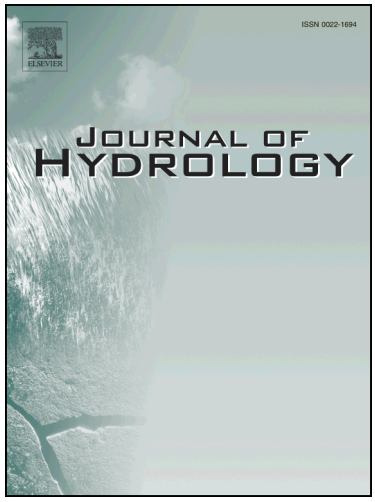

Research papers

A Soil Moisture Estimation Framework Based on the CART Algorithm and Its Application in China

Jiaqi Han, Kebiao Mao, Tongren Xu, Jingpeng Guo, Zhiyuan Zuo, Chunyu Gao

PII: S0022-1694(18)30379-2
DOI: https://doi.org/10.1016/j.jhydrol.2018.05.051
Reference: HYDROL 22828

To appear in: *Journal of Hydrology*

Received Date: 22 August 2017
Revised Date: 17 May 2018
Accepted Date: 21 May 2018

Please cite this article as: Han, J., Mao, K., Xu, T., Guo, J., Zuo, Z., Gao, C., A Soil Moisture Estimation Framework Based on the CART Algorithm and Its Application in China, *Journal of Hydrology* (2018), doi: https://doi.org/10.1016/j.jhydrol.2018.05.051

A Soil Moisture Estimation Framework Based on the CART Algorithm and Its Application in China

Jiaqi Han[1], Kebiao Mao[1,2,3]*, Tongren Xu[4], Jingpeng Guo[1], Zhiyuan Zuo[1], Chunyu Gao[1]

1 National Hulunber Grassland Ecosystem Observation and Research Station, Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences, Beijing 100081, China; gaochunyu@caas.cn; zuozhiyuan2004@126.com;
2 State Key Laboratory of Remote Sensing Science, Institute of Remote Sensing and Digital Earth Research, Chines Academy of Science, Beijing 100086, China;
3 College of Resources and Environments, Hunan Agricultural University, Changsha 410128, China;
4 School of Geography, Beijing Normal University, Beijing 100086, China; 11112011088@bnu.edu.cn
* Correspondence: maokebiao@caas.cn

Abstract: Soil moisture is an important parameter associated with the land-atmosphere interface and is highly influenced by multiple factors. Previous studies have provided an effective mechanism for accurately estimating soil moisture by building a global estimation model that comprehensively integrates multiple factors at a local scale. However, a global model is inefficient for accurately estimating soil moisture at a large or even global scale because of the complex surface features that make it difficult to fit data globally. Furthermore, inconsistencies in the spatial integrity between multisource data and the mismatch between the training space and application space decrease the generalizability of the model, which may lead to unreasonable soil moisture values in certain areas. This study proposes a "pyramid" framework that integrates multiple factors from different sources using the classification and regression tree (CART) algorithm, a machine learning method, to estimate soil moisture at a high spatial resolution (1 km). The framework considers soil moisture as a response variable and several factors, such as precipitation, soil properties, and temperature, as explanatory variables. The framework uses piecewise fitting instead of global fitting and avoids the generation of unreasonable values. A k-fold cross-validation approach using "hold-out" years was used to assess the performance of the soil moisture estimation framework for the summer period. The results show that the performance