Models and Data

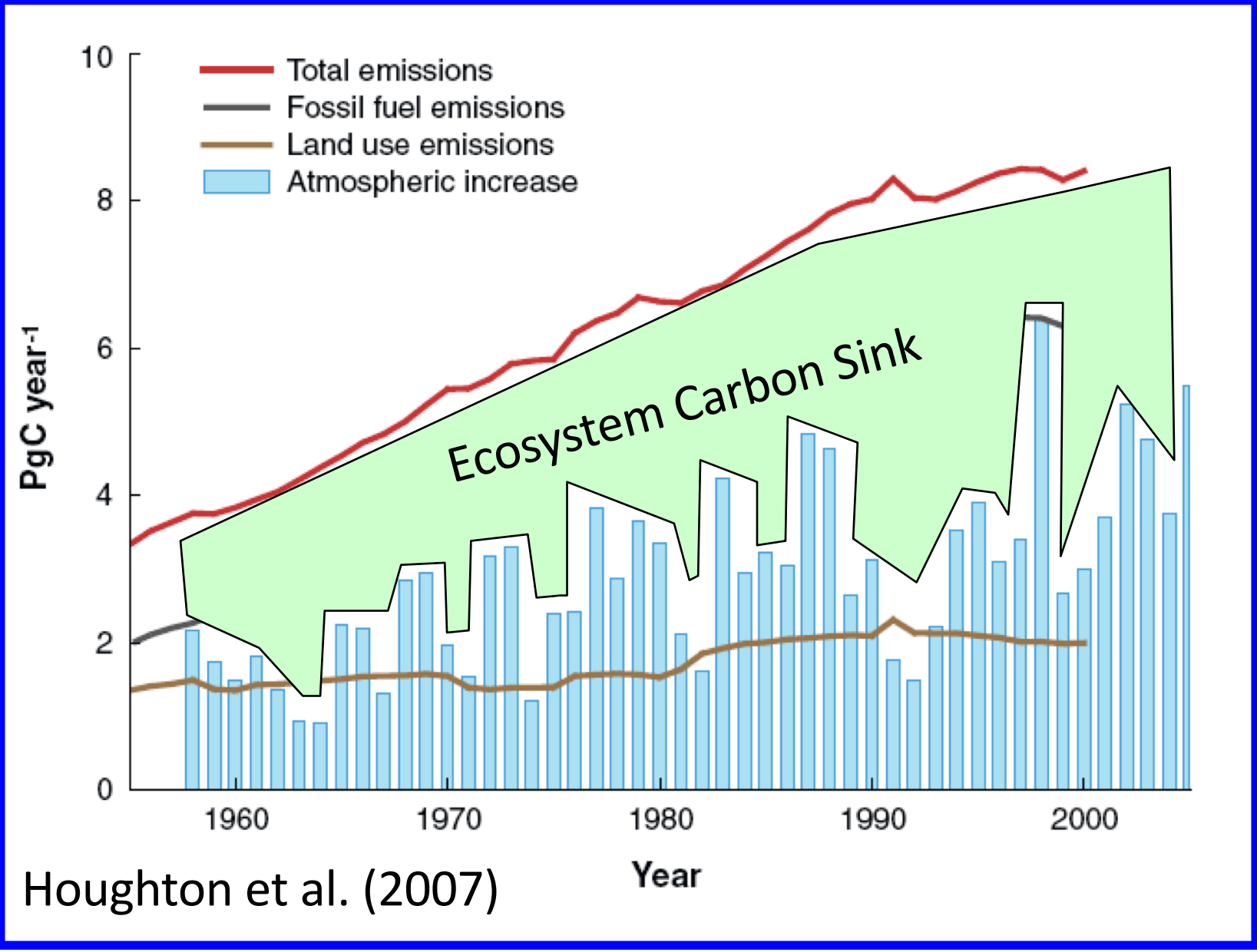
A Gentle Introduction

Why am I here?

The Big Picture

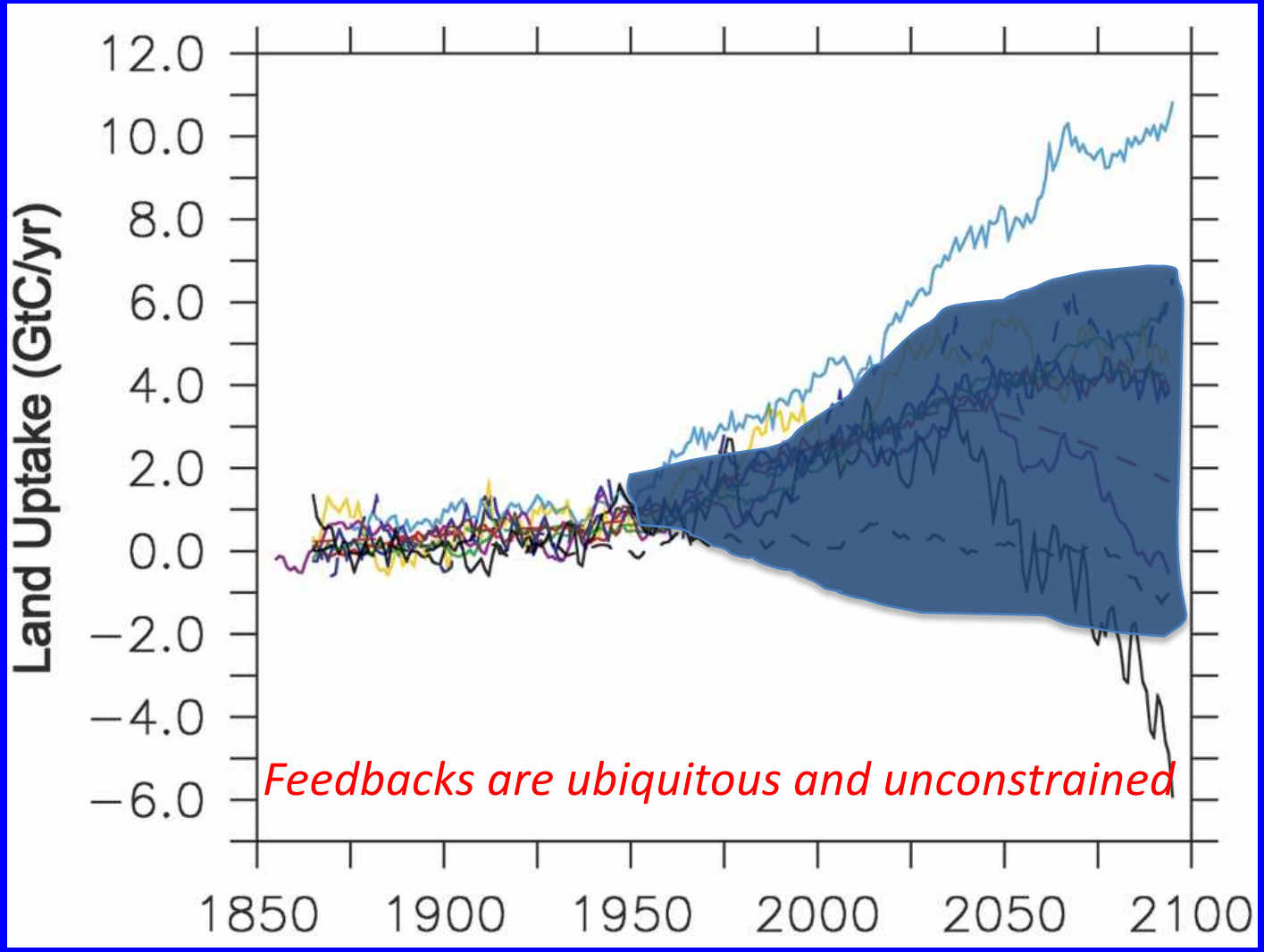
- GIVEN:
  - Interannual variability in the growth rate of atmospheric greenhouse gases is **driven by terrestrial carbon cycle**
  - Long-term accumulation of greenhouse gases and global energy balance is strongly mediated by *direct and indirect* feedbacks of biospheric (land+ocean) vegetation distribution and carbon cycling to future variability in climate, nutrients/water, and land use.
- WE NEED TO CONSIDER:
  - Diagnosing and predicting these responses is **inherently regional** given the covariation and long-term adaptation of species and climate across space.
  - Complexity of feedbacks requires integration of observations and experiments with **sufficiently** complex, structurally correct, and well-tuned models
  - Questions: What is the right spatial/temporal scale? What is a sufficiently complex, correct, and well-tuned model?

Biology drives Physics



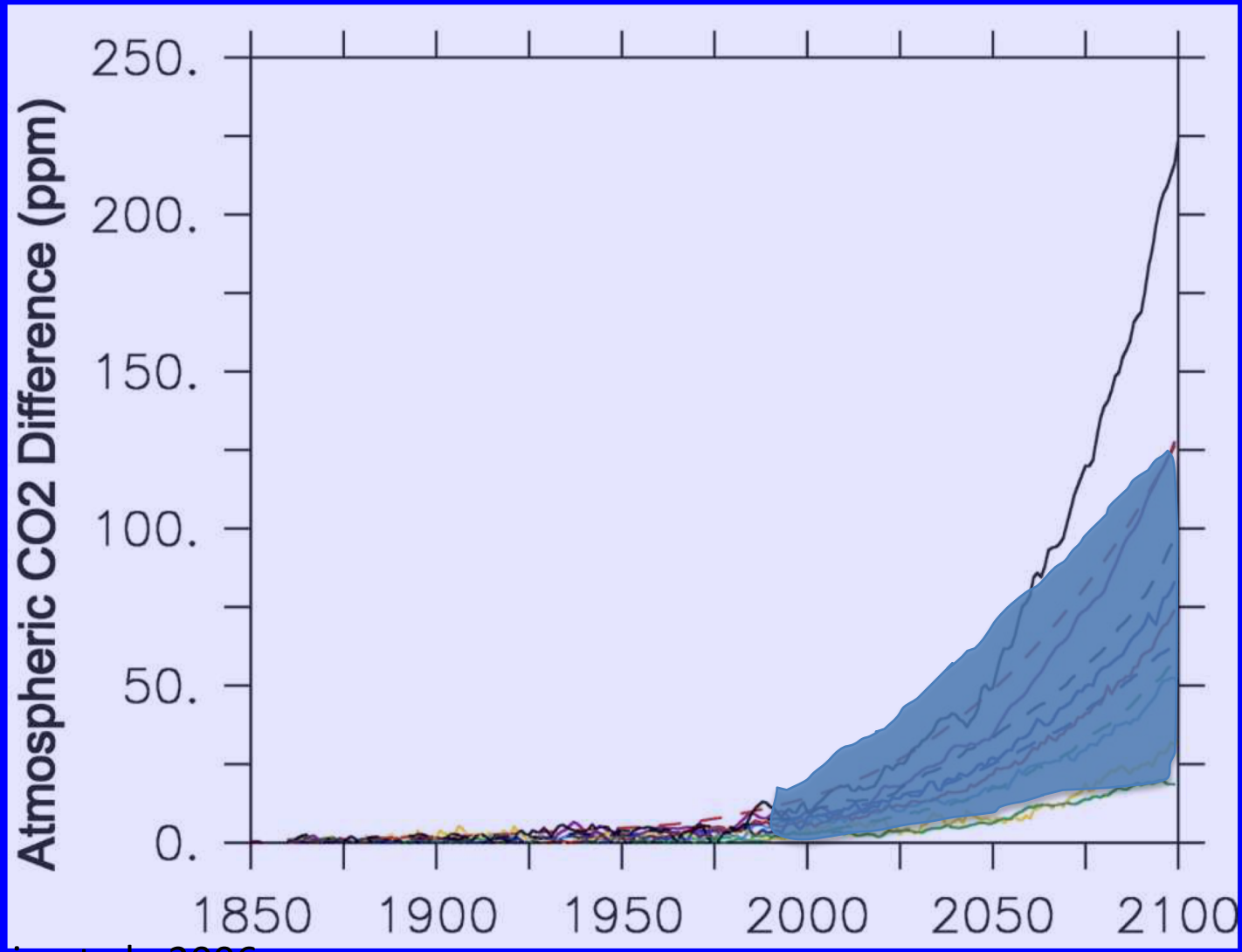
Houghton et al. (2007)

And we fail at modeling it…



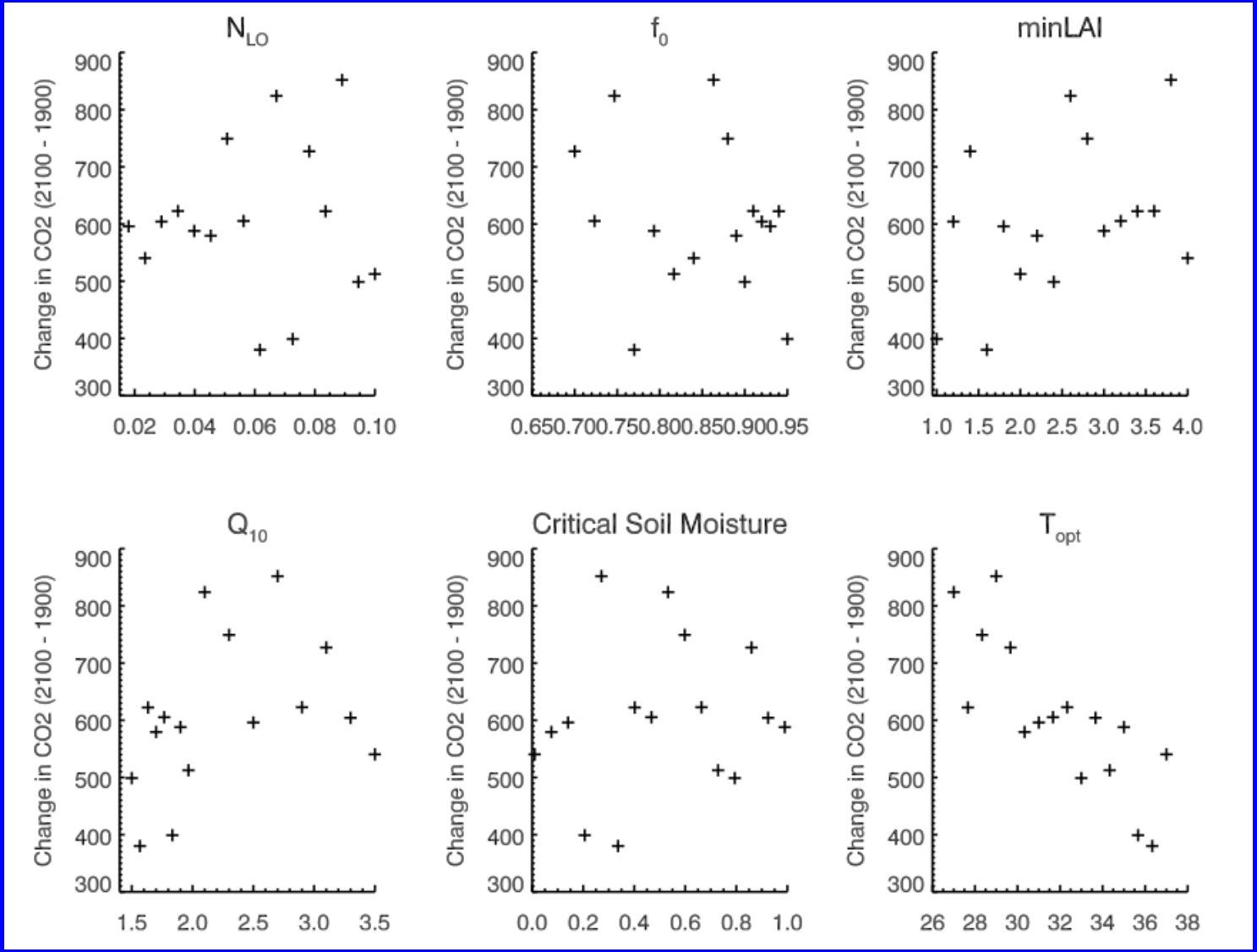
Friedlingstein et al., J Clim, 2006

It matters…



Friedlingstein et al., 2006

Parameter uncertainty!



Booth et al., 2012, ERL

Some terms

- Model
- Parameters
- Traits
- Plant Functional Type
- States
- Initial conditions
- Fluxes
- Forcing
- Data assimilation
- Observation operator
- Bayesian
- Prior
- Posterior
- Likelihood
- Model emulation
- Validation data
- Machine learning
- Sensitivity
- Training data
- Sto(y)chastic
- Markov Chain
- Gaussian
- Uncertainty
- Error covariance matrix
- Kalman filter
- Data Fusion
- Equifinality
- Prediction
- Projection

Some topology of ecosystem models

- Heuristic models
  - Penman-Montieth, Montieth/Moncrieff GPP model
- Big-leaf vs sunlit/shaded big-leaf vs canopy vs individual model
  - LAI = Leaf area index, LAD = leaf area density, mechanistic photosynthesis (Farquhar) vs empirical (Jarvis), allocation,
- Soil turnover vs process-based soil
  - "fast" and "slow" pools vs soil structural vs nutrient cycling vs microbial loop
- Vegetation Demographic / "Dynamic" model
  - Age since disturbance, height, mortality, growth, recruitment
- Land surface model (or "offline" model)
  - Energy balance
- Land-atmosphere coupled model
  - Weather/climate models, "coupler", "ESM"
- Hydrologic model
  - Routing of water in rivers, groundwater, conductances
- Empirical models
  - Machine learning, emulated models, big box models,
- Ensemble models
  - Sensitivity to initial conditions (chaos) vs forcing/parameters
- Lots of acronyms: Biome-BGC, ED2, JULES, SIPNET, CLM/CESM, ACME, WRF-NOAH, LPJ-GUESS, LANDIS-II, FUBAR, TRIFFID, ORCHIDEE, SWAT
- Parsimony

Why models and data?

- Old way:
  - Make one model
  - Guess some parameters
  - Make up some drivers
  - Compare to your favorite data source
  - Publish the best comparisons
  - Attribute discrepancies to error
  - Be happy

Why models and data?

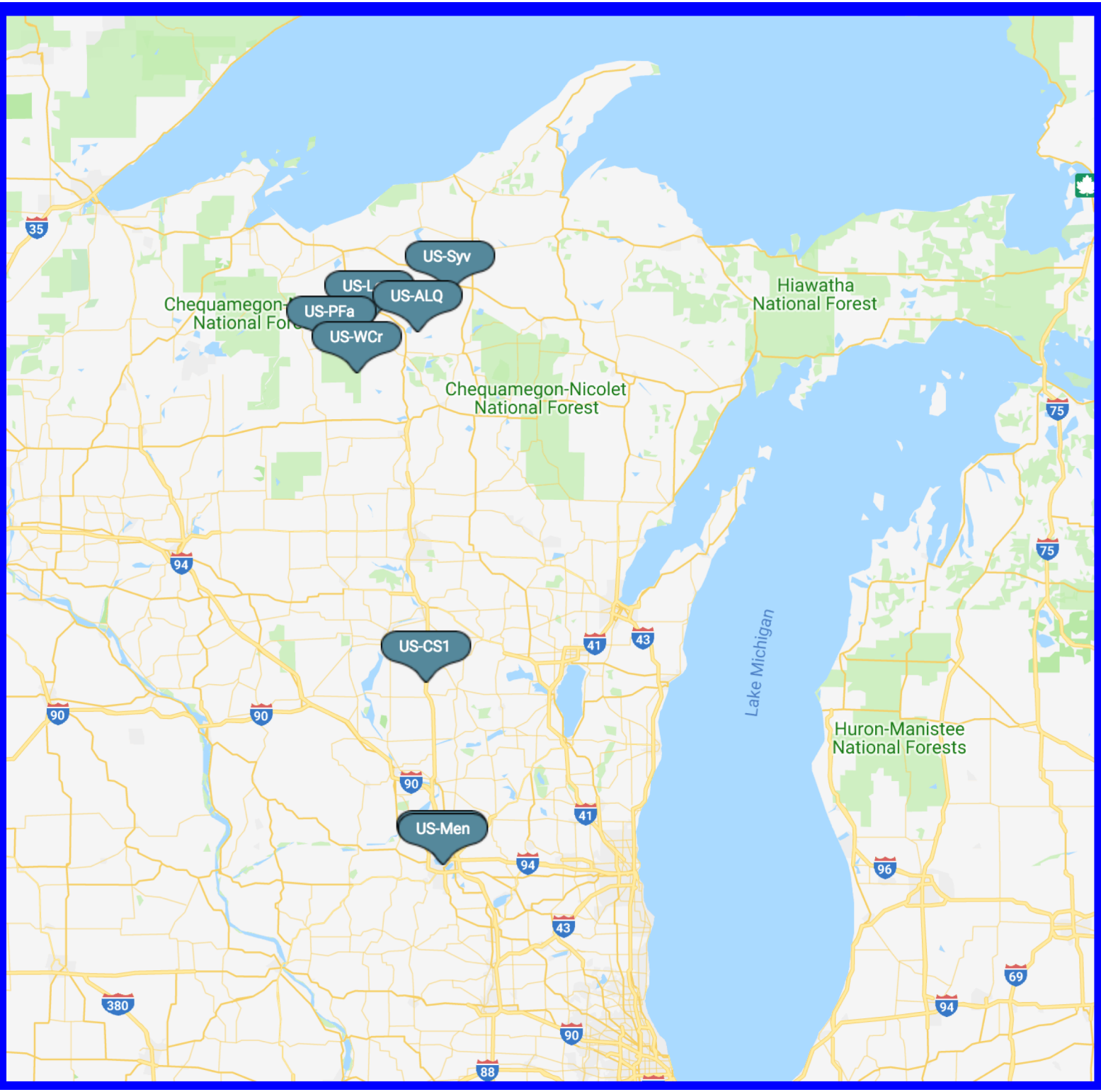
- New way:
  - Quantify likelihood or probability of multiple model(s) being consistent with multiple observations and instrument + sampling uncertainty
  - Find how various models or parameters (working hypotheses) cannot explain observations accounting for both model and data confidence
  - Learn something about fundamental interactions
  - Publish the discrepancies and knowledge gained
  - Work harder, be slightly less happy, but generate more knowledge

Basis of data assimilation

- Goal: Find the collection of model process, state and/or parameter sets that are consistent with data
- Bayes' rule: P(model | data) =
  - P(model) * P(data | model) / P(data)
- Implementations typically rely on estimating model sensitivity through ensemble runs and calculating likelihood using a set of training data observations
  - Least squares regression, Stratified random sampling, Markov Chain Monte Carlo (MCMC), Ensemble Kalman Filter (EKF)

What might DA be good for?

- Turn diverse jumbled mess of data into useful tests of hypotheses and improve diagnosis and prediction of regional carbon/water cycling/vegetation dynamics and climate and land use impacts!
- Testbed: Chequamegon Heterogenous Ecosystem Energy-Balance Study Enabled by a High-Density Array of Detectors 2019
  - #CHEESEHEAD19
- Observing system simulation experiment (OSSE)


US-Syv
US-L
US-ALQ
US-PFa
US-WCr
US-CS1
US-Men
Chequamegon-
National For
Chequamegon-Nicolet
National Forest
Hiawatha
National Forest
Lake Michigan
Huron-Manistee
National Forests
35
94
90
90
90
88
380
41
43
41
43
94
75
75
96
69
94

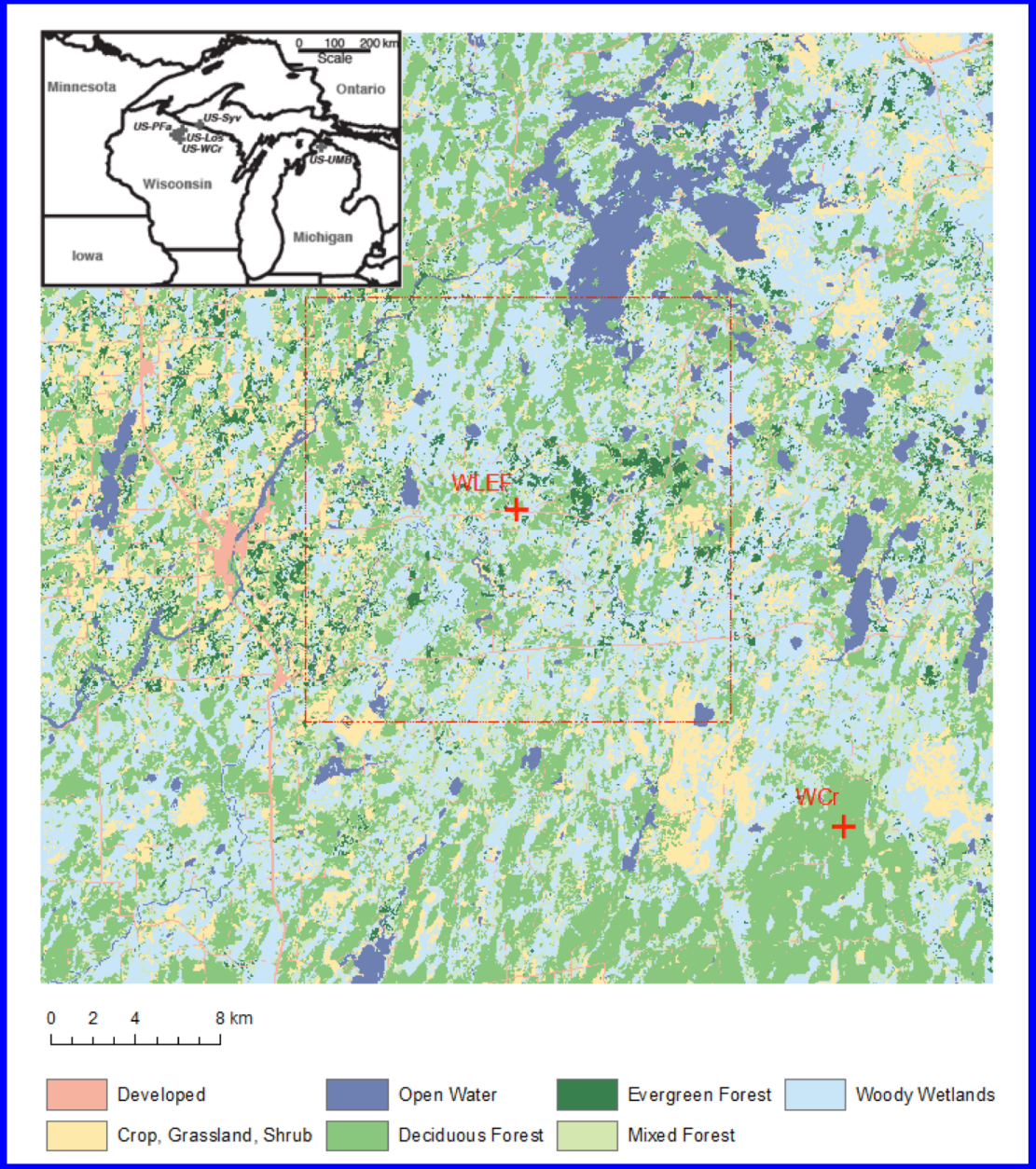
Why Regional?

- Spatial interpolation/extrapolation
- Evaluation across scales
- Landscape level controls on biogeochem.
- Understand drivers of spatial variability
- Emergent properties of landscapes

Park Falls/Chequamegon-Nicolet National Forest region, Wisconsin

Park Falls WLEF tower (US-PFa)
EC fluxes at 30, 122, 396 m
NOAA tall tower greenhouse gas site
COSMOS soil moisture
TCCON column GHG

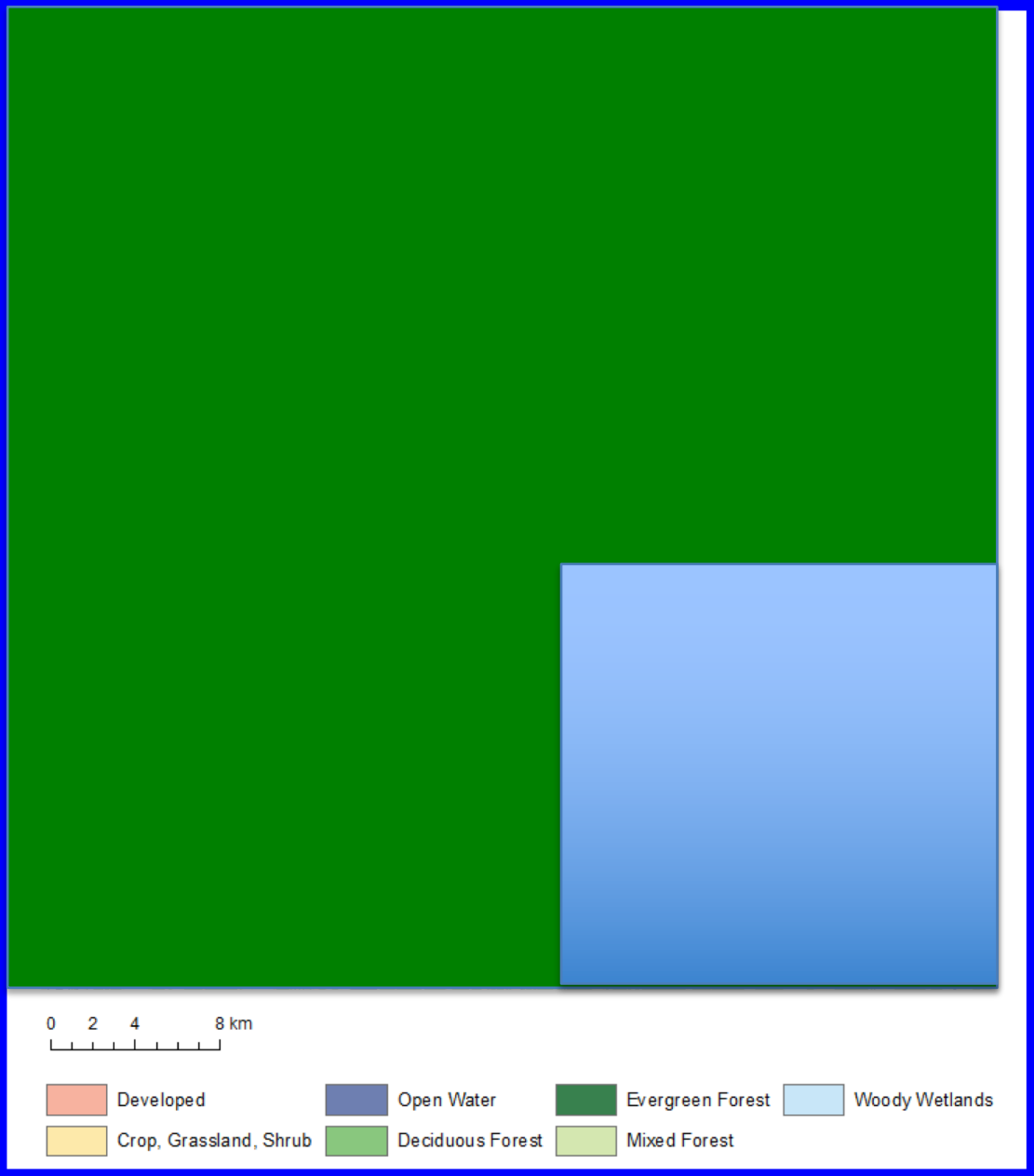
Credit: Matt Rydzik (U Wisconsin)



Desai et al., 2015, AFM

Park Falls/Chequamegon-Nicolet National Forest region, Wisconsin

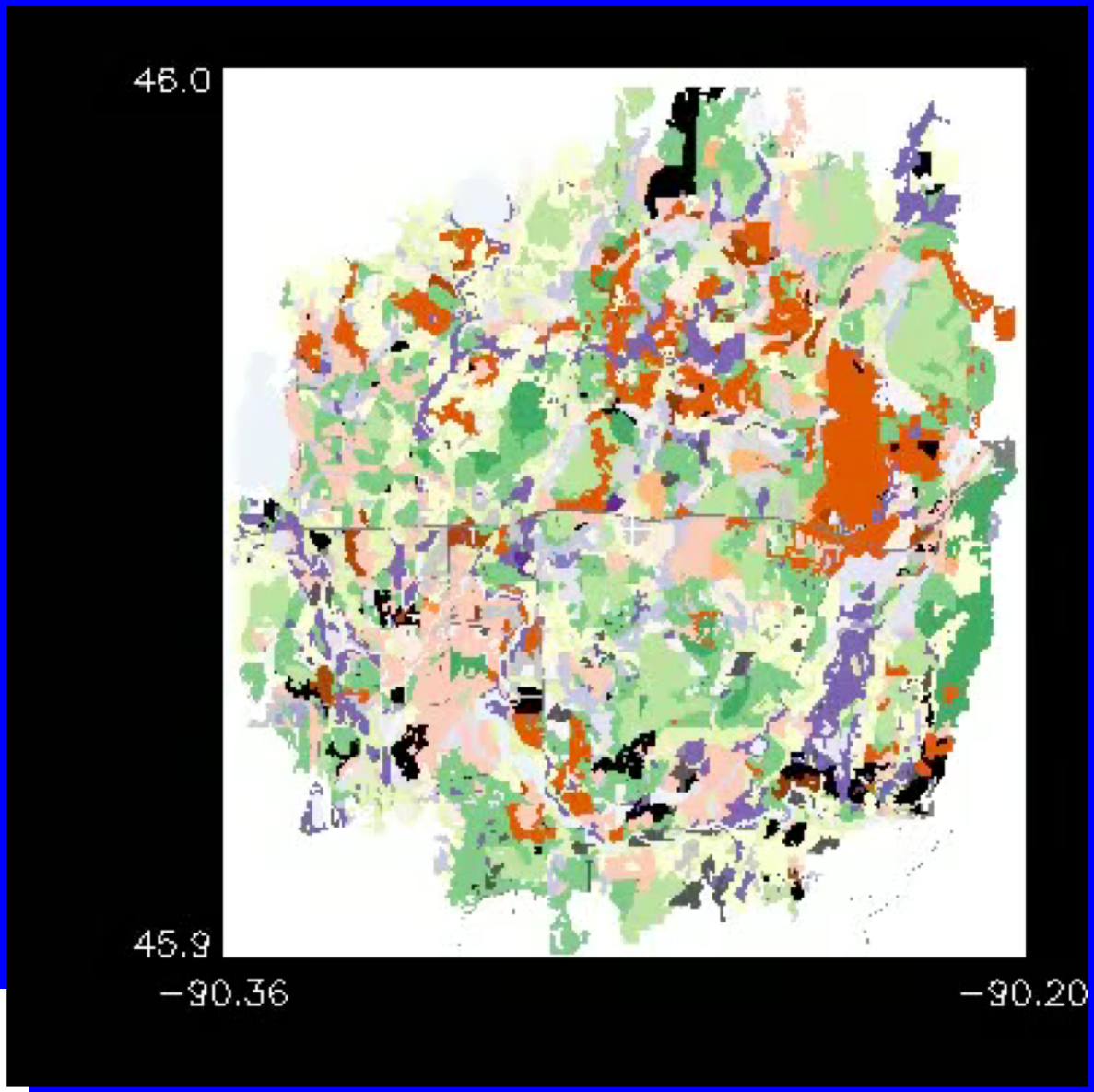
Park Falls WLEF tower (US-PFa)
EC fluxes at 30, 122, 396 m
NOAA tall tower greenhouse gas site
COSMOS soil moisture
TCCON column GHG

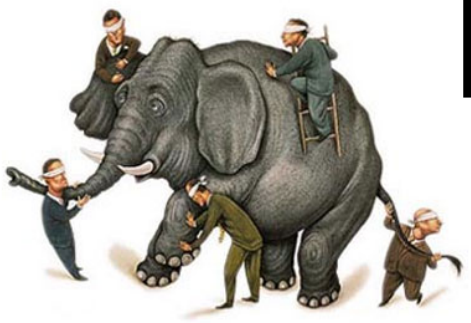


Credit: Matt Rydzik (U Wisconsin)

Desai et al., 2015, AFM

Flux towers see the trees for the forest…



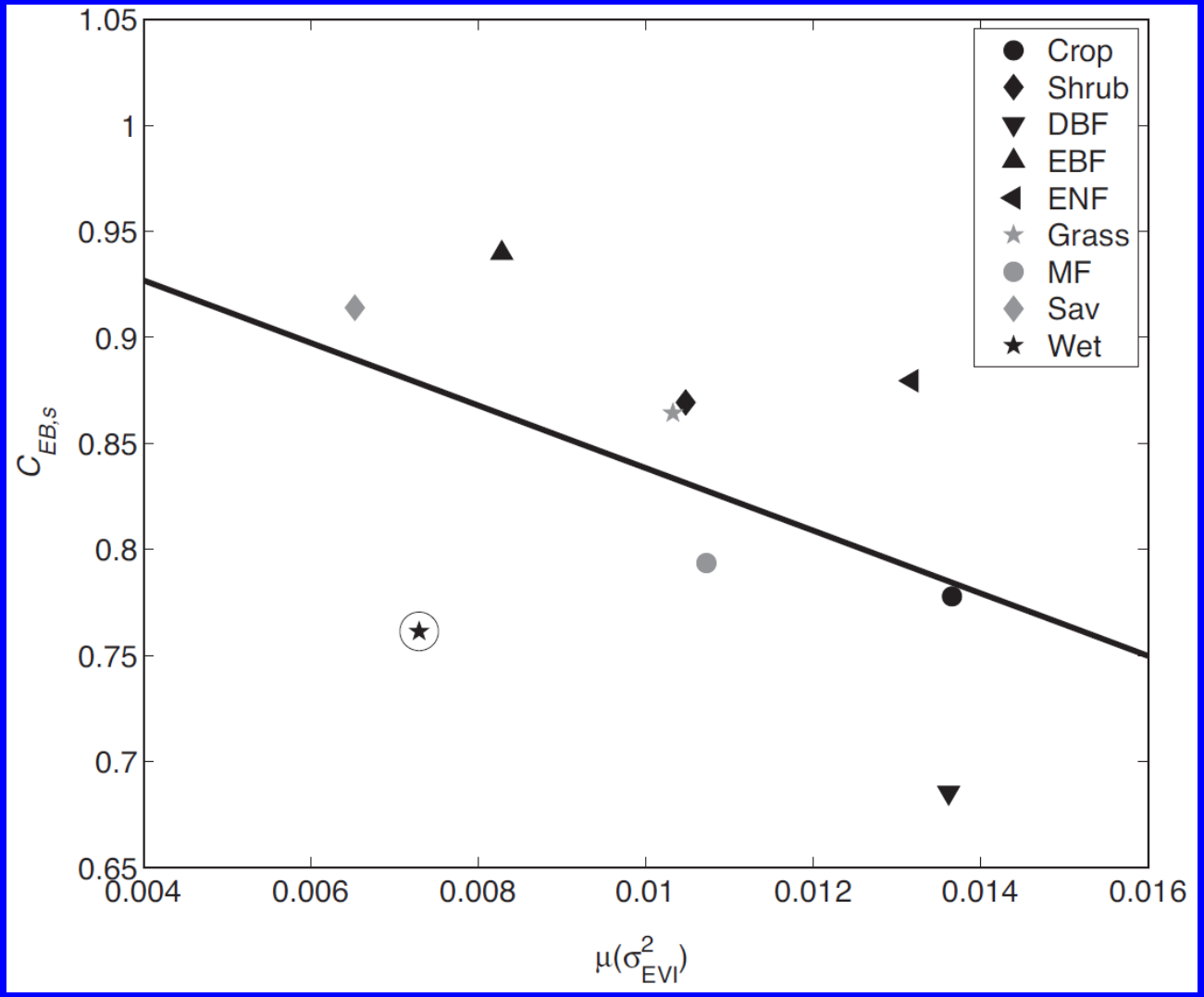


Adopted from a version by HaPE Schmid (KIT)

Paul Stoy is almost always right

EBC=
H+Le

Rnet-G



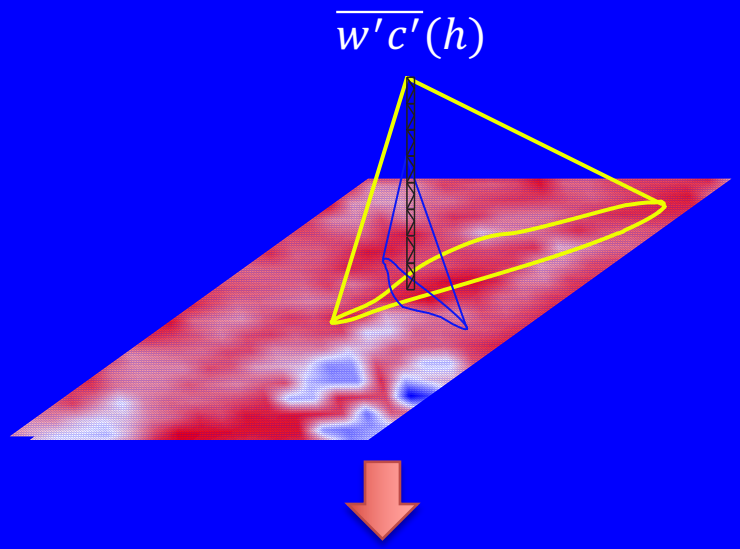
Greenness spatial variance

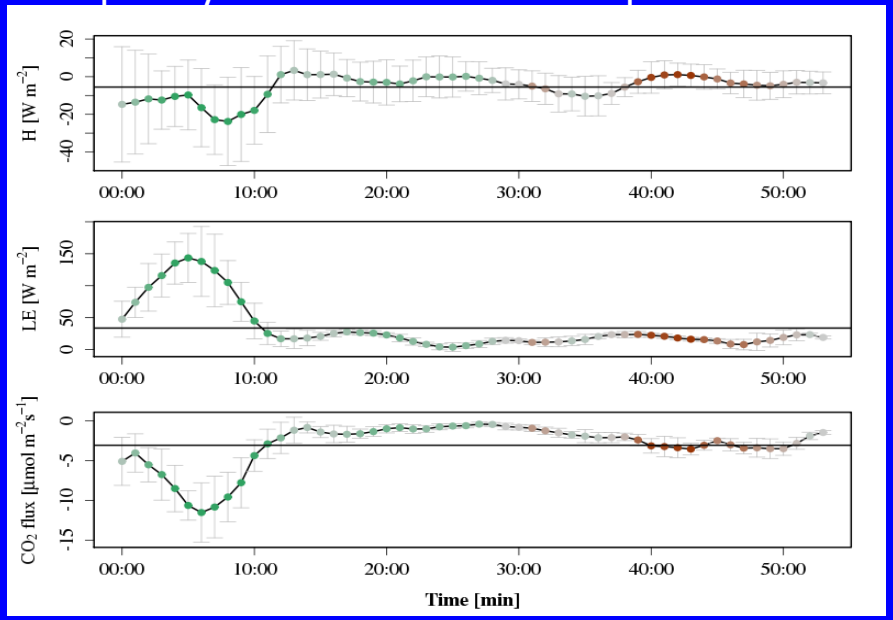
Stoy et al., 2013, AFM

Environmental Response Function (ERF) scaling method

Metzger et al., 2013, Biogeosci , Xu et al., 2017, AFM, Metzger, 2018, AFM, Xu et al., 2018, AFM

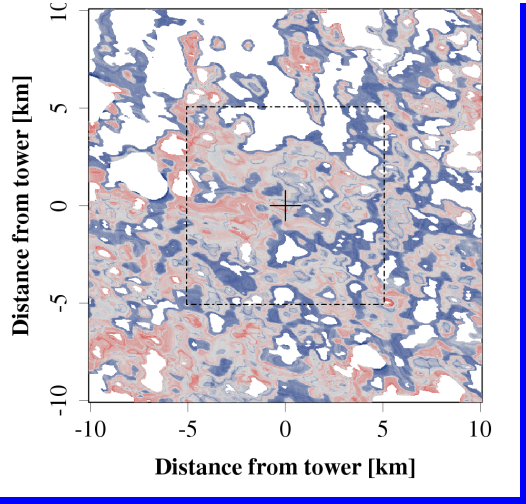
High frequency time series of flux response and drivers



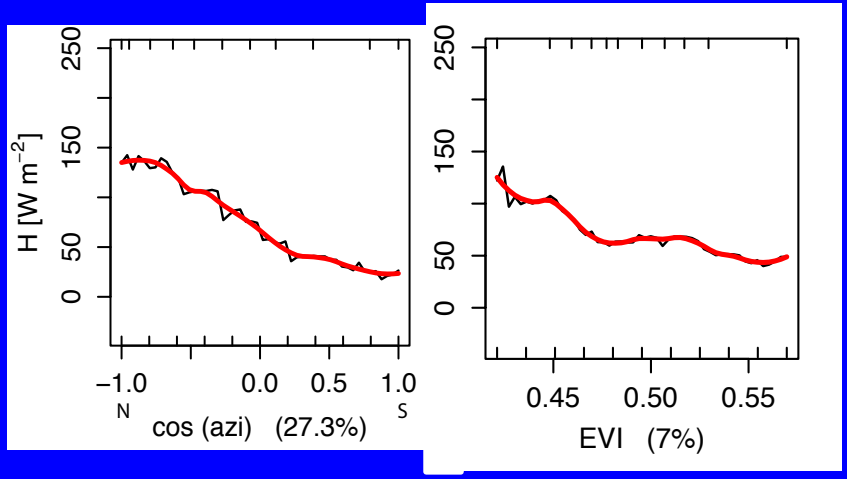


$$\frac{1}{4L^2}\int_{-L}^{+L}\int_{-L}^{+L}\overline{w'c'}(h)\ dx\ dy$$

Domain-projected turbulent flux at measurement level

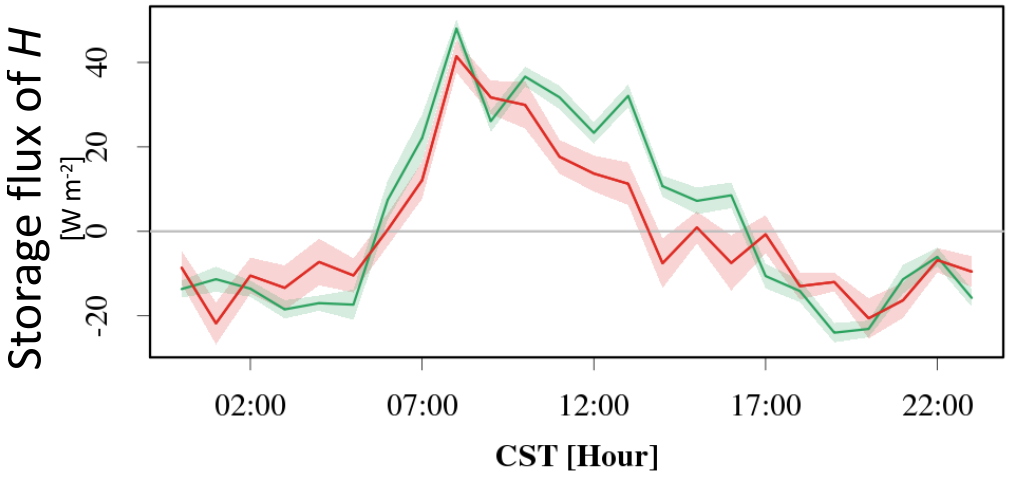


Extracted relationships

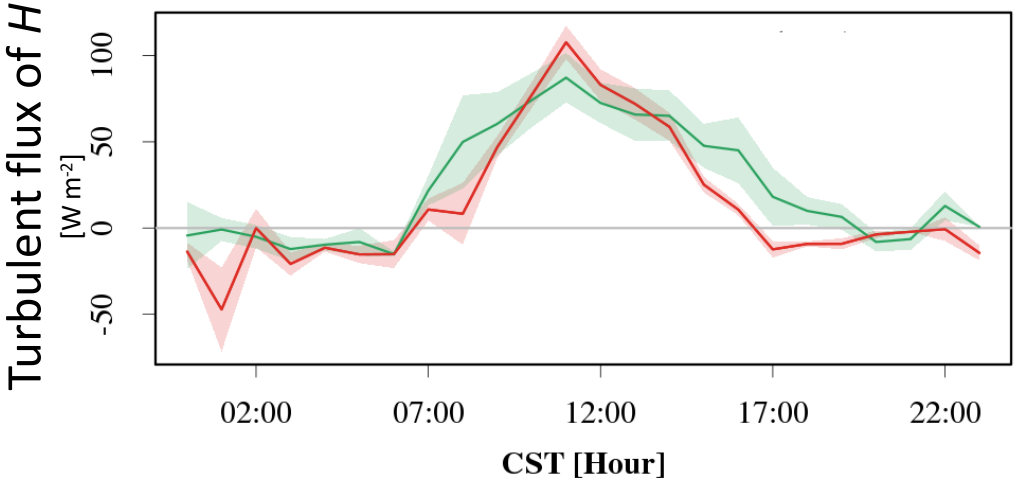


Does rectified surface atmosphere exchange help EBC and location bias?

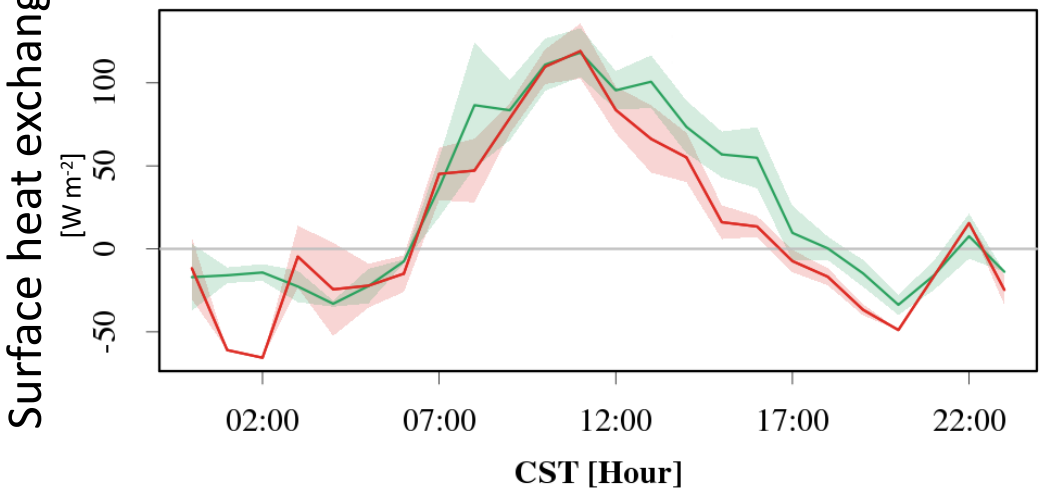
Tower: -0.3 W/m^2 Volume: 2.4 W/m^2



Tower: 19.4 W/m^2 Volume: 24.1 W/m^2



Tower: 18.5 W/m^2 Volume: 25.8 W/m^2

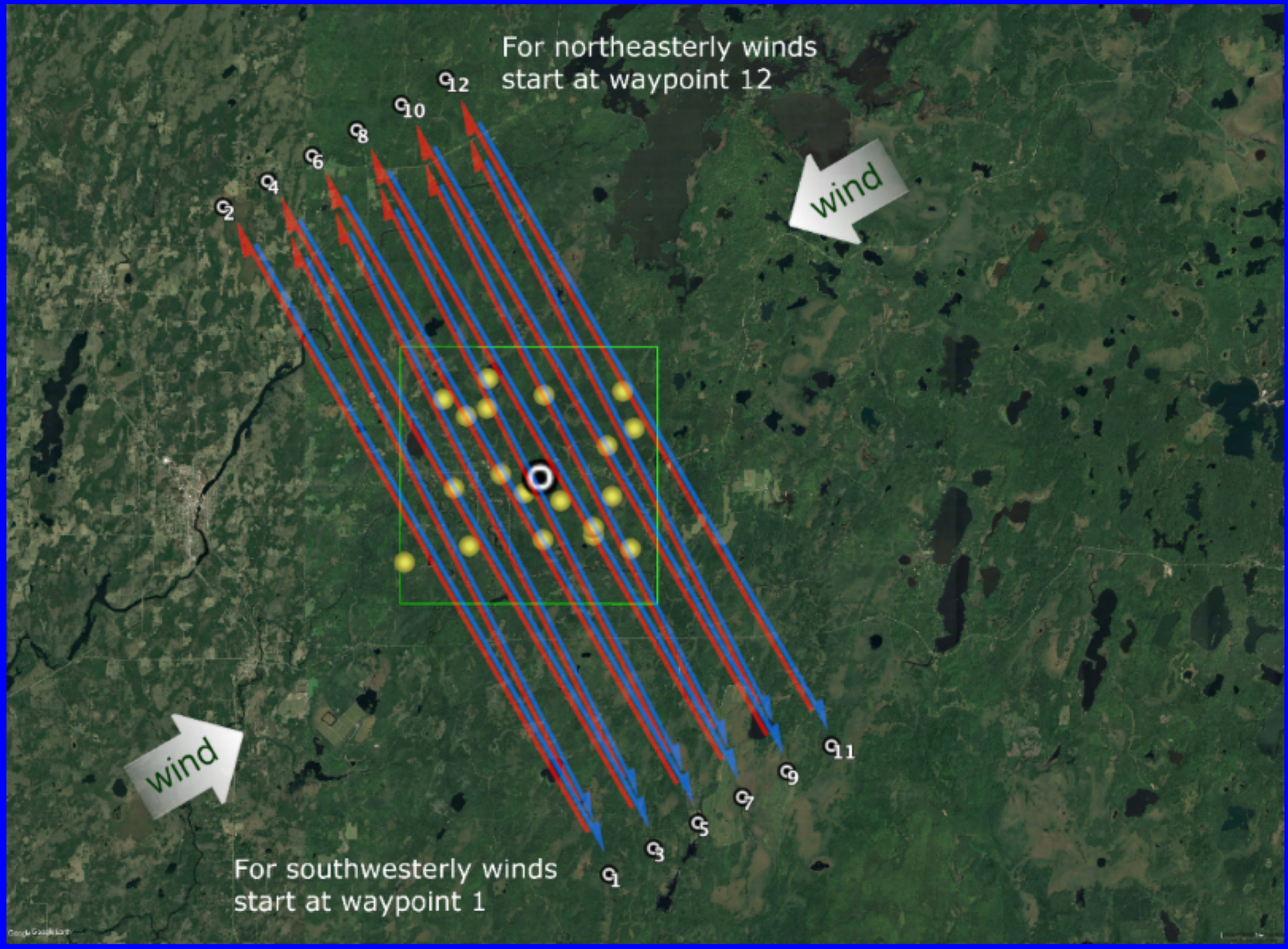


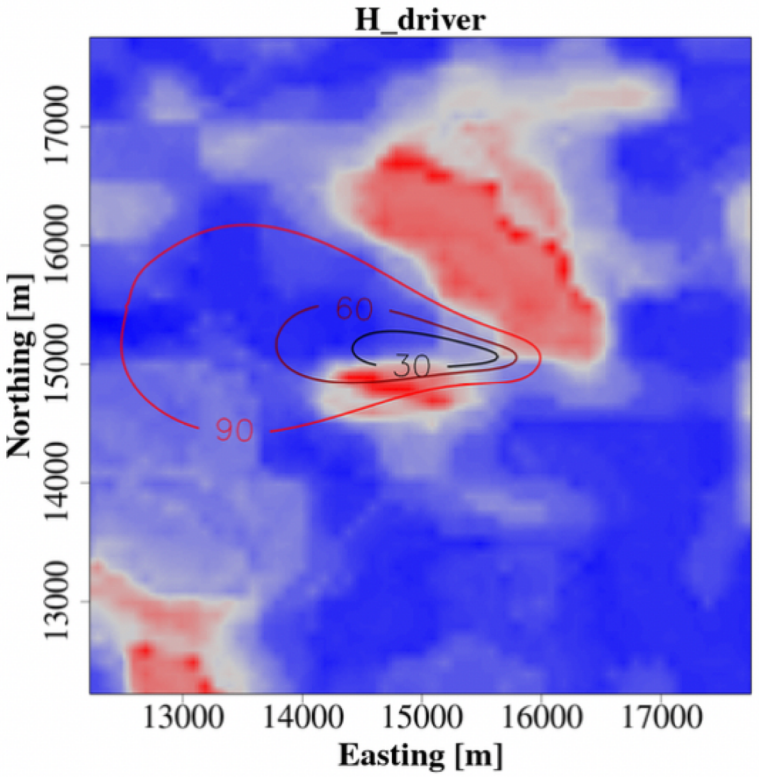
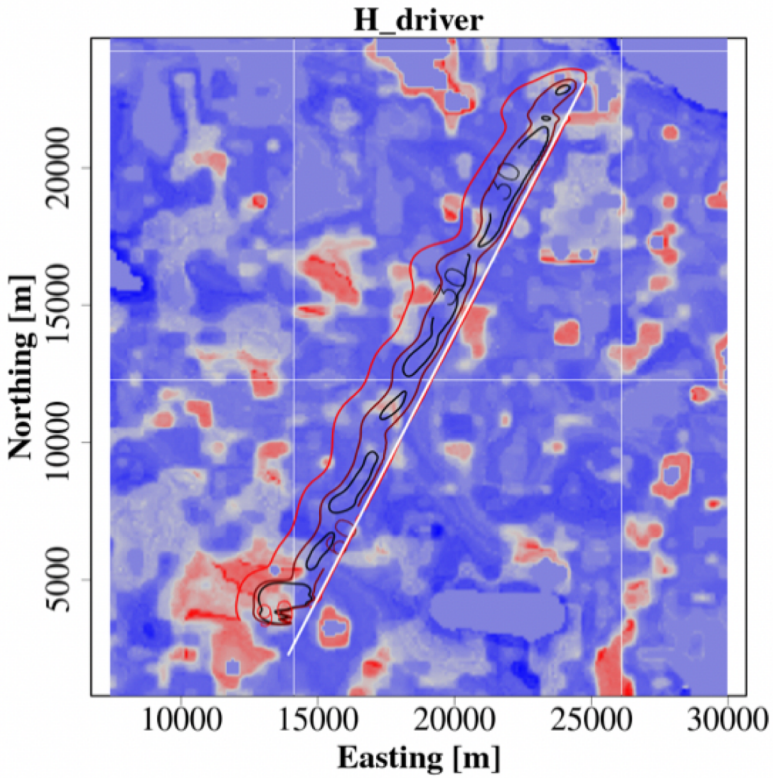
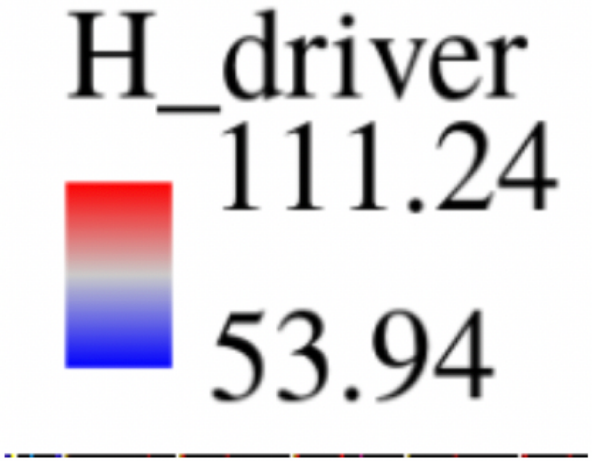
Volume-rectified: + 7.3 W/m^2 potentially help close energy balance!

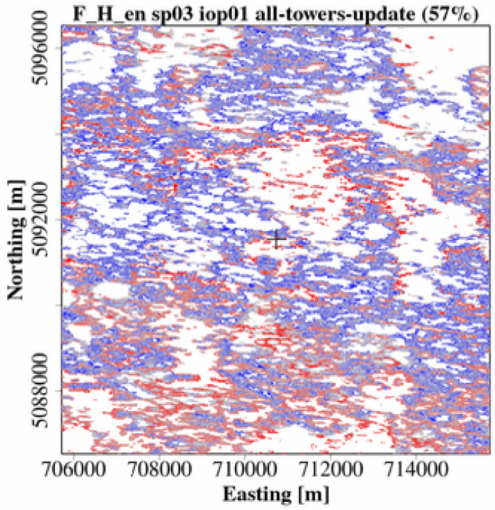
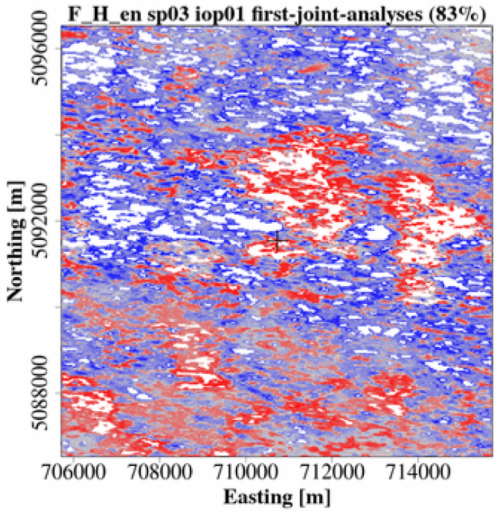
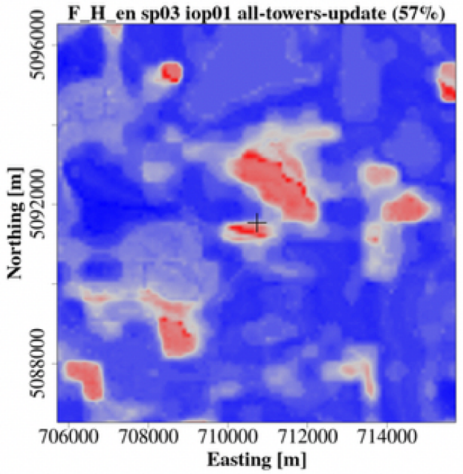
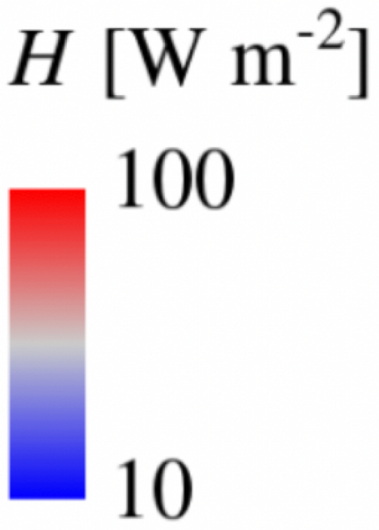
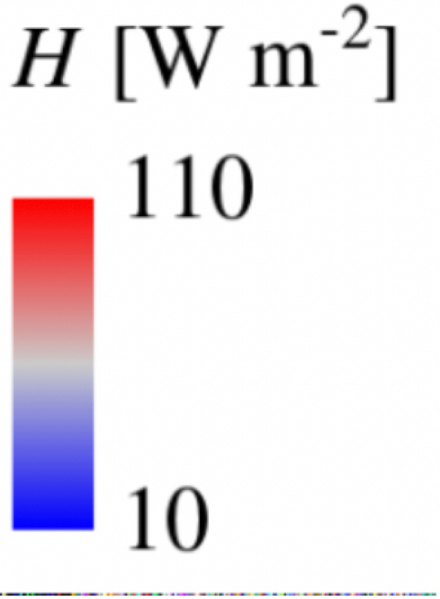
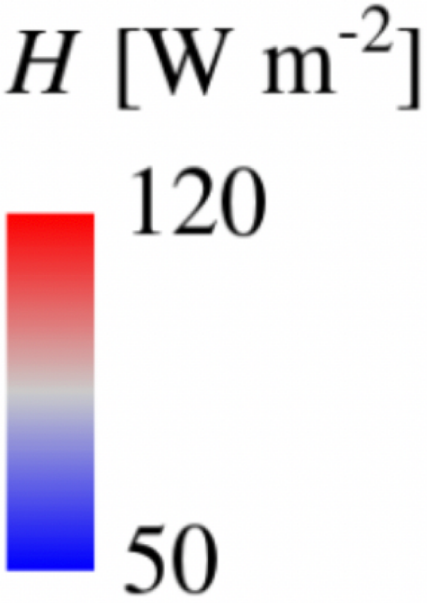
Location bias!

tower-observed

volume-rectified


For northeasterly winds
start at waypoint 12
wind
wind
For southwesterly winds
start at waypoint 1

WLEF @120m	UWKA @100m (leg_88_rep12)
	
	

tower projection	tower + aircraft projection	original LES "truth"
		
		

The Future

- 1-6 degree C warming in next 100 years, concentrated in high latitude, wet areas get wetter, dry areas get drier in frequency and intensity
- Atmospheric CO_2 at 550-950 by 2100, maybe even double or triple by 2200?
- Population grows at least 30% more, food demand increases faster as diets change - plants will always need to provide food, fiber, lumber, medicine
- Meteorites, alien invasions, nuclear winter…
- Urgent need for ecosystem science to be able to make credible predictions about complex interactions and thresholds of rapid change in ecosystem dynamics and nutrient cycling
- Model-data assimilation has to be a tool in your workbench

Thanks!

- Ankur Desai, University of Wisconsin-Madison
  - desai@aos.wisc.edu, @profdesai
  - http://flux.aos.wisc.edu

Photo by Brian Butterworth