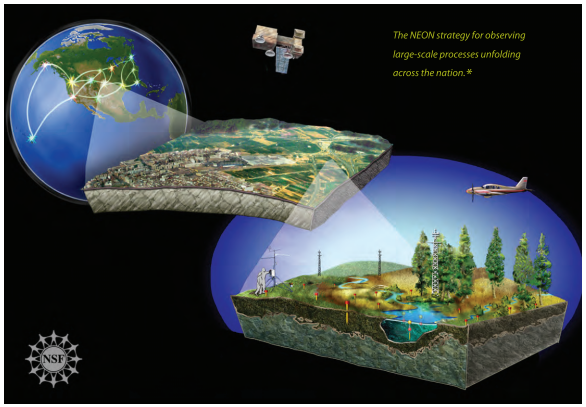
INTEGRATING NEON DATA WITH EXISTING MODELS: AN EXAMPLE WITH THE COMMUNITY LAND MODEL

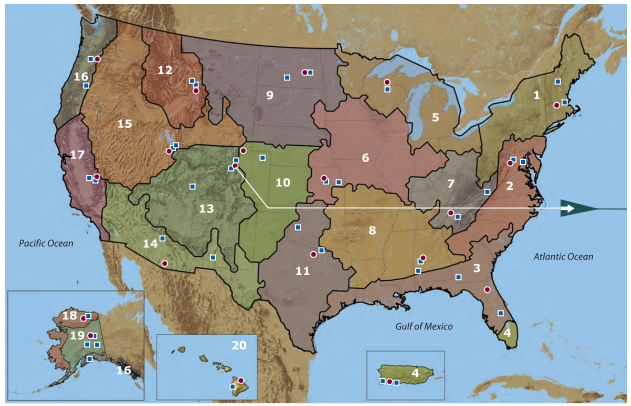
Edmund M. Hart[1], Andy Fox[1],Steve Berukoff[1], Tim Hoar[2]

1. National Ecological Observatory Network
2. Institute for Mathematics Applied to Geosciences (iMAGE), National Center for Atmospheric Research

Overview

- What is NEON?
- Data assimilation research test-bed (DART) and the CLM
- NEON Informatics architecture





What is NEON?

The overarching goal of NEON is to enable understanding and forecasting of climate change, land use change, and invasive species on continental-scale ecology *by providing infrastructure* to support research in these areas.

- ~14,000 automated terrestrial sensors
- 45 Tb of data/year
- 584 level 1 data products
- 95 different measurement types
- 60 different sites

NEON sites

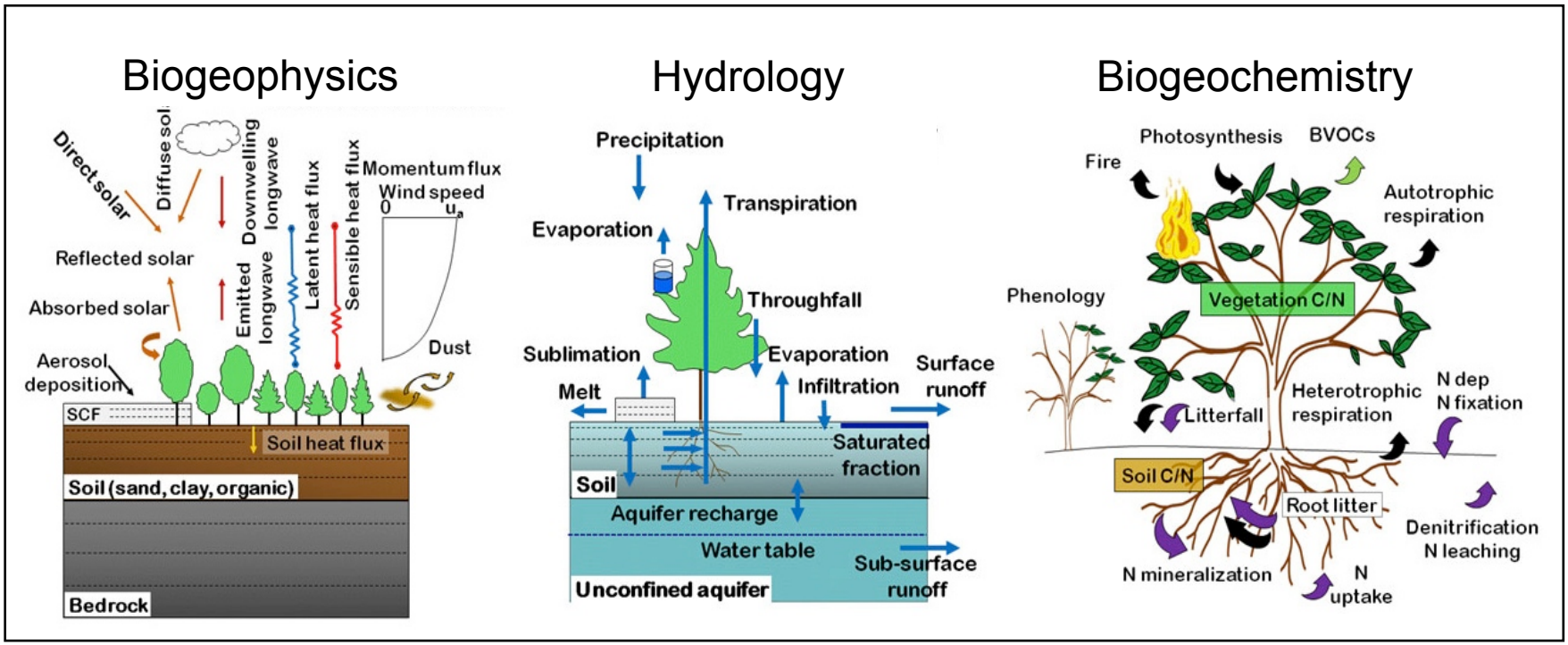
Ecological forecasting

1. **Predicting the most likely future state of an ecological system**
   i. Relevant for short-term forecasts
   ii. Systems own dynamics most strongly govern change over time
   iii. i.e. forecasting the likely spread of invasive species

2. **Predicting the most likely future state, given a decision today**
   i. Relevant when alternate management actions or scenarios are considered
   ii. i.e. forecasting likely impacts on biodiversity from alternate wildfire mitigation schemes

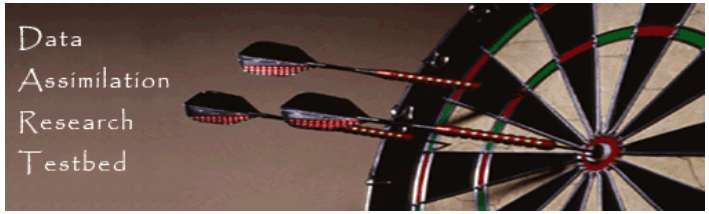
CLM – Community Land Model

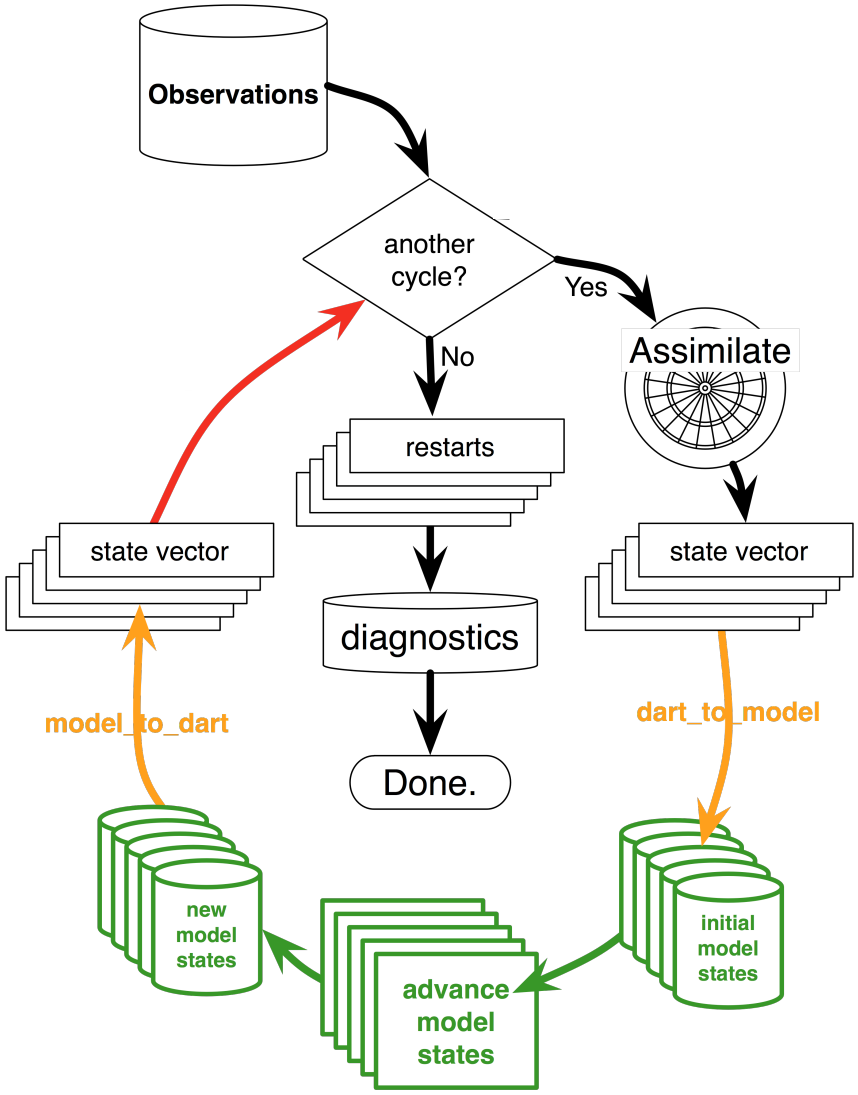
Simulates ecoclimatological processes in each grid cell.



Data Assimilation Research Testbed (DART)

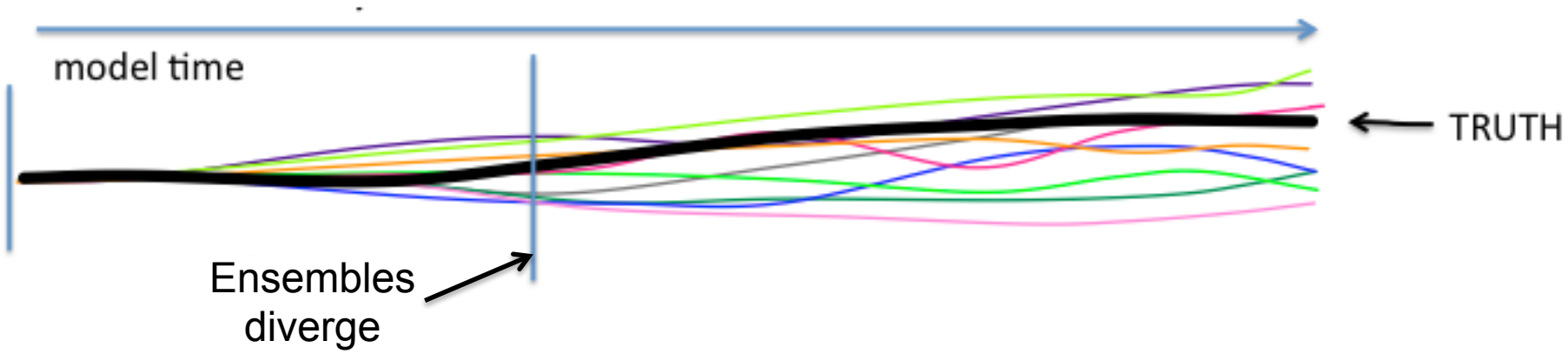
- DART is a community facility for ensemble DA
- Uses a variety of flavors of filters
  - Ensemble Adjustment Kalman Filter
- Many enhancements to basic filtering algorithms
  - Adaptive inflation
  - Localization





Data assimilation

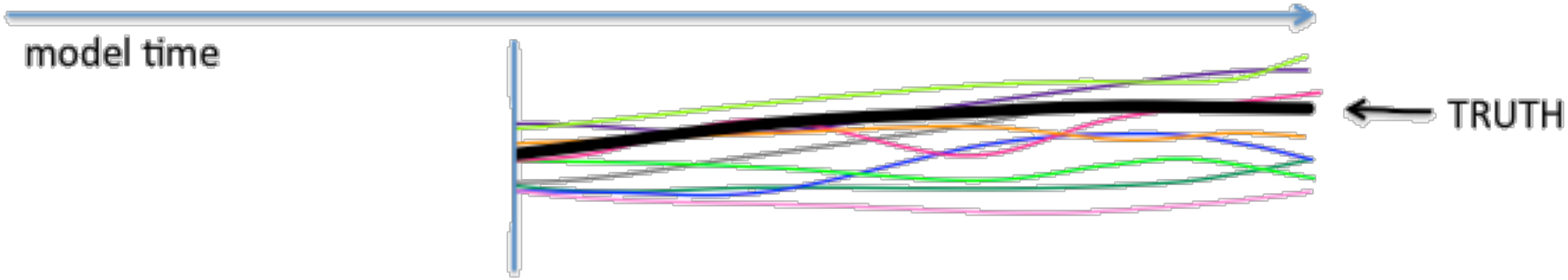
- Testing the success of the DART-CLM implementation with perfect model experiments



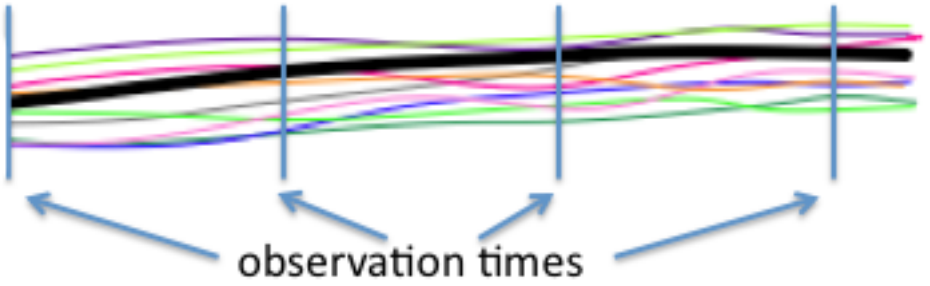
- Each line represents the evolution of individual instances of CLM
- Pick one instance and declare it the truth
- Generate synthetic observations from this 'truth', adding a prescribed noise/uncertainty

Data assimilation

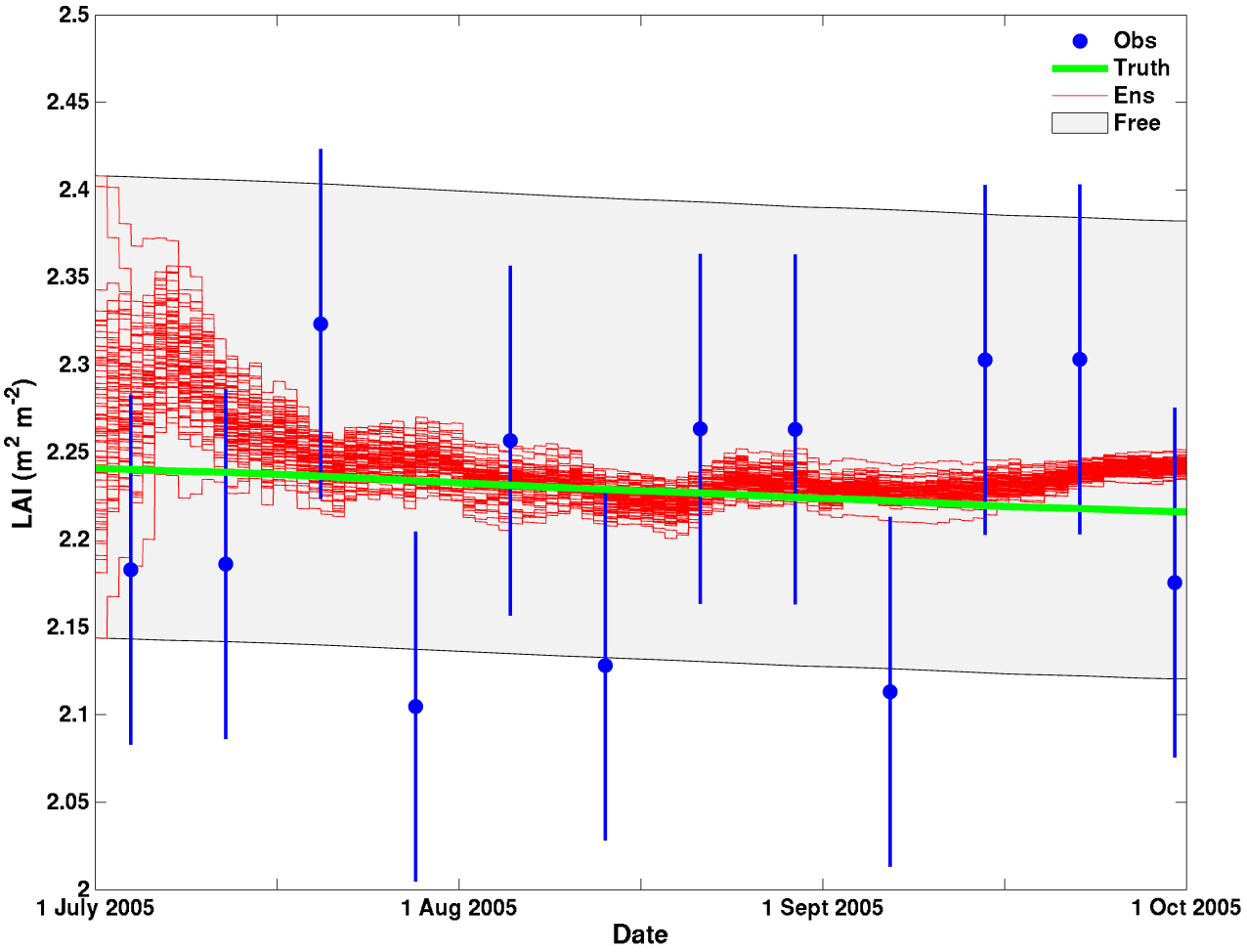
- Without assimilation:
  - frequently ensemble spread will grow



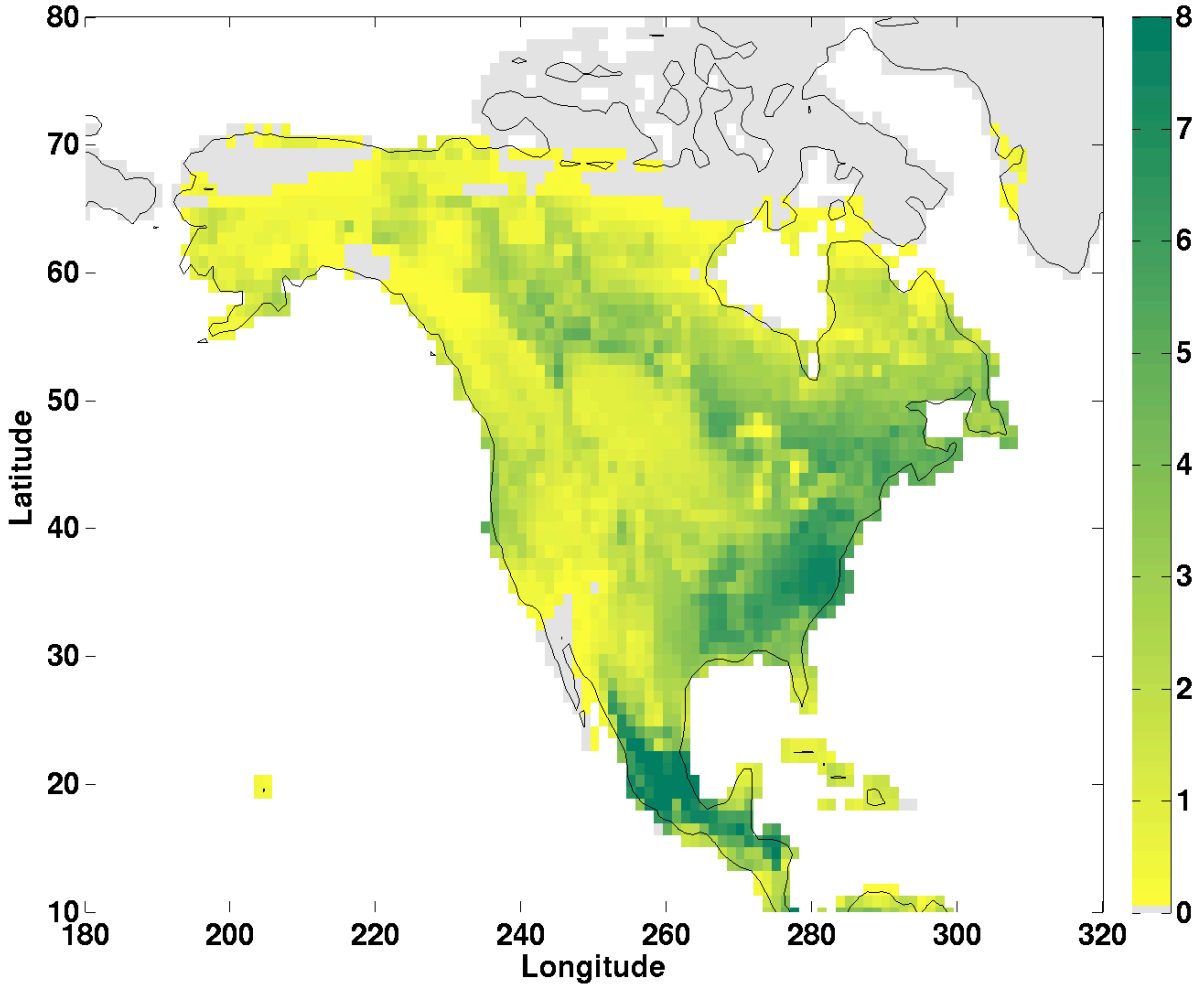
- With assimilation:
  - Ensemble spread remains stable, is small enough to be informative, but does not collapse away from truth



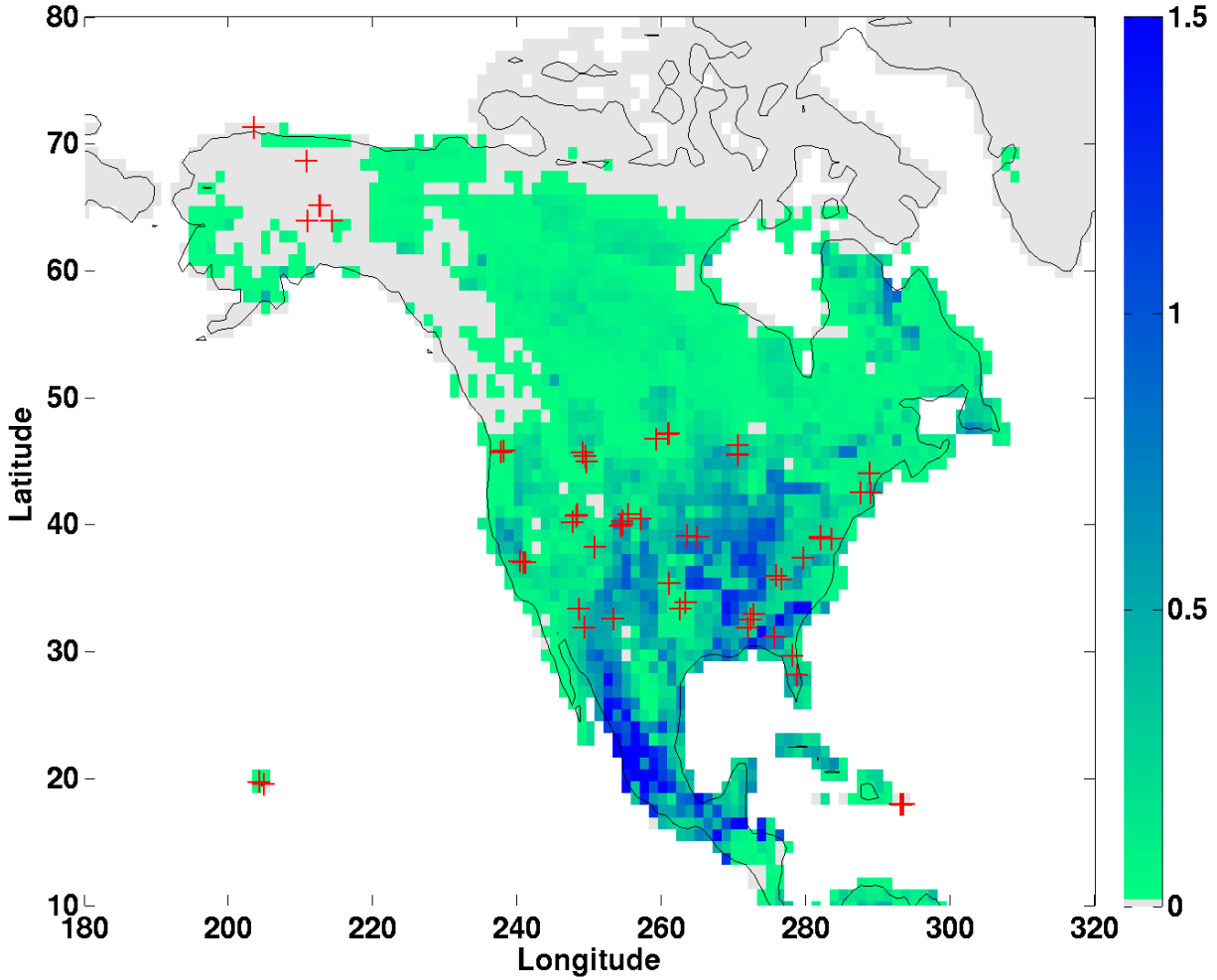
LAI – observations every 8 days



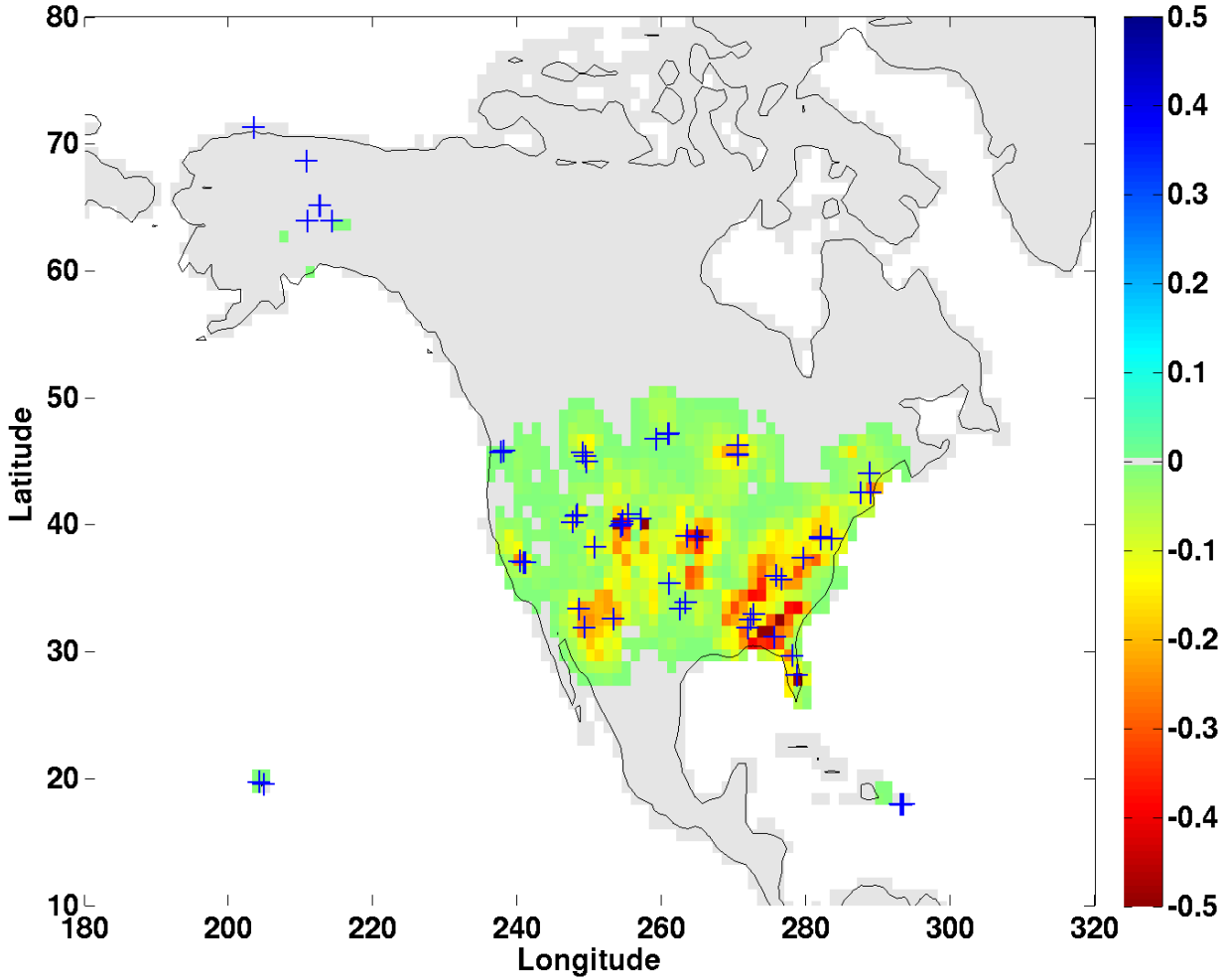
Mean LAI from 80 ensemble members



LAI spread from 80 ensemble members



Reduction in LAI ensemble spread



Guts of DA - Ameriflux

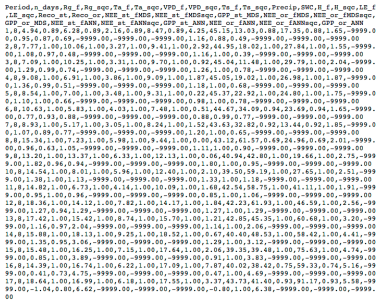
L4 data from NC Ameriflux

```
Period,n_days,Rg_f,Rg_sqc,Ta_f,Ta_sqc,VPD_f,VPD_sqc,Ts_f,Ts_sqc,Precip,SWC,H_f,H_sqc,LE_f
,LE_sqc,Reco_st,Reco_or,NEE_st_fMDS,NEE_st_fMDSsqc,GPP_st_MDS,NEE_or_fMDS,NEE_or_fMDSsqc,
GPP_or_MDS,NEE_st_fANN,NEE_st_fANNsqc,GPP_st_ANN,NEE_or_fANN,NEE_or_fANNsqc,GPP_or_ANN
1,8,4.94,0.89,6.28,0.89,2.16,0.89,8.47,0.89,4.25,45.15,13.03,0.88,17.35,0.88,1.65,-9999.0
0,0.95,0.87,0.69,-9999.00,-9999.00,-9999.00,1.16,0.88,0.49,-9999.00,-9999.00,-9999.00
2,8,7.77,1.00,10.06,1.00,3.27,1.00,9.41,1.00,2.92,44.95,18.02,1.00,27.84,1.00,1.55,-9999.
00,1.08,0.97,0.48,-9999.00,-9999.00,-9999.00,1.16,1.00,0.39,-9999.00,-9999.00,-9999.00
3,8,7.09,1.00,10.25,1.00,3.31,1.00,9.70,1.00,0.92,45.04,11.48,1.00,29.79,1.00,2.04,-9999.
00,1.29,0.99,0.74,-9999.00,-9999.00,-9999.00,1.26,1.00,0.78,-9999.00,-9999.00,-9999.00
4,8,9.08,1.00,6.91,1.00,3.86,1.00,9.09,1.00,1.87,45.05,19.02,1.00,26.98,1.00,1.87,-9999.0
0,1.36,0.99,0.51,-9999.00,-9999.00,-9999.00,1.18,1.00,0.68,-9999.00,-9999.00,-9999.00
5,8,8.54,1.00,7.00,1.00,3.48,1.00,9.31,1.00,0.22,45.37,22.92,1.00,24.80,1.00,1.75,-9999.0
0,1.10,1.00,0.66,-9999.00,-9999.00,-9999.00,0.98,1.00,0.78,-9999.00,-9999.00,-9999.00
6,8,10.63,1.00,5.83,1.00,4.03,1.00,7.48,1.00,0.51,44.67,34.09,0.94,23.69,0.94,1.65,-9999.
00,0.77,0.93,0.88,-9999.00,-9999.00,-9999.00,0.88,0.99,0.77,-9999.00,-9999.00,-9999.00
7,8,8.93,1.00,5.17,1.00,3.05,1.00,8.24,1.00,1.52,43.63,32.82,0.92,13.44,0.92,1.85,-9999.0
0,1.07,0.89,0.77,-9999.00,-9999.00,-9999.00,1.20,1.00,0.65,-9999.00,-9999.00,-9999.00
8,8,15.34,1.00,7.23,1.00,5.98,1.00,9.44,1.00,0.00,43.12,61.57,0.69,24.96,0.69,2.01,-9999.
00,0.96,0.63,1.05,-9999.00,-9999.00,-9999.00,1.11,1.00,0.90,-9999.00,-9999.00,-9999.00
9,8,13.20,1.00,13.37,1.00,6.33,1.00,12.13,1.00,0.06,40.94,42.80,1.00,19.66,1.00,2.75,-999
9.00,1.82,0.96,0.94,-9999.00,-9999.00,-9999.00,1.80,1.00,0.95,-9999.00,-9999.00,-9999.00
10,8,14.54,1.00,8.01,1.00,5.96,1.00,12.40,1.00,2.10,39.50,59.19,1.00,27.65,1.00,2.51,-999
9.00,1.38,1.00,1.13,-9999.00,-9999.00,-9999.00,1.33,1.00,1.18,-9999.00,-9999.00,-9999.00
11,8,14.82,1.00,6.73,1.00,4.14,1.00,10.09,1.00,1.68,42.54,58.75,1.00,41.11,1.00,1.91,-999
9.00,0.95,1.00,0.96,-9999.00,-9999.00,-9999.00,0.85,1.00,1.06,-9999.00,-9999.00,-9999.00
12,8,18.36,1.00,14.12,1.00,7.82,1.00,14.17,1.00,1.84,42.23,61.93,1.00,46.59,1.00,2.56,-99
99.00,1.27,0.94,1.29,-9999.00,-9999.00,-9999.00,1.27,1.00,1.29,-9999.00,-9999.00,-9999.00
13,8,17.42,1.00,15.42,1.00,8.74,1.00,15.70,1.00,1.21,42.85,45.35,1.00,60.68,1.00,3.20,-99
99.00,1.16,0.97,2.04,-9999.00,-9999.00,-9999.00,1.14,1.00,2.06,-9999.00,-9999.00,-9999.00
14,8,15.88,1.00,18.13,1.00,9.25,1.00,18.52,1.00,0.67,40.40,48.53,1.00,58.42,1.00,4.41,-99
99.00,1.35,0.95,3.06,-9999.00,-9999.00,-9999.00,1.29,1.00,3.12,-9999.00,-9999.00,-9999.00
15,8,15.48,1.00,16.25,1.00,7.15,1.00,17.64,1.00,2.06,39.35,39.48,1.00,75.63,1.00,4.74,-99
99.00,0.85,1.00,3.89,-9999.00,-9999.00,-9999.00,0.91,1.00,3.83,-9999.00,-9999.00,-9999.00
16,8,14.39,1.00,16.74,1.00,6.22,1.00,17.09,1.00,7.87,40.02,38.42,0.75,59.33,0.74,5.16,-99
99.00,0.41,0.73,4.75,-9999.00,-9999.00,-9999.00,0.47,1.00,4.69,-9999.00,-9999.00,-9999.00
17,8,18.64,1.00,16.99,1.00,6.18,1.00,17.55,1.00,3.37,43.73,41.40,0.93,91.17,0.93,5.58,-99
99.00,-1.04,0.80,6.62,-9999.00,-9999.00,-9999.00,-0.80,1.00,6.38,-9999.00,-9999.00,-9999.
00
```

Guts of DA - Ameriflux

How to make sense of it.

L4 data from NC Ameriflux



LEVEL 4 – VARIABLE DESCRIPTION

Variables description:
Level 4 data are obtained from the level 3 products, data are ustar filtered, gap-filled using different methods and partitioned. Datasets are also aggregated from daily to monthly. Flags with information regarding quality of the original and gapfilled data are added.

Half hourly dataset variables description:

- Month: from 1 to 12
- Day: day of the month
- Hour: from 0 to 23.5, indicates the end of the half hour of measurement
- DoY: decimal day of the year
- Rg_f: global radiation filled [W m-2]
- Rg_fqc: global radiation quality flags: 0 = original, 1 = category A (most reliable), 2 = category B (medium), 3 = category C (least reliable). (Refer to Reichstein et al. 2005 Global Change Biology for more information)
- Ta_f: air temperature filled [°C]
- Ta_fqc: air temperature quality flags: 0 = original, 1 = category A (most reliable), 2 = category B (medium), 3 = category C (least reliable). (Refer to Reichstein et al. 2005 Global Change Biology for more information)
- VPD_f: vapour pressure deficit [hPa]
- VPD_fqc: vapour pressure deficit quality flags: 0 = original, 1 = category A (most reliable), 2 = category B (medium), 3 = category C (least reliable). (Refer to Reichstein et al. 2005 Global Change Biology for more information)
- Ts_f: soil temperature filled [°C]
- Ts_fqc: soil temperature quality flags: 0 = original, 1 = category A (most reliable), 2 = category B (medium), 3 = category C (least reliable). (Refer to Reichstein et al. 2005 Global Change Biology for more information)
- Precip: precipitation [mm]
- SWC: soil water content [%vol]
- H_f: sensible heat flux filled [W m-2]
- H_fqc: sensible heat flux quality flags: 0 = original, 1 = category A (most reliable), 2 = category B (medium), 3 = category C (least reliable). (Refer to Reichstein et al. 2005 Global Change Biology for more information)
- LE_f: latent heat flux filled [W m-2]

Guts of DA - Ameriflux

How to get data into DART

NAMELIST

This namelist is read from the file `input.nml`. Namelists start with an ampersand '&' and terminate with a slash '/'. Character strings that contain a '/' must be enclosed in quotes to prevent them from prematurely terminating the namelist.

```
&level4_to_obs_nml
   text_input_file = 'textdata.input',
   obs_out_file    = 'obs_seq.out',
   year            = -1,
   timezoneoffset  = -1,
   latitude        = -1.0,
   longitude       = -1.0,
   elevation       = -1.0,
   flux_height     = -1.0,
   maxgoodqc       = 3,
   verbose         = .false.
   /
```

Contents	Type	Description
text_input_file	character(len=128)	Name of the Level 4 ASCII file of comma-separated values. This may be a relative or absolute filename.
obs_out_file	character(len=128)	Name of the output observation sequence file.
year	integer	The year of the observations in the Level 4 text file.
timezoneoffset	real	the time zone offset (in hours) of the station. The tower observation times are local time, we need to convert them to GMT.
latitude	real	Latitude (in degrees N) of the tower.
longitude	real	Longitude (in degrees E) of the tower. For internal consistency, DART uses longitudes in the range [0,360]. An input value of -90 will be converted to 270, for example.
elevation	real	surface elevation (in meters) of the tower.
flux_height	real	height (in meters) of the flux instrument on the tower.
maxgoodqc	real	maximum value of any observation quality control flag to pass through to the output observation sequence. Keep in mind that *filter* has the ability to discriminate on the value, so there is really little to be gained by rejecting them during the conversion.
verbose	logical	Print extra information during the *level4_to_obs* execution.

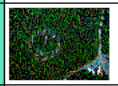
Guts of DA - Ameriflux

How to get data into DART

NAMELIST

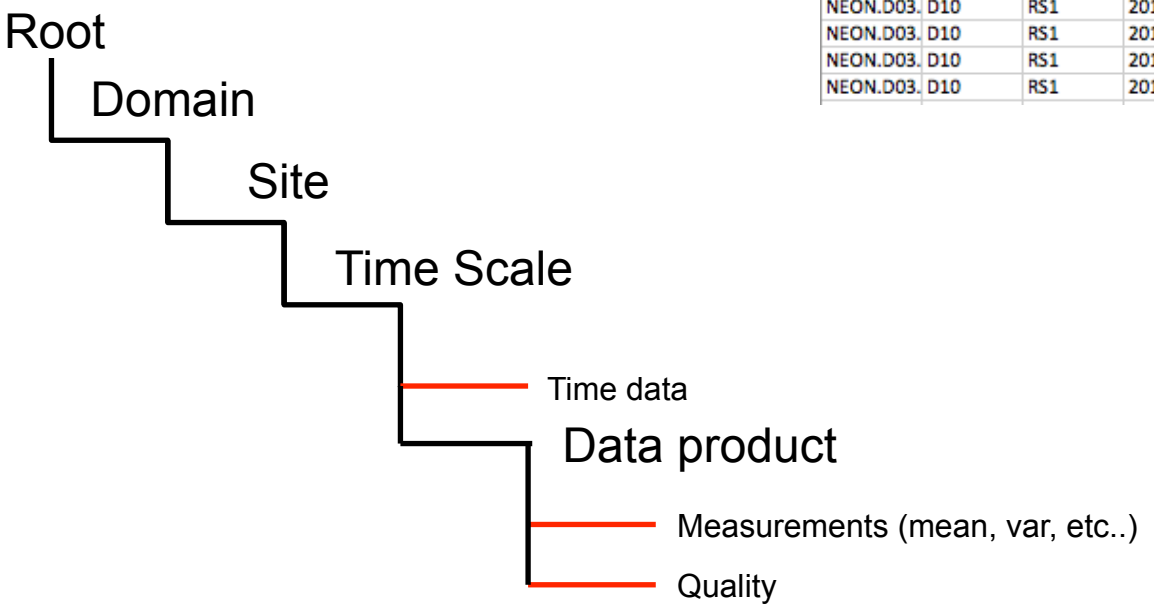
This namelist is read from the file `input.nml`.
from prematurely terminating the namelist.

```
&level4_to_obs_nml
   text_input_file = 'textdata.input',
   obs_out_file    = 'obs_seq.out',
   year            = -1,
   timezoneoffset  = -1,
   latitude        = -1.0,
   longitude       = -1.0,
   elevation       = -1.0,
   flux_height     = -1.0,
   maxgoodqc       = 3,
   verbose         = .false.
   /
```

General Site Information	
Site_id:	US-Dk3
Site_name:	Duke Forest - loblolly pine
Tower_team:	PI: Gaby Katul - gaby@duke.edu - Duke University Affiliate: Paul Stoy - paul.stoy@montana.edu - Montana State University Affiliate: Ram Oren - ramoren@duke.edu - Duke University
Latitude:	35.9782
Longitude:	-79.0942
Elevation(m):	163.00
Network	AmeriFlux
IGBP:	ENF (Evergreen Needleleaf Forests)
Climate_Koeppen:	Cfa (Humid Subtropical: mild with no dry season, hot summer)
Mean Annual Temp (degrees C):	14.36
Mean Annual Precip. (mm):	1170
Years Of Data Available:	Ameriflux: 11 (Duration: 1998 - 2008) FLUXNET: 5 (Duration: 2001 - 2005)
Description:	The site was established in 1983 following a clear cut and a burn. Pinus taeda L. (loblolly pine) seedlings were planted at 2.4m by 2.4m spacing and ecosystem development has not been managed after planting. Canopy height increased from 16m in 2001 to 18m in 2004. The canopy is comprised primarily of P. taeda with some emergent Liquidambar styraciflua L. and a diverse and growing understory with 26 different woody species of diameter breast height 42.5 cm. The flux tower lies upwind of the CO2-enriched components of the free atmosphere carbon enrichment (FACE) facility located in the same pine forest. EC instrumentation is at 20.2m on a 22m tower.
Acknowledgment:	
Site image(s):	
VE Cutouts:	Max Zoom :: Mid 1 Zoom :: Mid 2 Zoom :: Min Zoom

File formats

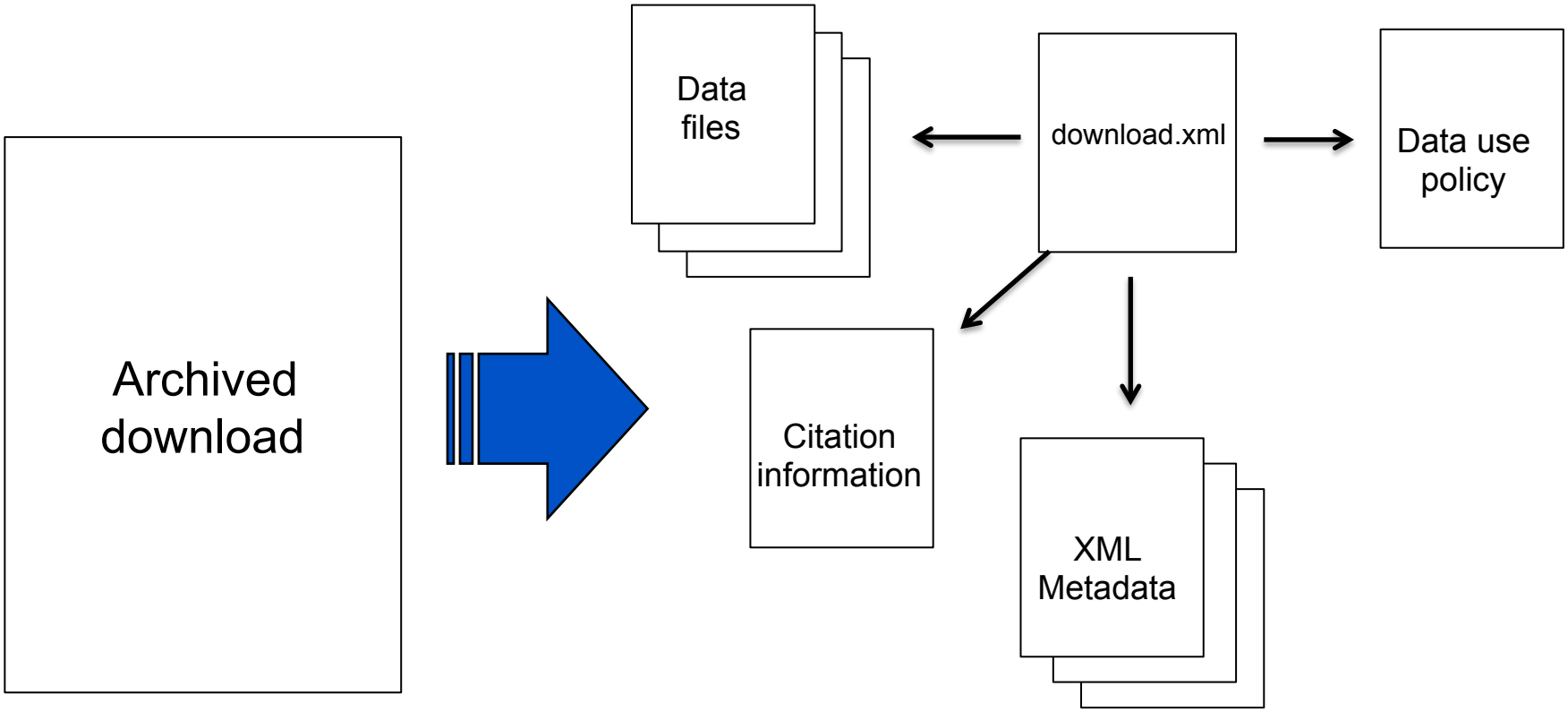
HDF5 format schematic



CSV alternative

dataProduc	domainID	siteID	startDate	endtDate	mean	standardEr	count	finalQF	min
NEON.D03.	D10	RS1	2013-03-24	2013-03-24	256.1836	0.0101	60	0	253.5888
NEON.D03.	D10	RS1	2013-03-24	2013-03-24	250.4642	0.0103	60	0	247.5259
NEON.D03.	D10	RS1	2013-03-24	2013-03-24	244.9080	0.0101	60	0	241.8125
NEON.D03.	D10	RS1	2013-03-24	2013-03-24	239.6961	0.0111	60	0	237.7858
NEON.D03.	D10	RS1	2013-03-24	2013-03-24	233.7573	0.0114	60	0	229.2461
NEON.D03.	D10	RS1	2013-03-24	2013-03-24	226.0407	0.0110	60	0	223.6695

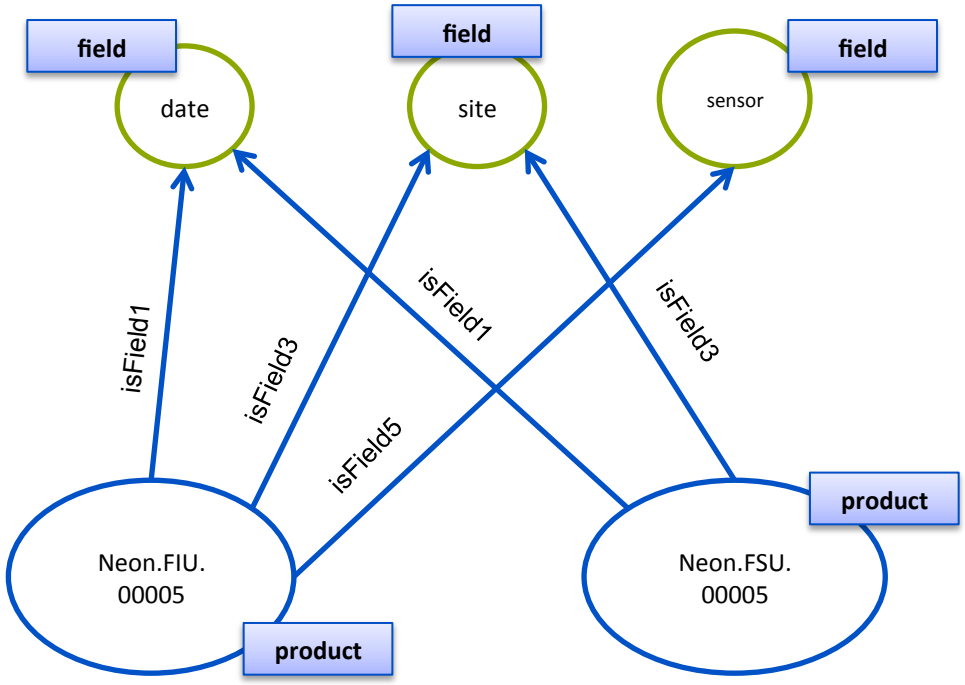
User payload



Managing terms

Graph database of terms, data products, and metadata

Oracle SQL DB of measurements

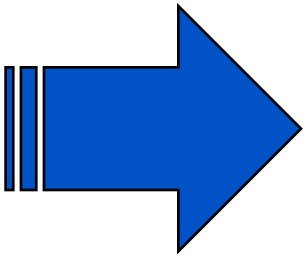


Linking NEON data and DART

Data terms managed in GraphDB

Standardized formats across heterogeneous measurements

Standardized user payloads w/ metadata

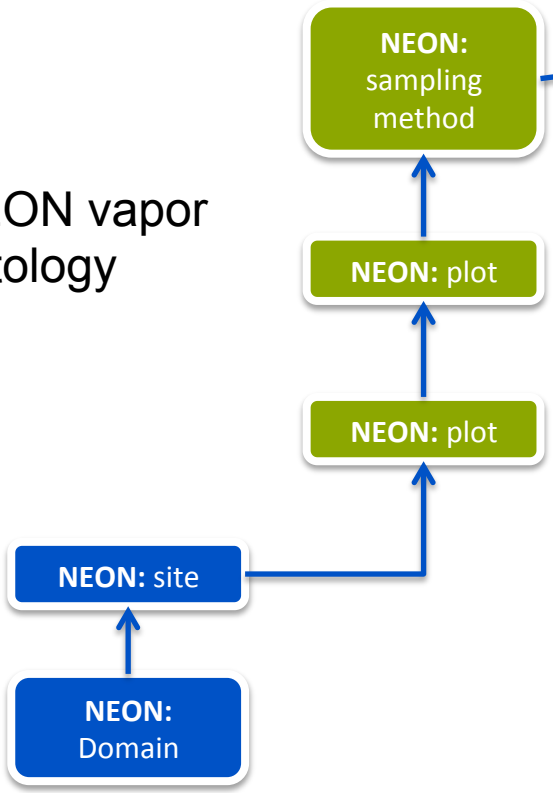


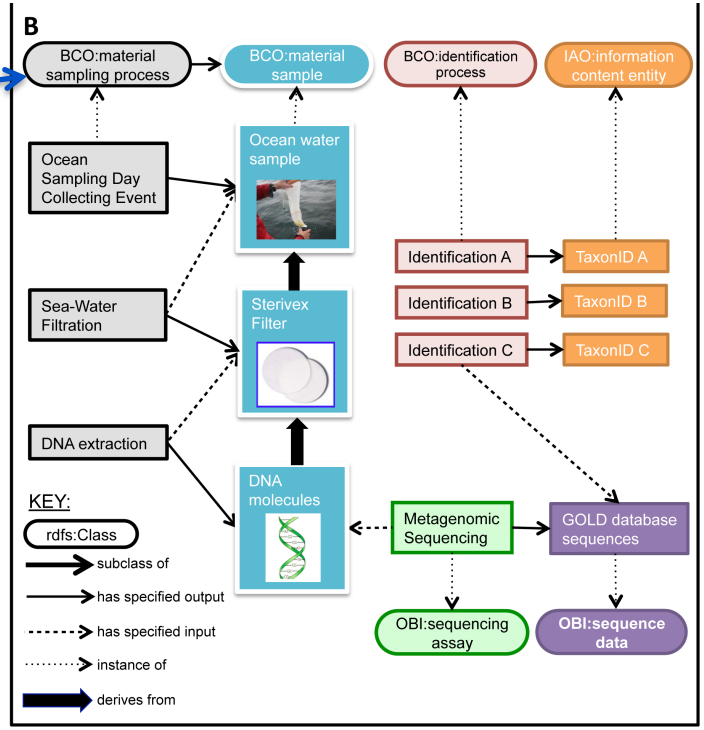
Programmatic pipelines into DART via NEON data tools

The future….

- Semantic ontologies describe NEON observation process
- Link with developing and existing community standards
- SPARQL endpoints and API's allow magic!

NEON vapor ontology

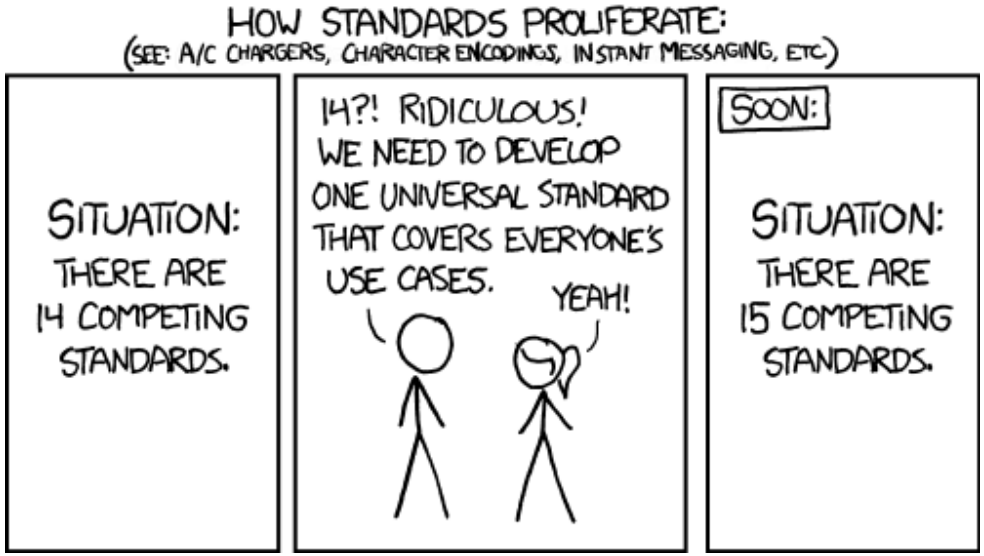




Walls et al, 2014, Figure 4B

Thanks

- Andy Fox
- Steve Berukoff
- Brian Miles



National Ecological Observatory Network, Inc.

NCAR
NATIONAL CENTER FOR ATMOSPHERIC RESEARCH

@distribecology

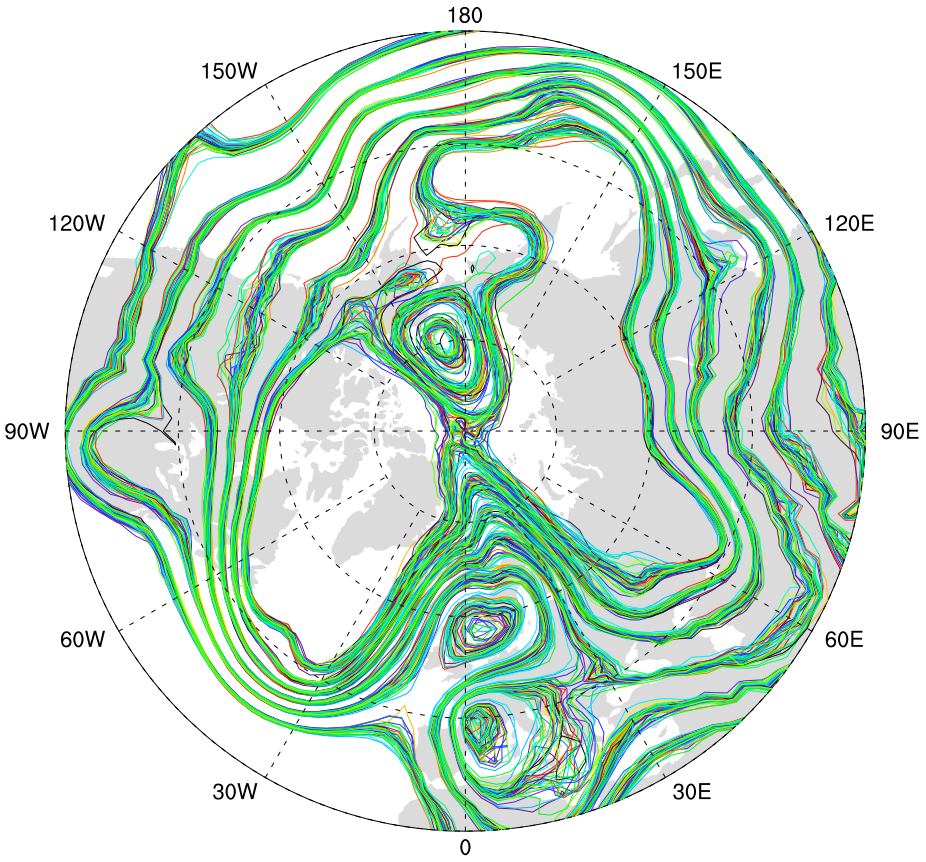
@emhart

Observing system simulation experiment

- 80 member, 6 hourly climate reanalysis available, 1998 – 2010
- Each forces separate CLM ensemble member at 1° x 1°
- Generates spread in the land model states
- At 60 NEON sites observe:

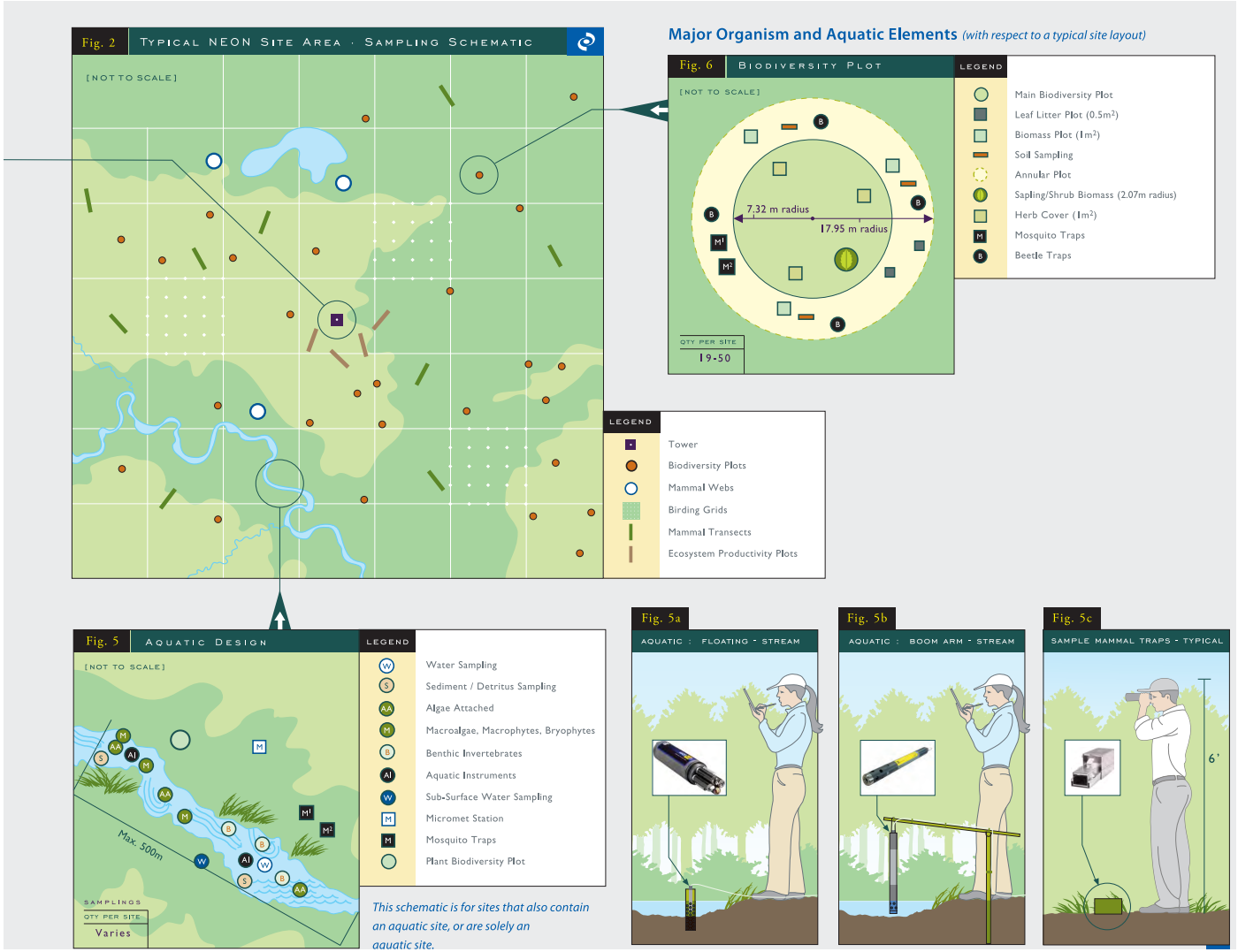
i. **Leaf area index**
ii. **Leaf nitrogen concentration**
iii. **Net Ecosystem Productivity**
iv. **Evapotranspiration**

- **175,000 observations a month**

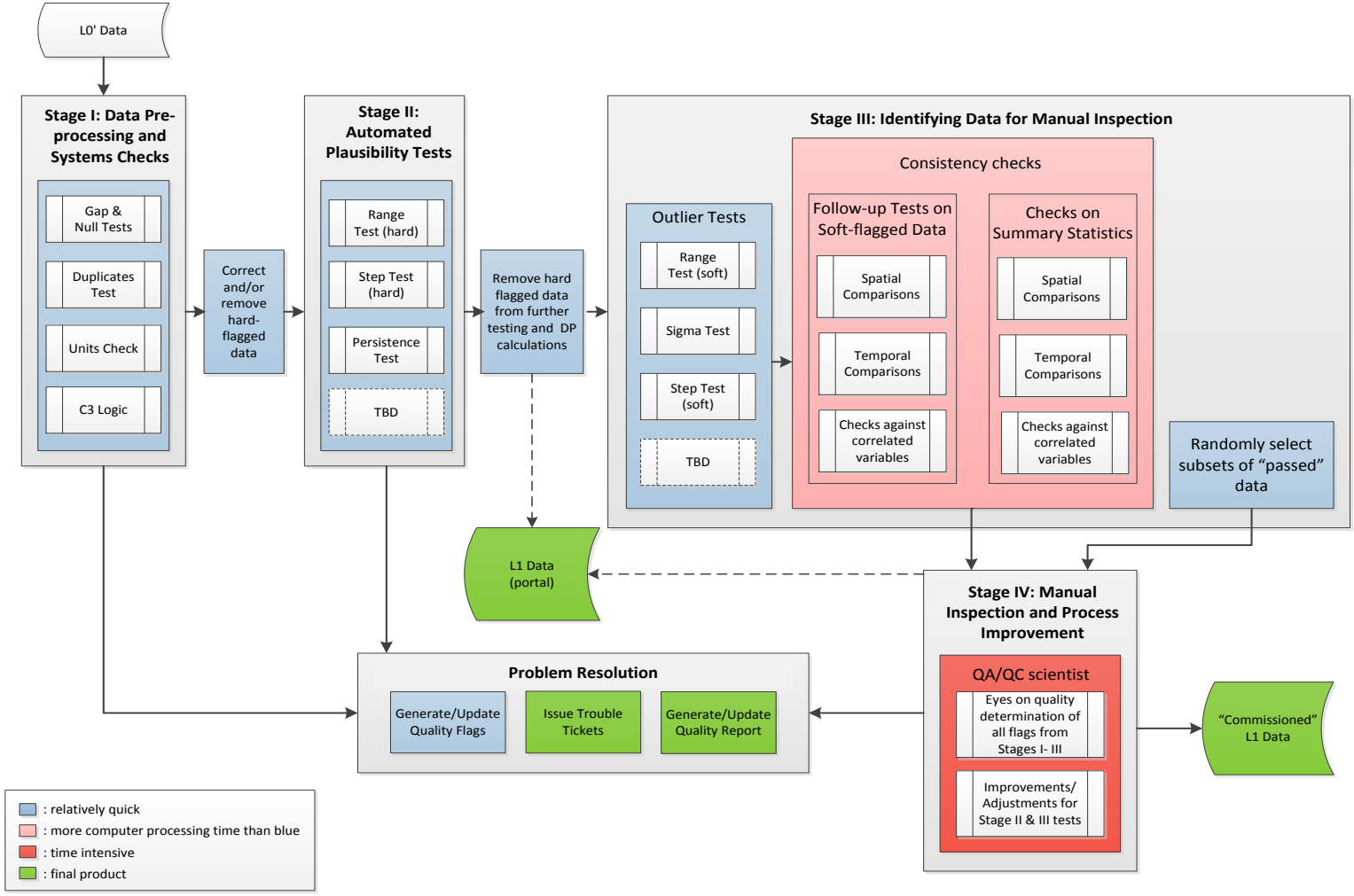


500 hPa GPH
Feb 17 2003

A NEON site



Automated tower data collection



NEON Tower measurements

NEON site at Ordway-Swisher Biological Station (Photo Credit: NEON)