

Exploring the Tropospheric Response to Stratospheric Variability Using Lagged Quantile Regression

Kathrin Finke, Abdel Hannachi

1. Objective
2. Quantile Regression
3. Application to Reanalysis and the Stratosphere -Troposphere relation
4. Conclusion

1. Objective

- Stratospheric variability can affect the tropospheric circulation for an extended time period (e.g. Baldwin and Dunkerton (2001))
- Different regimes of the tropospheric circulation (e.g. low and high pressure in a certain area) may be influenced by the stratosphere to a different degree

Statistical tool is needed that is able to capture the impact of the stratosphere on various tropospheric weather patterns

- Standard linear regression methods can only estimate the stratosphere-troposphere relation based on the mean of the tropospheric variables distribution



Quantile Regression (QR)

- ✓ Robust against outliers
- ✓ allows to model relation of stratosphere and various quantiles of the tropospheric response variable

2. Quantile Regression

- QR model for any τ th for i variables (x) that may have an effect on the response variable (y) with a corresponding regression coefficient $\beta(i)$. (Koenker and Bassett, 1978; Koenker and Hallock, 2001)

$$Q_{\tau}(y_i) = \beta_0(\tau) + \beta_1(\tau)x_{i1} + \dots + \beta_p(\tau)x_{ip} \quad i = 1, \dots, n$$

- The stratospheric variable, i.e. the predictor is represented by x and the tropospheric response variable is represented by y .
- Hence, various quantiles of the tropospheric variable conditioned on the stratospheric state can be analysed by using the QR model
- choosing $\tau=0.5$ estimates the median regression slope (similar to linear regression)
- only one variable at a time is used in the present study to avoid overfitting: $i=1$

2. Quantile Regression

- The unknown parameters for a specific quantile are estimated by minimization of the median absolute deviation (MAD).
- ρ_τ is a loss function, giving the asymmetric weights to the error ϵ in dependence of the quantile and the overall sign of the error (Koenker and Bassett, 1978; Koenker and Hallock, 2001)

$$MAD = \frac{1}{n} \sum_{i=1}^n \rho_\tau (y_i - (\beta_0(\tau) + \beta_1(\tau)x_{i1} + \dots + \beta_p(\tau)x_{ip}))$$
$$\rho_\tau(\epsilon) = \tau \max(\epsilon, 0) + (1 - \tau) \max(-\epsilon, 0)$$

- Using MAD makes the method robust against outliers in the distribution

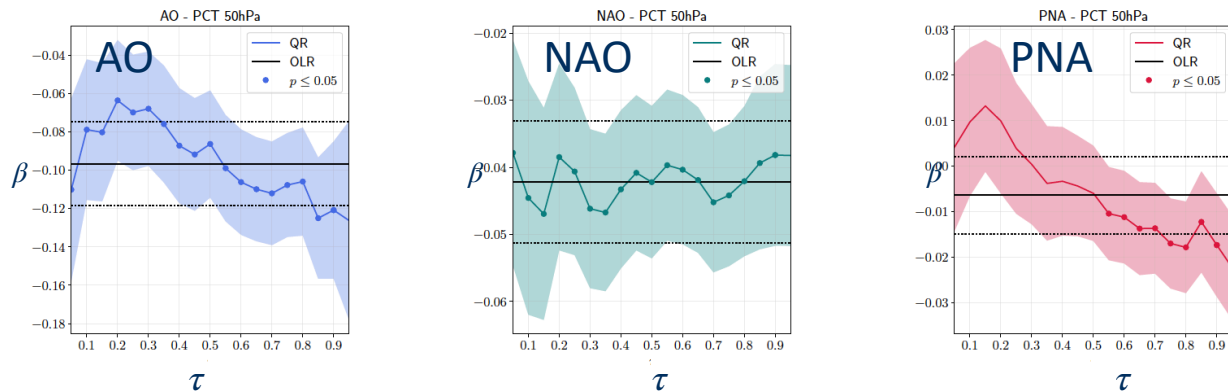
3. Application to reanalysis - Data

- ERA5 data north of 20°N
- Arctic oscillation index (AO), North Atlantic Oscillation index (NAO) and Pacific-North American teleconnection pattern (PNA) index from NOAA
- daily averages from 6-hourly data for December-April 1979-2018
- Fields: T2m, SLP, Z500, T50 on a 1°x1° latitude-longitude grid
- Anomalies are calculated by subtracting the long-term (1979-2018) mean at each grid point
- 5-day running mean is applied to all data

3. Application to reanalysis - Indices

- Stratospheric Index (PCT50)
 - 50hPa, latitude-weighted, polar cap temperature anomalies (averaged over 65°N-90°N)
as proxy for stratospheric polar vortex strength
- Tropospheric Index
 - AO, NAO, PNA index as the tropospheric index
 - each grid point of reanalysis field as the tropospheric index `
- Lagged Quantile regression
 - The lagged relation during DJF is examined by letting the tropospheric index lag the stratospheric one by up to 60 days

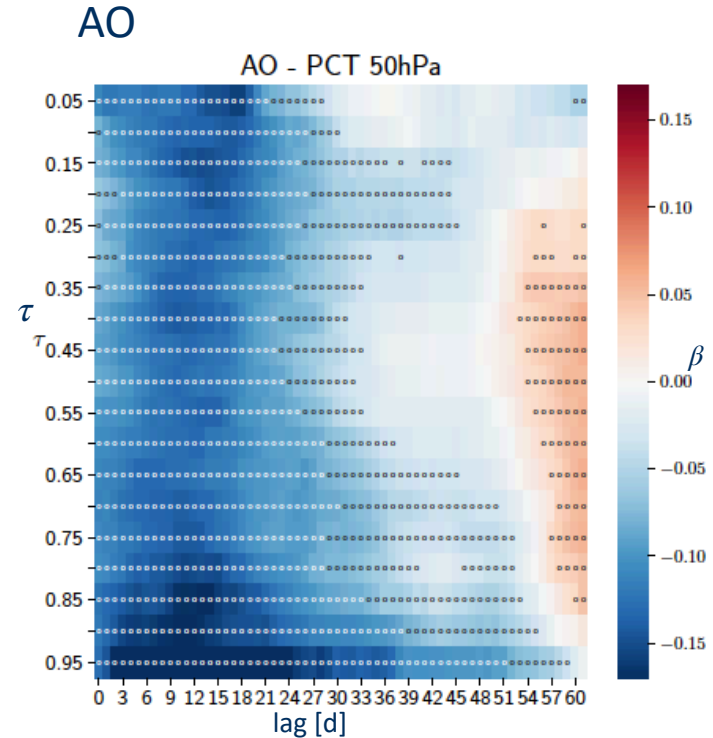
PCT50 vs AO, NAO & PNA (no lag)



- **AO:** significant QR coefficients lie outside of the 95% confidence interval of the linear regression for $\tau=[0.2, 0.25, 0.3]$ and $\tau=[0.85, 0.9, 0.95]$
 - ➔ relation between PCT50 and AO for these quantiles can likely not be modelled with linear regression
 - ➔ relation is strongest during the strongly positive regime of the AO
- **NAO:** All QR coefficients lie within the 95% confidence interval of the linear regression
 - ➔ linear regression is sufficient relative to QR
- **PNA:** significant QR coefficients lie outside of the 95% confidence interval of the linear regression coefficient for $\tau=[0.75, 0.8, 0.9, 0.95]$
 - ➔ relation between PCT50 and PNA for these quantiles can likely not be modelled with linear regression
 - ➔ relation non-significant during negative phase of PNA

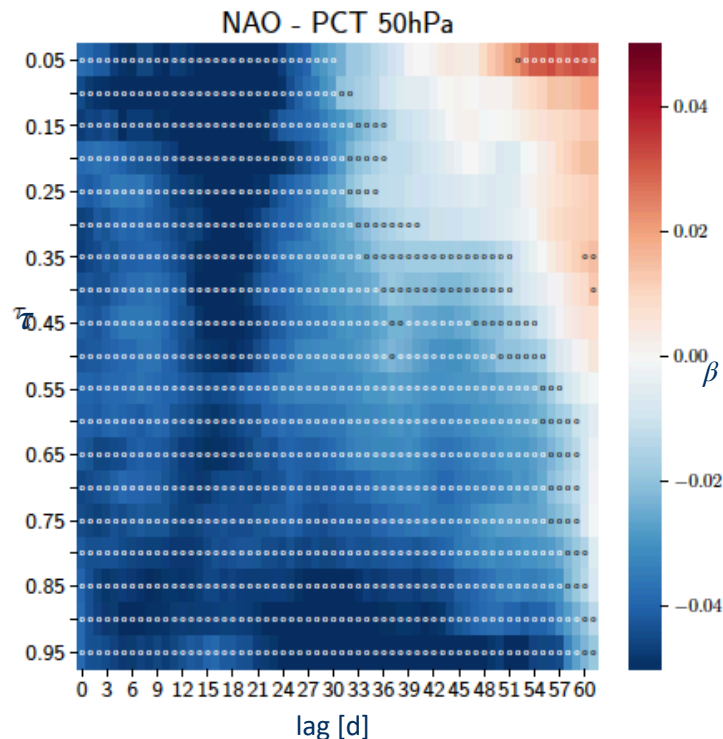
PCT50 vs AO: lag 0-60 days

- ➔ negative AO pattern (low quantiles) shows significant negative relation to PCT50 up until about 30 days, being strongest around a 20 day lag
- ➔ positive AO pattern (high quantiles) shows significant negative relation to PCT50 up until about 60 days which is stronger compared to lower quantile
- ➔ significant positive relation between PCT50 and quantiles 0.3-0.85 of AO distribution from about 55 days lag which is not apparent

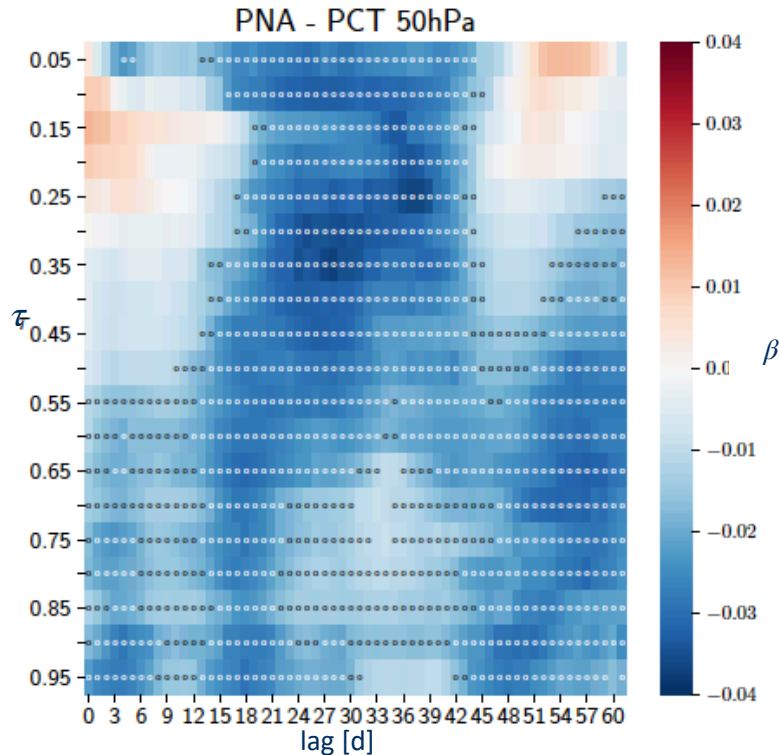


PCT50 vs NAO: lag 0-60 days

- ➔ negative NAO pattern (low quantiles) shows significant negative relation to PCT50 up until about 30 days
- ➔ positive NAO pattern (high quantiles) shows significant negative relation to PCT50 up until about 60 days that is stronger than for the negative NAO
- ➔ relation for lower and higher quantiles is stronger compared to median quantile
- ➔ significant positive relation between PCT50 and lowest quantile of NAO distribution from about 55 days lag

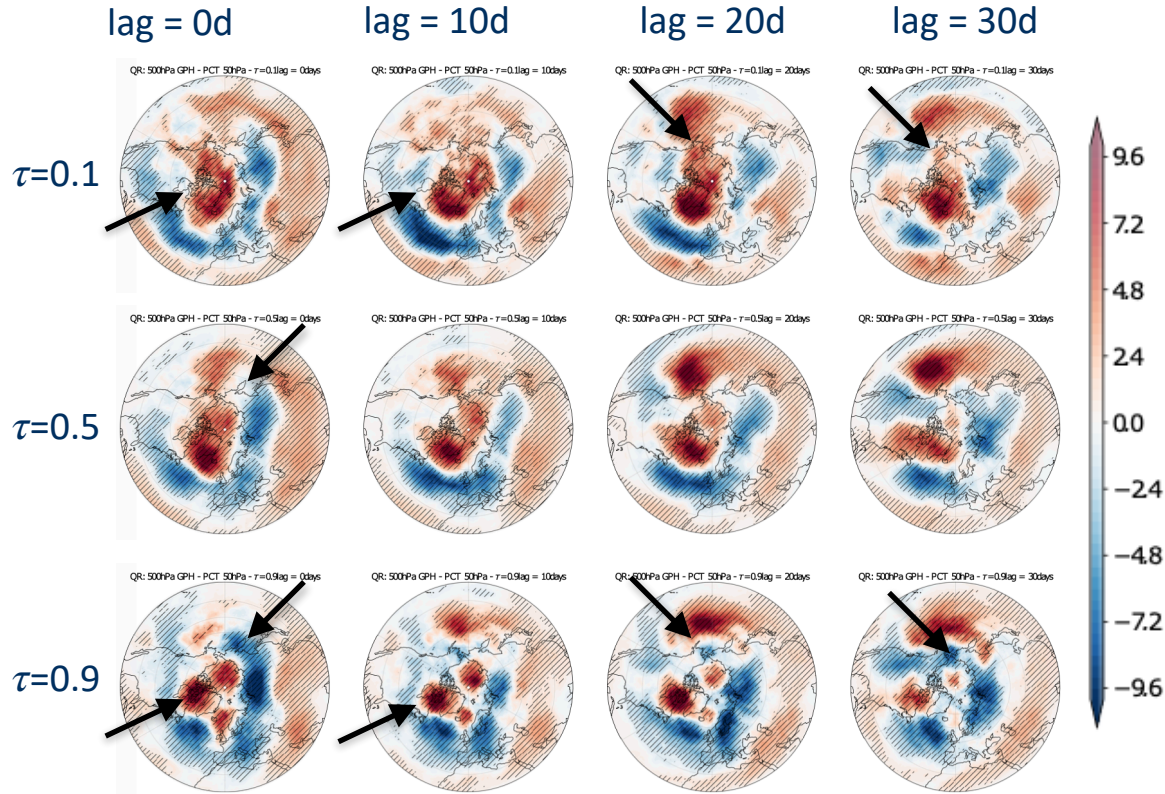


PCT50 vs PNA: lag 0-60 days



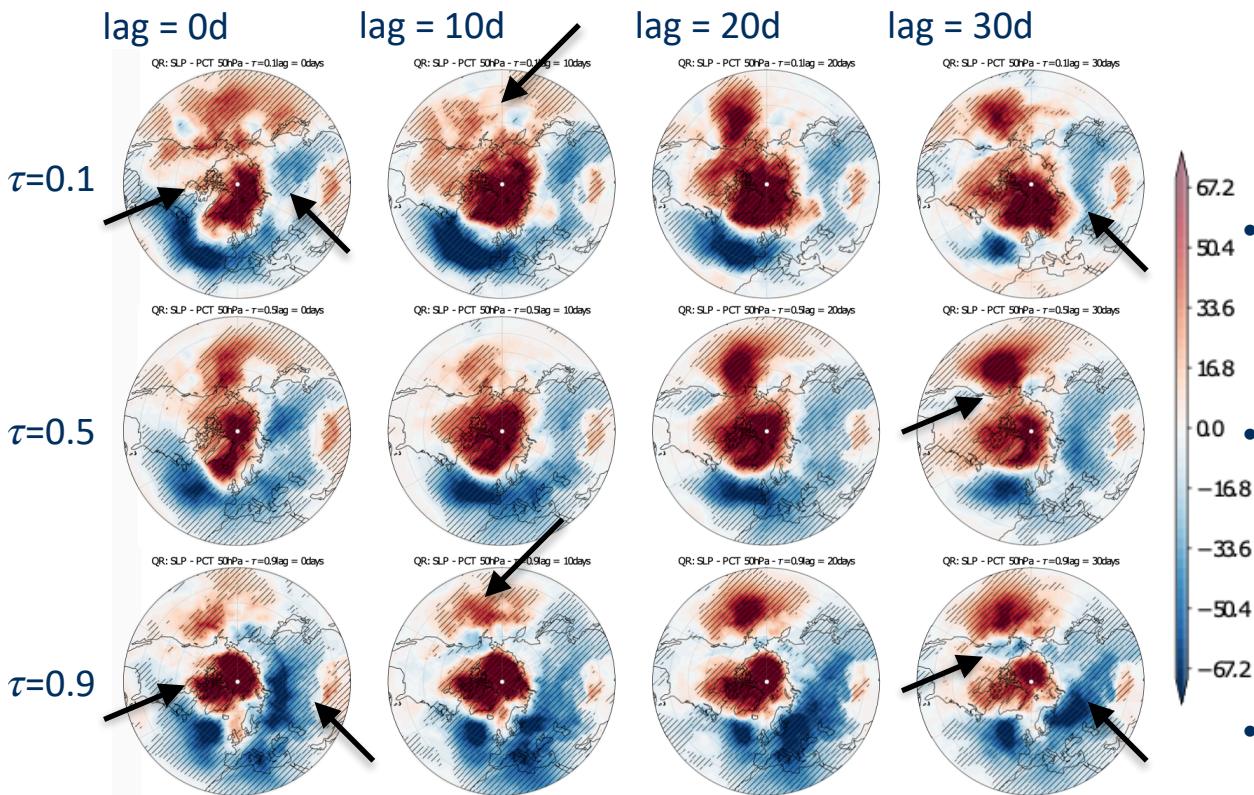
- ➔ negative PNA pattern (low quantiles) shows significant negative relation to PCT50 with a lag between 15 and about 40 days
- ➔ positive PNA pattern (high quantiles) shows significant negative relation to PCT50 up until about 60 days

3. Application to reanalysis - PCT50 vs z500



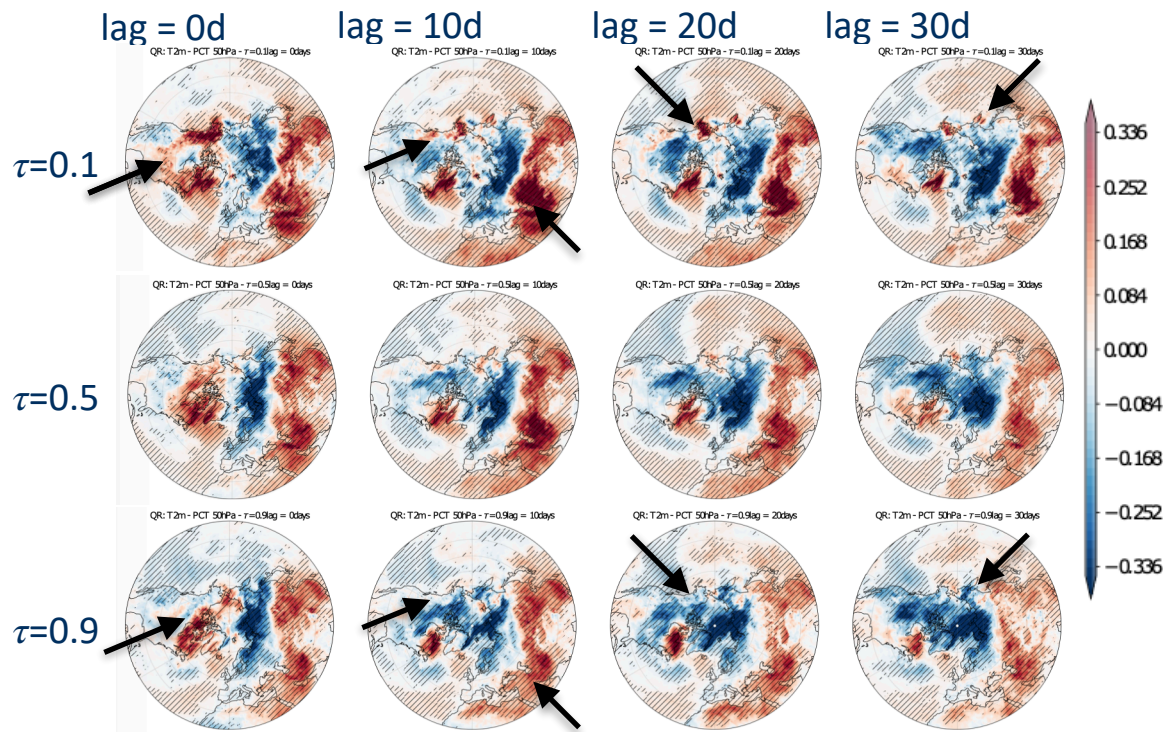
- z500 patterns can be explained by PCT50 in a similar way in vast regions across the three quantiles but with a change in strength of the regression coefficient
- in several regions (some of them marked) spatial shifts of the regression patterns are visible across the quantiles
- particularly the regression coefficients for the 0.9 quantile show a different relation with PCT50 in several regions compared to the median and the lower quantile
- differences get more apparent when looking at the lagged QR's

3. Application to reanalysis - PCT50 vs SLP



- SLP can be explained by PCT50 in a similar way in vast regions across the three quantiles but with a change in strength of the regression coefficient
- in several regions (some of them marked) spatial shifts of the regression patterns is visible across the quantiles
- similar to the z500 analysis the regression coefficients for the 0.9 quantile show different relations with PCT50 in several regions compared to the median and the lower quantile
- differences appear for all lags

3. Application to reanalysis - z500 vs T2m



- T2m can be explained by PCT50 in a similar way in vast regions across the three quantiles but with a change in strength of the regression coefficient
- in some regions (some of them marked) are shifts of the regression patterns visible across the quantiles
- the regression coefficients for the 0.9 quantile show different relations with PCT50 in several regions compared to the median and the lower quantile

- differences appear for all lags!

4. Conclusion

QR yields supplementary information about the relation between stratosphere and troposphere that cannot be inferred from ordinary linear regression models

Outlook:

- QR will be used for various other fields or indices in connection with the stratosphere-troposphere relation i.e. wave-activity fluxes, blocking indices, ..
- QR is a promising statistical tool that has potential in being used for extended range prediction

References

- Baldwin, M. P., and T. J. Dunkerton (2001), Stratospheric harbingers of anomalous weather regimes, *Science*, 294 (5542), 581-584, doi: 10.1126/science.1063315.
- Koenker, R., and K. F. Hallock (2001), Quantile Regression, *Journal of Economic Perspectives*, 15 (4), 143-156, doi:10.1257/jep.15.4.143.
- Koenker, R., & Bassett, G. (1978). Regression Quantiles. *Econometrica*, 46(1), 33-50. doi:10.2307/1913643
- NOAA; Climate Prediction Center, NAO, AO and PNA indices, accessed on 09.10.2020.