PIRE: Taming Water in Ethiopia

Lake Level Estimation using Machine Learning and Physically-based Approaches in Lake Tana, Ethiopia

Z. Dokou, N. Reljin, M. Kheirabadi, A. Bagtzoglou, and E. Anagnostou

Civil and Environmental Engineering
Biomedical Engineering

Presentation and Research Outline

Estimation of monthly water levels of Lake Tana, Ethiopia

Physically-based model

January 2001 – December 2012

Data-driven model - SVR

Comparative analysis under various SVR scenarios

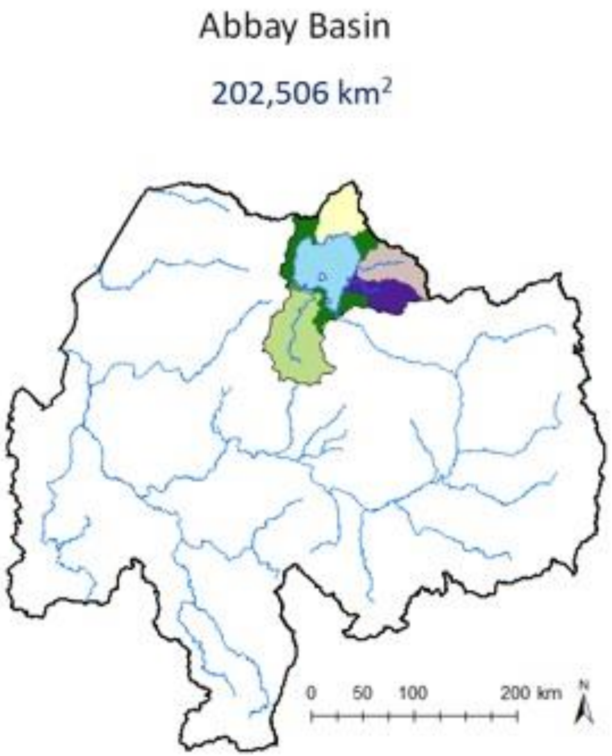
Errors, applicability, ease of use and computational speed

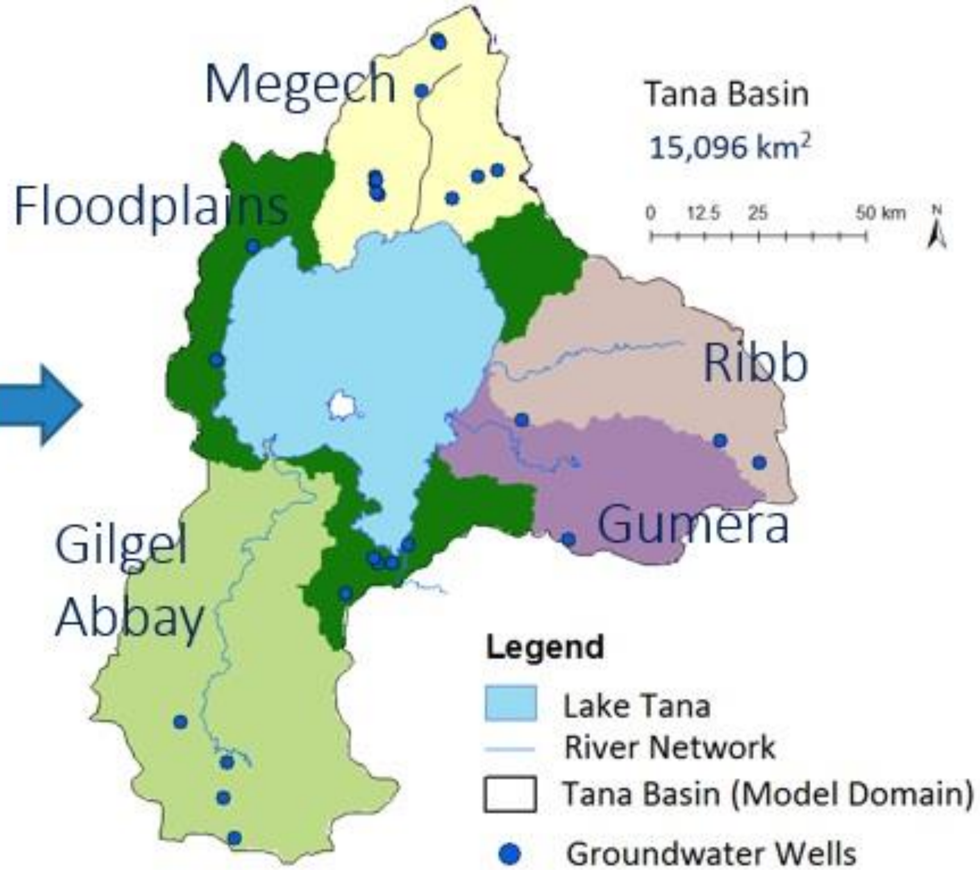
Summary and Conclusions

PIRE

NEUREL 2018

Location of Blue Nile and Lake Tana

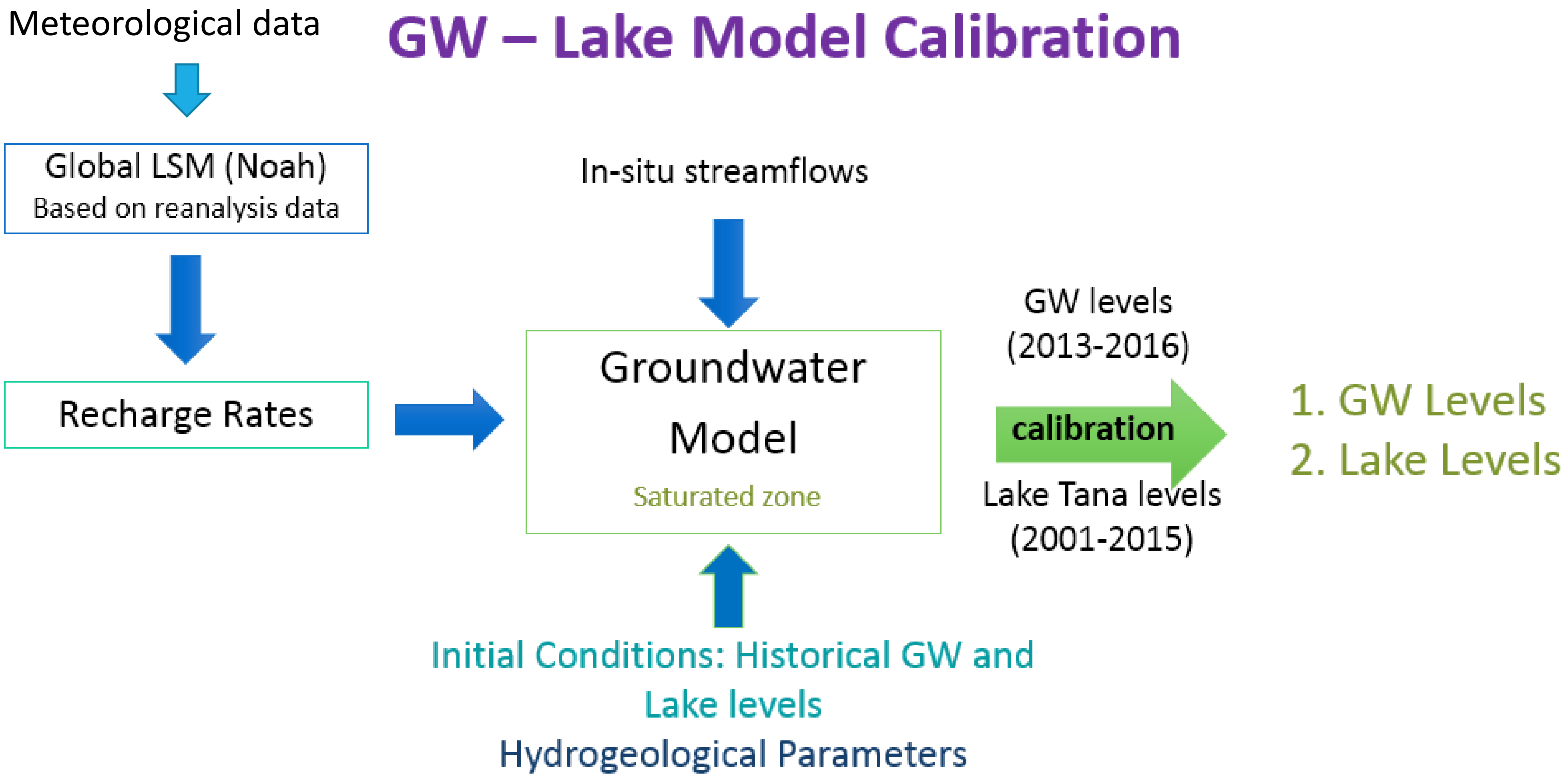




PIRE

NEUREL 2018

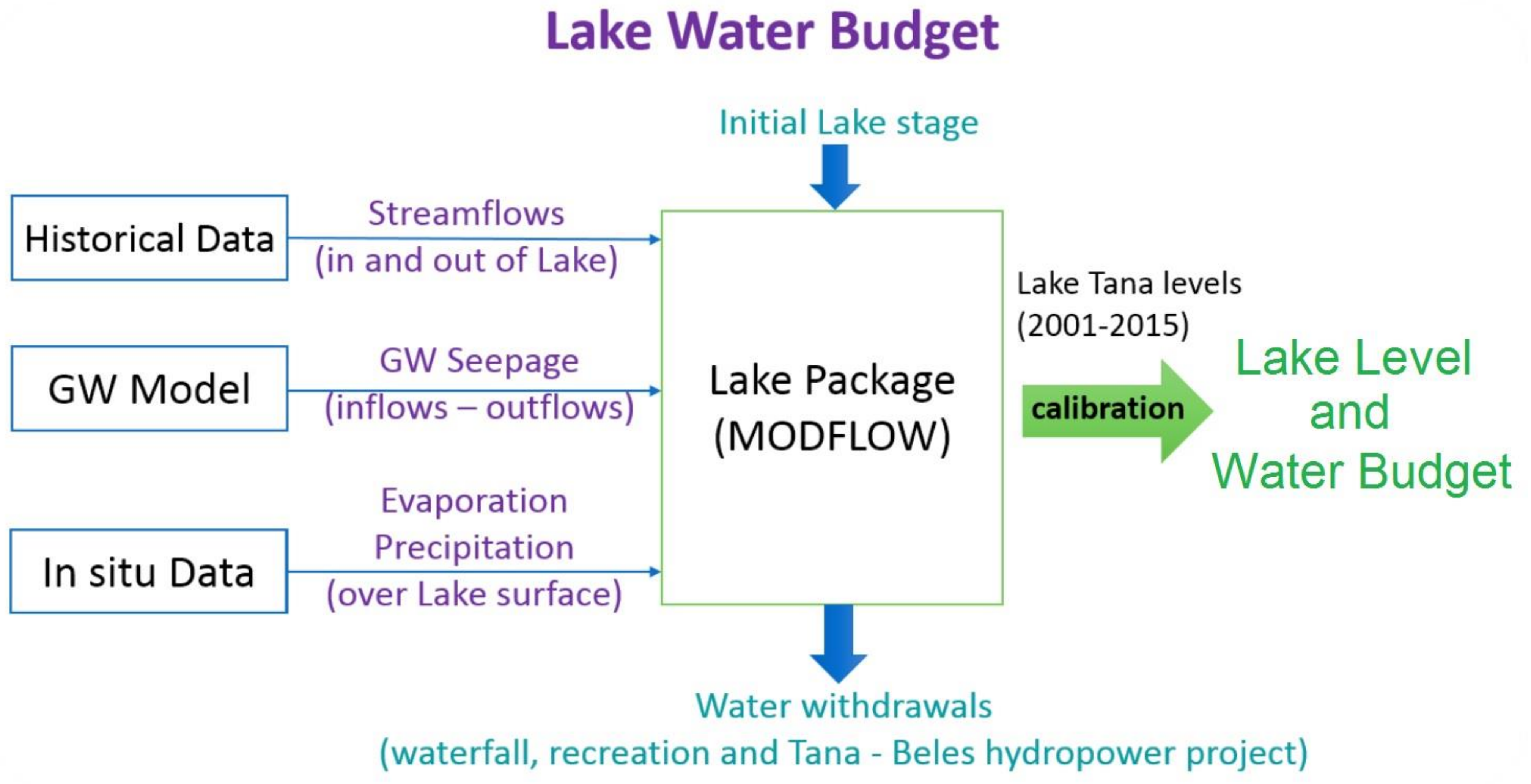
Modeling Framework: Physically-based Model



PIRE

NEUREL 2018

Modeling Framework: Physically-based Model

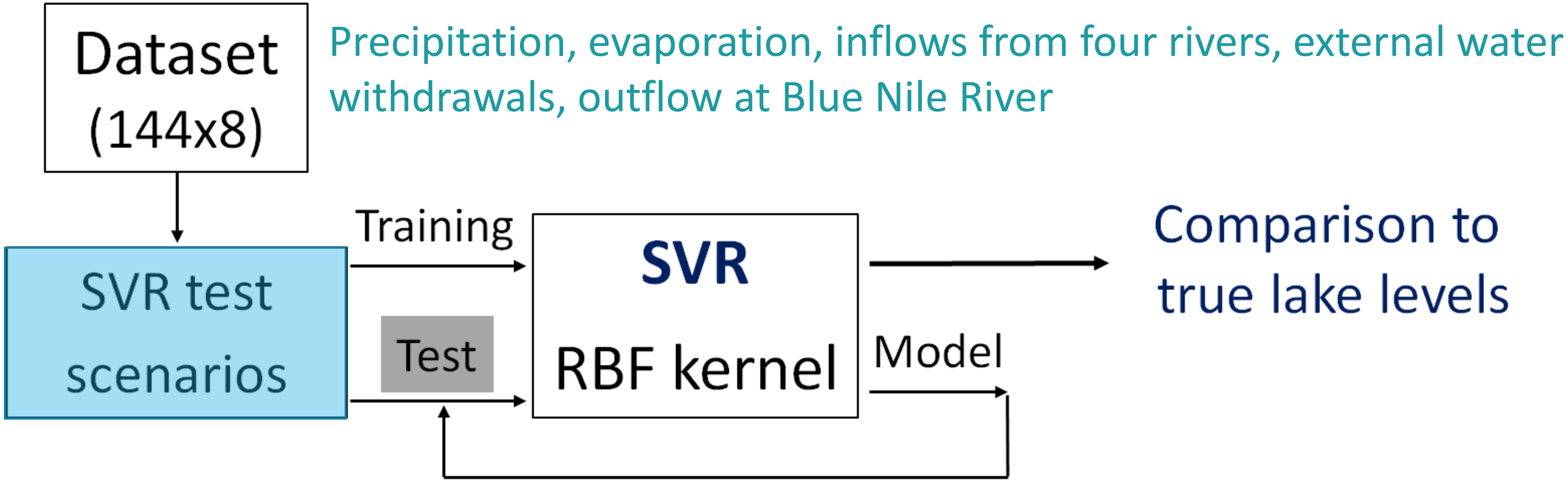


PIRE

NEUREL 2018

Modeling framework: Data-driven model

Support Vector Regression (SVR)



Radial basis function (RBF) kernel:

$$K\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right) = \exp\left(-\gamma \left\|\boldsymbol{x}_i - \boldsymbol{x}_j\right\|^2\right)$$

$$\gamma = \frac{1}{2\sigma^2}$$

γ = {0.001, 0.005, 0.01, 0.05, 0.1, 0.15, 0.2}

C = {0.01, $1\cdot10^i$, $2\cdot10^i$, $5\cdot10^i$}, i = 0,1,…,4 (regularization parameter)

LIBSVM toolbox v3.22

PIRE

NEUREL 2018

SVR Test Scenarios

10-fold cross-validation (permuted)

↓

12-fold cross-validation (un-permuted, each year is a test set)

↓

¼ Hold out (Test years: 2004, 2006, 2010)

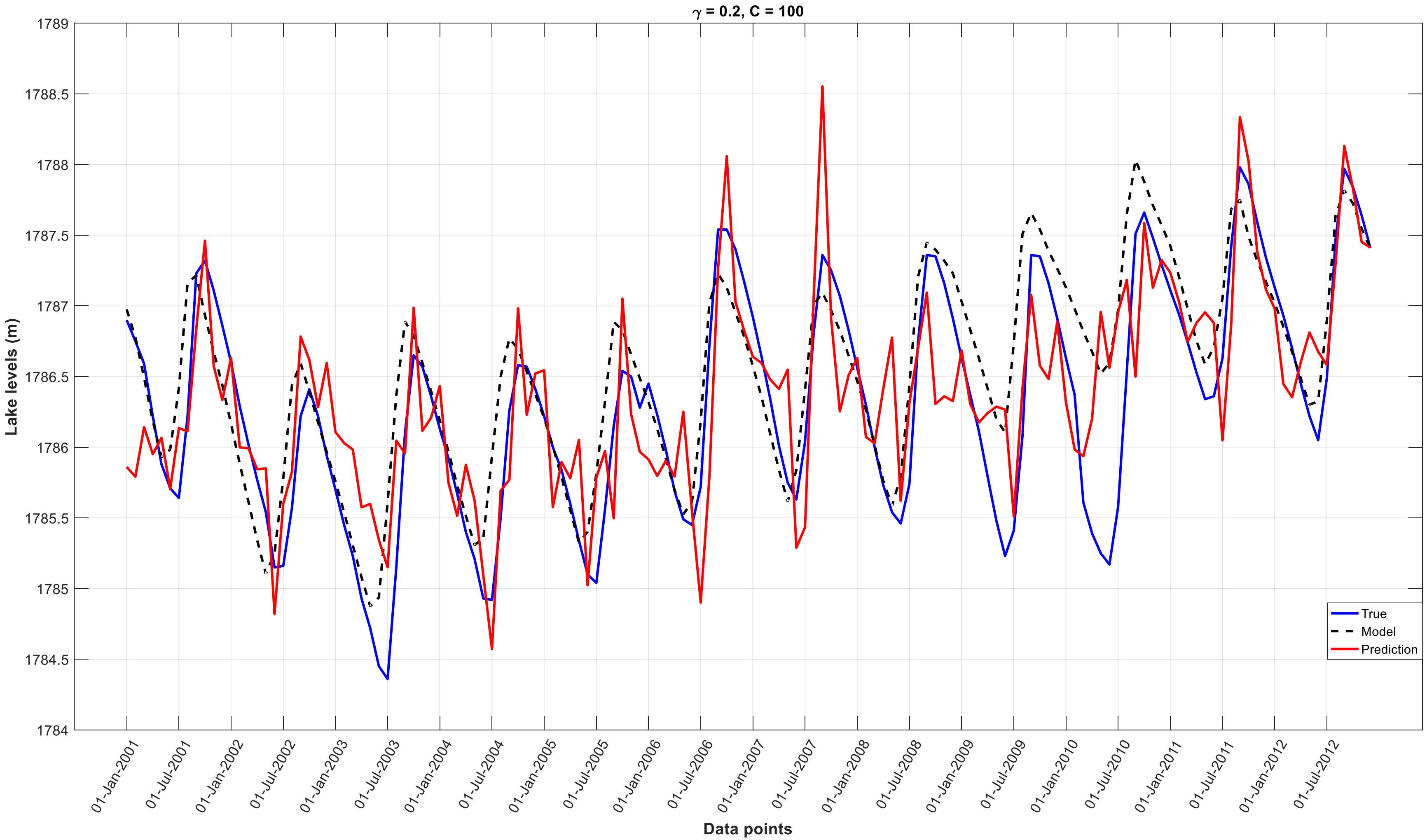
↓

12 1/12 Hold out algorithms (each year is a test year)

PIRE

NEUREL 2018

Scenario 1: 10-fold cross-validation

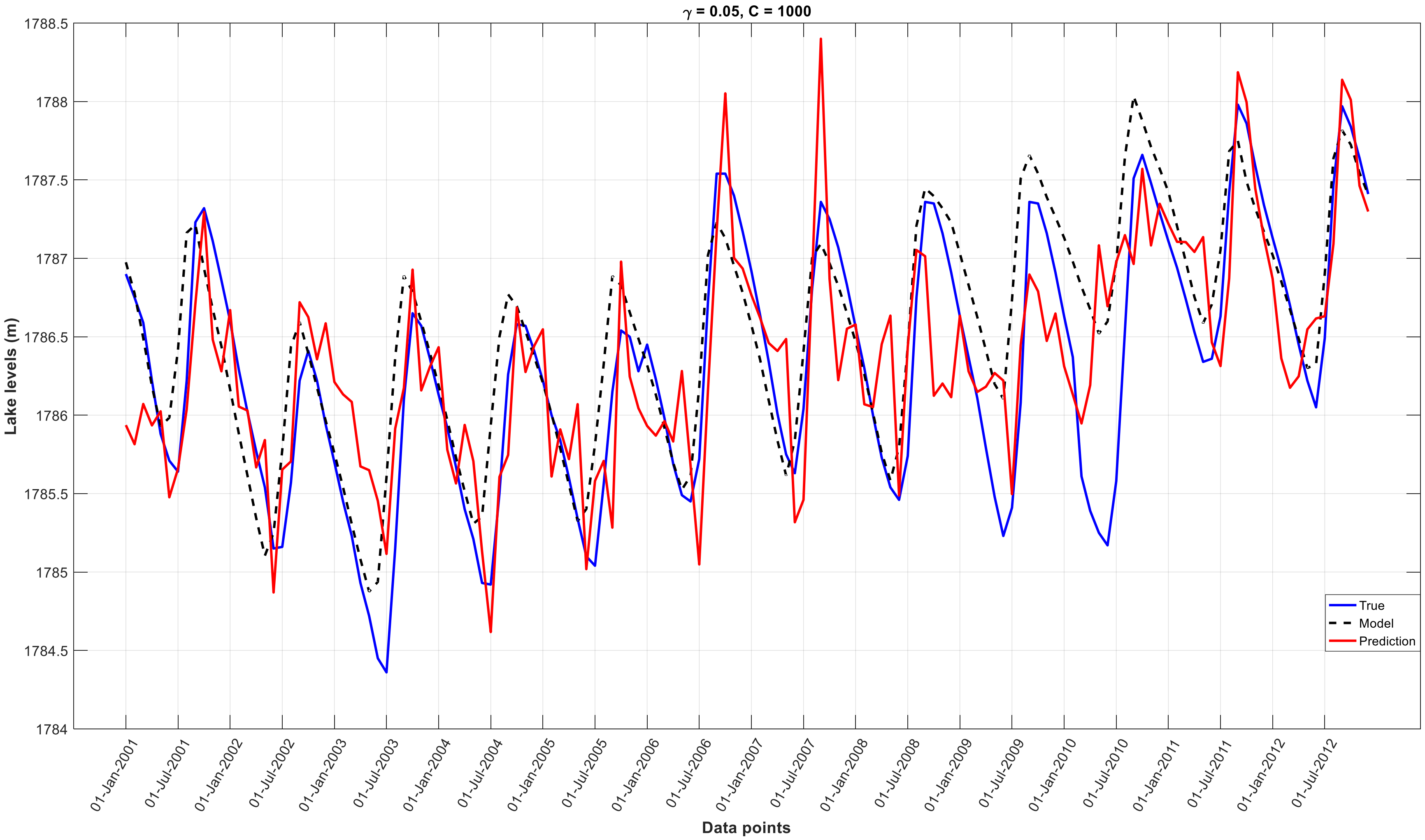


	MSE (m^2)	RMSE (m)	NRMSE	Mean error (m)	r	Bias (m)
γ = 0.2, C = 100	0.2788	0.5280	0.0296	0.4161	0.7636	0.065
Model	0.2376	0.4875	0.0273	0.3438	0.8376	0.212

PIRE

NEUREL 2018

Scenario 2: 12-fold cross-validation

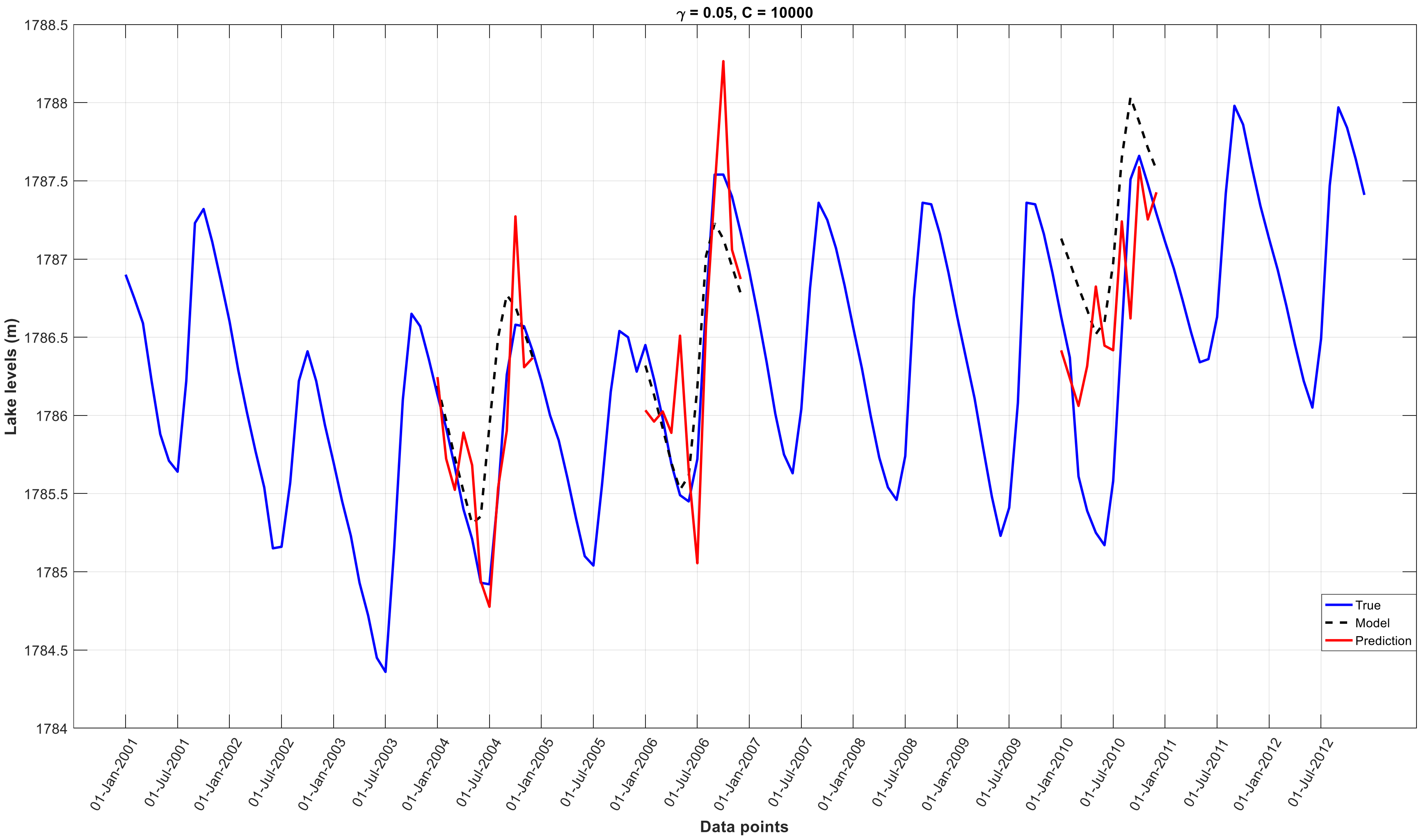


	MSE (m^2)	RMSE (m)	NRMSE	Mean error (m)	r	Bias (m)
γ = 0.05, C = 1000	0.2759	0.5252	0.0294	0.4089	0.7639	0.055
Model	0.2376	0.4875	0.0273	0.3438	0.8376	0.212

PIRE

NEUREL 2018

Scenario 3: ¼ Hold out (Test years: 2004, 2006, 2010)

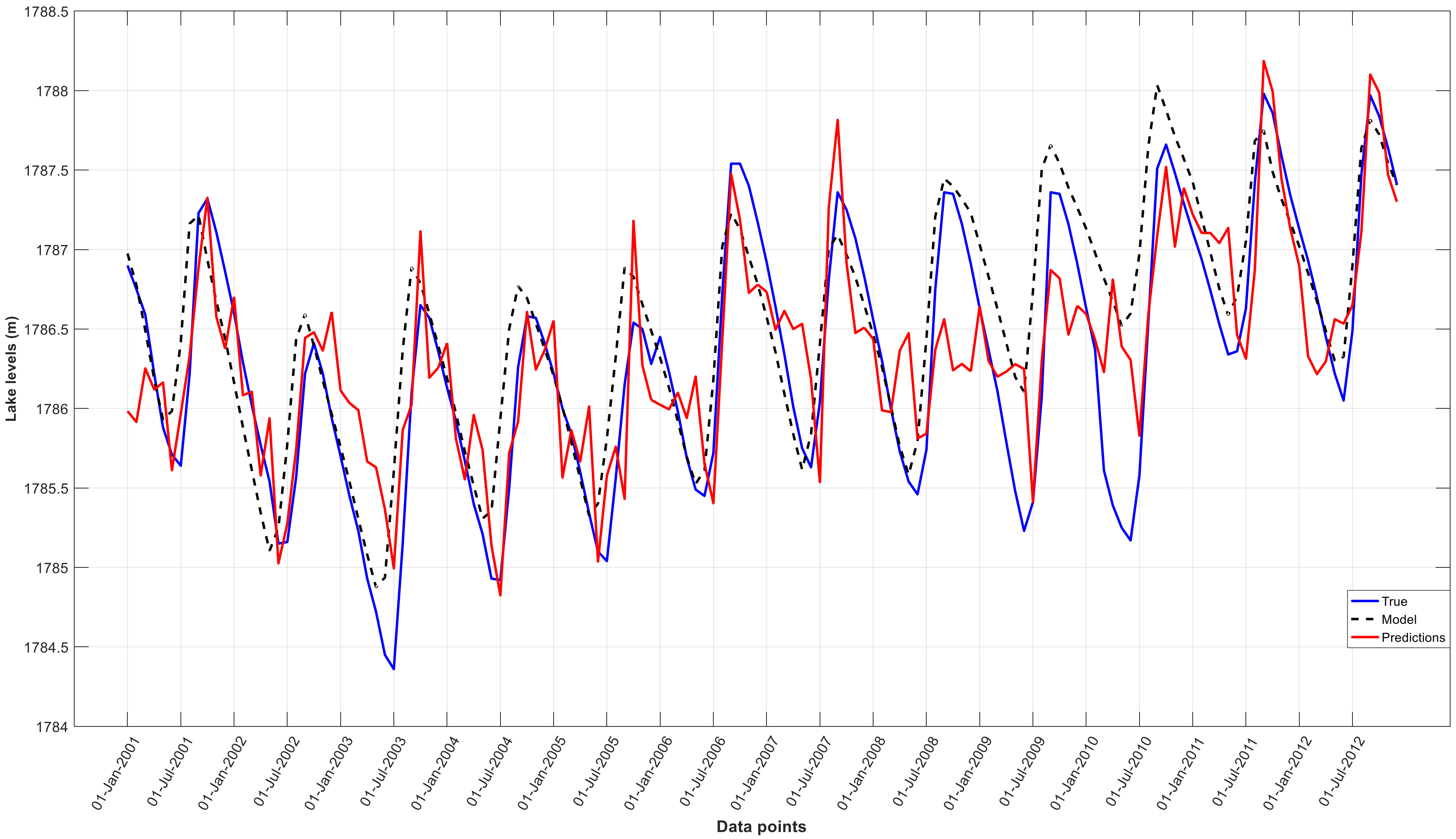


	MSE (m²)	RMSE (m)	NRMSE	Mean error (m)	r	Bias (m)
γ = 0.05, C = 10000	0.3082	0.5552	0.0311	0.4115	0.7730	0.139
Model	0.4030	0.6349	0.0355	0.4554	0.7709	0.348

PIRE

NEUREL 2018

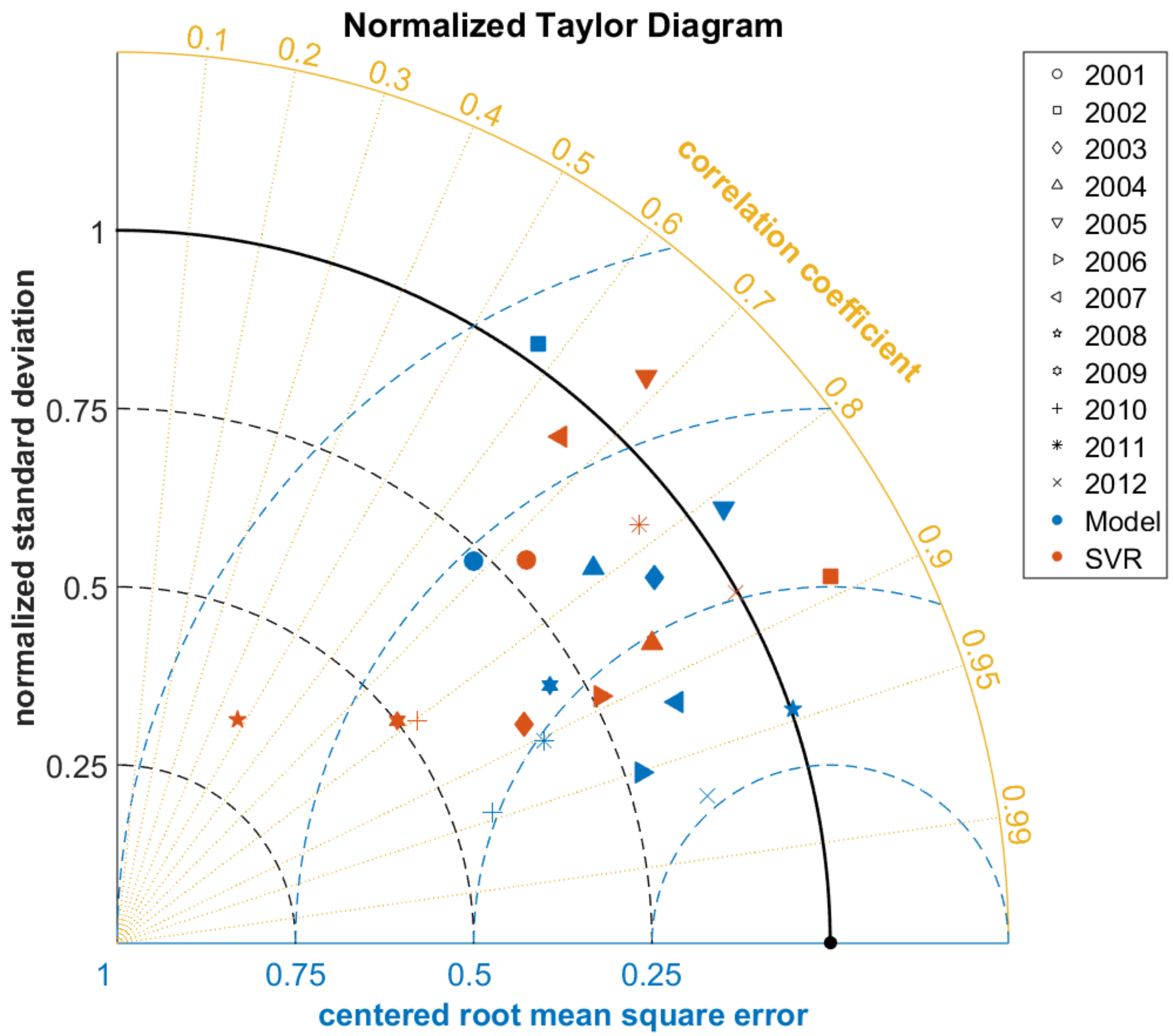
Scenario 4: 12 1/12 Hold out algorithms



PIRE

NEUREL 2018

Scenario 4: 12 1/12 Hold out algorithms



PIRE

NEUREL 2018

Summary and Conclusions

- The levels of Lake Tana were estimated using two approaches:
  - a physically-based groundwater model, coupled with a lake and streamflow modules
  - a data-driven algorithm which uses support vector regression with RBF kernel
- Both techniques achieved satisfactory results in estimating the lake levels in terms of various statistical metrics.
- The physically-based model outperformed the data-driven model in all but the bias metric.
- The data-driven model has multiple competitive advantages:
  - reduced computational effort,
  - shorter training/calibration time,
  - requires the selection of fewer model parameters
- Next step: explore forecasting capabilities of the data-driven model using incremental SVR

PIRE


THANK YOU
Хвала