The parallel system for integrating impact models and sectors (pSIMS)

Neil Best

XSEDE13, San Diego

AgMIP The Agricultural Model Intercomparison and Improvement Project

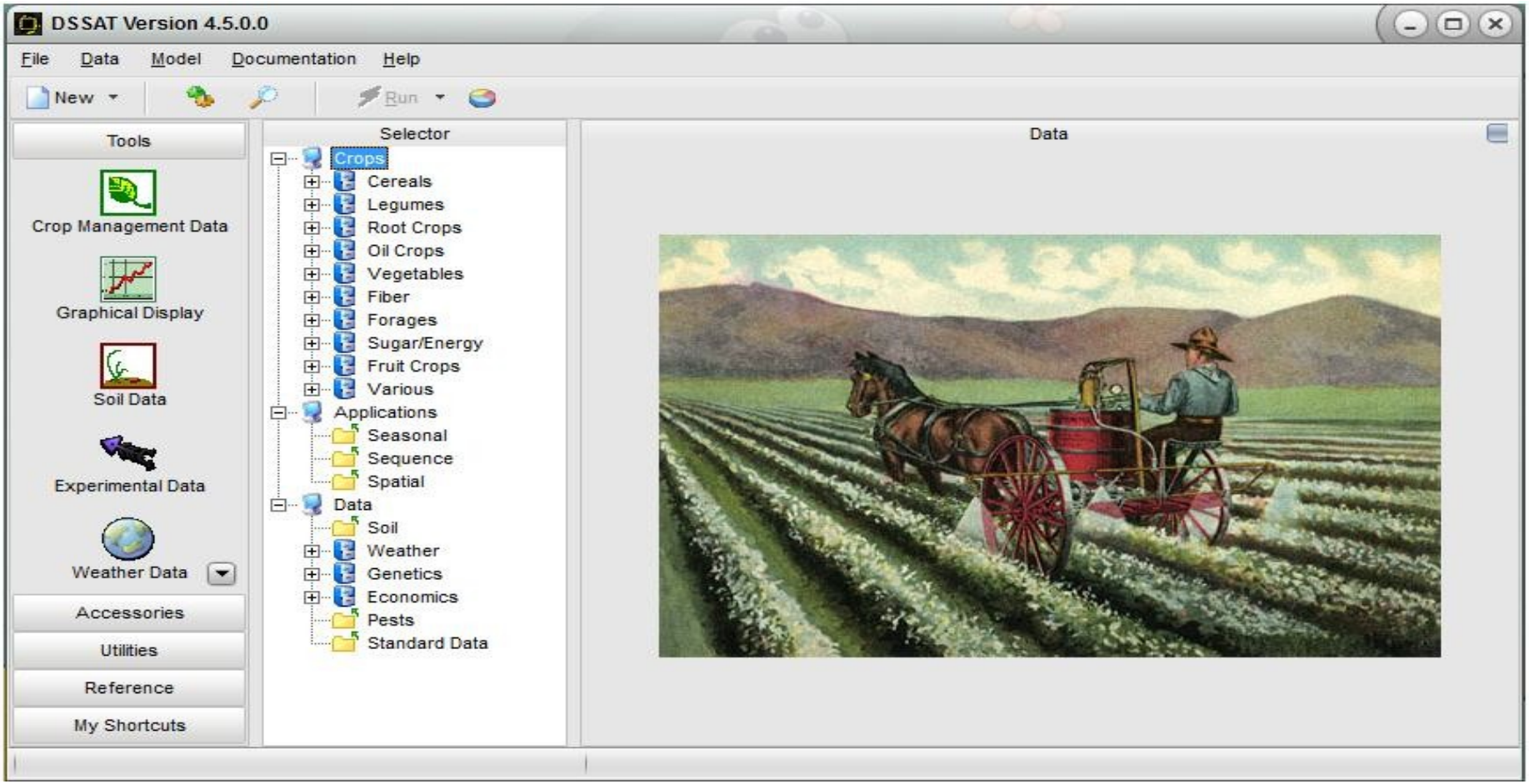
ISI-MIP

pSIMS contributors, collaborators and users

- pSIMS developers
  - Joshua Elliott (funded to work on pSIMS through NSF RDCEP center and SEES fellowship)
  - Michael Glotter (funded to work on pSIMS through NSF graduate student fellowship)
  - Neil Best (funded to work on pSIMS through NSF RDCEP, DOE, and Monsanto)
  - David Kelly (funded to work on pSIMS through NSF RDCEP center)
  - James Chryssanthacopoulos (Columbia U; funded to work on pSIMS through USAID)
  - Kanika Jhunjhnuwala (NZ Landcare; funded to work on pSIMS through New Zealand)
- Close collaborators, contributors, and users of pSIMS outputs
  - NASA GISS/Columbia: Cynthia Rosenzweig and Alex Ruane; University of Florida: Jim Jones, Ken Boote, Senthold Asseng, Cheryl Porter; Iowa State: Jerry Hatfield, Fernando Miguez; IFPRI: Jerry Nelson, Ricky Robertson; NZ Landcare: James Lennox, Adam Daigneault; PNNL JGCRI: Page Kyle, Cesar Izaurralde, Kate Calvin; U. East Anglia: Delphine Deryng
- ISI-MIP collaborators
  - C. Muller, K. Frieler, M. Konzmann, D. Gerten, J. Heinke, C. Gellhorn, C. Schmitz, V. Huber, F. Piontek, L. Warszawski, J. Schewe, A. Levermann, H. Lotze-Campen, H.J. Schellnhuber, M. Mengel, S. Ostberg, S. Schaphoff: PIK; M. Florke, S. Eisner: U. of Kassel; Y. Wada, M.F.P. Bierkens: U. of Utrecht; C. Folberth, H. Yang: EAWAG; S.N. Gosling: U. of Nottingham; I. Haddeland: NVE; P. Havlik, N. Khabarov, M. Lomas, H. Valin: IIASA; W. Franssen, F. Ludwig, H. van Meij: Wageningen; S. Fujimori, Y. Masaki, K. Nishina: NIES, Japan; A. Arneth, S. Olin: Lund; H. Kim , Y. Satoh: U. of Tokyo; E. Schmid, E. Stehfest: Boku; T. Stacke: Max Planck Inst. for Meteorology; R. Pavlick: Max Planck Inst. for Biogeochemistry; Q. Tang: IGSNRR; D. Wisser: U. of Bonn; P. Ciais: LSCE; P. Doll: Goethe U.; P. Falloon, B. Fekete, D. Hemming: Hadley Centre; A.D. Friend, T. Oki: U. of Cambridge; K. Neumann: PBL; T. Pugh: Karlsruhe Inst. of Technology; D.B. Clark: Centre for Ecology & Hydrology; F. Colon-Gonzalez, A.M. Tompkins: Abdus Salam International Centre for Theoretical Physics; A. Morse: U. of Liverpool; Z. Tessler: CUNY; H. Ahammad, E. Heyhoe: DAFF; M. Von Lampe: OECD; D. Mason d'Croz: IFPRI; A. Tabeau, H. van Meijl: LEI; D. van der Mensbrugghe: FAO; R. Sands: USDA; D. Willenbockel IDS

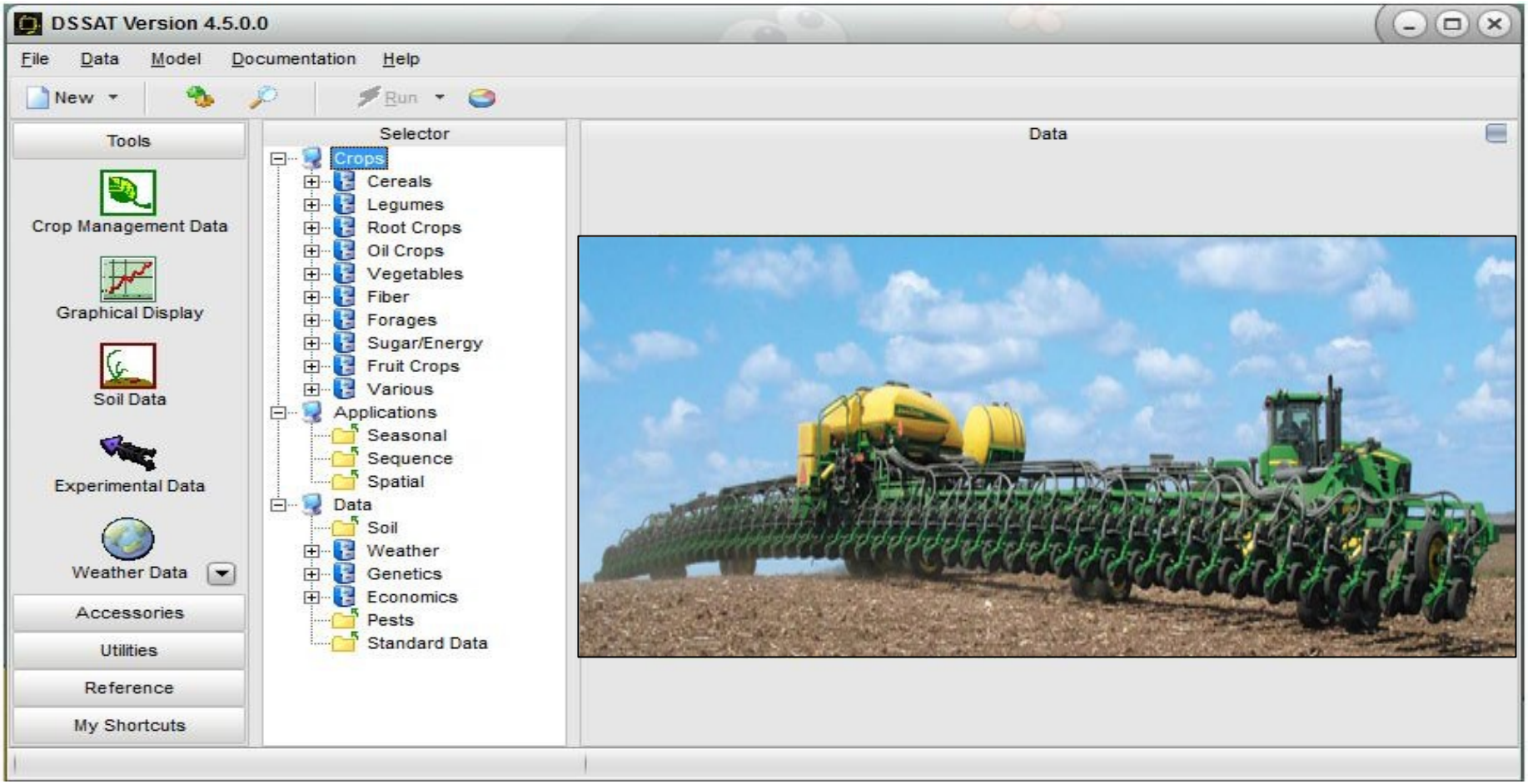
Traditional field-scale crop models

Decision Support System for Agrotechnology Transfer (DSSAT)
is a software application program that comprises crop simulation models for over 28 crops

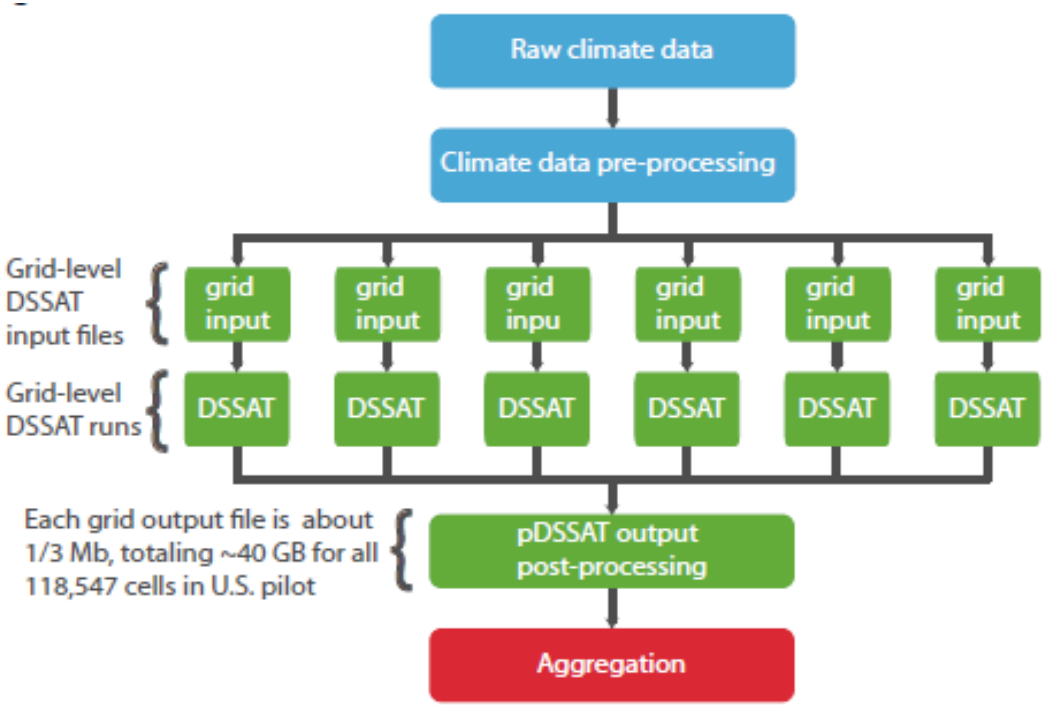


High Performance Computing

Goal: Leverage high-performance computing to facilitate large-scale crop yield and climate impact assessments without sacrificing model detail or representational complexity.



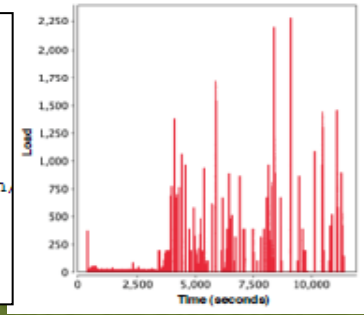
Prototype: parallel DSSAT (pDSSAT)

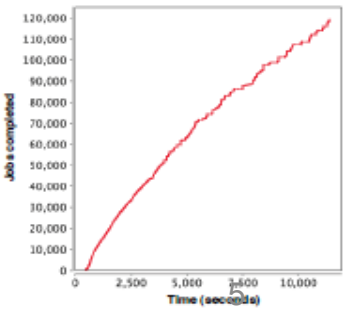


- Simulations are parallelized with the Swift parallel scripting language.
- Typical campaign involves 10^5 - 10^6 individual simulation calls
- Input data for each campaign runs 100 Gb – 1 Tb
- Highly scalable w/ utilization of >4k simultaneous compute cores typical.
- Actively running on many systems
  - **UC3 – 2k core open UChicago cluster**
  - **Beagle – 151 teraflop Cray XE6 supercomputer**
  - **OSG – distributed SC network; users leveraging 1M CPU-hours/day**

A simulation "campaign" is specified by a parameter file:

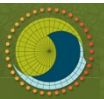
```
model               dssat45
scenarios           /scratch/01503/davidkel/psims/data/scenarios/rcp8p5_mai
weather             /scratch/01503/davidkel/psims/data/wth/HadGEM2-ES/rcp8p5
refdata             /scratch/01503/davidkel/psims/data/common
bindata             /scratch/01503/davidkel/psims/bin
bintransfer
/scratch/01503/davidkel/psims/bin/DSCSM045.tar.gz,/scratch/01503/davidkel/psims/bin,
executable          DSCSM045.EXE:A:X1234567.MZX
outtypes            .OUT,.MZX
postprocess         ./gen_part_dssat.pl:--num_scenarios:20:--years:150:--
variables:PDAT,ADAT,MDAT,CWAM,HWAM,IRCM,PRCM,ETCM:--crop:mai
```





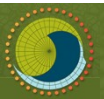
Coordinated multi-sector multi-model climate impact simulation

- Must support
  - dozens of observational weather, reanalysis, and climate model based data products through standardized input data
    - created versatile .psims.nc format and data pipeline for ingest of arbitrary data formats
  - simulations with multiple models and sectors driven by the same data (coordinated)
    - created easy campaign specification interface and high-performance modular compute system
  - processing, standardization, and aggregation of large amounts of custom model outputs
    - creating standardized output formats and methodologies for aggregation of impact measures to arbitrary administrative and environmental boundaries

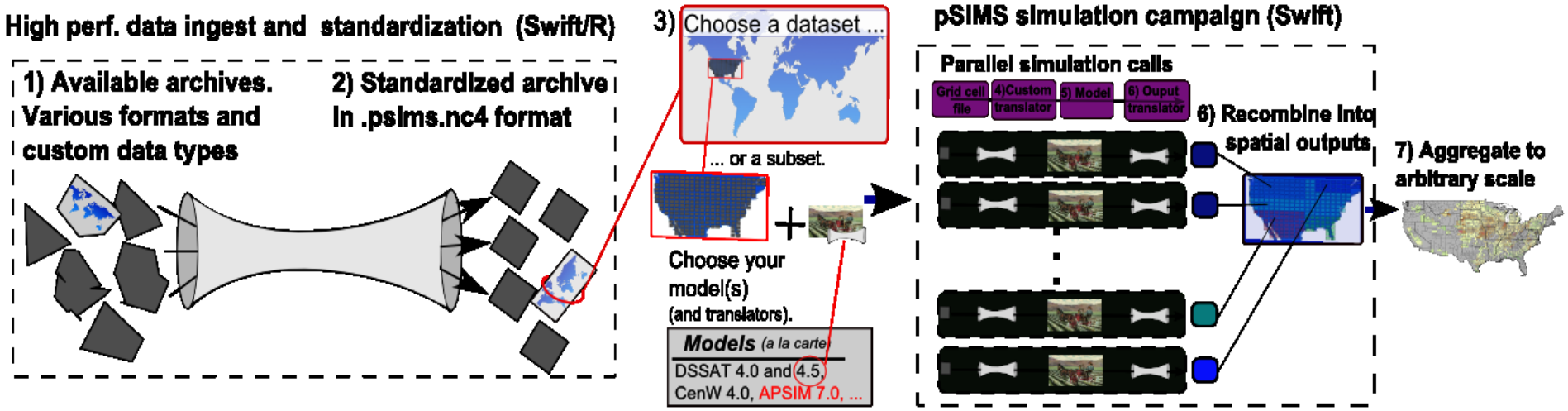


Coordinated multi-sector multi-model climate impact simulation

- Support ***research use cases*** at a wide variety of
  - skill levels:
    - from undergrad students
    - to skilled researchers with development teams and large-scale resources
  - problem sizes:
    - from county, state, or national level crop-climate impact assessments
    - to global scale multi-GCM multi-model ensemble simulations with dozens of species



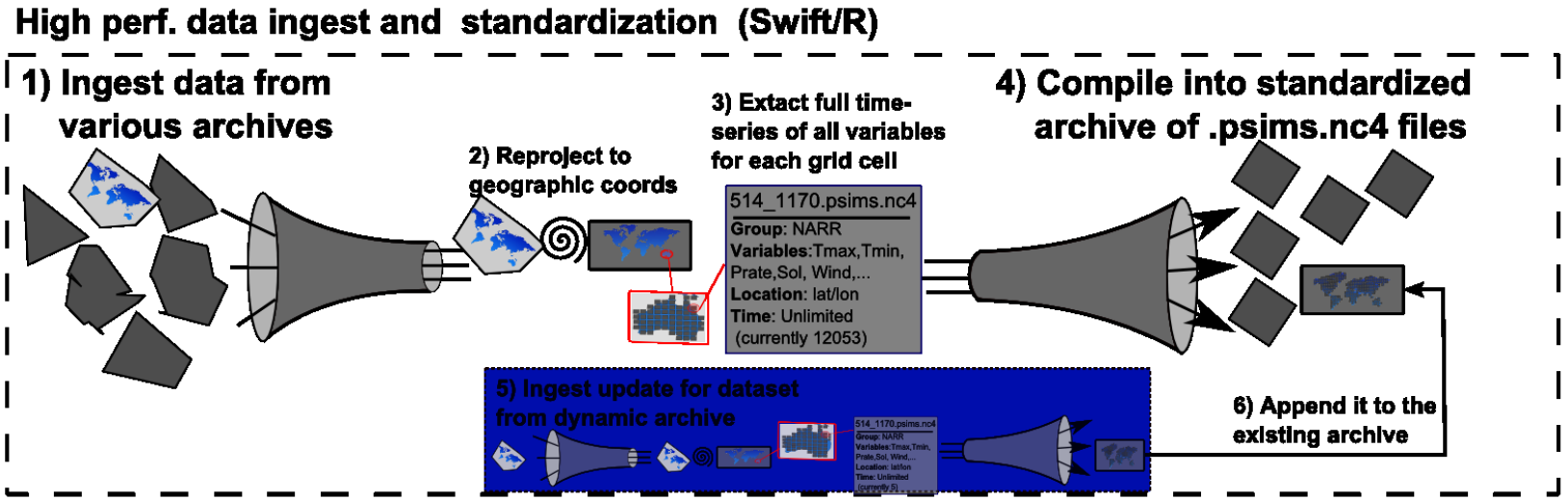
The parallel System for Integrating Impacts Models and Sectors (pSIMS)



- Model choices available so far include
  - DSSAT v4.0 and 4.5 (currently supports ~28 distinct crop species)
  - APSIM v7.5 (currently supports ~30 crop, grass, and tree species)
  - CenW 4.0 (currently supports ~5 tree and grass species)
  - *Expected to include some support for EPIC before release in 2014.*

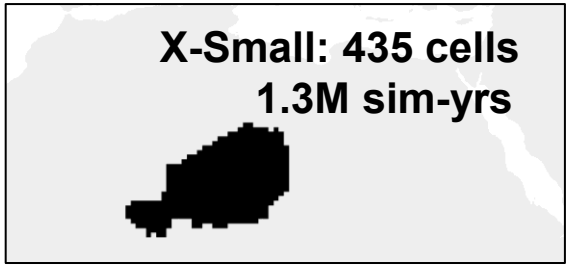
Framework in use by three groups at U Chicago, Columbia U, and New Zealand Landcare. Open code release planned early in 2014.

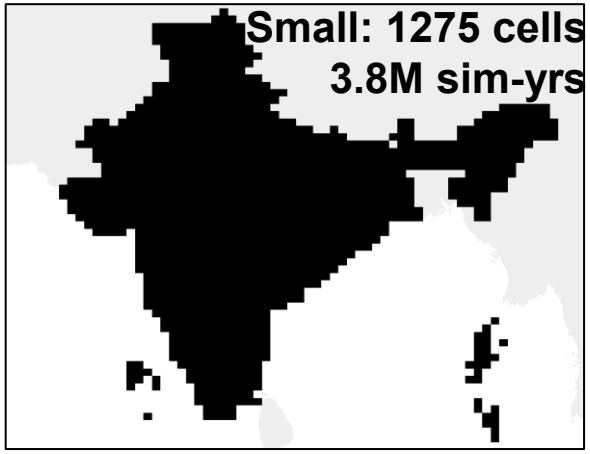
pSIMS input data pipelines and archive

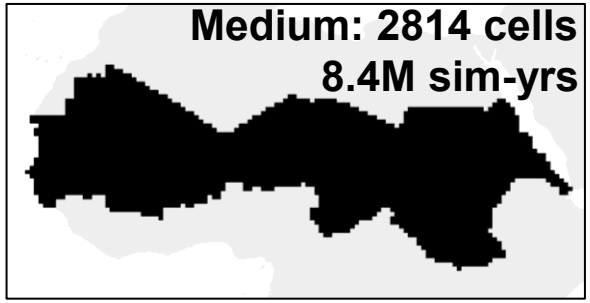


- **Challenge**: transpose data from large spatial rasters to single point time-series required for model runs.
- **Solution**: create new data standard (.psims.nc)
  - optimized for point-based impact models that require daily weather
  - without sacrificing the benefits of formats like netcdf
    - efficient compression, self-describing files, embedded meta-data, and big actively developed low-level operator toolkits like NCO and CDO.
  - open source code base for high-performance ingest and update of data sources using R and Swift.

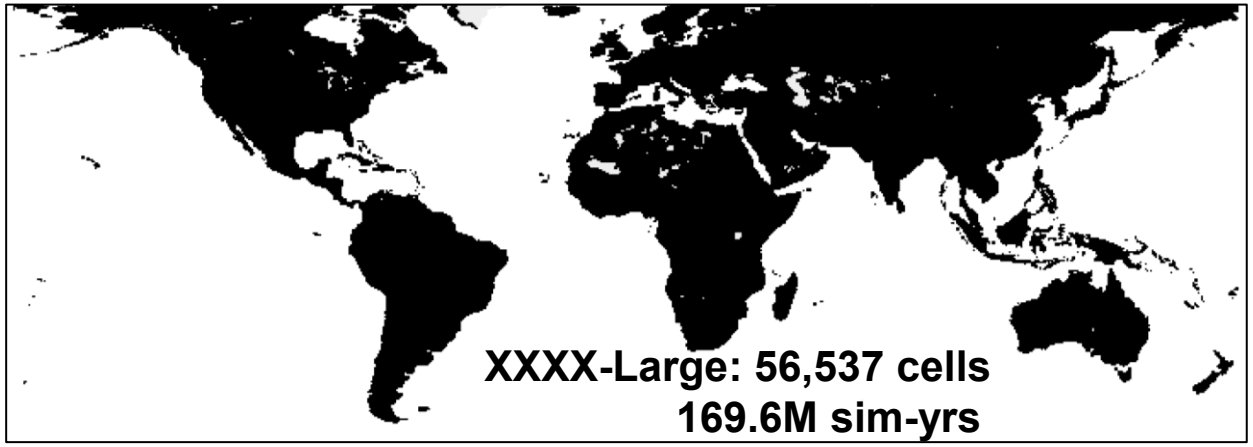
Versatile design of pSIMS "campaigns" allow for assessments at any scale





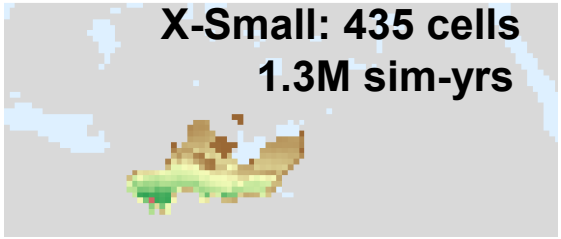


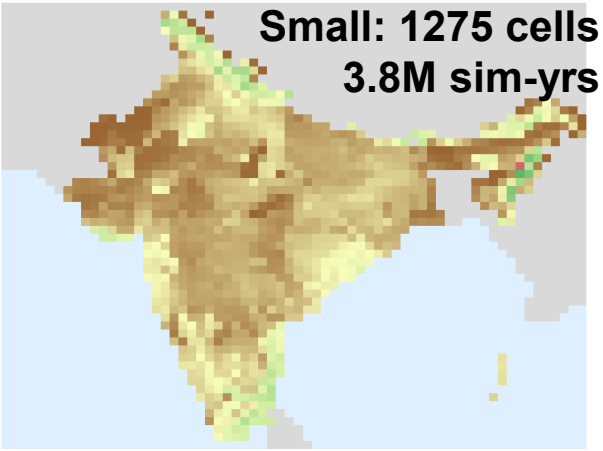
- For each cell we might run 20 management scenarios (or treatments) for 150 years (1950-2099) each, varying irrigation, fertilizer application rates, cultivar choice, and CO2 effects, for a total of 3000 sim-yrs per cell.

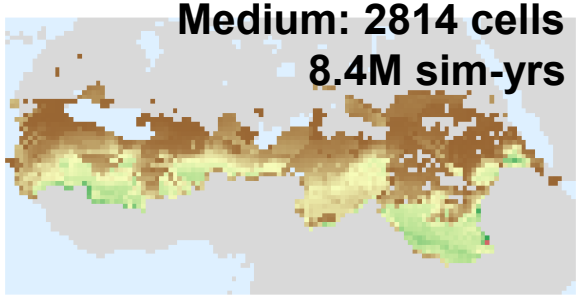


Global campaign run in 1 hour on O(1k) cores of a modern HPC resource, takes >1 month if run in serial.

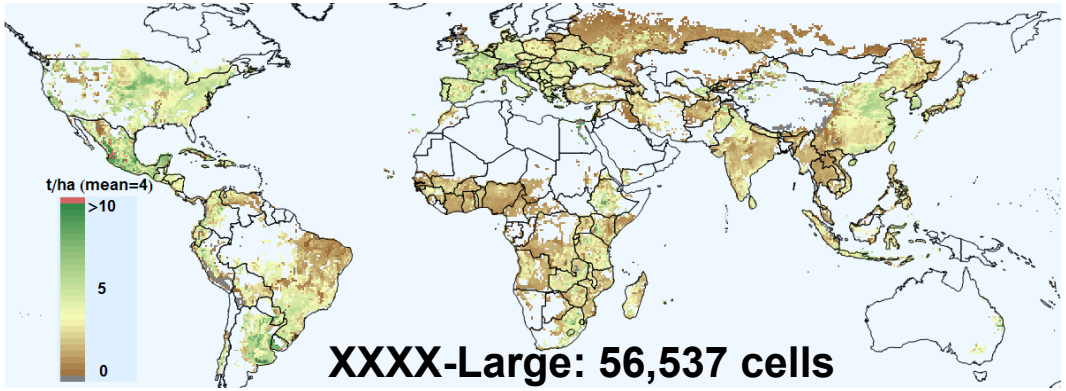
Versatile design of pSIMS "campaigns" allow for assessments at any scale







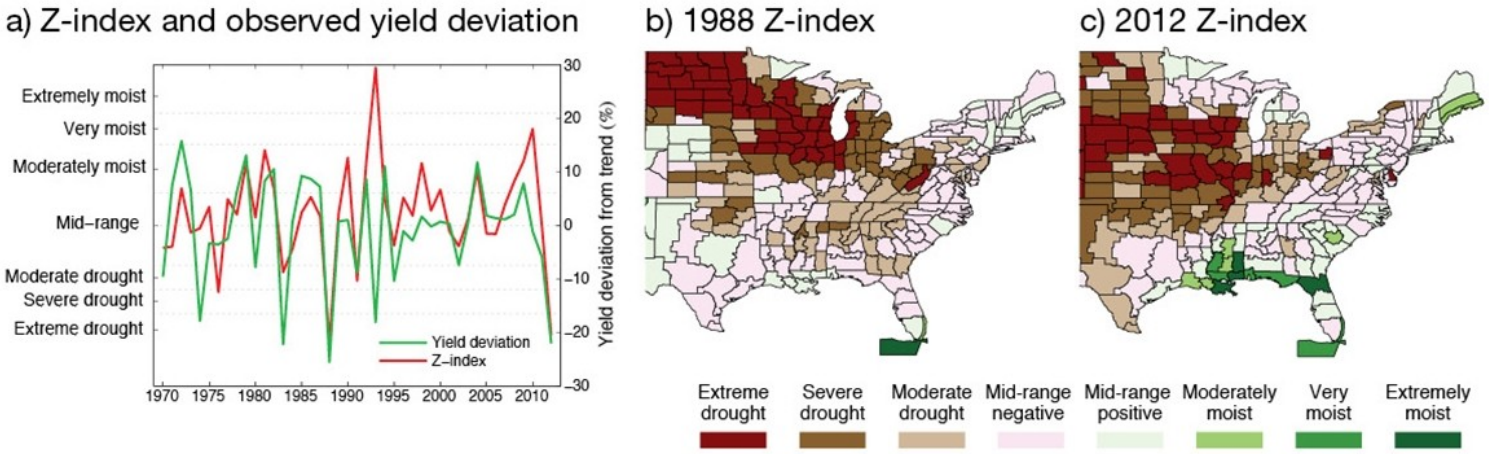
- For each cell we might run 20 management scenarios (or treatments) for 150 years (1950-2099) each, varying irrigation, fertilizer application rates, cultivar choice, and CO2 effects, for a total of 3000 sim-yrs per cell.



Global campaign run in 1 hour on O(1k) cores of a modern HPC resource, takes >1 month if run in serial.

Predicting 2012

- pSIMS as a probabilistic forecast tool for decision support, risk management, and counterfactual analysis at seasonal time-scales.
- Can mechanistic crop models be used to “Forecast” drought year yields (like 2012)?
- Compare 2012 drought in detail with 1988

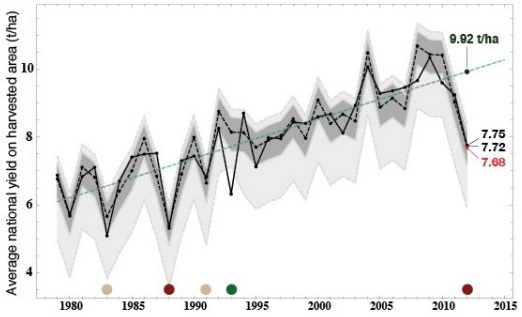


Elliott, J., M. Glotter, N. Best, A.C. Ruane, K.J. Boote, J.W. Jones, J.L. Hatfield, C. Rosenzweig, L. A. Smith, and I. Foster. **Predicting and adapting to the agricultural impacts of large-scale drought** (2013). Submitted/ RDCEP WP series.

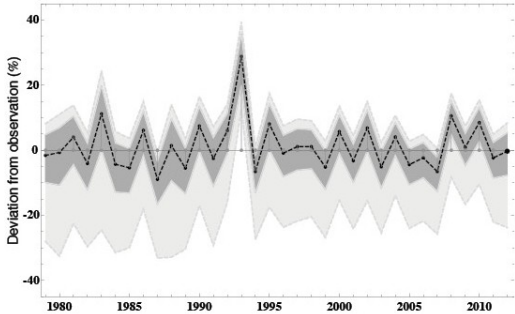
Predicting 2012

- Hindcast 1979-2011 with detailed parameterizations of technology
- Counterfactual of 1988 with 2012 weather and 2012 with 1988 weather, shows that 2012 drought was notably more severe.
- ~5 percent (of trend value) better/worse in 2012/1988 if they had 1988/2012 weather.
- Accuracy at state level comparable to USDA NASS forecasts.

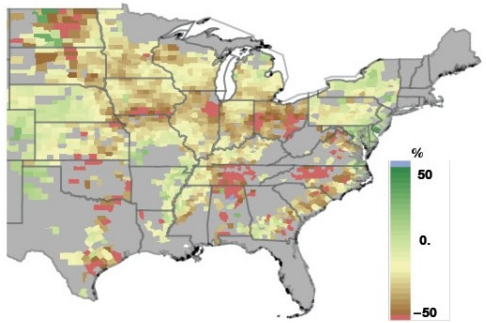
a) Observed and simulated yields



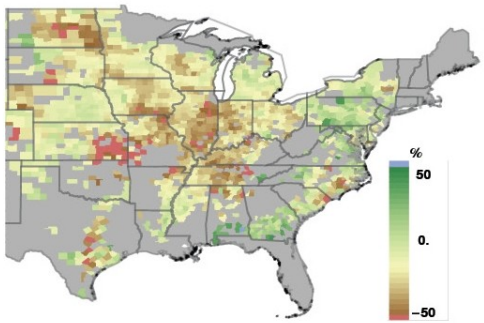
b) Deviation of simulated yields



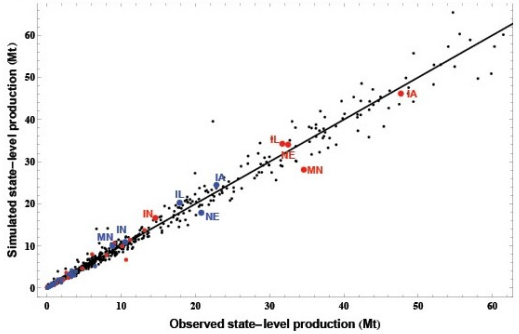
c) Predicted 1988 deviation from county trend



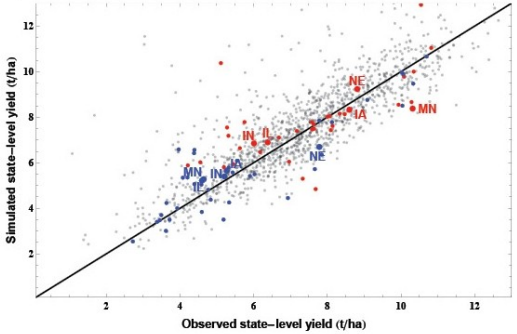
d) Predicted 2012 deviation from county trend



a) Observed and simulated production



b) Observed and simulated yield



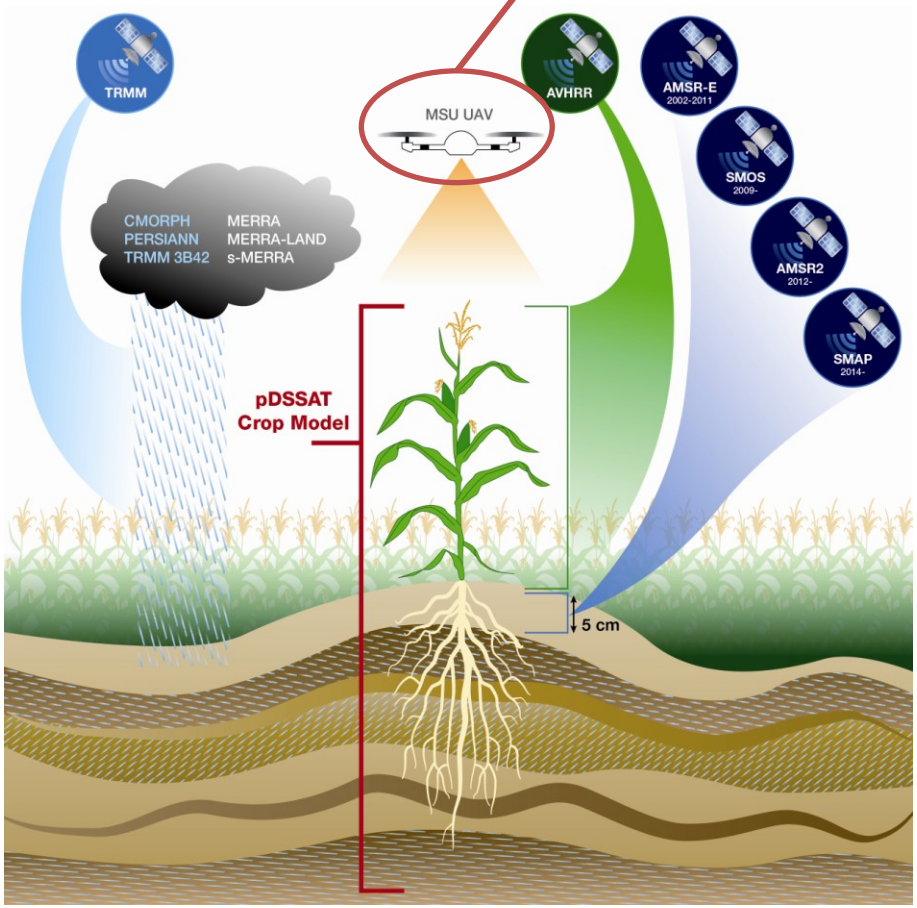
Elliott, J., M. Glotter, N. Best, A.C. Ruane, K.J. Boote, J.W. Jones, J.L. Hatfield, C. Rosenzweig, L. A. Smith, and I. Foster. **Predicting and adapting to the agricultural impacts of large-scale drought** (2013). Submitted/ RDCEP WP series.

2013/14: Regional! More models!

- Finalizing pAPSIM and prototype pEPIC for coordinated multi-model simulations in pSIMS.
- Developing new regional studies
  - India (with Andy Challinor),
  - China (with Hong Yang),
  - Africa (with Lyndon),
  - South America (with Cesar Izaurralde),
  - North America (with Ken Boote, Claudio Stockle, Cesar, etc.).

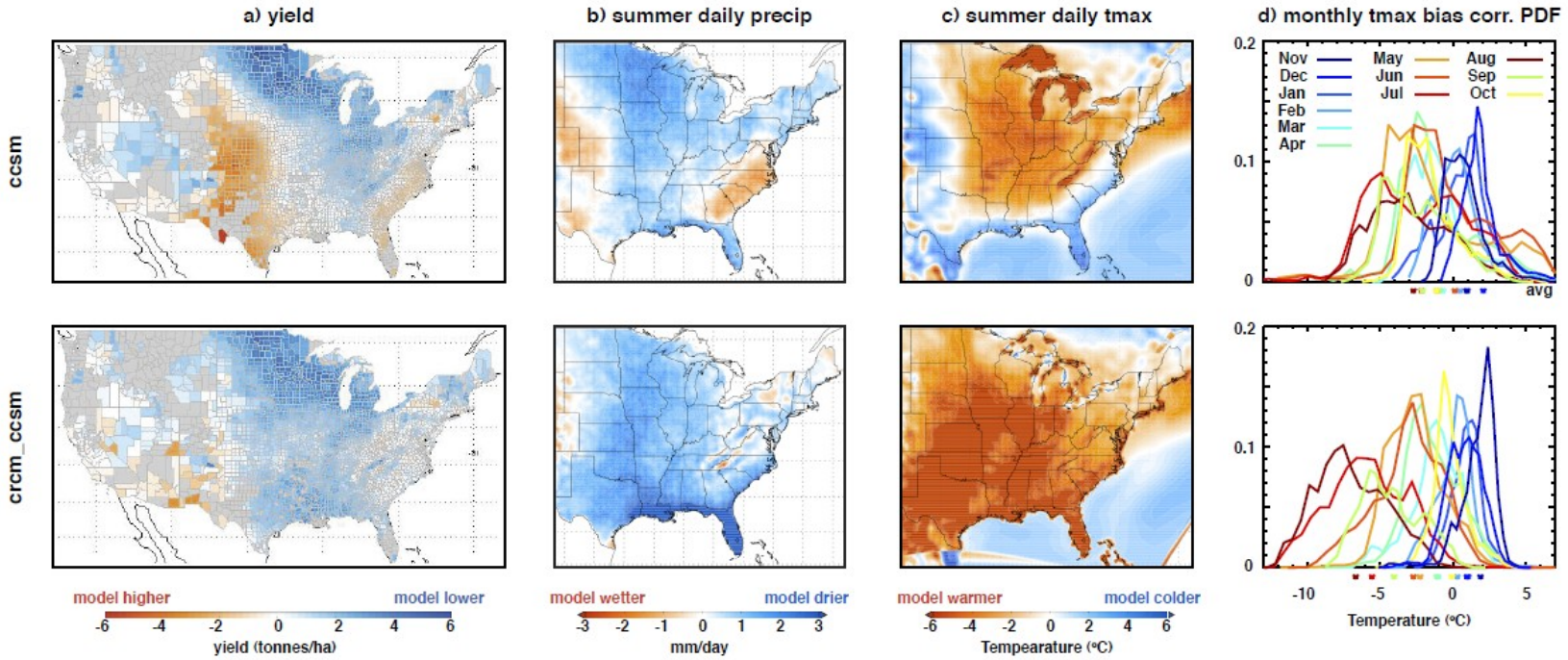
2013/14: Water!

- Develop special focus projects on climate extremes, water, and technology
- Both in general and as part of the regional projects (mostly around drought)



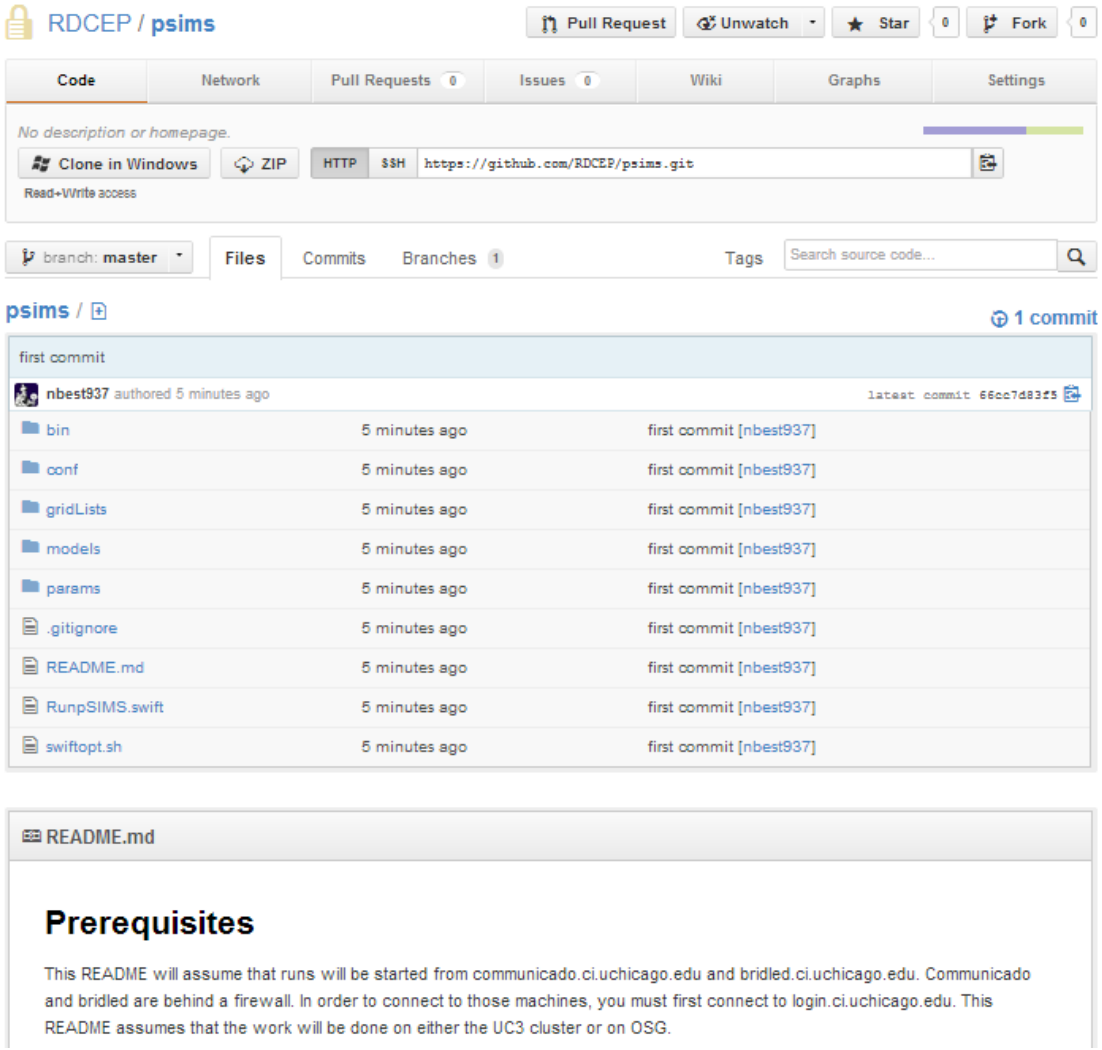
Crop models as a filter for comparing climate data products

- RCMs tend to “fix” errors near topographical regions
- Most differences eliminated after applying a bias correction
- Often makes other problems worse (for complicated reasons)
- Suggest comp. expense of RCMs may not be warranted for ag impacts



pSIMS software tools and outreach

- pSIMS source code managed on github
- 3 model choices supported (pDSSAT versions 4.0 and 4.5, pCenW, and APSIM)
- 4 crops tested in pDSSAT and 2 tree species tested in pCenW
- Coming soon: more crops, more models, new data sources for soil and weather, …



pSIMS in the iPlant Discovery Environment and Galaxy Science Portal

- Working with folks and CI and TACC, we have prototyped the inclusion of pSIMS into the iPlant DE and Galaxy
- These portals enable users without expertise to use pSIMS with various HPC resources