The Supplement of

Evaluating Multi-Year Prediction Proficiency for Decision Support in Finland's Energy Industry

SURAJ D POLADE AND HILPPA GREGOW

Finnish Meteorological Institute

S1 Method

The accuracy of this decade-long prediction for the upcoming 1–5 years is evaluated through both deterministic and probabilistic forecasts for four meteorological variables: temperature and precipitation during the extended cold season (November - March) and summer (May-August) against the E-OBS observation dataset (Cornes et al. 2018). Given their substantial influence on renewable energy generation and consumption, our analysis focuses exclusively on these two variables. Furthermore, we emphasise these variables due to their availability in subsequent real-time predictions provided by the WMO Global Annual to Decadal Climate Update (GADCU). Decadal forecasts, with hindcast (retrospective) accessibility, offer an advantage in evaluating forecasting accuracy compared to traditional climate predictions. The hindcast experiment is conducted in the same way as real-time forecasts but for past dates where observation datasets are available to assess the skill of decadal forecasts, as opposed to climate simulations. Numerous approaches are available for developing multi-model forecasts, yet consensus on the most effective method is lacking (Hemri et al. 2020). In this study, we chose the simplest approach, wherein equal weights are assigned to all forecast models. The deterministic forecast is produced by averaging ensemble means from each forecast system. In contrast, the probabilistic forecast is formed by averaging probability density functions of all forecast models for each tercile, with equal weights assigned. The decadal predictions offer forecasts initiated each year for the subsequent 5–10 years; in this analysis, we specifically focus on the forecast for years 1–5. To clarify, the forecast run commenced in 2000, utilizing the forecast until 2005 (2001 to 2005). Subsequently, the 2006 (2006 to 2010) forecast initiated from the run in 2005, and this pattern continued for subsequent years.

In assessing forecast quality, we used anomalies to mitigate systematic errors caused by differences in climatologies between forecast systems and reference datasets. To facilitate a meaningful comparison, we standardized the data from both the ensemble systems and E-OBS reference dataset onto a common grid with a spatial resolution of 0.25. These anomalies were calculated relative to the 1991–2020 climatology (the GADCU reference period). We employed three important metrics for our evaluation: the anomaly correlation coefficient (ACC), root-mean-square error skill score (RMSSS), and ranked probability skill score (RPSS). The evaluation was performed for the period 1991–2020. The ACC measures the linear relationship between simulated and observed anomalies, with a perfect forecast earning a score of 1, while values near zero or in the negative range indicate lower forecast skill. RMSSS, based on root-mean-square error, assesses the improvement or deterioration of forecast errors compared to a reference forecast. Positive RMSSS indicates greater forecast skill, whereas a negative value suggests the reference forecast is more skilful. In the domain of probabilistic forecasting, we used RPSS to measure enhancements or the lack thereof in probabilistic forecasts compared to a reference forecast across lower, middle, and upper terciles. Similar to RMSSS, a positive RPSS indicates higher forecast skill, while a negative RPSS suggests lower skill relative to the reference. Probabilistic forecast skill is assessed within equiprobable categories, including below the lower tercile, between the lower and upper terciles, and above the upper tercile. Importantly, ACC and RPSS are insensitive to biases in mean and variance, while the RMSSS is sensitive and accounts for the signal's amplitude. In order for the near-term probabilistic decadal predictions to be of value, it is essential to rectify the systematic errors within the forecasting system. We implemented calibration postprocessing techniques based on Doblas-Reyes et al. (2005), encompassing adjustments and variance inflation of predictions aimed at mitigating bias, enhancing overall forecast accuracy, and refining the reliability of the forecasts.

Table S1. Specifications of Forecast Systems Contributing to the GADCU and DCPP-A Components of CMIP6. The spatial resolution is shown for the atmospheric grid.

Forecast System	Institution	Spatial resolution	Members
CanESM5	CCCma	2.8° X 3 2.8°	10
MPI-ESM1-2-LR	DWD	1.9° X 3 1.9°	16
EC-Earth3	BSC	0.7° X 3 0.7°	10
MIROC6	MIROC	1.4° X 3 1.4°	10
MRI-ESM2.0	MRI	1.125° X 1.125°	10
NorCPM1	NCC	1.9° X 3 2.5°	16

S2 Application of decadal prediction.

Figure S1 presents an illustrative climate service product tailored for the energy sector, with a focus on terciles. This figure showcases a time series of calibrated forecasts for multi-year averaged temperature and precipitation during both the extended winter and summer seasons within the GADCU decadal prediction framework, spanning the period from 2023 to 2027. These predictions were initialized in November 2022 and represent averages across Finland. The corresponding observational values are also included in the time series, covering the hindcast period from 1971 to 2020. Temperature data is presented as absolute anomaly values relative to the 1991–2020 reference period, while precipitation is expressed as a percentage change.

The forecast for the extended cold season's temperature proves reliable, with 80 % of observations falling within an 80 % ensemble spread. A linear correlation of 0.38 is observed between the observational data and the ensemble mean model prediction. In 2023, there's a higher probability (51 %) of the upper tercile for temperature, and the average temperature exceeds the reference period by 0.8 degrees Celsius. Similarly, for the subsequent 1–5 years, predictions indicate a probability of more than 50 % for the upper tercile, with an average temperature surpassing the reference period by 0.8 degrees Celsius. The summer forecast also displays reliability, with a correlation of 0.37 between the reference and ensemble mean. Similar to the extended winter, both 1- and 1–5-year predictions show an enhanced chance of the upper tercile, with probabilities of 51 % and 43 %, respectively, and the average temperature exceeding the reference period by 0.3 degrees Celsius.

In the extended winter, the precipitation forecast is reliable, with a correlation of 0.31 between the reference and ensemble mean. There's a higher probability of precipitation falling between the middle and upper terciles for 2023, while the 1–5-year projection indicates a 40 % likelihood of being in the upper tercile, with an overall possibility of increasing precipitation by more than 2 % compared to the reference period. Conversely, the precipitation projection is less reliable during the summer, with a correlation as low as 0.14. Nonetheless, there is an enhanced chance of the upper terciles for both the 1-year and 1–5-year forecasts, with average precipitation expected to increase by 0.6 % and 2 %, respectively.

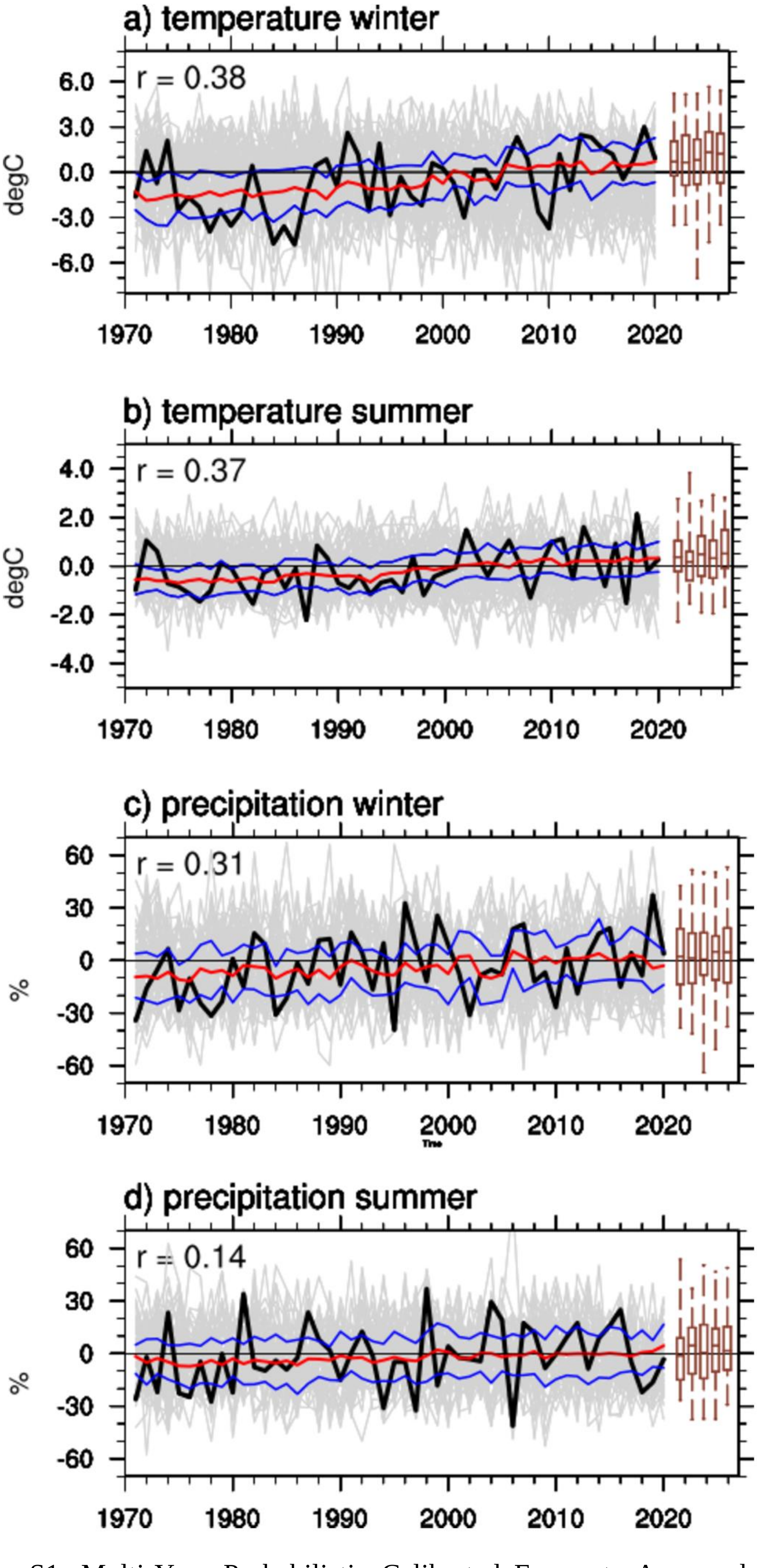

Fig S1: Multi-Year Probabilistic Calibrated Forecasts Averaged for Temperature in (a) Extended Winter and (b) Summer, and for Precipitation in (c) Extended Winter and (d) Summer across Finland.

References

Doblas-Reyes, F. J., Hagedorn, R. and Palmer, T., 2005: The rationale behind the success of multi-model ensembles in seasonal forecasting – II. Calibration and combination. *Tellus*, 57A, 234–252, https://doi.org/10.3402/tellusa.v57i3.14658.

Cornes, R., van der Schrier, G., van den Besselaar, E.J.M., Jones, P., 2018: An Ensemble Version of the E-OBS Temperature and Precipitation Datasets, J. Geophys. Res. Atmos., 123, https://doi.org/10.1029/2017JD028200.

Hemri, S., and Coauthors, 2020: How to create an operational multi-model of seasonal forecasts?. *Climate Dyn*, 55, 1141– 1157, https://doi.org/10.1007/s00382-020-05314-2.