Optimisation of the Global Calculator using Monte Carlo Markov Chains and Genetic Algorithms

Jorge Garcia

Imperial College London,
Department of Earth Science and Engineering

Master's Thesis in Applied Computational Science and Engineering,
undertaken for Total

August 23, 2020

Supervisors: Dr Alan Irving (Total) and Dr Adriana Paluszny (Imperial College)

http://github.com/acse-jg719 - jg719@ic.ac.uk

Abstract

This investigation aims to generate different climate change mitigation pathways with the "Global Calculator" – a model used to forecast the world's energy, food and land systems to 2050. Performing a constrained multiobjective optimisation of the model's input parameter space yields alternative pathways to sustainability. The key challenge of exploring an ample parameter space (8e70 different parameter combinations) efficiently is addressed by three methods: Monte-Carlo Markov Chains (MCMC), Genetic Algorithms (GA) and Covariance Matrix Adaption Evolutionary Strategy (CMA-ES). Firstly, the probability distribution of the model inputs that maximises the probability of satisfying climate and cost objectives is derived using MCMC. Then, the multiobjective optimisation problem is tackled using GA, thus enabling the generation of bespoke climate-friendly and economically viable pathways. Total's sustainable pathway has been calculated for two conflicting objectives and twelve input constraints. Lastly, CMA-ES and artificial neural networks (ANNs) have been adapted to the Global Calculator but failed to yield efficiency gains. The challenge of updating the calculator with real-time data is proposed as further work.

Keywords: Multiobjective optimisation, Genetic algorithms, Global Calculator, Environmental change, Energy modelling.

1 Introduction

This section introduces the calculator, the purpose of optimising it, and the methods considered in doing so. The Global Calculator is a tool used to inform the climate debate at an international level. As stated by The Paris Agreement's long-term temperature goal [31], the increase in global average temperature must be below 2 °C [14]. Policymakers, business leaders, NGOs and researchers use the Global Calculator to design their version of the future up to 2050, see the implications for the climate, and take business decisions in the present accordingly [4].

The Global Calculator is a free, open-source, and interactive model of energy, land and food systems to 2050. It can be used to assess a wide range of climate change mitigation pathways. It has 48 input 'levers' which are organised in clusters and around 60 outputs. Each input lever has four levels of ambition (minimal effort, ambitious effort, very ambitious effort and extreme effort).

Currently, users of the calculator create their pathways by selecting lever combinations by hand. Trial-and-error confirms the difficulty of achieving the 2 °C warming limit. This can yield suboptimal lever combinations. This investigation aims to generate optimal lever combinations based on user-specified constraints. The calculator's parameter space consists of around 8e70 possible lever combinations. The key challenge addressed is the efficient navigation of such a broad parameter space under various potentially conflicting constraints.

Quickly navigating a large parameter space is a common challenge in the field of Data Science and Game Artificial Intelligence (AI). This investigation contributes to the state of the art by adapting methods from these fields to the Global Calculator.

2 Literature review

The use of carbon calculators as a tool for informing climate debate has become available for different purposes and scales, including private companies and NGOs, and national and global calculators led by governments [15]. This section presents a brief literature review of the Global Calculator, the opportunity of optimising it and the optimisation approaches that could be used in doing so.

The Global Calculator has been used to inform policy, business decisions, and research [4]. In 2019, the UN wrote a report that relied on the results of the calculator [8]. Many research publications use the calculator to aid their analyses, as seen in [5], [6], [7] and [9]. Organisations have developed and published their pathways, including Shell, Mott-MacDonald, the World Nuclear Association and the Vegan Society [21]. Four benchmark pathways to sustainability have been published by the authors of the Global Calculator in [1], [2] and [3]. These include "distributed effort", "consumer reluctance", "low action on forests" and "consumer activism".

Users of the Global Calculator seek to generate bespoke optimal pathways. Probability distributions for the inputs of the calculator that maximise the probability of meeting user-specified optimisation objectives can be derived to help them do so. These distributions can be approximated by methods such as MCMC [34], importance sampling [35], the forward-backwards algorithm [36], and sequential Monte Carlo (SMC) [37]. MCMC is chosen as it is the most straightforward to implement and has succeeded in solving similar problems in Bayesian statistics [11], computational physics [16], computational biology [17] and computational linguistics [18]. MCMC methods approximate a posterior distribution given a prior distribution of the model inputs, a likelihood function, and a set of observations [22]. It is a technique commonly used in the numerical approximation of multi-dimensional integrals [19]. Adapting and applying MCMC to the Global Calculator by imposing constraints through fixed observations would add to the literature.

Calculating optimal pathways that satisfy multiple user-specified constraints requires the efficient navigation of the Global Calculator's parameter space. A similar research challenge frequently appears in games – where the action space is nearly continuous, thus leading to exorbitant branching factors [20]. Classic tree search methods were originally used to play simple board games such as Chess and Checkers and are unsuitable for high branching factors. Instead, the field of Game AI has successfully tackled this problem by using methods such as evolutionary strategies, reinforcement learning and Monte Carlo tree search [38][10].

Finding climate-friendly and economically viable pathways involves dealing with the trade-off between conflicting constraints. The field of multiobjective optimisation deals with similar challenges through the simultaneous optimisation of more than a single objective function [23]. In the context of the calculator, this could be maximising economic viability and minimising climate emissions. Traditional methods such as weighted sum, e-constraint and weighted method have been replaced by evolutionary strategies [24]. Within this category, genetic algorithms stand out as the most popular, having previously tackled multiobjective optimisation problems in engineering [25]. CMA-ES is considered as an alternative multiobjective optimisation method [26] and has the potential to outperform GA [32]. Adapting and applying GA and CMA-ES to perform a constrained optimisation of the Global Calculator would contribute to the literature.

Lastly, using ANNs to perform non-linear regression could enable a successful approximation of the Global Calculator, thus speeding up calculations. ANNs are widely used as function approximators [13], as seen in [39] and [27]. An upper bound for the number of neurons of the network is provided in [12]. The amount of training data required for a reasonable prediction accuracy is variable and depends on the underlying model [28]. Developing a framework to model carbon calculators using ANNs would add to the literature.

3 Methodology

This section outlines the technical considerations evaluated in choosing the methods used in this investigation.

3.1 Interacting with the Global Calculator: Web scraping

The Global Calculator can be accessed via its web version [21], Excel spreadsheet [29], or Ruby code [30]. Each of these platforms is evaluated according to three metrics: efficiency, reliability and user-friendliness, as seen in Table 1. The Excel spreadsheet behaves prohibitively slowly, taking around two minutes per iteration. On the other hand, interacting the Ruby code proved extremely difficult in practice. Interacting with the web version is the best compromise between efficiency, reliability and user-friendliness. A bespoke web-scraping library has been built to interact with the calculator's web version. It enables the user to easily, efficiently, and reliably communicate with the Global Calculator's web tool via Python.

	Efficiency	Reliability	User-friendliness
Excel spreadsheet	~ 2 min/iteration	High	High
Ruby code	< 1s/iteration	High	Very low
Web version (selected)	~ 2s / iteration	Medium	High

Table 1: Comparison of platforms to interact with the Global Calculator.

3.2 Exploring the Global Calculator: Preliminary analysis

A preliminary quantitative and qualitative exploration has been conducted to better understand the Global Calculator.

- **Quantitative exploration - Mono-parameter sensitivity analysis and output correlation matrix:** The mono-parameter sensitivity analysis aims to quantify how sensitive the model outputs are to changes in the input parameters. Each lever is moved by an equal distance at a time while measuring the variations in per capita GHG emissions and changes in GDP, as shown in Algorithm 3 (Appendix A). The output correlation matrix is calculated by generating 10000 random outputs bounded between 1.5 and 3.5 using the Global Calculator.

- **Qualitative exploration - Lever information and data sources:** The data source and relevance of levers with a highest climate sensitivity are investigated. These levers include quantity of meat, global population, livestock (pasture-fed), type of meat, livestock (grains/residues fed), wastes and residues, wind, solar, crop

yields, calories consumed, electric and hydrogen, mode, fossil fuel efficiency, nuclear, CCS (industrial), CCS (power) and coal/oil/gas.

3.3 Studying the inputs of the Global Calculator: Monte Carlo Markov Chains

Applying MCMC to the Global Calculator yields the posterior distribution of model inputs that maximises the likelihood of meeting the two core conflictiving objectives – minimising climate impact and economic cost. Adjusting the parameters of MCMC is a highly empirical task. A well-adjusted MCMC implementation shall have a high acceptance rate and low autocorrelation coefficient of accepted samples. High autocorrelation values indicate a lack of independence between accepted samples, which can result in ill-posed posterior distributions of the model inputs. The steps below have been followed to adapt the method to the Global Calculator, resulting in Algorithm 1.

To impose the two objectives, a set of 200 observations is randomly drawn from a 2D Gaussian distribution with its mean centred around the climate and cost targets. This is illustrated in Figure 5 (Appendix A). The standard deviation is determined empirically to maximise the acceptance rate and minimise the autocorrelation coefficient of accepted samples.

The likelihood function must as well be a 2D Gaussian. Its standard deviation is directly related to the acceptance rate. This value can be changed to adjust the weight given to the observations. Standard deviation values of 1,000,000 per capita GHG emissions and 100% change in GDP worked well in practice. The constrained observations and likelihood function are shown in Figure 6 (Appendix A).

Interacting with the Global Calculator yields prior knowledge about the empirical distribution of temperature and cost values. A set of 300 different random lever combinations is generated, bounding their values between 1.5 and 3.5 to avoid unrealistic levels of ambition. Fitting a Gaussian distribution to the resulting model outputs results in Figure 7 (Appendix A). Ready-made MCMC samplers cannot be applied to the Global Calculator as the map between model inputs and outputs is unknown. As such, a discrete variation of the Metropolis-Hasting sampler is used. Each lever is moved by some distance according to the probability distribution shown in Table 2.

Jump length	Probability of jump
0.1	0.3
-0.1	0.3
0.2	0.15
-0.2	0.15
0.3	0.05
-0.3	0.05

Table 2: Probability distribution of lever jumps.

Algorithm 1: Monte Carlo Markov Chains adapted to the Global Calculator with 2 constraints.

```
procedure MCMC(μ_prior, σ_prior, μ_current, σ_likelihood ← [[1000000, 0], [0, 100]],
               Observations, Current_levers, Constraints,
               Iterations)
    Accepted_levers ← []                          Destination of accepted lever combinations
    for _ in range(Iterations):
        Current_levers, μ_proposal ← sampler(Current_levers, Constraints)    Calculator outputs
        Likelihood_ratio = 2D_Gaussian(μ_proposal,σ_likelihood)*observations / 2D_Gaussian(μ_current,σ_likelihood)*observations    Likelihood ratio
        Prior_ratio = PDF(μ_proposal)~2D_Gaussian(μ_prior,σ_prior) / PDF(μ_current)~2D_Gaussian(μ_prior,σ_prior)    Prior ratio
        Acceptance_probability = Likelihood_ratio*Prior_ratio    Acceptance prob.
        if random() < Acceptance_probability:    Accepted value?
            μ_current ← μ_proposal    Update outputs
            Accepted_levers ← Current_levers    Save lever values
    return Accepted_levers
end procedure
```

3.4 Optimising the Global Calculator: Genetic algorithms

GAs are used to perform the multiobjective optimisation of the Global Calculator's inputs. The optimiser takes different types of constraints, as listed in Table 3. In addition, the optimal lever combination must not raise any of the warnings listed in Table 4.

Type of constraint	**Input value**	**Input range**	**Output value**	**Climate and GDP**
Description	Setting a lever to a fixed value.	Bounding the values of a lever within a range.	Setting an output value to a fixed value.	Ensuring climate impact is below 2C and minimise cost.

Table 3: Constraint types accepted by the GA optimiser

Warning name	*Description*
Energy security	The energy supply does not meet the energy demand.
Electricity	The energy supply exceeds the energy demand.
Technological development	Ambitious technologies are used, such as GHGR.
Resources	More resources than the world has available are required.
Land use	More land than the world has available or a large decrease in forest area.
Abatement effort	Lever values are set to unrealistic levels.
Climate	The 2C warming target is not met.

Table 4: Warning types raised by the Global Calculator.

The optimisation constraints are incorporated in the cost function by calculating the Mean Squared Error (MSE) between current and target values. The outputs of the Global Calculator have different units and scales, so they must be normalised to balance the cost function. This is achieved by approximating the standard deviation of each output by generating 100 random lever combinations.

The selection operator finds the two lever combinations with the lowest cost values. The crossover operator performs uniform crossover with these two to yield a new lever combination. The mutation operator introduces randomness in the crossover process to avoid getting stuck in local minima. A mutation rate of 20% has shown to work well in practice. The outputs of the Global Calculator are sensitive to changes in the levers, which makes pre-made mutation operators perform poorly. A simple bespoke operator is created, where the lever value that mutates is randomly increased or decreased by 0.1. This is illustrated in Table 5, where each new lever value is chosen from either parent with probability 0.5. Green cells show the values passed on to the child, whereas red ones correspond to those discarded. Blue cells correspond to mutated values. Algorithm 2 shows the multiobjective GA optimiser applied to the Global Calculator.

Parent 1	3.2	2.5	2.9	3	2.8	2	2.8	3.3	3.5	2.5	2.3	2.5
Parent 2	1.7	3.4	2.6	2	3.4	1.7	2.2	1.7	2.1	1.6	2.7	2.3
Child	**1.7**	**2.5**	**3**	**2**	**2.8**	**2**	**2.2**	**1.8**	**3.5**	**1.6**	**2.3**	**2.4**

Table 5: GA crossover and mutation of two parents to yield child.

Algorithm 2: Genetic Algorithm applied to the Global Calculator, multi-constraint/multi-objective

```
procedure GA(Population_size, Iterations, Constraints)
    Output_values ← []                          Destination of output values
    Fitness_values ← []                         Destination of fitness values
    Lever_values ← []                           Destination of lever values
    for _ in range(Population_size)             Generate initial population
          Output_values, Lever_values ← Generate_random_chromosome(Constraints)
    for _ in range(Iterations):                 Perform GA iteration
          Fitness_values ← Calculate_fitness(Lever_values, Constraints)    Calculate fitness
          Parents ← Fittest(Lever_values, Fitness_values)                  Find fittest 2
          Lever_values ← Crossover_and_mutation(Parents, Constraints)      Crossover & mutate
          Output_values ← Read_outputs(Lever_values)                       Read outputs
    return Lever_values, Output_values
end procedure
```

3.5 Attempting to improve the optimisation of the Global Calculator: Covariance Matrix Adaption Evolutionary Strategy

CMA-ES can be used to tackle multiobjective optimisation problems and has the potential for converging faster than GA. The method updates the covariance matrix depending on the current state of the system. The map between the Global Calculator's inputs and outputs is unknown. As such, a bespoke method for updating the covariance matrix is required.

A map between standard deviation and probability values of lever jumps is created by fixing the jump probabilities and generating random lever combinations around a benchmark pathway. The jump probability values used are shown in Table 6. The algorithm used to adapt CMA-ES to the Global Calculator is too long to be included. It can be found in the *Iterate* method, as part of the *"CMA-ES Analysis.ipynb"* file.

	Choice 1	Choice 2	Choice 3	Choice 4	Choice 5	Choice 6	Choice 7	Choice 8	Choice 9	Choice 10	Choice 11	Choice 12
Probability of ± 0.1 jump	0	0	1	0	0	0	0.33	0.5	0.66	0.33	0.5	0.66
Probability of ± 0.2 jump	0	1	0	0.33	0.5	0.66	0	0	0	0.66	0.5	0.33
Probability of ± 0.3 jump	1	0	0	0.66	0.5	0.33	0.66	0.5	0.33	0	0	0

Table 6: CMA-ES map between probabilities and jump distance.

3.6 Attempting to model the Global Calculator using Artificial Neural Networks

Approximating the Global Calculator with a neural network would result in a speed-up in the rest of the methods. According to the universal approximation theorem, ANNs can approximate any function. To approximate the Global Calculator, the design of the neural network has been kept as simple as possible. The network takes 48 inputs and returns 55 outputs, and has two fully connected layers of size 100 and 70 (without including bias terms). Leaky ReLU has been chosen as the activation function. The optimisation is carried out using Adam's method. The outputs of the network are in different scales and rarely become zero, so the scale-free Mean Absolute Percentage Error (MAPE) error metric has been used. The learning rate is analysed later on. Lastly, the training data fed to the network has been generated by randomly sampling the input space of the Global Calculator.

4 Results

4.1 Preliminary analyses

Figures 1 and 2 show the sensitivity of GHG emissions per capita and changes in GDP to variations in the input levers. Levers related to food and demographics have a high climate impact. Levers related to demographics and

transport have a significant impact on economic output.

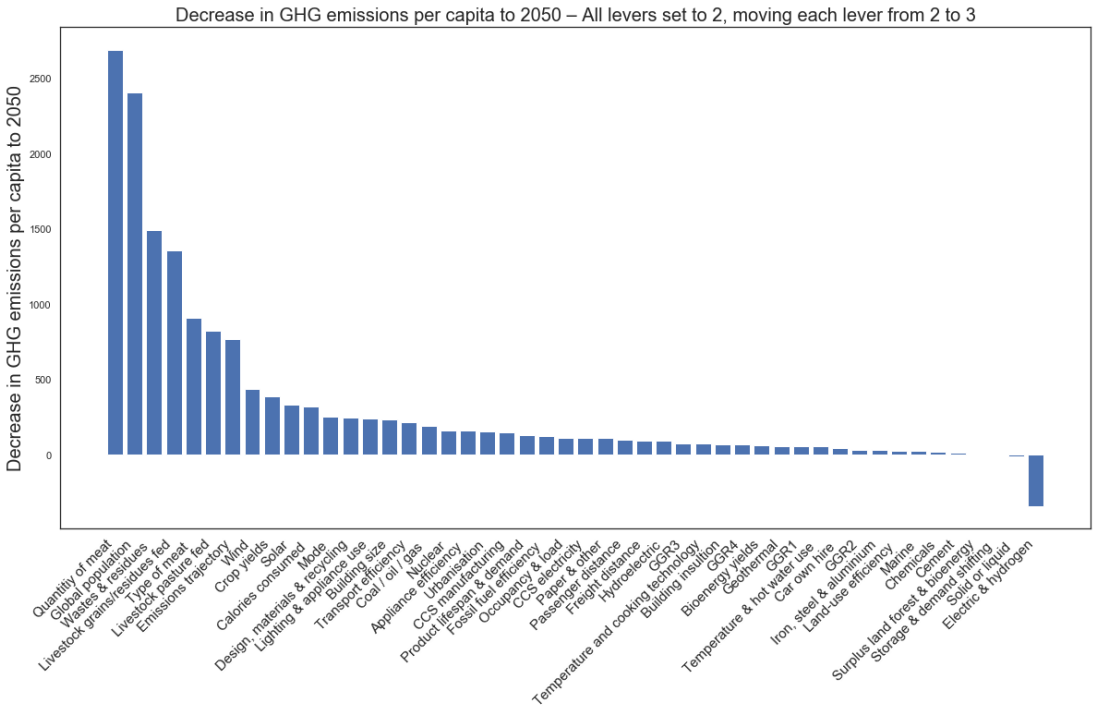


Figure 1: Mono-parameter sensitivity analysis of GtCO2 per capita decreases.

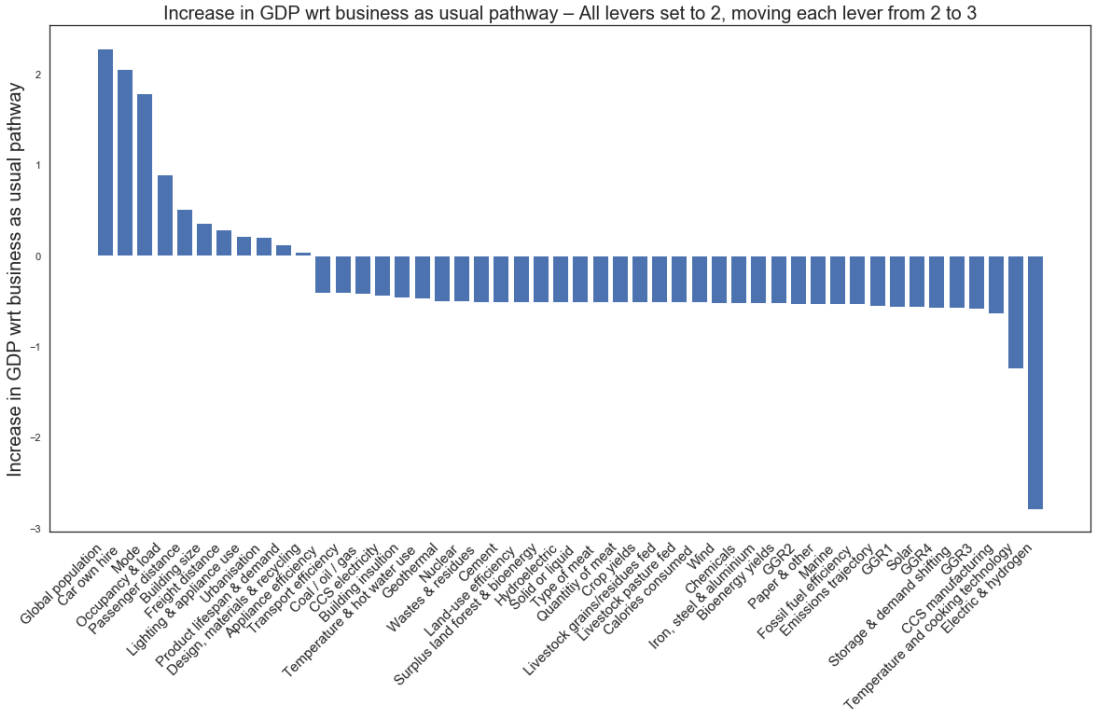


Figure 2: Mono-parameter sensitivity analysis of changes in GDP.

The output correlation matrix is presented in Figure 15 (Appendix C). White cells represent no correlation between outputs, while blue and red cells correspond to positive and negative correlations, respectively. Some of the output correlations of the Global Calculator are expected. For example, forest area is inversely correlated to GHG emissions per capita, and bioenergy area is positively correlated to bioenergy supply.

Other correlations are not straightforward, such as the nearly perfect inverse correlation between the lifespan of a refrigerator in urban areas and the demand for consumer packaging. To understand these relationships, Figure 15 (Appendix C) shows the connections between model outputs. Green edges correspond to a strong positive correlation and magenta to a negative one. Lastly, the qualitative analysis of input levers has been conducted by Dr Alan Irving (Total). It can be found in *"Qualitative lever analysis.pptx"*.

4.2 Web scraping library

The process of developing the web scraping library has been highly empirical, adapting the methods to the Global Calculator's architecture. The library has been thoroughly tested to ensure its robustness. Its performance depends on the local broadband capacity and the Global Calculator's server capacity. The interaction with the website is parallelised by running several browser windows concurrently. The results are shared between notebooks by reading and writing to the same XLS files.

4.3 Monte Carlo Markov Chain

MCMC has been run for over 24 hours, resulting in 21,300 different proposed parameter combinations. The posterior distribution of each input lever is shown in Figure 18 (Appendix E). On the other hand, the accepted model outputs yield the posterior distribution shown in Figure 3. This figure shows a random walk around the solution space bounded by the constraints.

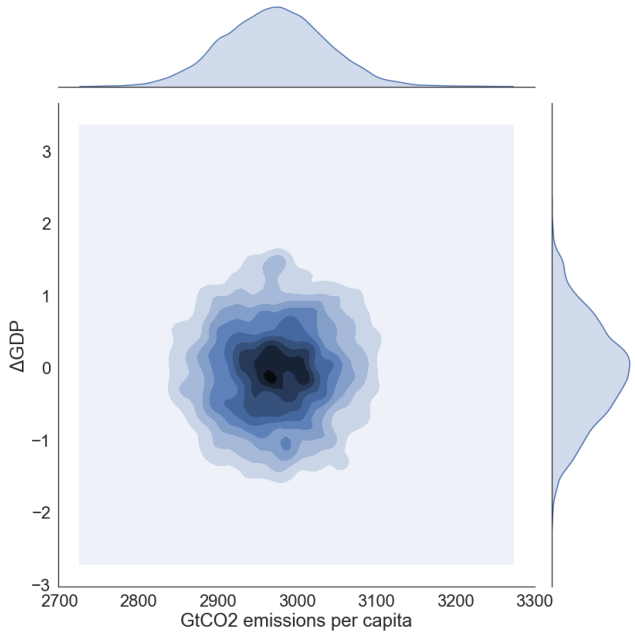


Figure 3: MCMC posterior distribution of accepted model outputs.

Around half of the model outputs proposed are rejected. Figure 8 (Appendix A) shows the acceptance rate as a function of parameter proposals, oscillating around 0.5. The autocorrelation coefficients of the accepted model outputs are shown in Figure 9 (Appendix A). Accepted GTCO2 emissions per capita show a lower autocorrelation than accepted GDP changes.

The correlation matrix of lever values that maximise the probability of meeting the optimisation objectives is presented in Figure 16 (Appendix D). Correlations between lever values appear to be frequent, suggesting that most inputs are related to each other. The graph shows input combinations that might work best together. Figure 17 (Appendix D) extends this idea by showing the connections between accepted model inputs, where green corresponds to strong positive correlations and magenta to negative ones. For example, wind energy and bioenergy yields are inversely correlated. This shows that a high value of wind energy should be accompanied by a low value of bioenergy yields to maximise the probability of meeting the optimisation objectives.

On the other hand, a subset of the paired density and scatter plot matrix is presented in Figure 19 (Appendix F). It includes levers related to fossil fuels and renewables. The chart presents a guide for selecting input lever combinations that maximise the probability of satisfying the optimisation objectives. For example, a high level of solar energy combined with a low level of marine energy shall maximise such a probability.

4.4 Genetic Algorithms

The GA optimiser has succeeded in generating Total's sustainable pathway. The constraints used in doing so are summarised in Table 7. These values are highly subjective and do not represent any particular corporate strategy. The evolution of the average cost per chromosome and the computing time are analysed to determine an appropriate population size. Figure 10.1 (Appendix A) shows how the average cost per chromosome of different population sizes decreases as the number of generations is increased. For a fixed number of generations, a population size of 5 fails to minimise the cost per chromosome as much as sizes of 10, 20 and 40 do. Figure 10.2 (Appendix A) shows the computing time taken as the generations are increased. As expected, the relationship between the number of lever combinations per generation and computing time taken is linear. A population size of 20 is selected as a sensible trade-off between cost-minimisation and computing time.

Constraint type	Constraint description	Value
Output	Climate impact	*Minimise*
Output	Economic impact	*Minimise*
Input	Global population	[1.6, 2.0]
Input	Electric & hydrogen	[2.8, 3.2]
Input	CCS (manufacturing)	[1, 2]
Input	CCS (electricity)	1
Input	Nuclear	[1.5, 2]
Input	Wind	[2.6, 3.0]
Input	Solar	[3.0, 3.4]
Input	Calories consumed	[2, 3]
Input	Quantity of meat	[2, 3]
Input	Type of meat	[2, 3]
Input	Livestock grains/residues fed	[1.8, 2.2]
Input	Land-use efficiency	[1.8, 2.2]

Table 7: Total's optimisation constraints.

Figure 11 (Appendix A) shows the evolution of the model outputs generated by the GA optimiser after 9 generations towards an optimal lever combination. The figure shows how the two conflicting objectives of minimising the GHG emissions per capita and change in GDP are satisfied. As can be seen in Figure 4, the GA minimises the total cost of each generation.

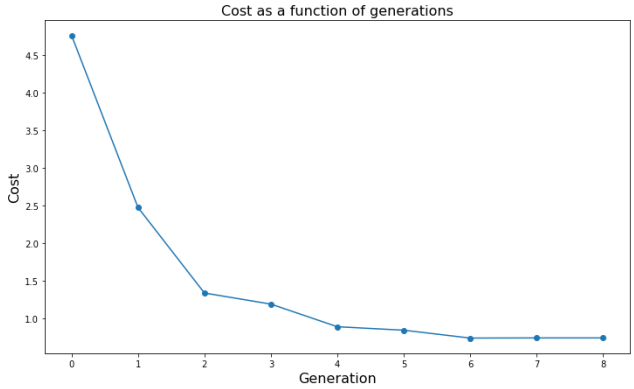


Figure 4: MSE minimisation using GA for Total's pathway

The complete forecast produced by Total's 2050 sustainable pathway can be accessed at [33]. Table 9 (Appendix G) contains the lever values of Total's pathway, the four benchmark scenarios (distributed effort, consumer reluctance, low action on forests and cosumer activism), and the business as usual scenario: IEA 6DS. Total's pathway yields a forecast to 2050 of 2835 GtCO2 per capita, which is below the 2C warming target of 3010 GtCO2 per capita. It also estimates an improvement in economic output of 3.7% of the GDP to 2050 with respect to the business as usual scenario.

Figure 20 (Appendix H) shows the lever values corresponding to Total's sustainable pathway and the business as usual pathway. Most of Total's levers have been set to a more ambitious level than the business as usual pathway. Together with other sustainable pathways, Total's lever combination suggests that ambition efforts must be increased overall to meet the climate warming target while minimising economic cost.

4.5 CMA-ES

This method has failed to converge even for the most straightforward constraint scenarios. Figure 12 (Appendix A) shows the model outputs of an initial lever combination (blue) and 20 variations of it (black) generated by randomly increasing or decreasing each lever by 0.1. The graph shows that the output variance is too high, thus making the adaption of the covariance matrix unfeasible. As a result, the method generates scattered output proposals that fail to satisfy the two conflicting objectives.

4.6 Attempting to model the Global Calculator using Artificial Neural Networks

After trying to model the Global Calculator using ANNs, the results indicate that this is not feasible in practice when dealing with the web tool. Figure 13 (Appendix A) shows how the Mean Average Percentual Error changes for different learning rates as the epochs are increased. Learning rates above 0.5 make both the SGD and Adam optimisers diverge. Learning rates below 0.3 are too slow. The slopes of the three curves suggest that, in the best-case scenario, several millions of training samples would be needed to have an accurate model approximator.

5 Discussion and further work

The mono-parameter sensitivity analyses showed the climate and cost impact of each lever. The levers that yield the most substantial decrease in per capita GHG emissions are quantity of meat, global population and wastes & residues. The climate impact of the electric and hydrogen lever depends on the split in the electricity generation (renewables or fossil fuels) that powers electric vehicles. On the other hand, the levers that produce the highest economic impact include global population, car own hire and mode. The output correlation matrix has shown non-obvious correlations between outputs of the Global Calculator such as the nearly perfect inverse relationship between the lifespan of a refrigerator in urban areas and the demand for consumer packaging. Further work would involve determining whether these strong correlations in the outputs of the calculator imply causality.

The qualitative exploration of levers has provided a glimpse into the intricacy of the model. Levers are interrelated in a non-linear and sophisticated way. Attempting to extrapolate values to the year 2050 based on 2011 and 2020 values has not been particularly useful as there is little variation between lever levels. This qualitative analysis has been translated into constraints used to generate Total's sustainable pathway. Next steps would include finding updated data sources where possible and devising a scalable way for updating the Global Calculator with such data.

A bespoke web scraping library has been developed to interact with the Global Calculator. This is the most appropriate way of interacting with the model given the time constraints. The library allows the user to automate the interactions with the web tool. Its time performance depends on the local internet connection and the load of the Global Calculator's server. Parallelising the interaction by running several sessions concurrently further reduces the computing time.

The input levers to the model have been analysed using MCMC. In doing so, the climate and cost objectives have been translated into observations. The prior distribution, likelihood function and sampler have been empirically adjusted to maximise the acceptance rate and minimise the autocorrelation of the output parameters. The method has run for over 24 hours, resulting in 21,300 proposed lever combinations and model outputs. Around 50% of these have been accepted. Cost outputs present a higher autocorrelation than climate outputs as a result of the lower sensitivity of GDP changes. The correlation length could be estimated from the autocorrelation functions to ensure the independence of accepted samples.

The probability distributions of input levers that maximise the probability of meeting the climate and economic cost objectives are calculated. Similarly, the correlation matrix and paired density and scatter plot matrix of lever values that maximise the probability of meeting the optimisation constraints are derived. These charts help users of the Global Calculator select individual lever values and lever combinations that are likely to yield sustainable and economically viable scenarios. Furthermore, the values from the posterior distributions of model inputs have been used as input range constraints for the optimiser, slightly improving its convergence. The posterior distribution of model outputs is derived too, showing how the accepted values oscillate around the climate and cost objectives. Adapting MCMC to higher dimensions would allow the user to specify further constraints.

A GA optimiser has been implemented to solve this multiobjective optimisation problem. It enables the user to specify different types of input and output constraints, and returns lever combinations that satisfy them. Its mutation operator and rate, and crossover operator have been adapted to the Global Calculator to maximise the method's rate of convergence. Its population size has been set to 20 as a compromise between accuracy and efficiency. The contributions to the cost function from different constraints have been balanced by normalising them. The number of generations needed before convergence depends on the number of constraints set, but heavily constrained problems have been solved in at most nine generations. Further work would involve using a quantitative approach to determine optimal GA parameters.

CMA-ES has been heavily modified to adapt it to the Global Calculator. The method has failed to converge, yielding scattered parameter proposals. This is a result of the high sensitivity of the model outputs to changes in the levers. Further work would involve decreasing the minimum variance by randomly varying subsets of chromosomes. Additionally, the mutation operator used in GA could become adaptive by using a similar approach as in CMA-ES.

An ANN has been used to try to approximate the Global Calculator. The results suggest that this is not feasible in practice as it would require millions of training samples. Further work would involve optimising the network architecture to minimise the number of training samples required, as well as finding an efficient way to sample millions of training data points from the calculator.

Total's pathway to sustainability is calculated by setting the constraints and running the GA optimiser. Its lever values suggest that meeting the climate warming target while minimising economic impact is realistic but requires an increase in the current effort level in most sectors.

6 Conclusions

The first aim of this investigation was to explore and learn more about the Global Calculator. This has been addressed by adapting MCMC to the model and calculating correlation maps of the model's inputs and outputs, as well as probability distributions of model inputs that maximise the likelihood of a climate-friendly and economically viable future.

The second aim was to perform a multiobjective optimisation of the model inputs. GA has succeeded in this endeavour, allowing the user to specify multiple conflicting objectives. This yields bespoke optimal pathways to sustainability. CMA-ES has failed to converge and ANNs have shown that a massive amount of training data would be required to reach an acceptable accuracy.

The final aim was to generate Total's 2050 pathway. The optimiser returned an optimal input combination that accounts for two conflicting objectives and twelve input constraints. Total is a major energy player and climate is at the core of its strategy. This pathway contributes to its mission of moving towards a more sustainable future.

7 Code metadata

Software version	v.1.0
Permanent link to code	https://github.com/acse-2019/irp-acse-jg719
Legal Software License	MIT
Code versioning system used	git
Software code languages, tools, and services used	python (>=3.6), Jupyer Notebook (>6.1.3)
Installation requirements & dependencies	numpy (>= 1.19.1), pandas (>= 1.1.0), matplotlib (>= 3.3.1), torch (>= 1.6.0), selenium (>= 4.0.0)
Operating system requirements	MacOS, Windows, Linux
Link to developer documentation	https://jg719.github.io/IRP-acse-jg719-documentation/
Support email	jg719@ic.ac.uk

Table 8: Code metadata.

8 Bibliography

[1] "An early view on the results of the 2050 Calculator International Outreach", UK Department of Energy & Climate Science, 2015. [Accessed June 24 2020].

[2] "Prosperous living for the world in 2050: insights from the Global Calculator", UK Department of Energy & Climate Change, 2015. [Accessed June 24 2020].

[3] "The Global Calculator Sector metrics from 2 degree pathways", UK Department of Energy & Climate Change, 2020. [Accessed June 24 2020].

[4] A. Strapasson et al., "Modelling carbon mitigation pathways by 2050: Insights from the Global Calculator", Energy Strategy Reviews, vol. 29, p. 100494, 2020. Available: 10.1016/j.esr.2020.100494.

[5] A. Strapasson, J. Woods, H. Chum, N. Kalas, N. Shah and F. Rosillo-Calle, "On the global limits of bioenergy and land use for climate change mitigation", GCB Bioenergy, vol. 9, no. 12, pp. 1721-1735, 2017. Available: 10.1111/gcbb.12456.

[6] A. Elizondo, V. Pérez-Cirera, A. Strapasson, J. Fernández and D. Cruz-Cano, "Mexico's low carbon futures: An integrated assessment for energy planning and climate change mitigation by 2050", Futures, vol. 93, pp. 14-26, 2017. Available: 10.1016/j.futures.2017.08.003.

[7] "Editorial, Environmental Development, December 2019.", Environmental Development, vol. 32, p. 100471, 2019. Available: 10.1016/j.envdev.2019.100471.

[8] "United nations expert group meeting on population, food security, nutrition and sustainable development for sustainable development", United Nations Secretariat, Department of Economic and Social Affairs, 2019. [Accessed June 24 2020].

[9] M. Bailera, P. Lisbona, S. Espatolero and L. Romeo, "Power to Gas technology under Spanish Future Energy Scenario", Energy Procedia, vol. 114, pp. 6880-6885, 2017. Available: 10.1016/j.egypro.2017.03.1828.

[10] G. Yannakakis and J. Togelius, Artificial Intelligence and Games. 2018.

[11] M. Kasim, A. Bott, P. Tzeferacos, D. Lamb, G. Gregori and S. Vinko, "Retrieving fields from proton radiography without source profiles", Physical Review E, vol. 100, no. 3, 2019. Available: 10.1103/physreve.100.033208.

[12] S. Ferrari and R. Stengel, "Smooth function approximation using Neural Networks", IEEE Transactions on Neural Networks, vol. 16, no. 1, pp. 24-38, 2005. Available: 10.1109/tnn.2004.836233.

[13] L. Shiyu and R. Srikant, "Why deep neural networks for function approximation?", Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, 2017. [Accessed June 24 2020].

[14] J. Rogelj et al., "Paris Agreement climate proposals need a boost to keep warming well below 2C", Nature, vol. 534, no. 7609, pp. 631-639, 2016. Available: 10.1038/nature18307.

[15] A. Strapasson, J. Woods and A. Donaldson, "Bulb Calculator: An Independent Review", Report commissioned by Bulb Energy Limited, 2020. [Accessed June 24 2020].

[16] A. Gupta and J. Rawlings, "Comparison of parameter estimation methods in stochastic chemical kinetic models: Examples in systems biology", AIChE Journal, vol. 60, no. 4, pp. 1253-1268, 2014. Available: 10.1002/aic.14409.

[17] "Bayesian methods: a social and behavioral sciences approach", Choice Reviews Online, vol. 40, no. 06, pp. 40-3452-40-3452, 2003. Available: 10.5860/choice.40-3452.

[18] D. Scott, "Introducing Monte Carlo Methods with R by Christian P. Robert, George Casella", International Statistical Review, vol. 78, no. 3, pp. 476-477, 2010. Available: 10.1111/j.1751-5823.2010.00122-29.x.

[19] F. Briol, C. Oates, M. Girolami, M. Osborne and D. Sejdinovic, "Probabilistic Integration: A role in Statistical Computation?", Statistical Science, vol. 34, no. 1, pp. 1-22, 2019. Available: 10.1214/18-sts660.

[20] S. Mohandas, "A.I. for games with high branching factor", 2018 International CET Conference on Control, Communication, and Computing (IC4), 2018. [Accessed June 24 2020].

[21] "The Global Calculator spreadsheet v.3.99.0", Tool.globalcalculator.org, 2020. [Online]. Available: http://tool.globalcalculator.org. [Accessed: 19- Aug- 2020].

[22] F. Rigat and A. Mira, "Parallel hierarchical sampling: A general-purpose interacting Markov chains Monte Carlo algorithm", Computational Statistics & Data Analysis, vol. 56, no. 6, pp. 1450-1467, 2012. Available: 10.1016/j.csda.2011.11.020.

[23] M. Caramia and P. Dell'Olmo, "Multi-objective Optimization", Multi-objective Management in Freight Logistics, pp. pp 21-51, 2020. Available: https://doi.org/10.1007/978-3-030-50812-8 2 [Accessed 27 August 2020].

[24] D. Gong, Y. Liu and G. Yen, "A Meta-Objective approach for many-objective Evolutionary Optimization", Evolutionary Computation, vol. 28, no. 1, pp. 1-25, 2020. Available: 10.1162/evco a 00243.

[25] H. Fu and P. Liu, "A Multi-Objective Optimization model based on Non-Dominated Sorting Genetic Algorithm", International Journal of Simulation Modelling, vol. 18, no. 3, pp. 510-520, 2019. Available: 10.2507/ijsimm18(3)co12.

[26] H. Chen, R. Cheng, J. Wen, H. Li and J. Weng, "Solving large-scale many-objective optimization problems by covariance matrix adaptation evolution strategy with scalable small subpopulations", Information Sciences, vol. 509, pp. 457-469, 2020. Available: 10.1016/j.ins.2018.10.007.

[27] X. Dou and T. Liang, "Training Neural Networks as Learning Data-Adaptive Kernels: Provable Representation and Approximation Benefits", Journal of the American Statistical Association, pp. 1-14, 2020. Available: 10.1080/01621459.2020.1745812.

[28] L. Song, "On the Complexity of Learning Neural Networks," 31st Conference on Neural Information Processing Systems (NIPS 2017).

[29] Department of Energy & Climate Change, "The Global Calculator," GOV.UK, 28-Jan-2015. [Online]. Available: https://www.gov.uk/government/publications/the-global-calculator. [Accessed: 19-Aug-2020].

[30] Decc, "decc/global-calculator," GitHub. [Online]. Available: https://github.com/decc/global-calculator. [Accessed: 19-Aug-2020].

[31] Unfccc.int. 2020. [online] Available at: https://unfccc.int/process-and-meetings/the-paris-agreement/the-paris-agreement [Accessed 21 August 2020].

[32] M. Gregory, S. Martin and D. Werner, "Improved Electromagnetics Optimization: The covariance matrix adaptation evolutionary strategy.", IEEE Antennas and Propagation Magazine, vol. 57, no. 3, pp. 48-59, 2015. Available: 10.1109/map.2015.2437277.

[33] "The Global Calculator — spreadsheet v.3.99.0", Tool.globalcalculator.org, 2020. [Online]. Available: http://tool.globalcalculator.org/globcalc.html?levers=i3233q333i2pj3qqrmh33ji31fsimvolr3o32msgmc111112p22 11111111/dashboard/en. [Accessed: 27- Aug- 2020].

[34] B. Thijssen and L. Wessels, "Approximating multivariate posterior distribution functions from Monte Carlo samples for sequential Bayesian inference", PLOS ONE, vol. 15, no. 3, p. e0230101, 2020. Available: 10.1371/jour-

nal.pone.0230101.

[35] V. Elvira, L. Martino, D. Luengo and M. Bugallo, "Generalized Multiple Importance Sampling", Statistical Science, vol. 34, no. 1, pp. 129-155, 2019. Available: 10.1214/18-sts668.

[36] V. Apidopoulos, J. Aujol and C. Dossal, "Convergence rate of inertial Forward–Backward algorithm beyond Nesterov's rule", Mathematical Programming, vol. 180, no. 1-2, pp. 137-156, 2018. Available: 10.1007/s10107-018-1350-9.

[37] S. Rouchier, M. Jiménez and S. Castaño, "Sequential Monte Carlo for on-line parameter estimation of a lumped building energy model", Energy and Buildings, vol. 187, pp. 86-94, 2019. Available: 10.1016/j.enbuild.2019.01.045.

[38] K. De Jong, "Evolutionary computation: a unified approach", GECCO '20: Proceedings of the 2020 Genetic and Evolutionary Computation Conference Companion, pp. Pages 327–342, 2020. Available: https://doi.org/10.1145/3377929.3389871.

[39] T. Qin, K. Wu and D. Xiu, "Data driven governing equations approximation using deep neural networks", Journal of Computational Physics, vol. 395, pp. 620-635, 2019. Available: 10.1016/j.jcp.2019.06.042.

Appendix A: Other relevant figures.

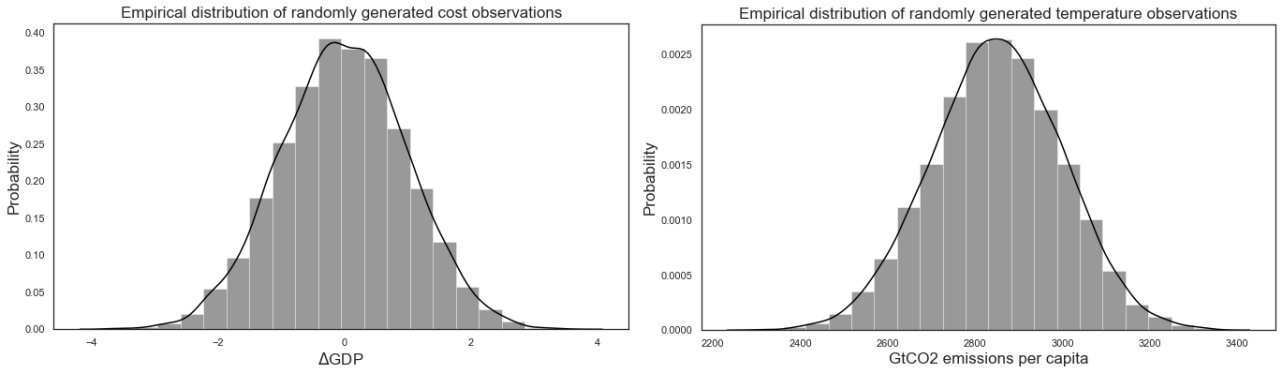


Figure 5: Empirical distribution of constraints - Randomly generated temperature and cost values.

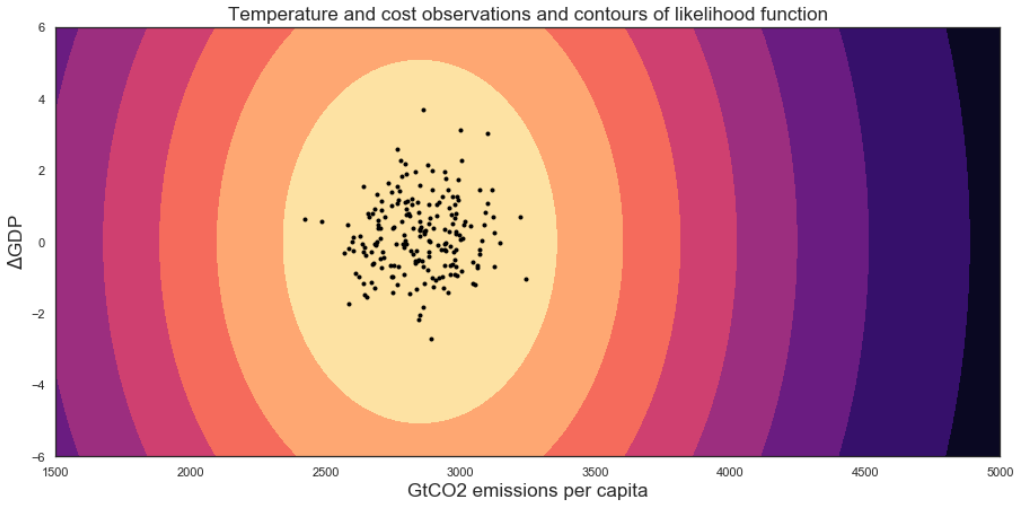


Figure 6: MCMC observations and contour plot of the likelihood function.

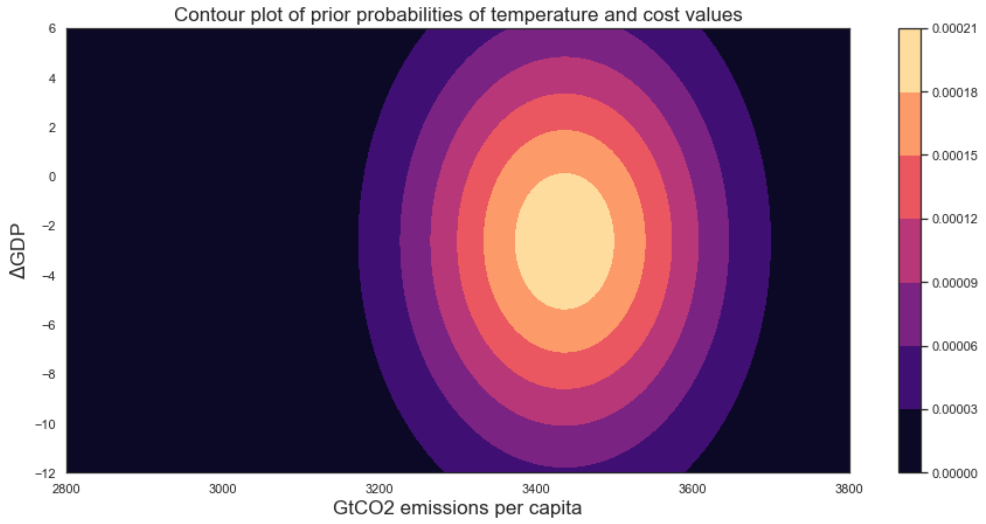


Figure 7: MCMC contour plot of empirical prior distribution.

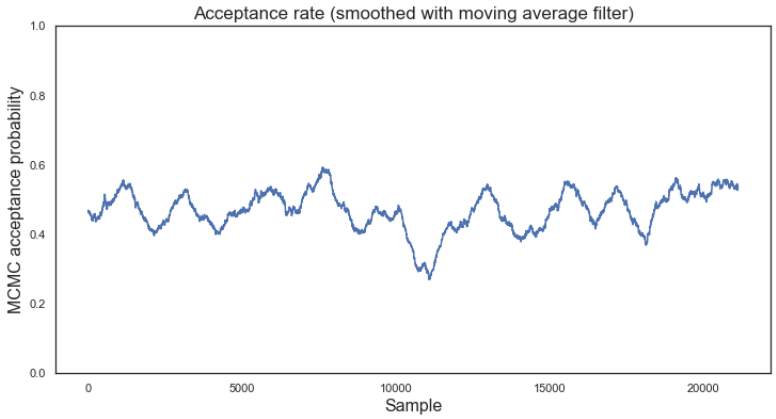


Figure 8: MCMC acceptance rate as a function of the number of proposed model parameters.

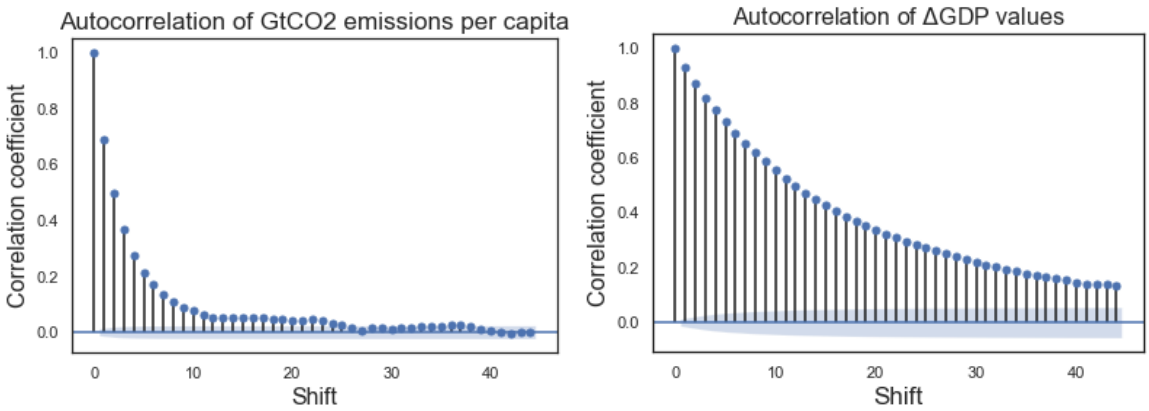


Figure 9: Autocorrelation function of GtCO2 emissions per capita and changes in GDP values.

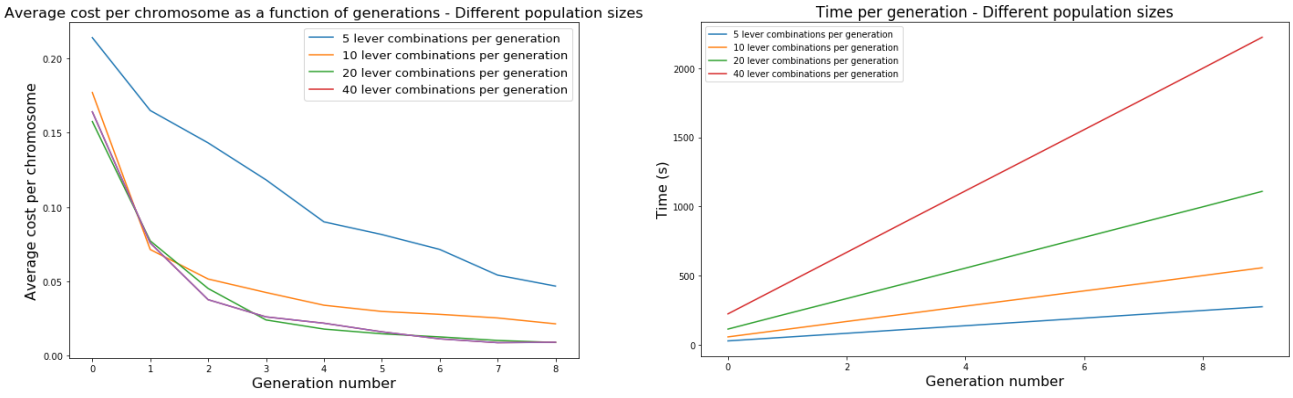


Figure 10: GA average cost per chromosome and computing time as a function of generations for different population sizes.

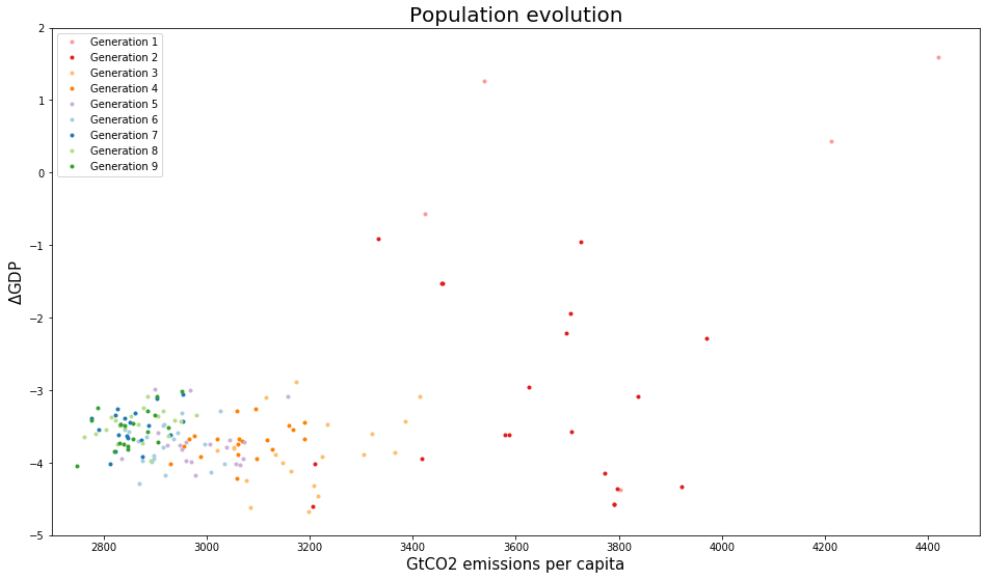


Figure 11: GA population evolution of Total's pathway - GHG emissions per capita and changes in GDP

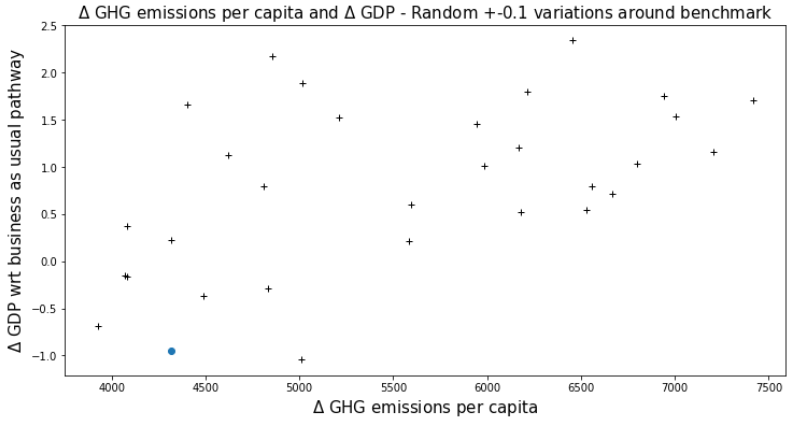


Figure 12: Variations in GHG emissions per capita and GDP changes around benchmark pathway for random lever moves.

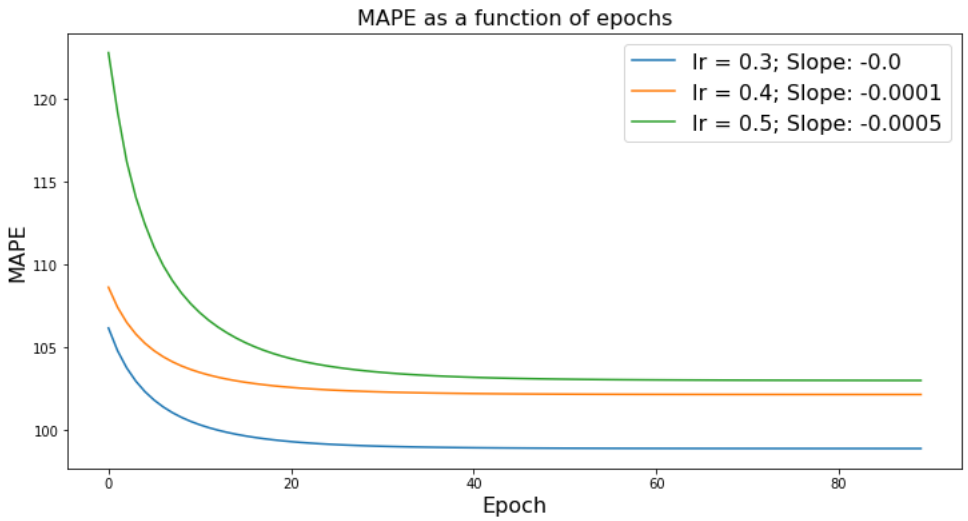


Figure 13: ANN evolution of MAPE as a function of epochs.

Appendix B: Mono-parameter sensitivity analysis algorithm.

Algorithm 3: Mono-parameter sensitivity analysis.

```
L: Input lever values                              shape: [number of input levers, 1]
Emissions_values ← []                              Destination of per capita GHG emissions
Cost_values ← []                                   Destination of GDP changes
procedure Read_outputs(lever_values)               Given a set of lever values, read outputs
    emissions ← TheGlobalCalculator(lever_values).emissions
    cost ← TheGlobalCalculator(lever_values).cost
    return emissions, cost
end procedure
procedure
L[:, 1] ← 2                                        Set all levers to 2
for i in range(len(L[:, 1])):                      Set lever to 3, read output and set back to 2
    L[i, 1] ← 3
    Emissions_values, Cost_values ← Read_outputs(L[:, 1])
    L[i, 1] ← 2
end for
end procedure
```

Appendix C: Model outputs – Correlation matrix and map

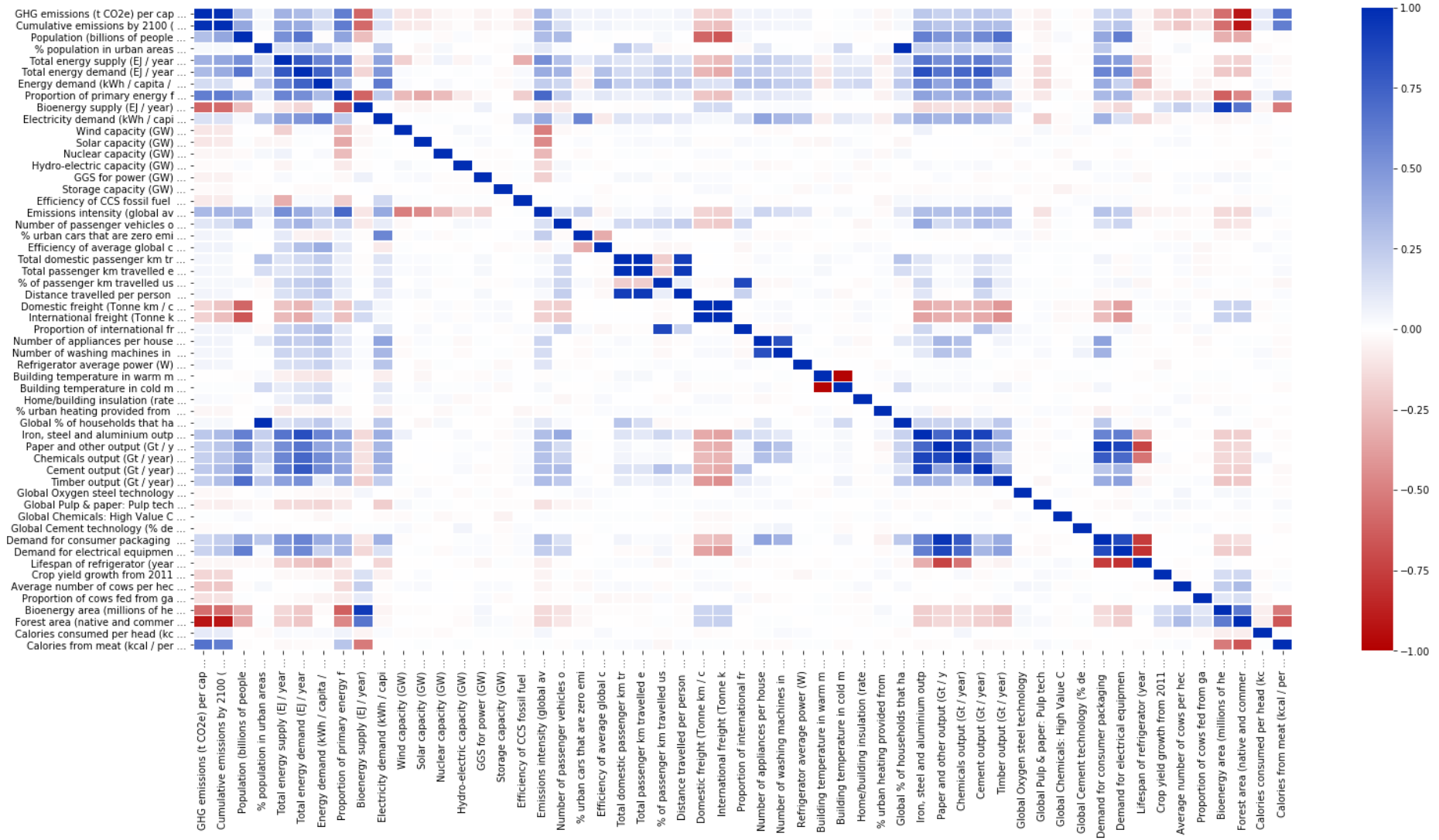


Figure 14: Correlation matrix of model outputs.

Figure 15: Correlation map of model outputs.

Appendix D: MCMC accepted model inputs - Correlation matrix and map

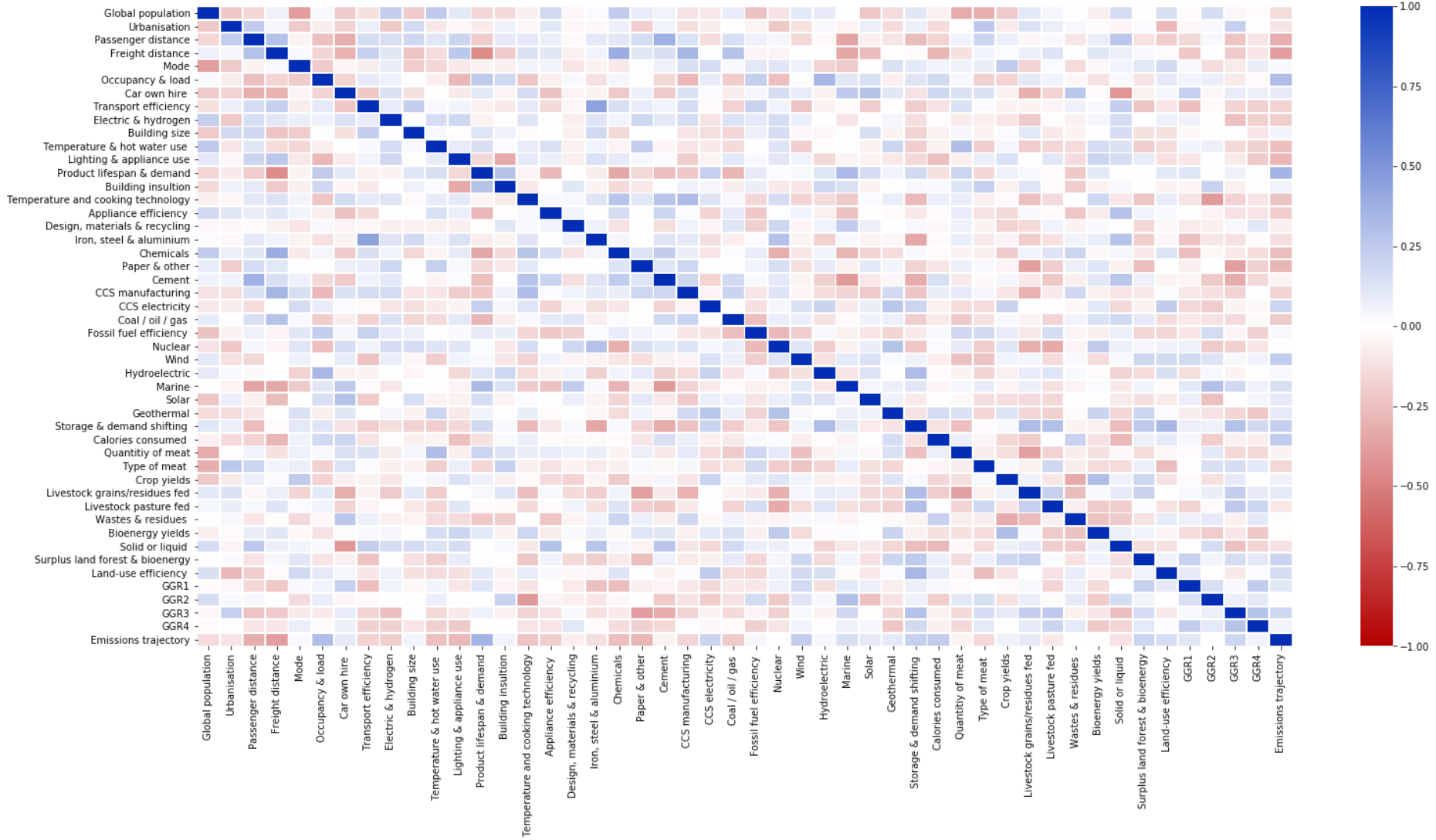
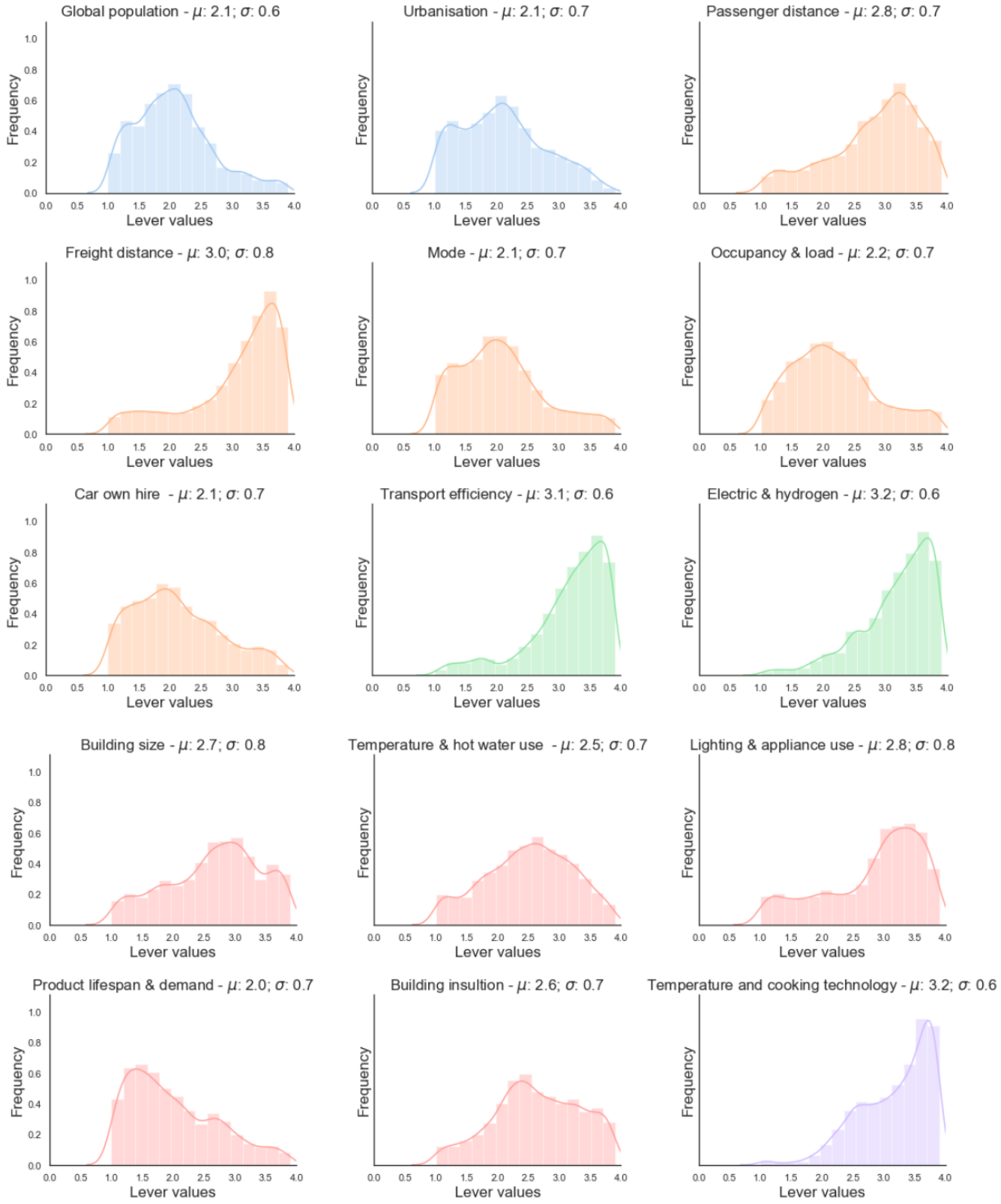
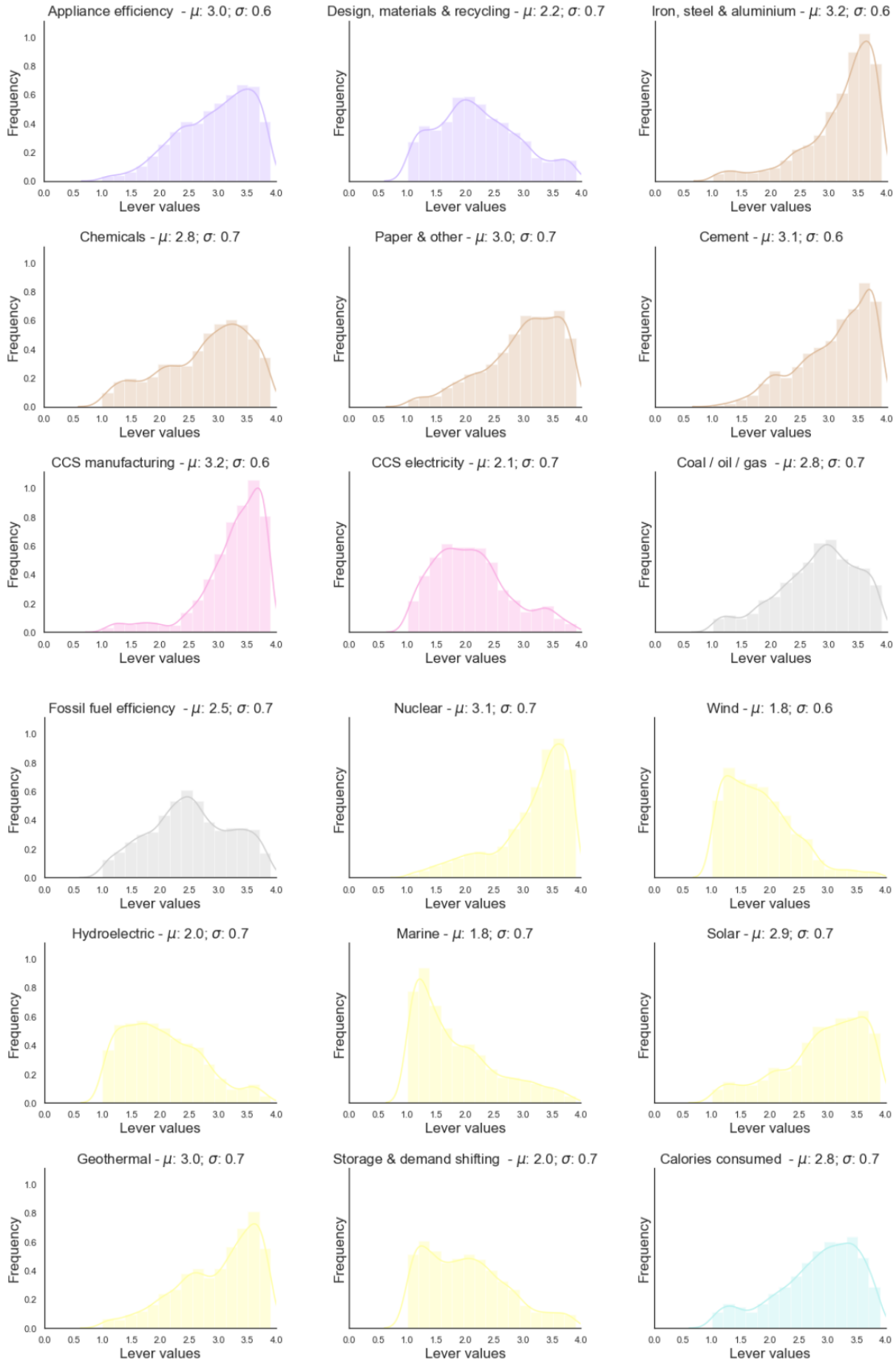


Figure 16: Correlation matrix of model inputs.

Figure 17: Correlation map of model inputs.

Appendix E: Posterior distributions of model inputs derived using MCMC




Appliance efficiency - μ: 3.0; σ: 0.6
Design, materials & recycling - μ: 2.2; σ: 0.7
Iron, steel & aluminium - μ: 3.2; σ: 0.6
Chemicals - μ: 2.8; σ: 0.7
Paper & other - μ: 3.0; σ: 0.7
Cement - μ: 3.1; σ: 0.6
CCS manufacturing - μ: 3.2; σ: 0.6
CCS electricity - μ: 2.1; σ: 0.7
Coal / oil / gas - μ: 2.8; σ: 0.7
Fossil fuel efficiency - μ: 2.5; σ: 0.7
Nuclear - μ: 3.1; σ: 0.7
Wind - μ: 1.8; σ: 0.6
Hydroelectric - μ: 2.0; σ: 0.7
Marine - μ: 1.8; σ: 0.7
Solar - μ: 2.9; σ: 0.7
Geothermal - μ: 3.0; σ: 0.7
Storage & demand shifting - μ: 2.0; σ: 0.7
Calories consumed - μ: 2.8; σ: 0.7
Frequency
Lever values

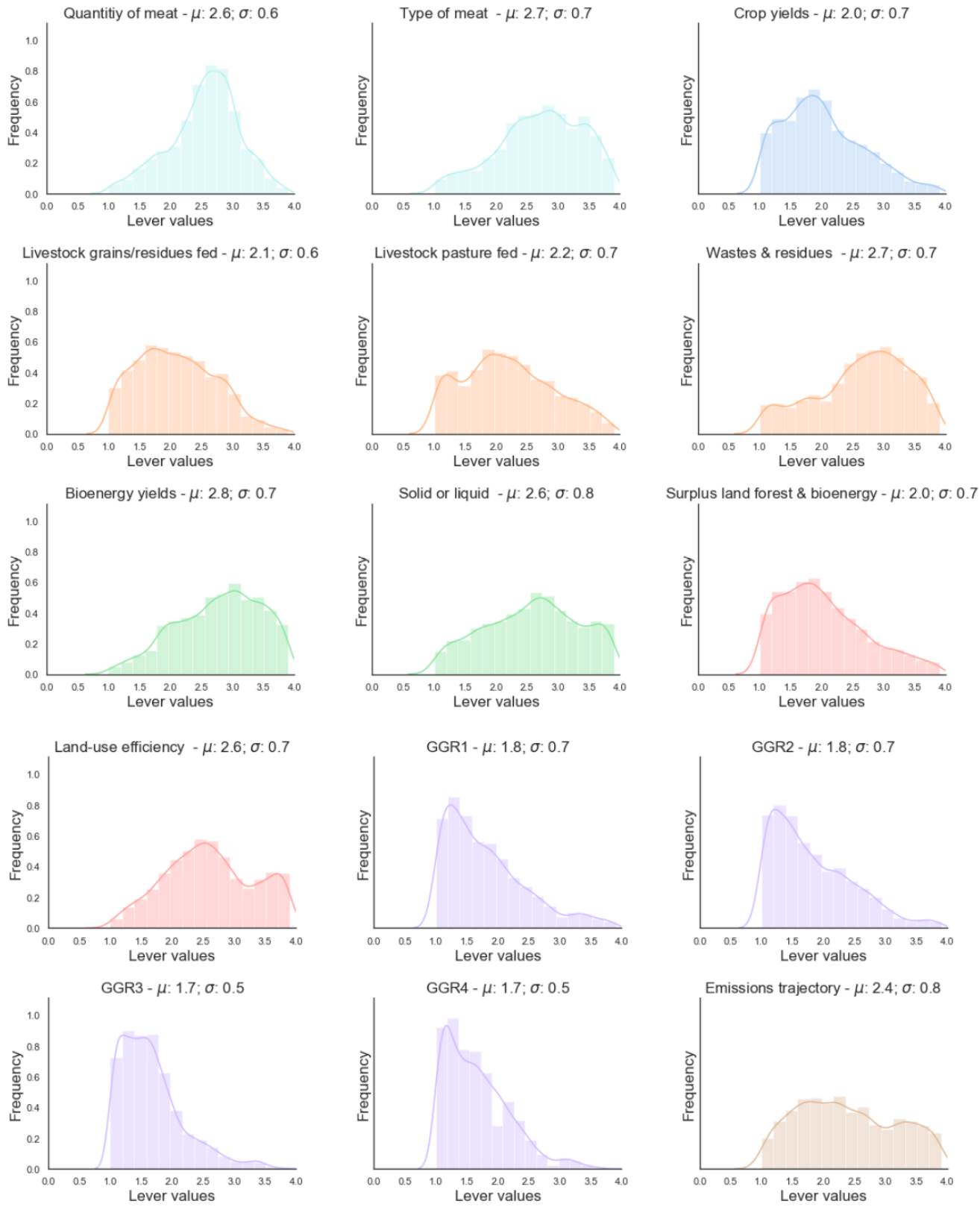


Figure 18: Posterior distributions of model inputs.

Appendix F: Paired density and scatter plot matrix of accepted MCMC model inputs

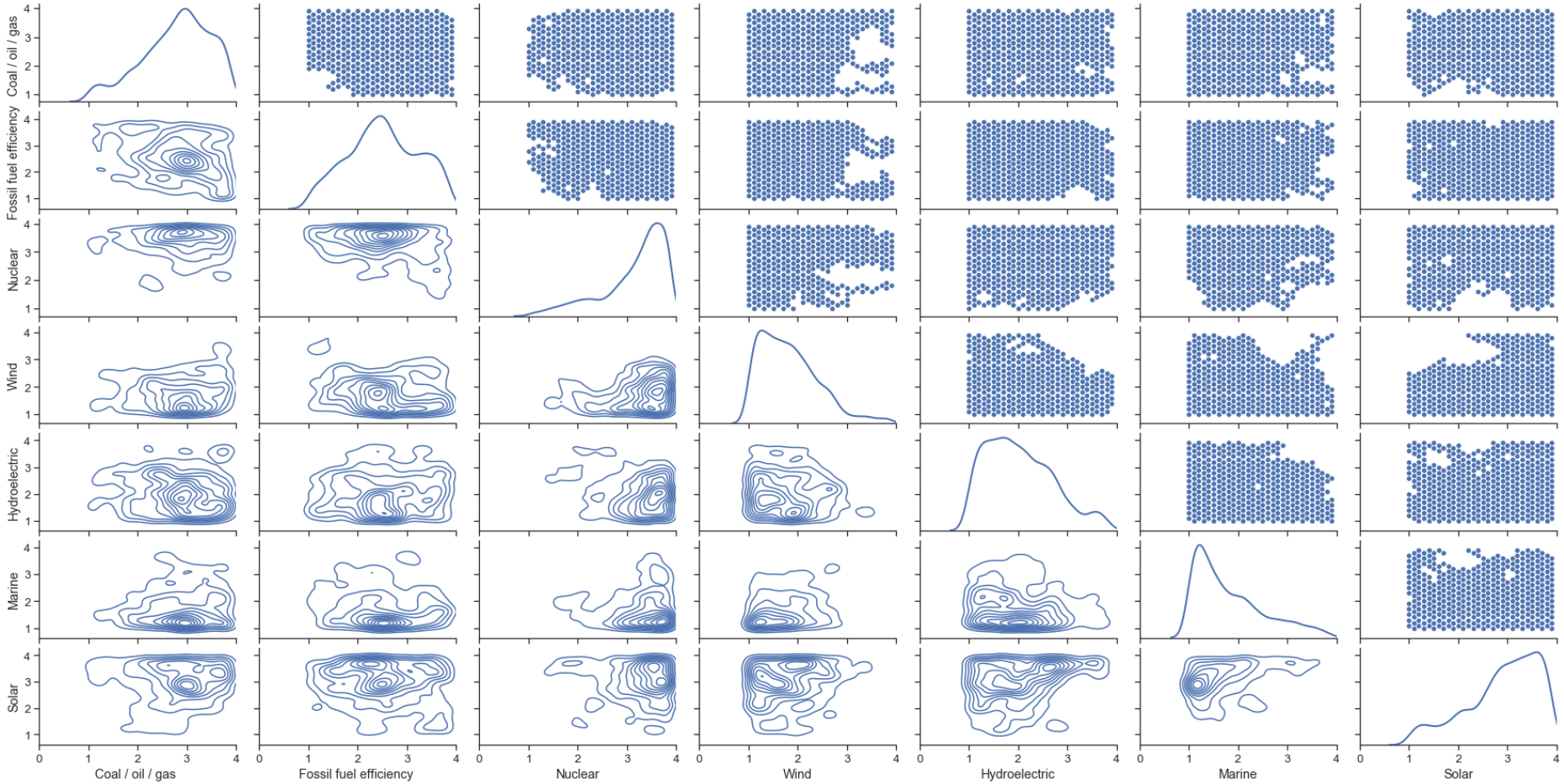


Figure 19: Paired density and scatter plot matrix of lever combinations accepted by MCMC.

Appendix G: Total's sustainable pathway and benchmark pathways.

Lever	Total's pathway	Distributed effort	Consumer reluctance	Low action on forests	Consumer activism	IEA 6DS (approx.)
Global population	1.8	2.0	2.0	2.0	2.0	2.0
Urbanisation	3.0	2.0	2.0	2.0	2.0	2.0
Passenger distance	2.0	2.7	2.7	2.7	2.7	2.7
Freight distance	3.0	1.5	1.5	1.5	1.5	1.5
Mode	3.0	2.4	2.4	2.4	3.0	2.4
Occupancy and load	2.6	1.4	1.4	1.4	2.0	1.4
Car own or hire	3.0	2.0	2.0	2.0	2.4	2.0
Efficiency	3.0	2.8	2.8	3.0	3.0	1.4
Electric and hydrogen	3.0	2.8	2.0	3.0	3.0	1.0
Building size	1.8	3.0	3.0	3.0	3.0	3.0
Temperature & hot water use	2.0	1.1	1.1	1.1	1.1	1.1
Lighting, cooking & appliance use	2.5	1.4	1.4	1.4	1.4	1.4
Building insulation	1.9	2.8	2.0	3.0	3.0	1.0
Temperature, cooking & lighting technology	3.0	2.8	2.0	3.0	3.0	1.0
Appliance efficiency	2.6	2.8	3.0	3.0	3.0	1.0
Product lifespan	2.6	1.0	1.0	1.0	2.0	1.0
Design, material switch & recycling	2.7	2.8	2.0	3.0	3.0	1.2
Iron, steel & aluminium	2.2	2.8	3.0	3.0	2.0	2.0
Chemicals	1.7	2.8	3.0	3.0	2.0	1.2
Paper and other	3.0	2.8	3.0	3.0	2.0	2.0
Cement	3.0	2.8	3.0	3.0	2.0	1.2
Carbon capture and storage (ind.)	1.9	2.8	3.0	3.0	2.0	1.0
Coal / oil / gas	1.8	2.8	3.0	3.0	3.0	2.3
Fossil fuel efficiency	3.0	2.8	3.0	3.0	3.0	3.0
Carbon capture and storage (power)	1.0	2.8	3.0	2.0	3.0	1.0
Nuclear	1.5	2.8	3.0	2.8	2.0	1.7
Wind	2.8	2.8	2.7	3.0	2.0	1.5
Hydroelectric	1.8	2.8	2.7	3.0	2.0	1.9
Marine	2.2	2.8	2.7	3.0	2.0	1.3
Solar	3.1	2.8	2.0	3.0	3.0	1.2
Geothermal	2.4	2.8	2.7	3.0	2.0	1.4
Storage and demand shifting	2.1	2.8	2.7	3.0	2.0	1.5
Calories consumed	2.7	2.0	2.0	2.0	2.0	2.0
Quantity of meat	3.0	2.0	2.0	2.0	2.2	2.0
Type of meat	2.4	2.0	2.0	2.0	3.0	2.0
Crop yields	3.0	2.8	3.0	2.0	2.0	1.7
Land-use efficiency	2.0	2.8	3.0	3.0	3.0	2.5
Livestock (grains/residues fed)	2.2	2.8	3.0	2.0	1.5	2.0
Livestock (pasture fed)	2.8	2.8	3.0	3.0	3.0	3.0
Bioenergy yields	1.6	2.8	3.0	3.0	2.0	3.0
Solid or liquid	2.2	2.8	3.0	2.0	2.0	1.5
Surplus land (forest & bioenergy)	1.2	2.8	3.0	3.0	2.0	2.0
Biochar	1.0	1.0	1.0	1.0	1.0	1.0
Direct air capture	1.0	1.0	1.0	1.0	1.0	1.0
Ocean fertilisation	1.0	1.0	1.0	1.0	1.0	1.0
Enhanced weathering (oceanic)	1.0	1.0	1.0	1.0	1.0	1.0
Enhanced weathering (terrestial)	1.0	1.0	1.0	1.0	1.0	1.0
Wastes and residues	2.0	2.8	2.0	2.5	3.0	1.5
Emissions trajectory	2.5	2.7	2.8	3.0	2.8	2.3
Atmospheric CO2 fraction	2.0	2.0	2.5	2.0	2.0	2.0
Confidence in climate models	B	B	B	B	B	B

Table 9: Total's sustainable pathway and other benchmark pathways.

Appendix H: Total's sustainable pathway compared to business as usual pathway

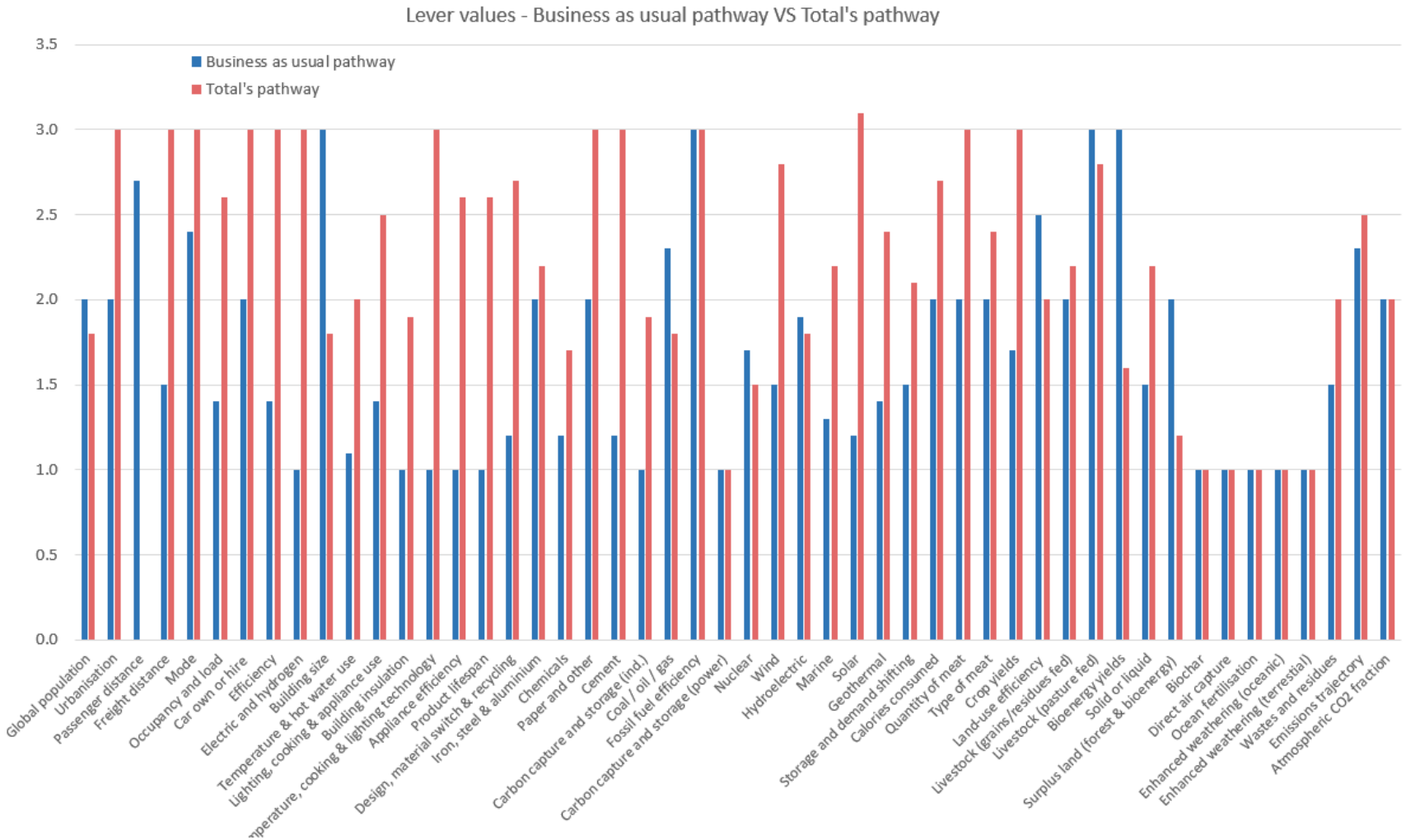


Figure 20: Total's pathway VS Business as usual pathway.