Land Use Harmonization Model Documentation

The land-use harmonization code-set includes multiple scripts for pre-processing input data from HYDE and Integrated Assessment Models (IAMs), as well as code for the Global Land-use Model (GLM) which computes all gridded land-use states and transitions 1500-2100 associated with the HYDE and IAM data. This document gives additional details about the code and models in the land-use harmonization public code release. It is not intended to be a "user's guide", rather an explanation of the various components of the model and the general workflow that is used to run the model.

The land-use harmonization workflow is as follows:

1. **Pre-process the HYDE data**
   Since all future land-use states and transitions must be harmonized with the historical land-use states and transitions, the data for the historical period (starting in 1500, 1700, or 1850) must first be generated with GLM. To run GLM for the historical period, the HYDE data is first pre-processed (using the scripts in `HYDE_preprocessing`) to generate annual half-degree grids of the area fraction of cropland, pasture, urban land, ice-water area, and natural vegetation (including both primary and secondary land) and vegetation class (forest or non-forest) for each grid-cell.

2. **Run historical GLM simulations**
   Once the input grids have been generated in step 1 above, GLM is run for the historical period to compute the annual, gridded transitions between land-use states, as well as the age, area, and biomass of secondary land. The shell script `run_glm.sh` specifies various options for the GLM simulation, and then executes the GLM code. Options that are set within the shell script include the historical start date, preference for primary or secondary land for land-use conversions, and whether or not to include wood harvesting and/or shifting cultivation. The shell script compiles and executes the GLM code in five stages (to ensure there is enough memory for the data stored in GLM). The GLM code that is compiled and executed for the historical simulation is contained in `glm.c` and `glm_urban.c` (the latter includes urban land-use, the former doesn't).

3. **Pre-process the IAM data**
   Once the historical GLM simulation has run, the future simulations using the IAM data are run. The first step is to pre-process the IAM data so that it is "harmonized" with the historical data (i.e., so that cropland, pasture, and urban gridcell fractions vary continuously across 2005 while preserving as much information from IAMs as possible) and to ensure that the data is in the

needed format for input into GLM. The wood harvest data from the IAMs is also pre-processed to generate suitable input files for GLM. The scripts to carry out these pre-processing steps are in:
`IAM_preprocessing/IMAGE`
`IAM_preprocessing/MiniCAM`
`IAM_preprocessing/AIM`
`IAM_preprocessing/MESSAGE`

4. **Run future GLM simulations**
   After pre-processing the IAM data, GLM is then run for the future period from 2005-2100 to compute the annual, gridded transitions between land-use states, as well as the age, area, and biomass of secondary land. This is done using the same shell script `run_glm.sh` that was used in Step 2, but configured for a "future simulation". The various options that GLM uses are set in this shell script and the particular IAM dataset that is being used is also chosen. The shell script compiles and executes the GLM code in two stages – the first for the years 2005 to 2050, and the second for the years 2050 to 2100. The GLM code that is used for the future simulations is contained in `glm.future.c` and `glm_urban.future.c`.