Data Fusion Revisited

Paul Preckel

September 11, 2013

Land Allocation Challenge

- Current state of the art
  - Limited data available at fine grained spatial level (our focus is on “pixel” level)
  - Actual land use data is available for more aggregate regions (sums across pixels)
- Reconciliation approaches
  - “Downscaling”
  - Entropy-based with economic competition priors
  - Discrete choice models

Land Allocations Issues

- Not all procedures are fully documented
- Discovering the provenance of some data can be a challenge
  - Ad hoc data adjustments by analysts – not all of these are well-known and documented
  - Use of outputs of predecessor models as data – are we modeling models or physical processes?

Data Fusion Vision – Land Allocation

- Features
  - Procedure
    - Open source
    - Modular to permit experimentation with alternative models
  - Data
    - Publically available
    - Minimally processed
    - Provenance well documented
    - Layered to facilitate incremental improvement

Data Hierarchy

- Fundamental challenge – data is available at different levels of granularity
  - Pixel-level data may have different pixel sizes, needing procedures for consistently pixelating to a standard (variable) size (elevation, slope, distance to markets, satellite images, etc.)
  - Actual land use is generally only available for aggregate areas – sometimes can be pixelated (e.g. lakes) and sometimes not (e.g. land in a specific crop)

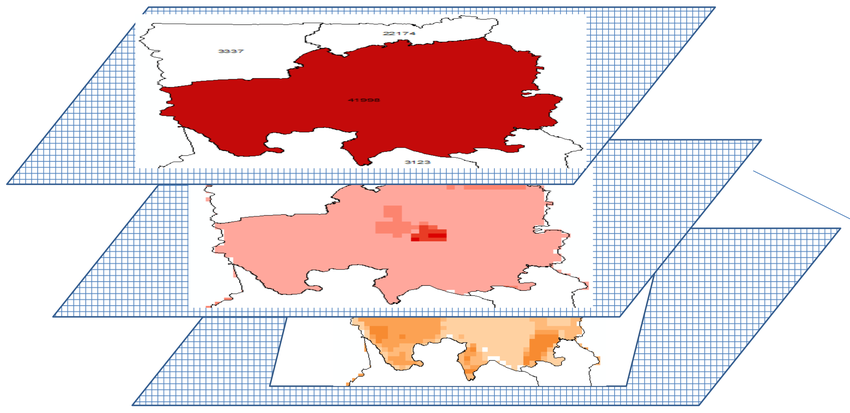
Procedures

- Ideas circulating
  - Downscaling
  - Discrete choice modeling
  - Entropy
  - Calibration
- Perhaps these could be combined to create approaches to land allocation projection

One Concept

- Possible procedure
  - Inventory pixelated land attribute data
  - Use downscaling to reconcile pixilation at a consistent level of spatial granularity
  - Modify discrete choice estimation procedures to fit aggregate land use observations
  - Use the discrete choice model (DCM) to project a land allocation prior
  - Use entropy-based or calibration approach with DCM prior and constrain aggregate totals

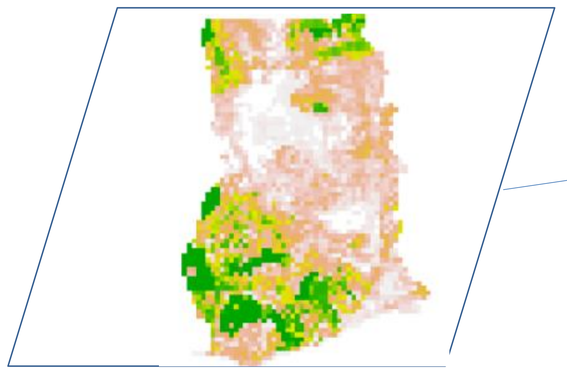
Pixelated Data Layers



Disc. Choice Model

Pr(use *i* on pixel *n*)

Model Calibration



Map w/Admin Area Totals for Land Use

Factors to Consider

- Spatial scope – need to ensure adequate variability in independent variables
- Land use scope – may need to focus on aggregations of land uses (e.g. coarse grains, oil seeds, etc.)
- Validation essential
  - Ground truthing
  - Out of sample validation

Factors to Consider (cont'd.)

- All processing of should be internal to the estimation program including:
  - Raw land attribute data
  - Aggregate region land use observations
- Computer language should be
  - Flexible enough to handle data processing and estimation
  - Well-known
  - Easily accessible and capable of self documentation of the program