SHORT COMMUNICATION

Representative soil profiles for the Harmonized World Soil Database at different spatial resolutions for agricultural modelling applications

Peter G. Jones
Waen Associates, Y Waen, Islawrdref, Dolgellau, Gwynedd, LL40 1TS, United Kingdom
E-mail: p.jones@cgiar.org
Tel +44 1341 423561

Philip K. Thornton*
CGIAR Program on Climate Change, Agriculture and Food Security (CCAFS), ILRI, PO Box 30709, Nairobi 00100, Kenya
E-mail: p.thornton@cgiar.org
Tel +44 131 667 1960

ABSTRACT

Agricultural modellers often need detailed soil profile data with which to run their models. We combine an extensive soil profile database with the Harmonized World Soil Database, a 30 arc-second raster database of soil information worldwide, and describe a statistical process to identify representative soil profiles for each of its 188 distinct soil types at different spatial resolutions. We then outline a method to cluster the soils in the Harmonized World Soil Database to produce soils maps at coarser resolution, and we describe derived global soils maps at spatial resolutions of 5 and 10 arc-minutes, which may be more practical for some large-scale modelling studies. The derived data files allow a user to select any point or area on land and then to access the set of soil profiles pertaining to the mapping unit selected, which are available in a format suitable for use in modelling applications. In situations where the user has little or no other information about the soils in the region of study, the methods described can be used to produce plausible soil profile information based on the most up-to-date global soils map currently available.

Keywords: WISE 3.1, soil profile, DSSAT, cluster, MarkSimGCM, CMIP5

* Corresponding author

1. Introduction

The availability of suitable input data continues to be a serious constraint to undertaking many applied research activities in the realm of agriculture, particularly in developing countries. Previously, we described a tool, MarkSimGCM, which allows the stochastic generation of daily weather data that are characteristic of current climatologies and, to some extent characteristic, of future climatologies (Jones and Thornton, 2013). These data can then be used to drive any agricultural model that requires daily (or otherwise aggregated) weather data. In addition to weather, another primary determinant of agricultural production is the soil. As for weather data, accessibility to and availability of appropriate soils data for agricultural modelling may pose serious constraints. Furthermore, the format in which soils data are available may constrain their widespread use, added to which is the problem of different soil classification schemes, which may or may not be translatable one to another.

Here we use the Harmonized World Soil Database (HWSD), the most up-to-date world soil map (FAO, 2012). It incorporates a data table of 48,148 soil profile descriptions related to the various soils associated with each mapping unit, at a spatial resolution of 30 arc-seconds (approximately 1 km at the equator). The data have been derived from a variety of sources, and are quite comprehensive. Unfortunately for many users, the data are grouped into broad topsoil and subsoil categories. For some purposes such as crop modelling, however, a soil profile is required with a full set of horizon data.

Below we describe an analysis that uses soil profile data from a large database and identifies the most statistically representative soil profile for each soil type in the HWSD. We then outline a method to cluster the soils in the HWSD to produce soils maps at coarser resolutions. We describe modifications to an existing tool, MarkSimGCM, that provide the user with soil profile data for any location, along with daily weather data for current and future climatologies. We have also updated MarkSimGCM with more recent climate model output; we describe this also. The weather and soils data can be used directly for a wide range of purposes. We briefly illustrate the use of the soils data set, and comment on its limitations.

2. Materials and methods

2.1 Using the WISE database to define representative soil profiles

The WISE databases contain soil profile data from many places around the world (Batjes, 2008; 2009). These data have been converted for use with the DSSAT (Decision Support System for Agrotechnology Transfer) crop modelling system (Jones et al., 2003; Gijsman et al., 2007; Chaves and Hoogenboom, 2014). From the most recent version, WISE 3.1, Chaves and Hoogenboom (2014) constructed a data file with 9618 soil profiles in DSSAT format. All soil profiles were classified according to both the original and the most recent legends of FAO's soil map of the world (FAO, 1974; 1988). From this set of soil profiles, one question of interest is, which profile is the most representative of each soil type? If we can define representative profiles for each soil type, we can assign appropriate profiles to each mapping unit.

We proceeded as follows. Each soil profile contains both character and numeric data (Table 1, from Chavez and Hoogenboom, 2014). Tier 1 and tier 2 variables were available for each horizon within the profile. The varying number of horizons made direct comparisons difficult between profiles. Three horizons were therefore chosen: the top, bottom and the one closest to the middle. Some soil types such as Lithosols had only one profile and some had only two. In the first case the algorithm triplicated the horizon and in the second case the second horizon was duplicated. This ensured consistency of the distance measure across profiles and simplified programming.

Taking all available numeric values and eliminating those with missing values and those that did not vary across the sample of profiles gave a minimum of 37 values for profile comparison; some soil types had more than 37. All variables were normalised by dividing by the range and the average of each set of profiles for each soil type was calculated. Then for each profile the Euclidean distance (in 37-dimensional space) from the average was calculated. The most representative profile was chosen as the profile with the minimum distance from the mean.

In most cases, the full binomial key to the soil type was available and so the set of profiles applies to a specific soil (e.g., Eutric Cambisol). In some cases the modifier is not given, just

the major soil type (e.g., Cambisol). In such cases all profiles classified as Cambisols were used for finding the typical profile. Typical profiles were found for both the FAO (1974) and the FAO (1988) legend keys.

Several soil types are not present in Chavez and Hoogenboom (2014) but do exist in the HWSD data (FAO, 2012): Folic Histosols, Luvic Gypsisols, Gypsic Kastanozems, Glosic Chernozems, Albic Lixisols, Urbic Anthrosols, Gelic Andosols, Aric Anthrosols, Stagnic Lixisols, Gelic Podzoluvisols, Gelic Planosols, and Andic Gleysols. Of these, the Gelic and Aric modifiers relate more to climate than to the actual soil profile, and are unlikely to be good arable land. The others are more problematic as they could potentially be used for cropping. It was decided to provide a reference profile for all of them for completeness, based on the representative profile of the appropriate major soil type.

Two further soils, Takyric Solonchaks and Takyric Yermosols, deserve mention. There are two profiles of the first and one of the second present in the Chavez and Hoogenboom (2014) dataset, but it was judged that these represented too few profiles from which to draw a representative profile. In the case of the Solonchaks the profiles were quite different, one being a heavy clay soil and the other a silty clay. Takyric soils are unlikely to be good agricultural soils and these have been assigned the representative profile of the major soil type.

The representative soil profile identifiers are stored in the file "HWSD_consolidated_class.txt", which contains 48,148 records and corresponds one-to-one with the HWSD database (see Table 2). The 188 reference profiles in DSSAT format are stored in the file "Consolidated.sol" (all distribution files are listed in Table 3).

2.2 Processing the HWSD to produce 10- and 5-minute grids

In section 2.1 above we describe how a DSSAT-formatted profile has been assigned to each of the soil types in each mapping unit of the HWSD (FAO, 2012). This is obviously the best possible representation, as the appropriate set of soil types corresponds exactly with the mapping unit. A spatial resolution of 30 arc-seconds may not be appropriate for some modelling projects, however, particularly if large areas are involved. In this section, we

describe the production of HWSD-based soil maps at coarser resolutions of 5 and 10 arc-minutes, which may be more acceptable in these cases.

We investigated the approach of simply taking the modal mapping unit from the 30 arc-second map in the larger pixel at 5 and 10 arc-minutes. We then mapped the percentage of each pixel represented by the modal pixel. Values less than 15% were common, especially in areas of the HWSD where the soil mapping was most detailed. We judged this approach to be inappropriate, so we proceeded as follows. First, we carried out a pixel-by-pixel inventory of the soils present with the share of each in the pixel from the 30 arc-second HWSD. These Mapping Units identities (MUID) in each pixel were listed in a file along with the percentage of the area covered by each one. The percentages were scaled to represent the percentages of actual soils (in HWSD they include shares for non-soil inclusions) and soils with shares of less than 1% of the pixel were discounted. This simplified and standardized the percentage scale for soils and allowed comparison for clustering purposes.

Next, we clustered soils using a simple leader algorithm without cluster average recalculation (Hartigan, 1975). The distance measure is calculated over an array of 188 values (the number of representative profiles in the consolidated profile list), which are the percentage shares of each soil present in the comparison. If a soil is present in only one of the 5-minute and the 10-minute MUID lists, then the contribution is the squared value of the share of the soil present. If the soil is present in both lists then the contribution is the squared difference of the share percentages. The squared measure was divided by the number of common soils. This conveniently yields a root mean square distance, which is zero if all soils are common with the same shares and very large if no soils are common. We used repeated trial distances to determine appropriate cut-offs. For both the 5-minute and 10-minute maps, we restricted the total number of clusters to slightly fewer than 32,768 or 2^{15}. The reason for this is to be able to express the index as a 2-byte (positive) integer. We include distribution files for two resolutions, 5 and 10 arc-minutes. Choice of resolution involves a trade-off between precision (similarity to the set of soil profiles in the original HWSD) and number of soil types in each pixel (which will affect simulation time).

The clustering procedure groups the pixels with the most similar combinations of soils and percentage covers by Euclidean distance of the soil percentages of common soils; they are not all exactly the same within a cluster. To harmonise these, we needed to simplify to one

representation per cluster. We did this by adding in all the soils from each pixel with the same key, averaging the percentage shares and eliminating soils with less than a 1% share of the resulting mapping unit key (rounded down). The final result is an index file for each resolution (Table 3). The number of lines in each file is three greater than the number of soil index keys because key 0 is water and key 1 is non-soil such as glacier, rock, and urban area; the first record of the file is a header record including details of the index. The actual soil mapping unit keys therefore run from 2 to 31488 for the 10 minute grid and 2 to 32368 for the 5 minute grid.

2.3 Updating MarkSimGCM to include data from CMIP5 climate models

In Jones and Thornton (2013) we described a web-based tool that runs with Google Earth (see www.google.co.uk/intl/en_uk/earth/). It utilises a generalised downscaling and data generation method that takes the outputs of a General Circulation Model and allows the stochastic generation of daily weather data that are to some extent characteristic of future climatologies, using a weather generator, MarkSim (Jones and Thornton, 1993). The first version of this tool used data from six of the climate models and scenario runs carried out for the 2007 Fourth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC, 2007). These data were from the Coupled Model Intercomparison Project phase 3 (CMIP3) of the World Climate Research Programme (WCRP). The tool has now been updated to include data from 17 climate models from the next CMIP phase (CMIP5) that were used in the Fifth Assessment Report in 2014 (IPCC, 2013). We obtained yearly GCM data interpolated by bilinear interpolation from the original pixel sizes to 1-degree pixels (the models used are shown in Table 4). Historical data ran from 1961 to 2005, and projection data ran from 2006 to 2099. We calculated monthly means for rainfall, maximum and minimum air temperature, all expressed as anomalies from the historical mean, and then bias corrected to WorldClim v1.3 (Hjsmans et al., 2005). The GCM data provide annual deviations for the years 2006 to 2099. We fitted fourth-order polynomials for each variable to every pixel through time. The methods used are otherwise as described in Jones and Thornton (2013).

3. Results

3.1 Soils data

Table 5 illustrates the soils clustering process; it shows a cluster located in southern Paraguay in an area of broad scale cultivation in Itapica to the north-west of the town of General Delgado. This example shows the grouping of seven different soil sequences when some of the cluster distances are quite marginal. The first pixel at (737, 701) is the type pixel (i.e., it defines the cluster centroid and is located at 57.2°W, 26.8°S). All the other pixels are closest to this one in 'soil %' space not necessarily in geographic terms, although in this case they are close but not contiguous. Ferric Acrisols, Haplic Ferralsols, and Eutric Gleysols constituted 90% of the area and the minor soils, Dystric Planosols, Eutric Regosols, Eutric Podzoluvisols and Ferric Lixisols, a further 7%. The soils eliminated from the mapping unit key are shown in bold italics and each constitutes less than 1% of the mapping unit area.

Results of the soil clustering process at the two resolutions are shown in Table 6, which shows statistics of fit of the derived soil clusters to the 30 arc-second HWSD soils map for 20 randomly-selected 1-degree latitude-longitude pixels. We calculated various statistics to estimate the information loss from the high resolution to the lower ones; the Euclidean distance of the soil proportions to judge the largest deviation, the standard deviation of the soil proportions and the proportion of soils included in error due to the clustering process. As expected, the higher resolution (5 arc-minute) clusters provide a somewhat better fit than the lower resolution clusters, as measured by the smaller mean Euclidean distance to the cluster centroid, the smaller root mean square (RMS) % deviation, and the smaller residual share of soils not in the original soils map, but none of these differences were significant. As noted above, choice of resolution involves a trade-off between somewhat better-fitting clusters and speed of simulation. This is illustrated in Figure 1, for a grid cell of size 1 degree latitude by 1 degree longitude located in southern Ethiopia (pixel number 9 in Table 6). The percentage of each pixel classified as Haplic Calcisol is mapped at three resolutions: 30 arc-seconds (the native resolution of the HWSD and thus the most detailed), 5 arc-minutes, and 10 arc-minutes.

Details of the files associated with the soil keys and data set are listed in Table 3. It should be noted that the percentage shares in the index key files are percentages of the soil cover in the index unit. This is not the same as the percentages of the mapping unit in the original HWSD. The user thus needs the percentage cover image along with the key image. This is a direct result of the need for a standardized measure of soil percentage share in order to cluster the soil sequences. It may also be noted that some pixels have a percentage less than 100%.

There are various reasons for this, including the presence of sea in coastal pixels, and of rock and urban areas in other pixels. While the structure of the key index file is identical for both 10-minute and 5-minute grids, there is no correspondence between the mapping keys, so they should not be mixed up.

FORTRAN modules were developed to output soil profile data as described in section 2, along with modifications to the graphical user interface in Google Earth. The revised MarkSimGCM is freely available at http://gisweb.ciat.cgiar.org/MarkSimGCM/. Soil profile data in DSSAT format are produced for the location selected, one file per soil profile, with an indication of the percentage cover of the pixel in each soil type.

3.2 Weather data

Weather data generation operates in a similar way to that described in Jones and Thornton (2013): the user may choose any combination, and any number, of the 17 climate models for a time-slice centred on any year between 2010 and 2099, for any one of four Representative Concentration Pathways (RCPs). The RCPs are four greenhouse gas concentration trajectories adopted by the IPCC (2013), covering a range (low to high) of emission scenarios. One new feature of the tool is the ability to plot a climate diagram for the location and model(s) selected, giving the user a quick overview of the current (or possible future) climate at the point of interest. As before, there is an option to generate daily data that are representative of current conditions as in WorldClim; in this way, MarkSimGCM operates as an updated version of the CD-based release of MarkSim (Jones et al., 2002). Output is produced in two formats: as annual charts of daily rainfall, maximum and minimum air temperatures and solar radiation; and as annual data files that are fully compatible with the DSSAT crop modelling suite (Jones et al., 2003).

3.3 Summary of modifications to MarkSimGCM

To clarify the relationship between the version of MarkSimGCM described in Jones and Thornton (2013) and the version discussed here, Figure 2 shows the linkages between the various elements and inputs. The Google Earth-based version of MarkSimGCM itself now produces DSSAT soil files as well as weather files. An updated stand-alone version of MarkSIMGCM with the CMIP5 climate model data is available at the website ccafs-

climate.org, along with documentation, and this version is available for users who want to generate weather data for multiple locations or grids. Alternatively, the user can develop his or her own scripts to utilise the MarkSimGCM executable as needed. In the same way, DSSAT soil profiles for multiple locations and at different resolutions can be produced in stand-alone mode, again with the use of appropriate scripts or programming languages.

4. Discussion and conclusions

A considerable amount of work continues to be undertaken on gridded simulations of the impacts of climate change on agricultural systems, along with evaluations of different options that can help households adapt (see, for example, Elliott et al., 2014). At the same time, there are strong indications that climate change impacts on agricultural systems, particularly in developing countries, are being under-estimated (Hertel and Lobell, 2014; Thornton et al., 2014). The need for gridded simulations is not going to decline any time soon. There are several key improvements in input data sources that are needed, including land-use and agricultural management information as well as soils information. The additions to the MarkSim GCM tool described here in relation to soil data provision are a small contribution towards this, by making the most up-to-date global soils database more accessible to researchers wanting to run their own simulations.

In relation to the updating of the climate models that MarkSim GCM now uses, we reiterate the downscaling issues which we highlighted in Jones and Thornton (2013): in particular, there is considerable lack of agreement between the projections of different climate models in some regions, and understanding is still limited of what the local-level impacts of climate change may be, and the adequacy of different downscaling techniques is still to be determined. Nevertheless, MarkSimGCM can provide weather data for possible future climatologies that agricultural impact modellers can use with care.

Acknowledgements

Support for this work was provided by the CGIAR program on Climate Change, Agriculture and Food Security (CCAFS) from the CGIAR Fund, AusAid, Danish International Development Agency, Environment Canada, Instituto de Investigação Científica Tropical, IrishAid, Netherlands Ministry of Foreign Affairs, Swiss Agency for Development and Cooperation, Government of Russia, UK Aid, and the European Union, with technical support from the International Fund for Agricultural Development. We are very grateful to Oliver Brown for assistance with the GCM data, and to Gerrit Hoogenboom and Bernardo Chaves for their work with the WISE soils data.

References

Batjes, N.H., 2008. ISRIC-WISE Harmonized Global Soil Profile Database (Ver. 3.1) Report 2008/02, ISRIC-World Soil Information, Wageningen. Online at www.isric.org/Webdocs /ISRIC_Report_2008_02.pdf

Batjes, N.H., 2009. Harmonized soil profile data for applications at global and continental scales: updates to the WISE database. Soil Use and Management 25, 124-127.

Chaves, B., Hoogenboom, G., 2014. Strengthening soil databases for climate change and food security modeling applications. CCAFS Research Report. Copenhagen, Denmark: CGIAR Research Program on Climate Change, Agriculture and Food Security (CCAFS).

Collier, M.A., Jeffrey, S.J., Rotstayn, L.D., Wong, K.K-H., Dravitzki, S.M.,Moeseneder, C., Hamalainen, C., Syktus, J.I., Suppiah, R., Antony, J., El Zein, A., Atif, M., 2011. The CSIROMk3.6.0 Atmosphere-Ocean GCM: participation in CMIP5 and data publication. MODSIM 2011, Perth, 12–16 December 2011.

Collins, W.J., Bellouin, N., Doutriaux-Boucher, M., Gedney, N., Halloran, P., Hinton, T., Hughes, J., Jones, C.D., Joshi, M., Liddicoat, S., Martin, G., O'Connor, F., Rae, J., Senior, C., Sitch, S., Totterdell, I., Wiltshire, A., Woodward, S., 2011. Development and evaluation of an Earth-System model-HadGEM2. Geoscientific Model Development 4 (4), 1051–1075.

Donner, L.J., and 40 others, 2011. The dynamical core, physical parameterizations, and basic simulation characteristics of the atmospheric component AM3 of the GFDL Global Coupled Model CM3. Journal of Climate 24 (13), 3484-3519.

Dufresne, J.L., and 60 others, 2013. Climate change projections using the IPSL-CM5 Earth System Model: from CMIP3 to CMIP5. Climate Dynamics 40 (9-10), 2123-2165.

Dunne, J.P., John, J.G., Adcroft, A.J., Griffies, S.M., Hallberg, R.W., Shevliakova, E., Stouffer, R.J., Cooke, W., Dunne, K.A., Harrison, M.J., Krasting, J.P., Malyshev, S.L., Milly, P.C.D., Phillipps, P.J., Sentman, L.T., Samuels, B.L., Spelman, M.J., Winton, M., Wittenberg, A.T., Zadeh N., 2012. GFDL's ESM2 global coupled climate–carbon earth system models.

Part I: physical formulation and baseline simulation characteristics. Journal of Climate 25, 6646–6665.

Elliott, J. and 26 others, 2014. Constraints and potentials of future irrigation water availability on agricultural production under climate change. Proceedings of the National Academy of Sciences 111 (9), 3239-3244.

FAO / UNESCO, 1974. FAO-UNESCO Soil Map of the World. Volume 1. Legend. UNESCO, Paris.

FAO / UNESCO / ISRIC, 1988. FAO-UNESCO Soil Map of the World. Revised Legend. World Soil Resources Report 60, FAO, Rome.

FAO / IIASA / ISRIC / ISSCAS / JRC, 2012. Harmonized World Soil Database (version 1.2). FAO, Rome, Italy and IIASA, Laxenburg, Austria.

Gijsman, A.J., Thornton, P.K., Hoogenboom, G., 2007. Using the WISE database to parameterize soil inputs for crop simulation models. Computers and Electronics in Agriculture 56, 85-100.

Hartigan, J.A., 1975. Clustering Algorithms. John Wiley & Sons, New York. Pp 74-83.

Hertel, T.W., Lobell, D.B., 2014. Agricultural adaptation to climate change in rich and poor countries: Current modeling practice and potential for empirical contributions. Energy Economics 46, 562-575.

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.

IPCC, 2007. Summary for Policymakers. In: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Solomon, S., D. Qin, M. Manning, Z. Chen, M.

Marquis, K.B. Averyt, M.Tignor and H.L. Miller (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

IPCC, 2013. Summary for Policymakers. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Jones, J.W ., Hoogenboom, G., Porter, C. H., Boote, K.J ., Batchelor, W.D ., Hunt, L.A., Wilkens P.W., Singh, U., Gisjman A.J., Ritchie, J.T ., 2003. The DSSAT cropping system model. European Journal of Agronomy, 18, 235-265.

Jones, P.G., Thornton, P.K., 1993. A rainfall generator for agricultural applications in the tropics. Agricultural and Forest Meteorology 63, 1-19.

Jones, P.G., Thornton, P.K., 2013. Generating downscaled weather data from a suite of climate models for agricultural modelling applications. Agricultural Systems 114, 1-5.

Jones, P.G., Thornton, P.K., Diaz, W., Wilkens, P.W., 2002. MarkSim: A Computer Tool that Generates Simulated Weather Data for Crop Modeling and Risk Assessment. CD-ROM Series, with manual. Centro Internacional de Agricultura Tropical (CIAT), Cali, Colombia.

Kirkevag, A., Iversen, T., Seland, O., Debernard, J.B., Storelvmo, T., Kristjansson, J.E. , 2008. Aerosol-cloud-climate interactions in the climate model CAM-Oslo. Tellus A 60(3), 492–512.

Schmidt, G.A., and 35 others., 2006. Present day atmospheric simulations using GISS ModelE: comparison to in-situ, satellite and reanalysis data. Journal of Climate 19, 153-192.

Seland, O., Iversen, T., Kirkevag, A., Storelvmo, T., 2008. Aerosol-climate interactions in the CAM-Oslo atmospheric GCM and investigation of associated basic shortcomings. Tellus A 60(3), 459–491.

Song, Z., Qiao, F., Song, Y., 2012. Response of the equatorial basin-wide SST to wave mixing in a climate model: an amendment to tropical bias, Journal of Geophysical Research 117, C00J26.

Thornton, P.K., Ericksen, P.J., Herrero, M., Challinor, A. J., 2014. Climate variability and vulnerability to climate change: a review. Global Change Biology 20 (11), 3313-3328.

Watanabe, M., Suzuki, T., O'ishi, R., Komuro, Y., Watanabe, S., Emori, S., Takemura, T., Chikira, M., Ogura, T., Sekiguchi, M., Takata, K., Yamazaki, D., Yokohata, T., Nozawa, T., Hasumi, H., Tatebe, H., Kimoto, M., 2010. Improved climate simulation by MIROC5: mean states, variability, and climate sensitivity. Journal of Climate 23, 6312–6335.

Watanabe, S., Hajima, T., Sudo, K., Nagashima, T., Takemura, T., Okajima, H., Nozawa, T., Kawase, H., Abe, M., Yokohata, T., Ise, T., Sato, H., Kato, E., Takata, K., Emori, S., Kawamiya, M., 2011. MIROC-ESM2010: model description and basic results of CMIP5-20c3m experiments. Geoscientific Model Development 4 (4), 845–872.

Wu, T., 2012. A mass-flux cumulus parameterization scheme for largescale models: description and test with observations. Climate Dynamics 38, 725–744.

Yukimoto, S. 2012. A new global climate model of Meteorological Research Institute: MRI-CGCM3 – Model description and basic performance. Journal of the Meteorological Society of Japan 90a, 23–64.

444 Table 1. Soil data input for a daily time step crop simulation model (Chaves and
445 Hoogenboom, 2014)
446
447
448
449

Characteristic	Definition	Units
	General Data	
SLTX	Texture code of surface layer	
SLDP	Soil depth	cm
SLDESCRIP	Soil description or local classification	
COUNTRY	Country name	
LAT*	Latitude	
LONG*	Longitude	
SCSC FAMILY	Soil Class	
	Profile Data	
SCOM*	Soil color (Munsell color system)	
SALB*	Albedo	
SLU1*	Evaporation limit	cm
SLDR*	Drainage rate	Fraction day^{-1}
SLRO*	Runoff curve number	
SLNF	Mineralization factor	0-1 scale
SLPF	Soil fertility factor	0-1 scale
SMHB	pH in buffer determination method	
SMPX	Extractable phosphorus determination code	
SMKE	Potassium determination code	
	First tier	
SLB	Depth until base of layer	cm
SLMH	Master horizon	
SLLL*	Lower limit of plant extractable soil water	cm^3cm^{-3}
SDUL*	Drained upper limit	cm^3cm^{-3}
SSAT*	Saturated upper limit	cm^3cm^{-3}
SRGF*	Root growth factor	0-1 scale
SSKS*	Saturated hydraulic conductivity	cm h^{-1}
SBDM*	Bulk density (moist)	g cm^{-3}
SLOC	Soil organic carbon concentration	%
SLCL	Clay (<0.002 mm)	%
SLSI	Silt (0.002 to 0.05 mm)	%
SLCF	Coarse fraction (>2mm)	%
SLNI	Total nitrogen concentration	%
SLHW	pH in water	
SLHB	pH in buffer	
SCEC	Soil cation exchange capacity	Cmol(+)kg^{-1}
SADC	Soil adsorption coefficient (anion exchange cap.)	0-1 scale

	Second tier	
SLPX	Extractable soil phosphorus concentration	$mgkg^{-1}$
SLPT	Total soil phosphorus concentration	$mgkg^{-1}$
SLPO	Soil organic phosphorus concentration	$mgkg^{-1}$
CACO3	Soil $CaCO_3$ concentration	%
SLAL	Soil aluminum concentration	$mgkg^{-1}$
SLFE	Soil iron concentration	$mgkg^{-1}$
SLMN	Soil manganese concentration	$mgkg^{-1}$
SLBS	Soil base saturation	%
SLPA	Soil phosphorus isotherm A	$mmol\ kg^{-1}$
SLPB	Soil phosphorus isotherm B	$mmol\ kg^{-1}$
SLKE	Exchangeable potassium soil concentration	$cmol(+)\ kg^{-1}$
SLMG	Exchangeable magnesium concentration	$cmol(+)\ kg^{-1}$
SLNA	Exchangeable sodium concentration	$cmol(+)\ kg^{-1}$
SLSU	Soil sulfur concentration	$cmol(+)\ kg^{-1}$
SLEC	Soil electric conductivity	dSm^{-1}
SLCA	Soil calcium concentration	$cmol(+)\ kg^{-1}$

* Calculated variables

Table 2. Contents of file "HWSD_consolidated_class.txt". The file contains 48,148 records and corresponds one-to-one with the HWSD database (FAO, 2012).

Variable	Meaning
ID	HWSD record ID
MU_GLOBAL	HWSD mapping unit number
SHARE	HWSD share % of the mapping unit for this profile
SEQ	HWSD sequence number of profile in mapping unit
SU_SYM74	HWSD FAO soil symbol 1974
SU_SYM90	HWSD FAO soil symbol 1988
WISE_PROFILE	10-character WISE profile identifier
WISE_KEY	Numeric key to wise profile
SOIL_NAME	Soil name, according to either FAO (1974) or FAO (1988)
FLAG	Soil profile origin: 0 Not a soil. This includes water bodies, rock outcrops etc. 1 Missing soil type with the profile of the major soil class provided. 2 Profile obtained from the complete group of profile for the major soil type. 3 Profile obtained from the set of appropriate profiles,

480 Table 3. List of files associated with the HWSD soil profiles

481

Name	Contents
HWSD_consolidated_class.txt	See Table 2. File contains 48,148 records in a 1:1 correspondence the HWSD database
Consolidated.sol	188 profiles in key order in DSSAT profile format
10-minute files	
HWSD_10min_keys.txt	Contains the key index in 31487 records. Format: key id, number of soils, and soil key, share percentage for each
HWSD_10min_keys.rst	The key image, use for the key index. IDRISI[1] format, binary
HWSD_10min_keys.rdc	IDRISI[1] document file for the above image file
HWSD_10min_percent_soil.rst	Contains the percentage of the pixel covered by soil. Soil areas may sum to a few percent less than indicated in this image. IDRISI[1] format, binary
HWSD_10min_percent_soil.rdc	IDRISI[1] document file for the above image file
5-minute files	
HWSD_5min_keys.txt	Contains the key index in 32367 records. Format: key id, number of soils, and soil key, share percentage for each
HWSD_5min_keys.rst	The key image, use for the long key index. IDRISI[1] format, binary
HWSD_5min_keys.rdc	IDRISI[1] document file for the above image file
HWSD_5min_percent_soil.rst	Contains the percentage of the pixel covered by soil. Soil areas may sum to a few percent less than indicated in this image. IDRISI[1] format, binary
HWSD_5min_percent_soil.rdc	IDRISI[1] document file for the above image file

482
483

484 1. IDRISI GIS software, http://www.clarklabs.org/

485

486

487

488

489

490

491

492

493

Table 4. Atmosphere-Ocean General Circulation Models from CMIP5 in the updated version of MarkSimGCM.

	Model	Institution	Resolution, Lat x Long °	Reference
1	BCC-CSM 1.1	Beijing Climate Center, China Meteorological Administration	2.8125 x 2.8125	Wu (2012)
2	BCC-CSM 1.1(m)	Beijing Climate Center, China Meteorological Administration	2.8125 x 2.8125	Wu (2012)
3	CSIRO-Mk3.6.0	Commonwealth Scientific and Industrial Research Organisation and the Queensland Climate Change Centre of Excellence	1.875 x 1.875	Collier et al. (2011)
4	FIO-ESM	The First Institute of Oceanography, SOA, China	2.812 x 2.812	Song et al. (2012)
5	GFDL-CM3	Geophysical Fluid Dynamics Laboratory	2.0 x 2.5	Donner et al. (2011)
6	GFDL-ESM2G	Geophysical Fluid Dynamics Laboratory	2.0 x 2.5	Dunne et al. (2012)
7	GFDL-ESM2M	Geophysical Fluid Dynamics Laboratory	2.0 x 2.5	Dunne et al. (2012)
8	GISS-E2-H	NASA Goddard Institute for Space Studies	2.0 x 2.5	Schmidt et al. (2006)
9	GISS-E2-R	NASA Goddard Institute for Space Studies	2.0 x 2.5	Schmidt et al. (2006)
10	HadGEM2-ES	Met Office Hadley Centre	1.2414 x 1.875	Collins et al. (2011)
11	IPSL-CM5A-LR	Institut Pierre-Simon Laplace	1.875 x 3.75	Dufresne et al. (2013)
12	IPSL-CM5A-MR	Institut Pierre-Simon Laplace	1.2587 x 2.5	Dufresne et al. (2013)
13	MIROC-ESM	Atmosphere and Ocean Research Institute (University of Tokyo), National Institute for Environmental Studies, Japan Agency for Marine-Earth Science and Technology	2.8125 x 2.8125	Watanabe et al. (2011)
14	MIROC-ESM-CHEM	Atmosphere and Ocean Research Institute (University of Tokyo), National Institute for Environmental Studies, Japan Agency for Marine-Earth Science and Technology	2.8125 x 2.8125	Watanabe et al. (2011)
15	MIROC5	Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute (University of Tokyo), National Institute for Environmental Studies	1.4063 x 1.4063	Watanabe et al. (2010)
16	MRI-CGCM3	Meteorological Research Institute	1.125 x 1.125	Yukimoto (2012)
17	NorESM1-M	Norwegian Climate Centre	1.875 x 2.5	Kirkevag et al. (2008); Seland et al. (2008)

Table 5. A sample cluster of pixels constituting mapping unit 549 in the 10-minute derived HWSD map.

Col	Row	Dist	nsoil	key1	%	key2	%	key3	%	key3	%	key4	%	key5	%	key6	%
737	701	0.00	6	3	48.9	63	20.4	78	19.6	140	5.5	123	2.8	161	2.8		
739	703	1.67	6	3	45.8	78	20.9	63	19.1	140	7.1	123	3.6	161	3.6		
738	703	3.26	6	3	55.0	63	22.9	78	17.1	140	2.5	123	1.2	161	1.2		
738	701	3.87	5	3	52.0	63	21.7	78	13.0	140	7.9	161	5.3				
745	702	7.23	7	3	51.4	63	19.9	78	13.2	122	4.9	***125***	4.3	99	3.9	***127***	2.4
744	703	7.23	6	3	52.3	63	23.7	78	13.1	***125***	5.7	***127***	3.2	99	1.9		
740	703	7.57	5	3	58.0	63	24.2	78	14.5	***22***	1.5	***29***	0.8				
739	702	8.33	4	3	59.9	63	24.9	78	15.0	***22***	0.1						
738	702	8.40	3	3	60.0	63	25.0	78	15.0								
737	702	8.60	6	3	32.9	78	26.3	63	13.7	140	13.6	123	6.8	161	6.8		
744	702	9.00	4	3	57.1	63	17.8	78	15.0	122	10.1						

Col, column number in the image; Row, row number in the image; Dist, cluster centroid distance; nsoil, number of soils in the pixel; key, soil key ID; %, percentage coverage in the pixel. Bold italic keys show soils whose coverage is <1% of the mapping unit and therefore omitted from the cluster.

Key	%	Profile	FAO soil type
3	52.1	WIACCI0016	Ferric Acrisol
63	21.2	WIFRBR0137	Haplic Ferralsol
78	16.6	WIGLNO0007	Eutric Gleysol
140	3.4	WIPLBR0613	Dystric Planosol
161	1.3	WIRGBR0433	Eutric Regosol
122	1.4	WIPDBY0030	Eutric Podzoluvisol
123	1.3	WILXBJ0618	Ferric Lixisol

Table 6. Statistics of fit of derived soil clusters at 10 (10') and 5 (5') arc-minute resolutions to the 30 arc-second HWSD soils map for 20 randomly-selected 1-degree latitude-longitude quadrats.

#	Latitude	Longitude	Euclidean Distance		RMS % Deviation		Residual Share[1]	
			10'	5'	10'	5'	10'	5'
1	31.302	-5.236	2.677	3.245	0.947	1.147	0.876	0.130
2	-23.036	-60.027	0.895	0.841	0.248	0.233	0.116	0.066
3	43.883	-99.520	1.225	0.695	0.707	0.402	0.481	0.374
4	16.720	-97.537	2.395	0.954	1.071	0.426	0.746	0.772
5	4.594	-75.532	1.343	1.175	0.425	0.372	0.542	0.517
6	-28.694	-70.401	0.459	0.280	0.132	0.081	0.085	0.165
7	68.089	-93.575	1.030	0.764	0.595	0.441	0	0
8	7.190	-65.703	0.947	1.861	0.424	0.832	0.601	1.124
9	5.043	42.606	1.741	0.641	0.658	0.242	1.149	0.654
10	-25.341	25.130	1.397	2.377	0.285	0.485	1.342	0.470
11	55.411	58.623	1.548	1.217	0.489	0.385	0.808	0.666
12	-21.527	-48.630	0.979	0.829	0.399	0.338	0.915	0.781
13	-27.266	151.235	2.403	0.660	0.566	0.156	1.714	1.399
14	60.092	108.564	6.219	2.319	2.351	0.876	0.115	0.514
15	-24.433	-58.537	0.747	0.498	0.264	0.176	0.058	0.022
16	-30.776	-68.087	1.115	0.733	0.372	0.244	0.086	0
17	-19.437	-57.295	0.935	0.395	0.296	0.125	0	0.108
18	42.665	104.869	0.504	1.639	0.206	0.669	0.481	1.118
19	-17.601	31.279	2.306	1.031	0.640	0.286	0.728	0.325
20	51.489	42.559	0.872	1.026	0.356	0.419	0.229	0.586
		Mean	1.586	1.159	0.5715	0.4167	0.5536	0.4895
		Variance	1.6110	0.5819	0.2336	0.0752	0.2362	0.1661
		Std deviation	1.2693	0.7628	0.4833	0.2743	0.4860	0.4076

1: Residual share of soils not in the 30 arc-second HWSD map

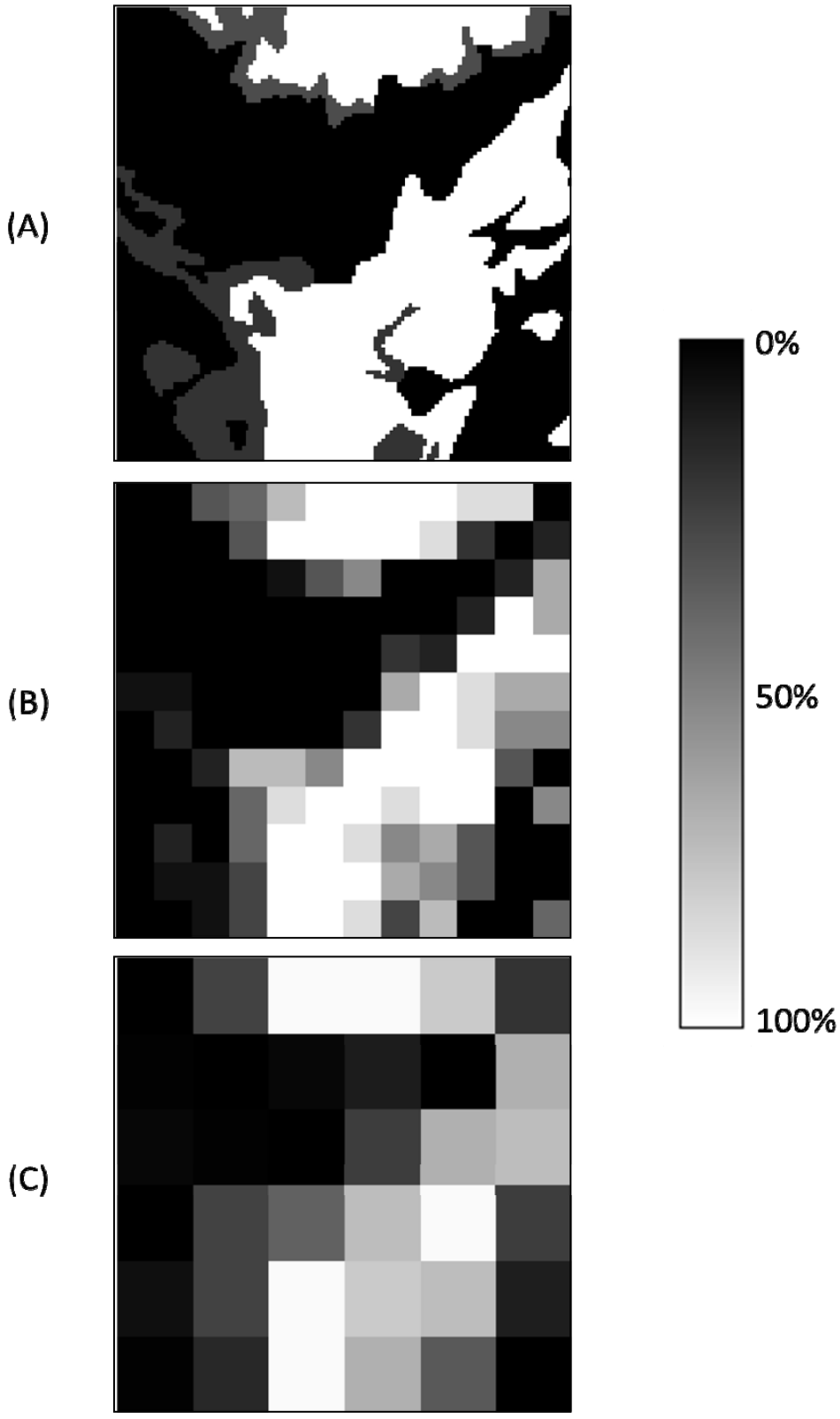

Figure 1. Percentage of Haplic Calcisols in pixels within a one-degree grid cell at three resolutions: (A) 30 arc-seconds, (B) 5 arc-minutes, (C) 10 arc-minutes. The lower left-hand corner of the grid cell is located at 4.33° N, 42.33° E.

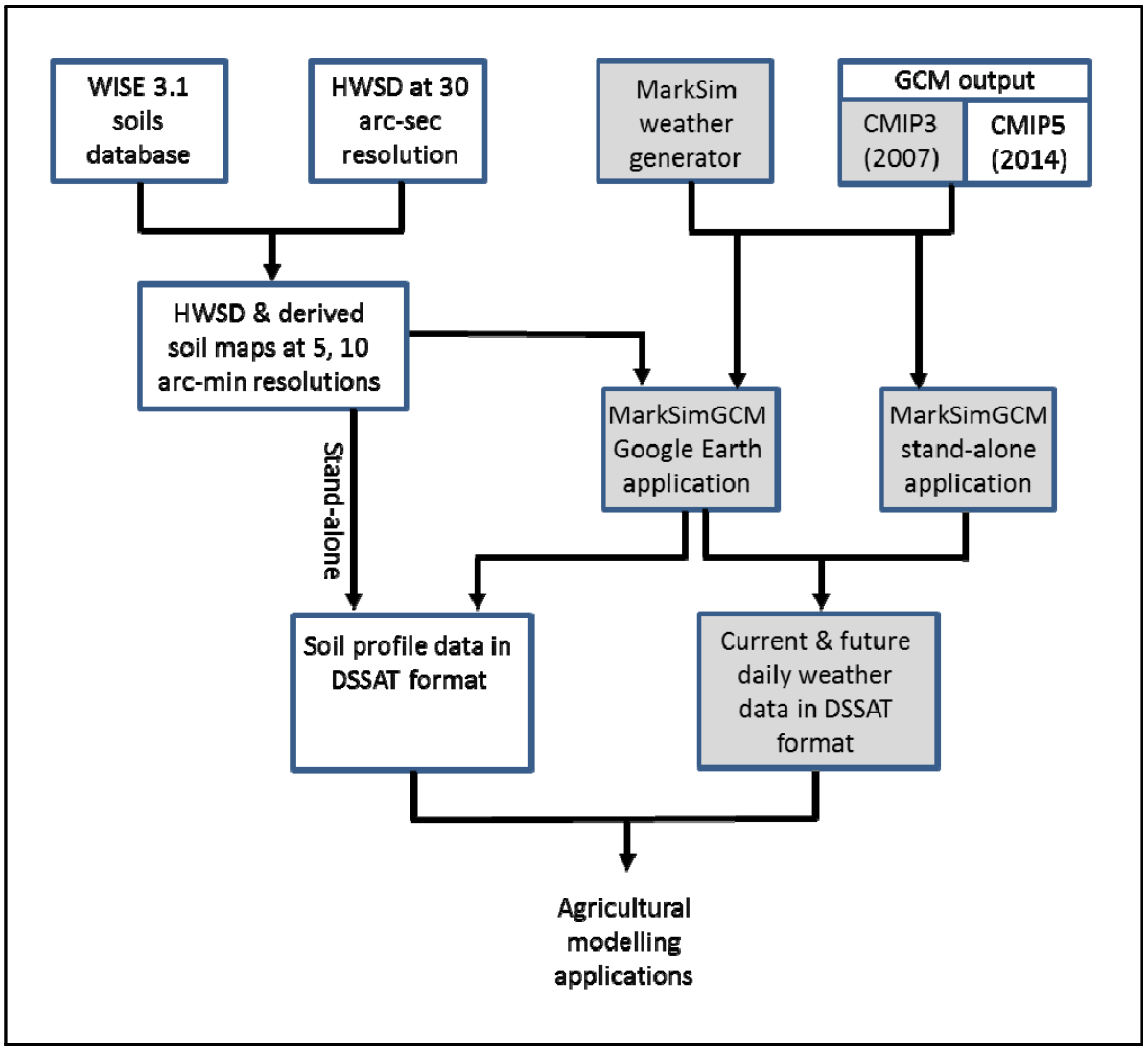

Figure 2. The different elements of the modified MarkSimGCM application. Boxes in grey outlined in Jones and Thornton (2013); the MarkSimGCM grey boxes have been modified to deal with CMIP5 data. The white boxes are outlined here.