



Multi-model forecast quality assessment of CMIP6 decadal predictions

Carlos Delgado-Torres¹, Markus G. Donat^{1,2}, Nube Gonzalez-Reviriego¹, Louis-Philippe Caron^{1,3}, Panos J. Athanasiadis⁴, Pierre-Antoine Bretonnière¹, Nick J. Dunstone⁵, An-Chi Ho¹, Klaus Pankatz⁶, Andreas Paxian⁶, Núria Pérez-Zanón¹, Margarida Samsó Cabré¹, Balakrishnan Solaraju-Murali¹, Albert Soret¹, and Francisco J. Doblas-Reyes^{1,2}

¹Barcelona Supercomputing Center (BSC), Barcelona, Spain

²Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain

³Ouranos, 550 Sherbrooke St W, Montreal, QC, Canada

⁴Centro Euro-Mediterraneo sui Cambiamenti Climatici, Bologna, Italy

⁵Met Office Hadley Centre, FitzRoy Road, Exeter, EX1 3PB, UK

⁶Business Area of Climate and Environment, Deutscher Wetterdienst, Offenbach (Main), Germany

Decadal climate predictions are a new source of climate information for inter-annual to decadal time scales, which is of increasing interest for users. Forecast quality assessment is essential to identify windows of opportunity (e.g., variables, regions, and lead times) with skill that can be used to develop a climate service and inform users in several sectors. Also, it can help to monitor improvements in current forecast systems. The Decadal Climate Prediction Project Component A (DCPP-A) of the Coupled Model Intercomparison Project Phase 6 (CMIP6) now provides the most comprehensive set of retrospective decadal predictions from multiple forecast systems. The increasing availability of these simulations leads to the question of how to best post-process the raw output from the forecast systems so that the most useful and reliable information is provided to users.

This work evaluates the quality of deterministic and probabilistic forecasts for spatial fields of near-surface air temperature and precipitation, and time series of the Atlantic multi-decadal variability index (AMV) and global near-surface air temperature anomalies (GSAT) generated from all the available decadal predictions contributing to CMIP6/DCPP-A (169 members from 13 forecast systems). The predictions generally show high skill in predicting temperature and the AMV and GSAT time series, while the skill is more limited for precipitation. Also, different approaches for building a multi-model forecast are compared (pooling all ensemble members versus combining the averages from individual forecast systems), finding small differences. Besides, the multi-model ensemble is compared to the individual forecast systems. The best system usually provides the highest skill. However, the multi-model ensemble is a reasonable choice for not having to select the best system for each particular variable, forecast period and region. Furthermore, the decadal predictions are compared to the uninitialized historical climate simulations (195 members from the same forecast systems as the decadal prediction members) to estimate the impact of

initialization. An added value is found for temperature over several ocean and land regions, and for the AMV and GSAT time series, while it is more reduced for precipitation. Moreover, the full DCP-A ensemble is compared to a sub-ensemble of predictions that could be provided in near real-time for a potential operational product generation. The comparison shows a benefit of using a large ensemble over several regions, especially for temperature. Finally, the implications of these results in a climate services context are discussed.