TIGGE/S2S CHALLENGE

User-oriented variables and their prediction in the TIGGE/S2S databases

One of the aims of the World Meteorological Organisation (WMO) is to encourage better communication between forecasters and users. The "High Impact Weather" (HIW) project and the "Societal and Economic Research Applications" (SERA) working group are examples where this desire is put into practice. This communication between forecasters and users will clearly be facilitated if it can be focused around user-oriented forecast variables. Simple examples are windspeed-cubed for energy production, and the multi-variate "Discomfort Index" (which combines temperature and humidity) for the health sector. Recently the Joint WWRP/WGNE Working Group on Forecast Verification Research (JWGFVR) ran a challenge to develop user-oriented forecast verification metrics, and the present challenge takes much of its inspiration from this. The aim here is to use existing user-oriented variables and/or to develop new ones which are purely meteorological, and to evaluate these variables using the TIGGE/S2S databases. Key questions to address include:

- How reliable and skilful are forecasts of these user-oriented variables?
- How do users calibrate these forecasts, and how could this calibration be improved?
- How can model developers learn from forecast calibration?

Predictability concepts

Ideally, ensemble forecasts should be 'reliable', and as 'sharp' as possible. For a reliable system, of the occasions when a particular weather event is predicted with any given probability p, it should occur with frequency p. This attribute is clearly important for users in their decision-making processes. Forecasts can be adjusted to achieve reliability – on average – but this is not so easily achieved in a flow-specific sense. Ideally, we would like the raw forecasts to be as reliable as possible.

While maintaining (or improving) reliability, the forecast community's aim is to reduce uncertainty in the initial conditions so that the outcome is as 'refined' (*i.e.* deterministic) as possible, within the limits of predictability. This requires more observational information, or more efficient extraction of information from the existing observations. For reliable forecast systems, refinement is equivalent to forecast 'sharpness' (associated with the 'spread' or standard deviation of the ensemble).

'Proper' scores, such as the Brier Score and Continuous Rank Probability Score (CRPS), reward improvements in reliability and refinement, and are therefore important for keeping ensemble forecast development on-track. In operational ensemble forecasting, it is probably fair to say that proper scores are largely applied to linear and univariate functions of model output. For example, the CRPS of 850hPa temperature (T850) is a headline score at the European Centre for Medium-Range Weather Forecasts (ECMWF), and it is monitored through the course of system development. While this is a very useful score of the forecast system as a whole, T850 it not the most relevant variable for many forecast users.

Key questions of relevance to this challenge are how reliable and skilful are TIGGE/S2S forecasts of more user-oriented variables?

Calibration concepts

From a pragmatic point-of-view, forecasts with any degree of refinement, will never be completely reliable, let-alone perfectly skilful. There is evidence that post-calibration of raw model output, and multi-model approaches, can improve forecast performance. Key questions of relevance to this challenge are how well does this calibration preserve multi-variate dependencies, spatial and temporal relationships between variables, and plausible "trajectories" over the course of the forecast?

Challenge

We would like participants to use/invent suitable, relatively simple, variables derived from ensemble forecast output which broadly represent the meteorological constraints important for a chosen user. For this challenge, it is important that the variables only involve meteorological information and do not include feedbacks with the decision-making process. The aim would be to evaluate proper scores of these variables, and their reliability components, over the range forecast lead-times available within the TIGGE and/or S2S databases. Calibration techniques could then be applied and evaluated, with the aim of improving forecast performance and improving feedback to model developers. Such a process should facilitate better two-way communication between the research and user communities.

Examples of user-oriented variables and questions to address

There are probably at least as many potential derived variables of forecast output as there are users of forecast information, but here is a short set of examples to motivate the challenge.

1. Wind damage and wind power

The prediction of windspeed V is often evaluated for ensemble forecast systems but the destructive force of the wind is proportional (in simple models) to the square of the wind-speed, and wind-power generation is proportional (in simple models) to the cube of the wind-speed. If an ensemble system is reliable for forecasts of V, then it should be reliable for forecasts of V^2 and V^3, but how does the CRPS differ for these different powers of V? What are the useful spatio-temporal scales to evaluate these variables? What lead-time is predictive skill no better than that of a forecast based on the climatological frequency? Can calibration lead to more reliable and/or more skilful forecasts, and why?

2. Health

According to the US National Weather Service, heat is the number one weather-related killer in the United States. Data shows that heat causes more fatalities per year than either floods, lightning, tornadoes or hurricanes. The Discomfort Index (DI) is a measure of how hot it feels when factoring in the effect that relative humidity has on one's ability to loose heat through sweating. DI is defined as $T - 0.0055(100 - RH)(T - 14.5)$, where T is the temperature (*e.g.* at a height of 2m above the surface) and RH is the relative humidity (in %). How do forecast scores of DI depend on the lead-time and the key thresholds shown in Table 1? How reliable are such forecasts? How can calibration help?

DI range (°C)	Implications
DI≤21	No discomfort
21≤DI<24	Under 50% population feels discomfort
24≤DI<27	Over 50% population feels discomfort
27≤DI<29	Most of population suffers discomfort
29≤DI<32	Everyone feels severe stress
32≤DI	State of medical emergency

Table 1. Classification of the human thermal comfort during summer according to the discomfort index values. From Stathopoulou et al (2005) Thermal remote sensing of Thom's Discomfort Index (DI): comparison with in situ measurements. Proc. SPIE 5983, Remote Sensing for Environmental Monitoring, GIS Applications, and Geology V, 59830K (29 October 2005); doi: 10.1117/12.627541

3. Agriculture

Plant growth can be strongly dependent on the temperature. A simple model of growth is that it is zero below a given threshold (T_b; e.g. 7°C for cool-season cereal crops such wheat), and then increases nearly linearly with temperature above this threshold. The 'Growing Degree Units' (GDU) variable can be used to evaluate this growth model. GDU is defined as $\sum_{Day=1}^{n} \mathrm{Max}\left(\frac{T_{\max}+T_{\min}}{2} - T_b,\ 0\right)$ where $T_{\max}$ is the maximum daily temperature and $T_{\min}$ is the minimum daily temperature. Hence, for three days where $\left(\frac{T_{\max}+T_{\min}}{2} - T_b\right)$ is 2, 1, -1, respectively, the cumulative GDU value would be 2+1+0=3. This derived variable could be calculated and scored for all periods up to the maximum lead-time of the TIGGE (and/or S2S) models. Again, how reliable are such forecasts? When – if at all – is predictive skill lost and how can calibration help?

4. Transport

In air travel, fuel economy is an important consideration from both profitability and environmental points of view. Fuel is used to over-come drag. Lift-induced drag is the essential component of drag required to keep the plane aloft; it is roughly inversely proportional to the speed of the aircraft and, when integrated over a scheduled flight window, a simple model would imply that this requires a fixed amount of fuel to generate the required lift. The remaining 'parasitic' aspect of drag – including 'form drag' and 'skin-friction drag' - grows with the square of the plane's speed relative to the air. Given a forecast of cruising altitude (*e.g.* 200hPa) winds $\boldsymbol{v}_a(\boldsymbol{x}, t)$, a minimal-fuel flightpath $\boldsymbol{v}_p(t)$ could be deduced by minimising the integral of the parasitic drag $\int_{Dep}^{Arr} \left|\boldsymbol{v}_p - \boldsymbol{v}_a\right|^2 dl$ (subject to the flight schedule and other constraints). With reasonable scaling assumptions, a key component of the uncertainty in the fuel requirement will be associated with the integral of the predicted along-track winds $\int_{Dep}^{Arr} \boldsymbol{v}_a \cdot \boldsymbol{dl}$. While standard evaluation assesses forecasts of winds at a point, here we see that it is useful to evaluate reliability, spread, skill, etc., on spatial scales equivalent to the distances between international airports (up to thousands of km). As a simple approximation, we could consider the great circle flightpath joining London and New York City, evaluate the integral of 200 hPa zonal wind for TIGGE forecasts for 'tomorrow', and use analysed (step 0) winds for the 'truth'. How reliable and skilful are these 'fuel requirement forecasts'?

Summary

The "Predictability, Dynamics and Ensemble Forecasting" (PDEF) working group, and the "Subseasonal-to-Seasonal Prediction" (S2S) project, of the WMO would like to encourage better communication between forecasters and users through further development of user-oriented meteorological variables, and their evaluation within the TIGGE/S2S databases. These variables can be relevant to any chosen user community (just a few possible examples were given above – another key one might be associated with the ideal conditions for diseases to develop).

We would like to encourage submission of abstracts on such derived variables to the "Workshop on Predictability, dynamics and applications research using the TIGGE and S2S ensembles", to be held at the European Centre for Medium-Range Weather Forecasts (ECMWF) from 2-5 April 2019. At this workshop, a break-out group will specifically focus on user-oriented forecast variables, and their verification.