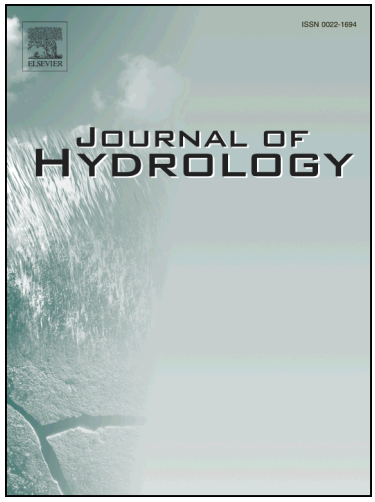

Research papers

Daily pan evaporation modeling from local and cross-station data using three tree-basedmachine learning models

Xianghui Lu, Yan Ju, Lifeng Wu, Junliang Fan, Fucang Zhang, Zhijun Li

PII: S0022-1694(18)30749-2
DOI: https://doi.org/10.1016/j.jhydrol.2018.09.055
Reference: HYDROL 23152

To appear in: *Journal of Hydrology*

Received Date: 15 July 2018
Revised Date: 6 September 2018
Accepted Date: 26 September 2018

Please cite this article as: Lu, X., Ju, Y., Wu, L., Fan, J., Zhang, F., Li, Z., Daily pan evaporation modeling from local and cross-station data using three tree-basedmachine learning models, *Journal of Hydrology* (2018), doi: https://doi.org/10.1016/j.jhydrol.2018.09.055

Daily pan evaporation modeling from local and cross-station data using three tree-based machine learning models

Xianghui Lu[a], Yan Ju[a], Lifeng Wu[a,b], Junliang Fan[b,*], Fucang Zhang[b], Zhijun Li[b]

[a]*School of Hydraulic and Ecological Engineering, Nanchang Institute of Technology, Nanchang 330099, China*

[b]*Key Laboratory of Agricultural Soil and Water Engineering in Arid and Semiarid Areas of Ministry of Education, Northwest A&F University, Yangling 712100, China*

* Corresponding author.

Dr. Junliang Fan

College of Water Resources and Architectural Engineering, Northwest A&F University,

Yangling 712100, China

E-mail: nwwfjl@163.com

Abstract

Accurate estimation of pan evaporation (E_p) is required for many applications, e.g., water resources management, irrigation system design and hydrological modeling. However, the estimation of E_p for a target station can be difficult as a result of partial or complete lack of local meteorological data under many conditions. In this study, daily E_p was estimated from local (target-station) and cross-station data in the Poyang Lake Watershed of China using four empirical models and three tree-based machine learning models, including M5 model tree (M5Tree), random forests (RF_s) and gradient boosting decision tree (GBDT). Daily meteorological data during 2001-2010 from 16 weather stations were used to train the models, while the data from 2011 to 2015 were used for testing. Two cross-station