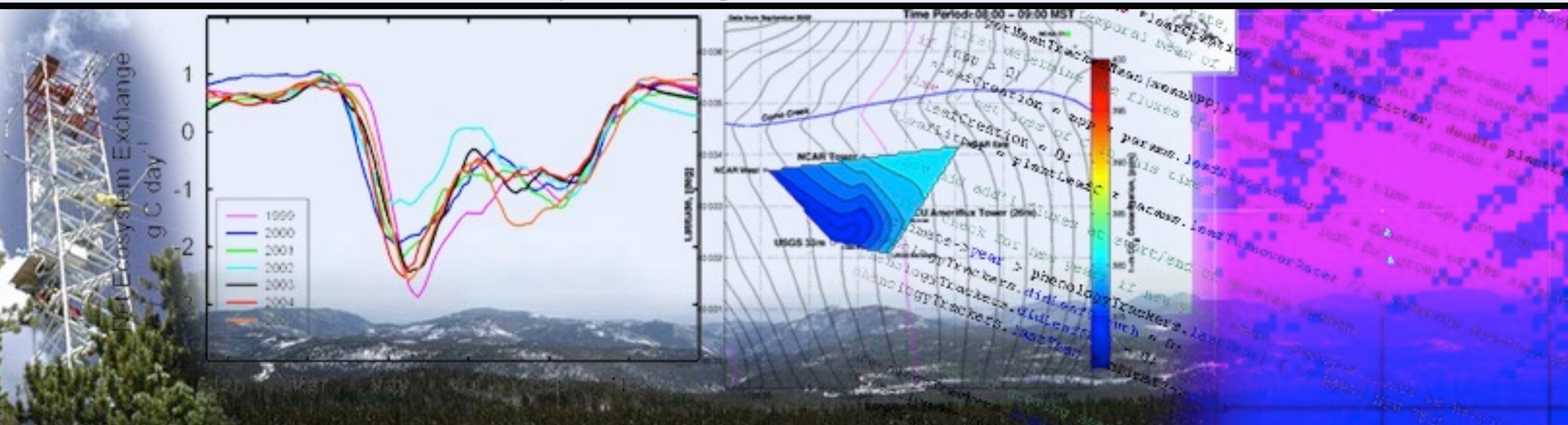
Data Assimilation for Vegetated Ecosystems

Dave Moore

July 2022

University of Arizona, School of Natural Resource & Environment

davidjpmoore@email.arizona.edu



Very briefly

HOW TO MAKE A DATA ASSIMILATION SYSTEM

You will need...

- A **model** (today we can use SIPNET)
  - The model must estimate the type of data you wish to assimilate.
  - It must also be appropriate for the system
- The **data** (estimate the UNCERTAINTY!)
  - The data can be any data (today we'll stick to flux data mostly)
- An **estimator** (some iterative 'cost' function)
  - A way of finding the 'optimal parameter set'

Don’t forget the acronym!

- **A**ssimilatio**N** **K**alman filter for **U**nderstory **R**espiration (ANKUR)
- **K**alman **I**nteractive **M**odel (KIM)
- **M**ontecarlo **I**nitiated **K**alman **E**stimator (MIKE)
- **D**ata **A**ssimilation of **N**itrogen L**I**mination of **C**arbon **A**ssimilation (DANICA)
- **AN**nual **D**rought**Y** **R**ealtime **E**xperimental **W**etlands (ANDREW)
- **D**ata **A**ssimilation of **V**egetated **E**cosystem**s** (DAVEs)
- **E**cosystem **D**emography (ED)
- **T**hermal **R**andom **E**vaporation from **V**egetation (TREV)

A couple of examples of applying data assimilation that illustrate some pitfalls to address some science questions

1. Optimizing fluxes

2. Inverse parameter estimation

3. Using an optimized model to test model structure (Big hypotheses)

Ecosystem models

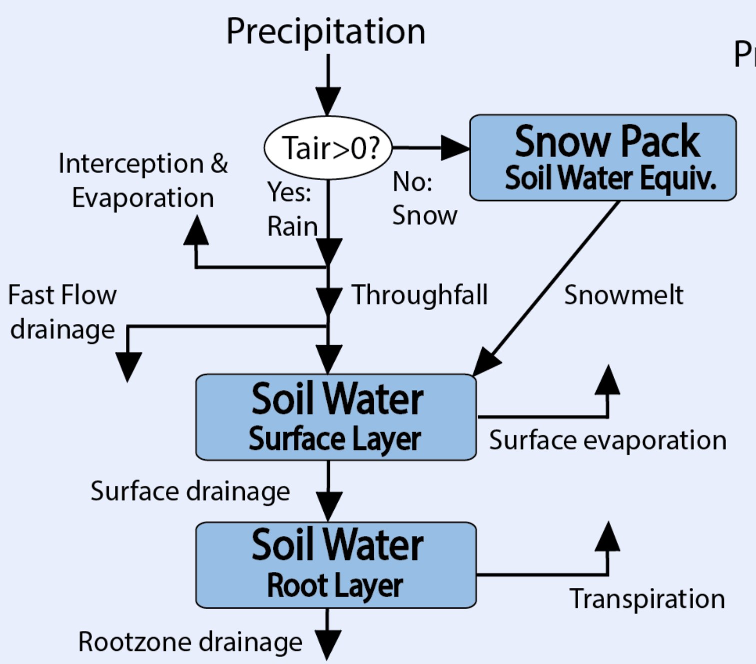
All models are wrong

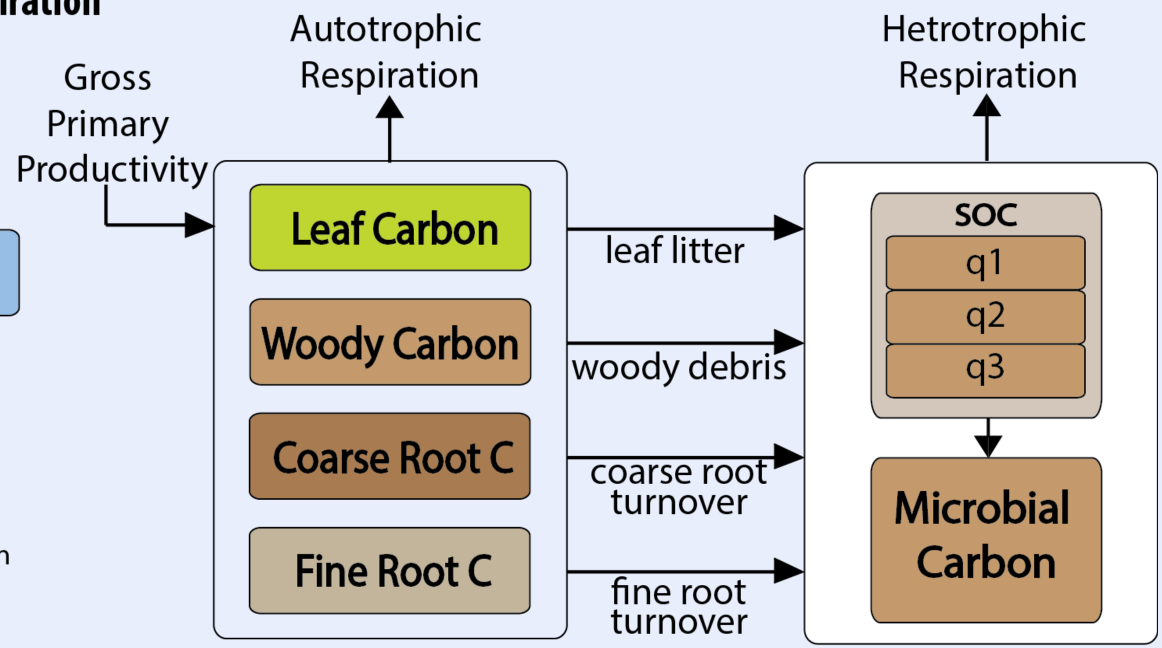
(but some of them are useful)

What type of model would be useful?

- I like the model to be process based
  - So we can learn from failure & try to predict
- Should be simple to avoid over-fitting
  - Few parameters
  - Also runs quickly!
- Needs to calculate the data you want to assimilate
  - So we can directly compare data to the model output
- Needs to be driven by readily measured climate variables
  - If you want to use it all flux sites

The Simplfied Photosynthesis and EvapoTranspiration (SIPNET) data assimilation system





- Twice-daily time step (day & night)
- Goal: keep model as simple as possible

Photosynthesis:

f (Leaf C, T_{air}, VPD, PAR, Soil Moisture)

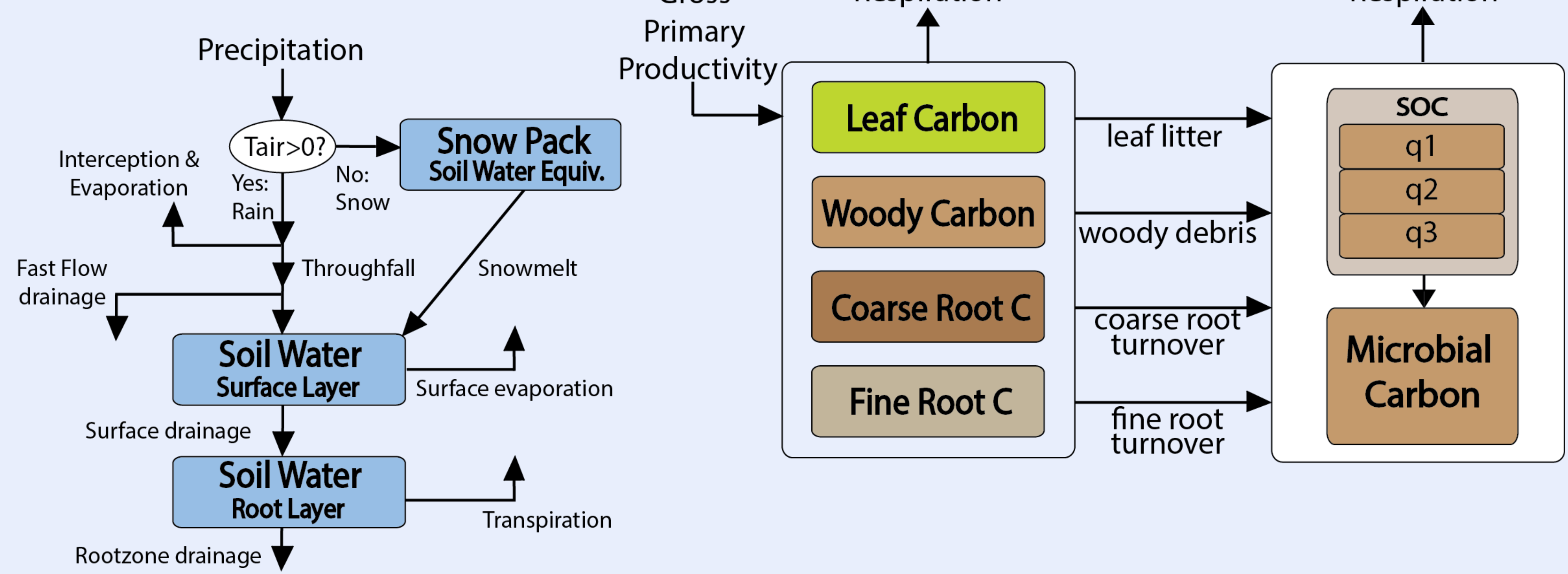
Autotrophic Respiration:

f (Plant C, T_{air})

Heterotrophic Respiration:

f (Soil C, T_{soil}, Soil Moisture)

The Simplfied Photosynthesis and EvapoTranspiration (SIPNET) data assimilation system

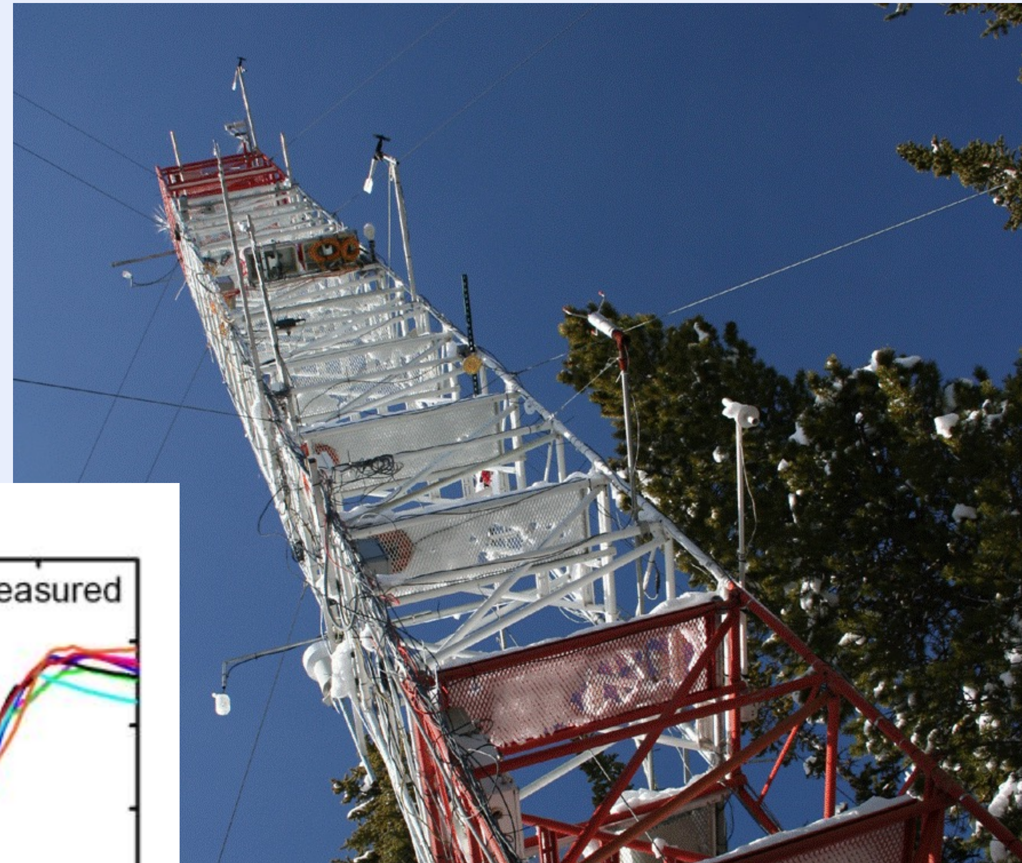


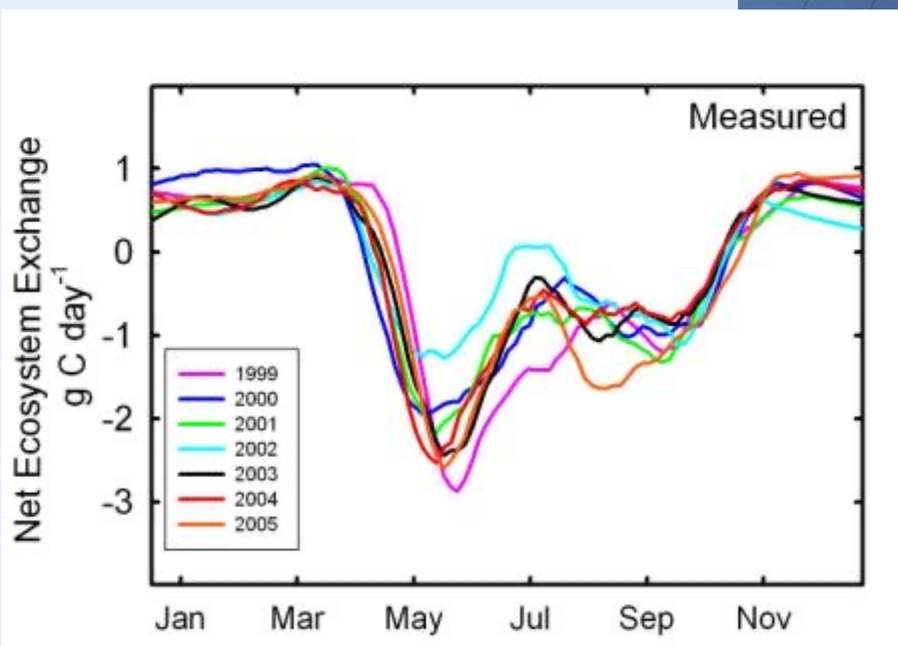
Driven by 8 climate variables
(1) average air temperature,
(2) average soil temperature
(3) Precipitation
(4) PAR
(5) atmospheric vapor pressure
(6) atmospheric vapor pressure deficit
(7) vapor pressure deficit between the soil and the atmosphere
(8) wind speed

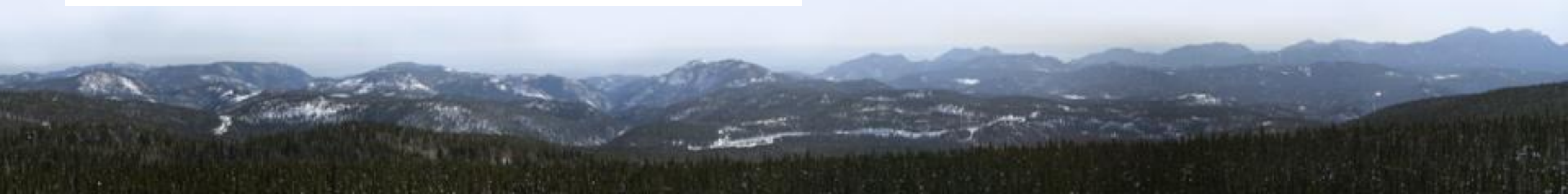
Section 1

OPTIMIZING FLUXES

Niwot Ridge, CO





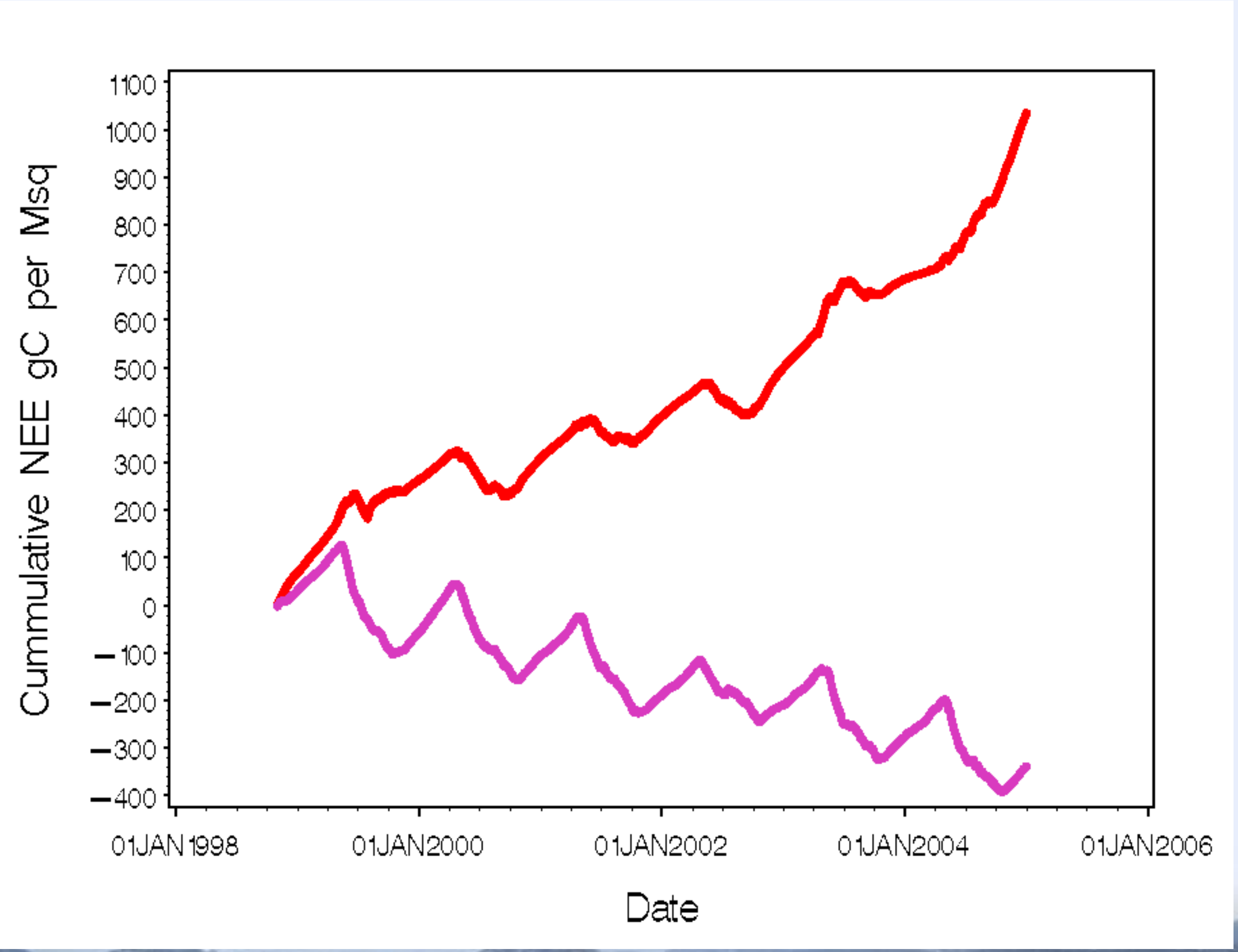


Parameterization*

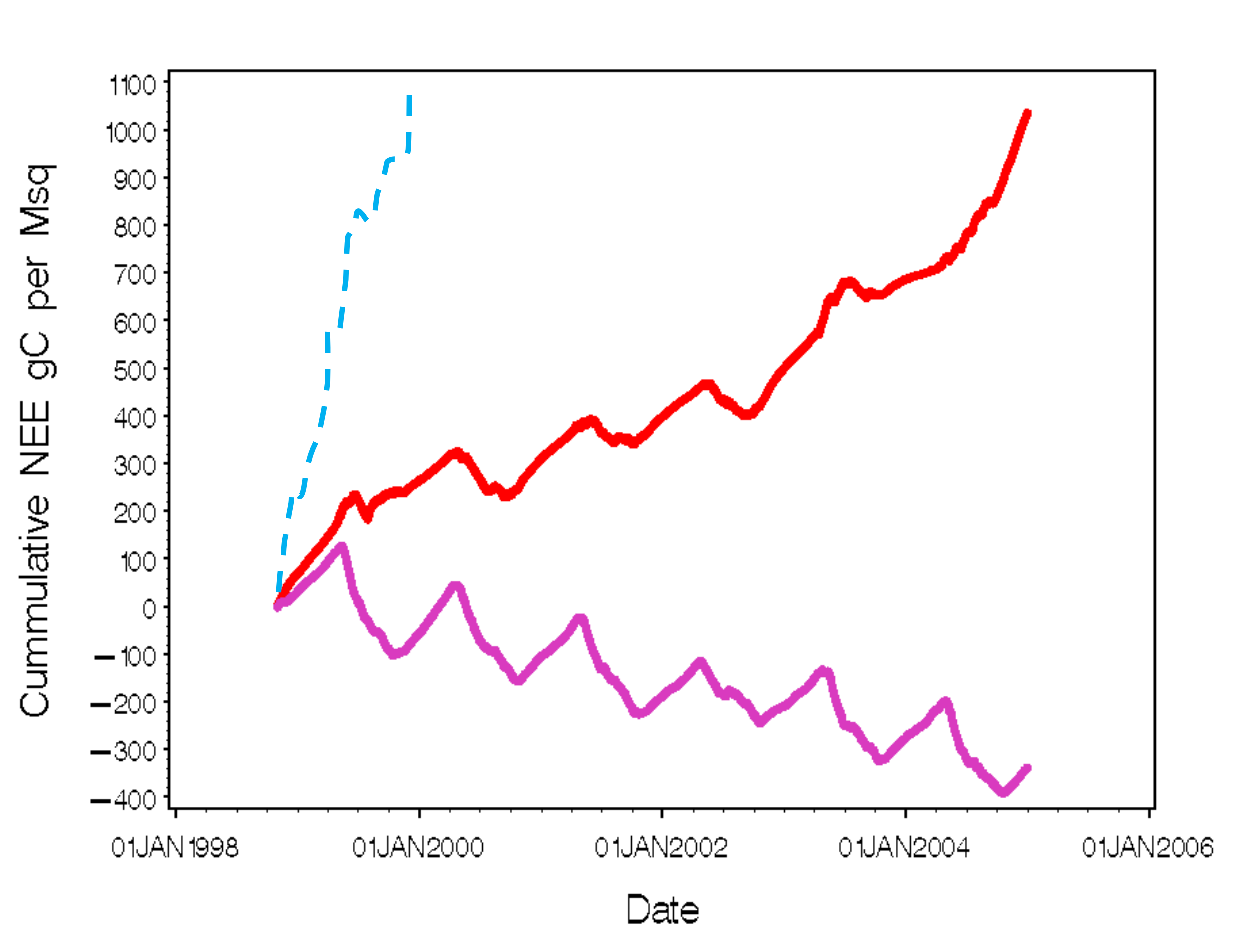
- First guess parameters
  - Diligent field work
  - Long hours of field work
  - ~~Guess work~~
  - Wisdom drawn from long experience at working at a site

*We'll come back to this later

Graph default SIPNET output plus observed fluxes



Graph default SIPNET output plus observed fluxes



For reference – an Audi V8

SIPNET at Niwot Ridge

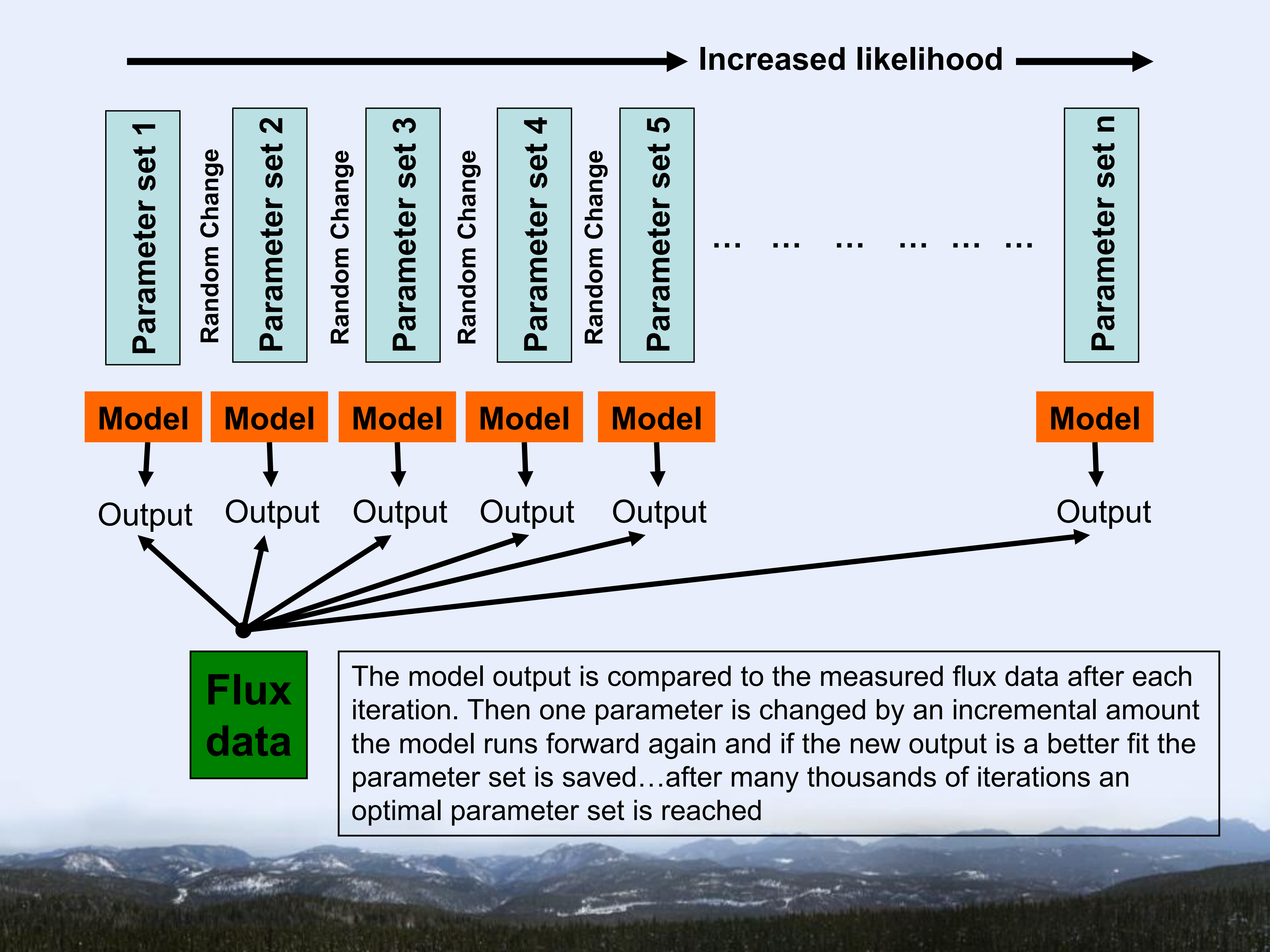
WHAT HAPPENS WHEN WE ASSIMILATE NEE ESTIMATES FROM THE TOWER?

Model-data error is defined in terms of likelihood (L), and minimizing this error is like maximizing the likelihood:

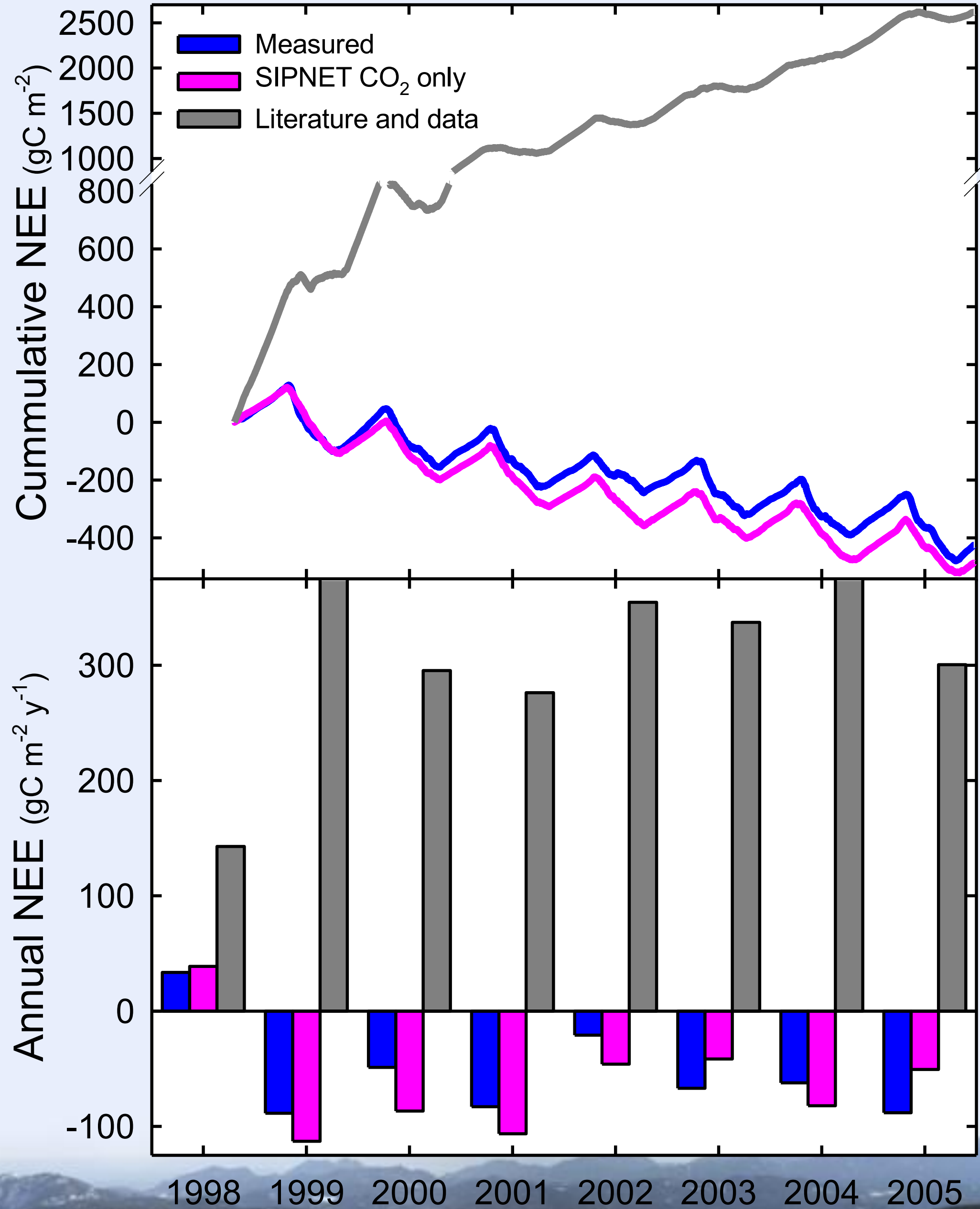
$$L = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma}} e^{-(x_{\mathrm{i}} - \mu_{\mathrm{i}})^2 / 2\sigma^2}$$

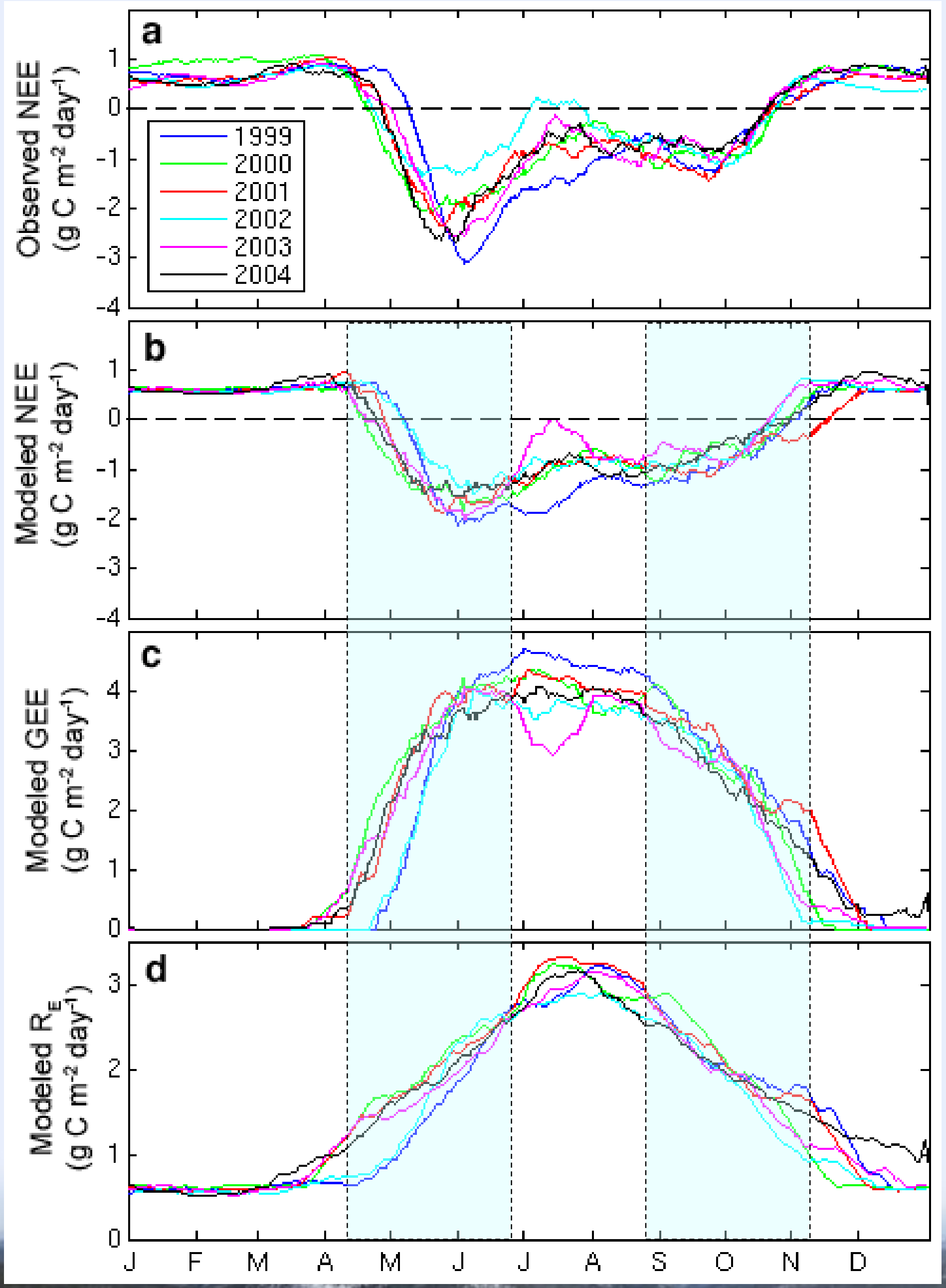
where n is the number of data points and is the standard deviation on <u>each</u> data point.

The smaller the model residual the better
The larger the number of points the better


Increased likelihood
Parameter set 1
Random Change
Parameter set 2
Random Change
Parameter set 3
Random Change
Parameter set 4
Random Change
Parameter set 5
… … … … … …
Parameter set n
Model
Model
Model
Model
Model
Model
Output
Output
Output
Output
Output
Output
Flux data
The model output is compared to the measured flux data after each iteration. Then one parameter is changed by an incremental amount the model runs forward again and if the new output is a better fit the parameter set is saved…after many thousands of iterations an optimal parameter set is reached

Large improvement in the model's ability to represent measured fluxes





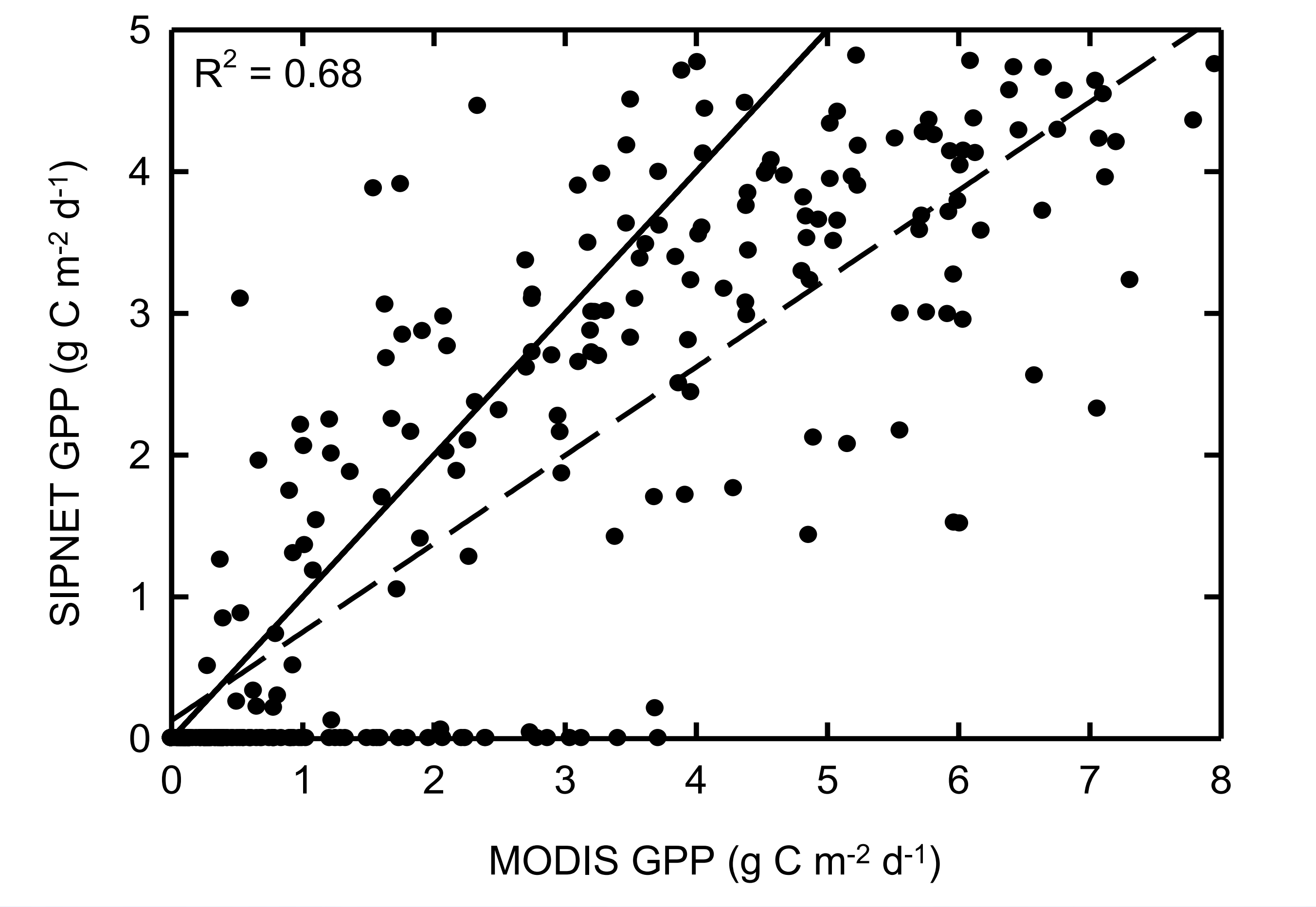
Observed and modeled NEE and components of NEE for the six-year observation period

The observed NEE was taken from the Niwot Ridge eddy flux record. The modeled NEE, GEE and R_E were derived from the SIPNET model conditioned on the entire six year record of the eddy fluxes (from Sacks et al. 2007)

MODIS at Niwot Ridge

SCALING ECOSYSTEM PROCESSES WITH SATELLITES

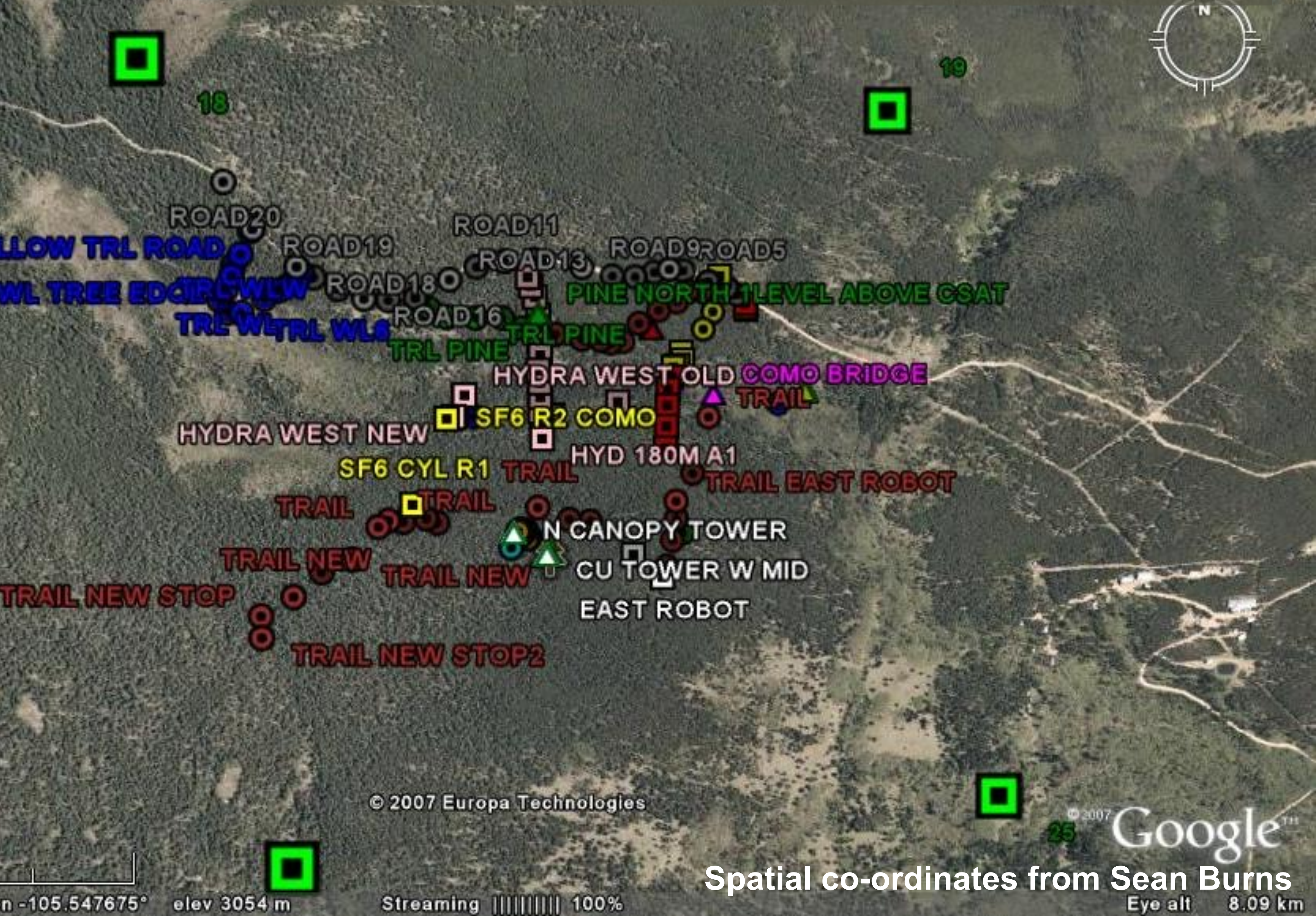
GPP estimated from the SIPNET model conditioned on tower fluxes and GPP estimated using MODIS



Moore and Monson (2008 unpublished)

MODIS Moderate Resolution Imaging Spectroradiometer
Subset pixels for Niwot Ridge, CO

MODIS Subset pixels for Niwot Ridge, CO

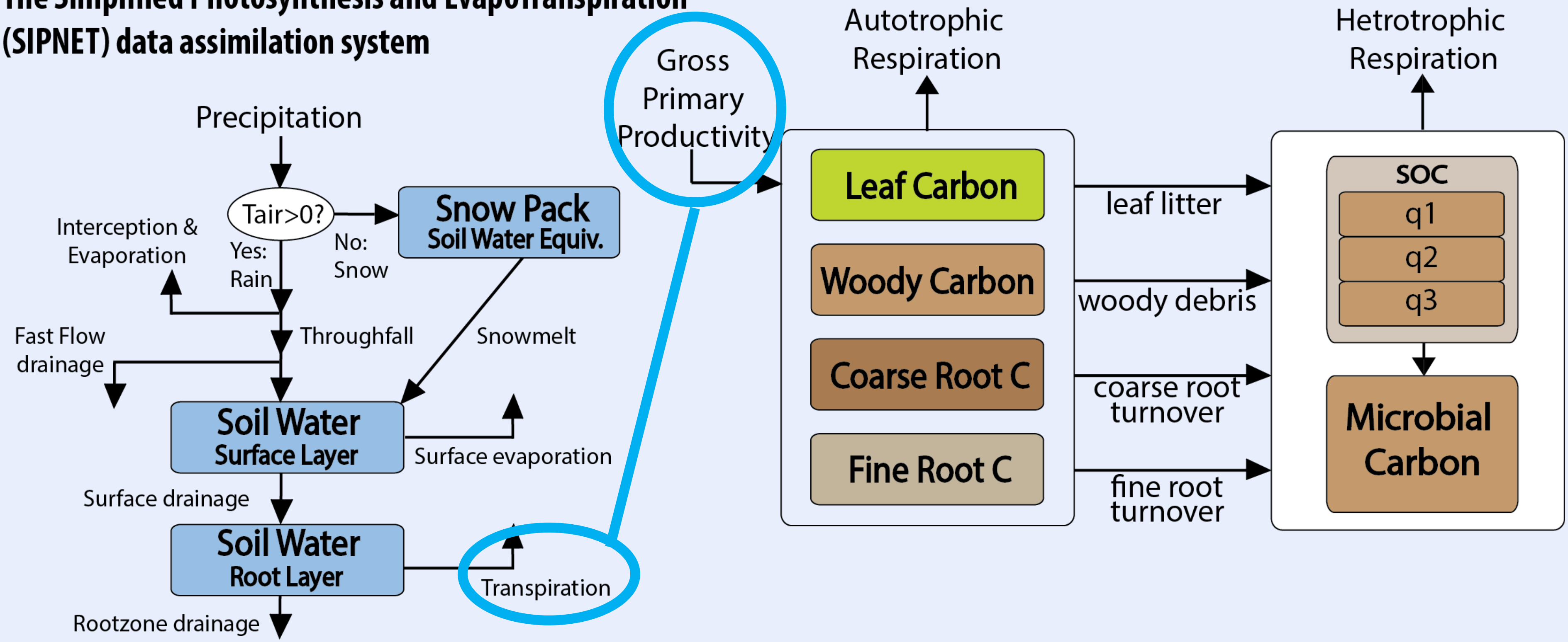


Spatial co-ordinates from Sean Burns

SIPNET at Niwot Ridge

WHAT HAPPENS WHEN WE ASSIMILATE NEE AND ET ESTIMATES FROM THE TOWER?

The Simplfied Photosynthesis and EvapoTranspiration (SIPNET) data assimilation system



Driven by 8 climate variables

(1) average air temperature,

(2) average soil temperature

(3) Precipitation

(4) PAR

(5) atmospheric vapor pressure

(6) atmospheric vapor pressure deficit

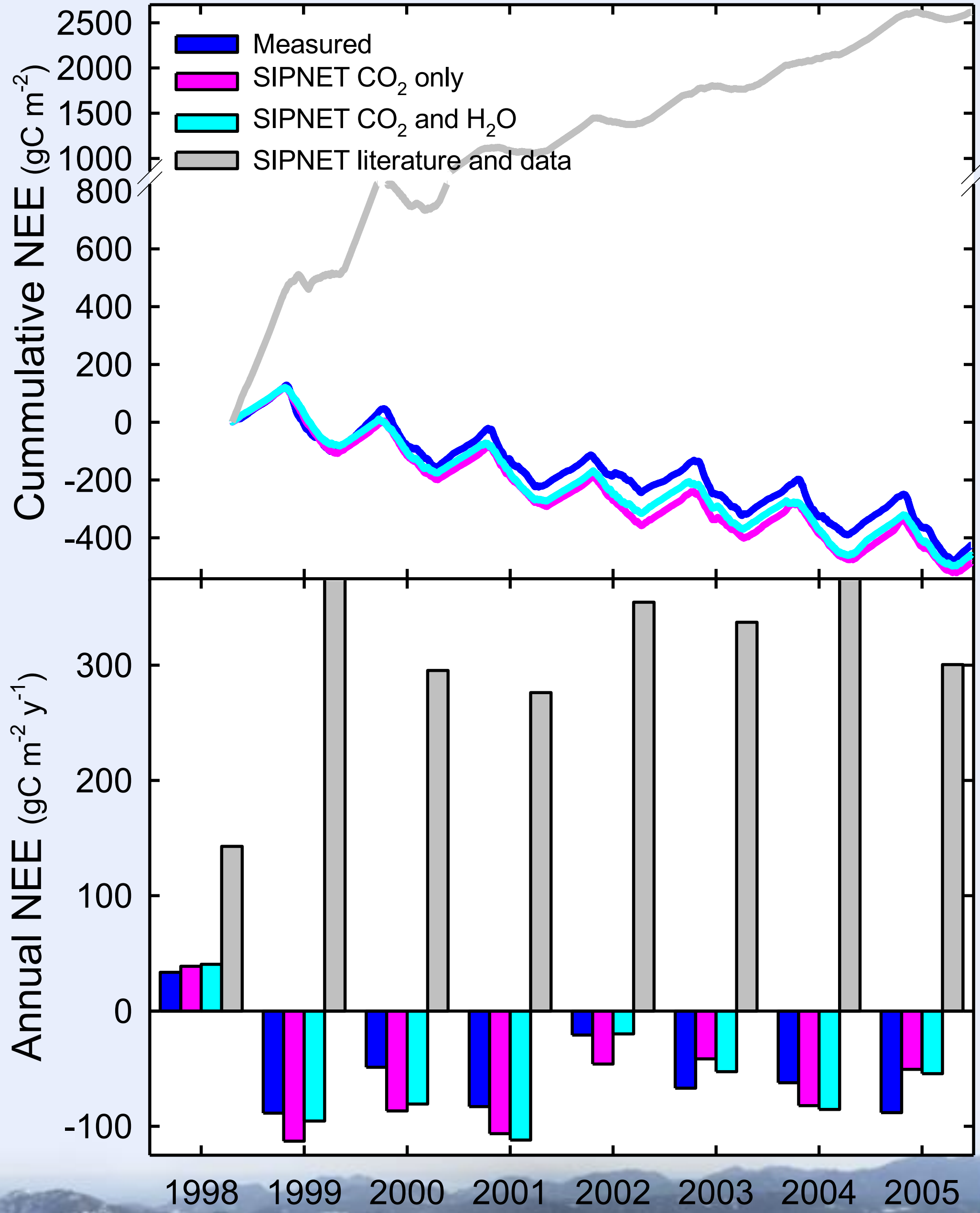
(7) vapor pressure deficit between the soil and the atmosphere

(8) wind speed

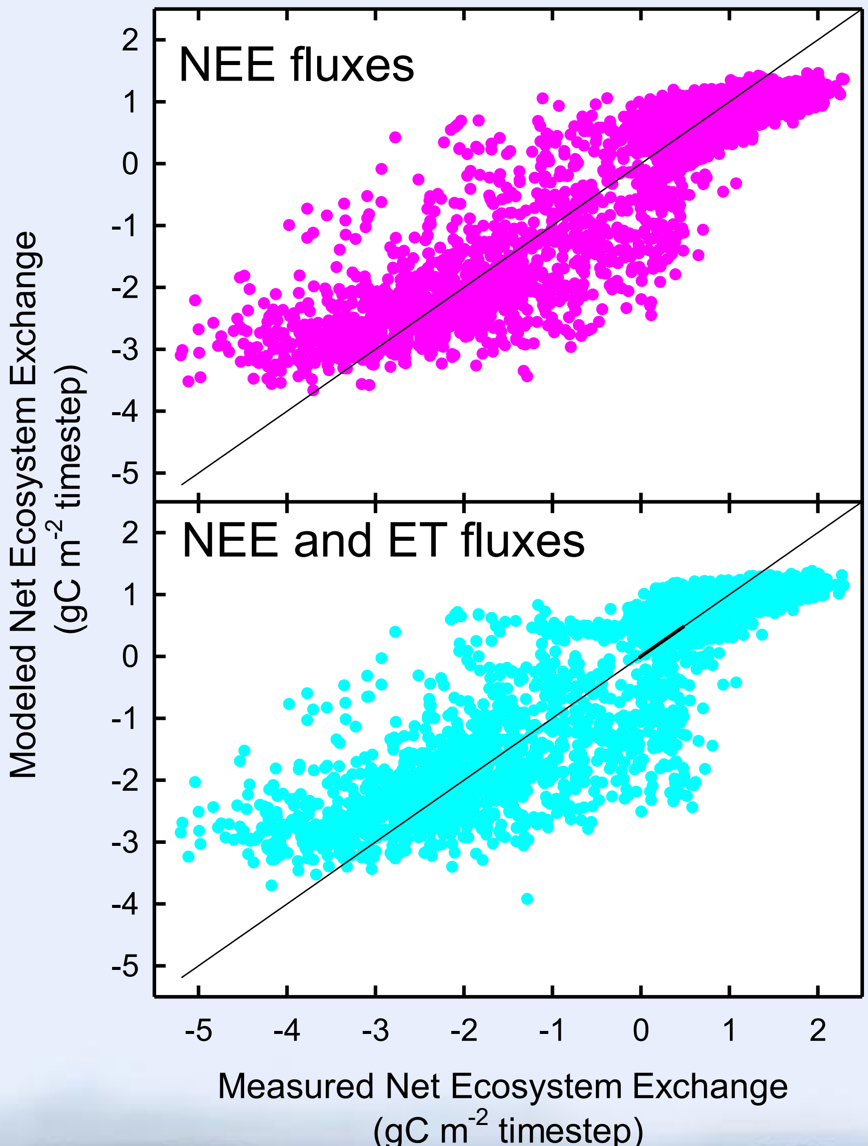
SIPNET, driven by climate data, can replicate the measured NEE fluxes

- Estimating parameters using literature based values yields poor estimates of NEE

- Using data assimilation to estimate model parameters results in NEE estimates which closely match the measured fluxes



Moore et al 2008

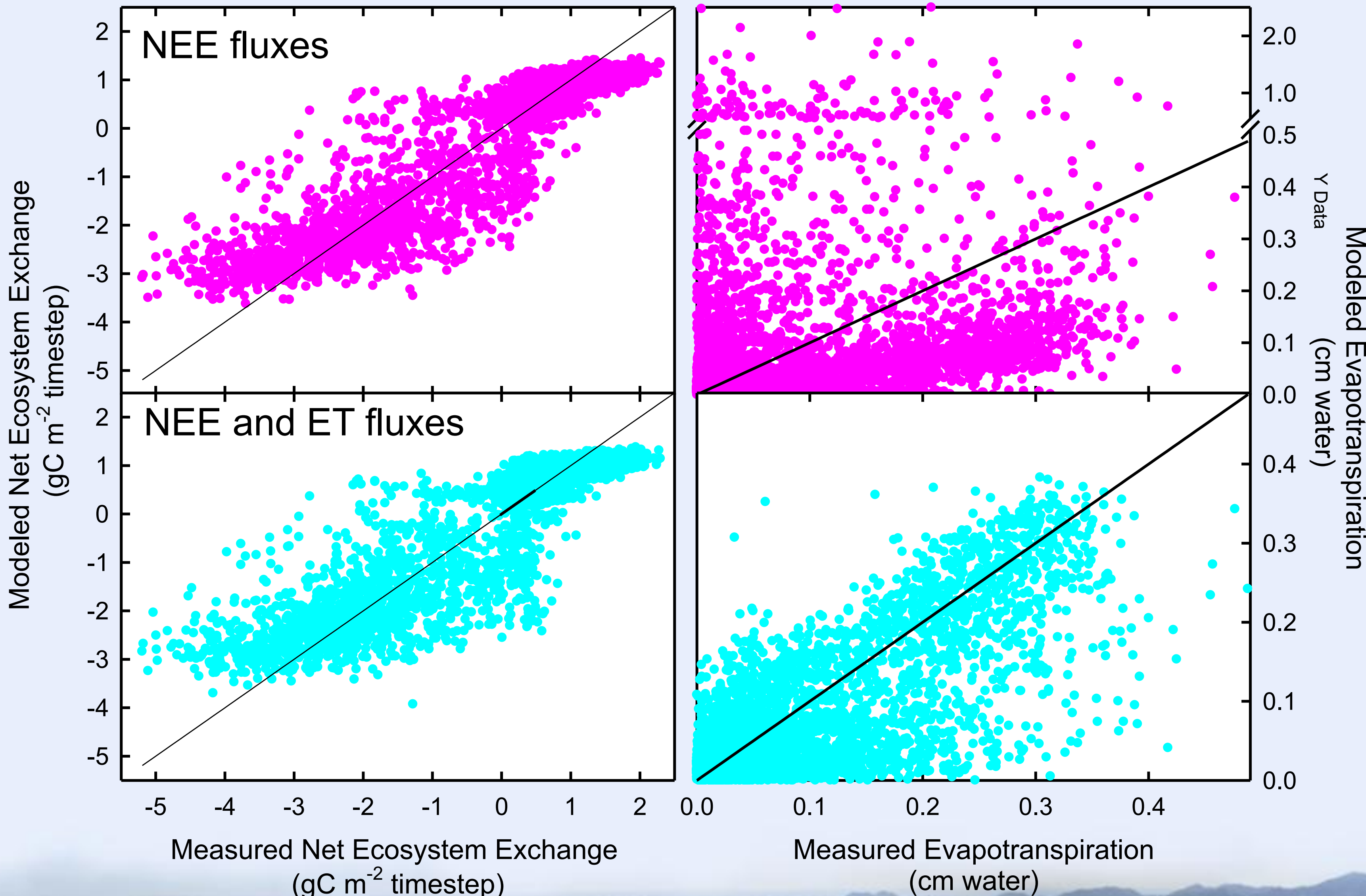


Modelled fluxes fall along the one to one line

Assimilating CO_2 and ET together seems to make NO Difference to the NEE flux.

Both parameter sets result in scatter and both fail at the extremes

Moore et al 2008



Moore et al 2008

We need to be careful

If we push the model around sometimes we cause other part of the model to behave incorrectly

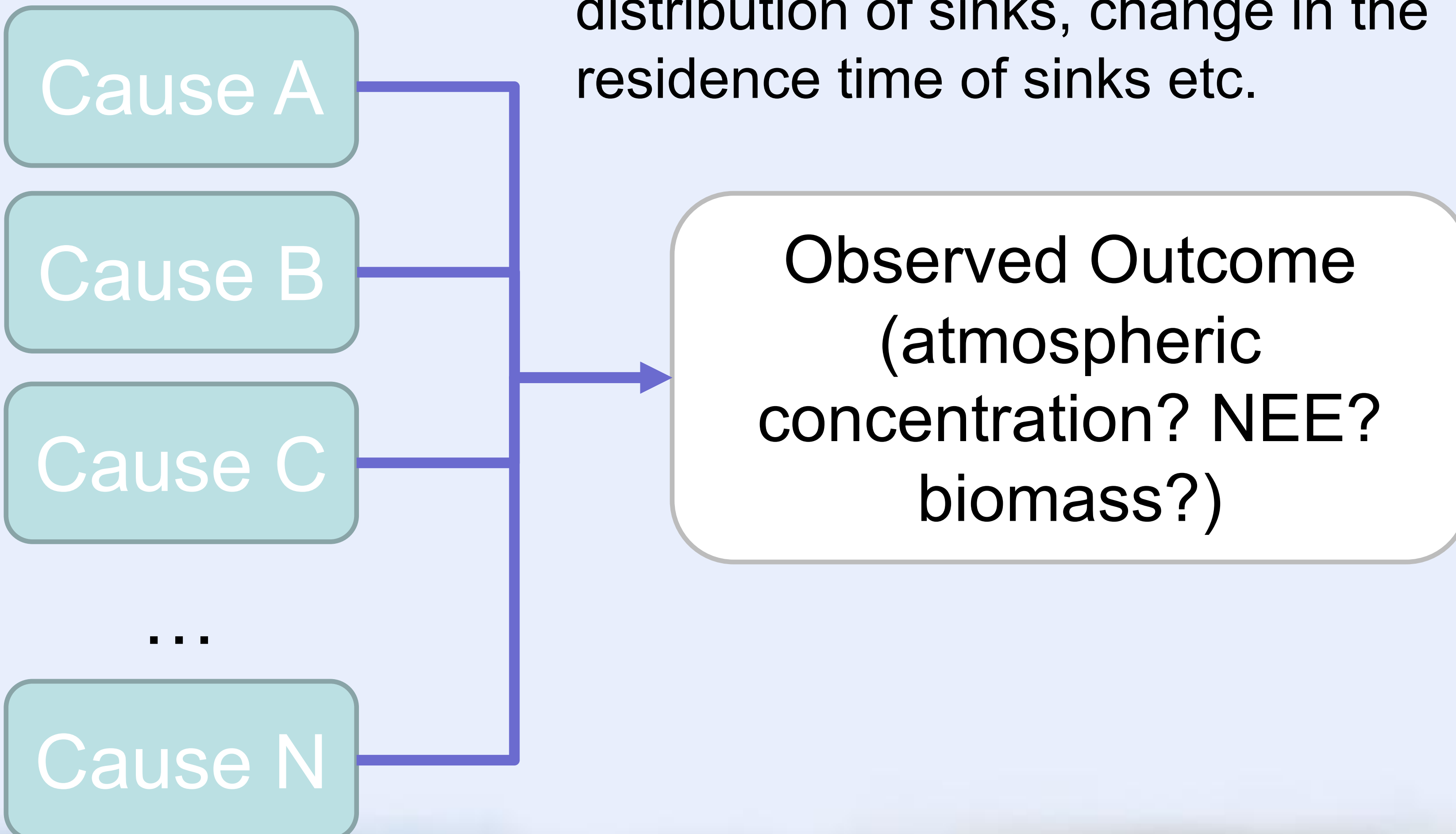
- Choose carefully which processes you are interested in

- Choose carefully which parameters you can constrain

- Make very careful measurements (garbage in = garbage out)

Equifinality

There are many ways to explain a change in flux – increased uptake, reduced release, change in the distribution of sinks, change in the residence time of sinks etc.

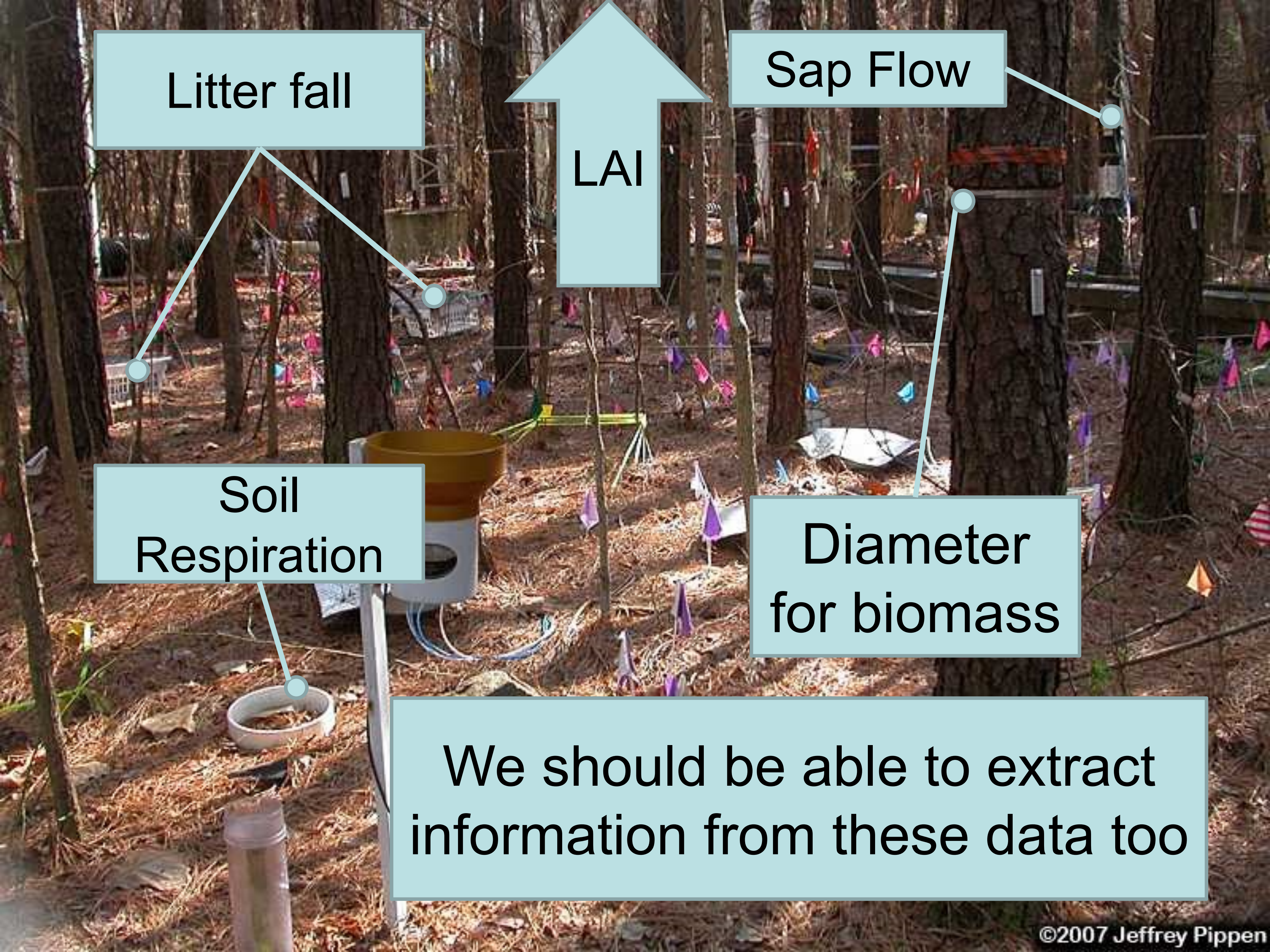


SIPNET at Niwot Ridge

HOW CAN WE CHECK TO SEE IF OUR ASSIMILATION MODELS ARE A GOOD REPRESENTATION?

Trust…but verify

Data used to tune a model cannot be used to validate the same model.



Litter fall
LAI
Sap Flow
Soil Respiration
Diameter for biomass
We should be able to extract information from these data too
©2007 Jeffrey Pippen

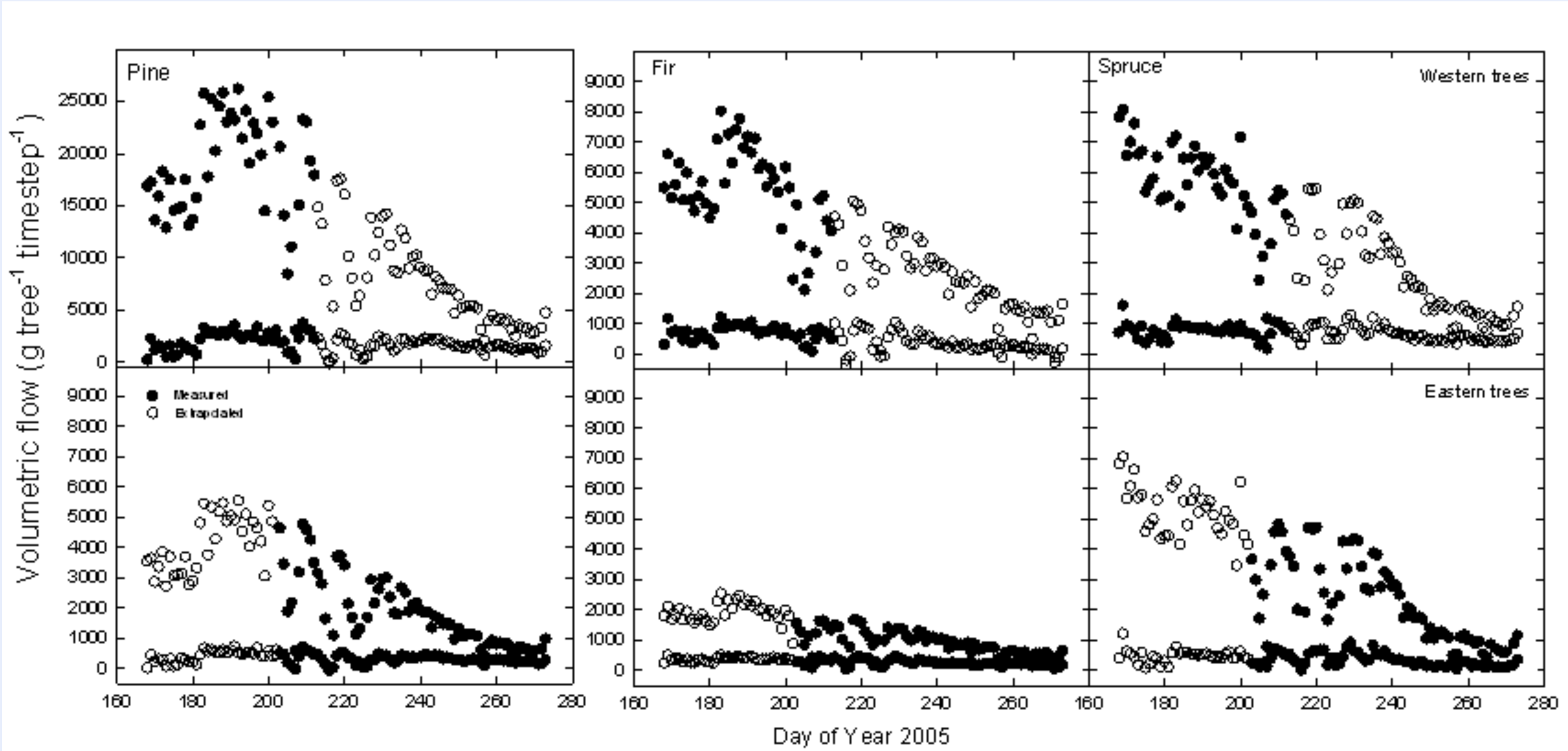
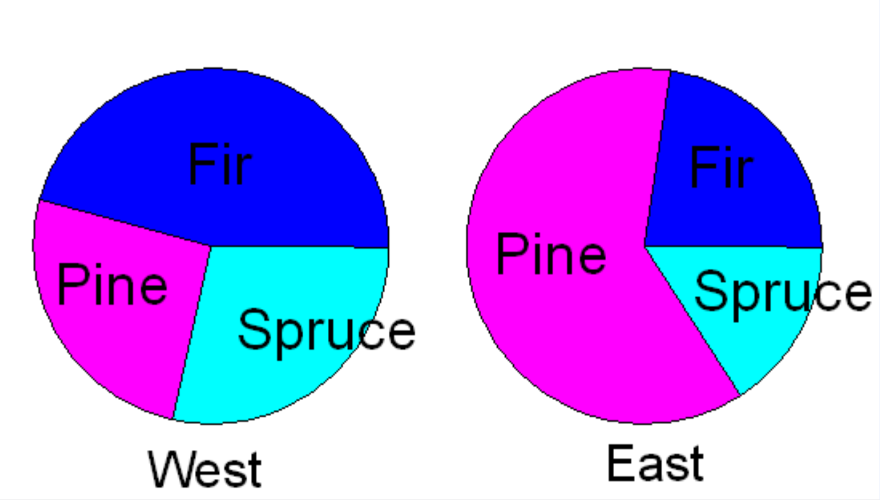
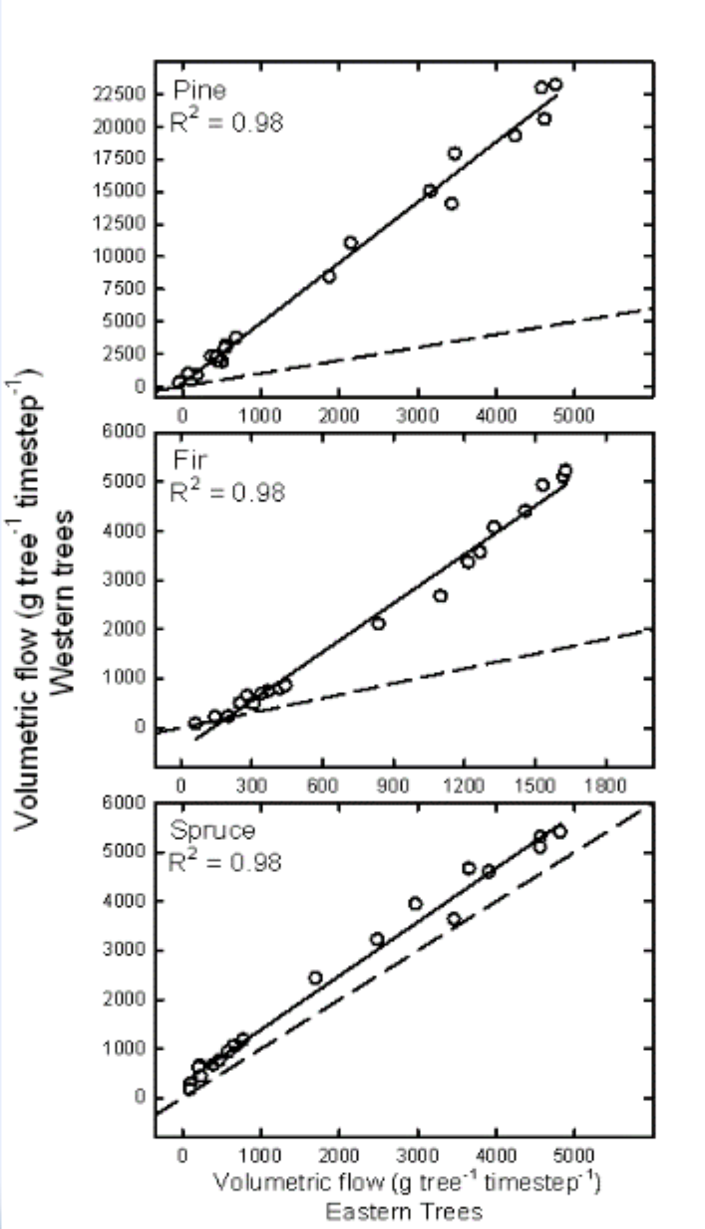


Fig: Volumetic sap flow of Pine, Fir and Spruce trees in Western (upper panals) and Eastern (lower panals) portions of the Niwot Ridge experimental forest. The sap flow values are separated into night and day to allow comparison with SIPNET model output. The points along the base of each graph represent night time transpiration. Closed symbols represent the average total volumetric flux for the time step, open symbols represent data calculated from the above linear regressions. ***Note Pines in the West are on a different scale***

Data: Jia Hu

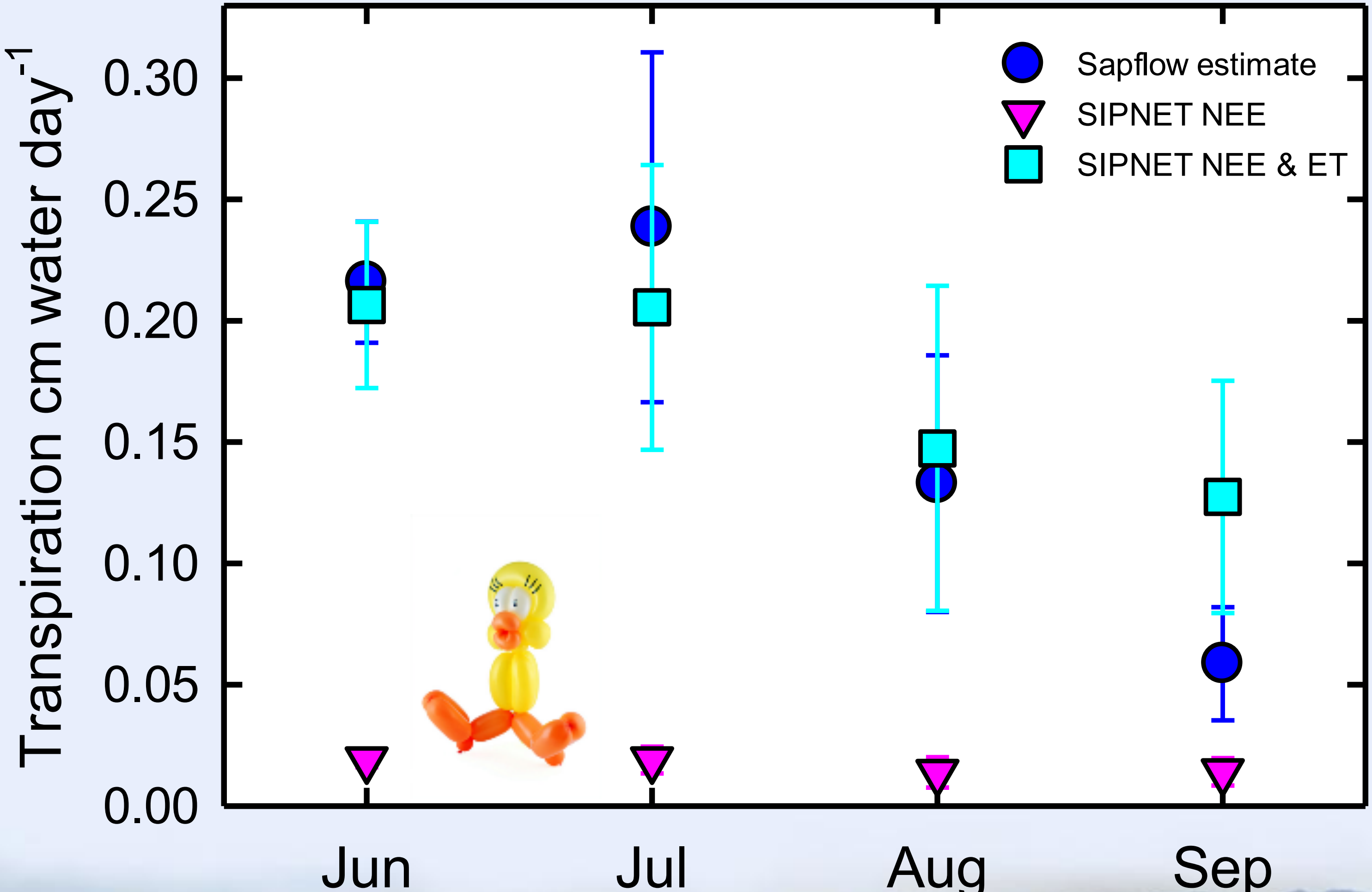
Volumetric sap flow summed by time step (day and night) in the eastern and western portions of the Niwot Ridge experimental forest. Solid lines represent predicted values based on linear regression of Western vs Eastern flows for each species. Dashed lines show the one to one relationship. ***Note Pines in the West are on a different scale***



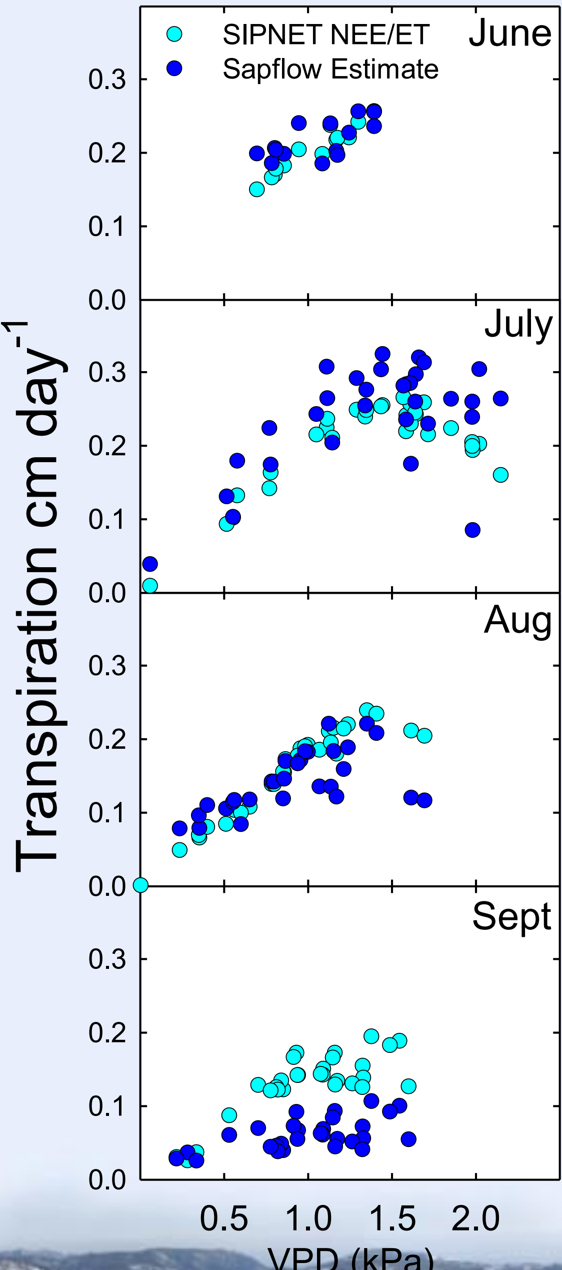


Data: Jia Hu

Using only NEE to parameterize the model gives VERY poor estimates of Transpiration



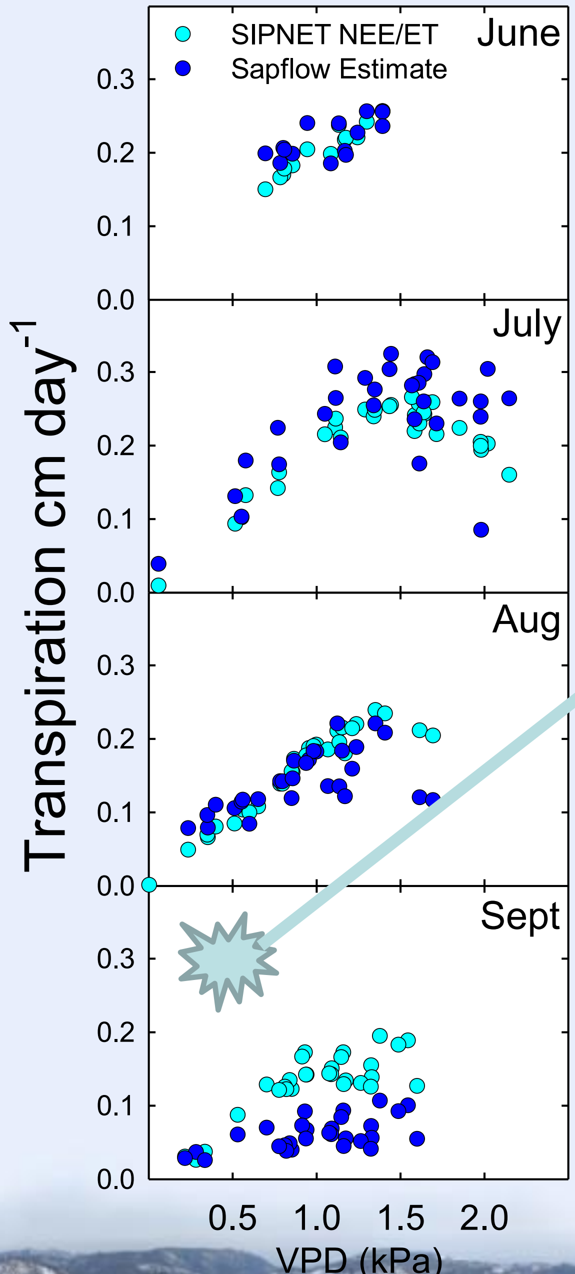
Moore et al 2008

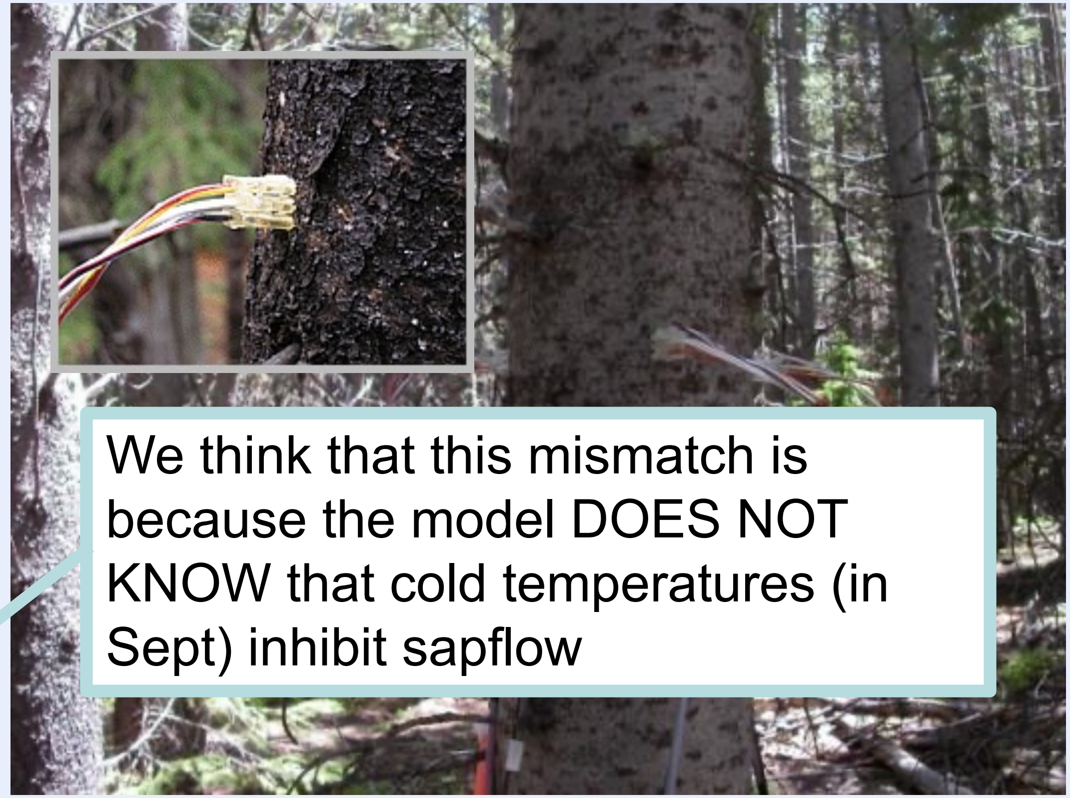


Optimized Model reproduces measured transpiration
This could be used to predict water use in different temperature and precipitation regimes

Moore et al 2008

Moore et al. (2008) Ag. Forest Met.



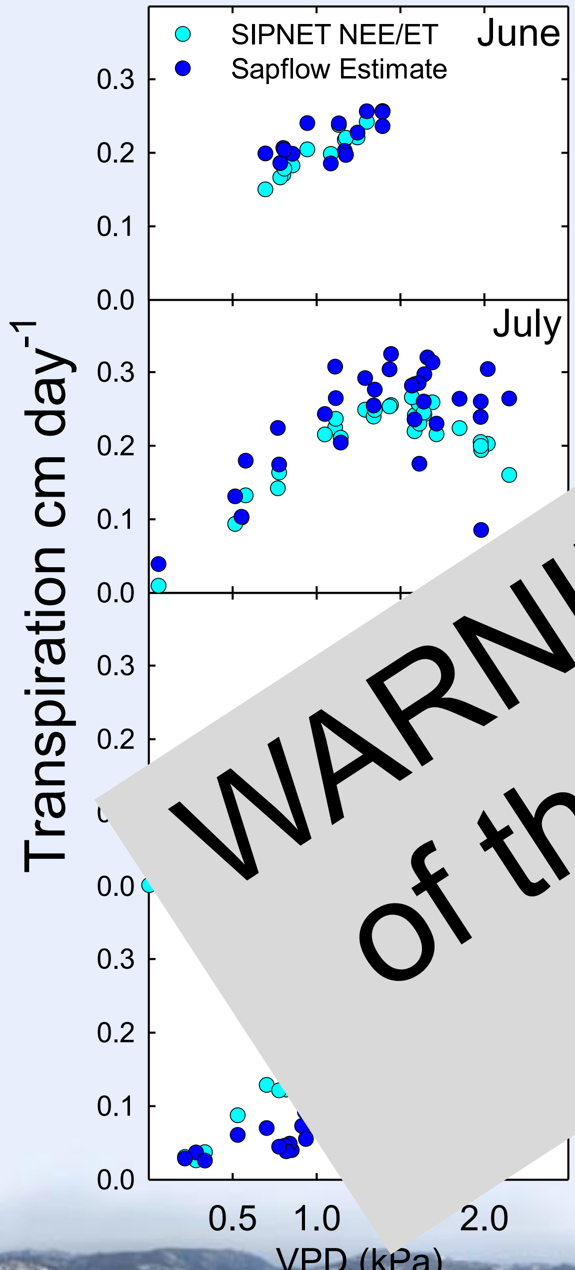


We think that this mismatch is because the model DOES NOT KNOW that cold temperatures (in Sept) inhibit sapflow

Optimized Model reproduces measured transpiration

This could be used to predict water use in different temperature and precipitation regimes

Moore et al. (2008) Ag. Forest Met.

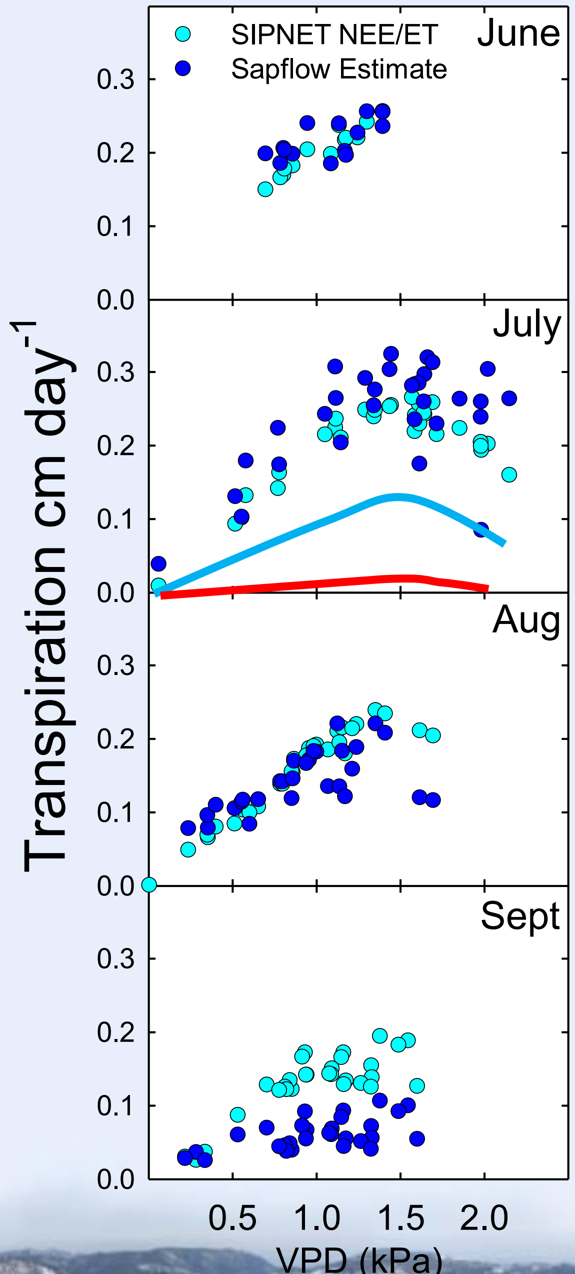


WARNING !!!!! Neither of these estimates is correct!

...ized Model reproduces measured transpiration

...his could be used to predict water use in different temperature and precipitation regimes

Moore et al. (2008) Ag. Forest Met.



Optimized Model reproduces measured transpiration

This could be used to predict water use in different temperature and precipitation regimes

Moore et al. (2008) Ag. Forest Met.

Ray Leunning always said that a flux scientist should "Know thy site"

For the same reason "Know thy model"

How are the carbon and water cycles linked in SIPNET?

```
if (potGrossPsn < TINY) { // avoid divide by 0
    *trans = 0.0; // no photosynthesis -> no transpiration
    *dWater = 1; // dWater doesn't matter, since we don't have any photosynthesis
  }

  else {
    wue = params.wueConst/vpd;
    potTrans = potGrossPsn/wue * 1000.0 * (44.0/12.0) * (1.0/10000.0);
    // 1000 converts g to mg; 44/12 converts g C to g CO2, 1/10000 converts m^2 to cm^2

    removableWater = soilWater * params.waterRemoveFrac;
    if (climate->tsoil < params.frozenSoilThreshold) // frozen soil - less or no water available
      removableWater *= params.frozenSoilEff; /* frozen soil effect: fraction of water available if soil is
       frozen
 (assume amt. of water avail. w/ frozen soil scales linearly with amt. of
 water avail. in thawed soil) */
    if (removableWater >= potTrans)
      *trans = potTrans;
    else
      *trans = removableWater;
```

How are the carbon and water cycles linked in SIPNET?

Water use efficiency is calculated using an estimated constant modified by Vapor Pressure Deficit

wue = params.wueConst/vpd

potTrans = potGrossPsn/wue

dWater = Trans/potTrans

*gpp = potGrossPsn * dWater;

How are the carbon and water cycles linked in SIPNET?

wue = params.wueConst/vpd

$\boxed{\text{potTrans = potGrossPsn/wue}}$

dWater = Trans/potTrans

*gpp = pot

Potential transpiration is calculated as the ratio of Potential Gross Photosynthesis and Water Use Efficiency

How are the carbon and water cycles linked in SIPNET?

If there is enough water Transpiration is the same as potential Transpiration… if water is limiting Transpiration is reduced accordingly the ratio dWater is a measure of this reduction

wue = params.wueConst/vpd

potTrans = potGrossPsn/wue

$\boxed{\text{dWater = Trans/potTrans}}$

*gpp = potGrossPsn * dWater;

How are the carbon and water cycles linked in SIPNET?

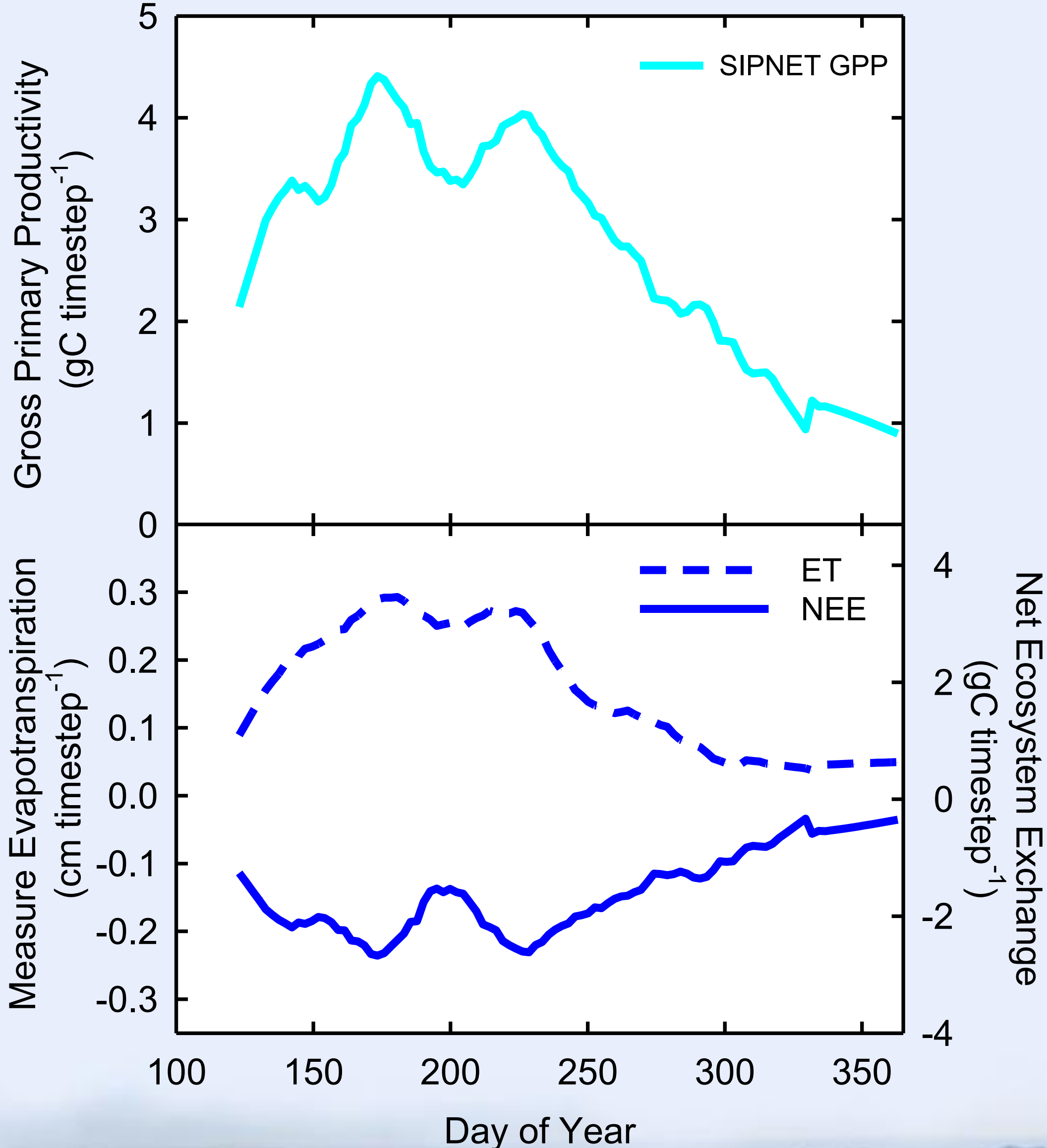
wue = params.wueConst/vpd

potTrans = potGrossPsn/wue

dWater = Trans/potTrans

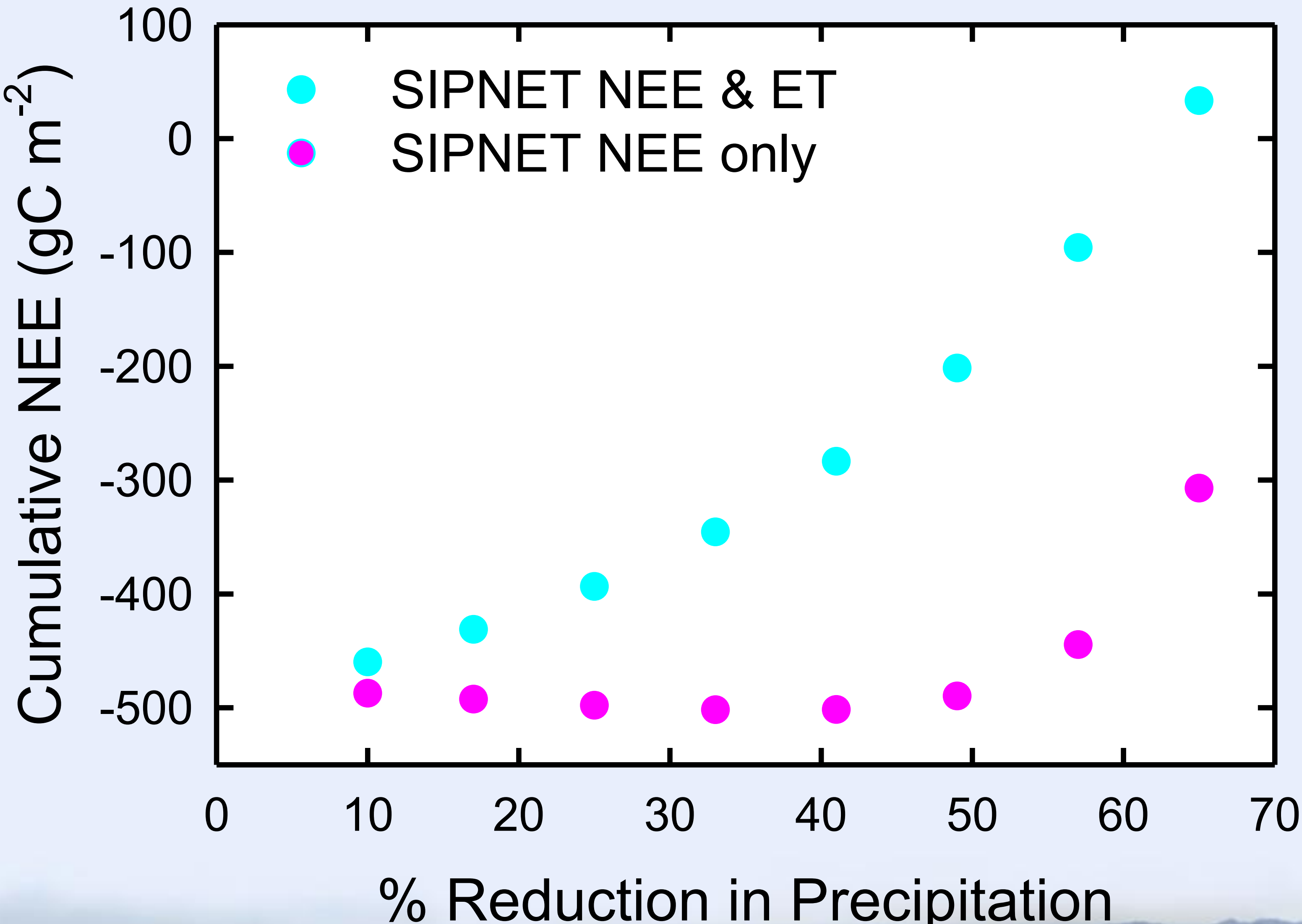
*gpp = potGrossPsn * dWater;

GPP is calculated as Potential Gross Photosynthesis modified by the ratio of potential transpiration to actual transpiration (i.e. GPP is reduced if there is insufficient soil water)



The optimization alters parameters dealing with Water Use Efficiency and Canopy resistance to increase it's estimate of transpiration at the expense of evapo-transpiration to mirror GPP.

Moore et al. (2008) Ag. Forest Met.



Moore et al. (2008) Ag. Forest Met.

Conclusion

- ~~Using NEE and ET gives me a correct estimate of Transpiration~~
- There is INFORMATION in the ET data which can tell us something about Transpiration.

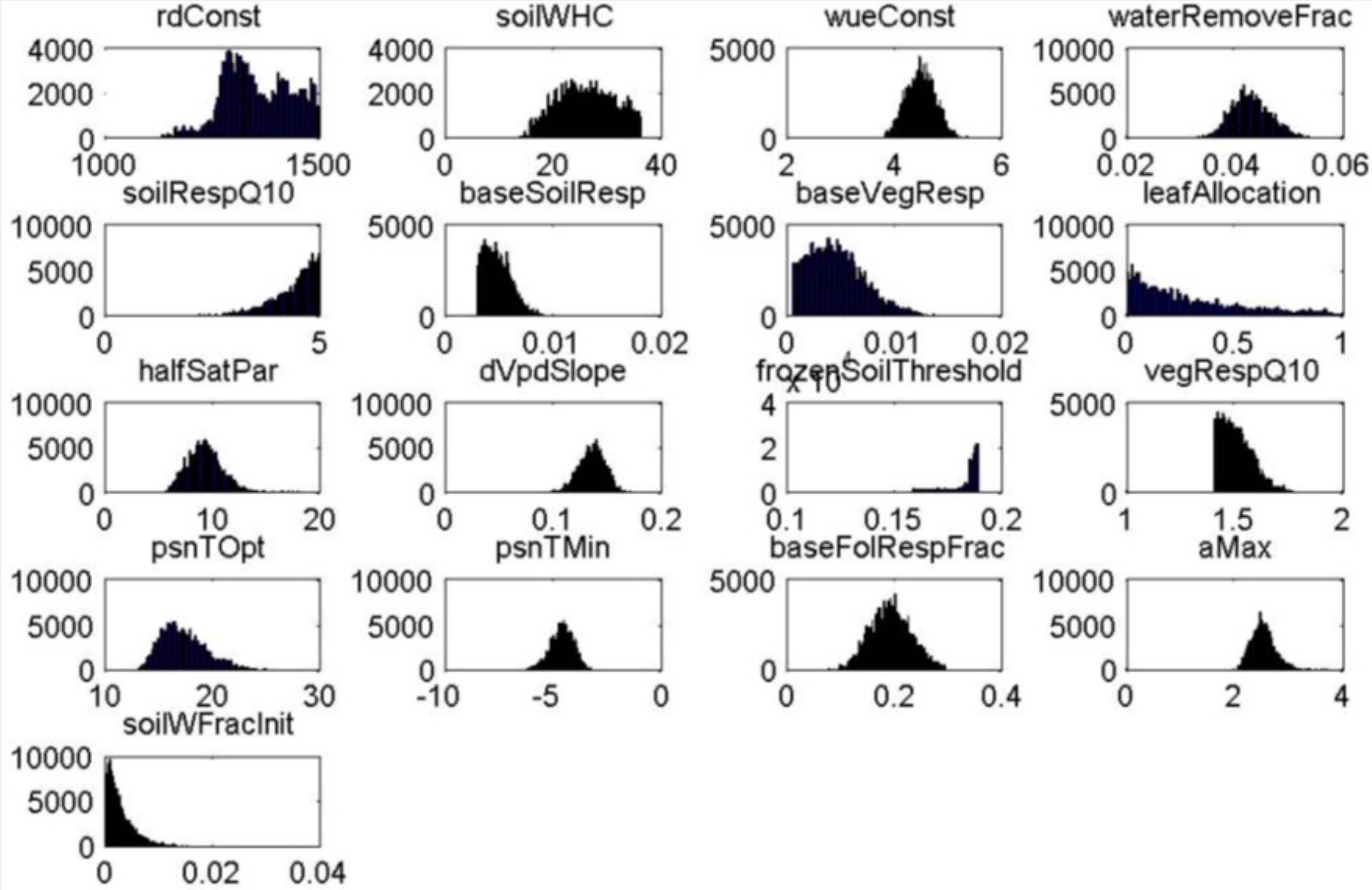
Section 2

ESTIMATING PARAMETERS AND INITIAL STATES

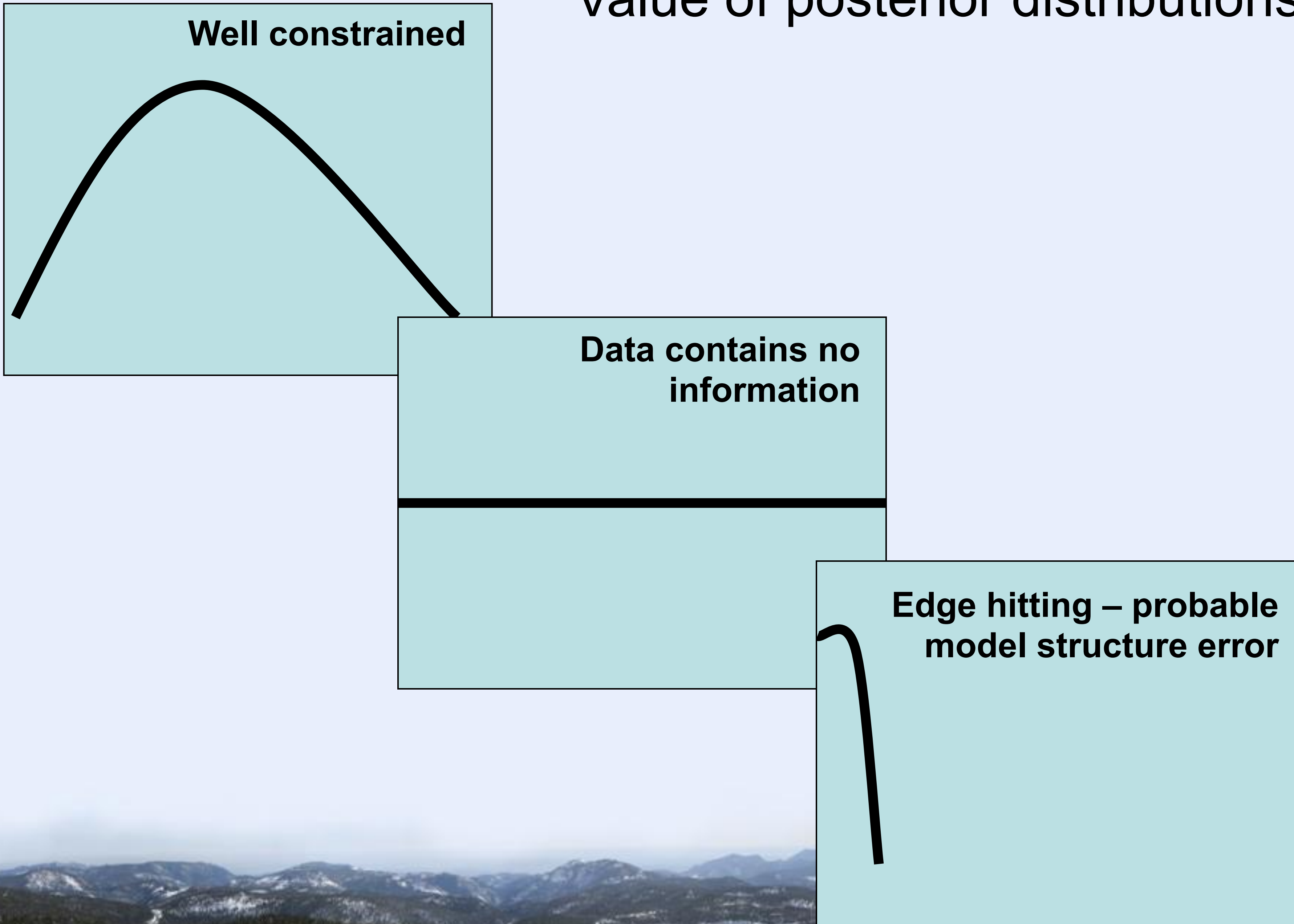
Table1: SIPNET parameters and initial conditions that are allowed to vary in the optimization, and their allowable ranges. The ranges assume a uniform prior distribution.

Symbol	Definition	Range
Initial Pool Values:		
$W_{S,0}$	Initial soil moisture content (fraction of $W_{S,c}$)	0 – 1
Photosynthesis/Respiration Parameters:		
A_{max}	Maximum net CO_2 assimilation rate (nmol CO_2 g^{-1} (leaf biomass) s^{-1})	0 – 34
K_F	Foliar maintenance respiration as fraction of A_{max} (no units)	0.05 – 0.30
T_{min}	Minimum temperature for photosynthesis (°C)	-8 – 8
T_{opt}	Optimum temperature for photosynthesis (°C)	5 – 30
$Q10_V$	Vegetation respiration Q_{10} (no units)	1.4 – 2.6
T_s	Soil temperature at which photosynthesis and foliar respiration are shut down (°C)	-5 – 5
K_{VPD}	Slope of VPD-photosynthesis relationship (kPa^{-1})	0.01 – 0.25
$PPFD_{1/2}$	Half saturation point of PPFD-photosynthesis relationship (mol m^{-2} day^{-1})	4 – 27
NPP_L	Fraction of NPP allocated to leaf growth (no units)	0 – 1
K_A	Wood respiration rate at 0°C (g C g^{-1} C yr^{-1})	0.0006 – 0.06
K_H	Soil respiration rate at 0°C and moisture-saturated soil (g C g^{-1} C yr^{-1})	0.003 – 0.6
$Q10_S$	Soil respiration Q_{10} (no units)	1.4 – 5
Moisture Parameters:		
f	Fraction of soil water removable in one day (no units)	0.001 – 0.16
K_{WUE}	VPD-water use efficiency relationship (mg CO_2 kPa g^{-1} H_2O)	0.01 – 109
$W_{S,c}$	Soil water holding capacity (cm (precipitation equivalent))	0.1 – 36
R_d	Scalar relating aerodynamic resistance to wind speed (no units)[a]	1 – 1500

Symbol	Description	units	Optimized Parameter		Allowed Range	
			CO	CW	low	upp
soilWFracInit	Initial soil moisture content	fraction of Soil Water Content	**0.39**	**0.80**	0	1
aMax	Maximum net CO2 assimilation rate	nmol CO_2 g^{-1} leaf biomass s^{-1}	4.74	4.94	0	34
baseFolRespFrac	Foliar maintenance respiration as a fraction of A_{max}	-	0.10	0.13	0.05	0.3
psnTMin	Minimum temperature for photosynthesis	^oC	*-2.91*	*-3.64*	-8	8
psnTOpt	Optimal temperature for photosynthesis	^oC	*14.59*	*18.75*	5	30
vegRespQ10	Vegetation Respiration Q_{10}	-	1.45	1.41	1.4	2.6
frozenSoilThreshold	Soil temperature at which photosynthesis and foliar respiration are shut down	^oC	0.02	0.02	-5	5
dVpdSlope	Slope of VPD-photosynthesis relationship	kPa^{-1}	0.12	0.15	0.01	0.25
halfSatPar	PAR at which photosynthesis is half A_{max}	E m^{-2} d^{-1}	7.34	8.17	4	27
leafAllocation	Fraction of mean NPP allocated to leaves	-	0.42	0.52	0	1
baseVegResp	Wood respiration rate at 0^oC	gC g^{-1} PlantC d^{-1}	0.03	0.03	0.0006	0.06
baseSoilResp	Wood respiration rate at 0^oC without moisture stress	gC g^{-1} soilC d^{-1}	0.01	0.00	0.003	0.6
soilRespQ10	Soil Respiration Q_{10}	-	5.00	4.69	1.4	5
waterRemoveFrac	Fraction of water removable in a timestep	-	0.04	0.05	0.001	0.16
wueConst	VPD-water use efficiency relationship	(mg CO_2 kPa g^{-1} H_2O)	**85**	**8**	0.01	109
soilWHC	Soil water holding capacity	cm water equivalent	**4.19**	**17.90**	0.1	36
rdConst	Scalar relating aerodynamic resistance to wind speed		**37.17**	**1467**	1	1500


rdConst
4000
2000
0
1000
1500
soilWHC
4000
2000
0
0
20
40
wueConst
5000
0
2
4
6
waterRemoveFrac
10000
5000
0
0.02
0.04
0.06
soilRespQ10
10000
5000
0
0
5
baseSoilResp
5000
0
0
0.01
0.02
baseVegResp
5000
0
0
0.01
0.02
leafAllocation
10000
5000
0
0
0.5
1
halfSatPar
10000
5000
0
0
10
20
dVpdSlope
10000
5000
0
0
0.1
0.2
frozenSoilThreshold
x 10
4
2
0
0.1
0.15
0.2
vegRespQ10
5000
0
1
1.5
2
psnTOpt
10000
5000
0
10
20
30
psnTMin
10000
5000
0
-10
-5
0
baseFolRespFrac
5000
0
0
0.2
0.4
aMax
10000
5000
0
0
2
4
soilWFracInit
10000
5000
0
0
0.02
0.04

Value of posterior distributions



SIPNET at Niwot

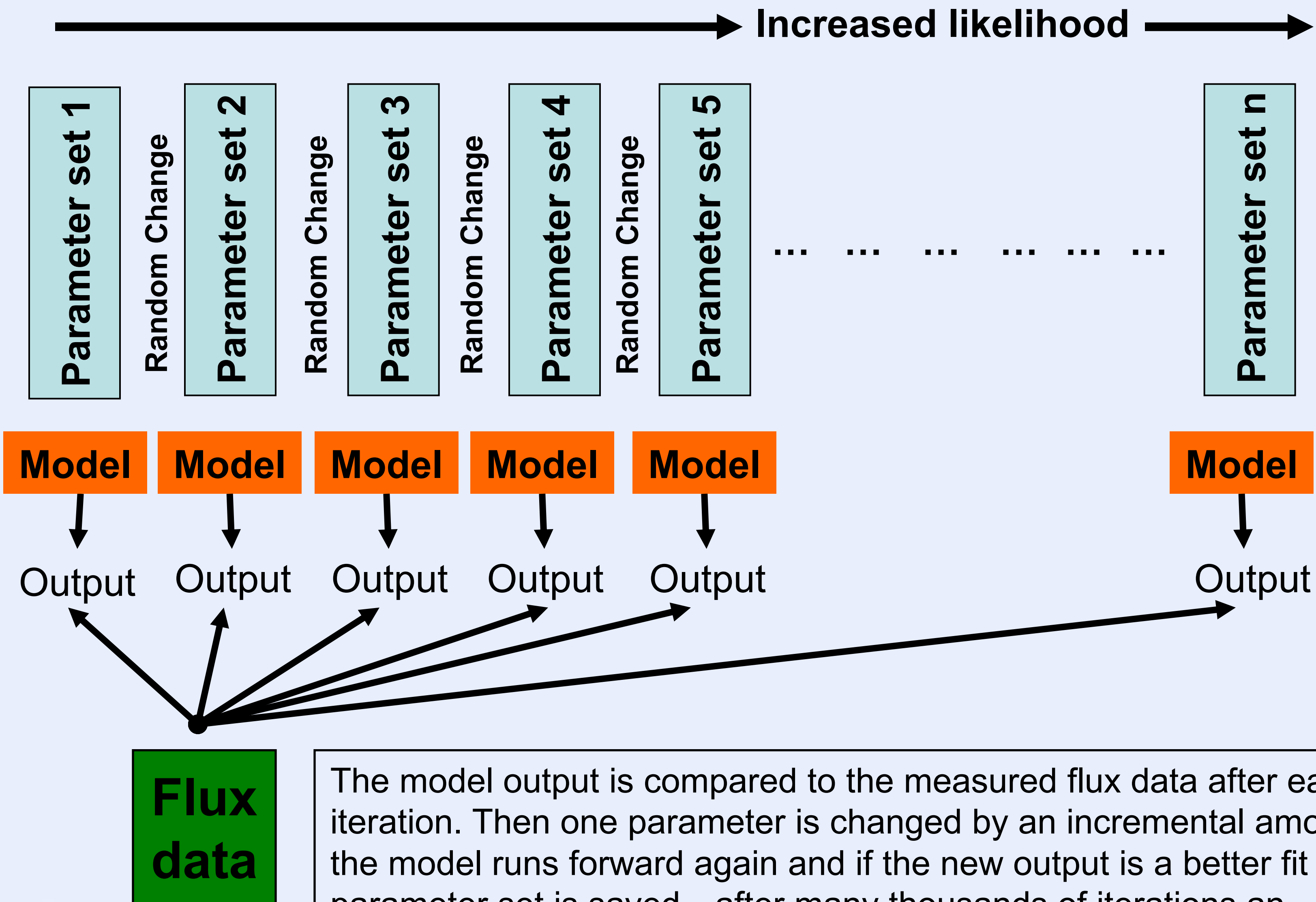
USING AN OPTIMIZED MODEL TO TEST MODEL STRUCTURE

How do we know there’s a problem?

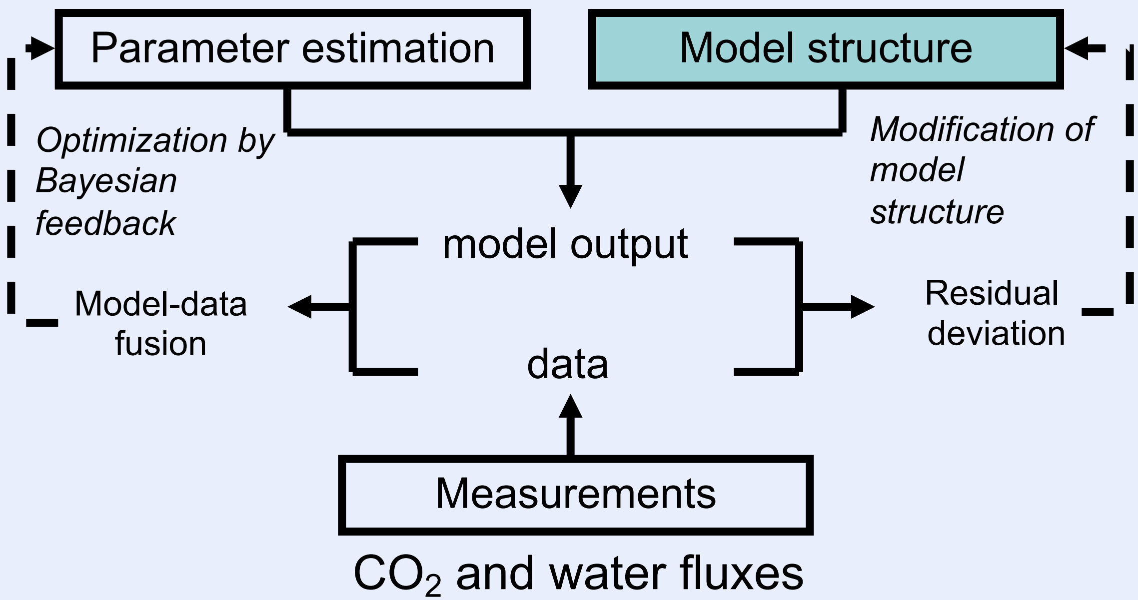
- Parameters cannot be optimized effectively (edge hitting parameters)
- Pattern to the mismatch between model and data.
  - Does the pattern of residuals look like another process?

Variation in model structure

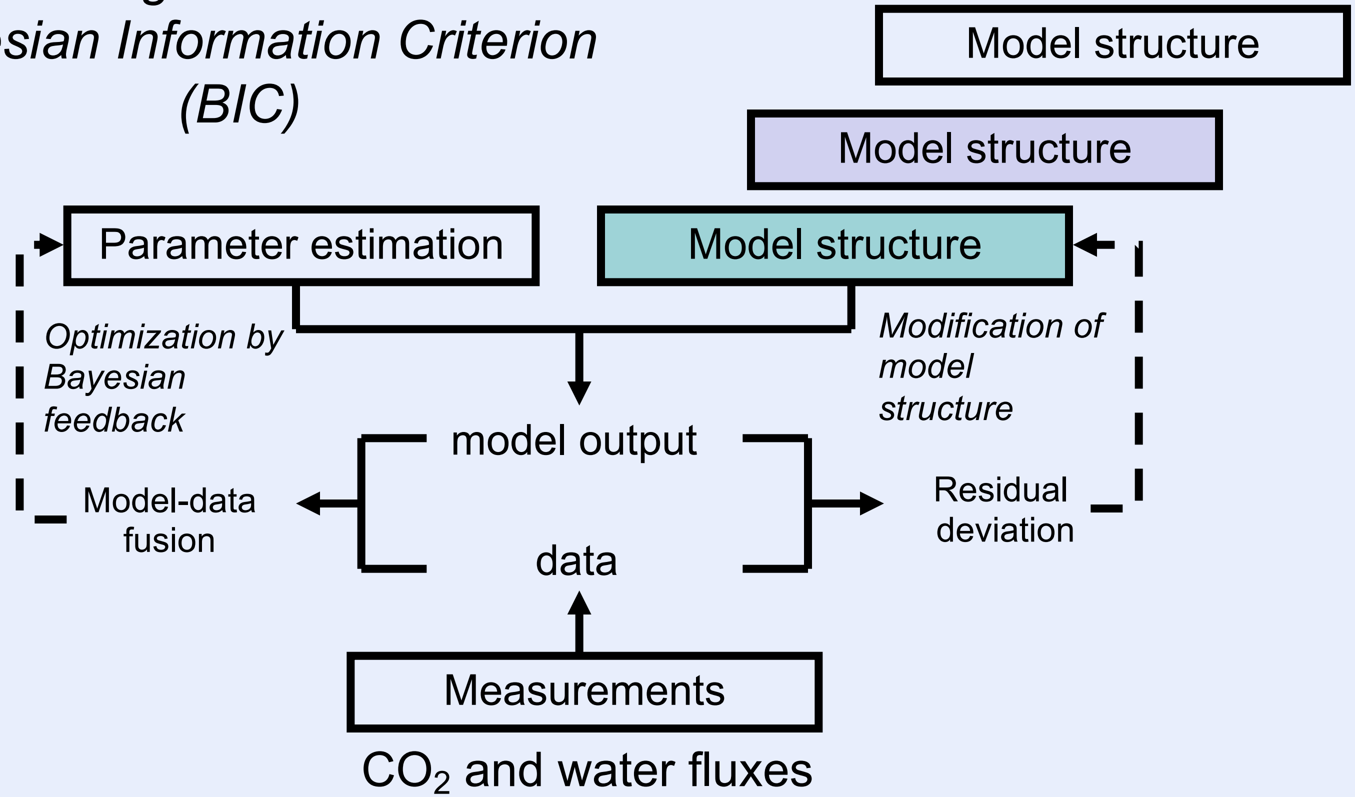
- What’s the best way to model a process or a set of processes?
- Phenological methods (Harvard Forest Braswell et al 2005 – Richardson et al various)
- Variations in how respiration is modelled (Niwot Ridge, Sacks 2006, 2007)
- Below ground carbon cycling (Niwot Ridge, Zobitz et al 2008)



The model output is compared to the measured flux data after each iteration. Then one parameter is changed by an incremental amount the model runs forward again and if the new output is a better fit the parameter set is saved…after many thousands of iterations an optimal parameter set is reached


Parameter estimation
Model structure
Optimization by Bayesian feedback
Modification of model structure
model output
data
Model-data fusion
Residual deviation
Measurements
CO_2 and water fluxes

Compare using information criterion.
Bayesian Information Criterion (BIC)



BIC (Bayesian Information Criterion) = -2 . LL + K . ln (n), where LL is the log likelihood, K is the number of free parameters, and n is the number of data points used in optimization

BIC (Bayesian Information Criterion) =

$$-2 . LL + K . \ln (n)$$

where LL is the log likelihood, K is the number of free parameters, and n is the number of data points used in optimization

Smaller is better!

Fewer number of *(free)* parameters is better

Fewer points is better

	Base model	No winter-time shutdown of psn., foliar resp.	**Seasonal R_H**	Add'l litter pool	Moisture-independent R_H
Best log likelihood[a]	-2404.2	-2614.7	**-2374.0**	-2407.6	-2415.7
RMS error[b]	0.555	0.597	**0.550**	0.556	0.558
# free parameters	32	31	**35**	35	32
BIC[c]	5063.4	5476.5	**5027.0**	5094.1	5086.4

Model-data comparison statistics from running five versions of SIPNET using the best parameter set retrieved from the optimization on each model. See text for description of model variations.

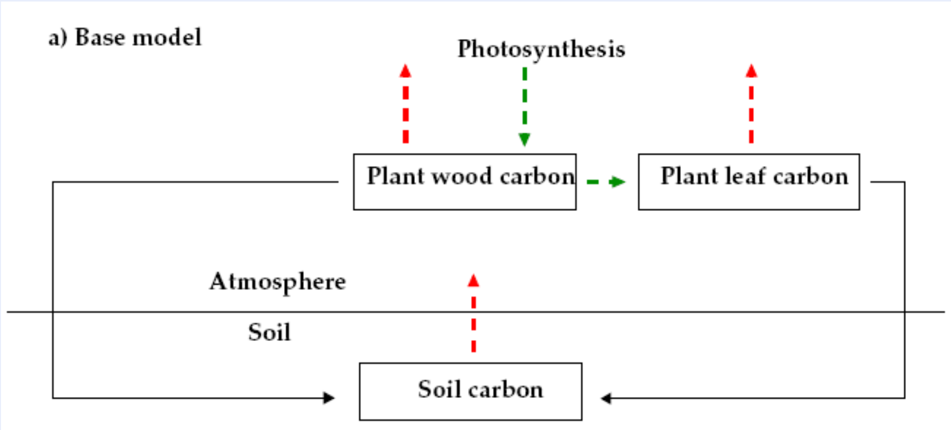
(a) Larger (i.e. closer to zero) numbers mean greater likelihood.

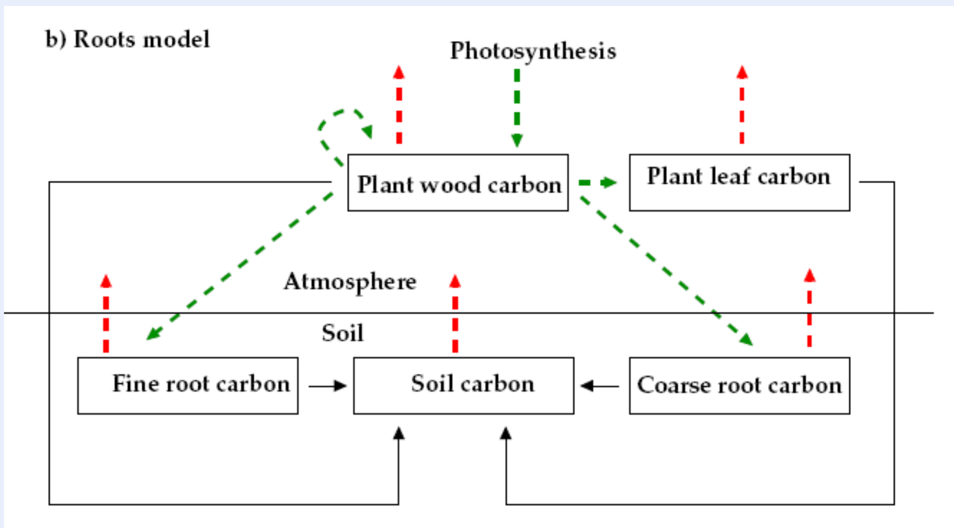
(b) Root mean square error in g C m-2 over a single time step.

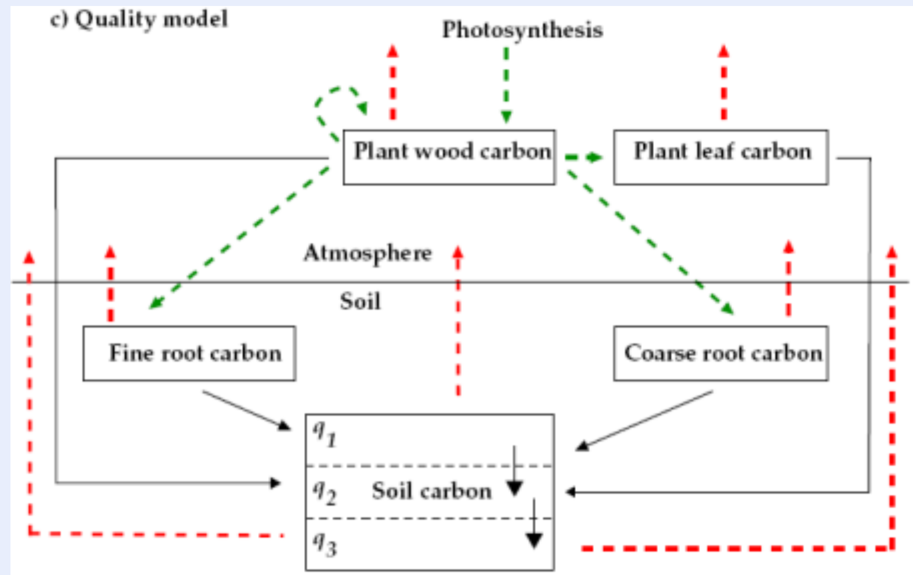
(c) BIC (Bayesian Information Criterion) = -2 . LL + K . ln (n), where LL is the log likelihood, K is the number of free parameters, and n is the number of data points used in optimization (2894). A lower BIC indicates a model with greater support from the data.

Sacks et al 2006 Global Change Biology, 12: 240-259

Three model structures for dealing with below ground C cycling

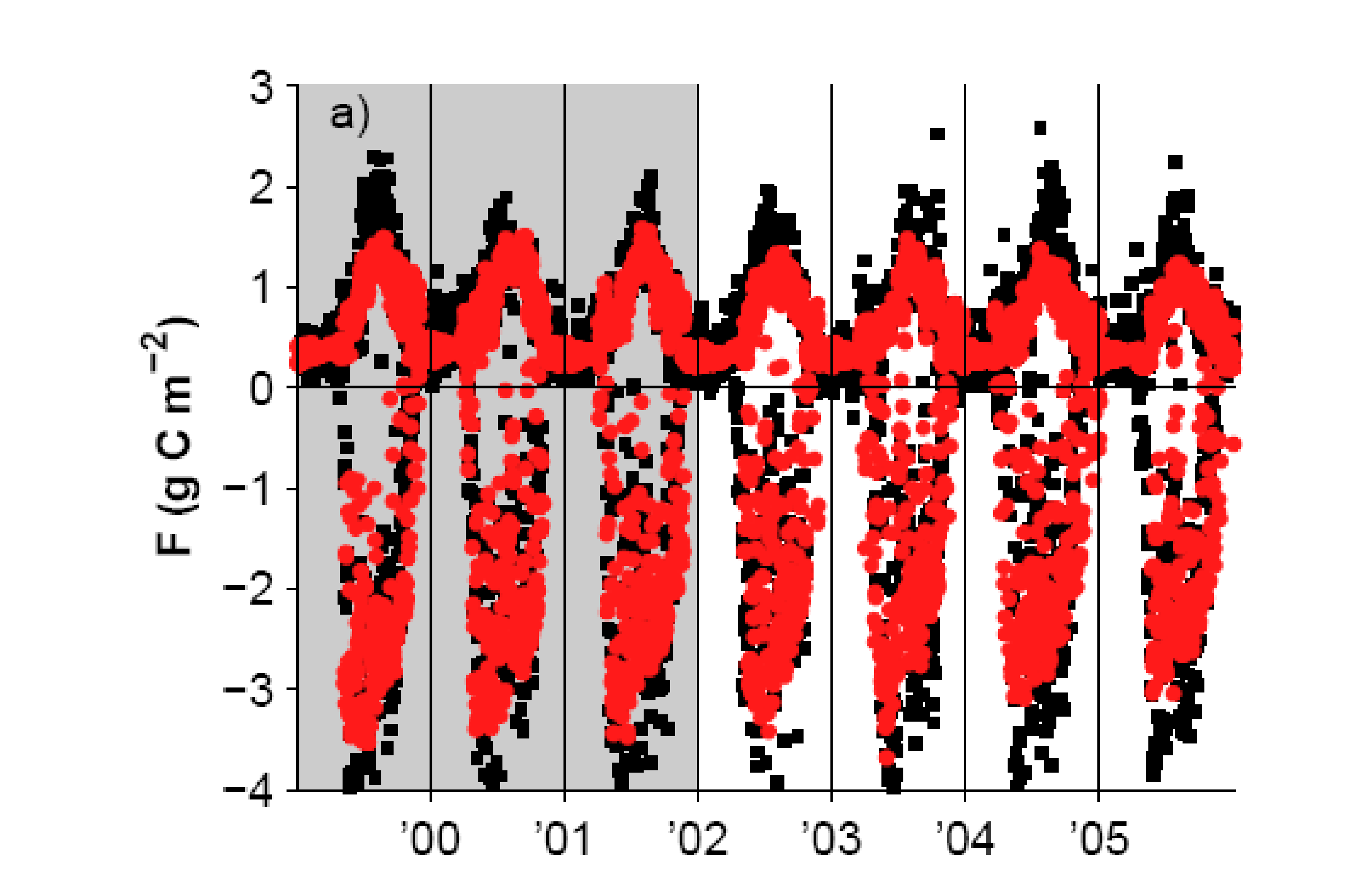






Zobitz et al 2008 Ecosystems

Model is optimized based on the first three years of data and used to predict the remaining years



Zobitz et al. (2008) Ecosystems

Table 4: Model comparisons using the optimized parameter set retrieved from each model run.

Model:	**Base**	**Roots**	**Quality**	**Microbes**
Log likelihood (LL):	-1462.5	-1437.8	-1423.7	-1634.7
Root mean square error:	0.45	0.44	0.45	0.48
Number of data points (n):	4463	4463	4463	4463
Number of parameters (M):	17	23	23	24
BIC†:	3068	$\boxed{3041}$	3069	3471

Validation data were used to calculate these values. The root mean square error is calculated from the squared difference between the measured and modeled difference for F and ET. (†): The Bayesian information criterion (BIC) equals $-2LL + M\ln(n)$. A lower BIC indicates a model with greater support from the data.

Zobitz et al. (2008) Ecosystems

Contrast between Day (psn) and Night (no psn) allows separation ← Separation of NEE into GPP and Re

(Sacks et al 2006, 2007)

Responses of NEE to precipitation change → **Seasonal co-ordination of GPP and ET (obs) allows a reasonable response to be extracted**

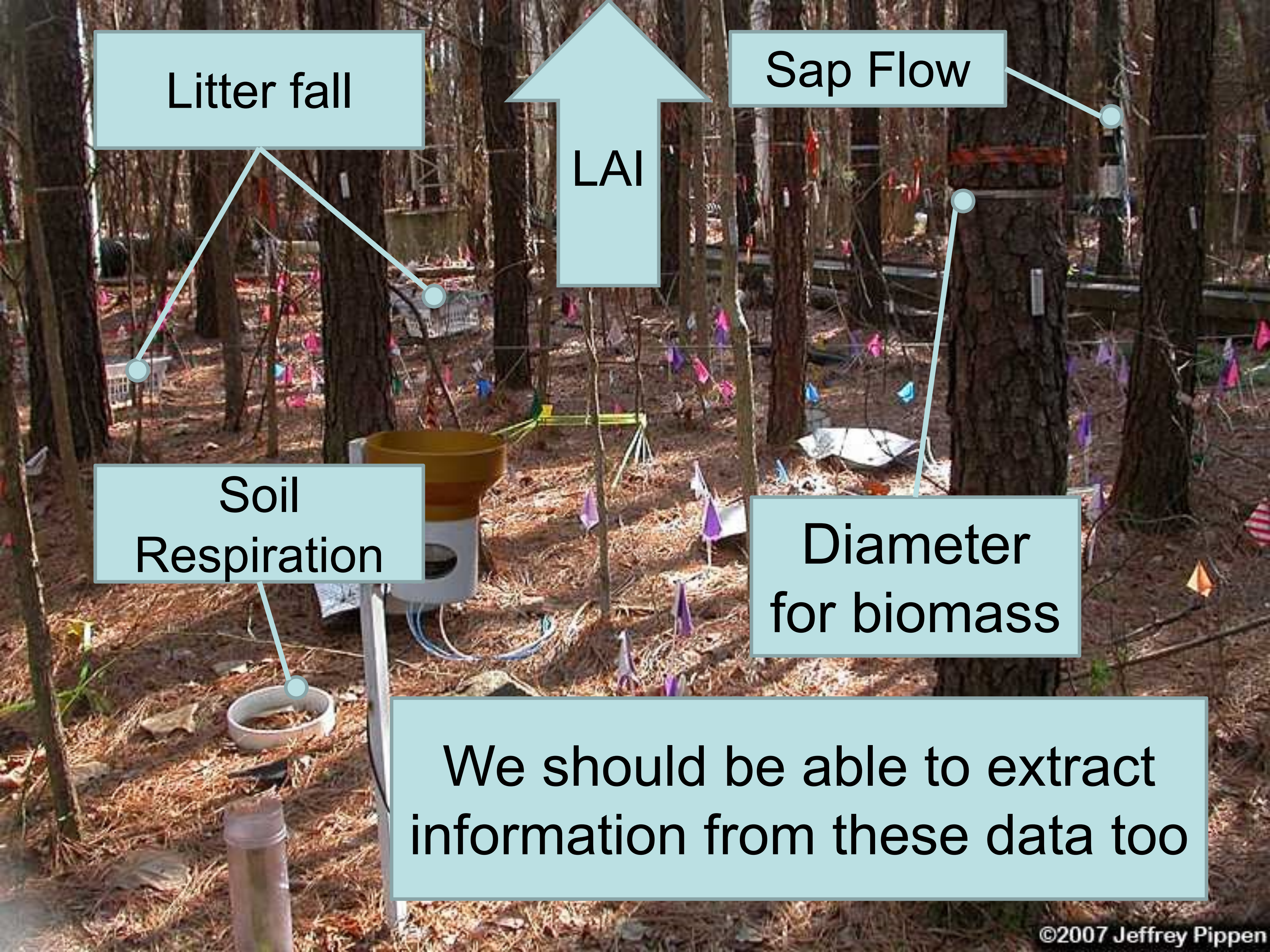
(Moore et al 2008)

Above and below ground processes confounded in tower based measurements ← Flux data alone does not constrain below ground processes well

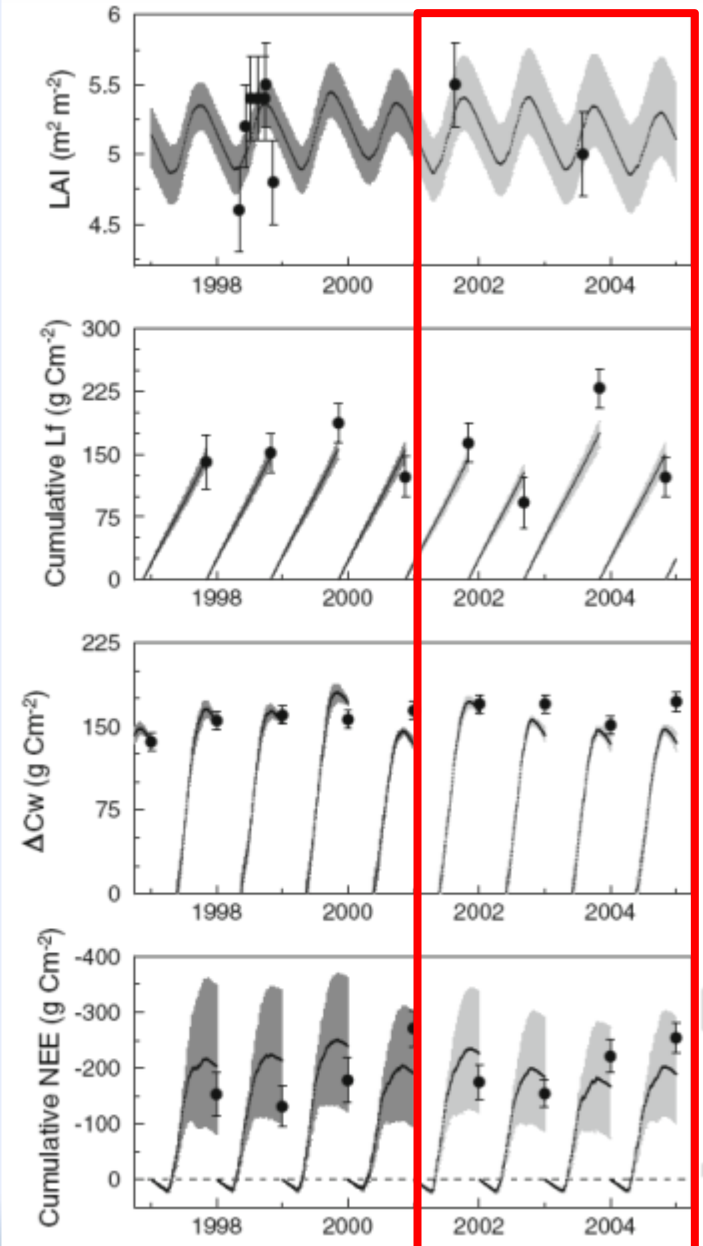
(Zobitz et al 2008)

NEE does not constrain long term processes → **Biomass, Soil Resp, LAI, litterfall can be used to constrain different parameters in the model**

Richardson *et al*. 2010


Litter fall
LAI
Sap Flow
Soil Respiration
Diameter for biomass
We should be able to extract information from these data too
©2007 Jeffrey Pippen

Howland Forest



LAI

NOTE: Calibration / **validation**

Litterfall

Woody Biomass

NEE

Fig. 4 Time series. Modeled leaf area index, *LAI*; litterfall, L_f (cumulative since last collection); annual woody biomass increment, ΔC_W; and annual cumulative net ecosystem exchange, (*NEE*) of carbon, with uncertainties (90% confidence interval), for the Howland Forest. Modeling was conducted with the DALEC model, constrained (calibration period 1997–2000; validation period 2001–2004) with a variety of different data streams (Run 8 in Table 2); actual measurements are indicated by *filled circles*, with *error bars* indicating estimated measurement uncertainties. For observed cumulative NEE, the annual sum was estimated by gap-filling the 30-min eddy covariance record using a standard empirical model

Richardson et al (2010) Estimating parameters of a forest ecosystem C model with measurements of stocks and fluxes as joint constraints. *Oecologia DOI:* 10.1007/s00442-010-1628-y

We can evaluate models at multiple timescales and using multiple datasets

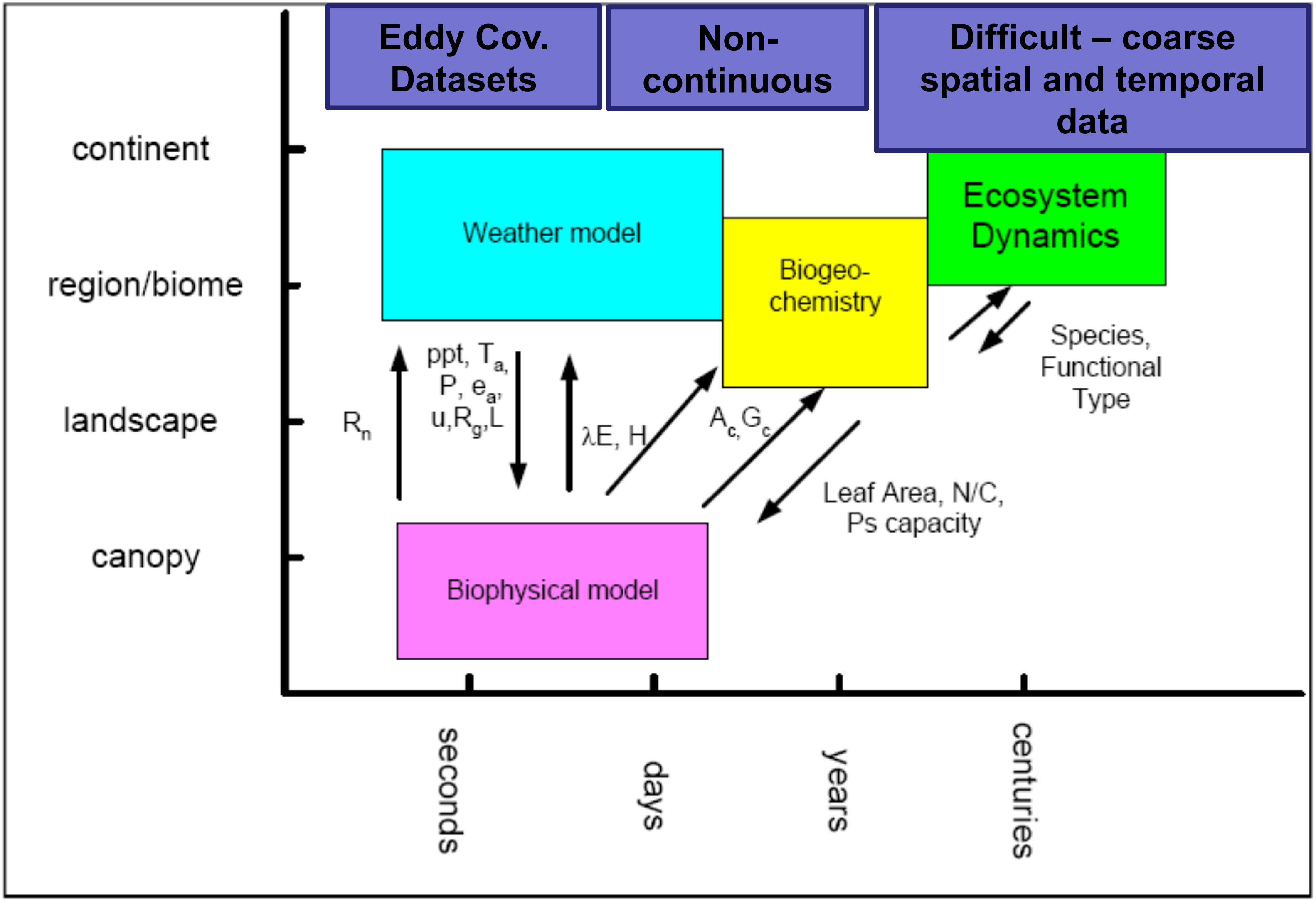
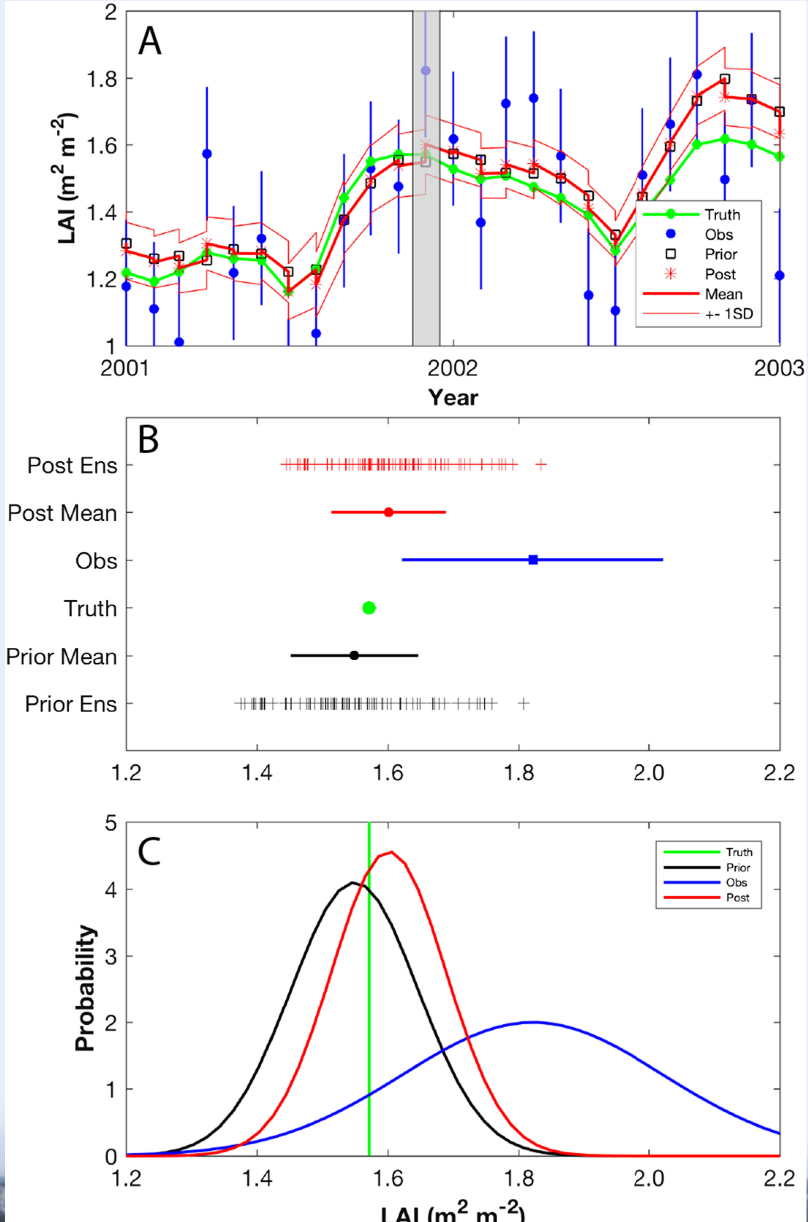


Diagram modified from: Dennis Baldocchi

DOING DATA ASSIMILATION WITH A LAND SURFACE MODEL

Fox, A. M., Hoar, T. J., Anderson, J. L., Arellano, A. F., Smith, W. K., Litvak, M. E., ... & Moore, D. J. (2018). Evaluation of a data assimilation system for land surface models using CLM4. 5. *Journal of Advances in Modeling Earth Systems*, *10*(10), 2471-2494.

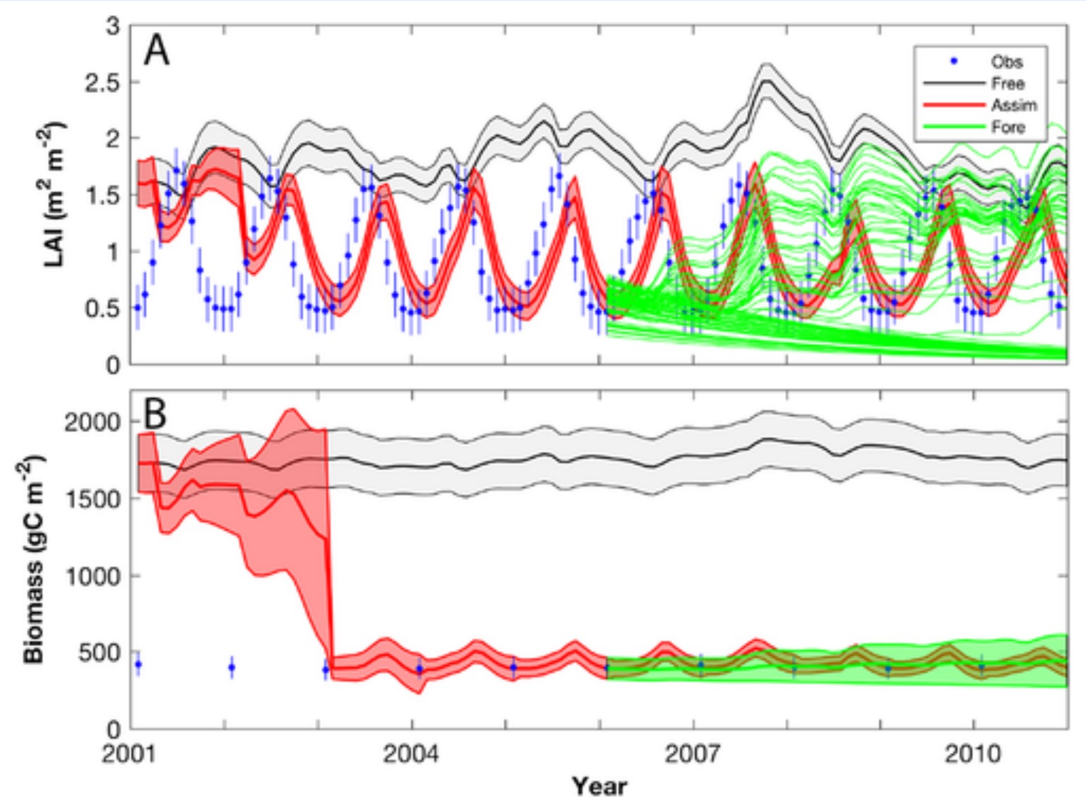


We assimilated LAI from MODIS into CLM for one of Marcy's sites in New Mexico

- Assimilating LAI and BIOMASS shows that the model is quite biased (compared to the observations)

- Assimilation removes that bias

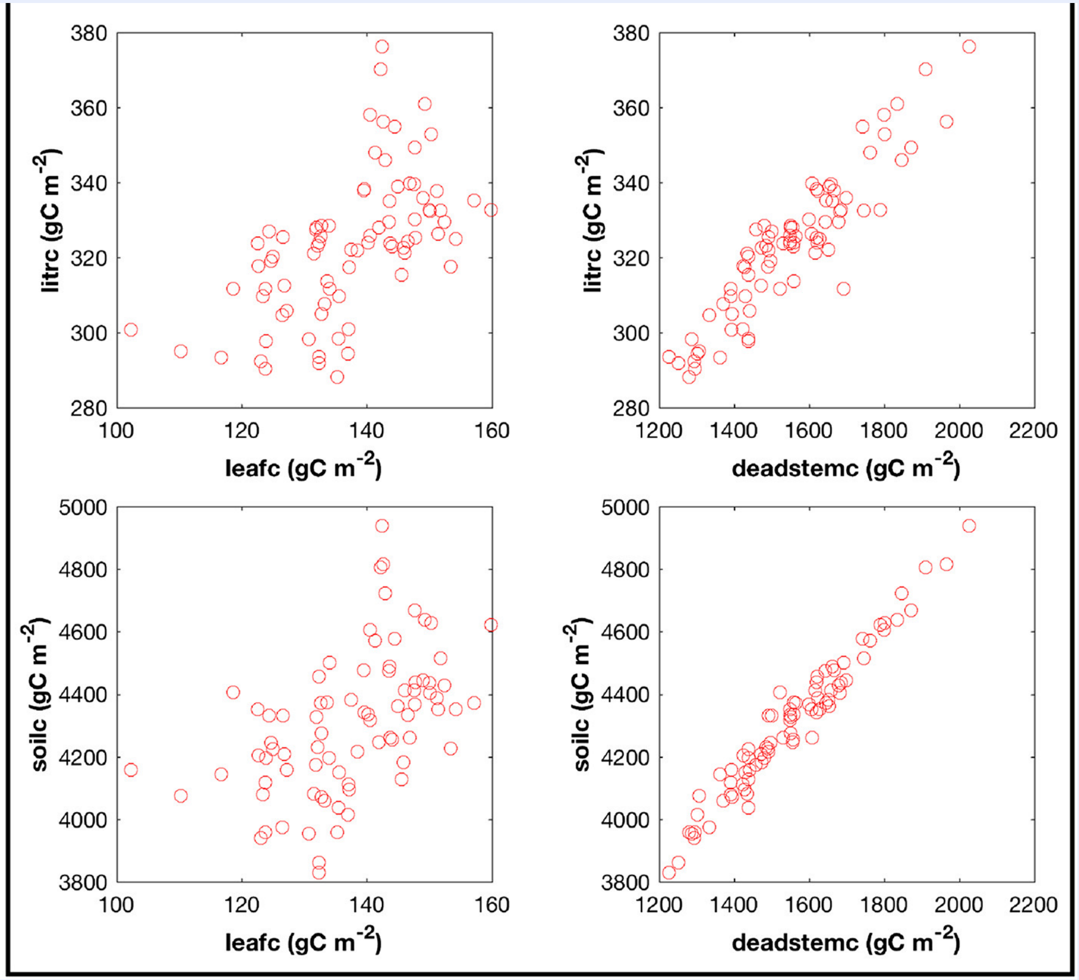
- We can then examine what the model needed to do so that it matched the BIOMASS and LAI



Fox et al 2018

•Assimilating satellite derived observations of leaf area allows us to estimate quantities that we DO NOT observe

•Here we show litter carbon and soil carbon that the model infers based on the correlation between LAI (leafC) and Biomass (deadstem)



Fox et al 2018

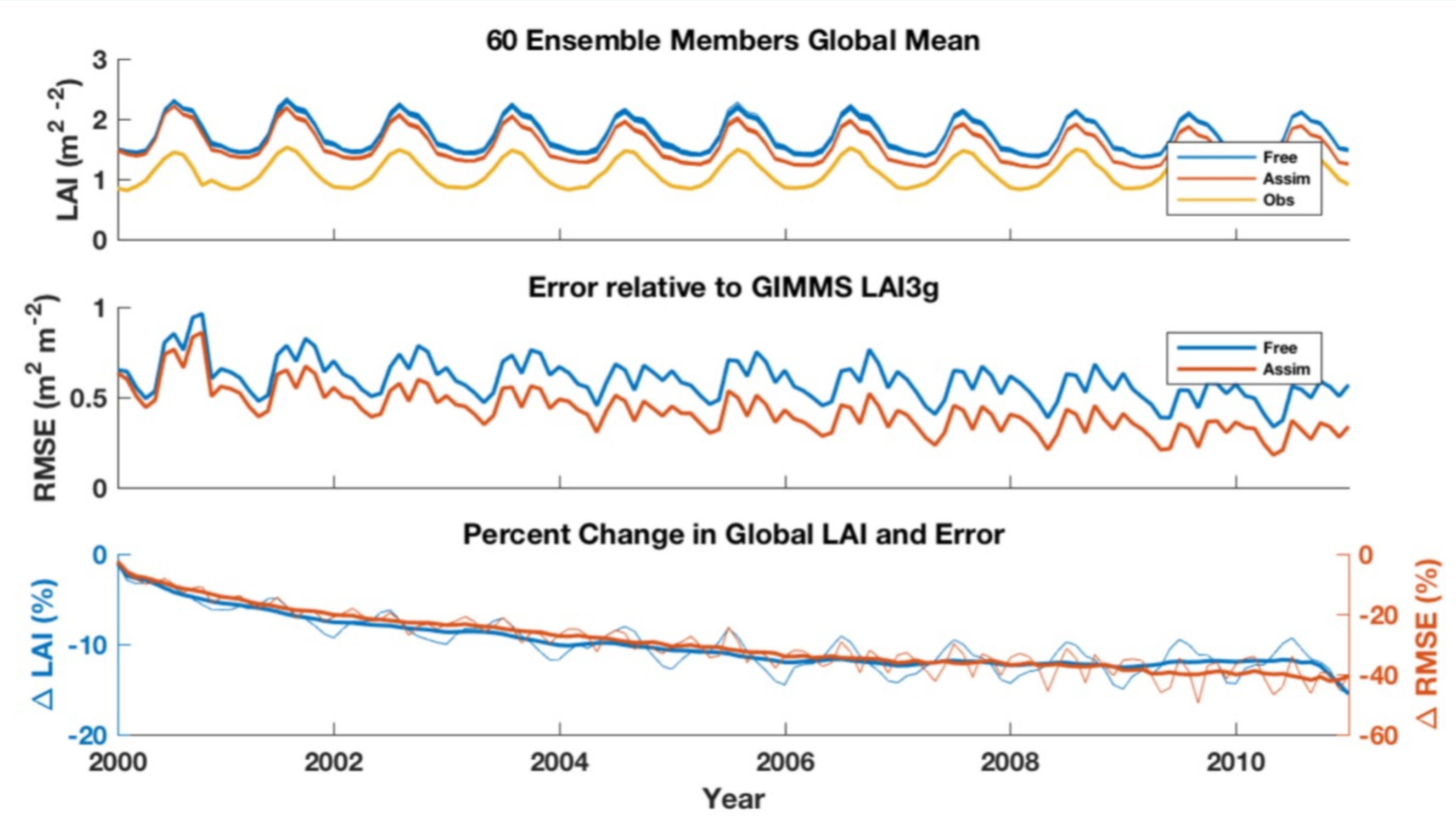
GO BIG OR GO HOME!!!!

DOING DATA ASSIMILATION WITH A LAND SURFACE MODEL FOR THE GLOBE

Fox, A. M., Huo, X., Hoar, T. J., Dashti, H., Smith, W. K., MacBean, N., ... & Moore, D. J. (2022) Assimilation of global satellite leaf area estimates reduces modeled global carbon uptake and energy loss by terrestrial ecosystems. *Journal of Geophysical Research: Biogeosciences*, e2022JG006830.

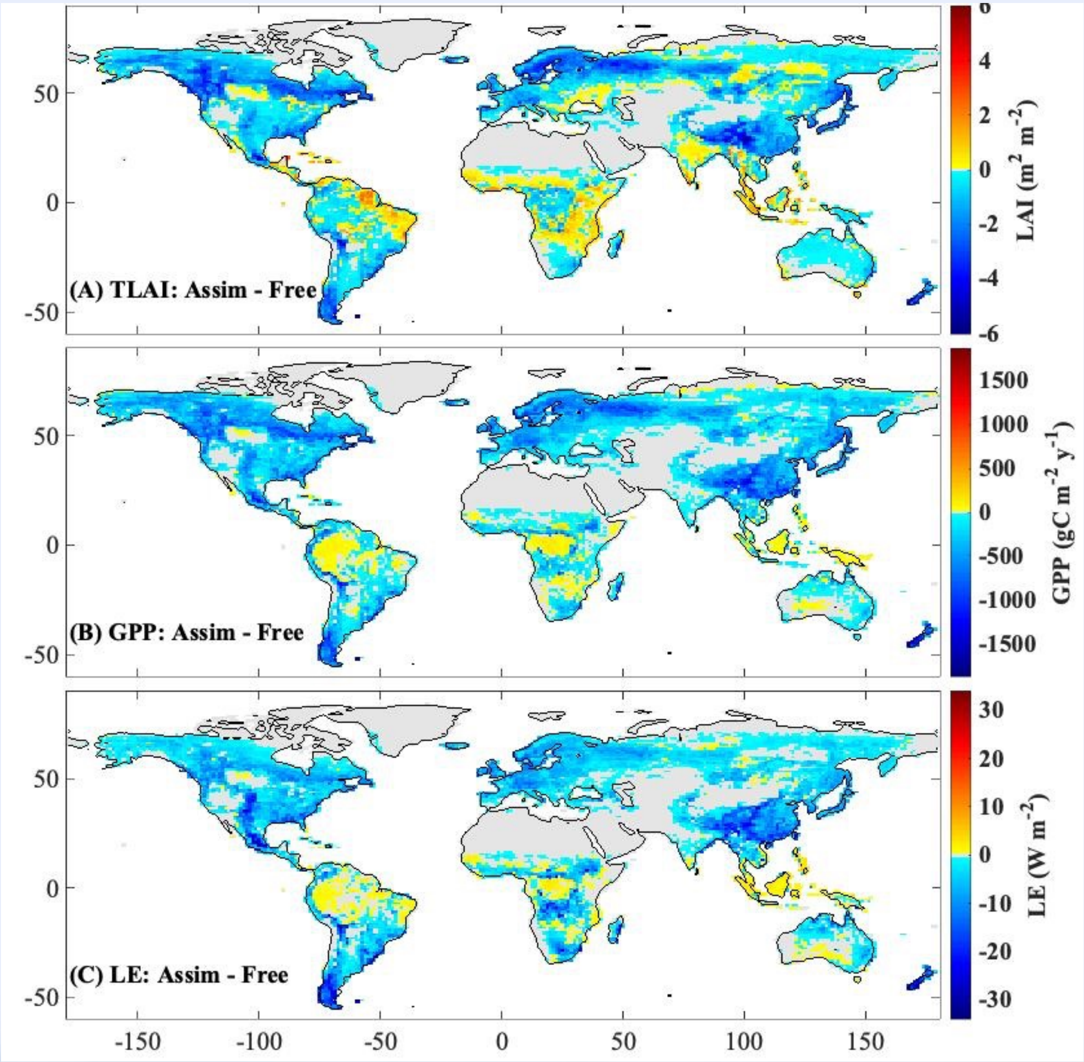
Assimilating global LAI estimates (LAI3g)

- We can also assimilate LAI for the entire globe and examine the consequences for fluxes



Fox, Huo et al 2022

Assimilating global LAI estimates (LAI3g)

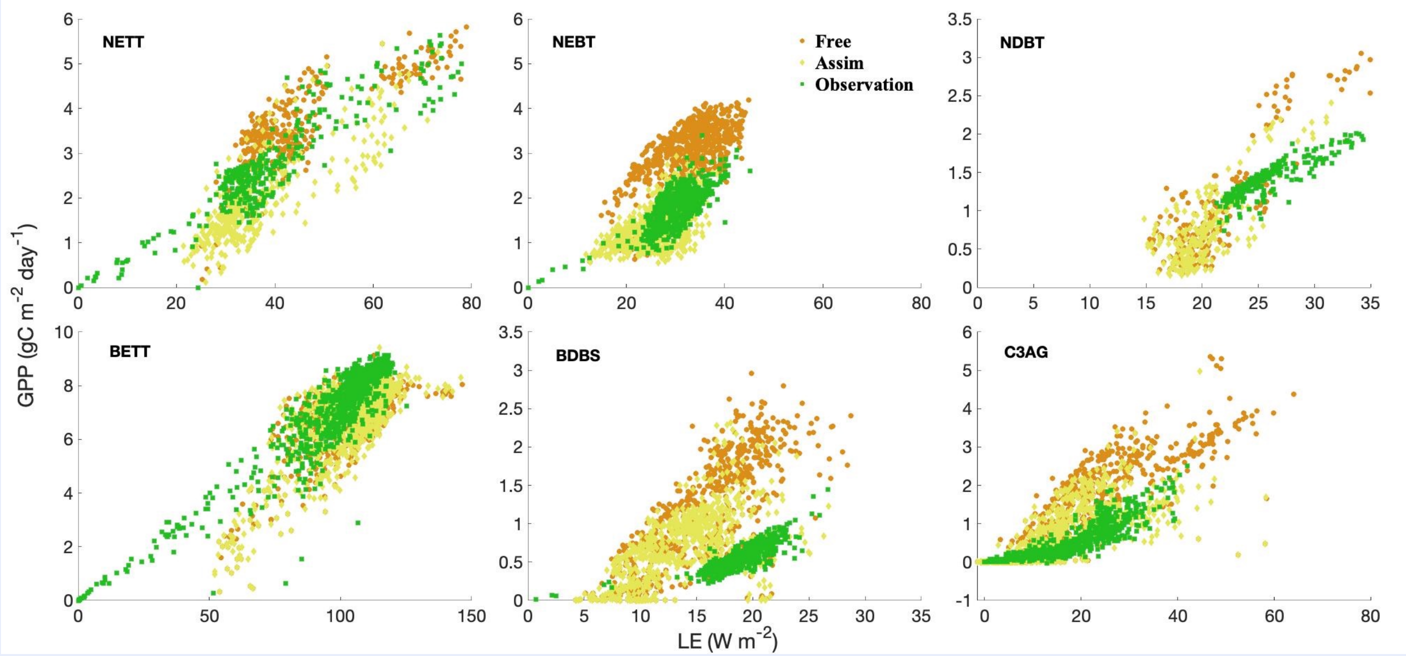


- Assimilating satellite derived observations of leaf area on average reduced **Community Land Model** estimates of Leaf Area Index

- This **reduced global estimates of gross primary production by 18% and latent heat flux by 6%,** improving fit to independent data sets

Fox, Huo et al 2022

Assimilating global LAI estimates (LAI3g)



Suggestions that the default CLM has poorly parameterized GPP to LE relationships for some Plant Functional Types

To check the credibility of the results we compare the fluxes in CLM against the FLUXCOM product – this allows us to compare the fluxes at the right scale – however we need to be mindful that the scaling procedure introduces new factors! Maybe there's some circularity here!

Fox, Huo et al 2022

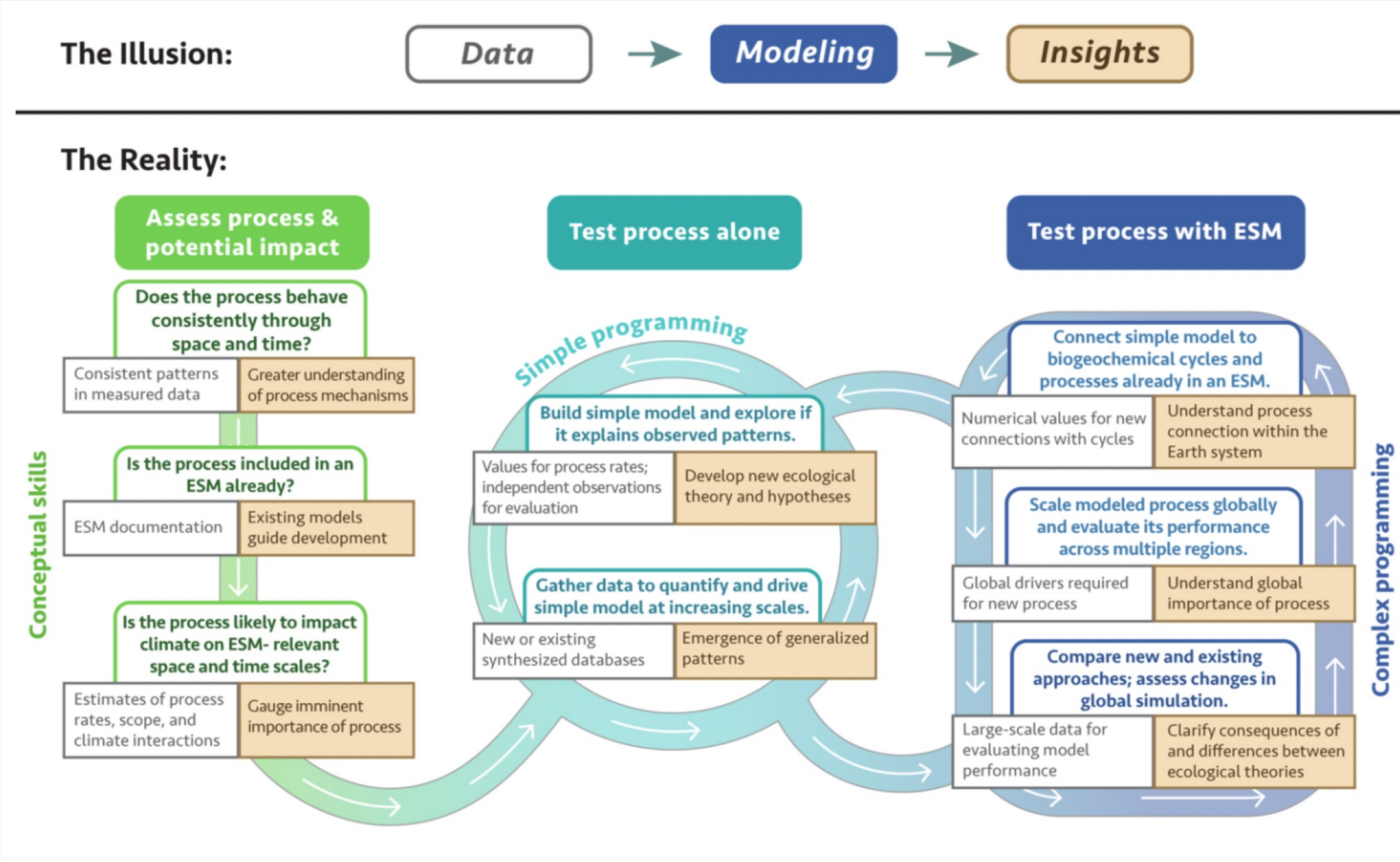
Is CLM too simple?

WE COULD JUST KEEP ADDING PROCESSES FOR EVER – MAKING MORE COMPLEX AND MORE UNWIELDY MODELS PROBABLY NOT A GOOD IDEA

There is a misconception that models just suck in data and produce insights.
While this has happened in the past, we think it’s better to have a more integrated approach

The old way!

Kyker Snowman et al 2021



Kyker Snowman et al 2021

Criteria for adding new processes to Earth System Models

1)New ecological processes should influence Earth's climate on a large scale or that the process must result in changes to the carbon, water or energy balance of ecosystems.

2) Any new process cannot require more of the model than the model can currently provide. For example, leaching of nutrients cannot be added to a model without a nutrient cycle.

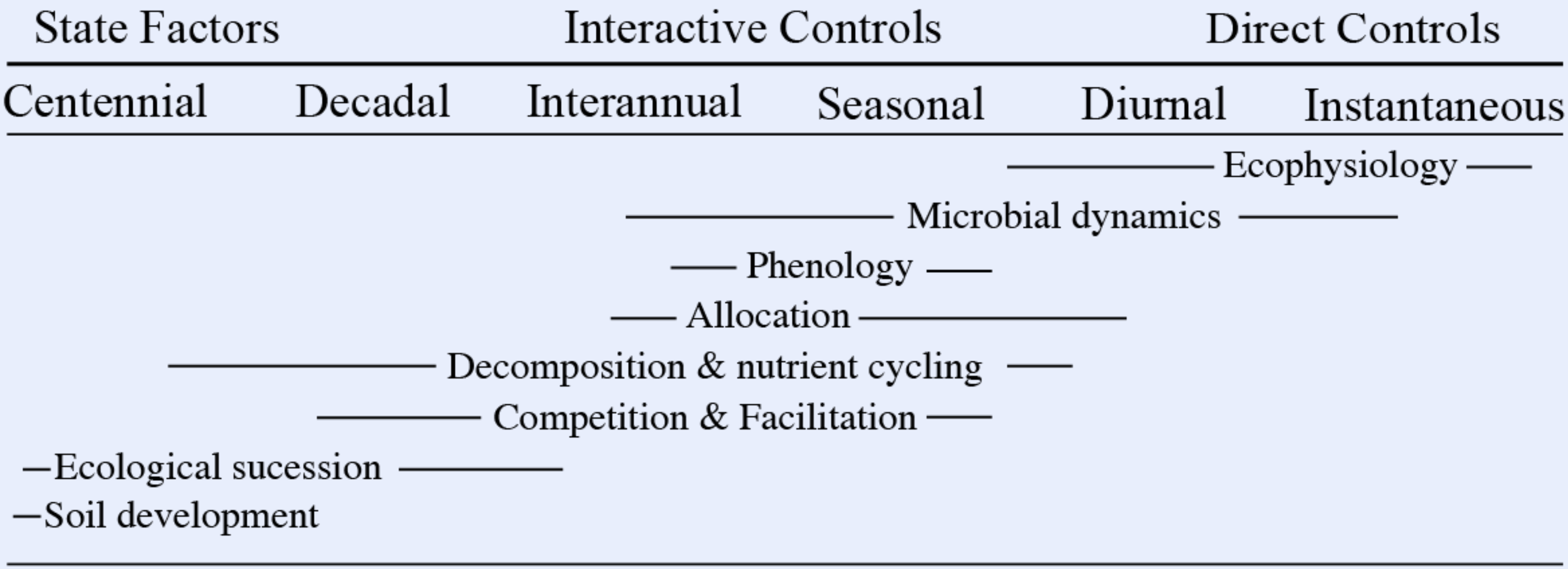
3) there should be sufficient understanding of the process and data to test the process globally; adding poorly established theory or theory that cannot be independently verified will cause potentially serious and unquantifiable bias.

4) the new processes must be governed by mathematics that are within reach of our current computational capacity

5) there must be a dedicated community of researchers to develop, test and maintain the process in the model.

Kyker Snowman et al 2021

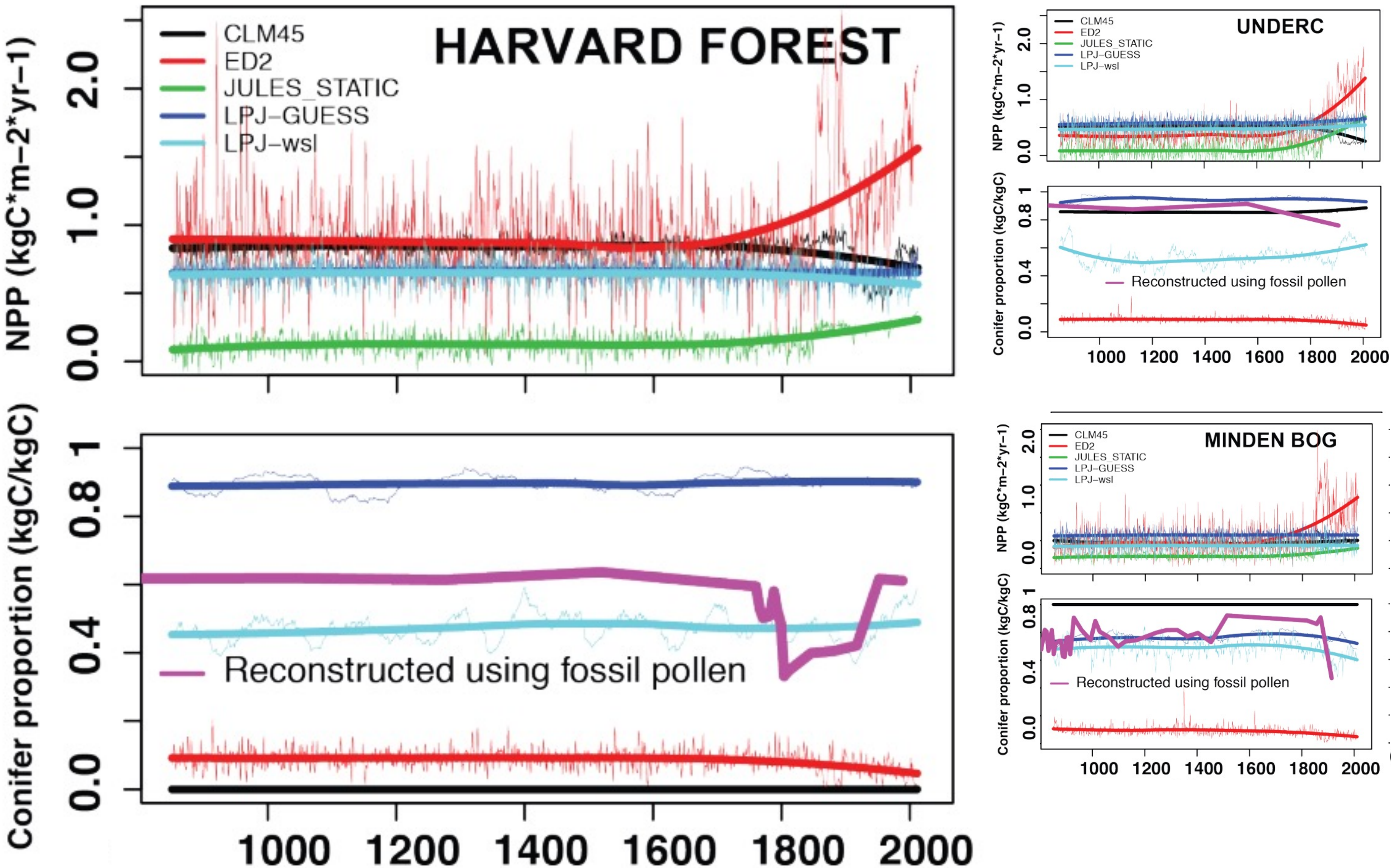
Some candidate Land Surface Model processes to investigate



When studying fluxes, ecological understanding is very useful.

Commentary on Kyker Snowman et al 2021

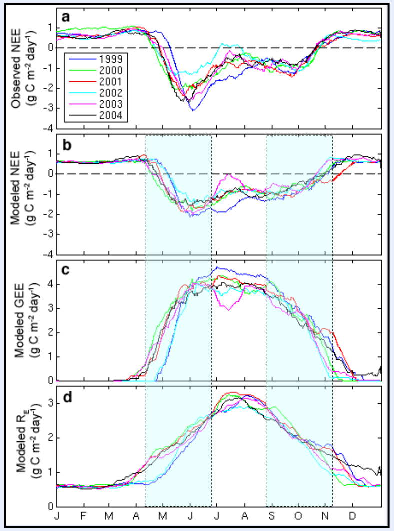
Long term vegetation dynamics



Liu, Moore et al *AGU 2014*

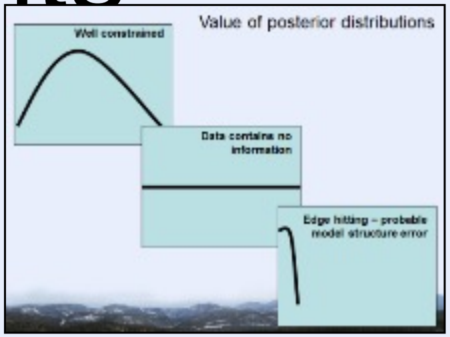
Some take home points

- Defensible estimates of GPP and Rtot can be extracted from NEE, though some extreme values are missed (errors?)

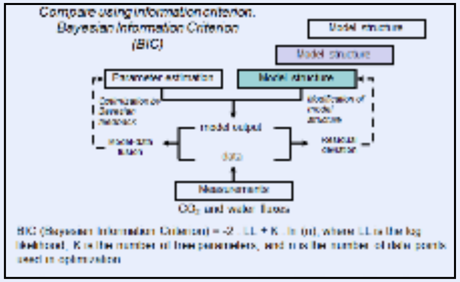


- We can get the right answer for the wrong reasons!
- Single datasets can only constrain some of the parameters and can lead to spurious results..

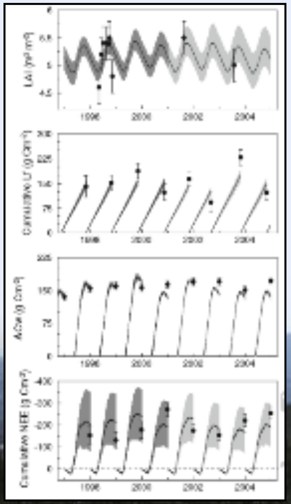
Some take home points



- Retrieved parameter distributions help us understand how good our constraints are



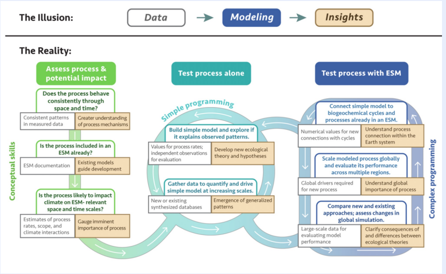
- Model structures can be tested by comparing the data-model mismatch of optimized parameter sets



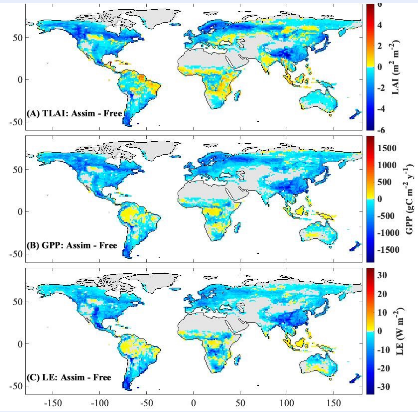
- Multiple data sets can be used to cross constrain parameters and processes.

Some take home points

- Not all models NEED to be complex – we need to think carefully about which processes we need to add



- We can assimilate STATES like LAI and biomass and then examine what the impact on the fluxes are



- We can carry this out at the site, regional or global level.