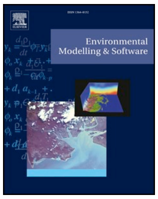
Environmental Modelling and Software 119 (2019) 374–378

Contents lists available at ScienceDirect

Environmental Modelling & Software

journal homepage: www.elsevier.com/locate/envsoft



Advancing the application of a model-independent open-source geospatial tool for national-scale spatiotemporal simulations

Jing Huang[a,b,*], Laura Scherer[b], Kang Lan[a], Fu Chen[c], Kelly R. Thorp[d]

[a] College of Life Science and Engineering, Southwest University of Science and Technology (SWUST), Mianyang, 621010, China
[b] Institute of Environmental Sciences (CML), Leiden University, Leiden, 2333 CC, the Netherlands
[c] College of Agronomy, China Agricultural University, Beijing, 100193, China
[d] USDA-ARS, U.S. Arid Land Agricultural Research Center, Maricopa AZ, 85138, USA

ARTICLE INFO

Keywords:
Geospatial simulation
AquaCrop
High spatiotemporal resolution
Water consumption
Yield

ABSTRACT

Growing demands for geospatial application of environmental models have led to tool development for conducting simulations spatially. The model-independent, open-source tool “Geospatial Simulation” (GeoSim) has been developed previously. Based on previous applications at field scale, this study advances GeoSim application for national-scale and multi-year simulations. The widely-applied AquaCrop model was implemented by GeoSim to simulate wheat yield and irrigation requirements on a daily step across China from 2000 to 2009. The spatial inputs required by AquaCrop were minimized and 6915 unique response units were identified among the primary 116,801 polygons. It took around 20 h to perform the 10-year simulations. Post-processing of simulation outputs permitted mapping at the original 5 arc-minute resolution. The novel methods developed in this study demonstrate new opportunities for efficiently managing national-scale and multi-year simulations with high resolution. They render AquaCrop more suitable for studies on the water-food nexus at large scales, which are more policy-relevant.

1. Introduction

Over the past decades, applications of agroecosystem models have rapidly expanded, and the models have been applied in numerous areas such as resource use and efficiency, food security, and environmental performance (Holzworth et al., 2015). There are more than 100 crop models available, and they have been applied for the simulation of about 150 crops or land uses (Rivington and Koo, 2010). With the expanding application domain, models are increasingly applied at diverse spatiotemporal scales (Bryan, 2013; Folberth et al., 2012; Liu, 2009; Zhao et al., 2013). This has required either spatial simulation capability to be added to the models, or constructing model wrappers which can provide this capability. In recent years, progress has been made on the development of geographic information systems (GIS) to handle geoprocessing tasks, store geo-spatial data, manage input and output data from the simulation model and visualize the results spatially (Bryan, 2013; Holzworth et al., 2015; Thorp and Bronson, 2013; Thorp et al., 2008). A few agroecosystem models have been improved as GIS-based modelling systems allowing simulations at multiple locations to be run simultaneously with dynamic interactions, e.g. Apollo-DSSAT for DSSAT (Thorp et al., 2008), GEPIC for EPIC (Liu, 2009), and Grid-Parallel-APSIM for APSIM (Zhao et al., 2013).

However, many GIS-based modelling systems have been developed for a specific application or model. These systems are model-dependent and are useful only for limited purposes. In addition, some systems, which may be proprietary, computer platform dependent, now obsolete, or cost prohibitive, are not available for all users (Thorp and Bronson, 2013). To avoid these problems, Thorp and Bronson (2013) developed a tool called Geospatial Simulation (GeoSim), which is a model-independent open-source system for managing point-based model simulations at multiple locations. The tool has been demonstrated on the widely-applied AquaCrop and DSSAT models (Memic et al., 2018; Thorp and Bronson, 2013). Although GeoSim does not restrict model applications to any specific spatial scale, the past applications of GeoSim were all conducted at the field scale. To date, no study has demonstrated GeoSim for broader spatial scales, although it can potentially be applied as such.

There is a growing demand for large-scale and high-resolution modelling of crop yields and water consumption, as the concerns for food and water security continue to increase under global change (McNeill et al., 2017). To demonstrate GeoSim's applicability to satisfy such demand, this study used GeoSim to implement the widely-applied

---

[*] Corresponding author. College of Life Science and Engineering, Southwest University of Science and Technology (SWUST), Mianyang, 621010, China
E-mail address: huang.jing@swust.edu.cn (J. Huang).

https://doi.org/10.1016/j.envsoft.2019.07.003
Received 28 November 2018; Received in revised form 23 June 2019; Accepted 6 July 2019

FAO AquaCrop model for simulating wheat yield and irrigation water requirements across China from 2000 to 2009. Other tools such as AquaCrop-GIS (Lorite et al., 2013) and AquaCrop-OS (Foster et al., 2017) have also been designed to implement AquaCrop for multiple simulations. However, these tools work with old AquaCrop versions and have limitations, such as low simulation efficiency or need for programing skills, which can constrain the model applicability. The objective of this study was 1) to demonstrate that GeoSim can efficiently work with AquaCrop for modelling at large spatiotemporal scales with high resolution and 2) to guide users on how to explore the flexibility of GeoSim on setting up the simulations to improve efficiency. The outputs of national wheat yield and irrigation water requirements can be further applied to examine implications for an integrated management of water resources and food production in China. While beyond the scope of this study, GeoSim also has the potential to work with other models following a similar approach. As such, it facilitates addressing global challenges such as food security and environmental security by efficiently extending models for spatial simulations.

2. Methods

2.1. GeoSim and AquaCrop

GeoSim was designed to facilitate spatial simulations for any point-based model which uses ASCII files for input and output (Thorp and Bronson, 2013). It was developed as a plug-in for Quantum GIS (QGIS, https://www.qgis.org), and both of these software programs are open-source and freely available. GeoSim provides an interface to run point-based models using geospatial data contained in a QGIS database. Six tools are currently available in GeoSim. These tools enable the system to prepare and process the required shapefile, to manage input and output data for the polygons in that shapefile, and to optimize model parameters to minimize errors.

FAO's AquaCrop (http://www.fao.org/aquacrop) is a water-driven dynamic model which derives biomass gain as a function of water consumption under rain-fed or different irrigation conditions on a daily step (Raes et al., 2009; Steduto et al., 2009). The model performance has been widely tested for numerous crops under diverse environments and agricultural production systems around the world (Vanuytrecht et al., 2014). In order to evaluate the model performance and judge the suitability of the model for water and food security research, field experiments were conducted to calibrate and validate the AquaCrop model for wheat and maize cropping systems in China (Han et al., 2019). It showed reasonable agreement between simulated crop yields and observed yields, indicating that the calibrated AquaCrop model is robust at reproducing potential yields under full irrigation (Han et al., 2019).

2.2. Pre-processing of spatial inputs

To present a national case study of a 10-year simulation, relevant data for the simulations of wheat yield and irrigation water requirement were collected for the period from 2000 to 2009, in which all the required data are available. Six types of input data were used: 1) crop distribution, 2) climate data, 3) crop parameters, 4) soil parameters, 5) initial soil water conditions, and 6) management data. All the unique input files required by AquaCrop were prepared separately, prior to conducting simulations with GeoSim.

A shapefile (vector data) of national wheat distribution was converted from a raster dataset with a resolution of 5 arc-minutes in the year 2000 (section 1.1 in Appendix A). Climate files were prepared according to the AquaCrop formats based on daily data from 825 meteorological stations provided by the National Meteorological Information Centre (NMIC, http://data.cma.cn). The coordinates of the meteorological stations were used to generate a shapefile containing the station codes (section 1.2 in Appendix A). The default file for atmospheric CO_2 concentration in the AquaCrop database was used. Two crop types (winter and spring wheat) were distinguished by the data obtained from the NMIC and were contained in the attribute table of the shapefile with meteorological station codes. Crop files for both spring and winter wheat containing numerous parameters were prepared according to the AquaCrop formats (section 1.3 in Appendix A). Day numbers, which indicate the first and last days of cropping and simulation periods, were calculated for China's 41 agro-ecological zones (AEZ) and included in the attribute table of the AEZ shapefile (Fig.A3 in Appendix A). The crop type, which indicates whether the crop was spring or winter wheat, was also included in the AEZ shapefile. A soil shapefile containing 39 soil codes was prepared with national coverage (section 1.4 in Appendix A). Furthermore, 39 soil files containing soil hydraulic parameters for each soil code in the shapefile were prepared according to the AquaCrop formats. Initial soil water contents were assumed to be at field capacity (section 1.5 in Appendix A). To model the irrigation water requirement of wheat under full irrigation, the "determination of net irrigation requirements" option was selected in AquaCrop to create an irrigation file (section 1.5 in Appendix A). Other management effects, such as fertilizer application and ground surface cover, were disregarded. Groundwater was not considered due to lacking detail in the national dataset.

All the prepared climate, crop, soil, irrigation, and initial condition files were stored in a directory (can be several directories). The flexibility of GeoSim permits a user to pass any spatial input data to the model files, such as climate data and soil parameters. However, to minimize the input spatial variables, this case study applied GeoSim by passing spatial data only to AquaCrop's project file (*.PRM) which controls AquaCrop simulations (Fig. 1). The directory (or directories), where input files were stored in, was specified in AquaCrop's project file (*.PRM). Then GeoSim was used to pass only the names of these files as spatial variables to the project file, rather than passing the specific numerous parameters for each input file. To further reduce the spatial variables, the climate files for each simulation were named as the same code of the meteorological station from where the climate data were. GeoSim passed a station code to different locations where the climate file names were required in the project file.

2.3. Identification of unique response units

The shapefiles for wheat distribution, meteorological station codes (including crop types), AEZ-based day numbers, and soil codes were used to create a base polygon shapefile required by GeoSim. By intersecting these layers, a new shapefile with 116,801 polygons was generated. However, efficiency of spatial simulations could be improved by combining some of these polygons. For example, because the wheat distribution information was only used to indicate the location of wheat cropping, this spatial attribute information did not need to be passed to AquaCrop. The only required attributes for geospatial AquaCrop simulations, as managed by GeoSim, were meteorological station codes,

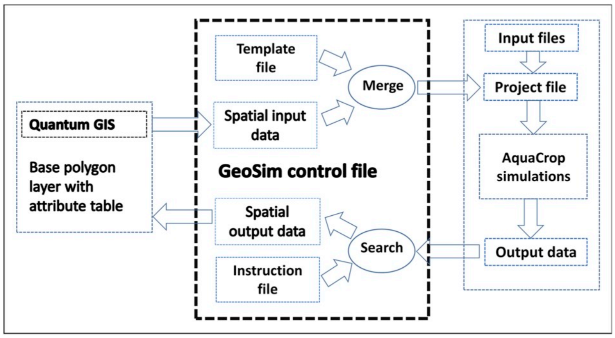
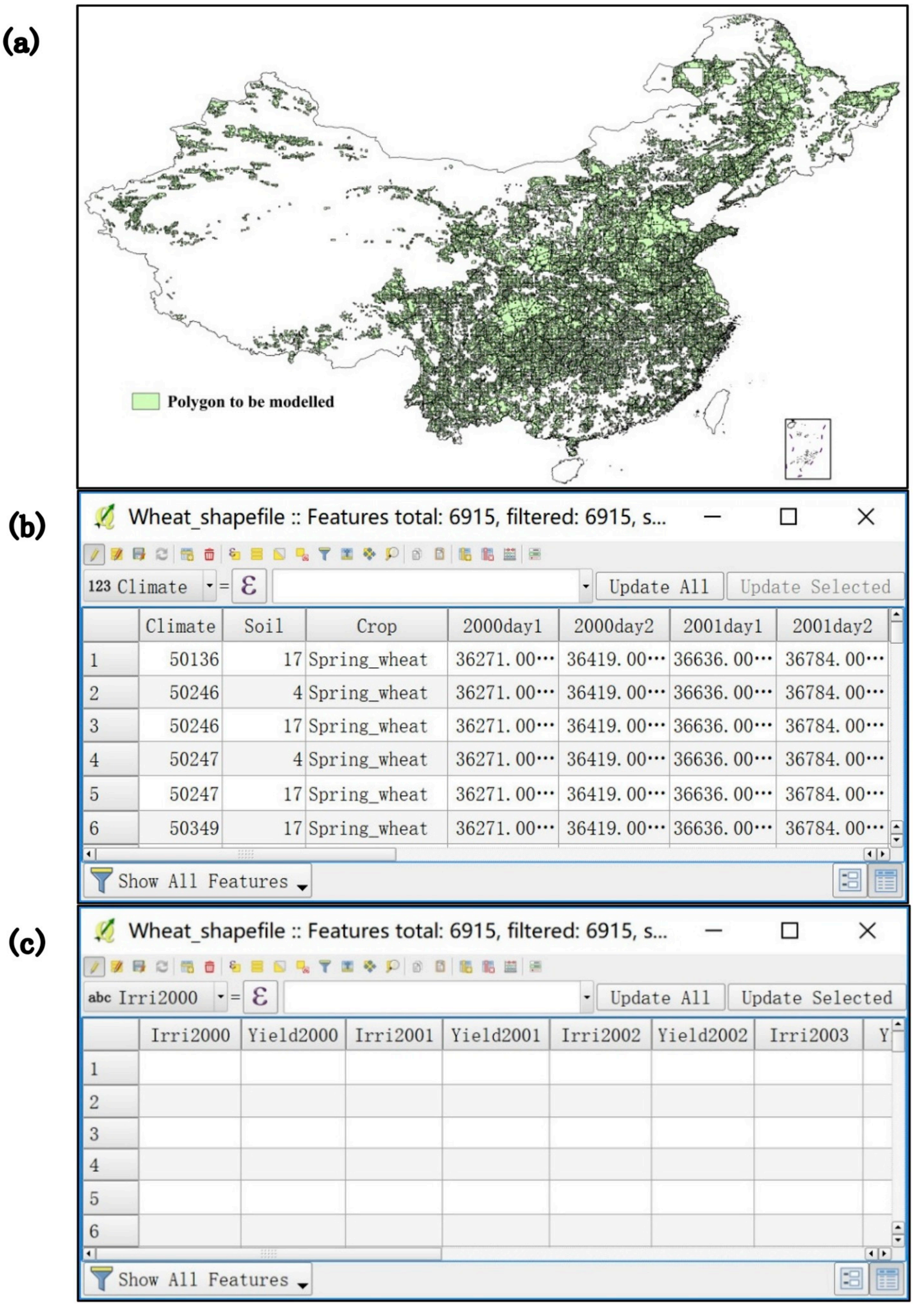


Fig. 1. Framework of GeoSim application for AquaCrop, adapted from (Thorp and Bronson, 2013).



Wheat_shapefile :: Features total: 6915, filtered: 6915, s...

	Climate	Soil	Crop	2000day1	2000day2	2001day1	2001day2
1	50136	17	Spring_wheat	36271.00…	36419.00…	36636.00…	36784.00…
2	50246	4	Spring_wheat	36271.00…	36419.00…	36636.00…	36784.00…
3	50246	17	Spring_wheat	36271.00…	36419.00…	36636.00…	36784.00…
4	50247	4	Spring_wheat	36271.00…	36419.00…	36636.00…	36784.00…
5	50247	17	Spring_wheat	36271.00…	36419.00…	36636.00…	36784.00…
6	50349	17	Spring_wheat	36271.00…	36419.00…	36636.00…	36784.00…

Wheat_shapefile :: Features total: 6915, filtered: 6915, s...

	Irri2000	Yield2000	Irri2001	Yield2001	Irri2002	Yield2002	Irri2003	Y
1								
2								
3								
4								
5								
6								

Fig. 2. The base polygon shapefile (a) and excerpts from the attribute table (b and c). The first three columns "Climate", "Soil" and "Crop" in (b) indicate meteorological station codes, soil codes, and wheat types respectively, while remaining 20 columns indicate the day numbers, i.e. the first and last day numbers for the 10 years. (c) shows the blank spaces to be overwritten by the results from the output files. There were 20 columns for receiving all the target results, yield and irrigation estimates, for each of 10 years.

wheat types, soil codes, and day numbers. As some polygons shared these attributes in common, there was opportunity to eliminate redundant AquaCrop simulations. To identify the unique response units, which were defined as polygons with unique combination of meteorological station code, crop type, soil code, and day numbers, the "Dissolve" tool in QGIS was applied to merge polygons with the same attribute values. Finally, only 6915 polygons remained in the base shapefile to be modelled (Fig. 2a). This demonstrates an advantage of using GeoSim within QGIS. By using QGIS tools to identify identical spatial zones, the efficiency of simulations managed by GeoSim could be improved. The input information for each polygon were appended in the different columns of the attribute table (Fig. 2b), while 20 blank columns (two outputs with 10 years) were set up to receive output data (Fig. 2c).

2.4. *Template, instruction, and control files*

In this study, the template file (*.gst) is a replicate of the AquaCrop's project file (*.PRM). Within the template file, "unique codes" were included at the locations of day numbers and the names of climate, crop, and soil files. As this case study conducted simulations in 10 years, there were 10 sections indicating the simulation for each year in the template file. Accordingly, there were 20 lines of commands in the instruction file (*.gsi) to retrieve the irrigation and yield results for each of the 10 years. The names of the attributes in the base shapefile and their corresponding unique codes in the template file were provided in

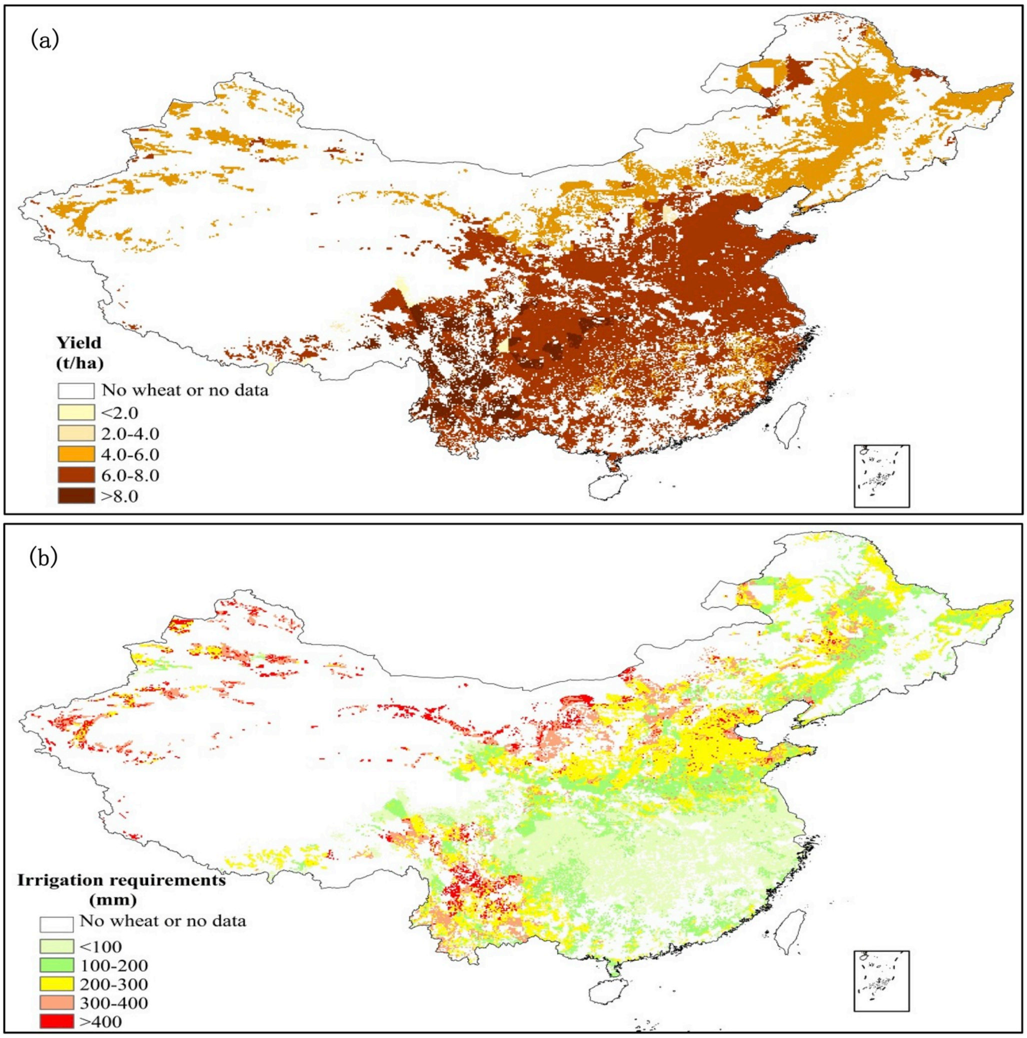


Fig. 3. Average wheat yield (a) and (b) irrigation water requirements during 2000–2009.

the control file (*.gsc). Excerpts from the template, instruction and control files are presented in Appendix A. Further information for developing these GeoSim files can be found in Thorp and Bronson (2013) and the GeoSim manual. After setting up the GeoSim files, GeoSim conducted the simulations for all the polygons in the base shapefile.

2.5. *Post-processing of spatial outputs*

As the model completed the simulations, results for wheat yield and irrigation requirement in each of the 10 years for each polygon were transferred from the model output files to the attribute table of the base shapefile. Then the "Vector Geoprocessor" tool within GeoSim was applied to calculate the mean values of irrigation and yield from 2000 to 2009, and the results were appended to the base shapefile.

To disaggregate the unique response units back to the original resolution, we intersected the base shapefile with the primary wheat distribution layer. This new shapefile contained 116,801 polygons as opposed to 6915 polygons used during the modelling. The large number of spatial units equaled those resulting from the intersection of wheat distribution, meteorological station codes (including crop types), AEZ-based day numbers, and soil codes layers (section 2.3), and also contained all the result attributes. The yield and irrigation attributes in the new shapefile were then converted to raster datasets with a resolution of 5 arc-minutes. By pairing native QGIS tools with GeoSim to improve the efficiency of spatial simulations, a novel method to develop national-scale maps of wheat yield and irrigation requirements was demonstrated.

3. Results and discussion

3.1. *Case study*

The total running time to complete 69,150 simulations (6915 polygons with 10 years) was about 20 h with single-core processing. The grid-based results of wheat yield and irrigation requirement with a resolution of 5 arc-minutes are presented in Fig. 3. Depending on the research goals, further analysis and application of these maps can provide implications for decision-making related to water management, food security, and land use management. The modelling quality depends on the input data and AquaCrop model algorithms rather than GeoSim. Because GeoSim serves to help users manage the geospatial model inputs and outputs, it does not change any input data or model algorithms within AquaCrop. An in-depth analysis of the model outputs is beyond the scope of this paper. By spatially applying AquaCrop using GeoSim, the detailed spatial and temporal variability of crop yield and irrigation requirements was easily demonstrated across China.

3.2. *Advancement of GeoSim application*

Previous applications of GeoSim at the field scale used the tool to pass spatial input data to AquaCrop's soil file, initial soil water file, and irrigation files (Thorp and Bronson, 2013). However, for large spatiotemporal simulations at the national scale, a simpler approach was to pass spatial input data only to AquaCrop's project file. By setting up all of AquaCrop's input files in a specific directory, GeoSim was used to pass only four types of spatial variables (meteorological station codes, crop types, soil codes, and day numbers) to the project file. Although GeoSim can be used to pass any spatial input data required by AquaCrop, passing data to the climate and soil files would have resulted in a massive and cumbersome base shapefile and required several template files to receive these data. By directly incorporating spatial input data into unique climate and soil files and using GeoSim to adjust file names assigned to each spatial zone, this study demonstrated an efficient way to use a base shapefile with few columns and used only one template file. Thus, the effort for pre-processing the GeoSim files was substantially reduced. It also demonstrated the flexibility of GeoSim to be used in numerous different ways depending on the goals of simulation analysis.

To reduce the modelling time, 6915 unique response units were identified in the base shapefile among the 116,801 primary polygons. Because QGIS allowed the post-processing of outputs, the original resolution of the analysis could be regained. By conducting simulations only for the unique spatial zones and subsequently expanding the simulation results to the primary zones, simulation efficiency was improved by over 16 times, while the primary spatial resolution was not changed. These pre-processing and post-processing activities demonstrated the advantage and flexibility of using GeoSim in combination with native QGIS functionality to achieve efficient simulation analyses at the national scale.

3.3. Comparison with alternative AquaCrop wrappers

Compared with AquaCrop-GIS (Lorite et al., 2013), which is currently recommended by FAO for a high number of simulations (http://www.fao.org/aquacrop), GeoSim significantly reduced the simulation time. AquaCrop-GIS uses an Excel spreadsheet to control AquaCrop, and it creates all the possible combinations of inputs for each polygon. Thereby, AquaCrop-GIS wastes time by doing irrelevant simulations. For example, it will do 5,393,700 simulations (6915 polygons × 39 soils × 2 wheat types × 10 years) when applied for our case study. In contrast, GeoSim only passes polygon information to AquaCrop that is of interest for the study. Therefore, GeoSim required only 69,150 AquaCrop simulations, almost 80 times less. Moreover, it may be impossible for a modern desktop computer to complete all the 5,393,700 simulations with AquaCrop-GIS in a reasonable time. It would also be time consuming to split the work into several parts. Also, selecting targeted results from numerous output files would be cumbersome and time consuming with AquaCrop-GIS, but GeoSim automates this.

Foster et al. (2017) developed the AquaCrop-OS model, which also facilitates large numbers of AquaCrop simulations. However, this model requires users to have programming experience and it has its own format requirements for input files and users must define separate sets of input files for each simulation run. Unlike AquaCrop-OS, GeoSim does not change the formats of any input files required by AquaCrop. All GeoSim does is to help users manage the geospatial model inputs and outputs within QGIS. A user who knows how to conduct the standalone AquaCrop simulations can easily learn to apply GeoSim for geospatial simulations.

4. Conclusions

By using native QGIS geo-processing functions with GeoSim, the efficiency of AquaCrop for conducting national-scale simulations was improved. The flexibility of GeoSim permits its implementation in numerous different ways depending on the goals of the analysis. Other tools exist for spatial extension of AquaCrop, but none could perform with the ease and efficiency of GeoSim. This study demonstrated a novel approach to apply GeoSim for large-scale spatiotemporal simulations with high resolution, which renders AquaCrop more valuable for studying the national water-food nexus. GeoSim is a model-independent tool and it can always work with the latest version of the model. Also, as GeoSim was developed as an open-source software, users can improve or expand the source code for customized purposes.

Declarations of interest

None.

Acknowledgements

This work was supported by China's National Key Research and Development Program (grant number 2016YFD0300210) and the Longshan Academic Talent Research Supporting Program of SWUST (grant numbers 18LZXT06 and 18LZX449). Jing Huang is grateful for the scholarship she received from the China Scholarship Council (grant number 201808510050).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envsoft.2019.07.003.

References

Bryan, B.A., 2013. High-performance computing tools for the integrated assessment and modelling of social-ecological systems. Environ. Model. Softw 39, 295–303. https://doi.org/10.1016/j.envsoft.2012.02.006.

Folberth, C., Yang, H., Wang, X., Abbaspour, K.C., 2012. Impact of input data resolution and extent of harvested areas on crop yield estimates in large-scale agricultural modeling for maize in the USA. Ecol. Model. 235–236, 8–18. https://doi.org/10.1016/j.ecolmodel.2012.03.035.

Foster, T., Brozović, N., Butler, A.P., Neale, C.M.U., Raes, D., Steduto, P., Fereres, E., Hsiao, T.C., 2017. AquaCrop-OS: an open source version of FAO's crop water productivity model. Agric. Water Manag. 181, 18–22. https://doi.org/10.1016/j.agwat.2016.11.015.

Han, T., Zhao, Y.H., Huang, J., Wang, C., Ai, X.K., Chen, F., Chu, Q.Q. Applicability Evaluation of AquaCrop Model for Winter Wheat & Summer Maize Cropping System in the Heilonggang River Valley. Journal of China Agricultural University. (In Chinese with English abstract) (Accepted).

Holzworth, D.P., Snow, V., Janssen, S., Athanasiadis, I.N., Donatelli, M., Hoogenboom, G., White, J.W., Thorburn, P., 2015. Agricultural production systems modelling and software: current status and future prospects. Environ. Model. Softw 72, 276–286. https://doi.org/10.1016/j.envsoft.2014.12.013.

Liu, J., 2009. A GIS-based tool for modelling large-scale crop-water relations. Environ. Model. Softw 24, 411–422. https://doi.org/10.1016/j.envsoft.2008.08.004.

Lorite, I.J., García-Vila, M., Santos, C., Ruiz-Ramos, M., Fereres, E., 2013. AquaData and AquaGIS: two computer utilities for temporal and spatial simulations of water-limited yield with AquaCrop. Comput. Electron. Agric. 96, 227–237. https://doi.org/10.1016/j.compag.2013.05.010.

McNeill, K., Macdonald, K., Singh, A., Binns, A.D., 2017. Food and water security: analysis of integrated modeling platforms. Agric. Water Manag. 194, 100–112. https://doi.org/10.1016/j.agwat.2017.09.001.

Memic, E., Graeff, S., Claupein, W., Batchelor, W.D., 2018. GIS-based spatial nitrogen management model for maize: short- and long-term marginal net return maximising nitrogen application rates. Precis. Agric. 20, 295–312. https://doi.org/10.1007/s11119-018-9603-4.

Raes, D., Steduto, P., Hsiao, T.C., Fereres, E., 2009. Aquacrop—the FAO crop model to simulate yield response to water: II. main algorithms and software description. Agron. J. 101, 438–447. https://doi.org/10.2134/agronj2008.0140s.

Rivington, M., Koo, J., 2010. Report on the Meta-Analysis of Crop Modelling for Climate Change and Food Security Survey. CCAFS, Copenhagen, Denmark.

Steduto, P., Hsiao, T.C., Raes, D., Fcreres, E., 2009. AquaCrop—the FAO crop model to simulate yield response to water: I. Concepts and underlying principles. Agron. J. 101, 426–437. https://doi:10.2134/agronj2008.0139s.

Thorp, K.R., Bronson, K.F., 2013. A model-independent open-source geospatial tool for managing point-based environmental model simulations at multiple spatial locations. Environ. Model. Softw 50, 25–36. https://doi.org/10.1016/j.envsoft.2013.09.002.

Thorp, K.R., DeJonge, K.C., Kaleita, A.L., Batchelor, W.D., Paz, J.O., 2008. Methodology for the use of DSSAT models for precision agriculture decision support. Comput. Electron. Agric. 64, 276–285. https://doi.org/10.1016/j.compag.2008.05.022.

Vanuytrecht, E., Raes, D., Steduto, P., Hsiao, T.C., Fereres, E., Heng, L.K., Garcia Vila, M., Mejias Moreno, P., 2014. AquaCrop: FAO's crop water productivity and yield response model. Environ. Model. Softw 62, 351–360. https://doi.org/10.1016/j.envsoft.2014.08.005.

Zhao, G., Bryan, B.A., King, D., Luo, Z., Wang, E., Bende-Michl, U., Song, X., Yu, Q., 2013. Large-scale, high-resolution agricultural systems modeling using a hybrid approach combining grid computing and parallel processing. Environ. Model. Softw 41, 231–238. https://doi.org/10.1016/j.envsoft.2012.08.007.