The road to assimilating climate hindcasts into paleo-climate reanalyses

Updates about the roadtrip here:

UNIVERSITÄT BERN

OESCHGER CENTRE CLIMATE CHANGE RESEARCH

Jörg Franke[1,2], **Martin Wegmann[1,2]**, Lorenz Hilfiker[1,2] & Stefan Brönnimann[1,2]

1 Institute of Geography, University of Bern *2 Oeschger Centre for Climate Change Research, University of Bern*

TAKE-HOME MESSAGE

- **We thought it would be exciting to try and assimilate ML-based hindcasts into future versions of our paleo-climate reanalysis.**
- **Why? Because this has the potential to decrease reanalysis uncertainties in remote regions and cold seasons.**
- **So we produced some hindcasts with a simple CNN trained with model data only and they seem to point to the right climate anomalies.**
- **But actually assimilating them will be a big challenge and we do not know yet if they indeed decrease reanalysis uncertainty.**

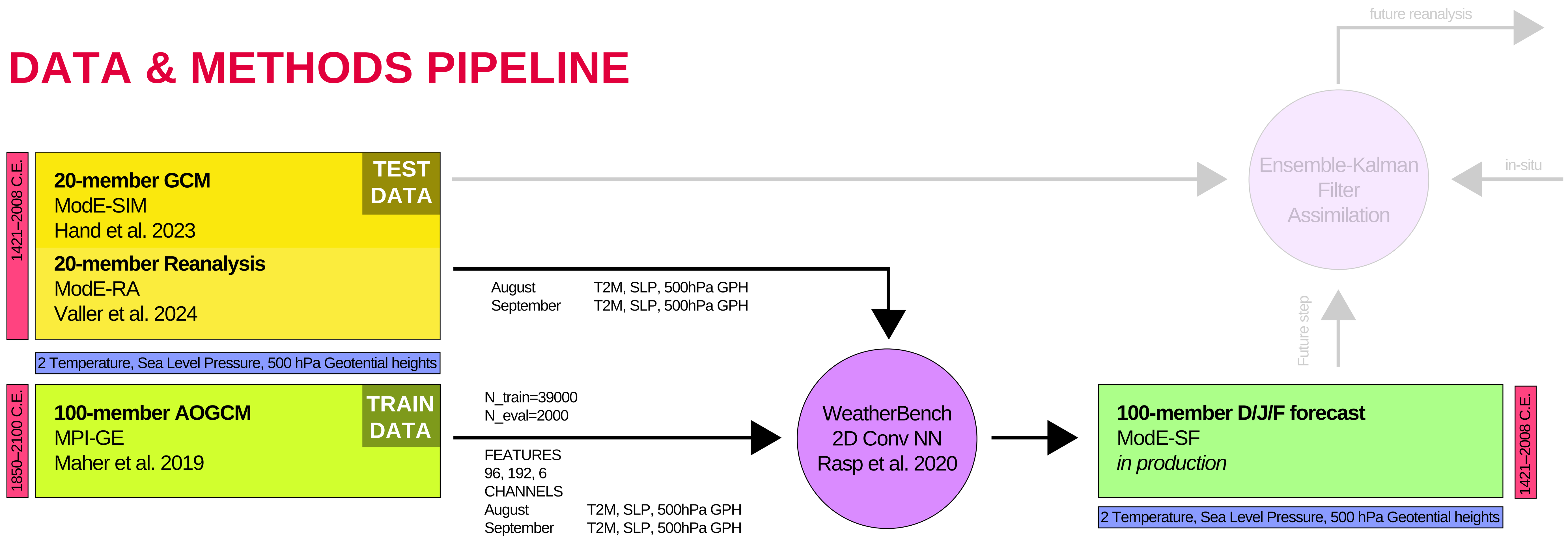
PROBLEM & GAP

- To understand long-term climate variability, analysing longer time periods than what is covered by modern instruments is necessary.
- However, to do so, global climate datasets with a high temporal resolution are required. For this, paleo-climate reanalyses are an essential solution.
- State-of-the-Art paleo-climate reanalyses assimilate observations, proxies & documentary data. Unfortunately, this information is not evenly distributed.
- Climate information density about the past is skewed towards a) warm seasons and b) regions with higher population densities.
- This introduces uncertainties regarding our knowledge of past climates.

A CREATIVE SOLUTION?

- Given that we trust the information in warm seasons more, can we leverage that fact and infer winter information from this global, gridded, multi-variable initial state?
- Here we hypothesize that global, gridded machine-learning seasonal forecasts for boreal winter carry more „useful" climate information for assimilation than the baseline background GCM climate/climatology field for winter.
- Assimilating those forecasts would solve two problems: 1) cover regions of low information density and 2) cover seasons of low information density.

DATA & METHODS PIPELINE



FIRST RESULTS

- Even though we lack a ground truth for most of the time period, we would hope that hindcasts show the same deviations from the GCM for each time step than the assimilation/reanalysis does.
- This gives us a baseline trust for the hindcast quality. If the hindcast point to completely opposite anomalies, that would be suspicious.
- That said, both products could produce similar „unreal" anomalies. Moreover, it is hard to pinpoint if anti-correlation in anomalies points towards an issue in the reanalysis or the hindcast.

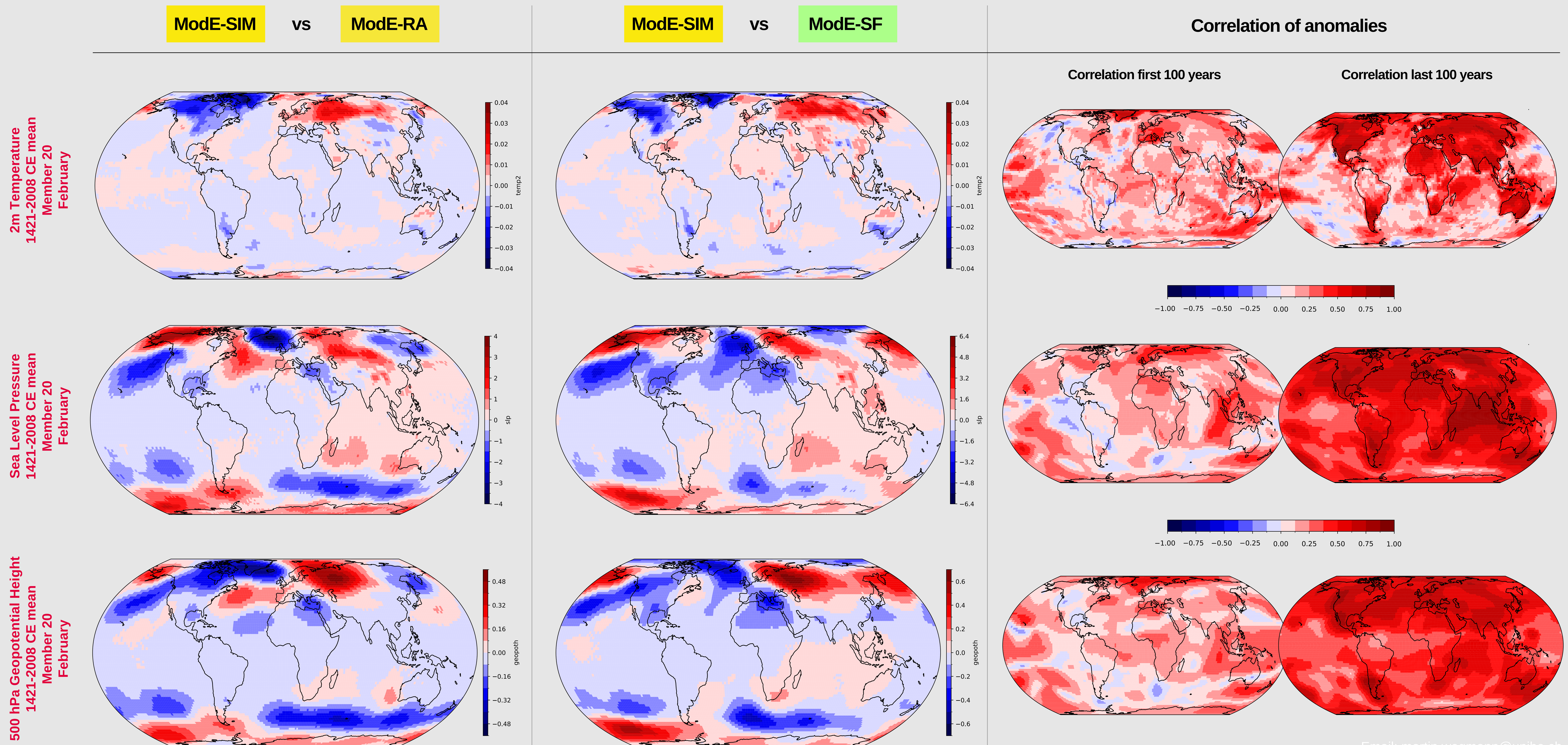
To the right you **find diagnostics for 1 out of 20 members in the ModE family and one forecast month, which highlights the above points. Similar results can be found for each ensemble member.**

We are happy if:

- Structure and magnitude of the anomalies look similar between ModE-SIM vs ModE-RA and ModE-SIM vs ModE-SF
- The correlation of both anomaly types over time is significantly positive. Improvement of this correlation over time is also expected, since initial conditions of the hindcasts are improving over time.

What we still do not know:

- Where and when are regions and episodes of opportunity for the hindcasts to provide value in an assimilation. For example: Boundary condition SSTs in the 20th century are more trustworthy than 2m temperature hindcasts, but maybe not so much in the 15th century.
- How to avoid conflict between in-situ data and hindcasts in the assimilation process.
- If in-situ data availability and hindcast skill are complementary. We sure hope they are.
- Which variable of the hindcast is the most robust to assimilate.



Key References

- Rasp, S., Dueben, P. D., Scher, S., Weyn, J. A., Mouatadid, S., & Thuerey, N. (2020). WeatherBench: a benchmark data set for data-driven weather forecasting. Journal of Advances in Modeling Earth Systems, 12(11), e2020MS002203
- Hand, R., Samakinwa, E., Lipfert, L., & Brönnimann, S. (2023). ModE-Sim–a medium-sized atmospheric general circulation model (AGCM) ensemble to study climate variability during the modern era (1420 to 2009). Geoscientific Model Development, 16(16), 4853-4866
- Valler, V., Franke, J., Brugnara, Y., Samakinwa, E., Hand, R., Lundstad, E., ... & Brönnimann, S. (2024). ModE-RA: a global monthly paleo-reanalysis of the modern era 1421 to 2008. Scientific Data, 11(1), 36.

Email: martin.wegmann@unibe.ch