

# Supplementary materials for the article “Machine learning for simulation of the urban heat island dynamics based on large-scale meteorological conditions”

Mikhail Varentsov<sup>1,2,\*</sup>, Mikhail Krinitskiy<sup>3,1</sup> and Victor Stepanenko<sup>1,2,4</sup>

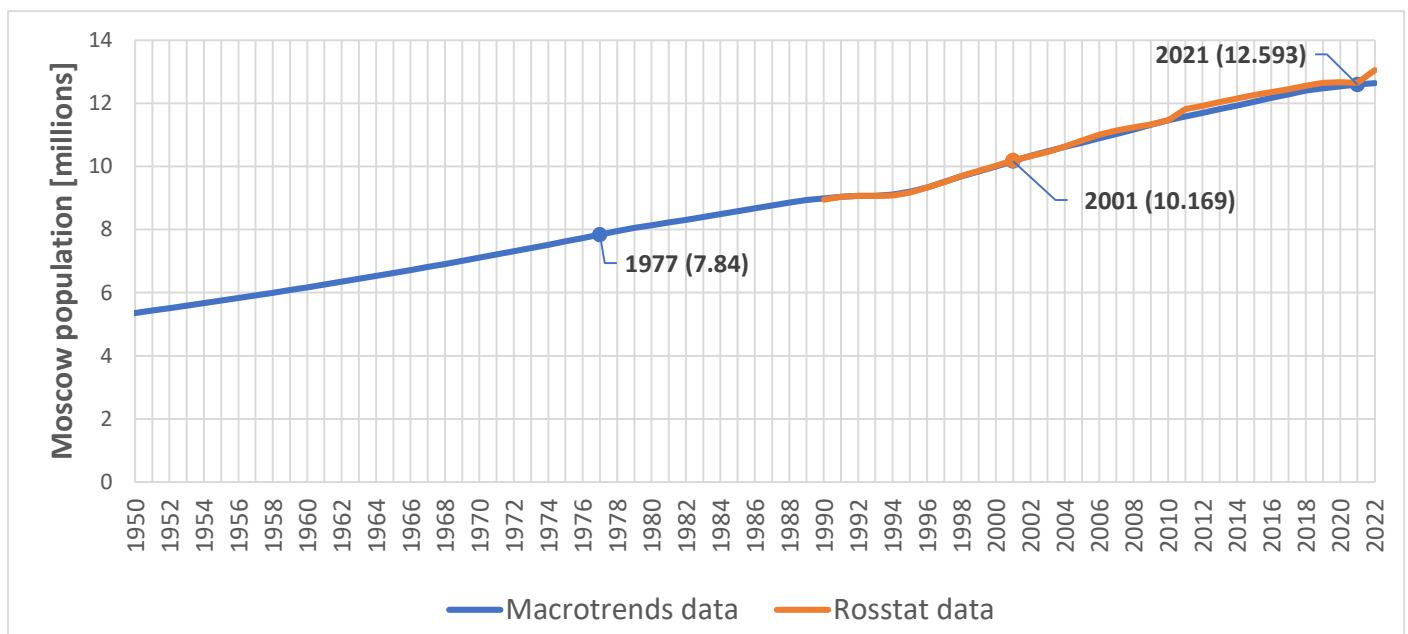
<sup>1</sup> Research Computing Center, Lomonosov Moscow State University, 1/4 Leninskiye Gory, 119991 Moscow, Russia;

<sup>2</sup> Moscow Center for Fundamental and Applied Mathematics, 1 Leninskiye Gory, 119991 Moscow, Russia;

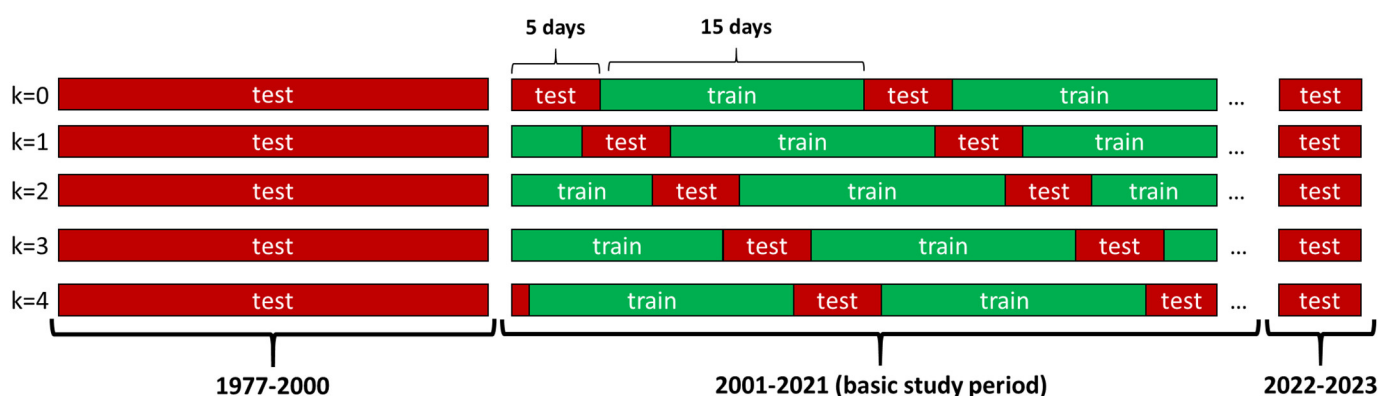
<sup>3</sup> Shirshov Institute of Oceanology, Russian Academy of Sciences, Moscow, Russia

<sup>4</sup> Faculty of Geography, Lomonosov Moscow State University, 1/4 Leninskiye Gory, 119991 Moscow, Russia;

\* Correspondence: mikhail.varentsov@srcc.msu.ru;



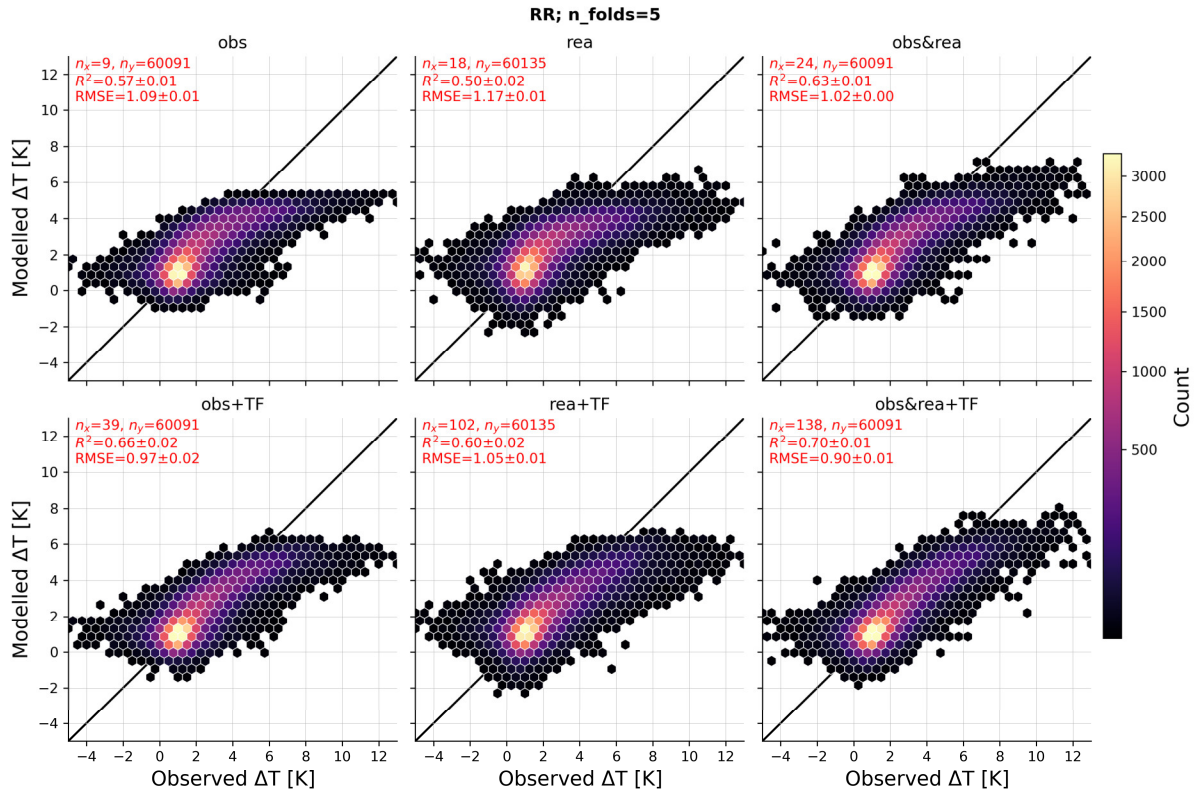
**Figure S1.** Recent changes in the population of the city of Moscow as an administrative unit according to open-access data from Rosstat, the Russian statistical agency (<https://rosstat.gov.ru/folder/12781>), which is available only since 1990, and according to open-access Macrotrends data (<https://www.macrotrends.net/cities/22299/moscow/population>), which is available for longer period.



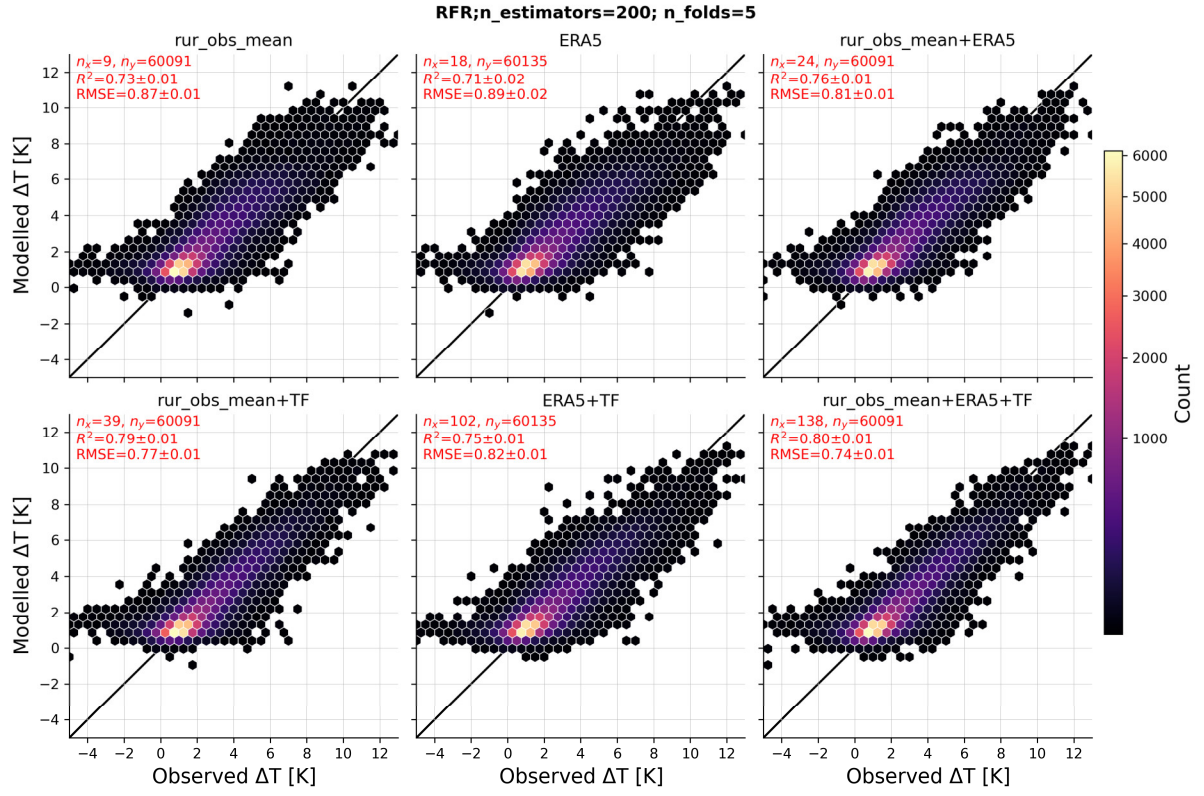
**Figure S2.** Scheme of the blocked k-fold method used to split the dataset into train and test subsets.

**Table S1.** Metrics of comparison between observation-based and reanalysis-based predictors: mean error of reanalysis with respect to observations (ME), root-mean square error (RMSE), ME and RMSE normalized by observed standard deviation (NME and NRMSE), correlation coefficient (R).

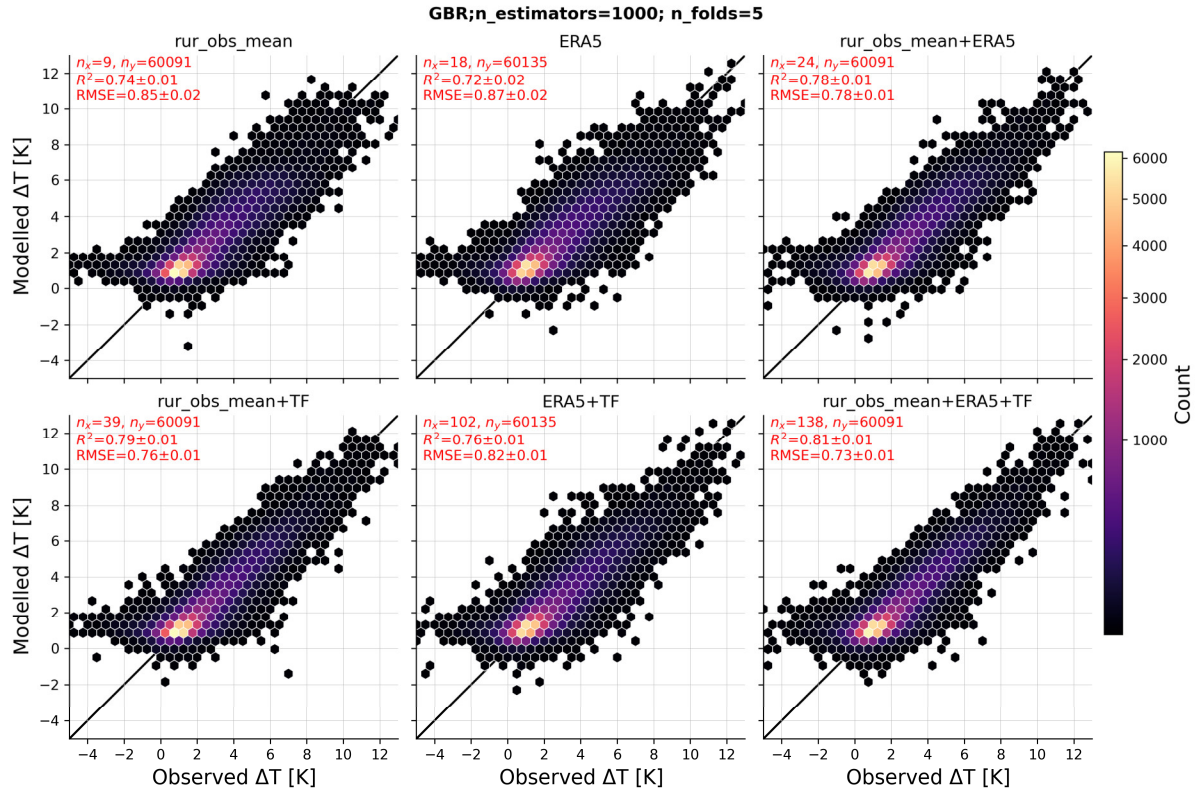
Predictor	Unit	ME	RMSE	NME	NRMSE	R
t2m		-0.28	0.84	-0.03	0.08	1
rh2m		-0.52	4.55	-0.03	0.28	0.96
tcc		0.06	0.25	0.19	0.82	0.72
lcc		0	0.31	0.01	0.86	0.69
vel10m		1.23	1.48	0.92	1.1	0.86



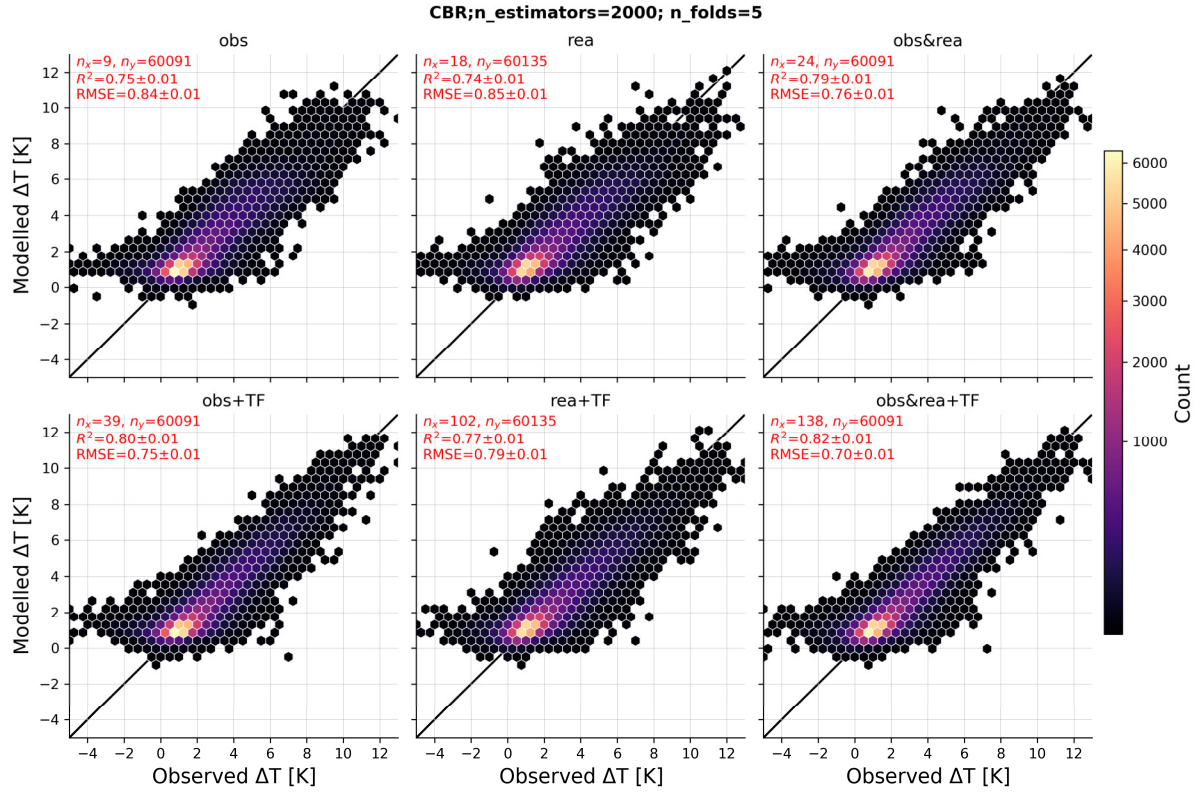
**Figure S3.** Hexbin plots showing comparison between observed and modelled UHI magnitude for the Ridge Regression (RR) baseline model; duplication of the Figure 2.



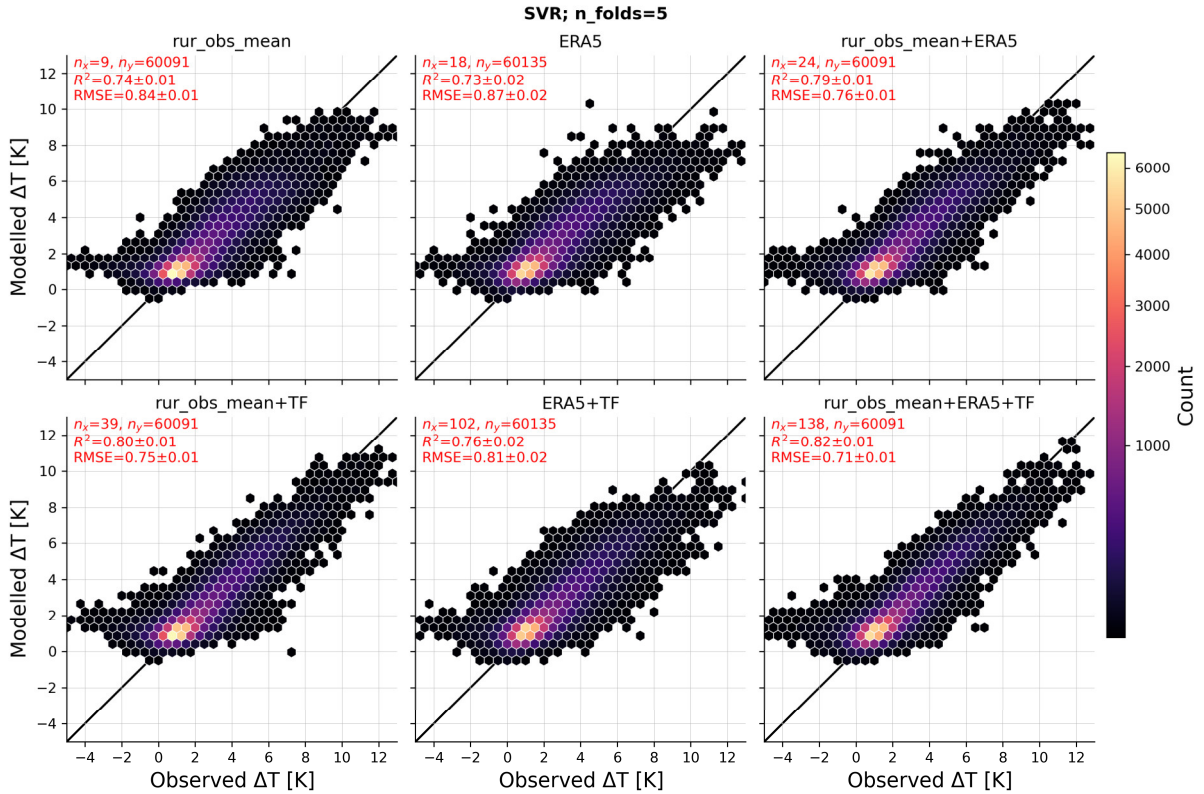
**Figure S4.** Same as Figure S3 for the most complex configuration of Random Forests Regression (RFR) model.



**Figure S5.** Same as Figure S3 for the most complex configuration of Gradient Boosting Regression (GBR) model.

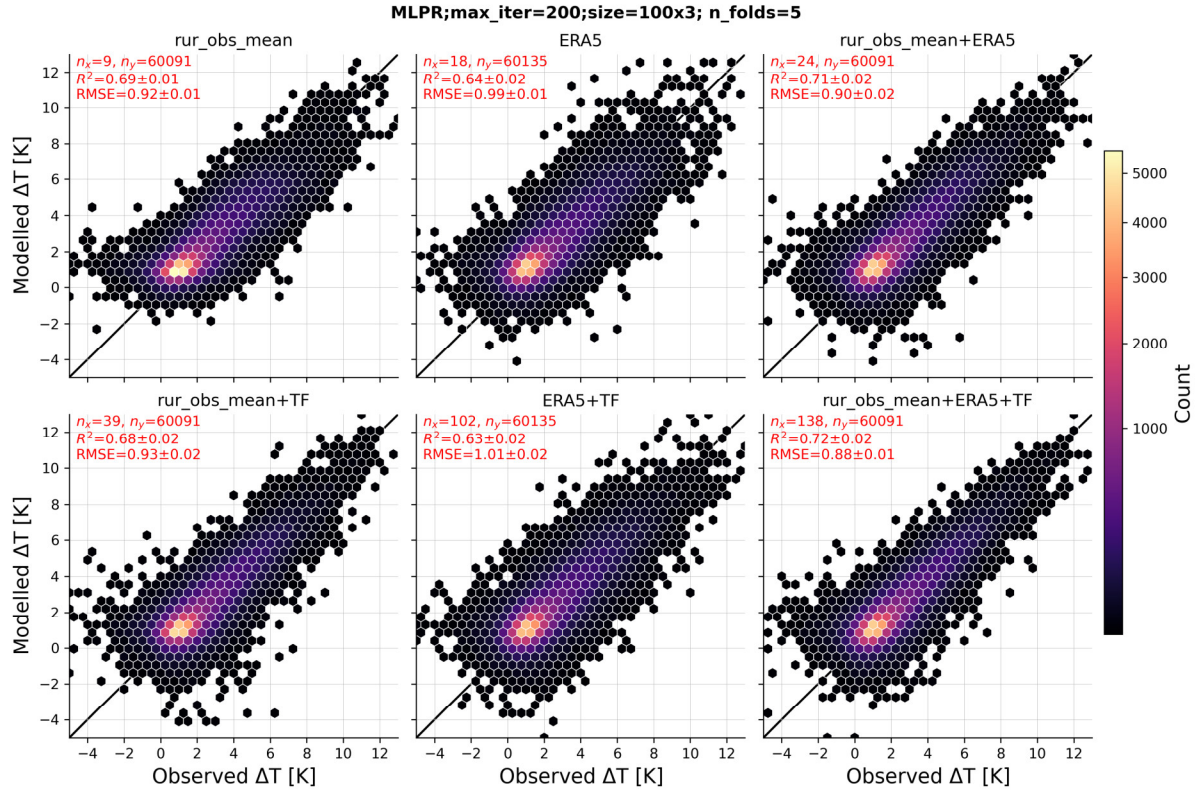


**Figure S6.** Same as Figure S3 for the most complex configuration of CatBoost Regression (CBR) model; duplication of the Figure 3.

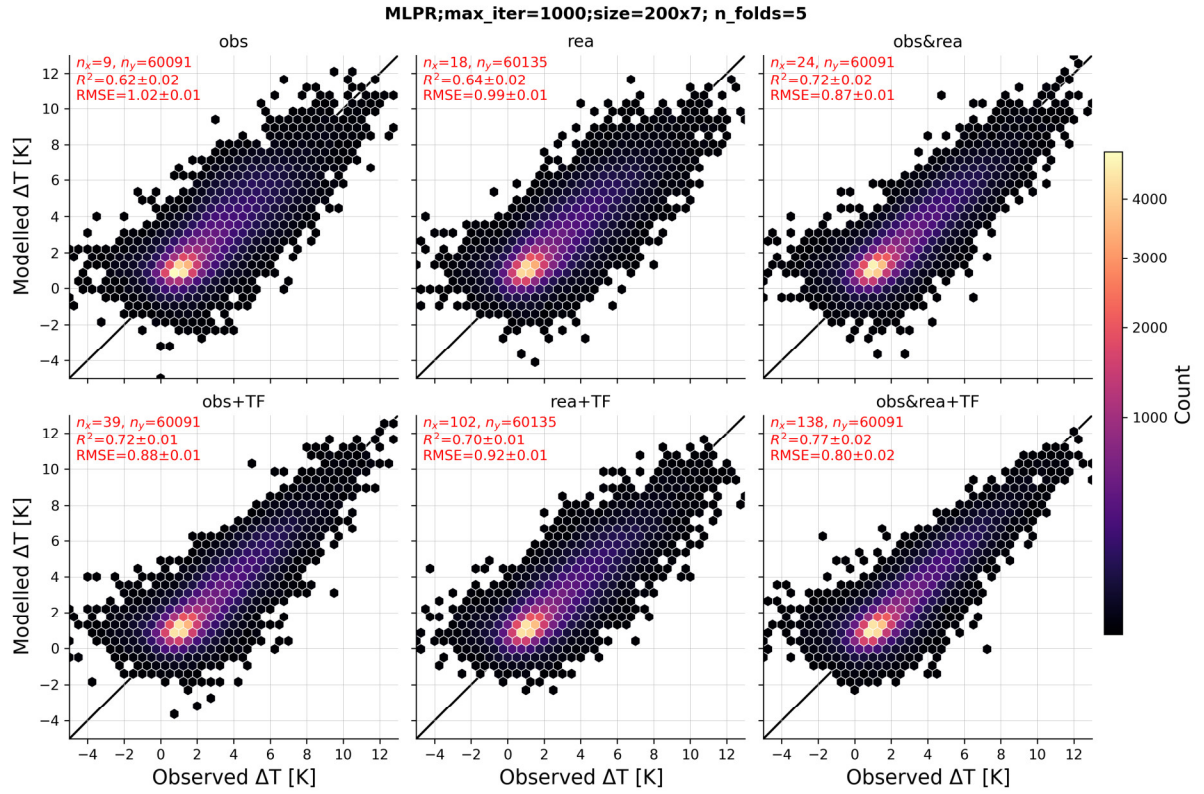


**Figure S7.** Same as Figure S3 for the Support Vector Regression (SVR) model.





**Figure S8.** Same as Figure S3 for the least complex configuration of the Multi-Layer Perceptron Regression (MLPR) model.



**Figure S9.** Same as Figure S3 for the most complex configuration of the Multi-Layer Perceptron Regression (MLPR) model; duplication of the Figure 4.