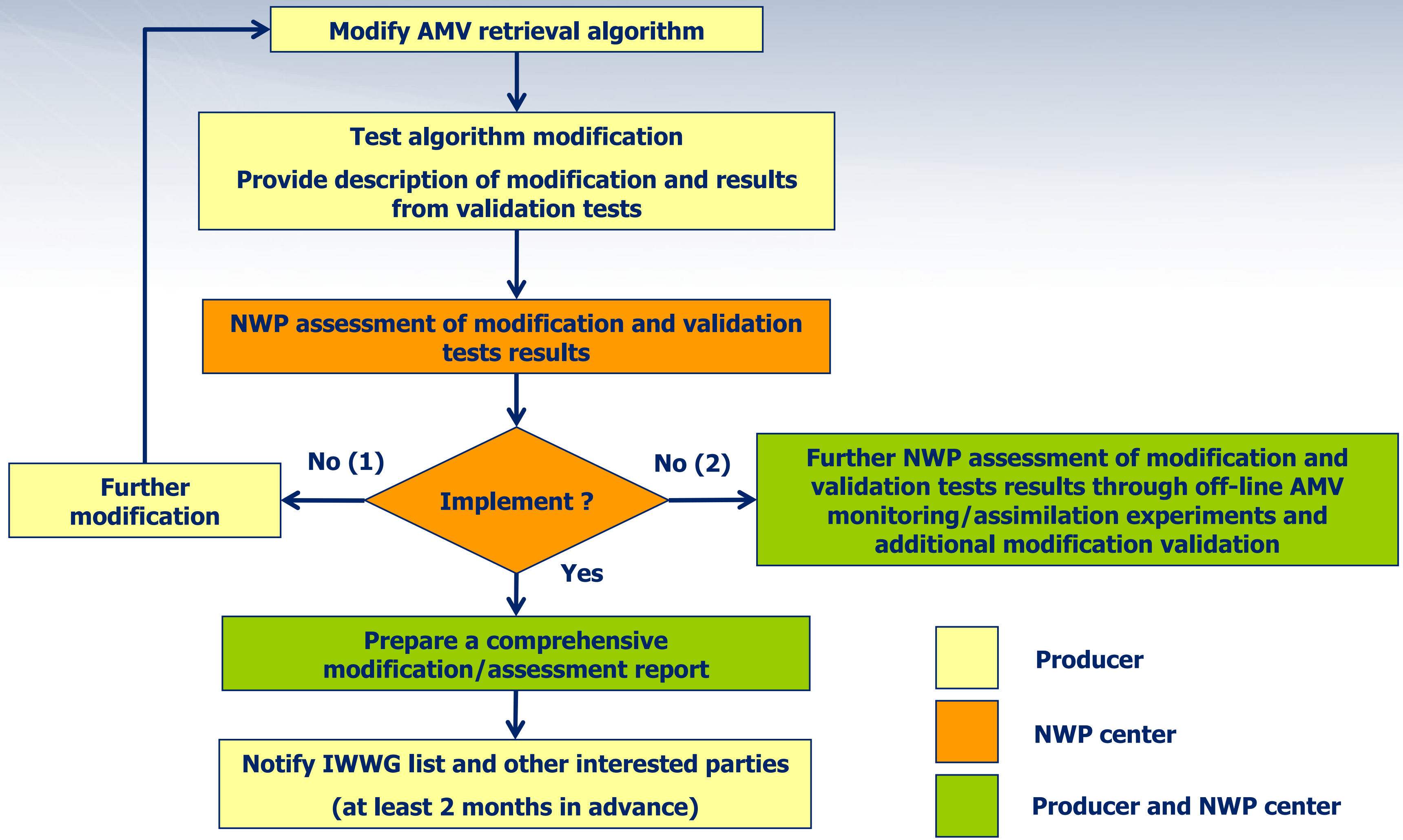
The Art of Evaluating AMV Algorithm Changes

Arthur de Smet (EUMETSAT)
Iliana Genkova (ECMWF)
Mary Forsythe (UK Met Office)
Niels Bormann (ECMWF)

Background

- Evaluation of AMV algorithm modifications has become less straightforward, as algorithms are more advanced and sophisticated, and user requirements tighten.
- Changes in operational AMV algorithms are implemented after evaluation performed by winds producers and a designated NWP centre.
- How can we optimise the Producer ←→ NWP centre *dialogue* ?
- How can we ensure "safe" algorithm modification implementation, without hampering progress ?

General Framework for AMV Algorithm Testing



Methods of Algorithm Evaluation

- Visual inspection
- Collocations with other observations, e.g., radiosonde
- Comparison against short-term forecasts or analyses from NWP
- Forecast impact trials in NWP
- Advanced observation diagnostics

Methods of Algorithm Evaluation

- Visual inspection
- Collocations with other observations, e.g., radiosonde
- Comparison against short-term forecasts or analyses from NWP
- Forecast impact trials in NWP
- Advanced observation diagnostics

Visual Inspection

- By expert meteorologist
- Compares new AMV products with operational ones
- Applies visualisation & analysis tool
- Very useful as sanity check

Methods of Algorithm Evaluation

- Visual inspection
- Collocations with other observations, e.g., radiosonde
- Comparison against short-term forecasts or analyses from NWP
- Forecast impact trials in NWP
- Advanced observation diagnostics

Collocations with other Observations

Follows established criteria (4th IWW, 1998, Saanenmoser, Switzerland):

Satellite wind and its companion radiosonde observation:

•must differ by no more than 1.5 hours

•must be within the distance of 150 km

•the height of the satellite wind and the height of the radiosonde observation must be within 25 hPa;

Evaluation in terms of (3rd IWW, 1996, Ascona, Switzerland):

•Sample size, Mean Speed, Speed Bias, Mean Vector Difference, Root-mean-square Vector Difference

•All / Low / Medium / High level winds

•Global, NH, TR, SH

•Spectral channel

Collocations with other Observations

Pros:

- Evaluation against independent, unbiased observations.

Cons:

•Very limited geographical and temporal sampling:

  - few radiosondes over sea
  - no diurnal signal

•A month or more of data required to obtain a sufficient number of collocations

•Collocation errors

Methods of Algorithm Evaluation

- Visual inspection
- Collocations with other observations, e.g., radiosonde
- Comparison against short-term forecasts or analyses from NWP
- Forecast impact trials in NWP
- Advanced observation diagnostics

Comparisons against forecasts or analyses

Evaluation: Same as with radiosondes.

Pros:

•Comparison available for every AMV: complete geographical and better temporal sampling possible

•A relatively short period of data (~ 2-3 weeks) usually yields meaningful statistics

•Stratification in several ways possible (e.g., height assignment method, spectral channel, etc.), while still maintaining reasonable sample sizes

Cons:

•Results may be affected by NWP model biases or errors

•Results may be affected by observational biases (AMV or other) affecting NWP analyses / short-term forecasts

Methods of Algorithm Evaluation

- Visual inspection
- Collocations with other observations, e.g., radiosonde
- Comparison against short-term forecasts or analyses from NWP
- **Forecast impact trials in NWP**
- Advanced observation diagnostics

Forecast Impact Trials in NWP

Evaluation in terms of forecast scores.

Pros:

•Makes good headlines – everyone wants to show a positive forecast impact.

Cons:

•Long trial period needed (≥ ~ 3 months) for robust results, as changes to existing AMVs are usually small compared to the unchanged rest of the observing system, the assimilation system, and the forecast model.

•Forecast scores are often difficult to interpret – a negative forecast impact is difficult to trace back to deficiencies in the AMV data or their usage.

•Only evaluates the sample of AMVs used in the assimilation system.

Methods of Algorithm Evaluation

- Visual inspection
- Collocations with other observations, e.g., radiosonde
- Comparison against short-term forecasts or analyses from NWP
- Forecast impact trials in NWP
- Advanced observation diagnostics

Advanced Observation Diagnostics

Aimed at estimating forecast sensitivity to particular observations (e.g., Cardinali 2009)

Pros:

•Aims to characterise forecast impact, i.e. the "headline item", for each observation type.

•Statistics can be stratified by various aspects, similar to FG/AN departure statistics.

Cons:

•Characterises impact on 24 h forecast only (i.e. not longer range) - the choice of verifying analysis can have a large impact on the apparent forecast score at this range.

•Significance tests not yet established.

•New method – we still need to learn more how to use it for diagnostics (and what sample sizes are required)

AMV Algorithm Evaluation at EUMETSAT

AMV Algorithm Evaluation at EUMETSAT

Source of algorithm changes:

- User requirement
- New scientific insights
- Anomaly in processing

AMV Algorithm Evaluation at EUMETSAT

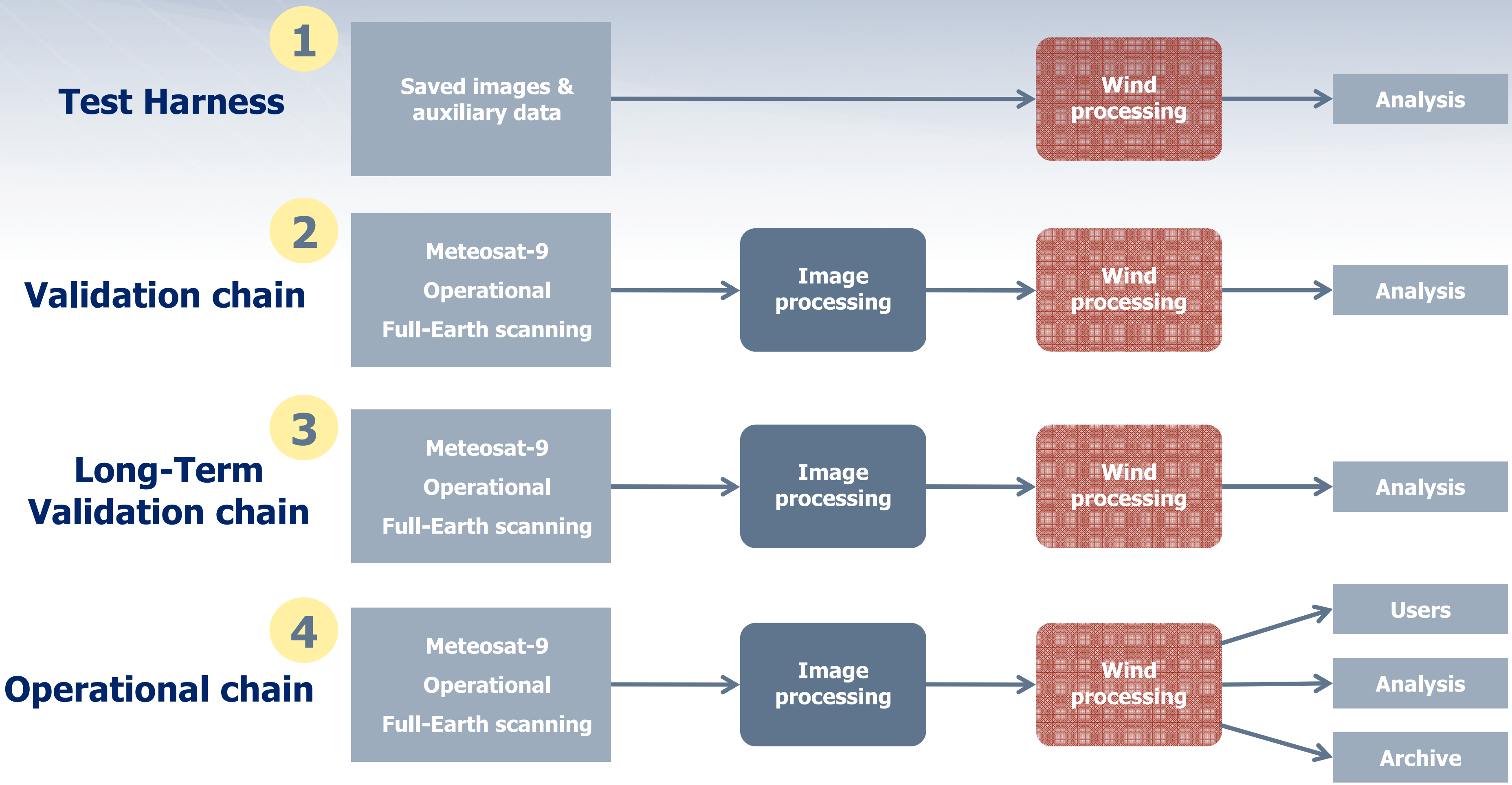
Source of algorithm changes:

- User requirement
- New scientific insights
- Anomaly in processing

Systems involved in evaluation process (operations side):

- Algorithm Test Harness
- Validation Processing Chain
- Long-Term Validation Processing Chain
- *Operational Processing Chain*

Current Working Practice



Algorithm Test Harness

Algorithm test harness

- Basic testing
- Sanity checks
- Not suitable for longer tests

Validation Processing Chain

Validation chain

2

Meteosat-9
Operational
Full-Earth scanning → Image processing → Wind processing → Analysis

- Stability testing
- Typically 1 – 3 weeks testing period
- Limited comparison against operational winds

Long-Term Validation Chain

3

Long-term validation chain

- Standard evaluation (collocations, CGMS statistics)
- Data sent to ECMWF / UKMO for NWP assessment
- Typically 1 – 2 months testing period
- Comparison against operational winds
- Not suitable for quick tests

Operational Chain

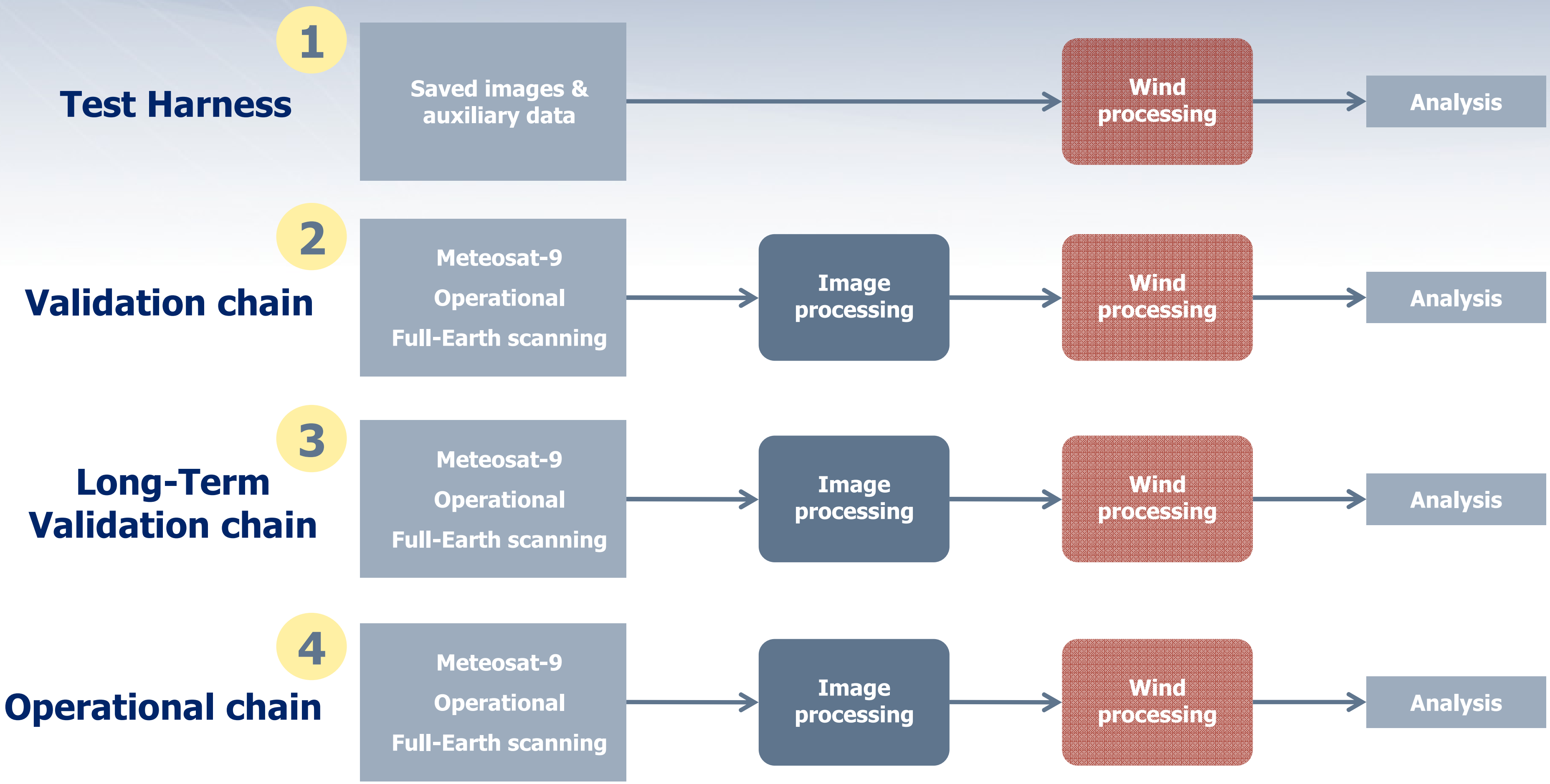
4

Operational chain

Meteosat-9 Operational Full-Earth scanning → Image processing → Wind processing → Analysis

- After ‘go ahead’ from ECMWF / UK Met Office
- Product Validation Report
- User notification 1-2 months in advance

Current Working Practice



Current Working Practice

Disadvantages:

- Only 1 processing chain for serious evaluation
- Limited availability, due to other tests:
  - Radiative Transfer Model
  - Scenes Analysis
  - Cloud Analysis
- Needs careful planning
- Near real time nature of processing: 1 hour worth of data needs 1 hour processing time

New Approach: Reprocessing Chain

New Approach: Reprocessing Chain

- Aimed at larger test runs (typically > 1 month)
- Direct comparison against saved wind data
- Repetition of test runs
- Fast: 1 month worth of data can be processed in 1-2 days

New Approach: Reprocessing Chain

- Aimed at larger test runs (typically > 1 month)
- Direct comparison against saved wind data
- Repetition of test runs
- Fast: 1 month worth of data can be processed in 1-2 days
- Will be introduced in the course of 2010

ECMWF Met Office EUMETSAT

Evaluation of AMV Algorithm Changes

The NWP Perspective

ECMWF

Met Office

EUMETSAT

Evaluation of AMV Changes, NWP Perspective (1)

What evaluation of AMV algorithm changes should be performed on the winds producers side ?

- What data to evaluate against ?
  - Other observations (e.g., radiosondes, aircraft); but sampling limited.
  - Short-term forecasts from NWP:
    - Allows comparison for every AMV, detailed investigations
    - Currently often underused on the winds producers side
  - All of the above with the same test/development dataset
    - Should be encouraged
    - How to implement it ?
- What is the recommended length of in-house test/development datasets ?

ECMWF Met Office EUMETSAT

Evaluation of AMV Changes, NWP Perspective (2)

What evaluation of AMV algorithm changes should be performed on the winds producers side ? (continued)

•How to better interpret and evaluate the impact on quality control ?

- Evaluation of AMVs before auto-editing / RFF for NESDIS / CIMSS winds
- Are more winds with a higher QI really an improvement ?
- QI characteristics vary between producers. Is it feasible to unify the QI implementation across AMV producers ?

•What should be achieved before moving on to evaluation in NWP ?

- An overall neutral or positive impact on radiosonde and short-term forecast comparison statistics is required from an NWP point of view

ECMWF Met Office EUMETSAT

Evaluation of AMV Changes, NWP Perspective (3)

How do we facilitate meaningful evaluation of AMV algorithm changes in NWP ?

•Frequency of changes ? Time-scales allowed for evaluation at NWP centres ?

- It is not feasible for NWP centres to evaluate many small changes with marginal impact. Bundle changes together ?

•Length of test datasets ? What is feasible for AMV producers ?

- 2-3 weeks or more required for meaningful statistics against short-term forecasts
- ~ 3 months or more required for meaningful forecast impact evaluation (ideally for different seasons)

ECMWF Met Office EUMETSAT

Evaluation of AMV Changes, NWP Perspective (4)

How do we facilitate meaningful evaluation of AMV algorithm changes in NWP ? (continued)

- Scope of evaluation, also given length of test datasets ?
  - Evaluation against short-term forecasts most important as most robust (can include data that is not assimilated)
  - Forecast impact evaluation useful, but should not be overemphasised. Effect of any AMV algorithm change likely to be small in comparison to the unchanged rest of the assimilation / observing system
  - Potential of diagnostics such as forecast sensitivity to observations.
- Different requirements for global and meso-scale users ?
- What kind of change documentation from winds producers is useful for NWP ?

Plenary discussion on process for testing and implementing operational changes

EUM/
Issue <No.>
<Date>