The challenges of Data Assimilation for Tropical Cyclones

Giovanna De Chiara, Linus Magnusson, Massimo Bonavita, Phil Browne, Eulalie Boucher, Mohamed Dahoui, Patrick Laloyaux, Anna Lombardi, Peter Lean, Michael Maier-Gerber, Magnus Lindskog, Benoit Vanniere, Ziga Zaplotnik, Sean Healy, Tony McNally

ECMWF



ECMWF Annual Seminar, 15th Sept, Reading, UK

Tropical Cyclones


Fire Island (Long Island – NY) 12/09/2026

Tropical cyclones are among the most devastating natural hazards

The damages are not only due to the strong wind and heavy rainfall, but secondary events such as storm surges, flooding, landslides play their role

Early warning systems are an important instrument that can help reducing TCs impact: NWP system is one of the elements

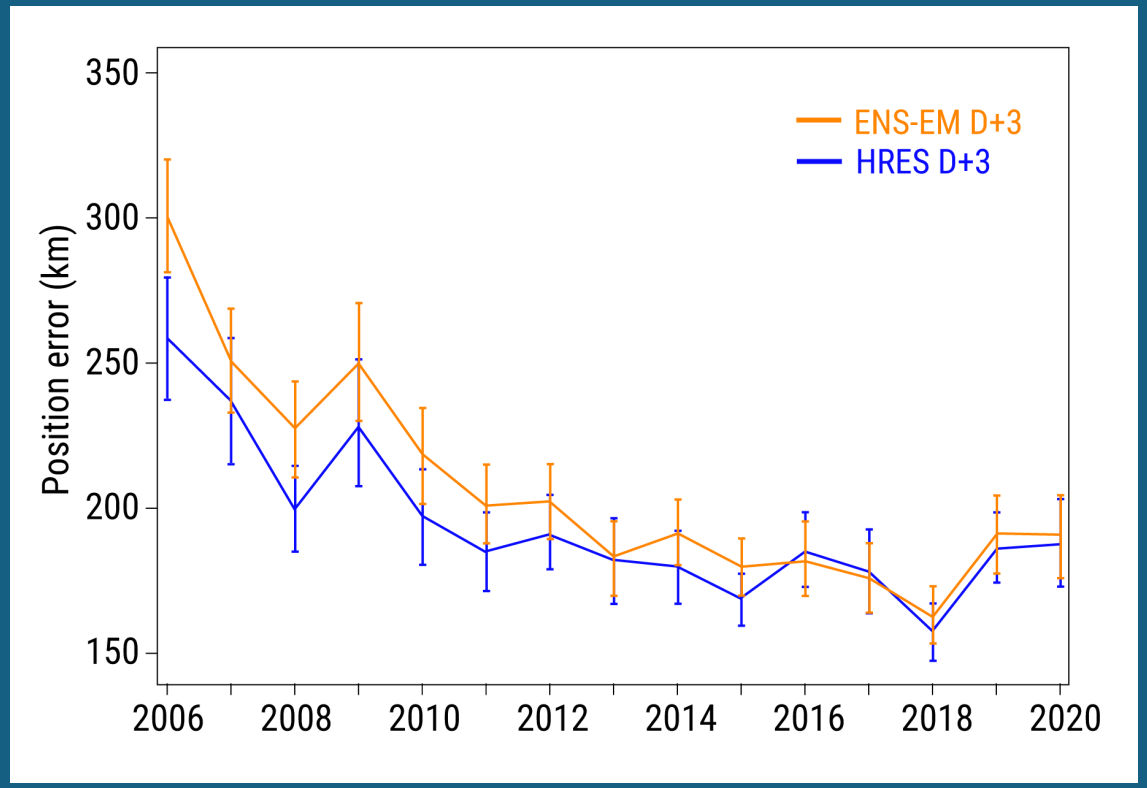
In NWP, we have two aspects to solve: TC **track** and TC **intensity**

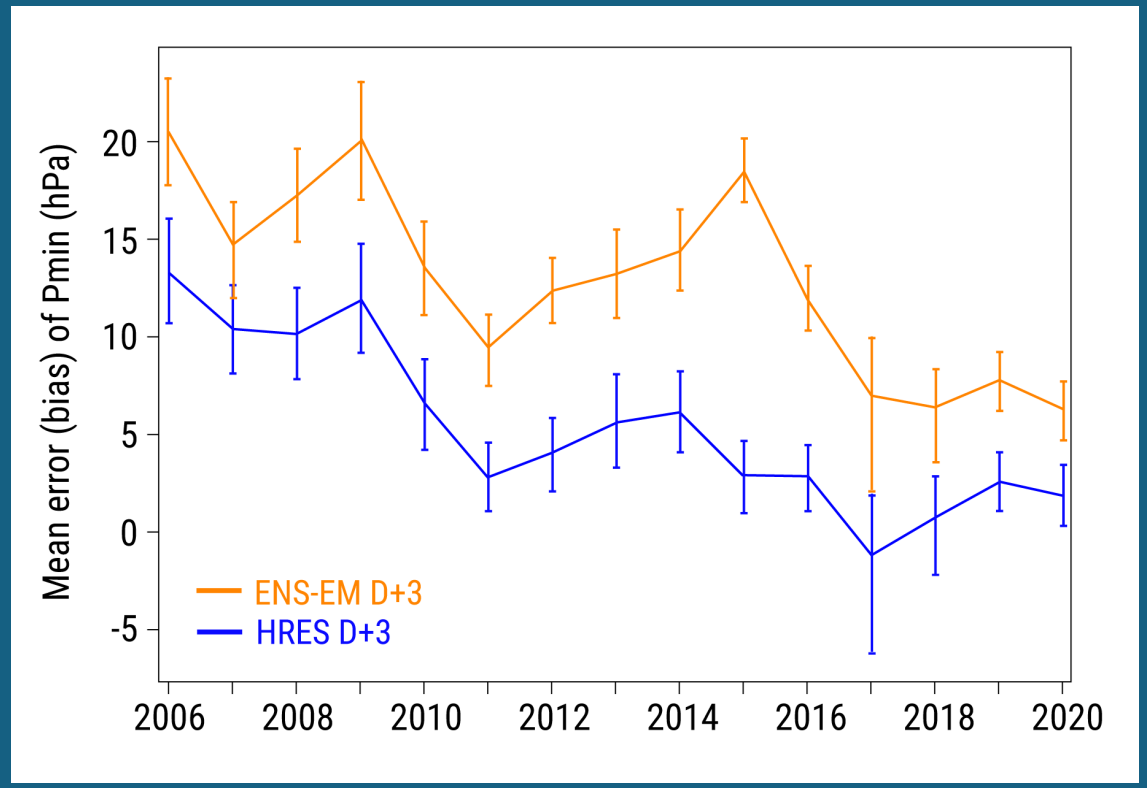
- we need to know in which direction a cyclone will move
- we need to get the intensity right

How good are we now?

TC forecast skills increased over the last 20+ years, mainly for the tracks

Position and Intensity day-3 forecast error over time (2006 – 2020)





POSITION ERROR at analysis time for different storm categories
226 Atlantic storm cases in 2020

All

Vmax < 50kt

Vmax 50- 63kt

Vmax > 64kt

- The majority of the cases has position error < 30 km
- The error is smaller for strong storms

We manage to pinpoint the position of well-defined systems

Magnusson et al., ECMWF TechMemo 888, 2021

The TC motion is determined by the *steering* flow, that is the large-scale atmospheric wind current that pushes a tropical cyclone across the ocean

INTENSITY ERROR at analysis time for different storm categories
226 Atlantic storm cases in 2020

All

Vmax < 50kt

- 100 out of 226 cases → very small error
- Larger errors (up to 40 hPa) for hurricanes
- In the majority of cases, the AN of Vmax is 5 m/s weaker than the Best Track

Vmax 50- 63kt

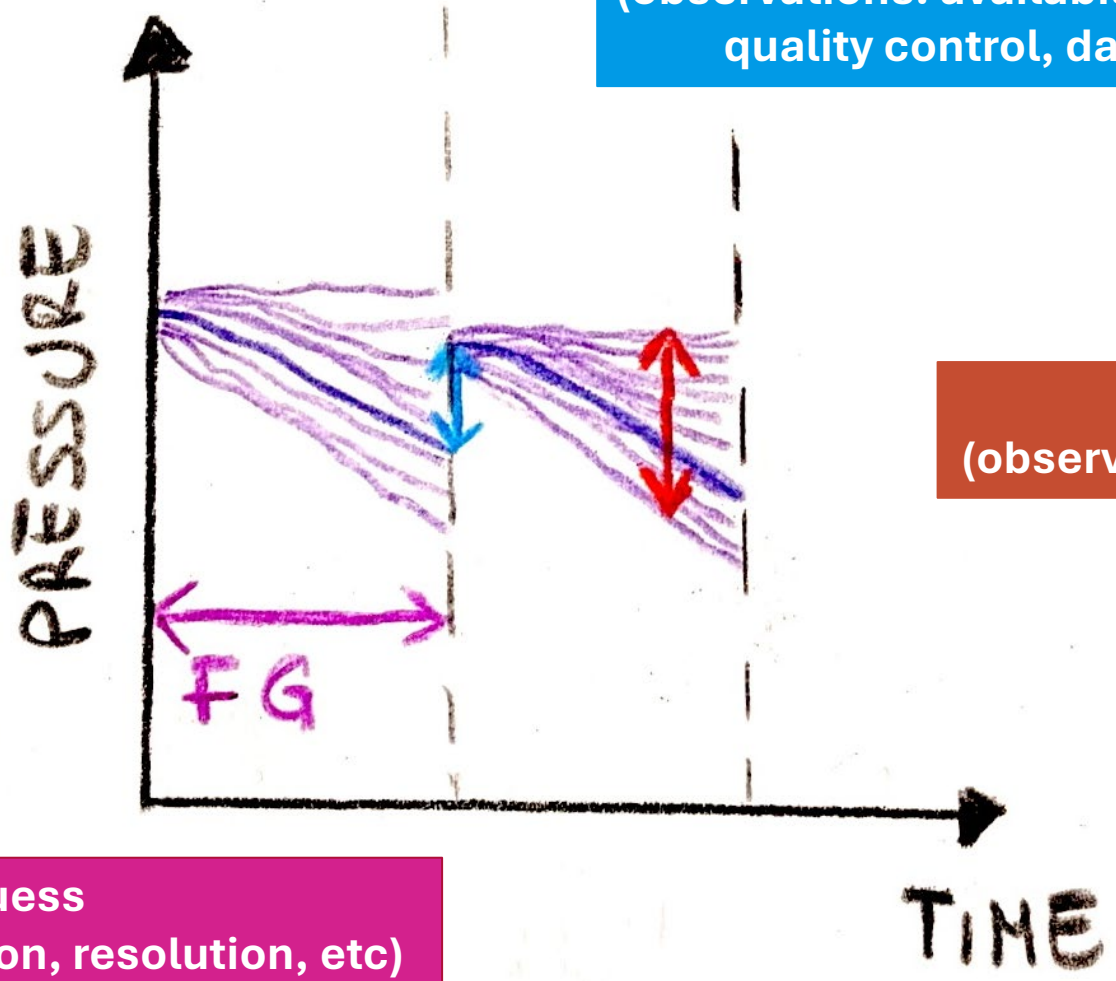
Vmax > 64kt

The largest initial intensity errors usually occur in strong TCs

Magnusson et al., ECMWF TechMemo 888, 2021

Tropical Cyclones recipe

Analysis Increments
(observations: availability and error characterization, quality control, data assimilation methods)

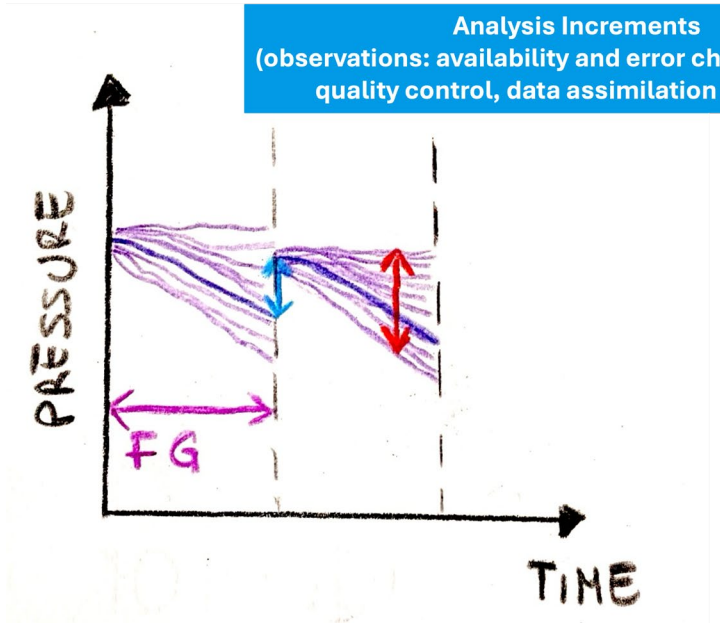


Ensemble Spread
(observations + model uncertainty)

First Guess
(model: parametrization, resolution, etc)

Tropical Cyclones and data assimilation

Several ingredients are involved



Observations

- Do we have enough observations close to the centre of the storm and in the outflow?
  - ✓ Challenging having conventional obs close to the TC centre
- Are the observations able to solve the spatial scale - especially sharp spatial gradients?
- How do we deal with the them?
  - ✓ Quality control, observation error and observation weight

Data Assimilation methods

- Are all the components of the Earth System coupled?
- Do we use the right resolution in the inner loop to solve the scales?

Quality Control and observation error

In data assimilation, observation errors are commonly assumed to be Gaussian.

The observation error and the background error together determine the weight of an observation in the assimilation:

- ✓ *Large** observation error → analysis draws less closely to the observations → smaller increment
- ✓ *Small** observation error → analysis draws more closely to the observations → larger increment

* Large/small relative to background errors

The same observation can have a small impact or big impact depending on how the background errors are distributed

The background error plays an important role → large BG errors can lead to:

- False rejection of a good observation
- Missed rejection of a bad observation

We apply two layers of QC:

First-Guess Check:

$$Y_{obs} - H(X_b) > L * \sqrt{(\sigma_o{}^2 + \sigma_b{}^2)}$$

Where:

$\sigma_o{}^2$ is the assumed observation error variance

$\sigma_b{}^2$ is the background error variance interpolated to the observation location

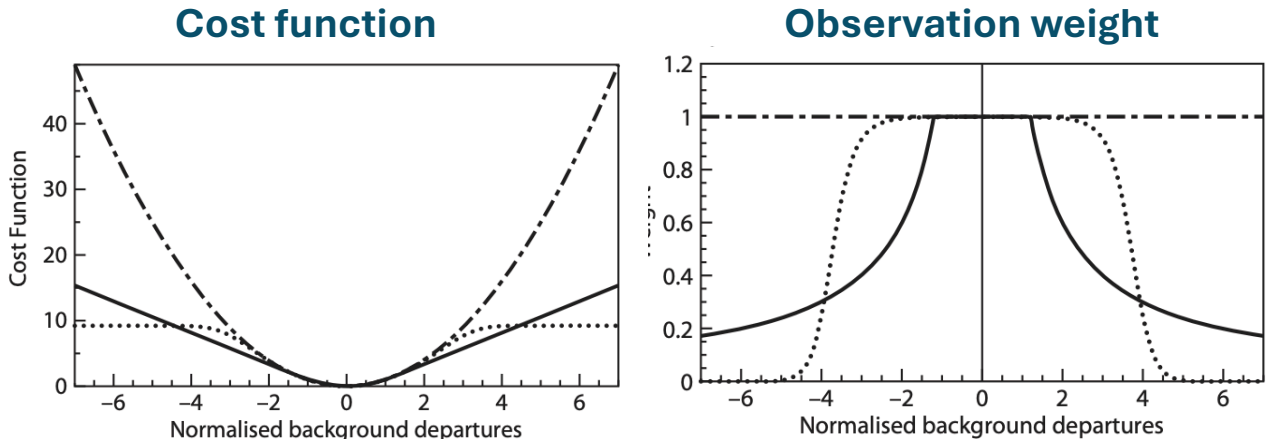
L tolerance factor (typically between 3 and 5)

In a "business as usual" situation, observations with large departure, i.e. 40 hPa, would be rejected without hesitation.

In extreme cases like TCs we should probably not!

Quality Control and observation weight

Variational Quality Control: assign a weight to each observation based on the departures and prescribed observation error



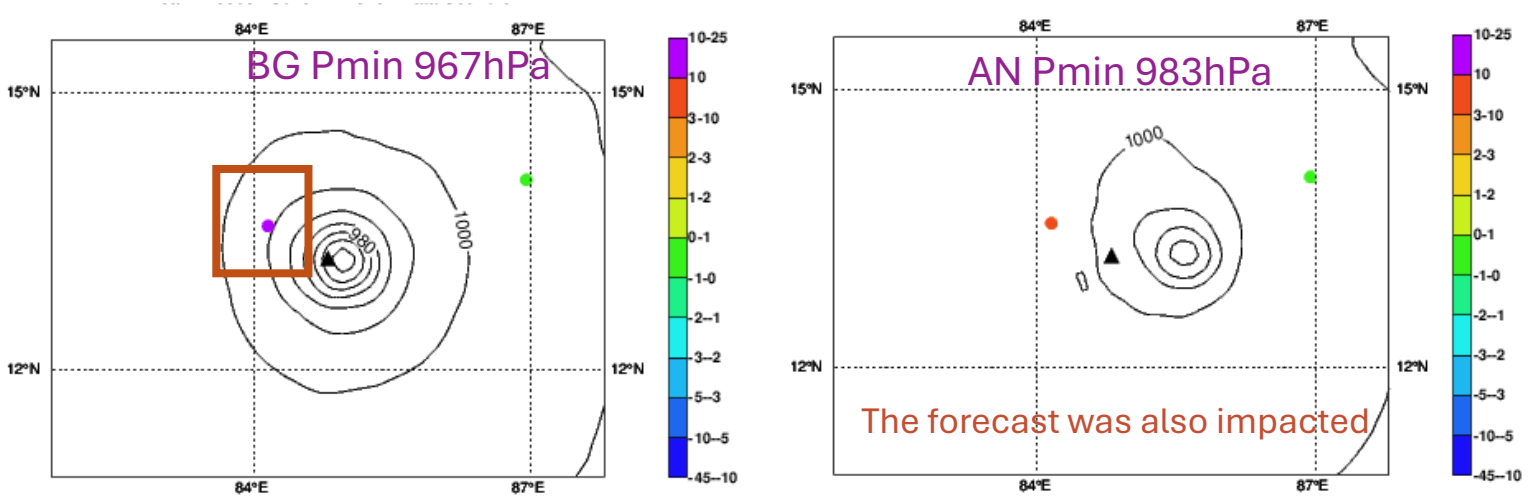
With Huber Norm, in case of large departures, instead of rejecting, we give the observations a lower weight so collectively they can influence and improve the analysis

Huber Norm is currently applied to most of the conventional observations (not to dropsondes geopotential height!)

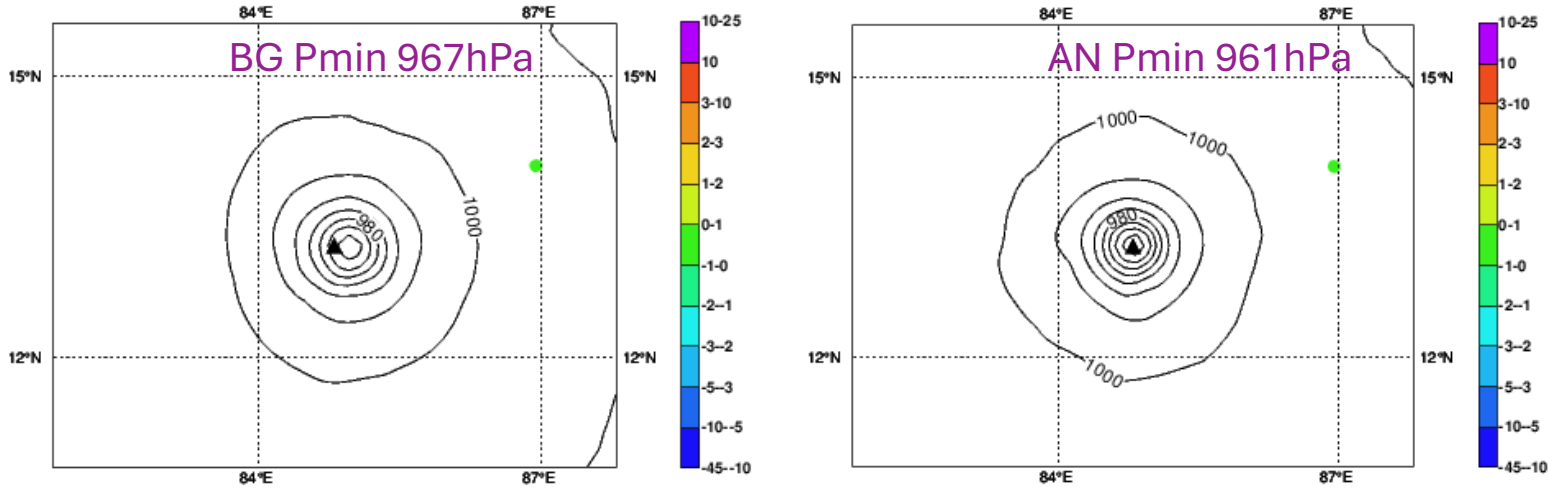
Tavolato and Isaksen (2014): https://www.ecmwf.int/sites/default/files/elibrary/2014/12572-use-huber-norm-observation-quality-control-ecmwf-4d-var.pdf

Quality Control of Observations: why is it important?

TC Fani 20190430 12UTC: a **faulty moored buoy** passed the quality control



Removing the buoy the analysis went down of more than 20hPa

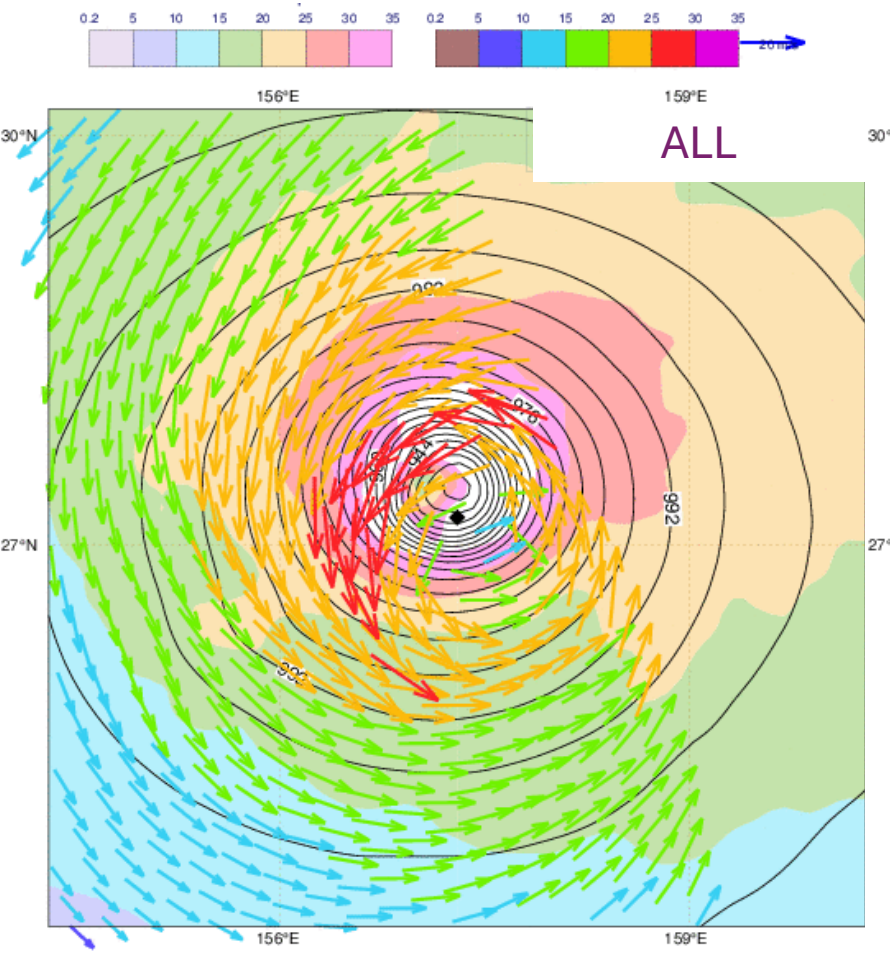


Scatterometer and TC

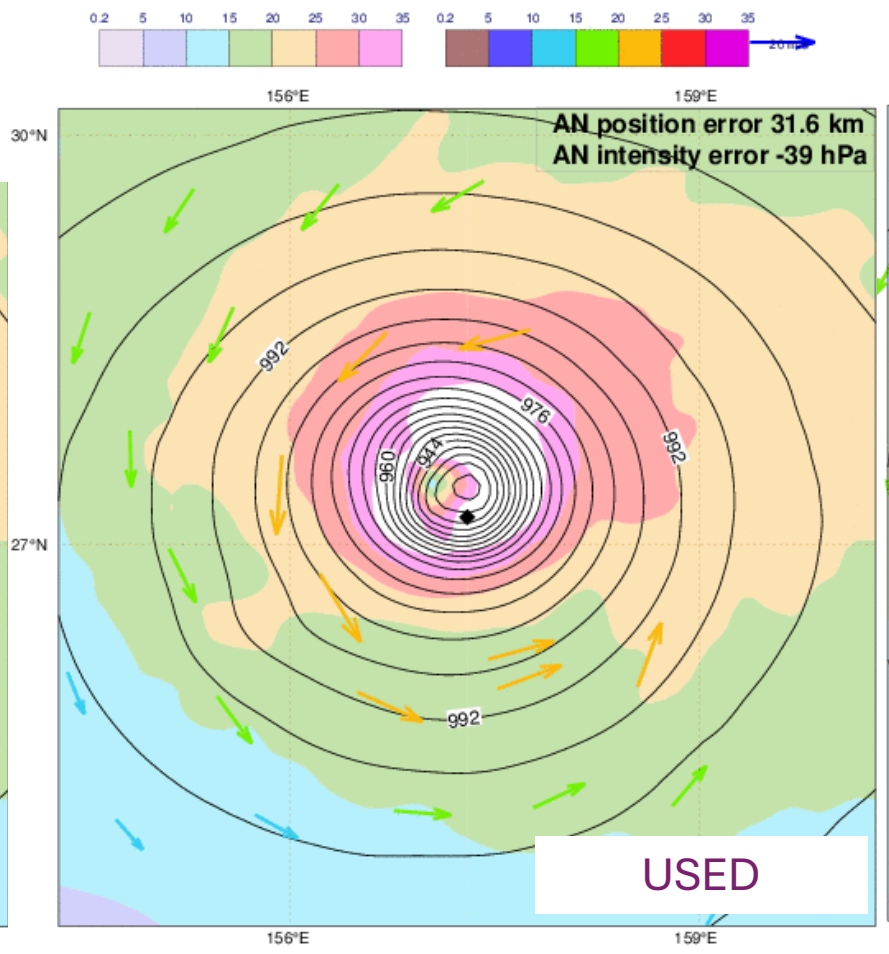
TC KILO – 2015090812 ASCAT-A Observations

Less observations due to:

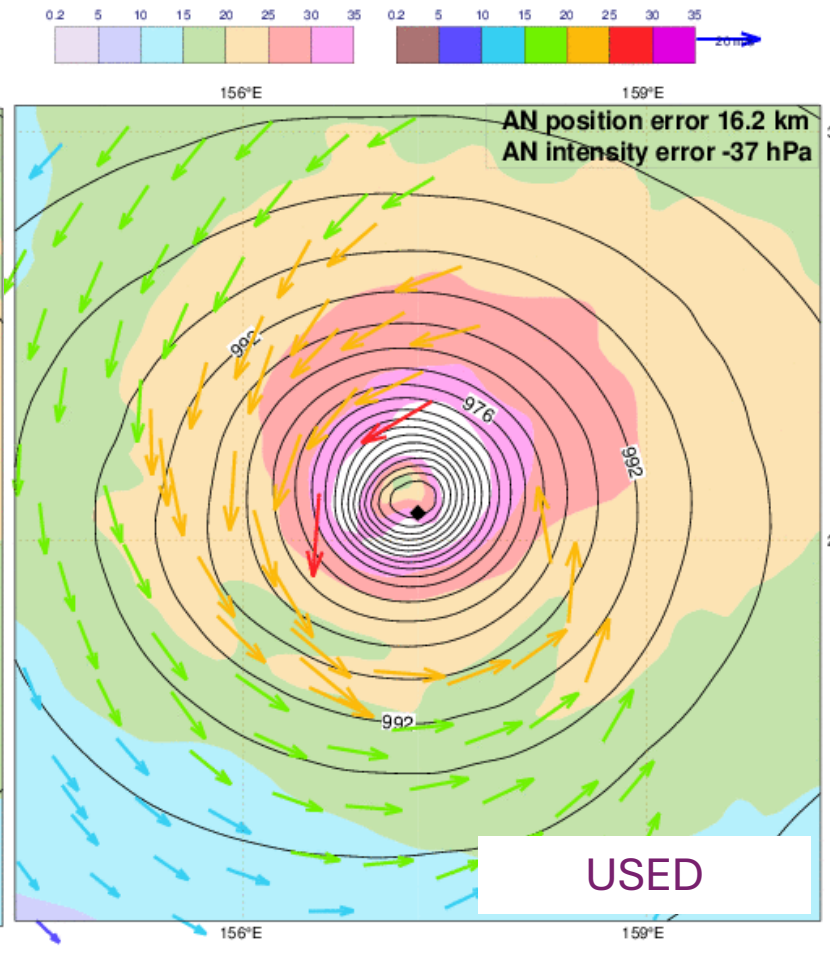
- Thinning



Thinning = 4



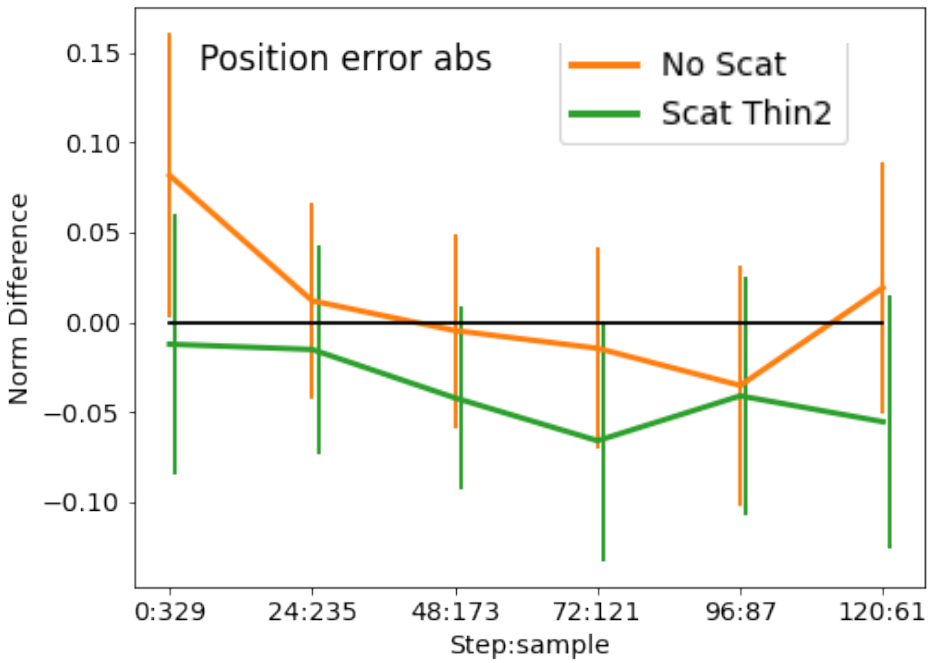
Thinning = 2



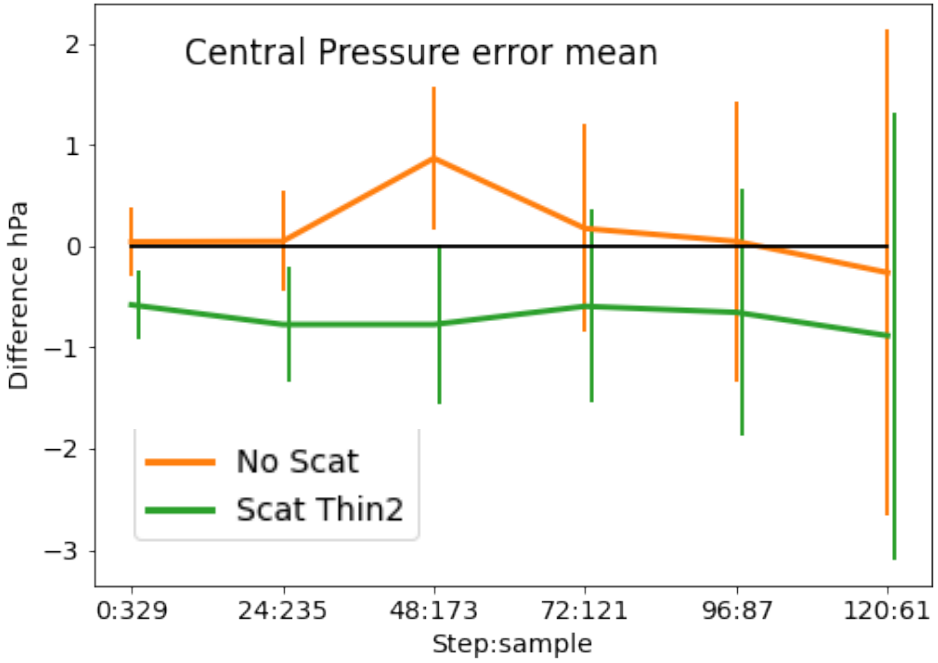
Scatterometer and TC

Impact of ASCAT reduced thinning on TCs

Normalized differences versus SCAT THINNING 4





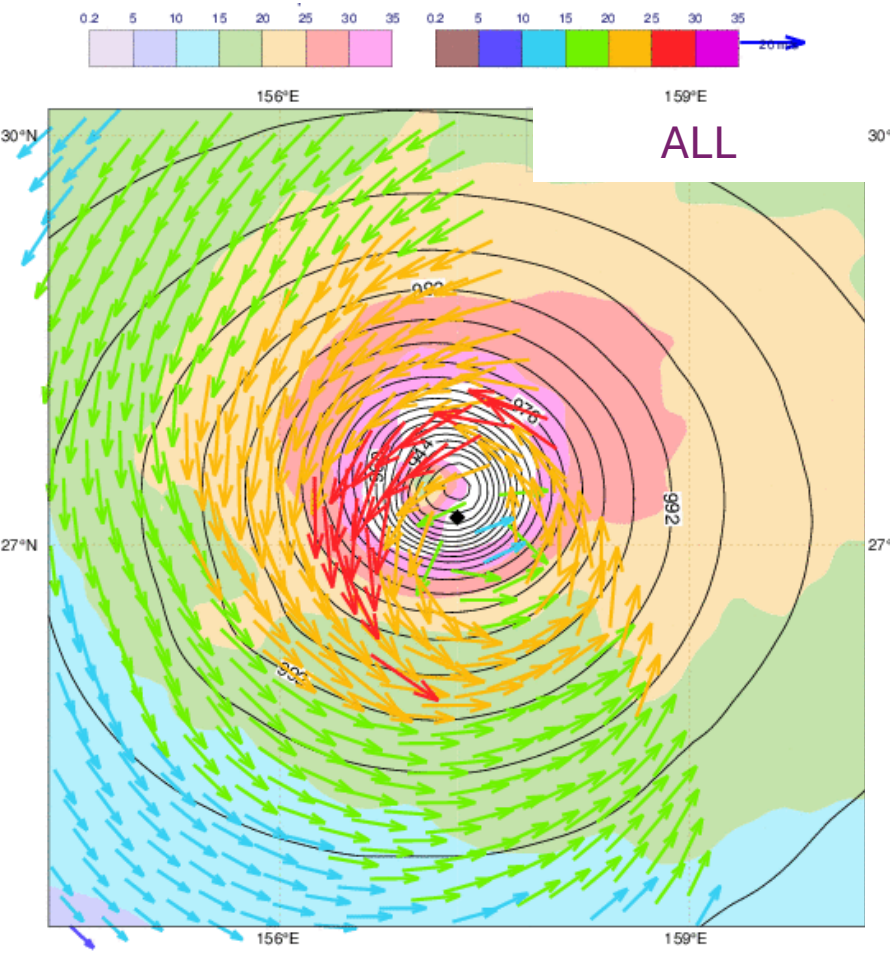


Scatterometer and TC

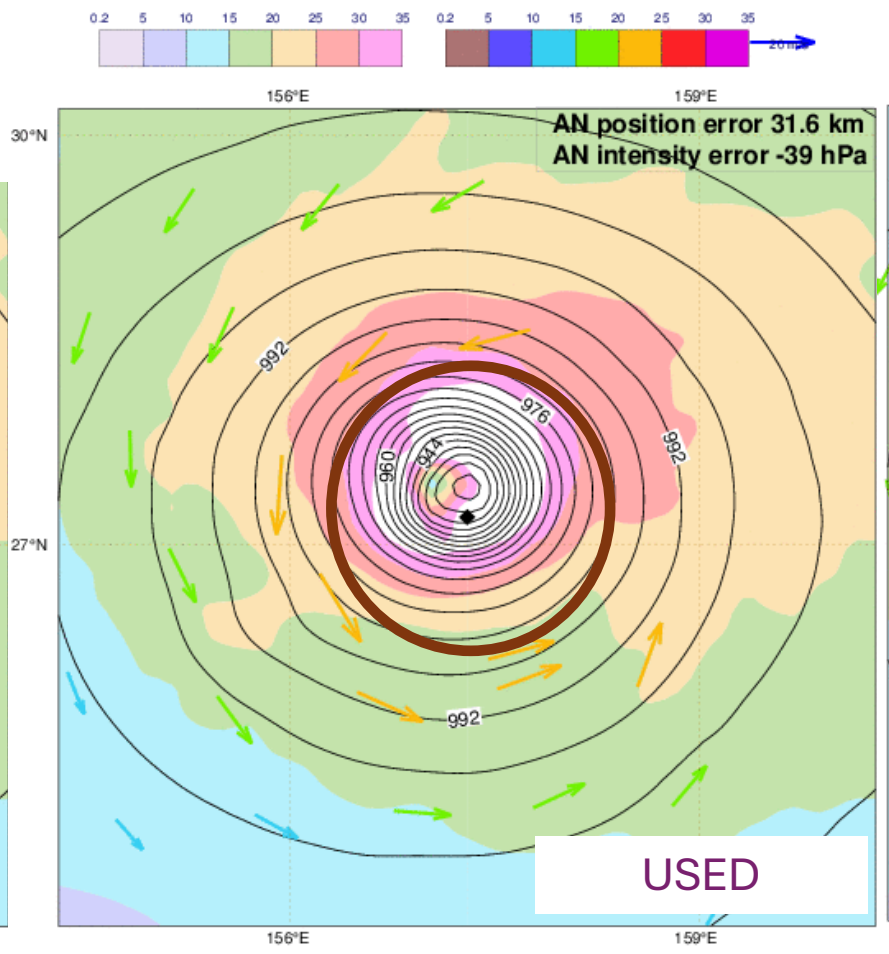
TC KILO – 2015090812 ASCAT-A Observations

Less observations due to:

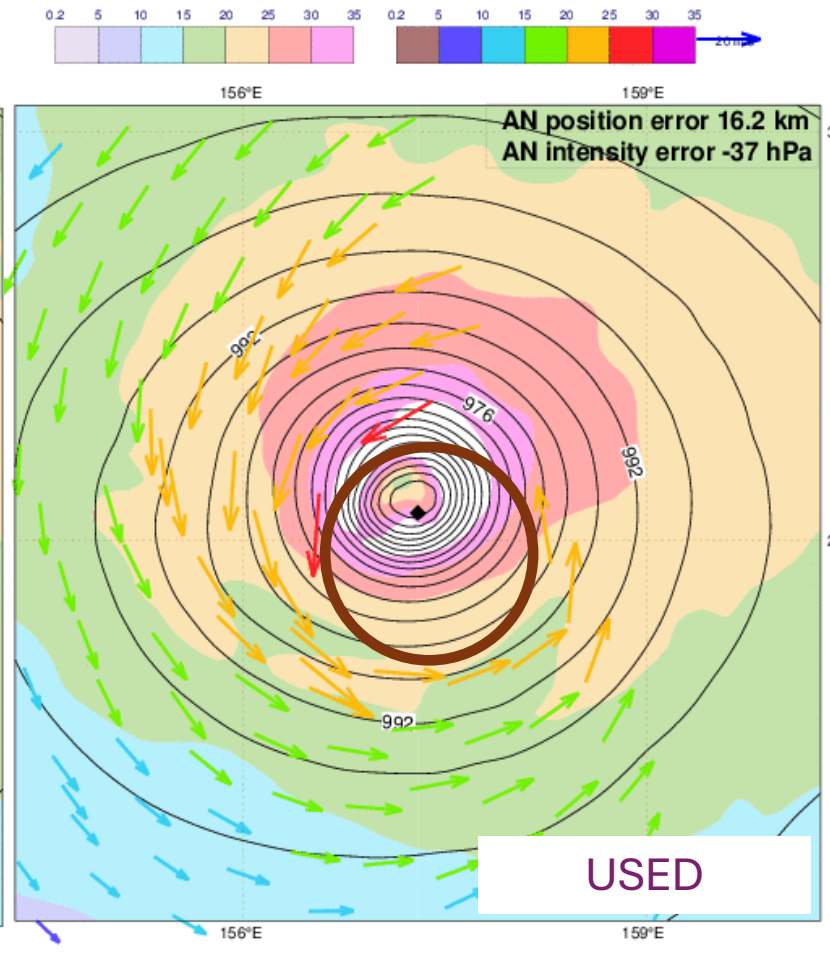
- Thinning
- VarQC



Thinning = 4



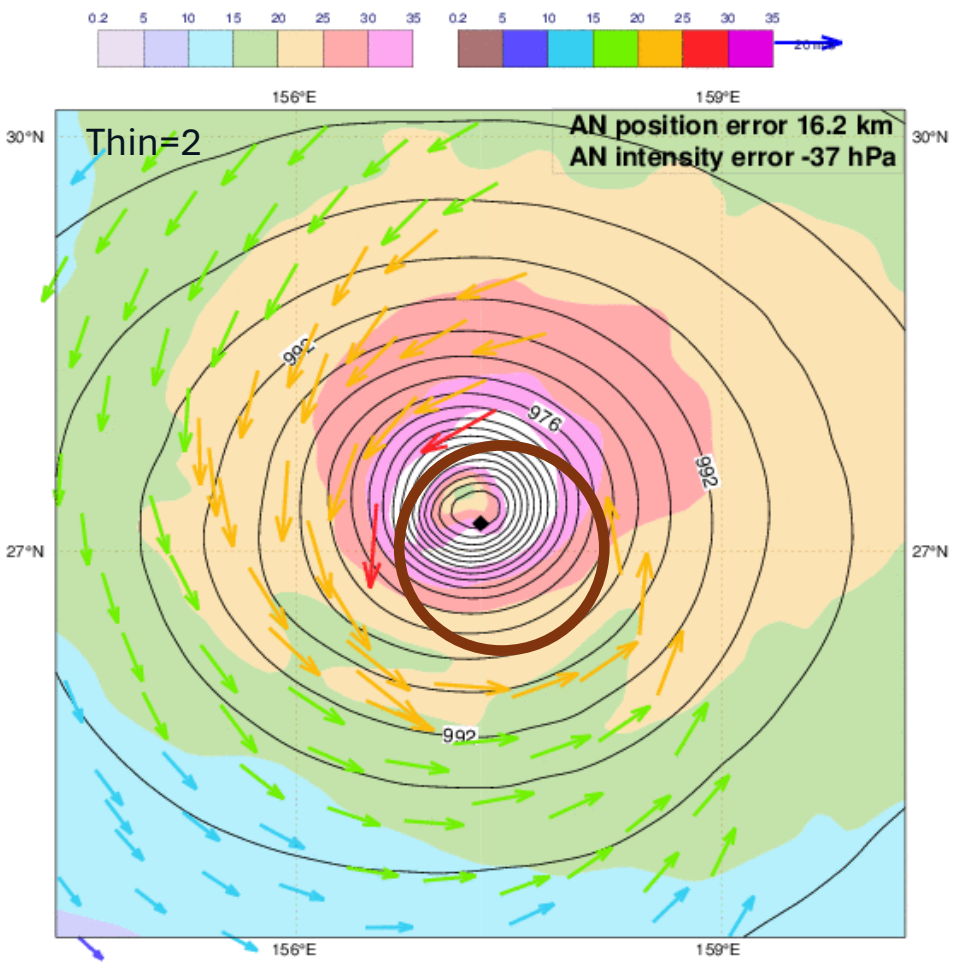
Thinning = 2



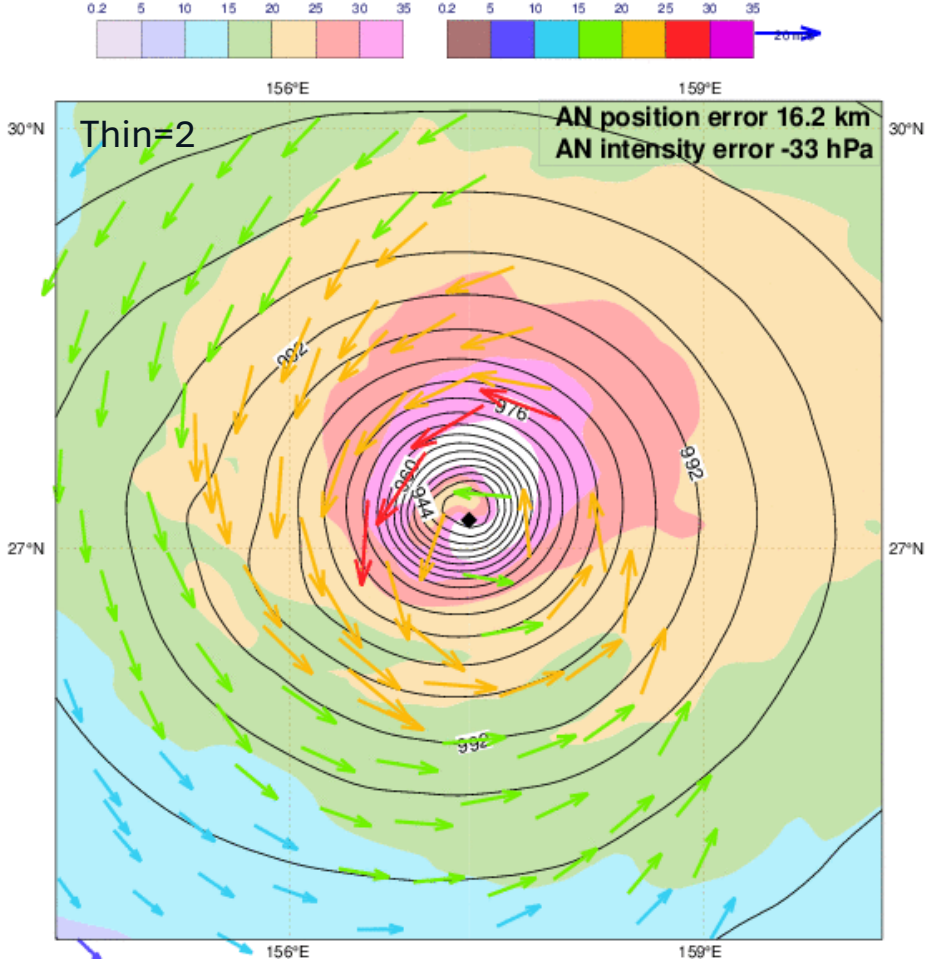
Scatterometer and TC

TC KILO – 2015090812 ASCAT-A Observations

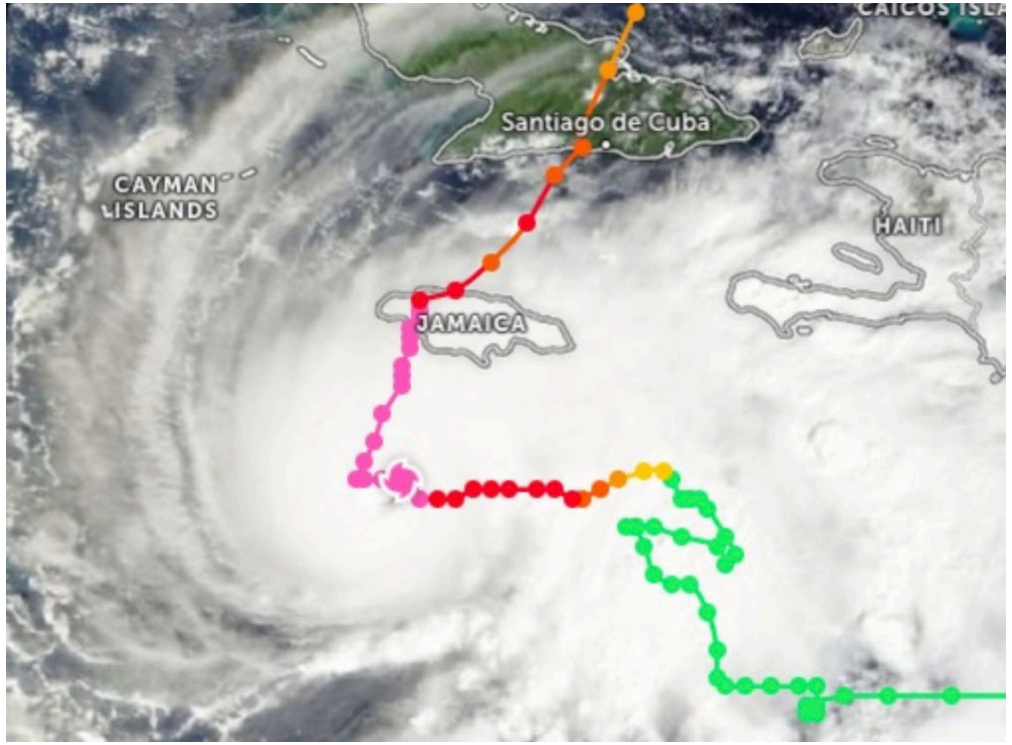
VarQC



Huber Norm



TC Melissa: Major Hurricane



NHC best track values:

NATIONAL HURRICANE CENTER and
CENTRAL PACIFIC HURRICANE CENTER
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION

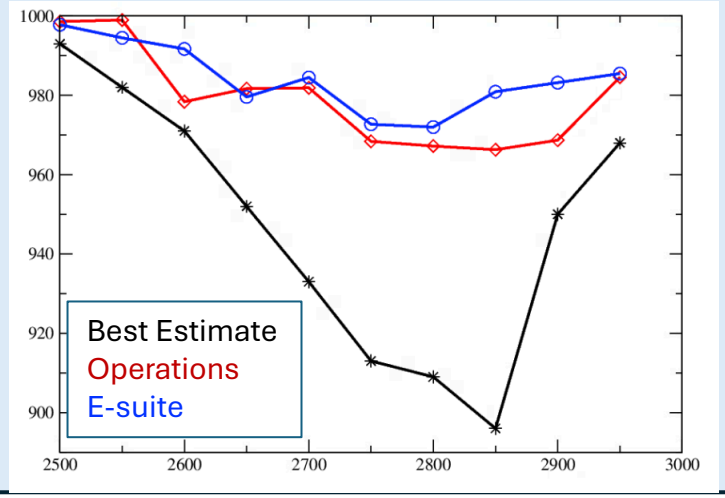
ANALYSIS & FORECASTS | DATA & TOOLS | EDUCATIONAL RESOURCES | ARCHIVES | ABOUT | SEARCH

Hurricane MELISSA

```
The next position update will be provided with the intermediate
advisory at 200 PM EDT (1800 UTC).

SUMMARY OF 100 PM EDT...1700 UTC...INFORMATION
-----------------------------------------------
LOCATION...18.1N 78.0W
ABOUT 25 MI...40 KM SE OF NEGRIL JAMAICA
ABOUT 145 MI...235 KM S OF GUANTANAMO CUBA
MAXIMUM SUSTAINED WINDS...185 MPH...295 KM/H
PRESENT MOVEMENT...NNE OR 25 DEGREES AT 9 MPH...15 KM/H
MINIMUM CENTRAL PRESSURE...892 MB...26.34 INCHES
```

Tropical Cyclone Melissa central pressure (hPa)



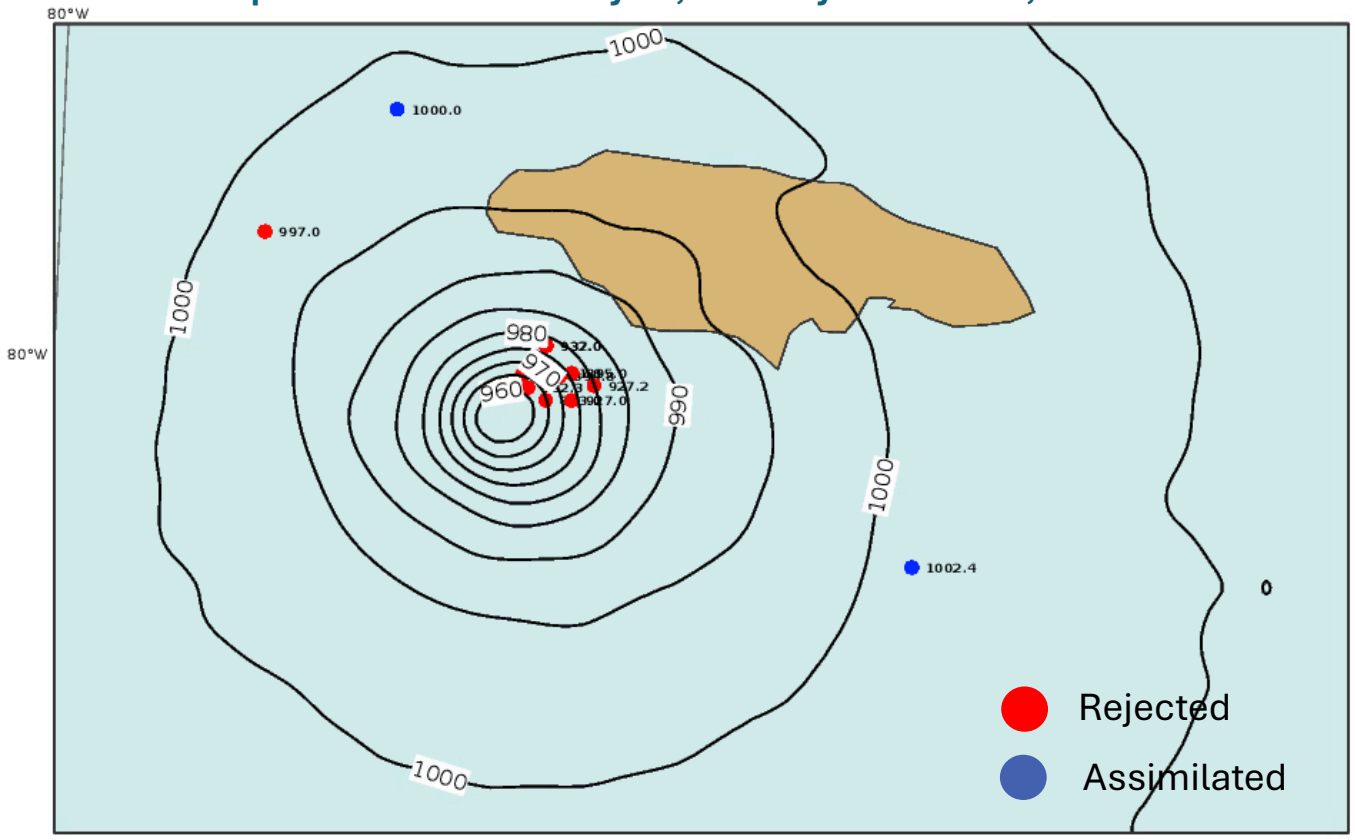
Dropsondes and TC

From dropsondes we assimilate:

- Wind, Temperature, Humidity (Huber Norm)
- Geopotential Height values within 50m from the surface - similar to Pmsl (VarQC)

TC Melissa

Operational MSLP analysis, Tuesday 28 October, 12UTC

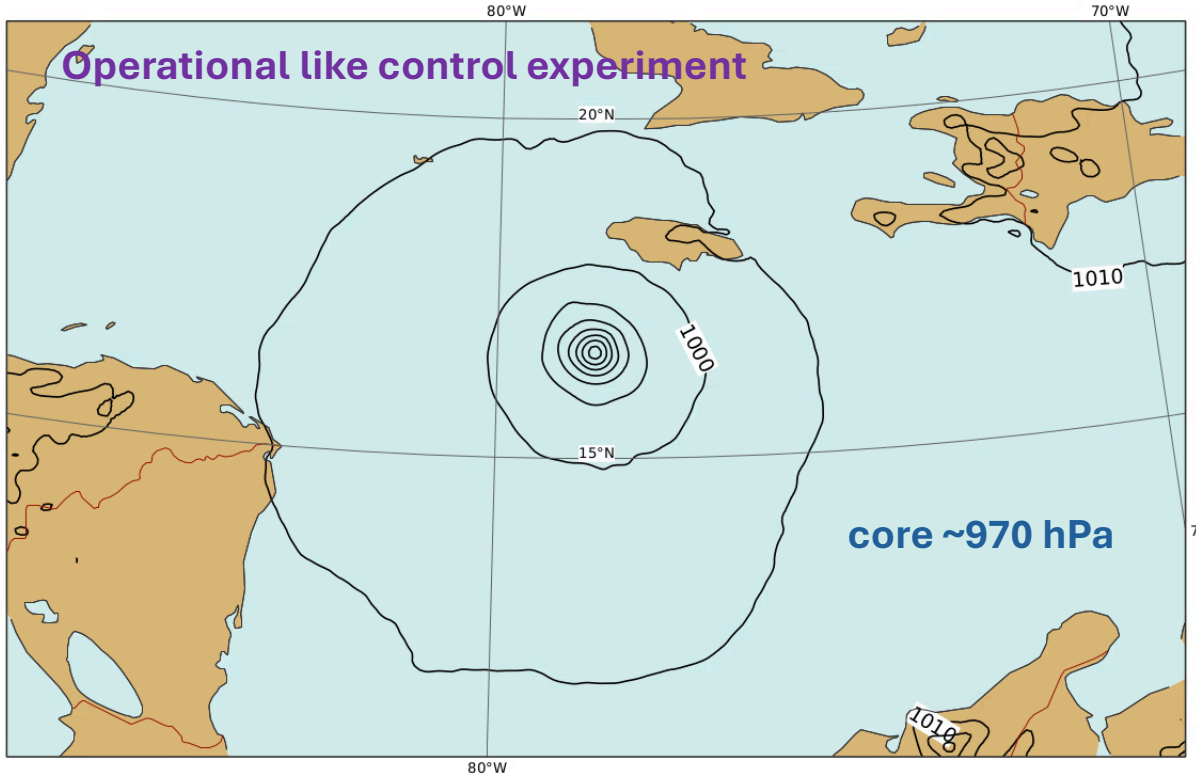


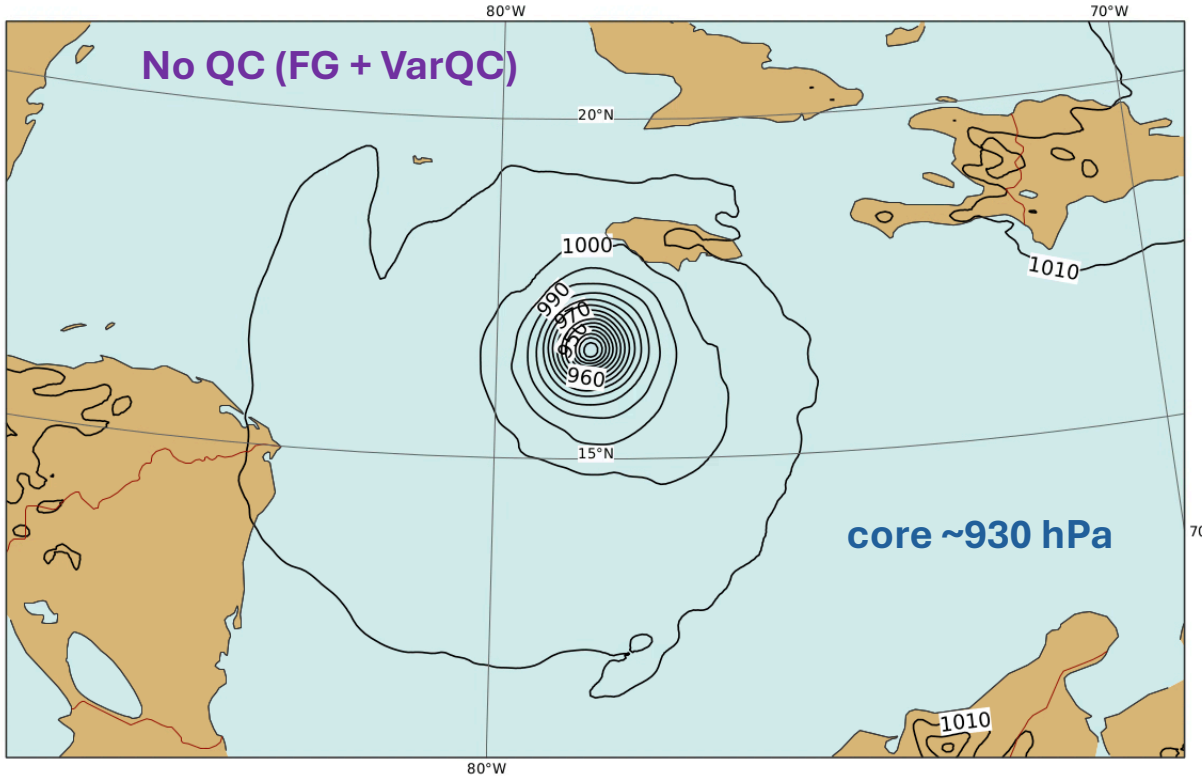
- Dropsonde geopotential height observations were rejected by the quality control (in the centre of the storm)
- First-Guess departures were up to ~50 hPa equivalent.

Dropsondes and TC

Sensitivity experiments: remove all the QC (FG + VarQC) ➔ we force all the observations to be used ➔ deeper analysis
The signal persisted up to 48 hours

TC MELISSA: MSLP **analysis**, Tuesday 28 October, 00 UTC



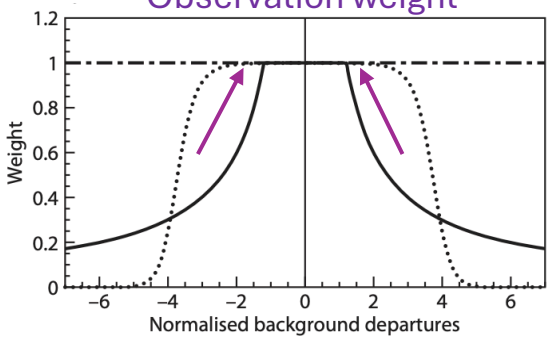


Dropsondes and TC

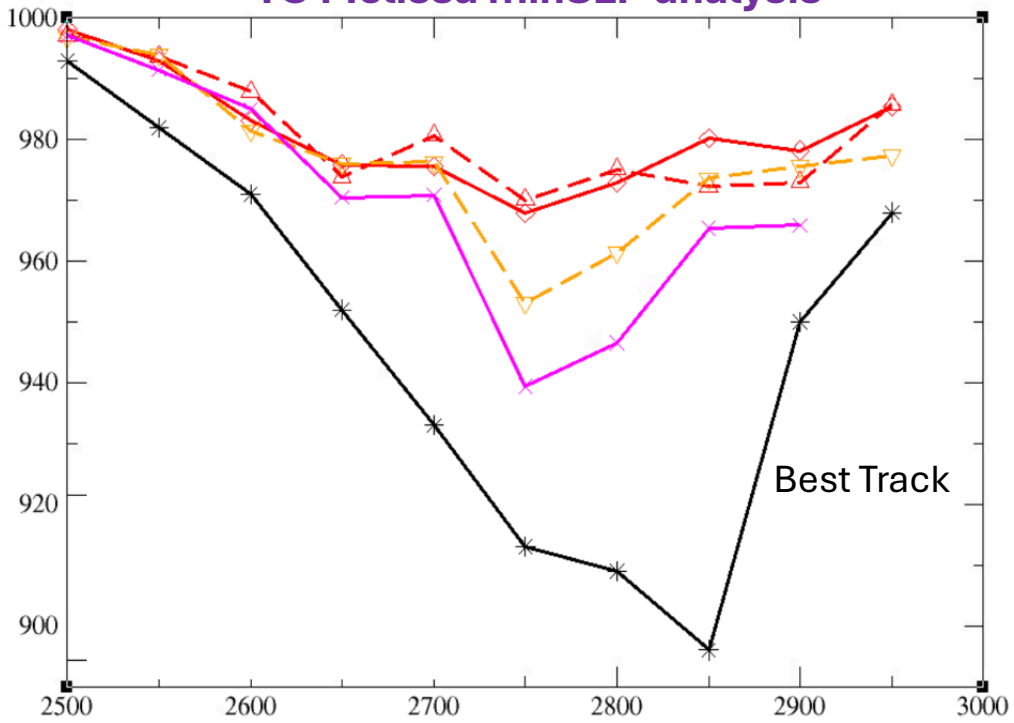
Additional sensitivity experiments relaxing the QC: no FG check but different setup options for the Huber Norm

No FGchk
No FGchk + VarQC Huber Norm lim 3
No FGchk + VarQC Huber Norm lim 6
No FGchk + VarQC Huber Norm lim 10

Observation weight



TC Melissa minSLP analysis



- Additional geopotential height observations improve the analyses of TC Melissa even with a relaxed QC
- This is a safer compromise, a much more controlled situation that removing the QC completely

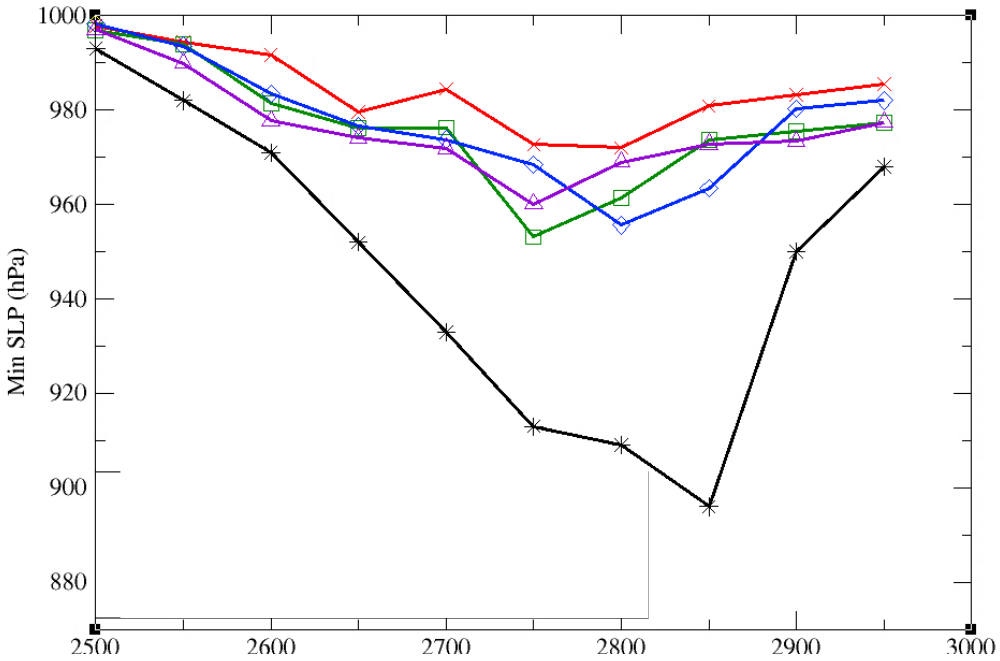
Dropsondes and TC

The computation of the background error for the geopotential height (Z) is not optimal near the surface: QC is too tight

Sensitivity experiment testing a new computation of Z for dropsondes:

e-suite
No Z FG check, Huber norm left/right=6
New Z FG check, Huber norm left/right=6
New Z FG check, Huber norm left/right=6, Higher resolution

TC Melissa minSLP analysis



- By using the new FG check for Z, large part of the improvement seen disabling the QC completely is recovered (15-20 hPa)
- This new configuration (the blue one) will be tested on other TCs and will be considered for operational implementation

Assimilation of the Best Track Pmin

BAROS (Best Approximate of Real Ocean Surface pressure)

At least four times a day, each of the Regional Specialized Meteorological Centres (RSMCs) and Tropical Cyclone Warning Centres (TCWCs) produce **estimates** of the **position** and **intensity** for all present TCs in their basin

They are subjective human assessment of the TC centre location, intensity, and structure produced using all the observations available (satellites and aircraft missions in the Atlantic) at the time of analysis

These “observations” (estimates!) are often referred to as Best Track

Other NWP centres (UK Met Office, JMA, BOM, NCAR/NCEP) use the TC bogus observations

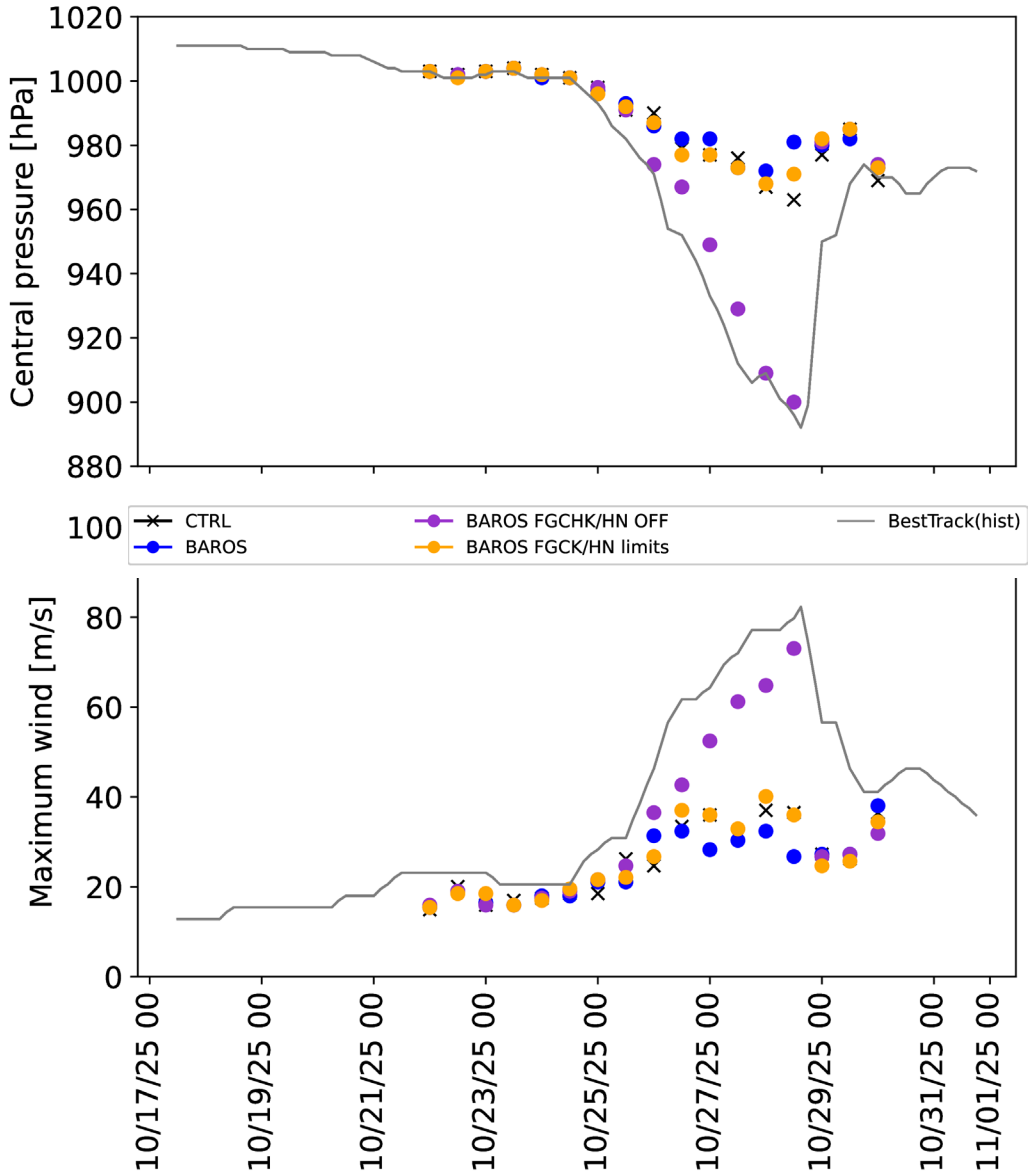
Can the best track Pmin help improving TC analysis and forecasts in IFS?

Sensitivity experiments assimilating the minimum central pressure from the Best Track (called BAROS at ECMWF!) were run showing promising results.

Tests were performed using different observation errors, interpolation time, FG-check and Huber Norm configurations.

TC Melissa: central pressure and max wind speed at analysis time

Tests with different QC configurations (observation error: 0.47 hPa)

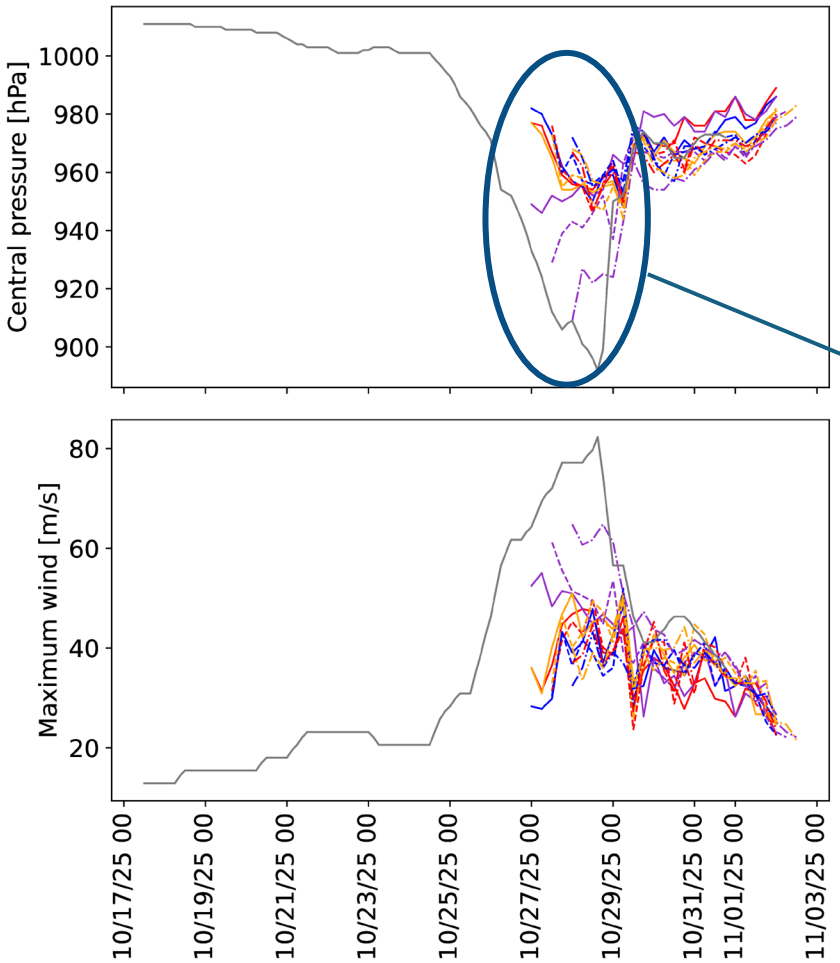


X CTRL (BAROS out)
BAROS IN
BAROS IN – relaxed FG check limits and HN limits
BAROS IN - FG check OFF / HN OFF

Removing completely the QC, with 1 single observation we manage to push the analysis down to 900 hPa

TC Melissa: central pressure and max wind speed forecast

forecasts starting at different AN times

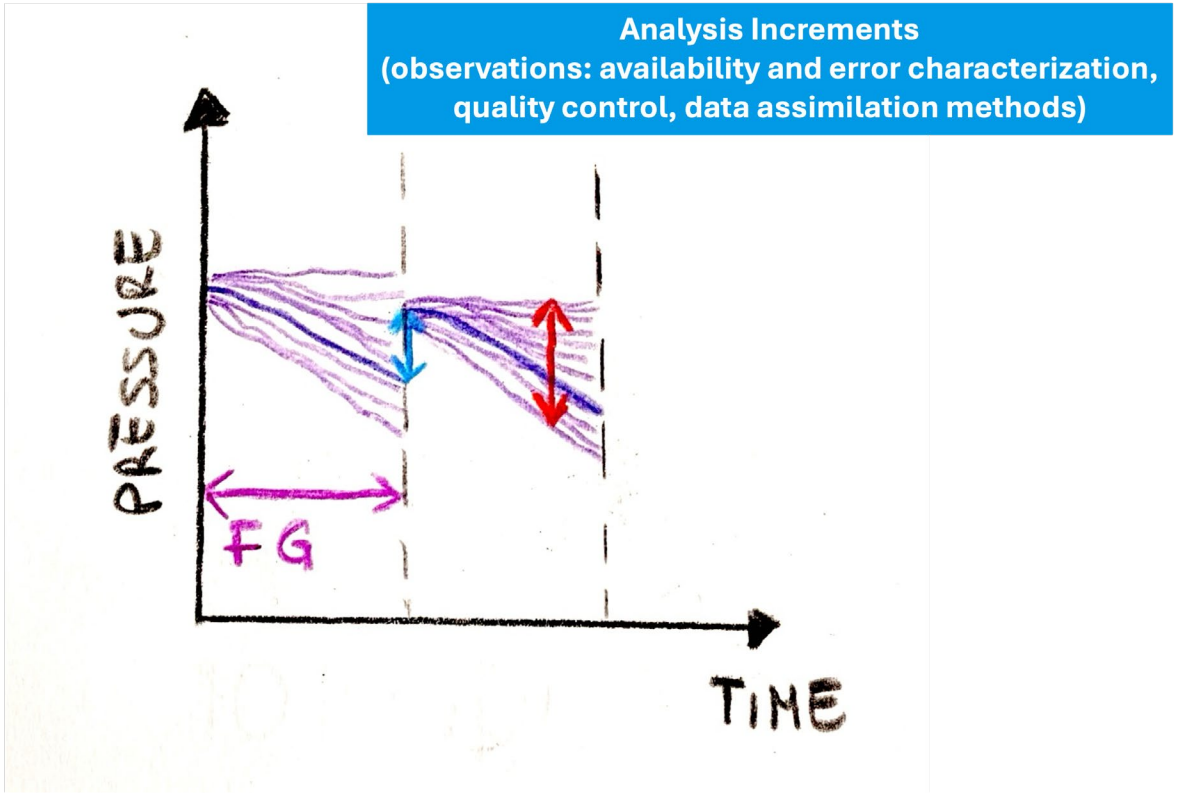


CTRL
BAROS
BAROS FGCK/HN OFF
BAROS FGCK/HN limits
BestTrack

- Better forecast thanks to a better analysis
- The model is not able to predict an intensification of the storm

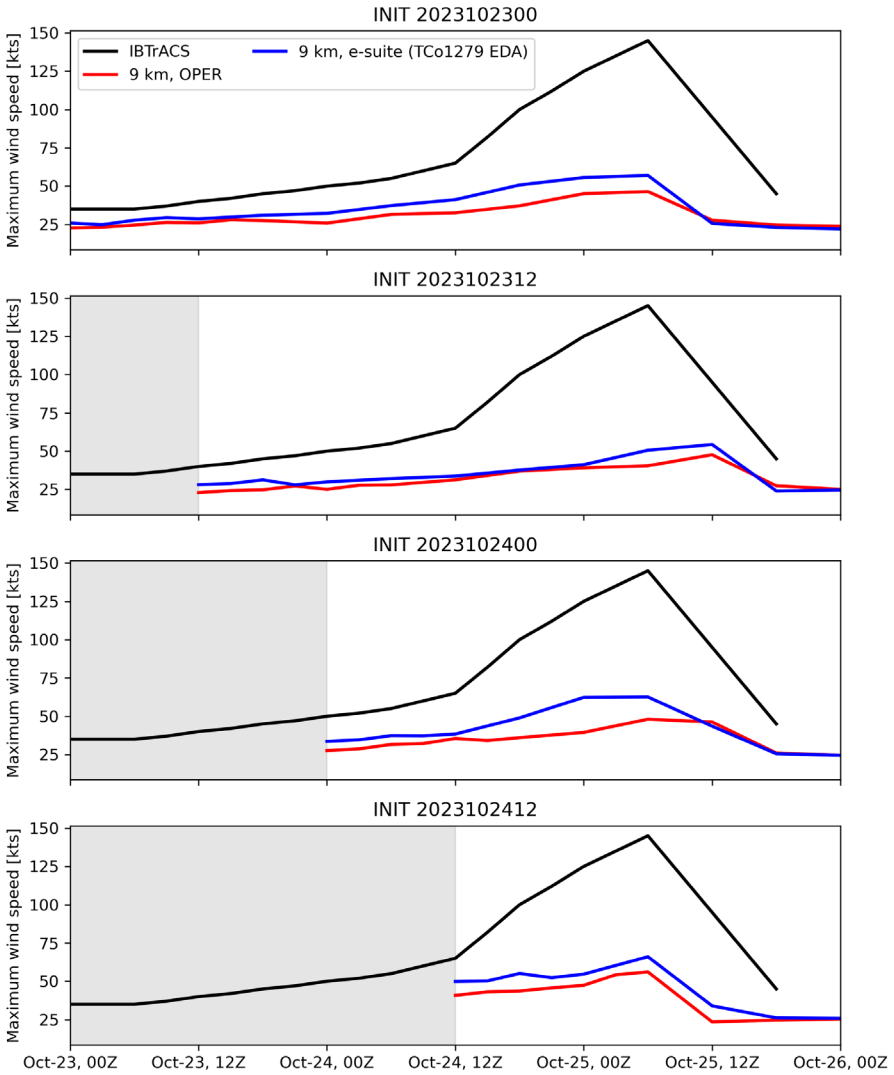
We cannot remove the QC completely, but how can we build a system that is reliable in all the situations?

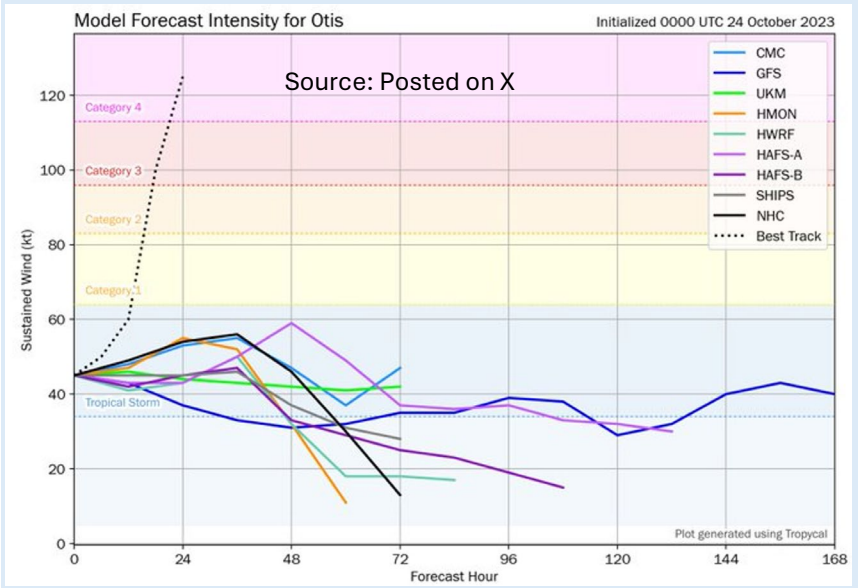
More extensive study will be run over a various TC seasons to optimize the QC scheme and the observation errors



DA methodology: 4D-Var Resolution

TC Otis - October 2023





None of the global or regional deterministic models or machine learning models were able to ”capture” the rapid intensification
→ **a major forecast bust**

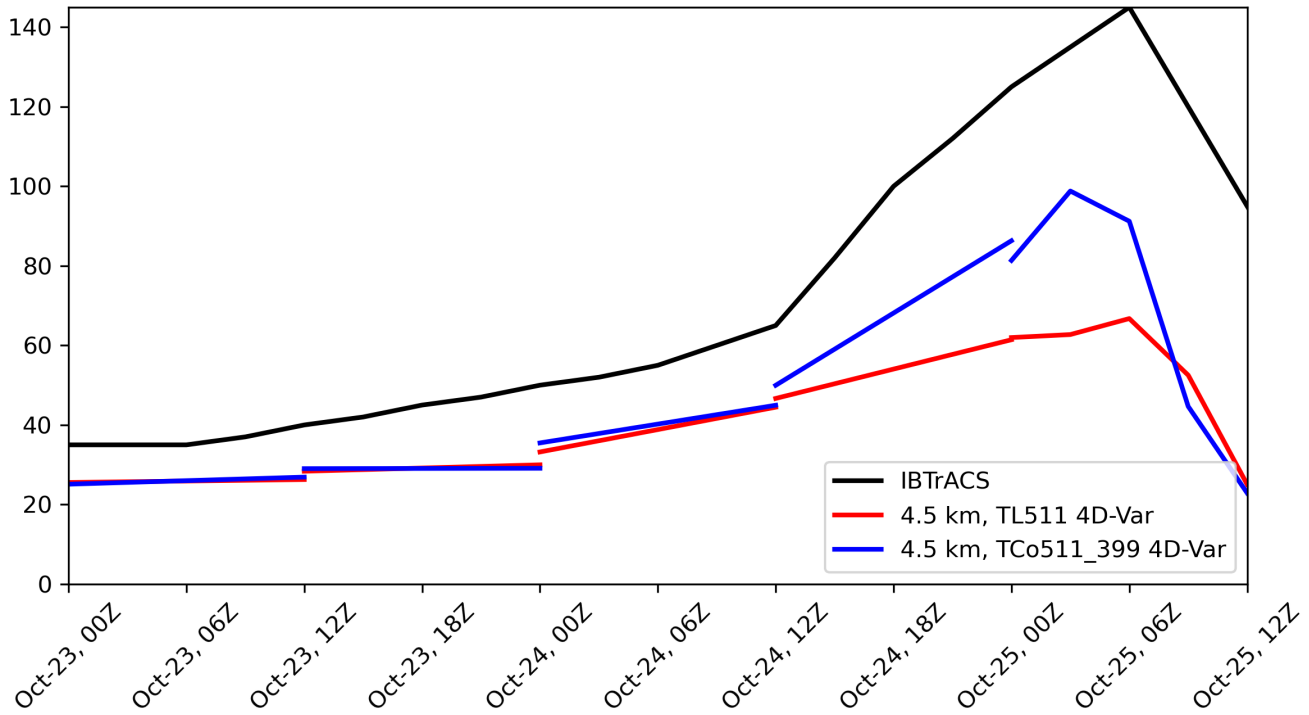
- East-Pacific hurricane which Hit Mexico on October 25th as a Cat5 (Pmin 923 hPa, Vmax 270 km/h)
- Rapid intensification from a tropical storm to Cat-5 in less than 24h
- IFS was not able to “capture” the rapid intensification ➔ due to deficiencies of the FC model
- Not many observations close to the centre of the storm: ASCAT and 1 reconnaissance flight (dropsondes) available but rejected (or given low weight) due to large departure
- The e-suite (Cy49r1) did slightly better than the operational IFS (Cy48r1) ➔ larger EDA relative vorticity spread in the e-suite

TC Otis (October 2023)

The 4D-Var was not able to properly capture the large spatial gradients in the core of TC Otis with the operational-resolution linearized models (inner loop)

Does a resolution of 4D-Var minimisation help this case?

Several tests have been done reducing the outer and the inner loop resolution

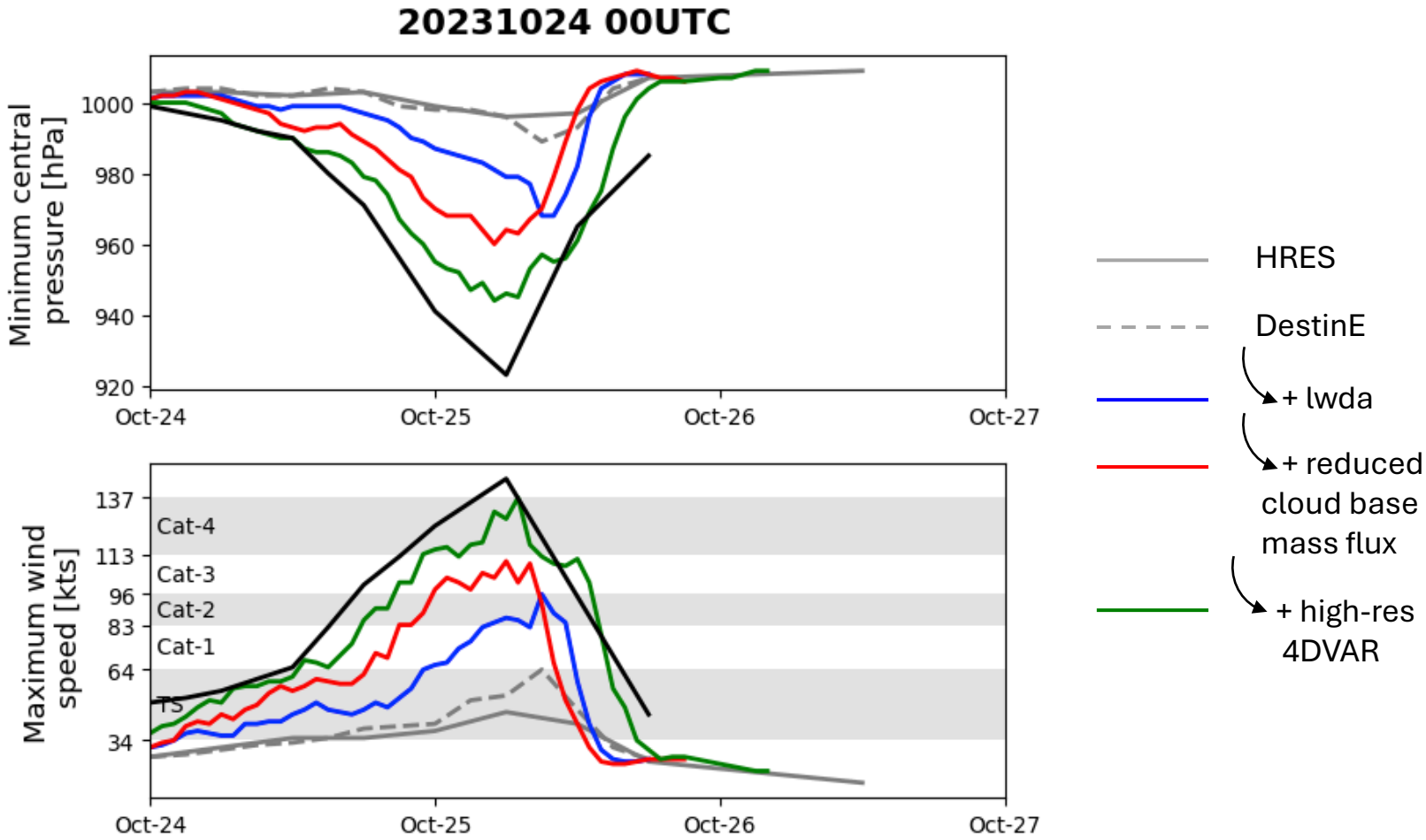


It does help!

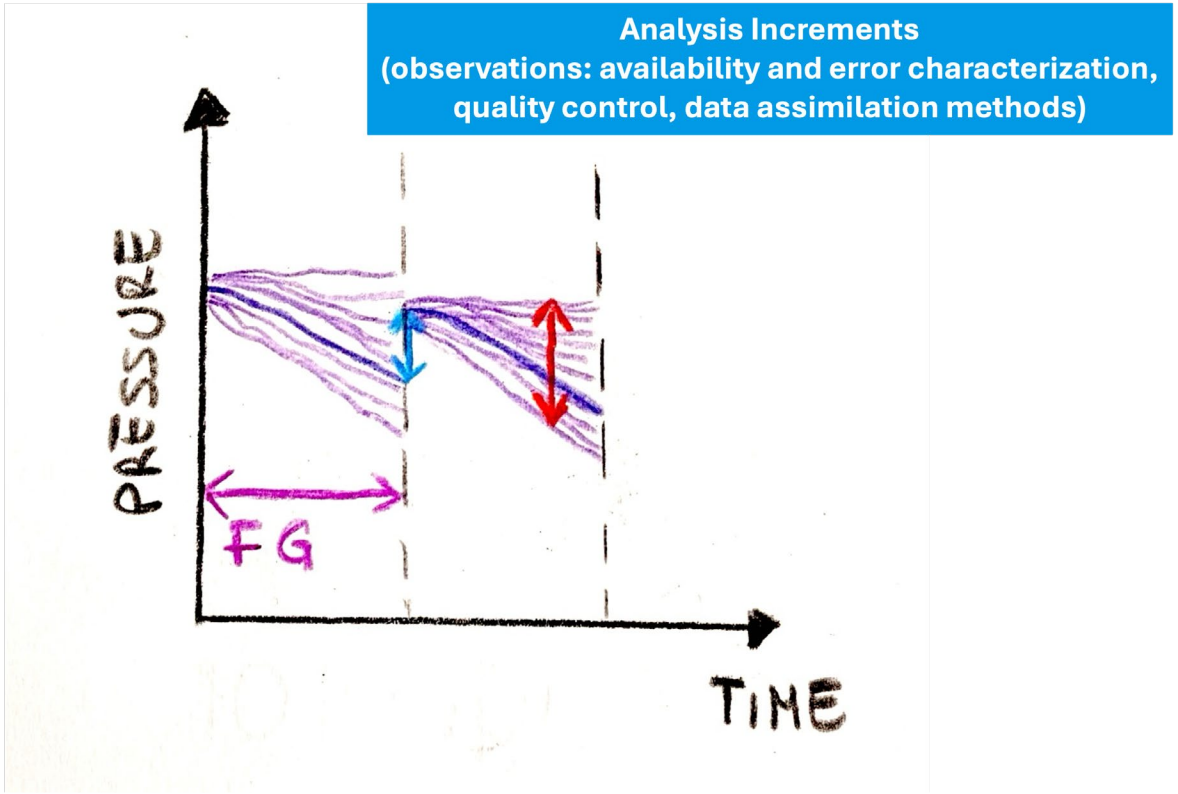
With high-res 4D-Var ➔ more pronounced warm core ➔ more intense low-level cyclone, more divergence at upper-level, but little difference of low-level convergence.

Predicting the supposedly unpredictable TC Otis

Combining successively extra observations, reduced parametrized convection and higher resolution 4D-Var dramatically increases the forecast skill of Otis.



+20 hPa due to assimilated data, 20 hPa due to resolution of the 4D-Var and 20 hPa due to convection settings



DA methodology: Atmosphere-Ocean coupling

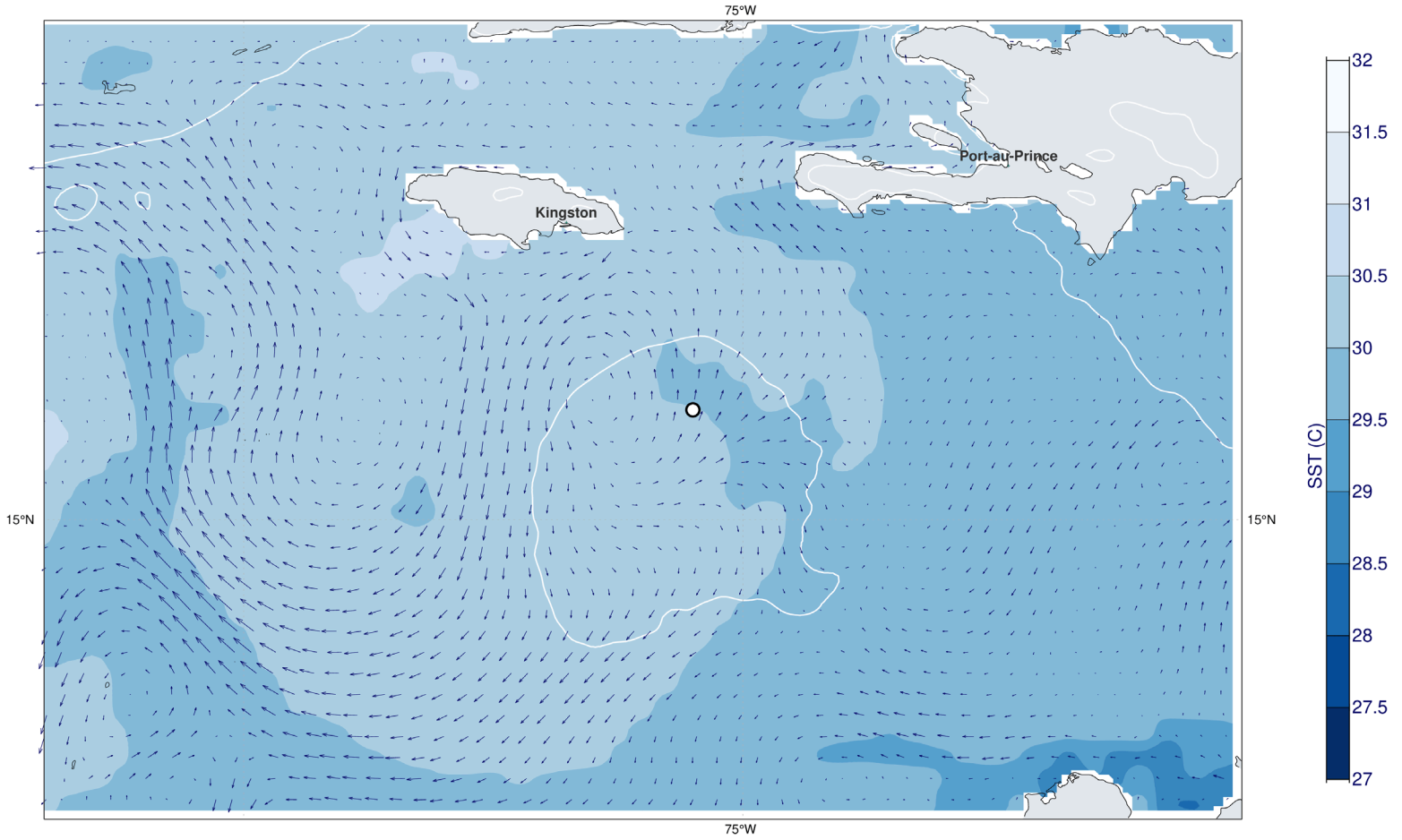
Impact of ocean coupling on TC

In a coupled system the SST and the ocean currents are updated every 6 (or 12) hours from the ocean components (in an uncoupled system the SST id updated daily).

TC Melissa 24-28 October 2025

TC melissa - AN 20251024 0UTC
Best Track mslp: 1001hPa

Coupled Run: SST (shaded area), Ocean Currents (vectors) and Best Track minSLP (white circle)



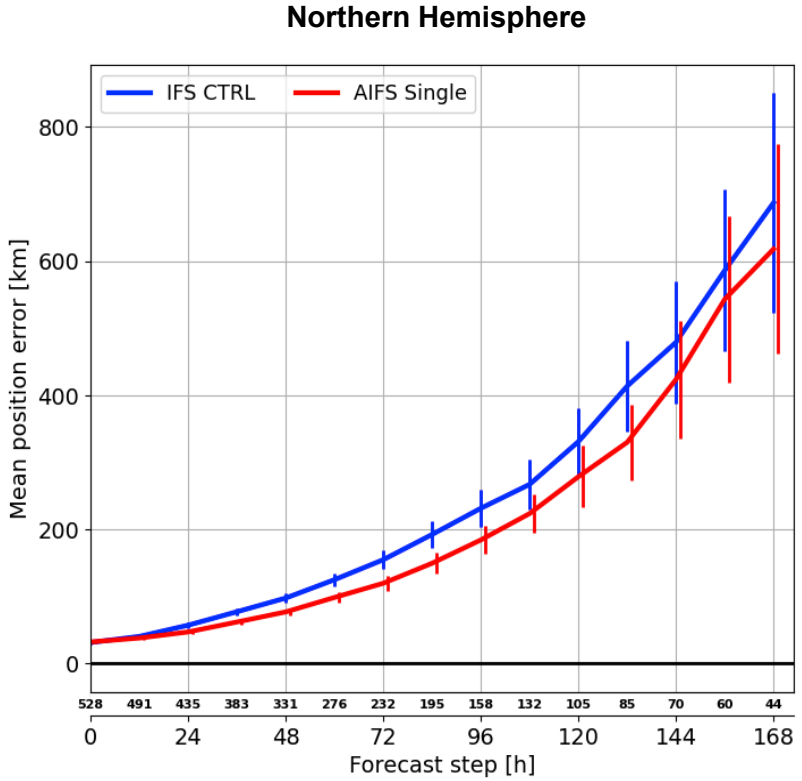
In the coupled system we expect:

- More realistic ocean-atmosphere processes
- SST more consistent with the atmospheric state (i.e. cold wake)
- Ocean currents impact the atmosphere through the ocean-air interface and interface observations (like scatterometers)

Tropical Cyclones in Global Data-Driven Forecasting Models:

AIFS and AI-DOP

AIFS Single vs IFS Control: TC track verification



Verification period: Jun - Nov 2025

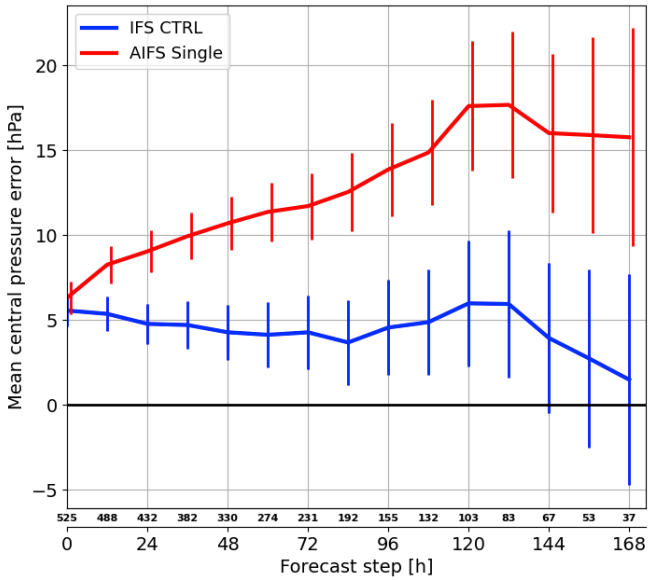
Standard Verification
- Against IBTRACS
- Homogeneous sample
- TC present at initialisation

- AIFS Single had significalty better tracks than IFS Control up to 6 days ahead.
- Compared to previous seasons, differences start to diverge at shorter lead times.

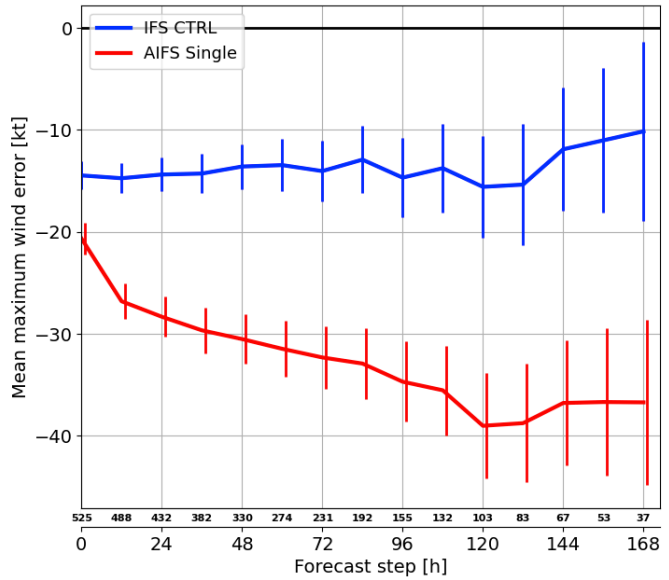
Data-driven forecasting models have more accurate and consistent tracks
Track improvements mainly result from improvements in steering flow
Smaller steering flow errors translate into reduced TC motion bias

AIFS Single vs IFS Control: TC intensity verification

Central pressure bias



Wind speed bias



Verification period: Jun - Nov 2025

Standard Verification
- Against IBTRACS
- Homogeneous sample
- TC present at initialisation

- TCs in AIFS show a weak-intensity bias in central pressure and wind, and this bias grows with forecast lead time
- This increase is partly due to TCs continuously expanding in size in AIFS, unlike in IFS

Data-driven forecasting models have larger intensity bias, and inherit limited physical consistency from ERA5

More info in the AIFS blog post !
DOI: 10.21957/d7d335141c

Research ongoing to improve the TC intensity:

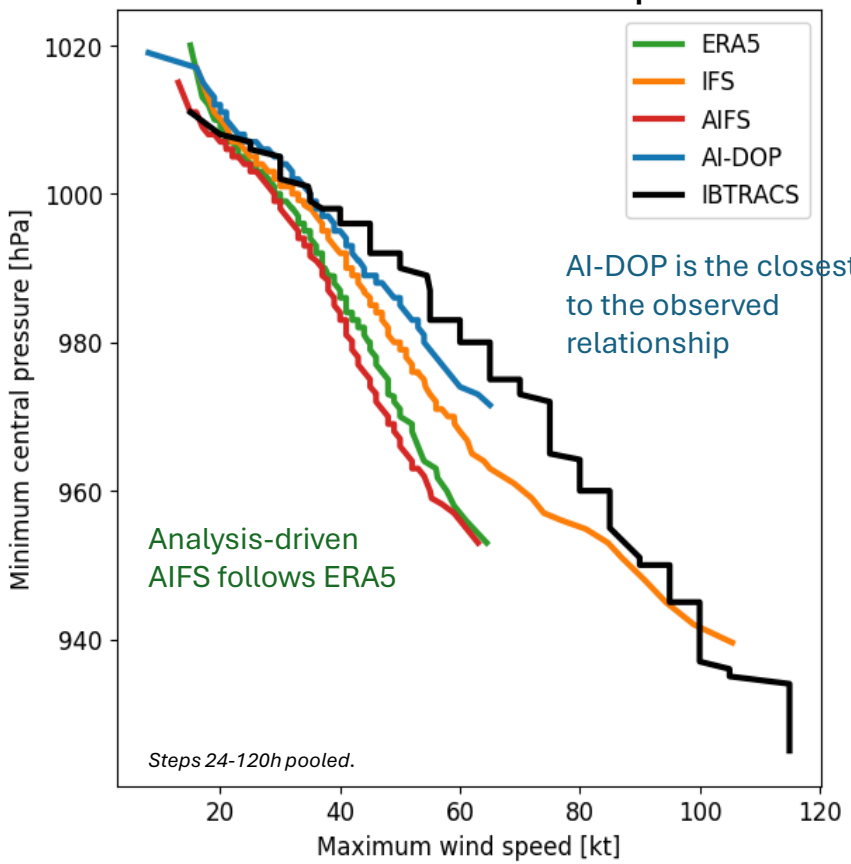
- Impact of training steps on TC forecast performance: training done with a single-step forecasts (6h) instead of the autoregressive forecasts - up to 72h (Rollout!) ➔ smaller bias than IFS from day 2 forecast

Preliminary results !

AI-DOP – a first look at TC verification

Verification period: 1 Aug – 10 Oct 2022
AI-DOP 36 km Resolution (N320)

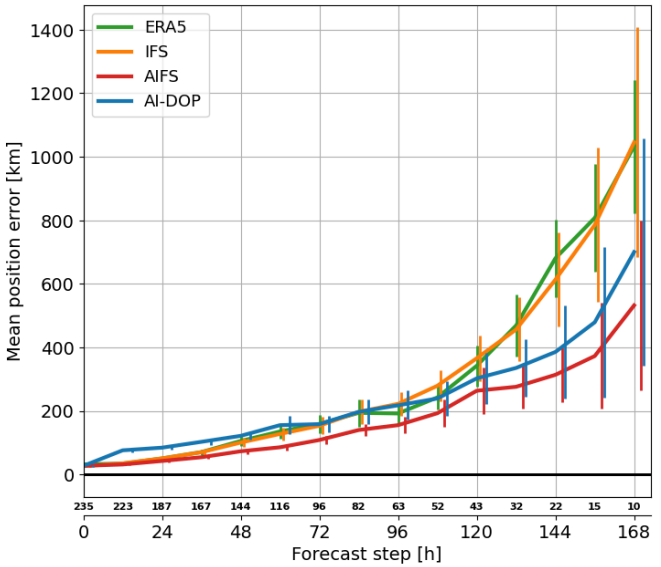
Pressure-wind relationship



AI-DOP is the closest to the observed relationship

Analysis-driven AIFS follows ERA5

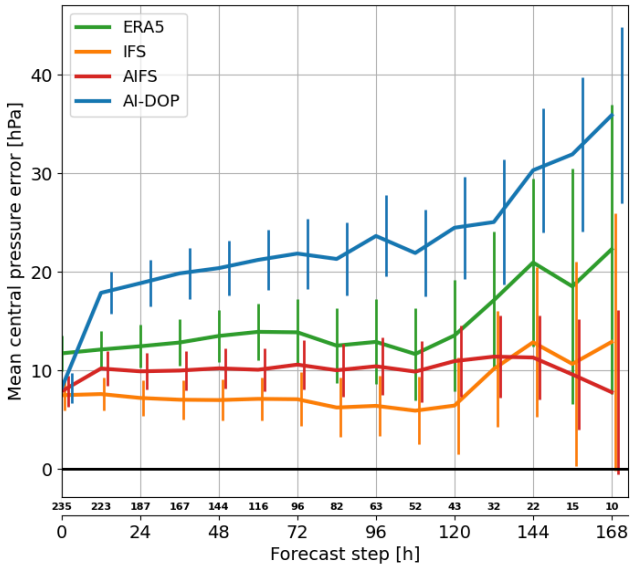
Position



Larger track errors in short range (resolution related?)

In medium-range AI-DOP beats IFS and is similar to AIFS

Central pressure



Currently bad for central pressure (too weak)

A new AI-DOP model version, at higher resolution and with additional observations (i.e. scatterometer backscatter, best track, etc) is in preparation

Conclusions

Substantial progresses on TC skills over the years but still ”room for improvements”

NWP outputs are essential for other ‘downstream hazard models’: flooding, storm surge, etc

Recent studies at ECMWF showed main improvements using higher resolution 4D-Var: we need to find a balance between computational costs and fc skills

We need high fidelity physical model: a better model would lead to a better forecast ➔ smaller observation departures ➔ less obs rejection

We need to better exploit observations close to the centre of the storm and use new observations:

- ✓ New/more observations ➔ BAROS & dropsondes ongoing activities
- ✓ How do we build a QC system that is reliable for all the situations, that can deal with large differences and trust the right one? Could we probably use AI for the QC?
- ✓ Should we take more risks relaxing the QC?
- ✓ Check what other NWP centres do

TCs and Global Data-Driven Forecasting models:

- ✓ AIFS has better TC tracks than IFS; not same results on the TC intensity but research is ongoing to improve it
- ✓ AI-DOP: preliminary results showed better wind-pressure relationship - a new AI-DOP version with more obs soon ready



Thank you!