THE ART OF EVALUATING AMV ALGORITHM CHANGES

Arthur de Smet [1], Iliana Genkova [2], Mary Forsythe [3], Niels Bormann [2]

[1] EUMETSAT, Eumetsat-Allee 1, D-64295 Darmstadt, Germany
[2] Iliana Genkova, ECMWF, Shinfield Park, Reading, UK
[3] Met Office, Fitzroy Road, Exeter, UK

ABSTRACT

This paper describes the current practice of evaluating changes in Atmospheric Motion Vector (AMV) algorithms, as well as the corresponding difficulties. It gives a detailed overview of the various steps at both the side of the producer of AMVs and at the user's side. The overview describes the practice at EUMETSAT, as the AMV producer, and at ECMWF and the UK Met Office, as the main users of the EUMETSAT AMV products. Two questions feature in this paper: [1] How to optimise the dialogue between the producer of the AMVs and the main users, i.e. the NWP centres [2] How to ensure the implementation of safe algorithm modifications, without hampering progress? These questions were raised at the 10th International Winds Workshop and were the starting point of a lively discussion. This paper includes the main points of the discussion.

1. INTRODUCTION

The evaluation of modifications in Atmospheric Motion Vector (AMV) algorithms has become less straight-forward, as the algorithms are nowadays more advanced and sophisticated, and the user requirements have tightened.
The whole process of AMV evaluation requires direct involvement of both the producer of the AMV product and the users, the most important of which are the NWP centres. This implies a thorough verification of the algorithm changes by both parties. The whole process usually takes at least three months in total and requires good levels of communication between both. Only after approval by the producer and the NWP centre, will the modified AMV algorithms become operational.

The structure of this paper is as follows. Section 2 describes the various methods of algorithm evaluation that are currently in use. Section 3 contains an overview of the algorithm testing facilities used by EUMETSAT. The NWP perspective of the evaluation process, which was the introduction to the ensuing discussion at the 10th International Winds Workshop, is described in Section 4. The results of the discussion are listed in the final section.

2. METHODS OF ALGORITHM EVALUATION

Figure 1 visualises the general framework for AMV algorithm testing, representing the working practice at EUMETSAT on the one hand, as producer of AMV products, and ECMWF and the UK Met Office on the other hand, as users of the AMV data.
There are five distinct methods of evaluation, which will be performed by the producer and the user and will eventually lead to a go or no-go decision. The following sub-sections will discuss these methods in detail.

Visual inspection

A meteorological expert at EUMETSAT, with a profound knowledge of atmospheric motions, will subject the new AMV products to a visual inspection. This implies a comparison of the new AMVs to the operational ones, using a visualisation and analysis tool. This method has proved to be very useful as a sanity check.
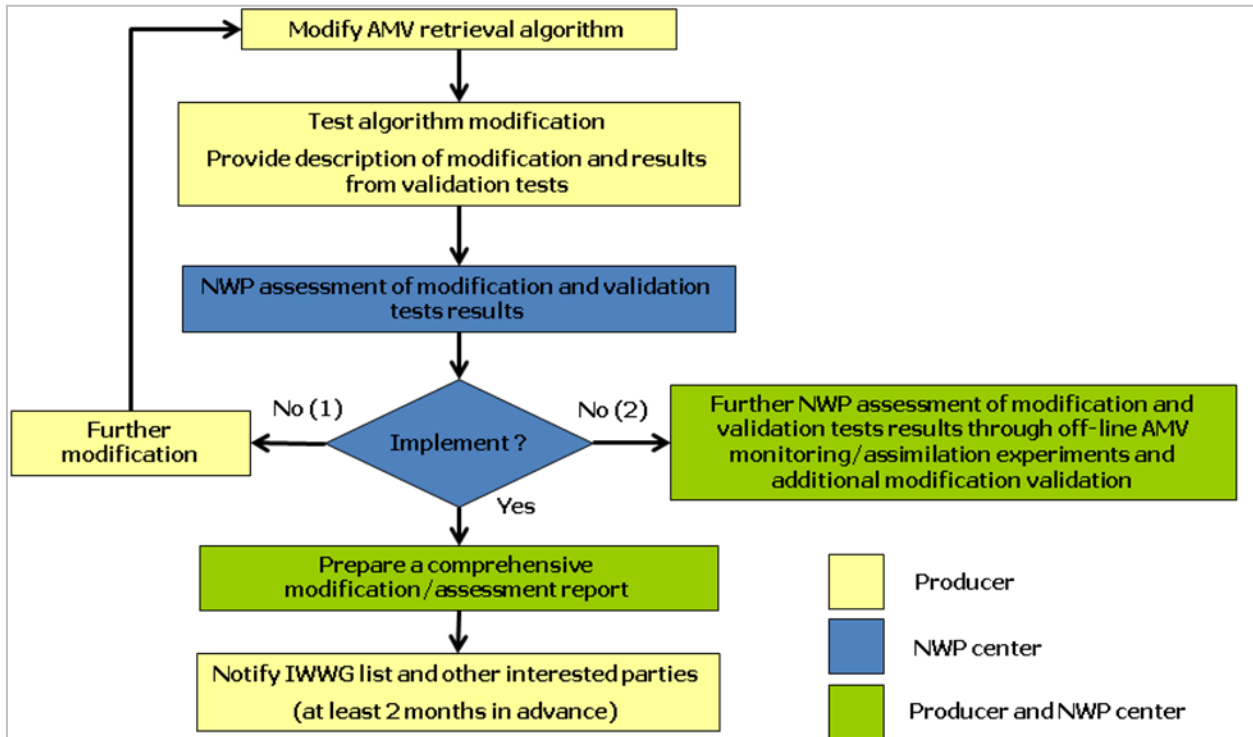

Figure 1   **General framework for AMV algorithm testing.**

Collocations with other observations

Most AMV producers collocate their winds with radiosonde observations and publish the associated statistics. Other sources of observations are air craft and wind profilers. The collocations follow a set of criteria that was established at the 4th International Winds Workshop, in 1998 at Saanenmoser, Switzerland. The criteria apply to each pair of satellite wind and radiosonde observation and prescribe that:
- They must differ by no more than 90 minutes,
- They must be within a horizontal distance of 150 km to each other,
- Their heights must differ by no more than 25 hPa.

Individual collocations may not always be useful, because one can expect differences due to the fact that radiosonde observations are more representative of the local wind field, whereas the AMVs represent larger scale winds (the extent of which depends on the AMV processing target sizes). It is the longer term statistics that are more useful and may reveal biases. Satellite wind producers nowadays derive the following statistics:
- Number of collocations,
- Mean AMV speed,
- Mean radiosonde wind speed,
- Speed bias,
- Mean vector difference,
- Root-mean-square vector difference.

A stratification of this information is possible along the following lines:
- Atmospheric level: low – medium – high – all levels;
- Geographic area: Northern hemisphere – Southern hemisphere – Tropics – global;
- Spectral channel: e.g., infrared channels, water vapour channels, visible channels.

The main advantage of collocating AMVs against radiosonde observations is that it represents an evaluation against independent, unbiased observations.

A major disadvantage is the limited geographical coverage of the radiosonde observations, which shows an uneven distribution over land and very few observations over sea. The temporal sampling of the radiosondes is also limited, which implies that a possible diurnal signal will not reveal itself in the statistical record.

A limitation that is important in the context of algorithm evaluation is the fact that at least one or two months of collocation statistics are required before one can make a valid judgement about the algorithm change.

Comparison against short-term forecasts

Atmospheric profiles from short-term NWP forecasts or analyses can be collocated with satellite winds in the same way as radiosonde observations. The main advantage of this is the complete geographical coverage and the better temporal sampling of the forecast profiles. This enables a comparison for every AMV, with the consequence that a relatively short period (two or three weeks) is sufficient to obtain meaningful statistics. A stratification of the statistics is possible in several ways (e.g., height assignment method, spectral channel, etc.), while still maintaining reasonable sample sizes.

Forecast model data are not free from biases and errors though, and these will have an impact on the collocation results. A potential error source is the biases in the observational data (AMV data and others) that are used by the NWP forecasts and analyses.

Forecast impact trials

An established method is to assimilate the new AMVs in the forecast model and perform a so-called forecast impact study. The evaluation then follows directly from the forecast scores. A distinct advantage of this method is that it makes good headlines, at least when there is a discernible positive impact. But this method requires a long trial period of at least three months for robust results, as the changes to the AMVs will usually be small compared to the other, unchanged parts of the observing system, as well as to the assimilation system and the forecast model.

One difficulty is the interpretation of the forecast scores: a negative forecast impact is not easy to trace back to deficiencies in the AMV data or their usage. Moreover, this method evaluates only the sample of AMVs that is actually used in the assimilation system, which is typically only five percent of the full set of AMV data.

Advanced observation diagnostics

New methods to evaluate the impact of observations on NWP forecasts are being developed all the time. One of the latest methods aims at estimating the forecast sensitivity to particular observations (Cardinali, 2009). This enables the characterisation of the forecast impact of each observation type.

Since this is a new method there is only limited experience with using it as a diagnostics tool. The current methodology is limited to 24 hour forecasts only.

3. ALGORITHM EVALUATION AT EUMETSAT

There is a constant need to improve the algorithms that lie at the heart of the AMV derivation. The origin of this can be a new scientific insight, or an anomaly that was found in one of the processing steps. Whatever the reason, changing an algorithm will always require a thorough evaluation. This is not just necessary to convince oneself that the new or modified algorithm works as expected, but also to identify potential side-effects. If the modification leads to changes in one or more essential output parameters, like the height assignment for example, the users of the end products must be involved in the evaluation cycle.

EUMETSAT recognises three levels of algorithm evaluation in the operational environment and, directly corresponding to these, three validation entities. The following sub-sections describe these.

Basic testing

The *Algorithm Test Harness* is a test tool available on the development system and allows basic testing. It relies on saved image data. One of benefits of this is that one can rerun a test case several times. This makes it the tool of choice when assessing, for example, the impact of modifying a set-up parameter. It is also an ideal tool for debugging and for performing sanity checks, like array bound checking.

It is not suitable for test runs that span more than a few hours of satellite image data.

Basic validation

The next step in the sequence of algorithm validation is the *Validation Processing Chain*. This supports near real-time processing and mimics the *Operational Processing Chain*. It is relatively easy to use and allows a comparison of the AMV products against the operational AMV products. A test period of one to three weeks is usually sufficient to assess the stability and correctness of the software changes.

Since this is the main validation entity for all meteorological algorithms, as well as for the image processing software, one can never assume that any changes in the AMV parameters are due to the modifications in the AMV algorithms only. The images and the basic products on which the AMV processing relies (scenes and cloud products) may all differ from those on the operational processing chain. This limits the value of a direct comparison of the AMV products against the operational ones.

Long term validation

The *Long-Term Validation Processing Chain* addresses this last point. It is under strict configuration control and works with operational images only. It is the ideal tool to assess the impact of algorithm changes on the meteorological products (AMV and others). The standard practice is to plan any algorithm changes in advance and to prevent the modification of more than one meteorological product at a time, unless they are completely independent. This allows a direct comparison between the modified product and the operational one.

This validation entity has the same functional components as the operational processing chain, including AMV collocations and CGMS statistics. This enables longer term analysis.

If an NWP assessment of the algorithm changes is necessary, the products from this processing chain will be transferred to ECMWF and the UK Met Office. The typical test period length for such an assessment is one or two months.

When the outcome of the NWP assessment is positive and a decision is taken to go live with the new algorithm, the users will be notified one to two months before the actual change date.

The complete evaluation procedure on the *Long-Term Validation Processing Chain*, including a description of the modifications in the algorithms as well as the test results, will be published in a so-called Product Validation Report. This will include the results from the NWP assessment.
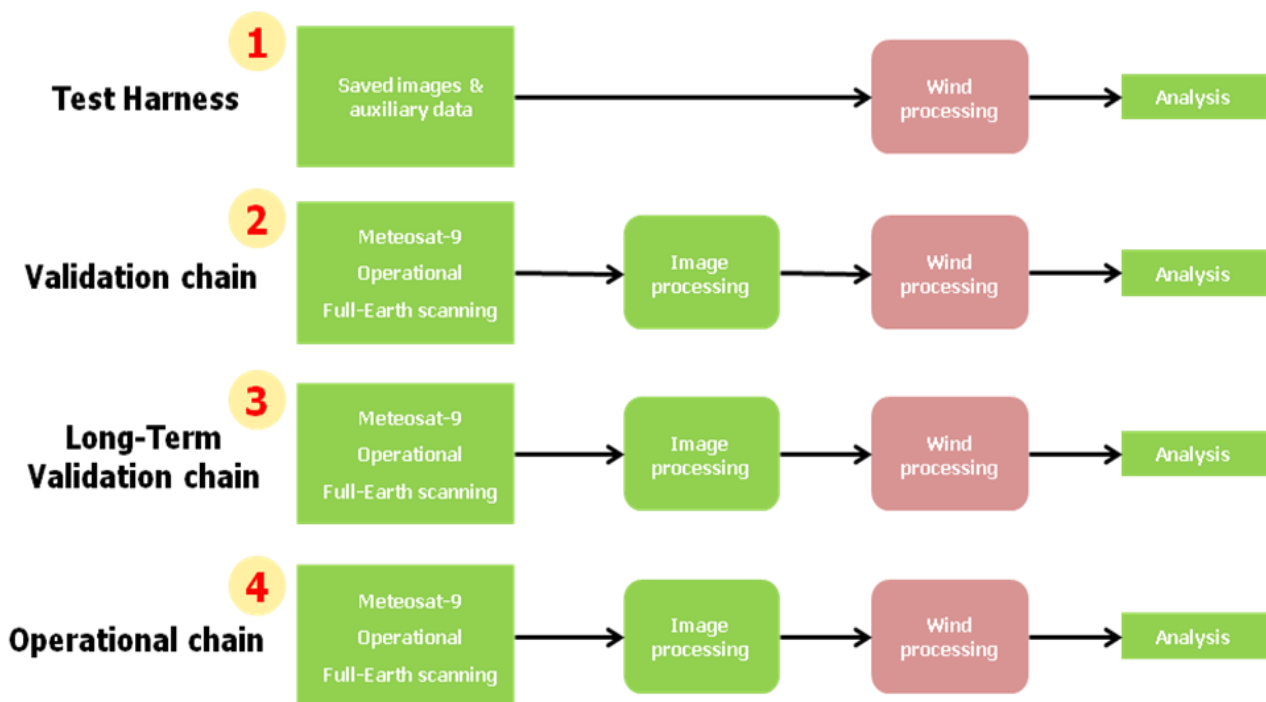

Figure 2   Sequence of algorithm validation at EUMETSAT

It is clear that the current practice of algorithm evaluation at EUMETSAT is not ideal. There is only one processing chain for serious testing. But the availability of this chain for assessing the impact of a change in one particular product is limited, as other products use the same processing chain for evaluation purposes. Another drawback is the near real time nature of the image and product processing: one hour worth of data needs one hour of processing.

Reprocessing chain
These limitations have been recognised for a long time. The answer to it will be introduced in the course of 2010 and is called the *Reprocessing Chain*. It has a structure that bears no similarity with the operational and validation processing chains. It relies on archived images, rather than on near real time imagery. This has two distinct advantages. Firstly, any test run can be repeated many times. This enables the test conductor to assess the impact of gradually changing a tuning parameter, or to divide a complex change into a number of small modifications, the effects of which are easier to assess. Secondly, no longer being limited by near real time processing, it is much faster. The current expectation is that one month worth of data can be processed within two days.

4. DIALOGUE BETWEEN PRODUCER AND USER

A clear dialogue between the producer of satellite winds and the users, in the first place the NWP centres, is an important condition to the successful improvement of AMV algorithms.
There is room for improvement here. There is currently no formal procedure that describes the various steps in the evaluation process. This may lead to expectations on one side that are not recognised by the other side.
The AMV producer feels sometimes overwhelmed by the results of the forecast impact study: the report sent by the NWP centre usually contains many plots, with ambiguous signals.
The NWP centres, on the other hand, feel that the producers of satellite winds could put more effort in the comparison against short-term forecasts.

5. NWP PERSPECTIVE

AMV producers evaluate their winds routinely against independent observations, like radiosonde and aircraft data. The limited sampling, both in space and time, poses a serious problem.
The NWP community believes that collocations against short-term forecasts are currently underused on the AMV producer side. They would encourage an increased effort to include this type of collocations. An advantage is that they allow a comparison for every AMV.
The assessment of the impact of AMV changes on the forecasts is not trivial. It involves the analysis of several months' worth of data and an impact study, all of which are not routine operations.
It is therefore suggested by ECMWF and the UK Met Office that an NWP evaluation can only be initiated after the AMV producer has shown an overall neutral or positive impact on radiosonde collocations and short-term forecast comparison statistics.

Considering the efforts of forecast impact studies it is not feasible for NWP centres to evaluate many small changes with a marginal impact. It is more efficient to bundle changes together.

There is a tendency to attach much importance to the results of forecast impact studies. One should be careful with this though. The effect of any AMV algorithm change is likely to be small in comparison to the unchanged part of the assimilation and observing system. The relative importance of satellite winds has decreased over the last two decades, due to the introduction of other kinds of observations, many of which originate from satellite instruments.

6. DISCUSSION

One of the important factors to consider when evaluating AMV algorithm modifications is the spatial representativeness of the validating data set, as well as the related scaling issues associated with a forecast impact study. In this regard, inter-comparisons of AMVs with RAOBs and short-term forecasts are two complementary approaches. NWP centres strongly recommend that the producers conduct inter-comparisons and provide results from analysis of the first guess (or forecast) departures, in terms of biases and standard deviations.

Ideally, if the first guess and analysis departures are improved, then a forecast impact study will be set. By definition, the AMV Quality Indicator (QI) includes a forecast consistency component. EUMETSAT reported using it internally (in house) for monitoring the quality of new AMV data sets, however it was recognised that such analysis have not been distributed to users. Producers were encouraged to consider outputting this QI component and exploring it as a mean to evaluate algorithm modifications.

NWP centres strongly recommend experiments to be longer than three months and when possible data from two seasons to be included. At CMA, only FY2-C AMVs are validated against RAOBs, because their extraction times coincide with RAOBs launch times. FY2-D AMVs assume the same quality without further validation. There are also plans to compare FY2 winds to AMVs from JMA.

Two novel diagnostics tools – Forecast and Analysis Sensitivity to observations, have been developed at ECMWF (Cardinali, 2009). They allow for investigating the impact of the assimilated observations by narrowing down to spectral channel, pressure range, etc. The Analysis Sensitivity tool will soon be in operational use at ECMWF.

Auxiliary cloud information at the time of AMV extraction was recommended as a valuable means to understanding the quality of the product. However, at this time only EUMETSAT has an operational cloud analysis product generated prior to the AMV extraction. After the launch of the GOES-R satellite, the same could be done at NESDIS as well. The above described modification validation chain (see Section 3) was accepted by JMA for near future implementation, however they will not have access to cloud information.

When a forecast impact study is conducted, the impact of the new AMV is assessed by verification against analysis. The verification can be done against:
- An operational analysis (which does not include the tested data set),
- The experiment's own analysis (which includes the tested data), or
- Observations.

The choice of verification analysis sometimes determines the magnitude of the impact the tested data has on the forecast, however, very seldom the impact changes from positive to negative or vice versa. When the impacts are very weak, all three verification approaches are applied for best evaluation of the impact study. Finally, producers inquired whether the same assessment criteria are used for any algorithm change. While in general the evaluation measures are the same – first guess and analysis departures, mean wind analysis maps, forecast error maps and scatter plots – it was agreed that depending on the specific algorithm change or geographical coverage of the AMV set, more tailored in depth investigations would benefit the quality of the AMVs.

7. BIBLIOGRAPHIC REFERENCES

Cardinali, C., (2009) Monitoring the observation impact on the short-range forecast. Quarterly Journal of the Royal Meteorological Society, **135**, Issue 638, pp 239-250