Science plan

S2S sub-project on verification

Third draft: 14 January 2015

Objectives:

• Recommend verification metrics and datasets for assessing forecast quality of S2S forecasts

• Provide guidance for a potential centralized verification effort for comparing forecast quality of different S2S forecast systems, including the comparison of multi-model and individual ensemble systems and consider linkages with users and applications

In order to successfully achieve these objectives the subproject will need to address a number of issues including:

i) identification of current practices in the emerging field of S2S forecast verification and of suitable reference verification datasets (including gridded, satellite, in-situ ground observations as well as model-independent analyses). It is worth noting that precipitation is a key parameter for assessing monsoon activity as well as user-oriented products (e.g. active and break rainfall phases and wet/dry spells) and therefore will require specific guidance on reference verification datasets. Temperature, humidity, wind, evaporation and soil moisture are other variables of great relevance to energy, agriculture and health activities that deserve specific guidance particularly because of the difficulty in obtaining appropriate verification datasets of some of those variables;

ii) identification of quantities or variables to be verified, both user-relevant and those that will facilitate forecast system development, including the temporal and spatial scales associated with each variable;

iii) provision of guidance on minimum hindcast standards for meaningful verification practice (e.g. length of hindcast period and number of ensemble members);

iv) dealing with the challenge of computing forecast probabilities under limited (small) ensemble sizes in the hindcasts;

v) promotion of subseasonal forecasting systems intercomparison verification efforts and comparison of different ensemble forecasting systems configurations to evaluate the benefits of the multi-model ensemble approach;

vi) promotion of the use of novel verification measures adequate for S2S forecasts (e.g. probabilistic measures [Weigel et al. 2008, Weigel and Mason 2011; Ferro 2014] and spatial methods that provide performance information for forecasts with coherent structures [Gilleland et al. 2009] if the spatial resolution of the forecasts is found to allow such detailed spatial verification;

vii) promotion of the use of verification metrics that are meaningful to users, e.g. the use of user-relevant thresholds when verifying probabilistic forecasts;

viii) promotion of the use of bias correction methodologies to handled model systematic errors, especially when comparing forecast quality of models with different spatial resolutions.

viii) provision of guidance for computing verification measures confidence intervals in order to properly address the associated sampling uncertainties and produce statistically meaningful comparisons between forecasting systems, particularly when dealing with auto-correlated (i.e. not fully independent) hindcasts.

ix) promotion of conditional verification practices such as verification conditional on MJO, ENSO, NAO, IOD and SAM phases as well as on particular weather regimes.

x) promotion of the use of appropriate verification measures (e.g. Stephenson et al. 2008; Ferro and Stephenson 2011) when dealing with extreme events (e.g., heat waves, cold snaps, drought, and extended rainy conditions).

Science questions:

• What forecast quality attributes are important when verifying S2S forecasts and how they should be assessed? Which verification methods and forecast attributes are appropriate for reporting S2S forecast quality to users, and which provide added insight into forecast system development and improvement? For example, for the latter, examining spread-skill relationships may help inform on optimal forecast initialization and ensemble generation schemes for subseasonal prediction.
• How should issues of short hindcast period availability and reduced number of ensemble members in hindcasts compared to real-time forecasts be dealt with when constructing probabilistic skill measures?
• How can we best identify windows of forecast opportunity, including assessing the contributions of climate drivers, such as the MJO and ENSO, to S2S forecast skill (e.g. consider skill assessment conditioned on ENSO phases)? This will require adequate samples (including hindcasts) to allow subsetting of the data to provide meaningful verification.
• Which verification methods are most appropriate for the verification of extreme events, particularly given challenges associated with their rarity, small sample sizes and large uncertainties?
• How can we best verify active and break rainfall phases and wet/dry spells in current S2S forecast systems?
• How can we best address verification in a seamless manner, for comparing forecasts across timescales?

For addressing the questions dealing with verification methods, links with the Joint Working Group on Forecast Verification Research will be established. Verification is essentially a cross-cutting theme in the S2S project, necessary in all the sub-projects, but for addressing the questions dealing with climate processes and user needs, specific links with the S2S Monsoon, Africa and Extremes sub-projects will be established.

Proposed tasks/activities and tentative deliverables

2015-2016: Identify entry points and consult with user communities to define verification needs. Possible entry points include the S2S Africa Sub-project, the WWRP SERA working group, the WMO Commission for Agricultural Meteorology (CAgM), the WMO Programme on Hydrological Forecasting for Water Resources Management (HFWR), and the Research Program on Climate Change and Food Security (CCAFS).

2015: Literature review of S2S forecast verification research and survey on current S2S verification practices in operational centers, the latter performed by this sub-project contacting the operational centers.

2015-2016: Agree on and propose a minimum set of verification measures for intercomparison of subseasonal forecasting systems and provide guidance/recommendation for S2S forecast verification practice. The proposed measures should include basic verification metrics such as weekly rainfall amount, which may have not been assessed yet in a coordinated way. A coordinated verification effort already exists for operational seasonal forecasts through the Lead Centre for Standardised Verification of Long-Range Forecasts (LC-SVSLRF), responsible for collecting verification information from the WMO designated Global Producing Centers of long range forecasts (GPCs) and displaying it in standardised formats. For operational seasonal forecasts the Global Data-Processing and Forecast System (GDPFS) manual (Attachment II.8) provides the

standards for verification. For short and extended-range forecasts the GDPFS manual (Attachment II.7) provides the standards for verification. This S2S subproject on verification has the opportunity to contribute to establish the standards for S2S forecast verification and also link the verification activities performed by both research and operational communities.

2015-2018: Facilitate intercomparison of S2S forecasting systems and evaluation of the benefits of the multi-model ensemble approach. This intercomparison and beneficial evaluation will be promoted by encouraging the research community to start investigating the hindcasts and delayed realtime forecasts of different S2S forecasting systems as soon as these become available in the S2S database at ECMWF. Both historical and event based (case study) intercomparison assessments will be encouraged.

References:

Ferro CAT, Stephenson DB, 2011: Extremal Dependence Indices: improved verification measures for deterministic forecasts of rare binary events. *Weather and Forecasting*, 26, 699-713, doi:10.1175/WAF-D-10-05030.1.

Ferro CAT, 2014: Fair scores for ensemble forecasts. *Quarterly Journal of the Royal Meteorological Society*, 140, 1917-1923, doi: 10.1002/qj.2270

Gilleland, E., D. Ahijevych, B.G. Brown, B. Casati, and E.Ebert, 2009: Intercomparison of spatial forecast verification methods. *Weather and Forecasting*, **24**, 1416-1430.

Stephenson DB, Casati B, Ferro CAT, Wilson CA, 2008: The extreme dependency score: a non–vanishing measure for forecasts of rare events. *Meteorol Appl* 15:41–50

Weigel, A.P., D. Baggenstos, M.A. Liniger, F. Vitart, and C. Appenzeller, 2008: Probabilistic verification of monthly temperature forecasts. *Monthly Weather Review*, **136**, 5162-5182.

Weigel, A.P., and S. Mason, 2011: The Generalized Discrimination Score for ensemble forecasts. *Monthly Weather Review*, **139**, 3069-3074.