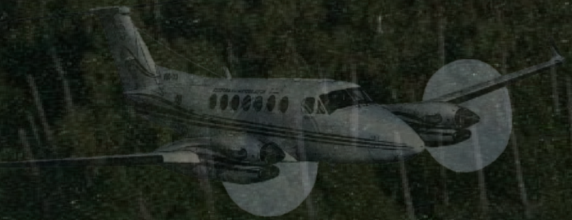
Integrating Multi-Source Data with Machine Learning Techniques to Upscale Wetland Carbon Dioxide Fluxes



By:

Abdullah Al Fazari, Dr. Caiyun Zhang & Dr. Xavier Comas

Date & Time:

Thursday, April 24th, 2025, at 2:05 PM

Event:

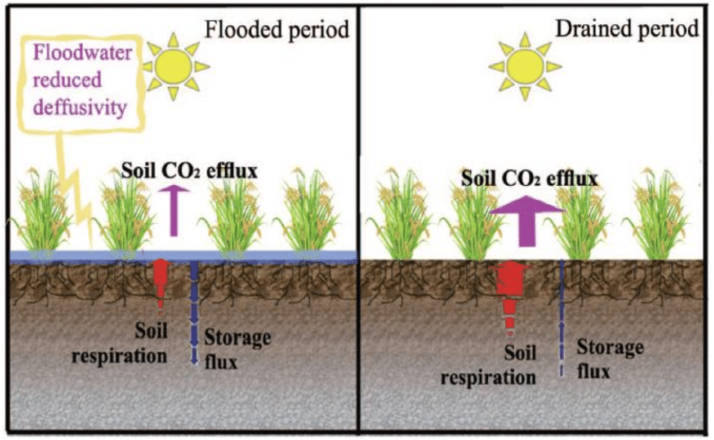
GEER Conference, Coral Spring, FL

Introduction

Background and Objectives

Research Background

- Carbon dioxide (CO_2), one of the Greenhouse Gases (GHG) emitted both by human activities and natural processes, is the primary driver of climate change (Solomon et al., 2009).
- Wetland soil can hold up to or even more than 40% of carbon (Lal et al., 1995).
- Wetlands play a critical role in controlling the global carbon cycle and quantifying their carbon budgets, including storage, uptake, and emissions (Lu et al., 2017).
- Dry wetlands release more CO_2 but less CH_4. In contrast, wet wetlands release less CO_2 but more CH_4 (Waddington and Price, 2000).
- If air temperature increases, then water depth decreases.
- CO_2 emissions increased in a linear relationship as the water level decreased.



Comparison of soil CO2 flux processes under the flooded and the drained conditions. Source: (Sardar et al., 2018).

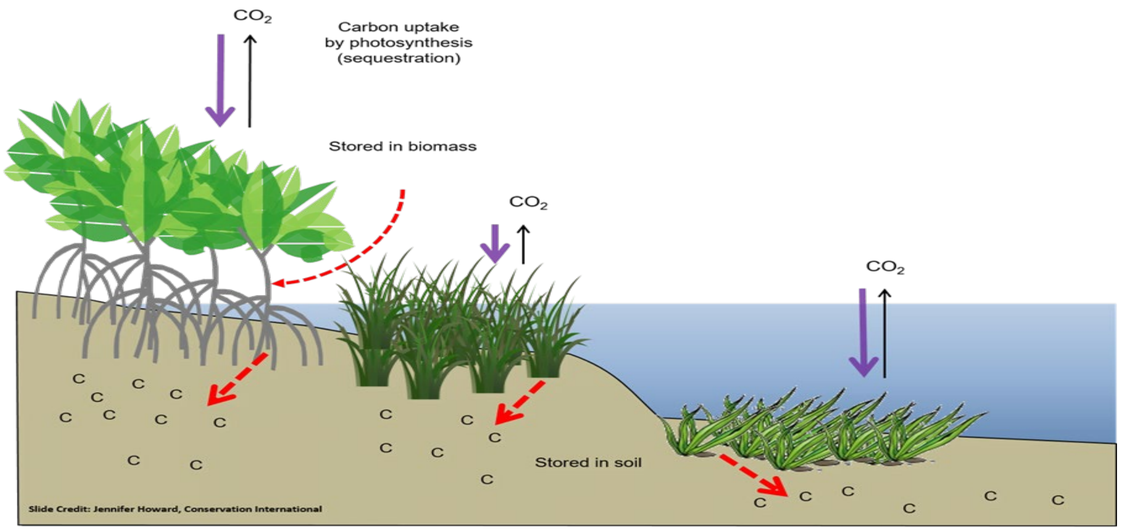


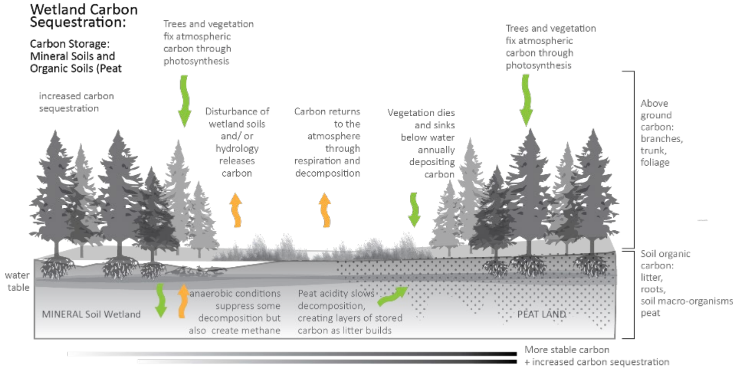
Figure 1. In intact coastal wetlands (from left to right: mangroves, tidal marshes, and seagrasses), carbon is taken up via photosynthesis (purple arrows) where it gets sequestered long term into woody biomass and soil (red dashed arrows) or respired (black arrows).
Source: (Howard et al. 2017).

- **Why** do we need to quantify and measure CO_2 in wetlands?

To assess and understand carbon dynamics (emission, absorption, and sequestration).

Assess environmental factors, and their impacts (e.g., hydrology, air temperature, soil, vegetation)

Improve climate models.



Source: https://bwsr.state.mn.us/carbon-sequestration-wetlands

How to measure CO2?

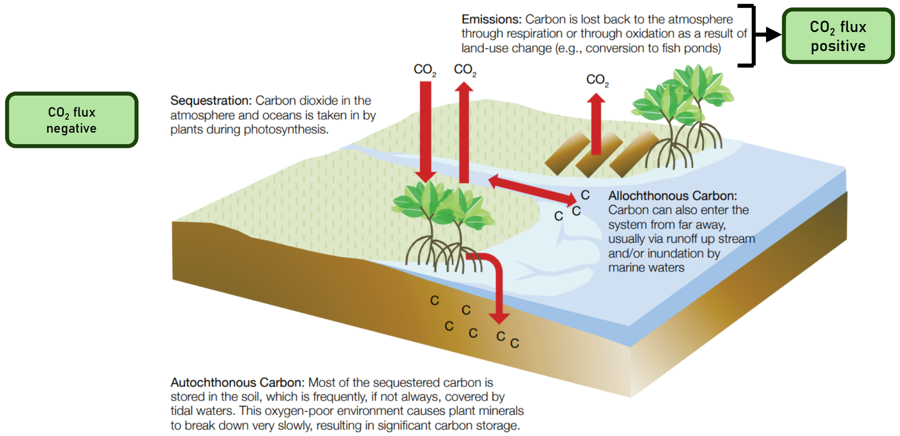
Net Ecosystem CO_2 Exchange (NEE) = Reco – GPP

NEE indicates the net exchange of CO_2

+ NEE = releases CO_2,

- NEE = absorbs CO_2

(Bubier et al., 2002; Chapin et al., 2011).



Mechanisms by which carbon moves into and out of wetlands. Source: modified from Howard et al. 2014.

Research Gap

- Ground-based CO_2 monitoring techniques, are among the most reliable and accurate methods. However, it has limitations in terms of spatial coverage and can be both time-consuming and costly.
- Not all areas within the wetlands are accessible anytime due to many logistical challenges.
- Operating and maintaining the ground-based techniques over the long term requires allocated budgets and specialized human resources.
- In situ measurements tend to have noticeable gaps in their records, which lack continuous records of measurements.
- Applications in upscaling CO_2 fluxes using multiple sources, like spaceborne and airborne, combined with an OBIA approach and machine learning algorithms for wetlands remain limited.
- Lack of comprehensive studies that examine the seasonal variations of CO_2 fluxes in the Everglades using such an integrated approach, considering the newly airborne data of CO_2.


EC tower

Main Objective

- Synthesize multi-source data to develop an innovative model to characterize carbon dioxide fluxes throughout multiple seasons from 2022 to 2023 in the Big Cypress National Preserve (BCNP) and Everglades National Park (ENP) in the Florida Everglades.

Sub Objective

1) Develop an object-based machine learning ML framework to estimate CO_2 fluxes by linking tower and airborne flux measurements with satellite observations.

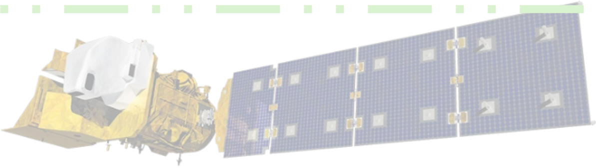
2) Evaluate the effectiveness of the new airborne CO_2 dataset from NASA's BlueFlux mission across different ecosystems in terms of EC tower measurements in the Everglades.

3) Explore the value of environmental variables (e.g., air temperature, water table depth, and land use) for CO_2 fluxes upscaling.

4) Identify the effectiveness of different satellite products (Landsat 8, Harmonized Landsat Sentinel-2 (HLS)) for flux upscaling.

5) Assess the performance of the different ML algorithms and ensemble analysis for estimating CO_2 fluxes.

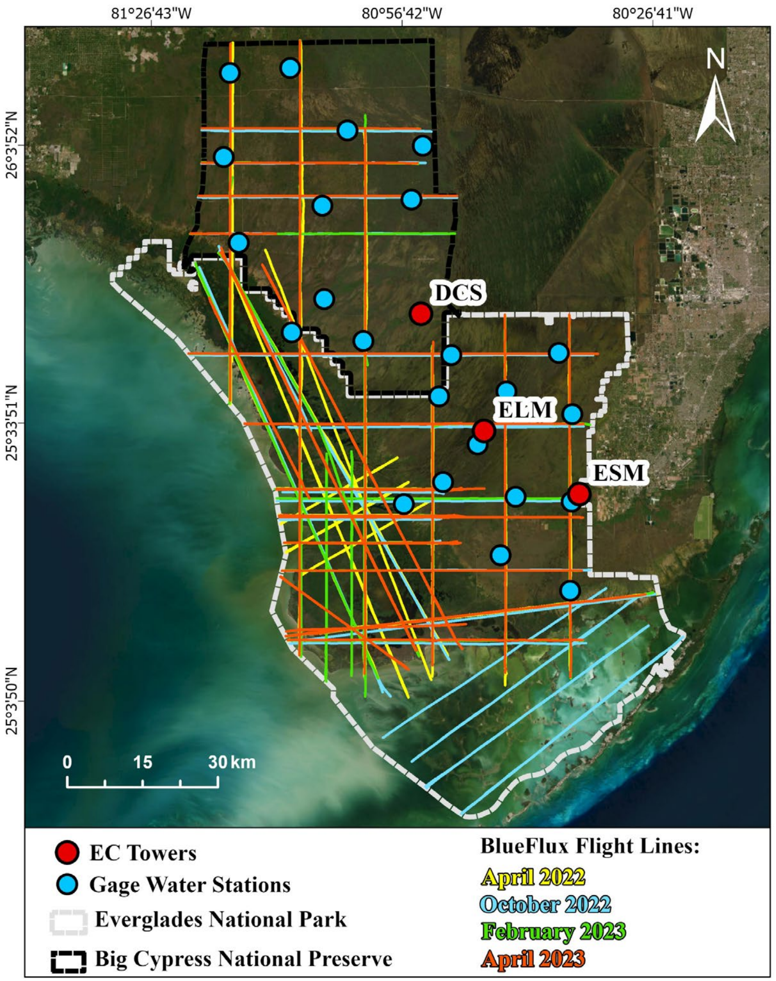
6) Generate maps of CO_2 fluxes across ENP and BCNP to identify hot spots of flux exchanges and document the seasonal and interannual flux variations.



STUDY AREA
AND
DATA

Study Area

- Big Cypress National Preserve (BCNP): Area: ~2960 km^2
- Everglades National Park (ENP): ~6000 km^2
- Climate: Subtropical with distinct wet and dry seasons; hot, humid summers and mild winters.
- Ecosystem Types: Includes sawgrass marshes, cypress swamps, sloughs, and mangroves.
- Temperature: Northern region stays above 27°C (April–October), southern region (March–November).
- Rainfall: 70% occurs from mid-May to November, with hurricanes and droughts impacting water levels.



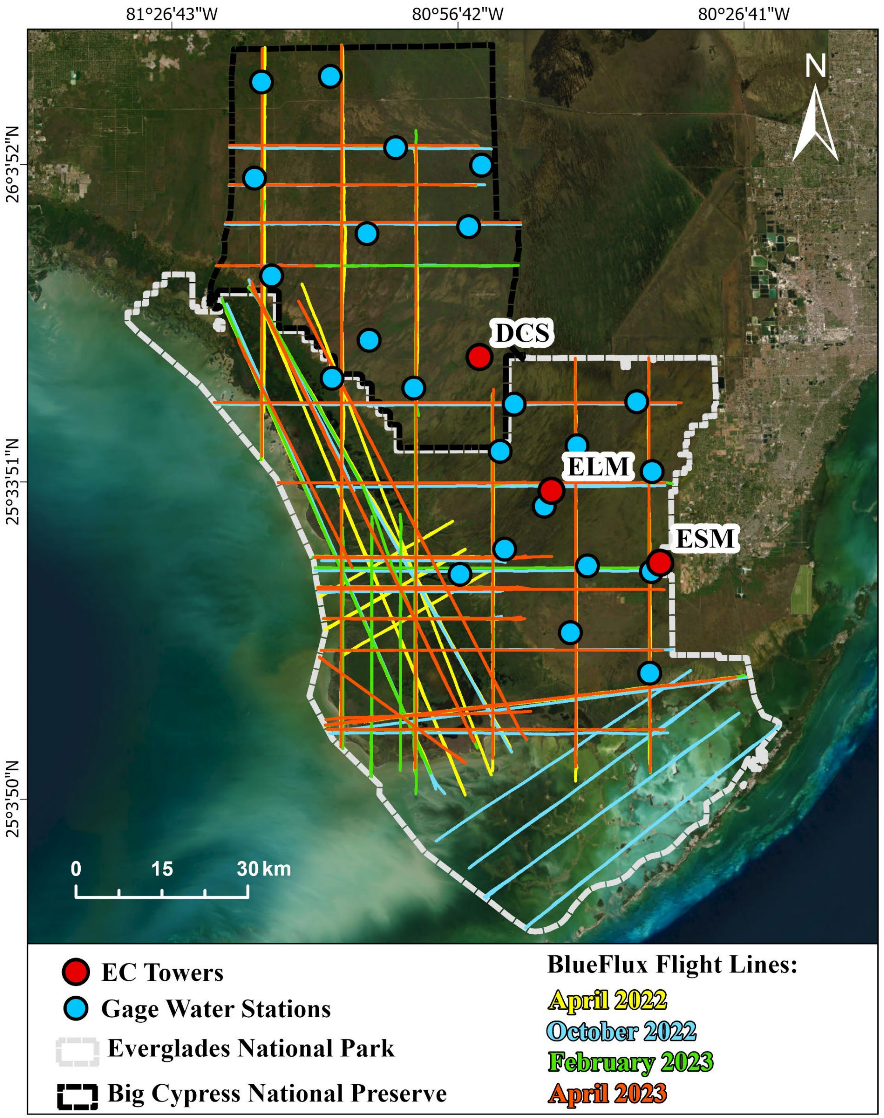
The study areas (ENP and BCNP).

Datasets

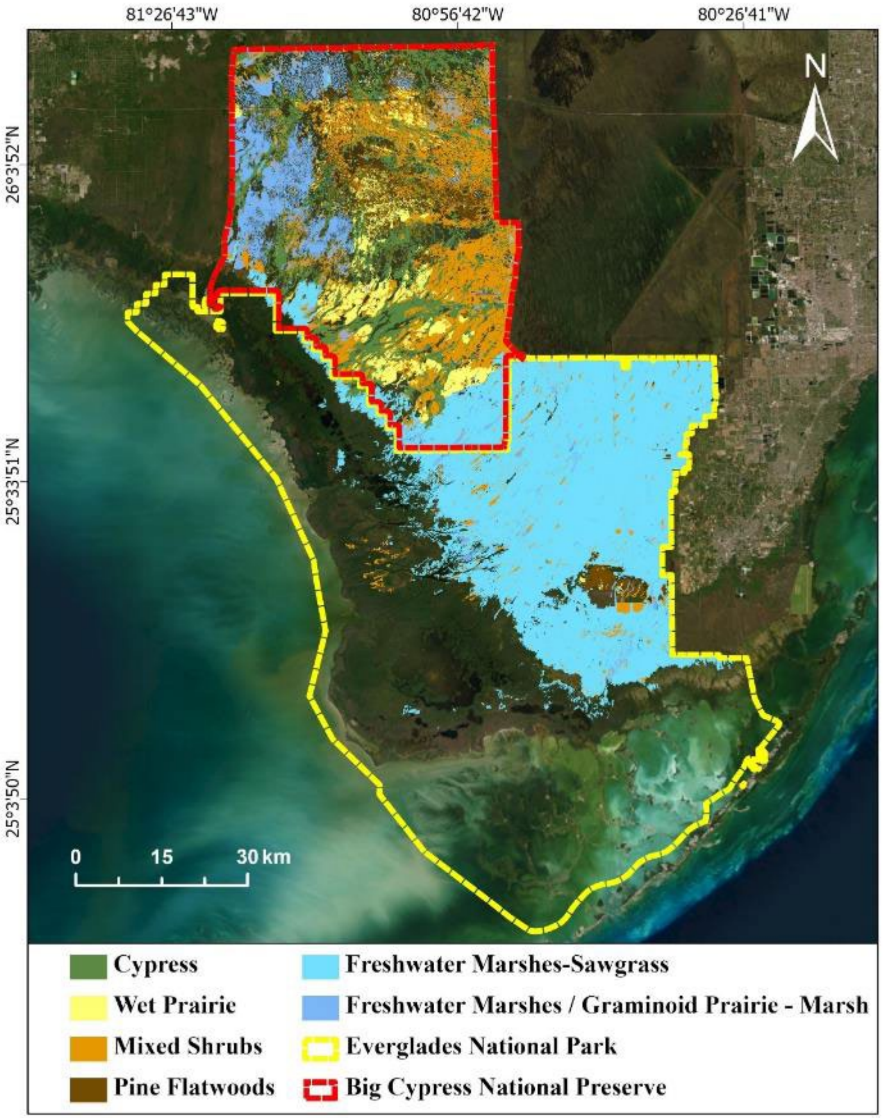
- ❖ AmeriFlux Eddy Covariance (EC tower)
- ❖ Airborne BlueFlux Data
- ❖ Landsat 8 imagery (Bands 2, 3, 4, 5, 6, 7)
- ❖ Harmonized Sentinel-2 MSI Imagery (Bands 2, 3, 4, 8, 11, 12)
- ❖ Land Use Land Cover (LULC) Dataset (SFWMD Geospatial Open Data portal, can be accessed at (https://geo-sfwmd.hub.arcgis.com).
- ❖ Water Level Data. (EDEN)
- ❖ Air Temperature Data. (Landsat thermal band 10)
- ❖ LiDAR data, Digital Elevation Model (DEM), was downloaded from USGS Lidar Explorer Map https://apps.nationalmap.gov

Flight Period
April 19 - April 26, 2022
October 14 - October 20, 2022
February 5 - February 13, 2023
April 13 - April 19, 2023

Table 4: The four flight campaigns by BlueFlux.

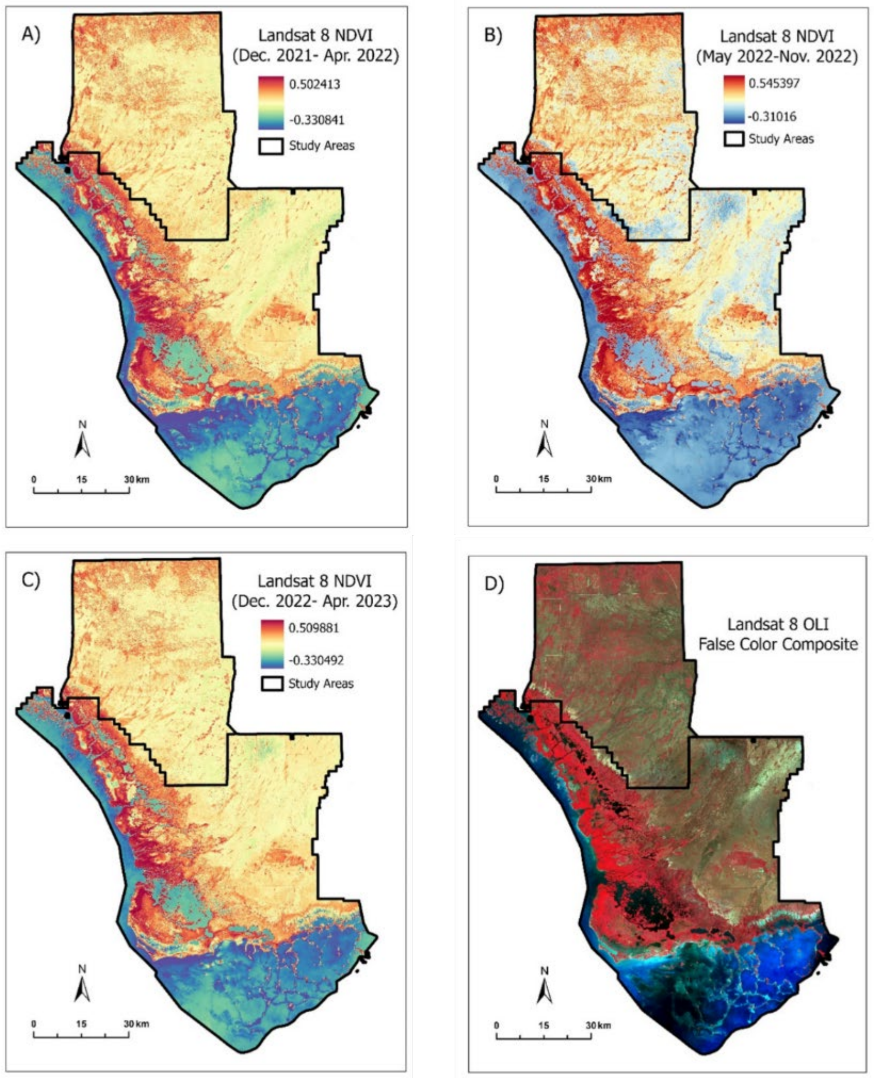


The study area, with the datasets.



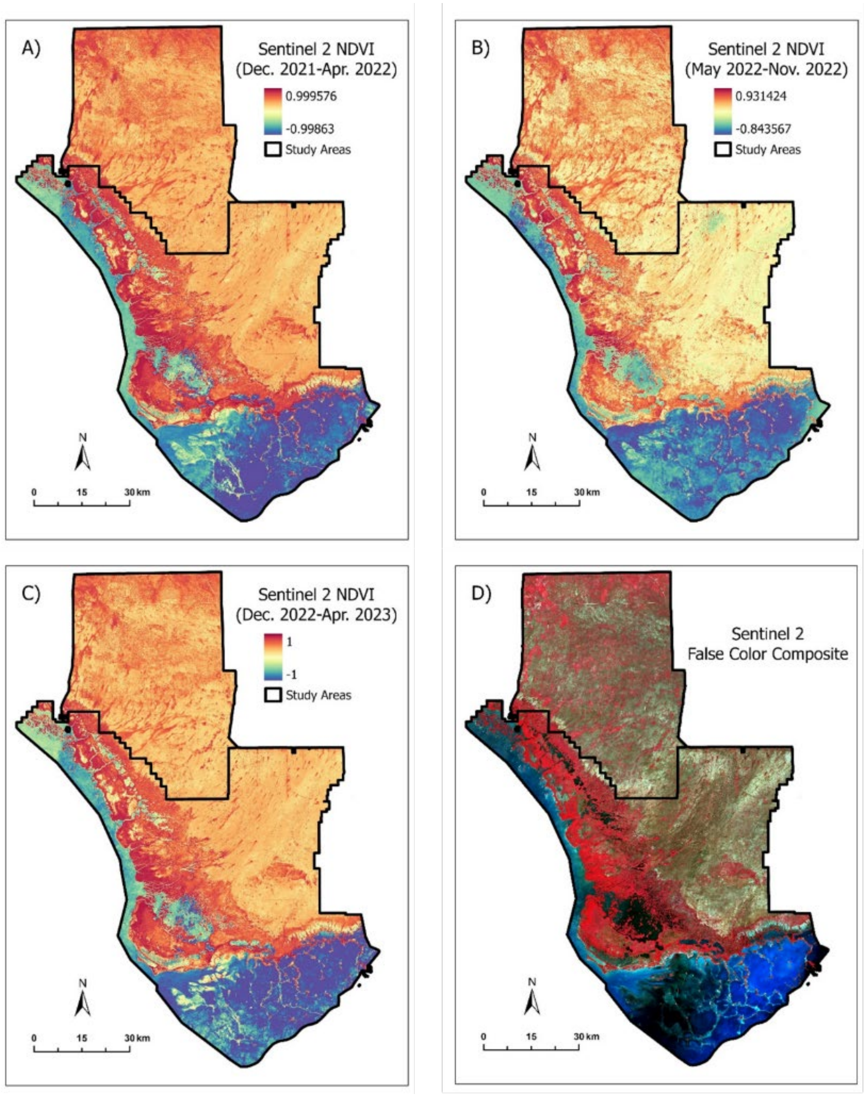
The land use land cover of the Everglades National Park and Big Cypress National Preserve.

NDVI Variations Across Seasons



The Landsat 8 NDVI variations for the study area across the three different time periods.

A) NDVI for the period December 2021-April 2022, B) NDVI for the period May 2022-November 2022, C) NDVI for the period December 2022-April 2023, and D) a false-color composite.



The NDVI variations of the Harmonized Sentinel-2 MSI for the study area across the three different time periods.

A) NDVI for the period December 2021-April 2022, B) NDVI for the period May 2022-November 2022, C) NDVI for the period December 2022-April 2023, and D) a false-color composite.

Methodology

Framework for Upscaling CO_2 Estimations

- Image Segmentation
- Data Matching
- Machine Learning Modeling for Upscaling CO_2 Estimations
- Accuracy Assessment
- CO_2 Mapping Products Generation

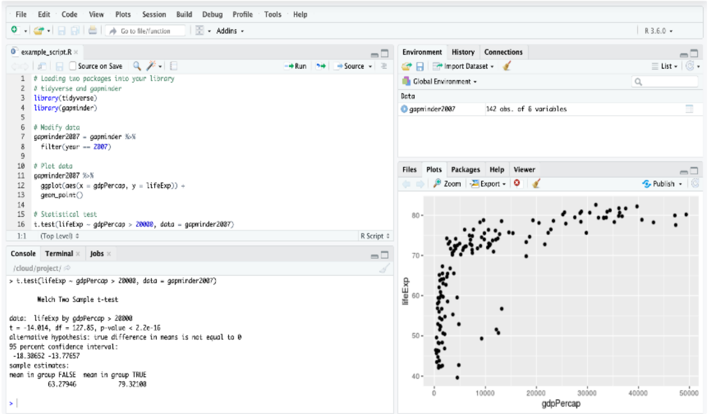
Machine Learning Modeling - Algorithms

No.	Machine Learning Technique	Library and package in R	References
1	Random Forest (RF)	library (randomForest); package: randomForest	(Breiman, 2001)
2	Support Vector Machine (SVM)	library (kernlab); package: svm	(Karatzoglou et al., 2004)
3	k-Nearest Neighbor (k-NN)	library (class); package: knn	(Venables and Ripley, 2013)
4	Multiple Linear Regression (MLR)	Library (stats); package: lm	(Draper and Smith, 1998)
5	Extreme Gradient Boosting (XGB)	library (xgboost); package: xgboost	library (xgboost); package: xgboost
6	Weighted Ensemble Analysis	library(caretEnsemble); package: caretEnsemble	(Deane-Mayer and Knowles, 2019)
7	Meta-Ensemble Analysis	library(SuperLearner); package: SuperLearner	(van der Laan et al., 2007)

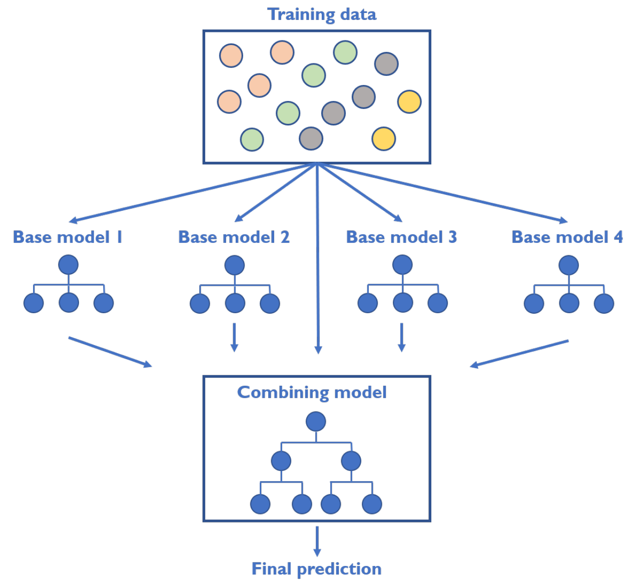
Table 5 The machine learning techniques with their respective libraries and packages in R.

		B2	B3	STD_B8	STD_B11	STD_B12	STD_B3	STD_B4	STD_B2	NDVI_Mean	NDVI_STD	DEM_Mean	DEM_STD	WTL_Mean
1	3	351.180555555556	435.131944444444	712.529570401108	270.35982886881	140.904715637515	108.413303594251	67.061831943562	49.4865515931374	0.580221	0.167668	-0.574804	0.476304	0.139118
2	4	329.766666666667	436.296825396825	569.302642228167	284.878158101787	147.098676386323	92.4437365880801	97.7987163727738	58.0275687079748	0.206125	0.239102	-0.56205	0.439007	0.139118
3	9	275.149560117302	328.41935483871	343.254855771228	228.77766397281	146.523162962997	120.234406104551	128.994093712832	73.1754384147887	0.234997	0.144781	-0.560506	0.386816	0.139118
4		345.773529411765	461.514705882353	440.042480658103	263.207251144377	153.39564360355	173.667411686177	178.353723626532	95.3375334257192	0.173974	0.174593	-0.545853	0.362724	0.139118
5	3	678.275510204082	1040.58673469388	431.968943753306	228.029768638343	136.388835836429	269.379057576432	260.515577706235	168.482908393279	0.138842	0.110106	-0.442762	0.295958	0.139118
6	5	306.196787148594	363.343373493976	236.083726296761	132.004790433316	86.2211570671677	85.237522884555	97.1330141109777	58.099013858794	0.243702	0.084713	-0.378138	0.239589	0.153046
7	3	342.047244094488	538.543307086614	659.354301417826	270.657946276195	171.43307466355	97.9021893303585	100.885462261681	50.4680624459733	0.731147	0.121165	-0.303492	0.497616	0.139118
8	2	570.238866396761	874.692307692308	416.414883964155	148.347845810291	102.909752859109	160.340529771443	136.670928109958	106.488341284312	0.262637	0.11327	-0.294985	0.233746	0.139118
9	3	254.363468634686	341.621771217712	653.959203685703	371.327919961844	234.62545225201	126.458224678131	137.346791766642	69.0583727723235	0.651924	0.132245	-0.288326	0.232209	0.153046
10	2	406.553805774278	487.062992125984	303.724138309263	290.920811415434	200.061169293984	102.734150813048	114.413546747108	68.0445015526811	0.344609	0.066345	-0.287414	0.168173	0.139118

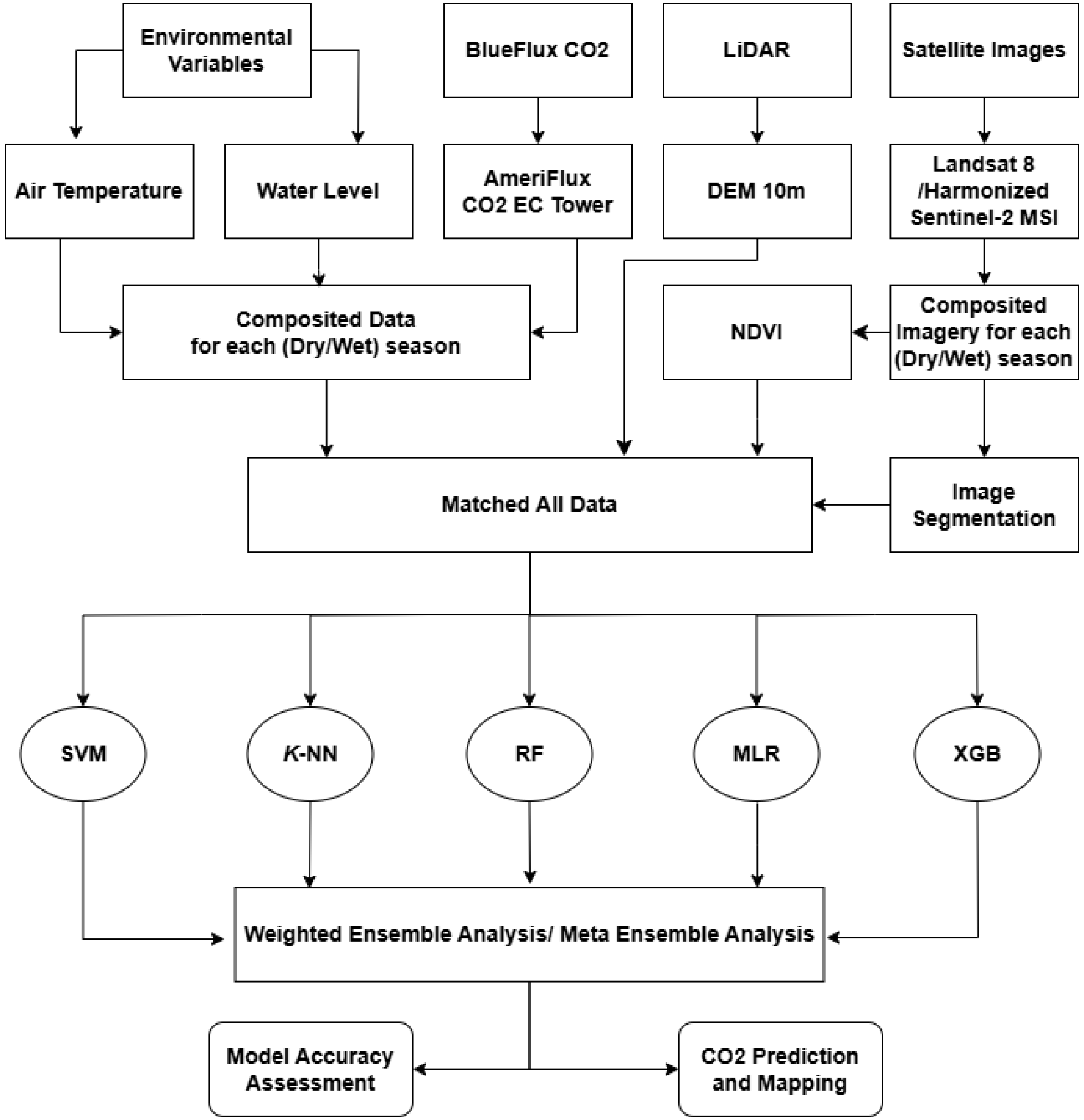
RStudio software



Source: https://argoshare.is.ed.ac.uk/healthyr_book/what-is-rstudio.html

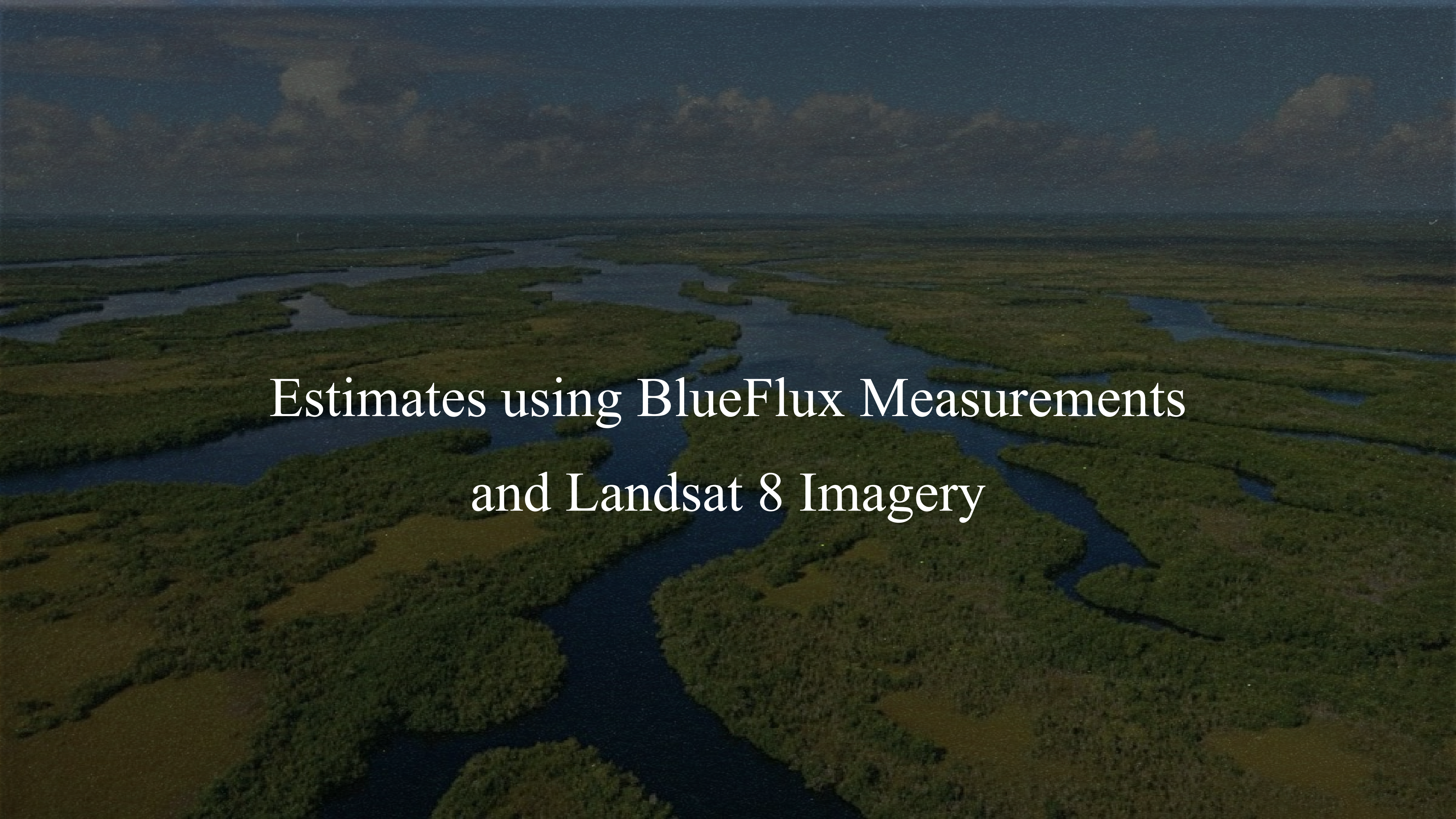


Source: https://hyperskill.org/learn/step/15391

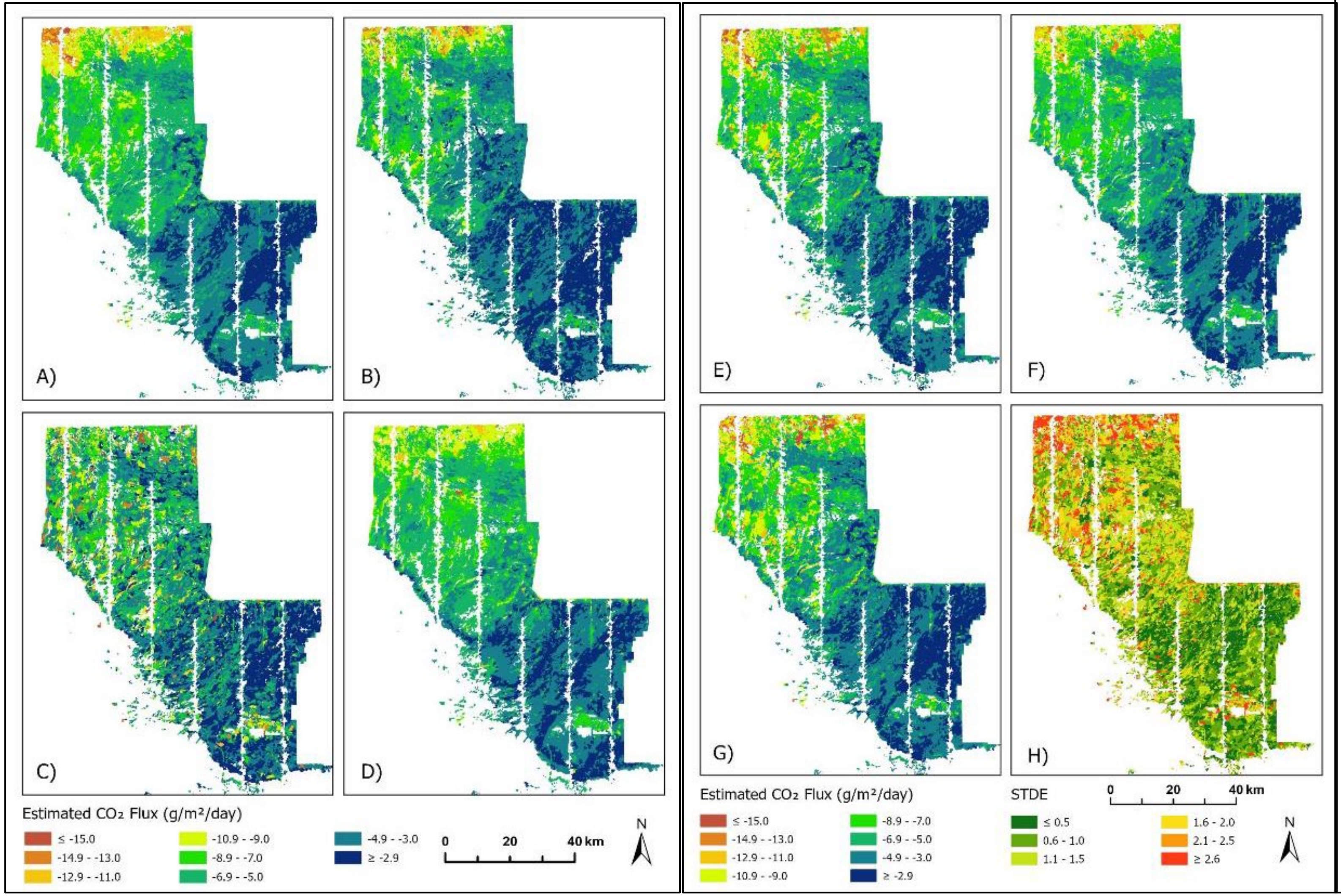


The flowchart illustrates the methodology for upscaling CO_2 fluxes from both the AmeriFlux EC towers and the BlueFlux airborne measurements using Landsat 8 and Harmonized Sentinel-2 MSI imagery and machine learning models.

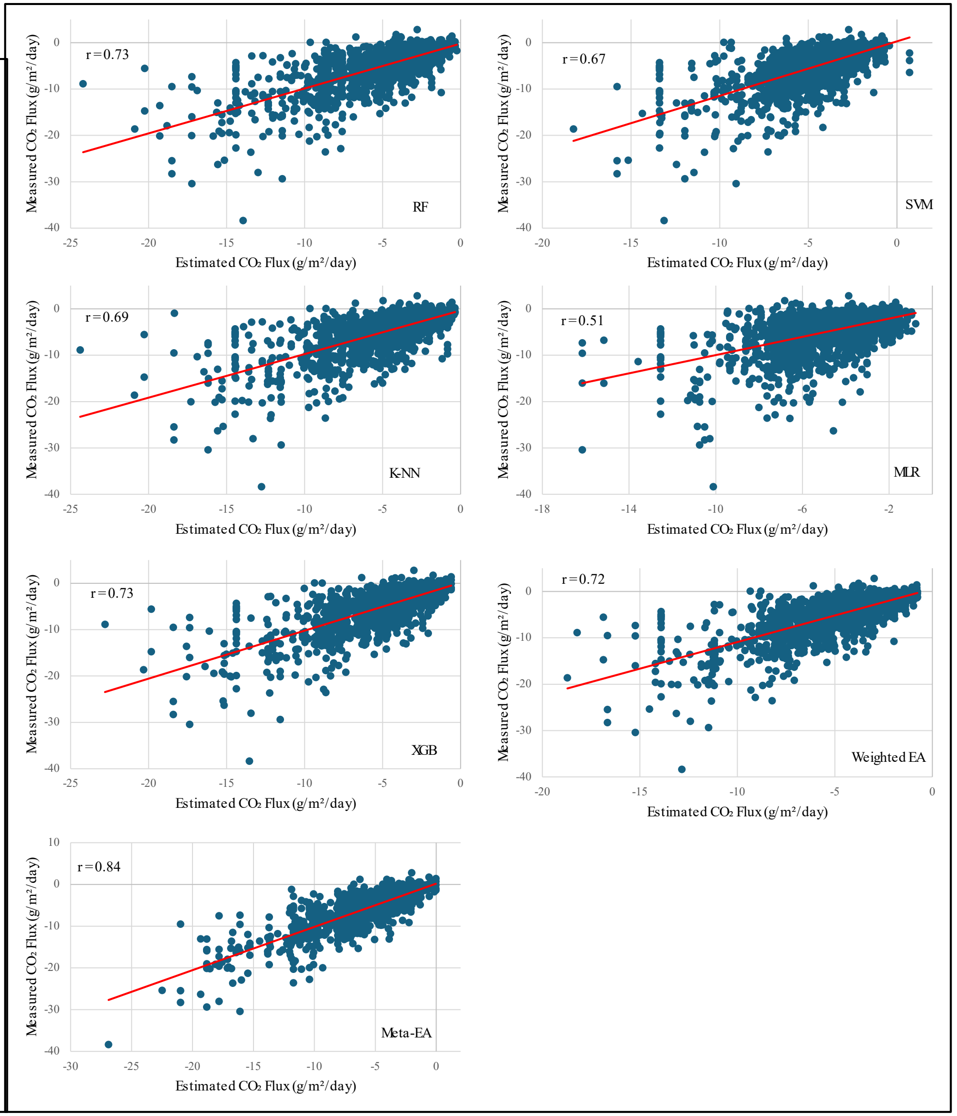
Results and Discussion


Estimates using BlueFlux Measurements
and Landsat 8 Imagery

CO_2 Flux Estimation (Landsat 8: December 2021-April 2022)

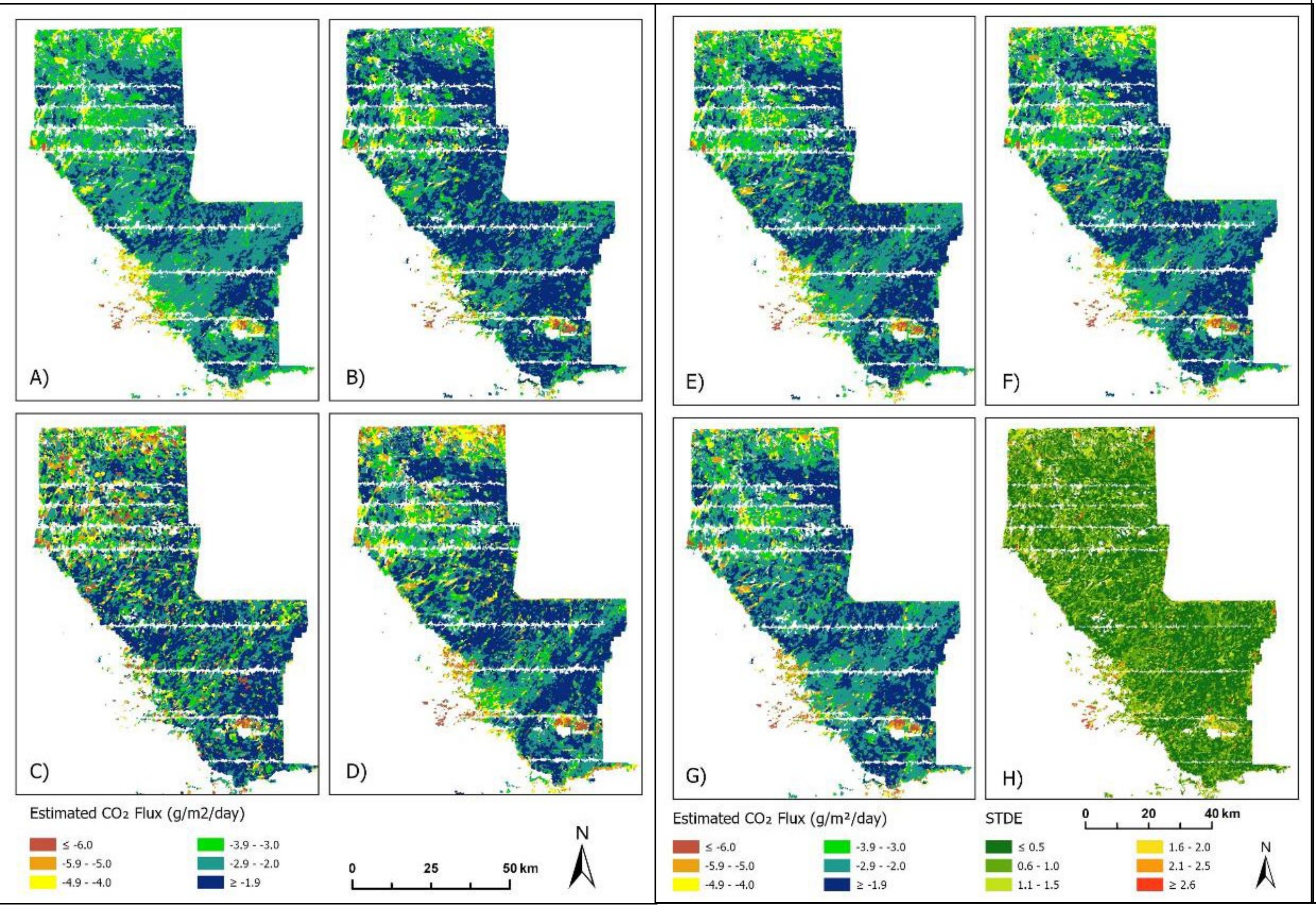


Estimated mean CO2 flux (g/m2/day) in BCNP and ENP from December 2021 to April 2022 using A) RF, B) SVM, C) KNN, D) MLR, E) XGB, F) Weighted Ensemble Analysis, G) Meta-Ensemble, and H) STDE by the BlueFlux data and Landsat 8 imagery.

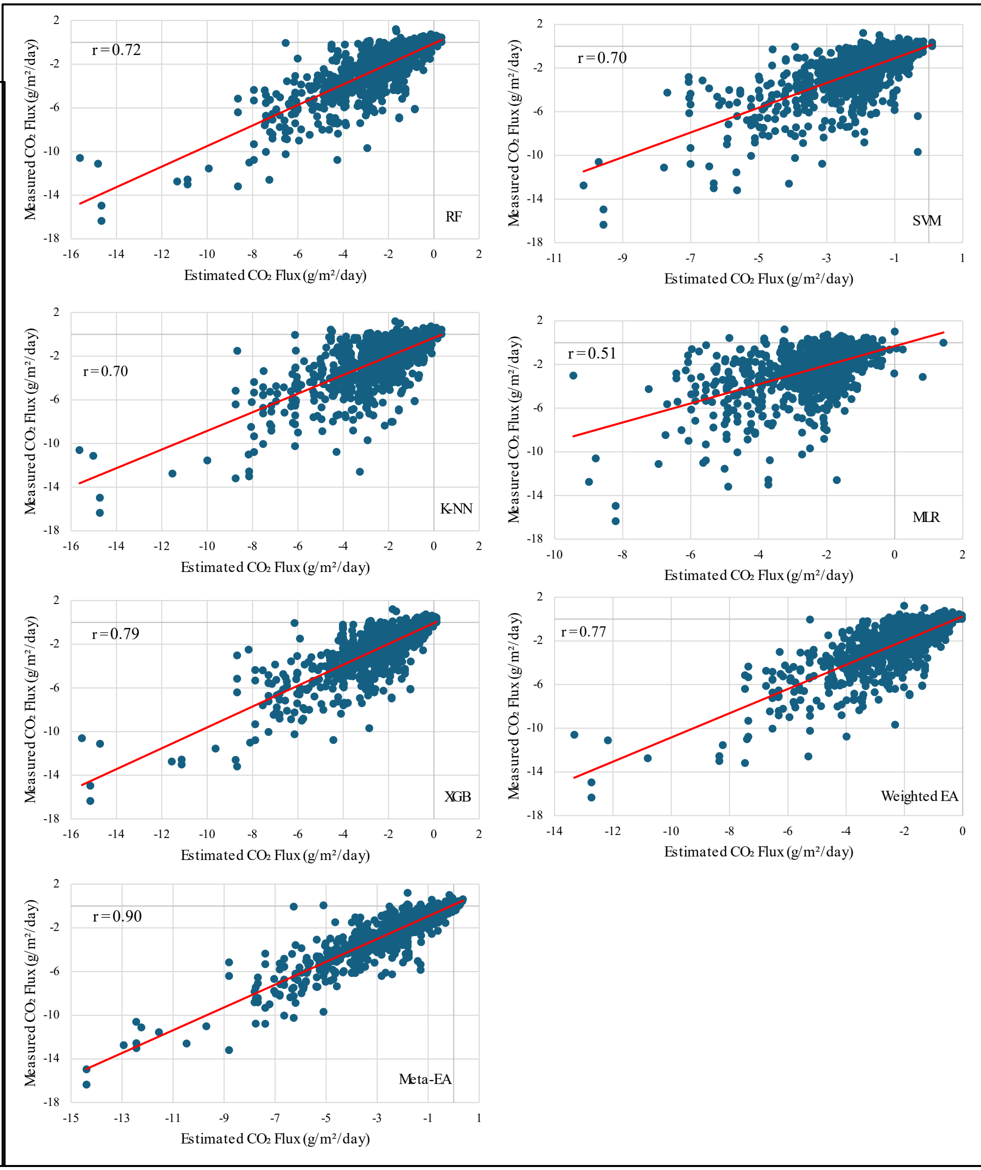


Scatterplot of estimated and measured CO2 flux (g/m2/day) and corresponding regression line for the season December 2021 to April 2022 by the BlueFlux data and Landsat 8 imagery.

CO_2 Flux Estimation (Landsat 8: May–Nov 2022)



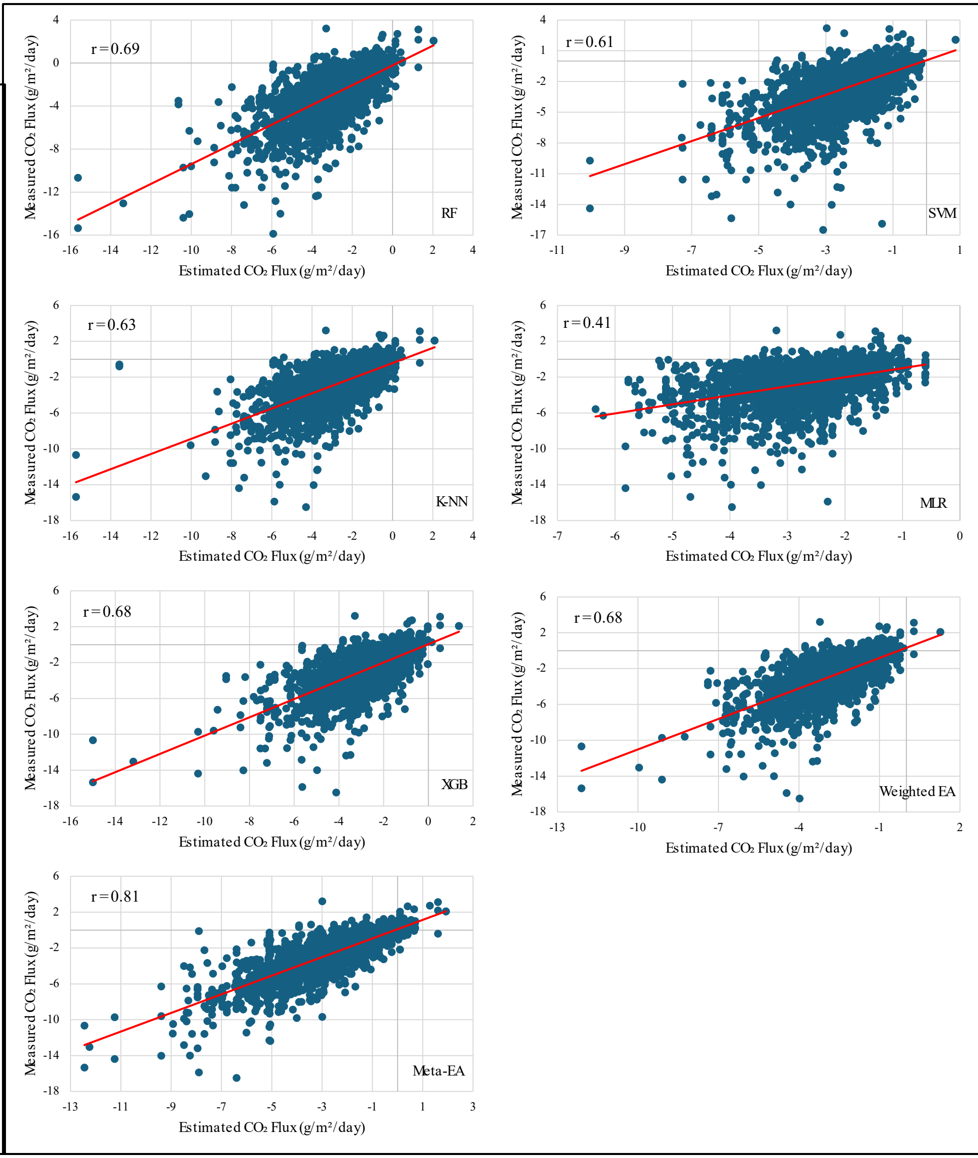
Estimated mean CO2 flux (g/m2/day) in BCNP and ENP from May 2022-November 2022 using A) RF, B) SVM, C) KNN, D) MLR, E) XGB, F) Weighted Ensemble Analysis, G) Meta-Ensemble, and H) STDE by the BlueFlux data and Landsat 8 imagery.



Scatterplot of estimated and measured CO2 flux (g/m2/day) and corresponding regression line for the season May 2022-November 2022 by the BlueFlux data and Landsat 8 imagery.

CO_2 Flux Estimation (Landsat 8: December 2022-April 2023)

Estimated mean CO2 flux (g/m2/day) in BCNP and ENP from December 2022-April 2023 using A) RF, B) SVM, C) KNN, D) MLR, E) XGB, F) Weighted Ensemble Analysis, G) Meta-Ensemble, and H) STDE by the BlueFlux data and Landsat 8 imagery.

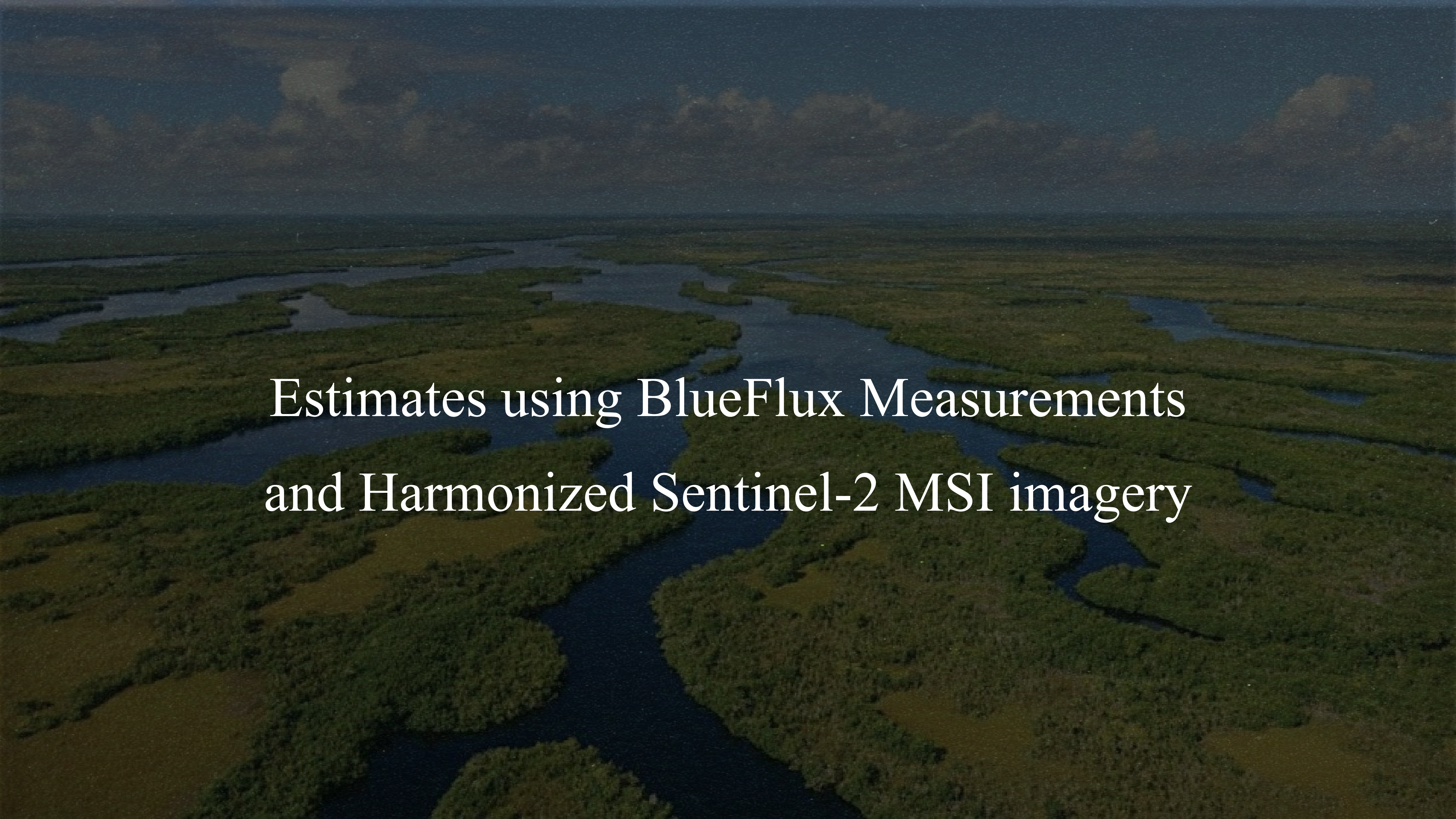


Scatterplot of estimated and measured CO2 flux (g/m2/day) and corresponding regression line for the season December 2022-April 2023 by the BlueFlux data and Landsat 8 imagery.

Estimates using BlueFlux Measurements and Landsat 8 Imagery

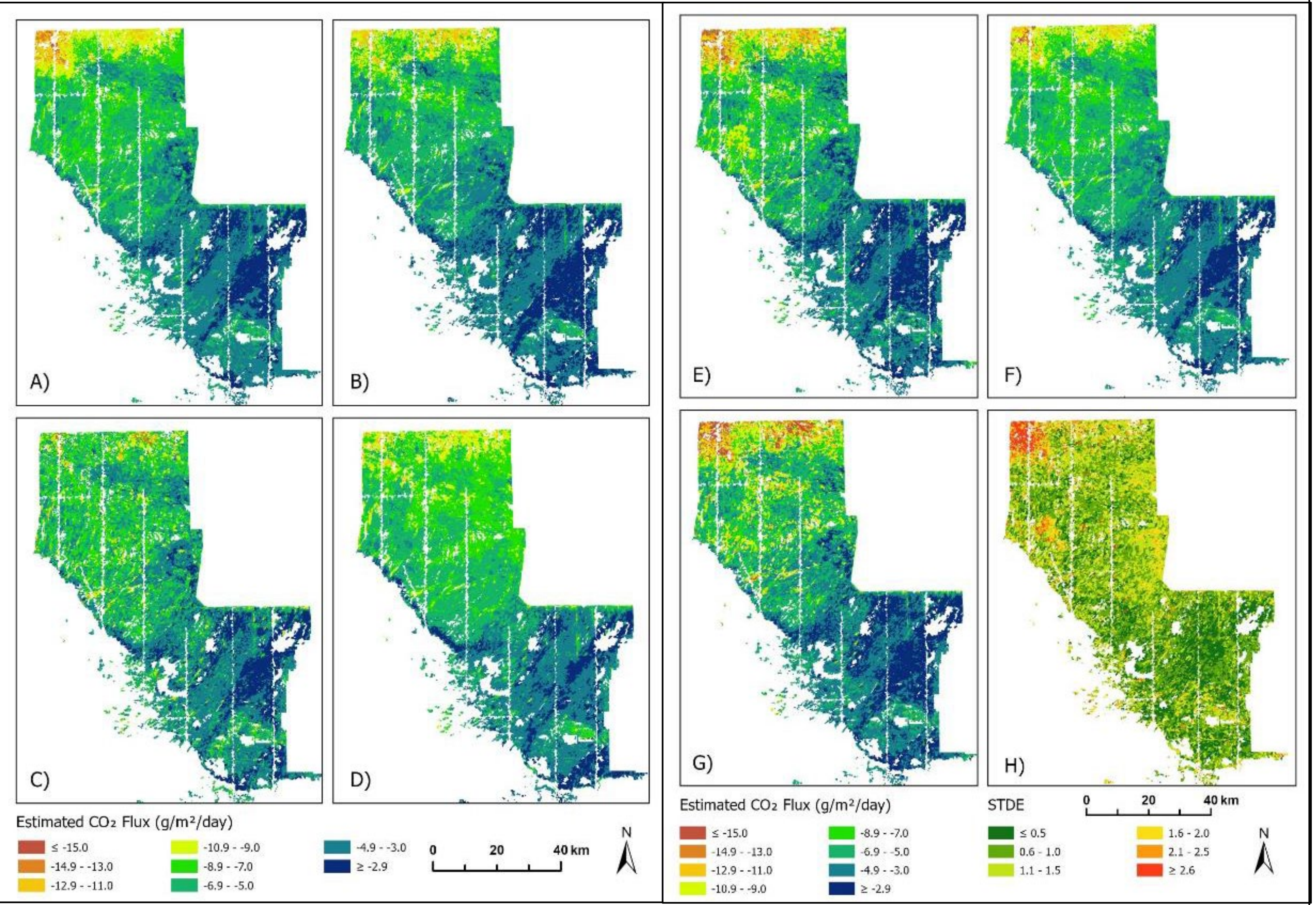
Season	Metric	RF	SVM	KNN	MLR	XGB	Weighted EA	Meta-EA
December 2021-April 2022	*r*	0.73	0.67	0.69	0.51	0.73	0.72	**0.84**
	MAE (g/m2/day)	1.93	2.06	2.05	2.53	1.95	1.98	1.52
	RMSE (g/m2/day)	2.84	3.12	3.00	3.54	2.83	2.89	2.24
May 2022-November 2022	*r*	0.79	0.70	0.70	0.51	0.79	0.77	**0.90**
	MAE (g/m2/day)	0.95	1.10	1.14	1.40	0.96	1.01	0.65
	RMSE (g/m2/day)	1.37	1.64	1.61	1.93	1.37	1.42	0.98
December 2022-April 2023	*r*	0.69	0.61	0.63	0.41	0.68	0.68	**0.81**
	MAE (g/m2/day)	1.12	1.21	1.22	1.46	1.15	1.16	0.88
	RMSE (g/m2/day)	1.56	1.74	1.70	1.97	1.58	1.59	1.28

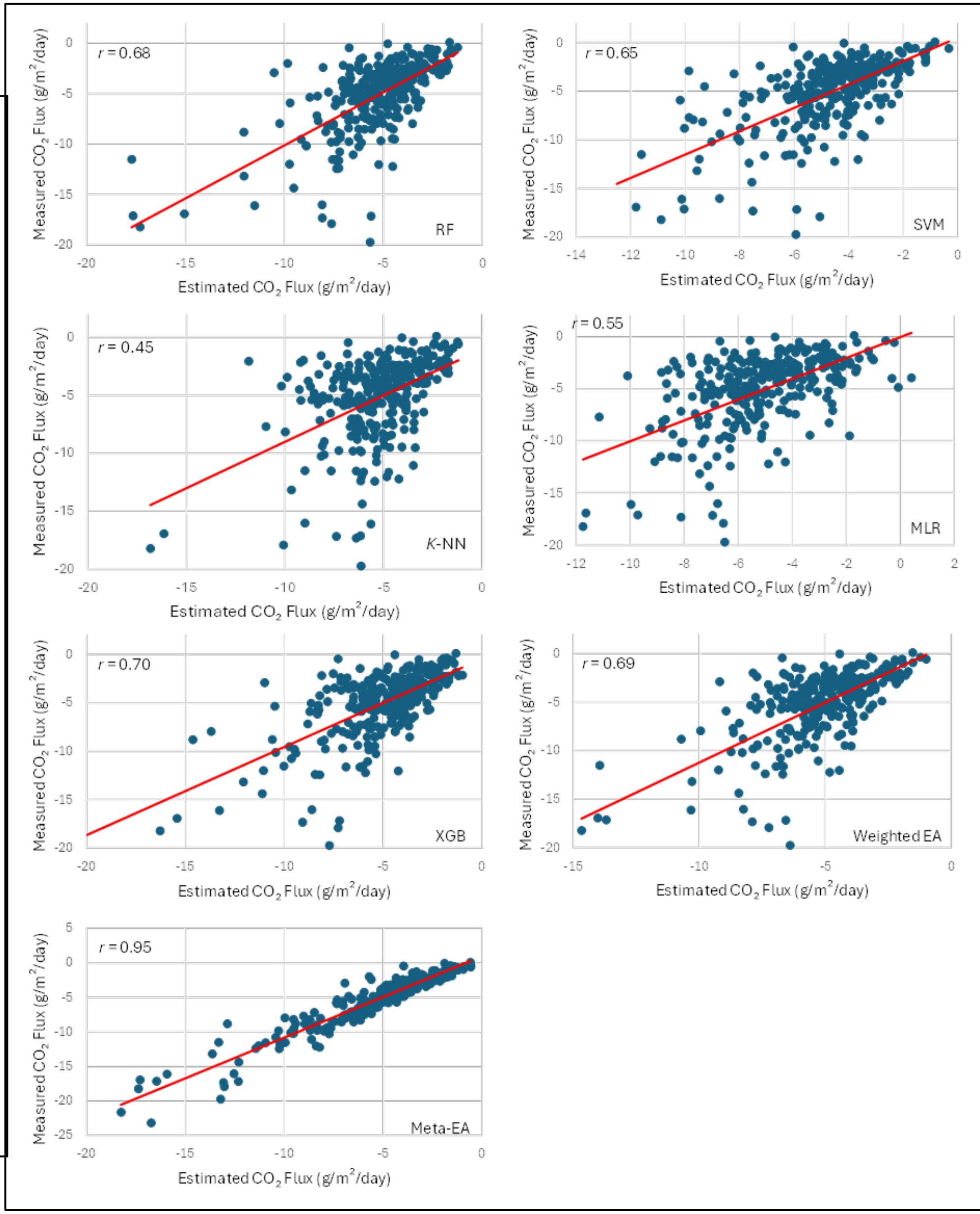
Table Predicted CO2 accuracies using BlueFlux data and Landsat imagery for the study area across the three study seasons.



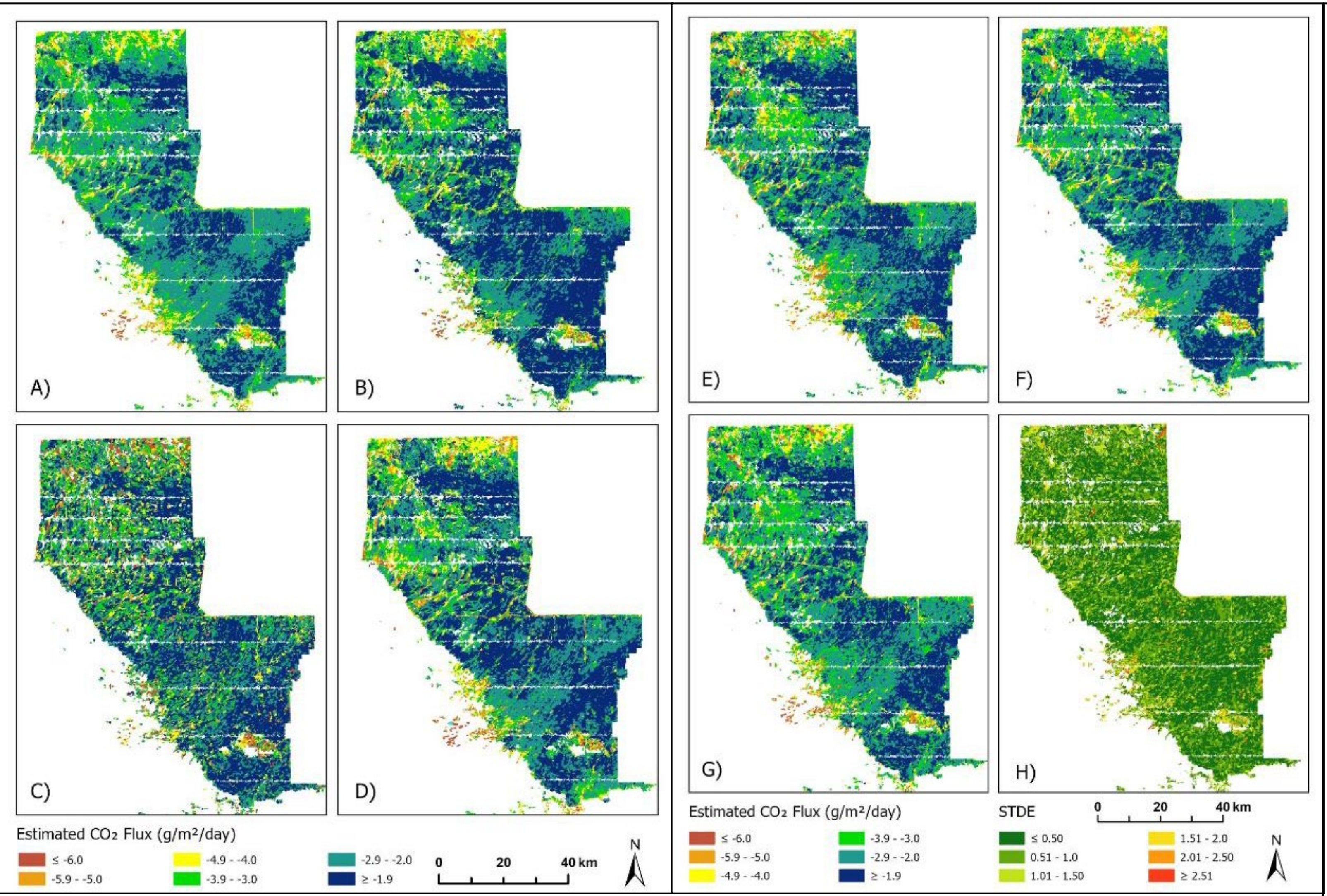
Estimates using BlueFlux Measurements and Harmonized Sentinel-2 MSI imagery

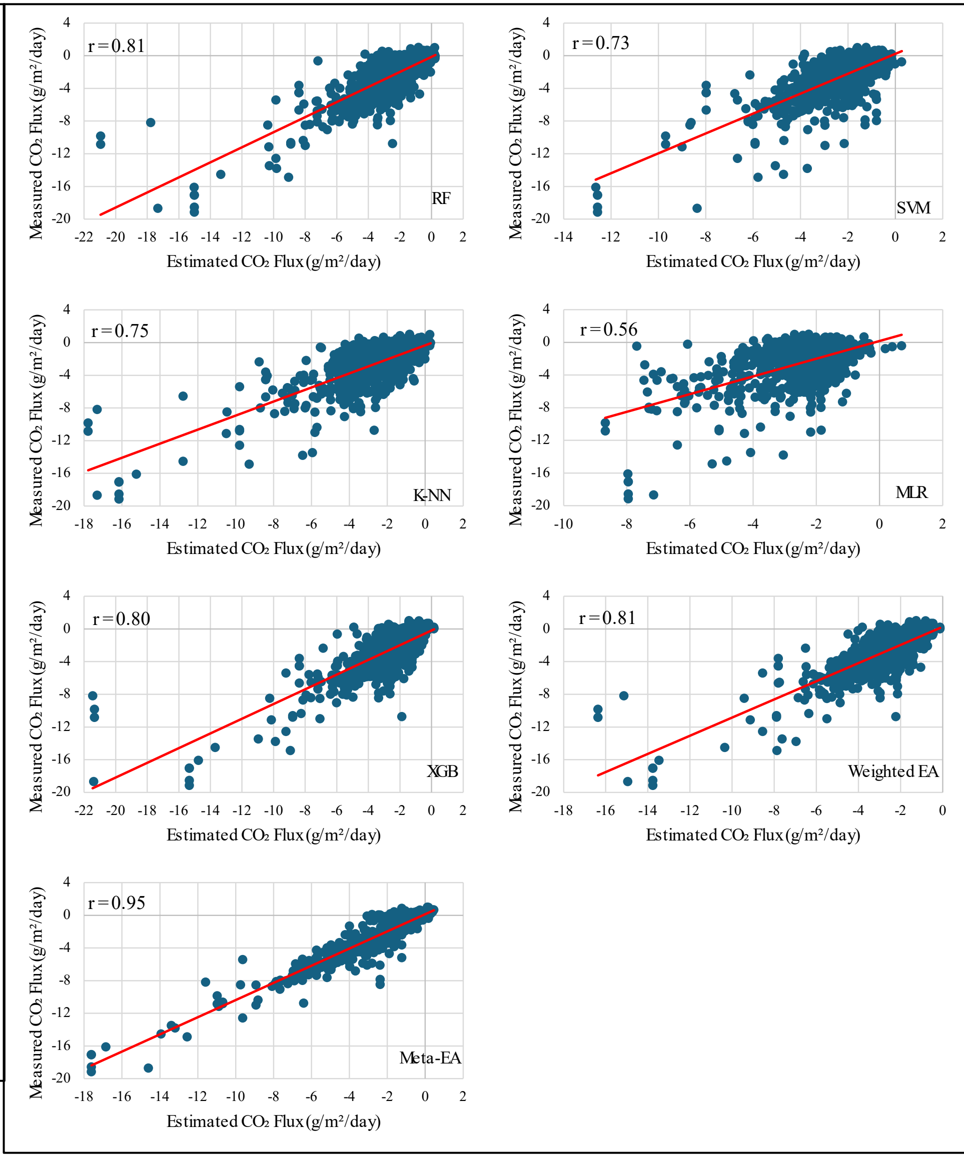
CO_2 Flux Estimation (HLS imagery: December 2021-April 2022)



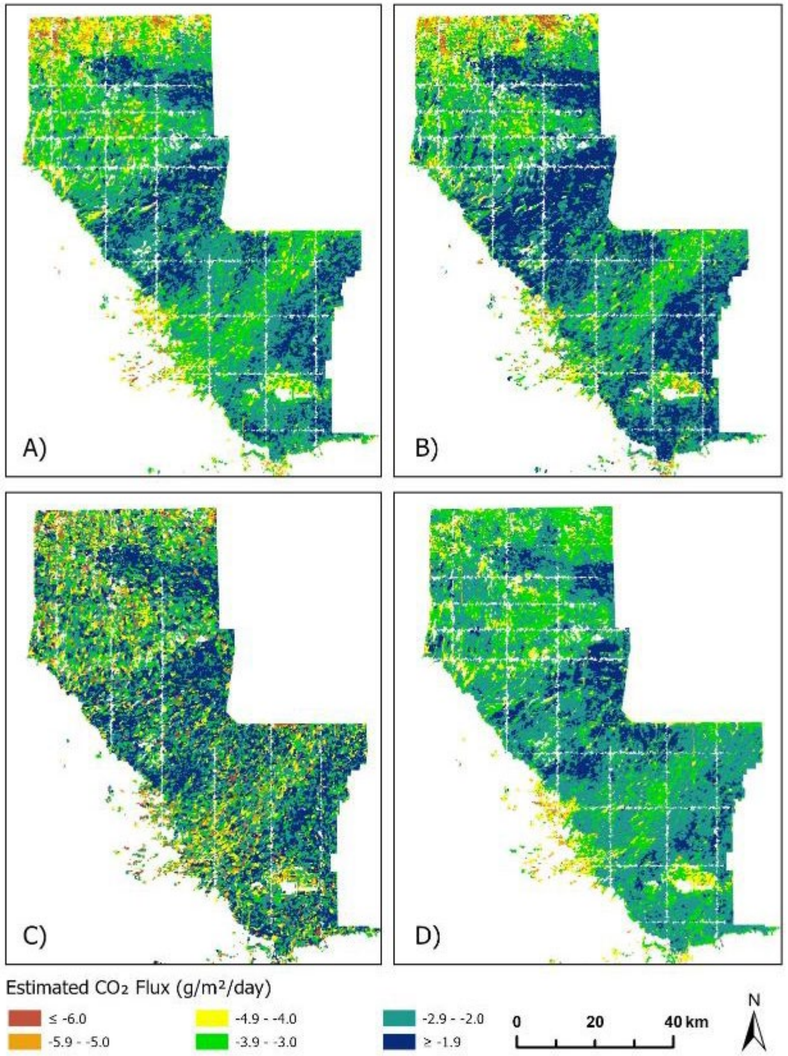


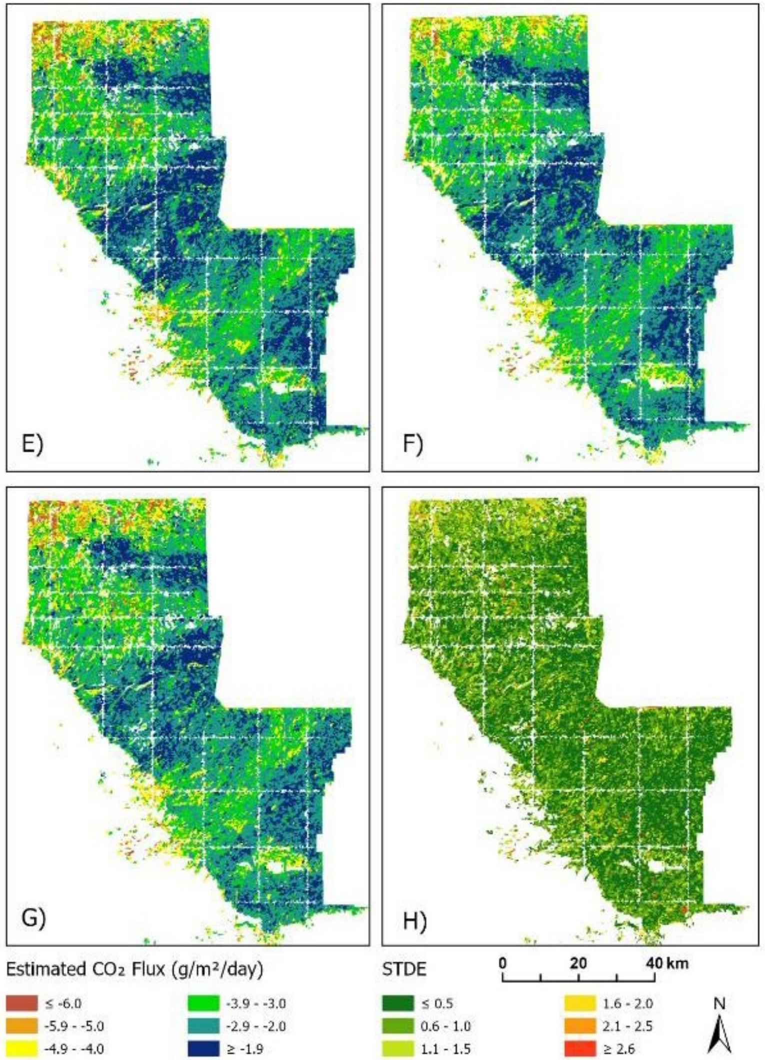
CO_2 Flux Estimation (HLS imagery: May–Nov 2022)

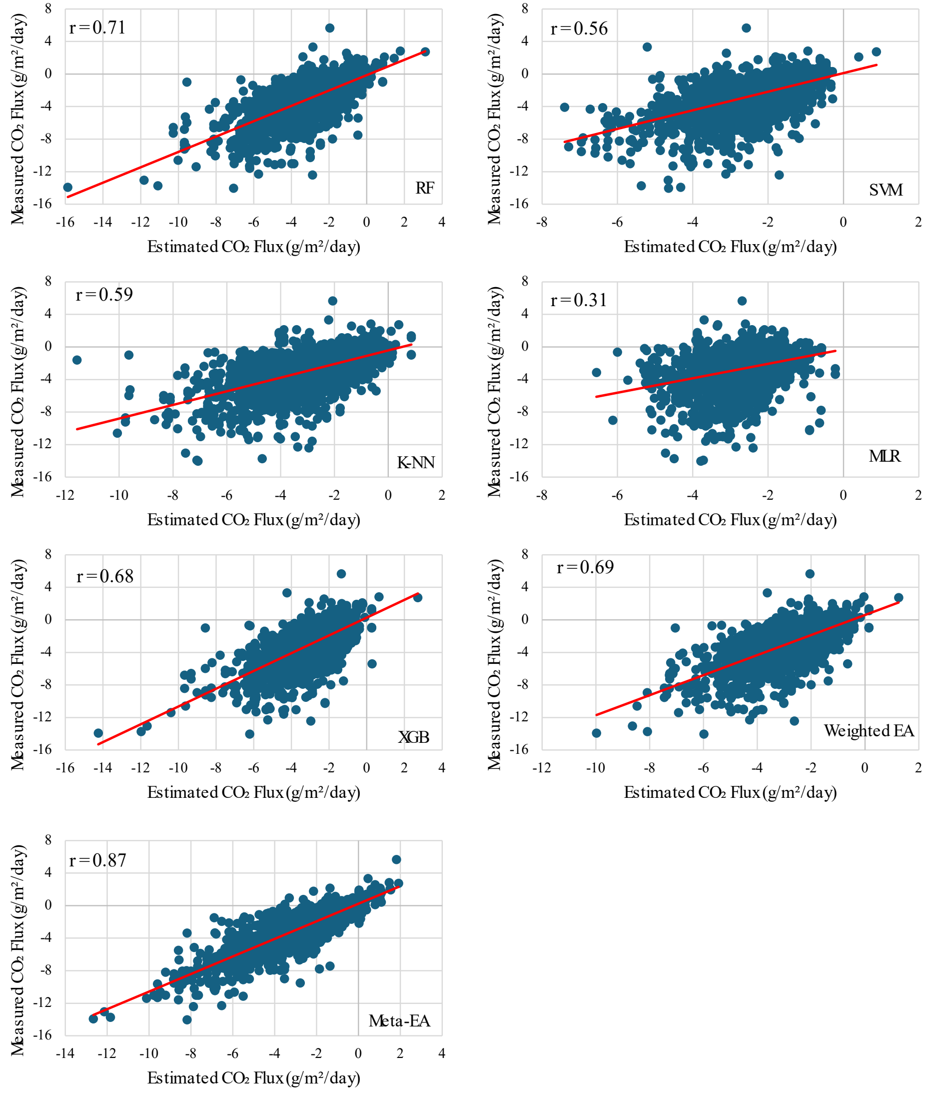




CO_2 Flux Estimation (HLS imagery: December 2022-April 2023)



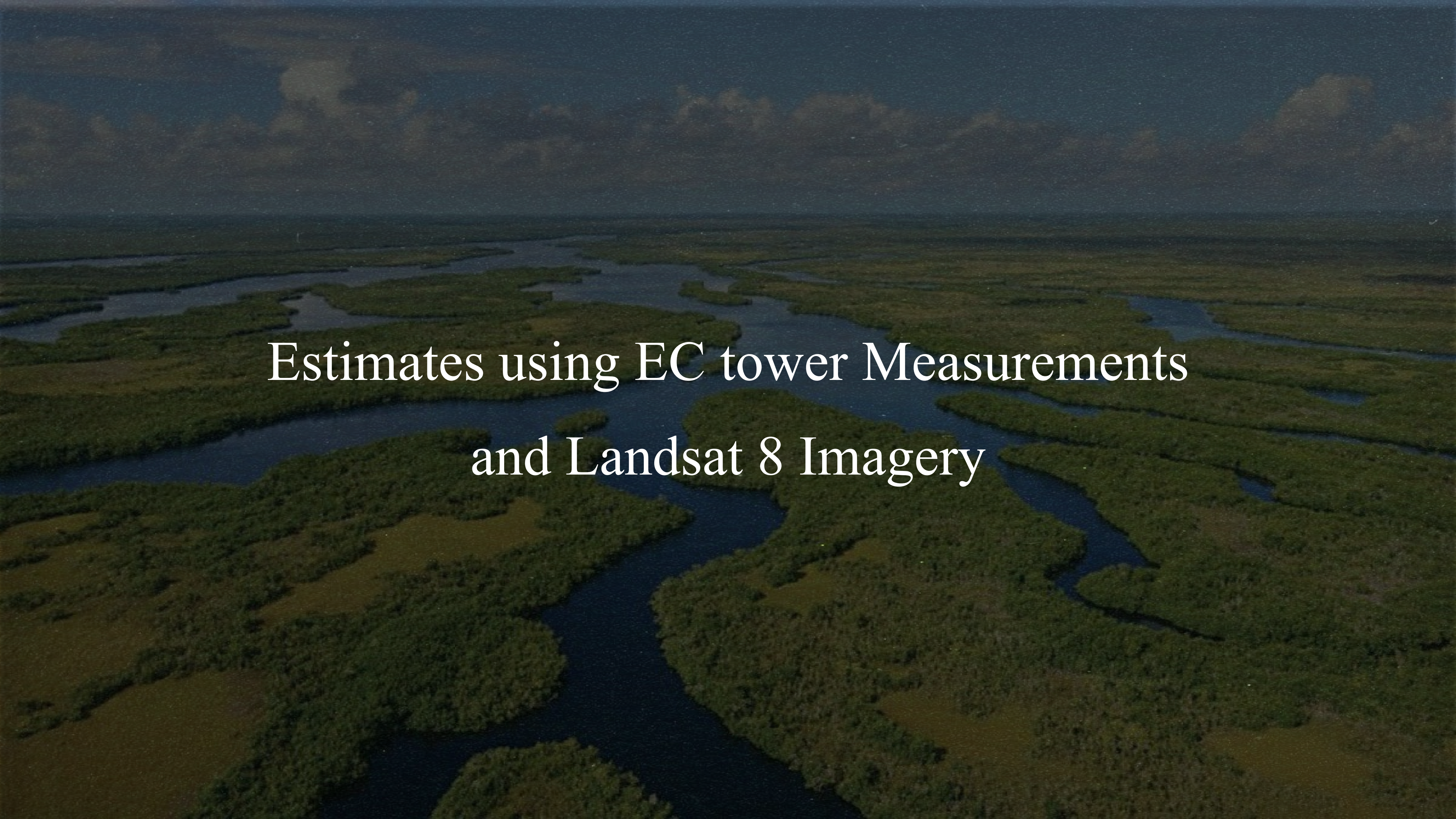




Estimates using BlueFlux Measurements and Harmonized Sentinel-2 MSI imagery

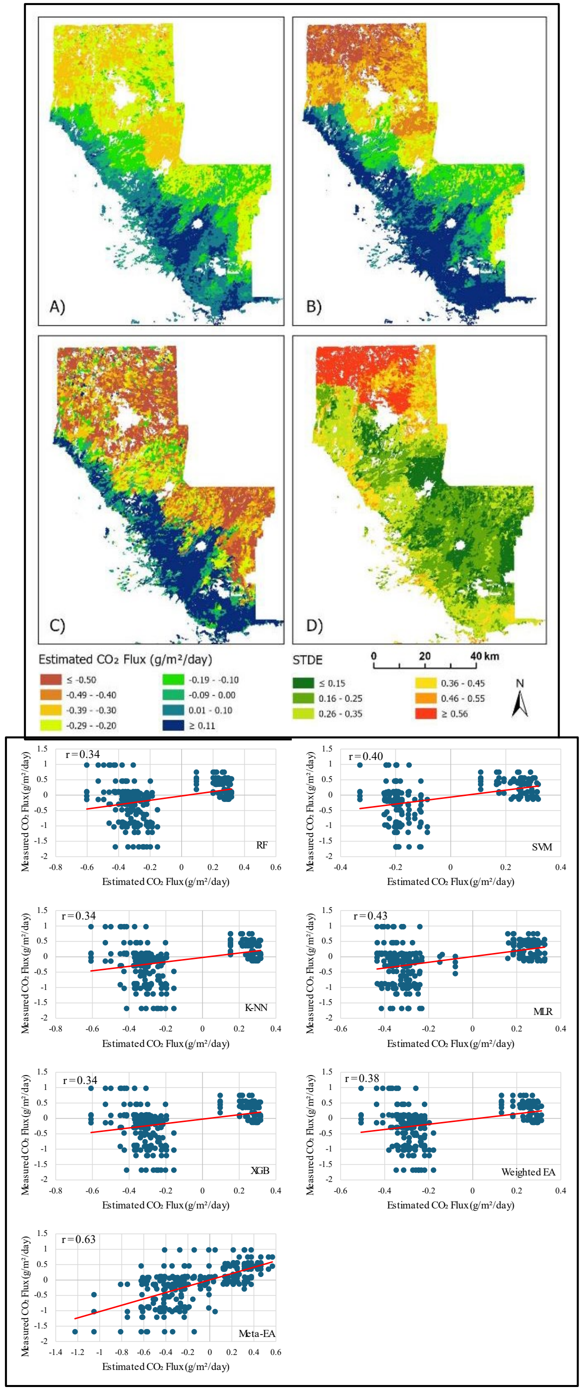
Season	Metric	RF	SVM	KNN	MLR	XGB	Weighted EA	Meta-
December 2021-April 2022	r	0.68	0.65	0.45	0.55	0.70	0.69	**0.95**
	MAE (g/m2/day)	1.99	2.02	2.39	2.29	1.95	1.99	0.87
	RMSE (g/m2/day)	2.77	2.94	3.39	3.13	2.71	2.78	1.28
May 2022-November 2022	r	0.81	0.73	0.75	0.56	0.80	0.81	**0.95**
	MAE (g/m2/day)	0.97	1.16	1.19	1.44	1.00	1.04	0.50
	RMSE (g/m2/day)	1.44	1.73	1.65	2.04	1.49	1.47	0.80
December 2022-April 2023	r	0.71	0.56	0.59	0.31	0.68	0.69	**0.87**
	MAE (g/m2/day)	1.09	1.27	1.26	1.52	1.15	1.15	0.72
	RMSE (g/m2/day)	1.51	1.80	1.74	2.04	1.57	1.58	1.06

Table Predicted CO2 accuracies using BlueFlux data and Harmonized Sentinel-2 MSI imagery for the study area across the three study seasons.

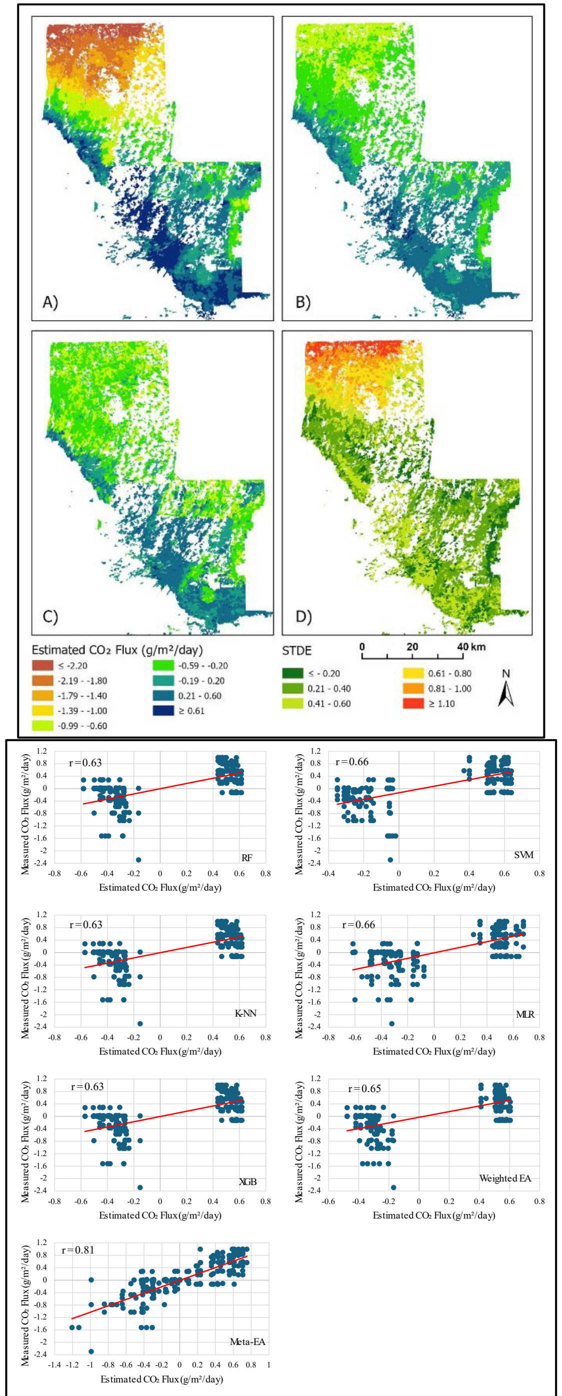


Estimates using EC tower Measurements and Landsat 8 Imagery

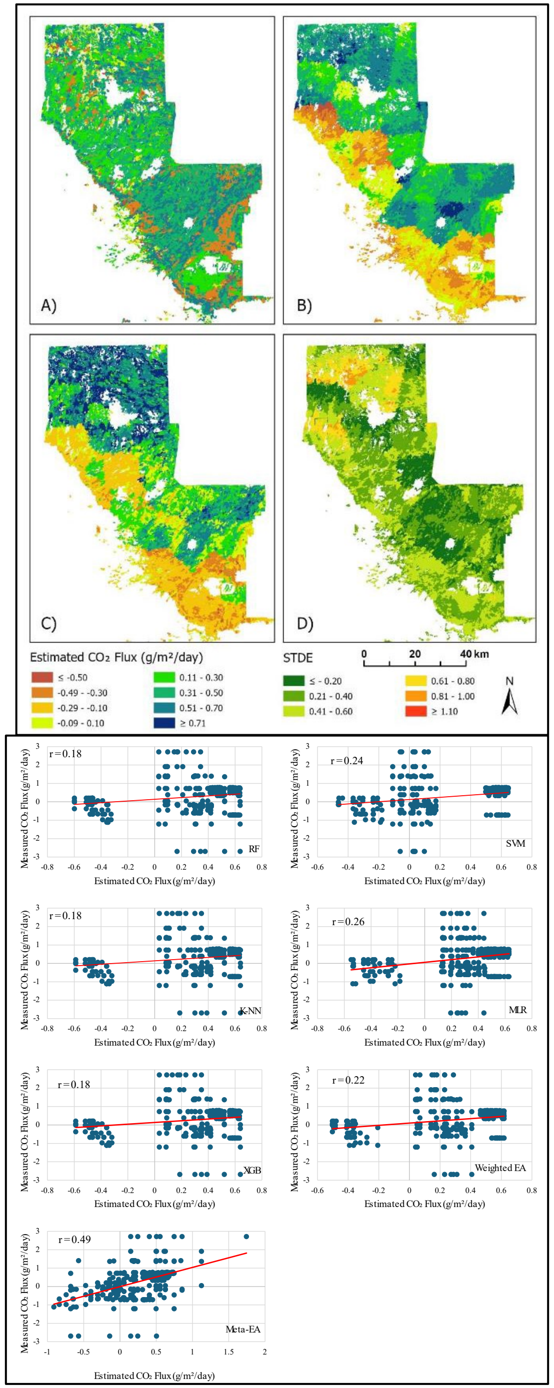
December 2021-April 2022



May–November 2022



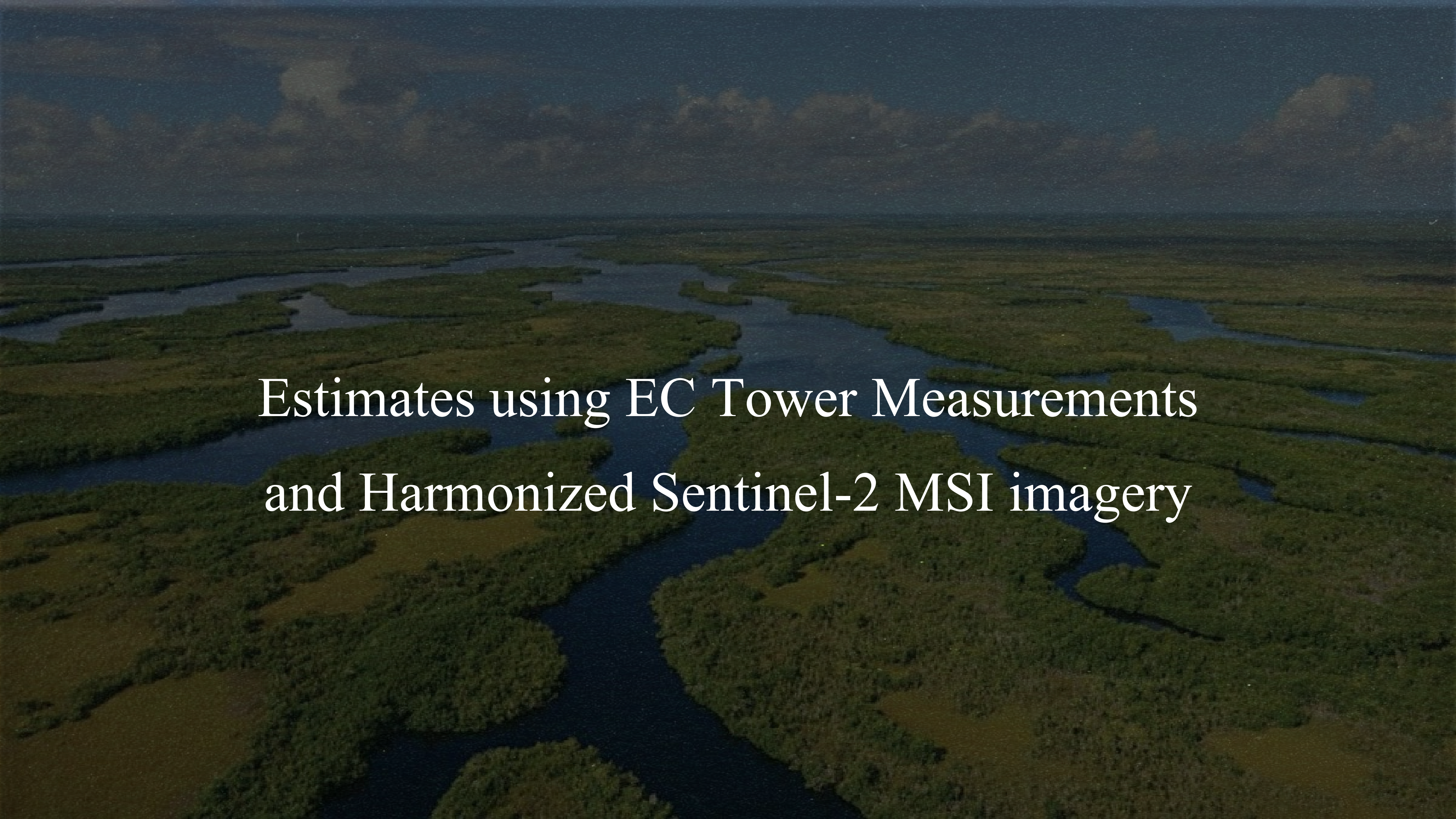
December 2022-April 2023



Estimates using EC tower Measurements and Landsat 8 Imagery

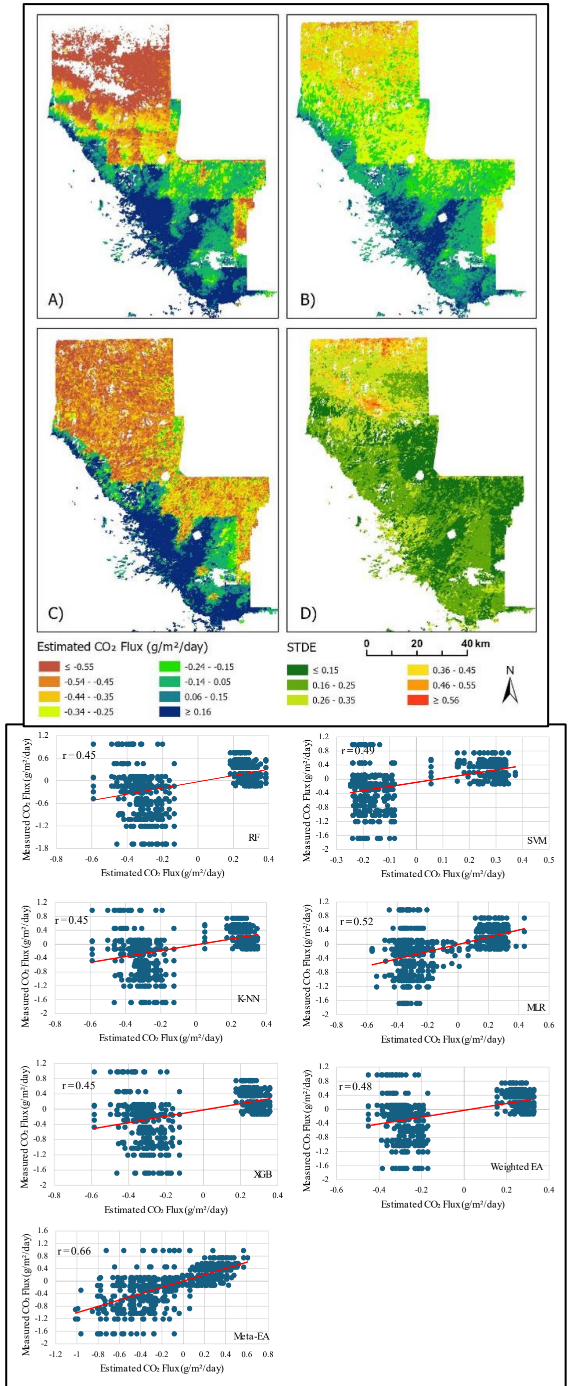
Season	Metric	RF	SVM	KNN	MLR	XGB	Weighted EA	Meta-EA
December 2021-April 2022	*r*	0.34	0.40	0.34	0.43	0.34	0.38	**0.63**
	MAE (g/m²/day)	0.43	0.40	0.43	0.41	0.43	0.42	0.35
	RMSE (g/m²/day)	0.56	0.54	0.56	0.53	0.56	0.54	0.45
May 2022-November 2022	*r*	0.63	0.66	0.63	0.66	0.63	0.65	**0.81**
	MAE (g/m²/day)	0.37	0.33	0.37	0.36	0.37	0.36	0.27
	RMSE (g/m²/day)	0.47	0.47	0.47	0.45	0.47	0.46	0.35
December 2022-April 2023	*r*	0.18	0.24	0.18	0.26	0.18	0.22	**0.49**
	MAE (g/m²/day)	0.57	0.54	0.57	0.55	0.57	0.55	0.49
	RMSE (g/m²/day)	0.86	0.85	0.86	0.83	0.86	0.85	0.75

Table Predicted CO2 accuracies using EC towers data and Landsat 8 imagery for the study area across the three study seasons.

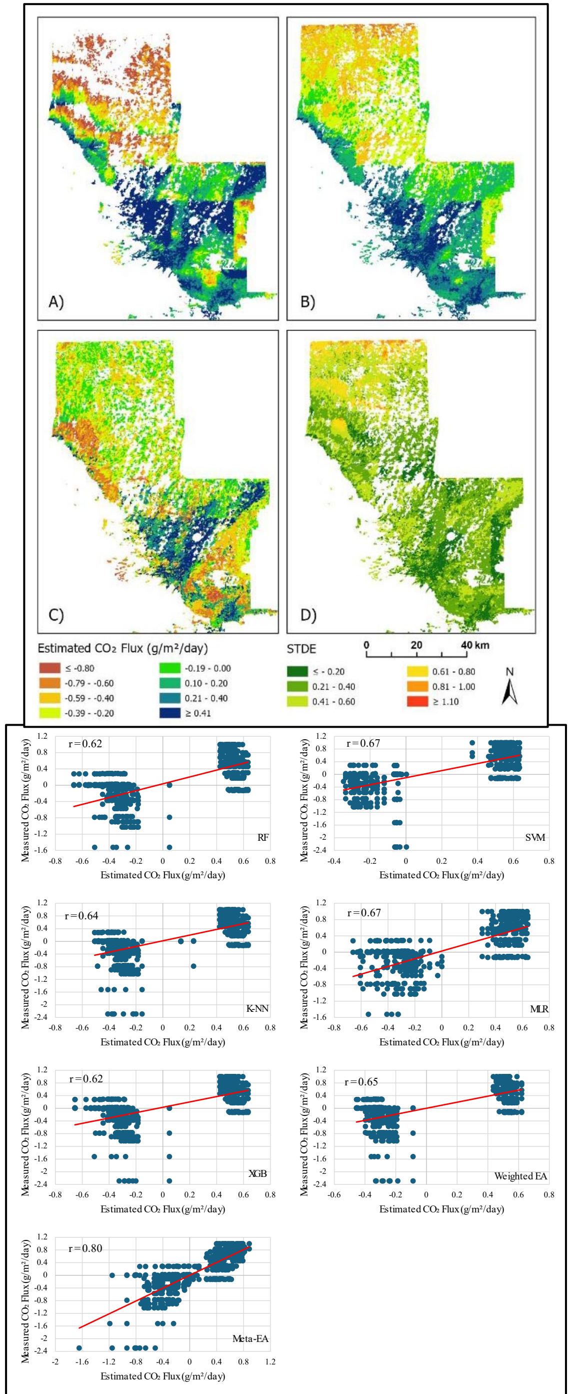


Estimates using EC Tower Measurements and Harmonized Sentinel-2 MSI imagery

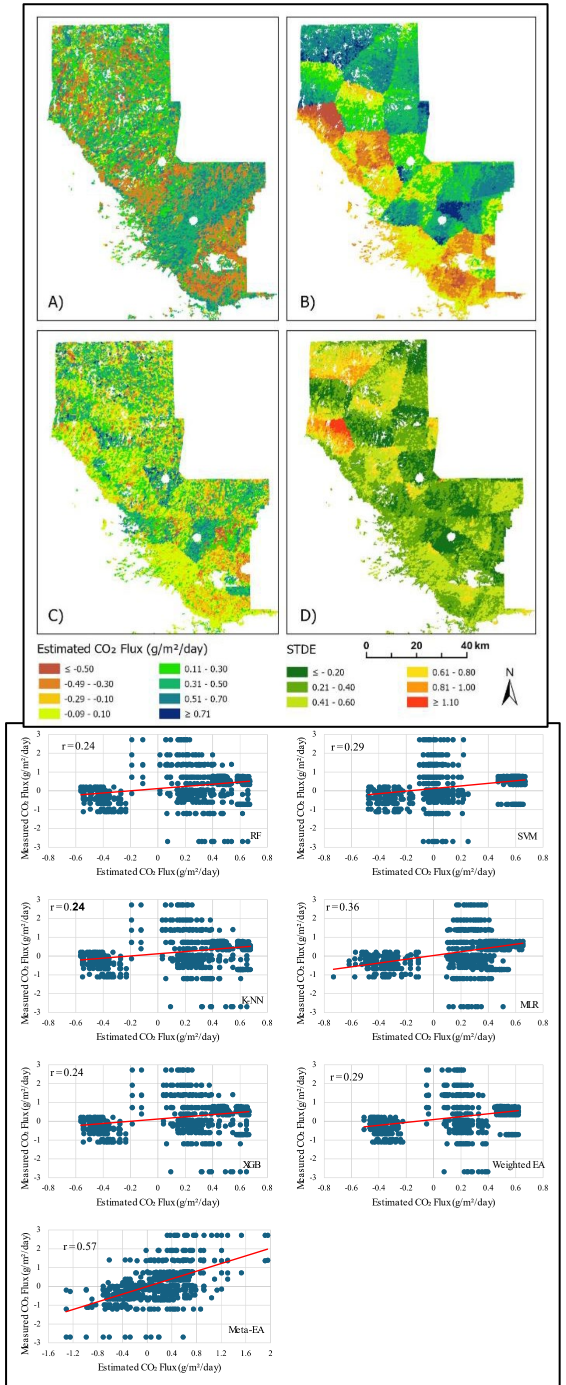
December 2021-April 2022



May–November 2022



December 2022-April 2023



Estimates using EC Tower Measurements and Harmonized Sentinel-2 MSI imagery

Season	Metric	RF	SVM	KNN	MLR	XGB	Weighted EA	Meta-EA
December 2021-April 2022	*r*	0.45	0.49	0.45	0.52	0.45	0.48	**0.66**
	MAE ($g/m^2/day$)	0.37	0.35	0.37	0.35	0.37	0.36	0.29
	RMSE ($g/m^2/day$)	0.49	0.48	0.49	0.47	0.49	0.48	0.41
May 2022-November 2022	*r*	0.62	0.67	0.64	0.67	0.62	0.65	**0.80**
	MAE ($g/m^2/day$)	0.37	0.30	0.36	0.35	0.37	0.34	0.27
	RMSE ($g/m^2/day$)	0.49	0.46	0.48	0.46	0.49	0.47	0.37
December 2022-April 2023	*r*	0.24	0.29	0.24	0.36	0.24	0.29	**0.57**
	MAE ($g/m^2/day$)	0.57	0.54	0.57	0.55	0.57	0.55	0.48
	RMSE ($g/m^2/day$)	0.86	0.85	0.86	0.81	0.86	0.84	0.72

Table Predicted CO2 accuracies using EC towers data and Harmonized Sentinel-2 MSI imagery for the study area across the three study seasons.

CONCLUSION

Key Findings

- ❖ Airborne BlueFlux data outperformed EC tower data in all seasons.
- ❖ Harmonized Sentinel-2 MSI imagery provided smoother and improved results over Landsat 8.
- ❖ Meta-Ensemble Analysis achieved the highest accuracy across all data sources and seasons.
- ❖ K-NN and MLR excelled with EC tower data due to simpler models fitting better with limited data.
- ❖ Wet season (May-Nov) produced the best results for CO_2 flux estimations across all scenarios.
- ❖ BlueFlux data with Sentinel-2 MSI imagery is a reliable method for upscaling CO_2 flux in wetland ecosystems.
- ❖ Future work should expand the use of Airborne data and ensemble models for better accuracy and refine carbon monitoring efforts across longer time periods.
- ❖ Carbon uptake rate and emission vary between wetland types. Coastal wetlands have the capacity to absorb substantial CO2 (Chmura, 2013) but release very low levels of CH4 (Chmura et al., 2003).


THANK YOU!