prediction.

# 6.    Data assimilation

## 6.1.    Introduction

As for any other weather event, the accuracy of TC forecasts is dependent on the accuracy of the initial conditions (Chen et al., 2019). The skill of a TC forecast is sensitive to the initial conditions in both the TC itself and the environmental dynamic and thermodynamic conditions that influence the track and structure (and hence the intensity).

Some of the fundamental challenges in creating accurate initial conditions of the TC structure are connected to the presence of sharp spatial gradients in the meteorological variables, together with the scarcity of observations inside the TC and of its active environment. Available in-situ observations (from buoys, aircraft, and ships) are limited to a few surface pressure measurements. Dropsondes

from aircraft reconnaissance and surveillance flights are only typically available in the Atlantic basin. It is therefore important to assimilate all observations, direct and remote, to define the best initialisation of TCs and their environment. The evolution of the observing system and the continuous improvement of data assimilation methods are key in improving TC analyses and therefore forecasts. Substantial efforts at ECMWF and other agencies are continuously dedicated to improving data assimilation methodologies, quality control and data selection. In this context, observing system experiments are useful for the following reasons: (i) to identify existing data sources with more impact and for which additional resources need to be allocated to improve their usage; and (ii) to promote the need for extra sources of data to be assimilated.

In contrast to many other NWP centres, ECMWF does not apply any specific data assimilation strategy targeted for TCs. For example, the Met Office assimilates the Best Track estimates as surface pressure observations and JMA applies a similar strategy. In the regional hurricane models, more attention is paid to the initialisation of the TCs. COAMPS-TC applies a special vortex initialisation scheme (Komaromi et al., 2021). For HWRF, an initialisation scheme for TCs is used. Until recently, increments from the HWRF assimilation in the TC core were masked, and the initialisation solely relied on vortex initialisation. Since 2019, instead a wave-number filtering of increments is applied (Christophersen et al., 2021). Importantly, HWRF now also assimilates flight-level data, airborne Doppler radar data, and dropsonde data from reconnaissance flights, and their TC forecasts have accordingly improved substantially in recent years (Cangialosi et al., 2020; Zawislak et al., 2021). The open question here is whether ECMWF should adopt any of these initialisation methods.

The observation systems used at ECMWF, described in Section 2, are continuously evolving, due to the launching and decommissioning of satellites and instruments. As instruments become older, the changes in the bias structures also need attention. Furthermore, the conventional observation system is non-stationary, and continuous monitoring is needed to minimise the risks in the data assimilation of problematic observations. For satellite observations, there is an ongoing move from only assimilating data in clear-sky conditions to all-sky (Geer et al., 2018). Another ongoing change is the increased availability of GNSS-RO data, further enhanced by the recent COSMIC-2 satellites. Since January 2020, ECMWF is also assimilating wind observations from the exploratory mission Aeolus. However, due to computational limitations together with the risk of correlated observation errors, satellite data are usually thinned before assimilation. In features with sharp gradients like TCs, it is necessary to ask if more information could be extracted from the datasets.

In recent years, coupled ocean-atmosphere assimilation has gained momentum. Given that some of the fastest change rates in SST occur under TCs, we expect that coupled data assimilation will play an important role in TC structure and intensity change prediction, and the representation of the ocean surface and subsurface currents, temperature, and salinity.

For this report, we have completed a range of parallel observation impact and coupled data assimilation experiments to explore the impact from recent and ongoing activities. The experiments were performed for the period of 15 August - 21 September 2020, to cover the most active period of the Atlantic hurricane season and a few key typhoons in the western Pacific (listed in Appendix B). For forecasts initialised every 12 hours, it resulted in more than 200 verified cases at the initialisation time and close to 100 cases in 3-day forecasts. We note that ECMWF regularly runs observing system

experiments (OSE); however, these are usually with reduced resolution (e.g., Lawrence et al., 2019). For TCs, we expect the resolution to have a significant impact, and so the experiments for this report are run at the current operational resolution of approximately 9 km grid spacing. To reduce the cost, the experiments only include the 12-hour long-window cycling (LWDA analyses), and all experiments use the background errors derived from the operational EDA.

This section is organised as follows. In Section 6.2, we will give a summary of progress at ECMWF in 4D-var and ensemble of data assimilation. Following this, we examine the observational coverage in and around TC Laura (2020) in Section 6.3, and the impact of individual observation types on TC performance in Sections 6.4-6.6. We then explore the possibility to assimilate the Best Track estimates in Section 6.7. The impact of variants of coupled data assimilation is reported in Section 6.8. In Section 6.9, we review the sampling of initial uncertainties around TCs in the ECMWF ensemble. Finally, a summary of the key points is provided in Section 6.10.

## 6.2. Progress and challenges in 4D-Var and ensemble data assimilation

During the past few years, incremental upgrades to the ECMWF system have gradually improved the realism of TC structure and intensity predictions and allowed better usage of available high-quality observations. Despite these advances, the initialisation of TCs remains challenging. A few challenges are described in this sub-section.

Since 2011, ECMWF has been using an ensemble of data assimilations (EDA) to estimate background errors and provide initial perturbations to the ensemble forecasts (Bonavita et al., 2016). Since 2016 the resolution of the EDA is TCo639 (Holm et al., 2015), which improved the intensity of TCs, and since 2019 the system contains 51 ensemble members (Lang et al., 2019). An example of the evolution of the position and Pmin in the HRES analysis and the first ten EDA members is illustrated in Figure 29 for TC Laura. During the first phase when the Laura was a weak Tropical Storm, relatively large spread in the analysis position is evident. This example is consistent with the increased initial position errors for weak TCs reported in Section 5. An additional uncertainty for this case is the effect from the terrain to find a minimum in the pressure field, resulting in large errors while passing Puerto Rico, Hispaniola, and Cuba. Once Laura started to intensify into a hurricane west of Cuba, the position among the members converged. For the intensity (both Pmin and Vmax), some differences among the analysis members are evident during the most intense phase (Figure 30). The difference in resolution between HRES-LWDA (long-window analysis) and EDA is evident, with a TC that is much more intense in HRES-LWDA but still weaker than the Best Track estimate.
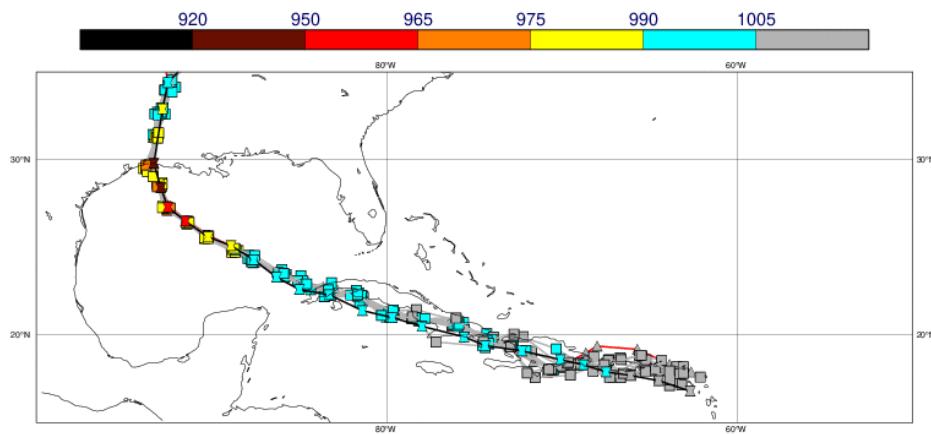
*Figure 29: Position and intensity for TC Laura in concatenated HRES LWDA analyses (red line, triangle), the 10 first EDA members (grey lines, square) and Best Track (black, hourglass). The colour of the symbol shows Pmin.*
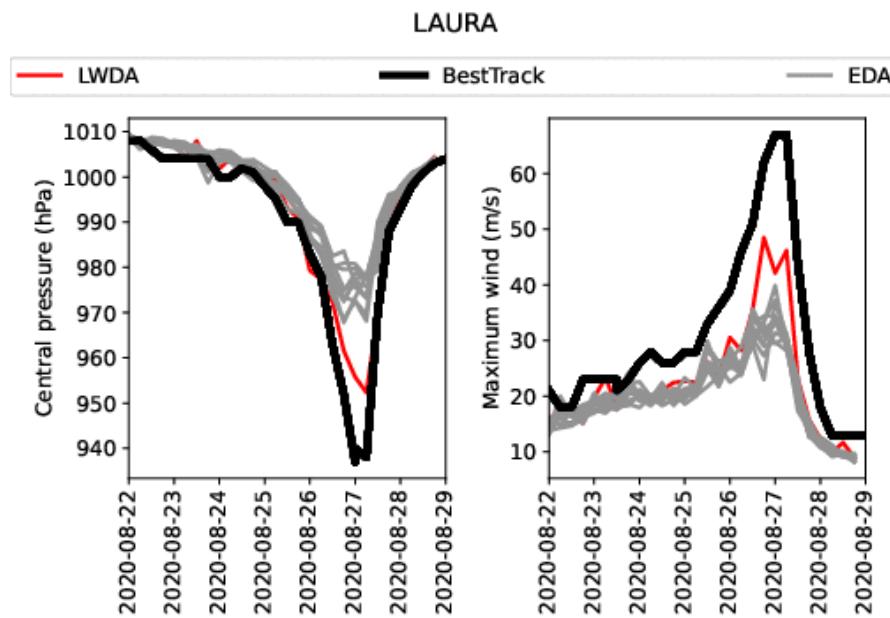


*Figure 30: Intensity in terms of Pmin (left) and Vmax (right) for TC Laura in HRES LWDA analysis (red), the 10 first EDA members (grey) and BestTrack (black). The colour of the symbol shows Pmin.*

Even if the non-linear model used in the 4D-Var data assimilation update trajectories (outer-loop trajectories) is run with 9 km resolution, the increments are calculated at a lower resolution (~50 km in the last minimisation). This implies that the sharp gradients and detailed structures present in the analysed TC in the vicinity of its centre are largely the product of the forecast model. The incremental 4D-Var approach aims to adjust the large scale whilst retaining finer scale gradients, but if the TC is mislocated in the background the increments may smooth the gradients. Another candidate for the

smoothing here is that the background error covariance structures may be broader than the real covariance structures in the core of the TCs.

Another consequence of sharp gradients near TCs is that small discrepancies between the observed and predicted structure can result in very large discrepancies between available observations and their model counterparts. As discussed in Bonavita et al. (2017), the interplay between extreme observations, large background forecast errors and linear assumptions in incremental 4D-Var can be problematic near TCs. With increases in the resolution in the EDA, the intensity of the TCs is more realistic (Holm et al., 2015). This leads to larger (and more realistic) ensemble spread/background errors. However, this also makes the analysis update more exposed to extreme and unrepresentative observations, e.g., of wind speed from dropsondes in the eyewall of a TC. The adaptive first guess quality control (as discussed in Bonavita et al., 2017) mitigated the issue, in particular for dropsonde winds, but occasional problems still occur. Further improvements need to be explored (e.g., improved QC, activation of VarQC in the first minimisation, etc.). Discrepancies between the predicted and observed position can also lead to wrong selection of observations to be used by the analysis (e.g., selection of the wrong ambiguity in the case of scatterometer observations, as discussed in Section 2).

Another challenge was the issue of multiple local minima in the analysis and first few hours of the forecast of surface pressure in TCs, caused by the EDA and the increased model resolution (Bonavita et al., 2017). This issue was mostly cured by a reduction of the resolution of the EDA-derived background errors used in 4D-Var and a non-homogeneous noise filtering technique together with an adaptive, stricter quality control technique applied to dropsonde wind observations.

An occasional issue is the assimilation of SYNOP/BUOY observations of surface pressure. As there are normally only small day-to-day fluctuations in surface pressure in the tropics, faulty (or misplaced) observations could go unnoticed or cause the bias correction algorithm to converge to erroneous values. Once the TC approaches the observed position, the faulty surface pressure observation starts affecting the assimilation of the TC with undesirable consequences (typically filling up the TC). One example was for the TC Leslie (2017) east of Australia, where a faulty position of a ship observation had a severe impact on the data assimilation. Another example is TC Surigae (2021) where a SYNOP station had the wrong position. One forecast just before the cyclone reached the location of the mis-placed SYNOP (17 April 12UTC) and just after (18 April 00UTC) is shown in Figure 31. For the latter forecast, the intensity is much reduced, and the track forecast had large errors. In the experiment without the SYNOP station, both the intensity and track are much better. In this type of case, the observations were given low weight in the quality control but were still allowed to have a significant impact during the first minimisation in 4D-Var. More generally, it is an open question whether tighter QC should be applied in the first 4D-Var minimisation, and it should be investigated.
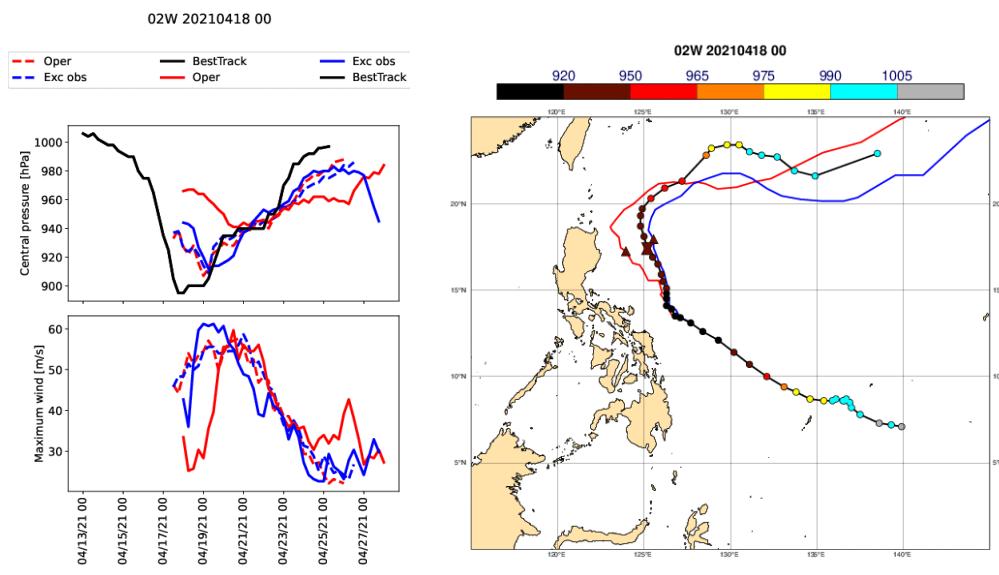
*Figure 31: Pmin (left-top) and Vmax (left-bottom) forecasts from 17 April 2021 12UTC (dashed) and 18 April 2021 00UTC (solid), and track forecast from 18 April 00UTC (right) for TC Surigae. Operational forecasts (red), experiment without misplaced SYNOP (blue) and Best Track (black). Symbols in right panel indicate Pmin for Best Track every 6ᵗʰ hours and at 21 April 00UTC in experiments (triangles) and BestTrack (hourglass).*

Another problem with the assimilation of surface pressure is the lack of stiffness of the current adaptive bias correction algorithm, which can react too quickly to large departures. If a cyclone passes a buoy with a too weak intensity in the first guess forecast, a bias correction will be applied to the buoy in the next assimilation cycle, and it can then have an impact for several days. A potentially more robust VarBC approach to the surface bias correction problem is currently being assessed.

## 6.3. Observations and their impact on Hurricane Laura

Figure 32 shows the observation coverage from one assimilation cycle over Hurricane Laura (27 August 00UTC) for a selection of observation platforms described in Section 2. The analysis time coincided with the most intense phase of Laura just before landfall. It was a busy period with reconnaissance flights as seen in the coverage of dropsondes (Figure 32(a)). For ASCAT, both Metop-B and Metop-C had direct overpasses of the TC (Figure 32(b)). With the operational thinning of 100 km, it resulted in just 6 wind vectors inside the 1000hPa isobar of the TC. Figure 32(c) and Figure 32(d) respectively show AMSU-A and microwave humidity sounders (MHS) for one tropospheric channel each. Operationally, AMSU-A is assimilated in clear-sky mode while MHS is in all-sky. As the contributing satellites are almost the same in the plots (NOAA-15 has a functioning AMSU-A but not an MHS instrument), the impact on the coverage of all-sky assimilation is clearly seen with cloud-masked observations for AMSU-A close to Laura (only two observations inside 1000 hPa). However, also for MHS the number of observations is relatively sparse due to thinning to approximately 130 km. For radiances from geostationary satellites, ECMWF currently only assimilate observations in clear-sky mode. This results in very few used observations near TCs (Figure 32(e)). From the same instruments Atmospheric Motion Vectors (AMV) are also derived from cloud and water-vapour

features. However, even if TCs contains a lot of clouds to trace, the number of used observations in the case of Laura is very few (none within the 1000hPa isobar, Figure 32(f), due to cloud tops above the limit for GOES-16 AMV and/or lack of contrasts in cloud signatures (see Section 2.4). In the bottom row GNSS-RO (Figure 32(g)) and Aeolus (Figure 32(h)) are presented. For GNSS-RO there are a few observations present in the outskirts of the TC. Aeolus measures the wind in one line tilted under the satellite, and one such passage was close but not immediately over the TC. All these panels together show that there are very different types of coverage from different observation platforms, with only a limited number directly inside or over the TC.
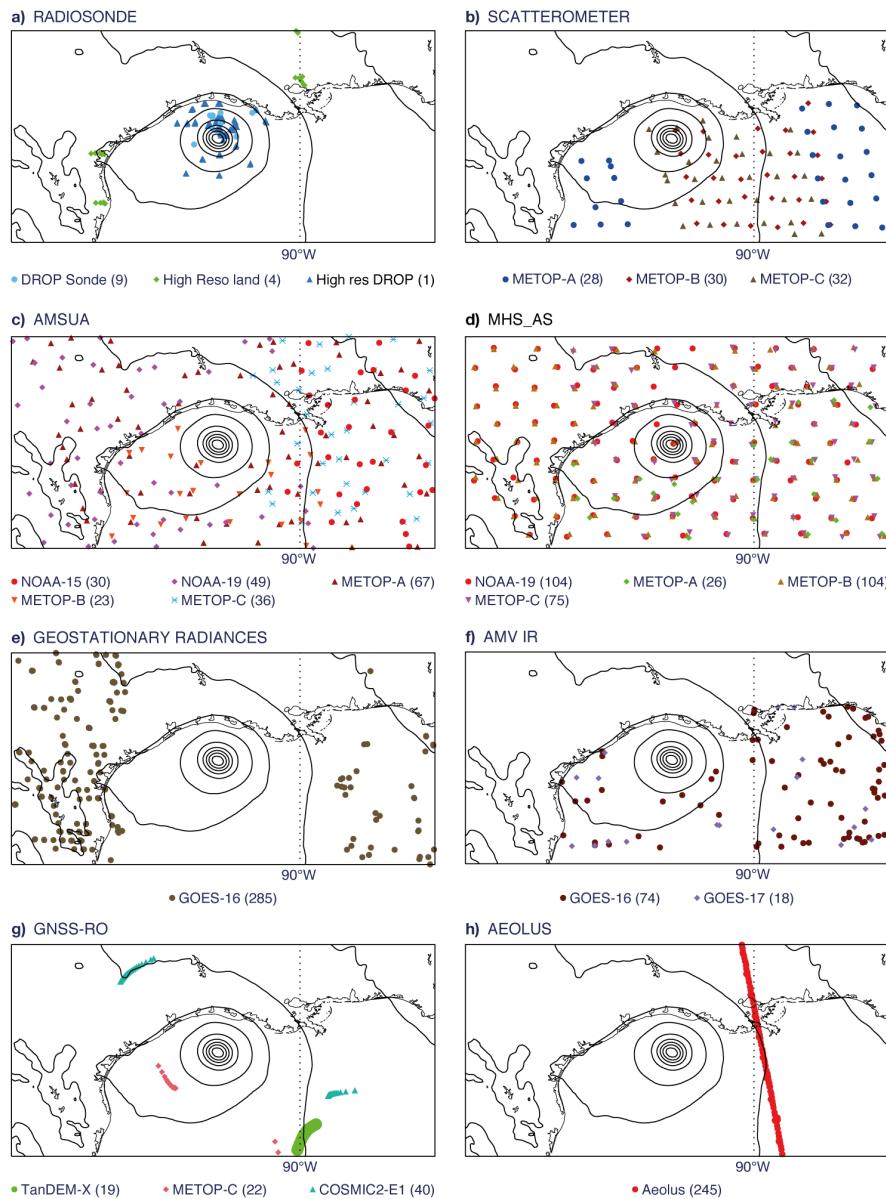
*Figure 32: Observation coverage for the operational (6-hour window) assimilation on 27 August 00UTC for TC Laura for (a) radiosondes and dropsondes, (b) ASCAT, (c) AMSU-A channel 5, (d) microwave humidity sounder channel 4 in all-sky mode, (e) IR radiances from geostationary, (f) atmospheric motion vectors from geostationary, (g) GNSS-RO and (h) Aeolus.*

To illustrate the impact of assimilating the dropsonde observations, Figure 33 shows the intensity forecasts for TC Laura initialised at four consecutive times: 25 August 12UTC and 26 August 00UTC just before the RI (left), and 26 August 12UTC and 27 August 00UTC during the intense phase of Laura (right). The plots include forecasts initialised from the operational configuration of LWDA (blue) and an experiment without dropsonde data (green). For the forecasts initialised just before the intensification, the intensity was best predicted by the operational configuration, indicating that the dropsondes provided useful information. This result is in line with a similar experiment reported in Magnusson et al. (2019a). However, for the initialisation on 27 August 00UTC, the intensity is

degraded in the LWDA control analysis, with Pmin changing from 941hPa in the 12 h background forecast to 951hPa in the analysis, while the Best Track estimate was 940hPa. At the same time, the experiment without dropsondes only gave a slight degradation. The dropsonde impact will be further discussed later in this section.

Degradation of the intensity for the analysis during intense phases been noted on several occasions by analysts at ECMWF. One could therefore ask the question how the intensity would look if no observations were assimilated in the vicinity of the TC. A parallel assimilation experiment was run for Laura between 22-27 Aug, during which all observations within a 4 x 4-degree box centred on Laura were withheld (Figure 33, orange line). It is evident that the denial of these data local to the moving TC substantially degraded the intensity forecasts. For the 25 August 12UTC and 26 August 00UTC initialisations, the track forecast was also worse (not shown). These results illustrate for this TC that the observations close to the TC that are assimilated are beneficial to the forecasts, but that there are potential problems during the most intense phase.
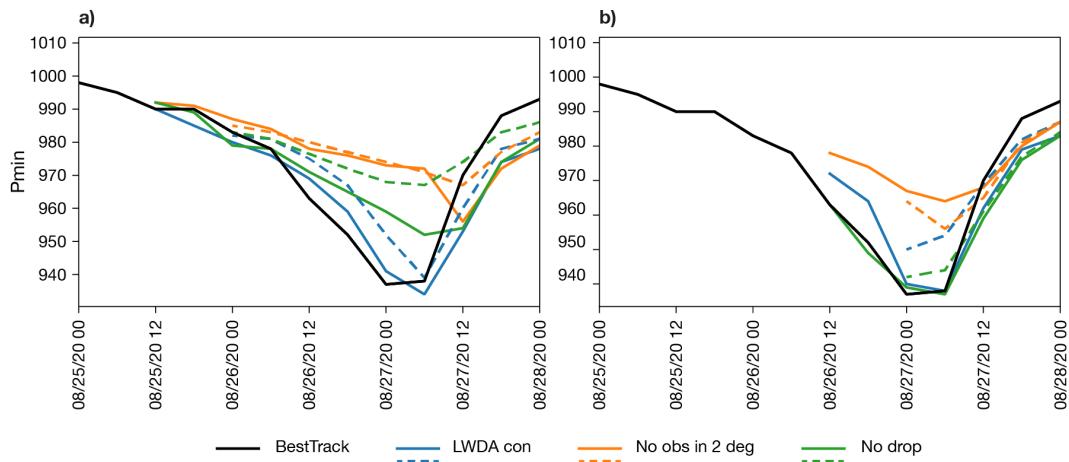


*Figure 33: Pmin in forecasts for Hurricane Laura in experiments initialised from LWDA control experiment (blue), assimilation without observations inside a 4 x 4-degree box around the cyclone (orange) and without dropsonde observations (green). Forecasts initialised 25 Aug 12UTC (solid, left), 26 Aug 00UTC (left, dashed) and 26 Aug 12UTC (right, solid) and 27 Aug 00UTC (right, dashed).*

## 6.4. Impact of assimilating dropsondes from reconnaissance flights

As introduced in Section 2, dropsondes are regularly deployed from reconnaissance missions, both inside the core of the TC and the surrounding environment. A case example of the impact was presented in Section 6.3, and here we briefly provide an overview of progress at ECMWF and results from the 2020 season.

ECMWF assimilates temperature, wind and humidity from dropsondes. In recent years, several experiments at ECMWF have aimed to quantify the impact from dropsondes on TC forecasts. While Magnusson et al. (2019a) reported a positive impact of dropsondes during the intensification phases of TC Harvey (2017) and Maria (2017), it was difficult to find any statistically significant improvement averaged over the full 2017 season. Since 2019, some of the observations are reported in the BUFR

format that contains information about the position of each reading and makes it possible to account for the drift (Ingleby et al., 2020). At the same time, the impact of accounting for the drift showed a slight positive impact on the intensity forecast.

Despite the difficulties due to COVID-19, the 2020 Atlantic hurricane season resulted in a record number of reconnaissance missions. Figure 34 provides results from an experiment in which all dropsonde observations were withheld from the assimilation during the 2020 test period. Although the impact on position error is neutral (not shown), an improvement to the average Pmin absolute error is found for the experiment with dropsonde observations, especially for a small number of intensifying TCs of initially moderate strength (50-64 kt, e.g., Laura prior to 25 August 12UTC). This is consistent with the results of Magnusson et al. (2019a) where positive impacts were found for two rapidly intensifying TCs (Harvey and Maria in 2017).  Positive impacts were usually not found for initially very weak TCs (< 50 kt), or TCs of hurricane strength (>64 kt). Further research is needed into the corrections to the analysis fields due to the dropsondes, and whether this impact is carried over into the forecasts. One could also speculate that additional impact can be achieved in the future with higher resolution and sharper covariance structures in the data assimilation, which would permit analysis corrections that are closer to the observed data from the dropsondes.
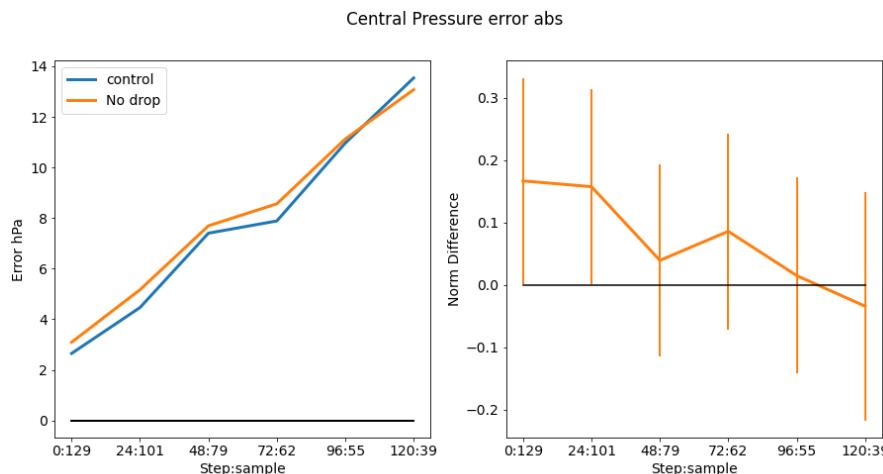


*Figure 34: Mean absolute errors for Pmin for experiment without dropsondes for cyclones in the Atlantic. Normalised difference to operational configuration in the right panel with 95% confidence interval plotted as error bars.*

## 6.5.    Impact of satellite observations: recent progress at ECMWF

As discussed in Section 2, ECMWF is using a range of different satellite observations in the data assimilation. In McNally et al. (2014), the impact of satellite data on the track for Sandy (2012) was discussed, and for that case the satellite data were found to be crucial for the prediction of Sandy's landfall over the United States east coast. Similar results for Irma (2017) were demonstrated in Magnusson et al. (2019a). In both experiments, large volumes of available satellite data were removed from the data assimilation system for a short period before the genesis of the TCs.

The impact of a selection of satellite observing systems on TC forecasts has also been assessed for a recent set of global observing system experiments (Bormann et al., 2019). While these experiments were run at a lower spatial resolution, the observation usage reflects that of the operational system, including the quality control and data thinning applied to satellite data. The experiments covered a period of eight months over two seasons (1 June - 30 Sept 2016, and 1 Dec 2017 - 31 March 2018), and TC statistics were calculated over all basins. Several aspects are noteworthy in the impact on position and intensity error (Figure 35): The experiment denying all microwave radiances shows the clearest impact, with statistically significant increase in error up to day 3 without these observations. This observing system was also found to have one of the largest impacts on general headline scores in Bormann et al. (2019). It is notable that MW radiances provide regular observations near the core of the TC through the all-sky assimilation of humidity-sensitive MW radiances (cf. Figure 32(d)). While it is not clear that the impact is due to the near-core observations, it provides a plausible mechanism, and the impact of near-core observations has also been noted in assessments of FSOI statistics (not shown). Other observing systems also show statistically significant benefits for specific forecast ranges, with some of the most consistent improvements for position errors found for AMVs. We note that scatterometer data also provide near-core observations, and while they were not considered in Bormann et al. (2019), experiments with scatterometer data together with microwave data are presented for 2020 in Section 6.6.

Another aspect is also apparent from these experiments: the ECMWF system shows robustness to denials of single observing systems for TC forecasts. Even removing sizeable contributions to the global observing system such as all microwave observations translates to a loss of only around 6 h for TC forecasting. In contrast, the overall strong benefit of all satellite data combined has previously been demonstrated, for instance, in McNally et al. (2014). The results for single observing systems shown here also highlight the challenge of verifying improvements in TCs.
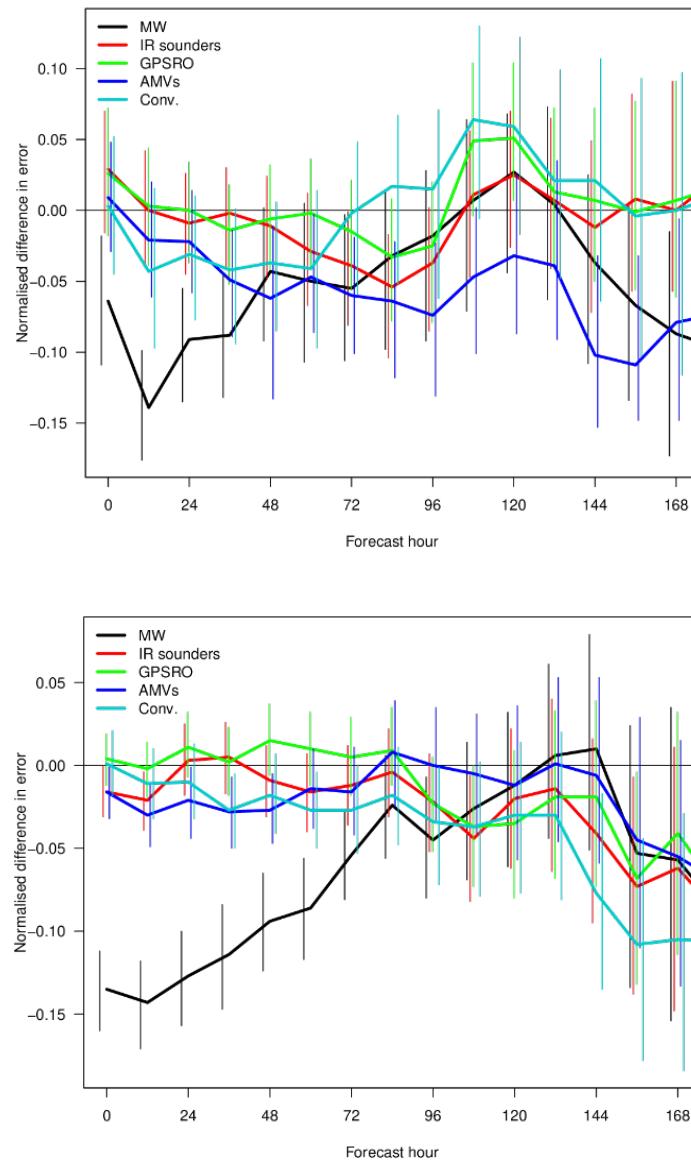
*Figure 35: Normalised difference(control minus experiment) in the position (top) and absolute intensity error (bottom) from the observing system experiments considered in Bormann et al. (2019). These measure the impact from denying selected observations from an otherwise full observing system. Considered are all satellite MW radiances (MW, black), all hyperspectral IR sounders (red), bending angles from GPS radio occultation (GPSRO, green), AMVs (blue), and in-situ observations, encompassing all aircraft data, radio- and drop-sondes, synops, etc (Conv., cyan). Statistics are based on experimentation over 8 months (1 June - 30 Sept 2016, and 1 Dec 2017 - 31 March 2018), with the number of TC forecast cases considered ranging from 700 in the short range to over 110 at day 7. Error bars indicate statistical significance at the 95 % confidence level.*

## 6.6.    Impact of satellite observations: 2020 experiments

Several new observing system experiments with satellite observations were performed for the special 37-day period in Autumn 2020. A control experiment ("LWDA Con") based on the operational observation usage is used as a reference for all the observing system experiments in this section and paper. In Figure 36, the denial of the following observation types which were assimilated operationally in 2020 is evaluated: (i) satellite wind observations from Aeolus; and (ii) GNSS-RO data from COSMIC-2. To put these results in perspective, these impacts are compared against an experiment in which all microwave observations are denied, similar to the experiment of Bormann et al. (2019) illustrated above in Figure 35. As we remove observations in these experiments, we expect the error to increase for observation types with large impact. A "positive" impact for the denial experiments refer to increased errors when the observations are removed from the experiment. In addition to the denial experiments, Figure 36 also includes an experiment with the expected improvements from all-sky assimilation of AMSU-A temperature sounders upon the operational configuration. This change to the all-sky assimilation is planned for implementation in Cycle 47r3 in late 2021. However, as all the individual systems by themselves are a small part of the full observation set, one should expect relatively small impacts.  While the focus in this section is on TC predictions, results for the impact over the full tropics are provided in Appendix C. As the results in Appendix C are based on verification against the operational analysis, the results are not straightforward to interpret for short lead times. However, one can note that all instruments discussed here have a statistically significant positive impact for 700hPa winds in the tropics for 3-day forecast, as an example.

As expected, the largest impacts on the TC position and Pmin are found when all microwave observations are removed (Figure 36) and the results are in line with the low-resolution experiment over a longer period presented in Figure 35. Since the humidity sounders and microwave imagers are assimilated in all-sky mode, including near the TC (as seen in Figure 32), this could be one reason for the positive impact on Pmin. An examination of the impact of adding all-sky assimilation of AMSU-A temperature sounders reveals a slight improvement in the position error for one-day predictions, and no statistically significant benefit for the Pmin error. However, a similar experiment during a longer period of the 2019 TC season at lower resolution provided a positive impact on both position and intensity forecasts (Duncan et al., 2021).

For COSMIC-2, statistically significant improvements at the 95% level are found for Pmin for the first two days and still on the positive side out to five days (Figure 36, right). However, the magnitude of the average improvement compared with microwave is small, while the impact on position forecasts is neutral (Figure 36, left).

For Aeolus, the results are neutral for both position and Pmin. However, as seen in Figure 32, with the line measurement from one instrument, only a few crossings over a centre of a TC are expected in a 40-day sample, which makes it difficult to reach a statistically robust impact. For TC Teddy, the system was fortunate to achieve three straight passages over the TC. This provides an opportunity for a further evaluation of this case in the future.
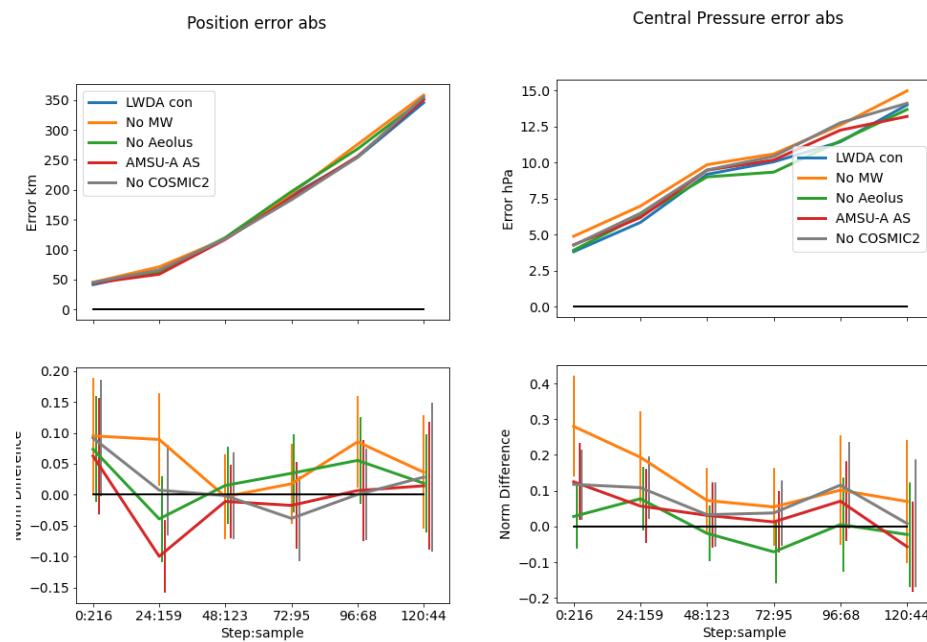
*Figure 36: Mean absolute errors for position (left) and Pmin (right) for different satellite observation system experiments (experiment minus control). Normalised difference to operational configuration in the bottom panels with 95% confidence interval plotted as error bars.*

As discussed in Section 2, surface winds from scatterometer instruments have been assimilated operationally for many years. Since 2020, ECMWF has been assimilating ASCAT winds from three Metop satellites, which are provided with an original grid spacing of 25 km and thinned to 100 km. An observation error of 1.5 m/s is assigned to the wind components. While providing high quality observations, sometimes a wrong direction is presented. This can be due to observation errors, the retrieval process or ambiguity removal issues, with the latter being the most common source of wind direction problems. Since the ambiguity removal process is based on the short-range forecast (background) wind fields, in case of TCs, this issue usually happens when we have large first guess errors in wind direction. An example of erroneous wind vectors presented to the 4D-Var is shown in Figure 30 for TC Paulette where a couple of wrong directions can be found north-east of the TC. However, these wind vectors are the closest to the background, and the alternative solutions do not appear to be better. Most likely, in this specific case, the problem could be due to either observation error in an area with a quite chaotic ocean surface, or an issue in the retrieval process. The incorrect selections lead to very large differences between the observations and background forecast, and the variational quality control (VarQC, Andersson and Jarvinen, 1999) then correctly assigns zero weight to these observations, as seen for this case in Figure 38, where the problematic vectors are not among the active observations). However, we also see that other (correct) observations of winds above 20 m/s were rejected. To reduce the number of rejected ASCAT observations the usage of Huber Norm instead of the VarQC has undergone some preliminary investigations (De Chiara et al., 2018) and this could be a promising line of future research.
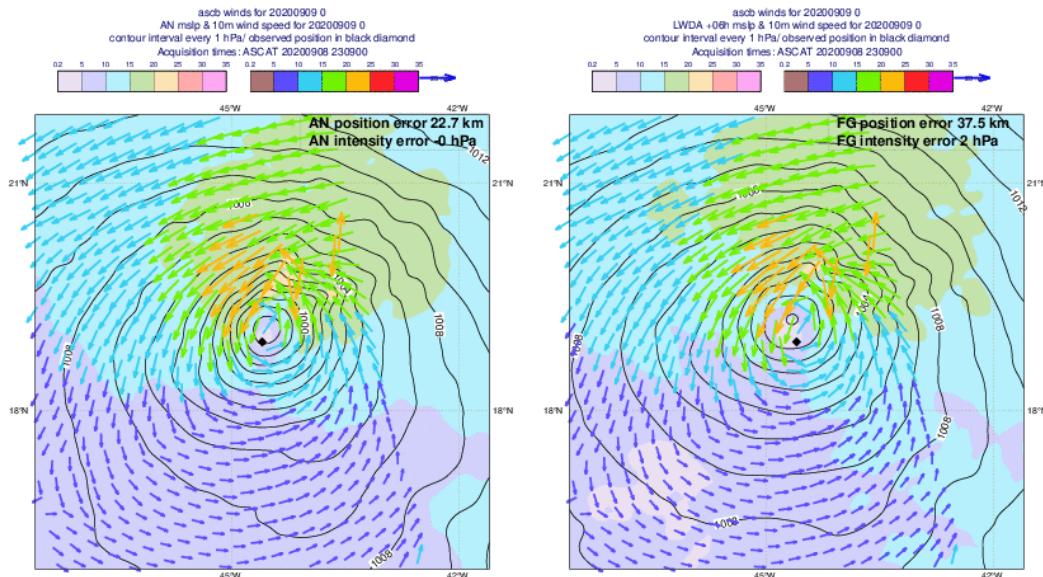
*Figure 37: Analysis (left) and first-guess (right) of MSLP (contour) and wind speed in m/s (shading), and **all** ASCAT observations (wind vectors) coloured by wind speed for TC Paulette on 9 September 00UTC in the operational analysis. Best Track position indicated as diamond symbol.*
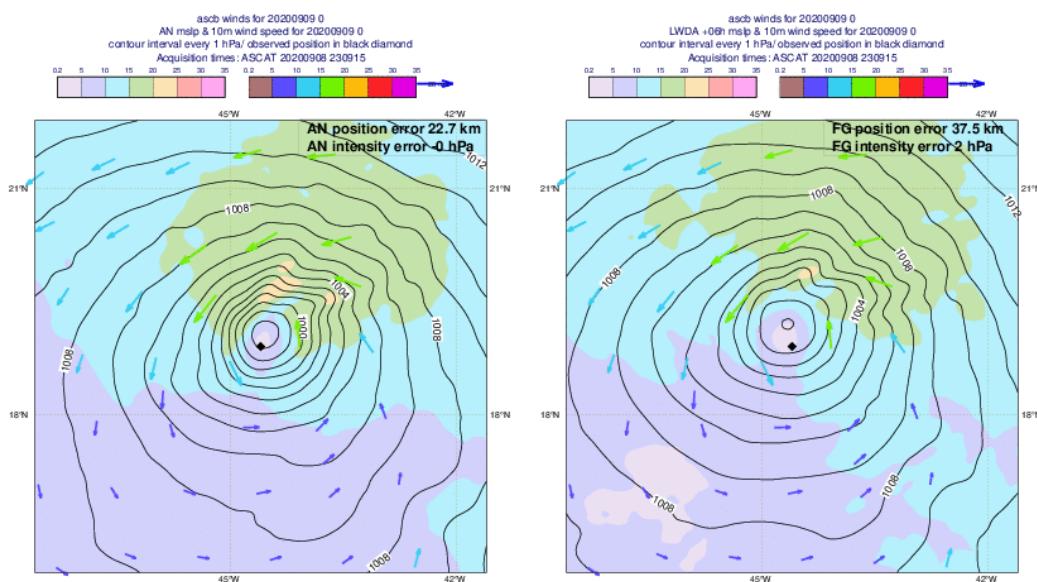


*Figure 38: Same as Figure 37 but for **active** observations.*

Due to the challenges (ambiguity selection, thinning, wind saturation) in using ASCAT winds close to TCs, one can ask the question of whether they are beneficial in TC analysis and forecasting. In operations, ASCAT observations are thinned, with one in every four observations assimilated along and across track, as seen when comparing Figure 37 (all observations) and Figure 38 (active observations). This results in a final horizontal sampling of 100 km. To explore the sensitivity to the assimilation of ASCAT winds, two experiments with a different thinning of the observations have been completed. As the reduced thinning increases the correlation of observation errors, this is compensated for by increasing the observation errors.

One experiment ("Scat Thin2") is thinning the observations by a factor of 2 (instead of 4 in operations) and uses an observation error of 2.25 m/s. The other experiment ("Scat No thin") has no thinning and an observation error of 3.75 m/s. Note that increasing the observation error also leads to a relaxed quality control. Another experiment with all ASCAT winds removed was also conducted. The results from all experiments are presented in Figure 39 and Figure 40. First, the removal of all scatterometer observations is found to degrade the forecasts of Pmin out to 48 h (Figure 39, orange line). Next, while the experiment "Scat Thin2" only shows a slight (non-significant) reduction in Pmin error (Figure 39, green line), the biases in Pmin are reduced (Figure 40, green line), strengthening analyses and forecasts of hurricane-force TCs out to two days (not shown). This is more clearly illustrated in the scatterplot in Figure 41, where the analysis Pmin values for deep TCs is usually lower than in the experiment analysis ("Scat Thin2") than in the LWDA Control analysis ("LWDA con"). The mechanisms behind this result will be further explored. Similar results, albeit more modest, are found in the experiment with no thinning and higher observation errors ("Scat No thin").
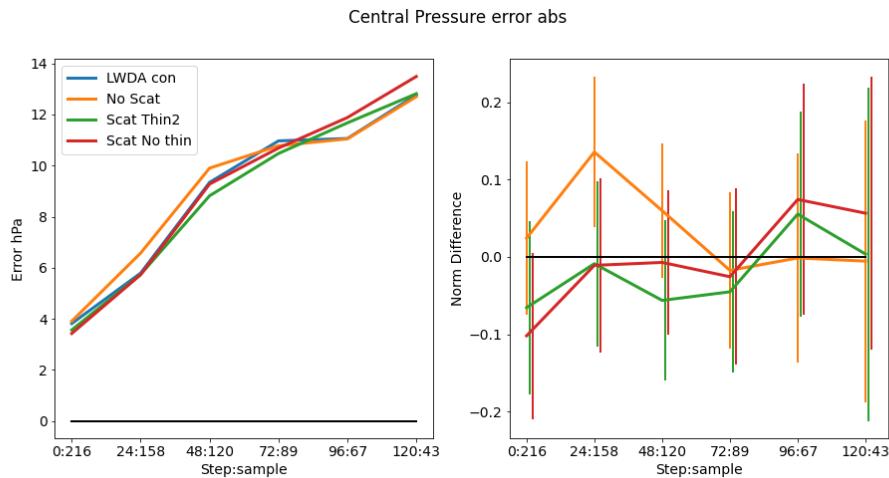


*Figure 39: Mean absolute error for Pmin for experiment with different ASCAT configurations. Normalised difference to operational configuration in the right panel with 95% confidence interval plotted as error bars.*

Central Pressure error mean
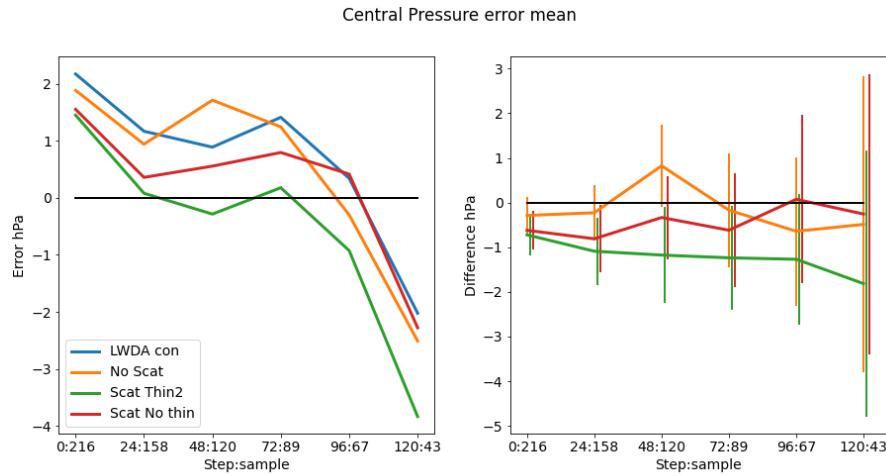
*Figure 40: Mean error for Pmin for experiment with different ASCAT configurations. Difference to operational configuration in the right panel with 95% confidence interval plotted as error bars.*
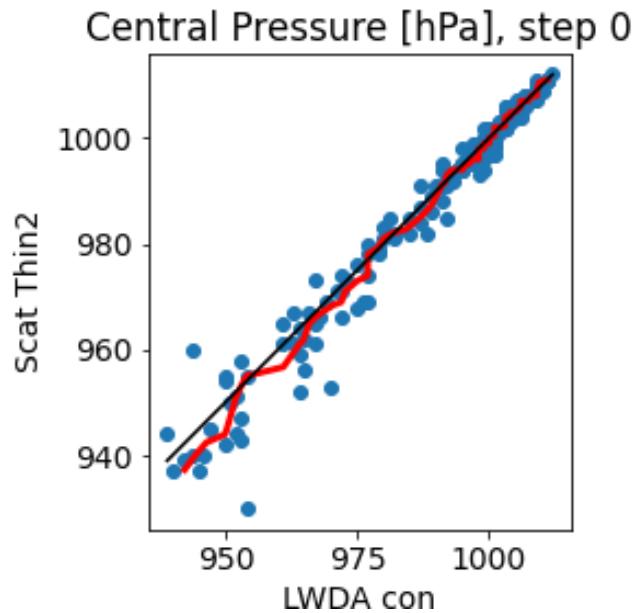


*Figure 41: Scatterplot of Pmin in control vs reduced thinning of ASCAT with a factor of 2 ("Scat Thin2"). Quantile-quantile matching is shown in red.*

## 6.7.    Assimilation of Best Track

Most of the other NWP centres make use of the information in the Best Track in the data assimilation or initialization. Some centres apply a vortex relocation scheme to move the TC according to the

estimated position. Other centres like the Met Office (Heming, 2016) infer the estimate as a surface observation of pressure in the data assimilation. A similar approach was also undertaken in the ERA5 back-extension to 1950. However, severe issues were detected during this approach in cases where the background forecast lacked a TC or was far from the truth. This led to the development and implementation of a safer Best Track observations assimilation for the pre-satellite era and a renewal of the production of the ERA5 back extension.

In the case of the Met Office, they interpolate the values between the six-hourly estimates to produce hourly observations to assimilate, to increase the impact. To test the impact on assimilating Best Track, we have conducted experiments that follow the operational approach at the Met Office. The Best Track position and intensity has been linearly interpolated from six-hourly (from "operational" Best Track) and three-hourly data (from IBTRACS files) into one-hourly sampling and is used as hourly surface pressure observations. There is also a potential difference in the TC life-length spanned by these two sources. The results are also compared with an experiment only assimilating the six-hourly estimations. Observation errors have been set to 3.5 hPa for Atlantic TCs (which are usually observed by both satellites and reconnaissance aircraft) and 6 hPa for TCs in the Pacific. For hourly interpolated values, the observation errors are inflated by 1.2 to account for interpolation errors. When the TC intensity is constant for more than 12 hours, the errors are inflated by a factor of 1.5.

Figure 42 illustrates the position and Pmin mean absolute errors for these three experiments. For the position error, all three experiments lead to an improvement of 3-5% for one to two-day forecast, but do not pass the 95% significance level. The magnitude is somewhat lower compared to the improvement found at the Met Office (Heming, 2016, table 3). For the Pmin, the impact on MAE is positive but non-significant, while the bias is significantly reduced at the analysis time for both experiments using 1-hourly interpolation (not shown).The positive impact for the bias is found in making strong TCs more intense at the analysis time, as seen in the scatterplot of Pmin values in Figure 43. The main contributions here are TC Maysak and TC Haishen in the north-western Pacific. While the impact is positive in the northwest Pacific, the experiment seems to degrade the forecasts in the Atlantic (not shown). One could speculate that the difference in intensity distribution, the estimate of observation error, or availability of dropsondes in the Atlantic may cause this difference. However, the variability from case to case is large and none of the results passed the 95% significance level.
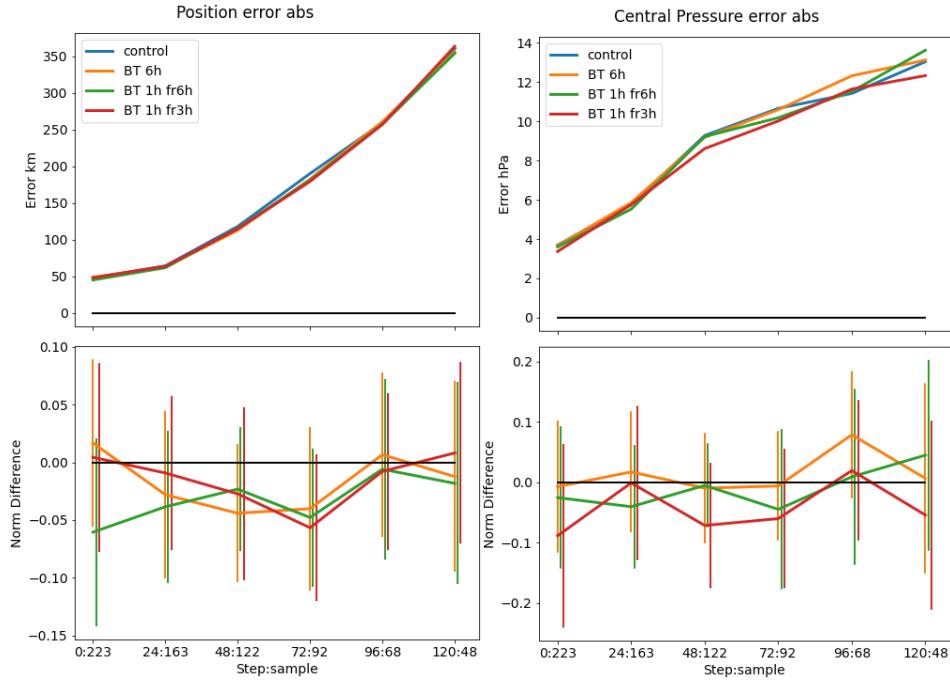
*Figure 42: Mean absolute errors for position (left) and Pmin (right) for experiments with assimilating BestTrack estimates. Normalised difference to operational configuration in the bottom panels with 95% confidence interval plotted as error bars.*
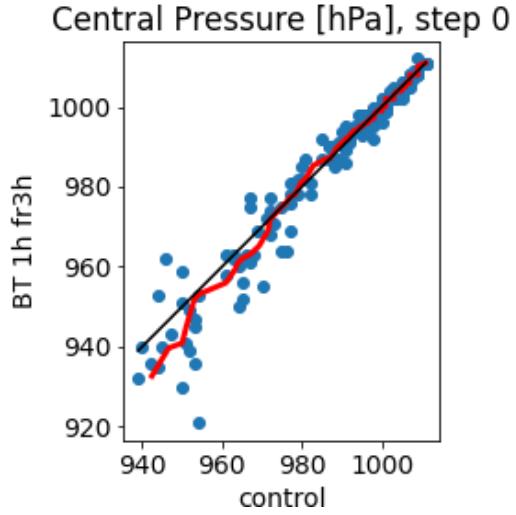


*Figure 43: Scatterplot of Pmin in control vs assimilation BT from 3h estimates. Quantile-quantile matching is shown in red.*

Applying this procedure poses several difficulties. One lies in the estimation of the observation error. This is expected to vary according to the availability of reconnaissance flights etc. (Torn and Snyder, 2012; Landsea and Franklin, 2013). The estimate of the observation error also has an impact on the

VarQC. If the Best Track observation is too far from the first guess forecast value, the observation will have very little or no impact due to the quality control. Another lies in the exposure to human errors and technical faults in the reporting chain, that can lead to assimilating erroneous values. As noted above, non-human observations can also suffer from difficulties with estimating observation errors, biases, and occasional technical faults. The variability in the results from the three different experiments also indicates a sensitivity to interpolation and selection of cyclones to include (pre- and post-stages of the TCs). All these aspects need a careful consideration before this type of observation is put into operations, which could be a time-consuming exercise.

## 6.8.    Ocean Aspects: Coupled Data Assimilation

An important aspect is the treatment of the SST, given that rapid upper-ocean cooling may occur in the wake of TCs. Since 2013 (ensemble) and 2018 (HRES), the ECMWF atmospheric forecasts have been coupled to a dynamical ocean model (see Section 7). Since 2019, the SST boundary conditions for the atmospheric analysis have been taken in the tropics from the ocean data assimilation system, providing a weakly coupled Atmosphere-Ocean data assimilation system (Browne et al., 2019). In the extra-tropics the SST boundary conditions continue to come from the OSTIA (Good et al., 2020) observational product. For observations of SST, ECMWF currently relies on the OSTIA product (see Section 2), which is nudged into the ocean data assimilation at the top level in the ocean model. However, the latency in the availability of the observational product could lead to errors in the initial SST in the wake of TCs.

Figure 44 illustrates the problem with OSTIA just after the passage of Hurricane Teddy (2020). During 21 September, Teddy passed over a cluster of buoys, launched from aircraft ahead of the storm. Around the same time, TC Teddy passed over the area where TC Paulette passed a week before. The figure shows SST from a 2-day forecast valid 22 September 00UTC (top-left) and the OSTIA analysis (top-right) valid at the same time. Both panels include buoy observations. The figure also includes panels with SST from satellite products valid around the same time. The 2-day forecast agrees well with the buoys about the strength of the cold wake created by the TC. However, the used OSTIA analysis lacks the trace of the TC and it is much warmer than the buoys, although the wake after Paulette is still visible. There are two reasons behind the lack of signal from TC Teddy. First, as a daily averaged product it is necessarily received one day behind real time. The other issue is seen in the panels with the SST from the satellites. Due to cloudiness the information is missing from the regions where we expect to see the cooling. However, this case shows that the information from the model can be valuable to fill in the missing parts in the wake of the TC.

The solution to both the observation latency and difficulty of satellites to see through the optically deep clouds around a TC is to use the ocean model as part of the analysis system. This is a cornerstone of the benefits of the weakly coupled ocean-atmosphere data assimilation system which is currently operational. To show the impact of ocean atmosphere coupled assimilation configurations we show in Figure 44 and Figure 45 a number of experiments: an LWDA control (LWDA con) with SST data from OCEAN5 used between +-20° and OSTIA elsewhere, an experiment where this region is extended to +-30° (Partial 30), an experiment (OSTIA) where we do not use any SST from OCEAN5 but rather use OSTIA globally, and finally an experiment where we have outer loop coupling active within the +-30° region (LWDA Outloop con). The middle right panel of Figure 44

shows that when the cold wake of a TC occurs within the region of coupled SST analysis, the ocean model is able to capture signals which can be seen in the in situ observations that remain hidden from the satellites.
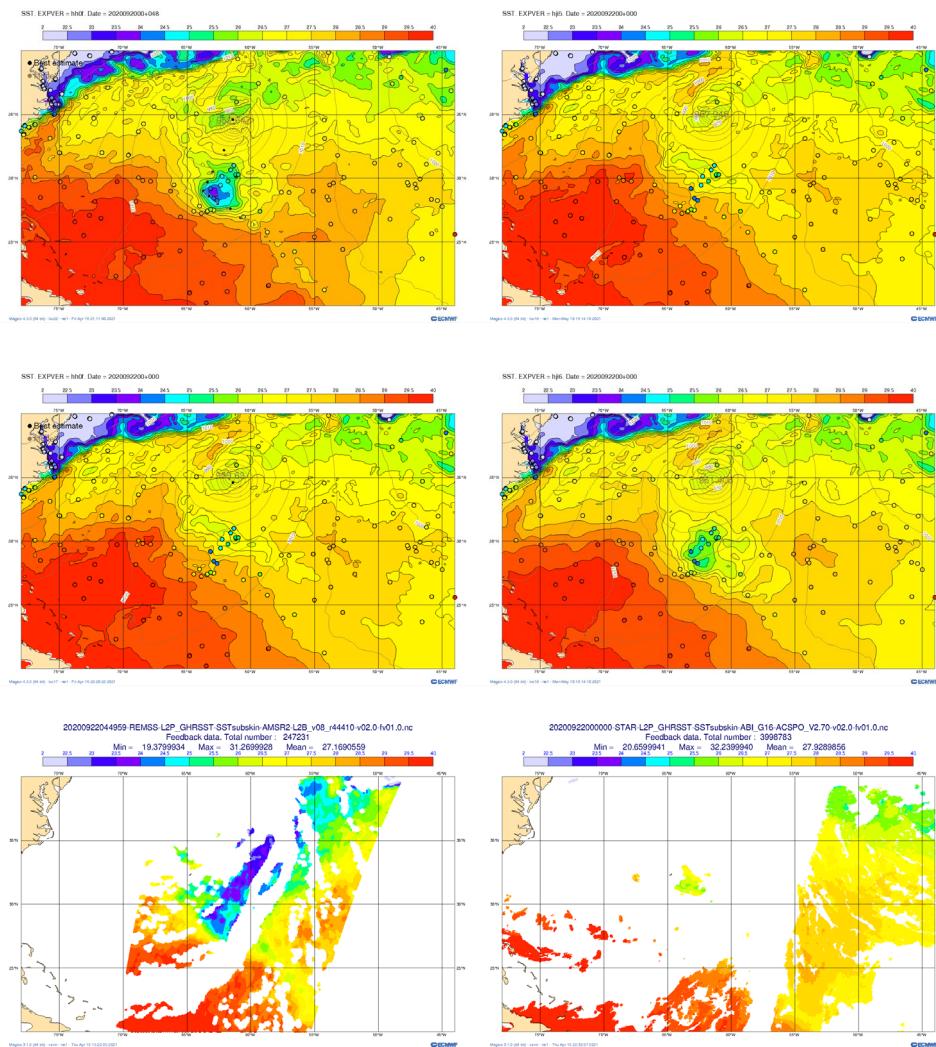


*Figure 44: SST valid 22 September 00UTC from a 2-day forecast (top-left) and from OSTIA (top-right). SST initial conditions for 22 September 00UTC based on the operational configuration with OSTIA SST mainly used ("partial coupling") north of 20N (middle-left) and an experiment with the the boundary moved to 30N (middle-right). SST observations from buoys are marked in circles. SST from AMSR2 valid 22 Sept 04UTC (bottom-left) and from GOES-16 valid 00UTC (bottom-right)*

Figure 45 shows the mean absolute error for position and Pmin for various experiments in different configurations of coupled assimilation and hence different treatments of SST. The impact of the changes on the future of the TC is small on average. However, as the TC mainly affects the wake, other parameters could be affected, as may future TCs if they pass over the cooler wake. The similarity in TC performance with different coupled DA systems is encouraging as it helps show that

coupled DA is not passing ocean model biases to the atmosphere which would show up as changes to the steering flow and therefore degradations to track and intensity forecasts.
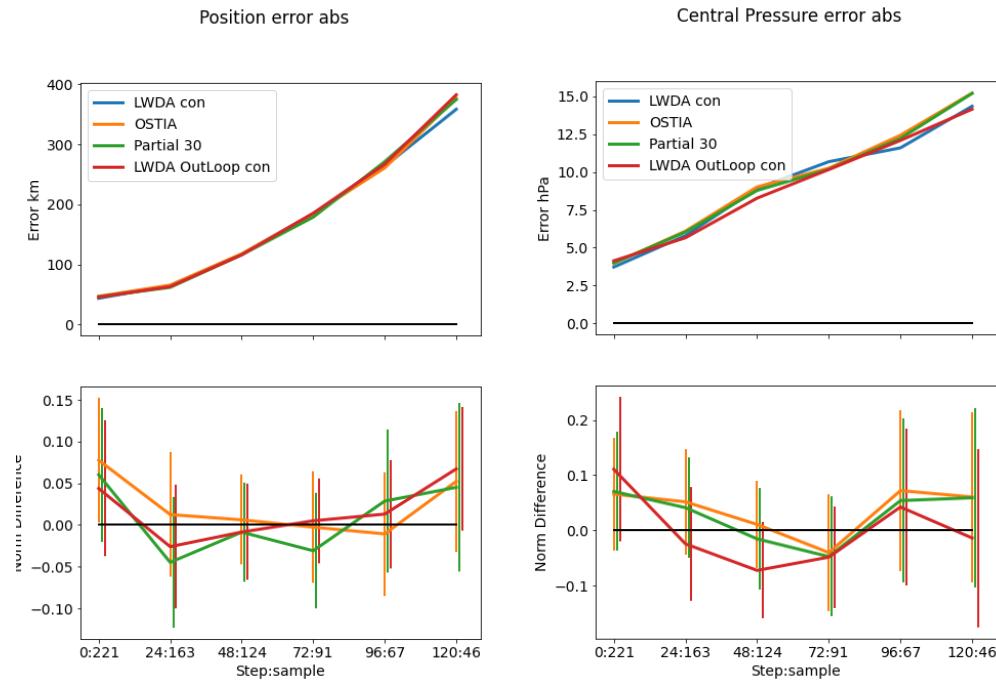


*Figure 45: Mean absolute errors for position (left) and Pmin (right) for various coupled DA experiments.*

## 6.9. Sampling initial uncertainties for tropical cyclones

The perturbed initial conditions for the ensemble members are generated by adding perturbations based on a combination of EDA and singular vector perturbations (SVs, e.g. Leutbecher and Palmer, 2008) to the deterministic high-resolution analysis (Buizza et al. 2008, Lang et al. 2021). To increase the spread of the ENS TC track forecasts, SVs are targeted on each reported TC (Barkmeijer et al., 2001, Puri et al., 2001). At ECMWF, SVs are computed at relatively low (T42) horizontal resolution. Hence, perturbations are more focused on the large-scale steering flow around the TC (Yamaguchi and Majumdar, 2010). For a detailed discussion of the impact of the different perturbation methods and their relative contributions to the ENS TC track and intensity forecasts spread, see Lang et al. (2012).

To determine the SV TC target regions, the reported position together with the TC tracks of the previous ensemble run are taken into account (see Leutbecher and Palmer, 2008). The maximum number of targeted SVs for TCs is set to six. If there are more TCs active at the same time than the maximum number of targeted SV sets, target regions that are close together are combined.

This led to problems during the very active 2020 season. Because of the large number of active TCs, very big target regions were created, which encompassed multiple TCs. This is exemplified by the 17 September 2020 in Figure 46, when ten different tropical systems were active. In the Atlantic, Teddy,

Vicky and Sally was active together with three INVEST systems which a day later became Wilfred, Alpha and Beta. As a result, one SV calculation region was created for Teddy, Wilfred, Alpha and Beta, and perturbation amplitudes associated with each individual system were markedly reduced for several dates (Figure 47, left panel). Increasing the maximum number of allowed target regions to 12 prevents the creation of overly large target regions (comparing the panels in Figure 46) and increases the SV perturbation amplitude in the vicinity of the TCs that would otherwise have been assigned the same target region (comparing the panels in ). In the case of Teddy (the most intense system), the initial perturbations were increased in the southern edge of the sub-tropical anti-cyclone to the north-east of Teddy, a structure the future track for Teddy is sensitive to. As expected, the increased perturbation amplitude leads to an increase in TC track spread in these cases, while otherwise the track spread would have been artificially reduced (not shown).
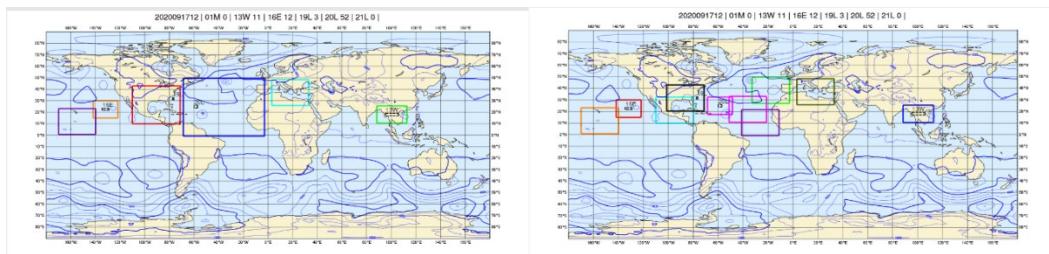


*Figure 46: Tropical cyclone target regions (20200917 12UTC); a) maximum number of 6 regions and b) maximum number of 12 regions.*
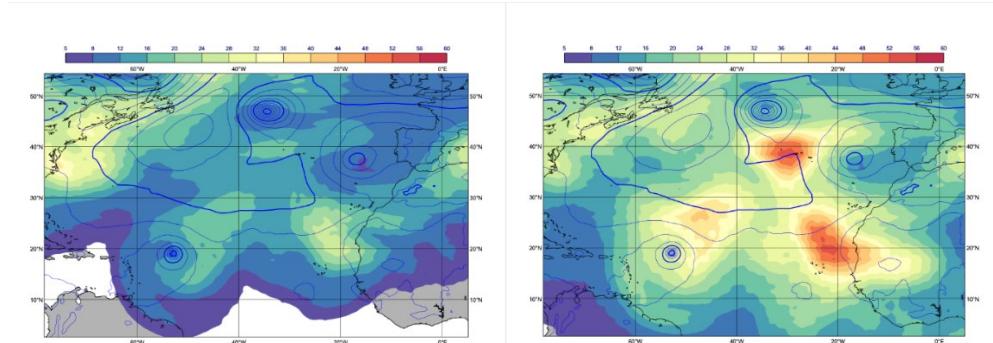


*Figure 47: Vertically integrated total energy of the singular vector initial perturbation and MSLP of control forecast at initial time (2020-09-17-12 00 UTC); a) maximum of 6 target regions and b) maximum number of 12 target regions.*

## 6.10.  Summary

In this section, recent progress at ECMWF in data assimilation in the context of TC analysis and forecasting has been reviewed. Several data assimilation experiments were completed during a special 37-day period in Autumn 2020, run at the same resolution as the operational system at the time (47r1). These parallel experiments provided a unified framework to evaluate and understand the roles of different types of observational data, in the context of the same TCs under investigation. These were conducted with the recognition that any single observation system has a small impact, and that all observing systems combined in the operational system have a substantial collective impact. The goal

of the individual experiments was thus to provide insights on how to make the best use of the observations that we already have, and to identify the context behind which any specific datasets are especially useful.  Based on the 2020 experiments, the results are summarized as follows:

- Analysis values of the position are generally accurate (within 50 km of the Best Track), except for a few weaker TCs in which there is uncertainty in the actual position (Section 5.3).

- Analysis values of the intensity (Pmin) can be substantially weaker than the observed or estimated values, especially in intensifying TCs.

- Faulty surface pressure observations occasionally deteriorate the ECMWF analysis of TCs. A similar risk also exists for ASCAT observations with wrong selection of ambiguity.

- Targeted experiments for TC Laura showed that observations within a 4x4 degree box centred on the TC are important for predictions of both position and intensity.

- The impact of assimilating dropsonde data on position forecasts is neutral while being beneficial for Pmin forecasts. The gains are evident in a small number of cases of intensifying TCs, consistent with the results of Magnusson et al. (2019a).

- Some satellite datasets in the assimilation (Atmospheric Motion Vectors; scatterometers) are sparse in the vicinity of TCs, due to thinning of the data, cloud-top height restrictions (AMV) and saturation of signal for wind speeds above 30 m/s (ASCAT).

- Withholding all satellite-based microwave observations significantly degrades the TC forecasts out to two days.

- The individual impact of each other satellite dataset by itself is marginal and of lower magnitude than the denial of all microwave data, as expected since it comprises a very large fraction of the observational network. Some positive impacts are evident for recently added COSMIC-2.

- The nature of sampling from Aeolus (being only one instrument on one satellite) makes it difficult to reach statistically significant results, and a sub-sampling of cases with good passes might give more insights.

- The assimilation of interpolated hourly central pressure and position data from the Best Track, following the practice at the Met Office, has a small but overall positive impact on intensity for stronger TCs in the north-western Pacific basin, but not in the Atlantic basin.

- An increase in the number of TC target regions for ensemble initialization leads to larger and more relevant initial perturbations for each individual TC during particularly active periods.

- The impact usually varies depending on the initial intensity of the TC.

- The results can differ from basin to basin.

## 6.11.    Discussion and future directions

TCs comprise processes on multiple spatial scales ranging from O(1)-O(100) km, and these scale decompositions also vary depending on the stage of the TC. Questions arise on how effectively the observational data can capture these processes; how the data assimilation scheme is able to exploit these data and provide realistic initial conditions for TCs; how meaningful insights can be derived on observation impact; and how initial conditions and their uncertainty can be appropriately represented. The remainder of this final section discusses these facets and suggests ways forward with observations and data assimilation within this context.

Although the experiments reported in this section have shown that several observing platforms have contributed to improving TC intensity analyses and forecasts, it is not yet known whether the signal is from the assimilation of observations near the core of the TC, or from the surrounding conditions (e.g., environmental humidity and vertical wind shear) important for intensity changes. In parallel, the question of representativeness of the observations, especially in the inner core of the TC where the processes are small-scale and vary rapidly, needs addressing.  Similarly, the volume of observations that are of high quality but are rejected from the assimilation warrant investigation.

For dropsonde data, the intriguing result of why they improved intensification forecasts, albeit for a small sample, warrants further investigation, together with the more general question of how they are contributing to the inner-core structure. One open question here is if the horizontal scale of the data assimilation is still too coarse to take the full advantage of this fine-scale information in areas with sharp gradients in the wind fields. It is also worth noting that other aircraft data, such as flight-level data from the aircraft itself, and remotely sensed data such as airborne Doppler radar are now being routinely assimilated in NOAA's operational systems and are demonstrating benefit. An initial step at ECMWF is under way to explore the impact of flight-level data or additional surface wind products that can help to anchor the position and low-level wind field.  In the longer term, the assimilation of airborne Doppler radar data may be feasible, subject to the availability and maintenance of not only the data but the observation operator and associated error characteristics.

The NHC deploys the NOAA G-IV aircraft around TCs for synoptic surveillance, and they continuously seek to improve their strategies for targeting the aircraft, to optimise the use of dropsonde data for assimilation at operational centres including ECMWF (Michael Brennan, personal communication, 2020). Additional studies to investigate the useful distances and effectiveness of synoptic-scale sampling around the TC in the ECMWF system will help guide NHC's sampling strategy.

The global network of satellite observations is large and complex, with the backbone comprising geostationary satellites and low-Earth orbit satellites. Given that TCs spend most of their lifetimes over the ocean, and aircraft data are usually only available (and are limited in volume) in the Atlantic basin, the combined assimilation of many types of satellite data is especially important for TC predictions.

The positive impact of microwave observations on TC forecasts is due in part to the large volume of observations from several instruments and satellites sampling different times of the day. We also expect that the positive impact on the intensity is coming from the effort in recent years to assimilate observations in all-sky conditions and not only clear sky (Geer et al., 2018). This effort continues with the introduction of all-sky assimilation of AMSU-A temperature sounding radiances, and further investigation on how the all-sky assimilation can be further improved is warranted.

It is important to ask what more can be done with current observations. Most satellite datasets, including scatterometer winds, radiances, and AMVs, are thinned substantially in the operational assimilation. For example, ASCAT winds, whose native grid of wind retrievals is 25 km, are thinned to 100 km, thereby yielding only very sparse data near the TC. The experiment that reduced the thinning before the assimilation further elevated the positive impact of ASCAT. In fact, we now plan to implement the 50 km scatterometer thinning configuration in Cy48r1 based on these results. There

is also more spatial information available in MW radiances, including on warm core structures (Tian and Zou, 2016) down to scales of around 20-50 km. Further work is recommended to establish the optimal spacing of satellite data in which mesoscale features within the TC (including its outflow) are captured, while accounting for correlated observation errors; ambiguities and bias correction would also likely need to be accounted for. One avenue for exploration would be to conduct experiments in which the spatial and temporal density of satellite data in the immediate environment of a TC is increased in the assimilation.

For microwave imagers, some sensors are only passively monitored because in its current configuration the assimilation system is not able to successfully use the additional information. With the addition of some of the sensors that are currently passive (for example, Windsat, DMSP-F15 SSMI and DMSP-F18 SSMIS), temporal coverage of TCs from these instruments could be substantially improved, provided the issues preventing us from using more of these instruments can be overcome.

Infrared radiances are thinned on 100 km scales and are currently only assimilated in clear-sky and fully overcast situations (a very limited number of scenes, particularly around TCs). All-sky infrared assimilation is looking increasingly feasible both in the IFS (e.g., Geer et al., 2019) and in the convective assimilation context (e.g., Sawada, et al. 2019). The combination of geostationary infrared observations at fine time and space scales, and all-sky assimilation, could be very helpful to TC initialisation; these are some of the main observations used in deriving best track estimates. Current satellite observations could therefore provide far more detail on the temperature, humidity and cloud and precipitation structure of TCs than is currently being assimilated.

The GNSS-RO measurements have a limb geometry, and this limits the horizontal resolution of the measurement to hundreds of km, which is clearly not ideal in the context of TCs. In addition, the sampling of the GNSS-RO observations is still quite sparse despite the availability of COSMIC-2 (see Figure 32(g)). Some new GNSS-RO measurements will become available from Sentinel-6a and Spire in 2021-22, but we will also lose Metop-A GRAS. The new operational processing of Metop GRAS measurements, scheduled for implementation in September 2021, should improve the quality of these measurements in the troposphere. In addition, ECMWF will investigate updating the uncertainty model used to weight the GNSS-RO data in the 4D-Var system.

Aeolus has been demonstrated to provide positive impact on forecasts of tropical wind and temperature (Rennie et al. 2021), however showing a positive impact on TCs is proving more difficult. This is probably due to the length of the OSEs and the limited spatial coverage of one instrument. A much longer OSE using reprocessed Aeolus data is underway to try to determine the impact of Aeolus on TCs and extreme weather (an ESA funded project). There are some residual biases for Aeolus winds varying with altitude and wind-speed which are expected to be improved with future reprocessed Level-2B winds. It is possible that the reduced biases will increase the impact, making TC impact more detectable. The early phases of a possible operational EUMETSAT Doppler Wind Lidar mission are underway; so any improvements in the way we assimilate Aeolus or in the data processing may benefit NWP in the future.

Additional satellite platforms targeting TCs, which are being tested in limited-area assimilation, could be considered for testing the ECMWF system. One example is SAR images of surface winds, which

are similar to ASCAT images. Additionally, new <u>small satellites</u> are being developed and deployed, at a small fraction of the cost of the more conventional satellites. Although they are in their infancy, these new types of satellites may hold promise for further improving analyses and forecasts at ECMWF, if the data are timely and of high quality. One example is the TROPICS (Time-Resolved Observations of Precipitation structure and storm Intensity with a Constellation of Smallsats) constellation (Blackwell et al., 2018), which will be launched in 2022, following successful deployment of a pathfinder satellite in June 2021. TROPICS contains microwave soundings with a higher temporal sampling of TCs than is presently achievable. Another promising technology is CYGNSS, whose mission was recently extended to 2023. However, the CYGNSS measurements are less accurate at high windspeeds and are not yet available in near-real-time. The introduction of new observations brings additional costs for the system and needs to be motivated.

The investigations of observation impact in this report have highlighted the question of how to evaluate the impact of new observations in the most efficient way for TCs. There are several caveats in evaluating observation impact in TCs. Firstly, with the huge number of observations present in the operational system, it is difficult to reach a statistically significant impact from a single observation type. There is also the impact of previous assimilation cycles, remote versus local impact, and nonlinear effects in the weights given to the observations by the quality control. Given that the results are often a mixture of improvements and degradations for specific cases, it is important to consider a sufficiently large sample to achieve robust overall results. However, since TCs are not common events, a long integration period would be necessary. Such experiments are often not affordable computationally with the operational resolution, which was motivated in this special study by the fact that TCs are dominated by convective-scale and mesoscale processes. It remains an open question as to how experiments at lower resolution affect the observation impact on analyses and forecasts of TC position, intensity, and structure. For example, the effect of reduced thinning of observations might be under-played in low resolution experiments. As an alternative, the Forecast Sensitivity Observation Impact (FSOI) method can be utilised. Examples of the use of FSOI for TC cases can be found in Duncan et al. (2021). Even if no results from FSOI have been presented in this report, an exploratory study suggested that the method needs to be further adapted to reach reliable results for tropical cyclone impact when aggregated over many cases.

There are several aspects of the <u>data assimilation methodology</u> that could be further explored. With processes on multiple spatial scales, the fidelity of the data assimilation scheme in spatially and temporally resolving TCs in its outer and inner loops requires investigation. The scales also vary with the life-cycle of the cyclone. For example, the onset of RI usually occurs when a TC is a disorganised depression or weak tropical storm, with individual convective bursts dominating until a coherent vortex is formed and begins to axisymmetrize and intensify. These convective bursts and intensification processes occur on timescales of a few hours, within or on the edges of the 12-hour optimisation window in 4D-Var. On the other hand, a mature TC that is not undergoing major fluctuations (such as eyewall replacement cycles) may be resolved both spatially and temporally in the model, with lower wavenumbers dominating. The consequences of deficient resolution that may preclude the effective use of observations on finer time and space scales include: a) the inability to correctly represent spatial correlations in the observation errors; b) overly broad scales in the background errors, and the necessity to further tailor the background errors to TC conditions; and c)

the time and space scales of TC evolution may in some instances be too fast and small for the current system, with its 12 h window and around 50 km increment scale. Experiments to increase both the horizontal resolution of the data assimilation and reduce the length of the assimilation window (and/or to introduce a weak-constraint approach enabling increments to be generated on closer to 3 h timescales) may shed new light on the potential benefits to TC analysis and forecasting, especially for important cases of small, intensifying TCs which occurred several times in 2020 and whose intensity forecasts were poor.

Another assumption of the data assimilation methodology is the tangent linear model, whose fidelity requires investigation especially for small-scale features whose evolution is highly nonlinear. The physical tangent linear structures and subsequent increments in TCs warrant further analysis.

A separate but similarly important issue is the handling of erroneous observations that can severely impact the assimilation. Although several updates have been made in recent years to make the system more stable around TCs, occasional problems with erroneous observations still occur.

Another aspect to further investigate is the quality control. When the difference between an observation and the background is very large, the observation is not used in the analysis so as to avoid any erroneous outliers that degrade the analysis. VarQC was implemented at ECMWF in 1999 (Andersson and Jarvinen, 1999) and is currently applied to most satellite observations and to some conventional ones. But sometimes the large departure could be due to errors in the background, not in the observations. This is more commonly the case for rejected scatterometer wind observations close to the centre of a TC: if the location in the background is misplaced, the scatterometer observations close to the TC's centre have very large departures and are therefore rejected (the observations are given no weight in the assimilation). Usage of the Huber Norm (Tavolato and Isaksen, 2015) as alternative approach to VarQC is a good candidate for ASCAT winds; the Huber Norm allows observations with large departures to be included in the analysis even if with a reduced weight (which is a function of the magnitude of the departure). Preliminary investigations on the ASCAT winds with a few previous model cycles gave mixed results (from neutral to positive; De Chiara et al., 2018). It may be worth testing this again in combination with reduced thinning.

We have also discussed the aspect of coupled data assimilation and the generation of SST analysis. This aspect is most important in TCs as the temperature cooling rate is strong, and yet clouds limit satellite observations of the surface temperature.

The SVs targeted at TCs are important for the reliability of ensemble forecasts of TC track, and increasing the number of SV target regions prevents spread collapse in crucial periods during which many TCs are active at the same time. Future work will assess the impact of increasing the horizontal resolution of the SV computations, and how to improve the generation of the perturbed ensemble members in the context of ensemble data assimilation. Currently, EDA perturbations are constructed from short-range forecasts. The perturbations are then re-centred on the high-resolution deterministic analysis, which can lead to spurious structures in case of tropical cyclones (Lang et. al., 2015). The impact of alternative re-centring options in the EDA and multi-level Monte Carlo methods will also be evaluated.

# 7.    Modelling of tropical cyclones

## 7.1.    Introduction

In this section, recent and ongoing developments in the ECMWF atmospheric and ocean modelling system are presented in the context of TCs. During the 2020 Atlantic hurricane season, model Cy47r1 was operational. While working on this report, Cy47r2 was implemented with an increase of vertical levels for the ensemble from 91 to 137, and the move from double precision to single precision in both HRES and ENS. At the same time, model Cy47r3 was in preparation.

While the large-scale features of TCs can be captured by global models, the core of TCs is dominated by convective-scale and mesoscale processes with sharp gradients, which are at (and beyond) the limit of what a global model can accurately simulate operationally in 2021. To account for these small-scale processes, km-scale limited area models are used as the main tool for TC intensity forecasts. Examples include AROME run operationally at Météo-France with 2.5 km resolution over five different domains in the tropics, and the HWRF (2 km) and COAMPS-TC (4 km) regional models that are run globally at NOAA and the United States Navy respectively. The benefits of this higher resolution are to further resolve processes in the boundary layer, eyewall and rainbands, yielding improved predictions of Pmin and Vmax, and potentially the track due to more realistic TC structures interacting with the environment. However, it should be noted that improvements to the forecast quality are not automatically gained due to an increase in resolution, but go hand in hand with updates to the model physics and the dynamical core of the model.

With increased resolution, the distinction between the model dynamics (the resolved transport) and physical parametrizations (the subgrid-scale transport and mixing) is becoming more blurred, a challenge often referred to as the 'grey-zone'-problem. This is a wide topic for research, for example on the need of parametrized convection, and in Section 7.2 some relevant ongoing developments at ECMWF and their impacts on TC forecasts are highlighted. The recent development of the new moist physics package for the forthcoming Cy47r3 is also presented, as are the effects of a possible future resolution of 4 km and how convection is handled on this scale. Improvements to the model dynamics and the physics-dynamics coupling are ongoing, and the impact of this on TC forecasts is presented. Finally, in Section 7.2, an update on the future developments of the dynamical core towards a non-hydrostatic finite volume model is provided.

TCs develop and are maintained by heat and moisture fluxes from the ocean and are sensitive to the momentum flux at the air-sea interface. It is therefore important to accurately represent these processes with the coupling to the dynamical wave and ocean models, and relevant work is discussed in Section 7.3.

With the imperfections in modelling the evolution of TCs, it is important to also simulate the model uncertainties in the ensemble system to obtain a more reliable ensemble forecast. Some recent and ongoing developments in this field at ECMWF are highlighted in Section 7.4.

In Section 5 we highlighted challenges in simulating the observed relation between Pmin and Vmax, as the model has too slow propagation speeds and too slow intensification and weakening rates. In this section, we examine average errors (biases) and evaluate if any of these aspects are improved. One has to keep in mind that the compensation of errors can lead to a small bias for the wrong reason. For

example, we can ask if the current version of HRES with 9 km grid-spacing and effective resolution of roughly 30 km should have a bias in TC intensity close to zero as seen in Figure 9.

As in Section 5, most experiments reported on in this Section were conducted for the 37-day period 15 August-21 September 2020. The baseline experiment is a 9 km (TCo1279) coupled forecast with Cy47r1, as used operationally in autumn 2020. All experiments are initialised from the operational analysis.

## 7.2.    Model developments in physics and dynamics

As TCs are one of the phenomena with the strongest near-surface winds and gradients on the globe, they are one of the most challenging test cases for the numerical aspects of the model. The strong horizontal and vertical winds push some of the theoretical assumptions, which guarantee stable and efficient performance of the model numerical algorithms, to their limits. For example, it was found that the numerical algorithm used to compute the departure points of the semi-Lagrangian (SL) advection scheme needs more iterations to converge in the high wind speed and high wind shear area of TCs than elsewhere (Diamantakis and Magnusson, 2016). Analysis showed that in such regions, the Lipschitz condition (deformational Courant number smaller than 1) which needs to be satisfied to guarantee convergence of the departure-point iterative procedure is close to its limit. A failure to converge results in mislocated air parcels and large errors in the predicted flow. For the long timesteps that are permitted by the good stability and dispersion properties of the semi-Lagrangian method, convergence of this iterative procedure is slow, and five iterations were found to be required. The numerical convergence can also be improved with a shorter timestep but at the expense of a much larger increase in the overall computational cost.

A reduction in the **model timestep** affects not only dynamical but also physical processes. Therefore, the model is sensitive to the timestep choice, which may have an impact on the TC core pressure and track. Figure 48 includes results from an experiment with the model time-step reduced from 450s ("9km-oper") to 300s ("300s"). In previous experiments we have found occasionally deeper TCs with smaller timesteps, but in this sample the mean difference is small and slightly to the weaker side (not shown). In the past we have also seen indications of a sensitivity for propagation speed to the time-step going from 450s to 225s resulting in ~0.1 m/s faster TCs. For this experimentation, evaluating the average speed for TCs between step 24 and 48 hours, we find a weaker sensitivity (decrease in bias from -0.30 to -0.28 m/s). As seen in Appendix A, the tropical RMSE for 700hPa temperature and winds are slightly degraded with a shorter time-step. There is currently no other systematic improvement known in the overall model forecast skill to justify shortening the timestep and raising the model cost.

There are several ongoing dynamical core developments that could potentially affect TCs. The first one concerns the issue of **slow convergence of the departure-point iterations** that was mentioned earlier in this section. A reformulation of this algorithm (Diamantakis and Vana, 2021), in a geocentric Cartesian framework, allows one to simplify these computations and to accelerate their convergence with no impact on TC track and forecast error. This is illustrated in Figure 48 where a single-precision version of the IFS (as operational in Cy47r2) that uses the "new SETTLS" scheme is compared with a double-precision version of the operational scheme ("9km-oper") for departure points. Both experiments use model Cy47r1 and identical physics, but the former is almost twice as

fast as the latter (due to single precision and use of less iterations) and performs equally well in terms of predictive skill.

In the IFS the **Coriolis force** is only included for horizontal motions. However, in Liang and Chan (2005) it was argued that including the Coriolis force due to vertical motion in the model should lead to a shift of the TC propagation towards the south-west. This result is relevant as the IFS currently has a propagation speed that is slow in comparison with observations, and with a frequently observed drift to the right for weak westward propagating TCs in the Atlantic (as illustrated in Figure 20). In Tort and Dubos (2014), a reformulation of the shallow-atmosphere equations that retains both the vertical and horizontal components of the Coriolis force is proposed. This reformulation is currently being tested in the IFS and the results are presented in Figure 48. Whilst the impact on position and Pmin absolute error is small, we find an impact on the propagation speed bias from -0.30 m/s in the 9 km control experiment ("9km-oper") to -0.24 m/s in the experiment with the change to the Coriolis force ("Coriolis"), and the difference seems to be larger for low-latitude storms. As the magnitude of the bias is small compared with the variability in propagation speed, further investigation is needed to understand the robustness of this result.
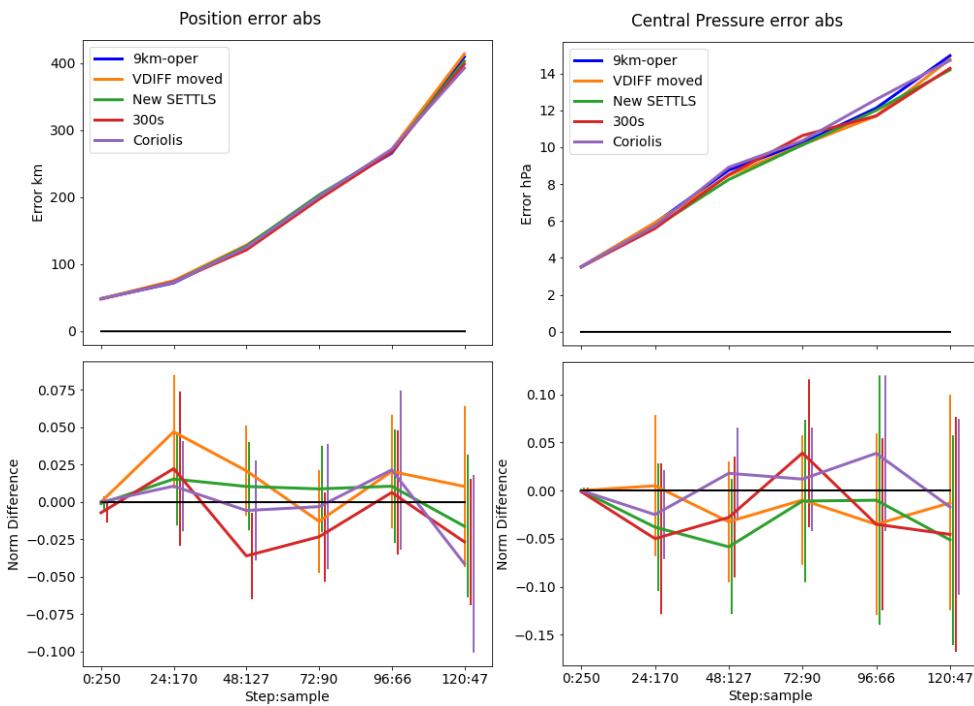


*Figure 48: Mean absolute error for position (left) and Pmin mean absolute error (right) for experiments with changes to the model dynamics. The bottom plots show the normalised difference in error to the 9km-oper experiment.*

A **major upgrade of the moist physics** has been prepared for Cy47r3 (Bechtold et al., 2020). One main driver of this development is the future grid resolution upgrade of the ensemble to O(5 km). With this in mind, the complicated interactions between the turbulence in the lowest part of the atmosphere, convective motions and the cloud physics are parameterised as simply, efficiently, accurately and scale-independently as possible. An important component of this upgrade is a revision

to the convective closure that takes into account, beyond the convective instability, the total moisture advection, allowing a more realistic representation of mesoscale convective systems in terms of rainfall statistics and propagation (Becker et al., 2021).

However, during initial evaluations of the physics upgrade, a large sensitivity of TC core pressure to the convective closure became apparent. TCs and especially the eyewall constitutes an extreme case of total moisture advection, and an excessive increase in convective stabilisation will lead to TCs that are too shallow. A reasonable fix to this problem in Cy47r3 was to exclude the moisture advection term in regions where the vertically integrated saturation fraction exceeds a threshold of 0.94 marking the transition to resolved moist overturning.

The forecast position errors (Figure 49), Pmin errors (Figure 50), Vmax errors (Figure 51) and wind-pressure relationship (Pmin versus Vmax, Figure 52) are illustrated for model runs with the operational Cy47r1 moist physics ("9km-oper") and the new moist physics ("9km-newMP") that forms the basis for Cy47r3. The forecasts have also been run with a resolution of 4 km (TCo2559) representative of a possible future resolution upgrade, for both the operational moist physics ("4km-oper") and the new moist physics ("4km-newMP").

The lowest position error is produced by the "4km-newMP", while both the 9 and 4 km runs of the newMP reduce the position error compared to the respective operational runs, in particular during the first three days (Figure 49(a),(b)). This result might be an effect of the improved large-scale winds in the tropics with the new moist physics, as shown in Appendix C.

The improved position forecasts for the 4 km resolution experiment with new moist physics ("4km-newMP") are encouraging. Next, the corresponding evaluations of the intensity are considered. An illustration of the improvement to the surface wind structure is provided in panels (c) and (d) of Figure 49, for the same 60 h forecast of Hurricane Laura for which the corresponding operational 9 km forecast was illustrated in Figure 5. From Figure 49(c), the TC is much stronger in the "4km-newMP" experiment than in the operational forecast, with a sharper eyewall. This is an encouraging result, as it demonstrates the ability of the 4 km model to realistically simulate the most intense hurricanes. Comparing Figure 49(d) against Figure 5(right), the maximum wind speed in the "4 km-newMP" simulation is accordingly closer to the NHC Best Track value of Vmax, and there is a stronger inflow (blue line). Overall, this encouraging example suggests the potential for overall improvement in intensity forecasts for the entire sample, which will be described next.
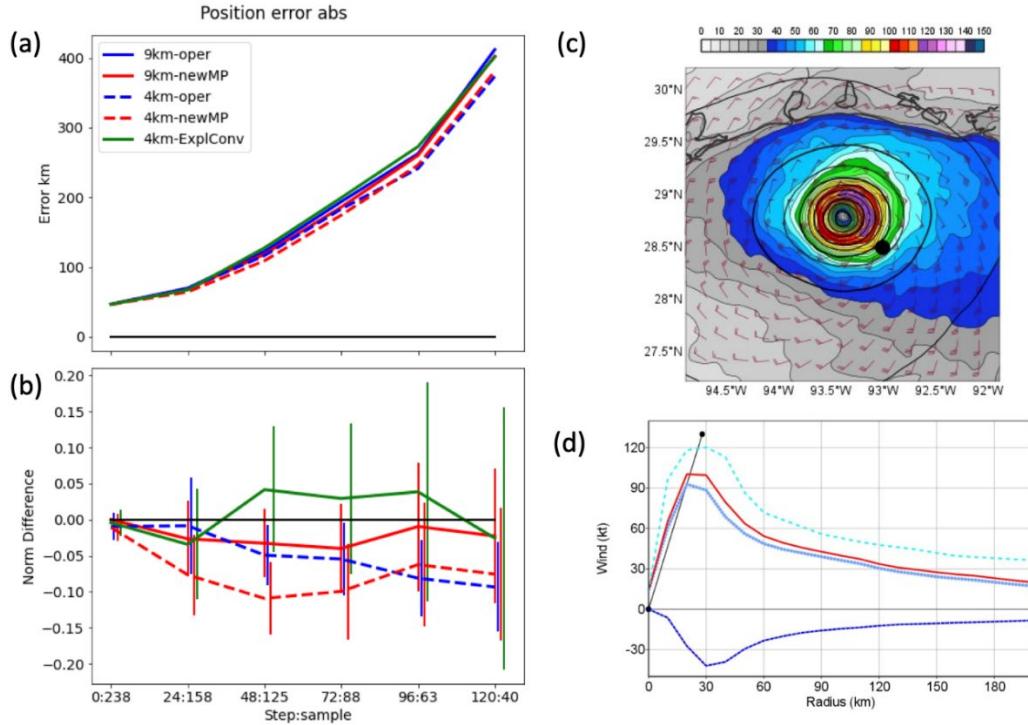
*Figure 49: (a) Mean absolute error for position for experiments with the new moist physics package and experiments with 4-km resolution. (b) Normalised difference in error to the 9km-oper experiment. (c) Surface wind structure in Hurricane Laura, for a 60 h 4km-newMP experimental forecast initialized from the operational analysis on 1200 UTC 24 August 2020. Black contours: Forecast MSLP. Wind barbs: 10 m winds (kt). Shading: 10 m wind speed (kt). Black dot: NHC Best Track position. (d) Radial mean of total (red), tangential (light blue, dotted) and radial (blue) 10-metre wind speed, and maximum wind speed at each radius (cyan, dashed) for 4km-newMP experiment. Black dot: NHC Best Track value of Vmax and the radius to Vmax.*

The operational configuration of HRES ("9km-oper") has a bias in Pmin close to zero (Figure 50) although, as presented in Section 5.5, this is due to the cancelling out between a low Pmin bias for initially strong TCs and a high Pmin bias for initially weak TCs (Figure 21). The newMP with 9 km resolution ("9km-newMP") produces weaker TCs on average due to the moisture convergence change. The mean absolute error in Pmin is slightly lower for this configuration out to 3 days, and these improvements are mostly evident in initially strong TCs (in which the operational forecasts have a low - i.e., too deep - Pmin bias, and the new moist physics reduces this bias). When the resolution is increased to 4 km, the TCs are deeper in both physics configurations. The operational moist physics ("4km-oper") produces TCs that are too deep, which increases the mean absolute Pmin error with respect to the operational 9 km simulations. In contrast, the 4 km simulations with new moist physics ("4km-newMP") brings the overall bias back near zero and produces the lowest mean absolute error of all configurations presented in this paper.
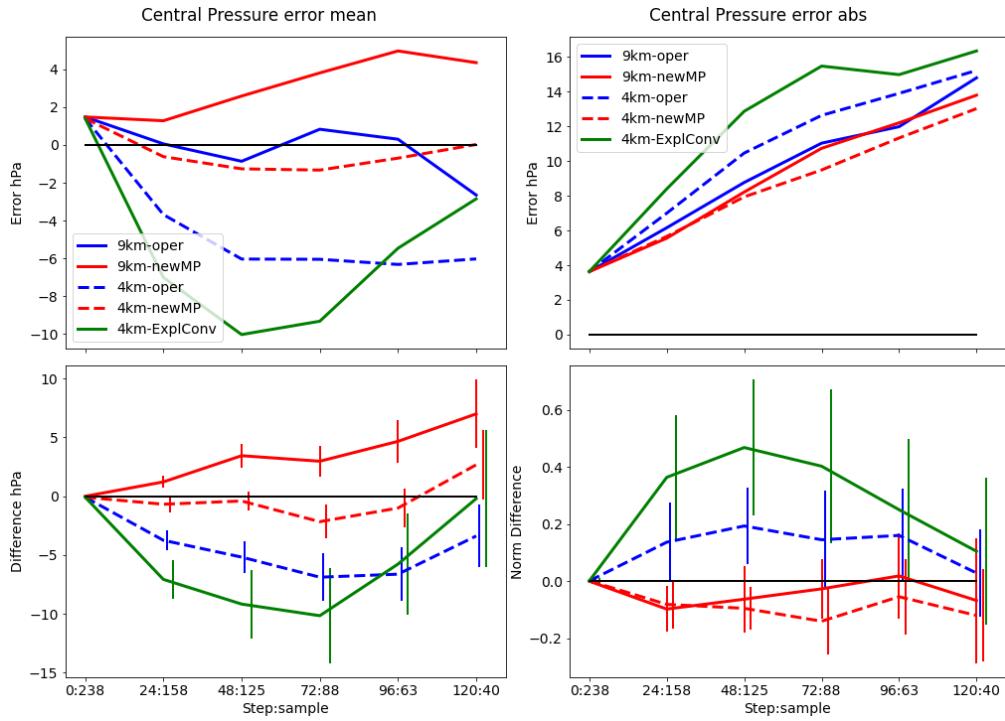
*Figure 50: Mean error (left) and mean absolute error for position (right) for Pmin for experiments with the new moist physics package and experiments with 4-km resolution. The bottom plots show the (normalised in right) difference in error to the 9km-oper experiment.*

The corresponding results for Vmax are presented in Figure 51. Given that Vmax is the key intensity metric used by most operational TC forecasting centres, and it is generally recognized that models of resolution 4 km or less are required for meaningful predictions of this metric, the 4 km experiments here are of particular interest. The results are impressive. Not only does the increase in resolution to 4 km substantially reduce the too weak Vmax bias, but the mean absolute error in Vmax for two-four-day forecasts is reduced by ~7 kt. For the Atlantic TCs in the experiment sample, these improvements are statistically significant (compared with "9km-oper", using a one-tailed t-test) at the 85% level for all forecast times, and usually the 90% or 95% levels. The improvements are most evident and significant for the cases in the sample that were initially weak TCs (Vmax < 50 kt), for all forecast times out to five days (not shown). These improvements due to the 4 km resolution corroborate those for Pmin, although unlike for Pmin, no clear distinction in MAE between "4km-oper" and "4km-newMP" is evident for Vmax probably due to a lower bias for 4km-oper. Overall, the level and consistency of improvement in both Pmin and Vmax when the resolution was increased to 4 km with newMP was not evident in any of the other modelling or data assimilation experiments.
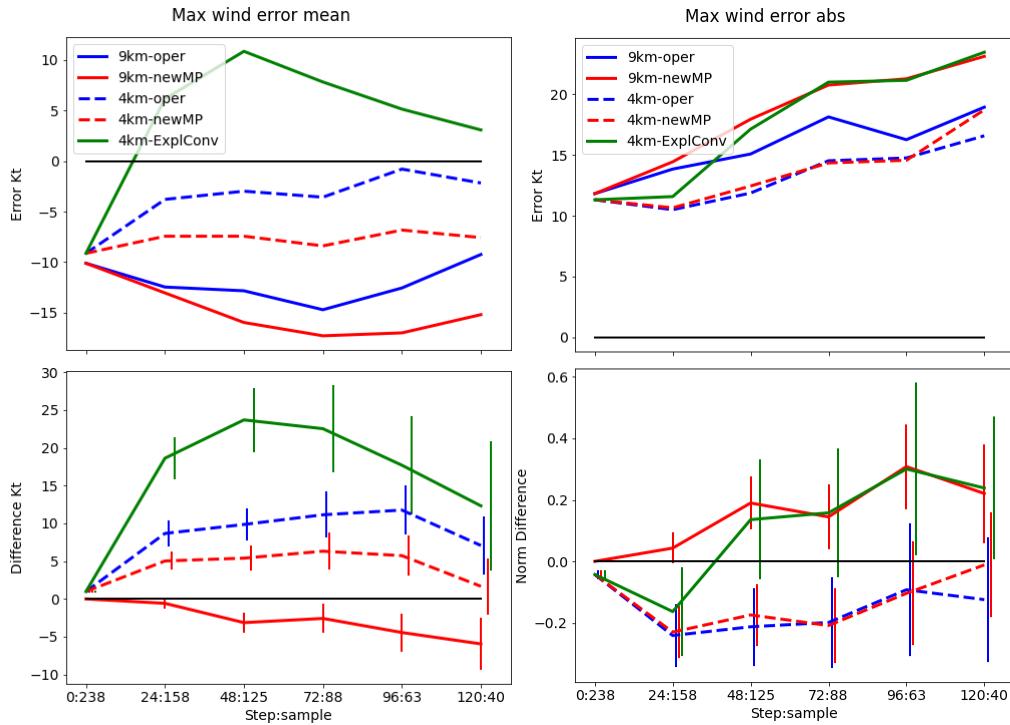
*Figure 51: Mean error (left) and mean absolute error (right) for Vmax for experiments with the new moist physics package and experiments with 4-km resolution. The bottom plots show the (normalised in right) difference in error to the 9km-oper experiment.*

For the pressure-wind relation (Pmin vs Vmax, Figure 52), all configurations underestimate Vmax on average compared with the Best Track. However, one needs to keep in mind the difference in temporal and spatial sampling for the maximum wind in the Best Track, as the model represents a grid box average over a full time-step. As will be discussed later in this section, at ECMWF the pressure-wind relation has improved substantially since the introduction of cycle 47r1. Increasing the model resolution from 9 km to 4 km improves the pressure-wind further for Vmax > 50 kt, and substantial improvements are especially evident at high intensities, where the dashed lines are significantly closer to the black line in Figure 52. This suggests an improved gradient wind balance relation, which is the dominant horizontal balance in intense TCs.
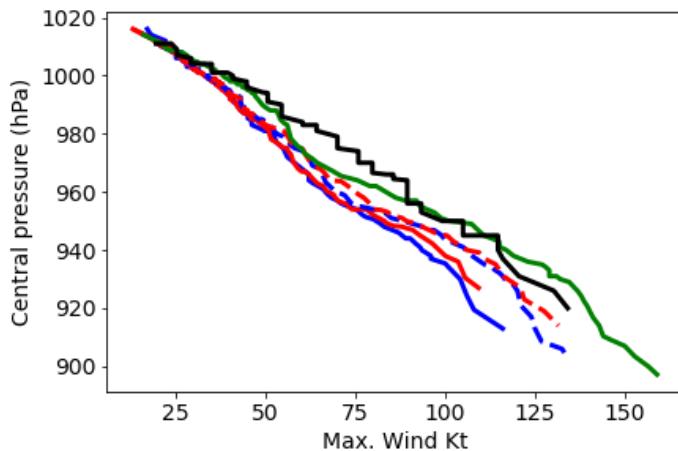
*Figure 52: Wind-pressure relation for experiments with the new moist physics package and experiments with 4-km resolution evaluated for time-steps 24-120h. For legend, see Figure 51 (Best Track in black).*

Table 2 shows the number of tracked feature points for all basins relative to the 9 km control experiment. (Note that these numbers are not compared with the Best Track as the ECMWF tracker includes pre-TC and post-TC stages.) Here we find that the runs with newMP (both 4 km and 9 km) produce ~10% fewer features compared with their old physics counterparts. Comparing 4 km resolution with 9 km resolution, we find a similar number of features, but about 30% more cases of hurricane intensity with the higher resolution. Further diagnostics of the TC activity is needed to validate the benefit of the changes (see Section 4.3).

*Table 2: Number of features relative to 9km control experiment for different cyclone strengths.*

|  | >8 m/s | >17 m/s | >32 m/s |
|---|---|---|---|
| 9km | 1 | 1 | 1 |
| 9km-newMP | 0.87 | 0.89 | 0.93 |
| 4km | 1.02 | 1.24 | 1.31 |
| 4km-newMP | 0.90 | 1.06 | 1.21 |
| 4km-ExplConv | 1.24 | 1.60 | 1.69 |

Going towards km-scale model resolutions, one open question is the handling of deep convection. In the DYAMOND project (Judt et al., 2021) a 4 km version of the **IFS without parameterised** convection was compared with other global models with similar resolution. We have therefore also run here the newMP with the deep convection switched off, to let the model explicitly resolve the convection ("4km-ExplConv"). This required a reduction of the model time step from 300 to 180 s to

avoid model failures due to excessive vertical velocities. The experiment with explicitly resolved deep convection provided slightly higher errors in the position (Figure 49), and far too intense TCs (Figure 50 and Figure 51). However, the wind-pressure relation is improved with 4km-ExplConv compared with 4km-newMP (Figure 52). This result is in line with experiments for Polar Lows where simulations with 5 km resolution and explicit convection produced stronger and more localised wind maxima compared with simulations with parameterised convection (Hallerstig et al., 2021).

The impacts of increased model resolution and deep-convection parameterisation are illustrated for one forecast for TC Laura (Figure 53). For the 4 km experiment with explicit deep convection, Pmin almost reached 900 hPa when the Best Track was around 940 hPa. Among the other experiments, the best results are seen for 4km-newMP, in line with the verification results for the full sample. This is further illustrated by the comparison of simulated infrared satellite images at the most intense time of TC Laura, with a satellite image from GOES-16 (Figure 54). Here the 4km-newMP has a clear eye of the cyclone, in line with the real image. For 4km with explicit convection, the eye is hidden by thick clouds and the TC is smaller in scale, resulting in too small a radius of maximum wind (not shown).
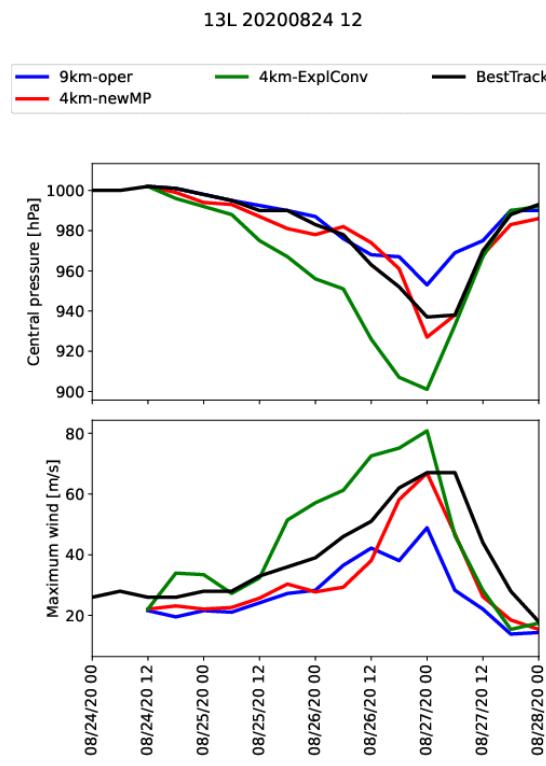


Figure 53: Intensity (Pmin) forecasts for TC Laura in forecasts from 24 August 12UTC.
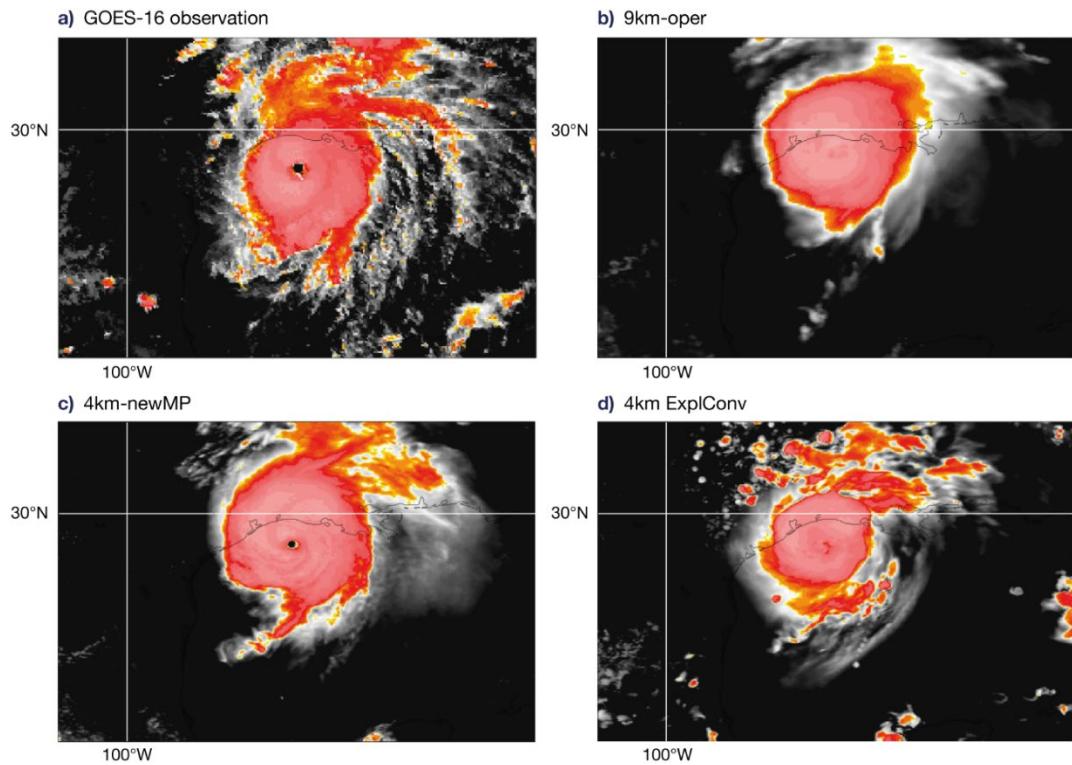
*Figure 54: Satellite images for TC Laura on 27 Aug 00UTC from GOES-16 (a), and simulated images from forecasts initialised 24 Aug 12UTC in 9km-Oper (b), 4km-newMP (c) and 4-km with explicit convection (d).*

For the propagation speed, averaged between step 24 and 48 hours, we find for this sample a bias of -0.28 m/s for the control experiment (9 km). Note that the numbers are slightly different to the ones from the dynamical core experimentation due to homogenisation over a different set of experiments. 9km-newMP increases this bias to -0.32 m/s. Both 4 km experiments decrease this bias to -0.20 m/s (4km) and -0.26 m/s (4km-newMP). However, as documented in Appendix C, both newMP experiments have the lowest RMSE for 700hPa winds in the tropics. Furthermore, 4km-ExplConv results in a lowest propagation speed bias (-0.09 m/s) while being the experiment by far with the highest RMSE for wind errors at 700hPa in the tropics. One can note that the two experiments with the strongest intensity bias also have the fastest propagation of the TCs. As discussed above, we did see some sensitivity to introducing the vertical component of the Coriolis force (reduction from -0.30 to -0.24 m/s). Further investigations with a larger forecast sample are needed to better understand these differences, and how it relates to different types of TCs (strong vs. weak, low-latitude vs high-latitude).

Experiments have also been undertaken to investigate sensitivities in the **coupling between the dynamics and physics** in the model. Results with the new moist physics show that the quality of the TC forecast is sensitive to the coupling of the diabatic processes parametrization to the model dynamical core. Following suggestions made by Beljaars (2016) and results of the theoretical study of Termonia and Hamdi (2007), the vertical diffusion scheme has been moved to the very end of the sequential physics chain. This allows the convection scheme to directly respond to the explicit

tendencies from dynamics, while having the vertical diffusion scheme moved at the end stabilizes the whole physics. This theoretically justified option currently represents only an experimental design requiring a double call of the vertical diffusion scheme – before convection, to provide BL fluxes to convective closure and then second time at the end of physics. Furthermore, changes in the coupling between the vertical diffusion scheme and the radiation, showed promising results in terms of forecast skill were also tested on TCs. Including both coupling changes shows a small positive impact with respect to the new moist physics (see VDIFF experiment in Figure 48). More optimal alternatives in terms of computational cost for the coupling between the convection and vertical diffusion scheme may be investigated in future.

Another open question is the impact of moving to a **non-hydrostatic dynamical core**. However, tests with the IFS at 9 km horizontal resolution indicate much larger sensitivities to the special numerical algorithms, which need to be activated to integrate in a stable fashion the more demanding non-hydrostatic equation set, than to relaxing the hydrostatic approximation itself. For example, recent experimentation shows that the physics-dynamics coupling of the so called Iterative Centred Implicit scheme (ICI) used by the non-hydrostatic model needs improvement to be competitive with the scheme used in the hydrostatic dynamical core. The TC performance is degraded also in a hydrostatic experiment with the ICI activated to a similar level as the non-hydrostatic experiments, compared to the operational hydrostatic model without the ICI scheme (not shown). Interestingly, while the ICI scheme deepens the TCs, the maximum wind in the TCs is almost unchanged, leading to a significant degradation in the maximum wind vs. minimum pressure relationship.

As a further development, the **IFS-FVM non-hydrostatic dynamical core** employs a fundamentally different finite-volume design with inherent conservation and robust semi-implicit integration (Kühnlein et al., 2019). The IFS-FVM foundations lie in cloud-scale large-eddy simulation, potentially offering numerical techniques for NWP at kilometre-scale resolutions and beyond. Finite-volume methods are common among the newest generation of NWP models such as ICON of DWD, but formulations differ significantly in details. An experimental IFS-FVM forecast configuration using the IFS Cy43r3 physical parametrizations has been developed, which enables re-forecast studies (as yet uncoupled) with this new model formulation, including tropical cyclone cases. In 2020, a first single case study of TC Irma (2017) verified the IFS-FVM against the established IFS as well as the best track data (see Kühnlein et al., 2020). The TC case study has been very useful to demonstrate the basic skill, as well as to identify open issues and specific aspects for further refinement of the IFS-FVM forecast configuration, particularly in initialization and physics coupling. Further systematic comparison studies against the hydrostatic and non-hydrostatic spectral-transform IFS will follow. To enable the highest possible future resolutions, the IFS-FVM is currently undergoing revision in terms of its software infrastructure towards emerging computing technologies (Bauer et al., 2020).

## 7.3.    Ocean and wave modelling

Since 2018, all ECMWF forecasts are produced with a **coupling to the NEMO dynamical ocean model**. One of the main motivations for making this step was improved intensity in TCs. Strong TCs lead to SST cooling due to vertical mixing, enhanced upwelling, and heat and moisture flow to the atmosphere. Historically, when ECMWF forecasts predicted too weak TCs (see Section 1) due to the lower horizontal resolution, the lack of ocean coupling was not crucial due to absence of very strong

wind. Also, the lack of SST cooling might have partly compensated intensity errors due to low resolution. However, with increasing model resolution, cases of too deep TCs (Pmin too low) started to appear, most pronounced in the north-western corner of the North-western Pacific basin. Mogensen et al. (2017) investigated this issue and related it to the lack of ocean coupling, which had the strongest effect on the shallow, warm layers in the ocean such as south of Japan. The positive effect of ocean coupling has also been found at other global modelling centres, such as the Met Office (Vellinga et al., 2020).

Figure 55 includes results for experiments with ("9km-oper") and without ocean coupling ("FC uncoup"). On average, the experiment without ocean coupling led to too strong TCs and increased absolute errors. The clearest example in this sample from Aug-Sept 2020 was TC Haishen, whose path took it over an area south of the Japanese island Kyushu with low ocean heat content, but initial warm SST of around 30°C. This is illustrated in Figure 56 where we have plotted the mean temperature in the uppermost 300 m as a measure of upper ocean heat content. The bottom panels of Figure 56 show the evolution of SST in the coupled experiment (left) and the uncoupled experiments (right) overlayed with SST observations from drifters and ships. The coupled model was able to simulate the cold wake after the TC even though the cooling was less than what is observed. As the uncoupled model lacked the physics needed for the air-sea interaction, it continued to be driven by the prescribed warm SSTs leading to the over-intensification of the forecast. It is worth noting that TC Haishen behaves in a very similar way to TC Neoguri 2014 which was discussed in detail in Mogensen et al. (2017).
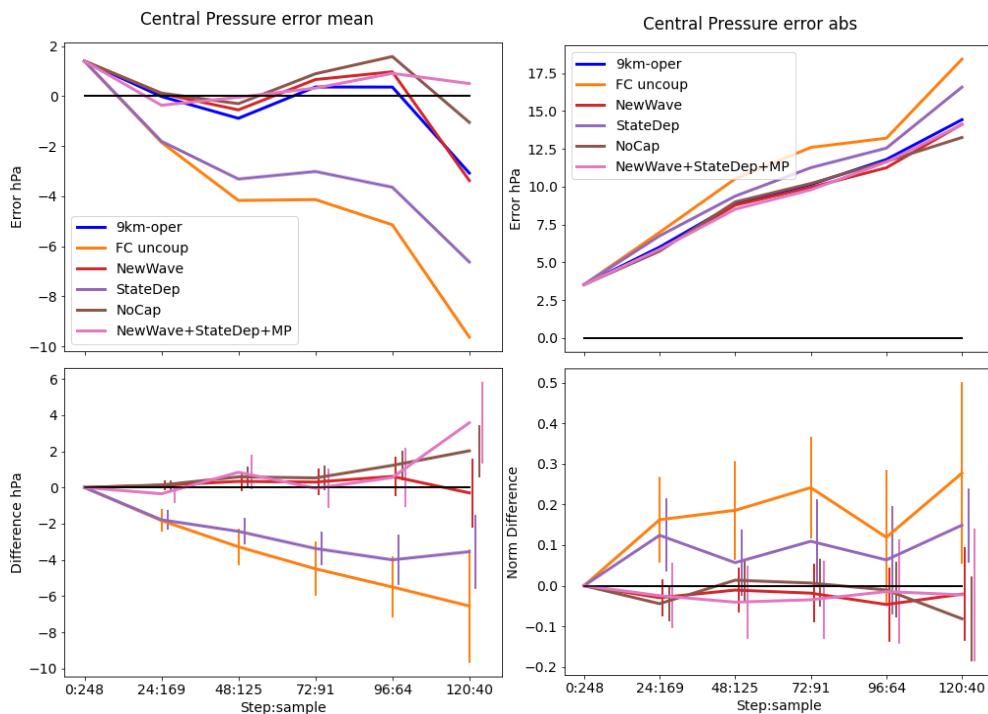


*Figure 55: Mean error (left) and mean absolute error (right) for Pmin for various experiments related to ocean and waves. The bottom plots show the (normalised in right) difference in error to the 9km-oper experiment.*
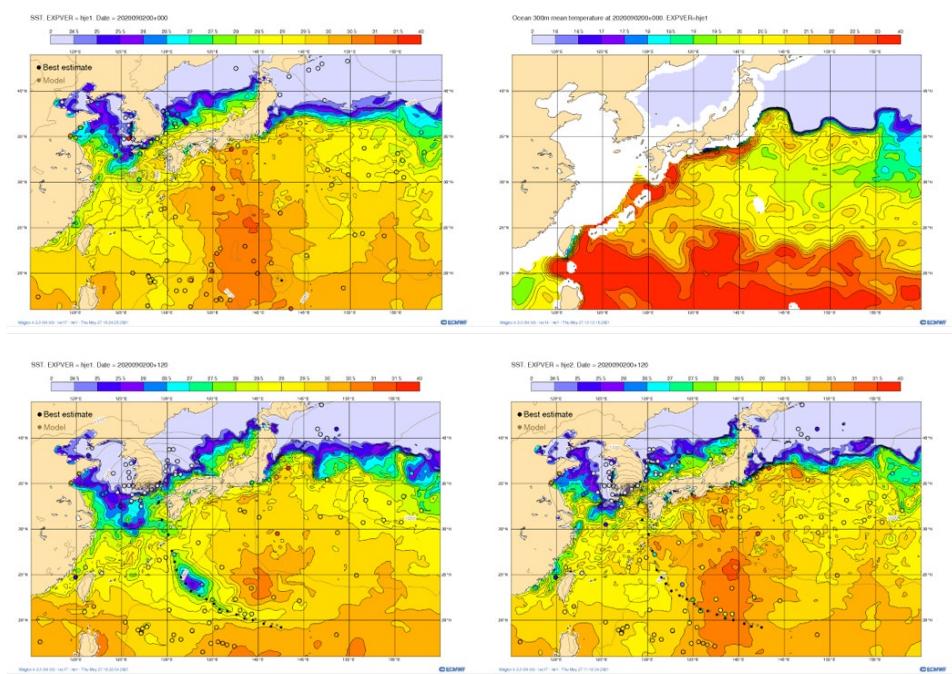
*Figure 56: Sea surface temperature (top left) and mean upper 300 m temperature (top right) both valid at 20200902 00z and SST from coupled (bottom left) and uncoupled forecasts (bottom right) valid 120 hours later. Black dots indicate Best Track for TC Haishen and coloured dots are SST from observations.*
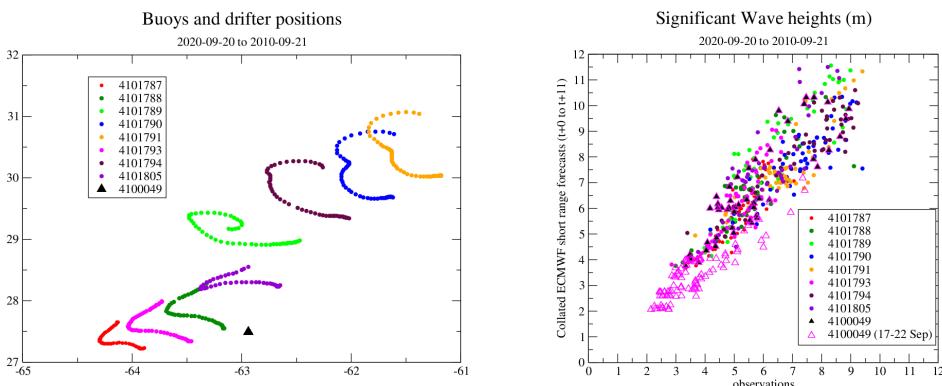


*Figure 57: Observation positions for buoy and drifter after passage of TC Teddy (left) and significant wave height co-located in short HRES forecasts and observations.*

All ECMWF forecasts are coupled to a dynamic ocean wave model. Besides providing users with sea state parameters, as detailed as the full two-dimensional wave spectrum, the two-way coupling to the wave model ensures a physically more consistent representation of air-sea interactions, linking the lower atmosphere to the upper ocean. As an example of wave evaluation for one cyclone, Figure 57 shows significant wave height from HRES and observations under TC Teddy, from an array of buoys seen in Figure 44, around 28°N, 62°W. For this case there is a general good match between modelled

and observed wave height. For the buoys closest to the centre of the cyclone centre path (purple and light green), the wave height was somewhat over-estimated in HRES.

Prior to IFS Cy47r1, ECMWF forecasts underestimated maximum wind speed for intense TCs even given the correct central pressure. While there are many different factors that could account for this behaviour, a strong candidate is that it could be linked to the parametrization of **momentum exchange at the ocean surface.** This momentum exchange is dependent on the state of the waves (sea state) and not just the surface winds, resulting in a range of values for the drag coefficient over the oceans for similar wind speeds. It is achieved by coupling the atmosphere with an ocean wave model.

Over the last decade, it has been suggested that the drag coefficient should tail off for strong winds. In the IFS, the momentum exchange with the sea surface is modelled via a dependency of the roughness length ($z_0$) on the surface stress. This expression accounts for both low and high wind regimes. At low wind speed, the sea surface becomes aerodynamically smooth and $z_0$ is determined by viscosity. At high wind speed, Charnock's relation is used, in which $z_0$ is expressed as a function of surface stress, air density, gravitational acceleration and a sea-state-dependent Charnock parameter. In ECMWF's wave model, the Charnock parameter depends on the state of development of the resolved waves and a tuneable parameter ($\alpha_b$) which represents the impact of unresolved short waves (background roughness beyond the highest frequency resolved by the model) on the overall surface stress. Until Cy47r1, this parameter had a constant value.

Observational evidence that the drag coefficient should be much lower for high winds suggests that the coupling between the ocean surface and the wind above becomes less efficient at transferring momentum for high winds. For this to happen, it was realised that the Charnock parameter should be considerably smaller in the case of high winds (above 35 m/s). This was achieved by reducing $\alpha_b$ for strong wind speeds in Cy47r1. This development is the result of internal ECMWF work informed by discussions with scientists at Météo-France and the US National Centers for Environmental Prediction (NCEP).

The effect of the drag change in Cy47r1 is evident in the pressure-wind (Pmin versus Vmax) relation in Figure 58, where the experiment "NoCap" has the formulation from the previous cycle without the cap of the drag. The wind-pressure relation is much less consistent with observations for the experiment "NoCap" for wind speeds above 70 Kt, compared to 9km-oper. However, it could also be seen that all experiments still underestimate the winds in a similar way for speed between 60-70 Kt. This indicates that more than just a capping of the surface drag is necessary to solve this issue.
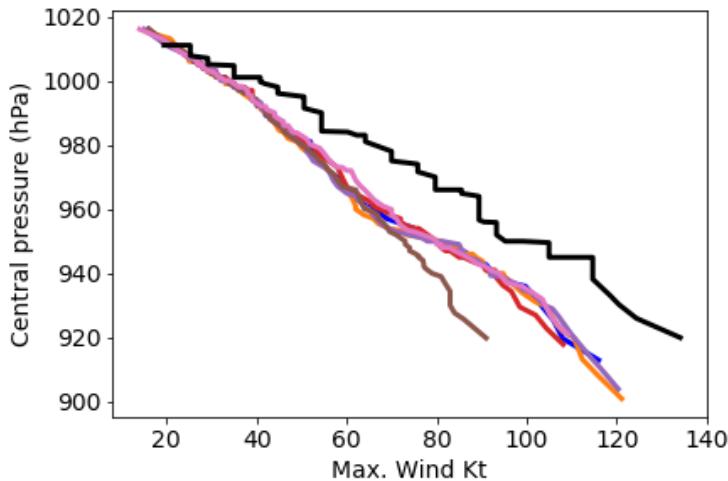
*Figure 58: Wind-pressure relation for experiments with the different ocean configurations evaluated for time-steps 24-120h. For legend, see Figure 55.*

We are currently exploring an alternative approach to the simple capping of the Charnock parameter. The current parametrization for the wind input into waves is not properly accounting for the impact of the short gravity-capillary waves, nor is the nonlinear effect of the wave spectrum on the growth rate of wind waves accounted for. Recent work as presented in Janssen and Bidlot (2021), is proposing a simple model for the impact of gravity-capillary waves on the wave growth and is extending the current parametrization to include the nonlinear effect. Both effects yield a reduction of the drag that is function of the sea state. Because the impact of the new approach affects all wind regimes, it will require full meteorological testing. However, for the current experimentation, we see that the "NewWave" simulation gives similar results as the reference (Figure 55, Figure 58).

The sea state also has an impact on the latent and sensible heat fluxes via the exchange coefficients for heat and moisture. From recent observation campaigns (Brut et al., 2004; Cook and Renfrew, 2015), there are indications that the exchange coefficients for heat and moisture should have a stronger dependency on the wind speed (i.e., the sea state) than currently modelled. Recent sensitivity experiments have shown the potential impact on the atmosphere (Jansen and Bidlot, 2018). In the ongoing work, we have quantified the impact of this new parameterization on TCs ("StateDep"). With this change, more heat and moisture can be extracted from the ocean, the TC gets deeper, and surface winds stronger. It is worth stressing that such a change would not have been feasible without an interactive ocean, as the TC-induced cooling of the ocean is needed to avoid further over deepening of TCs. There still appears to be some over intensification. However, when we combine "NewWave" and "StateDep" with the new moist physics (see Figure 55), the mean error for Pmin is close to zero (and similar to control), as the new moist physics reduces the intensity. This illustrates that there are several configurations that can give a close to zero bias for Pmin, and highlights the care needed when interpreting the results due to compensating errors.

We also note that the effect of spray as generated by whitecaps and breaking waves is still ignored and should considered in future work (Wu et al. 2015).

## 7.4.    Sampling model uncertainties

At ECMWF, the Stochastically Perturbed Parametrisation Tendency scheme (SPPT) is used to simulate model uncertainties in ensemble forecasts (Buizza et al., 1999). The idea behind SPPT is to perturb the total physics tendency, that is, the sum of the tendencies from all physical parametrization schemes, such as convection, radiation, etc. The perturbations are made by multiplying the total tendency with a factor determined from a two-dimensional (2D) Gaussian random field with prescribed space and time decorrelation scales. The currently operational implementation of SPPT at ECMWF is described in Leutbecher et al. (2017) and Lock et al. (2019).

Figure 59 shows the mean absolute error of Pmin during July-November 2020 for operational HRES (the performance of Cy47r2 HRES was very similar, not shown), ensemble control (ENS-CF) and one ensemble member (ENS-PF1), from the pre-operational ensemble experimentation for Cy47r2. For the Pmin, the ENS-CF and ENS-PF1 has a larger positive bias (too weak TCs) than HRES as expected from the lower horizontal resolution. We find the ENS-PF1 has somewhat stronger TCs than the ENS-CF for all lead-times. We suspect this difference is coming from the use of the SPPT scheme to simulate model uncertainties. The wind-pressure relation is similar for the ENS-PF1 and ENS-CF (not shown).
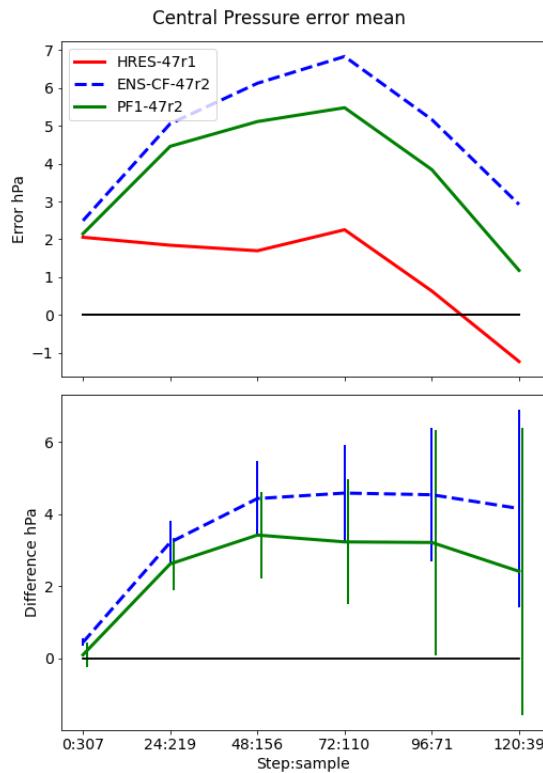


*Figure 59: Pmin mean error for HRES and control forecast and one perturbed member from cy47r2. The bottom plots show the difference in error to the HRES.*

Table 3 shows the number of cyclonic features relative to operational HRES. With the lower horizontal resolution for the ensemble we expect fewer features, especially for the stronger categories. This is confirmed and the ENS-CF has only 45% of the number of hurricanes in HRES. Comparing ENS-CF and ENS-PF, we find more features in the perturbed members and the signal is stronger for the weaker categories. This result is in line with results from System 5 seasonal forecasts (Stockdale et al., 2018) and long simulations presented in Vidale et al. (2021).

*Table 3: Number of cyclonic features relative to 9km HRES forecasts for different cyclone strengths for step 24-120h.*

|                  | >8 m/s | >17 m/s | >32 m/s |
|------------------|--------|---------|---------|
| HRES (9km)       | 1      | 1       | 1       |
| ENS-CF (18km)    | 0.92   | 0.80    | 0.45    |
| ENS-PF (18km)    | 1.13   | 0.95    | 0.56    |

ECMWF is currently developing a Stochastically Perturbed Parametrizations scheme (SPP), which represents model uncertainty by introducing stochastic perturbations directly into the physical parametrization schemes. In contrast to SPPT, SPP preserves conservation properties of the parametrization schemes within the model column. While the first implementation of SPP produces less skilful medium-range probabilistic forecasts than SPPT, a recently revised SPP configuration is now competitive with SPPT (Lang et al., 2021). The revised version from the scheme introduces perturbations to additional quantities and modifies the probability distributions sampled by the scheme compared to the original implementation of SPP documented in Ollinaho et al. (2017) and Leutbecher et al. (2017).

Lang et al. (2021) compared the ensemble spread for TCs in lower-resolution ensemble forecasts (TCo399, approx. 30 km horizontal grid spacing) and found that SPP results in larger TC intensity (Pmin) ensemble spread compared to SPPT, while intensity forecast errors are very similar, leading to a better spread-error relation. TC track errors and ensemble spread are also very similar for SPPT and SPP.

A further scheme that is under development explores how to represent model uncertainty arising from the IFS dynamical core: STOCHDP introduces stochastic perturbations to the departure point (DP) calculation in the semi-Lagrangian transport scheme (Leutbecher et al., 2017). As discussed in Section 7.2, in Diamantakis and Magnusson (2016) the iterative DP calculation was shown to converge most slowly (indicating greatest uncertainty) in regions associated with more complex flow. In particular, the authors presented a case study of TC Neoguri (2014), which demonstrated that where the windspeeds and Lipschitz numbers are largest, the DP calculation is slowest (and in places even fails) to converge.

In STOCHDP, the DP convergence rate is used to scale the stochastic perturbations that are added to the DP calculation. An examination of the TC Neoguri case with STOCHDP active has shown that the scheme successfully generates ensemble spread which tracks and develops with the TC (see Lock, ECMWF Annual Seminar 2020). Work is ongoing to explore the scheme sensitivities to the choice of perturbations and model resolution.

## 7.5.    Summary of results

In this section, several recent and ongoing developments to the IFS model have been discussed, both in the physics and dynamics of the atmosphere, and aspects related to the ocean. The simulation of model uncertainties has also been discussed. All these aspects have largely been discussed in the context of processes in the vicinity of TCs, which occur mostly on the convective-scale and mesoscale (of order 1 km up to a few hundred km).

Particular attention has focused on a range of modelling experiments during the special 37-day period in autumn 2020, which were conducted using the same initial conditions as in operations. We therefore note that the challenges described in Section 5 in creating accurate operational analyses of TCs carry over to the modelling experiments, as the modelled TC is dependent on the realism of the initial conditions. Cognizant of this caveat, these parallel experiments have nevertheless provided a unified framework to understand the roles of the different modelling tests, with the same TCs under investigation. The intent in some of the tests is not necessarily to improve the forecast, but instead to understand the behaviour of a particular modification in several cases. A summary of the finding of the 2020 experiments is provided here, with further discussion and future directions in the final sub-section.

Generally, as was the case for the data assimilation experiments, the modelling experiments did not produce dramatically different forecasts from the (near) operational version that was used as the control.  This is as desired, since it demonstrates robustness and a lack of volatility in the model.

The reduction of the timestep (in the 9 km configuration) did not result in a systematic improvement to the TC forecast skill, and hence these results do not provide a justification to shorten the timestep and raise the model cost.

The single-precision test performed as well (with new SETTLS scheme) as the (47r1) double-precision test, encouragingly suggesting that overall degradations are not likely when single-precision is used. This is consistent with Cy47r2 that included the change to single-precision in HRES.

The introduction of the contribution to the Coriolis force from the vertical motion provided some modest improvements to the track forecasts of weak TCs in the Atlantic basin. The experiment showed a decreased propagation speed bias, a result that will be followed up with a longer experiment period to increase the robustness.

The change to the moist physics package at 9 km resolution, which will be included in the 47r3 cycle, provided improved TC position forecasts.  Statistically significant improvements to Pmin forecasts were found in the Atlantic basin, especially for cases being hurricanes (> 64 kt) at the initialisation time.

The experiments in which the resolution was increased to 4 km provided the most significant combined improvements to Pmin, Vmax, and the wind-pressure relation out of all the 2020 experiments. The usage of the new moist physics package in the 4 km experiment yielded further improvements in Pmin, with a reduction in the strong bias for initially strong TCs, and a reduction in the weak bias for initially weak TCs. As we expect the intensity bias to asymptotically approach zero with increasing resolution, we see this as a positive result for the new moist physics. The experiment with explicitly resolved deep convection provided slightly higher position errors, and the TCs were far too intense.

The experiment with no coupling of the atmospheric model to the ocean yielded significantly degraded results, especially for the minimum pressure of hurricane-strength TCs. This result confirms the importance of the coupling in the operational system. Since Cy47r1, the wave drag has been modified for extreme wind speeds to improve the wind-pressure relation for tropical cyclones, a result that also was confirmed in this report.

For the ensemble we find on average stronger TCs with the use of the SPPT scheme and a higher number of cyclonic features, something that needs further investigation to understand. Experiments with the SPP scheme shows promising impact on the TC prediction in terms of generating ensemble spread.

## 8.    Discussion and future directions

The benefits of increasing the model resolution from 9 km to 4 km are evident in the Pmin and Vmax forecasts, and the wind-pressure relation. This is especially clear when the new moist physics package is used. However, as the model resolution is increased, the traditional boundary between model physics and dynamics becomes blurred. As the resolution approaches 4 km, it opens the question of whether parametrized deep convection should continue to be used, or whether to let the dynamics explicitly resolve the convection. The explicit convection experiment in this section suggests that parametrized convection is still beneficial for TC simulation at this resolution, as the simulations without the parametrization created too strong TCs. Related to the research on **slow convergence of the departure-point iterations**, the impact of a new IFS development that introduces a more accurate algorithm to find the departure points based on the 4th order Runge-Kutta Lobatto IIIA scheme will be investigated. Future model resolutions also bring up the question of using a non-hydrostatic dynamical core. However, this choice has implications for the coupling to the model physics that are currently being explored. In the longer term, the aim is to move the dynamical core to a finite-volume model, which is non-hydrostatic.

As part of the ongoing INCITE20 project, 1 km global simulations (Wedi et al., 2020) have also been conducted for two seasons covering Nov-Feb 2019 and Aug-Oct 2019. The TC statistics of these seasonal simulations are currently being evaluated in collaboration with Reading University and compared with the 9 km resolution equivalents. A special case has also been simulated at 1 km with higher temporal resolution information as a reference example (Hurricane Dorian in late August 2019). Although only a single simulation, it indicates that the track forecast was good when started from the interpolated 9 km operational analysis six days earlier, and the storm had an increased

intensity compared to the operational run, but it also shares features of the described 4 km simulations with deep convection simply switched off such as a cloud filling of the cyclone eye.

Regarding ocean processes, the ECMWF HRES forecast saw an improvement in the intensity forecast with the introduction of ocean coupling in 2018, a result that is confirmed by the no-coupling experiment for our special 37-day period in 2020.  In 2020, a change to the ocean drag in the wave model led to improved forecasts of Vmax, and therefore a better pressure-wind relation. Further developments are under way to improve the heat and momentum exchange under extreme conditions, such as TCs. One needs to be aware of that these types of developments require a coupling to an ocean model to give realistic results in terms of energy exchanges.

The current ocean model has a resolution of around 0.25° globally, which puts it in the eddy permitting regime. To have a fully eddy resolving model would require a resolution of 1/12° or better, which is unrealistic to implement operationally until the next HPC after the Atos upgrade. However, to gain a better understanding of the ocean response of coupled models an informal collaboration with Météo-France, the Met Office and NRL has been established to compare different models with the observational ALAMO float data deployed by the US Naval Academy and Woods Hole Oceanographic Institute. By comparing very high-resolution limited area coupled models with global coarser resolution coupled models, we hope to be able to quantify the importance of horizontal atmospheric and oceanographic resolution for the ocean response in TC conditions.

We have investigated sensitivities to some of the key challenges found in Section 5, such as the slow propagation speed bias, wind pressure-relation and issues in capturing intensification and weakening rates. For the propagation speed, we found improved biases in the 4 km experiment with explicit deep convection. However, this experiment had at the same time the strongest intensity bias, and the worst tropical winds in general. But we also saw positive results from the experiment considering the Coriolis force due to vertical motion, a result that we will explore further.

The wind-pressure relation diagnostic helps to understand model deficiencies in simulating the extreme winds under TC conditions. The change in ocean drag provided significant improvements to the relation in Cy47r1. In the experiments with 4 km resolution, we found further improvement. Together with the result for the ENS control forecast presented in Section 5, this suggests a resolution dependence on capturing the relation. We also found an even stronger relation in the experiment with explicit convection, which could be related to convective wind structures. Finally, in all experiments, there is an underestimation of the relation for TCs in the range 25-40 m/s, something that needs further investigation.

All the steps discussed above are important to explore in order to deliver a km-scale ensemble forecast system with high quality for extreme events such as TCs. Although we have so far mainly investigated standard TC metrics such as position error, Pmin and Vmax, the modelling of fields that contribute to useful probabilistic forecasts of TC storm surge and flooding will become more important, as will be discussed in Section 9.

# 9.    Forecast products

## 9.1.    Current and future forecast products for tropical cyclones

Based on the tracker output (described in Section 3), ECMWF produces different graphical products. For each active TC reported at the initial forecast time by RSMC/TCWC, a combined product with the strike probability or track plume map up to ten days (or out to six days for 06/18UTC forecast runs introduced operationally in May 2021) together with Lagrangian metgrams for intensity (Pmin and Vmax) is computed. One example is shown in Figure 60 (with the track plume). Recently, the TC products from 06/18UTC forecast runs were introduced and can be accessed in the web open-charts or via dissemination together with 00/12UTC forecasts.
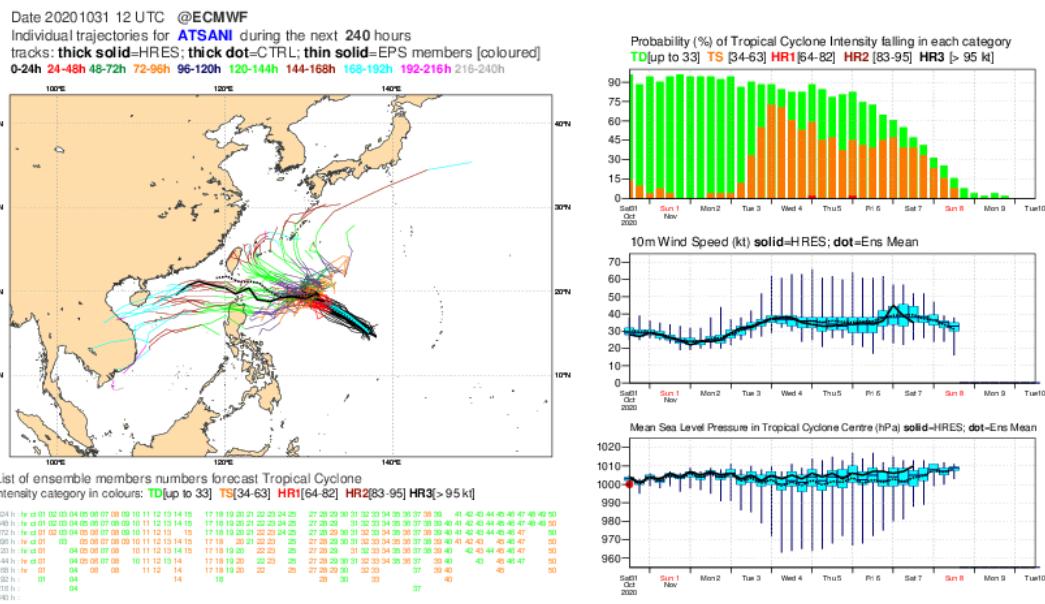


*Figure 60: Lagrangian metgram for TC Atsani from 31 October 12UTC 2020.*

ECMWF is also producing TC activity maps that include genesis (probability maps) both for medium-range forecasts (with a 48-hour time window) and extended-range forecasts (with 7-day accumulation period).

For the extended-range and seasonal forecasts, statistics of basin-wide weekly/seasonal Tropical Storm frequency and ACE (up to 30 days, only Mon and Thu) are also produced.

The ECMWF TC tracking and products are today restricted to the WMO official TC basins. This means that the system does not produce tracks for occasional cyclones in the southern Atlantic, or for "Medicanes" in the Mediterranean. For the Medicanes, the products from the extra-tropical cyclone database (CDB, Section 3.3) can be used, but it could also be considered to extend the TC infrastructure to the Mediterranean.

In the 12-24 h or so spanning an extra-tropical (ET) transition event, Lagrangian products from both the tropical and extratropical cyclone trackers (CDB) are normally available (see Section 3). If users want to continue seeing plots in a similar (tropical) format, they should continue with the TC tracker

output, but must recognise that at some point after ET it will look like the cyclone has decayed, when in fact it may have intensified a lot, as an extratropical feature. If, conversely, users want to see the full life cycle and try to identify a nominal "time of ET" they are recommended to use the CDB plots to see when the TC, denoted as a barotropic low (black dot), changes to a frontal wave cyclone (orange dot). That feature should then continue on as a trackable feature right through its extratropical phase, encompassing any re-intensification (likely reverting to barotropic low form) and also subsequent decay. The main CDB limitation is that it currently only covers part of the tropical north Atlantic, so there are many TCs for which it cannot be used.

As discussed in previous sections, the TC size based on the wind radii forecasts of the 34, 50 and 64 kt became available to users in July 2020. One example of a possible future product for ensembles was shown in Figure 6, and one example following the TC in HRES is shown for hurricane Dorian (2019) in Figure 61. The chart displays the 34 kt wind radii forecast of HRES every 12 h. The product is available from HRES and ENS forecasts in BUFR for dissemination, but not yet in the chart catalogue.
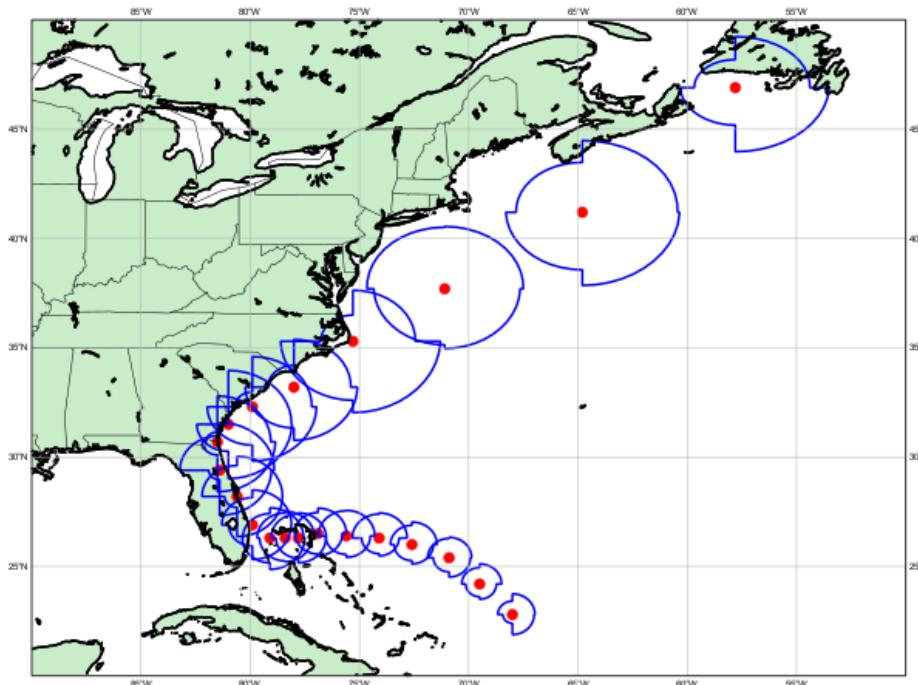


*Figure 61: HRES wind radii forecast for the 34 kt wind threshold out to 240 h, initialised at on 30 August 00UTC 2019. The red dots indicate the predicted centre of hurricane Dorian at 12 h intervals.*

While the TC products only comprise the position and intensity (core pressure, maximum surface wind speed and wind radii) no hazard products associated with TCs are currently available. Work is planned to explore ways to better use forecasts that will help users assess TC hazards. Figure 62 shows an example of the probability of 24 h maximum wind gusts above 20 m/s within a 250 km radius, centred on the position forecast of TC Tauktae (2021) as it progressed northwards, parallel to the west coast of India before landfall near Diu in the state of Gujarat. This highlights the coastal

regions exposed to the impact of cyclone TC Tauktae. Rainfall is also a significant TC-related hazard (discussed further in Section 9), and future forecast products may include charts specifically for TC rainfall.
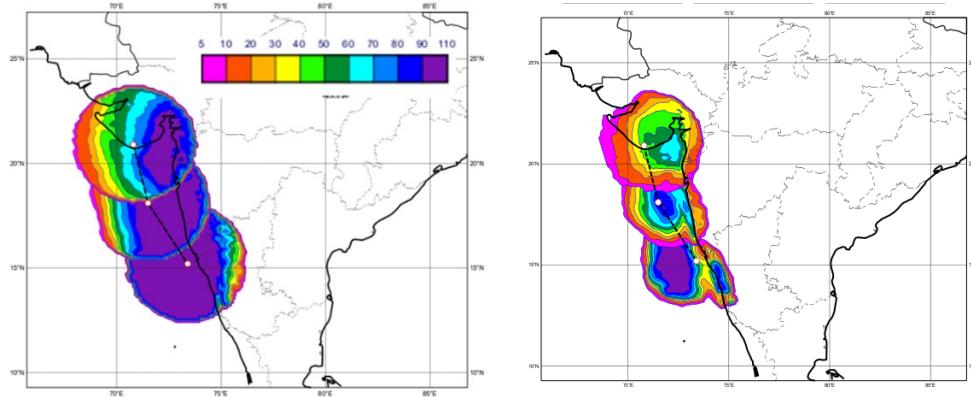


*Figure 62: Probability (%) of 24 h maximum wind gust exceeding 20 m/s (above Tropical Storm strength threshold – 17 m/s) centred at 0000 UTC on 16, 17 and 18 May 2021, based on the forecast initiated at 1200 UTC on 15 May (left) and probability of total precipitation exceeding 100 mm/24 h (right) valid for the same dates. Symbols connected by a dashed line correspond to the HRES position forecast (close to the ENS mean track) valid at the same dates. Probability circles are produced for a radius of 300 km.*

## 9.2.    Future clustering of tropical cyclone tracks

Despite the significant progress made in position forecasts of TCs in recent years as a result of better NWP models, cases of large uncertainty regarding the track still occur (e.g., Magnusson et al., 2019b), especially when the TC passes close to a bifurcation point in the steering flow. Figure 60 shows such a case for TC Atsani (2020). While some ENS members depict the TC moving towards the Philippines, other members had a curvature to the north. In these situations, cluster analysis techniques can provide better guidance to forecasters under multi-modality track scenarios (Tsai and Elsberry, 2013; Kowaleski and Evans, 2020), and help to depict intensity uncertainty for example.

For extended-range TC prediction, clustering of TC tracks also has potential. Such a product would help address some of the issues linked to model biases (e.g., too rare TC landfall in extended-range and seasonal forecasts because of a too early TC track curvature) by using the dynamical model to predict the probability of a given cluster and a statistical method to predict its impact (e.g., probability of landfall). This dynamical-statistical method would be similar to the use of weather regimes in the Euro-Atlantic sector to predict the probability of extreme events in Europe.

Camargo et al. (2021) applied a clustering technique, described in Gaffney et al. (2007), to observed TC tracks and the ECMWF extended-range reforecasts produced in 2018 over the North Atlantic. This analysis indicates that there are clear spatial differences; the ECMWF model having an additional cluster with recurring tracks close to the African coast, with characteristics that do not correspond to observations. In addition, the relative population of the clusters is not realistic in the ECMWF extended-range forecasts, with, for instance, too few TCs in the cluster containing Caribbean TCs. This study also showed that the clusters have different levels of predictability since

they are differently modulated by sources of predictability such as the MJO, ENSO, AMM (Atlantic Meridional mode) and the NAO.

## 9.3. Discussion

Based on TC tracks, ECMWF provides a set of products freely available on the ECMWF website for different time-ranges. However, there are a lot of other products related to tropical cyclones that are currently not produced by ECMWF. As the tracks are also provided to external users, via GTS or FTP, there are many other sites that display ECMWF TC forecasts, either separate or as a part of a multi-model ensemble. The open question here is how much resource ECMWF should spend on increasing the graphical product catalogue for TCs.

# 10. Applications: Impact Forecasting

## 10.1. Introduction

Often, the first thing that comes to mind when considering TC impacts is the winds, and this has typically been the focus of intensity forecasting. However, the most dangerous impacts of TCs are often those related to water, including various types of flood hazard: flash flooding, river flooding, and flooding from storm surge and extreme ocean waves. These water-related hazards require additional modelling capabilities to provide forecasts of storm surge and rainfall-induced inundation. As discussed in Section 4, traditionally the focus in terms of TC predictability and predictive skill has been on track prediction, and is increasingly moving towards structural characteristics and intensity as model resolution has increased and it has become possible to capture mesoscale features. Increasingly, emphasis is also moving towards hazard- and impact-based forecasting, aiming to narrow the gap between science and decision-making for early action based on forecast information. For TCs, this means incorporating forecasts of not only track and intensity, but also wind fields, precipitation and flooding, and another step further, risk information such as populations and infrastructure exposed to these hazards.

The downstream hazard models required for predicting water-related TC hazards rely on accurate forecasts of rainfall and other meteorological variables from numerical weather prediction models and highlight the importance of accurate rainfall prediction for TCs. It is known that rainfall amounts are not directly related to the intensity of TCs, but the translation speed and size of the TC alongside topography and geography of the landfall region are instead key. Therefore, predicting TC rainfall relies on several factors: TC track, intensity, size, structure, and interactions with land and the wider atmospheric circulation (Titley et al., 2021). Furthermore, flood severity from TCs depends on additional factors including the TC duration, total precipitation, and catchment characteristics such as antecedent soil moisture and orography height and gradient (Titley et al., 2021). An important avenue for future work is to assess the predictability and predictive skill of TC precipitation and flood forecasts, in order to fully understand whether forecast skill is enough for accurate and useful impact-based forecasting.

Current capabilities, ongoing work and future directions for hazard forecasting and impact modelling in the context of TCs are discussed in the following subsections, and Section 9.5 provides examples and case studies of the use of ECMWF forecasts for such applications and decision-making.

## 10.2.    River Flooding

One aspect of hazard forecasting for water-related TC impacts is forecasting river flow, including increased river flow and flooding from TCs. As part of the Copernicus Emergency Management Service (CEMS), ECMWF produces global flood forecasts, using a hydrological model forced by the ECMWF ensemble forecasts (The Global Flood Awareness System, GloFAS). GloFAS v3.1 (Copernicus, 2021; www.globalfloods.eu), implemented on 26 May 2021, uses a hydrological modelling system based on the open source global LISFLOOD model (Van der Knijff et al., 2010). LISFLOOD is a spatially distributed rainfall-runoff-routing model, developed at the European Commission's Joint Research Centre (JRC). ECMWF's 00UTC ensemble forecasts (for precipitation, temperature, dew point temperature, 10-metre u and v wind components and surface net solar and thermal radiations) are used to drive the LISFLOOD model, which calculates a complete water balance and produces forecasts of runoff at each grid cell (0.1° resolution). It then routes this runoff through the river network, producing forecasts of river flow. The initial conditions for GloFAS are primarily taken from the GloFAS-ERA5 river flow reanalysis, produced using the GloFAS modelling chain forced with ERA5 data. Harrigan et al. (2020) provide an overview of GloFAS-ERA5 production and evaluation, based on the GloFAS v2.1 modelling chain, and the GloFAS Wiki (Copernicus, 2021) provides details of the different GloFAS model versions.

Each new GloFAS forecast is compared against flood thresholds, calculated from the GloFAS-ERA5 reanalysis for various return periods, to provide the probability of exceeding various flood severity thresholds, out to 30 days ahead. These probabilities are provided through different products including maps, hydrographs and persistence plots, amongst various other hydrological and meteorological forecast layers. This approach based on exceedance probabilities limits the influence of systematic biases, which are expected in regions where the model remains uncalibrated. At present, GloFAS does not provide forecasts for coastal or pluvial flooding, which are also of significant concern during TC landfalls.

GloFAS forecasts have been used operationally for forecasting river flooding from TCs, in combination with a detailed flood inundation model in order to predict population exposure (see Section 9.5 below). Further research as part of an ongoing collaboration between ECMWF and the University of Reading, is evaluating the ability of GloFAS to forecast fluvial flooding from TCs by examining all parts of the forecast chain, including the track, intensity, precipitation and hydrological components. Recent research published in Titley et al. (2021) has identified key factors that influence the severity of TC flooding, such as translation speed and river catchment characteristics, and ongoing work aims to verify ensemble precipitation forecasts in TCs and highlight the main influences on predictability of fluvial flooding from TCs, in order to improve flood forecasts and their use for decision-making during TCs. Initial work based on a case study of Hurricane Iota (2020) indicated that in one river, the key driver of good flood forecasts was the accurate prediction of the broader onshore atmospheric flow into the mountains and was less sensitive to the prediction of landfall location. In another river, successful flood forecasts relied on the accurate prediction of landfall location and of heavy rainfall close to the TC centre, which is less predictable at longer lead times (Figure 63).
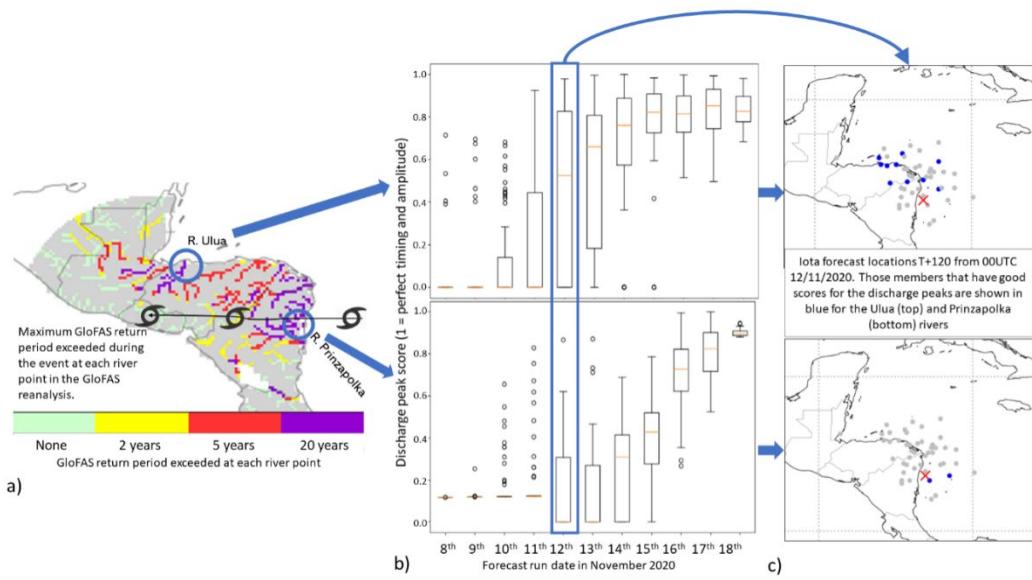
*Figure 63: Initial analysis of the GloFAS forecast predictability for Hurricane Iota: a) the GloFAS return period exceeded at each river point in the Iota flood event, using the GloFAS-ERA5 reanalysis; b) box-whisker plot showing the range of discharge peak scores across the ensemble for two river points (circled in blue in (a)) for all forecast runs in the lead up to the flood event; c) Iota track forecast centre locations at T+120 from 0000 UTC 12th November 2020, with the ensemble members that displayed good forecast skill highlighted in blue.*

As part of a wider evaluation and discussion of flood forecast bulletins for Cyclones Idai and Kenneth in Mozambique in 2019 (discussed further in Section 9.5), Emerton et al. (2020) evaluated the chain of forecasts from TC tracks and rainfall to river flow and flood inundation forecasts. The location of the storm is key for both precipitation and flood forecasts, and it is important to consider that the track forecasts indicate only the centre of the TC, but winds and rain associated with the storm extend much further. This was a consideration when these TCs made landfall, as in both cases, track forecasts indicated that the storms would continue to move further inland before dissipating. However, both Idai and Kenneth stalled after making landfall, resulting in sustained periods of heavy rainfall over the same region rather than smaller rainfall amounts spread over a larger region. This stalling was only picked up ~one day ahead and resulted in uncertainty in the rainfall and flood forecasts, in terms of the probability of severe flooding and the rivers which were likely to flood. This is seen in Figure 64(b), which shows under-estimation of rainfall at two days ahead (red shading) nearer the coast where Idai stalled (and over the Mozambique Channel), and an over-estimation further west due to track forecasts indicating the storm would move further inland. This had implications for decision-making based on the associated flood forecasts, and again highlights the importance of more evaluation and diagnostics accounting for various aspects of TC forecasts including hazards beyond the track and intensity.
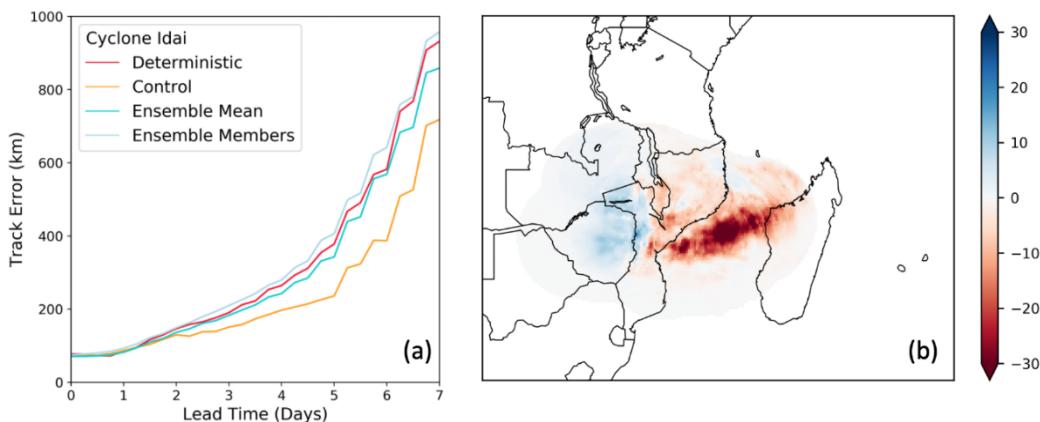
*Figure 64: (a) Track location errors (km) with lead time for Cyclone Idai, Mozambique March 2019 and (b) TC-related precipitation errors (mm/day) for all 2-day ensemble mean precipitation forecasts throughout the lifecycle of Cyclone Idai, where red indicates an under-estimation of rainfall, and blue an over-estimation, compared to GPM IMERG satellite rainfall data. Figure adapted from Emerton et al. (2020).*

## 10.3.    Inundation Forecasting/Impact Modelling

Beyond forecasts of river flow, in order to move towards hazard and impact forecasting, it is important to provide information for decision-making purposes on the inundation area expected from the predicted river flows, and on the associated risks, such as population and critical infrastructure that may be exposed to flooding.

In addition to forecast products indicating predicted river flow and threshold exceedance probabilities, GloFAS has a risk mapping component showing the estimated maximum predicted inundation extent at a 1 km scale, and associated level of risk based on the exposed population (using the Global Human Settlement Layer dataset) and timing of maximum flood hazard. These products are available for river points with an upstream area greater than 5000km$^2$, where the flood hazard is predicted to exceed the 10-year return period threshold during the 30-day forecast horizon. Flood risk is estimated per global administrative region and provided through the 'rapid impact assessment' forecast layer of the GloFAS interface (Figure 65). Additional information on land cover types and critical infrastructure (using data from e.g., the European Space Agency and OpenStreetMap) is also provided for each region with an expected impact. The risk is assessed to be higher for events predicted with a shorter lead time, and with larger exposed populations. Further details are provided via the GloFAS Wiki (Copernicus, 2021), and limitations of this approach and the global datasets used are discussed on the GloFAS website (Copernicus EMS, 2021). While these products are not specific to TCs, they provide added information at the global scale that is applicable for predicting river flooding impacts from landfalling TCs.
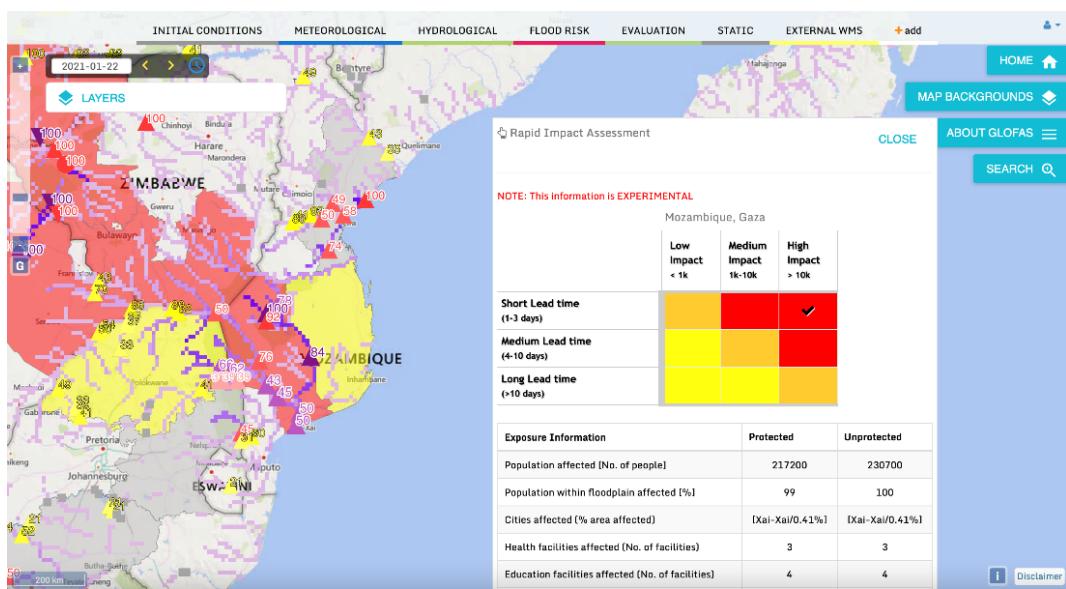
*Figure 65: Screenshot of the GloFAS interface for the forecast from 22 January 2021, 1 day ahead of Cyclone Eloise's landfall in Mozambique. The map indicates rivers where flooding is expected to exceed the 20-year return period (purple shading, where darker colours indicate higher probabilities), and the rapid impact assessment layer (red and yellow shading). The box provides additional rapid impact assessment information, shown here for the Gaza administration region in Mozambique, indicating a high potential impact at a short lead time on the impact matrix, and estimates of the population affected.*

In the future, it may be possible to incorporate flood inundation mapping at much higher resolutions e.g., through assimilation of satellite derived inundation extents, or use of river flow forecasts to drive a dynamical hydraulic model and provide inundation forecasts that directly relate to the probabilistic river flow forecasts. It will also be important to incorporate further information on critical infrastructure and vulnerability, which can be significantly impacted by both winds and flooding during TCs and post longer-term risks beyond the duration of the storm. A similar approach to that described here has been used operationally to forecast flood inundation and population exposure from TCs, in collaboration with the University of Bristol, for early humanitarian action and response. This is discussed in section 9.5.

## 10.4. Storm Surge Forecasting

With rising sea levels and increasing populations in coastal regions (Neumann et al., 2015), flooding from storm surge, which can be destructive and dangerous, is a key consideration for TC hazard and impact forecasting. The inundation risk from storm surge is also affected by other factors such as astronomical tides, waves, river flow and precipitation (Kohno et al., 2018). As discussed in Section 5, the ECMWF forecasts include an ocean wave model. ECMWF does not produce any storm surge forecasts, but other institutions run storm surge models using forcing from ECMWF meteorological and ocean forecasts. There are also steps to build full inundation model systems based on forcing from ECMWF forecasts (Zhou et al., 2020). This is important for risk communication, in order to convert

forecasts of storm surge into estimates of total sea level and inundation, but it requires detailed information on land elevation which is not always available (Kohno et al., 2018).

There are two distinct approaches to the use of atmospheric forcing for storm surge modelling. The first is to use only the forecasts of TC position and create an artificial wind field around the TC, and the second is to use more realistic wind fields directly from the forecasts.

The first approach uses the track forecasts for TCs and derives wind fields based on parametric models, for example using empirical pressure profiles such as Holland (1980) or Fujita (1952) to define the sea level pressure in the region of the TC. Gradient wind relations can then be used to estimate the wind fields, and adding the forward motion of the TC gives an asymmetric wind field. This approach has been used historically because NWP systems were unable to provide a realistic enough wind field to produce useful storm surge forecasts. While TC forecast skill for track and intensity has been continuously improving, it is often insufficient for accurate storm surge predictions (Kohno et al., 2018), and as such, these parametric modelling approaches are popular. A similar approach is being trialled by HR Wallingford using ECMWF's forecasts as part of a pilot study by the UK government for early humanitarian action, which is discussed in Section 9.5. This trial looks at moving from using a deterministic TC track forecast from the relevant RSMC, to additionally using ensemble TC track forecasts from ECMWF to provide probabilistic storm surge forecasts. For this application, the Holland (1980) profile model is used to determine the wind field, and the forward motion of the storm is added to give an asymmetric wind field, with the wind directions corrected based on the inflow angel. The TELEMAC2D model is then used to simulate storm surge. There is further potential to use ECMWF's new wind radii forecast products to provide a more accurate TC structure.

However, NWP systems moving to higher resolutions and providing more realistic simulations of TC winds provides new opportunities to use the wind fields directly from NWP models to force storm surge models. This second approach is particularly desirable in cases when TC structures vary from the standard structure used in parametric models. Moreover, it is known that the state of development of the wave field can have quite some impact on the actual storm surge intensity (Bertin et al., 2015). Therefore, ECMWF coupled forecasts can also provide consistent information on the sea state and surface fluxes for better storm surge forecasts.

Development of a storm surge forecasting system based on NWP, with global coverage (GLOSSIS, the Global Storm Surge Forecasting and Information System, (Deltares, 2021)), is also ongoing. This configuration uses meteorological forcing and the Global Tide and Surge Model (GTSM), alongside Delft-FEWS, to produce 10-day water level and storm surge forecasts for ~16,000 coastal segments around the globe. The model uses an unstructured grid, with coastal areas represented at ~5 km resolution and open oceans at 50 km resolution, and provides return period exceedances based on return periods derived from reanalysis data. GLOSSIS currently makes use of NOAA's GFS forecasts for the meteorological forcing, but also has functionality to use ECMWF's forecasts alongside track forecasts from the JTWC.

ECMWF has received a request by the Norwegian met service to explore the feasibility of using NEMO capabilities to provide storm surge information derived from the ensemble runs. This activity is currently at a very early stage but might prove useful in deriving storm surge early warnings at a

global scale. Additionally, the JRC has been developing a new unstructured mesh model to simulate storm surge globally, forced with ECMWF's operational HRES forecasts and ERA5 for testing. While this new model is currently under development, in future it could be linked with GloFAS to provide information on coastal flooding, which, as mentioned in Section 9.2, is not currently provided by GloFAS but would be further beneficial for forecasting flood impacts from TCs.

## 10.5. Examples of forecast uptake for decision-making

ECMWF's medium-range TC forecast products are used by national meteorological services and the RSMCs around the globe for operational TC forecasting, warning and decision-making, alongside seasonal forecasts of TC activity in each ocean basin. As mentioned in Section 9.2, forecasts of rainfall and flooding from ECMWF and GloFAS are also used for hazard and impact forecasting ahead of tropical cyclones. Here we provide some recent examples of the combined use of ECMWF TC and associated hazard forecasts for decision-making purposes.

### 10.5.1. Forecast-based Financing using GloFAS flood forecasts

GloFAS forecasts have been used to take early action ahead of flood events since 2015, when humanitarian action was triggered by a GloFAS flood forecast for the first time, ahead of flooding in Uganda, allowing aid to be distributed before flooding began. Typically, humanitarian aid reaches communities after a disaster has occurred, but the forecast-based financing programme (IFRC, 2021) enables access to humanitarian funding based on forecast information and risk analysis, using specific forecast thresholds that trigger the release of pre-agreed financial resources set out in an Early Action Protocol (EAP). In some regions, it is possible to trigger early humanitarian action for flooding from tropical cyclones based on forecasts. Research into the use of GloFAS forecasts for early humanitarian action is carried out at the University of Reading through the FATHUM project, in collaboration with CEMS and ECMWF.

### 10.5.2. Emergency flood bulletins for tropical cyclones

Researchers at the Universities of Reading and Bristol in the UK collaborated with ECMWF in 2019 to assist in providing flood forecast information for TCs Idai and Kenneth, which made landfall in Mozambique with devastating impacts. After TC Idai made landfall in March 2019, the President of Mozambique declared a state of emergency and requested international assistance. The UK government's Foreign, Commonwealth and Development Office (FCDO, previously the Department for International Development, DFID) tasked a team of scientists from the two Universities to produce real-time flood forecast bulletins to support humanitarian decision-making during the flooding that followed Idai's landfall. Using ECMWF TC track and rainfall forecasts and GloFAS flood forecasts, with the support of ECMWF, the Universities produced daily bulletins detailing the expected location and timing of landfall, alongside the rivers most at risk of flooding and the predicted timing of the flood peak and recession. This was based on the forecasts and methods described in Section 9.3. ECMWF also provided GloFAS river flow forecast data, which was used to produce inundation and population exposure information using a model framework similar to that described in Section 9.3, based on a hydrodynamic model with the LISFLOOD-FP code, and global scale gridded population data from the High-Resolution Settlement Layer (HSRL) dataset.

The objective of the bulletins was to facilitate decision-making, such as the distribution of resources ahead of flooding, and increase understanding of the situation and the areas most at risk of flooding and when, to be used alongside forecast and warning information from the mandated regional and national authorities. The flood bulletins, modelling framework, and an evaluation of the forecasts and the bulletins themselves, are described in detail in Emerton et al. (2020).

Six weeks after TC Idai struck Mozambique, the UN Office for the Coordination of Humanitarian Affairs (UN OCHA) and the UK government requested re-activation of these emergency flood bulletins ahead of TC Kenneth, which was forecast to impact northern Mozambique. This allowed forecast information to be used for humanitarian decision-making ahead of the TC and consequent flooding, and further led to the introduction of a pilot project by FCDO to operationalise the production of emergency flood bulletins by scientists at the Universities of Reading and Bristol, and HR Wallingford, with the support of ECMWF. This pilot project was activated a number of times during 2020 and 2021, including for TCs Iota, Amphan and Eloise. ECMWF provided forecasts, data, expertise and training in relation to this pilot project. This pilot project also allowed for the inclusion of storm surge forecasting by HR Wallingford, based on official track forecasts from the RSMC in La Réunion. As discussed in Section 9.3, this storm surge forecasting is being tested using ECMWF TC track forecasts, with the potential to further include the new wind radii products. The inclusion of storm surge information allows a more complete overview of the hazards associated with TCs, from both riverine and coastal flooding. At present, the outcomes of the pilot project are being evaluated in consideration of the next steps for the production of such bulletins.

### 10.5.3.  The ARISTOTLE Consortium

ECMWF has also provided emergency reports for six TCs that included flood impacts since October 2020, using GloFAS flood forecasts and a range of TC forecast products for track and winds, including from ECMWF and the Met Office. These reports were produced as part of the ARISTOTLE (All Risk Integrated System TOwards Transboundary hoListic Early-warning) consortium, of which ECMWF is a partner alongside 17 other institutions. ARISTOTLE is a multi-hazard partnership covering expertise across multiple hazard groups (severe weather, flooding, forest fires, volcanoes, earthquakes and tsunamis). The institutions in each hazard group work in rotation to provide routine monitoring reports (three times per week) and emergency reports (within three hours) when activated by the European Commission's Emergency Response Coordination Centre (ERCC). ARISTOTLE is funded by the EU Directorate General for Humanitarian Aid and Civil Protection (DG ECHO) and is set to continue for the next five years.

### 10.6.  Summary

Current capabilities, ongoing work and future directions for TC hazard forecasting and impact modelling have been discussed in this section. Often, the focus for TC forecasting is on the prediction of TC track and intensity, which are key in terms of decision-making and in terms of the accuracy of other forecast products and models. However, water-related hazards due to intense and prolonged rainfall, and flooding including flash flood, river flood and coastal flood hazards, can be the most dangerous aspects of TCs. It is therefore important to move towards provision of forecast information that can be used for these hazards. Hazard forecasting and impact modelling often requires the use of

additional models such as hydrological, hydraulic and storm surge models, typically driven with the output of NWP models such as ECMWF's IFS. In this section, we provided an overview of the forecasting capabilities for river flow and river flood forecasting, and inundation modelling and exposure estimates, through the Copernicus Emergency Management Service (CEMS) Global Flood Awareness System (GloFAS), in the context of TCs. ECMWF does not currently have capabilities in storm surge forecasting, but other organisations make use of ECMWF forecasts to drive storm surge models. Some recent examples of the application of ECMWF forecasts for decision-making were discussed further, with a focus on hazard and impact forecasting for TC flooding. One aim of the WMO/WWRP project HIWeather is to evaluate the full forecast production chain for extreme events, including TCs, and ECMWF is participating in this work.

In order to provide useful input to downstream models, ECMWF needs to make sure that both the forecast quality and the forecast output is accurate. For the forecast quality, more focus on the precipitation and wind extent around TCs is needed in the evaluation, and model development is necessary. Regarding the model output, close communication with forecast users is necessary to make sure that the model output frequency, parameters and products meet their needs.

## 11.    Concluding remarks and avenues for improvement

### 11.1.    Current progress and challenges

The performance of forecasts for TCs needs to be viewed from the perspective of the **overall performance of the forecasting system**. The performance of tropical winds and the extra-tropical waveguide will have a large impact on TC propagation and the humidity over the tropical oceans. However, there are some aspects that are more germane to TCs than other systems. Expanding on and updating two recent articles on ECMWF activities (a report on data assimilation by Bonavita et al., 2017 and an overview article by Magnusson et al., 2019), the aspects of the overall forecasting system most relevant to TCs have been presented in this report and are summarised below.

**Observations** are used for multiple purposes, including the "Best Track" estimation of TC position and structure (provided by RSMCs) which is used for verification; case study diagnostics; and of most importance to ECMWF's forecasts, data assimilation (summarised further below). For TCs, there is an especially large reliance on satellite data. Additionally, specially tasked aircraft are sometimes deployed, mostly in the Atlantic basin, to provide data on the TC and its environment. Ongoing challenges with observational data include the timely processing of incoming data, quality control schemes, handling faulty observations, managing observation operators and specifying their errors, prescribing errors and biases of the observations (which may vary as a platform ages), and accounting for correlated observation errors. Moreover, old observation platforms can cease to exist, sometimes suddenly, and new observation platforms are introduced each year.

**Data assimilation** at ECMWF involves the ingestion of over 800 million observations each day, of which more than 60 million are processed by the data assimilation.  Given that forecasts of TC **position** are determined by the steering flow over a large region, several million observations have an influence on these forecasts. It is therefore difficult to quantify the impact from single observation systems on forecasts of TC position. For short-range forecasts of TC intensity (Pmin or Vmax), an

accurate initial representation of the conditions within and near the TC is especially necessary. However, the hostile environment within a TC and the cloudy conditions around the TC substantially limit the coverage of both conventional and satellite data. Here, the impact of single observation systems might be more easily detected. From the special observing system experiments conducted for the August-September 2020 period, improvements to the short-range intensity forecast errors were evident due to the assimilation of all-sky microwave observations, COSMIC2 radio occultation, and surface wind scatterometer observations.

Several recent improvements to the **data assimilation methodology** at ECMWF have provided benefits for TC analyses and forecasts. In the vicinity of TCs, data assimilation is challenging due to the sharp gradient and lack of observations. One recent improvement has been to increase the **resolution of the Ensemble of Data Assimilations** (EDA), which improved the background error statistics for TCs (Holm et al, 2015). However, this also led to larger weights for observations close to the TC. As the drift of dropsondes was not accounted for, very large observation errors occurred that occasionally had a severe effect on the analysis. This issue was mitigated with **adaptive first guess quality control.** Also, since the introduction of the BUFR format for dropsondes has made it possible to report the position for each observation, the position was updated during the ascent. Even if it is difficult to measure the impact on the overall skill, these improvements are expected to avoid damaging situations. A remaining danger for assimilation of data within TCs is occasionally **faulty observations**, from surface synoptic observations (SYNOP), buoys, or incorrect positions of ships or land stations. On a few occasions in recent years, the assimilation of these faulty observations has had a severe impact on TC initial conditions.

Data assimilation schemes in the ECMWF system and other global NWP systems all possess substantial **limitations in TCs**. These limitations include the resolution of the data assimilation scheme, some components (such as tangent linear models) that assume linear dynamics, and background error covariances that may not reflect the main dynamic and thermodynamic balances in TCs. This is further compounded by the relative paucity of observations within the TC. These limitations lead to several consequences. For an intensifying TC, especially one that is small and/or intensifying rapidly, small-scale processes are especially important. The successive analysis cycles are often found to be **unable to keep up with the actual intensification of a TC**. As a result, the short-range forecast that provides the background field for assimilation is too weak, and the subsequent analysis TC structure (and intensity) drifts even further behind the actual TC structure (and intensity). These difficulties are not so severe for large, relatively symmetric TCs with eyewalls that are resolvable. Some similar challenges exist in weakening TCs, where the weakening rate in successive analyses is not as sharp as the actual weakening rate.

Challenges also exist in the **use of observations** within the TC. In the assimilation experiments, we have not yet investigated specifically how observations served to correct the analysis structure in the TC. Equally importantly, we would need to investigate those observations that did not pass the quality control. Given the very large discrepancies between observed values and those in the background field, especially for an intense hurricane, several high-quality observations may fail the quality control process or not be weighted substantially in the assimilation. However, this is likely a necessary practice to prevent overly drastic changes to the analysis fields, which could cause more harm than if these extreme observations were excluded.

To help provide initial conditions of TCs that are more consistent with forecasters' estimates of the (working) Best Track, other NWP centres directly **assimilate Best Track estimates**. These estimates include uncertainties, and as they are subjective, different practices can lead to different error characteristics. Following the method at the Met Office (Heming, 2016), new assimilation experiments were conducted here to use the Best Track estimation of central pressure as a regular surface observation. From these experiments, no statistically significant results were consistently evident, although improvements did occur for a small sample of strong TCs in the north-west Pacific basin.

Recent **model developments** have led to the improvements of TC forecasts. The introduction of **ocean coupling** has improved intensity forecasts, via avoiding over-deepening of TCs. The introduction of a **cap on the surface drag coefficient** during very high wind speed regimes has helped improve the forecast of the maximum wind speed, and the wind-pressure relation. With these recent improvements, it is also time to raise the bar and evaluate the **structure of the surface wind speed**. Since Cy47r1, the tracker output includes the radius of different wind speed thresholds (including tropical storm force and hurricane force).

In the 2020 operational configuration of HRES with 9 km resolution (Cy47r1), the bias of the central pressure (Pmin) appeared close to zero. However, there were strong conditional biases where initially weak TCs developed a bias toward too shallow TCs (underpredicting intensification rates), and initially strong TCs developed a bias towards too deep TCs (underpredicting weakening rates). As expected, the bias in the maximum wind speed (Vmax) was consistently low, regardless of the initial intensity of the TC. We note that Vmax can be dominated by small-scale (O(1 km)) processes.

The question of whether an increase in the **model resolution** improves TC intensity forecasts has been widely investigated in the NWP community. In this study, experiments with 4 km resolution have been tested over the special 37-day period in August-September 2020. Using the 2020 operational physics package, the 4 km experiments provided TCs that were too deep on average. However, when the 9 km and 4 km experiments were repeated with the **new physics package** intended for Cy47r3, the intensity was on average reduced, leading to an increased bias at 9 km resolution but a reduced bias for 4 km. The intriguing result was that the mean absolute error improved for both resolutions, by reducing the conditional biases discussed above. We should therefore be ready for an increase in intensity bias with Cy47r3 in the knowledge that it is probably due to unresolved structures in the 9 km resolution model and compensation of errors. However, we still believe that there are several compensating errors in the model that need to be further explored.

The wind-pressure relation, which has been a long-standing problem in ECMWF forecasts and other global models, was improved in Cy47r1 as mentioned above. In the 4 km experiment the relation is further improved, demonstrating that the model resolution is important to predict not only the maximum wind speed (Vmax) and central pressure (Pmin), but also the relationship between the two. However, there is still an underestimation for wind speeds of 30-40 m/s, which will need to be further investigated in the context of modelling the drag from ocean waves.

A long-standing **systematic error in ECMWF TC track forecasts** is a too slow propagation speed on average (Figure 32 in Haiden et al., 2021) and a tendency to curve too much to the right of the observed track in the tropics. This bias is difficult to target as it depends on the sample of test cases,

and it is relatively small in magnitude compared with the random component of the error. The same argument holds in the context of a visual inspection of weak, westward-moving TCs in the Atlantic basin that often drift slightly to the right (north) of the actual track, even at early forecast times. Although there are difficulties in finding statistical evidence, one can expect these biases to affect forecasts of not only TCs but extra-tropical transitions and downstream impacts in the mid-latitudes. In our experimentation, we found encouraging results by including the Coriolis effect due to vertical motion. This is expected from theoretical work (Liang and Chan, 2005), and seems to reduce the bias in the IFS but more experimentation is needed to confirm the results.

For **long-range predictions** of TC activity, the forecasts do not reproduce the pattern over the Atlantic. They overestimate the activity over the central Atlantic, and underestimate the number of TCs in the Gulf of Mexico. This is an obstacle to the goal of predicting the risk of landfalling TCs in the coming season. An ECMWF team that was recently tasked with investigating the seasonal forecasts found a trend of increased wind shear over the tropical Atlantic in the model that is not present in the reanalysis. This has led to an underestimation of the TC activity in recent years. This finding about the wind structure also opens the much wider question about the global warming response in the model and might also have implications for climate models.

Predicting TC **genesis** is important both in the medium- and extended-range. However, it is difficult to draw general conclusions about the performance, as the variability in multi-scale mechanisms and predictability between individual cases is very high. The predictability varies from more than a week in advance to not even capturing the TC at the formation time. In situations of higher predictability, there is often a relatively large-scale, coherent precursor disturbance. In situations of lower predictability, the precursor disturbance is often less defined, and there may be a greater dependence on bursts of convection on short time scales.

## 11.2.    Avenues for future improvement

Based on evolving user needs, the challenges described in the individual sections and summarised above, and the continuous progress in TC forecasting at other NWP centres around the world, several recommendations for improvement are provided in this closing sub-section. Some of the items below are related to planned activities at ECMWF and some are to consider for future plans.

First, a **structured evaluation framework** for TCs that advances beyond the current framework is proposed, to monitor progress and identify further challenges:

- For **position**, evaluate errors for paired and homogeneous samples, including metrics for along-track and cross-track forecast biases, and propagation speed.
- For **intensity**, evaluate the minimum central pressure, the maximum surface wind, and the wind-pressure relation.
- For **surface wind structure**, continue the recent development of evaluation techniques for wind radii (34 kt, 50 kt, 64 kt), and explore the radius of maximum wind.
- For **other aspects of TC structure**, explore verification against infrared and other satellite images by using simulated images.
- Include **TC activity** in the routine verification, investigating both predictive skill and bias, and accounting for environmental variables (such as wind) that affect the detection threshold.

- Probabilistic verification of **TC genesis**, using related tools to those used for TC activity.
- For **impact related metrics**, investigate the possibility of rainfall verification.

New methods to **evaluate model structures** require investigation, given that the Best Track values only represent a small number of characteristics related to the surface pressure and wind. Examples include comparisons of simulated infrared images versus actual images, comparisons against dropsonde and other aircraft data, and comparisons against specialised satellite data such as SAR. For the evaluation of both future assimilation and model improvements, TCs put special constraints on the testing procedure since the resolution is expected to be especially important.

For **impact forecasting**, accurate modelling of the structure of the wind and precipitation fields is necessary. An additional requirement in **storm surge** forecasting is for accurate predictions of the timing of landfall and the corresponding wind fields, given that the storm surge is superposed on the astronomical tide. Ultimately, for **flood** forecasts based on both storm surge and TC precipitation, the different model components will form a basis for inundation models that aim to predict water levels down to street resolution. However, this will not only demand high-quality TC forecasts, but also the processing of ensemble data at high spatial and temporal resolutions.

For the **data assimilation**, several aspects were not tested here, due to the substantial effort and research required. It includes **modifications to the data assimilation** scheme itself, such as the resolution of the loops in the 4D-Var algorithm, error covariance matrices that account for the dynamics of the TC structures and the observation quality control. We believe these aspects could have a relatively strong impact on TC analyses and forecasts, and should therefore be evaluated with special attention.

For **observation usage in the data assimilation**, the following priorities have been identified:

- Increase the resilience to faulty surface observations in the tropics.
  - The avoidance of such cases will be targeted via the continued development of the observation monitoring and quality control, with help from machine learning algorithms.
- Increase the use of observations close to the TC.
  - Examples include adaptive thinning of satellite observations to capture finer-scale structures in TCs, a continued move to all-sky assimilation of satellite radiances (e.g. infrared channels) and exploring the increased use of atmospheric motion vectors near the TC.
- Explore new satellite products.
  - Examples with marine wind information at high windspeed include passive L-band (SMOS, SMAP, CIMR), building on work carried out in the SMOS collaboration, and Synthetic Aperture Radar (SAR), building on the work already done at Météo-France.
- Continue the initial exploration of assimilation of Best Track.
  - The work should be conducted with the awareness that finding the "correct" way to use the data can be a time-consuming task.

> o   Establish contact with Met Office and other centres to exchange knowledge on the topic

Several other new and future satellite observation types and algorithms may enhance analyses and therefore forecasts of TCs.  Examples include new LEO satellites, targeted assimilation of fine-scale structures now discernible in geostationary satellites, rapid-scan AMVs, future Lidars, ocean surface wind retrievals, small satellite constellations such as TROPICS and CYGNSS, and new remotely sensed and in situ ocean observations.  The effective assimilation of all these observation types not only depends on their accuracy, but their availability on the GTS in real-time together with the provision of error characteristics and observation operators.

Further investigation is required on how reconnaissance and surveillance aircraft data are enhancing the analysis, and whether new methodologies would make better use of these specialised, high-quality observations. In addition to the in-flight data mentioned in the bullet point above, the large volume of aircraft data that are not presently assimilated, including airborne Doppler radar data, may substantially improve analysis structures of TCs. Future datasets to be considered for assimilation include Stepped Frequency Microwave Radiometer (SFMR) data at the ocean surface, and data from a variety of unmanned aircraft.

The ongoing **model development and experimentation** at ECMWF is expected to provide long-term improvements to TC forecasts. These include not only the resolution and physics, but also continued improvements to ocean coupling, wave modelling, and representation of the atmospheric boundary layer. Continued experimentation with nonhydrostatic modelling, a finite-volume dynamical core, and explicit convection, is strongly encouraged in the context of improving TC forecasting. Some priorities to address include:

*   Improved modelling of the propagation speed
*   Further investigation of the tendency to a right-hand track bias
*   Understand the causes behind the too slow intensification and decay of TCs
*   Understand the deficiencies in simulating the TC climatology for Atlantic sub-basins in seasonal forecasts

For **seasonal forecasting**, it is also important to further investigate the weak teleconnections from ENSO and local SST. It is also important to further understand the global warming trend in TCs in the model compared with reanalysis data, a topic that is of importance to the wider climate modelling community. All these aspects need to be worked on in collaboration with the model developers.

**Ensemble perturbation methods** require evaluation in the context of TC prediction. These include the EDA method, moist singular vectors that are targeted over TCs, and model perturbation methods. Results here indicate that quite often the best track falls outside the ensemble plume, something that can be related to the track biases mentioned above. It is therefore important to further assess the reliability of the ensemble track forecasts and how they can be improved in the future.

The ensemble prediction system benefits from all model and data assimilation improvements. However, in the current configuration the lower resolution for the ensemble affects the TC predictions, especially for intensity. The current plan is to bring the ensemble resolution to 9-11 km in 2022-2023. Figure 66 illustrates the impact of increased ensemble resolution to 9 km for TC Laura.

The prediction of the RI is much improved. Such resolution increase is also expected to improve the ensemble spread for the intensity.
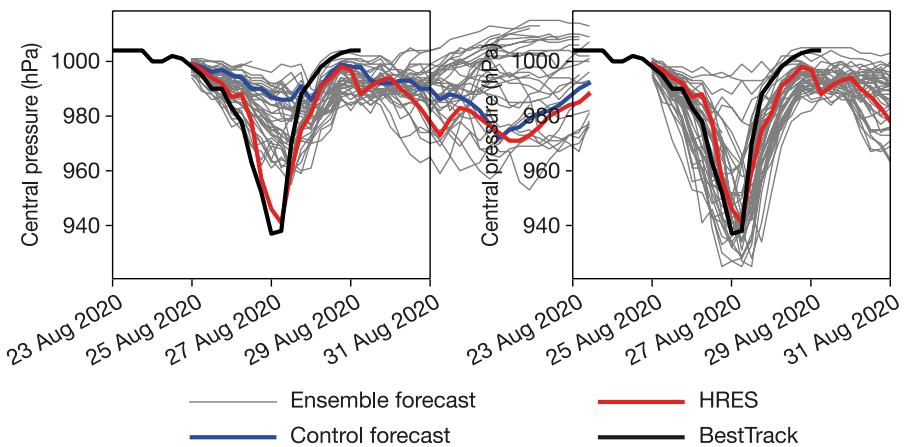


*Figure 66: Forecasts for TC Laura initialised 25 August 00UTC of Pmin comprising the operational ensemble forecast, control forecast, high-resolution forecast (HRES) and Best Track data (left) and HRES and preliminary. Best Track data with the experimental ensemble with the same resolution as HRES (right).*

For **usage of forecast output**, the following recommendations are provided:

- Increase collaborations across multiple centres on **TC tracking methods**
- Introduction of forecast products targeted towards the **impact** of the TCs
    - Rainfall and wind gusts in TCs are examples
    - The flooding impact from TCs needs to be validated
- Work together with the groups that use ECMWF forecasts for **storm surge modelling**
    - This is necessary to ensure the ECMWF forecasts are used optimally

**Machine learning** may help address some of the challenges described in this section. For example, a new project will explore machine learning to find drivers of predictability related to TC genesis, and also to be used as a forecasting aid.

All these points are important to address in collaboration with ECMWF Member States, other NWP centres and the wider research community. Examples include exploring high-resolution ocean modelling together with Météo-France and the Met Office. The Destination Earth programme would also provide opportunities to explore some of the listed aspects.

Although this report is extensive, there are inevitably parts that have not been covered. An example is the impact from aerosols on TCs. With the forecasting in Copernicus Atmospheric Monitoring Service (CAMS), this is an area that could be explored in the future. We have also not covered reanalysis aspects of TCs, something that is challenging especially for the pre-satellite era.

To conclude, substantial progress continues to be made in ECMWF's forecasts of TCs. As global user demands increase and diversify, and global and regional modelling of TCs at other centres continues to improve quickly, several avenues require investigation and improvement. These involve more

advanced evaluation and diagnostic methods, improvements to observational usage, and continuous advances in the data assimilation and modelling systems in several directions.

## 12.    Acknowledgements

## 13.    References

Agusti-Panareda, A., Thorncroft, C. D., Craig, G. C., & Gray, S. L. (2004). The extratropical transition of hurricane *Irene* (1999): A potential-vorticity perspective. Quarterly Journal of the Royal Meteorological Society, 130(598), 1047–1074. https://doi.org/10.1256/qj.02.140

Andersson, E., & Järvinen, H. (1999). Variational quality control. Quarterly Journal of the Royal Meteorological Society, 125(554), 697–722. https://doi.org/10.1002/qj.49712555416

Baker, A. J., Hodges, K. I., Schiemann, R. K. H., & Vidale, P. L. (2021). Historical variability and lifecycles of North Atlantic midlatitude cyclones originating in the tropics. Journal of Geophysical Research: Atmospheres, 126, e2020JD033924. https://doi.org/10.1029/2020JD033924

Barkmeijer, J., Buizza, R., Palmer, T.N., Puri, K. and Mahfouf, J.-F. (2001). Tropical singular vectors computed with linearized diabatic physics. Q.J.R. Meteorol. Soc., 127: 685-708. https://doi.org/10.1002/qj.49712757221

Bauer, P, Quintino, T, Wedi, N, Bonanni, A, Chrust, M, Deconinck, W, Diamantakis, M, Düben, P, English, S, Flemming, J, Gillies, P, Hadade, I, Hawkes, J, Hawkins, M, Iffrig, O, Kühnlein, C, Lange, M, Lean, P, Marsden, O, Müller, A, Saarinen, S, Sarmany, D, Sleigh, M, Smart, S, Smolarkiewicz, P, Thiemert, D, Tumolo, G, Weihrauch, C, Zanna, C, Maciel, P. (2020). The ECMWF Scalability Programme: Progress and Plans. ECMWF Tech Memo 857

Bechtold, P., Forbes, R., Sandu, I., Lang, STK., Ahlgrimm, M. (2020). A major moist physics upgrade for the IFS. ECMWF Newsletter 164. https://www.ecmwf.int/en/newsletter/164/meteorology/major-moist-physics-upgrade-ifs

Becker, T., Bechtold, P. & Sandu, I. (2021). Characteristics of convective precipitation over tropical Africa in storm-resolving global simulations. Q.J.R. Meteorol. Soc., Accepted. https://doi.org/10.1002/qj.4185

Beljaars, A. (2015). Time step dependence of wind errors in storm conditions, ECMWF RD memo RD16-041.

Bergman, D. L., Magnusson, L., Nilsson, J., Vitart, F. (2019). Seasonal forecasting of tropical cyclone landfall using ECMWF's System 4, Weather and Forecasting, 34 (5), 1239-125, https://doi.org/10.1175/WAF-D-18-0032.1

Bertin, X., Li, K., Roland, A., & Bidlot, J.-R. (2015). The contribution of short-waves in storm surges: Two case studies in the Bay of Biscay. *Continental Shelf Research*, *96*, 1–15. https://doi.org/10.1016/j.csr.2015.01.005

Bidlot, J.-R., Prates, F., Ribas R., Mueller-Quintino, A., Crepulja, M., Vitart, F. (2020). Enhancing tropical cyclone wind forecasts, ECMWF Newsletter 164. https://www.ecmwf.int/en/newsletter/164/meteorology/enhancing-tropical-cyclone-wind-forecasts

Biswas, M.K., D. Stark & L. Carson, (2018). GFDL Vortex Tracker Users's Guide V3.9a.

Blackwell, WJ, Braun, S, Bennartz, R, et al. (2018). An overview of the TROPICS NASA Earth Venture Mission. Q J R Meteorol Soc.144 ( Suppl. 1): 16– 26. https://doi.org/10.1002/qj.3290

Blake, E. S., and D. A. Zelinsky, (2018). Hurricane Harvey (AL092017), National Hurricane Center Tropical Cyclone Report, https://www.nhc.noaa.gov/data/tcr/AL092017_Harvey.pdf

Bloemendaal, N. et al. (2021). Adequately reflecting the severity of tropical cyclones using the new Tropical Cyclone Severity Scale, Environ. Res. Lett. 16 014048, https://iopscience.iop.org/article/10.1088/1748-9326/abd131/meta

Brammer, A., & Thorncroft, C. D. (2015). Variability and evolution of african easterly wave structures and their relationship with tropical cyclogenesis over the eastern atlantic. Monthly Weather Review, 143(12), 4975–4995. https://doi.org/10.1175/MWR-D-15-0106.1

Brut, A., Butet, A., Durand, P., Caniaux, G., & Planton, S. (2005). Air–sea exchanges in the equatorial area from the EQUALANT99 dataset: Bulk parametrizations of turbulent fluxes corrected for airflow distortion. Quarterly Journal of the Royal Meteorological Society, 131(610), 2497–2538. https://doi.org/10.1256/qj.03.185

Bonavita, M., Hólm, E., Isaksen, L., & Fisher, M. (2016). The evolution of the ECMWF hybrid data assimilation system. Quarterly Journal of the Royal Meteorological Society, 142(694), 287–303. https://doi.org/https://doi.org/10.1002/qj.2652

Bonavita, M., Dahoui, M., Lopez, P., Prates, F., Hólm, E., De Chiara, G., Geer, A., Isaksen, L., & Ingleby, B. (2017). On the initialization of tropical cyclones. ECMWF Tech Memo 810

Bormann, N., Lawrence, H., Farnan, J. (2019). Global observing system experiments in the ECMWF assimilation system. ECMWF Tech Memo 839

Brammer, A. and C. D. Thorncroft, 2015: Variability and evolution of African easterly wave structure and the relationship with tropical cyclogenesis over the eastern Atlantic. *Mon. Wea. Rev*., **143**, 4975-4995.

Browne, P. A., de Rosnay, P., Zuo, H., Bennett, A., & Dawson, A. (2019). Weakly Coupled Ocean-Atmosphere Data Assimilation in the ECMWF NWP System. Remote Sensing, 11(234), 1–24. https://doi.org/10.3390/rs11030234

Buizza, R., Leutbecher, M. and Isaksen, L. (2008). Potential use of an ensemble of analyses in the ECMWF ensemble prediction system. Quarterly Journal of the Royal Meteorological Society, 134, 2051–2066. URL: https://doi.org/10.1002/qj.346.

Camargo, S., F. Vitart, C.-Y. Lee and M. Tippett. (2021). Skill, predictability, and cluster analysis of Atlantic hurricanes in the ECMWF monthly forecasts. Submitted to Monthly Weather Review.

Cangialosi, J. P., and C. W. Landsea, 2016: An examination of model and official National Hurricane Center tropical cyclone size forecasts. *Wea. Forecasting*, **31**, 1293–1300

Cangialosi, J. P., E. Blake, M. DeMaria, A. Penny, A. Latto, E. N. Rappaport, and V. Tallapragada, 2020: Recent progress in tropical cyclone intensity forecasting at the National Hurricane Center. Wea. Forecasting, 35, 1913–1922

Chen, J. H., Lin, S. J., Magnusson, L., Bender, M., Chen, X., Zhou, L., Xiang, B. (2019). Advancements in Hurricane Prediction With NOAA's Next-Generation Forecast System, Geophysical Research Letters 46 (8), 4495-4501

Christophersen, H., J. Sippel, A. Aksoy, and N. L. Baker, 2021: Tropical Cyclone Data Assimilation. AGU Geophysical Monograph Series: "Earth's Climate and Weather: Dominant Variability and Disastrous Extremes".  Under Review.

Cione, J. J. and Coauthors, 2020: Eye of the Storm: Observing Hurricanes with a Small Unmanned Aircraft System. *Bull. Amer. Meteor. Soc.*, **101**, E186-E205.

Cook, P. A., & Renfrew, I. A. (2015). Aircraft-based observations of air-sea turbulent fluxes around the British Isles: Observations of Air-Sea Fluxes. Quarterly Journal of the Royal Meteorological Society, 141(686), 139–152. https://doi.org/10.1002/qj.2345

Copernicus (2021), *Global Flood Awareness System - Copernicus Services - ECMWF Confluence Wiki*. Retrieved 23 June 2021, from https://confluence.ecmwf.int/display/COPSRV/Global+Flood+Awareness+System

Copernicus EMS. (2021). *GloFAS Rapid Risk Assessment*. Global Flood Awareness System; Copernicus. Retrieved 23 June 2021, from https://www.globalfloods.eu/technical-information/glofas-impact-forecasts

Cotton, J., Francis, P., Heming, J., Forsythe, M., Reul, N., & Donlon, C. (2018). Assimilation of SMOS L-band wind speeds: Impact on Met Office global NWP and tropical cyclone predictions. Quarterly Journal of the Royal Meteorological Society, 144(711), 614–629. https://doi.org/https://doi.org/10.1002/qj.3237

Coughlan de Perez, E., van den Hurk, B., van Aalst, M. K., Jongman, B., Klose, T., & Suarez, P. (2015). Forecast-based financing: An approach for catalyzing humanitarian action based on extreme weather and climate forecasts. Natural Hazards and Earth System Sciences, 15(4), 895–904. https://doi.org/https://doi.org/10.5194/nhess-15-895-2015

Davis, C. A. (2018). Resolving tropical cyclone intensity in models. Geophysical Research Letters, 45(4), 2082–2087. https://doi.org/10.1002/2017GL076966

De Chiara, G., English S., (2016). "SMOS Hurricane wind speed analysis", Report for ESA contract 4000101703/10/NL/FF/fk CCN5.

De Chiara, G, Isaksen, L., English, S. (2018). ECMWF support for EPS/ASCAT ocean wind assessment", Final Report Contract No. EUM/CO/15/4600001497/JF

Deltares (2021). Global storm surge information system(GLOSSIS). Retrieved 23 June 2021, from https://www.deltares.nl/en/projects/global-storm-surge-information-system-glossis/

DeMaria, M., Franklin, J. L., Onderlinde, M. J., & Kaplan, J. (2021). Operational forecasting of tropical cyclone rapid intensification at the national hurricane center. Atmosphere, 12(6), 683. https://doi.org/10.3390/atmos12060683

Diamantakis, M. and Magnusson, L. (2016). Sensitivity of the ECMWF model to Semi-Lagrangian departure point iterations, Monthly Weather Review, 144 (9), 3233-3250

Diamantakis, M. and F. Vana, (2021). A fast converging and concise algorithm for computing the departure points in semi-Lagrangian weather and climate models, submitted in Q. J. R. Meteor. Soc.

Duncan, D. I., Bormann, N., Geer, A. J., and Weston, P. (2021). Assimilation of AMSU-A in All-sky Conditions. 57, EUMETSAT/ECMWF Fellowship Programme Research Report.

Duong, Q.-P., Langlade, S., Payan, C., Husson, R., Mouche, A., & Malardel, S. (2021). C-band sar winds for tropical cyclone monitoring and forecast in the south-west indian ocean. Atmosphere, 12(5), 576. https://doi.org/10.3390/atmos12050576

Doyle, J.D., C.A. Reynolds, C. Amerault, J. Moskaitis, (2012). Adjoint sensitivity and predictability of tropical cyclogenesis. J. Atmos. Sci., 69, 3535-3557.

Doyle, J. D. and Coauthors, 2014: Tropical cyclone prediction using COAMPS-TC. *Oceanography*, **27**, 104–115

van den Dool, H., Becker, E., Chen, L.-C., & Zhang, Q. (2017). The probability anomaly correlation and calibration of probabilistic forecasts. Weather and Forecasting, 32(1), 199–206. https://doi.org/10.1175/WAF-D-16-0115.1

Dvorak, V. F., (1984). Tropical cyclone intensity analysis using satellite data. NOAA Tech. Rep. NESDIS 11, 47 pp.

Elless, T. J., & Torn, R. D. (2018). African easterly wave forecast verification and its relation to convective errors within the ecmwf ensemble prediction system. Weather and Forecasting, 33(2), 461–477. https://doi.org/10.1175/WAF-D-17-0130.1

Emerton, R., Cloke, H., Ficchi, A., Hawker, L., de Wit, S., Speight, L., Prudhomme, C., Rundell, P., West, R., Neal, J., Cuna, J., Harrigan, S., Titley, H., Magnusson, L., Pappenberger, F., Klingaman, N., & Stephens, E. (2020). Emergency flood bulletins for Cyclones Idai and Kenneth: A critical evaluation of the use of global flood forecasts for international humanitarian preparedness and response. International Journal of Disaster Risk Reduction, 50, 101811. https://doi.org/10.1016/j.ijdrr.2020.101811

Evans, C., et al. (2017). The Extratropical Transition of Tropical Cyclones. Part I: Cyclone Evolution and Direct Impacts, Monthly Weather Review, 145(11), 4317-4344. https://doi.org/10.1175/MWR-D-17-0027.1

Fujita, T. (1952). Pressure distribution within typhoon. Geophys. Mag., 23, 437–451.

Gaffney, S. J., A. W. Robertson, P. Smyth, S. J. Camargo, and M. Ghil, (2007). Probabilistic clustering of extratropical cyclones using regression mixture models. Clim. Dyn., 29, 423–440, 643 doi:10.1007/s00382-007-0235-z

Geer, A. J., Lonitz, K., Weston, P., Kazumori, M., Okamoto, K., Zhu, Y., Liu, E. H., Collard, A., Bell, W., Migliorini, S., Chambon, P., Fourrié, N., Kim, M., Köpken-Watts, C., & Schraff, C. (2018). All-sky satellite data assimilation at operational weather forecasting centres. Quarterly Journal of the Royal Meteorological Society, 144(713), 1191–1217. https://doi.org/10.1002/qj.3202

Geer, A. J., Migliorini, S., & Matricardi, M. (2019). All-sky assimilation of infrared radiances sensitive to mid- and upper-tropospheric moisture and cloud. Atmospheric Measurement Techniques, 12(9), 4903–4929. https://doi.org/10.5194/amt-12-4903-2019

Gall, R., Franklin, J., Marks, F., Rappaport, E. N., & Toepfer, F. (2013). The hurricane forecast improvement project. Bulletin of the American Meteorological Society, 94(3), 329–343. https://doi.org/10.1175/BAMS-D-12-00071.1

Good S, Fiedler E, Mao C, Martin MJ, Maycock A, Reid R, Roberts-Jones J, Searle T, Waters J, While J, Worsfold M. (2020). The Current Configuration of the OSTIA System for Operational Production of Foundation Sea Surface Temperature and Ice Concentration Analyses. Remote Sensing. 2020; 12(4):720. https://doi.org/10.3390/rs12040720

Gray, W. (1979). Hurricanes: Their formation, structure and likely role in the tropical circulation, Meteorology over the Tropical Oceans, D. B. Shaw, Ed., Royal Meteorological Society, 155– 218.

Haiden, T., Janousek, M., Vitart, F., Ben-Bouallegue, Z., Ferranti, L., Prates, C., & Richardson, D. (2021). Evaluation of ECMWF forecasts, including the 2020 upgrade. ECMWF Tech Memo 880 https://doi.org/10.21957/6NJP8BYZ4

Hallerstig, M., Magnusson, L., Kolstad, E. W., & Mayer, S. (2021). How grid-spacing and convection representation affected the wind speed forecasts of four polar lows. Quarterly Journal of the Royal Meteorological Society, 147(734), 150–165. https://doi.org/10.1002/qj.3911

Harrigan, S., Zsoter, E., Alfieri, L., Prudhomme, C., Salamon, P., Wetterhall, F., Barnard, C., Cloke, H., &

Pappenberger, F. (2020). GloFAS-ERA5 operational global river discharge reanalysis 1979–present [Preprint]. Hydrology and Soil Science – Hydrology. https://doi.org/10.5194/essd-2019-232

Heming, J. T. (2016). Met Office Unified Model Tropical Cyclone Performance Following Major Changes to the Initialization Scheme and a Model Upgrade. Weather and Forecasting 31.5, 1433-1449, https://doi.org/10.1175/WAF-D-16-0040.1.

Heming, J. T., Prates, F., Bender, M. A., Bowyer, R., Cangialosi, J., Caroff, P., Coleman, T., Doyle, J. D., Dube, A., Faure, G., Fraser, J., Howell, B. C., Igarashi, Y., McTaggart-Cowan, R., Mohapatra, M.,

Moskaitis, J. R., Murtha, J., Rivett, R., Sharma, M., … Xiao, Y. (2019). Review of recent progress in tropical cyclone track forecasting and expression of uncertainties. Tropical Cyclone Research and Review, 8(4), 181–218. https://doi.org/10.1016/j.tcrr.2020.01.001

Hewson, T. D., & Titley, H. A. (2010). Objective identification, typing and tracking of the complete life-cycles of cyclonic features at high spatial resolution: Objective identification, typing and tracking of the complete life-cycles of cyclonic features. Meteorological Applications, 17(3), 355–381. https://doi.org/10.1002/met.204

Hodges, K. I. (1995). Feature tracking on the unit sphere. Monthly Weather Review, 123(12), 3458–3465. https://doi.org/10.1175/1520-0493(1995)123<3458:FTOTUS>2.0.CO;2

Hodges, K. I. (1999). Adaptive constraints for feature tracking. Monthly Weather Review, 127(6), 1362–1373. https://doi.org/10.1175/1520-0493(1999)127<1362:ACFFT>2.0.CO;2

Hodges, K., Cobb, A., & Vidale, P. L. (2017). How well are tropical cyclones represented in reanalysis datasets? Journal of Climate, 30(14), 5243–5264. https://doi.org/10.1175/JCLI-D-16-0557.1

Hodges, K. I., Klingaman, N. P. (2019). Prediction Errors of Tropical Cyclones in the Western North Pacific in the Met Office Global Forecast Model, Weather and Forecasting, 34(5), 1189-1209. https://doi.org/10.1175/WAF-D-19-0005.1

Holland, G. J. (1980). An analytic model of the wind and pressure profiles in hurricanes. *Monthly Weather Review*, *108*(8), 1212–1218. https://doi.org/10.1175/1520-0493(1980)108<1212:AAMOTW>2.0.CO;2

Holm, E. V., Bonavita, M., Magnusson, L., (2015). Improved spread and accuracy in higher-resolution Ensemble of Data Assimilations. ECMWF Newsletter 145, https://www.ecmwf.int/sites/default/files/elibrary/2015/14589-newsletter-no145-autumn-2015.pdf

Horn, M., Walsh, K., Zhao, M., Camargo, S. J., Scoccimarro, E., Murakami, H., Wang, H., Ballinger, A., Kumar, A., Shaevitz, D. A., Jonas, J. A., Oouchi, K. (2014). Tracking Scheme Dependence of Simulated Tropical Cyclone Response to Idealized Climate Simulations, Journal of Climate, 27(24), 9197-9213. https://doi.org/10.1175/JCLI-D-14-00200.1

Huffman, G., Bolvin, D., Braithwaite, D., Hsu, K., Joyce, R., Xie, P. (2014), Integrated Multi-satellitE Retrievals for GPM (IMERG), version 4.4. NASA's Precipitation Processing Center, accessed 31 March, 2015, ftp://arthurhou.pps.eosdis.nasa.gov/gpmdata/

IFRC (2021). *Forecast-Based Financing*. Retrieved 23 June 2021, from https://www.forecast-based-financing.org/about/

Ingleby, B., Prates, F., Isaksen, L., Bonavita, M. (2020). Recent BUFR dropsonde data improved forecasts. ECMWF Newsletter 162.

Janssen, P. (1997). Effect of surface gravity waves on the heat flux. ECMWF Tech Memo 239, https://doi.org/10.21957/80E0INRGX

Janssen, P. A. E. M., & Bidlot, J.-R. (2018). Progress in operational wave forecasting. Procedia IUTAM, 26, 14–29. https://doi.org/10.1016/j.piutam.2018.03.003

Janssen, P., & Bidlot, J.-R. (2021). On the consequences of nonlinearity and gravity-capillary waves on wind-wave interaction. ECMWF Tech Memo 882, https://doi.org/10.21957/B88TQJD6Q

Jones, S. C., Harr, P. A., Abraham, J., Bosart, L. F., Bowyer, P. J., Evans, J. L., Hanley, D. E., Hanstrum, B. N., Hart, R. E., Lalaurette, F., Sinclair, M. R., Smith, R. K., & Thorncroft, C. (2003). The Extratropical Transition of Tropical Cyclones: Forecast Challenges, Current Understanding, and Future Directions, Weather and Forecasting, 18(6), 1052-1092. https://doi.org/10.1175/1520-0434(2003)018<1052:TETOTC>2.0.CO;2

Judt, F., Klocke, D., Rios-Berrios, R., Vanniere, B., Ziemen, F., Auger, L., Biercamp, J., Bretherton, C., Chen, X., Düben, P., Hohenegger, C., Khairoutdinov, M., Kodama, C., Kornblueh, L., Lin, S.-J., Nakano, M., Neumann, P., Putman, W., Röber, N., … Zhou, L. (2021). Tropical cyclones in global storm-resolving models. Journal of the Meteorological Society of Japan. Ser. II, 99(3), 579–602. https://doi.org/10.2151/jmsj.2021-029

Jurlina, T., Baugh, C., Pappenberger, F., Prudhomme, C. (2020). Flood hazard risk forecasting index (FHRFI) for urban areas: The Hurricane Harvey case study. Meteorol Appl.; 27:e1845. https://doi.org/10.1002/met.1845

Keller, J. H., et al. (2019). The Extratropical Transition of Tropical Cyclones. Part II: Interaction with the Midlatitude Flow, Downstream Impacts, and Implications for Predictability, Monthly Weather Review, 147(4), 1077-1106. https://doi.org/10.1175/MWR-D-17-0329.1

Knapp, K. R., Kruk, M. C., Levinson, D. H., Diamond, H. J., & Neumann, C. J. (2010). The international best track archive for climate stewardship (Ibtracs): Unifying tropical cyclone data. Bulletin of the American Meteorological Society, 91(3), 363–376. https://doi.org/10.1175/2009BAMS2755.1

Kohno, N. Dube, S. K., Entel, M., Fakhruddin, S.H.M., Greenslade, D., Leroux, M.-D., Rhome, J., Ba Thuy, N. (2018). Recent Progress in Storm Surge Forecasting, TCRR, Vol 7, 128-139, https://www.sciencedirect.com/science/article/pii/S2225603219300207

Komaromi, W. A., P. A. Reinecke, J. D. Doyle, and J. R. Moskaitis, 2021: The Naval Research Laboratory's Coupled Ocean-Atmosphere Mesoscale Prediction System – Tropical Cyclone Ensemble (COAMPS-TC Ensemble). Wea. Forecasting, 36, 499-517.

Kowaleski, A. M., & Evans, J. L. (2020). Use of multiensemble track clustering to inform medium-range tropical cyclone forecasts. Weather and Forecasting, 35(4), 1407–1426. https://doi.org/10.1175/WAF-D-20-0003.1

Kowaleski, A. M., Morss, R. E., Ahijevych, D., Fossell, K. R. (2020). Using a WRF-ADCIRC Ensemble and Track Clustering to Investigate Storm Surge Hazards and Inundation Scenarios Associated with Hurricane Irma, Weather and Forecasting, 35(4), 1289-1315, https://doi.org/10.1175/WAF-D-19-0169.1

Kühnlein C., W. Deconinck, R. Klein, S. Malardel, Z. P. Piotrowski, P. K. Smolarkiewicz, J. Szmelter, and N. P. Wedi, 2019. FVM 1.0: a nonhydrostatic finite-volume dynamical core for the IFS. Geosci. Model Dev., 12, 651–676.

Kühnlein C. et al., Towards global weather forecasting with IFS-FVM, ECMWF Annual Seminar 2020, https://events.ecmwf.int/event/167/contributions/1367/attachments/799/1410/AS2020-Kuehnlein.pdf

Kumler-Bonfanti, C., Stewart, J., Hall, D., Govett, M. (2020). Tropical and Extratropical Cyclone Detection Using Deep Learning, Journal of Applied Meteorology and Climatology, 59(12), 1971-1985. https://doi.org/10.1175/JAMC-D-20-0117.1

Landsea, C. W., & Franklin, J. L. (2013). Atlantic hurricane database uncertainty and presentation of a new database format. Monthly Weather Review, 141(10), 3576–3592. https://doi.org/10.1175/MWR-D-12-00254.1

Lang, S.T.K., Leutbecher, M. and Jones, S.C. (2012), Impact of perturbation methods in the ECMWF ensemble prediction system on tropical cyclone forecasts. Q.J.R. Meteorol. Soc., 138: 2030-2046. https://doi.org/10.1002/qj.1942

Lang, S.T.K., Bonavita, M. and Leutbecher, M. (2015), On the impact of re-centring initial conditions for ensemble forecasts. Q.J.R. Meteorol. Soc., 141: 2571-2581. https://doi.org/10.1002/qj.2543

Lang, STK, Lock, S-J, Leutbecher, M, Bechtold, P, Forbes, RM. (2021), Revision of the Stochastically Perturbed Parametrisations model uncertainty scheme in the Integrated Forecasting System. Q J R Meteorol Soc. 2021; 147: 1364– 1381. https://doi.org/10.1002/qj.3978

Lang, S.T.K., Dawson, A., Diamantakis, M., Dueben, P., Hatfield, S., Leutbecher, M., Palmer, T., Prates, F., Roberts, C.D., Sandu, I., Wedi, N. (2021), More Accuracy with Less Precision, submitted to Q.J.R. Meteorol. Soc.

Lawrence, H., Bormann, N., Sandu, I., Day, J., Farnan, J., & Bauer, P. (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

Lee, C., Camargo, S. J., Vitart, F., Sobel, A. H., & Tippett, M. K. (2018). Subseasonal Tropical Cyclone Genesis Prediction and MJO in the S2S Dataset, Weather and Forecasting, 33(4), 967-988, https://doi.org/10.1175/WAF-D-17-0165.1

Lee, C., Camargo, S. J., Vitart, F., Sobel, A. H., Camp, J., Wang, S., Tippett, M. K., & Yang, Q. (2020). Subseasonal Predictions of Tropical Cyclone Occurrence and ACE in the S2S Dataset, Weather and Forecasting, 35(3), 921-938. https://doi.org/10.1175/WAF-D-19-0217.1

Liang, X., Chan, J. C. L. (2005). The Effects of the Full Coriolis Force on the Structure and Motion of a Tropical Cyclone. Part I: Effects due to Vertical Motion, Journal of the Atmospheric Sciences, 62(10), 3825-3830, https://doi.org/10.1175/JAS3545.1

Lillo, S.P., Parsons, D.B. (2017). Investigating the dynamics of error growth in ECMWF medium-range forecast busts. Q.J.R. Meteorol. Soc., 143: 1211-1226. https://doi.org/10.1002/qj.2938

Liu, Z., Wang, H., Zhang, Y. J., Magnusson, L., Loftis, J. D., Forrest, D. (2020). Cross-scale modeling of storm surge, tide, and inundation in Mid-Atlantic Bight and New York City during hurricane Sandy, 2012, Estuarine, Coastal and Shelf Science 233, 106544

Leonardo, N. M., & Colle, B. A. (2020). An Investigation of Large Along-Track Errors in Extratropical Transitioning North Atlantic Tropical Cyclones in the ECMWF Ensemble, Monthly Weather Review, 148(1), 457-476. https://doi.org/10.1175/MWR-D-19-0044.1

Leonardo, N. M., & Colle, B. A. (2021). An Investigation of Large Cross-Track Errors in North Atlantic Tropical Cyclones in the GEFS and ECMWF Ensembles, Monthly Weather Review, 149(2), 395-417. https://doi.org/10.1175/MWR-D-20-0035.1

Leutbecher, M. and Palmer, T. N. (2008) Ensemble forecasting. J. Comput. Phys., 227, 3515–3539. URL: https://doi.org/10.1016/ j.jcp.2007.02.014.

Leutbecher, M., et al. (2017). Stochastic representations of model uncertainties at ECMWF: State of the art and future vision, QJRMS, 143, 2315-2339

MacLeod, D., Easton-Calabria, E., de Perez, E. C., & Jaime, C. (2021). Verification of forecasts for extreme rainfall, tropical cyclones, flood and storm surge over Myanmar and the Philippines. Weather and Climate Extremes, 33, 100325. https://doi.org/10.1016/j.wace.2021.100325

Magnusson, L., Bidlot, J. R., Lang, S. T. K., Thorpe, A., Wedi, N., & Yamaguchi, M. (2014). Evaluation of medium-range forecasts for hurricane Sandy. Monthly Weather Review, 142(5), 1962-1981.

Magnusson, L. , Dahoui, M., Velden, C. S., Olander, T. L., (2017). A fresh look at tropical cyclone intensity estimates. ECMWF Newsletter 152. https://www.ecmwf.int/en/newsletter/152/news/fresh-look-tropical-cyclone-intensity-estimates

Magnusson, L., Bidlot, J.-R., Bonavita, M., Brown, A., Browne, P., De Chiara, G., Dahoui, M., Lang, S. T. K., McNally, T., Mogensen, K. S., Pappenberger, F., Prates, F., Rabier, F., Richardson, D. S., Vitart, F. Malardel, S. (2019a). ECMWF activities for improved hurrivane forecasts, BAMS, 100 (3), 445-458

Magnusson, L., Doyle, J. D., Komaromi, W. A., Zhang, F., Torn, R., Tang, C. K., Chan, C. L., Yamaguchi, M. (2019b). Advances in understanding difficult cases of track forecasts, Tropical Cyclone Research and Review, 8(3), 109-122. https://doi.org/10.1016/j.tcrr.2019.10.001

Magnusson, L. and Bidlot, J.-R. (2020). Challenges in forecasting Hurricane Lorenzo. ECMWF Newsletter, 162, 2-3.

Maloney, E. D., & Hartmann, D. L. (2000). Modulation of Eastern North Pacific Hurricanes by the Madden–Julian Oscillation, Journal of Climate, 13(9), 1451-1460. https://doi.org/10.1175/1520-0442(2000)013<1451:MOENPH>2.0.CO;2

McNally, T., Bonavita, M., & Thépaut, J. (2014). The Role of Satellite Data in the Forecasting of Hurricane Sandy, Monthly Weather Review, 142(2), 634-646. https://doi.org/10.1175/MWR-D-13-00170.1

Mogensen, K. S., Magnusson, L. and Bidlot, J.-R. (2017). Tropical cyclone sensitivity to ocean coupling in the ECMWF coupled model, Journal of Geophysical Research: Oceans, 122, 4392-4412,

Mouche, A., Chapron, B., Knaff, J., Zhao, Y., Zhang, B., Combot, C. (2019). Copolarized and cross-polarized SAR measurements for high-resolution description of major hurricane wind structures:

Application to Irma category 5 hurricane. Journal of Geophysical Research: Oceans, 124, 3905–3922. https://doi.org/10.1029/2019JC015056

NCAR. (2021). THORPEX Interactive Grand Global Ensemble (TIGGE) Model Tropical Cyclone Track Data, https://doi.org/10.5065/D6GH9GSZ.

Neumann, B., Vafeidis, A. T., Zimmermann, J., & Nicholls, R. J. (2015). Future coastal population growth and exposure to sea-level rise and coastal flooding—A global assessment. PLOS ONE, 10(3), e0118571. https://doi.org/10.1371/journal.pone.0118571

Olander, T. L., and C. S. Velden, 2007: The advanced Dvorak technique: Continued development of an objective scheme to estimated tropical cyclone intensity using geostationary infrared satellite imagery. *Wea. Forecasting*, **22**, 287—298.

Omranian, E., Sharif, H., & Tavakoly, A. (2018). How well can global precipitation measurement (Gpm) capture hurricanes? Case study: hurricane harvey. Remote Sensing, 10(7), 1150. https://doi.org/10.3390/rs10071150

Puri, K., Barkmeijer, J. and Palmer, T.N. (2001), Ensemble prediction of tropical cyclones using targeted diabatic singular vectors. Q.J.R. Meteorol. Soc., 127: 709-731. https://doi.org/10.1002/qj.49712757222

Rappaport, E. N., Franklin, J. L., Avila, L. A., Baig, S. R., Beven, J. L., Blake, E. S., Burr, C. A., Jiing, J.-G., Juckins, C. A., Knabb, R. D., Landsea, C. W., Mainelli, M., Mayfield, M., McAdie, C. J., Pasch, R. J., Sisko, C., Stewart, S. R., & Tribble, A. N. (2009). Advances and challenges at the national hurricane center. Weather and Forecasting, 24(2), 395–419. https://doi.org/10.1175/2008WAF2222128.1

Rennie, M.P., Isaksen, L., Weiler, F., de Kloe, J., Kanitz, T. & Reitebuch, O. (2021) The impact of Aeolus wind retrievals on ECMWF global weather forecasts. Quarterly Journal of the Royal Meteorological Society, 1–32. Available from: https://doi.org/10.1002/qj.4142

Reul, N., Chapron, B., Zabolotskikh, E., Donlon, C., Mouche, A., Tenerelli, J., Collard, F., Piolle, J. F., Fore, A., Yueh, S., Cotton, J., Francis, P., Quilfen, Y., & Kudryavtsev, V. (2017). A new generation of tropical cyclone size measurements from space. Bulletin of the American Meteorological Society, 98(11), 2367–2385. https://doi.org/10.1175/BAMS-D-15-00291.1

Richardson, D. S., Cloke, H. L., & Pappenberger, F. (2020). Evaluation of the consistency of ecmwf ensemble forecasts. Geophysical Research Letters, 47(11). https://doi.org/10.1029/2020GL087934

Riemer, M. and Jones, S.C. (2014). Interaction of a tropical cyclone with a high-amplitude, midlatitude wave pattern: Waviness analysis, trough deformation and track bifurcation. Q.J.R. Meteorol. Soc., 140: 1362-1376. https://doi.org/10.1002/qj.2221

Roberts, M. J., Camp, J., Seddon, J., Vidale, P. L., Hodges, K., Vanniere, B., Mecking, J., Haarsma, R., Bellucci, A., Scoccimarro, E., Caron, L.-P., Chauvin, F., Terray, L., Valcke, S., Moine, M.-P., Putrasahan, D., Roberts, C., Senan, R., Zarzycki, C., & Ullrich, P. (2020). Impact of model resolution on tropical cyclone simulation using the highresmip–primavera multimodel ensemble. Journal of Climate, 33(7), 2557–2583. https://doi.org/10.1175/JCLI-D-19-0639.1

Rodwell, M., Ferranti, L., Haiden, T., Magnusson, L. et al. (2015). New developments in the diagnosis and verification of high-impact weather forecasts. ECMWF Tech Memo 759, https://doi.org/10.21957/36B0T3MKE

Ruf, C. S., Atlas, R., Chang, P. S., Clarizia, M. P., Garrison, J. L., Gleason, S., Katzberg, S. J., Jelenak, Z., Johnson, J. T., Majumdar, S. J., O'Brien, A., Posselt, D. J., Ridley, A. J., Rose, R. J., & Zavorotny, V. U. (2016). New ocean winds satellite mission to probe hurricanes and tropical convection. Bulletin of the American Meteorological Society, 97(3), 385–395. https://doi.org/10.1175/BAMS-D-14-00218.1

Ruston, B., & Healy, S. (2021). Forecast Impact of FORMOSAT -7/ COSMIC -2 GNSS Radio Occultation Measurements. Atmospheric Science Letters, 22(3). https://doi.org/10.1002/asl.1019

Sanabia, E. R., & Jayne, S. R. (2020). Ocean observations under two major hurricanes: Evolution of the response across the storm wakes. AGU Advances, 1, e2019AV000161. https://doi.org/10.1029/2019AV000161

Sawada, Y., Okamoto, K., Kunii, M., & Miyoshi, T. (2019). Assimilating every-10-minute himawari-8 infrared radiances to improve convective predictability. Journal of Geophysical Research: Atmospheres, 124(5), 2546–2561. https://doi.org/10.1029/2018JD029643

Schäfler, A., et al. (2018). The north Atlantic waveguide and downstream impact experiment, BAMS, 99 (8), 1607-1637, https://doi.org/10.1175/BAMS-D-17-0003.1

Skofronick-Jackson, G., Petersen, W. A., Berg, W., Kidd, C., Stocker, E. F., Kirschbaum, D. B., Kakar, R., Braun, S. A., Huffman, G. J., Iguchi, T., Kirstetter, P. E., Kummerow, C., Meneghini, R., Oki, R., Olson, W. S., Takayabu, Y. N., Furukawa, K., & Wilheit, T. (2017). The global precipitation measurement (Gpm) mission for science and society. Bulletin of the American Meteorological Society, 98(8), 1679–1695. https://doi.org/10.1175/BAMS-D-15-00306.1

Stockdale, T. and co-authors. (2018). SEAS5 and the future evolution of the long-range forecast system. ECMWF Tech Memo 835

Tang, C. K., Chan, J. C. L., & Yamaguchi, M. (2021). Large tropical cyclone track forecast errors of global numerical weather prediction models in western north pacific basin. Tropical Cyclone Research and Review, S2225603221000242. https://doi.org/10.1016/j.tcrr.2021.07.001

Tavolato, C., & Isaksen, L. (2015). On the use of a Huber norm for observation quality control in the ECMWF 4D-Var. Quarterly Journal of the Royal Meteorological Society, 141(690), 1514–1527. https://doi.org/10.1002/qj.2440

Termonia, P., & Hamdi, R. (2007). Stability and accuracy of the physics—Dynamics coupling in spectral models. Quarterly Journal of the Royal Meteorological Society. https://doi.org/10.1002/qj.119

Tian, X., & Zou, X. (2016). ATMS- and AMSU-A-derived hurricane warm core structures using a modified retrieval algorithm. Journal of Geophysical Research: Atmospheres, 121(21), 12,630-12,646. https://doi.org/10.1002/2016JD025042

Titley, H., Yamaguchi, M., Magnusson, L. (2019). Current and potential use of ensemble forecasts in operational TC forecasting: results from a global forecaster survey, Tropical Cyclone Research and Review, 8(3), 166-180.

Titley, H. A., Bowyer, R. L. Cloke, H. L. (2020). A global evaluation of multi-model ensemble tropical cyclone track probability forecasts. Q J R Meteorol Soc., 146: 531– 545. https://doi.org/10.1002/qj.3712

Titley, H. A., Cloke, H. L., Harrigan, S., Pappenberger, F., Prudhomme, C., Robbins, J. C., Stephens, E. M., & Zsoter, E. (2021). Key factors influencing the severity of fluvial flood hazard from tropical cyclones. Journal of Hydrometeorology, 1(aop). https://doi.org/10.1175/JHM-D-20-0250.1

Torn, R. D., & Snyder, C. (2012). Uncertainty of tropical cyclone best-track information. Weather and Forecasting, 27(3), 715–729. https://doi.org/10.1175/WAF-D-11-00085.1

Torn, R. D., T. J. Elless, P. Papin, C. A. Davis, (2018). The sensitivity of TC track forecasts within deformation steering flows. Mon. Wea. Rev., 146, 3183-3201

Tort, M., & Dubos, T. (2014). Dynamically consistent shallow-atmosphere equations with a complete Coriolis force: Non-Traditional Shallow-Atmosphere Equations. Quarterly Journal of the Royal Meteorological Society, 140(684), 2388–2392. https://doi.org/10.1002/qj.2274

Tsai, H.-C., & Elsberry, R. L. (2013). Detection of tropical cyclone track changes from the ECMWF ensemble prediction system: TC TRACK CHANGES IN ECMWF ENSEMBLE. Geophysical Research Letters, 40(4), 797–801. https://doi.org/10.1002/grl.50172

Van der Grijn, G., Paulsen, J. E., Lalaurette, F., Leutbecher, M. (2005). Early medium-range forecasts of tropical cyclones. ECMWF Newsletter 102, https://www.ecmwf.int/sites/default/files/elibrary/2004/14623-newsletter-no102-winter-200405.pdf

Van Der Knijff, J. M., Younis, J., & De Roo, A. P. J. (2010). LISFLOOD: A GIS-based distributed model for river basin scale water balance and flood simulation. *International Journal of Geographical Information Science*, *24*(2), 189–212. https://doi.org/10.1080/13658810802549154

Vellinga, M., Copsey, D., Graham, T., Milton, S., Johns, T. (2020). Evaluating Benefits of Two-Way Ocean–Atmosphere Coupling for Global NWP Forecasts, Weather and Forecasting, 35(5), 2127-2144. https://doi.org/10.1175/WAF-D-20-0035.1

Vidale, P. L., Hodges, K., Vannière, B., Davini, P., Roberts, M. J., Strommen, K., Weisheimer, A., Plesca, E., & Corti, S. (2021). Impact of stochastic physics and model resolution on the simulation of tropical cyclones in climate gcms. Journal of Climate, 34(11), 4315–4341. https://doi.org/10.1175/JCLI-D-20-0507.1

Vitart, F., Anderson, J. L., Stern, W. F. (1997). Simulation of Interannual Variability of Tropical Storm Frequency in an Ensemble of GCM Integrations, Journal of Climate, 10(4), 745-760. https://doi.org/10.1175/1520-0442(1997)010<0745:SOIVOT>2.0.CO;2

Vitart, F., Anderson, D., & Stockdale, T. (2003). Seasonal forecasting of tropical cyclone landfall over mozambique. *Journal of Climate*, *16*(23), 3932–3945. https://doi.org/10.1175/1520-0442(2003)016<3932:SFOTCL>2.0.CO;2

Vitart, F. (2009). Impact of the Madden Julian oscillation on tropical storms and risk of landfall in the ECMWF forecast system: MJO composites of model tropical storms. Geophysical Research Letters, 36(15), n/a-n/a. https://doi.org/10.1029/2009GL039089

Vitart, F., Prates, F., Bonet, A., & Sahin, C. (2012). New tropical cyclone products on the web. ECMWF Newsletter 130, https://doi.org/10.21957/TI1191E2

Wedi, N. P., Polichtchouk, I., Dueben, P., Anantharaj, V. G., Bauer, P., Boussetta, S., Browne, P., Deconinck, W., Gaudin, W., Hadade, I., Hatfield, S., Iffrig, O., Lopez, P., Maciel, P., Mueller, A., Saarinen, S., Sandu, I., Quintino, T., & Vitart, F. (2020). A baseline for global weather and climate simulations at 1 km resolution. Journal of Advances in Modeling Earth Systems, 12(11). https://doi.org/10.1029/2020MS002192

Wilks, D. S. (2006), Statistical Methods in the Atmospheric  Sciences.2$^{nd}$ Academic Press, 627 pp

WMO (2013), Verification methods for tropical cyclone forecasts, WWRP/WGNE Joint Working Group on Forecast Verification Research, November 2013

Wu, L., Rutgersson, A., Sahlée, E., & Larsén, X. G. (2015). The impact of waves and sea spray on modelling storm track and development. Tellus A: Dynamic Meteorology and Oceanography, 67(1), 27967. https://doi.org/10.3402/tellusa.v67.27967

Yamaguchi, M., Majumdar, S. J. (2010). Using TIGGE Data to Diagnose Initial Perturbations and Their Growth for Tropical Cyclone Ensemble Forecasts, Monthly Weather Review, 138(9), 3634-3655. https://doi.org/10.1175/2010MWR3176.1

Yamaguchi, M., Vitart, F., Lang, S. T. K., Magnusson, L., Elsberry, R. L., Elliott, G., Kyouda, M. and Nakazawa, T. (2015). Global Distribution of the Skill of Tropical Cyclone Activity Forecasts on Short- to Medium-Range Time Scales. Wea. Forecasting, 30, 1695–1709.

Yamaguchi, M., Ishida, J., Sato, H., Nakagawa, M. (2017). WGNE Intercomparison of Tropical Cyclone Forecasts by Operational NWP Models: A Quarter Century and Beyond, Bulletin of the American Meteorological Society, 98(11), 2337-2349. https://doi.org/10.1175/BAMS-D-16-0133.1

Zawislak, J., Rogers, R. F., Aberson, S. D., Alaka, G. J., Alvey, G., Aksoy, A., Bucci, L., Cione, J., Dorst, N., Dunion, J., Fischer, M., Gamache, J., Gopalakrishnan, S., Hazelton, A., Holbach, H. M., Kaplan, J., Leighton, H., Marks, F., Murillo, S. T., … Zhang, J. A. (2021). Accomplishments of noaa's airborne hurricane field program and a broader future approach to forecast improvement. Bulletin of the American Meteorological Society, 1–79. https://doi.org/10.1175/BAMS-D-20-0174.1

## Appendix A: List of Abbreviations

| | |
|---|---|
| COAMPS-TC | Coupled Ocean/Atmosphere Mesoscale Prediction System - Tropical Cyclones (USA) |
| ENS | Operational ECMWF ensemble forecast |
| ENSO | El Niño and the Southern Oscillation |
| GTS | Global Telecommunications System |
| HRES | Operational high-resolution ECMWF forecast |
| HWRF | Hurricane Weather Research and Forecasting model (USA) |
| IBTrACS | International Best Track Archive for Climate Stewardship |
| JMA | Japan Meteorological Agency |
| MDR | Main Development Region in Atlantic basin (10-20N, 20-80W) |
| MHS | Microwave Humidity Sounder |
| MJO | Madden-Julian Oscillation |
| MSLP | Mean Sea Level Pressure |
| NASA | National Aeronautics and Space Administration (USA) |
| NCEP | National Centers for Environmental Prediction (USA) |
| NHC | National Hurricane Center (USA / NOAA) |
| NOAA | National Oceanographic and Atmospheric Administration (USA) |
| NWP | Numerical Weather Prediction |
| Pmin | Minimum central surface pressure in the tropical cyclone |
| RSMC | WMO Regional Specialized Meteorological Center |
| RI | Rapid Intensification |
| SEAS5 | ECMWF's Seasonal Forecasting System |
| SFMR | Stepped Frequency Microwave Radiometer |

| SST | Sea Surface Temperature |
|---|---|
| TC | Tropical Cyclone |
| TCWC | Tropical Cyclone Warning Centre |
| UKMO | UK Met Office |
| Vmax | Maximum sustained surface wind speed at any location in the tropical cyclone |
| WGNE | WMO Working Group on Numerical Experimentation |
| WMO | World Meteorological Organisation |
| WWRP | World Weather Research Program |

## Appendix B: List of Tropical Cyclones in Special Experiment Period

The cases highlighted in grey refer to tropical cyclones with a Pmin lower than 980 hPa. Each of these cases lasted for at least 108 hours and provided the most substantial contributions to the statistics.

Vmax refers to the maximum sustained surface wind speed averaged over 1 minute, except for the western north Pacific tropical cyclones (marked by *) in which the averaging is over 10 minutes.

| TC Code | TC Name | Genesis Time | Final Time | Duration (h) | Maximum Vmax (kt) | Minimum Pmin (hPa) |
|---|---|---|---|---|---|---|
| 13L | Laura | 2020-08-20-00 | 2020-08-29-00 | 216 | 130 | 937 |
| 14L | Marco | 2020-08-20-12 | 2020-08-25-00 | 108 | 65 | 991 |
| 15L | Omar | 2020-09-02-00 | 2020-09-05-12 | 84 | 35 | 1003 |
| 16L | Nana | 2020-09-01-12 | 2020-09-04-00 | 60 | 65 | 994 |
| 17L | Paulette | 2020-09-07-12 | 2020-09-16-06 | 210 | 90 | 965 |
| 18L | Rene | 2020-09-07-18 | 2020-09-14-12 | 162 | 40 | 1001 |
| 19L | Sally | 2020-09-12-18 | 2020-09-17-06 | 108 | 95 | 966 |
| 20L | Teddy | 2020-09-14-06 | 2020-09-23-00 | 210 | 120 | 945 |
| 21L | Vicky | 2020-09-14-12 | 2020-09-17-18 | 78 | 45 | 1001 |
| 22L | Beta | 2020-09-18-18 | 2020-09-22-18 | 96 | 55 | 993 |

| 23L | Wilfred | 2020-09-18-18 | 2020-09-21-00 | 54 | 35 | 1006 |
|-----|---------|---------------|---------------|-----|------|------|
| 11E | Fausto | 2020-08-16-00 | 2020-08-17-12 | 36 | 35 | 1004 |
| 12E | Genevieve | 2020-08-16-12 | 2020-08-21-12 | 120 | 115 | 950 |
| 13E | Hernan | 2020-08-24-00 | 2020-08-28-18 | 114 | 40 | 1001 |
| 14E | Iselle | 2020-08-24-12 | 2020-08-30-18 | 150 | 50 | 997 |
| 15E | Julio | 2020-09-03-12 | 2020-09-07-06 | 90 | 40 | 1004 |
| 16E | Karina | 2020-09-10-00 | 2020-09-17-00 | 168 | 50 | 996 |
| 08W | Higos | 2020-08-17-18 | 2020-08-19-06 | 36 | 60* | 992 |
| 09W | Bavi | 2020-08-21-18 | 2020-08-27-12 | 138 | 85* | 950 |
| 10W | Maysak | 2020-08-28-00 | 2020-09-03-06 | 150 | 95* | 935 |
| 11W | Haishen | 2020-08-31-12 | 2020-09-07-12 | 168 | 105* | 910 |
| 13W | Noul | 2020-09-15-18 | 2020-09-18-12 | 66 | 45* | 992 |

# Appendix C. Special Experiments: Verification Scores for Temperature and Winds in the Tropics

The appendix presents root-mean-square (RMSE) errors for the tropics (20°N-20°S). The plots show normalised RMSE differences against the control experiment, where positive values mean that experiment performed better than the control. The verification is against the operational analysis. For data assimilation experiments, this will favour experiments similar to the operational configuration for short lead-times.

Figure 67 shows results for satellite observation experiments, corresponding to Figure 36. As the experiments introduce difference to the verifying analysis, a negative impact is seen for short lead times. We expect the impact to be negative from the experiments where we remove observations, and significant differences remain into day 2-3 for several experiments, indicating a benefit at 700hPa in the tropics of assimilating these.

Figure 68 shows the results for the scatterometer experiments from Figure 39. Again, we see a negative impact for the shortest lead-times as mentioned above. But here we see improvements after Day 1 with reduced thinning and increased error for the experiment without any ASCAT data.

Figure 69 includes results for model dynamics experiments and Figure 70 for model physics and resolution experiments. For the dynamics experiments (corresponding to Figure 48) the results are mainly neutral, but with an interesting improvement from the move of the vertical diffusion calculation for longer lead-times. For the new moist physics (Figure 70) we find a positive impact for the tropical winds and temperature for both 4 km and 9 km resolution, corresponding to Figure 49 to Figure 51. The experiment with explicit convection (no parameterised convection) and 4 km resolution is clearly worse than the control experiment.

Figure 71 shows the results for the experiment without ocean coupling and for different wave model experiments, corresponding to Figure 55. By removing the ocean coupling we see a negative impact on the forecast scores. For the new state dependent wave physics we see a degradation in the tropics, but together with the new moist physics is still results in a positive impact.
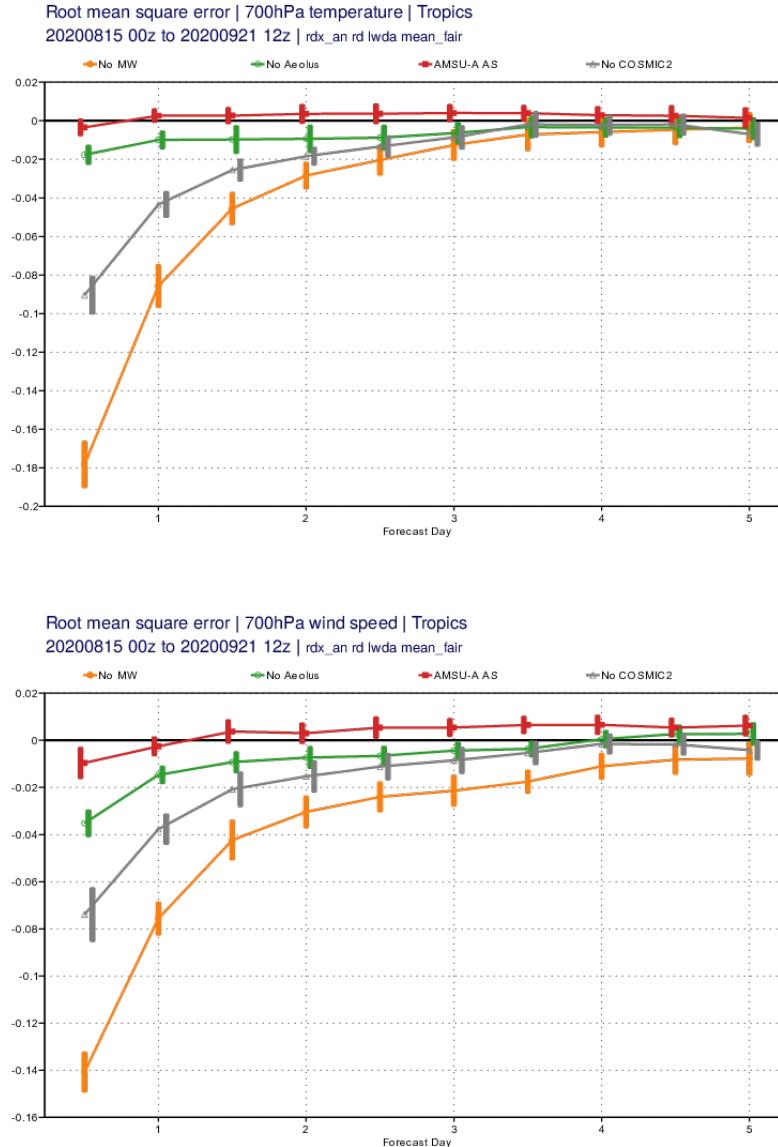


Figure 67: Normalised RMSE difference to control for satellite observation experiments for 700hPa temperature (top) and 700hPa wind speed (bottom) in the Tropics. (Positive means experiment better than control.
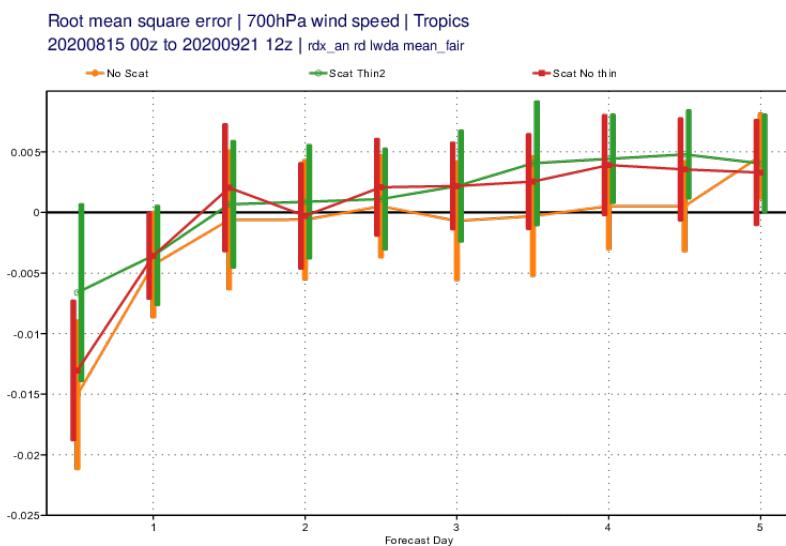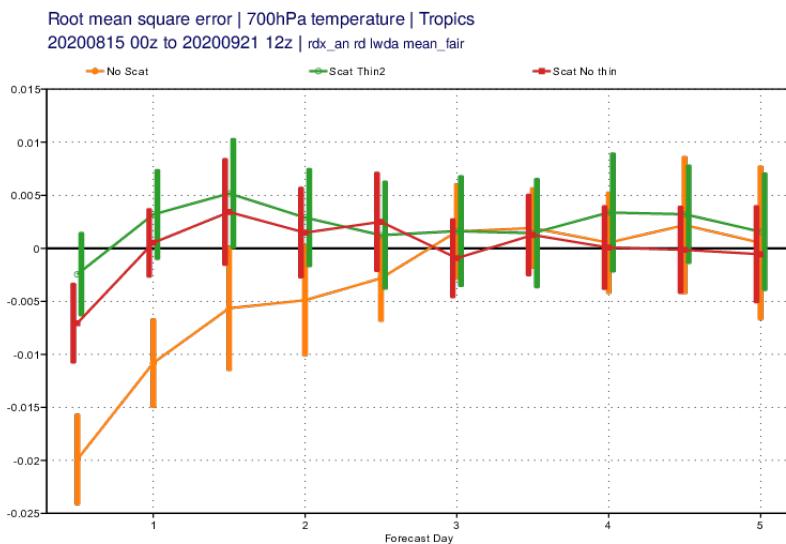
*Figure 68: Normalised RMSE difference to control for scatterometer experiments for 700hPa temperature (top) and 700hPa wind speed (bottom) in the Tropics. (Positive means experiment better than control.*
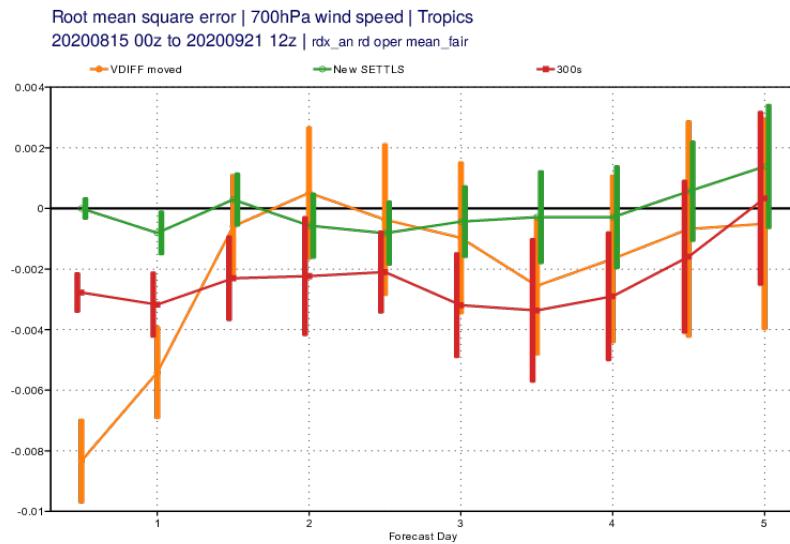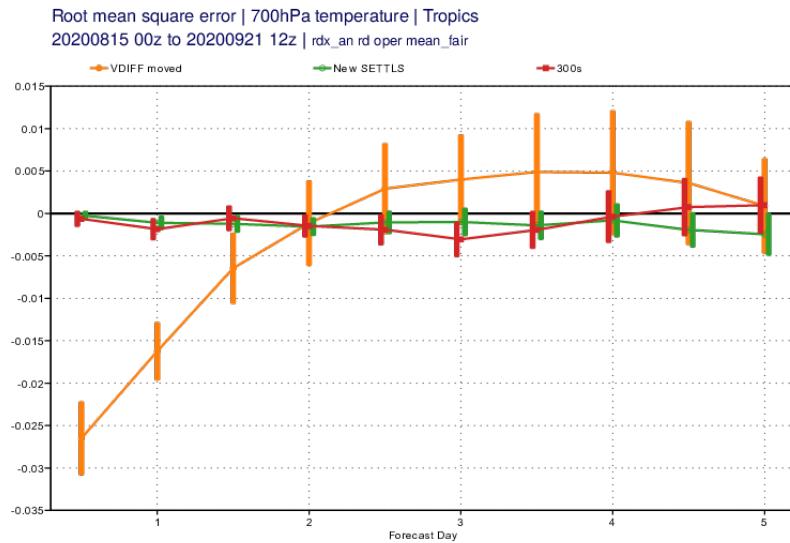
*Figure 69: Normalised RMSE difference to 9km-oper for model dynamics experiments for 700hPa temperature (top) and 700hPa wind speed (bottom) in the Tropics. (Positive means experiment better than control.*

Technical Memorandum No.888
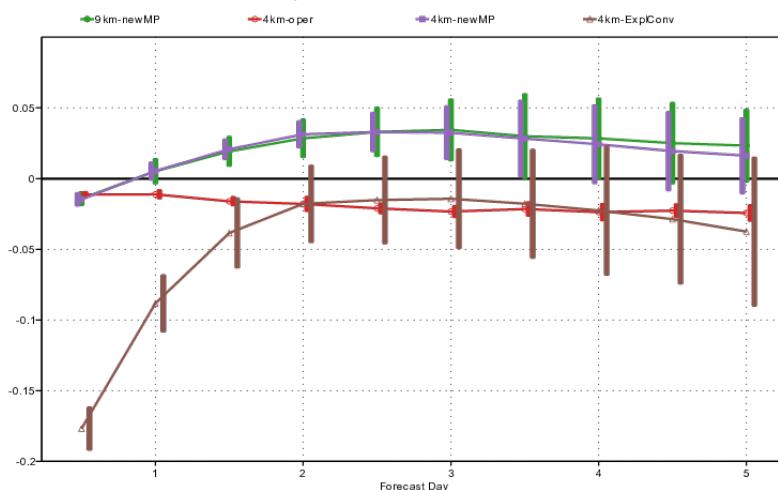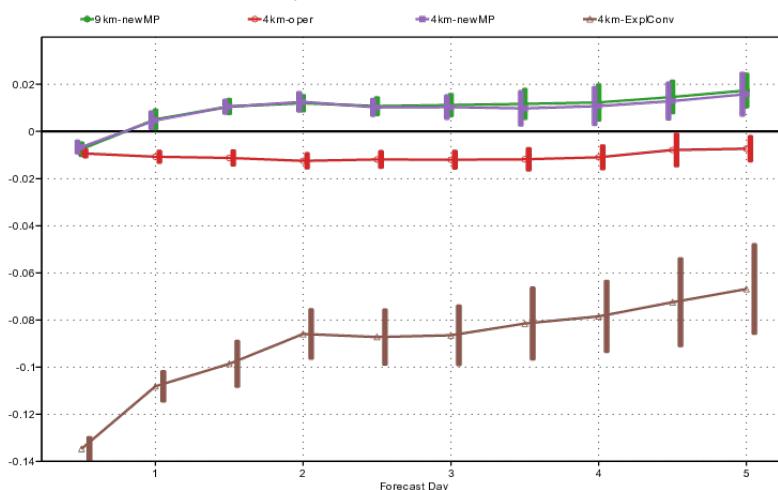
*Figure 70: Normalised RMSE difference to 9km-oper for model physics and resolution experiments for 700hPa temperature (top) and 700hPa wind speed (bottom) in the Tropics. (Positive means experiment better than control.*
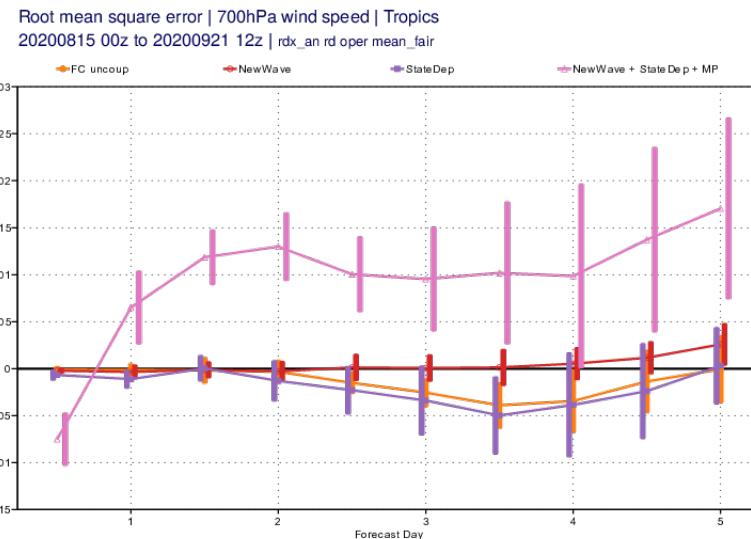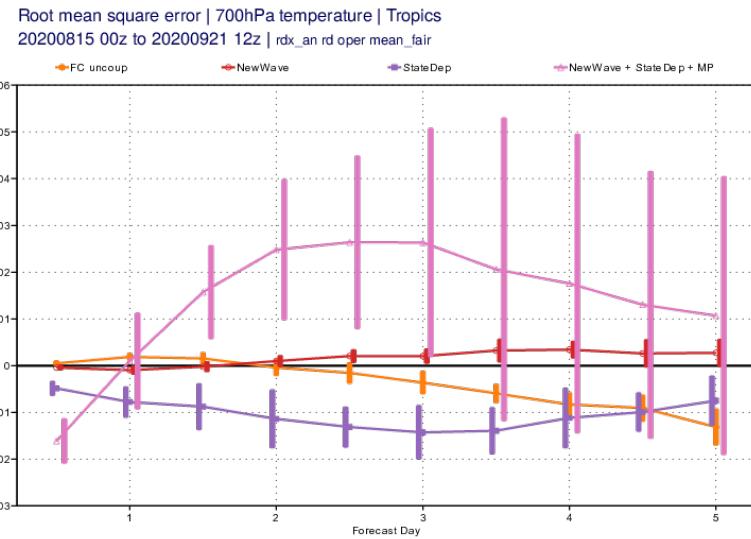
*Figure 71: Normalised RMSE difference to 9km-oper for ocean-related model experiments for 700hPa temperature (top) and 700hPa wind speed (bottom) in the Tropics. (Positive means experiment better than control.*