Developments in Data Assimilation and use of Observations at the Met Office

David Simonin on behalf of the NG-PAO's Team

ECMWF
Workshop on data assimilation: initial conditions and beyond

Content:

- Introduction
- Observation processing: JOPA
- Data assimilation
  - JADA-Global
  - JADA-Regional
- Summary

Introduction

In late 2010's the Met Office launched the **Next Generation Modelling Systems Programme** (NGMS) to reformulate and redesign the Met Office's existing systems to deliver, a unified Earth environment prediction framework: **Momentum™**.

Momentum will allow the Met Office and its partners to fully exploit future generations of supercomputer for the benefits of society.

GungHo Atmosphere Science Project (GHASP)	LFRic development	NG Marine systems
NG coupling project	**NG observation processing & data assimilation**	Visualisation & Analysis Tools
Post processing & verification	Research workflows	R2O

MOSAC 2019 & 2024

Introduction

Scope and deliverables

NEXT GENERATION JOPA PROCESSING AND ASSIMILATION OF OBSERVATINON

JEDI-based Observation Processing Application

Replicate our current Observation Processing science for atmospheric and marine data assimilation.

This is like for like replacement → Expecting a neutral impact

NEXT GENERATION JADA PROCESSING AND ASSIMILATION OF OBSERVATINON

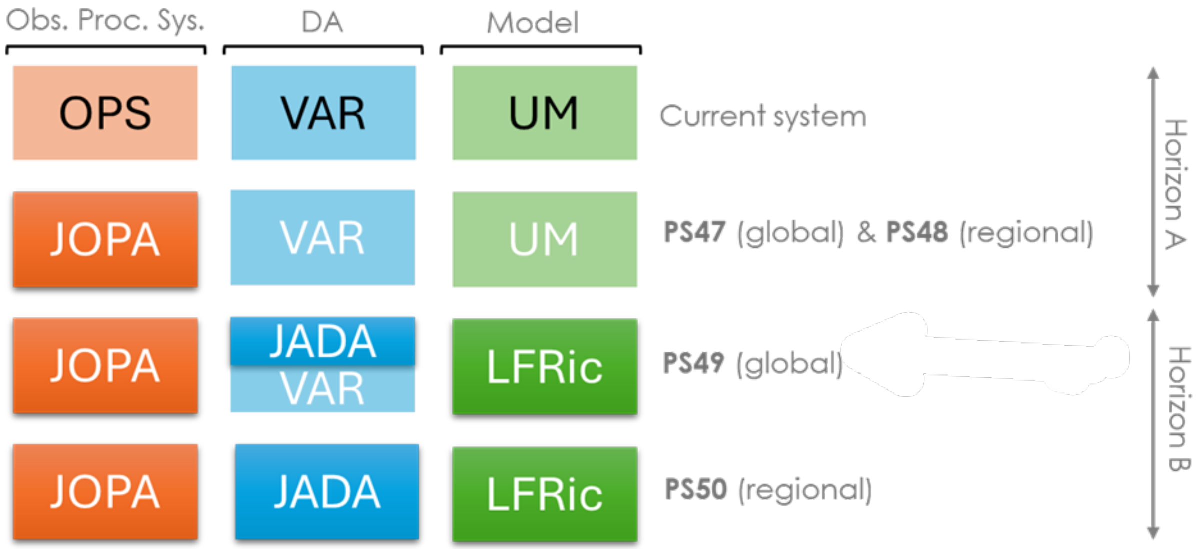
JEDI-based Application for Data Assimilation

Strategy: "putting ensembles at the heart of everything we do" for Global and Regional NWP

New science: Only use well-proven science – operational or published

Start of parallel Suite

- *PS47: now*
- *PS48: in ~1 year*
- *PS49: in ~Sept 2027*
- *PS49: in ~Sept 2028*



Developments in Data Assimilation and use of Observations at the Met Office

Introduction

Capitalize on investments

Community engagement and collective resilience

- Enhance collaborations to fast-track academic research into operations (e.g. UoR joint-positions)
- Train future DA workforce in JEDI (UoR annual DA course)

Versatile and portable:

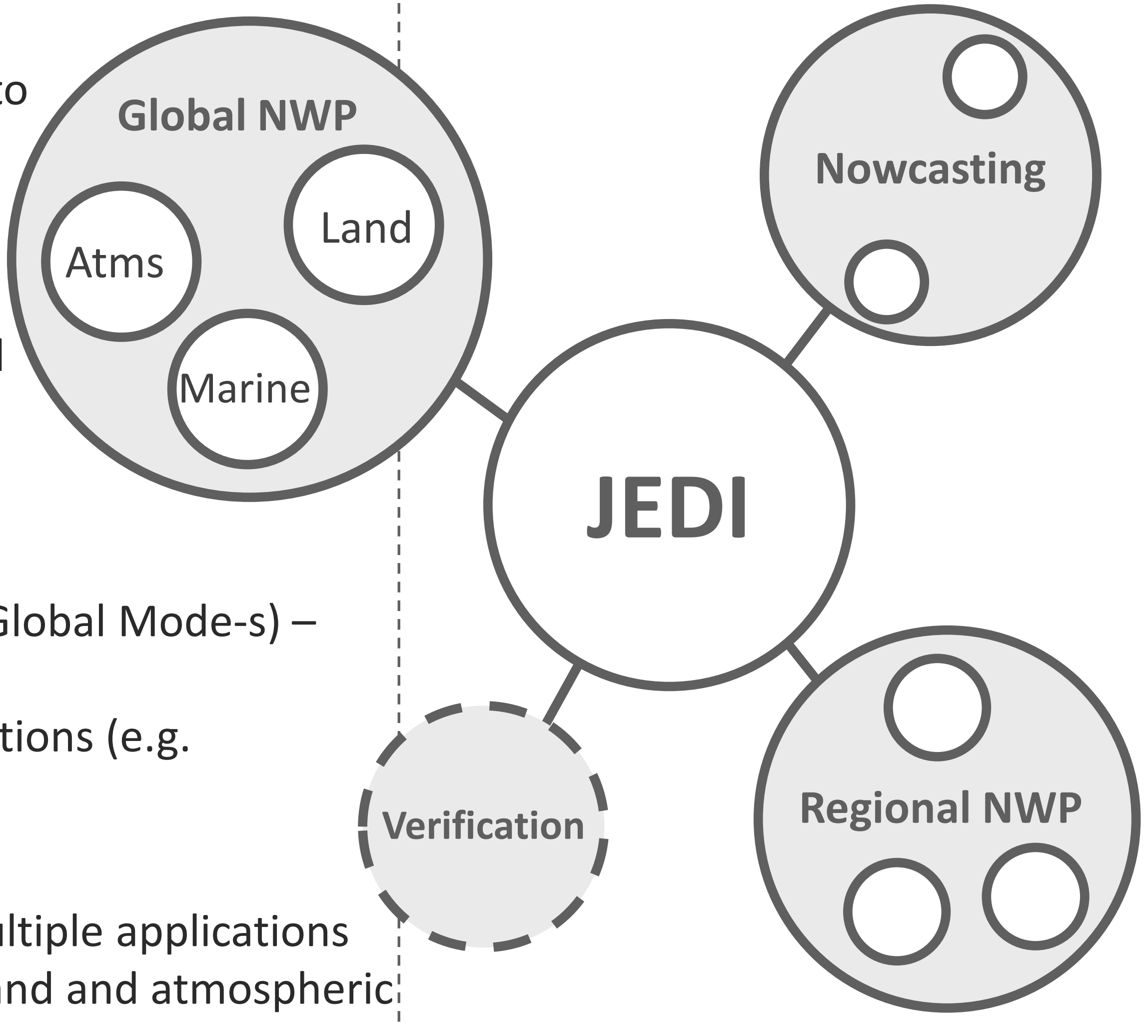
- Wide range of platforms: Laptop, Linux cluster, HPC, Cloud
- Develop and test new science with toy models
- Same code, different models/domains

Continuous scientific innovation

- Faster introduction of observations into operations (e.g. Global Mode-s) – even without code changes
- Rapid pool through of scientific advancements into operations (e.g. multiscale, nonlinear, continuous DA)

“Rosetta Stone” to coupled DA

- Remove technical barrier to coupled DA – one tool for multiple applications
- Improve research synergy on coupled DA from – ocean, land and atmospheric groups



JOPA: JEDI-based Observation Processing Application

Martins, A. Maycock, B. Candy, C. Harlow, C. Mao, C. Thomas, D. Davies, D. Rundle, F. Carminati, H. Lawrence, J. Cotton, J. Hocking, J. Waller, L. Hawkness-Smith, M. Wlasak, M. Forsythe, M. Shin, M. Thurlow, M. Cooke, N. Bowler, O. Lomax, O. Lewis, P. Levens, S. Good, S. Migliorini, S. Sandbach, S. Newman, T. Searle, W. Tennant, Y. Pradhan, R. Nixon-Hill, T. Tobin, K. Franklin, M Dell, R. McLaughlin

In collaboration with the Joint Center for Satellite Data Assimilation

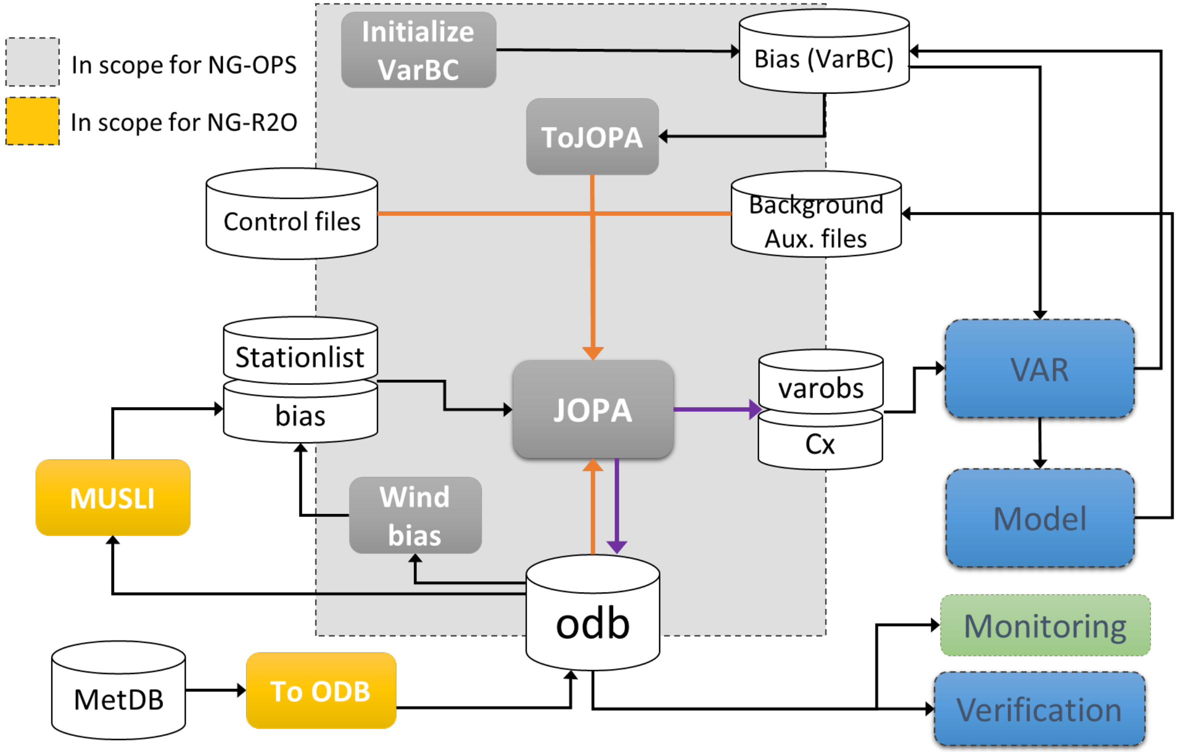
Scope – Global & Regional

Replicate our current Observation Processing science

Requirements

- Provide a system to allow the Met Office and its partners to fully exploit the opportunity provided by the future generations of supercomputer.
- Develop a versatile system that improve:
  - turnaround of new observation,
  - pool through to operation.
- Provide a system suitable for both
  - global and regional domains,
  - current and future DA system and model
    - UM with VAR | LFRic with JADA

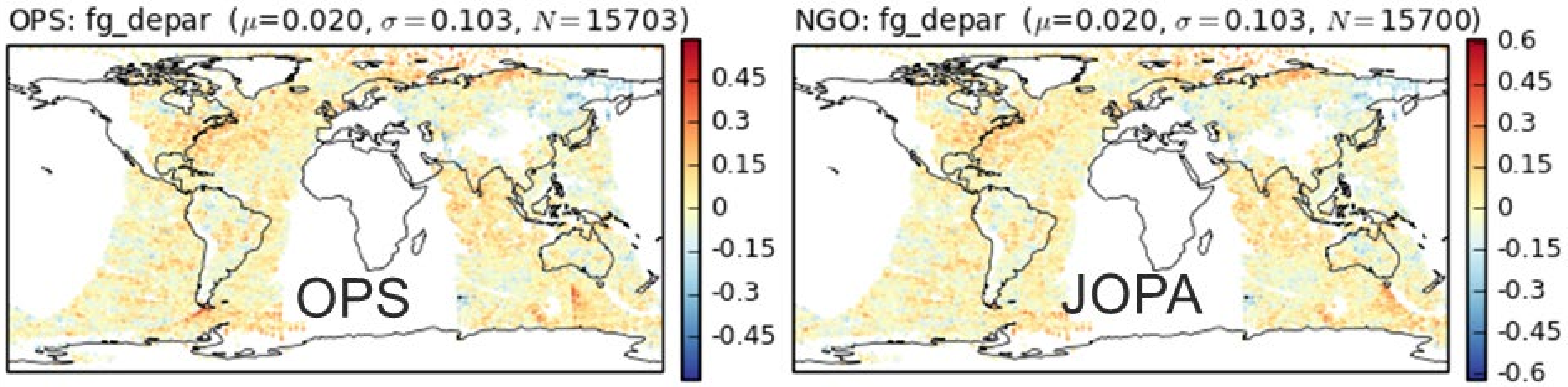
JOPA: Jedi-based Observation Processing Application



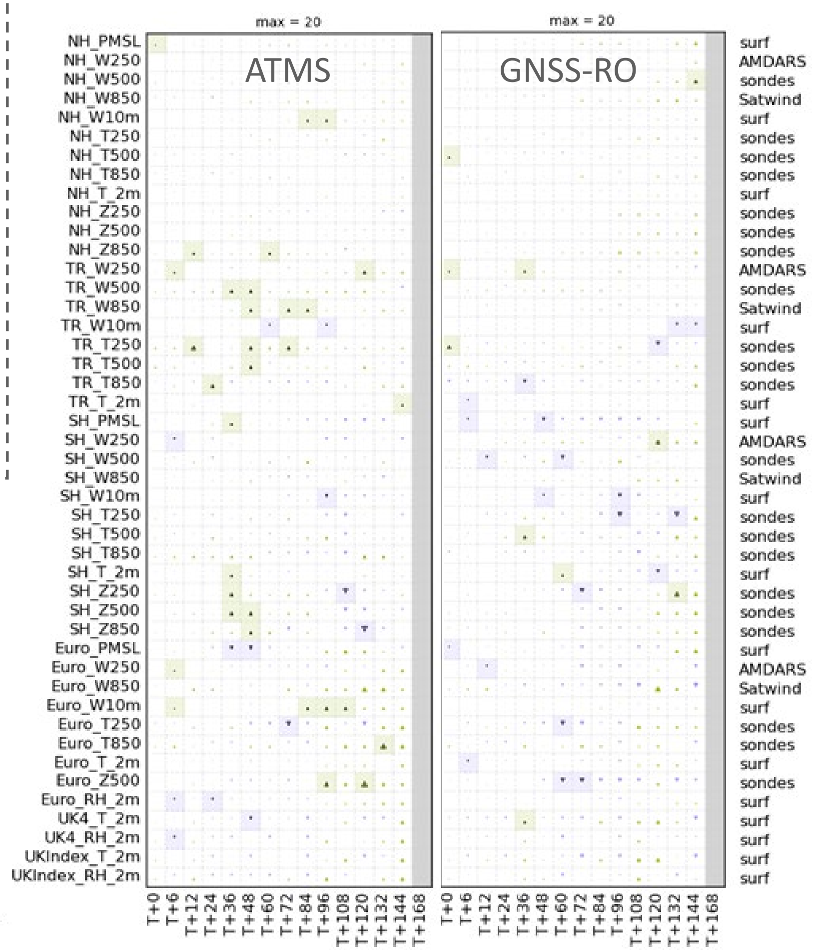
Outcomes

Porting and trialling

- **Looking for a neural impact**
- **All global and >90% of regional observations** (atoms and ocean) have been ported.
  - Great difficulty with reproducing thinning procedure
  - No more Latent Heat Nudging for the regional (direct assimilation instead).
  - Porting of observation for Land-DA has started
- New capability:
  - “Bouquet of GeoVals” will enable a full representation of the observation geometry (foot-print, radar gate, etc) in H(x).



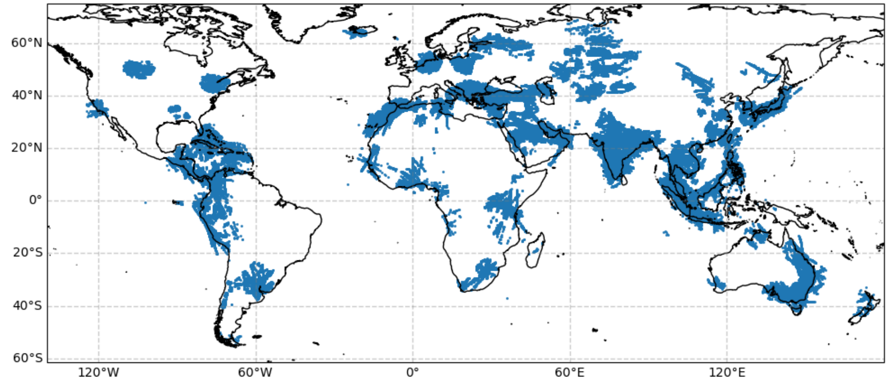
Brightness temperature departure ATMS Ch8 - 20210313T0600Z

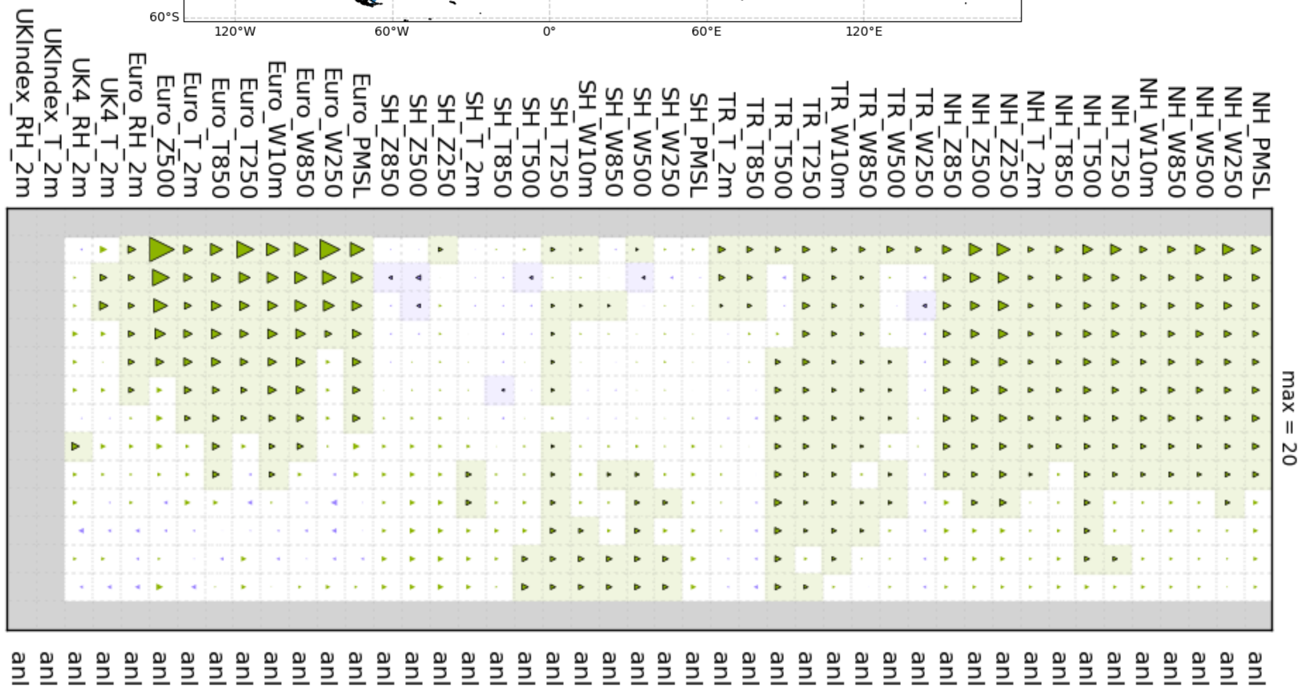


% Difference (JOPA vs Control) - RMSE against observation 20210608 to 20210920

Outcomes

Typical global mode-s coverage (without EMADDC observation)





Path to Operation

- A new ecosystem suitable for re-analysis has been developed
- We already see the benefit of the simplicity and efficiency of the new system:
    - while introducing new observation (global mode-s),
    - while investigating forward operator (Radar).
- **PS47** (now): JOPA-Global
- **PS48** (new year): JOPA-Regional

% Difference against control – overall 0.72%
RMSE against own analysis
01/03/2024 to 31/05/2024
(Elliott Warren).

JADA: JEDI-based Application for Data Assimilation
Global NWP

M. Wlasak, R. Rawlins, T. Payne, S. Sandbach, T. Vockerodt, J. Colcough,T. Hill, L. Milazzo, S. Pullen, S. Migliorini, C. Thomas, J. Waller, G. Inverarity, K. Huxtable, T. Leung, M. Destouches, N. Bowler, F. Carminati, T. Hutchins, O. Aljabar, M. Cooke, O. Lomax, J. Cotton, O. Lewis, M. Shin.

In collaboration with the Joint Center for Satellite Data Assimilation

Global-Scope

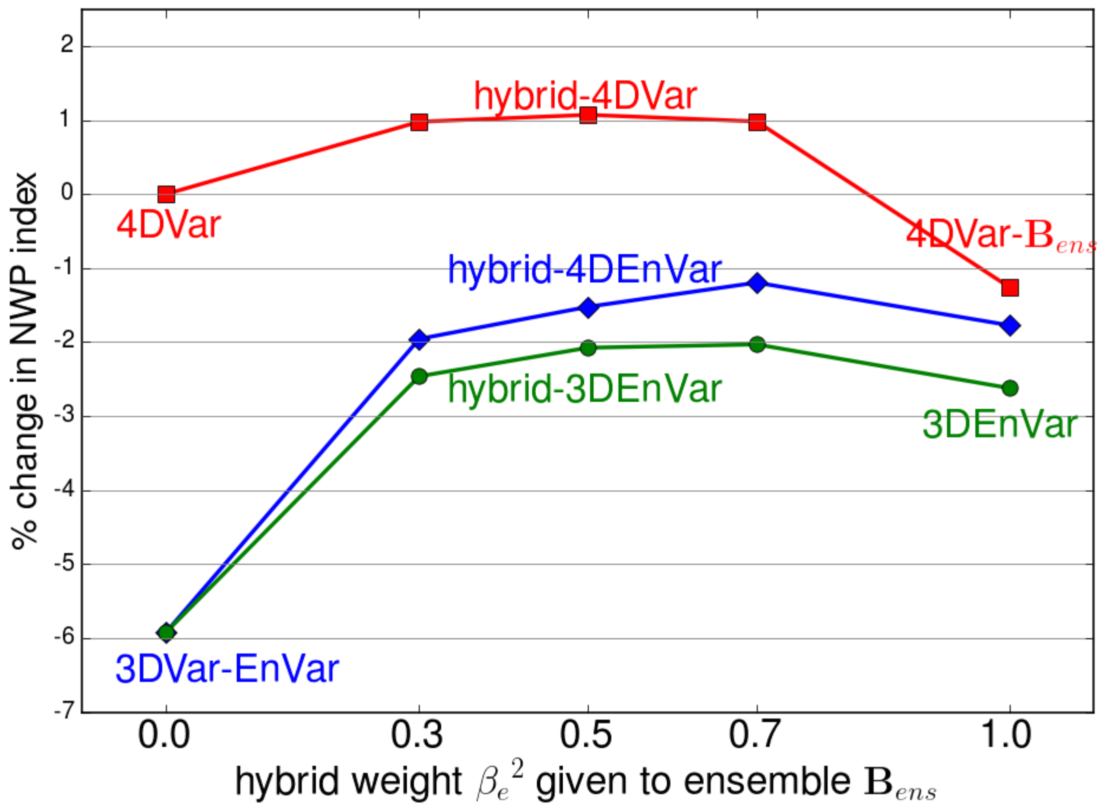
"putting ensembles at the heart of everything we do"

Requirements

- A fully ensemble-based NWP system
- Focus only on well-proven science (operational or published)
- To provide up to date LBC for our regional NWP system
- Suitable for next generation HPC

Designed choices

- New system to be based around hybrid-4DVar, with
  - a hybrid background error covariance,
  - a hybrid-Tangent Linear Model (HTLM),
  - a **R**apid **U**pdate **C**ycle.
- Ensemble approach considered:
  - Control-Pert ensemble based around 4DVAR?
  - Ensemble of independent 4DVAR ?



JADA-Global

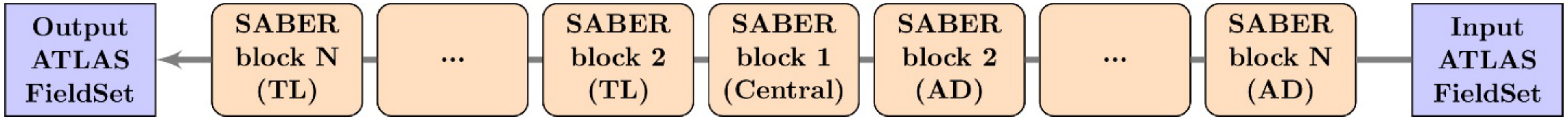
hybrid background error covariance

Background error covariances are fundamental to data assimilation as they **spreads** the information from the observations based on the prior uncertainty in the background state.

Like our current system we are developing a “**Hybrid**“ background error covariance model:

Hybrid **B** = Static **B** + Flow-dependent **B**

In JEDI, covariance models are represented by their primary constituent elements and put together through recipes; making them very easy to develop, modify and test.



$$B = L_N \; \dots L_2 \, L_1 \, E \, L_1^T \, L_2^T \dots L_N^T$$

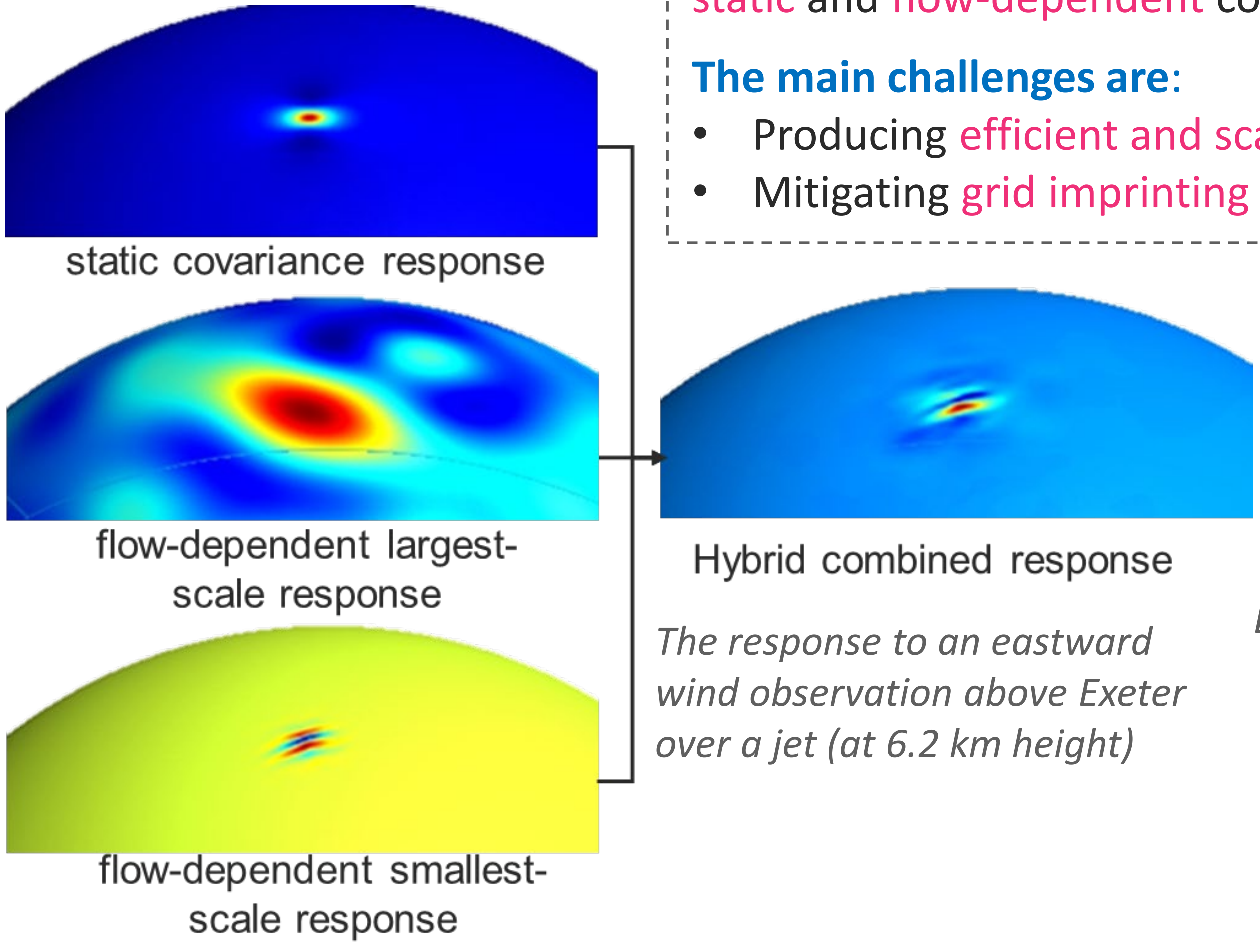
JADA-Global

hybrid background error covariance

The Met Office, similar to current system, uses **a spectral representation** within both the static and flow-dependent covariance models.

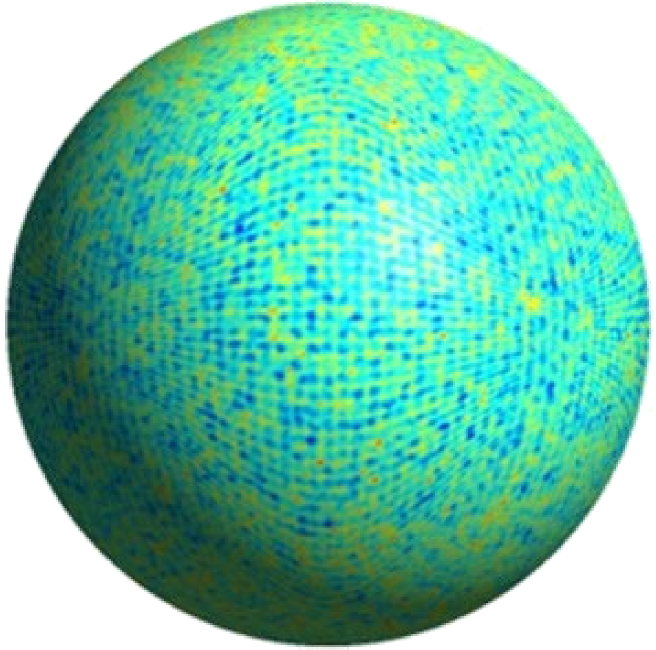
The main challenges are:

- Producing efficient and scalable code while keeping scientific accuracy.
- Mitigating grid imprinting when going from Gaussian to cubedSphere.



The response to an eastward wind observation above Exeter over a jet (at 6.2 km height)

Example of grid imprinting Eastward wind variance estimate at surface level



Development of TLM/H-TLM

A key feature of 4D-Var is the linearisation of the model (+adjoint)

A large proportion of the work in 4D-Var goes into building and maintaining this.

- Linearising the dynamical core of the model is relatively easy (line-by-line differentiation), **but** linearising the physical parametrisations is very difficult and challenging. → linear model often only accounts for a limited range of the physical parametrisation.

Solution

Taking advantage of previous experience using ensemble differences to reconstruct a TLM

- Replace coded TLM (of dynamics + physics) with coded TLM for dynamics only (simplified tangent linear model - STLM).
- Derive corrective coefficients representing the contribution of the physic using an ensemble.

Basic H-TLM principal

Suppose we have available a simplified TLM M_{i-1} so that

$$\delta\vec{x}(t_i)^- = M_{i-1}\delta\vec{x}(t_{i-1})$$

We can use an ensemble to find a correction N_i which will update $\delta\vec{x}(t_i)^-$ at time t_i

$$\delta\vec{x}(t_i)^+ = N_i\delta\vec{x}(t_i)^-$$

Payne, Monthly Weather Review, Vol 149, pages 3-19, 2021 A Hybrid Differential-Ensemble Linear Forecast Model for 4D-Var

Development of TLM/H-TLM

The introduction of the hybrid TLM is a major scientific change for the MO

PRO

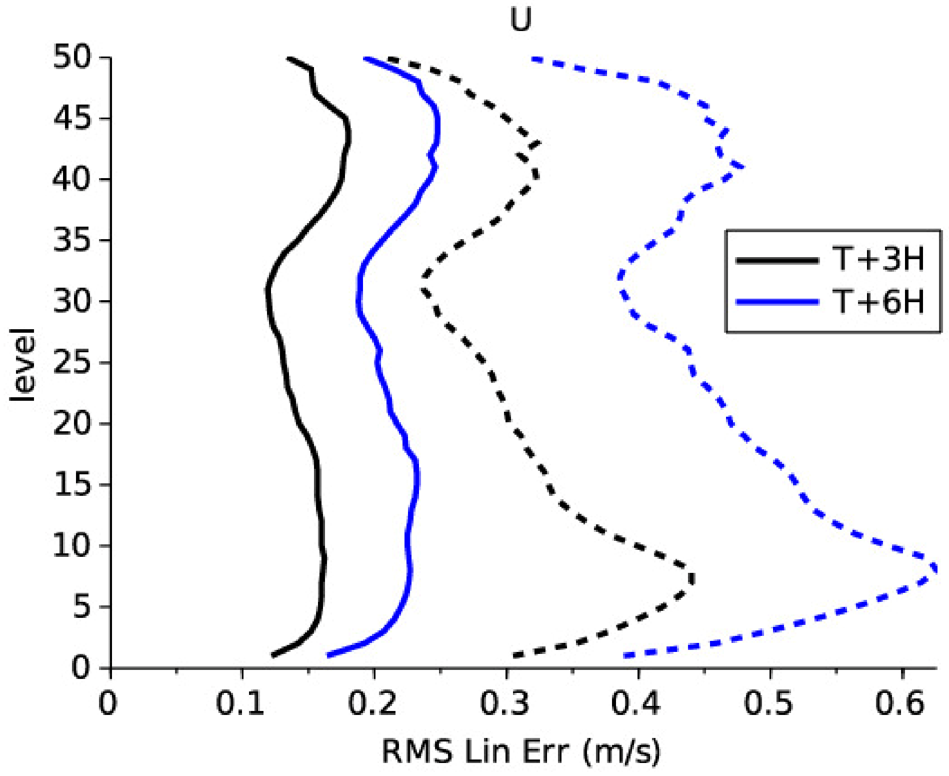
- All physics included without any complex decision-making
- Dramatic reductions in linearisation error
    - Large reductions in RMS T+6 forecast- obs (see figure)
- Relatively simple to code
- Automatically keeps up to date with changes in nonlinear model physics

CON

- Additional computational and I/O cost – however coefficients can be computed ahead of time, because observations are not needed.

Status

- There is a generic implementation (toys model, FV3, LFRic, …) and ready to be used.
- Waiting for our simplified tangent linear model to be available



Comparison of linearisation error using current PF Model (dashed) and hybrid TLM (solid).
- Tim Payne -

JADA-Global

Cycling 3DVAR

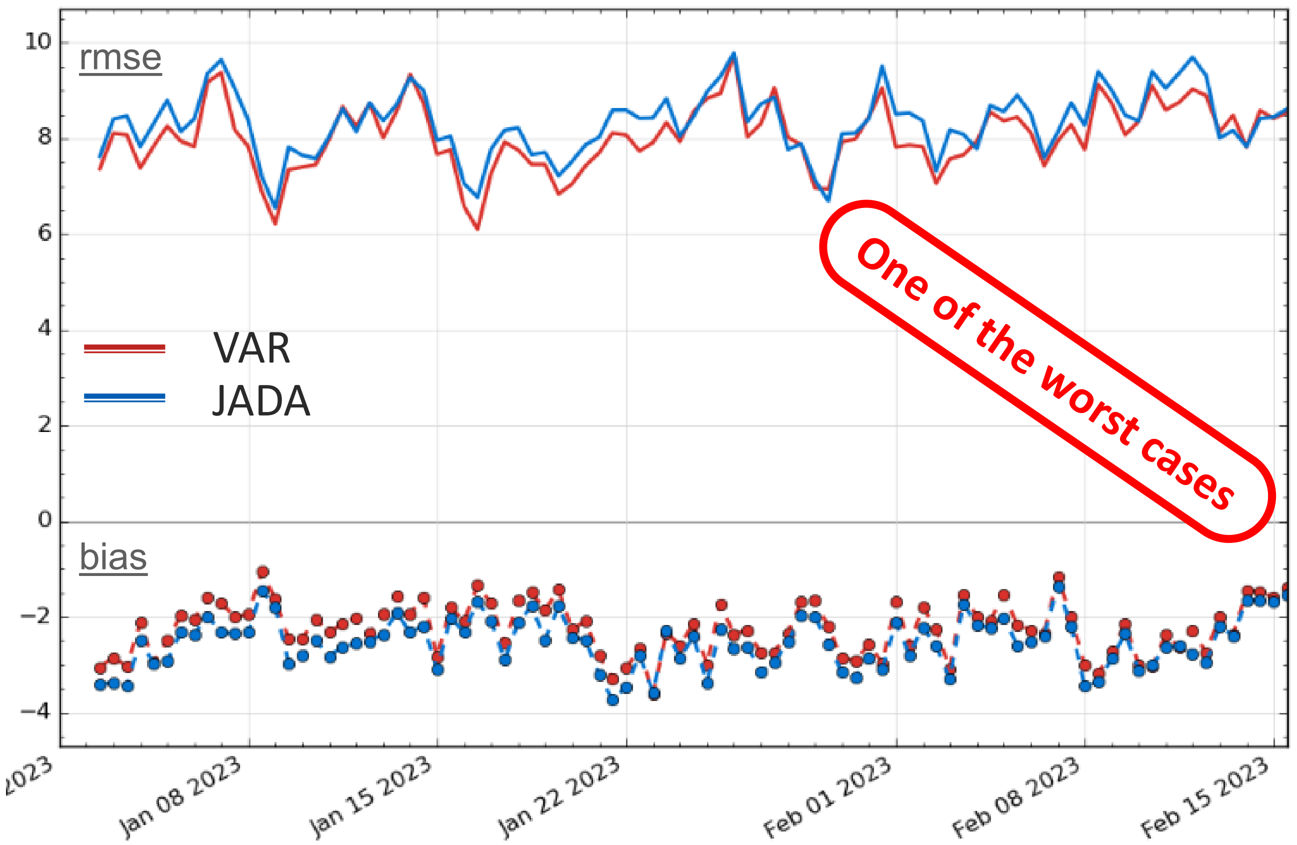
Evaluation of JADA 3DVar:

- >11 observation types available
- Resolution C224 (LFRic) / N320 (UM) → ~40km
- No VarBC (Variational Bias Correction) yet

Outcomes:

- VAR: slightly better verification performance
- JADA: better o-b (obs – background) diagnostics
- Overall, VAR and JADA have similar performance
  - Only minor differences, no anomalies
  - JADA is consistent with what we expect under such circumstances

RMSE: Geopotential Height (m) @ 500hPa Southern Hemisphere against ECMWF analysis | T+24 | 11 Obs. types



JADA-Global

Which Ensemble Data Assimilation

Two approaches for our Ensemble of DA:

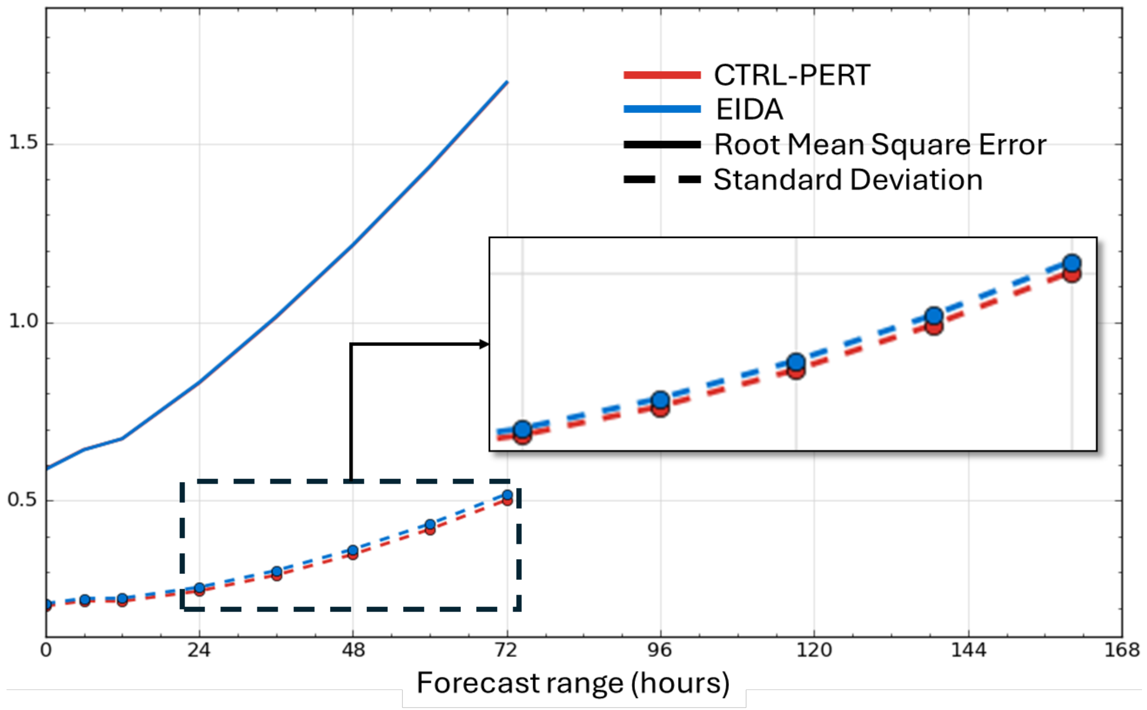
- (Option 1) Ctrl-Pert (Ctrl = 4DVar, Pert = 3DVar)
  - Cheap to run & recentring around control
  - However, simplification/approximation solution
- (Option 2) EIDA: Ens of Independent 4DVars
  - More accurate & suitable for extreme events
  - Every ensemble member forecast as good as the other
  - However, expensive to run

Decision depends on:

- model resolution,
- ensemble size,
- number of outer-loops,
- I/O, memory footprint,

as well as resources available for DA in the critical path.

Temperature (K) @ 500hPa, North Hemisphere
Equalized and Meaned between 20230101 and 20230131
Against ECMWF analysis



Our current development path will enable us to commit to one of the choices as late as possible

Met Office

JADA: JEDI-based Application for Data Assimilation
Regional NWP

J. Waller, C. Thomas, T. Vockerodt, M. Wlasak, O. Lewis, O. Aljabar, K. Franklin, T. Hutchins, R. Nixon-Hill, M. Dell

In collaboration with the Joint Center for Satellite Data Assimilation, NOAA and the Bureau of Meteorology

JADA-Regional

JADA-Regional - Scope

"putting ensembles at the heart of everything we do"

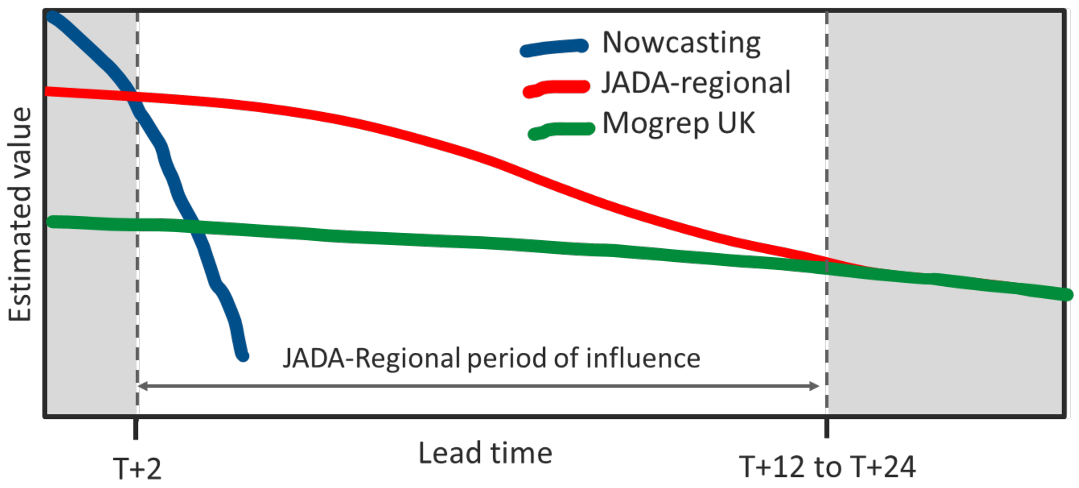
Requirements

- A fully ensemble-based NWP system.
- Sit between our nowcasting and downscaler system.
- Focus only on well-proven science (operational or published)
- Focus on convective scales and providing flow dependent information.
- Harvest as much as possible existing JEDI's capabilities.

Designed choices

New system to be based around Ensemble Kalman Filter, with

- Large scale blending from hourly global run (IC & LBC)
- Priority on high resolution, high coverage observation
- Improve observation error representation



JADA-Regional – Areas of investigation and development

Filter flavours

Stochastic

- Use perturbed observations to update each ensemble member individually.
- Calculate mean and covariance of updated ensemble.

→ Can provide scenarios, but require more ensemble members (>40-50)

Deterministic

- Use linear and Gaussian theory to update ensemble mean and analysis perturbations.

→ Require less ensemble members, but provides mean solution, and associated 'ensemble spread'

Effective ensemble size

- We will have cross-validation to reduce risk of ensemble collapse.

Information on different scales

Assimilation will analyse small scale, but we must ensure that information on all scales is accurate. Will investigate:

- Large scale blending | Multiscale localization

Complex Observation Error

Better error representation by:

- Introducing different error distributions (e.g. inverse-gamma) for bounded variable.
- Implement capability to handle a full observation error covariance matrix

Summary

Summary

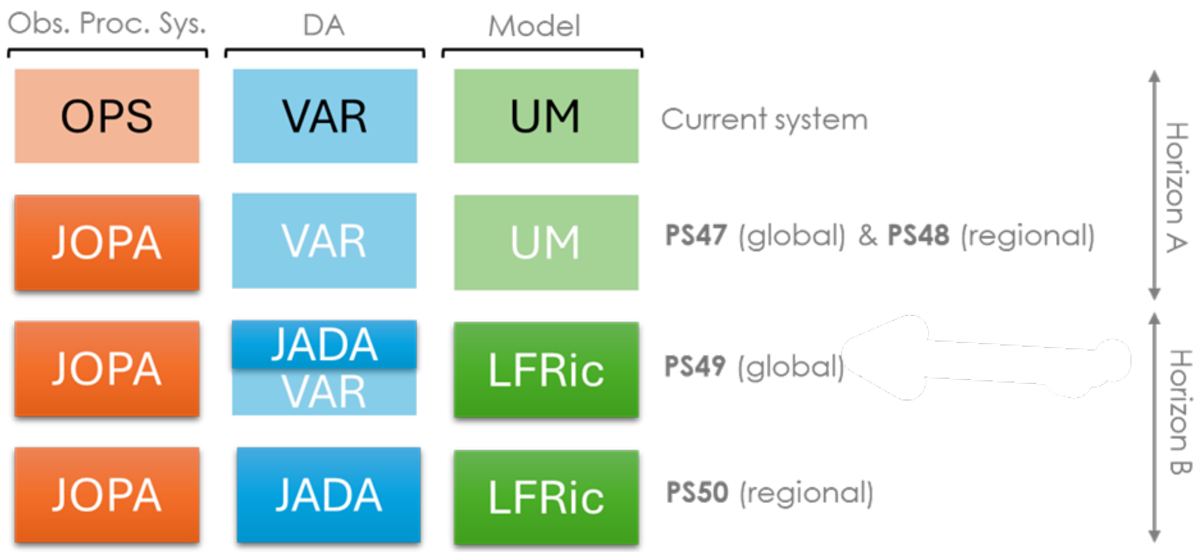
Our new observation processing and data assimilation system based on JEDI will enable the Met Office to fully exploit future generations of supercomputer.

This work is achieved in close collaboration with the Joint Center for Satellite Data Assimilation (JCSDA).

We have already seen benefits from JEDI: turnaround of new observation, improve collaboration (e.g. University) and provide a very portable infrastructure

Start of parallel Suite

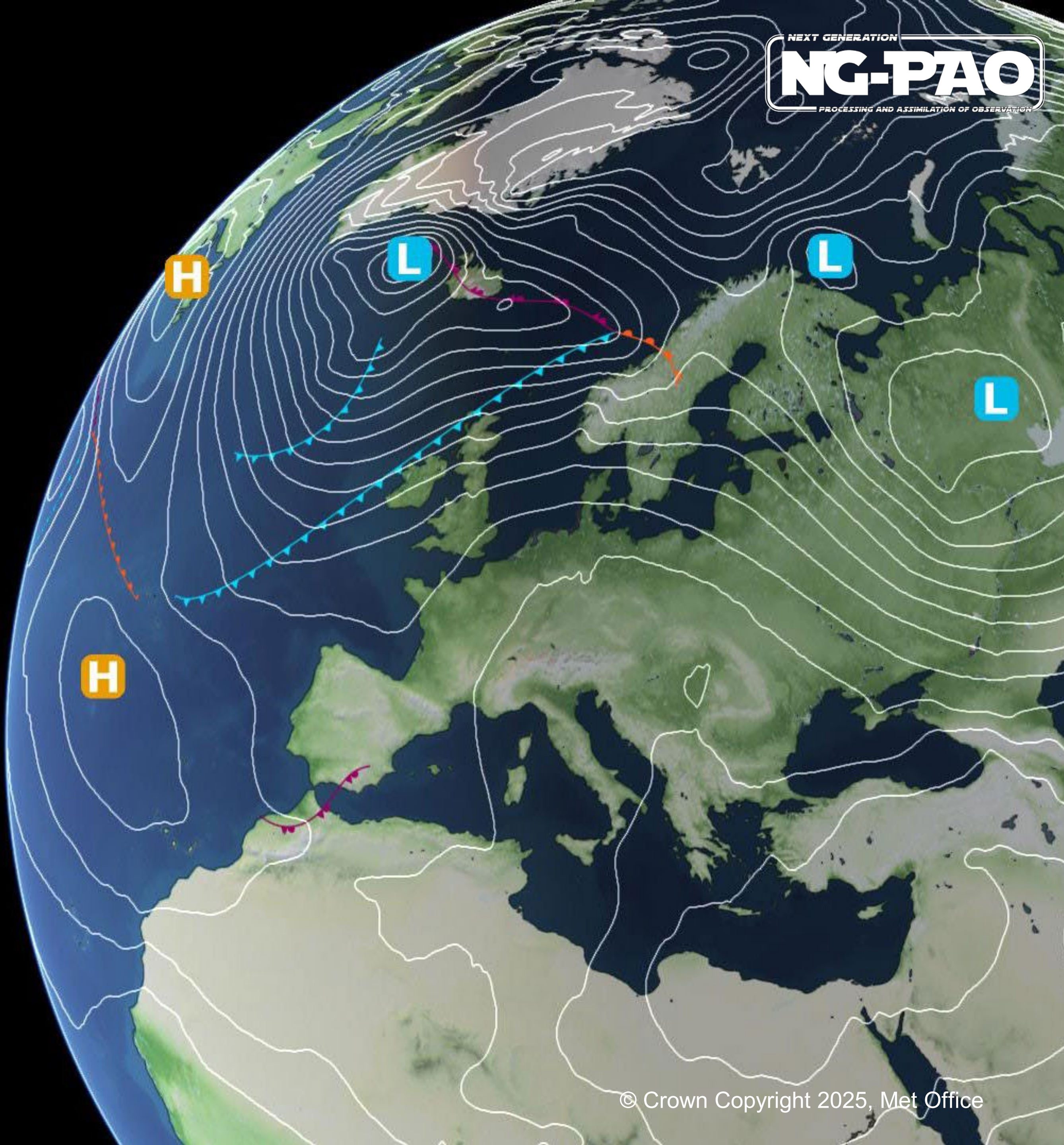
- *PS47: now*
- *PS48: in ~1 year*
- *PS49: in ~Sept 2027*
- *PS49: in ~Sept 2028*



Met Office

Thank you

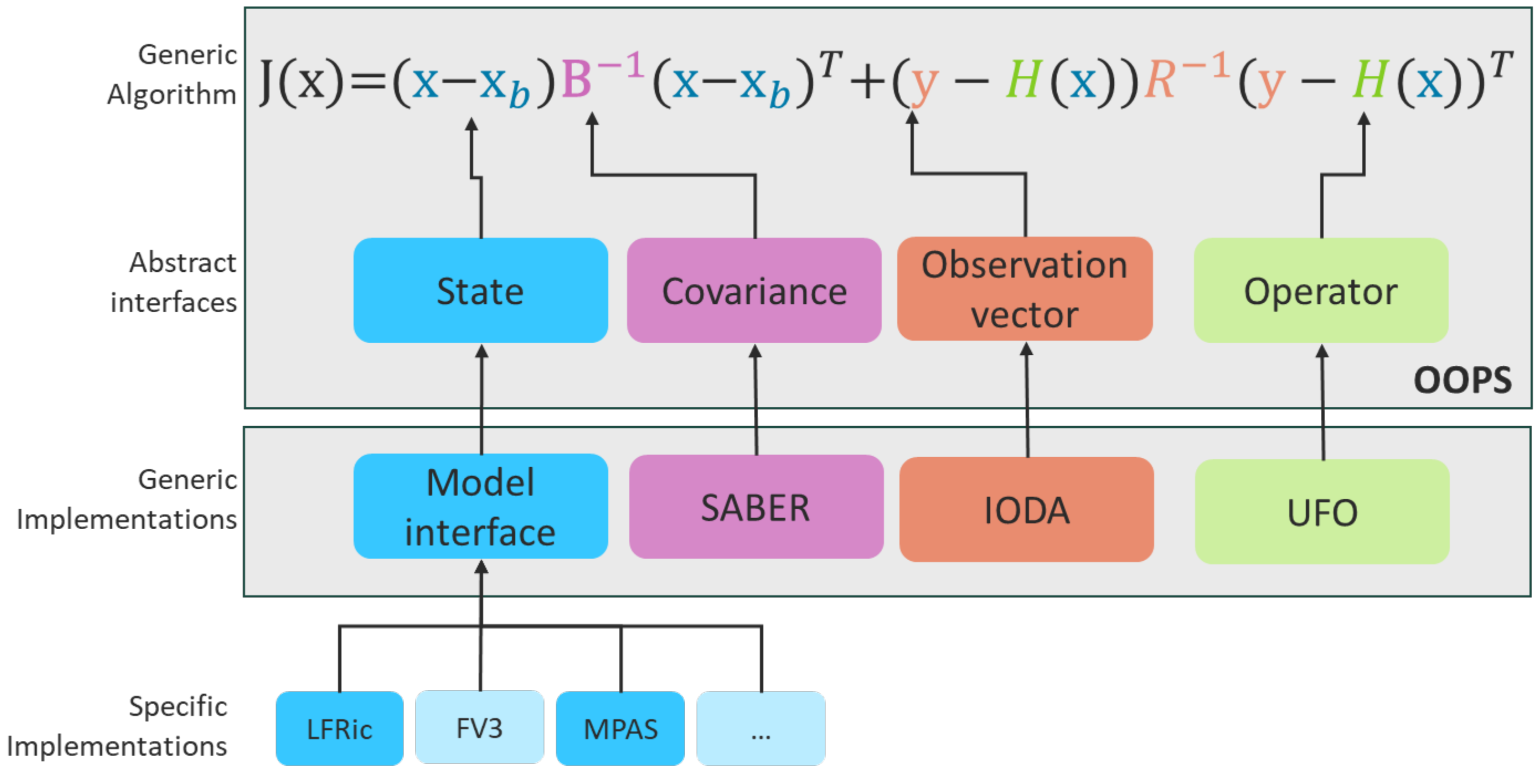
Any Questions?

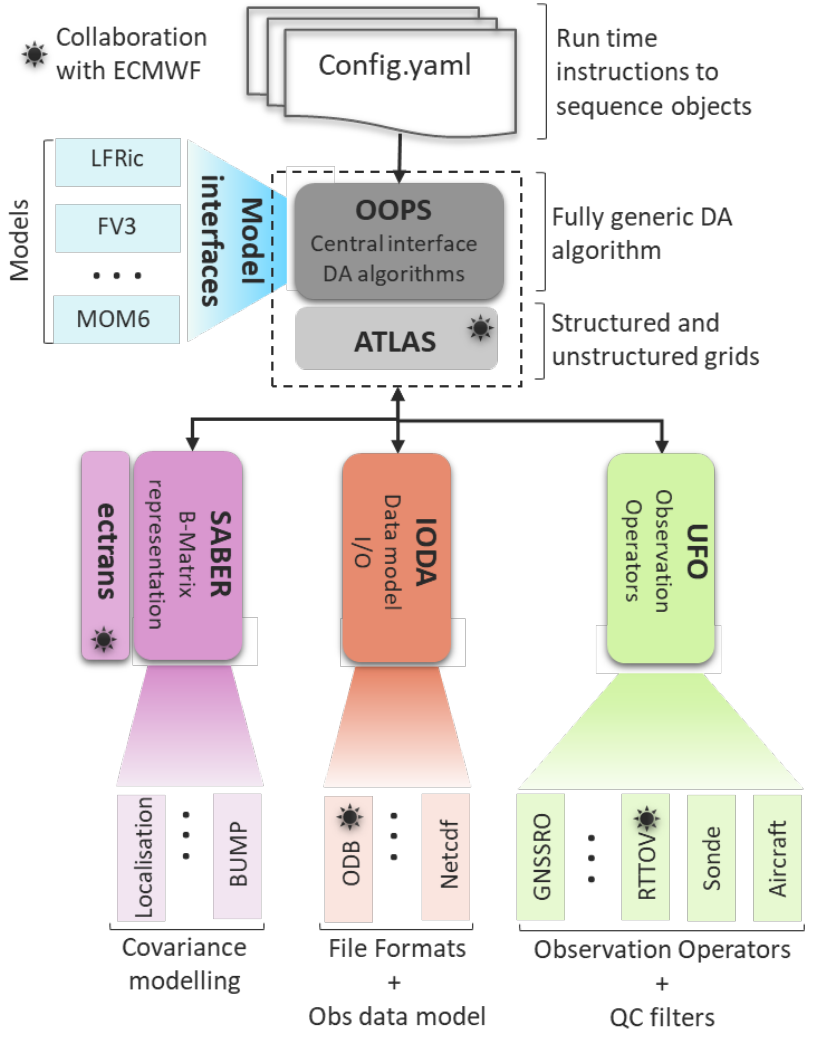


Additional Slides

JEDI

JEDI - Structure



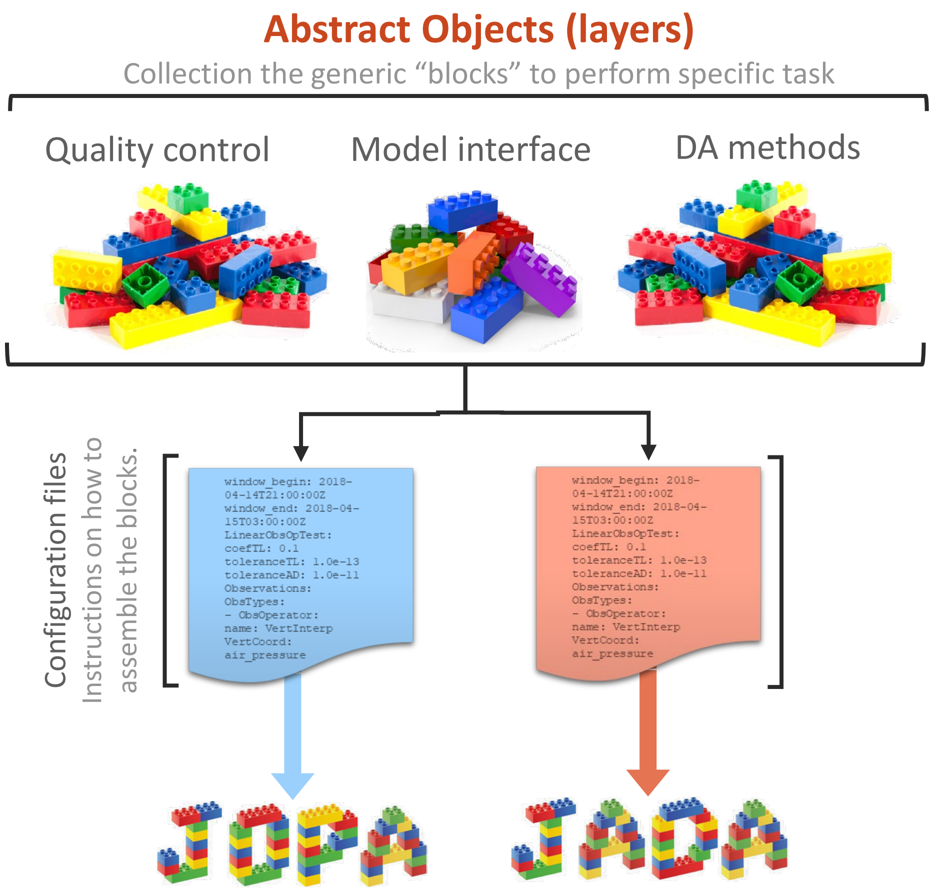


JEDI - Structure

Our current **observation processing and data assimilation system**, is very static:

- the logical chain of processing is explicitly coded,
- any processing changes require a code change, with the need to handle inevitable side effects.

Our new system (JOPA and JADA) is considerably more flexible and "dynamic". It relies on JEDI's underlying design and its abstract layers to build any application, applying any scientific setting from a simple configuration.



JEDI - Example of flexible configurations

The science settings are purely expressed in the configuration file using sequencing of generic procedures.

User defined diagnostics flag

```
- filter: Create Diagnostic Flags
  flags:
  - name: NEW-FLAG-DEFINITION
     initial value: false

- filter: RejectList
  where:
  - variable: {name: MetaData/stationId}
  [...]
  actions:
  - name: set
    flag: NEW-FLAG-DEFINITION
  - name: reject
```

Repetition of procedures

```
- filter: Parameter Substitution
  section to repeat:
    filter: Poisson Disk Thinning
    [...]
  repetitions:
  - min_horizontal_spacing:
    - { "0": 2000, "1": 1000 }
    - { "0": 3000, "1": 2000 }
```

Flexible scientific configuration

```
- filter: Gaussian Thinning
    where:
    - variable:
        name: MetaData/SubTypeNum
      is_in: 30500, 30600, 30700
    horizontal_mesh: 120.0
    [...]
```

⬇

```
- filter: Gaussian Thinning
    where:
    - variable:
        name: MetaData/SubTypeNum
      is_in: 30500, 30600
    horizontal_mesh: 120.0
    [...]
```

```
- filter: Gaussian Thinning
    where:
    - variable:
        name: MetaData/SubTypeNum
      is_in: 30700
    horizontal_mesh: 200.0
    [...]
```

Create and assigned a new variable

```
- filter: Variable Assignment
    where:
    - variable:
        name: MetaData/SubTypeNum
      is_in: 30700
    [...]
    assignments:
      - name: DerivedObsValue/windSpeed
        type: float
        function:
          name: ObsFunction/Arithmetic
          options:
            variables:
            - name: ObsValue/windSpeed
            coefs: [0.514444] # convert kts to m/s
```

There are also:

- Multiple variable transforms
- Collection of H(x)
- …

We currently porting our LAND observation with almost no code change, just creative Yaml.