

Journal Pre-proof

A dataset of large ensemble of CMIP6-based transient climate scenarios for impact assessment in Great Britain

Mikhail A. Semenov , Nimai Senapati , Kevin Coleman ,
Adrian L. Collins

PII: S2352-3409(25)00425-1
DOI: <https://doi.org/10.1016/j.dib.2025.111695>
Reference: DIB 111695



To appear in: *Data in Brief*

Received date: 8 January 2025
Revised date: 13 May 2025
Accepted date: 15 May 2025

Please cite this article as: Mikhail A. Semenov , Nimai Senapati , Kevin Coleman ,
Adrian L. Collins , A dataset of large ensemble of CMIP6-based transient climate scenarios for impact
assessment in Great Britain, *Data in Brief* (2025), doi: <https://doi.org/10.1016/j.dib.2025.111695>

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2025 Published by Elsevier Inc.

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

ARTICLE INFORMATION

Article title

A dataset of large ensemble of CMIP6-based transient climate scenarios for impact assessment in Great Britain

Authors

Mikhail A. Semenov ^{a*}, Nimai Senapati ^{a*}, Kevin Coleman ^b, Adrian L. Collins ^c

Affiliations

^a Sustainable Soils and Crops, Rothamsted Research, West Common, Harpenden, AL5 2JQ, United Kingdom.

^b Net Zero and Resilient Farming, Rothamsted Research, West Common, Harpenden, AL5 2JQ, United Kingdom.

^c Net Zero and Resilient Farming, Rothamsted Research, North Wyke, Okehampton, Devon EX20 2SB, United Kingdom.

Corresponding author's email address and Twitter handle

mikhail.semenov@rothamsted.ac.uk; nimai.senapati@rothamsted.ac.uk; @nimai_senapati

Keywords

Climate change; transient climate scenarios; CMIP6 ensemble; LARS-WG weather generator; climate change impact assessment; downscaling; SSPs

Abstract

Under the extant threat of climate change, impact assessment studies are essential to investigate and quantify the severity of the potential impacts, and support recommendations for mitigation strategies with foresight. Future climate change scenarios are therefore crucial for underpinning impact studies. Here, transient climate scenarios are important as they provide a more realistic and dynamic evolution of future climate conditions over time, rather than only static climate scenarios. It is also important to downscale climate projection of Global Climate Models (GCMs) from coarse spatial and temporal resolution to local scale site-specific daily climate scenarios which have a sufficiently large number of years or realisations for accounting for inter-annual variability and detecting rare extreme climatic events. In the new dataset presented herein, transient future climate scenarios were generated at 26 representative sites across Great Britain (GB) using the Long Ashton Research Station Weather Generator (LARS-WG 8.0), based on climate projections from a subset of five GCMs from the latest Coupled Model Intercomparison Project Phase 6 (CMIP6) ensemble and two emission scenarios. For each site, 100 realisations of continuous transient time series of daily weather (minimum air temperature, maximum air temperature, rainfall and solar radiation) over the period 2020 to 2090 were generated. The use of a subset of five GCMs reduces computational requirements substantially

for impact assessments, while allowing quantification of uncertainties in impacts related to uncertainty in future climate projections arising from GCMs. The dataset can be used to underpin assessments of future climate change risk and vulnerability, and their temporal patterns and progressive changes over time. Our data are designed to be used as a time series of climatic input to impact models for climate change assessments continuously over time related to various fields and disciplines, including land and water resources, agriculture and food production, soil carbon cycle, ecology and epidemiology, and human health and welfare. Various key stakeholders, such as researchers, breeders, farm managers, social and public sector advisers, policymakers and planners, may benefit from this new transient dataset for investigating, forecasting, designing and prioritising adaptive and mitigation strategies under changing climate.

SPECIFICATIONS TABLE

Subject	<u>Earth & Environmental Sciences</u>
Specific subject area	Climate change; transient climate scenarios; downscaling of climate data from GCM and generating daily climate scenarios at local scale
Type of data	Table (.dat file and .st file) Raw model generated data (txt files)
Data collection	<p>CMIP6 ensemble data: Future climate projections from the CMIP6 ensemble were obtained from the Copernicus Climate Data Store. https://cds.climate.copernicus.eu/cdsapp#!/dataset/projections-cmip6?tab=form</p> <p>Observed historical weather: Observed historical daily weather data at 26 representative sites across GB was obtained from the UK Met Office [1].</p> <p>Baseline and transient future climate scenarios: One hundred yearly realisations of daily weather (minimum air temperature, maximum air temperature, rainfall and solar radiation) for baseline, and one hundred realisations of transient daily climate scenarios at the 26 selected sites were generated for the period 2020-2090 using a stochastic weather generator (LARS-WG 8.0) [2].</p>
Data source location	<p>Primary data sources:</p> <p>(i) Observed historical climate data: UK Met Office [1].</p> <p>(ii) Future climate projections: CMIP6 ensemble, Copernicus Climate Data Store. https://cds.climate.copernicus.eu/cdsapp#!/dataset/projections-cmip6?tab=form</p>
Data accessibility	<p>The dataset [3] is available from -</p> <p>Repository name: Zenodo</p>

	Data identification number: https://doi.org/10.5281/zenodo.14040993 Direct URL to data: https://zenodo.org/records/14040993
Related research article	None

VALUE OF THE DATA

- The dataset [3] provides continuous transient future climate scenarios on daily time scale with sufficiently large numbers of realisations at local scale across GB, based on climate projections from a subset of five GCMs from the CMIP6 ensemble and two emission scenarios (SSP2-4.5 and SSP5-8.5). Each transient scenario for each site consists of 100 realisations of daily weather time series during the period 2020 to 2090. This large ensemble of CMIP6-based transient climate scenarios provides essential data, such as daily minimum and maximum air temperatures, rainfall and solar radiation, for climate change impact studies at local scale across GB for understanding the dynamic evolution of the impacts over time due to the time-dependent nature of these scenarios.
- The dataset could underpin assessments of future climate change risk, particularly their temporal patterns and progressive changes over time, for example risk and vulnerability of natural ecosystems, agroecosystems, water resources and infrastructure systems to heat waves, floods and droughts.
- The sufficiently large number of realisations in this new transient climate dataset could enable accounting of inter-annual variability and detection of rare extreme climatic events within various time scales (e.g., day, week, month, year, etc.) during 2020-2090.
- The use of a subset of GCMs reduces computational requirements substantially for impact assessments, while allowing explicit estimation of uncertainty propagation in impact assessments due to inherent uncertainties in GCMs.
- This new dataset can be used as a time series of climatic input to impact models for climate change assessments continuously over time in various fields, including land and water resources, agriculture and food production, soil carbon cycle, ecology and epidemiology, and human health and welfare. This dataset could help in explicit investigating and quantification of climate change legacy effects and non-linear feedback mechanisms over time.

- Various stakeholders, such as researchers, breeders, farm managers, social and public sector advisers, policymakers and planners, could benefit from this new transient climate dataset for designing and prioritising adaptive and mitigation strategies depending on the temporal patterns of impacts.

BACKGROUND

Climate change is the single most critical issue of the 21st century [4]. Under ongoing climate change, impact studies are essential for assessing the severity of the potential impacts and supporting recommendations for mitigation strategies in advance. Future climate change scenarios are therefore crucial as an input for impact studies. Static future climate scenarios for a specific period, for example very near (e.g., 2030) or near future (2050), are important to estimate the climate change impacts at a particular point of time in details, for example climate scenarios as in Semenov et al. [5]. However, time-dependent climate change scenarios, i.e. climate time series or so-called transient climate scenarios, are critically important to investigate temporal patterns of the impacts, since they provide a more realistic and dynamic evolution of future climate conditions over time [6,7], rather than only static scenarios.

Climate change impact assessments often require a large ensemble of transient climate scenarios downscaled to local-scale, where each ensemble member represents plausible long weather series. A sufficiently large number of realisations is also needed for accounting for inter-annual variability and detecting rare extreme climatic events [5]. However, only very few realisations of transient climates are usually available for each GCM due to the high computational costs. Here, a potential viable solution can involve employing a stochastic weather generator [2] to downscale GCMs projections at local scale and generate a large ensemble of transient daily site-specific climate scenarios with a sufficiently large number of realisations based on a minimum workable subset of GCMs, covering the entire distribution and sensitivity of climate projections from GCMs ensemble. Such datasets are currently unavailable for GB. To address such a need, a large ensemble of the latest CMIP6-based [8] daily transient climate scenarios was generated for impact assessments in GB.

DATA DESCRIPTION

This dataset [3] provides future continuous transient climate scenarios on daily time scale with sufficiently large numbers of realisations at local scale across GB, based on climate projections from the latest CMIP6 ensemble. The dataset consists of 100 realisations of continuous transient time series of daily weather over the period 2020 to 2090 at 26 representative sites across GB, based on a subset of five GCMs (ACCESS-ESM1-5, CNRM-CM6-1, HadGEM3-GC31-LL, MPI-ESM1-2-LR, and MRI-ESM2-0) selected from CMIP6 ensemble and two emission scenarios (SSP2-4.5 and SSP5-8.5). The dataset is

deposited as a zip file “RFF.GB26.100TR.ZIP” in Zenodo, with notations indicating the name of the research project (Resilient Farming Futures: RFF), Great Britain (GB), 26 representative sites, 100 realisations and transient (TR) climate scenarios (Fig. 1). RFF.GB26.100TR.ZIP consists of 26 zipped folders for the 26 representative sites across GB. The file name for each site starts with the site abbreviation followed by “5GCM100TR”, indicating five selected GCMs and 100 realisations of transient climate scenarios (Fig. 1). Each site folder includes a total of 1002 files; viz., 1000 transient future climate scenarios over the time span 2020-2090 (5-GCMs \times 2-SSPs \times 100 realisations) [.dat files], one baseline climate [.dat file], and one site meta-file [.st] (Fig. 1 and 2). The name of the future climate scenarios file (.dat) begins with the site abbreviation, followed by the name of the GCM, SSP, and the realisation number from 1 to 100. The name of the baseline climate file (.dat) consists of the site abbreviation followed by “WG”, meaning baseline generated by LARS-WG. The baseline climate at each site contains 100 realisations of daily weather for a typical year representing the observed period 1985-2015. The site meta file name starts with the site abbreviation followed by the abbreviation “WG”, meaning generated by LARS-WG. The site file contains information about the site (latitude, longitude and altitude), and climate scenarios (format, name and sequence of climatic variables). The baseline and each transient climate scenario contain four variables, viz., daily minimum air temperature, maximum air temperature, rainfall and solar radiation. The name, abbreviation, and units of the climatic variables in the dataset are described in Table 1. Fig. 3 and 4 show an example of a baseline and transient climate scenario for a single weather variable (maximum air temperature) at the Rothamsted (RR) site.

Table 1. Climatic variables, abbreviations and units in the transient climate dataset.

Variable	Year	Day of year	Minimum air temperature	Maximum air temperature	Rainfall	Solar radiation
Abbreviation	YEAR	JDAY	MIN	MAX	RAIN	RAD
Unit	N/A	N/A	°C	°C	mm day ⁻¹	MJ m ⁻² day ⁻¹

EXPERIMENTAL DESIGN, MATERIALS AND METHODS

Sites selection across the GB. For this transient climate dataset [3], 26 representative sites across GB (Fig.5) were selected from an available 85 climate stations within the Met Office network [1]. These sites were selected based on: (1) the missing data in observations should be <10% for temperature and precipitation, and (2) sites should be representative of diverse locations and climates across GB, providing broad and even coverage of its arable land [5]. According to the Köppen-Geiger climate classification for GB, climate predominantly falls into the Cfb climate type, which is a warm temperate climate with a maritime influence. This indicates GB experiences relatively mild winters and cool summers with no significant dry season and at least 75% of the precipitation occurring evenly

throughout the year. While Cfb is the dominant climate, some areas, particularly in the south and southeast, might see shifts towards more Mediterranean-like climates (Csa or Cs) due to increasing temperatures.

Baseline climate. The daily observed weather during the period 1985-2015 are available at each selected site [1,5]. To generate a baseline at each site, observed weather data were used to estimate site-specific climatic parameters required by a stochastic weather generator (LARS-WG 8.0) [2]. Adequate reproduction of climatic variability and change by LARS-WG, including extreme climatic events at local scale, have been reported by various previous studies [5, 9-12]. To generate a reasonably large number of years of the baseline, for accounting for inter-annual variability and detecting climatic extreme events, 100 years of daily weather representing typical years for the period 1985-2015 at each site were generated using LARS-WG, based on estimated site-specific parameters; hereafter, defined as the 'baseline climate'. The baseline climate has statistical characteristics similar to the observed weather at each site, with probability distributions close to those of the observed climate. LARS-WG automatically computes various statistics, including KS-test statistics, t-statistics and f-statistics, and their corresponding p-values which allows to compare observed and generated weather [2].

Future transient climate scenarios based on CMIP6 ensemble. Future transient climate scenarios were based on five GCMs from the latest CMIP6 ensemble [8]. A total of 15 GCMs from the CMIP6 ensemble were integrated into LARS-WG 8.0 [2]. Based on their performance over northern Europe including the UK, climate sensitivity and the distribution of GCMs, a subset of the best five performing GCMs (Table 2) was selected for the present dataset [5,13]. The selected five GCMs are evenly distributed among the 15 GCMs, capturing uncertainty in climate projections from the CMIP6 ensemble [5]. The use of a subset of five GCMs substantially reduces computational time for impact assessment studies, while allowing explicit estimation of uncertainties in impacts related to uncertainty in future climate projections due to the GCMs. Two future emission scenarios, defined as Shared Socioeconomic Pathways (SSPs), were selected for this transient climate dataset to cover the range of possible future development of anthropogenic drivers of climate change; viz., (i) an intermediate GHG emission scenario: SSP2-4.5 – Middle of the Road (medium challenges to mitigation and adaptation), with an additional radiative forcing of 4.5 W m^{-2} by the year 2100, and (ii) a very high GHG emission scenario: SSP5-8.5 – Fossil-fuelled Development – Taking the Highway (high challenges to mitigation, low challenges to adaptation), with an additional radiative forcing of 8.5 W m^{-2} by the year 2100 [14]. LARS-WG downscales climate projections from the GCMs and incorporates changes at local scale in the mean climate, climatic variability and extreme events derived from the GCMs by

modifying the statistical distributions of the weather variables at each site with delta-changes in climatic variables derived from the GCMs [2,10,15]. The monthly outputs from GCMs were used to calculate delta-changes which were used to perturb site-specific parameter distributions generated by LARS-WG. For each site, GCM and SSP, 100 realisations of transient daily weather for the period 2020-2090 were generated using LARS-WG 8.0, based on climate projections from the five selected GCMs.

Table 2 The five Global Climate Models (GCMs) from the Coupled Model Intercomparison Project Phase 6 (CMIP6) multi-model ensemble used in the construction of the transient climate scenarios for Great Britain.

GCM	Country of origin	Research centre	Grid resolution: latitude x longitude	Reference
ACCESS-ESM1-5	Australia	Commonwealth Scientific and Industrial Research Organisation (CSIRO)	1.25° x 1.875°	[16]
CNRM-CM6-1	France	Centre National de Recherches Meteorologiques (CNRM), Centre Europeen de Recherche et de Formation Avancee en Calcul Scientifique (CERFACS)	1.40° x 1.406°	[17]
HadGEM3-GC31-LL	UK	UK Met Office Hadley Centre (MOHC)	1.25° x 1.88°	[18]
MPI-ESM1-2-LR	Germany	Max Planck Institute for Meteorology (MPI-M)	1.39° x 1.41°	[19]
MRI-ESM2-0	Japan	Meteorological Research Institute (MRI)	1.113° x 1.125°	[20]

LIMITATIONS

The potential uncertainty in transient climate scenarios due to inherent uncertainties in GCMs was captured by using an ensemble of 5 GCMs. However, only one weather generator, LARS-WG, was used. It might be beneficial to quantify uncertainty related to the use of different weather generators for construction of transient climate scenarios. This will allow to estimate overall uncertainty from both sources in future transient climate scenarios for impact assessment.

ETHICS STATEMENT

The authors have read and follow the ethical requirements for publication in Data in Brief and confirming that the current work does not involve human subjects, animal experiments, or any data collected from social media platforms.

CRedit AUTHOR STATEMENT

Mikhail A. Semenov: Conceptualization, Methodology, Software, Validation and Data Curation. **Nimai Senapati:** Conceptualization, Writing - original draft, Writing - review & editing. **Kevin Coleman:** Conceptualization, Writing – review & editing. **Adrian L. Collins:** Conceptualization, Writing – review & editing.

ACKNOWLEDGEMENTS

We acknowledge the CMIP6 community for providing the climate model data, retained and globally distributed in the framework of the Earth System Grid Federation (ESGF). Rothamsted Research receives strategic funding from the UK Research and Innovation-Biotechnology and Biological Sciences Research Council (UKRI-BBSRC) through the Resilient Farming Futures institute strategic programme [BB/X010961/1, BBS/E/RH/230004A] and the Delivering Sustainable Wheat strategic programme [BB/X011003/1].

DECLARATION OF COMPETING INTERESTS

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES

- [1] Met Office, The UK climate - Synoptic and climate stations. UK. 2023. <https://www.metoffice.gov.uk/research/climate/maps-and-data/uk-synoptic-and-climate-stations>
- [2] Semenov, M.A. LARS-WG stochastic weather generator for climate change impact assessment (LARS-WG 8.0), Zenodo, 2024. <https://doi.org/10.5281/zenodo.10632391>.
- [3] M.A. Semenov, N. Senapati, A.L. Collins. A large ensemble of CMIP6-based transient climate scenarios for impact assessment in Great Britain. (1.1) [Data set], Zenodo, 2024. <https://doi.org/10.5281/zenodo.14040993>
- [4] IPCC. Summary for Policymakers. In: Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (Eds.)], Geneva, Switzerland, 2023, pp. 1-34. doi: 10.59327/IPCC/AR6-9789291691647.001.
- [5] M.A. Semenov, N. Senapati, K. Coleman, A.L. Collins. A dataset of CMIP6-based climate scenarios for climate change impact assessment in Great Britain. Data in Brief, 55 (2024), 110709. <https://doi.org/https://doi.org/10.1016/j.dib.2024.110709>

- [6] A.S. Meyer, et al. Temporal dynamics of climate change exposure and opportunities for global marine biodiversity. *Nat. Commun.* 15 (2024), 5836. <https://doi.org/10.1038/s41467-024-49736-6>
- [7] M. Mudelsee. Trend analysis of climate time series: A review of methods. *Earth-Sci. Rev.* 190 (2019) 310-322. <https://doi.org/https://doi.org/10.1016/j.earscirev.2018.12.005>
- [8] V. Eyring, et al. Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.* 9 (2016), 1937-1958. <https://doi.org/https://doi.org/10.5194/gmd-9-1937-2016>
- [9] M.W. Gitau, S. Mehan, T. Guo. Weather generator effectiveness in capturing climate extremes. *Environ. Process.* 5 (2018), S153-S165. <https://doi.org/https://doi.org/10.1007/s40710-018-0291-x>
- [10] M.A. Semenov, et al. ELPIS: a dataset of local-scale daily climate scenarios for Europe. *Clim. Res.* 44 (2010), 3-15. <https://doi.org/https://doi.org/10.3354/cr00865>
- [11] M.A. Semenov. Simulation of extreme weather events by a stochastic weather generator. *Clim. Res.* 35 (2008), 203-212. <https://doi.org/https://doi.org/10.3354/cr00731>
- [12] M.A. Semenov, R.J. Brooks, E.M. Barrow, C.W. Richardson. Comparison of the WGEN and LARS-WG stochastic weather generators for diverse climates. *Clim. Res.* 10 (1998), 95-107. <https://doi.org/10.3354/cr010095>
- [13] C. Bradshaw, A. Cottrell, T. Crocker, P. Falloon. Assessing CMIP6 climate model performance over the UK to support Resilient Farming Futures. Met Office Hadley Centre, UK (2023).
- [14] K. Riahi, et al. The Shared Socioeconomic Pathways and their energy, land use, and greenhouse gas emissions implications: An overview. *Glob. Environ. Change.* 42 (2017), 153-168. <https://doi.org/https://doi.org/10.1016/j.gloenvcha.2016.05.009>
- [15] M.A. Semenov, E.M. Barrow. Use of a stochastic weather generator in the development of climate change scenarios. *Clim. Change*, 35 (1997), 397-414. <https://doi.org/10.1023/a:1005342632279>
- [16] T. Ziehn, et al. CSIRO ACCESS-ESM1.5 model output prepared for CMIP6 CMIP. Earth System Grid Federation (2019). <https://doi.org/10.22033/ESGF/CMIP6.2288>
- [17] A. Voldoire. CNRM-CERFACS CNRM-CM6-1 model output prepared for CMIP6 CMIP. Earth System Grid Federation (2018). <https://doi.org/10.22033/ESGF/CMIP6.1375>
- [18] M. Roberts. MOHC HadGEM3-GC31-LL model output prepared for CMIP6 HighResMIP. Earth System Grid Federation (2017). <https://doi.org/10.22033/ESGF/CMIP6.1901>

- [19] K.-H. Wieners, et al. MPI-M MPI-ESM1.2-LR model output prepared for CMIP6 ScenarioMIP ssp245. Earth System Grid Federation (2019). <https://doi.org/10.22033/ESGF/CMIP6.6693>
- [20] S. Yukimoto, et al. MRI MRI-ESM2.0 model output prepared for CMIP6 CMIP. Earth System Grid Federation (2019). <https://doi.org/10.22033/ESGF/CMIP6.621>

Figures

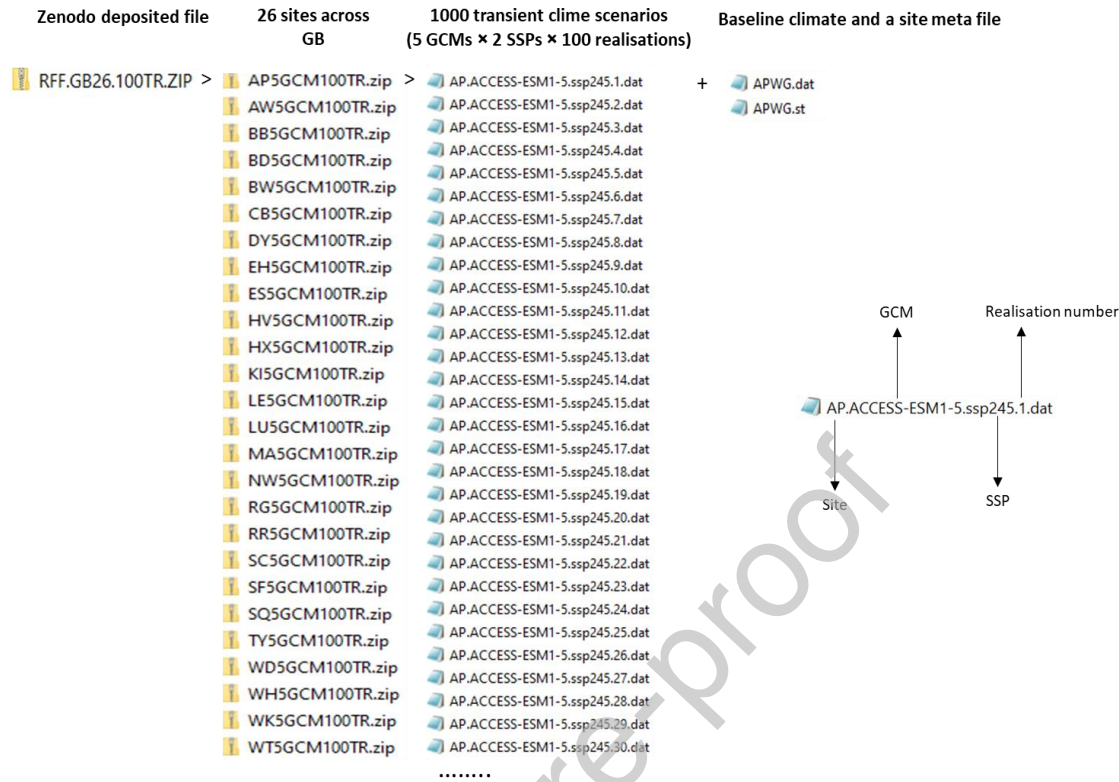


Fig 1. Description of the transient climate dataset [3].

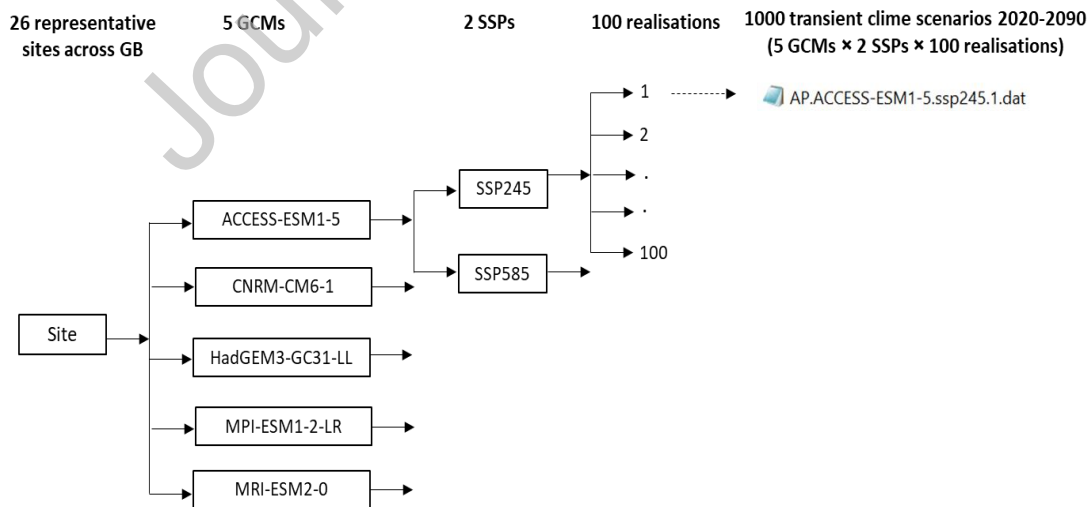


Fig 2. The structure of the transient climate scenarios spanning 2020 to 2090 at a representative site in the transient climate dataset [3].

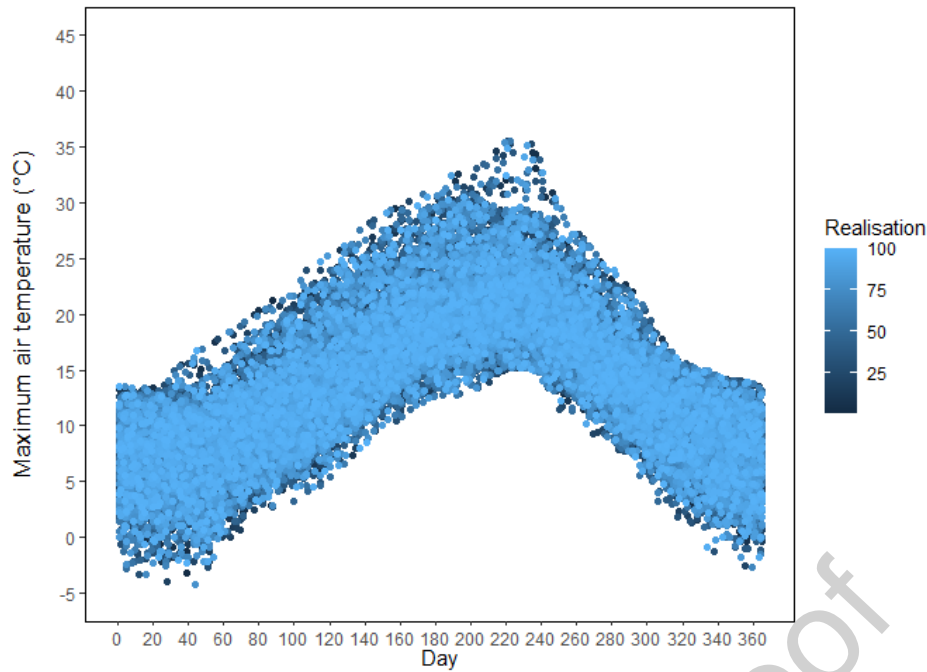


Fig. 3. An example of a baseline climate at the Rothamsted (RR) site, representing 100 realisations of one climatic variable, i.e., daily maximum air temperature ($^{\circ}\text{C}$), in a typical year representing the observed period 1985-2015 [data from the baseline climate file RRWG.dat inside the site folder RR5GCM100TR.zip].

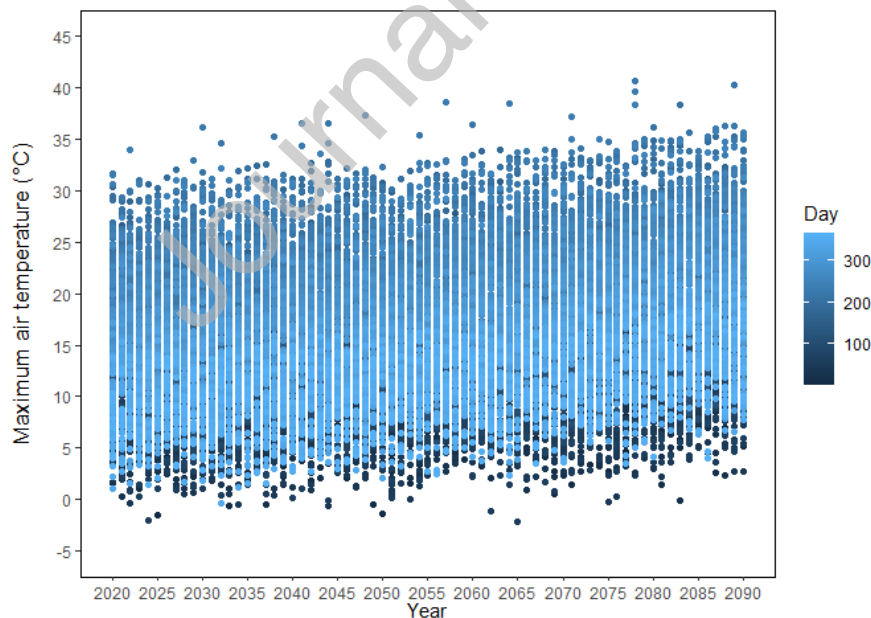


Fig. 4. An example of a transient climate scenario at the Rothamsted (RR) site, representing the first realisation of one climatic variable, i.e., daily maximum air temperature ($^{\circ}\text{C}$), during 2020 to 20290, based on one GCM (HadGEM3-GC31-LL) and one emission scenario (SSP5-8.5) [data from the climate scenario file RR.HadGEM3-GC31-LL.ssp585.1.dat inside the site folder RR5GCM100TR.zip].

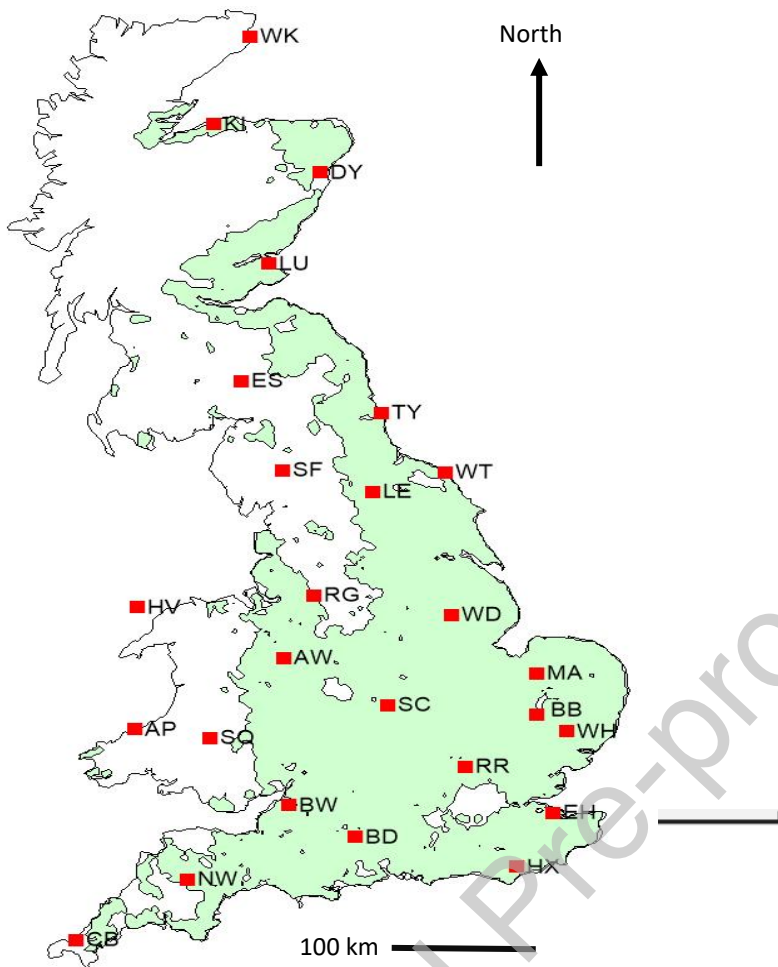


Fig.5. The selected representative sites in the transient climate dataset across Great Britain [3]. The green shading represents arable land. Please see details of the site information in Semenov et al. [5].