A Latin Hypercube Sampling Utility: with an application to an Integrated Assessment Model

Dominique van der Mensbrugghe[†]

11 April 2023

Abstract: This paper describes the use of a utility that creates a Latin Hypercube Sample (LHS). The LHS approach to sampling has had wide applicability as it represents a Monte Carlo strategy that limits sample size and therefore computer time to study the outcomes of simulations under uncertainty. The utility is a new version of the LHS utility that has been publicly available from Sandia National Labs since the early 2000s. Beyond the recoding from FORTRAN to C/C++, the new version of the utility has some additional features including new output options and additional statistical distributions. This paper demonstrates the use of the new utility by coupling it to an integrated assessment (IAM) model which is derived from the Meta21 model developed by (Dietz et al., 2021). The Meta21 model, beyond demonstrating the LHS approach to taking into account uncertainty, also has many components that can be readily integrated into global economic models that track greenhouse gas emissions—a simple climate module, economic impacts derived from sea-level and temperature rises and bio-physical tipping points such as the Amazon dieback. The utility and the code to the IAM model are available as supplementary materials.

JEL Codes: C15, C65, O44, Q51, Q54
Keywords: Monte Carlo, Latin Hypercube Sampling, Integrated Assessment

[†] The Center for Global Trade Analysis, Department of Agricultural Economics, Purdue University. The paper has been prepared for the 26th Annual Conference on Global Economic Analysis, Bordeaux, France, 14-16 June 2023. Corresponding e-mail: vandermd@purdue.edu.

1 Introduction

This paper describes the use of a utility that creates a Latin Hypercube Sample (LHS). The LHS approach to sampling has had wide applicability as it represents a Monte Carlo strategy that limits sample size and therefore computer time to study the outcomes of simulations under uncertainty. The utility is a new version of the LHS utility that has been publicly available from Sandia National Labs (Swiler and Wyss (2004)) since the early 2000s. Beyond the recoding from FORTRAN to C/C++, the new version of the utility has some additional features including new output options and additional statistical distributions. This paper demonstrates the use of the new utility by coupling it to an integrated assessment (IAM) model which is derived from the Meta21 model developed by Dietz et al. (2021). The Meta21 model, beyond demonstrating the LHS approach to taking into account uncertainty, also has many components that can be readily integrated into global economic models that track greenhouse gas emissions—a simple climate module, economic impacts derived from sea-level and temperature rises and bio-physical tipping points such as the Amazon dieback. The utility and the code to the IAM model are available as supplementary materials.

The LHS utility has been compiled as a stand-alone 'exe' file for Windows-based computers,[1] and can readily be embedded in a complex workflow. It has five output options for the sample: (1) a 'CSV' file; (2) an 'XML' file, which is ready for use in Excel; (3) a 'GMS' text file for GAMS; (4) a 'GDX' file, also for use with GAMS; and (5) the original file format of the FORTRAN code. The LHS utility includes over 40 statistical distributions and the ability to impose a correlation relation across two or more of the sampled variables. A summary user guide is available in the supplementary materials and is meant to document this version of the LHS utility— and for most users should be sufficient—, but the more complete documentation is available in Swiler and Wyss (2004), save for the new features included with this version of the LHS utility.

A subsequent section illustrates the use of the LHS utility and describes an integrated assessment model derived from the Meta21 model (Dietz et al. (2021)) but with a subset of the tipping points from the original model.[2] The Meta21 model has many similarities with the DICE model (Nordhaus (2017)), but with a number of additional features, notably the incorporation of tipping points. Greenhouse gas emissions are largely exogenous in the Meta21 model—with the exception of those linked to tipping points such as the Amazon dieback—but the model has many features that can be easily incorporated in a standard global computable general equilibrium (CGE) model that track greenhouse gas emissions. The Meta21 model also demonstrates how to incorporate random events in a model. In the Meta21 model the random events are related to tipping points. But examples in economics abound such as recessions, pandemics, natural catastrophes, conflicts, etc. The original model is 'coded' in Excel and uses the external Excel add-in called @Risk to generate the random sample.[3] The final section describes how to couple the LHS utility with the IAM and highlights some of the key findings from the uncertainty analysis.

[1] The C/C++ code is intended to be fully ANSI compatible and should compile readily on UNIX-based computers including the Macintosh operating system.

[2] The model code, in GAMS, includes all of the features of the full Meta21 model, but the more complex features of the model are not described herein.

[3] @Risk uses the same type of sampling methodology as the LHS utility.

2 Latin Hypercube Sampling

2.1 An introduction to Latin Hypercube Sampling

Monte Carlo methods are well established techniques to explore model or simulation uncertainty when the underlying model is based on uncertain parameters (but with an assumed distribution). For each uncertain parameter, a draw of a random deviate is taken from the known distribution—and if there are k uncertain parameters, each draw will consist of selecting a random deviate for each of the k uncertain parameters. Latin Hypercube Sampling (LHS) (McKay et al. (1979)) was introduced as an efficient sampling method, where efficiency allows for limiting the size of the sample while at the same time insuring that the entire sample space is covered.

Before explaining LHS, we describe first the technique for generating random deviates. If a distribution has an easily invertible cumulative distribution function (CDF), the first step is to generate a random deviate from the Uniform distribution over the range $[0, 1]$ and then to use the inverse function to generate the random deviate.[4] This is the so-called inverse transform method.

We highlight the technique using the Logistic distribution, which has a readily invertible CDF. Its CDF is given by:

$$F(x) = \frac{1}{1 + e^{-(x-\mu)/s}}$$

where μ is the location parameter and s is the scale parameter. The inverse of the CDF, also known as the quantile function, is easily derived for any value of F, say p:

$$x = Q(p) = \mu + s \ln\left(\frac{p}{1-p}\right)$$

Figure 1 plots an example Logistic CDF with parameters $\mu = 9$ and $s = 3$. We randomly select 5 deviates from the Uniform distribution: $0.12, 0.36, 0.54, 0.79, 0.95$, visible along the y-axis. We then generate the random deviates for the Logistic distribution by locating the Uniform deviates on the y-axis of the CDF and evaluating the corresponding Logistic deviates, respectively 3.0, 7.3, 9.5, 13.0, and 17.8, easily derived from the Logistic quantile function.

In the standard Monte Carlo method, a random deviate is sampled each time from the entire domain of the random variable, i.e. over $[0, 1]$. In the case of selecting a sample of 5, nothing guarantees that the sample will cover the full range. For example, 4 of the numbers could cluster around 0.65 with a fifth at 0.35. LHS sampling breaks the $[0, 1]$ range into 5 equal segments of width 0.2 (for a sample of 5). Sampling is done within each of the sub-ranges, guaranteeing that the sample will cover (almost) the entire range. In the more general case, with a sample size of n, LHS sampling divides the $[0, 1]$ range into equivalent sub-ranges with a width of $1/n$ and samples within the sub-ranges. The efficiency argument is that LHS allows to have a smaller sample size than standard Monte Carlo sampling, while ensuring that the full potential sample range is sampled. The technique is implemented for each random variable.

Figure 2 highlights a simple example of the potential benefits of LHS sampling.[5] The example uses the Logistic distribution with a location parameter of 0 and a scale parameter of 2. The mean of this distribution is 0 and the standard deviation is 3.628.[6] The x-axis of each graph, shows the

[4] Not all CDF's have an easy expression for their corresponding quantile function, including the Normal distribution. In the case of the latter, the inverse error function is readily available in most software packages. The Beta distribution is an example where the inverse CDF is not readily available. In these cases, other numerical techniques are deployed to generate random deviates.

[5] The GAMS code that generates the data for the figure is available upon request.

[6] $\sigma = \sqrt{\frac{(\pi s)^2}{3}}$, where s is the scale parameter.

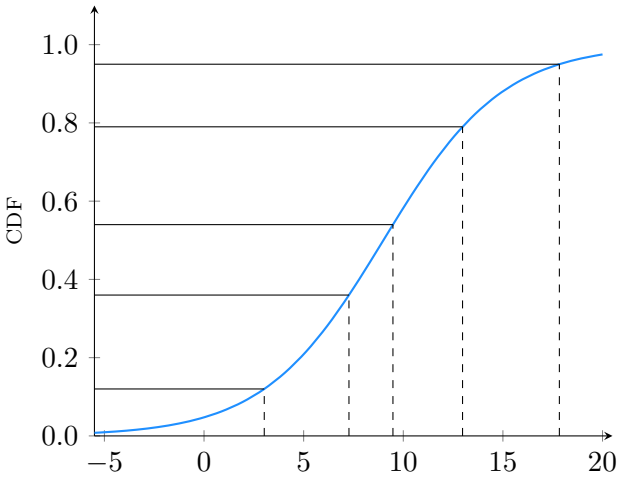
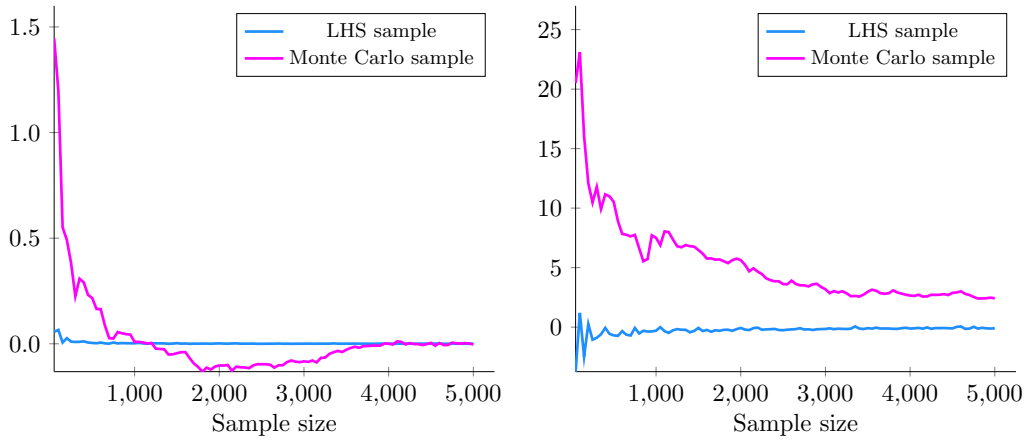


Figure 1: The Logistic distribution CDF and selecting random deviates

number of draws for each sample, ranging from 50 to 5000 in steps of 50. The left-panel shows the deviation of the sample mean from the expected mean of 0 and the right-panel shows the percent deviation of the standard deviation from the expected standard deviation. Both figures clearly indicate that the LHS samples very quickly converge to the expected moments of the distribution. The Monte Carlo simulations show persistent deviations in the standard deviation, even with 5,000 draws.[7]



(a) Deviation from the expected mean (0)

(b) Deviation from expected standard deviation, %

Figure 2: Comparison of LHS versus Monte Carlo sampling for a Logistic(0,2) distribution

As coded in the LHS utility, the sample for each variable is in ascending order, as the samples are derived from each sub-range in ascending order. The deviates could be further randomized by

[7] The advantage of LHS sampling with many uncertain parameters is disputed, though part of the issue relates to memory and computer time, which is disappearing as an issue. A blog post by Lonnie Chrisman (https://lumina.com/latin-hypercube-vs-monte-carlo-sampling/) summarizes part of this discussion and provides additional references.

sampling without replacement each of the deviates of each variable. The LHS utility does something a bit different. In the absence of user-specified correlations, it re-orders each column of deviates in such a way that the resulting correlation matrix across variables is (nearly) the identity matrix, while preserving the marginal distribution of each column. This is called the Iman-Conover method (Iman and Conover (1982)). The method is explained intuitively and with a numerical example in the Appendix. The same method is used if the user specifies a correlation matrix for all or some of the variables—a correlation of 0 is assumed for all pairs of variables for which a correlation is not user-provided.

In summary, the LHS method for Monte Carlo sampling, uses a method of stratified sampling, which is an efficient method to generate a sample that is nearly sure to cover the entire sample space. In addition, the deviates of each variable are ordered in such a way to (nearly) match a desired correlation scheme, which in the absence of any user-imposed correlations is the identity matrix.

2.2 An introduction to the LHS utility

The LHS utility is a program that generates a LHS sample based on a user-prepared input file describing the contours of the LHS cube: the list of random variables with their respective distributions and distribution parameters, and optionally correlations across pairs of the random variables. The input file is a simple text file and its contents is described in the accompanying LHS User Guide. The output of the LHS utility is a data file containing the sample. The user has several output options including a CSV file, an XML file (readily loaded into Excel), a text-based GAMS data file and a GDX file.[8]

The LHS utility is a full recoding of the original utility, which has been developed in FORTRAN and can be downloaded from https://dakota.sandia.gov/downloads.[9] The re-coding has been done in C/C++ using mostly ANSI-C standards and thus should be readily compilable on other platforms, notably Linux and macOS.[10]

The two versions have been extensively tested and produce virtually the same sample.[11] The key advantage of the new version is the relative ease of coding in C/C++ compared with FORTRAN and the built-in memory management. While the FORTRAN code relies on a set of limits to the size of matrices, the C/C++ code uses dynamic memory management.[12]

The C/C++ version, in addition, has new features including additional distributions and new (and aforementioned) output options that facilitate embedding the utility in a workflow. The specific new features are:

- Ten new distributions: Cauchy, Dagum, Fréchet, Gompertz, Gumbel, Laplace, Lévy, Logistic, Kumaraswamy and Rayleigh[13]
- Four additional output options: CSV, XML, GMS and GDX[14]

[8] The new software is fully backward compatible and the default output format is still supported.

[9] The download includes the full suite of software in Sandia National Lab's Dakota package. The LHS code is located in dakota-6.16.0-public-src-cli\packages\external\LHS. Interested users can contact the author for a copy of the FORTRAN code, which includes a few additional features and a *Makefile* to compile with GFortran.

[10] The new Mersenne Twister algorithm for generating random numbers requires a compiler that recognizes non-standard integer types, notably `unsigned long long`.

[11] The largest differences, and they are negligible, relate to deviates for the Beta distribution. The C/C++ code uses a different set of routines for the latter.

[12] The original FORTRAN code in fact crashed when attempting to generate the LHS sample for the IAM model. It had a relatively low limit on the number of potential user-based correlations.

[13] The latest version of the FORTRAN code includes the Fréchet and Gumbel distributions.

[14] We also provide GEMPACK code that automates conversion of the CSV output into a GEMPACK HAR file.

	Average seconds per sampling		Index relative to C/C++ w/ output	
	w/ output	w/o output	w/ output	w/o output
C/C++	8.4	6.0	1.0	0.7
FORTRAN	24.7	21.8	2.9	2.6

Table 1: Benchmarking of the LHS software

- A new optional random number generator based on the Mersenne Twister algorithm described in Matsumoto and Nishimura (1998) and Nishimura (2000)[15]

Most of the features of the LHS utility are described in the accompanying User Guide. Users who are interested in the full features of the LHS utility (apart from the new ones) are referred to the original documentation in Swiler and Wyss (2004).

Benchmarking suggests that the C/C++ code is significantly faster than the FORTRAN code. Both are compiled with the Mingw compilers, so we would not expect to see such large differences. The benchmark involves sampling for the Meta21 simulation described in the next section. The Meta21 sample is large and with a significant number of correlated distributions.[16] The sampling is run 100 times for both the C/C++ and FORTRAN code. It is also run with and without saving the sample. The output file size is over 150MB, so it is possible that C/C++ is more efficient in writing than FORTRAN. The benchmarking results are summarized in Table 1. The FORTRAN version is around 2.9 times slower than the C/C++ code, and the speed difference is not linked to the writing of the output sample.

3 The Integrated Assessment Model

Integrated assessment models (IAMs) of climate change integrate economic models that track GHG emissions with bio-physical models that at a minimum generate changes in the global mean temperature, which subsequently impact the economy via changes in economic potential such as crop yields, infrastructure, labor productivity, etc. Figure 3 provides a schema for the typical IAM model, which has four components: (1) the socio-economic module, which could be a CGE model; (2) which generates emissions; (3) that impact atmospheric chemistry and the global energy balance, i.e., temperature; and (4) that impacts economic potential. In the context of ongoing research under the aegis of the Intergovernmental Panel on Climate Change (IPCC), most IAMs are using drivers and storylines called the shared socio-economic pathways (SSPs)—a set of five distinct storylines with variegated pathways for GDP and population growth. The SSPs are coupled with the so-called Representative Concentration Pathways (RCPs)—that represent different pathways for climate, with changes in the global mean temperature as one of the indicators. The system will also be influenced by policies—mitigation polices that will change the emission pathways, and/or adaptation policies that will modify climate-induced impacts.

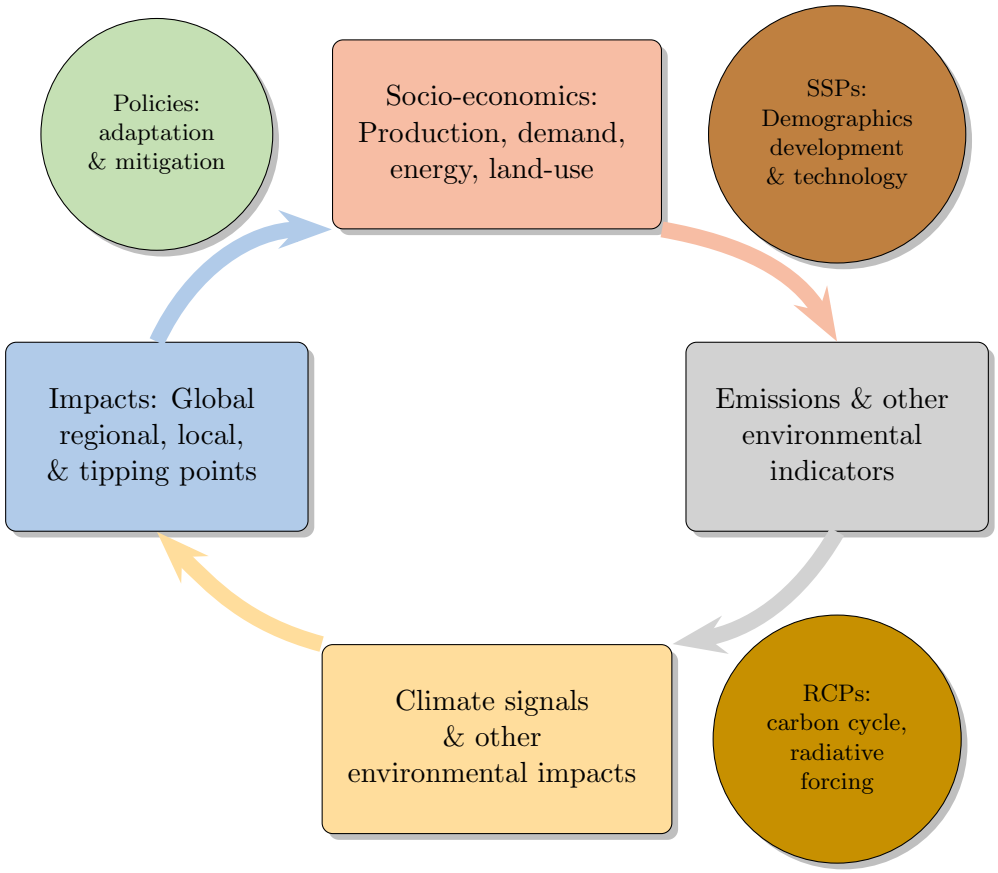
This section will describe one such IAM, the Meta21 model Dietz et al. (2021). The Meta21 model is in the same spirit as Nordhaus' DICE/RICE suite of models (Nordhaus (2017)), Hope's PAGE model (Hope and Schaefer (2016)) and the FUND model (Anthoff and Tol (2014)). Like these models, Meta21 links emissions to the carbon cycle, to forcing and temperature change

[15] The version of this algorithm in the LHS code is explicitly based on a 64-bit architecture, which may limit its portability.

[16] The full sample is a matrix of dimension 10000×582 and a correlation matrix of dimension 582×582. The benchmark was done under Windows 10 under normal operations (on a non-synced drive) using an Intel i9 12th generation CPU and 64 GB of RAM.

and then to economic damages. The economic and emissions module of the Meta21 model are highly simplified and are largely exogenous in fact. GDP is driven by exogenous assumptions on total factor productivity (TFP) and savings and is only impacted by climate-induced changes in TFP. Similarly, emissions are largely exogenous—not even responding to changes in GDP—though impacted by some of the modeled tipping points such as the Amazon dieback. The key focus of this section is on the climate, tipping points and impact modules of the Meta21 model, which can be readily incorporated in more complex economic models such as a global CGE model.

Figure 3: Integrated Assessment



This section describes a significant portion of the Meta21 model, though leaves out some of the more complicated modules. The full description of the model and its parameterization are described in the Dietz et al. (2021) supplementary materials. The focus of Meta21 is on tipping points, which include the following:

- The thawing of permafrost, which leads to additional carbon and methane emissions, labeled *PCF*
- The dissolution of methane hydrates which leads to additional methane emissions, labeled *OMH*
- The Amazon dieback, which leads to a jump in carbon emissions, labeled *AMAZ*
- Ice sheet melting, which impacts sea-level rise. There are four sources: thermal expansion, small ice-sheets and glaciers, the Greenland ice-sheet (GIS) and the West Antarctic ice-sheet (WAIS)

- Changes in the Atlantic Meridional Overturning Circulation (AMOC), also called the thermohaline circulation. AMOC ends up impacting the change in country-level average temperatures

- The weakening of the Indian summer monsoon (ISM), which leads to an additional damage effect for India

In addition to these tipping points, the Meta21 model also captures changes in the planet's albedo from the change of ice cover—leading to adjustments in the change in the global mean surface temperature. A final important feature of the model is that it allows for interaction effects across the tipping points.

The model captures two sources of uncertainty. The first is captured by the uncertainty in the model's parameters, for example the climate sensitivity parameter. Meta21 captures this uncertainty with an explicit description of each parameter's probability distribution function (PDF), and in some cases also describes the correlation across parameter estimates, for example between equilibrium climate sensitivity and transitory climate sensitivity. This uncertainty is modeled using standard Monte Carlo techniques with random draws, albeit using the LHS methodology. The second type of uncertainty deals with the start of the tipping point. In each future year, there is a non-zero probability that a tipping point occurs. The probability itself increases with temperature. The event's start is sampled using the Binomial distribution with a single draw and the temperature-sensitive probability. Once an event starts, it is irreversible.

The original Meta21 model is 'coded' in Excel and is linked to the external @Risk package for the Monte Carlo draws. The Excel file contains the distribution information for the uncertain parameters, including their associated correlation assumptions. The draws for the start of events occur independently.[17]

The version of the Meta21 model implemented herein only uses part of the full Meta21 model to illustrate the coupling with the LHS package. This version excludes the more complex components of the Meta21 model, including the weakening of the Indian summer monsoon, and the impact of a changing albedo on temperature changes and the interaction effects.[18]

The model will be described in its recursive dynamic structure—starting with emissions, the carbon cycle, radiative forcing, the energy balance model i.e., temperature, the tipping points, damages and the economic module. This section will conclude with a brief overview of the parameterization assumptions.

3.1 The Emission/Climate Module

Carbon emissions

Global carbon emissions in any given year are composed of four items in the Meta21 model:

- Exogenous carbon emissions that are linked to one of the specific representative concentration pathways (RCP, Emi^r)

- Carbon emissions linked to permafrost processes (PCF)

[17] In principle, one could forego @Risk and use LHS to generate the sample of random deviates and link the relevant cells in the spreadsheet to the LHS-generated sample. The start of a random event can be generated with the Excel function `BINOM.INV(1, A2, RAND())`, where A2 represents the relevant probability. In essence this describes a Bernouilli trial with an outcome of 0 or 1. A macro would be needed to automate the running of the Monte Carlo simulations and save the desired results.

[18] We are also leaving out the assessment of non-economic damages on the social cost of carbon.

- Carbon emissions linked to Amazon dieback ($AMAZ$)
- A shock to emissions that sets emissions on a different path from the RCP pathway—for example a pulse of emissions in a specific year (Emi^x). The latter is used to calculate the social cost of carbon (SCC).

$$Emi_{CO2,t} = Emi^r_{CO2,t} + PCF_{CO2,t-1} + AMAZ_{CO2,t-1} + Emi^x_{CO2,t} \tag{1}$$

Carbon emissions are measured in $GtCO_2$. The impacts of perma-frost and the Amazon dieback are lagged (for recursivity purposes) and are described below.

Carbon cycle

The carbon cycle is an adaptation of the FaIR 2.0 model (Leach et al. (2021)). The FaIR model includes four carbon sinks or pools that include geological processes, deep ocean invasion/equilibration, biospheric uptake/ocean thermocline invasion and rapid biospheric uptake/ocean mixed-layer invasion. Though the labels for the pools are suggestive, the pools and their parameterization are mathematical constructs that summarize a complex set of interactions that include chemical reactions in the atmosphere and oceans and the fluxes across a number of carbon sinks that include vegetation and soils. Thus each pool represents a specific impact on part of the atmospheric concentration and the aggregate atmospheric concentration is the sum across all of the pools. The carbon stock in each 'pool', indexed by cb, is equal to the previous period's concentration adjusted by some rate of decay, δ^c, plus some fraction, α^e, of current CO_2 emissions, equation (2).[19]

One of the innovations of the FaIR model is to make the decay rate a function of the level of concentration—as the absorption rate in each 'pool' depends on the level of carbon saturation. The coefficient α^c, determined below, is the state-dependent variable that accounts for adjustments in the decay rate. Note that the adjustment is uniform across all pools. Equation (3) then determines the total stock of carbon in the atmosphere (in $GtCO_2$), including the pre-industrial stock converted from parts per million to $GtCO_2$, where χ^{ppm}_{co2} is the conversion factor.[20] Equation (4) converts the atmospheric stock of carbon into parts per million.

$$\begin{aligned} R_{cb,t} &= \alpha^e_{cb} Emi_{CO2,t} + \left(1 - \frac{\delta^c_{cb}}{\alpha^c_{t-1}}\right) R_{cb,t-1} \\ &= \alpha^e_{cb} Emi_{CO2,t} + \left(1 - \frac{1}{\alpha^c_{t-1}\tau^c_{cb}}\right) R_{cb,t-1} \end{aligned} \tag{2}$$

$$StockGt_t = \sum_{cb} R_{cb,t} + \chi^{ppm}_{co2} Conc^{pre}_{co2,t} \tag{3}$$

$$Conc_{co2,t} = StockGt_t / \chi^{ppm}_{co2} \tag{4}$$

[19] In the original FaIR documentation, the motion equation for the pools is given by:

$$\frac{dR}{dt} = \alpha^e Emi_t - \frac{1}{\alpha^c \tau^c} R_t$$

One discrete approximation of this equation is therefore:

$$R_t = \alpha^e Emi_t + \left(1 - \frac{1}{\alpha^c \tau^c}\right) R_{t-1}$$

[20] $7.81 = 2.13 \times 44/12$

Equation (5) determines cumulative anthropogenic CO_2 emissions from the beginning of time. It is initialized at 521.45 gtC in 2009. Equation (6) determines the carbon stock levels in all non-atmospheric carbon pools (soils, oceans, etc.) derived from anthropogenic carbon emissions, i.e. it is total cumulative emissions less the contemporaneous level in the atmosphere. This is measured in gtC.

$$CEmi_{co2,t} = CEmi_{co2,t-1} + \chi_C^{co2} Emi_{CO2,t} \tag{5}$$

$$CStock^x_{co2,t} = CEmi_{co2,t} - \chi_C^{co2} \sum_{cb} R_{cb,t} \tag{6}$$

The original version of the model calibrated the α^c parameter to two long-term conditions relating to a variable linked to the integrated impulse response function (`iIRF100`), which represents the 100-year average airborne fraction of a pulse of CO_2 (Joos et al. (2013)).

The first of these conditions, derived from a physical relation, is the following:

$$iIRF100_t = \sum_{cb} \alpha^e_{cb} \alpha^c \tau_{cb} \left[1 - \exp\left(-\frac{100}{\alpha^c \tau_{cb}}\right)\right]$$

The second condition, derived from a reduced form statistical relation, is:

$$iIRF100_t = r^0 + r^C CStock^x_{co2,t} + r^T Temp^{atm}_t \tag{7}$$

which links *iIRF100* to the carbon stock level and temperature. The two equations together determine *iIRF100* and α^c, though α^c is only determined implicitly. In the supplementary materials of Dietz et al. (2021) and in Leach et al. (2021), it is posited that α^c is approximately exponential in *iIRF100* and can be expressed as:

$$\alpha^c_t = \alpha^1 \exp\left(\alpha^2 iIRF100_t\right) \tag{8}$$

where α^1 and α^2 are calibrated parameters.

Methane emissions and concentration

Methane emissions, similar to carbon emissions, are on an RCP-based reference trajectory, which in the Meta21 model are adjusted by (lagged) methane emissions linked to permafrost processes and to ocean methane hydrates (OMH). Atmospheric concentration of methane is based on a single-box model with a decay rate given by δ^{ch4}. Emissions are converted from metric tons to parts per billion (ppb) using the conversion factor $\chi^{ppbch4}_{mt} = 2.78$.

$$Emi_{ch4,t} = Emi^r_{ch4,t-1} + PCF_{ch4,t-1} + OMH_{t-1} \tag{9}$$

$$Conc_{ch4t} = Conc^{pre}_{ch4} + \left(1 - \delta^{ch4}\right)\left(Conc_{ch4,t-1} - Conc^{pre}_{ch4}\right) + Emi_{ch4,t}/\chi^{ppbch4}_{mt} \tag{10}$$

Forcing

Radiative forcing is the sum of impacts from a number of drivers. In the case of the Meta21 model, there are three drivers—two of which are modeled explicitly: carbon and methane atmospheric concentration— and the third is exogenous, though dependent on the selected RCP.

Carbon

Carbon concentration is assumed to have a logarithmic impact on radiative forcing with no interactions with other forcings. The key parameter is *f2xco2*, which represents the radiative forcing resulting from a doubling of carbon concentration from its pre-industrial level. It is treated as a random variable with a mean value of 3.45. Further discussion of the sampling strategy will be reserved until later as it is linked to the parameterization of the energy balance module (EBM).

$$Forc_{co2,t} = f2xco2 \times \log\left(\frac{Conc_{co2,t}}{Conc_{co2}^{pre}}\right)/\log(2) \tag{11}$$

Methane

The radiative impact of methane concentration depends on the square root of the current concentration, relative to the square root of the pre-industrial methane concentration. The key parameter is α^{fch4}. There is nonetheless an interaction effect between methane concentration and the initial concentration of nitrous oxide (N_2O). The interaction expression is given in equation (13).

$$\begin{aligned} Forc_{ch4,t} &= \alpha^{fch4}\left[\sqrt{Conc_{ch4,t}} - \sqrt{Conc_{ch4}^{pre}}\right] \\ &- \left[f\left(Conc_{ch4,t}, Conc_{n2o,t0}\right) - f\left(Conc_{ch4,t0}, Conc_{n2o,t0}\right)\right] \end{aligned} \tag{12}$$

where the methane/nitrous oxide interaction relationship is given by:

$$\begin{aligned} f\left(Conc_{ch4,t}, Conc_{n2o,t}\right) &= 0.47\log\Bigg(1 + 2.01E^{-5}\left(Conc_{ch4,t}Conc_{n2o,t}\right)^{0.75} \\ &+ 5.31E^{-15}\left(Conc_{ch4,t}Conc_{n2o,t}\right)^{1.52}\Bigg) \end{aligned} \tag{13}$$

Total radiative forcing

Exogenous forcing is provided in the RCP database for four RCPs.[21] In this case we write the total radiative forcing (RF) equation, as the sum of radiative forcing over the three drivers—carbon, methane and the exogenous shifter.

$$RF_t = \sum_{em} Forc_{em,t} \tag{14}$$

The Energy Balance Module

The energy balance model from DICE, FaIR and others can take the same form:

$$T_t = [1 + A]\, T_{t-1} + M\, RF_t \tag{15}$$

where T represents the (change) in temperature in time t and is defined over two or more temperature boxes.[22] In the case of DICE and FaIR 1.5 there are two boxes, essentially the atmosphere (and the surface and shallow ocean) and the (deep) ocean. In FaIR 2.0, the ocean box is split into shallow

[21] In the GAMS code, it is attached to the emission labeled `FGAS` and is fixed.

[22] It is a discrete approximation to the state-space form, which is $\dot{T} = A\,T + M\,RF$.

and deep. The matrix A represents the conveyance of energy across the boxes and the matrix M represents the impact of the contemporaneous radiative forcing on atmospheric temperature. In the FaIR 2.0 model, the matrix A takes the following form:

$$A = \begin{pmatrix} -(\lambda + \kappa_2)/C_1 & \kappa_2/C_1 & 0 \\ \kappa_2/C_2 & -(\kappa_2 + \varepsilon\kappa_3)/C_2 & \varepsilon\kappa_3/C_2 \\ 0 & \kappa_3/C_3 & -\kappa_3/C_3 \end{pmatrix}$$

where the parameter C is the heat capacity of each box, the κ parameters are the heat exchange coefficients, λ is a feedback parameter and ε allows to account for transient warming. M is a 3×1 vector with $1/C_1$ in the first cell.[23]

The Meta21 model uses a two-box model, the parameterization of which will be described below. The change in temperature derived from equation (15) is further adjusted by a factor influenced by the changing albedo. The albedo adjustment has been dropped from this version.

Country temperatures

The change in global mean surface temperature is downscaled to the country level. Equation (16) determines the ratio of national mean surface temperature with respect to global mean surface temperature. The term in the logarithm represents the change in the global mean surface temperature from it's pre-industrial level. The first term is the change from pre-industrial level to the level in 2010,[24] and the second term represents the incremental change in temperature in time t relative to 2010. The α^g and β^g parameters are estimated using OLS by Dietz et al. (2021).[25] The temperature level in each country, TL, is then equal to the global temperature adjusted by the downscaling factor μ^g. In addition, each country's temperature is impacted by the the slowdown of the Atlantic Meridional Overturning Circulation, T_c^h, which is described below.

$$\mu_{c,t}^g = \alpha_c^g + \beta_c^g \ln\left(\left(TL_{2010}^{atm} - 18.825\right) + \left(T_t^{atm} - T_{t_0}^{atm}\right)\right) \tag{16}$$

$$TL_{c,t} = \mu_{c,t}^g \left(TL_{2010}^{atm} + T_t^{atm} - T_{t_0}^{atm}\right) + T_{c,t}^h \tag{17}$$

Summary

This concludes the emissions/temperature module. The sequence as described above is recursive given the implementation of the lag structure, with the exception of the α^c adjustment used to adjust the decay parameters in the carbon boxes. The formulas for *iIRF100* and α^c should come after updating the temperature equation.

3.2 The Impact/Tipping points Module

Changes in the Atlantic Meridional Overturning Circulation (AMOC)

The timing of the tipping point for the slowdown of the Atlantic Meridional Overturning Circulation is based on a probabilistic specification that is a function of the temperature anomaly. The hazard rate is given by b^{AMOC}.[26] The impact from the AMOC slowdown is felt differentially across

[23] The DICE parameterization can be converted to this matrix notation by setting κ_2 to `C3` (in the DICE code), λ is equal to `fco22x/t2xo2`, C_1 is equal to `1/C1`, C_2 is equal to `C3/C4` and κ_3 is 0.

[24] The notation TL is used to designate the temperature level, whereas the notation T refers to the temperature anomaly.

[25] The downscaling parameters α^g and β^g are fixed.

[26] In the full model the probability may be impacted by interaction effects, ψ_{amoc}^x.

countries. This has been estimated using four general circulation models (GCMs) for a given impulse of fresh water from higher precipitation and ice melt from Greenland. The results of these GCM runs is a permanent change in the long-run temperature across countries, $T^h_{c,max}$. The long-run temperature change occurs over time, with an annual increment given by Δ^{AMOC}. Equation (20) thus provides the country-specific temperature anomaly due to the AMOC slowdown, $T^h_{c,t}$. This anomaly is added to the country temperature anomaly due to the change in mean global temperature, see equation (17) above.

$$p_t^{AMOC} = \min\left(\psi^x_{amoc,t}\left(1 - \exp\left(-b^{AMOC}\right)\right), 1\right) \tag{18}$$

$$I_t^{AMOC} = \begin{cases} 1 & \text{if } I_{t-1}^{AMOC} = 1 \\ \texttt{randBinomial}(1, p_t^{AMOC}) & \text{if } I_{t-1}^{AMOC} = 0 \end{cases} \tag{19}$$

$$T^h_{c,t} = T^h_{c,t-1} + \begin{cases} 0 & |T^h_{c,t-1}| \geq |T^h_{c,max}| \\ \frac{T^h_{c,max}}{\Delta^{AMOC}} & |T^h_{c,t-1}| < |T^h_{c,max}| \end{cases} \tag{20}$$

Ocean methane hydrates (OMH)

The probability of the beginning of the release of methane from ocean methane hydrates is given in equation (21), where b_{OMH} is the so-called hazard rate. Given the probability, in each year t, a single random draw from the binomial distribution is taken with the given probability. The indicator function takes the value of 1 for every year subsequent to a positive draw, i.e. the release of the methane hydrate starts in the first year of the positive draw, and is assumed to be irreversible.

Figure (4) depicts the probability density for a sample of 100,000 random draws for the range of temperature change from 0 to 3°C. The red line shows the level of the probability of the methane release starting for each change in temperature—it starts at 0 and rises to over 30% when the temperature change is 3°C. The mode of the PDF is at around 0.8-0.9°C. Given that the current estimated level of temperature change is already over 1°C, this parameterization implies a relatively early start to the release of the methane hydrates.

$$p_t^{OMH} = 1 - \exp\left(-b^{OMH} T_t^{atm}\right) \tag{21}$$

$$I_t^{OMH} = \begin{cases} 1 & \text{if } I_{t-1}^{OMH} = 1 \\ \texttt{randBinomial}(1, p_t^{OMH}) & \text{if } I_{t-1}^{OMH} = 0 \end{cases} \tag{22}$$

Once the emissions commence, a constant level of emissions is released each year until the given stock of sequestered methane runs out, equation (23), where $CH4^{OMH}_{Max}$ is the initial stock of sequestered methane and Δ^{OMH} is the annual release.

$$OMH_t = CH4^{OMH}_{Max}/\Delta^{OMH} \quad \text{if } I_t^{OMH} = 1 \text{ and } \sum_{t'=t_0}^{t-1} OMH_t < CH4^{OMH}_{Max} \tag{23}$$

Permafrost carbon feedback (PCF)

The specification of the permafrost module derives from Kessler (2017). It is based on a two-stage process. In the first stage, near-surface permafrost thaw is a linear function of temperature change relative to the reference year (i.e. 2010). PF^{ext} in equation (24) is the area of permafrost remaining

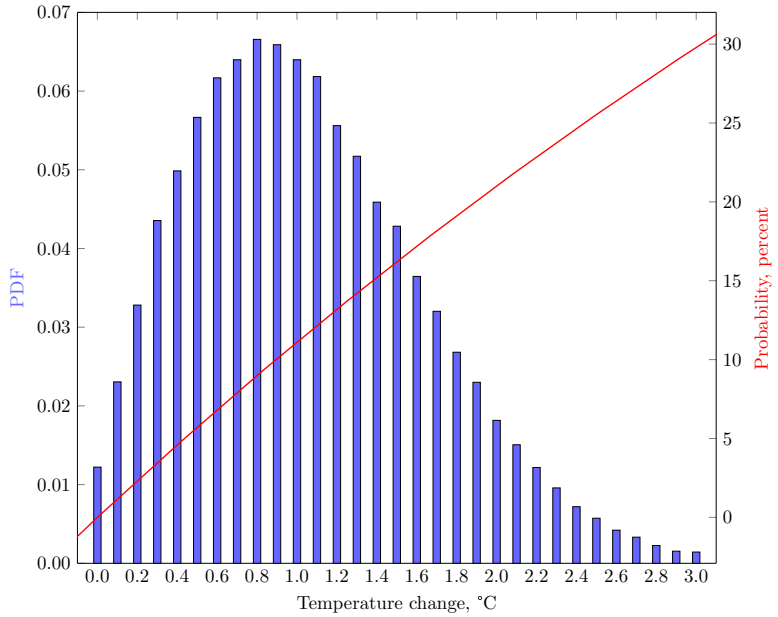


Figure 4: Sampling for OMH, where $b^{omh} = 0.118$

at time t, and β^{pf} is the sensitivity of permafrost thaw with respective to temperature change. Note that $PF_t^{ext} = PF_t^{area}/PF_{t0}^{area}$. The amount of carbon in the freshly thawed permafrost is the product of the total stock of carbon locked in the near-surface northern circumpolar permafrost region, C^{pf}, and the freshly thawed permafrost area, equation (25).

$$PF_t^{ext} = 1 - \beta^{pf}\left(T_t^{atm} - T_{t0}^{atm}\right) \tag{24}$$

$$CThaw_t^{pf} = -C^{pf}\left(PF_t^{ext} - PF_{t-1}^{ext}\right) \tag{25}$$

The stock of carbon is divided into two reservoirs: one is passive that releases no carbon, and another that is active that decomposes exponentially. $CumCThaw^{pf}$ in equation (26) is the cumulative release of emissions in year t from the active reservoir, where χ^{psv} is the passive proportion of the total reservoir and τ^{pf} is the rate of decomposition of the active reservoir. The decomposed emissions generate carbon and methane emissions in fixed proportion. Equation (27) represents the level of both emissions, where φ^{ch4} is the proportion of methane emissions. Permafrost carbon emissions are measured in gtC and are thus converted to $gtCO_2$, the same units as carbon emissions. Methane emissions are converted from gtC to mt NH_4, where the conversion factor is $1000 \times 16/12$.[27]

$$CumCThaw_t^{pf} = (1 - \chi^{psv}) \sum_{t'=t0}^{t} CThaw_{t'}^{pf}\left(1 - e^{(t'-t)/\tau^{pf}}\right) \tag{26}$$

$$PCF_{em,t} = \begin{cases} \left(1 - \varphi^{ch4}\right)\left(CumCThaw_t^{pf} - CumCThaw_{t-1}^{pf}\right)/\chi_C^{co2} & \text{if } em = co2 \\ \varphi^{ch4}\left(CumCThaw_t^{pf} - CumCThaw_{t-1}^{pf}\right)\chi_{mtNH4}^{gtC} & \text{if } em = ch4 \end{cases} \tag{27}$$

[27] The atomic weight of carbon is 12 and that of methane is 16.

The Amazon rainforest dieback module (AMAZ)

The Amazon rainforest dieback module is implemented similar to methane hydrates. There is a given quantity of carbon sequestered in the forest. When the dieback 'starts', the quantity is released into the atmosphere at a constant level each year until it is exhausted.[28] The probability of the start of the dieback increases with temperature, but the year the dieback starts is a random event based on the probability. The coefficient ψ^x is an interaction effect between the Amazon dieback and other tipping points, which is described further below.

$$p_t^{AMAZ} = \min\left(\psi^x_{amaz,t}\left[1-\exp\left(-b^{AMAZ}\max\left(0,T_t^{atm}-1\right)\right)\right],1\right) \tag{28}$$

$$I_t^{AMAZ} = \begin{cases} 1 & \text{if } I_{t-1}^{AMAZ}=1 \\ \texttt{randBinomial}(1,p_t^{AMAZ}) & \text{if } I_{t-1}^{AMAZ}=0 \end{cases} \tag{29}$$

$$AMAZ_t = CO2_{Max}^{AMAZ}/NYear^{AMAZ} \text{ if } I_t^{AMAZ}=1 \& \sum_{t'=t_0}^{t-1} AMAZ_{t'} < CO2_{Max}^{AMAZ} \tag{30}$$

The Greenland ice sheet module (GIS)

The specification of the disintegration of the Greenland Ice Sheet is derived from Kessler (2017) and Nordhaus (2019). Equation (31) posits a relation between the volume of GIS and an equilibrium temperature anomaly, T_t^{eGIS}. The GIS volume is measured as an index between 0 and 1, where the volume is at an initial (and maximum) level at 1, and fully melted at 0. When the equilibrium temperature anomaly is 0 the volume is maximized, i.e. V^{GIS} is equal to 1. On the other hand, when the volume is fully melted, the equilibrium temperature anomaly is equal to T^{max}, which Nordhaus has estimated to be 3.4°C. The default assumption is that this relation is linear and thus η^{eGIS} equals 1. The rate of melting, equation (32), depends on the difference between the actual and equilibrium temperature anomalies and the GIS volume. The default value for β^{GIS} is -0.0000106 and $\eta^{GIS}=0.2$. Equation (33) updates the GIS volume, only if the melt rate is negative, and, in which case, an interaction effect is taken into account. Equation (34) posits a linear relation between the GIS volume and sea-level rise, where the default value for SLR_{max}^{GIS} is 7 (m), i.e. when fully melted, sea-level rise would be 7 meters.

$$T_t^{eGIS} = T^{max}\left(1-V_t^{GIS}\right)^{\eta^{eGIS}} \tag{31}$$

$$\Delta V_t^{GIS} = \beta^{GIS}\texttt{sign}\left(T_{t-1}^{atm}-T_{t-1}^{eGIS}\right)\left(T_{t-1}^{atm}-T_{t-1}^{eGIS}\right)^2\left(V_{t-1}^{GIS}\right)^{\eta^{GIS}} \tag{32}$$

$$V_t^{GIS} = V_{t-1}^{GIS} + \begin{cases} \psi^x_{gis,t}\Delta V_t^{GIS} & \text{if } \Delta V_t^{GIS}<0 \\ 0 & \text{if } \Delta V_t^{GIS}\geq 0 \end{cases} \tag{33}$$

$$SLR_t^{GIS} = SLR_{max}^{GIS}\left(1-V_t^{GIS}\right) \tag{34}$$

[28] The factor $NYear^{AMAZ}$ represents the time it will take to release the carbon in years—unlike the methane hydrate factor Δ^{OMH}, which represents the amount released each year.

The West Arctic ice sheet module (WAIS)

Dietz et al. (2021) cite Diaz and Keller (2016) for the specification of the disintegration of the West Arctic ice sheet (WAIS). Similar to the methane hydrates and the Amazon dieback, the trigger for the start of the disintegration of the West Arctic ice sheet is based on a random start year that is a function of a probability. Once the disintegration starts, the annual sea-level rise generated by WAIS is a constant amount, R^{WAIS}.

$$p_t^{WAIS} = \min\left(b^{WAIS}\psi_{wais,t}^{x}\left(T_t^{atm} - 0.6\right)^2, 1\right) \tag{35}$$

$$I_t^{WAIS} = \begin{cases} 1 & \text{if } I_{t-1}^{WAIS} = 1 \\ \texttt{randBinomial}(1, p_t^{WAIS}) & \text{if } I_{t-1}^{WAIS} = 0 \end{cases} \tag{36}$$

$$\Delta SLR_t^{WAIS} = R^{WAIS} I_t^{WAIS} \tag{37}$$

$$SLR_t^{WAIS} = \sum_{t'=t_0}^{t} \Delta SLR_{t'}^{WAIS} \tag{38}$$

The sea-level rise module (SLR)

Sea-level rise is generated by four sources: the aforementioned melting of the Greenland and West Arctic ice sheets, the melting of glaciers and small icecaps, and thermal expansion. The latter two have an annual melt that is linear in the temperature anomaly. The variable SLR^{THX} represents the aggregate sea level rise from the latter two sources. Total sea-level rise, SLR^{TOT}, is the sum across all sources.

$$SLR_t^{THX} = SLR_{t-1}^{THX} + \left(\rho^{THX} + \rho^{GSIC}\right) T_t^{atm} \tag{39}$$

$$SLR_t^{TOT} = SLR_t^{THX} + SLR_t^{GIS} + SLR_t^{WAIS} \tag{40}$$

3.3 The Economic Module

Damages

In this version of the Meta21 Model there are two sources of damage—change in temperature and sea-level rise.[29]

Temperature damage is a quadratic expression in the change in (country) temperature since 1990. The source of the parameter values is Burke et al. (2015). Dietz et al. (2021) describe their transformation of the Burke et al. (2015) estimates. The study further provides an estimate of the mean and standard deviation for the β coefficients. Sampling of the β parameters is undertaken with a supplied correlation matrix across countries and between the two β parameters. There are over 75,000 coefficients in the correlation matrix.

$$Dam_{c,t}^{Temp} = \begin{cases} 0 & \text{if } ifDamage = \texttt{No} \\ \beta_c^1\left(TL_{c,t} - TL_{c,1990}\right) + \beta_c^2\left(TL_{c,t} - TL_{c,1990}\right)^2 & \text{if } ifDamage = \texttt{Yes} \end{cases} \tag{41}$$

[29] In the full model, there are also additional damages in India derived from the changing of the summer monsoon.

Sea-level rise damage is a linear function of sea-level rise. The sampling strategy is to assume a Triangular distribution of damages based on simulations with the CIAM model (Diaz (2016)).

$$Dam_{c,t}^{SLR} = \begin{cases} 0 & \text{if } ifDamage = \texttt{No} \\ \chi_c^{slr} SLR_t^{TOT} & \text{if } ifDamage = \texttt{Yes} \end{cases} \tag{42}$$

GDP and consumption

Pre-damage GDP per capita is described in equation (43). It is equal to the previous period's GDP augmented by an exogenous growth rate, γ^x. The exogenous growth rate comes from one of four SSPs.[30] The use of the SSPs and their extension beyond 2100 is described in the supplemental materials of (Dietz et al. (2021)). The equation allows for two cases and a convex combination of the two. In one case, when φ is set to 1, the damages only impact GDP levels as there is no memory in the system relative to damages in the previous period. In other words, pre-damage GDP per capita is equal to the exogenous GDP per capita as given by the SSPs. In the other case, the growth is applied to actual GDP in the previous period, in which case damages influence the growth of GDP, not just the level. The default level for the weight, φ is 0.5, it is sampled using a Triangular distribution bounded by 0 and 1 with a mode of 0.5.

$$y_{c,t}^g = \left(1 + \gamma_{c,t}^x\right) \left[\varphi y_{c,t-1}^x + (1-\varphi) y_{c,t-1}\right] \tag{43}$$

Post-damage GDP per capita is described in equation (44), which reflects the compound damage from temperature and sea-level rise. The damage from temperature is adjusted for the growth rate, as it is assumed to shift the growth rate additively. Consumption per capita is equal to GDP per capita less savings. The rate of savings is calibrated to 2010 levels derived from the World Bank's WDI databank—and are fixed over time at the reference level.

$$y_{c,t} = y_{c,t}^g \left(1 - \frac{Dam_{c,t}^{Temp}}{1 + \gamma_{c,t}^x}\right) \left(1 - Dam_{c,t}^{SLR}\right) \left(1 - Dam_{c,t}^{ISM}\right) \tag{44}$$

$$c_{c,t} = (1 - s_c)\, y_{c,t} \tag{45}$$

Utility (on a per capita basis), is given in equation (46), which uses a constant-elasticity-of-substitution expression and where η represents the elasticity of the marginal utility of consumption.[31] The parameter η is sampled from a Triangular distribution bounded by 0.5 and 2 with a mode of 1.5.

$$u_{c,t} = \frac{1}{1-\eta} \left(c_{c,t}\right)^{1/(1-\eta)} \tag{46}$$

Aggregations and the social cost of carbon

The country results can be aggregated to any set of regional aggregations, defined by the set R. Set R must at least contain the world as one aggregate region designated by the label `WLD`.

[30] The Shared Socio-Economic Pathways (SSPs) have been developed by the Integrated Assessment Modeling Community (IAMC) for the purposes of assessing the economics of climate change through 2100. Five SSPs were developed by three members of the IAMC—four of the SSPs are available as part of the Meta21 distribution. See for example Dellink et al. (2017), for a description of the SSPs developed at the OECD.

[31] In the full model, consumption is also adjusted by a factor that reflects the economic value of non-market damages, based on a specification attributed to Manne and Richels (2005).

Equations (47) and (48) define respectively aggregation population and aggregate consumption per capita. Equation (49) defines discounted aggregate utility, where the discounting is relative to a reference year, t_r, which is currently set at 2020. The parameter ρ is the pure rate of time preference (PRTP). The Excel file refers to EV as equivalent regional consumption per capita. Global welfare is defined in equation (51). The social cost of carbon, defined as $SCC_t = \frac{\partial W/\partial E_t}{\partial W/\partial c_t}$, is approximated by injecting 10 gt of CO_2 emissions in 2020. It is measured as the change in global welfare relative to the reference year global per capita consumption. PRTP is sampled from a Triangular distribution bounded by 0.01% and 2% with a mode of 1%.

$$Pop^a_{R,t} = \sum_{c \in R} Pop_{c,t} \tag{47}$$

$$c^a_{R,t} = \sum_{c \in R} Pop_{c,t} c_{c,t} / Pop^a_{R,t} \tag{48}$$

$$U^a_{R,t} = \sum_{c \in R} u_{c,t} Pop_{c,t} \left(1+\rho\right)^{-(t-t_r)} \tag{49}$$

$$EV^a_{R,t} = \left(\frac{(1-\eta) U^a_{R,t}}{Pop^a_{R,t}} \right)^{1/(1-\eta)} \tag{50}$$

$$W^g = \sum_{t \geq t_r} U^a_{\texttt{WLD},t} \tag{51}$$

$$SCC = -\frac{\left(W^g_x - W^g_r\right) / Emi^x_{co2,t_t}}{\left(c^a_{\texttt{WLD},t_r}\right)^{-\eta}} \tag{52}$$

3.4 Parameterization

This section will mostly focus on the parameters subject to the LHS sampling. Further details on the parameterization can be found in the original Meta21 Excel workbook and the supplemental materials of Dietz et al. (2021).

Carbon cycle

Table 2 provides the main sampling assumptions for the carbon cycle parameters. The column labeled 'X' represents the parameter as used in this document, and the column labeled 'Code' represents the code used for the LHS utility (and corresponds to the naming convention of the Meta21 Excel file). The middle panel shows the expected mean and standard deviation for each of the uncertain parameters based on the input parameters, and the right panel shows the actual mean and standard deviation from the sample.[32] In most cases, the differences are small—less than 0.1% in absolute terms. The exceptions are the log-normal distributions where there can be more significant deviations in the standard deviation.

Though the LHS sample is quite good at representing the targeted marginal distribution of each of the individual parameters of the carbon cycle, it does not perform well in capturing the desired correlation across parameters. Tables 3 and 4 provide the targeted and sample correlations, respectively, for the carbon cycle parameters. The most severe example of the deviation between the

[32] The actual input parameters for the LHS utility can be seen in the `Meta21.in` file, which is the input file for the LHS utility.

Table 2: Sampling assumptions for carbon cycle parameters

			Expected		Sample	
X	Code	Distribution	mu	σ	mu	σ
α^e_{cb1}	alpha0	LogNormal	0.2349	0.1364	0.2349	0.1362
α^e_{cb2}	alpha1	Logistic	0.224	6.16E-05	0.224	6.17E-05
α^e_{cb4}	alpha3	Levy	∞	∞	0.298	1.95
δ^c_{cb2}	rho1	LogNormal	0.0036	0.0121	0.0036	0.0106
δ^c_{cb3}	rho2	Kumaraswamy	0.0285	0.0079	0.0285	0.0079
δ^c_{cb4}	rho3	LogNormal	0.249	0.298	0.249	0.296
r^0	r_0	Normal	32.4	2.55	32.4	2.55
r^C	r_C	Normal	0.019	0.0015	0.019	0.0015
r^T	r_T	Normal	4.165	0.328	4.165	0.328

Table 3: Targeted correlations for the carbon cycle parameters

	alpha0	alpha1	alpha3	rho1	rho2	rho3
alpha0	1					
alpha1	-0.6348	1				
alpha3	0.0876	-0.1113	1			
rho1	0.5673	-0.1757	-0.1047	1		
rho2	-0.0190	0.3920	-0.5641	0.0686	1	
rho3	-0.3432	-0.4030	-0.1319	0.0686	-0.2532	1

target and sample correlation, is the correlation between alpha3 and rho2. The desired correlation is -0.56 and the sample correlation is around 0. This requires further investigation, but as noted in Swiler and Wyss (2004), there can be inconsistencies in the desired correlations that cannot be overcome when sampling.

Forcing and energy balance model parameters

The sampling strategy for the forcing and energy balance model parameters is provided in Table 5.[33] Similar to the carbon cycle parameters, the sampled distribution modes line up well with the desired modes, with the largest deviation, around 1.1%, for the mean of the C_0 parameter. Tables 6 and 7 provide, respectively, the targeted and sample correlations for the energy balance model parameters. The deviations between the two are much less pronounced than for the correlations across the carbon cycle parameters—particularly for the two pairs with large correlations, i.e., between xi_3 and f2xco2 and between f2xco2 and t2xco2.

[33] Note that the parameter t2xco2, also denoted as *ECS* is only used in the version of the model that has the Albedo effect on temperature.

Table 4: Sample correlations for the carbon cycle parameters

	alpha0	alpha1	alpha3	rho1	rho2	rho3
alpha0	1					
alpha1	-0.5922	1				
alpha3	-0.0107	0.0023	1			
rho1	0.3438	-0.0945	-0.0035	1		
rho2	-0.0203	0.3843	-0.0164	0.0398	1	
rho3	-0.2339	-0.1334	-0.0047	-0.0482	-0.1903	1

Table 5: Sampling assumptions for the forcing and EBM parameters

			Expected		Sample	
X	Code	Distribution	mu	σ	mu	σ
α^{fch4}	afch4	Triangular	0.0360	0.0016	0.0360	0.0016
$1/C_1$	xi_1	Pareto	0.1400	0.0291	0.1400	0.0290
$1/C_0$	C_0	Pareto	128.0496	∞	126.6281	259.9582
ξ_3	xi_3	Triangular	0.7457	0.1738	0.7457	0.1738
f2xco2	f2xco2	Normal	3.4594	0.4367	3.4593	0.4369
ECS	t2xco2	Normal	3.2531	0.8003	3.2531	0.8002

Table 6: Targeted correlations for the energy balance model parameters

	xi_1	C_0	xi_3	f2xco2	t2xco2
xi_1	1				
C_0	-0.0445	1			
xi_3	-0.4372	-0.1198	1		
f2xco2	0.0139	-0.0397	-0.4623	1	
t2xco2	-0.1934	-0.0802	0.0655	0.6512	1

Table 7: Sampled correlations for the energy balance model parameters

	xi_1	C_0	xi_3	f2xco2	t2xco2
xi_1	1				
C_0	-0.0184	1			
xi_3	-0.3383	-0.0528	1		
f2xco2	0.0095	-0.0064	-0.4505	1	
t2xco2	-0.1676	-0.0175	0.0623	0.6510	1

Table 8: Sampling assumptions for the tipping point parameters

X	Code	Distribution	Expected		Sample	
			mu	σ	mu	σ
φ^{ch4}	propCH4	Normal	0.0230	0.0060	0.0230	0.0060
β^{pf}	beta_PF	Normal	0.1720	0.0261	0.1720	0.0261
C^{pf}	C_PF	Normal	1035	76.553	1035	76.553
χ^{psv}	propPassive	Normal	0.4	0.0550	0.4	0.0550
τ^{psv}	Tau_PF	Normal	70	30	70	30
$NYEAR^{AMAZ}$	NYear_AMAZ	Triangular	103.3333	52.4934	103.3332	52.4932
β^{GIS}	Beta_GIS	Normal	-1.06E-05	4.88E-08	-1.06E-05	4.88E-08
R^{WAIS}	Rate_WAIS	LogNormal	0.0033	0.0017	0.0033	0.0017

Table 9: Sampling assumptions for the tipping point parameters

X	Code	Distribution	Expected		Sample	
			mu	σ	mu	σ
φ	yDamWeight	Triangular	0.5	0.2041	0.5	0.2041
η	eta	Triangular	1.3333	0.3118	1.3333	0.3118
ρ	rho	Triangular	0.0103	0.0039	0.0103	0.0039

Tipping point parameters

Table 8 presents the sampling strategy for the uncertain parameters for the various tipping points. The underlying distribution functions are relatively regular and there are very little deviations between the targeted and samples modes. None of these parameters is subject to a targeted correlation.[34]

Economic module parameters

Table 9 presents the sampling strategy for the uncertain parameters for the economic module. All three of the uncertainty parameters follow a triangular distribution and there is virtually no deviation between the targeted and sampled modes.

In addition to these global variables, there are also two sets of country-specific uncertain parameters. The linear damage parameter for sea-level rise, χ^{slr}, is assumed to be characterized by a Triangular distribution, with country-specific estimates for the total range and the mode. The damage from changing temperatures is specified using a quadratic function. Both the linear and quadratic coefficients, respectively β_1 and β_2 are assumed to be characterized by a Normal distribution, with country-specific estimates of the mean and standard deviation. In addition, Dietz et al. (2021) provides a correlation matrix for all pairs of the coefficients, i.e., $\rho_{c,c'}^{i,j}$, where i and j take the values of 1 and 2, respectively for β^1 and β^2, and c, c' represents all country pairings.

4 An application

This section describes an application of the LHS tool coupled to the IAM model, i.e., the (partial) Meta21 model.

[34] The AMOC tipping point is not subject to a sampling strategy—though there are four sets of estimates for the parameters provided by four GCMs.

4.1 Sampling

The sampling is done with LHS using the distributional assumptions described earlier. There are 9 uncertain parameters in the carbon cycle module—some of which have an identified correlation matrix; there are 6 uncertain parameters in the forcing and EBM module—again, some of which are assumed to be correlated; there are 8 independent tipping point uncertain parameters; and there are 3 uncertain parameters in the economic module. The damage module contains a very large number of uncertain parameters due to the fact that the damage parameters—of which there are three—are country specific. In addition the sampling takes into account the correlation matrix for the temperature-based damage coefficients. The input file for the LHS utility has over 75,000 lines in order to take into account the very large correlation matrix across the country-based temperature damage coefficients. Despite the size of the resulting LHS cube, the software operates in under 13 seconds. The resulting cube has 582 random variables sampled 10,000 times.

4.2 Simulating

There are two core simulation files. The first, `Meta21.gms`, initializes the framework: defining dimensions, user-specific flags, and central parameter estimates for each module. In addition, it can solve the Meta21 model with the central parameter estimates and without the stochastic elements of the tipping points. In the absence of the stochastic tipping points, the model assumes a fixed starting point for the tipping elements, subject to the temperature profile. This file also reads two other GAMS file, one called `Meta21Header`, which contains the bulk of the definitions of the core dimensions for the model, and `Meta21Model`, which contains the model specification and is the second core simulation file. The final part of `Meta21.gms` prepares the stochastic simulations. It reads the sample file, created by the LHS utility, as well as a file that facilitates the mapping of the sample variable names to the GAMS names (for indexed variables). It also initializes the random number seed for each of the simulations, i.e., all 10,000. This allows to use the same seed for the two simulations that calculate the SCC, i.e., without and with the extra pulse of CO_2. Thus step 1 is to run `Meta21.gms`, and to save its image using the GAMS `save` option.[35]

The Monte Carlo runs are driven by the file `Meta21Stoch`. It uses the GAMS option `restart` to start the Monte Carlo runs. By default, Meta21Stoch will run all 10,000 simulation sequentially.[36] At the beginning of each of the 10,000 runs, the uncertain parameters are initialized to one draw from the full sample and the random seed is initialized. When calculating the SCC, i.e., when running the simulation with the extra pulse of CO_2 emissions, it is assumed that the onset of the tipping point is the same as the initial simulation, i.e., without the extra pulse.

4.3 Results

The social cost of carbon

With the current configuration for the simulations with tipping points, the simulation fails (for numerical) reasons for 183 observations, i.e. just under 2% of the sample size. Figure 5 presents the histogram for the remaining SCC values greater than 0 and less than $500. This removes some 17 negative values, of which 10 are less than $100 in absolute terms, and 51 values above $500, of

[35] The file can be run any number of times to test the model and its parameterization. Since the focus here is on the Monte Carlo runs, we ignore any results generated by running `Meta21.gms` and its main aim is simply to setup the framework for the Monte Carlo runs.

[36] The supplemental materials contains an alternative method to run the stochastic simulations, which uses a Python script to use multi-processing capabilities. This reduces run time to under 1 hour, compared to 4 hours with sequential processing.

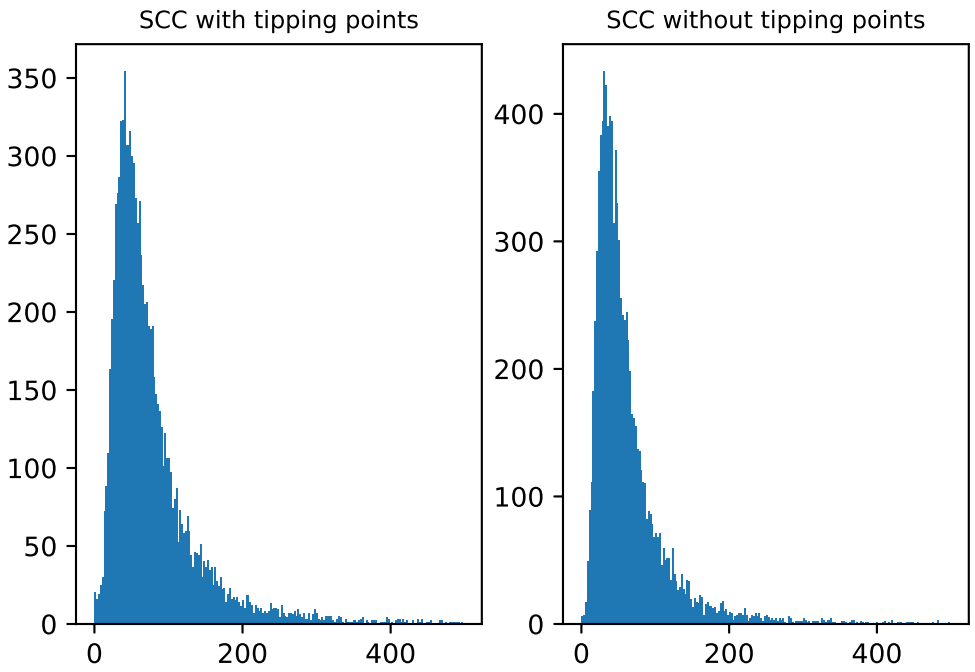


Figure 5: Histogram of SCC

which 6 are greater than $1000. There are also some unusual pairings and it might be useful to drill down to see if there are plausible explanations. For example, sample number 1885 has an SCC of $82 without the tipping points, but over $45,000 when the tipping points are included. Another extreme example is for sample number 8947. With tipping points, the SCC is -$89 and around -$6.5 million without the tipping points.

Figure 6 summarizes the SCC for the two configurations—with and without the tipping points. The median value of the SCC with tipping points is $60, and without tipping points is $48. The latter is somewhat lower than the value of $52 in Dietz et al. (2021), part of which could be attributed to differences in prices and exchange rates, we are using $2010 prices and purchasing power parity (PPP) exchange rates,[37] with the remainder attributable to changes in sampling strategy and removal of outliers. Thus, the inclusion of these tipping points raises the estimate of the SCC by around 26%, similar in magnitude to the increase in Dietz et al. (2021).

Tipping points start year

Table 10 describes the average start year for the uncertain tipping points. The parameterization assumes that the AMOC and OMH starting points are expected to start almost immediately, with the average start year for WAIS around 2058 and a much later start year for the Amazon dieback. The table also provides the frequency of the tipping point within the 2010/2200 time frame over the 10,000 samples. With the exception of the Amazon dieback, the tipping points are present in almost the entire sample, with the Amazon dieback occurring in roughly 50% of the samples.

[37] Dietz et al. (2021) inflate the SCC results by 20% to render the values in 2020 prices and exchange rates. Using the same factor would push our SCC value to $57.

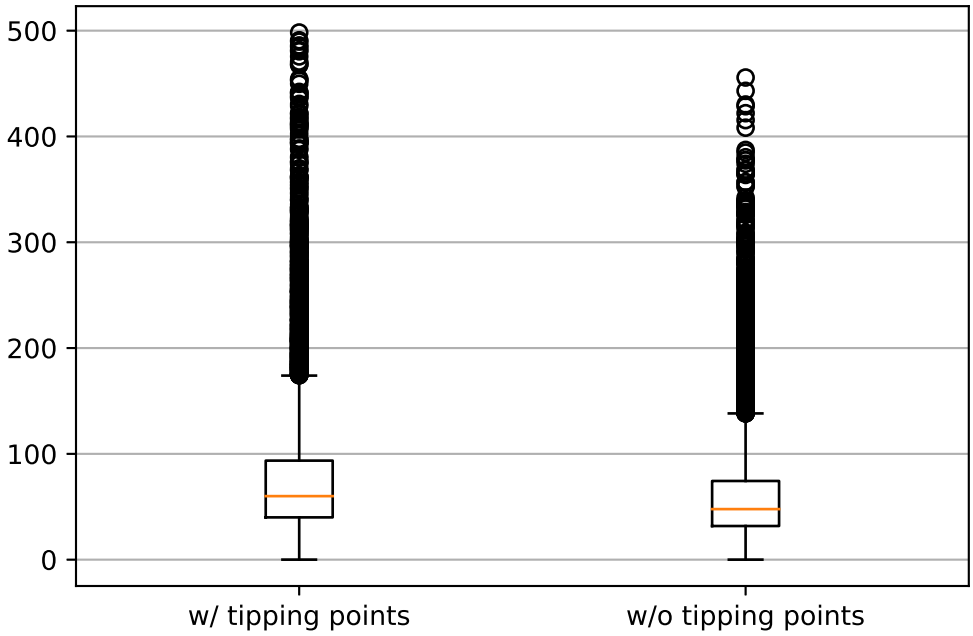


Figure 6: Box and whisker plot of SCC

Tipping point	Average start year	Frequency
AMOC	2012	9749
OMH	2016	9743
WAIS	2058	9743
AMAZ	2110	4483

Table 10: Tipping points start year

The distribution of the start years can be gleaned from Figure 7. The start years are heavily skewed towards the left, with the exception of the Amazon dieback, which has a relatively uniform distribution.[38]

Change in temperature

Figure 8 shows the change in temperature across the two configurations. The outer-most curves show the maximum and minimum temperature profiles across the 10,000 samples. The inner curves show the 95 percentile range. The middle curve shows the average. With the tipping points, the average temperature change could reach 4 °C by 2200, with a probably range between 2-6 °C, and a maximum of 10 °C. Without the tipping points, the ranges are more narrow—notably because 3 of the 5 tipping points have a direct impact on atmospheric concentrations of greenhouse gases.

[38] N.B. The x-axis differs across the tipping points.

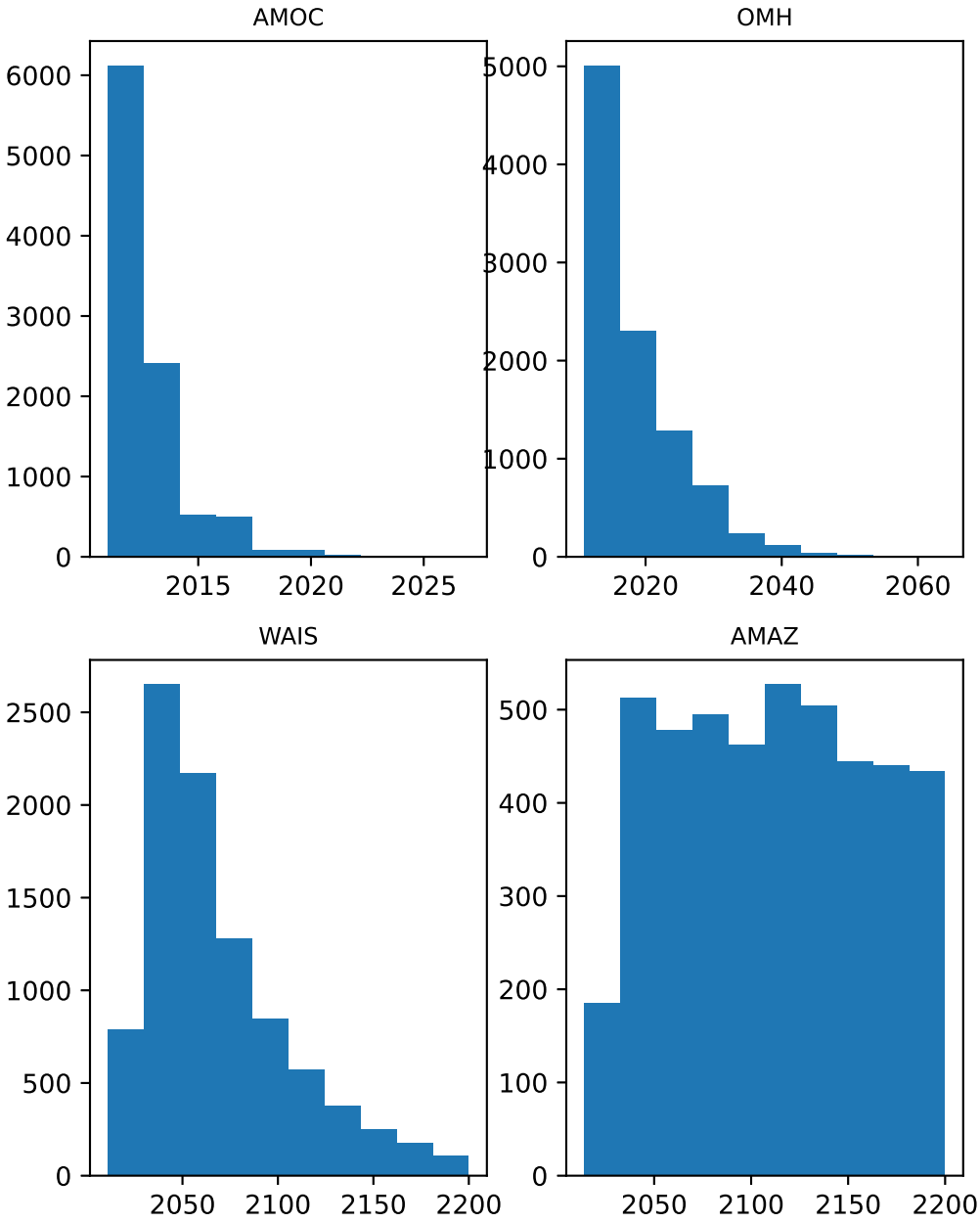


Figure 7: Tipping points start year

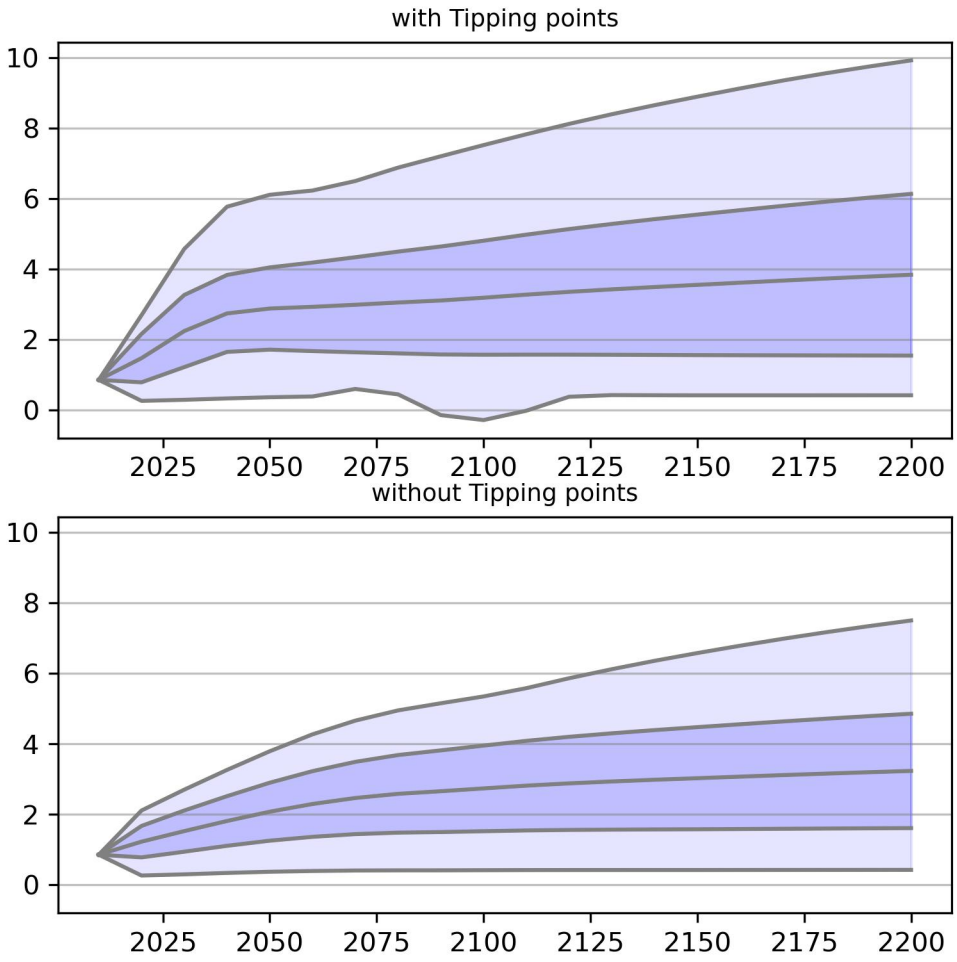


Figure 8: Change in temperature, °C

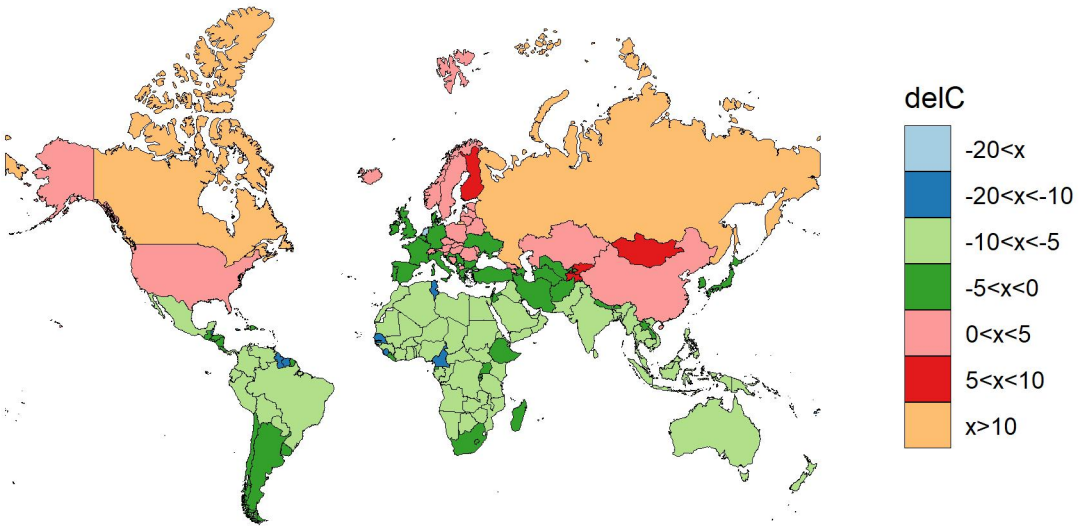


Figure 9: Average change in consumption due to climate damages including tipping points

Impacts on consumption

We conclude by assessing the impacts of damages and the tipping points on consumption. Figure 9 shows the average change in consumption in 2100 across countries relative to a no-damage baseline. There are a handful of countries that could benefit on average from rising temperatures—most visibly in the upper latitudes, e.g., Canada, Finland, Mongolia and Russia, but also China and the United States. The countries that would perceive the highest negative shocks are not visible as these are mostly small island nations. Other countries that would suffer damages in excess of 10% include Guyana and Suriname in South America, and Cameroon, Senegal, Sierra Leone, and Tunisia in Africa.

Figuree 10 shows the impacts of including the tipping points in the assessment of the average damages. The impact tends to be concentrated in the range of additional damages of between 0 and 5% relative to the model outcomes when the tipping points are ignored. There are one or two countries that may see improvement when tipping points are included, e.g., Tajikistan, and again, the most impacted countries tend to be the small island states.

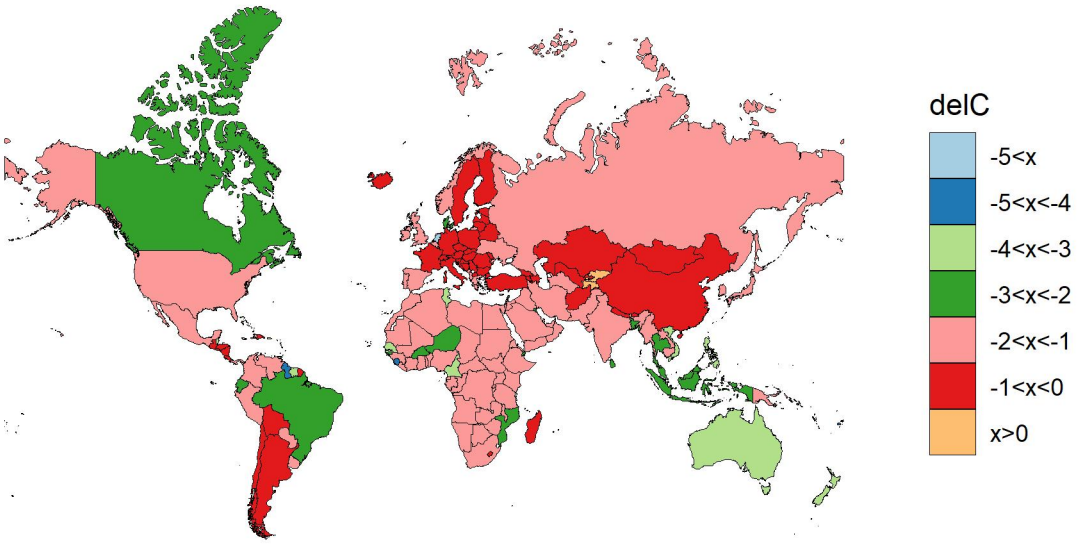


Figure 10: Average change in consumption from including tipping points

5 Concluding remarks

This paper introduces a new version of a well-documented program, if perhaps little known among economists, to generate a Latin Hypercube Sample, including, if desired assumptions on the co-variability of the uncertain variables. The new program, coded in C/C++, has been tested exhaustively and also includes additional functionality compared with the original FORTRAN code, notably additional distributions and output options. The LHS utility can be readily embedded in a workflow, which links a simulation model with a Monte Carlo-style sample of random deviates with the desired distributional characteristics.

As an example, the LHS utility was coupled with the Meta21 integrated assessment model (Dietz et al. (2021)). The Meta21 model has two distinct features related to Monte Carlo simulations. The first relates to a large number of uncertain parameters, for which the LHS utility was utilized to generate a random sample of 10,000 observations—including *a priori* assumptions correlations on a large number of the uncertain parameters. The second relates to the uncertain timing of the onset of the specific tipping points, modeled using a hazard rate and a draw from a binomial distribution.

There are many avenues for further exploration. Within the confines of the Meta21 model, there are multiple configurations of the model, which could be used to unpack in greater depth the tipping points—alternative RCP/SSP combinations, the relative importance of specific tipping points, alternative specifications for any of the components of the model, etc. Beyond Meta21, there are opportunities to port some of the components—notably the modeling of tipping points and the damage components, to other IAMs, including CGE models. Finally, generation of random deviates that capture as broadly as possible the degree of uncertainty in the simulation framework is a

necessary, but only first step in uncertainty analysis. Additional analysis is needed to determine the level of influence of uncertain parameters on key indicators—such as the social cost of carbon. This additional analysis could provide guidance on where to focus future research to reduce uncertainty. Examples of this type of analysis in the context of climate change include Lobell et al. (2013), Butler et al. (2014), Anderson et al. (2014) and Gillingham et al. (2018).

Acknowledgments

The initial development of this new version of the LHS utility was done a number of years ago while the author was working at the World Bank, so thanks to Hans Timmer who provided the space for its development. Needless to say, the delay in making the utility public is the author's alone. Let me also acknowledge help from Simon Dietz for elucidating some of the complexities of the Meta21 model. A final thanks go to Steven Dirkse for helping with the parallelizing of the Monte Carlo simulations.

Bibliography

Anderson, B., E. Borgonovo, M. Galeotti, and R. Roson. 2014. "Uncertainty in Climate Change Modeling: Can Global Sensitivity Analysis Be of Help?" *Risk Analysis*, 34(2): 271–293. doi:10.1111/risa.12117.

Anthoff, D., and R.S.J. Tol. 2014. "The Climate Framework for Uncertainty, Negotiation and Distribution (FUND), Technical Description, Version 3.9." Mimeo. http://www.fund-model.org/files/documentation/Fund-3-9-Scientific-Documentation.pdf.

Burke, M., S.M. Hsiang, and E. Miguel. 2015. "Global non-linear effect of temperature on economic production." *Nature*, 527: 235–239. doi:10.1038/nature15725.

Butler, M.P., P.M. Reed, K. Fisher-Vanden, K. Keller, and T. Wagener. 2014. "Identifying parametric controls and dependencies in integrated assessment models using global sensitivity analysis." *Environmental Modelling & Software*, 59: 10–29. doi:10.1016/j.envsoft.2014.05.001.

Dellink, R., J. Chateau, E. Lanzi, and B. Magné. 2017. "Long-term economic growth projections in the Shared Socioeconomic Pathways." *Global Environmental Change*, 42: 200–214. doi:10.1016/j.gloenvcha.2015.06.004.

Diaz, D., and K. Keller. 2016. "A Potential Disintegration of the West Antarctic Ice Sheet: Implications for Economic Analyses of Climate Policy." *American Economic Review*, 106(5): 607–611. doi:10.1257/aer.p20161103.

Diaz, D.B. 2016. "Estimating global damages from sea level rise with the Coastal Impact and Adaptation Model (CIAM)." *Climatic Change*, 137: 143–156. doi:10.1007/s10584-016-1675-4.

Dietz, S., J. Rising, T. Stoerk, and G. Wagner. 2021. "Economic impacts of tipping points in the climate system." *Proceedings of the National Academy of Sciences*, 118(34). doi:10.1073/pnas.2103081118.

Gillingham, K., W. Nordhaus, D. Anthoff, G. Blanford, V. Bosetti, P. Christensen, H. McJeon, and J. Reilly. 2018. "Modeling Uncertainty in Integrated Assessment of Climate Change: A Multimodel Comparison." *Journal of the Association of Environmental and Resource Economists*, 5(4): 791–826. doi:10.1086/698910.

Helton, J.C., and F.J. Davis. 2002. "Latin Hypercube Sampling and the Propagation of Uncertainty in Analyses of Complex Systems." Sandia National Laboratories, Sandia Technical Report No. SAND2001-0417. doi:10.2172/806696.

Hope, C., and K. Schaefer. 2016. "Economic impacts of carbon dioxide and methane released from thawing permafrost." *Nature Climate Change*, 6: 56–59. doi:10.1038/nclimate2807.

Horridge, M., and K. Pearson. 2011. "Systematic Sensitivity Analysis with Respect to Correlated Variations in Parameters and Shocks." Global Trade Analysis Project (GTAP), Department of Agricultural Economics, Purdue University, West Lafayette, IN, GTAP Technical Paper No. 30. https://www.gtap.agecon.purdue.edu/resources/res_display.asp?RecordID=3496.

Iman, R.L., and W.J. Conover. 1982. "A distribution-free approach to inducing rank correlation among input variables." *Communications in Statistics - Simulation and Computation*, 11(3): 311–334. doi:10.1080/03610918208812265.

Joos, F., R. Roth, J.S. Fuglestvedt, G.P. Peters, I.G. Enting, W. von Bloh, V. Brovkin, E.J. Burke, M. Eby, N.R. Edwards, T. Friedrich, T.L. Frölicher, P.R. Halloran, P.B. Holden, C. Jones, T. Kleinen, F.T. Mackenzie, K. Matsumoto, M. Meinshausen, G. Plattner, A. Reisinger, J. Segschneider, G. Shaffer, M. Steinacher, K. Strassmann, K. Tanaka, A. Timmermann, and A.J. Weaver. 2013. "Carbon dioxide and climate impulse response functions for the computation of greenhouse gas metrics: a multi-model analysis." *Atmospheric Chemistry and Physics*, 13(5): 2793–2825. doi:10.5194/acp-13-2793-2013.

Kessler, L. 2017. "Estimating the economic impact of the permafrost carbon feedback." *Climate Change Economics*, 8(2): 1–23. https://www.jstor.org/stable/90009412.

Leach, N.J., S. Jenkins, Z. Nicholls, C.J. Smith, J. Lynch, M. Cain, T. Walsh, B. Wu, J. Tsutsui, and M.R. Allen. 2021. "FaIRv2.0.0: a generalised impulse response model for climate uncertainty and future scenario exploration." *Geoscientific Model Development*, 14(5): 3007–3036. doi:10.5194/gmd-14-3007-2021.

Lobell, D.B., U.L.C. Baldos, and T.W. Hertel. 2013. "Climate adaptation as mitigation: the case of agricultural investments." *Environmental Research Letters*, 8(1): 015012. doi:10.1088/1748-9326/8/1/015012.

Manne, A.S., and R.G. Richels. 2005. *Merge: An Integrated Assessment Model for Global Climate Change*, Boston, MA: Springer US. pp. 175–189. doi:10.1007/0-387-25352-1_7.

Matsumoto, M., and T. Nishimura. 1998. "Mersenne Twister: A 623-Dimensionally Equidistributed Uniform Pseudo-Random Number Generator." *ACM Transactions on Modeling and Computer Simulation*, 8(1): 3–30. doi:10.1145/272991.272995.

McKay, M.D., R.J. Beckman, and W.J. Conover. 1979. "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code." *Technometrics*, 21: 239–245. doi:10.1080/00401706.1979.10489755.

Nishimura, T. 2000. "Tables of 64-Bit Mersenne Twisters." *ACM Transactions on Modeling and Computer Simulation*, 10(4): 348–357. doi:10.1145/369534.369540.

Nordhaus, W. 2019. "Economics of the disintegration of the Greenland ice sheet." *Proceedings of the National Academy of Sciences*, 116(25): 12261–12269. doi:10.1073/pnas.1814990116.

Nordhaus, W.D. 2017. "Revisiting the social cost of carbon." *Proceedings of the National Academy of Sciences*, pp. . doi:10.1073/pnas.1609244114.

Swiler, L.P., and G.D. Wyss. 2004. "A User's Guide to Sandia's Latin Hypercube Sampling Software: LHS UNIX Library Standalone Version." Sandia National Laboratories, Sandia Technical Report No. SAND2004-2439. doi:10.2172/919175.

Appendix: The LHS algorithm

The origin of LHS is attributed to the work of McKay (McKay et al. (1979)) and Iman and Conover (Iman and Conover (1982)), which is very nicely described by Helton and Davis (Helton and Davis (2002)). LHS has two components. The first is intended to cover as much of the sampling space as possible by using a simplified version of stratified sampling. The key idea behind this is to reduce the number of samples—particularly for complex models. In random sampling, each random deviate is selected over the entire range of possible outcomes, i.e., over the probability range of $[0, 1]$. The LHS technique divides the $[0, 1]$ range into n sub-intervals of equal length set to $1/n$, where n is the number of deviates being requested in the sample. In this sense, LHS covers the entire probability range. For example, for a sample size of 10, the procedure is certain to select a deviate from the $[0, 0.1]$, whereas their is no guarantee of this outcome from pure random sampling. This simplified stratified sampling is used for each random variable, and hence LHS is sure to cover the entire range of the multi-variate sample.

The second component is the pairing of the deviates so as to match a targeted correlation matrix for the random variables. In the absence of a user-specified correlation matrix, the algorithm pairs the deviates to target a zero-correlation matrix, or as close as possible. The pairing part of the algorithm has several components, which are described below. Let X be a matrix of the initial random deviates. It is of dimension $n \times k$, where n is the sample size and k is the number of random variables. Given the sampling strategy, each column of X is sorted in ascending order. Let C be the targeted correlation matrix, in the absence of a user-specified list of correlations, it is the identity matrix. It has the dimensions of $k \times k$, is symmetric and semi-positive definite. Therefore, it can be split using the Cholesky factorization, which is a lower triangular matrix, say H, with the property $C = HH'$.[39]

The second step involves developing an auxiliary matrix of random deviates, based on the so-called van der Waerden scores. This matrix will be transformed such that it has the same correlation as the target correlation, and once this is accomplished, its rank scores will be used to derive the same rank for the original sample.

The van der Waerden scores represent the inverse of the normal standard deviation and each column will have the same initial distribution, until fully randomized. For example, in the case of $n = 10$, each column has the following values until randomized:

$$\begin{bmatrix} -1.335178 \\ -0.908458 \\ -0.604585 \\ -0.348756 \\ -0.114185 \\ 0.114185 \\ 0.348756 \\ 0.604585 \\ 0.908458 \\ 1.335178 \end{bmatrix}$$

which is the inverse of the cumulative distribution function of the standard normal distribution at each cumulative interval defined by $1/11$. After randomization, column by column, the matrix S, may look something like the following:

[39] Horridge and Pearson (2011) use a similar technique for their method of systematic sensitivity analysis using Gaussian quadrature and allowing for imposing a covariance matrix on the uncertain parameters.

Table 11: Example of a sample definition for an EBM model

Name	Distribution	Parm 1	Parm 2	Parm 3
xi_1	Pareto	5.907	0.11628	
C_0	ParetoAlt	1.7062	53.0	
xi_3	Triangular	0.5	0.5	1.23723
f2xco2	Normal	3.45938	0.43674	
t2xco2	Normal	3.25312	0.80031	

$$S = \begin{bmatrix} -0.11419 & -1.33518 & -0.34876 & -0.34876 & 0.90846 \\ 0.11419 & 0.34876 & 0.11419 & 0.34876 & -0.90846 \\ -0.90846 & 0.60459 & -0.90846 & -0.60459 & 0.60459 \\ -1.33518 & 0.11419 & 0.60459 & -0.90846 & -0.34876 \\ 0.90846 & -0.60459 & -1.33518 & -0.11419 & -0.11419 \\ 0.60459 & -0.34876 & 0.34876 & 0.90846 & 1.33518 \\ -0.34876 & 1.33518 & -0.60459 & 1.33518 & -1.33518 \\ 1.33518 & -0.11419 & 0.90846 & -1.33518 & -0.60459 \\ 0.34876 & 0.90846 & 1.33518 & 0.11419 & 0.11419 \\ -0.60459 & -0.90846 & -0.11419 & 0.60459 & 0.34876 \end{bmatrix}$$

for a sample of 5 random variables. From the matrix S, which to emphasize, has nothing to do with the sample itself, we can derive its correlation matrix $C^s = Corr(S)$, and the subsequent Cholesky decomposition ($C^s = H^s (H^s)'$) and the inverse of the Cholesky decomposition, call the latter $M = (H^s)^{-1}$. The following step transforms S to S^1:

$$S^1 = S (HM)'$$

where $(HM)'$ is a projection matrix, based on the targeted Cholesky matrix and the inverse of the Cholesky matrix derived from the van der Waerden scores. S^1 it turns out will have the desired correlation matrix, i.e., its correlation matrix will be C, the targeted correlation matrix. The final step is to assume that the matrix S^1, of transformed van der Waerden scores, has the same rank correlation as the sample of random deviates. So the algorithm proceeds to calculate the rank order of each column of S^1 and uses that rank order to re-order the cells in each column of X.

The following numerical example will show how the algorithm works in practice. The example is for a set of five random variables with the attributes in Table 11. This is an example for the energy balance model (EBM) of the Meta21 model, which drives the change in temperature component. In addition, the components are assumed to be linked to a correlation matrix as below.

$$C = \begin{bmatrix} & xi\_1 & C\_0 & xi\_3 & f2xco2 & t2xco2 \\ xi\_1 & 1.00000 & & & & \\ C\_0 & -0.04451 & 1.00000 & & & \\ xi\_3 & -0.43716 & -0.11978 & 1.00000 & & \\ f2xco2 & 0.01392 & -0.03966 & -0.46228 & 1.00000 & \\ t2xco2 & -0.19343 & -0.08016 & 0.06549 & 0.65122 & 1.00000 \end{bmatrix}$$

The targeted Cholesky matrix is given by:

$$H = \begin{bmatrix} & xi\_1 & C\_0 & xi\_3 & f2xco2 & t2xco2 \\ xi\_1 & 1.00000 & & & & \\ C\_0 & -0.04451 & 0.99901 & & & \\ xi\_3 & -0.43716 & -0.13937 & 0.88852 & & \\ f2xco2 & 0.01392 & -0.03907 & -0.51956 & 0.85343 & \\ t2xco2 & -0.19343 & -0.08886 & -0.03540 & 0.74060 & 0.63635 \end{bmatrix}$$

The initial sample of random deviates is given below.[40] The rank correlation is exactly 1 for all pairs of columns as all of the deviates are in ascending order in each column, and the Spearman correlation matrix is also nearly the identity matrix.

$$X = \begin{bmatrix} xi\_1 & C\_0 & xi\_3 & f2xco2 & t2xco2 \\ 0.11646 & 56.14932 & 0.51226 & 2.63749 & 2.11869 \\ 0.11920 & 59.22047 & 0.55074 & 3.05827 & 2.28394 \\ 0.12332 & 65.15213 & 0.58688 & 3.13678 & 2.81773 \\ 0.12652 & 65.61294 & 0.62132 & 3.32291 & 2.85096 \\ 0.12781 & 77.97449 & 0.67191 & 3.45491 & 3.10905 \\ 0.13417 & 82.40018 & 0.76344 & 3.52562 & 3.39424 \\ 0.14078 & 95.69734 & 0.79359 & 3.60265 & 3.55156 \\ 0.14771 & 108.00303 & 0.87995 & 3.78290 & 3.91509 \\ 0.15354 & 153.88674 & 0.94936 & 3.89073 & 3.92973 \\ 0.17777 & 272.65331 & 1.07535 & 4.59178 & 4.34054 \end{bmatrix}$$

The correlation matrix for S is given by:

$$C^s = \begin{bmatrix} & xi\_1 & C\_0 & xi\_3 & f2xco2 & t2xco2 \\ xi\_1 & 1.00000 & & & & \\ C\_0 & -0.16433 & 1.00000 & & & \\ xi\_3 & 0.16526 & 0.17962 & 1.00000 & & \\ f2xco2 & -0.04612 & 0.21876 & -0.21009 & 1.00000 & \\ t2xco2 & -0.01611 & -0.56745 & -0.03087 & -0.03345 & 1.00000 \end{bmatrix}$$

and the corresponding Cholesky matrix is:

$$H^s = \begin{bmatrix} & xi\_1 & C\_0 & xi\_3 & f2xco2 & t2xco2 \\ xi\_1 & 1.00000 & & & & \\ C\_0 & -0.16433 & 0.98641 & & & \\ xi\_3 & 0.16526 & 0.20963 & 0.96371 & & \\ f2xco2 & -0.04612 & 0.21409 & -0.25666 & 0.94136 & \\ t2xco2 & -0.01611 & -0.57795 & 0.09644 & 0.12141 & 0.80104 \end{bmatrix}$$

from which we derive the inverse matrix:

$$M = \begin{bmatrix} & xi\_1 & C\_0 & xi\_3 & f2xco2 & t2xco2 \\ xi\_1 & 1.00000 & & & & \\ C\_0 & 0.16660 & 1.01378 & & & \\ xi\_3 & -0.20772 & -0.22052 & 1.03765 & & \\ f2xco2 & -0.04553 & -0.29068 & 0.28292 & 1.06229 & \\ t2xco2 & 0.17222 & 0.80205 & -0.16781 & -0.16101 & 1.24837 \end{bmatrix}$$

[40] This will depend on the random sample function and the initial seed.

We now have all the elements to transform the S matrix:

$$S^1 = S\,(HM)' = \begin{bmatrix} xi\_1 & C\_0 & xi\_3 & f2xco2 & t2xco2 \\ -0.11419 & -1.36616 & 0.20236 & 0.01005 & 0.19013 \\ 0.11419 & 0.36713 & -0.08598 & 0.23054 & -0.41555 \\ -0.90846 & 0.50155 & -0.45556 & -0.45183 & 0.24884 \\ -1.33518 & -0.04715 & 1.38002 & -1.12544 & -0.66758 \\ 0.90846 & -0.50155 & -1.61302 & 0.46806 & -0.49921 \\ 0.60459 & -0.27950 & 0.04923 & 0.82639 & 1.55553 \\ -0.34876 & 1.30972 & -0.78274 & 1.13261 & 0.14081 \\ 1.33518 & 0.04715 & 0.01497 & -1.35900 & -1.52515 \\ 0.34876 & 0.96259 & 0.69972 & -0.42452 & 0.40673 \\ -0.60459 & -0.99378 & 0.59100 & 0.69313 & 0.56545 \end{bmatrix}$$

The correlation matrix of S^1 matches the targeted correlation matrix for the sample. The rank ordering of each of the columns of S^1 is given by:

$$R^1 = \begin{bmatrix} xi\_1 & C\_0 & xi\_3 & f2xco2 & t2xco2 \\ 5 & 1 & 7 & 5 & 6 \\ 6 & 7 & 4 & 6 & 4 \\ 2 & 8 & 3 & 3 & 7 \\ 1 & 5 & 10 & 2 & 2 \\ 9 & 3 & 1 & 7 & 3 \\ 8 & 4 & 6 & 9 & 10 \\ 4 & 10 & 2 & 10 & 5 \\ 10 & 6 & 5 & 1 & 1 \\ 7 & 9 & 9 & 4 & 8 \\ 3 & 2 & 8 & 8 & 9 \end{bmatrix}$$

Based on R^1, we re-arrange the rows of X to yield:

$$X^1 = \begin{bmatrix} xi\_1 & C\_0 & xi\_3 & f2xco2 & t2xco2 \\ 0.12781 & 56.14932 & 0.79359 & 3.45491 & 3.39424 \\ 0.13417 & 95.69734 & 0.62132 & 3.52562 & 2.85096 \\ 0.11920 & 108.00303 & 0.58688 & 3.13678 & 3.55156 \\ 0.11646 & 77.97449 & 1.07535 & 3.05827 & 2.28394 \\ 0.15354 & 65.15213 & 0.51226 & 3.60265 & 2.81773 \\ 0.14771 & 65.61294 & 0.76344 & 3.89073 & 4.34054 \\ 0.12652 & 272.65331 & 0.55074 & 4.59178 & 3.10905 \\ 0.17777 & 82.40018 & 0.67191 & 2.63749 & 2.11869 \\ 0.14078 & 153.88674 & 0.94936 & 3.32291 & 3.91509 \\ 0.12332 & 59.22047 & 0.87995 & 3.78290 & 3.92973 \end{bmatrix}$$

which is the final sample. Each column has the same sample distribution as the corresponding column in X, but the deviates have been arranged so that the correlation matrix is 'near' the targeted correlation matrix. In this case, the correlation matrix is:

$$C^1 = \begin{bmatrix} & xi\_1 & C\_0 & xi\_3 & f2xco2 & t2xco2 \\ xi\_1 & 1.00000 & & & & \\ C\_0 & -0.17602 & 1.00000 & & & \\ xi\_3 & -0.31037 & -0.26483 & 1.00000 & & \\ f2xco2 & -0.32630 & 0.55056 & -0.29066 & 1.00000 & \\ t2xco2 & -0.26453 & 0.00238 & 0.12102 & 0.45300 & 1.00000 \end{bmatrix}$$

The 'fit' is not so good in this example, in part because the sample size is very small. Using the LHS tool with a sample size of 20,000, we get the following correlation matrix. Though not exactly the target, it is much closer than that with the sample size of 10. As explained in Swiler and Wyss (2004), there are limitations on the targeted correlations.

$$C^2 = \begin{bmatrix} & xi\_1 & C\_0 & xi\_3 & f2xco2 & t2xco2 \\ xi\_1 & 1.00000 & & & & \\ C\_0 & -0.04270 & 1.00000 & & & \\ xi\_3 & -0.42080 & -0.11510 & 1.00000 & & \\ f2xco2 & 0.01130 & -0.04040 & -0.44650 & 1.00000 & \\ t2xco2 & -0.18610 & -0.07590 & 0.05960 & 0.63470 & 1.00000 \end{bmatrix}$$