October 15, 2010

A Global, Spatially-Explicit, Open-Source Data Base for Analysis of Agriculture, Forestry, and the Environment: Proposal and Institutional Considerations[1][2]

Thomas Hertel, Purdue University
Wolfgang Britz, University of Bonn
Noah Diffenbaugh, Stanford University
Navin Ramankutty, McGill University
Nelson Villoria, Purdue University

With additional contributions from
Stanley Wood, IFPRI
Stefan Siebert, University of Bonn
Glenn Hyman, CIAT
Andrew Nelson, IRRI

[1] This proposal and its scientific review were made possible through the financial support of the Foresight Programme from the UK Secretary of State for Business, Innovation and Skills.

[2] The authors would like to thank a team of reviewers who have had important input into this proposal, including: John Antle, Jelle Bruinsma, Neil Conklin, Rick Farnsworth, Raymond Florax, Lowell Hardin, Kim Hjort, Glenn Hyman, Suchada Langley, David Lobell, Gerald Nelson, Wolfram Schlenker, Josef Schmidhuber, Brent Sohngen, Elke Stehfest, Sally Thompson, Patrick Westhoff , and Stanley Wood.

Abstract

After reviewing the current state of the art in spatially explicit, global data base infrastructure for the analysis of agriculture, land use and the environment, we propose new infrastructure to support researchers working in this area. The proposed effort would do three things: 1) gather 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) employ *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) develop a *data portal*, including new tools for providing data in a variety of convenient formats to the global research community. Distinguishing features of this effort will be its focus on transparency, documentation and peer-review, quality control and sustainability over time. The majority of such previous efforts have only focused on the data portal element – they rely on other researchers to provide them the data. By integrating data collection, development, and provision into one project, we will be able to support the global research community as it seeks to understand the long-run sustainability of the global agricultural system.

Table of Contents:

1 Motivation

The recent global commodity crisis underscored the vulnerability of the global food system to shocks from extreme weather events, energy markets, and obstruction of the smooth functioning of commodity markets via export bans and other measures designed to force instability onto global markets. We have learned that a "perfect storm" in which all these factors coincide can have a devastating impact on the world's poor. As we look ahead several decades, will such commodity price spikes become more commonplace in the future? What investments and policies need to be put in place to mitigate the frequency, intensity and/or impacts of such events? In order to answer these questions, existing projection and modeling approaches need to be linked in analysis and new ones developed, and to do so we need a new global data base infrastructure for the food system.

In the past few years, there has been a convergence of interest in the global farm and food system and its contributions to feeding the world's population as well as to ensuring the environmental sustainability of the planet. The number of mouths which the world must feed is expected to increase by another 2 billion to 9 billion people by the middle of this century. When coupled with nutritional improvements for those currently living on less than $2/day, this translates into a very significant rise in the demand for agricultural production. At the same time, growing use of biomass for energy generation introduces further demand in agricultural markets. This is all occurring as water, a key input into agricultural production, is diminishing in availability in many parts of the world. Indeed, a recent report by McKinsey & Co. (2009) predicts that, absent efficiency improvements, one-third of the world's population will live in river basins where the water demands exceed sustainable, reliable supplies by more than 50%. This will have important implications for irrigated agriculture which currently accounts for 42% of global crop production and 70% of global freshwater withdrawals (Bruinsma 2009).

Agriculture and forestry are also increasingly envisioned as key sectors for climate change mitigation policy. Farming and land use change – much of it induced by agriculture - currently account for about one-third of global greenhouse gas emissions, and any serious attempt to curtail these emissions will involve changes in the way agricultural land is used, as well as limits on expansion of farming into areas currently under forest. Limiting deforestation is also critical to preserving the planet's biodiversity. Finally, agriculture is likely to be one of the sectors most sharply affected by the inevitably changing climate. This will likely shift the pattern of global comparative advantage in agriculture, and may well reduce the productivity of farming in precisely those regions of the world where malnutrition is most prevalent, while increasing yield variability.

All of these factors point toward the need for a better understanding of the spatial distribution of agricultural production, resource use and associated greenhouse gas emissions. Current and future climate predictions are also critical elements of any such knowledge infrastructure. The goal of the proposed project is to make available to the world's research community an open source, spatially explicit, global data base on agriculture and the environment to advance research on the long run sustainability of the world's food system.

While there now are extensive open-source global spatial data bases on climate –largely driven by the Intergovernmental Panel on Climate Change, there are no comparable data bases on agriculture. While many rich data sets on agriculture exist, they tend to be regional or national in scope, they are not compatible with one another, and they are often not publicly available; when they are available, they are not practically accessible to researchers due to technical challenges. This situation has greatly inhibited our ability to adequately address contemporary policy issues including: climate change mitigation, environmental impacts of biofuels, offsite pollution from agriculture, and the preservation of biodiversity, among others. All of these questions require knowledge of local conditions facing agricultural producers, including soil, water, topography, climate, and current farming practices. At the same time, many of these questions are global in scope. Combining such site-specific detail with global coverage is therefore critical to advancing our understanding of these important issues.

The availability of coordinated climate model experiments to the international community in a single archive has revolutionized climate science. Similar benefits could be achieved in the agricultural research community through a new global, spatially-explicit Database that brings together a diversity of datasets into a common framework. Such a Database would have great benefit in a broad set of research communities, especially if it could provide data structures and formats unified across diverse disciplinary domains.

2 Need for the Project

A natural question to ask at this point is: Hasn't this already been done? Indeed, there are many impressive efforts currently underway which seek to address part or all of the challenges outlined in the previous section. These ongoing projects can be grouped into two categories: (1) efforts to compile already existing global databases and to develop a web-based interface to provide easy access to users, and (2) efforts to develop entirely new databases open to the global community. It is very inexpensive to set up a web page, and, as a result, the number of efforts underway in the first category is growing rapidly. We will devote relatively less attention to these projects, because our proposal focuses on the development of new databases and dissemination strategies.

One of the most important efforts to develop spatially explicit, global data for long run analysis of agriculture and the environment is the GAEZ model and associated data bases developed over the past two decades by FAO and IIASA and recently updated in the form of "GAEZ-2009" (IIASA/FAO 2010). This "offers a standardized framework for the characterization of climate, soil and terrain conditions relevant to agricultural production which can be applied globally at sub-national levels." (IIASA/FAO 2010 p.21) The resulting geo-referenced land resources data base is maintained at 5 arc-minute resolution (although some data are only available at a more aggregate level, while still other data are available at the 30 arc-second level) and permits the assessment of crop-specific growing conditions at a local level. GAEZ model equations translate these attributes into crop growth outcomes for both rainfed and irrigated lands. This framework has been used to look at a wide range of issues, including future water requirements in agriculture, the impact of climate change on agriculture, and the impact of biofuels expansion on global land use. In short, it is highly relevant for the questions at hand.

The most important drawback of the GAEZ framework is that it is not publicly available. As such, it is of little use to researchers outside of the team at IIASA that maintains this framework. One of the authors of this proposal (Hertel) participated in a peer-review of the GAEZ framework, held at FAO in June, 2010.[3] This review underscored the need to make the basic inputs and outputs from the GAEZ framework publicly available. It also highlighted the fact that, despite its high degree of sophistication, the GAEZ framework does not avail itself of the latest data and developments in the fields of global land use and environmental analysis. Two examples of such data limitations were immediately apparent to peer-reviewers. The first pertains to the use of climate data and predictions from the Intergovernmental Panel on Climate Change (IPCC) third assessment report – as opposed to the more recent and much improved CMIP3 archive used in the IPCC's fourth assessment report, and upon which all current research on climate change impacts is based. This feature alone is likely to discredit any climate change analysis based on this framework, despite its many other valuable features.

A second drawback of GAEZ pertains to the data and methods used to assess current land cover and forest density. In contrast to GAEZ-2009, Ramankutty et al. (2008) and Monfreda et al. (2008) used: (a) a much richer collection of subnational agricultural statistics (for ~16000 administrative units, and (b) sophisticated spatial disaggregation techniques that made rich use of widely available remote sensing data. Similarly, Hansen et al. (2003) used MODIS satellite data to derive percentage tree cover at 500 m resolution globally. Finally, while the GAEZ framework has been described in several internal reports and on websites, it has not been published in peer-reviewed outlets. In short, despite the enormous effort expended on, and experience garnered through, the GAEZ system over more than 20 years, it has not resulted in a usable, peer-reviewed data base available to the global research community. The specific limitations highlighted in the recent FAO review underscore the need for getting input and contributions from a wide range of experts working in the field of agriculture, environment and climate change.

There are many other important efforts aimed at producing spatial data bases focusing on agriculture and the environment, including the research into global irrigation and water use at the Universities of Bonn and Frankfurt[4], the History Database of the Global Environment of The Netherlands Environmental Assessment Agency[5], ongoing research at the Universities of McGill and Minnesota[6], as well as collaborative efforts such as AgroMAPS which involved

---

[3] Based on this review, we reached the following conclusions: (a) this is a most impressive framework, (b) it is complex and highly specialized in its application – even the FAO staff members involved with GAEZ for the last wo decades cannot utilize it on their own, (c) this is a unified framework, with no clear distinction drawn between data, intermediate results, parameters and model; this makes it difficult for an outsider interested in using parts of the framework to distinguish between source data which they might wish to drawn on for other purposes, and intermediate data (e.g., sub-national yields) which are the result of predictions made by the GAEZ model equations, and which are therefore not appropriate for use as primary data in other investigations, and (d) while highly sophisticated, the GAEZ framework does not avail itself of all the latest data and developments in the fields of global land use and environmental analysis.

[4] http://www.geo.uni-frankfurt.de/ipg/ag/dl/forschung/Irrigation/index.html

[5] http://www.pbl.nl/en/themasites/hyde/index.html

[6] http://www.geog.mcgill.ca/~nramankutty/index.html; http://environment.umn.edu/gli/index.html

contributions from FAO, IFPRI and the University of Wisconsin (Box 1). These projects have generally produced one-time snapshots, and do not have a plan in place for regularly updating the data. Other projects, such as IFPRI-Minnesota Harvest Choice[7] are focused in specific regions and seek to go into much greater depth on those regions than envisioned in the present proposal. Many of these efforts would benefit from greater coordination by creating synergies and avoiding duplication of efforts.

There have also been some initiatives to store datasets related to the agricultural system in order to facilitate spatially explicit research. FAO's *Geonetwork*, a portal to spatial data, contains information from a wide variety of sources. Most of these data are not global in scope[8]. Moreover, their resolutions are different. This is indicative of most current data base projects which generally represent a compendium of existing data sets[9]. In other words, they contact various researchers around the world, get their data, and put it on a website. They aim to be sites for "one-stop-shopping". They are not generating new data. Moreover, since these efforts involve different groups of scientists, following different data base conventions, the data sets are usually not compatible with each other[10].

Given the volume and diversity of spatial data available today, the unique opportunity of a new database lies in bringing together a variety of datasets in a uniform format within a coordinated context. Those authors studying the impacts of climate change on agriculture must typically gather data from a variety of sources and assemble those datasets in a common data format on a common geographical grid. This can be a laborious task even before considering the massive climate model output datasets that are necessary for exploring the range of possible future climate outcomes. Further, when those climate model datasets are added to the mix, the magnitude of the task is amplified, as researchers must navigate the model archive, acquire the data from each realization of each model for each emissions scenario, 'bias-correct' the model output so that it mimics historical outcomes, and then translate each of those climate model realizations to a common grid over a common time-period. Assembly of these diverse datasets into a single database using a common framework (data format, grid, time dimension, etc.) will

---

[7] http://harvestchoice.org/production/systems/distribution focuses on in sub-Saharan Africa and South Asia.

[8] The only truly global data sets that are publicly available and relate to agriculture seem to be the Farming Systems maps developed by FAO, and the global crop suitability maps created by FAO/IIASA. These global data represent a minute fraction of what we hope to compile through our efforts. When global data bases are available, such as data for land cover, it is typically just a global land-cover classification data set, based on remote sensing or a compilation of various maps (e.g., Olson or Matthews database).

[9] For example the Global Change Master Directory (http://gcmd.nasa.gov/) compiles various datasets from diverse sources. A similar initiative is carried out by Columbia University's Center for International Earth Science Information Network (http://www.ciesin.columbia.edu/). The Distributed Active Archive Center for Biogeochemical Dynamics (http://daac.ornl.gov/ is a NASA data center offering data produced by NASA supported projects. The Global Resource Information Database (GRID) established by the United Nations Environment Programme (UNEP), provides access to an extensive collection of useful maps (e.g., example, http://www.grida.no/). The Carbon Dioxide Information Analysis Center (CDIAC) of the U.S. Department of Energy (http://cdiac.ornl.gov/) compiles an impressive collection of data related to climate and carbon cycle.

[10] Another effort to disseminate data is the Consortium for Spatial Information of the Consultative Group for International Agriculture Research (CGIAR CSI) (http://www.cgiar-csi.org/about-csi/who-we-are) which organizes and distributes the data created by the CGIAR Centers.

remove a major barrier to research that requires – or would benefit from – synthesis of diverse arrays of physical, biological, and socioeconomic data bases.

A related challenge, in addition to the labor and expertise required to build an individual database for each study, is that many options exist for each of the data processing steps (e.g., cleaning, re-gridding, etc.). In the current, *ad-hoc* environment, researchers must build their own individual databases for each project and there is little control over – or even reporting of – the data processing choices that are made. This makes the resulting research output largely non-comparable. Assembly of these diverse datasets into a single database using a common framework will help to create a more uniform approach to data synthesis, and increase quality control across the community. It is important to note that, while this coordination will be particularly beneficial to the general purpose user, this will not limit the creativity of the disciplinary specialist, for whom the original data sources will still be available.

As we project into the future, we envision the Database becoming a vehicle for standardizing geo-referenced information. Indeed, as the institutional setting of the database matures, we envision the creation of new datasets. The Database can also become a forum for discussing crucial issues such as how to assess the uncertainty of the available data and mobilizing the resources needed to improve the quality of these data sets. Nowhere is the need for improved data quality more important than in the area of data on forest carbon stocks. As documented in the recent *Resources For the Future* study of global forest inventories (Waggoner 2009), the current quality of global data on forest area and stocks is very poor. And the underlying satellite measurement systems are extremely limited. While it will no doubt be some time before the level of accuracy in these data bases is satisfactory, we draw some hope from the experience of the Global Trade Analysis Project (GTAP – see Box#2 below) which encountered similar data challenges in the areas of international trade and protection data when the project was initiated in 1992. Now, nearly two decades later, a group of data contributors and sponsors have grown up around that project, and their contributions have produced a global data base which is deemed suitable as input for trade policy analysis and decision making.

In summary, the proposed effort will do three things: 1) Gather national and sub national statistics from various statistical agencies around the world to put together a **consistent** global data set on agriculture and land use; 2) Employ **spatial disaggregation methods** which utilize satellite remote sensing technology (e.g., Ramankutty et al., 2008) and spatial statistics (You and Wood 2006) to produce geographically-explicit gridded data which can be used alongside the administrative unit data; and 3) **Develop infrastructure** (including data base standards and a data portal) for the exchange and distribution of spatial data pertaining to agriculture, forestry and the environment. This infrastructure will include data base and documentation standards, a data portal, and tools for providing data in a variety of convenient formats to the global research community[11] 4) Focus on **transparency** (documentation, meta-data) and sustainability (updates) and **quality control** (cross-checking between data sets, standards test when new data are

---

[11] (E.g., http://www.sage.wisc.edu/iamdata; http://harvestchoice.org/ production/systems/distribution).

uploaded). Most previous efforts have only focused on the data portal element – they rely on other researchers to provide them the data, operating on a 'catch as catch can' basis. By integrating data collection, development, and provision into one project, we will be able to better address the needs of various users.

Box 1. The Agro-MAPS Initiative: Success through collaboration

Agro-MAPS (Mapping of Agricultural Production Systems) was a collaborative project between FAO, IFPRI, and the Center for Sustainability and Global Environment at the University of Wisconsin (http://www.fao.org/landandwater/agll/agromaps/interactive/page.jspx). The project compiled statistics on crop area, yields, and production for countries around the world, aggregated by subnational administrative units, for *ca*. 2000. One of the authors of this proposal was a Co-PI on this project (Ramankutty).

The Agro-MAPS team initially focused on developing a common framework and protocol for compiling the data. Then, each team took charge of a different region of the world and compiled statistics from censuses and estimates of various national statistical agencies. The project was completed within two years. The data have since been spatially disaggregated, as documented in two widely cited studies (Monfreda, Ramankutty, and Foley 2008; You and Wood 2006)). These data are being widely used in variety of global change studies (e.g., Wisser et al., 2008; Johnston et al., 2009; Portman et al., 2010; Li et al., 2010, Neumann et al., 2010), thus demonstrating their usefulness. (See several of the studies in the Annex to this proposal.)

What lessons can we learn from the success of the Agro-MAPS project? First, the project was initiated due to a convergence of interest between the three organizations, who were all interested in compiling such a database. So there was strong incentive for the groups to collaborate in order to reduce the workload for each, and there was also intellectual leadership from all three groups. Second, the groups respected and were confident of each other's contributions. Third, the project started from a well-developed initial framework and data collection protocol. Fourth, outside funding for a portion of the project -- provided by the Millennium Ecosystem Assessment -- helped the investigators to justify their effort expended on this joint undertaking.

Unfortunately, there was no mechanism put in place to update the Agro-MAPS effort on an ongoing basis. Since the initial effort to compile data for *ca*. 2000, there have been no coordinated follow-up efforts and this critical data base on global land use is becoming outdated. Maintaining such an effort on an ongoing basis will require additional funding and the commitment of key institutions to systematizing the collection and compilation of the requisite subnational data.

3 Scope of the Database

The Database facilitates linking spatially-explicit data to satisfy the needs of a wide range of users. To be immediately usable, at relatively low cost, the information in the Database must be consistent across datasets and sources. Such consistency should be defined by a set of minimum rules regarding spatial extent, spatial resolution, the possibility of connecting different datasets, and the crucial distinction between data inputs and model outputs. These rules should in turn be flexible enough to facilitate the integration of datasets built by disparate research groups while allowing for the incorporation of new developments as the Database, and the science behind it, evolves. To ensure the broadest possible access, the Database products will be offered free of charge and delivered in readily accessible formats. To define the scope of the Database we focus on three dimensions: users, contents, and technical capabilities. We now turn to discuss each of these dimensions.

3.1 Users: Who will benefit from the Database?

The Database will benefit the community of researchers who need to access and integrate geospatial data on agriculture, forestry, socioeconomic variables, and climate model experiments in order to understand the adaptability of agricultural systems to climate change, changing food demands, potential biofuel expansion, growing demands for biodiversity, and urbanization.

This community of researchers encompasses several disciplines such as climatology, geography, agronomy, and economics. We illustrate the potential reach of the proposed Database through a series of examples of research applications presented in the Annex to this proposal. These examples are by no means exhaustive, but serve to highlight prominent studies of several issues including:

- *Global climate and carbon cycle modeling*: Global climate modelers (e.g., the LUCID project[12] (Pitman et al. 2009); and numerous others) need global, historical data sets on land-cover change with information on broad land-cover classes (forest, cropland, grassland, etc.) In addition to access to broad land-cover change information, Global carbon cycle modelers (McGuire et al. 2001; Shevliakova et al. 2009), also need information on detailed land-cover transition dynamics (e.g., forest to cropland, cropland to pastures, pasture to forest, etc.), as well as information on carbon stocks.
- *Environmental impacts of biofuel policies*: One of the most controversial aspects of recent biofuel mandates in the US and the EU has been the associated land use impacts. The original estimates by Searchinger et al. (2008) for US corn ethanol suggested that inclusion of GHG releases due to biofuel-induced land conversion might result in a doubling of emissions from substituting ethanol for gasoline. Subsequent research by Hertel et al. (2010) has shown that these estimates are highly sensitive to data on land cover the responsiveness of land use change to cropland returns. The proposed Database

[12] http://www.ileaps.org/index.php?option=com_content&task=view&id=99&Itemid=200

would greatly increase the quality of such analysis by improving both the base data on land cover as well as the data used for econometric estimation of land use change parameters.

- *Econometric assessments of the impacts of climate change on agricultural productivity*: Researchers using statistical methods to assess the impacts of variation in climate on agricultural productivity require spatially explicit observations on agricultural yields, harvested areas, weather, soil quality, terrain slope, land cover, special regimes (such as national parks), to name just a few. Success in identifying key thresholds associated with potential non-linear responses to climate change hinges on the quality of the data (Schlenker and Roberts 2009; Schlenker and Lobell 2010). Econometricians are also interested in deriving variables from geographic and socioeconomic dimensions of the data base such as distance to markets which can be equally important in explaining agricultural yields.[13]
- *Climate model driven assessment of the impacts of climate change*: These studies combine the outputs of coordinated climate models with crop simulation models to provide assessments of future changes in climate on agricultural production and hence patterns of trade and consumption. Recent examples are offered by IIASA (Fischer 2009) and IFPRI (Nelson et al. 2010). Currently participation in such exercises is not available to the average researcher, as the overhead in developing and reconciling all the associated data bases is too large. However, this is an area that would benefit from more involvement by independent researchers.[14]
- *Estimation of future production potential:* Future production potential depends in part on the "yield gap" – that is, the difference between observed yields for a given crop at a given location and the maximum potential yield attainable given current technology. Lobell, Cassman, and Field (2009) review the methods currently in use to assess the yield gap. Licker et al. (2010) utilize a global data set on harvested area and yields to estimate yield gaps globally for major crops. This information is important to agri-businesses and governments interested in boosting future production. When combined with information on input use (e.g., Potter et al. 2010) and economic determinants of yields it can be used to predict which factors are currently limiting yields from attaining their potential (Neumann et al. 2010).
- *Understanding the role of irrigation in global agriculture:* Spatially-explicit time series of irrigated land use are required in many disciplines because irrigation has multidimensional effects. Crop yields are much larger in irrigated agriculture as compared to rainfed agriculture (Siebert and Döll 2010) so that it is required to know whether crops are irrigated or not to develop realistic patterns of crop yields and crop production potential (Bruinsma, 2009, Table 10; Neumann et al. 2010, Figure 4), as well as to analyze spatio-temporal patterns of food security as well. Irrigation is also the major water using sector with a global share of 72% in water withdrawals and 92% in consumptive use (Döll 2009). In addition, irrigation occurs at the interface between land

---

[13] Geographic variables offering exogenous variation for use as instrumental variables are also important. For example, Chomitz and Gray (1996) use soil quality as an instrument for road location.

[14] For example, see the discussion by Hertel (2010) regarding the economic assumptions made in these models.

and atmosphere and increased crop evapo-transpiration results in a significant surface cooling effect (Boucher, G Myhre, and A Myhre 2004; Kueppers, Snyder, and Sloan 2007; Lobell, Bonfils, et al. 2008; Puma and Cook 2010) and modifications of the South Asia Summer Monsoon circulation (Eungul Lee et al. 2009; Saeed, Hagemann, and Jacob 2009) which again effects glacier extent and water resources in highly populated basins like the Ganges, Brahmaputra, Indus, Yangtze and Yellow river basins (Immerzeel, van Beek, and Bierkens 2010). Therefore patterns of irrigation water use are requested as input to an increasing number of climate models.

- *Evaluation of the tradeoffs in ecosystem services associated with agricultural land use practices*: Green et al. (2005) raise the possibility that the best way to preserve biodiversity may be to farm existing lands more intensively. This so-called "land-sparing" strategy towards environmental conservation has introduced a new wrinkle into the biodiversity debate. The proposed data base will be useful in shedding light on the environmental tradeoffs associated with more intensive cultivation of existing lands vs. area expansion.

The foregoing represents a partial list of potential areas of application. How will this Database contribute to their research productivity? This is nicely illustrated by the widespread use of the Ramankutty et al. (2008) and Monfreda et al. (2008) gridded data sets on land cover and land use, spearheaded by one of the authors on this proposal (Ramankutty). These data are available online, free-of-charge, in open and common file formats. They are well-documented in peer-reviewed publications and form a critical foundation for several of the studies cited in the foregoing list.

Another recent example of the tremendous impact such a publicly available, standardized data base can have on the process of discovery may be drawn from the climate change community. This shared data base is formed from the outputs of coordinated climate model experiments – a key component of any climate change impacts study. These experiments form the primary resource for researchers who are not climate modeling experts, and such coordinated experiments have formed the backbone of recent assessments of the Intergovernmental Panel on Climate Change (IPCC). With the advent of the Fourth Assessment Report, the results of these coordinated climate model experiments have been made available in centralized archives. The availability of these data to the broad international community has literally revolutionized the science of climate change, enabling thousands of researchers from disciplines ranging from fluid dynamics to economics to directly access a large suite of climate model experiments generated by climate modeling centers around the world. However, these archives comprise massive data volumes from numerous experiments using two dozen different climate models employing a variety of different geophysical grids. Access to, and refinement of, individual variables of interest for analysis of agricultural impacts of climate change therefore requires expertise in data acquisition, subsetting, and interpolation. Therefore, "post-processing" of the primary climate model output into uniform formats, and archival of those post-processed data in a central database alongside other agriculturally relevant data, would be a great benefit to the large community of researchers whose expertise lie outside of the climate modeling domain.

By integrating a comprehensive set of variables (discussed in the next section), the proposed Database will offer researchers the opportunity to access information that otherwise may be inaccessible due to the lack of time, skills, or equipment[15], thereby bringing new talents and ideas to bear on these pressing questions. Besides lowering the costs of entry into the field of spatially explicit analysis of global agriculture, a centralized and consistent Database will facilitate the standardization of the procedures used to process and integrate different pieces of information. By reducing the costs of assembling the data, available resources can be focused on modeling and analysis efforts instead of on database construction. This makes the research process more efficient, with the added benefit of facilitating the comparison of model outputs and results. In short, the proposed Database will benefit the research community by encouraging competition and growth on the analytical side, even as researchers share a common state-of-the art platform for global data.

3.2 Contents: What is in the Database?

To be successful, the Database must supply the inputs necessary for existing as well as evolving global models of agriculture, climate and the environment (see, for example, the recent volume edited by Hertel, Rose, and Tol, 2009). This means it must be global in coverage, consistent in treatment of variables, and exhaustive in its coverage of key determinants of agricultural productivity and environmental impacts. It should also support state of the art statistical/econometric research aimed at better understanding the physical and economic relationships underpinning the global agricultural system and its impacts on the environment (e.g.,Lobell, Burke, et al. 2008). For this reason, a time series dimension is also important. Simple cross-section analysis which is all that is possible with most currently available spatial data sets can often be misleading due to the role of omitted variable bias. Indeed, this can even lead to getting the sign of climatic effects on agricultural productivity wrong (Schlenker, Hanemann, and Fisher 2005)

The dataset is global in scale, however, it will benefit from contributions and enhancements by regional collaborators. Indeed, in many cases, global datasets are the product of assembling sub-global or regional datasets (e.g., AgroMaps). Therefore, it will be essential to encourage national contributions/updates/upgrades, as well as active participation of research groups with an explicit regional focus (e.g. Africa or Latin America). Indeed, such regional partnerships will be essential in Africa, where current international information on land cover and agricultural production is of dubious quality. Here project such as *Harvestchoice* will be key partners. Encouraging, the contribution of updated components of the global data bases will be a key piece of the data base

---

[15] By way of example, consider the following: one of the authors of this proposal needed to use the weather data from the Climate Research Unit for an econometric exercise. At the time the research was being undertaken, these data were available as large ASCII files containing half-degree grids of monthly information for the period 1960-2002. The average size of these files was 400 MB. To access them using standard commercial GIS software was exceedingly laborious, as it required bursting the dataset in monthly files (504 of those!) and adding to each one an appropriate header to be recognized as an ASCII raster. Another possibility was to compile a piece of Fortran code available at the CRU website, but this posed a problem for users without Fortran experience. Luckily, the files were available elsewhere on the web, translated into NetCDF format. Today, CRU makes these data available in NetCDF format, thereby greatly facilitating the extraction and subsetting of these important datasets. In short, simple differences in data format can present a formidable barrier to entry in this field.

architecture to be further discussed below. Over time, the success of the Database can stimulate the use of globally consistent methodologies for obtaining information, thereby facilitating the process of assembling regional datasets.

All the datasets will be available on a gridded basis. Global data are often collected in various spatial formats. Biophysical data are often collected at point locations (e.g., climate, soils), but with the advent of remote sensing, these data are also available in spatially gridded formats (e.g., land cover). Socio-economic data are most often collected using sample surveys involving specific locations or households, but are almost always aggregated to administrative units (e.g., a county) for confidentiality reasons. Compilation of these different types of data without further processing will result in data with different spatial formats, at different spatial resolutions, which would be nearly impossible to use by individual researchers. Therefore, we plan to put much effort into making data available in a consistent, spatially gridded (raster) format. If data is produced at a fairly high spatial resolution, this would allow for use by a wide range of researchers who can either aggregate the data to a coarser gridded resolution for use in their biophysical models, or to larger political units for use in socioeconomic models. An example of a utility that automates such aggregation is the IAMDATA website (http://www.sage.wisc.edu/iamdata/).

In many cases creation of a common, gridded format will require disaggregating data that is only available at lower levels of resolution (i.e, subnational administrative units). Various disaggregation methods have been developed, with different levels of sophistication (e.g., K. Klein Goldewijk, Van Drecht, and Bouwman 2007; Navin Ramankutty et al. 2008; Monfreda, Navin Ramankutty, and Jonathan A. Foley 2008; Liu et al. 2010) that utilize satellite data and other ancillary data such as population density, climate, etc. For many purposes this can be harmless. However, for statistical exercises, gridding data can convey a false sense of variability while introducing spatial correlation between observational units To be compatible with the many potential users of the Database, it is necessary to fully document gridding procedures, and whenever possible, include complementary data to inform the researcher about the underlying variability. For example, the CRU data provide geo-referenced information on meteorological stations so users can form their own opinion about the accuracy and completeness of the available data. In the case of socioeconomic data, the information should be available at the most disaggregated level provided from the original source; for instance, census data in the US should be available at the county level. This will help to clarify the highest level of disaggregation at which variation in the data is meaningful.

The database will have *minimal standards* for geographical resolution, product/land use class coverage. Meta-data according to accepted standards must be provided.

A unique advantage of the Database is documentation through both full-fledged documents and metadata. This will allow users to discriminate between data inputs and outputs. As illustrated by the literature review in the Appendix, this distinction is application-specific. For example, climate model outputs can be the inputs of a climate impact assessment. Likewise, potential yield estimates from a crop model can be inputs in a yield gap study.

A tentative listing of key variables is shown in Table 1. All of the variables would be assembled at their maximum attainable level of resolution. The goal would also be to build a time series for all of these variables that change over time at annual scale. The reader will note that this data base focuses on variables which can be measured in physical units. Ultimately, it would be appealing to include comprehensive economic data, including prices as well as economic values associated with agricultural outputs and inputs. However, obtaining these data has proven problematic even for the UN-FAO and therefore it is beyond the scope of the current proposal.

3.3 Technical capabilities: How will the users interact with the Database?

There are many options to deliver the Database, however, the ideal underlying technology must ensure (1) the ability of the database to relate disparate datasets through relational database queries (2) the possibility for researchers to upload their datasets so that they can be readily related to what is already stored in the database, and (3) capabilities for connecting with other databases.

The Database should be hosted on a web server with database and geospatial capabilities. The standard language for database handling is SQL. To be useful the system should allow the implementation of relational databases, thus the compatibility of the underlying datasets is crucial. This implies that either the grid cells are standardized, or "on-the-fly" methods of re-sampling are included. The web server must allow the possibility of uploading data that meets the minimum standards discussed in the previous section.

Another aspect to consider is the ability of the Database to connect with datasets stored elsewhere. One option is to have copies of all the datasets listed below in one physical place (server). Another is to allow querying different places. For example, temperature and precipitation from the CRU could be queried simultaneously with the data on production and yields from Monfreda et al., and perhaps the urban-rural population data from CIESIN. Of course, this will require tremendous coordination efforts and can be seen as future step.

Another technical aspect relates to the accessibility of the Database. Here, we plan to follow two principles: (1) All the datasets are available free of charge for non-profit use as long as the sources are acknowledged and (2) these are delivered in formats widely used by the research community.. By virtue of (1) we ensure accessibility to the broadest research community, however, contributors to the dataset can decide precisely what to distribute. Item (2) ensures the compatibility of downloaded data with a large number of software applications and finally.

Undoubtedly, other data centers such as the IPCC Data Centre[16], the IIASA-FAO GAEZ framework[17], Earth System Atlas[18], and CIESIN[19], will be valuable sources of information for designing the technical infrastructure of the database.

---

[16] http://www.ipcc-data.org/

[17] http://www.iiasa.ac.at/Research/LUC/GAEZ/index.htm

[18] http://www.earthsystematlas.org/

Table 1. Variables to be included in the proposed Database.

• Agricultural production and management:: o Crop production, harvested area (implying also yields) and cropland area; information on fallow land, multiple cropping, crop failures and temporary pastures will be necessary to link the harvested area to the cropland cover. Also distinguish area covered by tree crops. o Crop inputs, starting with irrigation, fertilizer use, pesticide use, and extending to other inputs, as available. o Crop calendars (sowing dates, harvest dates) o Crop cultivar distribution o Livestock numbers, including the distinction between grazing animals and confined feeding and derived outputs (meat and milk yields), including raising animals. o Pasture land and grazing land, with a sharper distinction between the two
• Physical Environment: o Soil type, carbon content, pH o Topography (Elevation, Slope, Aspect) o Watershed o Temperature on a daily/monthly basis: average and daily min/max o Precipitation on a daily/monthly basis: average and daily min/max o
• Forests: o Forest type o Age class distribution o Timber production o Carbon stock: both above and below ground o Accessibility data o Management costs o Timber value or price o Ownership status o Ecosystem type o Proportion of land under sustainable management o Forest accessibility
• Socioeconomic variables: o Population (Urban vs. Rural) o GDP (from Gecon project) o Roads, navigable rivers, and lakes o Irrigation o Areas under special regimes (national parks, etc.)

[19] http://www.ciesin.org/

4 How will the Database work?

4.1 Institutional design

The success of the Database will hinge critically on the institutional design of the underlying project. In this section, we summarize some basic lessons drawn from three prominent previous experiences in which the authors have been intimately involved, namely the Coupled Model Intercomparison Project (CMIP3), the Global Trade Analysis Project (GTAP), and AgroMaps.

Lesson 1: Investment in the technical infrastructure and human capital that enable access to a broad community of researchers can revolutionize a field of inquiry.
In the case of the CMIP inter-model comparison effort, the initial generations of this archive were focused almost exclusively on the climate modeling endeavor, and data access was mostly limited to those participating in the inter-comparison exercises. As a result, while the collaboration was very important for the development of physical climate models, the scientific impact outside of that community was limited. However, there was a huge, step-wise transition in impact between CMIP2 and CMIP3. The defining difference is that a group of leaders within CMIP agitated to make the dataset available and – critically – that funding agencies saw the potential benefits and agreed to support the archiving effort. And it was also important that the individual institutions at which the climate modeling groups reside were willing to support the extra effort to archive the results of the experiments.

Lesson 2: While the availability of data can enable scientific innovation, it can also limit the kind of science that is done.
In the context of climate models, not only is a class of science emerging that is oriented around analyzing multi-climate-model experiments, but those analyses are necessarily limited by the classes of variables that have been archived. As a result, while we may learn much about those classes of variables, other aspects of climate are not explored simply because of choices that were made about which variables to store (and, critically, which not to store). However, if the database is designed to be flexible and the people involved in the database construction and management are engaged with the community, then successive generations of the database can potentially overcome these limitations.

Lesson 3: By having access to a broader, more comprehensive, more user-friendly database, the community is likely to contribute valuable insights about future improvements to the database.
The existence and availability of the coordinated model database has even spurred development and synthesis of observational datasets that were not needed (or perceived to be needed) prior to the database being available to the broad community. In the case of the CMIP effort, the agricultural research community has provided valuable feedback about the types of climate variables that are needed for the most sophisticated research (e.g., sub-daily-scale temperature throughout the year). It is therefore critical that the design of a new coordinated database be built to be flexible in the face of unforeseen developments in both data availability and data demand.

Related to this point, the GTAP experience has shown that the mere existence of a consistent database has triggered mechanisms for improving data quality. Viewed in the context of this spatial data base proposal, the credibility of some datasets such as those on forest stocks is quite low, as previously noted. The network effects of the Database can help to increase the demand for quality, offering incentives and new mechanisms for producing better data.

Lesson 4: The availability of an interesting, coordinated and well-managed database can spur the demand for greater data volume.

A related lesson is that the availability of an interesting, coordinated and well-managed database can spur the demand for greater data volume. The CMIP3 archive comprises approximately 36 Terabytes of data. However, due to the high demand, the total data downloaded is approximately 15 times the data being stored. Further, for the next CMIP effort (CMIP5, which will support the IPCC 5th Assessment Report), the expectation is that the data volume will exceed 1 Petabyte. This increase in expected data volume results partly from the interest in generating a more complicated and comprehensive set of climate model experiments. However, it also results partly from the excitement in the community about the value of the CMIP3 database, and the community's requests for archiving more data in order to enable more sophisticated analyses. As a result, the plan for CMIP5 is not only to conduct a larger suite of climate model experiments, but also to archive more fields at higher time resolution for a much larger subset of the simulations than was archived in CMIP3.

This level of data volume clearly creates infrastructure challenges. However, it also creates opportunity for both physical and virtual infrastructure development. Because of the high impact of the CMIP3 database and the large data volume expected for CMIP5, a variety of experts in fields outside of climate science are investigating ways to archive and analyze the CMIP data most efficiently. In addition to helping to make valuable climate model data accessible to the agricultural research community, these infrastructure lessons – and technologies – will likely be very valuable for the construction and operation of a new global, spatially explicit database for agricultural research.

Lesson 5: Incentives, incentives, incentives.

Absent special initiatives such as those being proposed here, there is little incentive for development and maintenance of databases. Scientists are typically rewarded for publications, and it is hard to publish datasets. Development of an original database also typically requires far more effort than a scientific publication using secondary data. So if an individual scientist does invest in a development of an original database, they have a strong incentive to keep it to themselves until they have fully exploited its publication potential – at which point it is often out-of-date. Public institutions often have a mandate to provide data to the research community, and many do an admirable job of this. Unfortunately, they are also limited by their founding charter. So that the UN-FAO maintains data on food and agriculture, but not on climate, while the Climate Change Research Unit, which maintains the climate data, does not offer comparable data on agriculture and land use. In addition, publicly provided data bases often become stagnant over time, as they have a strong incentive to keep doing things the way they have done in the past, and are slow to respond to the rapidly changing needs of the research community.

The Global Trade Analysis Project (GTAP), founded 20 years ago by one of the authors of this proposal (Hertel), provides an interesting study in incentive design (Box 2). GTAP's main activity is the production of a global, economic data base used for analysis of national impacts of trade and environmental policies. Individual contributors are offered privileged access to the data base (others are charged for the data base and receive it later in the production cycle). The work of data contributors is documented in the core data publication and receives extensive citations; significant data contributors are recognized as GTAP Research Fellows, and their work is highlighted at GTAP courses and conferences. In short, data developers are veritable celebrities in the GTAP network.

Core support for the GTAP activity comes from consortium of national and international agencies, of which more will be discussed below under the governance section. However, from the standpoint of incentives, continued support for the project is contingent on continued updates and improvements to the database. While we do not envision charging for the spatial data base proposed here, we do envision a pre-release phase during which the data are vetted and data contributors and sponsors have privileged access to the data base. A public release will follow this pre-release with a lag of 6 months to a year. This can be adjusted to ensure that all those who have something substantial to contribute will have an incentive to do so.

Lesson 6: Databases supported by solid institutions help to ensure sustainability, facilitate updating, and act as gravitational centers of scientific discovery.

The success of the coordinated, multi-climate-model experiment and freely available archive is such that, although it clearly requires tremendous effort on the part of the climate modeling groups, pressure now exists for climate modeling groups to participate and have their institutions represented, rather than be absent from the database. This is true for the global climate modeling groups involved in CMIP, as well as for regional climate modeling groups involved in efforts such as CORDEX. Likewise, in the broader scientific community, reviewers and editors treat analysis of the climate model archives as a critical component of studies that examine some aspect of physical climate change -- even to the point that there is an expectation that authors will make use of whatever appropriate archived climate model output is available in order to produce a publishable paper.

Box 2. The Global Trade Analysis Project (GTAP): A Potential Role Model?

The Global Trade Analysis Project was founded in 1992 and offers a highly successful example of global data base infrastructure – albeit at the national-level – for analysis of trade and environmental policy issues. At the time of its founding, global economic modeling was confined to a few, well-resourced institutions around the world which had invested significant resources into the necessary data base and software infrastructure (Powell 2007). Estimates of the potential global welfare gains from the Uruguay Round of global trade talks which was underway at the time varied from $40 billion to more than $1 trillion. However, results from these models were not generally comparable, and modelers explaining these differences tended to emphasize differences in model specifications – as opposed to differences in the underlying data bases and policy scenarios. It was virtually impossible to test these hypotheses about underling sources of difference, since none of the groups allowed for replication of their findings by those outside the institution. In short, the global trade community was roughly in the same place as the global agricultural modeling community is with respect to global economic impacts of climate change (Alexandratos 2010).

Once the diverse modeling groups analyzing the World Trade Organization's (WTO) Uruguay Round adopted a common GTAP data base and a common GTAP-based set of trade policy scenarios, the range of welfare gains dropped dramatically, varying by a factor of roughly twenty percent, depending on the modeling assumptions utilized. Clearly the great majority of the differences in these results were due to the use of different data sets – in particular, different measures of current protection levels, as well as the shocks being dictated by the WTO agreement (Martin and Winters 1997).

GTAP was established with the express objective of lowering the entry barriers to global economic analysis, thereby involving a greater variety of individuals from many countries around the world – both as data contributors and as users of the data base. Key features of its design include: (1) a consortium of leading national and international institutions involved in global economic analysis which provide financial support as well as advice and data access for GTAP; (2) a Center at Purdue University which acts as the collection point for data and which constructs the global data base; (3) regular courses designed to introduce new network members to the data base and modeling framework; (4) an annual conference where GTAP-related research is presented -- recent conferences were hosted by the United Nations in Malaysia (2010), Chile (2009), Finland (2008) and attracted 200-250 participants; (5) an active network, knitted together through the Purdue-based web site and listserv, as well as a peer-reviewed GTAP technical paper series.

Since its establishment in 1992, GTAP has published 7 public releases of its global data base, which is currently used by more than 8,000 individuals in 150 countries around the world. Its use has expanded beyond trade policy to include studies of global environmental issues, including climate change. However, it is not oriented towards spatially resolved data. *The current proposal would not tie into GTAP in any direct fashion.* Rather it simply seeks to emulate aspects of the institutional design of this very successful project.

4.2 Institutional infrastructure

Based on the forgoing analysis, as well as our extensive discussions with those currently working in this field, we believe that there are several levels of potential engagement/funding for such an effort. We explore these here, beginning with the simplest approach and ending with a full-blown Database Center:

a) Ignore the lessons listed above and just **maintain a web page**: This web page would contain links to the various data bases, pertinent to this project, that are currently posted on the web.

   i. This has the advantage of being a very low cost option. Indeed, the cost is so low that numerous such efforts have been springing up on the web already.
   ii. *However, the low costs are accompanied by low benefits*; the data will remain largely inaccessible to the general research community without considerable IT investment. Different data sets will be provided in different formats, with different mnemonics, different spatial resolution. Quality management is not possible in such an environment (e.g., limited meta-data attached to data sets). Resources are not sufficient to stimulate new contributions of data.
   iii. There is no real continuity offered here. Indeed, there is a real danger that data sets will be lost if providers delete the data from the web.
   iv. Resource requirements: roughly one-half FTE/year.

b) Invest in a few **full time staff, at distributed locations**:

   i. At a minimum this would involve funding *two individuals*. By having two individuals involved over time, there is some diversification and continuity if one person leaves.
   ii. They could be *located anywhere in the world*, and need not be in the same location. They would ideally be chosen to represent the *diverse interests of the project*, e.g. natural sciences vs. social sciences or geography, but the most important thing would be for them to have significant data base skills. They deliver monthly reports to the governing body of this project; annual reports to the scientific committee.
   iii. Rather than simply accepting whatever is available on the web, these permanent staff members *actively network* with those working in this field to *identify new data bases*. Once secured, these individuals *convert* these data bases to a *common format* and *collect and check meta-data* before putting them on the web.
   iv. Maintain a *listserv* linking those with an interest in the project/data bases. (I.e. take the lead in responding to questions.)
   v. Resource requirements:
      a. 2 FTEs
      b. Travel budget for 3 conferences/year for each individual
      c. Rental of a server 'in the clouds'

c) **Distributed network of nodes, with one administrative center**:

i. Add to the prior option, a full time project coordinator and a half-time IT person.
ii. A feature of this option is the use of advanced cyber-infrastructure for hosting and delivering the contents of the database. In particular, a platform such as Purdue's HUBzero™ Platform for Scientific Collaboration[20] would be needed. Through this type of hub, data providers will be able to network and share information, as well as to create, publish and access interactive visualization tools. The use of search facilities, conversion tools, and GIS functionality will greatly enhance the usability of the Database. Moreover, the possibility of enhancing such tools with computational power is facilitated by the connections between HUBzero and large clusters of computing resources.
iii. While there is an administrative center, the team of researchers involved in data base construction is distributed across nodes. Each node represents an international research leader in some aspect of spatial data base development/analysis.
iv. *Project coordinator* is an active champion for the project, seeks to build network and encourage new contributions. Additional focus on data consolidation, statistical work to harmonize data sets and to fill missing observations. Actively identifies data gaps, attracts third party funds by becoming involved in research projects to enlarge the data base and improve its quality.
v. Resource requirements:
    a. 1 FTE/year/node, with 1.5 FTEs at the host institution
    b. Travel budget to bring members together once a year
    c. Cyber-infraestructure costs. (around $50,000/year).

d) **Foresight Center**: This Center would be based on a scale approaching the GTAP Center with a Director, several full time staff members, visiting fellows, a full time IT person, and possibly a full time administrative person.

i. In addition to the activities listed under the previous option (c), the Center would foster an *active network* by *offering courses*, organizing an *annual conference,* maintaining a mailing list, etc.
ii. Visiting fellows, supported by Center staff, would undertake *major improvements to global data bases.*
iii. *Center staff could facilit*ate future projects along these lines; visiting research fellows could lead one-off improvements to the data base/IT infrastructure; Center would solicit grants to pay for such improvements as well as methodological developments (e.g., data base fusion a la Ramankutty).

[20] M. McLennan, R. Kennell, "HUBzero: A Platform for Dissemination and Collaboration in Computational Science and Engineering," Computing in Science and Engineering, 12(2), pp. 48-52, March/April, 2010.

iv. Center Director would solicit national data base contributions from individual countries.

5 Governance and Sustainability: How can we ensure continuity?

It is clear that considerable thought must be put into setting up the governance structure for the Database. We believe there are two key elements required for this project to be effective over time: the *Scientific Committee* and the *Governing Board*. The Scientific Committee will comprise leading experts in spatially explicit analysis of global agriculture, forestry, environment and climate issues. We have already made some progress in identifying potential members of this committee who were peer-reviewers of this proposal (see cover page for a list of scientific reviewers.). This group can be expanded, with each individual serving a 3 year term, allowing for rotation of new individuals onto the committee. A total of roughly a dozen members should allow for suitable diversity of backgrounds without becoming unwieldy. The responsibility of the Scientific Committee will be to make recommendations on data base content and methods of delivery.

The Governing Board will oversee management of the project. They will be responsible for hiring the full time staff, or delegating such hiring to the Project Director, as well as selecting the host site(s) for this Project. The Governing Board will solicit funds for the Database and provide long term strategic direction for this Project, in particular ensuring that the needs of the global policy making community are met. Governing Board members will represent the major donors behind the project, with two “members at large” having the role of serving the broader public interest.

6 Where do we go from here? A pilot project proposal with “multiple nodes”

This is a very ambitious project, as the external reviewers of this proposal have frequently reminded us. Accordingly, we propose to begin with a two year, pilot project which will demonstrate proof of concept, and which will provide a more limited set of data to the global scientific community. This pilot project will also allow a half dozen data base ‘champions’ to take the lead in this work – each one operating as an independent node, linked together, both bilaterally and through the administrative center, which will be housed at Purdue University during the pilot phase. After the pilot phase, many of the most difficult uncertainties associated with a project of this scale should be resolved, or at least clearly defined. We propose to begin by focusing on land cover, agricultural production, and land use – leaving the more challenging issue of forest stocks to phase two (the full-scale project).

An issue brought up by many of the reviewers of this proposal is the need for individuals who will champion the Database initiative. Therefore, in the pilot phase of this project we propose to start out with a "multiple nodes" model (closest in design to choice c in the institutional design section above – although with more modest funding for each node). The initial goal will be to provide support to individual nodes to expand current activities at each of these sites, with the explicit goal of making the fruits of their labor available to the global research and policy making community in a standardized format. By carefully selecting a set of nodes, this approach also ensures that there is a leader at each institution with a vested interest in the continued development of the data base. As the project develops, we anticipate that these research leaders will become project champions.

The two year pilot project will also establish a framework for long-term collaboration between the nodes. Currently we have agreement from the following leading researchers to serve as research nodes (areas of expertise in parentheses): Navin Ramankutty -- McGill University (global land cover and agricultural land use), Stefan Siebert -- Bonn University (global irrigation), Stanley Wood -- IFPRI (land cover, agriculture and land use data for Africa), Glenn Hyman -- International Institute for Tropical Agriculture (CIAT: land cover, land use and agricultural data for Latin America), Andrew Nelson – International Rice Research Institute (IRRI: land cover, land use and agricultural data for Asia) and Noah Diffenbaugh -- Stanford University (global climate data – both historical and prospective). Project coordination in the pilot phase will be undertaken at Purdue University, which has extensive experience in managing global data consortia and which is also the home for the HUBzero platform for scientific collaboration (http://hubzero.org). At the end of the pilot phase of the project, all of the nodes as well as the project coordination site will be re-evaluated in the context of a larger, phase two effort.

We estimate that the cost of this pilot phase of the project will be roughly $600,000/year over the two year period. Detailed plans of activities for each of the six nodes are provided in Annex 2. These plans outline current data availability, as well as data commitments in the pilot phase and beyond. As this pilot phase develops, donors will have the opportunity to review progress and discuss continued and funding for the future. In short, the pilot phase will provide an opportunity for both the nodes and the funders to test the system first before making long-term commitments to this project.

Conditional on the success of the pilot phase, a next step will be to seek funding to support formation of a scientific committee which will review progress under the pilot project and lay out the detailed specifications of the full scale data base and the underlying IT architecture. Simultaneously we envision establishment of a governing board which will include representatives from the major sponsor agencies as well as key “players” in the field of global agriculture, forestry and the environment. The governing board would oversee a call for proposals to host this global data base. The initial contract would be for 5 years, renewable for another 5 years pending successful performance reviews.

7 Concluding remarks

There is no better way to understand the value of this type of data base than to see it at work. Annex 1 contains a review of several recent applications using geospatial data to analyze the global agricultural and forestry system which have yielded important scientific insights and played an important role in policy making. Annex 2 details the specific proposals made by each of the proposed hubs in the pilot project.

Annex I. Some applications using geospatial data

This Annex documents the kind of scientific advances which can only be accomplished with access to spatially explicit data on agriculture, forestry and environment. The proposal outlined above would greatly lower the entry barriers to undertaking such studies, thereby increasing the pool of talent available for working on these important issues. In addition, the quality of many of these studies would be improved by drawing on a higher quality, richer, global spatial data base. Many of these authors had access to only a portion of the data discussed in this proposal – and often the data they use was only available for a specific country/region. In short, this proposal would have a major impact on the quality and quantity of research conducted on global agriculture, forestry and environmental issues.

Examples from the biofuel policy modeling arena

Two recent applications of simulation model analyses of the environmental impacts of biofuels, both of which are built on spatial data are provided in figures 1 and 2. The first focuses on GHG emissions from land use change, while the second highlights impacts on nitrogen runoff due to changes in the pattern of livestock production as well as intensification of crop production.

Global land use impacts of US maize ethanol production

Figure 1 is taken from the Hertel et al. 2010 study of the global land use impacts of US maize ethanol production. In that work, land cover and use are modeled at the level of Agro-Ecological Zones (AEZs) such that land cover change due to increased feedstock production as well as production of displaced crops in other regions depends on competing activities within the same AEZ. From the top panel we can see that the land cover change arises disproportionately in the US – the origin of the shock – and comes disproportionately from pasture land. Somewhat surprisingly, there is modest forest reversion in many regions. At the margin, forestry activity shifts to AEZs where there is little direct competition with feedstocks or competitor crops. The lower panel translates these land use changes into GHG emissions, reported in terra-grams of CO2 equivalent. Here, we see that, despite the predominance of pasture in land conversion, the bulk of emissions come from forest conversions. These results were used by the California Air Resources Board to aid in formulating their Low Carbon Fuel Standard, which sought to account for GHG releases from indirect land use change.

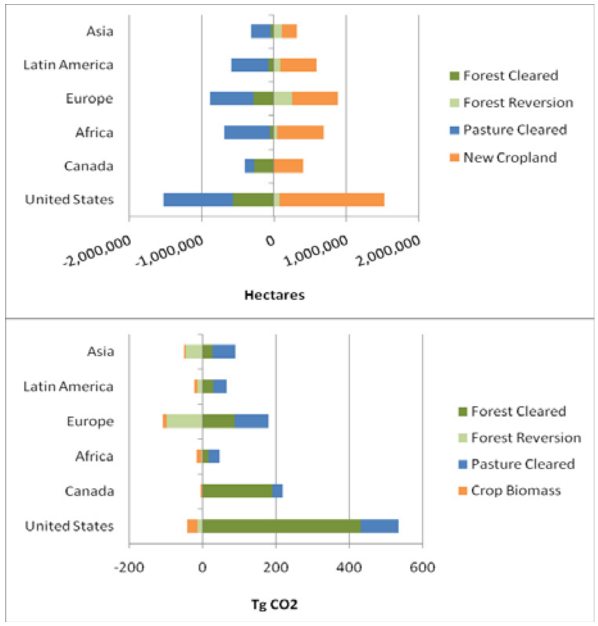

Figure 1. Global land conversion and associated GHG Emissions due to increased US maize ethanol production of 50.15 gigaliters per year at 2007 yields, by region. Source: Hertel et al. 2010.

Land use impacts of biodiesel expansion in the EU

Figure 2, taken from Britz and Hertel (2009), reports changes in nutrient surplus in the wake of an expansion of EU biodiesel production to satisfy 6.25% of liquid fuels for transport in the EU. This is a massive expansion, which also triggers indirect land use changes such as those highlighted in Figure 1. However, it also results in an intensification of production. Figure 2 reports the resulting changes in nutrient surplus – and hence potential nitrogen run-off -- in kg/ha, by NUTS region in the EU. The variation in potential local environmental impacts is striking, ranging from more than 3kg/ha increases in the rapeseed producing regions of the EU, to potential environmental improvements in other parts of the EU where the main impact is through reductions in livestock numbers (due to higher feed costs).

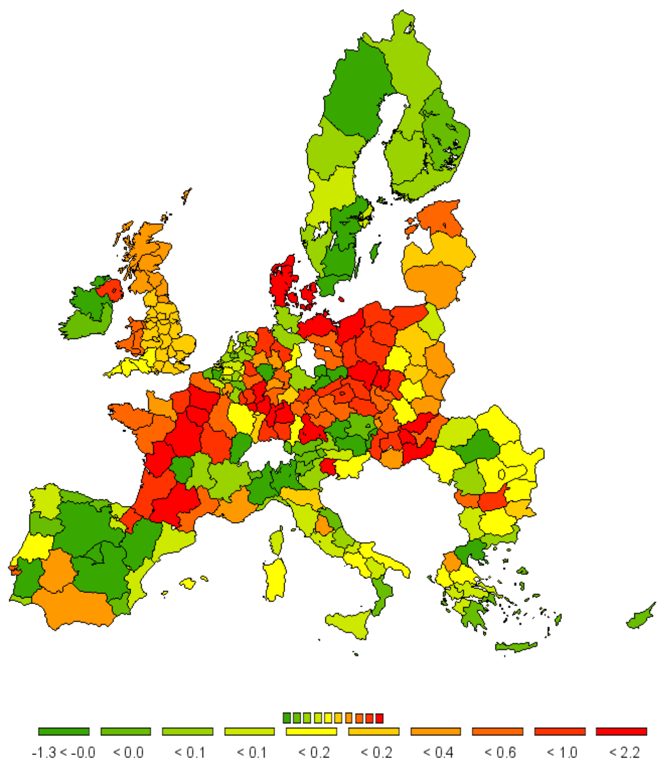

Figure 2. Changes in nutrient surplus in kg/ha due to EU biofuel mandates. Source: Britz and Hertel (2009).

Assessment of climate change impacts:

Impacts of climate change in sub-Saharan Africa

Spatial databases on weather and agricultural productivity have proven useful for estimating cross-country agricultural production functions that explicitly link weather to productivity thus allowing predicting the impacts of climate change.

An interesting example is provided by Schlenker and Lobell (2010) who combine historical crop production and weather data generating a model of yield response to climate change for key crops in a number of Sub-Sahara African countries, using data from 1961-2000. A methodological challenge of this work is to match annual production data available only at the national level (from FAOSTAT) with gridded weather data available at daily or monthly intervals. The authors tackle this mismatch by combining the gridded datasets of precipitation and temperature (NCC o CRU 2.1) with the percentage of each grid-cell under cultivation from Monfreda et al. (2008), and grid-cell specific growing seasons (Lobell, Burke, et al. 2008), thus obtaining annual values used as independent variables in a panel regressions.

The parameter estimates of these regressions are then used to predict changes in yield using the output of 16 climate change models. An advantage of the econometric approach is that it yields confidence intervals around the predictions that take into account both, uncertainty from the climate models and uncertainty from the parameter estimates. This is shown in Figure 3, which displays the predicted impacts by country, where crops (maize, sorghum, millet, and groundnuts) are shown in the rows, and the three columns give the 5 percentile, mean, and 95 percentile of the impacts. While the impacts are generally less severe for the low fertilizer subsample (all except South Africa and Zimbabwe), they are still statistically significant, i.e., even the 95th percentile of the damage distribution is still negative (has a red color) in the right column for most crops and countries. This occurs despite rainfall increases in several of the climate projections, particularly in Eastern Africa.

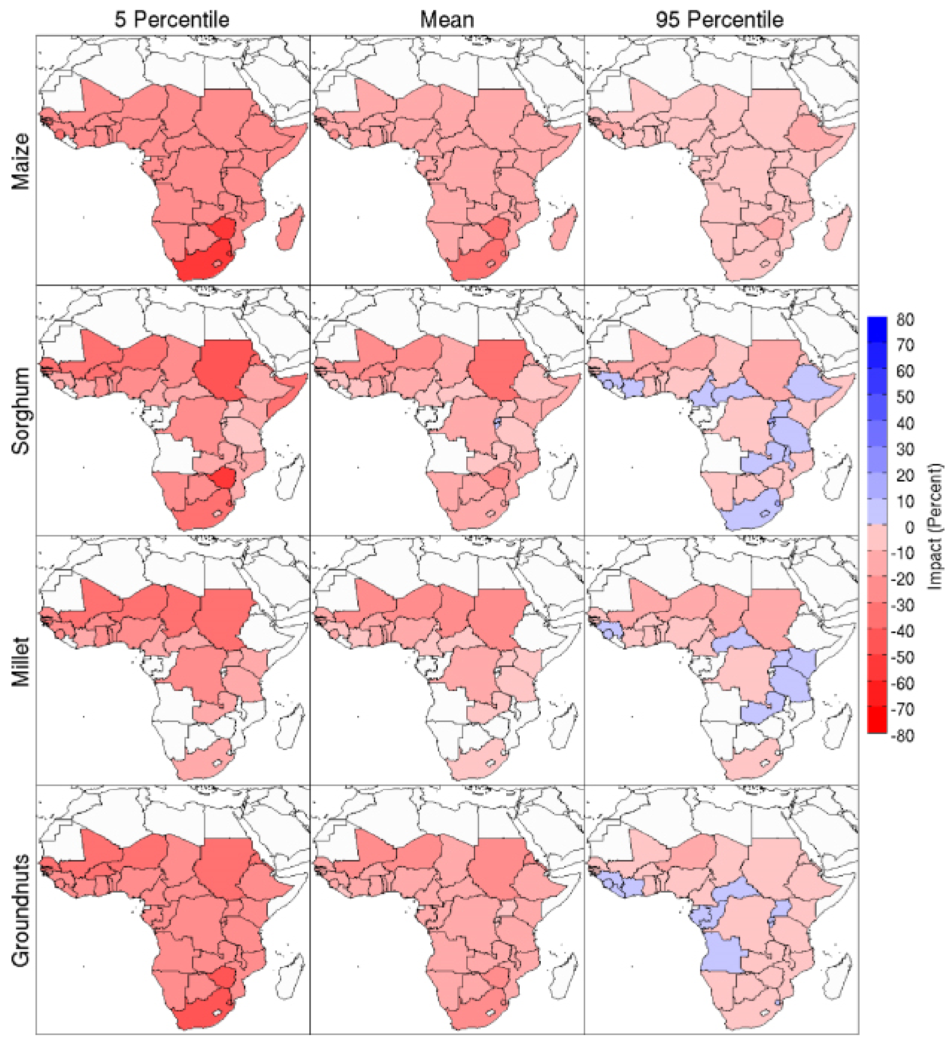

Figure 3. Distribution of impacts from climate change by country (per cent yield change). Mean impacts (middle column) as well as the 5 and 95 percentile (left and right column, respectively) are shown. Each row represents one crop. Source: Schlenker and Lobell (2010).

Nonlinear effects of temperature on US agricultural yields

Another example of econometric work that exploits spatially disaggregated datasets is provided by Schlenker and Roberts (2009). They pair a panel of US counties' yields for corn, soybeans, and cotton with a fine-scale weather dataset that incorporates the whole distribution of temperatures within each day and across all days in the growing season. Their results are summarized in Figure 4, which shows that yields increase with temperature up to 29° C for corn, 30° C for soybeans, and 32° C for cotton but that temperatures above these thresholds are very harmful. As is evident from the plots, the slope of the decline above the optimum is significantly steeper than the incline below it. The authors find the same nonlinear and asymmetric relationship when they isolate either time-series or cross-sectional variations in temperatures and yields. They suggest that this evidences limited historical adaptation of seed varieties or management practices to warmer temperatures because the cross-section includes farmers' adaptations to warmer climates and the time-series does not.

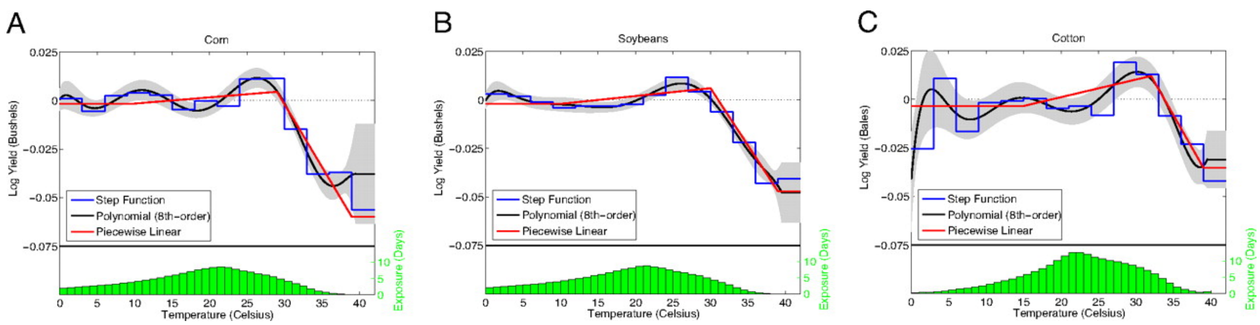

Figure 4. Nonlinear relation between temperature and yields for corn, soybeans, and cotton. Graphs at the top of each frame display changes in log yield if the crop is exposed for one day to a particular 1° C temperature interval where we sum the fraction of a day during which temperatures fall within each interval. The 95% confidence band, after adjusting for spatial correlation, is added as gray area for the polynomial regression. Curves are centered so that the exposure-weighted impact is zero. Histograms at the bottom of each frame display the average temperature exposure among all counties in the data. Source: Schlenker and Roberts (2009).

Climate impacts on the wine industry

The premium wine industry is highly valued worldwide. In the U.S., the California industry alone is valued at more than $45 billion annually ((White et al. 2006); Wine Institute of California www.wineinstitute.org). Although a large suite of conditions contributes to the quality of premium wines (White et al. 2009), the highest quality winegrapes are grown within a relatively narrow temperature envelope (e.g., (Jones et al. 2005; White et al. 2006)). White et al. (2006) have used known relationships between temperature and wine quality, geographic range, and plant physiology to develop screening criteria for premium wine production in the U.S. The suitability of temperature for premium wine production in a given year is linked to the seasonal heat accumulation, the mean seasonal temperature, the seasonal diurnal temperature range, and the occurrence of heat and cold stress. The screening criteria thus consist of terms measuring 1) the heat accumulation during the growing season (expressed in growing degree days (GDD) above a threshold); 2) the mean growing season temperature; 3) the growing season diurnal temperature range; 4) the ripening season diurnal temperature range; 5) the occurrence of heat stress in the growing season; 6) the occurrence of heat stress in the ripening season; 7) the occurrence of cold stress during the autumn and spring; and 8) the occurrence of cold stress during the winter.

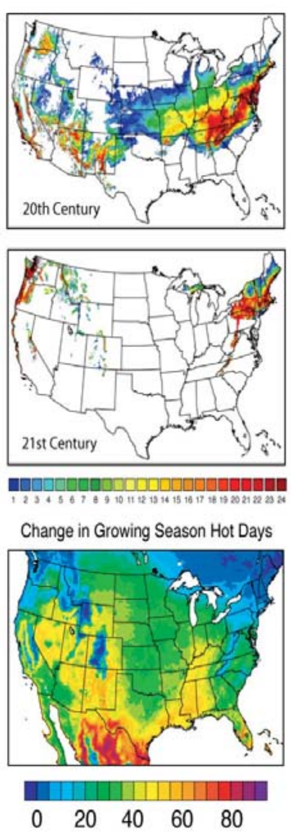

Figure 5. Premium wine suitability, expressed as the number of suitable years (max 24), along with the change in hot days. Source: White et al (2006).

White et al. (2006) developed four categories of temperature suitability, with the tolerance for heat and cold stress determining the four categories (heat-tolerant/cold-tolerant, heat-tolerant/cold-intolerant, heat-intolerant/cold-tolerant, heat-intolerant/cold-intolerant). The long-term suitability of each category is expressed as the fraction of years for which the screening criteria are met at each location (e.g., Figure 5 for the heat-tolerant/cold-tolerant category). Analyses based on the Daymet daily temperature data and the climate model simulations of Diffenbaugh et al. (2005) suggest that global warming could result in substantial reductions in the overall suitability in the United States, including substantial reductions in what are now the highest value areas. However, the analyses also suggest that high quality temperature areas could exist in the Northwest and Northeast, although these areas currently experience challenges associated with humidity and heavy precipitation during the growing season.

Climate change and the incidence of agricultural pests

Relationships between invertebrate physiology and temperature have been well-studied for decades. Diffenbaugh et al. (2008) used temperature-physiology relationships from the literature to develop screening criteria for a suite of pests known to infest corn crops in North America (European corn borer, northern corn rootworm and western corn rootworm), and one cosmopolitan pest known to infest corn and other crops, including soybeans and tomatoes (corn earworm). The range of these pests in a given year is linked to the heat accumulation during the year, along with the severity of cold limitation during the cold season. The screening criteria thus consist of a heat accumulation term (expressed in growing degree days (GDD) above a threshold) and a cold limitation term (expressed in chilling hours below a threshold for corn earworm, northern corn rootworm and western corn rootworm, and annual absolute minimum temperature for European corn borer). In a given year, a location is considered suitable for a pest taxon if that taxon's minimum heat accumulation is exceeded and that taxon's cold limitation threshold is not reached. The long-term prevalence of each pest is expressed as the fraction of years for which the screening criteria are met at each location. Analyses based on the Daymet daily temperature data and the climate model simulations of Diffenbaugh et al. (2005) suggest that global warming could result in expansion of maize pest ranges in the United States, along with the potential for additional pest generations within a given season (e.g., Figure 6 for corn earworm.

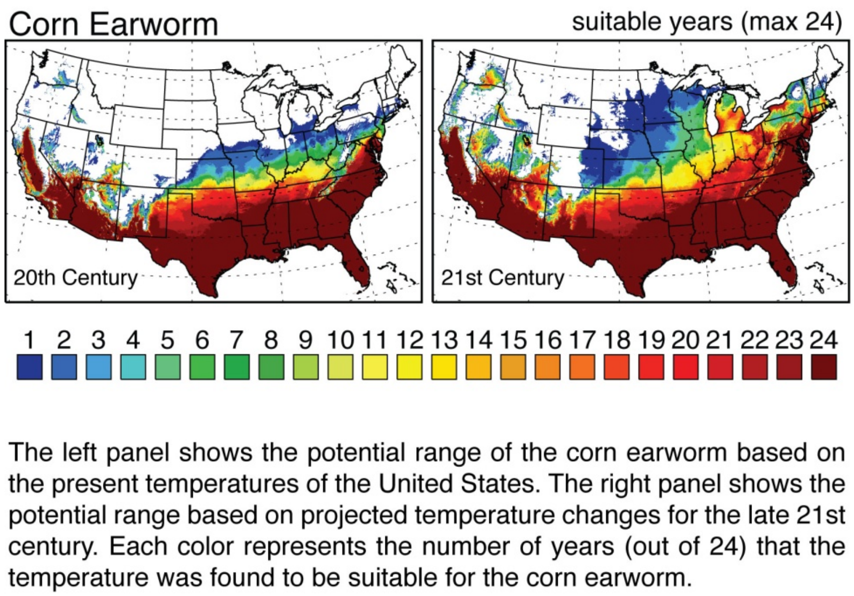

Figure 6. Climate change and the incidence of agricultural pests. Source: Diffenbaugh et al. (2008).

Examples on estimating future production potential

Estimating global yield gaps using maximum observed yields

Recently, a spatially explicit, global data set on crop yields and harvested area has become available (Monfreda, Navin Ramankutty, and Jonathan A. Foley 2008) which has permitted researchers to assess the global yield gap for specific crops, based on a comparison of yields under specific agronomic and climatic conditions. This is in the spirit of the Lobell, Cassman, and Field (Lobell, Cassman, and Field 2009) survey category of studies comparing observed yields to maximum yields by other farmers – albeit comparing grid cell averages across the entire globe. Two papers utilizing this data set to examine the yield gap were available to me at the time of writing this paper. The first, by Licker et al. (2010), establishes a matrix of 100 different global climatic and agronomic conditions based on growing degree days and soil moisture availability. They then order grid cells based on observed yields (from lowest to highest yields) and cumulate harvested area within these "climate zones" until they reach the 90th percentile, the yield of which they use as their maximum potential yield estimate under current climate, technology and economic circumstances. They then compute the yield gap fraction as *(1 – Actual yield/Climatic potential yield)* by crop and grid cell and plot these results (see Figure 7). The pattern of yield gaps varies considerably by crop. Yield gaps for maize are small in North America and Western Europe, but extremely large in Africa, and also quite large in Eastern Europe. For soybeans, yield gaps are low in the US Corn Belt as well as quite low in Brazil. Yield gaps for wheat are small in Western Europe, larger in the US and quite large in Eastern Europe and Russia. The authors also compute the potential for global production increases, based on currently harvested area, provided all grid cells reached their maximum climatic yield potential, as revealed in the current data base. These increases by crop are as follows: maize (50%), rice (40%), soybeans (20%), and wheat (60%), suggesting that, given the right incentives, much of the increased demand for cereals and oilseeds in 2050 could be met using existing technology.[21]

[21] Bruinsma (2009) performs a similar exercise using the FAO/IIASA Global AEZ framework in order to isolate the "transferable" portion of existing yield gaps. Controlling for crop variety and AEZ characteristics, he computes the maximum attainable yield in each country and compares this to average yields for rainfed wheat. With a few exceptions, the differences between attainable and actual average yields are quite substantial in most of the countries examined. In Argentina, for example, average actual yield is 2.6 t/ha vs. 4.6 t/ha attainable yield. In the Ukraine, this gap is even larger (2.5 t/ha vs. 7.1 t/ha attainable).

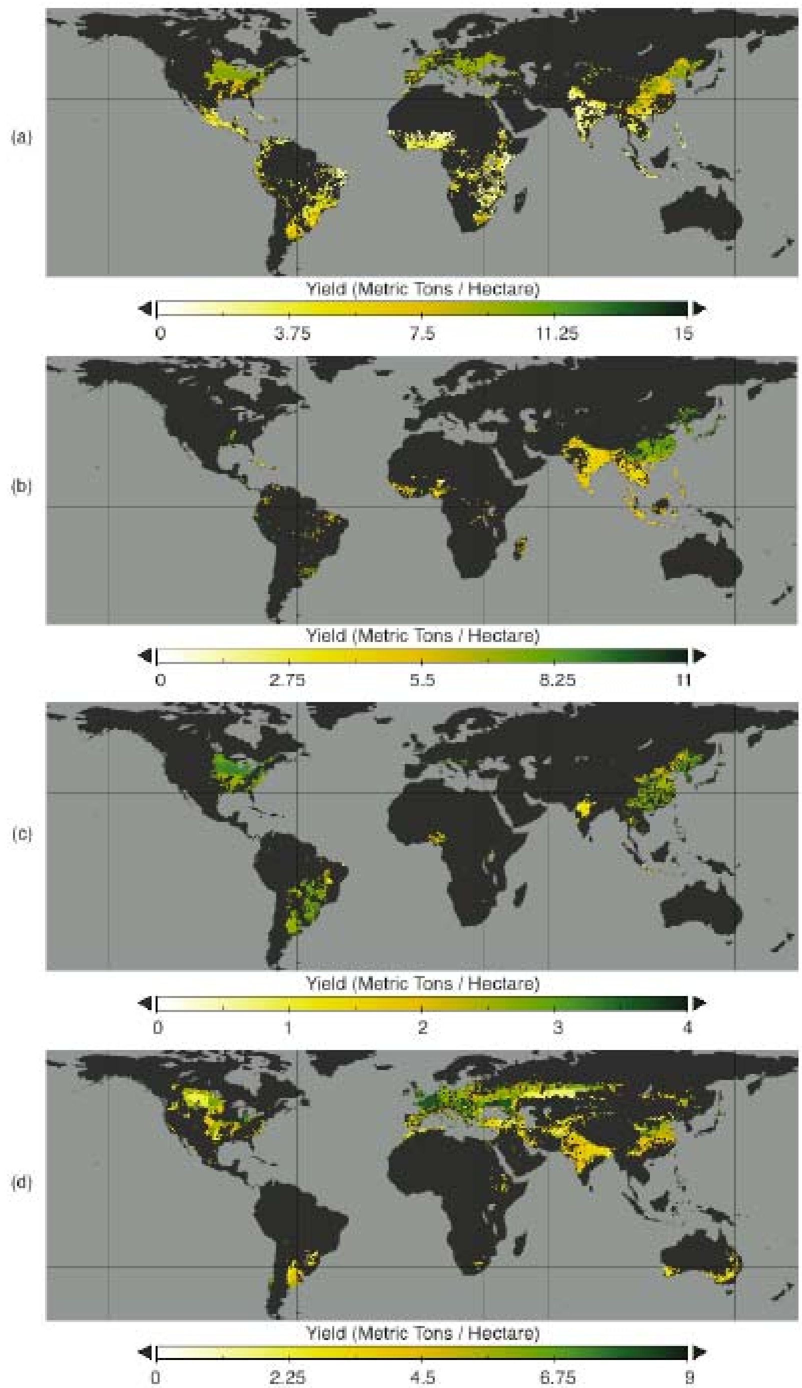

Figure 7. Global patterns of climatic potential yield . The figure shows the 90th percentile yields from each climate zone, across the globe. According to Licker et al. (2010), these are reasonable estimate of the "maximum" yields achievable today – given the modern distribution of yields across the different climate zones of the planet. Source: Licker et al. (2010).

Estimating global yield gaps using a production frontier approach

Given the considerable potential for global yield increases on existing lands, the question arises: What would it take to close this yield gap? This is addressed by the second paper using the Monfreda et al. (2008) data set. Neumann et al. (2010) estimate a frontier production function for global grain production in which the climatic and agronomic variables set the frontier for each grid cell, and a variety of physical and socio-economic variables are used to explain deviations from this frontier. The latter include: irrigation, slope, labor force (population) density, accessibility to markets, and a so-called "market influence" variable. Not surprisingly, their results suggest that the constraining factors (causes of deviations from potential yields) vary considerably by region. Neumann et al.'s (2010). Figure 8 highlights the regions with very low efficiency for wheat, maize and rice, and links these with the predominant constraints in the region. For example, in the case of maize throughout the Guinea-Savanna zone of Africa, they identify market influence and accessibility as key constraints. This is in agreement with recent analysis of that same region by the World Bank (2009) which concludes that poor infrastructure and high transport costs impede the effective demand for land in much of sub-Saharan Africa. Neumann et al. (2010) also find that market influence and irrigation are critical constraints in Central America and South Asian maize production. (The issue of irrigation will be discussed further below.)

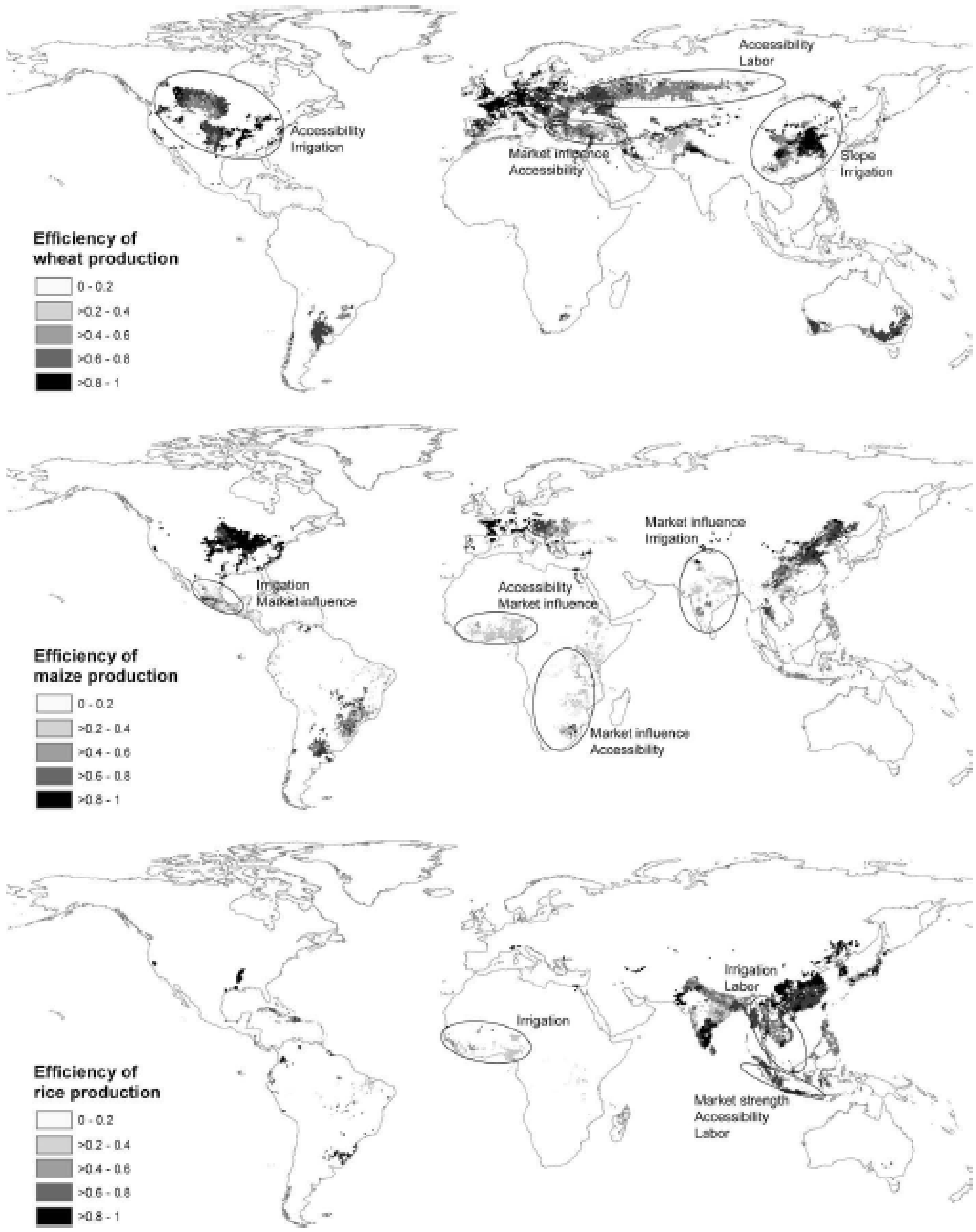

Figure 8. Efficiencies of wheat (Map 1), maize (Map 2) and rice (Map 3) production with the most determining factors per world-region. Source: Neumann et al. (2010).

Estimating the yield potential of biofuel crops

Biofuel crops are considered an important solution to climate change mitigation, and many countries have already mandated biofuels as part of future fuel mixes. However, there is a raging debate surrounding the potential of biofuels. Estimates of future biofuel production potentials have typically assumed certain values for crop yields available from the literature. A recent publication by Johnston et al. (2009) compared the crop yield values used in these studies to actual crop yields realized in the field as derived from the Monfreda et al. (2008) data set. They found that studies of biofuel production potential have typically overestimated crop yields, thereby overstating the fossil-fuel offset potential of biofuels in the future (Figure 9).

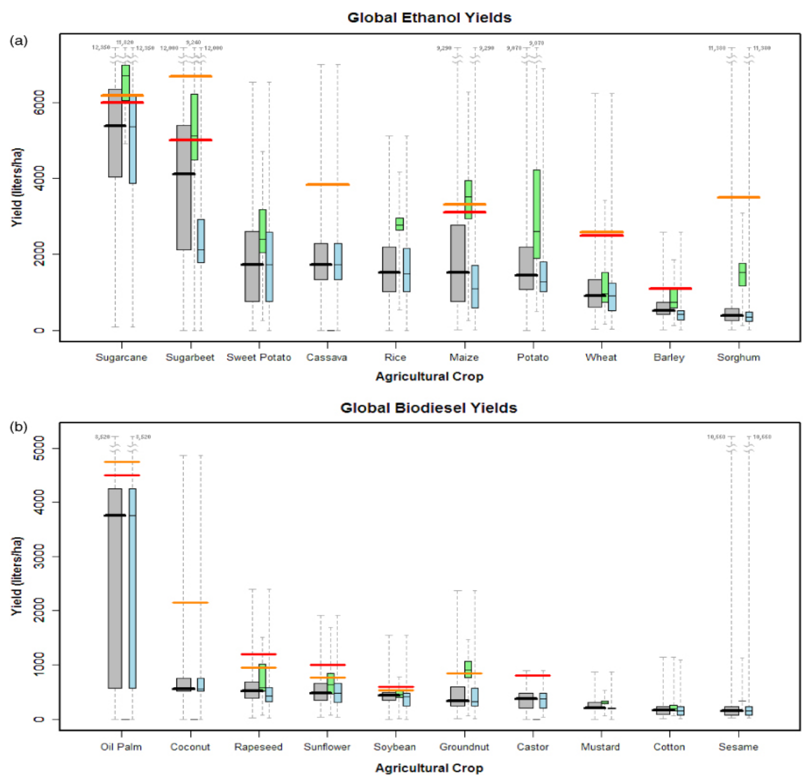

Figure 9. Revised estimates of global biofuel crop yields. (a) Global ethanol yields, (b) global biodiesel yields. (Box plots represent the variation of yields for common biofuel crops. Here we show results averaged for the entire globe (gray), developed countries (green) and developing countries (blue). The horizontal black bars represent median yields, and the boxes are bound vertically by 25th percentile yields on the bottom and 75th percentile yields on the top. The whiskers (in light gray) represent the absolute minimum and maximum yield values recorded in the M3 cropland datasets. The red and orange bars offer comparisons to two previous examples of biofuel feedstock yield estimates, reported by the Worldwatch Institute and Brown (complete citations in the original paper) respectively. Please note, the M3 results for sorghum are compared to Brown's estimate for *sweet* sorghum, which is a different variety with higher sugar content than is normally grown and traded commercially.) Source: Jhonston et al. (2009).

Time series data on land cover and the environmental impacts of land use

For many purposes it is important to have time series data on land cover change. A recent examples of particular interest is concerned with the unprecedented pace at which land cover is changing. Lepers et al. (2005) document these trends and highlight deforestation "hotspots" using a synthesis of available information on land-cover changes at the regional to global scale from 1981 to 2000 based exclusively on existing data sets. As shown in Figure 10, Most of these hotspots are in the tropics, with the Amazon leading the way, followed by Southeast Asia. Central Africa and Central America. Russia, too, shows some deforestation hotspots in their analysis. These authors also conclude that the areas with the greatest amounts of degradation – often from multiple sources – arise in the Middle East and near Asia. The authors note that many parts of the world are not adequately represented in the available data sets, so it is possible that rapid change is occurring in locations that are not identified in the maps. Based on the land cover data they conclude the deforestation mostly takes place at the edge of large forest areas and along major transportation networks (e.g., along the southern Amazon basin), and there are different trajectories of land-cover change in different parts of the world (e.g., decrease in cropland in temperate areas and increase in the tropics).

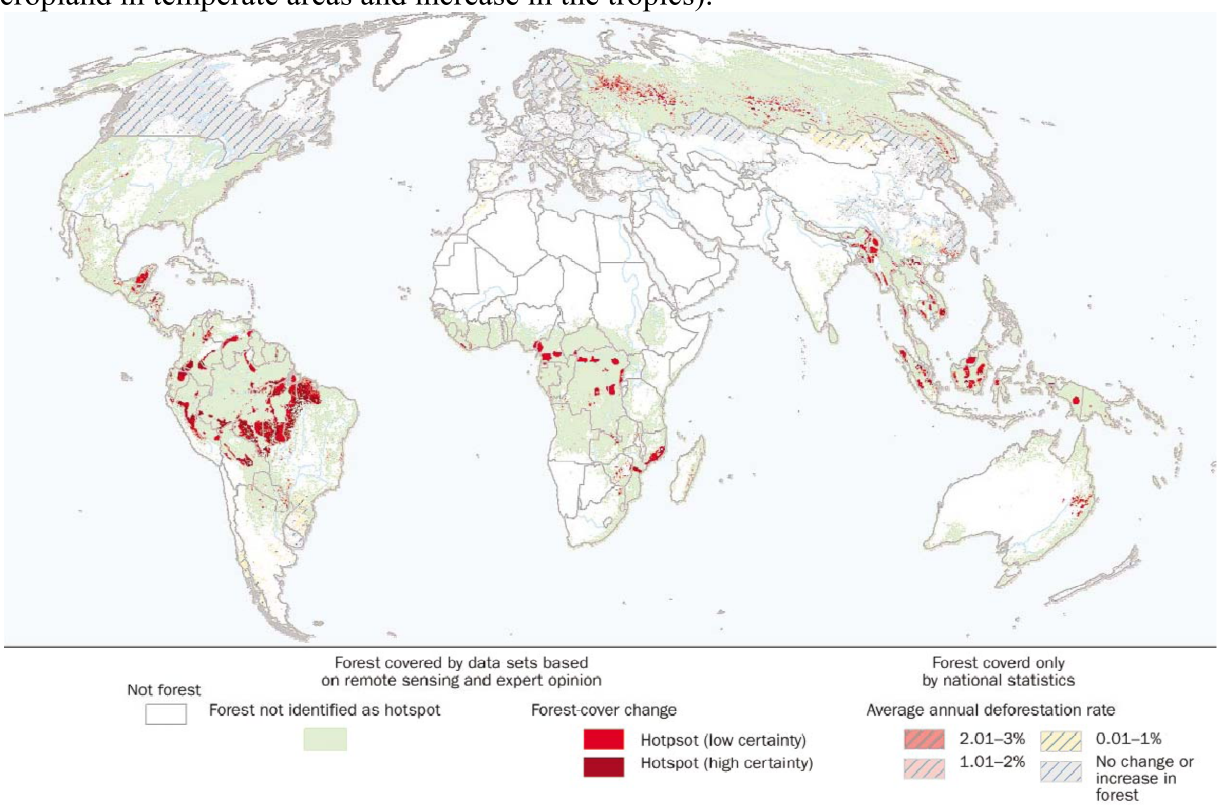

Figure 10. Main areas of forest-cover changes over 20 years (1980-2000). The map identifies, for each "forested" grid cell, whether it was considered as a main area of forest-cover changes by the input data sets. The color code represents the reliability (estimated in terms of convergence of evidence) of the information (i.e., the frequency of detection as a hotspot relative to the number of data sets covering the area).

Resource outlook: land and water

A recent paper undertaken by Fischer (2009) for the 2009 FAO food summit asks: how much could a large scale, global, sustained bioenergy program affect global land use? Under Fischer's "TAR-V1" scenario, the mandatory, voluntary and indicative targets for biofuel use announced by both developed and developing countries are implemented by 2020 – boosting production to twice current levels. Second generation technologies are assumed to become available after 2015, and are only gradually deployed under this scenario. This results in 2020 cereal prices which are 38% above baseline, falling off to 27% above baseline by 2050, when second generation biofuels assume a larger share of the total. The price rise is widespread and crops prices increase by roughly the same amount as do cereals (35% in 2030 and 27% in 2050). This price rise boosts cultivated land globally by 38 Mha in 2020 and 48 mill ha (or about 2.8%) by 2050. The very strong price impact, and quite small increase in land supply, suggests rather inelastic supply elasticities in the long run. This is important, since price increases of this magnitude could have a significant impact on global poverty (Ivanic and Martin 2008).

Along with land, water is perhaps the most precious agricultural input. Fischer et al. (2007) explore the effects of climate change on water available for irrigation by combining crop water requirements (obtained from the FAO/IIASA Agro-ecological Zone model discussed in Section 2), climate change scenarios from the HadCM3 and CSIRO general circulation models, and two alternative projected socioeconomic scenarios (until 2080), one with mitigation and the other without it (see Riahi, Grübler, and Nakicenovic 2007 for fuller details)

. The unmitigated scenario (SRES A2-r) "describes a very heterogeneous world (…) where (…) fertility patterns across regions converge only slowly, which results in continuously increasing global population" moreover, "economic development is primarily regionally oriented and per capita economic growth and technological change are more fragmented and slower than in other [scenarios]". This "scenario assumes the highest levels of urbanization rates and largest income disparities, both within cities (e.g., between affluent districts and destitute 'favelas') and between urban and rural areas." On the other hand, the mitigated scenario (SRES-B2) "describes a world in which the emphasis is on local solutions to economic, social, and environmental sustainability. It is a world with continuously increasing population at a rate lower than in A2 (above), intermediate levels of economic development, and less rapid and more diverse technological change than in the" more optimistic scenarios. In this scenario, technological progress in alternative sources of energy and increasing scarcity of easy-access fossil resources reduces GHG emissions in the long run. The simulations performed by Fischer et al. suggest that the global impacts of climate change on irrigation water can be as large as the changes implied solely by the unmitigated socioeconomic scenario; moreover, mitigation can have large water saving impacts in the presence of climate change; finally, the authors find that the benefits of mitigation vary significantly across regions.

Bruinsma (2009) uses geospatial data to assess the balance between water supply and demand in different regions of the world. He documents the extreme variability of climatic conditions facing the developing countries (average annual precipitation varies from a low 160 mm per year in the arid regions of Near East/North Africa to a high precipitation of about 1530 mm per year in Latin America). Moreover, his data suggests that those countries suffering from low precipitation and therefore most in need of irrigation are also those where water resources are naturally scarce. Bruinsma (2009) estimates irrigation water withdrawals in 2050 assuming a modest increase in the efficiency of water usage. At the global level, the increase in irrigation water withdrawal is expected to grow by a modest 11 percent, from the current 2620 km3/yr to 2906 km3/yr in 2050. For most developing countries, water use in irrigation currently represents a relatively small part of their total water resources and there remains a significant potential for further irrigation development. Given the relatively low expected increase in water needs, this situation is unlikely to change. However, locally and in some in the Near East/North Africa region there are already several water shortages.

Modeling the influence of global land cover change on climate and on the terrestrial carbon cycle

Changes in global land cover can influence regional and global climate through biogeophysical pathways[22], by altering the reflectivity of the land surface (albedo), changing the partitioning of energy at the surface, changing the rooting depth and therefore the amount of water pumped into the atmosphere, and altering surface roughness which changes wind speeds. Numerous climate modeling studies have been undertaken to estimate the global impact of land use change (e.g., T. N. Chase et al. 2000; Bounoua et al. 2002; Matthews et al. 2003; Brovkin et al. 2006; Betts et al. 2007; Findell et al. 2006) and reported mainly that the global effects were minimal, but that there were significant regional impacts. However, in spite of this plethora of modeling studies, there has been no consensus on exactly how, and in which regions, land use change changes would influence climate.

In a recent, important study, Pitman et al. (2009), compared seven different climate model simulations driven by identical changes in global land cover. The land cover change data in the models were derived from Ramankutty and Foley (1999) for changes in cropland, and from Goldewijk (2001) for changes in pastures. The study found that there was little agreement between the models in terms of climate impacts (Figure 11). While five of the models simulated statistically significant cooling in the regions of land-cover change, one simulated warming. In terms of processes, three models simulated increases in evaporative flux, while three simulated decreases. The study suggested that the lack of agreement between the models was a result of the different ways in which the models represented the altered land cover, and in the different ways they represented land surface processes. Thus, they concluded that the IPCC 5th assessment will face a dilemma in wanting to include land-cover changes because they significantly influence regional climates, but it would be unfeasible to impose a common land-cover change simulation across multiple models.

Land cover changes, mainly the clearing of forests for agricultural land, contribute to ~20% of the total emissions of carbon dioxide. This is one of the reasons that REDD (Reducing Emissions from Deforestation and Degradation) has become a popular initiative for climate change mitigation. However, despite enormous progress in estimating deforestation using remote sensing, there is still widespread uncertainty in deforestation rates, and in estimates of carbon emissions (R. DeFries and F. Achard 2002; Houghton 2003).

---

[22] Land cover changes also influence climate indirectly, through their emissions of greenhouse gases (known as the biogeochemical pathway), but these effects are already implicitly included in global climate models in a lumped fashion, where the resultant atmospheric carbon dioxide concentration is influenced by both fossil-fuel and land use emissions.

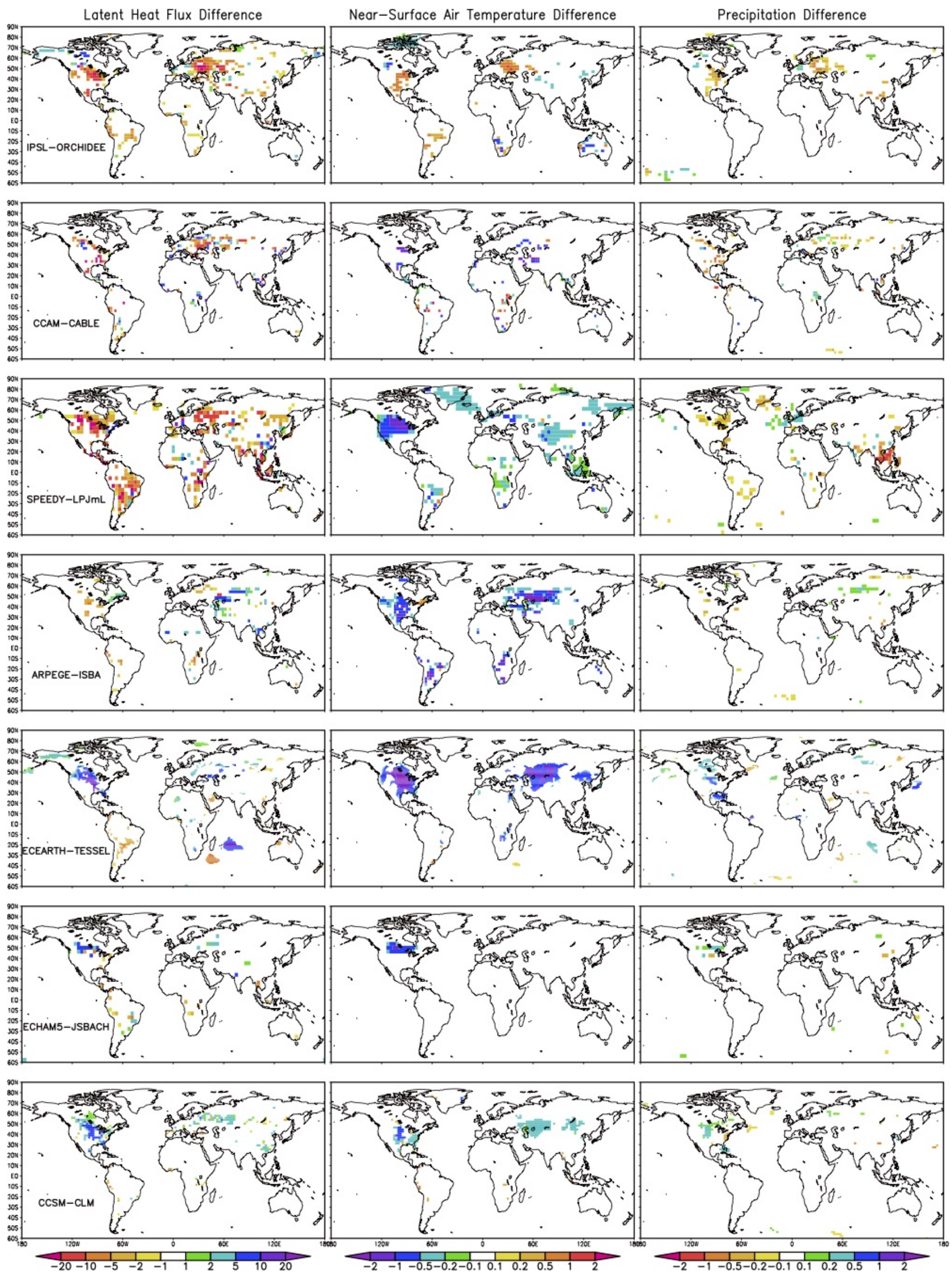

Figure 11. Change in summer evaporative flux (left), surface air temperature (middle) and precipitation (right) due to changes in land cover, as simulated by seven different models (rows). Only areas with statistically significant changes are shown (95% confidence). (Figure 2 from Pitman et al. 2009)

In a recent paper, Ramankutty et al. (2007) argued that we need to make progress in several fronts in order to reduce the uncertainty in carbon emissions, including better estimates of deforestation rates, and in the dynamics of land cover following deforestation (i.e., land transitions). Since the publication of this paper, there has been continuing progress on several of these issues. Recently, Hansen et al. (2010) used MODIS satellite data to estimate global *gross* forest cover loss from 2000-2005. But, Kurz (2010) placed this work in context, suggesting that gross forest cover loss is only one of numerous pieces of the puzzle that is necessary to estimate carbon loss. In a forthcoming study, Gibbs et al (submitted) made some progress in filling some pieces of this puzzle by estimating the land-cover transitions that result from deforestation. Gibbs and colleagues analyzed over 100 Landsat scenes across the tropics to identify the land-cover “sources” of expanding agricultural land. They found that, between 1980 and 2000, more than 55% of new agricultural land expanded at the expense of intact forests, and another 28% replaced degraded forests. In

Figure 12, we show an example of their analysis for Tropical America.

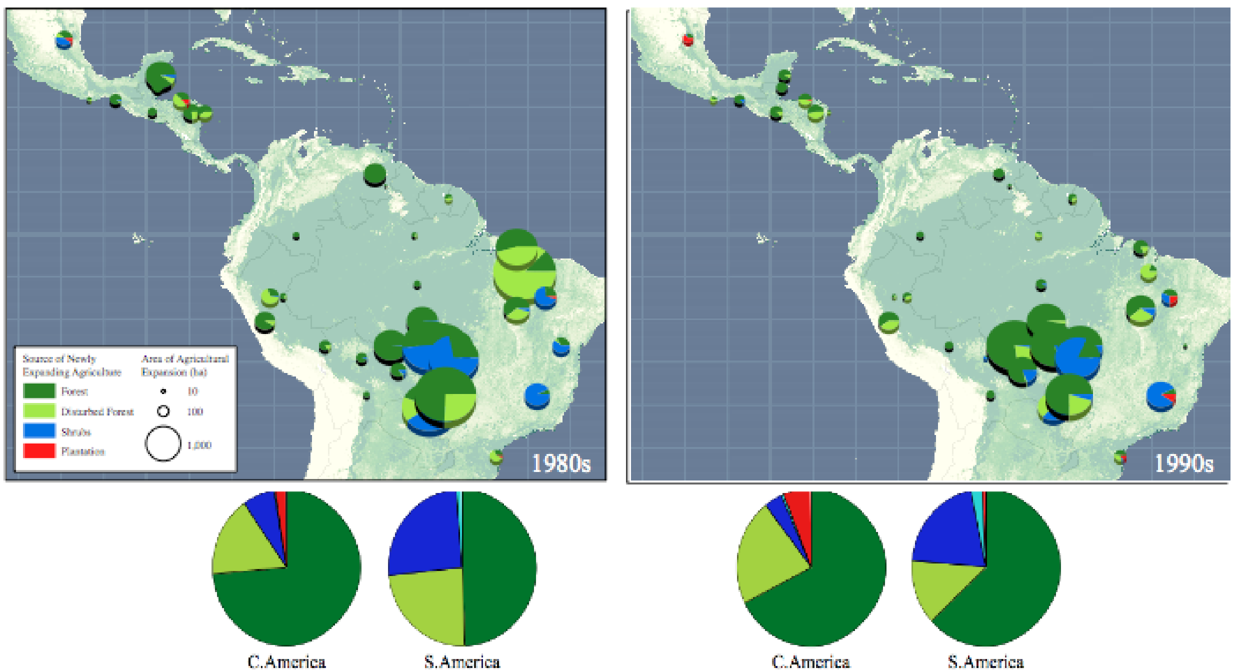

Figure 12. Sources for newly expanding agricultural land in tropical America in the 1980s and 1990s. The pie charts represent the relative proportions across broad regions and for individual Landsat sites, and are scaled according to the size of the agricultural land expansion.

Irrigation, Climate, and Crop Productivity

Impact of irrigation on the South Asian summer monsoon

South Asia is one of the most densely irrigated regions of the world. Irrigation results in increased evapotranspiration and in a cooling of the surface, thus decreased surface sensible heat flux and increased latent heat flux (Boucher et al., 2004; Kueppers et al., 2007; Lobell et al., 2008; Puma and Cook, 2010). Lee et al. (2009) and Saeed et al., (2009) show in recent studies, that this also has a significant impact on the South Asian Summer Monsoon (SASM).

Lee et al. (2009) analysed NDVI time series for the Indian subcontinent for the period 1982-2003 and found that decreased July surface temperature in the Indian subcontinent (an expected result of increased evapotranspiration due to irrigation and increased vegetation) caused a reduced land–sea thermal contrast, which is one of the factors driving the monsoon, and therefore weakens the monsoon circulation. They conclude that a weak early Indian Summer Monsoon appears to be at least partially a result of irrigation and the resultant increased vegetation and crop activity prior to the monsoon.

Saeed et al., (2009) included irrigation into a regional climate model for the Indian subcontinent and found that this consideration of irrigation removed biases of temperature and sea level pressure over the Indus basin resulting in less differential heating between land and sea masses. This in turns reduced the westerlies entering into land from Arabian Sea, hence creating conditions favourable for currents from Bay of Bengal to intrude deep into western India and Pakistan that have been unrealistically suppressed before. They concluded that the representation of irrigated water is unavoidable for realistic simulation of south Asian summer monsoon and its response under global warming (see Figure 13).

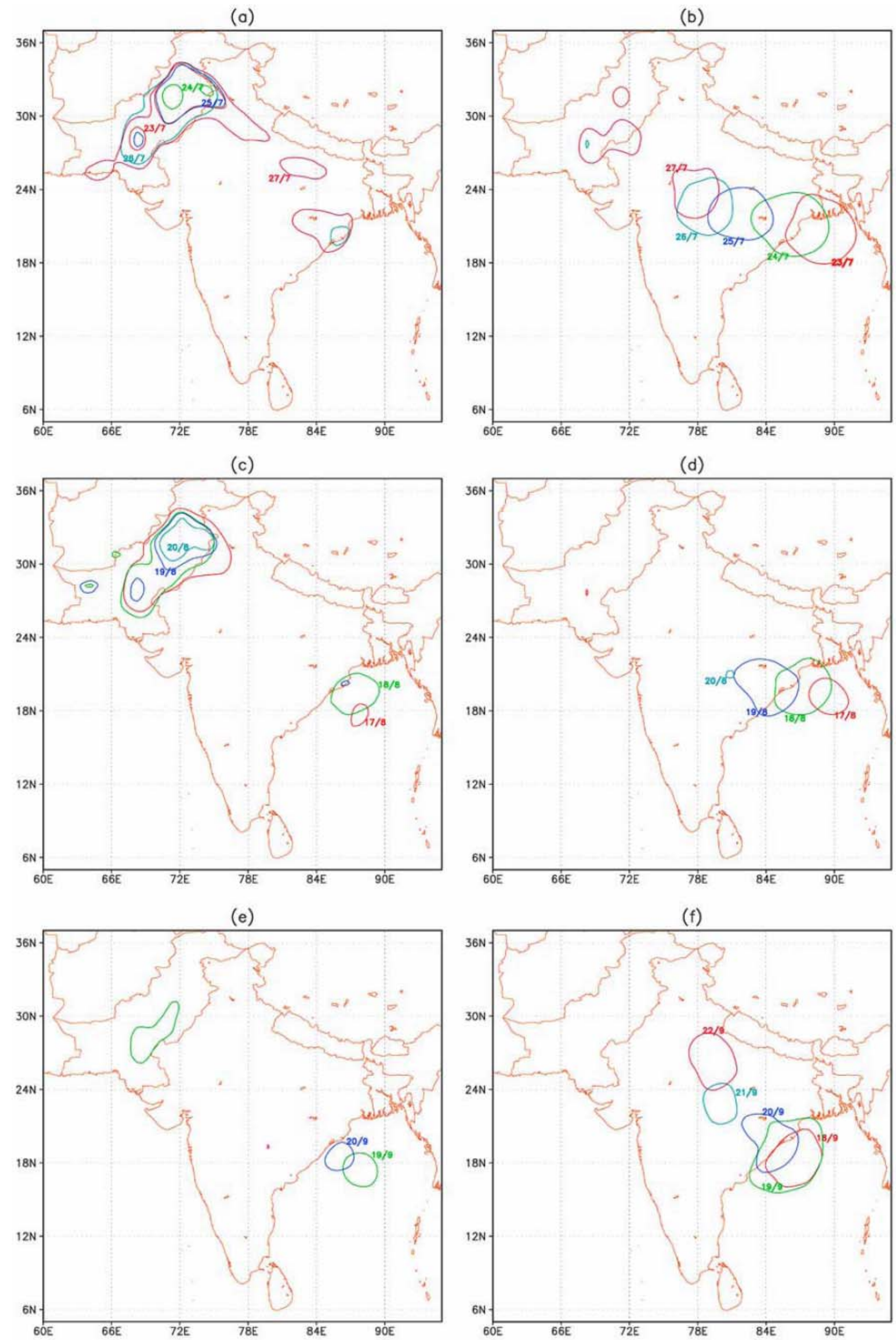

Figure 13. Monsoon depressions/lows development and movement: July 23–27, 1991 at 993 hPa MSLP (a) REMO-baseline, (b) REMO-irrigation; Aug 17–20, 1990 at 996 hPa MSLP (c) REMO-baseline (d) REMO-irrigation; and Sep 18–22, 1991 at 999 hPa MSLP (e) REMO-baseline (f) REMO-irrigation. (Figure 3 from Saeed et al., 2009)

Crop production from irrigated and rainfed land

Irrigation is used to intensify agricultural production. In general, crop yields are higher under irrigated conditions as compared to rainfed conditions because of avoiding or reducing drought stress of cultivated crops. Siebert and Döll (2010) quantified in a recent paper this effect by applying a newly developed crop water model for 26 different crop classes at 5 arc-minute resolution for the period 1998-2002. They found that total crop water use in that period was 6,685 $km^3\ yr^{-1}$, of which blue water use (irrigation water) was 1,180 $km^3\ yr^{-1}$, green water use (precipitation water) of irrigated crops was 919 $km^3\ yr^{-1}$ and green water use of rainfed crops was 4,586 $km^3\ yr^{-1}$. Average crop yield of irrigated cereals was 442 Mg km^{-2} while average yield of rainfed cereals was only 266 Mg km^{-2}. They furthermore summarized that if currently irrigated crops were not irrigated, global production of dates, rice, cotton, citrus and sugar cane would decrease by 60%, 39%, 38%, 32% and 31%, respectively. 43% of cereal production was on irrigated land, and without irrigation, cereal production on irrigated land would decrease by 47%, corresponding to a 20% loss of total cereal production. The largest cereal production losses would occur in Northern Africa (66%) and Southern Asia (45%). These potential production losses are illustrated in Fig. 14 for irrigated cereal production and for total cereal production on irrigated and rainfed land.

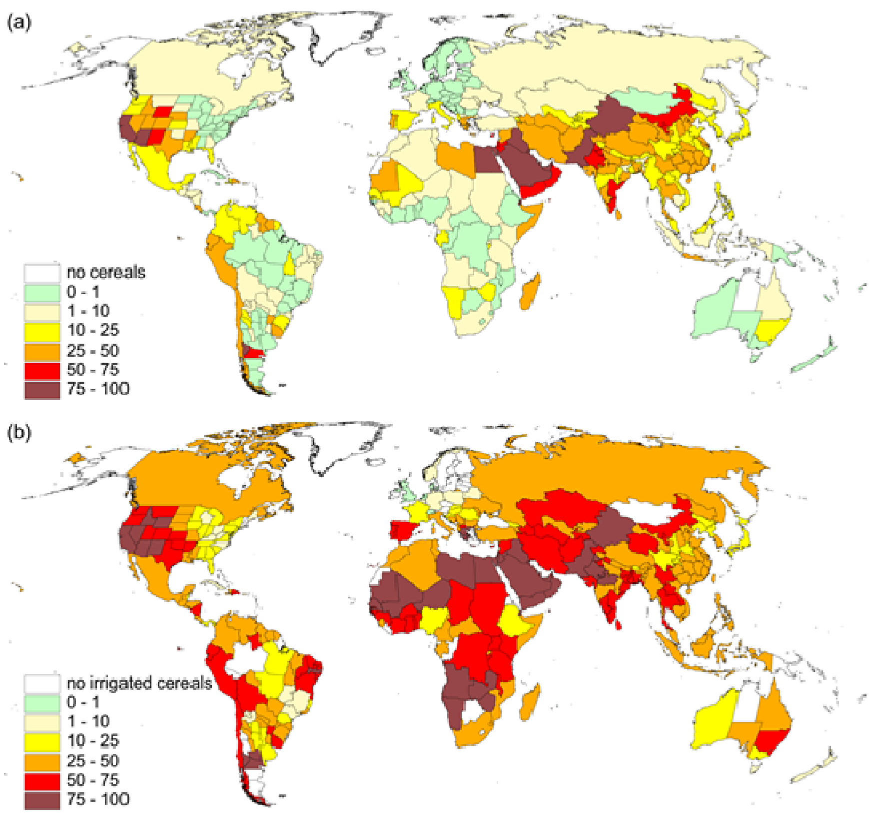

Figure 14. Loss in total cereal production (a) and in irrigated cereals production (b) if no irrigation water were applied compared to actual cereal production computed by the Global Crop Water Model for the period 1998–2002. (Figure 8 from Siebert and Döll, 2010)

Annex II. Proposal for a pilot project to compile consistent global and regional data on agriculture and climate

Agriculture presents a dilemma. On the one hand, increased food production per capita over the last few decades has lifted millions of people out of starvation. On the other hand, agriculture is a major driver of global environmental change. Balancing the tradeoffs between the need to increase agricultural production (for both food and fuel) while minimizing its environmental impacts is a major challenge of the 21st century.

Global analysis of the tradeoffs requires the development of global databases. The FAO provides data on agriculture land cover and use since 1961 at the national level (FAOSTAT database), but such national-level data are inadequate for the detailed spatial analysis necessary to understand environmental degradation and for economic analyses. While high-resolution data are now available from satellite remote sensing, they are thematically weak in their descriptions of agriculture. In summary, current efforts to provide global data operationally are insufficient to address the major challenges related to agriculture.

A few efforts have made some progress toward addressing this challenge, in research mode. By using statistical methods to combine remotely-sensed data, a vast compilation of agricultural inventory data, and other ancillary data products, a few recent studies have developed new geographically-explicit data sets describing agricultural land cover and use in far greater thematic and spatial detail than previously available (Ramankutty et al, 2008; You et al., 2009; Portman et al., 2010). In the AgroMAPS project, IFPRI, FAO, and the University of Wisconsin collaborated to compile global subnational statistics on crop harvested area and production. However, these studies have been one-off efforts, and continued support for operationalizing them does not exist.

This is a proposal for a pilot study at several different nodes around the world, to co-ordinate the compilation of global data on agriculture and climate. The pilot study will serve as proof of concept to test the feasibility of operationalizing such an effort in the long term. The pilot study will occur over a two-year period, with the aim of supplementing the FAO data with subnational level statistics, and combining it with satellite remote sensing information to develop global gridded data products.

The different nodes will release data in two phases: while all the data will be shared internally between the nodes during phase 1 (first two years), only a limited set of data will be released publicly at the end of phase 1 in order to allow the various nodes to conduct additional science with the data and to protect student thesis/dissertations, and also to allow additional quality checks. Then, if there is a phase 2, the groups will inter-compare and harmonize their data sets further, and develop and release other derived products. The different nodes have expertise in different elements of the work – some are focused on compiling the best available subnational statistics on agricultural land, while others are focused on developing the best spatial data set on agricultural land use and climate. This is reflected in the types of data proposed for release at the end of phase 1. However, in phase 2, the nodes will co-ordinate better to release consistent global data sets at both the best available subnational administrative-level as well as in spatially-explicit formats.

A global database of agriculture land cover and use from 1961-2010. Navin Ramankutty, McGill University.

(In collaboration with Jonathan A. Foley, Univ. of Minnesota)

The pilot study at McGill University, in collaboration with colleagues at the University of Minnesota, will contribute to the update and new development of historical data on agricultural land use/cover. Our data, previously known as the SAGE data set, and currently referred to as M3LAND (Montreal-Minnesota-Madison LANd use Database) (Monfreda et al., 2008; Ramankutty and Foley, 1999; Ramankutty et al., 2008), have been used by in a wide variety of Earth system analysis (Betts et al., 2007; Bondeau et al., 2007; Gordon et al., 2005; McGuire et al., 2001; Van Minnen et al., 2009), as well in economic analysis of the climate mitigation options involving land use as well as in estimating the impacts of biofuel mandates (Golub et al., 2009; Taheripour et al., 2010).

Here, we propose to update our most commonly available data sets to the 1961-2010 time period. This will involve updating the data to a more recent year, but will also use a much richer compilation of subnational statistics (enabled by this project), as well as improvements in the methodology of reconstruction. We will further provide historical reconstructions for some of our more recent datasets that are currently only available for the Year 2000.

Table 2. Objectives of the pilot project at McGill University

Variables	Data available before the start of the project	Data to be released at the end of project phase 1*	Data to be released at the end of project phase 2 (if funded)
Cropland area Pasture area	Global 5 arc-minute grids for the 1961-2007 period, developed using a limited set of subnational statistics (Ramankutty and Kimball, in prep.).	Global 5 arc-minute grids for the 1961-2010 period, enriched with a richer collection of subnational statistics.	Census or survey-based subnational statistics for 1961-2010 at one administrative level below the nation (e.g., state, province) for 25-30 major agricultural nations of the world, and the European Union.
Crop harvested area Crop production	Global 5 arc-minute grids for the Year 2000 (Monfreda *et al.*, 2008).	Global 5 arc-minute grids for the 1961-2010 period, developed using a limited set of subnational statistics.	Census or survey-based subnational statistics for 1961-2010 at one administrative level below the nation, for 10-15 major crops, for 25-30 major nations of the world growing each crop and the EU.
Fertilizer application rates	Global 0.5 degree resolution grids for the Year 2000 of N/P/K fertilizer application rates, developed using IFA/FAO national statistics (Potter *et al.*, 2010).	Global 5-arc minute resolution grids for the Year 2000 of N/P/K fertilizer application rates, developed using a richer collection of subnational statistics.	Census or survey-based subnational statistics for Year 2000 at one administrative level below the nation (e.g., state, province) for 25-30 major agricultural nations of the world, and the European Union.
Livestock numbers	Global 3-arc minute grids for the Year 2000 and 2005 (Wint and Robinson, 2007).	Global 5 arc-minute grids for the 1961-2010 period, developed using a limited set of subnational statistics.	Census or survey-based subnational statistics for 1961-2010 at one administrative level below the nation, for ~5-10 major animals for 25-30 major nations of the world stocking each animal.

- Note that there will likely be gaps in the data because censuses/estimates are not always available consistently for all countries for all years. We will try to compile all available data and fill gaps as much as possible.

Resources sought

Postdoctoral Fellow
Undergraduate research assistants (x 2)
Travel (3 trips per year total for PI & postdoc)
Other (office supplies, computer software, publication charges, etc.)

A global irrigation database for the period 1961-2010. Stefan Siebert, University of Bonn.

(In collaboration with the Land and Water Division of FAO, Rome, Italy)

The pilot study at University of Bonn aims to further develop a global database of irrigation of agricultural land for the period 1961-2010 by supplementing census based irrigation statistics provided by FAO (FAOSTAT, AQUASTAT) with subnational statistics and to combine these data to historical irrigation maps and to the agricultural land use data base developed at the same time at McGill University. Spatially-explicit time series of irrigated land use are required in many disciplines because irrigation has multidimensional effects. Crop yields are much larger in irrigated agriculture as compared to rainfed agriculture (Siebert and Döll, 2010) so that it is required to know whether crops are irrigated or not to develop realistic patterns of crop yields and crop production which are again important to downscale socio-economic variables like GDP or to analyze spatio-temporal patterns of food security. Irrigation is also the major water use sector with a global share of 72% in water withdrawals and 92% in consumptive use (Döll, 2009). Therefore it is very important for analyses of water resources and water use to know the extent of irrigated land. Irrigation occurs at the interface between land and atmosphere and increased crop evapotranspiration results in a significant surface cooling effect (Boucher et al., 2004; Kueppers et al., 2007; Lobell et al., 2008; Puma and Cook, 2010) and modifications of the South Asia Summer Monsoon circulation (Lee, 2009; Saeed, 2009) which again effects glacier extent and water resources in highly populated basins like the Ganges, Brahmaputra, Indus, Yangtze and Yellow river basins (Immerzeel et al., 2010). Therefore patterns of irrigation water use are requested as input to an increasing number of climate models.

The Global Map of Irrigation Areas (Siebert et al., 2005 updated by Siebert et al., 2006) was used in the studies mentioned before to define the extent of irrigated land. This data set has been developed in a cooperation of the node partners at University of Bonn with the Land and Water Division of FAO and shows the area equipped for irrigation at 5 arc minute resolution around year 2000 (http://www.fao.org/nr/water/aquastat/irrigationmap/index.stm). Currently a project is running to update this data set and to compile subnational statistics on the water source for irrigation, area actually irrigated and irrigation water use around year 2005 (http://www.lap.uni-bonn.de/forschung/global-irrigation-map/detailed-project-information). Furthermore the PI at University of Bonn has been involved in the development of time series of area equipped for irrigation by country (Freydank and Siebert, 2008) and in the development of a data set showing the extent of irrigated and rainfed crops around year 2000 (Portmann et al., 2010).

During the two year pilot study the data mentioned before will be combined to develop time series of subnational irrigation statistics distinguishing area equipped for irrigation, area actually irrigated and harvested area of irrigated crops for the period 1961-2010 (Table 3). These terms have a different meaning but are nevertheless mixed in census-based statistics which often simply refer to “irrigated land”. Exchange with the node partners having a regional focus (IFPRI, CIAT) will help to improve the data base for Latin America and Africa where availability of irrigation data is weaker as compared to countries in Asia, Europe or Northern America. A new methodology will be developed to define the historical extent of irrigated land within the subnational units contained in the data base by using digitized historical irrigation maps. Another focus will be on the development of methods to combine the agricultural land use database developed by the node partners at McGill University with the irrigation database developed at University of Bonn University to subdivide agricultural land use consistently into irrigated and rainfed cropland and pasture. The methods will be demonstrated for a limited number of countries. In a second phase of this project (if funded) the methods developed in the first phase will be applied at the global scale by using the data collected in the first phase (Table 3). This will require intensive error checking and a harmonization of the different data. Emphasis will be on the development of operational tools to enable continuous updates of the data base. Products developed in this project will be released as harmonized global open access database.

Table 3. Objectives of the pilot project at the University of Bonn

Variables	Data available before the start of the project	Data to be released at the end of project phase 1*	Data to be released at the end of project phase 2 (if funded)
Area equipped for irrigation (AEI)	Census-based statistics for 26 909 subnational administrative units for year 2000, global grid at 5 arc-minute resolution for year 2000 (Siebert et al., 2006); time series of AEI per country for the period 1900-2003 (Freydank and Siebert, 2008)	Global 5 arc-minute grids of AEI for the period 1961-2010, not harmonized with McGill cropland and pasture extent	Global 5 arc-minute grids of AEI for the period 1961-2010, harmonized with McGill cropland and pasture extent
Area actually irrigated (AAI)	Census-based statistics for more than 5000 subnational administrative units, around year 2000 (Siebert et al., in press)	Time series 1961-2010 of area actually irrigated per country	Global 5 arc-minute grids of AAI for the period 1961-2010, harmonized with McGill cropland and pasture extent
Irrigated and rainfed harvested area of major crops	Global 5 arc-minute grids for 26 crop classes, around year 2000 (Portmann et al., 2010)	Improved and updated database of harvested area of irrigated and rainfed crops per country, for major irrigating countries 1 administrative level below the nation	Global 5 arc-minute grids of harvested area of irrigated and rainfed crops for the period around year 2005, harmonized with McGill harvested crop area
Irrigated and rainfed production of major crops	Global 5 arc-minute grids for 26 crop classes, around year 2000, model output of the Global Crop Water Model GCWM (Siebert and Döll, 2010)	Improved and updated database of irrigated and rainfed crop production per country, for major irrigating countries 1 administrative level below the nation	Global 5 arc-minute grids of harvested area of irrigated and rainfed crops for the period around year 2005, harmonized with McGill crop production, model output based on an improved version of GCWM

- Note that there will likely be gaps in the data because censuses/estimates are not always available consistently for all countries for all years. We will try to compile all available data and fill gaps as much as possible.

Resources sought

PhD student
Undergraduate research assistant
Travel (3 trips per year total for PI & PhD student)
Other (office supplies, data access licenses, publication charges, etc.)

A global climate database for the 20th and 21th Century. Noah Diffenbaugh, Stanford University.

Climate model experiments help to understand the vulnerability of agricultural systems to climate change as well as the potential effectiveness of different agricultural adaptation options. Coordinated climate model experiments form the primary resource for researchers who are not climate modeling experts, and such coordinated experiments have formed the backbone of recent assessments of the Intergovernmental Panel on Climate Change (IPCC). At present, the results of these coordinated climate model experiments are stored in centralized archives. The availability to the broad international community has literally revolutionized the science of climate change, enabling thousands of researchers from disciplines ranging from fluid dynamics to economics to directly access a large suit of climate model experiments generated by climate modeling centers around the world.

However, these archives comprise massive data volumes from numerous experiments using numerous different climate models employing a variety of different geophysical grids. Access to and refinement of individual variables of interest therefore requires expertise in data acquisition, subsetting, and interpolation. Therefore, "post-processing" of the primary climate model output into uniform formats, and archival of those post-processed data in a central database alongside other agriculturally relevant data, would be a great benefit to the large community of researchers whose expertise lie outside of the climate modeling domain. Thus, a first contribution of this pilot project is to extract the most relevant climate fields from the latest framework for coordinated climate change experiments (CIMP5) in formats readily usable by the research communities targeted by the Database.

A second contribution of this pilot study is to increase the spatial and temporal resolution of the climate model outputs. In general, climate models provide the advantage of full spatial coverage over the geographic domain that is being simulated. The most widely used climate models are general circulation models (GCMs). These models include component models of the global atmosphere and ocean (including the land surface and sea ice), and provide output for all land and ocean surface of the globe. These global-scale models enable numerical experiments that test the response of the climate system to changes in forcing, including both human- and non-human-caused changes. GCM experiments are now typically run for centuries to millennia, including multiple realizations to constrain internal climate system variability and multiple forcing scenarios to constrain the response to different levels and pathways of climate system forcing. Thanks largely to the Coupled Model Intercomparison Project (CMIP), large suites of GCM experiments generated at modeling centers around the world are now available to the international community (http://cmip-pcmdi.llnl.gov/). The third generation of the intercomparison project (CMIP3) provided the backbone for the recent Fourth Assessment Report of the IPCC. (See details of the CMIP3 and forthcoming CMIP5 experiments and available data below.)

In the time dimension, climate models have very high resolution. However, the output are often only saved at course time resolution (usually monthly scale), in part because of the data storage requirements associated with saving daily and sub-daily fields for long experiments, and because of a perceived lack of demand for higher time resolution fields. However, in large part from communication with communities working on the impacts of climate variability and change on agriculture and ecosystems, there is an increasing effort to archive high time resolution output, to the point that climate model output are becoming available with a greater combination of spatial coverage and time resolution than the observational datasets.

However, in contrast to both spatial coverage and temporal resolution, climate models are limited in their spatial resolution. This limitation is due in large part to computational limitations. As a result, the "standard" global climate model output of ~200 km resolution is course compared to the heterogeneity of many variables that are important for agriculture (e.g., temperature, precipitation, soil properties, land fragmentation). High-resolution nested climate models offer the capability to simulate limited areas at higher spatial resolutions. The current state-of-the-science provides resolutions of 20-50 km over continental-scale areas for multi-decadal simulations. A number of coordinated experiments have been conducted over specific regions such as Europe (PRUDENCE http://prudence.dmi.dk/main.html and ENSEMBLES http://ensembles-eu.metoffice.com/) and North America (NARCCAP http://www.narccap.ucar.edu/). Currently, a new coordinated experiment is being launched to cover all inhabited regions of the globe at high spatial resolution with multiple nested climate models (CORDEX http://copes.ipsl.jussieu.fr/RCD_Projects/CORDEX/CORDEX.html). Integration of these high-resolution climate model experiments will provide climate change information at finer spatial scales than are available from the more widely used global climate models, albeit over limited geographic regions. In addition, there are efforts under way to make the coarse resolution GCM experiments available at time and space resolutions – and in data acquisitiong formats – that are more compatible with the stated needs of the agricultural research community. One example is the project by Ed Maurer and colleagues, which includes an effort to correct the biases in the CMIP3 GCM fields (http://gdo-dcp.ucllnl.org/downscaled_cmip3_projections/dcpInterface.html).

Against this background, the pilot project at Stanford University has three main objectives: (1) to extract the most relevant climate fields from the latest framework for coordinated climate change experiments (CIMP5) in formats readily usable by the research communities targeted by the Database, (2) to make available climate model outputs with higher the spatial and temporal resolution, and (3) to provide a bias-corrected set of climate model variables.

Table 4. Objectives of the pilot project at Stanford University

Variables	Data available before the start of the project	Data to be released at the end of project phase 1	Data to be released at the end of project phase 2
Simulated monthly climate fields (temperature, precipitation, surface winds, incoming solar radiation)	20^{th} and 21^{st} century CMIP3 GCM output	20^{th} and 21^{st} century CMIP5 GCM output	Possibility of 20^{th} and 21^{st} century high-resolution nested climate model output
Simulated daily temperature and precipitation fields		20^{th} and 21^{st} century CMIP5 GCM output	Possibility of 20^{th} and 21^{st} century high-resolution nested climate model output
Bias-corrected temperature and precipitation fields		Bias-corrected monthly CMIP3 GCM output	Bias-corrected monthly and daily CMIP5 GCM output; possibility of bias-corrected monthly and daily high-resolution nested climate model output

Resources sought

Postdoctoral Fellow
Travel (2 trips per year)

Land Use in Latin America and the Caribbean (LAC), 1990 – 2010. Glenn Hyman, International Center for Tropical Agriculture (CIAT).

Growing population and international trade has driven vast agricultural land-use changes in Latin America and the Caribbean over the last two decades. The growth of the beef and soy industries are two prominent examples of recent land use changes related to international food and feed demand, especially from Asia. The area planted to soybeans has increased two and half times over the last two decades. Pasture area doubled in Brazil and grew by 25% in the rest of Latin America between 1980 and 2005 (FAO, 2009). These changes are expected to intensify. Well watered and with available land compared to other world regions, Latin America may have the highest potential to respond to global food demand from population growth over the next half-century. According to FAO (2002), the region uses less than one third of its potentially arable land, less than any other world region.

Agricultural land use changes impact on climate change, soil erosion, biodiversity loss and water resources degradation, all serious problems in Latin America. Mitigating climate change through avoided deforestation must focus on Latin America. Over 60 percent of the world's tropical deforestation between 2000 and 2005 occurred in the region (Hansen et al., 2008). Adaptation to climate change will require a new orientation to agricultural research, one that requires data on past and future land-use change to understand effects on agricultural technology (Jones et al., 2007). Soil erosion threatens the sloping lands of the Andes and Central America as well as prime agricultural land in the Argentina and southern Brazil (Harden and Hyman, 2007). Threats to ecosystems and biodiversity are expected to increase in the region (Jarvis et al., 2009). Addressing these environmental problems will require a better understanding agricultural land-use and land use change.

Despite the clear need for agricultural land use assessment of Latin America, the capacity to conduct research suffers from a lack of consolidated data sets across the entire region and the lack of data comparability between countries. These problems are similar to those found in other developing regions such as Africa and Asia, and even in the developed regions of North America and Europe.

The International Center for Tropical Agriculture proposes a pilot study of agricultural land-use in Latin America and the Caribbean to be carried out as part of a larger global effort. The study would build on CIAT's previous research on crop mapping and its' participation in the AgroMaps initiative by acquiring major crop data sets that could be added to the previous collection (Hyman, 1999; Table 1). The effort should focus on the major crops grown in the region, especially those commodities with high international demand. It should pay special attention to the role of livestock and pastures, including information on cultivated pastures where available. Livestock population numbers are available in most agricultural censuses and surveys. However, the area in pastures is less commonly reported. When reported, it rarely distinguishes between natural and cultivated pastures. But both natural and improved pasture area information is needed for agricultural land-use change analyses, particularly related to analyses of deforestation and carbon stocks (e.g. White et al., 1999). Some improved pastures in the South American savannas sequester considerably more carbon than natural pastures (Fisher et al., 1994). The proposed study should aim to consolidate data at the second administrative level below the nation (equivalent to US counties). In the first phase the project should improve on progress already made in crop mapping. The pilot phase should provide consistent data on livestock populations across the regions. It should inventory and collect all available information on pasture areas, including improved pastures. The Latin American study would support the larger global study in developing gridded data products. A future phase of the project could focus on estimating pasture area based on livestock population (as was done in, for example, Barona et al., 2009)

Table 5. Objectives of the pilot project at CIAT

Variables	Data available before the start of the project	Data to be released at the end of project phase 1*	Data to be released at the end of project phase 2 (if funded)
Harvested area of all major crops for which data is available, with the aim of collecting more than 95% of the cultivated area, from census data and government estimates.	Data for at least one date at subnational resolutions, since 1990: Costa Rica, El Salvador, Mexico, Bolivia, Brazil, Chile, Colombia, Ecuador, Paraguay, Peru and Uruguay.	Existing data but not online or widely published: Belize, Honduras, Guatemala, Nicaragua, Argentina, Guyana, French Guiana, Surinam, Venezuela, Cuba, Haiti, Trinidad and Tobago. Online data that needs to be linked to maps and processed: Colombia, Ecuador, Jamaica, Puerto Rico, Dominican Republic,	Data described for phase 1 converted to raster grids by combining survey and census data with remote sensing.
Production of all major crops to match harvested area data.	Same as harvested area above.	Same as harvested area above.	Data described for phase 1 converted to raster grids by combining survey and census data with remote sensing.
Population of the 5 to 10 most important livestock species, and their animal unit equivalents.	Livestock population numbers generally come from the same sources mentioned for production and area.	We would expect to be able to release livestock data at sub-national resolutions for all countries except small Caribbean islands.	Data described for phase 1 converted to raster grids by combining survey and census data with remote sensing.
Natural pasture and rangeland area	Only a handful of countries produce this data	We expect to release more than a handful of countries after an extensive data search in phase 1.	Estimates of pasture and rangeland area derived from livestock population numbers.
Improved and cultivated pasture area	Only a handful of countries produce this data; Brasil reports improved pastures for a few dates.	We expect that more than a few countries will have some information on cultivated pastures.	Report on state of knowledge of the magnitude and distribution of improved pastures.

Table 6. Sources by Country

Country	Website	Agriculture or crop censuses		CONTACTS
		Date	STATUS	
BELICE	www.statisticsbelize.org.bz	2000	Exists but not yet acquired	info@statisticsbelize.org.bz
COSTA RICA	www.inec.go.cr/#	1984	Acquired	informacion@inec.go.cr
EL SALVADOR	www.minec.gob.sv	1971	Acquired	carzel@minec.gob.sv
GUATEMALA	www.ine.gob.gt	2002	Exists but not yet acquired	info-ine@concyt.gob.gt
HONDURAS	www.ine-hn.org	1993	Exists but not yet acquired	gti@ine-hn.org
MÉXICO	www.inegi.gob.mx	1991	Acquired	
NICARAGUA	www.inec.gob.ni	2001	Exists but not yet acquired	webmaster@inec.gob.ni
PANAMÁ	www.contraloria.gob.pa	2001	Acquired	cgrdec@contraloria.gob.pa
ARGENTINA	www.indec.mecon.gov.ar	2002	Exists but not yet acquired	ventas@indec.mecon.gov.ar
BOLIVIA	www.ine.gov.bo	1999	Exists but not yet acquired	ceninf@ine.gov.bo
	CIRCA	2005	Acquired	CIFOR
Country	**Website**	**Agriculture or crop censuses**		**CONTACTS**
		Date	**Status**	
BRASIL	www.ibge.gov.br	1996-2006	Acquired	ibge@ibge.gov.br
CHILE	www.ine.cl	1997	Acquired	alejandro.sabag@ine.cl
COLOMBIA	www.dane.gov.co	2000	Acquired	dane@dane.gov.co
	http://sigotn.igac.gov.co/sigotn/	2006,2007, 2008	Online	
ECUADOR	http://www.inec.gov.ec/web/guest/ecu_est/est_agr/cen_agr	2000	Online	
	www.inec.gov.ec	2003	Acquired	inec1@ecnet.ec
GUYANA	http://www.statisticsguyana.gov.gy/	1978	Exists but not yet acquired	
GUYANA FRAN.	www.recensement.insee.fr/	2000	Exists but not yet acquired	
PARAGUAY	www.dgeec.gov.py	1991	Acquired	info@dgeec.gov.py
PERU	www.inei.gob.pe	1994	Acquired	infoinei@inei.gob.pe
	INEI	2008	Acquired	CIFOR
SURINAM	http://www.statistics-	1981	Exists but not yet acquired	

	suriname.org/			
URUGUAY	www.ine.gub.uy	2000	Acquired	difusion@ine.gub.uy
VENEZUELA	www.ine.gov.ve	1998	Exists but not yet acquired	cen2000@platino.gov.ve
CUBA	www.cubagob.cu	?	Exists but not yet acquired	webmaster@one.gov.cu
HAITÍ	http://www.ihsi.ht/	1971,1982	Exists but not yet acquired	
JAMAICA	http://statinja.gov.jm/	1996,2007	Exists but not yet acquired	info@statinja.com
PUERTO RICO	http://www.censo.gobierno.pr/	1997, 2002	Online	
REP. DOMINICANA	www.one.gov.do	2000	Online	ofic.estadis@codetel.net.co
TRINIDAD/TOBAGO	http://www.cso.gov.tt/cso/statistics/agriculture.aspx	2004	Exists but not yet acquired	

Resources sought

Research Fellow (33% time commitment)
research assistants (x 2, 33% time)
Travel (to Southern Cone and Caribbean for Research Fellow in 1st year; to USA/Canada and Final meeting site for PI in first and second year)
Other (office supplies, computer software, publication charges, etc.)

Land Use in Asia for the period 1960-2010. Andrew Nelson, International Rice Research Institute (IRRI).

Rice is the dominant crop in Asia, whether it is in monoculture systems like the intensively farmed delta areas of Vietnam, Thailand and Myanmar, the vast rice/wheat belts of the Indo Gangetic Plain and the Yangtze River valley, or the many other systems that can be found across Asia such as rice/rye, rice/pulses, rice/groundnut and rice/aquaculture. In short mapping rice areas - an area in which IRRI has a long history – by necessity means mapping the vast majority of agricultural lands in Asia. Over 90% of the world's rice is produced in Asia, rice systems sustain the livelihood of more than 1 billion people, many of them in Asia, and rice is a key sources of calories for over 3.5 billion people. However rice cultivation is also a water intensive system and one of the largest sources of methane emissions. Developing and maintaining a spatially detailed database of rice farming systems would feed directly into a global database of agricultural systems, providing scientific information on food security, land use change, water scarcity and the environmental impact of agriculture.

The pilot study at the International Rice Research Institute aims to build upon pioneering work at IRRI in the 1980's and 1990's on the dominant rice and rice/wheat systems of Asia (Huke and Huke, 1982; 1988; 1997) to develop an Asia wide database agricultural land for the period 1960-2010. It would also draw upon more recent remote sensing data such as IRRIs MODIS derived map of rice systems in South Asia for years 2000 and 2009 (Gumma et al, submitted).

IRRI's social science division has long collected rice area, yield and production statistics at subnational level from all countries in Asia from Pakistan in the west to Japan in east. Information is available at the 2nd level administrative level for all countries and at the 3rd level for many others. With data going back to the 1940's in some countries, including information on irrigated and rainfed areas, seasonal yields and varieties, this is a valuable resource which needs to be made more widely available. IRRI has offices in thirteen countries across Asia and is well placed to expand efforts in the collection of subantional data to other important crops in the region.

The first phase of this proposed study would aim to consolidate the historical rice system information in order to provide consistent time series of crop areas and production figures. It would also expand the MODIS based mapping of rice systems to South East and East Asia. The second phase would develop agricultural systems maps based on MODIS type data such as the number of cropping seasons, length and startdate of season for years 2000 to 2015 and aim to fuse this with subnational crop area estimates.

Paragraph on previous work

IRRI has worked extensively on mapping rice systems in Asia from pioneering work carried out by Huke and Huke in the 1980s and 1990s (Huke and Huke, 1982; 1988; 1997) that mapped out irrigated, rainfed and upland rice systems, the human geography of rice and agroclimatic regions across Asia. More recent activities have included MODIS derived rice systems for years 2000 and 2009 for South Asia (Gumma et al, submitted).

Andrew Nelson is a GIS expert with a Ph.D. in geography who specialises in the development and coordination of spatial research projects for natural resource management at continental and global scales. During the last 15 years he has worked in several countries for international, multi-disciplinary research institutes such as the European Commission's Joint Research Centre, the World Bank, CIAT, UNEP and FAO. His research covers a broad range of development and environmental topics that require innovative spatial solutions, including; integrated modelling of environmental and socio-economic issues, environmental alert systems, access to markets, population modelling, poverty mapping and protected area monitoring. He has contributed to 50 publications. His recent work has featured in Science, New Scientist and The Economist.

He has worked on the generation of several global and regional datasets, including:

- the UNEP population databases for Latin America and Caribbean (UNEP, 2000) and Africa (UNEP, 2004),
- the Global Rural Urban Mapping Project with CIESIN, CIAT and IFPRI,
- the CSI-SRTM, a void filled version of the NASA SRTM global digital elevation model and
- the first global map of accessibility or travel time to markets (Nelson, 2008)
- gRoads, a consortium for developing a spatial global roads database

Table 7. Objectives of the pilot project at IRRI

Variables	Data available before the start of the project	Data to be released at the end of project phase 1*	Data to be released at the end of project phase 2 (if funded)
Harvested area for major crops in the region including rice, wheat, maize, beans, pulses, oil palm	Subnational annual data going back to the 1940s for some countries for rice areas	Subnational data for all major crops in Asia for 1980-2000	Data coverage expanded to 1960-2010
Production and yield information for major crops in the region including rice, wheat, maize, beans, pulses, oil palm under irrigated and rainfed conditions	Subnational annual data going back to the 1940s for some countries for rice areas	Subnational annual data for all major crops in Asia for 1980-2000	Data coverage expanded to 1960-2010 and trends in irrigation mapped out
MODIS based maps of agricultural systems in Asia	500m resolution map of rice systems in South Asia including rice/wheat and rice/pulse areas for years 2000 and 2009 and irrigated and rainfed areas	500m resolution maps for South, SE and East Asia for the major agricultural systems for year 2000 and 2010	Cropping season, and key phonological information for the major agricultural areas at a resolution to be determined (>500m and <5000m)

- Note that there will likely be gaps in the data because censuses/estimates are not always available consistently for all countries for all years. We will try to compile all available data and fill gaps as much as possible.

Resources sought

Scientist (25% time commitment)
Post doc fellow (100% time commitment)
Travel (to South, SE and E Asia for PDF in 1st year; to USA/Canada and Final meeting site for PI in first and second year)
Other (office supplies, computer software, publication charges, etc.)

Land Use in Sub-Saharan Africa, 2000-2010, Stanley Wood, International Food Policy Research Institute (IFPRI).

(In collaboration with the CGIAR Consortium for Spatial Information)

This node proposal focuses on the contribution of agricultural production data collection, harmonization, development, validation and sharing for a single region, sub-Saharan Africa (SSA). Focus on a single region supports both broader and deeper treatment of a coherent sub-set of countries compared to global efforts. With well-planned and conscientious harmonization, however, regional products can enhance the reliability of global datasets, and promote nested regional and sub-regional analysis within a compatible global analytical framework.

The IFPRI team has considerable experience in spatial data development and application to support agricultural policy analysis and development investment decision making (Wood and Pardey 1998, Wood et al. 1999, Wood et al. 2003). Team members were instrumental in the development of both the GRUMP and Agro-Maps global database initiatives, and developed the entropy-based optimization approach utilized in the Spatial Production Allocation Model, SPAM (You and Wood 2006, You et al. 2009)

As its contribution to the Global Foresight database initiative, IFPRI proposes to develop two entirely new databases (see B and C below) for SSA, that build out from the existing SPAM database (A) and, in particular, extend and integrate crop and livestock systems data (B) as well as adding a range of smallholder household and production system attributes and an associated production system typology to support further economic evaluation (C).

The primary goal in moving beyond measures of production distribution and performance that implicitly treat agriculture as a homogenous production process (or at best, a rainfed and an irrigated production process) is to recognize and provide some accounting framework for the wide range of conditions (e.g., varietal selection, cropping patterns and rotations, agronomic practices, and input mix and amount) under which crops and livestock are produced. While this depth of data development calls for adequately enlightened and sustained investment, there are significant advantages to gaining a better appreciation of the physical and socioeconomic heterogeneity of agriculture. The maintained hypothesis in developing regional datasets is that they can provide more robust platforms to test the likely local and aggregate consequences of change arising from environmental, technological or policy drivers than can datasets assembled within a global purview. However, by proper coordination between efforts at different geographical scales and levels of system characterization, it is likely that the relevance and reliability of both global and regional datasets will be improved. Furthermore, more coherent, consistent and reliable two-stage, nested regional and global evaluations become feasible.

The scientific and development goals of the proposed Foresight database initiative are highly congruent with those of the HarvestChoice project also supported by the IFPRI Team[23].

[23] See www.harvestchoice.org and http://labs.harvestchoice.org/

HarvestChoice is funded by the Bill and Melinda Gates Foundation with a mandate to generate and disseminate strategic, public-goods knowledge to inform agricultural development investments in SSA that improve the productivity and profitability of smallholder agriculture and improve the welfare of smallholder households. In recognition of many common data and analytical goals, HarvestChoice will co-fund the activities and identified in this proposal so as to deliver more deliverables sooner than would otherwise be the case.

Resources sought (includes Co-Financing from HarvestChoice)

Senior Scientists (10% FTE time commitment)
GIS Expert (50% FTE time commitment)
Statistician (50% FTE time commitment)
Travel

Table 8. Objectives of the pilot project at IFPRI/HarvestChoice

Variables	Data available before the start of the project	Data to be released at the end of project phase 1*	Data to be released at the end of project phase 2 (if funded)
A. Crop production system distribution and performance	SPAM Global 2000 Ver 2. Area, yield and production for 20 rainfed (commercial, smallholder) and irrigated crops for 2000 (You at al. 2009)	SPAM SSA 2005 Ver 1. Area, yield and production for rainfed (commercial and smallholder) and for 10 staple and cash crops[a] for 2005 based on second admin level (e.g. US county equivalent) statistical and household survey data. Incorporates validation feedback from CSI members and other partners on Ver 2 data,and harmonized with other nodes (10km grid)	SPAM SSA 2010 Ver 1. Area, yield and production for 15 primary rainfed (commercial and smallholder) and irrigated staple and cash crops and livestock products[b] for 2010 based on second level statistical and household survey data. Incorporating validation feedback from CSI members on Ver 3 data and cross-node harmonization (10km grid)
B. Integrated crop and livestock system mapping.	i.Dixon et al 2001 ii. SPAM Ver 2 iii. Ramankutty et al 2007 iv. Sere & Steinfeld Livestock system maps (Steinfeld and Sere, 1996, Kruska 2002) v. Crop livestock system maps (Notenbaert et al 2009)	Ver 1. SSA Integrated mapping of the distribution of SSA smallholder crop and livestock systems 2005 (focusing on identifying major crop, crop-livestock, and pastoral systems) (10km grid)	Ver 2. SSA Integrated crop and livestock system mapping 2010, including major cropping patterns and the productivity of both crops and livestock products. (10km grid)
C. Smallholder household and production system characterization	Smallholder household and production system characterization briefs for selected countries; Ethiopia, Zambia, Zimbabwe (HarvestChoice, in preparation) and SSA (Internal strategy review analysis for the Gates Foundation).	Smallholder household and production system attributes for at least 13 SSA countries[c] accounting for >80% of production (HH demographics, agricultural enterprises, input use, marketed shares, income share from agriculture, and share of expenditure on food) based on household survey data analysis. (10km grids)	Smallholder household/ production system typologies for SSA c.2004-2010 based on HH demographics, expenditure quintiles (US$ PPP), ag/non-ag, crop/ livestock production enterprises, input use, and market participation. Typology shares by gridcell (compatible with A and B results)

- Note that there will likely be gaps and (documented) incompatibilities in some variables since the underpinning sub-national statistics and nationally-representative household surveys/censuses are not available or consistent for all countries or years.
- *a.* Maize, sorghum, millet, rice, wheat, cassava, sweet potato, bean, pigeon pea, cow pea, chick pea, groundnut, coffee, cocoa, cashew
- *b.* As *a.* plus yams, banana & plantains, onions, tomatoes, beef, milk, goat/sheep meat, draught power.
- *c.* Ethiopia, Kenya, Tanzania, Uganda, Rwanda, Malawi, Zambia, Mozambique, Ghana, Nigeria, Mali, Burkina Faso, Senegal.

References to Annex II

Barona, E, N Ramankutty, G Hyman and O Coomes. 2010. The role of pasture and soybean in deforestation of the Brazilian Amazon. Environmental Research Letters. 5 (April-June 2010). Available from http://iopscience.iop.org/1748-9326/5/2/024002/fulltext

Betts, R. A., P. D. Falloon, K. K. Goldewijk, and N. Ramankutty, Biogeophysical effects of land use on climate: Model simulations of radiative forcing and large-scale temperature change, Agricultural and Forest Meteorology, 142(2-4), 216-233, 2007.

Bondeau, A., P. C. Smith, S. Zaehle, S. Schaphoff, W. Lucht, W. Cramer, D. Gerten, H. Lotze-Campen, C. MÜLler, and M. Reichstein, Modelling the role of agriculture for the 20th century global terrestrial carbon balance, Glob. Change Biol., 13(3), 679-706, 2007.

Boucher, O., Myhre, G., Myhre, A., Direct human influence of irrigation on atmospheric water vapour and climate, Climate Dynamics, 22(6-7), 597-603, 2004.

DeFries, R, T Rudel, M Uriarte and M Hansen. 2010. Deforestation driven by urban population growth and agricultural trade in the twenty-first century. Nature Geoscience. published online: 7 February 2010 | DOI: 10.1038/NGEO756

Dixon, J.A., D.P. Gibbon, A. Gulliver. Farming systems and poverty: improving farmers' livelihoods in a changing world. FAO and World Bank. Rome and Washington DC. 2001.

Döll, P., Vulnerability to the impact of climate change on renewable groundwater resources: a global-scale assessment, Environmental Research Letters, 4(3), 035006, 2009.

FAO. 2002. World Agriculture: towards 2015/2030. Food and Agriculture Organization of the United Nations. Rome.

FAO. 2009. The State Of Food and Agriculture: Livestock in the Balance. Food and Agriculture Organization of the United Nations. Rome.

Fisher, M., I. Rao, M. Ayarza, C. Lascano, J. Sanz, R. Thomas and R. Vera. 1994. Carbon storage by introduced deep-rooted grasses in the South American savannas. Nature. 371:236-238.

forcing by land-use change, Geophysical Research Letters, 34(3), L03703, 2007.

Freydank, K., and Siebert, S., Towards mapping the extent of irrigation in the last century: time series of irrigated area per country, Frankfurt Hydrology Paper 08, Institute of Physical Geography, University of Frankfurt, Frankfurt am Main, Germany, 2008.

Golub, A., T. Hertel, H.-L. Lee, S. Rose, and B. Sohngen, The opportunity cost of land use and the global potential for greenhouse gas mitigation in agriculture and forestry, Resource and Energy Economics, 31(4), 299-319, 2009.

Gordon, L. J., W. Steffen, B. F. Jonsson, C. Folke, M. Falkenmark, and A. Johannessen, Human modification of global water vapor flows from the land surface, Proc. Natl. Acad. Sci. U. S. A., 102(21), 7612-7617, 2005.

Gumma, M.K., Nelson, A., Thenkabail, P. and Singh, A.S (submitted) A 500m resolution map of rice systems in South Asia for the year 2000. Remote Sensing of Environment

Hansen, M et al. 2008. Humid tropical forest clearing from 2000 to 2005 quantified by using multitemporal and multiresolution remotely sensed data. PNAS July 8, 2008 vol. 105 no. 27 9439-9444.

Harden, C and G Hyman. 2007. Agriculture and Soil Erosion. In Veblen, T., K. Young and A. Orme, (eds). The Physical Geography of South America. Oxford University Press.

Herrero, M., P. K. Thornton, A.M. Notenbaert, S. Wood, S. Msangi, H. A. Freeman, D. Bossio, J. Dixon, M. Peters, J. van de Steeg, J. Lynam, P. Parthasarathy Rao, S. Macmillan, B. Gerard, J. McDermott, C. Seré, and M. Rosegrant. Smart Investments in Sustainable Livestock Production: Revisiting Mixed Crop-Livestock Systems. Feb 2010. Science 327.

Hyman, G, 1999. Crop distribution mapping: applications and techniques for broad-scale analysis of crop geography. In: Pande, S., C. Johansen, J. Lauren, and F.T. Bantilan Jr. GIS analysis of cropping systems. Proceedings of an international workshop on harmonization of databases for GIS analysis of cropping systems in the Asia region. Cornell University, Ithaca, USA and ICRISAT, Patencheru, India. p. 91–96.

Immerzeel, W. W., van Beek, L. P. H., Bierkens, M. F. P., Climate Change Will Affect the Asian Water Towers, Science, 328, 1382-1385, 2010.

Jarvis, A, J Touval, M Schmidt, L Sotomayor and G Hyman. 2010. Assessments of threats to ecosystems in South America. Journal for Nature Conservation. 18:180-188.

Jones, P, A Jarvis, G Hyman, S Beebe and D Pachico. 2007. Climate proofing agricultural research investments. Journal of Semi-Arid Tropics Agricultural Research. Available from http://www.icrisat.org/Journal/SpecialProject/sp12.pdf.

Kueppers, L. M., Snyder, M. A., Sloan, L. C., Irrigation cooling effect: Regional climate

Lee, E., Chase, T. N., Rajagopalan, B., Barry, R. G., Biggs, T.W., Lawrence, P. J., Effects of irrigation and vegetation activity on early Indian summer monsoon variability, International Journal of Climatology, 29(4), 573-581, 2009.

Lobell, D. B., Bonfils, C. J., Kueppers, L. M., Snyder, M. A., Irrigation cooling effect on temperature and heat index extremes, Geophysical Research Letters, 35(9), L09705, 2008.

McGuire, A. D., S. Sitch, J. S. Clein, R. Dargaville, G. Esser, J. Foley, M. Heimann, F. Joos, J. Kaplan, D. W. Kicklighter, R. A. Meier, J. M. Melillo, B. Moore, I. C. Prentice, N. Ramankutty, T. Reichenau, A. Schloss, H. Tian, L. J. Williams, and U. Wittenberg, Carbon balance of the terrestrial biosphere in the twentieth century: Analyses of CO2, climate and land use effects with four process-based ecosystem models, Global Biogeochemical Cycles, 15(1), 183-206, 2001.

Monfreda, C., N. Ramankutty, and J. Foley, Farming the planet: 2. Geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000, Global Biogeochem. Cycles, 22, GB1022, doi:1010.1029/2007GB002947, 2008.

Monfreda, C., N. Ramankutty, and J. Foley, Farming the planet: 2. Geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000, Global Biogeochem. Cycles, 22, GB1022, doi:1010.1029/2007GB002947, 2008.

Notenbaert, A., M. Herrero, R. Kruska, L. You, S. Wood, P. Thornton, and A. Omolo. “Classifying livestock production systems for targeting agricultural research and development in a rapidly changing world”, Discussion Paper No. 19, International Livestock Research Institute, Nairobi, 2009. http://hdl.handle.net/10568/589%20

Otte, M.J., and P. Chilonda. Cattle and small ruminant production systems in sub-Saharan Africa: A systematic review. Livestock Information Sector Analysis and Policy Branch, FAO Rome. 2002.

Portmann, F. T., S. Siebert, and P. Döll, MIRCA2000 Global monthly irrigated and rainfed crop areas around the year 2000: A new high-resolution data set for agricultural and hydrological modeling, Global Biogeochem. Cycles, 24(1), GB1011, 2010.

Portmann, F. T., S. Siebert, and P. Döll, MIRCA2000 Global monthly irrigated and rainfed crop areas around the year 2000: A new high-resolution data set for agricultural and hydrological modeling, Global Biogeochem. Cycles, 24(1), GB1011, 2010.

Potter, P., N. Ramankutty, E. Bennett, and S. Donner, Characterizing the Spatial Patterns of Global Fertilizer Application and Manure Production, Earth Interactions, 14, 2010.

Puma, M. J., and Cook, B. I., Effects of irrigation on global climate during the 20th century, Journal of Geophysical Research – Atmospheres, 115, D16120, 2010.

R.E. Huke & E.H. Huke, 1982. Rice area by type of culture: South, Southeast, and East Asia. IRRI

R.E. Huke & E.H. Huke, 1988. Human Geography of rice in South Asia. IRRI

R.E. Huke & E.H. Huke, 1990. Human Geography of rice in Southeast Asia. IRRI

R.E. Huke, 1982. Agroclimatic and dry-season maps of South, Southeast, and east Asia. IRRI

R.E. Huke, R.E. & E.H. Huke, 1997. Rice area by type of culture. South, Southeast, and East Asia. A revised and updated database. IRRI

Ramankutty, N., A. T. Evan, C. Monfreda, and J. A. Foley, Farming the planet: 1. Geographic distribution of global agricultural lands in the year 2000, Glob. Biogeochem. Cycles, 22, GB1003, doi:1010.1029/2007GB002952, 2008.

Ramankutty, N., A. T. Evan, C. Monfreda, and J. A. Foley, Farming the planet: 1. Geographic distribution of global agricultural lands in the year 2000, Glob. Biogeochem. Cycles, 22, GB1003, doi:1010.1029/2007GB002952, 2008.

Ramankutty, N., A. T. Evan, C. Monfreda, and J. A. Foley, Farming the planet: 1. Geographic distribution of global agricultural lands in the year 2000, Glob. Biogeochem. Cycles, 22, GB1003, doi:1010.1029/2007GB002952, 2008.

Ramankutty, N., and A. Kimball, Three Centuries of Expansion: A Reconstruction of the Spatial Patterns of Cropland and Pastures from 1700 to 2007, Earth System Science Data, in prep.

Ramankutty, N., and J. A. Foley, Estimating historical changes in global land cover: Croplands from 1700 to 1992, Global Biogeochemical Cycles, 13, 997-1027, 1999.

Ramankutty, N., and J. A. Foley, Estimating historical changes in global land cover: Croplands from 1700 to 1992, Global Biogeochemical Cycles, 13, 997-1027, 1999.

Saeed, F., Hagemann, S., Jacob, D., Impact of irrigation on the South Asian summer monsoon, Geophysical Research Letters, 36, L20711, 2009.

Sere, C., Steinfeld, H., 1996. World livestock production systems: current status, issues and trends. Animal Production and Health Paper 127. FAO, Rome. 1996

Siebert, S., and Döll, P., Quantifying blue and green virtual water contents in global crop production as well as potential production losses without irrigation, Journal of Hydrology, 384(3-4), 198-217, 2010.

Siebert, S., Burke, J., Faures, J. M., Frenken, K., Hoogeveen, J., Döll, P., Portmann, F. T., Groundwater use for irrigation – a global inventory, Hydrology and Earth System Sciences 14, 1863-1880, 2010.

Siebert, S., Döll, P., Hoogeveen, J., Faures, J-M., Frenken, K., and Feick, S., Development and validation of the global map of irrigation areas, Hydrology and Earth System Sciences, 9, 535-547, 2005.

Siebert, S., Hoogeveen, J., and Frenken, K. (2006): Irrigation in Africa, Europe and Latin America - Update of the Digital Global Map of Irrigation Areas to Version 4. Frankfurt Hydrology Paper 05, Institute of Physical Geography, University of Frankfurt, Frankfurt am Main, Germany and Food and Agriculture Organization of the United Nations, Rome, Italy, 2006.

Taheripour, F., T. W. Hertel, W. E. Tyner, J. F. Beckman, and D. K. Birur, Biofuels and their by-products: Global economic and environmental implications, Biomass and Bioenergy, 34(3), 278-289, 2010.

Van Minnen, J. G., K. K. Goldewijk, E. Stehfest, B. Eickhout, G. van Drecht, and R. Leemans, The importance of three centuries of land-use change for the global and regional terrestrial carbon cycle, Clim. Change, 97(1-2), 123-144, 2009.

White, D, F Holman, S Fujisaka, K Reategui and C Lascano. 1999. Will intensifying pasture management in Latin America protect forest – or is it the other way around? In Angelsen, A and D. Kaimowitz, eds. Agricultural Technologies and Tropical Deforestation. CABI Publishing. Wallingford, UK.

Wood, S. and P.G. Pardey. 1998. Agroecological Aspects of Evaluating Agricultural R&D. Agricultural Systems 57: 1, 13-41.

Wood, S., K. Sebastian, and J. Chamberlin. 2003. Land quality, agricultural productivity, and food security: a spatial perspective. In K. Weibe, Land quality, agricultural productivity, and food security: biophysical processes and economic choices at local, regional, and global levels (pp. 47-110). Edward Elgar Publishing Limited, Cheltenham, UK/Northampton, MA.

Wood, S., K. Sebastian, F. Nachtergaele, D. Nielsen, and A. Dai. 1999. Spatial aspects of the design and targeting of agricultural development strategies. Environment and Production Technology Division Discussion Paper 44. International Food Policy Research Institute. Washington, D.C.

You, L. and S. Wood. 2006. An entropy approach to spatial disaggregation of agricultural production. Agricultural Systems 90 (1-3), 329-347

You, L., S. Wood, and U. Wood-Sichra, Generating plausible crop distribution maps for Sub-Saharan Africa using a spatially disaggregated data fusion and optimization approach, Agricultural Systems, 99(2-3), 126-140, 2009.

You, L., S. Wood, and U. Wood-Sichra, Generating plausible crop distribution maps for Sub-Saharan Africa using a spatially disaggregated data fusion and optimization approach, Agricultural Systems, 99(2-3), 126-140, 2009.

References

Alexandratos, Nikos. 2010. Critical Evaluation of Selected Projections. In . Rome, Italy.

Betts, Richard A., Peter D. Falloon, Kees Klein Goldewijk, and Navin Ramankutty. 2007. Biogeophysical effects of land use on climate: Model simulations of radiative forcing and large-scale temperature change. *Agricultural and Forest Meteorology* 142, no. 2-4 (February 12): 216-233. doi:10.1016/j.agrformet.2006.08.021.

Boucher, O, G Myhre, and A Myhre. 2004. Direct human influence of irrigation on atmospheric water vapour and climate. *Climate Dynamics* 22, no. 6 (June 1): 597-603.

Bounoua, L., R. DeFries, G. J. Collatz, P. Sellers, and H. Khan. 2002. Effects of Land Cover Conversion on Surface Climate. *Climatic Change* 52, no. 1 (January 1): 29-64. doi:10.1023/A:1013051420309.

Britz, Wolfgang, and Thomas W. Hertel. 2009. Impacts of EU biofuels directives on global markets and EU environmental quality: An integrated PE, global CGE analysis. *Agriculture, Ecosystems & Environment* In Press. doi:10.1016/j.agee.2009.11.003.

Brovkin, V., M. Claussen, E. Driesschaert, T. Fichefet, D. Kicklighter, M. Loutre, H. Matthews, N. Ramankutty, M. Schaeffer, and A. Sokolov. 2006. Biogeophysical effects of historical land cover changes simulated by six Earth system models of intermediate complexity. *Climate Dynamics* 26, no. 6 (May 1): 587-600. doi:10.1007/s00382-005-0092-6.

Bruinsma, Jelle. 2009. The resource outlook to 2050. By how much do land, water use and crop yields need to increase by 2050? In *Session 2: The resource base to 2050: Will there be enough land, water and genetic potential to meet future food and biofuel demands?* Rome, Italy, June 24.

Chase, T. N., R. A. Pielke Sr., T. G. F. Kittel, R. R. Nemani, and S. W. Running. 2000. Simulated impacts of historical land cover changes on global climate in northern winter. *Climate Dynamics* 16, no. 2 (February 8): 93-105. doi:10.1007/s003820050007.

Chomitz, Kenneth M., and David A. Gray. 1996. Roads, Land Use, and Deforestation: A Spatial Model Applied to Belize. *The World Bank Economic Review* 10, no. 3 (September): 487-512.

DeFries, R., and F. Achard. 2002. New estimates of tropical deforestation and terrestrial carbon fluxes: Results of two complementary studies. *LUCC Newsletter* 8: 7-9.

Diffenbaugh, Noah S, Christian H Krupke, Michael A White, and Corinne E Alexander. 2008. Global warming presents new challenges for maize pest management. *Environmental Research Letters* 3, no. 4 (10): 044007. doi:10.1088/1748-9326/3/4/044007.

Döll, Petra. 2009. Vulnerability to the impact of climate change on renewable groundwater resources: a global-scale assessment. *Environmental Research Letters* 4, no. 3 (7): 035006. doi:10.1088/1748-9326/4/3/035006.

Findell, Kirsten L., Elena Shevliakova, P. C. D. Milly, and Ronald J. Stouffer. 2006. Modeled Impact of Anthropogenic Land Cover Change on Climate. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.143.9162.

Fischer, Günther. 2009. How do climate change and bioenergy alter the long-term outlook for food, agriculture and resource availability? In *Session 2: The resource base to 2050: Will there be enough land, water and genetic potential to meet future food and biofuel demands?* Rome, Italy, June 24.

Fischer, Günther, Francesco N. Tubiello, Harrij van Velthuizen, and David A. Wiberg. 2007. Climate change impacts on irrigation water requirements: Effects of mitigation, 1990-2080. *Technological Forecasting and Social Change* 74, no. 7 (September): 1083-1107. doi:10.1016/j.techfore.2006.05.021.

Gibbs, H. K., A.S. Ruesch, F. Achard, M. Clayton, P. Holmgren, N. Ramankutty, and J.A. Foley. Pathways of Agricultural Expansion Across the Tropics: Implications for Forest Conservation and Carbon Emissions. *Proc Natl Acad Sci US* (Submitted).

Goldewijk, K. Klein, G. Van Drecht, and A. F. Bouwman. 2007. Mapping contemporary global cropland and grassland distributions on a 5 × 5 minute resolution. *Journal of Land Use Science* 2, no. 3: 167. doi:10.1080/17474230701622940.

Goldewijk, Kees Klein. 2001. Estimating global land use change over the past 300 years: The HYDE database. *Global Biogeochemical Cycles* 15: 417-434.

Green, Rhys E., Stephen J. Cornell, Jorn P. W. Scharlemann, and Andrew Balmford. 2005. Farming and the Fate of Wild Nature. *Science* 307, no. 5709 (January 28): 550-555. doi:10.1126/science.1106049.

Hansen, M. C, R. S Defries, J. R. G Townshend, M. Carroll, C. Dimiceli, and R. A Sohlberg. 2003. Global percent tree cover at a spatial resolution of 500 meters: first results of the MODIS vegetation continuous fields algorithm. *EARTH INTERACTIONS, 7, PAPER* 10. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.116.495.

Hansen, Matthew C., Stephen V. Stehman, and Peter V. Potapov. 2010. Quantification of global gross forest cover loss. *Proceedings of the National Academy of Sciences* 107, no. 19 (May 11): 8650 -8655. doi:10.1073/pnas.0912668107.

Hertel, Thomas, Steven K. Rose, and Richard S.J. Tol. 2009. Economic Analysis of Land Use in Global Climate Change Policy (Hardback). http://www.routledge.com/books/details/9780415773089/.

Hertel, Thomas W. 2010. The Global Supply and Demand for Agricultural Land in 2050: A Perfect Storm in the Making? In *AAEA Presidential Address*. Denver, CO, July 25.

Hertel, Thomas W., Alla G Golub, Andrew Jones, Michael O'Hare, Richard Plevin, and Daniel Kammen. 2010. Global Land Use and Greenhouse Gas Emissions Impacts of US Maize Ethanol: Estimating Market-Mediated Responses. *BioScience* 60, no. 3.

Houghton, R. A. 2003. Why are estimates of the terrestrial carbon balance so different? *Global Change Biology* 9, no. 4: 500-509. doi:10.1046/j.1365-2486.2003.00620.x.

IIASA/FAO. 2010. GAEZ 2009. Global Agro-ecological Zones. Model Documentation.

Immerzeel, Walter W., Ludovicus P. H. van Beek, and Marc F. P. Bierkens. 2010. Climate Change Will Affect the Asian Water Towers. *Science* 328, no. 5984 (June 11): 1382-1385. doi:10.1126/science.1183188.

Ivanic, Maros, and Will Martin. 2008. Implications of Higher Global Food Prices for Poverty in Low-Income Countries. *SSRN eLibrary* (April 1). http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1149097.

Johnston, Matt, Jonathan A Foley, Tracey Holloway, Chris Kucharik, and Chad Monfreda. 2009. Resetting global expectations from agricultural biofuels. *Environmental Research Letters* 4, no. 1 (1): 014004. doi:10.1088/1748-9326/4/1/014004.

Kueppers, Lara M., Mark A. Snyder, and Lisa C. Sloan. 2007. Irrigation cooling effect: Regional climate forcing by land-use change. *Geophysical Research Letters* 34 (February 7): 5 PP. doi:200710.1029/2006GL028679.

Kurz, Werner A. 2010. An ecosystem context for global gross forest cover loss estimates. *Proceedings of the National Academy of Sciences* 107, no. 20 (May 18): 9025 -9026. doi:10.1073/pnas.1004508107.

Lee, Eungul, Thomas N. Chase, Balaji Rajagopalan, Roger G. Barry, Trent W. Biggs, and Peter J. Lawrence. 2009. Effects of irrigation and vegetation activity on early Indian summer monsoon variability. *International Journal of Climatology* 29, no. 4 (3): 573-581. doi:10.1002/joc.1721.

Lepers, Erika, Eric F. Lambin, Anthony C. Janetos, Ruth S. DeFries, Frédéric Achard, Navin Ramankutty, and Robert J. Scholes. 2005. A Synthesis of Information on Rapid Land-cover Change for the Period 1981–2000. *BioScience* 55, no. 2: 115-124. doi:doi: 10.1641/0006-3568(2005)055[0115:ASOIOR]2.0.CO;2.

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?

Liu, Junguo, Liangzhi You, Manouchehr Amini, Michael Obersteiner, Mario Herrero, Alexander J. B. Zehnder, and Hong Yang. 2010. A high-resolution assessment on global nitrogen flows in cropland. *Proceedings of the National Academy of Sciences* 107, no. 17 (April 27): 8035 -8040. doi:10.1073/pnas.0913658107.

Lobell, David B., Celine J. Bonfils, Lara M. Kueppers, and Mark A. Snyder. 2008. Irrigation cooling effect on temperature and heat index extremes. *Geophysical Research Letters* 35 (May 7): 5 PP. doi:200810.1029/2008GL034145.

Lobell, David B., Marshall B. Burke, Claudia Tebaldi, Michael D. Mastrandrea, Walter P. Falcon, and Rosamond L. Naylor. 2008. Prioritizing Climate Change Adaptation Needs for Food Security in 2030. *Science* 319, no. 5863 (February 1): 607-610. doi:10.1126/science.1152339.

Lobell, David B., Kenneth G. Cassman, and Christopher B. Field. 2009. Crop Yield Gaps: Their Importance, Magnitudes, and Causes. *Annual Review of Environment and Resources* 34, no. 1 (11): 179-204. doi:10.1146/annurev.environ.041008.093740.

Martin, Will, and L. Alan Winters. 1997. *The Uruguay Round and the Developing Countries*. Cambridge University Press, January 28.

Matthews, H. Damon, Andrew J. Weaver, Michael Eby, and Katrin J. Meissner. 2003. Radiative forcing of climate by historical land cover change. *Geophysical Research Letters* 30 (January 1): 27-1.

McGuire, A. D., S. Sitch, J. S. Clein, R. Dargaville, G. Esser, J. Foley, M. Heimann, et al. 2001. Carbon balance of the terrestrial biosphere in the Twentieth Century: Analyses of CO2, climate and land use effects with four process-based ecosystem models. *Global Biogeochemical Cycles* 15, no. 1: PP. 183-206. doi:200110.1029/2000GB001298.

McKinsey & Co. 2009. *Charting Our Water Future: Economic frameworks to inform decision-making*. 2030 Water Resources Group: McKinsey & Co.

Monfreda, Chad, Navin Ramankutty, and Jonathan A. Foley. 2008. Farming the planet: 2. Geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000. *Global Biogeochemical Cycles* (March 1): 1:19.

Nelson, Gerald C., Mark W. Rosegrant, Jawoo Koo, Richard Robertson, Timothy Sulser, Tingju Zhu, Claudia Ringler, et al. 2010. *The Costs of Agricultural Adaptation to Climate Change*. International Food Policy Research Institute (IFPRI).

Neumann, Kathleen, Peter H. Verburg, Elke Stehfest, and Christoph Müller. 2010. The yield gap of global grain production: A spatial analysis. *Agricultural Systems* 103, no. 5 (June): 316-326.

Pitman, A. J., N. de Noblet-Ducoudré, F. T. Cruz, E. L. Davin, G. B. Bonan, V. Brovkin, M. Claussen, et al. 2009. Uncertainties in climate responses to past land cover change: First results from the LUCID intercomparison study. *Geophysical Research Letters* 36 (July 23): 6 PP. doi:200910.1029/2009GL039076.

Portmann, Felix T., Stefan Siebert, and Petra Döll. 2010. MIRCA2000—Global monthly irrigated and rainfed crop areas around the year 2000: A new high-resolution data set for agricultural and hydrological modeling. *Global Biogeochemical Cycles* 24 (March 13): 24 PP. doi:201010.1029/2008GB003435.

Potter, Philip, Navin Ramankutty, Elena M Bennett, and Simon D Donner. 2010. Characterizing the Spatial Patterns of Global Fertilizer Application and Manure Production. *Earth Interactions* 14, no. 2: 1-22.

Powell, Alan. 2007. Why How and When did GTAP Happen? What has it Achieved? Where is it Heading? http://www.gtap.agecon.purdue.edu/resources/res_display.asp?RecordID=2467.

Puma, M. J., and B. I. Cook. 2010. Effects of irrigation on global climate during the 20th century. *Journal of Geophysical Research* 115 (August 28): 15 PP. doi:201010.1029/2010JD014122.

Ramankutty, Navin, Amato T. Evan, Chad Monfreda, and Jonathan A. Foley. 2008. Farming the planet: 1. Geographic distribution of global agricultural lands in the year 2000 (January 17). http://www.agu.org/pubs/crossref/2008/2007GB002952.shtml.

Ramankutty, Navin, and Jonathan A. Foley. 1999. Estimating historical changes in global land cover: Croplands from 1700 to 1992. *Global Biogeochemical Cycles* 13, no. 4: PP. 997-1027. doi:199910.1029/1999GB900046.

Ramankutty, Navin, Holly K. Gibbs, Frédéric Achard, Ruth DeFries, Jonathan A. Foley, and R. A. Houghton. 2007. Challenges to estimating carbon emissions from tropical deforestation. *Global Change Biology* 13, no. 1: 51-66. doi:10.1111/j.1365-2486.2006.01272.x.

Riahi, Keywan, Arnulf Grübler, and Nebojsa Nakicenovic. 2007. Scenarios of long-term socio-economic and environmental development under climate stabilization. *Technological Forecasting and Social Change* 74, no. 7 (September): 887-935. doi:10.1016/j.techfore.2006.05.026.

Saeed, Fahad, Stefan Hagemann, and Daniela Jacob. 2009. Impact of irrigation on the South Asian summer monsoon. *Geophysical Research Letters* 36 (October 30): 7 PP. doi:200910.1029/2009GL040625.

Schlenker, Wolfram, W. Michael Hanemann, and Anthony C. Fisher. 2005. Will U.S. Agriculture Really Benefit from Global Warming? Accounting for Irrigation in the Hedonic Approach. *The American Economic Review* 95, no. 1 (March): 395-406.

Schlenker, Wolfram, and David B Lobell. 2010. Robust negative impacts of climate change on African agriculture. *Environmental Research Letters* 5, no. 1 (1): 014010. doi:10.1088/1748-9326/5/1/014010.

Schlenker, Wolfram, and Michael J. Roberts. 2009. Nonlinear temperature effects indicate severe damages to U.S. crop yields under climate change. *Proceedings of the National Academy of Sciences* 106, no. 37: 15594-15598. doi:10.1073/pnas.0906865106.

Searchinger, Timothy, Ralph Heimlich, R. A. Houghton, Fengxia Dong, Amani Elobeid, Jacinto Fabiosa, Simla Tokgoz, Dermot Hayes, and Tun-Hsiang Yu. 2008. Use of U.S. Croplands for Biofuels Increases Greenhouse Gases Through Emissions from Land-Use Change. *Science* 319, no. 5867 (February 29): 1238-1240. doi:10.1126/science.1151861.

Shevliakova, Elena, Stephen W. Pacala, Sergey Malyshev, George C. Hurtt, P. C. D. Milly, John P. Caspersen, Lori T. Sentman, Justin P. Fisk, Christian Wirth, and Cyril Crevoisier. 2009. Carbon cycling under 300 years of land use change: Importance of the secondary vegetation sink. *Global Biogeochemical Cycles* 23 (June 23): 16 PP. doi:200910.1029/2007GB003176.

Siebert, Stefan, and Petra Döll. 2010. Quantifying blue and green virtual water contents in global crop production as well as potential production losses without irrigation. *Journal of Hydrology* 384, no. 3-4 (April 30): 198-217. doi:10.1016/j.jhydrol.2009.07.031.

Waggoner, Paul E. 2009. *Forest Inventories: Discrepancies and Uncertainties*. Resources For the Future, August. http://ideas.repec.org/p/rff/dpaper/dp-09-29.pdf.html.

White, M. A., N. S. Diffenbaugh, G. V. Jones, J. S. Pal, and F. Giorgi. 2006. Extreme heat reduces and shifts United States premium wine production in the 21st century. *Proceedings of the National Academy of Sciences* 103, no. 30 (July 25): 11217-11222. doi:10.1073/pnas.0603230103.

You, Liangzhi, and Stanley Wood. 2006. An entropy approach to spatial disaggregation of agricultural production. *Agricultural Systems* 90, no. 1-3 (October): 329-347. doi:10.1016/j.agsy.2006.01.008.

You, Liangzhi, Stanley Wood, and Ulrike Wood-Sichra. 2009. Generating plausible crop distribution maps for Sub-Saharan Africa using a spatially disaggregated data fusion and optimization approach. *Agricultural Systems* 99, no. 2-3. Agricultural Systems: 126-140.