Supplementary Information

GPRChinaTemp1km: a high-resolution monthly air temperature dataset for China (1951–2020) based on machine learning

Qian He[1, 2], Ming Wang[1, 3], Kai Liu[1, 3], Kaiwen Li[1, 2], Ziyu Jiang[1, 2]

[1] Academy of Disaster Reduction and Emergency Management, Beijing Normal University, 100875 Beijing, China

[2] Faculty of Geographical Science, Beijing Normal University, 100875 Beijing, China

[3] The School of National Safety and Emergency Management, Beijing Normal University, 100875 Beijing, China

Correspondence to: Ming Wang (wangming@bnu.edu.cn)

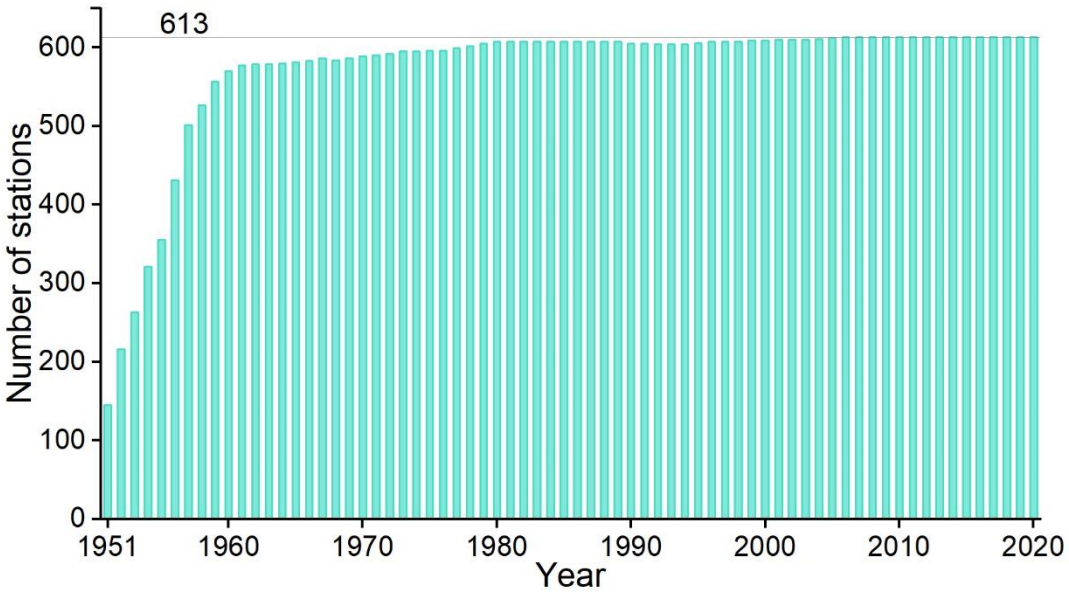

Figure S1: Number of weather stations in each year (from 1951 to 2020).

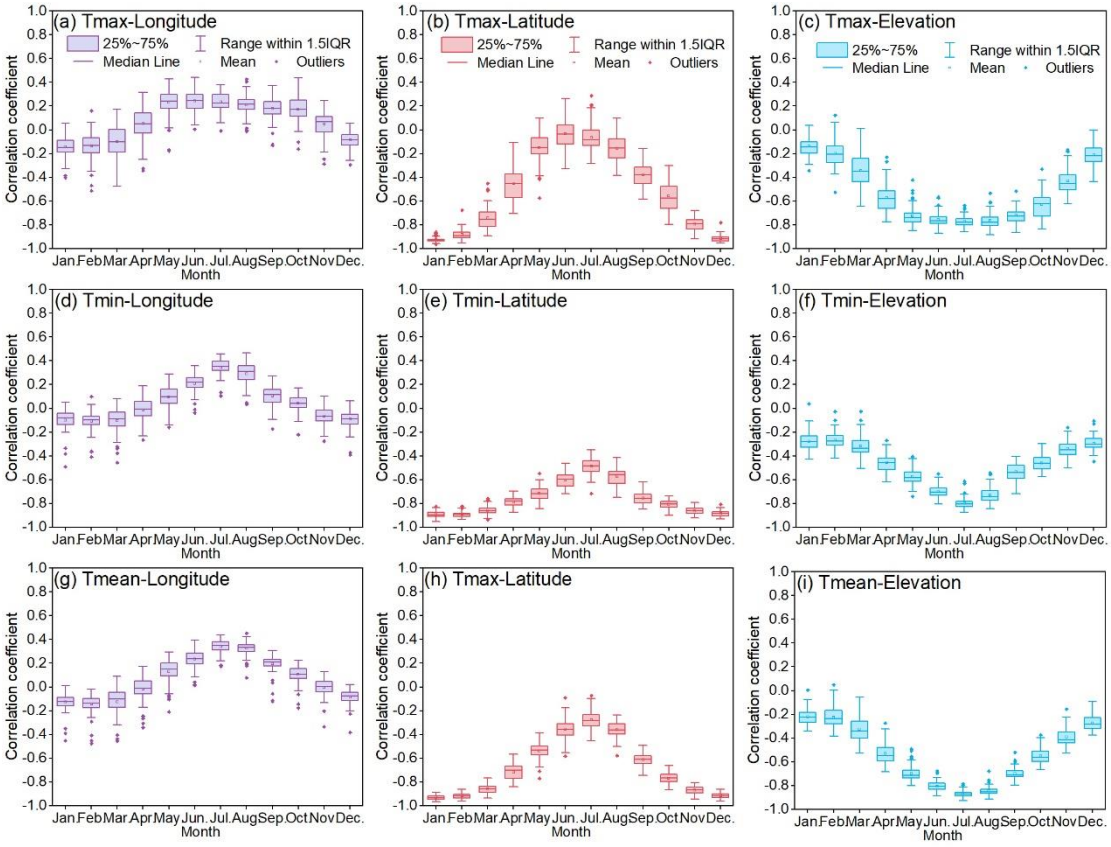

Figure S2: Correlation coefficients between temperature with longitude, latitude and elevation for each month from 1951-2020. a, b, c: monthly maximum air temperature (Tmax), c, d, e: monthly minimum air temperature (Tmin); g, h, g: monthly mean air temperature (mean).

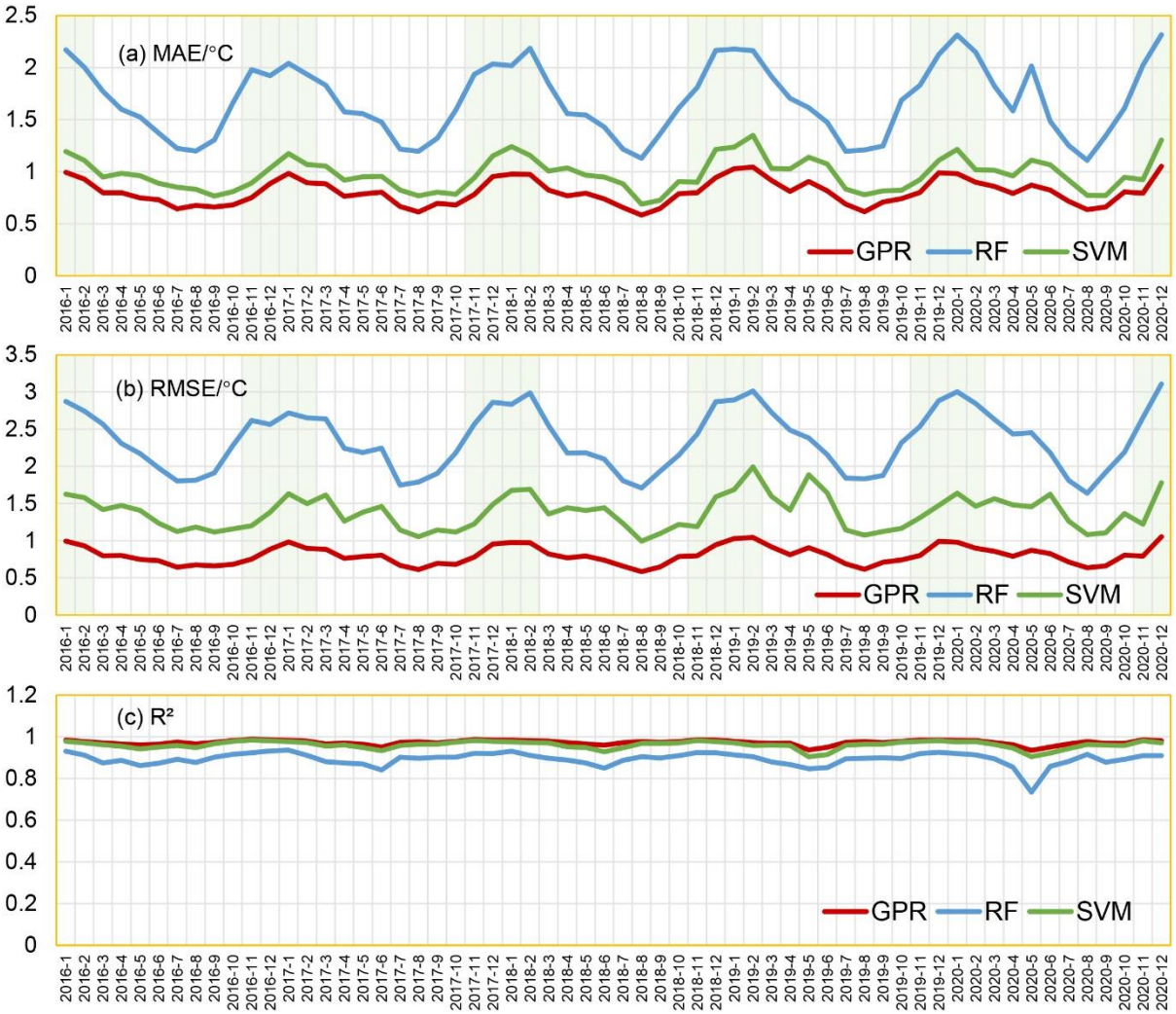

Figure S3: Zoomed-in accuracy metrics of Tmean for the period of January 2016 to December 2020.

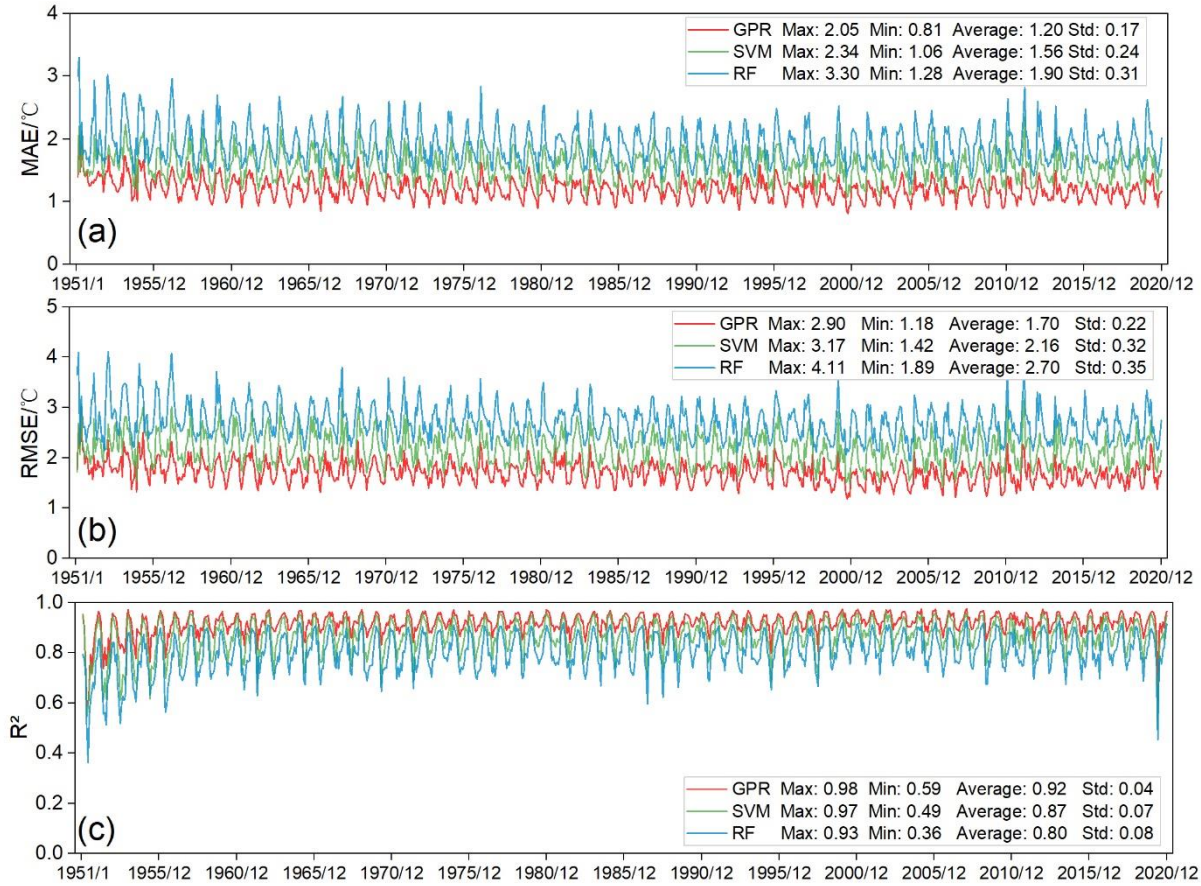

Figure S4: Mean absolute error (MAE), root mean square error (RMSE) and coefficients of determination (R^2) between observed Tmax and predicted Tmax by three machine learning models (GPR, SVM, RF) of testing weather stations over the time period from January 1951 to December 2020.

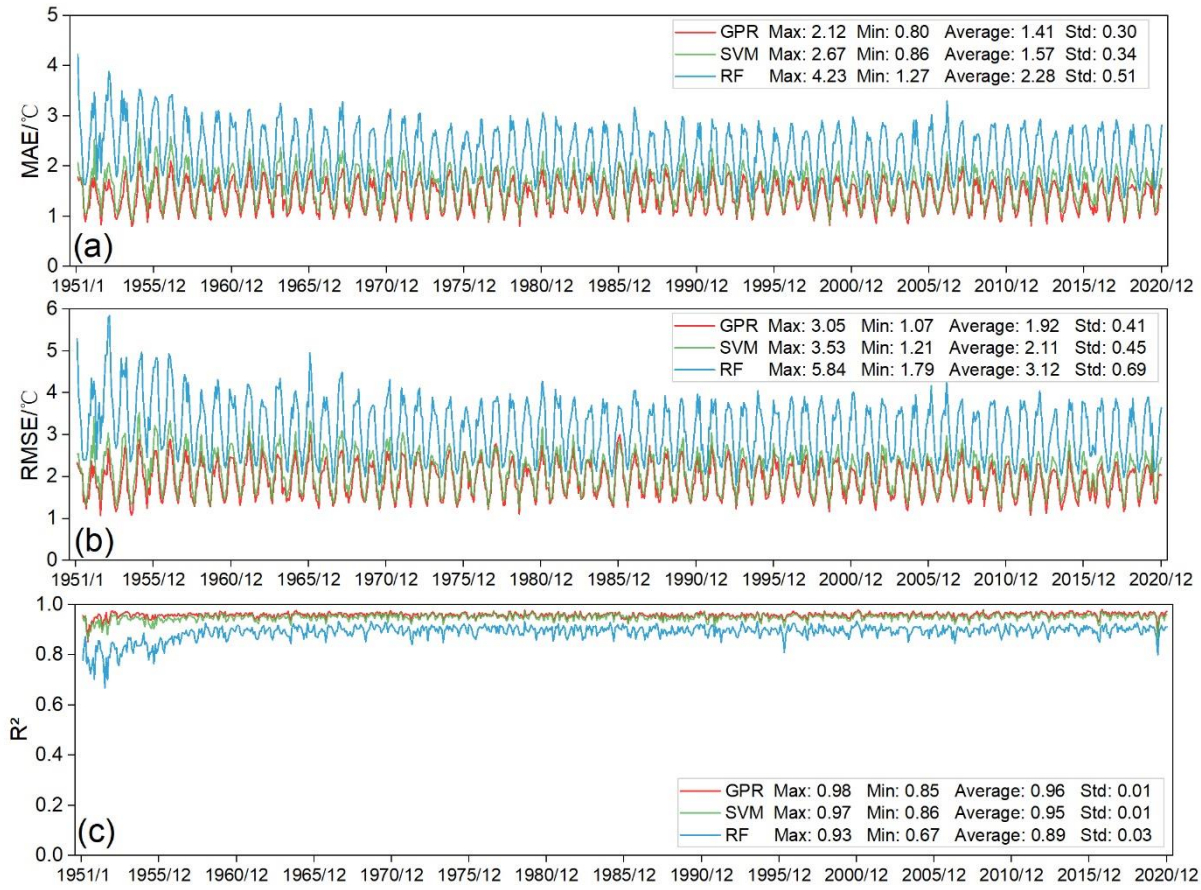

Figure S5: Mean absolute error (MAE), root mean square error (RMSE) and coefficients of determination (R2) between observed Tmin and predicted Tmin by three machine learning models (GPR, SVM, RF) of testing weather stations over the time period from January 1951 to December 2020.

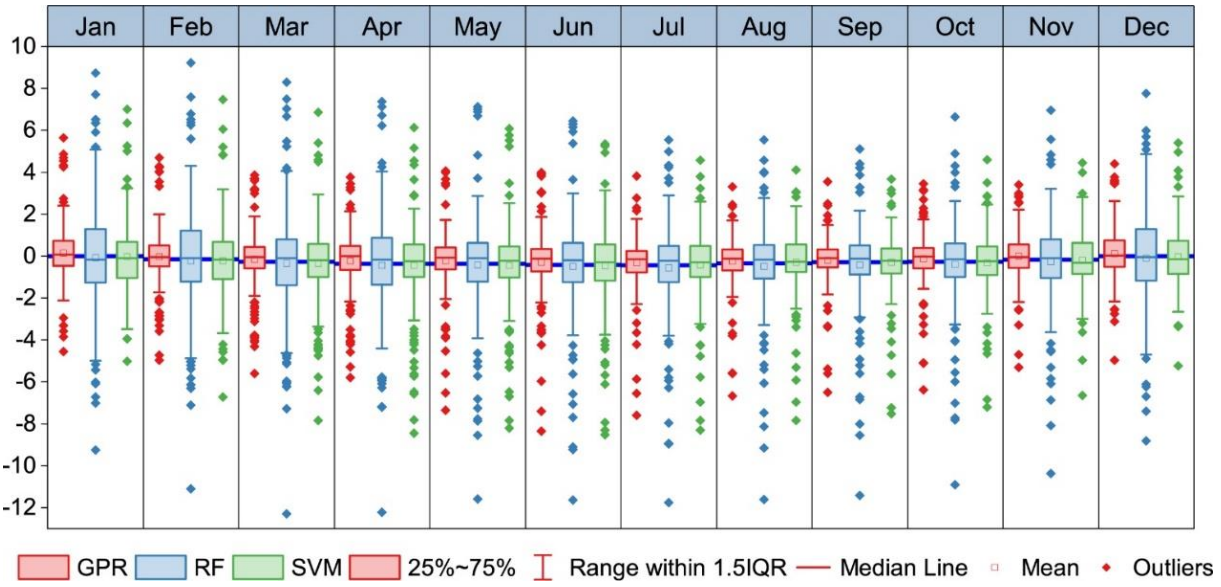

Figure S6: Residuals of the monthly Tmax predicted by the machine learning models with respect to in situ Tmax for the test meteorological stations. Note that the average of the residuals of Tmax from 1951–2020 for each test meteorological station is shown for each month.

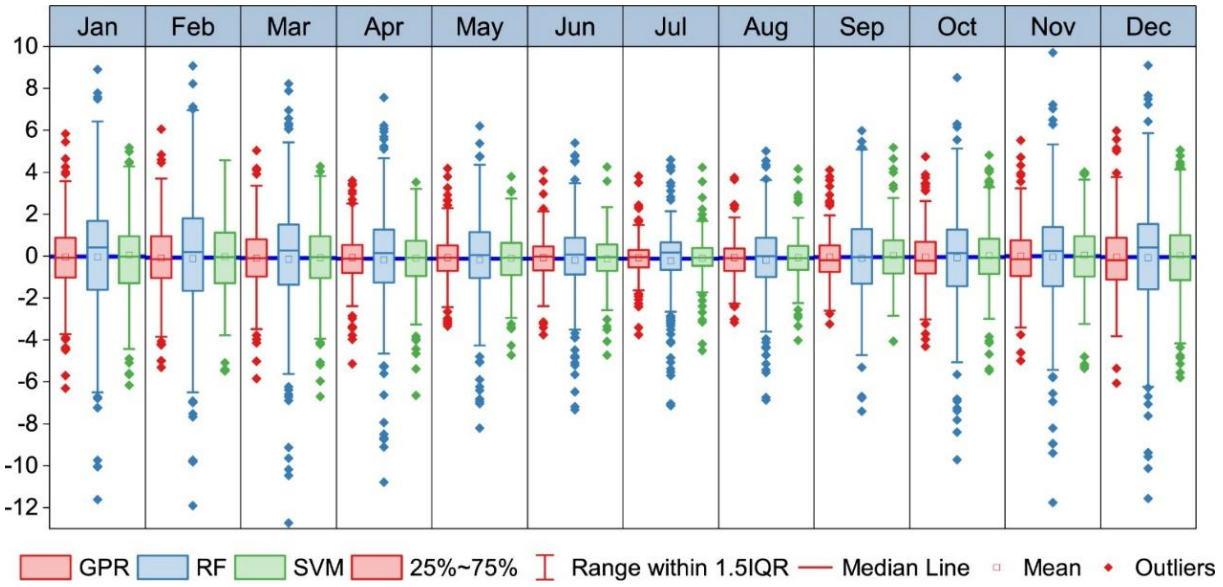

Figure S7: Residuals of the monthly Tmin predicted by the machine learning models with respect to in situ Tmin for the test meteorological stations. Note that the average of the residuals of Tmin from 1951–2020 for each test meteorological station is shown for each month.

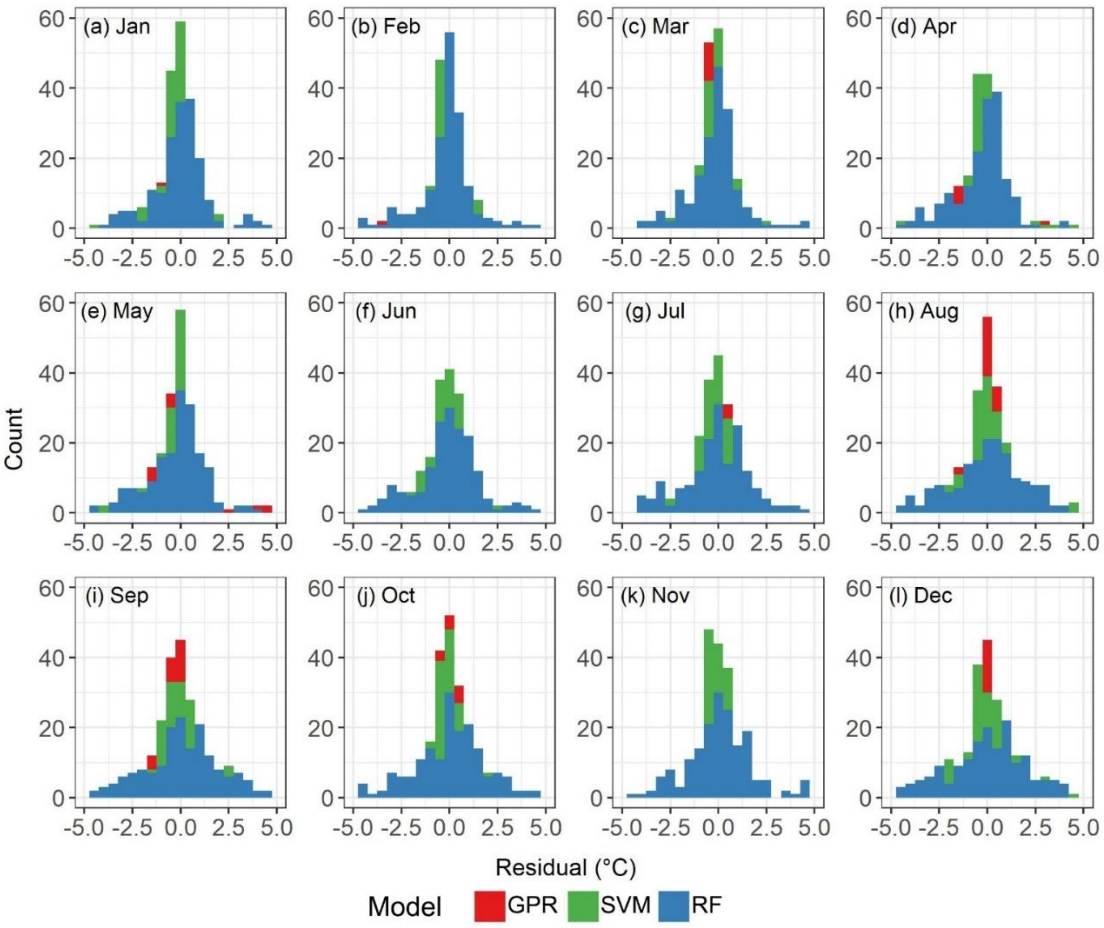

Figure S8: Frequency distribution of residuals for monthly Tmean using the machine learning methods during the period from 1951 to 2020. Note that the displayed residuals are the average of residuals of 70 years (1951-2020) for each month.

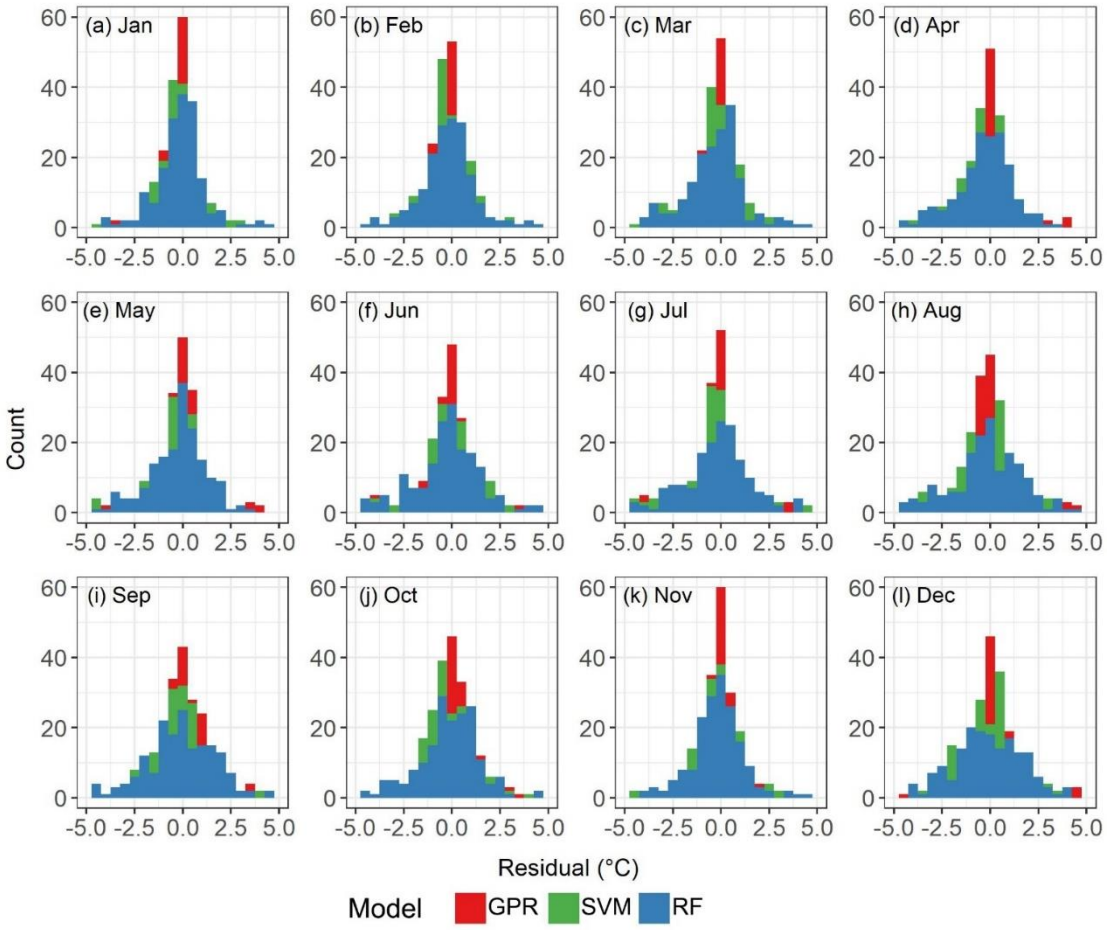

Figure S9: Frequency distribution of residuals for monthly Tmax using the machine learning methods during the period from 1951 to 2020. Note that the displayed residuals are the average of residuals of 70 years (1951-2020) for each month.

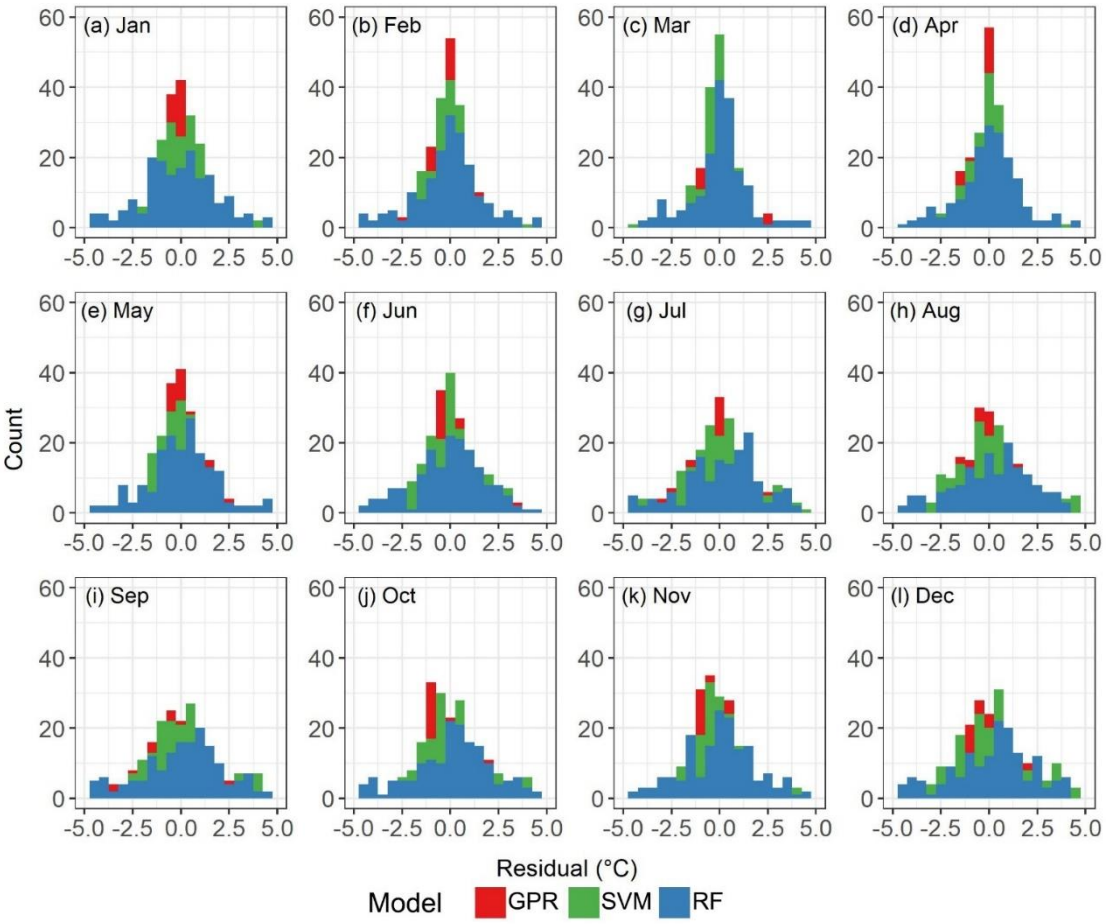

Figure S10: **Frequency distribution of residuals for monthly Tmin using the machine learning methods during the period from 1951 to 2020. Note that the displayed residuals are the average of residuals of 70 years (1951-2020) for each month.**

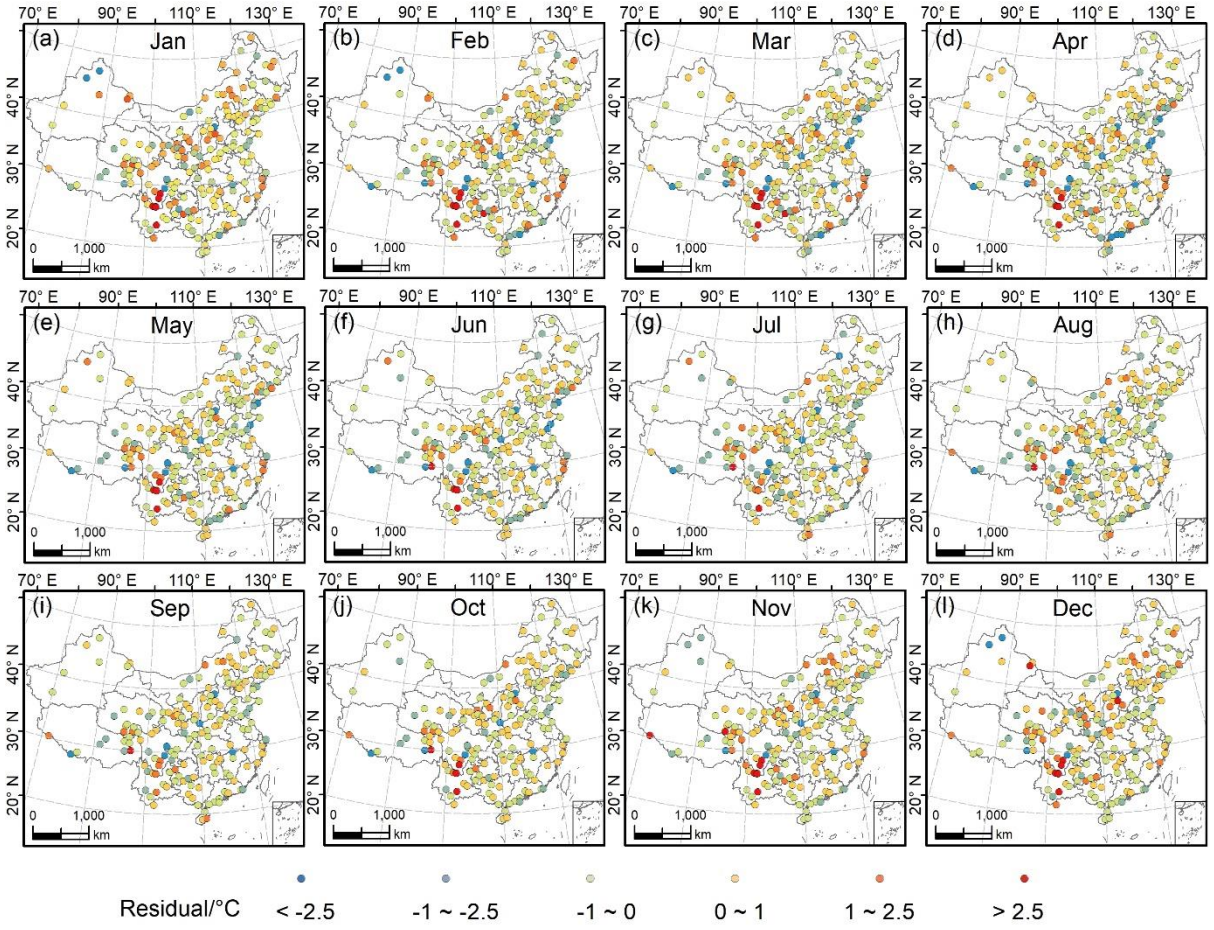

Figure S11: The spatial distribution of residuals between the observed Tmax and the predicted Tmax by GPR for the testing stations for each month. Note that the exhibited residuals are the average residual of 70 years from 1951- 2020 for each month.

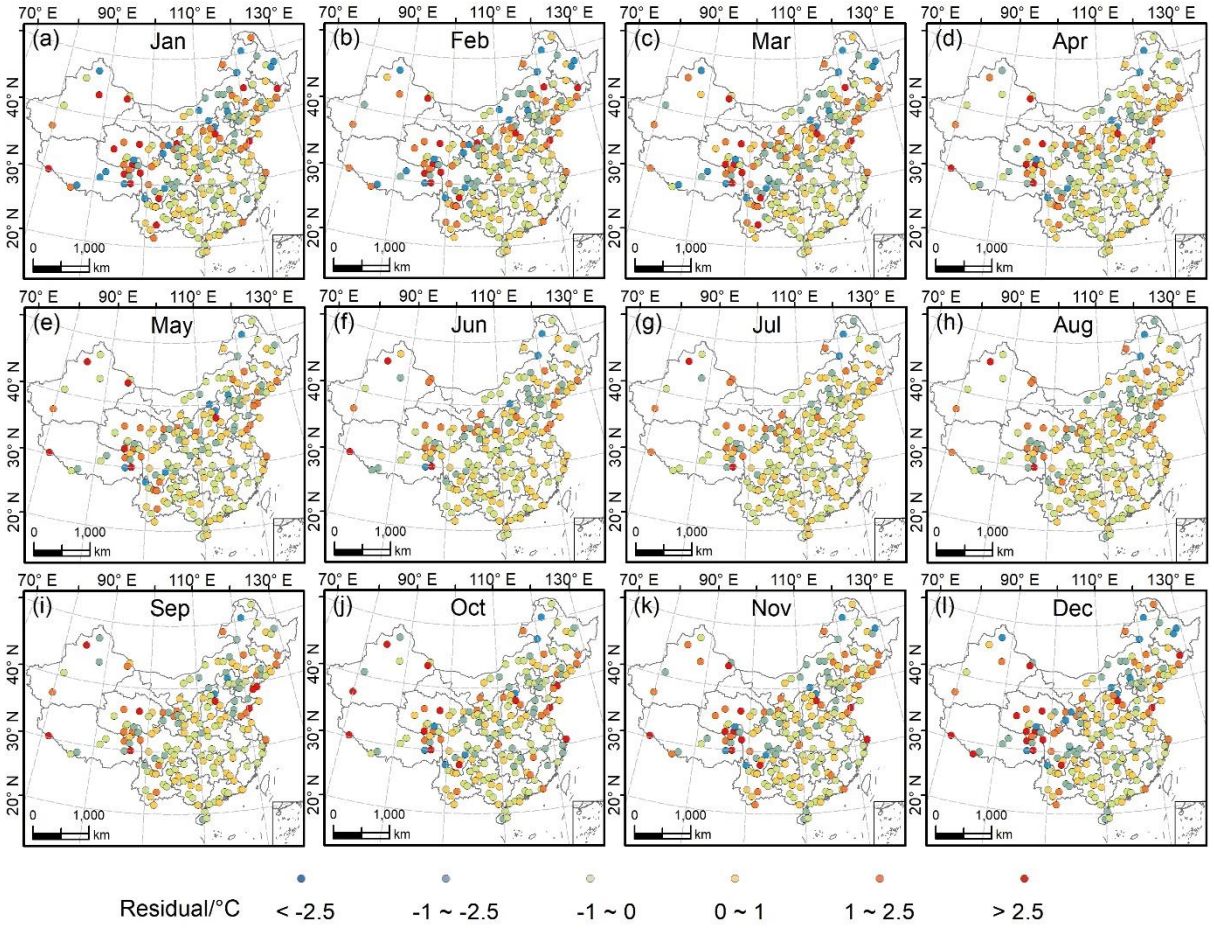

Figure S12: The spatial distribution of residuals between the observed Tmin and the predicted Tmin by GPR for the testing stations for each month. Note that the exhibited residuals are the average residual of 70 years from 1951- 2020 for each month.

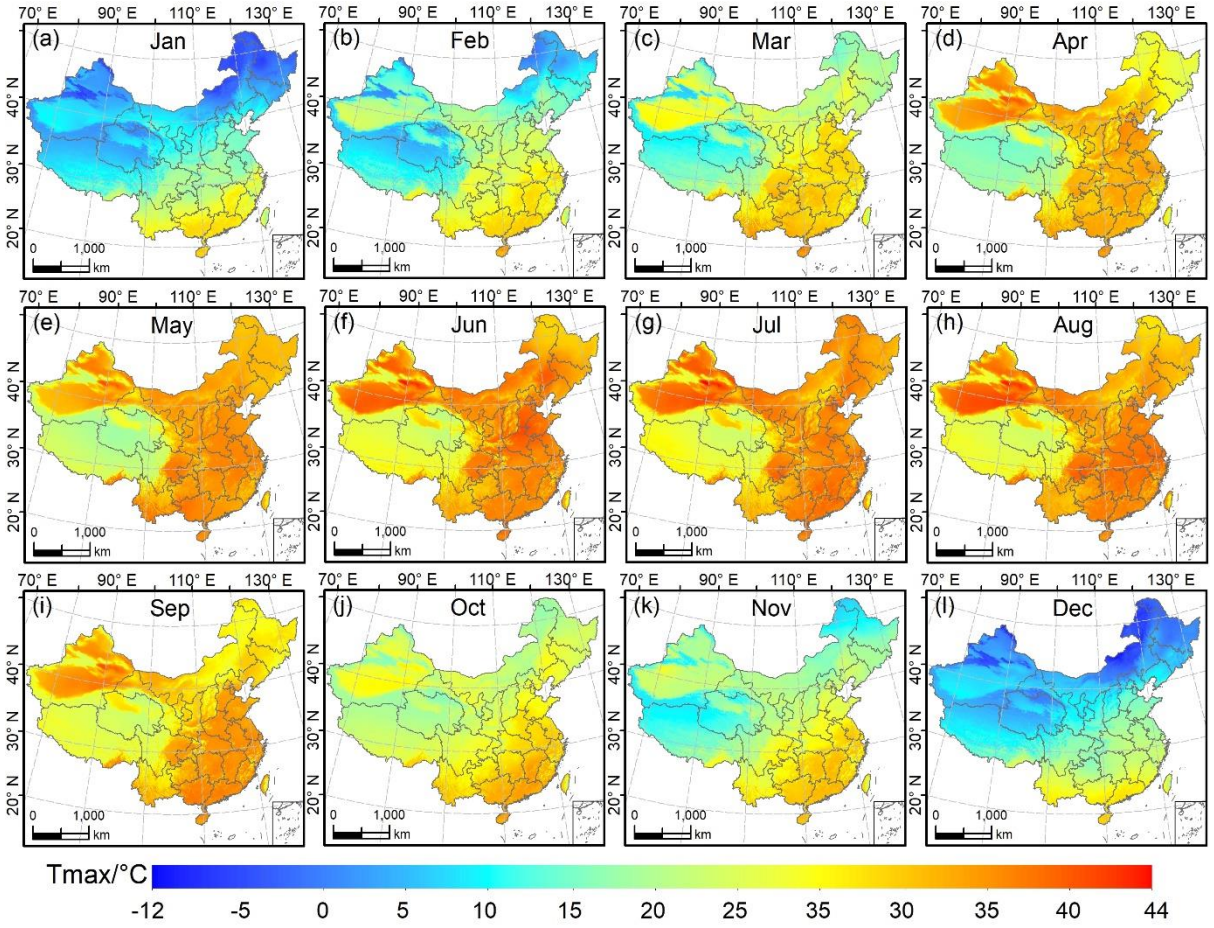

Figure S13: The spatial distribution maps of monthly Tmax predicted by GPR in China for each month in 2020.

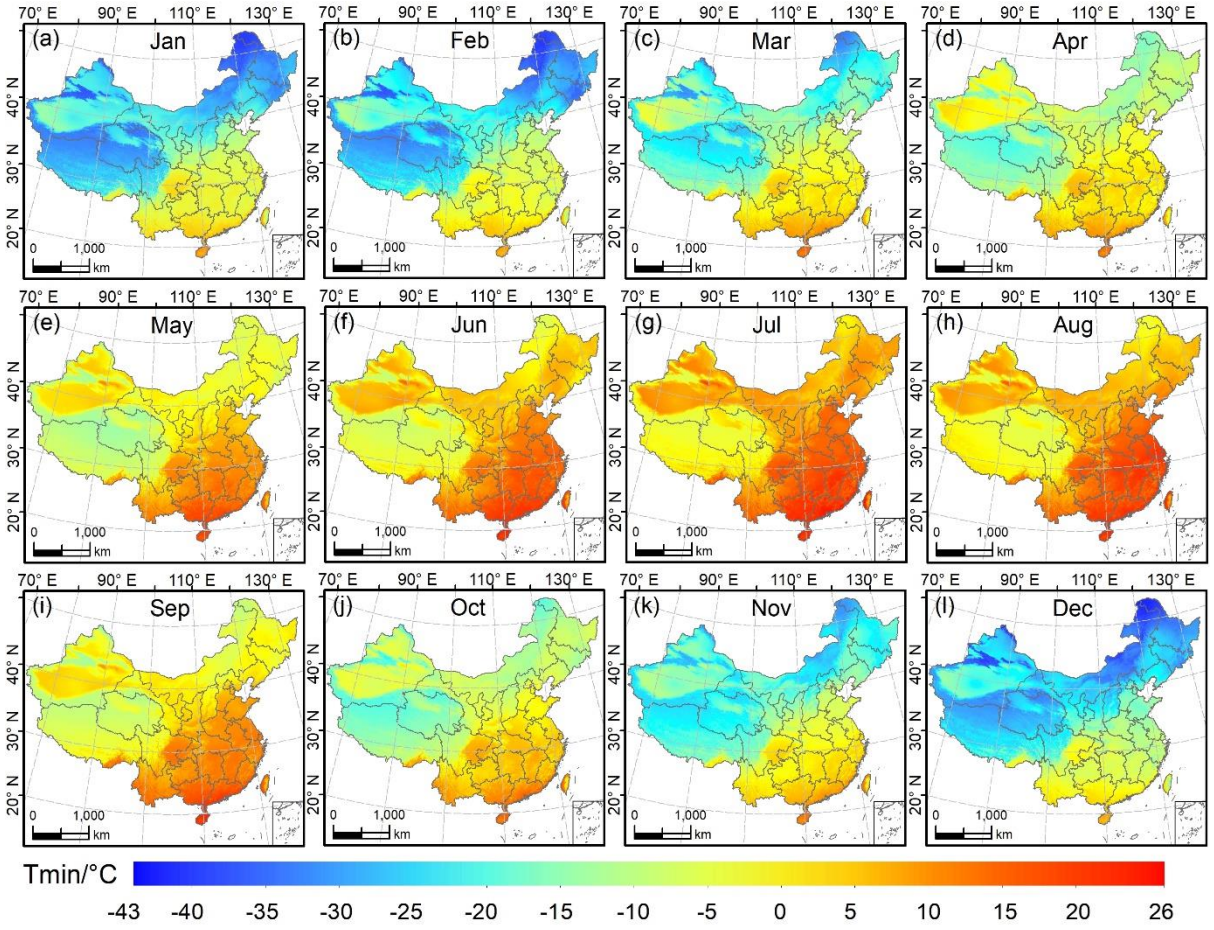

Figure S14: The spatial distribution maps of monthly Tmin predicted by GPR in China for each month in 2020.

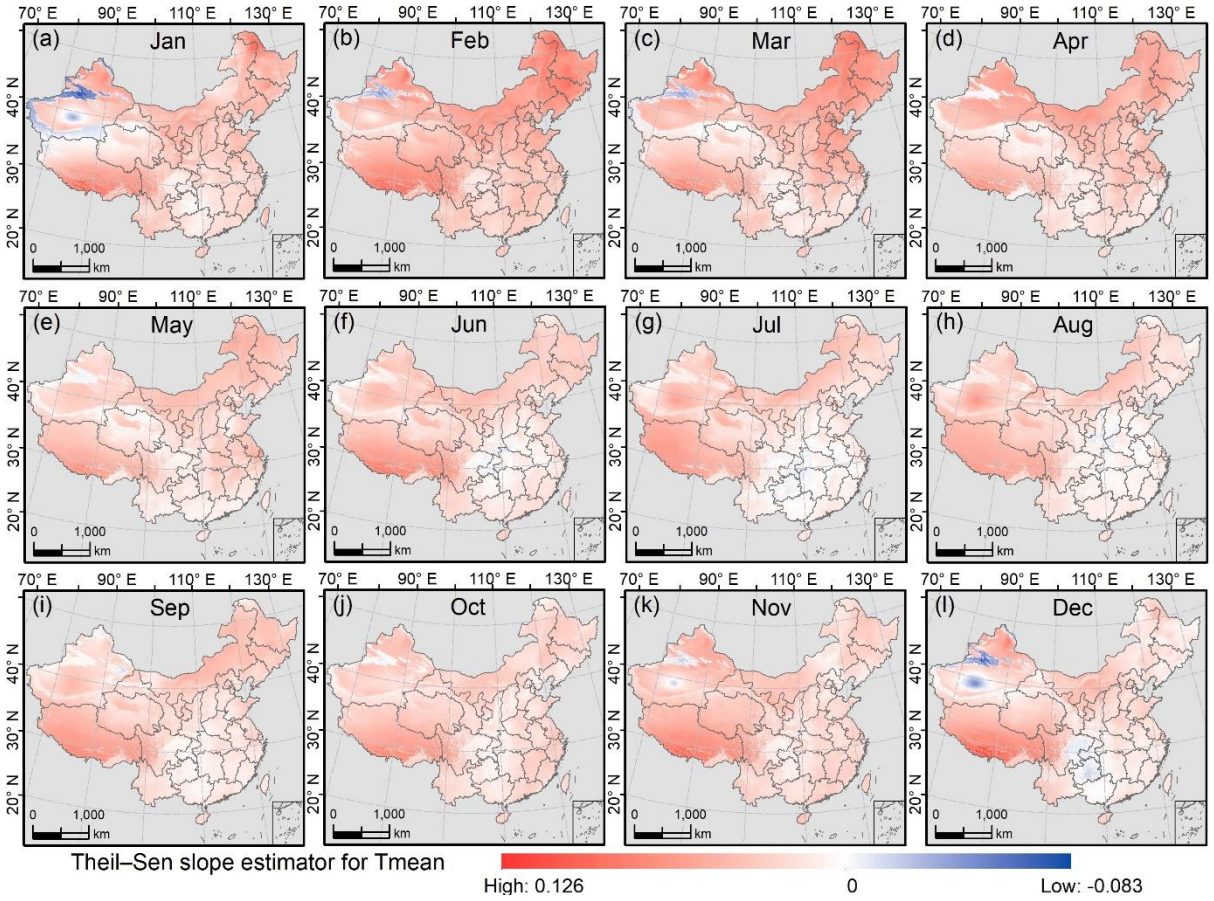

Figure S15: Theil–Sen (median) slope for Tmean over China (1951–2020) in each month.

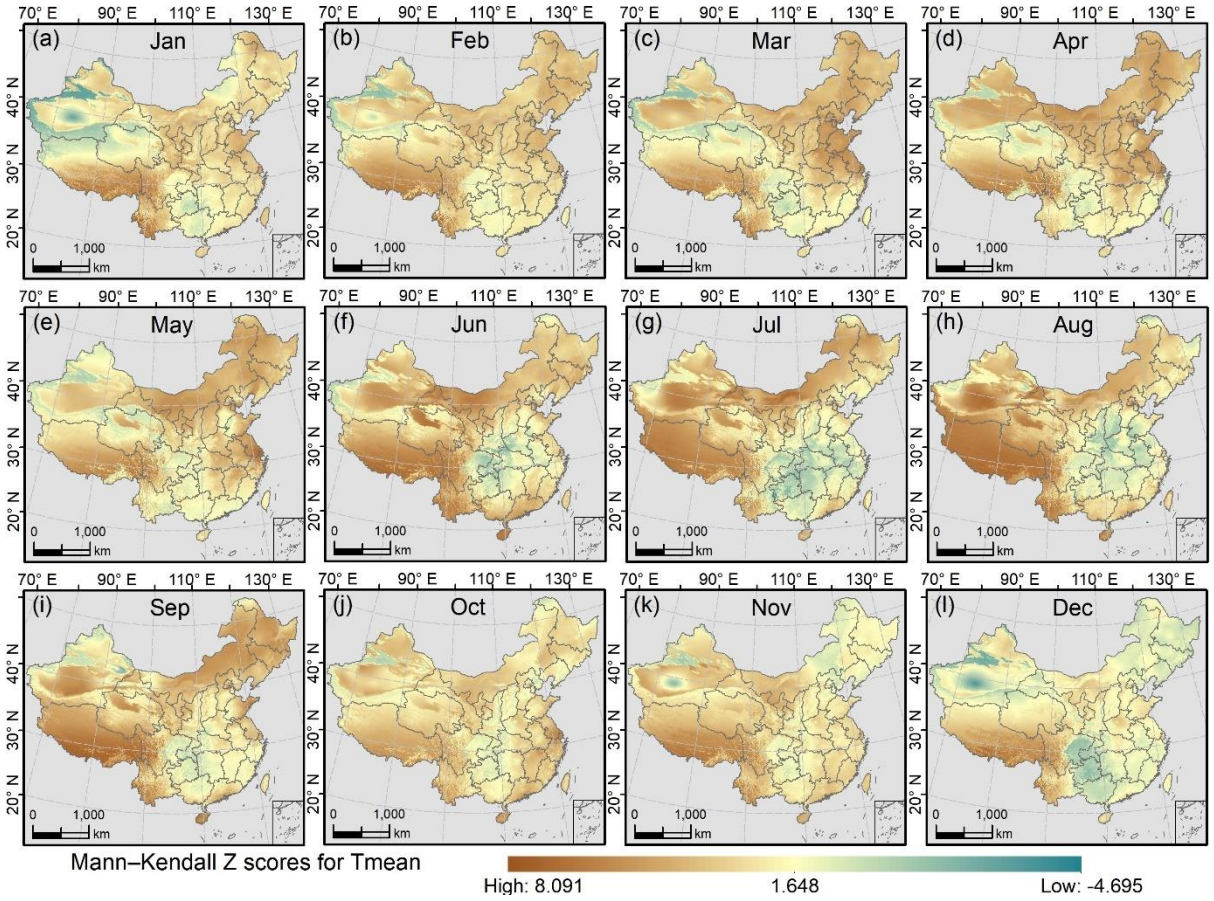

Figure S16: Mann–Kendall Z scores for Tmean over China (1951–2020) in each month.

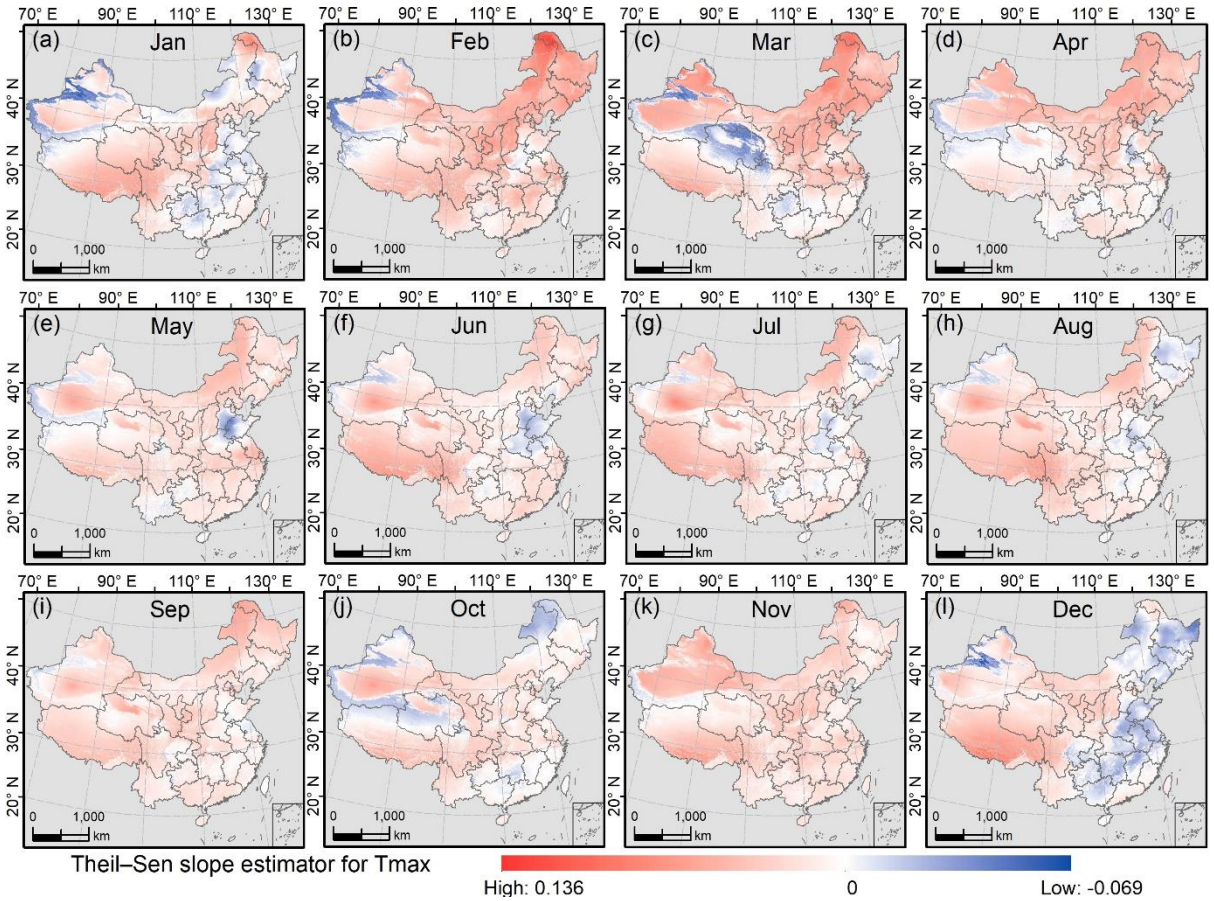

Figure S17: Theil–Sen (median) slope for Tmax over China (1951–2020) in each month.

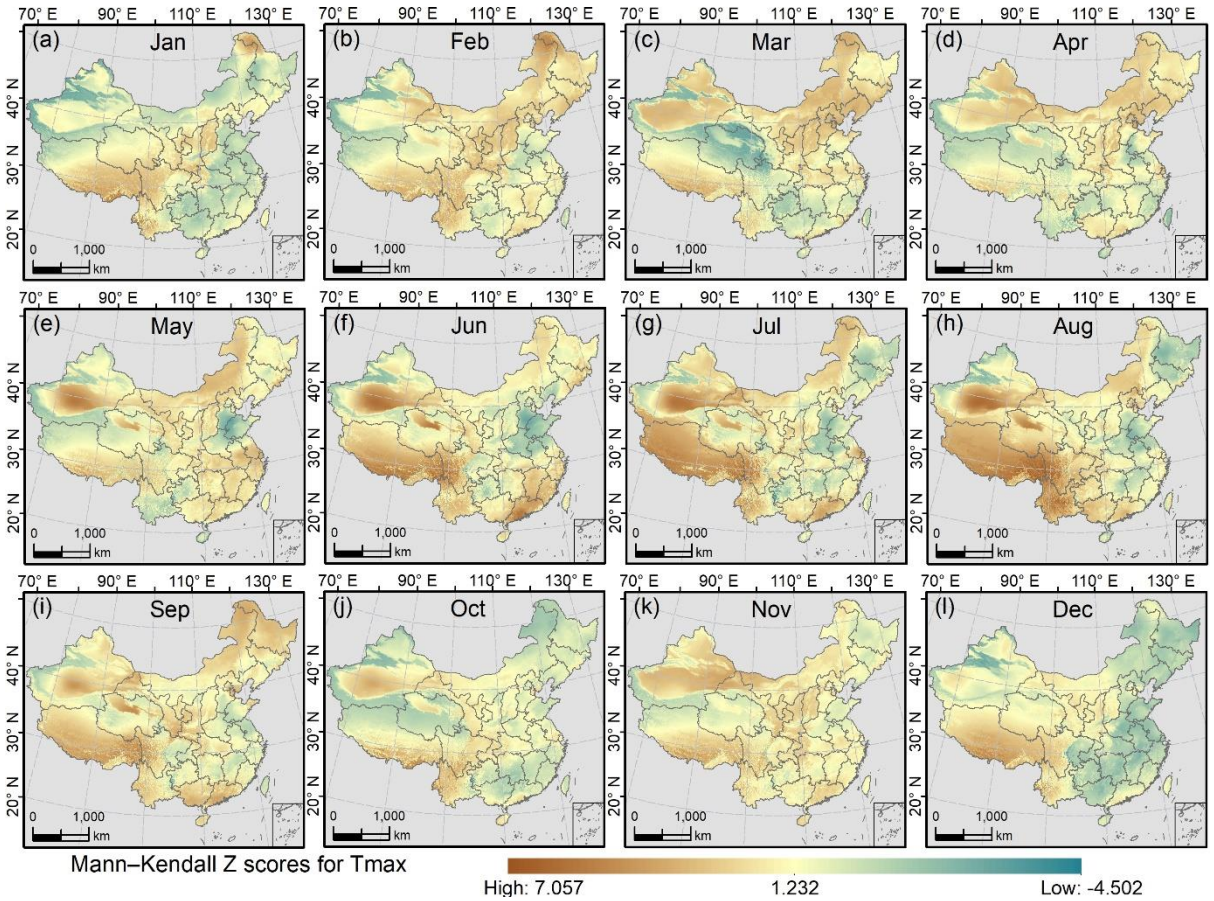

Figure S18: Mann–Kendall Z scores for Tmax over China (1951–2020) in each month.

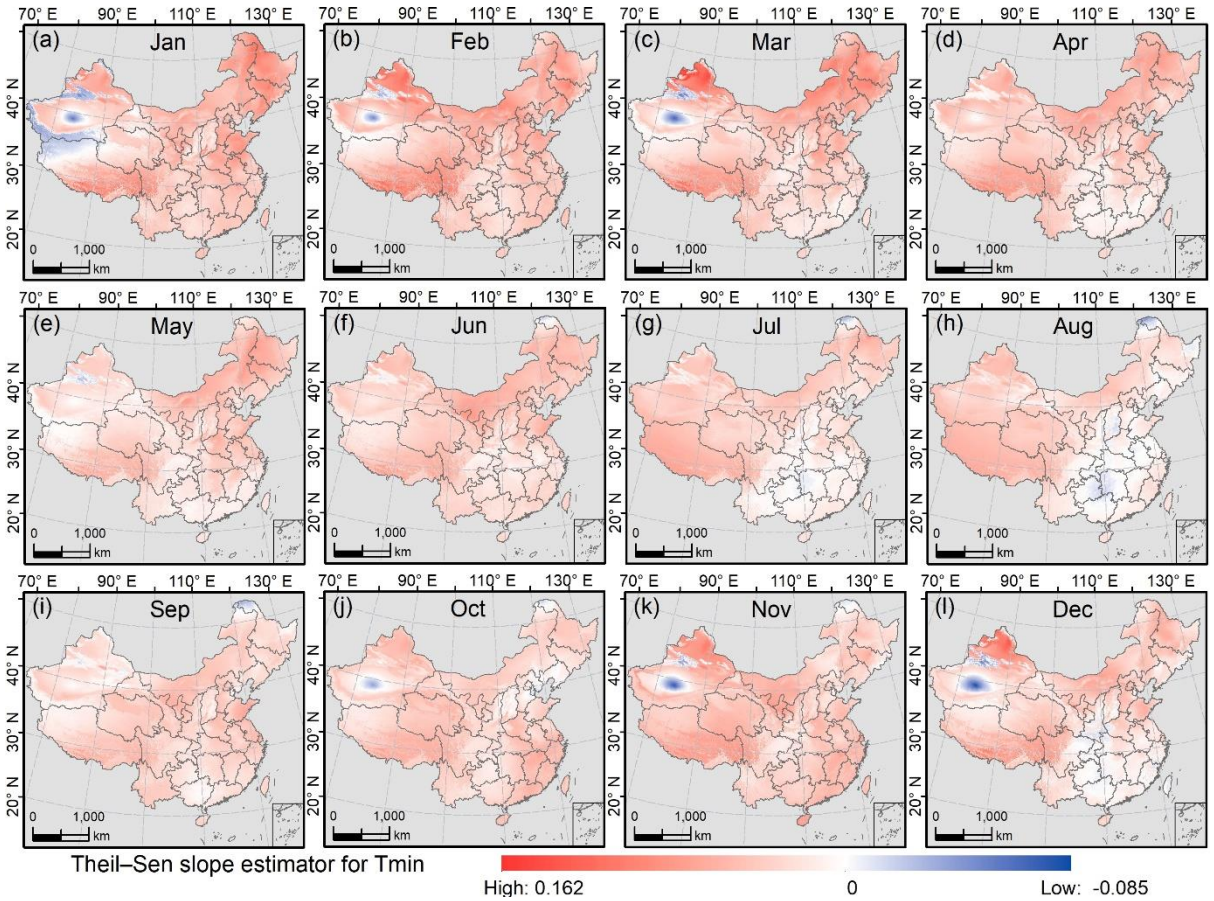

Figure S19: Theil–Sen (median) slope for Tmin over China (1951–2020) in each month.

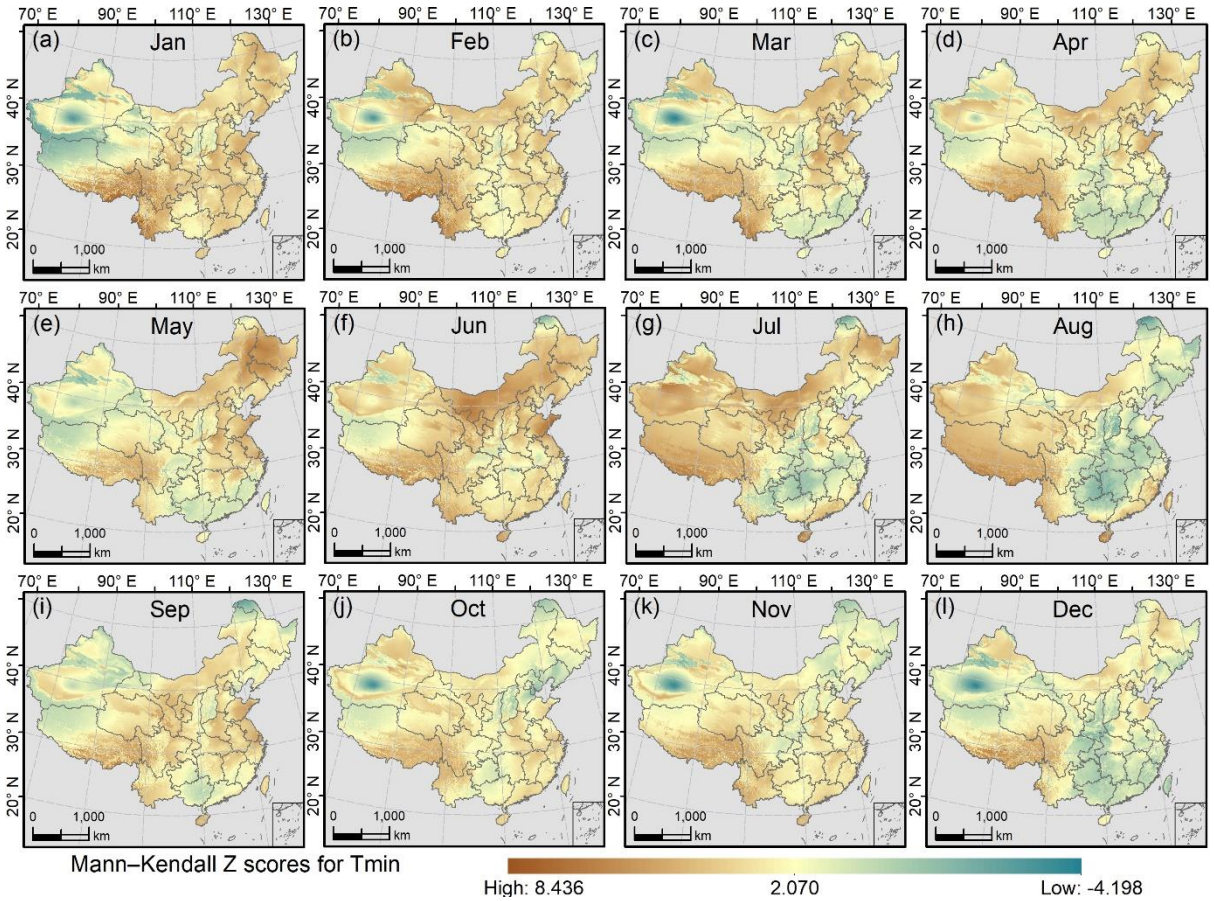

Figure S20: Mann–Kendall Z scores for Tmin over China (1951–2020) in each month.

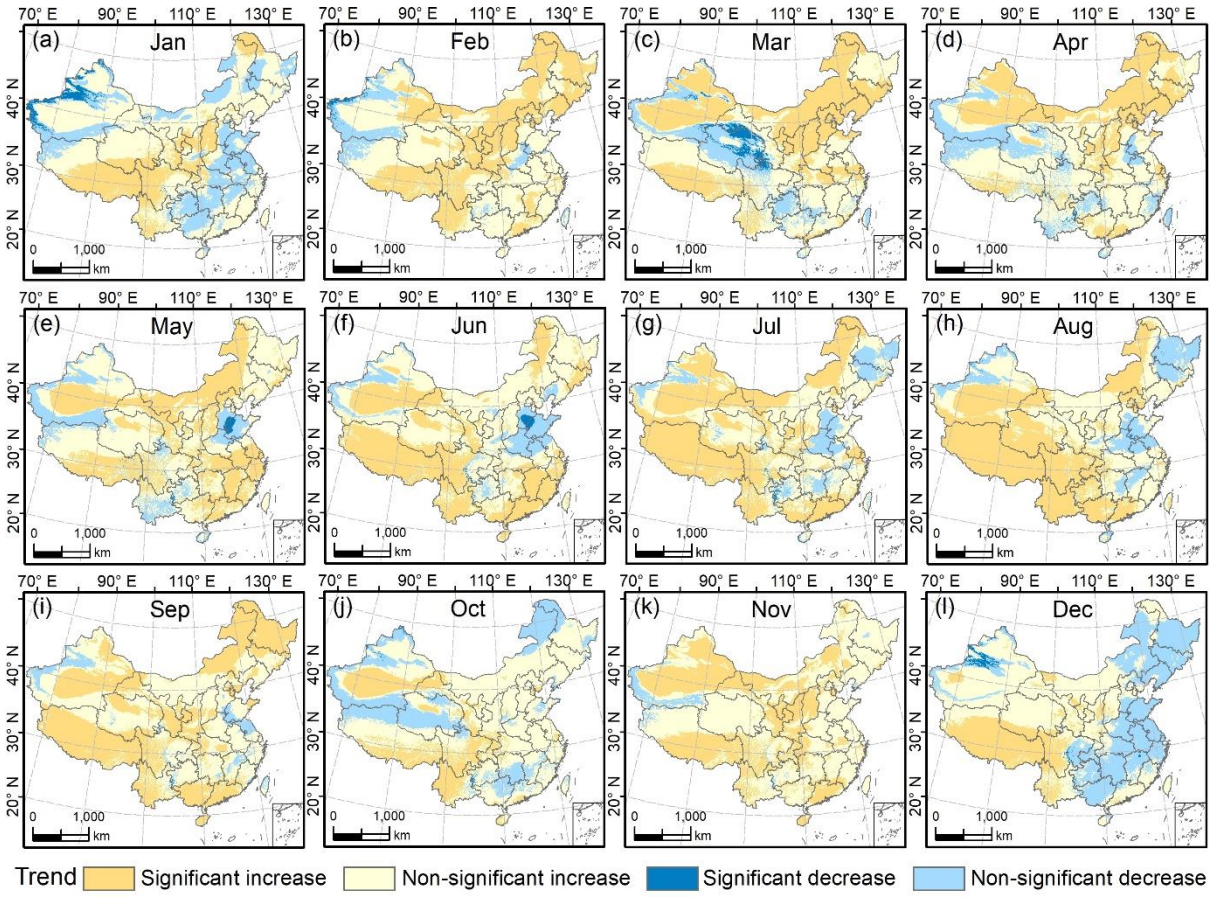

Figure S21: Monthly trends of Tmax change in China during the period 1951-2020 obtained by median Theil–Sen slope. The significance of trends is quantified by the Mann-Kendall statistical test at the 95% confidence level.

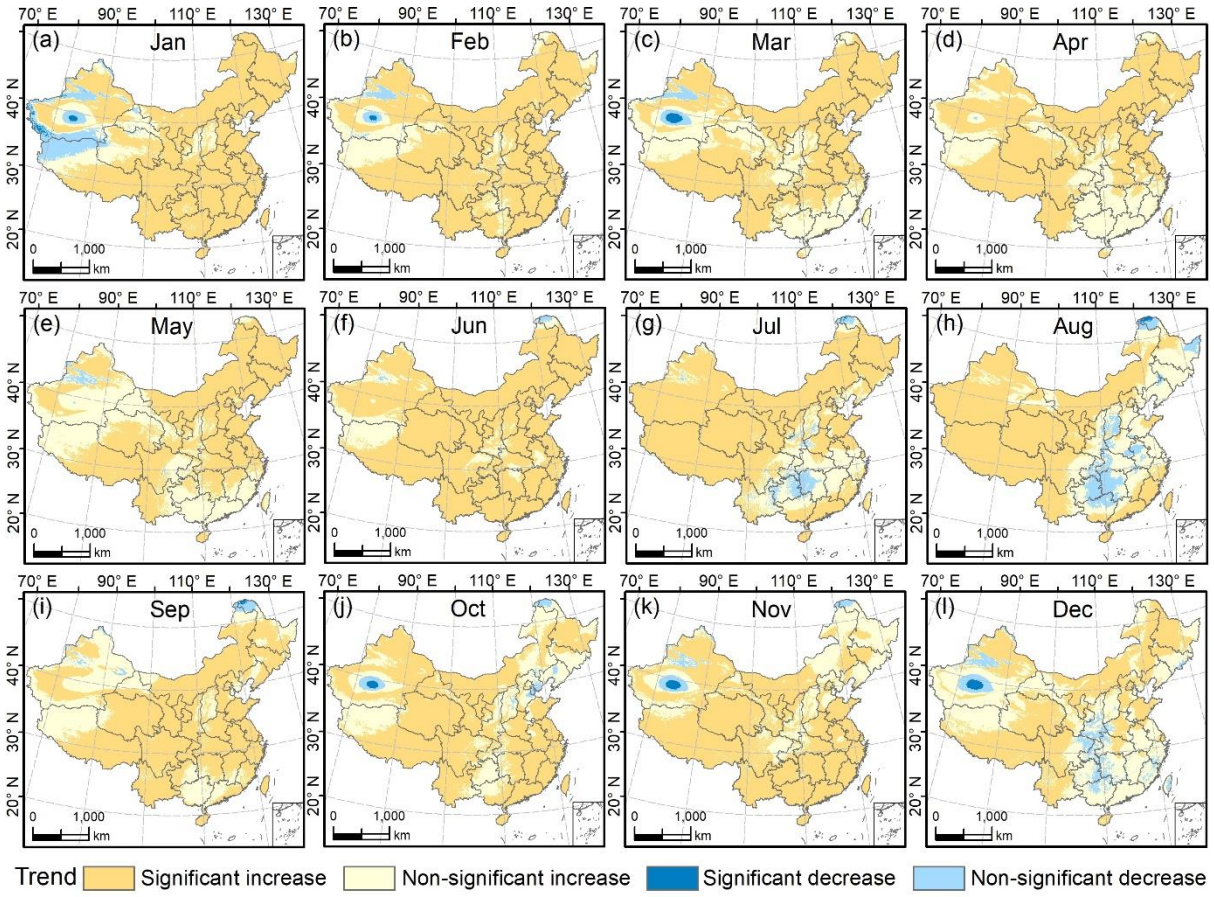

Figure S22: Monthly trends of Tmin change in China during the period 1951-2020 obtained by median Theil–Sen slope. The significance of trends is quantified by the Mann-Kendall statistical test at the 95% confidence level.

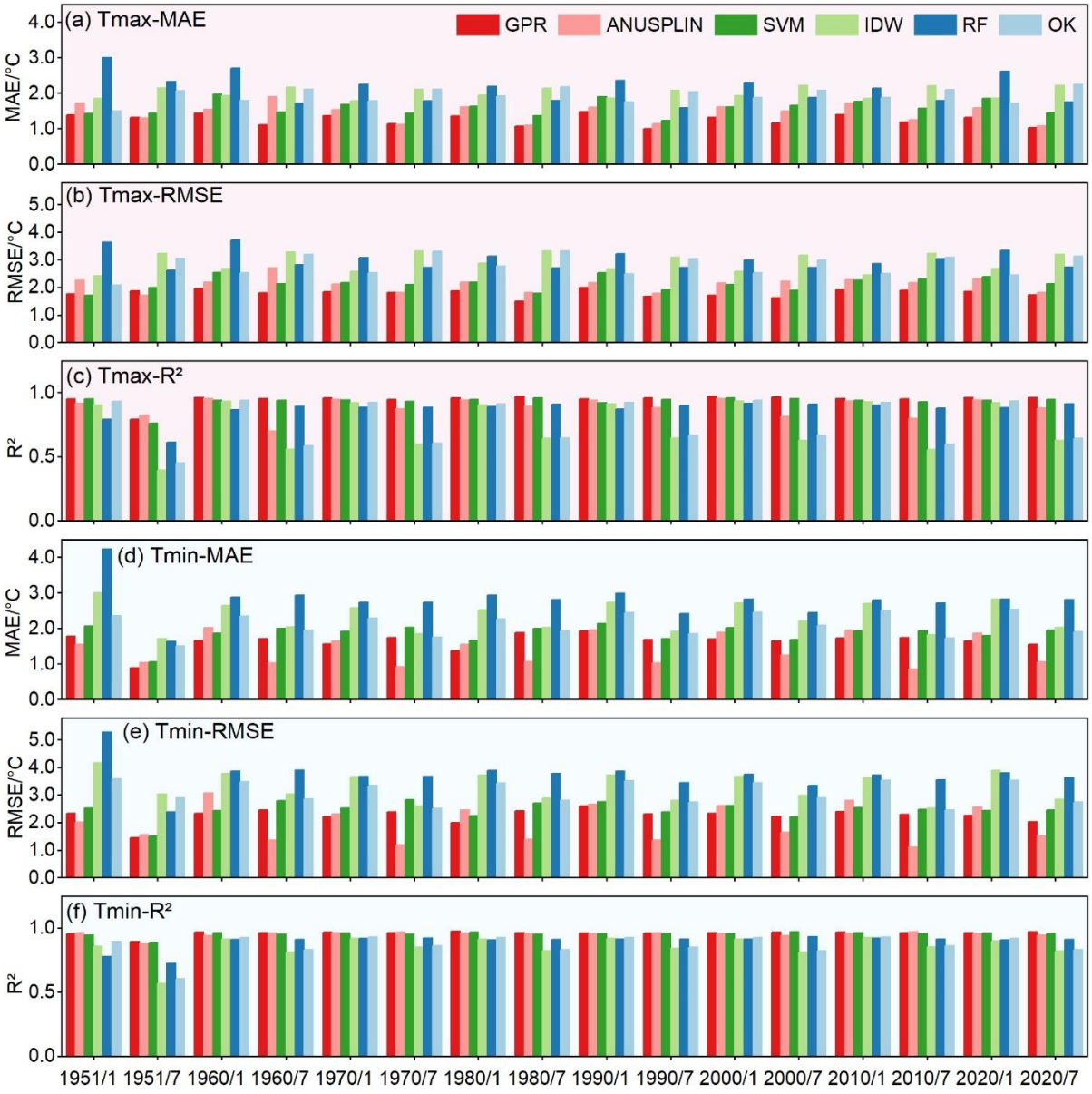

Figure S23: Accuracy of Tmax and Tmin derived by the machine learning methods and traditional methods for January and July from 1951 to 2020 with an interval of 10 years.

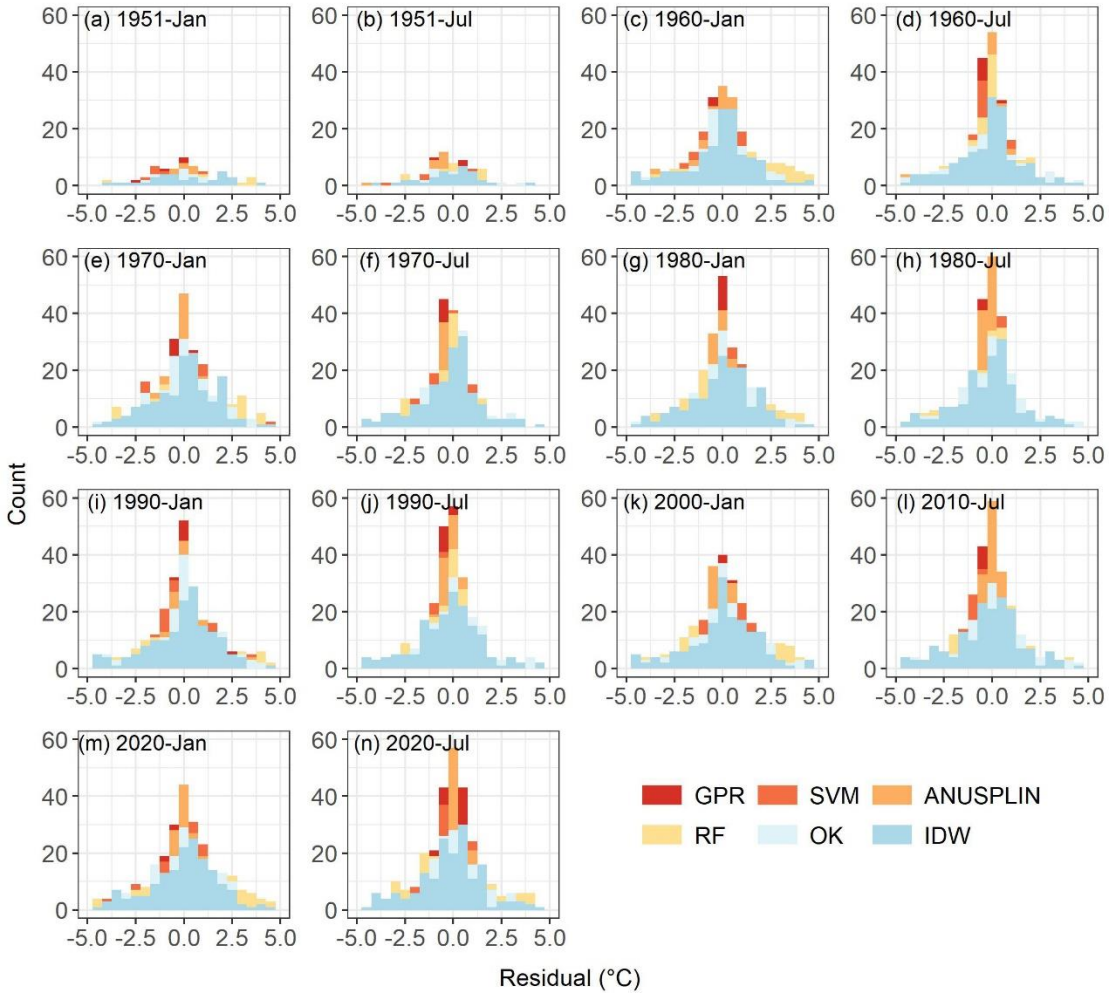

Figure S24: Frequency distribution of residuals for Tmean by three machine learning methods and three traditional methods for separate 8 years from 1951 to 2020 with an interval of 10 years.

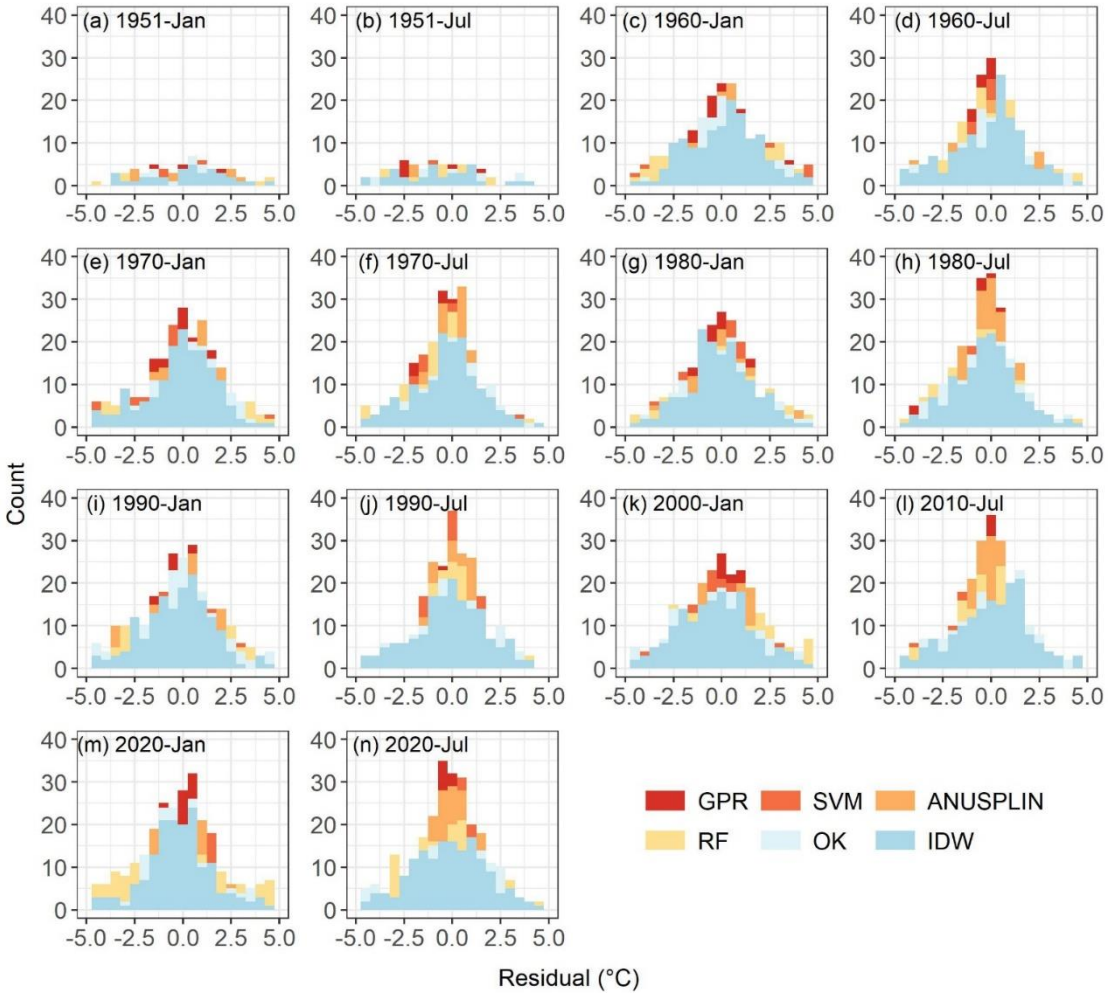

Figure S25: Frequency distribution of residuals for Tmax by three machine learning methods and three traditional methods for separate 8 years from 1951 to 2020 with an interval of 10 years.

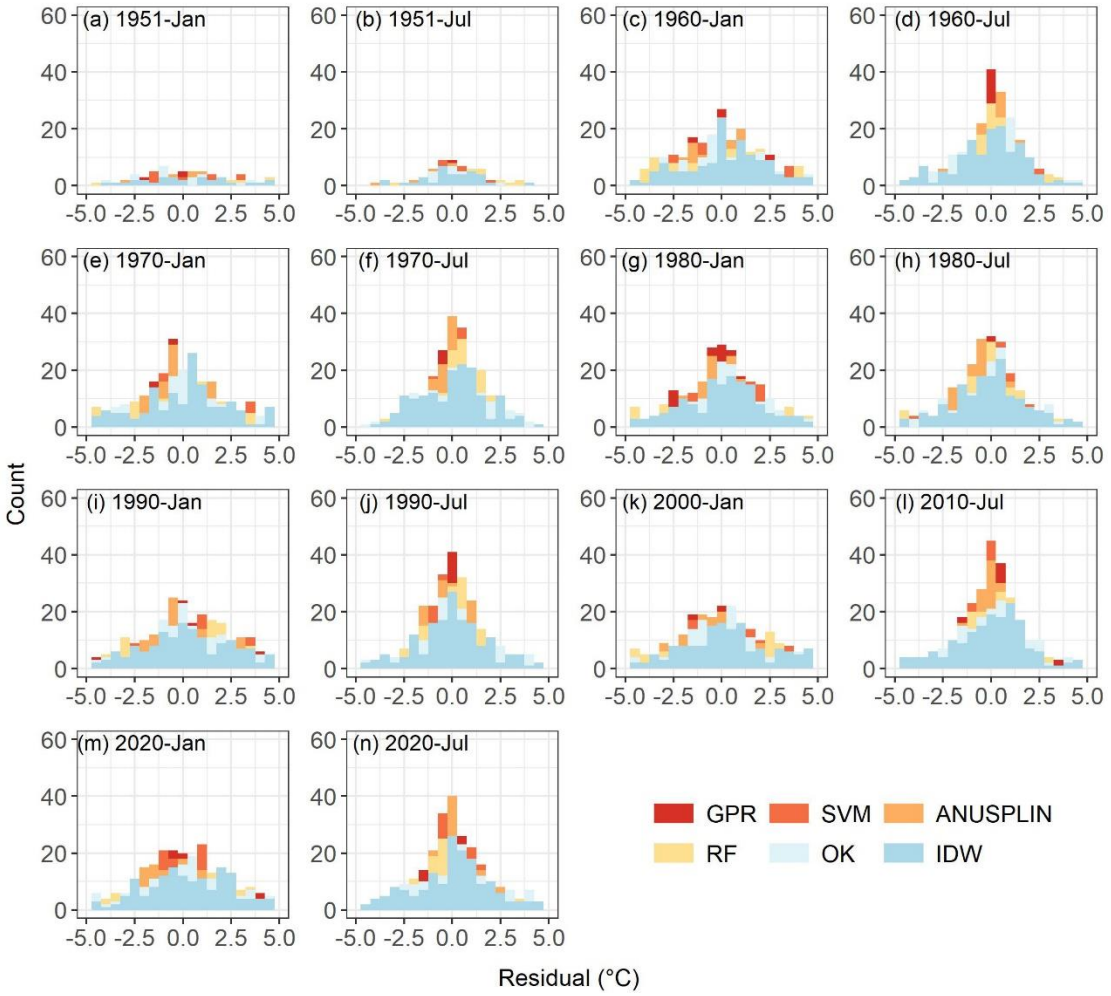

Figure S26: Frequency distribution of residuals for Tmin by three machine learning methods and three traditional methods for separate 8 years from 1951 to 2020 with an interval of 10 years.

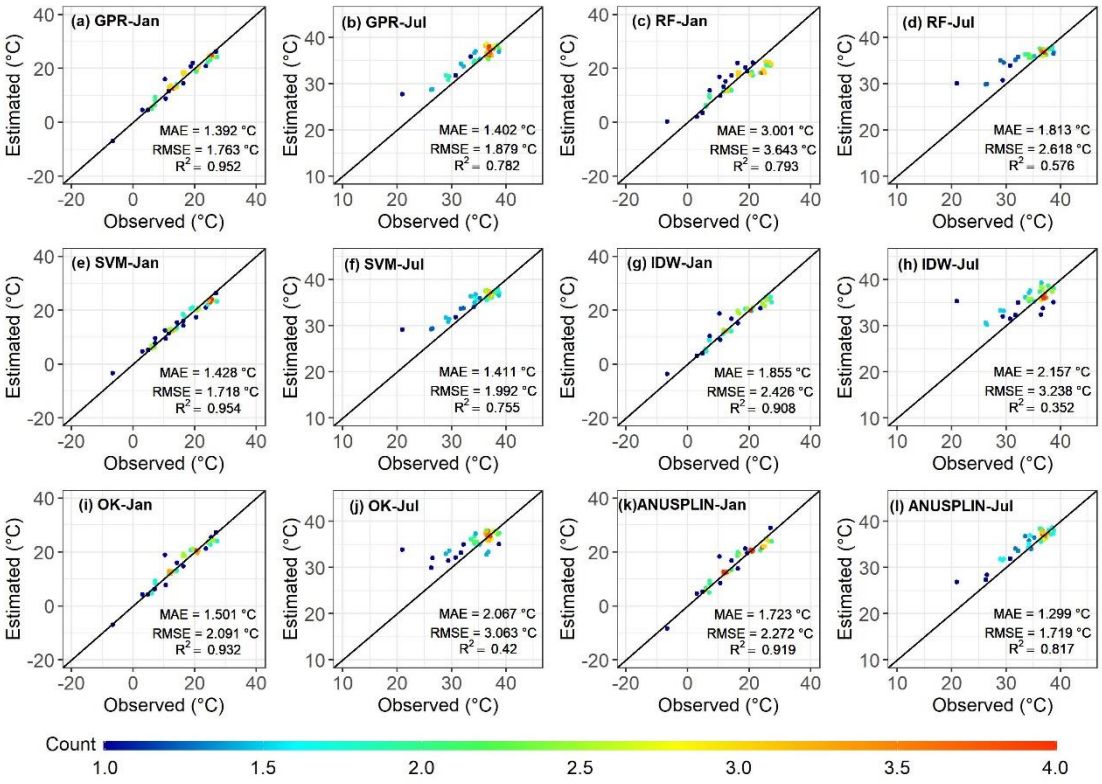

Figure S27: Scatterplot of estimated Tmax by machine learning models and traditional models against the observed monthly mean temperature of January and July in 1951.

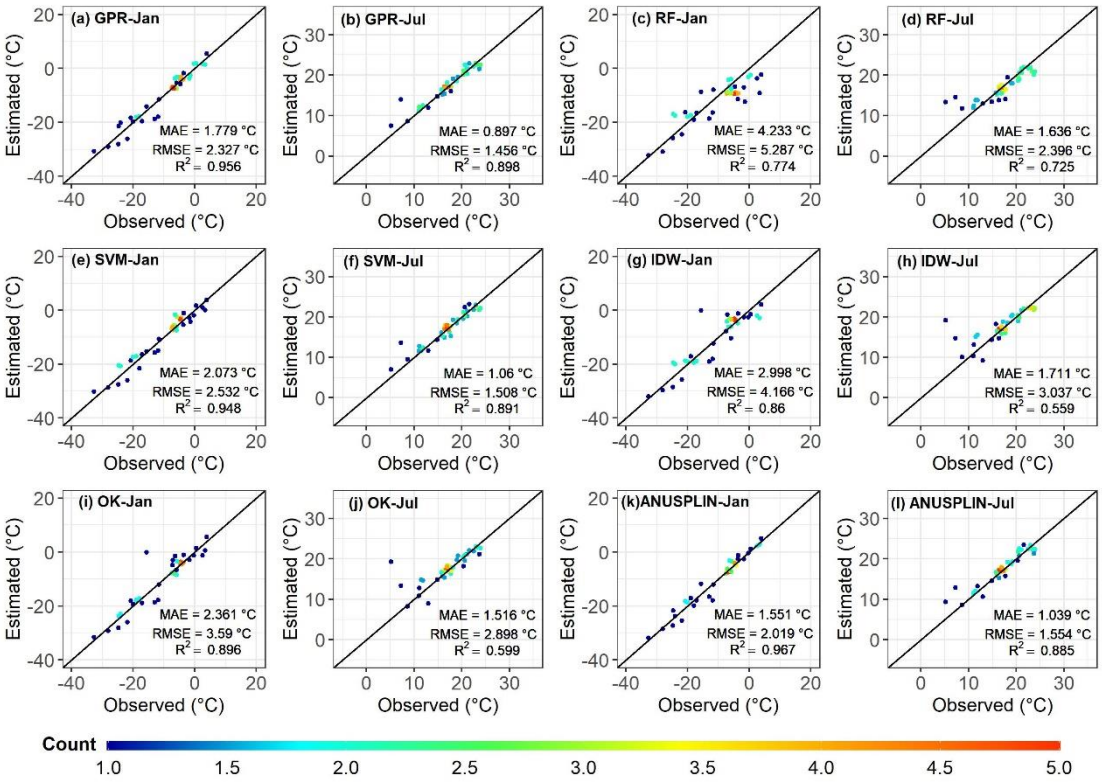

Figure S28: Scatterplot of estimated Tmin by machine learning models and traditional models against the observed monthly mean temperature of January and July in 1951.

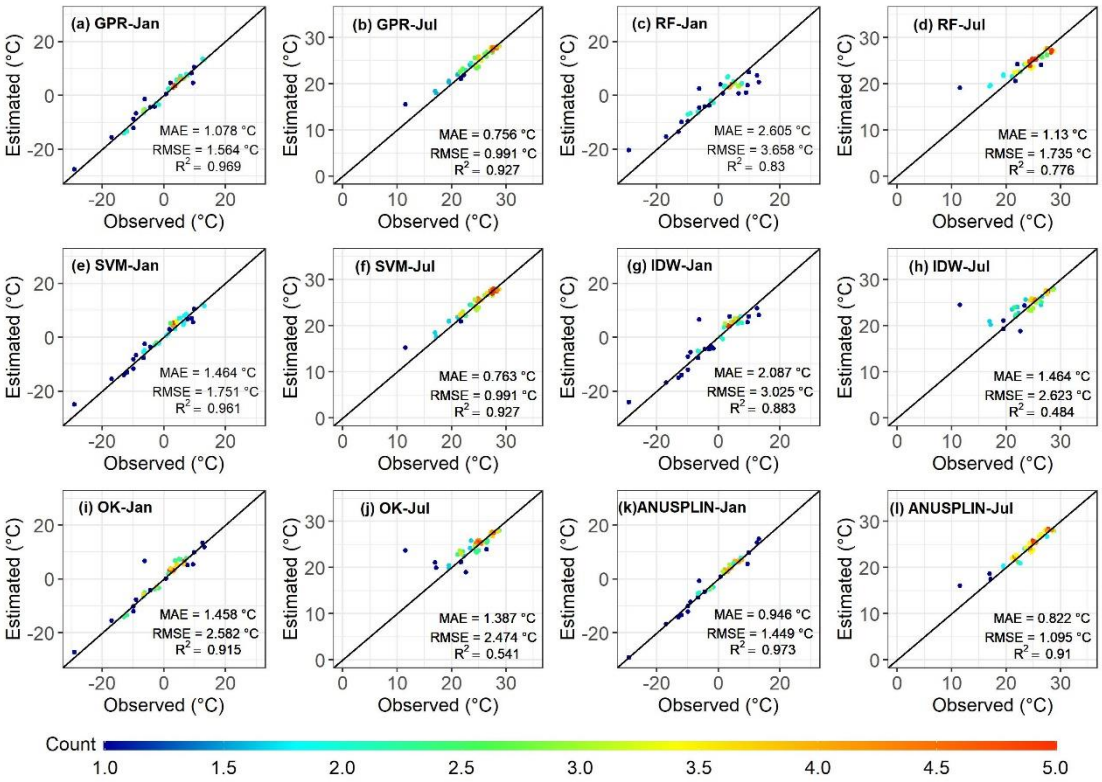

Figure S29: Scatterplot of estimated Tmean by machine learning models and traditional models against the observed monthly mean temperature of January and July in 1951.

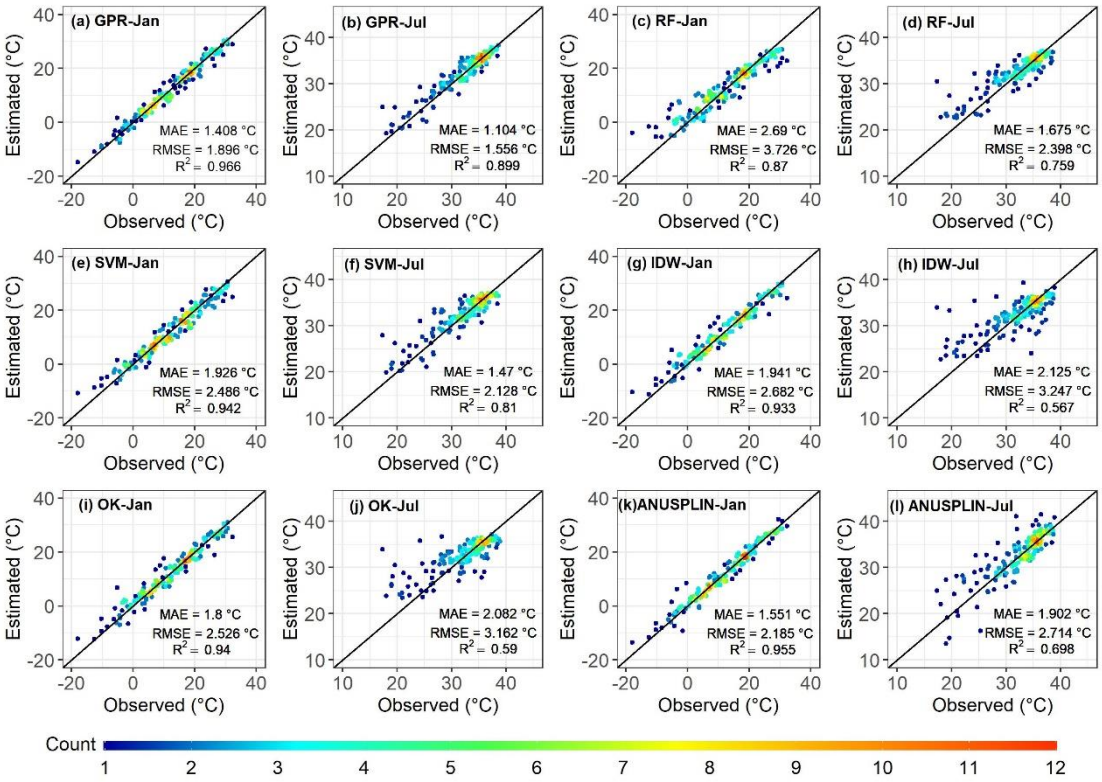

Figure S30: Scatterplot of estimated Tmax by machine learning models and traditional models against the observed monthly mean temperature of January and July in 1960.

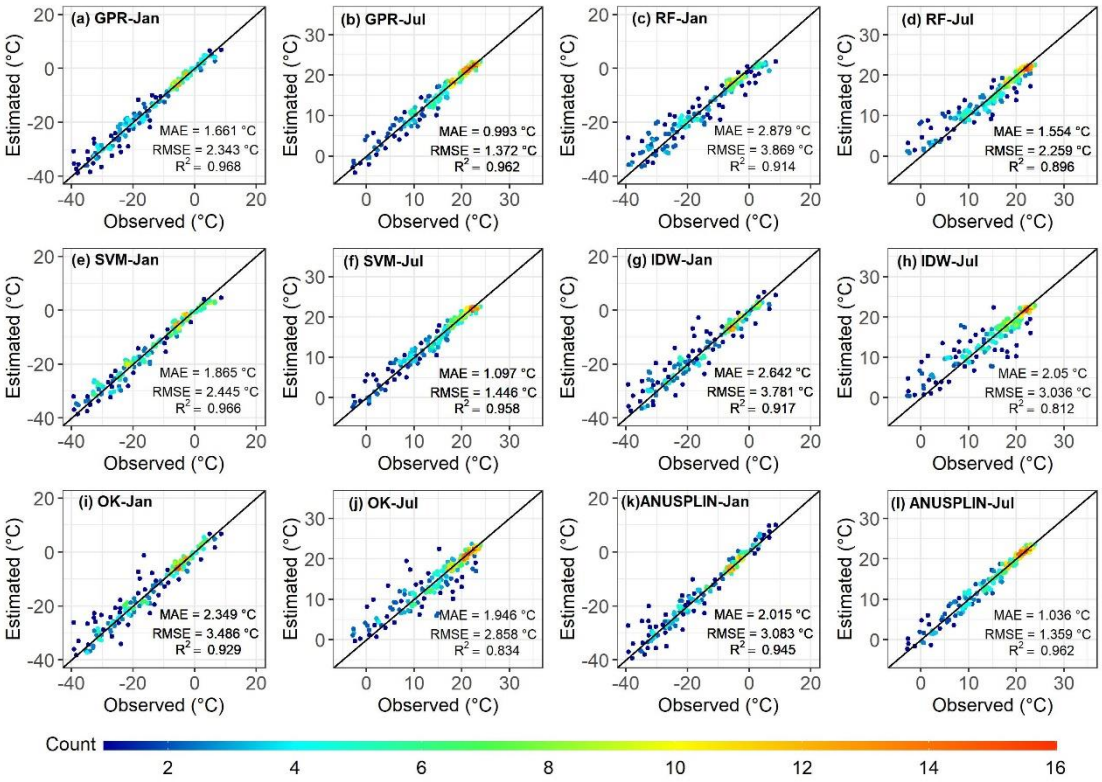

Figure S31: Scatterplot of estimated Tmin by machine learning models and traditional models against the observed monthly mean temperature of January and July in 1960.

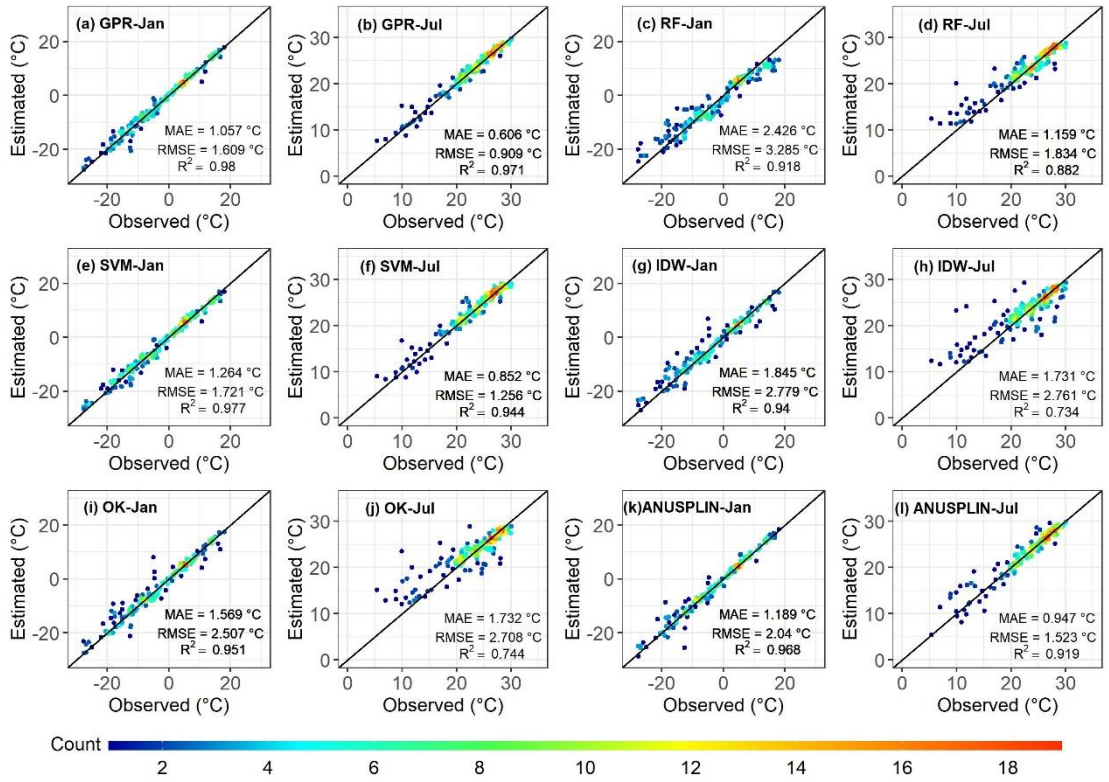

Figure S32: Scatterplot of estimated Tmean by machine learning models and traditional models against the observed monthly mean temperature of January and July in 1960.

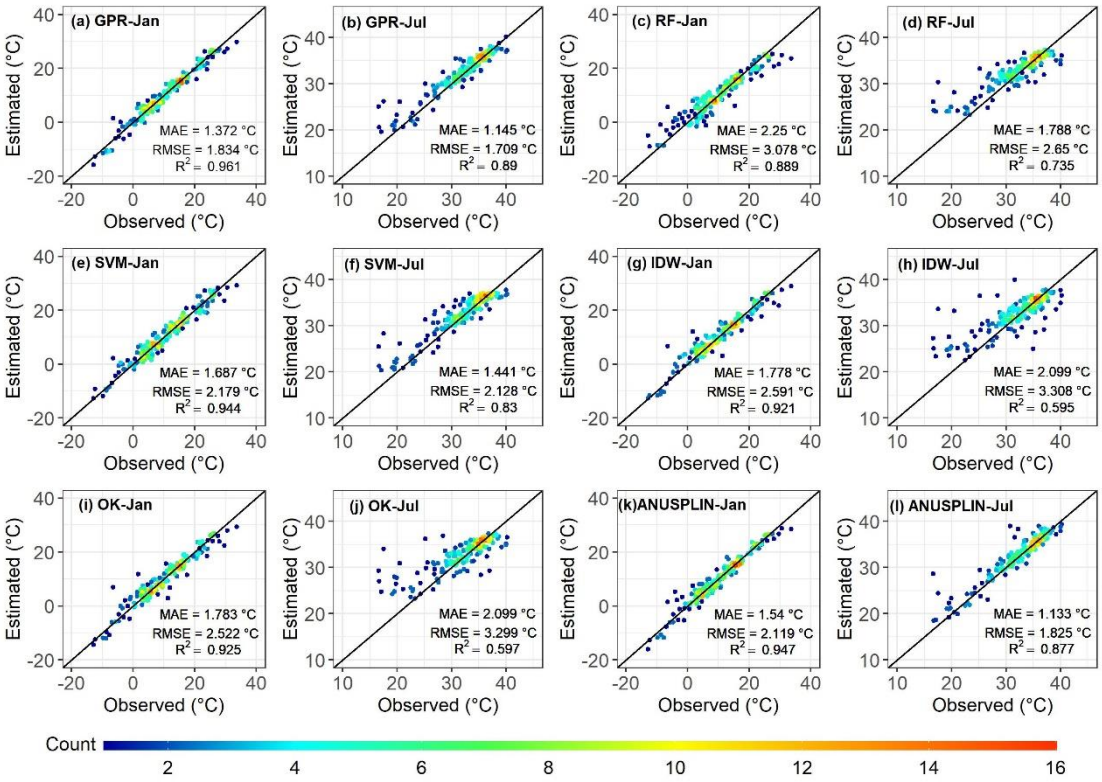

Figure S33: Scatterplot of estimated Tmax by machine learning models and traditional models against the observed monthly mean temperature of January and July in 1970.

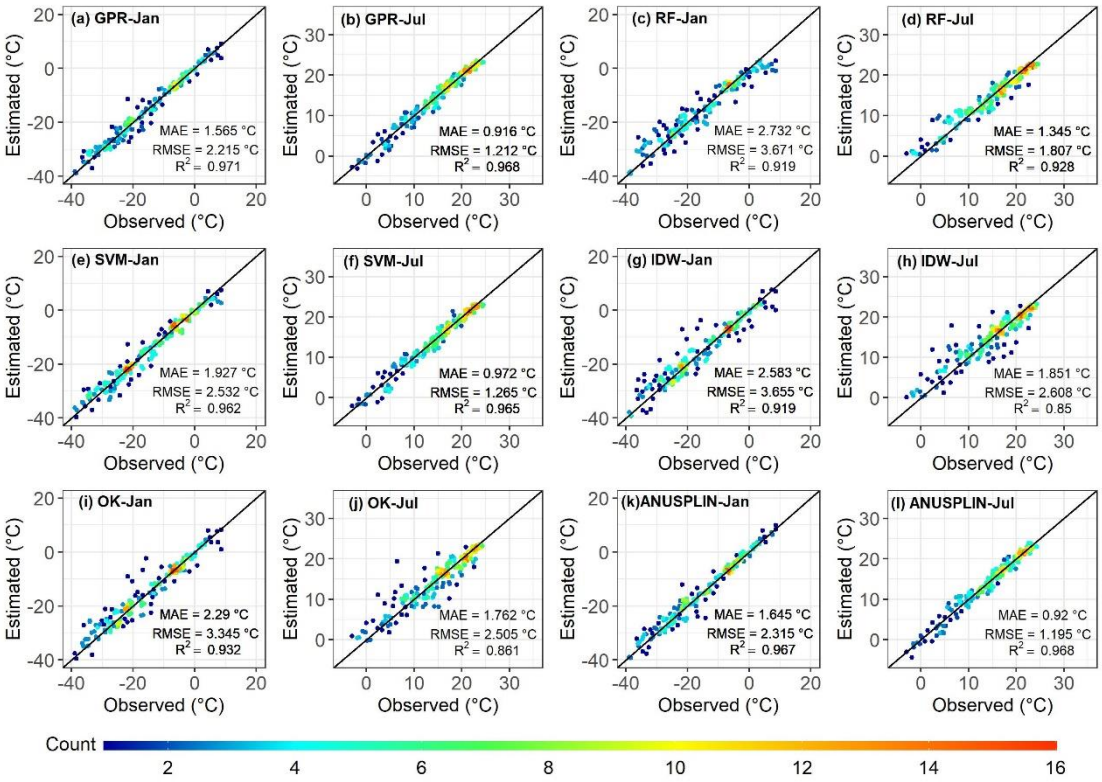

Figure S34: Scatterplot of estimated Tmin by machine learning models and traditional models against the observed monthly mean temperature of January and July in 1970.

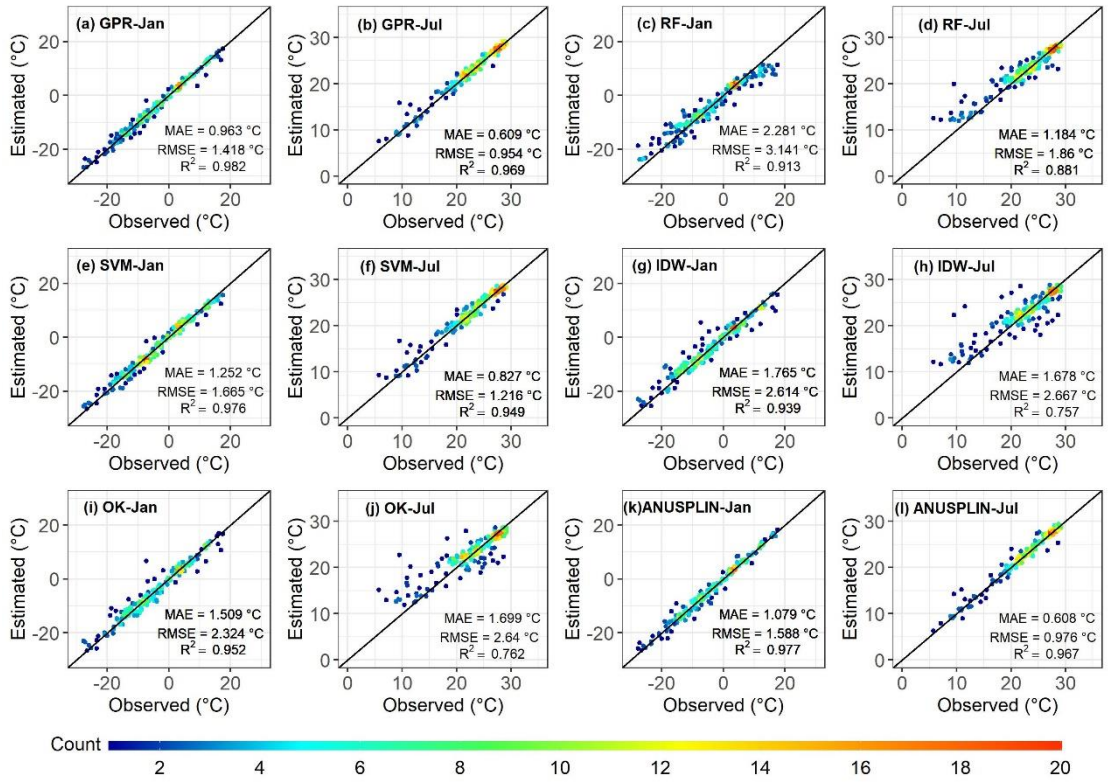

Figure S35: Scatterplot of estimated Tmean by machine learning models and traditional models against the observed monthly mean temperature of January and July in 1970.

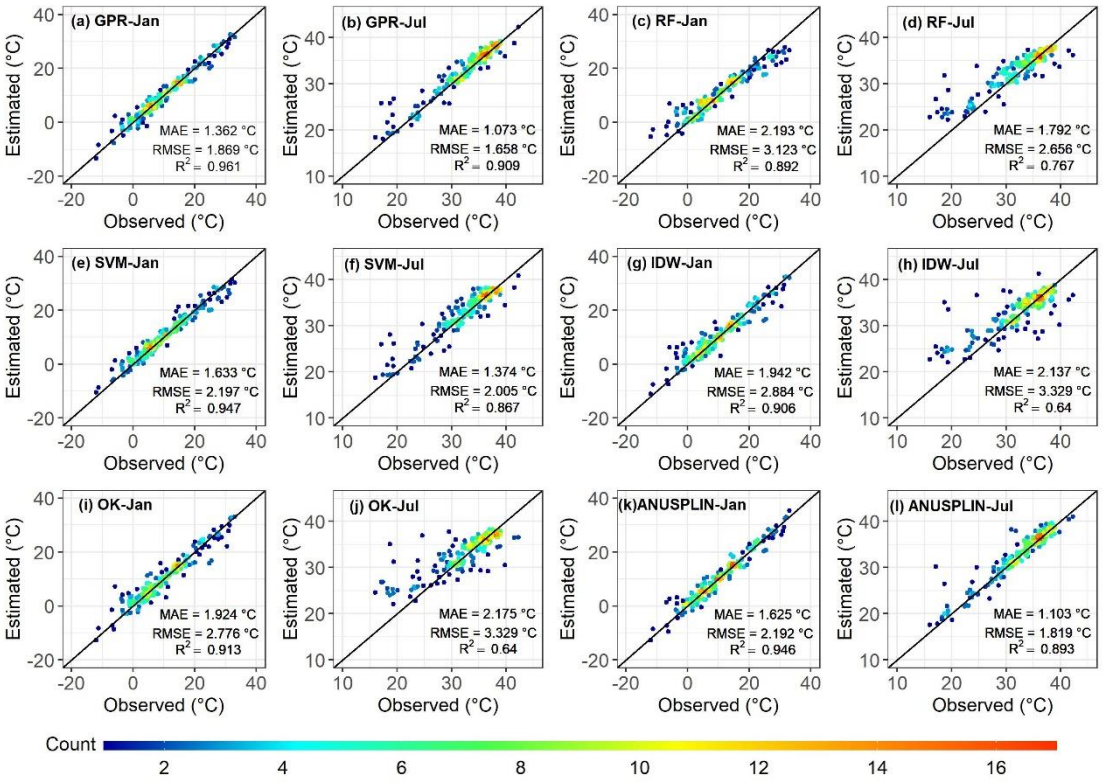

Figure S36: Scatterplot of estimated Tmax by machine learning models and traditional models against the observed monthly mean temperature of January and July in 1980.

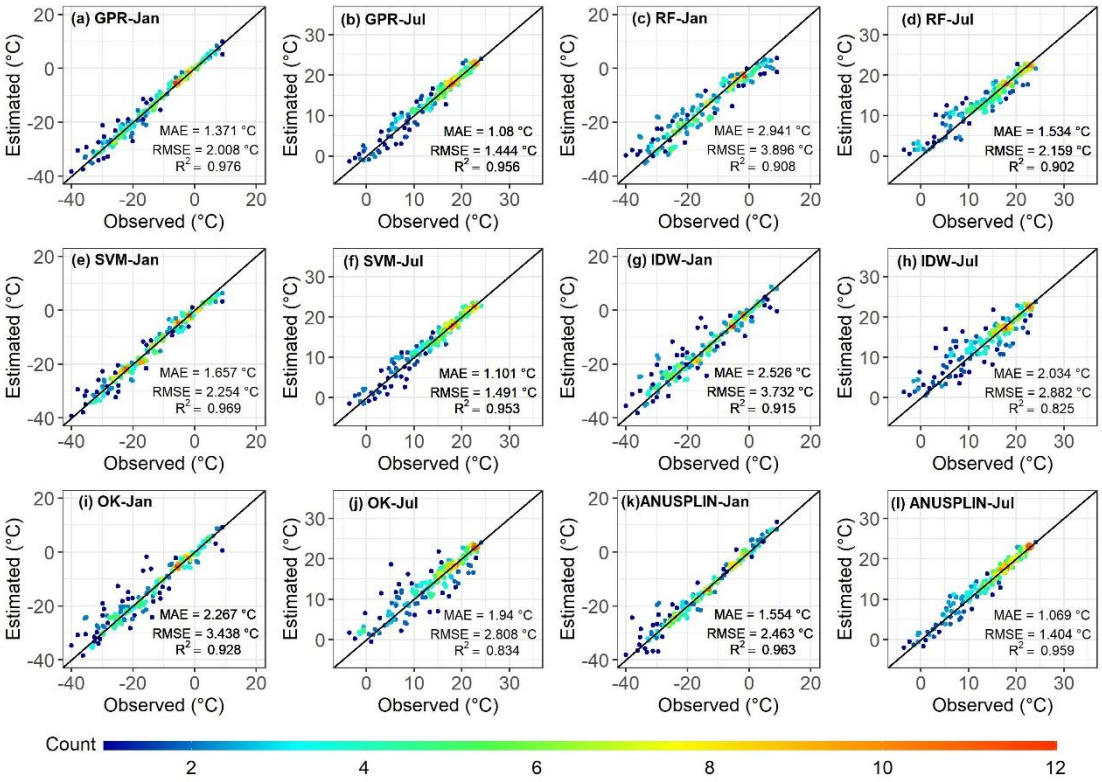

Figure S37: Scatterplot of estimated Tmin by machine learning models and traditional models against the observed monthly mean temperature of January and July in 1980.

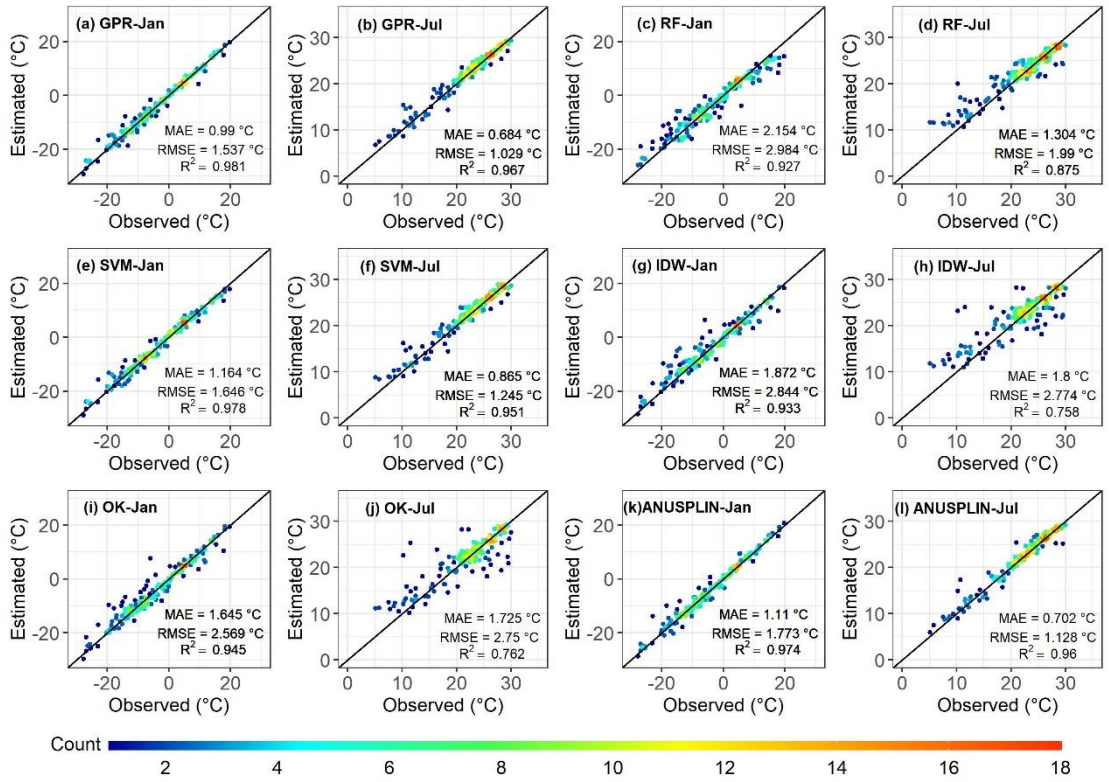

Figure S38: Scatterplot of estimated Tmean by machine learning models and traditional models against the observed monthly mean temperature of January and July in 1980.

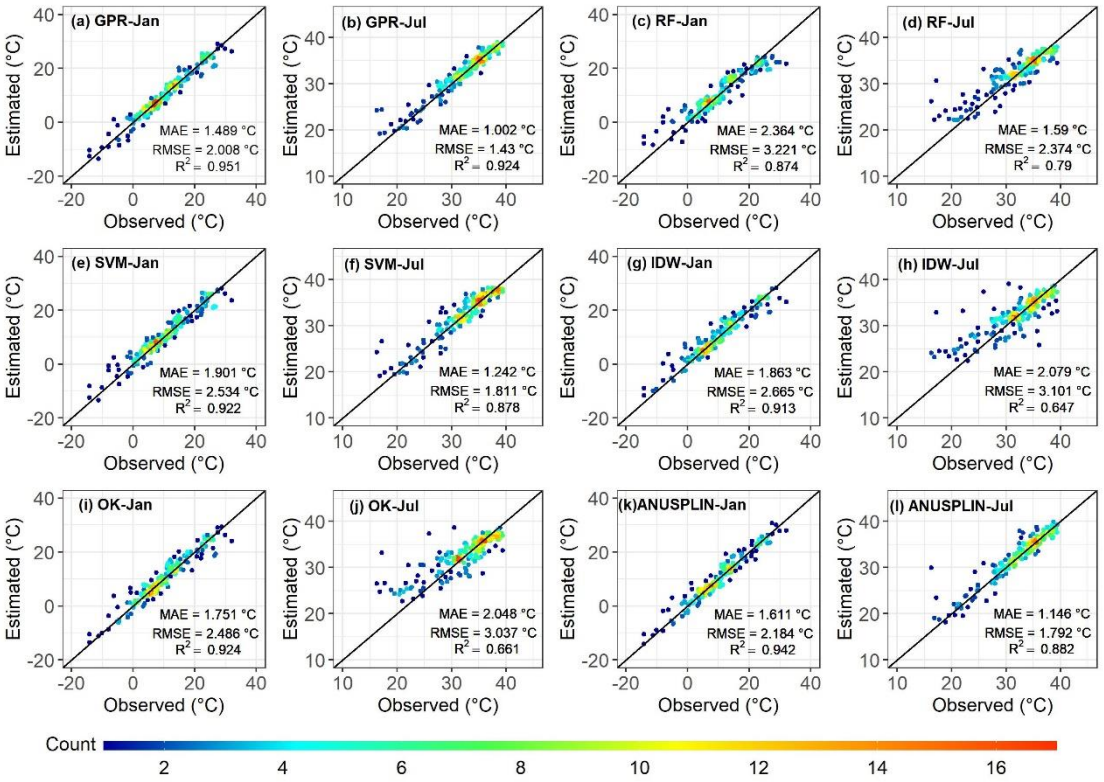

Figure S39: Scatterplot of estimated Tmax by machine learning models and traditional models against the observed monthly mean temperature of January and July in 1990.

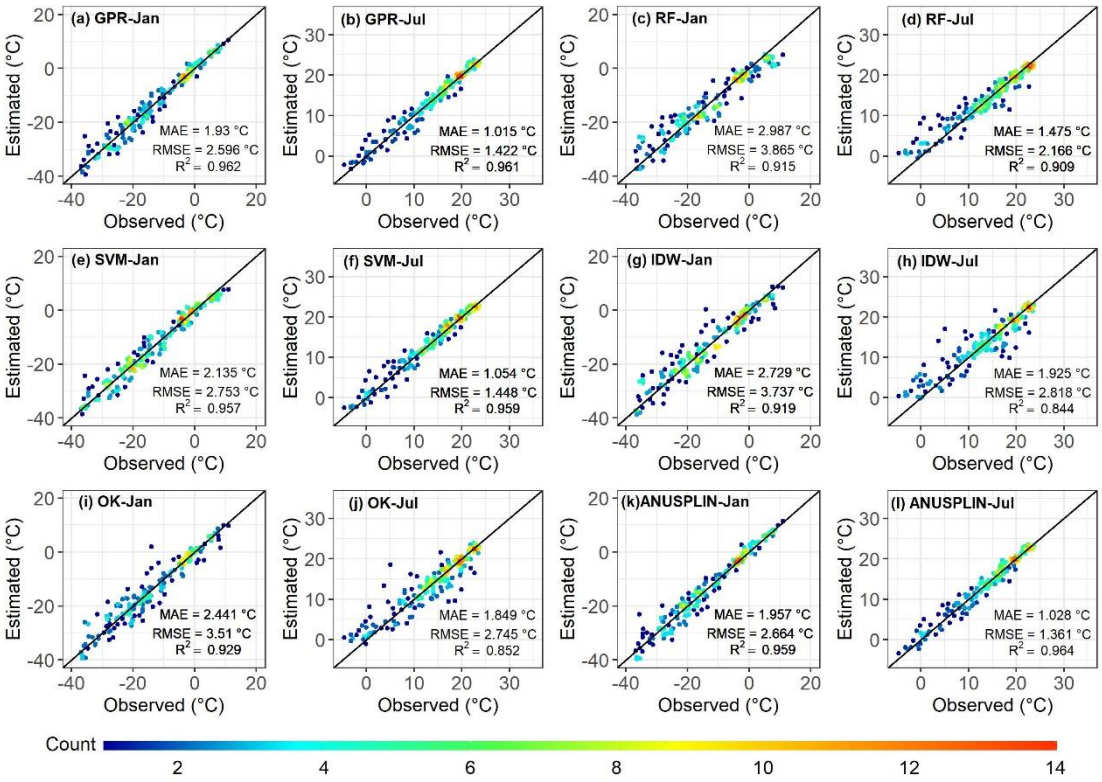

Figure S40: Scatterplot of estimated Tmin by machine learning models and traditional models against the observed monthly mean temperature of January and July in 1990.

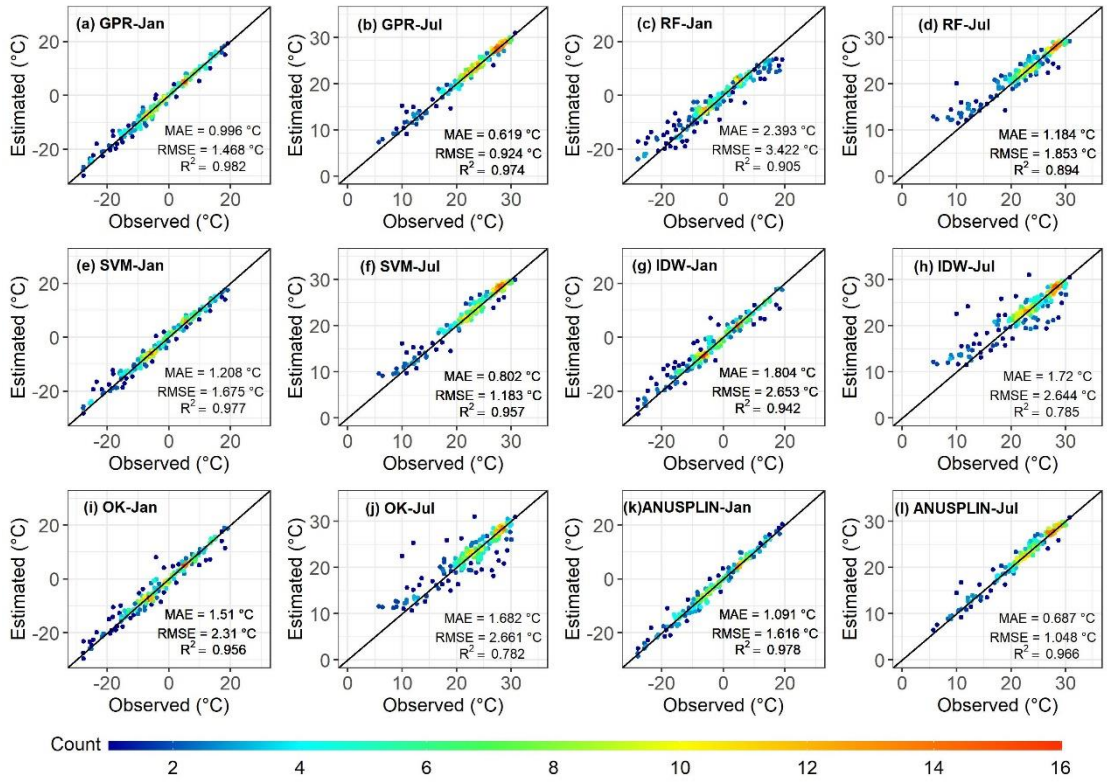

Figure S41: Scatterplot of estimated Tmean by machine learning models and traditional models against the observed monthly mean temperature of January and July in 1990.

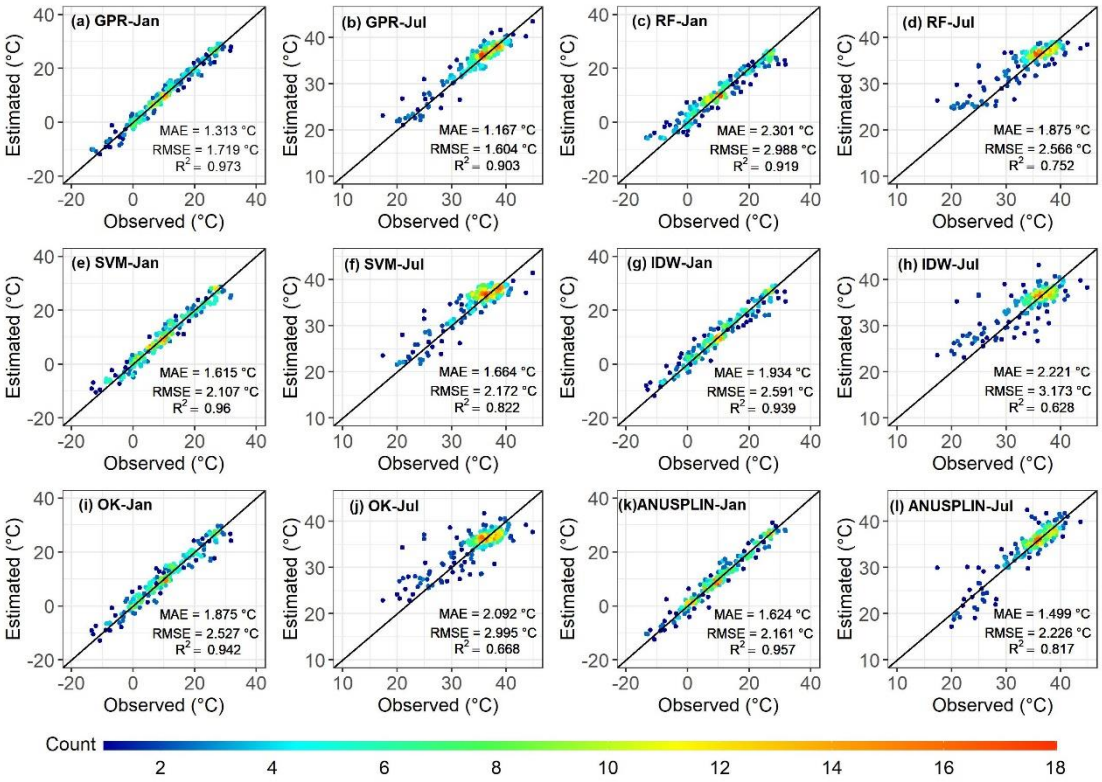

Figure S42: Scatterplot of estimated Tmax by machine learning models and traditional models against the observed monthly mean temperature of January and July in 2000.

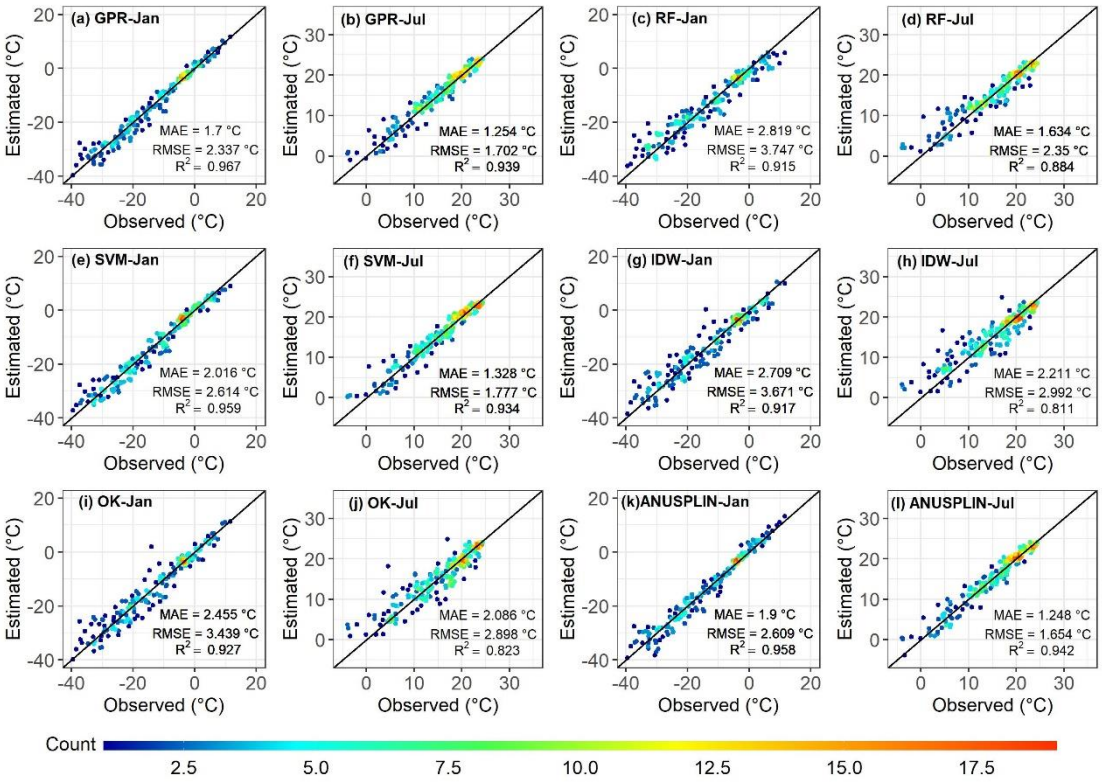

Figure S43: Scatterplot of estimated Tmin by machine learning models and traditional models against the observed monthly mean temperature of January and July in 2000.

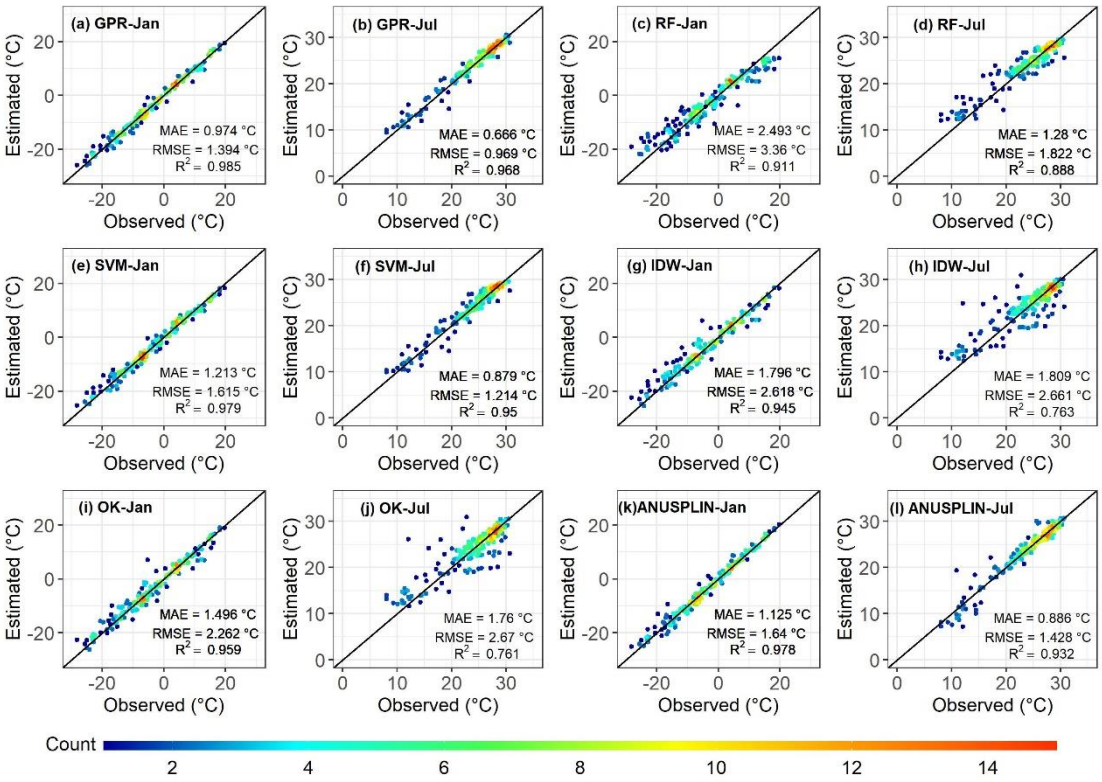

Figure S44: Scatterplot of estimated Tmean by machine learning models and traditional models against the observed monthly mean temperature of January and July in 2000.

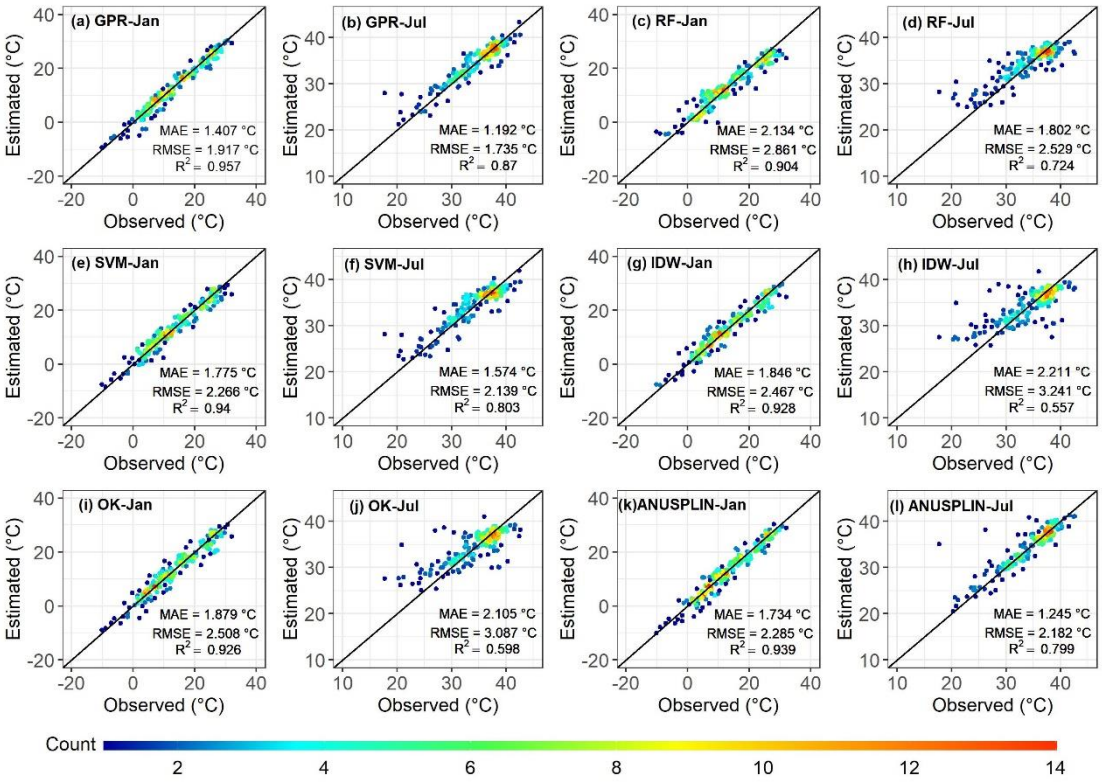

Figure S45: Scatterplot of estimated Tmax by machine learning models and traditional models against the observed monthly mean temperature of January and July in 2010.

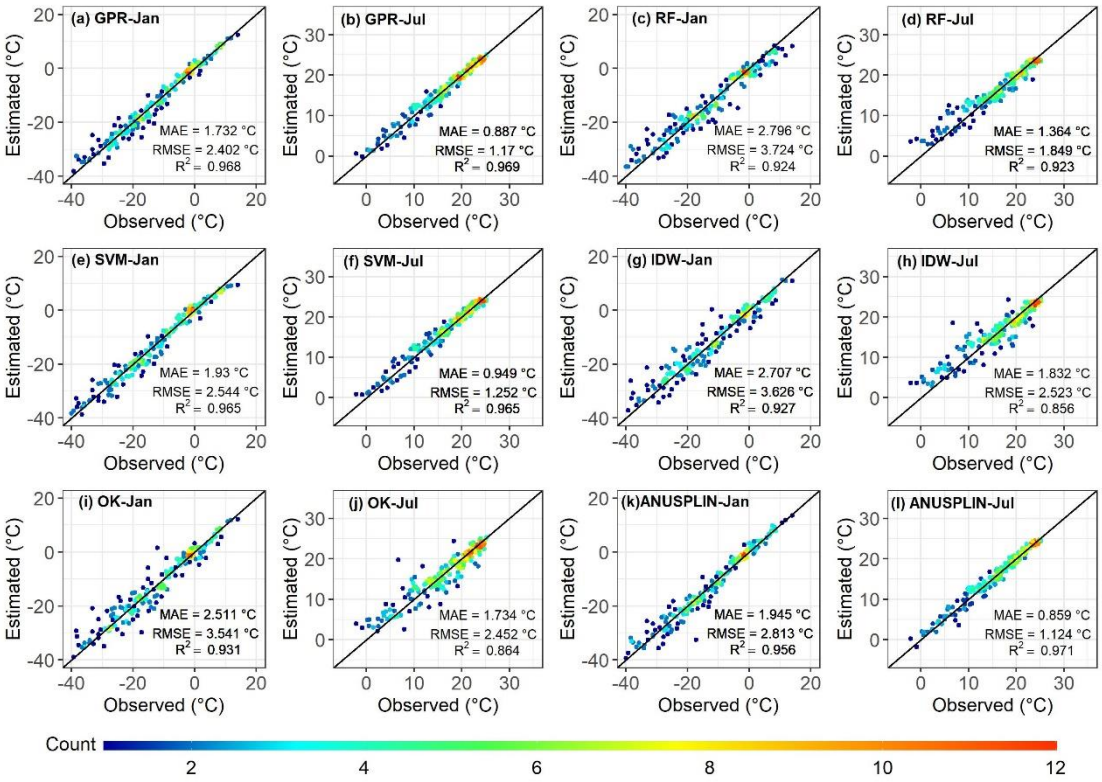

Figure S46: Scatterplot of estimated Tmin by machine learning models and traditional models against the observed monthly mean temperature of January and July in 2010.

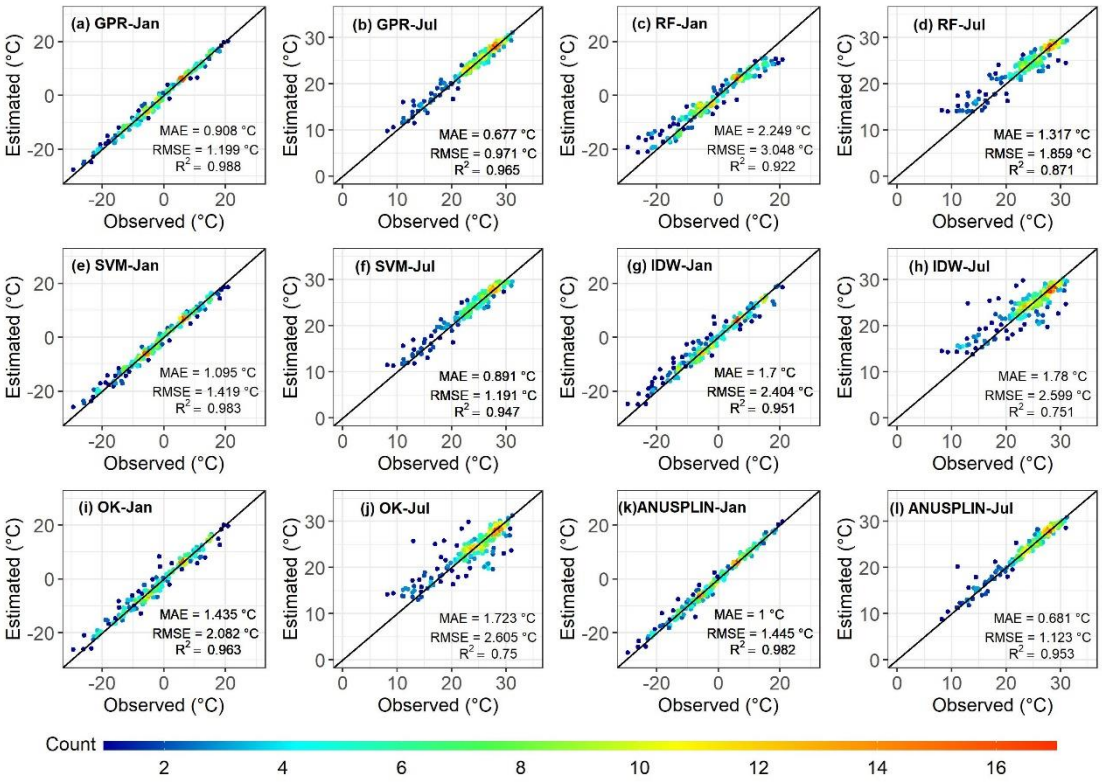

Figure S47: Scatterplot of estimated Tmean by machine learning models and traditional models against the observed monthly mean temperature of January and July in 2010.

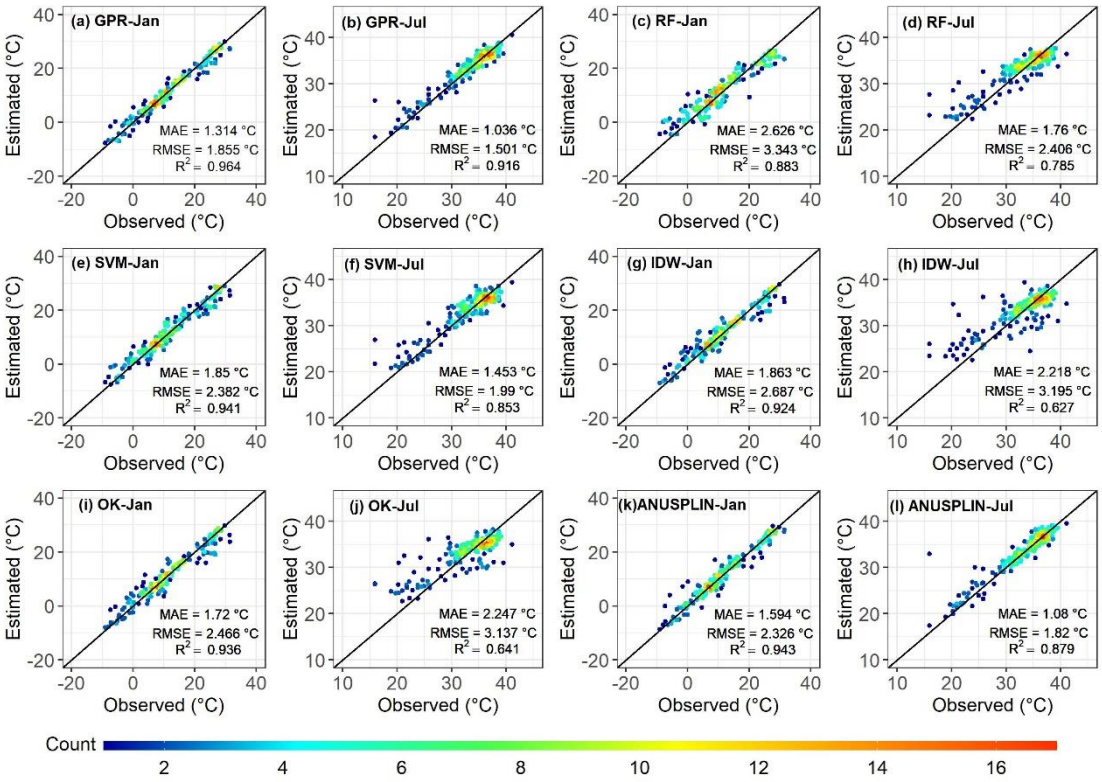

Figure S48: Scatterplot of estimated Tmax by machine learning models and traditional models against the observed monthly mean temperature of January and July in 2020.

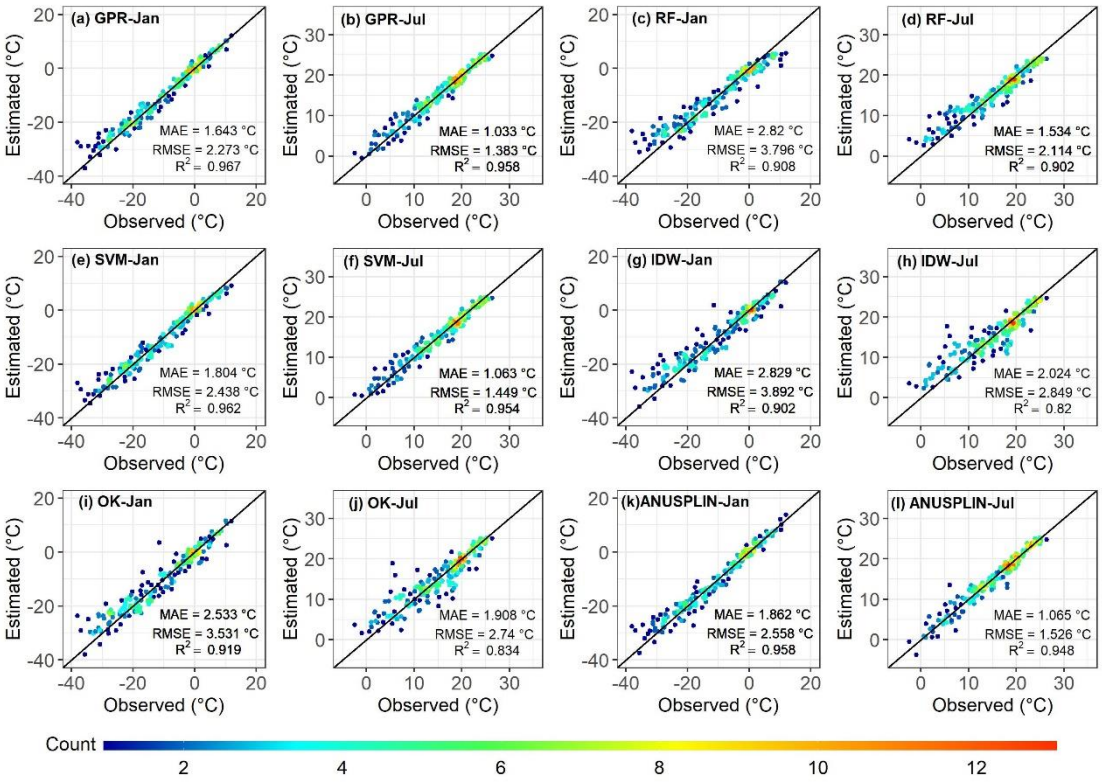

Figure S49: Scatterplot of estimated Tmin by machine learning models and traditional models against the observed monthly mean temperature of January and July in 2010.

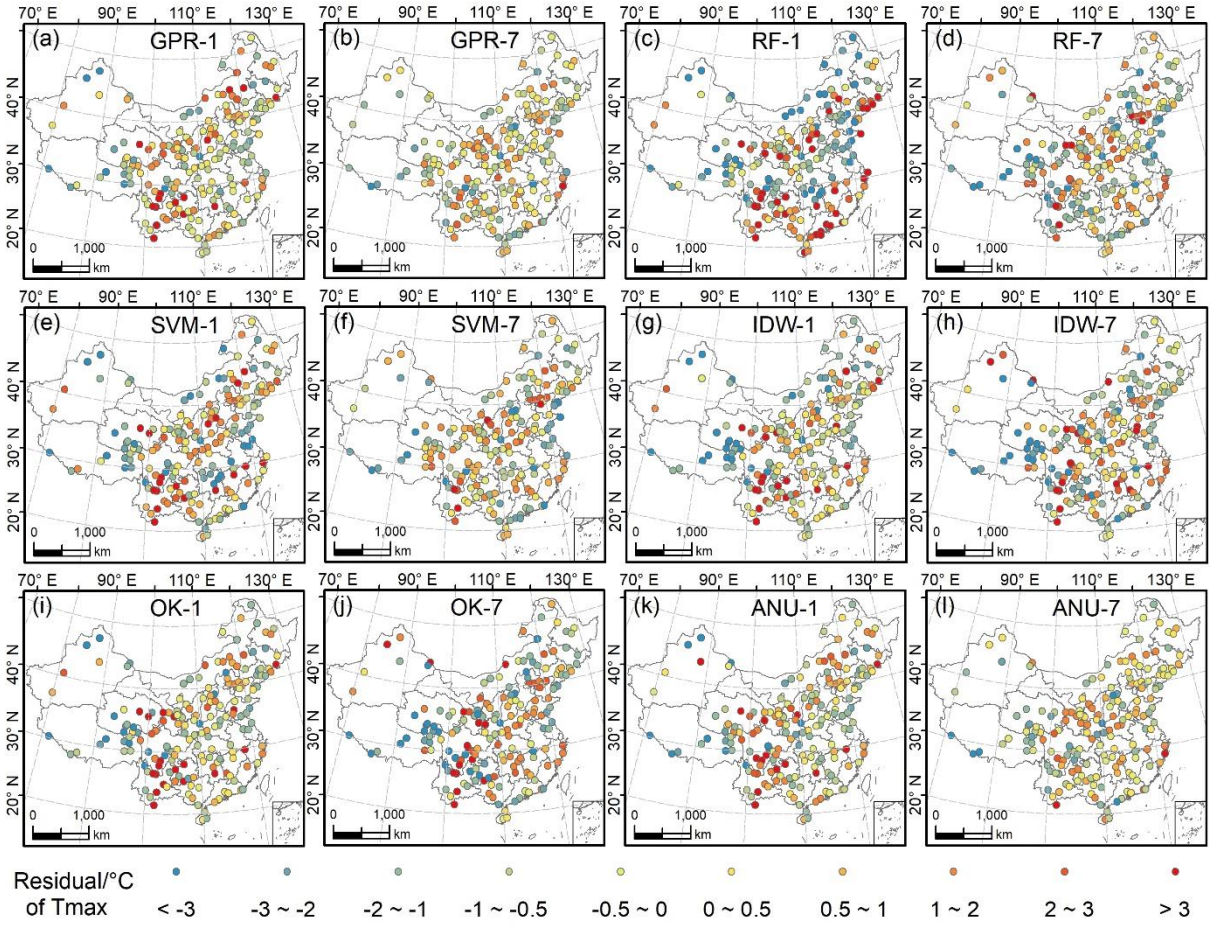

Figure S50: Comparison of the spatial distribution of residuals between machine learning and traditional methods for Tmax in January and July of 2020.

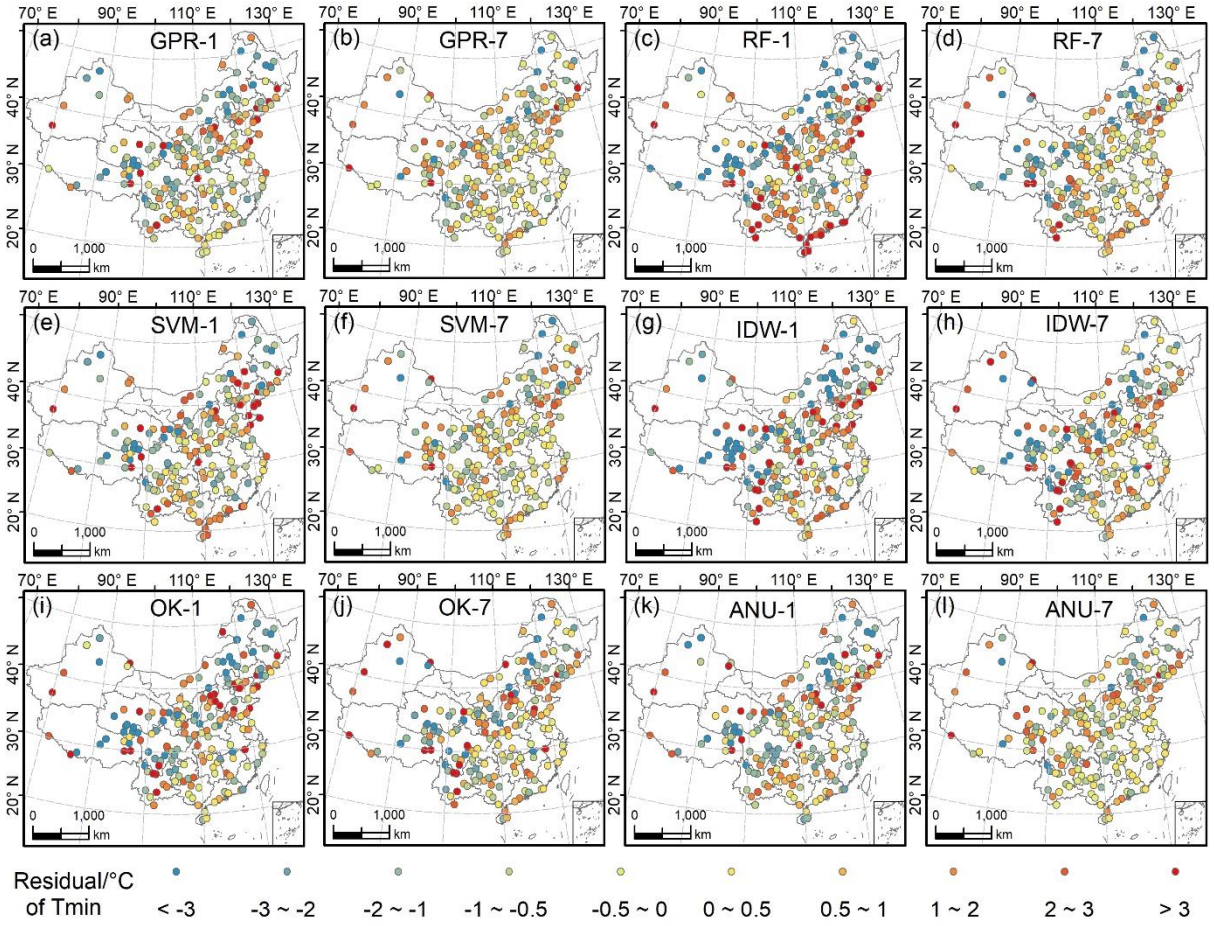

Figure S51: Comparison of the spatial distribution of residuals between machine learning and traditional methods for Tmin in January and July of 2020.

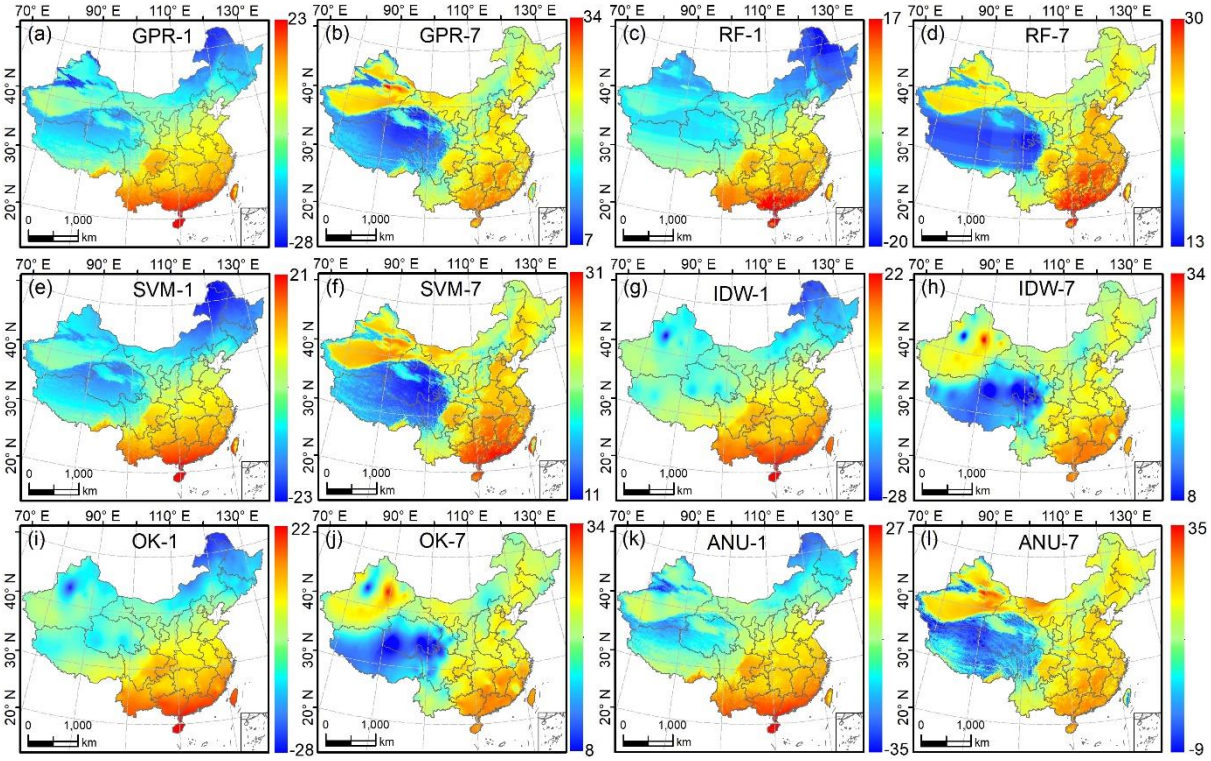

Figure S52: The spatial maps of monthly Tmean over China in January and July of 2020 (1, 7: January, July; unit: °C).

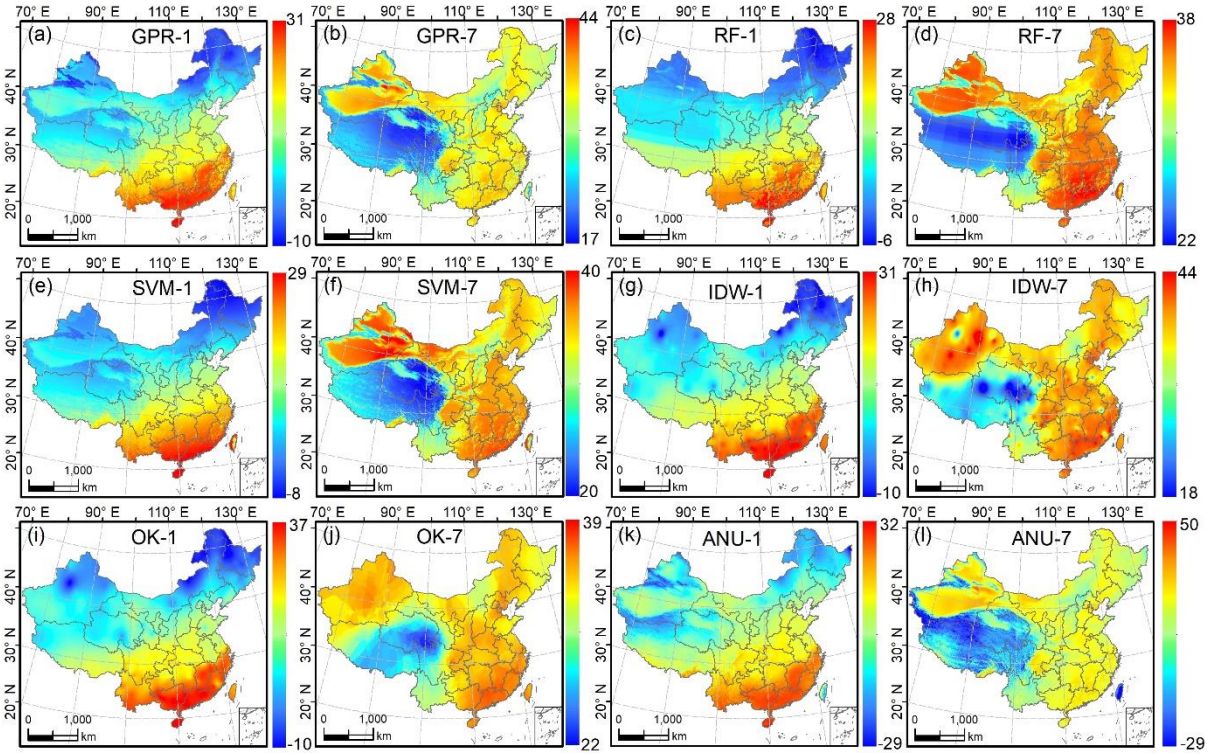

Figure S53: The spatial maps of monthly Tmax over China in January and July of 2020 (1, 7: January, July; unit: °C).

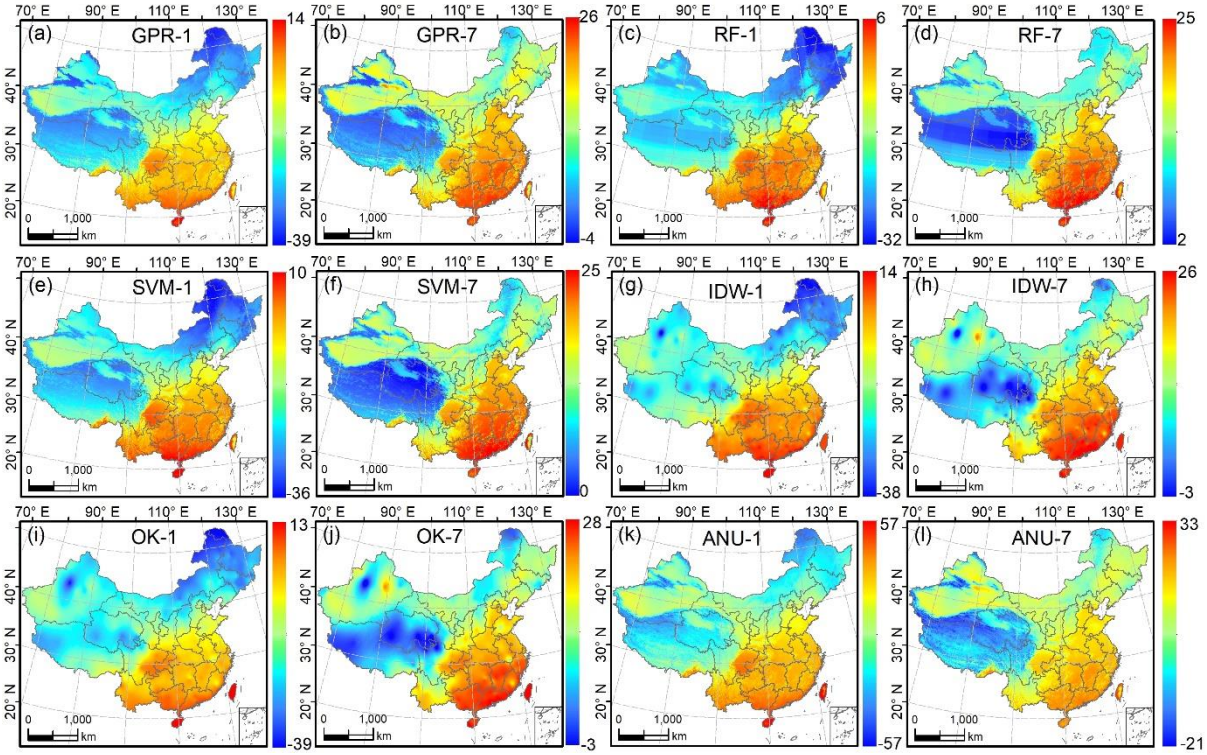

Figure S54: The spatial maps of monthly Tmin over China in January and July of 2020 (1, 7: January, July; unit: °C).

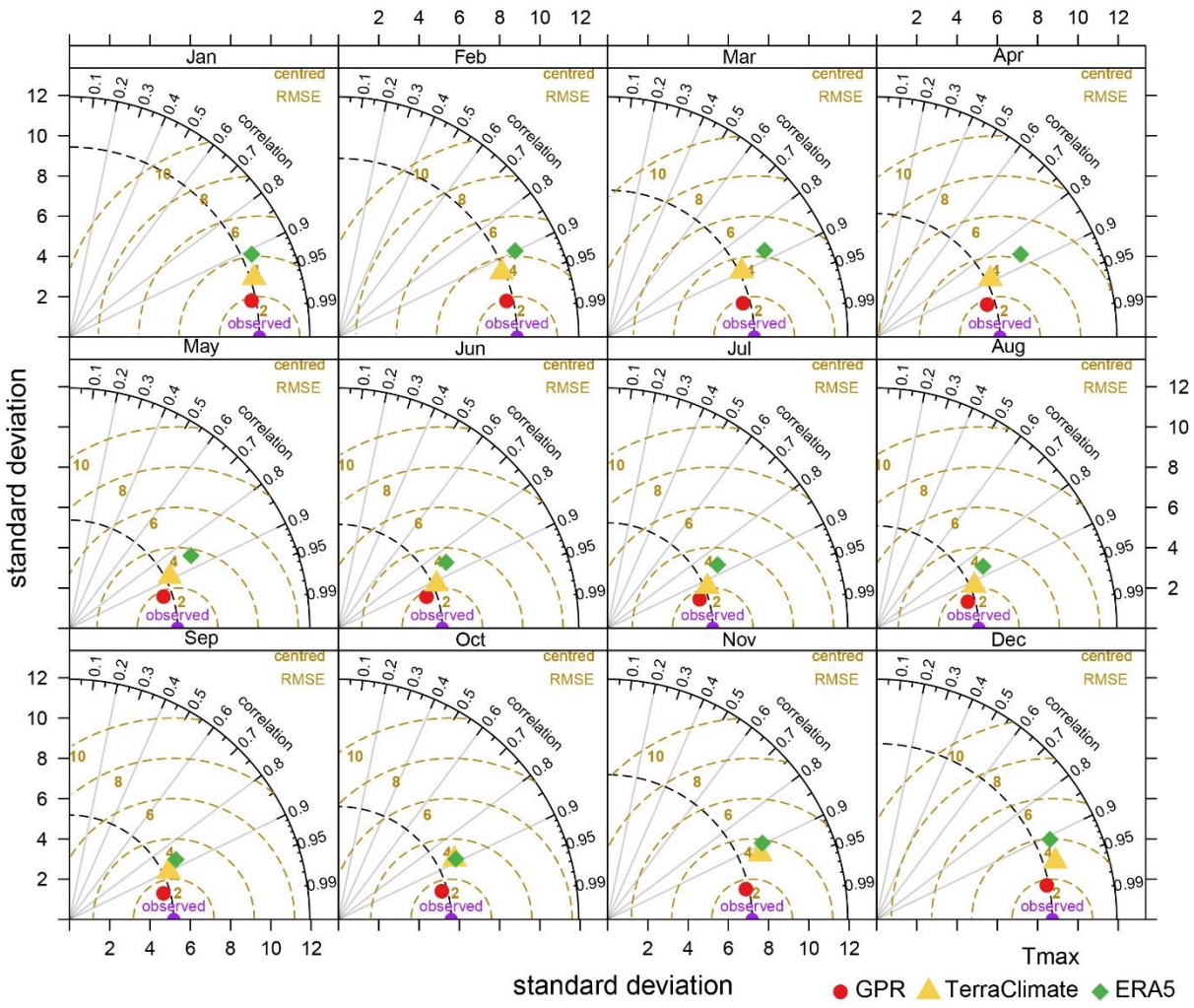

Figure S55: Taylor diagram displaying a statistical comparison with observations for Tmax between our products and other products for each month. The available time periods for the TerraClimate and ERA5 products are: 1958-01-01 - 2020-12-01, and 1979-01-01 - 2020-06-01, respectively. Considering the overlapping periods, we chose January 1979-December 2019 for comparing Tmax.

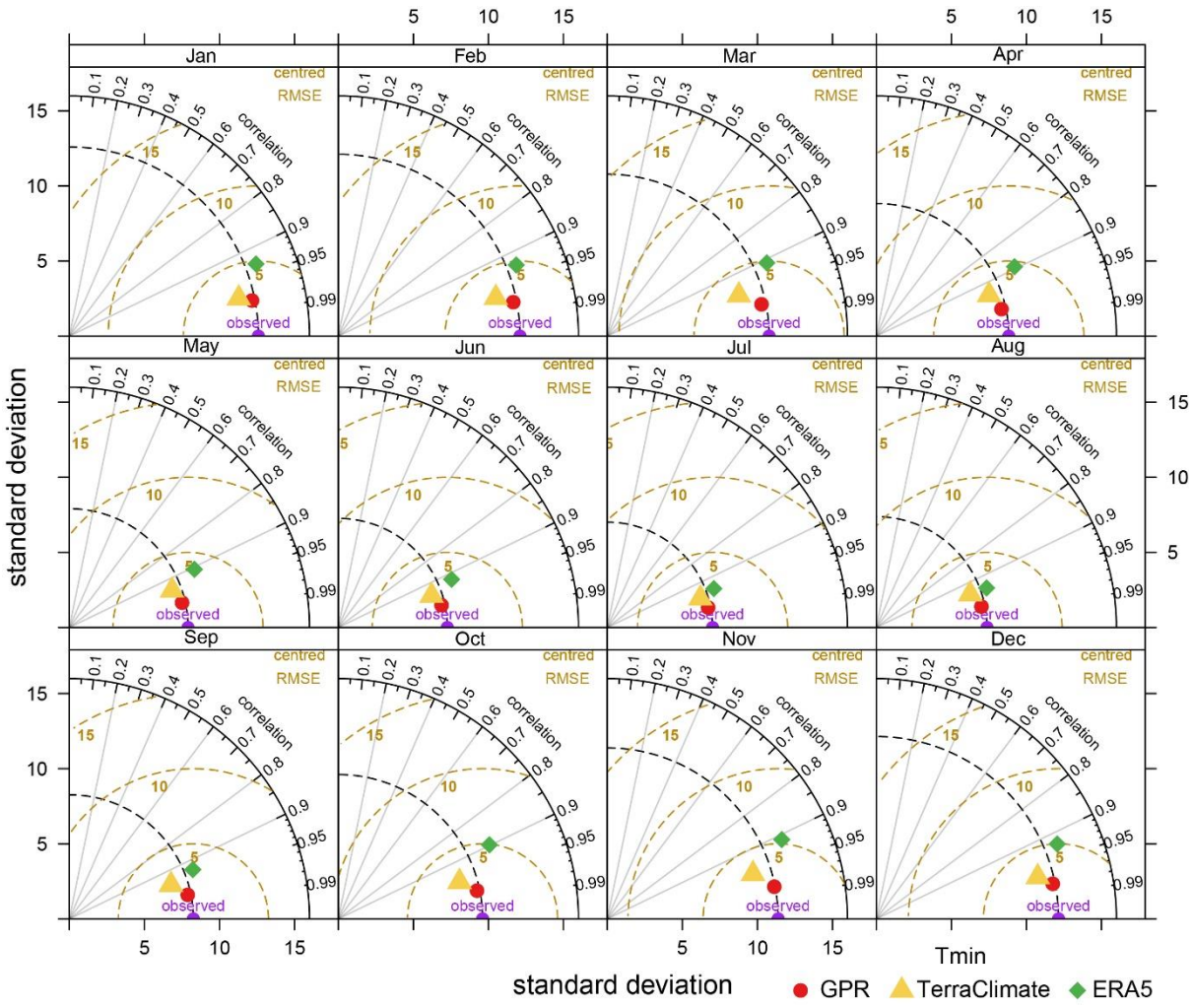

Figure S56: Taylor diagram displaying a statistical comparison with observations for Tmin between our products and other products for each month. The available time periods for the TerraClimate and ERA5 products are: 1958-01-01 - 2020-12-01, and 1979-01-01 - 2020-06-01, respectively. Considering the overlapping periods, we chose January 1979-December 2019 for comparing Tmin.

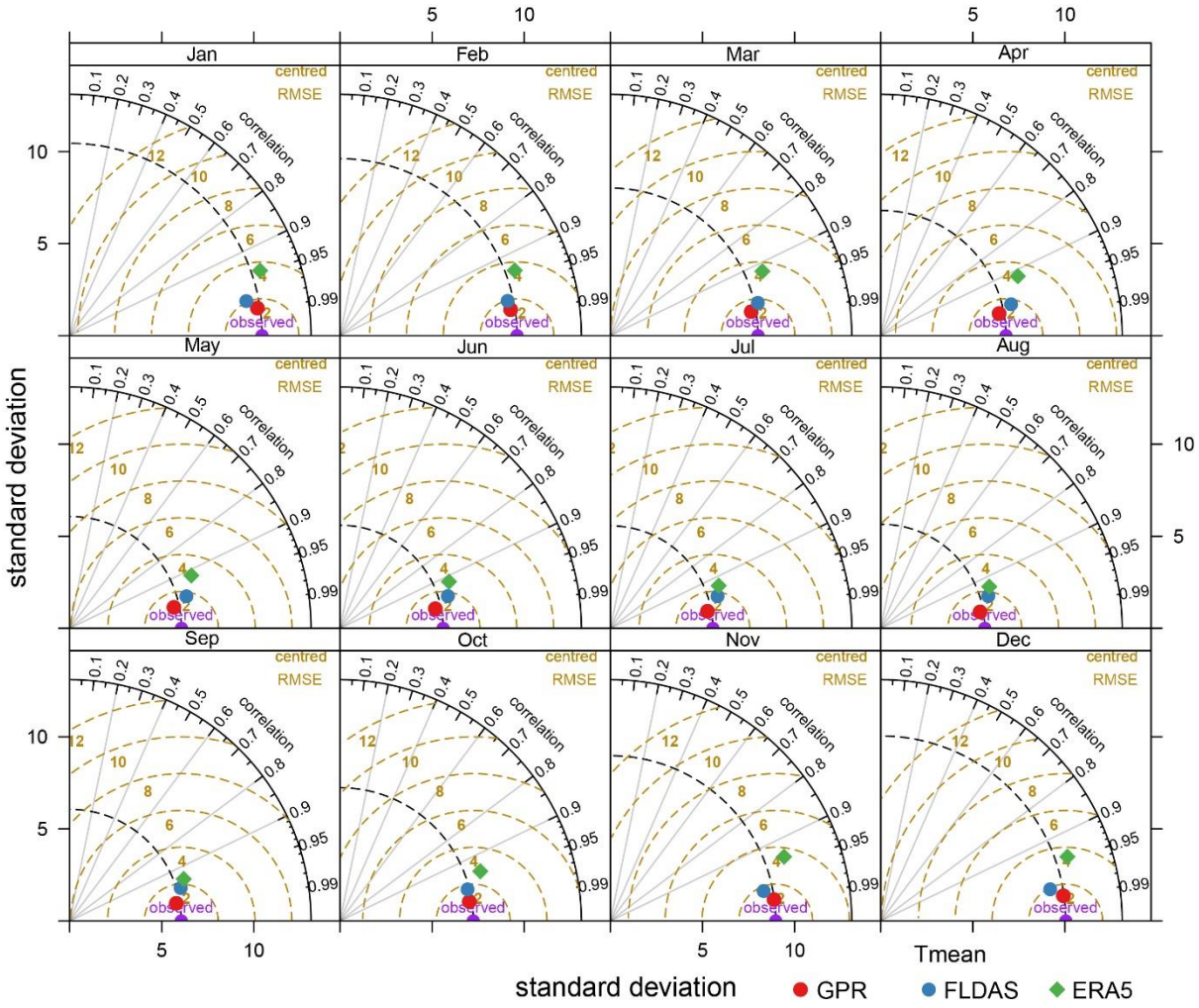

Figure S 57: Taylor diagram displaying a statistical comparison with observations for Tmean between our products and other products for each month. The available time periods for FLDAS and ERA5 products are: 1982-01-01 - 2021-05-01, and 1979-01-01 - 2020-06-01, respectively. Considering the overlapping periods, we chose January 1982 to December 2019 for comparing Tmean.