

The potential of machine learning for modeling spatio-temporal properties of water isotopologue distributions in precipitation

Kira Rehfeld¹³, Jonathan Wider¹³, Nadine Theisen¹³, Martin Werner², Ullrich Köthe³, Nils Weitzel¹

¹Institute of Environmental Physics, Ruprecht-Karls-Universität Heidelberg, Heidelberg, Germany

²Alfred-Wegener-Institut Helmholtz-Zentrum für Polar- und Meeresforschung, Bremerhaven, Germany

³Interdisciplinary Center for Scientific Computing, Ruprecht-Karls-Universität Heidelberg, Heidelberg, Germany



UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386



PAL
MOD

GERMAN
CLIMATE
MODELING
INITIATIVE

Contact:

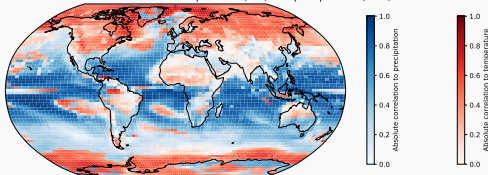
krehfeld@iup.uni-heidelberg.de

jwider@iup.uni-heidelberg.de

Motivation/Background

- Simulating stable water isotopologues within a climate model allows validating the model's hydrological cycle with proxy records. Isotope-enabled general circulation models (iGCMs) were developed for this purpose (compare e.g. Tindall, Valdes, and Sime, 2009, Werner et al., 2016)
- But only few iGCMs exist. Additionally, if isotopologue tracers are included, the simulation runtime increases (by a factor of about 2 to 3 for the HadCM3 iGCM).
- There exist correlations between isotopologues and other climate variables (Dansgaard, 1964). "Amount effect" with precipitation in tropics, "temperature effect" in polar regions.

Strongest absolute correlation with $\delta^{18}\text{O}$: Temperature (red) and precipitation (blue)



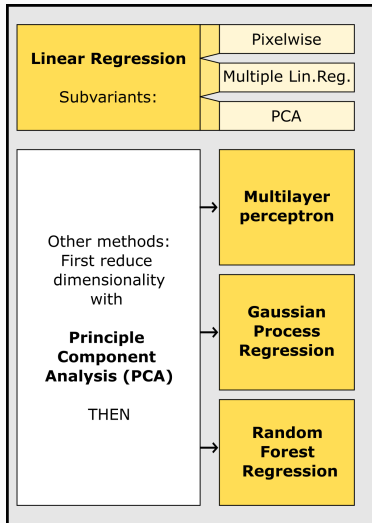
Plot created with yearly averages from iHadCM3 last millenium run (compare Bühler et al., 2020)

Idea & Description of Case Study

- Correlations to other climate variables $\Rightarrow$ **Idea: Predict isotope abundancies from other more commonly simulated climate variables.**
- Case study: Interannual data (annual averages) from a past millenium run (850 CE to 1850 CE) with the HadCM3 iGCM (Bühler et al., 2020).
- For each grid cell: Predict the abundance of a specific isotope ($\delta^{18}O$) in precipitation from annually averaged surface temperature, precipitation amount and surface pressure.
- Split the data set into a training set (900 years) and a test set (100 years). If not mentioned otherwise the sets are selected randomly (results do not differ qualitatively if sets are time-ordered).

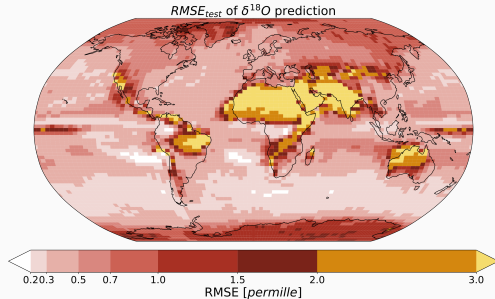
Methods

- Test several well-established machine learning algorithms (compare figure).
- Data is highdimensional (96×73 gridcells, 1000 timesteps)
⇒ Principle component analysis to reduce spatial dimensionality
- Test different choices for the predictor variables.
- Python libraries used:
 - Multilayer perceptrons:  Keras
 - All other methods: 



For theoretical details on the ML-methods used, see for example:

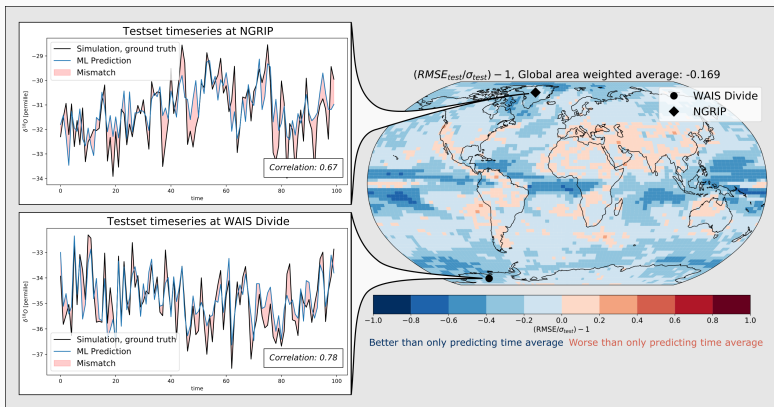
Multilayer perceptrons: Hastie, Tibshirani, and Friedman, 2001, chapter 11 **Gaussian processes:** Rasmussen and Williams, 2006 **Random Forest Regression:** Hastie, Tibshirani, and Friedman, 2001, chapter 15



- To measure the quality of the prediction we compute the $RMSE$ at each gridbox (for an example see above).
- Because $RMSE$ varies spatially (as does $\delta^{18}O$ variance σ in the simulation): Compute $\frac{RMSE_{test}}{\sigma_{test}} - 1$.
- This allows comparing to climatology (simply predicting each cell's time average at every timestep). Negative (positive) values show better (worse) predictions than baseline (see figure on next slide).

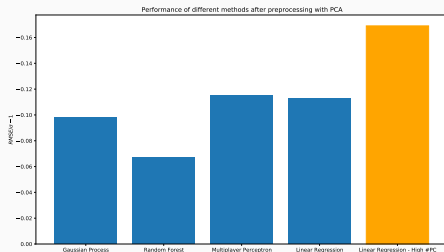
Results from best tested method - Spatial patterns

- Prediction accuracy varies spatially. Very good predictions over oceans in tropics (using "amount effect" and strong correlation with precipitation), but also in polar regions. Worse over continents.
- Included are time series at ice core drilling sites in Greenland (NGRIP) and the Antarctic (WAIS-Divide).



Results - Comparing methods

- PCA leads to big improvements when using linear regression and makes other methods computationally feasible.
- Best combination of predictors: Precipitation and surface temperature (no further gain, when adding surface pressure).
- When comparing methods while using a modest number of principle components (20 for each climate variable), many methods perform approximately equally well (blue).
- Increasing number of principle components (PCs) by an order of magnitude $\Rightarrow$ improved performance for linear regression (orange, "best method").
- Did not test other methods in this regime yet.



Conclusions and outlook

- Machine learning might open a pathway to make many (paleo-)climate simulations available for validation with isotope data.
- First results look promising in regions with ice core drilling programs.
- Important questions for further research:
 - State dependency: How well do the machine learning methods trained on one climatic setup (here: last millenium) perform in other situations?
 - Timescale dependency: How does the predictability vary on different timescales? Can we make use of temporal (auto-)correlations on shorter timescales?

References

- [1] J. C. Bühler et al. "Comparison of the oxygen isotope signatures in speleothem records and iHadCM3 model simulations for the last millennium". In: Climate of the Past Discussions 2020 (2020), pp. 1–30. doi: 10.5194/cp-2020-121.
- [2] W. Dansgaard. "Stable isotopes in precipitation". In: Tellus 16.4 (1964), pp. 436–468. doi: 10.3402/tellusa.v16i4.8993. eprint: <https://doi.org/10.3402/tellusa.v16i4.8993>.
- [3] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. The Elements of Statistical Learning. Springer Series in Statistics. New York, NY, USA: Springer New York Inc., 2001.
- [4] CE. Rasmussen and CKI. Williams. Gaussian Processes for Machine Learning. Adaptive Computation and Machine Learning. Cambridge, MA, USA: MIT Press, Jan. 2006, p. 248.
- [5] J. C. Tindall, P. J. Valdes, and L. C. Sime. "Stable water isotopes in HadCM3: Isotopic signature of El Niño–Southern Oscillation and the tropical amount effect". In: Journal of Geophysical Research: Atmospheres 114.D4 (2009). doi: <https://doi.org/10.1029/2008JD010825>. eprint: <https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2008JD010825>.
- [6] M. Werner et al. "Glacial–interglacial changes in H₂¹⁸O, HDO and deuterium excess – results from the fully coupled ECHAM5/MPI-OM Earth system model". In: Geoscientific Model Development 9.2 (2016), pp. 647–670. doi: 10.5194/gmd-9-647-2016.