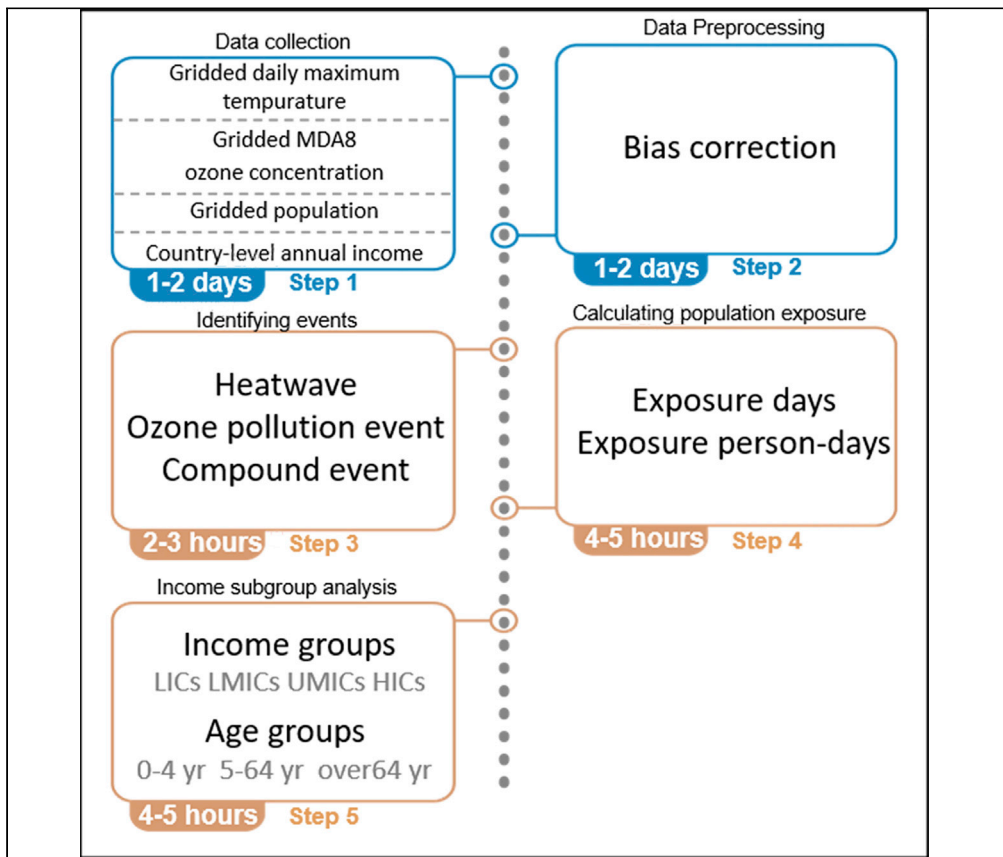## Protocol

Protocol for estimating exposure to compound heat wave and ozone pollution under future climate change

Kailai Lu, Jie Ban, Qing Wang, Tiantian Li

litiantian@nieh.chinacdc.cn

Highlights

We perform bias correction for the modeled temperature and ozone concentration

We project both compound days occurrence and population exposure under SSPs

We identify disparity in compound exposure among different income groups

We characterize the proportion of population exposed under high-level compound days

Here we describe the procedure for estimating exposure to the compound heatwave and ozone pollution under future climate scenarios. We first apply the daily-level temperature and ozone concentration across the world and perform bias correction by comparing the distribution of the modeled temperature and ozone concentration to the distribution of historical observation. Then we identify the heatwaves, ozone pollution events, and compound events. Finally, we combine the future exposure and population to identify the high-risk regions and populations.

Publisher's note: Undertaking any experimental protocol requires adherence to local institutional guidelines for laboratory safety and ethics.

Protocol

Protocol for estimating exposure to compound heat wave and ozone pollution under future climate change

Kailai Lu,[1,3,4] Jie Ban,[1,2,3,4] Qing Wang,[1] and Tiantian Li[1,5,*]

[1]China CDC Key Laboratory of Environment and Population Health, National Institute of Environmental Health, Chinese Center for Disease Control and Prevention, Beijing 100021, China

[2]Collaborative Innovation Center of Atmospheric Environment and Equipment Technology, Jiangsu Key Laboratory of Atmospheric Environment Monitoring and Pollution Control (AEMPC), Nanjing University of Information Science & Technology, Nanjing 210044, China

[3]These authors contributed equally

[4]Technical contact: rikuya1@outlook.com; banjie@nieh.chinacdc.cn

[5]Lead contact

*Correspondence: litiantian@nieh.chinacdc.cn
https://doi.org/10.1016/j.xpro.2023.102090

SUMMARY

Here we describe the procedure for estimating exposure to the compound heatwave and ozone pollution under future climate scenarios. We first apply the daily-level temperature and ozone concentration across the world and perform bias correction by comparing the distribution of the modeled temperature and ozone concentration to the distribution of historical observation. Then we identify the heatwaves, ozone pollution events, and compound events. Finally, we combine the future exposure and population to identify the high-risk regions and populations.
For complete details on the use and execution of this protocol, please refer to Ban et al. (2022).[1]

BEFORE YOU BEGIN

The protocol given below describes the overall design of the study, the specific steps present how to identify compound heatwave and ozone pollution events and calculate exposure days and exposure person-days.

Study design

This study considers 1995–2014 as the baseline period following the most recently used datasets and defines three future mid- and long-term projection periods, namely, the 2040s (2031–2050), the 2060s (2051–2070), and the 2080s (2071–2090). We adopt four different Shared Socioeconomic Pathways (SSP)-Representative Concentration Pathways (RCP) scenarios (SSP1-2.6, SSP2-4.5, SSP3-7.0, and SSP5-8.5), and select three defined future periods to investigate the changes in exposure to compound extremes under climate change. We estimate the number of event days and number of person-days of exposure with respect to compound events, ozone events, and heatwaves. We divide the countries into different groups based on their income levels, and then perform a comparative analysis to explore the disparities in exposure to compound events among countries.

Data collection

⏲ Timing: 1–2 days depending on file sizes

1. Temperature and Ozone Simulation Data.

   a. Set the data inclusion criteria for the climate simulation data, including daily maximum temperature and ozone concentration, as follows:

     i. The simulated data should be global.

     ii. Include the daily values for maximum temperature data, and the hourly values for ozone concentration data.

     iii. The simulation data needs include the historical simulations from 1995 to 2014, and the future predictions under four scenarios (SSP1-2.6, SSP2-4.5, SSP3-7.0, SSP5-8.5) from 2015 to 2090.

   b. Download the qualified datasets from the United Kingdom's Earth System Model (UKESM1) of CMIP6 (Database: https://esgf-node.llnl.gov/search/cmip6/). The resolution of this model is 1.25° × 1.875° (latitude × longitude).[2]

     i. Establish a grid with the same resolution as the UKESM1 model and extract the global land climate simulation data by using the longitude and latitude of the grid. The simulated data should be global.

     ii. Convert the hourly ozone into the daily maximum 8-h average ozone concentration.

     iii. For example, one of the datasets we download named 'tasmax_day_UKESM1-0-LL_ssp245_r1i1p1f2_gn_20150101-20491230.nc', it means this NC file contains the daily data of global daily maximum temperature under SSP2-4.5 scenarios from 2015 to 2049, and the file is obtained from UKESM1-0-LL model. The file name of hourly ozone concentration is like 'sfo3_AERhr_UKESM1-0-LL_ssp126_r1i1p1f2_gn_203001010030-203912302330.nc'.

2. Temperature and Ozone Observational Data.

   a. Download the observed global temperature data (1995–2014) from the ERA-Interim reanalysis database and apply zonal statistics to fit it into the same spatial resolution as the simulated data from UKESM1. This method aims to correct the daily maximum temperature simulation data.[3]

   b. For ozone concentrations, download the worldwide ozone-monitoring dataset from the Tropospheric Ozone Assessment Report (TOAR).[4] Although the dataset does not fully cover the present study areas, it is the best available observational data.

3. Population Data.

   a. Download the baseline and future SSP population data at 1-km resolution from NASA's Socioeconomic Data and Applications Center (SEDAC). We use population in the year 2000 as the baseline population and use the data in 2040, 2060, and 2080 under SSP1, SSP2, SSP3, and SSP5 as future population.[5]

   b. Apply the rasterstats library of Python 3.7 to count the population under each gridded climate dataset.

   c. Use the country-level age-specific population data provided in the SSP database (SSP Scenarios Population Data: https://secure.iiasa.ac.at/web-apps/ene/SspDb/dsd?Action=html page&page=10) to calculate the proportion of age-specific population (0–4, 5–64, and over 65 years old) in different countries, and then calculate the amount of age-specific population for each grid cell by combing country-level age proportion and gridded total population.

4. Gross National Income Data.

   a. To analyze the inequality among countries, download the latest per capita gross national income data from the World Bank website, and classify countries into low-income countries (LICs), lower-middle-income countries (LMICs), upper-middle-income countries (UMICs), and high-income countries (HICs).

   b. Assign the income level of the country to each grid cell such that it fully covers the grid cell. For the grid cells covered by different countries (such as the grid cells located at the boundary of different countries), assign the income level of the country with the largest area within the grid cell.

Data preprocessing

⏱ Timing: 1–2 days depending on file sizes and computer performance

5. Bias Correction.
   a. For temperature correction, refer to the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP) method for bias correction[6]:
      i. Calculate the monthly mean daily maximum temperatures (both observed and simulated) from January to December for each baseline period (1995–2014) under the grid. Then, correct the overall bias of the model by comparing the differences between the observed and simulated values for each month. Here, we take 01/15/2013 as an example. In this step, we calculate the monthly average temperature of all the January during 1995–2014 from both observational data and simulation data, respectively named Tmaxmean_Jan_Obs and Tmaxmean_Jan_Sim. Their difference (Δ Tmaxmean_Jan) will be used for the first correction of all January data.
      ii. Calculate the temperature difference (ΔTmax) between the daily maximum temperature and the monthly average temperature of each grid during the baseline period, and use the QMap method for mapping the daily ΔTmax of the simulated values in each month to the ΔTmax of the observed values, and correct the dispersion of the simulated data. Continue to take 01/15/2013 as the example, in this step, first, stat the daily maximum temperature of every January days from the observed and simulated values(named Tmax_Obs and Tmax_Sim), here, the simulated value of 01/15/2013 is named Tmax_Sim_20130115. Second, calculate the difference between them and the monthly mean value for the month (named ΔT_Obs and ΔT_Sim) respectively. And the difference between Tmax_Sim_20130115 and the monthly mean value in January 2013(Tmaxmean_JAN_2013_Sim) is namedΔT_Sim_20130115. Third, look for the position of ΔT_Sim_20130115 among all ΔT_Sim, assume that it ranks in the 60th percentile. Then find the value at the 60th percentile of ΔT_Obs, substitute it for ΔT_Sim_20130115, and add it to Tmaxmean_JAN_2013_Sim and Δ Tmaxmean_Jan, we get the corrected Tmax_Sim_20130115.
   b. For ozone concentration correction, perform bias correction based on the partial ISIMIP method:
      i. Calculate the monthly mean maximum daily 8-h average ozone simulated concentrations from January to December for each baseline period (1995–2014) under the grid and make it consistent with the missing spatiotemporal observations.
      ii. Correct the overall bias of the model by comparing the differences between the observed and simulated values for each month. Here we only perform bias correction for regions with observations.
   c. Use Taylor plots to demonstrate the effect of bias correction.
   d. Compared to the original output of the GCM model, the bias-corrected temperature changes in each grid cell across the globe, as shown in Figure 1.

> ⚠ CRITICAL: The observation data and simulation data need to be strictly corresponded in space and time. At the same time, this step requires a lot of computing resources, so we recommend using multithreaded code.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER*
Deposited data		
Temperature and Ozone Simulation Data	https://esgf-node.llnl.gov/search/cmip6/	N/A
Temperature Observation Data	https://data.isimip.org/	N/A
Ozone Observation Data	https://doi.pangaea.de/10.1594/PANGAEA.876108	N/A
Total Population Data	https://sedac.ciesin.columbia.edu/data/set/popdynamics-1-km-downscaled-pop-base-year-projection-ssp-2000-2100-rev01	N/A

(Continued on next page)

REAGENT or RESOURCE	SOURCE	IDENTIFIER*
Age-specific Population Data	https://secure.iiasa.ac.at/web-apps/ene/SspDb/dsd?Action=htmlpage&page=10	N/A
Gross National Data	https://www.worldbank.org/en/home	N/A
Software and algorithms		
R v.4.0.1	https://www.R-project.org/	N/A
Python 3.7	https://www.python.org/	N/A

*N/A: not available.

STEP-BY-STEP METHOD DETAILS

In this section, firstly, we illustrate the method of selection of thresholds and define the heatwaves and ozone pollution events, based on which we explain the method for identifying compound events. Then, we determine the exposure to the three types of events in terms of event days and person-days. Finally, we carry out regional statistical analysis for different countries and age subgroups.

Note: We perform all the steps in the statistical computing environment "Python".

Identifying events

⏱ Timing: 2–3 h depending on file sizes

This step uses Python to determine the temperature threshold based on the historical temperature data file and identify the heatwaves. In addition, this step uses Python to identify the ozone pollution events based on the WHO air quality guidelines for ozone. Based on these two types of events, identify the compound events as the overlapping days between heatwaves and ozone pollution events.
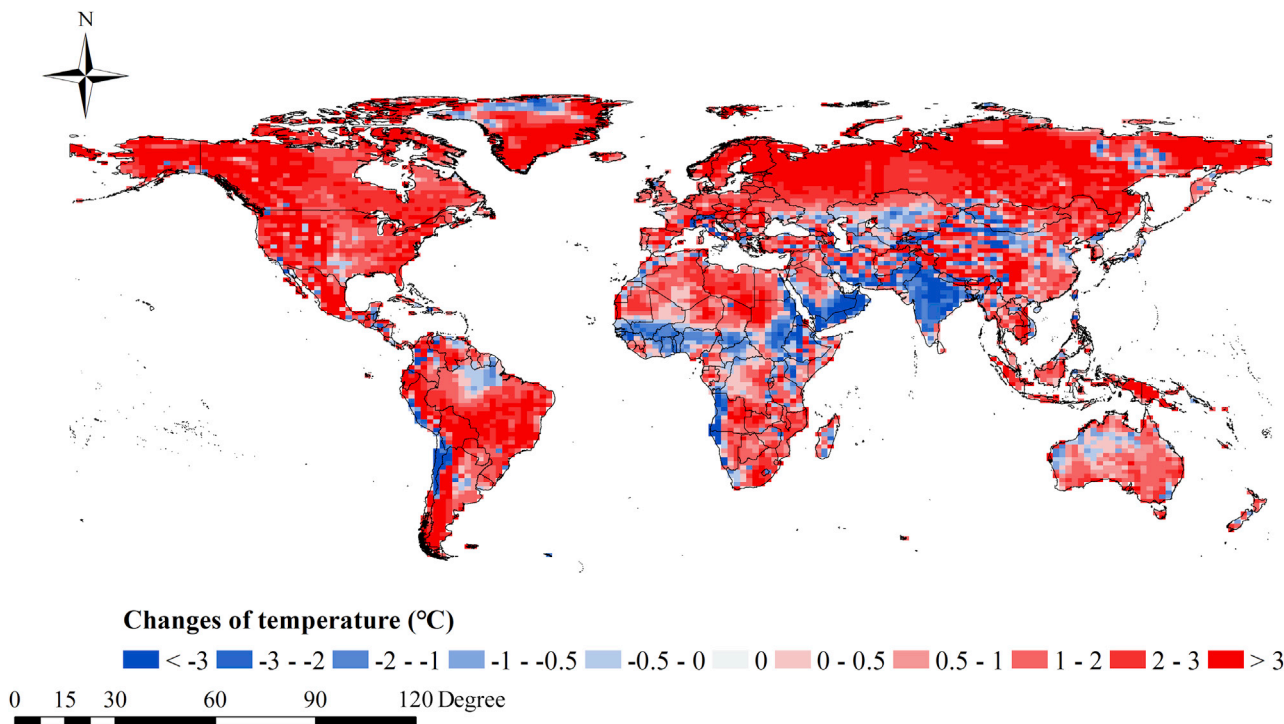

Figure 1. Annual changes of bias-corrected temperature compared to the original output temperature from UKESM1 model under SSP5-8.5 in 2080s

1. Heatwave threshold calculation.
   a. Organize the historical temperature data into a CSV file with each column representing one day and each row representing a grid cell.
   b. Threshold calculation.
      i. Based on the literature, select the 98th percentile of daily maximum temperature as the most suitable threshold for each grid cell.
      ii. Calculate the threshold using the df.T.quantile function from Pandas library.
2. Consider ozone pollution threshold as 100 μg/m^3, based on the 8-h limit in the WHO air quality guidelines.
3. Sort the temperature data and ozone concentration data of each period under each scenario into CSV files, and apply the Pandas package to discriminate events in each grid from the CSV file. Assign the grids in the same row that are continuously higher than the heatwave threshold/ozone pollution threshold as 1, and assign the remaining grids as 0.
4. Based on the results of the previous step, identify the compound events as the grid cell which is assigned the value as 1 in both heatwave and ozone pollution CSV files.

Calculating population exposure

⏱ Timing: 4–5 h depending on the file sizes

This step describes the calculations of exposure to compound events and their spatiotemporal trends. In this step, we adopt two types of exposure indicator including exposure days and exposure person-days. The indicator of exposure days only addresses the number of the natural occurrence days of the event, while the exposure person-days would consider population exposed to the event days.

5. Use exposure days as one of the exposure indicators. It is the sum of the number of event occurrence days in each grid cell under the corresponding scenario and period.
6. For the three types of events, compute the annual average number of event occurrence days for each grid cell, each scenario, and each period, and calculate their changes by comparing to the baseline period. We then display the statistical results mainly in the form of thematic maps.
7. Use exposure person-days as the exposure indicator. It is the product of the number of exposure days and the population in the same grid cell under the corresponding scenario and period.
8. For the population exposure to the three types of events, compute the annual average number of exposure person-days for each grid cell, each scenario, and each period, and calculate their changes by comparing to the baseline period. We then display the statistical results mainly in the form of thematic maps.
9. Display the globally averaged temporal variations of exposure days and exposure person-days under different scenarios in the form of a line chart.

   Note: For detailed format of all the maps and charts, please refer to Ban's study.[1]

Income subgroup analysis

⏱ Timing: 4–5 h depending on the file sizes

This step describes the method of evaluation of compound event exposure and compares the results of different countries grouped based on their income levels, as well as different age groups in each group of countries.

10. Divide all the grid cells into four categories according to the income levels defined by the World Bank and display the statistics in the form of bar charts.

11. Calculate the population exposed to compound events of different intensities in each country group.
    a. Here, we divide the intensity of the compound event into 4 grades, namely, 0 days, > 5 days, > 10 days, and > 20 days. This grading method can better reflect the inequality among different countries based on their income levels, especially in the SSP1-2.6 scenario.
    b. Calculate the proportion of the population subjected to different exposure intensities, and display the results in a bar chart.

For example, we match the population to the grid cells corresponding to each LIC with the number of days of exposure, and then count whether the number of days of exposure is greater than 0, 5, 10, or 20 days in turn. If in a certain period of a scenario, the total population corresponding to all grid cells with more than 5 days of exposure is 1 million, while the total population of the remaining grid cells is 3 million, then the proportion is 25%.

12. Statistical representation of data on the different age group populations exposed to compound events in different groups of countries.
    a. Compare the proportion of the population that may be exposed to compound events of different intensities as well as the proportion of age specific population (0–4, 5–64, and over 64 years old) in each country group that may be exposed to compound events.
    b. Display the results of different age groups in the form of a bar graph.

    Note: In our study, we only focused on different age subgroups and income subgroups. In fact, there are other disparities could be considered in the future research, such as urban/rural regions, different climatic zones.

EXPECTED OUTCOMES

The spatial and temporal trends of compound events and population exposure can be compared, and based on this, the inequality between countries at different income levels can be identified.

LIMITATIONS

The lack of ozone observations and simulation data could reduce the validity of bias-correction, which may finally lead uncertainties in the identification and prediction of ozone pollution events.

TROUBLESHOOTING
Problem 1
The available datasets of the future ozone concentration predicted by the CMIP6 models are limited. In our study, we only adopt one model because it is the only model meeting the criteria of providing both daily ozone concentration and temperature under four SSP-RCP scenarios in each period. Application of single model can undoubtedly bring uncertainty in projection. We should considered this problem when designing the study.

Potential solution
The good news is that more updated datasets from different CMIP6 models are coming out. Data availability is increasing in satisfying the requirements of multiple scenarios, fine spatial resolutions, and precise temporal scales. Researchers could properly select models according to their study design.

Problem 2
The bias correction of ozone concentration errors may lead to unexpected uncertainties due to the lack of observational ozone data in some regions. The observed and simulated values should be initially combined into a grid of the same resolution. However, it is hard to obtain complete long-term historical observational datasets for all the regions across the world. For example, in our study,

the global observation dataset covered less regions out of American and Europe. We should consider this problem before the step of data preprocessing-5 Bias Correction.

Potential solution
There are two ways to solve this problem. First, remove the simulated values corresponding to the missing observational ozone values to reduce the uncertainties brought by missing data during the bias correction step, which is simple and fast, but may lost study samples. Second, perform data imputation by constructing models to simulate the historical ozone concentrations based on related factors such as meteorological factors, pollution emission, and population distribution, etc. Then applying the improved historical datasets in bias correction step.

Problem 3
We perform the bias correction by adjusting the distribution of years of historical observation and projected data in each grid cell at daily level. Therefore, bias correction step requires for huge computation resources, which should be considered in the step of data preprocessing-5 Bias Correction.

Potential solution
High performance computing systems are required. Parallel programming is more suitable to run the global datasets with daily-level gridded concentration and temperature data from multiple locations.

Problem 4
Since there are no consistent definitions worldwide for heatwave, ozone pollution events or compound heatwave and ozone pollution event, we defined these three types of extreme events according to previous literatures and experiences. Therefore, variances may exist in the projection results if different definitions are applied. We should consider this problem when designing the study.

Potential solution
One way to address this issue is to adopt different definitions in the same study and leave one as main analysis and others as sensitivity analysis. In our study published on One Earth (https://doi.org/10.1016/j.oneear.2022.05.007), we applied two definitions of heatwave using relative threshold (temperature percentile) and absolute threshold (a fixed temperature), respectively.

Problem 5
In our study, we address the disparity among different income groups. However, there are other disparities worthy to be discussed, such as disparity between urban and rural regions, disparities among different climatic zones, etc. For example, severe heatwaves in urban areas due to heat-island effect may lead to higher possibility of ozone pollution and compound heatwave and ozone pollution, the urban-rural disparity may be important to consider. We should consider this problem in study design and Income Subgroup Analysis step.

Potential solution
Based on the study aims, researchers could design different types of subgroup analysis. First to obtain the subgroups classified as urban or rural population; then to calculate exposure of each subgroup. The SSP dataset provides the projected urban and rural population, which could support the urban-rural subgroup analysis.

RESOURCE AVAILABILITY
Lead contact
Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Dr. Tiantian Li (litiantian@nieh.chinacdc.cn).

Materials availability

This study did not generate new unique materials.

Data and code availability

The projected data on temperature and ozone concentration can be obtained from the CMIP6 database (CMIP6 Data: https://esgf-node.llnl.gov/search/cmip6/). The population data can be obtained from the Socioeconomic Data and Applications Center (Global 1-km Population Data: https://sedac.ciesin.columbia.edu/data/set/popdynamics-1-km-downscaled-pop-base-year-projection-ssp-2000-2100-rev01/maps/services). Data describing the projection results of exposure to compound events and the relative Python code have been deposited and are publicly available at Zenodo: (Code base: https://doi.org/10.5281/zenodo.6591120). Any additional information required for reanalyzing the data reported in this study is available from the lead contact upon reasonable request.

AUTHOR CONTRIBUTIONS

T.L. contributed to the study conception; T.L. and J.B. designed the study; J.B., K.L., and Q.W. implemented the methods; and J.B. and K.L. contributed to the data evaluation, code development, result visualization, and manuscript drafting. J.B. and K.L. contributed equally to this work. All authors have discussed and approved the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Ban, J., Lu, K., Wang, Q., and Li, T. (2022). Climate change will amplify the inequitable exposure to compound heatwave and ozone pollution. One Earth 5, 677–686. https://doi.org/10.1016/j.oneear.2022.05.007.

2. Sellar, A.A., Jones, C.G., Mulcahy, J.P., Tang, Y., Yool, A., Wiltshire, A., et al. (2019). UKESM1: description and evaluation of the UK Earth system model. J. Adv. Model. Earth Syst. 11, 4513–4558.

3. Lange, S. (2019). EartH2Observe, WFDEI and ERA-Interim Data Merged and Bias-Corrected for ISIMIP (EWEMBI) (GFZ Data Services). https://doi.org/10.5880/pik.2019.004.

4. Schultz, M.G., Schröder, S., Lyapina, O., Cooper, O.R., Galbally, I., Petropavlovskikh, I., von Schneidemesser, E., Tanimoto, H., Elshorbany, Y., Naja, M., et al. (2017). Tropospheric Ozone Assessment Report: database and metrics data of global surface ozone observations. Elem. Sci. Anth. 5, 58. https://doi.org/10.1525/elementa.244.

5. Jones, B., O'Neill, B.C., and Gao, J. (2020). Global 1-km downscaled population base year and projection grids for the shared socioeconomic pathways (SSPs), revision 01. https://sedac.ciesin.columbia.edu/data/set/popdynamics-1-km-downscaled-pop-base-year-projection-ssp-2000-2100-rev01.

6. Hempel, S., Frieler, K., Warszawski, L., Schewe, J., and Piontek, F. (2013). A trend-preserving bias correction – the ISI-MIP approach. Earth Syst. Dyn. 4, 219–236. https://doi.org/10.5194/esd-4-219-2013.