https://doi.org/10.5194/egusphere-2024-582
Preprint. Discussion started: 28 March 2024

Spatially aggregated climate indicators over Sweden (1860–2020), Part 1: Temperature

Christophe Sturm[1]

[1]Hydro-climatology research group, Swedish Meteorology and Hydrology Institute

Correspondence: Christophe Sturm (christophe.sturm@smhi.se)

Abstract.

Climate indicators are useful tools to synthesise climate information from multiple station time-series into a single national indicator. The method applied should be spatially representative and robust over time. We introduce a new method, based on Empirical Orthogonal Functions (EOF) during the calibration period 1961–2018, in order to reconstruct the climate indicator for temperature in Sweden for the full 1860–2020 period of available observations.

The new method delivers results in good overall agreement with the reference method (i.e. arithmetic mean from selected stations in the reference network). Discrepancies are found prior to 1900, primarily related to the reduced number of active stations: the robustness of the indicator estimation is assessed by an ensemble computation with added random noise, which confirms that the ensemble spread increases significantly prior to 1880.

The present study establishes that the 10-year running averaged temperature indicator rose from $-1.03°C$ in 1903 to $+1.19°C$ in 2010 (with respect to a mean value of $+4.64°C$ over the 1961–2018 calibration period), i.e. an increase by $+2.22°C$ in a century. The temperature difference between 1860 and 2020 was largest for winter (DJF) averages ($+3°C$) and minimal for summer ($+2°C$).

The leading EOF patterns illustrate the spatial modes of variability for climate variability, with a predominantly homogeneous, mono-modal distribution for temperature. For precipitation, the first EOF pattern displays more pronounced regional features (maximum over the West coast), which is completed by a north-south seesaw pattern for the second EOF. We illustrate that EOF patterns calculated from observation data reproduce the major features of EOF calculated from GridClim, a gridded data-set over Sweden, for annual and seasonal averages. The leading EOF patterns vary significantly for seasonal averages (DJF, MAM, JJA, SON) for temperature.

Finally, future developments of the EOF-method are discussed for calculating regional aggregated climate indicators, their relationship to synoptic circulation patterns and the benefits of homogenisation of observation series.

The EOF-based method to compute a spatially aggregated indicator for precipitation is presented in a companion article (Sturm, 2024a). The code and data for this study is available on Zenodo (Sturm, 2024b).

https://doi.org/10.5194/egusphere-2024-582
Preprint. Discussion started: 28 March 2024

1 Introduction

Direct observations over the instrumental period unequivocally illustrate the ongoing climate change, as cited in the latest IPCC (Physical science basis) report (Masson-Delmotte et al., 2021; Gulev et al., 2021)). Numerical methods that aggregate time-series from individual observation stations over larger area into a single climate indicator, are therefore a useful tool to represent climate change over e.g. an entire country. The goal of the present method is to compute a bias-free estimator, despite the fact that data availability (i.e. station activity) varies over time. In other words, the challenge is to define a uniform method, whose results for the early part of part of the record (with only few active observation stations, not necessarily spread uniformly over the territory) will be consistent with later results (with numerous, evenly spaced stations).

The Swedish Meteorology and Hydrology Institute (SMHI) has collected and compiled quality-ensured observations of temperature across Sweden since the eighteenth century: in the current study, we present a total of 993 time-series for temperature over the 1860–2020 period from observation stations across Sweden. We introduce a new method to calculate the respective climate indicator over Sweden, with help of a gridded climate data-set (GRIDCLIM (Andersson et al., 2021)), available over 1961–2018. First, missing observations during the 1961–2018 calibration period are replaced for selected observation stations (defined as calibration network). Second, the dominant modes of variability in the (now complete) calibration data-set of observations are analysed with Empirical Orthogonal Functions (EOF).

The analysis of the leading EOF patterns illustrate the dominant modes of variance in the temperature observations. We assess the similarities and differences between EOF computed from observation data versus a gridded data-set covering all of Sweden. In particular, we assess the effect of poorly sampled areas in the observation data-set, e.g. in high-elevation areas, on the leading EOF patterns.

In order to test the robustness of the new climate indicator, we blend random noise and random sub-sampling from the original station observations: this ensemble computation allows to define the 25% and 75% percentiles (in other words, the range between which 50% of all ensemble computations are found).

Finally, we discuss the discrepancies between the new EOF-based method and the original SMHI method, using an arithmetic mean of observations from the reference station network.

2 Data and methods

The present sections first describes the data-sets used in this study: SMHI's database for station observations over 1860–2020 called *SMHI-MORA*, and the gridded data-set *SMHI-GridClim* (Andersson et al., 2021) produced as a combination of *SMHI-MORA* observations with *UERRA* regional reanalysis over the 1961–2018 period. The data-sets *SMHI-MORA* and *SMHI-GridClim* will hereafter be referred to as MORA and GRIDCLIM.

Secondly, we describe the method currently implemented at SMHI to compute national climate indicators, based on an arithmetic mean of selected reference stations. Thirdly we introduce a new method based on empirical orthogonal functions (EOF), computed over the 1961–2018 period for which *UERRA* reanalysis is available. Leading EOF patterns established over the 1961–2018 period are used to reconstruct the climate indicator over the MORA period 1860–2020.

https://doi.org/10.5194/egusphere-2024-582
Preprint. Discussion started: 28 March 2024

Data-Set	Period	Y_{max}^{temp}	# Sta.	Perc.
Complete	Total		993	100%
	Initial	1860	32	3.2%
	Maximum	1962	422	42.5%
	Present-day	2020	293	29.5%
Calibration	Total		338	100%
	Initial	1860	18	5.3%
	$\text{GridClim}^{\text{start}}$	1961	193	57.1%
	Maximum	1977	249	73.7%
	$\text{GridClim}^{\text{end}}$	2018	151	44.7%
Reference	Total		35	100%
	Initial	1860	8	22.9%
	Maximum	2009	35	100%
	Present-day	2020	34	97.1%

Table 1. Number of stations found in the MORA database. The upper third represents the **Complete** MORA data-set, the middle third the **Calibration** sub-set, i.e. stations with at least 15 years coverage during 1961–2018, and the bottom third the **Reference** selected stations for the original SMHI methodology. For each subset, **"Total"** indicates the number of stations active at least once during the 1860–2020 period; **"Initial"** shows the number of stations active in 1860; **"Maximum"** the maximum of simultaneously active stations; **"Present-day"** the number of stations currently active. The complete time-evolution of active stations can be found in figures Fig. (1).

We finally describe a method to evaluate the robustness of various climate indicator estimators: the span of estimator realisations is computing by adding normally-distributed noise and restricting the computation to a subset of MORA stations.

2.1 Observational data sets

2.1.1 MORA observation station network

The MORA [1] database centralises meteorological and hydrological observations in Sweden. MORA is updated with real-time observations, and used operationally in several SMHI services.

Each station is identified with a unique numeric ID, with station name, geographic coordinates (longitude, latitude, altitude) as meta-data. If a station is relocated (e.g. if increasing nearby urbanisation leads to the station no longer), a new station ID is created; after a few years overlap, the old station is usually closed. The MORA thus contains many missing values, as illustrated in and Fig. (1). Table (1) summarises the maximum number of simultaneously active stations –which occurs in the

[1]The MORA acronym stands for "Meteorologiska observationer för realtid och arkiv", meaning "Meteorological observations for real-time and archive".

https://doi.org/10.5194/egusphere-2024-582
Preprint. Discussion started: 28 March 2024

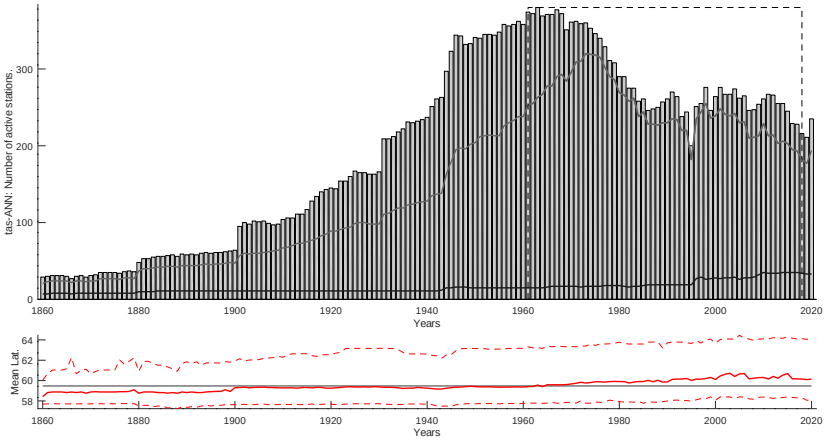

Figure 1. Upper plot: Number of active temperature stations in MORA over time (as bars). The dark grey line represents the number of active in the original reference station network; they light grey line represents the number of stations for the calibration network (i.e. individual stations being active at least 15 years during the calibration period 1961–2018, as highlighted by the dashed box). Lower plot: Median latitude for active stations in the calibration data-set over time (incl. the [25%–75%] bounds). The median latitude is used as a proxy for the distribution of the observation network.

early 1960's– amounts to 42.5% for temperature of the total number of stations in the network. This indicate that, on average, every station was replaced once or more under its lifetime.

The lower plot in Fig. (1) indicate the median latitude (as thick red line) and [25%; 75%] percentiles (in thin dashed red lines) of active stations over time. The median latitude is used in the present study as a proxy for the progressive extension of the observation network, in particular the increased station density in Sweden's northern regions over 1860–2020. It is worth noticing that the relative decrease of active observations in MORA since 1960 proportionally favoured observation stations in northern parts of Sweden.

Historical observations are continuously added to the MORA database, with ongoing digitalisation of historic observation reports (from printed records). Newly rescued observations are quality-controlled, but currently not (yet) homogenised: hence possible biases related to changes in measure instruments or modifications in the station's environment are not corrected. Historical observations for temperature are available at daily or monthly resolution. Monthly average are computed in MORA if less than 5 daily observations are missing, which are in this case interpolated.

The present study considers annually resolved observations, representing the annual average (hereafter referred to as *ANN*), or seasonal averages [2] (winter – *DJF*, spring – *MAM*, summer – *JJA* and autumn – *SON*). Each station (identified by a single station Id) is here considered as a single record: no "station coupling" was applied in the current study, i.e. "stitching" together

[2]In season's acronyms, the capital letter represents the first letter of the month (e.g. DJF: December–January–February).

https://doi.org/10.5194/egusphere-2024-582
Preprint. Discussion started: 28 March 2024

nearby stations (e.g. concatenating observations from a older station with a newer one close by, meant to replace it, in order to obtain longer time-series) (Joelsson et al., 2022, 2023).

Since the mid 19th-century, a network of stations covering all of Sweden using standardised measuring methods is available. In the present study, in line with previous investigations at SMHI, we select 1860 as initial year for climate indicator: it follows the establishment of the first meteorological observation network by the Swedish Academy of Science (Svenska Vetenskapsakademin) in 1858–1860. For the present study, we retrieved temperature observations for all available stations over the 1860–2020 period. Monthly values were retrieved from MORA; annual means were computed as arithmetic average, provided that no month was missing. The number of stations for temperature measurements are indicated in Table (1), while the evolution of active stations is shown in Fig. (1). In Table (1), the fist column indicates the number of stations in the MORA data-set, for temperature (column #4), The rows *Initial* and *Present-day* show the number active stations respectively at the start and the end of observation series. Additional rows $\text{GridClim}^{\text{start}}$ and $\text{GridClim}^{\text{end}}$ indicate the number of active stations during the 1961–2018 calibration period. The row *Maximum* indicates the year for which the number of active stations was maximal, along with the corresponding number of active stations.

The present study focuses on *annually-resolved* climate indicators, using annual and seasonal averages (winter DJF: December–February, sprint MAM: March–May, summer JJA: July–August, autumn SON: September–November).

2.1.2 GRIDCLIM gridded data-set

The GRIDCLIM project (Andersson et al., 2021), performed by SMHI, combines the regional European reanalysis UERRA (Schimanke and Service, 2019; Schimanke et al., 2019) with station observations to produce a uniform, bias-corrected gridded climate dataset. Observations for Sweden were retrieved from the MORA database maintained by SMHI, and similar observations were obtained for Norway and Finland from respective national meteorological agencies. Observations were completed from ECMWF MARS archive for neighbouring countries in the study domain (Denmark, Baltic States). Quality control on the observation data-sets were performed using the TITAN package (Båserud et al., 2020).

The original resolution of the UERRA reanalysis is 11 km, subsequently downscaled at 2.5 km with MEPS (Ono et al., 2021). The GRIDPP open-source software (Lussana, 2007), developed by the Norwegian Meteorological Institute, performed an Optimal Interpolation (OI) analysis to combine observations and the UERRA gridded first guess. The full GRIDCLIM data-set consists of $567 \times 823 = 466\,641$ grid-points, of which $69\,842$ cover the Swedish territory.

The GRIDCLIM data-set provides gridded estimates (at 2.5 km resolution) for following variables: 2-meter temperature, 2-meter relative humidity, daily maximum and minimum temperature, daily precipitation, and daily snow depth at hourly resolution. We restrict the current analysis to 2-m temperature (henceforth referred to as "temperature"). The data was first averaged at monthly scale, and further to annually-resolved annual and seasonal averages.

The GRIDCLIM data-set (Andersson et al., 2021) covers the period 1961–2018, in accordance with the availability of the UERRA reanalysis (Schimanke and Service, 2019; Schimanke et al., 2019). This is henceforth defined as the calibration period (1961–2018) in subsequent method description.

https://doi.org/10.5194/egusphere-2024-582
Preprint. Discussion started: 28 March 2024

In the following sections, GRIDCLIM data-set is used in two ways: GridClim-all represents all 69842 grid-cells of the GRIDCLIM data-set covering the Swedish territory (in its native EPSG:9001 Lambert Conic Conformal projection with 63° standard parallel, 2.5km x 2.5km resolution), is used to compute the EOF and SVD output. GridClim-all is also represented in Table (3) and corresponding graphs for comparison purposes. GridClim-sub is a subset of GridClim-all, whose grid-cells correspond to MORA calibration station network. GridClim-sub is used for the gap-filling of MORA calibration data-set (338 for temperature) prior to the EOF analysis. GridClim-ref is a subset of GridClim-all, with time-series taken from grid-cells corresponding to the MORA reference network SMHI-ref (39 for temperature).

2.2 Computing a national climate indicator

2.2.1 Emulating the operational SMHI method

SMHI operationally calculates climate indicators for temperature (Engström, 2022, 2023). The climate indicators are computed from a reference network of stations across the country (35 stations for temperature, cf. Table (1)), at annual and seasonal resolution, hereafter referred to as SMHI-ref data-set.

In the operational computations, the reference data consists of coupled stations (i.e. concatenating neighbouring stations to obtain longer time-series for a given "meta-station"). The data is subsequently homogenised in order to correct breaks, i.e. signal disruptions not deemed to be related to climate variability. These can include changes in mean value or seasonal amplitude, as a result of changes in station location, instrument calibration etc.

The original method to compute the climate indicator is the simplest way to synthesise multiple stations records as a single time-series: it uses an arithmetic average, as presented in Eq. (1), where $x_{i_{sta}}(t)$ represents an individual (incomplete) observation time-series and $n_{sta}(t)$ is a function counting active stations over time.

$$\begin{aligned} CI_{ori}(t) &= \frac{\sum_{i_{sta}=1}^{n_{sta}} x_{i_{sta}}(t)}{n_{sta}(t)} \\ \Leftrightarrow CI_{ori}(j) &= \frac{\sum_{i=1}^{n} \mathbf{X}(i,j)}{\sum_{i=1}^{n} (\mathbf{X}(i,j) \neq \text{NaN})} \end{aligned} \tag{1}$$

The method is described in further detail in the online documentation (Engström, 2022, 2023).

The original method to calculate climate indicator as an arithmetic average of available station is currently applied by the SMHI to the "reference network", with 35 (meta-)stations for temperature, cf. Table (1), lower half. In the results section, this data-set is referred to as SMHIRef.

In the present study, the original method was emulated by selecting the individual station with the longest record in each meta-station, without coupling and homogenisation. Hence, the emulated reference network in this study has sparser data coverage, thereby slightly different result, than the climate indicator provided by SMHI. In the results section, this data-set is referred to as GridClim-sub.

https://doi.org/10.5194/egusphere-2024-582
Preprint. Discussion started: 28 March 2024

2.2.2 EOF decomposition of 2-dimensional $[space, time]$ data-set

Standard linear algebra methods require the data-set to be complete, without any missing values (marked as not-a-number NaN in the $\mathbf{X}$ notation). The GRIDCLIM gridded data-set is per construction complete over the calibration period 1961–2018: the EOF decomposition can be performed immediately. The EOF decomposition method is described in Björnsson and Venegas (1997).

During the 1961–2018 calibration period, we select MORA stations having at least 15 year of data coverage. Initial tests revealed that the EOF computation could result in physically inconsistent results, when stations with short data-coverage and/or gaps (possibly with erroneous annual mean) were included in the input (see also discussion of Fig. (2)). Hence an additional criteria was introduced: we select only MORA stations who share at least 50% of variance with their corresponding grid-cell in GRIDCLIM (i.e. a correlation factor $r > \sqrt{0.5}$ with pair-wise removal of missing data). This defines the "Calibration" subset of MORA stations, whose dimension for temperature is given in Table (1).

The first step before gap-filling is to identify, for every station, which yearly record is available, defined by availability matrix $\mathcal{A}$. For each station, a linear regression (in least-square sense, i.e. minimising ε) is computed between available annual MORA station records $\mathbf{X}_{MORA}$ and the value in the corresponding GRIDCLIM grid-cell. Missing values in the MORA station record are substituted by the linear function applied to the GRIDCLIM value for the corresponding time-step, as shown in Eq. (2). Replacing missing values using linear regression is a fairly trivial method, which requires a complete (gridded) data-set. Alternative method include the use of homogenisation tools (Guijarro et al., 2023; Guijarro, 2024), or AI methods (Kadow et al., 2020).

$$\forall i \in [[1, n_{stations}]], \forall j \in [[1, n_{years}]],$$
$$\begin{cases} X_{MORA}(i,j) \neq \text{NaN} \Rightarrow & \mathcal{A}(i,j) = 1 \\ X_{MORA}(i,j) = \text{NaN} \Rightarrow & \mathcal{A}(i,j) = 0 \end{cases}$$

Linear regression MORA vs GRIDCLIM:

$$\forall i \in [[1, n_{stations}]], \forall j \in (\mathcal{A}(i,:) = 1), \tag{2}$$
$$\alpha^i \cdot \mathbf{X}_{\text{GridClim}}(i,:) + \beta^i + \varepsilon = \mathbf{X}_{MORA}(i,:)$$

Substitution of MORA missing values:

$$\forall i \in [[1, n_{stations}]], \forall j \in (\mathcal{A}(i,:) = 0),$$
$$\mathbf{X}_{MORA}(i,:) = \alpha^i \cdot \mathbf{X}_{\text{GridClim}}(i,:) + \beta^i$$

Hence, the calibration sub-set of MORA stations is now complete over the 1961–2018 calibration period. The data-set $\mathbf{X}_{MORA}$ can be centered, i.e. the 1961–2018 temporal mean for each station/grid-point is removed from the time-series. In other words, centring the data-set $\mathbf{X}$ means computing individual station anomalies.

https://doi.org/10.5194/egusphere-2024-582
Preprint. Discussion started: 28 March 2024

Centering the data-set prior to applying the EOF analysis provides the advantage of representing the spatial and temporal variability of the studied data-set (i.e. respective spatial patterns and time loadings) as departures from zero. It is worth noticing that absolute values can easily be obtained from the centered data-set (as both input or output), by adding the respective mean over the reference period:

$$\begin{aligned} &\forall i \in [[1, n_{\text{stations}}]], \forall t \in [[1961, 2018]] \\ &\qquad x_c^i(t) = x^i(t) - \frac{\sum_{t=1961}^{2018} x^i(t)}{2018 - 1961 + 1} \\ &\Leftrightarrow \mathbf{X}(i,:) = \mathbf{X}_c(i,:) + \overline{\mathbf{X}_{1961-2018}(i,:)} \end{aligned} \tag{3}$$

Hereafter, centered data-sets are marked as $\mathbf{X}_c$.

The Empirical Orthogonal Functions (EOF) method, equivalent to Principal Component Analysis (PCA), aims at decomposing the spatio-temporal variability in the data-set $\mathbf{X}_c$ as a series of spatial patterns (hereafter referred to as spatial EOF patterns) $\mathbf{EOF}$ (Björnsson and Venegas, 1997), associated to its time expansion coefficients $\mathbf{A}$.

$$\begin{aligned} \text{Covariance matrix: } &\mathbf{R} = \mathbf{X}_c^T \cdot \mathbf{X}_c \\ \text{Eigenvalue problem: } &\mathbf{EOF} \cdot \mathbf{R} = \mathbf{EOF} \cdot \Lambda \\ &\Leftrightarrow \mathbf{R} = \mathbf{EOF} \cdot \mathbf{\Lambda} \cdot \mathbf{EOF}^{-1} \\ \text{Time expansion coefficient: } &\mathbf{A} = \mathbf{X}_c \cdot \mathbf{EOF} \end{aligned} \tag{4}$$

By construction, each pattern in the $\mathbf{EOF}$ matrix is "orthogonal" to each other, in other words spatially uncorrelated. Accordingly, all time expansion vectors ($\mathbf{A}$) are temporally uncorrelated. The eigenvalue matrix (Λ) is a diagonal matrix, indicating the variance associated to each EOF mode. Following property of the $\mathbf{EOF}$ is worth mentioning: $\mathbf{EOF} \cdot \mathbf{EOF}^T = \mathbf{EOF} \cdot \mathbf{EOF}^{-1} = \mathbf{Id}$, hence $\mathbf{EOF}^T = \mathbf{EOF}^{-1}$.

As a result, the original matrix $\mathbf{X}_c$ can be identically reconstructed based on its decomposition in spatial patterns ($\mathbf{EOF}$) and associated time expansion coefficients ($\mathbf{A}$). Assuming that the original data-set $\mathbf{X}_c$ has n stations/grid-points, and m time steps (as columns), the number of unique modes in the present study is set to $p = \min(n, m)$, i.e. in practice the number of time steps.

$$\begin{aligned} &\mathbf{X}_c = \mathbf{A} \cdot \mathbf{EOF}^T \\ \Leftrightarrow &\mathbf{X}_c = \sum_{i=1}^{p} \overrightarrow{\mathbf{a}^i} \cdot \overrightarrow{\mathbf{eof}^T i} \end{aligned} \tag{5}$$

The EOF decomposition was applied over the 1961–2018 calibration period for the GRIDCLIM data-set, restricted to the Swedish territory (Fig. (2)). Analogue results of the EOF decomposition on the MORA Calibration data-set for the 1961–2018 period is shown in (Fig. (3)).

https://doi.org/10.5194/egusphere-2024-582
Preprint. Discussion started: 28 March 2024

2.2.3 Estimating the climate indicator and related uncertainties

The formulation in Eq. (5) can be extended beyond the 1961-2018 calibration period, despite missing values found in the longer 1860–2020 record. Hereafter, $\mathbf{X}^{\star}$ represents a matrix $\mathbf{X}$ with missing values, and $\widehat{\mathbf{A}}$ the estimator of its time expansion coefficients. In other words, $\widehat{\mathbf{A}}$ is an approximation of the time expansion coefficient matrix $\mathbf{A}$ for the entire study period 1860–2020 and full rank of Λ (i.e. p, in our case the number of years in the calibration period, cf. Eq. (4)), with two limitations:

Missing values: $\widehat{\mathbf{A}}$ can only be computed from available observations in $\mathbf{X}^{\star}$. As illustrated in Fig. (1), the number of active stations decreases continuously prior to the calibration period. In other words, the dominant modes in $\widehat{\mathbf{A}}$ can only be captured if the distribution of active stations represents a sufficient sampling of the corresponding spatial pattern $\mathbf{EOF}$.

Stationarity: The current reconstruction methods relies on the implicit assumption that the dominant modes established during the 1961–2018 calibration period are time-independent, or at least reasonably represent the relative behaviour of observed variables during 1860–1961.

Eq. (6) expresses the new method for estimating the gap-filled observation data-set, based on the EOF decomposition of the MORA calibration subset $\mathbf{X}_c^{MORA,Cal}$.

Time expansion coefficients:

$$\begin{aligned} \mathbf{A} \simeq & \mathbf{X}_c^{\star} \cdot \mathbf{EOF} \\ \Leftrightarrow \widehat{\mathbf{A}} = & \sum_{i=1}^{p} \mathbf{X_c^{\star}}^i \cdot \mathbf{eof}^i \\ \Rightarrow \forall i \in [[1,p]] & \\ \widehat{\mathbf{a}^i(t)} = & \sum_{x_c^i(t) \neq \mathrm{NaN}} x_c^i(t) \cdot \mathbf{eof}^i(t) \end{aligned}$$

Data-Set reconstruction:

$$\widehat{\mathbf{X}_c}^{EOF} = \widehat{\mathbf{A}} \cdot \mathbf{EOF}^{-1} \tag{6}$$

Here, the notation $\mathbf{EOF}^{-1}$ represents the pseudo-inverse, since $\mathbf{EOF}$ is not a square matrix.

Since most of the variance is comprised in the leading modes (given that, by construction, eigenvalues in Λ are listed in decreasing order), we chose to restrict the reconstruction of $\widehat{\mathbf{A}}$ to the 10 leading modes (cf. Table Table (2)). Furthermore, leading modes ($\mathbf{eof}^{MORA}(\#1-3)$ as shown in Figures Fig. (2)) display spatial patterns consistent with climate phenomenon, while modes with lower eigenvalues mostly display numerical "noise" (i.e. patterns without obvious physical significance). Hence restricting to the 10 leading modes reduces the risk of spurious, physically inconsistent artefacts in $\widehat{\mathbf{A}}^{MORA}$. Hence, Eq. (6) is applied hereafter with $p = 10$.

Analogously to Eq. (1) defining CI_{ori}, the EOF-based climate indicator CI_{EOF} can be expressed as:

https://doi.org/10.5194/egusphere-2024-582
Preprint. Discussion started: 28 March 2024

$$\forall j \in [[1, 2020-1860+1]], CI_{EOF}(j) = \frac{\sum_{i=1}^{n} \widehat{\mathbf{X}_c}^{EOF}(i,j)}{\sum_{i=1}^{n} (\mathbf{X}_c(i,j) \neq \text{NaN})} \quad (7)$$

Compared to the original method (Eq. (1)), the new methods has the advantage that available observations contribute to the estimated indicator according to their "weight" in the 10 leading modes. Hence the EOF-based method (Eq. (7)) can be considered as a weighted, rather than arithmetic average of station data (Eq. (1)). Thus, it has the potential of being less sensitive to changes in station coverage, as shown in Figures Fig. (1). The effect of damped variability when station coverage decreases (mentioned in the previous paragraph) is corrected in the estimation of CI_{EOF} by considering the number of active stations in the denominator ($\sum_{i=1}^{n} (\mathbf{X}_c(i,j) \neq \text{NaN})$ is equivalent to $\sum_{i=1}^{n} \mathcal{A}(i,j)$ in Eq. (2)).

In order to test the robustness of the climate indicator estimator, we perform an ensemble computation to assess the effect of sub-sampling and added noise.

$$\forall k \in [[1, n_{ens}]], CI^{k}_{ens}(j) = \frac{\sum_{i \in \text{(k-th random subset)}} \left(\widehat{\mathbf{X}_c}(i,j) + \varepsilon^{k}(i,j)\right)}{\sum_{i \in \text{(k-th random subset)}} (\mathbf{X}_c(i,j) \neq \text{NaN})} \quad (8)$$

We assume hereafter that the daily temperature observations are normally distributed around 0, with a standard deviation of $\varepsilon_{\text{temp}} = 0.5°\text{C}$. This is a conservative estimate, based on the fact that early thermometers were graduated with $0.5°\text{C}$ increments. The corresponding measurement uncertainty for monthly means are reduced by a factor $\sqrt{30} \cdot \sqrt{n_{\text{month}}}$, i.e. the average amount of daily measurements in a monthly mean. n_{month} represents the number of monthly records in the annually-resolved average: for annual means, $n_{\text{month}} = 12$, for seasonal means, $n_{\text{month}} = 3$. Hence the random noise function ε is normally distributed, with a standard deviation of $\sigma^{\text{temperature}}_{ANN} = \frac{0.5}{\sqrt{30} \cdot \sqrt{n_{\text{month}}}} = 0.026\,°\text{C}$ for annual temperature and $\sigma^{\text{temperature}}_{seas} = 0.053°\text{C}$ for seasonal means (DJF, MAM, JJA, SON). The presumed measurement error for monthly observations is here applied to annual values (i.e. a 12-member mean): we consider thus the random noise factor to be in the upper range of physically reasonable values.

The second aspect to be assessed is the impact of sample size on the ***CI*** climate indicator estimate. A random subset of stations within the calibration network, including three times as many stations as in the reference network were used to compute the ***CI***, with a new realisation of the random noise function ε. In other words, Table (1) indicates that the reference network contains $n_{Ref} = 35$ stations for temperature; the ***CI*** for each ensemble member is thus computed from $n = 3 \times n_{Ref}$ randomly chosen stations within the calibration network.

The procedure above is repeated 100 times; the 25% and 75% percentiles are computed and presented as thin lines on Fig. (6), representing the difference from the original SMHIref (ΔCI). The same procedure was applied to evaluate the 25% and 75% percentiles for RefNetwork, albeit using 100 realisations with half (i.e. $n = \frac{n_{Ref}}{2}$) of available stations in the reference network.

https://doi.org/10.5194/egusphere-2024-582
Preprint. Discussion started: 28 March 2024

3 Results

The first sections describe the EOF patterns computed for temperature, from the GRIDCLIM and MORA calibration data-set, as computed from Eq. (4).

Following figures represent the leading EOF modes of annual temperature from GRIDCLIM (Fig. (2)); annual temperature from MORA (Fig. (3). seasonal temperature from GRIDCLIM (Fig. (4)).

Similarly, the SVD decomposition was performed for temperature for the GRIDCLIM and MORA data-sets. The EOF and SVD patterns are virtually identical. Hence the corresponding figures are shown in the supplementary material.

	Data-Set	Method	$\lambda(\#1)$	$\lambda(\#2)$	$\lambda(\#3)$	Sum
T	GRIDCLIM	EOF	92.2%	4.5%	0.9%	97.6%
T	MORA	EOF	82.9%	3.7%	1.3%	87.9%

Table 2. Portion of explained variance λ for the leading three modes with the EOF analysis of GRIDCLIM and MORA over the 1961–2018 calibration period. The sum of the variance explained by the first 3 modes is indicated in the last column.

3.1 EOF patterns for temperature

By construction, plain (i.e. unrotated) spatial EOF patterns (without prior detrending) are expected to display a uni-modal distribution in the first mode, a bi-modal distribution in the second, and a tri-modal in the third. This behaviour is clearly apparent in Figures Fig. (2).

Fig. (3) represent the EOF patterns for the MORA data-set. The polygons represent the Delaunay triangulation of the MORA stations in the calibration subset; it is worth noticing that, despite the area of the polygons vary (with a tendency to increase in northern Sweden), each time-series has the same weight in the EOF method.

Table (2) indicates how much of the total variance is expressed in the leading three modes (as obtained from matrix Λ in Eq. (4)). For *temperature*, the EOF first mode is strongly dominant: up to 92.2% of the variance is expressed in the first mode. This result confirms previous observation: temperature tends to vary fairly homogeneously across all of Sweden. This aspect is clearly an advantage to define a robust climate indicator for the country: the computation will be less sensitive to the location of stations, and less affected if the stations are temporarily inactive.

3.1.1 Annual EOF patterns for temperature

Beyond this initial statement, it becomes obvious that GRIDCLIM ($\lambda^{EOF}_{\text{GridClim}}(\#1) = 92.2\%$, $\sum_{i=1}^{3} \lambda^{EOF}_{\text{GridClim}}(i) = 97.6\%$) has significantly higher portion of variance explained in its first modes than MORA ($\lambda^{EOF}_{MORA}(\#1) = 82.9\%$, $\sum_{i=1}^{3} \lambda^{EOF}_{MORA}(i) = 87.9\%$). Several factors explain it: GRIDCLIM contains more homogeneous values, as a result of assimilating observations with re-analyses. By representing mean temperature for each grid-cells, GRIDCLIM also fails to represent extreme values measured at point stations in MORA.

https://doi.org/10.5194/egusphere-2024-582
Preprint. Discussion started: 28 March 2024

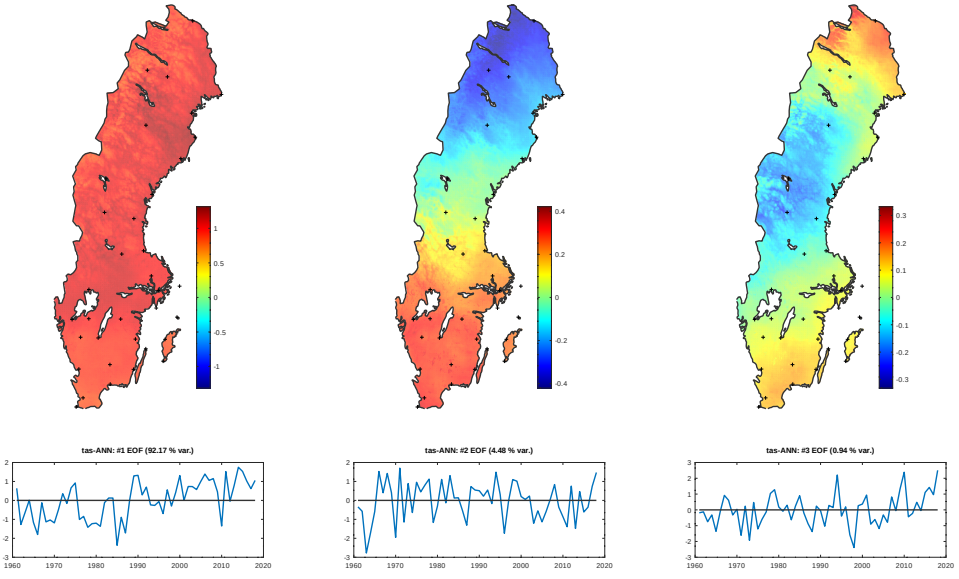

Figure 2. Leading three EOF patterns for GRIDCLIM over Sweden ($\mathbf{eof}(\#1-3)$, upper row), with their associated time expansion vectors ($\mathbf{a}(\#1-3)$, lower row) for *temperature*. The unit for the product $\mathbf{a}(\#1-3)\cdot\mathbf{eof}^{\boldsymbol{T}}(\#1-3)$ is in °C. Black crosses indicate the location of MORA stations in the reference network.

Fig. (3) represents the leading EOF patterns for the MORA *temperature* data-set (top row) and their associated time-expansion coefficients (according to Eq. (4) and Eq. (5)).

For mode #1 of temperature, the spatial EOF patterns from MORA resembles that of GRIDCLIM, despite some stations standing out; the respective time expansion coefficient are similar (lower graphs in Fig. (2) and Fig. (3)), both in term of variability and trend.

For EOF mode $\mathbf{eof}^{MORA}$ #2 and #3, the same conclusion holds: the GRIDCLIM spatial EOF pattern and time expansion coefficients are generally accordingly represented in the EOF pattern for MORA, but it differs slightly in mountain regions. It is beyond the scope of the present study to assess the causes of the discrepancy: it could be related to insufficient sampling of high-altitude locations in the MORA station network in GRIDCLIM.

3.1.2 Seasonal EOF patterns for temperature

The leading seasonal EOF modes for temperature in the GRIDCLIM data-set are shown in Fig. (4); by construction, EOF#1 displays a mono-modal distribution, EOF#2 a bi-modal etc. The overall geographic pattern is similar to the annual EOF (cf. Fig. (2)).

https://doi.org/10.5194/egusphere-2024-582
Preprint. Discussion started: 28 March 2024

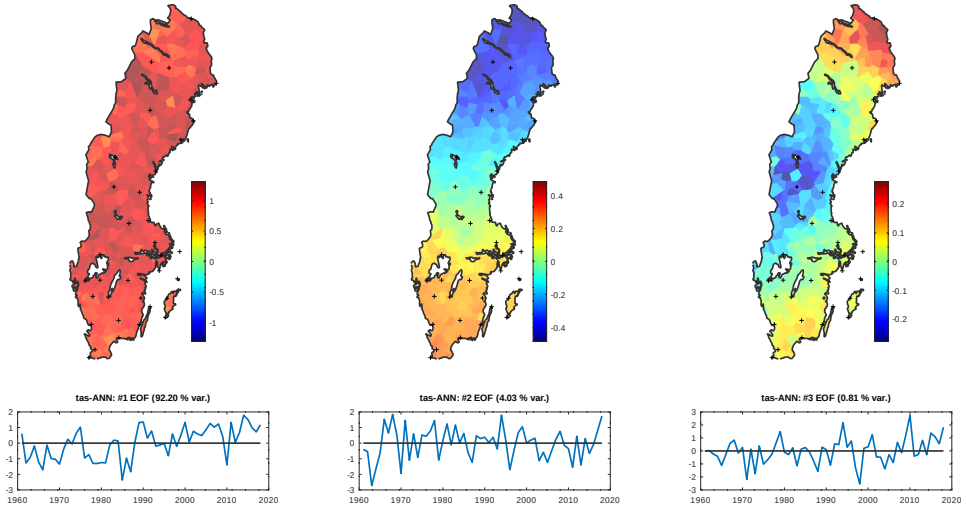

Figure 3. Leading three EOF patterns for MORA (Calibration network) over Sweden ($\mathbf{eof}^{MORA}(\#1-3)$), with their associated time expansion vectors ($\mathbf{a}^{MORA}_{EOF}(\#1-3)$) for *temperature*.

We can however notice that $\mathbf{eof}_{DJF}(\#1)$ for winter (i.e. DJF, upper left plot) has an amplitude about twice of $\mathbf{eof}_{ANN}(\#1)$, as illustrated by the colour-bar. The feature highlight topographic features: when the mean temperature over Sweden increases, low-land areas in the vicinity of mountains will experience a stronger temperature increase than surrounding areas.

The summer (JJA) $\mathbf{eof}_{JJA}(\#1)$ distinguishes itself from other seasons (and the annual mean), as the portion of explained variance ($\lambda_{JJA}(\#1) = 80\%$) is significantly less than the annual ($\lambda_{ANN}(\#1) = 92\%$). During this season, a stronger second mode ($\lambda_{JJA}(\#2) = 13\%$) reflects the North-South seesaw pattern, which is less pronounced in annual means: when Southern Sweden gets warmer, North-Western Sweden tends to get cooler.

The analysis of seasonal time expansion coefficients $\mathbf{a}$ (according to Eq. (4)) reveals seasonal differences. $\mathbf{a_{ANN}(\#1)}$ displays a distinctive shift to warmer (annual) temperatures around 1990. The same behaviour is identifiable for winter (DJF) and spring (MAM), however not for summer (JJA) nor autumn (SON).

3.2 Reconstructing a national climate indicator *CI*

Several methods are evaluated to estimate the climate indicator (***CI***) over the 1961–2018 calibration period and the whole study period 1860–2020, as listed in Table (3). For each climate indicator (temperature and precipitation), the characteristics of the *reference* (used for the SMHI reference method) and the *calibration* (used in the present EOF method) observation networks are listed in Table (1), and illustrated in Fig. (1).

https://doi.org/10.5194/egusphere-2024-582
Preprint. Discussion started: 28 March 2024

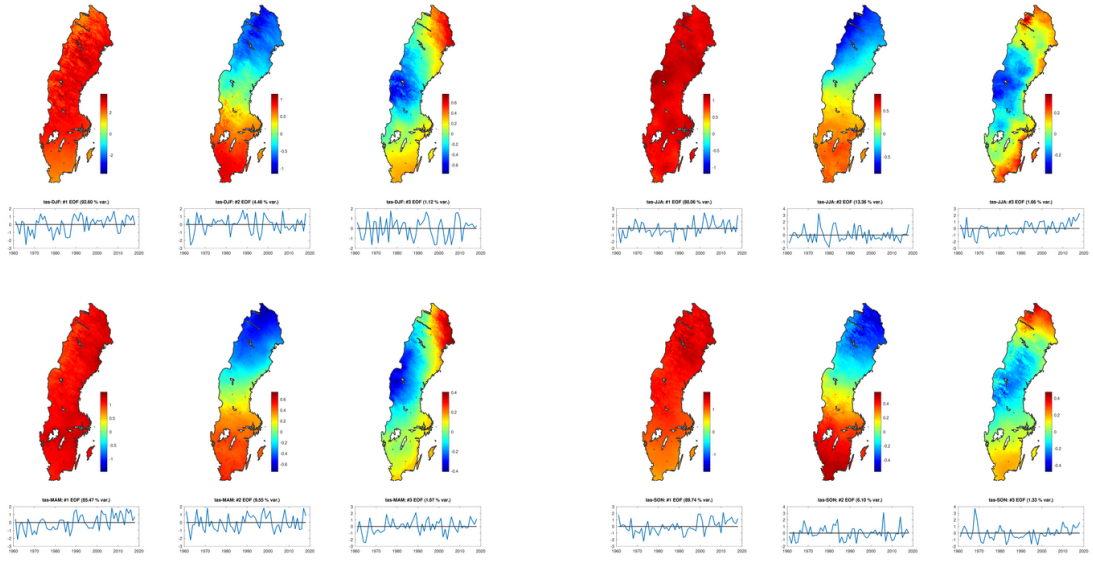

Figure 4. Leading three EOF seasonal patterns for GRIDCLIM *temperature* over Sweden ($\mathbf{eof}(\#1-3)$, upper row), with their associated time expansion vectors ($\mathbf{a}(\#1-3)$, lower row) for *temperature*. The unit for the product $\mathbf{a}(\#1-3)\cdot\mathbf{eof}^{\boldsymbol{T}}(\#1-3)$ is in °C. Black crosses indicate the location of MORA stations in the reference network.

As a preliminary comment, the EOF (SVD) based ***CI*** estimates are generally very close to the SMHI reference indicator SMHI-ref. In order to investigate the differences between ***CI*** estimates computed in the present study with SMHI-ref, we introduce $\Delta CI[\mathbf{X_c}]$, the departure from SMHI-ref, defined in Eq. (9).

The definition of all ***CI*** methods, including their corresponding legend for figures in this section, are summarised in Table (3). The upper part of the table lists indicators that apply to the 1961–2018 calibration period, while ***CI*** described in the lower part apply to the entire study period 1860–2020.

$$\begin{cases} CI_{ori}^{Ref} & = CI_{ori}[\mathbf{X}^{\star}(\text{Reference network})] \\ \Delta CI[\mathbf{X_c}] & = CI[\mathbf{X_c}] - CI_{\text{SMHI-ref}}^{centered} \end{cases} \tag{9}$$

The calibration period 1961–2018 presents the advantage of having a complete GRIDCLIM data-set, as well as comparing gap-filled $\mathbf{X_c}$ and original $\mathbf{X_c^{\star}}$ MORA data-sets. The present section evaluates the performance climate indicator (***CI***) estimates with EOF (respectively SVD) methods. We also assess the representativity of the *reference*, *calibration* MORA station networks for Sweden's climate, compared to a ***CI*** computed over all grid-points in GRIDCLIM covering Sweden.

Following estimations of ***CI*** are available over the calibration period:

https://doi.org/10.5194/egusphere-2024-582
Preprint. Discussion started: 28 March 2024

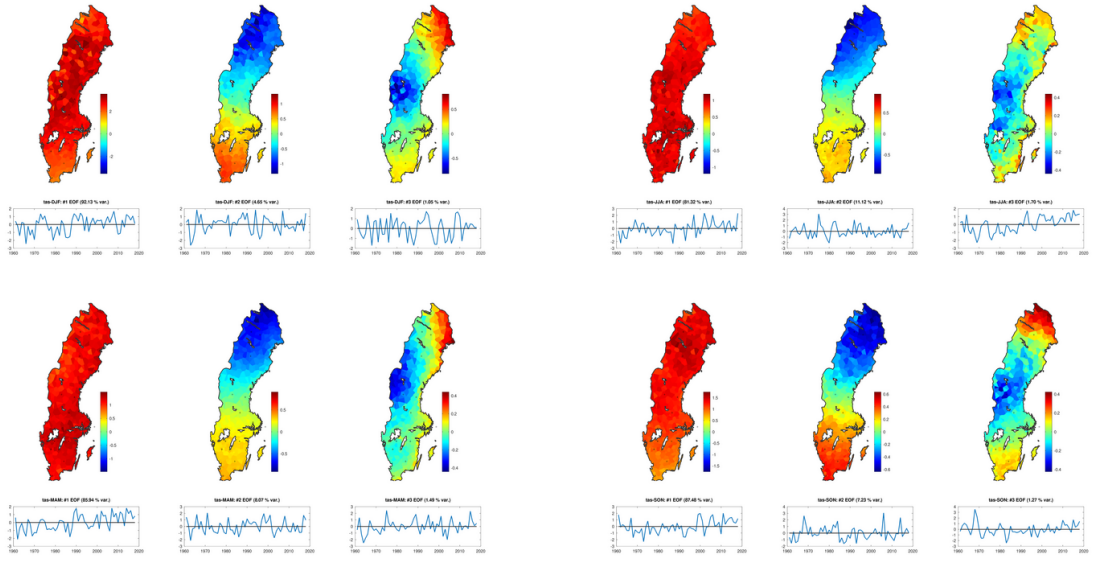

Figure 5. Leading three EOF seasonal patterns for MORA *temperature* over Sweden ($\mathbf{eof}(\#1-3)$, upper row), with their associated time expansion vectors ($\mathbf{a}(\#1-3)$, lower row) for *temperature*. The unit for the product $\mathbf{a}(\#1-3)\cdot\mathbf{eof}^{\boldsymbol{T}}(\#1-3)$ is in °C. Black crosses indicate the location of MORA stations in the reference network.

- $CI_{\text{SMHI-ref}}$: current ***CI*** definition used by SMHI, calculated as the arithmetic mean of observations for selected MORA observations (i.e. a sub-set of the MORA calibration network). $CI_{\text{SMHI-ref}}$ is not calculated in the present study; only the climate indicator is available, without access to the underlying station data (including station coupling, partial homogenisation and correction). Therefore, the centered climate indicator, in this particular case, is obtained as: $CI^{centered}_{\text{SMHI-ref}} = CI^{abs}_{\text{SMHI-ref}} - \overline{CI^{abs}_{\text{SMHI-ref}}}\Big|_{1961-2018}$
- $CI_{\text{GridClim-ref}}$: arithmetic average of time-series from the GRIDCLIM data-set for grid-points where *reference* stations are located. $CI_{\text{GridClim-ref}}$ is primarily meant to assess effect of time-space averaging order. Unlike $CI_{\text{SMHI-ref}}$, the spatial average is performed on centered station records $CI^{centered}_{\text{GridClim-ref}} = \dfrac{\sum_{i_{sta}=1}^{n_{sta}} \mathbf{X}^{\text{GridClim}}(ind_{\text{Ref-Network}},:)}{n_{sta}}$
- $CI_{\{\text{Sub,All}\}\text{-GridClim}}$: arithmetic average of time-series from the GRIDCLIM data-set for grid-points where *calibration* stations are located (for $CI_{\text{Sub-GridClim}}$), or all grid-points covering Sweden, as shown in e.g. Fig. (2) (for $CI_{\text{All-GridClim}}$).
- $CI_{\text{EOF-cal}}$: average over the EOF-reconstructed $\widehat{\mathbf{X}^{\text{EOF}}_{\mathbf{MORA}}}$, according to Eq. (7).

The definition of ΔCI over the 1961–2018 calibration period, including their corresponding legends for figures Fig. (6) and following, are summarised in the upper Table (3).

https://doi.org/10.5194/egusphere-2024-582
Preprint. Discussion started: 28 March 2024

Label	Mathematical expression and Description	Nb. stations	Date range	Colour
GridClim-ref	$CI_{ori}[\mathbf{X}_c^{\text{GridClim}}(Reference\,network)]$, arithmetic average for all GRIDCLIM grid-points ∈ reference network	39	1961-2018	black
GridClim-sub	$CI_{ori}[\mathbf{X}_c^{\text{GridClim}}(Calibration\,network)]$, arithmetic average for all GRIDCLIM grid-points ∈ calibration network	338	1961-2018	yellow
GridClim-all	$CI_{ori}[\mathbf{X}_c^{\text{GridClim}}(Gridpoints\,over\,Sweden)]$, arithmetic average for all GRIDCLIM grid-points ∈ Sweden	69,842	1961-2018	green
EOF-cal	$CI_{EOF}[\mathbf{X}_c^{MORA}(Calibration\,Network)]$, EOF-based method applied to the gap-filled SMHI-MORA data for the calibration network	338	1961-2018	cyan
SMHI-ref	$CI_{ori}[\mathbf{X}^{\star}(Reference\,network)] - \overline{CI_{ori}}\vert_{1961-2018}$, original SMHI indicator, centered to its 1961–2018 mean, for MORA data in the reference network	39	1860-2020	grey
EOF-rec	$CI_{EOF}[\mathbf{X}^{\star\,MORA}_c(Calibration\,Network)]$, EOF-based method applied to the SMHI-MORA data for the calibration network	338	1860-2020	red

Table 3. Description of the labels for Figures with corresponding label and colour. The number of stations for the reference and calibration networks are indicated for temperature. The mathematical formalism follows the definitions in the text, where $\mathbf{X}_c^{MORA}$ refer to the centered SMHI-MORA data-set, and $\mathbf{X}_c^{\text{GridClim}}$ refer to the gridded GRIDCLIM data-set.

3.2.1 Comparing SMHI-ref and GridClim-ref over 1961 – 2018

Before comparing ***CI*** estimators with a larger network (GridClim-sub and GridClim-all) with GridClim-ref, let us evaluate GridClim-ref versus SMHI-ref, operationally used by SMHI. The method to compute the GridClim-ref ***CI*** is similar, but not identical to the SMHI-ref ***CI***. To be consistent with EOF estimates, the GridClim-ref is based on the longest single time-series in the MORA data-set, rather than the coupled 'pseudo-station' used in SMHI-ref (in which nearby stations are 'stiched together' to obtain longer continuous time-series, as explained in Joelsson et al. (2023). Furthermore, the centered climate indicator for GridClim-ref (and all other ***CI*** computed in the current study) is computed as the spatial mean of centered station time-series; on the other hand, SMHI-ref is computed as the arithmetic mean of absolute station observations, from which the ***CI*** mean over the 1961–2018 calibration period was subtracted (according to Eq. (9)), since individual station data used for SMHI-ref calculation were no longer available.

The fact that SMHI-ref is computed with absolute values from observation stations makes it more sensitive to inhomogeneities: if, despite the coupling procedure, a given station is interrupted, the computed ***CI*** will be affected, especially if the multi-annual mean temperature of this station departs significantly from the ensemble mean.

https://doi.org/10.5194/egusphere-2024-582
Preprint. Discussion started: 28 March 2024

Fig. (6) enable to look in more details at systematic differences between GridClim-ref and SMHI-ref representing ΔCI, the departure from SMHI-ref (according to Eq. (9)). ***CI*** estimates by SMHI-ref and GridClim-ref agree within at least a $\Delta CI_{\text{temp}} \in \pm 0.1°\text{C}$ for temperature over the 1961–2018 calibration period. This sets a first constraint on the ***CI*** accuracy, thus a metrics for the performance of EOF-based ***CI*** estimates.

3.2.2 How does spatial sampling affect the 1961 – 2018 climate indicator ?

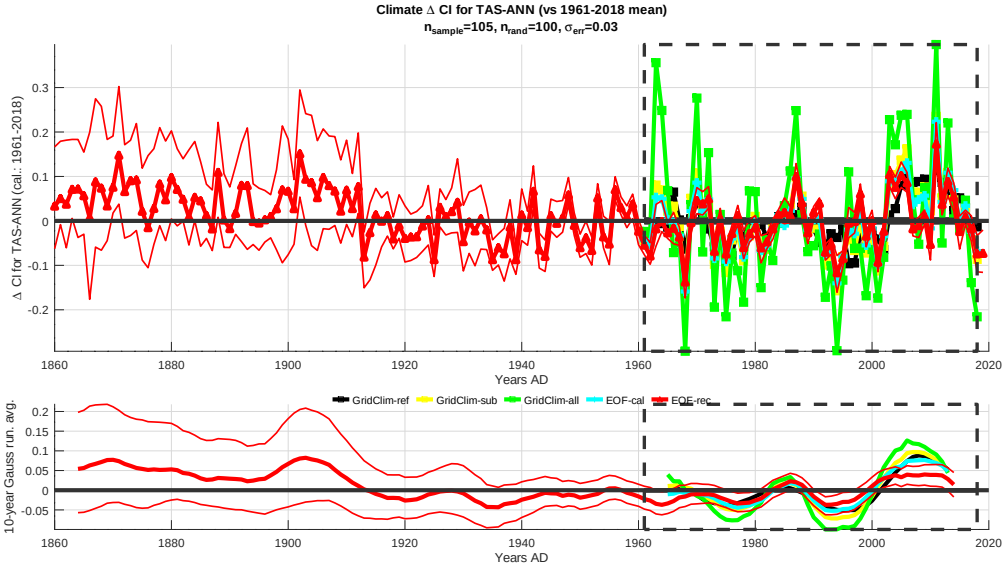

Figure 6. ΔCI, i.e. departures from centered original SMHI indicator, for **temperature** $\Delta CI[\mathbf{X_c}] = CI[\mathbf{X_c}] - CI^{centered}_{\text{SMHI-ref}}$. Labels are identical to Figure Fig. (6) .

Similar to SMHI-ref, we calculated 3 ***CI*** as arithmetic means: GridClim-ref (discussed in the previous section, with GRIDCLIM values at the same locations as the SMHI-ref network), GridClim-sub, with GRIDCLIM values at locations corresponding to the MORA calibration network, and finally GridClim-all, using all GridClim grid-points covering the Swedish territory (with the same resolution as Fig. (2)).

CI estimators using the calibration network of 338 stations for temperature all show minor, synchronised departures from SMHI-ref, regardless of their computation method: arithmetic for GridClim-sub (yellow), EOF-based for EOF-cal (cyan, with gap-filled $\mathbf{X}$ values) and EOF-rec (red, with missing observations $\mathbf{X}^{\star}$). Since their ΔCI is similar to GridClim-ref, we conclude that ***CI*** estimated from the calibration network (1098 stations for precipitation) are equivalent; the difference to SMHI-ref is likely related to the difference of average method (spatial average of absolute values minus ***CI*** mean over 1961–2018 for SMHI-ref, versus spatial average of departure from 1961–2018 for GridClim-ref, GridClim-sub, EOF-cal and EOF-rec).

https://doi.org/10.5194/egusphere-2024-582
Preprint. Discussion started: 28 March 2024

However, we see a significant change for the arithmetic ***CI*** in GridClim-all, compared to GridClim-ref (black line) and GridClim-sub (yellow line): the ***CI*** is computed from all grid-cells covering Sweden (GridClim-all, in green line), stands out with a larger amplitude in inter-annual variability. In other words, annual ***CI*** values with local minima (maxima) i GridClim-sub appear generally lower (higher) in GridClim-all.

GridClim-all, using all 69 842 GRIDCLIM grid-points covering the Swedish territory (c.f. pixel resolution in Fig. (2)) captures more of the inter-annual climate variability than the 39 stations in the SMHI-ref network for temperature. This is consistent with the EOF patterns shown on Fig. (2) and Fig. (3): the 3 dominant modes in Figure Fig. (2) show large variations in remote areas (e.g. Jämtland in central-western Sweden, Norrbotten close to the Finnish border). Such patterns are qualitatively well captured by the calibration network (Fig. (3), upper row): the leading EOF patterns ($\mathbf{eof}^{MORA}(\#1-3)$ are similar to ($\mathbf{eof}^{GridClim}(\#1-3)$, with a strong correlation of their respective time expansion coefficients (lower plots in Fig. (2) and Fig. (3) for temperature).

However the difference in amplitude between the GridClim-all and GridClim-ref can be related to the under-sampled remote areas with large variations. Hence an important conclusion: the definition of a 'national' observation-based climate indicator is dependent on the station network on which it relies. In other words, a national network of stations covering most of the territory is not *per se* a guaranty that all regions are equally represented in the 'national' climate indicator, and may differ from a gridded, land-covering approach. Vose et al. (2014) applied a methodology similar to GridClim-all: a spatially interpolated data-set for temperature and precipitation observations for the contiguous US, with elevation-dependent dependence, proved to make a significant difference the previous ***CI*** version.

The similarity between GridClim-ref, with 39 stations for temperature, and GridClim-sub with 338 stations, supports the sub-sampling method used for the ***CI*** robustness (Equation Eq. (8)). Given that a 10-time sub-sampling of carefully selected stations (i.e. the reference network for GridClim-ref compared to the calibration network for GridClim-sub) yield comparable results, we consider that selecting an ensemble of GridClim-sub subsets, each consisting of a randomly chosen subset with 3-time as many stations as GridClim-ref, is reasonable to evaluate the ***CI*** robustness.

In summary, the difference between GridClim-ref and GridClim-sub for temperature on Fig. (6) are less than $0.1°\mathrm{C}$, i.e. smaller than discrepancies between GridClim-ref and SMHI-ref. However, ΔCI for GridClim-all significantly larger (up to $0.3°\mathrm{C}$), which also shows in absolute values in Fig. (8).

3.2.3 Comparing EOF-estimated climate indicators to the SMHI reference indicator for 1860–2020: temperature

Earlier discussed results illustrate how EOF patterns are computed, based on the GRIDCLIM data-set (with grid-points covering all of Sweden) or the MORA calibration data-set (selected observation series). By construction, the EOF represent the dominant modes of variability over the calibration period 1961–2018.

As a reminder, $\mathbf{X_c}^{\star}$ represents a matrix of temperature observation with missing values ; $\mathbf{X_c}$ indicates that observations are centered, i.e. the station average over the calibration period 1961–2018 $\overline{\mathbf{X}_{cal.}(i)}$ has been subtracted, according to Eq. (3). $\widehat{\mathbf{A}}$ the estimator of its time expansion coefficients, from which the original data-set $\mathbf{X_c}^{\star}$ can be estimated as $\widehat{\mathbf{X_c}^{EOF}}$, according to Eq. (6).

https://doi.org/10.5194/egusphere-2024-582
Preprint. Discussion started: 28 March 2024

The EOF patterns $\mathbf{eof}^{\mathbf{MORA}}(\#\mathbf{1}-\mathbf{10})$, established over the calibration period 1961–2018 on the MORA calibration data-set, allow to reconstruct a climate indicator for study period 1860–2020: available observations prior to 1961 are projected on the respective $\mathbf{eof}^{\mathbf{MORA}}(\#\mathbf{1}-\mathbf{10})$ patterns to obtain the estimated observation matrix $\widehat{\mathbf{X}^{\mathbf{MORA}}}$, $\widehat{\mathbf{X}^{\mathbf{MORA}}}$ is then spatially averaged to define the estimated $\mathbf{CI_{EOF}}$, according to Eq. (7).

In the present section, we focus on the reconstruction of the climate indicator over the full experiment period 1860–2020. The centered observational MORA data-set $\mathbf{X_c^\star}$, including its missing values, is projected onto the EOF pattern (Equation Eq. (6) for EOF to produce an estimate of the complete data-set $\widehat{\mathbf{X}_c}^{EOF}$. The latter is used in Equation Eq. (7) to provide the ***CI*** based on the EOF result. Note that only values originally available in the MORA data-set $\mathbf{X_c^\star}$ are used, hence the denominator in Equation Eq. (7). In other words, the EOF method performs a weighted mean of time-series in the MORA data-set $\mathbf{X_c^\star}$, with weights determined from the spatial variability established during the 1961–2018 calibration period. In particular, we discuss following ***CI*** reconstructions over the full experiment period 1860–2020:

Fig. (8) presents the reconstructed ***CI*** for temperature EOF-rec and SVD-rec, compared to the original SMHI-ref (in grey). At first glance, there is a very good agreement between SMHI-ref and the EOF-based reconstructions EOF-rec and SVD-rec, both for inter-annual (upper panel) and decadal variability (lower panel). This result is consistent with the high portion of variance explained by the first mode of variability: $\lambda(\#1) > 80\%$ in Table (2). The first spatial EOF pattern $\mathbf{eof}^{MORA}(\#1)$ (shown on Figures Fig. (2), Fig. (3), upper left panel) is uni-modal (by construction), and furthermore fairly uniform: the ***CI*** reconstruction from a limited number of stations (as for SMHI-ref) is therefore likely to be very close that from a larger station network (as for EOF-rec). The same conclusion holds for the SVD reconstruction ($\mathbf{svd}^{MORA}(\#1)$, Figure Fig. (A2)).

In other words, we consider the EOF method for estimating the temperature ***CI*** as fully consistent with SMHI-ref, currently used by SMHI. The 10-year Gaußian moving average (lower panel) further indicates that decadal variability is portrayed by all different ***CI***: a largely continuous warming from 1860–1935, followed by an abrupt cooling in the early 1940s. Temperatures across Sweden remain variable until 1980. The 1980–2020 period shows a consistent warming trend (the strongest on instrumental records), with a $2°\mathrm{C}$ warming over the last 40 years.

Equation Eq. (8) illustrates a metrics for the robustness of the computed ***CI***: it takes into account the impact of measurement error (symbolised by the normally distributed noise ε added) and sub-sampling. Since only available measurements (i.e. non-NaN elements in the $\mathbf{X_c^\star}$ matrix) are used in EOF-rec, the 25%–75% percentile range implicitly mirrors the data availability: if less stations are active, e.g. in the early parts of the record, the added noise ε is less likely to cancel out, hence the 25%–75% becomes larger. This phenomenon is most obvious in the 10-year smoothed ***CI*** (lower panel): while the 25%–75% range is small and constant over time during the 1961–2018 calibration period (during which the number of active stations does not vary significantly, cf. Fig. (1)), it increases gradually for earlier periods, reaching up to $0.4°\mathrm{C}$ for the 10-year smoothed ***CI*** prior to 1880. Despite the increasing scarcity of observations prior to 1900, the uncertainty due to measurement error and geographic sampling appears smaller than the climate signal in annual and decadal temperature ***CI***. Calculating the temperature ***CI*** from at most 338 instead of 39 stations thus does not introduce a systematic bias.

On Fig. (6), ΔCI values for EOF-rec display 3 large outliers (in 1864, 1871 and 1879), by up to $0.75°\mathrm{C}$. Given the small number of stations active at the time, it is difficult to conclude if such discrepancies are significant, nor which of the EOF-rec

https://doi.org/10.5194/egusphere-2024-582
Preprint. Discussion started: 28 March 2024

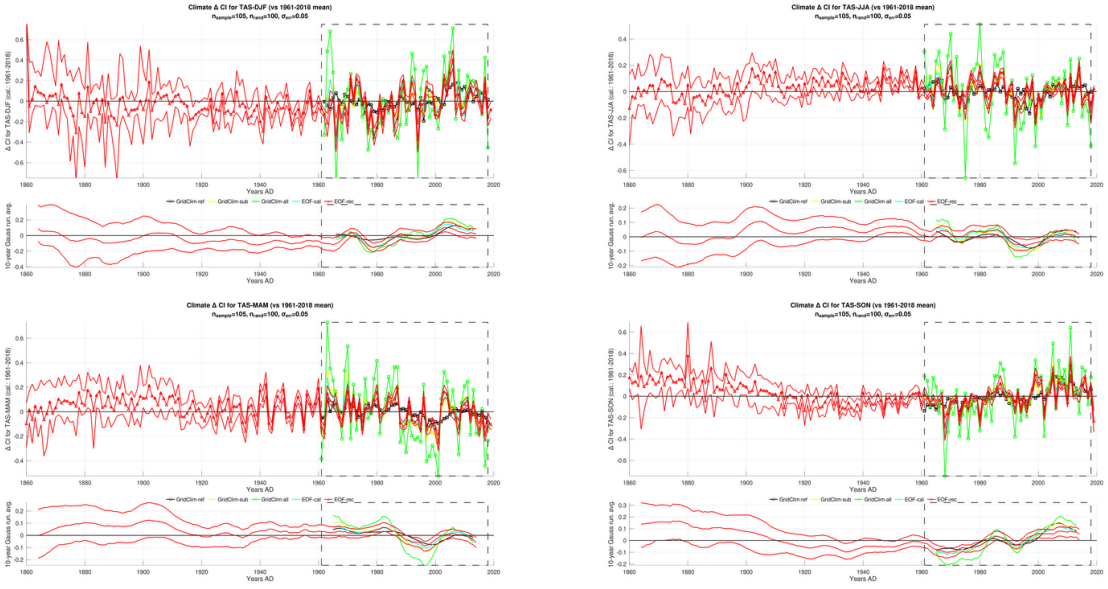

Figure 7. ΔCI, i.e. departures from centered original SMHI indicator, for **temperature** $\Delta CI[\mathbf{X_c}] = CI[\mathbf{X_c}] - CI^{centered}_{\text{SMHI-ref}}$. Labels are identical to Figure Fig. (6) .

and SMHI-ref values are physically more relevant. It is worth noticing that the 3 outliers occur for the coldest years in 1860–1880. This is not surprising, since the spread of extreme cold temperatures across Sweden is larger than for warm extremes: the inclusion (or not) of a station with a strong negative anomaly may therefore have a large impact on the ***CI*** value. It is furthermore interesting to notice that the EOF-rec 25%-75% percentile envelope includes the SMHI-ref value, for all 3 outliers: this further confirms the hypothesis that such large punctual discrepancies might be related to a single station observation, which becomes pregnant when only few stations are active.

3.2.4 Seasonal EOF reconstructions of the climate indicator 1880–2020: temperature

Fig. (7) presents EOF-rec, the EOF-based reconstruction of temperature ***CI*** for seasonal averages. Since it is beyond the scope of this study to investigate *per se* the differences in the seasonal ***CI***, we focus here on the differences between EOF-rec and SMHI-ref, i.e. ΔCI_{EOF}, presented in Fig. (7).

When analysing seasonal EOF-rec, one needs to keep in mind that the EOF patterns $\overrightarrow{\mathbf{eof}^{\text{seas}}}$, defined in Eq. (5), are computed independently over the 1961–2020 period for each seasonal average $\mathbf{X}_c^{\text{seas}}$. Hence the difference for the annual $\Delta CI^{\text{ANN}}_{EOF}$ cannot be obtained as an average of seasonal $\Delta CI^{\text{seas}}_{EOF}$. In other words:

$$\forall iSeas = \text{DJF, MAM, JJA, SON}, \Delta CI^{\text{ANN}}_{EOF} \neq \frac{\sum \Delta CI^{iSeas}_{EOF}}{4} \tag{10}$$

https://doi.org/10.5194/egusphere-2024-582
Preprint. Discussion started: 28 March 2024

The largest discrepancies between EOF-rec and SMHI-ref for the annual indicator, shown in Fig. (6), were found prior to 1880: ΔCI_{EOF} for individual years sporadically exceed 0.5 °C for the national climate indicator. As a consequence of such spurious differences, the decadal Gaußian average prior to 1880 does not express any physical signification.

The result is comparable for seasonal $\Delta CI_{\text{EOF-rec}}$ in Fig. (7). In general, seasonal ΔCI_{EOF} display discrepancies similar to the annual ***CI*** (± 0.5°C). GridClim-all displays the largest discrepancies found during the 1961–2018 calibration period (± 0.75°C). A conclusion can be drawn from the discussion of EOF patterns in Fig. (2) versus Fig. (3): the leading modes of variability for the complete GRIDCLIM data-set reveal more details, which are not all captured in the MORA calibration network. Thus, for some individual years where under-sampled region weigh strongly in the national mean, the $\Delta CI_{\text{GridClim-all}}$ will display larger absolute values.

The prime objective of the EOF method was to better capture the climate variability, and thereby be less sensitive to the observation network configuration. Fig. (7) demonstrates that the EOF-method delivers a national ***CI*** which is more representative for all of Sweden that the reference method: while EOF-rec does not capture all the amplitude of $\Delta CI_{\text{GridClim-all}}$ signal (which cannot be expected from EOF-rec, based on the calibration network coverage), the $\Delta CI_{\text{EOF-rec}}$ displays identical signs to $\Delta CI_{\text{GridClim-all}}$.

The robustness of the ***CI*** reconstruction is assessed by adding random noise and selecting a sub-sample of $\widehat{\mathbf{X}_c}$, according to Eq. (8). Thin lines in Fig. (7) represent the 25% and 75% of a 100-member ensemble of random realisations. The span between the 25% and 75% percentiles increases the further back in time the ***CI*** is computed. This is consistent with the evolution of the number of active stations shown in Fig. (1), upper sub-figure: the fewer stations are active for a given year, the more will random noise affect the $\Delta CI_{\text{EOF-rec}}$.

For all seasons, the 25% and 75% percentiles of $\Delta CI_{\text{EOF-rec}}$ include the $\Delta CI = 0$ line (i.e. X-axis), at least for the Gaußian 10-year running average. In other words, there is a 50% probability that EOF-rec does not significantly differ from SMHI-ref – according to Eq. (8), with $\varepsilon_T^{\text{seas}} = 0.053$°C measurement accuracy for seasonal averages (corresponding to a fairly conservative estimate of a daily measurement precision $\varepsilon_T^{\text{seas}} = 0.5$°C).

The 10-year Gaußian filter (i.e. lower row in sub-plots in Fig. (7)) for seasonal $\Delta CI_{\text{EOF-rec}}$ shows a limited, albeit persistent negative bias for DJF and SON temperatures for the 1915–1970 period. On the other hand, JJA temperatures have positive departures over 1905–1960.

The 1900–1960 period corresponds to a steady expansion of MORA network for temperature (Fig. (1), upper plot): the number of active stations increases from less than 100 in 1900 to over 400 in 1960. During this period, the station density has improved in northern Sweden, as illustrated by the 75% percentile of active stations' latitude (Fig. (1), lower plot).

As illustrated in Fig. (4), the leading EOF pattern $\mathbf{eof}_{DJF}^{\text{GridClim}}(\#1)$ has its highest values in northern Sweden; the same holds for $\mathbf{eof}_{SON}^{\text{GridClim}}(\#1)$. Hence, a higher station density in northern Sweden, as captured by EOF-rec, can explain the discrepancy with SMHI-ref (whose station network has remained fairly constant during 1900-1960: black line in Fig. (1), upper plot).

The situation is different for summer (JJA) temperature. $\mathbf{eof}_{JJA}^{\text{GridClim}}(\#1)$ does indeed slightly higher values in mountain regions (Jämtland, Sarek region), but differences with central Sweden are not as pronounced as for DJF and SON. However, the second mode $\mathbf{eof}_{JJA}^{\text{GridClim}}(\#2)$ – characterised by a clear North-South dipole, explains more than twice of the signal variance

https://doi.org/10.5194/egusphere-2024-582
Preprint. Discussion started: 28 March 2024

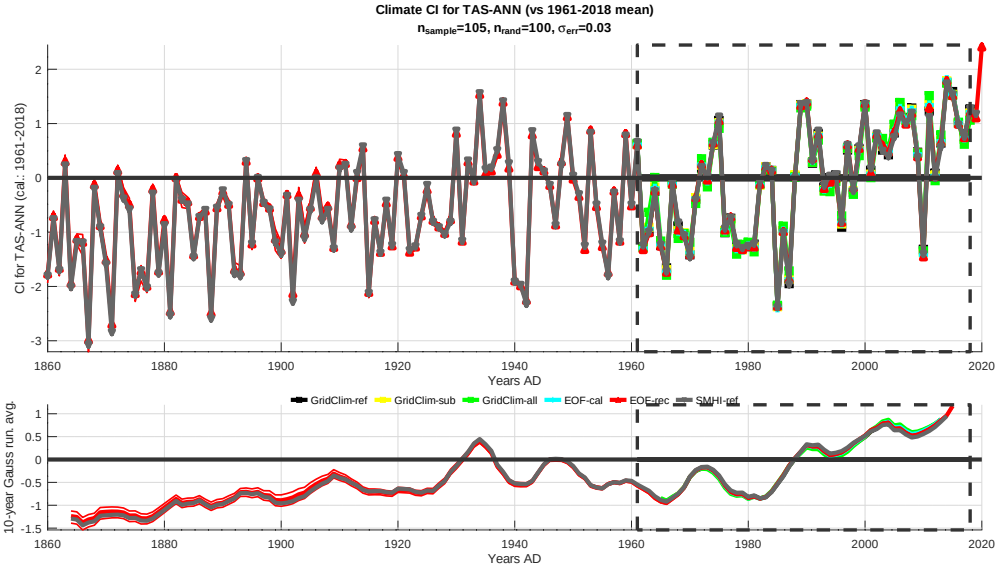

Figure 8. Estimates of the annual climate indicator ***CI*** for **temperature** with various methods. The ***CI*** is represented as thick lines, whereas the 25% and 75% percentiles of the robustness ensemble CI_{ens} are shown as thin lines of the same colour. The labels for various ***CI*** are indicated in Table Table (3) .

($\lambda_{JJA}(\#2) = 13\%$) than for DJF and SON. The stronger influence of the second (bipolar) mode thus explains the discrepancy between EOF-rec and SMHI-ref.

For all seasons, $\Delta CI_{\text{EOF-rec}}$ exhibits large values. However, the associated span of 25%–75% percentiles indicate that this results may not be significant. Therefore, they are not discussed in detail in the present study.

4 Discussion

4.1 Benefits of the EOF-based climate indicators

The EOF methods were designed to emulate the original the original SMHI climate indicators (hereafter referred to as ***CI***) for temperature. Indeed, the resulting ***CI*** reconstruction (referred to as EOF-rec) are in close agreement with the original ***CI***, computed as the arithmetic mean of selected observation time-series (referred to as SMHI-ref).

The present study introduces the EOF method as an alternative ***CI*** estimate: instead of deriving the ***CI*** as an *arithmetic* mean of available time-series, it is computed as a *weighted* mean. The corresponding weighting coefficients are established using independent (i.e. orthogonal) modes of variability in the observations during the calibration period. The weighing coefficients are derived from EOF (empirical orthogonal functions) patterns, noted $\mathbf{eof}$.

https://doi.org/10.5194/egusphere-2024-582
Preprint. Discussion started: 28 March 2024

Figure 9. Reconstructed seasonal absolute ***CI*** for **temperature** and the centered original SMHI indicator $\Delta CI[\mathbf{X_c}] = CI[\mathbf{X_c}] - CI^{centered}_{\text{SMHI-ref}}$. Labels are identical to Figure Fig. (8) .

We choose a calibration period extending from 1961–2018, which corresponds to the period covered by the gridded climate data-set GridClim. From a theoretical point of view, using a weighted ***CI*** method makes its estimates less sensitive to changes in observation availability over time, i.e. how the spatial density of stations in the observation network evolves over time. Such an automated method also allows to use the full observation database, instead of solely hand-selected stations in the reference network.

Our results confirm that SMHI-ref ***CI***, based on a reference network of 39 stations for temperature, is well representative for a larger ensemble of 338 temperature stations, referred here as calibration network, used for EOF-rec. In other words, SMHI-ref and EOF-rec present largely equivalent results for the ***CI*** estimation, provided that sufficient meteorological stations are active, i.e. after 1901. This is a novel result *per se*, since the reference network used in SMHI-ref had not been thoroughly evaluated against the full observation database.

Based on the EOF method, we also suggests a metrics to evaluate the range of likely ***CI*** values. By adding random noise and sub-sampling in an ensemble computation, we can plot the 25% and 75% percentiles in addition to the median value for the complete data-set. The spread between the 25% and 75% percentiles reflects mainly the data availability, and becomes significantly larger prior to 1900.

https://doi.org/10.5194/egusphere-2024-582
Preprint. Discussion started: 28 March 2024

Furthermore, the EOF analysis enables to visualise the dominant modes of variability in the MORA observation data-set, and compare it to the gridded data-set GRIDCLIM. The EOF patterns illustrate that the seasonal variability for temperature is significantly different from annual means.

Finally, the EOF-method has been tested in the present study for ***CI*** temperature. Since this method is fully automated, it lends itself in principle to other climate variables for which SMHI provides an aggregated climate indicator, e.g. precipitation (as presented in Sturm (2024a)), snow depth or length of the growing season.

4.2 Implication for decadal climate variability over Sweden

Despite being relatively small ($\pm 0.1°C$), differences between the EOF-based ***CI*** EOF-rec and the reference SMHI-ref are relevant for the analysis of decadal climate variability over Sweden, especially for seasonal means (as shown in Fig. (8) and Fig. (9)).

Fig. (6) and Fig. (7) suggest that SMHI-ref may underestimate spring (MAM) and autumn (SON) temperatures by $0.1°C$ prior to 1920. During summer (JJA), SMHI-ref appears to underestimate temperature by $0.05°C$ for the 1900–1960 period. On the other hand, temperatures in winter (DJF) and autumn (SON) estimated by EOF-rec are $0.1°C$ lower than SMHI-ref.

During the 1961–2018 calibration period, the warming trend for mean annual temperature over Sweden is estimated by SMHI-ref to $+1.9°C$ between 1980 and 2020. EOF-rec suggests a stronger trend: $+2.1°C$. The trend for winter (DJF) temperatures is more pronounced: $+3.5°C$ for SMHI-ref, which increases to $+3.9$ in EOF-rec. On the other hand, EOF-rec suggests a smaller temperature increase for the 1960–2020 for spring (MAM), and to a lesser extent for summer (JJA). The difference between the SMHI-ref and EOF-rec are predominantly related to spatial sampling: the larger number of observation stations used for EOF-rec as compared to SMHI-ref (summarised in Table (1)) enables to better represent climate variability in Sweden's northern regions. Since northern Sweden experiences a stronger recent warming (in connection to polar amplification (Masson-Delmotte et al., 2021; Gulev et al., 2021)), EOF-rec captures a stronger temperature increase than SMHI-ref. The latter is even more pronounced for GridClim-all, which calculates the ***CI*** for all grid-points over Sweden for the 1961–2018 period. SMHI-ref is thus likely to represent a smaller amplitude for decadal climate variability (in particular lesser trend for recent warming) compared to other national indicators computed from gridded data-sets Vose et al. (2014).

4.3 Comparison to climate indicators in other countries

Vose et al. (2014) presents the climate division data-set for the conterminous US for temperature since 1895. Station data is interpolated using thin-plate smoothing spline method, accounting for invariant predictors such as elevation. A national indicator can be derived from area-weighted average of "division units" constituting the grid.

The method for deriving a area-weighted climate indicator for the contiguous is analog to the GridClim-all ***CI*** presented in this study, i.e. the mean of all grid-points in the GRIDCLIM data-set covering Sweden.

Furthermore, Vose et al. (2014) indicate that a version 2 of the US ***CI*** led to overall lower estimates of mean monthly temperatures compared to version 1. During 1960–2014, the difference between the old and new version is fairly constant

https://doi.org/10.5194/egusphere-2024-582
Preprint. Discussion started: 28 March 2024

at $-0.4°\mathrm{C}$. The difference steadily increases prior to 1960, reaching $0.9°\mathrm{C}$ in 1895. The difference in monthly temperatures between both data-sets displays a distinct seasonal cycle.

Vose et al. (2014) indicate that changes in the temperature ***CI*** is primarily related to better sampling (i.e. increased density of observation stations). This conclusion is in agreement with this study, illustrated in Fig. (2): high-elevation regions in Sweden display a particular behaviour in $\mathbf{eof}^{GridClim}(\#1)$, which is under-sampled in the MORA network EOF-rec and absent in the reference network SMHI-ref. Hence, an area-weighted ***CI*** over Sweden is likely to yield significantly different results than the current station-based ***CI***, both for annual and seasonal estimates.

Meteo-Swiss Begert et al. (2005) provides a set of 12 homogenised monthly temperature series since 1864. Climate indicators are available as time-series for each reference observation station, or as interpolated maps over the Swiss territory (since 1871). The Swiss meteorological service adopts a hybrid approach (maps and individual observation time-series), which differs from SMHI's aggregated national climate indicator.

Kaspar et al. (2017) presents a review of climate observation in Germany since 1881. Available observations are interpolated on a 1 km x 1 km grid, which allows for altitude-dependent correction of the considered climate variable.

In the United Kingdom (Office, 2024), climate indicators are available for England, Wales, Scotland and Northern Ireland, as well as the entire UK since 1881, based on the HadUK-grid. In France (Meteo-France, 2024), monthly indicators are available for each department since 1852, as well as long homogenised time-series for selected observation stations.

This brief overview reveals that most national weather agencies adopt a hybrid approach to produce climate indicators. Long, homogenised time-series provide a long-term base-line for climate change in time, which is complemented by gridded data-sets to represent its spatial variability. The present study shows the relevance of the EOF method to better capture the spatial of climate variability, as a first step towards delivering region-specific climate indicators. The homogenisation of climatological observations is an ongoing task at SMHI (Joelsson et al., 2022, 2023).

4.4 Comparison to studies of historic climate variability in Fennoscandia

Bergström and Moberg (2002) present daily air temperature and pressure series for Uppsala for 1722–1998, located in Central Sweden (70 km north of Stockholm). In accordance with this study, Bergström and Moberg (2002) report lower maximum daily temperatures between 1880-1900 and 1950–1970. It is interesting to notice that the minimum daily temperature does not display similar variations, but a strong positive trend since 1980. Furthermore, the daily temperature amplitude (i.e. the difference between daily maximum and minimum temperatures) shows a mostly continuous decrease since 1880, particularly marked after 1980. The current study focuses on annual and seasonal means of mean daily temperature; conclusions by Bergström and Moberg (2002) provide a rationale to apply the EOF methods to annual/seasonal means of daily maximum and minimum temperatures and temperature amplitude: these complementary ***CI*** could illustrate the climate variability in Sweden in more details.

Wilcke et al. (2020) set the long-lasting high-pressure systems that occurred in summer 2018 over Sweden, causing extreme warm and dry conditions, in historical context. Wilcke et al. (2020) used as reference the Stockholm temperature series (Moberg and Bergström, 1997; Moberg et al., 2003) spanning 263 years (1756–2018); the MORA data-set of temperature observations

https://doi.org/10.5194/egusphere-2024-582
Preprint. Discussion started: 28 March 2024

1860–2018, the same as the one used in this study, was also used, along with 5 long (150-year) climate simulation by CMIP global climate models. Temperature anomalies ΔT were calculated for the periods 1861–1890 and 1989-2018, using the 1981–2010 as deference. The probability density functions for summer temperature anomalies show a distinctive difference between 1861–1890 and 1989–2018: for the southern half of Sweden, the shift in median values amounts to $\Delta T = +1\,^{\circ}\mathrm{C}$, and $\Delta T = +1.5\,^{\circ}\mathrm{C}$ for the northern half of Sweden, which is consistent with the $\Delta T = +1.25\,^{\circ}\mathrm{C}$ in Fig. (9). These results are consistent with a recent attribution study of extreme events in Sweden, using temperature observations in Sweden since 1882 (Holmgren and Kjellström, 2023). Wilcke et al. (2020) further emphasises that summer temperature changes are larger in the northern Sweden than in its southern half, which is consistent with Fig. (4): the analysis of extreme events throughout the 1860–2020 period therefore requires a detailed representation of the temperature patterns, which the EOF method is more likely to represent than the SMHI reference (arithmetic) method.

4.5 Persistence of leading EOF modes

Kjellström et al. (2022) present a large-scale atmospheric clustering affecting Scandinavian temperature and precipitation between two climate normals (1961–1990 and 1991–2020). The frequency of circulation types display significant changes between 1961–1990 and 1991-2020; the change in occurrence frequency differs also for individual months. They also report strong inter-annual correlations between the North Atlantic Oscillation (NAO) index and temperature ***CI***, in particular for winter (DJF) with an $R^2 = 60\%$; the ***CI*** for summer (JJA) temperature however is much smaller $R^2 = 5\%$. For precipitation, the winter (DJF) ***CI*** has the strongest correlation ($R^2 = 20\%$). It is worth noticing that the correlation coefficients between the NAO-index and the national ***CI*** for temperature, for each season, are very close between the 1961–1990 and 1991-2020 period.

These results are highly relevant for the present study. Since the occurrence of various circulation types are proven to differ between the 1961–1990 and 1991–2020 normal periods, it is reasonable to assume that the circulation type distribution is also different for previous 30-year periods. The underlying assumption for this study is that the EOF pattern, calculated over the 1961–2018 calibration period, is deemed sufficiently persistent to be applied over the entire 1860-2020 study period. Further research is thus needed to evaluate if the leading EOF modes $\mathbf{eof}(\#1-10)$ for temperature vary significantly over 30-year periods. In that case, the EOF method could be adapted by identifying a predictor representing circulation types (e.g. based on mean sea-level pressure reconstructions since 1850 (Allan and Ansell, 2006; Ansell et al., 2006; Gallego et al., 2005)). A particular set of EOF patterns $\mathbf{eof}(\#1-10)$ can be assigned to dominant clusters of the predictor, which could be used for ***CI*** reconstruction EOF-rec for earlier periods.

The North Atlantic Oscillation (NAO) is commonly accepted as the dominant circulation type affecting climate variability in Sweden, in particular during winter. The fact that correlation R^2 between the NAO-index and temperature is similar between 1961–1990 and 1991-2020 indicate that, despite a change in the occurrence of circulation type, the relation between NAO and temperature/precipitation ***CI*** is persistent. In other words, while the circulation types change, the internal variability (i.e. the leading EOF patterns) remains mostly unchanged.

Balting et al. (2021) show, using and EOF analysis on $\delta^{18}O$ from tree cellulose across Europe, that the dominant modes of climate variability are stable since 1850. Furthermore, the second EOF, resembling features shown in Fig. (A3), is associated

https://doi.org/10.5194/egusphere-2024-582
Preprint. Discussion started: 28 March 2024

to summer European patterns. Hence, it would be interesting to investigate if $\mathbf{eof}^{MORA}(\#2)$ in Fig. (A3), can specifically be related to blocking situations over Sweden, which are associated to extreme conditions (e.g. during summer 2018).

Observation data presented in this study can be used to assess the correlation of circulation indices with temperature time-series over Sweden, for annual and seasonal averages, over 30-year time-slices between 1860 and 2020. Several reconstructions of NAO since 1850 are available (Cropper et al., 2014, 2015; Comas-Bru and Hernández, 2018; Hanna et al., 2022; Jacobeit et al., 2001b; Slonosky et al., 2000; Zveryaev, 2006), or the last millennium (Gouirand et al., 2007; Jones et al., 2001). Comas-Bru and Hernández (2018) also provides estimates of the East Atlantic (EA) and Scandinavian (SCA) circulation patterns since 1851 to present. We however would advise that observations in the MORA database be processed at a monthly resolution (instead of annually resolved ANN, DJF, MAM, JJA, SON averages) for such an analysis.

Jacobeit et al. (2001a, 2003) present zonal circulation indices (NAO: North Atlantic Oscillation, CEZ: Central European Zonal Index) for Europe over the 1780–1995 period. The NAO index shows a period until the 1850s with accumulating negative anomalies in winter whereas, in summer, positive are prevailing during 1820–1910. This is in contrast with recent evolution with positive anomalies prevailing during winter (1920–1970) and negative ones during summer (1970–1995).

Numerous studies have assessed the impact of synoptic circulation (mostly NAO) on precipitation and temperature patterns across Scandinavia since 1850 (Hänsel, 2020; Ansell et al., 2006; Gallego et al., 2005; Hanna et al., 2022; Alvarez-Castro et al., 2018; Fleig et al., 2015; Jones and Mann, 2004; Philipp et al., 2007; Slonosky et al., 2000; Kyselý, 2007, 2008). While it is beyond the scope of the present study to assess the dependence of observed temperature to synoptic circulation patterns, it would be interesting to investigate whether the leading EOF modes $\mathbf{eof}^{MORA}$ differ significantly for e.g. positive/negative phase composites of the NAO.

Such an analysis would however require to analyse the temperature and precipitation observations at monthly resolution, instead of annually resolved (ANN, DJF, MAM, JJA, SON) as in the present study: Massei et al. (2007), based on a wavelet analysis of the daily NAO index, demonstrate that a large portion of the power-spectrum of NAO variability is found at sub-annual frequencies.

5 Conclusions

The present study introduces a new method to aggregate observation time-series into a single climate indicator, which is evaluated for temperature over Sweden. A prerequisite for the new method is to emulate the climate indicator operationally used by the Swedish Meteorology and Hydrology Institute (SMHI), based on an arithmetic mean of all station time-series.

5.1 Added value of station coupling and homogenisation

The present study focuses on the individual station measurements in the MORA database maintained by SMHI. On one hand, it has the advantage of considering all available observations; on the other hand, it comprises only few station with long time-series. SMHI Climatology experts (including the authors of the SMHIref indicator (Alexandersson, 1986)) have used the 'coupling' technique to stitch together similar station records into a long, single 'pseudo-station'; however, such a process has

https://doi.org/10.5194/egusphere-2024-582
Preprint. Discussion started: 28 March 2024

long been a time-intensive, person-operated effort; it has thus been restricted to a limited number of carefully chosen stations (e.g. the SMHI reference network). A recent development by Joelsson et al. (2022) enables an automatic, objective coupling routine. The present methodology can be improved by using the coupling routine prior to applying the EOF method.

The quality control applied to all observations stored in the MORA database aims at insuring that the instrumental value is correctly digitised; however it does not account for biases related to changes in station location, instruments, surroundings etc. An *a posteriori* correction can be performed using homogenisation, such as BaRT/Homer (Joelsson et al., 2023) or Climatol (Guijarro et al., 2023), both homogenisation tools being used at SMHI. Homogenisation is particularly recommend when working with coupled records, since the 'stitching' is likely to introduce variations that are not related to climate, hence considered inhomogeneities.

Further developments of the present method will investigate the impact of coupling and homogenisation on instrumental observations on the estimation of the climate indicators. An automated method for station coupling (Joelsson et al. (2022, 2023)) enables an optimal compromise between the maximum number of different time-series and the longest times-series. The coupling procedure will increase the number of time-series retained in the calibration data-set EOFcal that extend to the early parts of the record. in other words, the light grey line in Fig. (1) is likely to be closer to the total number of active stations (represented as staples). Furthermore, the number of missing values during the 1961–2018 calibration period is likely to be reduced, hence the gap-filling procedure (prior to the EOF analysis) is less likely to introduce numerical artefacts in the indicator reconstruction.

However, possible inhomogeneities (introduced by the coupling and/or other events, e.g. station relocation) will remain in the calibration data-set EOFcal. These can be removed with the homogenisation procedure. This additional step will illustrate to which extent the national climate indicator is sensitive to (presumably randomly distributed) inhomogeneities.

Finally, the homogenisation toolboxes Climatol (Guijarro et al., 2023) and Bart/HOMER (Joelsson et al., 2023) include a gap-filling feature (based on similarities with neighbouring stations), which potentially can deliver complete time-series over the entire experiment period 1860–2020. It is however questionable to which extend the gap-filling algorithm delivers physically significant results over such long periods, when the station availability drops to low numbers (e.g. in the early part of the record). The present method, in particular the ensemble computation with randomly distributed noise, would be a suitable tool to evaluate the robustness of the indicator estimation depending on the pre-processing steps (station coupling and homogenisation (Joelsson et al., 2022)), the choice of the calibration period and the climate variable (temperature, precipitation and potentially other observational data-sets).

5.2 Further development of the SVD approach

As illustrated in this study, the EOF and SVD methods are largely equivalent in the present objective: defining a ***CI*** primarily based on the MORA station network. However, the SVD method has the potential to be developed further. Most national weather agencies deliver regionalised climate indicators (Vose et al., 2014; Begert et al., 2005; Kaspar et al., 2017; Office, 2024; Meteo-France, 2024).

https://doi.org/10.5194/egusphere-2024-582
Preprint. Discussion started: 28 March 2024

Eq. (A3) only makes use of $\mathbf{SVD}^{MORA}$ (defined in Eq. (A1)) to estimate $\widehat{\mathbf{X}_c}^{SVD}$, the MORA data-set over 1860–2020. The SVD method allows to use the other term in Eq. (A1): the spatial pattern $\mathbf{SVD}^{\text{GridClim}}$ and associated time expansion coefficients $\mathbf{A}^{\text{GridClim}}$. By construction (Björnsson and Venegas, 1997), the SVD time expansion coefficients $\mathbf{A}^{MORA}$ and $\mathbf{A}^{\text{GridClim}}$ are related to each other, with eigenvalue matrix Λ^{SVD} defined in Eq. (A1).

$$\mathbf{A}^{MORA} \cdot \left(\mathbf{A}^{\text{GridClim}}\right)^T = \Lambda^{SVD}$$
$$\mathbf{A}^{\text{GridClim}} = \left(\left(\mathbf{A}^{MORA}\right)^{-1} \cdot \Lambda^{SVD}\right)^T$$

where the notation $\left(\mathbf{A}^{MORA}\right)^{-1}$ refers to a pseudo-inverse, since $\mathbf{A}^{MORA}$ is not a square matrix. Based on this result, the SVD method makes it possible to estimate a reconstructed gridded data-set $\widehat{\mathbf{G_c}}^{\text{GridClim}}$:

$$\begin{aligned}
&\widehat{\mathbf{G}_c}^{SVD} = \widehat{\mathbf{A}}^{\text{GridClim}} \cdot \left(\mathbf{SVD}^{\text{GridClim}}\right)^{-1} \\
\Leftrightarrow\ &\widehat{\mathbf{G}_c}^{SVD} = \left(\left(\widehat{\mathbf{A}}^{MORA}\right)^{-1} \cdot \Lambda^{SVD}\right)^T \cdot \left(\mathbf{SVD}^{\text{GridClim}}\right)^{-1} \\
\Leftrightarrow\ &\widehat{\mathbf{G}_c}^{SVD} = \left(\left(\mathbf{X_c^\star} \cdot \mathbf{SVD}^{MORA}\right)^{-1} \cdot \Lambda^{SVD}\right)^T \cdot \left(\mathbf{SVD}^{\text{GridClim}}\right)^{-1}
\end{aligned} \tag{11}$$

New climate indicators based on $\mathbf{G}_c^{SVD}$ can potentially be used to emulate the GridClim-all over the entire 1860–2020 period, provided that the projection of MORA $\mathbf{X}^\star$ on its SVD patterns is sufficient to capture the (dominant) climate variability over the entire Swedish territory. If that is the case, it would be possible to define regional climate indicators based on $\mathbf{G}_c^{SVD}$, even in regions where no direct measurements are available. This assumption might be hard to fulfil in early parts of the records, where only few stations are available (cf. Fig. (1)), and would require additional research to assess the significance of the method. The present method could extended by using regional climate simulations over the 1860–2020 period in substitution for the single GRIDCLIM data-set used in this study; the principle of applying SVD to combine (even sparse) observations with a gridded data-set remains valid. This is a promising way to characterise Sweden's pre-industrial climate, to the best precision the available observation evidence allows.

This example illustrates that relatively simple linear algebra methods, such as EOF and SVD, have the potential to deliver valuable results for the analysis of climate observations. The analysis can be further refined using more advanced method, such as Principal Oscillation Patterns (POP), as presented in Storch et al. (1995).

Code and data availability. The code and processed data used in this study are freely available on Zenodo (Sturm, 2024b).

The computations for the present study are performed using the open-source software OCTAVE, which is mostly compatible with MATLAB. The code is run with OCTAVE version 7.3.0 on a Linux computer. It further requires following OCTAVE packages:

- *io*, version 2.6.4

https://doi.org/10.5194/egusphere-2024-582
Preprint. Discussion started: 28 March 2024

- *netcdf*, version 1.0.16
- *stat*, version 1.4.3
- *mapping*, version 1.4.1

The OCTAVE code is available freely available at. Since the final processed data used in this study is available (cf. below), the user may be able to run the code without the full list of required OCTAVE packages.

The data, processed as required for this study, is freely available for download. The archive format is OCTAVE's binary format, which can be loaded into an OCTAVE session using the *load* command. The processed data files, including the result of the EOF computation, is organised in separate files for each annual and seasonal means (annual – ANN, and seasonal: winter – DJF, spring – MAM, summer – JJA, autumn – SON).

Author contributions. The current study was designed, performed and written by the first (and only) author.

Competing interests. The author declares that they have no conflict of interest.

Appendix A: Supplementary material

https://doi.org/10.5194/egusphere-2024-582
Preprint. Discussion started: 28 March 2024

A1 Number of seasonally active stations

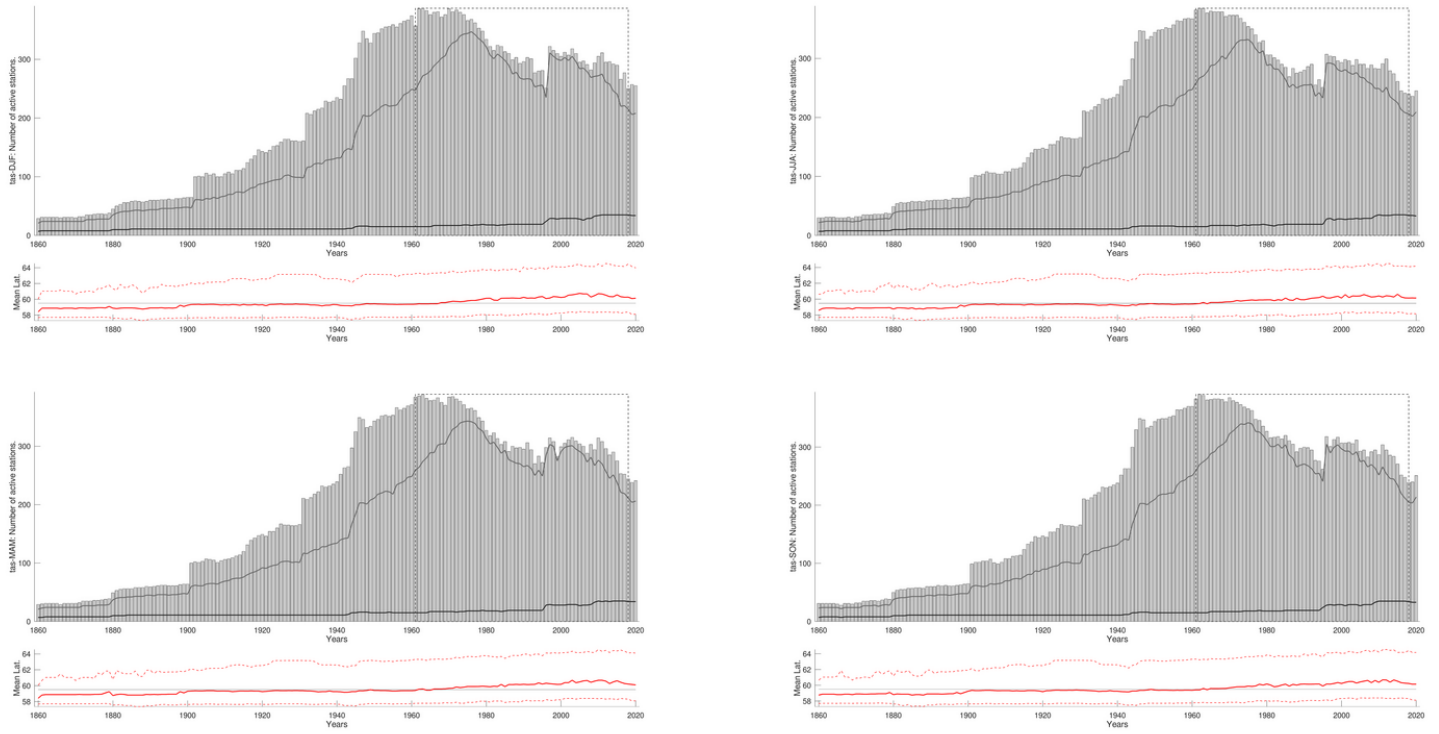

Figure A1. Upper plot: Number of active temperature stations in MORA over time (as bars), for each season (cf. Fig. (1)). The dark grey line represents the number of active in the original reference station network; they light grey line represents the number of stations for the calibration network (i.e. individual stations being active at least 15 years during the calibration period 1961–2018, as highlighted by the dashed box). Lower plot: Median latitude for active stations in the calibration data-set over time (incl. the [25%–75%] bounds). The median latitude is used as a proxy for the distribution of the observation network, in particular its coverage of Sweden's Norther regions.

https://doi.org/10.5194/egusphere-2024-582
Preprint. Discussion started: 28 March 2024

A2 SVD decomposition of the coupled MORA and GRIDCLIM data-sets

In section 2.2.2, the EOF decomposition was described for a single independent data-set (MORA or GRIDCLIM). The EOF method can be further developed using the Singular Value Decomposition (SVD) method described hereafter. In the present study, the leading EOF patterns $\mathbf{eof}(\#1-3)$ were, in all but one case, virtually identical to their SVD counterparts $\mathbf{svd}(\#1-3)$ for both the MORA calibration network and the complete GRIDCLIM data-set.

The "Singular Value Decomposition" (SVD) is a further development of the EOF method (Eq. (4)), where the leading modes no longer express the dominant variability in a single data-set, but the shared variability modes that two data-sets have in common. The SVD requires that both data-sets have the same number of time-steps (i.e. the same number of columns), but may have different spatial dimensions (i.e. number of rows).

Being able to jointly assess the common variability, over the same period, of two data-sets with different spatial extent is a particular benefit of the SVD method. In the present case, the (climate) variability of a given MORA station is no longer solely compared to its corresponding grid-cell in the GRIDCLIM data-set (e.g. as was done in the gap-filling procedure); the SVD enables to assess how the 338 MORA time-series for observed temperature linearly relate to the $69\,842$ individual time-series from GRIDCLIM grid-cells covering the Swedish territory. Since regional-scale climate variability is (presumably) well captured in the GRIDCLIM re-analysis product, the SVD method thus enables to isolate corresponding trends in the MORA data-set – regardless of how well the one-to-one correspondence between the observed and simulated signal matches for any single location.

$$\text{Covariance matrix: } \mathbf{R} = \mathbf{X}_c^{MORA} \cdot \left(\mathbf{X}_c^{\text{GridClim}}\right)^T$$

$$\text{Singular Value Decomposition: } \mathbf{R} = \mathbf{SVD}^{MORA} \cdot \Lambda^{SVD} \cdot \left(\mathbf{SVD}^{\text{GridClim}}\right)^T$$

Time Expansion Coefficients:

$$\begin{cases} \mathbf{A}^{MORA} = & \mathbf{X}_c^{MORA} \cdot \mathbf{SVD}^{MORA} \\ \mathbf{A}^{\text{GridClim}} = & \mathbf{X}_c^{\text{GridClim}} \cdot \mathbf{SVD}^{\text{GridClim}} \end{cases} \tag{A1}$$

Analogously to Eq. (4), the portion of the explained (common) variance can be retrieved from eigenvalues in diagonal matrix Λ^{SVD}. Similar to Eq. (5), the original $\mathbf{X}_c^{MORA}$ and $\mathbf{X}_c^{\text{GridClim}}$ matrices can be identically reconstructed from the results of the SVD:

$$\begin{cases} \mathbf{X}_c^{MORA} = \mathbf{A}^{MORA} \cdot \left(\mathbf{SVD}^{MORA}\right)^T \\ \mathbf{X}_c^{\text{GridClim}} = \mathbf{A}^{\text{GridClim}} \cdot \left(\mathbf{SVD}^{\text{GridClim}}\right)^T \end{cases} \tag{A2}$$

The first three leading modes of the SVD for the MORA calibration subset are presented in Fig. (A2) for temperature: the top row shows the spatial SVD patterns $\left(\mathbf{svd}^{MORA}\right)^{[1:3]}$, the bottom row the corresponding time expansion coefficients

https://doi.org/10.5194/egusphere-2024-582
Preprint. Discussion started: 28 March 2024

$\left(\overrightarrow{\mathbf{a}^{MORA}}\right)^{[1:3]}$. The portion of the variance explained by each mode is obtained from eigenvalues in matrix $\mathbf{\Lambda}^{\mathbf{SVD}}$, and displayed in the legend.

Analogously to Eq. (6), the original data-set with missing values $\mathbf{X}_c^{\star}$ can estimated using SVD patterns as $\widehat{\mathbf{X}_c}^{SVD}$.

$$\widehat{\mathbf{X}_c}^{SVD} = \widehat{\mathbf{A}}^{MORA} \cdot \left(\mathbf{SVD}^{MORA}\right)^{-1} \tag{A3}$$

The climate indicator ***CI*** can thus be computed according to Eq. (7). Since $\mathbf{eof}^{MORA}$ and $\mathbf{svd}^{MORA}$ are virtually identical, the reconstructed EOF-rec and SVD-rec cannot be distinguished from each other. Hence SVD-rec is not shown on Fig. (6) and related figures.

https://doi.org/10.5194/egusphere-2024-582
Preprint. Discussion started: 28 March 2024

A3 Leading SVD patterns for SVD-rec

Figure A2. Leading three SVD patterns for MORA (Calibration network) over Sweden ($\mathbf{svd}^{MORA}(\#1-3)$), with their associated time expansion vectors ($\mathbf{a}_{SVD}^{MORA}(\#1-3)$) for temperature.

https://doi.org/10.5194/egusphere-2024-582
Preprint. Discussion started: 28 March 2024

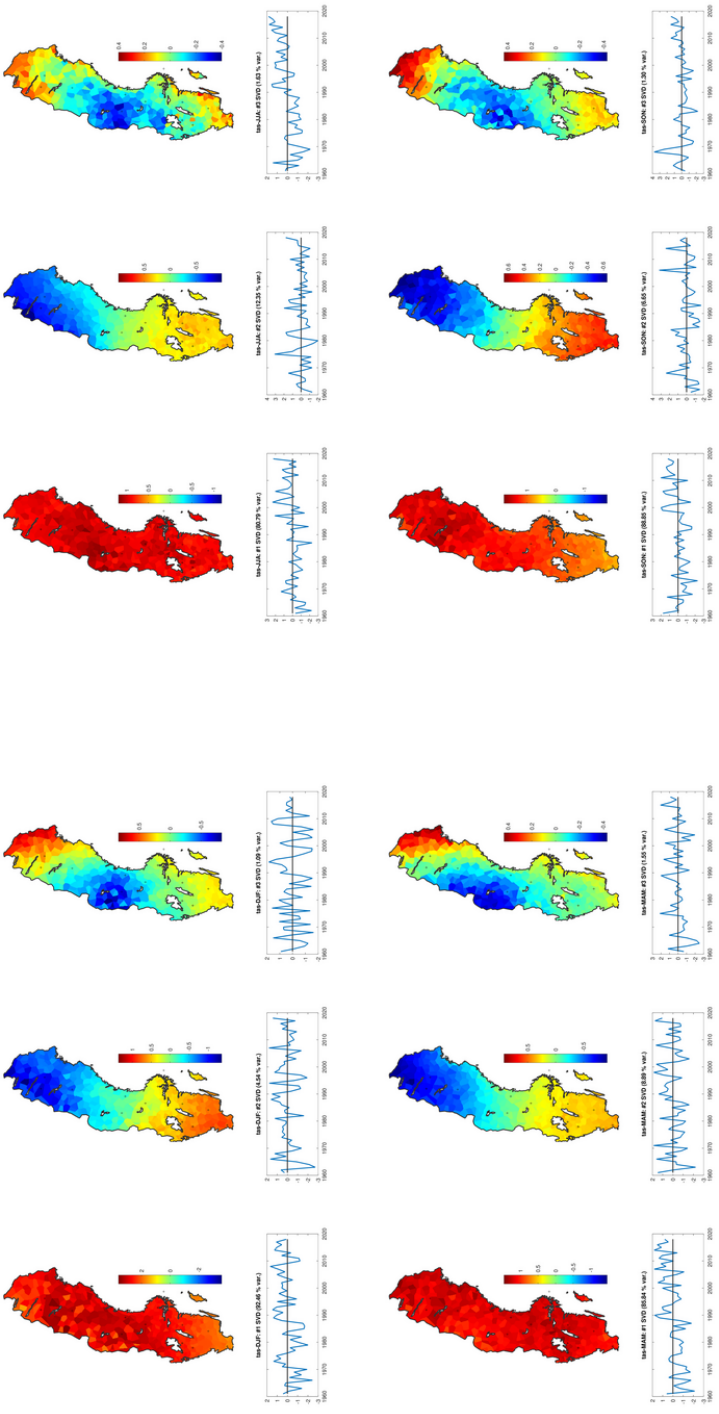

Figure A3. Analogue to Figure Fig. (A2) , for seasonal SVD patterns of *temperature* instead annual. Seasons are defined as DJF (winter, upper left), JJA (summer, upper right), MAM (spring, lower left) and SON (autumn, lower right).

https://doi.org/10.5194/egusphere-2024-582
Preprint. Discussion started: 28 March 2024

Acknowledgements. The manuscript has greatly benefited from comments and suggestions by Peter Berg, Magnus Joelsson, Erik Engström, Semjon Schimanke and colleagues. Furthermore, no evaluation of historic measurements over Sweden would be possible without the dedication of Lennart Wern, Sverker Hellström and colleagues to digitise, quality-check and organise observations in the MORA database.

https://doi.org/10.5194/egusphere-2024-582
Preprint. Discussion started: 28 March 2024

References

Alexandersson, H.: A homogeneity test applied to precipitation data, Journal of Climatology, 6, 661–675, https://doi.org/10.1002/joc.3370060607, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/joc.3370060607, 1986.

Allan, R. and Ansell, T.: A New Globally Complete Monthly Historical Gridded Mean Sea Level Pressure Dataset (HadSLP2): 1850–2004, Journal of Climate, 19, 5816–5842, https://doi.org/10.1175/JCLI3937.1, publisher: American Meteorological Society Section: Journal of Climate, 2006.

Alvarez-Castro, M. C., Faranda, D., and Yiou, P.: Atmospheric Dynamics Leading to West European Summer Hot Temperatures Since 1851, Complexity, 2018, e2494 509, https://doi.org/10.1155/2018/2494509, publisher: Hindawi, 2018.

Andersson, S., Bärring, L., Landelius, T., Samuelsson, P., and Schimanke, S.: SMHI Gridded Climatology, Tech. Rep. 118, SMHI, Norrköping, Sweden, https://orcid.org/0000-0003-4387-6232, 2021.

Ansell, T. J., Jones, P. D., Allan, R. J., Lister, D., Parker, D. E., Brunet, M., Moberg, A., Jacobeit, J., Brohan, P., Rayner, N. A., Aguilar, E., Alexandersson, H., Barriendos, M., Brandsma, T., Cox, N. J., Della-Marta, P. M., Drebs, A., Founda, D., Gerstengarbe, F., Hickey, K., Jónsson, T., Luterbacher, J., Nordli, O., Oesterle, H., Petrakis, M., Philipp, A., Rodwell, M. J., Saladie, O., Sigro, J., Slonosky, V., Srnec, L., Swail, V., García-Suárez, A. M., Tuomenvirta, H., Wang, X., Wanner, H., Werner, P., Wheeler, D., and Xoplaki, E.: Daily Mean Sea Level Pressure Reconstructions for the European–North Atlantic Region for the Period 1850–2003, Journal of Climate, 19, 2717–2742, https://doi.org/10.1175/JCLI3775.1, publisher: American Meteorological Society Section: Journal of Climate, 2006.

Balting, D. F., Ionita, M., Wegmann, M., Helle, G., Schleser, G. H., Rimbu, N., Freund, M. B., Heinrich, I., Caldarescu, D., and Lohmann, G.: Large-scale climate signals of a European oxygen isotope network from tree rings, Climate of the Past, 17, 1005–1023, https://doi.org/10.5194/cp-17-1005-2021, publisher: Copernicus GmbH, 2021.

Begert, M., Schlegel, T., and Kirchhofer, W.: Homogeneous temperature and precipitation series of Switzerland from 1864 to 2000, International Journal of Climatology, 25, 65–80, https://doi.org/10.1002/joc.1118, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/joc.1118, 2005.

Bergström, H. and Moberg, A.: Daily Air Temperature and Pressure Series for Uppsala (1722–1998), Climatic Change, 53, 213–252, https://doi.org/10.1023/A:1014983229213, 2002.

Björnsson, H. and Venegas, S. A.: A manual for EOF and SVD analyses of climatic data, CCGCR Report 97, Centre for Climate and Global Change Research – McGill University, http://www.geog.mcgill.ca/gec3/wp-content/uploads/2009/03/Report-no.-1997-1.pdf, 1997.

Båserud, L., Lussana, C., Nipen, T. N., Seierstad, I. A., Oram, L., and Aspelien, T.: TITAN automatic spatial quality control of meteorological in-situ observations, in: Advances in Science and Research, vol. 17, pp. 153–163, Copernicus GmbH, https://doi.org/10.5194/asr-17-153-2020, iSSN: 1992-0628, 2020.

Comas-Bru, L. and Hernández, A.: Reconciling North Atlantic climate modes: revised monthly indices for the East Atlantic and the Scandinavian patterns beyond the 20th century, Earth System Science Data, 10, 2329–2344, https://doi.org/10.5194/essd-10-2329-2018, publisher: Copernicus GmbH, 2018.

Cropper, T., Hanna, E., Valente, M. A., and Jónsson, T.: A daily Azores–Iceland North Atlantic Oscillation index back to 1850, Geoscience Data Journal, 2, 12–24, https://doi.org/10.1002/gdj3.23, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/gdj3.23, 2015.

Cropper, T. E., Hanna, E., Valente, M. A., and Jónsson, T.: A Daily Azores-Iceland North Atlantic Oscillation Index back to 1850, https://doi.org/10.5281/zenodo.9979, 2014.

https://doi.org/10.5194/egusphere-2024-582
Preprint. Discussion started: 28 March 2024

Engström, E.: SMHI Klimatindikator för nederbörd över Sverige, https://www.smhi.se/klimat/klimatet-da-och-nu/klimatindikatorer/klimatindikator-nederbord-1.2887, 2022.

Engström, E.: SMHI Klimatindikator för temperatur över Sverige, https://www.smhi.se/klimat/klimatet-da-och-nu/klimatindikatorer/klimatindikator-temperatur-1.2430, 2023.

Fleig, A. K., Tallaksen, L. M., James, P., Hisdal, H., and Stahl, K.: Attribution of European precipitation and temperature trends to changes in synoptic circulation, Hydrology and Earth System Sciences, 19, 3093–3107, https://doi.org/10.5194/hess-19-3093-2015, publisher: Copernicus GmbH, 2015.

Gallego, D., Garcia-Herrera, R., Ribera, P., and Jones, P. D.: Seasonal mean pressure reconstruction for the North Atlantic (1750–1850) based on early marine data, Climate of the Past, 1, 19–33, https://doi.org/10.5194/cp-1-19-2005, publisher: Copernicus GmbH, 2005.

Gouirand, I., Moberg, A., and Zorita, E.: Climate variability in Scandinavia for the past millennium simulated by an atmosphere-ocean general circulation model, Tellus A: Dynamic Meteorology and Oceanography, 59, 30–49, https://doi.org/10.1111/j.1600-0870.2006.00207.x, publisher: Taylor & Francis _eprint: https://doi.org/10.1111/j.1600-0870.2006.00207.x, 2007.

Guijarro, J. A.: The climatol R package, https://climatol.eu/, 2024.

Guijarro, J. A., López, J. A., Aguilar, E., Domonkos, P., Venema, V. K. C., Sigró, J., and Brunet, M.: Homogenization of monthly series of temperature and precipitation: Benchmarking results of the MULTITEST project, International Journal of Climatology, 43, 3994–4012, https://doi.org/10.1002/joc.8069, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/joc.8069, 2023.

Gulev, S., Thorne, P., Ahn, J., Dentener, F., Domingues, C., Gerland, S., Gong, D., Kaufman, D., Nnamchi, H., Quaas, J., Rivera, J., Sathyendranath, S., Smith, S., Trewin, B., von Schuckmann, K., and Vose, R.: Changing State of the Climate System, in: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, edited by Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J., Maycock, T., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B., pp. 287–422, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, https://doi.org/10.1017/9781009157896.004, type: Book Section, 2021.

Hanna, E., Cropper, T. E., Hall, R. J., Cornes, R. C., and Barriendos, M.: Extended North Atlantic Oscillation and Greenland Blocking Indices 1800–2020 from New Meteorological Reanalysis, Atmosphere, 13, 436, https://doi.org/10.3390/atmos13030436, number: 3 Publisher: Multidisciplinary Digital Publishing Institute, 2022.

Holmgren, E. and Kjellström, E.: Exploring extreme event attribution by using long-running meteorological observations, EGUsphere, pp. 1–28, https://doi.org/10.5194/egusphere-2023-2879, publisher: Copernicus GmbH, 2023.

Hänsel, S.: Changes in the Characteristics of Dry and Wet Periods in Europe (1851–2015), Atmosphere, 11, 1080, https://doi.org/10.3390/atmos11101080, number: 10 Publisher: Multidisciplinary Digital Publishing Institute, 2020.

Jacobeit, J., Jones, P., Davies, T., and Beck, C.: Circulation Changes in Europe since the 1780s, in: History and Climate: Memories of the Future?, edited by Jones, P. D., Ogilvie, A. E. J., Davies, T. D., and Briffa, K. R., pp. 79–99, Springer US, Boston, MA, ISBN 978-1-4757-3365-5, https://doi.org/10.1007/978-1-4757-3365-5_5, 2001a.

Jacobeit, J., Jönsson, P., Bärring, L., Beck, C., and Ekström, M.: Zonal Indices for Europe 1780–1995 and Running Correlations with Temperature, Climatic Change, 48, 219–241, https://doi.org/10.1023/A:1005619023045, 2001b.

Jacobeit, J., Wanner, H., Luterbacher, J., Beck, C., Philipp, A., and Sturm, K.: Atmospheric circulation variability in the North-Atlantic-European area since the mid-seventeenth century, Climate Dynamics, 20, 341–352, https://doi.org/10.1007/s00382-002-0278-0, 2003.

https://doi.org/10.5194/egusphere-2024-582
Preprint. Discussion started: 28 March 2024

Joelsson, L. M. T., Sturm, C., Södling, J., Engström, E., and Kjellström, E.: Automation and evaluation of the interactive homogenization tool HOMER, International Journal of Climatology, 42, 2861–2880, https://doi.org/10.1002/joc.7394, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/joc.7394, 2022.

Joelsson, L. M. T., Engström, E., and Kjellström, E.: Homogenization of Swedish mean monthly temperature series 1860–2021, International Journal of Climatology, 43, 1079–1093, https://doi.org/10.1002/joc.7881, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/joc.7881, 2023.

Jones, P. D. and Mann, M. E.: Climate over past millennia, Reviews of Geophysics, 42, https://doi.org/10.1029/2003RG000143, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2003RG000143, 2004.

Jones, P. D., Osborn, T. J., and Briffa, K. R.: The Evolution of Climate Over the Last Millennium, Science, 292, 662–667, https://doi.org/10.1126/science.1059126, publisher: American Association for the Advancement of Science, 2001.

Kadow, C., Hall, D. M., and Ulbrich, U.: Artificial intelligence reconstructs missing climate information, Nature Geoscience, 13, 408–413, https://doi.org/10.1038/s41561-020-0582-5, number: 6 Publisher: Nature Publishing Group, 2020.

Kaspar, F., Mächel, H., Jacob, D., and Kottmeier, C.: Beobachtung von Klima und Klimawandel in Mitteleuropa und Deutschland, in: Klimawandel in Deutschland: Entwicklung, Folgen, Risiken und Perspektiven, edited by Brasseur, G. P., Jacob, D., and Schuck-Zöller, S., pp. 17–26, Springer, Berlin, Heidelberg, ISBN 978-3-662-50397-3, https://doi.org/10.1007/978-3-662-50397-3_3, 2017.

Kjellström, E., Hansen, F., and Belušić, D.: Contributions from Changing Large-Scale Atmospheric Conditions to Changes in Scandinavian Temperature and Precipitation Between Two Climate Normals, Tellus A: Dynamic Meteorology and Oceanography, 74, 204–221, https://doi.org/10.16993/tellusa.49, number: 1 Publisher: Stockholm University Press, 2022.

Kyselý, J.: Implications of enhanced persistence of atmospheric circulation for the occurrence and severity of temperature extremes, International Journal of Climatology, 27, 689–695, https://doi.org/10.1002/joc.1478, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/joc.1478, 2007.

Kyselý, J.: Influence of the persistence of circulation patterns on warm and cold temperature anomalies in Europe: Analysis over the 20th century, Global and Planetary Change, 62, 147–163, https://doi.org/10.1016/j.gloplacha.2008.01.003, 2008.

Lussana, C.: metno/gridpp, https://github.com/metno/gridpp, original-date: 2015-01-28T12:58:02Z, 2007.

Massei, N., Durand, A., Deloffre, J., Dupont, J. P., Valdes, D., and Laignel, B.: Investigating possible links between the North Atlantic Oscillation and rainfall variability in northwestern France over the past 35 years, Journal of Geophysical Research: Atmospheres, 112, https://doi.org/10.1029/2005JD007000, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2005JD007000, 2007.

Masson-Delmotte, V., Zhai, P., Pirani, S., Connors, C., Péan, S., Berger, N., Caud, Y., Chen, L., Goldfarb, M., and Scheel Monteiro, P. M.: IPCC 2021: Summary for policymakers., Tech. rep., IPCC, Geneva, https://www.ipcc.ch/report/ar6/wg1/chapter/summary-for-policymakers/, publisher: Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2021.

Meteo-France: Données climatologiques de base - mensuelles, https://meteo.data.gouv.fr/datasets/6569b3d7d193b4daf2b43edc, 2024.

Moberg, A. and Bergström, H.: Homogenization of Swedish temperature data. Part III: the long temperature records from Uppsala and Stockholm, International Journal of Climatology, 17, 667–699, https://doi.org/10.1002/(SICI)1097-0088(19970615)17:7<667::AID-JOC115>3.0.CO;2-J, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/%28SICI%291097-0088%2819970615%2917%3A7%3C667%3A%3AAID-JOC115%3E3.0.CO%3B2-J, 1997.

Moberg, A., Alexandersson, H., Bergström, H., and Jones, P. D.: Were southern Swedish summer temperatures before 1860 as warm as measured?, International Journal of Climatology, 23, 1495–1521, https://doi.org/10.1002/joc.945, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/joc.945, 2003.

https://doi.org/10.5194/egusphere-2024-582
Preprint. Discussion started: 28 March 2024

Office, M.: UK temperature, rainfall and sunshine time series, https://www.metoffice.gov.uk/research/climate/maps-and-data/uk-temperature-rainfall-and-sunshine-time-series, 2024.

Ono, K., Kunii, M., and Honda, Y.: The regional model-based Mesoscale Ensemble Prediction System, MEPS, at the Japan Meteorological Agency, Quarterly Journal of the Royal Meteorological Society, 147, 465–484, https://doi.org/10.1002/qj.3928, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/qj.3928, 2021.

Philipp, A., Della-Marta, P. M., Jacobeit, J., Fereday, D. R., Jones, P. D., Moberg, A., and Wanner, H.: Long-Term Variability of Daily North Atlantic–European Pressure Patterns since 1850 Classified by Simulated Annealing Clustering, Journal of Climate, 20, 4065–4095, https://doi.org/10.1175/JCLI4175.1, publisher: American Meteorological Society Section: Journal of Climate, 2007.

Schimanke, S. and Service, C. C. C.: UERRA regional reanalysis for Europe on pressure levels from 1961 to 2019, https://doi.org/10.24381/CDS.92221551, 2019.

Schimanke, S., Undén, P., Isaksson, L., Edvinsson, L., Ridal, M., Olsson, E., Hopsch, S., and Andersson, S.: Copernicus regional reanalysis for Europe, in: EMS Annual Meeting Abstracts; EMS: Berlin, Germany, https://doi.org/10.24381/cds.dd7c6d66, 2019.

Slonosky, V., Jones, P., and Davies, T.: Variability of the surface atmospheric circulation over Europe, 1774–1995, International Journal of Climatology, 20, 1875–1897, https://doi.org/10.1002/1097-0088(200012)20:15<1875::AID-JOC593>3.0.CO;2-D, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/1097-0088%28200012%2920%3A15%3C1875%3A%3AAID-JOC593%3E3.0.CO%3B2-D, 2000.

Storch, H. v., Bürger, G., Schnur, R., and Storch, J.-S. v.: Principal Oscillation Patterns: A Review, Journal of Climate, 8, 377–400, https://doi.org/10.1175/1520-0442(1995)008<0377:POPAR>2.0.CO;2, publisher: American Meteorological Society Section: Journal of Climate, 1995.

Sturm, C.: Spatially aggregated climate indicators over Sweden (1860–2020), part 2: Precipitation, Climate of the Past Discussions, 2024a.

Sturm, C.: Spatially aggregated climate indicators over Sweden (1860–2020): scripts and data, https://doi.org/10.5281/zenodo.10888129, 10.5281/zenodo.10888129, 2024b.

Vose, R. S., Applequist, S., Squires, M., Durre, I., Menne, M. J., Williams, C. N., Fenimore, C., Gleason, K., and Arndt, D.: Improved Historical Temperature and Precipitation Time Series for U.S. Climate Divisions, Journal of Applied Meteorology and Climatology, 53, 1232–1251, https://doi.org/10.1175/JAMC-D-13-0248.1, publisher: American Meteorological Society Section: Journal of Applied Meteorology and Climatology, 2014.

Wilcke, R. A. I., Kjellström, E., Lin, C., Matei, D., Moberg, A., and Tyrlis, E.: The extremely warm summer of 2018 in Sweden – set in a historical context, Earth System Dynamics, 11, 1107–1121, https://doi.org/10.5194/esd-11-1107-2020, publisher: Copernicus GmbH, 2020.

Zveryaev, I. I.: Seasonally varying modes in long-term variability of European precipitation during the 20th century, Journal of Geophysical Research: Atmospheres, 111, https://doi.org/10.1029/2005JD006821, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2005JD006821, 2006.