# Online Resource 3: Data processing, data sources, sensitivity, and agreement with observed SOC stocks.

## Data processing and sources

*Processing of spatially explicit data* — Raster maps were obtained from various sources using an equally spaced longitude-latitude raster but differing in resolution. Quantitative data (*e.g.*, temperature) was resampled to a halfdegree grid using averages. Qualitative data (e.g., land use) was resampled using the modal value. We used GRASS 6.4 (http://grass.osgeo.org) to process geographical data and R (R Development Core Team, 2011) for mapping and post-processing.

*Climate* — We used WorldClim 1.4 at 5 arc minutes resolution to obtain mean monthly temperatures and precipitation and to calculate annual degree days and the aridity index. WorldClim represents the climate for the period 1950–2000 (Hijmans et al., 2005 , http://www.worldclim.org). WorldClim was generated through interpolation from average monthly climate data from weather stations on a 30 arc second grid. Ensemble predictions for mean monthly temperature and precipitation anomalies as the medians of the output of 16 GCM for the A1B emission scenario 2070-2099 were obtained at 30 arc minutes ("50 km") resolution from ClimateWizard (http://www.climatewizard.org, accessed 2010-08-17, Adam and Lettenmaier, 2003; Meehl et al., 2007; Wood et al., 2004). We applied anomalies to climate data at 5 arc minutes resolution and then averaged to 30 arc minutes resolution. Regions of strong changes are shown in Figs. 1–2.

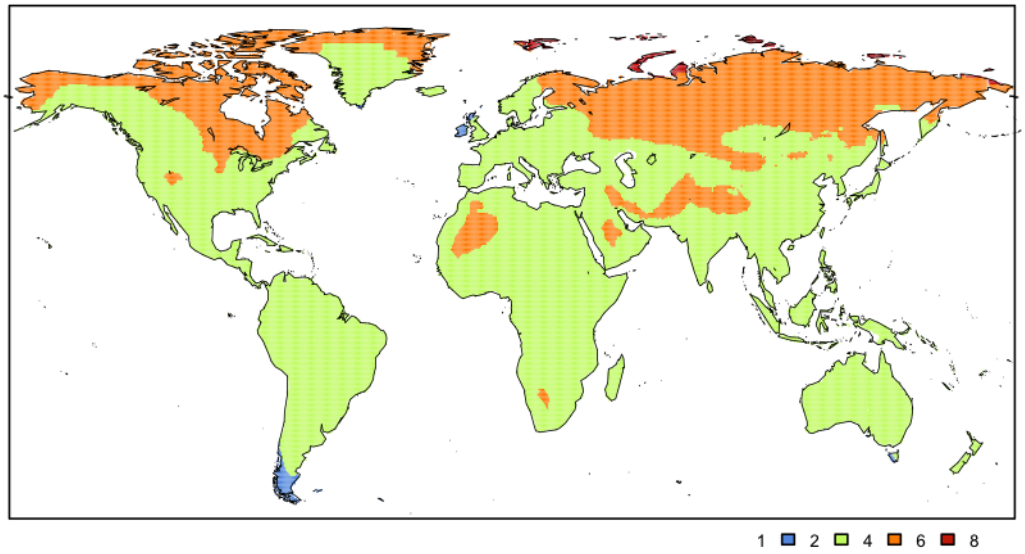


Fig. S3.1. Mean absolute changes in temperature (K) of each month of the year ($\sum(|\Delta T_M|)/12$).

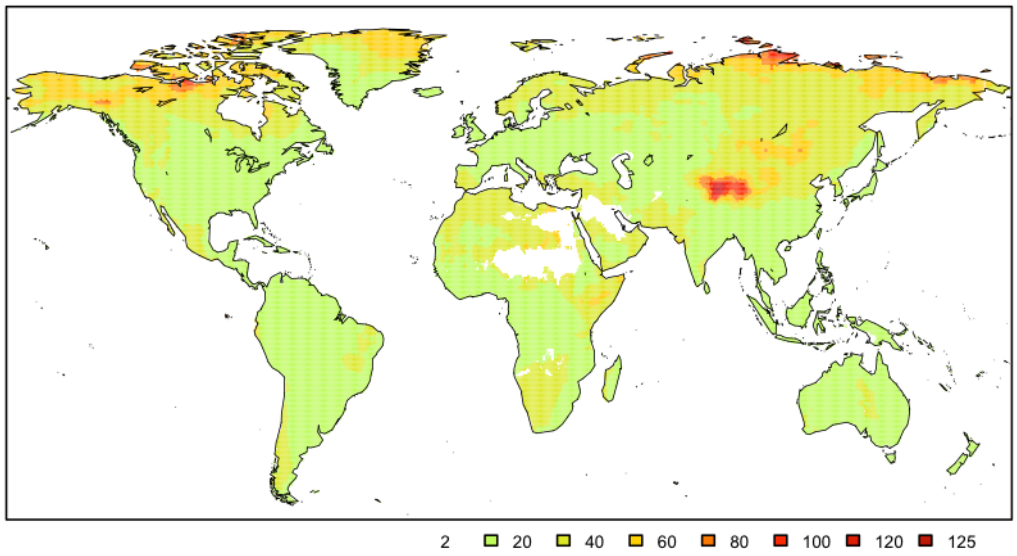


Fig. S3.2. Mean absolute changes in monthly precipitation (mm) of each month of the year ($\sum(|\Delta P_M/P_M|)/12$).

*Soil* — C content, bulk density, soil depth, CEC, and pH($H_2O$) were extracted from the Harmonized World Soil Database (HWSD) 1.1 (FAO et al., 2009) at 30 arc seconds resolution. Bulk density of organic soils and SOC were corrected as described in Köchy (Köchy et al., 2014). SOC stocks were calculated separately for 0-30 cm and 30-100 cm soil depth and summed to 1 m or a smaller depth indicated in the HWSD. C stocks for primary and secondary soil types within each soil mapping unit were weighted by their relative area. Similarly, we calculated a weighted CEC across soil layers and soil types for each soil mapping unit (Fig. S3.3). We averaged pH across soil types only for the top soil layer because most of the decomposition processes occur there. Presence of permafrost and constraints of oxygen availability were obtained from HWSD supplementary soil quality data 'SQ4' at 5 arc minutes resolution (Fischer et al., 2008). Oxygen constraints take into account soil type, soil texture, soil phases, and terrain slope.

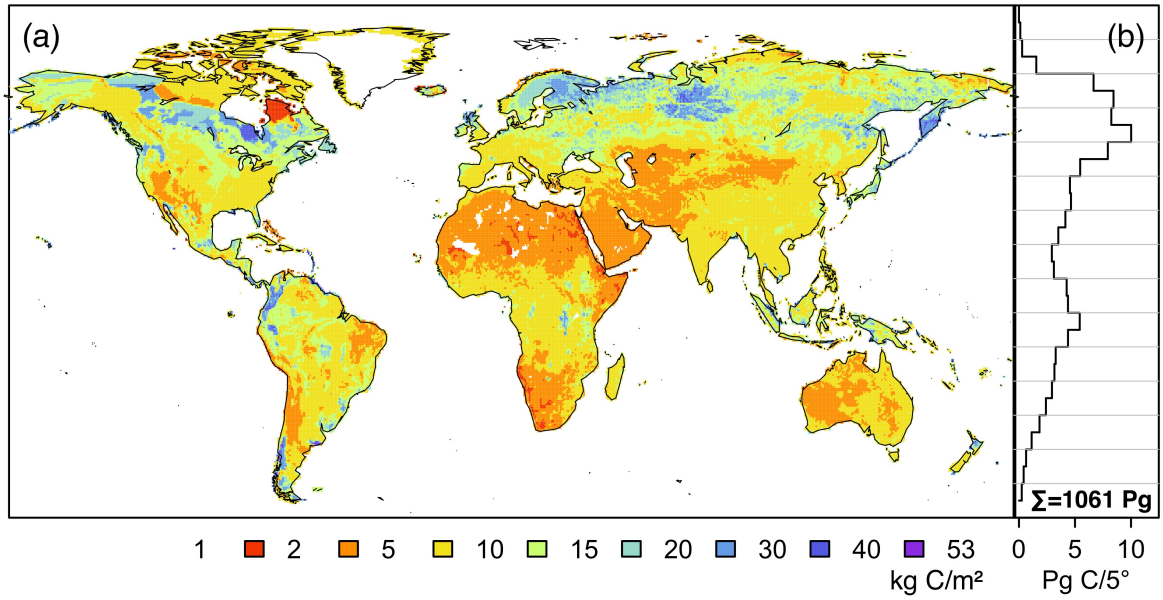


Fig. S3.3. Global stock (a) and mass (b, per 5° latitude) of organic carbon in the top 100 cm of the terrestrial soil calculated from HWSD v.1.1-modified (from Köchy et al., 2014).

*Vegetation cover* — Zonal (potential natural) vegetation (Fig. S3.4) and land use is based on corresponding products at 5 arc minutes resolution (Sterling and Ducharne, 2008). These were ultimately synthesized from output of the BIOME3 vegetation model (Haxeltime and Prentice, 1996), the IGBP DISCover land cover map referring to 1992 (Ramankutty and Foley, 1999), and additional sources. We reclassified the zonal vegetation types of Sterling & Ducharne to match the classes for biome transition of Gonzalez et al. (2010) (Online Resource 4, Table S4.1). The azonal vegetation classes of Sterling & Ducharne were aggregated to built-up, annual crops, and pasture. We set the azonal vegetation class to 'plantation' when the forestry map of Erb et al. (2007) indicated >50% land use by forestry or when the Global Land Cover Characterization map (version 2.0, resampled to 5 arc minutes, http://edcdaac.usgs.gov/glcc/glcc.html, representing conditions in 1992/1993) indicated tree crops (classes 95 and 96 of the Global Ecosystems legend). We set the azonal vegetation class to 'wetland' when the Global Lakes and Wetlands Database (level 3, resampled to 5 arc minutes, Lehner and Döll, 2004) indicated the dominance of swamp forest/flooded forest, coastal wetland, bog/fen/mire, or 50-100% wetland.

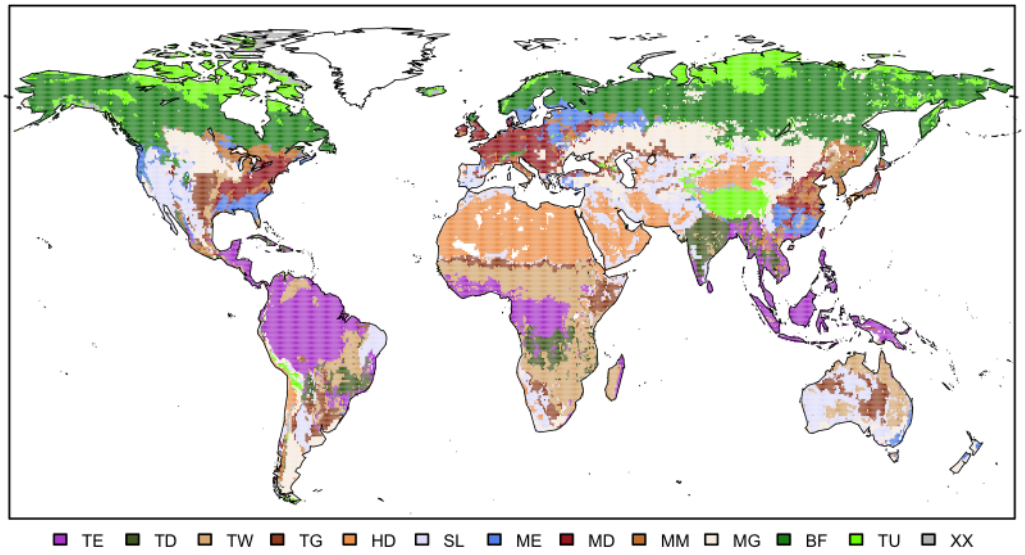


Fig. S3.4. Global distribution of vegetation zones used in this study. TE: tropical evergreen forest, TD: tropical deciduous forest, TW: tropical woodland, TG: tropical grassland, HD: hot deserts, SL: shrubland, ME: temperate evergreen forest, MD: temperate deciduous forest, MM: temperate mixed forest, MG: temperate grassland, BF: boreal forest, TU: tundra, XX: bare, ice, rocks.

*Net primary productivity* of the zonal and azonal vegetation for contemporary conditions is based on maps $NPP_0$, and $NPP_t$ of Haberl et al. (2007). $NPP_0$ represents the productivity of the zonal vegetation simulated by the dynamic global vegetation model LPJ for the climate of 1902–2002 on a 0.5° raster. $NPP_t$ represents NPP remaining in ecosystems after harvest calculated from average LPJ results for 1998–2002, resampled to 5 arcminutes, and modified according to harvest and resource use statistics from FAOStat (Haberl et al., 2007). When the azonal vegetation of a grid cell was 'built-up' (not included in Haberl et al., 2007), we set $NPP_t$ to zero.

For environmental conditions under the A1B emission scenario, we used simulations of $NPP_0$ and multiplied it with the harvest factor $NPP_t/NPP_0$. For the enhanced-NPP target scenario we used the $NPP_0$ simulated by the process-based LPJ dynamic global vegetation model (Sitch et al., 2003; Gerten et al., 2004) with the CCSM3 global climate model as a driver averaged across years 2075–2099 (Fig. S3.2a). We used LPJ because it was used by Haberl et al. (2007) for calculating $NPP_0$. We used CCSM3 because it was similar (by visual inspection of range and pattern) to the ensemble mean climate change predictions. For the limited-NPP target scenario we used the NCEAS model (Del Grosso et al., 2008) to calculate $NPP_0$ (Fig. S3.2b). The NCEAS model uses mean annual precipitation (MAP), mean annual mean temperature (MAT), and tree dominance of the vegetation as predictors of $NPP_0$ as follows:

$NPP_0 = 6116 \times [1 - \mathbf{exp}(-6.05\ 10^{-5} \times MAP)]$ for non-tree dominated systems and
**min**[$f$(MAP), $f$(MAT)] for tree-dominated systems with
$f(MAP = 0.551 \times MAP^{1.055}/\mathbf{exp}(0.000306 \times MAP)$ and
$f(MAT) = 2540/[1 + \exp(-4.77 \times 10^{-5} \times MAT)]$.
We capped MAP at 6000 mm to prevent the prediction of unrealistically low $NPP_0$ in perhumid regions and restricted the range of $NPP_0$ to 1.5 kg/m², the highest value on the scale of Haberl et al. (2007) for $NPP_0$.

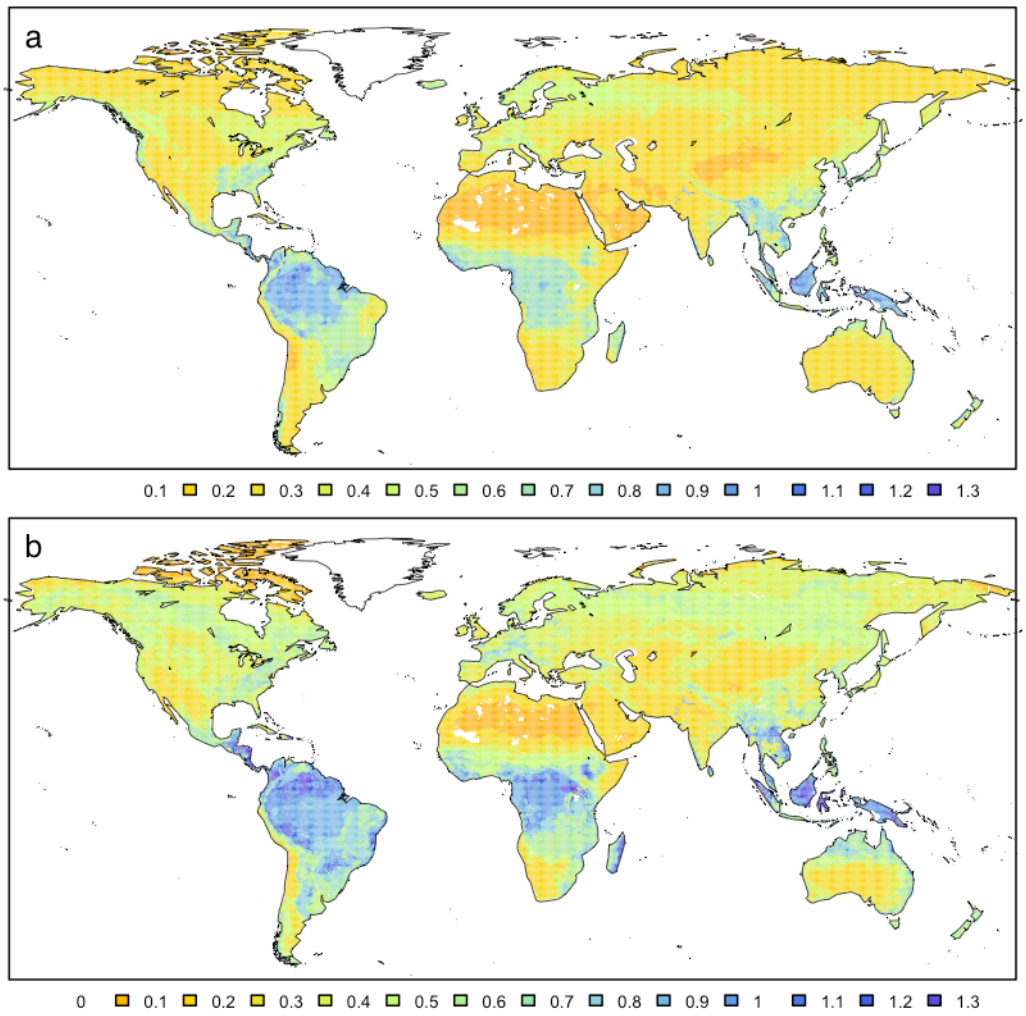


Fig. S3.5. Distribution of NPP after harvest ($NPP_t$, kg C/m$^2$) in a) the limited NPP and b) the enhanced NPP scenario.

## Soil model, summary equation

The total amount of C in the fast and slow pool after one year are

$C_{fast,t+1} = (C_{fast,t} + NPP_t)\cdot(1-fmf\cdot F_f)\cdot(1-to_{slow})$ and

$C_{slow,t+1} = C_{slow,t}\cdot(1-fmf\cdot F_s) + C_{fast,t+1}\cdot(1-fmf\cdot F_s)$.

The initial value of $C_{fast}$ is 0 and the initial value of $C_{slow}$ is the accessible fraction (*af*) of the initial total C stock ($C_0$) or $C_0 \cdot af$.

For *n* years, the equations can be expressed as

$$C_n = NPP \cdot \left( \frac{1-[(1-to_{slow})\cdot(1-fmf\cdot F_f)]^{n+1}}{1-(1-to_{slow})\cdot(1-fmf\cdot F_f)} - 1 \right)$$

$$+ \frac{NPP\cdot to_{slow}\cdot(1-fmf\cdot F_f)^{n+1}}{1-(1-to_{slow})\cdot(1-fmf\cdot F_f)} \cdot \left( \frac{1-\left[\frac{1}{1-fmf\cdot F_s}\right]^{n+1}}{1-\frac{1}{1-fmf\cdot F_s}} - \frac{1-\left[(1-to_{slow})\cdot\frac{1-fmf\cdot F_f}{1-fmf\cdot F_s}\right]^{n+1}}{1-(1-to_{slow})\cdot\frac{1-fmf\cdot F_f}{1-fmf\cdot F_s}} \right)$$

$$+ C_0 \cdot af \cdot (1-fmf\cdot F_s)^n.$$

## Calibration, further details

HWSD-SOC stocks > 10 kg/m² tended to be slightly underestimated by one class level (Fig. S3.3), whereas stocks < 2 kg/m$^2$ tended to be overestimated. 93% of all observed values were within 1 and 99% were within 2 standard deviations of the predicted value. We checked randomly selected points where the prediction was off by > 10 kg/m$^2$. The model overestimated HWSD-C stocks where HWSD indicated severe $O_2$-constraints but the pixels were not classified as wetlands. The model underestimated HWSD-SOC stocks in continental-climate crop regions due to a low harvest fraction and in dry shrublands. We note that our calibration procedure implicitly assumes that the SOC stock as calculated from HWSD is in equilibrium with $NPP_t$ and contingent on the land use derived from other maps. This assumption is not likely to be true for each grid cell.

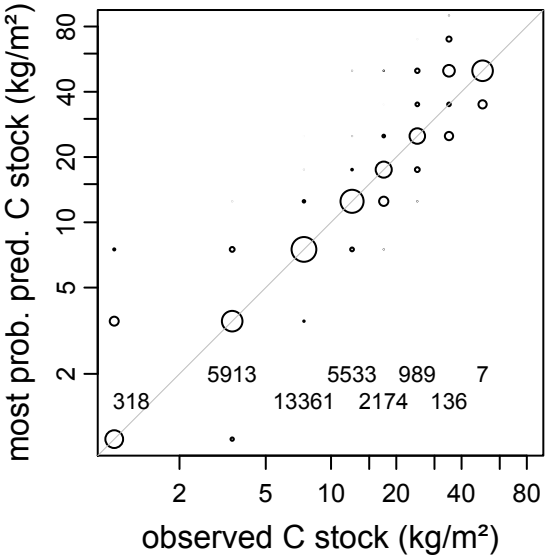


Fig. S3.6. Relative proportion (circle diameter) of the most probable predicted C stock classes with number of cases in each class of observed (HWSD) C stock in the validation set (50% of all points).

## Sensitivity

We express the sensitivity of the reference SOC stock as the percent reduction of its variance due to setting a class of each of the other variables to certain (100% certainty) (Pearl, 1991). We use the LPJ network variant for assessing sensitivity because of its greater climatic detail. The greatest sensitivity of the total reference SOC stock was to $NPP_t$ (36%) and $NPP_0$ (30%), followed by vegetation zone (20%) and HWSD-SOC (16%) and *fmf* (15%). Sensitivity to vegetation zone was high because it was

linked to many variables in the network to express regional subsets, which reduced the overall number of class combinations. Plant types (12%), land use (8%), pH (7%), and remaining NPP fraction (6%) had smaller effects on variance reduction. Sensitivity of the remaining variables, including degree days and $fmf_{climate}$ were <5%.

## References


Adam J. C., Lettenmaier D. P. (2003) Adjustment of global gridded precipitation for systematic bias. Journal of Geophysical Research 108, 1-14. doi: 10.1029/2002JD002499

Del Grosso S., Parton W., Stohlgren T., Zheng D., Bachelet D., Prince S., Hibbard K., Olson R. (2008) Global potential net primary production predicted from vegetation class, precipitation, and temperature. Ecology 89, 2117-2126. doi: 10.1890/07-0850.1

Erb K.-H., Gaube V., Krausmann F., Plutzar C., Bondeau A., Haberl H. (2007) A comprehensive global 5min resolution land-use dataset for the year 2000 consistent with national census data. Journal of Land Use Science 2, 191-224. doi: 10.1080/17474230701622981

FAO, IIASA, ISRIC, ISSCAS, JRC (Eds) (2009) Harmonized World Soil Database (version 1.1). FAO and IIASA, Rome, Italy, and Laxenburg, Austria,

Fischer G., Nachtergaele F., Prieler S., van Velthuizen H. T., Verelst L., Wiberg D. (2008) Global agro-ecological zones assessment for agriculture (GAEZ 2008). IIASA, Laxenburg, Austria and FAO, Rome, Italy.,

Gerten D., Schaphoff S., Haberlandt U., Lucht W., Sitch S. (2004) Terrestrial vegetation and water balance - hydrological evaluation of a dynamic global vegetation model. Journal of Hydrology 286, 249-270. doi: 10.1016/j.jhydrol.2003.09.029

Gonzalez P., Neilson R. P., Lenihan J. M., Drapek R. J. (2010) Global patterns in the vulnerability of ecosystems to vegetation shifts due to climate change. Global Ecology and Biogeography 19, 755-768. doi: 10.1111/j.1466-8238.2010.00558.x

Haberl H., Erb K. H., Krausmann F. *et al.* (2007) Quantifying and mapping the human appropriation of net primary production in earth's terrestrial ecosystems. Proceedings of the National Academy of Sciences of the United States of America 104, 12942-12947. doi: 10.1073/pnas.0704243104

Haxeltime A., Prentice C. I. (1996) BIOME3: An equilibrium terrestrial biosphere model based on ecophysiological constraints, resource availability, and competiion among plant functional types. Global Biogeochemical Cycles 10, 693-709. doi: 10.1029/96GB02344

Hijmans R. J., Cameron S. E., Parra J. L., Jones P. G., Jarvis A. (2005) Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology 25, 1965-1978. doi: 10.1002/joc.1276

Köchy M, Freibauer A, Hiederer R (2014, in preparation) Soil carbon mass and stocks for the tropics, permafrost regions, wetlands, and the globe. In: Grunewald S, Thompson J (eds) Global Digital Soil Mapping in a Changing World. Springer.

Lehner B., Döll P. (2004) Development and validation of a global database of lakes, reservoirs and wetlands. Journal of Hydrology 296, 1-22. doi: 10.1016/j.jhydrol.2004.03.028

Meehl G. A., Covey C., Delworth T., Latif M., McAvaney B., Mitchell J. F. B., Stouffer R. J., Taylor K. E. (2007) The WCRP CMIP3 multi-model dataset: A new era in climate change research. Bulletin of the American Meteorological Society 88, 1383-1394. doi: 10.1175/BAMS-88-9-1383

Pearl J. (1991) Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann, San Mateo, CA, U.S.A.

*R* Development Core Team. (2011) *R*: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Ramankutty N., Foley J. A. (1999) Estimating historical changes in land cover: An analysis of global croplands data. Global Biogeochemical Cycles 12, 667-685. doi: 10.1029/98GB02512

Sitch S., Smith B., Prentice I. C. *et al.* (2003) Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model. Global Change Biology 9, 161-185. doi: 10.1046/j.1365-2486.2003.00569.x

Sterling S., Ducharne A. (2008) Comprehensive data set of global land cover change for land surface model applications. Global Biogeochemical Cycles 22, GB3017. doi: 10.1029/2007GB002959

Wood A. W., Leung L. R., Sridhar V., Lettenmaier D. P. (2004) Hydrologic implications of dynamical and statistical approaches to downscaling model outputs. Climatic Change 62, 189-216. doi: 10.1023/B:CLIM.0000013685.99609.9e