SRNWP - Verification
Programme proposal

Prepared by:	SRNWP Expert Team
Purpose:	To Present the Draft SRNWP Verification Optional Programme
Action proposed:	Council to consider, comment and approve
Majority required:	Simple majority and participation of at least a third of EUMETNET's Members

Council requested a proposal for SRNWP Verification. This proposal has been prepared by the SRNWP Expert Team and submitted by the Met Office.

Proposal for EUMETNET/SRNWP Programme for

Verification and Model Intercomparison for European models

by

SRNWP Expert Team on diagnostics, validation and verification

Submitted to the EUMETNET Council by the Met Office

Table of content

1. Background

At the meeting "A Vision for Numerical Weather Prediction in Europe", jointly organised by the Met Office and the ECMWF, which took place at the ECMWF the 15th-17th of March 2006, several recommendations have been made at the request of the EUMETNET Council in order to increase cooperation and efficiency in NWP in Europe (see EUMETNET Document EMN/C27/Doc12).
Under the theme "Improved Framework for Collaboration", one of the recommendations reads: "Initiate the reorganisation of the SRNWP Programme" in order to enlarge its scope and strengthen its activity. This recommendation also proposes "the definition of specific projects in a similar way to the ECSN Programme".
The same Document also stipulates the following: "SRNWP Programme will define at its next general meeting a draft programme proposal for spring 2007".
The SRNWP general meeting took place the12th of October 2006 in Zurich and three specific projects were defined (See the minutes of the meeting under http://srnwp.cscs.ch/Annual_Meetings/2006/Report2006.htm).

One of these three specific projects was the ***development of a common verification package and the realization of an operational model intercomparison***. After consideration by a redaction committee, the proposal was not submitted to the 30th EUMETNET Council in April 2007. It was felt by the redaction committee that the scope of the proposal needed to be reconsidered. Whilst construction of a new common verification package could lead to benefits in ensuring consistent methods at all centres and in allowing access to verification methods by those NMSs without the resources to develop their own packages, it was felt that this was not required to deliver the intercomparison. Many centres had developed their own packages in recent years and were unlikely and unwilling to devote time and resources to repeating the development for a common package. No candidate NMS indicated its willingness to take the responsibility for developing the package. The original proposals were unlikely to deliver the common intercomparison results in the near future. The committee also felt that the challenge of verifying higher resolution models needed more research and evaluation before including newer techniques in a common package. The original proposal also included setting up a data hub of non-GTS data. This depends upon identifying a NMS willing to host this. The redaction committee concluded that this was a more ambitious aim that would be better addressed by establishing a separate programme. Instead the responsible member and the Expert team would seek to ensure that all information regarding the sources of possible data for verification was disseminated and its use encouraged. This would be especially applicable to the verification of higher resolution models.

This is a revised proposal from the SRNWP Expert Team on diagnostics, validation and verification, in which the objectives are staged. It concentrates on quick wins so that more information about the relative strengths and weaknesses of the 4 consortia models can identified earlier to inform the future development of the models. The stages are:

1. Intercomparison of 4 operational consortia reference models using existing verification systems at one or more centres, up to a maximum of 4. Collection and dissemination of results.
2. Incorporate more operational models in the comparison including high resolution models
3. Exchange new methods/code developed for verification of high resolution models
4. Encourage use of and access to radar and other non-conventional non-GTS observations in verification

Based on these activities the following outputs are going to be delivered:

- **Deliverable D1: Operational verification comparison of one version of each of the 4 regional models of Europe (available for all the participating members).**
- **Deliverable D2: Additional intercomparison of other versions of the consortia models including high resolution models**
- **Deliverable D3: Inventory and recommendations of “new” scale selective verification methods.**
- **Deliverable D4: Catalogue of sources of non-GTS data.**

Main Objective of this project: Realization of an operational model intercomparison of the 4 consortia models

Thanks to the existence of Consortia, we have in Europe only four basic operational regional models: ALADIN, HIRLAM, COSMO and the limited area version of the Unified Model. The first aim is to deliver a meaningful intercomparison of operational deterministic forecasts from a “reference version” of each of the 4 consortia models. Although there are many different configurations and operational models in Europe, they are all based upon one of these basic regional modelling systems. Rather than seek a comprehensive verification comparison of all possible operational limited area models, which would entail identifying several verification areas common to these, we seek to set a baseline based on the reference regional versions with the largest domains. Other operational models could be added later in stage 2. A thorough comparison of our four regional models in order to find out their respective strengths and weaknesses would foster a general improvement of our models (they all have weaknesses) by re-design or replacement of some parts.

The operational exchanges of WMO CBS scores of global models are a good way for a national Weather Service active on the global scale to monitor the quality of its own model in comparison to the others. It allows it to see what comes from its own modifications and what comes from a more predictable atmosphere during a given period.

For limited area models, score exchanges are not systematically organized because the way scores are computed in the different NMS is not unique and because the simulation domains are not the same. One exception is provided by the Meteorological Office, which runs a routine comparison over the British Isles of the operational precipitation forecasts computed by the four major European regional models. The details are given in Appendix 1. A quick win in verification comparison would be to extend this to other parameters than precipitation.

The four regional model forecasts could be verified by one or more NMS using their current verification systems. This is in the same spirit as the WMO CBS exchange of global scores, except that forecasts are exchanged and scores applied consistently by each verification NMS. The exchange format for the Met Office comparison is currently GRIB1. If the interoperability recommends GRIB2 this could be adopted.

Most NMS's today have their own verification package, usually calculating the same standard scores for similar quantities. Whilst ideally we might wish to construct a new common verification package for adoption by all participants so that different implementations and choices are minimised, it is not a pre-requisite for meaningful intercomparison. The WMO CBS verification scheme for evaluating and comparing Global forecast models has been successfully followed for a number of years. This defines the parameters and common verification area but allows each centre to perform its own verification using its own package. In this spirit, we propose that the four regional model forecasts could be verified by one or more NMS using their current verification systems. The forecasts from a representative reference version of the model of each consortium are verified by the centre responsible for that version. In addition the forecasts from the other three centres will be exchanged and verified using that centre's system. This would ensure that the same observations and quality control are used for all the forecasts *at each verifying centre* although there would still be differences between centres. Thus, if the verification is independently done at all centres, 4 models will be verified at each of 4 centres using their existing verification packages. All the results will be exchanged amongst the 4 participants. One centre will coordinate the amalgamation and display of the results via a password-protected web-page .A consensus of the 4 verification packages can also be derived. If fewer centres choose to perform the verification, the comparison will still be valid but less will be known about the influence of the observational data selection on the verification.

The use of other model forecasts within different forecasting display and verification systems has been slowed down in the past by the lack of interoperability between our models. The SRNWP "Interoperability between the European Models" project is addressing this. Whilst this verification proposal should not take any decisions in conflict with the interoperability recommendations, whilst awaiting that project to define exchange formats, parameters and domains, interim formats for the exchange of the forecast data for verification could be used. For example the current European Model precipitation intercomparison performed by the Met Office uses GRIB1 data format. . If the interoperability recommends GRIB2 this would be adopted.

2. Technical details of an operational model intercomparison (Deliverable D1)

2.1 Models to be compared

The verification and the comparison are principally open to all the operational versions of the ALADIN, HIRLAM, COSMO and Unified Model of the Participating Members.
In this first step, the model intercomparison will be limited to:

- The North Atlantic - Europe (NAE) version of the Unified Model run by the Met Office
- The HIRLAM reference version, as run by the Finnish Meteorological Institute
- The ALADIN-France model run by Meteo-France and possibly an ALADIN-LACE reference run so that ALADIN products may be compared over a larger domain
- The European area version of the COSMO model as run by Deutscher Wetterdienst (DWD).

Only the 00 UTC forecasts will be verified. The forecast range will be 48 hours. The 12 UTC forecasts may be added later.

2.2 Forecast delivery

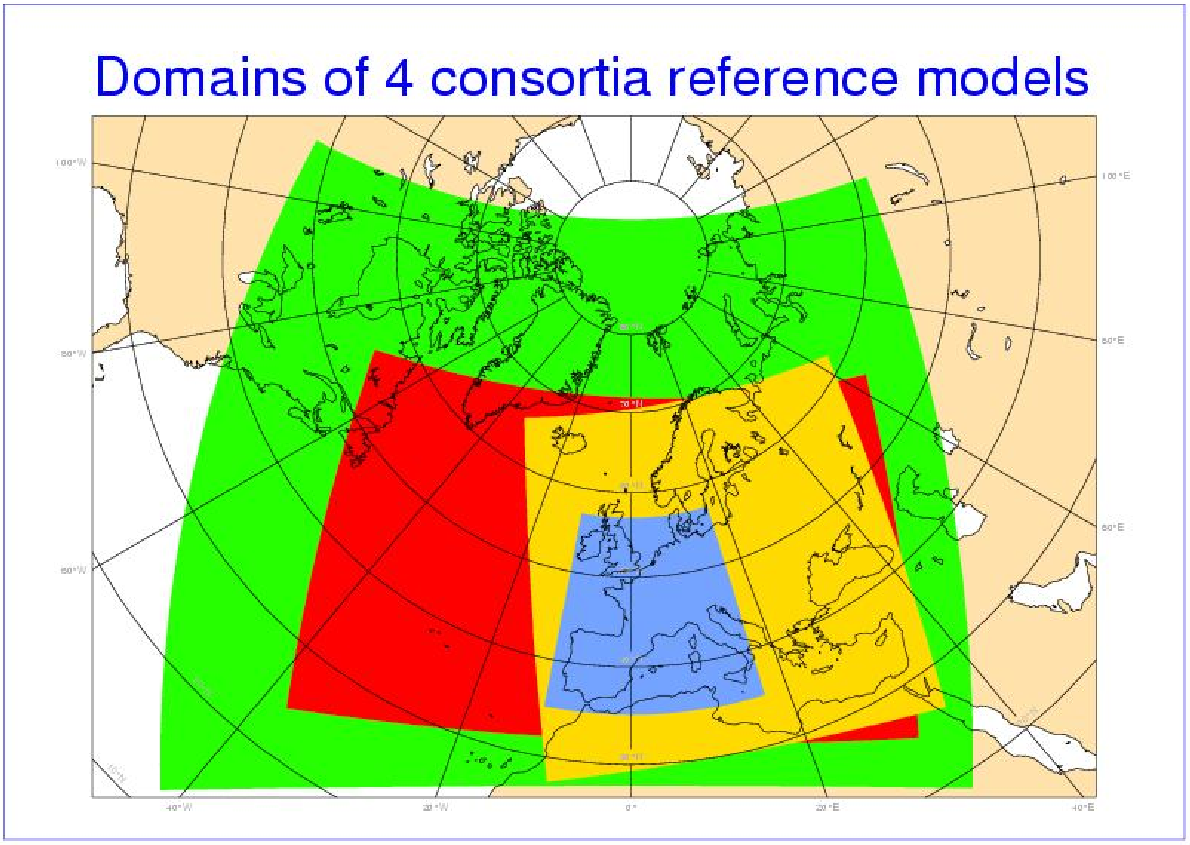
The fields of the parameters to be verified will be sent to the Responsible Member within 48-hour of the nominal data time. This is to allow the verification tasks to be included within operational schedules.

Forecast fields of the parameters to be verified will eventually be delivered in the format of the common model output that the EUMETNET Interoperability Project will define. The interface programme will be developed by the Project Interoperability. ***However in the interim it is proposed to use GRIB1 format.***

The model outputs for all the parameters to be verified (precipitation excepted) are requested from T+0 to T+48h, at 6h intervals. For precipitation, accumulated precipitation for the 8 time intervals (+0/+6), ... , (+42/+48) are needed.

2.3 Verification domain

For model intercomparison, verification will be made over the largest possible common domain of the participating models, excluding lateral boundary and extension zones.



2.4 Types of observation to be used

The forecasts will be compared against SYNOP station reports. In addition, radar estimates of surface precipitation will be used as an alternative or complement to SYNOP station reports of precipitation totals. The use of the radar composites developed for operational production at the Met Office in the frame of the OPERA Programme will be the preferred choice.

2.5 Verification methods

Each verifying centre will use its own package. It will be a mandatory requirement that the methods and observations used, including the quality control employed, will be documented and provided to all participating centres. As far as is practical, differences between methods and approaches will be minimised by agreement between the centres. Where there is a reasonable choice between alternative approaches comparison of the verification of the same model forecast by the different centres may yield important insights in the strengths of each choice.

The model forecasts (apart from precipitation) will be projected onto the synoptic stations' locations. This may be nearest grid point or by bilinear interpolation. Height adjustments will also be used. Precipitation forecasts, radar estimates and precipitation reported in SYNOPs will be area-meaned to a common coarsest resolution rotated latitude-longitude grid.

2.6 Scores and forecast variables to be verified

- For the variables mean sea level pressure, temperature and wind speed, the scores to be produced are bias, root mean square error and skill score (with respect to persistence).

- For the norm of the wind vector difference, the scores to be produced are the root mean square error and the skill score (with respect to persistence)

- For all the above variables, the scores will be computed every 6 hours, i.e. for +00, +06, ... , +42, +48.

- The ECMWF high-resolution analyses will be used for persistence in order to allow a fair common reference.

- For precipitation, the scores to be produced are frequency bias, equitable threat score (ETS), log-odds ratio and the Peirce (Hansen-Kuipers) skill score against persistence.

- 6-, 12- and 24-hourly accumulated total precipitation will be verified.

Geographical distributions and time-series, as well as monthly, seasonal and yearly means will be produced for all the scores of all the parameters verified.

It would also be desirable, for the variables mean sea level pressure, temperature, wind speed and norm of the wind vector difference, the monthly means of the bias and root mean square error will be computed at each individual station for the hours +36 and +48.

2.7 Dissemination of results

All the verification results of the model intercomparison will be published on the web site of the Responsible Member under password protection.

Participating Members to this Project will be entitled to receive the password.

2.8 Addition of other models including high resolution models

After stage 1 has been established and the verification results and methods reconciled, it is straightforward to expand the comparison to include more centres' models (**deliverable D2**). This could either be through an expanded exchange of model forecasts between the additional centres and the verifying centres, or, more likely, a comparison done within consortia of their models against the reference model of the 4 consortia. This will not always be possible due to the limited regional domains of some models, as there will be no common area.

3. Exchange new methods/code developed for verification of high resolution models

It is now a well-established fact that the verification of the precipitation forecasts of models with km-scale resolution cannot be adequately performed at that scale with the traditional scores. If it is done, the results are often systematically inferior to the results of models with coarser, even much coarser horizontal resolutions. The development of new verification methods for the precipitation forecasts suitable for km-scale models is today a very active field of research. It is necessary to assess one these new methods that NMS could use for their model version with the highest resolution. It is quite a new field of research. The interpretation of the results of these methods is still the object of discussion and, as for the traditional scores; each of them shows only one characteristic of the precipitation behaviour. At the present it would not be meaningful in this Proposal to dictate to the Responsible Member the method he should use. In Appendix 2, a few hints are given on some of these methods.

The SRNWP expert team on diagnostics, validation and verification will discuss and organise the evaluation of the newer scale-selective methods, and make recommendations (**deliverable D3**). It will also encourage and organise the exchange of methods and code to implement them between consortia. Recent efforts to verify high resolution models are reported in Amodei M. and J. Stein (2008) for ALADIN, Mittermaier (2006) and Roberts (2008) for the Unified Model, Theis et al (2005) for COSMO, and Kok et al, 2008 for HIRLAM.

4. Encouraging the use of non-GTS observation data in verification

Standard observation data like SYNOP, TEMP, METAR, etc are collected in many databases in Europe, not only at each NMS and at the ECMWF, but also by projects as, for example, the EU Project EUROGRID or the EUMETNET Programme "European Climate Assessment & Dataset".

Beside these GTS-data, there are in Europe large amounts of meteorological observations that do not circulate: they come from stations that are not registered by WMO, which often belong to counties or provinces. Their data remain local or inside national borders. The high-density rain gauge networks are the best example. Although they normally belong to National Meteorological Services, no large-scale exchange of their data takes place in Europe. High-resolution observing networks - particularly the rain gauge networks - have been established in the past for climatic purposes: to better know the climate of a region or of a country. Modern high-resolution observing networks made of automatic observing stations serve primarily the knowledge of the present weather but also the climatology.

We are presently witnessing a tremendous increase in spatial resolution of the NWP models: models of some 4 km resolution are already operational; models with resolution between 1 and 3 km are in a pre-operational stage. It is thus easy to understand that

these high-resolution data become very important for the verification of the results of these high-resolution models. However, it is generally very difficult to access these data for stations outside national borders. Some efforts have already been done to collect these data, particularly the high-density precipitation measurements. Examples:
- DWD collects for the needs of the Consortium COSMO the non-GTS precipitation data of Germany, Switzerland, Poland and Northern Italy
- ECMWF collects the non-GTS precipitation data of its Members
- The EUMETNET Programme "European Climate Assessment & Dataset" collect non-GTS precipitation data, but not with a high spatial resolution
- The BALTEX meteorological data centre (***http://www.gkss.de/baltex/data/bmcd.html***) collects, in addition to synop data, data on evaporation, precipitation, radiation, snow depth, soil moisture, and soil temperature for north-western Europe.

Missing in Europe is a hub that would centralised all the non-GTS meteorological data, as we today already have in the frame of EUMETNET a hub for radar data and a hub for wind profiler data.

The absence of such a hub can be well noticed today in Europe by the fact that when a project is submitted, the first foreseen task is often "to establish a data base for high resolution meteorological data". And when the project is accepted, several NMS Directors receive a letter from the project leader asking for non-GTS data (cf. Project ELDAS, Project ENSEMBLES, Programme ECA&D, etc.).

In the original Programme, it was planed to have a data hub for non-GTS meteorological data, but it was not intended to create a new data centre. The idea was to supplement an already existing observation database with the maximum of verified non-GTS observation data. This hub would be an extension of the observation database of a NMS or of the ECMWF. Whilst this may still be a worthwhile aim it is felt that this should be in the context of a separate project on data and quality control for use in both data assimilation and verification. The routine use of such data in verification will progress in tandem with its greater exploitation in data assimilation which is an active area of research and development.

In the context of this programme it is proposed that a catalogue of sources of non-GTS data is established (**deliverable D4**) and published so that participating centres may easily gain greater access to these. Studies in the use of this data will also be encouraged within the development of the high resolution verification methods.

5. Duties of the Responsible Member

The Responsible Member shall

- for the Model Intercomparison
    - organise the exchange of forecasts from the 4 reference models
    - coordinate the participating verification centres
    - verify the reference models using its verification package
    - produce the graphics and compute the consensus verification scores
    - maintain up-to-date the model intercomparison pages on its web site

- store on its computer system all the verification results

- for the use of the non-GTS observing data in verification
  - Establish a catalogue of data sources
  - Publicise verification studies and routine use of such data
  - motivate the NMS to provide their non-GTS observation data for verification use

6. Reporting

The Responsible Member shall send quarterly reports and annual reports to the Programme Manager of the SRNWP Programme reflecting the state of

- the advancement of the model intercomparison
- the availability and exploitation of non-GTS data

The quarterly reports have to insist on the difficulties encountered and make propositions on how these difficulties could be solved or their effects mitigated. Also the Responsible Member will present the progress of the Programme during the annual EWGLAM/SRNWP meetings

7. Start and length of the Programme

The Programme should start the 1st of November 2008 and end the 31st of October 2010. The operational model intercomparison should start as soon as a few scores are ready.

8. Costs per year

Costs of the Responsible Member for
- The set up and operation of the model intercomparison facility, inclusive computing and archiving costs
- the maintenance of the web pages of the model intercomparison results

0.3 Full time equivalent scientist:	€ 30'000.-
Travel expenses of the 0.3 full time equivalent scientist:	€ 2'000.-
Total cost per year:	**€ 32,000.-**

Appendix 1

European Mesoscale model Intercomparison of Precipitation (EMIP)
The current precipitation intercomparison done at the Met Office is described here.

Models and forecasts verified
The current models included in the precipitation verification are:

- The 12km NAE model run at the Met Office
- The 22km reference HIRLAM model run by FMI
- The 9km ALADIN-France model run by Meteo France
- The 7km COSMO-EU model run by Deutscher Wetterdienst

The intercomparison aims to verify the various models against the UK NIMROD radar-rainfall composite over a large part of the UK. The forecasts from 00UTC are verified. Only daily (24h) precipitation is compared. The intercomparison has data from January 2004 to present. At present the comparison is made at the coarsest model resolution, the HIRLAM at 22 km, with finer resolution models area-mean summed to the coarser grid.

At present, four scores derived from contingency tables are displayed: the frequency bias, Equitable Threat Score (ETS), the log-odds ratio and a new experimental score called the Extreme Dependency Score.

Mean scores since January 2004 to present are plotted against precipitation thresholds. The thresholds are 0.125, 0.25, 0.5, 1.0, 2.0, 4.0, 8.0, 12.0, 16.0, 20.0, 24.0, 32.0 and 48.0 mm. Time series of the scores are also produced and monthly contingency tables available to download for users' own use. The results are displayed on the Met Office External web, under password protection.

Appendix 2

New methods of verification for precipitations of km-scale models

There are several methods presently in development or put recently into operation; those recognised by specialists as particularly interesting are:

Fuzzy verification Methods

There are several "fuzzy" approaches to verification of high resolution models, some detailed in papers published in the literature, some outlined in conference and workshop proceedings. Recently a review of the approaches has been made by Ebert (2007), emphasising the framework in which the approaches can be compared and assessed.. "Fuzzy" verification be applied under many different forms, some of them being even very simple. However, there is still no consensus yet on the most useful or appropriate form to be applied. A consensus may emerge from research and development of the methods and activities such as the Intercomparison of methods applied to WRF forecasts (see http://www.ral.ucar.edu/projects/icp/index.html).

The scale intensity method

The scale intensity method of Casati et al. (2004), used by Mittermaier (2006) in the Met Office is also a very powerful method and temporal aggregations of the results are possible.

Structure , Amplitude, location (SAL) method

This is another new interesting approach by Wernli et. al. (2008). It decomposes the errors in three components: structure, amplitude and location - named SAL. It has the advantage that the area can be freely chosen, i.e. river catchment areas.

References

Amodei M. and J. Stein, 2008: Deterministic and fuzzy verification methods for a hierarchy of numerical models. In revision for *Meteorological Applications*

Ebert, E E, 2007. Fuzzy Verification of High Resolution Gridded Forecasts: A review and Proposed Framework. Submitted to *Meteorological Applications*

Casati, B, Ross, G, D B Stephenson, 2004: A new intensity-scale approach for the verification of spatial precipitation forecasts. *Meteorological Applications* **11,** 141-154.

Kok, K. , Wichers Schreur,B. , and D. Vogelezang, 2008: Valuing information from mesoscale forecasts. In revision for *Meteorological applications*.

Mittermaier, M P, 2006: Using an intensity-scale technique to assess the added benefit of high-resolution model precipitation forecasts *Atmos. Sci Letters* **7,** 36-42

Roberts, N. and H. Lean, 2008. Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Monthly Weather Review*, **136,** 78-97.

Theis, S, A. Hense and U. Damrath, 2005: probabilistic precipitation forecasts from a deterministic model: a pragmatic approach. *Meteorological Applications* **12** , 257-268

Wernli, H., P. Paulat, M. Hagen, C. Frei, 2008: SAL - a novel quality measure for the verification of quantitative precipitation forecasts. To appear in *Monthly Weather Review.*