Remote Sensing of Environment 284 (2023) 113335

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Paddy rice methane emissions across Monsoon Asia

Zutao Ouyang [a,*], Robert B. Jackson [a,b,c], Gavin McNicol [a,d], Etienne Fluet-Chouinard [a], Benjamin R.K. Runkle [e], Dario Papale [f,g], Sara H. Knox [h], Sarah Cooley [a,i], Kyle B. Delwiche [a,j], Sarah Feron [a,k], Jeremy Andrew Irvin [l], Avni Malhotra [m], Muhammad Muddasir [f], Simone Sabbatini [f,g], Ma. Carmelita R. Alberto [n], Alessandro Cescatti [o], Chi-Ling Chen [p], Jinwei Dong [q], Bryant N. Fong [r], Haiqiang Guo [s], Lu Hao [t], Hiroki Iwata [u], Qingyu Jia [v], Weimin Ju [w], Minseok Kang [x], Hong Li [y], Joon Kim [z], Michele L. Reba [r], Amaresh Kumar Nayak [aa], Debora Regina Roberti [ab], Youngryel Ryu [ac], Chinmaya Kumar Swain [aa], Benjei Tsuang [ad], Xiangming Xiao [ae], Wenping Yuan [af], Geli Zhang [ag], Yongguang Zhang [w]

[a] Department of Earth System Science, Stanford University, Stanford, CA, USA
[b] Woods Institute for the Environment, Stanford University, Stanford, CA, USA
[c] Precourt Institute for Energy, Stanford University, Stanford, CA, USA
[d] Department of Earth and Environmental Sciences, University of Illinois Chicago, Chicago, IL, USA
[e] Department of Biological and Agricultural Engineering, University of Arkansas, Fayetteville, AR, USA
[f] Department for Innovation in Biological, Agro-food and Forest systems (DIBAF), University of Tuscia, Viterbo, Italy
[g] Euro-Mediterranean Center on Climate Change (CMCC), IAFES division, Viterbo, Italy
[h] Department of Geography, The University of British Columbia, Vancouver, British Columbia, Canada
[i] Department of Geography, University of Oregon, Eugene, OR, USA
[j] Department of Environmental Science, Policy And Management, UC Berkeley, Berkeley, CA, USA
[k] Campus Fryslan, University of Groningen, Groningen, Netherlands
[l] Department of Computer Science, Stanford University, Stanford, CA, USA
[m] Department of Geography, University of Zurich, Zurich, Switzerland
[n] International Rice Research Institute, Laguna, Philippines
[o] European Commission, Joint Research Center (JRC), Ispra, Italy
[p] Taiwan Agricultural Research Institute, Wufeng, Taichung, Taiwan
[q] Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China
[r] USDA, ARS Delta Water Management Research Unit, Jonesboro, AR, USA
[s] Fudan University, Ministry of Education Key Laboratory for Biodiversity Science and Ecological Engineering, and Coastal Ecosystems Research Station of the Yangtze River Estuary, Shanghai, China
[t] Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters (CIC-FEMD) / Jiangsu Key Laboratory of Agricultural Meteorology, Nanjing University of Information Science and Technology, Nanjing, China
[u] Department of Environmental Science, Shinshu University, Matsumoto, Japan
[v] China Meteorological Administration, Institute of Atmospheric Environment, Shenyang, China
[w] International Institute for Earth System Sciences, Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing University, Nanjing, Jiangsu 210023, China
[x] National Center for AgroMeteorology, Seoul, South Korea
[y] Faculty of Architecture and Urban Planning, Chongqing University, Chongqing, China
[z] Seoul National University, Landscape Architecture & Rural Systems Engineering, Seoul, South Korea
[aa] ICAR-National Rice Research Institute Cuttack, Odisha, India
[ab] UFSM Federal University of Santa Maria, Santa Maria, Brazil
[ac] Department of Landscape Architecture and Rural Systems Engineering, Seoul National University, Seoul, South Korea
[ad] Department of Environmental Engineering, National Chung-Hsing University, Taichung, Taiwan
[ae] Department of Microbiology and Plant Biology, Center for Earth Observation and Modeling, University of Oklahoma, Norman, OK, USA
[af] Sun Yat-sen University, School of Atmospheric Sciences, Guangzhou, China
[ag] Agricultural University, College of Land Science and Technology, Beijing, China

* Corresponding author.
E-mail address: ouyangzt@stanford.edu (Z. Ouyang).

https://doi.org/10.1016/j.rse.2022.113335
Received 11 June 2022; Received in revised form 23 October 2022; Accepted 27 October 2022
Available online 14 November 2022
0034-4257/© 2022 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

ARTICLE INFO

Edited by Jing M. Chen

Keywords:
Remote sensing
Climate change
Greenhouse gas emission
Machine learning
Eddy covariance

ABSTRACT

Although rice cultivation is one of the most important agricultural sources of methane (CH_4) and contributes ~8% of total global anthropogenic emissions, large discrepancies remain among estimates of global CH_4 emissions from rice cultivation (ranging from 18 to 115 Tg CH_4 yr^{-1}) due to a lack of observational constraints. The spatial distribution of paddy-rice emissions has been assessed at regional-to-global scales by bottom-up inventories and land surface models over coarse spatial resolution (e.g., > 0.5°) or spatial units (e.g., agro-ecological zones). However, high-resolution CH_4 flux estimates capable of capturing the effects of local climate and management practices on emissions, as well as replicating in situ data, remain challenging to produce because of the scarcity of high-resolution maps of paddy-rice and insufficient understanding of CH_4 predictors. Here, we combine paddy-rice methane-flux data from 23 global eddy covariance sites and MODIS remote sensing data with machine learning to 1) evaluate data-driven model performance and variable importance for predicting rice CH_4 fluxes; and 2) produce gridded up-scaling estimates of rice CH_4 emissions at 5000-m resolution across Monsoon Asia, where ~87% of global rice area is cultivated and ~ 90% of global rice production occurs. Our random-forest model achieved Nash-Sutcliffe Efficiency values of 0.59 and 0.69 for 8-day CH_4 fluxes and site mean CH_4 fluxes respectively, with land surface temperature, biomass and water-availability-related indices as the most important predictors. We estimate the average annual (winter fallow season excluded) paddy rice CH_4 emissions throughout Monsoon Asia to be 20.6 ± 1.1 Tg yr^{-1} for 2001–2015, which is at the lower range of previous inventory-based estimates (20–32 CH_4 Tg yr^{-1}). Our estimates also suggest that CH_4 emissions from paddy rice in this region have been declining from 2007 through 2015 following declines in both paddy-rice growing area and emission rates per unit area, suggesting that CH_4 emissions from paddy rice in Monsoon Asia have likely not contributed to the renewed growth of atmospheric CH_4 in recent years.

1. Introduction

Atmospheric methane (CH_4) is a more potent greenhouse gas than carbon dioxide (CO_2) and has contributed approximately 0.5 °C to observed global warming (2010–2019 relative to 1850–1900) (IPCC, 2021). After a relatively stable period during 1999–2006, atmospheric methane concentration reached ~1908 ppb in May 2022 (Dlugokencky, 2022). Anthropogenic emissions from agriculture have been proposed as one of the main sources contributing to this revived CH_4 increase (Nisbet et al., 2016; Schaefer et al., 2016; Zhang et al., 2021). Rice cultivation is one of the most important CH_4 sources from agriculture and contributed an estimated 8% of total global anthropogenic emissions for the 2008–2017 decade (Saunois et al., 2020). However, large discrepancies exist among estimates of global CH_4 rice emissions, ranging from 18 to 115 Tg CH_4 yr^{-1} (Frankenberg et al., 2005; Saunois et al., 2020; Yan et al., 2009; Zhang et al., 2016).

Both bottom-up and top-down approaches have been used to estimate CH_4 emissions from paddy rice. Top-down approaches use atmospheric CH_4 measurements with transport model inversions to infer surface CH_4 emissions. Top-down approaches also require prior emissions estimates and the spatial distribution of paddy rice fields, which are usually derived from bottom-up approaches (Bergamaschi et al., 2007; Bloom et al., 2010; Jacob et al., 2016; Saunois et al., 2020). Bottom-up approaches include both inventory methods drawing on region-specific emission factors (Yan et al., 2009) and process-based land surface models that simulate grid-based CH_4 emissions (Zhang et al., 2016), each scaled by the emitting area of paddy rice. However, universal emission factors in inventory methods typically average flux variability across environmental heterogeneities and climate dynamics and thus limit accurate predictions of CH_4 emissions. Process-based models consider multiple environmental factors and land-surface heterogeneities, but current models for paddy rice (e.g., Dynamic Land Ecosystem Model-DLEM and DeNitrification-DeComposition-DNDC) remain at coarse spatial resolution (e.g., usually >0.5°) (Wang et al., 2021; Zhang et al., 2016) and lack constraints and validation from longer-term and ecosystem-scale observations. New and independent data-driven estimates of paddy rice CH_4 emissions from observations that represent diverse management and climate conditions are important for benchmarking top-down and bottom-up estimates, reconciling differences between different estimates, and refining parameterization in process-based models (Jung et al., 2020).

Eddy covariance (EC) methods measure the quasi- continuous exchange of carbon (CO_2 and CH_4), water, and energy flux between the land surface and the atmosphere at ecosystem scales (Baldocchi, 2014), and can be combined with rigorous scaling methods and remote sensing data to produce spatially detailed landscape emission estimates. Coordination within the EC flux community has resulted in the formation of the international data network FLUXNET that provides standardized and gap-filled EC flux data of CO_2, water, and energy (Baldocchi, 2014; Papale, 2020). FLUXNET CO_2 data have been combined with remote sensing data and rigorous scaling methods to produce spatially detailed carbon flux estimates over large geographic scales for benchmarking and informing earth system models (Jung et al., 2020; Xiao et al., 2014).

An analogous global CH_4 synthesis for CH_4 fluxes (i.e., FLUXNET-CH_4) was produced more recently (Delwiche et al., 2021; Knox et al., 2019), allowing data-driven CH_4 flux products using eddy covariance measurements. One study, for instance, developed data-driven CH_4 emission products for northern-latitude wetlands (>45°N) (Peltola et al., 2019). Due to the scarcity of synthesized EC data for CH_4 flux in paddy rice (FLUXNET-CH_4 only contains 7 rice sites), no data-driven modeling approach to our knowledge has been used to spatially upscale rice-paddy CH_4 eddy covariance data regionally or globally.

To fill these gaps, we compiled and synthesized CH_4 flux data in paddy rice fields from 23 globally distributed EC sites and developed the first data-driven gridded paddy rice CH_4 emission maps (RiceCH_4). Our approach combines EC measurements and remote-sensing-based predictors from 2001 to 2015 across Monsoon Asia, which contains ~87% of the global paddy rice area and ~ 90% of global rice production (Zhang et al., 2020). We generated CH_4 emission maps at 8-day intervals, an interval length that is short enough to show the seasonality of CH_4 emissions and also matches the 8-day intervals of MODIS remote sensing products that provide key biophysical predictors of CH_4 emissions. We also produced maps at 5-km resolution to reveal detailed spatial distribution at regional and continental scales, which are unavailable from inventory-based methods. Overall, the higher temporal and spatial resolution of our products allows finer examination of the spatial-temporal variations of paddy-rice CH_4 emissions, in comparison with previous products and process-based studies, and better inputs as emissions priors in top-down inversion studies.

2. Data and methods

2.1. Eddy covariance data synthesis

We collected data from 23 paddy rice EC flux sites (Fig. 1, Supplementary Table 2), seven from the FLUXNET-CH_4 Version 1.0 dataset (Delwiche et al., 2021), 15 with data obtained directly from the site investigators, and one digitized (CN-JSY) from plots in publications (Ge et al., 2018). Most of our sites were in Monsoon Asia (14) and North America (7), with additional sites in Italy and Brazil.

Half-hourly flux and meteorological measurements were collected for all sites except CN-CMC, CN-SJP, CN-LHP, and CN-JSY. All data were quality-controlled, standardized, and post-processed using the procedures for the FLUXNET-CH_4 database (Delwiche et al., 2021). Briefly, we used the REddyProc package in R (Wutzler et al., 2018) to filter flux values with low friction velocity (u*) and to fill gaps in CO_2 and energy fluxes and in meteorological variables including air temperature (TA), incoming shortwave (SWIN) and relative humidity (RH), vapor pressure deficit (VPD), pressure (PA), precipitation (P), and wind speed (WS) using the marginal distribution sampling method (Reichstein et al., 2005). We then partitioned net CO_2 fluxes into gross primary production (GPP) and ecosystem respiration (ER) using both the nighttime (Reichstein et al., 2005) and daytime methods (Lasslop et al., 2010). Lastly, gaps in CH_4 flux were filled using a random-forest algorithm specifically developed for CH_4 time-series from both wetlands and rice paddy sites (Irvin et al., 2021). This method uses all available 30-min predictors measured at the site and an additional three temporal variables (decimal, sine, and cosine of day of year), where the gap-filling result is evaluated using a nested cross-validation procedure applied to artificial gaps (Irvin et al., 2021). The test scores of gap-filling are listed in Supplementary Table 1. Gap-filling performed well in 16 of our 23 sites ($R^2 > 0.65$) but performed only moderately well in the remaining seven sites ($0.37 < R^2 < 0.65$), with an overall mean R^2 of 0.74 ± 0.15 (mean ± standard deviation) among all sites.

Daily flux and meteorological measurements were collected for CN-CMC and CN-JSY, and 8-day flux and meteorological measurements were collected for CN-SJP and CN-LHP. The daily or 8-day data were computed by site investigators based on marginal distribution sampling (MDS)-gap-filled half-hour data.

Eight-day aggregates of CH_4 fluxes and other EC tower measurements were computed (sum for precipitation and mean for other variables) for all sites to match the 8-day composites of MODIS remote-sensing products. Since gap-filling performance decreases for longer gaps relative to shorter gaps (Irvin et al., 2021), only 8-day intervals with >40% of half-hourly CH_4 observations or >50% daily data available were included in our training data to further ensure data quality.

2.2. Predictor data preparation

A total of 175 predictors (including lagged terms of some variables) were explored to predict CH_4 flux, including site measurements of meteorological data (18 out of 175), MODIS remote sensing data (116 out of 175), other geospatial data on climate, soil, and topography (37 out of 175), and four artificial temporal variables (4 out of 175). Details on the full set of potential predictors are provided in Supplementary Table 3.

Meteorological variables include air temperature (TA), shortwave incoming radiation (SWIN), relative humidity (RH) and vapor pressure deficit (VPD), all of which were computed from tower measurements directly. These variables have been shown previously to be well correlated with CH_4 flux (Dai et al., 2019; Knox et al., 2016, 2021; Hwang et al., 2020; Li et al., 2019). We only tested local meteorological variables when they were measured at all sites and relevant spatial datasets are easily available (e.g., from European Centre for Medium-Range Weather Forecasts (ECMWF) re-analysis or other reanalysis datasets). Some potentially important environmental/meteorological variables, such as soil water content, and water table depth that are important CH_4 controlling variables (Delwiche et al., 2021; Knox et al., 2021), were not considered because data were missing across some sites and geospatial datasets did not exist for upscaling. Additionally, we considered the potential radiation at the top of the atmosphere that can be computed based on time, latitude, and longitude, because it embeds information on seasonal cycles and was also found to be useful in previous CH_4 flux upscaling (Peltola et al., 2019).

MODIS remote sensing data include land surface temperature and 12 indices related to surface vegetation, water, and soil conditions (Table 1). MODIS surface temperatures for both night and daytime were extracted from MOD11A2 (8-day intervals at 1000 m resolution) and converted from Kelvin to Celsius, whereas other indices were computed based on band-based references from MOD09A1 (8-day intervals at 500 m resolution). We applied quality control based on internal quality flags of MOD09A1 and MOD11A2. Specifically, questionable 8-day observations under cloudy, high view angle, or high solar zenith angle conditions were excluded. Short gaps (1–2 8-day time steps) in the data due to quality control removals were filled using linear interpolation for both MODIS land surface temperature and indices, and long gaps (>3 8-day time steps) were filled using the 2001–2015 mean seasonal cycles. MODIS data at the tower locations used for model training were

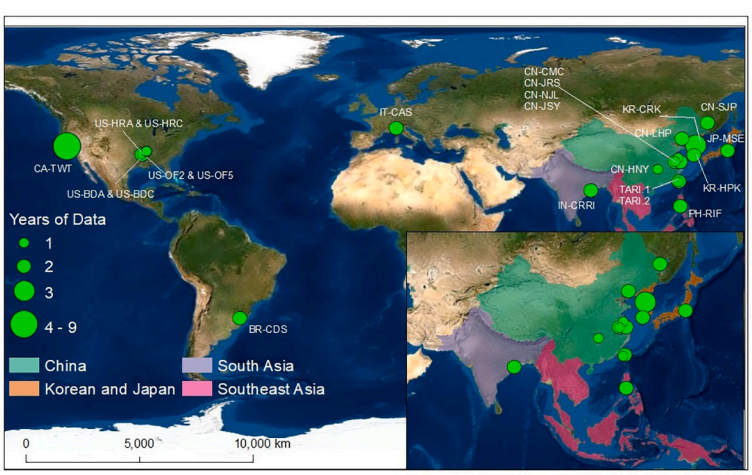

Fig. 1. Global locations of 23 distributed paddy rice EC flux sites, and (inset) the Monsoon Asia study area with its four sub-regions used to compare with other estimates of methane emissions from paddy rice. Although Mongolia is often considered to be part of monsoon Asia, it is excluded for its sparse rice cultivation and lack of rice calendar information (i.e., planting and harvesting dates). More information about the 23 sites is found in Supplementary Table 2.

Table 1
MODIS indices explored in this study with their formula and background references. B1 to B7 are MODIS bands where RED=B1, NIR1 = B2, BLUE = B3, GREEN=B4, NIR2 = B5, SWIR1 = B6, SWIR2 = B7. Many of these indices were originally developed and tested using Landsat images; we replaced them with MODIS bands with similar ranges of wavelength.

Indices	Full Name	Formula	References
NDVI	Normalized Vegetation Index	(NIR-RED)/(NIR + RED)	
EVI	Enhanced Vegetation Index	EVI = 2.5 * ((NIR1 - RED) / (NIR1 + 6 * RED - 7.5 * BLUE +1)	(Huete et al., 2002)
NDWI	Normalized Water Index	(NIR1- NIR2)/(NIR1+ NIR2)	(Gao, 1996)
SRWI	Simple Ratio Water Index	NIR1/NIR2	(Zarco-Tejada and Ustin, 2001)
LSWI	Land Surface Water Index	(NIR1 - SWIR1) / (NIR 1+ SWIR1)	(Xiao et al., 2002)
STI	Simple Ratio Tillage Index	SWIR1/SWIR2	(Van Deventer et al., 1997)
NDTI	Normalized Difference Tillage Index	(SWIR1-SWIR2)/ (SWIR1-SWIR2)	(Van Deventer et al., 1997)
CRC	Crop Residue Cover Index	(SWIR1 -BLUE)/ (SWIR1 + BLUE)	(Sullivan et al., 2006)
CRCm	Modified CRC index	(SWIR1 -GREEN)/ (SWIR1 + GREEN)	(Sullivan et al., 2006)
SAVI	Soil Adjusted Vegetation Index	SAVI = (NIR1-RED) *(1 + 0.5)/(RED+NIR1 + 0.5)	(Huete, 1988)
SATVI	Soil Adjusted Total Vegetation Index	(1 + 0.5) *(SWIR1-RED)/ (SWIR1 + RED+0.5)- SWIR2/2	(Marsett et al., 2006)
NDSVI	Normalized Difference Senescent Vegetation Index	(SWIR1 - RED)/(SWIR1 + RED)	(Qi et al., 2002)

extracted using the AppEEARS platform (https://lpdaac.usgs.gov/tools/appeears/). However, when prepared for spatial upscaling, the gridded data were processed in Google Earth Engine at a resolution of 5000-m where the value in each cell is the average value over all paddy rice covered area at 500-m resolution in that grid. We chose the 5000-m resolution for upscaling because it is still a relatively finer-scale resolution compared to previous flux machine learning (Peltola et al., 2019) or process-based modeling studies (Zhang et al., 2016) but at the same time can reduce spatial gaps in MODIS predictors through spatial aggregation. Gaps of time-series of MODIS images (<5%) at 5000-m resolution were filled using a simple linear interpolation method.

Other geospatial predictors are a set of 37 static variables relevant to bioclimate, soil, and topography from either modeled or assimilated reanalysis products. These variables include nineteen bioclimatic variables extracted from WorldClim 2.0 (Fick and Hijmans, 2017), two variables on total nitrogen and sulfur deposition (Lamarque et al., 2013), five soil variables from SoilGrids (Hengl et al., 2017), and 11 topographic variables from Earth Environment Topography (Amatulli et al., 2018) (See Supplementary Table 3 for more details). Static variables are more likely to cause spatial overfitting than dynamic variables (Meyer et al., 2018), therefore, we only considered them as ancillary predictors. When prepared for upscaling, any selected static variables were resampled to 5000 m resolution.

Additionally, four temporal variables associated with rice calendars were created to mimic a generic seasonal cycle and tested for predictive performance. The four variables are the number of 8-day intervals since the planting date (WOS_0), the number of 8-day intervals to the harvesting date (WOS_1), the length of the growing season (i.e., SLength: the number of 8-day intervals between the planting and harvesting date), and the sine function of the decimal growing season length (Sine (WOS_0/SLength)). If any of these variables are selected as important variables, the corresponding gridded map could be computed based on maps of the rice calendar.

2.3. Pretreatment of spatial representativeness

Spatial representativeness issues may arise when modeling with MODIS data, when a single image pixel is larger than the flux footprints and includes heterogeneous land covers that are different from the targeted land cover in the footprints (Chu et al., 2021). In such cases, modeling CH_4 flux with MODIS variables may not work well because the MODIS pixels do not directly represent the paddy rice area under EC measurements. Based on visual image check and site-based knowledge, we identified a few of our sites as being relatively more affected by this spatial misrepresentation and applied site specific processes to reduce the problem in training data. Specifically, we excluded pre-2010 data from US-TWT because the paddy rice fields covered a limited extent before 2010 (Knox et al., 2016). We used additional MODIS data from three neighboring pixels covered by similar rice cultivation (i.e., similar management and rice variety) to reduce the spectral contribution from open water at JP-MSE, because a large proportion of open water exists in the overlapping 500-m MODIS pixel. TARI1 measured a small paddy rice field within the overlapping MODIS pixel among other crops, so we averaged its MODIS index with that of a neighboring pixel that covers a large proportion of paddy rice similar to the one under measurement. A similar scale problem affected the site PH-RIF, but we failed to find an adjacent MODIS pixel that covers similar paddy rice fields. However, weekly LAI was sampled at PH-RIF. We then used Landsat-derived EVI and NDVI to establish a linear regression with measured LAI and then further used this relationship to compute 8-day EVI/NDVI for replacing MODIS EVI/NDVI. Other remote sensing indices for PH-RIF were still extracted from MODIS images directly. US-HRA and US-HRC were under alternate wetting and drying manipulations in 2015 and 2016, and 2016 respectively, which are different from adjacent paddy-rice applying traditional continuous flooding within the overlapping MODIS pixel. Therefore, data in 2015 and 2016 for US-HRA, and data in 2016 for US-HRC were excluded. US-BDA was under alternate wetting and drying manipulations in 2015 but adjacent paddy-rice in the overlapping MODIS pixel was applying continuous flooding, therefore we also excluded 2015 data from US-BDA. US-OF2 and US-OF5 were under traditional continuous flooding but a proportion of their neighboring fields are under alternate wetting and drying manipulation. This mismatch will potentially cause a low spatial representation of the overlapping MODIS pixel to the flux footprint. However, since there is no way to eliminate the issue, we used these two sites without any pretreatment of spatial representativeness. For more details on sites with spatial representativeness issues between MODIS pixels and EC flux footprints, see Appendix A in the supplementary.

2.4. Calendar and distribution of paddy rice

Maps of paddy rice were used to mask non-rice pixels prior to extracting predictors for upscaling methane fluxes. We used annual paddy rice distribution maps at 500-m resolution in Monsoon Asia from 2001 to 2015 that were produced using time series MODIS data and a phenology-based algorithm (Xiao et al., 2005; Zhang et al., 2017a; Zhang et al., 2020). The paddy rice maps were validated for both total areas using FAO statistical data and for pixel-level classification accuracy using higher resolution Landsat-based paddy rice maps (Zhang et al., 2020).

Maps of the rice calendar were used to remove predictions that fell outside of rice cultivation periods. Unlike natural wetlands, rice is usually cultivated for part of the year with a long winter fallow season and sometimes is rotated with other crops such as winter wheat. Our upscaling approach should not account for periods when other crops are planted beyond the rice growing season. Moreover, our training data consists primarily of flux measurements during the rice-growing season, as many sites only measure methane emissions after rice is planted and before rice is harvested. Therefore, in this study, we focus on emissions during rice growing periods, including short transition periods between

two or more crop seasons but excluding the longer winter fallow seasons, which can also generate CH_4 emissions especially when sites were flooded (Knox et al., 2016; Reba et al., 2019). RiceAtlas, which provides a spatial database of global paddy rice calendars (Laborte et al., 2017), was used to create binary masks for each 8-day interval during a year to tell whether each grid was in or out of rice cultivation periods. A few spatial gaps exist in RiceAtlas for India and China, which we filled using averages of neighboring spatial units with similar climates.

2.5. Machine learning model development

We used the random forest (RF) regression algorithm (Breiman, 2001) to produce upscaled flux predictions (Fig. 2). RF models consist of a large ensemble of regression trees where each tree is built by training it with a random subset of training data and predictors. The prediction of an RF model is the average of all the predictions made by individual regression trees in the forest, thus taking full advantage of ensemble means to decrease the noise of the prediction. RF has been widely adopted to predict CO_2, energy, and CH_4 fluxes, and shows similar performance when compared to other machine learning algorithms including neural network, Cubist, and Support Vector Machine (Irvin et al., 2021; Tramontana et al., 2016; Xu et al., 2018; Zhang et al., 2021a). RF models were developed using the 'caret' (Kuhn, 2008) and 'ranger' packages (Wright and Ziegler, 2015) in R including essential steps of cross validation, predictor selection, and hyper-parameter tuning.

2.5.1. Cross validation

A nested leave-one-cluster-out spatial cross validation scheme (LOCOC validation, hereafter) was applied during training for both predictor selection and hyper-parameter tuning. The 23 sites were grouped into 16 clusters of one or multiple sites. Sites within a 20-km distance were grouped as a cluster. One exception was for CN-HNY, which was grouped with CN-CMC despite their large distance because CN-HNY has few flux measurements but was under similar management as CN-CMC and has similarly high CH_4 emissions. In each round of the model development, 16 RF models were trained with data from one cluster held out for the purpose of independent testing and the

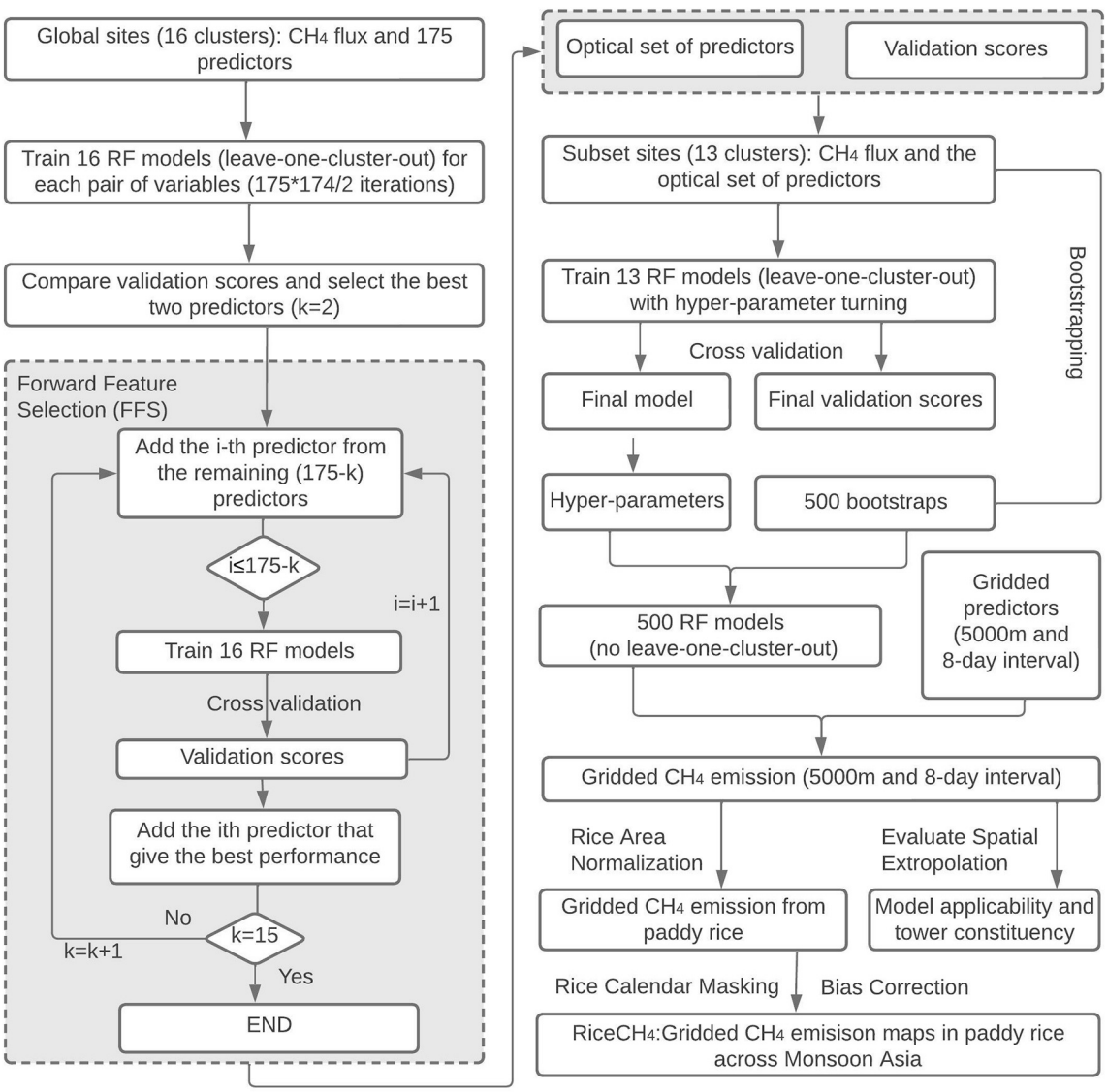

Fig. 2. Flowchart showing the processes from the processed CH_4 flux and predictor data to produce the upscaled gridded CH_4 emission maps (RiceCH_4) in paddy rice across Monsoon Asia, i.e., RiceCH_4 using random forest (RF). The cross validation applied here is the leave-one-cluster-out cross validation, as introduced in section 2.5.1. The gridded predictors based on MODIS remote sensing are aggregated into 5000-m resolution with non-paddy rice pixels masked out prior to aggregation. The rice area normalization means multiplication of the percentage of paddy rice in area for each grid. The cross validation scores from forward feature selection based on global sites were used to determine which sites out of Monsoon Asia were included in the final models for upscaling Monsoon Asia paddy rice.

remaining 15 clusters were used for 5-fold internal cross-validation. Folds were formed based on site clusters rather than random splits, i.e., all data from a given cluster falls either entirely within or out of a single fold. Model performance was evaluated by comparing predicted and observed CH_4 fluxes (from the held out cluster only) using the coefficient of determination (R^2) and mean absolute error (MAE). An instance of RF model (i.e., with prescribed predictors and hyperparameters) was regarded better if the average MAE score from all 16 hold-out tests based on 16 trained RF models was smaller than the alternative. Nash-Sutcliffe Efficiency (NSE) (Nash and Sutcliffe, 1970) was also computed as an integrative measure of model performance for the final selected model. NSE above zero corresponds to a model performance better than simply taking the average of the data while NSE below zero indicates that the model performance is worse than the mean of validation data. Performance was not only evaluated with respect to 8-day data, but also on-site mean flux (mean of all 8-day flux from all years as a proxy to annual mean flux, because not all sites have measured the whole growing season, and it is not possible to compute the real annual mean flux for validation). The ability to accurately predict spatial variability in annual site mean CH_4 fluxes is important, as it is the total or mean emissions that can help estimate paddy rice source contributions to the global CH_4 budget (Saunois et al., 2020).

2.5.2. Predictor selection

An optimal set of predictors used in the final upscaling model was identified using a forward feature selection (FFS) method (Meyer and Pebesma, 2021). FFS starts with the best one (or two) predictor(s) based on the performance metric (cost function) computed on validation data, then gradually adds one new variable at each step that maximizes the model performance - the one that gives the largest increase in model performance (minimizes cost function) on validation data by comparing all candidates. We started with the best pairs of predictors (i.e., LSTn and MaxLSTd, selected after comparing all pairs of candidate predictors) and followed with 15 further single FFS steps. More FFS steps are not needed as no improvement on performance metrics was observed after a few steps. Within each FFS iteration, random forest models were trained following LOCOC validation as explained in **2.5.1**. Variable selections were based first on the validated MAE score, then R^2 score, because squared error metrics (e.g., R^2) are more sensitive to outliers and highly skewed data, which are characteristic of CH_4 flux data (Morin, 2019). FFS requires considerable computing time because of its many iterations. To reduce the computation time of each FFS iteration, we adopted a set of fixed values of hyperparameters (the number of variables to possibly split at each node (*mtry*) is set to the square root of the total number of predictors at each step, and the minimum node size (*min.node.size*) is fixed to 5 as recommended for regression problems, the number of trees (*ntree*) is set to 100), rather than turning them with multiple tries.

2.5.3. Hyperparameter tuning

Hyperparameters including the number of trees (*ntree*), number of variables to possibly split at each node (*mtry*), and minimum node size (*min.node.size*) were tuned to further optimize random-forest models on the set of optimal predictors identified through FFS. The same LOCOC validation scheme was applied during hyperparameter turning as in FFS. A full hyperparameter grid-search (including *ntree*, *mtry*, and *min.node.size*) was performed, which allowed for trees of varying depth and complexity. The final hyperparameter values were determined based on performance scores through cross validation.

2.5.4. Bias correction

Machine learning (ML) regression models can suffer from 'regression to the mean', which results in over-prediction (positive bias) for small values and under-prediction (negative bias) for high values. Our model has a low mean bias (see 3.2 Model validation and performance) among training samples, but still over-predicts small values and underestimates high values. Multiple methods, ranging from a simple regression of observed on estimated values (Song, 2015) to a second machine learning model used to estimate residuals (Zhang and Lu, 2012) can be used to correct bias in regression based machine learning. In this study, we applied a linear equation based on the *Z*-score transform method (Belitz and Stackelberg, 2021) to correct bias based on prediction compared to observation during cross-validation. This method is simple and linear, yet the performance on bias-correction can approximate more complicated methods such as empirical distribution machining and machine learning based methods (Belitz and Stackelberg, 2021).

Given the original RF prediction for CH_4 flux (denoted by Y_{ML}) and the observed CH_4 flux (denoted by Y_{OBS}), the bias corrected prediction (denoted by Y_{ZZ}) is:

$$Y_{ZZ} = a^* Y_{ML} + b$$

where $a = \sqrt{var(Y_{OBS})}/\sqrt{var(Y_{ML})}$ and $b = E(Y_{OBS}) - a * E((Y_{ML})$. E is the expected value operator.

2.6. CH_4 flux upscaling

A final model for CH_4 flux upscaling was trained using the determined hyperparameters and optimal set of predictors using 16 sites (13 of the 16 clusters) in Monsoon Asia as well as IT-CAS and BR-CDS in Italy and Brazil respectively. The selection of sites in training the final model was made to maximize temporal-spatial coverage of training flux without significantly diminishing cross-validation accuracy scores. All sites in Monsoon Asia are included since that is the extent of our upscaling predictions. We then added other sites/clusters outside Monsoon Asia to increase temporal-spatial representativeness of training conditions if its testing R^2 score (when it is a hold-out sample) was less than one standard deviation away from the mean R^2 of all sites (see also **3.2 Model validation and performance**). Using this criterion, all US sites were excluded from the training data. Similar folded cross-validation was still applied based on clusters when training the final model, i.e., clusters are either entirely included or excluded during training and internal validation.

The uncertainty around the predicted fluxes was estimated by generating 500 predictions from 500 independent RF models trained using bootstrap samples from the available training data. This bootstrapping enables us to make 500 predictions for each grid cell and time step in the upscaled CH_4 flux map, which are then summarized into a standard deviation around the predicted mean flux.

The bootstrapped models were applied to 8-day time series of global grids for the final set of selected optimal predictor variables in the period 2001–2015 at a spatial resolution of 5000-m. All gridded predictors were prepared and processed consistently using MODIS sinusoidal projection at 5000-m resolution (note: all maps in this paper were reprojected to WGS-84 geographic coordinates). Finally, the upscaled methane fluxes were masked at each time step using both paddy rice distribution and calendar information to restrict upscaled flux to represent only flux emitted from paddy rice under cultivation (i.e., winter fallow seasons are excluded).

2.7. Model applicability and tower constituency

Cross validation evaluates model extrapolation performance but is ultimately limited by the environmental conditions captured by training data. Spatial upscaling applies model predictions to a much larger spatial (Stell et al., 2021) and temporal (Chu et al., 2017) domain, risking extrapolation beyond cross validation conditions which may reduce prediction accuracy. The sparsity of our training sites (16 sites) was complemented by long time-series data at each site as the dynamic conditions can represent different "space" (time for space), but it is still important to evaluate spatial representativeness of the training data with respect to the multivariate predictor space of the model at different

periods (Villarreal et al., 2018; Villarreal and Vargas, 2021). In this study, we evaluate the spatial representativeness of the training conditions by mapping 1) grid-based dissimilarity, and 2) tower constituency, as described below.

The grid-based dissimilarity score was defined as the minimum Euclidean distance between each grid cell to flux tower combinations (all 8-day data in the training sites) in predictor space, normalized by the mean distance among flux towers (Meyer and Pebesma, 2021). Distances were computed based on the eight selected optimal predictors that were firstly rescaled to between zero and one, and then weighted in proportion to their average variable importance in the random forest model. Dissimilarity score was evaluated at each 5000-m grid with non-zero paddy-rice area at an 8-day time-step. If a dissimilarity score is less than one, the new data point's distance to its nearest tower in predictor space is closer than the average distance among towers, suggesting a high chance of high-quality interpolation. If the dissimilarity score is greater than one, the difference to the nearest training data point is larger than the average distance between all training data pairs, suggesting high chances of low-quality extrapolation. We define good model applicability (low risk of extrapolation) for a grid if its dissimilarity score is less than the 95% percentile (~0.3) of the dissimilarity from training sites.

A tower's constituency is estimated as the geographical area that is most analogous in predictor space to the tower. To map the tower constituency, each 5000 m-grid was assigned as a constituent of the site that was closest in predictor space (Hargrove and Hoffman, 2004). The percentage of area coverage by each tower's constituency was then computed to identify the importance of the tower.

3. Results and discussion

3.1. Model predictors

We identified the 8 best predictors from the 175 total variables in the final model based on FFS, according to a combined consideration of reduction in MAE and increase in R^2 (Fig. 3). The final best eight predictors were: satellite-estimated land-surface temperature at night (LSTn), LSTn lagged with a one 8-day interval (LSTn_LAG1) and two 8-day intervals (LSTn_LAG2), annual maximum satellite-estimated land-surface temperature during daytime (MaxLSTd), annual maximum crop residual cover index (MaxCRC), available soil water-holding capacity (ASWC), simple ratio water index lagged by one 8-day interval (SRWI_LAG1), and soil adjusted total vegetation index lagged by three 8-day intervals (SATVI_LAG3).

Four of the final eight predictors were temperature-related and were the highest in importance score (Fig. 3), thus temperature and its lagged terms are considered to be the dominant factors in our model for predicting CH_4 emissions. Temperature as the single most important predictor of CH_4 emissions is consistent with the results of various global syntheses and regional data-driven upscaling based on multiple sites in natural wetlands (Peltola et al., 2019), as well as in predictions of temporal dynamics in single site-based studies (Dai et al., 2019; Knox et al., 2016; Li et al., 2019; Xiao et al., 2017). Night-time surface temperature was also selected as a more important predictor than daytime surface temperature, probably because nighttime surface temperature is more correlated to soil temperature (Huang et al., 2020) which influences the microbial processes controlling CH_4 production and oxidation and subsequent soil diffusion and ebullition (Knox et al., 2019, 2021).

Biomass-related vegetation indices were also included in the final predictive model. Substrate availability influences CH_4 production potential because it fuels methanogenesis (Delwiche et al., 2021; Knox et al., 2021; Sturtevant et al., 2016; Xiao et al., 2017). Thus, gross primary productivity (GPP) is often found to be an important variable controlling substrate availability as a proxy for recent organic carbon supply to the soil. Although we were unable to include GPP directly in our initial exploratory datasets because of a lack of GPP measurements at two of our sites and the lack of gridded GPP products for paddy rice, in particular, we did include many vegetation indices that are proxies of GPP or NPP, such as EVI and NDVI, CRC, SATVI, LSWI. The model selected at least two greenness/biomass related indices: MaxCRC and SATVI_LAG3. SATVI instead of the commonly used EVI or NDVI was likely selected because of the better ability of soil-adjusted vegetation indices to minimize soil influences on canopy spectra (Huete, 1988; Qi

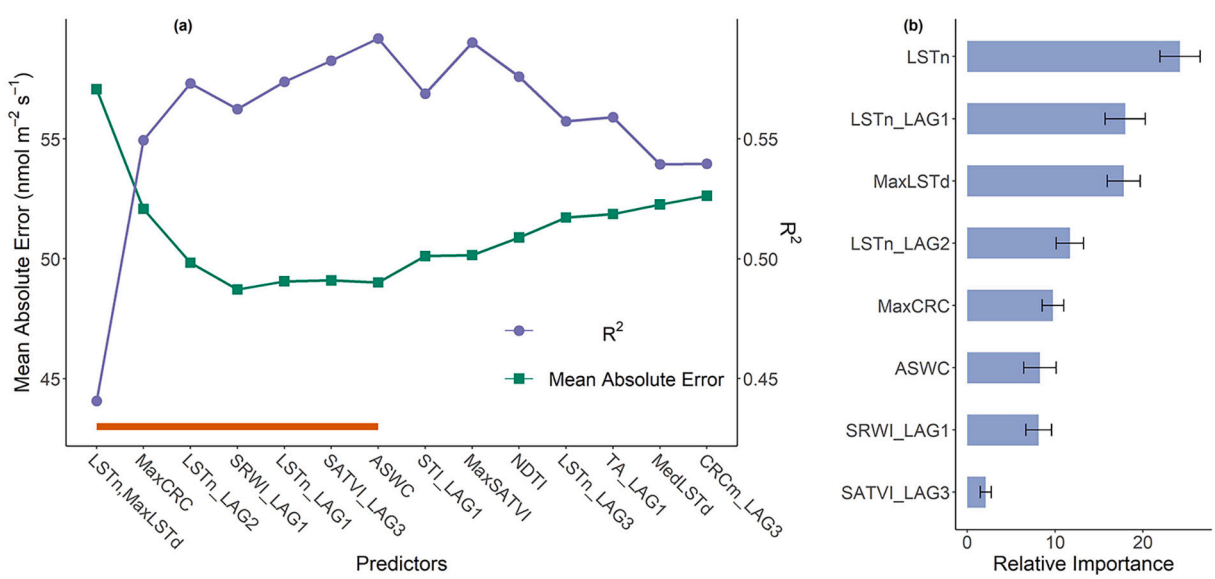

Fig. 3. (a) Evolution of model cost function (mean absolute error (MAE) and R^2) during forward feature selection. The best-performing feature pair was added first (i.e., LSTn and MaxLSTd), then a single additional predictor was added in each forward step from the remaining predictors if it reduces most (or increase least) of the MAE. The horizontal bar (orange) length encompasses the 8 final model predictors, after which MAE begins to increase and R^2 begins to decline. (b)Variable importance ranked using the permutation importance method. LSTn: Land surface temperature during night; LSTn_LAG1:LSTn lagged with one 8-day interval; LSTn_LAG2: LSTn lagged two 8-day intervals; MaxLSTd: annual maximum satellite measure land surface temperature during daytime; MaxCRC: annual maximum crop residual cover index; ASWC: available soil water holding capacity; SRWI_LAG1: simple ratio water index lagged by one 8-day interval; SATVI_LAT3: soil adjusted total vegetation index lagged by three 8-day intervals. The results presented here are based on cross-validation on 16 clusters (23 sites) globally.

et al., 1994); this effect is important for short vegetation such as rice, especially during early growing stages when soil is visible from above the canopy. Residual biomass or brown litter can provide organic substrate to methane producers, too. Crop residues such as straw and stubble are commonly left in some paddy rice fields after harvest and can stimulate methane emissions (Liou et al., 2003; Yan et al., 2005). The crop residual index (CRC) was designed to differentiate soil and crop residues but also reflect green vegetation signals during the growing season. Thus, the annual maximum of the crop residual index (MaxCRC) likely cannot capture the amount of crop residues left in the field after harvest, but rather suggests the peak value of combined brown and green biomass/vegetation. Additionally, the water-relevant index SRWI can also reflect the amount of green biomass during the growing season as leaf thickness and leaf area affect the ability of satellites to see water underneath the canopy (Zarco-Tejada and Ustin, 2001).

Water regimes are also widely recognized as regulators of methane emissions in wetlands and paddy rice (Knox et al., 2021; Runkle et al., 2019; Yan et al., 2005). However, without full information on water-table dynamics and irrigation practices at the gridded scale, we could not incorporate them into our predictive model for spatial upscaling. Nevertheless, a water index should capture surface water conditions affected by water management. The model selected SRWI from various water indices, which was also adopted in previous research for wetland CH_4 emission upscaling (Peltola et al., 2019). SRWI has been shown to be correlated with wetland water-table depth and thus reflects surface water dynamics (Meingast et al., 2014). Moreover, the soil water-holding capacity (ASWC) (Hengl et al., 2017) quantifying the static amount of water that the soil can hold was selected as a second variable related to the water regime.

We found that relationships between the selected predictors and CH_4 emission are non-linear (Supplementary Fig. 1). A partial exponential dependence of CH_4 was observed against LSTn, LSTn_LAG1, and LSTn_LAG2, reconfirming the dominance of temperature on CH_4 fluxes. The partial dependence of CH_4 on other predictors is more complicated (Supplementary Fig. 1), suggesting the overlapping effects of multiple mechanisms of CH_4 production, consumption and transport. Furthermore, the model selected lagged terms of temperature and vegetation and water indices, suggesting lagged effects on methane emissions, which have also been observed previously (Chang et al., 2021; Knox et al., 2021).

3.2. Model validation and performance

The model achieved a moderate to high (>0.4) hold-out test score of R^2 for most clusters, with a mean R^2 of 0.48 ± 0.26 (mean ± standard deviation) among all 16 clusters (Fig. 4). The clusters with lower scores included US-HRA, US-HRC, US-OF2, US-OF5, US-BDA, US-BDC, and US-TWT, and PH-RIF (i.e., cluster 10). Many of these sites with low validation scores have low spatial representativeness of flux footprints by the MODIS pixels (Supplementary Appendix A, Figs. A1-A7), because the overlapping MODIS pixels cover additional crops and management practices. On the other hand, clusters of sites showing good testing

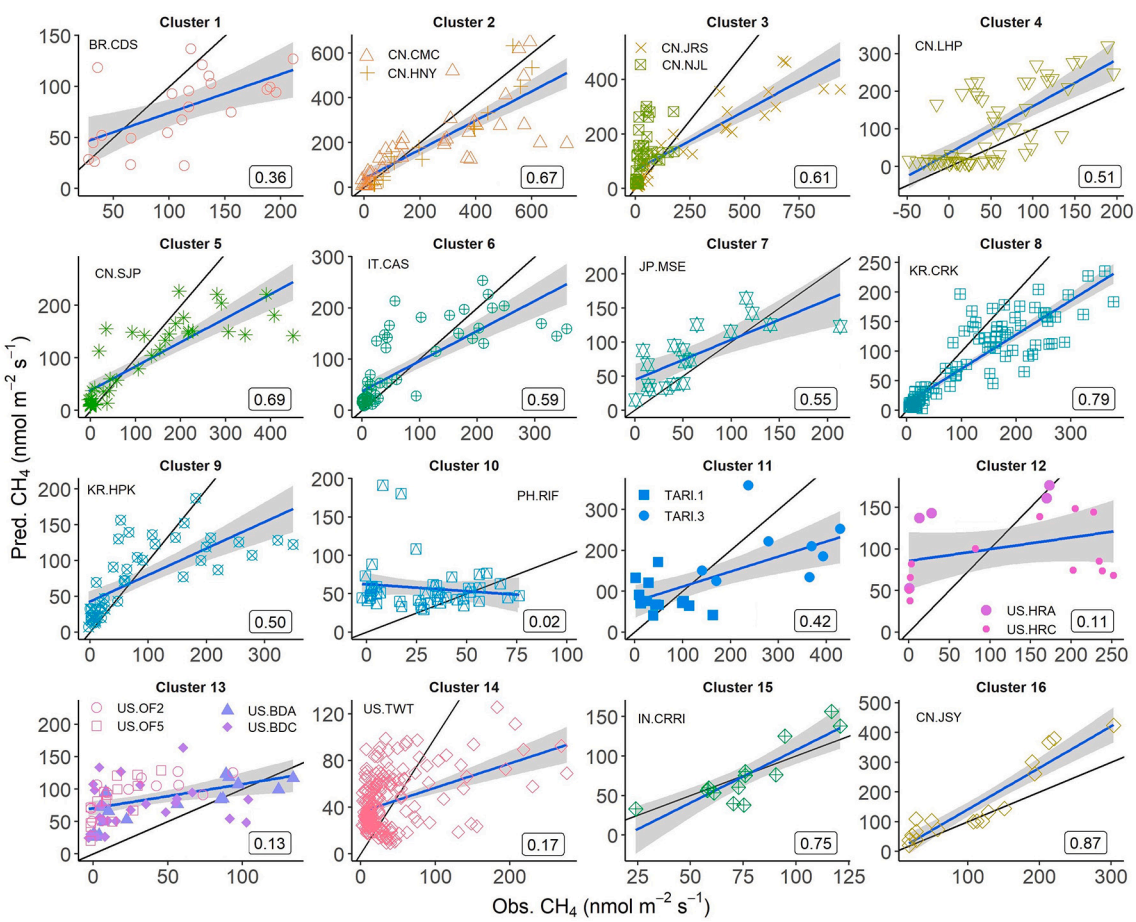

Fig. 4. Independent test score of R^2 (coefficient of determination) on each single hold-out cluster data when the model was trained on all other 15 clusters using leave-one-cluster-out cross validation during model development using all 16 clusters of sites that are distributed globally. For clusters containing multiple sites (2, 3, 11, 12 and 13), the shape and color of points depict individual sites. Clusters 11, 12, and 13 containing all sites in the United States were excluded before training the final upscaling model because they are outside Monsoon Asia and have low cross-validation scores. Note changes in axis scaling among sub-plots; also note one-to-one line in black relative to the line of best fit with gray confidence intervals.

scores generally had good spatial representativeness of the MODIS pixel to the overlapping flux footprints (e.g., CN-LHP shown in Supplementary Appendix A, Fig. A8).

The performance of our model during cross validation was comparable to or even better than recent efforts upscaling wetland CH_4 fluxes with similar data-driven approaches trained on EC measurements (Peltola et al., 2019). When pooling together all 16 hold-out clusters, the model can predict 8-day CH_4 fluxes with a MAE of 49.7 nmol m^{-2} s^{-1}, NSE of 0.57, and R^2 of 0.58 (Fig. 5(a)), which is slightly higher than the recent Northern latitude upscaling on wetland CH_4 flux based on 25 eddy covariance towers (Peltola et al., 2019) and global upscaling of CO_2 flux (Tramontana et al., 2016). However, our performance did not reach the accuracy achieved in global upscaling of gross primary productivity ($R^2 > 0.7$) and ecosystem respiration ($R^2 > 0.6$) (Jung et al., 2020; e.g., Tramontana et al., 2016). This may be because CH_4 fluxes are more variable and episodic than CO_2 fluxes and because CH_4 fluxes are less seasonally predictable due to non-linear regulation from biophysical variables with a greater influence of lagged effects (Chang et al., 2021; Knox et al., 2021). The importance of including lagged effects to improve the prediction of CH_4 flux was demonstrated by our model including multiple lagged predictors during feature selection. The model also captured geographic differences better than seasonal differences, as the cross validation achieved an NSE of 0.63 and R^2 of 0.69 for site-mean flux (Fig. 5(b)), higher than those same metrics for 8-day flux predictions (0.57 and 0.58 respectively).

Because our target study area is Monsoon Asia, we wanted to evaluate the tradeoff between including low-performance sites outside of this region to increase temporal-spatial representativeness and adversely impacting the model performance of our final upscaling RF model. We excluded a site outside of Monsoon Asia if its testing R^2 score (when it is a hold-out sample as shown in Fig. 4) was more than one standard deviation lower than the mean R^2 of all sites (i.e., <0.22). Applying this criterion, all sites in the United States were excluded, i.e., sites in clusters 12, 13, and 14 as shown in Fig. 4. Cross-validation scores show that excluding the low performance sites in the United States improved the overall model performance (Fig. 5), with the NSE increasing from 0.57 to 0.59 for 8-day fluxes and from 0.63 to 0.69 for site-mean fluxes, and nMAE decreasing from 0.6 to 0.53 for 8-day fluxes and from 0.41 to 0.31 for site mean fluxes.

3.3. Temporal dynamics of CH_4 emissions across monsoon Asia

Across all the paddy rice in Monsoon Asia, our upscaling model depicted notable seasonal dynamics and annual trends for total CH_4 emissions (Fig. 6). Seasonal CH_4 emissions peaked in late July and early August (the 27-29th 8-day interval) (Fig. 6a), driven by both the highest temperatures of the year and the largest paddy rice area under cultivation in Asia (Fig. 6a) (Laborte et al., 2017).

Estimated annual total CH_4 emissions in Monsoon Asia increased from 2001 to 2007, then declined from 2007 through 2015, a result consistent with the bilinear trends in paddy-rice cultivation area (Fig. 6 (b)). Although a decline in planted paddy rice area apparently

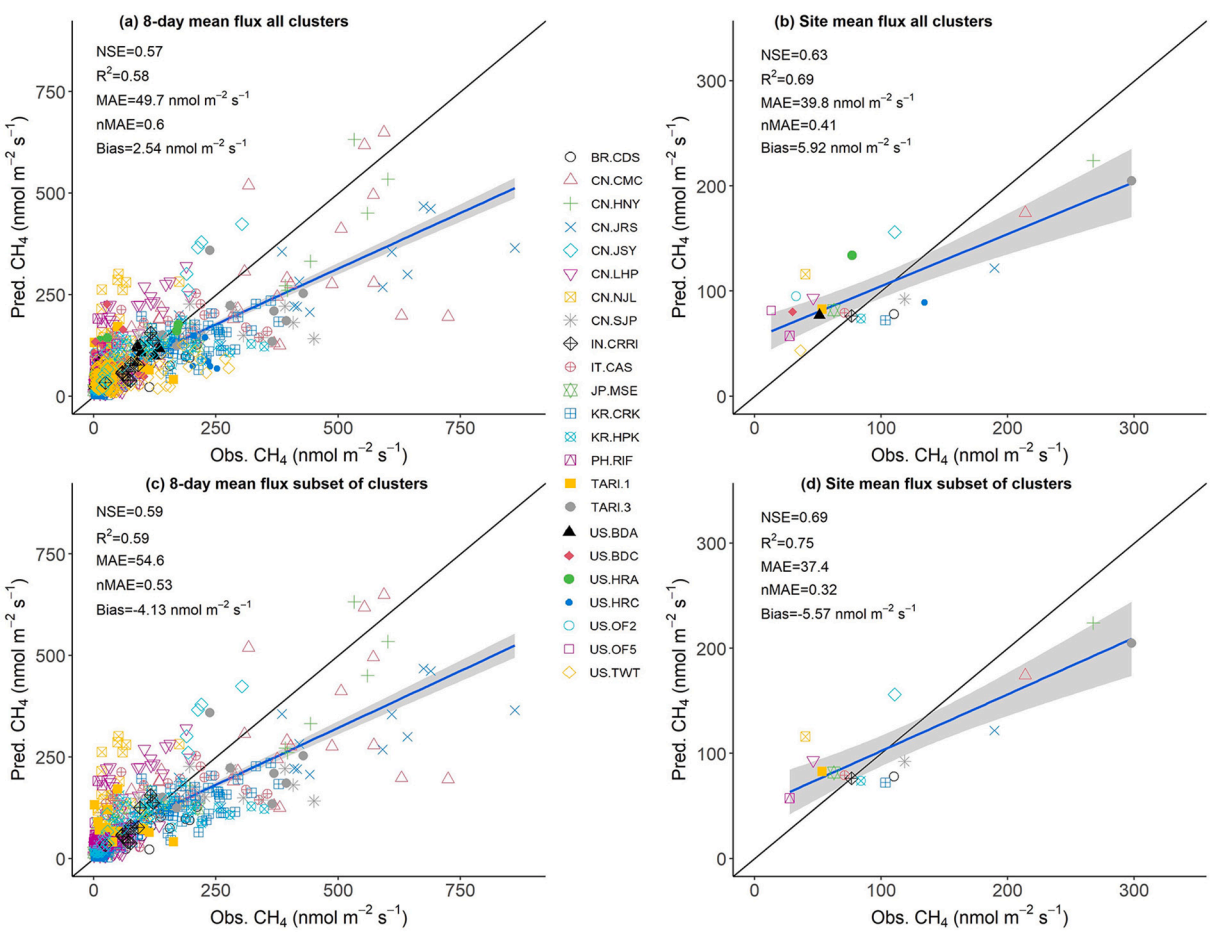

Fig. 5. Model predicted (Pred.) CH_4 flux versus observed (Obs.) CH_4 flux on all hold-out data for (a) 8-day CH_4 flux, and (b): site mean CH_4 flux (both including all original 16 clusters from 23 sites) (c) 8-day CH_4 flux and (d): site mean CH_4 flux (both excluding US-TWT, US-BDA,US-BDC, US-OF2, US-OF5, US-HRA, and US-HRC sites). NSE, Nash-Sutcliffe Efficiency (NSE); R^2: coefficient of determination; MAE: mean absolute error; nMAE: mean absolute error normalized by the average observed CH_4 flux; Bias: the average predicted CH_4 flux minus the average observed CH_4 flux. The black line shows the 1:1 line. Note this plot shows cross-validation results without bias corrections.

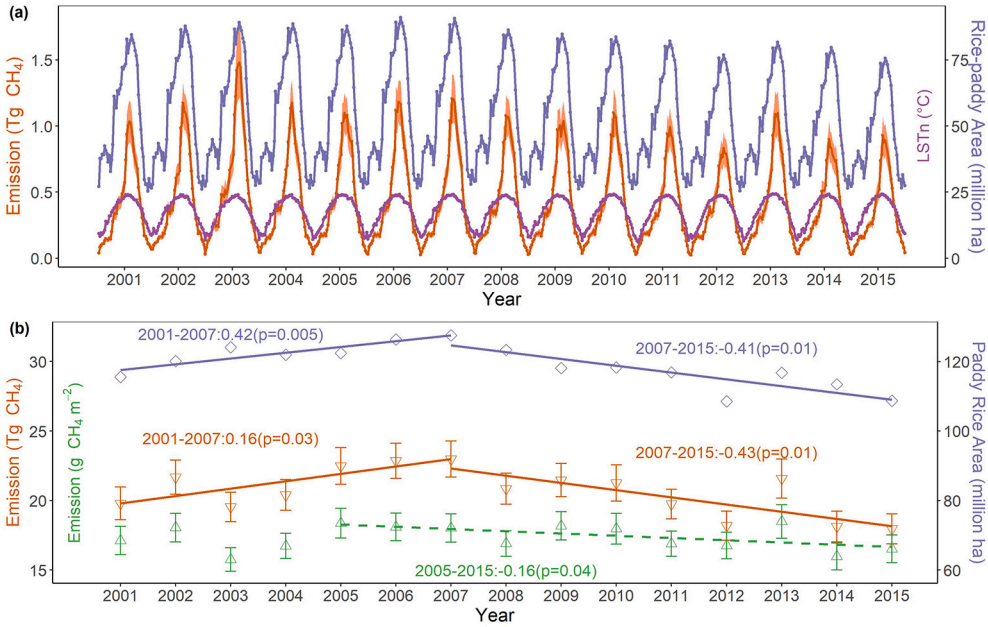

Fig. 6. (a) The 8-day emissions of CH_4 from all paddy rice in Monsoon Asia, overlaid with the paddy rice area under cultivation and the average land surface temperature at night across all paddy rice. (b) The annual total and area normalized emission of CH_4 (error bars show standard deviation based on 500 bootstraps), and paddy rice area. The temporal trends in periods of 2001–2007 and 2007–2015 were drawn for paddy-rice area and total emissions, and a trend for emission rates (area-normalized emissions) was drawn for the period 2005–2015. The linear trends (i.e., slope) and p values are labeled for each period. Error bars of emissions represent standard deviation based on 500 bootstrapped upscaling. Note in (a) the label of year on the x-axis is positioned in the middle of the year rather than at the beginning of the calendar year.

dominated the cause of a decline in CH_4 emission after 2007 in our estimate, it was not the sole causing factor. The emission rate (emission per unit of area) computed based on our estimates also declined from 2005 to 2015 (Fig. 6(b)), suggesting effects from other factors, such as variations in climate and agronomic management. Different trends are observable before and after 2007, with regional CH_4 emissions increasing from 2001 through 2007 and then declining from 2007 to 2015 (Fig. 6b). The renewed increase in atmospheric CH_4 concentration since 2007 has little consensus as to its cause (Allen, 2016; Nisbet et al., 2014, 2016; Schaefer et al., 2016; Schwietzke et al., 2017; Turner et al., 2019). Some studies suggest that biogenic sources from agriculture may be the key contributor to this renewed growth in atmospheric CH_4 (Nisbet et al., 2016; Schaefer et al., 2016). Paddy rice is one of the main agricultural sources of CH_4 and Monsoon Asia grows ~87% of global

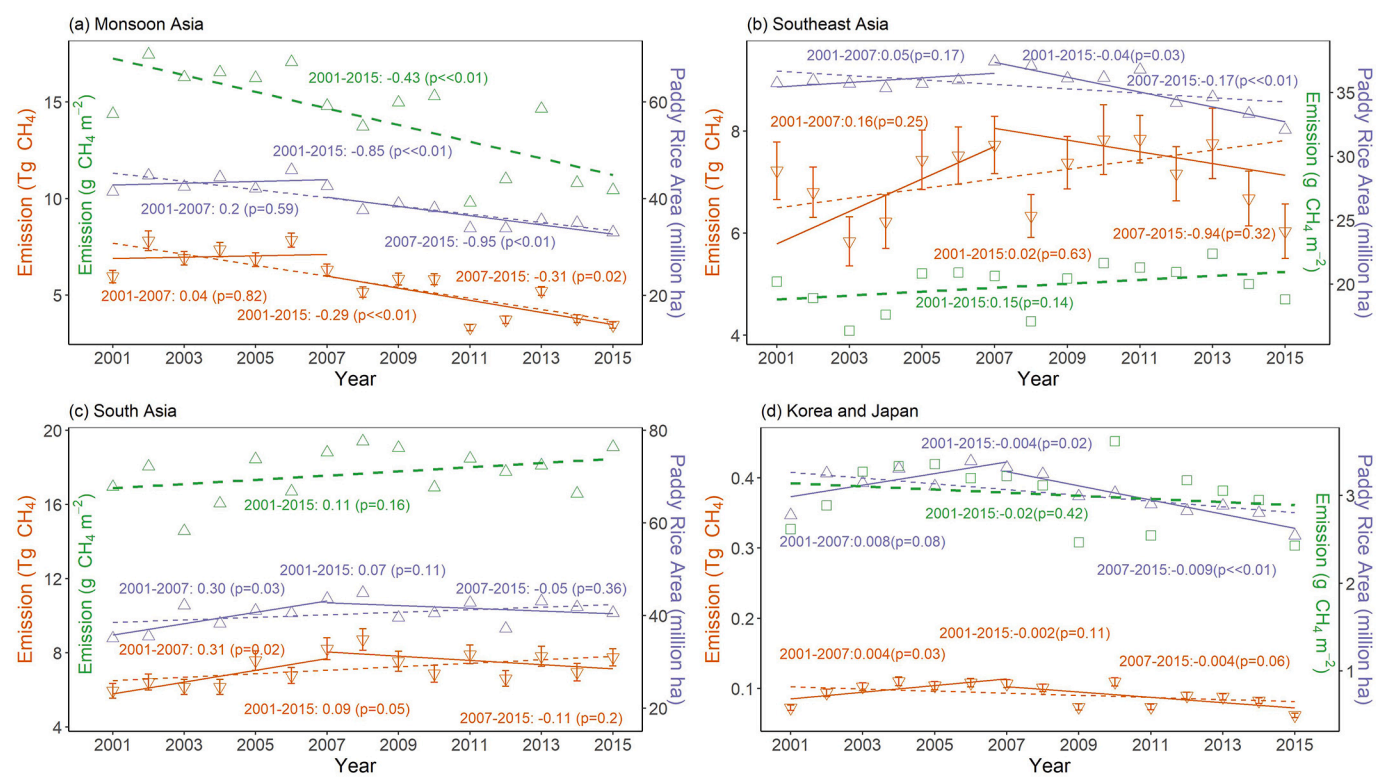

Fig. 7. Annual total CH_4 emissions (Tg CH_4), emission rate (i.e., area-normalized emission (g CH_4 m^{-2})), and paddy rice area (million ha) in (a) China, (b) Southeast Asia, (c) South Asia, and (d) Korea and Japan. Trends in 2001–2007, 2007–2015, and 2001–2015 are drawn for total emission and paddy rice area, while trends in 2001–2015 are drawn for normalized emission rate. The linear trends (i.e., slope) and p values are labeled for each period. Error bars represent standard deviation based on 500 bootstrapped upscaling.

paddy rice. Our estimated decrease in CH_4 emissions from paddy rice in Monsoon Asia since 2007 through 2015 would support the notion that rice paddies are likely not contributing substantially to the regrowth in atmospheric CH_4 concentrations during the same period.

Contrasting trends in annual CH_4 emissions exist among the four subregions of Monsoon Asia: China, South Asia, Southeast Asia, and Korea-Japan. China and Korea-Japan show similar declining trends after 2007, and they contributed most to the total methane emission reduction in Monsoon Asia after 2007 (Fig. 7). A declining trend of total cultivated paddy rice in both China and Korea-Japan after 2007 was apparently one of the main causes of reduced emissions. However, we also estimated a declining trend of emission rate after 2005 (0.2 g CH_4 m^{-2} yr^{-1})), which is likely related to the northeastward shift of paddy rice area in China during recent years (Wang and Hijmans, 2019; Zhang et al., 2017a) and the increasing use of intermittent irrigation schemes in both China and Korea and Japan (Bo et al., 2022; Xu et al., 2016; Zhang et al., 2017b; Hwang et al., 2020). Methane emissions in South Asia increased before 2007 but showed no trend after 2007, same as the paddy rice area over South Asia that increased before 2007 but then remained relatively stable after 2007. This finding is consistent with an earlier report showing little year-to-year variability of rice methane emissions during 2010–2015 in India (Ganesan et al., 2017), the largest rice producer in South Asia. Southeast Asia shows no significant trends of total emissions before or after 2007. While its total area of paddy rice declined after 2007, the average annual total emission after 2007 is higher than that before 2007.

3.4. Spatial patterns of emissions

The spatial pattern of CH_4 emissions closely matches the spatial pattern of paddy rice area distributions (Fig. 8). This is not surprising as the existence of paddy rice is the main driver of CH_4 emissions at large spatial scales, though climate, soil, and agro-management can be important local drivers of emissions. The most prominent hotspot of CH_4 emission is the Ganges River Delta in India and Bangladesh (~20% of the total emission in Monsoon Asia by our estimate), which is consistent with results from previous process-based modeling (Zhang et al., 2016) where the highest emission rates were observed in the same region. The Ganges River Delta region uses intensive paddy rice cultivation with two to three rice seasons as well as high temperatures, which can lead to high annual CH_4 emissions. The highest CH_4 emission rates per unit area were also found in Ganges River Delta in process-based modeling. The Mekong River Delta in Southeast Asia, the Red River Delta in Vietnam, the Ayeyarwady Delta in southern Myanmar, and the Yangtze River Delta in southern China were also hotspots of CH_4 emissions (Fig. 8, see Supplementary Fig. 2 for the locations of these deltas). These coastal alluvial delta plains not only have high temperatures but also dense paddy rice areas. Among these hotspots, the Yangtze River Delta in China was predicted to have lower than average CH_4 emissions probably due to lower temperatures, fewer rice cultivation seasons within the year, and more widely adopted intermittent irrigation scheme, though this delta area has similar dense rice areas as the other major tropical deltas. Aside from these alluvial plains, the Indo-Gangetic Plain in northwest India, the Chengdu plain in central China, and northeast

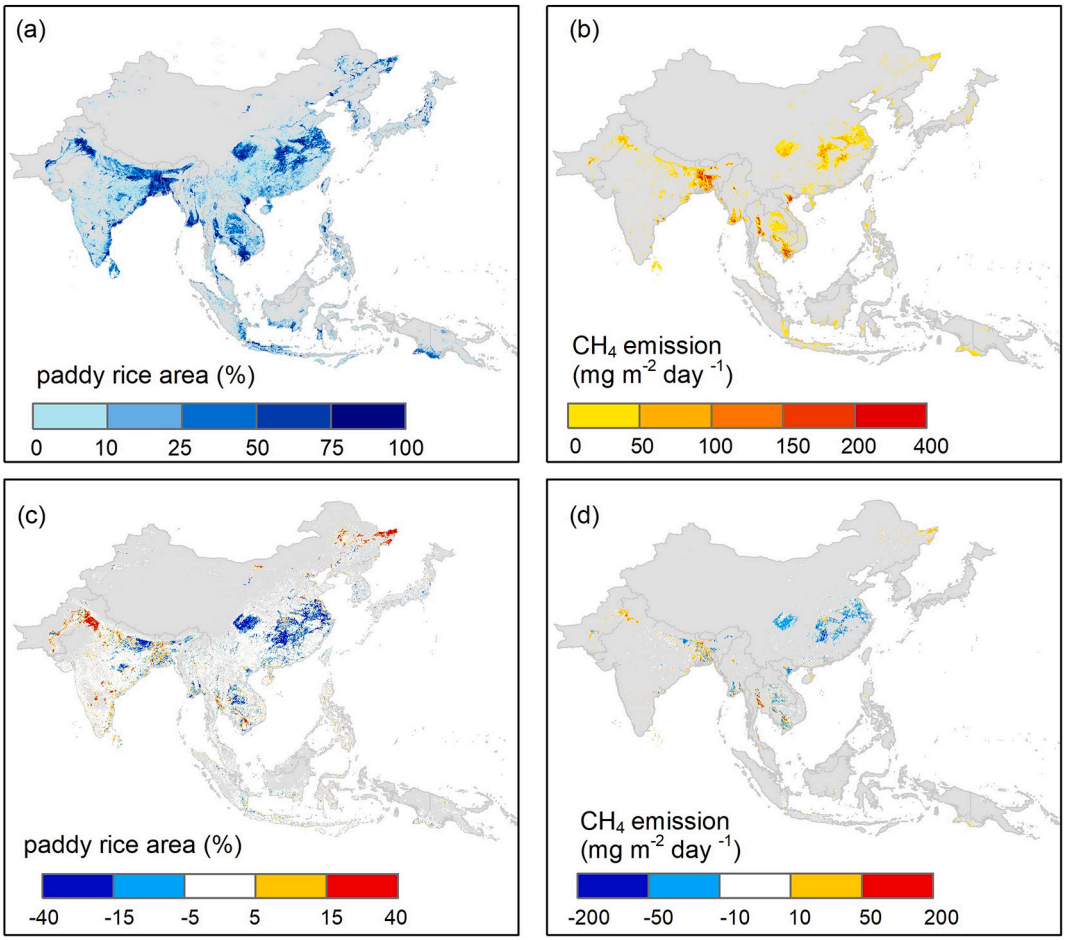

Fig. 8. The spatial distribution at 5000-m resolution of (a) the multi-year (2001–2015) annual average percentage of paddy rice area from 2001 to 2015, (b) the multi-year annual (2001–2015) average CH_4 emissions in paddy rice area. (c) The difference of average annual paddy rice area between 2008 and 2015 and 2001–2007, and (d) the difference of average annual methane emissions in paddy rice between 2008 and 2015 and 2001–2007.

China also appear as hotspots of CH_4 emissions due to intensive rice cultivation there.

There was a substantial decline of paddy rice area after 2007 compared with before 2007 in Yangtze Plain of southern China, Chengdu plain of Central China, and eastern Thailand, and accordingly a decrease of methane emissions (Fig. 8). Changes of paddy rice densities, however, do not necessarily lead to the same change of methane emissions, suggesting again multiple regulating factors on methane emissions beyond rice area alone. Decline of paddy rice area also occurred in the northeast of India where neighbors Nepal and the east of Thailand and correspondingly a decrease of methane emissions. The northeast of China (mainly in Heilongjiang Province), on the other hand, has observed increased paddy rice area and estimated methane emissions. However, despite an increase in paddy rice area observed for northwest India, modeled methane emissions did not increase proportionally to the increase of paddy rice area. This may be attributable to

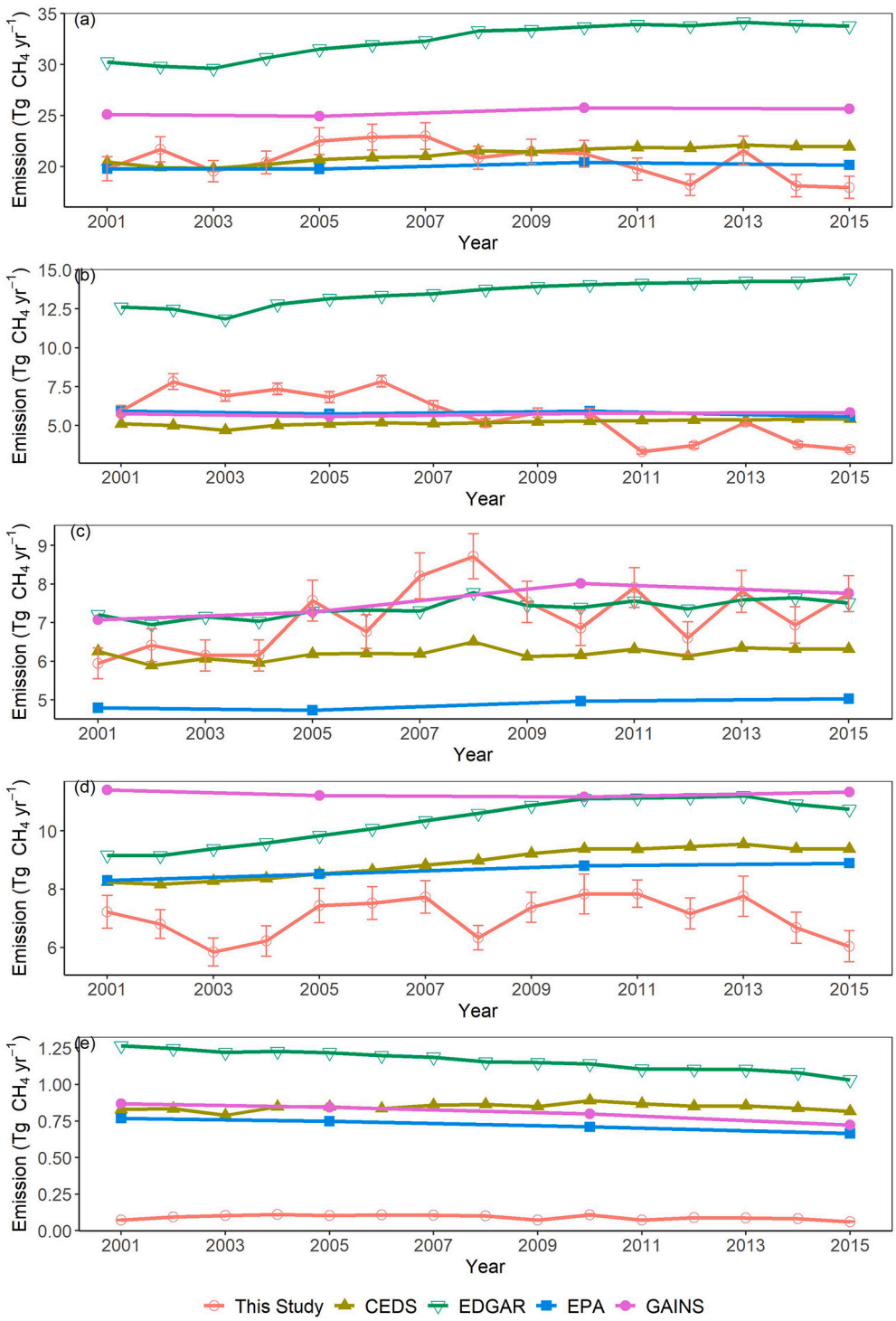

Fig. 9. Comparison of total paddy rice CH_4 emissions estimated in this study to four major inventory-based estimates for (a) the entire Monsoon Asia, and for four sub-regions in (b) China, (c) South Asia, (d) Southeast Asia, and (e) Korea-Japan. Error bars shown for our study represent standard deviation based on 500 bootstrapped upscaling. The inventory versions used were CEDS (v2021_04_21), EDGAR (v60), and GAINS (GAINS-Eclipse6).

the decline of the water table in the region that forced more water-saving technologies in rice field irrigation in recent years (Humphreys et al., 2010). In the Chao Phraya River Basin of Thailand, we observed no obvious increase of paddy rice area but, nevertheless, a substantial increase of CH_4 emissions. Further investigation is needed to understand the cause of this estimated increase of CH_4 emissions. It may be a combined effect from water management and climate changes, i.e., continuous flooding is still common in this region (Maraseni et al., 2018) and climate warming can lead to increase of emission under continuous flooding (Minamikawa et al., 2016).

3.5. Comparison with inventories

Our estimate of total CH_4 emissions associated with rice cultivation for Monsoon Asia is in the lower range of the inventory-based estimates, although our observed trends pre- and post-2007 diverge from those from inventories (i.e., CEDS and EDGAR) with estimations lower than all the inventories in the recent years (Fig. 9). Our average annual CH_4 emissions from paddy rice in monsoon Asia is 20.6 ± 1.1 Tg yr^{-1} for 2001–2015, which falls at the low end of the average annual emission range (20–32 Tg yr^{-1}) from four major inventory-based estimates (i.e., EPA, CEDS, GAINS, and EDGAR). Within this period, the highest annual emissions of our estimate occurred in 2007 at 23.0 ± 1.3Tg yr^{-1} and the lowest emissions occurred in 2015 at 17.9 ± 1.1 Tg yr^{-1}. Among the inventories, the estimates from EPA and CEDS are very close to ours, but GAINS and EDGAR's estimates are ~20–50% larger. Our finding that emissions increased until 2007 and then declined afterward contrasts strongly with both CEDS and EDGAR inventories showing an increasing trend CH_4 emissions since 2005 and 2003 respectively to 2015. EPA and GAINS only report emissions every 5 years, and they both show a slight emission decline from 2010 to 2015 (1% and 0.4% respectively), while our estimates show a 15% decline from 2010 to 2015.

The difference between our estimates of total CH_4 emission and the inventory-based estimate can be attributed to multiple reasons. Firstly, our low estimate of total emissions may partly derive from the lower paddy-rice area we use from MODIS relative to the FAO-based rice paddy area (Zhang et al., 2020) that was used by all other inventories. Two reasons may explain the underestimation of MODIS paddy-rice area. One reason is that the approach used to produce MODIS paddy-rice was based on detecting flooding signals (Zhang et al., 2020), which may miss some rain-fed paddy rice that are included in FAO statistics. The other reason is that the 500 m resolution of MODIS data may miss the detection of some small paddy rice fields. Secondly, the inventories adopted a large variation of CH_4 emission rates from rice paddies in different regions and different management conditions (e.g., organic amendment and irrigation scheme), which largely impacted the range of estimates of CH_4 emissions among themselves. The higher emission estimates for GAINS and EDGAR can be partly explained by their higher emission rates used for certain regions (Höglund-Isaksson, 2005; Janssens-Maenhout et al., 2019; Peng et al., 2016). Moreover, the inventory estimates adopted constant emission factors for each management type in a region that may fail to account for the declining trend due to growing adoption of water-drainage and drawdown practices over time in East Asia (Bo et al., 2022; Xu et al., 2016; Zhang et al., 2017b), due to outdated or inaccurate information of the area of paddy rice under different management. Using a constant emission rate for the same management type in a region also ignores the variation of emission rates caused by heterogeneous local environmental factors. In contrast, our model estimated grid-based emission rates completely based on local variables, which addressed at least partially the spatial variation of emission rates within a region. Although our model does not directly consider management types, the management practices such as organic amendments or water irrigation schemes might be partly captured in our model through their effects on surface temperature, vegetation index, and water index used as input.

At the scale of our four subregions (as shown in Fig. 1) the results are different among areas. Our estimate is in the range of the majority of the inventories in China and South Asia while it is lower than all the inventories in Southeast Asia and Korea and Japan. In this last case however, the magnitude of the emission is very low and probably inside the overall uncertainties. In terms of trends, the decline we estimated in China in the second part of the period was reported only by the EPA and GAINS inventories but at a very low rate. However, the stabilization of the emissions in South Asia and the declining trend we estimated in Korea and Japan after 2010 is consistent across all the inventories and our estimates. Inter-annual variability was much higher in our estimates than in the inventories, likely because our approach modeled high resolution variations in paddy-rice area and emission rates both in time and space.

Our results suggest that methane emissions in monsoon Asia have not only declined during 2001–2015 but may be also lower than previously thought. Since the inventories considered herein are largely based on IPCC Tier 1/2 methods, these varying discrepancies to our estimates reflect the large uncertainties for using different approaches on determining region-specific emission rates, rice cultivated area, and season length.

3.6. Future directions to improve data-driven upscaling

There is usually a geographic bias in spatial coverage of training flux sites (e.g., Papale et al., 2015), therefore increasing new measurements in under-measured areas is desirable to improve future data-driven upscaling. To assess the impact of such geographic bias on the model's ability to extrapolate beyond training conditions, we evaluated the spatial representativeness of the training sites with respect to the multivariate predictor space (Villarreal et al., 2018; Villarreal and Vargas, 2021) using dissimilarity and tower constituency (see 2.6 in Data and Methods). The dissimilarity map shows that fortunately, most of the study area (>97%) has good model applicability (i.e., low chances of extrapolation beyond training conditions) across different times of the year (Fig. 10). However, most of our training sites are located in the temperate zone, while only one site (PH-RIF) is located in the tropical zone. We observed that while IN-CRRI tends to dominate the constituency map in tropical South and Southeast Asia during both summer and winter, PH-RIF can have a significant share of the tower constituency space during spring and winter (Fig. 11). The low cross-validation score of 8-day samples at PH-RIF hints at possible model inaccuracy over tropical South and Southeast Asia for seasonal dynamics. Therefore, the under-measured areas of tropical South and Southeast Asia are prime locations for expansion of EC measurement of CH_4 flux in paddy rice to improve future upscaling efforts.

Our model development processes suggest that it is important to consider the spatial representativeness of EC flux footprints to remote sensing data in training datasets (Chu et al., 2021; Kong et al., 2022). Towers in paddy rice are usually a few meters (3-6 m) in height with a dynamic flux footprint area (a few hundred m^2) smaller than the area of the overlapping MODIS pixel (2500 m^2). Therefore, when the MODIS pixel covers extra different land types from the measured paddy rice within EC footprints, it introduces dissimilarity of biophysical properties between the two that hurts the model performance. Therefore, until finer resolution remote sensing products are available, where possible we recommend new towers to be located in the middle of large and homogeneous paddy rice areas for assisting modeling with coarse resolution remote sensing data. The heterogeneity around existing towers will be better accommodated in the future by leveraging the new generation of remote sensing products at both very high spatial and temporal resolution (e.g., 1-10 m resolution with daily or weekly revisit) to spatially match the boundaries of flux footprints more closely (Fisher et al., 2020; Kim et al., 2006; Yang et al., 2020). Future advances in the constellation of Cubesats (e.g., Planet) and data fusion techniques (e.g., MODIS-Landsat fusion, MODIS-Sentinel Fusion) are particularly promising in this regard.

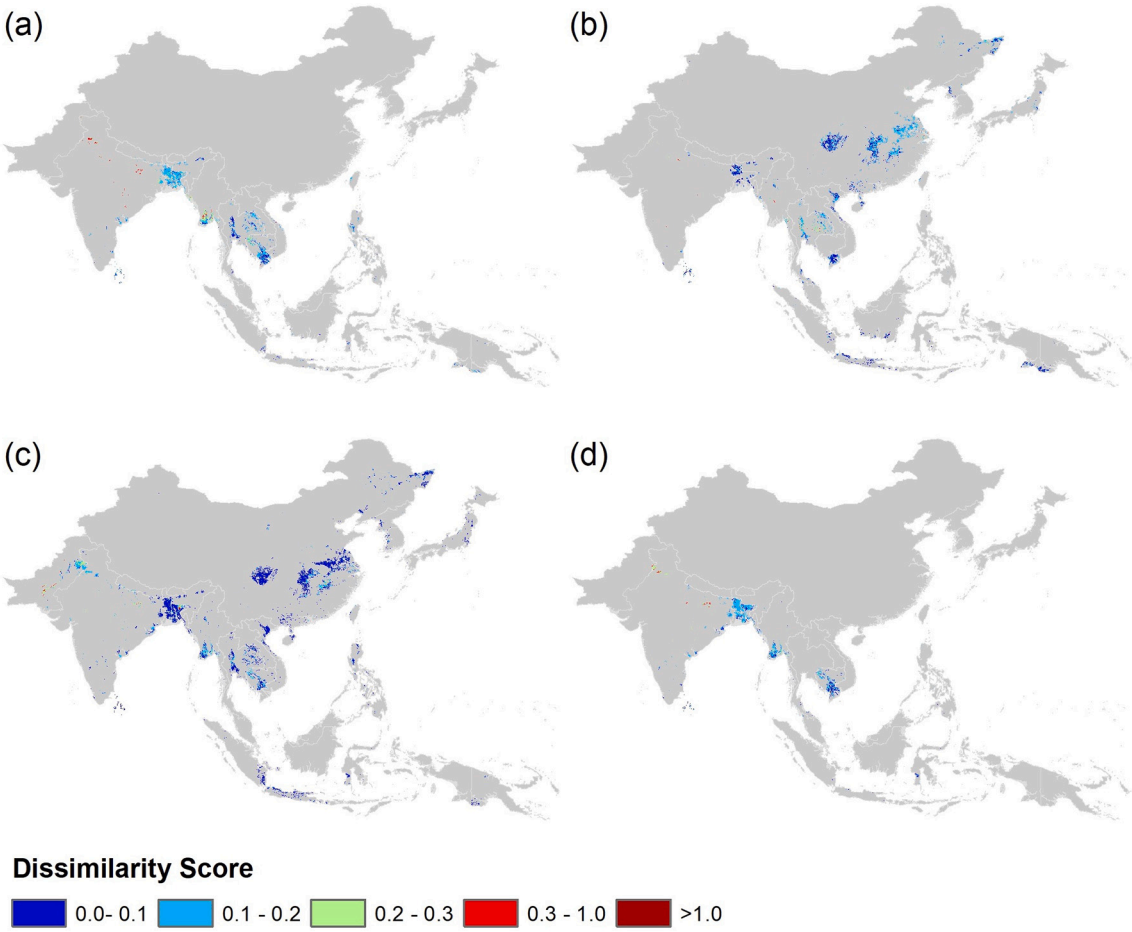

Fig. 10. Pixel dissimilarity score for (a) the 6th 8-day, (b) the 17th 8-day, (c) the 29th 8-day, and (d) the 41th 8-day in 2010, shown here as examples representing different seasons. Dissimilarity score is below the applicability threshold (0.3, i.e., extrapolation is unlikely) in most areas in all seasons. In winter and spring, the dissimilarity score in a small proportion of areas in South Asia is higher than 0.3.

It is also possible to improve data-driven models by including additional biophysical/biochemical predictors. Many studies have suggested that proxy variables of the availability of substrate for CH_4 production, such as gross primary productivity (GPP) or net primary productivity (NPP), are important predictors of CH_4 flux (Dai et al., 2019; Knox et al., 2021; Li et al., 2019; Schütz et al., 1990). However, due to limitations of site data and a lack of gridded products exclusively for rice, we did not consider GPP or NPP. Instead, we included several vegetation indices (e.g., EVI, NDVI, SAVI, and SATVI) that can be proxies of GPP, and our predictor selection process selected SATVI, but its importance is secondary to other variables (Fig. 2b). Indeed, the previous study for wetland CH_4 upscaling also did not identify GPP or vegetation indices (i.e., EVI) as important predictors either (Peltola et al., 2019). Moreover, Knox et al. (2019) did not find GPP as an important predictor of cross-site CH_4 emission variability in their multi-site synthesis study. Therefore, GPP/NPP might be important as within-site drivers of CH_4 emissions, but their effect on cross-site differences may be less important or may be confounded by other factors such as temperature. Latent heat flux, which is correlated to the plant-mediated transportation ability of gasses, has been found to influence or correlate with methane emission (Dai et al., 2019; Knox et al., 2021; Sturtevant et al., 2016) but was also not explored in this study for the same reason as GPP/NPP. Nevertheless, future studies should test the abilities of GPP/NPP and other fluxes (e.g., respiration and latent heat) to predict CH_4 fluxes, particularly if gridded products exclusively for paddy-rice would be available from global carbon and energy flux upscaling efforts such as FLUXCOM (Jung et al., 2020). Water table depth and soil moisture are important methane drivers too (Knox et al., 2021). Our inclusion of the water index LSWI can indirectly capture part of water dynamic information, but future studies should try to directly consider water table depth and soil moisture when such data becomes available at a large spatial scale.

Future studies should also consider the management of paddy rice (e.g. plowing, tillage, fertilization, and irrigation), which affects the magnitude and timing of CH_4 emissions (Hou et al., 2020; Runkle et al., 2019; Yan et al., 2009) but could not be thoroughly tested in our upscaling because of the scarcity of spatially-resolved data. Previous efforts have mapped different agronomic managements, such as manure application (Carlson et al., 2016) and synthetic nitrogen fertilizer application rates (Houlton et al., 2019). However, the spatial resolution of these products is still too coarse (based on country or province/state levels, and occasionally county levels) and typically not time-resolved, thus limiting their application for dynamic higher-resolution applications such as ours. As smart-farming tools spread, and statistical data is collected at finer scales in the future, researchers may combine citizen-science and crowdsourcing techniques to produce dynamic maps of agronomic management at regional-to-field scale which would be usable in CH_4 flux upscaling. Remote sensing can and will also play an important role in improving mapping of agronomic management (Bégué et al., 2018), especially for irrigation practices (Chen et al., 2018; Huang et al., 2021). New generations of sensors, including presently available Planet CubeSat, SMAP, and Sentinel 1 & 2 optical and SAR images, and other planned SAR sensors such as Tandem-L and NISAR Satellites, can provide great opportunities for mapping regional and global irrigation regimes and soil moisture conditions.

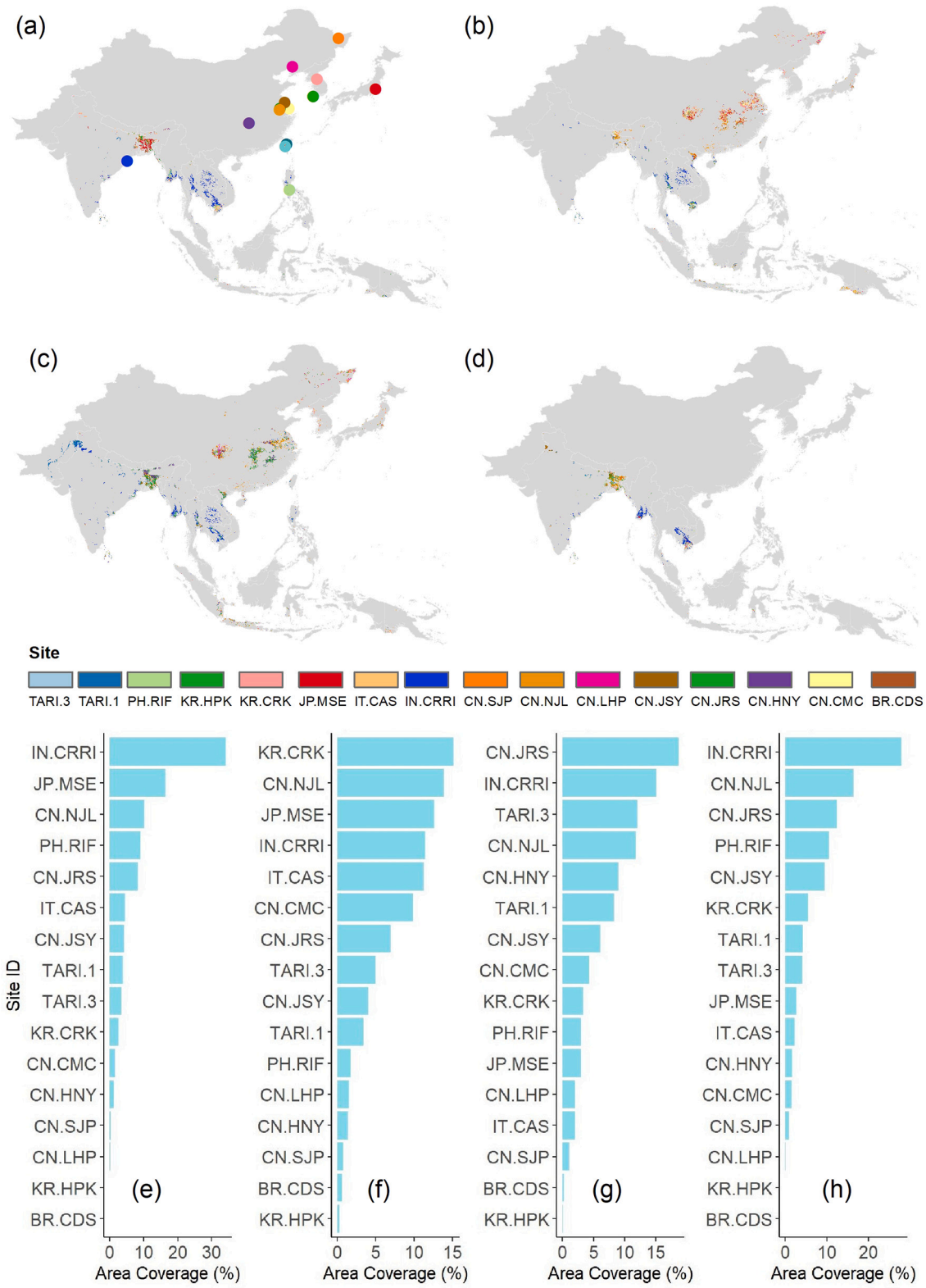

Fig. 11. Maps of site constituencies that identify the site from which training conditions have the largest importance on predictions outside of training locations on (a) the 6th 8-day, (b) the 17th 8-day, (c) the 29th 8-day, and (d) the 41th 8-day in 2010, and constituencies ranked by descending percent global coverage in terms of areas according to the constituency maps shown in Fig. S9. (e) the 6th 8-day, (f) the 17th 8-day, (g) the 29th 8-day, and (h) the 41th 8-day in 2010. The four selected dates represent mid spring, summer, autumn, and winter seasons of Monsoon Asia.

Calendars of rice cultivation are important to be improved for masking growing season and fallow season emissions. While rice plantation schedules generally follow the seasonality of the local climate, artificial planting and harvesting dates can still vary by a few weeks across regions with similar climates, leading to different season lengths and environmental or management conditions. The rice calendar data from riceAtlas were used in this study to mask predictions outside the rice season, but riceAtlas is not sufficiently detailed both in time and space (as it provides the earliest date, latest date, and peak date based on national survey statistics for spatial units at first (e.g., country) or second (e.g., state or province) administrative levels). We adopted the peak dates provided by riceAltas for calendar masking in this study, which could introduce errors and uncertainties at finer spatial scales. Therefore, it is important for future studies to consider more accurate rice cropping calendars and produce rice calendar data at finer spatial details. Again, new generations of remote sensing satellites (such as Planet CubeSat and/or Sentinel-1A images) provide new opportunities to map rice calendars at higher resolution utilizing time series images and analysis (Moeini Rad et al., 2019), which in the future can help improve predicting methane emissions in paddy rice.

4. Conclusions

We produced to our knowledge the first gridded CH_4 emission product for paddy rice in Monsoon Asia countries based on upscaling of ground-based eddy covariance CH_4 flux measurements using remote sensing predictors, at 8-day steps from 2001 to 2015. We predict average annual paddy rice CH_4 emissions of 20.6 ± 1.1 Tg yr^{-1} for 2001–2015 which is at the lower range of previous inventory-based estimates (20–32 Tg yr^{-1}). Our annual emission estimates also reveal that CH_4 emission from paddy rice may have been declining over time, especially after 2007, suggesting the CH_4 emission from paddy rice in Monsoon Asia has likely not contributed to the renewed growth of CH_4 in the atmosphere in recent years. We explored 175 predictors in predicting CH_4 through machine learning, and found that temperature, biomass and water related indices are most important for CH_4 prediction in paddy rice, but future studies should consider incorporating variables regarding carbon substrates, carbon fluxes, and crop management and calendars. Our network of 23 towers also highlights the need to expand ecosystem-scale CH_4 flux measurement to paddy rice in tropical parts of South Asia and Southeast Asia. All gridded emissions products are available at https://doi.org/10.5281/zenodo.7145497 and can be used to compare to other bottom-up or top-down studies or used as priors in inversion modeling.

Author contribution

RBJ acquired funding for the project. ZO, RBJ, BRKR, and DP conceived of the project and reviewed and edited the manuscript in all stages. ZO, EFC, and GM designed and created the gridded predictor and final methane products. ZO and GM developed the machine learning approach. ZO wrote the original draft with significant contributions from GM, EFC, SHK, SC, KBD, SF, and AM. JAI gap-filled methane flux data and edited the methodology. XX, YR, HI, WY, and YZ reviewed the draft. MM and SS initiated and contributed to data collection. JD, GZ, and XX provided the MODIS paddy rice distribution maps. MCRA, AC, CLC, BNF, HG, LH, HI, QJ, WJ, MK, HL, JK, MLB, AKN, DRR, YR, CKS, BT, WY, YZ were principal investigators that contributed flux data from one or more of the analysis sites. All authors have reviewed the final version of the paper.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

We acknowledge primary support from the Gordon and Betty Moore Foundation (grant GBMF5439, "Advancing Understanding of the Global Methane Cycle"; Stanford University) and from the John Wesley Powell Center for Analysis and Synthesis of the US Geological Survey ("Wetland FLUXNET Synthesis for Methane" working group). The work is dedicated to Muhammad Muddasir, suddenly and tragically died during his last year of PhD on CH_4 emissions from rice paddies.

We acknowledge Dr. Jiquan Chen from Michigan State University and Dr. Changliang Shao from Chinese Academy of Agricultural Sciences for helping data collection from Site PIs. We acknowledge Dr. Carlo Trotta from CMCC for advising on data standardization and Dr. Ken Belitz from USGS for sharing bias correction algorithms for machine learning models. We acknowledge Dr. Lena Höglund-Isaksson for sharing CH_4 emission data from rice cultivation from GAINS-ECLIPSE6. Benjamin R. K. Runkle was supported by the US National Science Foundation CBET CAREER Award 1752083.Y. Ryu was supported by the National Research Foundation of Korea (NRF-2019R1A2C2084626) and Ministry of Environment of Korea (2022003640002). YG. Zhang was supported by the National Research Foundation of China (NSFC-42071388). D. Papale thanks the Open-Earth-Monitor European Union's Horizon Europe research project (GA 101059548). KoFlux sites (KR.CRK and KR.HPK) were supported by National Research Foundation of Korea (NRF-2018R1C1B6002917) and Rural Development Administration of Korea (PJ014892022022). JP-Mse site was supported by the Ministry of Agriculture, Forestry and Fisheries, Japan through a research project ("Development of technologies for mitigation and adaptation to climate change in Agriculture, Forestry and Fisheries", FY 2010–2014), and also by the Ministry of Environment, Japan through the Global Environment Research Account for National Institutes ("Sensor Network and Data Processing Automation for Evaluation of the Carbon Cycle Change in Terrestrial Ecosystems", FY 2012–2016).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.rse.2022.113335.

References

Allen, G., 2016. Biogeochemistry: rebalancing the global methane budget. Nature 538, 46–48. https://doi.org/10.1038/538046a.

Amatulli, G., Domisch, S., Tuanmu, M.-N., Parmentier, B., Ranipeta, A., Malczyk, J., Jetz, W., 2018. A suite of global, cross-scale topographic variables for environmental and biodiversity modeling. Sci Data 5, 180040.

Baldocchi, D., 2014. Measuring fluxes of trace gases and energy between ecosystems and the atmosphere - the state and future of the eddy covariance method. Glob. Chang. Biol. 20, 3600–3609.

Bégué, A., Arvor, D., Bellon, B., Betbeder, J., de Abelleyra, D., Ferraz, R.P.D., Lebourgeois, V., Lelong, C., Simões, M., Verón, S.R., 2018. Remote sensing and cropping practices: a review. Remote Sens. https://doi.org/10.3390/rs10010099.

Belitz, K., Stackelberg, P.E., 2021. Evaluation of six methods for correcting bias in estimates from ensemble tree machine learning regression models. Environ. Model Softw. 139, 105006.

Bergamaschi, P., Frankenberg, C., Meirink, J.F., Krol, M., Dentener, F., Wagner, T., Platt, U., Kaplan, J.O., Körner, S., Heimann, M., Dlugokencky, E.J., Goede, A., 2007. Satellite chartography of atmospheric methane from SCIAMACHY on board ENVISAT: 2. Evaluation based on inverse model simulations. J. Geophys. Res. https://doi.org/10.1029/2006jd007268.

Bloom, A.A., Palmer, P.I., Fraser, A., Reay, D.S., Frankenberg, C., 2010. Large-scale controls of methanogenesis inferred from methane and gravity spaceborne data. Science 327, 322–325.

Bo, Y., Jägermeyr, J., Yin, Z., Jiang, Y., Xu, J., Liang, H., Zhou, F., 2022. Global benefits of non-continuous flooding to reduce greenhouse gases and irrigation water use without rice yield penalty. Glob. Chang. Biol. https://doi.org/10.1111/gcb.16132.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

Carlson, K.M., Gerber, J.S., Mueller, N.D., Herrero, M., MacDonald, G.K., Brauman, K.A., Havlik, P., O'Connell, C.S., Johnson, J.A., Saatchi, S., West, P.C., 2016. Greenhouse gas emissions intensity of global croplands. Nat. Clim. Chang. 7, 63–68.

Chang, K.-Y., Riley, W.J., Knox, S.H., Jackson, R.B., McNicol, G., Poulter, B., Aurela, M., Baldocchi, D., Bansal, S., Bohrer, G., Campbell, D.I., Cescatti, A., Chu, H., Delwiche, K.B., Desai, A.R., Euskirchen, E., Friborg, T., Goeckede, M., Helbig, M., Hemes, K.S., Hirano, T., Iwata, H., Kang, M., Keenan, T., Krauss, K.W., Lohila, A., Mammarella, I., Mitra, B., Miyata, A., Nilsson, M.B., Noormets, A., Oechel, W.C., Papale, D., Peichl, M., Reba, M.L., Rinne, J., Runkle, B.R.K., Ryu, Y., Sachs, T., Schäfer, K.V.R., Schmid, H.P., Shurpali, N., Sonnentag, O., Tang, A.C.I., Torn, M.S., Trotta, C., Tuittila, E.-S., Ueyama, M., Vargas, R., Vesala, T., Windham-Myers, L., Zhang, Z., Zona, D., 2021. Substantial hysteresis in emergent temperature sensitivity of global wetland CH4 emissions. Nat. Commun. https://doi.org/10.1038/s41467-021-22452-1.

Chen, Y., Lu, D., Luo, L., Pokhrel, Y., Deb, K., Huang, J., Ran, Y., 2018. Detecting irrigation extent, frequency, and timing in a heterogeneous arid agricultural region using MODIS time series, landsat imagery, and ancillary data. Remote Sens. Environ. 204, 197–211.

Chu, H., Baldocchi, D.D., John, R., Wolf, S., Reichstein, M., 2017. Fluxes all of the time? A primer on the temporal representativeness of FLUXNET. J. Geophys. Res. Biogeosci. 122, 289–307.

Chu, H., Luo, X., Ouyang, Z., Chan, W.S., Dengel, S., Biraud, S.C., Torn, M.S., Metzger, S., Kumar, J., Arain, M.A., Arkebauer, T.J., Baldocchi, D., Bernacchi, C., Billesbach, D., Black, T.A., Blanken, P.D., Bohrer, G., Bracho, R., Brown, S., Brunsell, N.A., Chen, J., Chen, X., Clark, K., Desai, A.R., Duman, T., Durden, D., Fares, S., Forbrich, I., Gamon, J.A., Gough, C.M., Griffis, T., Helbig, M., Hollinger, D., Humphreys, E., Ikawa, H., Iwata, H., Ju, Y., Knowles, J.F., Knox, S.H., Kobayashi, H., Kolb, T., Law, B., Lee, X., Litvak, M., Liu, H., Munger, J.W., Noormets, A., Novick, K., Oberbauer, S.F., Oechel, W., Oikawa, P., Papuga, S.A., Pendall, E., Prajapati, P., Prueger, J., Quinton, W.L., Richardson, A.D., Russell, E.S., Scott, R.L., Starr, G., Staebler, R., Stoy, P.C., Stuart-Haëntjens, E., Sonnentag, O., Sullivan, R.C., Suyker, A., Ueyama, M., Vargas, R., Wood, J.D., Zona, D., 2021. Representativeness of Eddy-covariance flux footprints for areas surrounding AmeriFlux sites. Agric. For. Meteorol. 301–302, 108350.

Dai, S., Ju, W., Zhang, Y., He, Q., Song, L., Li, J., 2019. Variations and drivers of methane fluxes from a rice-wheat rotation agroecosystem in eastern China at seasonal and diurnal scales. Sci. Total Environ. https://doi.org/10.1016/j.scitotenv.2019.07.012.

Delwiche, K.B., Knox, S.H., Malhotra, A., Fluet-Chouinard, E., McNicol, G., Feron, S., Ouyang, Z., Papale, D., Trotta, C., Canfora, E., Cheah, Y.-W., Christianson, D., Alberto, M.C.R., Alekseychik, P., Aurela, M., Baldocchi, D., Bansal, S., Billesbach, D. P., Bohrer, G., Bracho, R., Buchmann, N., Campbell, D.I., Celis, G., Chen, J., Chen, W., Chu, H., Dalmagro, H.J., Dengel, S., Desai, A.R., Detto, M., Dolman, H., Eichelmann, E., Euskirchen, E., Famulari, D., Fuchs, K., Goeckede, M., Gogo, S., Gondwe, M.J., Goodrich, J.P., Gottschalk, P., Graham, S.L., Heimann, M., Helbig, M., Helfter, C., Hemes, K.S., Hirano, T., Hollinger, D., Hörtnagl, L., Iwata, H., Jacotot, A., Jurasinski, G., Kang, M., Kasak, K., King, J., Klatt, J., Koebsch, F., Krauss, K.W., Lai, D.Y.F., Lohila, A., Mammarella, I., Belelli Marchesini, L., Manca, G., Matthes, J. H., Maximov, T., Merbold, L., Mitra, B., Morin, T.H., Nemitz, E., Nilsson, M.B., Niu, S., Oechel, W.C., Oikawa, P.Y., Ono, K., Peichl, M., Peltola, O., Reba, M.L., Richardson, A.D., Riley, W., Runkle, B.R.K., Ryu, Y., Sachs, T., Sakabe, A., Sanchez, C.R., Schuur, E.A., Schäfer, K.V.R., Sonnentag, O., Sparks, J.P., Stuart-Haëntjens, E., Sturtevant, C., Sullivan, R.C., Szutu, D.J., Thom, J.E., Torn, M.S., Tuittila, E.-S., Turner, J., Ueyama, M., Valach, A.C., Vargas, R., Varlagin, A., Vazquez-Lule, A., Verfaillie, J.G., Vesala, T., Vourlitis, G.L., Ward, E.J., Wille, C., Wohlfahrt, G., Wong, G.X., Zhang, Z., Zona, D., Windham-Myers, L., Poulter, B., Jackson, R.B., 2021. FLUXNET-CH4: a global, multi-ecosystem dataset and analysis of methane seasonality from freshwater wetlands. Earth Syst. Sci. Data 13, 3607–3689.

Dlugokencky, E., 2022. Global CH4 Monthly Means [WWW Document]. Global Monitoring Laboratory. URL (accessed 9.16.22). https://gml.noaa.gov/ccgg/trends_ch4/.

Fick, S.E., Hijmans, R.J., 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. Int. J. Climatol. 37, 4302–4315.

Fisher, J.B., Lee, B., Purdy, A.J., Halverson, G.H., Dohlen, M.B., Cawse-Nicholson, K., Wang, A., Anderson, R.G., Aragon, B., Arain, M.A., Baldocchi, D.D., Baker, J.M., Barral, H., Bernacchi, C.J., Bernhofer, C., Biraud, S.C., Bohrer, G., Brunsell, N., Cappelaere, B., Castro-Contreras, S., Chun, J., Conrad, B.J., Cremonese, E., Demarty, J., Desai, A.R., De Ligne, A., Foltýnová, L., Goulden, M.L., Griffis, T.J., Grünwald, T., Johnson, M.S., Kang, M., Kelbe, D., Kowalska, N., Lim, J.-H., Maïnassara, I., McCabe, M.F., Missik, J.E.C., Mohanty, B.P., Moore, C.E., Morillas, L., Morrison, R., Munger, J.W., Posse, G., Richardson, A.D., Russell, E.S., Ryu, Y., Sanchez-Azofeifa, A., Schmidt, M., Schwartz, E., Sharp, I., Šigut, L., Tang, Y., Hulley, G., Anderson, M., Hain, C., French, A., Wood, E., Hook, S., 2020. ECOSTRESS: NASA's next generation mission to measure evapotranspiration from the international space station. Water Resour. Res. 56 e2019WR026058.

Frankenberg, C., Meirink, J.F., van Weele, M., Platt, U., Wagner, T., 2005. Assessing methane emissions from global space-borne observations. Science 308, 1010–1014.

Ganesan, A.L., Rigby, M., Lunt, M.F., Parker, R.J., Boesch, H., Goulding, N., Umezawa, T., Zahn, A., Chatterjee, A., Prinn, R.G., Tiwari, Y.K., van der Schoot, M., Krummel, P.B., 2017. Atmospheric observations show accurate reporting and little growth in India's methane emissions. Nat. Commun. https://doi.org/10.1038/s41467-017-00994-7.

Gao, B.-C., 1996. NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. Remote Sens. Environ. 58, 257–266.

Ge, H.-X., Zhang, H.-S., Zhang, H., Cai, X.-H., Song, Y., Kang, L., 2018. The characteristics of methane flux from an irrigated rice farm in East China measured using the eddy covariance method. Agric. For. Meteorol. 249, 228–238.

Hargrove, W.W., Hoffman, F.M., 2004. Potential of multivariate quantitative methods for delineation and visualization of ecoregions. Environ. Manag. 34 (Suppl. 1), S39–S60.

Hengl, T., Mendes de Jesus, J., Heuvelink, G.B.M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M.N., Geng, X., Bauer-Marschallinger, B., Guevara, M.A., Vargas, R., MacMillan, R.A., Batjes, N.H., Leenaars, J.G.B., Ribeiro, E., Wheeler, I., Mantel, S., Kempen, B., 2017. SoilGrids250m: global gridded soil information based on machine learning. PLoS One 12, e0169748.

Höglund-Isaksson, L., 2005. GAINS Model for Greenhouse Gases – Version 1.0: Methane (CH4). International Institute for Applied Systems Analysis (IIASA).

Houlton, B.Z., Almaraz, M., Aneja, V., Austin, A.T., Bai, E., Cassman, K.G., Compton, J.E., Davidson, E.A., Erisman, J.W., Galloway, J.N., Gu, B., Yao, G., Martinelli, L.A., Scow, K., Schlesinger, W.H., Tomich, T.P., Wang, C., Zhang, X., 2019. A world of cobenefits: solving the global nitrogen challenge. Earth's Future 7, 865–872.

Hou, P., Yu, Y., Xue, L., Petropoulos, E., He, S., Zhang, Y., Pandey, A., Xue, L., Yang, L., Chen, D., 2020. Effect of long term fertilization management strategies on methane emissions and rice yield. Sci. Total Environ. 725, 138261.

Huang, R., Huang, J.-X., Zhang, C., Ma, H.-Y., Zhuo, W., Chen, Y.-Y., Zhu, D.-H., Wu, Q., Mansaray, L.R., 2020. Soil temperature estimation at different depths, using remotely-sensed data. J. Integr. Agric. 19, 277–290.

Huang, X., Runkle, B.R.K., Isbell, M., Moreno-García, B., McNairn, H., Reba, M.L., Torbick, N., 2021. Rice Inundation Assessment Using Polarimetric UAVSAR Data. Earth Space Sci 8 (e2020EA001554).

Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., Ferreira, L.G., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sens. Environ. 83, 195–213.

Huete, A.R., 1988. A soil-adjusted vegetation index (SAVI). Remote Sens. Environ. 25, 295–309.

Humphreys, E., Kukal, S.S., Christen, E.W., Hira, G.S., Balwinder-Singh, Sudhir-Yadav, Sharma, R.K., 2010. Chapter five - halting the groundwater decline in north-West India—which crop technologies will be Winners? In: Sparks, D.L. (Ed.), Advances in Agronomy. Academic Press, pp. 155–217.

Hwang, Y., Ryu, Y., Huang, Y., Kim, J., Iwata, H., Kang, M., 2020. Comprehensive assessments of carbon dynamics in an intermittently-irrigated rice paddy. Agric. For. Meteorol. 285–286, 107933.

IPCC, 2021. Climate Change 2021: The Physical Science Basis. IPCC.

Irvin, J., Zhou, S., McNicol, G., Lu, F., Liu, V., Fluet-Chouinard, E., Ouyang, Z., Knox, S. H., Lucas-Moffat, A., Trotta, C., Papale, D., Vitale, D., Mammarella, I., Alekseychik, P., Aurela, M., Avati, A., Baldocchi, D., Bansal, S., Bohrer, G., Campbell, D.I., Chen, J., Chu, H., Dalmagro, H.J., Delwiche, K.B., Desai, A.R., Euskirchen, E., Feron, S., Goeckede, M., Heimann, M., Helbig, M., Helfter, C., Hemes, K.S., Hirano, T., Iwata, H., Jurasinski, G., Kalhori, A., Kondrich, A., Lai, D.Y. F., Lohila, A., Malhotra, A., Merbold, L., Mitra, B., Ng, A., Nilsson, M.B., Noormets, A., Peichl, M., Rey-Sanchez, A.C., Richardson, A.D., Runkle, B.R.K., Schäfer, K.V.R., Sonnentag, O., Stuart-Haëntjens, E., Sturtevant, C., Ueyama, M., Valach, A.C., Vargas, R., Vourlitis, G.L., Ward, E.J., Wong, G.X., Zona, D., Alberto, M. C.R., Billesbach, D.P., Celis, G., Dolman, H., Friborg, T., Fuchs, K., Gogo, S., Gondwe, M.J., Goodrich, J.P., Gottschalk, P., Hörtnagl, L., Jacotot, A., Koebsch, F., Kasak, K., Maier, R., Morin, T.H., Nemitz, E., Oechel, W.C., Oikawa, P.Y., Ono, K., Sachs, T., Sakabe, A., Schuur, E.A., Shortt, R., Sullivan, R.C., Szutu, D.J., Tuittila, E.-S., Varlagin, A., Verfaillie, J.G., Wille, C., Windham-Myers, L., Poulter, B., Jackson, R.B., 2021. Gap-filling eddy covariance methane fluxes: comparison of machine learning model predictions and uncertainties at FLUXNET-CH4 wetlands. Agric. For. Meteorol. 308–309, 108528.

Jacob, D.J., Turner, A.J., Maasakkers, J.D., Sheng, J., Sun, K., Liu, X., Chance, K., Aben, I., McKeever, J., Frankenberg, C., 2016. Satellite observations of atmospheric methane and their value for quantifying methane emissions. Atmos. Chem. Phys. 16, 14371–14396.

Janssens-Maenhout, G., Crippa, M., Guizzardi, D., Muntean, M., Schaaf, E., Dentener, F., Bergamaschi, P., Pagliari, V., Olivier, J.G.J., Peters, J.A., Others, 2019. EDGAR v4. 3.2 global atlas of the three major greenhouse gas emissions for the period 1970–2012. Earth System Science Data 11, 959–1002.

Jung, M., Schwalm, C., Migliavacca, M., Walther, S., Camps-Valls, G., Koirala, S., Anthoni, P., Besnard, S., Bodesheim, P., Carvalhais, N., Chevallier, F., Gans, F., Goll, D.S., Haverd, V., Köhler, P., Ichii, K., Jain, A.K., Liu, J., Lombardozzi, D., Nabel, J.E.M., Nelson, J.A., O'Sullivan, M., Pallandt, M., Papale, D., Peters, W., Pongratz, J., Rödenbeck, C., Sitch, S., Tramontana, G., Walker, A., Weber, U., Reichstein, M., 2020. Scaling carbon fluxes from eddy covariance sites to globe: synthesis and evaluation of the FLUXCOM approach. Biogeosciences. https://doi.org/10.5194/bg-17-1343-2020.

Kim, J., Guo, Q., Baldocchi, D., Leclerc, M., Xu, L., Schmid, H., 2006. Upscaling fluxes from tower to landscape: overlaying flux footprints on high-resolution (IKONOS) images of vegetation cover. Agric. For. Meteorol. https://doi.org/10.1016/j.agrformet.2004.11.015.

Knox, S.H., Bansal, S., McNicol, G., Schafer, K., Sturtevant, C., Ueyama, M., Valach, A.C., Baldocchi, D., Delwiche, K., Desai, A.R., Euskirchen, E., Liu, J., Lohila, A., Malhotra, A., Melling, L., Riley, W., Runkle, B.R.K., Turner, J., Vargas, R., Zhu, Q., Alto, T., Fluet-Chouinard, E., Goeckede, M., Melton, J.R., Sonnentag, O., Vesala, T., Ward, E., Zhang, Z., Feron, S., Ouyang, Z., Alekseychik, P., Aurela, M., Bohrer, G., Campbell, D.I., Chen, J., Chu, H., Dalmagro, H.J., Goodrich, J.P., Gottschalk, P., Hirano, T., Iwata, H., Jurasinski, G., Kang, M., Koebsch, F., Mammarella, I., Nilsson, M.B., Ono, K., Peichl, M., Peltola, O., Ryu, Y., Sachs, T., Sakabe, A., Sparks, J.P., Tuittila, E.-S., Vourlitis, G.L., Wong, G.X., Windham-Myers, L., Poulter, B., Jackson, R.B., 2021. Identifying dominant environmental predictors of

freshwater wetland methane fluxes across diurnal to seasonal time scales. Glob. Chang. Biol. 27, 3582–3604.
Knox, S.H., Jackson, R.B., Poulter, B., McNicol, G., Fluet-Chouinard, E., Zhang, Z., Hugelius, G., Bousquet, P., Canadell, J.G., Saunois, M., Others, 2019. FLUXNET-CH 4 synthesis activity: objectives, observations, and future directions. Bull. Am. Meteorol. Soc. 100, 2607–2632.
Knox, S.H., Matthes, J.H., Sturtevant, C., Oikawa, P.Y., Verfaillie, J., Baldocchi, D., 2016. Biophysical controls on interannual variability in ecosystem-scale CO 2 and CH 4 exchange in a California rice paddy. J. Geophys. Res. Biogeosci. 121, 978–1001.
Kong, J., Ryu, Y., Liu, J., Dechant, B., Rey-Sanchez, C., Shortt, R., Szutu, D., Verfaillie, J., Houborg, R., Baldocchi, D.D., 2022. Matching high resolution satellite data and flux tower footprints improves their agreement in photosynthesis estimates. Agric. For. Meteorol. 316, 108878.
Kuhn, M., 2008. Building predictive models in R using the caret package. J. Stat. Softw. 28, 1–26.
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., Nelson, A., 2017. RiceAtlas, a spatial database of global rice calendars and production. Sci Data 4, 170074.
Lamarque, J.-F., Dentener, F., McConnell, J., Ro, C.-U., Shaw, M., Vet, R., Bergmann, D., Cameron-Smith, P., Doherty, R., Faluvegi, G., Ghan, S.J., Josse, B., Lee, Y.H., MacKenzie, I.A., Plummer, D., Shindell, D.T., Stevenson, D.S., Strode, S., Zeng, G., 2013. Multi-model mean nitrogen and sulfur deposition from the Atmospheric Chemistry and Climate Model Intercomparison Project (ACCMIP): Evaluation historical and projected changes. Atmos. Chem. Phys. 13, 7997–8018. https://doi.org/10.5194/acp-13-7997-2013.
Lasslop, G., Reichstein, M., Papale, D., Richardson, A.D., Arneth, A., Barr, A., Stoy, P., Wohlfahrt, G., 2010. Separation of net ecosystem exchange into assimilation and respiration using a light response curve approach: critical issues and global evaluation. Glob. Chang. Biol. 16, 187–208.
Li, H., Guo, H.-Q., Helbig, M., Dai, S.-Q., Zhang, M.-S., Zhao, M., Peng, C.-H., Xiao, X.-M., Zhao, B., 2019. Does direct-seeded rice decrease ecosystem-scale methane emissions?—A case study from a rice paddy in Southeast China. Agric. For. Meteorol. 272–273, 118–127.
Liou, R.-M., Huang, S.-N., Lin, C.-W., Chen, S.-H., 2003. Methane emission from fields with three various rice straw treatments in Taiwan paddy soils. J. Environ. Sci. Health B 38, 511–527.
Maraseni, T.N., Deo, R.C., Qu, J., Gentle, P., Neupane, P.R., 2018. An international comparison of rice consumption behaviours and greenhouse gas emissions from rice production. J. Clean. Prod. 172, 2288–2300.
Marsett, R.C., Qi, J., Heilman, P., Biedenbender, S.H., Carolyn Watson, M., Amer, S., Weltz, M., Goodrich, D., Marsett, R., 2006. Remote sensing for grassland management in the Arid Southwest. Rangeland Ecol. Manage. 59, 530–540.
Meingast, K.M., Falkowski, M.J., Kane, E.S., Potvin, L.R., Benscoter, B.W., Smith, A.M.S., Bourgeau-Chavez, L.L., Miller, M.E., 2014. Spectral detection of near-surface moisture content and water-table position in northern peatland ecosystems. Remote Sens. Environ. 152, 536–546.
Meyer, H., Pebesma, E., 2021. Predicting into unknown space? Estimating the area of applicability of spatial prediction models. Methods Ecol. Evol. 12, 1620–1633.
Meyer, H., Reudenbach, C., Hengl, T., Katurji, M., Nauss, T., 2018. Improving performance of spatio-temporal machine learning models using forward feature selection and target-oriented validation. Environ. Model Softw. 101, 1–9.
Minamikawa, K., Fumoto, T., Iizumi, T., Cha-Un, N., Pimple, U., Nishimori, M., Ishigooka, Y., Kuwagata, T., 2016. Prediction of future methane emission from irrigated rice paddies in Central Thailand under different water management practices. Sci. Total Environ. 566–567, 641–651.
Moeini Rad, A., Ashourloo, D., Salehi Shahrabi, H., Nematollahi, H., 2019. Developing an automatic phenology-based algorithm for Rice detection using Sentinel-2 time-series data. IEEE J. Select. Top. Appl.Earth Observ. Remote Sens. 12, 1471–1481.
Morin, T.H., 2019. Advances in the Eddy covariance approach to CH 4 monitoring over two and a half decades. Journal of geophysical researchBiogeosciences. https://doi.org/10.1029/2018jg004796.
Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I—A discussion of principles. J. Hydrol. 10, 282–290.
Nisbet, E.G., Dlugokencky, E.J., Bousquet, P., 2014. Methane on the rise—again. Science 343, 493–495.
Nisbet, E.G., Dlugokencky, E.J., Manning, M.R., Lowry, D., Fisher, R.E., France, J.L., Michel, S.E., Miller, J.B., White, J.W.C., Vaughn, B., Bousquet, P., Pyle, J.A., Warwick, N.J., Cain, M., Brownlow, R., Zazzeri, G., Lanoisellé, M., Manning, A.C., Gloor, E., Worthy, D.E.J., Brunke, E.-G., Labuschagne, C., Wolff, E.W., Ganesan, A.L., 2016. Rising atmospheric methane: 2007–2014 growth and isotopic shift. Glob. Biogeochem. Cycles. https://doi.org/10.1002/2016gb005406.
Papale, D., 2020. Ideas and perspectives: enhancing the impact of the FLUXNET network of eddy covariance sites. Biogeosciences 17, 5587–5598.
Papale, D., Black, T.A., Carvalhais, N., Cescatti, A., Chen, J., Jung, M., Kiely, G., Lasslop, G., Mahecha, M.D., Margolis, H., Merbold, L., Montagnani, L., Moors, E., Olesen, J.E., Reichstein, M., Tramontana, G., Gorsel, E., Wohlfahrt, G., Ráduly, B., 2015. Effect of spatial sampling from european flux towers for estimating carbon and water fluxes with artificial neural networks. J. Geophys. Res. Biogeosci. 120, 1941–1957.
Peltola, O., Vesala, T., Gao, Y., Räty, O., Alekseychik, P., Aurela, M., Chojnicki, B., Desai, A.R., Dolman, A.J., Euskirchen, E.S., Friborg, T., Göckede, M., Helbig, M., Humphreys, E., Jackson, R.B., Jocher, G., Joos, F., Klatt, J., Knox, S.H., Kowalska, N., Kutzbach, L., Lienert, S., Lohila, A., Mammarella, I., Nadeau, D.F., Nilsson, M.B., Oechel, W.C., Peichl, M., Pypker, T., Quinton, W., Rinne, J., Sachs, T., Samson, M., Schmid, H.P., Sonnentag, O., Wille, C., Zona, D., Aalto, T., 2019. Monthly gridded data product of northern wetland methane emissions based on upscaling eddy covariance observations. Earth Syst. Sci. Data. https://doi.org/10.5194/essd-11-1263-2019.
Peng, Piao, Bousquet, Ciais, Li, 2016. Inventory of anthropogenic methane emissions in mainland China from 1980 to 2010. Atmos. Clim. Sci. 16, 14545–14562. https://doi.org/10.5194/acp-16-14545-2016.
Qi, J., Chehbouni, A., Huete, A.R., Kerr, Y.H., Sorooshian, S., 1994. A modified soil adjusted vegetation index. Remote Sens. Environ. 48, 119–126.
Qi, J., Marsett, R., Heilman, P., Bieden-bender, S., Moran, S., Goodrich, D., Weltz, M., 2002. RANGES improves satellite-based information and land cover assessments in Southwest United States. Eos, Trans. Am. Geophys. Union. https://doi.org/10.1029/2002eo000411.
Reba, M.L., Fong, B.N., Rijal, I., 2019. Fallow season CO2 and CH4 fluxes from US mid-south rice-waterfowl habitats. Agric. For. Meteorol. 279, 107709.
Reichstein, M., Falge, E., Baldocchi, D., Papale, D., Aubinet, M., Berbigier, P., Bernhofer, C., Buchmann, N., Gilmanov, T., Granier, A., Grunwald, T., Havrankova, K., Ilvesniemi, H., Janous, D., Knohl, A., Laurila, T., Lohila, A., Loustau, D., Matteucci, G., Meyers, T., Miglietta, F., Ourcival, J.-M., Pumpanen, J., Rambal, S., Rotenberg, E., Sanz, M., Tenhunen, J., Seufert, G., Vaccari, F., Vesala, T., Yakir, D., Valentini, R., 2005. On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm. Glob. Chang. Biol. 11, 1424–1439.
Runkle, B.R.K., Suvočarev, K., Reba, M.L., Reavis, C.W., Smith, S.F., Chiu, Y.-L., Fong, B., 2019. Methane emission reductions from the alternate wetting and drying of Rice fields detected using the Eddy covariance method. Environ. Sci. Technol. 53, 671–681.
Saunois, M., Stavert, A.R., Poulter, B., Bousquet, P., Canadell, J.G., Jackson, R.B., Raymond, P.A., Dlugokencky, E.J., Houweling, S., Patra, P.K., Ciais, P., Arora, V.K., Bastviken, D., Bergamaschi, P., Blake, D.R., Brailsford, G., Bruhwiler, L., Carlson, K. M., Carrol, M., Castaldi, S., Chandra, N., Crevoisier, C., Crill, P.M., Covey, K., Curry, C.L., Etiope, G., Frankenberg, C., Gedney, N., Hegglin, M.I., Höglund-Isaksson, L., Hugelius, G., Ishizawa, M., Ito, A., Janssens-Maenhout, G., Jensen, K.M., Joos, F., Kleinen, T., Krummel, P.B., Langenfelds, R.L., Laruelle, G.G., Liu, L., Machida, T., Maksyutov, S., McDonald, K.C., McNorton, J., Miller, P.A., Melton, J.R., Morino, I., Müller, J., Murguia-Flores, F., Naik, V., Niwa, Y., Noce, S., O'Doherty, S., Parker, R.J., Peng, C., Peng, S., Peters, G.P., Prigent, C., Prinn, R., Ramonet, M., Regnier, P., Riley, W.J., Rosentreter, J.A., Segers, A., Simpson, I.J., Shi, H., Smith, S. J., Steele, L.P., Thornton, B.F., Tian, H., Tohjima, Y., Tubiello, F.N., Tsuruta, A., Viovy, N., Voulgarakis, A., Weber, T.S., van Weele, M., van der Werf, G.R., Weiss, R. F., Worthy, D., Wunch, D., Yin, Y., Yoshida, Y., Zhang, W., Zhang, Z., Zhao, Y., Zheng, B., Zhu, Q., Zhu, Q., Zhuang, Q., 2020. The global methane budget 2000–2017. Earth Syst. Sci. Data 12, 1561–1623.
Schaefer, H., Mikaloff Fletcher, S.E., Veidt, C., Lassey, K.R., Brailsford, G.W., Bromley, T. M., Dlugokencky, E.J., Michel, S.E., Miller, J.B., Levin, I., Lowe, D.C., Martin, R.J., Vaughn, B.H., White, J.W.C., 2016. A 21st-century shift from fossil-fuel to biogenic methane emissions indicated by 13CH₄. Science 352, 80–84.
Schütz, H., Seiler, W., Conrad, R., 1990. Influence of soil temperature on methane emission from rice paddy fields. Biogeochemistry 11, 77–95.
Schwietzke, S., Sherwood, O.A., Bruhwiler, L.M.P., Miller, J.B., Etiope, G., Dlugokencky, E.J., Michel, S.E., Arling, V.A., Vaughn, B.H., White, J.W.C., Tans, P. P., 2017. Corrigendum: upward revision of global fossil fuel methane emissions based on isotope database. Nature 543, 452.
Song, J., 2015. Bias corrections for random forest in regression using residual rotation. J. Korean Stat. Soc. 44, 321–326.
Stell, E., Warner, D., Jian, J., Bond-Lamberty, B., Vargas, R., 2021. Spatial biases of information influence global estimates of soil respiration: how can we improve global predictions? Glob. Chang. Biol. 27, 3923–3938.
Sturtevant, C., Ruddell, B.L., Knox, S.H., Verfaillie, J., Matthes, J.H., Oikawa, P.Y., Baldocchi, D., 2016. Identifying scale-emergent, nonlinear, asynchronous processes of wetland methane exchange. J. Geophys. Res. Biogeosci. 121, 188–204.
Sullivan, D.G., Truman, C.C., Schomberg, H.H., Endale, D.M., 2006. Evaluating techniques for determining tillage regime in the Southeastern Coastal Plain and Piedmont.
Tramontana, G., Jung, M., Camps-Valls, G., Ichii, K., Raduly, B., Reichstein, M., Schwalm, C.R., Arain, M.A., Cescatti, A., Kiely, G., Merbold, L., Serrano-Ortiz, P., Sickert, S., Wolf, S., Papale, D., 2016. Predicting carbon dioxide and energy fluxes across global FLUXNET sites with regression algorithms. Biogeosci. Discuss. 1–33.
Turner, A.J., Frankenberg, C., Kort, E.A., 2019. Interpreting contemporary trends in atmospheric methane. Proc. Natl. Acad. Sci. U. S. A. 116, 2805–2813.
Van Deventer, A.P., Ward, A.D., Gowda, P.H., Lyon, J.G., 1997. Using thematic mapper data to identify contrasting soil plains and tillage practices. Photogramm. Eng. Remote Sens. 63, 87–93.
Villarreal, S., Guevara, M., Alcaraz-Segura, D., Brunsell, N.A., Hayes, D., Loescher, H.W., Vargas, R., 2018. Ecosystem functional diversity and the representativeness of environmental networks across the conterminous United States. Agric. For. Meteorol. 262, 423–433.
Villarreal, S., Vargas, R., 2021. Representativeness of FLUXNET sites across Latin America. J. Geophys. Res. Biogeosci. 126 https://doi.org/10.1029/2020jg006090.
Wang, H., Hijmans, R.J., 2019. Climate change and geographic shifts in rice production in China. Environ. Res. Commun. 1, 011008.
Wang, Z., Zhang, X., Liu, L., Wang, S., Zhao, L., Wu, X., Zhang, W., Huang, X., 2021. Estimates of methane emissions from chinese rice fields using the DNDC model. Agric. For. Meteorol. 303, 108368.
Wright, M.N., Ziegler, A., 2015. ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. arXiv [stat.ML]..

Wutzler, T., Lucas-Moffat, A., Migliavacca, M., Knauer, J., Sickel, K., Šigut, L., Menzer, O., Reichstein, M., 2018. Basic and extensible post-processing of eddy covariance flux data with REddyProc. Biogeosciences 15, 5015–5030.

Xiao, J., Ollinger, S.V., Frolking, S., Hurtt, G.C., Hollinger, D.Y., Davis, K.J., Pan, Y., Zhang, X., Deng, F., Chen, J., Baldocchi, D.D., Law, B.E., Altaf Arain, M., Desai, A.R., Richardson, A.D., Sun, G., Amiro, B., Margolis, H., Gu, L., Scott, R.L., Blanken, P.D., Suyker, A.E., 2014. Data-driven diagnostics of terrestrial carbon dynamics over North America. Agric. For. Meteorol. https://doi.org/10.1016/j.agrformet.2014.06.013.

Xiao, Qitao, Zhang, Mi, Zhenghua, Hu., Gao, Yunqiu, Cheng, Hu., Liu, Cheng, Liu, Shoudong, et al., 2017. Spatial variations of methane emission in a large shallow eutrophic Lake in subtropical Climate. Journal of geophysical researchBiogeosciences. https://doi.org/10.1002/2017jg003805.

Xiao, X., Boles, S., Frolking, S., Salas, W., Moore, B., Li, C., He, L., Zhao, R., 2002. Observation of flooding and rice transplanting of paddy rice fields at the site to landscape scales in China using VEGETATION sensor data. Int. J. Remote Sens. 23, 3009–3022.

Xiao, X., Boles, S., Liu, J., Zhuang, D., Frolking, S., Li, C., Salas, W., Moore, B., 2005. Mapping paddy rice agriculture in southern China using multi-temporal MODIS images. Remote Sens. Environ. https://doi.org/10.1016/j.rse.2004.12.009.

Xu, T., Guo, Z., Liu, S., He, X., Meng, Y., Xu, Z., Xia, Y., Xiao, J., Zhang, Y., Ma, Y., Song, L., 2018. Evaluating different machine learning methods for upscaling evapotranspiration from flux towers to the regional scale. J. Geophys. Res. 123, 8674–8690.

Xu, Y., Zhan, M., Cao, C., Tian, S., Ge, J., Li, S., Wang, M., Yuan, G., 2016. Improved water management to reduce greenhouse gas emissions in no-till rapeseed–rice rotations in Central China. Agric. Ecosyst. Environ. 221, 87–98.

Yang, Y., Anderson, M., Gao, F., Hain, C., Noormets, A., Sun, G., Wynne, R., Thomas, V., Sun, L., 2020. Investigating impacts of drought and disturbance on evapotranspiration over a forested landscape in North Carolina, USA using high spatiotemporal resolution remotely sensed data. Remote Sens. Environ. 238, 111018.

Yan, X., Akiyama, H., Yagi, K., Akimoto, H., 2009. Global estimations of the inventory and mitigation potential of methane emissions from rice cultivation conducted using the 2006 intergovernmental panel on climate change guidelines. Glob. Biogeochem. Cycles. https://doi.org/10.1029/2008gb003299.

Yan, X., Yagi, K., Akiyama, H., Akimoto, H., 2005. Statistical analysis of the major variables controlling methane emission from rice fields. Glob. Chang. Biol. 11, 1131–1141.

Zarco-Tejada, P.J., Ustin, S.L., 2001. Modeling canopy water content for carbon estimates from MODIS data at land EOS validation sites. In: IGARSS 2001. Scanning the Present and Resolving the Future. Proceedings. IEEE 2001 International Geoscience and Remote Sensing Symposium (Cat. No.01CH37217), vol.1., pp. 342–344.

Zhang, B., Tian, H., Ren, W., Tao, B., Lu, C., Yang, J., Banger, K., Pan, S., 2016. Methane emissions from global rice fields: magnitude, spatiotemporal patterns, and environmental controls. Glob. Biogeochem. Cycles 30, 1246–1263.

Zhang, C., Brodylo, D., Sirianni, M.J., Li, T., Comas, X., Douglas, T.A., Starr, G., 2021. Mapping CO2 fluxes of cypress swamp and marshes in the Greater Everglades using eddy covariance measurements and landsat data. Remote Sens. Environ. 262, 112523.

Zhang, G., Lu, Y., 2012. Bias-corrected random forests in regression. J. Appl. Stat. 39, 151–160.

Zhang, G., Xiao, X., Biradar, C.M., Dong, J., Qin, Y., Menarguez, M.A., Zhou, Y., Zhang, Y., Jin, C., Wang, J., Doughty, R.B., Ding, M., Moore 3rd, B., 2017. Spatiotemporal patterns of paddy rice croplands in China and India from 2000 to 2015. Sci. Total Environ. 579, 82–92.

Zhang, G., Xiao, X., Dong, J., Xin, F., Zhang, Y., Qin, Y., Doughty, R.B., Moore, B., 2020. Fingerprint of rice paddies in spatial–temporal dynamics of atmospheric methane concentration in monsoon Asia. Nat. Commun. 11, 1–11.

Zhang, W., Sun, W., Li, T., 2017. Uncertainties in the national inventory of methane emissions from rice cultivation: field measurements and modeling approaches. Biogeosciences 14, 163–176.

Zhang, Z., Poulter, B., Knox, S., Stavert, A., McNicol, G., Fluet-Chouinard, E., Feinberg, A., Bousquet, P., Canadell, J.G., Ganesan, A., Hugelius, G., 2021. Anthropogenic emission is the main contributor to the rise of atmospheric methane during 1993–2017. Nat. Sci. Rev. 9 (5), nwab200. https://doi.org/10.1093/nsr/nwab200.