Statistical Approximation of High-Dimensional Climate Models

Alena Miftakhova, University of Zurich

Thomas S. Lontzek, RWTH Aachen

Kenneth L. Judd, Hoover Institution, Stanford

Karl Schmedders, University of Zurich

November, 2018

Outline

Integrated Assessment Modelling

Climate models (ESM)

- simulate major components of the climate system
- infer system response to exogenous factors
- no room for interaction with economy

Economic models

- model socio-economic development
- assess the consequences of social decisions
- do not consider effect on climate system

Integrated models (IAM)

- study interference of the two systems
- assess costs associated with climate change
- infer optimal climate policies

Problem Addressed

Challenge of IAMs

- trade-off between the complexities of climate module and economic module
- already involved number of state variables prohibits further extensions

$\Rightarrow$ reduction in the complexity of the climate part would simplify the computations and create more room for the economic modeling

Our aims:

- find a low-dimensional approximation of the climate system
- suggest an efficient technique for such approximation

MAGICC model (Meinshausen, Raper, Wigley, 2011)

"Reduced complexity" climate model

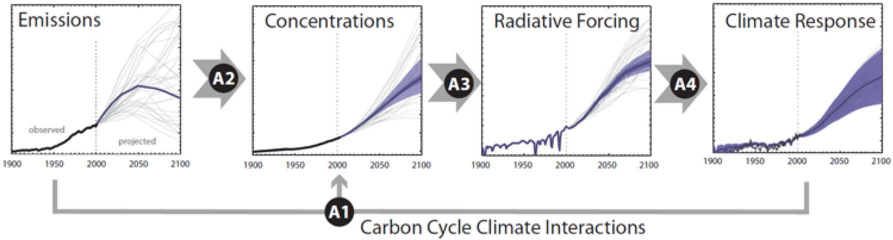
- couples atmosphere-ocean and carbon cycle models
- takes emissions as input and projects future climate response
- calibrated to 20 full complexity climate models and 10 carbon cycle models
- included into several IAMs (IMAGE, MESSAGE, MiniCAM)

What is special about it?

- can emulate 200 paths for each scenario
- any of 400+ parameters can be adjusted
- possibility to introduce arbitrary scenarios

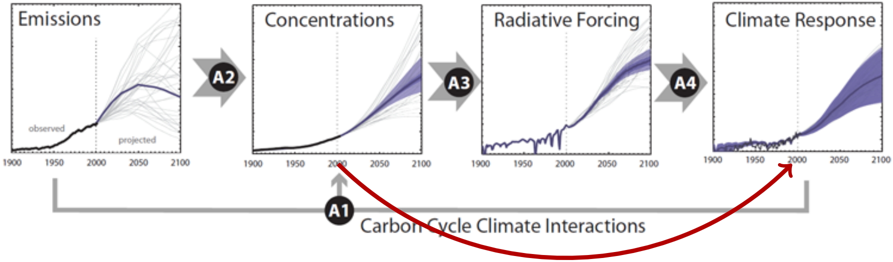
$\Rightarrow$ MAGICC chosen as a benchmark

MAGICC model: four-blocks structure



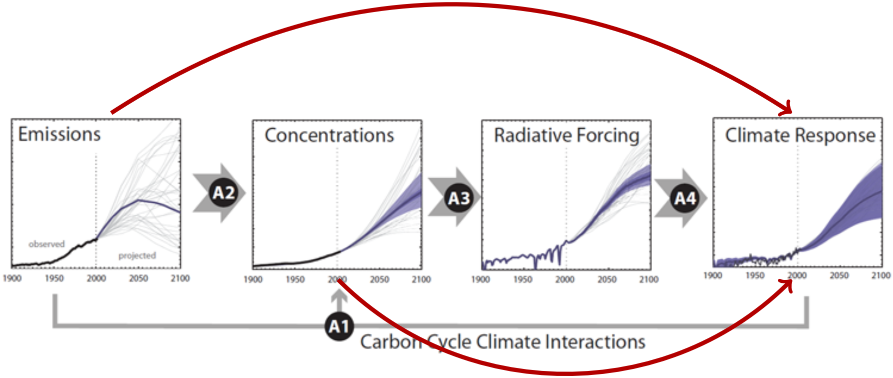
Source: wiki.magicc.org

Reducing complexity of emulation



Source: wiki.magicc.org

Reducing complexity of emulation



Source: wiki.magicc.org

Design of Input Scenarios

General ideas

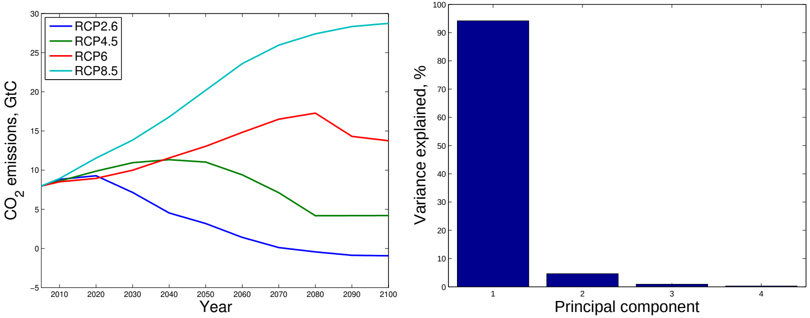
- input scenarios should extract as much information as possible from the unknown complex system
- running additional scenarios is often expensive; therefore, the number of required scenarios should be minimized
- the common approach would be either to create arbitrary scenarios (Castruccio et al., 2014) or to take the conventionally used ones (Meinshausen et al., 2011)

Our approach:

design few orthogonal scenarios and observe the response.

Informativeness of RCP scenarios

Are the RCP scenarios good for training the models?



Employing Chebyshev Polynomials

If we can design any emissions paths, why not make them orthogonal?

Chebyshev polynomials

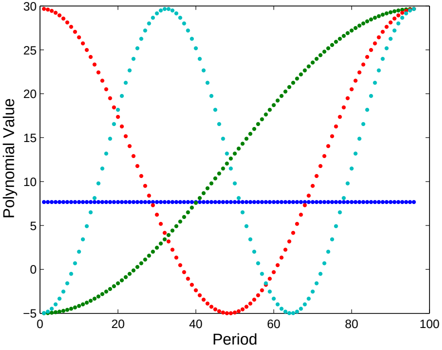
- general form

$$F_n(x) = \cos(n \arccos(x)), x \in [-1, 1]$$

- satisfy discrete orthogonality condition given that the function values are taken on N optimal nodes x_k:

$$\sum_{k=0}^{N-1} F_i(x_k) F_j(x_k) = \begin{cases} 0 & : i \neq j \\ N & : i = j = 0 \\ N/2 & : i = j \neq 0 \end{cases}$$

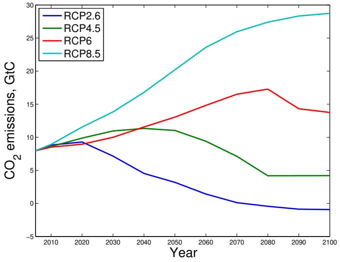
Design of Orthogonal Scenarios



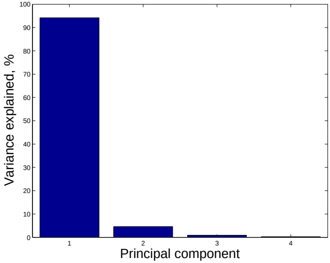
- to preserve zero correlation, values are taken on the optimal nodes
- strings of 96 polynomial values constitute emissions paths in years 2005-2100
- the scenarios are scaled to the range of RCPs; the zero-degree polynomial is kept at 2004 emissions level
- other gases are kept at the average level across the 4 RCPs

Informativeness of scenarios

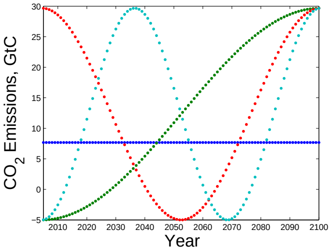
RCP Scenarios



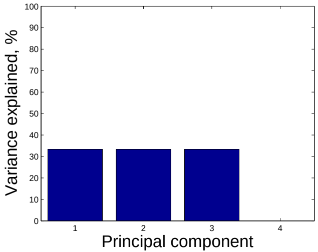
RCP Principal Components



Chebyshev Scenarios



Cheb. Principal Components

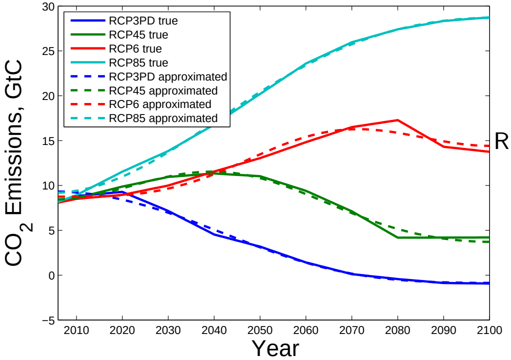


Approximation of RCP scenarios

How well do Chebyshev scenarios approximate RCP scenarios?

$$E_{RCP}(t) \approx \sum_{k=0}^{m} a_k F_k(t)$$

$E_{RCP}(t)$ - CO_2 emissions in RCP scenario in year t,
$F_k(t)$ - scaled value of Chebyshev polynomial of degree k at node t.



RMSE: $0.36 GtCyr^{-1}$

General Model

Dynamic model

Lagged Dependent Variable in addition to exogenous covariates is employed to fit the data

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \sum_{j=2}^{J} \beta_j X_{j,t-1} + \varepsilon_t$$

Y_t - predicted variable in year t, $X_{j,t}$ - $\mathrm{j_{th}}$ covariate in year t.

Errors

assumed to follow ARMA(1,1) process:

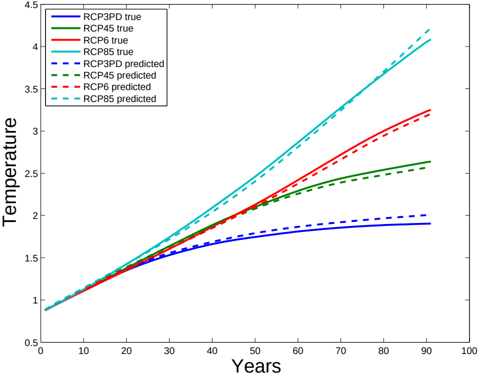
$$\varepsilon_t = a\varepsilon_{t-1} + bu_{t-1} + u_t$$

where $u_t \sim N(0, \sigma^2)$.

Results: Concentrations to Temperature

$$T_t = \beta_0 + \beta_1 T_{t-1} + \beta_2 S_{t-1} + \varepsilon_t$$

T_t - temperature anomaly in year t,
S_t - CO_2 concentrations in year t.



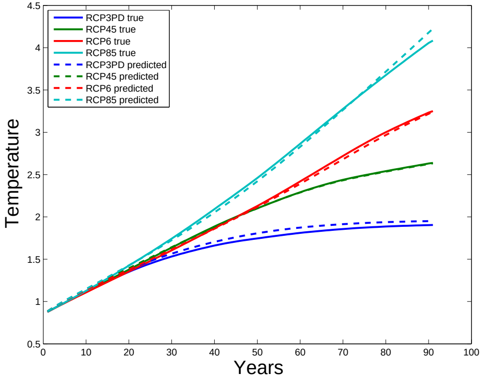
Parameter	Value
β_0	0.1230
β_1	0.6820
β_2	0.0286
a	0.9873
b	0.1832
σ	0.0018

RMSE: 0.0426℃

Results: Emissions to Temperature

$$T_t = \beta_0 + \beta_1 T_{t-1} + \beta_2 Cum_{t-1} + \varepsilon_t$$

T_t - temperature anomaly in year t,
Cum_t - Cumulative CO_2 emissions from 1765 to year t.



Parameter	Value
β_0	0.2500
β_1	0.7650
β_2	0.3632
a	0.9805
b	0.2128
σ	0.0022

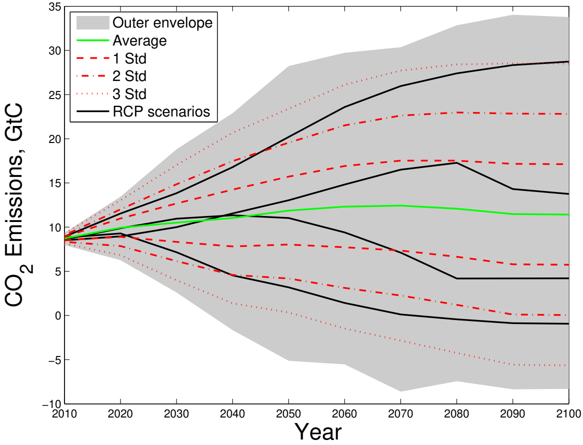
RMSE: 0.0338℃

Simulations: Setting

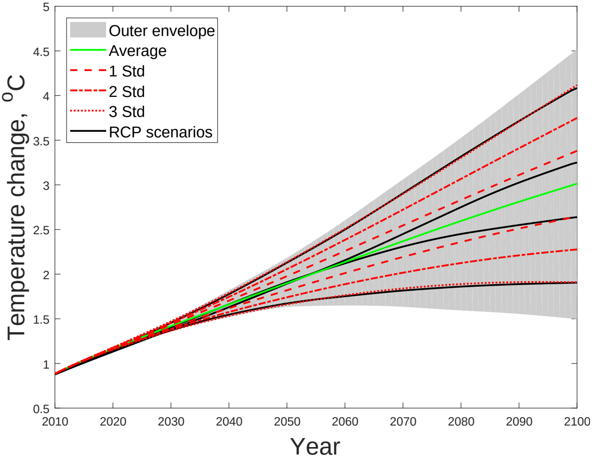
How does Chebyshev training perform on arbitrary RCP-like scenarios?

$$E_t = E_{t-1} + \varepsilon_t; \; \varepsilon_t \sim N(\mu_t, \sigma_t)$$

where μ_t and σ_t for each year are estimated from the four RCPs.



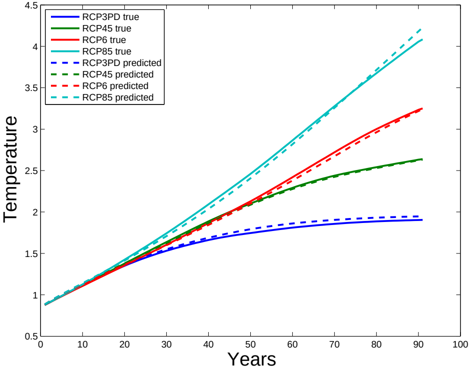
Simulations: Results



	Chebyshev training	**RCP training**
Average RMSE	0.0214	0.0396

Minimal Number of Scenarios

Can we reduce the number of scenarios to save the runs of the climate model?



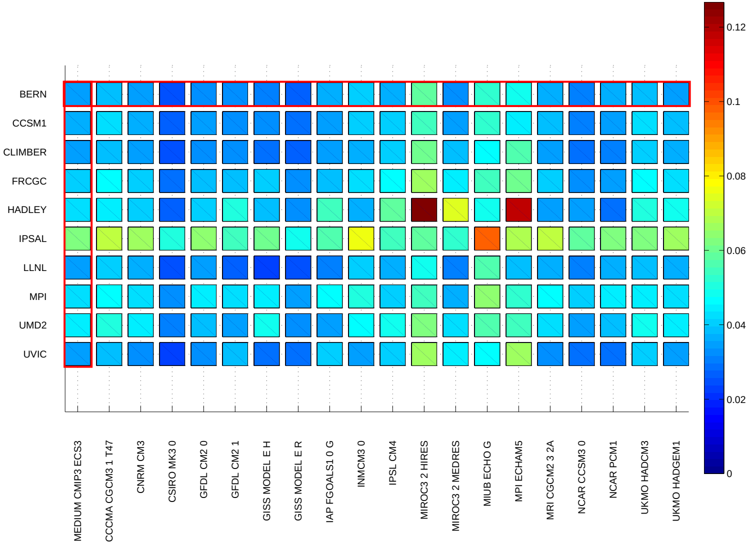
Number of polynomials	1	2	**3**	4
Average RMSE	0.1927	0.0635	**0.0334**	0.0338

Other Settings: Outline

Addressing model uncertainty

- so far only default setting of the climate and carbon cycle parameters in MAGICC was used
- however, there is a great variability among the complex models in their projections of the climate response
- we calibrate the model to 20 different climate models and 10 carbon cycle models
- the average error of prediction across all 200 settings is 0.043°C

Other Settings: Results



Summary

In our study we...

- emphasize the importance of designing the initial scenarios efficiently, prior to the complex model runs
- propose the construction of orthogonal emissions paths that do not carry any repetitive information from one to another
- as an example, construct an approximation of MAGICC and show that the temperature levels can be inferred immediately from the CO_2 emissions data within a one-line model that performs well on the conventional scenarios
- ensure that our approach suits various settings of the climate and carbon cycle parameters