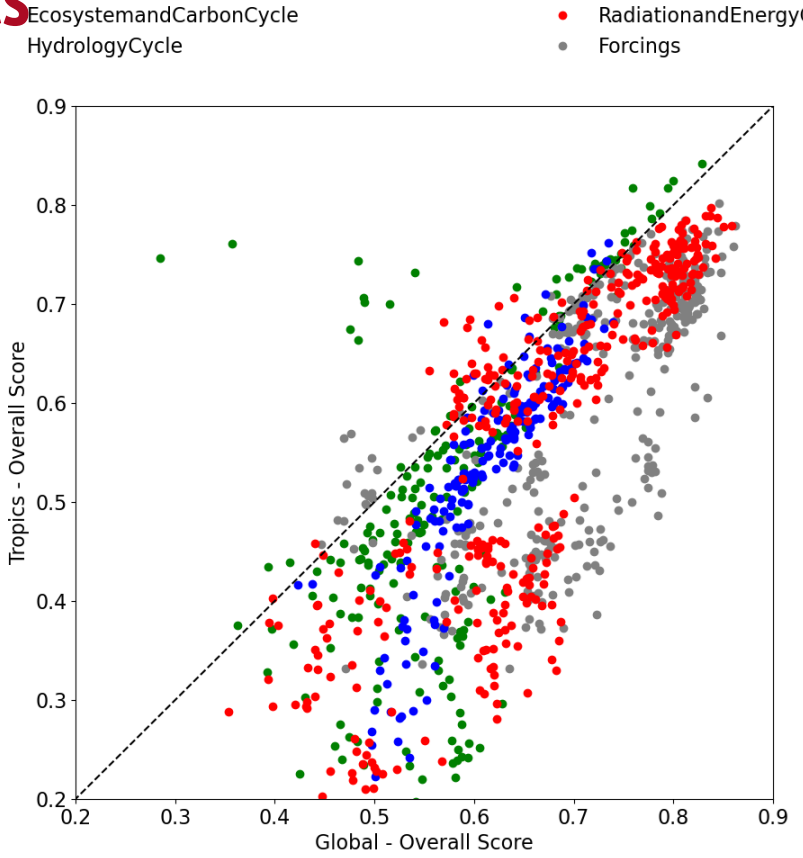
# Methodological Developments in the International Land Model Benchmarking (ILAMB) Effort

*Nathan Collier, Forrest Hoffman, Dave Lawrence*

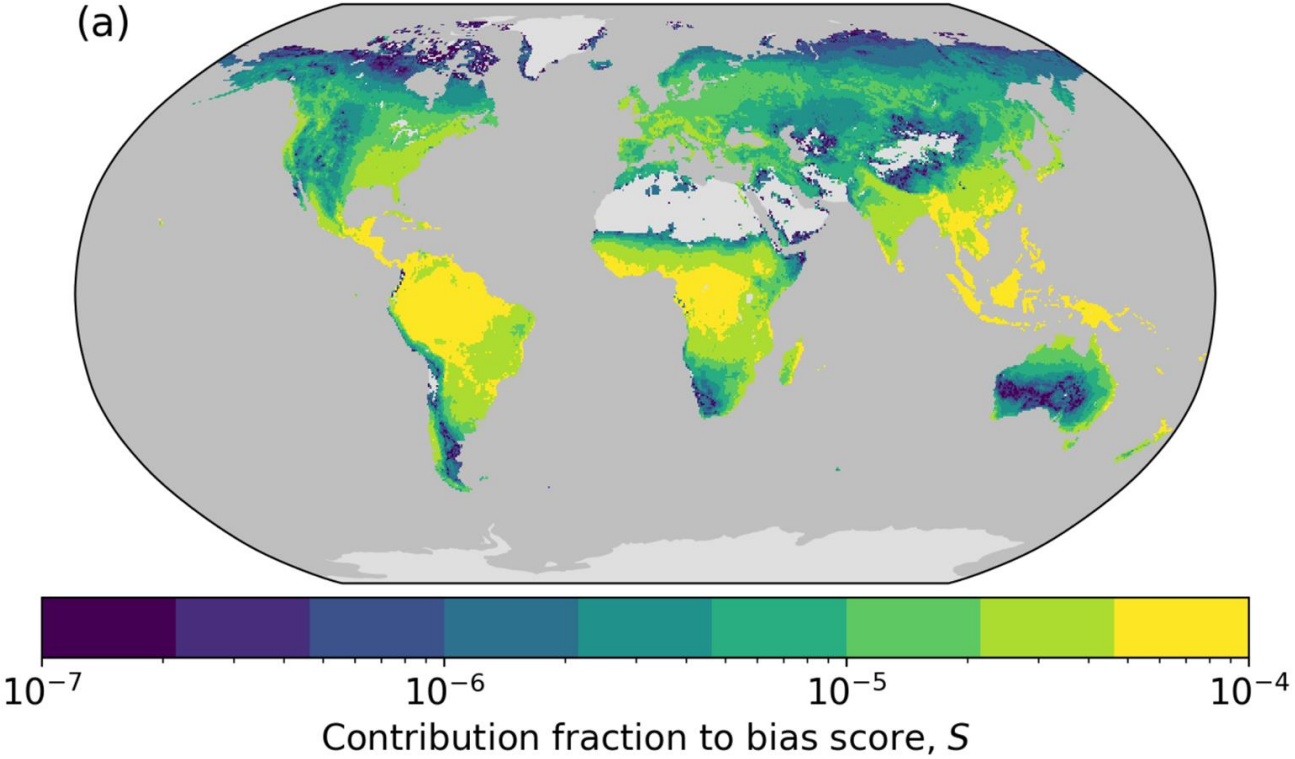
RUBISCO

# Heavily Influenced by Tropics

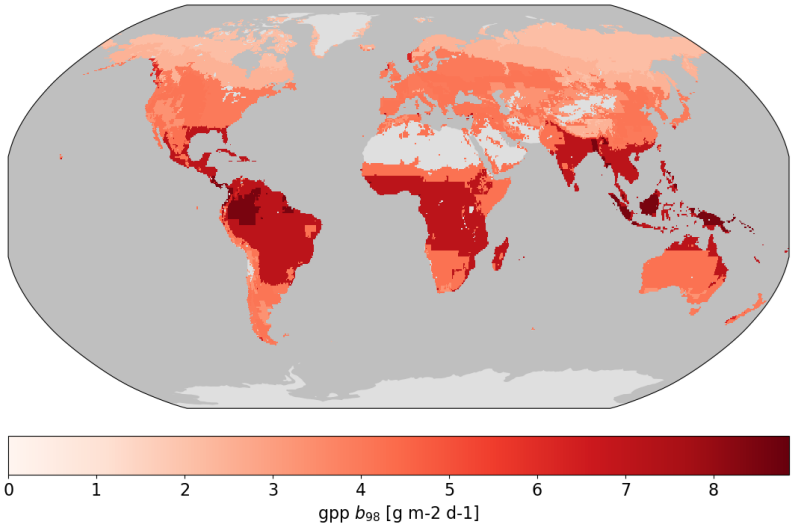
- We have observed that our current scoring methodology favors performance in the tropics.
- Plot shows that for many datasets, the tropics score correlates strongly to the global score
- This is due to our choice of normalizing errors by the variability of the reference data and the use of *mass weighting*



# Bias score influence map: gpp | FLUXCOM



# A Change in How We Normalize Errors

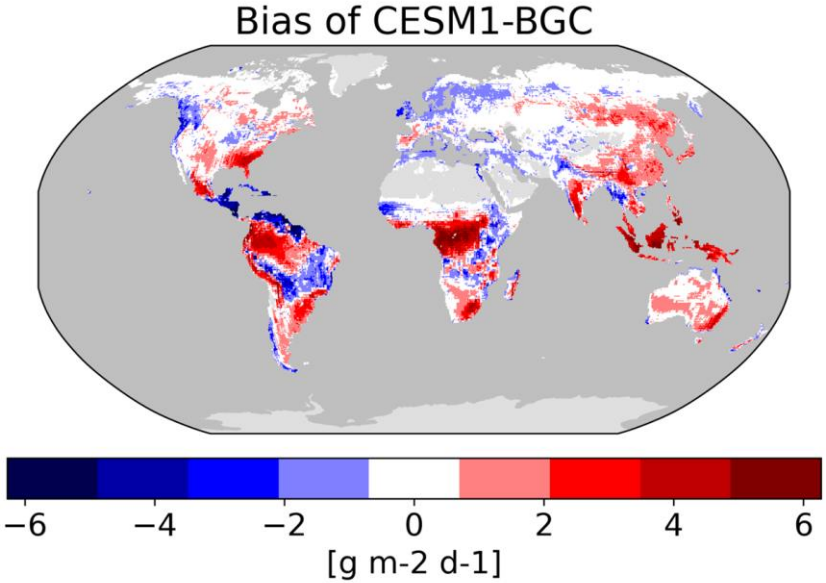


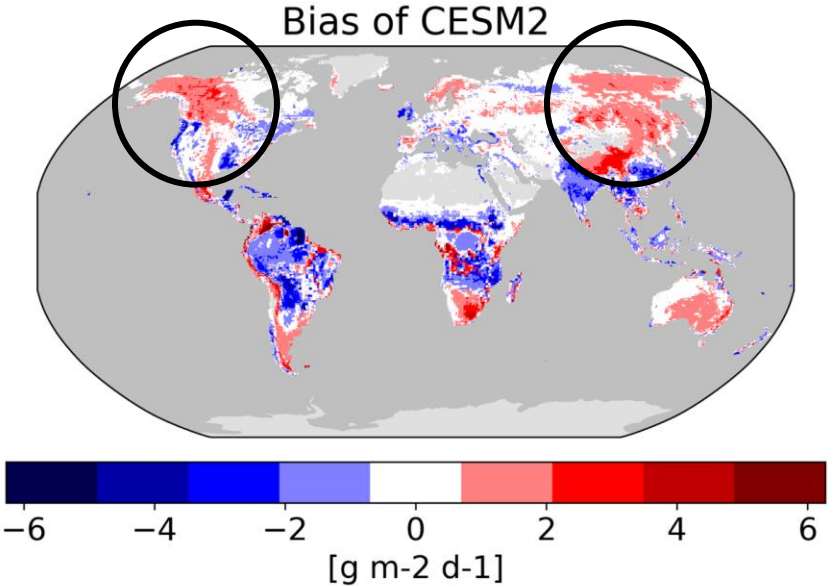

The gpp 98th error quantile within Whittaker biomes across CMIP5v6 models.

- The goal is to make errors from different areas of the globe comparable.
- Select a set of regions which represent biomes in which errors can be treated as commensurate in order of magnitude.
- Inside each region, for each variable, and across a selection of models, compute the 98th quantile of |bias(x)| with respect to all datasets for that variable.

# CESM2 | gpp | FLUXCOM

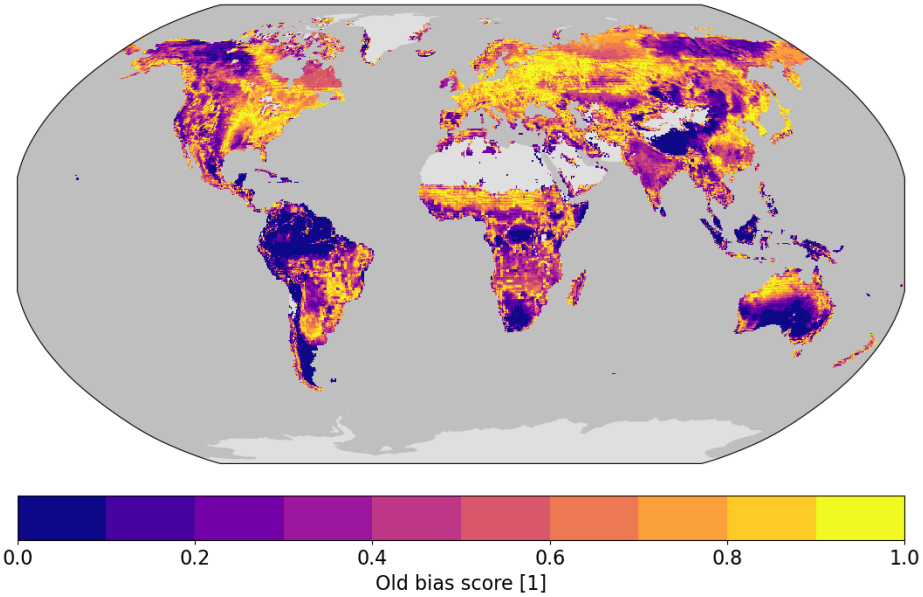
Notice larger bias in high latitudes, anomalous among CMIP models

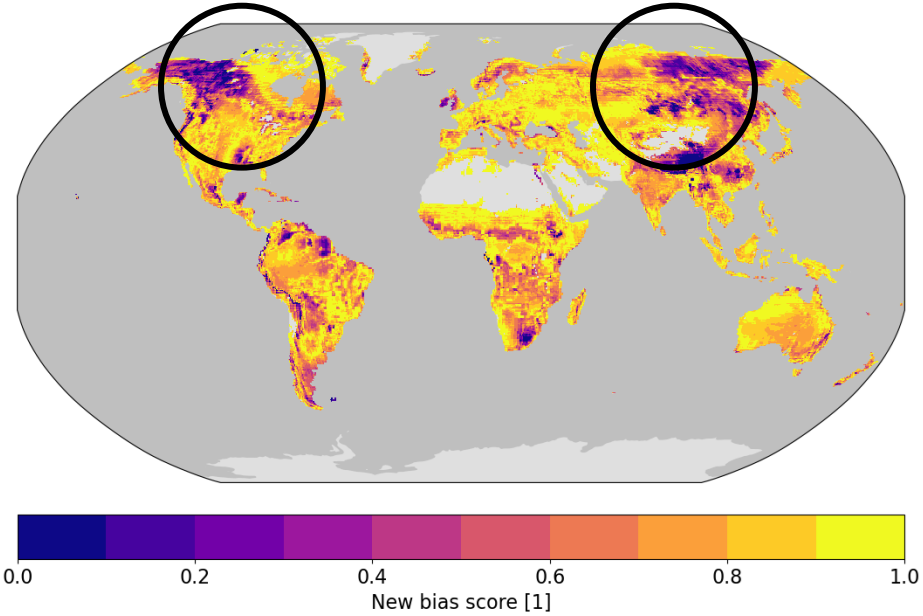


# CESM2 | gpp | FLUXCOM

With the new methodology, these areas light up clearly



# Other Ways to Use ILAMB Data

```
In [1]: import intake
   ...: cat = intake.open_catalog("https://raw.githubusercontent.com/nocollier/intake-ilamb/main/ilamb.yaml")

In [2]: cat['
```

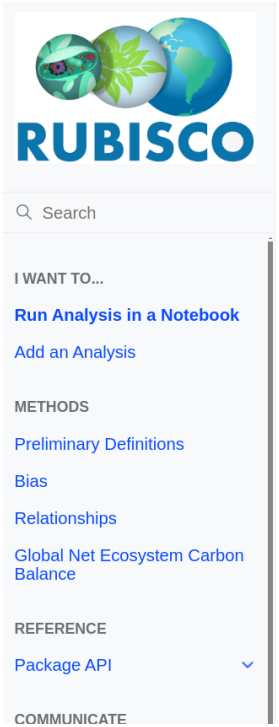
| | | |
|---|---|---|
| 'albedo \| CERESed4.1' | 'lai \| MODIS' | 'rlns \| GEWEX.SRB' |
| 'albedo \| GEWEX.SRB' | 'mrro \| CLASS' | 'rlns \| WRMC.BSRN' |
| 'biomass \| ESACCI' | 'mrro \| Dai' | 'rlus \| CERESed4.1' |
| 'biomass \| NBCD2000' | 'mrro \| LORA' | 'rlus \| FLUXNET2015' |
| 'biomass \| Thurner' | 'mrsos \| WangMao' | 'rlus \| GEWEX.SRB' |
| 'biomass \| Tropical' | 'nbp \| GCP' | 'rlus \| WRMC.BSRN' |
| 'biomass \| US.FOREST' | 'nbp \| Hoffman' | 'rns \| CERESed4.1' |
| 'burntFractionAll \| GFED4.1S' | 'nee \| FLUXCOM' | 'rns \| CLASS' |
| 'cSoil \| HWSD' | 'nee \| FLUXNET2015' | 'rns \| FLUXNET2015' |
| 'cSoil \| NCSCDV22' | 'pfext \| NSIDC' | 'rns \| GEWEX.SRB' |
| 'co2 \| NOAA.GMD' | 'pr \| CLASS' | 'rns \| WRMC.BSRN' |
| 'dtr \| CRU4.02' | 'pr \| CMAPv1904' | 'rsds \| CERESed4.1' |
| 'evspsbl \| GLEAMv3.3a' | 'pr \| FLUXNET2015' | 'rsds \| FLUXNET2015' |
| 'evspsbl \| MOD16A2' | 'pr \| GPCCv2018' | 'rsds \| GEWEX.SRB' |
| 'evspsbl \| MODIS' | 'pr \| GPCPv2.3' | 'rsds \| WRMC.BSRN' |
| 'fBNF \| DaviesBarnard' | 'reco \| FLUXCOM' | 'rsns \| CERESed4.1' |
| 'gpp \| FLUXCOM' | 'reco \| FLUXNET2015' | 'rsns \| FLUXNET2015' |
| 'gpp \| FLUXNET2015' | 'regions_continental \| ILAMB' | 'rsns \| GEWEX.SRB' |
| 'gpp \| WECANN' | 'regions_continental \| IPCC' | 'rsns \| WRMC.BSRN' |
| 'hfdsl \| CLASS' | 'regions_global_land \| ILAMB' | 'rsus \| CERESed4.1' |
| 'hfls \| CLASS' | 'regions_global_land_no_ant \| ILAMB' | 'rsus \| FLUXNET2015' |
| 'hfls \| DOLCE' | 'regions_whittaker_biomes \| ILAMB' | 'rsus \| GEWEX.SRB' |
| 'hfls \| FLUXCOM' | 'rhums \| CRU4.02' | 'rsus \| WRMC.BSRN' |
| 'hfls \| FLUXNET2015' | 'rhums \| ERA5' | 'swe \| CanSISE' |
| 'hfls \| WECANN' | 'river_basins \| Dai' | 'tas \| CRU4.02' |
| 'hfss \| CLASS' | 'rlds \| CERESed4.1' | 'tas \| FLUXNET2015' |
| 'hfss \| FLUXCOM' | 'rlds \| FLUXNET2015' | 'tasmax \| CRU4.02' |
| 'hfss \| FLUXNET2015' | 'rlds \| GEWEX.SRB' | 'tasmin \| CRU4.02' |
| 'hfss \| WECANN' | 'rlds \| WRMC.BSRN' | 'twsa \| GRACE' |
| 'lai \| AVH15C1' | 'rlns \| CERESed4.1' | |
| 'lai \| AVHRR' | 'rlns \| FLUXNET2015' | |

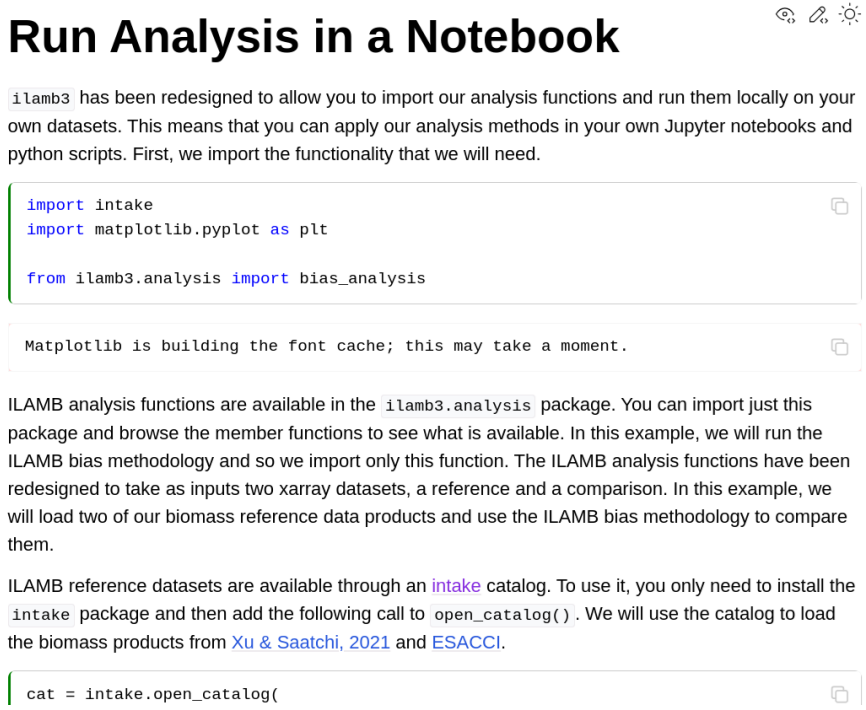
# Other Ways to Use ILAMB Data

```
In [1]: import intake
   ...: cat = intake.open_catalog("https://raw.githubusercontent.com/nocollier/intake-ilamb/main/ilamb.yaml")

In [2]: gpp = cat['gpp | WECANN'].read()

In [3]: gpp
Out[3]:
<xarray.Dataset>
Dimensions:      (time: 108, nb: 2, lat: 180, lon: 360)
Coordinates:
  * time         (time) object 2007-01-16 12:00:00 ... 2015-12-16 12:00:00
  * lat          (lat) float64 89.5 88.5 87.5 86.5 ... -86.5 -87.5 -88.5 -89.5
  * lon          (lon) float64 -179.5 -178.5 -177.5 -176.5 ... 177.5 178.5 179.5
Dimensions without coordinates: nb
Data variables:
    time_bounds  (time, nb) object 2007-01-01 00:00:00 ... 2016-01-01 00:00:00
    gpp          (time, lat, lon) float64 9.969e+36 9.969e+36 ... 9.969e+36
Attributes:
    title:         Water, Energy, and Carbon with Artificial Neural Networks ...
    version:       1
    institutions:  Columbia University
    source:        Solar Induced Fluorescence (SIF), Air Temperature, Precipi...
    history:       \n2020-11-02: downloaded https://avdc.gsfc.nasa.gov/pub/da...
    references:    \n@ARTICLE{Alemohammad2017,\n  author = {Alemohammad, S. H...
    comments:      \ntime_period: 2007-01 through 2015-11; temporal_resolutio...
    convention:    CF-1.8
```

# ILAMB v3: On the Horizon

- Shift to xarray as a base object.
- Adapt to the way researchers are working.
- Working from the bottom up and making soft *releases* as we go.
- Each new capability will be fully documented.
- Great time to get me your wish lists.



## Run Analysis in a Notebook

`ilamb3` has been redesigned to allow you to import our analysis functions and run them locally on your own datasets. This means that you can apply our analysis methods in your own Jupyter notebooks and python scripts. First, we import the functionality that we will need.

```
import intake
import matplotlib.pyplot as plt

from ilamb3.analysis import bias_analysis
```

```
Matplotlib is building the font cache; this may take a moment.
```

ILAMB analysis functions are available in the `ilamb3.analysis` package. You can import just this package and browse the member functions to see what is available. In this example, we will run the ILAMB bias methodology and so we import only this function. The ILAMB analysis functions have been redesigned to take as inputs two xarray datasets, a reference and a comparison. In this example, we will load two of our biomass reference data products and use the ILAMB bias methodology to compare them.

ILAMB reference datasets are available through an intake catalog. To use it, you only need to install the `intake` package and then add the following call to `open_catalog()`. We will use the catalog to load the biomass products from Xu & Saatchi, 2021 and ESACCI.

```
cat = intake.open_catalog(
```

https://github.com/rubisco-sfa/ilamb3