



# Emerging climate signals in the Lena River catchment: a non-parametric statistical approach

Eric Pohl<sup>1,2</sup>, Christophe Grenier<sup>2</sup>, Mathieu Vrac<sup>2</sup>, and Masa Kageyama<sup>2</sup>

<sup>1</sup>Department of Geosciences, University of Fribourg, 1700 Fribourg, Switzerland

<sup>2</sup>Laboratoire des Sciences du Climat et de l'Environnement (LSCE/IPSL),  
UMR CEA-CNRS-UVSQ, 91120 Gif-sur-Yvette, France

**Correspondence:** Eric Pohl (eric.pohl@unifr.ch)

Received: 15 July 2019 – Discussion started: 10 September 2019

Revised: 20 March 2020 – Accepted: 7 April 2020 – Published: 29 May 2020

**Abstract.** Climate change has far-reaching implications in permafrost-underlain landscapes with respect to hydrology, ecosystems, and the population's traditional livelihoods. In the Lena River catchment, eastern Siberia, changing climatic conditions and the associated impacts are already observed or expected. However, as climate change progresses the question remains as to how far we are along this track and when these changes will constitute a significant emergence from natural variability. Here we present an approach to investigate temperature and precipitation time series from observational records, reanalysis, and an ensemble of 65 climate model simulations forced by the RCP8.5 emission scenario. We developed a novel non-parametric statistical method to identify the time of emergence (ToE) of climate change signals, i.e. the time when a climate signal permanently exceeds its natural variability. The method is based on the Hellinger distance metric that measures the similarity of probability density functions (PDFs) roughly corresponding to their geometrical overlap. Natural variability is estimated as a PDF for the earliest period common to all datasets used in the study (1901–1921) and is then compared to PDFs of target periods with moving windows of 21 years at annual and seasonal scales. The method yields dissimilarities or emergence levels ranging from 0 % to 100 % and the direction of change as a continuous time series itself. First, we showcase the method's advantage over the Kolmogorov–Smirnov metric using a synthetic dataset that resembles signals observed in the utilized climate models. Then, we focus on the Lena River catchment, where significant environmental changes are already apparent. On average, the emergence of temperature has a strong onset in the 1970s with a monotonic

increase thereafter for validated reanalysis data. At the end of the reanalysis dataset (2004), temperature distributions have emerged by 50 %–60 %. Climate model projections suggest the same evolution on average and 90 % emergence by 2040. For precipitation the analysis is less conclusive because of high uncertainties in existing reanalysis datasets that also impede an evaluation of the climate models. Model projections suggest hardly any emergence by 2000 but a strong emergence thereafter, reaching 60 % by the end of the investigated period (2089). The presented ToE method provides more versatility than traditional parametric approaches and allows for a detailed temporal analysis of climate signal evolutions. An original strategy to select the most realistic model simulations based on the available observational data significantly reduces the uncertainties resulting from the spread in the 65 climate models used. The method comes as a toolbox available at [https://github.com/pohleric/toe\\_tools](https://github.com/pohleric/toe_tools) (last access: 19 May 2020).

## 1 Introduction

High latitudes experienced pronounced climate change, for example, in the form of warming air temperatures and precipitation regime shifts (Cohen et al., 2018). This manifests in far-reaching impacts on the livelihoods of permafrost communities (Crate et al., 2017), hydrological systems (Gautier et al., 2018; Karlsson et al., 2012; Prowse et al., 2010; Vey et al., 2013; Walvoord and Striegl, 2007; Yang et al., 2002), the evolution of permafrost, including changes in landforms (Boike et al., 2016), and feedbacks with the global carbon cy-

cle (Beermann et al., 2017; Hope and Schaefer, 2016; Schuur et al., 2015). The Lena River catchment in eastern Siberia is one of the largest watersheds in Siberia and provides a major contribution to the Arctic Ocean. It is a perfect example of a permafrost landscape that is prone to and highly sensitive to the impacts of climate change. Available air temperature and precipitation records in this region extend back more than 100 years and provide a database to investigate local trends and variability in climate in more detail.

Despite a general warming trend, a strong spatial and temporal variability is apparent over northeastern Eurasia (Desyatkin et al., 2015; Fedorov et al., 2014b; Gorokhov and Fedorov, 2018) and in the high latitudes in general (Mahlstein et al., 2011). A few locations show no apparent trend over the available long-term records (Fedorov et al., 2014b). Gorokhov and Fedorov (2018) focus on the region of Yakutia (Sakha Republic) and find positive temperature and precipitation trends for the region as a whole for the period 1966–2016. However, spatial and temporal variability is apparent in the form of a stronger warming trend in winter compared to summer (0.4 to 1 °C per decade compared to 0.1 to 0.4 °C per decade) and a negative precipitation trend in the northern region (−8 mm per decade) in contrast to increasing positive trends towards the south (~ 16 mm per decade). In addition, air and ground temperatures co-evolve with strong spatial heterogeneity (Fedorov et al., 2014b; Romanovsky et al., 2010), potentially associated with changes in regional precipitation and snow cover dynamics (Romanovsky et al., 2010, and references therein).

Such changes propel landscape transitions that are not necessarily linear. For instance, the interactions between meteorological forcing and the ground thermal regime in the permafrost-underlain region are complex due to thermal effects, including phase change in the freeze–thaw cycles and insulation effects of snow cover (Walvoord and Kurylyk, 2016). The impacted hydrological cycle already shows a systematic shift towards an increase in the intensity and duration of floods, higher frequency of large floods, and disappearing small floods (Gautier et al., 2018). More changes in the hydrological regime can be expected in the future through geomorphological changes, especially the formation of thermokarst lakes (Fedorov et al., 2014a; Ulrich et al., 2017). Most thermokarst lakes are initiated endorheic but might aggregate and connect to the river network with increasing permafrost thaw.

However, the spatiotemporal variability and heterogeneous evolution of different climate variables raise questions about the regional magnitude of climate change and how much of the observed variability can be attributed to natural climate variability or to human activities. Additionally, it renders an overarching assessment of how permafrost will evolve under climate change and what this means for the climate system as a whole difficult. The individual analysis of the key variables temperature and precipitation constitutes a first step to approach this problem. The identification of how

these variables have individually evolved with respect to their natural variability give insights into the complex and direct and indirect interactions in the Earth system. Ultimately, this is needed for a comprehensive understanding of the system and an assessment of resulting implications under continuing climate change. It further constitutes a prerequisite for the planning and execution of possible adaptation and mitigation actions that are needed to cope with the environmental and socio-economic impacts in a timely manner.

As a result, considerable effort has been put in the development of methods to investigate and identify when climate departs or emerges from its natural state or variability (time of emergence – ToE). ToE studies cover a wide spectrum of applications, from the most common climate variables like near surface air temperature and precipitation (Giorgi and Bi, 2009; King et al., 2015; Lehner et al., 2017; Mora et al., 2013) to climate extremes (King et al., 2015; Maraun, 2013; Scherer and Diffenbaugh, 2014) and sea level rise (Lyu et al., 2014). There are several methods to calculate the ToE (e.g. Sui et al., 2014, and references therein), depending on the available data sources and the specific purpose of the study. Two major aspects are at the front line of research. The first concerns the methodology, and the second concerns the database on which to perform the analysis.

### 1.1 ToE approaches

The ToE is defined as the timing when a climate signal, such as temperature or precipitation, permanently exceeds its natural variability (e.g. Giorgi and Bi, 2009; Hawkins and Sutton, 2012). Several existing methods rely on separating signal  $S$  (climate change) and noise  $N$  (natural variability). Such approaches may require a high level of parameterization (Lehner et al., 2017; Sui et al., 2014), for example, to define natural variability, to define a threshold for the  $S/N$  ratio, or to separate signal from noise. Additionally, some meta-parameters are needed, such as the size of moving windows or the selection of the period that is considered as reference time (e.g. preindustrial conditions). The variability of a variable within a reference period can be addressed by means of standard deviation (e.g. Hawkins and Sutton, 2012; Lehner et al., 2017; Mahlstein et al., 2011) or by the total observed range in values (e.g. Mora et al., 2013). Signals tested for emergence are somehow filtered to eliminate decadal and lower frequency variability, e.g. by means of moving averages (e.g. Lehner et al., 2017) or polynomial fitting (e.g. Hawkins and Sutton, 2012), and are then compared to the derived reference period variability. Other approaches are based on statistical tests that compare, for example, the distributions between a reference and a target period (King et al., 2015; Mahlstein et al., 2011, 2012). Mahlstein et al. (2012) and King et al. (2015), for example, used the Kolmogorov–Smirnov test (KS test) with a defined significance level to test the statistical similarity between reference and target period distributions. The KS test is based on

a continuous distance metric, i.e. the maximum difference between two cumulative density functions, but it has so far always been used in combination with a significance level.

All existing ToE methods are by definition a test, either on the exceedance of an  $S/N$  ratio threshold or based on a statistical significance level. As such, they require a parameterization, which can be a drawback in terms of objectivity and transferability. For instance, dealing with a set of different climate variables may lead to different distribution models, where different dataset record lengths affect the behaviour of statistical tests and filtering operations. The development of a non-parametric approach is appealing because results are not impeded by the choice of parameters as in the case of parametric approaches.

### 1.2 Data basis for ToE studies

The second major aspect of ToE research concerns the data basis. Observational datasets facilitate ToE studies that focus on changes which have already occurred. Direct observational data are the most accurate estimates but come with the downside of data gaps and limited spatial coverage. Reanalysis datasets assimilating observational data provide extended spatiotemporal coverage. Their continuous spatiotemporal coverage is an advantage over meteorological station data, but this comes at the cost of some biases with respect to the real observations (Khan et al., 2008; Serreze and Hurst, 2000). Possible ToE methods for these data types rely on a statistical analysis of their signal's evolution over time. In some cases, including the present study, continuous time series are compulsory, excluding data from meteorological stations with interrupted observations.

Ensembles of climate model simulations (CSs) provide estimates ranging from the past to the future and come with specific data structures. These structures are, in some cases, needed to address the effects of internal climate variability (Hawkins and Sutton, 2012; Lehner et al., 2017; Mora et al., 2013) or allow the utilization of preindustrial control runs, i.e. a forcing corresponding to preindustrial conditions (e.g. Karoly and Wu, 2005). The difference between model runs with different anthropogenic forcing scenarios and the control runs can provide an estimate for the effect of anthropogenic forcing on the climate (e.g. King et al., 2015; Knutson et al., 2013; Lyu et al., 2014). However, sometimes large CS ensemble spreads (e.g. Knutson et al., 2013; Koven et al., 2013) introduce considerable uncertainties in ToE estimates (Deser et al., 2012; Hawkins et al., 2014; King et al., 2015). In order to reduce the model spread, a preselection of CSs can be made based on a comparison between CSs and observations, e.g. by comparing the variability of certain variables from observations to those in the CS in the region of interest (e.g. Mahlstein et al., 2011). Alternatively, weights can be given to individual CSs based on how similar a model's internal structure is and how well they represent observational data (Knutti et al., 2017). Identifying a robust and objective

function for the selection of a CS to reduce uncertainty is, however, difficult, as it depends on available observational data and means to assess model similarity (e.g. Knutti et al., 2017; Leloup et al., 2008).

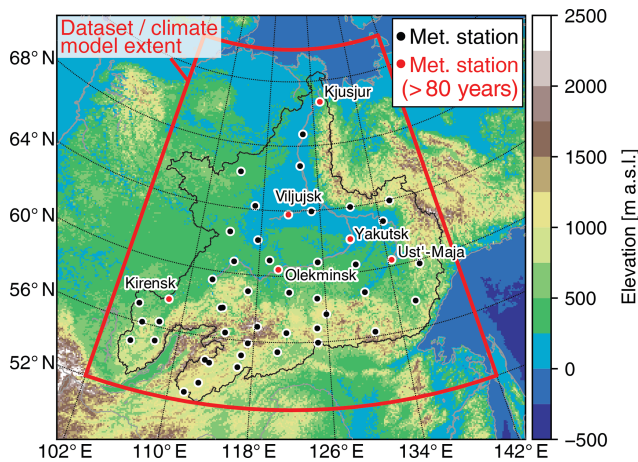
### 1.3 Aims

This study presents a novel ToE approach, allowing for the investigation of the actual evolution of emergence over time. This differs from other methods in the form of tests that provide either the indication of emergence or not. The approach is applied to near-surface air temperature ( $T$ ) and precipitation ( $P$ ) in the Lena River catchment, where changes in landscape (Crate et al., 2017) and hydrological behaviour (Gautier et al., 2018; Yang et al., 2002) are already apparent, and accounts for the variables' importance in the hydrological cycle and impact on permafrost evolution. The study is designed to utilize available observational data from meteorological stations, reanalysis data, and an ensemble of CSs from the Coupled Model Intercomparison Project phase 5 (CMIP5; Taylor et al., 2012). This multi-step, multi-source approach allows for comparison between obtained estimates from the most reliable (in situ) to the most uncertain (CS) data sources.

We test how such an approach can reduce uncertainty of ToE estimates by introducing a non-parametric method based on an adapted Hellinger distance metric (Hellinger, 1909). The method does not constitute a test, but it is a continuous metric that describes how far a climate signal, in the form of a time series, has emerged from its natural variability.

This approach is intentionally non-parametric by design in order to ensure transferability to other scientific fields and to other variables that inherit any kind of value distribution. Because the metric is derived as a continuous signal, it gives insights into how climate signals emerge from natural variability over time. This provides potentially added value to the general question of whether a signal has emerged or not based on a single test. Another strength of this approach is that it facilitates an in-depth analysis of how climate change emerges over time, and, in the process, it allows for selecting CSs that show an emergence consistent with real observations. Consequently, it allows for selecting the most realistic CS.

The succeeding sections present the method in detail, followed by the data sources and obtained results. We discuss the obtained results in the light of previous studies, as well as the unavoidable choices of meta-parameters in detail. The latter comprise the selection of a reference period, which is usually preindustrial conditions like 1881–1910 in Vautard et al. (2014) or 1860–1910 in King et al. (2015), to identify anthropogenic climate change and the window width to filter out natural and decadal climate variability of the climate signal. Finally, we present our conclusions on how the presented method provides a versatile tool for ToE studies and how it



**Figure 1.** Lena River catchment (black outline) on topographic map (colour coded) and position of short- and long-term meteorological stations used to test reanalysis and interpolated datasets. Of the long-term stations, Kjusjur has the lowest temporal coverage (less than 10 years) in the reference period 1901–1921.

can reduce uncertainty by the incorporation of observational and reanalysis datasets.

## 2 Methods

In the following section we present the methods for ToE detection, sensitivity analysis, and data selection. Our ToE method is a non-parametric metric and thus differs from previous approaches that are parameterized tests for emergence. Our metric describes emergence by measuring how data distributions in continuous target periods have changed with respect to a reference period. Like other approaches, it requires meta-parameter choices, like the start and end point of a reference period and window widths for target periods, for which we will present a sensitivity approach. Finally, the availability of actual long-term observations in the Lena River catchment (Fig. 1) allows for validating reanalysis and climate model simulation datasets for their potential to represent the same climate change evolution.

### 2.1 Time of emergence (ToE)

Our ToE method is based on a similarity metric between probability density functions (PDFs) described by Hellinger (1909). This metric belongs to a family of distance metrics (Cha, 2007) and can be roughly understood as the geometrical overlap of two PDFs (Fig. 2; Rust et al., 2010). The method has been used e.g. by Rust et al. (2010) to showcase similarities between distributions of circulation patterns obtained through different climate models.

As we want to describe the dissimilarity, i.e. how far a distribution has emerged from a reference one, we adjust the

writing and refer to the metric as the Hellinger distance (HD) according to

$$\text{HD}(QR) = \sqrt{1 - \int [Q(x)^{\frac{1}{2}} R(x)^{\frac{1}{2}}] dx}, \quad (1)$$

where  $Q(x)$  and  $R(x)$  are the PDFs of the target ( $Q$ ) and reference ( $R$ ) period, respectively. We use a Gaussian kernel density estimator (KDE) to derive the PDFs from the samples of  $Q$  and  $R$  and finally calculate the numerical approximation according to

$$\text{HD}(QR) = \sqrt{\frac{1}{2} \sum_{i=1}^d (\sqrt{Q_i} - \sqrt{R_i})^2}, \quad (2)$$

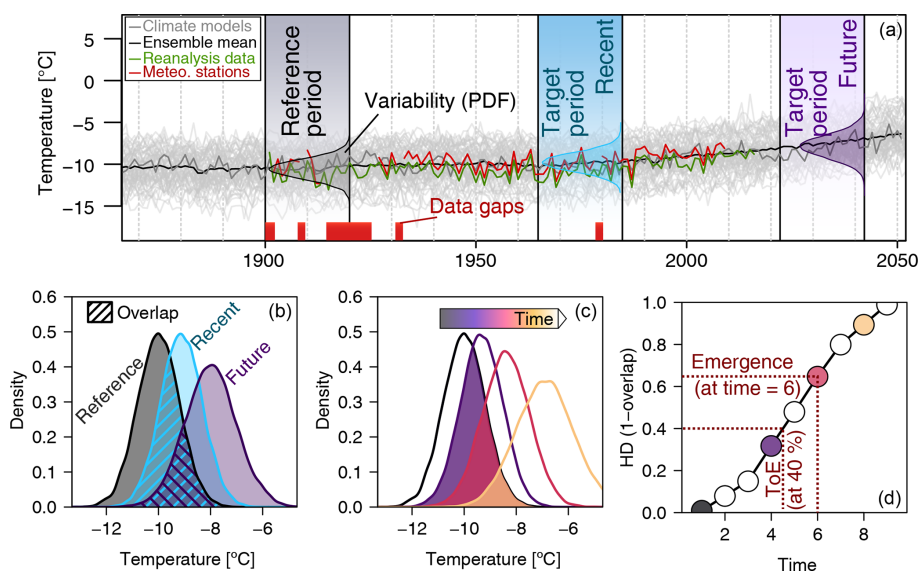
where  $Q_i$  and  $R_i$  are the densities of the PDFs at position  $i$  along a value range that corresponds to the minimum and maximum of the full time series of a variable, extended by the difference between these extremes in both directions. We use  $d = 200$  steps, equally incremented. Tests with more steps and further-extended minimum and maximum bounds resulted in insignificant changes (not shown).

The KDE introduces two meta-parameters for the shape of the kernel and the bandwidth (e.g. Scott, 2015). While the kernel shape normally has little impact, the bandwidth can have a strong impact on the obtained KDE–PDF value (Turlach, 1993; Scott, 2015). In contrast, the distance metric in the KS test does not require a PDF estimate. Therefore, we will show the performance against the KS test metric and the sensitivity of our approach with respect to the use of bandwidth selection in Sect. 4. In our approach we use an automatic bandwidth estimation that is based on Scott's factor (Scott, 2015), which is only dependent on the number of data points and dimensionality of the data. Therefore, the bandwidth stays fixed in this work for the calculation of the obtained ToE results, as we use fixed window sizes.

The HD can take values ranging between 0 (equal distributions and full overlap) and 1 (fully emerged distributions with no overlap). The outline of the method is presented in the schematic (Fig. 2). A climate signal will show a specific data distribution at each time step within a given time window for which a PDF is calculated (Fig. 2). The HD will increase both if a PDF with a same shape is shifted to higher or lower values and if its shape changes. The HD is calculated for each time step after the reference period stops. This results in a continuous time series of HD values or the level of emergence. This time series serves three purposes: (1) a level of emergence can be derived for any given time step, (2) the ToE can then be inferred based on a posteriori applied thresholds, and (3) different competing datasets can be tested for consistency based on their HD evolution (Fig. 2).

We additionally calculate the sign of change because emergence could also occur towards lower values (e.g. less precipitation). The sign (positive or negative) of change is





**Figure 2.** Schematic of the ToE method based on the Hellinger distance (HD). **(a)** Example of time series evolution from climate simulations, a meteorological station, and a reanalysis dataset. Natural variability as the PDF of the reference period and two example target periods with a window width of 21 years for recent and future assessment with a sketch of the corresponding PDFs. **(b–d)** The overlap between the PDFs of the reference and target periods **(b)**, a sketch of PDF evolution over time **(c)**, and the resulting HD as the dissimilarity of the target and subsequent reference PDFs (one minus the overlap). Exemplary determination of the ToE for a threshold of 40 % emergence or emergence at a chosen time step, respectively **(d)**.

calculated from the summed differences of the PDF densities ( $\Delta RQ$ ):

$$\Delta RQ = \sum_{i=1}^d (R_i - Q_i) \cdot bc_i, \quad (3)$$

where  $bc_i$  values are the actual values at the position  $i$  along the extended value range used in Eq. (2). The sign is 1 for  $\Delta RQ > 0$ ,  $-1$  for  $\Delta RQ < 0$ , and 0 for  $\Delta RQ = 0$ . We set the reference period to 1901–1921 and take values for the target periods in moving windows of 21 years too. We test different reference periods and number of years in a sensitivity analysis (see next section). The reference period contains the earliest 21 years commonly available for all datasets. The target periods are taken as a two-sided moving window around each year after the reference period stops, providing a distribution for each time step thereafter. The ToE method is applied independently to the reanalysis data and each individual CMIP5 CS. We follow previous studies by running our analysis not only at annual scale but also on the seasonal scale (winter – November to March – and summer – May to September) to highlight seasonal differences. The ToE values obtained are given as the year in the middle of the moving window (e.g. a ToE in 2000 corresponds to the target period 1990 to 2010 for a window width of 21 years). We finally test different reference periods and lengths of target periods in a sensitivity analysis (see next section).

## 2.2 Sensitivity analysis

Our method is non-parametric for the climate change detection, but like other methods it requires a set of meta-parameters. These can be divided into two groups. The first group concerns the choice of reference period and the time window for the PDF computation. This is an important issue because climate variability in the high latitudes is particularly strong (Mahlstein et al., 2011). Thus, it makes sense to test the influence of choosing different reference periods and window widths on the outcome of the ToE (Hawkins and Sutton, 2016). We test reference periods ending between 1915 and 1929 and different window widths ranging between 15 and 29 years. While choosing an earlier starting date makes the reference period more preindustrial, it also removes the ability to sample multi-decadal and internal variability. The final choice is consequently a compromise between the two. Similarly, the choice of longer window widths to choose data distributions limits the ability to detect the ToE at the end of the time series. We will present all tested combinations and discuss the derived first-order approximations of uncertainty related to this unavoidable selection of meta-parameters in Sect. 6.2.

The second group concerns the obtained PDFs using a KDE. Two meta-parameters are used for the KDE, namely the kernel type (e.g. Gaussian, triangular, etc.) and the bandwidth, which determines the smoothness of the resulting PDF. As mentioned before, the type of kernel usually has a low impact on the resulting PDF, whereas the bandwidth

can have a strong impact. We dedicate Sect. 4 to this analysis by generating synthetic data with exactly controlled intensity changes and the onset of change to test our approach against the distance metric used in the KS test.

### 2.3 Dataset selection

In order to obtain the most reliable estimates for the ToE, the best data choice would be measurements from long-term operating meteorological stations in the Lena River catchment. However, data gaps and a poor spatial coverage demand for alternative data sources to provide a spatially and temporally comprehensive analysis. We thus test three commonly used state-of-the-art reanalysis datasets for their actual representation of in situ temperature and precipitation records. In order to investigate the evolution of climate over the 21st century, we include a collection of CMIP5 climate simulations (Taylor et al., 2012) and test their performance by means of HD evolution (in the past) with respect to the reanalysis data.

The reanalysis datasets are tested against the records from the meteorological stations for near-surface air temperature ( $T$ ) and precipitation ( $P$ ) using ordinary least-square regression analysis. For each of the 49 stations in the Lena River catchment (Fig. 1), the corresponding pixel-based time series of either reanalysis dataset is extracted, and the performance in terms of explained variance ( $r^2$ ) is evaluated. The best-performing dataset is used in the subsequent steps.

For the analysis of the ToE in the future, we use both the whole set ( $n = 65$ ) of model simulations but also a subset ( $n = 10$ ). The subset is used to test whether it reduces uncertainty for ToE estimates compared to the use of the entire ensemble. The subset is chosen based on a comparison between the HD of reanalysis and climate model simulations. By comparing the HD evolution (0%–100%) instead of the actual values, we avoid possible bias issues in temperature and precipitation estimates within the CS. We use the Nash–Sutcliffe efficiency (Nash and Sutcliffe, 1970; Moriasi et al., 2007) as objective function for the selection. In contrast to the  $r^2$ , the Nash–Sutcliffe efficiency (NSE) adds a penalty for offsets between HD evolutions, according to

$$\text{NSE} = 1 - \frac{\sum_{i=1}^n (Y_i^{\text{R}} - Y_i^{\text{CS}})^2}{\sum_{i=1}^n (Y_i^{\text{R}} - Y^{\text{Rmean}})^2}, \quad (4)$$

where  $Y_i^{\text{R}}$  is the  $i$ th HD value of the used reanalysis dataset,  $Y_i^{\text{CS}}$  is the  $i$ th HD value of a climate model simulation, and  $Y^{\text{Rmean}}$  is the mean of the HD of the reanalysis dataset (Moriasi et al., 2007).

As we will show in the results, we had to question the validity of reanalysis data in some cases. To ensure confidence in the data, we made a further refinement by choosing five pixels within the Lena River catchment domain where meteorological stations provide long-term observations and allowed us to verify the quality of the reanalysis. Data records

for these five stations reach back into the reference period 1901–1921 and cover at least 10 years (see Fig. S1 in the Supplement). The corresponding five pixels were used to calculate the HD both for the reanalysis and each of the CSs. For the sake of completeness, however, we will present the HD evolution of the reanalysis data for the whole study area alongside.

## 3 Data

We focus on the two climate variables  $P$  and  $T$  for their importance in the hydrological cycle and for permafrost evolution and for their relatively good data availability.

### 3.1 Observational data

For observational data we use the All-Russia Research Institute of Hydrometeorological Information – World Data Centre (RIHMI-WDC; <http://meteo.ru/>, last access: 19 May 2020) dataset, compiled by Bulygina and Razuvaev (2012). The dataset comprises 49 stations within the catchment area of the Lena River (Fig. 1). Data were obtained as daily values and averaged and summed to monthly values of  $T$  and  $P$ , respectively. The longest records are available for site Yakutsk starting in 1834. All stations within the dataset have record gaps. The dataset provides data only for the locations of the meteorological stations.

### 3.2 Reanalysis data

#### 3.2.1 CRUNCEP v7

CRUNCEP v7 is a global forcing product (ds314.3; Viovy, 2018) used, for example, in the ORCHIDEE-MICT (Organizing Carbon and Hydrology in Dynamic Ecosystems: ameliorated Interactions between Carbon and Temperature) land surface model (Guimberteau et al., 2018). The dataset is derived through a combination of the annually updated CRU TS (Climatic Research Unit Time-Series) v3.24 monthly climate dataset (New et al., 2000) and NCEP (National Centers for Environmental Prediction) reanalysis (Kalnay et al., 1996). The CRU TS datasets are based on surface climate data anomalies from different quality-controlled datasets. They are combined with monthly climatologies and interpolated to provide full spatial and temporal coverage. The time coverage is from 1901 to 2016 in 6-hourly temporal and  $0.5^\circ$  spatial resolution. The data were resampled to monthly averages (sums) of 2 m air temperature (precipitation) and to a spatial resolution of  $2 \times 2^\circ$  to match other obtained datasets.

### 3.2.2 20th Century Reanalysis (V2c) (20CR)

The 20CR: Monthly Mean Single Level (Analyses and Forecasts) dataset (ds131.2; Compo et al., 2011; [http://www.esrl.noaa.gov/psd/data/gridded/data.20thC\\_ReanV2c.html](http://www.esrl.noaa.gov/psd/data/gridded/data.20thC_ReanV2c.html), last access: 19 May 2020) contains objectively analysed 4-dimensional weather maps and their uncertainty from the mid-19th century to 21st century. The dataset has a temporal coverage from 1851 to 2011 with a monthly temporal and a  $2 \times 2^\circ$  spatial resolution.

### 3.2.3 ERA-20C reanalysis (ERA20)

ERA-20C is a reanalysis product (ds626.0; ECMWF, 2014) of the European Centre for Medium Range Weather Forecast (ECMWF) of the 20th century, from 1900 to 2011. It assimilates observations of surface pressure and surface marine winds only. A coupled atmosphere–land–surface–ocean–wave model is used to reanalyse the weather, by assimilating surface observations. Data in monthly temporal resolution (monthly means of daily means) in  $2 \times 2^\circ$  spatial resolution were obtained.

## 3.3 Climate model data

We use a set of global climate scenarios from the Coupled Model Intercomparison Project phase 5 (CMIP5; Taylor et al., 2012), obtained through the R package “esd” (Benestad et al., 2015). The model predictions are bias-corrected through an empirical downscaling approach described in Benestad (2001). All models have historical natural and anthropogenic forcing and land use for the period 1861–2005 and the concentration pathway 8.5 (RCP8.5) thereafter until 2100. An overview of these model simulations is given in Table S1 in the Supplement.

## 4 Performance of the HD-based ToE

### 4.1 Comparison of the HD to the Kolmogorov–Smirnov distance metric

The most striking difference between the HD-based ToE approach and previous ones is the continuous character of the obtained metric. However, the KS test also utilizes a continuous metric, namely the maximum distance between the cumulative density functions (CDFs) of two samples, in the following referred to as the KS metric. In order to evaluate the additional value of the HD-based approach, we showcase how these two distance metrics compare to each other in terms of sensitivity and accuracy (Figs. 3 and 4). For this, we generated synthetic data with a controlled onset and strength in signal changes. We first use two datasets, one closely resembling a temperature time series of the utilized climate model datasets (type 1) and one that serves to showcase detection sensitivity (type 2). The type 1 and type 2 data are

normally distributed data with a fixed mean and standard deviation (SD) until the breakpoint year (1960). Thereafter, type 1 data have a fixed linear change (slope derived from an arbitrarily chosen pixel of one of the climate model simulations), while the SD stays constant. The type 2 data have a constant mean value after the breakpoint year but a continuous increase in standard deviation, reaching 2 times the reference SD at the end of the time series. We generate 5000 time series of each type of dataset and calculate the HD and KS metric.

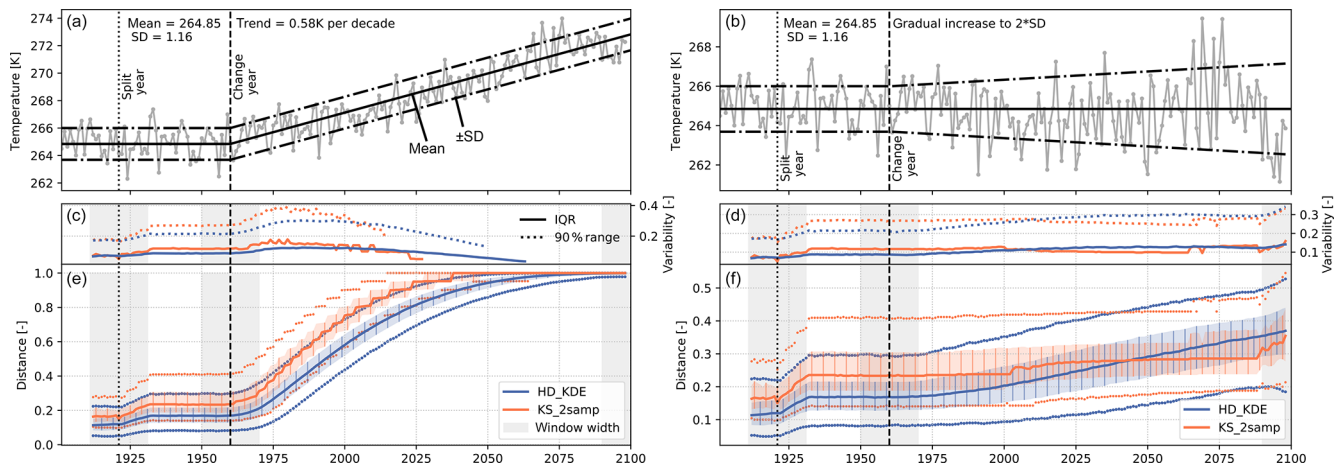
Figure 3 showcases one representation of each of the synthetic time series (panels a and b). The distance plots (panels c–f) show the median (bold line), interquartile range (shading), and the 5th–95th percentiles (points) to give a representative assessment of how the two distance metrics perform.

Generally speaking, the HD has a crucial advantage over the KS metric in terms of continuous change description and also in terms of accuracy. The left panel in Fig. 3 shows the co-evolution of the HD and KS metric for a time series with a pronounced trend. A step-function-like evolution is visible for the KS metric. This becomes even clearer if the change in the original signal is smaller (Fig. 3 – right). The interquartile range (IQR) of the HD based on 5000 samples of the time series is mostly lower than for the KS metric. Also, the 90 % range (5th–95th percentile) reaches overall lower values compared to the KS metric, as well as less variability along the time axis (the KS metric changes from a low to a high range within a few years).

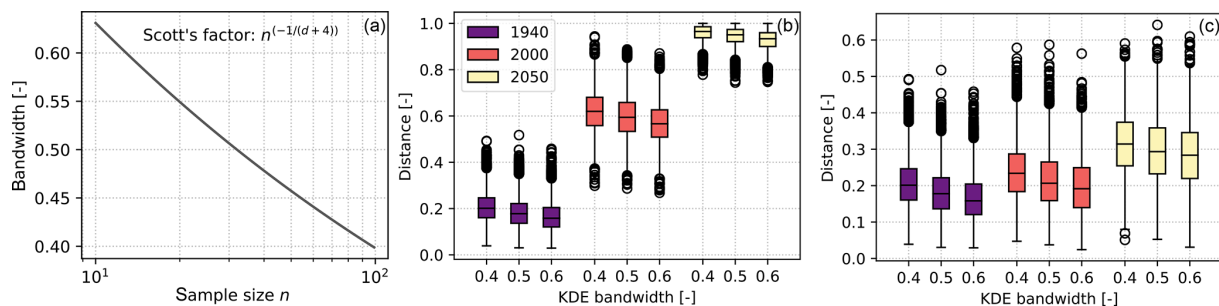
The right panels in Fig. 3 show a signal with slight changes (gradual increase of the standard deviation) and corresponding distances. The KS metric is not able to detect the change in a continuous way and only indicates change once a certain threshold is passed. The accuracy, i.e. the range in distance estimates based on the 5000 samples, is very similar in this case. The step-function-like evolution in the KS metric depends on the sample size, which determines the minimum dissimilarity increase ( $1/n$ , with  $n$  being the sample size). This step-function-like evolution is also clearly visible in the example with the strong onset of a trend (Fig. 3a, c, and e). Not shown are the minimum and maximum values that are possible. For the KS metric, these have a wider range because even a very slight shift of an otherwise equal distribution can cause a high KS metric. However, this does not happen often (not captured by the 90 % range).

### 4.2 KDE bandwidth sensitivity

To obtain the reference and target period PDFs, the utilized KDE is using Scott’s factor (Scott, 2015) for the automatic bandwidth selection (Fig. 4a). Even though it stays constant in our analysis with fixed window width and dimensionality, we test the possible impact on obtained HD estimates. For assumed common sample sizes for monthly to annual data between 10 and 100, Scott’s factor provides bandwidths between 0.4 and 0.6 (Fig. 4a). The resulting change in the HD



**Figure 3.** Comparison between the Hellinger distance using the KDE (HD\_KDE) and the Kolmogorov–Smirnov metric for the two-sample test (KS\_2samp). Two synthetic time series examples (a, b) and corresponding distance evolution (e, f). Interquartile range (IQR) and 5th–95th percentile range (c, d). Note the different scales in distances between the two types of data.



**Figure 4.** HD sensitivity to KDE bandwidth. (a) Automatic bandwidth selection in Python’s “scipy.stats” KDE (Virtanen et al., 2020) is based on Scott’s factor (Scott, 2015), where  $n$  is the sample size and  $d$  is the dimension. For 1-dimensional data, sample sizes between 10 and 100 correspond to bandwidths of 0.65 to 0.4. (b) Sensitivity in the HD for example type 1 (Fig. 3a, c, and e). (c) Sensitivity in the HD for example type 2 (Fig. 3b, d and f) at years 1940, 2000, and 2050.

is shown in Fig. 4 (panels b and c). We use again the two synthetic time series from before to show the change in the HD. For the relatively large range in sample sizes and resulting change in bandwidths, the overall change is in the range of only 5 % for both type 1 and type 2 data. A further analysis of how the bandwidth affects different types of signals is out of scope for this work. We do not explore the effects of different kernel shapes in addition to the bandwidth because of an inferior importance compared to the bandwidth (Bianchi, 1995).

However, we also tested the impact of strictly positive and strongly skewed distributions on the approach (Fig. S2). For small differences between such distributions, there is a positive bias resulting in an HD of at least 0.2 to 0.3. Once the HD reaches 0.3 and above, the bias to the actual HD of distributions becomes less than 10 % and becomes independent of the bandwidth.

## 5 Results

### 5.1 Dataset selection

The comparison of in situ data with CRUNCEP, 20CR, and ERA20 data shows differences in the reanalysis datasets’ performances (Fig. 5).  $T$  estimates of the datasets explain more than 95 % of the variance, but only CRUNCEP’s  $P$  estimates show a high correlation ( $r^2 = 0.85$ ) and limited bias with the observational data. Apart from the poor representation of the other datasets for  $P$ , 20CR also shows a systematic  $T$  underestimation (overestimation) in spring and autumn (summer and winter; Fig. 5). CRUNCEP provides the best estimates from the tested datasets for both target variables and is used in the following. As CRUNCEP results partially from direct station measurements, the best match is not surprising, even though we did not test which stations and which periods of the station data are incorporated or rejected. However, some artificial precipitation signals are apparent in the CRUNCEP dataset. These occur mainly in the northwestern part, where

**Table 1.** Nash–Sutcliffe efficiency statistics of the 10 best climate simulations with respect to the CRUNCEP data for each pixel encompassing a meteorological station with records of more than 10 years in the 1901–1921 reference period. Positive NSE in bold.

Station	Kirensk	Olyokminsk	Ust-Maya	Vilyuysk	Yakutsk
<i>Temperature</i>					
Annual					
NSE_mean	<b>0.58</b>	<b>0.58</b>	<b>0.61</b>	<b>0.43</b>	<b>0.62</b>
NSE_max	<b>0.79</b>	<b>0.81</b>	<b>0.74</b>	<b>0.74</b>	<b>0.73</b>
NSE_min	<b>0.44</b>	<b>0.43</b>	<b>0.53</b>	<b>0.26</b>	<b>0.53</b>
Summer					
NSE_mean	<b>0.27</b>	<b>0.35</b>	−0.02	<b>0.29</b>	<b>0.07</b>
NSE_max	<b>0.55</b>	<b>0.49</b>	<b>0.22</b>	<b>0.70</b>	<b>0.30</b>
NSE_min	<b>0.13</b>	<b>0.14</b>	−0.13	<b>0.01</b>	−0.08
Winter					
NSE_mean	<b>0.21</b>	<b>0.57</b>	<b>0.35</b>	−0.04	<b>0.38</b>
NSE_max	<b>0.46</b>	<b>0.71</b>	<b>0.62</b>	<b>0.44</b>	<b>0.59</b>
NSE_min	−0.08	<b>0.47</b>	<b>0.19</b>	−0.68	<b>0.29</b>
<i>Precipitation</i>					
Annual					
NSE_mean	<b>0.10</b>	−0.68	−1.32	−1.09	−0.89
NSE_max	<b>0.46</b>	<b>0.13</b>	−0.48	0.00	−0.19
NSE_min	−0.12	−1.18	−1.80	−1.94	−1.58
Summer					
NSE_mean	−0.20	−0.91	−0.50	−1.86	−1.18
NSE_max	<b>0.36</b>	<b>0.22</b>	<b>0.39</b>	−0.25	<b>0.24</b>
NSE_min	−0.56	−1.66	−0.95	−3.37	−2.40
Winter					
NSE_mean	−19.49	<b>0.07</b>	−16.88	−0.07	<b>0.24</b>
NSE_max	−1.36	<b>0.40</b>	−9.10	<b>0.20</b>	<b>0.54</b>
NSE_min	−31.83	−0.12	−20.60	−0.19	<b>0.08</b>

no stations with data records in the reference period exist (Figs. 1 and S1). For this region, the CRUNCEP  $P$  data show a strong artificial, annual repetitive pattern, with probable recycling of the same year, resulting in a very low inter-annual variability (Fig. S3). Here, the HD rapidly emerges to more than 40 % (Figs. 6 and 7 and Video 2). While the CRUNCEP  $T$  signals do not show a similar pattern that would be easy to identify, the inter-annual variability is also lower in the northeastern part compared to the rest of the study area (Fig. S3). Whether this implies an area-extensive bias in the CRUNCEP dataset for  $T$  is difficult to assess. The resulting differences in the HD for CRUNCEP based on the full dataset vs. the reduced dataset (pixels with validated long-term observations) are displayed side by side in Figs. 6 and 7 as solid and dashed green lines, respectively. The identified 10 best-performing model simulations based on either dataset are shown in Figs. S4 and S5, and the obtained NSE statistics derived from this analysis are shown in Table 1. The resulting HD differences are less than 10% emergence for  $T$  but in some cases more than 20 % for  $P$  (Figs. 6 and 7).

The obtained NSE values are presented in Table 1 (the corresponding graphs for HD evolution are available in Figs. S6 and S7). The NSE for  $T$  attests to a very good representation of the HD for some of the climate model simulations (0.73 to 0.81 for annual values), contrasting with a rather poor representation for  $P$  (below 0; Table 1). Based on this finding, we derive the set of best models based on temperature alone and use the same set for the ToE analysis of precipitation.

## 5.2 Temporal evolution of temperature and precipitation emergence

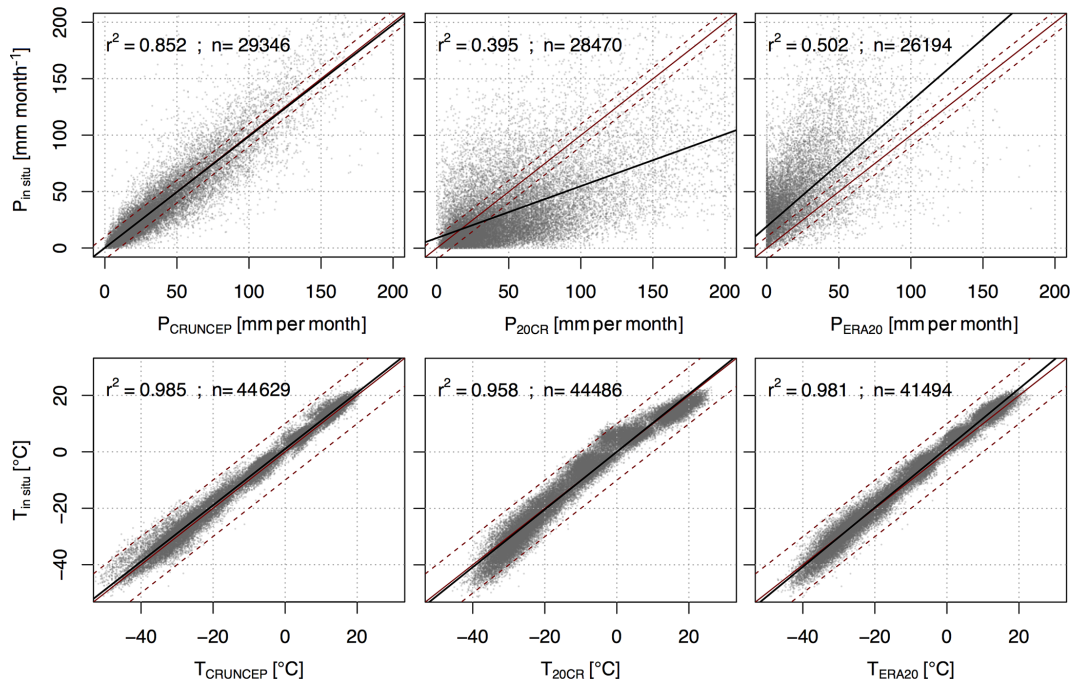
The evolutions of area-averaged annual  $T$  and  $P$  by means of CRUNCEP and the 65 CMIP5 CSs, as well as the model ensemble mean, are shown in Fig. 6. The CMIP5 ensemble mean temperature is in close agreement with CRUNCEP at the annual scale. The ensemble mean for precipitation overestimates the CRUNCEP signal, but some individual CSs are close to the CRUNCEP  $P$  estimates.

To highlight the effect of our sub-selection method for CSs, we present the study-area-averaged HD for the different data sources (1) CRUNCEP, (2) individual CSs, (3) average of the HD of all individual CSs, (4) ensemble mean, and (5) average of the HD of the 10 best CSs (Fig. 6). Video 1 and Video 2 show the spatiotemporal evolution for each of the datasets and seasons, respectively. In particular the HD of the ensemble mean is progressing very differently compared to the other datasets and shows decades earlier emergence (Fig. 6; see Sect. 6.1 for discussion). In contrast, the HD differences for both  $T$  and  $P$  between the average of all individual CSs and the average of the 10 best CSs are the lowest and show a similar evolution. Individual CSs may show a very different evolution and different regional patterns, which is also highlighted in Video 1 and Video 2. The videos include the single-best-performing CS to showcase the higher spatiotemporal variability of individual CSs compared to the averaged ones.

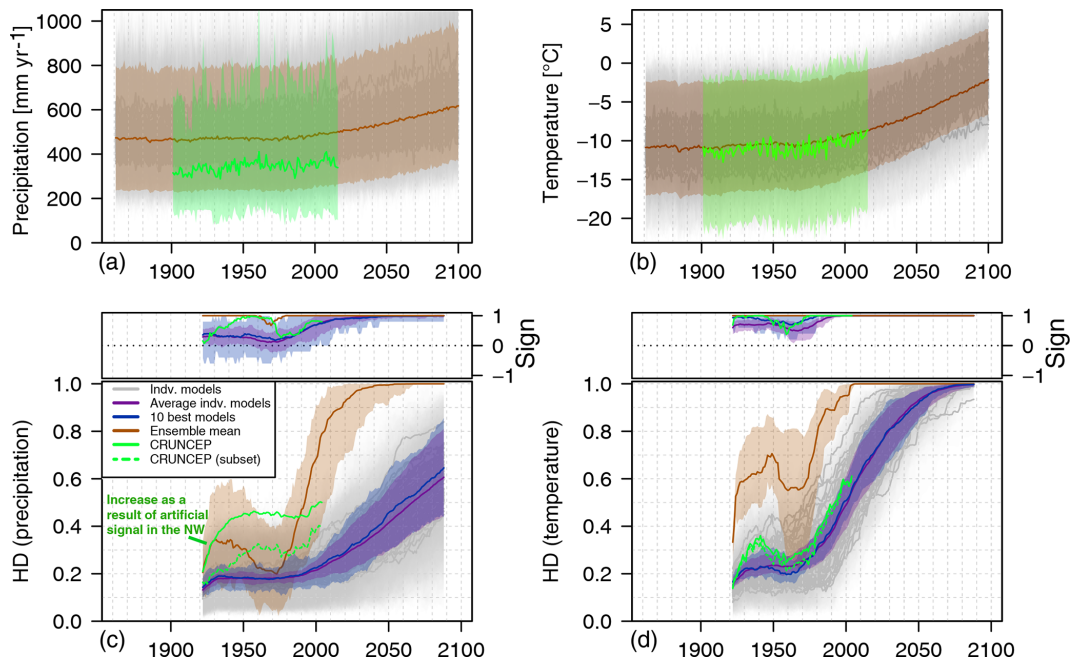
The  $T$  signals show the most prominent evolution and the most significant emergence. The emergence patterns for CRUNCEP and all individual CSs are very similar (Fig. 6). The HD shows a continuous increase starting in the 1960s. For CRUNCEP this increase is preceded by an initial HD increase at the beginning of the target period and stagnation thereafter until the 1960s. In contrast, the HD increase based on individual CSs indicates little change (< 30 %) until the 1960s and 1970s with respect to the reference period. The CRUNCEP signal emerges above 60 % by 2004 (last data point). The average HD of individual CSs and the 10 best CSs reach 90 % emergence in the 2040s and near 100 % emergence by the end of the time series (2089). In stark contrast to that is the HD based on the ensemble mean, which shows a 100 % emergence already by 2004.

For  $P$ , the evolution of the CRUNCEP signal and individual CSs, as well as the corresponding HD, show more significant uncertainties and are less well-defined (Fig. 6).

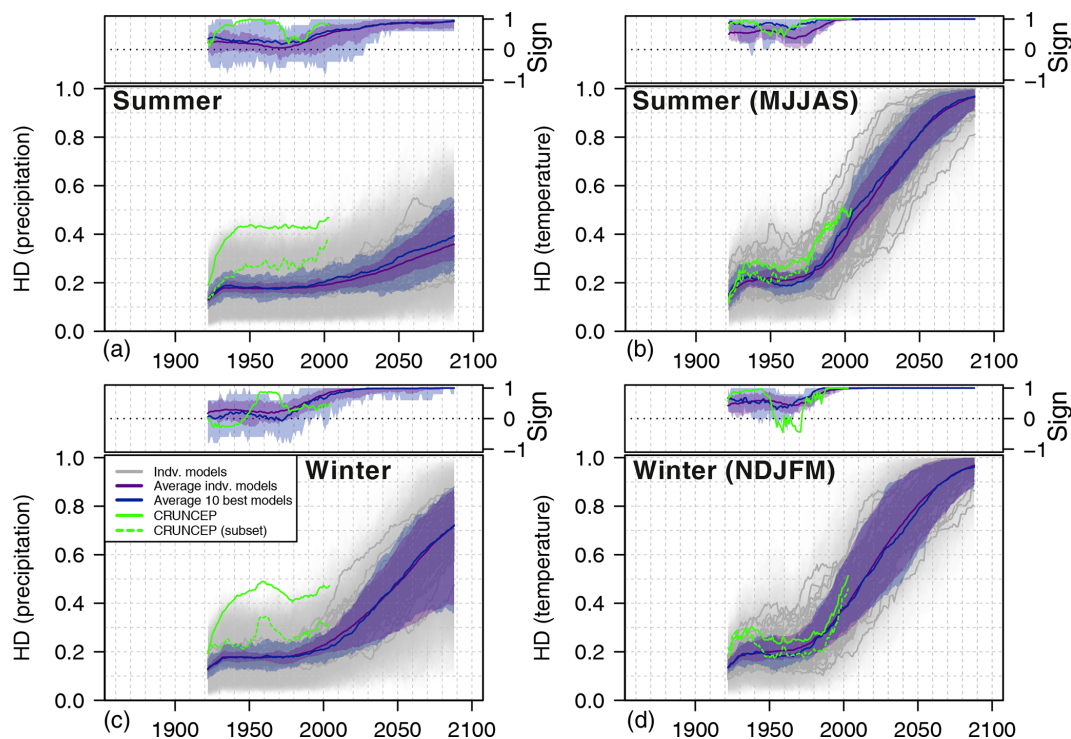




**Figure 5.** Comparison of three reanalysis datasets with in situ records (RIHMI-WDC) for monthly values. Red solid line is the 1 : 1 line; red dashed line is  $\pm 10$  mm for precipitation and  $\pm 10$  °C for temperatures.



**Figure 6.** Area-averaged  $T$  and  $P$  signal evolution, emergence as the Hellinger distance (HD), and the sign of emergence. (a, b) Evolution of summed annual precipitation (a) and mean annual temperature (b) over the entire catchment (red outline in Fig. 1). (c, d) Evolution of the HD with the sign of emergence. Shading indicates the value range over all pixels in the study area. Dashed line for CRUNCEP shows HD evolution based only on the five pixels where meteorological stations cover more than 10 years in the reference period to eliminate data issues; see also the text and Supplement for data issues of CRUNCEP. The smoothed signal of the ensemble mean (a, b) results in a strong and early emergence (c, d) that is not seen in any of the individual models.



**Figure 7.** Summer (May–September) (a, b) and winter (November–March) (c, d) emergence as the Hellinger distance (HD) and the sign of emergence. Shading indicates the value range over all pixels in the study area. Dashed line for CRUNCEP shows HD evolution based only on the five pixels where meteorological stations cover more than 10 years in the reference period to eliminate data issues.

The ensemble mean shows an emerging positive signal from 2000 onwards. The HD for CRUNCEP shows early strong emergence in the northeastern parts and to a lesser degree regionally across the entire domain (Video 2), which is related to the aforementioned data issues in the CRUNCEP dataset. The average HD of all individual CSs and of the 10 best CSs show an almost identical evolution until the 2000s when the HD shows a distinct departure reaching around 60 % emergence by 2089.

The sign change for both  $T$  and  $P$  is permanently positive once 40 % and 30 % emergence is reached, respectively. Before that, until the 1970s, around 60 % to 80 % of the pixels show a positive trend for  $T$  and 50 % to 60 % for  $P$  (Fig. 7).

The seasonal (summer and winter) evolutions generally show the same trend as the annual ones, but some differences are apparent. Most striking is the stronger regional variability in the HD for  $T$  in winter compared to summer (vertical shading in Fig. 7). For  $P$ , the seasonal difference is striking. An overall emergence of  $\sim 70$  % in winter compares to  $< 40$  % in summer. The corresponding area-wide mean ToE and corresponding changes in  $T$  and  $P$  are summarized in Table 2. The biggest ToE differences between summer and winter are apparent for  $P$  (20–29 years), whereas for  $T$  there is only a maximum difference of 1 year. The ToE of  $T$  for annual values is 11–15 years earlier compared to summer and

**Table 2.** The area-wide ToE based on the mean HD of all 65 CMIP5 model simulations and the corresponding change in temperature or precipitation at different emergence levels (HD) and seasons.

Emergence level	ToE (years)			Change ( $T$ in $^{\circ}\text{C}$ or $P$ in mm)		
	30 %	40 %	50 %	30 %	40 %	50 %
$T$ (annual)	1981	1992	2001	0.75	1.11	1.48
$T$ (summer)	1992	2005	2016	0.83	1.19	1.57
$T$ (winter)	1991	2004	2015	1.48	2.12	2.77
$P$ (annual)	2034	2049	2061	49.08	73.38	98.27
$P$ (summer)	2055	2067	2073	46.16	66.80	87.99
$P$ (winter)	2026	2041	2053	14.77	21.68	28.99

winter. For  $P$ , the annual ToE is in between the winter (earliest) and summer ToE.

### 5.3 Spatial and seasonal variability

The spatial variability in the ToE over the study area (vertical shading in Figs. 6 and 7) is displayed as maps in Figs. 8 and 9 for three different emergence levels (30 %–50 %) and the three temporally aggregated periods (annual, summer, and winter). The corresponding changes in  $T$  and  $P$  for a ToE at a given emergence level are shown in Figs. 10 and 11. Due to the nearly identical evolution of the ToE based on

the mean HD of either all individual CSs or the 10 best CSs (cf. Figs. 6 and 7, Video 1, and Video 2), we only display the results for the former.

The annual and winter analyses for  $T$  show the generally earlier ToE in the northeast compared to the southwest (Fig. 8). The summer pattern is almost reversed with an earlier ToE in the south and a later ToE in the north.

The strongest variability in the ToE for a given emergence level can be seen in the winter analysis, where the earliest and latest ToE can lie more than 30 years apart. Corresponding changes in  $T$  for a given ToE strongly depend on the spatial location and the season. For example, 50 % emergence in the  $T$  signal corresponds to 1.2 °C for the annual analysis in the south, contrasting with 3.6 °C for winter in the west (Fig. 10). Based on the temporal aggregation, up to a 2-fold difference in  $T$  change can be observed for a given emergence level (annual vs. winter; Fig. 10 and Table 2).

For  $P$ , there is also a north–south gradient with later ToE values in the south (Fig. 9). In addition to that, a pronounced later ToE along the eastern catchment boundary is visible in winter, whereas annual and summer ToE do not show such a pronounced feature. Most striking for  $P$  are the strong ToE differences between the seasons, locally with up to a 50-year earlier ToE in winter compared to summer.

Corresponding changes in  $P$  for a given ToE relate to the pronounced seasonality with the highest moisture supply in summer. This results in up to a 4-fold stronger increase in summer compared to winter, adding to the roughly 2-fold regional differences (Fig. 11).

Comparison of the ToE between the two variables  $T$  and  $P$  shows strong differences that locally reach 80 years (Figs. 8 and 9). The area-averaged differences between 30 % and 50 % emergence correspond roughly to a doubling of change in  $T$  and  $P$  for any season (Table 2).

#### 5.4 Sensitivity analysis

Although the developed approach for ToE computation relies on PDFs and is basically non-parametric, the method requires two meta-parameters with a potential outcome on the obtained emergence and ToE: target period (split year) and window width to calculate the PDFs of the reference and target periods (Fig. 12). The model simulations' internal maximum deviations for the tested meta-parameter combinations of window width and the end of the target period (split year) are around  $\pm 4$  years on average, for both for  $T$  and  $P$ . In contrast, the inter-model differences are up to 70 years at low emergence levels, which can be seen in Figs. 6 and 7. No particularly abrupt increase in the ToE for a specific year or window width is apparent. The more dominant parameter on the outcome of the ToE is the window width for  $P$ , as can be seen in the horizontal gradient in Fig. 12. For  $T$ , a stronger variability between the simulations and at different emergence levels is present (not shown), and the resulting average sensitivity in Fig. 12 is less pronounced than for  $P$ .

A slight change from a gradient with a later ToE for a late split year and short window width at 30 % emergence to a generally later ToE mainly based on split year length can be seen. The latter is represented by the vertical gradient. No particular year or window width can be identified to have a significant impact on the ToE estimates for either variable.

For both variables, the sensitivity to either meta-parameter based on all 65 CSs or the 10 best CSs is equally low. However, the standard deviation of ToE estimates is reduced by up to 6 years for the case of the 10 best simulations (Figs. S8 and S9). Derived ToE sensitivities for the full set of CSs, and the subset are very similar and reflect the similarity presented in Figs. 6 and 7 (cf. Figs. S8 and S9). The average patterns for both the 65 and 10 best CSs also largely resemble the pattern for CRUNCEP (Fig. S10). However, a sharper contrast for CRUNCEP between split years and window widths and a stronger impact on the range in the ToE are apparent. ToE estimates for low emergence levels reach up to  $\pm 9$  years for  $T$ , which is also the maximum range found amongst all individual CSs for different emergence levels. In summary, the found maximum variability resulting from the meta-parameter choice is very low ( $\pm 4$  years) in comparison to the inter-model variability (up to 70 years) and is well below commonly reported ToE bin sizes, i.e. time intervals ( $\sim 20$  years) to classify a regions' ToE.

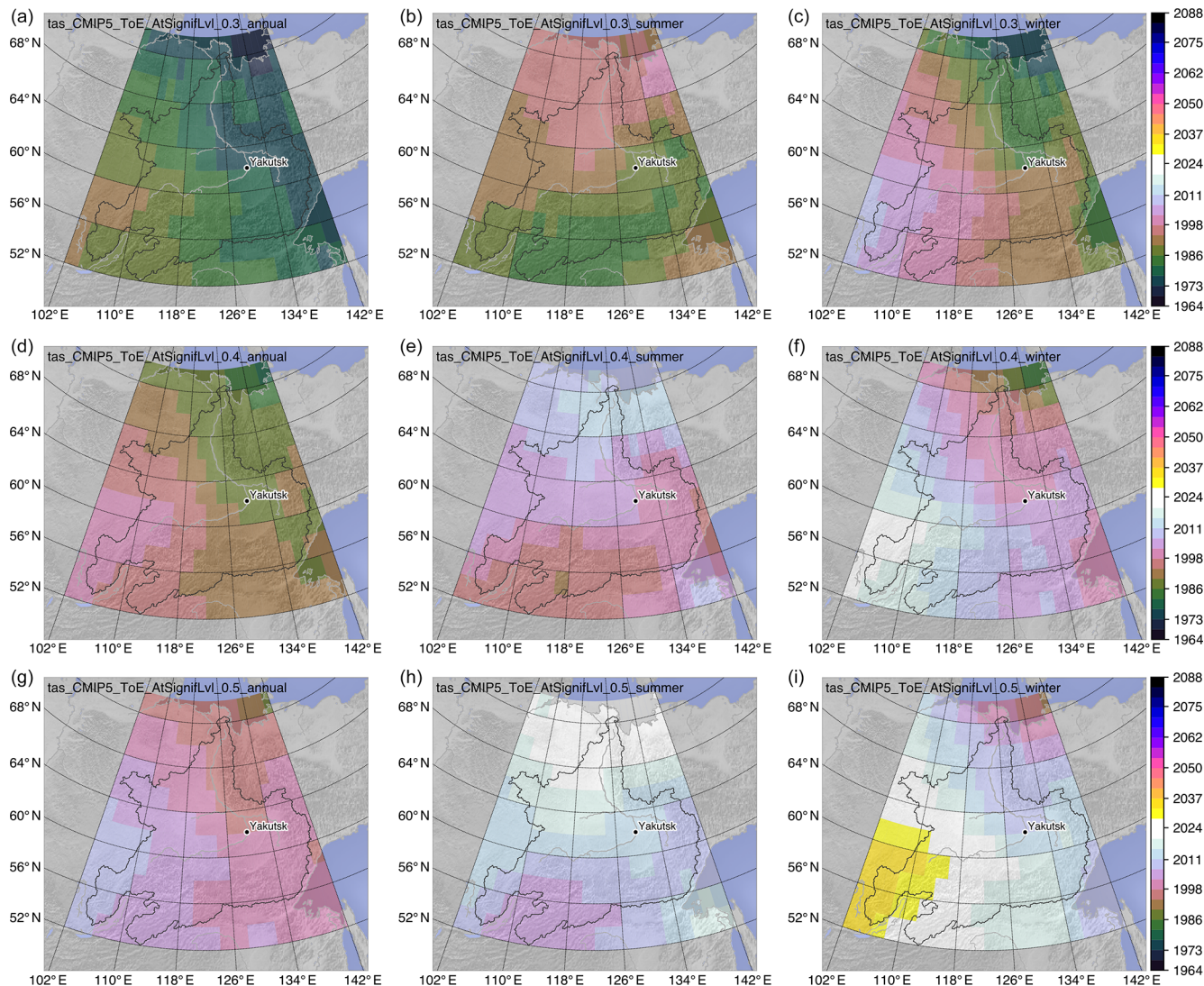
## 6 Discussion

The results showcase a strong variability between the temporal evolution of emergence and derived ToE values of the two tested climate variables  $T$  and  $P$ . Large differences also occur between the three temporal aggregations: annual, winter, and summer. These differences highlight the complexity in the climate system and emphasize that there cannot be a single answer to the general questions if and how much climate change has emerged in eastern Siberia.

### 6.1 Method

The ToE method applied in this study provides an innovative way to investigate climate change evolution and its emergence. It is different from existing ToE methods that rely on tests and are based either on the exceedance of an  $S/N$  threshold (e.g. Hawkins and Sutton, 2012) or a statistical significance level (e.g. Mahlstein et al., 2011). The new approach provides a continuous measure of emergence. This has advantages and disadvantages compared to previous methods. Striking benefits are that it facilitates the comparison of the evolution for different datasets (Figs. 6 and 7), allowing to rank and select climate simulations whose emergence signatures correspond the closest to observational data. This is a big difference with respect to preselection procedures based on statistical comparison (e.g. Mahlstein et al., 2011) or on weighting schemes that compare model simi-





**Figure 8.** Time of emergence for temperature (tas) according to 30 % (a–c), 40 % (d–f), and 50 % (g–i) emergence for annual (a, d, g), summer (b, e, h), and winter (c, f, i) values. Values are the mean over all individually determined ToE values for each of the 65 climate simulations.

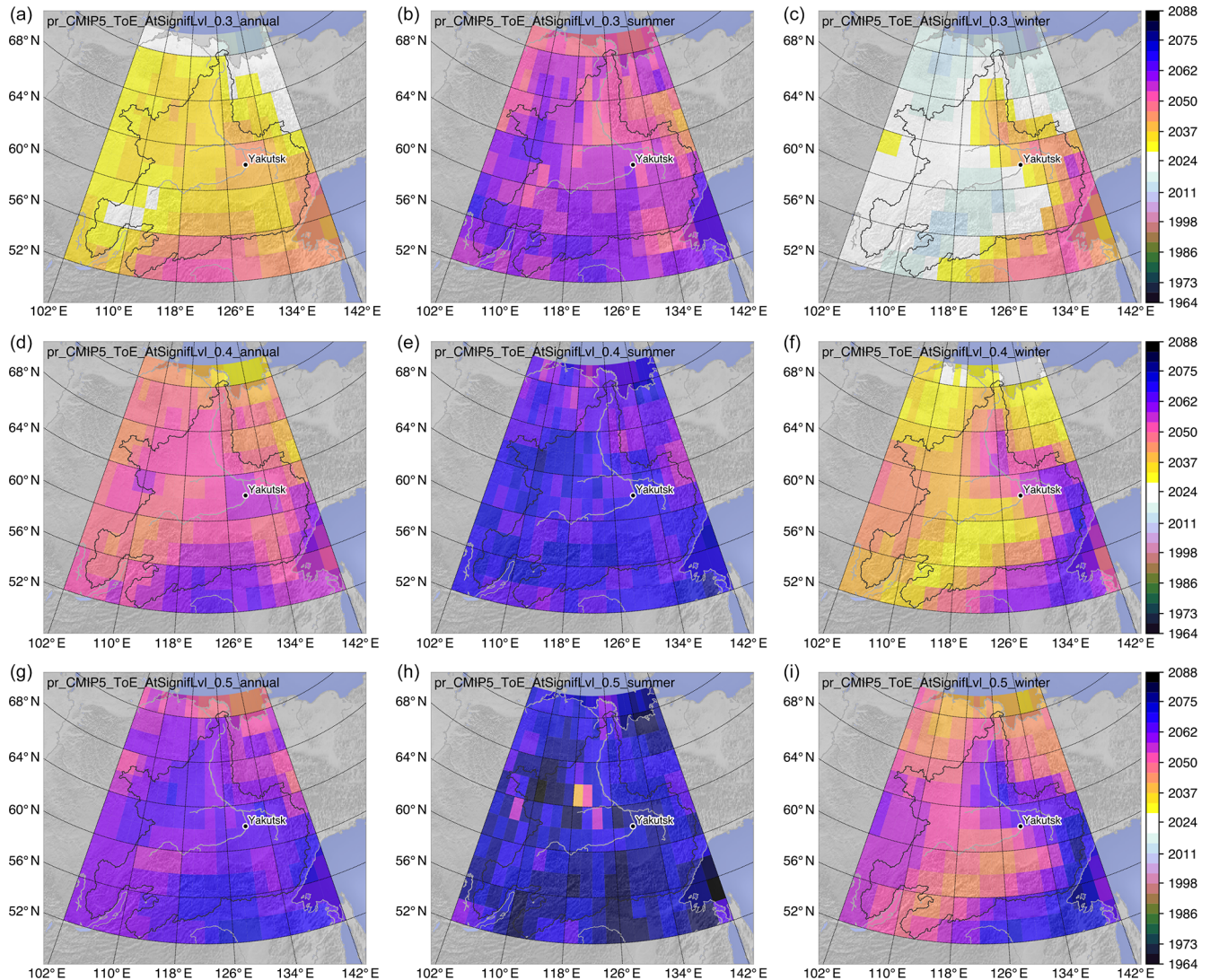
larities and the ability of a CS to represent observational data (Knutti et al., 2017). The expected downside of the developed method is the need to define an emergence level a posteriori in the present case. However, it should be kept in mind that other methods also require a threshold in form of an  $S/N$  ratio or a significance level for statistical tests. In our case, the information about the significance is directly provided by the value of emergence and allows answering questions like what a halfway-emerged climate looks like compared to initial conditions. The main difficulty might lie in finding a connection between the level of emergence of a climate variable under investigation and how it relates to possible environmental and socio-economic impacts. This certainly requires expert knowledge of changes which have already occurred or

are observed to be ongoing that involve complex interactions in permafrost landscapes.

The Hellinger distance shows particular advantages over the KS metric in a direct comparison. It is able to detect very small changes (Fig. 4), and the detection limit is not dependent on the sample size as is the KS metric, which produces a step-function-like evolution. However, the approach also comes at the price of a higher computational cost, e.g. through the calculation of PDFs using a KDE.

A very surprising finding is that independent of whether we calculate the average HD of the subset of the best models or of all CSs, the derived emergence and ToE estimates show only a few years difference, despite several decades differences between the HD of individual CSs (Figs. 6 and 7). The strongest impact of the sub-selection on the results is





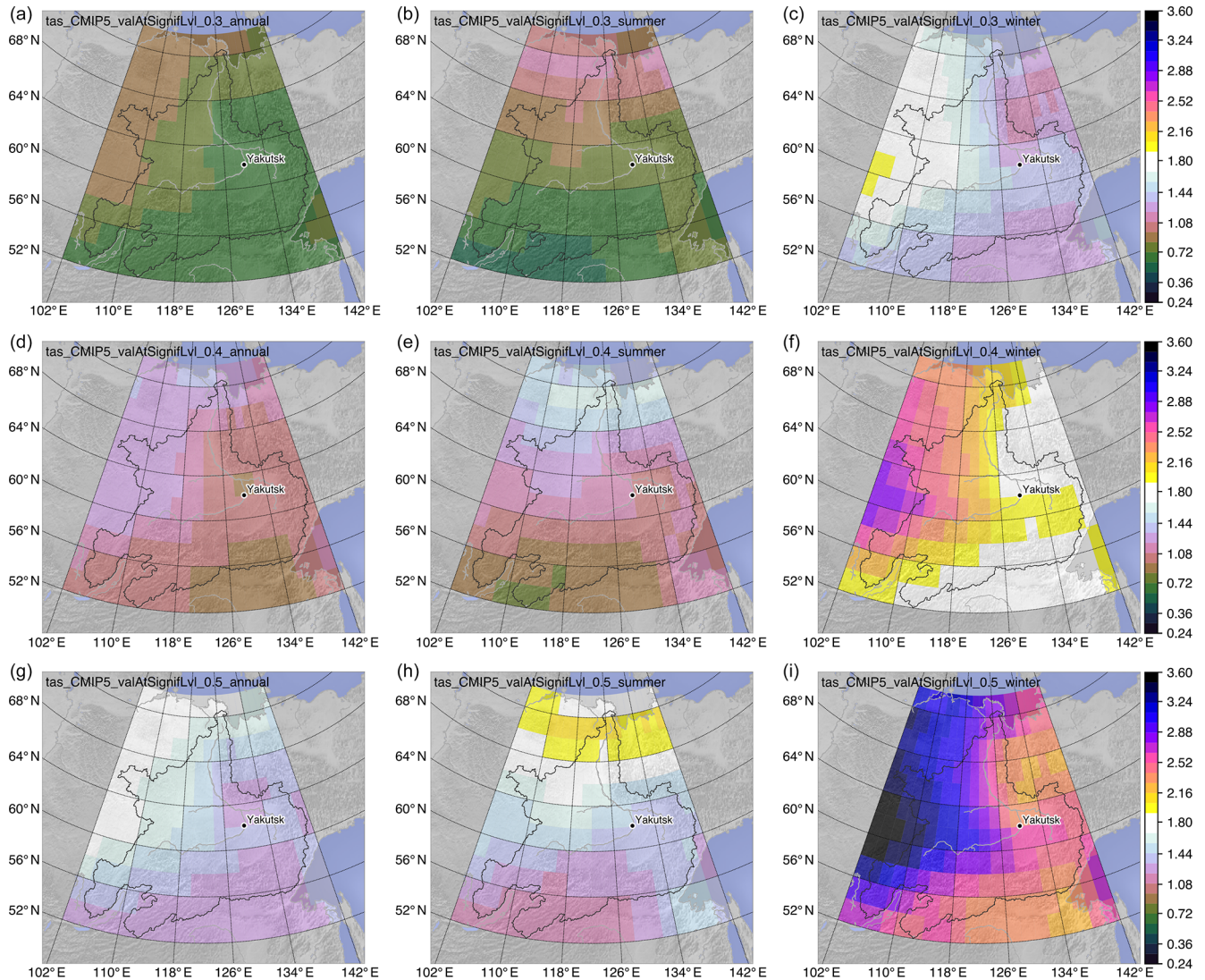
**Figure 9.** Time of emergence for precipitation (pr) according to 30 % (a–c), 40 % (d–f), and 50 % (g–i) emergence for annual (a, d, g), summer (b, e, h), and winter (c, f, i) values. Values are the mean over all individually determined ToE values for each of the 65 climate simulations. Artefacts at 50 % emergence in summer (earlier ToE than for 40 %) are due to the limited number of model simulations with emergence.

the reduced ToE variability between individual HD evolutions (Figs. S8 and S9). A rather low impact on the ToE from choosing a preselected number of CSs that best match the observations was also reported by Mahlstein et al. (2011). In stark contrast, applying the method on the ensemble mean yields significantly earlier and stronger emergence (Fig. 6), resulting from the extreme narrow range of the filtered signal. This is similar to the muting of internal climate variability through having multiple model runs using the same climate model (e.g. Deser et al., 2016). Resulting PDFs are very narrow, and exceedance occurs more rapidly than we can observe in any of the individual signals, including CRUNCEP (Figs. 6 and 7). In the present case, this muting is inconclusive because it results from the averaging over differ-

ent climate models with different internal variability. As our method is designed to specifically detect the change of a signal with respect to its natural variability, the presence of variability is a prerequisite.

The development of our method was made with the intention to have a wide range of applications, including nearly all types of time series data. Like the application of the KS test (King et al., 2015; Mahlstein et al., 2012), PDFs can be obtained for any data distribution, and their overlap as a measure for emergence can be easily understood. The resulting emergence as a time series provides the advantage over previous methods to investigate how a signal has emerged in detail. How different datasets and the utilization of different temporal resolutions (e.g. monthly data) affect the determi-





**Figure 10.** Temperature (tas) change ( $^{\circ}\text{C}$ ) corresponding to 30 % (a–c), 40 % (d–f), and 50 % (g–i) emergence for annual (a, d, g), summer (b, e, h), and winter (c, f, i) values. Values are the mean over all individually determined changes for each of the 65 climate simulations.

nation of the ToE should be explored in more detail in the future.

## 6.2 Sensitivity

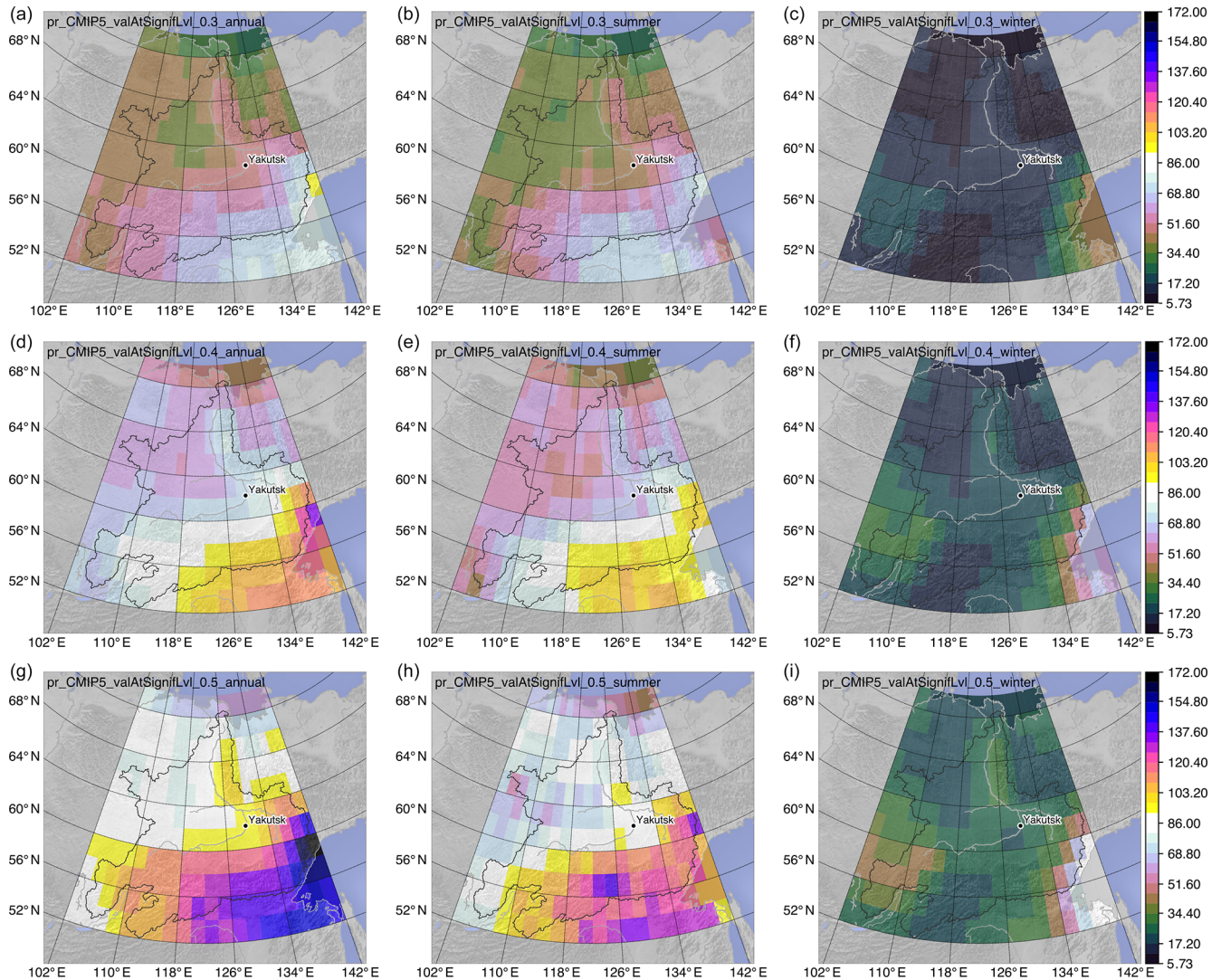
The expected uncertainty from the needed meta-parameter selection of window widths and the reference period (cf. Hawkins and Sutton, 2016) has a rather negligible impact on the overall outcome of the ToE compared to the differences resulting from the spread between individual CSs (Fig. 12): the study-area-averaged variability of  $\pm 4$  years across all meta-parameter combinations contrasts with up to 70-year differences between individual CSs. The analysis revealed some systematic patterns in the form of a dominant gradient in vertical direction for  $T$ , and a horizontal gradi-

ent for  $P$  (Fig. 12), providing insights into some important aspects related to the data and the method itself.

A longer reference period and ToE accompanied later, as can be seen for temperature (Fig. 12), indicate mainly that a trend towards increased values at the end of the reference period is present. Extending the reference period provides a wider PDF and higher values. The target periods will stay overlapping longer, and the ToE occurs later. A reverse situation with lower values towards the end is apparent in single cases only (not shown) so that the vertical gradient is reversed.

The gradient towards a later ToE for smaller window widths for precipitation (Fig. 12) is somewhat counterintuitive, as a small window size implies more variability. It results from local minima (low-precipitation years) that strongly impact the PDFs in the target period. They can thus





**Figure 11.** Precipitation (pr) change (mm) corresponding to 30 % (a–c), 40 % (d–f), and 50 % (g–i) emergence for annual (a, d, g), summer (b, e, h), and winter (c, f, i) values. Values are the mean over all individually determined changes for each of the 65 climate simulations.

again become similar to the PDF of the reference period. Consequently, an earlier continuous exceedance is not treated as permanent, and the finally obtained ToE is later for a small window width. Longer window widths will cause the extreme values to have a less significant impact on the PDF. The resulting dissimilarity stays above the threshold, and the derived ToE is earlier even if the initial threshold was crossed later. The same reason seems to cause the earlier emergence for annual  $T$  and  $P$  values, where a single extreme month has a relatively low impact on the annual PDF compared to its stronger impact on the seasonal PDF (Table 2).

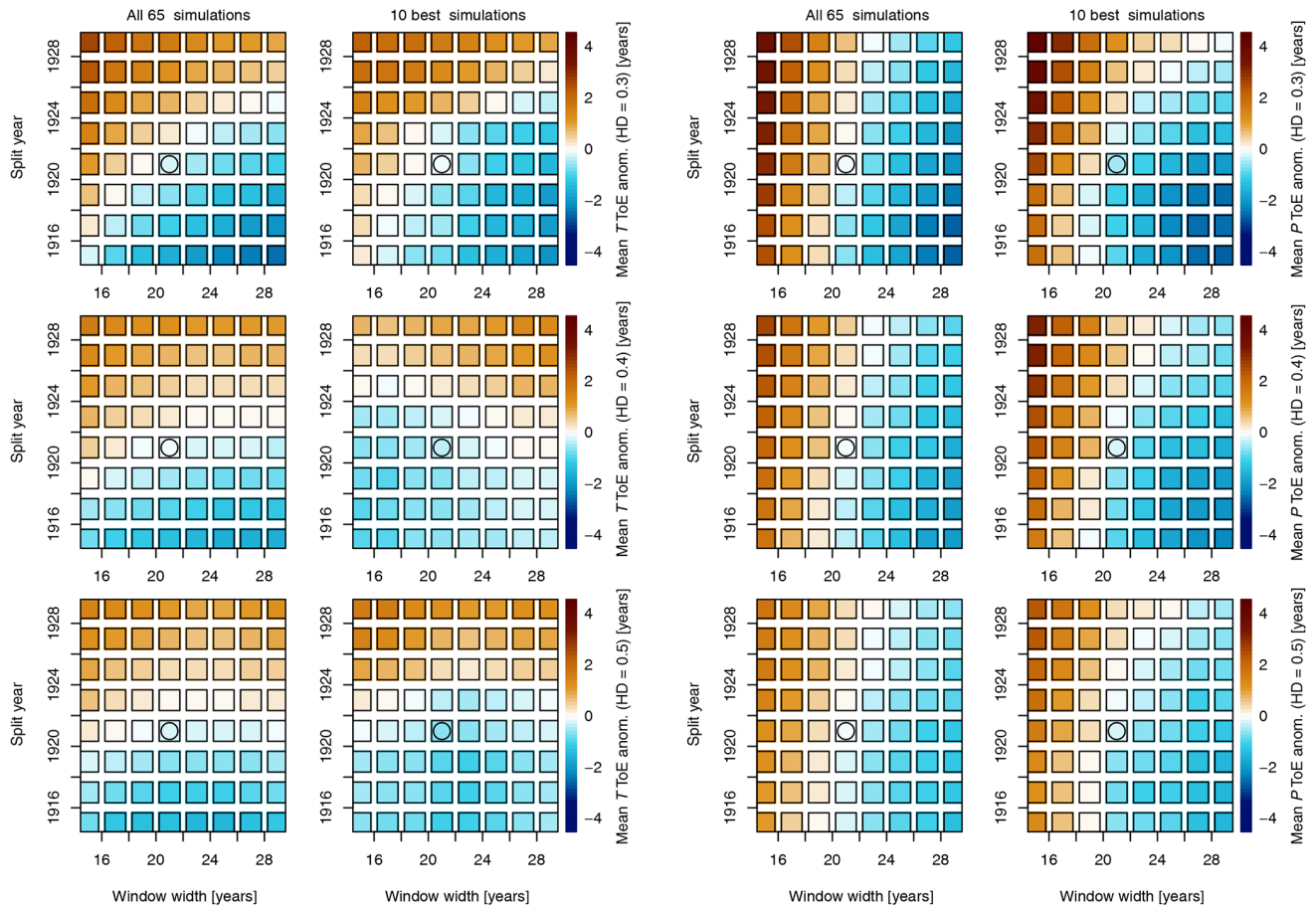
We assume a low impact on the uncertainty from the KDE-based determination of the PDFs for the HD calculation. Even though we have not tested the impact of the bandwidth on the different window widths in the sensitivity analysis, we have shown that for similar data and even larger ranges

in sample sizes (10 to 100), the overall uncertainty is in the range of  $\pm 5\%$  (Sect. 4 and Fig. 4). Since our approach also shows a low bias for non-normal distributions once a certain HD is reached (in the presented case 0.3; Sect. 4), we assume a wide range of applications. An option to turn off the automatic bandwidth determination is possible in the code implementation. This provides the possibility to test how this meta-parameter affects other types of data.

In summary, the sensitivity analysis is a valuable and relatively easy-to-apply tool to explore how a specific dataset and a combination of meta-parameters influence ToE estimates.

### 6.3 Data

Our initial selection of reanalysis data through comparison with observational data has shown good agreements for  $T$  but, with the exception of CRUNCEP, a very weak repre-



**Figure 12.** Impact of window width and split year on the ToE for  $T$  (left two columns) and  $P$  (right two columns) as the mean deviation from the mean over all combinations of window width and split year. Individual left columns for all 65 CSs and individual right columns for the subset of 10 best CSs. Rows show different emergence levels (30 %–50 %). The average ToE values and standard deviations are available in Figs. S8 and S9.

resentation for  $P$ . In combination with the systematic bias for the warmest and coldest temperatures for 20CR (Fig. 5), this also requires a cautious selection of CSs based on observational data in the region. As CRUNCEP results from interpolated observational data, the good match is not surprising. How well this dataset represents the actual conditions for the times where there are no measurements available remains unsolved. The apparent recycling of a single year in the CRUNCEP time series (Sect. 5.1) and the resulting standard deviation close to zero (Fig. S3) indicate that the data are biased or unreliable in the northwestern part.

Interestingly, the selected best CSs are from different models within the ensemble. That is despite some of the selected best CSs belonging to models that are represented with several runs in the ensemble (cf. Figs. S5 and S4 and Table S1), meaning that internal climate variability within the models of the ensemble plays an important role for the case presented here and potentially other ToE methods. It also stresses the benefit of ensembles to include multiple runs of a model

because it additionally helps other approaches to identify internal climate variability (Deser et al., 2016). While the HD comparison to select CSs for  $T$  shows very good matches (Table 1 and Fig. S6), the imperfect matches for  $P$  imply a high level of uncertainty that is difficult to assess (Fig. S7). The best indicator suggesting some reliability is the fact that the sensitivity for CRUNCEP (Fig. S10) shows similar patterns compared to the ensemble of CSs (Fig. 12). This pattern match can be interpreted as both datasets having a similar variability and distribution of extreme values, as well as an overall similar trend, as discussed in Sect. 6.2. However, the presented results for  $P$  should be treated with caution. Climate model simulations and reanalysis data need to be improved to derive regionally reliable estimates, which in turn are needed to investigate the physical processes in the Earth system that can aid decision making.

#### 6.4 ToE

ToE values are with respect to the reference period (1901–1921) and thus slightly later than otherwise chosen pre-industrial reference periods (e.g. 1881–1910 in Vautard et al., 2014, or 1860–1910 in King et al., 2015) but longer than in ToE studies focusing on observational data. There is no way to avoid this selection in the current study. The chosen period is the earliest possible one to have a basis for the comparison of the observational data and CMIP5 model simulations.

Data issues are almost always due to the lack of data or data quality (e.g. Hawkins and Sutton, 2016). The sensitivity analysis (Fig. 12) shows that choices of reference periods between 1901–1915 and 1901–1929 have a relatively small impact on the obtained ToE and that uncertainties from the spread in individual CSs are an order of magnitude higher. Since we report the emergence as a continuous signal, the question arises as to when this signal should be considered as significantly different with respect to the reference period. In other words, how strongly does a PDF need to change from its initial shape and position to indicate a significantly emerged climate? An obvious way is to compare obtained results with previous ToE studies and with reported changes in climatic variables.

King et al. (2015) reported the ToE for the region of the Lena River between 1980 and 2000 for summer temperatures and between 1980 and 2000 and in a few occasions between 1960 and 1980 for winter temperatures. These ToE values were obtained through a KS test and using 1860–1910 as a reference period. The reported ToE values correspond to the pronounced onset in the HD signal (Fig. 7) and an emergence level of around 30% (Fig. 8). King et al. (2015) further reported ToE values for winter precipitation between 2000 and 2020 in the lowlands and for 20 to 40 years later in the east and southeast. The same spatial pattern is derived with our method. Again, the timing corresponds to an emergence level of around 30%. Mahlstein et al. (2011) reported temperatures corresponding to the statistically significant identified changes using the KS test with a reference period of 1900–1929. A direct comparison is difficult, as they report these temperatures for countries. However, their identified value of 1.1 °C for summer temperatures for Russia corresponds to the 30% emerged signal in our study (Fig. 10).

Comparisons with temperature (Desyatkin et al., 2015; Fedorov et al., 2014b) and precipitation trends (Gorokhov and Fedorov, 2018) are somewhat complicated due to the different starting points of the datasets. Trends in Gorokhov and Fedorov (2018) are with respect to the 1966–2016 period. As indicated in Fig. 6, the study-area-wide precipitation signal shows relatively high values in the 1960s, with a positive emergence for CRUNCEP and a decline thereafter (Figs. 6 and 7). The derived trends in Gorokhov and Fedorov (2018) start in this positive emergence and are consequently depicting a negative trend in the northern regions ( $\sim -8$  mm per

decade), where precipitation changes according to the CS are lowest (Fig. 11). Gorokhov and Fedorov (2018) still find increasing positive trends towards the south ( $\sim 16$  mm per decade). This north–south gradient is reflected by our results (Fig. 10) even though we cannot associate any trend value with a derived emergence level.

Fedorov et al. (2014b) reported generally stronger positive trends for temperatures in the eastern and southern mountain regions in our study area and lower trends in the lowlands and towards the east. A general overlay of earlier ToE values (Fig. 10) is visible for stronger trends, and vice versa. However, weaker trends in the most northern part and one of the strongest trends for Yakutsk in the lowland render a conclusive comparison difficult. Fedorov et al. (2014b) use a dataset with variable station record length, which might explain to some degree the discrepancies. In the end, such differences are expected given the variability in the CMIP5 model simulations and individual offsets to CRUNCEP (Fig. 6).

In relation to such evolutions in  $T$  and  $P$ , ground temperatures and hydrological conditions are especially impacted. Fedorov et al. (2014a) pointed out that in the 1950s high ground temperatures might have initiated thermokarst lake formation. The identification of periods in which a triggering event initiates a state change are not included in any ToE method, despite their potential for landscape changes, which in turn has far-reaching impacts on permafrost evolution (Crate et al., 2017; Grenier et al., 2018; Walvoord and Kurylyk, 2016; Westermann et al., 2017). However, Fedorov et al. (2014a) also mention that despite the early initiation, the main progression of lake formation occurred in the 1990s, which represents the previously mentioned time period where emergence levels reach 30%.

Warmer summer temperatures of 1 to 2 °C in the future (Fig. 10) imply a strong impact on the hydrology by means of potential evapotranspiration increase and the evolution of thermokarst lakes. It is, however, difficult to exactly identify how the co-emergence of  $T$  and  $P$  at different rates (Figs. 8 and 9) will affect the evolution of thermokarst lakes that are currently in equilibrium between precipitation and ground ice melt water input and evapotranspiration output. Karlsson et al. (2012) point out that an increase in  $T$  would likely increase lake bodies due to the more important input from ground ice melt. This is in agreement with conclusions by Fedorov et al. (2014a) for the formation of new thermokarst lakes. However, old almas lakes with reduced input from ground ice melt might undergo a reduction if evapotranspiration increases more than total precipitation influx. More recently, Ulrich et al. (2017) have shown through multiple regression analyses that, in particular, increasing winter precipitation and winter temperatures control lake area changes of young and old thermokarst lakes in central Yakutia. As these two variables show the strongest emergence (Table 2), an increase in thermokarst lake area and a resulting overall change in the hydrological system should be expected.

Mean annual discharge of the Lena River has only increased significantly in the most recent time period of 2006–2012 (Gautier et al., 2018). However, late-spring discharges during the ice breakup had already experienced a strong increase a decade earlier (1996–2005). These periods lag behind the ToE presented for  $T$  but precede the ToE for  $P$  at 30 % emergence (Fig. 7). Taking into account the mutual interactions between temperatures and precipitation, which results in snow cover and ground thermal insulation as well as snow stocks for melt (Karlsson et al., 2011; Westermann et al., 2017), systematic changes should occur as a result of the two. The onset of winter  $P$  emergence in the 1990s and more strongly thereafter would provide a possible explanation. It would also not contradict the strong positive emergence for  $P$  in the 1950s and 1960s (Fig. 7) that has not resulted in detectable flood events. The HD and the signal of change for the CRUNCEP data show that more precipitation (positive signal) occurred alongside more negative temperatures (negative signal), which would counteract strong melting events.

The implied changes in  $T$  and  $P$  at different emergence levels will certainly have significant impact on various environmental and socio-economic aspects. How much these changes, at 50 % emergence and more and, at different seasons, will impact the complex hydrological system is difficult to assess and should be explored further in the future. Such assessments require, however, a continuation and advance in the modelling of cryo-hydrological systems that allow for a better understanding of how the climate variables affect the involved processes (Walvoord and Kurylyk, 2016). This, in turn, requires for the continuation of measurement efforts in the large, remote, and difficult to access Arctic regions, where observational data are sparse.

## 7 Conclusions

We developed a novel method for the determination of climate change emergence. Its non-parametric character allows for its application on data with different types of data distributions, which we showcased for  $T$  and  $P$  in the Lena River catchment, and use with synthetic datasets. The strongest biases were found in a synthetic dataset for low changes in PDFs when the distributions are strictly positive and heavily skewed, which might be expected for high-frequency data, like hourly precipitation. Even then, once the distributions show an HD of 0.3, these biases fall below 10 % and attest a large application range of the approach. Unlike other ToE methods that rely on a threshold or statistical test, our method provides a continuous signal of emergence. This facilitates an extended analysis of the progression of climate change signals and provides a useful tool for comparing datasets regarding their similarity in describing climate change. It comes with the need for applying a threshold a posteriori. Comparison with ToE estimates from other stud-

ies indicates that an equivalent ToE occurs at an emergence level of around 30 % for both  $T$  and  $P$ .

A comparison of three commonly used state-of-the-art reanalysis datasets with observational data from meteorological stations has revealed a generally good agreement for  $T$ , but only the tested CRUNCEP data provided  $P$  estimates with little bias. Even within this dataset, we found artificial behaviour in the time period 1901–1921 for the  $P$  estimates, probably due to the limited number of meteorological stations operating at that time. In combination with the  $P$  intensity bias of many of the CSs, conclusions on the emergence of  $P$  are rendered uncertain.

Our method allowed us to compare the evolution of emergence of  $T$  and  $P$  from CRUNCEP with those of 65 climate model simulations taken from a CMIP5 ensemble. This provides an alternative to preselection methods based on dataset statistics or weighting schemes for climate models and simulations.

We obtain surprisingly similar emergence times independent of using either the mean emergence of all simulations or from our sub-selection of the 10 best-performing simulations. On the contrary, individual models show estimate differences of up to 70 years at low emergence levels. This provides confidence in using ensembles which are large enough rather than somehow-chosen sub-selections to identify the ToE if no or insufficient observational data are available. Nonetheless, the selection method presented here might provide means to discriminate the most reliable data sources in other more documented regions or contexts. The conclusion to include full climate ensembles rather than single simulations is supported by a consistent similarity between the full set and the subset of CSs in all applied cases ( $T$  and  $P$  for annual, summer, and winter). The differences in derived emergence for reanalysis and climate simulations, however, stress the need for model improvements and an effort for continuous observational data, which can be comprehensively utilized in the presented approach.

Finally, the methodology should be explored in the future to analyse further-impacted variables (e.g. ground temperatures and hydrological conditions) in the complex cryo-hydrological system to identify spatiotemporal links. Ultimately, these are needed to derive an understanding of how and when climate change will impact the numerous aspects of this system.

*Code availability.* The main code to process and analyse the data is available in the scripting language Python under the GitHub repository [https://github.com/pohleric/toe\\_tools](https://github.com/pohleric/toe_tools) (last access: 19 May 2020) (Pohl, 2020).

*Data availability.* 20th Century Reanalysis V2c data are provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their website at <http://www.esrl.noaa.gov/psd/> (last access:



19 May 2020) (Compo et al., 2011). CRUNCEP v7 data are available through registration at the website <https://rda.ucar.edu/datasets/ds314.3/> (last access: 19 May 2020) (Viovy, 2018). ERA20 data are available from ECMWF data servers through the Python module “ecmwfapi” at <https://pypi.org/project/ecmwf-api-client/> (last access: 19 May 2020) (ECMWF, 2019). The RIHMI-WDC observational dataset used in this study can be obtained through the website [https://cdiac.ess-dive.lbl.gov/ndps/russia\\_daily518.html](https://cdiac.ess-dive.lbl.gov/ndps/russia_daily518.html) (last access: 19 May 2020) (Bulygina and Razuvaev, 2012).

**Video supplement.** Video 1 (<https://doi.org/10.5446/42824>) shows the spatiotemporal evolution of emergence for temperature in the Lena River catchment for the different data sources. It shows, by row, (1) CRUNCEP, (2) the average emergence of all individual CSs, (3) the average of the HD of the 10 best CSs, and (4) the single-best-performing model to showcase the higher variability of individual models compared to the averaged evolutions. Columns from left to right represent the different temporal analyses of annual, summer, and winter. Blue dots indicate a negative sign of the emergence.

Video 2 (<https://doi.org/10.5446/42825>) is the same as Video 1 but for precipitation.

**Supplement.** The supplement is added as additional document and provides information about the spatiotemporal variability of datasets and gives a more detailed view on some statistics. The supplement related to this article is available online at: <https://doi.org/10.5194/hess-24-2817-2020-supplement>.

**Author contributions.** EP and CG designed the study. EP did the calculation and produced the figures, maps, and the toolbox. EP wrote the outline. All authors contributed to the discussion and refinement of the paper.

**Competing interests.** The authors declare that they have no conflict of interest.

**Acknowledgements.** We thank the IPSL-EUR postdoc initiative that initiated a workshop on the ToE issue. We acknowledge the discussions with Pascal Terray, Goulven Laruelle, Marco Gaetani, and Vincent Thieu. Special thanks go to Alexander Fedorov and Pavel Konstantinov from the Melnikov Permafrost Institute, Yakutsk, for providing data and discussion. We acknowledge the World Climate Research Programme’s Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modelling groups (listed in Table S1 of this paper) for producing and making available their model output. For CMIP the US Department of Energy’s Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led the development of the software infrastructure in partnership with the Global Organization for Earth System Science Portals. Support for the 20th Century Reanalysis project dataset is provided by the US Department of Energy Office of Science Innovative and Novel Computational Impact on Theory and Experiment (DOE INCITE) programme and

Office of Biological and Environmental Research (BER) and by the National Oceanic and Atmospheric Administration Climate Program Office.

**Financial support.** This research has been supported by the Agence Nationale de la Recherche (grant no. ANR-11-IDEX-0004-17-EURE-0006).

**Review statement.** This paper was edited by Thomas Kjeldsen and reviewed by Wolfgang Schwanghart and one anonymous referee.

## References

- Beermann, F., Langer, M., Wetterich, S., Strauss, J., Boike, J., Fiencke, C., Schirrmeister, L., Pfeiffer, E.-M., and Kutzbach, L.: Permafrost Thaw and Liberation of Inorganic Nitrogen in Eastern Siberia, *Permafrost Periglac. Process.*, 28, 605–618, <https://doi.org/10.1002/ppp.1958>, 2017.
- Benestad, R. E.: A comparison between two empirical downscaling strategies, *Int. J. Climatol.*, 21, 1645–1668, <https://doi.org/10.1002/joc.703>, 2001.
- Benestad, R. E., Mezghani, A., and Parding, K. M.: esd V1.0, Zenodo, <https://doi.org/10.5281/zenodo.29385>, 2015.
- Bianchi, M.: Bandwidth Selection in Density Estimation, in: *XploRe: An Interactive Statistical Computing Environment*, Springer New York, NY, 101–112, 1995.
- Boike, J., Grau, T., Heim, B., Günther, F., Langer, M., Muster, S., Gouttevin, I., and Lange, S.: Satellite-derived changes in the permafrost landscape of central Yakutia, 2000–2011: Wetting, drying, and fires, *Global Planet. Change*, 139, 116–127, <https://doi.org/10.1016/j.gloplacha.2016.01.001>, 2016.
- Bulygina, O. N. and Razuvaev, V. N.: Daily Temperature and Precipitation Data for 518 Russian Meteorological Stations, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tennessee, <https://doi.org/10.3334/CDIAC/cli.100>, 2012.
- Cha, S. H.: Comprehensive survey on distance/similarity measures between probability density functions, *Int. J. Math. Model. Meth. Appl. Sci.*, 1, 300–307, 2007.
- Cohen, J., Zhang, X., Francis, J., Jung, T., Kwok, R., Overland, J., Tayler, P. C., Lee, S., Laliberte, F., Feldstein, S., Maslowski, G., Henderson, G., Stroeve, J., Coumou, D., Handorf, D., Semmler, T., Ballinger, T., Hell, M., Kretschmer, M., Vavrus, S., Wang, M., Wang, S., and Blackport, R.: Arctic change and possible influence on mid-latitude climate and weather, A US CLIVAR white paper (No. 2018-1), edited by: Uhlenbrock, K., US CLIVAR Project Office, Washington, D.C., <https://doi.org/10.5065/D6TH8KGW>, 2018.
- Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Matsui, N., Allan, R. J., Yin, X., Gleason, B. E., Vose, R. S., Rutledge, G., Bessemoulin, P., Brönnimann, S., Brunet, M., Crouthamel, R. I., Grant, A. N., Groisman, P. Y., Jones, P. D., Kruk, M. C., Kruger, A. C., Marshall, G. J., Maugeri, M., Mok, H. Y., Nordli, Ø., Ross, T. F., Trigo, R. M., Wang, X. L., Woodruff, S. D., and Worley, S. J.: The Twentieth Century Reanalysis Project, *Q. J. Roy. Meteorol. Soc.*, 137, 1–28, <https://doi.org/10.1002/qj.776>, 2011.

- Crate, S., Ulrich, M., Habeck, J. O., Desyatkin, A. R., Desyatkin, R. V., Fedorov, A. N., Hiyama, T., Iijima, Y., Ksenofontov, S., Mészáros, C., and Takakura, H.: Permafrost livelihoods: A transdisciplinary review and analysis of thermokarst-based systems of indigenous land use, *Anthropocene*, 18, 89–104, <https://doi.org/10.1016/j.ancene.2017.06.001>, 2017.
- Deser, C., Knutti, R., Solomon, S., and Phillips, A. S.: Communication of the role of natural variability in future North American climate, *Nat. Clim. Change*, 2, 775–779, <https://doi.org/10.1038/nclimate1562>, 2012.
- 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, *J. Climate*, 29, 2237–2258, <https://doi.org/10.1175/JCLI-D-15-0304.1>, 2016.
- Desyatkin, R., Fedorov, A., Desyatkin, A., and Konstantinov, P.: Air temperature changes and their impact on permafrost ecosystems in eastern Siberia, *Therm. Sci.*, 19, 351–360, <https://doi.org/10.2298/TSCI150320102D>, 2015.
- ECMWF: ERA-20C Project (ECMWF Atmospheric Reanalysis of the 20th Century), Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory, <https://doi.org/10.5065/D6VQ30QG>, 2014.
- ECMWF – European Centre for Medium-Range Weather Forecasts: ecmwf-api-client, available at: <https://pypi.org/project/ecmwf-api-client/> (last access: 19 May 2020), 2019.
- Fedorov, A. N., Gavriliev, P. P., Konstantinov, P. Y., Hiyama, T., Iijima, Y., and Iwahana, G.: Estimating the water balance of a thermokarst lake in the middle of the Lena River basin, eastern Siberia, *Ecohydrology*, 7, 188–196, <https://doi.org/10.1002/eco.1378>, 2014a.
- Fedorov, A. N., Ivanova, R. N., Park, H., Hiyama, T., and Iijima, Y.: Recent air temperature changes in the permafrost landscapes of northeastern Eurasia, *Polar Sci.*, 8, 114–128, <https://doi.org/10.1016/j.polar.2014.02.001>, 2014b.
- Gautier, E., Dépret, T., Costard, F., Vermoux, C., Fedorov, A., Grancher, D., Konstantinov, P., and Brunstein, D.: Going with the flow: Hydrologic response of middle Lena River (Siberia) to the climate variability and change, *J. Hydrol.*, 557, 475–488, <https://doi.org/10.1016/j.jhydrol.2017.12.034>, 2018.
- Giorgi, F. and Bi, X.: Time of emergence (TOE) of GHG-forced precipitation change hot-spots, *Geophys. Res. Lett.*, 36, 1–6, <https://doi.org/10.1029/2009GL037593>, 2009.
- Gorokhov, A. N. and Fedorov, A. N.: Current Trends in Climate Change in Yakutia, *Geogr. Nat. Resour.*, 39, 153–161, <https://doi.org/10.1134/s1875372818020087>, 2018.
- Grenier, C., Anbergen, H., Bense, V., Chanzy, Q., Coon, E., Collier, N., Costard, F., Ferry, M., Frampton, A., Frederick, J., Gonçalves, J., Holmén, J., Jost, A., Kokh, S., Kurylyk, B., McKenzie, J., Molson, J., Mouche, E., Orgogozo, L., Pannetier, R., Rivière, A., Roux, N., Rühaak, W., Scheidegger, J., Selroos, J. O., Therrien, R., Vidstrand, P., and Voss, C.: Groundwater flow and heat transport for systems undergoing freeze-thaw: Intercomparison of numerical simulators for 2D test cases, *Adv. Water Resour.*, 114, 196–218, <https://doi.org/10.1016/j.advwatres.2018.02.001>, 2018.
- Guimberteau, M., Zhu, D., Maignan, F., Huang, Y., Yue, C., Dantec-Nédélec, S., Otlé, C., Jornet-Puig, A., Bastos, A., Laurent, P., Goll, D., Bowring, S., Chang, J., Guenet, B., Tifafi, M., Peng, S., Krinner, G., Ducharne, A., Wang, F., Wang, T., Wang, X., Wang, Y., Yin, Z., Lauerwald, R., Joetzer, E., Qiu, C., Kim, H., and Ciais, P.: ORCHIDEE-MICT (v8.4.1), a land surface model for the high latitudes: model description and validation, *Geosci. Model Dev.*, 11, 121–163, <https://doi.org/10.5194/gmd-11-121-2018>, 2018.
- Hawkins, E. and Sutton, R.: Time of emergence of climate signals, *Geophys. Res. Lett.*, 39, 1–6, <https://doi.org/10.1029/2011GL050087>, 2012.
- Hawkins, E. and Sutton, R.: Connecting Climate Model Projections of Global Temperature Change with the Real World, *B. Am. Meteorol. Soc.*, 97, 963–980, <https://doi.org/10.1175/BAMS-D-14-00154.1>, 2016.
- Hawkins, E., Anderson, B., Diffenbaugh, N., Mahlstein, I., Betts, R., Hegerl, G., Joshi, M., Knutti, R., McNeall, D., Solomon, S., Sutton, R., Syktus, J., and Vecchi, G.: Uncertainties in the timing of unprecedented climates, *Nature*, 511, E3–E5, <https://doi.org/10.1038/nature13523>, 2014.
- Hellinger, E.: Neue Begründung der Theorie quadratischer Formen von unendlichvielen Veränderlichen, *J. Rein. Angew. Math.*, 136, 210–271, 1909.
- Hope, C. and Schaefer, K.: Economic impacts of carbon dioxide and methane released from thawing permafrost, *Nat. Clim. Change*, 6, 56–59, <https://doi.org/10.1038/nclimate2807>, 2016.
- Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Leetmaa, A., Reynolds, R., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Jenne, R., and Joseph, D.: The NCEP/NCAR 40-Year Reanalysis Project, *B. Am. Meteorol. Soc.*, 77, 437–471, [https://doi.org/10.1175/1520-0477\(1996\)077<0437:TNYRP>2.0.CO;2](https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2), 1996.
- Karlsson, J. M., Bring, A., Peterson, G. D., Gordon, L. J., and Destouni, G.: Opportunities and limitations to detect climate-related regime shifts in inland Arctic ecosystems through eco-hydrological monitoring, *Environ. Res. Lett.*, 6, 1–9, <https://doi.org/10.1088/1748-9326/6/1/014015>, 2011.
- Karlsson, J. M., Lyon, S. W., and Destouni, G.: Thermokarst lake, hydrological flow and water balance indicators of permafrost change in Western Siberia, *J. Hydrol.*, 464–465, 459–466, <https://doi.org/10.1016/j.jhydrol.2012.07.037>, 2012.
- Karoly, D. J. and Wu, Q.: Detection of Regional Surface Temperature Trends, *J. Climate*, 18, 4337–4343, <https://doi.org/10.1175/JCLI3565.1>, 2005.
- Khan, V., Holko, L., Rubinstein, K., and Breiling, M.: Snow Cover Characteristics over the Main Russian River Basins as Represented by Reanalyses and Measured Data, *J. Appl. Meteorol. Clim.*, 47, 1819–1833, <https://doi.org/10.1175/2007JAMC1626.1>, 2008.
- King, A. D., Donat, M. G., Fischer, E. M., Hawkins, E., Alexander, L. V., Karoly, D. J., Dittus, A. J., Lewis, S. C., and Perkins, S. E.: The timing of anthropogenic emergence in simulated climate extremes, *Environ. Res. Lett.*, 10, 094015, <https://doi.org/10.1088/1748-9326/10/9/094015>, 2015.
- Knutson, T. R., Zeng, F., and Wittenberg, A. T.: Multimodel assessment of regional surface temperature trends: CMIP3 and CMIP5 twentieth-century simulations, *J. Climate*, 26, 8709–8743, <https://doi.org/10.1175/JCLI-D-12-00567.1>, 2013.
- Knutti, R., Sedláček, J., Sanderson, B. M., Lorenz, R., Fischer, E. M., and Eyring, V.: A climate model projec-

- tion weighting scheme accounting for performance and interdependence, *Geophys. Res. Lett.*, 44, 1909–1918, <https://doi.org/10.1002/2016GL072012>, 2017.
- Koven, C. D., Riley, W. J., and Stern, A.: Analysis of permafrost thermal dynamics and response to climate change in the CMIP5 earth system models, *J. Climate*, 26, 1877–1900, <https://doi.org/10.1175/JCLI-D-12-00228.1>, 2013.
- 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, *J. Climate*, 30, 7739–7756, <https://doi.org/10.1175/JCLI-D-16-0792.1>, 2017.
- Leloup, J., Lengaigne, M., and Boulanger, J. P.: Twentieth century ENSO characteristics in the IPCC database, *Clim. Dynam.*, 30, 277–291, <https://doi.org/10.1007/s00382-007-0284-3>, 2008.
- Lyu, K., Zhang, X., Church, J. A., Slagen, A. B. A., and Hu, J.: Time of emergence for regional sea-level change, *Nat. Clim. Change*, 4, 1006–1010, <https://doi.org/10.1038/nclimate2397>, 2014.
- Mahlstein, I., Knutti, R., Solomon, S., and Portmann, R. W.: Early onset of significant local warming in low latitude countries, *Environ. Res. Lett.*, 6, 1–6, <https://doi.org/10.1088/1748-9326/6/3/034009>, 2011.
- Mahlstein, I., Hegerl, G., and Solomon, S.: Emerging local warming signals in observational data, *Geophys. Res. Lett.*, 39, L21711, <https://doi.org/10.1029/2012GL053952>, 2012.
- Maraun, D.: When will trends in European mean and heavy daily precipitation emerge?, *Environ. Res. Lett.*, 8, 1–7, <https://doi.org/10.1088/1748-9326/8/1/014004>, 2013.
- Mora, C., Frazier, A. G., Longman, R. J., Dacks, R. S., Walton, M. M., Tong, E. J., Sanchez, J. J., Kaiser, L. R., Stender, Y. O., Anderson, J. M., Ambrosino, C. M., Fernandez-Silva, I., Giusseffi, L. M., and Giambelluca, T. W.: The projected timing of climate departure from recent variability, *Nature*, 502, 183–187, <https://doi.org/10.1038/nature12540>, 2013.
- Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., and Veith, T. L.: Model evaluation guidelines for systematic quantification of accuracy in watershed simulations, *T. ASABE*, 50, 885–900, 2007.
- Nash, E. and Sutcliffe, V.: River flow forecasting through conceptual models Part I – A discussion of principles, *J. Hydrol.*, 10, 282–290, [https://doi.org/10.1016/0022-1694\(70\)90255-6](https://doi.org/10.1016/0022-1694(70)90255-6), 1970.
- New, M., Hulme, M., and Jones, P.: Representing Twentieth-Century Space–Time Climate Variability. Part II: Development of 1901–96 Monthly Grids of Terrestrial Surface Climate, *J. Climate*, 13, 2217–2238, [https://doi.org/10.1175/1520-0442\(2000\)013<2217:RTCSTC>2.0.CO;2](https://doi.org/10.1175/1520-0442(2000)013<2217:RTCSTC>2.0.CO;2), 2000.
- Pohl, E.: ToE\_tools: A non-parametric method to calculate the Time of Emergence of climate signals, available at: [https://github.com/pohleric/toe\\_tools](https://github.com/pohleric/toe_tools), last access: 19 May 2020.
- Prowse, T., Shrestha, R., Bonsal, B., and Dibike, Y.: Changing spring air-temperature gradients along large northern rivers: Implications for severity of river-ice floods, *Geophys. Res. Lett.*, 37, 1–6, <https://doi.org/10.1029/2010GL044878>, 2010.
- Romanovsky, V. E., Drozdov, D. S., Oberman, N. G., Malkova, G. V., Kholodov, A. L., Marchenko, S. S., Moskalenko, N. G., Sergeev, D. O., Ukraintseva, N. G., Abramov, A. A., Gilichinsky, D. A., and Vasiliev, A. A.: Thermal state of permafrost in Russia, *Permafrost Periglac. Process.*, 21, 136–155, <https://doi.org/10.1002/ppp.683>, 2010.
- Rust, H. W., Vrac, M., Lengaigne, M., and Sultan, B.: Quantifying Differences in Circulation Patterns Based on Probabilistic Models: IPCC AR4 Multimodel Comparison for the North Atlantic, *J. Climate*, 23, 6573–6589, <https://doi.org/10.1175/2010JCLI3432.1>, 2010.
- Scherer, M. and Diffenbaugh, N. S.: Transient twenty-first century changes in daily-scale temperature extremes in the United States, *Clim. Dynam.*, 42, 1383–1404, <https://doi.org/10.1007/s00382-013-1829-2>, 2014.
- Schuur, E. A. G., McGuire, A. D., Schädel, C., Grosse, G., Harden, J. W., Hayes, D. J., Hugelius, G., Koven, C. D., Kuhry, P., Lawrence, D. M., Natali, S. M., Olefeldt, D., Romanovsky, V. E., Schaefer, K., Turetsky, M. R., Treat, C. C., and Vonk, J. E.: Climate change and the permafrost carbon feedback, *Nature*, 520, 171–179, <https://doi.org/10.1038/nature14338>, 2015.
- Scott, D. W.: Multivariate density estimation: theory, practice, and visualization, John Wiley & Sons, Inc., Hoboken, New Jersey, <https://doi.org/10.1002/9781118575574>, 2015.
- Serreze, M. C. and Hurst, C. M.: Representation of Mean Arctic Precipitation from NCEP–NCAR and ERA Reanalyses, *J. Climate*, 13, 182–201, [https://doi.org/10.1175/1520-0442\(2000\)013<0182:ROMAPF>2.0.CO;2](https://doi.org/10.1175/1520-0442(2000)013<0182:ROMAPF>2.0.CO;2), 2000.
- Sui, Y., Lang, X., and Jiang, D.: Time of emergence of climate signals over China under the RCP4.5 scenario, *Climatic Change*, 125, 265–276, <https://doi.org/10.1007/s10584-014-1151-y>, 2014.
- Taylor, K. E., Stouffer, R. J., and Meehl, G. A.: An Overview of CMIP5 and the Experiment Design, *B. Am. Meteorol. Soc.*, 93, 485–498, <https://doi.org/10.1175/BAMS-D-11-00094.1>, 2012.
- Turlach, B. A.: Bandwidth selection in kernel density estimation: A review, in: CORE and Institut de Statistique, Université catholique de Louvain, Louvain-la-Neuve, Belgium, 1993.
- Ulrich, M., Matthes, H., Schirmer, L., Schütze, J., Park, H., Iijima, Y., and Fedorov, A. N.: Differences in behavior and distribution of permafrost-related lakes in Central Yakutia and their response to climatic drivers, *Water Resour. Res.*, 53, 1167–1188, <https://doi.org/10.1002/2016WR019267>, 2017.
- Vautard, R., Gobiet, A., Sobolowski, S., Kjellström, E., Stegehuis, A., Watkiss, P., Mendlik, T., Landgren, O., Nikulin, G., Teichmann, C., and Jacob, D.: The European climate under a 2 °C global warming, *Environ. Res. Lett.*, 9, 034006, <https://doi.org/10.1088/1748-9326/9/3/034006>, 2014.
- Vey, S., Steffen, H., Müller, J., and Boike, J.: Inter-annual water mass variations from GRACE in central Siberia, *J. Geod.*, 87, 287–299, <https://doi.org/10.1007/s00190-012-0597-9>, 2013.
- Viovy, N.: CRUNCEP Version 7 – Atmospheric Forcing Data for the Community Land Model, available at: <http://rda.ucar.edu/datasets/ds314.3/> (last access: 19 May 2020), 2018.
- Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Jarrod Millman, K., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, I., Feng, Y., Moore, E. W., Vander Plas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., and SciPy 1.0 Contributors: SciPy 1.0: Fundamental Algorithms for Scientific Computing in

- Python, Nat. Meth., <https://doi.org/10.1038/s41592-019-0686-2>, 2020.
- Walvoord, M. A. and Kurylyk, B. L.: Hydrologic Impacts of Thawing Permafrost – A Review, *Vadose Zone J.*, 15, 1–20, <https://doi.org/10.2136/vzj2016.01.0010>, 2016.
- Walvoord, M. A. and Striegl, R. G.: Increased groundwater to stream discharge from permafrost thawing in the Yukon River basin: Potential impacts on lateral export of carbon and nitrogen, *Geophys. Res. Lett.*, 34, L12402, <https://doi.org/10.1029/2007GL030216>, 2007.
- Westermann, S., Peter, M., Langer, M., Schwamborn, G., Schirmer, L., Eitzel, B., and Boike, J.: Transient modeling of the ground thermal conditions using satellite data in the Lena River delta, Siberia, *The Cryosphere*, 11, 1441–1463, <https://doi.org/10.5194/tc-11-1441-2017>, 2017.
- Yang, D., Kane, D. L., Hinzman, L. D., Zhang, X., Zhang, T., and Ye, H.: Siberian Lena River hydrologic regime and recent change, *J. Geophys. Res.-Atmos.*, 107, ACL 14-1–ACL 14-10, <https://doi.org/10.1029/2002JD002542>, 2002.