AsiaRiceYield4km: Seasonal Rice Yield in Asia from 1995 to 2015

Huaqing Wu[1,2]★, Jing Zhang[1,2]★, Zhao Zhang[1,2], Jichong Han[1,2], Juan Cao[1,2], Liangliang Zhang[1,2], Yuchuan Luo[1,2], Qinghang Mei[1,2], Jialu Xu[2], Fulu Tao[3,4]

[1]Key Laboratory of Environmental Change and Natural Disasters, Ministry of Education Beijing Normal University, Beijing 100875, People's Republic of China
[2]School of National Safety and Emergency Management, Beijing Normal University, Beijing 100875 / Zhuhai 519087, People's Republic of China
[3]Key Laboratory of Land Surface Pattern and Simulation, Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, 100101, People's Republic of China
[4]College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049, People's Republic of China

★These authors contributed to the work equally and should be regarded as co-first authors.

Correspondence to: Zhao Zhang (zhangzhao@bnu.edu.cn)

Supplementary Methods

Tuning hyper-parameters is helpful to improve ML prediction accuracy (Shahhosseini et al., 2021). Here, the grid search method was used to find the optimal parameter combinations. Three hyper-parameters of RF were tuned: '*n_estimators*', '*max_features*', and '*min_samples_split*', and six for XGBoost: '*n_estimators*', '*max_depth*', '*min_child_weight*', '*eta*', '*gamma*', '*tree_method*'. For LSTM, the dropout rate was set as 0.2 and L2 regularization was used to reduce over-fitting. Besides, the early stopping patience was set as 50 to further reduce over-fitting. The hyper-parameter through the GridSearchCV package in Python 3.7.

Reference:

Shahhosseini, M., Hu, G., Huber, I., and Archontoulis, S. V.: Coupling machine learning and crop modeling improves crop yield prediction in the US Corn Belt, Sci. Rep., 11, 1606, https://doi.org/10.1038/s41598-020-80820-1, 2021.

Supplementary Tables

Table S1: Rice yield sources and the minimum yield available administrative units for countries in the study area

Country	Source	Names of the local administrative unit
Bangladesh	https://bbs.portal.gov.bd; http://www.brri.gov.bd	district
Cambodia	https://www.fao.org/in-action/countrystat/national-countrystat-sites/en/	province
China	http://www.stats.gov.cn/	country, district
India	https://data.gov.in/	division
Indonesia	https://www.bps.go.id/; https://www.bps.go.id/subject/53/tanaman-pangan.html#subjekViewTab5.html	province
Japan	https://www.e-stat.go.jp/	prefectural division
Malaysia	https://www.statistics.gov.my/v1_/	state
Myanmar	https://www.mmsis.gov.mm/	region, state
Nepal	https://nepalindata.com/	district
Pakistan	http://www.amis.pk/	district
Philippines	https://www.fao.org/in-action/countrystat/national-countrystat-sites/en/	lalawigan
Republic of Korea	https://kostat.go.kr/portal/korea/index.action	si, gu, gun
Thailand	http://web.nso.go.th/	province
Vietnam	http://www.gso.gov.vn/default_en.aspx/	province, municipality

Note: names of the local administrative unit presents the specific name of the administrative divisions. These names are collected from Wikipedia (https://en.wikipedia.org/w/index.php?title=List_of_administrative_divisions_by_country&oldid=1081175578, last accessed: 7 April 2022). Besides, all the links in the table can be accessed on 8 April 2022.

Table S2: Overview of the data required in this study

Dataset	Variables	Spatial resolution	Temporal resolution	Time range	Source
Paddy rice area map	Paddy rice area map	500m	Yearly	2000-2015	Han et al. (2022)
Rice yield	Rice yield of single, double, and triple seasons	County level	Yearly	1995-2015	Table S1
Crop calendar	Planting date, heading date, and harvesting date	4km	Yearly	1995-2015	Laborte et al. (2017)
Vegetation index	Leaf area index	0.05°	8 days	1995-2015	Xiao et al. (2016, 2013)
Climate variables	PDSI, Pre, Srad, Tmax, Tmin, Vap, Ws	4km	Monthly	1995-2015	Abatzoglou et al. (2018)
Soil properties	T_Sand, T_SILT, T_CLAY, T_BULK_DEN, T_OC, T_PH_H2O	30″	-	-	Wieder et al. (2014)
Elevation	Elevation	1km	-	-	Hastings et al., (1999)

Reference:

Abatzoglou, J. T., Dobrowski, S. Z., Parks, S. A., and Hegewisch, K. C.: TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015, Sci. Data, 5, 1–12, https://doi.org/10.1038/sdata.2017.191, 2018.

Han, J., Zhang, Z., Luo, Y., Cao, J., Zhang, L., Zhuang, H., Cheng, F., Zhang, J., and Tao, F.: Annual paddy rice planting area and cropping intensity datasets and their dynamics in the Asian monsoon region from 2000 to 2020, Agric. Syst., 200, 103437, https://doi.org/10.1016/j.agsy.2022.103437, 2022.

Hastings, D. A., Dunbar, P. K., Elphingstone, G. M., Bootz, M., Murakami, H., Maruyama, H., Masaharu, H., Holland, P., Payne, J., and Bryant, N. A.: The global land one-kilometer base elevation (GLOBE) digital elevation model, version 1.0 National Oceanic and Atmospheric Administration, National Geophysical Data Center [data set], 325, 80305–3328, 1999.

Laborte, A. G., Gutierrez, M. A., Balanza, J. G., Saito, K., Zwart, S. J., Boschetti, M., Murty, M. V. R., Villano, L., Aunario, J. K., Reinke, R., Koo, J., Hijmans, R. J., and Nelson, A.: RiceAtlas, a spatial database of global rice calendars and production, Sci. Data, 4, 170074, https://doi.org/10.1038/sdata.2017.74, 2017.

Wieder, W. R., Boehnert, J., Bonan, G. B., and Langseth, M.: Regridded harmonized world soil database v1. 2 ORNL DAAC [data set], https://doi.org/10.3334/ORNLDAAC/1247, 2014.

Xiao, Z., Liang, S., Wang, J., Chen, P., Yin, X., Zhang, L., and Song, J.: Use of general regression neural networks for generating the GLASS leaf area index product from time-series MODIS surface reflectance, IEEE Trans. Geosci. Remote Sens., 52, 209–223, https://doi.org/10.1109/TGRS.2013.2237780, 2013.

Xiao, Z., Liang, S., Wang, J., Xiang, Y., Zhao, X., and Song, J.: Long-time-series global land surface satellite leaf area index product derived from MODIS and AVHRR surface reflectance, IEEE Trans. Geosci. Remote Sens., 54, 5301–5318, https://doi.org/10.1109/TGRS.2016.2560522, 2016.

Table S3: The all predicted variables in four categories

Categories	Predicted variables	Abbreviation
TI	Year	-
CEC	Longitude	Lon
	Latitude	Lat
	Elevation	-
	Topsoil Sand Fraction	T_Sand
	Topsoil Silt Fraction	T_SILT
	Topsoil Clay Fraction	T_CLAY
	Topsoil Reference Bulk Density	T_BULK_DEN
	Topsoil Organic Carbon	T_OC
	Topsoil pH (H2O)	T_PH_H2O
CGP	The sum of LAI for whole growing period	Sum_LAI_WGP
	The sum of LAI for vegetative stage	Sum_LAI_VEP
	The sum of LAI for reproductive stage	Sum_LAI_REP
	The sum of Palmer Drought Severity Index for whole growing period	Sum_PDSI_WGP
	The sum of Palmer Drought Severity Index for vegetative stage	Sum_PDSI_VEP
	The sum of Palmer Drought Severity Index for reproductive stage	Sum_PDSI_REP
	The sum of Precipitation accumulated for whole growing period	Sum_Pre_WGP
	The sum of Precipitation accumulated for vegetative stage	Sum_Pre_VEP
	The sum of Precipitation accumulated for reproductive stage	Sum_Pre_REP
	The sum of srad for whole growing period	Sum_Srad_WGP
	The sum of srad for vegetative stage	Sum_Srad_VEP
	The sum of srad for reproductive stage	Sum_Srad_REP
	The sum of maximum temperature for whole growing period	Sum_Tmax_WGP
	The sum of maximum temperature for vegetative stage	Sum_Tmax_VEP
	The sum of maximum temperature for reproductive stage	Sum_Tamx_REP
	The sum of minimum temperature for whole growing period	Sum_Tmin_WGP
	The sum of minimum temperature for vegetative stage	Sum_Tmin_VEP
	The sum of minimum temperature for reproductive stage	Sum_Tmin_REP
	The sum of vapor pressure for whole growing period	Sum_Vap_WGP
	The sum of vapor pressure for vegetative stage	Sum_Vap_VEP
	The sum of vapor pressure for reproductive stage	Sum_Vap_REP

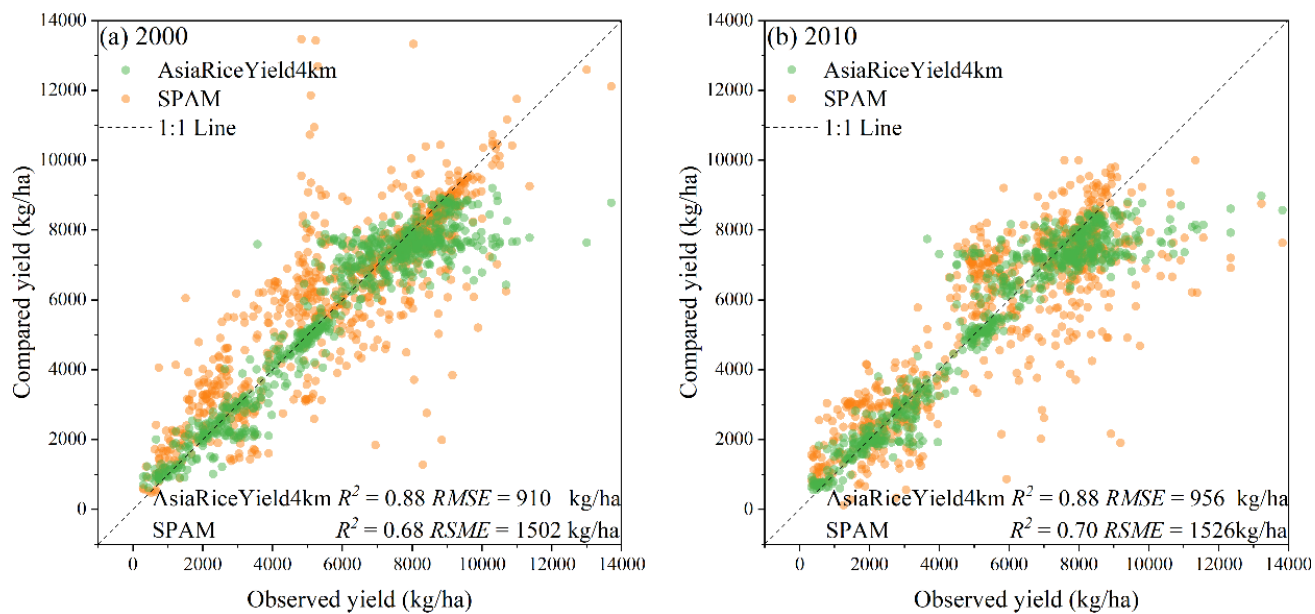|		
	The sum of for whole growing period	Sum_Ws_WGP
	The sum of for vegetative stage	Sum_Ws_VEP
	The sum of for reproductive stage	Sum_Ws_REP
EGP	The minimum LAI for whole growing period	Min_LAI_WGP
	The maximum LAI for whole growing period	Max_LAI_WGP
	The minimum Palmer Drought Severity Index for whole growing period	Min_PDSI_WGP
	The maximum Palmer Drought Severity Index for whole growing period	Max_PDSI_WGP
	The minimum Precipitation accumulated for whole growing period	Min_Pre_WGP
	The maximum Precipitation accumulated for whole growing period	Max_Pre_WGP
	The minimum downward surface shortwave radiation for whole growing period	Min_Srad_WGP
	The maximum downward surface shortwave radiation for whole growing period	Max_Srad_WGP
	The minimum temperature for whole growing period	Min_Tmax_WGP
	The maximum temperature for whole growing period	Max_Tmax_WGP
	The minimum temperature for whole growing period	Min_Tmin_WGP
	The maximum minimum temperature for whole growing period	Max_Tmin_WGP
	The minimum vapor pressure for whole growing period	Min_Vap_WGP
	The maximum vapor pressure for whole growing period	Max_Vap_WGP
	The minimum wind speed for whole growing period	Min_Ws_WGP
	The maximum for whole growing period	Max_Ws_WGP

Supplementary Figure

Figure S1: The accuracy of AsiaRiceYield4km and SPAM in 2000 and 2010.