Combining Large-scale Sensitivity Analysis in Computable General Equilibrium Models with Machine Learning:

An Example Application to policy supporting the bio-economy

By Wolfgang Britz, Jingwen Li and Linmei Shang, Institute for Food and Resource Economics, University of Bonn

- DRAFT: NOT FOR PUBLICATION -

Abstract

Policy design nowadays needs to consider goals related to multiple sustainability dimensions simultaneously. The Tinbergen principle suggests addressing each issue individually with a targeted policy measure. But each single resulting command-and-control, incentive or market-based policy is likely to affect also other policy goals, questioning overall policy coherence (cf. May et al. 2006). CGE modelling can contribute here by quantifying market-mediated impacts of changing policies against the given benchmark, considering key policy indicators such as income and its distribution, Green House Gas Emissions or land cover changes. But each single experiment considers one specific policy mix, from a public choice decision space which is extremely large, once we consider a larger set of policy domains (general economic and social policy, different environmental fields etc.). The growing interaction of global value chains implies that regional policy choices also increasingly affect sustainability outcomes elsewhere, a viewpoint increasingly addressed in policy impact assessments as well. Finally, impacts of policy choice in each region also depend on the policy chosen in others.

We combine large-scale sensitivity analysis, changing both policy instruments and key model parameters in different regions, focusing on key sustainability metrics, and fit a neural network to the results. The considered policy instruments are indirect tax rates changes relating to bio-economy sectors, while land supply and different substitution elasticities are subject simultaneously to sensitivity analysis. We find a very good fit to cases where only policy instruments are changed and still quite high once when also parameters are changed. We conclude from there that machine learning techniques are able to provide robust meta-models of CGEs and can be used to predict or even optimize over the response space of the CGE.

1. Introduction

Global warming adds new dimensions to global economic and political concerns related to welfare, inequality and international collaboration, and promoting bioeconomy has become a strategy for sustainable economic development in both developed and developing countries[1]. However, policy design related to Bioeconomy nowadays faces the challenge of dimensionality and conflicting goals.

The growing interaction of global value chains implies that regional policy choices increasingly affect sustainability outcomes elsewhere (Fuchs et al., 2020). Technology transfer and R&D support need to consider the linkages of local and global value chains and whether the bio-based components can be price competitive to enter the market (O'Brien et al., 2017; Sheppard et al., 2011; Zörb et al., 2018). International institutions, governments, and stakeholders at different positions on the global bio-based value chain may put different priorities for monitoring the sustainability of bioeconomy (Egenolf and Bringezu, 2019; Zörb et al., 2018). Individual countries pursue different national strategies to foster and develop a sustainable bioeconomy reflecting their national policy goals, which may however result in conflicts of goals at global level (Bracco et al., 2018; Dietz et al., 2018). When designing bioeconomy policies, multiple sustainability dimensions need to be considered simultaneously, including global spillover effects. The Tinbergen principle suggests addressing each issue individually with a targeted policy measure, but each single resulting command-and-control, incentive or market-based policy is likely to affect other policy goals, questioning overall policy coherence (May et al., 2006). Individual measures, such as promoting biomass use in a single industry, are unable to address multi-objective policy goals such as the UN-endorsed Sustainable Development Goals (SDGs) (Issa et al., 2019; Schütte, 2018), which relate to all three pillars of sustainability (Asada et al., 2020; Egenolf and Bringezu, 2019; O'Brien et al., 2017). Furthermore, growing international intra-industry linkages render ex-ante impact assessments of policies challenging which are needed for a rational policy design.

Computable General Equilibrium (CGE) modelling can contribute here by quantifying market-mediated impacts of changing policies against the given benchmark. The merit of

[1] e.g. EU's 2018 action plan on bioeconomy strategy, US 2012 National Bioeconomy Blueprint, Brazil's 2015 Chemistry Industry Development Plan and China's 2012 Bioindustry Development Plan all express the goal to replace fossil-based products by bio-based products and thus mitigate climate change.

CGE analysis is that it captures economic, environmental and social linkages in the global economy, considering key policy indicators such as income and its distribution, Green House Gas (GHG) Emissions or land cover changes, all relevant aspect when analysis bio-economy policies. For example, Hochman and Zilberman (2018) apply CGE analysis to examine the role of bioethanol exports of US reducing the balance of trade deficit, and employment effects of the use of fossil fuels and renewables for energy consumption. Further examples of using CGE models to analyze impact of policies on land use change, feedstock and energy price, and GHG emissions provide Britz and Hertel (2011), Timilsina and Mevel (2011) Timilsina et al. (2010), at global level, and McPhail and Babcock (2012), Laborde and Valin (2012), Mosnier et al. (2013), Wianwiwat and Asafu-Adjaye (2013), Weng et al. (2019) at country/region level.

However, existing CGE analyses usually can conduct a quite limited number of scenarios, often considering changes in one region and a few sectors only. Each single scenario has to define one specific policy mix. When structural changes in policy across a larger set of policy domains (general economic and social policy, different environmental fields, etc.) are considered, this is daunting exercise given the extremely large public choice decision space. The overall policy mix can tax and subsidize various transactions to different degrees, restrict the set of technologies firms are able to use or the set of products which can be marketed. Ideally, one would optimize simultaneously all potential measures given a social welfare function which assesses trade-offs between many goals, such as overall GDP, its distribution, various environmental indicators, contribution to growth and equality in other countries, etc. Leaving the challenge to define this welfare function aside, it is likely computationally impossible to optimize hundreds of tax rates and more so restrictions to technology simultaneously in a global CGE model.

Furthermore, analysis should address uncertainties in model parameterization and structure. Global CGE models comprise thousands of behavioral and other parameters which jointly define the outcomes of a policy experiment, asking for evolved approaches to sensitivity analysis (Hertel et al., 2007; Mary et al., 2019). But researchers have also the choice to aggregate the database differently, which can have distinct impact on results (Britz and van der Mensbrugghe, 2016), to use databases with additional detail (e.g. GTAP-Power, Peters (2016); GTAP-Water, Haqiqi et al. (2016)) and to change the configuration of the model. CGEBox (Britz and van der Mensbrugghe, 2018), to give an example, allows to choose among different nestings in production functions and, to employ different assumptions with regard to international trade, such as a Melitz's model for heterogeneous firms (Jafari and Britz, 2018), or MRIO differentiation drawing on Carrico et al. (2020). A user can further add "sub-models", such as GTAP-AEZ (Lee, 2005) focusing on land use based on land cover data from (Baldos et al., 2017), GTAP-E (Burniaux and Truong, 2002) on energy use, GTAP-AGR on agriculture sectors (Keeney and Hertel, 2005), and myGTAP (Walmsley and Minor, 2013). Any changes in model configuration will likely impact the outcomes of a shock in varying degrees (Britz and van der Mensbrugghe, 2018).

In here, we propose to combine large-scale sensitivity analysis with machine learning (ML) techniques to assess indicator outcomes of a large set of model runs, considering alternative policy mixes and model parameterizations to identify "quasi-optimal" policy choices while accounting for model uncertainty. This draws on the established combination of a Monte Carlo analysis and statistical methods for systematic sensitivity analysis of CGE results. Here, either known or assumed distributions around the benchmark values of parameters/exogenous variables drive the Monte Carlo experiment to derive a distribution of result variables (Abler et al., 1999; Hertel et al., 2007), or Monte Carlo experiments with random parameters provide a start point from which correlations and key parameters are derived (Anthoff and Tol, 2013; Belgodere and Vellutini, 2011; Mary et al., 2019). However, as Mary et al. (2019) suggest, methods such as regressions have strict assumptions on linearity and monotonicity, likely no appropriate for the response surface of a global CGE model. Artificial neural networks (Goodfellow et al., 2016) are good at capturing nonlinear functions in high dimensional spaces. They have been successfully employed as surrogate models in many research fields (Sun and Wang, 2019; Villa-Vialaneix et al., 2012).

Our approach is designed as follows. We derive a large data set of outcome based on Monte Carlo base sensitivity analysis with a CGE model to train neural networks. These can later serve as surrogate models to examine the impact of bioeconomy policies on various economic and environmental outcomes. The trained neural network can provide accurate and quick responses, much faster than running the CGE model itself. The large dataset from the Monte Carlo experiment can also be used as a searching space for the "quasi-optimal" solution of policy design under a CGE framework.

This study consists of the following steps:

- We define, based on literature, a limited set of sustainability indicators derived from outcome variables from the CGE model;
- We employ the integrated CGE modeling framework CGEBOX (Britz and van der

Mensbrugghe, 2018) to conduct a large set of comparative static counterfactual experiments for different combinations of bioeconomy policies, considering simultaneously parameter uncertainty;

- We train a neural networks model based on the sampled CGE outcomes to predict jointly the sustainability indicators, and use SHAP values to select the relative important policy variables.

2. Sustainability and Policy Metrics

We build a sustainability metrics system immediately applicable for a standard global CGE model, where indicators are quantified for each model region, derived from endogenous variables of CGE model. Shocks are defined in Monte-Carlo analysis by random draws of potential bioeconomy policy variables for each region. We define a scenario set as a closed set in R^n, with $k \in R^n$ policy instrument parameters (e.g. subsidies and ad valorem tariff rate) and $n \in R^n$ parameters (e.g. elasticity of substitution between biofuel and fossil-fuel sectors in intermediate input use).

The neoclassical paradigm of comparative advantages suggests that free trade optimizes economic welfare for all participating countries. However, this concept is insufficient since it ignores externalities such as knowledge spillovers and pollution, and does not take distributional impacts into account (Van den Bergh, 2010). For example, the use of crops for biofuel production increases prices of agricultural commodities, with unwanted effects on food security or land use change. This study therefore proposes the concept of a "sustainable comparative advantage" which considers besides cost advantages in production also externalities, with regard to environmental status and human living standards. It is applied here to two main "novel" bioeconomy industries, namely bioplastics and biofuels. The main objective is to analyze the impact of a huge set of potential trajectories of increased output of these sectors in each model region. The output increases come about by policies interventions, such as taxing fossil substitutes, subsidized production or use, increasing government spending on research etc. Both the mix and the intensity of these policy measures are subject to sensitivity analysis, and vary in each scenario across regions. Linked to these interventions are changes in the sustainability metrics.

The system of indicators should be capable of three main tasks: 1) to monitor the status quo of the global bioeconomy development drawing on available bioeconomy value chain datasets for CGE analysis; 2) to examine the impact of bioeconomy policies on economic, environmental and social sustainability with the well-trained neural networks; and, 3) to identify the sustainable comparative advantage of the main players globally.

Strategies and concepts for the bioeconomy are quite diversely designed and discussed worldwide, and debates about transition pathways to a bioeconomy consider their likely economic, social, and environmental sustainability impacts (Asada et al., 2020; Egenolf and Bringezu, 2019; O'Brien et al., 2017). With regard to economic aspects, sustainable agricultural production and food security are discussed extensively, as novel biomass uses compete with food production and are likely to increase food prices (McPhail and Babcock, 2012). This price increases reflect also limited land availability, where remaining buffers compete for the food (and linked to this feed) demand of a growing global population and increasing intermediate input demand for novel bio-based value chain (Hertel et al., 2013; O'Brien et al., 2017). From an environmental perspective, the contribution of policies which substitute fossil energy by biomass to Climate Change mitigation is questioned. For example, emissions from indirect land-use change associated with bioethanol production may be underestimated (Carriquiry et al., 2020; Mosnier et al., 2013; Rajagopal and Plevin, 2013). Regarding social impacts, how bioeconomy affects employment and inequality is disputed. The growing global integration of bio-economy value chains distributes cost and benefits of regional bio-economy policies geographically. Value added might be created in regions where no or little environmental damage is provoked, while induced land use change can cause loss of natural habitats and of carbon stocks elsewhere (O'Brien et al., 2017). There is also a vivid debate whether and where the bioeconomy might create jobs, especially if this will benefit rural communities/developing countries (Hochman and Zilberman, 2018).

Efforts in recent years summarize the existing assessment tools, conceptually and empirically, in order to inform effective and comprehensive bioeconomy strategies. This aims at building metrics systems which quantify how bioeconomy policies perform globally in terms of the national and international sustainability objectives (D'Adamo et al., 2020; Egenolf and Bringezu, 2019; Escobar and Britz, 2021; Karvonen et al., 2017; O'Brien et al., 2017; Robert et al., 2020; Ronzon and M'Barek, 2018). Drawing on conceptual frameworks at global level, reliable datasets, empirical models and statistical methods jointly play crucial roles to develop indicators of sustainability which quantify the relationship between these indicators and target values (Khanna et al., 2018; O'Brien et al., 2017; Zilberman et al., 2018). Examples include: Footprints for monitoring natural resource use and environmental impacts (land, forest, water, and carbon emission, etc) with the arms of footprint network accounting, multiregional input-output (MRIO) tables (Budzinski et al., 2017; Liobikiene et al., 2020; O'Brien et al., 2015; O'Donoghue et al., 2019; Tukker et al., 2016), Social

Accounting Matrix(SAM) frameworks (Fuentes-Saguar et al., 2017) and CGE frameworks (Escobar and Britz, 2021); Food security covering issues on protein supplies, food price volatility and import dependency (Egenolf and Bringezu, 2019); Bioeconomy competitiveness related to contribution to GDP, employment (O'Brien et al., 2017) and regional specialization represented as (Ronzon and M'Barek, 2018)'s location quotient for EU countries, etc.

-Table 1 insert here-

As illustrated in Table 1, we choose a bundle of variables from the CGE model to construct our sustainability indicator systems, ranging from national economic development and bioeconomy comparative advantages (gdp_pc, xs_bio, va_bio, and LQ_bio), social aspects like inequality and food security (gdp_sd, wagegap, cal, ps, foodself), to environmental impacts including land use change and Green House Gas emissions (xAezNest, ghg).

Among the outcome indicators, some are immediately ready from CGE model since they serve as endogenous variables in the equation system, e.g. Calorie intake per capita (cal)[2], landuse in ha by categories (xAEZNest), and real GDP per capita (gdp_pc), etc. Others require a simple summation over bio-sectors, ooly, e.g. bioeconomy value added (va_bio) and output (xs_bio). Some others need further calculations, e.g. bioeconomy location quotient for each country (LQ_bio) is calculated as a ratio of the proportion of people employed in bioeconomy in each country to the proportion of people employed in bioeconomy in all countries (Ronzon and M'Barek, 2018).

We aggregate the countries and regions in the GTAP Data Base Version 10 Aguiar et al. (2019) into 31 model regions which are partly individual countries (Table 2). We consider additionally established "novel" bio-economy value chains (ethanol, biodiesel and bio-based plastics) and consider policies of tax incentives and R&D support. Governments can subsidize the three bioeconomy commodities by changing indirect tax rates. Second, we pick two technology shift parameters, factor productivity of land by agriculture activities (techland), and labor productivity in bioeconomy sectors (techlabor). The elasticity of substitution between bio-based sectors and fossil-based sectors could be essential when examining the bioeconomy value chains, since they decide how easy it is to substitute inputs

[2] CGEBox calculate the calories, protein and fat content per food product from the FAO food balance sheets, and recalculate the daily per capita demands to avoid double counting caused by different accounting scheme of FAO.

in production, e.g. biofuels with fossil fuels. Thus, we select the elasticity parameter for plastics (σplas) and fuels (σfuel) as they are two represented bioeconomy today. Later on, we disaggregate these two sectors from total intermediate input in our CGEBox framework. Similarly, the elasticity of transformation between forest and agriculture land (ωAEZtop) can be crucial to the land use change and emission outcome. Land supply elasticity (ηf) which governs the total land supply with respect to average returns to land is also included in the sensitivity analysis.

3. CGE Modelling Framework

CGEBox is an open source framework to implement CGE models, its core based on GTAP Standard model and dataset (Hertel, 1997; Van der Mensbrugghe, can be extended to a range of modules[3]. We use a database derived from GTAP Version 10 (Aguiar et al., 2019) with 31 model regions (partially large individual countries) and full sectoral break-down, disaggregated further to consider ethanol, biodiesel (Taheripour et al., 2017), and bio-based plastics and more details for primary agriculture, in total featuring 64 sectors. The model departs from the comparative static GTAP Standard Model (Hertel, 1997), as implemented for Version 7 by (Britz and van der Mensbrugghe, 2018) in GAMS, adding a specific variant of the GTAP-AEZ (Lee, 2005) module together with CO2 and Non-CO2 accounting in order to catch the impact on land use and emissions from bioeconomy promotion policies. CO2 and Non-CO2 emissions data are from GTAP database. CO2 accounting in CGEBox uses total benchmark emissions related to domestic and import demand by the Armington agents as parameters, and calculate emissions by updating these totals based on the change in import domestic demand relative to the benchmark demands

-Figure 1 insert here-

The GTAP-AEZ(Lee, 2005) module allocates the total land use to 18 Agro-Ecological-Zones (AEZ) in each region with nested Constant Elasticity Transformation(CET) functions (See Figure 1. Within each AEZ, the top nest distributes total managed land to forestry and agricultural uses with the elasticity of transformation parameter ωAEZtop differed from AEZs. Then a second nest further disaggregates agricultural uses to pasture and crop activities with the elasticity of transformation parameter ωAEZgraz_agr. The top two tiers

[3] A documentation of model equations and implementation of CGEBox can be found in (Britz, 2019), http://www.ilr.uni-bonn.de/em/rsrch/cgebox/cgeboxGUI.pdf

use a volume preserving form of the CET (van der Mensbrugghe and Peters, 2016). A third tier distributes pasture to different animal activities and cropland to different crop activities. While there is no substitution among AEZs, a land stock equation with a supply elasticity ηf is used as in the standard GTAP model. Total land supply increase by the same proportion in all AEZs, driving by the numeraire normalized average returns to land in each country/region.

-Figure 2 insert here-

Based on the nesting of bioplastics and other chemicals from Escobar et al. (2018), Nong et al. (2020) and Escobar and Britz (2021) which assess impacts on GDP and green-house gases of emerging biomass based plastic production, we extend the production functions with a nesting which allows substitution between different biofuels and other fuels sectors with derived split-factors from the GTAPBIOF database (Taheripour et al., 2007). The standard GTAP database does not have a separate fuel sector, which means gasoline and diesel are part of the "p_c Petroleum & Coke: manufacture of coke and refined petroleum products". Thus, we substitute biofuels with the GTAP sector "p_c" for all activities' intermediate input demand in order to capture the innovation and promotion policies for a wider range of bioeconomy value chain, namely Ethanol, Bio-Diesel, and Bioplastics (Figure 2).

4. Systematic Sensitivity Analysis

4.1 Sampling process

We perform an unconditional systematic sensitivity analysis (SSA) by allowing all possible random variable combinations, rather than the traditional sensitivity analysis when only one parameter is permitted to vary from each run – which brings up the "curse of dimensionality". Given the high dimensionality of the design space (e.g. in our case we have different levels of subsidization of selected sectors in different countries, different R&D investments transformed to technology advancement in different primary and bioeconomy sectors), ten or even hundreds of thousands of runs are required to consider appropriately the overall policy design space. It is almost impossible to do this without exploiting parallelism for model running.

-Figure 3 insert here-

Figure 3 illustrates our steps of sampling process. The batch-mode facility of CGEBox can manage queued execution of experiments, solved in parallel buckets. Hence, we use several computing servers in parallel, and on each server, shocks are run in parallel at the same time. A shock file first chooses some regions randomly out of the 31 ones and then defines the ranges of uniform distributions for shocked exogenous variables. The batch-mode facility creates loops on each random shock, and we solve between four and ten shocks in parallel depending on the server. This brings the time for the average solve down to around 5s, when considering the use of multiple servers. A post-model file is used to calculate and store the only selected variables as shown in Table 1.

The Monte Carlo sampling process starts from an experiment which involves only the four policy variables: 2 tax shifters, dintx and mintx, which are drawn i.i.d from a uniform distribution between -5% and -10%[4]; 2 technology shifters, techland and techlabor, which are draw from a uniform distribution between 5% and 20%. The second-round experiment adds four elasticity parameters into the shock: ηf, ωAEZtop, σplas and σfuel, ranging from 50% decrease to 50% increase of their original level[5]. Accordingly, we trained two separate Neural Networks for both experiments.

After every ten thousand solves, we collect data from single runs and combine the results in a summary sample which is added to the training set of the surrogate model, to check if the neural networks can be well trained. A linear mapping of the stored variables is required since they may have different, and even more than two dimensions. Although we aggregate most of the indicator variables to the regional level and vary the policy variable and parameters only by region, we still arrive at 430 endogenous variables and 248 exogenous variables, which needed to be stored from each single run.

4.2 Machine learning and SHAP values

After gathering all the results from the CGE model experiments, we train neural networks to predict the multiple measurable outcome variables and use SHAP values to express their relative importance.

Artificial neural networks, belonging to the field of machine learning and deep learning, are mathematical models that mimic the fundamental structure and function of the biological nervous systems (Goodfellow et al., 2016). It has been shown that neural networks are well

[4] Subsidization is realised as a negative indirect tax shifter in CGEBox.

[5] The original omegaAEZtop for most of the AEZs equal to 0.2, while in some AEZs forest and agriculture land are perfect substitutes. We only change those that are not infinitely large. The original ηf is set as 0.25. σplas and σfuel equal to 10 and 0.5 originally.

suited to capture nonlinear functions in high dimensional spaces (Labb´e et al., 2009). A fully connected neural network usually consists of an input layer, multiple hidden layers and an output layer, and every layer contains nodes (i.e. neurons). The input layer consists of data that represents the problem (i.e. inputs or features) and the output layer shows the responses (i.e. outputs or targets to be predicted) with regard to the problem. Hidden layers, located between the input layer and output layer, perform nonlinear transformation of the inputs entered into the network and sequentially forward the information to the output layer. The connections between the neurons in a network are assigned with some weights which get updated during the training process based on the minimization of a loss function.

The training is conducted using the tensorflowKeras Python package (Abadi et al., 2016). The performance criteria of our neural networks is the mean R2 (calculated by the r2_score function in the sklearn Python package) over all outputs, since all of our outputs are of continuous type. Once the neural network is well-trained, we apply SHapley Additive exPlanations (SHAP) value to express feature importance (Lundberg and Lee, 2017), i.e. the importance of each input variable in predicting output variables. The SHAP value draws on a game theoretic approach to explain the output of a machine learning model. It is developed based on Shapley values (Shapley, 1953), and explains the prediction of an instance by computing the contribution of each feature to the prediction. The features of an instance act as players in a coalition. The prediction of this instance is regarded as the "payout" for all the "players" together, and SHAP method tries to distribute the "payout" to each "player" fairly by a permutation method.

We employ the KernelExplainer in the shap Python package to compute the SHAP values of each feature. The calculation of SHAP values by KernelExplainer comprise five steps (Molnar, 2020). First, sample coalitions are generated. Each sample is a coalition vector consisting K elements in corresponding to K features. An entry of 1 means that the corresponding feature is "present" and 0 that is "absent". For example, the vector (1, 1, 0, 0) means that the first and second features are present, while the other two features are absent. Second, the machine learning model predicts for each sample coalition by first converting this coalition to the original feature space. Here, a function is needed to convert the generated sample coalitions into valid instances. This function maps 1's to the corresponding value from the instance x that we want to explain, and it maps 0's to the value of another random instance that we sample from the data. Third, the algorithm computes the weight for each sample coalition with the SHAP kernel. SHAP weights the sampled instances according to the weight the coalition would get in the Shapley value estimation. Small coalition (few 1's) and large coalitions (many 1's) get the largest weights. The intuition behind this is: We learn most about the individual features if we can study their effects in isolation. The two extreme useful cases would be that the instance contains only this feature, or it misses only this feature. The weighting function of each instance can be found in (Lundberg and Lee, 2017). Fourth, the process fits the weighted linear model, based on the prediction and weight of each sample instance. Last, it returns Shapley values, i.e. the coefficients from the linear model.

Then SHAP repeats the same procedure for each instance from the original data set. delivering the feature importance of each feature in each instance. The importance of a feature across the whole data set can be measured by the mean absolute Shapley values.

5. Results

5.1 Performance of neural networks

The first neural network takes 124 inputs, which consider four categories i.e. taxdintx, taxmintx, techlabor and techland of all regions, and predicts 367 outputs including cal, gdp, and ghg of all regions. We used a one-shot sampling approach and fed all generated sample points (sample size = 10110), and 10% of them were used as an out-of-sample test data set. Our final selected neural network has one hidden layer with 700 neurons. The model was trained for 150 epochs with a mini-batch size of 16, Adam optimizer (default learning rate of 0.01), and the activation function of Rectified Linear Unit (ReLU) in each layer. This first model has achieved a very good performance on both the training (average R^2 = 0.997) and the test set (average R^2 =0.997).

The second neural network takes 248 inputs, considering different kinds of policy instruments and parameters, and predicts 430 outputs. Using the same hyperparameters with the first neural network, the performance of the second neural network is not as good as the first neural network: the average R2 values of the training and tests set are 0.908 and 0.911, respectively. We also find that the performance of this neural network is not improved when increasing the size of training set (Figure 7). This shows that this neural network cannot capture the underlying relationship when it has to consider additional parameter uncertainty, hinting at quite high and hard to predict sensitivity of the outcomes to these parameters. Changes in those parameters seem to lead to a fundamental change of the relationship between model inputs and outputs.

Thus, we further fine-tuned the second neural network. Due to the different randomization process within the training, we obtained a model that achieves a relatively good performance with average R^2 of 0.971 on of 0.965 on the training set, although the only hidden layer still has 700 neurons. This model was trained for 300 epochs instead of 150 epochs because we noticed that the model could be improved just by training for more epochs.

5.2 Results of SHAP values

-Figure 4 insert here-

-Figure 5 insert here-

Figure 4 shows the SHAP values of different features of the first neural network. We can see that features related to land efficiency (techland) are contributing the most to explaining the overall outputs, especially in China, EU28 and USA, followed by Indonesia, India and Rest of Latin America. On the contrary, the overall contributions of subsidies (taxdintx and taxmintx) and labor efficiency (techlabor) are not clearly distinguishable from each other. However, USA is leading the importance of domestic subsidies in explaining the overall outputs. Figure 5 confirms this result after we add all the parameter uncertainties into our experiment. Land efficiency still is the most important feature in explaining the overall performance of sustainability metrics. The impact of subsidies is overall small, but the SHAP value of domestic subsidies increase in the second neural network.

-Figure 6 insert here-

Figure 6 shows the SHAP values of the second neural network for all the elasticity parameters. It is clear that the land supply elasticity (ηf) plays the most important role. Similar to policy variables, EU28, China, USA are the top three when considering the importance of the land supply elasticity, followed by India, Indonesia, and Rest of Latin America.

The elasticity of substitution parameters between biofuels and fossil fuels (σfuel) have relatively high SHAP values, meaning that the neural network attributes relative high weights to them when predicting outcome variables. The elasticity of transformation parameters between forest land and agricultural land have very small SHAP values for some countries, while the elasticity of substitution in plastics industry (σplas) have SHAP values close to 0 for all regions.

6. Conclusion

CGE models have great potentials to inform even complex policy designs when combing them with statistic methods and machine learning models. To overcome the limitation that the counterfactual experiment in CGE models can only define one specific policy mix within each scenario, we propose a novel method which combines large-scale sensitivity analysis with machine learning. Based on our sustainability indicators derived from outcome variables of the CGE model, we conduct Monte Carlo sampling changing both policy variables and key model parameters. Based on surrogate neural network models, we can quite accurately and fast predict the sustainability outcomes of any policy mix under different model parameterization without running our CGE model. Based on SHAP values, we could identify the most sensitive parameters. In our application, the first, somewhat simpler neural network gives excellent prediction after 10,100 samples. The second neural network which also considers parameter uncertainty achieves quite well predictions with R2s greater than 99% for most of the outcome variables such as GDP per capita, inequalities, food securities, emissions and land use changes, except for bioeconomy growth (ss shown from Figure 8 – 12 in Appendices). The SHAP value suggest that changes in land productivity triggered by increased R&D spendings is the most important policy driver for sustainability outcomes, while the land supply elasticity is the most sensitive parameter.

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., et al. (2016). Tensorflow: A system for large-scale machine learning. In *12th {USENIX} symposiumon operating systems design and implementation ({OSDI} 16)*, pages

265–283.

Abler, D. G., Rodriguez, A. G., and Shortle, J. S. (1999). Parameter uncertainty in cge modeling of the environmental impacts of economic policies. *Environmental and Resource Economics*, 14(1):75–94.

Aguiar, A., Chepeliev, M., Corong, E. L., McDougall, R., and van der Mensbrugghe, D. (2019). The gtap data base: version 10. *Journal of Global Economic Analysis*, 4(1):1–27.

Anthoff, D. and Tol, R. S. (2013). The uncertainty about the social cost of carbon: A decomposition analysis using fund. *Climatic change*, 117(3):515–530.

Asada, R., Cardellini, G., Mair-Bauernfeind, C., Wenger, J., Haas, V., Holzer, D., and Stern, T. (2020). Effective bioeconomy? a mriobased socioeconomic and environmental impact assessment of generic sectoral innovations. *Technological Forecasting and Social Change*, 153:119946.

Baldos, U. L. et al. (2017). Development of gtap version 9 land use and land cover database for years 2004, 2007 and 2011. Technical report, Center for Global Trade Analysis, Department of Agricultural Economics .

Belgodere, A. and Vellutini, C. (2011). Identifying key elasticities in a cge model: a monte carlo approach. *Applied Economics Letters*, 18(17):1619–1622.

Bracco, S., Calicioglu, O., Gomez San Juan, M., and Flammini, A. (2018). Assessing the contribution of bioeconomy to the total economy: A review of national frameworks. *Sustainability*, 10(6):1698.

Britz, W. (2019). Cgebox–a flexible and modular toolkit for cge modelling with a gui. Institute for Food Resource economics, version November.

Britz, W. and Hertel, T. W. (2011). Impacts of eu biofuels directives on global markets and eu environmental quality: An integrated pe, global cge analysis. Agriculture, Ecosystems & Environment, 142(1-2):102– 109.

Britz, W. and van der Mensbrugghe, D. (2016). Reducing unwanted consequences of aggregation in large-scale economic models-a systematic empirical evaluation with the gtap model. *Economic Modelling*, 59:463–472.

Britz, W. and van der Mensbrugghe, D. (2018). Cgebox: A flexible, modular and extendable framework for cge analysis in gams. *Journal of GlobalEconomic Analysis*, 3(2):106–177.

Budzinski, M., Bezama, A., and Thr¨an, D. (2017). Monitoring the progress towards bioeconomy using multi-regional input-output analysis: The example of wood use in germany. Journal of Cleaner Production, 161:1–11.

Burniaux, J.-M. and Truong, T. P. (2002). Gtap-e: an energy-environmental version of the gtap model. *GTAP Technical Papers*, page 18.

Carrico, C., Corong, E., and van der Mensbrugghe, D. (2020). The gtap version 10a multi-region input output (mrio) data base.

Carriquiry, M., Elobeid, A., Dumortier, J., and Goodrich, R. (2020). Incorporating sub-national brazilian agricultural production and land-use into us biofuel policy evaluation. *Applied Economic Perspectives and Policy*, 42(3):497–523.

D'Adamo, I., Falcone, P. M., and Morone, P. (2020). A new socio-economic indicator to measure the performance of bioeconomy sectors in europe. *Ecological Economics*, 176:106724.

Dietz, T., Börner, J., Förster, J. J., and Von Braun, J. (2018). Governance of the bioeconomy: A global comparative study of national bioeconomystrategies. *Sustainability*, 10(9):3190.

Egenolf, V. and Bringezu, S. (2019). Conceptualization of an indicator system for assessing the sustainability of the bioeconomy. *Sustainability*,11(2):443.

Escobar, N. and Britz, W. (2021). Metrics on the sustainability of regionspecific bioplastics production, considering global land use change effects. *Resources, Conservation & Recycling, in press*.

Escobar, N., Haddad, S., Börner, J., and Britz, W. (2018). Land use mediated ghg emissions and spillovers from increased consumption of bioplastics. *Environmental Research Letters*, 13(12):125005.

Fuchs, R., Brown, C., and Rounsevell, M. (2020). Europe's green deal offshores environmental damage to other nations.

Fuentes-Saguar, P. D., Mainar-Causapé, A. J., and Ferrari, E. (2017). The role of bioeconomy sectors and natural resources in eu economies: A social accounting matrix-based analysis approach. *Sustainability*, 9(12):2383.

Goodfellow, I., Bengio, Y., Courville, A., and Bengio, Y. (2016). *Deep learning*, volume 1. MIT press Cambridge.

Haqiqi, I., Taheripour, F., Liu, J., and van der Mensbrugghe, D. (2016). Introducing irrigation water into gtap data base version 9. *Journal ofGlobal Economic Analysis*, 1(2):116–155.

Hertel, T., Hummels, D., Ivanic, M., and Keeney, R. (2007). How confident can we be of cge-based assessments of free trade agreements? *Economic Modelling*, 24(4):611–635.

Hertel, T., Steinbuks, J., and Baldos, U. (2013). Competition for land in the global bioeconomy. *Agricultural Economics*, 44(s1):129–138.

Hertel, T. W. (1997). Global trade analysis: modeling and applications.

Cambridge university press.

Hochman, G. and Zilberman, D. (2018). Corn ethanol and us biofuel policy 10 years later: A quantitative assessment. *American Journal of Agricultural Economics*, 100(2):570–584.

Issa, I., Delbrück, S., and Hamm, U. (2019). Bioeconomy from experts' perspectives–results of a global expert survey. *PloS one*, 14(5):e0215917.

Jafari, Y. and Britz, W. (2018). Modelling heterogeneous firms and nontariff measures in free trade agreements using computable general equilibrium. *Economic Modelling*, 73:279–294.

Karvonen, J., Halder, P., Kangas, J., and Leskinen, P. (2017). Indicators and tools for assessing sustainability impacts of the forest bioeconomy. *Forest ecosystems*, 4(1):2.

Keeney, R. and Hertel, T. (2005). Gtap-agr: A framework for assessing the implications of multilateral changes in agricultural policies. *GTAP Technical Papers*, page 25.

Khanna, M., Swinton, S. M., and Messer, K. D. (2018). Sustaining our natural resources in the face of increasing societal demands on agriculture: Directions for future research. *Applied Economic Perspectivesand Policy*, 40(1):38–59.

Labbé, B., Hérault, R., and Chatelain, C. (2009). Learning deep neural networks for high dimensional output problems. In *2009 International Conference on Machine Learning and Applications*, pages 63–68. IEEE.

Laborde, D. and Valin, H. (2012). Modeling land-use changes in a global cge: assessing the eu biofuel mandates with the mirage-biof model. *Climate Change Economics*, 3(03):1250017.

Lee, H.-L. (2005). The gtap land use data base and the gtape-aez model: incorporating agro-ecologically zoned land use data and land-based greenhouse gases emissions into the gtap framework.

Liobikiene, G., Chen, X., Streimikiene, D., and Balezentis, T. (2020). The trends in bioeconomy development in the european union: Exploiting capacity and productivity measures based on the land footprint approach. *Land Use Policy*, 91:104375.

Lundberg, S. and Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *NIPS*.

Mary, S., Phimister, E., Roberts, D., and Santini, F. (2019). A monte carlo filtering application for systematic sensitivity analysis of computable general equilibrium results. *Economic Systems Research*, 31(3):404– 422.

May, P. J., Sapotichne, J., and Workman, S. (2006). Policy coherence and policy domains. *Policy Studies Journal*, 34(3):381–403.

McPhail, L. L. and Babcock, B. A. (2012). Impact of us biofuel policy on us corn and gasoline price variability. *Energy*, 37(1):505–513.

Molnar, C. (2020). *Interpretable Machine Learning*. Lulu. com.

Mosnier, A., Havl´ık, P., Valin, H., Baker, J., Murray, B., Feng, S., Obersteiner, M., McCarl, B. A., Rose, S. K., and Schneider, U. A. (2013). Alternative us biofuel mandates and global ghg emissions: The role of land use change, crop management and yield growth. *Energy Policy*, 57:602–614.

Nong, D., Escobar, N., Britz, W., and Börner, J. (2020). Long-term impacts of bio-based innovation in the chemical sector: A dynamic global perspective. *Journal of Cleaner Production*, 272:122738.

O'Brien, M., Schütz, H., and Bringezu, S. (2015). The land footprint of the eu bioeconomy: Monitoring tools, gaps and needs. *Land Use Policy*, 47:235–246.

O'Brien, M., Wechsler, D., Bringezu, S., and Schaldach, R. (2017). Toward a systemic monitoring of the european bioeconomy: Gaps, needs and the integration of sustainability indicators and targets for global landuse. *Land Use Policy*, 66:162–171.

O'Donoghue, C., Chyzheuskaya, A., Grealis, E., Kilcline, K., Finnegan, W., Goggins, J., Hynes, S., and Ryan, M. (2019). Measuring ghg emissions across the agri-food sector value chain: The development of a bioeconomy input-output model. *International Journal on Food System Dynamics*, 10(1):55–85.

Peters, J. C. (2016). Gtap-e-power: an electricity-detailed economy-wide model. *Journal of Global Economic Analysis*, 1(2):156–187.

Rajagopal, D. and Plevin, R. J. (2013). Implications of market-mediated emissions and uncertainty for biofuel policies. Energy Policy, 56:75–82. Robert, N., Giuntoli, J., Araujo, R., Avraamides, M., Balzi, E., Barredo,

J. I., Baruth, B., Becker, W., Borzacchiello, M. T., Bulgheroni, C., et al. (2020). Development of a bioeconomy monitoring framework for the european union: An integrative and collaborative approach. New Biotechnology, 59:10–19.

Ronzon, T. and M'Barek, R. (2018). Socioeconomic indicators to monitor the eu's bioeconomy in transition. *Sustainability*, 10(6):1745.

Schütte, G. (2018). What kind of innovation policy does the bioeconomy need? *New Biotechnology*, 40:82–86.

Shapley, L. S. (1953). A value for n-person games. *Contributions to the Theory of Games*, 2(28):307–317.

Sheppard, A. W., Gillespie, I., Hirsch, M., and Begley, C. (2011). Biosecurity and sustainability within the growing global bioeconomy. *Current Opinion in Environmental Sustainability*, 3(1-2):4–10.

Sun, G. and Wang, S. (2019). A review of the artificial neural network surrogate modeling in aerodynamic design. *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering*,233(16):5863–5872.

Taheripour, F., Birur, D., Hertel, T., and Tyner, W. (2007). Introducing liquid biofuels into the gtap data base.

Taheripour, F., Levano, L. M. P., and Tyner, W. (2017). Introducing first and second generation biofuels into gtap 9 data base.

Timilsina, G. R., Beghin, J. C., Van der Mensbrugghe, D., and Mevel, S. (2010). *The impacts of biofuel targets on land-use change and food supply: A global CGE assessment*. The World Bank.

Timilsina, G. R. and Mevel, S. (2011). Biofuels and climate change mitigation: a CGE analysis incorporating land-use change. The World Bank.

Tukker, A., Bulavskaya, T., Giljum, S., de Koning, A., Lutter, S., Simas, M., Stadler, K., and Wood, R. (2016). Environmental and resource footprints in a global context: Europe's structural deficit in resource endowments. *Global Environmental Change*, 40:171–181.

Van den Bergh, J. C. (2010). Externality or sustainability economics? *Ecological Economics*, 69(11):2047–2052.

Van der Mensbrugghe, D. (2018). The standard gtap model in gams, version

7. Journal of Global Economic Analysis, 3(1):1–83.

van der Mensbrugghe, D. and Peters, J. C. (2016). Volume preserving ces and cet formulations.

Villa-Vialaneix, N., Follador, M., Ratto, M., and Leip, A. (2012). A comparison of eight metamodeling techniques for the simulation of n2o fluxes and n leaching from corn crops. *Environmental Modelling & Software*, 34:51–66.

Walmsley, T. and Minor, P. (2013). Mygtap model: a model for employing data from the mygtap data application—multiple households, split factors, remittances, foreign aid and transfers.

Weng, Y., Chang, S., Cai, W., and Wang, C. (2019). Exploring the impacts of biofuel expansion on land use change and food security based on a land explicit cge model: A case study of china. *Applied Energy*, 236:514–525.

Wianwiwat, S. and Asafu-Adjaye, J. (2013). Is there a role for biofuels in promoting energy self sufficiency and security? a cge analysis of biofuel policy in thailand. *Energy Policy*, 55:543–555.

Zilberman, D., Gordon, B., Hochman, G., and Wesseler, J. (2018). Economics of sustainable development and the bioeconomy. *Applied Economic Perspectives and Policy*, 40(1):22–37.

Zörb, C., Lewandowski, I., Kindervater, R., Göttert, U., and Patzelt, D. (2018). Biobased resources and value chains. In *Bioeconomy*, pages 75–95. Springer, Cham.

Appendix
Tables and Figures

Category	Variables in CGEBox	Description
Outcome indicators		
Economic growth	Regional income	Real GDP per capita (gdp_pc)
	Domestic production	Bioeconomy sectors output (xs_bio)
	Value added	Bioeconomy value added (va_bio)
	Region competitiveness	Location quotient(Employment) for bioeconomy (LQ_bio)
Inequality	International inequality	Real GDP standard deviation of all regions (gdp_sd)
	Wage gap between labors	Wage ratio of skilled to unskilled labor (wagegap)
Food security	Protein security	Calorie intake per capita (cal)
	Food price	Weighted average supply price for food sectors (ps)
	Food self-sufficiency	Value of food imports in total imports (foodself)
Land use change	Land use by categories	Forest, grazing, agriculture (xAezNest)
	Land supply	Total land supply of all AEZs (xAezNest)
Emission	GHG emission	Aggregate CO2 and nonCO2 emissions (ghg)
Policy instruments		
R&D support	Education on bioeconomy	TFP increase in labor efficiency (techlabor)
	Raw material productivity	TFP increase for primary sectors on land (techland)
Tax incentive	Subsidies on bio sectors	Indirect tax on domestic consumption (dintx)
		Indirect tax on imports (mintx)
Parameter uncertainties		
Land supply	Total Land supply	Total used land stock supply elasticity (η_f)
	Transformation within AEZs	Between managed forest and agricultural landuse (ωAEZ_{top})
Technology nest	Elasticity of substitution	Between bioplastics and fossil fuel based plastics (σ_{plas})
		Between biofuels and fossil fuels (σ_{fuel})

Table 1: Outcome and policy variables from CGEBox act as dependent and independent variables in neural networks

arg	Argentina	zaf	South Africa
aus	Australia	tha	Thailand
blr	Belarus	tur	Turkey
bra	Brazil	ukr	Ukraine
can	Canada	usa	United States of America
chn	China	irn	Iran Islamic Republic of
EU28	EU28	formRus	Former Russian Federation
ind	India	GulfMidEast	Gulf and Middle East
idn	Indonesia	MedAfr	Mediterranean Africa
jpn	Japan	AfrOther	Rest of Africa
kor	Korea	RestEastAsia	Rest of East Asia
mys	Malaysia	RestofEurope	Rest of Europe
mex	Mexico	SouthMidAm	Rest of Latin America
nzl	New Zealand	RestSouthAsi	Rest of South Asia
nga	Nigeria	RestSEAsia	Rest of South East Asia
phl	Philippines		

Table 2: Aggregation of countries in CGEBox

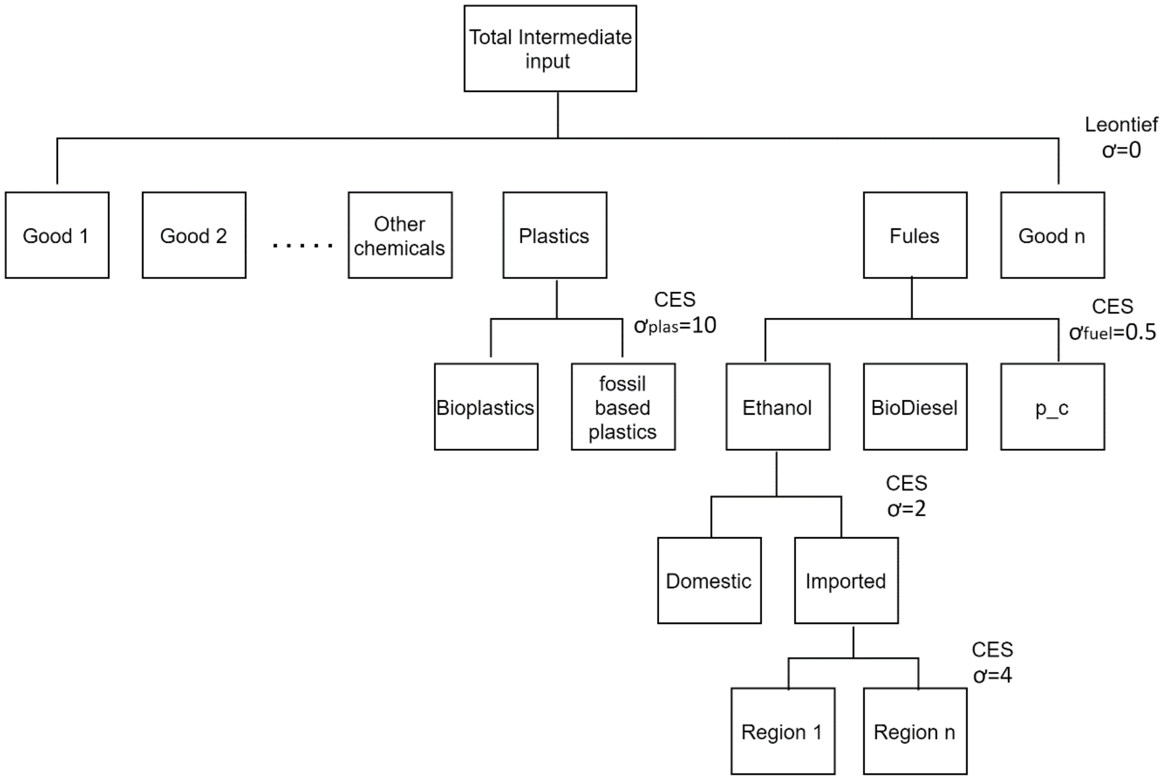

Figure 1: Each AEZ is disaggregated by CET with the elasticity of substitution ωAEZ

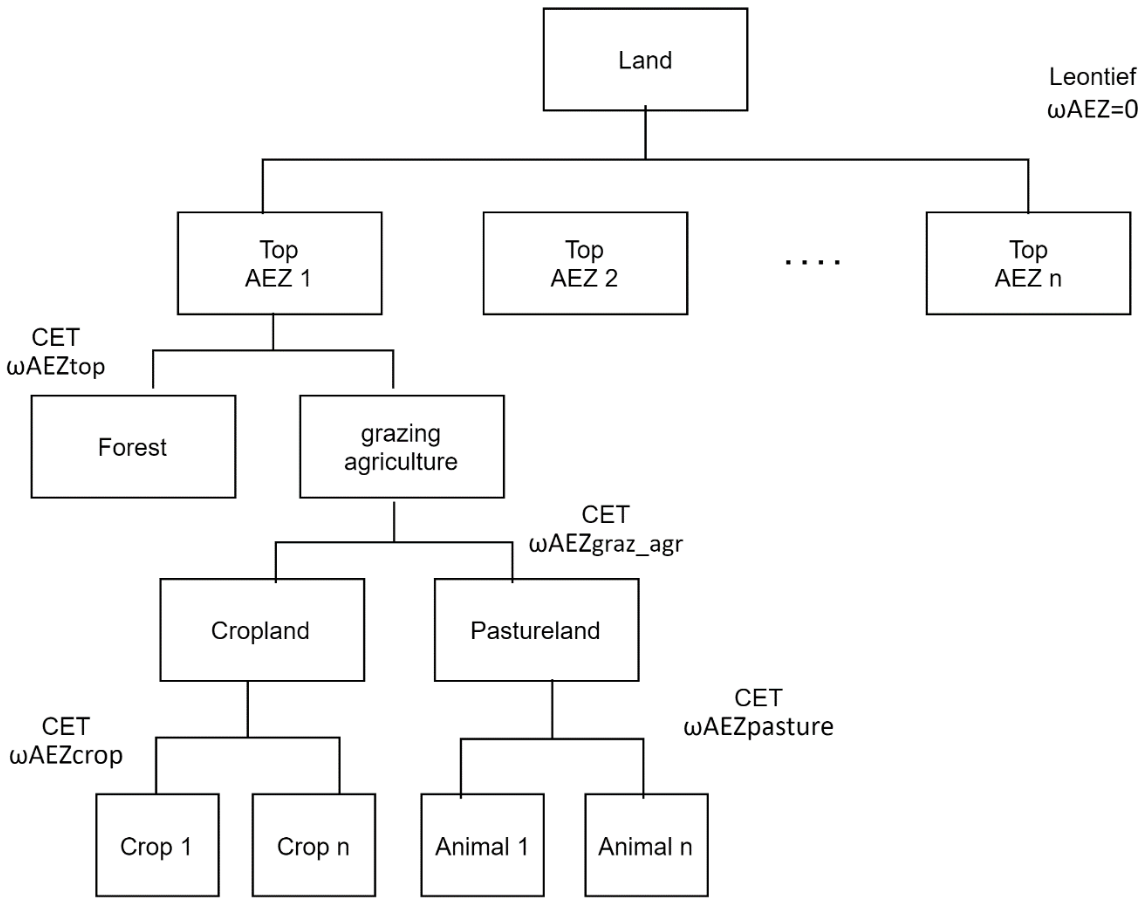

Figure 2: The total intermediate input is nested by Leontief function, all other levels of nestings use Constant Elasticity Supply (CES) function, with the elasticity of substitution denoted as σ.

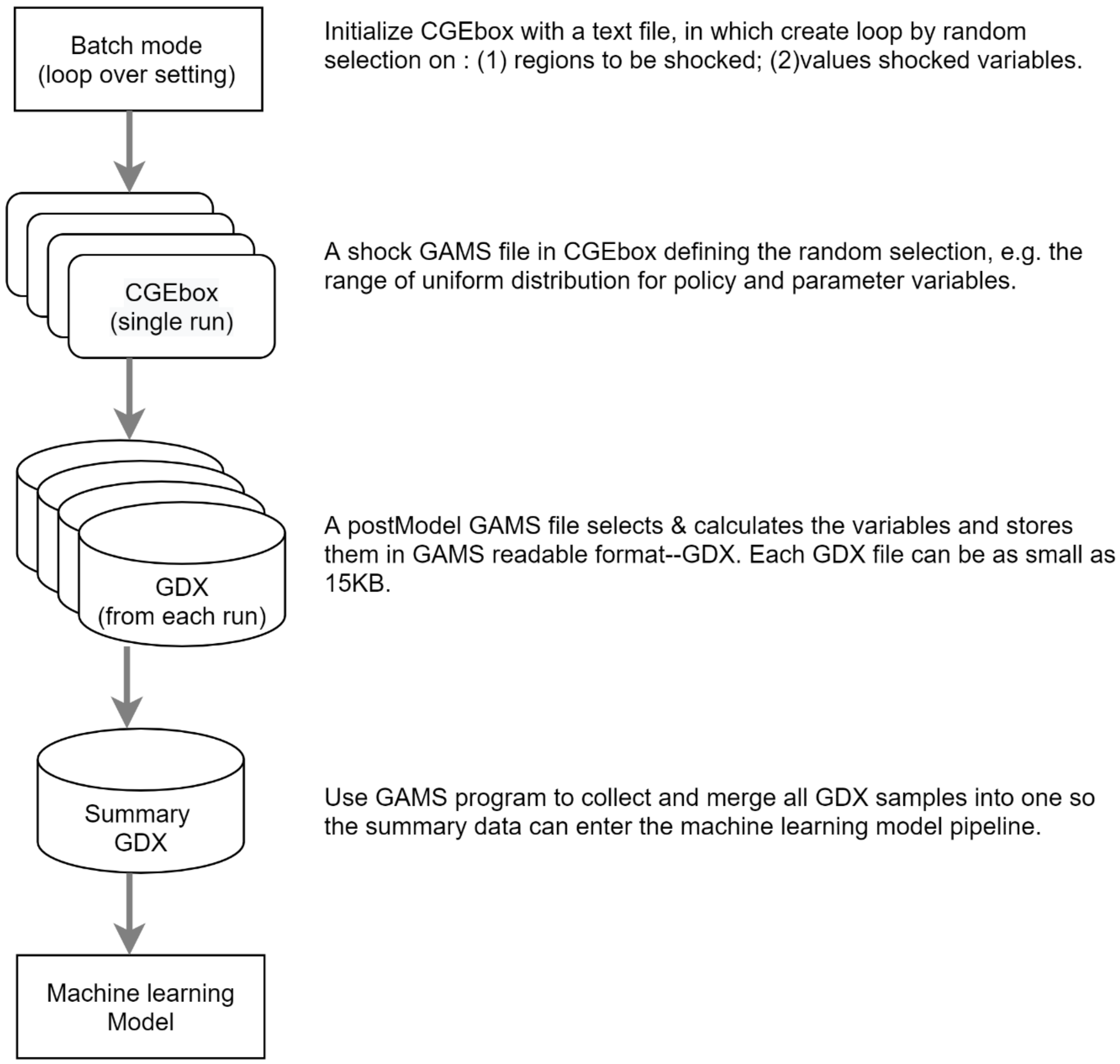

Figure 3: Flowchart of simulation and sampling method in CGEBox

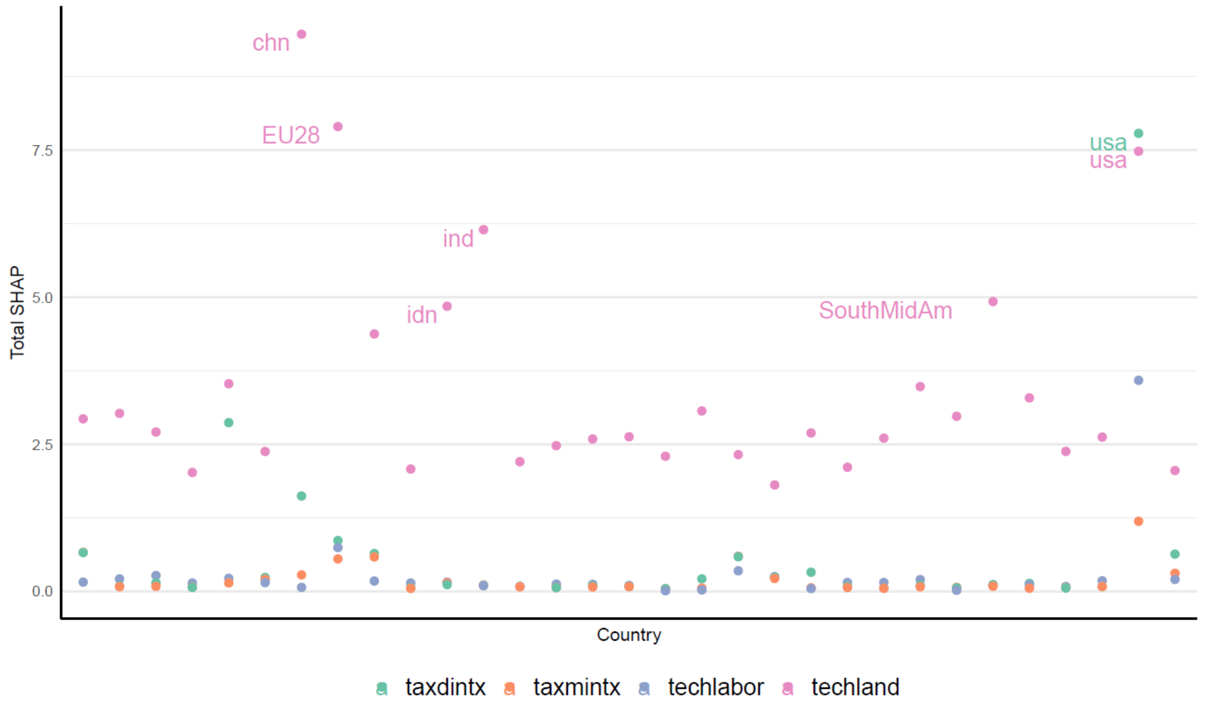

Figure 4: SHAP values by policy variables for the first neural network excluding parameter uncertainties. Note: Countries greater than 95% quantile are marked.

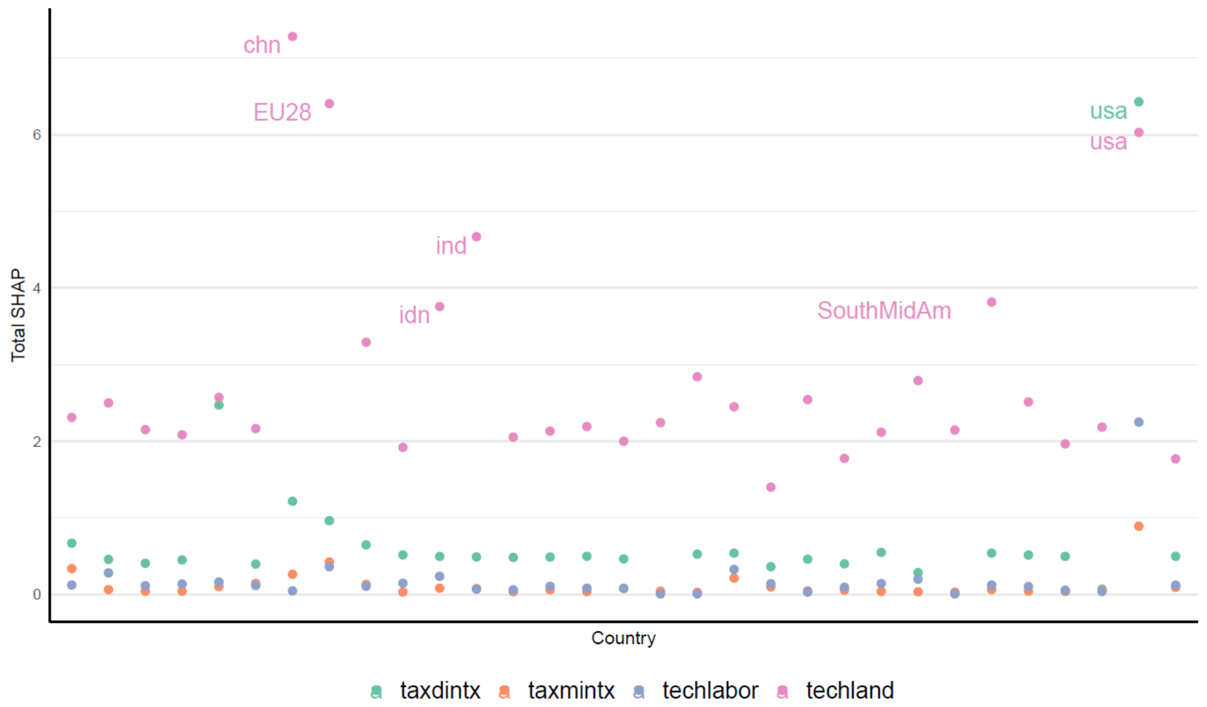

Figure 5: SHAP values by policy variables for the second neural network including parameter uncertainties. Note: Countries greater than 95% quantile are marked.

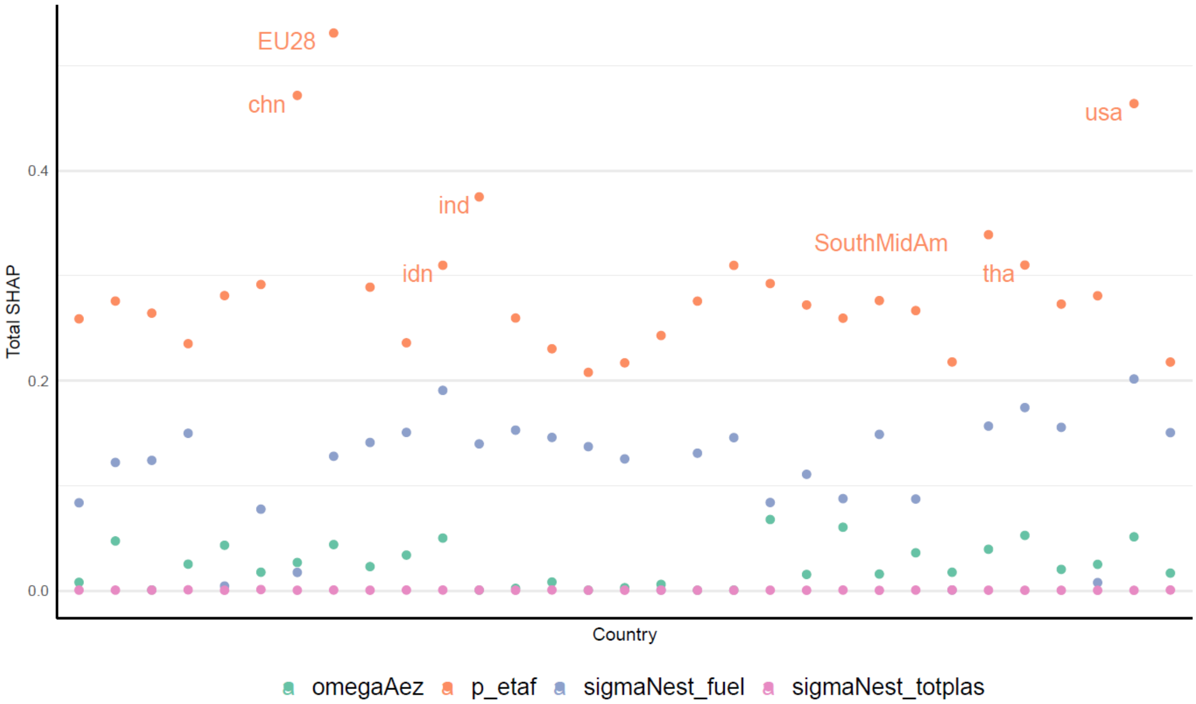

Figure 6: SHAP values by parameters for the second neural network. Note: Countries greater than 95% quantile are marked.

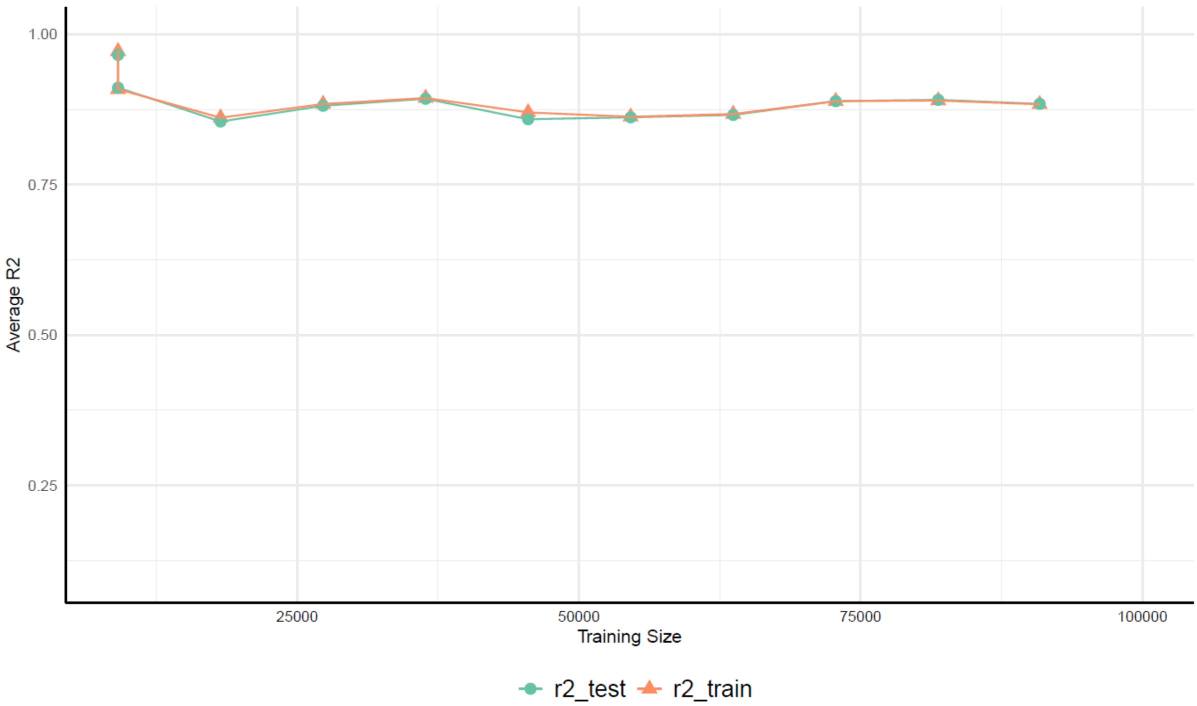

Figure 7: Total R2 including parameters.

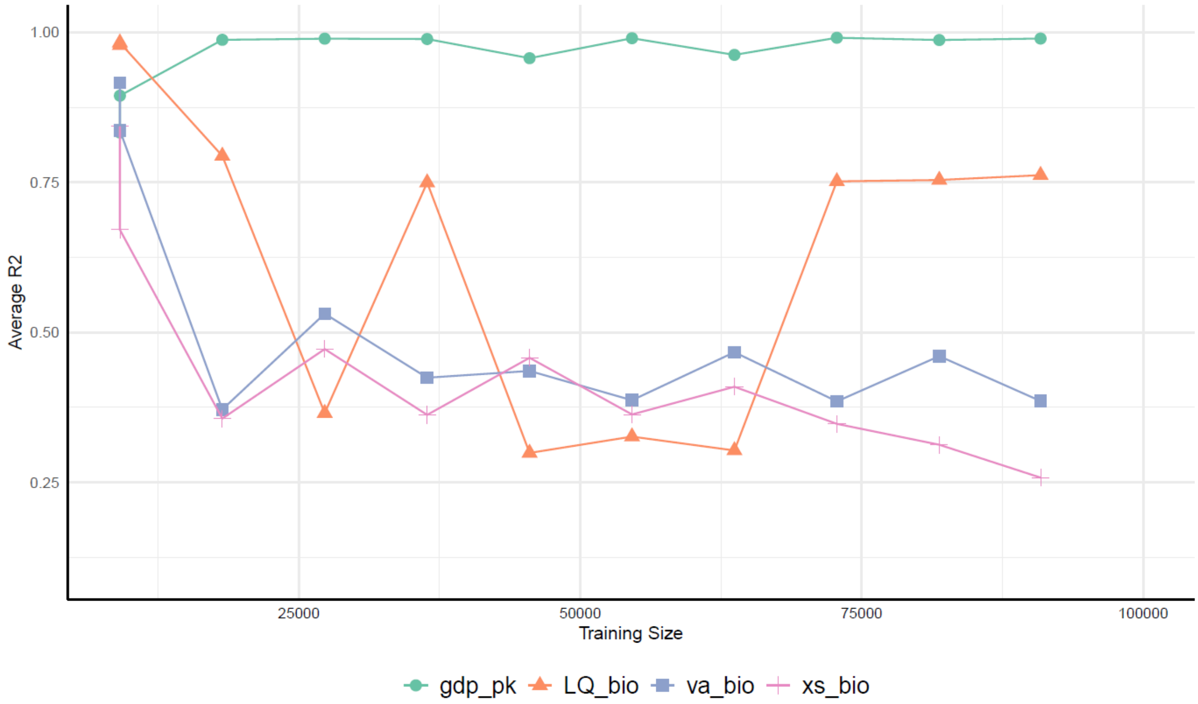

Figure 8: Economic growth performance including parameters.

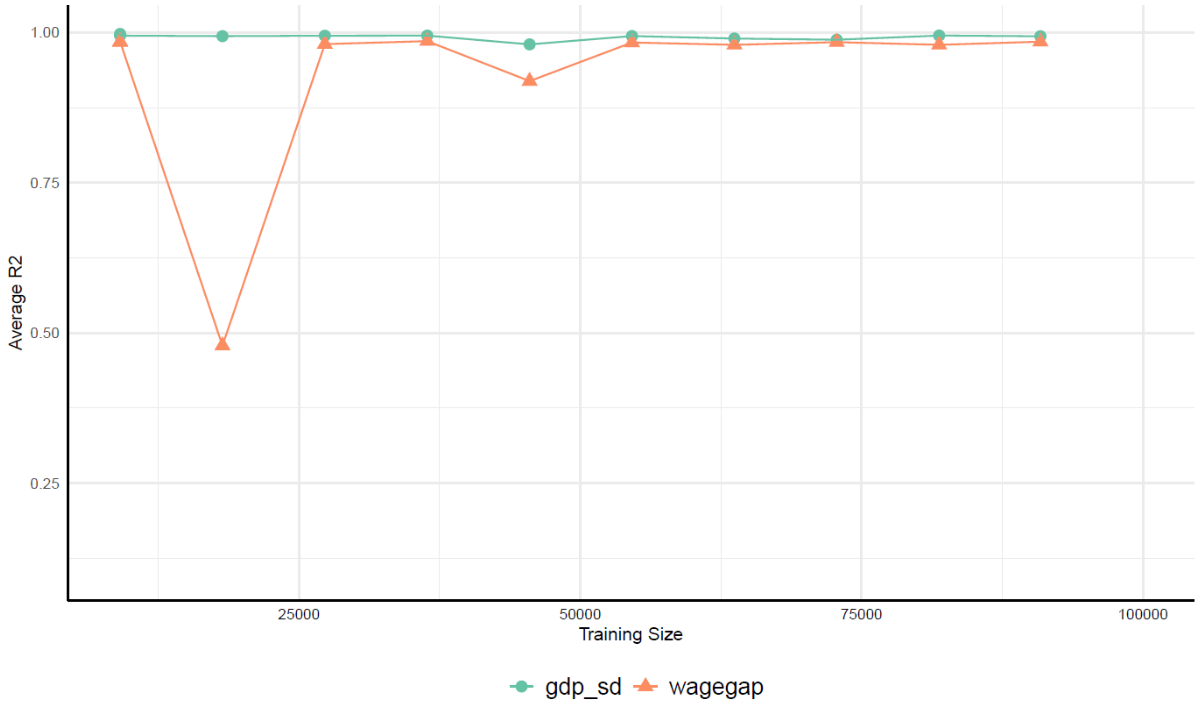

Figure 9: Inequality performance including parameters.

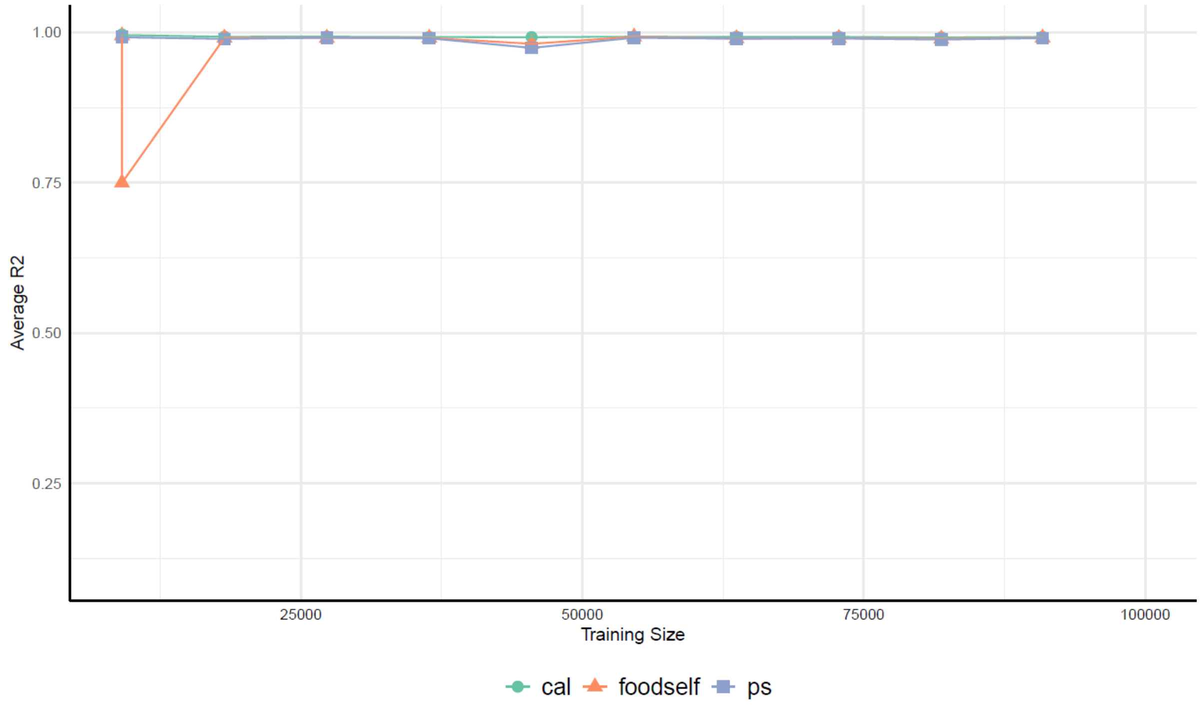

Table 10: Food security performance including parameters.

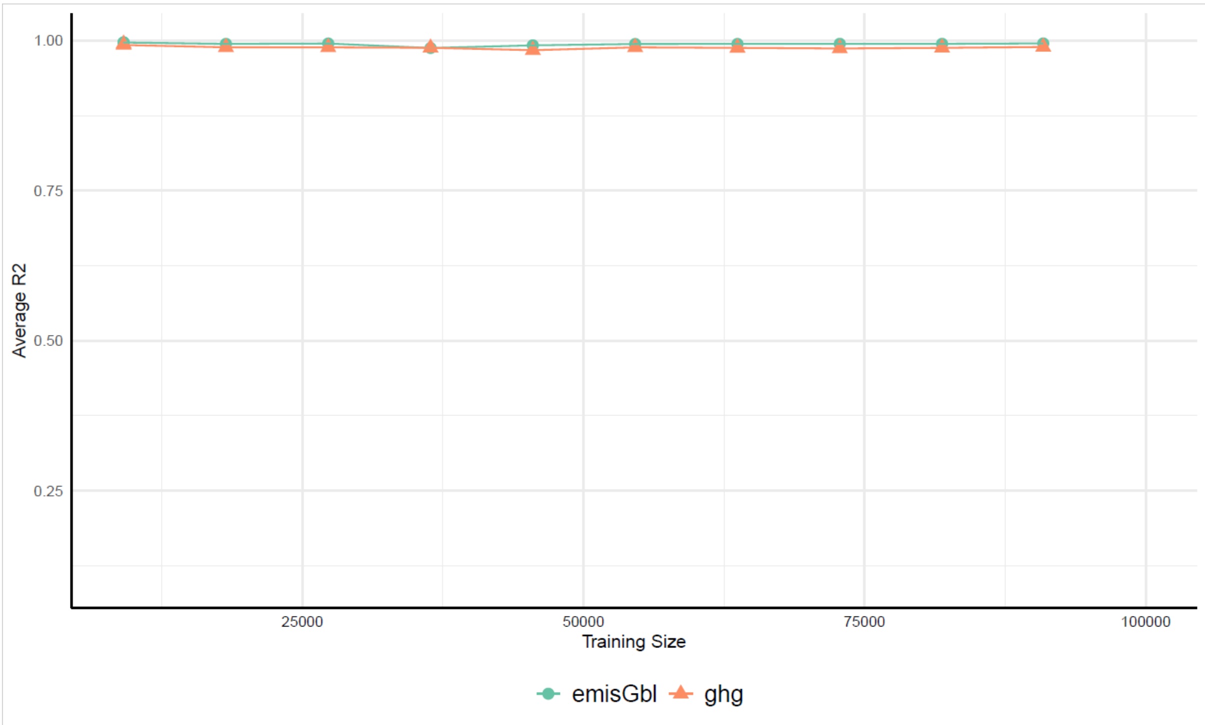

Table 11: GHG emission performance including parameters

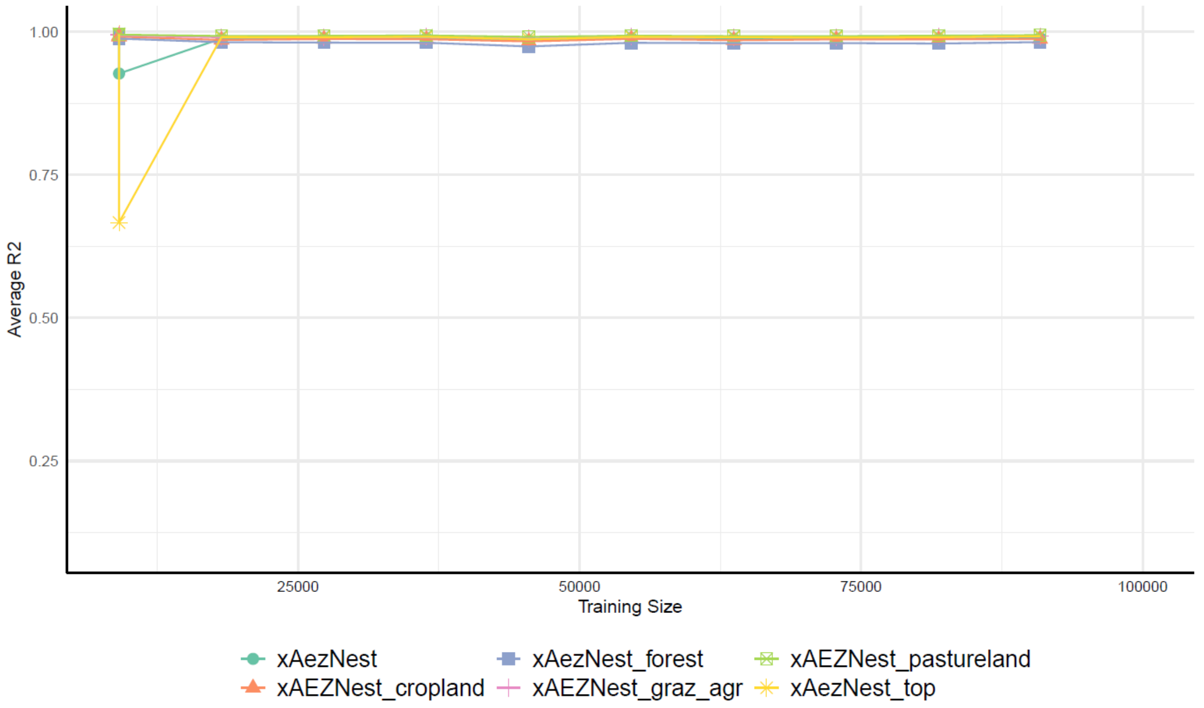

Table 12: Landuse change performance including parameters.