Exploiting interface observations in a coupled reanalysis system

Phil Browne, Patricia de Rosnay, Tony McNally, Sebastien Massart, Noureddine Semane, Sean Healy, Katerina Anesiadou, Alan Geer, Tracy Scanlon, plus many more

ECMWF | EUROPEAN CENTRE FOR MEDIUM-RANGE WEATHER FORECASTS

ECMWF

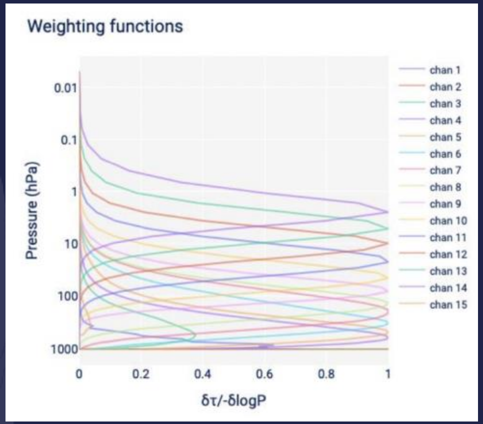
Outline

- Setting the scene
  - Interface observations sensitive to multiple Earth system components
  - Constraints of a Coupled Data Assimilation framework
- Assimilation of interface observations with an extended control vector
- Examples
  - 1: Sea surface temperature sensitive observations
  - 2: Altimeter range observations
  - 3: Sea ice sensitive observations
- Specific considerations for reanalysis

What are interface observations?

Interface observations are those which are **sensitive to**, or contain information about, **multiple** Earth system components. Typically, we mean the components which have historically been analysed separately

- T2m – screen level temperature
  - Is some interpolation between lowest atmospheric level (~10m) and the surface temperature
- Satellite window channels
  - Sensitive to Earth surface temperature and lower atmospheric column temperatures
- Altimeter delay measurements
  - Signal goes through atmospheric column (humidity) and reflects off surface
  - Tells us about atmospheric column and,
  - Distance of sea surface to satellite
  - Distance of sea ice surface to satellite – related to ice thickness/freeboard
  - Can give us spectral information on surface, i.e. wave heights and ice thickness distributions



The mathematics of the approach

A 4DVar cost function is famously (let's ignore Weak Constraint, Bias Correction, other important aspects to simplify the equations)

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{x}_b - \mathbf{x})^T B^{-1}(\mathbf{x}_b - \mathbf{x}) + \frac{1}{2}\sum_k (\mathbf{y} - \mathcal{H}_k(\mathcal{M}_k(\mathbf{x})))^T R_k^{-1}(\mathbf{y} - \mathcal{H}_k(\mathcal{M}_k(\mathbf{x})))$$

This leads to a gradient of the form

$$-\nabla J(\mathbf{x}) = B^{-1}(\mathbf{x}_b - \mathbf{x}) + \sum_k M_k^T H_k^T R_k^{-1}(\mathbf{y} - \mathcal{H}(\mathcal{M}(\mathbf{x})))$$

And we cycle our analysis with the model

$$\mathbf{x}_b = \mathcal{M}(\mathbf{x}_a)$$

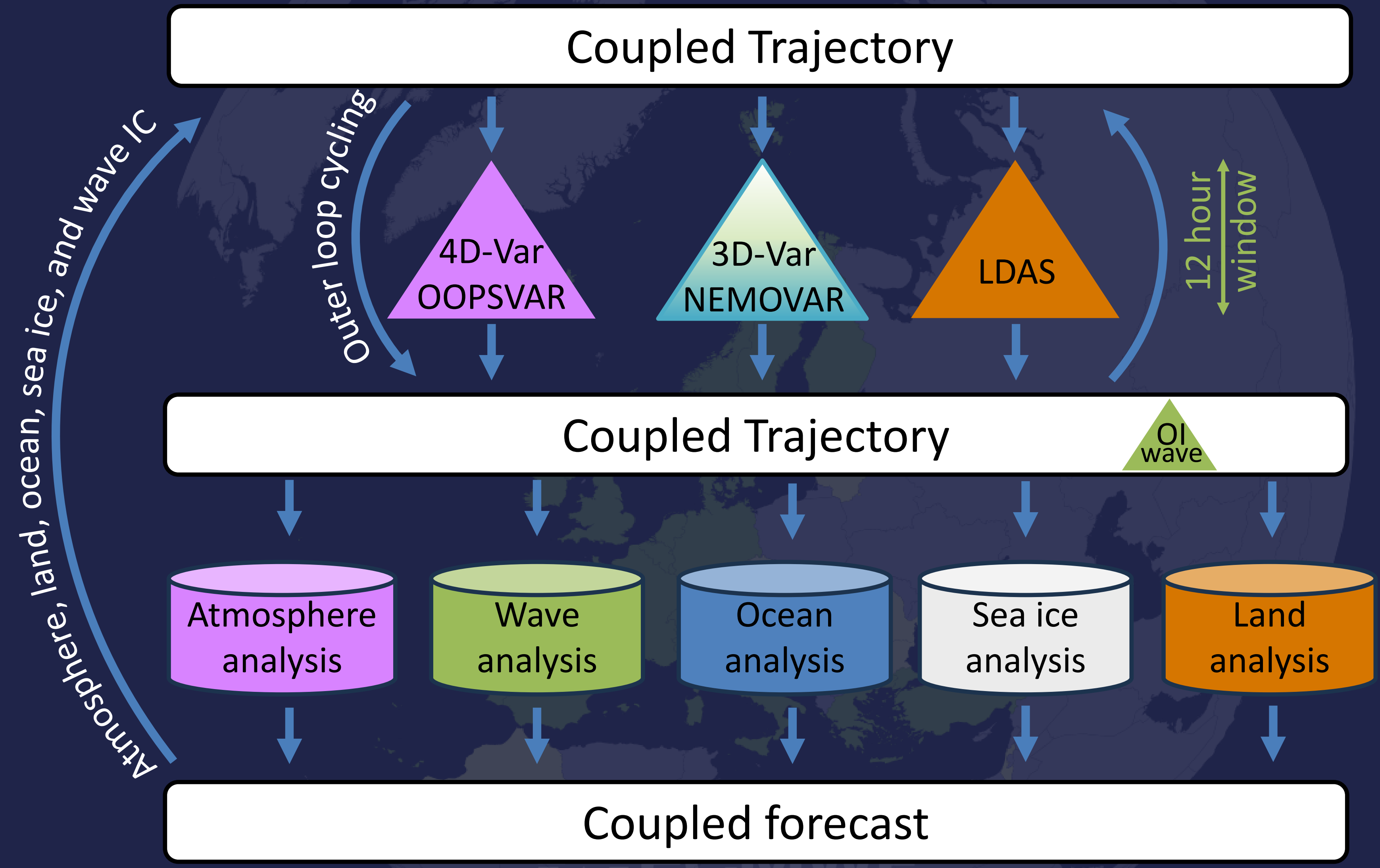
The mathematics of the approach

If the main operators in the system are uncoupled, or *block diagonal*, then the cost function in incremental form is *separable*, i.e.

$$J(\delta\mathbf{x}^m) = J\begin{pmatrix}\delta\mathbf{x}_1^m \\ \delta\mathbf{x}_2^m \\ \vdots \\ \delta\mathbf{x}_N^m\end{pmatrix} = \sum_i J_i(\delta\mathbf{x}_i^m)$$

This means we can bolt together different minimisations build up a coupled DA system within the framework of incremental 4DVar

Coupled DA setting at ECMWF within incremental 4DVar structure



The actual challenge

- Given an observation where $\mathcal{H}(x)$ is coupled, how do we use this effectively in a system built up of separable minimisations?

- If we are truly doing *coupled* DA, this is just DA with a longer state, so we get everything automatically

- Some approaches
    - CERA system – allow $\mathcal{M}(x)$ to do all the work

      Laloyaux, Patrick, et al. "CERA-20C: A coupled reanalysis of the twentieth century." *Journal of Advances in Modeling Earth Systems* 10.5 (2018): 1172-1195.

    - Interface solver – extend and overlap control vectors of the components

      Frolov, Sergey, et al. "Facilitating strongly coupled ocean–atmosphere data assimilation with an interface solver." *Monthly Weather Review* 144.1 (2016): 3-20.

- Our approach – use the eXtended Control Vector (XCV)
    - Pass data between components via through intermediate observation space departures
    - Combine this with the coupled $\mathcal{M}(x)$ of outer loop coupling

Extended Control Vectors (XCV)

We have different implementations of XCVs at ECMWF

- Observation space sink variables

  IFS Documentation CY28R1 - Part II: Data Assimilation, 2004 https://doi.org/10.21957/gpybdcbj

- 2D correlated fields

  Massart, S., Bormann, N., Bonavita, M., and Lupu, C.: Multi-sensor analyses of the skin temperature for the assimilation of satellite radiances in the European Centre for Medium-Range Weather Forecasts (ECMWF) Integrated Forecasting System (IFS, cycle 47R1), Geosci. Model Dev., 14, 5467–5485, https://doi.org/10.5194/gmd-14-5467-2021, 2021.

We can use these to compute observation space departures which transfer information between components

How does this work in practice?

We use an XCV for which we have

$$\mathbf{x}_{\mathrm{cv}} \in \mathbf{x}_1$$

And

$$\mathbf{x}_{\mathrm{cv}} = \mathcal{H}_{\mathrm{cv}}(\mathbf{x}_2)$$

Then first compute $\delta \mathbf{x}_1^0|_{\mathrm{xcv}}$

Then we say $(y - \mathcal{H}_{\mathrm{cv}}(\mathbf{x}_2^1)) = \delta \mathbf{x}_1^0|_{\mathrm{xcv}}$

Then compute $\delta \mathbf{x}_1^1$ and $\delta \mathbf{x}_2^1$ and repeat for the following outer loops.

So you can see this is a sequential methodology – the expected approach with outer loop coupling and separate minimization.

As a cartoon…

$$y - \mathcal{H}\left(\mathbf{x}^0\right) = y - \mathcal{H}\begin{pmatrix}\mathbf{x}_1^0 \\ \mathbf{x}_2^0\end{pmatrix}$$

$y - \mathcal{H}\left(\mathbf{x}^0\right)|_1$

$y - \mathcal{H}\left(\mathbf{x}^0\right)|_2$

OOPSVAR

NEMOVAR

$\delta\mathbf{x}_1^0$

$\delta\mathbf{x}_2^0$

$$y - \mathcal{H}\left(\mathbf{x}^1\right) = y - \mathcal{H}\begin{pmatrix}\mathbf{x}_1^0 + \delta\mathbf{x}_1^0 \\ \mathbf{x}_2^0 + \delta\mathbf{x}_2^0\end{pmatrix}$$

$y - \mathcal{H}\left(\mathbf{x}^1\right)|_1$

$$\begin{bmatrix} y - \mathcal{H}\left(\mathbf{x}^1\right)|_2 \\ \delta\mathbf{x}_1^0|_{\mathrm{xcv}} \end{bmatrix}$$

OOPSVAR

NEMOVAR

$\delta\mathbf{x}_1^1$

$\delta\mathbf{x}_2^1$

$$y - \mathcal{H}\left(\mathbf{x}^2\right) = y - \mathcal{H}\begin{pmatrix}\mathbf{x}_1^1 + \delta\mathbf{x}_1^1 \\ \mathbf{x}_2^1 + \delta\mathbf{x}_2^1\end{pmatrix}$$

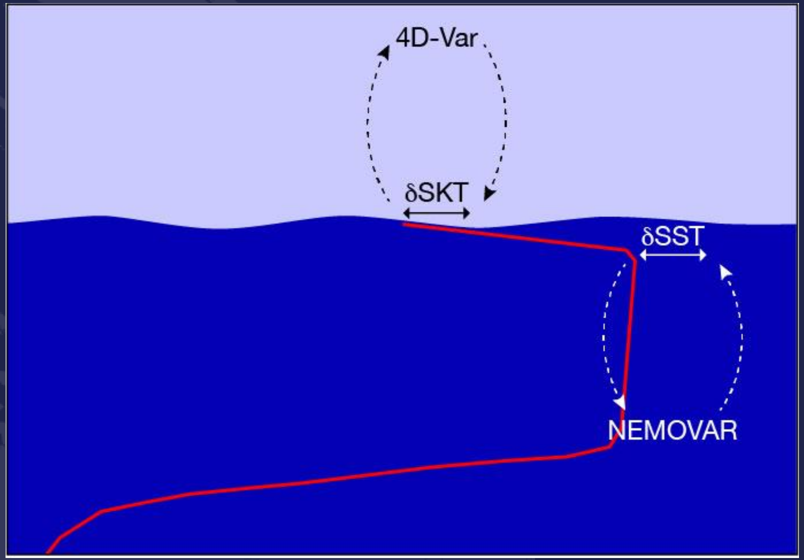
Examples of XCVs

Example 1: RADSST assimilation

x_{cv} = SKIN temperature

$\mathcal{H}_{cv}(x_2)$ = Ocean temp top level + Δ cold skin

$H_{cv}(x_2)$ = $H_{sst}(x_2) + \varepsilon$ where ε < 10E-8



Example 2: Coupled altimeter assimilation

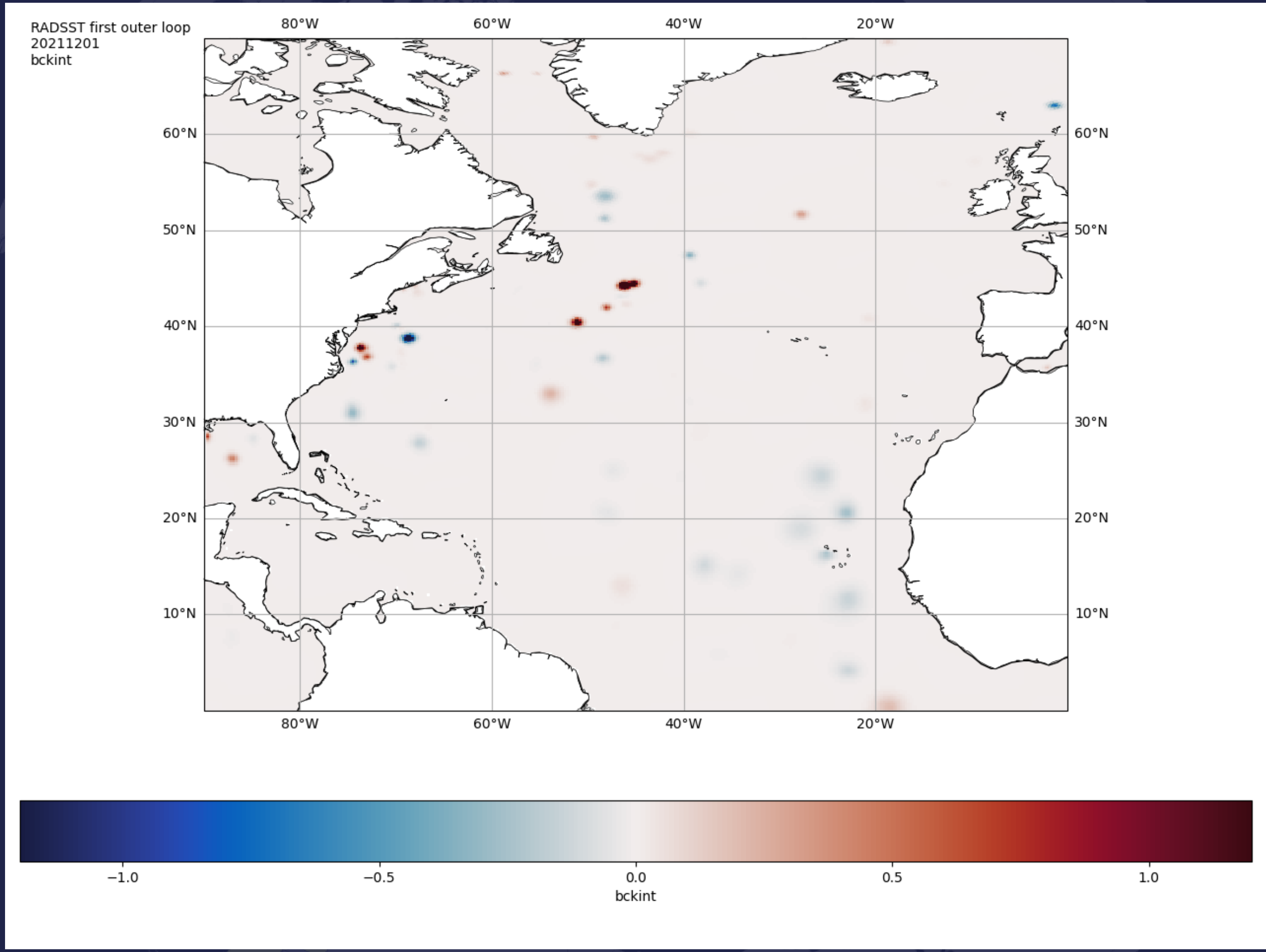
x_{cv} = Sea surface height (SSH)

$\mathcal{H}_{cv}(x_2)$ = Ocean model sea surface height above something + Δ f(MDT, geoid, fresh water, …)

$H_{cv}(x_2)$ = $H_{sla}(x_2) + 0$

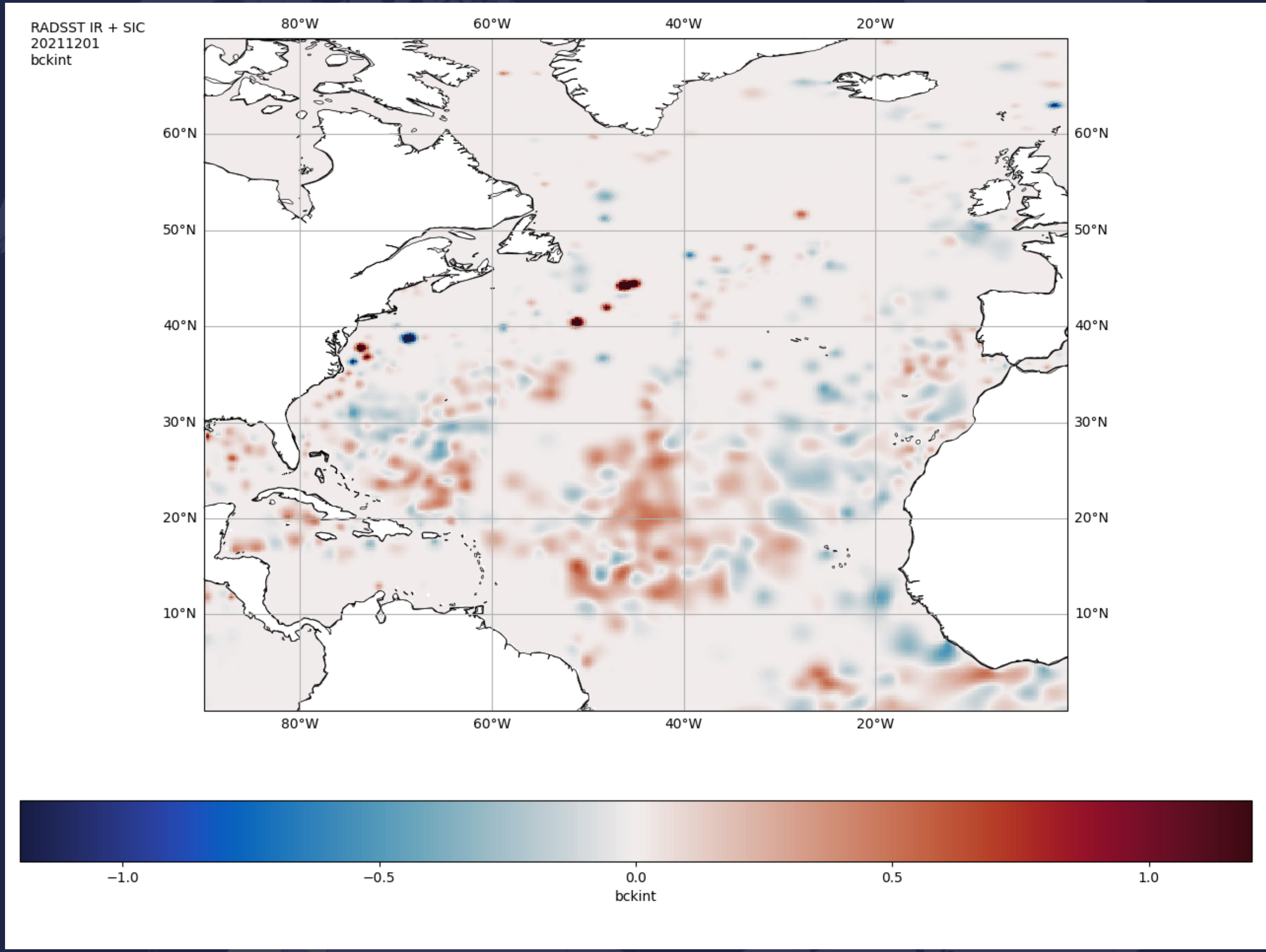
We have all the operators in place in the IFS and in NEMOVAR to do this – we just need to join them together.

Coupled SST increments from different observation sources



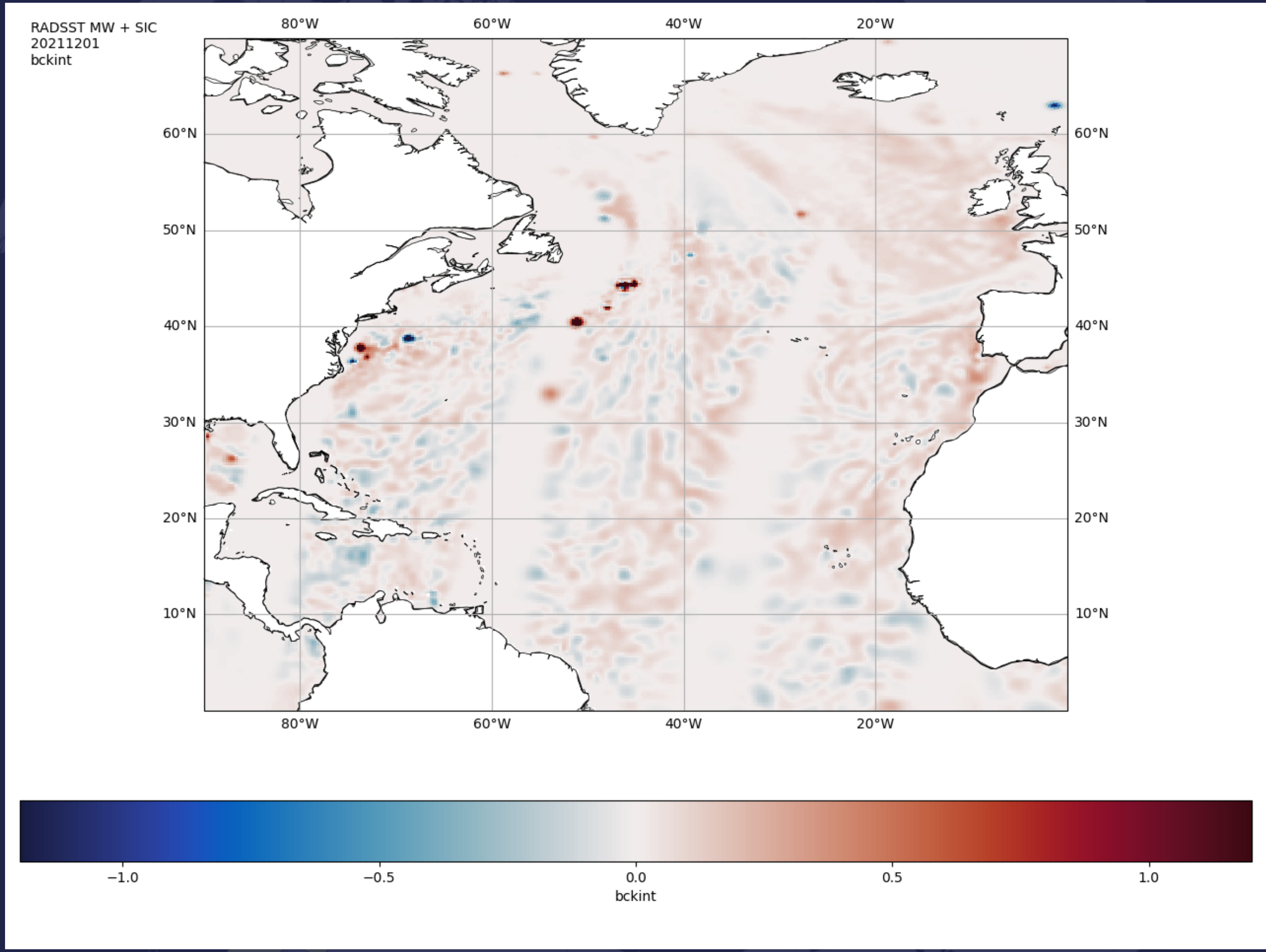
In situ + SLA

Coupled SST increments from different observation sources



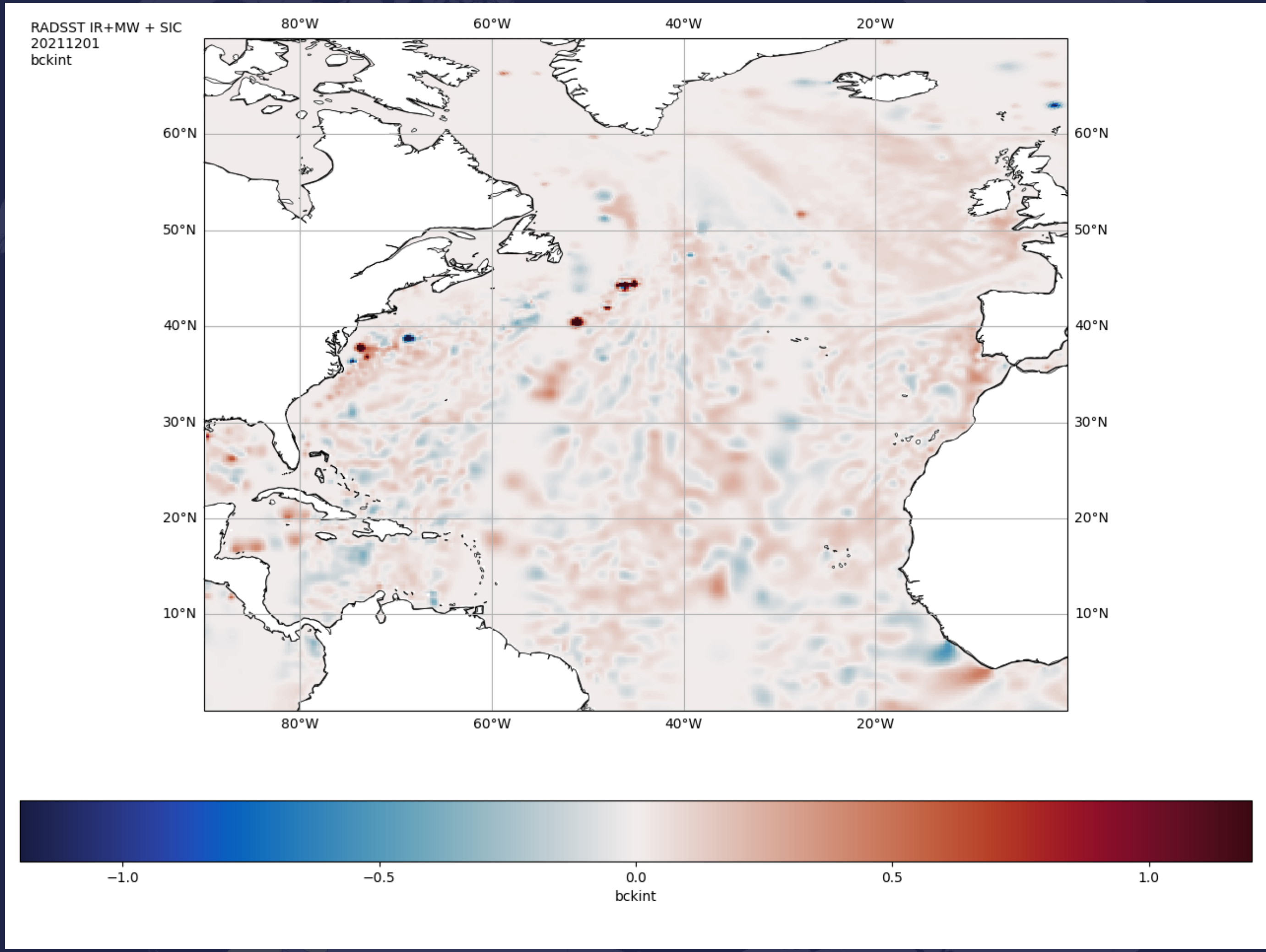
In situ + SLA + IR

Coupled SST increments from different observation sources



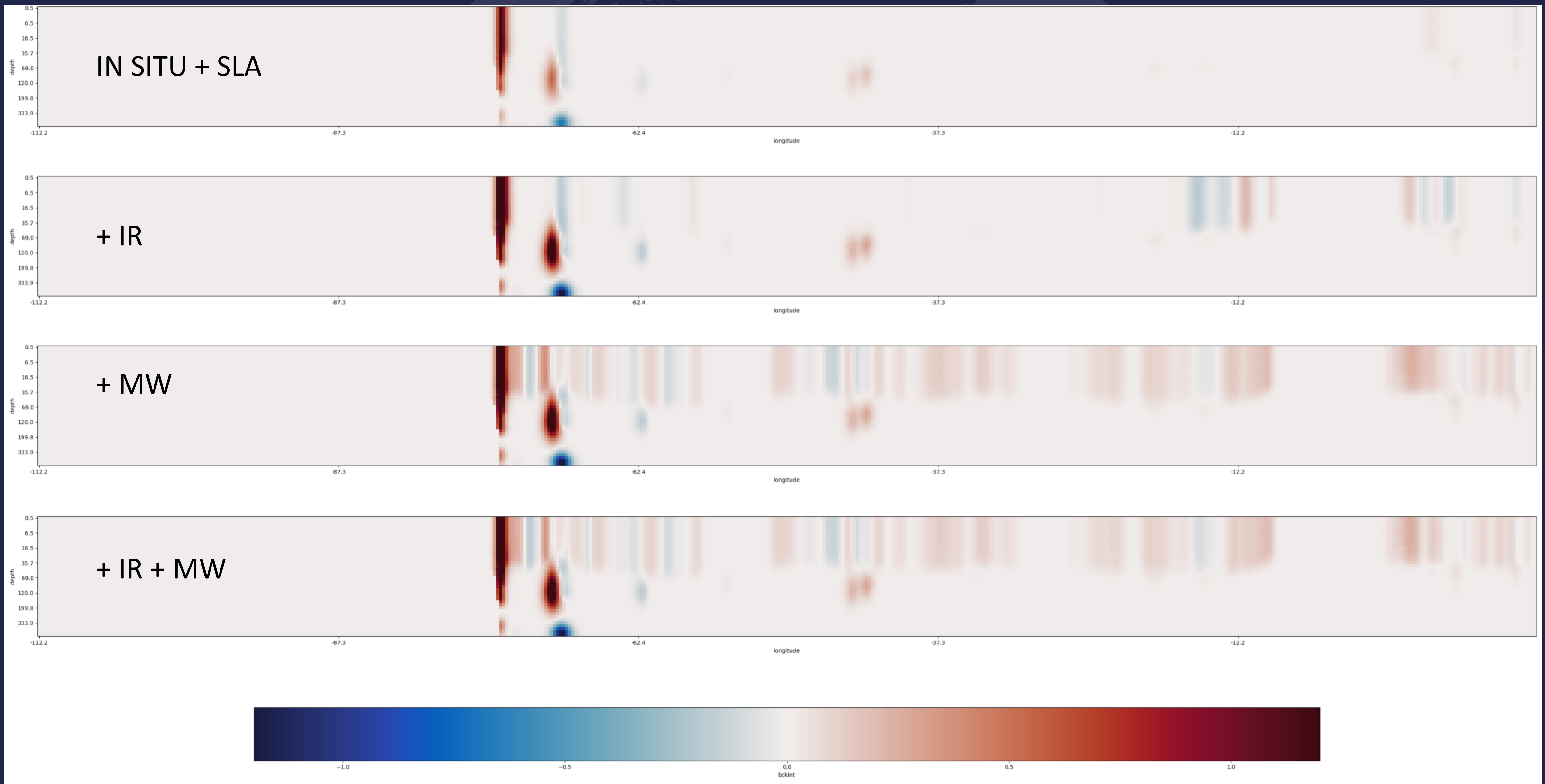
In situ + SLA + MW

Coupled SST increments from different observation sources

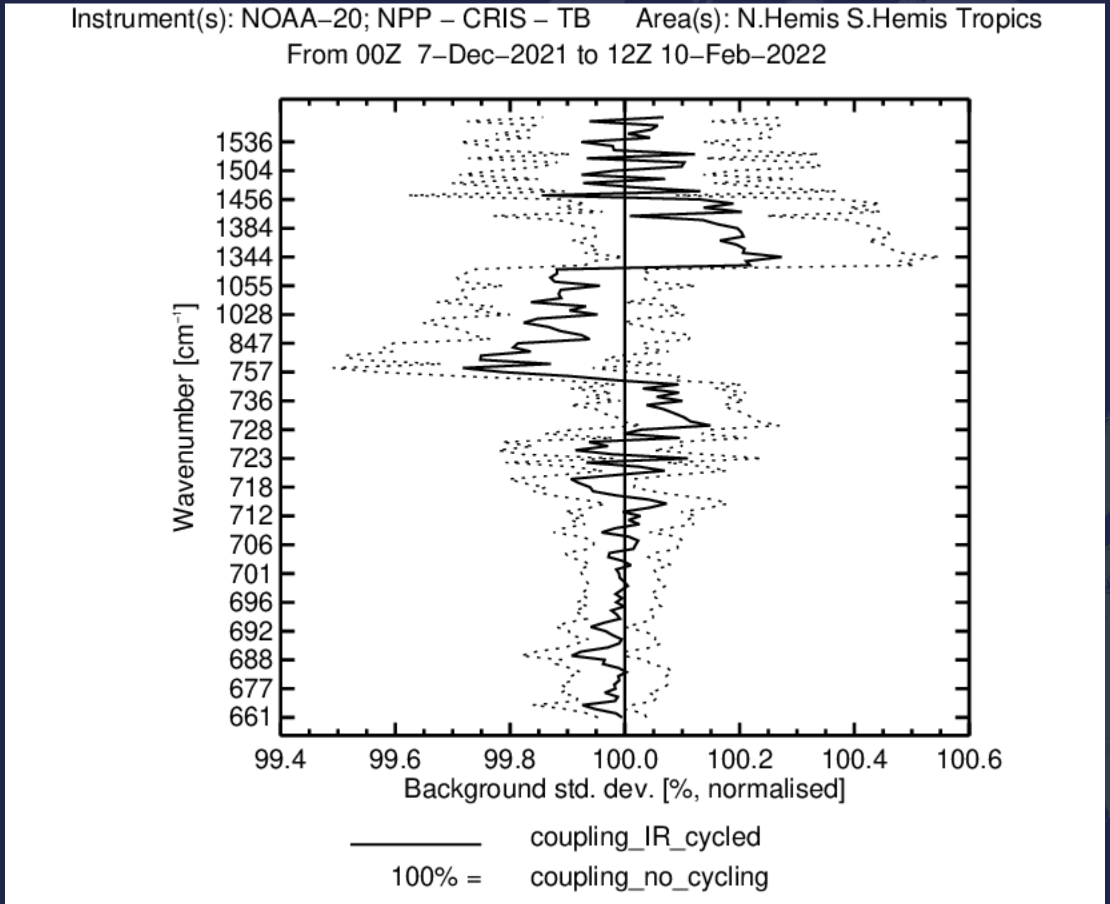


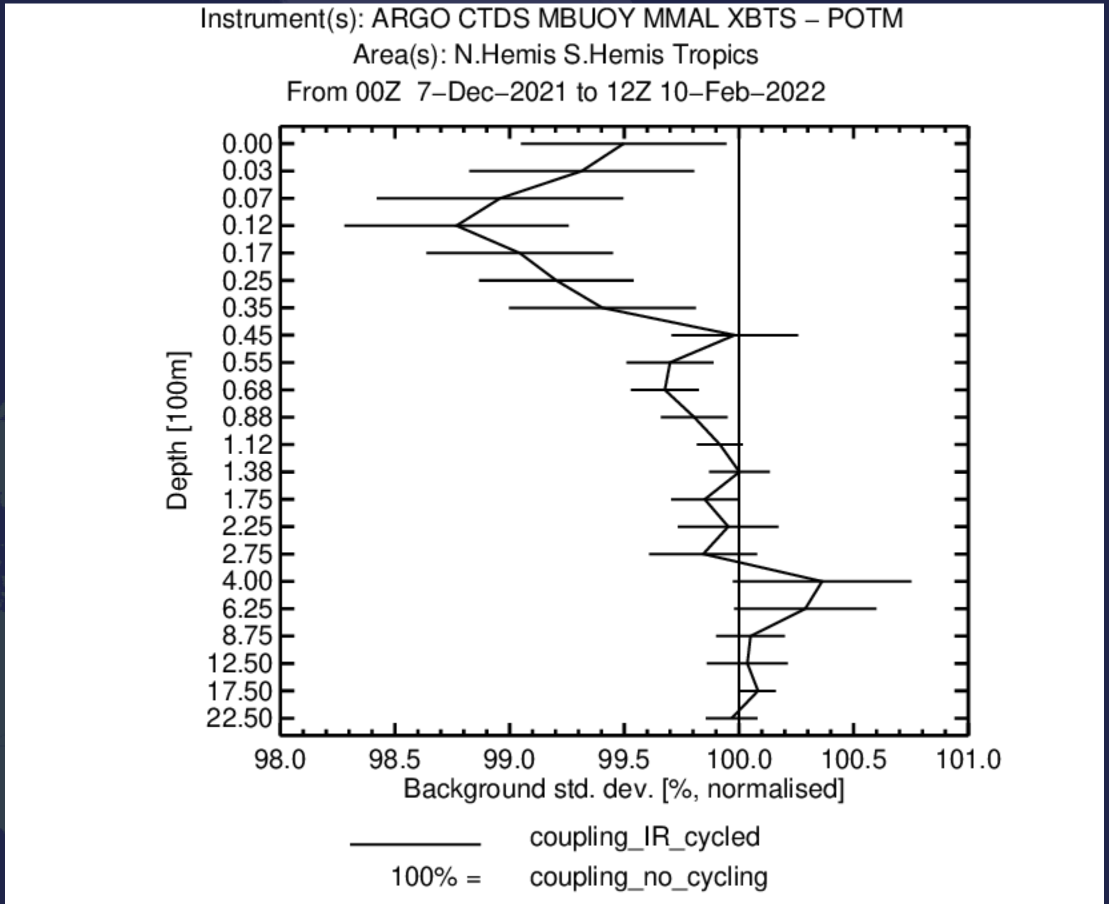
In situ + SLA +
IR + MW

Coupled 3D increments in the ocean from radiances



Verification against observations





Altimeter range observations

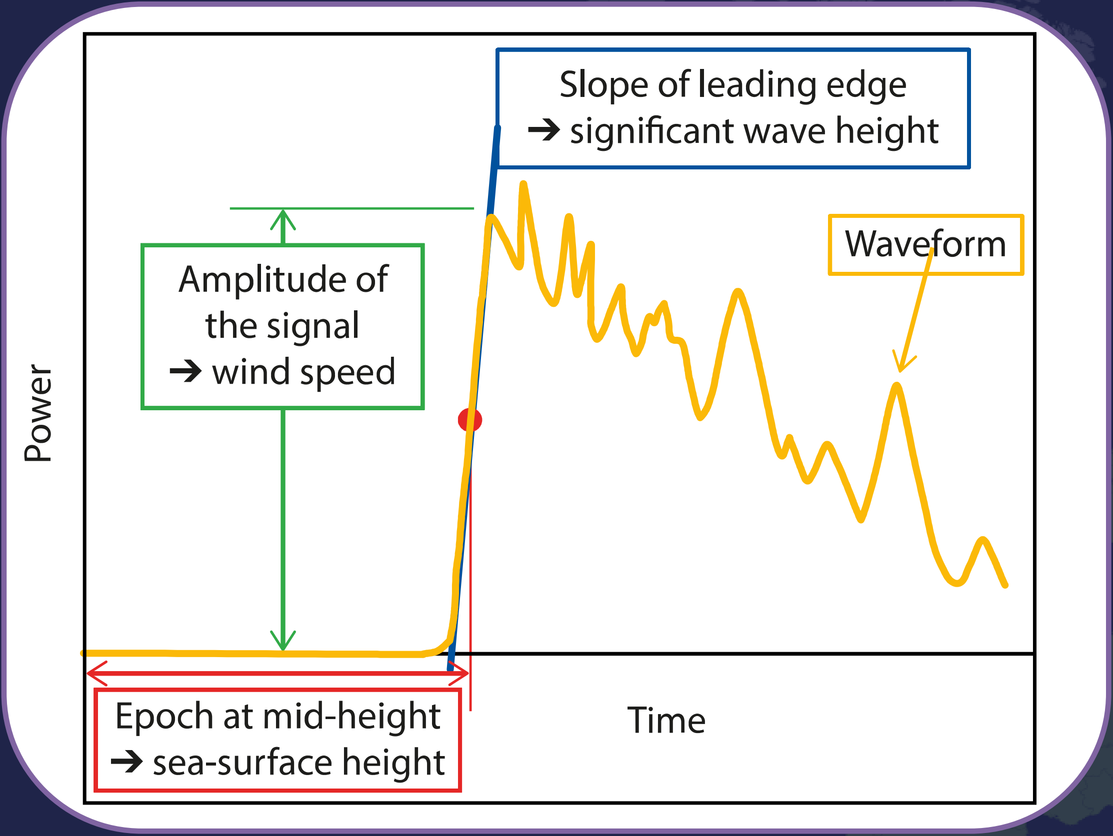


Figure from Saleh Abdalla
https://doi.org/10.21957/m2cx1w

The measurement here of "sea-surface height" is a time delay related to sea-surface height, not a direct measurement

It is also sensitive to the humidity in the column between the surface and the satellite

So it actually contains information on both total column water vapour and sea-surface height

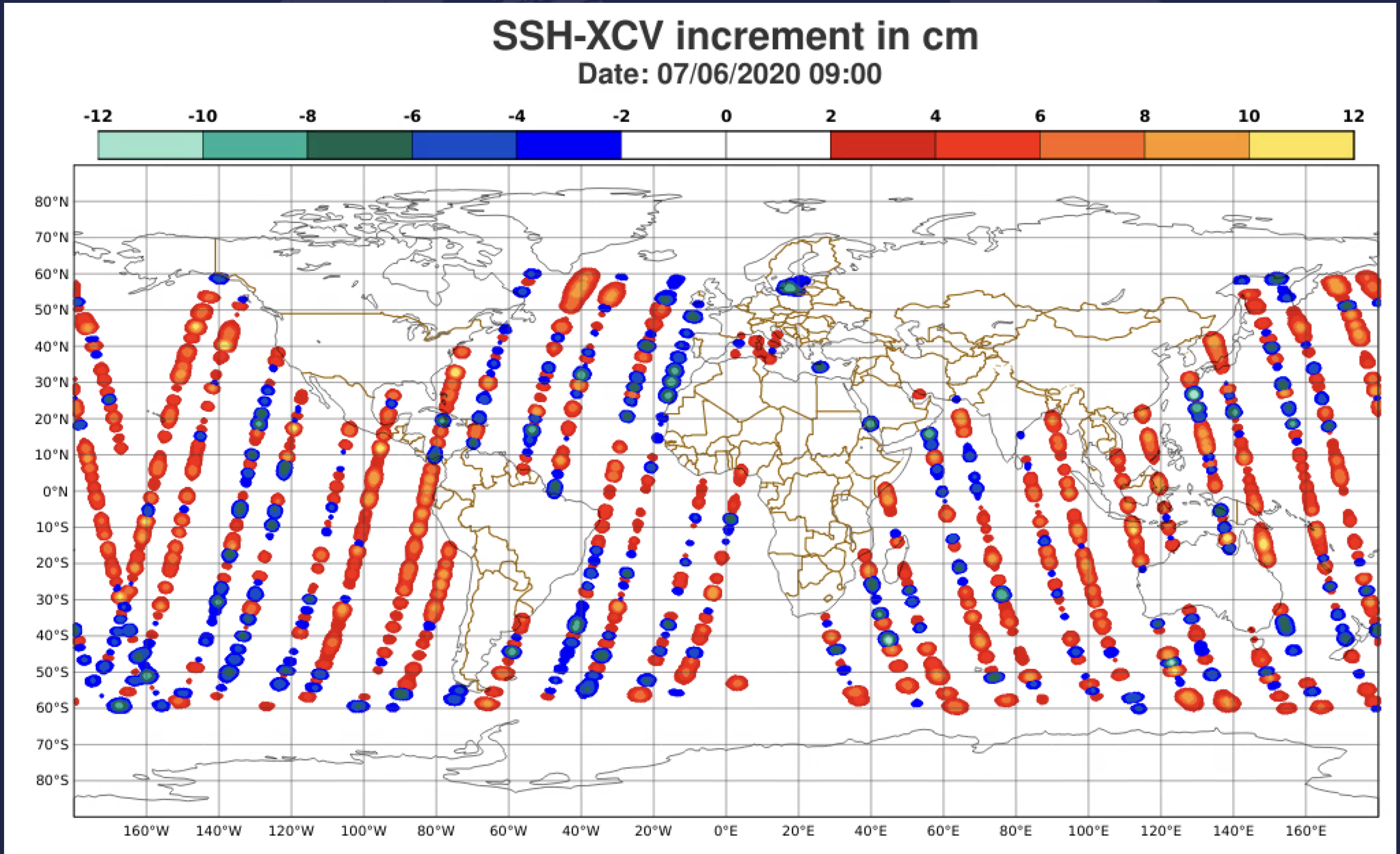
Typically, a microwave radiometer is flown to remove the atmospheric humidity effects

We want to extract both pieces of information with a coupled DA system

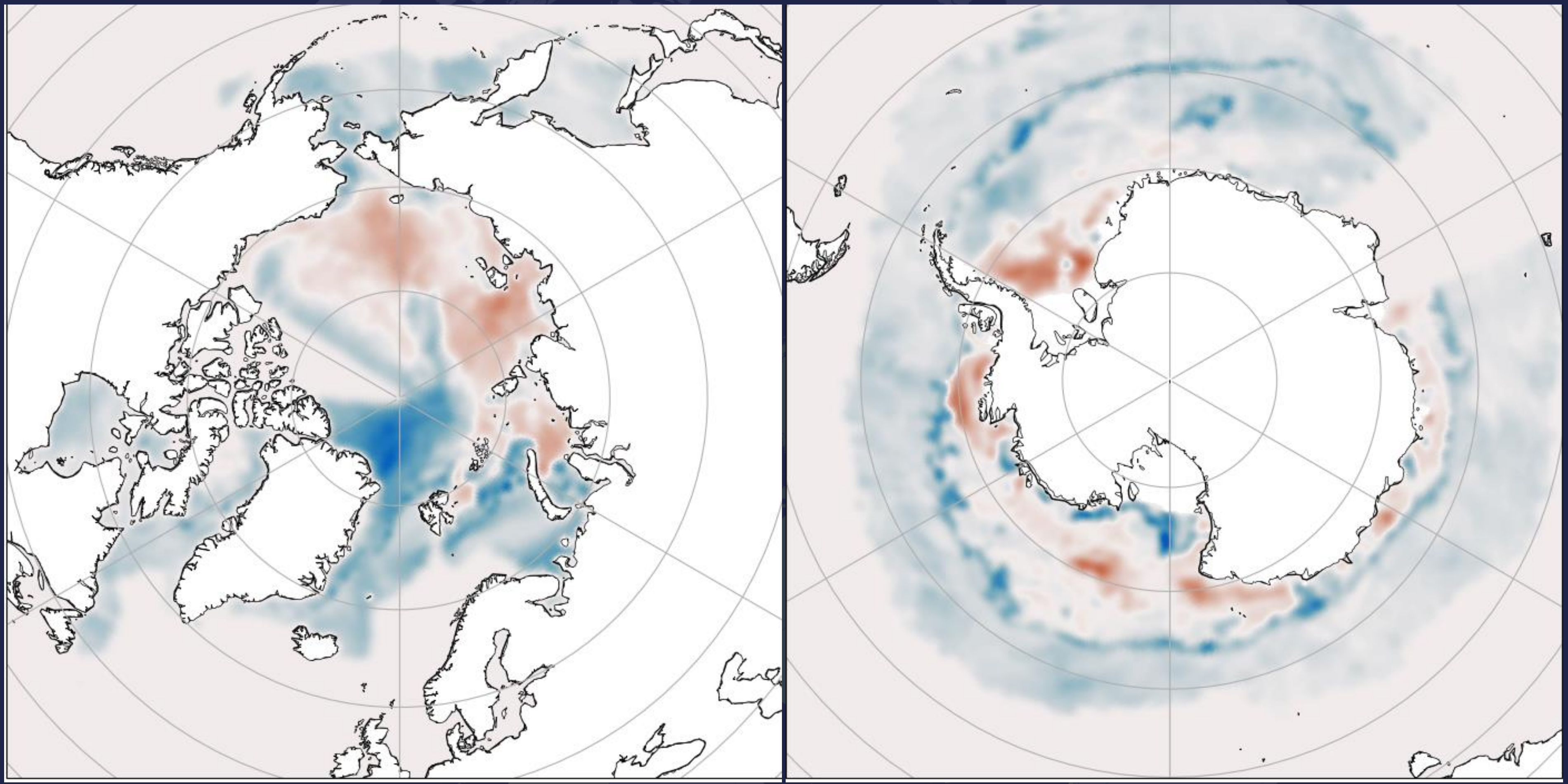
Altimeter range observations

- Want to assimilate Sea Surface to Satellite Delay (S3D)
- S3D is equivalent to the atmospheric delay of a signal starting from the ocean surface
  - This is the same as a ground based GPS measurement: we have an observation operator already!
- We extend the atmospheric control vector to include a Sea Surface Height XCV
- 4DVar then gives us an increment to atmospheric column and SSH

Coupled assimilation of altimeters – step 1 – XCV increments



Coupled Sea Ice Concentration increments from AMSR2 radiances



Reanalysis considerations

- For reanalysis we need to be aware of the changing observation network
- We want to run in periods where none of these interface observations are available
  - We need to be using "anchoring" conventional observations at the same time – see Bill Bell's/Nick Raynor's talk
- Performance will be defined by the quality of the B matrix
  - It must be flow dependent on both:
    - observation network changing timescales – see Laura Slivinski's talk
    - errors of the day – interface observations can capture very fast processes – see Hao Zuo's/François Counillon's talk
- Building the infrastructure is the "easy" part – making it perform well is a huge scientific effort