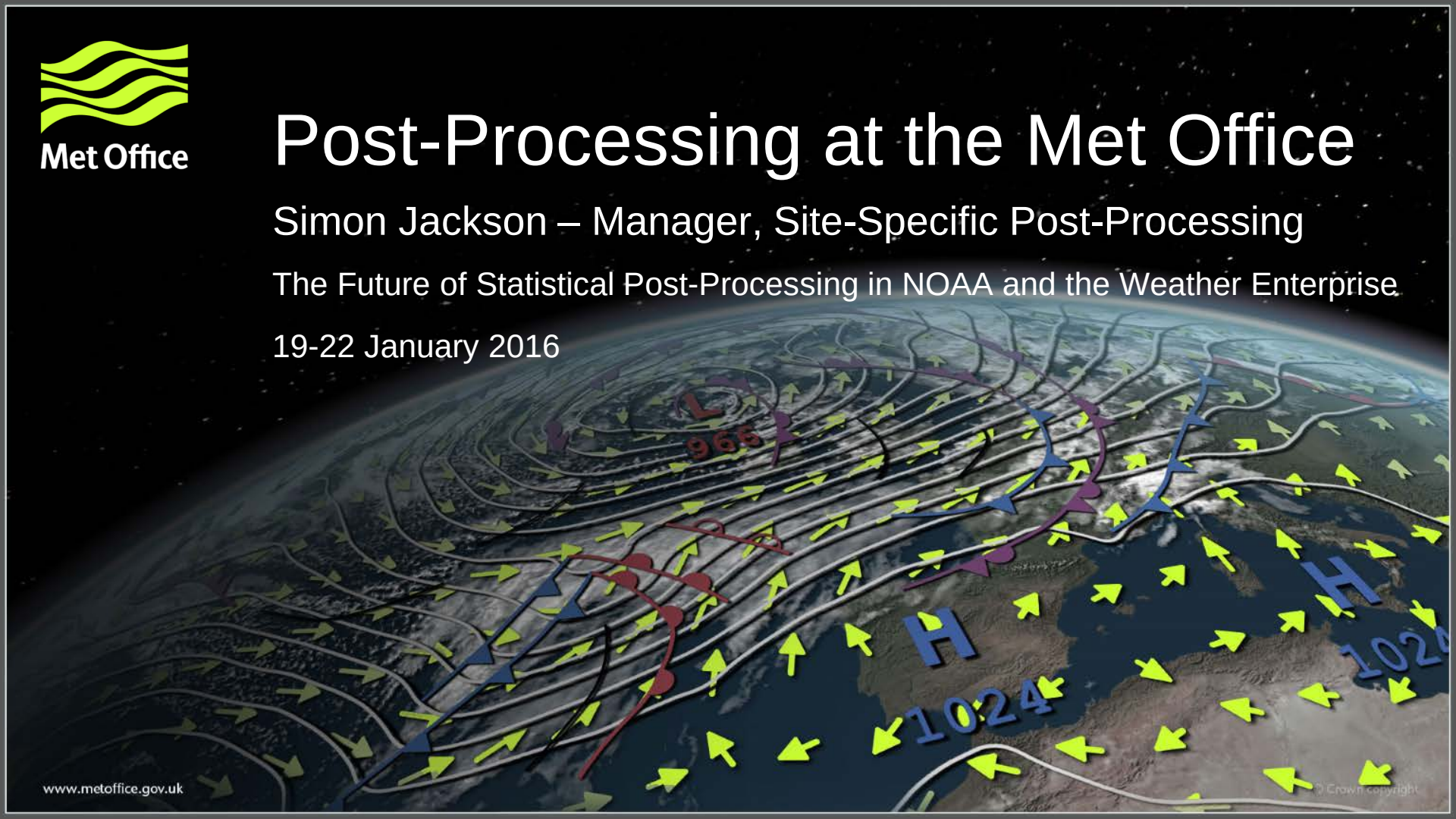
Post-Processing at the Met Office

Simon Jackson – Manager, Site-Specific Post-Processing

The Future of Statistical Post-Processing in NOAA and the Weather Enterprise

19-22 January 2016



With thanks to …

James Canvin, Paul Abernethy, Caroline Jones, Nina Schuhen, Bruce Wright, Andrew Bennett, Stephen Moseley, Ian Pearman, Marion Mittermaier, Nigel Roberts, Ken Mylne …

and to Matt Peroutka for inviting me!

Met Office

Where are we now?

Where next?

Beyond that?

Where are we now?

Separate, intertwined production chains for

- Gridded data
- Spot data
- Verification

Also, some legacy production chains that need to be ‘retired’

Post-Processing

- We have a complex, evolving suite of gridded NWP models
- We want useful, usable, accurate forecasts on grids and point locations
  - Agrees with what's happening now
  - Consistency from update to update
  - Most user doesn't need to know details of NWP suite
  - Makes best use of models

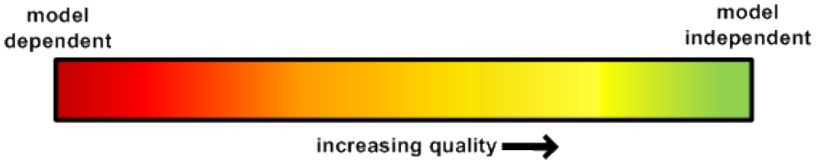
Post-Processing used to

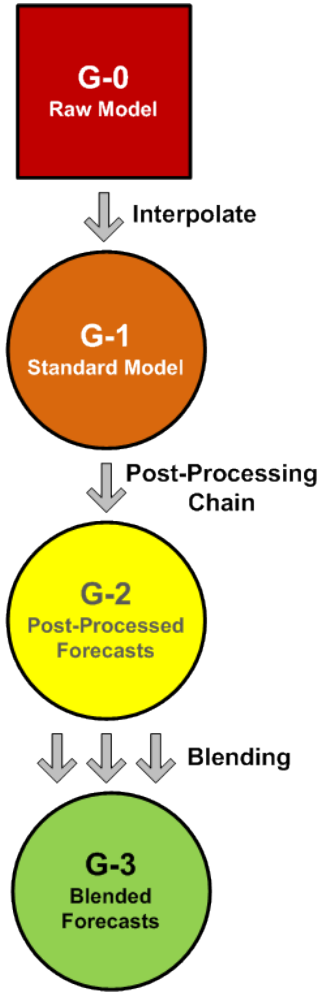
- Make best use of the available data
  - Physical downscaling
  - Statistical corrections
  - Blending
- Improve the delivery
  - Seamless dataset from nowcast to medium-range
  - Regular delivery times
  - Standard data formats and parameters

Processing Levels

This represents a progression:

- **in the application of scientific correction**
- **in data standardisation (decoupling from models)**





Closely-linked models

Research

Specialist Use

Downstream models

Research

Forecasters

Collaboration

Wholesale

Majority of Users

Gridded Post-Processing

Run models

Downscale to standard grids

Calculate diagnostics

No Kalman filter (yet)

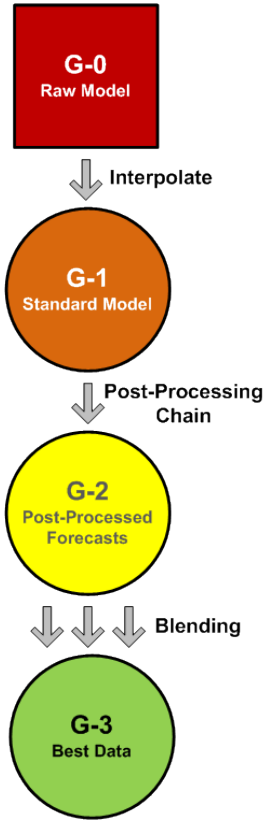
Processing of ensembles

No ensemble calibration (yet)

Upscaling and neighbourhood products

No blending (yet)

Many gridded products done this way for past 10 years



Site-Specific Post-Processing

Run models

Extract site data

Calculate diagnostics

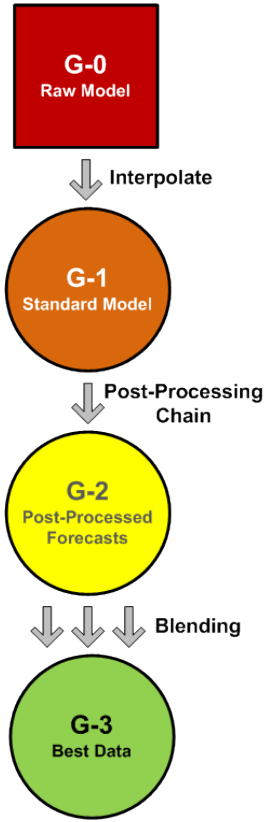
Run Kalman Filter

Processing of ensembles

No ensemble calibration (yet)

Blend with previous forecast to create update

Most of our site forecasts generated this way for past 5 years



These run on our HPC

- Mostly written in Fortran
- Includes code inherited/adapted from older systems
  - Bespoke / in-house file formats
- Used to generate most public and commercial products
- Gridded PP separate from Site-Specific PP → Inconsistencies
- Both manipulate data and match forecasts against observations
  - as does verification system

Where next?

Best Gridded Data

Non-operational prototype exists

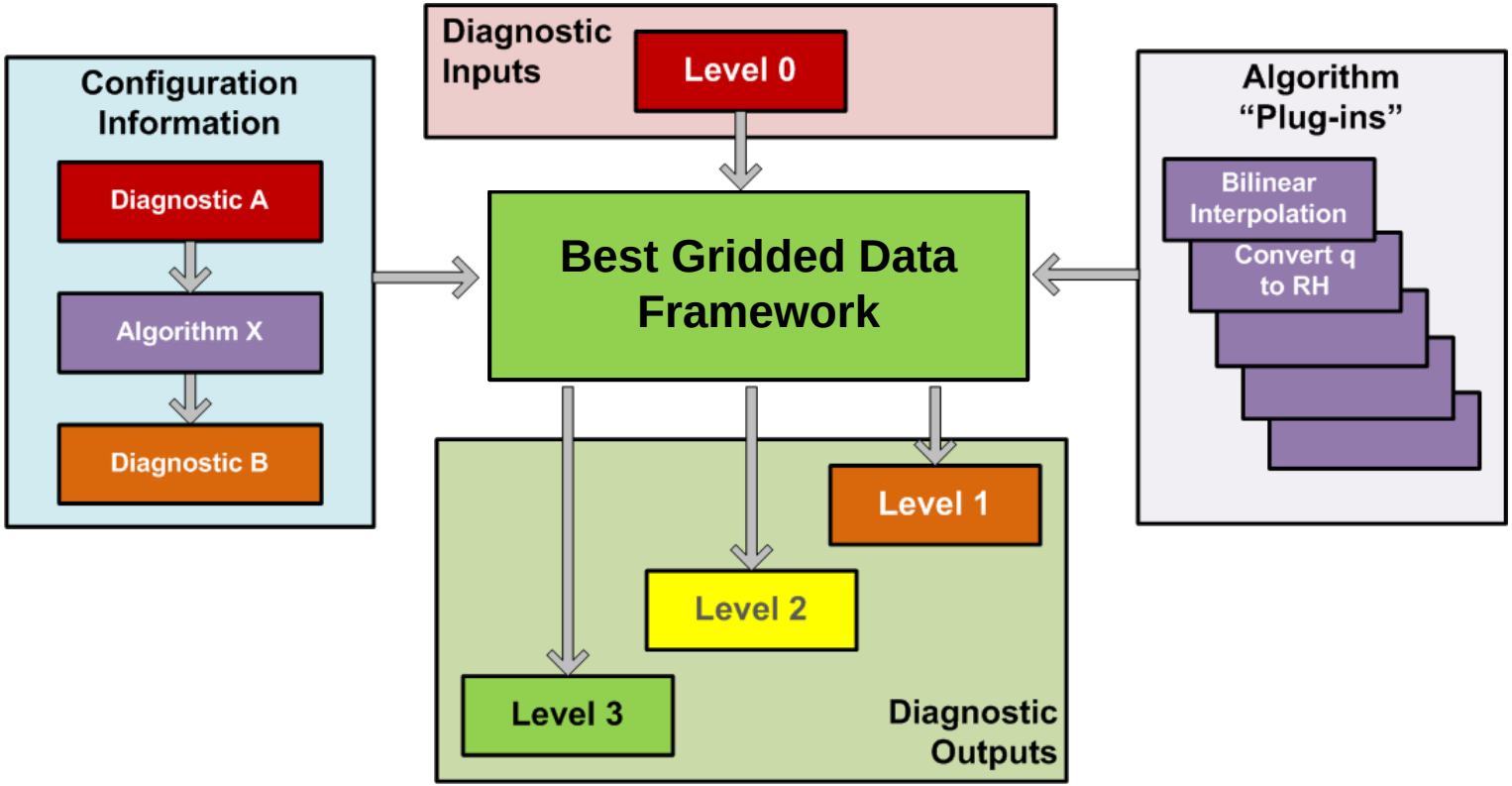
Similar methodology to site-specific process

- Includes physical downscaling
- Will include gridded Kalman Filter and ensemble calibration (EMOS)
- Includes blending of model cycles
- → less inconsistent with site forecasts

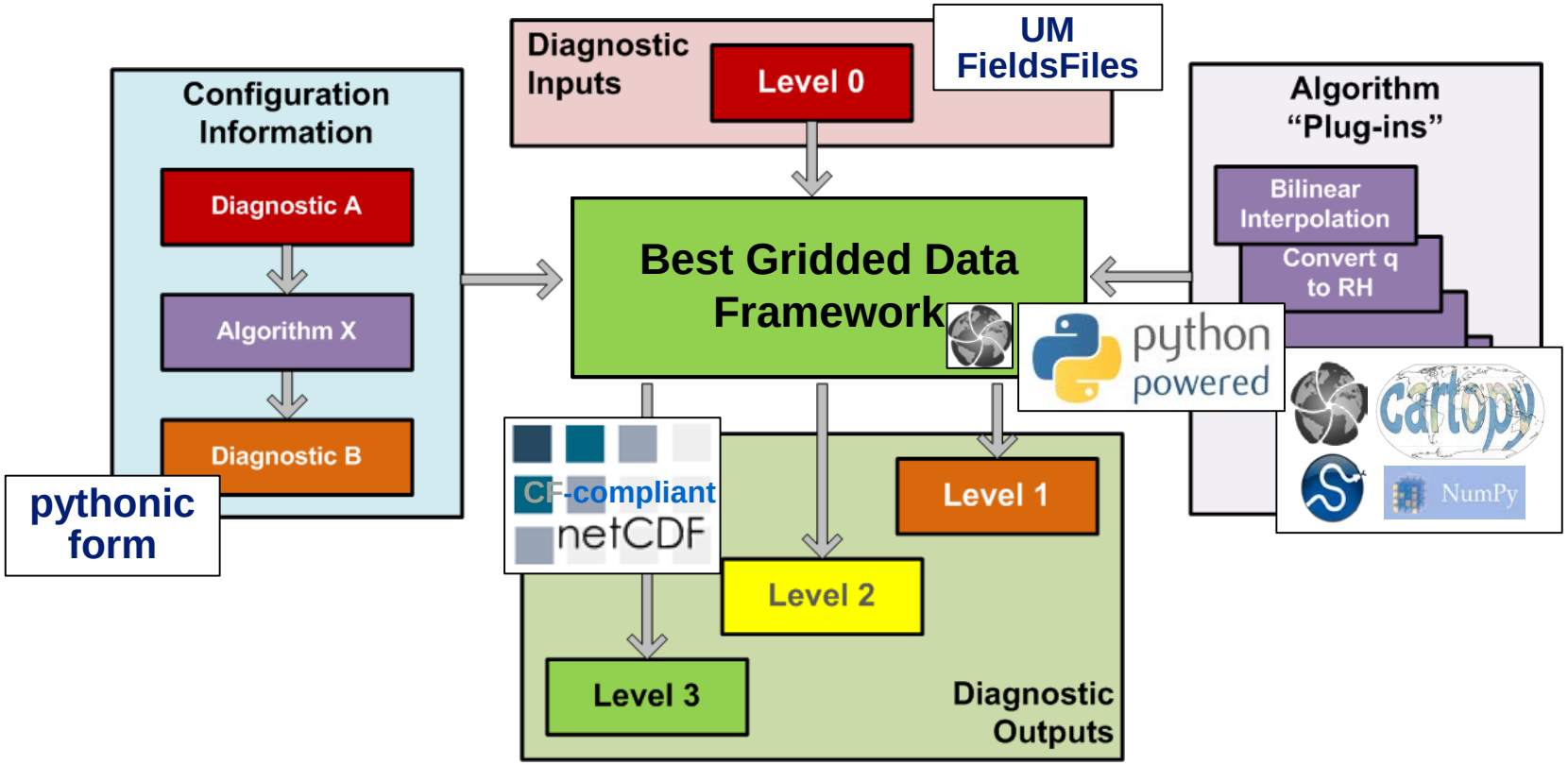
Written in python using Iris libraries

- Run on the HPC

Software Framework



Software Framework



Beyond that?

Post-Processing Review

Gridded PP, site-specific PP and verification all do similar things

- Distil large volumes of raw data into useful information
- Manipulate forecast data
- Match forecasts against ‘truth’

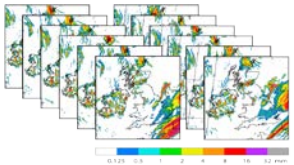
Systems developed independently

➔ inconsistency and duplication

Ongoing review of Post-Processing strategy is proposing an integrated approach

Integrated process could look like ...

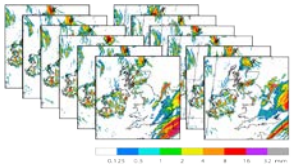
Met Office



Raw NWP forecasts

Verify

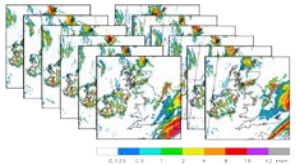
Physical post processing



Modified forecasts

Verify

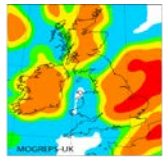
Statistical post processing



Modified forecasts

Verify

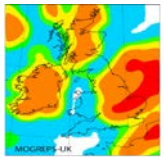
Neighbourhood processing



Probabilities

Verify

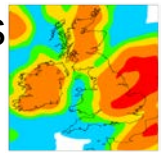
Statistical calibration



Adjusted probabilities

Verify

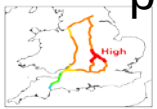
Probability blending



Blended probabilities

Verify

Generate gridded and point products



Verify

Choices will depend on ...

Stability of NWP suite

- Availability of training data
  - Adaptive Kalman Filter → deep machine learning

Scope and stability of user requirements

Compute resource & robustness of suites

Where in process to move from model grids to standard ones

Adoption of full probabilistic framework

Staff resource needed to monitor, maintain and develop system

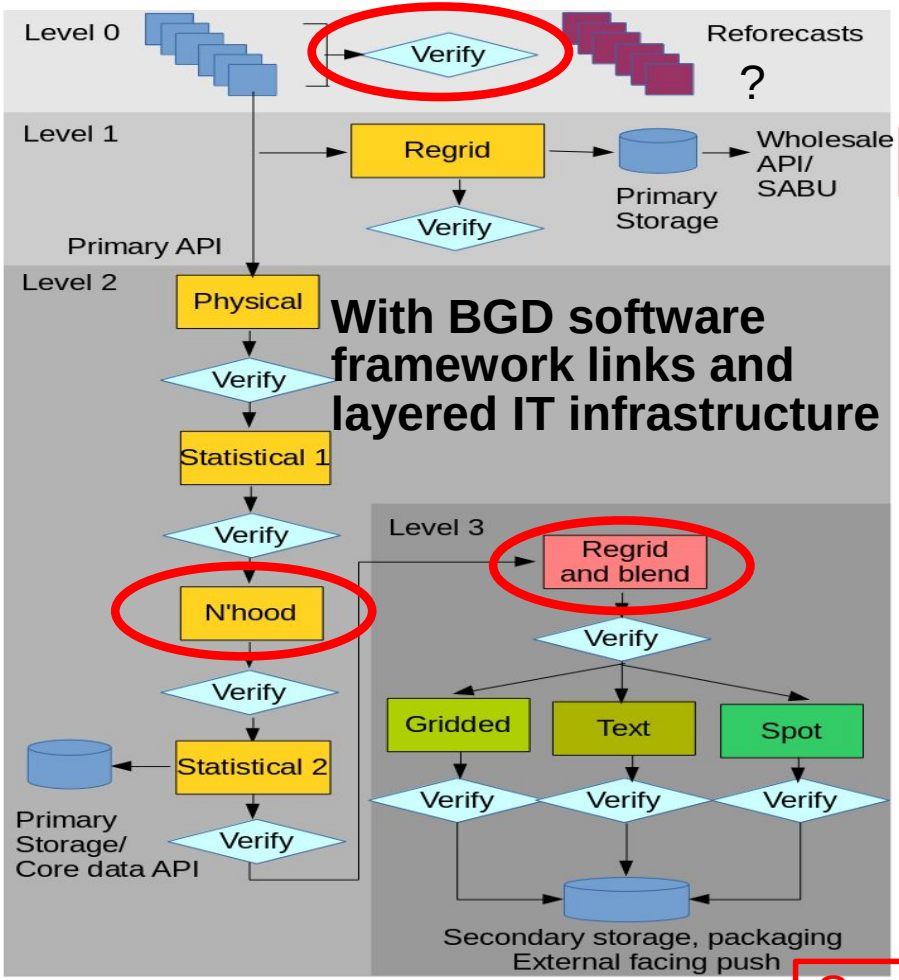
Accuracy vs precision vs consistency

Met Office

Any Questions?

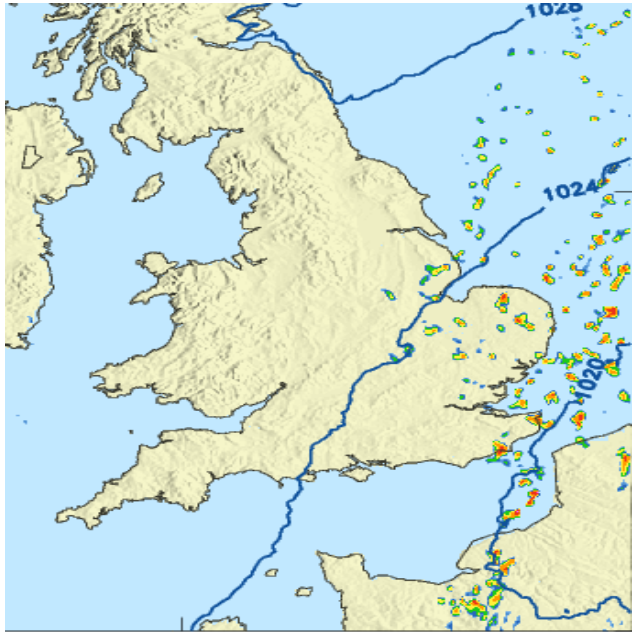
An initial outline of the new **single integrated post-processing chain** and how it fits in with BGD and the proposed IT infrastructure

Chain advocates the extraction of “spot” forecasts at the end of the processing chain



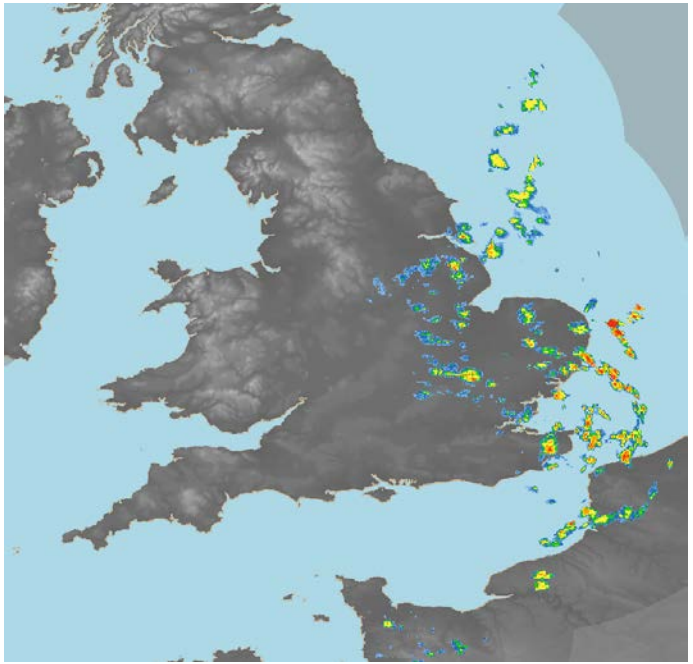
Great forecast

But the individual showers are mostly wrong



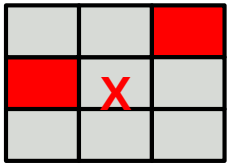
UKV forecast

UKV op Precipitation rate [mm/hr] and PMSL
Wednesday 1200Z 14/10/2015 (t+9h)

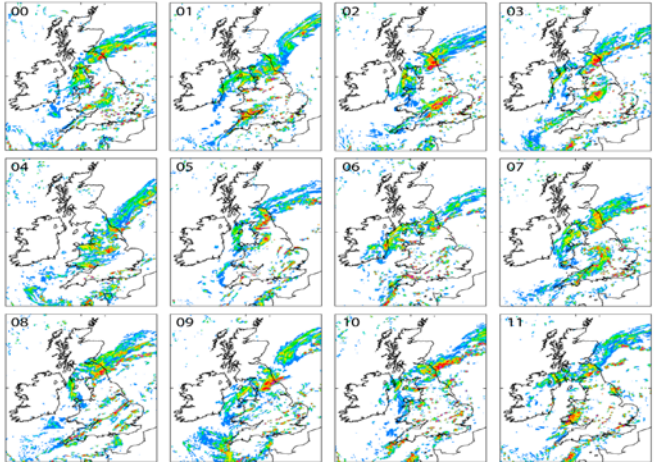


Radar

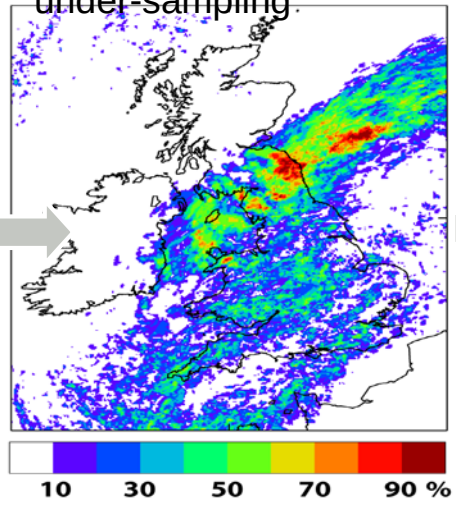
Making a probability forecast



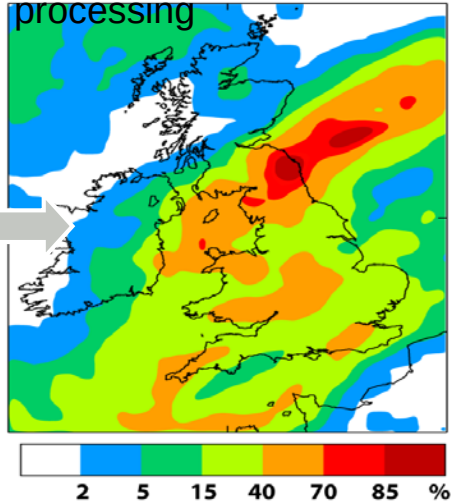
MOGREPS-UK



Gaps because of under-sampling



Smooth probabilities with additional neighbourhood processing

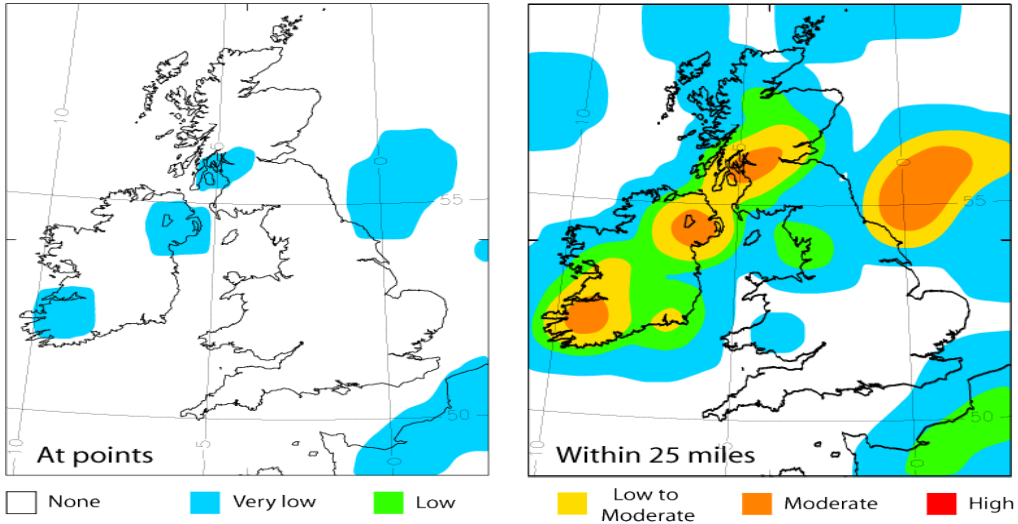


Spatial post processing (neighbourhood) essential

R&D neighbourhood size, adaptive methods, topography, blending probabilities

Some other considerations

Localised extremes



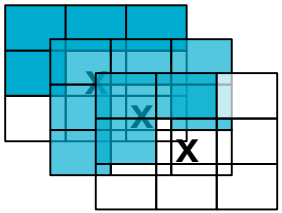
Topography

Spatial methods with small ensembles must deal with this

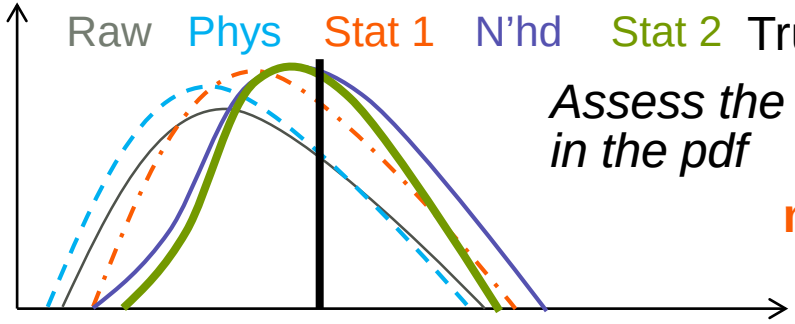
Verification at every stage

(when everything is on a grid)

Threshold exceeded

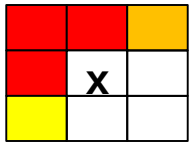


Same metrics at every stage



Assess the change in the pdf

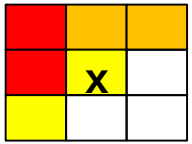
model-oriented and user-relevant



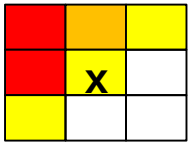
Raw

Framework is probabilistic.
Can extract and verify distributions at grid locations, both in space (when gridded truth is available) or at observing sites (using HiRA)

Stage 1



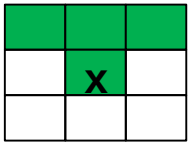
Stage 2



Check value added by each step

.....

Truth

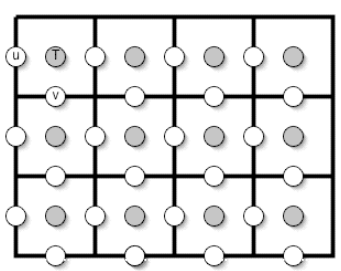


Met Office

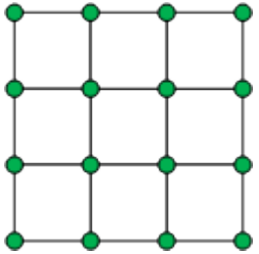
www.metoffice.gov.uk

Horizontal Grids

Proposed initial versions



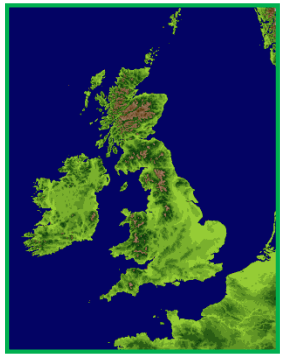
Staggered grid,
Model CRS



Regular grid,
Standard projection

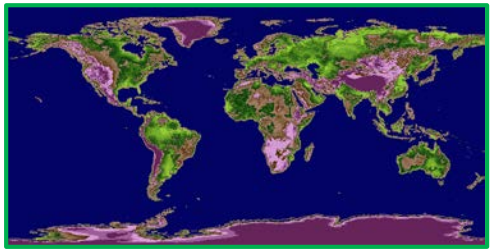
UK

2 km UKPP grid
OSGB National Grid



Global

~17 km grid (0.25°x 0.15°)
Equirectangular projection
(latitude-longitude)

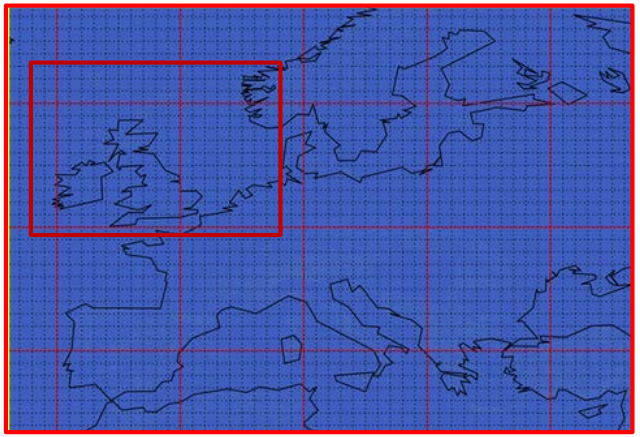


Europe

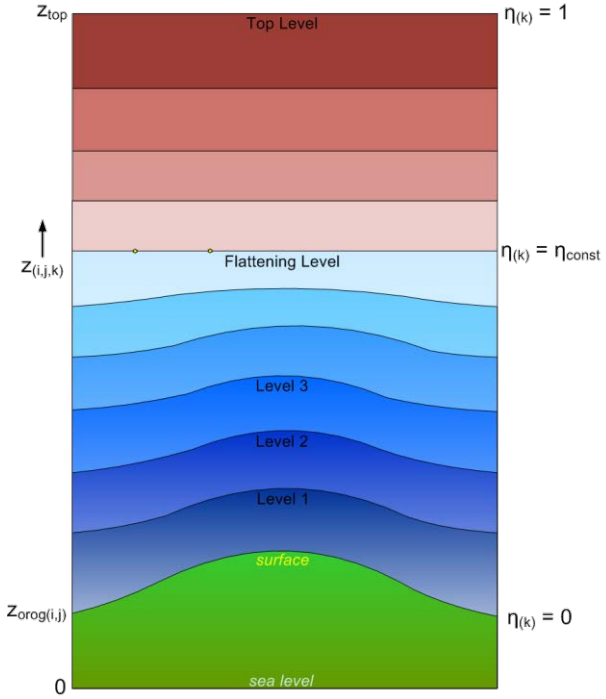
4 km Wholesale grid
(0.04° x 0.04°)
Equirectangular projection
(latitude-longitude)

UK & Surrounding Waters

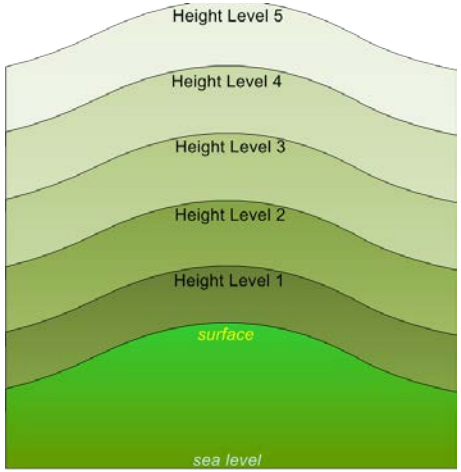
2 km grid (0.027° x 0.018°)
Equirectangular latitude-longitude projection



Vertical Levels

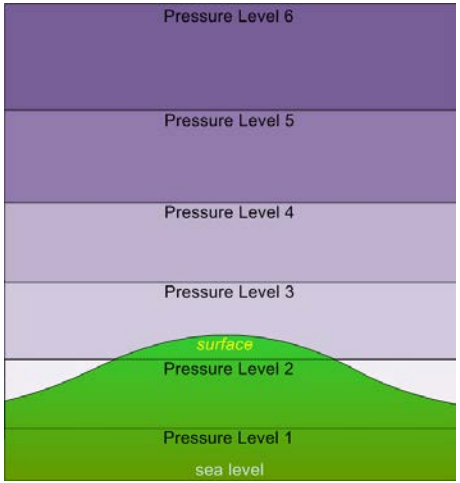


Hybrid height,
with Staggering



Height above Orography

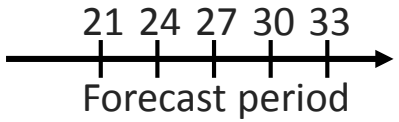
Initial version: 33 levels,
from 5 m up to 6 km



Pressure

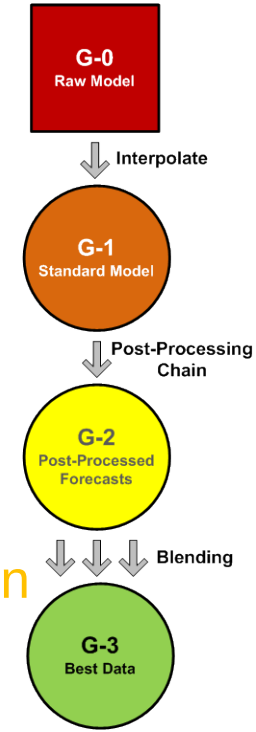
Initial version: 30 levels,
from 1000 hPa
up to 10 hPa

Temporal Aspects

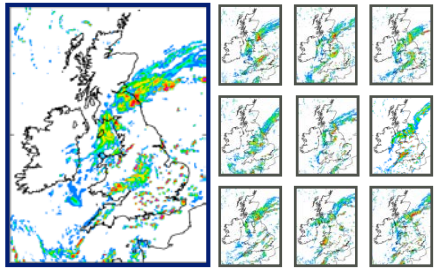


- Level 1 (model) and level 2 (post-processed) data made available as models complete
- Level 3 (blended) data re-issued hourly
  - sub-hourly updating still to be considered

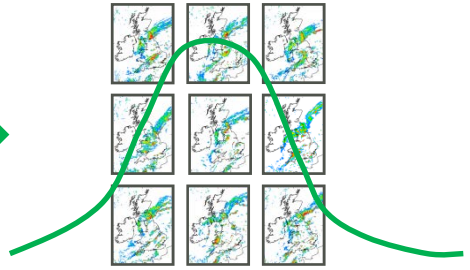
- Standard time steps out to 2 weeks
  - actual time steps to be confirmed for initial version
  - time interpolation where necessary



Ensembles



Deterministic forecast supported by ensemble members and derived probabilities



Ensembles are key members, **Percentiles** & derived probabilities

Level 1 (model) data made available as ensemble members

- Full set of members offered initially
- Versioning used to manage increases in available ensemble members

Level 2 (post-processed) and Level 3 (blended) data made available as a set of percentiles (representing a PDF):

- 13 percentiles: 5%, 10% 20% 25%, 30%, 40%, 50%, 60%, 70%, 75%, 80%, 90%, 95%
- Lower and upper bounds, mean, mode & SD

For multi-level diagnostics, just mean and SD

Site-Specific Process