JGR Biogeosciences

RESEARCH ARTICLE
10.1029/2020JG006002

A Remote Sensing Technique to Upscale Methane Emission Flux in a Subtropical Peatland

Caiyun Zhang[1], Xavier Comas[1], and David Brodylo[1]

[1]Department of Geosciences, Florida Atlantic University, Boca Raton, FL, USA

Key Points:
- A remote sensing upscaling approach is developed to model CH_4 emission flux in the subtropical Everglades wetland
- Models are constrained with sparse in situ measurements to produce consistent estimates for a dry and wet season
- CH_4 emissions in the wet season account for a threefold to fourfold increase in flux compared to the dry season

Correspondence to:
C. Zhang,
czhang3@fau.edu

Citation:
Zhang, C., Comas, X., & Brodylo, D. (2020). A remote sensing technique to upscale methane emission flux in a subtropical peatland. *Journal of Geophysical Research: Biogeosciences*, *125*, e2020JG006002. https://doi.org/10.1029/2020JG006002

Received 3 AUG 2020
Accepted 16 SEP 2020
Accepted article online 24 SEP 2020

Author Contributions:
Conceptualization: Caiyun Zhang, Xavier Comas
Data curation: Caiyun Zhang, Xavier Comas, David Brodylo
Formal analysis: Caiyun Zhang
Funding acquisition: Xavier Comas
Investigation: Caiyun Zhang
Methodology: Caiyun Zhang
Resources: Caiyun Zhang, Xavier Comas
Writing - original draft: Caiyun Zhang
Writing – review & editing: Xavier Comas, David Brodylo

Abstract Quantification of methane (CH_4) gas emission from peat is critical to understand CH_4 budget from natural wetlands under a climate warming scenario. Previous studies have focused on prediction and mapping of CH_4 emission flux using process-based models, while application of statistical-empirical models for upscaling spatially sparse in situ measurements is scarce. In this study, we developed an empirical remote sensing upscaling approach to estimate CH_4 emission flux in the Everglades using limited in situ point-based CH_4 emission flux measurements and Landsat data during 2013–2018. We spatially and temporally linked in situ data with Landsat surface reflectance based on temporally composite data sets and developed an object-based machine learning framework to model and map CH_4 emission flux. An ensemble analysis of two machine learning models, *k*-Nearest Neighbor (*k*-NN) and Support Vector Machine (SVM), shows that the upscaling approach is promising for predicting CH_4 emission flux with a R^2 of 0.65 and 0.87 based on a fivefold cross-validation for a dry season and wet season estimation, respectively. We generated emission flux map products that successfully revealed the spatial and temporal heterogeneity of CH_4 emission within the dominant freshwater marsh ecosystem in the Everglades. We conclude that Landsat is promising for upscaling and monitoring CH_4 emission flux and reducing the uncertainty in emission estimates from wetlands.

1. Introduction

Methane (CH_4) is one of the most important greenhouse gases impacting the global climate system (Dean et al., 2018). It plays a critical role in the Earth's carbon cycle and largely contributes to global warming due to a positive feedback between its production and air temperature (Dean et al., 2018). While CH_4 emissions come from multiple natural and anthropogenic sources, wetlands are considered the largest natural CH_4 source and account for 20–40% of the total global CH_4 emissions (Saunois et al., 2020). Quantification of wetland CH_4 emission is therefore essential to understand its role in carbon cycling and climate warming.

CH_4 is generated in wetlands when organic matter undergoes decomposition in the anaerobic conditions and is released to the atmosphere through three pathways: bubble ebullition, molecular diffusion, and transport via plants (Wang et al., 1996). Studies of CH_4 in wetlands concentrate on CH_4 emission estimation (Bartlett & Harriss, 1993; Wang et al., 1996), mechanisms of production and consumption (Segers, 1998), and its role in the global climate system (Dean et al., 2018). Traditionally, CH_4 flux data are acquired through time-consuming and labor-intensive point-based approaches at 1-m scale like chambers (e.g., Morin et al., 2017; Whiting & Chanton, 2001) and recent gas traps fitted with time-lapse cameras (Comas & Wright, 2012; Wright et al., 2018), or from eddy covariance towers at 100- to 1,000-m scale (e.g., Morin et al., 2017). In situ methods are spatially constrained and difficult in remote locations or harsh environments like wetlands. These factors in combination with the sensitive nature of CH_4 dynamics to environmental variables (e.g., temperature and atmospheric pressure) make characterization of the spatial and temporal variability of CH_4 emissions challenging. Complicated bottom-up/process-based models and top-down/atmospheric inversion approaches have been developed as an alternative to in situ approaches with varying success (e.g., Bloom et al., 2012; Miller et al., 2016; Zhang, Tian, et al., 2017; Zhang et al., 2012; Zhu et al., 2014). Large uncertainties in CH_4 emission estimates have been identified from these models due to insufficient in situ flux measurements, undefined model structures and parameterizations, and inaccurate wetland data sets (Saunois et al., 2020; Xu et al., 2016). In addition, estimations from these two types of models are not independent of each other since the top-down estimates usually rely on the prior

estimations obtained from bottom-up models (Peltola et al., 2019). CH_4 emission information has also been retrieved from the Greenhouse gases Observing Satellite (GOSAT), but large discrepancies are found between satellite observations and model simulations (Parker et al., 2018).

Statistical-empirical models are another alternative to estimate CH_4 emission flux based on the association between flux in wetlands and driver variables. For example, Peltola et al. (2019) applied a Random Forest machine learning approach to upscaling eddy covariance CH_4 flux observations in northern wetlands using 500- to 1,000-m Moderate Resolution Imaging Spectroradiometer (MODIS) products, wetland maps, and other gridded data sets. Machine learning-based models have proven valuable in gap filling of eddy covariance CH_4 flux data (e.g., Dengel et al., 2013; Kim et al., 2020; Morin et al., 2017) and upscaling eddy covariance observations in northern wetlands (Peltola et al., 2019). However, machine learning models have been barely applied for upscaling point-based CH_4 flux measurements using relatively high-resolution Landsat products for local/regional emission mapping. To our knowledge, a high-resolution (e.g., 30-m) CH_4 emission flux data set of wetlands like the Florida Everglades has not been generated.

The main objective of this study is, for the first time, to model CH_4 emission rates from peat soils using an ensemble analysis of machine learning models by combining 30-m Landsat data with a 6-year array of in situ measurements coarsely distributed across six study sites in the Florida Everglades. Optical remote sensing image products such as 500-m MODIS and 1,000-m Advanced Very High Resolution Radiometer (AVHRR) have been used for assisting with total CH_4 emission estimates from wetlands by providing wetland coverage, wetland habitat types, or vegetation fraction (e.g., Bansal et al., 2018; Takeuchi et al., 2003). Sensors like MODIS and AVHRR cannot properly capture the spatial variability in CH_4 emission flux in wetlands where emission flux is markedly heterogeneous. Finer-resolution 30-m Landsat data are suitable for local/regional applications and have been successfully used for vegetation change analysis (Zhang, Smith, et al., 2017) and marsh biomass modeling (Zhang et al., 2018) in the Florida Everglades. Bartlett et al. (1989) made the first effort to apply Landsat data for mapping the spatial variability of CH_4 flux in the Everglades National Park by assigning the field measured CH_4 flux to each habitat type classified from Landsat imagery. All these wetland CH_4 studies assume that in situ measured CH_4 emission rate is spatially constant either within a habitat type or entire wetland region for estimating the regional/global total CH_4 budget. This may lead to large uncertainties for regions delineated as the same habitat type (e.g., freshwater marshes) but have different composition of water, plants, or peat soils with varying emission rate. No attempt has been made to map CH_4 emission rate for a detailed delineation of the spatial pattern of CH_4 by directly linking the spectral reflectance from optical sensors with in situ CH_4 flux measurements. This is mainly attributed to the spatially sparse distribution of in situ measurements and high temporal variability of CH_4 fluxes, which makes a spatial and temporal linkage of in situ CH_4 data and satellite data challenging. Here we developed an innovative scheme to link these two data sets, which provided a unique opportunity to upscale the in situ point measurements, and the potential to characterize the spatial and temporal variability in CH_4 flux using 30-m Landsat data for the Florida Everglades. The specific objectives of this study were to (1) develop an ensemble approach to predict CH_4 emission flux using models developed from sparse in situ flux measurements and 30-m Landsat data; (2) assess the applicability of machine learning algorithms for CH_4 flux estimation; and (3) generate a spatially explicit CH_4 flux data set for the Florida Everglades, which would improve regional and global CH_4 emission estimates.

2. Study Area and Data

2.1. Study Area

Our study area is the Water Conservation Areas (WCAs) located in the Florida Everglades, which is the largest subtropical wetland in the United States (Figure 1). The Florida Everglades has been severely modified in the past century, resulting in many environmental issues in South Florida. Currently, the largest hydrological restoration project, the Comprehensive Everglades Restoration Plan (CERP), is underway to restore, preserve, and protect the Everglades (Everglades Restoration, 2020). WCAs with an approximate size of 3,457 km^2 are vast tracts of remnant Greater Everglades, which starts from Central Florida and ends at the tip of South Florida (Figure 1). WCAs are further divided into WCA-1, WCA-2A, WCA-2B, WCA-3A, and WCA-3B by levees and canals and are characterized by two peat soils: Everglades peat and Loxahatchee peat (Gleason & Stone, 1994). These WCAs were established to provide flood control and

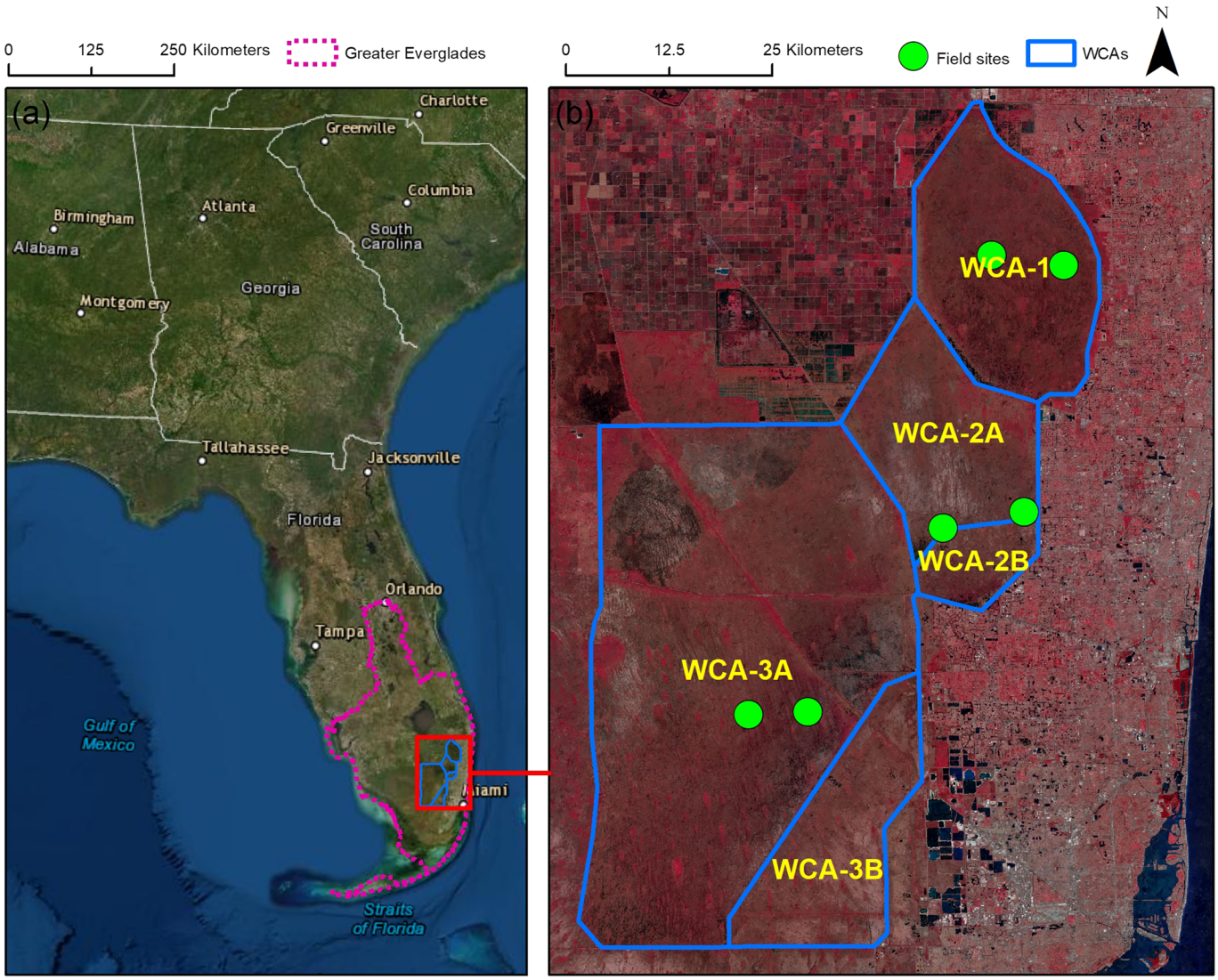

Figure 1. Study area (a) in the Greater Everglades, South Florida, USA, and (b) Water Conservation Areas (WCAs) displayed as a color infrared composite Landsat 8 image collected in 2018, and field sites for measuring CH_4 emission using modified gas traps (see section 3 for details).

water supply for South Florida and have been a principal beneficiary of the restoration. The study area is dominated by freshwater marshes (sawgrass and graminoid), which consist of ~92.3% of the total WCAs. South Florida is characterized by two seasons: the wet season (May–November) with a high temperature and abundant rainfall and the dry season (December–April) with a mild temperature and reduced precipitation. The annual mean high and low temperature is 83.4°F and 69.3°F, respectively. The average annual rainfall is around 60 in. with 70% received in the wet season and 30% received in the dry season.

2.2. Data

Data used in this study include in situ CH_4 emission rates over six sites during 2012–2018, as shown in Figure 1b; Landsat 8 Level 2 surface reflectance products; and a land use land cover data set. In situ point-based CH_4 emission rate data were collected as follows: 30 records from two sites at WCA-1 between 23 September 2016 and 13 October 2018; 49 records of two sites from WCA-2 between 1 July 2013 and 12 November 2014; and 56 records of two sites from WCA-3 between 25 June 2012 and 12 November 2014. In all cases, sites corresponded to freshwater marsh-dominated areas characterized by either Everglades or Loxahatchee peat. Detailed description of each site can be found in Wright and Comas (2016). Flux

measurements were estimated from entrapped biogenic gas volumes using modified gas traps and gas chromatography as described in section 3.

Geometrically and atmospherically corrected 30-m Landsat 8 Operational Land Imager (OLI) Level 2 surface reflectance products spanning 2013 to 2018 stored in Google Earth Engine (GEE) were used to spatially and temporally match with the in situ estimations for remote sensing model development. Seasonality is a major factor impacting CH_4 emission rate in the Everglades (Comas & Wright, 2014). Therefore, in addition to the Landsat data for model development, we also selected the 2018 Landsat data collected at the end of dry season (March–April) and wet season (October–November) and generated two seasonal composite images for mapping CH_4 emission rate for the entire WCAs. Six spectral bands (three visible, one near infrared, and two shortwave infrared) were used because these bands were valuable for mapping the Florida Everglades (Zhang, Smith, et al., 2017). Landsat 5 and 7 data products were also available for some periods but were not used in this study to reduce uncertainties in reflectance caused by multisensors on different platforms (Flood, 2014). Landsat 8 products are only available since 2013; thus, in situ measurements in 2012 were excluded. Other land cover types (shrubs, scrub, broadleaf forests, spoil areas, and canals) are also present in the study domain. Here we used a 2016 land use land cover data set to mask out nonfreshwater marshes so that the mapping was constrained to freshwater marshes only because our field sites were set up in the freshwater marshes. The land cover land use data set was produced by the South Florida Water Management District through manual interpretation of aerial photography techniques reported in Rutchey et al. (2008).

3. Methodology

3.1. In Situ Data Collection Approach

For in situ CH_4 emission rate estimation, we applied a modified gas trap method in the field and estimated gas fluxes from weekly to biweekly volumes of biogenic gas releases (Wright et al., 2018). We mounted traps in wooden platforms with pilings driven to the mineral soil to avoid soil disturbance during data collection. Gas traps consisted of inverted 20-cm plastic funnels connected to a clear PVC cylinder extending above the water surface. We analyzed collected gases by quantifying CH_4 and CO_2 composition using gas chromatography and finally converted weekly volumes into CH_4 emission rates (expressed as mg CH_4/m^2/day) to calibrate and validate the upscaling remote sensing models.

3.2. Remote Sensing Upscaling Framework

To upscale the in situ point-based CH_4 emission, we developed a framework to model and map CH_4 emission rate and predicted the total daily emission of CH_4 from the WCAs using Landsat products and in situ measurements, as displayed in Figure 2. We first generated composite data sets of in situ CH_4 emission rate of each site and Landsat reflectance for a specific time window (e.g., season) and then spatially and temporally matched two data sets, leading to a matched data set for CH_4 model development. We selected two contemporary machine learning algorithms, *k*-Nearest Neighbor (*k*-NN) and Support Vector Machine (SVM), and developed the empirical models with the in situ measurements as the dependent variable, and reflectance value as the independent variable. Our preliminary examination results showed these two machine learning algorithms worked better than other machine learning algorithms such as Artificial Neural Network and Random Forest. The performance of each model was evaluated using four statistical metrics: coefficient of determination (R^2), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Relative RMSE (RRMSE). If a model was found acceptable ($R^2 > 0.6$), it was used to estimate CH_4 emission rate for the entire study area. If two models were comparable, an object-based ensemble method was applied to predict and map the final CH_4 emission rate by combining the outputs of two models. The daily total emission of the entire WCAs was estimated by

$$\sum_{i=1}^{N} A_i E_i, \tag{1}$$

where N is the total number of image objects and A_i and E_i is the area and the estimated emission rate of the image object i, respectively. Major steps in the framework include data matching, model development, accuracy assessment, and CH_4 estimation and mapping. They are described below.

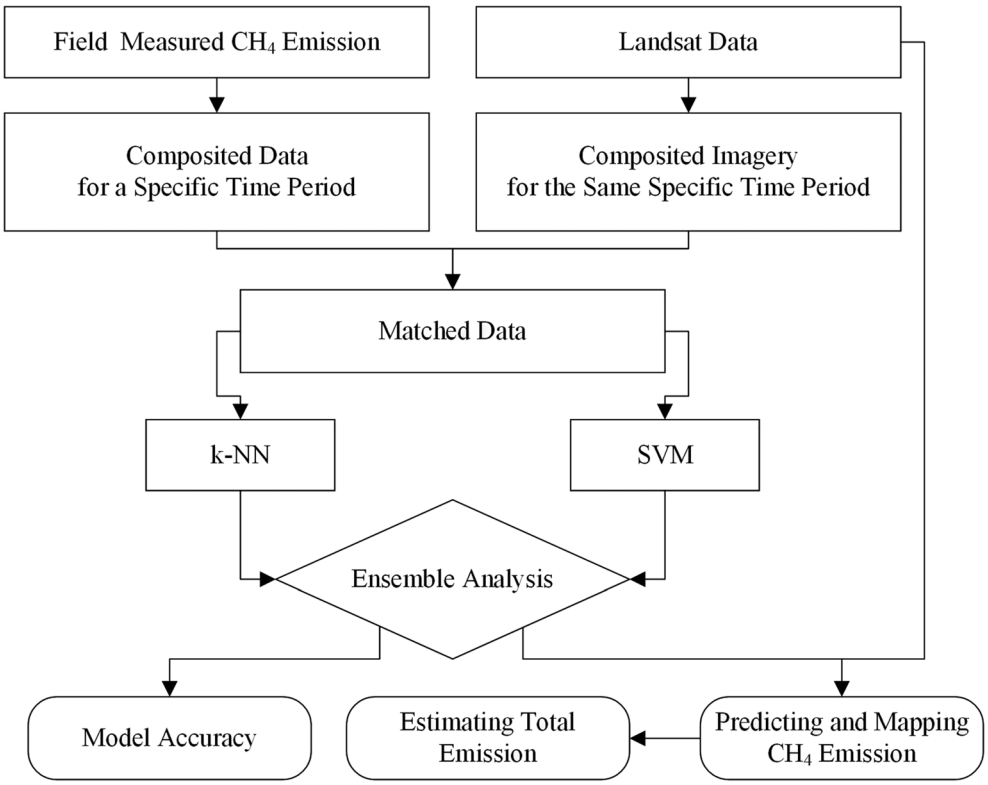

Figure 2. Methodology framework for upscaling in situ point-based CH_4 measurements using Landsat reflectance products.

3.2.1. Data Matching

In situ CH_4 emission rates were estimated for specific dates discontinuously during 2012–2018. Temporally, it is difficult to match the in situ estimation with Landsat data for a specific date because Landsat 8 has a temporal resolution of 16 days, and the CH_4 emission is highly dynamic. Cloud cover is another challenge to use a specific scene of Landsat in South Florida and other subtropical/tropical regions (Zhang, Smith, et al., 2017). To effectively match in situ estimations with Landsat data, multitemporal satellite images can be composited over a season (or other time period) to produce imagery that is representative of that period (Flood, 2013; Holben, 1986). This technique can largely reduce contamination by cloud or other problems and has been operationally applied in satellite product generation such as MODIS and AVHRR (Flood, 2013; Holben, 1986). This idea can be also transferred to the in situ CH_4 emission data set by calculating the average emission flux for a specific time period (e.g., at the quarterly or seasonal time scale), which is particularly relevant when inferring seasonal/annual dynamics of CH_4 flux. Here, the in situ daily CH_4 emission estimations were averaged to a specific period (season) using data collected during this specific period. The temporally averaged in situ data set was then spatially matched to the averaged Landsat imagery for the same time period. The composites can be the average, maximum, or minimum value of the time period. Here we selected the commonly used average value of in situ CH_4 estimations and spectral reflectance. This matching scheme could reduce data noise in both in situ and satellite data collection and cloud-contamination issues of Landsat. It therefore seems reasonable to spatially and temporally match two data sets acquired from different approaches and different times. This idea was expected to solve the bottleneck issue for upscaling in situ CH_4 measurements using satellite data. Emission flux has a high variation between the two seasons in WCAs; we thus separated the matched data set into two parts with one representing the dry season and the other representing the wet season. Different models were developed and applied for the two seasons.

3.2.2. Model Development, Accuracy Assessment, and CH_4 Emission Mapping

A previous study has shown that machine learning regression algorithms are powerful and valuable for aboveground marsh biomass estimation in the Everglades, and they have achieved significantly better

results than the traditional Multiple Linear Regression (MLR) approach (Zhang et al., 2018). Therefore, we selected two machine learning algorithms (*k*-NN and SVM) for CH_4 emission estimation and provided the comparison results in this study. *k*-NN is a relatively simple machine learning approach, and it estimates the prediction as a weighted average value with *k* spectrally nearest neighbors using a weighting method. SVM is a statistical learning approach (Vapnik, 1995). It transforms the input data into a high-dimensional feature space using a nonlinear kernel function to minimize training error and complexity of the model. These two machine learning models have been widely used in various fields such as remote sensing classification and modeling, as reviewed by Chirici et al. (2016) for *k*-NNs and Mountrakis et al. (2011) for SVMs. Each algorithm requires input parameters that might impact model performance. For example, SVM needs inputs of the selected kernels, precision, and penalty parameters. Here, all these inputs were tuned using an experimenter function integrated in an open-source machine learning modeling software package WEKA (Hall et al., 2009) to find an optimal model for CH_4 estimation. The experimenter function in WEKA can effectively tune an algorithm by changing the required input parameters and find the best model based on cross-validation results. For our matched data sets, *k* was specified as 2 in *k*-NN, and the radial basis function kernel in SVM was selected with the gamma as 5.0 and C as 3.0.

Previous studies also suggest that an ensemble approach is more robust for the final prediction than the application of a single model because each model has pros and cons (Zhang et al., 2018). Models can have a comparable performance but different predictions. Following the ensemble approach developed for aboveground sawgrass marsh biomass estimation in Zhang et al. (2018), the final CH_4 emission rate was estimated as

$$P = \sum \frac{r_i p_i}{\sum_{i=1}^{M} r_i}, \tag{2}$$

where P is the final estimation, p_i is the prediction from model i, r_i is the correlation coefficient (r) between in situ CH_4 measurements and predicted emission from model i, and M is the total number of models to be used in the ensemble approach.

We evaluated the performance of each model using the *k*-fold cross-validation technique, which splits matched samples into *k* subsets first, and then iteratively, one subset is used to test/validate the model, and other remaining subsets are used to train/calibrate the model. This technique can avoid the overfitting issue. Here *k* was set to 5. After iterations, predictions were generated for all reference samples, which were then used to calculate the statistical metrics, including R^2, MAE, RMSE, and RRMSE. The definition and calculation of these statistical metrics can be found in statistics textbooks and software packages.

If a model was acceptable, the emission rate was estimated for the entire study area using a Landsat composite or a scene collected for a specific date. The final estimation can be mapped at the pixel level, but in practice, researchers are less interested in a single pixel than a composition or a configuration of many pixels that comprise a meaningful landscape within an image. Mapping a patch/region is more useful. Thus, an object-based mapping approach was applied here. The object-based approach analyzes and maps objects rather than pixels, which reduces the "salt-and-pepper" effects in mapping heterogeneous landscapes and enhances the analysis accuracy in wetlands (Dronova, 2015). The remotely sensed pixel is affected by surrounding pixels, and speckles frequently occur due to intrinsic spectral variations. Using the averaged values of all pixels within an object/patch can reduce this effect (Dronova, 2015). Here we produced objects from the Landsat imagery using the multiresolution segmentation algorithm in eCognition Developer 9.0 (Trimble, 2014). The required inputs for this algorithm include a scale parameter, color/shape weights, and smoothness/compactness weights. Following Zhang, Smith, et al. (2017) using Landsat time series data for analyzing vegetation change in WCA-2A, we set color/shape weights to 0.9/0.1 so that spectral information would be considered most heavily for segmentation, smoothness/compactness weights to 0.5/0.5 so as to not favor either compact or noncompact segments, and the scale parameter to 50 that was identified as the optimal value for this study area. After segmentation, we extracted the mean reflectance of each object as the input for the prediction of CH_4 emission rate. The size of each object and the total number of objects were determined using Geospatial Information Science (GIS) methods, which were used in Equation 1 for the total daily emission estimation. Detailed description for generating image objects from Landsat data in the Florida Everglades can be found in Zhang, Smith, et al. (2017) and Zhang et al. (2018). Consequently, object-based CH_4 emission rate maps were produced for the dry and wet season of 2018, respectively.

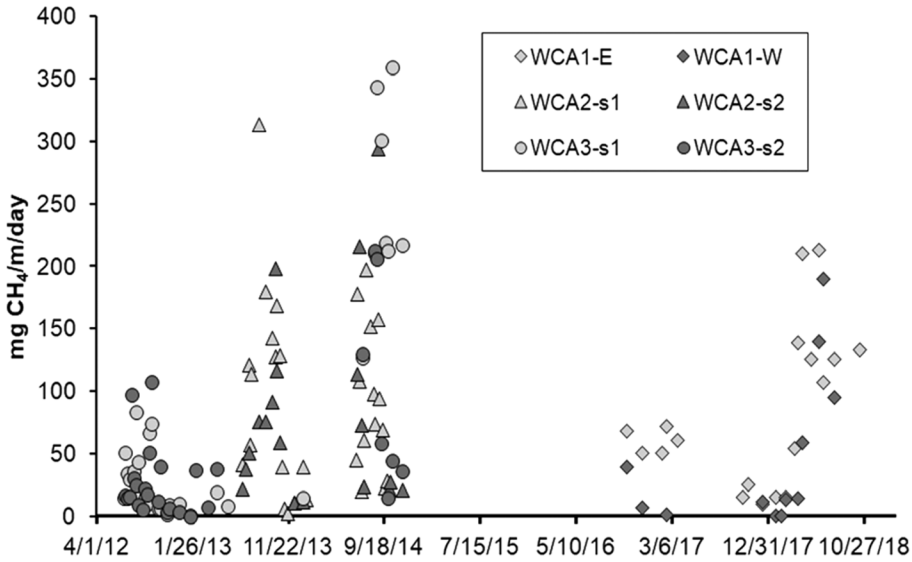

Figure 3. CH_4 flux estimates inferred from modified in situ gas traps (see section 3 for further details) at WCAs between 2012 and 2018. Gaps in data correspond to periods where data acquisition was not feasible due to issues with either site accessibility (i.e., conditions too dry for navigation) or funding availability.

4. Results

4.1. In Situ CH_4 Measurements and Model Performance

Based on in situ measurements, CH_4 emission rate varied from 0.2 to 213.5 mg CH_4/m^2/day at WCA-1, 6.1 to 216.1 mg CH_4/m^2/day at WCA-2, and 0.1 to 359.0 mg CH_4/m^2/day at WCA-3 (Figure 3). A distinct temporal difference in emission between two seasons was observed. In WCA-1, the averaged flux in the wet season was 109.3 mg CH_4/m^2/day, while in the dry season the flux was reduced to 40.8 mg CH_4/m^2/day. In WCA-2, the averaged emission varied from 16.5 mg CH_4/m^2/day in the dry season to 95.6 mg CH_4/m^2/day in the wet season. WCA-3 revealed a similar large change between two seasons with flux varying from 13.5 mg CH_4/m^2/day in the dry season to 75.3 mg CH_4/m^2/day in the wet season. The overall averages for all WCAs accounted for 23.6 and 93.4 mg CH_4/m^2/day during the dry and wet seasons, respectively, indicating a four-fold increase in flux in the wet season compared to the dry season.

Experimental analysis of two machine learning models revealed one outlier in the dry season and two outliers in the wet season. These outliers were identified based on the regression line on the scatter plot of estimated and measured flux. Outliers are points far away from the regression line. The outliers might be attributed to errors from either the field measurements or the Landsat data sets. Table 1 lists the model performance with the outliers excluded. The model worked better for the wet season estimation with a R^2 of 0.87 from *k*-NN and 0.75 from SVM, while the dry season models produced a R^2 of 0.63 from *k*-NN and 0.70 from SVM. This was expected because during the wet season, a stronger reflectance signature occurs due to the nature of vegetation phenology in the Everglades (Zhang et al., 2018). Also, there were fewer field samples

Table 1
Model Performance for Estimating CH_4 Emission Flux in WCAs

	Dry season			Wet season		
Statistical metrics	*k*-NN	SVM	EA	*k*-NN	SVM	EA
R^2	0.63	0.70	0.65	0.87	0.75	0.87
MAE (mg CH_4/m^2/day)	12.4	9.2	6.7	18.4	25.5	18.4
RMSE (mg CH_4/m^2/day)	17.2	12.0	15.7	26.7	33.7	24.3
RRMSE (%)	61.0	48.2	60.3	19.2	24.2	17.5

Note. EA = ensemble analysis.

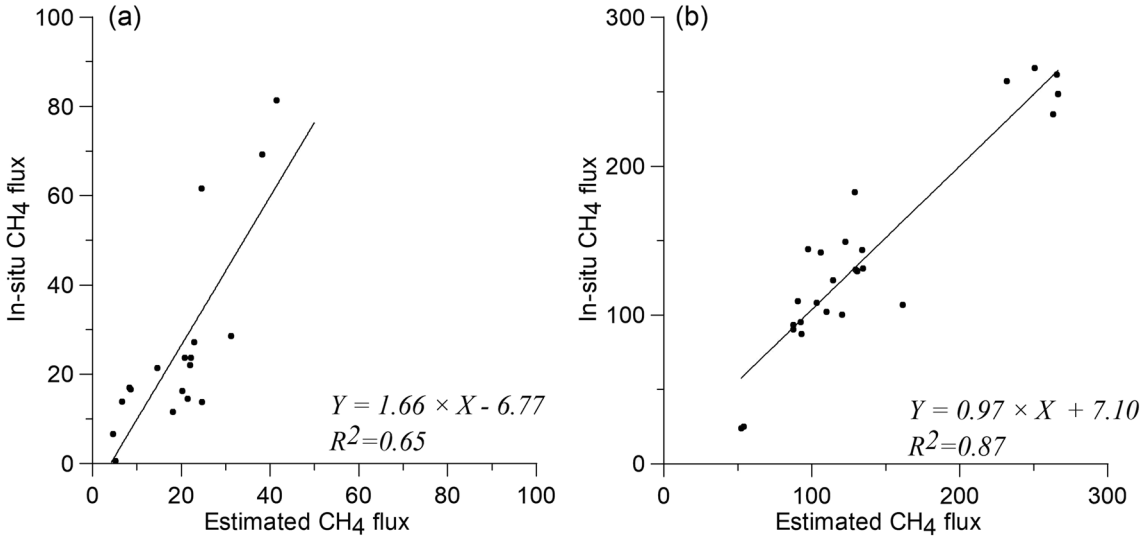

Figure 4. Scatter plot and regression between in situ measured and remote sensing estimated CH_4 emission flux (mg CH_4/m^2/day) for the (a) dry season and (b) wet season.

in the dry season due to the difficulties accessing the sites via airboat under dry conditions. Less training samples can reduce the model performance. They all produced better results than the MLR approach, which created a very low R^2 of less than 0.15 (results of MLR were not shown). Both SVM and *k*-NN achieved encouraging results, but negative values could be produced from the SVM model. Thus, an ensemble analysis of the outputs from the two models was conducted, and it slightly improved the estimation for the wet season with a R^2 of 0.87, MAE of 18.4 mg CH_4/m^2/day, RMSE 24.3 mg CH_4/m^2/day, and RRMSE of 17.5%. Scatter plots of the predicted CH_4 emission rate against in situ measures are provided in Figure 4. This is a further procedure to evaluate model performance. The results showed that models slightly overestimated the emission rate for the wet season while underestimated the flux for the dry season.

4.2. CH_4 Emission Rate Mapping and Total Emission Estimation

Ensemble analysis produced an encouraging estimation and thus was applied to estimate CH_4 emission rate for the entire study area, as shown in Figure 5. In general, seasonal emission maps demonstrate the sharp difference between the two seasons with overall lower fluxes in the dry season and higher fluxes in the wet season. During the dry season, most regions show a flux of less than 40 mg CH_4/m^2/day, as compared to larger than 140 mg CH_4/m^2/day during the wet season. Furthermore, the flux pattern in the wet season is more heterogeneous with values between 14 and 306 mg CH_4/m^2/day that contrast with a range between 0.3 and 100 mg CH_4/m^2/day in the dry season. Based on the maps, the average emission rate for the entire WCA was 32 mg CH_4/m^2/day for the dry season and 170 mg CH_4/m^2/day for the wet season. Estimated average values around the field sites were 25.3 mg CH_4/m^2/day (dry season) and 91.8 mg CH_4/m^2/day (wet season), representing a 72% increase in fluxes during the wet season. These values are consistent with the in situ measurements, which estimated an average of 23.6 and 93.4 mg CH_4/m^2/day for the dry and wet seasons or a 75% increase in fluxes during the wet season. The daily total emission was estimated as 113 and 597 tons in the dry season and wet season, respectively.

5. Discussion and Implications

Our field measured flux data sets revealed a strong temporal variability in CH_4 emission flux between the two seasons (dry vs. wet). While in high-latitude wetlands, intra-annual variation in CH_4 emission is mainly controlled by temperature (Whalen, 2005), in temperate wetlands, hydrological conditions (e.g., water level and soil moisture content) are the major factor influencing CH_4 emission flux (Bartlett & Harriss, 1993). In the Florida Everglades, studies have shown an increase in overall CH_4 emissions during the wet season (e.g., Comas & Wright, 2014; Villa & Mitsch, 2014), as supported by this study. Increased precipitation

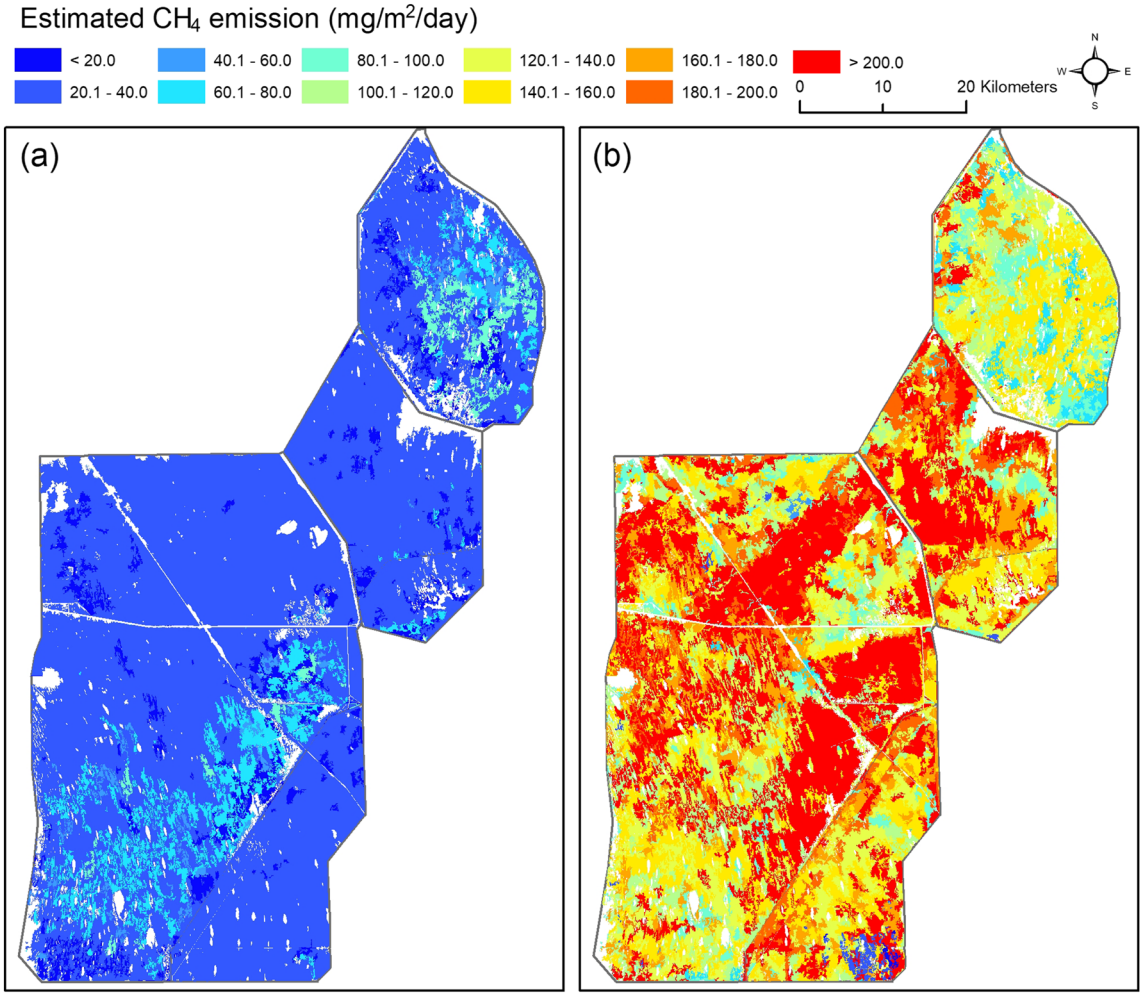

Figure 5. Maps of estimated CH_4 emission flux at the end of (a) dry season (March–April) and (b) wet season (October–November) in 2018 using the ensemble approach.

during the wet season of the Everglades largely affects the area of surface inundation, water table depth, and soil moisture content, which in turn promotes methanogenesis and enhances CH_4 production/emissions in peat soils (Bubier et al., 1993; Moore et al., 1990; Whalen, 2005).

Our field measurements also showed a pronounced spatial variability in CH_4 emission across sites. This can be partly attributed to the differences in peat type (i.e., Everglades vs. Loxahatchee peat) and associated differences in soil physical properties (such as porosity) and the thickness of the peat column (Comas & Wright, 2014; Wright & Comas, 2016). The composition of vegetation/habitat type is recognized as another major factor controlling the spatial variation of CH_4 emission in the Everglades based on previous field-based studies (Bartlett et al., 1989; Villa & Mitsch, 2014). Vegetation conditions (e.g., health status, density, and biomass) can directly impact peat accumulation and subsequent CH_4 production/release. For example, a study in an Everglades *Caldium* marsh has shown a strong correlation between CH_4 flux and plant biomass (Whiting et al., 1991). Given the coupling effects between vegetation, soil, and water conditions in wetlands to determine CH_4 production/emission, it seems logic to use optical sensors to predict CH_4 emission. This encourages us to explore the potential of using optical sensors to upscale field-based CH_4 flux measurements with the aim to reduce labor-intensive field work and to infer results into locations with limited or no flux measurements.

Spectral reflectance of vegetation from optical sensors can indirectly indicate subsurface marsh soil properties (Zhang et al., 2019). Landsat data have also proven powerful for characterizing the spatial and temporal

aspects of water inundation in the Everglades National Park (Rose & Rosendahl, 1983), mapping vegetation composition in WCA-2A (Zhang, Smith, et al., 2017), and quantifying marsh biomass in the coastal Everglades (Zhang et al., 2018). The capacity of Landsat for "seeing" vegetation composition, and ability for monitoring vegetation condition, phenology, and water inundation via regular revisiting, allows an indirect estimation of CH_4 emission flux in wetlands. For all these reasons, the approach presented in this study shows the potential of using Landsat for monitoring CH_4 flux in wetlands by integrating longer in situ time series flux data collected at limited sites with Landsat data.

The composite approach is valuable to reduce the cloud contamination issues found in satellite observations and mitigate some challenges in the upscaling. Our approach successfully revealed the spatial and temporal variation of flux within the dominant marsh ecosystem in the Everglades. A heterogeneous flux pattern was illustrated within the same marsh habitat type due to the heterogeneity of marsh density/height, species composition, and inundation condition. The models also captured well the increase in emissions during the wet season. Landsat data could explain 65% and 87% variance of field measured emission flux for the dry and wet season, respectively, using the ensemble approach. The low estimated emission flux during the dry season was not only consistent with our field data but also generally coherent with the flux range (7.5–81.9 mg CH_4/m^2/day) of inundated marshes in the Everglades National Park reported by Bartlett et al. (1989). The predicted threefold to fourfold increase in emissions during the wet season seems reasonable considering how seasonality affects CH_4 dynamics.

Large uncertainty may be caused by assigning one flux value to a single wetland type for global CH_4 emission estimates or to a single habitat type for regional emission estimates. For example, in Matthews and Fung (1987), the Florida Everglades represents a relatively homogeneous physical and biological environment; it is identified as nonforested swamp and assigned a CH_4 flux rate of 120 mg CH_4/m^2/day for global wetland CH_4 emission estimates. In Bartlett et al. (1989), a single flux value is assigned to each habitat type (e.g., sawgrass, mangrove, and hardwood hammock) for mapping the spatial variation of flux at the Shark River Slough test site in the Everglades to reduce the uncertainty for regional/global emission estimates. Our approach refines the heterogeneity of emission flux within the dominant freshwater marsh ecosystem to further enhance the local/regional/global wetland CH_4 emission estimates. The Landsat-based CH_4 flux products from our study provide an independent data set for assessing the large-scale process-based models and offer a linkage between the in situ-based local measurements and regional/global process-based estimations. Furthermore, our results indicate that changes in the accumulation and release of biogenic gases in peat soils are closely related to detectable changes in surface reflectance. This suggests that our approach could also potentially be applied to predict how specific changes in environmental variables (such as temperature or atmospheric pressure) may induce changes in biogenic gas accumulation and release at a scale of measurement not feasible for most field-based approaches.

Application of nonparametric machine learning regression models was also beneficial in the upscaling. A preliminary assessment of the linear MLR approach produced a poor estimation, indicating that the relationship between emission flux and reflectance was nonlinear due to the complicated coupling effects of vegetation, hydrology conditions, and soil properties on CH_4 production/release. Here both *k*-NN and SVM were successful and largely improved estimation with a R^2 larger than 0.6. Each algorithm has its pros and cons. For example, *k*-NN is relatively simple, but it requires users to select an appropriate distance metric to assess the similarity and weighting scheme to combine neighbors for predictions (Chirici et al., 2016). SVM can generalize the model well with limited training samples, but it also suffers from parameter specification issues that can largely impact the predictions (Mountrakis et al., 2011). Both algorithms produced a comparable accuracy based upon the statistical metrics, but they showed a diversity in the prediction. This was caused by the discrepancies in each algorithm. *k*-NN searches the best spectral match to estimate input, while SVM looks for an optimal hyperplane to minimize training errors. An ensemble analysis to combine the outputs of two models made the prediction more robust than the application of a single model. Therefore, the application of ensemble analysis was valuable in the modeling.

Multiple sources of error have been identified in the developed procedure, including the limited number of field sites, a lack of other environmental/hydrological data sets, limitation of Landsat data, and machine learning modeling. First, our in situ data were collected from six representative sites in WCAs. In theory, stratified sampling and planning schemes should be used to expand field data collection, but in practice,

setting up more in situ platforms and collecting CH_4 data representative of large wetlands like the Everglades is very challenging and time consuming, particularly given the difficult accessibility of some areas. Therefore, our sparse in situ measurements in terms of spatial coverage (six sites spread over almost 3,500 km^2) and temporal density (data sets scattered over 6 years) are considered the major uncertainty source in the flux mapping. This limitation was well illustrated by the relatively poor model performance in the dry season compared to the wet season since less field data were collected during the dry season. In addition, our measurements were all conducted in slough marsh areas, and a lack of data over ridge areas may also cause bias toward the flux mapping due to the characteristic "ridge-and-slough" landscape in the Everglades ecosystem. Increasing the number of field sites and/or longer time series data would make the estimation more accurate and robust. Alternatively, the use of more efficient methods like ground-penetrating radar (Comas & Wright, 2014; Wright & Comas, 2016) to infer gas dynamics in the Everglades can also increase spatial coverage. Despite this limitation, the approach presented here can serve as a testing ground to further develop larger-scale flux release models. Second, Landsat has the capacity to characterize vegetation, soil, and inundation conditions in the Everglades; however, the optical sensors cannot indicate water level in the wetlands. Water depth has been recognized as a critical factor dictating CH_4 flux in tropical and subtropical wetlands (Bartlett & Harriss, 1993). Including water depth data sets as well as other environmental variables (e.g., temperature) in the model might improve the estimation. Presently, the U.S. Geological Survey (USGS) provides water level products in the Everglades (USGS/EDEN, 2020) that could be potentially added into future models. Third, the in situ platform for each site was a few meters in length, and data from the platform were spatially matched to the 30-m Landsat data by assuming that CH_4 emission rate within the 900 m^2 pixel or mapped object was constant. Applying finer spatial resolution remote sensing data products like 10-m Sentinel 2A might improve the modeling and mapping result. Lastly, although the object-based ensemble approach achieved high accuracy for emission estimates based on two machine learning algorithms, uncertainties from data processing are inevitable, including model parameter specification, combining scheme used in the ensemble analysis and image segmentation for object-based mapping. Evaluating the effects of model selection and parameter inputs in the modeling procedure deserves further investigations.

6. Summary and Conclusions

We modeled CH_4 emission rate in a freshwater peatland by combining in situ flux measurements from gas traps with satellite observed surface reflectance using a machine learning based ensemble approach. We conclude that Landsat data are valuable for upscaling in situ CH_4 emission rate using contemporary machine learning modeling techniques and object-based mapping approach. The developed methodology framework can be applied to other wetlands, but a direct application of the developed models might be problematic because the model was calibrated by site-specific in situ data. An inclusion of more ancillary data such as water table, temperature, and precipitation might improve the estimation. We anticipate this work can stimulate further research on the application of optical satellite data to account for the heterogeneity of flux within wetland habitats and reduce uncertainties on emission estimates in general, and particularly in the Florida Everglades.

Data Availability Statement

Data are publicly available online (at http://www.geosciences.fau.edu/geophysics-lab/data-1.php).

Acknowledgments
This work was partially supported by NOAA (GC11-337), DOE (TES 10959421), USACE and USGS (Cooperative Agreement: Carbon Dynamics of the Greater Everglades). We thank Greg Mount, William Wright, Matthew McClellan, and Matt Sirianni for both laboratory and field support.

References

Bansal, S., Garg, J. K., Sharma, C. S., & Katyal, D. (2018). Spatial methane emission modelling from wetlands using geospatial tools. *International Journal of Remote Sensing*, *39*(18), 5907–5933. https://doi.org/10.1080/01431161.2018.1513182

Bartlett, D. S., Bartlett, K. B., Hartman, J. M., Harriss, R. C., Sebacher, D. L., Pelletier-Travis, R., et al. (1989). Methane emissions from the Florida Everglades: Patterns of variability in a regional wetland ecosystem. *Global Biogeochemical Cycles*, *3*(4), 363–374. https://doi.org/10.1029/GB003i004p00363

Bartlett, K. B., & Harriss, R. C. (1993). Review and assessment of methane emissions from wetlands. *Chemosphere*, *26*(1–4), 261–320. https://doi.org/10.1016/0045-6535(93)90427-7

Bloom, A. A., Palmer, P. I., & Reay, D. S. (2012). Seasonal variability of tropical wetland CH_4 emissions: The role of the methanogen-available carbon pool. *Biogeosciences*, *9*(8), 2821–2830. https://doi.org/10.5194/bg-9-2821-2012

Bubier, J. L., Moore, T. R., & Roulet, N. T. (1993). Methane emissions from wetlands in the midboreal region of northern Ontario, Canada. *Ecology*, *74*, 2240–2254. https://doi.org/10.2307/1939577

Chirici, G., Mura, M., McInerney, D., Py, N., Tomppo, E. O., Waser, L. T., et al. (2016). A meta-analysis and review of the literature on the k-Nearest Neighbors technique for forestry applications that use remotely sensed data. *Remote Sensing of Environment*, *176*, 282–294. https://doi.org/10.1016/j.rse.2016.02.001

Comas, X., & Wright, W. (2012). Heterogeneity of biogenic gas ebullition in subtropical peat soils is revealed using time-lapse cameras. *Water Resources Research*, *48*, W04601. https://doi.org/10.1029/2011WR011654

Comas, X., & Wright, W. (2014). Investigating carbon flux variability in subtropical peat soils of the Everglades using hydrogeophysical methods. *Journal of Geophysical Research: Biogeosciences*, *119*, 1506–1519. https://doi.org/10.1002/2013JG002601

Dean, J. F., Middelburg, J. J., Röckmann, T., Aerts, R., Blauw, L. G., Egger, M., et al. (2018). Methane feedbacks to the global climate system in a warmer world. *Reviews of Geophysics*, *56*, 207–250. https://doi.org/10.1002/2017RG000559

Dengel, S., Zona, D., Sachs, T., Aurela, M., Jammet, M., Parmentier, F. J. W., et al. (2013). Testing the applicability of neural networks as a gap-filling method using CH_4 flux data from high latitude wetlands. *Biogeosciences*, *10*(12), 8185–8200. https://doi.org/10.5194/bg-10-8185-2013

Dronova, I. (2015). Object-based image analysis in wetland research: A review. *Remote Sensing*, *7*(5), 6380–6413. https://doi.org/10.3390/rs70506380

Everglades Restoration. (2020). http://www.evergladesrestoration.gov/. Last data accessed on 25 March 2020.

Flood, N. (2013). Seasonal composite Landsat TM/ETM+ images using the Medoid (a multi-dimensional median). *Remote Sensing*, *5*(12), 6481–6500. https://doi.org/10.3390/rs5126481

Flood, N. (2014). Continuity of reflectance data between Landsat-7 ETM+ and Landsat-8 OLI, for both top-of-atmosphere and surface reflectance: A study in the Australian landscape. *Remote Sensing*, *16*, 7952–7970. https://doi.org/10.3390/rs6097952

Gleason, P. J., & Stone, P. (1994). Age, origin, and landscape evolution of the Everglades peatland. In S. M. Davis & J. C. Ogden (Eds.), *Everglades: The ecosystem and its restoration* (pp. 149–197). Delray Beach, Florida: St. Lucie Press.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: An update. *SIGKDD Explorations*, *11*, 10–18. https://doi.org/10.1145/1656274.1656278

Holben, B. N. (1986). Characteristics of maximum-value composite images from temporal AVHRR data. *International Journal of Remote Sensing*, *7*(11), 1417–1434. https://doi.org/10.1080/01431168608948945

Kim, Y., Johnson, M. S., Knox, S. H., Black, T. A., Dalmagro, H. J., Kang, M., et al. (2020). Gap-filling approaches for eddy covariance methane fluxes: A comparison of three machine learning algorithms and a traditional method with principal component analysis. *Global Change Biology*, *26*(3), 1499–1518. https://doi.org/10.1111/gcb.14845

Matthews, E., & Fung, I. (1987). Methane emission from natural wetlands: Global distribution, area, and environmental characteristics of sources. *Global Biogeochemical Cycles*, *1*(1), 61–86. https://doi.org/10.1029/GB001i001p00061

Miller, S. M., Commane, R., Melton, J. R., Andrews, A. E., Benmergui, J., Dlugokencky, E., et al. (2016). Evaluation of wetland methane emissions across North America using atmospheric data and inverse modeling. *Biogeosciences*, *13*(4), 1329–1339. https://doi.org/10.5194/bg-13-1329-2016

Moore, T. R., Roulet, N. T., & Knowles, R. (1990). Spatial and temporal variations of methane flux from subartic/northern boreal fens. *Global Biogeochemical Cycles*, *4*(1), 29–46. https://doi.org/10.1029/GB004i001p00029

Morin, T. H., Bohrer, G., Stefanik, K. C., Rey-Sanchez, A. C., Matheny, A. M., & Mitsch, W. J. (2017). Combining eddy-covariance and chamber measurements to determine the methane budget from a small, heterogeneous urban floodplain wetland park. *Agricultural and Forest Meteorology*, *237–238*, 160–170. https://doi.org/10.1016/j.agrformet.2017.01.022

Mountrakis, G., Im, J., & Ogole, C. (2011). Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, *66*(3), 247–259. https://doi.org/10.1016/j.isprsjprs.2010.11.001

Parker, R. J., Boesch, H., McNorton, J., Comyn-Platt, E., Gloor, M., Wilson, C., et al. (2018). Evaluating year-to-year anomalies in tropical wetland methane emissions using satellite CH_4 observations. *Remote Sensing of Environment*, *211*, 261–275. https://doi.org/10.1016/j.rse.2018.02.011

Peltola, O., Vesala, T., Gao, Y., Räty, O., Alekseychik, P., Aurela, M., et al. (2019). Monthly gridded data product of northern wetland methane emissions based on upscaling eddy covariance observations. *Earth System Science Data*, *11*, 1263–1289. https://doi.org/10.5194/essd-11-1263-2019

Rose, P. W., & Rosendahl, P. C. (1983). Classification of Landsat data for hydrologic application, Everglades National Park. *Photogrammetric Engineering and Remote Sensing*, *49*, 505–511.

Rutchey, K., Schall, T., & Sklar, F. (2008). Development of vegetation maps for assessing Everglades restoration progress. *Wetlands*, *28*(3), 806–816. https://doi.org/10.1672/07-212.1

Saunois, M., Stavert, A. R., Poulter, B., Bousquet, P., Canadell, J. G., Jackson, R. B., et al. (2020). The global methane budget 2000–2017. *Earth System Science Data*, *12*, 1561–1623. https://doi.org/10.5194/essd-12-1561-2020

Segers, R. (1998). Methane production and methane consumption: A review of processes underlying wetland methane fluxes. *Biogeochemistry*, *41*(1), 23–51. https://doi.org/10.1023/A:1005929032764

Takeuchi, W., Tamura, M., & Yasuoka, Y. (2003). Estimation of methane emission from West Siberian wetland by scaling technique between NOAA AVHRR and SPOT HRV. *Remote Sensing of Environment*, *85*(1), 21–29. https://doi.org/10.1016/S0034-4257(02)00183-9

Trimble (2014). eCognition Developer 9.0.1 Reference Book.

USGS/EDEN (2020). https://sofia.usgs.gov/eden/, accessed on 25 July, 2020.

Vapnik, V. N. (1995). *The nature of statistical learning theory*. New York: Springer-Verlag.

Villa, J. A., & Mitsch, W. J. (2014). Methane emissions from five wetland plant communities with different hydroperiods in the Big Cypress Swamp region of Florida Everglades. *Ecohydrology & Hydrobiology*, *14*(4), 253–266. https://doi.org/10.1016/j.ecohyd.2014.07.005

Wang, Z., Zeng, D., & Patrick, W. H. (1996). Methane emissions from natural wetlands. *Environmental Monitoring and Assessment*, *42*(1–2), 143–161. https://doi.org/10.1007/BF00394047

Whalen, S. C. (2005). Biogeochemistry of methane exchange between natural wetlands and the atmosphere. *Environmental Engineering Science*, *22*, 73–94. https://doi.org/10.1089/ees.2005.22.73

Whiting, G. J., & Chanton, J. P. (2001). Greenhouse carbon balance of wetlands: Methane emission versus carbon sequestration. *Tellus Series B: Chemical and Physical Meteorology*, *53*, 521–528. https://doi.org/10.1034/j.1600-0889.2001.530501.x

Whiting, G. J., Chanton, J. P., Bartlett, D. S., & Happell, J. D. (1991). Relationships between CH4 emission, biomass, and CO_2 exchange in a subtropical grassland. *Journal of Geophysical Research*, *96*(D7), 13,067–13,071. https://doi.org/10.1029/91JD01248

Wright, W., & Comas, X. (2016). Estimating methane gas production in peat soils of the Florida Everglades using hydrogeophysical methods. *Journal of Geophysical Research: Biogeosciences*, *121*, 1190–1202. https://doi.org/10.1002/2015JG003246

Wright, W., Ramirez, J. A., & Comas, X. (2018). Methane ebullition from subtropical peat: Testing an ebullition model reveals the importance of pore structure. *Geophysical Research Letters*, *45*, 6992–6999. https://doi.org/10.1029/2018GL077352

Xu, X., Yuan, F., Hanson, P. J., Wullschleger, S. D., Thornton, P. E., Riley, W. J., et al. (2016). Reviews and syntheses: Four decades of modeling methane cycling in terrestrial ecosystems. *Biogeosciences*, *13*(12), 3735–3755. https://doi.org/10.5194/bg-13-3735-2016

Zhang, B., Tian, H., Lu, C., Chen, C., Pan, S., Anderson, C., & Poulter, B. (2017). Methane emissions from global wetlands: An assessment of the uncertainty associated with various wetland extent data sets. *Atmospheric Environment*, *165*, 310–321. https://doi.org/10.1016/j.atmosenv.2017.07.001

Zhang, C., Denka, S., Cooper, H., & Mishra, D. R. (2018). Quantification of sawgrass marsh aboveground biomass in the coastal Everglades using object-based ensemble analysis and Landsat data. *Remote Sensing of Environment*, *204*, 366–379. https://doi.org/10.1016/j.rse.2017.10.018

Zhang, C., Mishra, D. R., & Pennings, S. (2019). Mapping salt marsh soil properties using imaging spectroscopy. *ISPRS Journal of Photogrammetry and Remote Sensing*, *148*, 221–234. https://doi.org/10.1016/j.isprsjprs.2019.01.006

Zhang, C., Smith, M., Lv, J., & Fang, C. (2017). Applying time series Landsat data for vegetation change analysis in the Florida Everglades Water Conservation Area 2A during 1996–2016. *International Journal of Applied Earth Observation and Geoinformation*, *57*, 214–223. https://doi.org/10.1016/j.jag.2017.01.007

Zhang, Y., Sachs, T., Li, C., & Boike, J. (2012). Upscaling methane fluxes from closed chambers to eddy covariance based on a permafrost biogeochemistry integrated model. *Global Change Biology*, *18*(4), 1428–1440. https://doi.org/10.1111/j.1365-2486.2011.02587.x

Zhu, Q., Liu, J., Peng, C., Chen, H., Fang, X., Jiang, H., et al. (2014). Modelling methane emissions from natural wetlands by development and application of the TRIPLEX-GHG model. *Geoscientific Model Development*, *7*(3), 981–999. https://doi.org/10.5194/gmd-7-981-2014