Overview of the Office of Biological and Environmental Research

NERSC BER Requirements Large Scale Computing and Storage Requirements for Biological and Environmental Research

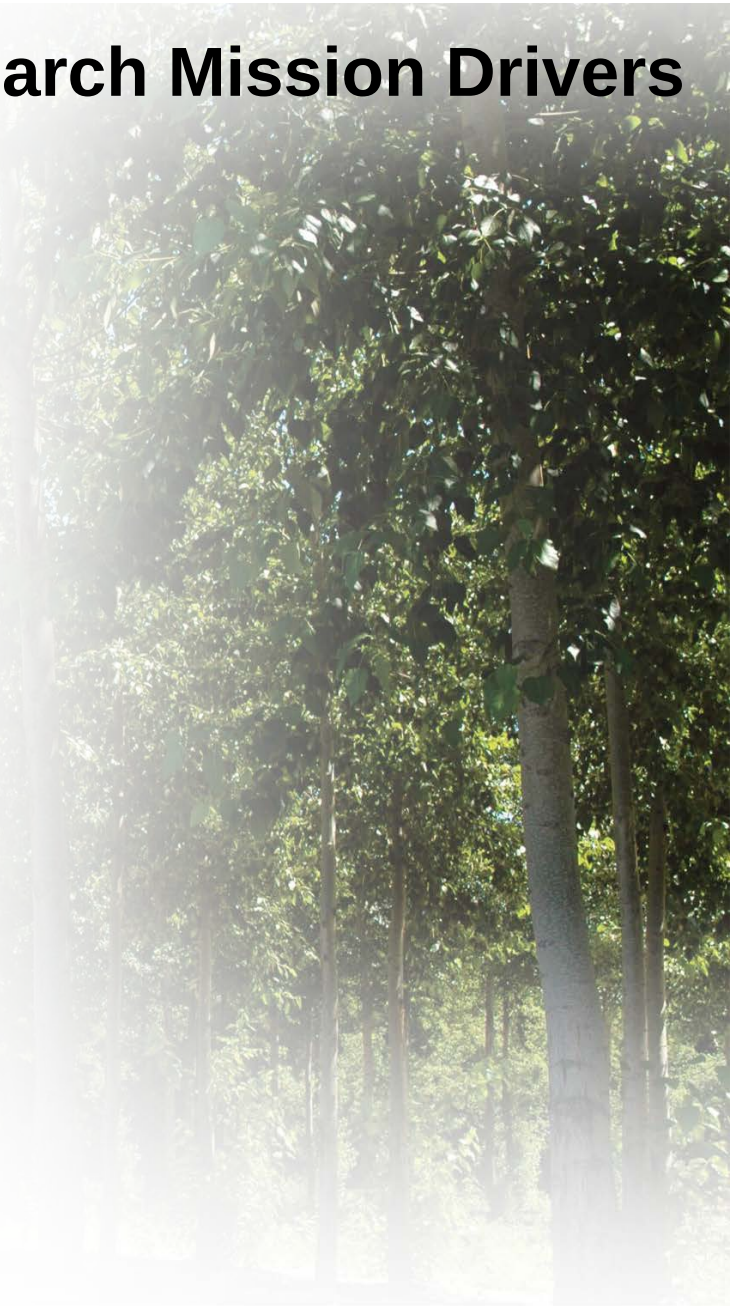
September 11-12, 2012
Rockville, MD

Susan K. Gregurick, Ph.D.
Program Manager, Computational Biosciences and Knowledgebase
Biological and Environmental Research

Biological and Environmental Research Mission Drivers

Provide the foundational science for:

- Supporting the development of biofuels as major, secure, and sustainable national energy resources
- Understanding potential effects of energy-related greenhouse gas emissions on Earth's climate and biosphere
- Predicting and controlling the cycling and mobility of materials in the subsurface and across key surface-subsurface interfaces in the environment
- Developing new tools to explore the interface of biological and physical sciences



Biological and Environmental Research Subprograms

Biological Systems Science

- Genomic Science
  - *Foundational Genomics*
  - *Genomics Analysis & Validation*
  - *Metabolic Synthesis & Conversion*
  - *Computational Bioscience*
  - o *Bioenergy Research Centers*
- Joint Genome Institute (JGI)
- Structural Biology Infrastructure
- Radiological Sciences

Climate and Environmental Sciences

- Atmospheric System Research
- Environmental System Science
  - *Terrestrial Ecosystem Science*
  - *Subsurface Biogeochemical Research*
- Climate & Earth System Modeling
- Atmospheric Radiation Measurement (ARM) Climate Research Facility
- Environmental Molecular Sciences Laboratory (EMSL)

Biological and Environmental Research

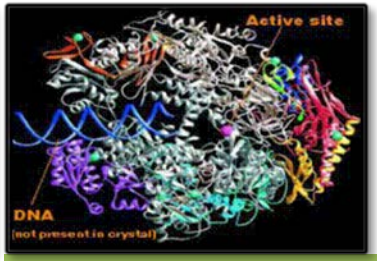
Understanding complex biological, climatic, and environmental systems across vast spatial and temporal scales

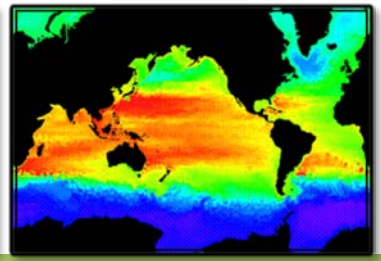
The Scientific Challenges:

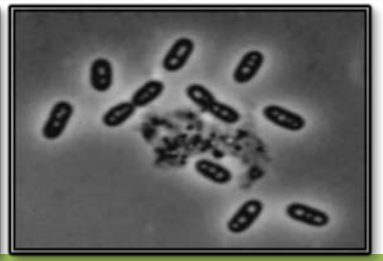
- Understand how genomic information is translated to functional capabilities, enabling more confident redesign of microbes and plants for sustainable biofuel production, improved carbon storage, or contaminant remediation.
- Understand the roles of Earth's biogeochemical systems (atmosphere, land, oceans, sea ice, subsurface) in determining climate so we can predict climate decades or centuries into the future, information needed to plan for future energy and resource needs.

FY 2013 Highlights:

- Clean energy biodesign of plant and microbial systems through development of new molecular toolkits for systems and synthetic biology research.
- Research and new capabilities to develop comprehensive environmental system models in the Arctic and tropics, regions especially vulnerable to rapid climate change.
- Continue support for the three DOE Bioenergy Research Centers, and operations of the Joint Genome Institute, the Environmental Molecular Sciences Laboratory, and the Atmospheric Radiation Measurement Climate Research Facility.

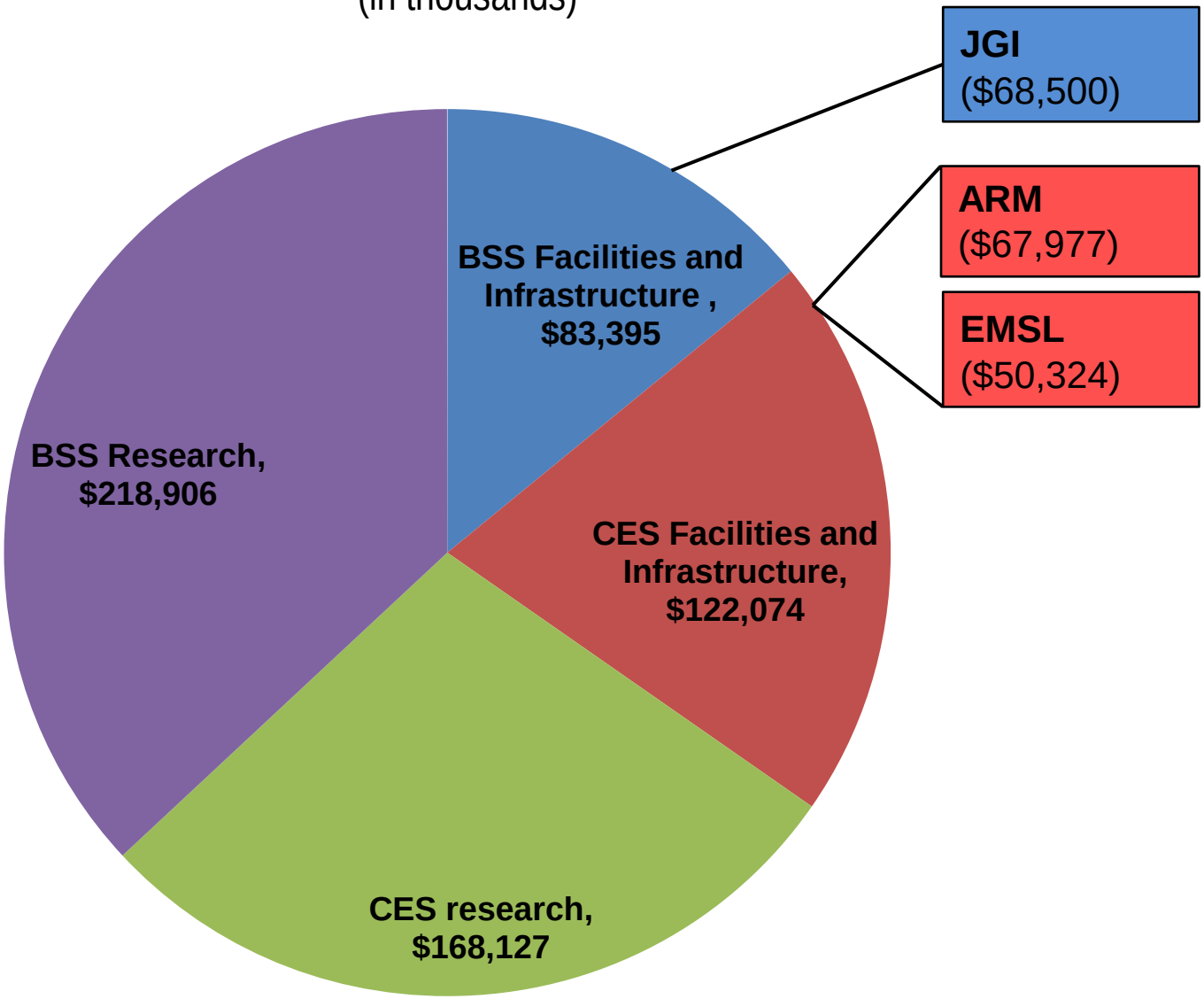


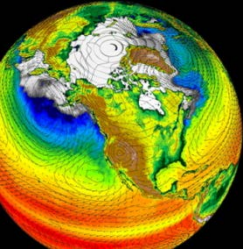


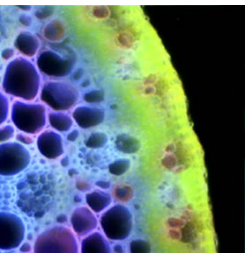
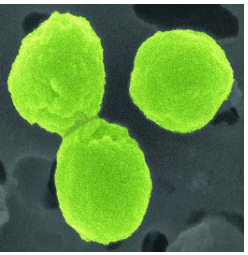


• FY 2012 BER Appropriation

(in thousands)

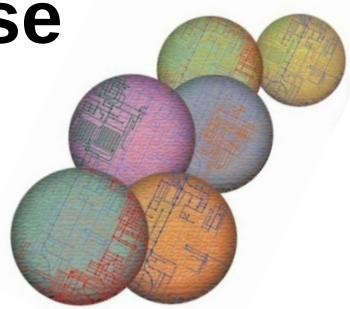




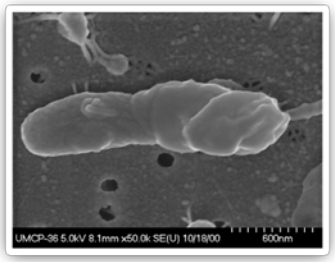




Biological Systems Science Division

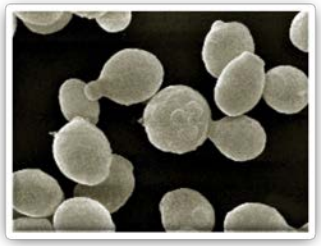
The Genomic Science Research Enterprise



- Foundational Genomics Research
  - Function and organization of complex biological (plant and microbe) systems; biodesign
- Genomics Analysis and Validation
  - Experimental validation and improvement of genome annotation and modeling



- Metabolic Synthesis and Conversion
  - Research on mechanisms and regulation of carbon storage in plant biomass or for sequestration



- Computational Biosciences
  - Enabling data integration and analysis with a systems biology knowledgebase
- Bioenergy Research Centers
  - Accelerate the development of clean and sustainable (bio)energy solutions

Bioenergy Research Centers

Highly focused, interdisciplinary, team-based transformational science with milestones:

BioEnergy Science Center (led by ORNL)

- Multi-institutional partnership with strategic focus on overcoming biomass recalcitrance as route to cost-effective cellulosic biofuels
- Consolidated bioprocessing

Great Lakes Bioenergy Research Center (led by U. WI, MSU)

- Re-engineering plants to produce more starches and oils
- Iterative experimental and modeling approaches
- Major research thrust on sustainability of biofuels

Joint BioEnergy Institute (led by LBNL)

- New pretreatment processes using room temperature ionic liquids
- Major research thrust on synthetic biology: re-engineering *E.coli* and yeast to produce hydrocarbons

jbei Joint BioEnergy Institute

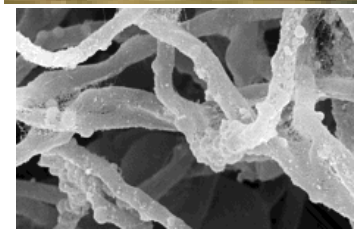
GREAT LAKES BIOENERGY RESEARCH CENTER

BESC BioEnergy Science Center

Joint Genome Institute

Meeting the DNA sequencing needs of the bioenergy, carbon cycle, and subsurface science communities

- Using high throughput tools, technologies and comparative analysis, the JGI serves as a discovery platform to understand the organization and function of complex genomes.
- Genome and metagenome expression and sequencing of microbes, plants, and other complex systems, such as microbial communities or the rhizosphere.
- Genome annotation, functional analysis and verification of genome-scale biological system models. Systems-level integration and validation of genomic data from multiple sequencing and functional analyses.



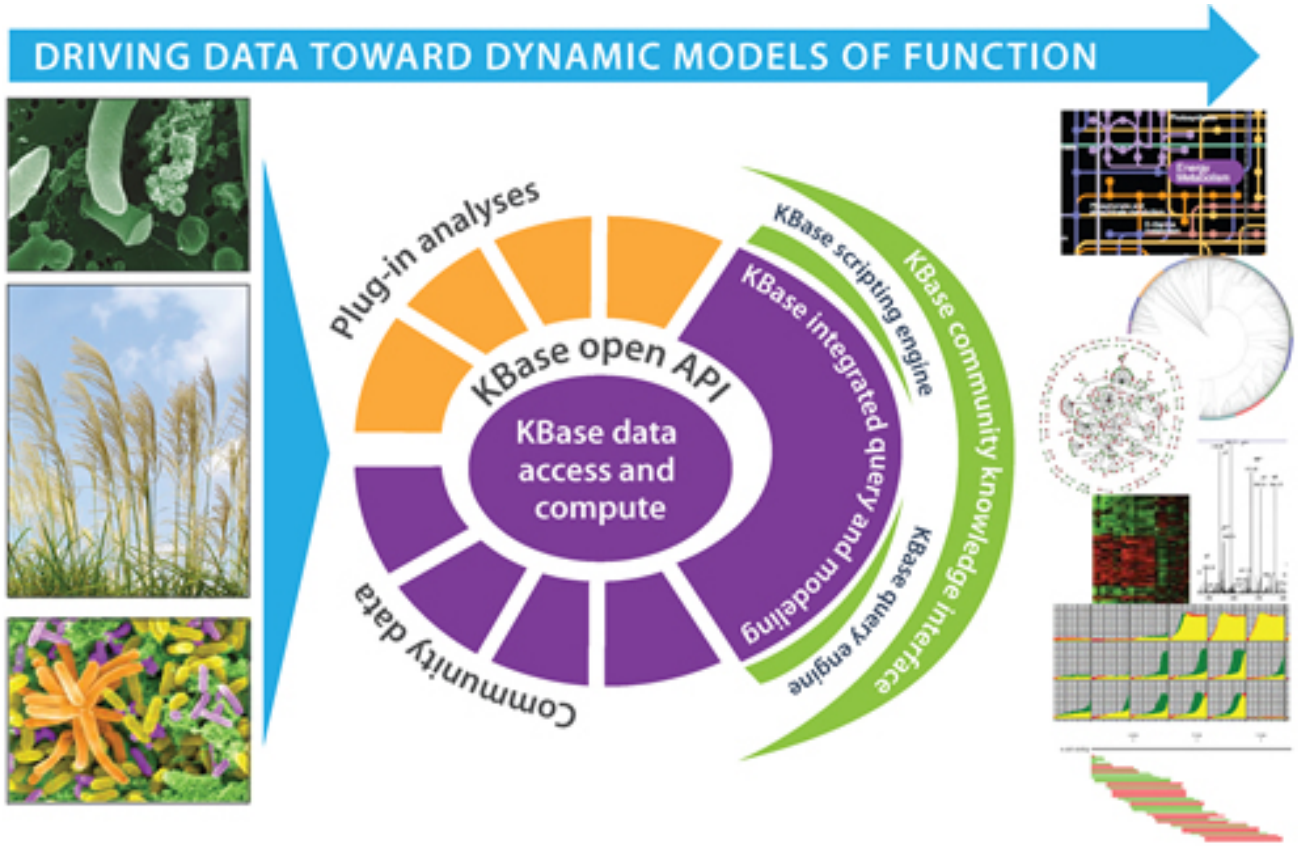
JGI DOE JOINT GENOME INSTITUTE US DEPARTMENT OF ENERGY OFFICE OF SCIENCE

DOE Joint Genome Institute

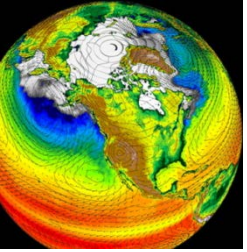
Enabling Advances in Bioenergy & Environmental Research

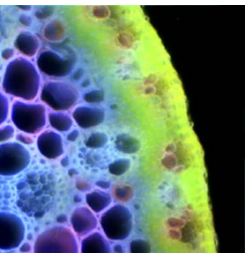
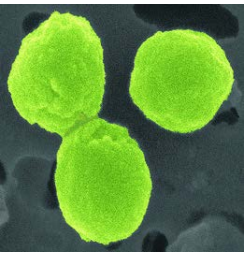
Kbase, working to build a(n)…

- ***Knowledge***base enabling ***predictive*** systems biology.
- Powerful modeling framework.
- Community-driven, extensible and scalable open-source software and application system.
- Infrastructure for integration and reconciliation of algorithms and data sources.



- Framework for standardization, search, and association of data.
- Resource to enable experimental design and interpretation of results.



Climate and Environmental Sciences Division

The Climate-Energy Nexus

Greenhouse gases (GHG) are emitted during energy production…
and climate change impacts energy production and the environment.

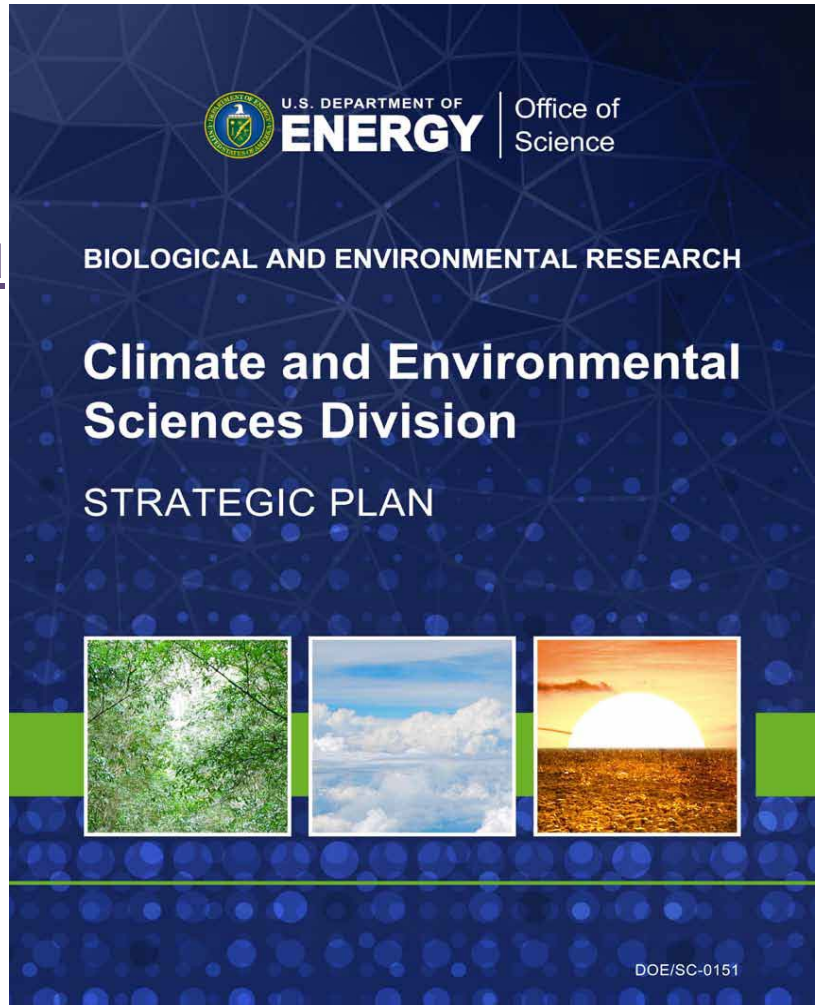
BER programs seek to:

- Understand the effects of GHG emissions on Earth's climate and the biosphere through:
  - World-leading capabilities in climate modeling
  - Unique capabilities in cloud and aerosol observations and process research
  - Ecosystem level research on climate change impacts and the carbon cycle.
- Build foundational science to support effective energy and environmental decision making

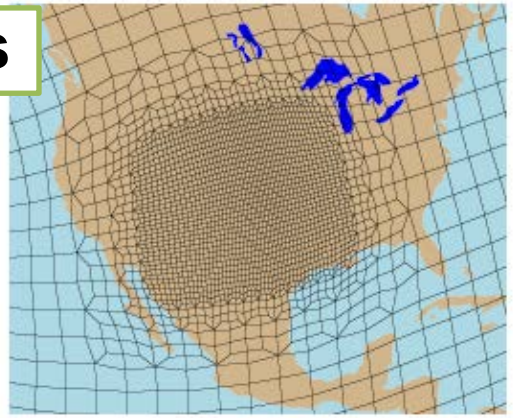
2012 CESD Strategic:

http://science.energy.gov/~/media/ber/pdf/CESD-StratPlan-2012.pdf

1. Synthesize new process knowledge and innovative computational methods advancing next generation, integrated models of the human-earth system.
2. Develop, test and simulate process-level understanding of atmospheric systems and of terrestrial ecosystems extending from bedrock to the top of the vegetative canopy.
3. Advance fundamental understanding of coupled biogeochemical processes in complex subsurface environments to enable systems-level prediction and control.
4. Enhance the unique capabilities and impacts of the ARM and EMSL scientific user facilities and other BER community resources to advance the frontiers of climate and environmental science.
5. Identify and address science gaps that limit translation of CESD fundamental science into solutions for DOE's most pressing energy and environmental challenges.



Earth System Modeling (ESM) Program Goals



Develop the Community Earth System Model (CESM) and community model evaluation tools:

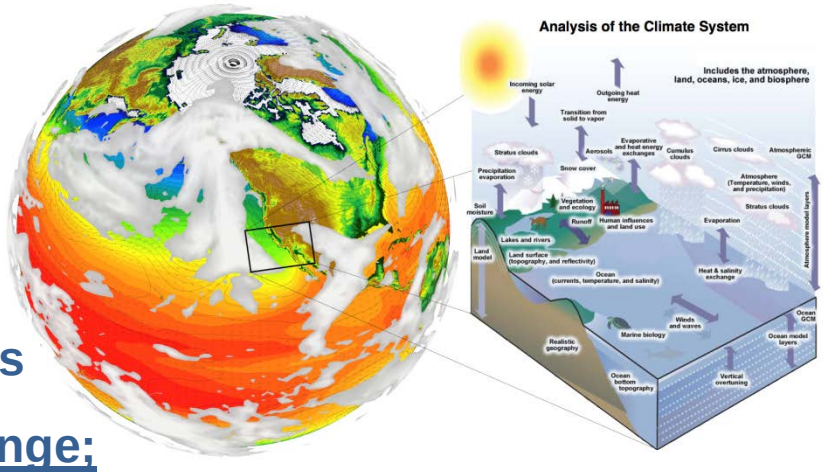
i. Community Atmosphere Model (CAM)
ii. Community Land Model (CLM)
iii. Model the full terrestrial-ocean-atmospheric carbon cycle
iv. Couple the human and natural systems within ESMs, with Integrated Assessment Research program
v. CESM ocean, sea-ice and land-ice systems, for sea-level rise projection

Construct "Generation+2" CESM, across DOE Laboratories:

i. SciDAC projects: Model (atmosphere, ocean, ice-sheets) including dynamical mesh
ii. Climate Science for a Sustainable Energy Future (CSSEF): Sophisticated "testbed" for model calibration with measurements
iii. Ultra-high resolution modeling

Regional and Global Climate Modeling (RGCM) Program

To develop a predictive understanding of the earth's climate, it's variability and change by focusing on the analysis of :



i. **Regions critical to the Earth's climate;**

ii. **Methods to obtain reliable information at higher spatial resolution;**

iii. **Coupled climate and earth system models to understand climate variability and change;**

- *Detection and Attribution* of Climate Change–
- *Extremes* event representation and attribution –
- *Feedbacks* and interactions between processes within the climate system
- Quantification of the *uncertainties*

Activities that support the analysis funded by the program include :

i. Development of metrics to inform model development

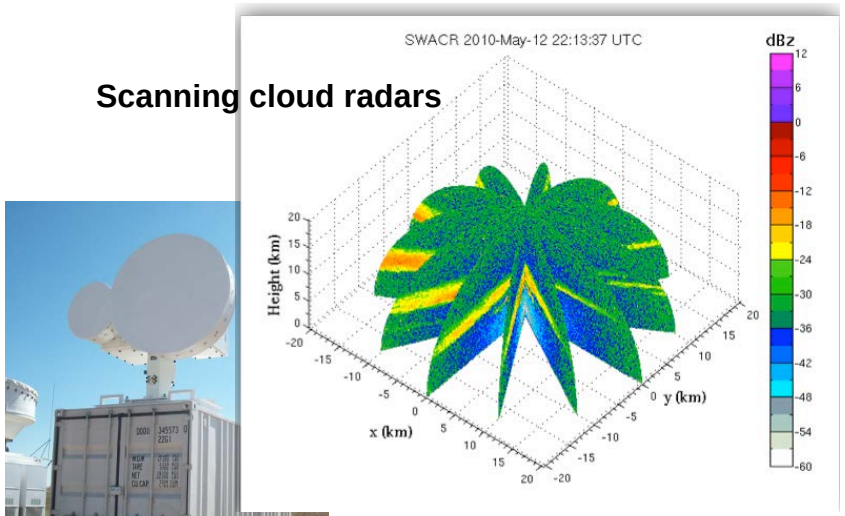
ii. Dissemination of data through the Earth System Grid Federation (ESGF)

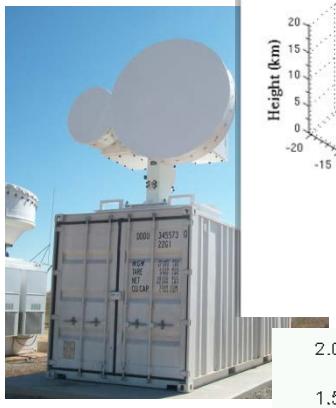
iii. Development of visualization and analysis tools to enable effective model analysis

ARM Climate Research Facility-Next Generation

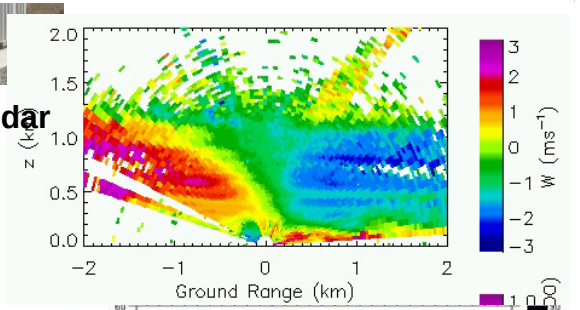
- ARM provides unique, continuous, long-term measurements for innovative research to address two largest uncertainties in climate models: the role of clouds and aerosols in climate change
- New instruments provide high-resolution, three-dimensional measurements of evolving cloud, aerosol and precipitation characteristics for improving model representations of aerosol-cloud – precipitation interactions.
- In FY 2013, ARM opens new sites in regions of high scientific interest, the Azores (marine clouds) and Alaska (Arctic clouds and aerosols over land, sea, and ice).

Scanning cloud radars

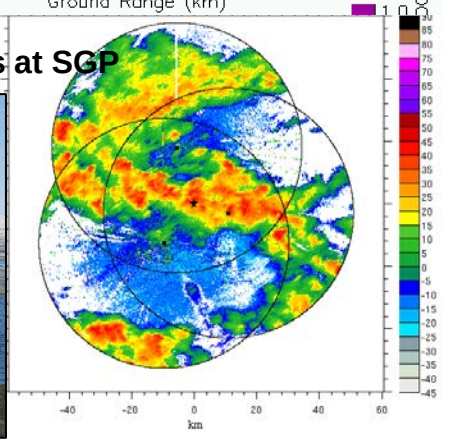




Doppler Lidar



Precipitation Radars at SGP



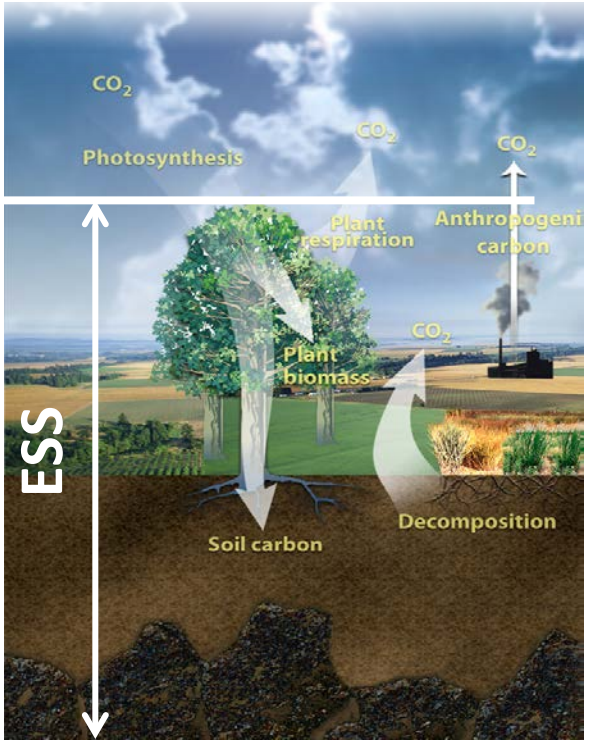
Environmental System Science (ESS)

GOAL: Develop a robust predictive understanding of terrestrial ecosystems extending from bedrock to the top of the vegetative canopy

CURRENT MODELS:

i. Community Land Model (CLM)

ii. Reactive Transport (RT) Codes:
- PFLOTRAN (SciDAC funded)
- ASCEM (EM funded)
- STOMP, TOUGH, etc.



FUTURE DIRECTIONS:
- **Link CLM with RT Codes in multiscale and extensible frameworks (e.g., NGEE)**
- **Develop genome enabled RT models of components**
- **Develop partnerships to leverage investments at SciDAC Institutes**

watersheds and their key

Focus:
carbon and nutrient cycling, biogeochemistry, contaminant transport

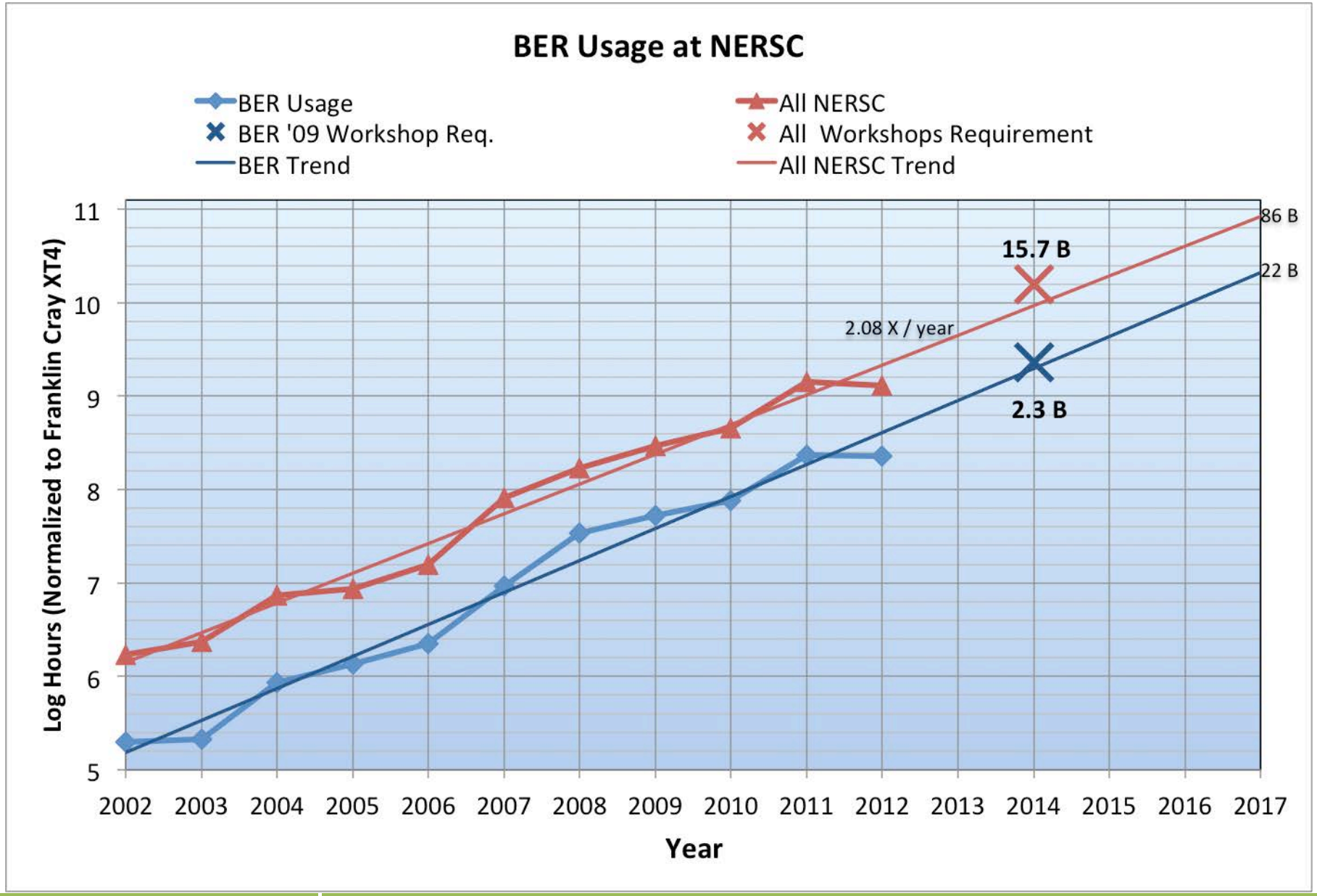
Next Generation Ecosystem Experiment

- Goal: **Develop Earth System Model simulation of Arctic Ecosystem evolution under climate change** by developing a process-rich ecosystem model, from bedrock to the top of the vegetative canopy, at the scale of an Earth System Model (ESM) grid cell (e.g. 30x30 km grid size)

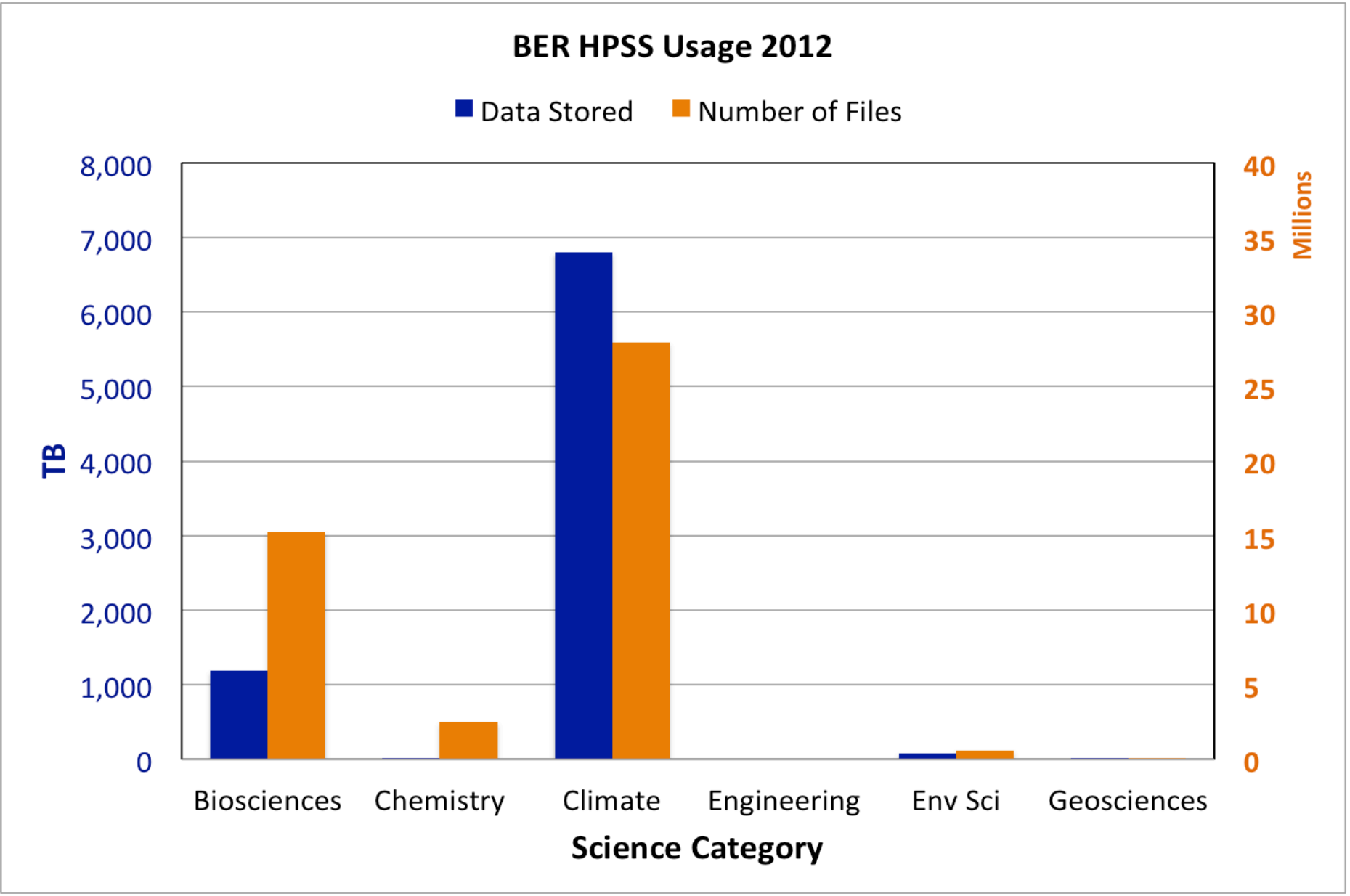
- Approach
  - Interdisciplinary, multi-scale approach to advance predictive understanding **through coupled modeling and process research**
  - Collaborative effort among DOE National Laboratories and universities, with opportunity for leveraging through external collaboration with other agencies

BER NERSC Usage and NERSC Enabled Science

Advances in DOE Science, Enable by NERSC



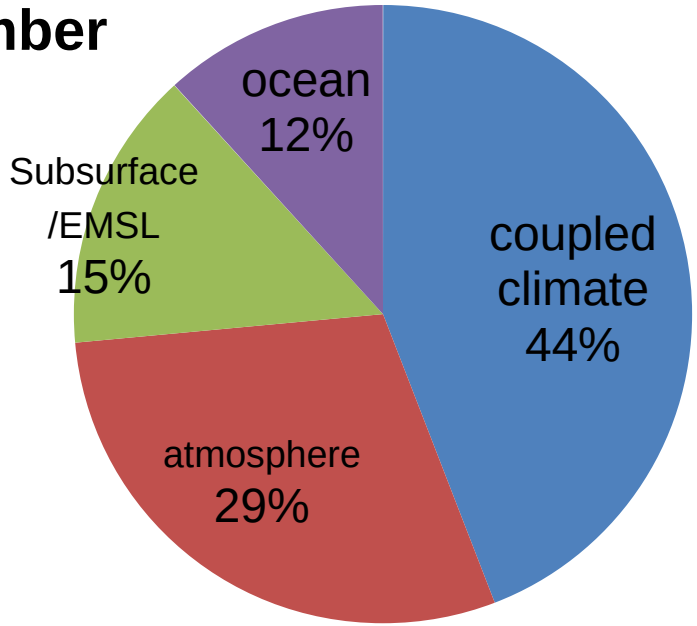
Advances in DOE Science, Enable by NERSC



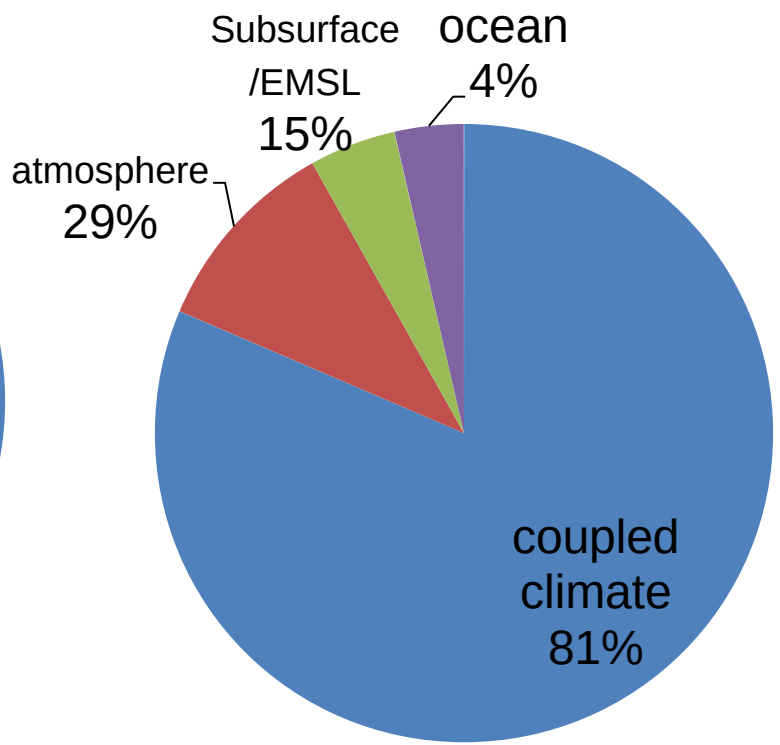
CESD NERSC allocations

- 63 Awards in 2012, 130M hours
- 34 of these (>300K hours each) use 111 hours

Number



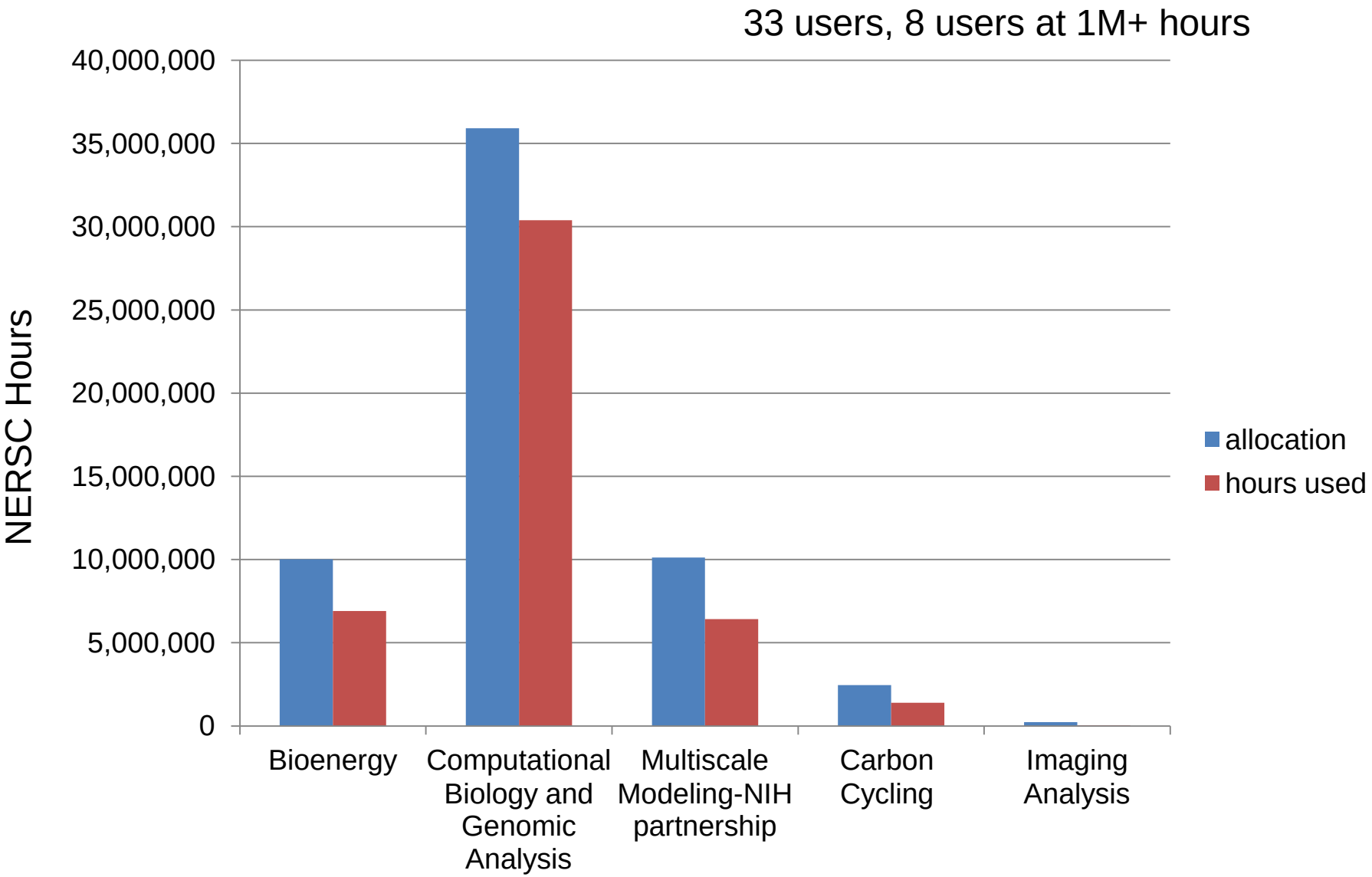
Hours



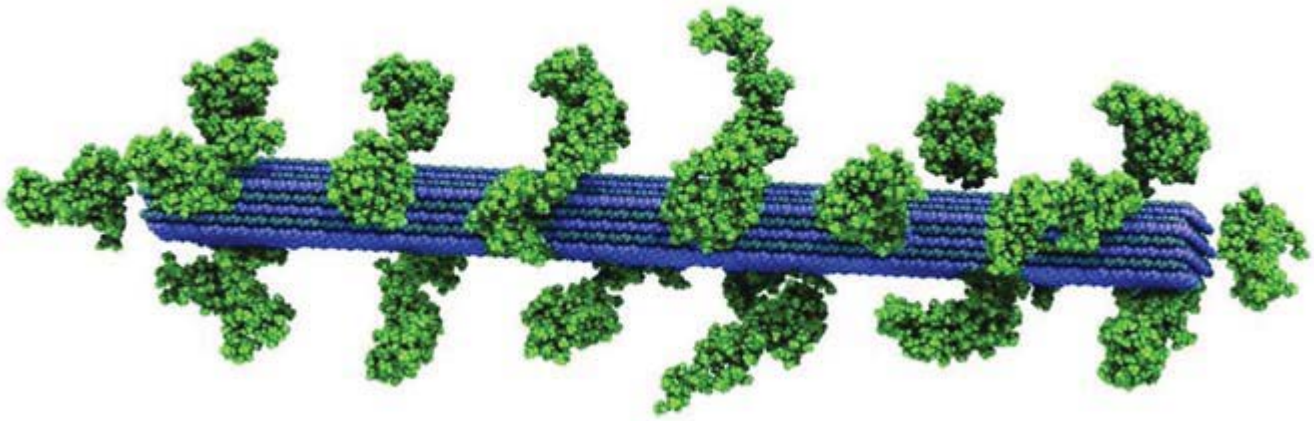
CESD Top 10 users (87M of 125M): 4Univ, 6Lab

1. Washington, 24.5M **Climate Change Simulations with CESM: Moderate and High Resolution Studies**
2. Compo,12.5M **Sparse Input Reanalysis for Climate Applications (SIRCA) 1850-2012**
3. Rasch,11M **Improving the Characterization of Clouds Aerosols and the Cryosphere in Climate Models**
4. Collins, 10M **Investigation of the Magnitudes and Probabilities of Abrupt Climate Transitions (IMPACTS)**
5. Leung, 6.4M **Development of Frameworks for Robust Regional Climate Modeling**
6. Boyle, 5.5M **Program for Climate Model Diagnosis and Intercomparison**
7. Stan, 5M **Simulations of Anthropogenic Climate Change Using a Multi-scale Modeling Framework**
8. Collins, 4.4M **Center at LBNL for Integrative Modeling of the Earth System (CLIMES)**
9. McClean,4.1M **Towards a Fine-Resolution Global Coupled Climate System for Prediction on Decadal/Centennial Scales**
10. McFarlane, 4M **CAM5 Parameter Sensitivity and Uncertainty Quantification for CSSEF**

Biological Systems Science NERSC Usage, 2012



BSSD Mission Focused Research enabled by NERSC



- JGI high throughput sequence assemble and computational annotation (23M hours used)
- Knowledgebase data storage and acquisition (new project)
- Computational Studies of Cellulose Degrading Enzymes (6.5M hours used)
- Force Field Calculations for Cellulose Modeling (360K hours used)
- Unveiling Microbial Carbon Cycling Processes in Key U.S. Soils using "omics“ (1.4M hours used)
- Computational Prediction of Transcription Factor Binding Sites (531K hours used)
- Molecular Dynamics Simulations of Protein Dynamics (7.5M hours used)

Thank You!

BER NERSC Program Managers

Dr. Dorothy Koch

dorothy.koch@science.doe.gov

Dr. Renu Joseph

renu.joseph@science.doe.gov

Dr. Susan Gregurick, Computational Biosciences and Knowledgebase

susan.gregurick@science.doe.gov