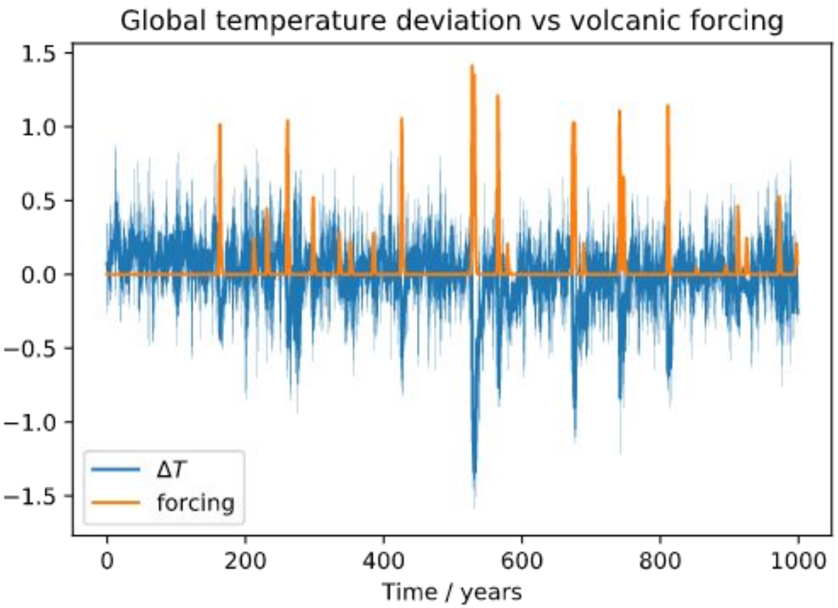
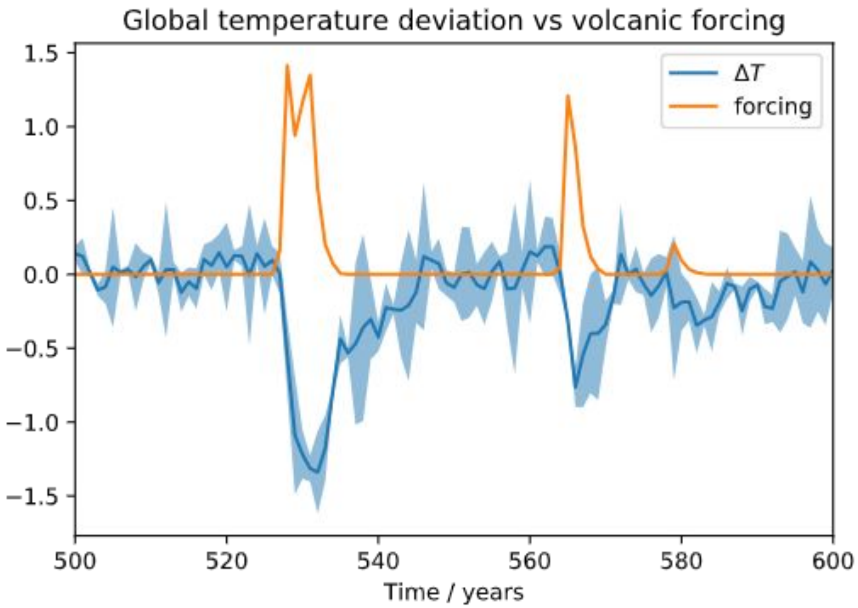
HIDA Remote Challenge 2020

Team Bayes

HiDA | HELMHOLTZ Information & Data Science Academy

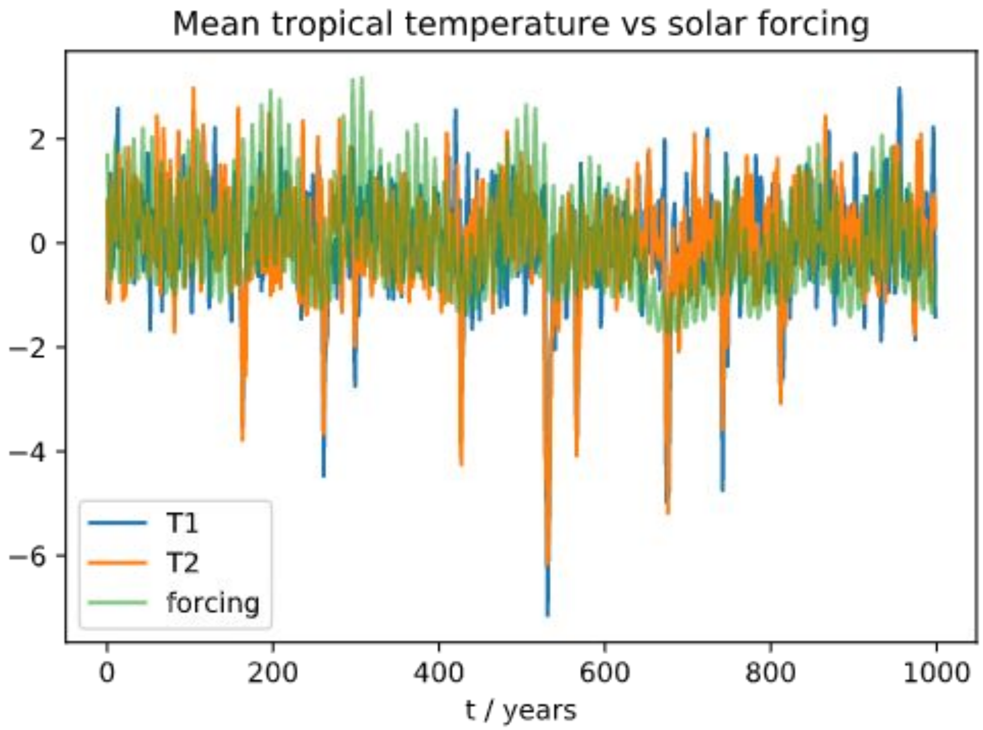
Volcanic forcing



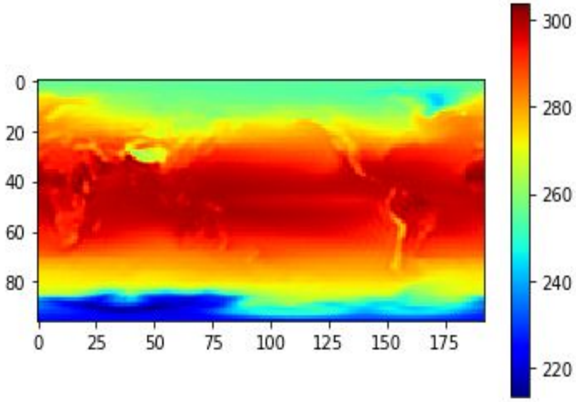


- High correlation
- Short time scale

Solar forcing

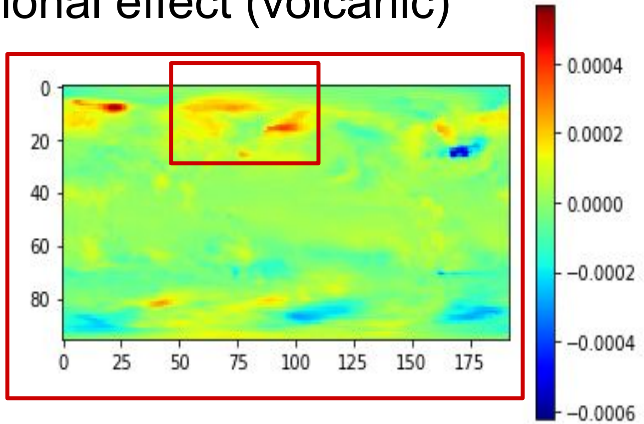


- Low correlation
- Long time scale



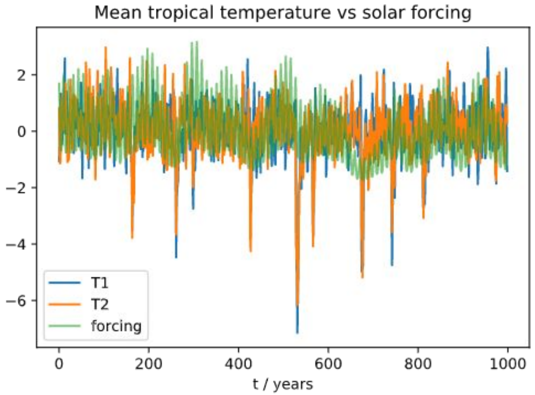
Renormalize data

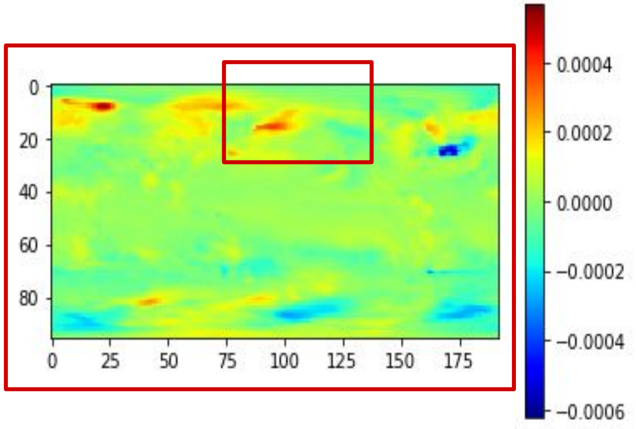
regional effect (volcanic)

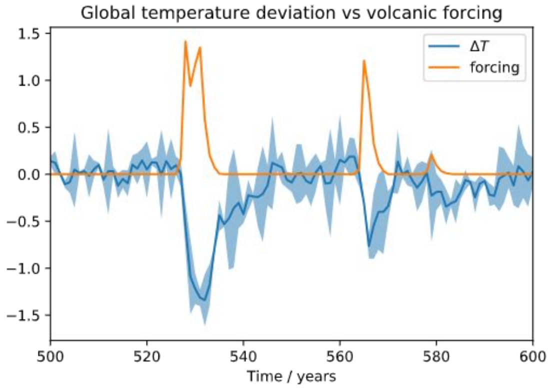


global effect (solar)

Solar and volcanic forcing







SOLAR

-Low correlation

-Long time scale

-Global effect

VOLCANIC

-High correlation

-Short time scale

-Regional effect

Why Bayes?

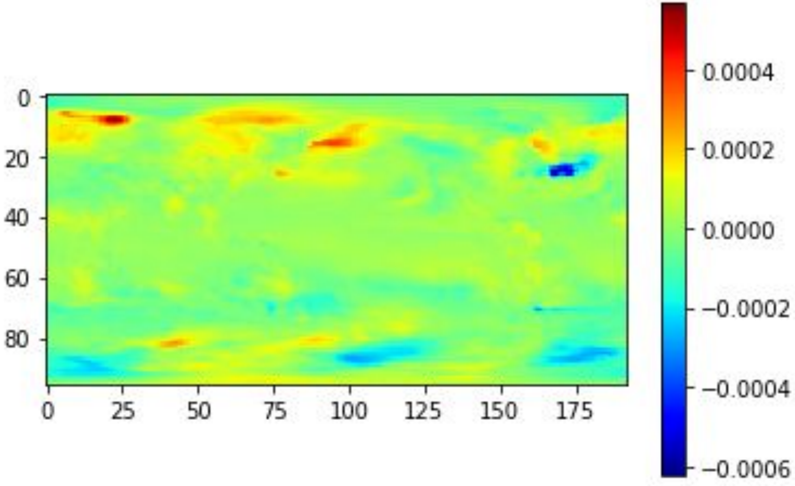
1) **LITTLE** data (just 2 observations)

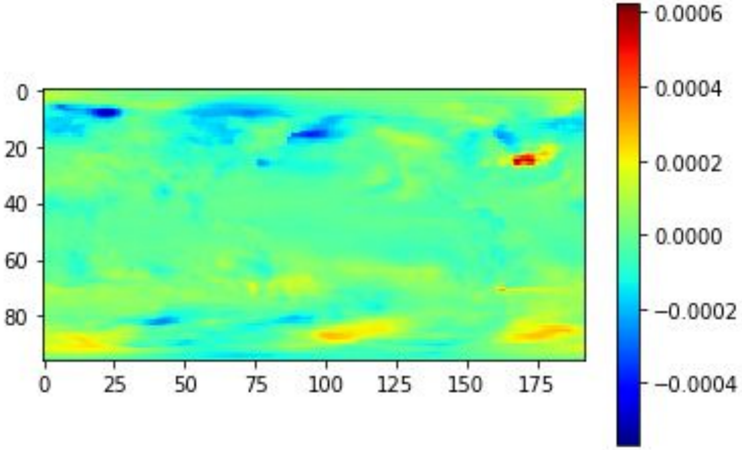
2) Include a-priori-knowledge
   a) forcing effects somewhat smooth but random
   b) different effect scales

3) Quantify uncertainties (they are huge)

4) Quantify evidence for solar forcing (it is little)

*mathematical details super easy, see mathematical appendix B: https://www.overleaf.com/read/fpdhmqjxwhqt





The Approach in a Nutshell

1) each simulation is a random field sample
2) find spatial correlation lengths
3) track spatial correlation lengths over time

track spatial correlation length over time

CHRIS' PLOT HIER

Questions for Problem 1

a) Has there been volcanic forcing last year ?

   i) Linear Relation of SQRT(ADO(t-1)) and $\Delta T(t)$
   ii) Fit coarse behaviour of $\Delta T(t)$ over longer time-scales
   iii) Bayesian model comparison if *2. alone* is enough or *1.+2.* are needed

   Answer: ingredients with GP and fit available

b) Was solar forcing higher or lower than average?
   or even: is solar forcing even detectable in data?

   Answer: we didn't find any usable correlation

Questions for Problem 2

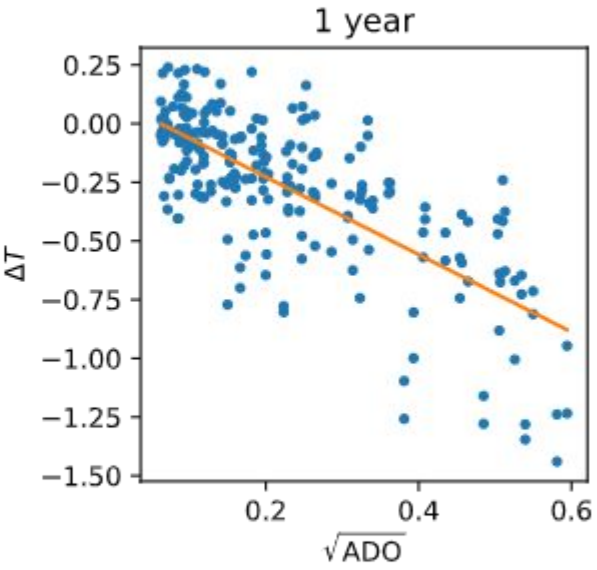
a) How do length-scales and noise look over time?

   Answer: look at GP regression over time (earlier slides).

b) What is different between tropical and polar regions?

   Answer: at least for volcano data, both regions were usable. In GP we see finer spatial scales near poles.

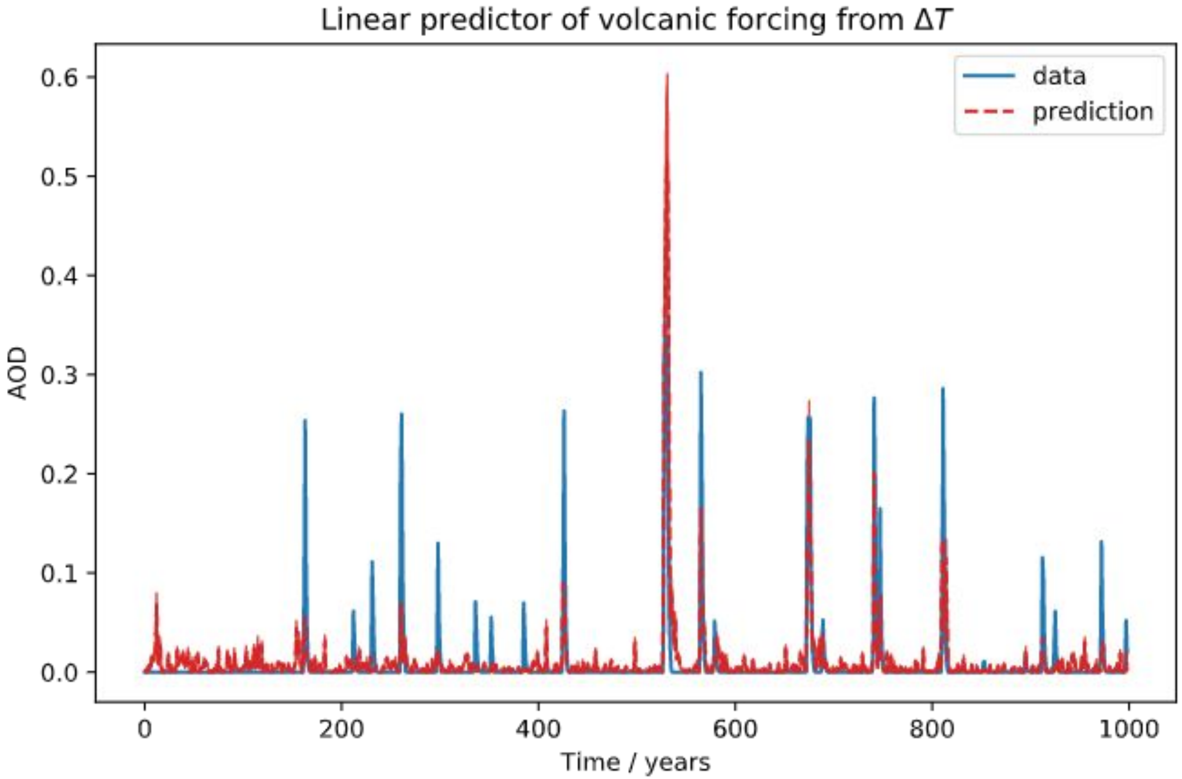
Fit of ΔT(t) in terms of ADO(t-1)



$$\Delta T(t) = -1.65\sqrt{\mathrm{ADO(t-1)}}$$

Std. Err. in coefficient: 0.1

Can also reconstruct volcano data from ΔT



Conclusion

- Model heterogeneity with Gaussian random fields
- Bayes gives uncertainties and model comparison
- solution lies in the scales
- implementation not trivial (not enough time)

APPENDIX A
Exploratory Data Analysis

Motivation of Approach

Not BIG DATA, but LITTLE DATA (only 2 simulations!)

- Unsupervised (essentially)
- Temperature field:
  - “Somewhat random but somehow smooth”
  - Spatial and temporal correlations
  - Model as Gaussian Random Field

BAYESIAN PROBABILITY THEORY

- “Little data theory”
- ONLY consistent and rigorous theory (mathematical proof)
- Quantifies uncertainties
- Includes a-priori-knowledge
  - smoothness, spatiotemporal correlations

The Approach in a Nutshell

1) ch year is a Gaussian Process sample (2 samples per year)
2) : different magnitudes and spatiotemporal scales
agnitudes not distinguishable
cales very different

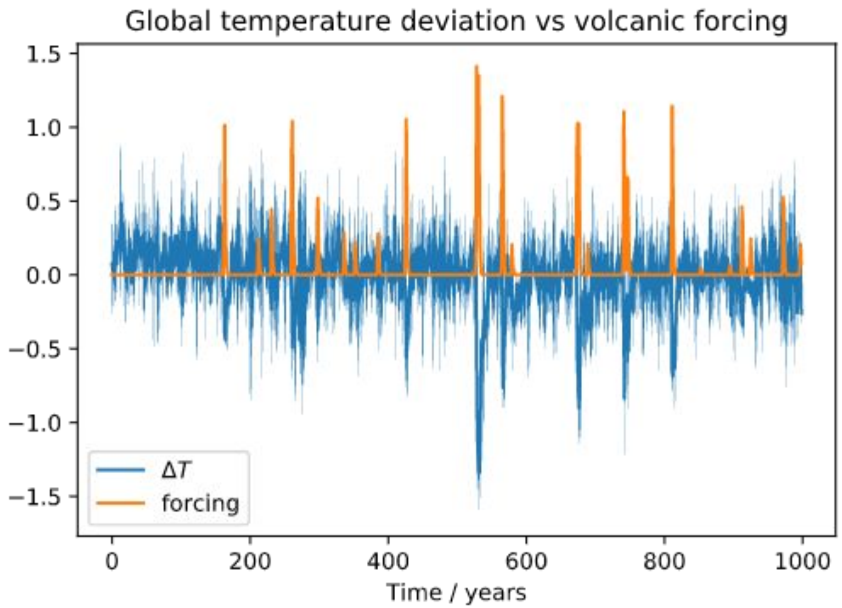
3) on for each year locally, i.e.
raw rectangle over region
stimate correlation parameters
ot correlation parameters over time
nall scale correlations indicate volcanic forcing
irge scale correlations indicate solar forcing

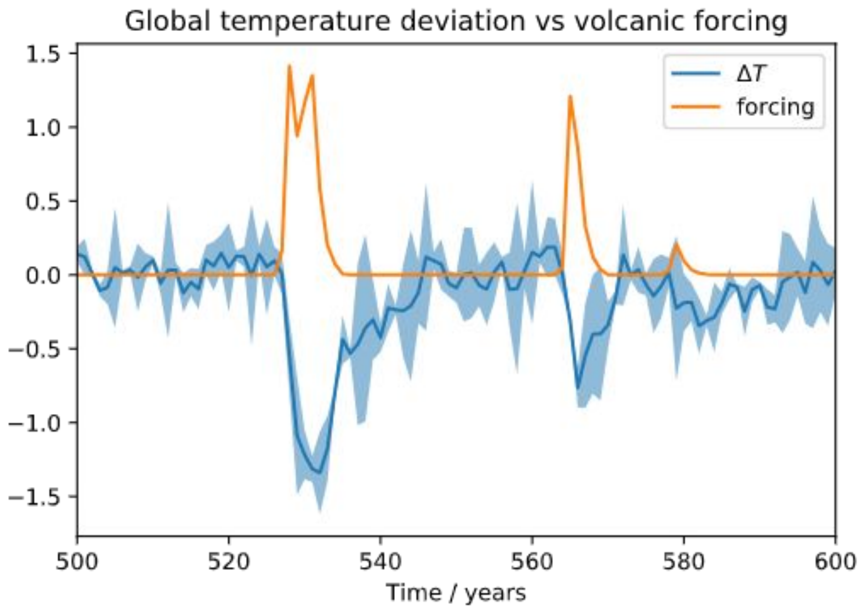
omparison:
Evidence for volcanic/solar forcing?
=
Evidence for small/large-scale correlations?

Problem 1

Can you find out **which periods** were **dominated by climate forcing** without knowing the exact dates of volcanic events or low solar activity? We know the solution to this question because we can match our simulations with time series data for changes in solar activity and volcanic outbreaks at the time, but we would like to be able to tell **just from the model**.

Effect of volcanic forcing

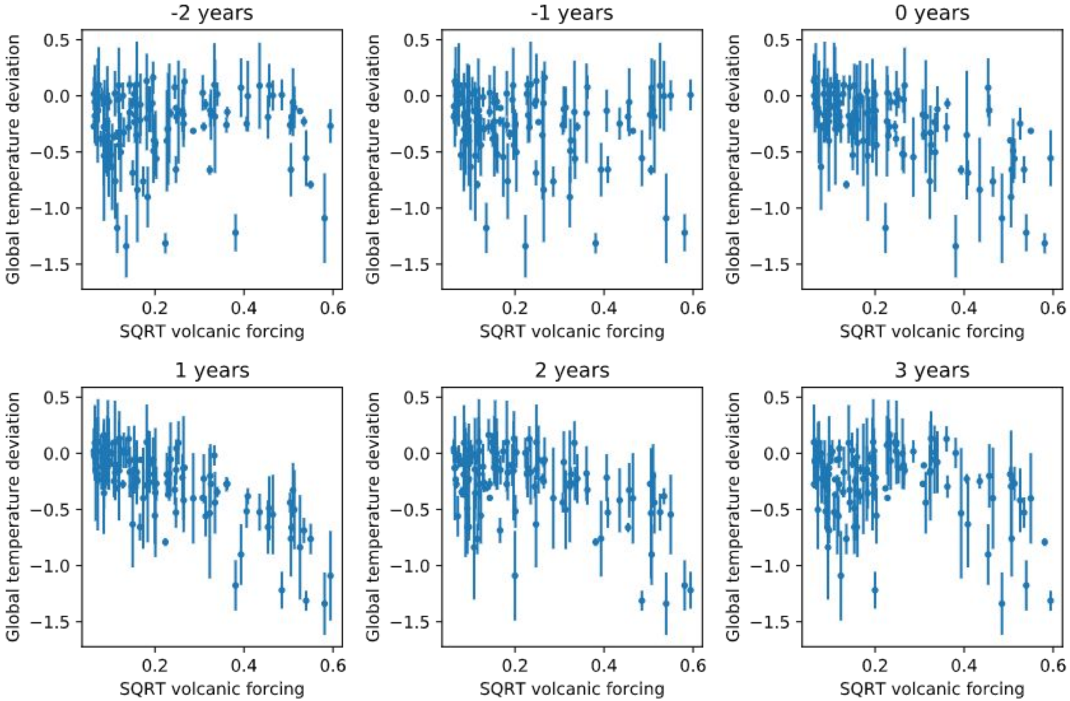




Lagged temperature response

Global temperature deviation

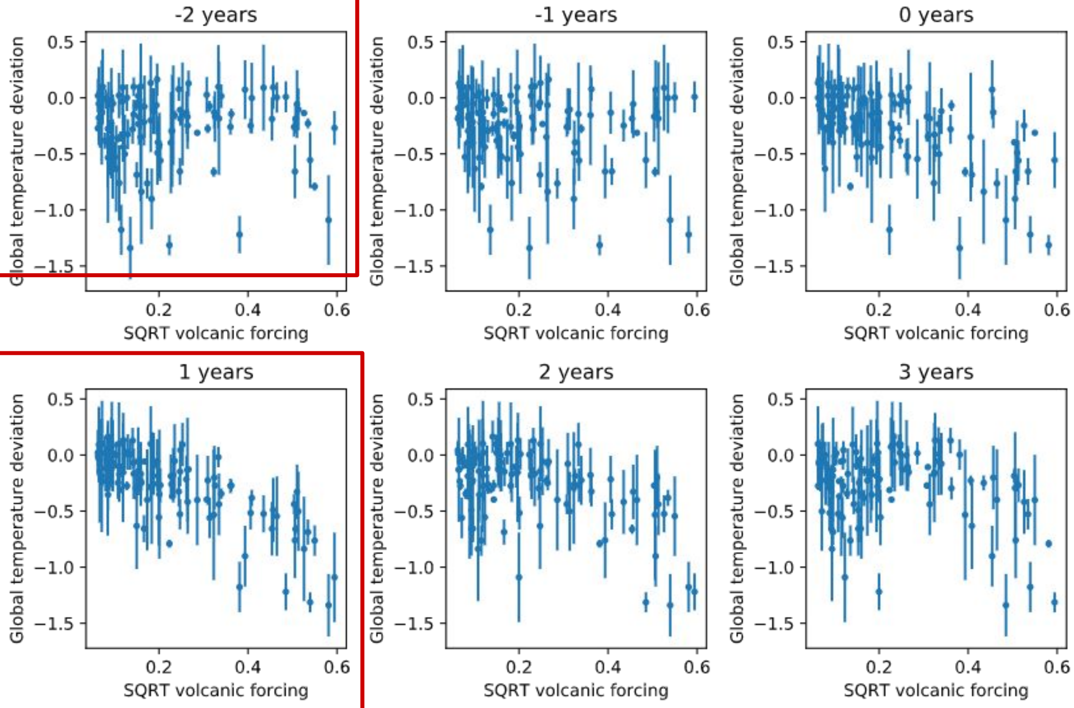
X years after **volcanic forcing**



Lagged temperature response

Global temperature deviation

X years after **volcanic forcing**



Conclusion:

Need only to look at lag of

1 year

(supported by statistical time-series analysis)

Questions

a) **Has there been volcanic forcing last year ?**

Has there been volcanic forcing (ADO) last year?

1. Fit relation between ADO(t-1) and ΔT(t)

2. Fit coarse behaviour of ΔT(t) over longer time-scales

3. Bayesian model comparison if *2. alone* is enough or *1.+2.* are needed

Has there been volcanic forcing (ADO) last year?

1. **Fit relation between ADO(t-1) and ΔT(t) (it looks linear in square-root)**

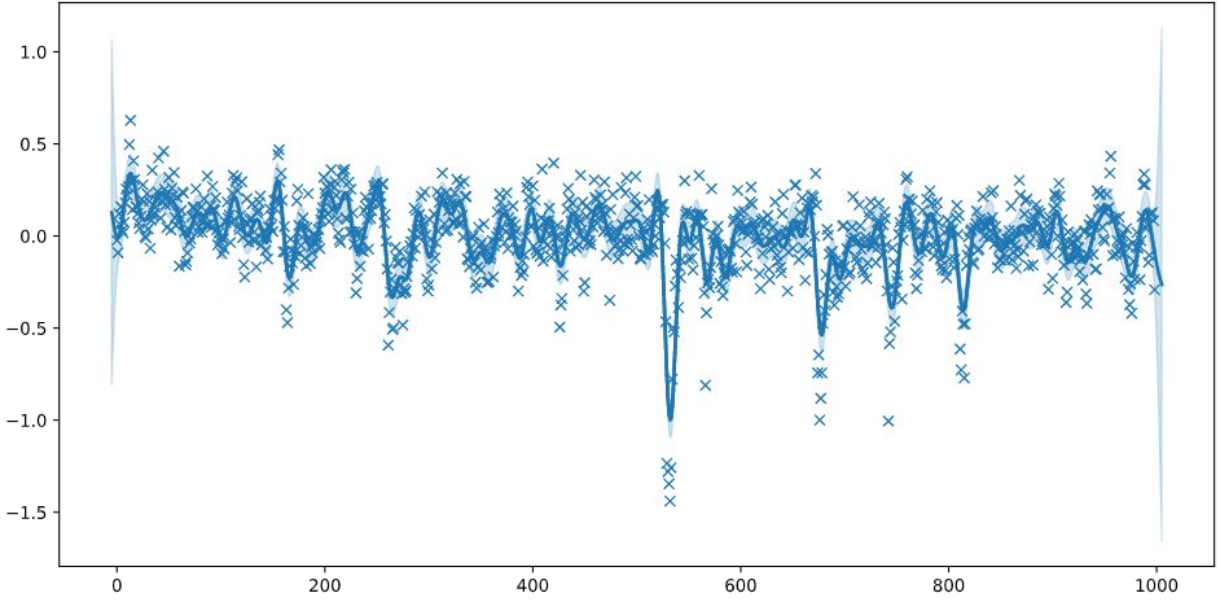
2. Fit coarse behaviour of ΔT(t) over longer time-scales

3. Bayesian model comparison if *2. alone* is enough or *1.+2.* are needed

Tropical vs Polar data

Fit behaviour of ΔT(t) over **several** time-scales

Gaussian process regression with two time-scales yields good fit, but may be too much. With single-timescale we get about the same likelihood and that fit:



Problem 2

Can you tell how the **spatial patterns change** for periods with **strong volcanic activity**? In other words do all regions in the two simulations show the same temperature decline or maybe even increase?

Please consider that we see **greater impact** of factors other than volcanos and solar activity in so called **extratropical** regions, the closer we move towards the poles, north of 30°N and south of 30°S. It would also be interesting to **test** whether **tropical regions** show a **more coherent temperature response** pattern in R1 and R2 simulations compared to extratropical regions.

Questions

a) **How do length-scales and noise look over time?**