~~The contribution~~ Considerable yet contrasting regional imprint of circulation ~~changes to~~ change on summer temperature trends ~~in~~ across the ~~northern~~ Northern hemisphere mid-latitudes~~: A multi-method quantification~~

Peter Pfleiderer[1], Anna Merrifield[2], István Dunkl[1], Durand Homer[3], Enora Cariou[4], Julien Cattiaux[4], Camps-Valls Gustau[3], and Sebastian Sippel[1]

[1]Institute for Meteorology, Leipzig University, Leipzig, Germany
[2]Institute for Atmospheric and Climate Sciences, ETH Zürich, Switzerland
[3]Image Processing Laboratory, Universitat de València, Spain
[4]Centre National de Recherches Météorologiques, Université de Toulouse, CNRS, Météo-France, Toulouse, France

Correspondence: Peter Pfleiderer (peter.pfleiderer@uni-leipzig.de)

Abstract. ~~The intensification of summer temperature and heat extremes is well documented and attributed to~~ Rising summer temperatures and more frequent heat extremes are well-documented outcomes of anthropogenic climate change. ~~There is, however, still a vivid debate about the influence of~~ However, the extent to which atmospheric circulation changes ~~on heat extremes. Over the northern hemispheric~~ contribute to these trends remains contested. Regional differences across the northern mid-latitudes ~~, considerable regional differences in summer temperatures have been observed. These differences have been linked to atmospheric circulation changes using statistical methods, but it remains challengingto evaluate such methods on multi-decadal time scales. Here, we evaluate different decomposition methods and systematically investigate circulation-induced summer temperature trends. For the evaluation of~~ suggest that circulation plays a role, yet robustly quantifying its contribution over multiple decades is very challenging. We address this by systematically testing statistical and machine learning decomposition methods ~~we use~~ with climate model simulations~~without external forcing but with atmospheric circulation nudged towards the circulation of a freely running forced simulation . We train the decomposition methods on the free-running forced simulation and compare its~~ . Specifically, we use unforced simulations with circulation nudged to match a forced simulation that includes anthropogenic emissions and land-use change. We apply decomposition methods to the forced simulations and compare their estimates of circulation-induced trends ~~to the trends simulated in the nudged circulation simulation. Most decomposition methods show skill in estimating the sign~~ with those found in nudged circulation simulations. Our analysis reveals that most methods accurately identify the direction of circulation-induced ~~trends but all methods underestimate the magnitudeof these trends. The application of tested decomposition methods confirms~~ changes, although they consistently underestimate their magnitude. Despite this limitation, the results demonstrate that circulation changes have ~~contributed substantially to an increase in summer heat over several~~ made a substantial contribution to summer temperature trends across the northern mid-latitudes. In Europe, a hotspot region, we estimate that up to half of the observed summer warming between 1979 and 2023 can be attributed to circulation trends. Furthermore, circulation trends have contributed to warmer summer temperatures over Western

North America, Central Siberia, Mongolia, Central China, and northeastern Canada. Yet, circulation changes have cooled summer temperatures over Eastern and Central North America, Eastern China, and Central Asia. Overall, our results, based on multiple methods, confirm a circumglobal mid-latitude ~~regions, including Europe~~pattern of considerable, yet contrasting, contributions of circulation changes to summer temperature trends.

1 Introduction

The intensity of ~~heat waves~~ heatwaves increases globally, causing ~~considerable~~ significant ecological and societal impacts. It is well understood that in a warming climate~~heat waves over~~ , heatwaves lasting a few days or warm summer seasons intensify, and this effect can be robustly attributed to anthropogenic activities (Seneviratne et al., 2021). The main reason for this intensification is of a thermodynamic nature: heat extremes occurring in a warmer atmosphere are also warmer. However, heat waves are not only determined by direct thermodynamic changes in a warming climate: Various effects can lead to regional trends in summer heat, including changes in atmospheric circulation (Teng et al., 2022; Rousi et al., 2022; Vautard et al., 2023; Singh et al., 2023), or ~~changes in soil moisture~~ feedbacks due to land-atmosphere interactions (Seneviratne et al., 2006). As a result, regional trends in summer average temperatures or summer heat extremes differ strongly across the world and even across the northern hemispheric mid-latitude land area. ~~An improved understanding in particular of effects due to atmospheric circulation changes and quantifying to which extent these changes are due to forced or internal variability is crucial for our understanding of past summer temperature trends (Merrifield et al., 2017), extreme events (Terray, 2021) and constraints on projected future temperatures.~~

In the mid-latitudes, large-scale circulation is a crucial driver for heat extremes (Rousi et al., 2022; Röthlisberger and Papritz, 2023) and there is great interest in understanding to what extent atmospheric circulation contributed to individual events (Cattiaux et al., 2010; Sippel et al., 2024; Zeder and Fischer, 2023), trends in heat extremes (Rousi et al., 2022; Singh et al., 2023) or seasonal temperatures (Teng et al., 2022). Forced changes in ~~the summer jet strength and structure alongside with~~ jet stream position and strength (Dong et al., 2022; Rousi et al., 2022; Woollings et al., 2023; Shaw and Miyawaki, 2024) , changes in storm track intensity ~~are one reason for forced circulation driven changes, yet model simulated trends are very diverse, not consistent across models and thus uncertain (Shepherd, 2014), nat. var however stays main factor here. Over mid-latitude Europe and parts of Eurasia, observed summer circulation has changed towards more anticyclonic conditions (Horton et al., 2015) with the increased occurrence of double jet stream states (Rousi et al., 2022)~~(Coumou et al., 2015; Chemke and Coum and the resulting changes in weather pattern frequencies (Horton et al., 2015; Hanna et al., 2018; Fabiano et al., 2021) are likely to affect local climate conditions (Pfleiderer et al., 2019). Over the observational record, these forced changes, however, are small compared to internal climate variability (Eyring et al., 2021). Estimating the contribution of atmospheric circulation changes to local temperature trends and quantifying the extent to which these changes are due to forced or internal variability is crucial for our understanding of past summer temperature trends (Merrifield et al., 2017; Teng et al., 2022; Vautard et al., 2023) and extreme events (Terray, 2021).

In some regions, the observed trends are falling ~~out of~~ outside the range of model-simulated expected trends (e.g., Western Europe, (Teng et al., 2022; Rousi et al., 2022; Vautard et al., 2023; Kornhuber et al., 2024)~~, but it is unclear if this only affects a few regions or whether this is a global phenomenon~~. Potentially, this ~~may indicate~~ indicates that model-simulated low-frequency variability in large-scale atmospheric circulation is too weak, or that a forced ~~circulation~~ change in circulation is missing in the models, notwithstanding the broad uncertainty across models (Shepherd, 2014). The missing low-frequency hypothesis is also ~~difficult~~ challenging to assess because the available observations are ~~rather~~ relatively short. Understanding past circulation changes and their contribution to temperature trends may provide an opportunity to constrain future changes in summer temperature trends, ~~as circulation presents a very large~~ acknowledging that future circulation changes remain a huge source of uncertainty (Topál and Ding, 2023; Fereday et al., 2018).

~~However, it is not straightforward to identify circulation-induced contributions to temperature trends (Deser et al., 2016). To empirically study circulation contributions to temperature trends one can either use i)~~ Identifying the contribution of atmospheric circulation to temperature trends is not straightforward (Deser et al., 2016). Two main approaches are commonly used. The first applies statistical or machine learning methods ~~that decompose the~~ to decompose observed or simulated temperature ~~trend in its thermodynamically induced and its circulation induced~~ trends into thermodynamic and circulation-driven components, often ~~framed as "dynamical adjustment " (Deser et al., 2016; Smoliak et al., 2015; Sippel et al., 2019)or ii) create~~ referred to as dynamical adjustment (Deser et al., 2016; Smoliak et al., 2015; Sippel et al., 2019). The second uses nudged circulation simulations~~where the circulation component is fixed~~, in which the circulation is prescribed and the thermodynamic component is removed ~~(see e.g., (Wehrli et al., 2018)). Although nudged~~ (Wehrli et al., 2018). Both options have their limitations: Nudged circulation experiments are limited by the representation of physical mechanisms in the climate model used~~, assessing how well~~. On the other hand, most statistical decomposition methods are ~~suited to estimate circulation-induced trends is challenging. Most of these methods are~~ designed to capture the relationship between daily circulation patterns and daily temperatures~~and they~~. They do indeed capture day-to-day variability very well~~, and good~~. Good skill is obtained on monthly or inter-annual time scales as well ~~(Smoliak et al., 2015; Sippel et al., 2019)(Cariou et al. , 2025).~~ (Smoliak et al., 2015; Sippel et al., 2019; Ca . Whether they can adequately capture a long-term trend is~~however more difficult~~, however, more challenging to test, because processes determining long-term trends may be distinctly different from those that determine short-term circulation variability~~;~~, and much fewer verification samples are available. Moreover, benchmarks for circulation-induced long-term trends have not been available ~~so far~~to date, and to our knowledge, no systematic comparison of dynamical adjustment methods has been ~~performed~~conducted.

In this study, we ~~provide a robust overview of circulation induced JJA temperature trends over the northern hemispheric~~ present a comprehensive assessment of circulation-induced summer temperature trends across the northern mid-latitudes ~~by making use of~~ using both statistical decomposition methods and nudged circulation experiments. We ~~address two specific research questions: First, we evaluate whether~~ focus on two key questions. First, can statistical-empirical methods ~~are able to correctly estimate circulation-induced~~ reliably estimate circulation-driven long-term trends when tested against a climate model benchmark~~. The climate model benchmark experiment is specificially designed as~~ ? To address this, we use a set of CESM2 nudged circulation simulations ~~, which allows comparison of~~ specifically designed to provide a reference for comparing

circulation-induced ~~trend components against statistical decompoistion~~ trends with those derived from statistical methods. Second, we identify circulation-induced ~~JJA~~ summer temperature trends across the northern hemispheric mid-latitudes in observations using four different statistical methods~~and~~ , as well as in CESM2 simulations that are nudged to the ERA5 circulation but driven without anthropogenic forcing.

2 Data & Methods

This study is mainly based on simulations of the fully coupled Community Earth System Climate Model, Version 2 (CESM2) (Danabasoglu et al., 2020), including simulations from the CESM2 large ensemble (Rodgers et al., 2021). In section 3.2 we apply decomposition methods to the European Centre for Medium-Range Weather Forecasts (ECMWF) Reanalysis v5 (ERA5) (Hersbach et al., 2020).

2.1 CESM2 nudged circulation simulations driven with CESM2 horizontal winds

To derive a benchmark for ~~the evaluation of the~~ evaluating the decomposition methods, we use nudged circulation experiments ~~that have been~~ conducted with the fully coupled ~~Community Earth System Climate Model, Version 2~~ CESM2 (Danabasoglu et al., 2020). First, three standard historical and future anthropogenic forcing experiments are simulated, following historical greenhouse gas emissions, aerosol emissions, and land use changes ~~in the period 1850-2014~~from 1850 to 2014, and anthropogenic forcings ~~following the SSP3-70 scenario~~ from then onwards ('hist+SSP370'). These simulations follow the protocol of the CESM2 large ensemble (Rodgers et al., 2021).

In a second step, for each of these simulations, a nudged circulation simulation is created for piControl forcing (e.g., no anthropogenic greenhouse gas and aerosol emissions, as well as no land use change). Each of these simulations starts with the same initial conditions as ~~their~~ its corresponding hist+ssp370 simulation~~;~~ . 6-hourly global meridional and horizontal winds throughout the atmosphere (at all vertical levels) ~~and globally are nudged~~ are nudged globally to their corresponding hist+ssp370 simulation. The nudging is achieved via a regular relaxation procedure described in the CAM6 handbook (camdoc) , and is similar to Topál and Ding 2023. These simulations will ~~be henceforth referred to~~ henceforth be referred to as 'piControl-nudged', as they lack ~~the~~ direct anthropogenic forcing~~but~~ . Still, through the nudging of atmospheric circulation, any potential anthropogenic forcing on ~~atmospheric circulation are present alongside with the identical~~ horizontal wind fields is present alongside the internal circulation variability of the atmosphere ~~.~~ from the hist+ssp370 simulation. Because land–atmosphere coupling influences near-surface circulation differently under varying climate forcings, it is recommended to apply nudging only at higher altitudes, allowing surface winds to evolve freely (Wehrli et al., 2018; Merrifield et al., 2019). Although these effects may influence individual events, the spatial patterns of long-term circulation trends remain robust regardless of whether near-surface winds are nudged (Singh et al., 2025).

~~Conceptual illustration of causal relationships influencing local temperatures in a freely running climate simulation (left) and in a piControl-nudged simulation (right). The arrow width indicates assumed importance of links. The blue rectangle highlights the processes studied here.~~

2.1.1 Conceptual Interpretation of CESM2 nudged circulation simulations as a benchmark for dynamical adjustment

In a free-running, fully coupled climate simulation, local ~~or~~ and regional temperatures are ~~affected both by thermodynamic forcing (sketched here as GMSTinfluence) as well as~~ shaped by both thermodynamic forcing, here represented by global mean surface temperature (GMST) as a proxy for the large-scale thermodynamic background conditions, and atmospheric circulation variability~~(e.g.; ~~, as well as their interactions (e.g., Deser et al. 2016, Fig. 1)~~, and interactions between both~~. Statistical dynamical adjustment methods ~~aim to separate those~~ seek to separate these influences, but ~~the skill of those methods is inherently difficult to evaluate in~~ their skill is challenging to evaluate within a coupled system. ~~Hence, the goal of our~~ To address this, we use 'piControl-nudged' simulations ~~is to derive benchmark simulations, in which only circulation is variable,~~ as benchmarks in which circulation varies while thermodynamic forcing is fixed ~~to~~ at pre-industrial ~~values. This will allow to~~ levels. This design allows us to isolate and evaluate the circulation contribution independently ~~from the trends in thermodynamic forcing. The conceptual interpretation of the 'piControl-nudged' simulations is sketched in~~ of thermodynamic trends. Conceptually (Fig. 1~~: Circulation variability is 'inherited' from its parent simulation , which we expect to determine~~), circulation variability is "inherited" from the parent simulation and is expected to dominate local and regional ~~temperatures to the largest extent. The inherited circulation variability may contain both internal variability from the parent simulation, but also potential~~ temperature responses. This inherited variability includes both internal fluctuations and possible forced circulation changes. ~~Moreover, we expect that some internal variability may remain in the 'piControl-nudged' simulations, for example due to ocean variability that may affect GMST~~ Some internal variability unrelated to atmospheric circulation may also remain, such as ocean-driven GMST fluctuations (arrow b in Fig. 1). Note that the decomposition into "thermodynamic" and "circulation-induced" changes is a simplification that overlooks important mechanisms for local temperatures. In Section 3.3, we discuss the implications of this simplification in more detail.

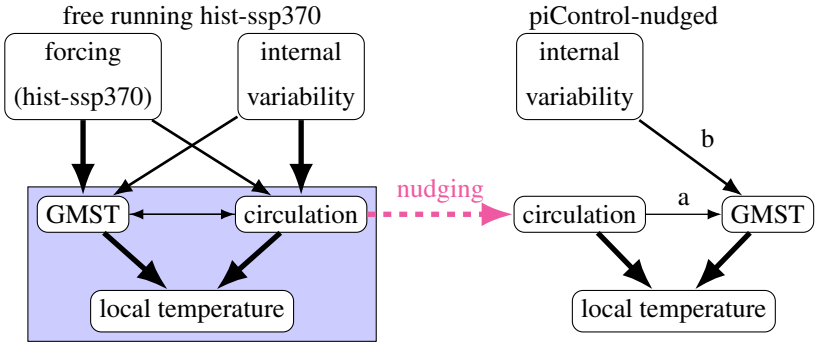

Figure 1. Conceptual illustration of causal relationships influencing local temperatures in a freely running climate simulation (left) and in a piControl-nudged simulation (right). The arrow width indicates the assumed importance of links. The blue rectangle highlights the processes studied here.

2.1.2 Illustration of CESM2 nudged circulation simulations

As expected, day-to-day variability in the nudged circulation run is closely related to ~~it's~~ its freely running counterpart from which the wind fields originated~~from~~. As shown in ~~figure ??~~Figure **??**, in the early period (left column of fig. **??**), geopotential height at 500 hPa and surface air temperature are nearly identical in the hist+ssp370 and the piControl-nudged run. In a warmer climate, ~~the~~ day-to-day variability ~~is still~~ remains highly correlated, but geopotential height and surface air temperatures are ~~relatively~~ uniformly shifted to higher values. ~~However, interpreting~~

Interpreting the GMST signal in the ~~'~~piControl-nudged ~~' runs is somewhat challenging. Despite the absence of~~ runs is not straightforward. Even without external forcing, small ~~trends in GMST can be found~~ GMST trends emerge over 40-year periods ~~in piControl simulations (compare figure~~ (Fig. 2b,c). These trends ~~may be partly forced by atmospheric circulation (all three 'piControl-nudged' simulations show weaker trends relative to the CESM2 large ensemble, Fig. 2). But some variations in GMST trends remain between the simulations. These variations in GMST may be driven by internal variability in the ocean dynamics.~~ likely reflect the combined influence of atmospheric circulation and internal ocean variability (see Fig. 1).

A comparison of 1979-2023 GMST trends in the piControl-nudged simulations with their corresponding freely running forced simulations indicates that both of these processes are relevant (figure 2~~): The run 1300~~ b). Run 1 is the simulation with the highest GMST trend in the freely running ~~as well as~~ configuration (0.25 K/decade) and near zero trend in the piControl-nudged ~~simulation while 1500 has lowest GMST trends in both simulations~~configuration. Runs 2 and 3 have lower GMST trends than Run 1 in the freely running scenario, and both show a cooling trend in the piControl-nudged scenario. The differences between freely running simulations and their corresponding piControl-nudged simulations are not constant, indicating that there is indeed an influence of low frequency internal variability that is not controlled by atmospheric circulation. ~~(Finally, while the GMST trends in the forced scenario are well within the distribution of the CESM2 large ensemble, GMST trends of piControl-nudged runs are low compared to a distribution obtained from freely running piControl simulations. This indicates that circulation changes as simulated in CESM2 (these 3 runs) have a negative contribution to GMST trends in the 1979-2023 period, but this is note the case in later periods.)~~

~~a: Global mean surface temperature in hist+ssp370 runs (orange) and in piControl-nudged runs (green). b: GMT trend over the period 1979-2023 (gray bar in a). c: GMT trend over the period 2025-2075 (blue bar in a). The cooler histogram depicts 500 44-year trends from piControl, the warmer historgram depicts 100(?) trends from the CESM2 large ensemble.~~

~~As a result, local temperature trends in piControl-nudged simulation are influenced by changes in atmospheric circulation as simulated in the corresponding freely running forced simulation and GMST changes that might be partially influenced by these circulation changes and partially by remaining internal variability in the nudged simulation (see figure 1).~~ Therefore, we do not expect ~~that~~ the circulation component estimated from the freely running forced simulation ~~matches~~ to match the trend found in the piControl-nudged simulation exactly. ~~However, we~~ In the ridge regression and DEA, we can account for the effect of these GMST trends in the piControl-nudged simulations. In general, we assume that the ~~effect~~ impact of GMST on local temperatures is ~~rather~~ relatively homogeneous around the ~~northern hemispheric~~ mid-latitudes ~~, and thus~~ of the northern hemisphere. Thus we expect that the spatial pattern relative to the mid-latitudinal mean is well captured.

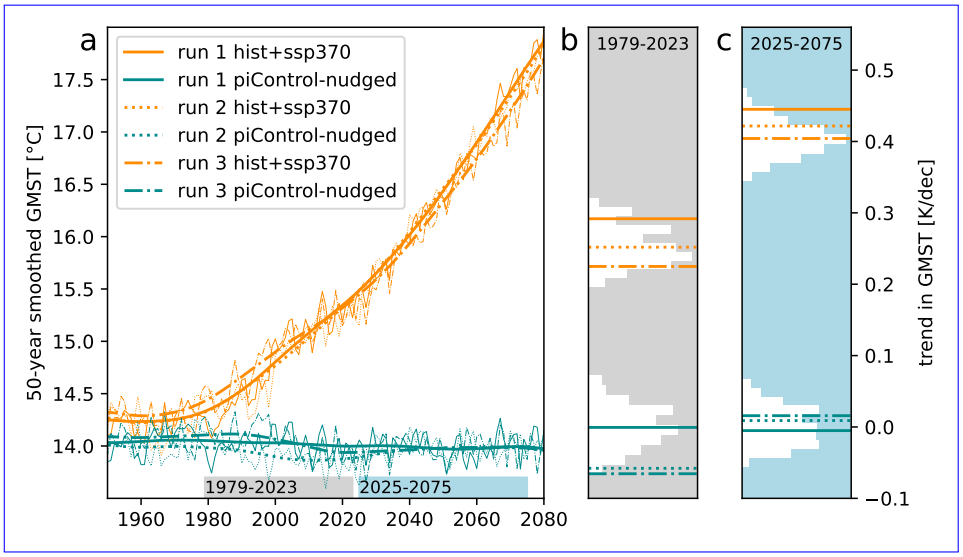

Figure 2. a: Global mean surface temperature in hist+ssp370 runs (orange) and in piControl-nudged runs (green). b: GMT trend over the period 1979-2023 (gray bar in a). c: GMT trend over the period 2025-2075 (blue bar in a). The cooler histogram displays 500 trends of the same length from piControl, while the warmer histogram displays 100 trends from the CESM2 large ensemble.

2.2 CESM2 nudged circulation simulations driven with ERA5 winds

To evaluate the use of the decomposition methods on observed circulation patterns, we created an additional benchmark simulation by nudging CESM2 to the horizontal wind fields from the reanalysis data ERA5 (Hersbach et al., 2020). This method relies on the same relaxation procedure and anthropogenic forcing as described before, with some modifications to the setup. Here, not only the piControl, but also the 'hist+SSP370' simulations are nudged to ERA5 ~~data~~ horizontal wind fields between 1940 and 2024. ~~In order to~~ To account for forcing-induced variability at the boundary later, the atmosphere is only nudged above 700 hPa~~(Wehrli et al., 2018).~~ , similar to (Wehrli et al., 2018).

The model is ~~also~~ run in an AMIP ~~setup with a prescribed ocean~~ configuration with prescribed ocean conditions from the Met Office Hadley Centre~~'~~'s sea ice and sea surface temperature ~~data set (HadISST, Rayner et al. (2003)). To create the sea surface temperature data for the~~ dataset (HadISST; Rayner et al. 2003). For the piControl simulation, ~~we removed~~ the forced component ~~from HadISST with a~~ was removed from HadISST using low-frequency pattern filtering (Wills et al., 2020). ~~The piControl sea ice data is produced~~ Sea ice concentration (SIC) was then estimated with a random forest model ~~based on the relationship between sea surface temperatures and sea ice. Somewhere I would mention that the potential downside this method is that CESM2 forced by winds that don't follow the models own climatology. This introduces a model bias .~~trained on the SST–SIC relationship in HadISST. To capture seasonal hysteresis, separate models were trained for the freezing season (Sept–Jan) and the melting season (Feb–Aug), and applied to piControl SSTs to generate corresponding SIC values. This approach reproduces daily temperature anomalies in the northern hemisphere well. However, nudging to observed winds and SSTs also introduces a temperature bias by displacing the model from its own climatology.

2.3 Dynamical adjustment methods

We test four methods designed to disentangle the dynamic effect from the thermodynamic effect: (1) ridge regression, (2) constructed circulation analogues, (3) direct effect analysis (DEA), and a neural network UNET (4).

Note that we apply the methods exactly as they were designed and used in other publications, and therefore different proxies for atmospheric circulation are used by different methods. We do not expect that the choice of variable to represent atmospheric circulation significantly affects the results. In Figure **??**, we show a sensitivity analysis for the ridge regression. In Section 3.3, we discuss differences between the methods and how they might affect the decomposition in more detail.

2.3.1 Ridge regression

For each ~~grid-cell~~grid cell, we train a linear regression model to predict daily mean temperatures in ~~JJA (Y)using as covariates~~ summer (June-July-August, hereafter referred to as JJA), Y, using GMST (yearly averaged) and the streamfunction ~~(Φ) at 500hPa in grid-cells in a rectangle of 40x40 degree around that grid-cell~~at 500 hPa in all grid cells G within a 40×40-degree rectangular area centered on that grid cell:

$$Y = \gamma_0 + \gamma_1 GMST + \sum_{j=2}^{G+1} \gamma_j \Phi_j + \epsilon, \tag{1}$$

where the streamfunction (Φ) is related to horizontal eastward (U) and northward (V) wind speeds as $u = \frac{\partial \Phi}{\partial y}$ and $v = \frac{\partial \Phi}{\partial x}$.

$$Y = \gamma_0 + \gamma_1 GMST + \sum_{j}^{grid-cells} (\gamma_j \Phi_j) + \epsilon$$

Ridge regression is a ~~linear method designed to deal with a high dimensionality of~~ regularized linear method that can deal with high-dimensional predictors (in our case, many correlated spatial locations with ~~stream function~~ streamfunction values, Φ_j). Ridge regression mitigates overfitting by introducing a penalty for model complexity~~(λ)~~, achieved through the shrinkage of regression coefficients. Shrinkage is determined by the sum of squared regression coefficients (known as L2 regularization) and a ridge regression parameter ~~(λ)~~, which controls the degree of shrinkage~~:~~

. The shrinkage term is added to the residual sum of squares (RSS) for the minimization:

$$\hat{\gamma} = \underset{\gamma}{\operatorname{argmin}} \{RSS + \lambda \sum_{i=1}^{p} \beta \|\gamma\|_2{}^2{}_i\}. \tag{2}$$

As a result, ridge regression solves a joint minimization problem, producing small but nonzero regression coefficients that are relatively evenly distributed among correlated predictors. The ~~tuning~~ regularization parameter λ ~~, which~~ dictates the extent of shrinkage ~~,~~ and is selected via cross-validation or knowledge about the noise variance. Notably, the intercept of the linear model as well as the GMST covariate remain ~~are~~ excluded from the shrinkage.

~~The streamfunction is related to horizontal eastward (U) and northward (V) wind speeds in the following way:~~

$$u = \frac{\partial \Phi}{\partial y}, v = \frac{\partial \Phi}{\partial x}$$

2.3.2 Constructed atmospheric circulation analogues technique

2.3.3 ~~Atmospheric circulation analogue technique~~

The atmospheric circulation analogue technique, introduced in Deser et al. (2016), is a linear dynamical adjustment method. It is designed to provide empirically derived estimates of climate trends induced by ~~"dynamics"~~ "dynamics" or atmospheric circulation patterns. The method has been used for a variety of applications, including trend assessments, variability analysis, performance weighting, and extreme event attribution ~~(Deser et al., 2016; Lehner et al., 2017; Merrifield et al., 2017; Guo et al., 2019; Mer~~ (Deser et al., 2016; Lehner et al., 2017; Merrifield et al., 2017; Guo et al., 2019; Terray, 2021).

The method is based on the ~~reconstruction of monthly mean atmospheric circulation fields, which we represent using~~ construction of a target monthly mean sea level pressure (SLP) ~~following Deser et al. (2016). Here, SLP analogues are selected from pools of~~ field (e.g., January 1980) using analogues. In the January 1980 example, analogues are SLP fields from other Januaries between 1850 and 2014 that resemble the target SLP pattern. The method is applied to every month in the record and proceeds as follows. First, the Euclidean distance between all SLP fields from the period of 1850-2014 and the target SLP field is computed. Euclidean distance is calculated at each grid point and averaged over the Northern hemisphere domain (20–90 °N, 0–360 °E). The N_a = 80 ~~possible choices sub-selected from the period 1850-2014~~closest SLP fields to the target are considered analogues. From the 80 ~~choices,~~ analogue choices, the target month is reconstructed using randomly selected subsets of N_s = 50 ~~SLP analoguesare selected to construct the target month, and the process~~ analogues. The process of choosing 50 out of 80 analogues and reconstructing the target SLP is repeated N_r = 100 times to obtain an average best estimate result.

~~Once the Euclidean distances are determined and 50 of the 80 possible SLP analogues are selected, we optimally reconstruct a~~ The target SLP field $\mathbf{X}_h$ ~~through a~~ is reconstructed through multivariate linear regression. The weight assigned to each SLP analogue, β, is computed through a singular value decomposition of a column vector matrix $\mathbf{X}_c$ containing the 50 selected analogues and can also be estimated using a Moore-Penrose pseudoinverse:

$$\beta = [(\mathbf{X}_c^T \mathbf{X}_c)^{-1} \mathbf{X}_c^T] \mathbf{X}_h \tag{3}$$

where β weights are ~~then applied to~~ applied to the corresponding monthly mean temperature fields, i.e., those from the same ~~months~~ month as the SLP ~~analogues to construct~~ analogue. The weighted linear combination of these fields defines the dynamic component of temperature for the target month. Before weighting~~the temperature fields, we remove~~, a quadratic trend ~~from the temperature record as a proxy for anthropogenic warming at each point in space~~ is removed from the full temperature record at each grid point to approximate the anthropogenic warming signal (Deser et al., 2016; Lehner et al., 2017). This is to approximate the unforced relationship between SLP and temperature one would expect to find in a preindustrial control simulation. ~~After repeating the dynamical adjustment protocol~~ The dynamic component of temperature is also computed N_r

= 100 times ~~, we average the results to define~~ for each month. Results are averaged, and once every month, the record is dynamically adjusted as described; we obtain a dynamic monthly mean temperature timeseries.

Note that the Moore–Penrose pseudoinverse implicitly deploys a "hard" regularization (i.e., kills directions with singular value exactly zero). In contrast, the previous ridge (Tikhonov) regularization imposes an explicit "soft" regularization (i.e., damps unstable directions even if singular values are just small, not zero).

We use the term "analogue~~"~~" to refer to a month with an SLP field close ~~, in terms of Euclidean distance,~~ to the SLP target. ~~Here, Euclidean distance is computed at each grid point and averaged over the Northern hemisphere domain (20–90 °N, 0–360 °E). This selection metric , therefore,~~ The Euclidean distance selection metric does not require an analogue to match the target month spatially over the whole domain. This step is necessary because, with ~~less~~ fewer than 200 ~~possible options~~available analogues, it is ~~extremely unlikely that a "perfect " match will exist for a particular~~ improbable to find a perfect match for any target month. ~~Van Den Dool (1994) estimated that around~~ Van Den Dool (1994) estimated that it would take on the order of 10^{30} years ~~would be required~~ to find two ~~similar (within observational uncertainty)~~ Northern Hemisphere circulation patterns ~~. Therefore~~similar within observational uncertainty. As a result, the method ~~must make the best use of~~ relies on imperfect analogues, which ~~may~~ can introduce spurious features or ~~affect~~ bias the amplitude of the estimated dynamic temperature trends.

Two analogue pool selection strategies are employed. The first is ~~an in-sample selection referred to as~~ the "leave-one-out~~" approach(cite lehnerNewEstimateTime2017);~~" approach, where analogues are selected from the ~~CESM piC-nudged run~~ free-running runs being dynamically adjusted, excluding the target month, during the 1850-2014 period ~~. The second is an out-of-sample approach;~~ (Deser et al., 2016; Lehner et al., 2017). Leave-one-out selection is used to estimate circulation-induced trends for comparison with the piControl-nudged simulations. In the second approach, analogues are selected from the entire 1850-2014 ~~periods~~ period of each of the ~~other CESM piC-nudged runs and averaged. For example, run 1300 is dynamically adjusted using analogues from run 1400, then dynamically adjusted using analogues from run 1500, and the two dynamic components are averaged. For~~ free-running runs and used to dynamically adjust ERA5~~, analogues are selected from all three CESM piC-nudged runs; dynamic components computed using~~ . The resulting three dynamical components of ERA5 computed using analogues from the free-running runs ~~1300, 1400, and 1500 analogues~~ are shown in ~~Supplementary Figure X.~~ Figure F1 and are averaged to produce the circulation-induced trend estimates in Figure 4.

2.3.3 DEA

Direct Effect Analysis (DEA) is a recently developed causal approach that aims at disentangling an outcome variable ~~$Y \in \mathbb{R}^d$~~ Y into a direct effect component, Y_{dir}, which represents the part of Y directly caused by some causal factor ~~$Z \in \mathbb{R}$~~Z, and an orthogonal component, ~~Y_{perp}~~Y_{orth}, which corresponds to the part of Y unaffected by Z. Both Z and Y may be influenced by other variables ~~$X \in \mathbb{R}^p$~~X, which act as confounders, and it is thus necessary to control for these covariates to get a correct estimate of the direct effect of Z on Y. This learned representation of Y can be seen as the result of encouraging conditional independence between Z and Y while controlling for X ~~(Durand et al., 2025a)~~[1](Durand et al., 2025b).

[1] ~~Submitted.~~

In this context, the outcome ~~Y is~~ Y represents the temperature field~~(~~, which is a random vector ~~where each dimension corresponds to a grid cell), and it is caused by Z, the GMST (monthly averaged)~~with one dimension per grid cell. The predictor Z is the monthly mean GMST, used as a proxy for ~~thermodynamical changes in temperature . We consider as covariates X~~ thermodynamic temperature changes. As covariates X, we include the leading Empirical Orthogonal Functions (EOFs) of ~~the~~ atmospheric circulation (Z500), denoted ~~as $\{p_j\}_{j=1}^{J}$~~ $\{p_j\}_{j=1}^{J}$ where J is selected through 5-fold cross-validation to maximize the $R2$ score.

Similar to the ridge regression approach described above, we assume the following linear model:

$$Y = \gamma b_0 + \mathbf{b}_0 1\text{GMST} + \sum_{j=1}^{J} {}_{j=2}^{J+1}\mathbf{b}_j p_j + \epsilon. \tag{4}$$

and get the optimal regression parameter matrix using a least squares algorithm. We obtain a matrix ~~$\mathbf{B} \in \mathbb{R}^{d\times(J+1)}$~~ $\mathbf{B}$ whose columns $\mathbf{b}_i$ encode how GMST and each EOF influence temperature across grid cells. We emphasize that this is a multivariate regression problem, where Y is a random vector—not a single variable—representing temperature values across multiple grid cells. Each dimension of Y corresponds to one spatial location. ~~The number J of EOFs is selected through 5-fold cross-validation to maximize the R^2 score.~~

To isolate the dynamical component ~~Y_{perp}~~ Y_{orth} of Y, we remove the part aligned with the GMST-related direction, as captured by ~~$\mathbf{b}_0\mathbf{b}_1$~~. This is achieved using the linear transformation ~~$Y_{\text{perp}} = \mathbf{P}^\top Y$, where $\mathbf{P} = \mathbf{I} - \frac{\mathbf{b}_0\mathbf{b}_0^\top}{\|\mathbf{b}_0\|^2}$~~ $Y_{\text{orth}} = \mathbf{P}^\top Y$, where $\mathbf{P} = \mathbf{I} - \frac{\mathbf{b}_1\mathbf{b}_1^\top}{\|\mathbf{b}_1\|^2}$. This transformation ensures that ~~Y_{perp}~~ Y_{orth} remains unaffected by any interventions on GMST, and thus represents the dynamical component of Y.

2.3.4 UNET

The ~~UNET~~ final method used in this paper is a convolutional neural network~~designed by Ronneberger et al. (2015) , a UNET~~ structure, recently proposed by Cariou et al. (2025) to link temperature variations to atmospheric circulation. The UNET architecture was initially introduced by Ronneberger et al. (2015) for biomedical image segmentation. It consists of two main ~~parts: the encoder , which captures the global features of the input~~ components: an encoder and a decoder. The encoder extracts global features from the input (in this case, circulation maps) by progressively reducing ~~its~~ spatial resolution while increasing ~~the depth of the feature maps via~~ feature depth through convolution and max-pooling layers~~, and the decoder , which~~ . The decoder then reconstructs the image using transposed convolutions. ~~These two partsare symmetrical and connected by~~ Symmetry between the two parts, combined with skip connections, ~~enabling~~ allows the network to preserve ~~the~~ and effectively reuse encoded information.

We use this architecture (see **??**) to estimate the ~~relationship between daily standardized sea-level pressure (SLP) maps (input) and the corresponding temperature anomalies (output). Temperature anomalies are obtained by removing an estimate of the daily non-stationary normal (Rigal et al. (2019)) from the raw temperature. The anomalous temperature of a given day (d)and a given year (y) can be written as $Ta(d,y) = T(d,y) - (f(d) + g(y) \times h(d))$. The second term corresponds to the non-stationary normal where $f(d)$ the mean seasonal cycle, $g(y)$ the yearly forced response and $h(d)$ the distortion of the annual cycle. The UNET is trained~~ part of daily temperature variations (T', the output) which can be explained by the

large-scale circulation, described by the sea level pressure (SLP, the input). Thus, we can write the UNET model as

$$T' = F(SLP). \quad (5)$$

We follow the methodology described in Cariou et al. (2025). Still, we extend the analysis to a larger spatial domain and train the UNET on daily data from 1850 to 2100 from 8 CESM2 transient simulations ~~.~~ (80% of the data are randomly selected for training, and the remaining 20% are used for ~~the validation. SLP maps are standardized by removing at each grid point the mean and standard deviation computed over all days from the 8 concatenated simulations. The parameters chosen for training are a batch size of 300,~~ validation). Since the UNET is trained on transient runs (historical and SSP), we must consider climate change in the relationship 5. The SLP is not detrended, assuming that in the CESM2 model, the forced responses in the SLP is small compared to the daily variability. This assumption is supported by Figures 2 and 3, which show that the three piControl-nudged experiments do not exhibit significant common trends. However, the forced response is substantial in terms of temperature. Thus, the ~~Adam optimizer with a learning rate of 0.001. The Mean Squared Error (MSE) is used for training and validation loss. The number of epochs is set to 100, but the process stops if the validation loss does not improve after 10 epochs. Then, the trained UNET is tested on the SLP maps of the~~ detrending is made following Rigal et al. (2019): temperature anomalies (T') are obtained by removing an estimate of the daily non-stationary normal containing both the mean seasonal cycle, which is not circulation-explained, and the climate change signal. The trained model is then tested on three CESM2 piC-nudged runs~~standardized with~~ , with SLP maps standardized using the same values as ~~for training~~ in the training process.

For ERA5, we use the UNET that was previously trained on CESM2 and ~~we~~ retrain it on ERA5 data from 1940 to 1978. This process is known as a fine-tuning method. SLP maps are standardized with mean and ~~stantard~~ standard deviation calculated on this training period, and the non-stationary normal is computed thanks to an estimate of ~~$g(y)$~~ the forced response obtained with Qasmi and Ribes (2022) method. Then we test it over the 1979-2023 period.

3 Results and Discussion

~~In the following, we~~ We evaluate each statistical method~~'~~'s estimate of circulation-induced mid-latitude JJA temperature trends. ~~To obtain those estimates, each statistical method is applied to obtain a circulation-induced temperature estimate from the~~ Estimated trends $\hat{y}$ are derived from CESM2 free-running hist+ssp370 ~~climate change simulations , and subsequently evaluated against the piControl-nudged CESM2~~ simulations ~~, which is used as a benchmark. The skill of each method is assessed by calculating~~ and compared against trends in CESM2 piControl-nudged simulations y, which serve as the benchmark. Method performance is assessed using four skill metrics: (i) the fraction of correctly identified trend signs, (ii) Pearson correlation ~~, that is a~~ (pattern correlation across the mid-latitudes), (iii) the coefficient of determination (~~R^2 score, that is the proportion of variance of circulation-induced trends captured by the statistical predictions),~~ $R^2 = 1 - \sum(y - \hat{y})^2 / \sum(y - \bar{y})^2$), and (iv) the regression slope between ~~the~~ predicted and benchmark trend estimates.

(i) This metric provides a general sense of whether the method can correctly capture the sign of the trend, which may be sufficient in specific contexts—for example, in climate change detection. (ii) Pearson correlation reflects how well the method

captures the spatial pattern of the trend. Some methods may systematically over- or underestimate the magnitude of trends, yet still accurately reproduce their spatial distribution. (iii) The coefficient of determination (R^2) is a widely used metric for spatial comparisons, as it accounts for the variance at each location and indicates how much of the observed variability is explained by the prediction. Yet, it is, in contrast to Pearson correlation, sensitive to any bias in the estimated average (Kvålseth, 1985); and hence, a statistical method may show a good spatial Pearson's correlation in its estimates but a poor R-squared score. (iv) The regression slope indicates whether the method tends to overestimate or underestimate the magnitude of trends.

3.1 Evaluation of circulation-induced trends in the historical period (1979-2023) in CESM2 nudged-circulation simulations

Over the period 1979-2023, JJA temperature trends in the piControl-nudged simulations range from -0.35 to 0.35 K per decade (figure 3 a,d,g). These trends are ~~heterogeneous throughout the northern hemispheric mid-latitudes with large regional patches of trends of the same sign~~organized in large regional clusters of alternating signs. Furthermore, the trend patterns differ considerably between the three piControl-nudged runs, indicating that in CESM2~~circulation induced~~ , circulation-induced trends are dominated by internal variability and that in CESM2, forced circulation changes are minor. Overall, JJA temperature trends are slightly higher in ~~the '1300' run~~ run 1, which is likely due to the positive GMT trends in the piControl-nudged runs during this period (see ~~figure~~ Figure 2).

Note that most of these trends are not statistically significant (see figure **??**). Since these trends mostly reflect internal climate variability, it is expected that, from a statistical point of view, the circulation-induced temperature changes at one location are not differentiable from noise. The spatially consistent trend patterns indicate that, although lacking statistical significance, these trends contain valuable information and are worth evaluating.

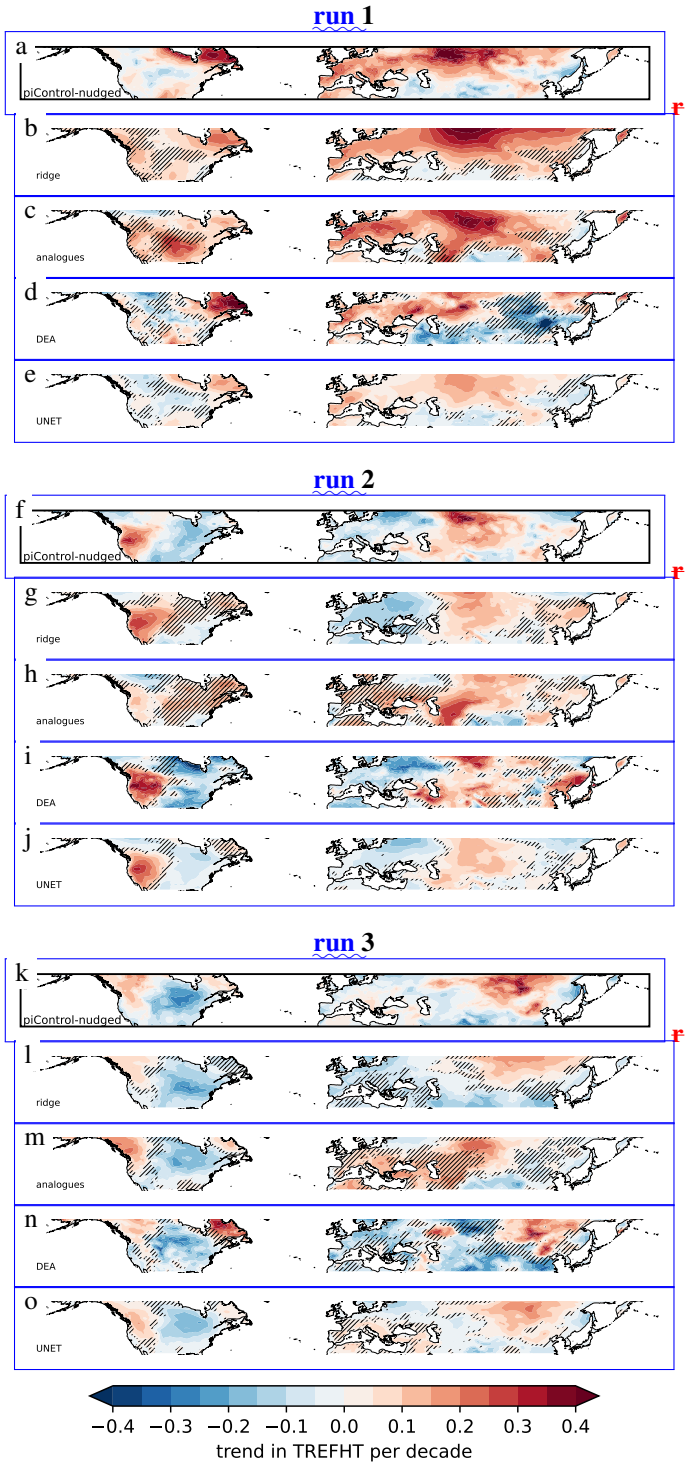

Figure 3. Trend in JJA temperatures over the period 1979-2023 in piControl-nudged (a,f,k) and predicted trends from different decomposition methods. For ~~the run 1300~~ runs 1 (a-e), ~~1400~~ 2 (f-j), and ~~1500~~ 3 (k-o). Estimates from the ridge regression (b, g, l), the analogues (c, h, m), DEA (d, i, n), and UNET (e, j, o). Areas where the predicted trend differs in sign from the piControl-nudged run are highlighted by black hatching.~~Besides the estimated trend maps from decomposition methods, small kernel density estimates of estimated trends (y-axis) versus simulated piControl-nudged trends (x-axis) are shown with x- and y-axis ranging from -0.5 K/dec to 0.5 K/dec.~~

Table 1. Evaluation metrics comparing trends in piControl-nudged simulations to estimates of circulation induced trends from statistical decomposition methods for land grid-cells between 30N-60N and the period 1979-2023. First block: percentage of correctly predicted ~~sign~~signs. Second block: Pearson correlation coefficient. Third block: Coefficient of determination. Fourth block: regression slope (as shown in figure **??**). See table **??** for the same evaluation over the period 2025-2075.

run	ridge	analogues	DEA	UNET
	correct sign			
all runs	**75%**	**65%**	**76%**	**84%**
1300	~~76~~75%	77%	73%	86%
1400	74%	56%	82%	84%
1500	~~76~~77%	63%	~~73~~74%	83%
	Pearson correlation (r)			
all runs	**~~0.74~~0.75**	**0.52**	**0.64**	**0.86**
1300	0.79	0.57	0.61	0.91
1400	~~0.65~~ 0.67	0.36	0.74	0.83
1500	0.75	0.41	0.58	0.89
	coefficient of determination (R2)			
all runs	**0.53**	**0.07**	**~~0.10~~0.08**	**0.66**
1300	~~0.56~~ 0.54	0.18	-0.19	0.50
1400	0.39	-0.28	~~0.25~~ 0.22	0.67
1500	~~0.47~~ 0.48	-0.07	-0.08	0.73
	regression slope			
all runs	**~~0.65~~0.66**	**0.43**	**~~0.73~~0.74**	**0.51**
1300	0.59	0.47	~~0.70~~ 0.71	0.47
1400	~~0.51~~ 0.54	0.26	~~0.96~~ 0.97	0.57
1500	~~0.54~~ 0.55	0.36	~~0.65~~ 0.68	0.58

~~Next, we discuss the performance of each statistical method.~~ Using the ridge regression trained on a forced simulation to predict the trends based on the streamfunction of the forced simulation and the GMT of the piControl-nudged run, we get a similar trend pattern as in the piControl-nudged run (figure 3 b,e,h). Over the mid-latitudinal land area, half of the variability in local temperature trends in the piControl-nudged run is explained by the ridge regression model (compare R2 score in table

1). For ~~three quarters~~ three-quarters of the grid-cells, the sign of the predicted trend is correct, and grid-cells for which the sign of the trend is not ~~predicted~~ indicated correctly are mostly grid-cells with small trends in the piControl-nudged simulation (and the prediction).

The ~~analog method also shows~~ analogue method reveals a positive correlation between predicted and simulated mid-latitude land trends~~and~~ , with a similar percentage of correctly ~~predicted signs of trends for the 1300 run~~ identified signs in Run 1 and slightly lower ~~performance for the two other runs. Note that in the analogues method we estimate the trend contribution coming from circulation and cannot account for contributions of GMST changes in the piControl-nudged simulation. It is therefore expected that the results are worse for the 1400~~ skill in Runs 2 and ~~1500 runs where a considerable negative trend in GMST was simulated~~3. Importantly, the analogue method only captures the circulation-driven component of the trend and does not account for GMST contributions in the piControl-nudged ~~runs.~~ simulations. Consequently, performance is lower in runs 2 and 3, where substantial negative GMST trends were simulated. It is important to note, however, that an offset largely influences the relatively poor R2 score in the mean circulation trend (overestimated warming), while the spatial pattern itself shows a rather good resemblance and Pearson correlation compared to the benchmark simulation (figure 3).

The DEA method performs well in estimating the sign of ~~circulation induced trends (~~circulation-induced trends with 76% ~~correct; see table~~ accuracy (see Table 1). Despite the relatively ~~good~~ strong correlation between the trend maps ($r = 0.64$), the coefficient of determination is close to zero. Estimates of ~~circulation induced~~ circulation-induced trends from DEA cover the full range of simulated piControl-nudged trends, including very high and very low trends. This is reflected by a relatively high regression slope between predicted and simulated trends (figure **??** and last block in table 1).

The UNET is performing the best of all tested methods here. With UNET, 84% of ~~trends~~ trend signs are predicted correctly~~,~~ ; it has the highest Pearson correlation coefficient (0.86) and the highest coefficient of determination (0.66). ~~As compared~~ In comparison to the DEA and ~~the~~ ridge regression, UNET ~~has a tendency~~ tends to predict lower ~~circulation induced~~ circulation-induced trends and rarely ~~exceeding~~ exceeds magnitudes of 0.2 K per decade. As shown in figure **??**~~d~~ , this leads to a systematic underestimation of the ~~magnitude of stronger trends; and a systematic overestimation of negative trends~~trend magnitude compared to the piControl-nudged simulation.

~~For the period~~ The evaluation over a different period (2025-2075~~the results are similar with all methods showing skill in projecting the sign of the trend and ... (do we need to go into detail here?)~~) yields similar skill metrics and confirms the above-discussed results (see Figure **??** and Table **??**).

Overall, UNET is the most accurate method ~~when it comes to explain~~ for explaining the variance in mid-latitude boreal summer temperature trends. The ridge regression and UNET ~~have the tendency to decompose the~~ tend to decompose temperature trends into a regionally smoothed pattern of circulation-induced temperature trends~~and have~~ , with a lower likelihood of predicting a strong trend of the wrong sign. DEA and the analogue method project strong trends of the wrong sign in some regions. DEA appears to be more ~~useful when it comes to estimate~~ helpful in estimating the potential magnitude of circulation-induced trends~~while UNET is rather conservativeand estimates generally to weaktrends~~, whereas UNET is relatively conservative, estimating trends that are generally too weak. The analogue method also shows skill in predicting the sign of trend~~but appears generally~~ , but appears typically less trustworthy when it comes to circulation-induced trends.

3.2 ~~Indentification~~ Identification of the circulation-induced boreal summer temperature trend (1979-2023) in ~~observations~~reanalysis

~~All decomposition methods applied to~~

Across both the ERA5 reanalysis ~~as well as~~ and the CESM2 simulations nudged to ERA5 winds~~suggest similar circulation induced~~ , all decomposition methods reveal similar circulation-induced trend patterns for ~~the period 1979-2023:~~ 1979–2023. Over Eurasia, a wave-like ~~trend pattern for the circulation induced trends is apparent with strong positive trends~~ structure emerges: strong warming over Central and Eastern Europe (around ~~30E~~30°E), cooling ~~trends~~ over Kazakhstan and western Siberia (~~between 60E and 90E~~60°–90°E), and warming ~~trends~~ again over Mongolia, eastern Siberia, and Central China (~~between 90E and 120E), and extending towards the Kamchatka peninsula~~ 90°–120°E), extending toward the Kamchatka Peninsula (Fig. 4). ~~Over North America~~Across North America, we observe a dipole pattern with positive trends ~~over~~ in the western part and negative trends ~~over the center~~ in the central and eastern parts~~;~~ , with positive trends again in the outermost ~~north-eastern partsagain~~northeastern parts. All decomposition methods as well as the CESM2 simulations nudged to ERA5 winds (figure 4j) agree on this broad trend pattern with only little regional deviations. The trend pattern identified with the statistical method is ~~thus~~ in good agreement with the pattern found in Teng et al. (2022).

~~From~~

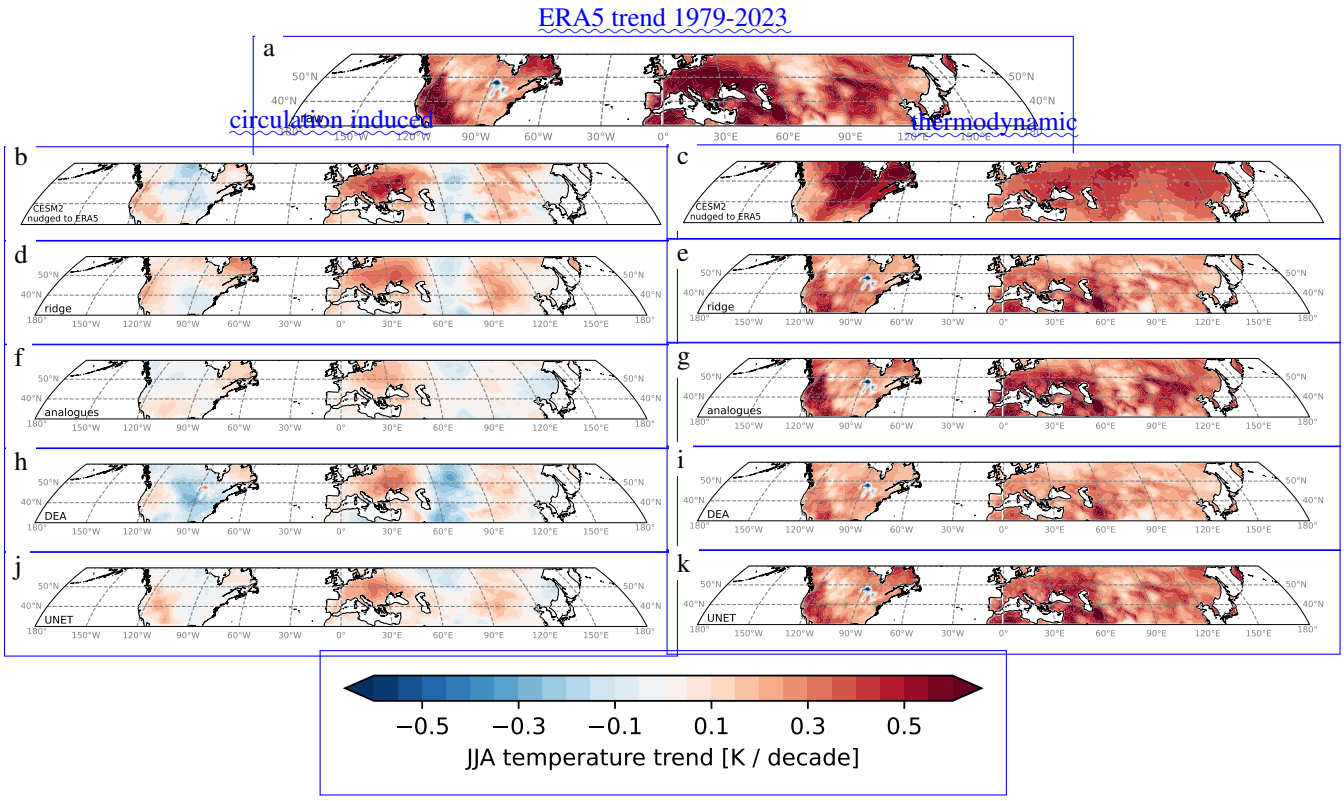

Figure 4. JJA mean temperature trends in ERA5 over the period 1979-2023 (a) decomposed in the circulation-induced (left) and thermodynamic (right) contribution for the ridge regression (d,e), DEA (f,g), analogues (h,i), UNET (j,k) and estimates from CESM2 simulation with horizontal winds nudged to ERA5 winds (b,c).

Based on our evaluation of decomposition methods against dedicated nudged experiments in the CESM2 setup, we would suggest ~~to give~~ giving more weight to the results from UNET ~~when it comes to~~ regarding the sign of ~~circulation induced~~ circulation-induced trends. For example, this would imply that the positive circulation induced trends over northern North America in the ridge regression are probably wrong (compare low skill of ridge regression in this region, figure 3).

Moreover, there remains ambiguity on the magnitude of the ~~above described~~ above-described trend pattern. The ridge regression, DEA, and UNET suggest a ~~circulation induced~~ circulation-induced trend of up to 0.3 K/dec over eastern Europe~~while the~~ . At the same time, the piControl simulations from CESM2~~that where~~ , which were nudged to ERA5 winds~~stronger circulation induced~~ , show stronger circulation-induced trends of up to 0.6 K/dec. In other regions, the nudged simulations and DEA ~~have stronger trends~~exhibit stronger trends, followed by the ridge regression and analogues~~and UNET suggest~~ , while UNET suggests somewhat weaker trends. From the evaluation of the decomposition methods, we know that all methods ~~except DEA~~ have indeed a tendency ~~of somewhat underestimating~~ to underestimate the magnitude of ~~circulation induced trends~~ circulation-induced trends somewhat, suggesting that the ~~circulation induced~~ circulation-induced trend over eastern Europe could be around 0.5 K/dec as ~~suggested~~ indicated by the nudged simulations~~and DEA~~.

In summary, our study confirms the highly variable mid-latitude boreal summer trend pattern found in ~~Teng et al. (2022)~~ Singh et al. (2023); Teng et al. (2022); Vautard et al. (2023) with five independent methods (four statistical methods and a nudged circulation simulation driven by ERA5 horizontal wind fields). The trend pattern ~~contains substantial~~ highlights several regional warming hotspots ~~in all of which a positive circulation contribution to warming has played a key role Teng et al. (2022). It is important to note that while the~~ where circulation has made a major positive contribution (Teng et al., 2022). While total boreal summer temperature trends are positive ~~all over~~ across the NH mid-latitudes, circulation has ~~contributed with cooling temperatures in large regions, in particular~~ driven cooling in large areas—most notably Central and Eastern North America, Central Eurasia, and~~to a minor extent~~, to a lesser extent, coastal eastern China. ~~Moderately~~ In these regions, the moderately positive total trends ~~in those regions are therefore due to a compensation of the~~ reflect a compensation between circulation-induced cooling ~~by a positive thermodynamic contribution~~and thermodynamic warming.

~~ERA5 trend 1979-2023circulation inducedthermodynamicJJA mean temperature trends in ERA5 over the period 1979-2023 (a) decomposed in the dynamic (left) and thermodynamic (right) contribution for the ridge regression (b,c), DEA (d,e), analogues (f,g) and CESM2 simulation with horizontal winds nudged to ERA5 winds (h,i).~~

3.3 Implications and limitations of statistical methods to isolate circulation effects on the time scales of climate trends

Many ~~previous studies aimed to identify~~ studies have sought to isolate circulation-induced components in ~~time series of climate variables, a methodology which is sometimes framed as 'dynamical adjustment ' (Smoliak et al.~~ climate time series, ~~2015; Deser et al., 2016; Guo et al 2019; Cariou et al., 2025; Singh et al., 2023; Safioti et al., 2017; Lehner; ...). The main idea~~ often referred to as dynamical adjustment (Smoliak et al., 2015; Deser et al., 2016; Guo et al., 2019; Cariou et al., 2025; Singh et al., 2023; Saffi . The core assumption is that circulation variability, primarily driven by internal processes, dominates temperature variability in many regions~~of the world and is related to internal variability; thermodynamic contributions are obtained as a residual (e.g., Deser et al. , 2016). The identification and separation of climate time series into~~ , while thermodynamic contributions can be derived as the residual (e.g. Deser et al., 2016). This separation of dynamic and thermodynamic ~~is a powerful tool for attribution (e.g., Shepherd 2014)~~components provides a powerful framework for climate attribution (e.g. Shepherd, 2014).

However, while different statistical methods ~~to obtain the~~ for obtaining circulation-induced components in climate time series are routinely evaluated on short time scales, the estimation of circulation-induced decadal trends has remained a challenge for the climate community and will likely ~~remain one~~continue to do so. This is because of five main reasons.

First, statistical methods have been found to perform very well on short time scales of day-to-day, month-to-month, or inter-annual variability ~~(Cariou et al. , 2025; Smoliak et al., 2015; Sippel et al., 2019).~~ (Cariou et al., 2025; Smoliak et al., 2015; Sippel et al., 201 . Yet, the difference in the performance on long (that is, trend) time scales versus short time scales has not been quantified so far~~, although~~ . Nevertheless, dynamical adjustment has been widely applied on the time scales of trends. A reduced performance on long time scales is expected, and indeed found in this study, because shorter time scales are dominated to the largest extent by circulation-induced variability, whereas on longer time scales other processes are becoming more dominant, such as land-atmosphere interactions ~~(e.g. Merrifield et al., 2017)~~ (e.g. Merrifield et al., 2017) or long-term warming, both of which may not be straightforward to account for.

Second, and partly related to the previous point, designing a method comparison for the identification of circulation-induced time series is challenging. This is because it is not immediately apparent what ~~are the components~~ the components are that the signal is decomposed ~~to; and~~ into, and which relevant mechanisms can be attributed to these components. In this study, we decompose a trend in local temperatures into a ~~"circulation induced"~~ "circulation-induced" component and a thermodynamic component without specifying to which of these ~~component~~ components changes in other important factors, such as ~~soil-moisture~~ soil moisture or aerosol concentrations, are attributed ~~to~~(see figure 5 for the example of land-atmosphere feedbacks). The different statistical methods evaluated here were initially developed for similar but slightly different research questions: The analogue method, for instance, was designed to separate the ~~"thermodynamic signal" from "~~"thermodynamic signal" from "circulation-induced variability~~".~~." Yet, it has been shown that summer land-atmosphere interactions remain largely in the residual, thermodynamic component due to the ~~way that the methodis set up (Merrifield, 2017)~~method's setup (Merrifield et al., 2017). On the other hand, machine learning methods such as the UNETs may partly implicitly identify land-atmosphere interactions as part of circulation variability, if circulation carries an imprint of land-atmosphere variability. In table 2 we summarize our thoughts on the treatment of land-atmosphere feedbacks in the different approaches.

~~Third, it is not straightforward to design~~

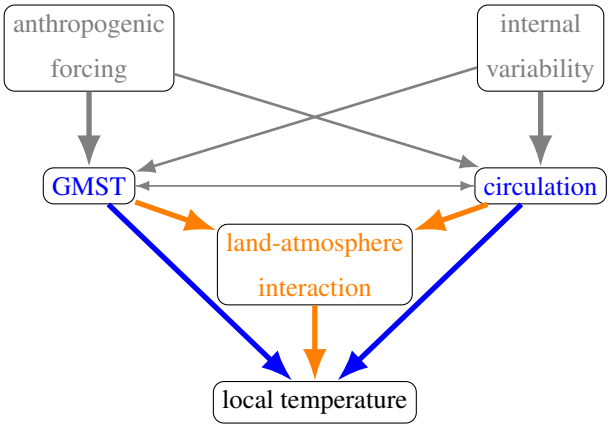

Figure 5. Conceptual illustration of causal relationships influencing local temperatures in a forced climate as in figure 1 but with an additional driver of local temperature. In this study we aim at decomposing the influence on local temperature into thermodynamic (GMST) and circulation induced contributions (in blue). Land-atmosphere interactions is a driver we do not explicitly model (in orange).

Third, designing a benchmark for the circulation-induced component ~~in time series of climate variables~~of climate time series, such as summer temperatures~~. Here, we have used a pi-control~~ , is a challenging task. In this study, we use a piControl nudged-circulation approach~~as a benchmark~~, where a climate model ~~was~~ is nudged to the horizontal winds of a forced ~~,~~ transient simulation. ~~We thus obtain~~ This setup provides circulation-induced changes ~~in an~~ within an otherwise unforced climate simulation. However, there may be factors of residual climate variability (such as ocean variability) or feedbacks between circulation and other factors, such as land-atmosphere coupling, that could still affect thermodynamical processes on climate over land ~~. However, based on the similarity between the pi-control~~ (see table 2). Additionally, summer temperatures in the

nudged circulation simulations ~~and the statistical methods , the climate model based benchmarkis valid for our purpose~~might be affected by nudging in other seasons. For example, circulation changes can influence soil moisture in late spring which would then have an impact on summer temperatures. Statistical decomposition methods do not use this information. Consequently, we have to admit that the nudged simulations are not a perfect benchmark. Further analysis is required to understand how these limitations affect our estimates of circulation-induced trends and whether a better-suited benchmark test could be designed.

~~Fourth, besides combining multiple lines of evidence, our study highlights the need of benchmarking efforts for statistical and machine learning approaches: without the evaluation against nudged circulation simulations one would conclude that different decomposition methods project similar trend patterns with some estimates showing a stronger version of the trend pattern than others. Evaluating which magnitude of the trend pattern is the most likely /plausible is challenging from the statistical analysis alone. Concluding that all~~

Table 2. Expectations on how land-atmosphere interactions might influence the decomposition into "circulation-induced" and "thermodynamic" contributions.

method	what is variability in land-atmosphere interactions attributed to?
ridge	In multiple linear regression, attribution depends on the collinearity of covariates (e.g., circulation) with the second-order effect (land–atmosphere interaction), determining whether it is assigned to GMST or circulation changes. This partitioning may vary by region. If there is no strong collinearity between the prevailing atmospheric circulation at the daily time scale and land-atmosphere interactions, which typically change at longer time scales (Merrifield et al., 2017), we expect the effects due to land-atmosphere interactions to remain in the residuals.
analogues	It is presumed that circulation analogues occur over a range of land surface states. The circulation-induced component of temperature is defined as an average across this range, which leaves the influence of the land surface on the atmosphere predominantly in the residual thermodynamic component (Merrifield et al., 2017). The land surface can induce a temperature anomaly and associated circulation pattern (e.g., a thermal low), and the analogue method could interpret this situation as circulation-induced rather than thermodynamic. Nudging all vertical levels of the atmosphere suppresses influence from the land surface to the atmosphere, so land-atmosphere interactions are likely to remain in the thermodynamic component of the piControl-nudged runs used as a benchmark in this study (Merrifield et al., 2019).
DEA	Because the approach removes the total effect of GMST without conditioning on land–atmosphere interactions, it may also eliminate the mediating effect of GMST operating through this pathway (but this effect is likely small and confined to trends that are colinear with GMST), whereas the mediating effect of atmospheric circulation is expected to be retained, given that the linear model has sufficient expressive capacity to capture these complex relationships.
UNET	The SLP is used as a predictor of the circulation. However, this variable may contain surface imprints which might affect the "circulation-induced" component. Therefore, land-atmosphere interactions may be partly predicted by the UNET architecture.
nudged simulations	Nudging is expected to separate the land-atmosphere interactions into a thermodynamically-driven, and a circulation-driven component (i.e., atmospheric imprints of land-atmosphere interactions are expected to be captured through nudging).

Regarding land-atmosphere interactions, we conclude that the effect on our estimates of circulation-induced trends varies between methods. This increases our confidence in the signals all methods agree on (e.g., circulation-induced warming over Europe). At the same time, there is no systematic (and consistent) difference in how statistical decomposition methods might be ~~underestimating the magnitude of the trend pattern would be impossible~~affected by land-atmosphere interactions in comparison to how land-atmosphere interactions might influence the nudged simulations. Therefore, the effect of land-atmosphere interactions cannot explain the systematic underestimation of the magnitude of circulation-induced trends in statistical decomposition methods (as compared to the nudged simulations).

~~Fifth~~Fourth, we present an evaluation of decomposition methods based on one set of nudged simulations from one ~~earth~~ Earth system model (CESM2). ~~Although the well documented~~ Despite the well-documented performance of CESM2, this is a flaw, as the strength of the links between atmospheric circulation patterns, GMST, and local temperatures might be misrepresented in the model. A ~~followup~~ follow-up study using multiple ESMs to create a benchmarking dataset would be crucial to ~~further~~ constrain our estimates of ~~circulation induced temperature trends .~~ circulation-induced temperature trends further. The use of such a multi-model ensemble would require adapting the different reconstruction methods slightly. In particular, the UNET would need to be pre-trained on a collection of multi-model data, rather than just CESM2. Preliminary tests conducted on Western Europe suggest that this does not degrade the quality of the reconstruction, especially when the fine-tuning step is applied to early ERA5 data.

~~Overall, our study shows that statistical methods are well-suited to~~ Finally, in addition to combining multiple lines of evidence, our study emphasizes the importance of benchmarking efforts for statistical and machine learning approaches. Without evaluating the results against nudged circulation simulations, one would conclude that different decomposition methods project similar trend patterns, with some estimates exhibiting a stronger version of the trend pattern than others. Evaluating which magnitude of the trend pattern is the most likely/plausible is challenging from the statistical analysis alone. Concluding that all decomposition methods applied to observations might be underestimating the magnitude of the trend pattern would be impossible.

Overall, our study demonstrates that statistical methods can effectively identify and separate circulation-induced temperature trends from residual ~~, thermodynamical trends. It is important to note, however, that the performance of those methods on time scales of climatic trends decreases relative to the higher performance found on short time scales. It is therefore crucial to account for this kind of uncertainty in potential subsequent studies that would use those~~ thermodynamic trends. However, their performance declines on climatic timescales compared to shorter timescales. This uncertainty should be carefully considered in future studies that use such estimates for attribution ~~of to derive constraints on future~~ or to constrain projections.

4 Conclusions and Outlook

In summary, our analysis targeted two specific research objectives and revealed two distinct findings: First, we evaluated whether statistical-empirical methods ~~are able to correctly~~ can accurately estimate circulation-induced long-term trends in the NH mid-latitudes ~~in~~ during boreal summer (and a residual ~~that is~~ dominated by thermodynamic trends) against a specifically

designed climate model benchmark of nudged circulation experiments. Four different statistical methods were tested~~and we showed~~ , and we demonstrated that each of these methods ~~is able to identify in general~~ can generally identify the large-scale pattern of circulation variability and changes, even though ~~the methods~~ they are typically trained and validated on short time scales (daily to seasonal). However, the methods showed differences in their ~~skill in which they reproduced~~ ability to reproduce the spatial trend pattern from the nudged circulation benchmark~~, and in the extent to which they partly underestimated the magnitude of circulation trends: With three quarters~~ . With three-quarters of correctly estimated signs of trends and coefficients of determination above 50%, the ridge regression and the UNET methods are performing sufficiently well for the purpose. The UNET has ~~overall the~~ the overall highest scores in most tested skill metrics. However, the UNET method tends to produce underdispersive results, that is, the magnitude of particularly ~~large (or weak)~~ strong circulation trends is often underestimated (~~or overestimated~~irrespective of the sign). DEA and circulation analogues have similar skill in predicting the sign of circulation-induced trends~~, but~~ . Still, due to the low coefficient of determination~~we would restrain~~ , we would refrain from interpreting the magnitude of regional trends estimated from these methods. Overall, identifying circulation-induced trends on climate time scales in the context of dynamical adjustment studies is ~~skillful, but~~ possible. Still, it does imply larger uncertainties than for the application on shorter time scales, which needs to be considered in future applications of the techniques.

Our second objective was to identify circulation-induced boreal summer temperature trends across the northern ~~hemispheric~~ mid-latitudes ~~in observations using four different~~ using four statistical methods and CESM2 simulations ~~that are nudged to the~~ nudged to ERA5 circulation without anthropogenic forcing. Large-scale ~~boreal~~ summer circulation trends and their ~~effects on temperature have been a topic of intense discussion (Teng et al. , 2022; Chemke and Coumou 2022; Rousi et al 2022; Vautard et al., 2023). Our analysis of circulation-induced trends in~~ temperature impacts have been widely debated (Teng et al., 2022; Chemke and Coumou, 2024; Rousi et al., 2022; Vautard et al., 2023). Analyzing ERA5 ~~over the period 1979-2023 confirms the positive circulation contribution to summer heat~~ for 1979–2023, we find positive circulation contributions to summer warming over Europe, ~~Western~~ western North America, and ~~over Mongolia. Following a wave structure,~~ Mongolia. In contrast, a wave-like pattern of circulation-induced cooling ~~has been identified~~ appears over Central Eurasia (~~Western~~ west Siberia and Kazakhstan) and Central ~~and~~ to Eastern North America.

~~We anticipate that the methods presented and evaluated in our present study will be applied in a variety of use cases in the future:~~

~~Dynamical adjustment of historical climate change; -> improved understanding and interpretation of which changes the distribution of climate variables like temperature are driven by dynamical vs. thermodynamical changes. AKA similar to Deser et al (2016; 2023) ... ; Attribution of~~ Beyond strengthening confidence in circulation-induced temperature changes, our evaluation also highlights systematic limitations of statistical decomposition methods. Improving understanding of their performance will enhance our ability to attribute regional climate trends. Isolating individual components of historical ~~changes~~ change is likely to ~~identify a stronger signal, especially when it comes to the attribution of~~ yield a stronger attribution signal, particularly for regional climate change~~; and to allow focussing~~ . Focusing on dynamical and ~~thermodynamical changes separately (e.g., Shepherd. ..).~~ thermodynamic changes separately is advantageous, as there are significant differences in the uncertainties of forced changes in these components (Shepherd, 2014). While several attribution studies of circulation changes

have been published ~~(Chemke and Coumou 2024);~~ (Coumou et al., 2015; Chemke and Coumou, 2024; Dong et al., 2022), uncertainties remain large, especially when it comes to ~~the attribution of the~~ attributing the downstream impacts of atmospheric circulation changes~~, such as whether *circulation-induced* temperature changes are due to forced or unforced variability. observations-based constraints on future projections separately for dynamical and thermodynamical components . -> likely relatively~~ . Being able to more robustly decompose a trend into a circulation-induced and a thermodynamic component should also help attribute circulation-induced temperature trends more effectively.

Finally, separating dynamical and thermodynamic components offers a pathway to constrain near-term climate projections using observation-based constraints. The thermodynamic constraint should be straightforward to identify~~thermodynamical constraints (because largely forced). large uncertainty for the dynamical components, but could be the basis for internal variability storylines (Liné et al 2024): If historical~~ , as it is mainly forced. There are different possibilities to constrain based on the dynamical component. With the assumption that circulation-induced ~~temperature change is~~ trends over the past decades were primarily due to internal ~~variability, likely reversal in the circulation-induced warming hotspots in the next 1-2 decades(e.g. Europe), while thermodynamical warming would continue~~climate variability, one would expect a reversal of the observed trend pattern over the coming decades. If circulation-induced temperature change is forced, both circulation-induced and ~~thermodynamical~~ thermodynamic trends would continue~~into the future. ..~~. Due to the considerable uncertainty in the forced circulation-induced changes, a storyline approach would be appropriate, explicitly treating different assumptions about forced atmospheric circulation changes and evaluating the potential outcomes of these scenarios (Shepherd, 2019; Liné et al., 2024).

. The code required to reproduce this study is available on https://github.com/peterpeterp/circ_contribution_to_JJA_trends.git.

Appendix A: Nudged circulation plots

. PP and SS conceived the study. PP wrote the manuscript with contributions from SS and all other authors. PP created all figures. PP contributed the results of the ridge regression. AM contributed the results for the constructed analogues. HD contributed the results for the DEA. EC and JC contributed the results for UNET. ID contributed the CESM2 simulations nudged to ERA5 winds.

. No competing interests are present.

. We thank Urs Beyerle for producing the nudged CESM2 simulations that were used to evaluate the statistical decomposition methods. ~~CESM2 large ensemble CESM2 model Development~~ We thank all the scientists, software engineers, and administrators who contributed to the development of CESM. We acknowledge the use of ERA5 reanalysis data provided by the Copernicus Climate Change Service (C3S). We acknowledge the CESM2 Large Ensemble Community Project for providing the model output used in this study. S.S. and P.P. acknowledge the project 'Artificial Intelligence for Enhanced Representation of Processes and Extremes in Earth System Models' (AI4PEX; grant agreement 101137682), funded by the EU's Horizon Europe program; and the climXtreme project funded by the German Federal Ministry of Education and Research (Phase 2, project PATTETA, grant number 01LP2323C). S.S. and I.D. acknowledge funding from the German Research Foundation's Heinz Maier-Leibnitz Prize 2024 for ~~early career~~ early-career researchers.

References

camdoc: Camdoc Documentation: 9 Physics Modifications via the Namelist, https://ncar.github.io/CAM/doc/build/html/users_guide/physics-modifications-via-the-namelist.html#nudging.

Cariou, E., Cattiaux, J., Qasmi, S., Ribes, A., Cassou, C., and Doury, A.: Linking European Temperature Variations to Atmospheric Circulation With a Neural Network: A Pilot Study in a Climate Model, Geophysical Research Letters, 52, e2024GL113 540, https://doi.org/10.1029/2024GL113540, 2025.

Cattiaux, J., Vautard, R., Cassou, C., Yiou, P., Masson-Delmotte, V., and Codron, F.: Winter 2010 in Europe: A Cold Extreme in a Warming Climate: COLD WINTER 2010 IN EUROPE, Geophysical Research Letters, 37, https://doi.org/10.1029/2010GL044613, 2010.

Chemke, R. and Coumou, D.: Human Influence on the Recent Weakening of Storm Tracks in Boreal Summer, npj Climate and Atmospheric Science, 7, 1–8, https://doi.org/10.1038/s41612-024-00640-2, 2024.

Coumou, D., Lehmann, J., and Beckmann, J.: The Weakening Summer Circulation in the Northern Hemisphere Mid-Latitudes, Science, 348, 324–327, https://doi.org/10.1126/science.1261768, 2015.

Danabasoglu, G., Lamarque, J.-F., Bacmeister, J., Bailey, D. A., DuVivier, A. K., Edwards, J., Emmons, L. K., Fasullo, J., Garcia, R., Gettelman, A., Hannay, C., Holland, M. M., Large, W. G., Lauritzen, P. H., Lawrence, D. M., Lenaerts, J. T. M., Lindsay, K., Lipscomb, W. H., Mills, M. J., Neale, R., Oleson, K. W., Otto-Bliesner, B., Phillips, A. S., Sacks, W., Tilmes, S., van Kampenhout, L., Vertenstein, M., Bertini, A., Dennis, J., Deser, C., Fischer, C., Fox-Kemper, B., Kay, J. E., Kinnison, D., Kushner, P. J., Larson, V. E., Long, M. C., Mickelson, S., Moore, J. K., Nienhouse, E., Polvani, L., Rasch, P. J., and Strand, W. G.: The Community Earth System Model Version 2 (CESM2), Journal of Advances in Modeling Earth Systems, 12, e2019MS001 916, https://doi.org/10.1029/2019MS001916, 2020.

Deser, C., Terray, L., and Phillips, A. S.: Forced and Internal Components of Winter Air Temperature Trends over North America during the Past 50 Years: Mechanisms and Implications*, Journal of Climate, 29, 2237–2258, https://doi.org/10.1175/JCLI-D-15-0304.1, 2016.

Dong, B., Sutton, R. T., Shaffrey, L., and Harvey, B.: Recent Decadal Weakening of the Summer Eurasian Westerly Jet Attributable to Anthropogenic Aerosol Emissions, Nature Communications, 13, 1148, https://doi.org/10.1038/s41467-022-28816-5, 2022.

Durand, H., Varando, G., and Camps-Valls, G.: Learning Causal Response Representations through Direct Effect Analysis, arXiv preprint arXiv:2503.04358, 2025a.

Durand, H., Varando, G., and Camps-Valls, G.: Learning Causal Response Representations through Direct Effect Analysis, https://doi.org/10.48550/arXiv.2503.04358, 2025b.

Eyring, V., Gillett, N. P., Achuta Rao, K. M., Barimalala, R., Barreiro Parrillo, M., Bellouin, N., Cassou, C., Durack, P. J., Kosaka, Y., McGregor, S., Min, S., Morgenstern, O., and Sun, Y.: Human Influence on 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. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B., book section 3, pp. 463–466, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2021.

Fabiano, F., Meccia, V. L., Davini, P., Ghinassi, P., and Corti, S.: A Regime View of Future Atmospheric Circulation Changes in Northern Mid-Latitudes, Weather and Climate Dynamics, 2, 163–180, https://doi.org/10.5194/wcd-2-163-2021, 2021.

Fereday, D., Chadwick, R., Knight, J., and Scaife, A. A.: Atmospheric Dynamics Is the Largest Source of Uncertainty in Future Winter European Rainfall, Journal of Climate, 31, 963–977, https://doi.org/10.1175/JCLI-D-17-0048.1, 2018.

Guo, R., Deser, C., Terray, L., and Lehner, F.: Human Influence on Winter Precipitation Trends (1921–2015) over North America and Eurasia Revealed by Dynamical Adjustment, Geophysical Research Letters, 46, 3426–3434, https://doi.org/10.1029/2018GL081316, 2019.

Hanna, E., Fettweis, X., and Hall, R. J.: Brief Communication: Recent Changes in Summer Greenland Blocking Captured by None of the CMIP5 Models, The Cryosphere, 12, 3287–3292, https://doi.org/10.5194/tc-12-3287-2018, 2018.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 Global Reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Horton, D. E., Johnson, N. C., Singh, D., Swain, D. L., Rajaratnam, B., and Diffenbaugh, N. S.: Contribution of Changes in Atmospheric Circulation Patterns to Extreme Temperature Trends, Nature, 522, 465–469, https://doi.org/10.1038/nature14550, 2015.

Kornhuber, K., Bartusek, S., Seager, R., Schellnhuber, H. J., and Ting, M.: Global Emergence of Regional Heatwave Hotspots Outpaces Climate Model Simulations, Proceedings of the National Academy of Sciences, 121, e2411258 121, https://doi.org/10.1073/pnas.2411258121, 2024.

Kvålseth, T. O.: Cautionary Note about R2, The American Statistician, 39, 279–285, https://doi.org/10.2307/2683704, 1985.

Lehner, F., Deser, C., and Terray, L.: Toward a New Estimate of "Time of Emergence" of Anthropogenic Warming: Insights from Dynamical Adjustment and a Large Initial-Condition Model Ensemble, Journal of Climate, 30, 7739–7756, https://doi.org/10.1175/JCLI-D-16-0792.1, 2017.

Liné, A., Cassou, C., Msadek, R., and Parey, S.: Modulation of Northern Europe Near-Term Anthropogenic Warming and Wettening Assessed through Internal Variability Storylines, npj Climate and Atmospheric Science, 7, 1–14, https://doi.org/10.1038/s41612-024-00759-2, 2024.

Merrifield, A., Lehner, F., Xie, S.-P., and Deser, C.: Removing Circulation Effects to Assess Central U.S. Land-Atmosphere Interactions in the CESM Large Ensemble, Geophysical Research Letters, 44, 9938–9946, https://doi.org/10.1002/2017GL074831, 2017.

Merrifield, A., Lehner, F., Lorenz, R., and Knutti, R.: Exploring Extended Warm Periods in an Observational Large Ensemble of Historical European Temperature, in: AGU Fall Meeting Abstracts, vol. 2020, pp. GC058–0005, 2020.

Merrifield, A. L., Simpson, I. R., McKinnon, K. A., Sippel, S., Xie, S.-P., and Deser, C.: Local and Nonlocal Land Surface Influence in European Heatwave Initial Condition Ensembles, Geophysical Research Letters, 46, 14 082–14 092, https://doi.org/https://doi.org/10.1029/2019GL083945, 2019.

Pfleiderer, P., Schleussner, C.-f., Kornhuber, K., and Coumou, D.: Summer Weather Becomes More Persistent in a 2 °C World, Nature Climate Change, https://doi.org/10.1038/s41558-019-0555-0, 2019.

Qasmi, S. and Ribes, A.: Reducing Uncertainty in Local Temperature Projections, Science Advances, 8, eabo6872, https://doi.org/10.1126/sciadv.abo6872, 2022.

Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P., Kent, E. C., and Kaplan, A.: Global Analyses of Sea Surface Temperature, Sea Ice, and Night Marine Air Temperature since the Late Nineteenth Century, Journal of Geophysical Research: Atmospheres, 108, https://doi.org/10.1029/2002JD002670, 2003.

Rigal, A., Azaïs, J.-M., and Ribes, A.: Estimating Daily Climatological Normals in a Changing Climate, Climate Dynamics, 53, 275–286, https://doi.org/10.1007/s00382-018-4584-6, 2019.

Rodgers, K. B., Lee, S.-S., Rosenbloom, N., Timmermann, A., Danabasoglu, G., Deser, C., Edwards, J., Kim, J.-E., Simpson, I. R., Stein, K., Stuecker, M. F., Yamaguchi, R., Bódai, T., Chung, E.-S., Huang, L., Kim, W. M., Lamarque, J.-F., Lombardozzi, D. L., Wieder, W. R., and Yeager, S. G.: Ubiquity of Human-Induced Changes in Climate Variability, Earth System Dynamics, 12, 1393–1411, https://doi.org/10.5194/esd-12-1393-2021, 2021.

Ronneberger, O., Fischer, P., and Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation, 2015.

Röthlisberger, M. and Papritz, L.: Quantifying the Physical Processes Leading to Atmospheric Hot Extremes at a Global Scale, Nature Geoscience, 16, 210–216, https://doi.org/10.1038/s41561-023-01126-1, 2023.

Rousi, E., Kornhuber, K., Beobide-Arsuaga, G., Luo, F., and Coumou, D.: Accelerated Western European Heatwave Trends Linked to More-Persistent Double Jets over Eurasia, Nature Communications, 13, 3851, https://doi.org/10.1038/s41467-022-31432-y, 2022.

Saffioti, C., Fischer, E. M., and Knutti, R.: Improved Consistency of Climate Projections over Europe after Accounting for Atmospheric Circulation Variability, Journal of Climate, 30, 7271–7291, https://doi.org/10.1175/JCLI-D-16-0695.1, 2017.

Seneviratne, S. I., Lüthi, D., Litschi, M., and Schär, C.: Land–Atmosphere Coupling and Climate Change in Europe, Nature, 443, 205–209, https://doi.org/10.1038/nature05095, 2006.

Seneviratne, S. I., Zhang, X., Adnan, M., Badi, W., Dereczynski, C., Di Luca, A., Vicente-Serrano, S. M., Wehner, M., and Zhou, B.: Chapter 11: Weather and Climate Extreme Events in a Changing Climate. In Climate Change 2021: The Physical Science Basis., 2021.

Shaw, T. A. and Miyawaki, O.: Fast Upper-Level Jet Stream Winds Get Faster under Climate Change, Nature Climate Change, 14, 61–67, https://doi.org/10.1038/s41558-023-01884-1, 2024.

Shepherd, T. G.: Atmospheric Circulation as a Source of Uncertainty in Climate Change Projections, Nature Geoscience, 7, 703–708, https://doi.org/10.1038/ngeo2253, 2014.

Shepherd, T. G.: Storyline Approach to the Construction of Regional Climate Change Information, Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 475, https://doi.org/10.1098/rspa.2019.0013, 2019.

Singh, J., Sippel, S., and Fischer, E. M.: Circulation Dampened Heat Extremes Intensification over the Midwest USA and Amplified over Western Europe, Communications Earth & Environment, 4, 1–9, https://doi.org/10.1038/s43247-023-01096-7, 2023.

Singh, J., Sippel, S., Gu, L., Knutti, R., and Fischer, E.: Externally Forced Circulation Changes Amplify Mid-Latitude Regional Heat Extremes in Climate Model Nudged-Circulation Experiments, https://doi.org/10.21203/rs.3.rs-7299009/v2, 2025.

Sippel, S., Meinshausen, N., Merrifield, A., Lehner, F., Pendergrass, A. G., Fischer, E., and Knutti, R.: Uncovering the Forced Climate Response from a Single Ensemble Member Using Statistical Learning, Journal of Climate, 32, 5677–5699, https://doi.org/10.1175/JCLI-D-18-0882.1, 2019.

Sippel, S., Barnes, C., Cadiou, C., Fischer, E., Kew, S., Kretschmer, M., Philip, S., Shepherd, T. G., Singh, J., Vautard, R., and Yiou, P.: Could an Extremely Cold Central European Winter Such as 1963 Happen Again despite Climate Change?, Weather and Climate Dynamics, 5, 943–957, https://doi.org/10.5194/wcd-5-943-2024, 2024.

Smoliak, B. V., Wallace, J. M., Lin, P., and Fu, Q.: Dynamical Adjustment of the Northern Hemisphere Surface Air Temperature Field: Methodology and Application to Observations, Journal of Climate, 28, 1613–1629, https://doi.org/10.1175/JCLI-D-14-00111.1, 2015.

Teng, H., Leung, R., Branstator, G., Lu, J., and Ding, Q.: Warming Pattern over the Northern Hemisphere Midlatitudes in Boreal Summer 1979–2020, Journal of Climate, 35, 3479–3494, https://doi.org/10.1175/JCLI-D-21-0437.1, 2022.

Terray, L.: A Dynamical Adjustment Perspective on Extreme Event Attribution, Weather and Climate Dynamics, 2, 971–989, https://doi.org/10.5194/wcd-2-971-2021, 2021.

Topál, D. and Ding, Q.: Atmospheric Circulation-Constrained Model Sensitivity Recalibrates Arctic Climate Projections, Nature Climate Change, 13, 710–718, https://doi.org/10.1038/s41558-023-01698-1, 2023.

Van Den Dool, H. M.: Searching for Analogues, How Long Must We Wait?, Tellus A, 46, 314–324, https://doi.org/10.1034/j.1600-0870.1994.t01-2-00006.x, 1994.

Vautard, R., Cattiaux, J., Happé, T., Singh, J., Bonnet, R., Cassou, C., Coumou, D., D'Andrea, F., Faranda, D., Fischer, E., Ribes, A., Sippel, S., and Yiou, P.: Heat Extremes in Western Europe Increasing Faster than Simulated Due to Atmospheric Circulation Trends, Nature Communications, 14, 6803, https://doi.org/10.1038/s41467-023-42143-3, 2023.

Wehrli, K., Guillod, B. P., Hauser, M., Leclair, M., and Seneviratne, S. I.: Assessing the Dynamic Versus Thermodynamic Origin of Climate Model Biases, Geophysical Research Letters, 45, 8471–8479, https://doi.org/10.1029/2018GL079220, 2018.

Wills, R. C. J., Battisti, D. S., Armour, K. C., Schneider, T., and Deser, C.: Pattern Recognition Methods to Separate Forced Responses from Internal Variability in Climate Model Ensembles and Observations, Journal of Climate, 33, 8693–8719, https://doi.org/10.1175/JCLI-D-19-0855.1, 2020.

Woollings, T., Drouard, M., O'Reilly, C. H., Sexton, D. M. H., and McSweeney, C.: Trends in the Atmospheric Jet Streams Are Emerging in Observations and Could Be Linked to Tropical Warming, Communications Earth & Environment, 4, 1–8, https://doi.org/10.1038/s43247-023-00792-8, 2023.

Zeder, J. and Fischer, E. M.: Quantifying the Statistical Dependence of Mid-Latitude Heatwave Intensity and Likelihood on Prevalent Physical Drivers and Climate Change, Advances in Statistical Climatology, Meteorology and Oceanography, 9, 83–102, https://doi.org/10.5194/ascmo-9-83-2023, 2023.