Supplement of Earth Syst. Sci. Data, 14, 3273–3292, 2022
https://doi.org/10.5194/essd-14-3273-2022-supplement

Supplement of

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

Qian He et al.

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

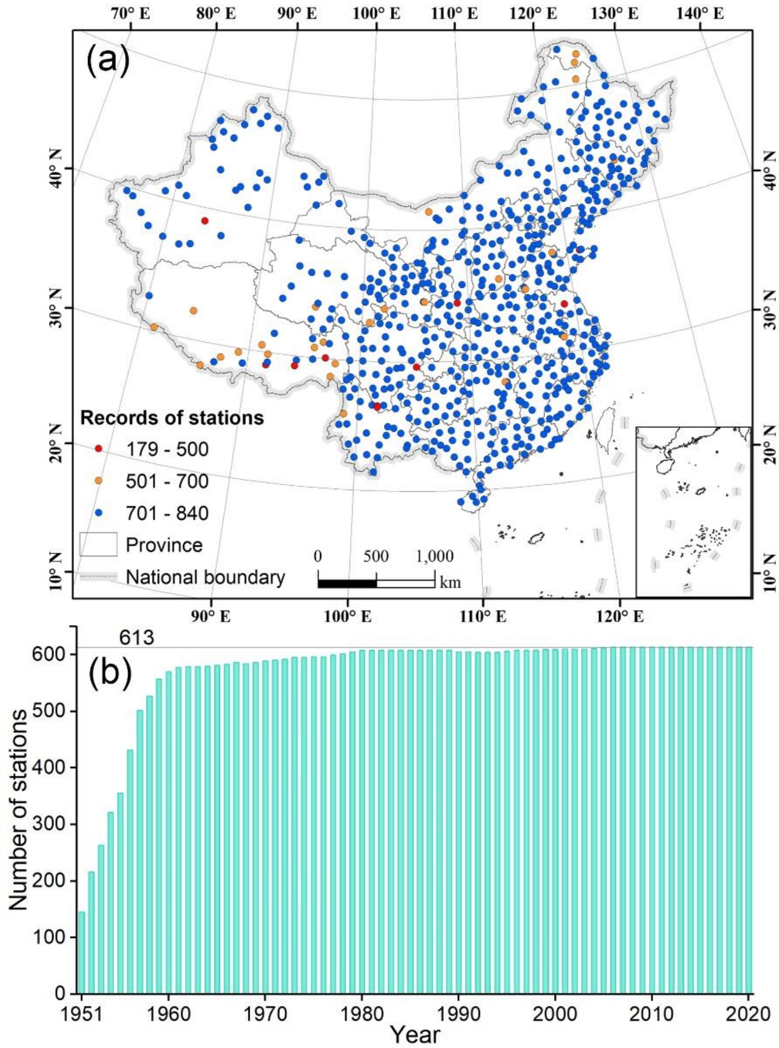

Figure S1: The records of each weather station (a) (and the number of stations in each year from 1951 to 2020 (b).

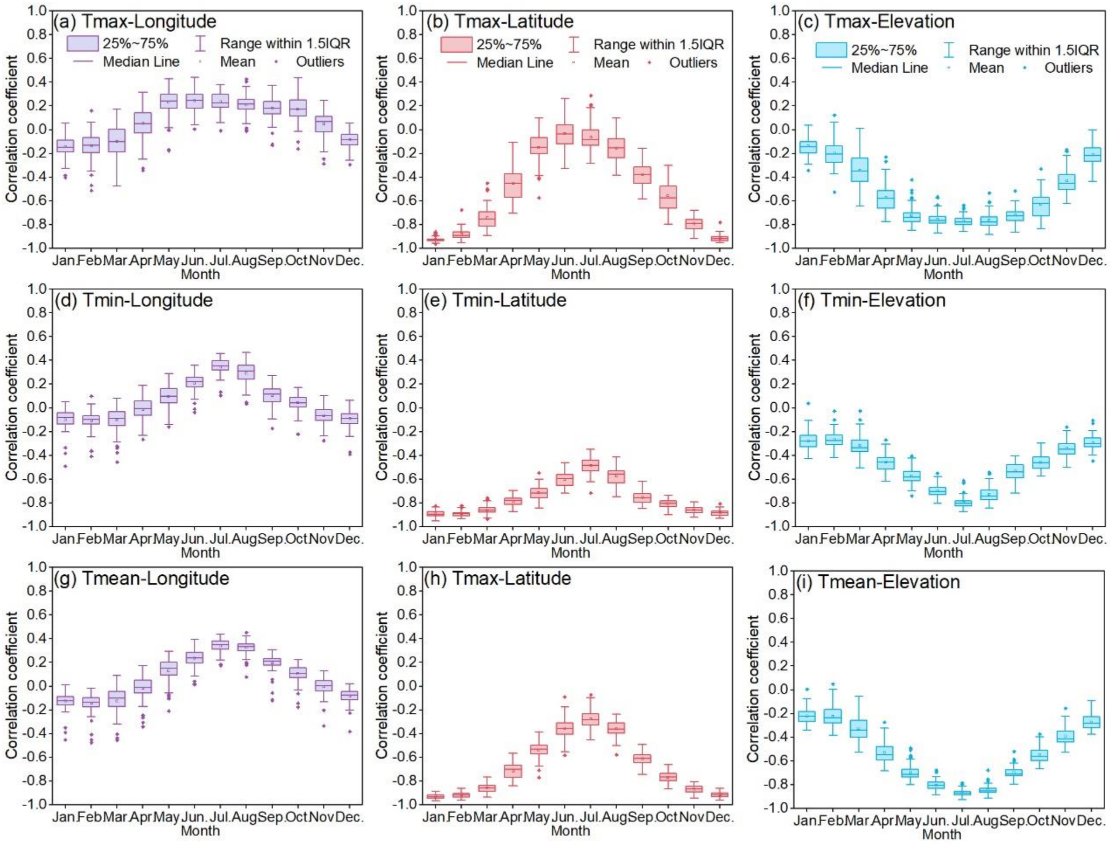

Figure S2: Correlation coefficients between temperature with longitude, latitude and elevation for each month from 1951-2020. a, b, c: monthly maximum air temperature (Tmax), c, d, e: monthly minimum air temperature (Tmin); g, h, g: monthly mean air temperature (mean).

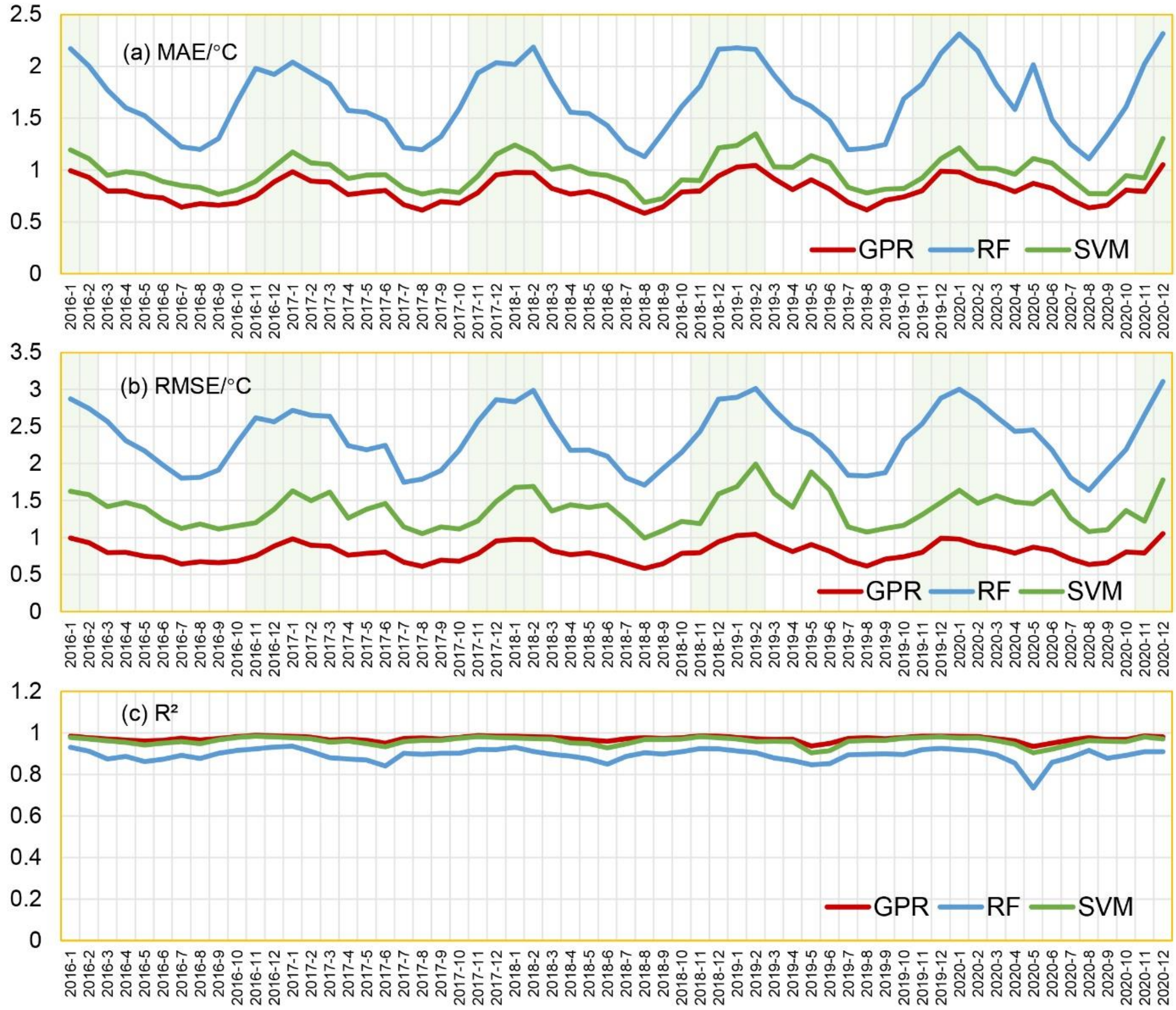

Figure S3: Zoomed-in accuracy metrics of Tmean for the period of January 2016 to December 2020.

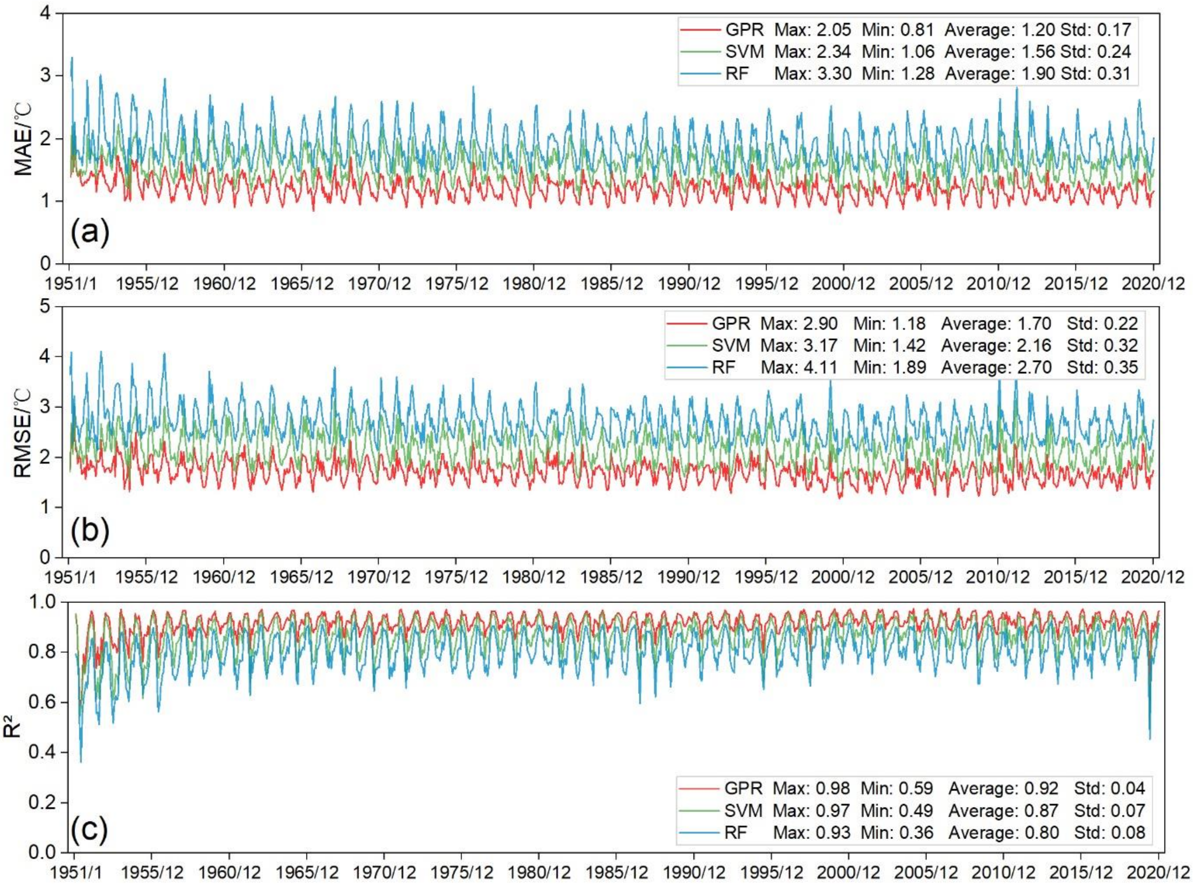

Figure S4: Mean absolute error (MAE), root mean square error (RMSE) and coefficients of determination (R^2) between observed Tmax and predicted Tmax by three machine learning models (GPR, SVM, RF) of testing weather stations over the time period from January 1951 to December 2020.

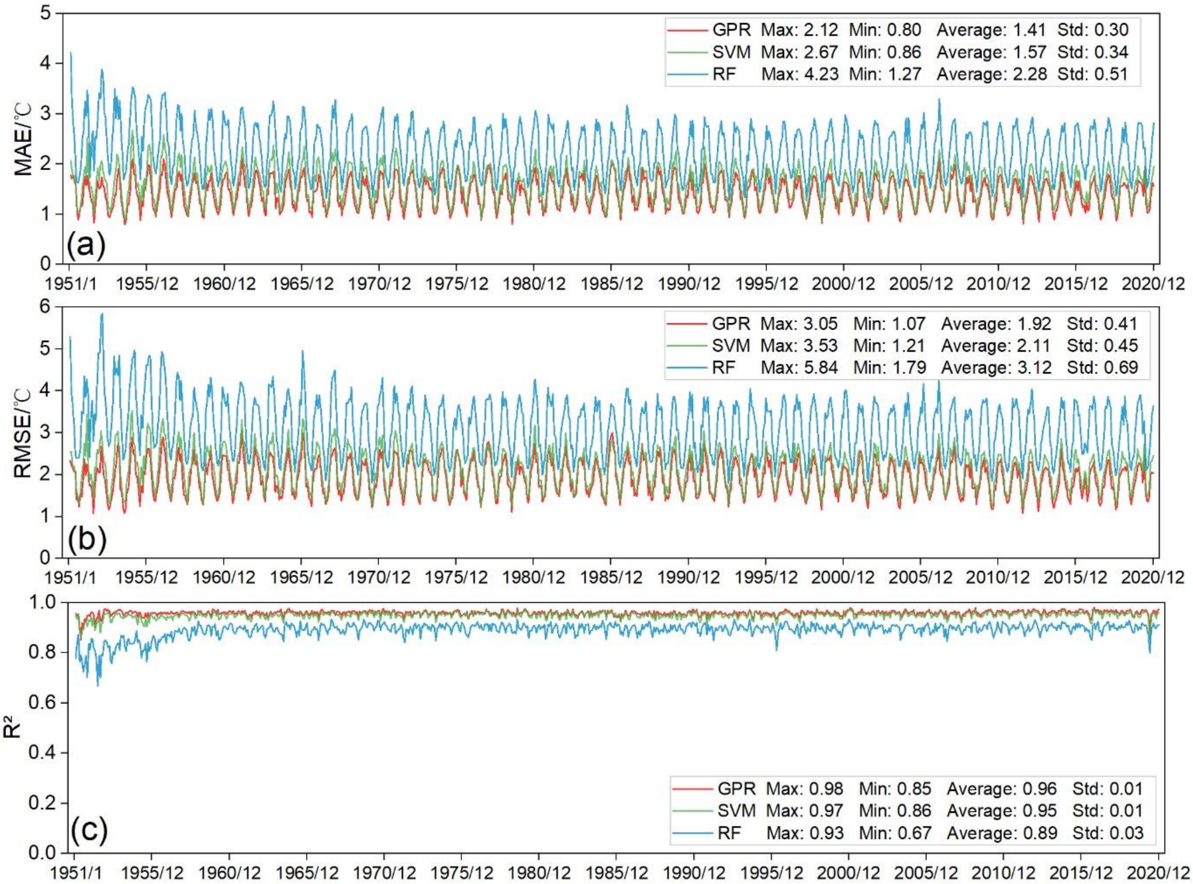

Figure S5: Mean absolute error (MAE), root mean square error (RMSE) and coefficients of determination (R2) between observed Tmin and predicted Tmin by three machine learning models (GPR, SVM, RF) of testing weather stations over the time period from January 1951 to December 2020.

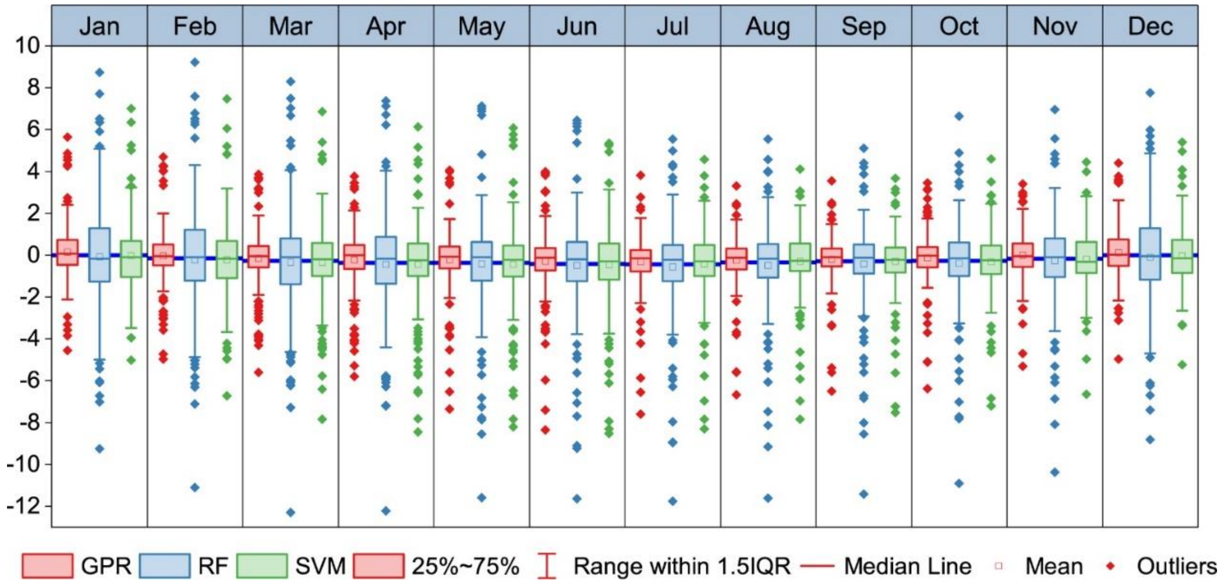

Figure S6: Residuals of the monthly Tmax predicted by the machine learning models with respect to in situ Tmax for the test meteorological stations. Note that the average of the residuals of Tmax from 1951–2020 for each test meteorological station is shown for each month.

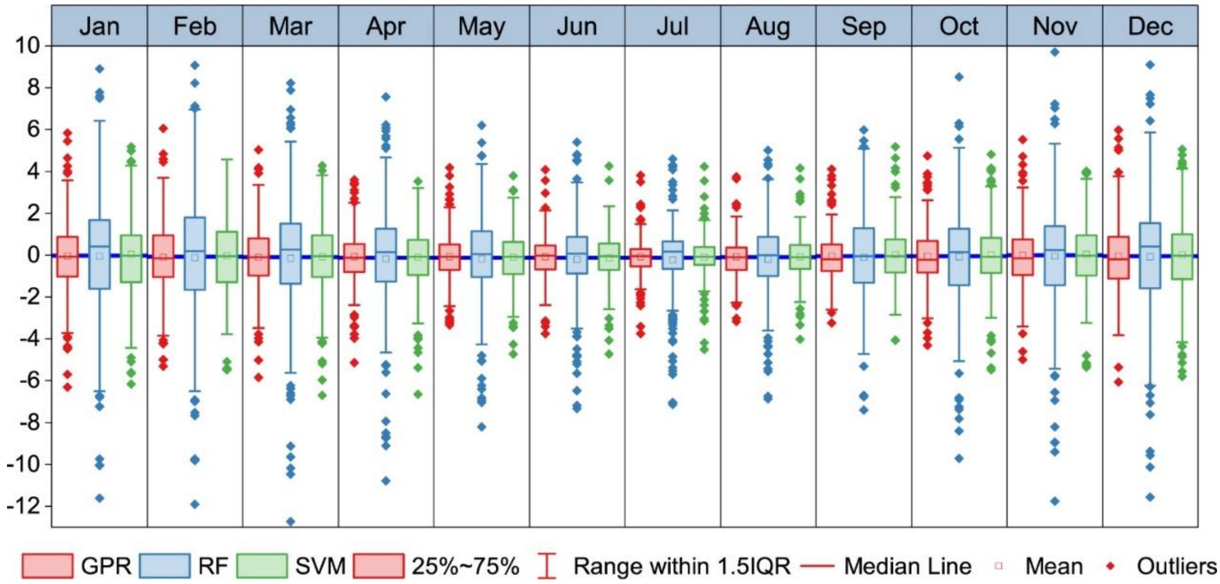

Figure S7: Residuals of the monthly Tmin predicted by the machine learning models with respect to in situ Tmin for the test meteorological stations. Note that the average of the residuals of Tmin from 1951–2020 for each test meteorological station is shown for each month.

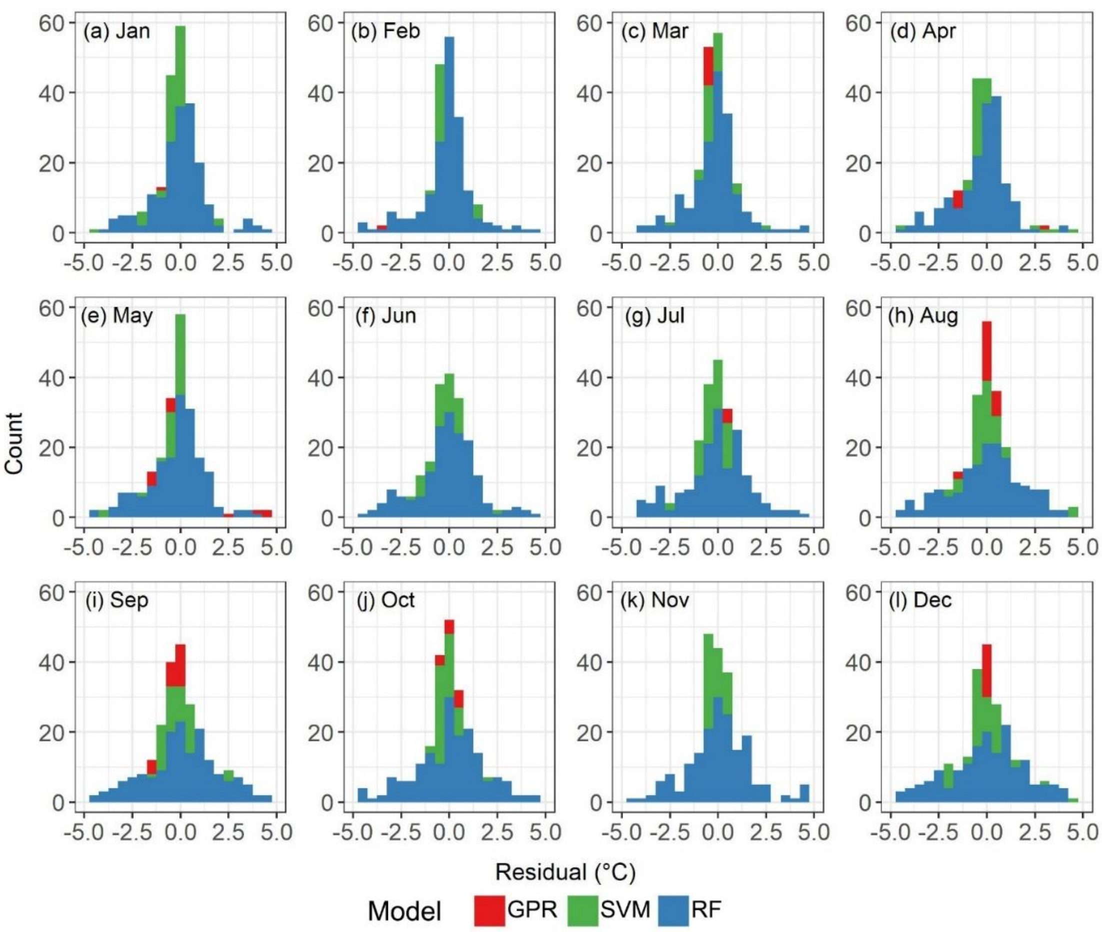

Figure S8: Frequency distribution of residuals for monthly Tmean using the machine learning methods during the period from 1951 to 2020. Note that the displayed residuals are the average of residuals of 70 years (1951-2020) for each month.

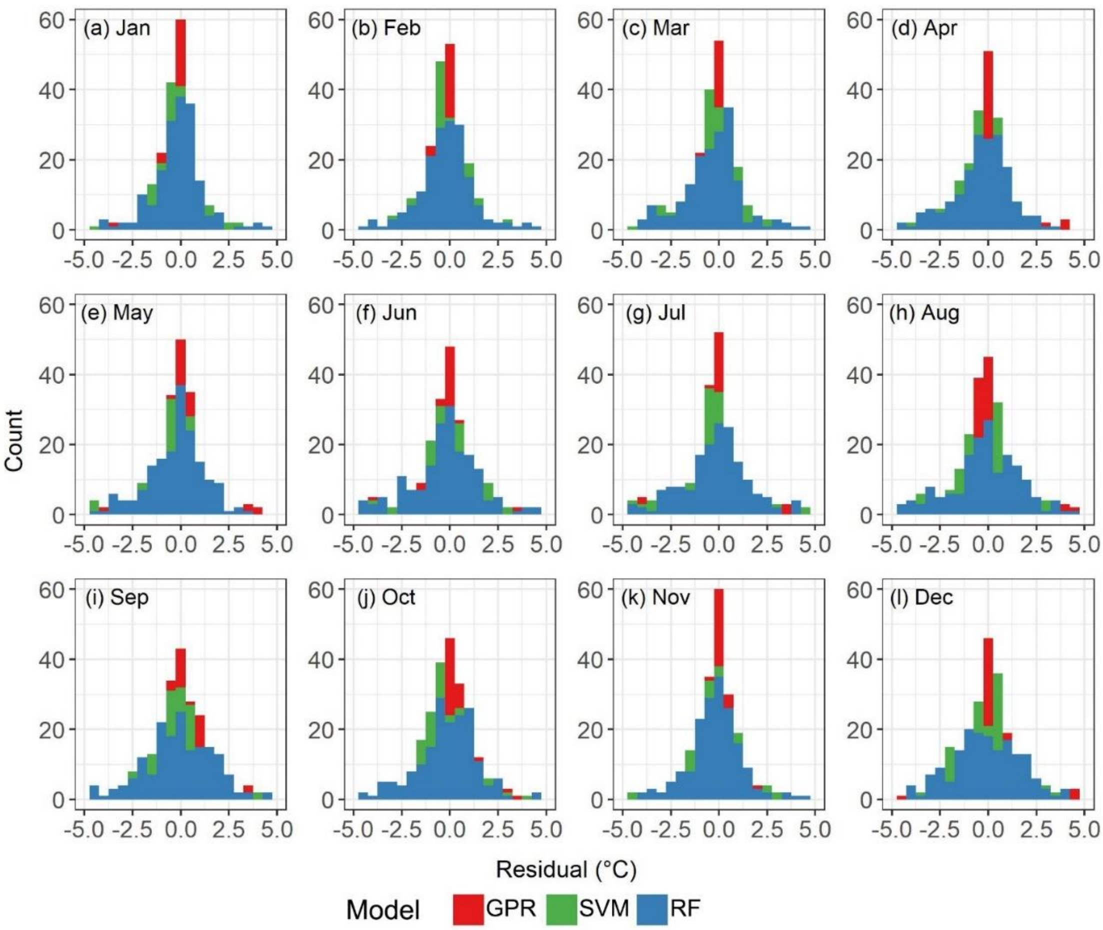

Figure S9: Frequency distribution of residuals for monthly Tmax using the machine learning methods during the period from 1951 to 2020. Note that the displayed residuals are the average of residuals of 70 years (1951-2020) for each month.

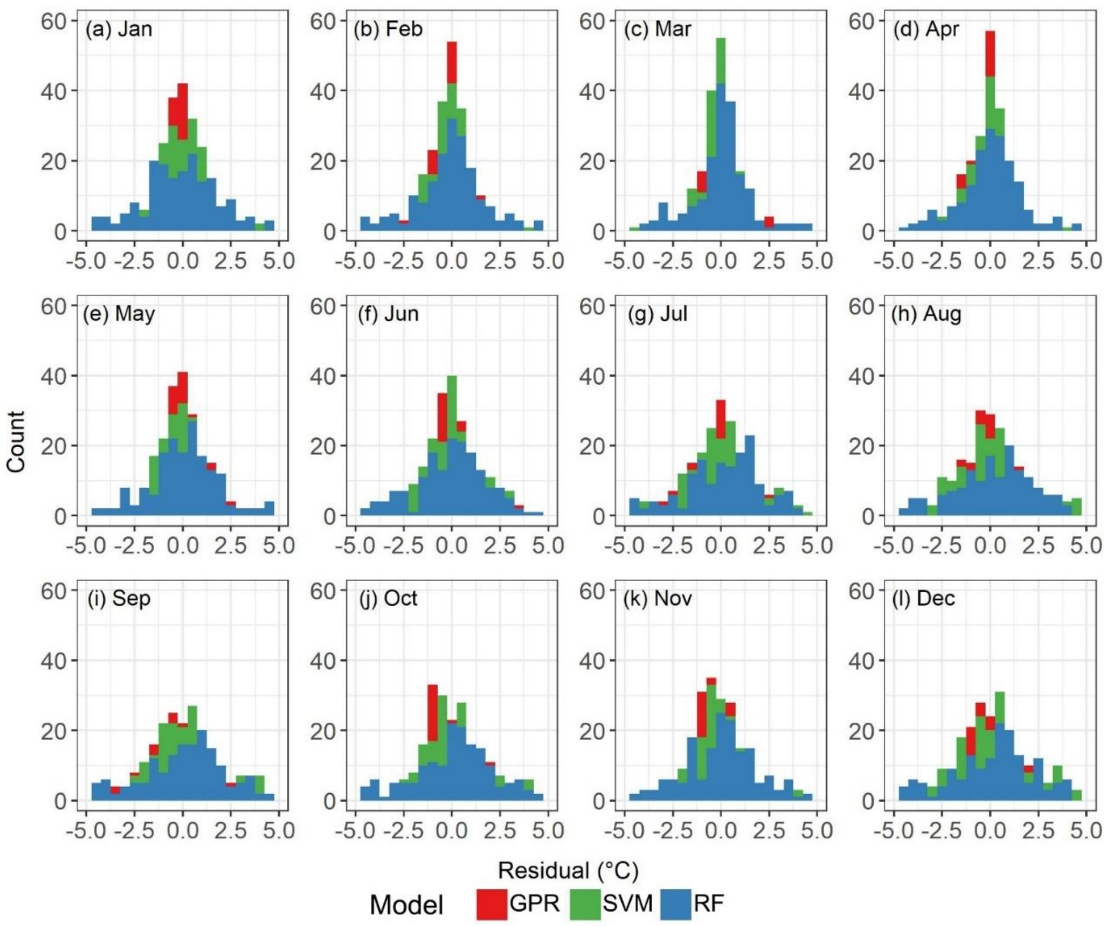

Figure S10: **Frequency distribution of residuals for monthly Tmin using the machine learning methods during the period from 1951 to 2020. Note that the displayed residuals are the average of residuals of 70 years (1951-2020) for each month.**

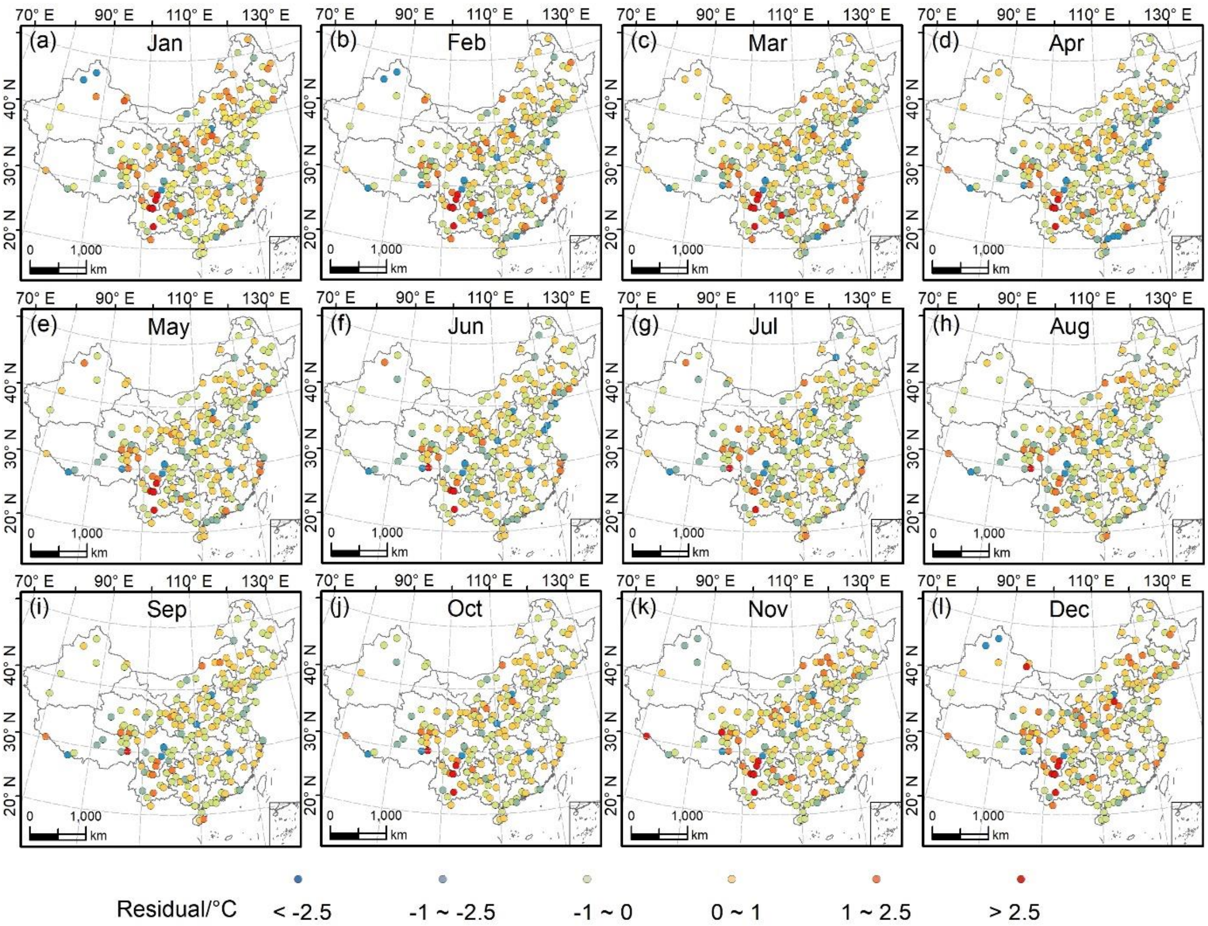

Figure S11: The spatial distribution of residuals between the observed Tmax and the predicted Tmax by GPR for the testing stations for each month. Note that the exhibited residuals are the average residual of 70 years from 1951- 2020 for each month.

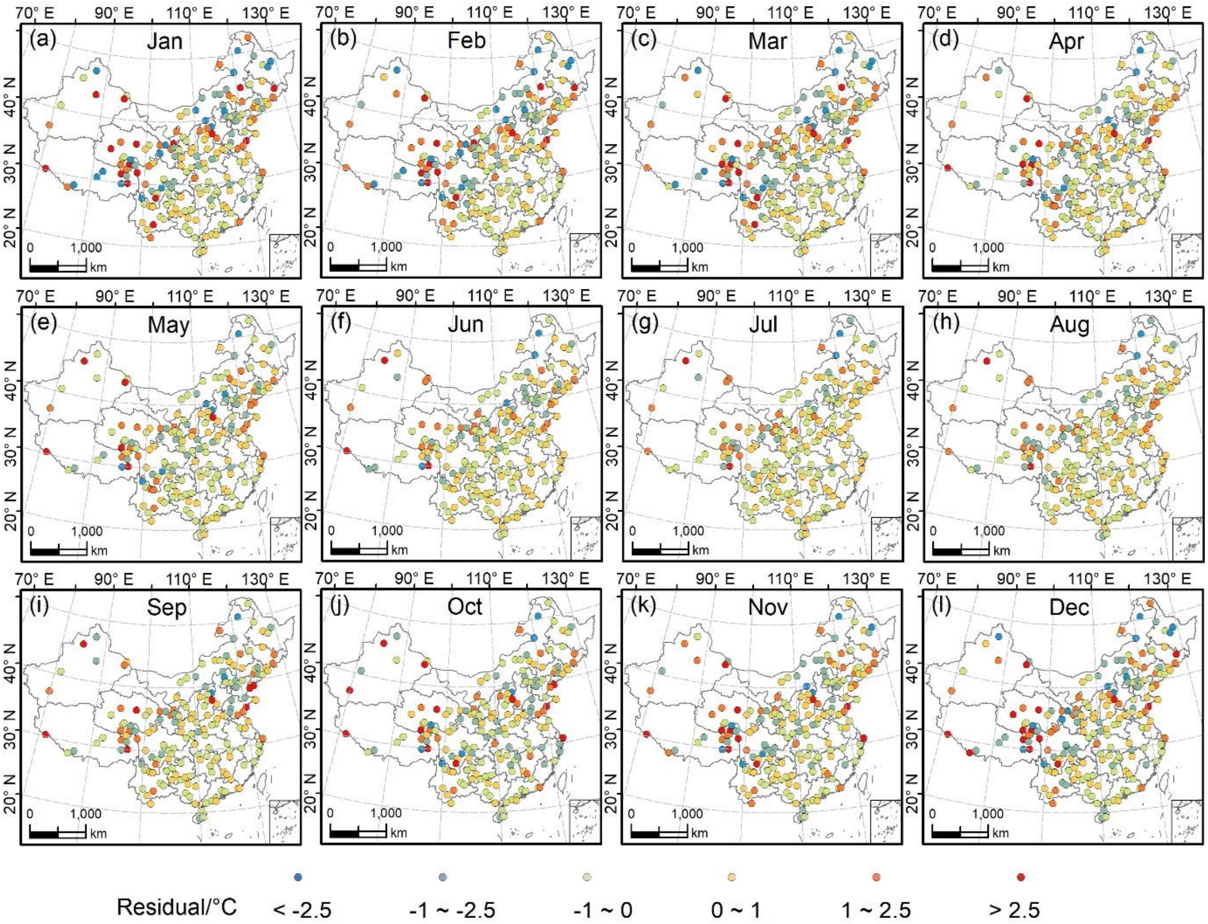

Figure S12: The spatial distribution of residuals between the observed Tmin and the predicted Tmin by GPR for the testing stations for each month. Note that the exhibited residuals are the average residual of 70 years from 1951- 2020 for each month.

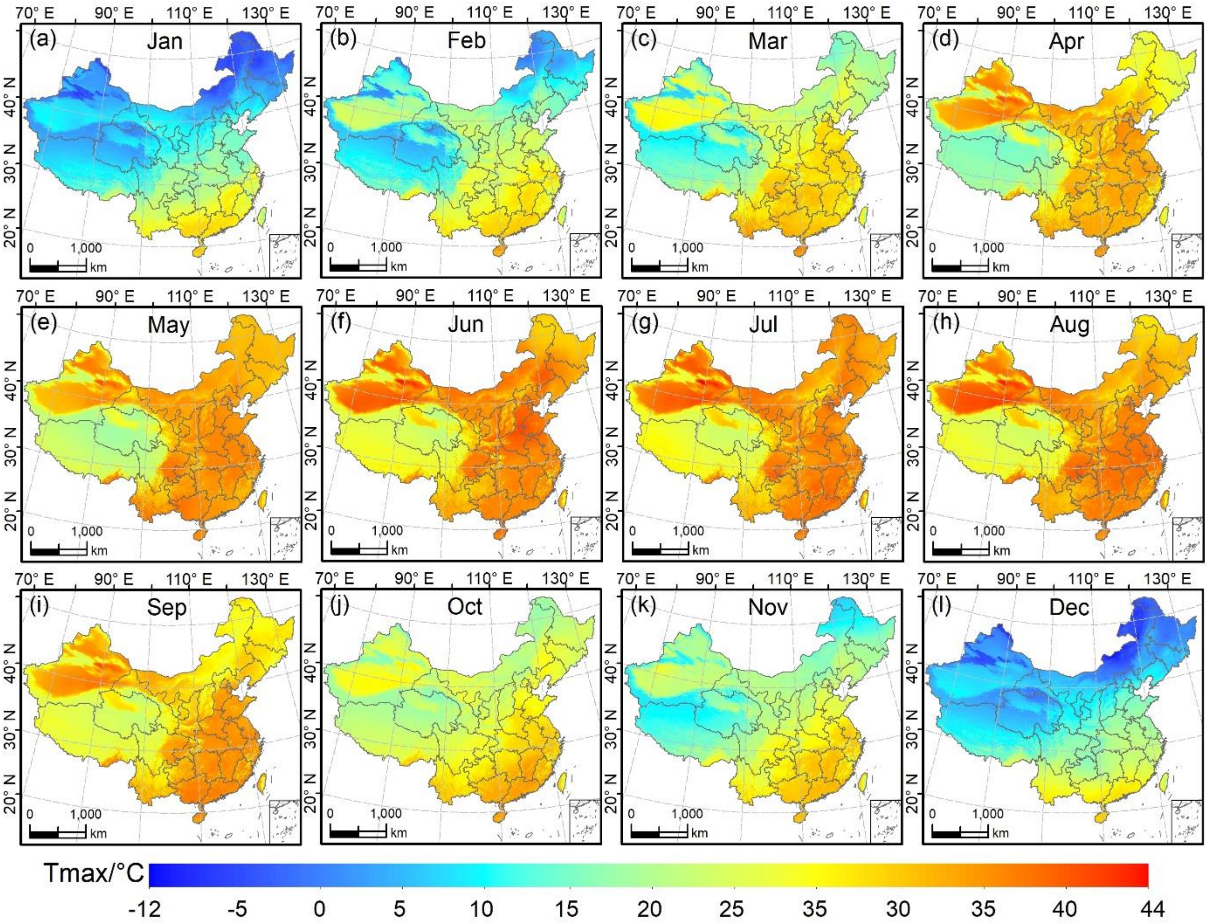

Figure S13: The spatial distribution maps of monthly Tmax predicted by GPR in China for each month in 2020.

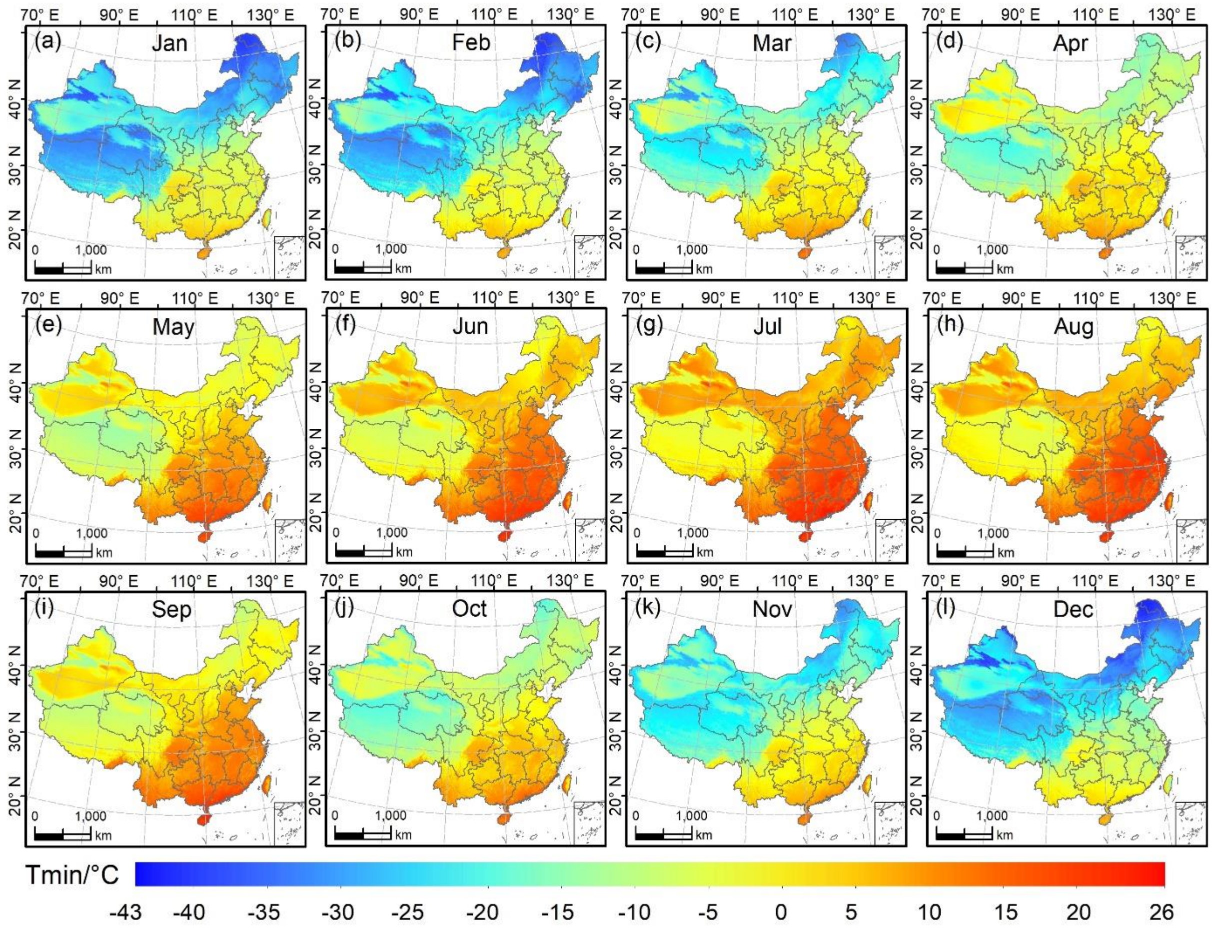

Figure S14: The spatial distribution maps of monthly Tmin predicted by GPR in China for each month in 2020.

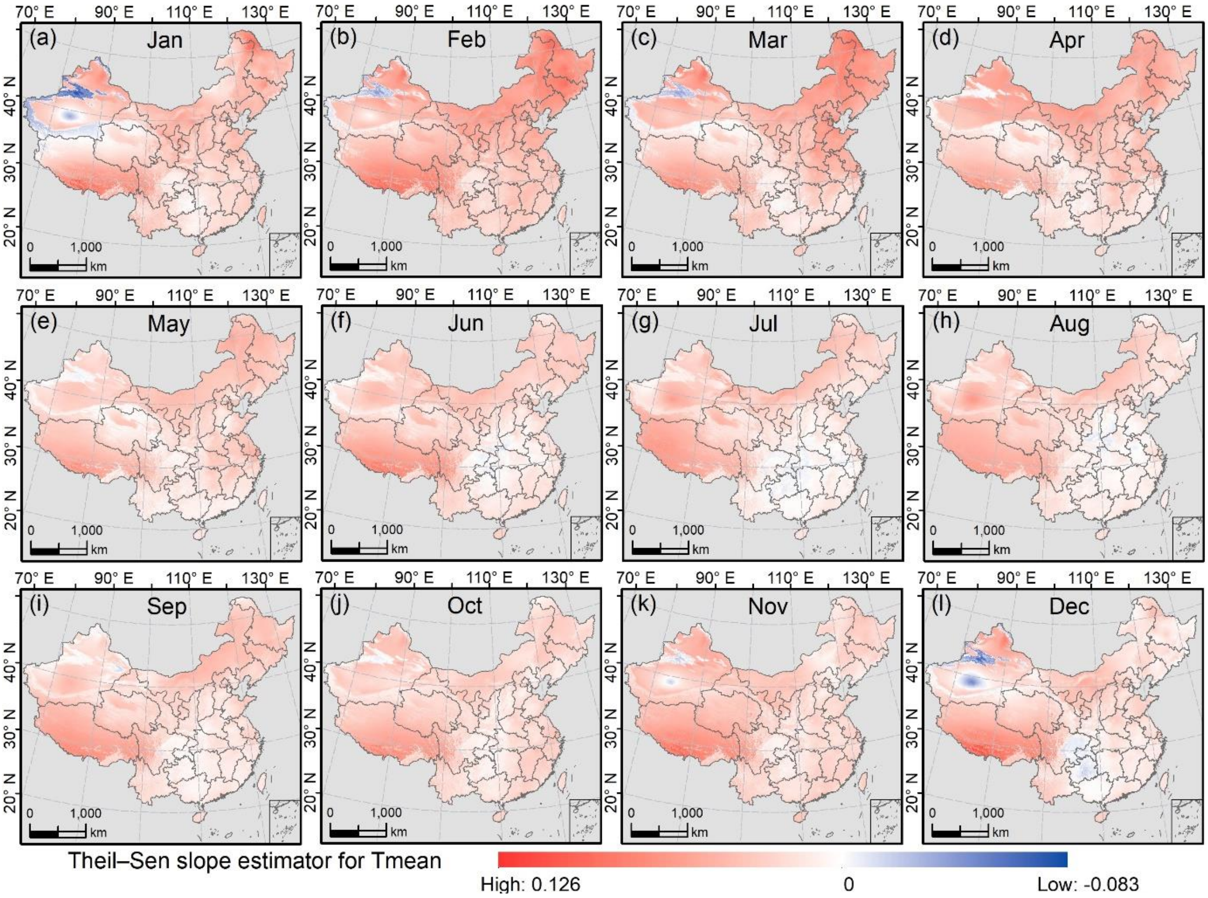

Figure S15: Theil–Sen (median) slope for Tmean over China (1951–2020) in each month.

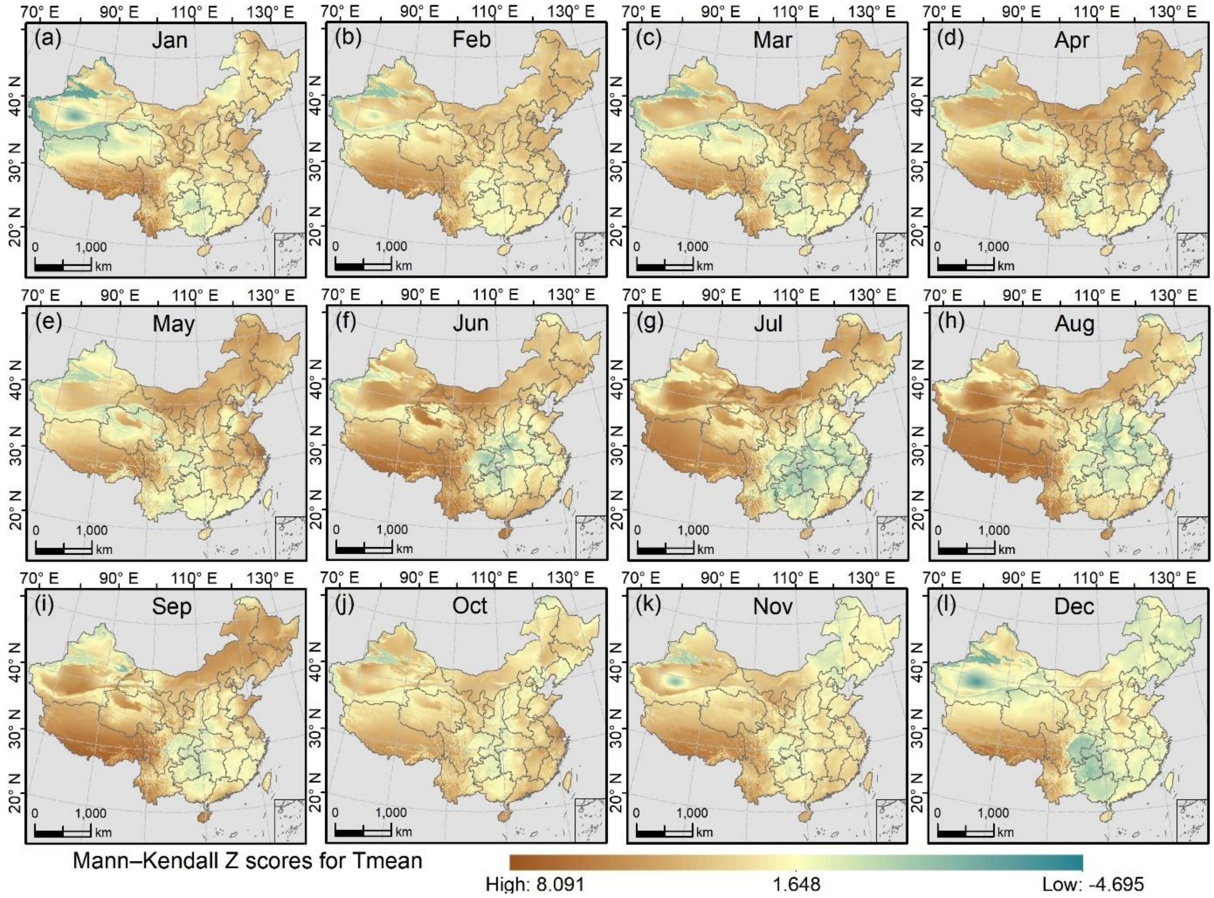

Figure S16: Mann–Kendall Z scores for Tmean over China (1951–2020) in each month.

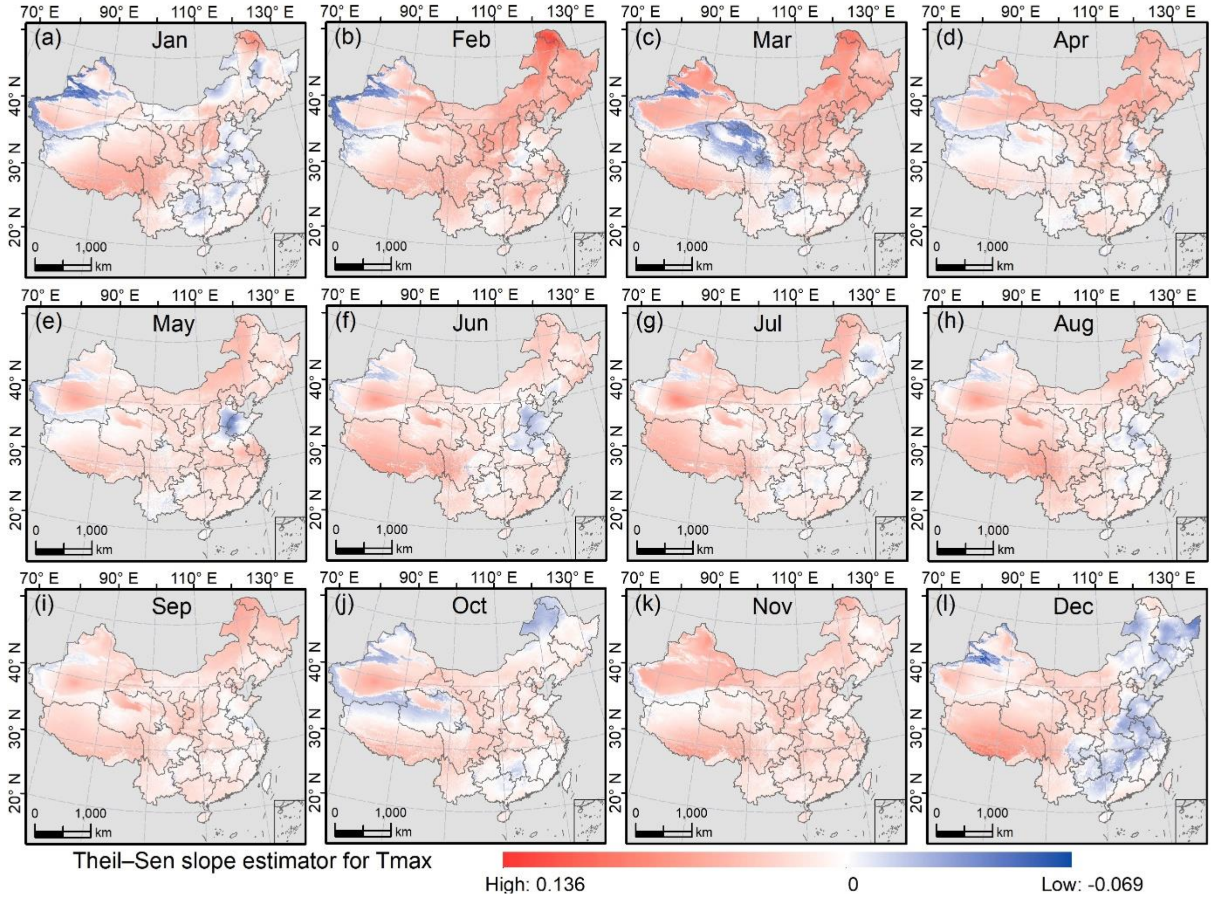

Figure S17: Theil–Sen (median) slope for Tmax over China (1951–2020) in each month.

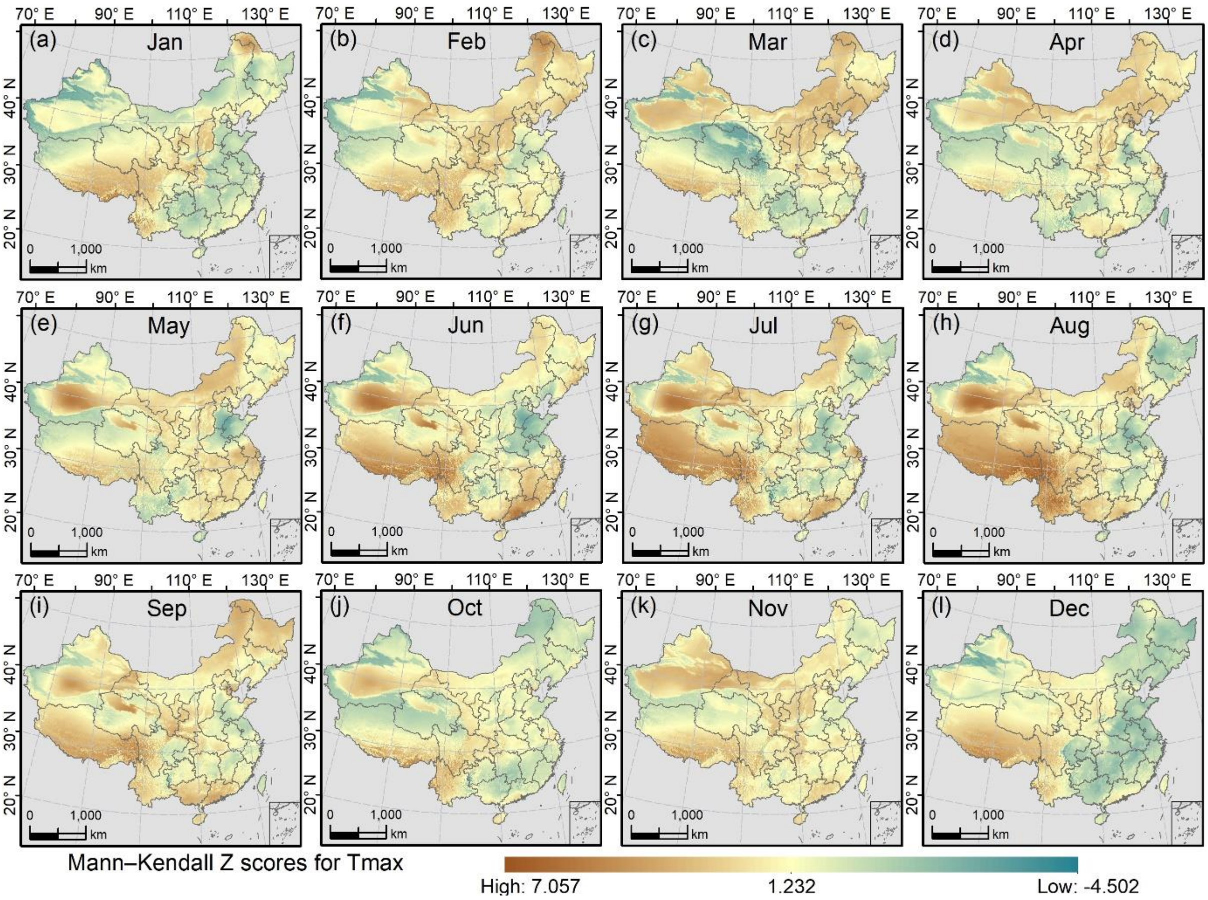

Figure S18: Mann–Kendall Z scores for Tmax over China (1951–2020) in each month.

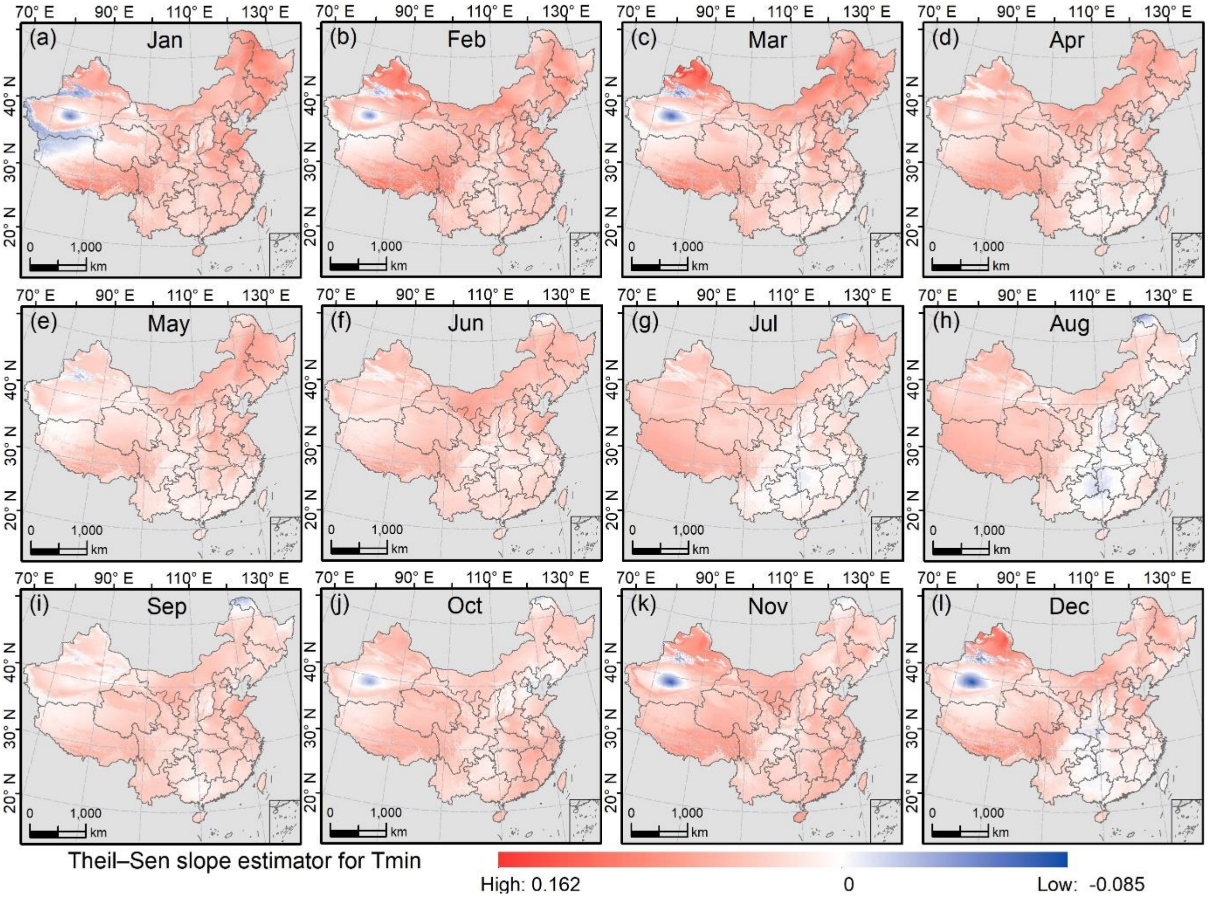

Figure S19: Theil–Sen (median) slope for Tmin over China (1951–2020) in each month.

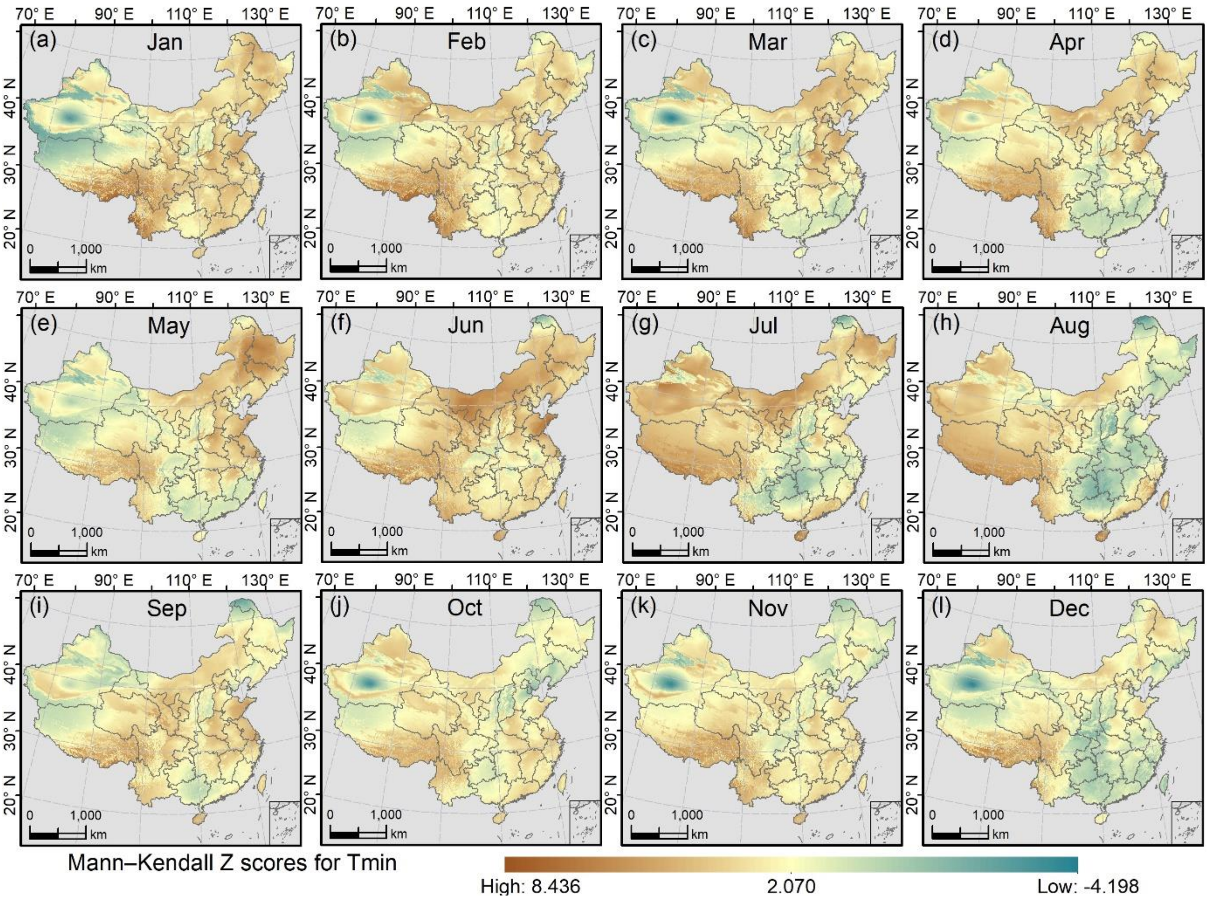

Figure S20: Mann–Kendall Z scores for Tmin over China (1951–2020) in each month.

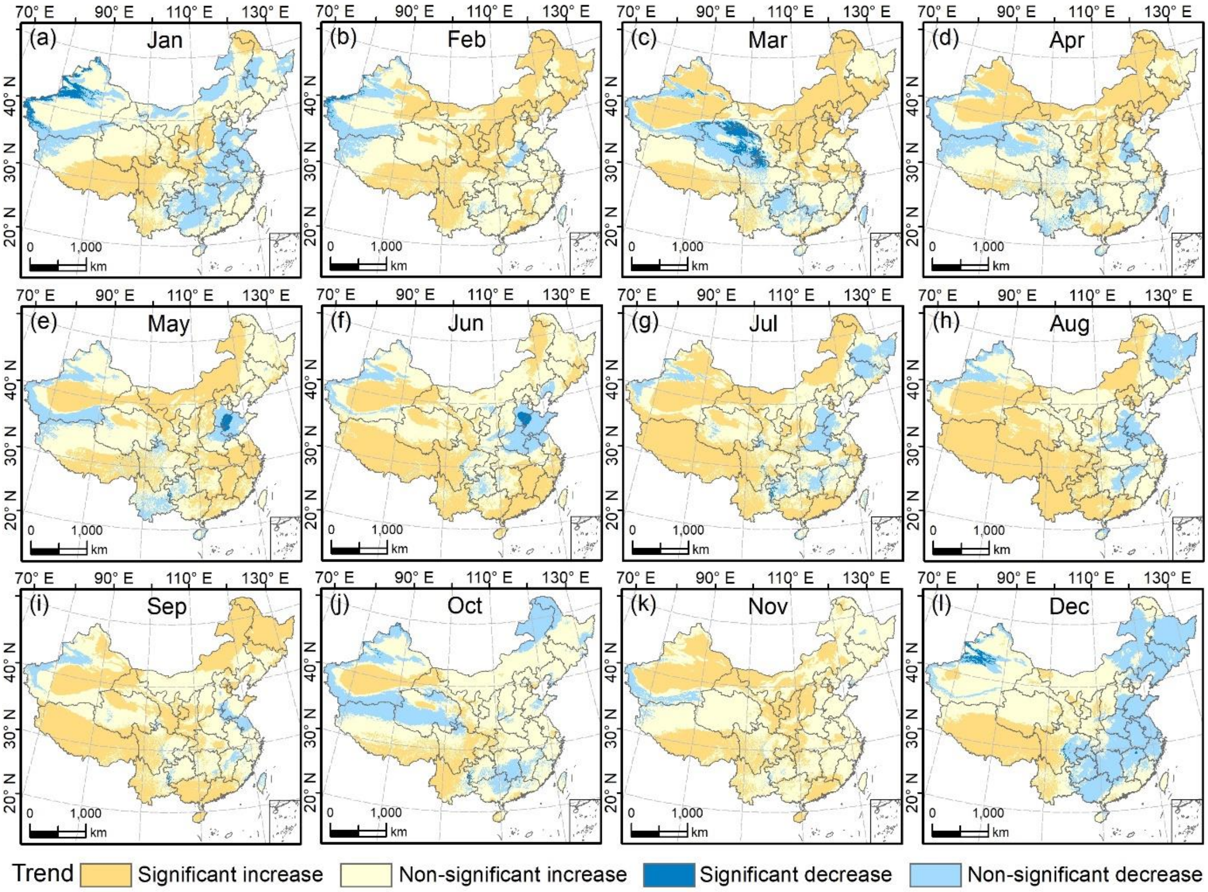

Figure S21: Monthly trends of Tmax change in China during the period 1951-2020 obtained by median Theil–Sen slope. The significance of trends is quantified by the Mann-Kendall statistical test at the 95% confidence level.

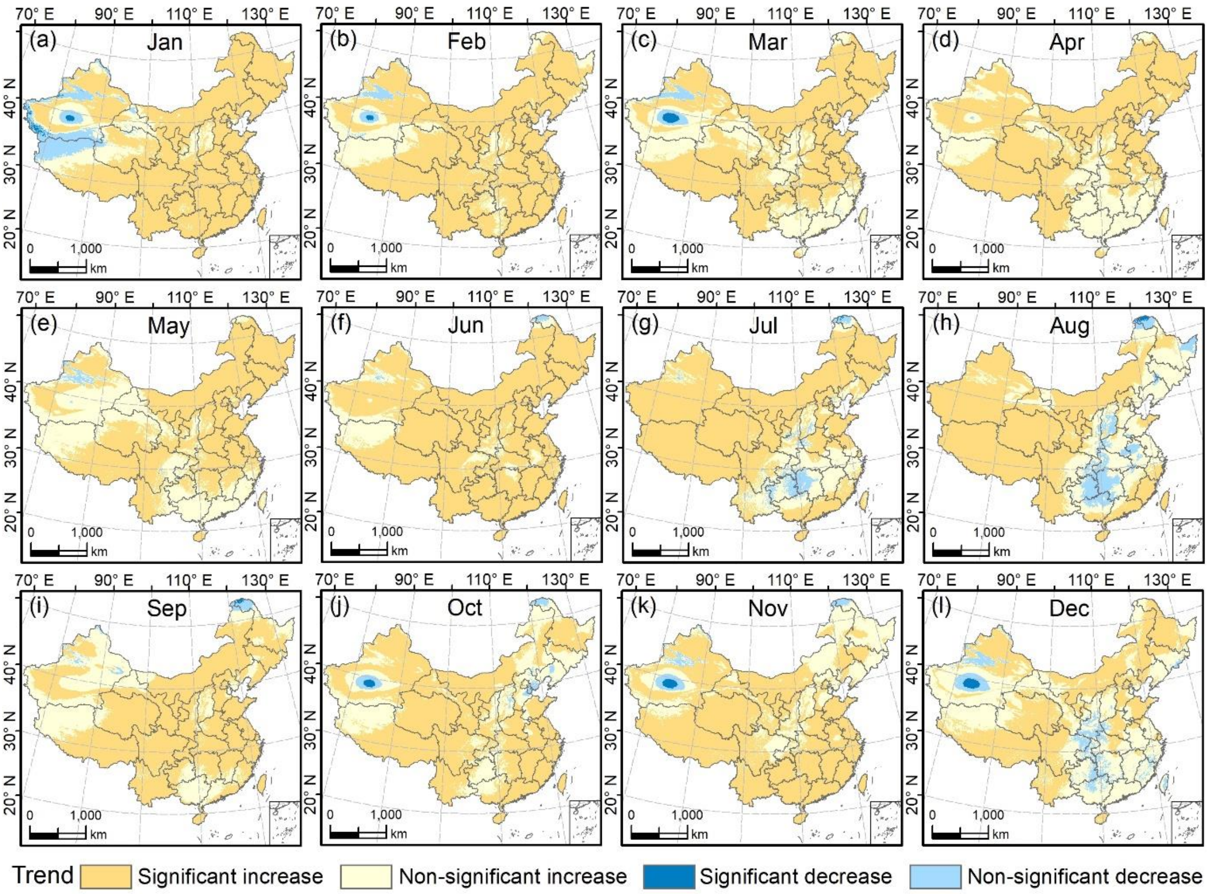

Figure S22: Monthly trends of Tmin change in China during the period 1951-2020 obtained by median Theil–Sen slope. The significance of trends is quantified by the Mann-Kendall statistical test at the 95% confidence level.

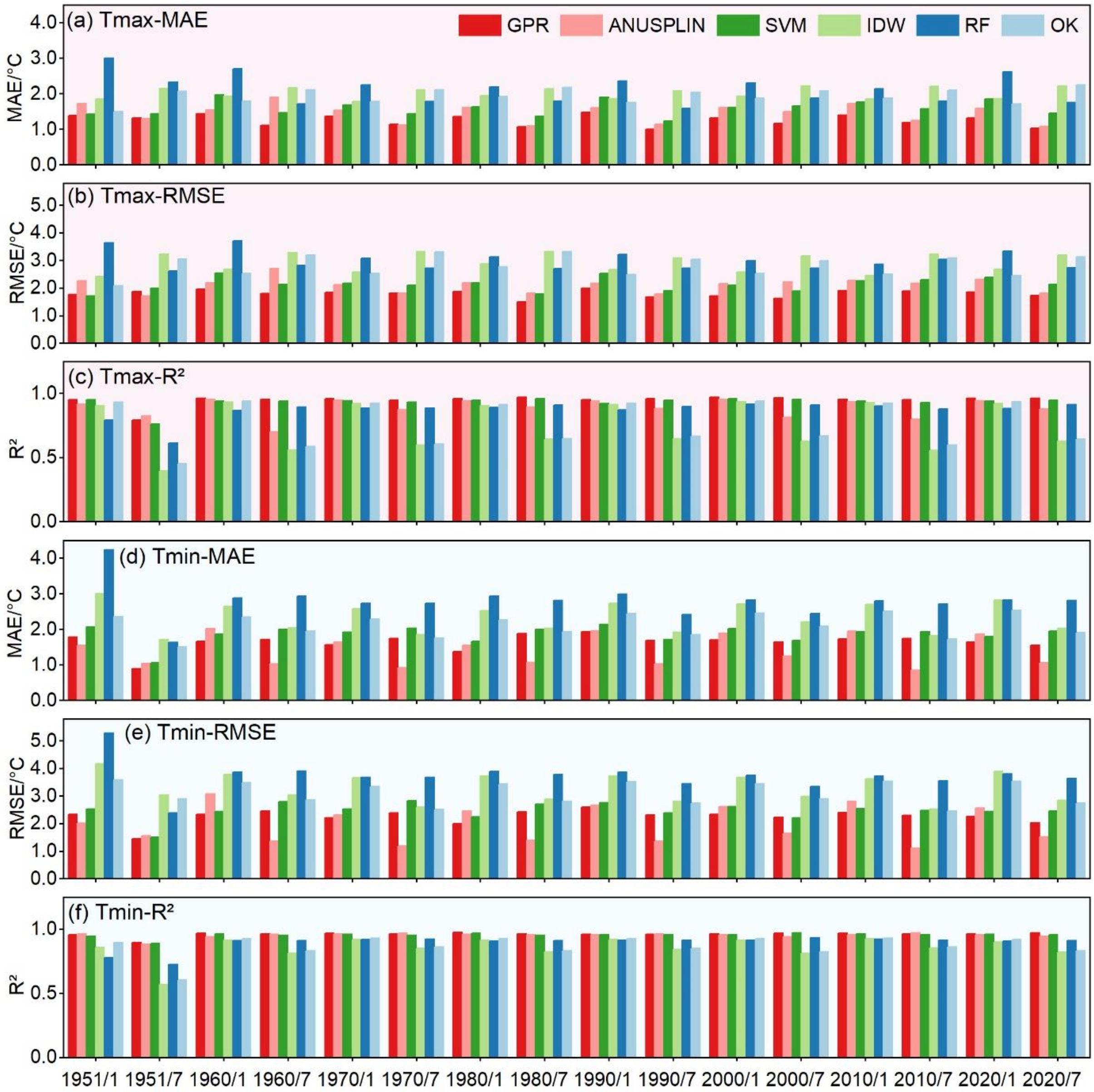

Figure S23: Accuracy of Tmax and Tmin derived by the machine learning methods and traditional methods for January and July from 1951 to 2020 with an interval of 10 years.

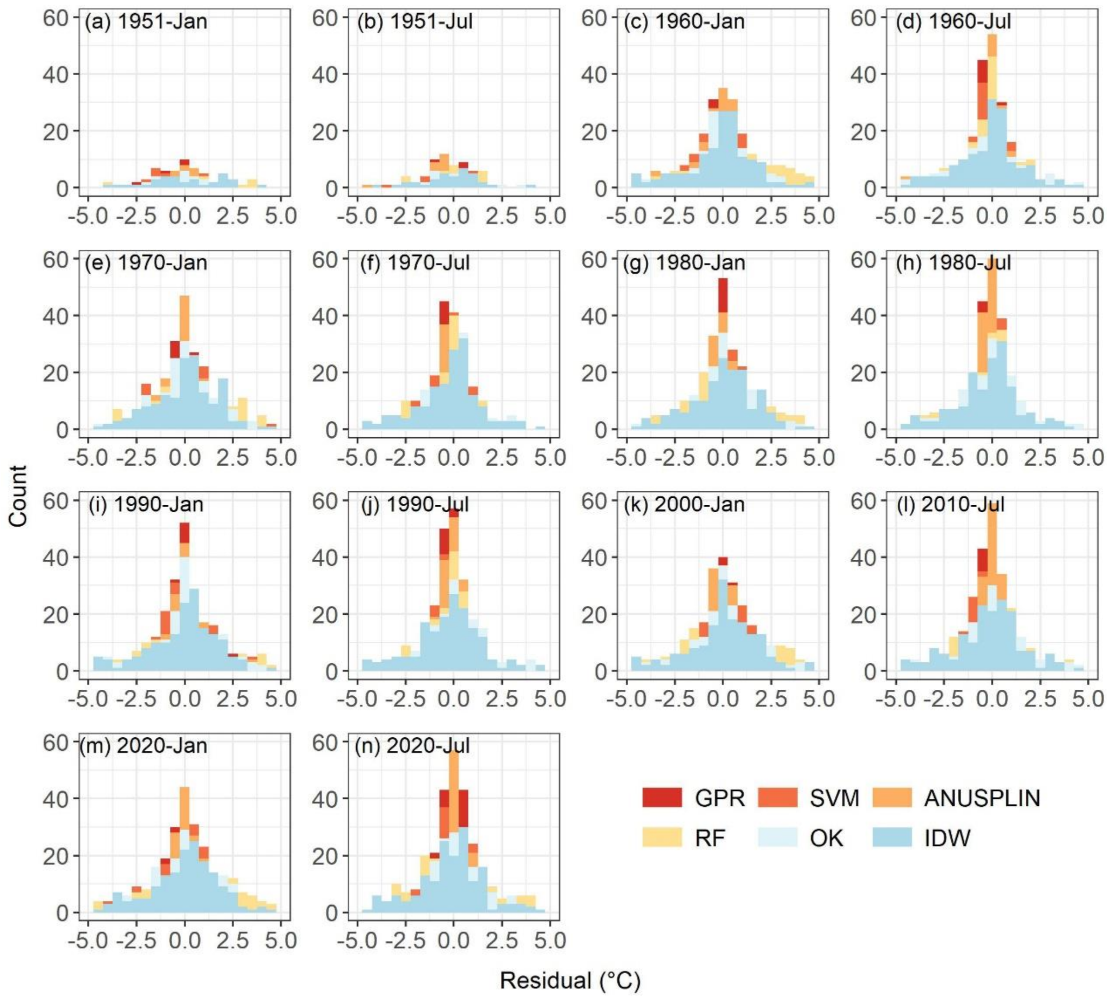

Figure S24: Frequency distribution of residuals for Tmean by three machine learning methods and three traditional methods for separate 8 years from 1951 to 2020 with an interval of 10 years.

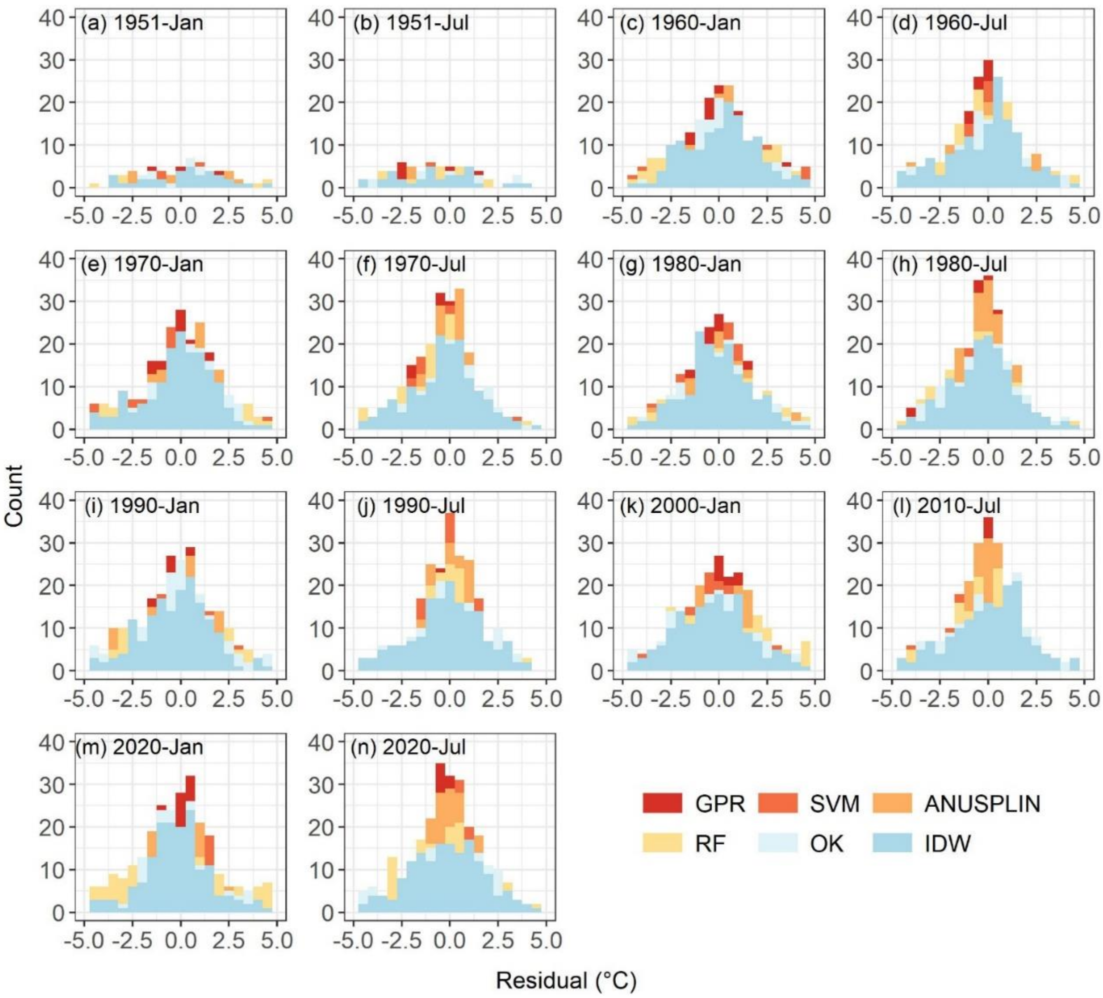

Figure S25: Frequency distribution of residuals for Tmax by three machine learning methods and three traditional methods for separate 8 years from 1951 to 2020 with an interval of 10 years.

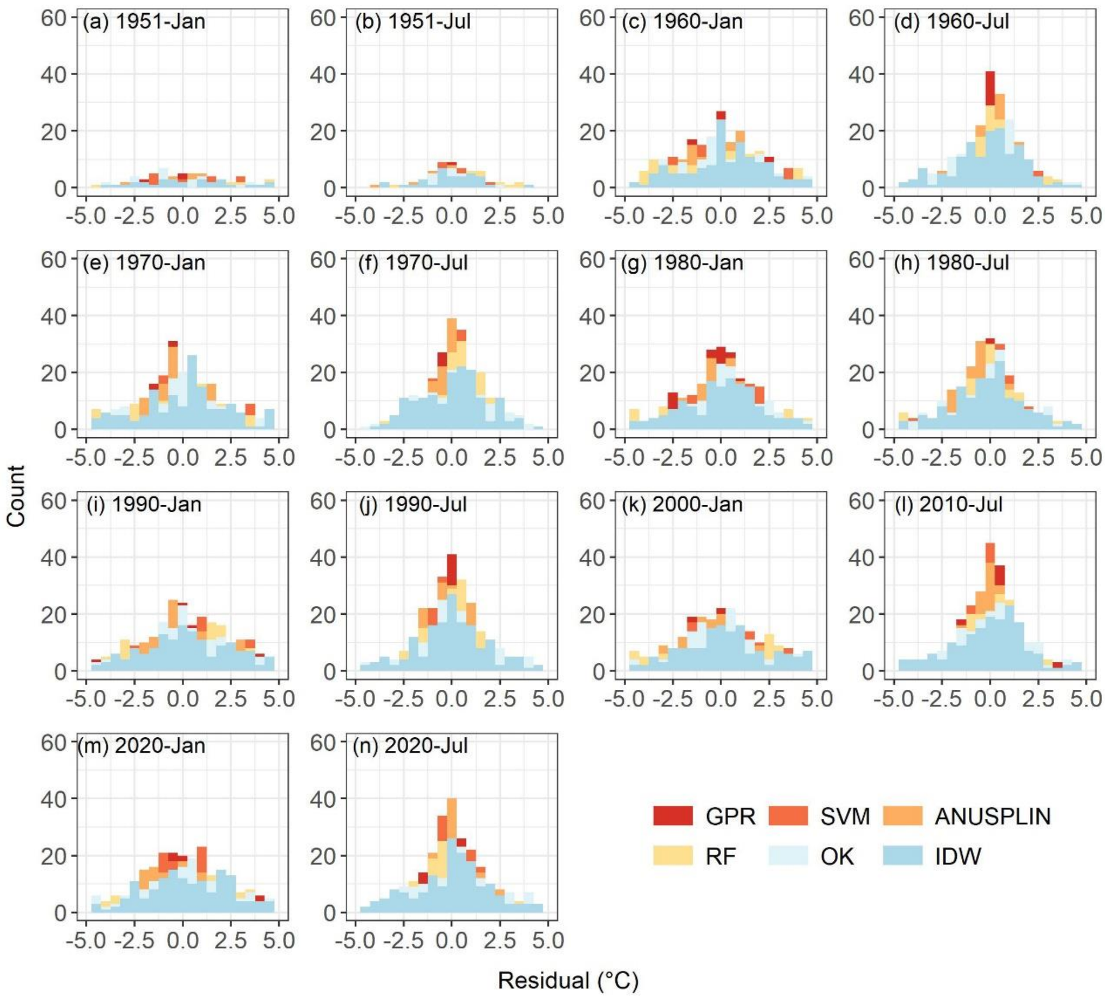

Figure S26: Frequency distribution of residuals for Tmin by three machine learning methods and three traditional methods for separate 8 years from 1951 to 2020 with an interval of 10 years.

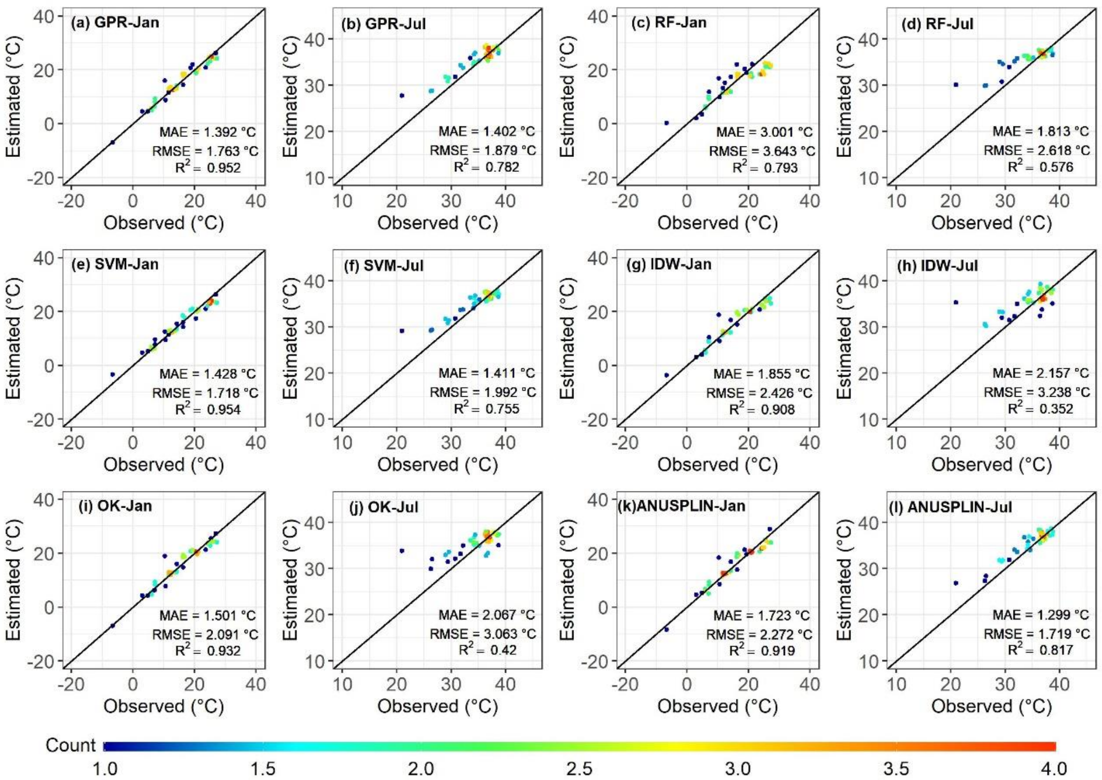

Figure S27: Scatterplot of estimated Tmax by machine learning models and traditional models against the observed monthly mean temperature of January and July in 1951.

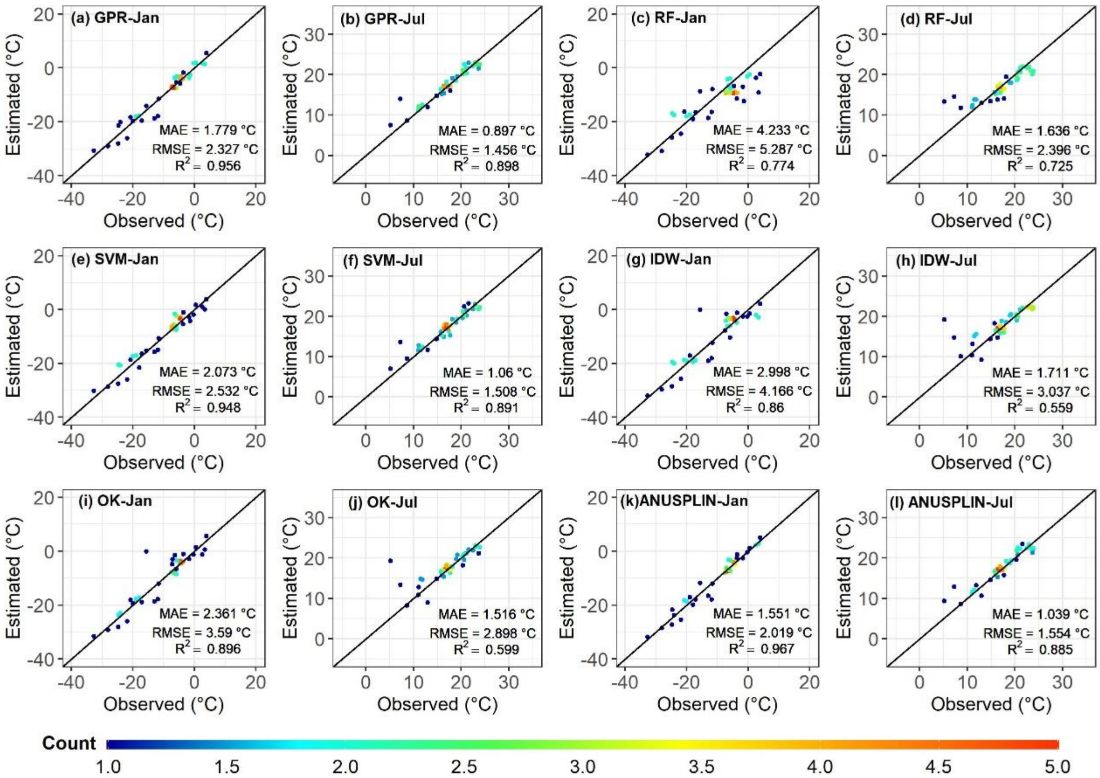

Figure S28: Scatterplot of estimated Tmin by machine learning models and traditional models against the observed monthly mean temperature of January and July in 1951.

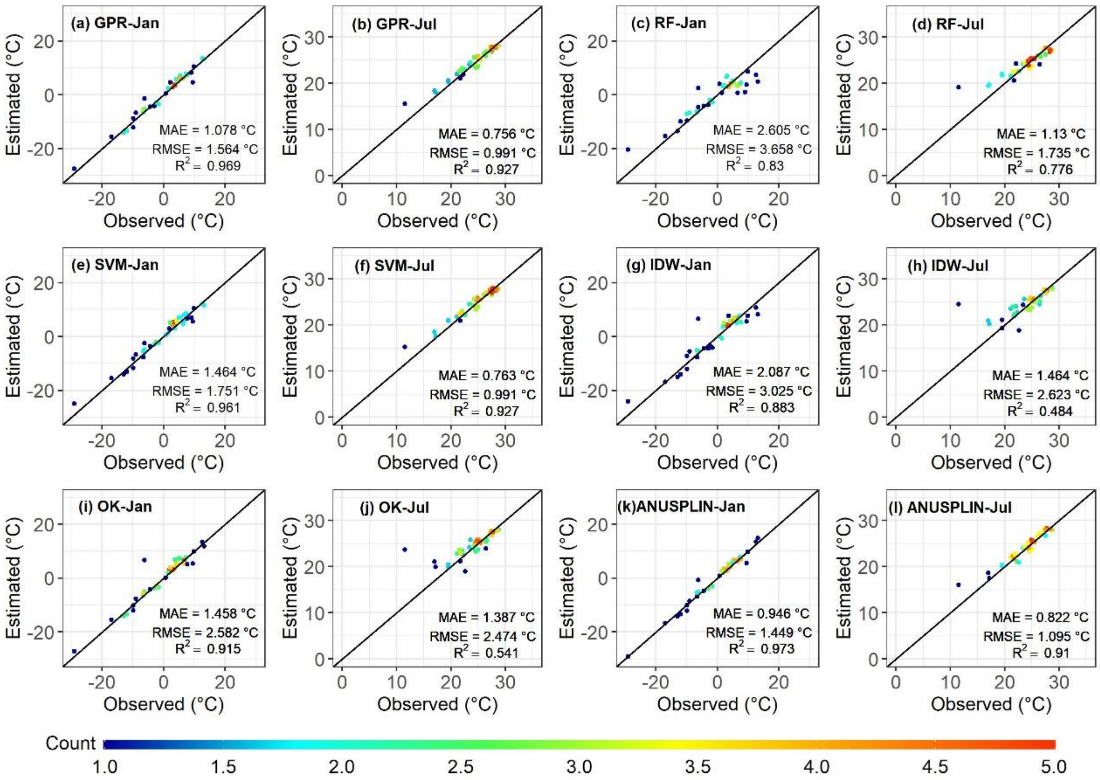

Figure S29: Scatterplot of estimated Tmean by machine learning models and traditional models against the observed monthly mean temperature of January and July in 1951.

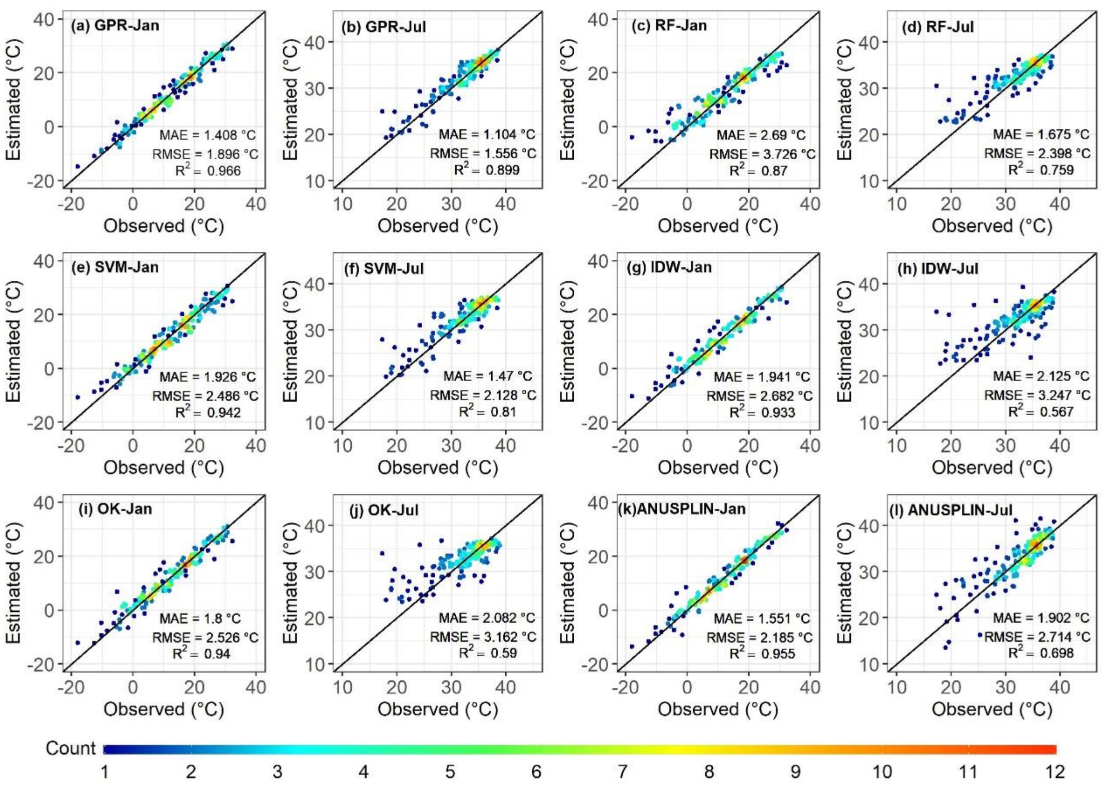

Figure S30: Scatterplot of estimated Tmax by machine learning models and traditional models against the observed monthly mean temperature of January and July in 1960.

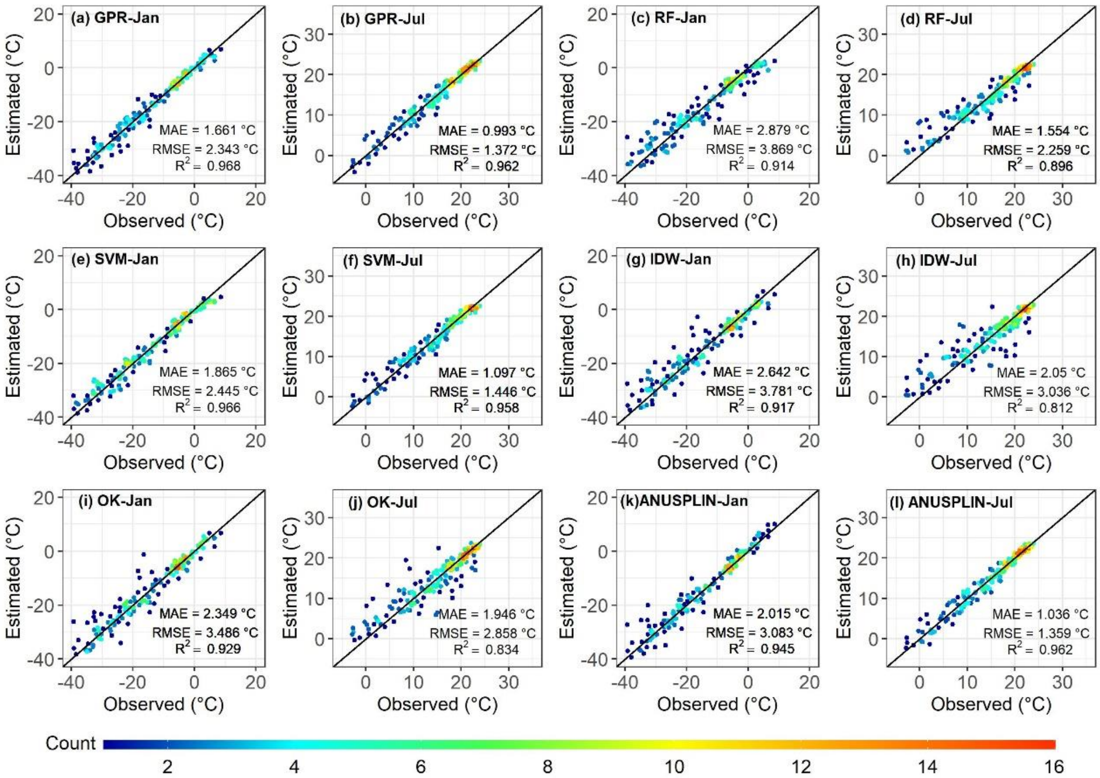

Figure S31: Scatterplot of estimated Tmin by machine learning models and traditional models against the observed monthly mean temperature of January and July in 1960.

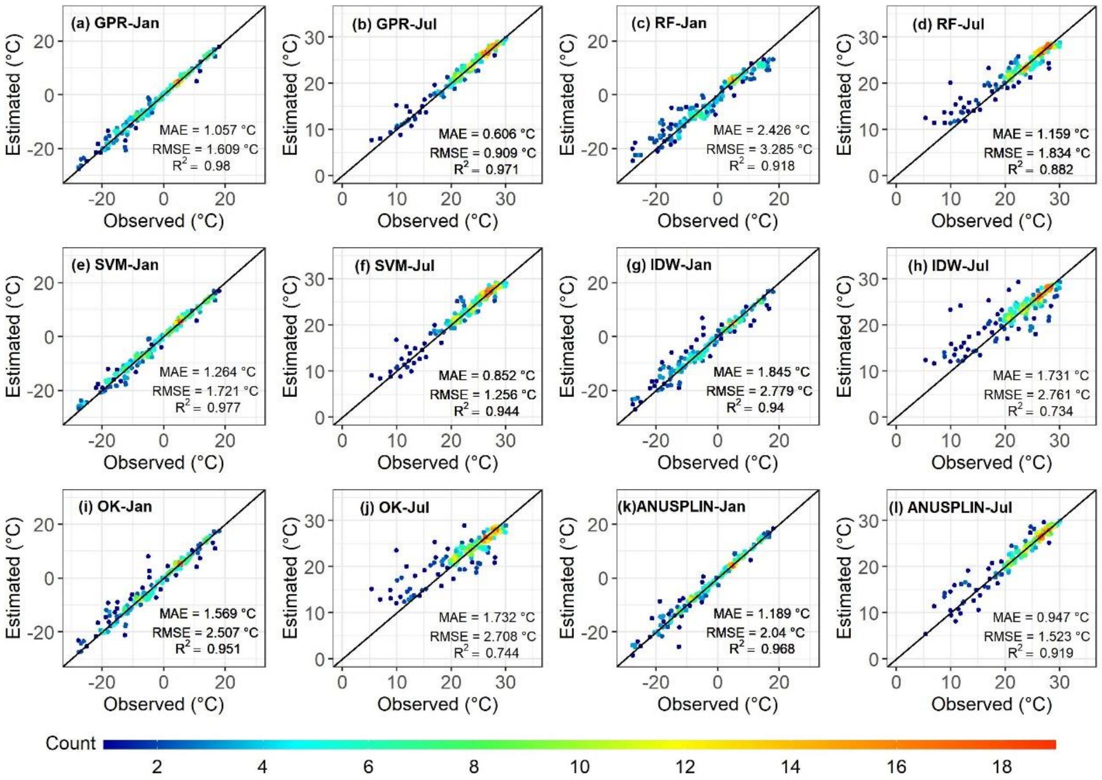

Figure S32: Scatterplot of estimated Tmean by machine learning models and traditional models against the observed monthly mean temperature of January and July in 1960.

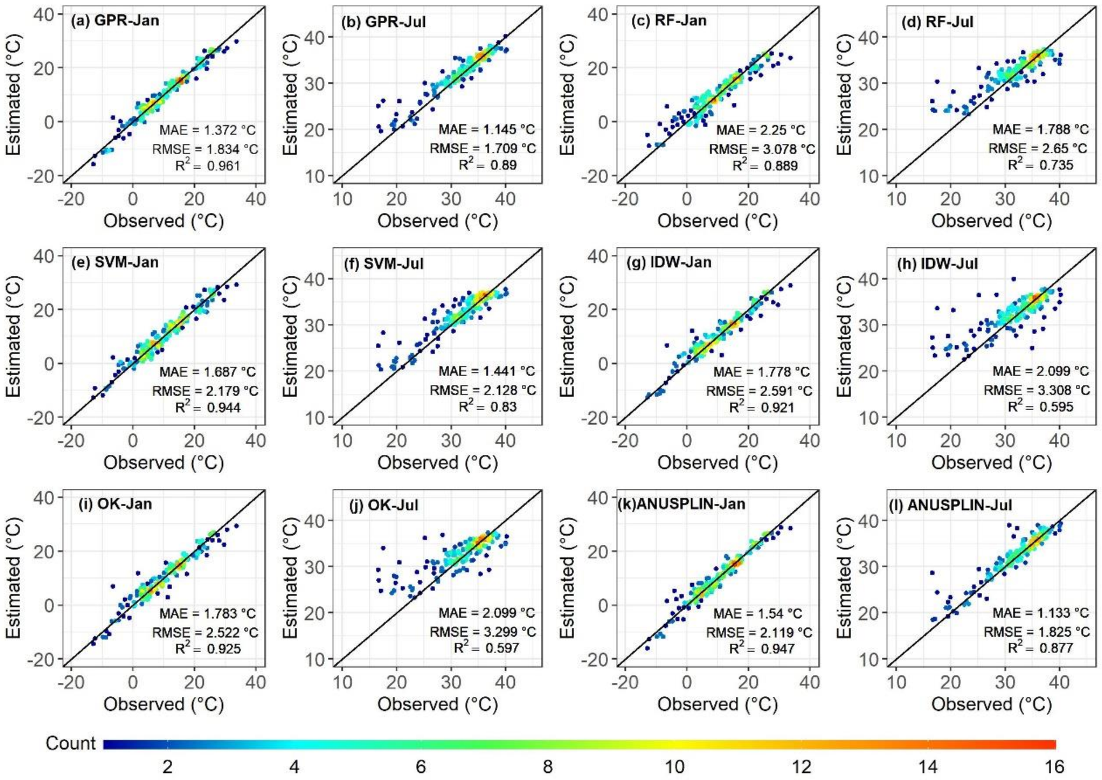

Figure S33: Scatterplot of estimated Tmax by machine learning models and traditional models against the observed monthly mean temperature of January and July in 1970.

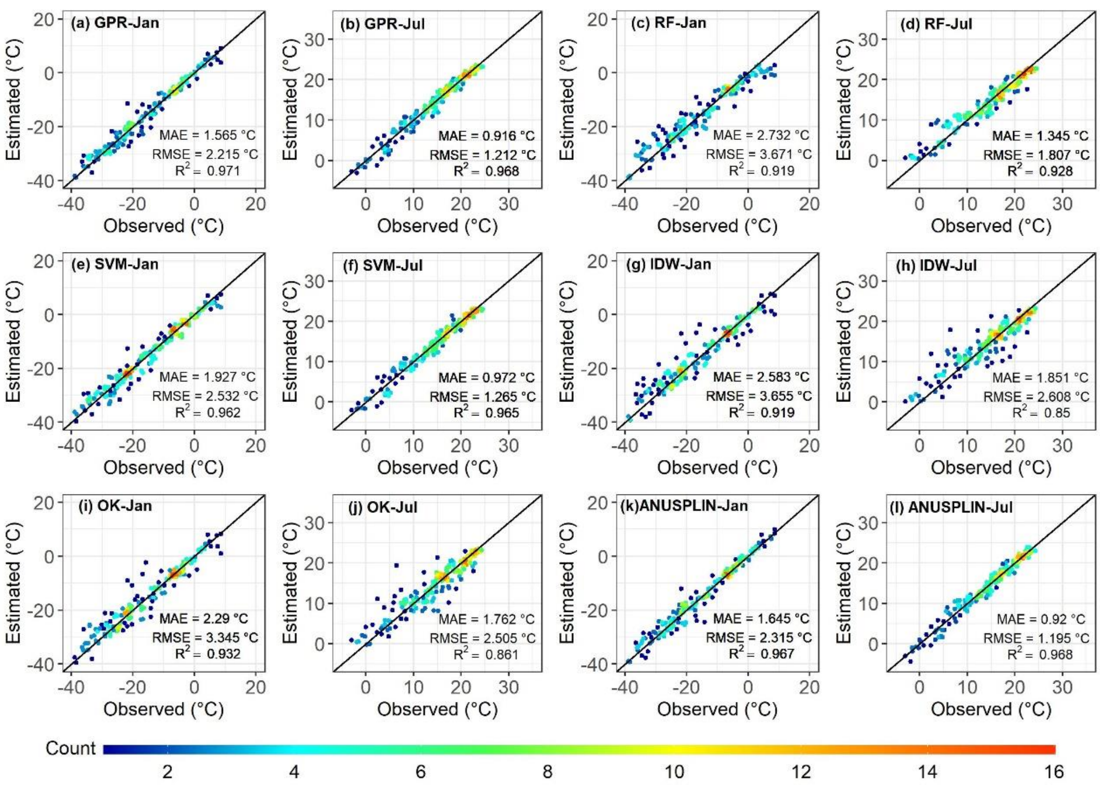

Figure S34: Scatterplot of estimated Tmin by machine learning models and traditional models against the observed monthly mean temperature of January and July in 1970.

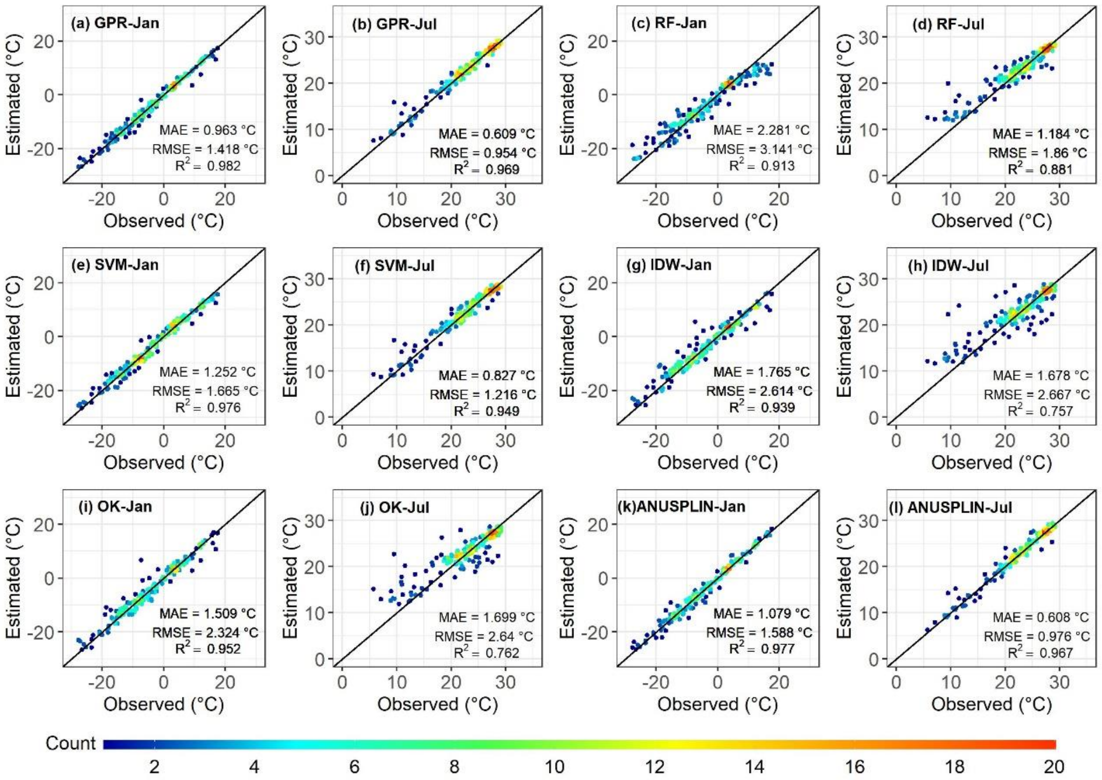

Figure S35: Scatterplot of estimated Tmean by machine learning models and traditional models against the observed monthly mean temperature of January and July in 1970.

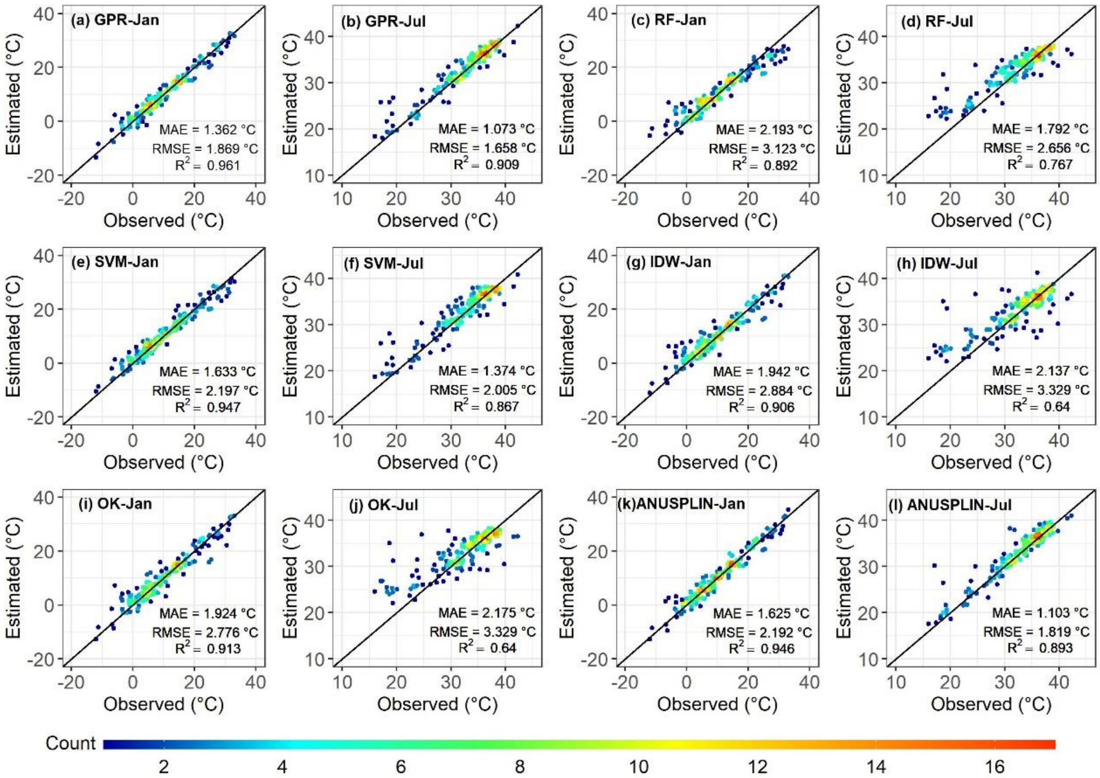

Figure S36: Scatterplot of estimated Tmax by machine learning models and traditional models against the observed monthly mean temperature of January and July in 1980.

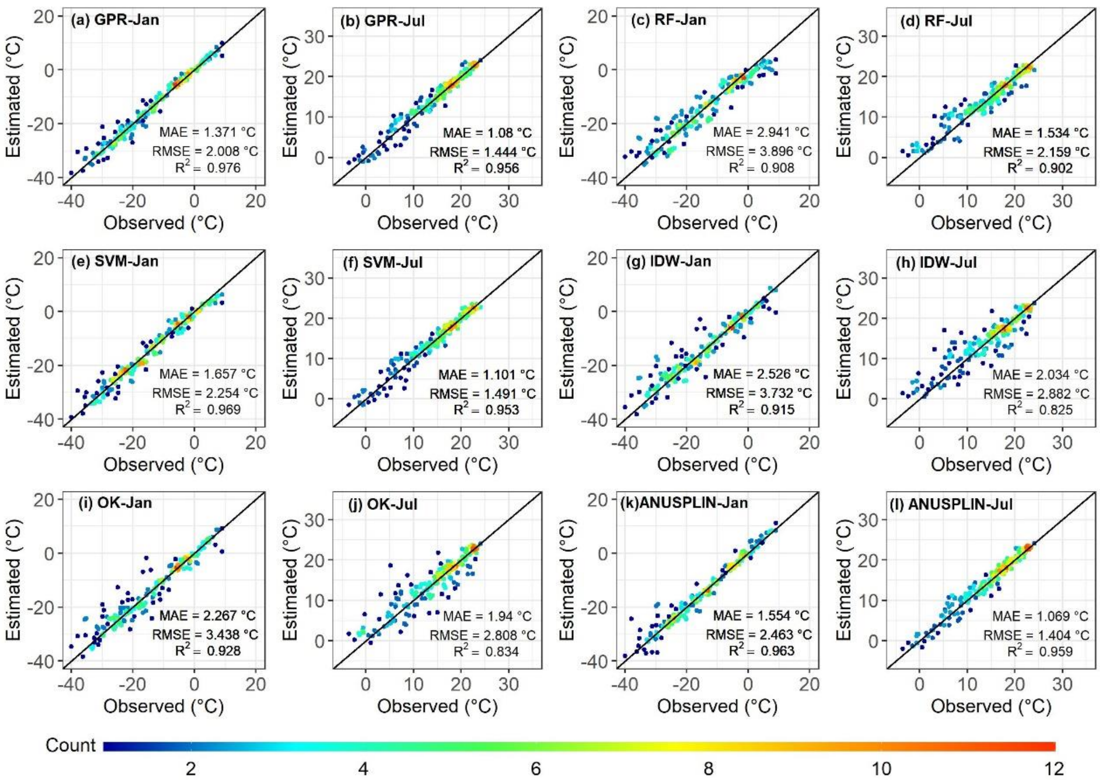

Figure S37: Scatterplot of estimated Tmin by machine learning models and traditional models against the observed monthly mean temperature of January and July in 1980.

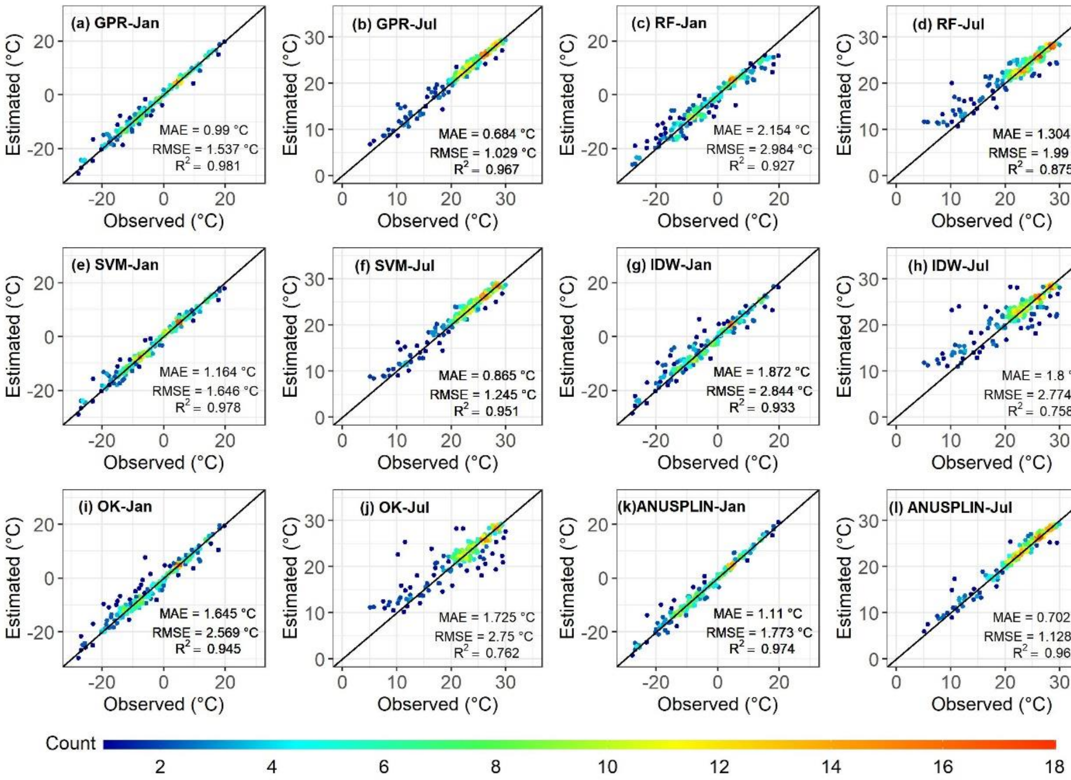

Figure S38: Scatterplot of estimated Tmean by machine learning models and traditional models against the observed monthly mean temperature of January and July in 1980.

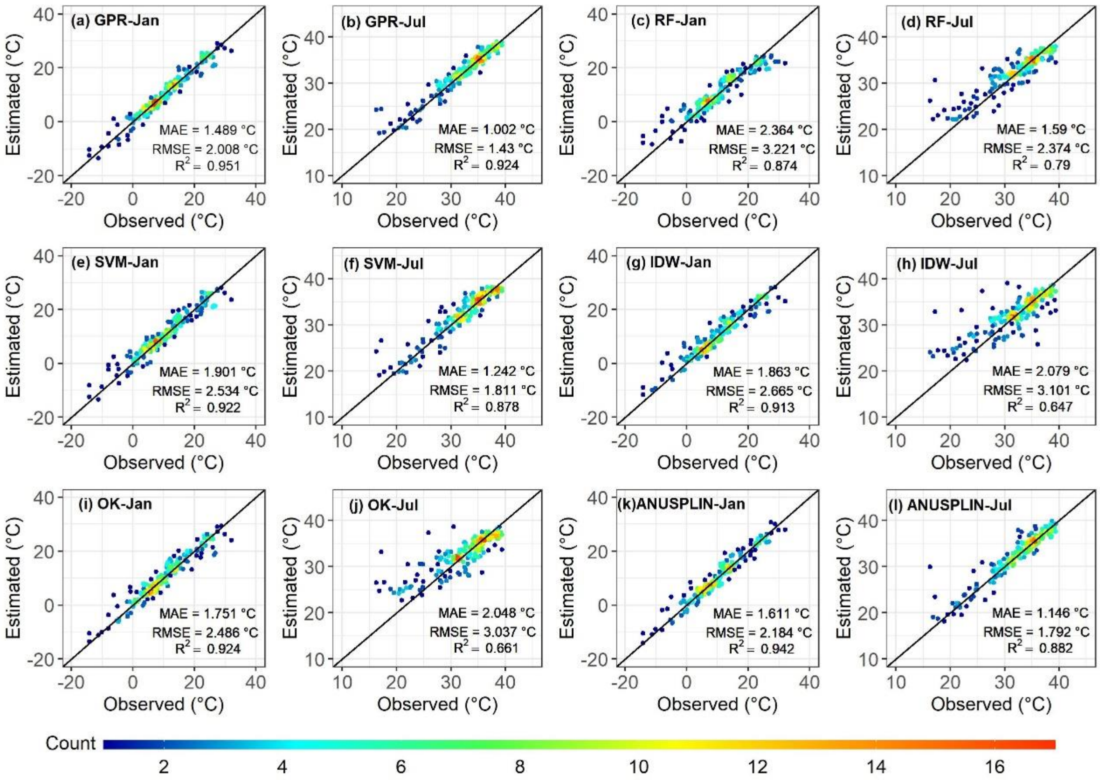

Figure S39: Scatterplot of estimated Tmax by machine learning models and traditional models against the observed monthly mean temperature of January and July in 1990.

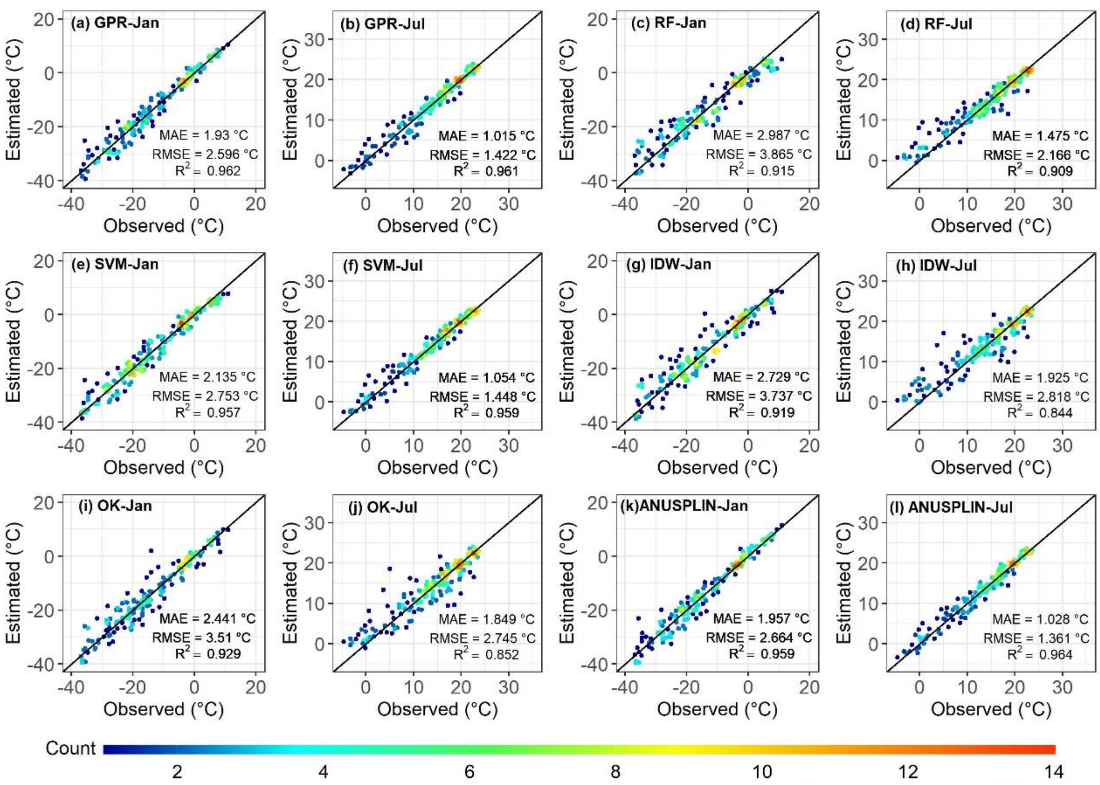

Figure S40: Scatterplot of estimated Tmin by machine learning models and traditional models against the observed monthly mean temperature of January and July in 1990.

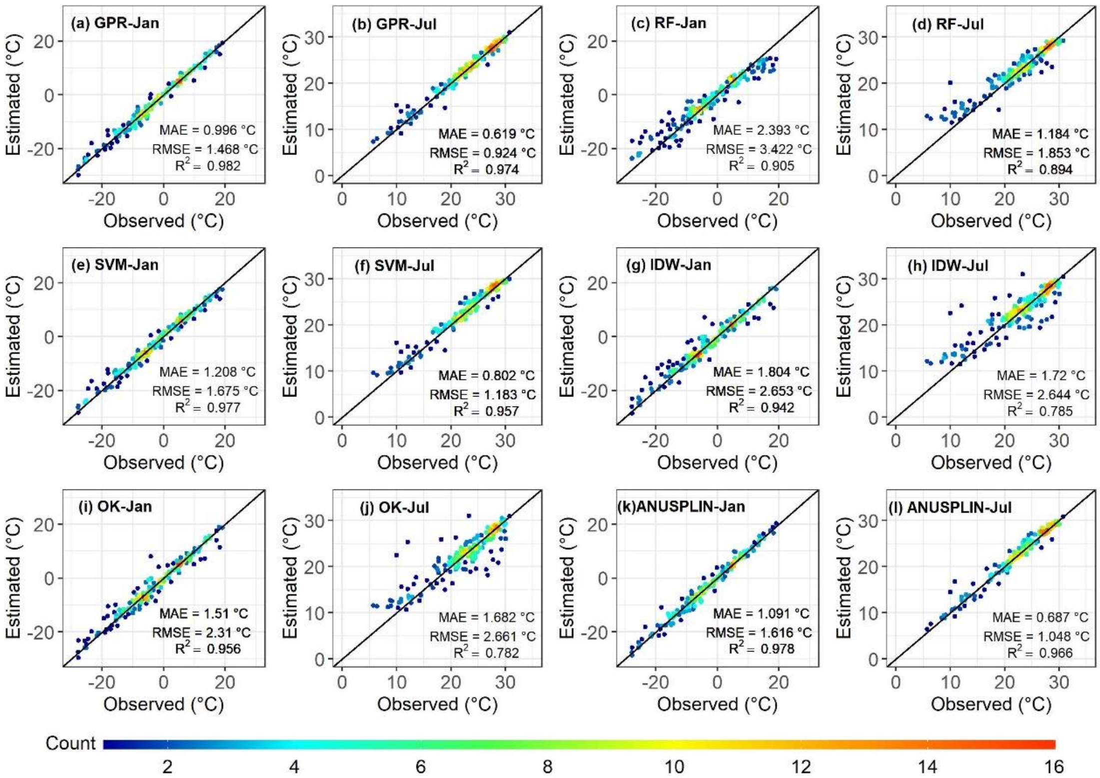

Figure S41: Scatterplot of estimated Tmean by machine learning models and traditional models against the observed monthly mean temperature of January and July in 1990.

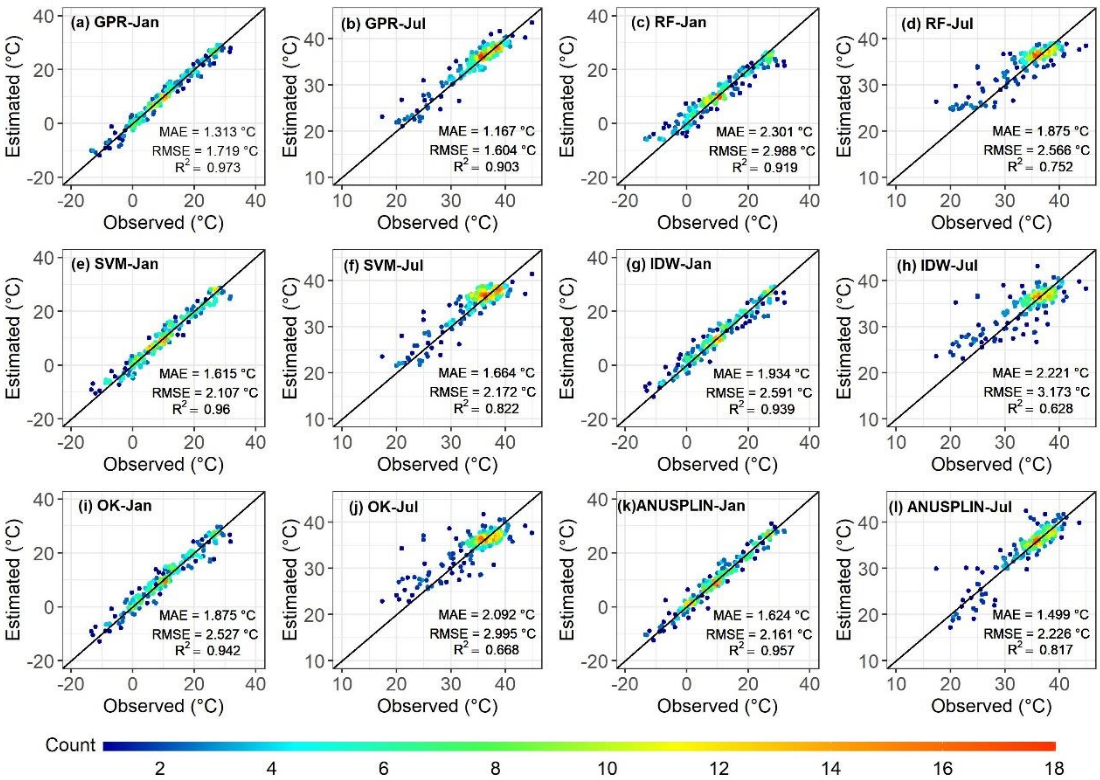

Figure S42: Scatterplot of estimated Tmax by machine learning models and traditional models against the observed monthly mean temperature of January and July in 2000.

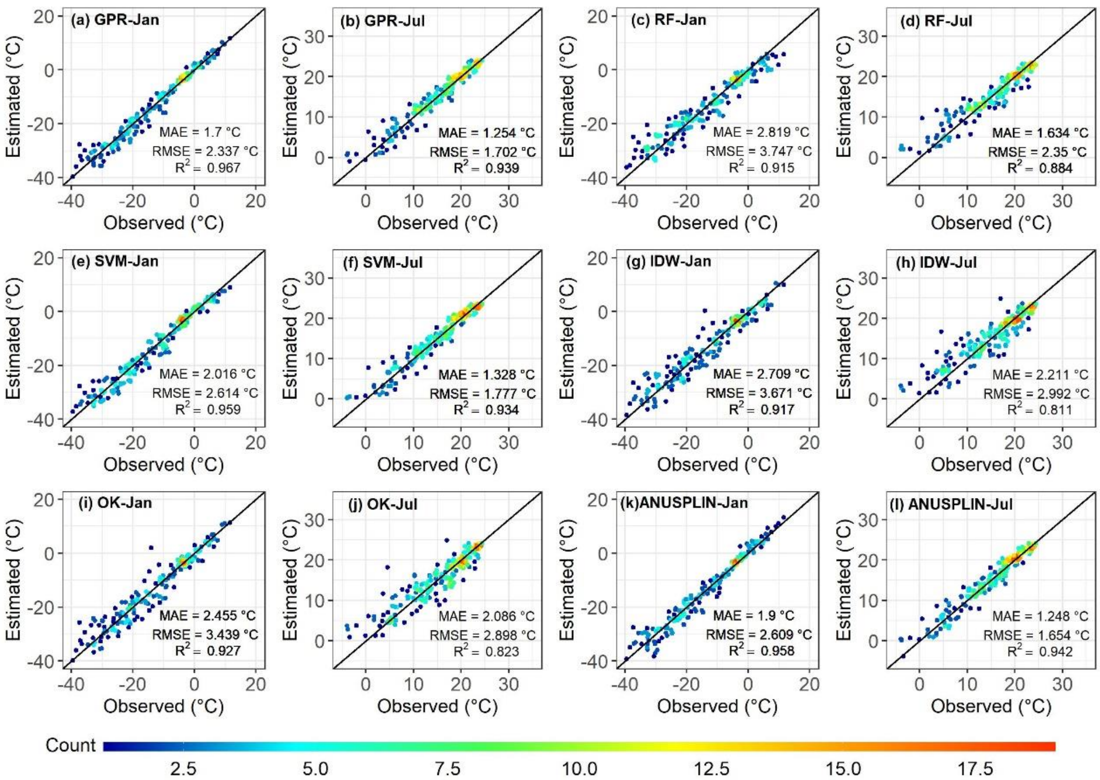

Figure S43: Scatterplot of estimated Tmin by machine learning models and traditional models against the observed monthly mean temperature of January and July in 2000.

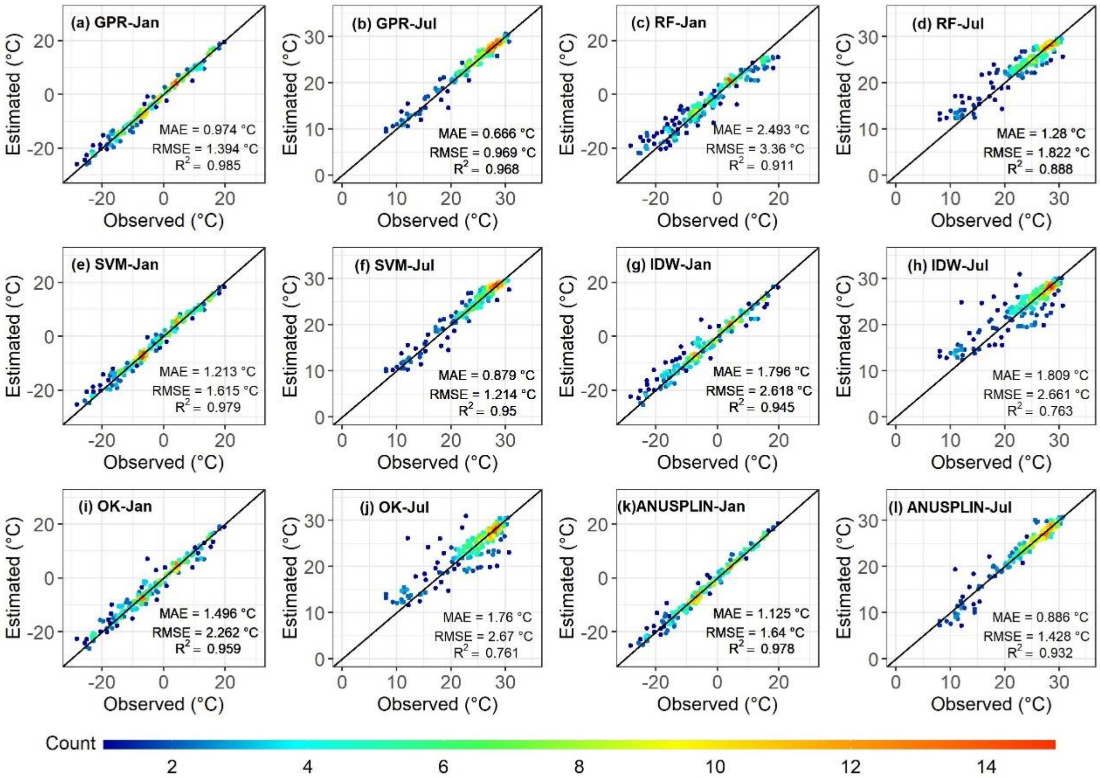

Figure S44: Scatterplot of estimated Tmean by machine learning models and traditional models against the observed monthly mean temperature of January and July in 2000.

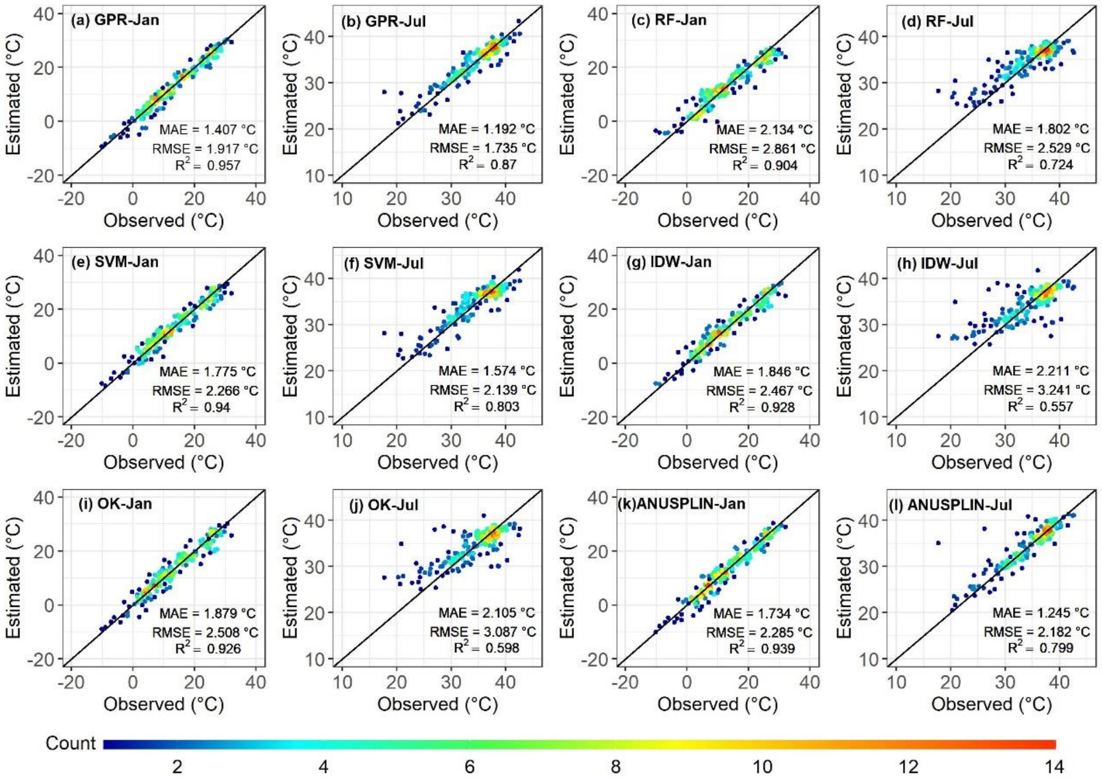

Figure S45: Scatterplot of estimated Tmax by machine learning models and traditional models against the observed monthly mean temperature of January and July in 2010.

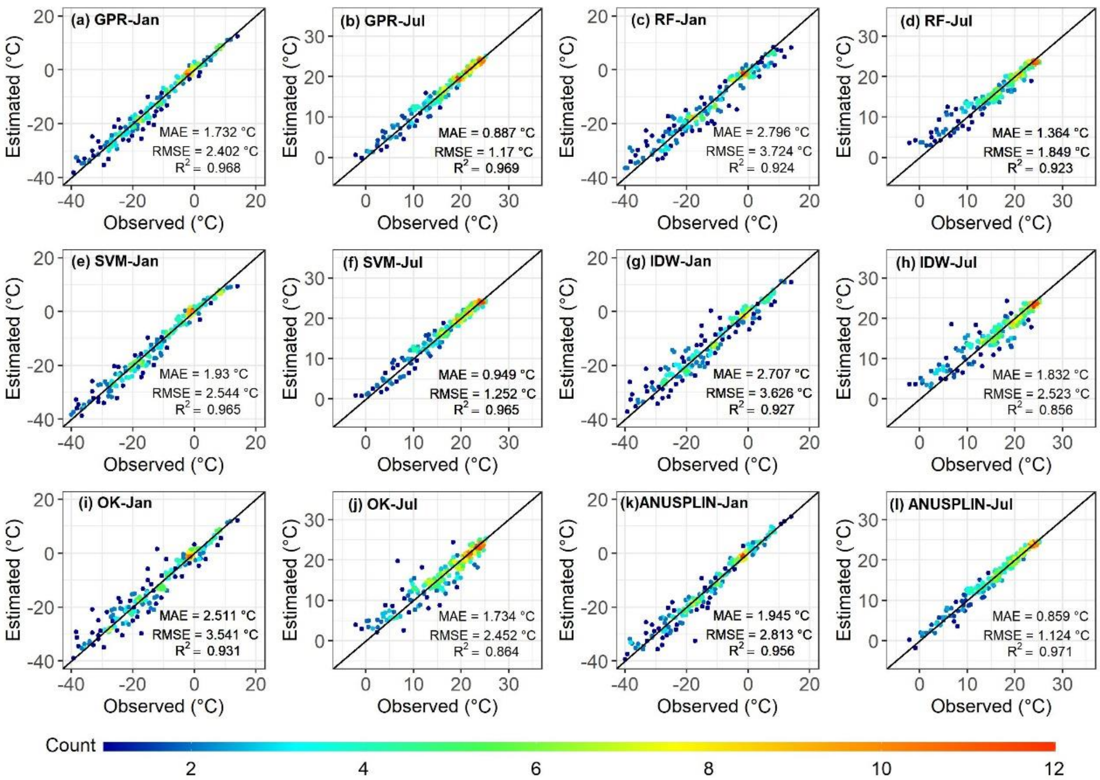

Figure S46: Scatterplot of estimated Tmin by machine learning models and traditional models against the observed monthly mean temperature of January and July in 2010.

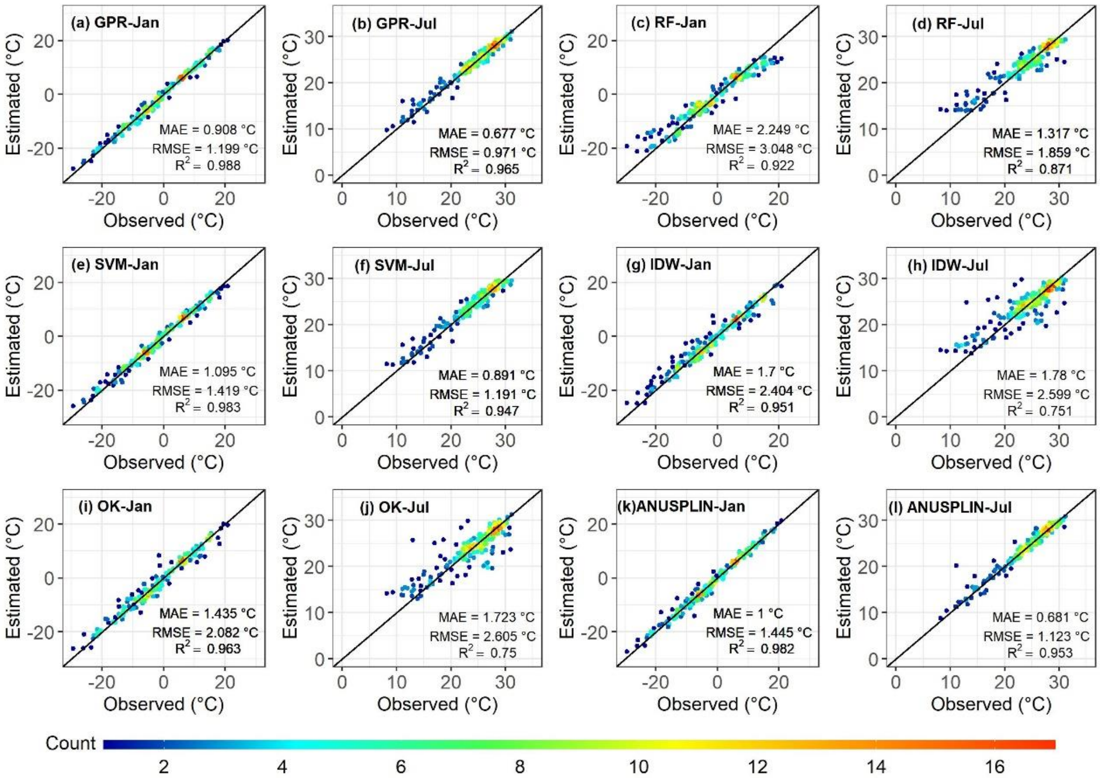

Figure S47: Scatterplot of estimated Tmean by machine learning models and traditional models against the observed monthly mean temperature of January and July in 2010.

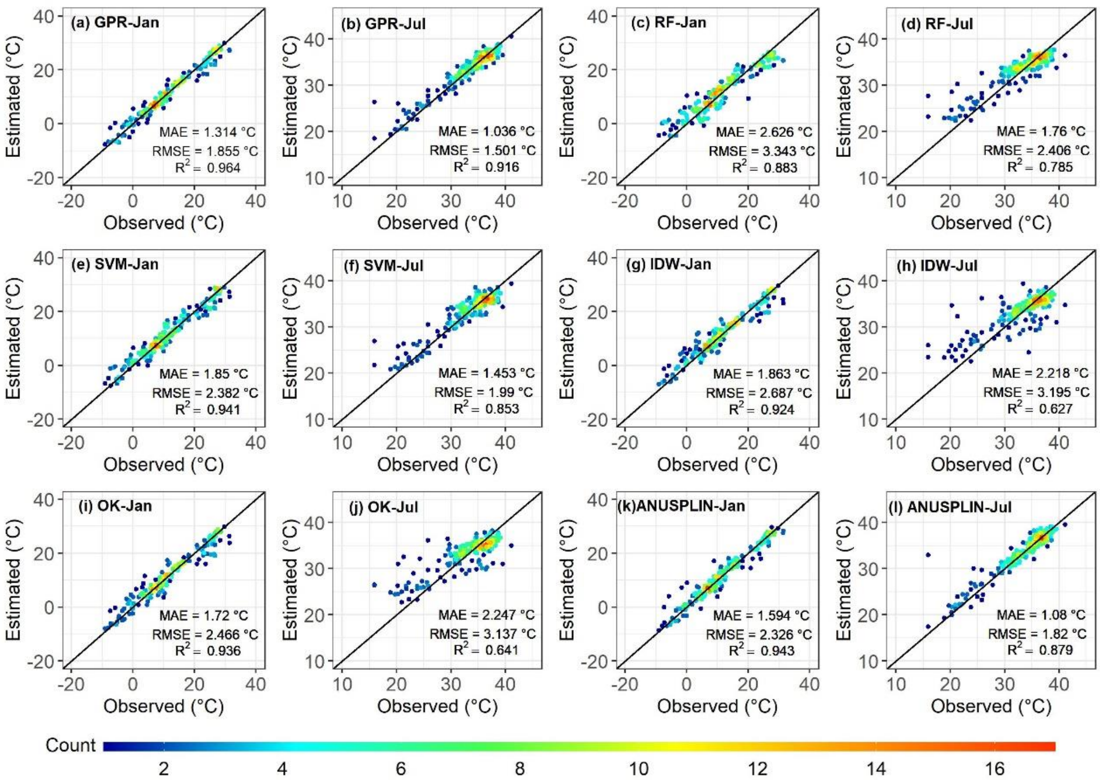

Figure S48: Scatterplot of estimated Tmax by machine learning models and traditional models against the observed monthly mean temperature of January and July in 2020.

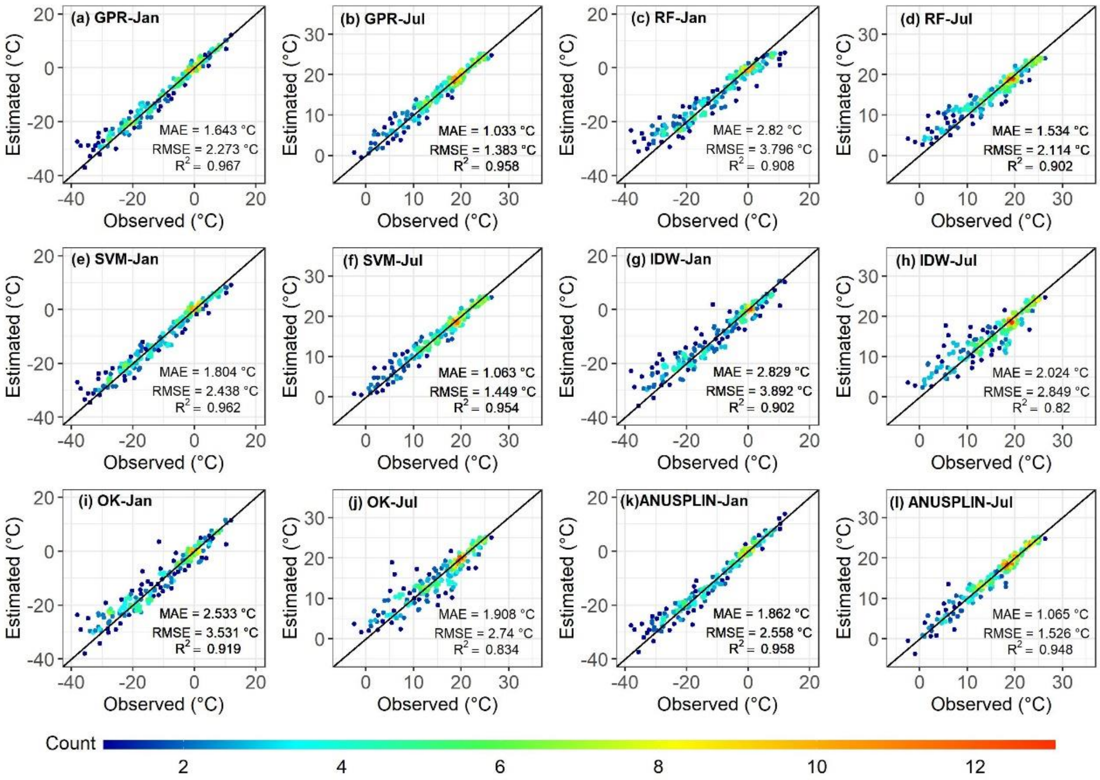

Figure S49: Scatterplot of estimated Tmin by machine learning models and traditional models against the observed monthly mean temperature of January and July in 2010.

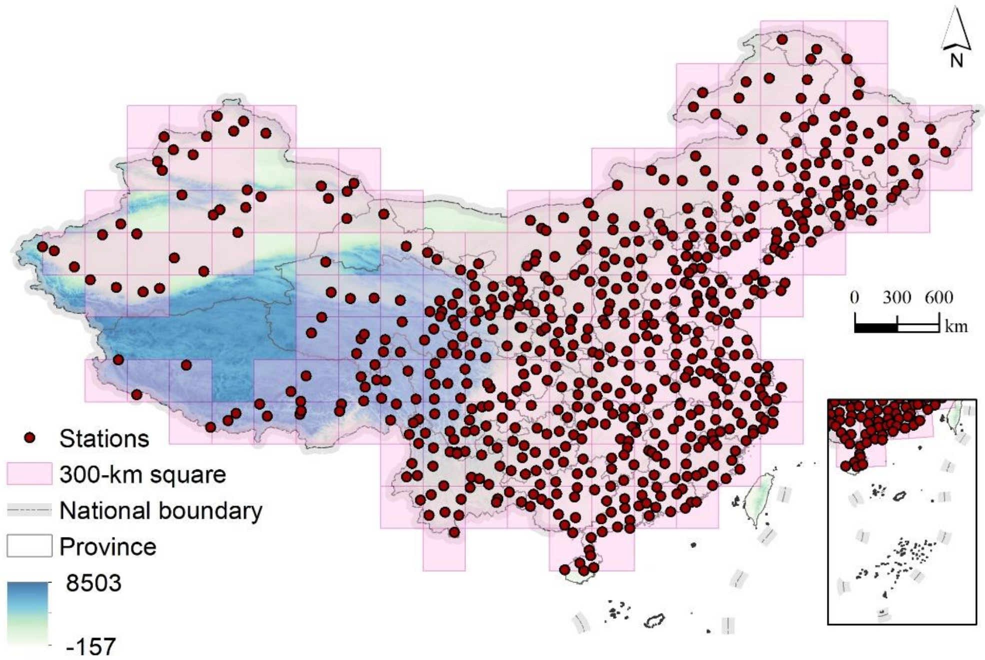

Figure S50: Illustration of the squares used for the experiments.

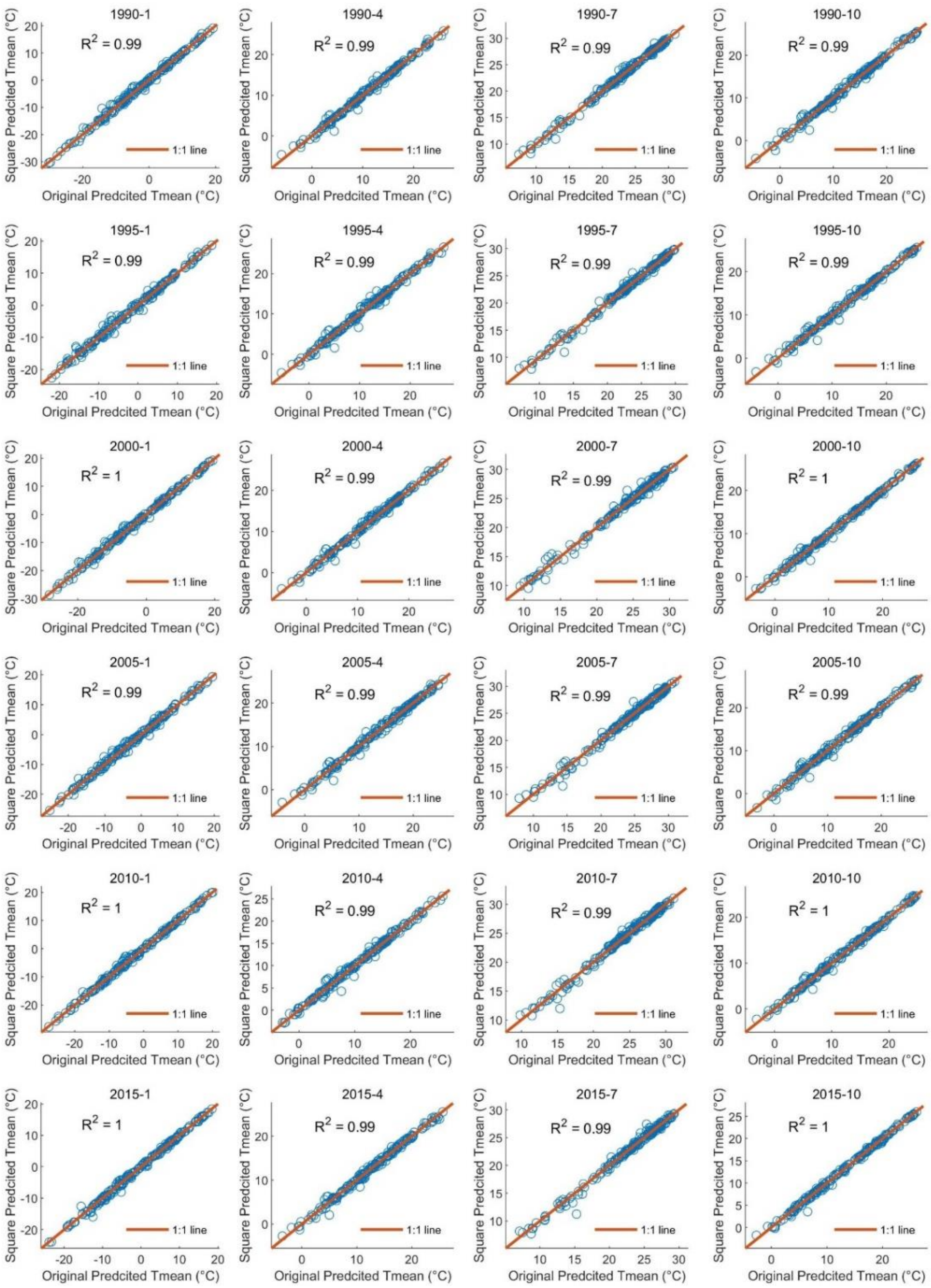

Figure S 51: Scatter plots between the original predicted Tmean and the predicted Tmean using the square for the testing stations.

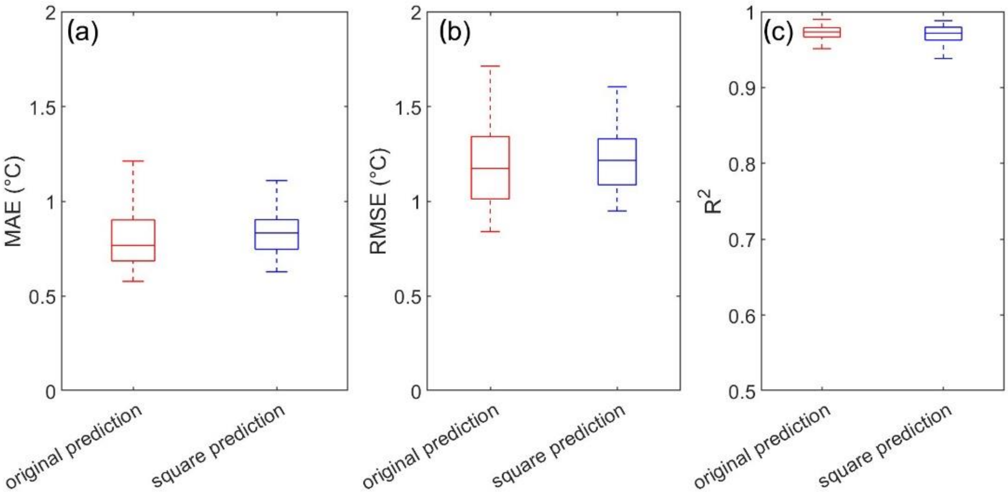

Figure S52: The accuracy comparison between the original prediction and the new prediction using the square strategy. The data are the Tmean for each month in 1980, 1983, 1985, 1988, 1990, 1993, 1995, 1998, 2000, 2003, 2005, 2008, 2010, 2013, 2015, and 2018.

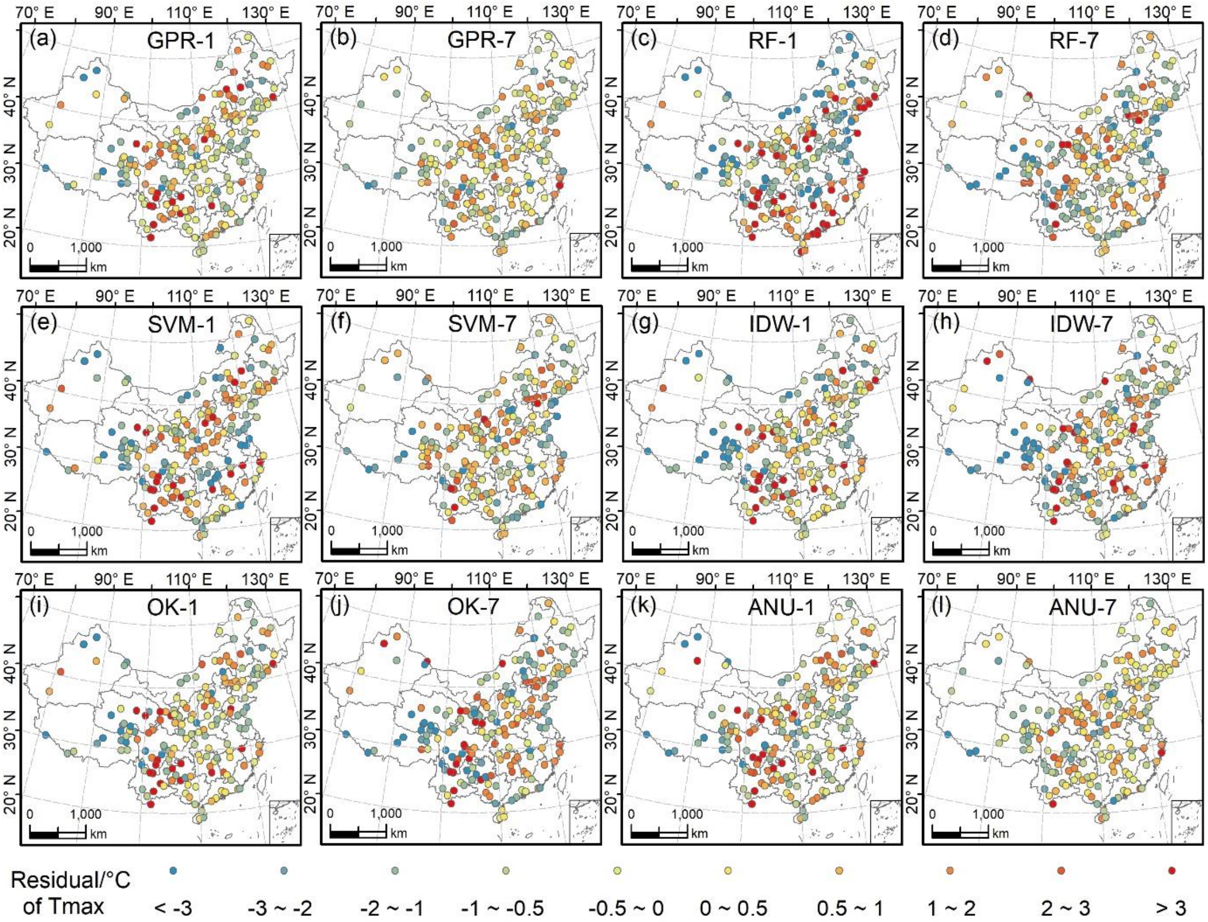

Figure S53: Comparison of the spatial distribution of residuals between machine learning and traditional methods for Tmax in January and July of 2020.

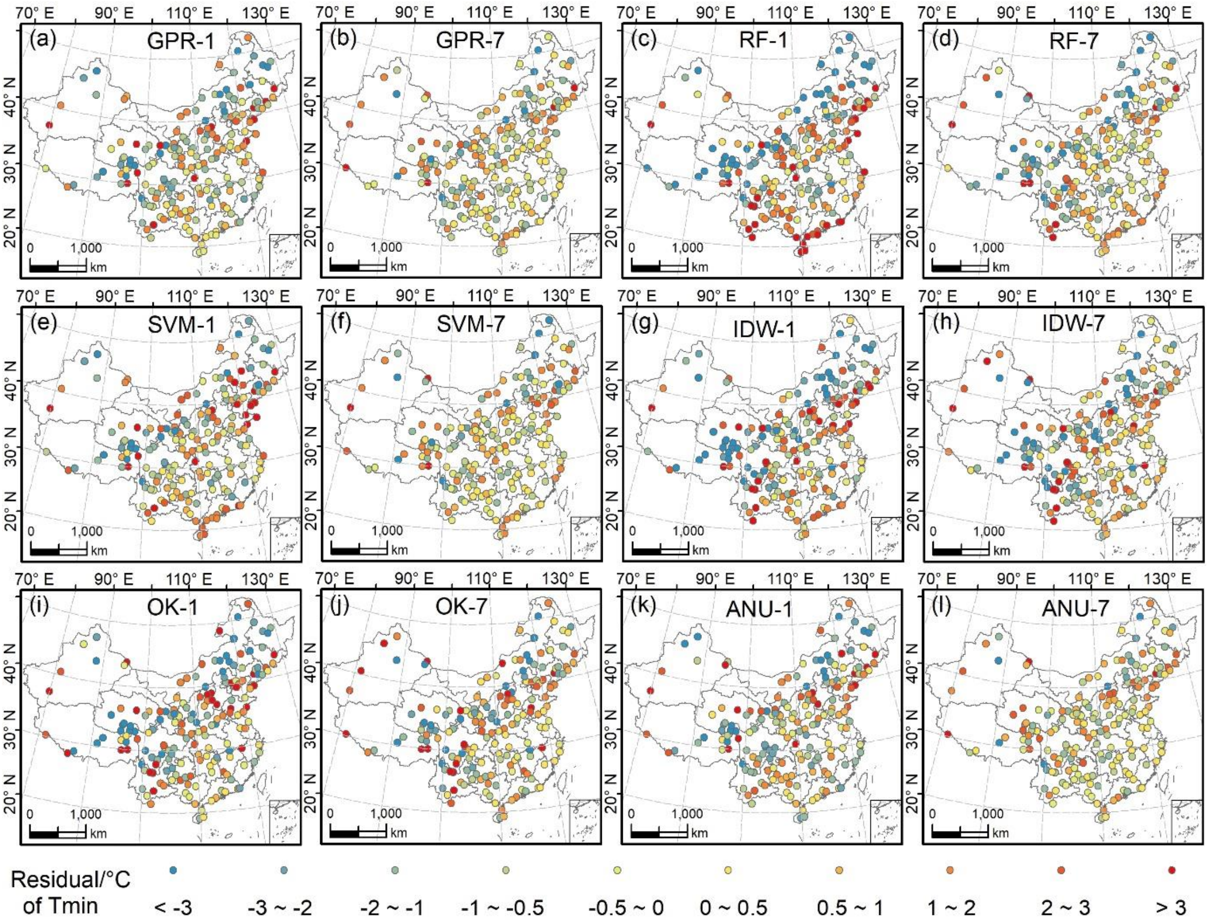

Figure S54: Comparison of the spatial distribution of residuals between machine learning and traditional methods for Tmin in January and July of 2020.

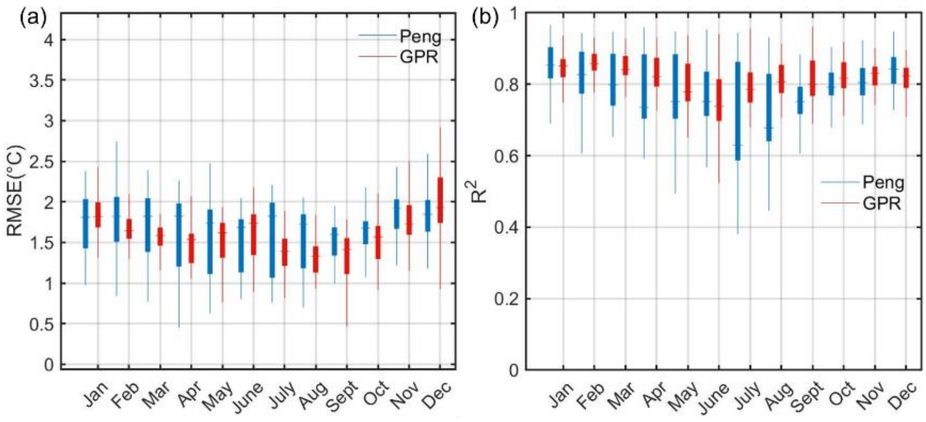

Figure S55: Comparison between the GPR data in our study and the Peng's data in the Tibetan Plateau.

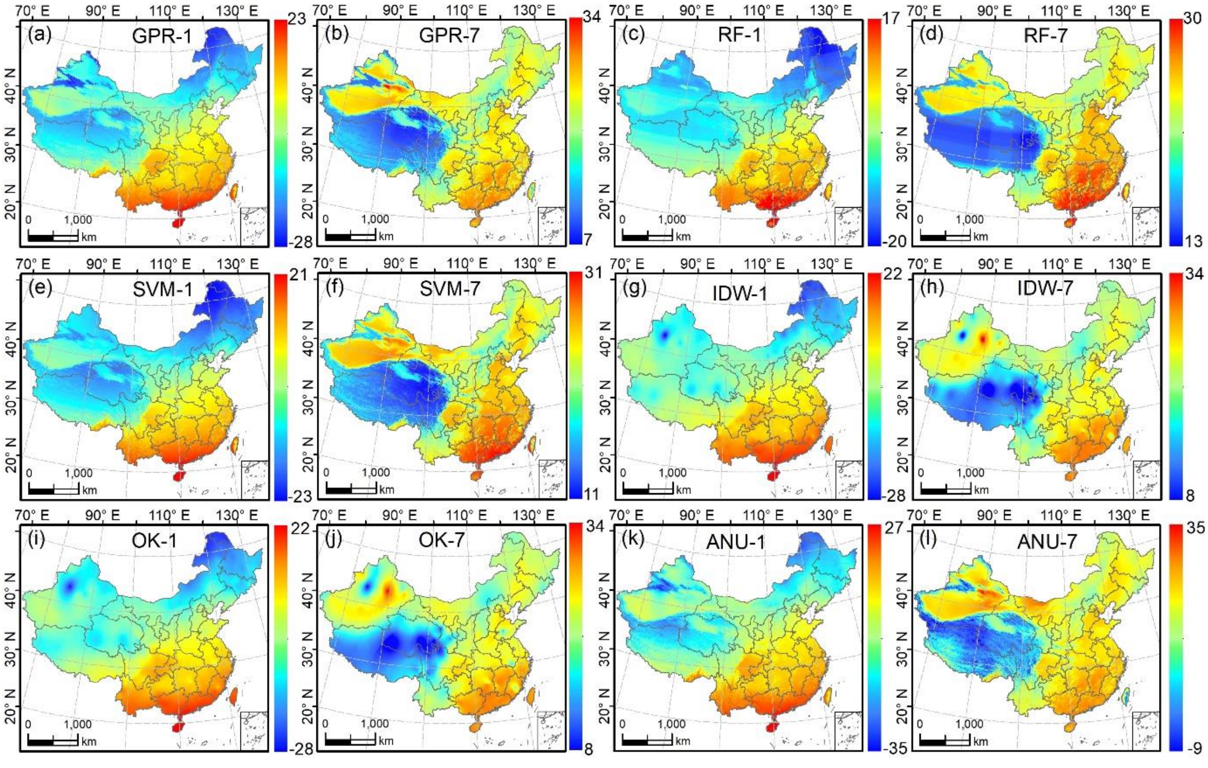

Figure S56: The spatial maps of monthly Tmean over China in January and July of 2020 (1, 7: January, July; unit: °C).

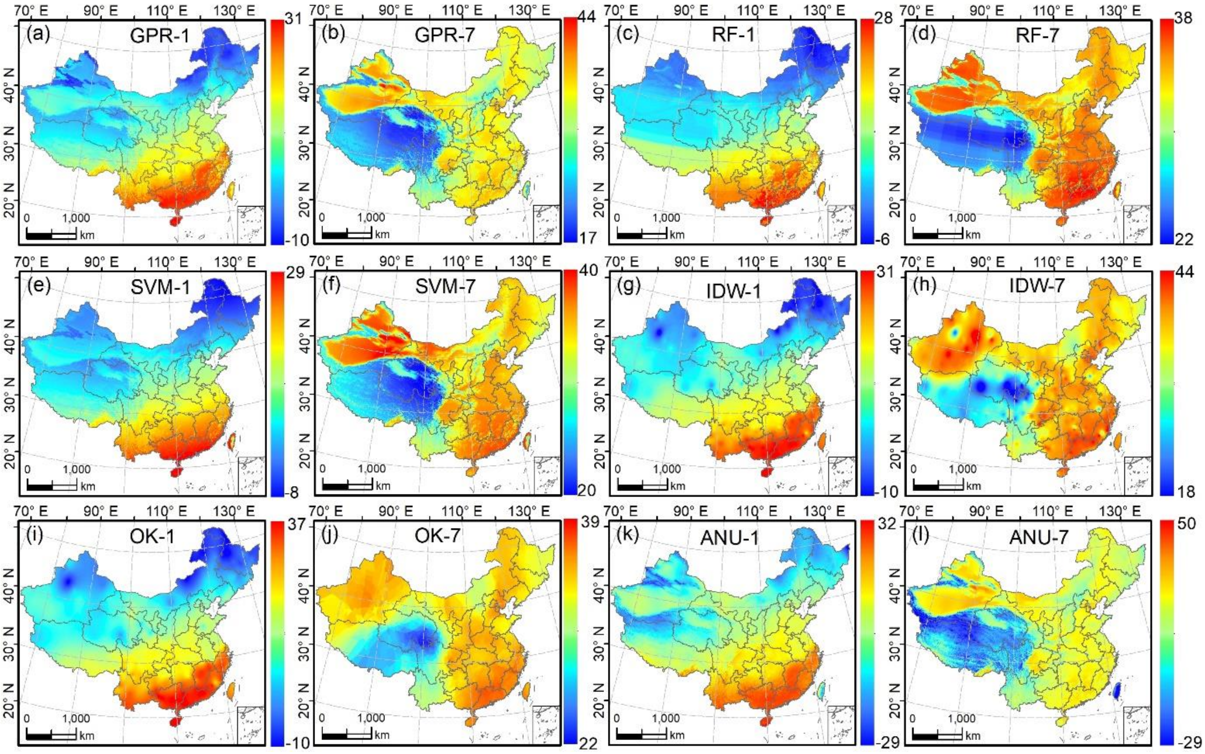

Figure S57: The spatial maps of monthly Tmax over China in January and July of 2020 (1, 7: January, July; unit: °C).

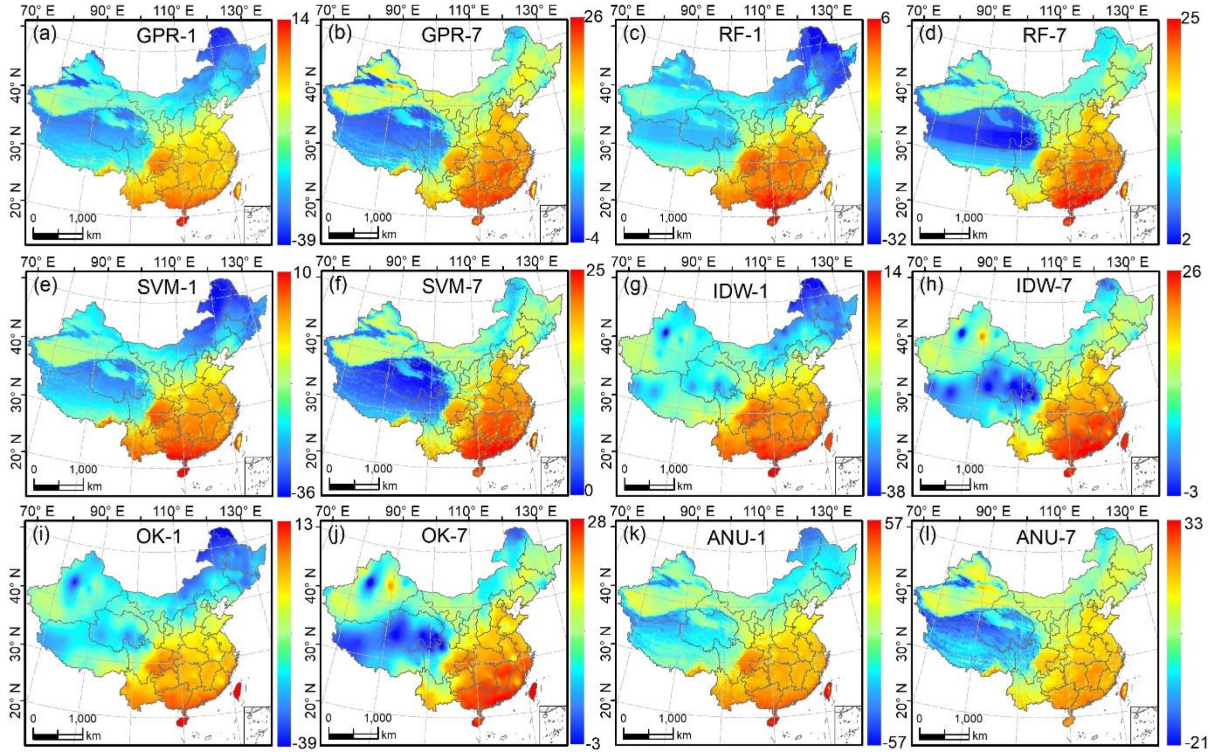

Figure S58: The spatial maps of monthly Tmin over China in January and July of 2020 (1, 7: January, July; unit: °C).

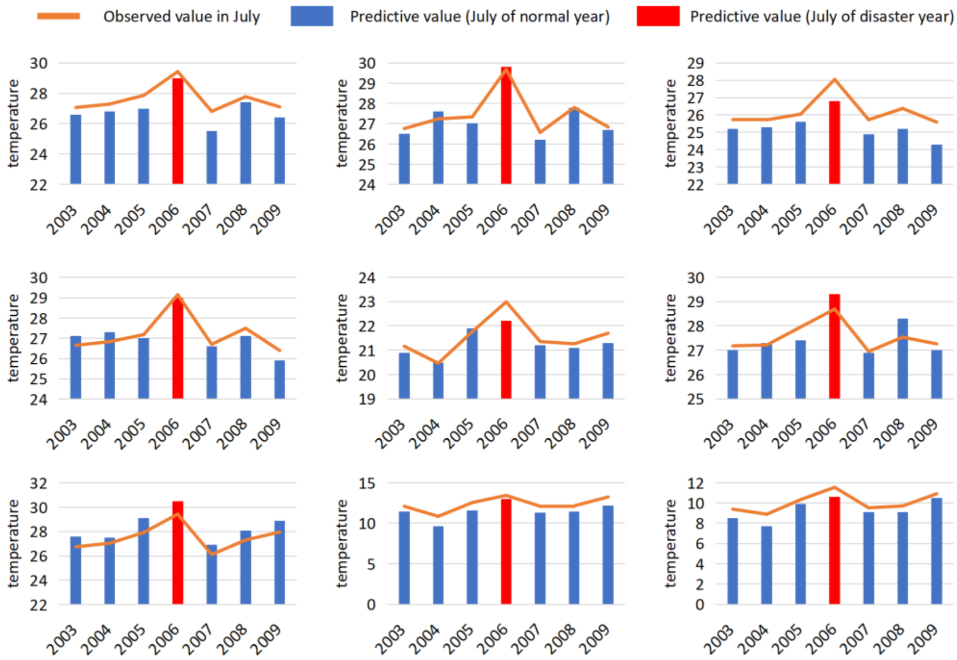

Figure S59: Mean temperature for 9 different testing stations in Sichuan province. We extracted our mean temperature data to the stations which were not used in the model training. We used nine stations in Sichuan and compared the mean temperature in July from 2003 to 2009. The temperature in 2006 is markedly higher than the neighbouring years, which means that our data can capture the anomalous condition (higher temperature) from the 2006 summer drought of eastern Sichuan Basin.

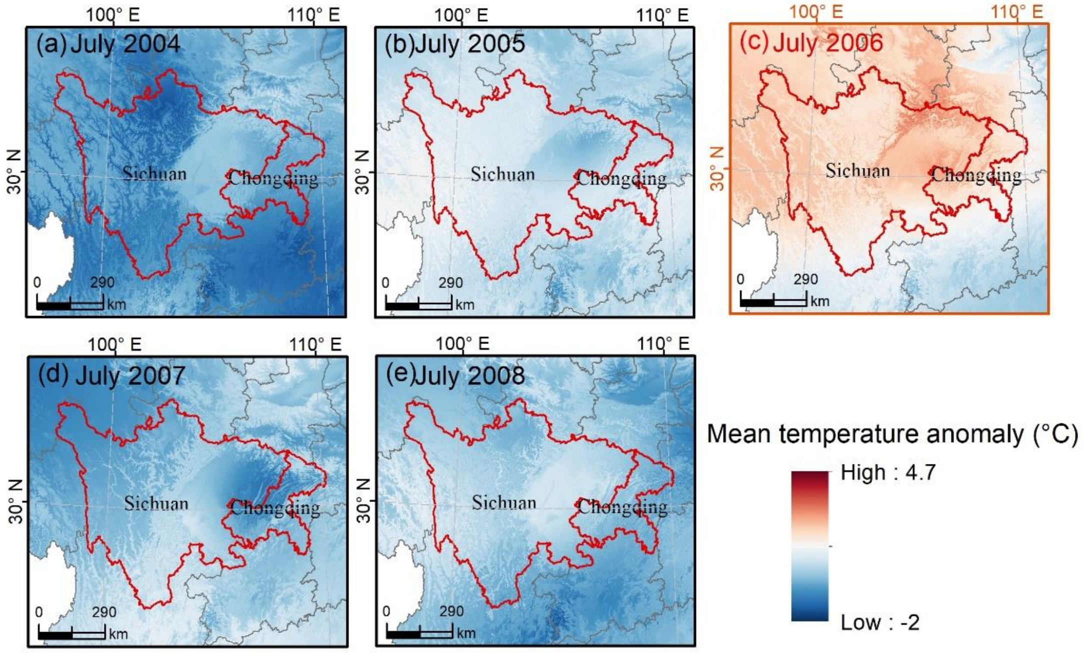

Figure S60: Comparison between the year with higher mean temperature anomaly (for Tmean) and the adjacent years. A drought event happened in Sichuan and Chongqing and neighbouring regions in July 2006 which is associated with heat waves (Li et al., 2011).

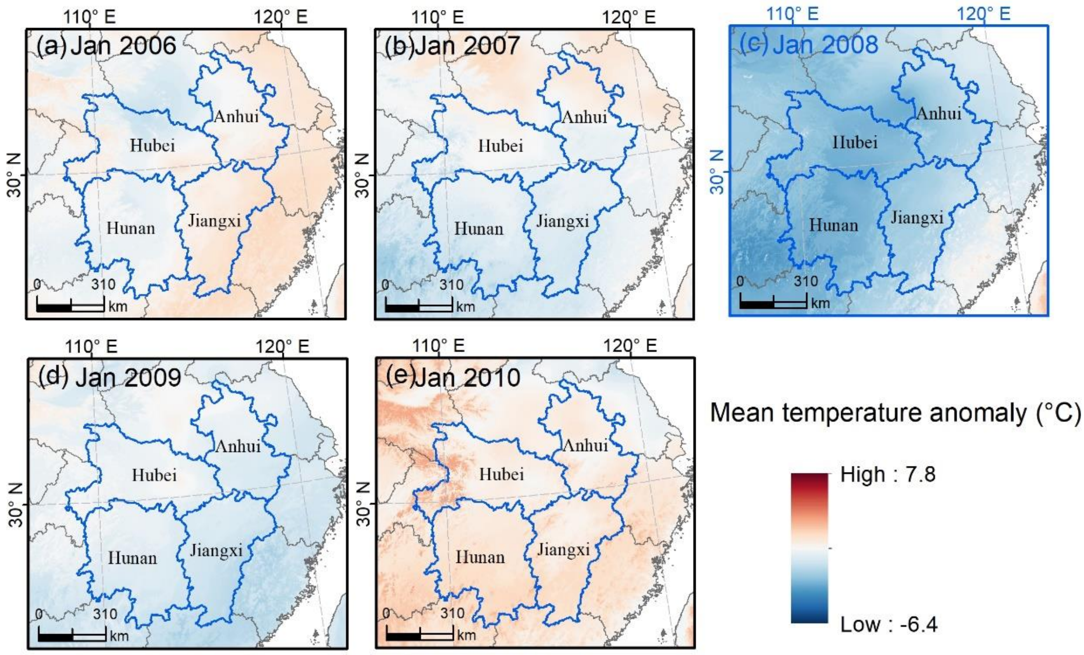

Figure S61: Comparison between the year with lower mean temperature anomaly (for Tmean) and the adjacent years. Hubei, Anhui, Hunan and Jiangxi and neighbouring regions were affected by cold temperatures in the 2008 Chinese winter storms (Liu et al., 2016; Zhou et al., 2014; Lu et al., 2010).

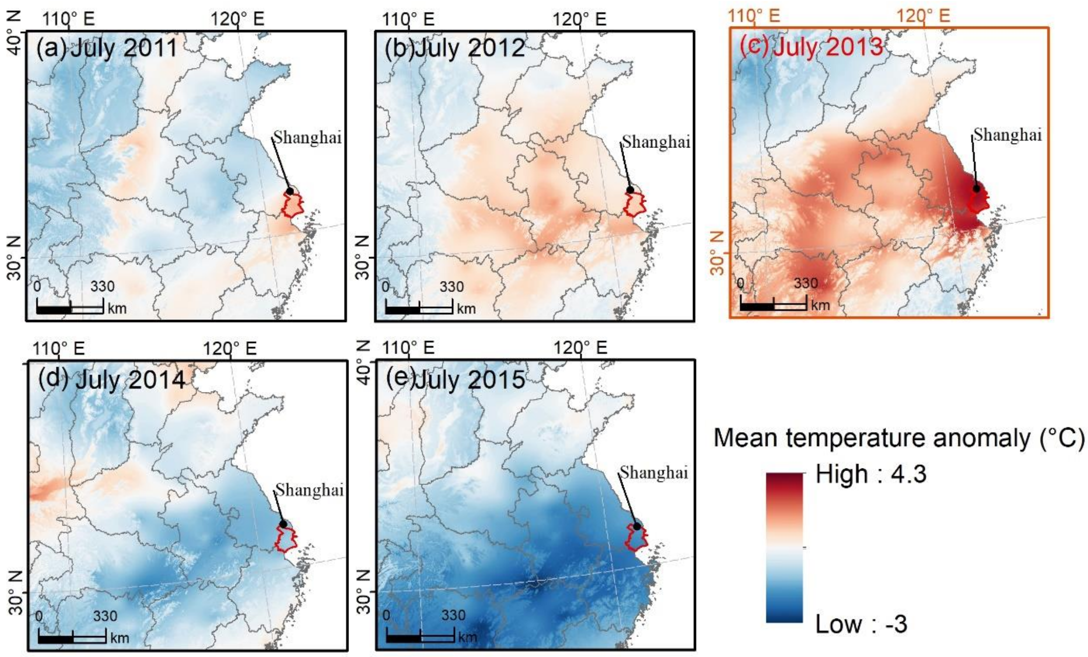

Figure S62: Comparison between the year with higher mean temperature anomaly (for Tmean) and the adjacent years. Heat waves hit Shanghai and neighbouring regions in July 2013 (Pu et al., 2017; Ding and Ke, 2015; Li et al., 2015; Jing-Bei, 2014).

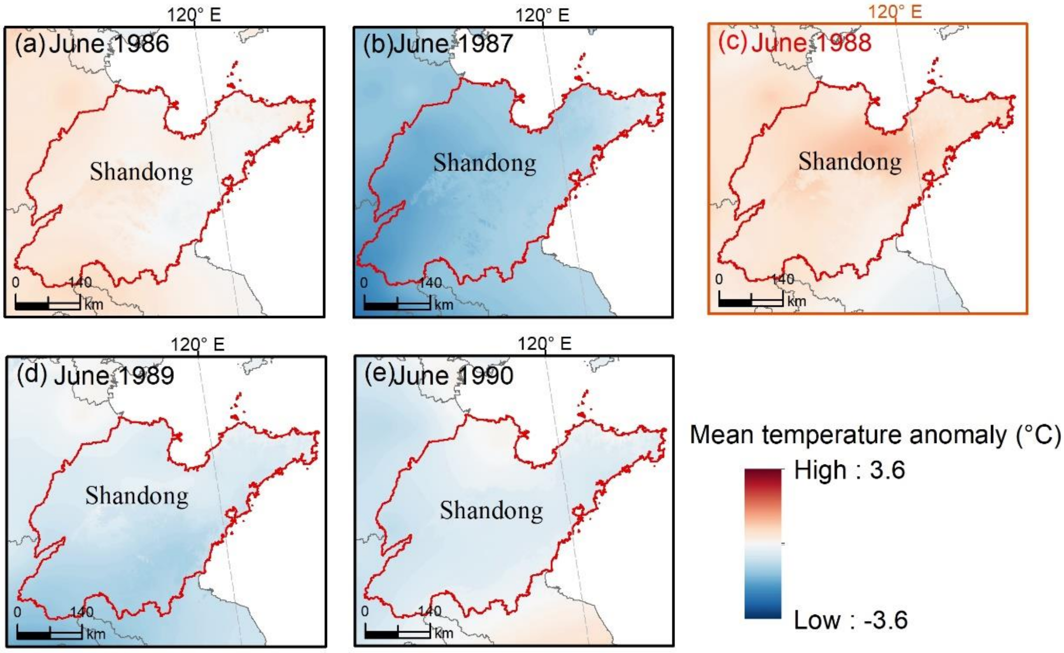

Figure S63: Comparison between the year with extremely high temperature and the adjacent years in Shandong. A drought event with heat wave happened in 1988 in Shandong province and neighbouring regions.

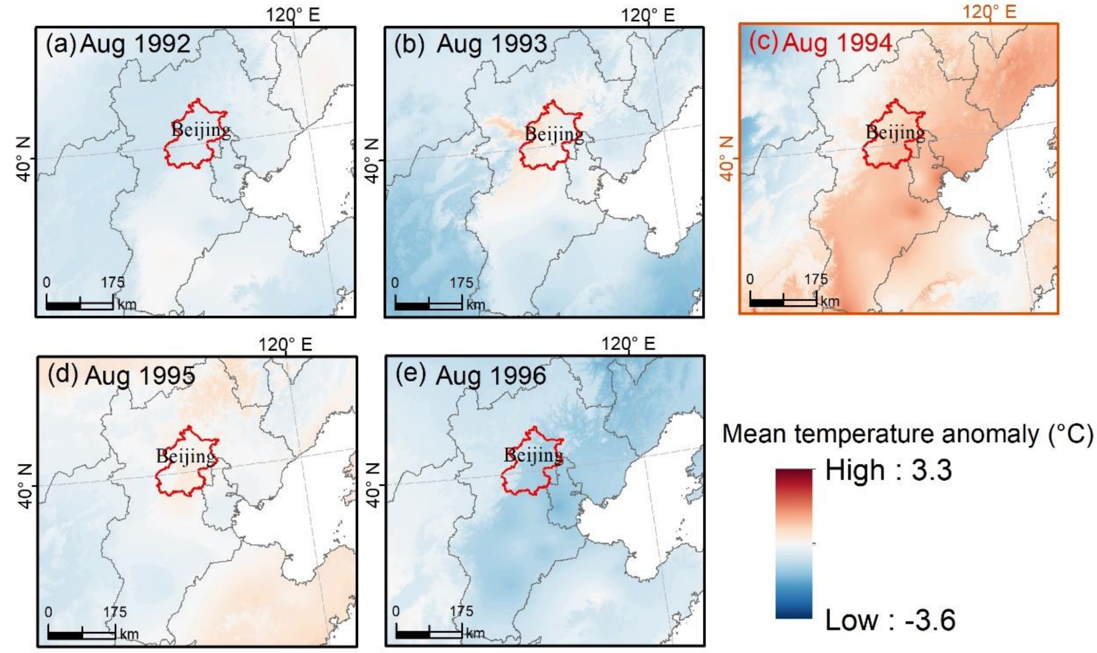

Figure S64: Comparison between the year with higher mean temperature anomaly (for Tmean) and the adjacent years. Heat waves happened in Beijing and the surrounding areas in August 1994 (Zhang et al., 2018; Park et al., 2012).

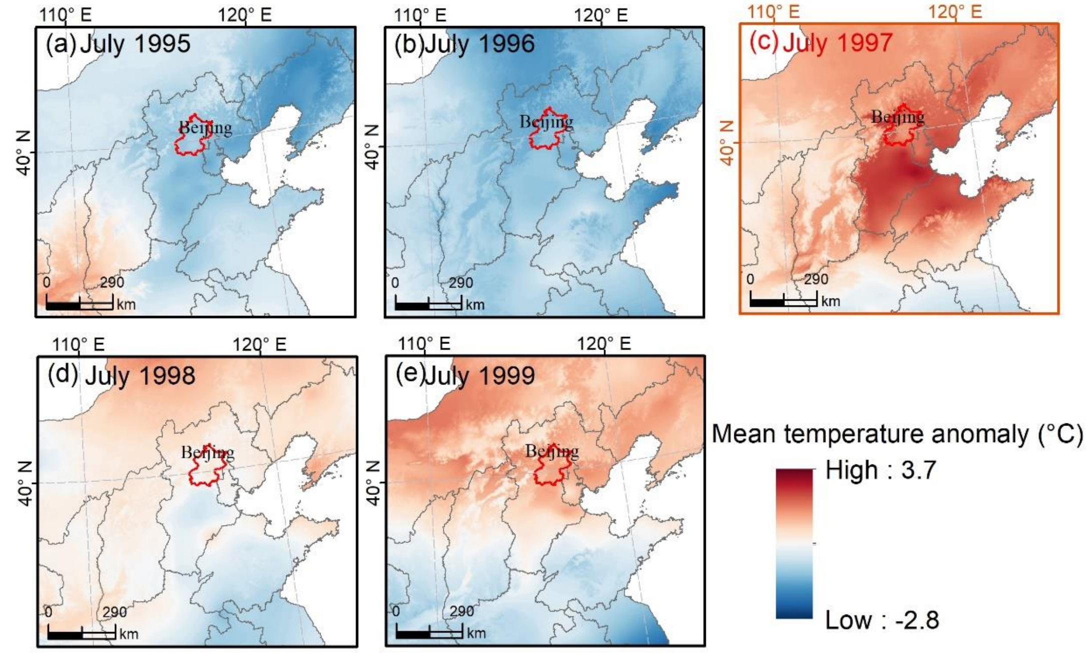

Figure S65: Comparison between the year with higher mean temperature anomaly (for Tmean) and the adjacent years. Heat waves hit Beijing and the neighbouring regions in July 1997 (Park et al., 2012; Zhang et al., 2018).

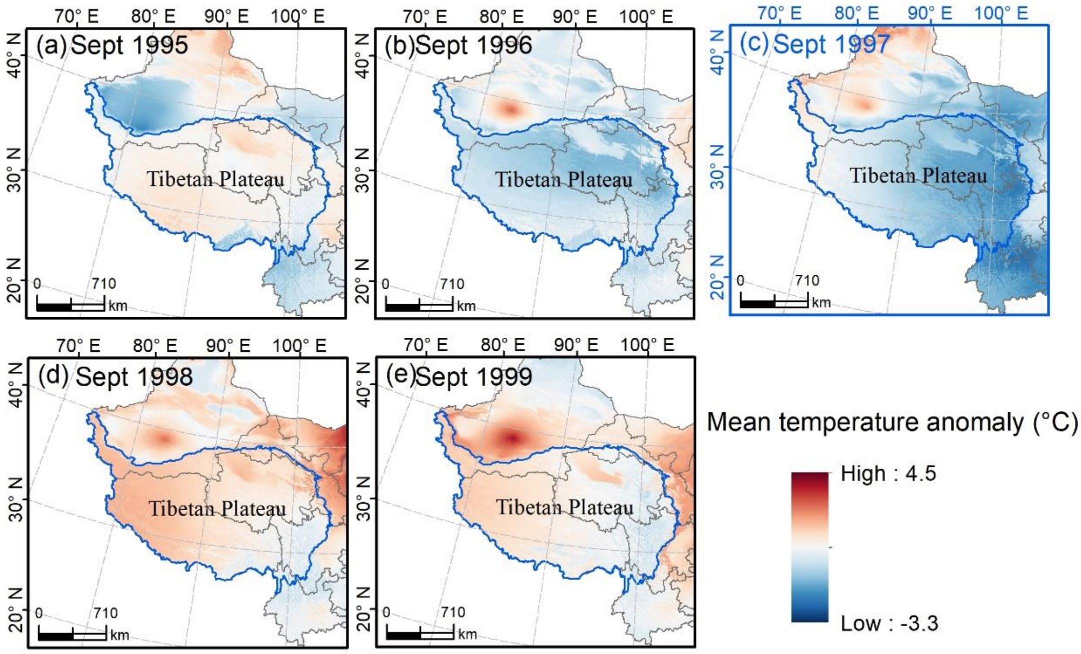

Figure S66: Comparison between the year with lower mean temperature anomaly (for Tmean) and the adjacent years. Cold waves hit the Tibetan Plateau and neighbouring regions in September 1997.

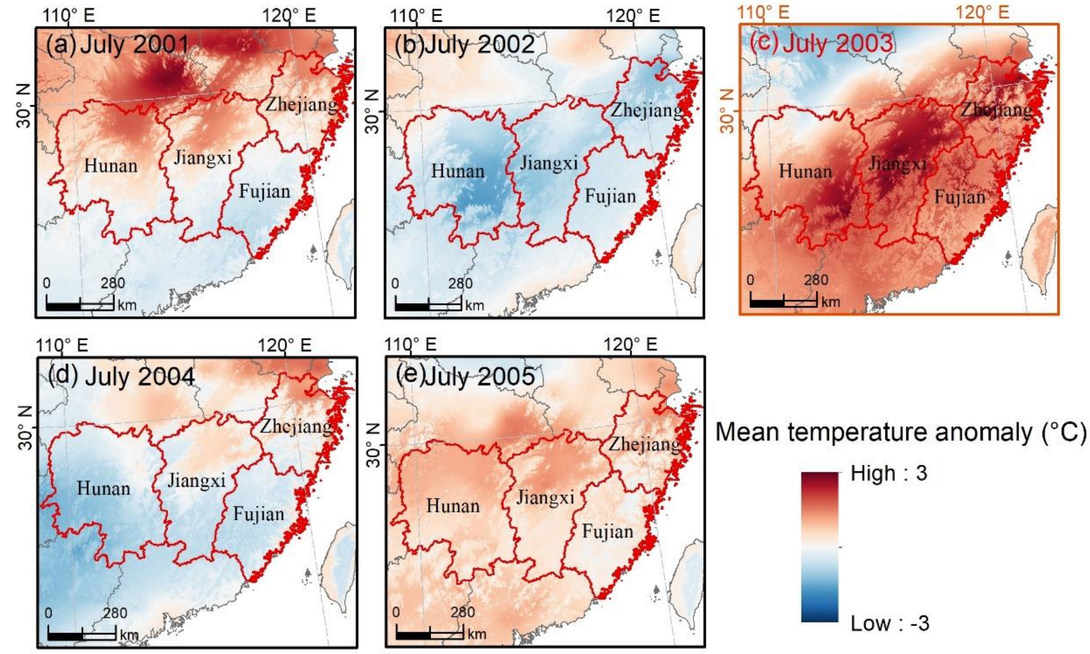

Figure S67: Comparison between the year with higher mean temperature anomaly (for Tmean) and the adjacent years. Drought happened in Hunan, Jiangxi, Zhejiang and Fujian provinces and neighbouring regions in July 2003 which is caused by the lack of precipitation and heat wave (Wang and Yan, 2021; Zhang et al., 2017; Ding and Ke, 2015). In 2001, the summer drought with high temperature happened in the southern China as well, such as Hubei, Zhejiang (Pandey et al., 2007, p.36–37). There were also heatwaves in the southern cities in 2005, for example Guangdong province (Yang et al., 2013).

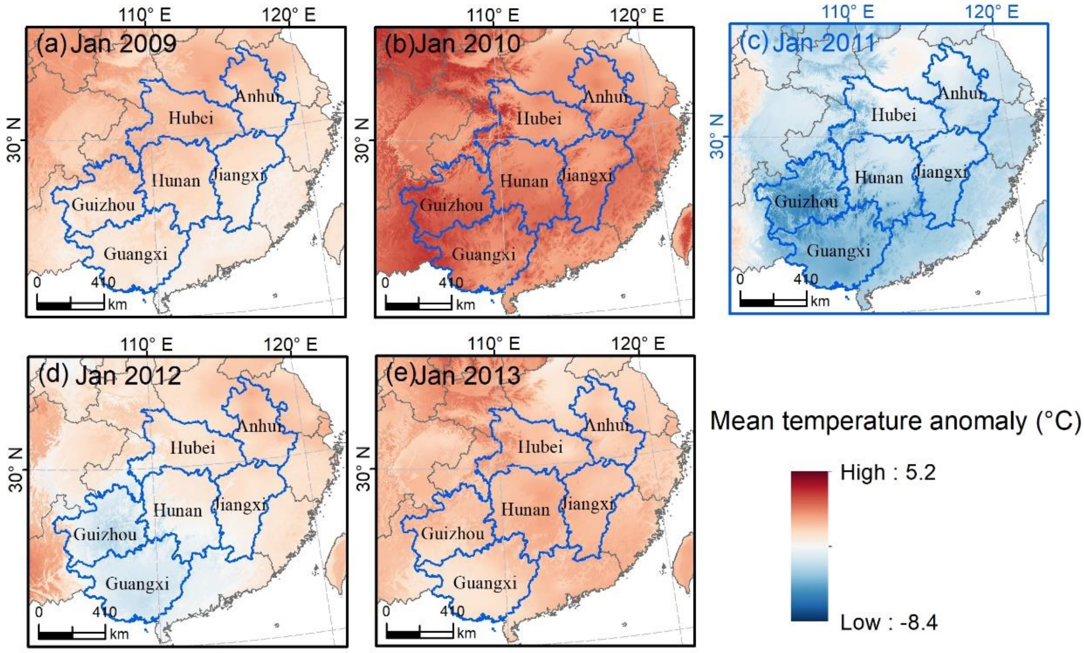

Figure S68: Comparison between the year with lower mean temperature anomaly (for Tmean) and the adjacent years. Guizhou, Guangxi, Hubei, Hunan, Anhui, and Jiangxi as well as the neighbouring regions were affected by cold waves in January 2011 (Qi et al., 2017; http://www.gov.cn/jrzg/2011-01/05/content_1778886.htm, last access: 18 May 2022).

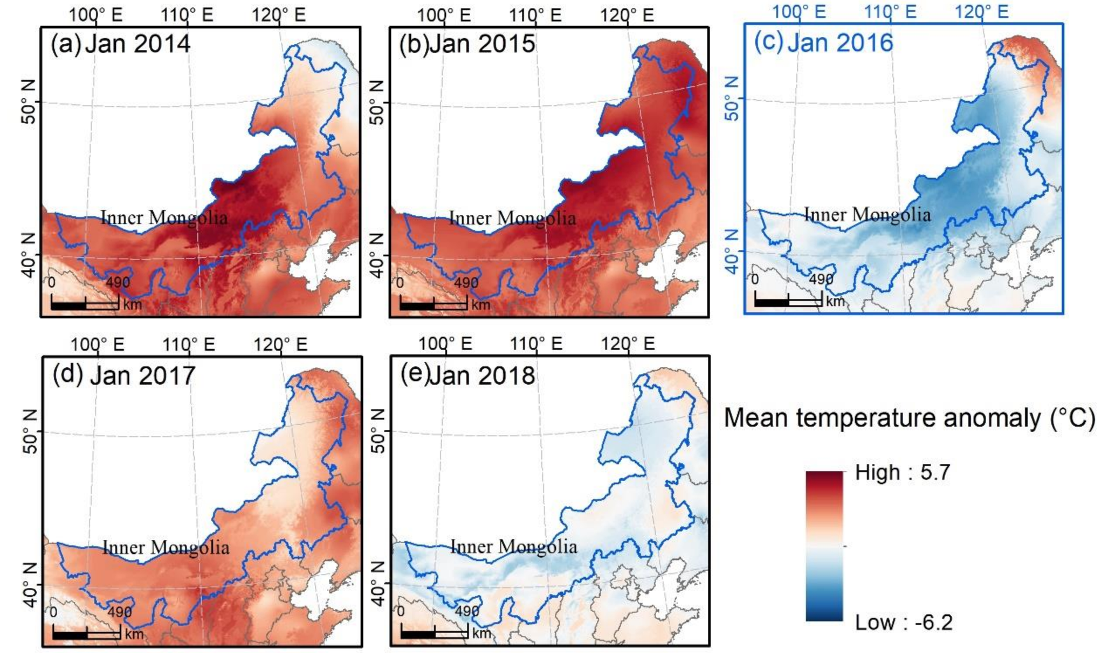

Figure S69: Comparison between the year with lower mean temperature anomaly (for Tmean) and the adjacent years. A cold wave happened in Inner Mongolia and neighbouring regions in January 2016 (Ma and Zhu, 2019; Jiang et al., 2018).

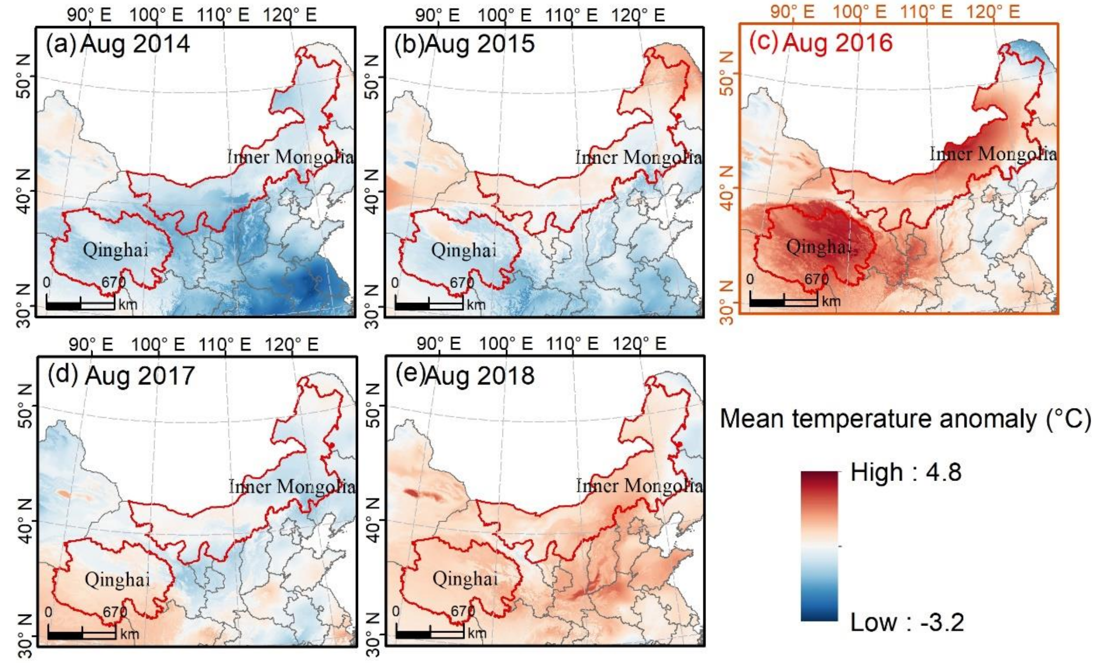

Figure S70: Comparison between the year with higher mean temperature anomaly (for Tmean) and the adjacent years. Extremely high temperatures occurred in Inner Mongolia and Qinghai and other neighbouring regions in August 2016 (http://www.cma.gov.cn/2011xwzx/2011xqxxw/2011xqxyw/201609/t20160902_320919.html, last access: 18 May 2022**).**

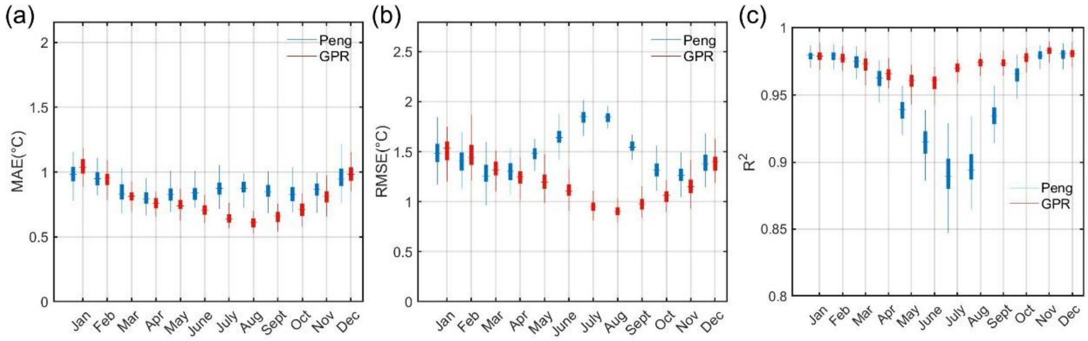

Figure S71: Accuracy comparison between the GPR data and the Peng's data for mean temperature.

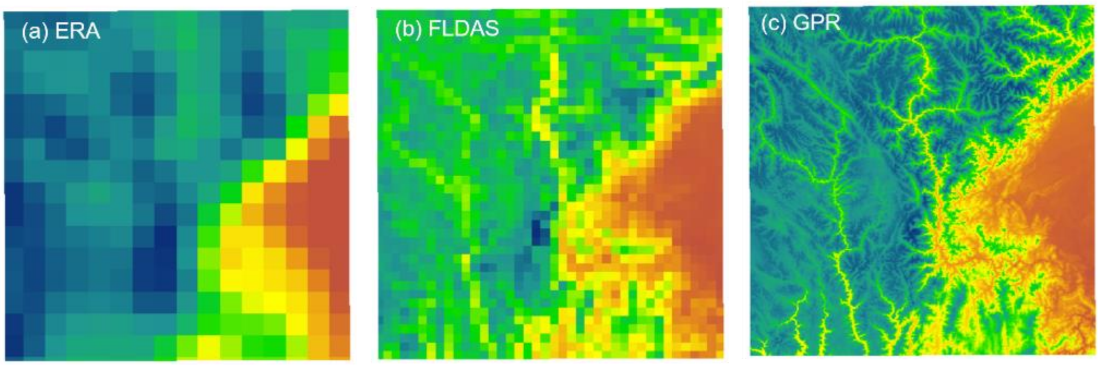

Figure S72: Comparison between the ERA, FLDAS and GPR datasets using the mean air temperature in July 2010. We displayed the mean temperature of July, 2010 in the same region using ERA, FLADS datasets and the GPR dataset generated in our study. The GPR data can provide more spatial details than ERA and FLDAS.

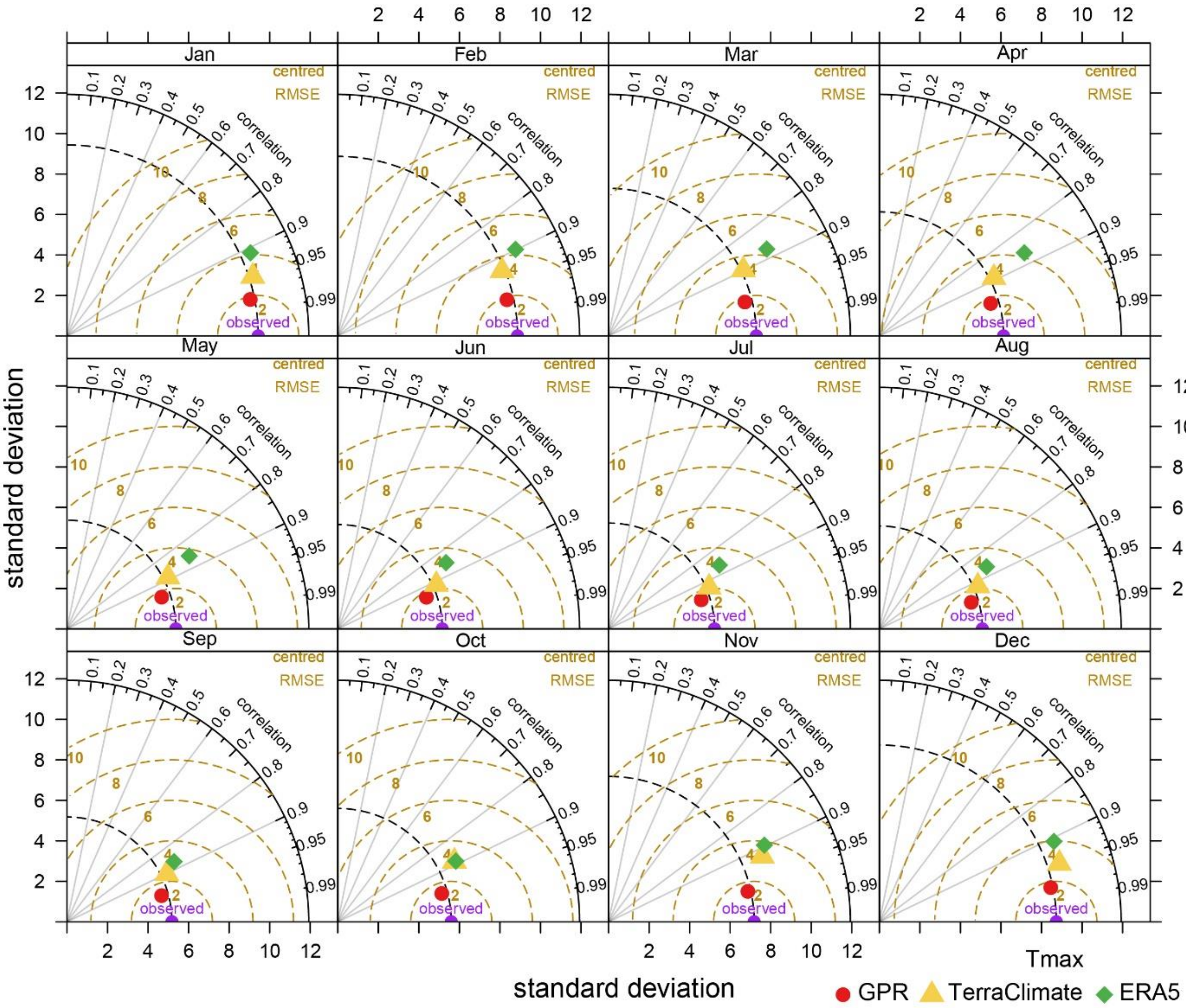

Figure S73: Taylor diagram displaying a statistical comparison with observations for Tmax between our products and other products for each month. The available time periods for the TerraClimate and ERA5 products are: 1958-01-01 - 2020-12-01, and 1979-01-01 - 2020-06-01, respectively. Considering the overlapping periods, we chose January 1979-December 2019 for comparing Tmax.

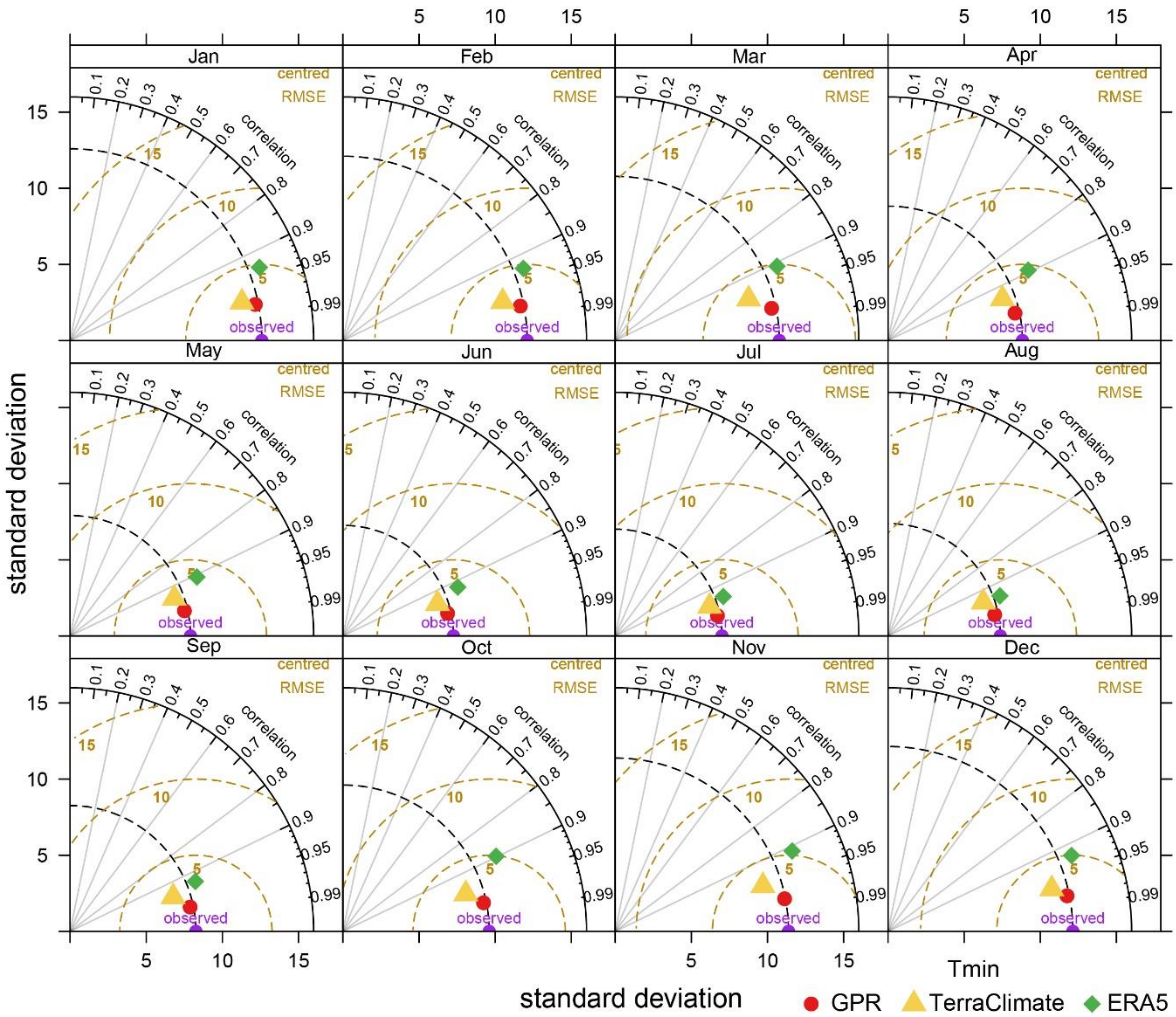

Figure S74: Taylor diagram displaying a statistical comparison with observations for Tmin between our products and other products for each month. The available time periods for the TerraClimate and ERA5 products are: 1958-01-01 - 2020-12-01, and 1979-01-01 - 2020-06-01, respectively. Considering the overlapping periods, we chose January 1979-December 2019 for comparing Tmin.

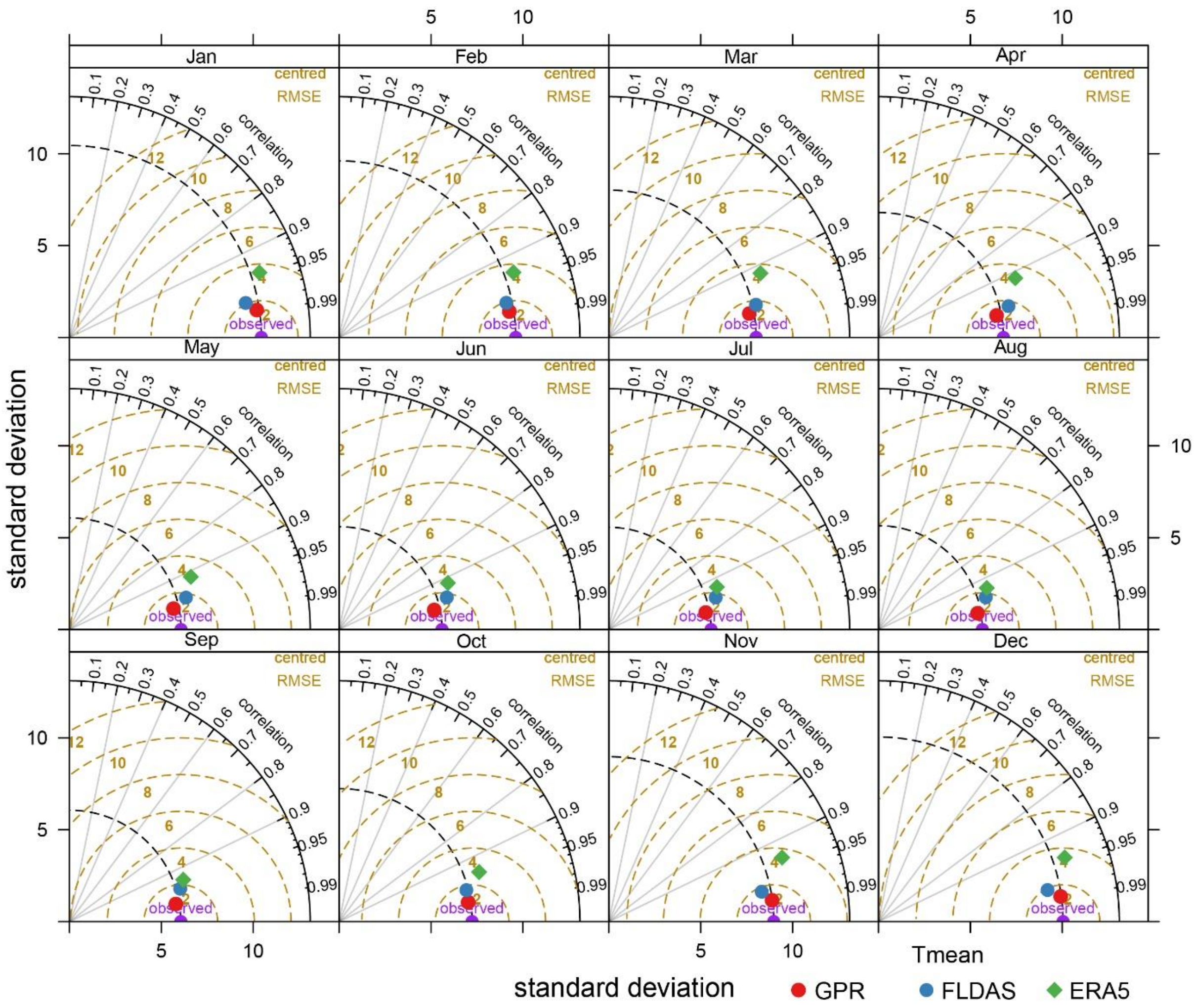

Figure S75: Taylor diagram displaying a statistical comparison with observations for Tmean between our products and other products for each month. The available time periods for FLDAS and ERA5 products are: 1982-01-01 - 2021-05-01, and 1979-01-01 - 2020-06-01, respectively. Considering the overlapping periods, we chose January 1982 to December 2019 for comparing Tmean.

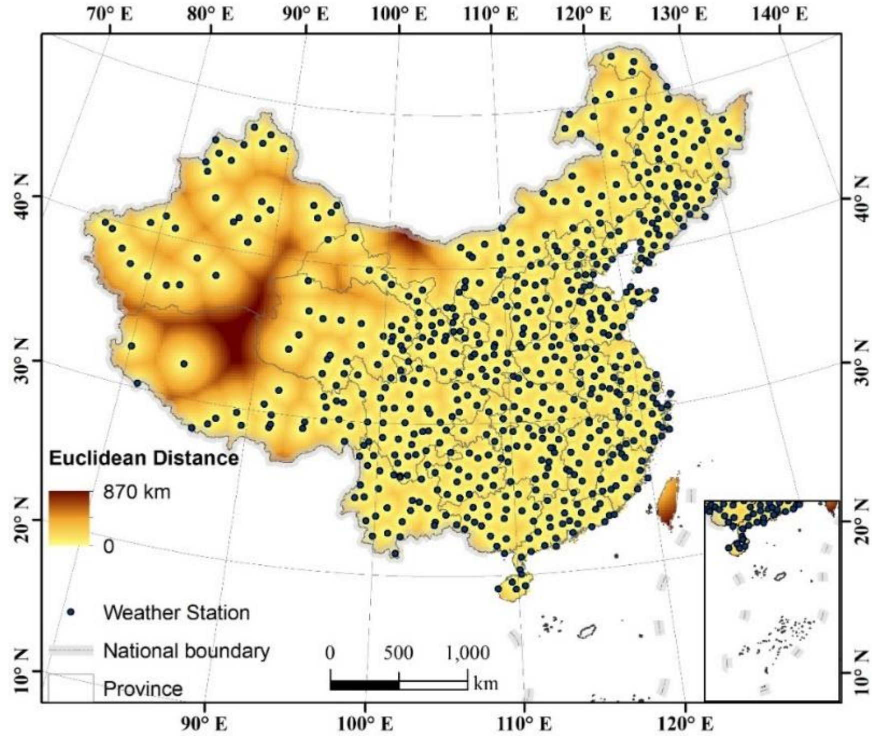

Figure S76: Euclidean distance of the weather stations.

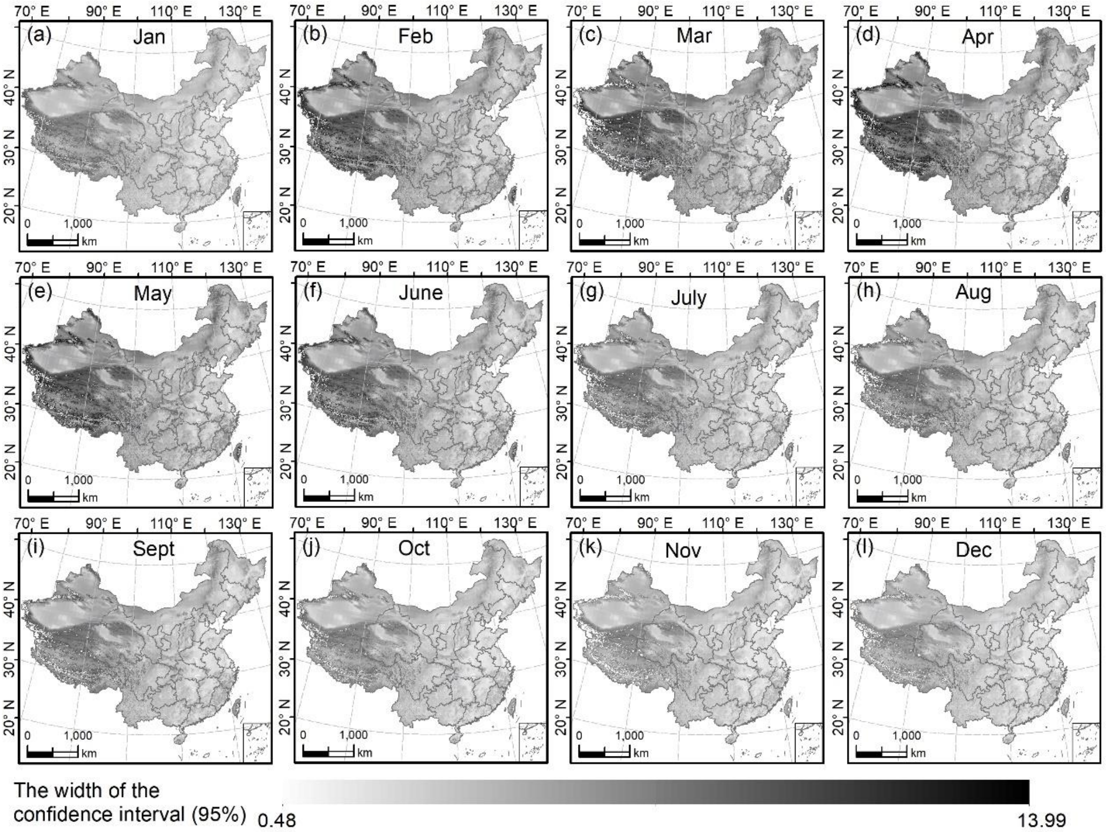

Figure S77: The spatial distribution of the width of the 95% prediction intervals (the upper limit minus the lower limit of the confidence interval) for 12 months in 2010.

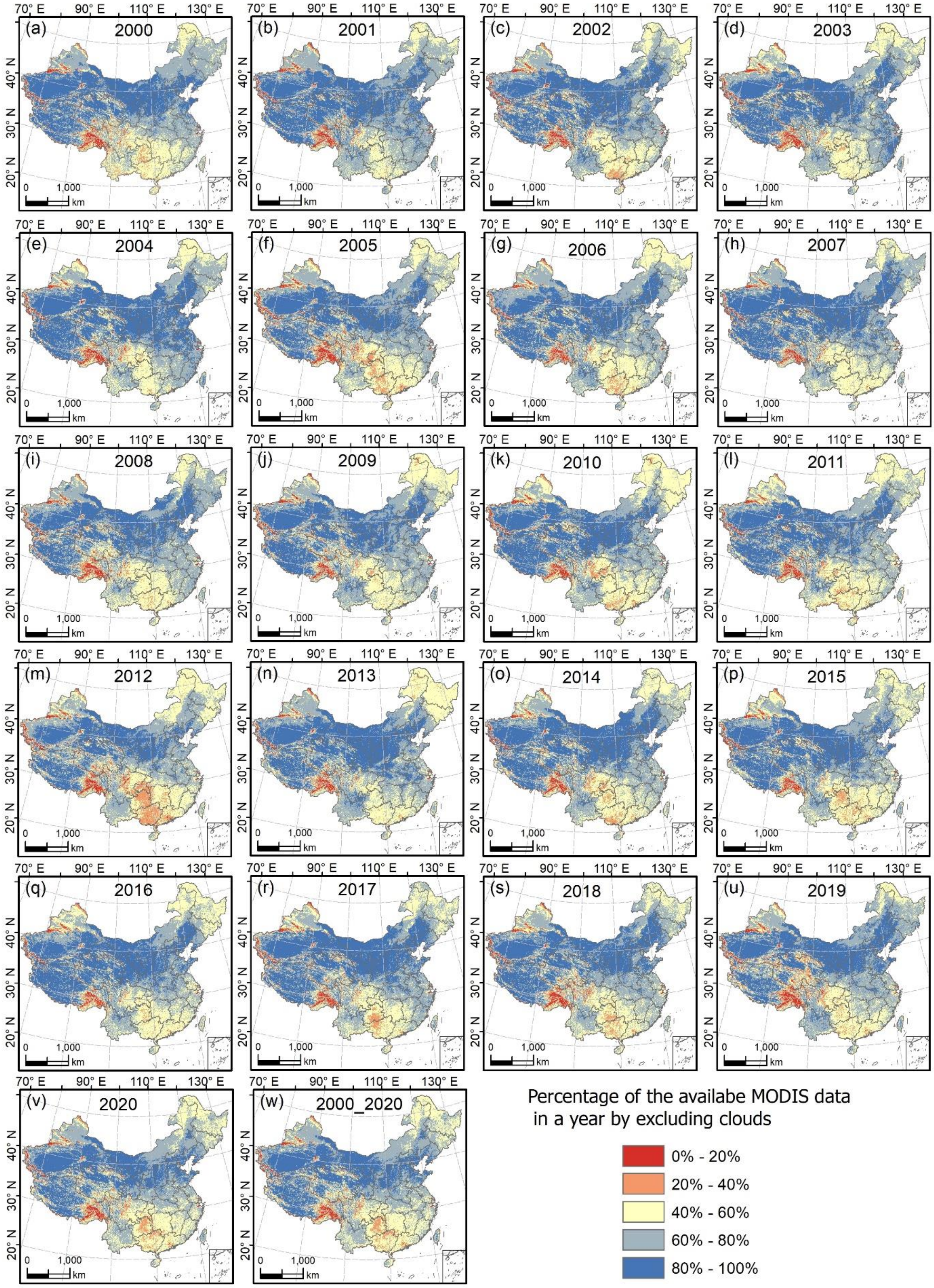

Figure S78: Spatial distribution of the percentage of the available MODIS images in each year (2000 - 2020) by excluding clouds.

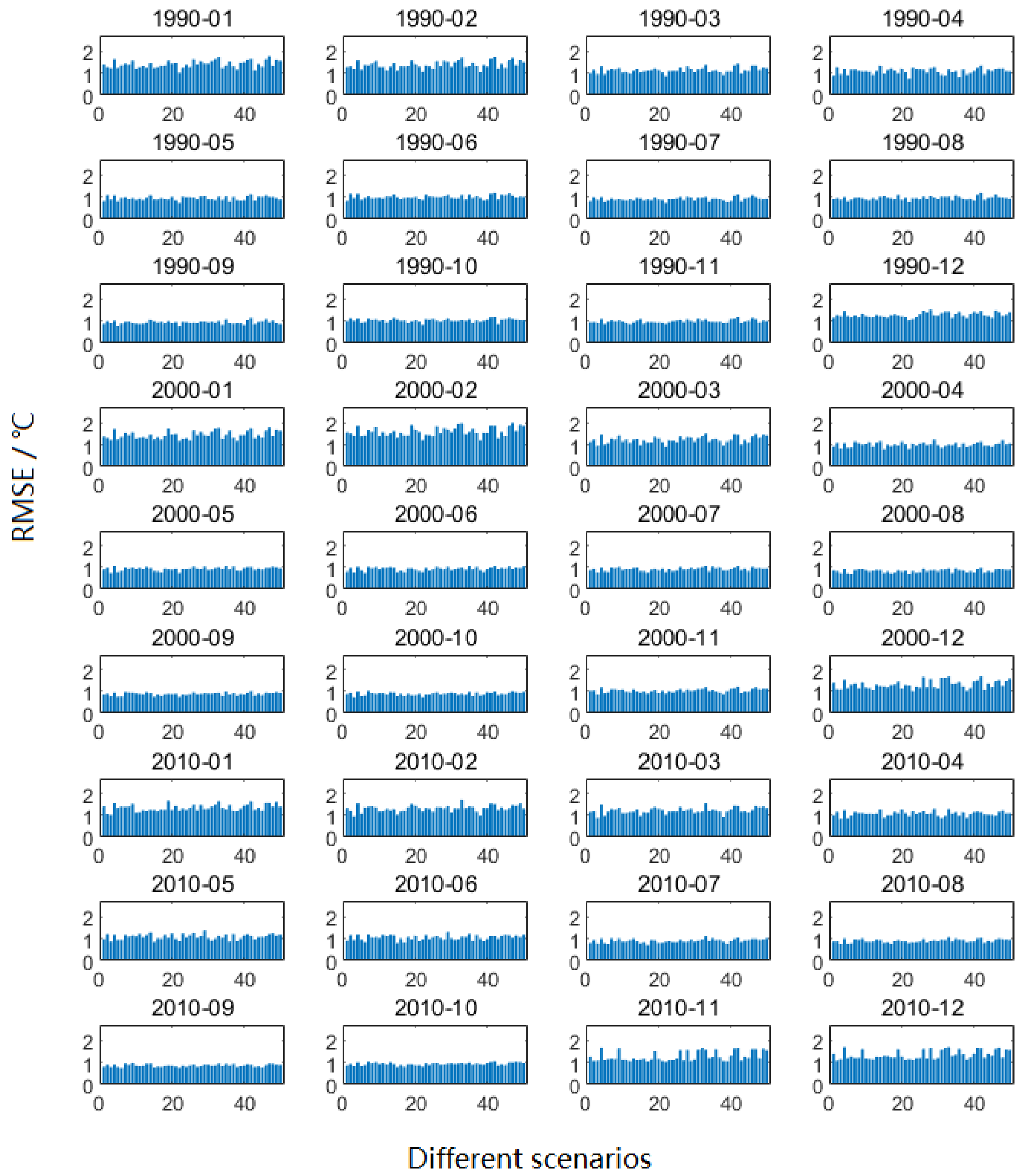

Figure S79: The RMSE using different testing and training datasets. We split the stations into testing and training stations in ArcGIS, which has considered the spatial distribution of the weather stations. We conducted a case study using the Tmean from 1990, 2000 and 2010 to figure out if the model output is sensitive to the choice of stations used in the test/training dataset. We conducted the experiment by randomly splitting the data into training and testing sets (7:3) 50 times using ArcGIS.

Figure S80: The R^2 using different testing and training datasets. We split the stations into testing and training stations in ArcGIS, which has considered the spatial distribution of the weather stations. We conducted a case study using the Tmean from 1990, 2000 and 2010 to figure out if the model output is sensitive to the choice of stations used in the test/training dataset. We conducted the experiment by randomly splitting the data into training and testing sets (7:3) 50 times using ArcGIS.

References

Ding, T. and Ke, Z.: Characteristics and changes of regional wet and dry heat wave events in China during 1960–2013, Theor Appl Climatol, 122, 651–665, https://doi.org/10.1007/s00704-014-1322-9, 2015.

Jiang, D., Xiao, W., Wang, J., Wang, H., Zhao, Y., Li, B., and Zhou, P.: Evaluation of the effects of one cold wave on heating energy consumption in different regions of northern China, Energy, 142, 331–338, https://doi.org/10.1016/j.energy.2017.09.150, 2018.

Jing-Bei, P.: An Investigation of the Formation of the Heat Wave in Southern China in Summer 2013 and the Relevant Abnormal Subtropical High Activities, 7, 286–290, https://doi.org/10.3878/j.issn.1674-2834.13.0097, 2014.

Li, J., Ding, T., Jia, X., and Zhao, X.: Analysis on the Extreme Heat Wave over China around Yangtze River Region in the Summer of 2013 and Its Main Contributing Factors, 2015, e706713, https://doi.org/10.1155/2015/706713, 2015.

Li, Y., Xu, H., and Liu, D.: Features of the extremely severe drought in the east of Southwest China and anomalies of atmospheric circulation in summer 2006, Acta Meteorol Sin, 25, 176–187, https://doi.org/10.1007/s13351-011-0025-8, 2011.

Liu, Y., Wang, M., Wang, W., Fu, H., and Lu, C.: Chilling damage to mangrove mollusk species by the 2008 cold event in Southern China, 7, e01312, https://doi.org/10.1002/ecs2.1312, 2016.

Lu, Q., Zhang, W., Zhang, P., Wu, X., Zhang, F., Liu, Z., and Dale, M. B.: Monitoring the 2008 cold surge and frozen disasters snowstorm in South China based on regional ATOVS data assimilation, Sci. China Earth Sci., 53, 1216–1228, https://doi.org/10.1007/s11430-010-3040-1, 2010.

Ma, S. and Zhu, C.: Extreme Cold Wave over East Asia in January 2016: A Possible Response to the Larger Internal Atmospheric Variability Induced by Arctic Warming, 32, 1203–1216, https://doi.org/10.1175/JCLI-D-18-0234.1, 2019.

Pandey, S., Bhandari, H. S., and Hardy, B.: Economic Costs of Drought and Rice Farmers' Coping Mechanisms: A Cross-country Comparative Analysis, Int. Rice Res. Inst., 36–37 pp., 2007.

Park, J.-K., Lu, R., Li, C., and Kim, E. B.: Interannual variation of tropical night frequency in Beijing and associated large-scale circulation background, Adv. Atmos. Sci., 29, 295–306, https://doi.org/10.1007/s00376-011-1141-1, 2012.

Pu, X., Wang, T. J., Huang, X., Melas, D., Zanis, P., Papanastasiou, D. K., and Poupkou, A.: Enhanced surface ozone during the heat wave of 2013 in Yangtze River Delta region, China, Science of The Total Environment, 603–604, 807–816, https://doi.org/10.1016/j.scitotenv.2017.03.056, 2017.

Qi L., Ma Q., and Zhang W.: Verification of forecasting capability of cold wave process in the winter of 2011 /2012 with GRAPES, 40, 791–802, https://doi.org/10.13878/j.cnki.dqkxxb.20150104001, 2017. (in Chinese)

Wang, J. and Yan, Z.: Rapid rises in the magnitude and risk of extreme regional heat wave events in China, Weather and Climate Extremes, 34, 100379, https://doi.org/10.1016/j.wace.2021.100379, 2021.

Yang, J., Liu, H. Z., Ou, C. Q., Lin, G. Z., Ding, Y., Zhou, Q., Shen, J. C., and Chen, P. Y.: Impact of Heat Wave in 2005 on Mortality in Guangzhou, China, Biomedical and Environmental Sciences, 26, 647–654, https://doi.org/10.3967/0895-3988.2013.08.003, 2013.

Zhang, S., Huang, G., Qi, Y., and Jia, G.: Impact of urbanization on summer rainfall in Beijing–Tianjin–Hebei metropolis under different climate backgrounds, Theor Appl Climatol, 133, 1093–1106, https://doi.org/10.1007/s00704-017-2225-3, 2018.

Zhang, Y., You, Q., Chen, C., and Li, X.: Flash droughts in a typical humid and subtropical basin: A case study in the Gan River Basin, China, Journal of Hydrology, 551, 162–176, https://doi.org/10.1016/j.jhydrol.2017.05.044, 2017.

Zhou, M. G., Wang, L. J., Liu, T., Zhang, Y. H., Lin, H. L., Luo, Y., Xiao, J. P., Zeng, W. L., Zhang, Y. W., Wang, X. F., Gu, X., Rutherford, S., Chu, C., and Ma, W. J.: Health impact of the 2008 cold spell on mortality in subtropical China: the climate and health impact national assessment study (CHINAs), Environmental Health, 13, 60, https://doi.org/10.1186/1476-069X-13-60, 2014.