Natural Resources Canada Ressources naturelles Canada

Assessing uncertainties related to satellite remote sensing indices to estimate Gross Primary Production

Ronny A. Hernández Mora.

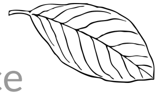
ronnyhdez

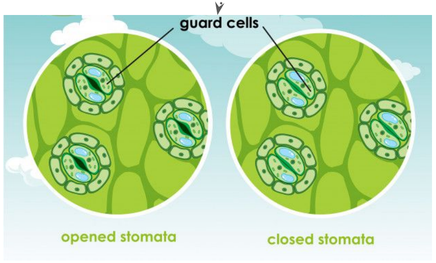
https://ronnyale.com/

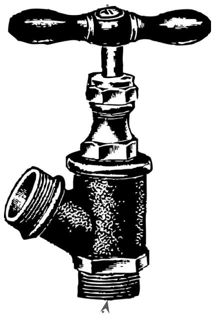
Gross Primary Production (GPP)

Depends on:
- Temperature
- Light
- Air humidity
- Water soil availability

Depends on:
- Stomatal conductance
- Humidity of the interior air of the leaf







Relation between the amount of water transpired per unit of carbon fixed

Leaf level: Well known!

Canopy level: Well known!

(but difficult)

Global level: WIP

Leaf spectral properties (Shull 1929; Gates 1965)
Leaf gas exchange (several before 70's)
RuBisCO (Farquhar 1979 and others)
Stomatal conductance (Ball et al 1987)
Chlorophyll fluorescence (Genty et al 1989)
PRI (Gamon et al 1992)
Modelling photosynthesis with transpiration (Collatz et al 1992)

LUE model (Monteith 1972 1977)
Two leaf model (Sinclair 1976)
Canopy photosynthesis EC (Baldocchi et al 1987)
Measuring canopy SIF (Meroni et al 2009; van der Tol et al 2009)
Evolving FLUXNET (Baldocchi et al 2001)

Landsat launched (1972)
Simple Biosphere model (Sellers 1985)
Global productivity models (90's)
MODIS photosynthesis and land products (Running et al 2004)
Machine learning global photosynthesis (Beeri et al 2010)
Global remote sensing of SIF (Frankenberg et al 2011)

How do we estimate GPP?

Eddy covariance

The eddy covariance technique has gained popularity in the last decade thanks to its methodology that allows the collection of large amounts of data in an automated way on the **exchange of heat, water vapor, carbon dioxide and other gases** between the surface of ecosystems and the low part of the atmosphere (Knauer et al., 2018)

$$NEE = GPP - ER_{day}$$

Satellites

Three methods:

- Vegetation Index models (VI)
- Light Use Efficiency models (LUE)
- Process-based models

NDVI

EVI

CCI

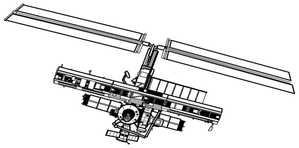
NIR_v

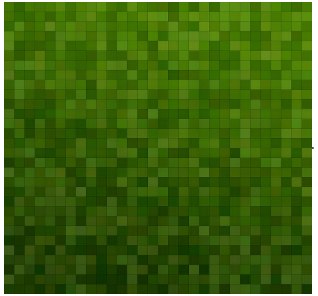
GPP = PAR * fAPAR * **LUE**

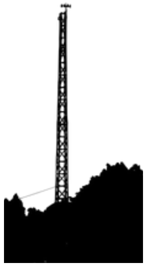
Global photosynthesis estimates uncertainties

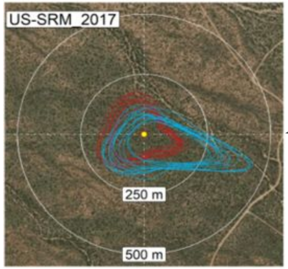
Annual sum values ranged from 112 to 169 PgC y^{-1}, interannual variability ranged from 0.8 to 4.4 PgC y^{-1}, and trends varied from 0.005 to 0.621 PgC y^{-2} (Anav *et al.*, 2015)

Research objective









NDVI, EVI, NIRv, and CCI

GPP

NEE = GPP−ER$_{day}$

Research objectives

Chap 2:

- To quantify and compare the uncertainty inherent in these models, particularly when the variation in environmental conditions (E) is controlled.
- To determine if a single VI outperforms other VIs reducing the errors when predicting GPP.

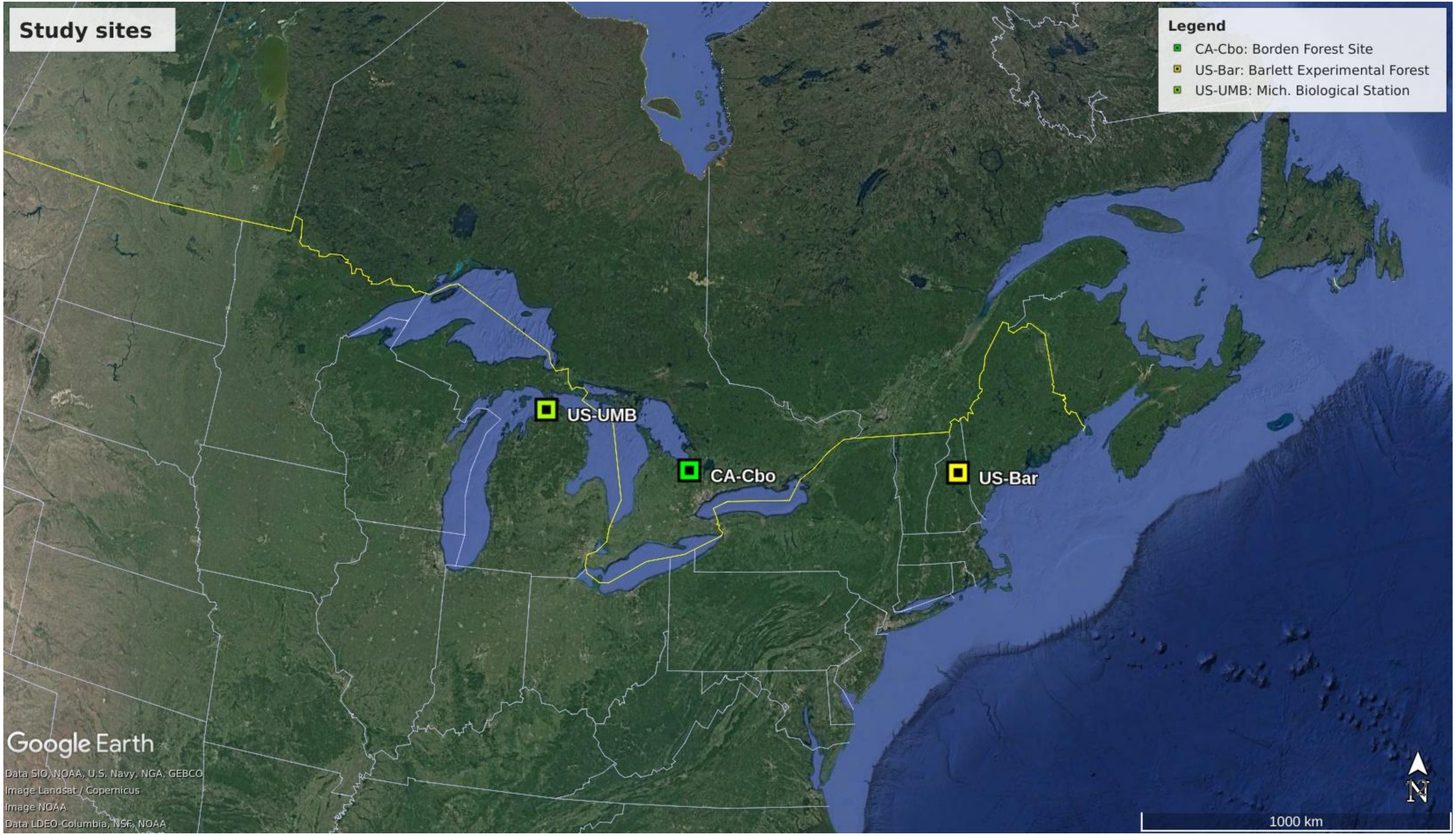
Chap 3:

- Does ML approaches perform better than traditional methods to predict GPP?
- What are the main drivers of predictions outcomes when using ML methods?

"Prediction" term should not be confused with forecasting, given that most of the models are not aiming at predicting into the future, instead the focus is to predict in the past or the present times.

Methods

Methods | Sites location



Methods | Sites data range available

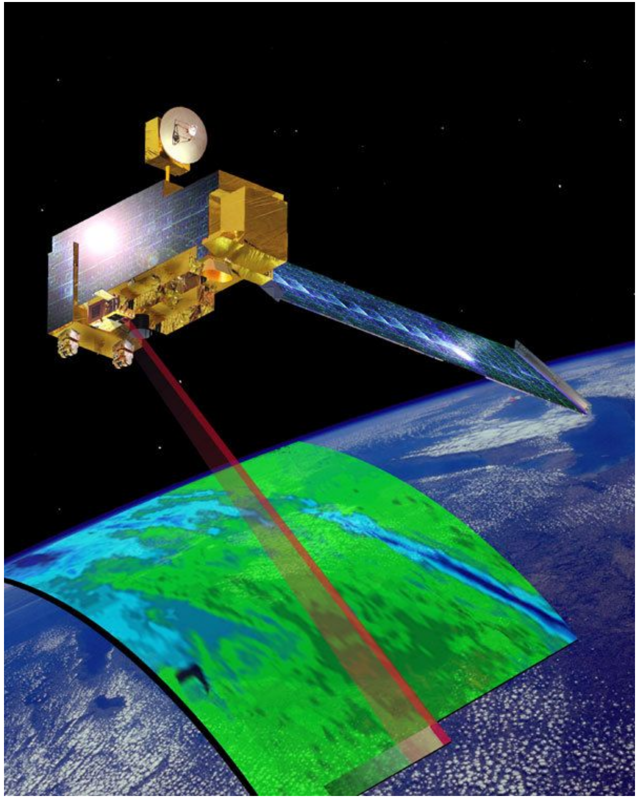
ONEFlux sites datasets description

Site	Data range available	Dataset name	Reference
Bartlett	Jan 2015 to Dec 2017	US-Bar: Barlett Experimental Forest (version: beta-3)	(Staebler 2019)
Borden	Jan 2015 to Jan 2022	CA-Cbo: Ontario - Mixed Deciduous, Borden Forest Site	(Richardson and Hollinger 2016)
Michigan	Jan 2015 to Jan 2018	US-UMB: Univ. of Mich. Biological Station (version: beta-4)	(C. Gough, Bohrer, and Curtis 2016)

Methods | MODIS

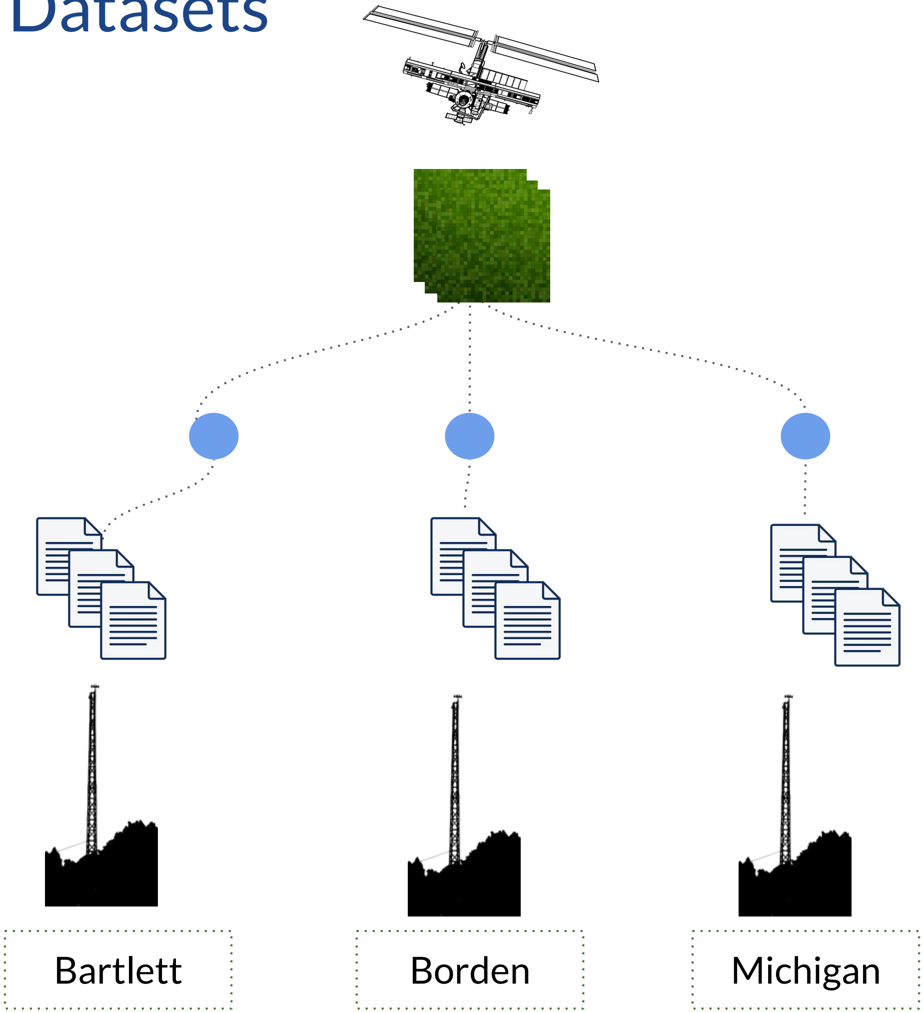
MODIS (MOD09GA.061 product) bands used to calculate the VIs

Name	Description	Resolution	Wavelength
sur_refl_01	Red	500 meters	620-670nm
sur_refl_02	NIR	500 meters	841-876nm
sur_refl_03	Blue	500 meters	459-479nm
sur_refl_04	Green	500 meters	545-565nm



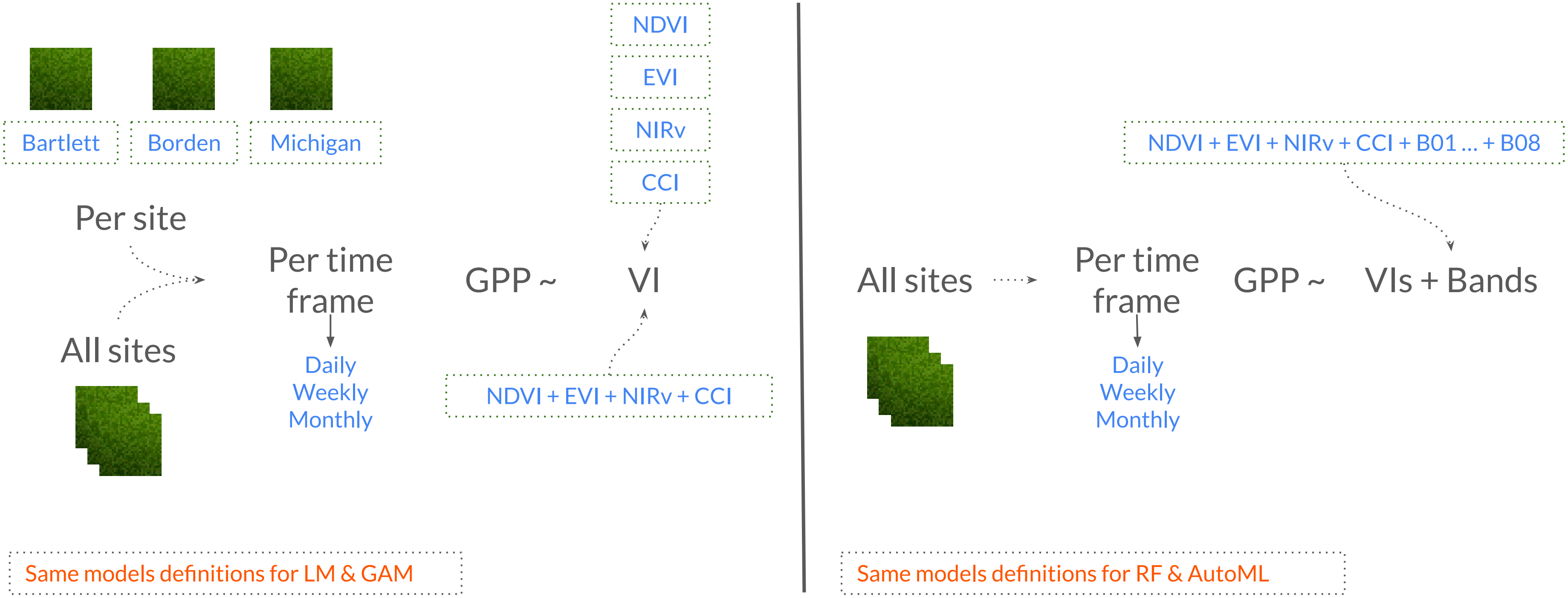
Comparison of meteorological observations and WRF model retrieved Cloud Top Height for the JEM-EUSO experiment. 2015/04/01. Tabone, Ilaria

Methods | Datasets



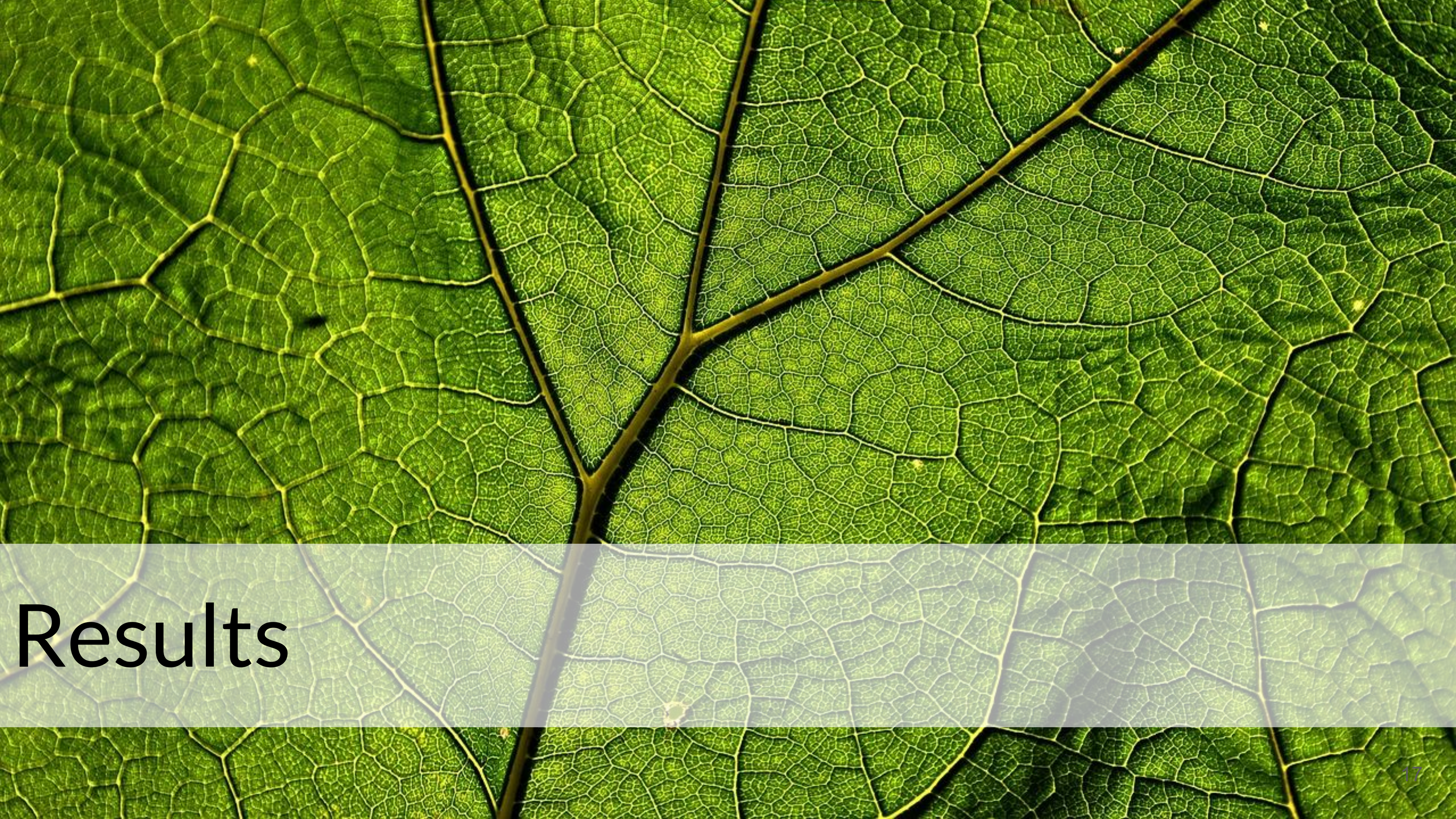
- Select pixels with highest data quality for each of the sites.
- Summarize pixels values in weekly and monthly (daily is the default)
- At least two years of records of GPP.
- Join GPP and MODIS datasets per date (daily, weekly, and monthly)

Data analysis



Chapter 2

Chapter 3



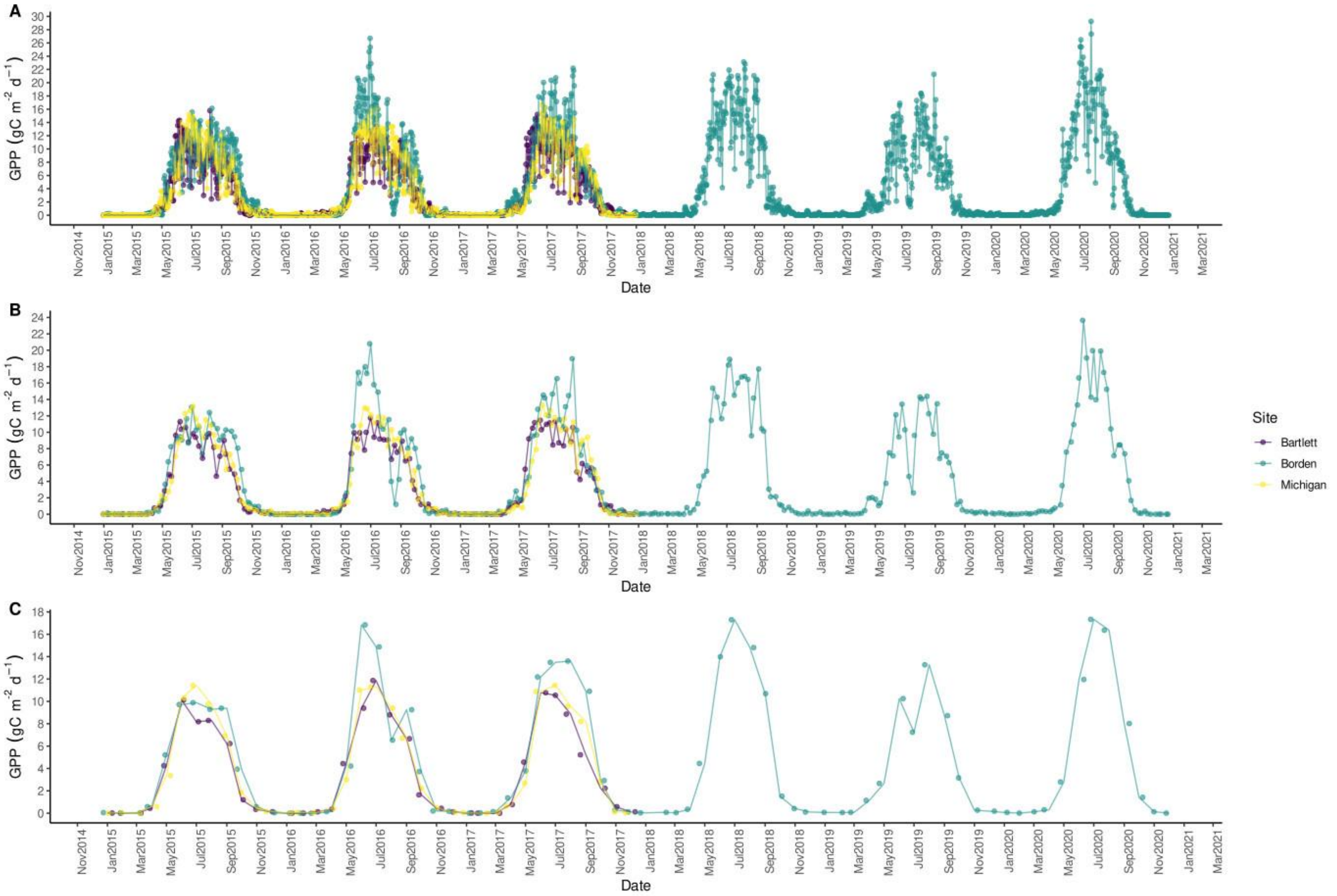
Results

Results | Sites GPP trends

Daily

Weekly

Monthly



Results | Linear models

Table 2.3: Summary of Linear models for GPP estimation using the vegetation indices on a monthly (a), weekly (b), and daily (c) basis. MAE and RMSE metrics units are gC m- ^{2} d- ^{1}

Monthly

(a)

Site	EVI			NDVI			NIRv			CCI			All		
	R2	MAE	RMSE	R2	MAE	RMSE	R2	MAE	RMSE	R2	MAE	RMSE	R2	MAE	RMSE
Bartlett	0.89	0.92	1.08	0.80	1.23	1.44	0.89	0.89	1.08	0.89	0.86	1.05	0.92	0.72	0.93
Michigan	0.86	1.05	1.35	0.72	1.62	1.92	0.86	1.05	1.33	0.87	1.02	1.29	0.94	0.72	0.90
Borden	0.75	1.92	2.54	0.76	1.95	2.46	0.75	1.90	2.50	0.79	1.82	2.32	0.80	1.72	2.24
All	0.57	2.21	2.87	0.64	1.95	2.62	0.59	2.18	2.81	0.74	1.69	2.26	0.78	1.63	2.04

Weekly

(b)

Site	EVI			NDVI			NIRv			CCI			All		
	R2	MAE	RMSE	R2	MAE	RMSE	R2	MAE	RMSE	R2	MAE	RMSE	R2	MAE	RMSE
Bartlett	0.79	1.16	1.54	0.64	1.62	2.03	0.77	1.21	1.61	0.79	1.23	1.56	0.82	1.06	1.43
Michigan	0.78	1.50	1.79	0.61	1.99	2.39	0.78	1.50	1.78	0.71	1.72	2.09	0.81	1.43	1.67
Borden	0.57	2.62	3.54	0.50	2.91	3.83	0.57	2.64	3.58	0.58	2.62	3.52	0.61	2.53	3.39
All	0.45	2.65	3.53	0.46	2.64	3.50	0.46	2.61	3.51	0.56	2.35	3.16	0.56	2.34	3.15

Daily

(c)

Site	EVI			NDVI			NIRv			CCI			All		
	R2	MAE	RMSE	R2	MAE	RMSE	R2	MAE	RMSE	R2	MAE	RMSE	R2	MAE	RMSE
Bartlett	0.81	1.42	1.98	0.66	2.01	2.68	0.80	1.44	2.05	0.83	1.42	1.88	0.87	1.20	1.67
Michigan	0.70	1.87	2.33	0.52	2.31	2.96	0.67	1.92	2.48	0.61	2.06	2.67	0.72	1.81	2.27
Borden	0.43	3.42	4.48	0.28	3.88	5.04	0.38	3.57	4.66	0.44	3.31	4.42	0.53	3.07	4.08
All	0.49	3.07	4.12	0.45	3.22	4.30	0.48	3.14	4.19	0.59	2.71	3.70	0.61	2.65	3.63

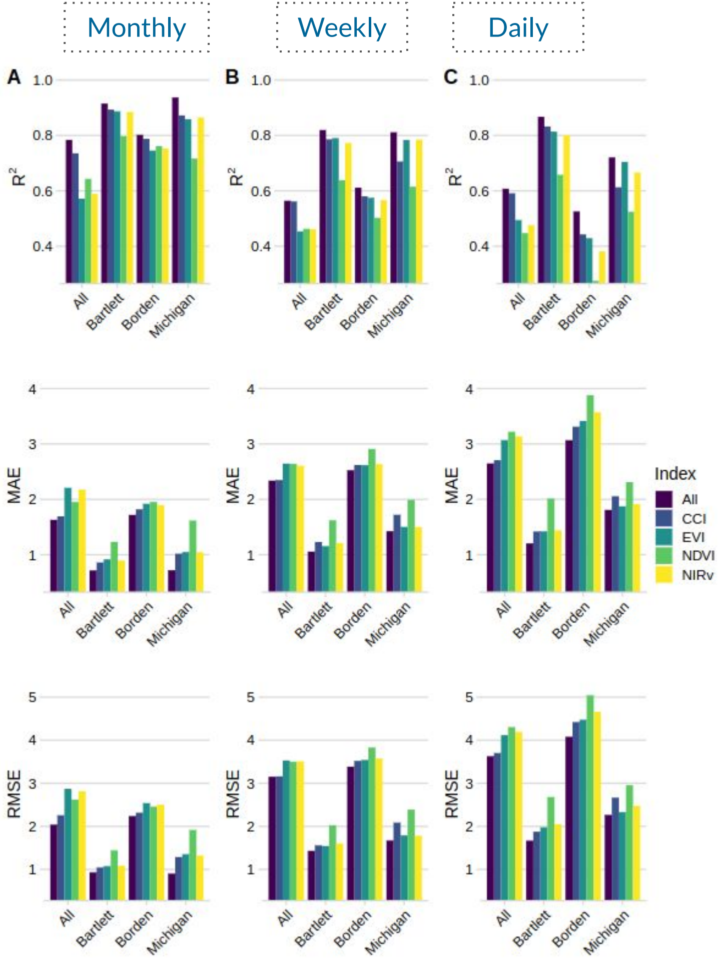


Figure 2.5: Summary of Linear models for GPP estimation using the vegetation indices on a monthly (a), weekly (b), and daily (c) basis. MAE and RMSE metrics units are gC m ^{2} d ^{1}

Chap. 2

Results | GAM models

Table 2.4: Summary of GAM models for GPP estimation using the vegetation indices on a monthly (a), weekly (b), and daily (c) basis. MAE and RMSE metrics units are gC m- ² d- ¹

Monthly

(a)

	EVI			NDVI			NIRv			CCI			All		
Site	R2	MAE	RMSE	R2	MAE	RMSE	R2	MAE	RMSE	R2	MAE	RMSE	R2	MAE	RMSE
All	0.66	1.91	2.50	0.72	1.72	2.27	0.67	1.89	2.47	0.75	1.60	2.16	0.65	2.39	3.37
Michigan	0.96	0.55	0.64	0.86	0.91	1.23	0.96	0.56	0.66	0.91	0.79	0.97	NA	NA	NA
Bartlett	0.88	0.92	1.08	0.78	1.23	1.44	0.88	0.89	1.08	0.91	0.67	0.87	NA	NA	NA
Borden	0.74	1.92	2.54	0.77	1.83	2.37	0.75	1.90	2.50	0.78	1.80	2.30	0.78	1.72	2.24

Weekly

(b)

	EVI			NDVI			NIRv			CCI			All		
Site	R2	MAE	RMSE	R2	MAE	RMSE	R2	MAE	RMSE	R2	MAE	RMSE	R2	MAE	RMSE
All	0.53	2.35	3.25	0.51	2.40	3.34	0.52	2.37	3.29	0.58	2.17	3.06	0.65	2.39	3.37
Michigan	0.84	1.20	1.49	0.67	1.61	2.16	0.82	1.30	1.61	0.75	1.46	1.88	0.86	1.05	1.33
Bartlett	0.81	1.03	1.45	0.63	1.62	2.03	0.79	1.10	1.51	0.79	1.13	1.50	0.82	0.98	1.39
Borden	0.59	2.51	3.41	0.50	2.90	3.82	0.58	2.57	3.49	0.58	2.62	3.52	0.61	2.47	3.35

Daily

(c)

	EVI			NDVI			NIRv			CCI			All		
Site	R2	MAE	RMSE	R2	MAE	RMSE	R2	MAE	RMSE	R2	MAE	RMSE	R2	MAE	RMSE
All	0.58	2.70	3.73	0.55	2.79	3.87	0.56	2.74	3.83	0.61	2.55	3.60	0.65	2.39	3.37
Bartlett	0.85	1.22	1.79	0.74	1.63	2.30	0.83	1.25	1.85	0.85	1.26	1.76	0.89	1.04	1.51
Michigan	0.74	1.67	2.13	0.58	2.01	2.73	0.67	1.84	2.43	0.64	1.92	2.55	0.79	1.45	1.92
Borden	0.50	3.22	4.17	0.42	3.41	4.47	0.46	3.33	4.33	0.45	3.27	4.36	0.55	2.94	3.92

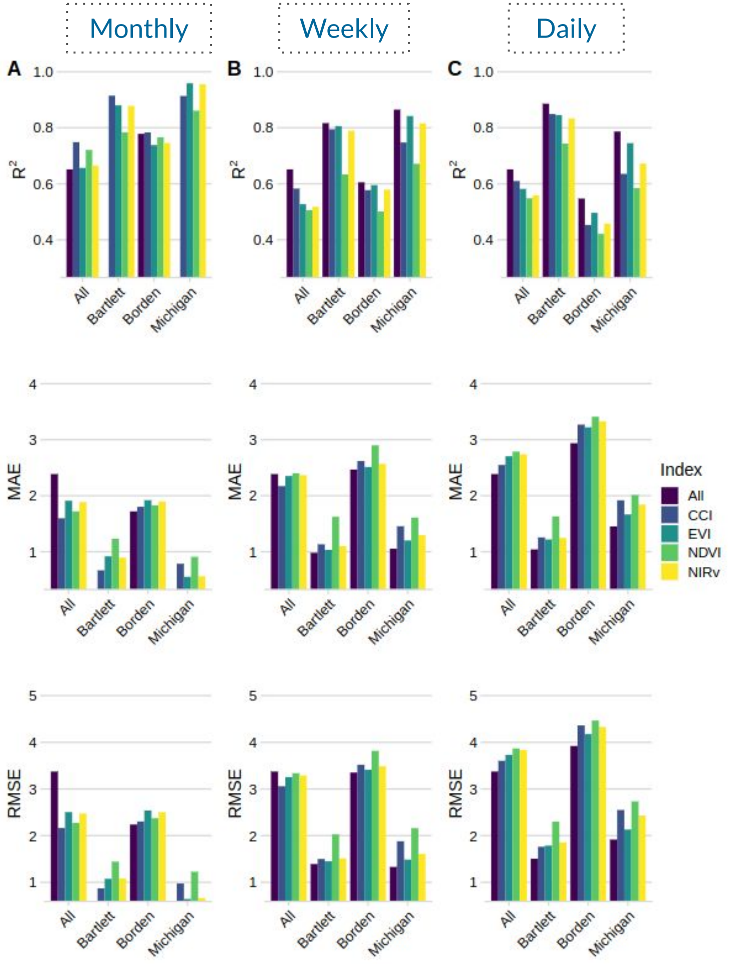


Figure 2.6: Summary of GAM models for GPP (gC m-² d-¹) estimation using the vegetation indices. Column A represents the metrics for the monthly models, B the weekly, and C the daily metrics. MAE and RMSE metrics units are gC m ² d ¹

Chap. 2

Results | GAM models

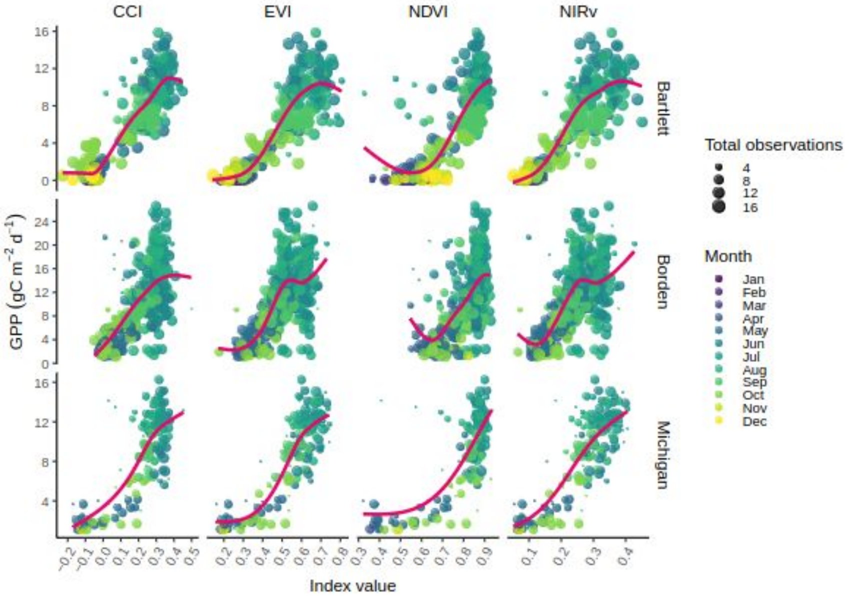


Figure 2.7: Scatterplot of MODIS 500m derived VIs and GPP with daily values. Every observation corresponds to the observed GPP from a flux tower site. Total observations corresponds to the number of observations used to obtain the mean of the vegetation index (NDVI, NIRv, CCI and EVI). The red line indicates the GAM fit.

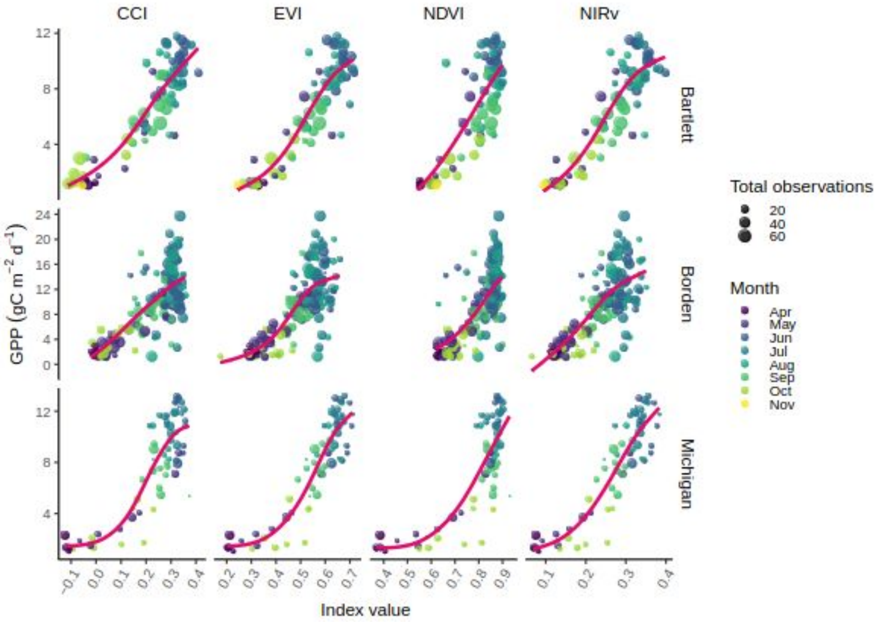


Figure 2.8: Scatterplot of MODIS 500m derived VIs and GPP with weekly values. Every observation corresponds to the observed GPP from a flux tower site. Total observations corresponds to the number of observations used to obtain the mean of the vegetation index (NDVI, NIRv, CCI and EVI). The red line indicates the GAM fit.

Chap. 2

Results | Linear models

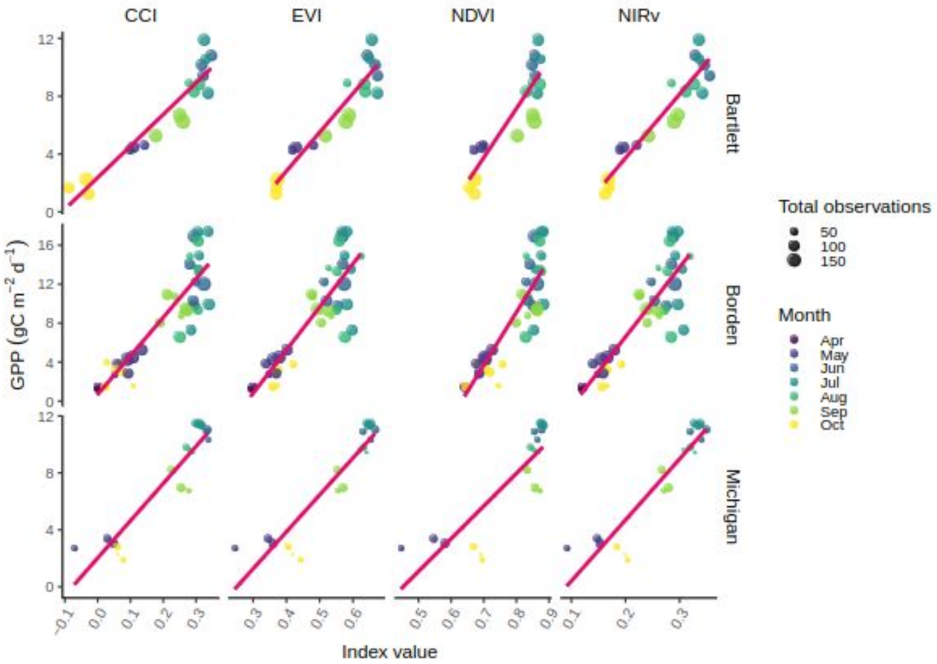
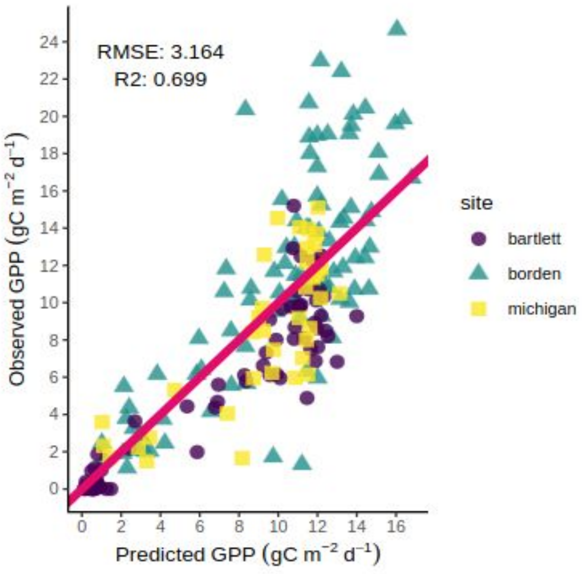


Figure 2.9: Scatterplot of MODIS 500m derived VIs and GPP with monthly values. Every observation corresponds to the observed GPP from a flux tower site. Total observations corresponds to the number of observations used to obtain the mean of the vegetation index (NDVI, NIRv, CCI and EVI). The red line indicates the GAM fit.

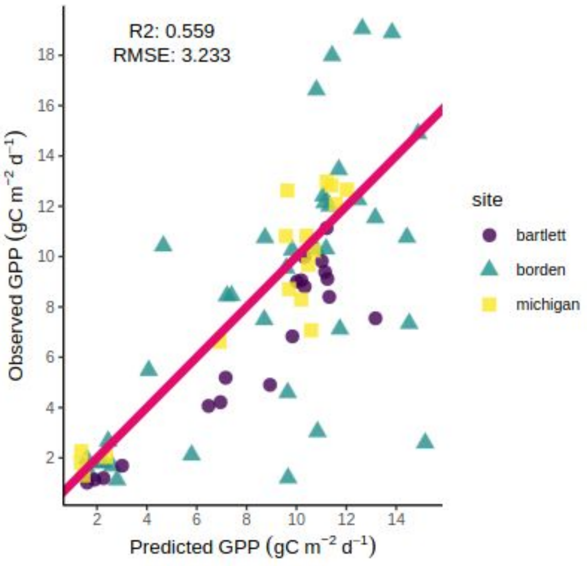
Chap. 2

Results | RF models

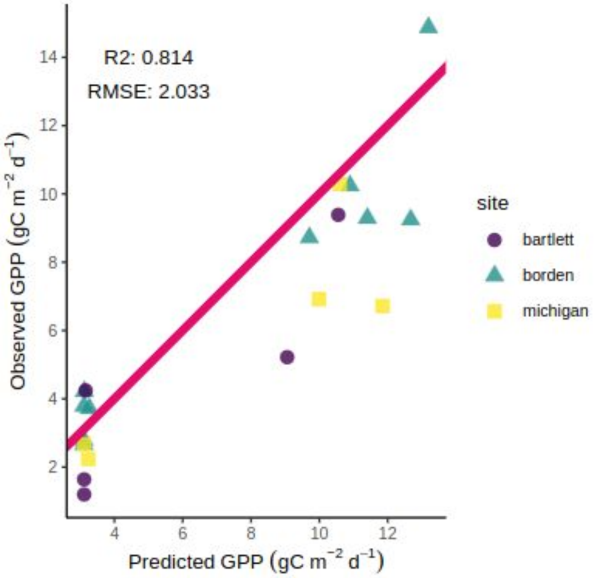
Daily



Weekly



Monthly



Chap. 3

Results | RF models

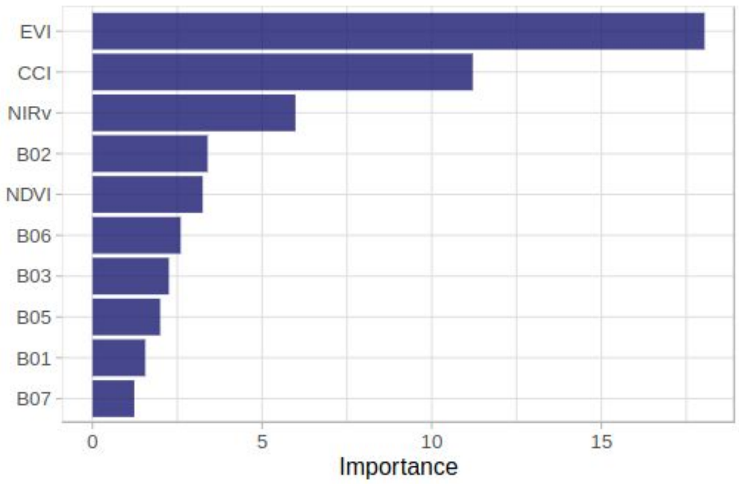


Figure 3.2: Variable of importance derived from the Random forest model for the daily values at 500 m spatial resolution model.

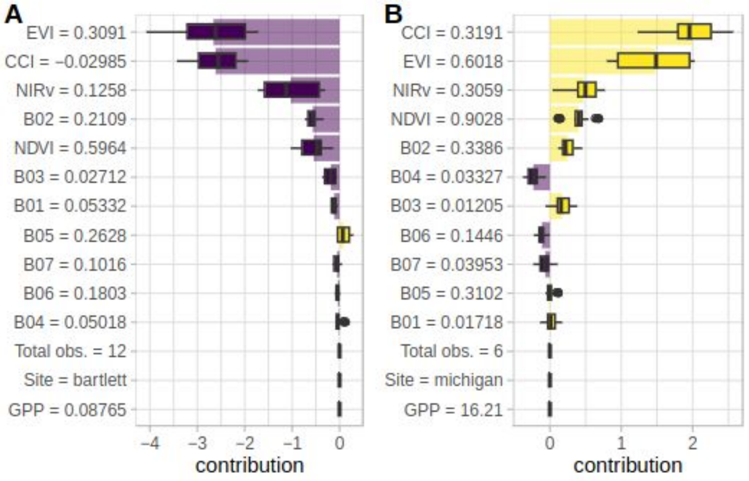


Figure 3.3: Shapley values derived from the Random forest model for the daily values at 500 m spatial resolution model. Predicted value for the low GPP value is 0.59 (A) and 12.7 for the selected high GPP value (B)

Daily

Chap. 3

Results | RF models

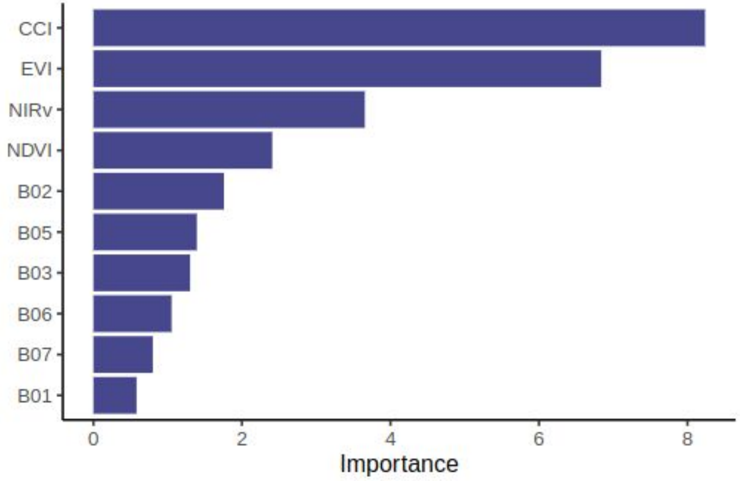


Figure 3.5: Variable of importance derived from the Random forest model for the weekly values at 500 m spatial resolution model.

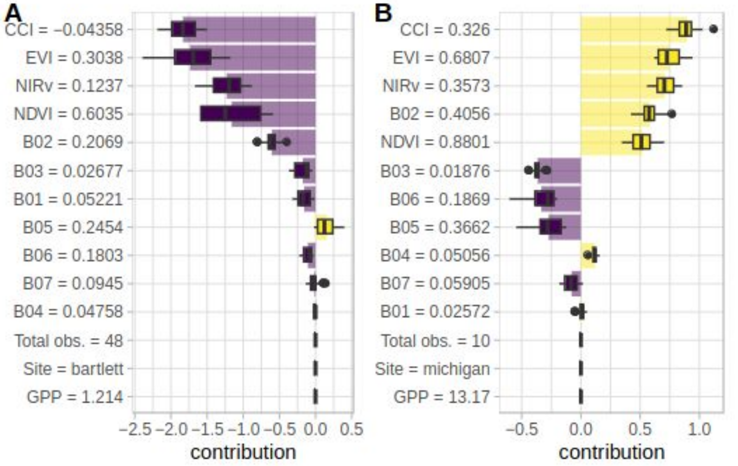


Figure 3.6: Shapley values derived from the Random forest model for the weekly values at 500 m spatial resolution model.

Weekly

Chap. 3

Results | RF models

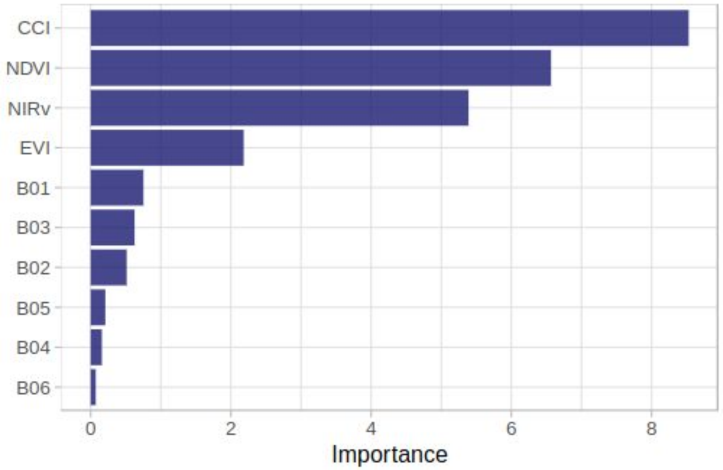


Figure 3.8: Variable of importance derived from the Random forest model for the monthly values at 500 m spatial resolution model.

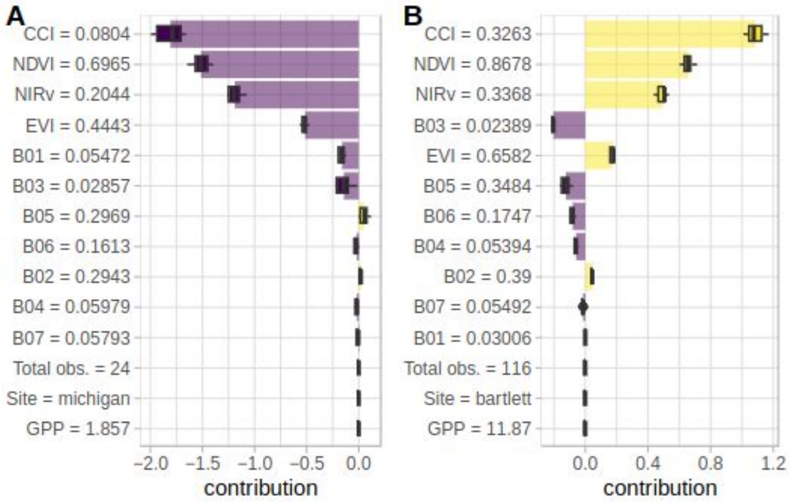


Figure 3.9: Shapley values derived from the Random forest model for the monthly values at 500 m spatial resolution model.

Monthly

Chap. 3

Results | AutoML

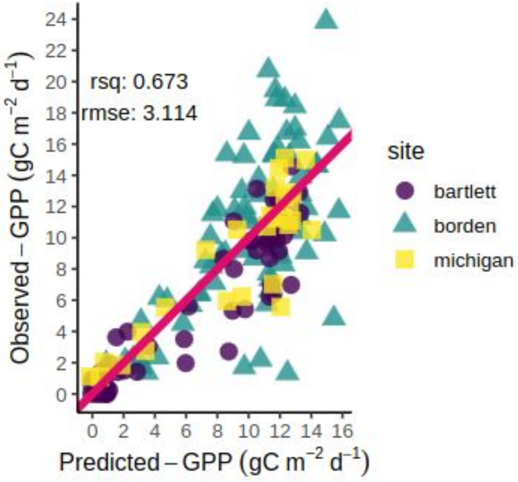


Figure 3.10: GPP observed and predicted values from the autoML for all the sites at a daily basis. The red line represents a 1:1 relation.

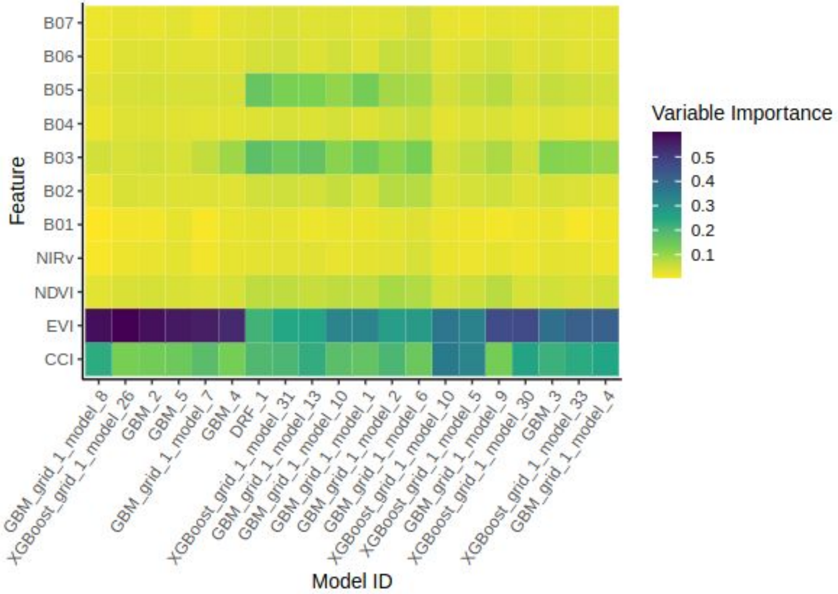


Figure 3.11: Variable of importance derived from the autoML model for the daily values at 500 m spatial resolution model.

Daily

Chap. 3

Results | AutoML

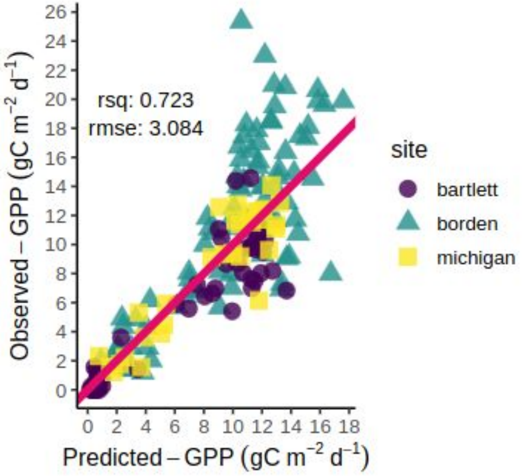


Figure 3.12: GPP observed and predicted values from the autoML for all the sites at a weekly basis. The red line represents a 1:1 relation.

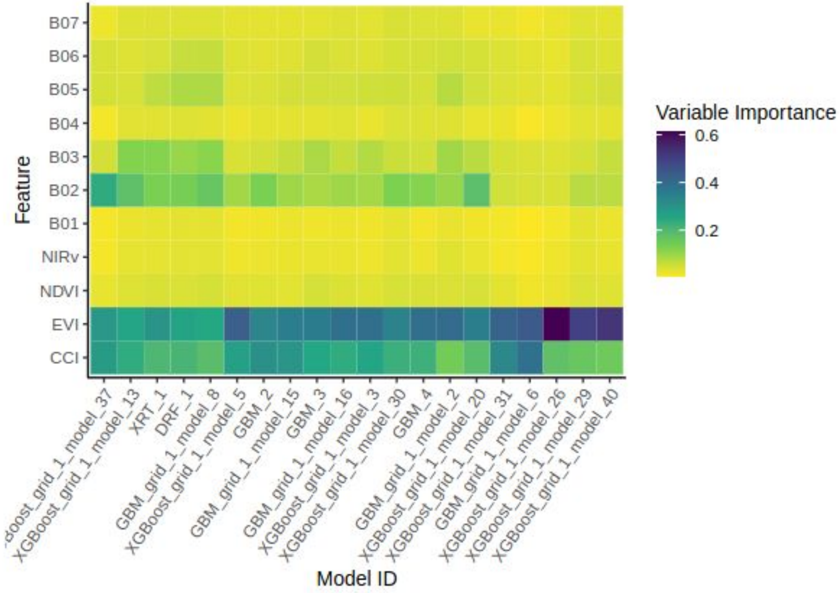


Figure 3.13: Variable of importance derived from the autoML model for the weekly values at 500 m spatial resolution model.

Weekly

Chap. 3

Results | AutoML

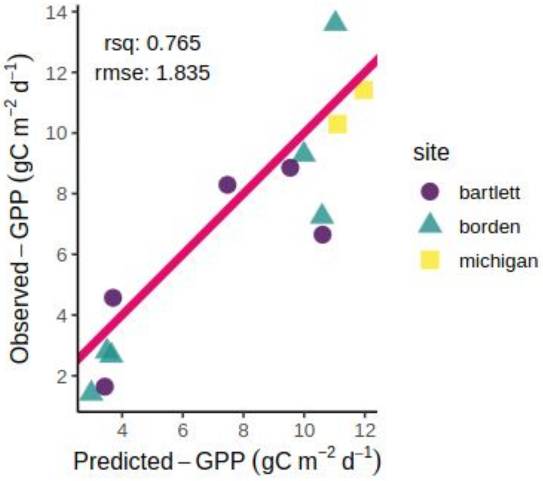


Figure 3.14: GPP observed and predicted values from the autoML for all the sites at a monthly basis. The red line represents a 1:1 relation.

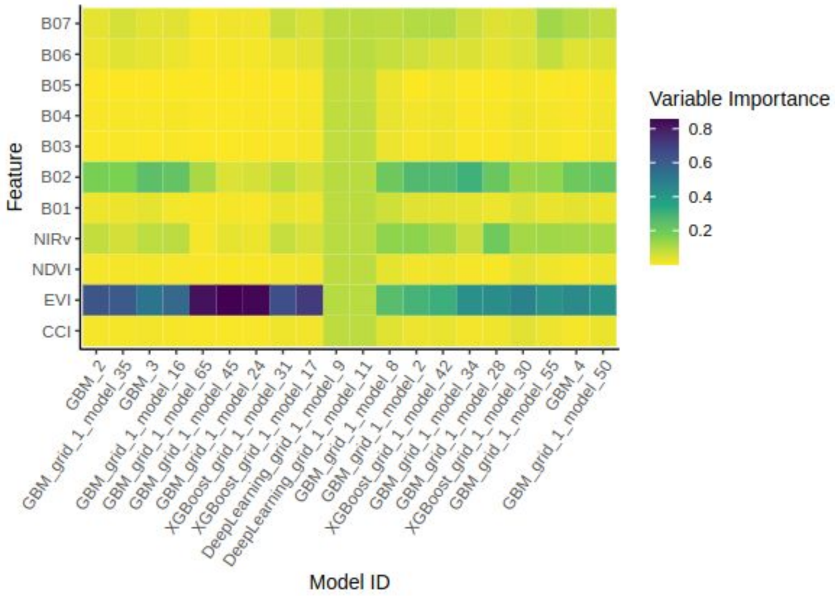


Figure 3.15: Variable of importance derived from the autoML model for the monthly values at 500 m spatial resolution model.

Monthly

Results | ML

Table 3.2.: Summary ML metrics

Variable	RF R2	RF RMSE	automl R2	automl RMSE
Daily	0.70	3.17	0.67	3.11
Weekly	0.55	3.23	0.72	3.08
Monthly	0.81	2.03	0.76	1.84

Conclusions and future work

- **No single VI** will be the best.
- Using all **VIs as covariates** can improve the prediction accuracy.
- While most VIs serve as proxies for vegetation properties, their representation of vegetation functioning has **limitations**.
- **Temporal aggregation** can have an impact on the prediction accuracy.
- **More bands** can capture more information for better predictions.

Scope & limitations

- Number of sites
- Longer time ranges
- Inclusion of more sensors
- Match flux footprint

Thanks!

CEOS
center for earth observation sciences
University of Alberta

Natural Resources Canada
Ressources naturelles Canada

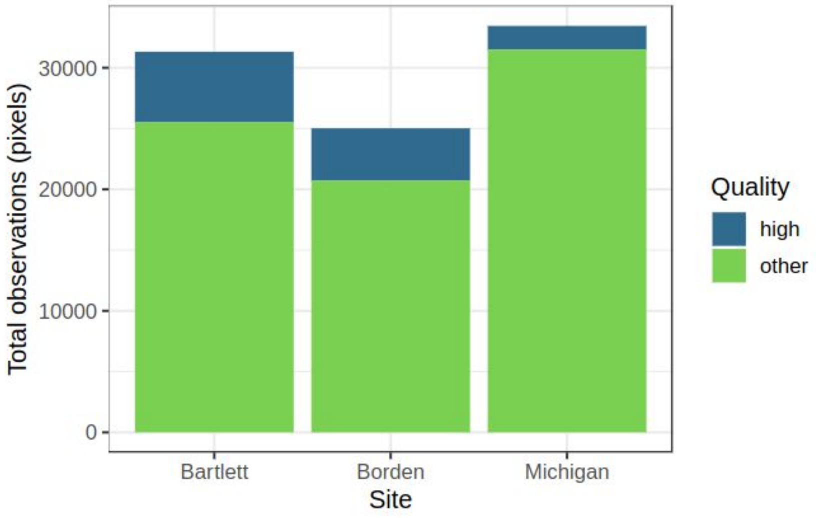
Canada

Ronny A. Hernández Mora.

ronnyhdez
http://ronnyale.com

More…

Results | MODIS quality observations



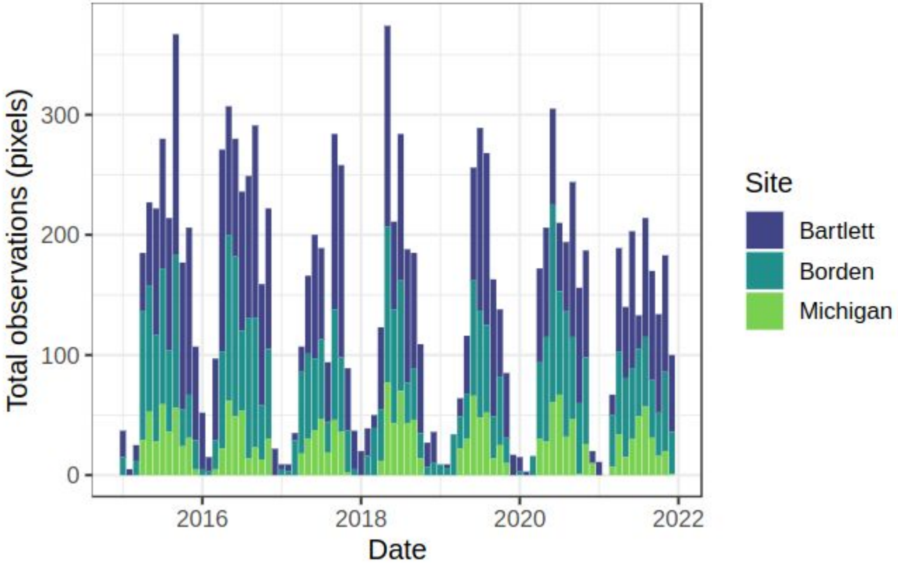


Table 2.2: ONEFlux Site characteristics overview

Variable	Site		
	Bartlett	Michigan	Borden
Mean annual temperature (°C)	6	5.5	7.4
Mean annual precipitation (mm)	1246	803	784
Elevation (m)	272	234	209
Dominant genera	Acer, Fagus, Betula	Populus, Quercus, Pinus	Acer, Pinus, Populus
Climate Koeppen[1]	Dfb	Dfb	Dfb

[1]**D** stands for the warm-summer continental or hemiboreal climate. **f** indicates that this climate has significant precipitation in all seasons. **b** indicates that the warmest month has an average temperature between 22°C and 28°C.

Methods | VIs

$$NDVI = \frac{NIR - Red}{NIR + Red} \tag{2.2}$$

$$NIRv = NIR \times \frac{NIR - Red}{NIR + Red} \tag{2.3}$$

$$EVI = 2.5 \times \frac{\text{NIR} - \text{Red}}{(\text{NIR} + 6 \times \text{Red} - 7.5 \times \text{Blue} + 1)} \tag{2.4}$$

$$CCI = \frac{Green - Red}{Green + Red} \tag{2.5}$$

Results | LM residuals distribution

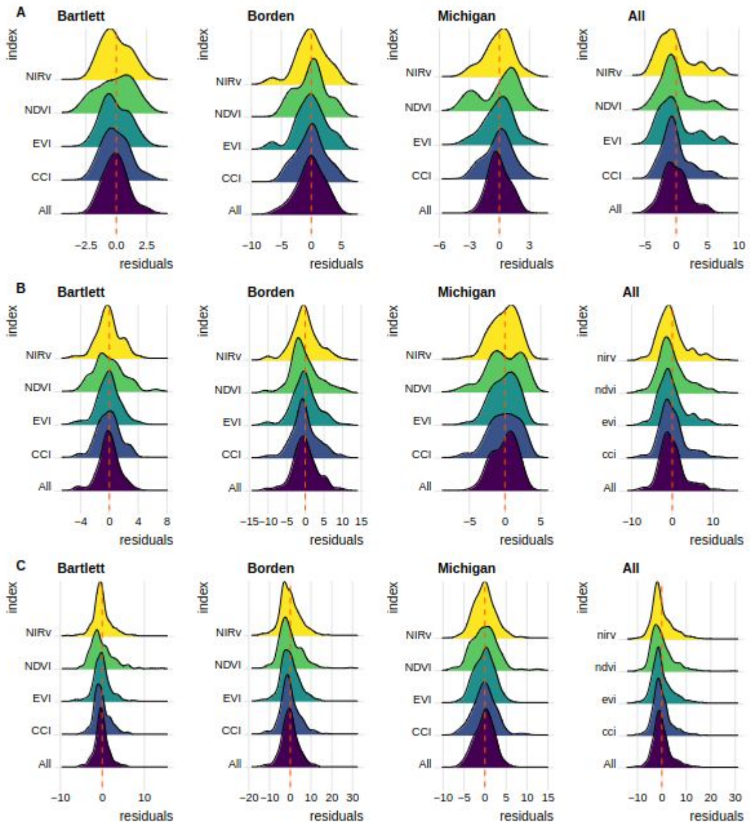


Figure 5.1: Residuals distributions for each of the LMs for GPP estimation using the vegetation indices on a monthly (a), weekly (b), and daily (c) basis.

Results | GAM residuals distribution

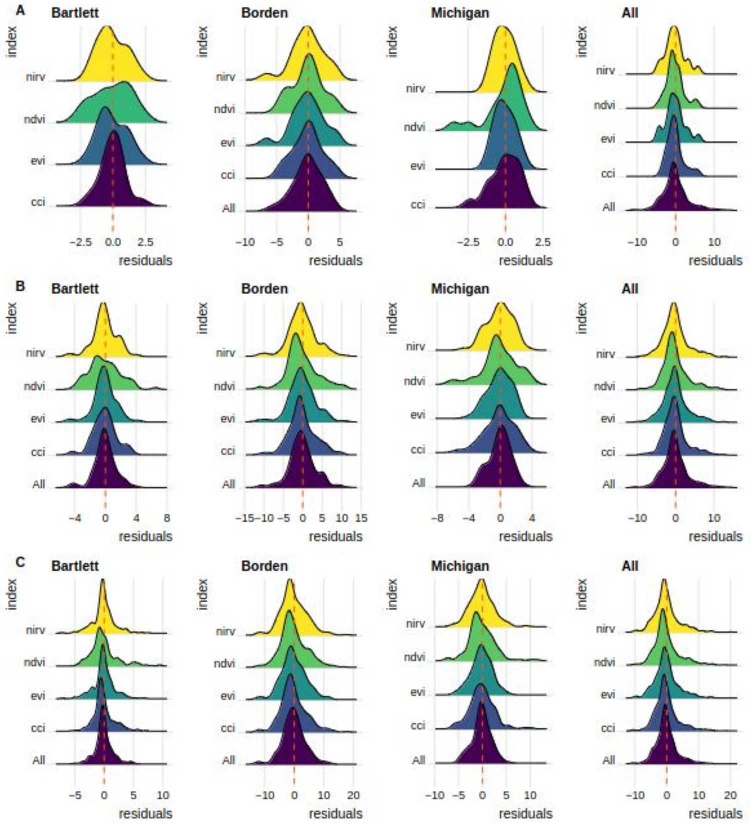
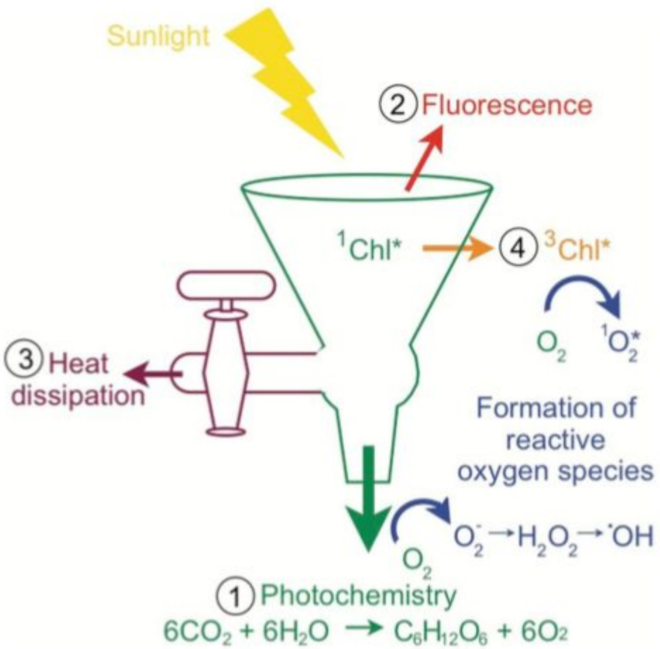


Figure 5.2: Residuals distributions for each of the GAMs for GPP estimation using the vegetation indices on a monthly (a), weekly (b), and daily (c) basis.

Fluorescencia, fisiología fotosintética de la hoja y estrés ambiental



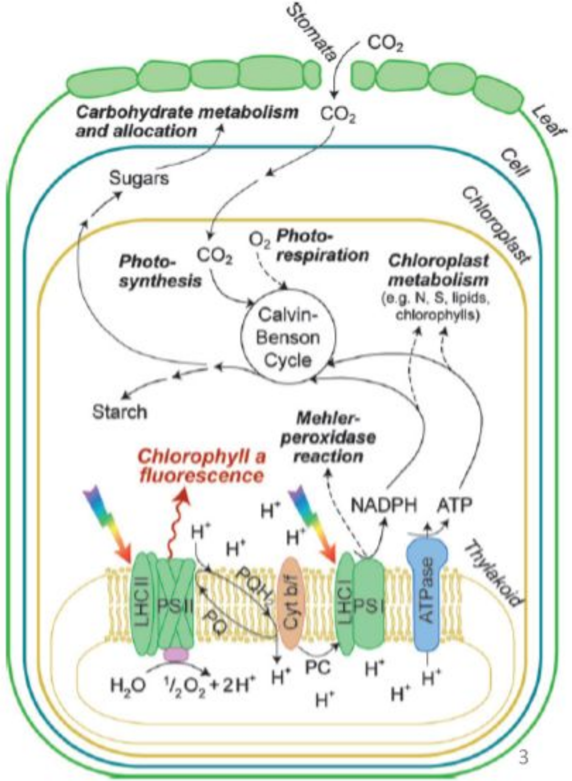


Table 1. Five hypothetical sets (cases) of 4 errors, and their corresponding totals, MAEs and RMSEs. Each e_i ($e_i = P_i - O_i$, $i = 1, 2, 3, 4$) is a hypothetical error value

Variable	Case 1	Case 2	Case 3	Case 4	Case 5
e_1	2	1	1	0	0
e_2	2	1	1	0	0
e_3	2	3	1	1	0
e_4	2	3	5	7	8
$\sum \lvert e_i \rvert$	8	8	8	8	8
MAE	2	2	2	2	2
$\sum \lvert e_i \rvert^2$	16	20	28	50	64
RMSE	2.0	2.2	2.6	3.5	4.0