Computers and Mathematics with Applications 64 (2012) 3717–3728

Contents lists available at SciVerse ScienceDirect

Computers and Mathematics with Applications

journal homepage: www.elsevier.com/locate/camwa

Modeling global temperature changes with genetic programming

Karolina Stanislawska [a], Krzysztof Krawiec [a,*], Zbigniew W. Kundzewicz [b,c]

[a] *Institute of Computing Science, Poznan University of Technology, Piotrowo 2, 60-965 Poznań, Poland*
[b] *Institute of Agricultural and Forest Environment, Polish Academy of Sciences, Poznań, Poland*
[c] *Potsdam Institute for Climate Impact Research, Potsdam, Germany*

ARTICLE INFO

Keywords:
Data-driven modeling
Unconstrained optimization
Evolutionary computation
Genetic programming
Global temperature modeling

ABSTRACT

We use genetic programming (GP), a variant of evolutionary computation, to build interpretable models of global mean temperature as a function of natural and anthropogenic forcings. In contrast to the conventional approach, which engages models that are physically-based but very data-demanding and computation-intense, the proposed method is a data-driven randomized search algorithm capable of inducing a model from moderate amount of training data at reasonable computational cost. GP maintains a population of models and recombines them iteratively to improve their performance meant as an ability to explain the training data. Each model is a multiple input–single output arithmetic expression built of a predefined set of elementary components. Inputs include external climate forcings, such as solar activity, volcanic eruptions, composition of the atmosphere (greenhouse gas concentration and aerosols), and indices of internal variability (oscillations in the Ocean-Atmosphere system), while the output is the large-scale temperature. We used the data from the period 1900–1999 for training and the period 2000–2009 for testing, and employed two quality measures: mean absolute error and correlation coefficient. The experiment showed that the models evolved by GP are capable to predict, based exclusively on non-temperature data, the global temperature more accurately than a reference approach known in the literature.

1. Introduction

The Earth is a very complex system composed of a large number of highly interconnected components. Relationships between the components are complicated and non-linear, and there are multiple feedback loops. Within the Earth system, the climate system is perhaps the most complex sub-system.

There are external drivers controlling the Earth's climate, such as the solar radiation, depending on the distance between Sun and Earth (with account of Earth's orbital patterns), solar activity, volcanic eruptions, properties of the atmosphere (content of greenhouse gases, dust and aerosols) and properties of the Earth's surface (albedo of the surface and water on and under the land surface). In addition, there are several patterns of oscillation in the Ocean-Atmosphere system, such as El Niño-Southern Oscillation (ENSO), North Atlantic Oscillation (NAO), Atlantic Multi-decadal Oscillation (AMO), Pacific Decadal Oscillation (PDO), that influence the climate.

Moreover, there are internal feedbacks (both negative and positive) in the system, diminishing or amplifying the effects of external drivers and generating variability. Examples of positive feedbacks include albedo change related to shrinkage of the cryosphere and methane emission from thawing permafrost. In case of warming, snow and ice areas decrease,

* Corresponding author.
E-mail addresses: karolina.stanislawska@gmail.com (K. Stanislawska), krawiec@cs.put.poznan.pl (K. Krawiec), kundzewicz@yahoo.com (Z.W. Kundzewicz).

0898-1221/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.camwa.2012.02.049

albedo decreases, less heat/radiation is reflected out into the space, and the Earth's surface gets warmer. As a consequence, permafrost thaws, methane (very strong greenhouse gas) is emitted, resulting in enhancement of the greenhouse effect, hence it gets even warmer. No wonder that such complex feedbacks made the climate system an excellent application area of, among others, chaos theory. Lorentz equations, the flagship model of deterministic chaotic dynamics, have been developed to describe the climate system.

Recently, the problem of climate change and fear for its serious negative impacts has gained vast theoretical and practical interest and high societal relevance. Tangible progress is expected towards better, and more straightforward, interpretation of large-scale temperature change in the future, necessary for informed policy making.

In this study, we approach the task of global temperature modeling using the methodology of genetic programming (GP, [1]), a variant of evolutionary computation devised for automated inference of explanatory models from data. Our main contribution is a specific method of adaptation of GP to the task of global mean temperature modeling (Section 3, supported by computational experiment on real temperature data (Sections 4.1 and 4.2) and interpretation of its results (Section 4.3).

2. Climate modeling

Modeling of global mean temperature is an example of a problem in climatology, a branch of geophysical sciences, that is hard to solve due to the difficulty to interpret the cause-effect chains, in a very complex system driven by multiple factors.

Many recent climate studies make use of simulations with the help of general (global) climate (circulation) models (GCMs) that represent mathematically the behavior of the global climate system and simulate the interactions of the oceans (temperature, salinity, currents), atmosphere (temperature, wind, clouds, water vapor, greenhouse gases, aerosols, atmospheric chemistry), land surface, including carbon cycle, biosphere, and water storage (also in cryosphere). Such climate models are based on an integration of systems of equations representing the basic principles of physics (fluid dynamics equations of Navier–Stokes, laws of Newton, Coriolis, thermodynamics), chemistry, and biology. They take account of incoming solar heat (short-wave radiation) as well as outgoing long-wave (infrared) radiation from the Earth to the space.

Climate models are derived from fundamental physical laws, which are subject to physical approximations appropriate for the large-scale climate system, and further approximated through discretization, using either the finite difference method or the spectral method. Furthermore, representation of the impacts of unresolved processes is required. Some physical processes occur at smaller (sub-grid) scales and cannot be properly modeled. Instead, properties of neglected sub-grid processes must be averaged over a larger scale in a technique known as parameterization. Parameterizations are used to include the effects of various processes, such as convection, cloud cover, land surface processes, albedo, and hydrology. Tuning is needed, because only some parameters can be measured, while others cannot, so that parameter values have to be adjusted, cf. [2].

Prognostic equations are integrated forward in time while diagnostic equations are evaluated from the simultaneous values of the variables. The models depict the climate using a three-dimensional grid over the globe. They often have a horizontal resolution of less than one to a few degrees in longitude and in latitude, 10–20 vertical layers in the atmosphere and 30 or more layers in the oceans. This makes more than a million grid cells. As the time step is of the order of minutes, the computational effort is gigantic, and computational resources become a critical factor that limits the working resolution of the model. On the other hand, the progress in computational technology since the 1960s made it possible to solve such large and complex computational problems of global climate modeling.

Moderate confidence in climate models results from the fact that their fundamentals are based on established physical laws, such as conservation of mass, energy and momentum, supported by a wealth of observations. Advanced climate models mimic essential physical mechanisms and internal feedbacks of the climate system. Such models have been found to reproduce broad observed features of recent and past global mean temperature (aggregate over all the grid cells).

2.1. Limitations of contemporary climate models

Global climate models have been extensively used to simulate observed climate change during the 20th century [3]. Such models were fed with combinations of natural and anthropogenic forcings and proved to be able to reproduce broad, large-scale, features of the observed Earth's climate of the past century (Fig. 1). However, they cannot mimic important details of observed temperature. This also holds for the global mean temperature — the spatial aggregate. For particular years, the difference between the black and the red line in Fig. 1 can be large, up to nearly 0.3 °C.

Many modeling advances have occurred over the past decades to climate models. In 1970s, models represented only atmosphere. Now they also include land-surface, ocean and sea ice, aerosols (sulfate and non-sulfate), carbon cycle, atmospheric chemistry and dynamic vegetation. The dynamical cores (advection, etc.) have been improved, and the horizontal and vertical resolutions of many models have been increased. However, despite the many model improvements, numerous issues remain and model results still show significant errors, even at large scales. One of the sources of errors is that several important small-scale processes cannot be represented explicitly in the models, and so must be included in approximate form as they interact with larger-scale features [2]. This is partly due to limitations in computing power, but also due to limitations in scientific understanding and lack of sufficient data to calibrate the models. Important is also the unavailability of detailed observations of some physical processes. For example, significant uncertainties are associated with