GEOSHARE:
Geospatial, Open-Source Hosting of Agriculture, Resources and Environmental Data

Presented by Thomas Hertel
Department of Agricultural Economics

Presentation to the Under Secretary for Research, Education and Economics, USDA
July 6, 2012: Washington, DC

What is GEOSHARE?

Mission: GEOSHARE develops maintains a freely available, global, spatially explicit database on agriculture, land use, and the environment accompanied by analysis tools and training programs for new scientists, decision makers, and development practitioners.

Vision: We envision a vibrant global network contributing to this shared infrastructure, enhancing capacity for analysis in developing countries, and applying these geospatial tools to guide decision making related to food security, land use, environmental sustainability and poverty reduction.

GEOSHARE Objectives

- *Provide a* globally consistent, temporally opportune, and locally relevant *database* for better decision making.
- *Assist decision makers*, policy analysts and researchers seeking to use geospatial data and analysis tools to inform activities relating to agriculture, poverty, land use and the environment.
- *Build capacity* throughout the world in individuals who can effectively bridge disciplines to make decisions and to identify solutions to complex resource use and development problems using geo-spatial data and analysis tools.

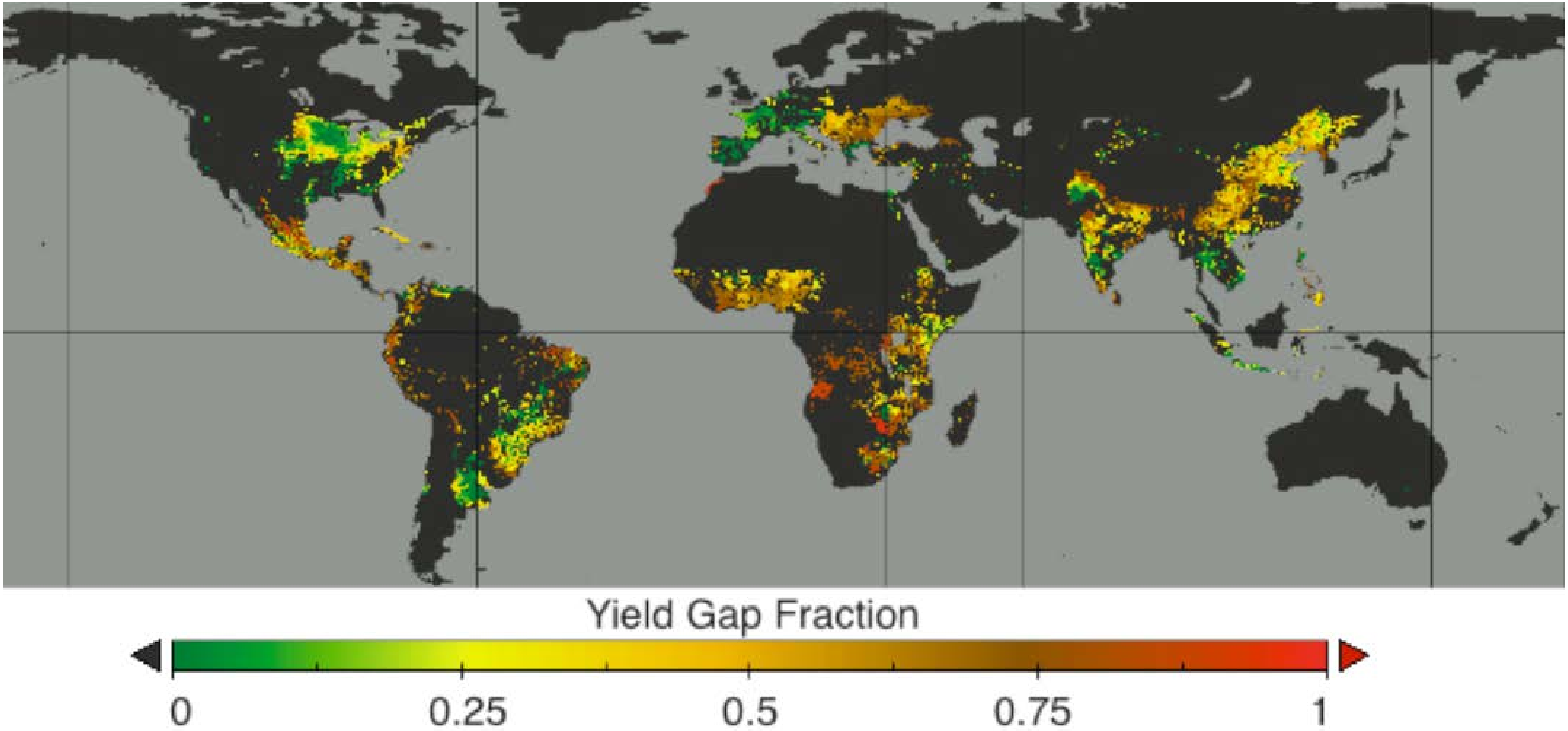
Motivation for GEOSHARE

- Feeding 9 billion people in 2050 in the face of a changing climate, while preserving the environment and eliminating extreme poverty, is one of the most important challenges facing us today.

- Agriculture is at the heart of this challenge:
  - Farming/land use change account for quarter of global GHG emissions
  - Land-based activities are very sensitive to climate change
  - Agriculture remains the predominant source of income for the world's poorest households and is therefore central to poverty reduction

- Key questions facing decision makers all require time series, spatial data:
  - Potential for boosting yields to meet projected growth in global demand
  - Optimal use of REDD+ funds to limit deforestation and sequester carbon
  - Impact of water scarcity on agricultural output
  - Impact of climate change on global agricultural productivity
  - Impacts of increasing climate extremes on vulnerability of the worlds' poor

Spatial detail is key in identifying yield gaps for crops (e.g., maize circa 2000)

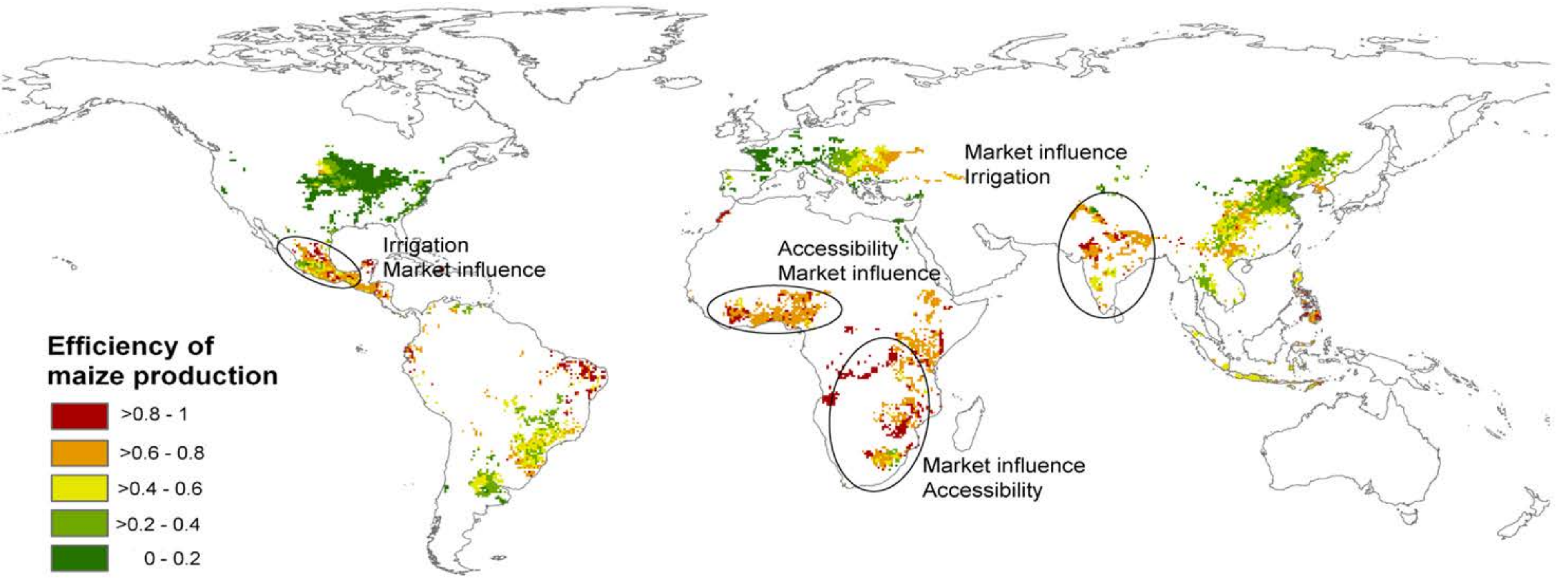


Gap = (*1 – Actual yield/Climatic potential yield*)

So 0 = on the production frontier, 1 = no productivity

Source: Licker et al. (2010)

As well as explaining their causes: Factors affecting maize production inefficiency

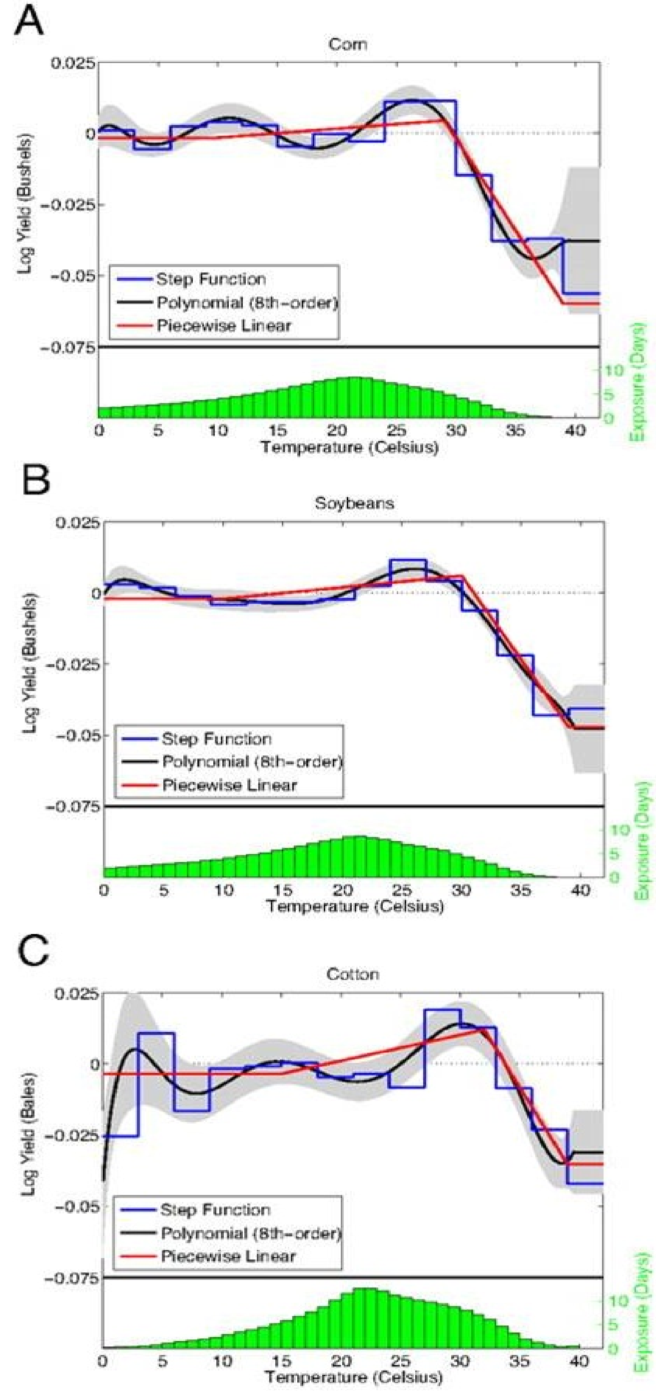


Darkened areas are more efficient – serve to "set the frontier".
Circled areas are inefficient; primary source of production inefficiency is identified

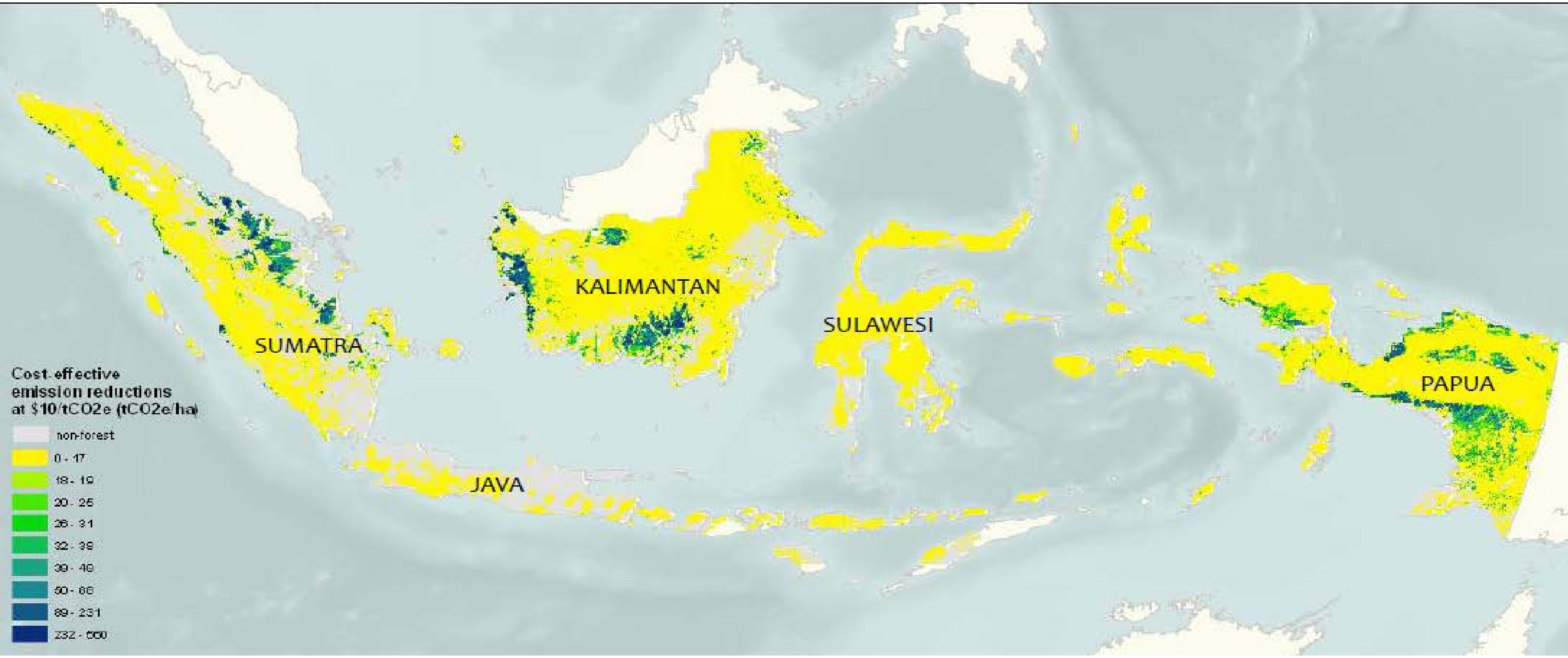
Source: Neumann et al. (2010), frontier production function estimated based on the Monfreda et al. (2008) data set

Combining time series and cross-section data is key to identifying threshold effects of climate on yields

- Schlenker and Roberts (2009):
  - pair US counties' crop yields with fine-scale weather dataset
  - incorporates the distribution of temperatures within each day and across all days in growing season
- Yields increase with temperature:
  - up to 29° C for corn
  - up to 30° C for soybeans
  - up to 32° C for cotton
- Temperatures above these thresholds are very harmful to yields



Time series spatial data are being used to design REDD payments in Indonesia



Land-cover response to carbon price of $10 tCO_2e paid for emission reductions below business-as-usual levels. Darker blue represents greater voluntary abatement of emissions from deforestation in response to incentive payments.

Source: Busch et al. (2012: PNAS), *Climate and revenue benefits of economic incentives to reduce emissions from deforestation in Indonesia*

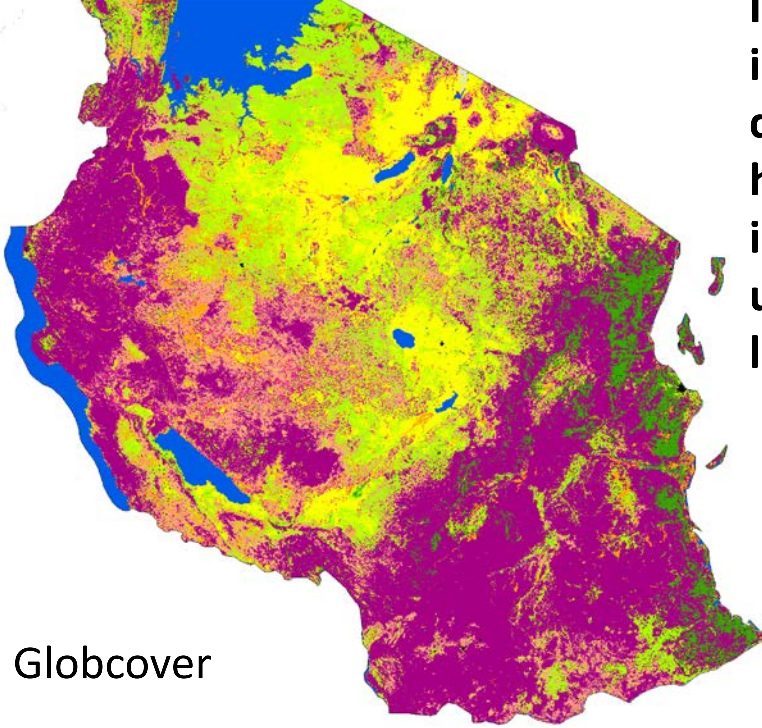
So what is the problem?

- Most spatial datasets:
  - Are regional or national, *not global;* while many of the problems are global in scope
  - Global (and many regional) data sets are *incompatible:* causes problems of inconsistency which are expensive and often impossible to resolve at the end of the data pipeline; compatibility must be designed in at outset
  - Present high barriers to entry

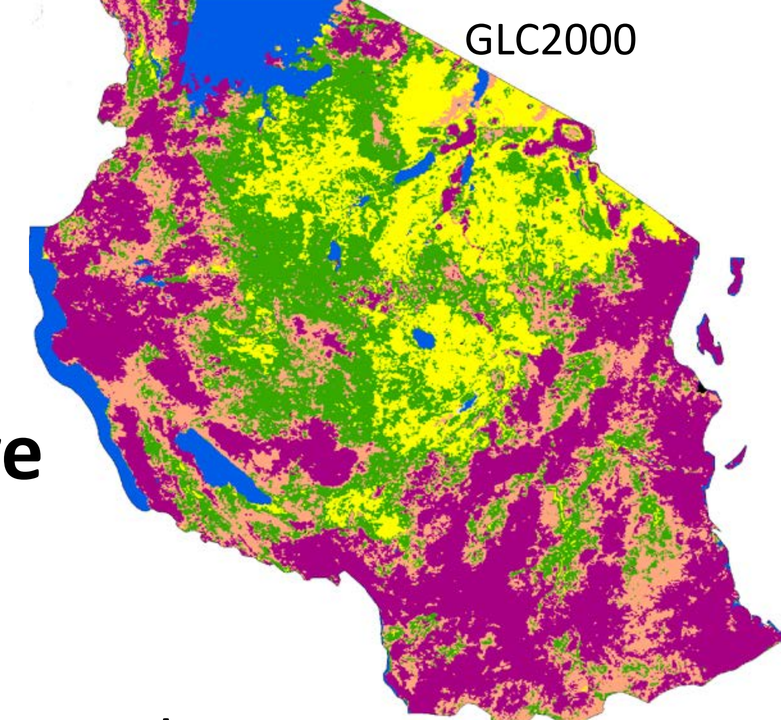
In spite of improving satellite data, there remain huge discrepancies in our understanding of land cover

Alternative views of Tanzania

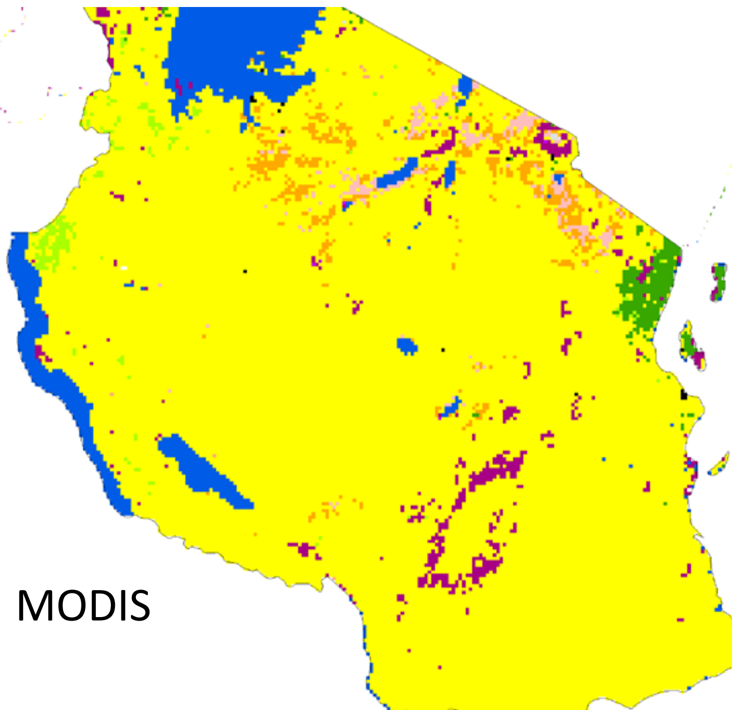
Provided by Stanley Wood



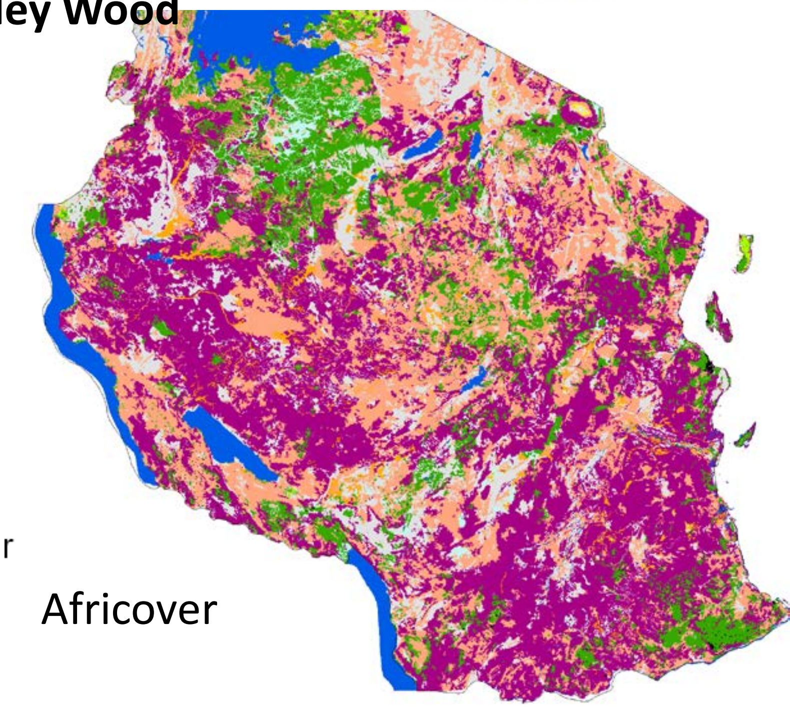
Globcover



GLC2000



MODIS



Africover

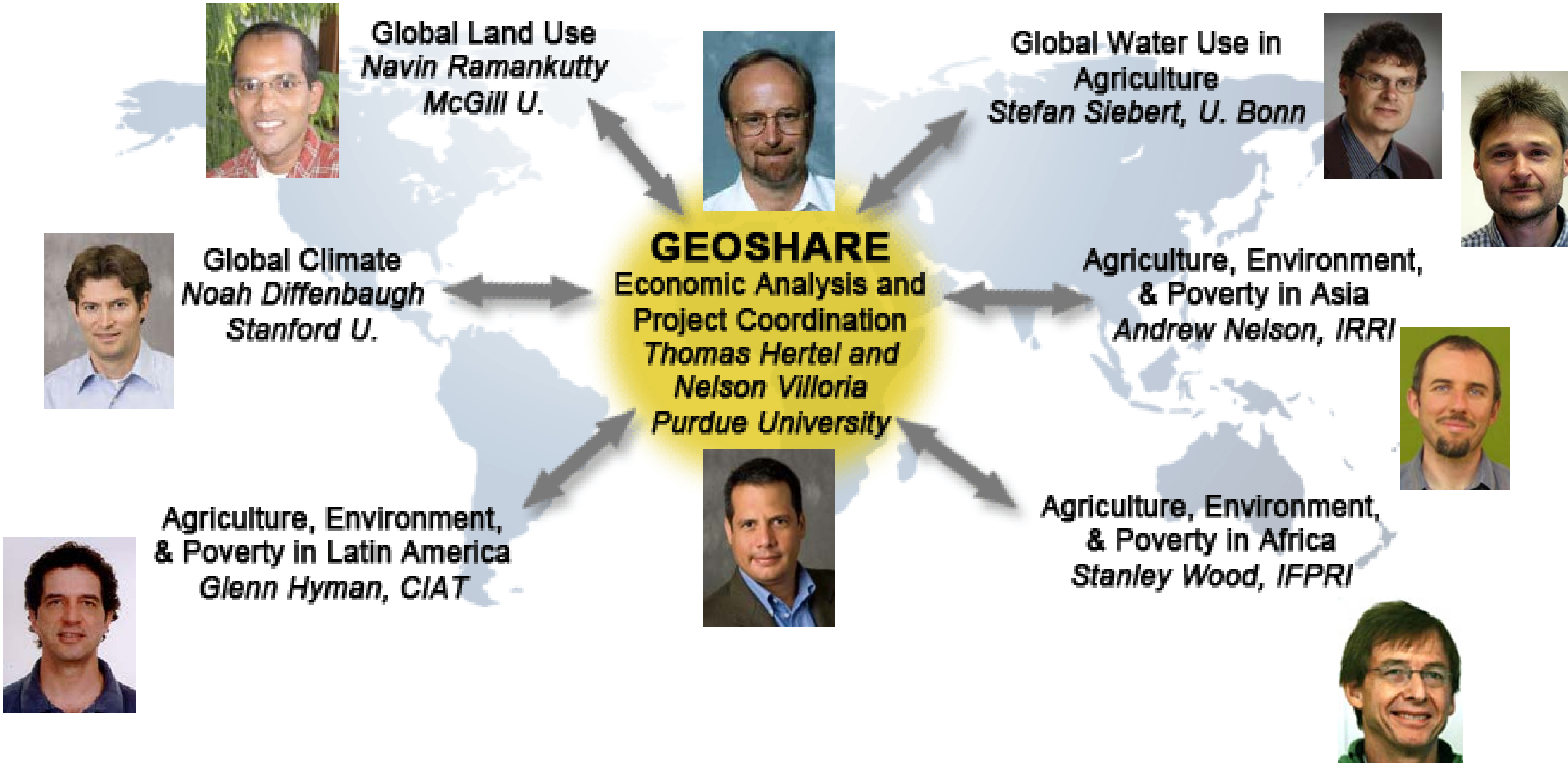
Our Diagnosis*

- The data currently available to understand how global and local phenomena affect the agriculture-environment-poverty nexus are insufficient to advance needed discovery/enable effective decision making

- This lack of information has greatly inhibited the ability of scientists, practitioners and policy makers to address the socio-economic and environmental impacts of contemporary policy issues related to poverty reduction and the long run sustainability of the world food system.

* Original proposal sponsored by the UK Foresight Programme. Available at http://www.agecon.purdue.edu/foresight/

GEOSHARE features a scalable structure with regional and global nodes



Additional nodes under consideration: Livestock, land tenure, population, regional node for Europe.

Data needed to understand tradeoffs between agriculture and the environment

Resources:
- Land (soil)
- Water
- Solar radiation
- Germplasm
- Energy
- Assets
- Population

Economic and institutional factors:
- Land tenure
- Market Access
- Credit Availability

Environmental and socio-economic outcomes:
- Poverty Alleviation
- Landscape protection
- Soil degradation
- Carbon fluxes
- Biodiversity

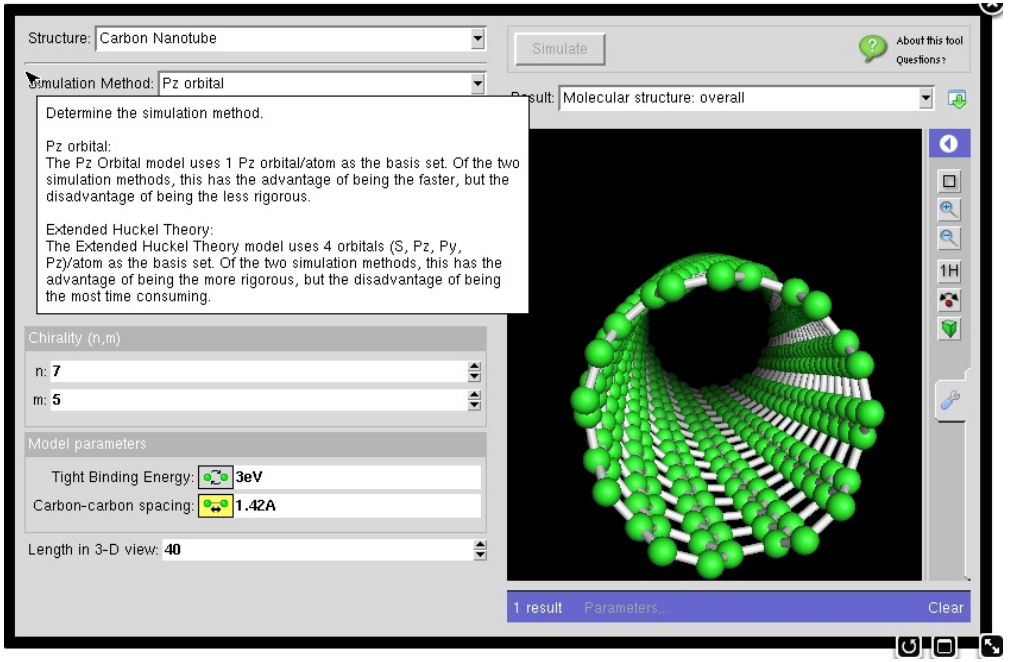
Agriculture and Forestry sectors

Management:
- Planting/harvest dates
- Irrigation
- Fertilization
- Plant protection
- Labor
- Mechanization

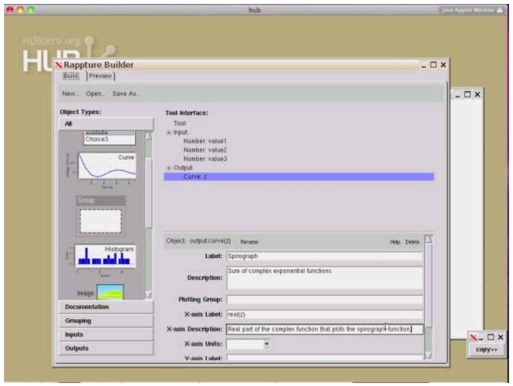
Marketed products:
- Food
- Fiber
- Fuel
- Timber
- Carbon credits

HubZero Cyberinfrastructure

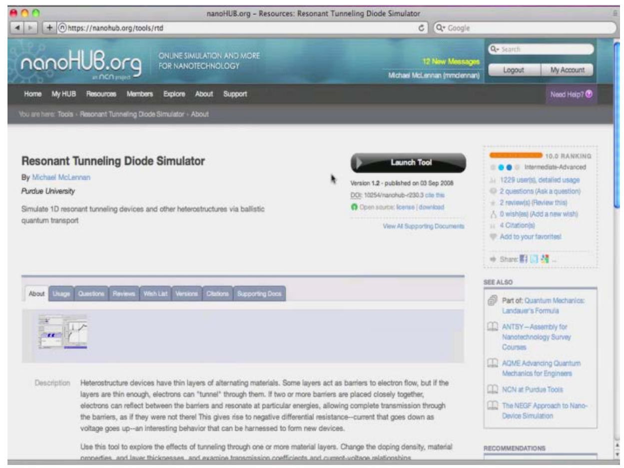
- Created by the NSF-funded Network for Computational Nanotechnology (In 2009 + 100,000 users launched 369,000 simulation runs from 160 different models)
- *Users* create, publish and access interactive visualization, simulation, and other analytical tools powered by cluster computing resources
- Focus on education and capacity building



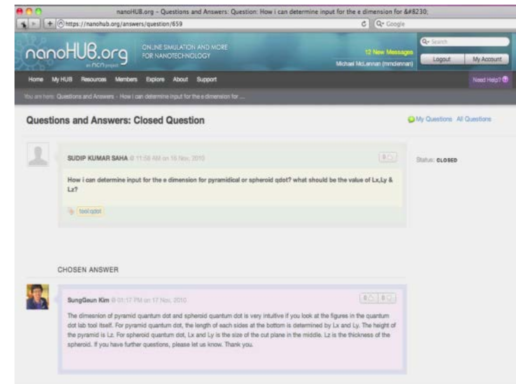
HUBzero facilitates analysis and dissemination of information

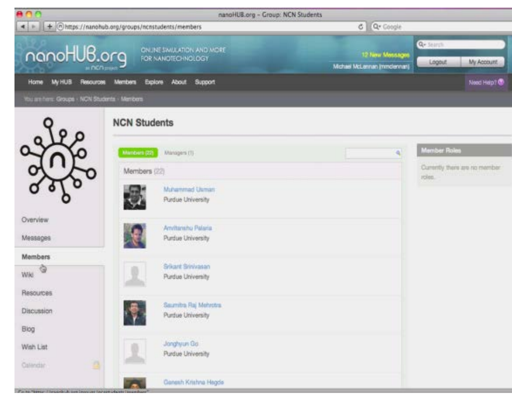


Build tools, create seminars and other resources



Publish online in restricted groups or for everyone

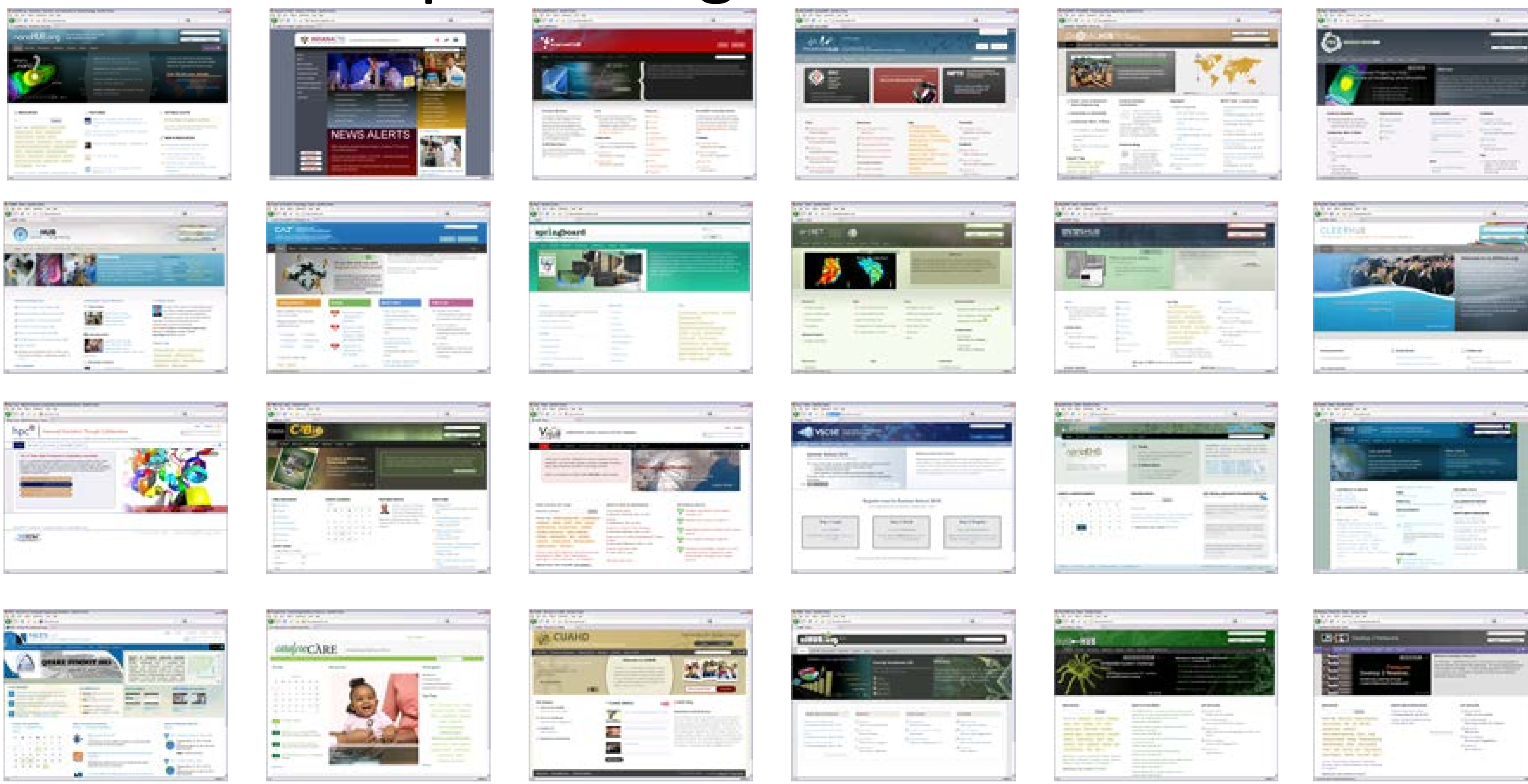




Share data, answer questions, work together in groups

Courtesy of M. McLennan, HUBzero

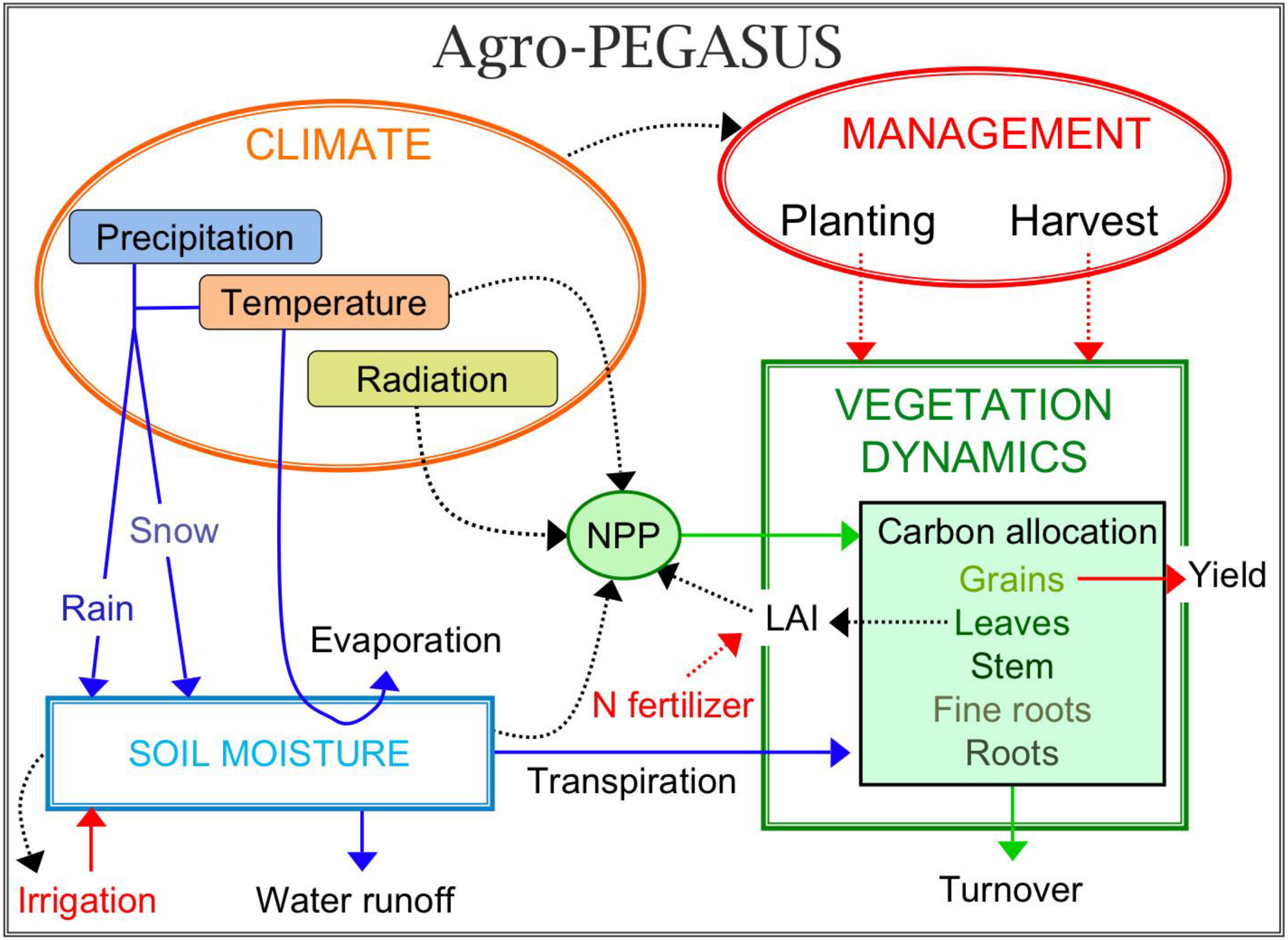
Usability –
HUBzero powering more than 30 hubs!



More underway…

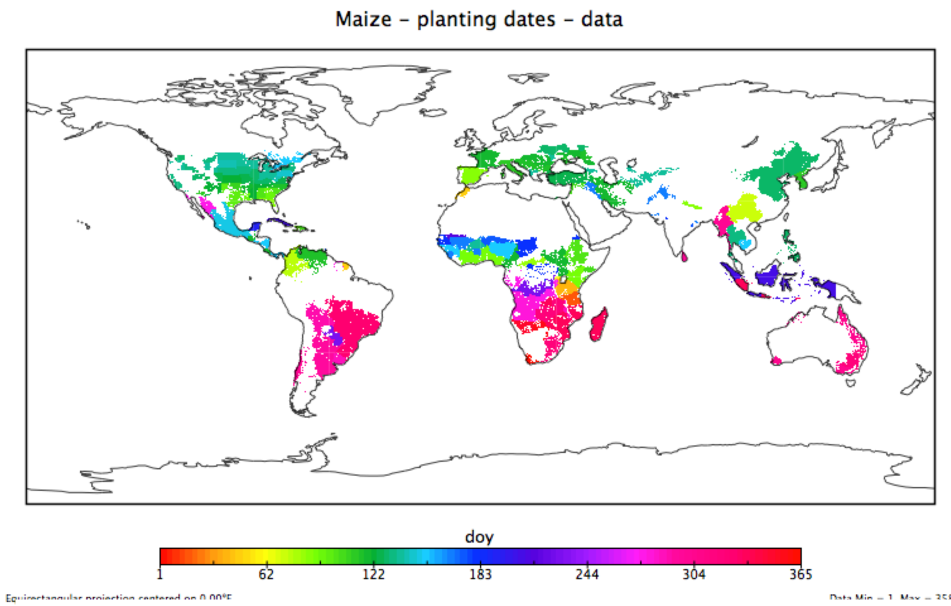
Courtesy of M. McLennan, HUBzero

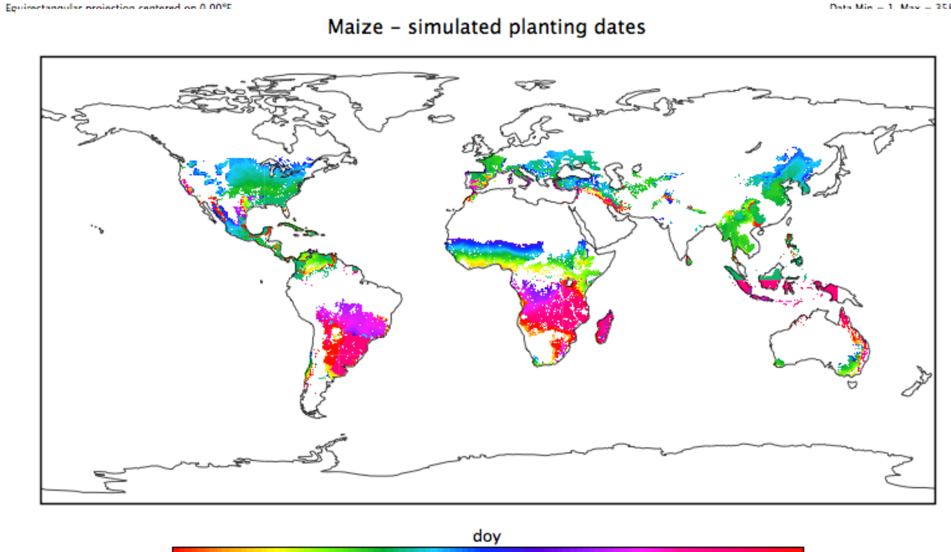
GEOSHARE Hub will offer analysis tools such as the Pegasus crop model



Source: Deryng, Sacks, Barford and Ramankutty, 2011, Global Biogeochemical Cycles

Pegasus validated at global scale

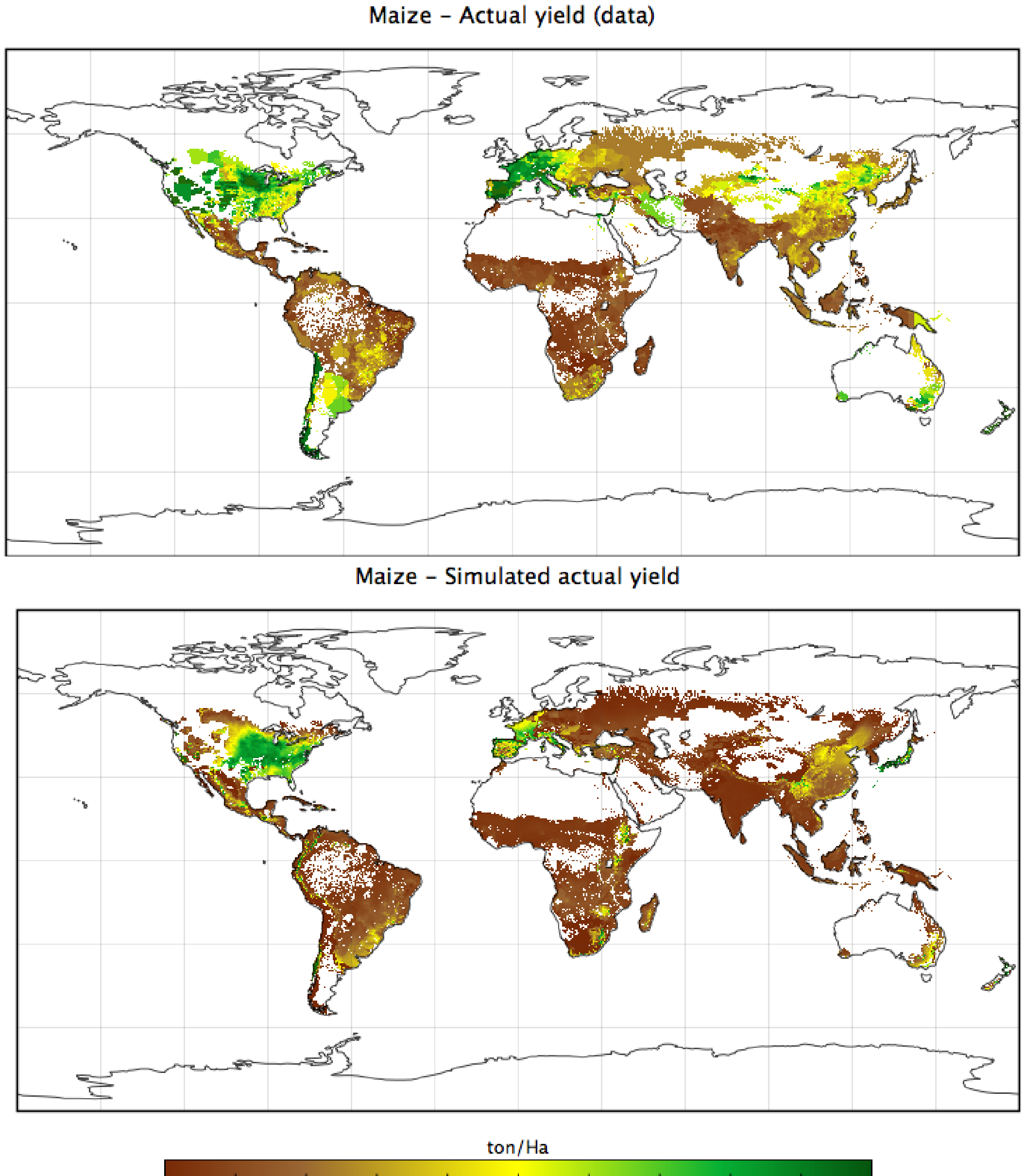




Planting (and harvest) dates agree for three-quarters global harvested area

Source: Deryng, Sacks, Barford and Ramankutty, 2011, Global Biogeochemical Cycles

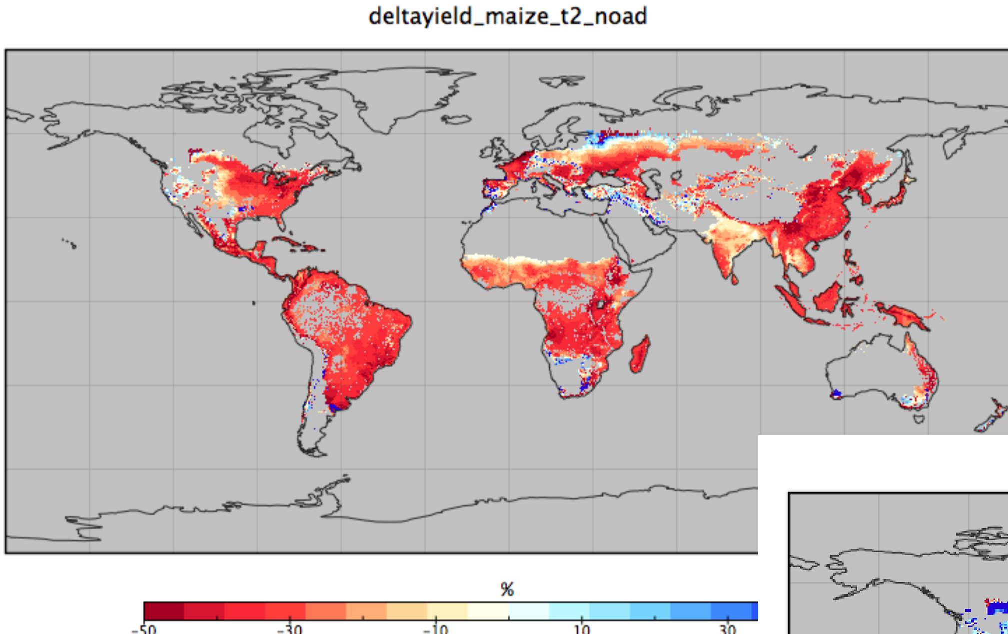
Pegasus validated at global scale



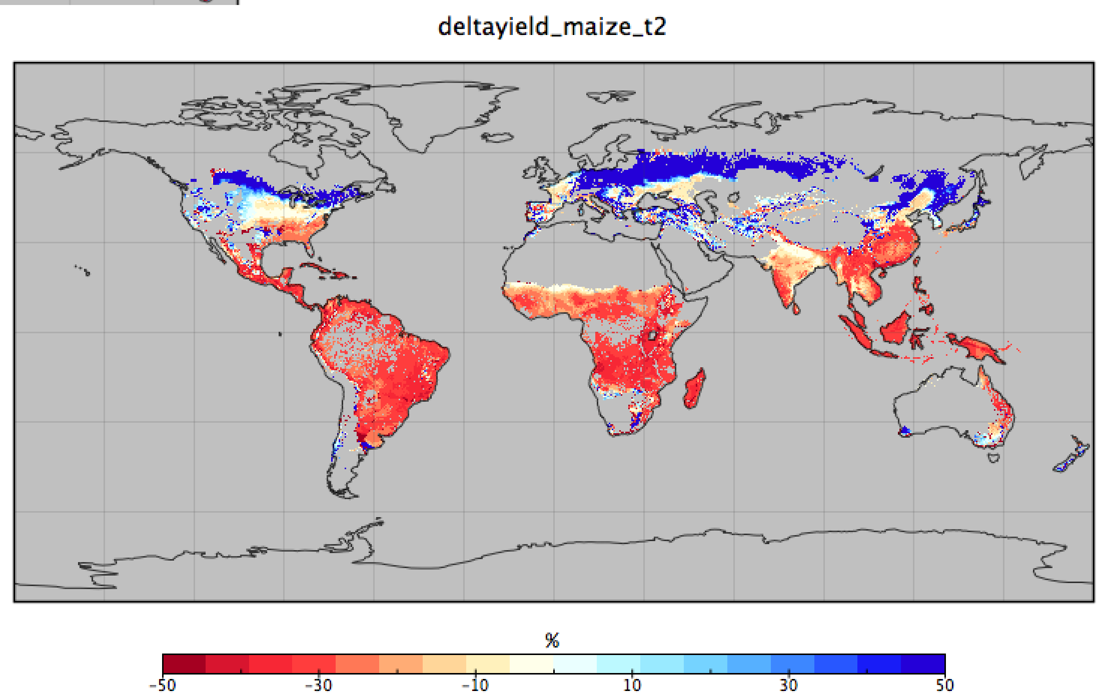
- Broad pattern of actual yields also replicated with this relatively simple (based on first principles) model
- Not obvious that this would be the case, since, unlike most crop models, Pegasus is not calibrated to a specific grid cell.

Source: Deryng, Sacks, Barford and Ramankutty, 2011, Global Biogeochemical Cycles

Pegasus can assess gridded climate impacts of 2C Temp rise



When do not permit adaptation, yields Fall globally.



Under biophysical adaptation, yields rise in temperate region; However, adaptation in the tropics is quite constrained and yields remain sharply reduced.

Source: Deryng, Sacks, Barford and Ramankutty, 2011, Global Biogeochemical Cycles

GEOSHARE is different from existing initiatives in that….

- It emphasizes development of time series, historical data for scientific analysis:
  - Data production, consistency, validation and interoperability are front and center
- HubZero technology goes beyond spatial zooming and summary statistics, facilitating:
  - Use of analysis tools 'in the clouds'
  - Networking and sharing of data
  - Technology transfer
  - Capacity building

GEOSHARE is complementary to existing initiatives, including:

- HarvestChoice, which has a strong regional focus, HarvestChoice Co-PI Stan Wood directs GEOSHARE's Africa region node
- Existing global data base infrastructure such as GlobalSoilMap.net and CMIP, which will be mapped to GEOSHARE
- FAOSTAT which will offer national comparisons
- GEO/GEOSS which set guidelines for geospatial data and GEO-GLAM for land use change

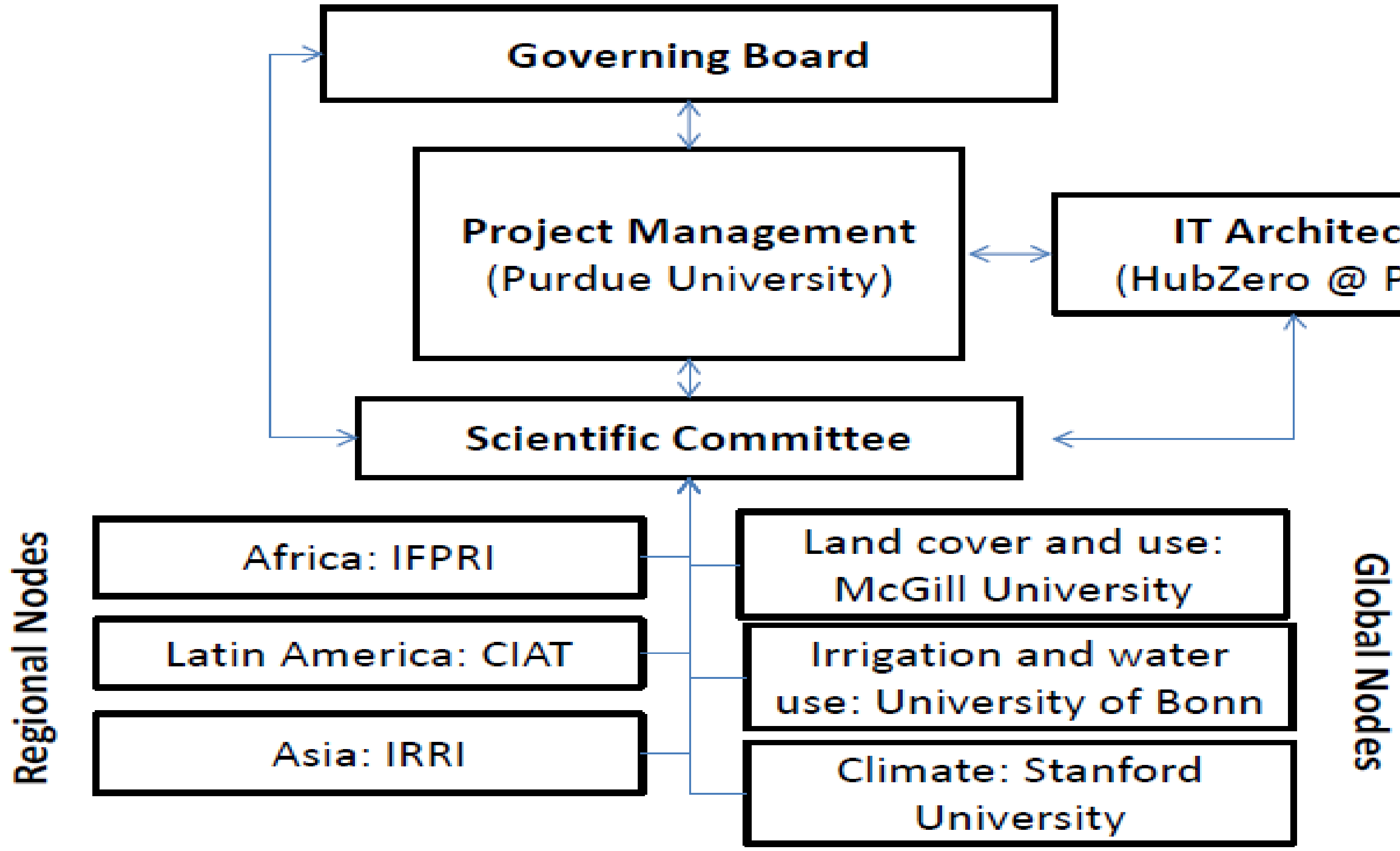
GEOSHARE is complementary to new initiatives on agricultural monitoring

- Includes new Gates-funded initiative focusing on Africa (Sachs et al., 2010)
- Three important types of complementarity:
  - Monitoring can ‘ground-truth’ GEOSHARE estimates
  - GEOSHARE can facilitate ‘scaling up’ site-specific findings, as well as choice of new monitoring sites
  - HubZero will facilitate biophysical and economic modeling to extrapolate from monitored to unobserved regions using comparable data inputs

GEOSHARE will be a source of input data for other projects

- *AgMIP*: global yield data are key for crop modeling; need rainfall/irrigation split
- Lobell et al (2011, Science) will use GEOSHARE data to update their estimates of climate impacts, disaggregating irrigated crops
- *GTAP* will use *GEOSHARE* data to disaggregate activity within countries to analyze agric impacts of trade and environmental policies

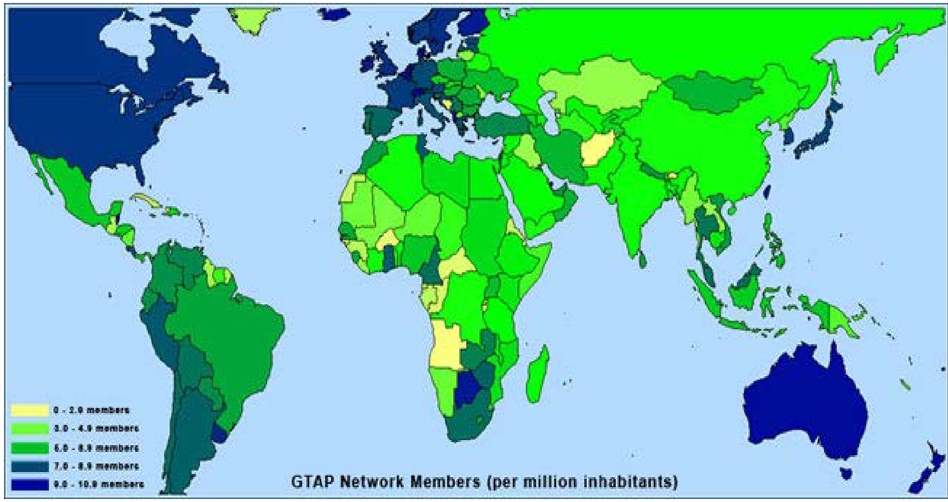
Proposed Governance Structure



GTAP as inspiration for GEOSHARE

- GTAP founded in 1992 at Purdue University by Dr. Hertel:
  - Aims to lower entry barriers and improve quality of global economic analysis
  - Offers annual short courses
  - Annual conference in 15th year; held June 27-29 at the WTO
  - Maintains analytical data base and supports dozens of different models
  - # countries has grown from 13 to 130
- GTAP network has grown rapidly
  - Now more than 9,000 in 150+ countries
  - 1,000 members in Africa
  - Most of national data bases sourced from network; data contributors 'on inside track'
- GTAP Consortium key to success
  - 28 sponsors represent leading national and international agencies
  - Diversified funding base provide stability
  - Advisory Board ensures that the project remains policy relevant
  - No single agency could accomplish what decentralized project now does each year

Region	Members	Network Perctage	Per Million Inhabtnts
Africa	1,020	11.7%	1.2
Asia	3,094	35.4%	0.8
Europe	2,059	23.6%	3.5
North America	1,725	19.8%	3.5
Oceania	340	3.9%	10.9
South America	494	5.7%	1.4



GEOSHARE Pilot has just begun

- Modest funding from three sources:
  - UK Department For International Development: $ 440k
  - UK Department for Environment, Food and Rural Affairs: $ 100k
  - USDA’s Economic Research Service: $100k
  - Additional funds from HubZero ($50k) and from CCAFS ($30k)
- Proof of concept:
  - Two countries in South Asia, six in Africa; regional case studies demonstrating support for decision makers in Asia and Africa
  - Integrate regional and global nodes – focus on irrigation/rainfed split, overlaid by data on poverty and land tenure
  - Delivery of data and decision tools (e.g. Pegasus crop model) through HubZero infrastructure
  - Design governance; assess costs and benefits; address ethical issues
- Donor’s forum in early 2014:
  - Demonstrate success of proof of concept from pilot project
  - Secure long term funding for GEOSHARE