GGCMI Phase3 soil input

ISIMIP proposes to use the GSWP3 soil data from HWSD, aggregated to 0.5° using method A, but we felt that this is not the best way to represent soils. To be consistent with ISIMIP, we also derive the values from the HWSD but we use a different algorithm to aggregate these to 0.5 arc-degrees.
With this procedure we are consistent with ISIMIP, as our soil data are based on HWSD and can be used by other sectors. The data provided here is cropland specific, but it can be easily adjusted to process data without the cropland weighting, if needed by other ISIMIP sectors.

aggregation

1. HWSD 1.12 downloaded, which comes at 30-arc-second, each raster cell is assigned to one mapping unit (MU_GLOBAL)
2. assign soil type to each raster cell by lookup table that lists various soil properties per MU_GLOBAL (there are 48148 MU_GLOBALs)
3. select dominant soil type within each 30-arc-minute (= 0.5 arc-degrees, the target resolution) using MIRCA2000 total harvested area as weights
    - replacing missing values in each MU_GLOBAL with the largest SHARE with those of the same MU_GLOBAL but smaller shares if available.
    - MIRCA2000 harvested areas come at 5 arc-minute spatial resolution, we are assuming that cropland is evenly distributed across all 30-arc-second pixels within each 5 arc-minute grid cell.
4. as several MU_GLOBALs can contribute to the dominant soil texture type within a grid cell, in a second step we are thus selecting the dominant MU_GLOBAL within the dominant soil texture class.
    - MU_GLOBAL entries can have several properties (i.e. several lines in lookup table with the same MU_GLOBAL). Here, we are using the HWSD feature "SHARE" to select the line of soil properties to use (select [first of] largest SHARE).
5. this gives us one single MU_GLOBAL per 30-arc-minute grid cell, which is the (cropland-weighted) dominant MU_GLOBAL on all cropland of the dominant soil texture class.
   At this point, we are filling data gaps by assigning the dominant MU_GLOBAL of the surrounding pixels with valid data to pixels that are included in the ISIMIP land mask, but have no value in HWSD (92 pixels missing). We are starting with the 8 neighboring cells and increase the radius by 1 until we have at least one valid MU_GLOBAL in the neighboring cells. In this step we are assuming that all cells are of equal size (which is not true, but an acceptable simplification, we hope, as the error we make by this nearest neighbor approximation is likely much larger than that imprecision)
6. in a final step, all other soil properties are taken from the HWSD look-up table that has values for each MU_GLOBAL
    - remember that missing properties were gap-filled if necessary from other variants of the same MU_GLOBAL (those with smaller SHAREs)
    - if a specific soil property was not available in any of the MU_GLOBAL variants, these properties are gap-filled using the nearest-neighbor approach as in step 5.
7. We then select that mapping unit to be the representative soil in the 0.5 degree pixel, so that each 0.5 degree pixel has one specific value for soil texture fractions, soil texture class, pH, CEC, ECE, carbonate content, gravel fractions etc.

notes on data usage

- We are using only the top-soil properties here, but if people need different parameters for the subsoil layers, these could be processed in the same way and can be extracted from the HWSD look-up table using MU_GLOBAL as the key
- Users should find all soil properties that were discussed in the GGCMI group in the netcdf. If additional properties are needed, these can be taken from the HWSD data base, simply selecting the corresponding value of the MU_GLOBAL. They should pay attention to the different options for MU_GLOBAL and select the one that has the largest value in column "SHARE".
- Those who need specific texture values rather than soil texture classes should take those supplied in the netcdf file (rather than using an average or otherwise representative texture mix per soil class).

versions

- all versions <1.0 are for internal discussions only.
- version 1.0 was published on Globus for the GGCMI groups on April 14, 2020.
- version 1.01 is as version 1.0 but was extended to include the parameter ISSOIL masking non-(mineral) soils. Released on Globus for the GGCMI groups on April 27, 2020 and sent to ISIMIP.
- version 2.0 with fixed bug in treating missing values (were interpreted as zero per default, which is wrong)
- version 2.1 with fixed bug for gap filling data (nearest neighbor extrapolation)
- version 2.2 with fixed gap filling issues for individual cases
- version 2.3 corrected units for impermeable_layer and root_obstacles

See #354 for the discussion.