



OPEN

Data driven pathway analysis and forecast of global warming and sea level rise

Jiecheng Song^{1✉}, Guanchao Tong¹, Jiayou Chao¹, Jean Chung², Minghua Zhang³, Wuyin Lin⁴, Tao Zhang⁴, Peter M. Bentler⁵ & Wei Zhu^{1✉}

Climate change is a critical issue of our time, and its causes, pathways, and forecasts remain a topic of broader discussion. In this paper, we present a novel data driven pathway analysis framework to identify the key processes behind mean global temperature and sea level rise, and to forecast the magnitude of their increase from the present to 2100. Based on historical data and dynamic statistical modeling alone, we have established the causal pathways that connect increasing greenhouse gas emissions to increasing global mean temperature and sea level, with its intermediate links encompassing humidity, sea ice coverage, and glacier mass, but not for sunspot numbers. Our results indicate that if no action is taken to curb anthropogenic greenhouse gas emissions, the global average temperature would rise to an estimated 3.28 °C (2.46–4.10 °C) above its pre-industrial level while the global sea level would be an estimated 573 mm (474–671 mm) above its 2021 mean by 2100. However, if countries adhere to the greenhouse gas emission regulations outlined in the 2021 United Nations Conference on Climate Change (COP26), the rise in global temperature would lessen to an average increase of 1.88 °C (1.43–2.33 °C) above its pre-industrial level, albeit still higher than the targeted 1.5 °C, while the sea level increase would reduce to 449 mm (389–509 mm) above its 2021 mean by 2100.

As of 2021, the global mean surface temperature had already risen by 1.21 °C over its pre-industrial level (1850–1900 average)¹, while the global mean sea level had increased by 82 mm compared to its 1986–2005 average². Global warming and sea level rise are two critical indicators of climate change. The record breaking summer of 2022 has sent an urgent and harsh reminder for action and resolution, with temperatures exceeding 40 °C in many parts of Europe and the USA³. Identifying the causal pathways leading to this change based on historical, open-box data and models that clarify the often complicated physical processes involved will provide more transparency on the role and significance of greenhouse gas emissions in driving climate change. In this work, we use the unified structural equation modeling (usem) approach developed by our research group⁴ to identify the climate change pathways leading to increased global mean surface temperature (GMST) and global mean sea level (GMSL) based on yearly historical data. Backtesting is performed to validate the pathway models. The confirmed pathway models are then used to forecast future GMST and GMSL based on (1) the unrestricted scenario assuming the global community will make no significant actions to contain the anthropogenic greenhouse gas emissions, (2) the COP26 scenario of restricted anthropogenic carbon dioxide (CO₂) and methane (CH₄) emissions, and (3) the SSP (Shared Socioeconomic Pathways) scenarios of greenhouse gases concentration projection. We have also modeled the link between global sea level to regional sea level and illustrated the impact of regional sea level rise on coastal metropolises such as New York City, USA, and Osaka, Japan, by 2050 and 2100.

Many previous studies have endeavored to establish causal relations and predictions for global mean surface temperature (GMST), and global mean sea level (GMSL)^{5–8}. Traditionally physical models utilizing complex process models of the general circulation of the planet's atmosphere and ocean have been employed to generate global climate change predictions, with strict boundary and starting conditions^{9–24}. These models can be used to comprehend the fundamental dynamics of the physical components of the natural climate phenomena, derive global temporal and spatial changes, and make predictions based on the future greenhouse gases emissions²⁵.

¹Department of Applied Mathematics and Statistics, State University of New York at Stony Brook, Stony Brook, NY 11794-3600, USA. ²Duke University, 2080 Duke University Road, Durham, NC 27708, USA. ³School of Marine and Atmospheric Sciences, State University of New York at Stony Brook, Stony Brook, NY 11794-5000, USA. ⁴Environmental and Climate Sciences Department, Brookhaven National Laboratory, Upton, NY 11973-5000, USA. ⁵Department of Statistics, University of California, Los Angeles, Los Angeles, CA 90095-1554, USA. ✉email: song.jiecheng@stonybrook.edu; jiecheng.song@gmail.com; wei.zhu@stonybrook.edu

Differences between simulations and observations are often expected, with significant differences usually caused by numerical approximations, nonlinear behaviors, unresolved small-scale processes, and variability in data^{26–28}. Therefore, the projections of the physical models can range widely²⁹.

In the last few decades, data-driven methods utilizing time series analysis and machine learning methods have been increasingly adopted to forecast global surface temperature or sea level^{5–7,30–41} and local climate^{28,42–48}. Data-driven models attempt to obtain evidence of externally driven climate change while minimizing the use of complex climate models. The advantages are usually simpler models and lower computational burdens. These models can be typically divided into two categories: cointegration approaches⁴⁹ and regression approaches⁵⁰. Cointegration is a statistical technique used to find the equilibrium connection between two or more non-stationary time series over the long run. The cointegration method assumes that non-stationary time series have a long-term relationship. These models, however, are not designed to capture the complex inter-relationships among climate variables⁴⁶. Unlike physical models, data-based models often suffer from data inconsistency and absence more severely. Machine learning methods, including deep learning, have been applied to examine global climate change recently^{28,30,31,41,44}. However, machine learning methods are often regarded as black boxes, with the model's underlying dynamics unseen to the users. In addition, machine learning methods require large amount of training data which are often unattainable in climate change studies.

In 2021, we published the first work⁷ on pathway analysis of global warming and sea level rise utilizing monthly historical data. In this paper, we have drastically extended the climate change network analysis by adopting much longer yearly data, representing the effect of greenhouse gas emissions in a more comprehensive manner, adding humidity and sunspot activities, and improving the pathway identification process by using both the system-wise and the equation-wise variable selection methods. Importantly, for the forecasting of GMST and GMSL based on the confirmed pathway models, we have focused on two scenarios: (1) without any restriction on anthropogenic greenhouse gas emission, and (2) with restriction on anthropogenic greenhouse gas emission as outlined in the 2021 United Nations Conference on Climate Change (COP26). To make our analysis more comparable, we also included SSP scenarios in this work. Finally, we have expanded the regional sea level projections beyond the US coastal line to include other global regions heavily affected by climate change.

Results

Path discovery and analysis of global mean surface temperature (GMST) and global mean sea level (GMSL). To analyze the causal pathways between GMST, GMSL, and other climate factors, we have considered 9 variables: (1) Global Mean Sea Level (GMSL), (2) Glaciers and Ice Sheets Mass Balance (Mass), (3) Arctic Sea Ice August Extent (Sealce), (4) Global Mean Surface Temperature with the sea ice area temperature measured by air (GMST), (5) Global Specific Humidity (Humidity), (6–8) Greenhouse Gases: CO₂, CH₄, and N₂O (which are represented jointly by their total global warming potential (GWP) to avoid multi-collinearity), and (9) Sunspot Number (SSN), a measurement of solar activity. We started with a full hypothetical unified structural equation model (uSEM)⁴ that includes all conceivable paths by only excluding those contradictory to common sense (Fig. 2a). The uSEM approach can incorporate both the contemporaneous and the longitudinal pathways, with the uSEM equation system consisting of a set of autoregressive distributed lags (ARDL) models. Subsequently, two variable selection methods were applied to select significant pathways based on historical data (Fig. 1).

First, a system-wise variable selection for the entire uSEM equation system was conducted using the adaptive LASSO (least absolute shrinkage and selection operator) in the regularized unified structural equation modeling (RuSEM) approach⁵¹. Secondly, an equation-wise variable selection was performed on each ARDL model of the uSEM equation system using the backward stepwise variable selection based on the Akaike information criterion (AIC). In our analysis, these two variable selection methods yielded the same selected model, confirming the robustness of the climate change pathways identified via uSEM. The identified uSEM system with significant paths was only refitted to obtain the estimated path coefficients (red for positive and blue for negative path coefficient values) and the corresponding p-values in parentheses (Fig. 2b). Backtesting using both the one-step forecast as well as the multi-step forecast methods was performed to further validate the identified climate change pathways. This uSEM path model is entirely data-driven, open-box with explicit equations, and with the minimum use of existing scientific knowledge for an intuitive and independent inference of the status and the

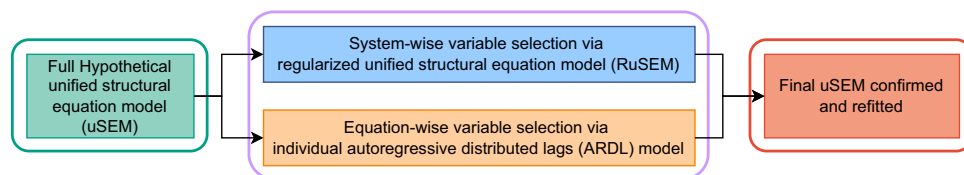


Figure 1. Flow chart of unified structural equation model (uSEM) pathway identification process using two variable selection methods, the system-wise regularized unified structural equation modeling (RuSEM) approach, and the equation-wise stepwise variable selection method using each individual uSEM equation, which is also an autoregressive distributed lags (ARDL) model. The fully hypothesized uSEM model and the final confirmed uSEM pathway model are depicted in Fig. 2a,b, respectively. For this work, the two variable selection methods identified a common final pathway model as shown in Fig. 2b below.

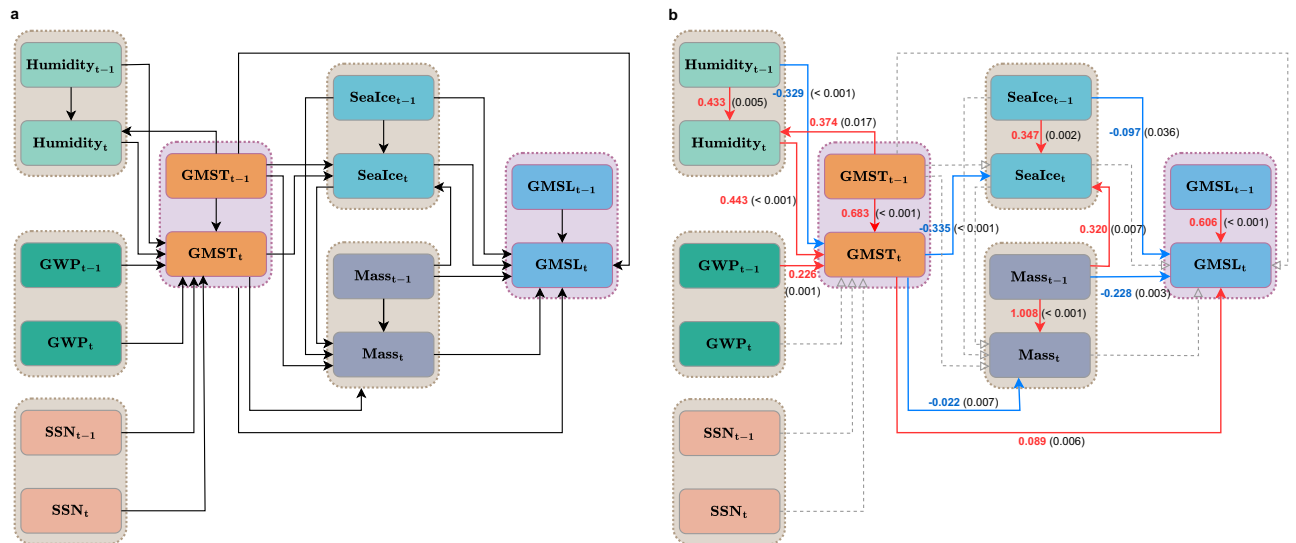


Figure 2. (a) Full hypothesized unified structural equation model (uSEM), and (b) the confirmed uSEM with significant pathways selected in unison by two variable selection methods—the system-wise regularized uSEM (RuSEM) approach and the equation-wise stepwise variable selection based on each ARDL model. Climate Factors included (from left top to right bottom) are: Global Specific Humidity (Humidity), Global Warming Potential (GWP), Sunspot Number (SSN), Global Mean Surface Temperature with sea ice area measured by air above sea ice (GMST), Arctic Sea Ice August Extent (Sea Ice), Glaciers and Ice Sheets Mass Balance (Mass), and Global Mean Sea Level (GMSL). In (a), the full hypothesized path model with all conceivable directed paths not contradictory to common sense as depicted by the black arrows. In (b), the final path model identified, significant positive or negative pathways are labeled with red or blue arrows with the corresponding path coefficients and p-values (in parentheses, 1-sided) labeled, while grey dashed arrows represent insignificant pathways at the significant level of 0.05 (1-sided).

future of our climate system in addition to the sophisticated physics-driven models which are often represented as a black-box to the laymen.

The data confirmed uSEM pathway as shown in Fig. 2b indicates that increases in greenhouse gas emission, in terms of the increased Global Warming Potential (GWP), would significantly increase the Global Mean Surface Temperature (GMST), while the Sunspot Number (SSN) has no significant impact on GMST. Humidity and GMST have a positive mutual interactive relationship. Furthermore, the increase in GMST would significantly decrease the Sea Ice coverage (SeaIce) and the Glaciers and Ice Sheets Mass Balance (Mass), which would in turn significantly increase the Global Mean Sea Level (GMSL). The melting glaciers also have a direct hand in the melting (or reduced formation) of sea ice, possibly due to the decreased ice sheet coverage leading to less reflective ice surfaces, or a decreased albedo, thus leading to a greater absorption (rather than reflection) of solar radiation that creates a positive feedback loop for increasing temperatures and the increased melting of ice sheets and sea ice. Sea ice is critical in the regulation of temperature, albedo, and ocean circulation, therefore a decrease in sea ice would lead to significant consequences for global climate. Notably, it has a significant role in driving the global ocean conveyor belt by sinking salty, denser water that forms underneath its surface, which circulates along the ocean floor and drives the flow of the warmer, less dense water on the surface⁵². An increase in GMST would also promote a higher GMSL directly besides its indirect role through Sea Ice decrease and Glacier melting.

Notably, the data-driven model yielded pathways highly consistent with recent scientific findings. GWP, for instance, has a positive impact on GMST, in agreement with the scientific verdict of the greenhouse effect^{53,54}. The positive mutual relationship between Humidity and GMST is consistent with the theory of the water vapor greenhouse effect and the water cycle theory, which states that warmer air contains more water vapor⁵⁵, thus increasing evaporation and reducing condensation. It is sensible that higher GMST results in melting of the glaciers, ice sheets, and sea ice due to warmer temperature^{56,57}. The negative impact from Mass or Sea Ice to GMSL can be explained similarly⁵⁸. Besides, the melting of Sea Ice may have contributed to GMSL rise indirectly through reduction in solar reflection and ocean thermal expansion. The positive impact from GMST to GMSL is consistent with the generally held view that the GMSL increase is partially driven by the thermal expansion of the oceans^{58,59}. In short, the data driven pathway analysis results align well with the principles of climate science. Detailed information about the data and methods, as well as detailed descriptions of the consistency between this data-driven approach and those based on physics, are shown in the Method section.

Forecast of global mean surface temperature (GMST) and global mean sea level (GMSL). The most recent United Nations Conference on Climate Change (COP26) was held in 2021 in Glasgow, Scotland, gathering almost 200 nations to discuss the critical issue of climate change and global strategy. To combat climate change, nations are urged to establish ambitious emission reduction plans. The conference participants agreed to a new climate deal, the Glasgow Climate Pact, aiming to keep GMST within 1.5 °C above its pre-industrial levels. According to the Glasgow Climate Conference⁶⁰, it is essential that the anthropogenic CO₂ emissions should

decrease by 45% by 2030 relative to its 2010 level, and the anthropogenic CH₄ emissions should decrease by 30% by 2030 relative to its 2020 level in order to achieve this goal. However, several studies^{61–63} have since questioned the sufficiency of the COP26 guidelines. Our research aims to answer the following questions by using pure data-driven models based on historic data: (1) What will the GMST and GMSL be by 2100 if the anthropogenic greenhouse gas emission is not controlled per COP26 regulations (namely, the unrestricted scenario)? (2) What will the GMST and GMSL be by 2100 if COP26 regulations are followed through (namely, the COP26 scenario)?

As shown in Fig. 2b, the GWP is solely incorporated in our path model as a predictor. In other words, no ARDL model in the uSEM equation system includes GWP as a response variable. For forecasting, we can model the trend of GWP changes from both anthropogenic and natural causes by fitting classic autoregressive integrated moving average (ARIMA) time series models, to historical data. The best fitting time series model for the historical data is an ARIMA(1,1,1) model. For forecasting under the COP26 scenario, our model calculates the future values of anthropogenic GWP of CO₂ and CH₄ using the COP26 guidelines assuming a constant rate of emission reduction while modeling the unrestricted component of GWP using the ARIMA(1,1,1) model. Permafrost can also contribute to future greenhouse gas emissions⁶⁴. To incorporate permafrost's effect in the forecast model, the forecast of CO₂ and CH₄ emissions from the Arctic permafrost under medium and low scenarios according⁶⁴ are added to the forecast of GWP under the unrestricted scenario and COP26 restricted emission scenario correspondingly. Details about permafrost greenhouse gas emission can be found in the Supplementary Materials.

Our uSEM pathway model (Fig. 2b) predicts that under the unrestricted scenario, GMST will increase to 1.97 °C above its pre-industrial level (or 0.76 °C above its 2021 level) by 2050, and 3.28 °C above its pre-industrial level (or 2.07 °C above its 2021 level) by 2100 (Fig. 3c). Using the 20-year mean level from 1986 to 2005 as the baseline, GMSL will increase to 246.72 mm (or 164.30 mm above its 2021 level) by 2050, and 655.25 mm (or 572.83 mm above its 2021 level) by 2100 (Fig. 3e). In contrast, under the COP26 scenario, GMST will increase to 1.66 °C above its pre-industrial level (or 0.45 °C above its 2021 level) by 2050, and 1.88 °C above its pre-industrial level (or 0.67 °C above its 2021 level) by 2100 (Fig. 3d). The GMSL will increase to 229.67 mm (or 147.25 mm above its 2021 level) by 2050, and 531.23 mm (or 448.81 mm above its 2021 level) by 2100 (Fig. 3f). Table 1 summarizes these results, with the inclusion of confidence intervals. This study and the projections indicate that the mandates set by COP26 will have a substantial influence on climate change mitigation. However, our results

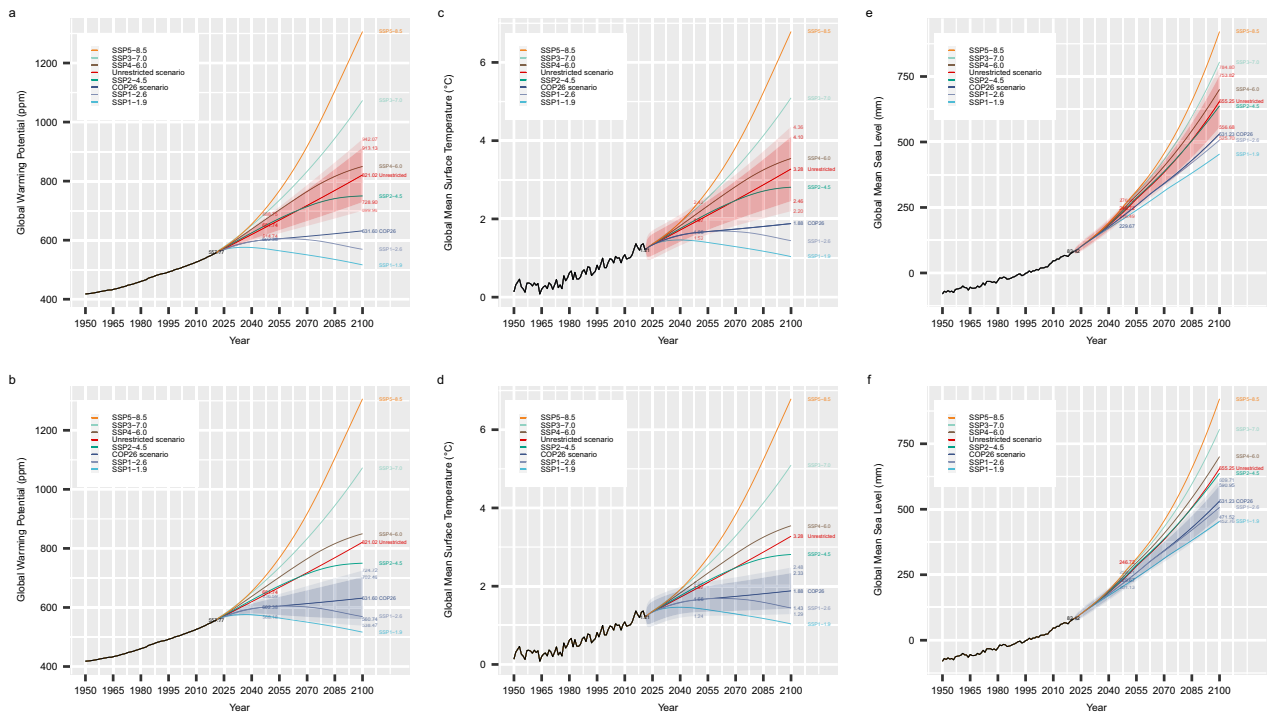


Figure 3. Data driven forecast of Global Warming Potential (GWP) (a,b), Global Mean Surface Temperature (GMST) (c,d), and Global Mean Sea Level (GMSL) (e,f) under the unrestricted scenario (red), the COP26 scenario (grey), SSP5-8.5 (orange), SSP3-7.0 (light-green), SSP4-6.0 (brown), SSP2-4.5 (dark-green), SSP1-2.6 (light-grey) and SSP1-1.9 (light-blue) from now till 2100, the uncertainties (forecast interval) under the unrestricted and the COP26 scenarios were shown by the red and grey shaded area. For the unrestricted and the COP26 forecast, values of the expected means and the 95% and 99% forecast intervals are shown for 2050 and 2100 respectively. Historical data before 2022 are shown in black. The modeling and forecast of GWP was based on an estimated ARIMA(1,1,1) model using historical data for the unrestricted scenario and integrating the proposed restriction for the COP26 scenario (with greenhouse gas emitted by Arctic permafrost added). The estimation and forecast of GMST and GMSL are based on the confirmed uSEM model as shown in Fig. 2b.

	Year	Unrestricted scenario			COP26 scenario		
		Mean	95% Forecast interval	99% Forecast interval	Mean	95% Forecast interval	99% Forecast interval
GMST (°C)	2050	1.97	1.62–2.31	1.52–2.42	1.66	1.34–1.97	1.24–2.07
	2100	3.28	2.46–4.10	2.20–4.36	1.88	1.43–2.33	1.29–2.48
GMSL (mm)	2050	246.72	223.72–269.72	216.49–276.95	229.67	207.96–251.38	201.13–258.21
	2100	655.25	556.68–753.82	525.70–784.80	531.23	471.52–590.95	452.76–609.71
	Year	SSP1-1.9	SSP1-2.6	SSP 2-4.5	SSP3-7.0	SSP4-6.0	SSP5-8.5
GMST (°C)	2050	1.43	1.66	2.02	2.32	2.17	2.43
	2100	1.04	1.44	2.81	5.09	3.54	6.79
GMSL (mm)	2050	216.15	229.47	250.33	268.39	259.03	271.93
	2100	453.92	507.36	638.03	805.73	701.11	921.64

Table 1. Predicted means and forecast intervals (95% and 99%) of GMST rise and GMSL rise by 2050 and by 2100 under unrestricted or COP26 restrictions on greenhouse gas emissions; and the predicted means of GMST rise and GMSL rise by 2050 and by 2100 under SSP1-1.9, SSP1-2.6, SSP2-4.5, SSP3-7.0, SSP4-6.0 and SSP5-8.5 scenarios. The baseline for GMST is the pre-industrial (1850–1900 average) level, while the baseline for GMSL is the 20-year mean value of GMSL from 1986 to 2005. The units of GMST are °C and the units of GMSL are mm.

indicate that the proposed plans may fall slightly short of keeping the global temperature rise below 1.5 °C over its pre-industrial level, as we forecast an average warming of 1.88 °C based on COP26 guidelines. Therefore, the COP26 guidelines are proven to be insufficient in keeping warming within 1.5 °C to curb the effects of severe climate change. To put in perspective, we have also included several Shared Socioeconomic Pathways (SSP) scenarios for the greenhouse emissions⁶⁵. We use future concentrations for CO₂, CH₄, and N₂O under the SSP scenarios^{66,67} to forecast the GMST and GMSL. The forecasts for the GMST and GMSL under different SSP scenarios by using our pathway model are also shown in Fig. 3. Under SSP 5-8.5 scenario, which is the most extreme scenario, GMST will increase to 2.43 °C above its pre-industrial (or 1.22 °C above its 2021 level) by 2050, and 6.79 °C above its pre-industrial (or 5.58 °C above its 2021 level) level by 2100. GMSL will increase to 271.93 mm (or 189.51 mm above its 2021 level) by 2050, and 921.64 mm (or 839.22 mm above its 2021 level) by 2100. As shown in Fig. 3, our projection of the unrestricted scenario is closest to the SSP 2-4.5 scenario, while our projection of COP26 scenario is closest to the SSP 1-2.6 scenario.

Forecast of regional mean sea level (RMSL). Regional sea level rise along the coastal lines is of great concern to the public and policymakers. For each major coastal region, we use the ARDL model, a versatile time series regression model, to predict the regional mean sea level (RMSL) based on its previous levels in time as well as the global mean sea level (GMSL) at the same and the prior time points. We studied eight coastal regions in total, with the New York City and Osaka results presented below. The other six regions are included in the Supplementary Materials. With stepwise variable selection, $RMSL_{t-1}$ and $GMSL_t$ are selected as significant predictors for all eight regional models. Based on the ARDL models, forecasts for New York City and Osaka are shown in Table 2 and Fig. 4.

The mean of 20 years of regional sea levels from 1986 to 2005 was taken as the baseline for the regional mean sea levels. By 2050, the New York City regional mean sea level (NYRMSL) will increase to 373.63 mm above the baseline, or equivalently 245.56 mm above its 2021 level under the unrestricted scenario, and these will reach 995.85 mm or 867.79 mm respectively by 2100. Under the COP26 scenario, by 2050, NYRMSL will increase to 348.34 mm above the baseline, or equivalently 220.28 mm above its 2021 level, and by 2100, these will be 808.25 mm or 680.20 mm, respectively. Under SSP 5-8.5, which is the most extreme scenario, by 2050, NYRMSL will increase to 411.23 mm above the baseline, or equivalently 283.16 mm above its 2021 level, and by 2100 these will be 1398.03 mm or 1269.97 mm, respectively (Fig. 4a,b).

Under the unrestricted scenario, the Osaka regional mean sea level (OSARMSL) will increase to 393.22 mm above the baseline, or 271.32 mm above its 2021 sea level by 2050, and these will reach 1061.39 mm or 939.49 mm by 2100. Under the COP26 scenario, OSARMSL will increase to 366.93 mm above the baseline, or 245.03 mm above its 2021 level by 2050, and these will reach 861.65 mm or 739.75 mm, respectively by 2100. Under SSP 5-8.5, which is the most extreme scenario, in 2050, OSARMSL will increase to 432.45 mm above the baseline, or equivalently 310.55 mm above its 2021 level, and by 2100 these will be 1488.58 mm or 1366.68 mm, respectively (Fig. 4c,d).

Considering the tidal effect, storm surge, and other climate factors, the maximum regional sea level can be much higher than the yearly mean sea level. The New York City (1920–2019) and Osaka (1961–2020) hourly mean sea level data were gathered to compute the yearly sea level fluctuation (the difference between the yearly maximum sea level and the yearly mean sea level). On the average, the yearly highest sea level is 1536.93 mm above the yearly mean sea level for New York City, while for Osaka it is 1138.95 mm. The Augmented Dickey–Fuller (ADF) tests confirm that the yearly fluctuations are stationary. On average, based on these assumptions and calculations, under the unrestricted scenario, the yearly highest sea level in 2100 could reach 2662.64 mm for New York City and 2326.49 mm for Osaka. Under the COP26 restricted emission scenario, these will be 2437.56 mm for New York City and 2098.60 mm for Osaka.

	Year	Unrestricted scenario			COP26 scenario		
		Mean (mm)	95% Forecast interval	99% Forecast interval	Mean (mm)	95% Upper forecast interval	99% Upper forecast interval
New York	2050	373.63	307.18–440.07	286.30–460.95	348.34	282.48–414.20	261.78–434.89
	2100	995.85	865.99–1125.71	825.19–1166.51	808.25	715.88–900.63	686.85–929.66
Osaka	2050	393.22	312.77–473.67	287.49–498.94	366.93	286.84–447.02	261.68–472.19
	2100	1061.39	935.24–1187.54	895.61–1227.18	861.65	763.65–959.65	732.86–990.44
	Year	SSP1-1.9 (mm)	SSP1-2.6 (mm)	SSP 2–4.5 (mm)	SSP3-7.0 (mm)	SSP4-6.0 (mm)	SSP5-8.5 (mm)
New York	2050	328.16	348.01	379.08	405.96	392.03	411.23
	2100	691.30	772.42	970.29	1223.20	1065.67	1398.03
Osaka	2050	345.81	366.56	398.99	427.07	412.51	432.45
	2100	737.11	823.84	1034.80	1303.12	1136.31	1488.58

Table 2. Predicted mean and forecast interval (95% and 99%) for the regional mean sea level (RMSL) rise at New York City and Osaka by 2050 and 2100 respectively under unrestricted or COP26 restrictions on greenhouse gas emissions; and the predicted means of the regional mean sea level (RMSL) rise at New York City and Osaka by 2050 and 2100 under SSP1-1.9, SSP1-2.6, SSP2-4.5, SSP3-7.0, SSP4-6.0 and SSP5-8.5 scenarios. The RMSL baselines for New York City and Osaka are the respective mean values of the 20-year RMSL for the two regions from 1986 to 2005. The units of RMSL are mm.

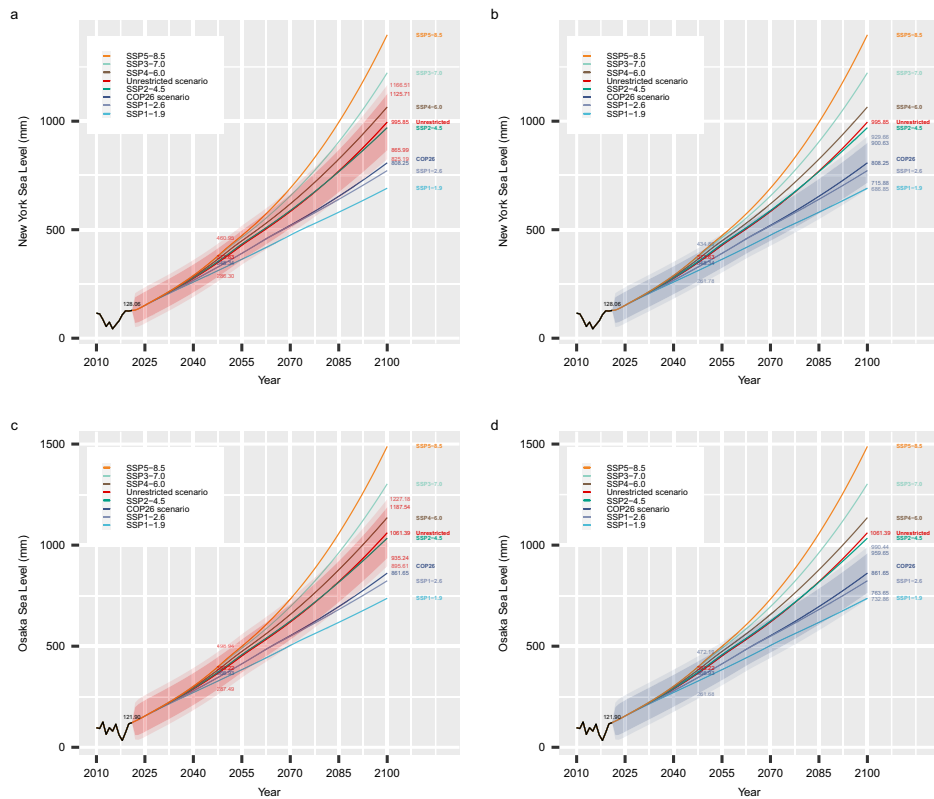


Figure 4. Regional mean sea level (RMSL) rise projections. The New York City (a,b) and Osaka (c,d) regional sea level rise projections under the unrestricted scenario (red), the COP26 scenario (grey), SSP5-8.5 (orange), SSP3-7.0 (light-green), SSP4-6.0 (brown), SSP2-4.5 (dark-green), SSP1-2.6 (light-grey) and SSP1-1.9 (light-blue) from now till 2100, the uncertainties (forecast interval) under the unrestricted and the COP26 scenarios were shown by the red and grey shaded area. For unrestricted and COP26 forecast, values of the expected means and the 95% and 99% forecast intervals are shown for 2050 and 2100 respectively. Historical data before 2022 are shown in black. The projections of the regional mean sea level are based on the ARDL models in Method section.

For New York City, if the yearly mean sea level will increase to 997.65 mm above the 2021 level (the upper 95% forecast interval bound under the unrestricted scenario) by 2100, according to historical hourly data, on average there will be 45.54 days per year the daily highest sea level will be higher than 2 m above its current level as shown in Fig. 5c. The 45.54 days of severe flooding will reduce to 6.29 days if the COP26 restrictions on greenhouse gas emissions can be followed through. For Osaka, if the yearly mean sea level increases to 1065.64 mm above its 2021 level (the upper 95% forecast interval bound under unrestricted scenario) by 2100, according to historical hourly data, on average we will see 3.55 severe flooding days that the daily highest sea level will be higher than 2 m above its current level as shown in Fig. 5f, while this will reduce to 0.38 day under the COP26 restricted emission scenario. In summary, we found that the COP26 resolutions on anthropogenic greenhouse gas reduction would be helpful in controlling the global mean surface temperature to a certain extent, although such effects are not as pronounced in regulating global and regional sea level rises.

Figure 5 uses the program Google Earth Pro to present a 3D simulation for the chosen areas in 2100 under the different scenarios. Although the sea level rise conditions in the 3D maps do not exactly match individual greenhouse gas emission scenarios in our analysis due to limitations in the Google map resolution, these maps do sufficiently convey the range of consequences of sea level rise from the spectrum of greenhouse gas emission scenarios included in this study. Osaka will be more severely affected than New York City due to the difference in altitude between these two cities.

Discussion

In this work, we presented a data driven approach to identify pathways leading to global warming and sea level rise. As expected, greenhouse gas emissions play a significant role in driving climate change. Furthermore, we found that the number of sunspots does not influence climate change significantly. We also found global surface temperature and humidity in a positive feedback loop with each other, while higher temperature induces higher sea level through the melting of sea ice and glaciers as well as ocean water thermal expansion. The melting glacier also has a direct hand in the melting of sea ice in addition to the higher temperature. Forecasting based on the pathway system reveals that without any intervention, global surface temperature will rise to a mean level of 3.28 °C above its pre-industrial level by 2100; whereas if the COP26 resolution on the reduction of anthropogenic greenhouse gas emission can be followed through, such increase would reduce to 1.88 °C, just short of the intended goal of 1.5 °C. The global sea level, however, will not be as drastically reduced by the COP26 resolution, with the global sea-level expected to rise an average of 448.81 mm above its 2021 mean following COP26 emission regulations, while 572.83 mm with no regulations. Regional sea level projections based on our models show that the increased global sea level will induce more frequent and severe flooding for the world's coastal regions including heavily populated metropolis such as New York City and Osaka. This calls for added protective measures in infrastructure to ensure that urban areas are more resilient to extreme flooding and tidal surges. Lastly, the data driven findings for the unrestricted scenario are found to be highly consistent with the

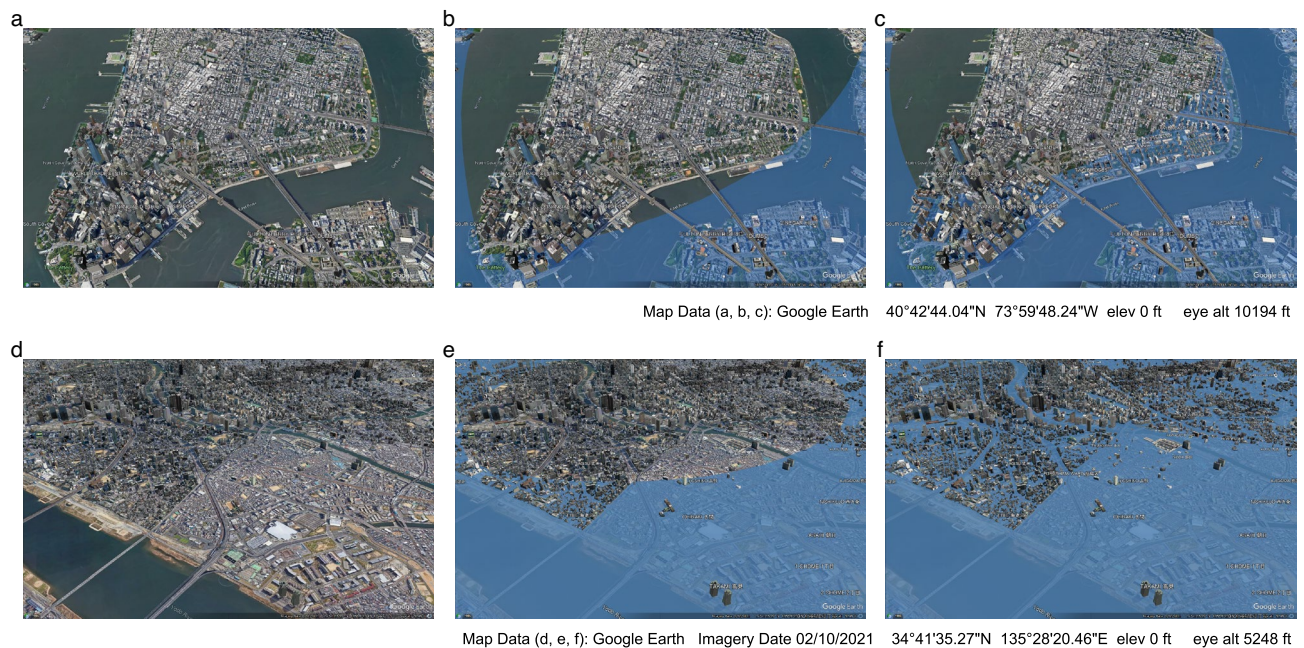


Figure 5. The 3D Google map simulation for regional mean sea level increase ranging from 0, to 1 and finally to 2 m for New York City (a–c) and Osaka (d–f) respectively, from its current level. According to our analysis, the New York City regional mean sea level will rise to 680.20 mm, 867.79 mm and 1269.97 mm higher than its current level under the COP26, unrestricted, and SSP5-8.5 scenarios respectively. The Osaka regional mean sea level will rise to 739.75 mm, 939.49 mm and 1366.68 mm higher than its current level under the COP26, unrestricted, and SSP5-8.5 scenarios, respectively.

2021 IPCC projections on climate change, underscoring the necessity to take drastic actions to mitigate future damages. The passage of the Inflation Reduction Act with \$369bn (£305bn) budgeted for climate action through the US Congress, the largest investment in US history—has brought a glimmer of hope in this critical moment of our collective fight to ensure a sustainable earth and future.

Finally, we acknowledge that the main obstacle to our analysis is the limited length of historical data available. Our model is a linear system due to data limitation, while the real-world climate processes are often non-linear. However, the short-term prediction accuracy of our system has been validated by back-testing based on historical data. Moving forward, we will update our forecasting system as more data are available. Furthermore, we will include more climate features such as ocean processes and paleoclimate time series as these data are increasingly procured and validated as technology progresses⁶⁸. We are hopeful to have a more panorama view of the climate change on earth through this incremental process.

Methods

Data. In this study, the following climate-related variables with yearly frequency from 1950 to 2021 were collected from multiple sources:

1. Global Mean Sea Level (GMSL) in mm was gathered from the mean of two data sources^{69,70}. The data for the year 2021 were collected and calibrated from GSFC (2021)².
2. Glaciers and Ice Sheets Mass Balance (Mass) in Gt (Gigaton) was estimated by two methods: (i) the product of global mean glacier mass in water equivalence from the world glacier monitoring service [1950–2021]⁷¹ and the total glaciers and ice sheets area⁷²; (ii) Summation of 17 mountain glaciers mass balance⁷³, Greenland and Antarctica ice sheet mass balance [2002–2020]⁷⁴, with missing values imputed by Kalman filter⁷⁵; overlapping period of estimation (i) and (ii) were calibrated by the average values.
3. Arctic Sea Ice August Extent (SeaIce) in Mkm² was gathered from two sources: (i) observed August Arctic sea-ice extent [1870–2008]⁷⁶, and (ii) observed north hemisphere sea ice extent in August [1979–2021]⁷⁷. The least squares method was adopted to calibrate and combine these two data sources.
4. Global Surface Temperature (with sea ice area measured by the air above sea ice) (GMST) in °C was obtained from Rohde (2020)¹.
5. Global Specific humidity (Humidity) in kg/kg (mass of water vapor per kilogram of moist air)⁷⁸.
6. Greenhouses gases: CO₂ [1950–1957]⁷⁹ [1958–2021]⁸⁰, CH₄ [1950–1983]⁸¹ [1984–2021]⁸², and N₂O [1950–1977]⁸¹ [1978–2021]⁸³. Global warming potential (GWP), which is measured by units of CO₂ equivalents in the environmental impacts, is adopted to analyze the effect of greenhouse gases in this study. The formulation to calculate GWP is shown in Eq. (1), following Intergovernmental Panel on Climate Change (IPCC) reports⁸⁴.

$$GWP = CO_2 + 28CH_4 + 265N_2O \quad (1)$$

7. Greenhouse gases concentration projection for CO₂, CH₄ and N₂O [2022–2100] under SSP scenarios^{66,67}.
8. Sunspot Number (SSN)⁸⁵.
9. Regional Mean Sea Level (RMSL) data were obtained from the Permanent Service for Mean Sea Level^{86,87}: New York City RMSL is obtained from the Battery station, and Osaka RMSL is obtained from the Osaka station.
10. New York City Hourly Water Level data were collected from the University of Hawaii Sea Level Center⁷⁰, and Osaka Hourly Water Level data were collected from Japan Oceanographic Data Center⁸⁸.

Augmented Dickey–Fuller (ADF) test results confirm all the variables are integrated of order 1, i.e. I(1) processes, based on historical data analysis.

Path analysis model identification and estimation for global mean surface temperature (GMST) and global mean sea level (GMSL). Structural equation modeling (SEM) refers to a family of multivariate procedures designed to infer causal relationships among a set of variables⁸⁹. It can be viewed as a system of multivariate regression equations with variables that can serve as both independent and dependent variables. Latent variables are also common although in this work we did not use any latent variables. Inference of SEM usually focuses on estimating the model implied population covariance matrix through the likelihood function:

$$F_{ML} = \log|\Sigma(\theta)| + tr(S\Sigma(\theta)^{-1}) - \log|S| - q, \quad (2)$$

where $\Sigma(\theta)$ is the model implied covariance matrix with θ representing model parameters including path coefficients, S is the sample variance–covariance matrix, and q is the number of parameters to be estimated. The unified structural equation model (uSEM)⁴, also known as Vector Autoregression-Structural Equation Modeling (VAR-SEM) or the Dynamic SEM modeling can be used to analyze the proposed causal associations among a set of time-series variables by incorporating both the contemporaneous relations and the longitudinal relations, simultaneously. Contemporaneous relations reflect relationships between variables at the same time point, while longitudinal temporal relations are defined as relationships between variables at different time points.

Variable selection for the hypothesized full uSEM model in this work was done in two approaches:

- (1) A system-wise variable selection based on the entire set of uSEM equations via the regularized unified structural equation modeling (RuSEM) method⁵¹, where regularization is done through the adaptive Lasso⁹⁰.

- (2) An equation-wise variable selection based on each individual uSEM equation, which can also be viewed as an autoregressive distributed lags (ARDL) model. The variable selection for each ARDL was done through the backward stepwise variable selection method.

The adaptive LASSO⁹⁰ penalizes parameters after each parameter is scaled by the un-penalized maximum likelihood estimators (MLE) in the following equation:

$$F_{\text{lasso}} = F_{\text{ML}} + \lambda \|\theta_{\text{ML}}^{-1} * \theta_{\text{pen}}\|_1, \quad (3)$$

where λ is the regularization parameter with a common initialization value of 0.1.

For equation-wise variable selection, each individual ARDL model has the following general form:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \dots + \beta_p y_{t-p} + \delta_0 x_t + \delta_1 x_{t-1} + \dots + \delta_q x_{t-q} + \epsilon_t, \quad (4)$$

where y_t is the endogenous variable at time t , x_t is the exogenous variable at time t , β_0 is the intercept, β_i , $i \in \{1, \dots, p\}$ and δ_j , $j \in \{0, \dots, q\}$ are the coefficients. The Akaike information criterion (AIC) is used to de-select unnecessary paths in the backward stepwise variable selection.

For our work, the final path model selected based on the system-wise and the equation-wise variable selection methods happen to be identical—signifying the robustness of our climate pathway result. The final path model, namely, the final refitted uSEM using only the significant pathways identified, is represented in the matrix form as follows, which can also be viewed as a system of ARDL time series regression equations:

$$\begin{bmatrix} Humidity_t \\ TEMP_t \\ Mass_t \\ Sealce_t \\ GMSL_t \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0.443 & 0 & 0 & 0 & 0 \\ 0 & -0.022 & 0 & 0 & 0 \\ 0 & -0.335 & 0 & 0 & 0 \\ 0 & 0.089 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} Humidity_t \\ TEMP_t \\ Mass_t \\ Sealce_t \\ GMSL_t \end{bmatrix} + \begin{bmatrix} 0.433 & 0.374 & 0 & 0 & 0 \\ -0.329 & 0.683 & 0 & 0 & 0 \\ 0 & 0 & 1.008 & 0 & 0 \\ 0 & 0 & 0.320 & 0.347 & 0 \\ 0 & 0 & -0.228 & -0.097 & 0.606 \end{bmatrix} \begin{bmatrix} Humidity_{t-1} \\ TEMP_{t-1} \\ Mass_{t-1} \\ Sealce_{t-1} \\ GMSL_{t-1} \end{bmatrix} + \begin{bmatrix} 0 \\ 0.226 \\ 0 \\ 0 \\ 0 \end{bmatrix} [GWP_{t-1}] + \begin{bmatrix} 0.034 \\ 0.029 \\ -0.052 \\ -0.026 \\ 0.047 \end{bmatrix} \quad (5)$$

GWP forecasts under COP26. The 26th United Nations Climate Change Conference (COP26), held in Glasgow, UK, from 31 October to 13 November 2021, has reached two key resolutions to reduce the anthropogenic greenhouse gas emissions: (1) to reduce the anthropogenic CO₂ (carbon dioxide) emission by 45% by 2030, compared to its 2010 levels, and (2) to reduce the anthropogenic CH₄ (methane) emissions by 30% by 2030, compared with its 2020 levels.

It is assumed that without anthropogenic emissions, the greenhouse concentration should remain stable, the same as last year. The global CO₂ concentration level in 2010 and 2009 was measured at 389.89 parts per million (ppm) and 387.34 ppm respectively, featuring a yearly increase of 2.54 ppm. With the assumption, we conclude all 2.54 ppm increase is anthropogenic, that is, due to human activities. Per COP26, for 2030, the expected concentration increase caused by anthropogenic emission of CO₂ will be $(1 - 0.45) * 2.54 \approx 1.40$ ppm. To simplify the matter, we have adopted a linear decrease, which means every year we will see the amount of emission decrease stays constant (we can also use other models such as exponential, which will not significantly change our results). The increase in CO₂ concentration in 2021 is 2.12 ppm. Therefore, if we can follow through the COP26 resolution, the annual increase of anthropogenic CO₂ concentration, from 2022 to 2030, in a linear decreasing pattern, will be (2.04, 1.96, 1.88, 1.80, 1.72, 1.64, 1.56, 1.48, 1.40) ppm. After 2030, it is assumed that the increase of CO₂ concentration will continue decreasing at the same rate until dropping to zero increase.

The global CH₄ concentration levels in 2020 and 2019 were measured at 1879.11 parts per billion (ppb) and 1866.60 ppb respectively, featuring a yearly increase of 12.51 ppb. With the assumption above, we conclude all 12.51 ppb is anthropogenic, that is, due to human activities. Per COP26, for 2030, the expected concentration increase caused by anthropogenic emission of CH₄ will be $(1 - 0.3) * 12.51 \approx 8.76$ ppb. To simplify the matter, we have adopted a linear decrease, which means every year we will see the amount of emission decrease stays constant (we can also use other models such as exponential, which will not significantly change our results). The increase in CH₄ concentration in 2021 is 16.52 ppb. Therefore, if we can follow through the COP26 resolution, the annual increase of anthropogenic CH₄ concentration, from 2022 to 2030, in a linear decreasing pattern, will be (15.66, 14.79, 13.93, 13.07, 12.21, 11.34, 10.48, 9.62, 8.76) ppb. After 2030, it is assumed that the increase of CH₄ concentration will continue decreasing at the same rate until dropping to zero increase.

Since no constraints were proposed for N₂O, it is assumed the N₂O concentration will increase following its historical trend, which is well fitted by an ARIMA(1,1,1) model. The uncertainty of the forecasted GWP level associated with the COP26 scenario was calculated under the assumption that the coefficient of variation ($CV = \sigma/\mu$, where σ is the standard deviation and μ is the mean) remains constant. The standard deviation in

the COP26 scenario is the corresponding standard deviation under the unrestricted scenario times the ratio of the mean of these two scenarios ($\sigma_C = \sigma_N \mu_C / \mu_N$). The final identified uSEM, the time series forecasts of GWP and its uncertainties, are used to forecast the future trend and uncertainties of GMST and GMSL until 2100 under the unrestricted scenario, the COP26 restriction scenario and the SSP scenarios. Meanwhile, backtesting tests using historical data were conducted to validate the forecast models. The backtesting used a 9:1 ratio of the training set to the test set. For the test data, both one-step ahead (predictors updated with true observed values at each time step) and multi-step ahead forecasting (predictors updated with forecasted values at each time step) were performed. Both results show that all the true values fall within the 99% forecast intervals validating the robustness of our model. Details are available in the Supplementary Materials.

Regional mean sea level (RMSL) forecasts. To forecast the RMSL, we have adopted the ARDL model with the stepwise variable selection method. The ARDL variable selection results are provided in Eq. (6) below:

$$\begin{aligned} NYRMSL_t &= 0.2477 + 0.2640NYRMSL_{t-1} + 1.1248GMSL_t \\ OSARMSL_t &= -2.4039 + 0.4461OSARMSL_{t-1} + 0.9129GMSL_t, \end{aligned} \quad (6)$$

where $NYRMSL_t$ is the regional mean sea level for New York City in year t , $OSARMSL_t$ is the regional mean sea level in Osaka in year t , $GMSL_t$ is the global mean sea level in year t . Results for the other six coastal locations examined are available in the Supplementary Materials. To reduce the impact of missing values in regional sea level data, we used a longer time span (1950–2015) for the training data when backtesting regional models. Model validation is confirmed by the backtesting results, showing that all the true values fall within the predicted range for all regions.

Consistency with physics-based models and forecasts. Although the data driven uSEM pathway modeling does not rely on physical theories and models, it still comes as a relief when we found our results highly consistent with the recent forecasts based largely on scientific theories. The IPCC 2021 climate change report⁹¹ has shown projections for global mean surface temperature (GMST) and global mean sea level (GMSL) rise under different scenarios. As the IPCC reports use a different method to assess greenhouse gas emissions than our model, we use Meinshausen et al.⁹² to compare the scenarios of our model with those of the IPCC. According to Meinshausen et al.⁹², the GWP concentration in 2100 will be 568.96 ppm in the SSP 1-2.6 scenario and 749.91 ppm in the SSP 2-4.5 scenario, which is comparable to our COP26 scenario forecast of 631.60 ppm and the unrestricted scenario forecast of 821.02 ppm, respectively. According to IPCC 2021, GMST under SSP2-4.5 scenario will be approximately 2.7 (2.1–3.5) °C by 2100, which is comparable to our unrestricted scenario projection of 3.28 (2.46–4.10) °C, and 1.8 (1.3–2.4) °C under SSP 1-2.6 scenario, which is close to our COP26 projection of 1.88 (1.43–2.33) °C under COP26 scenario. The GMSL is anticipated to rise by 464–784 mm under SSP2-4.5 scenario and 344–644 mm under SSP1-2.6 scenario in 2100 according to IPCC 2021, which agree well with our projection of 655.25 (556.68–753.82) mm for the unrestricted scenario and 531.23 (471.52–590.95) mm for the COP26 scenario. All estimates are calibrated to our baselines as mentioned before.

Per physics driven forecasts conducted by our group⁸, the GMSL will be 2.3 feet (701.04 mm) by 2100, which is highly consistent with our data driven forecast (655.25 mm) for the unrestricted scenario. The 95% confidence interval of the New York City regional sea level is estimated to range from 1.6 feet (487.68 mm) to 4.1 feet (1249.68 mm) above the baseline based on the physics driven models⁸, while the 95% confidence interval based on our data driven models is 865.99–1125.71 mm, residing completely inside the former confidence interval.

From a climate science point of view, the relationship between temperature and the specific humidity is not linear and outlined by the August–Roche–Magnus (ARM) formula^{93,94} as follows:

$$\begin{aligned} e_s &= 6.1094 \exp\left(\frac{17.625T}{T + 243.04}\right) \\ q &= \beta * RH * e_s / (p - (1 - \beta)e_s), \end{aligned} \quad (7)$$

where e_s is the saturation vapor pressure in hPa, T is the temperature in Celsius degree ($T = \text{GMST} + 14.105$ °C), q is the specific humidity in kg/kg, RH is the relative humidity ranging from 0 to 1, p is the atmospheric pressure in hPa (≈ 1013.25), and β is a constant coefficient (≈ 0.622). Although the ARM formula indicates a nonlinear relationship between temperature (T) and humidity (q), we can prove that this nonlinear relation can be well approximated by a linear relationship between T and q via Taylor Theorem, thus validating the sufficiency of our linear model structure.

According to the Taylor expansion, a function $q(T)$ that is infinitely differentiable at $T = T_0$ is equal to a power series of the form $q(T) = \sum_{i=0}^{\infty} a_i (T - T_0)^i$, where $a_i = \frac{q^{(i)}(T_0)}{i!}$. Therefore, we can write the ARM as:

$$q(T) = a_0 + a_1(T - T_0) + R_2(T). \quad (8)$$

According to the Taylor Theorem, the error term $R_2(T)$ satisfies $R_2(T) = \frac{q^{(2)}(T')}{2!} (T - T_0)^2$ for some T' in the range of T and T_0 . Choosing $T_0 = 14.62$ and using the first two terms as the linear representation $q_{lin}(T)$ of the ARM, we can show that the difference between the ARM formula ($q_{ARM}(T)$) and the linear representation ($q_{lin}(T)$) is:

$$|q_{ARM}(T) - q_{lin}(T)| = |R_2(T)| < 1.94 \times 10^{-4} RH < 1.94 \times 10^{-4}, \quad \text{for } T \in [13.80, 17.50]. \quad (9)$$

The temperature range includes the observed data from 1950 and the 99% prediction intervals of our model. Therefore, a linear representation is sufficient in approximating the non-linear relationship between GMST and Humidity given by the ARM formula. A numeric comparison is provided in the Supplementary Materials.

Reduced form and structural form of the uSEM. There are two equivalent ways to present the unified structural equation model (uSEM), the reduced form and the structural form. The main differences are: (1) In the reduced form, the right-hand (regressor) side features only terms at previous lags such $t-1$, etc.; while the structural form includes both terms at time t , and at the previous lags, $t-1$, etc., on the right-hand side. (2) For the reduced form, the errors of the uSEM are correlated, while for the structural form, the errors are uncorrelated.

In our work, we have shown the structural form for the pathway model, which is more intuitive in understanding the temporal autocorrelations as follows:

$$\begin{cases} Humidity_t = 0.034 + 0.433Humidity_{t-1} + 0.374GMST_{t-1} + \epsilon_{1t} \\ GMST_t = 0.029 - 0.329Humidity_{t-1} + 0.683GMST_{t-1} + 0.443Humidity_t + 0.226GWP_{t-1} + \epsilon_{2t} \end{cases}$$

where $corr(\epsilon_{1t}, \epsilon_{2t}) \approx 0$. We have applied the structural form for prediction as it is easier to compute as well thanks to the uncorrelated error structure. Worth to mention, the negative coefficient of from $Humidity_{t-1}$ to $GMST_t$ is caused by the multi-collinearity between $Humidity_{t-1}$ and $Humidity_t$, the interactive effect between Humidity and GMST are positive overall with both $Humidity_{t-1}$ and $Humidity_t$.

Additionally, $GMST_t$ and $Humidity_t$ affect each other simultaneously actually, which is represented by a correlation between the error terms $\{\epsilon_{1t}, \epsilon'_{2t}\}$ in the equivalent reduced form shown below:

$$\begin{cases} Humidity_t = 0.034 + 0.433Humidity_{t-1} + 0.374GMST_{t-1} + \epsilon_{1t} \\ GMST_t = 0.044 - 0.138Humidity_{t-1} + 0.849GMST_{t-1} + 0.226GWP_{t-1} + \epsilon'_{2t} \end{cases}$$

where the estimated correlation, $corr(\epsilon_{1t}, \epsilon'_{2t}) \approx 0.869$.

Data availability

Datasets containing Global Mean Sea Level matched for this study are available at Climate.gov: https://www.climate.gov/sites/default/files/Climate_dot_gov_dashboard_SeaLevel_Jan2021update.txt and <https://climate.nasa.gov/vital-signs/sea-level/> (Retrieved on May 20th, 2022). Datasets containing Glaciers and Ice Sheets Mass Balance for this study are available at World Glaciers Monitoring Service: <https://wgms.ch/global-glacier-state/>, Datasets underlying the publication: Global Glacier Mass Loss During the GRACE Satellite Mission: <https://doi.org/10.4121/13663433.v1>, and NASA's GRACE and GRACE Follow-On satellites: <https://doi.org/10.5067/TEMSC-3MJ62> (Retrieved on May 20th, 2022). Datasets containing Arctic Sea Ice August Extent for this study are available at National Snow and Ice Data Center: <https://nsidc.org/arcticseaicenews/sea-ice-tools/> and <http://poles.tpdc.ac.cn/en/data/fb3cfd6-9d2f-4664-b30b-c1c41d253ddb/> (Retrieved on May 20th, 2022). Datasets containing Global Mean Surface Temperature for this study are available at Berkeley Earth: <http://berkeleyearth.org/data/> (Retrieved on May 20th, 2022). Datasets containing Global Specific Humidity for this study are available at Copernicus: <https://cds.climate.copernicus.eu#!/home> (Retrieved on Jan 29th, 2022). Datasets containing CO₂ for this study are available at CDIAC <https://cdiac.ess-dive.lbl.gov/trends/co2/lawdome.html> and Scripps CO₂ Program https://scrippsco2.ucsd.edu/data/atmospheric_co2/mlo.html (Retrieved on May 20th, 2022). Datasets containing CH₄ for this study are available at NOAA: <https://www.ncei.noaa.gov/access/paleo-search/study/9959> and https://gml.noaa.gov/ccgg/trends_ch4/ (Retrieved on May 20th, 2022). Datasets containing N₂O for this study are available at NOAA: <https://www.ncei.noaa.gov/access/paleo-search/study/9959> and <https://gml.noaa.gov/hats/combined/N2O.html> (Retrieved on May 20th, 2022). Datasets containing future concentrations for CO₂, CH₄ and N₂O under SSP scenarios at University of Melbourne: <https://greenhousegases.science.unimelb.edu.au#!/ghg?mode=downloads> (Retrieved on Nov. 27th, 2022). Datasets containing Sunspot Number for this study are available at SILSO: <https://www.sidc.be/silso/datafiles> (Retrieved on May 20th, 2022). Datasets for all the Regional Mean Sea Level (RMSL) data is obtained from the Permanent Service for Mean Sea Level (PSMSL): <http://www.psmsl.org/data/obtaining/> (Retrieved on May 27th, 2022). Datasets containing New York City hourly water levels are available at <https://uhslc.soest.hawaii.edu/datainfo/> and Osaka at <https://jdoss1.jodc.go.jp/vpage/tide.html> (Retrieved on Apr 29th, 2022).

Code availability

The R codes used for the analysis and graphics of this paper are available upon request.

Received: 28 August 2022; Accepted: 1 March 2023

Published online: 04 April 2023

References

- Rohde, R. A. & Hausfather, Z. The Berkeley Earth Land/Ocean Temperature Record. *Earth Syst. Sci. Data* **12**, 3469–3479 (2020).
- MEASUREs. Global Mean Sea Level Trend from Integrated Multi-Mission Ocean Altimeters TOPEX/Poseidon, Jason-1, OSTM/Jason-2, and Jason-3 Version 5.1. <https://doi.org/10.5067/GMSLM-TJ151> (2021).
- Heating up. *Nat. Clim. Change* **12**, 693–693. <https://doi.org/10.1038/s41558-022-01462-x> (2022).
- Kim, J., Zhu, W., Chang, L., Bentler, P. M. & Ernst, T. Unified structural equation modeling approach for the analysis of multisubject, multivariate functional MRI data. *Hum. Brain Mapp.* **28**, 85–93 (2007).

5. Tabassum, A., Rabbani, M. & Omar, S. B. An approach to study on time series components and by using them to enumerate the height of sea level alteration for both Global Mean Sea Level (GMSL) and Bay of Bengal (BOB). In *2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT)* 1–7 (IEEE, 2019).
6. Hassan, K. M. A., Haque, M. A. & Ahmed, S. Comparative study of forecasting global mean sea level rising using machine learning. In *2021 International Conference on Electronics, Communications and Information Technology (ICECIT)* 1–4 (IEEE, 2021).
7. Chung, J., Tong, G., Chao, J. & Zhu, W. Path analysis of sea-level rise and its impact. *Stats* **5**, 12–25 (2021).
8. Zhang, M. *Sea-level Rise in New York in the 21st Century: Projection and Methodology* (New York State Resilience Institute for Storms and Emergencies, 2015).
9. Perlin, N., Kamenkovich, I., Gao, Y. & Kirtman, B. P. A study of mesoscale air–sea interaction in the Southern Ocean with a regional coupled model. *Ocean Model* **153**, 101660 (2020).
10. Lindsey, R. & Dahlman, L. Climate change: Global temperature. *Clim. Gov.* **16** (2020).
11. Mahmood, S. *et al.* The impact of time-varying sea surface temperature on UK regional atmosphere forecasts. *Meteorol. Appl.* **28**, e1983 (2021).
12. Hermans, T. H. *et al.* Projecting global mean sea-level change using CMIP6 models. *Geophys. Res. Lett.* **48**, e2020GL092064 (2021).
13. Schewe, J., Levermann, A. & Meinshausen, M. Climate change under a scenario near 1.5 C of global warming: monsoon intensification, ocean warming and steric sea level rise. *Earth Syst. Dyn.* **2**, 25–35 (2011).
14. Washington, W. M. *et al.* How much climate change can be avoided by mitigation?. *Geophys. Res. Lett.* **36**, L08703 (2009).
15. Shannon, S. R. *et al.* Enhanced basal lubrication and the contribution of the Greenland ice sheet to future sea-level rise. *Proc. Natl. Acad. Sci.* **110**, 14156–14161 (2013).
16. Wang, X. L. & Swail, V. R. Climate change signal and uncertainty in projections of ocean wave heights. *Clim. Dyn.* **26**, 109–126 (2006).
17. Wang, X. L., Swail, V. R., Zwiers, F. W., Zhang, X. & Feng, Y. Detection of external influence on trends of atmospheric storminess and northern oceans wave heights. *Clim. Dyn.* **32**, 189–203 (2009).
18. Robinson, A., Calov, R. & Ganopolski, A. Multistability and critical thresholds of the Greenland ice sheet. *Nat. Clim. Change* **2**, 429–432 (2012).
19. Goelzer, H. *et al.* Impact of Greenland and Antarctic ice sheet interactions on climate sensitivity. *Clim. Dyn.* **37**, 1005–1018 (2011).
20. Semtner, A. J. Modeling ocean circulation. *Science* **269**, 1379–1385 (1995).
21. Veeratomy, J., Condon, A. & van Ormondt, M. Forecasting storm surge and inundation: Model validation. *Weather Forecast.* **32**, 2045–2063 (2017).
22. Griffies, S. M. *et al.* OMIP contribution to CMIP6: Experimental and diagnostic protocol for the physical component of the Ocean Model Intercomparison Project. *Geosci. Model Dev.* **9**, 3231–3296 (2016).
23. Nowicki, S. M. J. *et al.* Ice Sheet Model Intercomparison Project (ISMIP6) contribution to CMIP6. *Geosci. Model Dev.* **9**, 4521–4545 (2016).
24. NASA, Nowicki, S. & Seroussi, H. Projections of future sea level contributions from the Greenland and Antarctic ice sheets: Challenges beyond dynamical ice sheet modeling. *Oceanography* **31** (2018).
25. Cramer, W., Yohe, G. *et al.* Detection and attribution of observed impacts. IPCC5 work group 2. *5th Assessment Report* 1–94 (2013).
26. Kurapov, A. L., Erofeeva, S. Y. & Myers, E. Coastal sea level variability in the US West Coast Ocean Forecast System (WCOFS). *Ocean Dyn.* **67**, 23–36 (2017).
27. Etala, P. Dynamic issues in the SE South America storm surge modeling. *Nat. Hazards* **51**, 79–95 (2009).
28. Krivec, T. *et al.* Data-driven method for the improving forecasts of local weather dynamics. *Eng. Appl. Artif. Intell.* **105**, 104423 (2021).
29. Jevrejeva, S. *et al.* Probabilistic sea level projections at the coast by 2100. *Surv. Geophys.* **40**, 1673–1696 (2019).
30. Balogun, A.-L. & Adebisi, N. Sea level prediction using ARIMA, SVR and LSTM neural network: assessing the impact of ensemble Ocean-Atmospheric processes on models' accuracy. *Geomat. Nat. Hazards Risk* **12**, 653–674 (2021).
31. French, J., Mawdsley, R., Fujiyama, T. & Achuthan, K. Combining machine learning with computational hydrodynamics for prediction of tidal surge inundation at estuarine ports. *Procedia IUTAM* **25**, 28–35 (2017).
32. Young, P. C. Data-based mechanistic modelling and forecasting globally averaged surface temperature. *Int. J. Forecast.* **34**, 314–335 (2018).
33. Haddad, M., Taibi, H. & Arezki, S. M. M. On the recent global mean sea level changes: Trend extraction and El Niño's impact. *C. R. Geosci.* **345**, 167–175 (2013).
34. Bagley, J. *et al.* The influence of photosynthetic acclimation to rising CO₂ and warmer temperatures on leaf and canopy photosynthesis models. *Global Biogeochem. Cycles* **29**, 194–206 (2015).
35. Putra, P. T. N., Yuliaty, L. & Lestari, E. K. An empirical study of climate change on carbon dioxide emission in ASEAN 4: Vector autoregression exogenous (VARX) approach. *Eko-Regional: Jurnal Pembangunan Ekonomi Wilayah* **14** (2019).
36. Cherry, J. A., McKee, K. L. & Grace, J. B. Elevated CO₂ enhances biological contributions to elevation change in coastal wetlands by offsetting stressors associated with sea-level rise. *J. Ecol.* **97**, 67–77 (2009).
37. Didenko, N., Skripnuk, D. & Miroyubova, O. Modeling the changes in global temperature due to pollution. *International Multi-disciplinary Scientific GeoConference: SGEM* **17**, 559–567 (2017).
38. Bhargawa, A. & Singh, A. Solar irradiance, climatic indicators and climate change—An empirical analysis. *Adv. Space Res.* **64**, 271–277 (2019).
39. Vermeer, M. & Rahmstorf, S. Global sea level linked to global temperature. *Proc. Natl. Acad. Sci.* **106**, 21527–21532 (2009).
40. Schmith, T., Johansen, S. & Thejll, P. Statistical analysis of global surface temperature and sea level using cointegration methods. *J. Clim.* **25**, 7822–7833 (2012).
41. Karamouz, M., Kia, M. & Nazif, S. Prediction of sea level using a hybrid data-driven model: New challenges after Hurricane Sandy. *Water Qual. Expo Health* **6**, 63–71 (2014).
42. Xie, M. *et al.* Dominant landscape indicators and their dominant areas influencing urban thermal environment based on structural equation model. *Ecol. Ind.* **111**, 105992 (2020).
43. Nieves, V., Radin, C. & Camps-Valls, G. Predicting regional coastal sea level changes with machine learning. *Sci. Rep.* **11**, 7650 (2021).
44. Alomar, M. K. *et al.* Data-driven models for atmospheric air temperature forecasting at a continental climate region. *PLoS One* **17**, e0277079 (2022).
45. Chattopadhyay, A., Mustafa, M., Hassanzadeh, P., Bach, E. & Kashinath, K. Towards physics-inspired data-driven weather forecasting: Integrating data assimilation with a deep spatial-transformer-based U-NET in a case study with ERA5. *Geosci. Model Dev.* **15**, 2221–2237 (2022).
46. Zhongming, Z. *et al.* New Data-Driven Global Climate Model Provides Projections for Urban Environments (2021).
47. Hosseini, R. H., Golian, S. & Yazdi, J. Evaluation of data-driven models to downscale rainfall parameters from global climate models outputs: The case study of Latyan watershed. *J. Water Clim. Change* **11**, 200–216 (2020).
48. Ortiz, L., Mustafa, A., Rosenzweig, B., Carrero, R. & McPhearson, T. Modeling urban futures: Data-driven scenarios of climate change and vulnerability in cities. In *Resilient Urban Futures* (eds Hamstead, Z. A. *et al.*) 129–144 (Springer International Publishing, 2021). https://doi.org/10.1007/978-3-030-63131-4_9.

49. Kaufmann, R. K., Kauppi, H., Mann, M. L. & Stock, J. H. Reconciling anthropogenic climate change with observed temperature 1998–2008. *Proc. Natl. Acad. Sci.* **108**, 11790–11793 (2011).
50. Stone, D. A. & Allen, M. R. Attribution of global surface warming without dynamical models: Attribution of observed warming. *Geophys. Res. Lett.* **32** (2005).
51. Lane, S. T. *Regularized Structural Equation Modeling for Individual-Level Directed Functional Connectivity* (The University of North Carolina at Chapel Hill, 2017).
52. NOAA. How does sea ice affect global climate? *National Ocean Service website* <https://oceanservice.noaa.gov/facts/sea-ice-climate.html> (2021).
53. *Chapter 1. IPCC Fourth Assessment Report* (IPCC, 2008).
54. Kweku, D. W. *et al.* Greenhouse effect: Greenhouse gases and their impact on global warming. *J. Sci. Res. Rep.* **17**, 1–9 (2018).
55. Chahine, M. T. The hydrological cycle and its influence on climate. *Nature* **359**, 373–380 (1992).
56. Zhongming, Z., Linong, L., Xiaona, Y., Wangqiang, Z. & Wei, L. Why we may be able to save the Greenland ice sheet (2018).
57. Kumar, A., Yadav, J. & Mohan, R. Global warming leading to alarming recession of the Arctic sea-ice cover: Insights from remote sensing observations and model reanalysis. *Heliyon* **6**, e04355 (2020).
58. Khan, A. A. Temporary removal: Why would sea-level rise for global warming and polar ice-melt? (2019).
59. Cazenave, A. & Remy, F. Sea level and climate: Measurements and causes of changes. *Wiley Interdiscip. Rev. Clim. Change* **2**, 647–662 (2011).
60. Lennan, M. & Morgera, E. The Glasgow Climate Conference (COP26). *Int. J. Mar. Coast. Law* **37**, 137–151 (2022).
61. Laybourn-Langton, L. & Smith, R. COP26 and health: Some progress, but too slow and not enough: The health community must step up its efforts to hold countries accountable for reducing greenhouse emissions and promoting adaptation. *Int. J. Med. Stud.* **9**, 255–256 (2021).
62. Zhu, Z. *et al.* Glasgow Climate Pact agreed at COP26—but is it enough? (2021).
63. O’Grady, C. Glasgow pact leaves 1.5 °C goal on life support (2021).
64. Schuur, E. A. G. *et al.* Permafrost and climate change: Carbon cycle feedbacks from the warming arctic. *Annu. Rev. Environ. Resour.* **47**, 343–371 (2022).
65. Riahi, K. *et al.* The shared socioeconomic pathways and their energy, land use, and greenhouse gas emissions implications: An overview. *Glob. Environ. Change* **42**, 153–168 (2017).
66. Meinshausen, M. *et al.* Historical greenhouse gas concentrations for climate modelling (CMIP6). *Geosci. Model Dev.* **10**, 2057–2116 (2017).
67. Meinshausen, M. *et al.* The shared socio-economic pathway (SSP) greenhouse gas concentrations and their extensions to 2500. *Geosci. Model Dev.* **13**, 3571–3605 (2020).
68. Franke, J. G. & Donner, R. V. Correlating paleoclimate time series: Sources of uncertainty and potential pitfalls. *Quatern. Sci. Rev.* **212**, 69–79 (2019).
69. Church, J. A. & White, N. J. Sea-level rise from the late 19th to the early 21st century. *Surv. Geophys.* **32**, 585–602 (2011).
70. Caldwell, P., Merrifield, M. & Thompson, P. Sea level measured by tide gauges from global oceans—The Joint Archive for Sea Level Holdings (NCEI Accession 0019568), Version 5.5, NOAA National Centers for Environmental Information, Dataset. *Centers Environ. Information, Dataset* **10**, V5V40S7W (2015).
71. W.G.M.S. Fluctuations of Glaciers Database. World Glacier Monitoring Service. <https://doi.org/10.5904/wgms-fog-2021-05> (2021).
72. Zemp, M. *et al.* Global glacier mass changes and their contributions to sea-level rise from 1961 to 2016. *Nature* **568**, 382–386 (2019).
73. Wouters, B., Gardner, A. S. & Moholdt, G. Global glacier mass loss during the GRACE satellite mission (2002–2016). *Front. Earth Sci.* **7**, 96 (2019).
74. Wiese, D., Yuan, D., Boening, C., Landerer, F. W. & Watkins, M. JPL GRACE and GRACE-FO Mascon Ocean, Ice, and Hydrology Equivalent Water Height Coastal Resolution Improvement (CRI) Filtered Release 06 Version 02 (2019).
75. Kalman, R. E. A new approach to linear filtering and prediction problems (1960).
76. Xiao, C. Autumn sea ice in Barents Kara Sea, Arctic (1289–1993) AD. *National Tibetan Plateau Data Center* <https://doi.org/10.11888/Meteoro.tpd.c.270926> (2019).
77. Fetterer, F., Knowles, K., Meier, W., Savoie, M. & Windnagel, A. Sea Ice Index, Version 3. <https://doi.org/10.7265/N5K072F8> (2017).
78. Hersbach, H. *et al.* ERA5 monthly averaged data on single levels from 1979 to present. *Copernicus Climate Change Service (C3S) Climate Data Store (CDS)* **10**, 252–266 (2019).
79. Etheridge, D. *et al.* Historical CO₂ records from the Law Dome DE08, DE08-2, and DSS ice cores. *Trends: A Compendium of Data on Global Change* 351–364 (1998).
80. Keeling, C. D. *et al.* Atmospheric CO₂ and 13 CO₂ exchange with the terrestrial biosphere and oceans from 1978 to 2000: Observations and carbon cycle implications. In *A History of Atmospheric CO₂ and Its Effects on Plants, Animals, and Ecosystems* 83–113 (Springer, 2005).
81. MacfarlingMeure, C. *et al.* Law dome CO₂, CH₄ and N₂O ice core records extended to 2000 years BP. *Geophys. Res. Lett.* **33**, L14810 (2006).
82. Dlugokencky, E. Global Monitoring Laboratory—Carbon Cycle Greenhouse Gases. *NOAA/GML*.
83. US Department of Commerce. NOAA Global Monitoring Laboratory—Halocarbons and other Atmospheric Trace Species.
84. IPCC (Intergovernmental Panel on Climate Change). Chapter 8: Anthropogenic and natural radiative forcing. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Rep. of the Intergovernmental Panel on Climate Change* (2013).
85. SILSO, W. D. C. Sunspot number and long-term solar observations. *On-line sunspot number catalogue, 2021* (2021).
86. Woodworth, P. & Player, R. The permanent service for mean sea level: An update to the 21st century. *J. Coast. Res.* **19**, 287–295 (2003).
87. Holgate, S. J. *et al.* New data systems and products at the permanent service for mean sea level. *J. Coast. Res.* **29**, 493–504 (2013).
88. Hydrographic Department: Maritime Safety Agency. Japan Oceanographic Data Center (JODC). (Brochure) (1987).
89. Bentler, P. M. & Bonett, D. G. Significance tests and goodness of fit in the analysis of covariance structures. *Psychol. Bull.* **88**, 588–606 (1980).
90. Zou, H. The adaptive lasso and its oracle properties. *J. Am. Stat. Assoc.* **101**, 1418–1429 (2006).
91. MassonDelmotte, V. *et al.* IPCC, 2021: Summary for policymakers. In *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge University Press, 2021).
92. Meinshausen, M. *et al.* The RCP greenhouse gas concentrations and their extensions from 1765 to 2300. *Clim. Change* **109**, 213–241 (2011).
93. Bolton, D. The computation of equivalent potential temperature. *Mon. Weather Rev.* **108**, 1046–1053 (1980).
94. Alduchov, O. A. & Eskridge, R. E. Improved Magnus form approximation of saturation vapor pressure. *J. Appl. Meteorol. Climatol.* **35**, 601–609 (1996).

Author contributions

J.S. and W.Z. conceived and designed the experiments; J.S., G.T., J.Chao and W.Z. collected and analyzed the data and produced the analysis results. All the authors contributed in writing and revising the paper (J.S., G.T., J.Chao, J.Chung, M.Z., W.L., T.Z., P.M.B. and W.Z.).

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-30789-4>.

Correspondence and requests for materials should be addressed to J.S. or W.Z.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023