Estimation of annual runoff using selected data machine learning algorithm

Authors: ujjwal singh[1], Rajani Kumar Pradhan[1], Shailendra Pratap[1], Martin Hanel[1], Ioannis Markonis[1], Sadaf Nasreen[1], and Petr Maca[1]

[1]Department of Water Resources and Environmental Modeling Czech University of Life Sciences Prague

Email: singh@fzp.czu.cz

Czech University of Life Sciences Prague
Faculty of Environmental Sciences

Runoff in the context of climate change

- Currently two-thirds of global population has been exposed to the global water scarcity.
- Anthropogenic activity and climate change are two main reason of water resource scarcity in Europe.
- Climate change alone poses a higher degree of threat on surface water security due to influence on precipitation and other essential climate variables that's lead to variability in water supplying many regions.
- Increasing population pressure on water resources has been adversely affected the health, sustainable development, and economy over Europe.
- The scarcity of surface runoff directly links with drought. 37% area of Europe has been affected by drought since three decades with triggering the socio-economical losses [1].

Input data and study area

- In this study, 50 years of reconstructed gridded data of precipitation [2] and temperature [3] has been taken for the whole of Europe.
- The gridded runoff E-RUN [4] has been taken as a benchmark dataset for training and validation of the overlap period.
- The E-RUN 25 years (1950-1975) data has selected for training and remaining 25 years data (1975-2000) for validation.
- In this study E-Budyko runoff has been ensemble mean of four different Budyko function [5-8] in order to minimise the each model biases.

Methodology

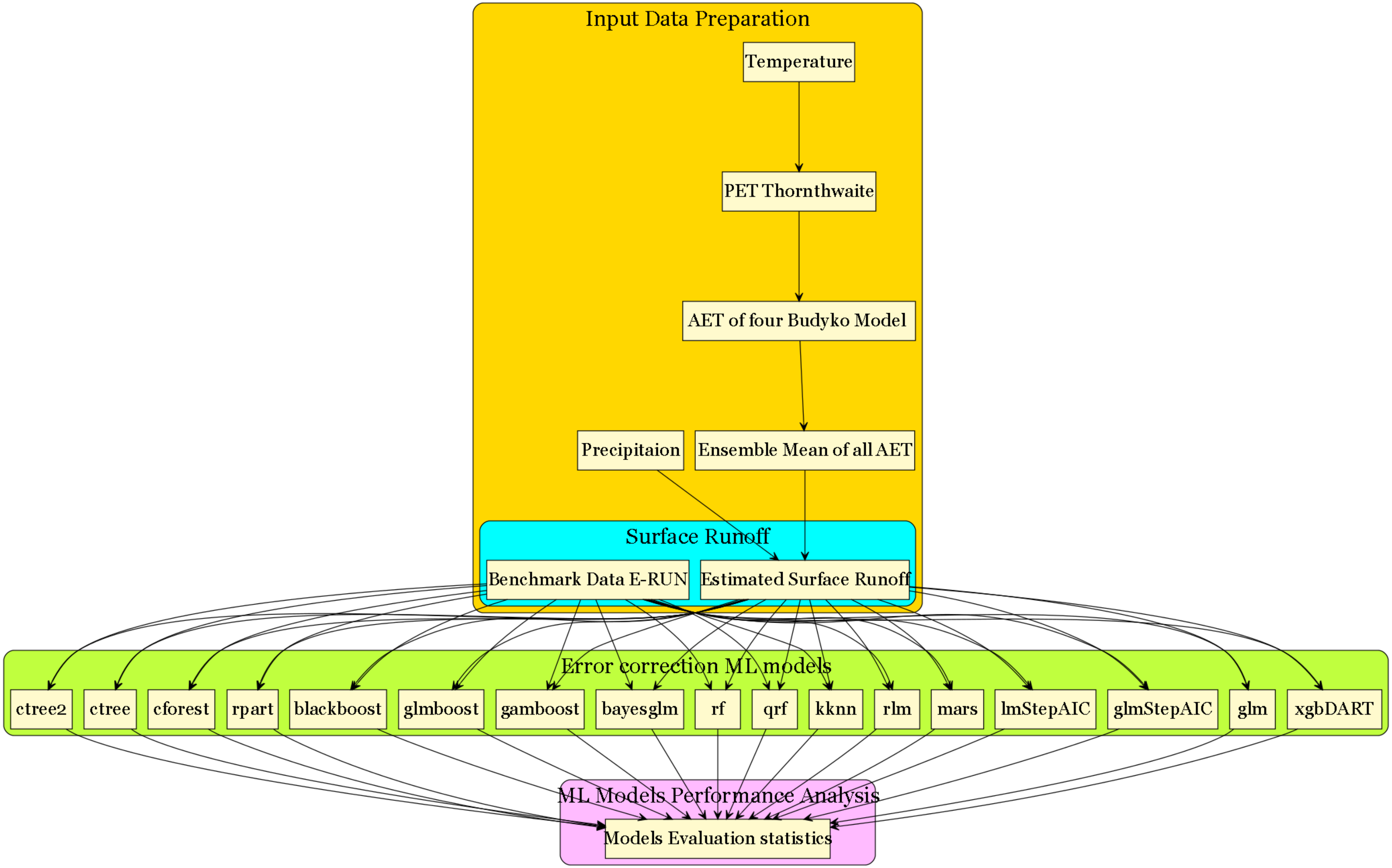

Fig.1 Flow chart of methodology

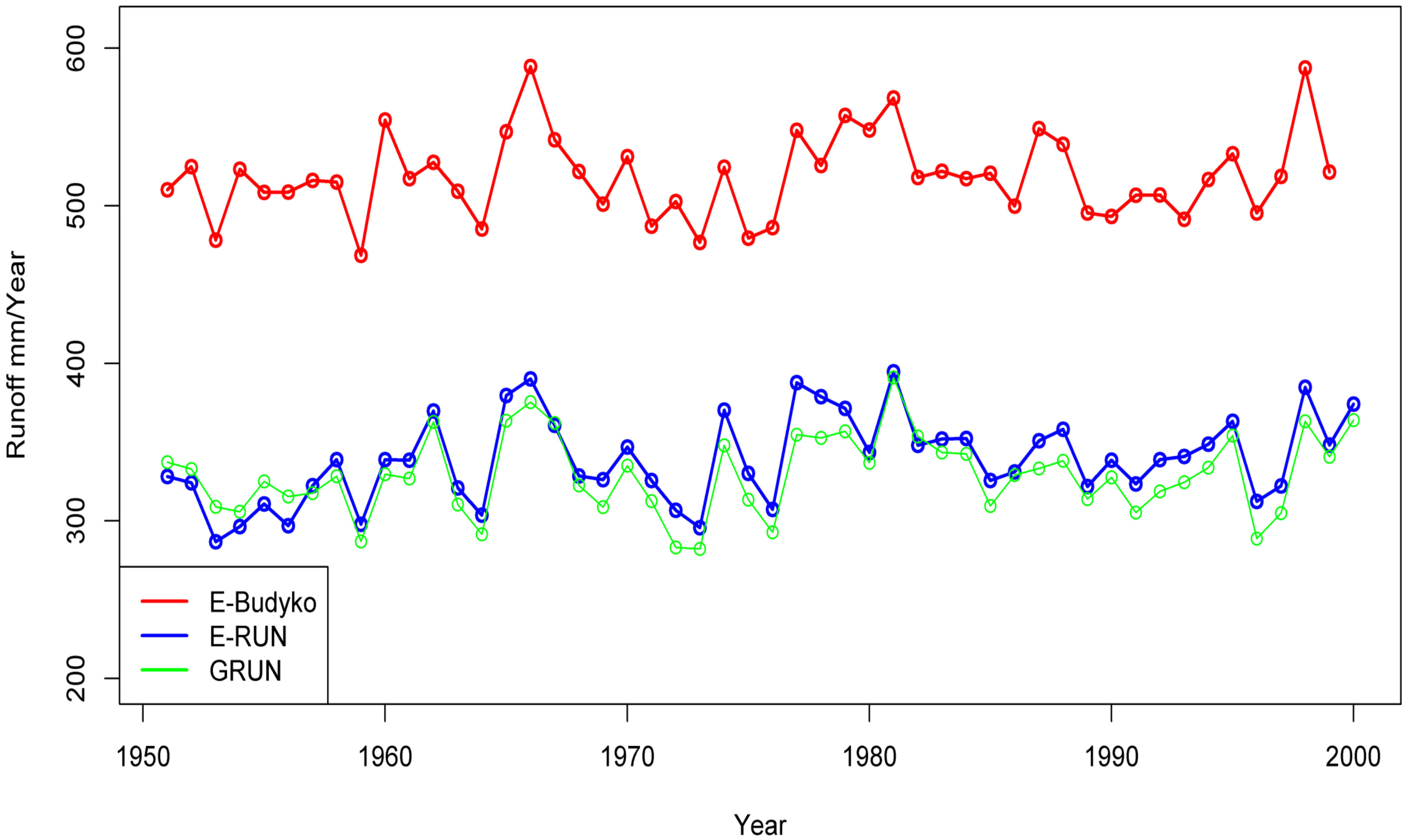

Fig.2 Spatial mean :time Series of Ensemble mean of Budyko models, E-RUN and GRUN data over Europe.

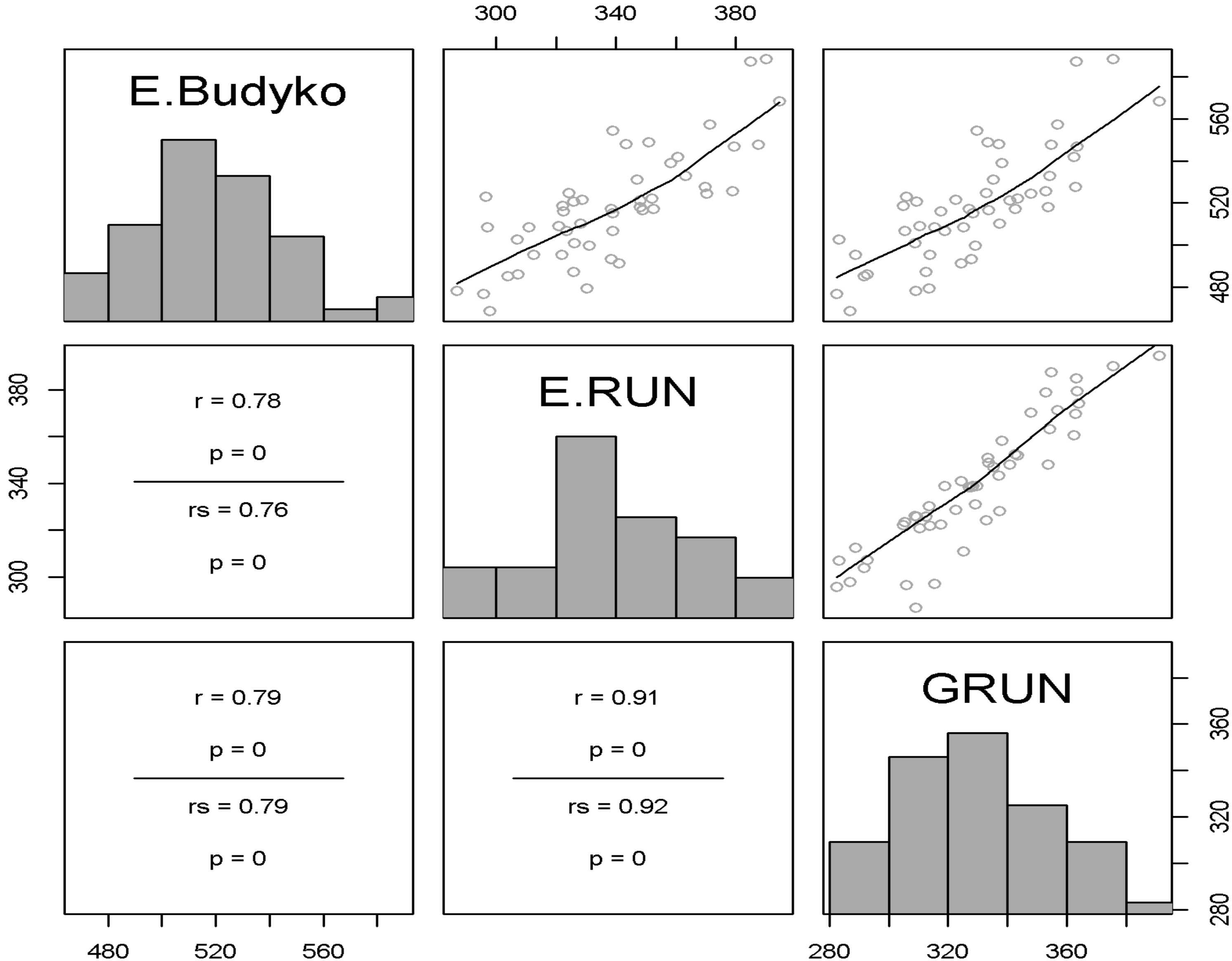

Fig.3 Correlation Matrix plot of spatial mean ,among Ensemble mean of Budyko, E-RUN and GRUN surface runoff data sets.

Why machine learning (ML) models in spite of Budyko physical model

- Temperature-based PET does not capture the precise accuracy of evaporative flux [9].
- Climate reconstructed data (precipitation and Temperature) has uncertainty [2].
- Fig.2 clearly represent the ensemble mean Budyko model has been not captured the realty of actual surface runoff due to above mentioned uncertainty(data and temperature based PET).
- Uncertainty in reconstructed data can be minimized using ML techniques.

Selected ML models for error correction in physical model of runoff

- ML models has own merit and demerit with respect to data nature.
- Diverse nature of ML models regression performed to understand the diversity of estimation with identification of best one.
- Each selected model tried to reduce the biases of the mean of four Budyko models.
- The evaluation statistics of training and validation of each ML models has more or less similar.

Individual ML models	R (spatial mean of Europe)	RMSE (spatial mean of Europe in mm/year)
Random Forest(rf)	0.9066816	37.67804
Bayesian Generalized Linear Model (bayesglm)	0.6397784	63.64268
Boosted Generalized Additive Model (gamboost)	0.6791987	61.49987
Boosted Generalized Linear Model (glmboost)	0.6397784	63.66092
Boosted Tree (blackboost)	0.6964298	62.47439
CART (rpart)	0.6724755	62.55502
Conditional Inference Random Forest (cforest)	0.6902587	60.47864
Conditional Inference Tree (ctree)	0.6802773	61.27214
Conditional Inference Tree (ctree2)	0.6789248	61.40228
eXtreme Gradient Boosting (xgbDART')	0.7326762	55.92445
Generalized Linear Model (glm)	0.6397784	63.64267
Generalized Linear Model with Stepwise Feature Selection (glmStepAIC)	0.6430418	63.64941
Linear Regression with Stepwise Selection (lmStepAIC)	0.6430418	63.64941
Multivariate Adaptive Regression Spline (earth)	0.6573511	62.95337
Robust Linear Model (rlm)	0.6396635	64.05868
k-Nearest Neighbors (kknn)	0.7490938	52.8055
Quantile Random Forest (qrf)	0.9077723	35.77893

Table.1 ML models evaluation statistics for validation data

Evaluation of models performance

- For evaluation of selected ML model E-RUN data has been used as a benchmark.
- ML model performance depend upon on the data property and its reliability.
- Quantile random forest performance is the best among all ML models.
- Random forest performance is also good due to the benchmark data reconstructed using random forest method.

Conclusion

- The selected machine learning models has a potential to reconstruct the past and future surface runoff.
- The historical estimated surface runoff is helpful to understand the decadal trend of surface runoff.
- Future reconstructed surface runoff is also helpful for government planning and policy with respect to water resource management.
- The selected model pixel wise temporal performance is also depend upon the number of data availability for training.

References

[1] M. Kossida, A. Kakava, A. Tekidou, A. Iglesias,M. Mimikouet al., “Vulnerability to water scarcity anddrought in europe,” 2017

[2] K. Beven, “A manifesto for the equifinality thesis,” Journal of hydrology, vol. 320, no. 1-2, pp. 18–36, 2006.

[3] A. Pauling, J. Luterbacher, C. Casty, and H. Wanner, “Five hundred years of gridded high-resolution precipitation reconstructions over europe and the connection to large-scale circulation,” Climate dynamics, vol. 26, no. 4, pp. 387–405, 2006.

[4] L. Gudmundsson and S. I. Seneviratne, “Observation-based gridded runoff estimates for europe (e-runversion 1.1),”Earth System Science Data, vol. 8, no. 2, pp.279–295, 2016. [Online]. Available: https://www.earth-syst-sci-data.net/8/279/2016

[5] P. Schreiber, “¨Uber die beziehungen zwischen demniederschlag und der wasserf ̈uhrung der fl ̈usse in mit-teleuropa,”Z. Meteorol, vol. 21, no. 10, pp. 441–452,1904.

[6] E. Ol’Dekop, “On evaporation from the surface of riverbasins,”Transactions on meteorological observations, vol. 4,p. 200, 1911.

[7] L. Turc, “Le bilan d’eau des sols: relations entre lespr ́ecipitations, l’ ́evaporation et l’ ́ecoulement,” Ph.D.dissertation, 1953.

[8] J. Pike, “The estimation of annual run-off from meteo-rological data in a tropical climate,”Journal of Hydrol-ogy, vol. 2, no. 2, pp. 116–123, 1964

[9] J. Sheffield, E. F. Wood, and M. L. Roderick, “Little change in global drought over the past 60 years,” Nature, vol. 491, no. 7424, pp. 435–438, 2012.