Modeling Coffee in Mexico: CAF2014 Geo-spatial Setup – Part II

Technical Note on the Implementation of the SAbERES Project **(Draft)**

Nikolay Khabarov, Christian Folberth, Rastislav Skalsky, Charlotte Gonzalez Abraham, Valeria Javalera Rincon

30 June 2025

Table of Contents

Context of the study....2
Onboarding of NASA GDEX dataset....2
Consideration of potential soil data refinements....3
1. SAGARPA/SADER-FAO soil database....3
2. Uncertainty in pedotransfer functions....4
Modeling representative management of small coffee plantations in Mexico....4
Conclusion and further work....5
References....6
Figures....9

Context of the study

In a series of project's internal weekly exchange and planning events, the IIASA team has identified several research tracks, which translated into the modeling of growing coffee in Mexico with the CAF2014 model [1]. Here we cover technical questions and document changes to the geo-spatial setup which are incremental to the earlier reports [2], [3], and [27].

Onboarding of NASA GDEX dataset

The multi-step process of onboarding (for justification see [27]) of the higher resolution climate dataset NASA GDEX [18] in its quality check stage revealed inconsistencies in the data. Reported numbers here are indicating the quality of the entire (global) dataset and therefore are not constrained to the territory of Mexico.

```
----------------------------
Conversion report summary:
  Error: 36 files with oversized values
  Error: 53756 files with negative sol. rad.
  Error: 111668 files with inconsistent temperature (Tmax < Tmin)
  Warning: 12664 (5%) files with bad rel. humidity (adjusted)
  Warning: 27685 (0%) cell-days with bad rel. humidity (adjusted)
----------------------------
```

(Technical reference: above is an excerpt from the `bin2epic` processing log for GFDL-CM4_gr1 GCM, historical period, including the time sub-period of just two years 1971-1972.)

The total number of cells (grid pixels) is 265272. While (1) over-sized values constitute a rather small share of all grid cells and can be excluded from the analysis; (2) negative solar radiation (occurring in approximately 20% of grid cells) could be formally fixed by setting it to zero; and (3) inconsistent temperature $T_{max} < T_{min}$ present in more tan 40% of grid cells could be formally fixed by swapping T_{max} and T_{min} values, - we decided not to apply such corrections, because we deem inconsistencies (2) and (3) affecting big parts of the dataset to be indicators of a more systematic potential problem in that data. That's why we had to abandon the idea of using climate data with higher spatial resolution for modeling.

Below is a respective information for the UKESM1-0-LL GCM, ssp126, including the time sub-period of just two years 2015-2016.

```
----------------------------
Conversion report summary:
  Error: 34 files with oversized values
  Error: 49736 files with negative sol. rad.
  Error: 8210 files with inconsistent temperature (Tmax < Tmin)
```

```
  Error: 311 files with negative wind speed values
  Warning: 8446 (3%) files with bad rel. humidity (adjusted)
  Warning: 28186 (0%) cell-days with bad rel. humidity (adjusted)
----------------------------
```

Consideration of potential soil data refinements

1. SAGARPA/SADER-FAO soil database

As reported in [27] there are data-related challenges with employing the GEO-BENE soil database [9] for the CAF2014 based modeling. In order to avoid missing a potential opportunity to improve that part, we have explored the recently published database of soil sampling in México [19]. That soil survey was carried out by the Secretariat of Agriculture, Livestock, Rural Development, Fisheries, and Food (SAGARPA; since 2018: Secretariat of Agriculture and Rural Development – SADER [21]) of Mexico in collaboration with the Food and Agriculture Organization of the United Nations (FAO) in the year 2012. The dataset consists of more than 4000 sampled points, which are shown on Figure 1 and zoomed as relevant to the arabica coffee production in Mexico in Figure 2. The Figure 2 illustrates a rather sparse coverage of the coffee producing areas by the sampling points, which is especially vivid for the Chiapas state. Therefore, the soils, which are relevant to coffee production are likely under-represented in the set of "compound samples surveyed on managed cropland" as the SAGARPA/SADER-FAO soil database is described in [19].

This sparse coverage warrants exploration of the soil variability by leveraging other sources of soil information in addition to those analyzed so far and also requires estimation of accuracy of the soil water characteristics that can be obtained from such sources (these are required for the CAF2014 modeling).

We consulted with the Harmonized World Soil Database [20] and checked a random point soil data in Chiapas (where according to Figure 2 (municipality-level) coffee is produced). Figure 3 illustrates few parameter values that are relevant to this location, which indicate considerable variability of soils at that location, whereas no conclusive information can be derived on just one single factual soil type and the whole set of various soils is deemed to be relevant at this geographic position.

The parameter AWC (Available water storage capacity) does never exceed the value 0.15 mm/mm in HWSD (see the HWSD documentation [20]). When checked at locations where according to Figure 2 (municipality-level) coffee is produced the values range 0.10-0.15 mm/mm (see e.g. Figure 4), whereas in the GEOBENE DB [9] the test location in Chiapas delivered values 0.25 mm/mm or greater (see the SAbERES Technical Note [2], Table 5). This is a considerable inconsistency (overestimation) in a key model's parameter informed by the two data sources[1]. At the

[1] An additional check of the more recent version of the HWSD v.2.0 [23] revealed a revised range for AWC with the maximum value for soils increased up to 0.175 mm/mm (see the the HWSD v.2.0 documentation [23], Table 2.2) except for soils consisting primarily of organic materials (histosols, see [22]), which are irrelevant for coffee production at high elevations in Mexico (see the map of histosols in [22]). Anyway, this increased value does not eliminate the discussed inconsistency.

same time, as opposed to that, the soil parameters set in CAF2014 (see soil water characteristics analyzed the earlier in Techical Note [2], WCFC=0.41, WCWP=0.26, resulting in AWC=0.41-0.26=0.15 mm/mm) are found to be perfectly compatible with HWSD data (point estimates; there is also admissible uncertainty range from 0.125 to 0.150 mm/mm when checked multiple locations in HWSD overlapping with the relevant map (coffee producing areas) on Figure 2). This is why we have adopted the approach of using a consistent default soil parametrization from CAF2014.

In order to consider additional "pro et contra" information related to that decision, we have also explored the plausible uncertainty range of the AWC estimates that can be derived based on the soil parameters available in the HWSD soil database.

2. Uncertainty in pedotransfer functions

According to the state-of-the-art method proposed by Saxton & Rawls for estimating soil water characteristics [24], there are uncertainties associated with the use of the pedotransfer functions (reproduced in Figure 5) employed to estimate Wilting Point (WP) and Field Capacity (FC), which correspond respectively to 1500 kPa and 33 kPa moisture. The uncertainty in WP and FC estimates is approximately ±30% and ±30% respectively as visually estimated for somewhat middle range values of WP=0.15 and FC=0.3. That uncertainty translates into a potential uncertainty of the AWC estimates within the range of (0.3-30%) - (0.15+30%) = 0.2 – 0.2 = 0.0 to (0.3+30%) - (0.15-30%) = 0.4 – 0.1 = 0.3. This potential range is by far exceeding the scale of inconsistency in the AWC estimates coming from the GEO-BENE soil database [9] and both most recent versions of the HWSD [20], [23] and indicates potential errors in thus obtained estimates: an "uncontrolled" variation potentially driving the data outside of the feasibility range. Additional concern is coming from the suggested pedotransfer functions' applicability limitation explicitly indicated by Saxton & Rawls [24]: "Samples with "extreme" values were omitted from the data. Excluded were those with [...] OM > 8 % (w) … " [24]. At the same time HWSD informs on the presence of soils with higher than 8% of OM (organic matter; see Figure 3; organic carbon is about 60% of organic matter - see [25] p.40).

This is why after completing this analysis we decided at this point to use a standard CAF2014 parametrization of the soil as a more robust approach and focus on the estimation of the climate (and management) signal as a relative (delta) yield to the historical baseline.

Modeling representative management of small coffee plantations in Mexico

The application of the default (small holders' BAU representative) management as described in [27] uniformly to all coffee growing areas in Mexico as per SIAP database [10] i.e.: planting/growing tree density, tree thinning/pruning, fertilizer application (nitrogen input) under default CAF2014 soil parametrization and under modeled very near future climate 2021-2030 (ISIMIP3b; gfdl-esm4_r1i1p1f1_w5e5_ssp585, subset) results in a reasonable reproduction of spatial distribution of historical yields (see Figure 7), especially under the following existing modeling limitations:

- coffee plant density parameter is not explicitly modeled by CAF2014, so some yield overestimates obtained through the modeling even for rather low N inputs can be explained by potential lower plant density than assumed by the model;
- the obtained numbers have to be understood as yield potentials due to not taking into account such disruptive factors lowering the yield as diseases.

These limitations manifest themselves in an apparent overestimation of coffee yields in the south-western part of Mexico (compared to other areas).

Conclusion and further work

The result obtained from CAF2014 can be deemed satisfactory under existing modeling and data limitations. Next steps in refining it could include:

- using additional climate models (GCMs) to explore the range of uncertainty induced in the modeling due to differences in climate re-projections for historical / projections for near-future time period;
- analysis of climate change projections for a set of SSPs/RCPs (shared socio-economic pathways) and medium and end of the century time periods consistently with [26].

References

[1] The first public release of CAF2014, a process-based model for coffee agroforestry systems. URL: https://zenodo.org/record/3608877 (Accessed: 2024-11-12)

[2] Khabarov, N., Folberth, C., Skalsky, R., Javalera Rincon, V., Modeling Coffee Agroforestry Systems with the Process-based Model CAF2014, Technical Note on the Implementation of the SAbERES Project (2024) https://user.iiasa.ac.at/~khabarov/2024-11-19-saberes/Technical-Note-SAbERES-IIASA-NK-Cofee-CAF2014-Modeling_v73.pdf (Accessed: 2025-03-31)

[3] Khabarov, N., Folberth, C., Skalsky, R., Javalera Rincon, V., Modeling Coffee in Mexico: Extending CAF2014 Model for Geo-spatial Applications, Technical Note on the Implementation of the SAbERES Project (2025) https://user.iiasa.ac.at/~khabarov/2025-03-31-saberes/Technical-Note-SAbERES-IIASA-NK-Cofee-CAF2014-Modeling-GeoSpatial_v96.pdf (Accessed: 2025-03-31)

[4] Juraj Balkovič, Marijn van der Velde, Rastislav Skalský, Wei Xiong, Christian Folberth, Nikolay Khabarov, Alexey Smirnov, Nathaniel D. Mueller, Michael Obersteiner. Global wheat production potentials and management flexibility under the representative concentration pathways. Global and Planetary Change, Volume 122, 2014, Pages 107-121, ISSN 0921-8181, https://doi.org/10.1016/j.gloplacha.2014.08.010

[5] Khabarov, N., Smirnov, A., Balkovič, J., Skalský, R., Folberth, C., Van Der Velde, M., & Obersteiner, M. (2020). Heterogeneous Compute Clusters and Massive Environmental Simulations Based on the EPIC Model. Modelling, 1(2), 215-224. https://doi.org/10.3390/modelling1020013

[6] Adriana Otero Arnaiz, Alexander Chinh, Coffee Annual, USDA, Date: May 17, 2024, Report Number: MX2024-0025, URL: https://apps.fas.usda.gov/_newgainapi/_api/_Report/DownloadReportByFileName?fileName=__Coffee+Annual_Mexico+City_Mexico_MX2024-0025.pdf (Accessed: 2025-03-31)

[7] ISIMIP/GGCMI Phase 3 Climate projections data set (ISIMIP3b) URL: https://protocol.isimip.org/#/ISIMIP3b/32-climate-related-forcing (Accessed: 2024-11-12)

[8] GMT 6: Wessel, P., Luis, J. F., Uieda, L., Scharroo, R., Wobbe, F., Smith, W. H. F., & Tian, D. (2019). The Generic Mapping Tools Version 6. Geochemistry, Geophysics, Geosystems, 20(11), 5556–5564. https://doi.org/10.1029/2019GC008515

[9] GEO-BENE global database for bio-physical modeling v. 1.0 (Concepts, methodologies and data) URL: https://geo-bene.project-archive.iiasa.ac.at/node/1734.html (Accessed: 2024-11-11)

[10] SIAP database: Datos Abiertos | Servicio de Información Agroalimentaria y Pesquera | gob.mx | En esta sección se encuentra la recopilación principal de datos abiertos del SIAP. URL: http://infosiap.siap.gob.mx/gobmx/datosAbiertos.php (Accessed: 2024-07-31)

[11] INEGI database: Instituto Nacional de Estadística y Geografía (INEGI), Mexico - Subnational Administrative Boundaries, URL: https://data.humdata.org/dataset/cod-ab-mex (Accessed: 2025-04-09)

[12] MAPSPAM 2010. Yu, Q., You, L., Wood-Sichra, U., Ru, Y., Joglekar, A. K. B., Fritz, S., Xiong, W., Lu, M., Wu, W., and Yang, P. (2020). A cultivated planet in 2010 – Part 2: The

global gridded agricultural-production maps, Earth Syst. Sci. Data, 12, 3545–3572, https://doi.org/10.5194/essd-12-3545-2020

[13] van Oijen, M., Dauzat, J., Harmand, J.-M., Lawson, G., & Vaast, P. (2010). Coffee agroforestry systems in Central America: I. A review of quantitative information on physiological and ecological processes. Agroforestry Systems, 80(3), 341–359. https://doi.org/10.1007/s10457-010-9294-y

[14] van Oijen, M., Dauzat, J., Harmand, J.-M., Lawson, G., & Vaast, P. (2010). Coffee agroforestry systems in Central America: II. Development of a simple process-based model and preliminary results. Agroforestry Systems, 80(3), 361–378. https://doi.org/10.1007/s10457-010-9291-1

[15] van Oijen, M., Haggar, J., Barrios, M., Büchi, L., Cerda, R., Cerretelli, S., López, E., de Melo Virginio Filho, E., & Ospina, A. (2022). Ecosystem services from coffee agroforestry in Central America: estimation using the CAF2021 model. Agroforestry Systems, 96(7), 969–981. https://doi.org/10.1007/s10457-022-00755-6

[16] Della Peruta, R., Mereu, V., Spano, D., Marras, S., Vezy, R., & Trabucco, A. (2025). Projecting trends of arabica coffee yield under climate change: A process-based modelling study at continental scale. Agricultural Systems, 227, 104353. https://doi.org/10.1016/j.agsy.2025.104353

[17] Rahn, E., Vaast, P., Läderach, P., van Asten, P., Jassogne, L., & Ghazoul, J. (2018). Exploring adaptation strategies of coffee production to climate change using a process-based model. Ecological Modelling, 371, 76–89. https://doi.org/10.1016/j.ecolmodel.2018.01.009

[18] NASA Earth Exchange Global Daily Downscaled Projections v.1.2 (NEX-GDDP-CMIP6) https://www.nccs.nasa.gov/services/data-collections/land-based-products/nex-gddp-cmip6 Documentation URL: https://www.nccs.nasa.gov/sites/default/files/NEX-GDDP-CMIP6-Tech_Note_4.pdf (Accessed: 2025-06-05)

[19] Arroyo-Cruz, C. E., Prado, B., Kolb, M., Mora-Palomino, L. N., Todd-Brown, K., & Guevara, M. (2025). Synthesis of a National Soil Dataset Across Productive Land in Mexico: The Importance of Making Existing Data Accessible. European Journal of Soil Science, 76(3). https://doi.org/10.1111/ejss.70116

[20] Harmonized World Soil Database v 1.2. URL: https://webarchive.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/ Accessed: 2025-06-11

[21] Secretariat of Agriculture and Rural Development (SADER) / Secretariat of Agriculture, Livestock, Rural Development, Fisheries, and Food (SAGARPA) https://en.wikipedia.org/wiki/Secretariat_of_Agriculture_and_Rural_Development Accessed on: 2025-06-25

[22] Histosols and the map of Histosols of the world, https://en.wikipedia.org/wiki/Histosol Accessed: 2025-06-25

[23] FAO & IIASA. 2023. Harmonized World Soil Database version 2.0. Rome and Laxenburg. https://doi.org/10.4060/cc3823en

[24] Saxton, K. E., & Rawls, W. J. (2006). Soil Water Characteristic Estimates by Texture and Organic Matter for Hydrologic Solutions. Soil Science Society of America Journal, 70(5), 1569–1578. https://doi.org/10.2136/sssaj2005.0117

[25] Fred Magdoff & Harold van Es. (2021) Building soils for better crops. Ecological management for healthy soils. Fourth Edition. Handbook Series Book 10. Sustainable Agriculture Research and Education Program. ISBN 9781888626193 https://www.sare.org/publications/building-soils-for-better-crops/amount-of-organic-matter-in-soils/

[26] Plan de Adaptación al Cambio Climático Productores Orgánicos de Calakmul. AC Campeche y Chiapas, Instituto Interamericano de Cooperación para la Agricultura (IICA) (Draft 2024-07-24; internal SAbERES documentation)

[27] Khabarov, N., Folberth, C., Javalera Rincon, V., Modeling Coffee in Mexico: CAF2014 Geo-spatial Setup – Part I, Technical Note on the Implementation of the SAbERES Project (Draft, 26 June 2025) URL: https://user.iiasa.ac.at/~khabarov/2025-06-30-saberes/Technical-Note-SAbERES-IIASA-NK-Cofee-CAF2014-GeoSpatial-Calibration-Part-1_v78.pdf

Figures

Figure 1. Sampled points (pink) from the SAGARPA/SADER-FAO soil database [19] mapped over the municipalities reported in the SIAP database [10] as producing coffee (production data are averaged over 2003-2023 and mapped to municipalities as delineated by INEGI [11]).

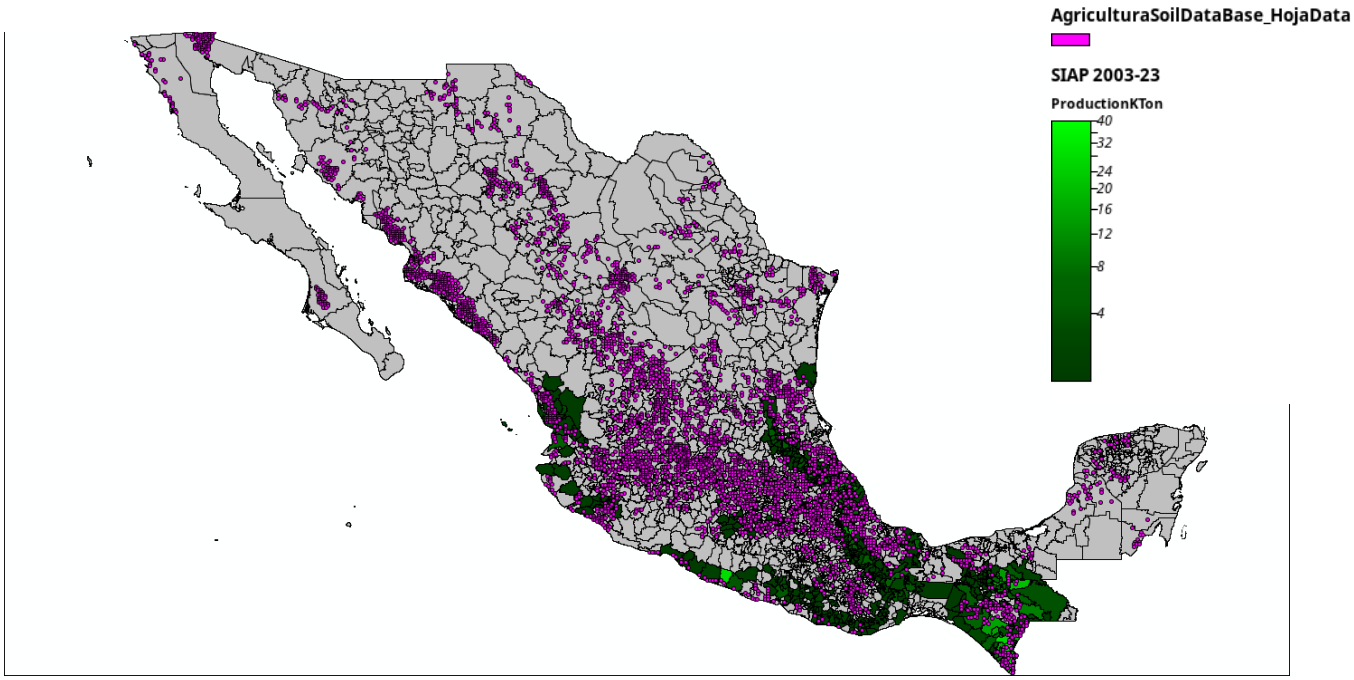


Figure 2. Sampled points (pink) from the SAGARPA/SADER-FAO soil database [19] mapped over the municipalities reported in the SIAP database [10] as producing coffee (production data are averaged over 2003-2023 and mapped to municipalities as delineated by INEGI [11]). Zoomed to the coffee producing part of Mexico.

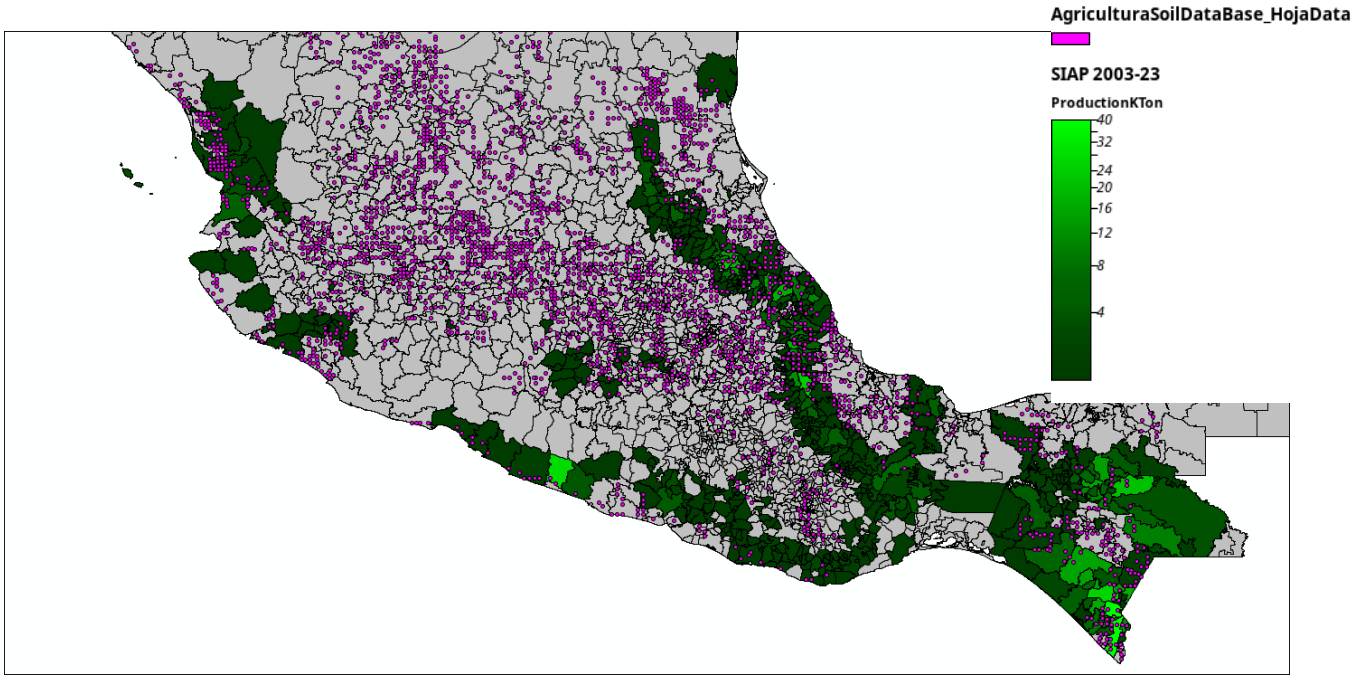


Figure 3. Harmonized World Soil Database [20] user interface showing a point soil data at a location in Chiapas state (small black cross on the map), where coffee is produced.

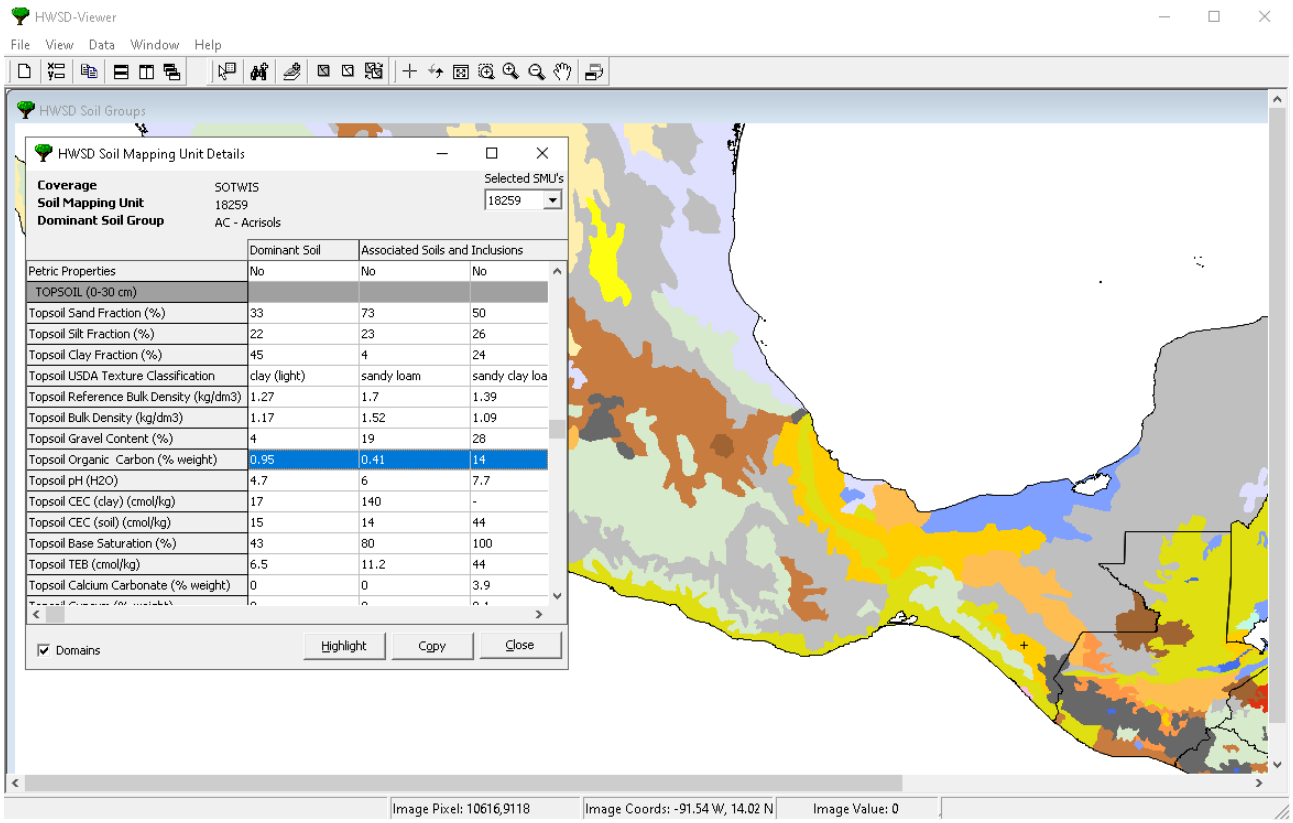


Figure 4. Harmonized World Soil Database [20] user interface showing a point soil data (notably AWC) at a location in municipality of Chiapas, where coffee is produced.

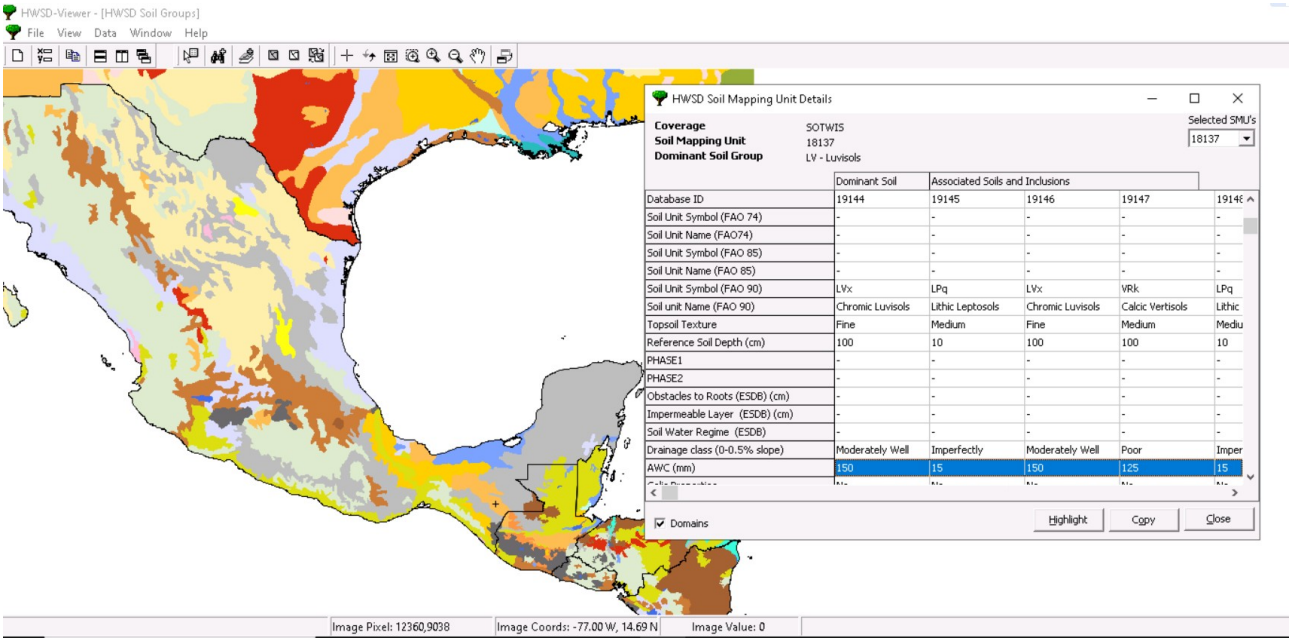


Figure 5. Soil water characteristics estimation method proposed by Saxton & Rawls (copied from [24]).

Table 1. Equation summary for soil water characteristic estimates.†

Variable	Equation	R^2/S_e	Eq.
	Moisture Regressions		
θ_{1500}	$\theta_{1500} = \theta_{1500t} + (0.14 \times \theta_{1500t} - 0.02)$	0.86/0.02	1
	$\theta_{1500t} = -0.024S + 0.487C + 0.006OM + 0.005(S \times OM) - 0.013(C \times OM) + 0.068(S \times C) + 0.031$		
θ_{33}	$\theta_{33} = \theta_{33t} + [1.283(\theta_{33t})^2 - 0.374(\theta_{33t}) - 0.015]$	0.63/0.05	2
	$\theta_{33t} = -0.251S + 0.195C + 0.011OM + 0.006(S \times OM) - 0.027(C \times OM) + 0.452(S \times C) + 0.299$		

Figure 6. Uncertainty in the Saxton & Rawls estimates of Wilting Point (moisture at 1500 kPa) and Field Capacity (moisture at 33 kPa) (copied from [24]).

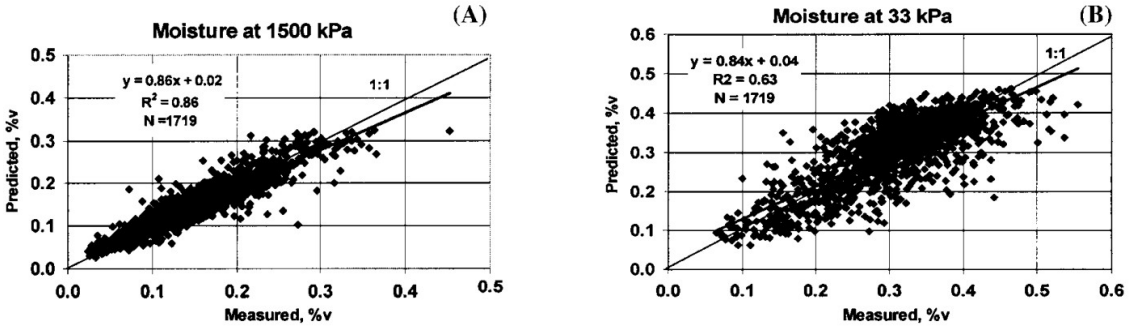


Figure 7. Left panel: Coffee yield in Mexico (ton fresh matter/ha/year) gridded at 0.5 arc-degree from municipality-level (SIAP DB [10]) averaged over 2003-2023 and mapped to municipalities as delineated by INEGI [11]. Right panel: coffee yields (ton fresh matter/ha/year) as estimated by geo-spatial version of CAF2014 "Rhaobi" model for the derived BAU-representative management and default CAF2014 soil water characteristics for the climate dataset ISIMIP3b gfdl-esm4_r1i1p1f1_w5e5_ssp585 daily 2021-2030.

