Chapter 8
Computation and Baseline: Efficient Methods for Solving a Large System of Equations for Projection and Scenario Analysis

Iman Haqiqi and Uris Lantz C. Baldos

The use of quantitative models is critical for sustainability studies because they provide geospatial insights and enable the analysis of the impact of climate extremes and conservation policies on food security and environmental sustainability. Traditional regional models may provide national-level insights, but they miss out on the crucial details that can be identified by using a geospatial approach. Explicit modeling of spatial outcomes enables targeted interventions and fosters more equitable solutions for creating a sustainable future. Additionally, these models can offer valuable insights into resource management and potential trade-offs between economic growth and environmental sustainability by explicitly incorporating land and water constraints.

However, the development and implementation of such models comes with significant computational hurdles. The large amount of data involved—coupled with the complex interactions between diverse economic, land, and water systems—creates a demanding challenge. These models may consist of a system of simultaneous equations with millions of unknowns with limited parallel computation possibilities. Traditional computational approaches often fail to efficiently handle the calculations required to solve these simulations.

To overcome these difficulties, the SIMPLE-G model offers a scalable and effective approach. This chapter introduces the key features and advantages of the SIMPLE-G computational approach and demonstrates how it can significantly contribute to gridded quantitative economic analysis. The chapter introduces and explains the key features and advantages of the SIMPLE-G computation and showcases how SIMPLE-G can contribute to advancing sustainability research and policy

I. Haqiqi (✉) · U. L. C. Baldos
Center for Global Trade Analysis, Department of Agricultural Economics, Purdue University, West Lafayette, IN, USA
e-mail: ihaqiqi@purdue.edu

I. Haqiqi, T. W. Hertel (eds.), *SIMPLE-G*,
https://doi.org/10.1007/978-3-031-68054-0_8

analysis, then discusses how the baseline and scenarios can be defined and implemented.

1 Computer Implementation

The SIMPLE-G database and model are prepared and solved with the GEMPACK modeling suite (Horridge et al. 2018). This software package is specifically designed to solve large-scale economic equilibrium models with numerous markets and agents and is used by economists to simulate the impacts of a wide range of policy issues, including trade, taxation, environmental regulation, and development. It is also used for dynamic projections of the economy. GEMPACK offers several advantages, including a user-friendly algebraic notation that is easy to learn and use, which makes it accessible even to users without a strong programming background. The database files can also readily store multiscale and multidimensional variables. Other attractive features of this software are discussed below, but the unique advantage of GEMPACK in the context of multiscale modeling is its ability to condense the model and later backsolve for key endogenous variables. This makes it a computationally efficient software package for solving SIMPLE-G models.

1.1 Condensation

Solution times can be substantial for an equilibrium model with many equations and with complex interconnections among the unknown variables (e.g., the market responds to farmer decisions even as the farmers respond to market outcomes).[1] A typical global SIMPLE-G application might include two million grid cells and 17 regions. A system of about 20 equations (recall Chap. 5) determines crop output for each of those grid cells, given grid-level exogenous settings and the endogenous price of output (which is the same for all cells in a given region), for a total of approximately 40 million grid-level equations. For each region, up to 100 other equations add up the grid-cell outputs to obtain total crop supply or interrelate region-level prices and quantities. The overwhelming majority of equations are clearly at the grid-cell level. A nonlinear system of 20 million equations would be impossibly slow to solve and might require enormous amounts of computer memory. We can greatly reduce the number of equations by substitution (i.e.,

[1] Researchers have designed multiple algorithms to reduce solution time. Most algorithms iterate between two phases: a linear algebra phase that solves a first-order approximation to the nonlinear equation system and a "formula" phase that updates variable values and recomputes coefficients of the linear system. In GEMPACK, solution time for the linear phase rises with the square or cube of the number of equations, while time for the formula phase tends to increase only linearly.

condensation). For example, we can write the fertilizer demand equation as follows (the grid index is omitted here):

$$q_N = q - a - a_N - \sigma(p_N - a_N - p - a). \tag{8.1}$$

Then, we could replace each occurrence of q_N in *other* equations by the righthand side of Eq. 8.1,

$$q - a - a_N - \sigma(p_N - a_N - p - a), \tag{8.2}$$

thereby dropping the first equation (8.1) from the system. By condensing out the grid-cell-specific variables, we can dramatically reduce the size of the model to be solved. After the system is solved, we can use the above equation to recover (or backsolve for) values of q_N.

Economic modelers often use such techniques, manually performing such substitutions in their model specification files. The drawback is that the necessary algebra is difficult, and the remaining equations become extremely complicated and nontransparent, especially when a large number of substitutions are performed. The GEMPACK software is able to solve the algebra to perform such substitutions (and their backsolves) automatically, reaping a performance gain while leaving the model specification (TABLO) file in its original, simpler, uncondensed form.

For SIMPLE-G, *all* grid-level equations are substituted out, leaving a regional-level system (for 17 regions) of modest size: 1700 equations. Such a system takes very little time to solve. However, the coefficients of the system involve grid-level calculations; the time taken to solve these is proportional to the number of grid cells. Hence, the solution time increases only linearly as a function of the number of grid cells (see Fig. 8.1).

1.2 Linearization

GEMPACK can automatically translate the original equation system into a linearized system (reformulated as a system of first-order partial differential equations). Alternatively, modelers can themselves specify conveniently interpretable linearized forms of the underlying behavioral equations. In this case, the modeler must also supply the requisite update formulas to ensure that all relevant variables are properly updated over the course of the nonlinear solution of this linear system. Clever representation of the model can facilitate condensation as well as a more rapid solution of the model. In SIMPLE-G, all of the cross-grid-cell interactions are transmitted through regional market prices, allowing us to substitute out all of the variables with a grid-cell index. Once we know the regional crop, fertilizer, irrigation capital, and other input prices, we can backsolve for crop output, input use, land prices, and the shadow price of irrigation water in each grid cell independently.

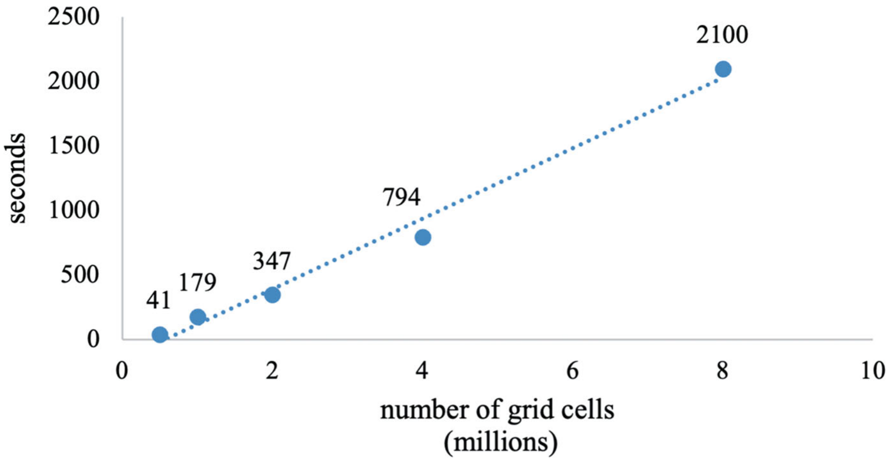

Fig. 8.1 Linear relationship between solution time for SIMPLE-G-US and number of grid cells Condensation allows users to solve a SIMPLE-G system of equations with 1 million grid cells (~10 million endogenous variables) on a laptop in a few minutes. CPU = core i5-7500, RAM = 16 GB

Since the model is nonlinear, the cost shares, price, and quantity variables must be updated at each step in the solution process. Consequently, the model is solved by multistep methods such as the Euler method or Gragg's modified midpoint method (Pearson 1991). The large system of linear equations is solved using sparse matrix techniques (Schiffmann and Jerie 2019). Richardson extrapolation is used to improve accuracy (Pearson 1991). This linearized approach has proven capable of accurately solving very large, nonlinear models (e.g., model in Fig. 8.1 contains eight million grid cells).

1.3 Decomposition

In addition to these features, GEMPACK has other extensions that prove invaluable in SIMPLE-G applications (see Part IV). It provides a way to formulate inequality constraints or nondifferentiable equations as complementarities (Bach and Pearson 1996), which can be important in sustainability analyses. It also offers a technique to decompose changes in model variables due to several shocks into components due to each individual shock (Harrison et al. 2000). We will illustrate this technique in Chap. 12.

1.4 Computation "In the Cloud"

The web application version of SIMPLE-G permits users to simulate, explore, and visualize the results of SIMPLE-G without installing the GEMPACK program or any visualization software. It works based on "containers" and Linux versions of GEMPACK programs run on the GeoHub server (https://mygeohub.org/groups/glassnet/res/tools). The web application also includes presolved experiments and demonstrations on how to run the model and analyze results based on the policy briefs presented at the 2018 Conference on Long Run Sustainability of US Agriculture (Haqiqi et al. 2018; Liu et al. 2018).

The latest version of the web application allows users to run their own experiments, which could range from global projections of food production and food demand to grid-level analysis. Users need to provide a GEMPACK header array (HAR) data file that includes the values for gridded shocks or growth rates and a GEMPACK command (CMF) text file that includes the configuration details. The tool can be used to explore the impacts of different scenarios of climate change, water scarcity, population growth, income growth, and biofuel mandates on food production, water resources, and land use. Ongoing improvements to the tool will make it more user friendly and easy to use. Figure 8.2 illustrates an example of SIMPLE-G results visualized on MyGeoHub.

It is also possible to link SIMPLE-G to models and data from different domains (e.g., hydrology, crop science, climate science, and ecology). This enables researchers to address complex, real-world grand challenge problems that require

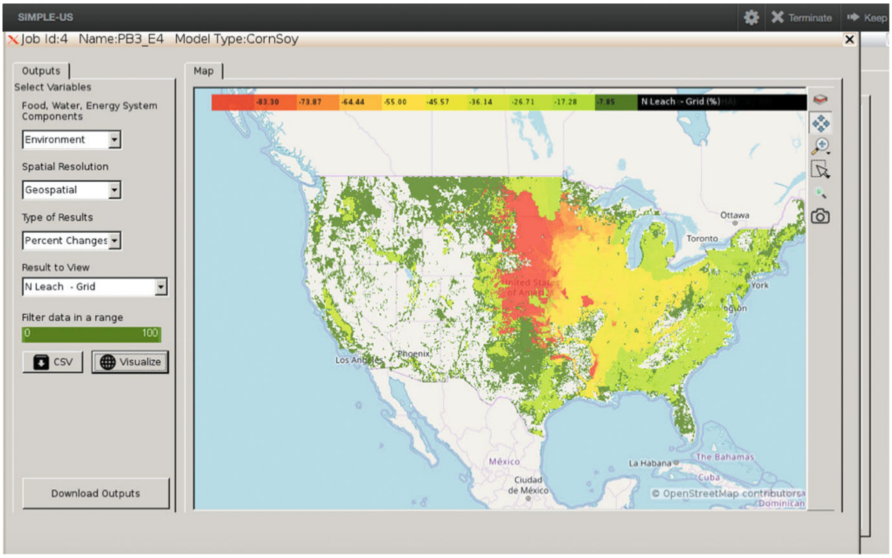

Fig. 8.2 Sample window from the SIMPLE-G web application at https://mygeohub.org/tools/simpleus

in-depth collaboration among researchers from multiple disciplines. Such linkages often involve harnessing multiscale and multidimensional data and models with unidirectional connections. For example, Agro-IBIS provides information to SIMPLE-G through the simulated yield response to N fertilizer. SIMPLE-G can also provide information on economically motivated changes in N application rates as inputs to Agro-IBIS to estimate N leaching under different scenarios.

It is also possible to couple SIMPLE-G with other models, which—in general—faces two major problems. First, it is difficult for collaborators to understand each other's data/models, a precondition for effective collaboration. Second, the infrastructure is usually unavailable to run the coupled systems. To address these challenges, Woo et al. (2022) designed C3F, a Collaborative Container-Based Model Coupling Framework, which can accelerate model integration and linking efforts by leveraging advanced cyberinfrastructures (e.g., high-performance computing and virtual containers). This framework is used to couple the Water Balance Model (WBM) and SIMPLE-G (Baldos et al. 2020; Grogan et al. 2022; Woo et al. 2022). While land-use decisions are exogenous in WBM and hydrologic water supply is exogenous in SIMPLE-G, they are used as bridges between the models. To address the differences in temporal and spatial resolutions, additional post-processing tools are developed to transfer the information at consistent scales.

2 Building a Baseline

Developing a baseline is critical for evaluating model outcomes and it serves as a point of reference for assessing the impact of policy interventions and/or different assumptions in model parameters. A comparison of baseline simulation results with a simulation based on different parameters or shocks and provides insights into the robustness of model predictions.

For counterfactual (what-if) analysis, a baseline is a reference point against which the simulation results of a scenario can be compared. By comparing the results of a model simulation to a baseline, it is possible to assess the impacts of different policy interventions, technological innovations, and climate change on the food system and agricultural water and land use.

Typically, the SIMPLE baseline captures changes in global food production and resource use given a trend in macroeconomic variables such as population and income growth as well as technological change. A baseline usually includes different regional rates for the drivers of global change. For example, it can include region-specific population growth rates as well as differentiated rates of income and technology growth.

There are many ways to define a model baseline. One typical baseline is the business-as-usual scenario, in which current trends in population, income, and technology are assumed to continue into the future. A baseline scenario could be based on historical data, projections of future trends, or a combination of both. The specific approach used will depend on the research question being addressed.

2.1 Historical Baseline

In the context of model validation, a historical baseline replicates the observed state of the system over a historical period. This baseline serves as a reference point for comparing model predictions with observations to assess the accuracy and reliability of the model. For SIMPLE, the historical baseline involves simulating the observed changes in crop production, cropland, and prices, given observed changes in the key drivers of food supply and demand.

2.2 Future Baseline

In SIMPLE and SIMPLE-G, the future baseline scenario typically requires inputs on population and per capita incomes, feedstock demand for biofuel use, and productivity growth trends for the crops, livestock, and processed food sectors. Other drivers can include future impacts of climate change on crop productivity as well as cropland supply shifters due to urbanization or regional demands for ecosystem services.

Constructing a future baseline involves gathering data from various sources to project future trends and dynamics within the system. Population, income, and biofuel production can be specified to follow long-run growth scenarios such as the Shared Socioeconomic Pathways (O'Neill et al. 2014) or other global economic projections.

2.3 Policies and Scenarios

The baseline serves as a foundation for analyzing the potential impacts of future technological advancements and climate change on the food system. Another use of the baseline is to assess the impacts of future sustainability policies, trade policies, or agricultural protection policies.

Constructing future baselines and alternative scenarios requires convergence science and multidisciplinary collaborations. For example, hydrological model simulations may be used to inform SIMPLE-G about the future availability of water or possible water sustainability policies. Agronomic crop model simulations may be used to inform SIMPLE-G about future yield trends for irrigated and rainfed crops. Additionally, population dynamic models can inform about future changes in food demand and the labor force. For some examples of baseline construction, see Chap. 10; Chap. 12 develops a groundwater sustainability scenario.

3 Summary

The process of going from complex ideas and theories to practical and useful conclusions is filled with difficulties and obstacles. This requires highly non-linear computable models of linked human and environmental systems. The high dimensionality of these models, large data requirements, and computational hurdles necessitate careful trade-offs between realism, tractability, and policy relevance. A theoretically simplified yet spatially detailed model like SIMPLE-G offers promising computational prospects for understanding relationships within environmental and human systems. In addition, SIMPLE-G computational method demonstrates the potential to overcome computational barriers of solving large-scale multi-system geospatial models.

The feasibility and efficiency of SIMPLE-G computation relies on condensation and linearization, two important techniques used in GEMPACK to reduce the complexity and size of the model equations and to solve them efficiently. Condensation involves simplifying the model by eliminating some equations and variables through substitution, which reduces the number of equations and variables that need to be solved. This makes the model more manageable and faster to run. On the other hand, linearization involves reformulating the model equations as a system of first-order partial differential equations that are linear in percentage changes or changes in variables. This allows for the use of a solution method which calculates the movements in the endogenous variables away from their initial values in response to movements in exogenous variables away from their initial values. These two techniques improve the computational efficiency and accuracy of the model solution.

Moving forward, continued research efforts focused on developing efficient solution algorithms, harnessing new data sources, and enhancing model interpretability are key to bridging the gap between the inherent complexity of these models and their practical utility in tackling sustainability challenges. Ultimately, we can harness the power of multi-scale multi-system models to better understand the interactions between economy, environment, and society, enabling informed and sustainable decision-making at all levels.

Acknowledgment and Competing Interests The authors acknowledge support from the U.S. Department of Energy, Office of Science, Biological and Environmental Research Program, Earth and Environmental Systems Modeling, MultiSector Dynamics under Cooperative Agreement DE-SC0022141; the National Science Foundation award #2118329: "NSF Institute for Geospatial Understanding through an Integrative Discovery Environment (I-GUIDE)," the United States Department of Agriculture AFRI grant #2019-67023-29679, "Economic Foundations of Long Run Agricultural Sustainability," and the National Science Foundation INFEWS award #1855937, "Identifying Sustainability Solutions Through Global-Local-Global Analysis of a Coupled Water-Agriculture-Bioenergy System."

The findings and conclusions presented in this chapter are those of the authors and should not be construed to represent any official determination or policy of the US Department of Agriculture (USDA), the US government, the Department of Energy (DOE), or the National Science Foundation (NSF). Furthermore, we declare that there is no conflict of interest related to this work.

References

Bach, Christian Friis, and Ken Pearson. 1996. *Implementing quotas in GTAP using GEMPACK or how to linearize an inequality*. Technical Paper 4. Department of Agricultural Economics, Purdue University, West Lafayette: Global Trade Analysis Project (GTAP). https://doi.org/10.21642/GTAP.TP04.

Baldos, Uris Lantz C., Iman Haqiqi, Thomas W. Hertel, Mark Horridge, and J. Liu. 2020. SIMPLE-G: A multiscale framework for integration of economic and biophysical determinants of sustainability. *Environmental Modelling & Software* 133: 104805. https://doi.org/10.1016/j.envsoft.2020.104805.

Grogan, Danielle S., Shan Zuidema, Alex Prusevich, Wilfred M. Wollheim, Stanley Glidden, and Richard B. Lammers. 2022. Water balance model (WBM) v.1.0.0: A scalable gridded global hydrologic model with water-tracking functionality. *Geoscientific Model Development* 15: 7287–7323. https://doi.org/10.5194/gmd-15-7287-2022.

Haqiqi, Iman, Laura C. Bowling, Sadia A. Jame, Thomas W. Hertel, Uris Baldos, and Jing Liu. 2018. Global drivers of land and water sustainability stresses at mid-century. *Purdue Policy Research Institute (PPRI) Policy Briefs* 4: 7.

Harrison, W. Jill, J. Mark Horridge, and K.R. Pearson. 2000. Decomposing simulation results with respect to exogenous shocks. *Computational Economics* 15: 227–249. https://doi.org/10.1023/A:1008739609685.

Horridge, J.M., Michael Jerie, Dean Mustakinov, and Florian Schiffmann. 2018. *GEMPACK manual*. Victoria University, Centre of Policy Studies/IMPACT Centre.

Liu, Jing, Thomas W. Hertel, Laura Bowling, Sadia Jame, Christopher Kucharik, and Navin Ramankutty. 2018. Evaluating alternative options for managing nitrogen losses from corn production. *Purdue Policy Research Institute (PPRI) Policy Briefs* 4: 9.

O'Neill, Brian C., Elmar Kriegler, Keywan Riahi, Kristie L. Ebi, Stephane Hallegatte, Timothy R. Carter, Ritu Mathur, and Detlef P. van Vuuren. 2014. A new scenario framework for climate change research: The concept of shared socioeconomic pathways. *Climatic Change* 122: 387–400. https://doi.org/10.1007/s10584-013-0905-2.

Pearson, K.R. 1991. *Solving nonlinear economic models accurately via a linear representation. Preliminary Impact Paper IP-55*. Parkville: University of Melbourne Impact Research Centre. https://doi.org/10.22004/ag.econ.295068.

Schiffmann, Florian, and Michael Jerie. 2019. Improving the performance of sparse LU decomposition in GEMPACK. In Warsaw, Poland.

Woo, Jungha, Lan Zhao, Danielle S. Grogan, Iman Haqiqi, Richard Lammers, and Carol X. Song. 2022. C3F: Collaborative container-based model coupling framework. In *PEARC '22: Practice and experience in advanced research computing*. Boston: Association for Computing Machinery. https://doi.org/10.1145/3491418.3530298.