The method used to produce the temperature forecasts

The 3-month probabilistic seasonal forecasts of minimum and maximum temperatures are produced by making use of the archived and real-time forecast output from two of the coupled ocean-atmosphere models administered through the North American Multi-Model Ensemble (NMME) project (https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/north-american-multi-model-ensemble). The models used here to produce multi-model forecasts are the GFDL-CM2p1-aer04 and the GFDL-CM2p5-FLOR-A06, with 10 and 12 ensemble members respectively. The model outputs (using March initial conditions) are calibrated before producing the forecasts for the target seasons (AMJ, MJJ and JJA) by making use of the respective hindcasts (or re-forecasts) for each of the target seasons. The hindcast period is over 38 years from 1982 through 2019. The calibration process for each ensemble member involves a normalization procedure relative to the ensemble mean that is applied separately on the different climate parameters (minimum and maximum temperature) and the two seasonal forecast models. Probability forecasts are then derived for three equi-probable categories (below-normal, near-normal and above-normal) from each individual model. These probability forecasts are then averaged (using an equal weight) to produce the multi-model seasonal forecasts.

The probabilistic forecasts at each of the 19 locations are extracted from the multi-model forecast by using a nearest neighbour method based on the radial distance between the location of interest (city or town) and the closest model grid-point. The same location extraction procedure is applied on the observed minimum and maximum temperatures, obtained from the Climate Prediction Center Global Daily Temperature dataset (CPC Global Temperature data provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their Web site at https://www.esrl.noaa.gov/psd/).

The probabilistic forecasts for each season and for each of 19 locations are presented as pie charts that reflect the predicted probabilities for the respective categories. Verification statistics are calculated over the 38 years and one of many verification parameters, the relative operating characteristic (ROC) score (x 1000), is presented next to the colour bars associated with the three categories. When that score value is HIGHER than 500 it means that over the 38-year verification period, the multi-model is able to discriminate that category from the other two. The four numbers below the colour bars represent (from top to bottom) the 25th, 33.3rd, 66.7th and 75th percentile values of the observed data over the same 38 years.