A Comparison of Using Dominant Soil and Weighted Average of the Component Soils in Determining Global Crop Growth Suitability

T. Avellan¹, F. Zabel¹, B. Putzenlechner¹ & W. Mauser¹

¹Ludwig-Maximilians-Universität Munich, Dept. for Geography and Remote Sensing, Munich, Germany

Correspondence: Tamara Avellan, Ludwig-Maximilians-Universität Munich, Dept. for Geography and Remote Sensing, Luisenstr. 37, Munich 80333, Germany. Tel: 49-89-2180-6689. E-mail: t.avellan@iggf.geo.uni-muenchen.de

Received: February 6, 2013	Accepted: March 15, 2013	Online Published: May 29, 2013
doi:10.5539/ep.v2n3p40	URL: http://dx.doi.org/10.5539/ep.v2n3p40	

Abstract

Soil parameters represent key data input for crop suitability analysis. Soil databases are complex offering soil mapping units made up of various component soils. In the case of the Harmonized World Soil Database there can be up to 8 component soils per unit. In roughly 1/3 of soil mapping units, the additional component soils take up more than 50% of the pixel share value. The soil parameter value estimate, such as pH, salinity and organic carbon content, may differ between the value of the dominant soil component and the weighted average of the values of all component soil. Understanding the effect of these differences on crop model outputs may allow quantifying the error. In this study, we show the changes in crop suitability of 15 crops while using the parameter value estimates of the dominant soils versus a weighted average of the component soils. In the case of the latter, global crop suitability amounts to 54.5% of the earth's land surface–1% more than when using the values of just dominant soils. Intrinsic regional differences in the quality of the soil database influence the distribution of crop suitability classes especially in areas where share values of the dominant soil are low. The uncertainty range for the use of dominant versus component soils on the overall global crop suitability could be considered to be 1%, while that of each suitability class can amount to up to 4%.

Keywords: crop suitability, HWSD, quality control, dominant soil, mapping units, component soils

1. Introduction

Ensuring food security for the global population is already challenging in current times and will be even more, when population rises up to around 8.3 billion by 2030 (UNDP, 2008). Enhanced food production relies on three factors: increased yield, enhanced cropping intensity and the expansion of agricultural land (FAO, 2003). In 2009, the total amount of agricultural and permanent crops amounted to 2.5 billion ha which equals about 19% of the earth's land surface (Bontemps, Defourny, Van Bogaert, Arino, & Kalogirou, 2009). In the last four decades of the past century, 172 million ha of land have been added in developing countries (FAO, 2003). To ensure global food security, an additional 120 million ha of converted land are projected to be necessary until 2030 and an extra 5% will be necessary up to 2050 (Bruinsma, 2009). Most land is expected to be transformed in South America and Sub Saharan Africa (Fischer, 2000).

Models based on climate and soil inputs can help discern the areas where crops can grow optimally for given natural conditions. Fischer et al. (2002) showed that roughly 2.8 billion ha are to some degree suitable for rain-fed agriculture and Avellan, Zabel, and Mauser (2012) showed that about a quarter of the earth's land surface is suitable to highly suitable for the rain-fed growth of 15 major crops (Avellan, Zabel, & Mauser, 2012; Fischer, 2002). Both authors base their different models (Global agro-ecological zones versus fuzzy logic crop suitability) on global soil and climate databases. However, global soil databases are scarce and rely on patchy soil sampling. Few sets exist, such as the Harmonized World Soil Database (HWSD) (FAO/IIASA/ISRIC/ISSCAS/JRC, 2009) and the ISRIC-WISE derived soil properties on a 5 by 5 arc minute grid (Batjes, 2006). Global Climate Datasets are more varied. Past climate data can be obtained from interpolated station data (WorldClim), reanalysed forecasts (ERA) or hind-casted climate models (ECHAM, HadCM).

Avellan et al. (2012) showed that the quality of climate inputs is quite homogenous while global soil databases can differ widely. The choice of the database can have a strong effect on the amount and distribution of crop suitable areas, leading to a 10% difference between the two most common global soil datasets (Avellan et al.,

2012). Soil databases are immensely complex and the quality of the data is geographically diverse. For example, the HWSD is made up of four different input databases–each covering different areas of the world, using different sampling and compilation methods (FAO/IIASA/ISRIC/ISSCAS/JRC, 2009) (see Figure 1). Each pixel can contain up to 8 component soils which may, in sum, have a larger share within the pixel than the dominant soil class (see mock up example in Figure 2). When taking component soil classes into account, the soil parameter value estimate for each given pixel may be different than that of the dominant soil mapping unit (i.e. dominant soil value for pH is 8, but that of the weighted average of all component soils is 7.8).

In order to enhance modelling results a balance between the quantity and quality of the used input parameters has to be maintained. While more parameters might refine the modelling results, poor quality parameters might, in fact, be counterproductive. A careful analysis of both the quality of the data as well as their influence on final results might inform the choice of parameters. In Avellan et al. (2012), we started our crop suitability analysis using only the parameter value estimates of the dominant soil mapping unit of the topsoil (0-30 cm) on a pixel by pixel basis. In comparison, the Global Agro-ecological zones studies, used soil parameters from all component soils, top- and subsoils (0-30 cm and 30 cm and below), phases as well as management practices (IIASA/FAO, 2012). It is clear to the authors that other parameters relevant to soil databases such as subsoil parameters (30 cm and below), including drainage, granularity or acidity, as well as phases and management practices can have drastic effects on crop growth (Benjamin, Nielsen, & Vigil, 2003; Kirchhof et al., 2000; Van den Akker, Arvidsson, & Horn, 2003).

To our knowledge, the use of parameters in crop suitability models has not been substantiated by the analysis of the quality of the data. The inclusion of factors is defended by referring to standard works (i.e. FAO manuals (FAO, 1976, 2007) or similar) without questioning the validity of the usage. It is our intent to enhance model complexity in a step-by-step approach while showing the error margins incurred. Analogous to the well-known uncertainty ranges of climate models we wish to demonstrate a similar approach in the use of crop suitability estimations. Here, we assessed the influence of the area-weighted average of the additional component soils of the soil mapping units of the topsoil, on the amount and distribution of crop suitable areas.

2. Materials and Methods

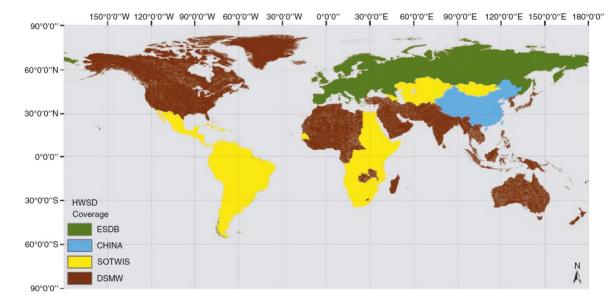


Figure 1. Distribution of the four underlying databases of the Harmonized World Soil Database (HWSD); European Soil Database (ESDB), Soil Map of China (CHINA), Soil and Terrain dataset (SOTWIS), Digital Soil Map of the World; adapted from (FAO, IIASA, ISRIC, ISS-CAS, & JRC, 2009)

2.1 Datasets

We used the following datasets and parameters at 30 arc seconds resolution (1 x 1 km at the equator):

Harmonized World Soil Database (HWSD, version 1.1): dominant and component soil mapping units of the topsoil (0-30cm) as input for eight parameter value estimates–textural class (USGS), coarse fragments volume (%), gypsum (%CaSO4), base saturation (%), pH, organic carbon (%), salinity (dS/m) and sodicity (%).

WorldClim dataset (Hijmans, Cameron, Parra, Jones, & Jarvis, 2005): mean annual temperature and mean annual precipitation

SRTM30 global digital elevation model (Farr et al., 2007; USGS, 2000): slope computed as percent rise.

Regions were defined for their economic relevance in global trade as a biophysical crop model was coupled to a Global Equilibrium Model in a subsequent step (Table 1).

TC 11 1	1 0 1	0.1	•
Tabla	l ('odina	ot the	ragiong
	 Coding 		16210115

Region code	Region name
AFR	Sub-saharan Africa
BEN	Belgium, Netherlands, Luxemburg
BRA	Brazil
CAN	Canada
CHI	China
FRA	France
FSU	Rest of former Soviet Republic
GBR	UK & Ireland
GER	Germany
IND	India
JPN	Japan
LAM	Rest of Latin America
MAI	Malaysia, Indonesia
MEA	Middle East, North Africa
MED	Spain, Portugal, Italy, Greece, Malta, Cyprus
MRC	Chile, Argentina, Uruguay, Paraguay
NAU	New Zealand, Australia
PAS1	Guayanas
PAS2	Iceland
PAS3	Switzerland
PAS4	Afghanistan, Pakistan
PAS5	Mongolia
REU	Austria, Estonia, Latvia, Lithuania, Poland, Hungary, Slovakia, Slovenia, Czech Republic, Romania, Bulgaria
RUS1	RUS1 (west)
RUS2	RUS2 (east)
SCA	Finland, Sweden, Denmark
SEA	Kambodscha, Laos, Thailand, Vietnam, Myanmar, Bangaldesh
USA	United States of America

2.2 Dominant vs. Component Soil Areas and Soil Parameter Value Estimates

Dominant soil is defined as the HWSD component soil with the largest share value irrespective of the fact that the other component soils together may have a larger share within one pixel. Soil parameter value estimates are the values each pixel has for a chosen parameter, i.e. pH, salinity, etc. In Figure 2 we have tried to show in a mock-up example how a pixel can be made up of several component soils and the effect the weighted average has on the parameter value estimate.

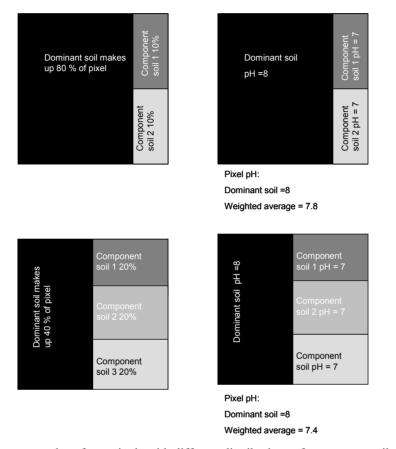


Figure 2. Mock-up examples of two pixels with different distributions of component soils (left); effect of using the weighted average on the overall parameter value estimate versus using that of the dominant soil (right)

We used GIS techniques to determine the area of prominence of dominant soils and compared it in size to that where component soils had higher percentages. We used Mondrian (version 1.2), an open source statistical analysis tool (University of Augsburg, 2012), to study the distribution of dominant soil units and component soil units. For the spatial representation of the soil units, a FORTRAN program was designed that allowed assigning the soil unit share to each pixel.

2.3 Determination of Crop Suitable Areas

We used the fuzzy logic approach as discussed in Avellan et al. (2012). Fuzzy classification methods define growth through membership functions and likelihoods (Burrough, MacMillan, & Deursen, 1992). The rationale behind this is that most soil parameters have a large error rate per se, due to sampling and handling errors, and crops are able to grow at various levels of these parameters (Rossiter, 1996). Thus strict Boolean classification systems may be too restrictive in growth ranges and areas. Fuzzy logic approaches have been used for a selected number of crops on limited study areas by other authors e.g. (Baja, Chapman, & Dragovich, 2002; Braimoh, Vlek, & Stein, 2004; Reshmidevi, Eldho, & Jana, 2009; Van Ranst, Tang, Groenemam, & Sinthurahat, 1996).

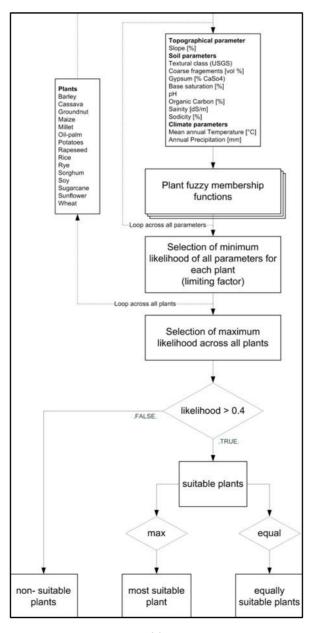
Raster-based soil, terrain and climate parameter values were matched on a sliding scale from 0 to 1 with their respective crop growth likelihoods as determined by (Sys, Van Ranst, Debaveye, & Beernaert, 1993) (Figure 3a). Subsequently, the most optimally matching crop was selected to be the most suitable for a given pixel. Each

component soil was assigned one fuzzy value (Figure 3b). Depending on the number of component soils in each soil mapping unit, up to 8 fuzzy values per pixel were assigned. These were aggregated based on their weighted share value of the respective soil mapping unit. Component soils with high share values end up with a stronger influence on the final fuzzy value.

Crop growth abilities were then categorized into four subsets as defined by Sys et al. (1993) and (FAO, 1976). Fuzzy value between:

- 1) 0–0.4 Pixel not suitable for crop growth (N) (none).
- 2) 0.4–0.6 Pixel marginally suitable for crop growth (S3).
- 3) 0.6–0.8 Pixel suitable for crop growth (S2).
- 4) 0.8–1 Pixel highly suitable for crop growth (S1).

Pixels are subsequently transformed into land surfaces according to their location on the globe through a FORTRAN programme. The total land surface is considered except Antarctica.



(a)

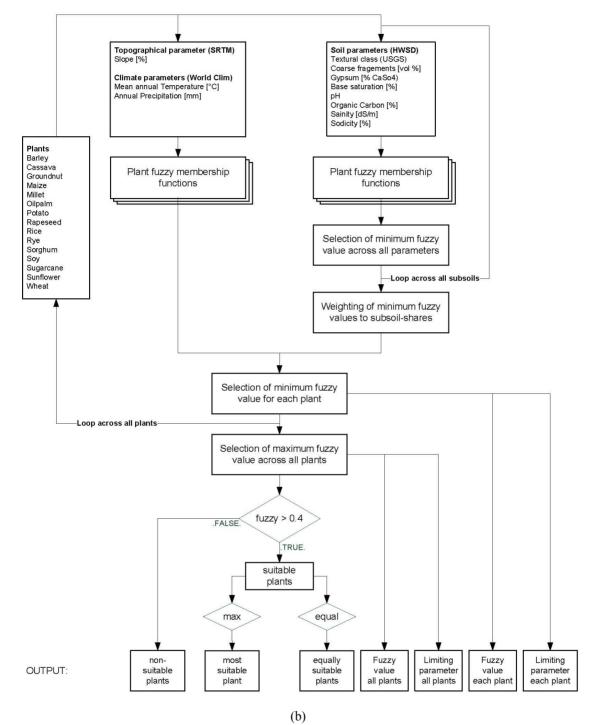


Figure 3. Overview of the methodology of fuzzy logic crop suitability analysis using just the parameter value estimates of a) the dominant soil (top) or b) of all component soils (bottom)

3. Results

3.1 Dominant vs. Component Soil Areas

In 64% of all pixel the dominant soil holds more than 50% of the pixel's share value. When looking at specific major soil groups, some only exist as dominant soil types (i.e. Is-Lithosols, Ns-Nitosols, U-Rankers and W-Planosols). Most soils comprise only two component soils in their soil mapping unit (i.e. dominant soil plus one additional component soil). Few cases exist where soil mapping units have 6 or more component soils. The

share value of the dominant soil component is very high in most of northern Asia, Greenland, the North America and large parts of Africa. These are areas where the dominant soil defines the parameter value estimate (grey areas in Figure 4). In the case of China, due to the way the database was produced, only one-the dominant-soil exists. In the Middle East, Central Asia, the Pacific and Australia, share values of the dominant soil component were very low. These are areas where the other component soils play a larger role in determining the parameter value estimates of the given pixel (black areas in Figure 4, see also mock up example in Figure 2). South America exhibits mostly areas with intermediate share values (data not shown explicitly).

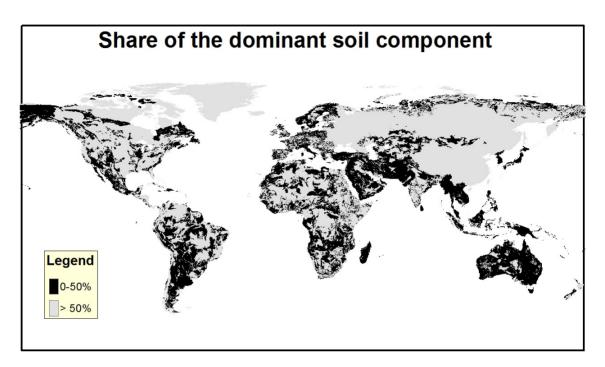


Figure 4. Analysis of shares and sequences of component soils. Grey areas represent soil mapping units where the share value of the dominant soil component holds more than 50%; Black areas are regions where the dominant soil component holds a share value of more than 50%

3.2 Determination of Crop Suitable Areas

While using the parameter value estimates of the dominant soil mapping units along with climate and terrain constraints, 9% of the earth's surface result in highly suitable (S1), 25% in suitable (S2) and 19% in marginally suitable (S3) areas (Figure 5). Barley (10.7%), wheat (5.6%), and oil palm (5.2%) are globally the most suitable crops (Figure 6) (Percentages of overall pixel, not of area).

While considering the parameter value estimates of all component soils in a given pixel, the area suitable for crop growth amounts to 54.5% of the earth's land surface excluding Antarctica. Roughly 4.5% can be categorised as highly suitable (S1); 27% and 23% can be classified as suitable (S2) and marginally suitable (S3), respectively (Figure 5). The most prominent crops were the same as when using dominant soils only, with adjustments in their overall percentages (barley-11.1%, wheat-6.5%, oil palm-5.9%) (Figure 6).

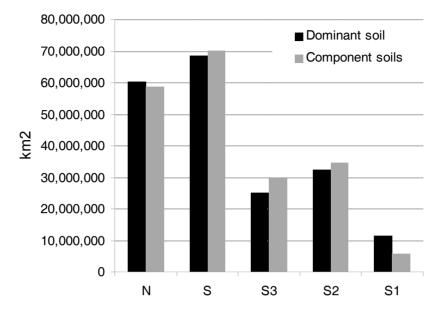
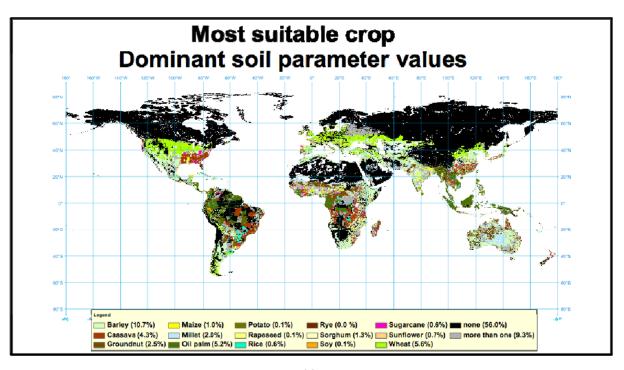


Figure 5. Amount of crop suitable areas while considering only dominant soils (black bars) or all component soils (grey bars). N-non suitable areas; S-sum of all suitable areas; S3-marginally suitable; S2-suitable; S1-highly suitable





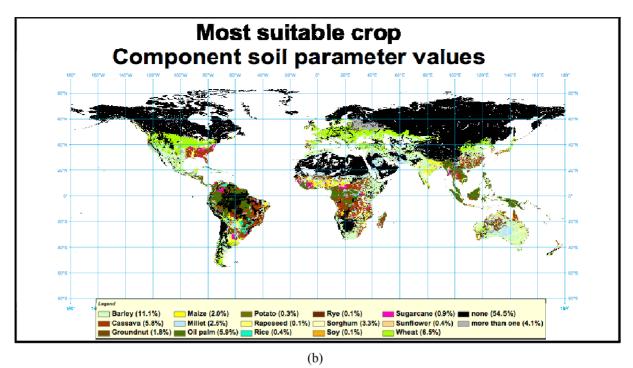


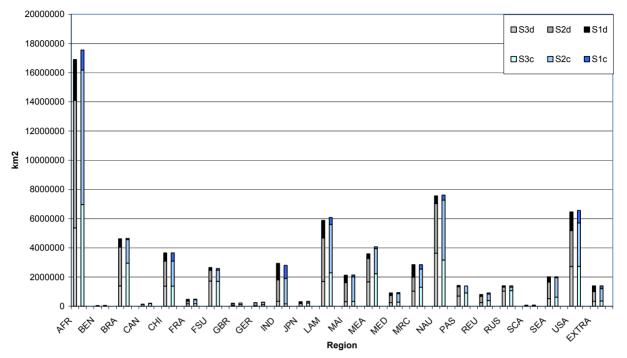
Figure 6. Distribution of the most suitable crop using all component soil parameter value estimates (top) or dominant soils (bottom) (% values represent the relative amount of pixel for that crop, not the relative area)

4. Discussion

In about 1/3 of the soil mapping units the share value of the dominant soil is less than 50% of the pixel. Its parameter value estimate, i.e. pH, salinity or organic carbon value, may not be the same as that of the weighted average of all component soils. In terms of crop suitability this translates in a 1% increase of crop suitable areas of the earth's land surface when using all soil components of the soil mapping unit of the topsoil. The global distribution of crops itself is marginally affected - the ranking of the top 5 crops with the highest amount of pixels remains equal. Changes in the distribution of the suitability classes are important. For instance, in the highly suitable areas, a reduction of 4.3% is observed when using component soils whereas the marginally suitable areas increased by 3.7% (Figure 5).

The model results reflect the qualitative differences of the underlying databases. The HWSD is an integrated patchwork of diverse datasets (see Figure 1). South and Central America, East Africa and parts of Central Asia are fed with the SOTWIS data which have the latest updates of soil samples (latest version of 2006). Europe and Russia is based on the datasets of the European Soil Database, a very comprehensive set (FAO/IIASA/ISRIC/ISSCAS/JRC, 2009). North America, West Africa, and large parts of Asia and Australia are still based on the outdated Digital Soil Map of the World (DSMW). Data for China was produced by assigning one soil class per pixel. No changes in crop suitability classes occurred for those areas which are composed of only one, the dominant, component soil, such as in the case of China. Changes in crop suitability classes were in general more prominent where dominant soil shares are below 50% such as in the Americas, Africa and Central Asia (Figure 7). For instance, while 867991 km² were assigned to be highly suitable when using the dominant soils in the Mercosur region (MRC), only 301704 km² were left in this category when using all component soils (Figure 7). Instead, 1300988 km² versus 1030206 km² were marginally suitable when using component soils or dominant soils only, respectively.

However, a linear relationship between data quality, more component soils and smaller share values in the dominant soils cannot be postulated. Some areas, in particular in the tropics, are predominantly composed of one dominant soil component with a share value of more than 50% but are based on "high quality" datasets (i.e. Amazon forest in Brazil based on SOTWIS, see grey areas in Figure 4). Other areas are based on "low quality" datasets, such as Australia on DSMW, and show large areas with several component soils and dominant soil shares below 50% (see black areas in Figure 4).



Regional distribution of crop suitable area classes

Figure 7. Region specific changes in crop suitability areas by categories using dominant soil parameter value estimates (d) or component soils (c). S3-marginally suitable, S2-suitable, S1-highly suitable

Now, how to make a choice of which dataset to use? The quality for all component soils is heterogenous; the effect on the extent and type of crop suitability minimal. The lack of consistent quality of global datasets is a known issue. A variety of research centres are working towards enhanced soil datasets and sampling, often in collaboration with many others such as in the Global Soil Initiative launched in 2011 (The Global Soil Partnership, 2011). In few cases of crop modelling some authors have undertaken extensive quality control of the underlying soil data and adapted it to their needs (Gijsman, Thornton, & Hoogenboom, 2007; Romero et al., 2012). This is very cumbersome and can only be carried out when sufficient expert staff is available for a specific target objective. However soil datasets are used widely by differing disciplines. We suggest explaining the inherent uncertainty attached to these datasets and lay open the error margin of their use. In this particular case, on the use of all component soils versus only the dominant soils we postulate that the error margin is of about 1% at a global scale.

It is clear to the authors that additional parameters can be used from the soil databases as well as a variety of other parameters such as refined climate datasets, in particular at the temporal scale. Knowledge on ethnicity, gender, management practices, adapted crops, irrigation, use of fertilizers and of the use of technology are all factors that influence the suitability of an area for agricultural purposes (FAO, 2007). Obtaining reliable data for these parameters may be even more challenging than for soil databases.

5. Conclusion

In this study, we intended to show the differences in model results when using all component soils for the analysis of crop suitability. This is important because it allows determining the level of uncertainty that modellers face when using current global soil databases. Including more parameters does not always mean better results. We showed that the distribution of the number of component soils of the HWSD is very heterogeneous on a geographical scale but is not linked to the quality of the underlying data subset. The error range for using either the dominant component soil versus all component soils could be considered to be 1%-the difference in crop suitable area between the two datasets. The margin of error varies according to the region and increases to up to 4% when looking at the individual suitability classes.

References

- Avellan, T., Zabel, F., & Mauser, W. (2012). The influence of input data quality in determining areas suitable for crop growth at the global scale–a comparative analysis of two soil and climate datasets. *Soil Use and Management*, 28(2), 249-265. http://dx.doi.org/10.1111/j.1475-2743.2012.00400.x
- Baja, S., Chapman, D. M., & Dragovich, D. (2002). A Conceptual Model for Defining and Assessing Land Management Units Using a Fuzzy Modeling Approach in GIS Environment. *Environmental Management*, 29, 647-661. http://dx.doi.org/10.1007/s00267-001-0053-8
- Batjes, N. (2006). *ISRIC-WISE derived soil properties on a 5 by 5 arc minute grid (version 1.1)* (No. 2006/02). Wageningen: ISRIC-World Soil Information.
- Benjamin, J. G., Nielsen, D. C., & Vigil, M. F. (2003). Quantifying effects of soil conditions on plant growth and crop production. *Geoderma*, 116(1-2), 137-148. http://dx.doi.org/10.1016/S0016-7061(03)00098-3
- Bontemps, S., Defourny, P., Van Bogaert, E., Arino, O., & Kalogirou, V. (2009). *GLOBCOVER 2009, Products Description and Validation Report*. ESA, Université catholique de Louvain.
- Braimoh, A. K., Vlek, P. L. G., & Stein, A. (2004). Land Evaluation for Maize Based on Fuzzy Set and Interpolation. *Environmental Management*, 33, 226-238. http://dx.doi.org/10.1007/s00267-003-0171-6
- Bruinsma, J. (2009). The resource outlook to 2050: By how much do land, water and crop yields need to increase by 2050 (p. 33). ROME: FAO.
- Burrough, P. A., MacMillan, R. A., & Deursen, W. (1992). Fuzzy classification methods for determining land suitability from soil profile observations and topography. *European Journal of Soil Science*, 43, 193-210.
- FAO. (1976). A framework for land evaluation. Food and Agricultural Organization of the United Nations.
- FAO. (2003). *World Agriculture: Towards 2015/2030: An FAO Perspective* (Jelle Bruinsma). Rome and London: FAO & Earthscan Publications Ltd.
- FAO. (2007). *Land evaluation: Towards a revised framework*. Food and Agricultural Organization of the United Nations.
- FAO/IIASA/ISRIC/ISSCAS/JRC. (2009). *Harmonized World Soil Database, Version 1.1.* FAO, Rome; IIASA, Laxenburg.
- Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., ... Alsdorf, D. (2007). The Shuttle Radar Topography Mission. *Rev. Geophys.*, 45, RG2004. http://dx.doi.org/10.1029/2005rg000183
- Fischer, G. V. (2000). Global Agro-Ecological Zones Assessment: Methodology and Results. Laxenburg.
- Fischer, G. V. (2002). *Global Agro-ecological Assessment for Agriculture in the 21st Century: Methodology and Results*. Rome, Laxenburg: Food and Agriculture Organization of the United Nations and International Institute for Applied Systems Analysis.
- 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. http://dx.doi.org/10.1016/j.compag.2007.01.001
- Hijmans, R., Cameron, S., Parra, J., Jones, P., & Jarvis, A. (2005). WORLDCLIM-a set of global climate layers (climate grids). Retrieved from http://www.worldclim.org/
- IIASA/FAO. (2012). Global Agro-ecological Zones (GAEZ v3.0). IIASA, Laxenburg, Austria and FAO, Rome, Italy. Retrieved from http://gaez.fao.org/Main.html#
- Kirchhof, G., So, H. B., Adisarwanto, T., Utomo, W. H., Priyono, S., Prastowo, B., ... Sanidad, W. B. (2000). Growth and yield response of grain legumes to different soil management practices after rainfed lowland rice. *Soil and Tillage Research*, 56(1-2), 51-66. http://dx.doi.org/10.1016/S0167-1987(00)00122-7
- Reshmidevi, T. V., Eldho, T. I., & Jana, R. (2009). A GIS-integrated fuzzy rule-based inference system for land suitability evaluation in agricultural watersheds. *Agricultural Systems*, 101, 101-109. http://dx.doi.org/10.1016/j.agsy.2009.04.001
- Romero, C. C., Hoogenboom, G., Baigorria, G. A., Koo, J., Gijsman, A. J., & Wood, S. (2012). Reanalysis of a global soil database for crop and environmental modeling. *Environmental Modelling & Software*, 35, 163-170. http://dx.doi.org/10.1016/j.envsoft.2012.02.018

- Rossiter, D. G. (1996). A theoretical framework for land evaluation. *Geoderma*, 72, 165-190. http://dx.doi.org/10.1016/0016-7061(96)00031-6
- Sys, C. O., Van Ranst, E., Debaveye, J., & Beernaert, F. (1993). *Land evaluation: Part III Crop requirements* (Vols. 1-3). Bruxelles (Belgium): Administration Generale de la Cooperation au Developpement.

The Global Soil Partnership. (2011). Retrieved from http://www.fao.org/globalsoilpartnership/home/en/

UNDP. (2008). World Population Prospects: The 2008 Revision. Population Division of the Department of Economic and Social Affairs of the United Nations Secretariat.

University of Augsburg. (2012). Mondrian. Retrieved from http://stats.math.uni-augsburg.de/mondrian/

USGS. (2000). SRTM_30 Version 2.1. Retrieved from http://dds.cr.usgs.gov/srtm/version2_1/SRTM30/

- Van den Akker, J. J., Arvidsson, J., & Horn, R. (2003). Introduction to the special issue on experiences with the impact and prevention of subsoil compaction in the European Union. *Soil and Tillage Research*, 73(1-2), 1-8. http://dx.doi.org/10.1016/S0167-1987(03)00094-1
- Van Ranst, E., Tang, H., Groenemam, R., & Sinthurahat, S. (1996). Application of fuzzy logic to land suitability for rubber production in peninsular Thailand. *Geoderma*, 70, 1-19. http://dx.doi.org/10.1016/0016-7061(95)00061-5

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/3.0/).