

Data-driven modeling decadal-to-centennial ENSO variability and its response to external forcing

*Aleksei Seleznev, Dmitry Mukhin, Andrey Gavrilov, and Alexander Feigin
Institute of Applied Physics of the Russian Academy of Sciences, Nizhny Novgorod, Russia*

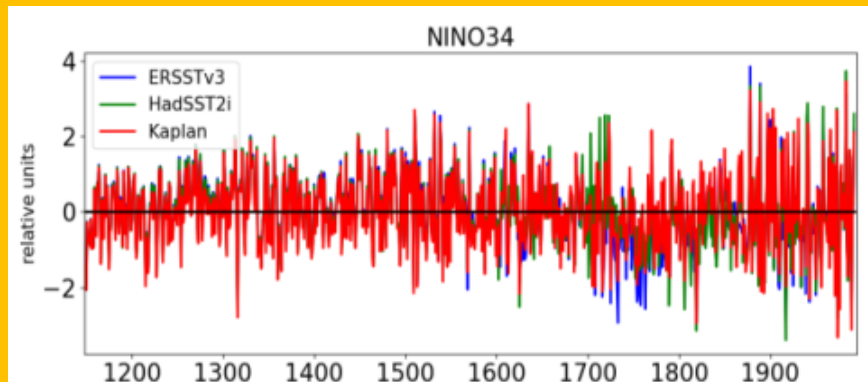
aseleznev@ipfran.ru

Abstract

We investigate the decadal-to-centennial ENSO variability based on nonlinear data-driven stochastic modeling. We construct data-driven model of yearly Niño-3.4 indices reconstructed from paleoclimate proxies based on three different sea-surface temperature (SST) databases at the time interval from 1150 to 1992. The data-driven model is forced by the solar activity and CO₂ concentration signals. We find the persistent antiphasing relationship between the solar forcing and Niño-3.4 SST on the bicentennial time scale. The dynamical mechanism of such a response is discussed.

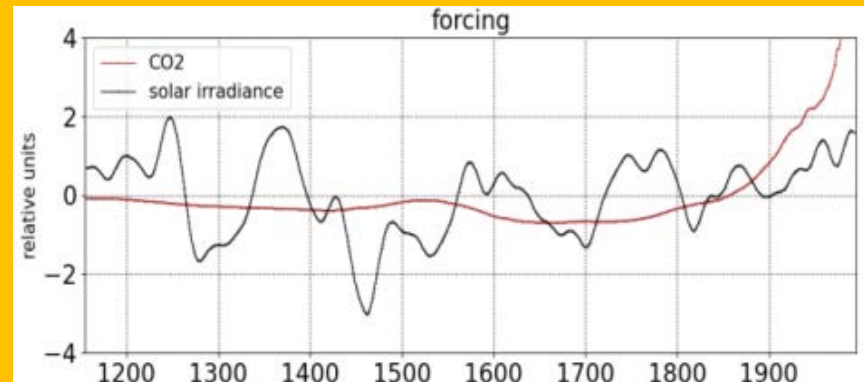
The work was supported by the Russian Science Foundation (Grant No. 20-62-46056)

¹ **Data:** yearly Niño-3.4 indices reconstructed from paleoclimate proxies based on three different sea-surface temperature databases



1. Emile-Geay, J., Cobb, K. M., Mann, M. E., & Wittenberg, A. T. (2013). Estimating Central Equatorial Pacific SST Variability over the Past Millennium. Part II: Reconstructions and Implications, *Journal of Climate*, 26(7), 2329-2352.
2. Steinhilber, F., J. Beer, and C. Frohlich (2009), Total solar irradiance during the Holocene, *Geophys. Res. Lett.*, 36, L19704, doi:10.1029/2009GL040142.
3. Malte Meinshausen, Elisabeth Vogel, Alexander Nauels, Katja Lorbacher, Nicolai Meinshausen, David Etheridge, Paul Fraser, Stephen A. Montzka, Peter Rayner, Cathy Trudinger, Paul Krummel, Urs Beyerle, Josep G. Canadell, John S. Daniel, Ian Enting, Rachel M. Law, Simon O'Doherty, Ron G. Prinn, Stefan Reimann, Mauro Rubino, Martin K. Vollmer, Ray Weiss (2016), Historical greenhouse gas concentrations, doi:10.5194/gmd-2016-169

^{2,3} **Forcing:** solar irradiance and air CO₂ concentration signals

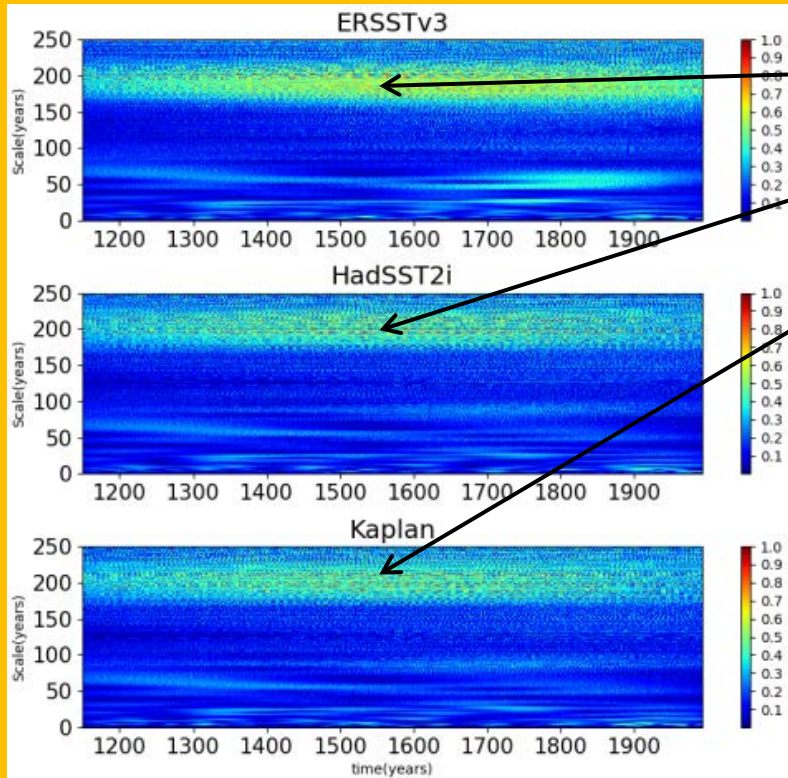


Data/forcing preparation and preprocessing

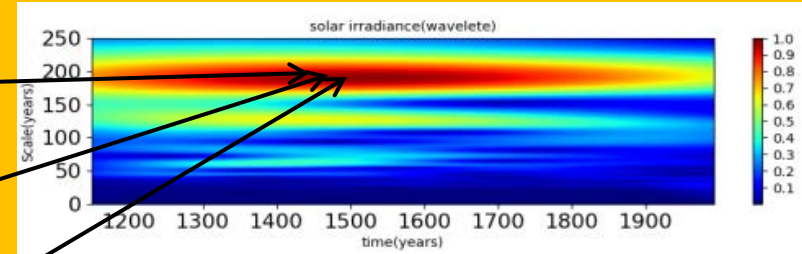
- Data and forcing are considered on the time interval from 1150 to 1992 (843-year series)
- Monthly air CO₂ concentration signal and signal of solar irradiance with 5-year step are resampled with 1-year step using spline interpolation

Wavelet analysis of the Niño-3.4 and solar irradiance signal

Niño-3.4 wavelets



Solar irradiance wavelet



- The 200-year solar activity cycle is well-documented:
Wagner, G., and Coauthors, 2001: Presence of the solar De Vries cycle (205 years) during the last Ice Age. *Geophys. Res. Lett.*, 28, 303–306.
- The distinct spectral peak near the 200-year period in all considered **Niño-3.4** data is observed. Is it a manifestation of solar irradiance?
- To clarify this question we construct the data-driven model

1 Nonlinear stochastic model

$$x_n = F(x_{n-1}, \dots, x_{n-L}, \mathbf{f}_n) + g \xi_n$$

- Niño-3.4 data: $\mathbf{X} = (x_1, \dots, x_N)$
- Two-dimensional forcing including both air CO2 concentration and solar irradiance signal: $\mathbf{f}_1, \dots, \mathbf{f}_N$
- Deterministic part F is parameterized via one-layer perceptron:

$$P^m(\mathbf{u}_n) = \sum_{i=1}^m \alpha_i \cdot \tanh(\boldsymbol{\omega}_i^T \cdot \mathbf{u}_n + \gamma_i), \mathbf{u}_n = (x_{n-1}, \dots, x_{n-L}, \mathbf{f}_n)$$

- L is memory depth, m is number of neurons

2 Bayesian optimality criterion

$\boldsymbol{\mu} = (\alpha, \gamma, \boldsymbol{\omega}, g)$ – model parameters

$\mathbf{H}_i = (L, m)$ – model hyperparameters



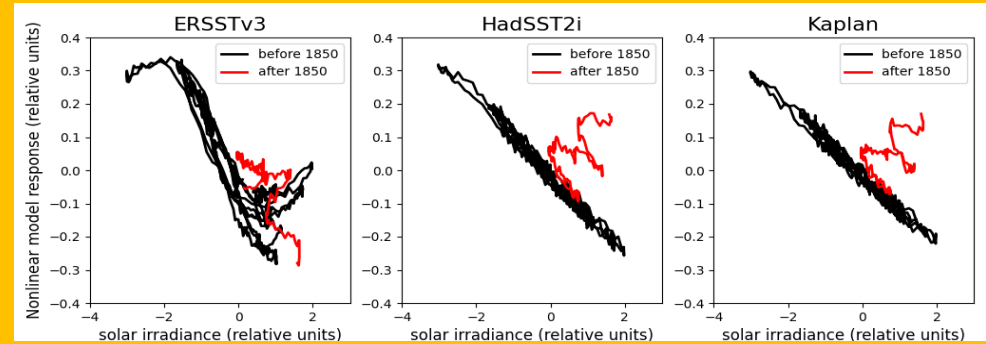
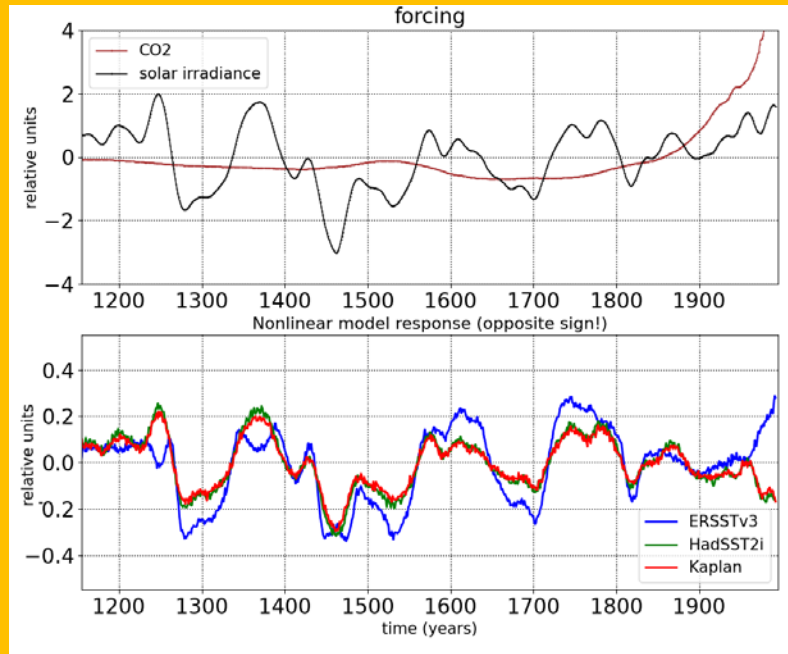
$P(\boldsymbol{\mu}, \mathbf{H}_i | \mathbf{X}) \propto P(\mathbf{X} | \boldsymbol{\mu}, \mathbf{H}_i) \cdot P_{\mathbf{H}_i}(\boldsymbol{\mu})$ – cost function for $\boldsymbol{\mu}$ (posterior PDF of the data $\mathbf{X}$ was produced by the model with $\boldsymbol{\mu}$ and $\mathbf{H}_i$)

$P(\mathbf{X} | \mathbf{H}_i) = \int P(\mathbf{X} | \boldsymbol{\mu}, \mathbf{H}_i) \cdot P_{\mathbf{H}_i}(\boldsymbol{\mu}) d\boldsymbol{\mu}$ – cost function for $\mathbf{H}_i$ (Bayesian evidence)

Optimal $\boldsymbol{\mu}$ and $\mathbf{H}_i$ corresponds to the maximum of the Bayesian evidence (please see Gavrilov et al. 2017 for more technical details)

1. Molkov Y. I., Loskutov E. M., Mukhin D. N., Feigin A. M. Random dynamical models from time series // Physical Review E. 2012. V. 85, №. 3. P. 036216.
2. Gavrilov A.S., Loskutov E.M., Mukhin D.N. Bayesian optimization of empirical model with state-dependent stochastic forcing // Chaos, Solitons and Fractals. 2017. V. 104. P. 327–337.

Results: solar irradiance vs model response to forcing (mean model output obtained by averaging 10000 stochastic realizations)

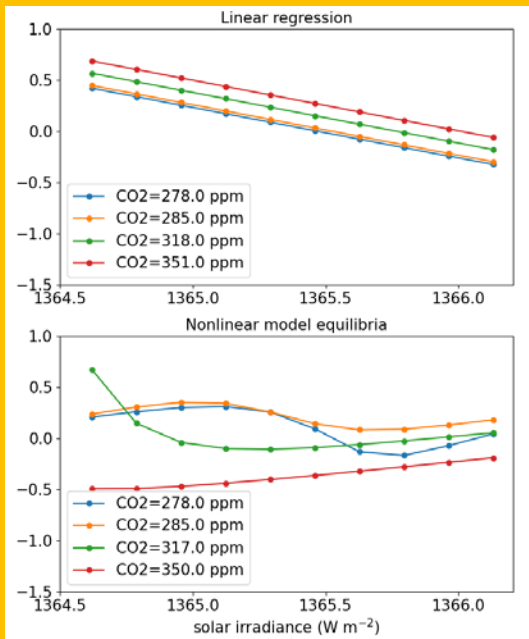


Main points:

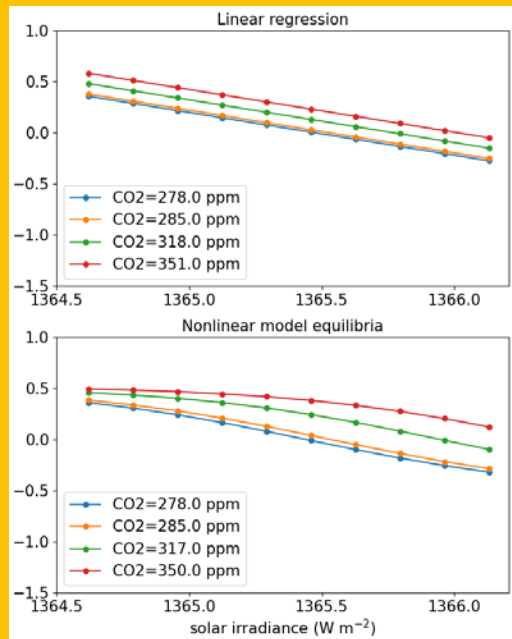
- According to the Bayesian criterion, the forced data-driven model is more optimal than the unforced one
- Optimal model hyperparameters (complexity) are not equivalent for different Niño-3.4 data ($L=3, m=3$ for ERSSTv3 and $L=2, m=1$ for HadSST2i and Kaplan)
- For all Niño-3.4 data, model response trend is opposite (antiphasing) to solar irradiance
- Model response is clearly nonlinear for Niño-3.4 data based on ERSSTv3
- Model response trend is changing after 1850 for all Niño-3.4 data

Results: equilibrium points of deterministic part of the model in depends on forcing

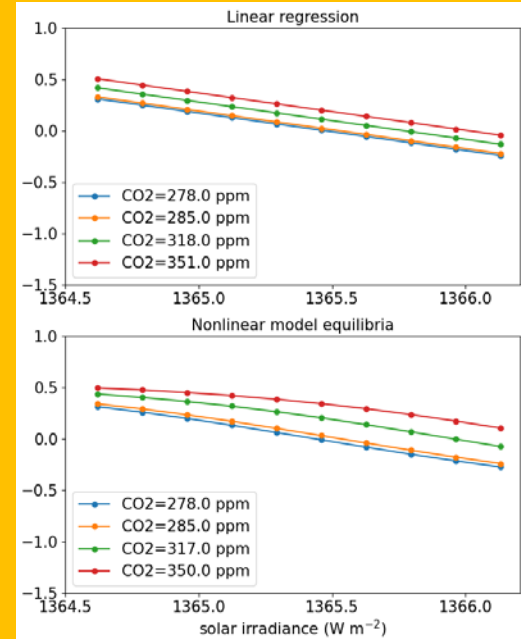
ERSSTv3



HadSST2i



Kaplan



Main points:

- For all Niño-3.4 data deterministic part F of the optimal model has a single stable equilibrium point (focus or node) in whole range of forcing values
- Linear regression model describes an additive contribution of forcing to Niño-3.4 data: $x_n = \hat{A} \cdot \mathbf{f}_n + b$
- For Niño-3.4 data based on ERSSTv3, equilibrium points of deterministic part of the model depend on forcing nonlinearly
- For Niño-3.4 based on HadSST2i and Kaplan data, the dependencies of equilibrium points of deterministic part of the model are close to linear regression

Main points

For all Niño-3.4 data:

- According to the Bayesian optimality criterion, including both the solar activity and CO2 concentration signals in the data-driven model is justified
- Model demonstrates persistent antiphasing relationship between solar forcing and Niño-3.4 on the bicentennial time scale

But!

- For **Niño-3.4** based on **ERSSTv3** data, nonlinear forcing influence on the model dynamics exists
- For **Niño-3.4** based on **HadSST2i** and **Kaplan** data, forcing contribution to the model dynamics is actually additive
 - Further analysis of the data uncertainties is required