**CONVERSION from LPJ-GUESS to CLM5**

In order to run the semi-realistic simulation, data from Tang et al. (2021) is taken as input to run the simulation in CLM5. This data is the output of LPJ-GUESS, a Dynamic Global Vegetation Model (DGVMs), a process-based model that simulates the vegetation dynamically based on a stochastic individual method – it can take into account the behavior of individuals depending on external conditions, as well as individual stochasticity (Lavallée et al., 2019). On the contrary the simulations in this work are run with CLM5, which treat a canopy as a single leaf.

The grid resolution and the PFT classification are different in the two models. While to match the grid a linear interpolation method is chosen to make the LPJGUESS grid to the coarser one of CLM5, for the PFT conversion a finer analysis is needed.

**PFT CONVERSION**

The following table lists a part of LPJ-GUESS PFTs, the one used in Tang et al. (2021). Some properties and a tentative conversion to CLM5 PFTs, based on distribution, growth form and leaf phenology are also shown

(sources…)

Table

Description automatically generated

**Main differences and challenges for conversion:**

* LPJGUESS has more PFTs, but many do not match exactly with no PFTs in CLM5.
* LPJGUESS has peatland PFTs, while CLM5 has just “natural PFTs”. The peatland PFTs and natural PFTs are normalized separately in LPJGUESS.
* LPJGUESS highlights the shade-intolerance property, not considered in CLM5.
* LPJGUESS has moss and lichen (PFT named CLM)
* In categorizing the shrubs, LPJGUESS priorities the height and the leaf phenology over the distribution, as CLM5 does instead. This makes hard to straight convert the PFTs.
* Some PFTs are ambiguous (GRT, EPDS, SPDS, CML). In LPJGUESS are considered belonging to the grass vegetation group (sources…), but it can be different for CLM5.
* CLM has also the bare ground as first PFT.

For all these listed reasons, a further analysis on the vegetation distribution is performed. The historical dataset from Tang et al. (2021) is considered for the comparison. It represents the foliage projection cover (FPC) averaged in the historical period 1971-2000. On the CLM5 side, the comparison is performed with the PFT fractional cover over each landunit from the input surface data file from a simulation in the year 2000 (further properties of the datasets?).

Before going into the distribution comparison, here is a list of the differences between the two models when performing this analysis:

* *A picture containing chart

  Description automatically generatedGrid resolution*: LPJGUESS has a finer resolution () than CLM5 (). A linear interpolation of the LPJGUESS variable over the CLM5 grid (center of each gridcell) is then performed.
* *Covered area:* data from LPJGUESS covers just a part of the land area above 45°N, while we have the whole grid range for CLM5. The data will be then compared in the common area. (why did they choose that area?)
* *Time*: while the LPJGUESS data comes from an average over thirty-year period, the CLM5 data is from single year (more details about the simulation that generated this surfdata file). But this shouldn’t be a big problem as we perfom a visual comparison and we assume that in those decades a possible northward shift is not quantitative relevant for this analysis.
* *Variable:* LPJGUESS uses FPC that represents the fraction of ground area occupied by the vertical projection of the foliage of each of the present PFTs (Deceukelier 2020-21). The sum of the FPCs of all PFTs cannot exceed 1. While in the surfdata file from CLM5 the occurrence of PFTs is the fractional cover over the landunit (% of the landunit). Are there differences? Maybe different way to define “cover”?
* *Variable values:* The sum of FPC for each grid cell is 0 ≤ x ≤ 1, while in CLM5 the sum of all PFTs is always 100% (or 1). This is because CLM5 considers the bare ground as PFT. In addition, the variable in CLM5 is evaluated over the landunit, but when dealing with the coastline cells the area considered is different in the two models. Knowing the land fraction in CLM5 we can simply rescale the data to be coherent.
* *PFT types:* as previously mentioned, LPJGUESS has also peatland PFTs modelled. After a quick graphical check, we could see that the only relevant ones (pmoss and WetGRS) do not correspond to any pattern of the natural PFTs in CLM5.

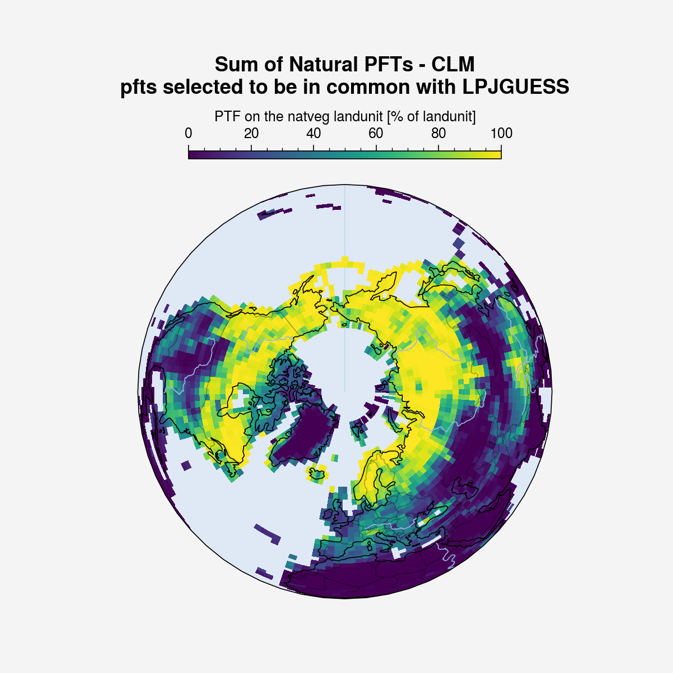
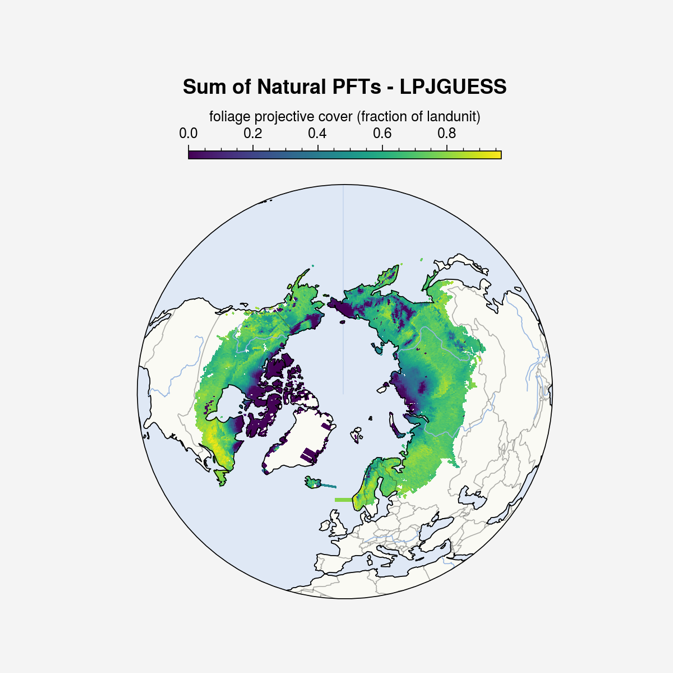


Figure 1: Basic differences between the two models: covered area, grid resolution, variable. For CLM5, just a selection of PFTs is shown (total sum would give 100% in all grid cells)