A Global, Spatially-Explicit, Open-Source Data Base for Analysis of Agriculture, Forestry, and the Environment

Presented by Thomas Hertel

Based on a joint proposal with Wolfgang Britz, Noah Diffenbaugh, Navin Ramankutty and Nelson Villoria

With additional contributions from Glenn Hyman, Andrew Nelson, Stefan Siebert and Stanley Wood

Proposal sponsored by the UK Foresight Programme.
Interdisciplinary research is challenging!
So you need great collaborators!
Outline of the Talk

• Importance of land-climate interactions
• Importance of time series spatial data to science and policy analysis
• So what is the problem?
• Outline of pilot effort
• Conclusions and funding requirements
Agriculture, forestry, land use and climate: Three distinct interactions

- Productivity is affected by climate change

- Land cover change and land use drive climate change: 30% of GHG emissions from land based activities/deforestation (Baumert, Herzog and Pershing, 2009)

- Agriculture and forestry offer low cost mitigation options: at $28/tCO2e tax, 50% of global GHG abatement could come from these sectors (Golub et al, 2009)
Time series spatial data are needed:
Identifying forest-cover hotspots concentrated in the tropics (1980–2000)

Source: Lepers et al. (2005)
Time series spatial data are valuable

Using 100+ Landsat scenes for 1980-2000 period, Gibbs et al found that, *globally more than 55% of new land expanded at the expense of intact forests, and 28% replaced degraded forests.*
Spatial detail is key in identifying yield gaps for crops (e.g., maize circa 2000)

\[ \text{Gap} = (1 - \frac{\text{Actual yield}}{\text{Climatic potential yield}}) \]

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
Impact of climate extremes on suitability for agricultural production (premium wines in USA)

Response of wine production driven by response of severe heat and response of severe heat varies spatially

Source: White et al. (2006)
So what is the problem?

- **Most spatial datasets are:**
  - Often regional or national, *not global*
  - May not be practically accessible
  - Their use requires considerable specialized IT knowledge
  - *Incompatibility* in various data sets causes huge problems
Modeling of many Earth system processes requires the combined use of different input data sets

=> consistency at the grid cell is essential

Source: Gruber & Galloway (2008)
Our Diagnosis

- Funded by the UK Government through Foresight Office
- Validated by peer-review of 20+ experts, including climatologists, geographers, agricultural economists, GIS specialists in the CGIAR Centers, and policy advisors:
  
  “the lack of an integrated, publicly available, global, explicitly spatial, time series database on land cover, land use and climate has inhibited scientists’ ability to assess issues relating to the long sustainability of agriculture and forestry in the face of climate change”

- While the policy issues we seek to address are global in scope, the assessments and solutions are invariably site-specific and hence local in nature
The proposed solution...

A database that:

1. Gathers national and sub-national statistics from various statistical agencies around the world to put together a consistent global data set, along with regional companion data sets, on agriculture and land use;

2. Employs spatial disaggregation methods, including the use of satellite remote sensing technology and spatial statistics to develop geographically-explicit gridded data on a global scale; and

3. Develops a data portal, including new tools for providing data in a variety of convenient formats to the global research community
Our partners and their expertise...

**Gridded Global Area & Production Data for Crops & Livestock**
Navin Ramankutty (Geographer)
McGill Univ., Canada

**Climate Change Models, resolution & bias correction**
Noah Diffenbaugh (Climatologist)
Stanford Univ., USA

**Irrigation**
Stefan Siebert (Crop Scientist)
Univ. of Bonn, Germany

**Co ordination, IT infrastructure**
Thomas Hertel & Nelson Villoria
Purdue Univ., USA
(Economics & IT)

**Land use in Asia**
Andrew Nelson (Geographer)
International Rice Research Institute, The Philippines

**Land use in Africa**
Stanley Wood (Economist)
International Food Policy Research Institute, Washington, DC

**LatAmerica land use**
Glenn Hyman (Geographer)
International Center for Tropical Agriculture, Colombia

Pilot Project: Global (in blue) & regional (in green) nodes. Coordination & data delivery at Purdue
Satellite data is continually improving!!
As is global coverage of agricultural censuses (Year 2000)

Source: Ramankutty et al., (2008)
Sub-national data greatly enrich our understanding of land use

Crop distribution - Maize

Statistics aggregated by 57 Countries

Statistics aggregated by 5690 Administrative units
Merging satellite and subnational census data can greatly enhance our knowledge of land use.

“Statistical Data Fusion”
Cropland cover changes: 1700-2000

Source: Ramankutty, N. McGill University
This is interesting… but can it be done?

Proposal draws inspiration as well as lessons in institutional design, from three previous projects focused on data collaboration, involving subsets of our authors:

- AgroMAPS: Collaboration between FAO, IFPRI and SAGE to collect and reconcile sub-national land use data for the world.
- CMIP3: Archive of results from the leading climate models; has formed the basis for hundreds of scientific papers written by scientists outside the those modeling groups.
Thank you for your attention!

•Proposing 2 year pilot project: $600K/yr

•Fund the six nodes and coordinating body to demonstrate ‘proof of concept’ and implement common standards (see Annex 2)

•After 2 years, meeting of potential donors:
  – Governing Board
  – Scientific Advisory Committee
  – Selection of sites for longer term project

•URL: http://www.agecon.purdue.edu/foresight/
Bibliography


Licker, Rachel, Matt Johnston, Carol Barford, Jonathan A. Foley, Christopher J. Kucharik, Chad Monfreda, and Navin Ramankutty. 2010. Mind the Gap: How do climate and agricultural management explain the “yield gap” of croplands around the world?


