Estimation of Terrestrial Global Gross Primary Production (GPP) with Satellite Data-Driven Models and Eddy Covariance Flux Data

JOANNA JOINER, YASUKO YOSHIDA, YAO ZHANG, GREGORY DUVEILLER, MARTIN JUNG, ALEXEI LYAPUSTIN, YUJIE WANG, COMPTON TUCKER

Joiner et al., Remote Sens. **2018**, *10*(9), 1346; https://doi.org/10.3390/rs10091346

Outline

- Motivation
- Basic Approach
- What didn't work so well
- What did work well
- What are the implications (global GPP)
- Summary

Outline

- **Motivation**
- Basic Approach
- What didn't work so well
- What did work well
- What are the implications (global GPP)
- Summary

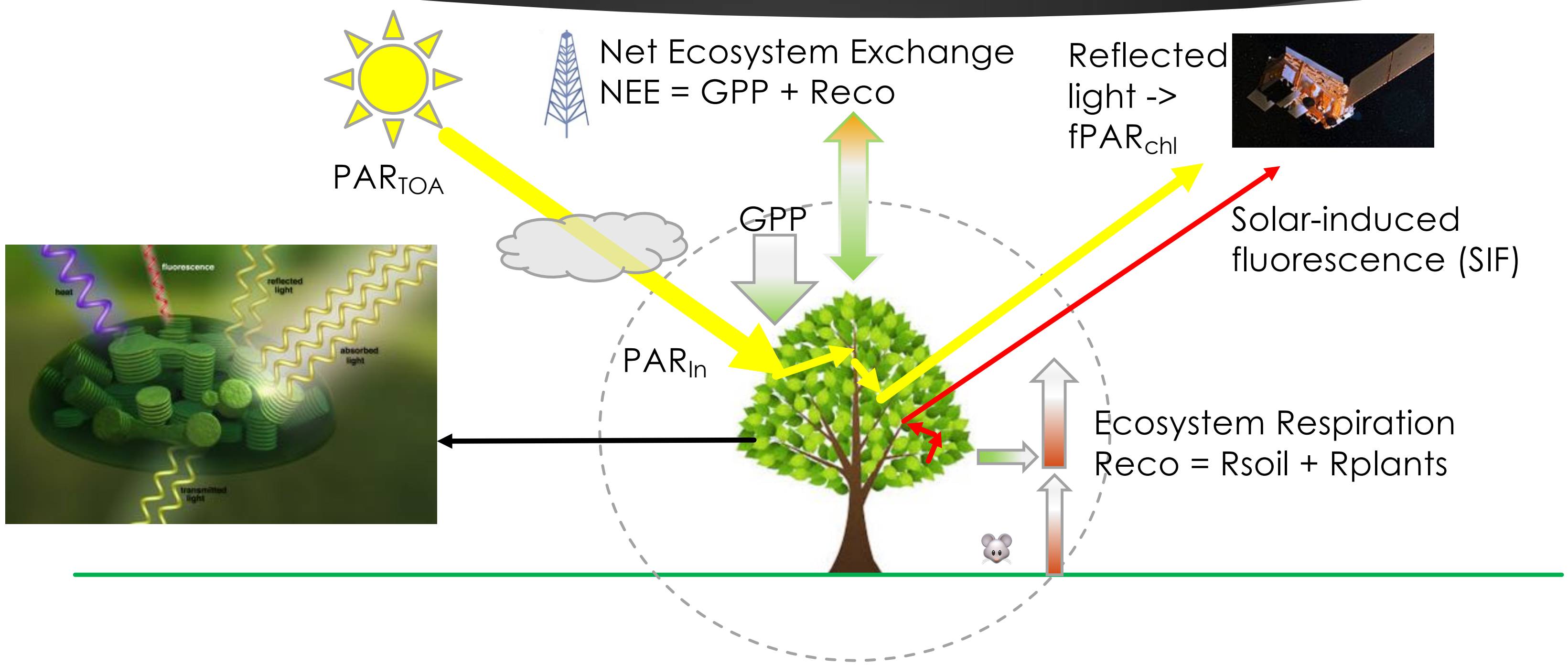
Motivation: What are current approaches to estimate global GPP? strengths/challenges?

- **Machine learning** typically trains to predict GPP from **eddy covariance data (100+ sites, gold standard)** with satellite and meteorological data inputs; works well but doesn't provide a physical basis for understanding; Which are the most important data?
- **Light-use Efficiency (LUE) models** use satellite and meteorological data:

  $GPP = PAR * fPAR_{chl} * LUE$,

  - $fPAR_{chl}$ = fraction of absorbed photosynthetically active radiation)
  - how good are the LUE parameterizations (e.g., do they create features that are (not) supported by eddy covariance observations)?
- **Terrestrial biology models (TBMs)** vary in their performance and need evaluation with data-driven approaches.
- **Assimilation and SIF-based estimates** (with optional downscaling) are newer approaches, but how good are they? need comparison with other approaches

Carbon fluxes in terrestrial ecosystems and relationship to satellite remote sensing



Motivation: Why do we need another GPP estimate? Questions to consider:

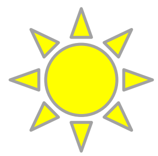
- How well can we do just with satellite reflectances? Angle (BRDF) adjusted?
- Use vegetation indices? and from what source, e. g., from composites?
- Deriving $fAPAR_{chl}$ (Q. Zhang et al.) with MODIS reflectances works well but computationally intensive to produce. Is there a simple data-driven approach (linear band combination)?
- What value can fluorescence (SIF) add?
- What are most important LUE drivers? Is there a straight-forward way to compute LUE?
- On what time scales can we accurately estimate GPP with satellite data? Monthly? Daily?
- What spatial scale is needed to evaluate or train with eddy covariance flux tower data?

Outline

- Motivation
- **Basic Approach (including input data sets)**
- What didn't work so well
- What did work well
- What are the implications (global GPP)
- Summary

Basic Light Use Efficiency (LUE) Approach

- LUE concept: GPP = PAR_{IN} * $fPAR_{chl}$ * LUE_p

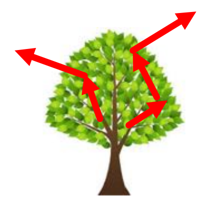


CERES or assimilation

MODIS or similar

Parameterization (meteorological, satellite data)

- Similar for SIF: SIF = PAR_f * $fPAR_{chl,f}$ * LUE_f * f_{esc}



- Many studies show GPP-SIF correlation

(weekly to monthly time scales, but scaling factor varies)

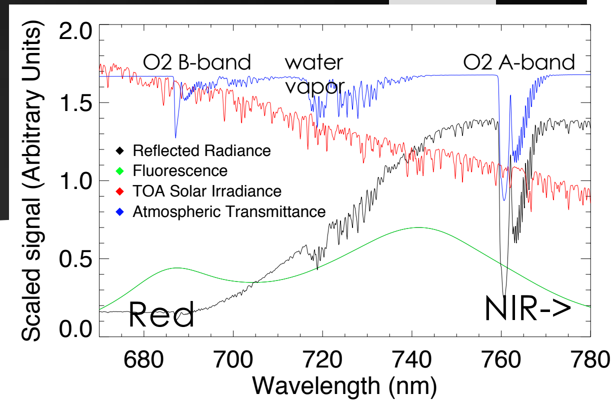
Considerations: SIF vs. Reflectances

- Large signal in reflectances (NIR – Red); Easy to measure (low spectral resolution bands), sensitive to clouds
- Small SIF signal requires high spectral resolution -> lower temporal and spatial resolution, but less sensitive to clouds
- Both sensitive to $fPAR_{chl}$, but SIF also sensitive to PAR_{IN}.
- SIF ~ linear with GPP (weekly to monthly), but scaling factor varies, complicated by f_{esc} and LUE_f
- SIF shows more sensitivity to high yield (C4) crops such as corn
- Can we utilize the assets of both?

Modified (simplified) LUE Approach

- 1) $GPP = fPAR_{chl} * PAR_{IN} * LUE$, where LUE is function of meteorological parameters.
- 2) $GPP =\sim fPAR_{chl} * PAR_{TOA}$ (* LUE'), where LUE' is optional & a function of reflectances
- LUE dependence on PAR is important (diurnal, cloud effects)!
- $PAR_{TOA} \alpha PAR_{IN} * LUE$ implicitly accounts for LUE variation with PAR to keep GPP nearly constant over a wide range of PAR (implied by earlier works for croplands by Peng and Gitelson)
- Majority of stress effects are reflected in $fPAR_{chl}$ (e.g., leaf curling, loss of chlorophyll) which occurs relatively quickly (within days to a few weeks of stress onset)
- LUE' can be (optionally) parameterized as a function of $fPAR_{chl}$ (or NDVI) to account for additional stress effects (moisture, temperature, radiation, nutrients) – accounts for very small amount of GPP variability

Vegetation indices as proxies for $fPAR_{chl}$: What are the issues?



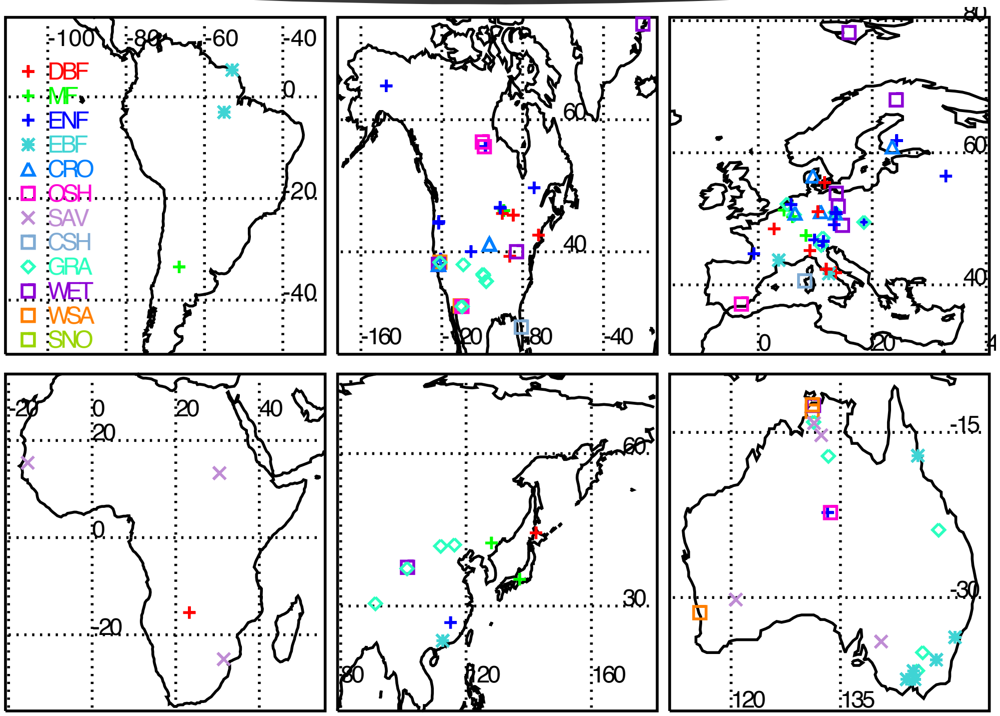
- NDVI= $(R_{NIR} - R_{red})/(R_{NIR} + R_{red})$; R: Reflectance
- NDVI /= 0 for background (soils, branches, etc.)
- Subtracting an constant offset from NDVI isn't the right thing to do (isn't necessary at high NDVI values). Instead, we subtract and phase out an offset value N_0 linearly with NDVI up to NDVI=0.7, i.e., NDVI'=NDVI – f(N_0).
- How to determine N_0? We tried using SIF (at lower spatial resolution) and GPP at flux sites, but ultimately settled on a constant value of 0.25.
- What about the NDVI "saturation problem" at high NDVI values?
- EVI is a similar 3 band index intended to improve upon NDVI (used in VPM)
- NIR_V=(NDVI – 0.08) * R_{NIR}
- How well does a simple linear combination of band reflectances work better?

Data sets used for GPP estimation

Input Dataset	Temporal Resolution	Spatial Resolution	Use
MCD43D reflectances *Schaaf et al.*	daily	$0.0083^\circ \times 0.0083^\circ$ ~1 km	Indices -> $fPAR_{CHL}$
GOME-2 SIF* (downscaled) *Duveiller & Cescatti, 2016*	monthly	$0.05^\circ \times 0.05^\circ$	GPP proxy (adjusted for daily PAR)
GOME-2 SIF *Joiner et al., 2013, 3016*	monthly	$0.5^\circ \times 0.5^\circ$	Identify high productivity areas
CERES or GEOS-DAS	any	$0.5^\circ \times 0.5^\circ$	PAR_{IN}
NIR_V x SW_{TOA}	any	any	GPP proxy
NDVI' x SW_{TOA}	any	any	GPP proxy
FLUXNET 2015 GPP	Daily, monthly	Site footprint ~1 km	Training, evaluation
VPM GPP (LUE model) *Y. Zhang et al., 2017*	monthly	$0.05^\circ \times 0.05^\circ$	Baseline evaluation
FLUXCOM GPP(ML) *Tramontana et al., 2016*	8-day	$0.083^\circ \times 0.083^\circ$	Baseline evaluation

FLUXNET 2015 Tier 1 eddy covariance sites (free and open use)

- Covers a wide range of ecosystems
- Sparse coverage over some continents and regions (e.g., tropical rain forests, high latitudes)



Outline

- Motivation
- Basic Approach
- **What didn't work so well**
- What did work well
- What are the implications (global GPP)
- Summary

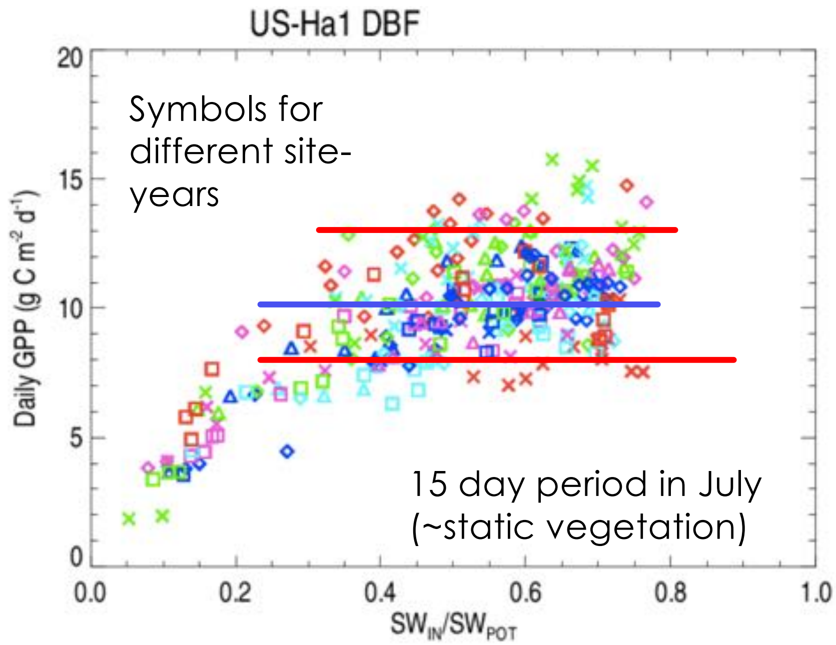
Things we tried that didn't work so well

- Using MOD09 reflectances (BRDF-adjusted data sets MCD43 NBAR and MAIAC worked better)
- Using MOD15 fPAR (and MOD17 GPP based on it)
- Using MOD13 NDVI (max value) composited data sets
- Using CERES/MERRA-2 SW as PAR_{IN} with constant or simple LUE (PAR_{TOA} works better)
- Using NIR_V without any account of PAR
- Accounting for SIF escape (SZA dependence) using simple MODIS reflectance-based approach

Why does using PAR_{TOA} work better?

- Plants are adjusting their efficiency such that GPP stays relatively constant over a wide range of PAR where most of the days fall

(Higher LUE at lower PAR; lower PAR corresponds to cloudy days with more diffuse light)

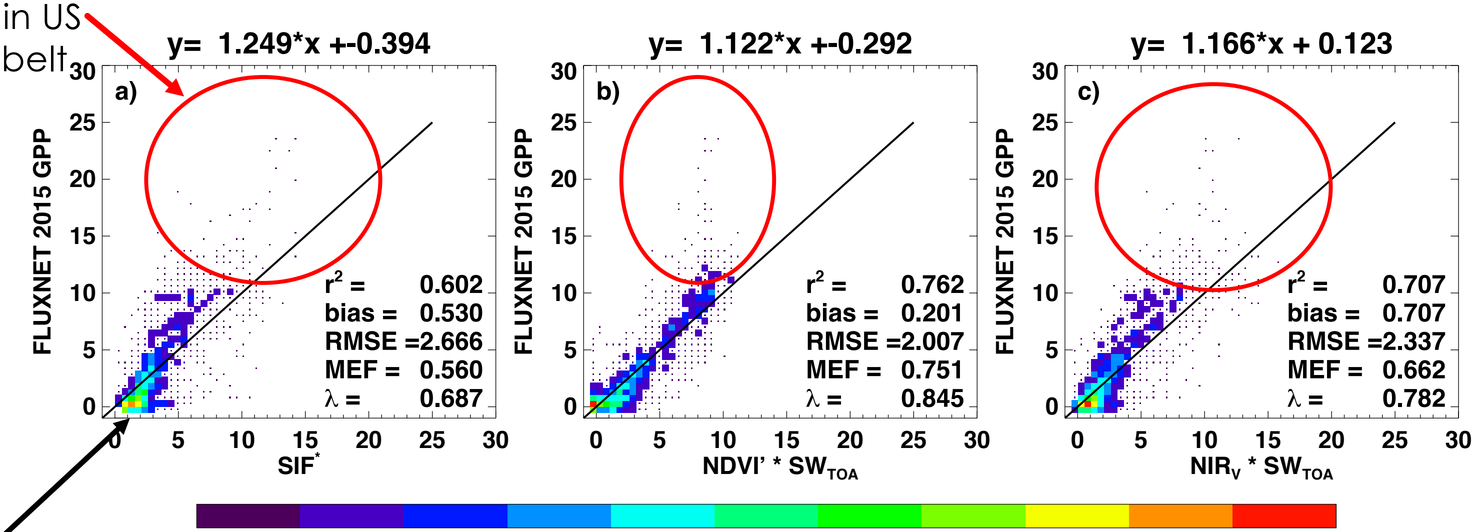


Outline

- Motivation
- Basic Approach
- What didn't work so well
- **What did work well**
- What are the implications (global GPP)
- Summary

Monthly GPP, 5 km: SIF* versus NDVI- and NIR_V-based (evaluation vs independent sites)

Mainly sites in US cornbelt

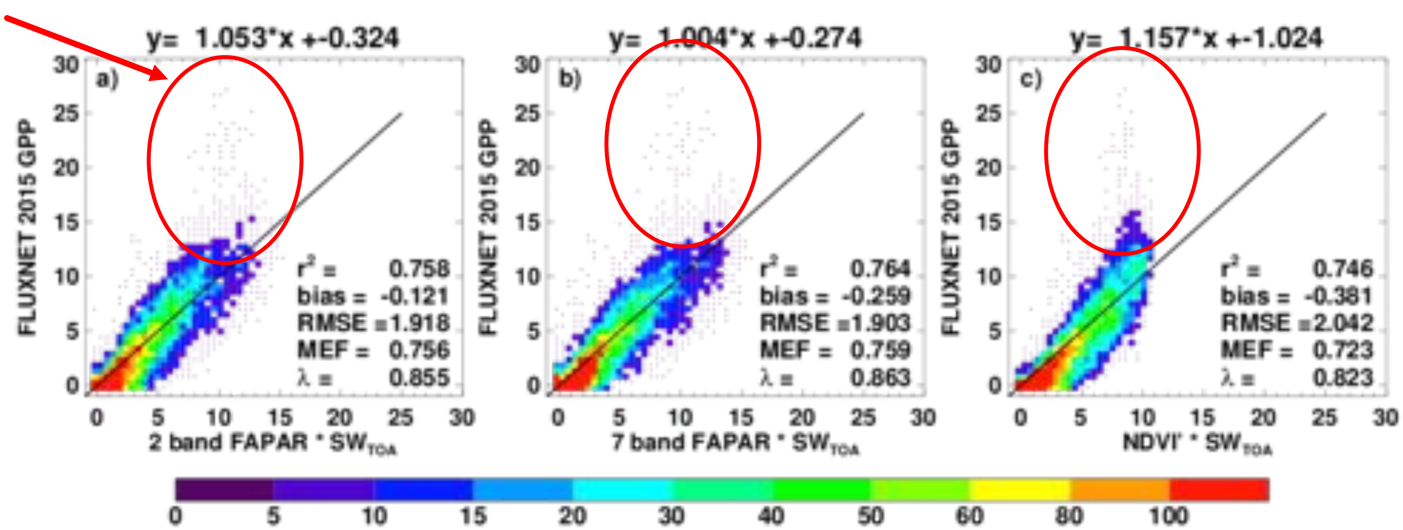


Zero-offset problem in early version of GOME-2 SIF (now corrected)

Number of points

8-day GPP, ~1 km, linear combination of bands versus NDVI-based

Mainly sites in US cornbelt



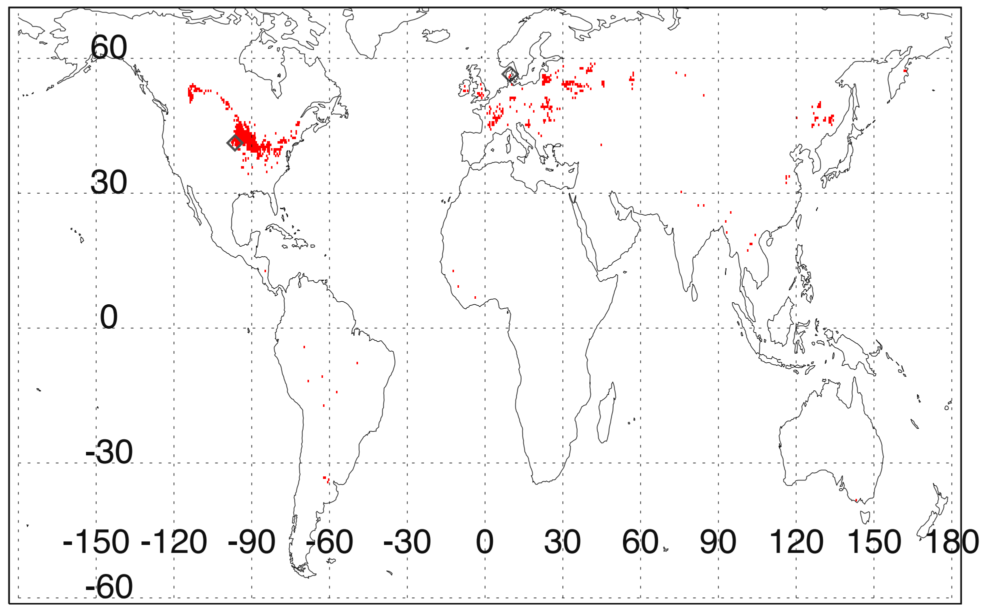
Number of points

Locations with high (normalized) SIF to NDVI ratios

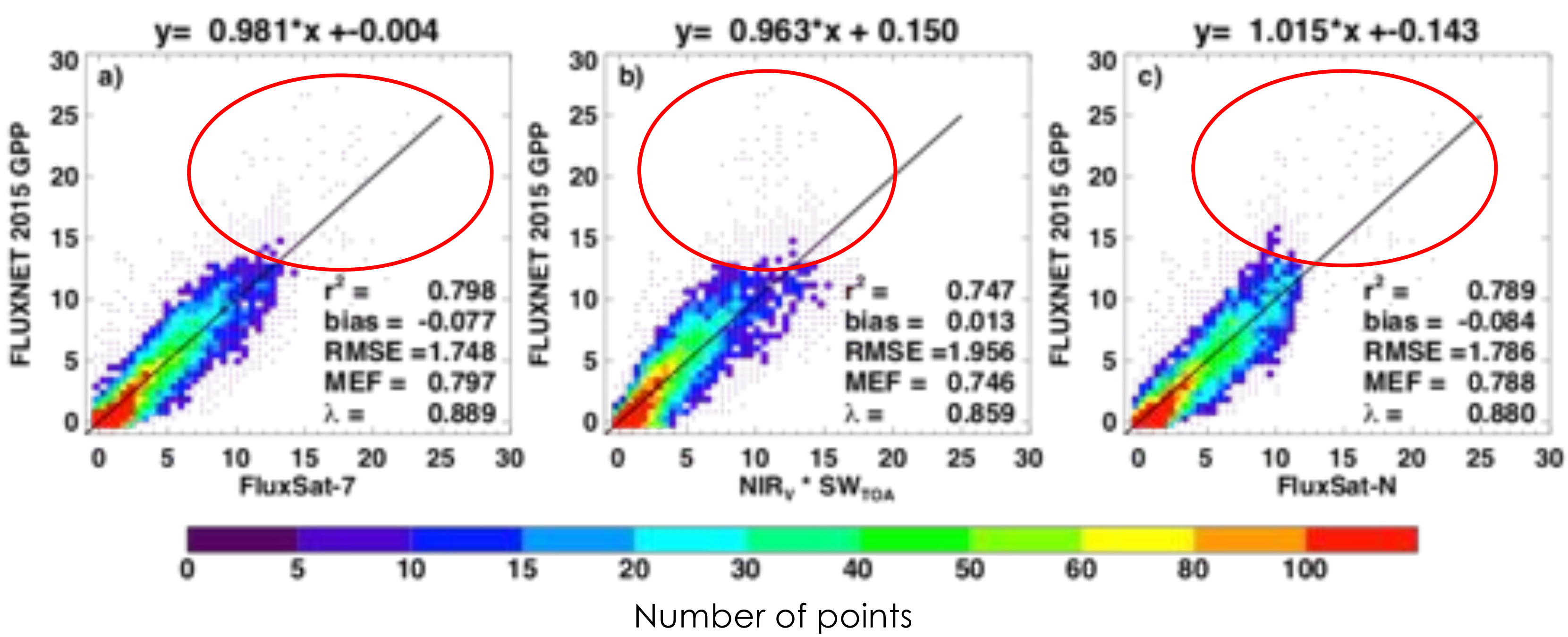
- separate fits for
  1. these points
  2. all others

- Additional parameterization of LUE to account for photoperiod and stress effect on LUE
  - Tried several approaches including use of PAR and NDVI
  - Best results with polynomial in NDVI

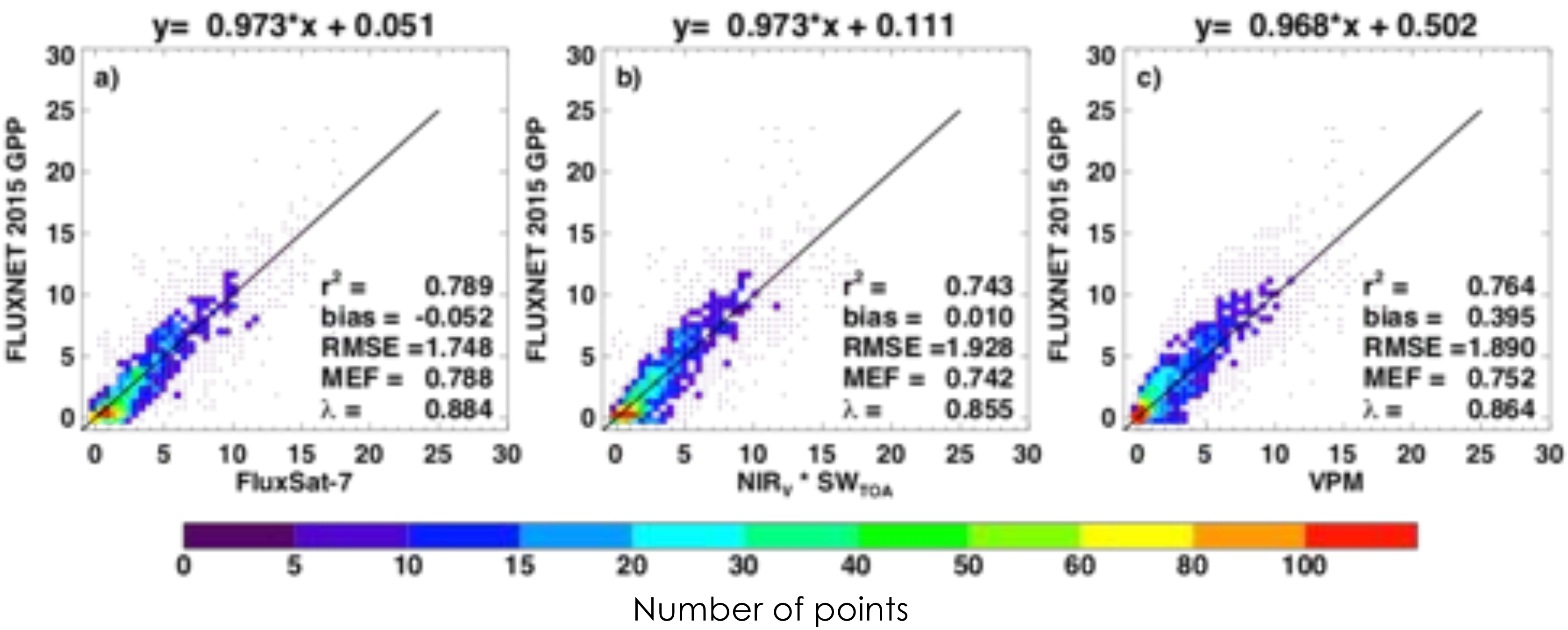
Dual fit, variable LUE - **"FluxSat" ("SatFlux")**



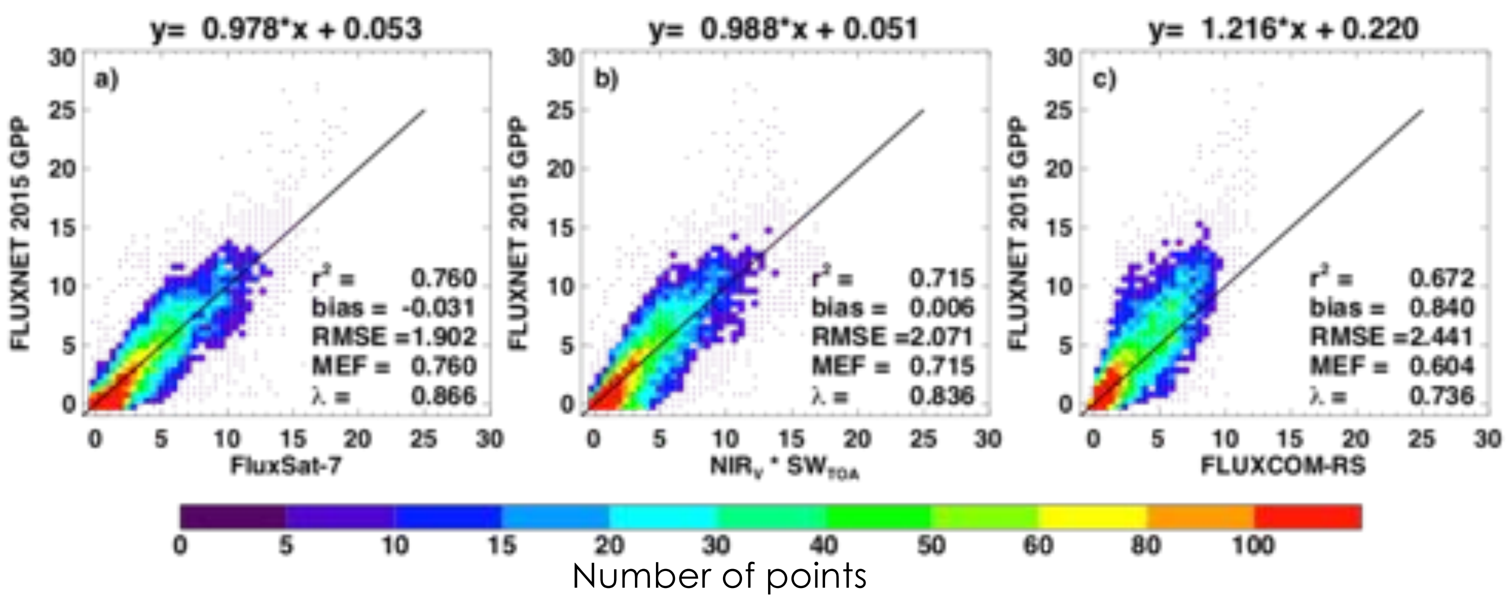
8-day 1 km GPP, dual fit (FluxSat)



Monthly, 5 km GPP comparison with VPM

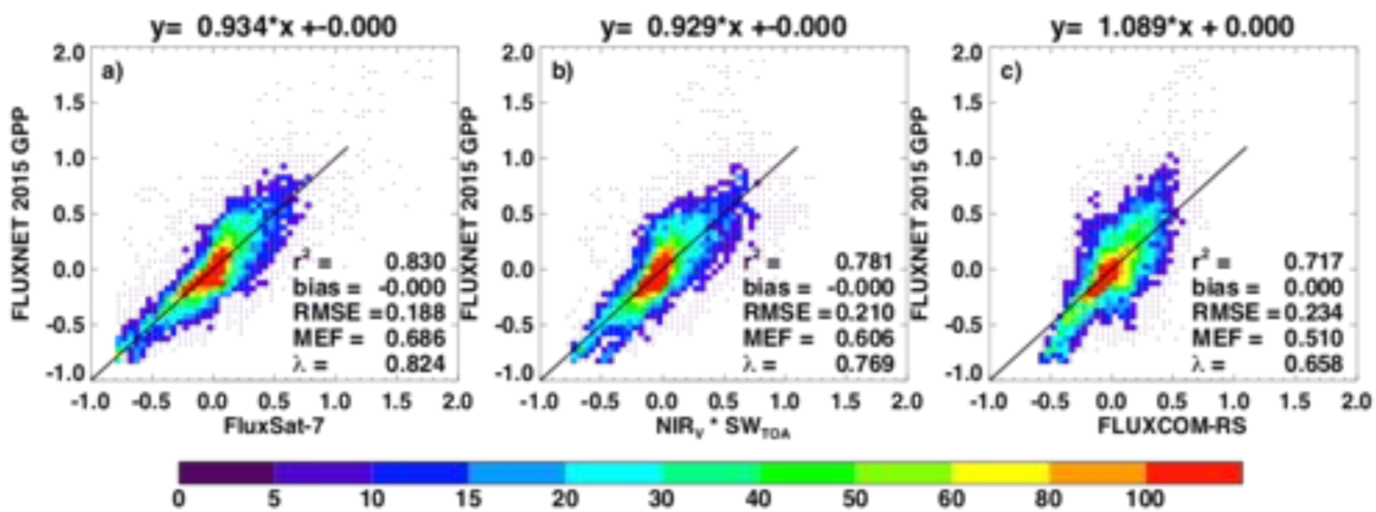


8-day, 8 km GPP comparison with FLUXCOM-RS



Training data set matters?! FLUXCOM used an older FluxNet data set.

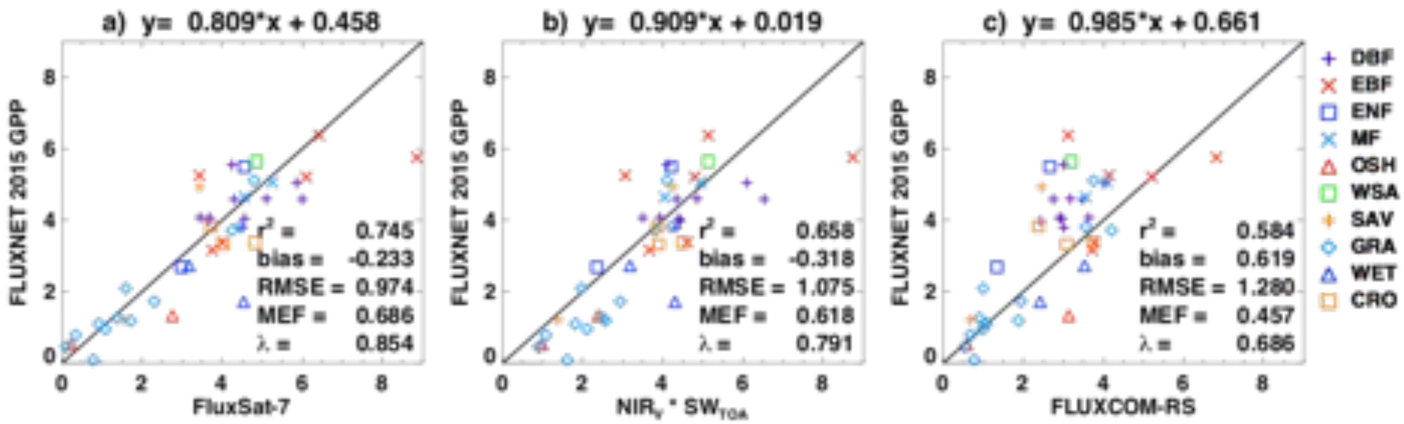
How well do the data sets capture interannual variations (anomalies)?



Number of points

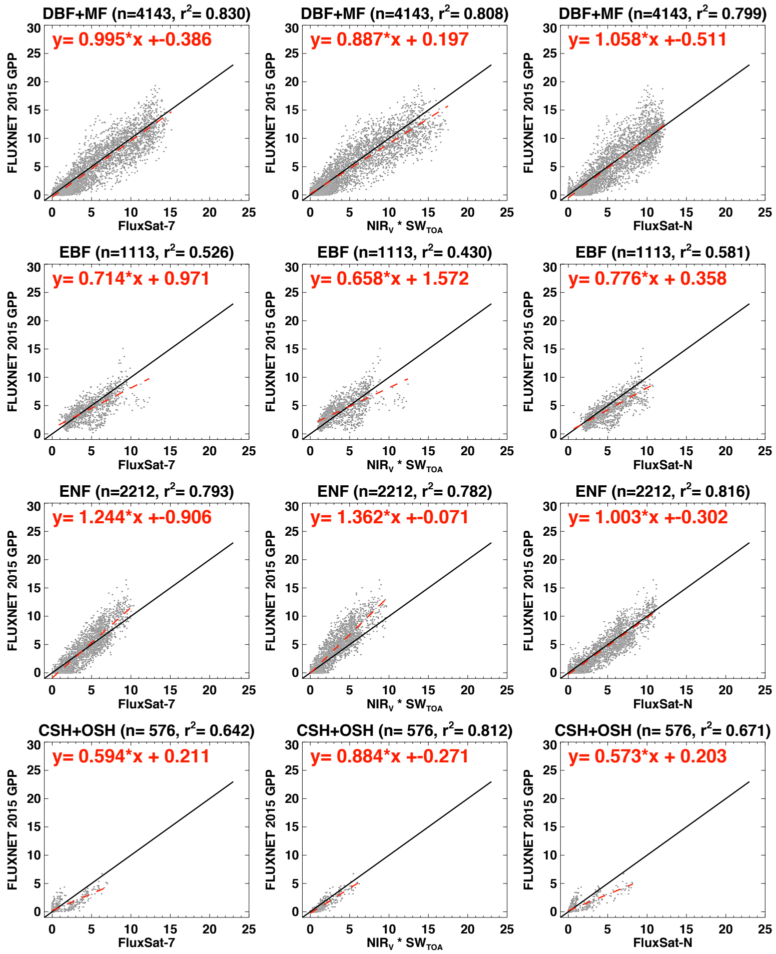
Anomalies normalized by GPP seasonal range (fractional, unitless)

How well do data sets capture site-to-site (spatial) variability?

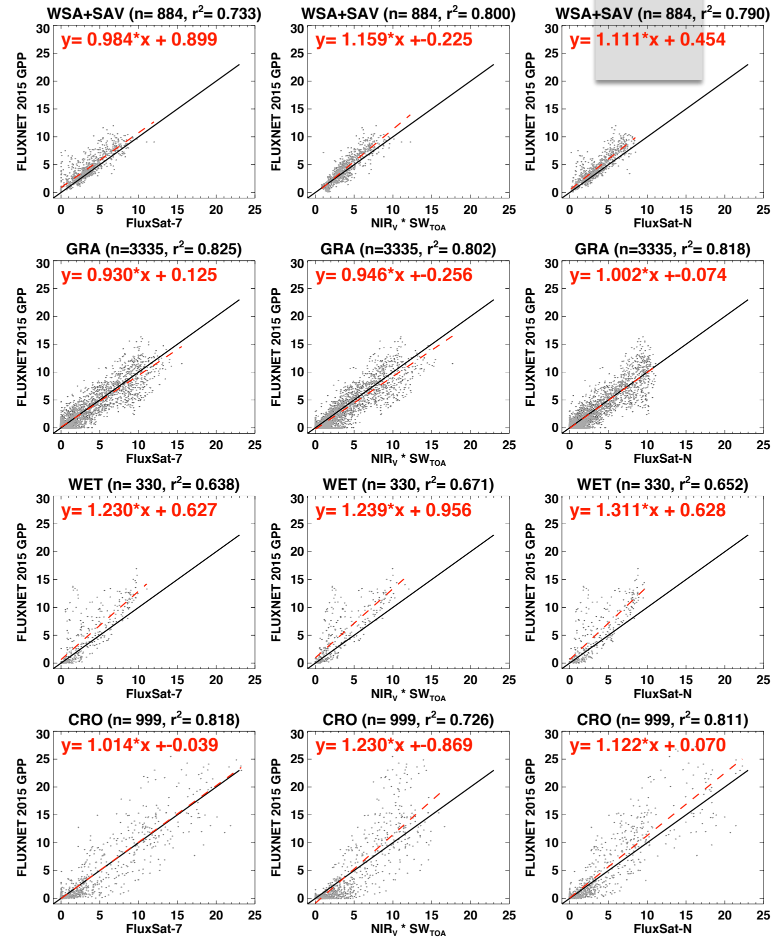


Each point is an annual mean for a single site

Black line: 1:1   Red line: fit



Each row is a different vegetation type

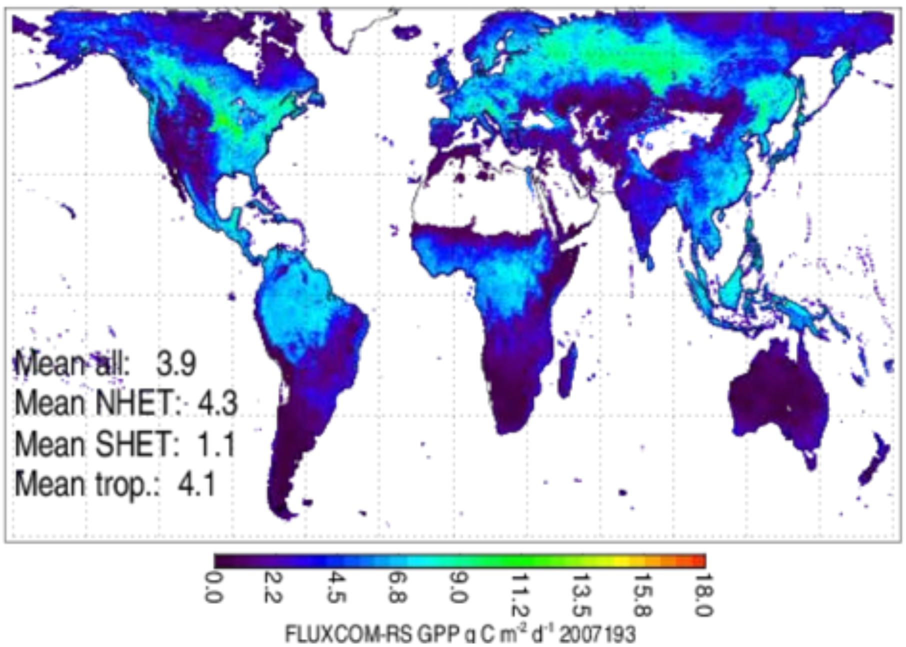


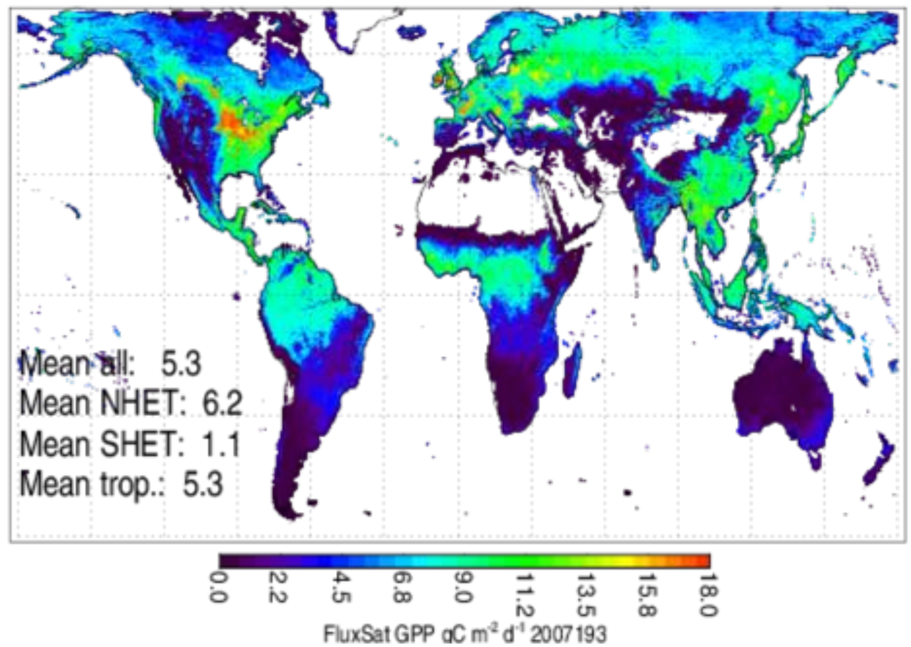
Outline

- Motivation
- Basic Approach
- What didn't work so well
- What did work well
- **What are the implications (global GPP)**
- Summary

July 2007 8-day average GPP

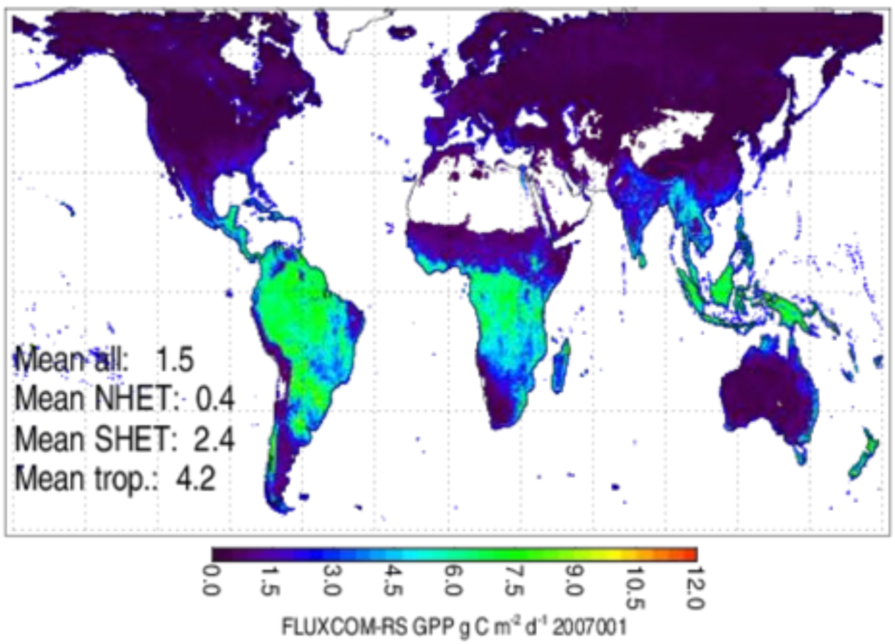
July 2007 FLUXCOM-RS **FluxSat-7**

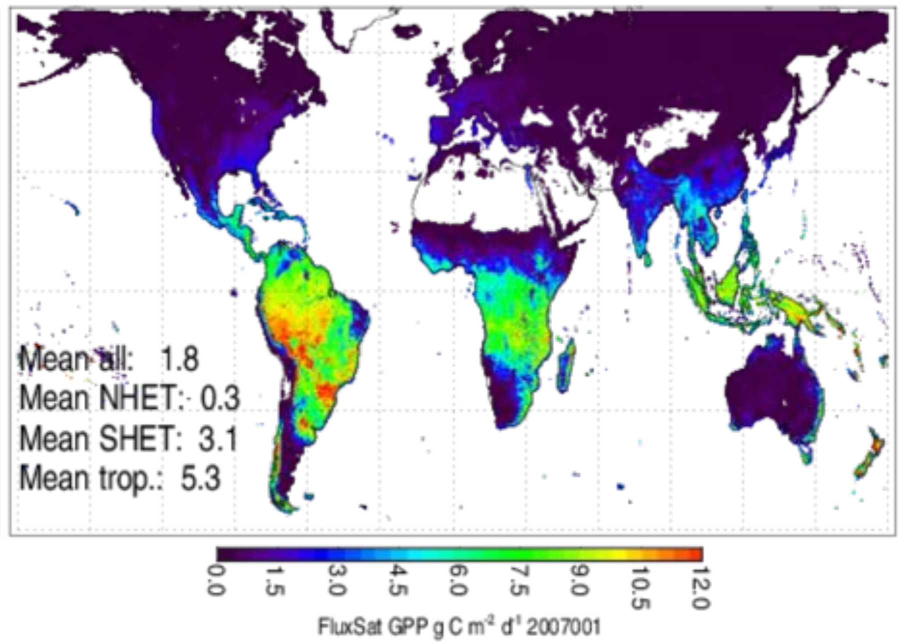




GPP 8 day average January 2007

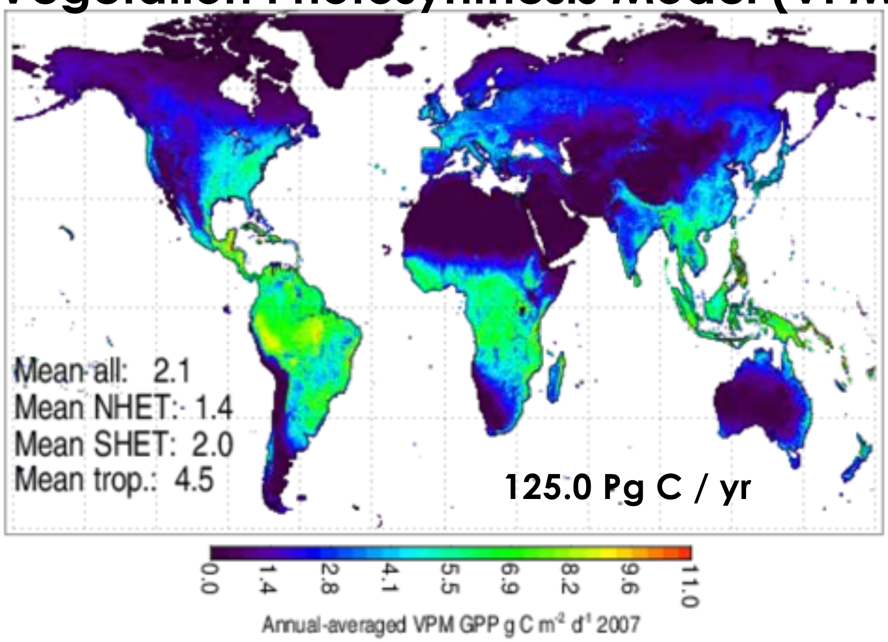
January 2007 **FLUXCOM-RS** **FluxSat-7**



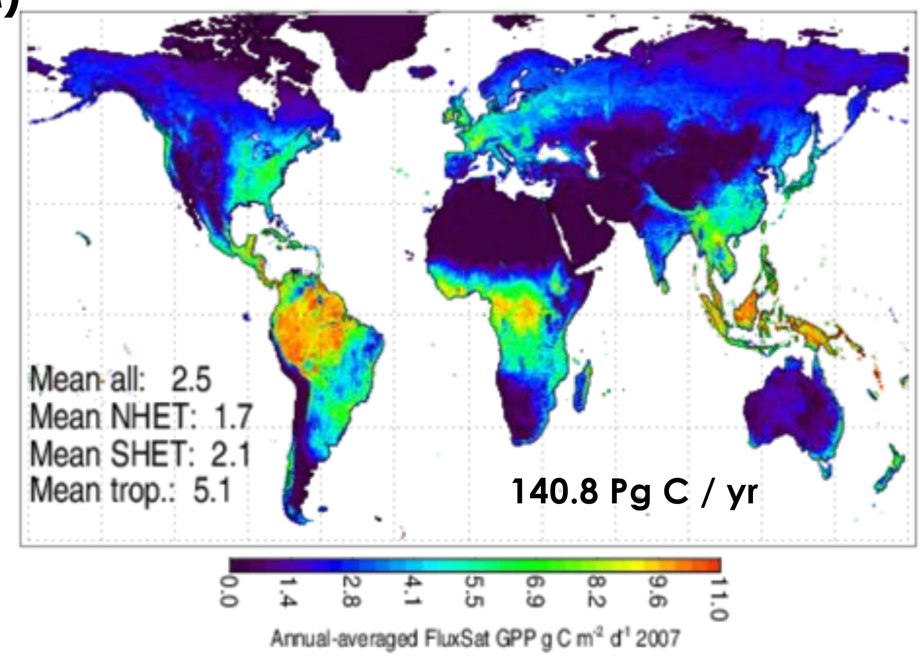


2007 Annual average GPP

Vegetation Photosynthesis Model (VPM)

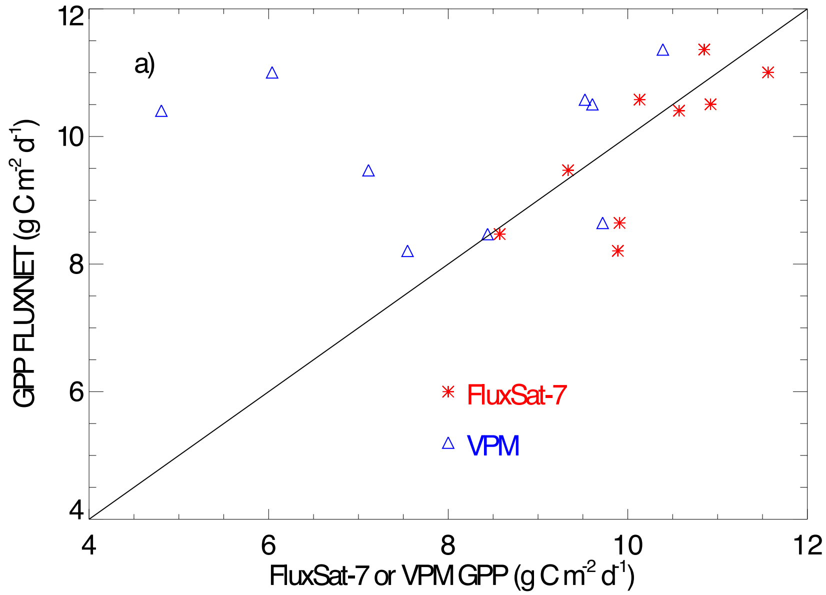


FluxSat-7

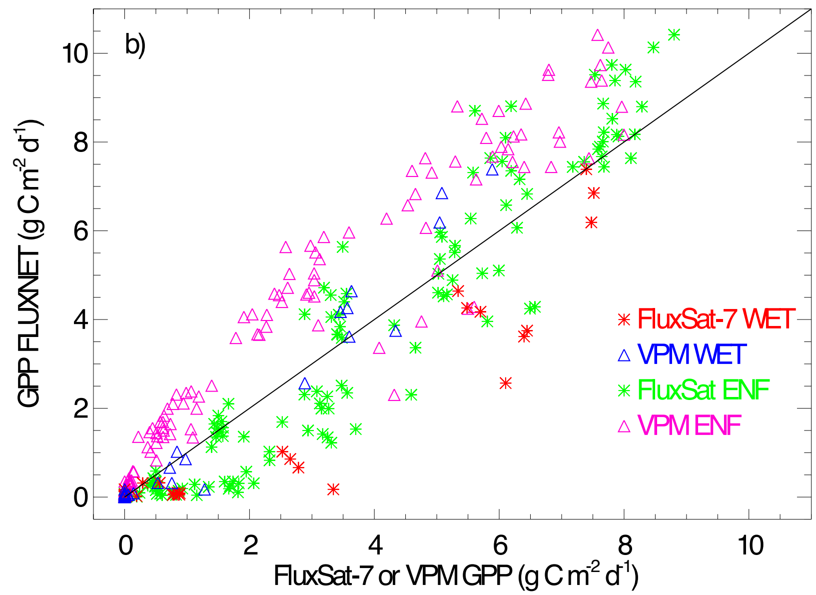


Are high FluxSat values in tropics and high latitudes supported by FLUXNET?

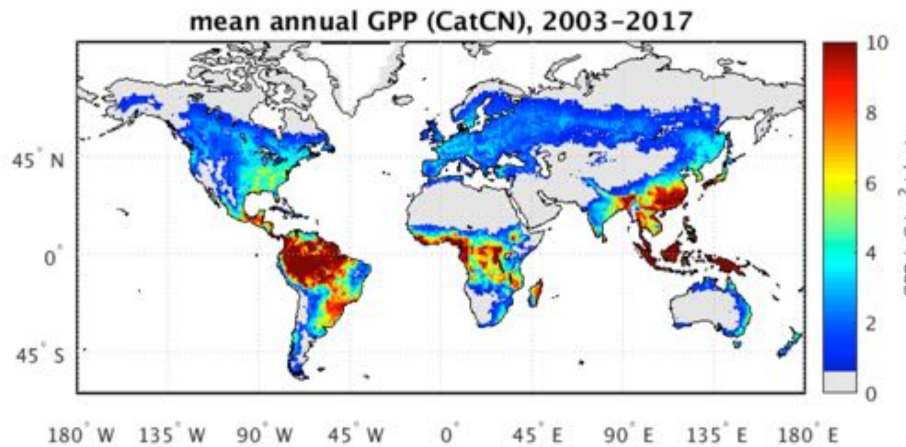
Tropical site (withheld from training)

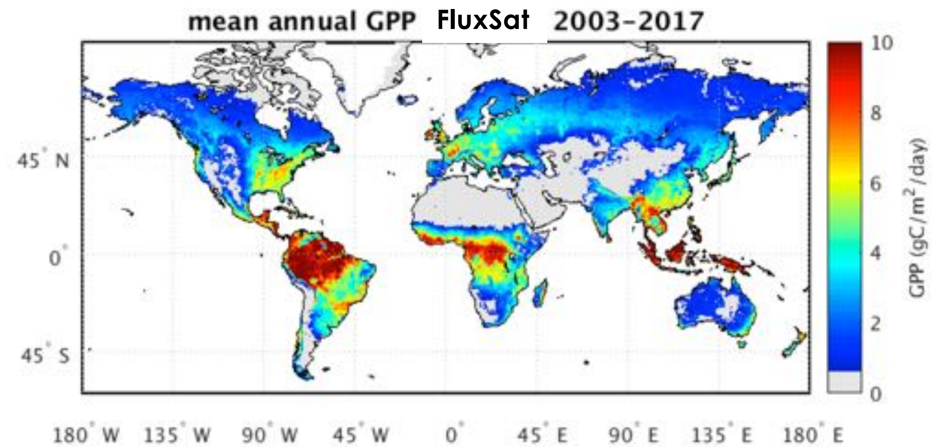


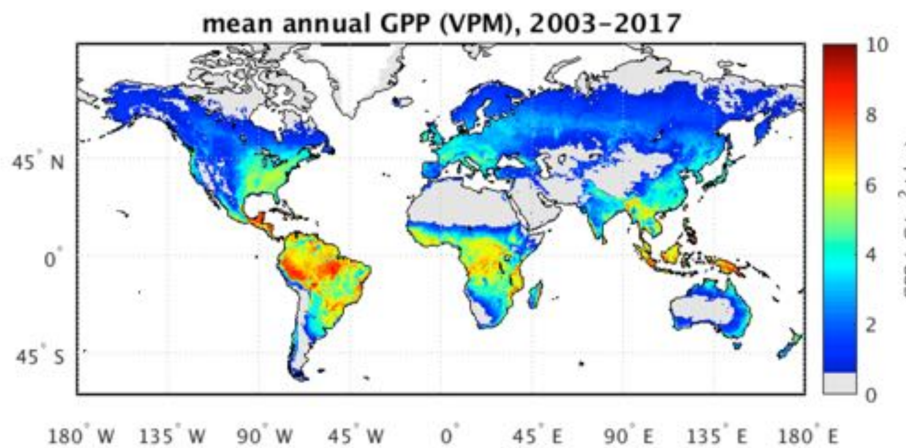
High latitude sites withheld from training



Annual mean GPP, model comparison (2003-2017)







FluxSat (SatFlux, version 1) is publicly available from the AVDC website:
https://avdc.gsfc.nasa.gov

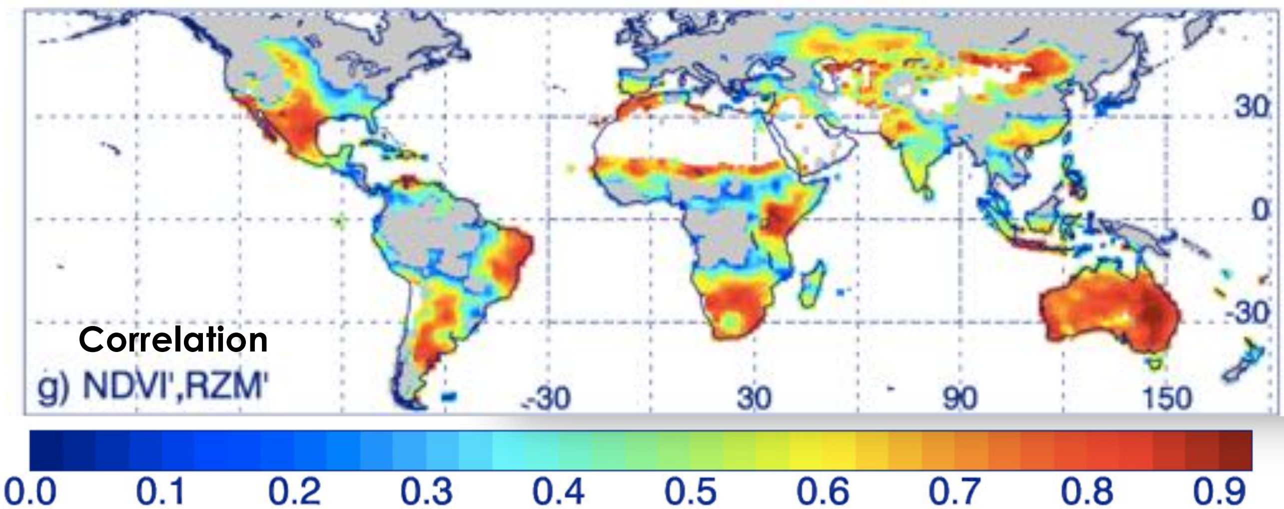
Go to Data and Archive menus

Courtesy Eunjee Lee, Fan-Wei Zeng, Randy Koster GMAO

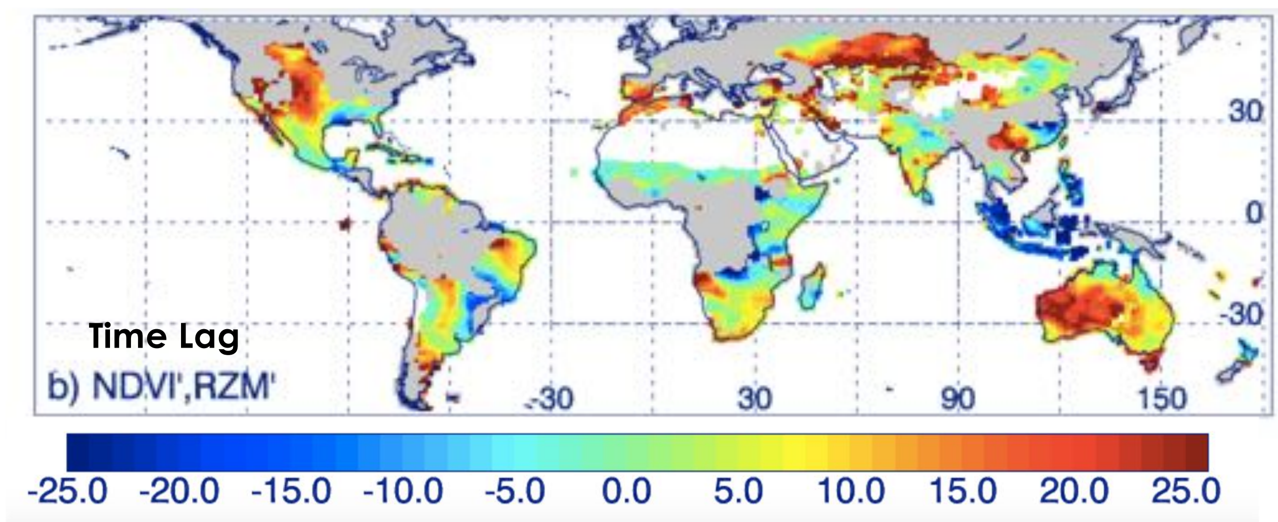
Summary

- We developed a simplified LUE approach to estimate global GPP with satellite data using assets of MODIS reflectanaces and SIF (FluxSat) trained using latest FLUXNET 2015 data set
- FluxSat performs as well or better than other more complex formulations (as compared with independent FLUXNET data)
- FluxSat estimates 2007 global annual GPP at **140.8 Pg C / yr -** generally higher than other satellite-based estimates but comparable to many TBM estimates.
- Still investigating details of multi-year data set and expect improvements in the future.
- Jury still out as to whether or not satellite driven SIF-based estimates can outperform reflectance-based GPP estimates.
- Need more flux towers in under-observed regions such as tropical rain forests

Backups

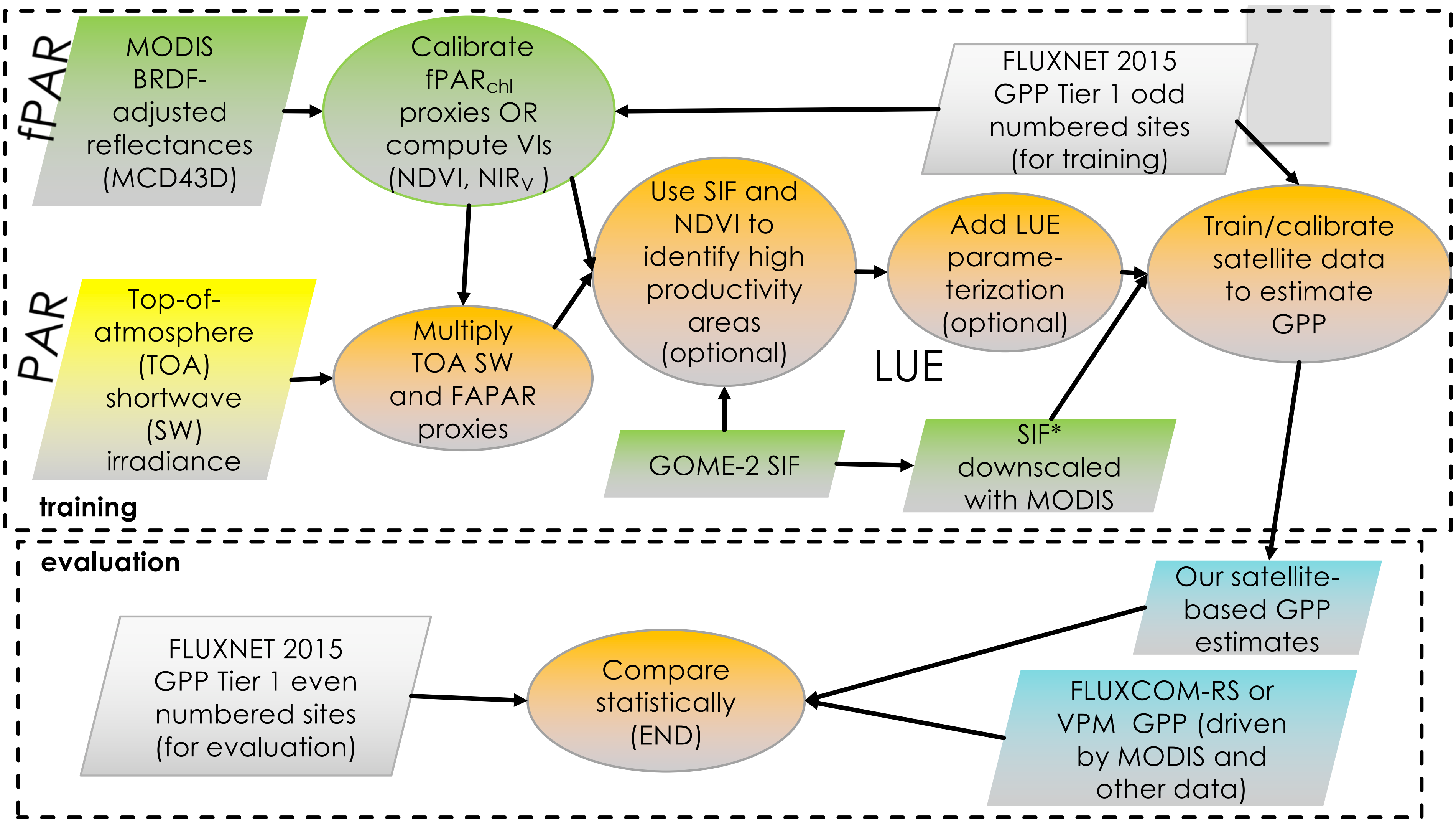


Correlation of NDVI and root-zone soil moisture (RZM) weekly anomalies (indicated by '). Gray areas are where correlations are not statistically significant. Highest correlations in semi-arid regions.



Time lags in days of NDVI with respect to root-zone soil moisture (RZM) weekly anomalies. Positive numbers are where NDVI' lags RZM'. Typically lags are days to a few weeks.

From Joiner et al., Rem. Sens. Environ., 2018



fPAR
MODIS BRDF-adjusted reflectances (MCD43D)
Calibrate $fPAR_{chl}$ proxies OR compute VIs (NDVI, NIR_V)
FLUXNET 2015 GPP Tier 1 odd numbered sites (for training)
PAR
Top-of-atmosphere (TOA) shortwave (SW) irradiance
Multiply TOA SW and FAPAR proxies
Use SIF and NDVI to identify high productivity areas (optional)
Add LUE parame-terization (optional)
Train/calibrate satellite data to estimate GPP
LUE
GOME-2 SIF
SIF* downscaled with MODIS
training
evaluation
Our satellite-based GPP estimates
FLUXNET 2015 GPP Tier 1 even numbered sites (for evaluation)
Compare statistically (END)
FLUXCOM-RS or VPM GPP (driven by MODIS and other data)