RESEARCH ARTICLE   **Open Access**

Impact of air–sea coupling on the probability of occurrence of heat waves in Japan

Akira Hasegawa[1]* , Yukiko Imada[2], Hideo Shiogama[1,3], Masato Mori[4], Hiroaki Tatebe[5] and Masahiro Watanabe[1]

Abstract

In extreme event attribution, which aims to answer whether and to what extent a particular extreme weather event can be attributed to global warming, the probability of an event is generally estimated through large ensemble simulations, using an atmospheric general circulation model (AGCM). In islands, such as Japan, it has been considered that surface air temperature (SAT) can be significantly affected by the surrounding sea surface temperature (SST), which mostly is affected by atmospheric circulation at mid- and high-latitudes. Therefore, the absence of SST responses to atmospheric variability in AGCMs impacts the estimation of the occurrence of extreme events, such as heat waves in Japan. In this study, we examined the impact of air–sea coupling on the probability of occurrence of severe heat waves that occurred in Japan in the summer of 2010 by analyzing the probability differences obtained from AGCM and coupled general circulation model (CGCM) large-ensemble experiments. The observed ocean temperature, salinity, and sea ice were assimilated in the 100-member CGCM experiments, as they were assigned as boundary conditions in the 100-member AGCM experiments. The SAT around Japan in the northern summer is largely related to the Bonin high, whose interannual variability is largely affected by the Silk Road and Pacific-Japan (PJ) pattern teleconnections in the real world. The SAT anomaly over Japan was related to the pressure variability due to the Silk Road and PJ patterns in the CGCM experiment. By contrast, the SAT over Japan simulated by AGCM was less sensitive to such pressure variability, and the SAT ensemble spread became narrower in AGCM. The results suggest that the probability of occurrence of the 2010 heat wave in Japan would tend to be underestimated by the AGCM ensemble compared to the CGCM ensemble, provided that the ensemble averages of the SAT anomalies were equal between CGCM and AGCM experiments. This study raised the issue of the absence of SST response to atmospheric variability in AGCMs, which can critically impact the estimation of extreme event probability, particularly in mid-latitude islands, such as Japan.

Keywords: Atmosphere-ocean coupling, Heat waves, Probability density function, Atmospheric internal variability, Silk Road wave train, Pacific-Japan pattern, Event attribution

1 Introduction

Extreme event attribution (EA) has been developed over the past decade to address questions regarding the impact of global warming on specific extreme weather events. Various types of approaches have been used to conduct EA, including observational analyses, model analyses, and multi-method studies (Stott et al. 2016; Easterling et al. 2016; Otto 2017). Shepherd (2016) highlighted two aspects of attribution questions: how anthropogenic climate change has altered the probability of occurrence of individual extreme events (risk-based approach), and how the severity (magnitude or intensity) of a particular event has changed due to climate change (storyline approach). Generally, the risk-based approach is based on two model experiments—for current conditions and a counterfactual world without anthropogenic

emissions (with pre-industrial concentrations of CO_2 and other greenhouse gases, including aerosols)—as described by Allen (2003). By preparing large ensembles to simulate the current climate and a counterfactual climate without anthropogenic climate change, the presumption of the shape of the extremal distribution is not required to compare the probability.

The selection of a numerical model depends on the most relevant type of natural variability for the case analyzed. In many cases, an atmospheric general circulation model (AGCM) has been used for EA studies because the timescale of the most concerning extreme weather events is usually within 1 month. Compared to this timescale, dominant intrinsic oceanic or sea ice variability has a considerably longer timescale, represented by interannual, decadal, and multi-decadal variability. Thus, the state of the ocean or sea ice during an extreme event can be considered the boundary condition for AGCM as a background climate condition of the extreme event in focus. However, in mid- and high-latitude oceans, atmospheric circulation is rather a driver of ocean circulation and considerably affects surface heat fluxes at shorter time scales. The atmospheric model led to unrealistic heat flux at the air–sea interface (Yu and Mechoso 1999) and the absence of important feedback processes (Kitoh and Arakawa 1999). These issues raise the question whether air–sea interaction critically affects the estimation of event probability in the EA approach, particularly for islands surrounded by ocean, such as Japan. Fischer et al. (2018) showed that sea surface temperature (SST)-forced models underestimate the temperature variability, whereas Uhe et al. (2016) indicated that the effect of air–sea interaction on extreme events is considerably smaller. Therefore, whether air–sea coupling critically affects an EA framework remains unclear. It also appears to depend on the location of the focus.

In this study, we propose a new framework to estimate the probability of extreme events, based on a conceptually more realistic approach, which considers both the longer time-scale background situation and the shorter air–sea coupling processes around mid-latitude small islands. It should be noted that a true probabilistic distribution of an observed event remains unknown because the observation is only a single realization among infinite possibilities. Thus, a numerical simulation is needed to deduce the distribution. Figure 1 shows a conceptual illustration explaining the composing elements of the variance of a probability density function (PDF) for a certain extreme index. When we use long-term coupled model intercomparison project (CMIP) type historical simulations (see Section 2) during some decades to draw a PDF (Fig. 1a, assuming single model), its variance is composed of an intrinsic natural variability of the ocean and atmosphere and the climatological tendency in

response to external forcings, such as greenhouse gases and aerosols. When a large ensemble of CMIP-type historical simulations is available, we can draw a PDF using only the data of a certain period (e.g., a certain month of a certain year) when an extreme event occurred (Fig. 1b). In this case, the external forcings are fixed to a specific level, and the variance of the PDF is composed of atmospheric and oceanic (sea ice) natural variabilities.

By contrast, when we use AGCM simulations with fixed boundary conditions for the observed ocean and sea ice focusing on the period of a certain extreme event (Fig. 1d), the variance of its PDF is composed only of atmospheric intrinsic natural variability, which has been the most used in many EA studies. In shorter time scales, which can be perceived by humans, probabilistic behaviors that can trigger extreme weather are induced by atmospheric intrinsic variability, whereas the effect of oceanic fast responses to the atmospheric stochastic variability should also be considered as a component of the probabilistic behaviors in the mid-latitude coastal regions. Given the fixed SST, the coastal air temperature might have a narrower variance. Thus, the conventional approach using AGCM might be inadequate to consider those probabilistic behaviors of the ocean.

In this study, we tested the estimation of extreme event probabilities using large-ensemble simulations of ocean data assimilation with a coupled general circulation model (CGCM) that can consider both short-term oceanic responses to atmospheric variability and prescribed long-term oceanic intrinsic variability (Fig. 1c). By comparing event probability between large ensembles of AGCM and the assimilated CGCM runs, we evaluated the impact of air–sea coupling on the shape of a PDF, particularly on its tail. The difference would be critical to the estimation of a risk ratio in EA studies. In this paper, we will not discuss the impact of air–sea coupling on a risk ratio but focus on the shape of PDFs, thus only using factual simulations.

As an example, we compared the probability of extreme events between large ensemble simulations of AGCM and CGCM, focusing on the extreme warm event that occurred in 2010 in Japanese islands, which are surrounded by the ocean. The surface air temperature (SAT) around Japan in the northern summer is largely related to the Bonin high. Two types of teleconnection patterns—the Pacific-Japan pattern (Nitta 1987; Kosaka and Nakamura 2010) and the Silk Road pattern (Enomoto et al. 2003; Enomoto, 2004)—largely affect the interannual variability of the Bonin high (e.g., Wakabayashi and Kawamura 2004; Yasunaka and Hanawa 2006). Thus, we focused on the representation of the causal relationship between SAT variability around Japan and the Silk Road and PJ teleconnections in AGCM and CGCM simulations.
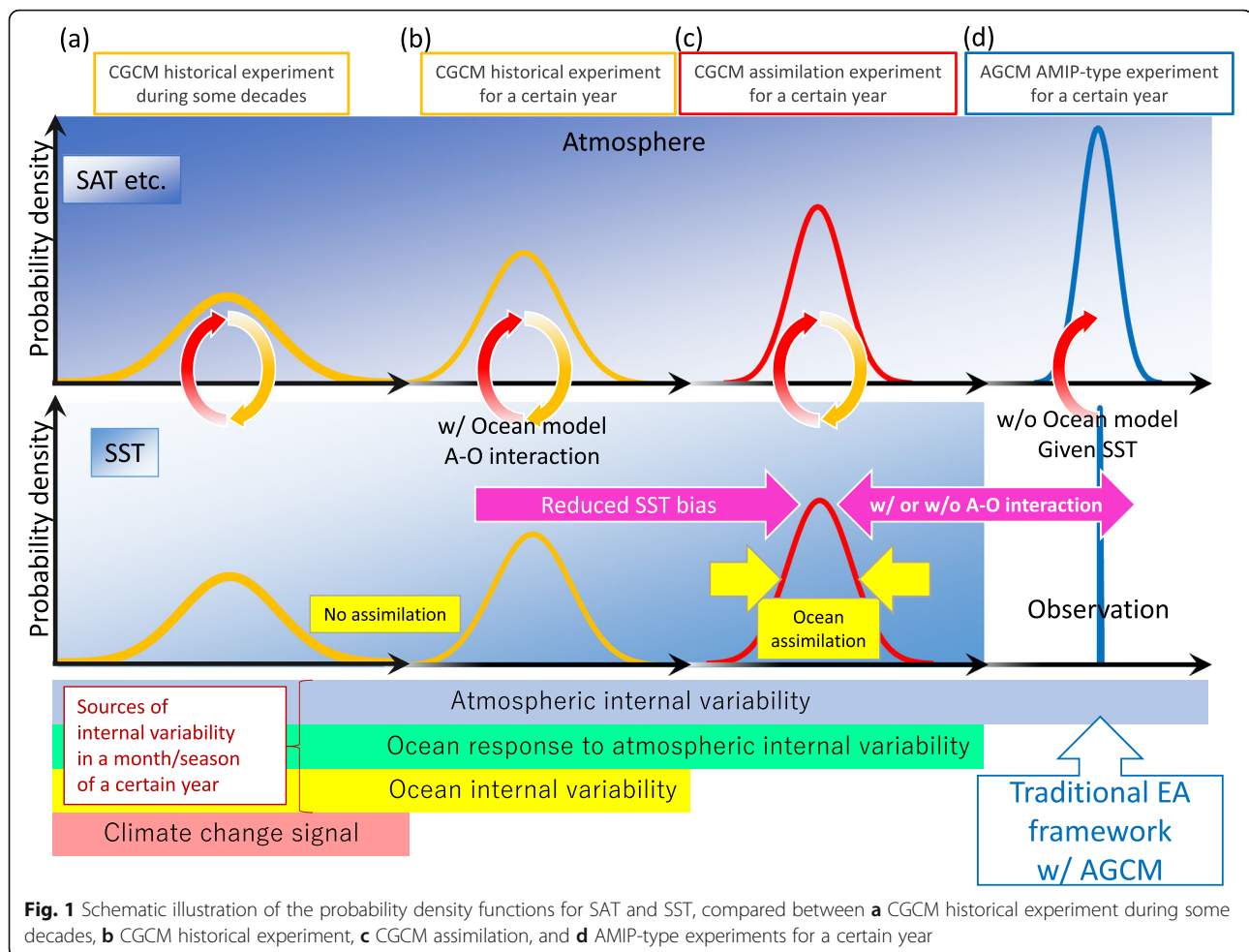
Fig. 1 Schematic illustration of the probability density functions for SAT and SST, compared between **a** CGCM historical experiment during some decades, **b** CGCM historical experiment, **c** CGCM assimilation, and **d** AMIP-type experiments for a certain year

The data, methods, and experimental design are described in the following section. The impact of air–sea coupling on the probability of the Japanese heat wave event in 2010 is described in Section 3. The different contributions of the PJ and Silk Road teleconnections between AGCM and CGCM simulations are discussed in Section 4, and the conclusion is presented in Section 5.

2 Methods and experimental design
We used the sixth version of the Model for Interdisciplinary Research on Climate (MIROC6), described by Tatebe et al. (2019) in more details. MIROC6 is composed of T85L81 atmosphere and 1L62 ocean. It is one of the CMIP phase 6 (CMIP6; Eyring et al. 2016) models.

To estimate the probabilities of extreme events, we prepared two types of large ensemble simulations with the atmospheric component of MIROC6 (MIROC6-AGCM) and the coupled version of MIROC6 (MIROC6-CGCM). Using MIROC6-AGCM, a 20-member ensemble of the Atmospheric Model Intercomparison Project (AMIP)-type experiment was conducted from 1979 to 2014, which was driven by the observed SST, sea ice concentration (COBE-SST2; Hirahara et al. 2014), and historical anthropogenic and natural external forcing agents. The probability of the Japanese heat wave was subsequently evaluated by increasing the ensemble size to 100 only for 2010, which corresponds to a factual simulation of a traditional EA framework.

To reproduce seasonal and interannual variabilities in the CGCM framework, we adopted 10-member full-field assimilation experiments, in which observed ocean temperature, salinity, and sea ice were assimilated for 1951–2014, which were conducted as an initialization process for the CMIP6 Decadal Climate Prediction Project (DCPP, Kataoka et al., 2020. under review). To increase the number of ensemble members from 10 to 100, we reran the full-field assimilation runs from February to December 2010 with 100 different combinations of atmosphere and ocean restart files. This full-field assimilation experiment is termed as CAssm.

For reference, CGCM runs without any data assimilation were also analyzed. A 50-member ensemble of the CMIP6 historical experiment (hereafter denoted as

CHist) was available from 1851 to 2014 as a no-assimilation CGCM case.

The anthropogenic and natural external forcing agents are common among AMIP, CHist, and CAssm experiments. SST and sea ice concentrations assimilated in CAssm are the same as the boundary conditions in AMIP runs.

For verification, we used COBE-SST2 (Hirahara et al. 2014), Global Precipitation Climatology (GPCP; Adler et al. 2003), and a 55 year Japanese reanalysis dataset (JRA-55; Kobayashi et al. 2015).

We defined SST around Japan (SST_{JP}) as the area average of SST in the longitude-latitude box (125–150° E, 25–50° N). The SAT averaged over Japanese islands is denoted as SAT_{JP}.

The Silk Road pattern index is defined as the second mode of empirical orthogonal function (EOF) analysis of detrended 200 hPa geopotential height in the box (30–50° N, 30–130° E) for each 30 year ensemble simulation

and JRA-55. The first mode of EOF corresponded to the Arctic Oscillation-like mode (not shown). The Pacific-Japan pattern index is defined as the second mode of EOF analysis of 850 hPa relative vorticity in the box (0–60° N, 100–160° E). We assumed the multiple linear regression of SAT_{JP} anomaly (y) as the response variable for each teleconnection pattern, as follows:
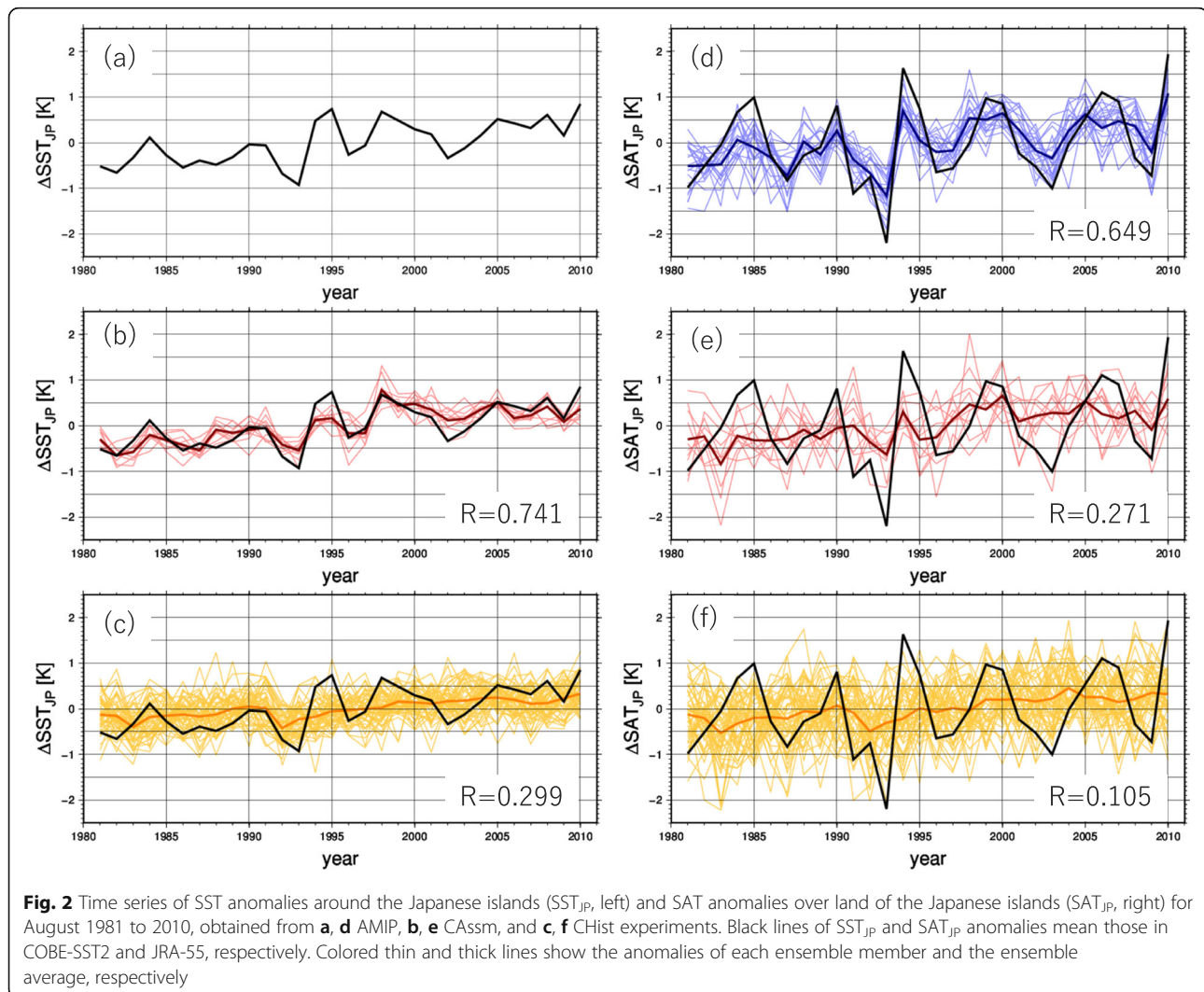
$$y = \beta_0 + \beta_{PJ} \cdot x_{PJ} + \beta_{SR} \cdot x_{SR} + \varepsilon$$

where x_{PJ} and x_{SR} are the PJ and Silk Road pattern indices defined above, β_{PJ} and β_{SR} are multiple regression coefficients for the PJ and Silk Road patterns, and β_0 and ε are the intercept and error terms, respectively.

3 Results

3.1 Thirty-year simulations

Figure 2a–c shows the time-series of SST_{JP} anomalies in August from 1981 to 2010 for AMIP, CAssm, and CHist

Fig. 2 Time series of SST anomalies around the Japanese islands (SST_{JP}, left) and SAT anomalies over land of the Japanese islands (SAT_{JP}, right) for August 1981 to 2010, obtained from **a**, **d** AMIP, **b**, **e** CAssm, and **c**, **f** CHist experiments. Black lines of SST_{JP} and SAT_{JP} anomalies mean those in COBE-SST2 and JRA-55, respectively. Colored thin and thick lines show the anomalies of each ensemble member and the ensemble average, respectively

experiments, respectively. The black line in each panel indicates the SST_{JP} anomaly in COBE-SST2 dataset, whereas the colored thick lines represent ensemble averages, and the colored thin lines represent the 20-, 10-, and 50-member ensembles for AMIP, CAssm, and CHist experiments, respectively. With regards to the ensemble mean response, as shown in Table 1, CAssm shows a relatively high reproductivity of the observed interannual SST_{JP} variability ($R \sim 0.74$) due to the ocean temperature assimilation, compared to CHist ($R \sim 0.30$) without assimilation. Table 1 shows that the averages and standard deviations of SST_{JP} in both CAssm and CHist are relatively lower than those in COBE-SST2, used as a boundary condition for JRA-55 and the AMIP experiment.

Figure 2d–f shows the time-series of SAT_{JP} anomalies in August from 1981 to 2010 for AMIP, CAssm, and CHist experiments, respectively. The black lines indicate the SAT_{JP} anomaly in the JRA-55 reanalysis dataset (Kobayashi et al. 2015). The number of ensemble members in AMIP (20) was twice that of CAssm (10). Nevertheless, SAT_{JP} in the AMIP ensemble shows a relatively smaller variance compared to that in the CAssm ensemble, indicating a relatively small noise in AMIP ensemble.

The ensemble mean response of SAT_{JP} in AMIP showed similar interannual variability as that found in the JRA-55 dataset, as shown in Fig. 2d, whereas CAssm showed a relatively weak coherency with JRA-55, as shown in Fig. 2e. Table 2 shows that the correlation coefficients of SAT_{JP} with JRA-55 are approximately 0.65 and 0.27 for AMIP and CAssm, respectively. Based on the comparison of the correlation coefficients, AMIP seems to be more reproducible. However, to deduce the PDF of extreme events, it can be said that AMIP overestimates a signal-to-noise ratio (underestimates the noise) because of the absence of the intraseasonal responses of the ocean to the atmospheric internal variability, which can affect the probability of SAT_{JP} extremes. Because of the absence of the oceanic intraseasonal responses to the atmosphere, SAT_{JP} in AMIP appeared to be strongly restricted around the observed SST field. Hence, we deduced the range of noise created by the intraseasonal internal variabilities of both atmosphere and ocean to estimate the probability of extreme events from large-ensemble simulations. It is considered that the absence of noises induced by the ocean could lead to overestimation of a signal-to-noise ratio, resulting in a larger

correlation in AMIP. Thus, we assume that CAssm with relatively larger noise and smaller correlation would capture a more realistic probabilistic distribution. By contrast, the interannual variability simulated by CHist is significantly different from that in JRA-55 ($R \sim 0.10$) because of the model-own internal variability, which is free from the long-term observed variation in the ocean and sea ice.

Figure 3a–i shows the SAT, sea level pressure (slp), and 200 hPa geopotential height (z200) anomalies for August 2010 (reference period of climatology: 1981–2010) in JRA-55, AMIP, and CAssm, respectively. Both observed and simulated SAT showed a positive anomaly around Japan in August 2010. On this date, the Japanese islands were covered by positive pressure anomalies at lower (slp) and upper (z200) levels (Fig. 3d–i). In the ensemble mean field (Fig. 3e, f, h, i), the simulated slp anomalies near Japan are off to the south and the wave trains along the subtropical jet are unclear. This is because noise components are compensated, whereas the observed field includes atmospheric noises. When we focus on the one specific member whose PJ and Silk Road indices are similar to those of the observation, the circulation patterns are comparable with JRA-55 (not shown).

Figure 3j shows that GPCP had a relatively low precipitation anomaly around Japan in August 2010. In Fig. 3k and l, low precipitation anomalies over the Japanese islands are also shown in AMIP and CAssm, respectively, but they are unclear. Both AMIP and CAssm represent the dry region on the east side of the Philippines, associated with the cold SST region (not shown) in Fig. 3k and l. The contrast between dry and wet regions of the maritime continent is more enhanced in AMIP than in CAssm and JRA-55. The correlation coefficient between SST and precipitation in August 1981–2010 is mostly positive over the western North Pacific in AMIP, and negative in CAssm and observation (not shown), as indicated by Wang et al. (2005). These differences between AMIP and CAssm indicate that CAssm could reproduce the water and energy cycles more reasonably than AMIP in the Asian-Pacific summer monsoon region.

The climatological error of the simulated SAT_{JP} in August 1981–2010 against the SAT_{JP} in JRA-55 was + 0.13 K for AMIP and CAssm, and − 0.28 K for CHist, as shown in Table 2. The standard deviation of SAT_{JP} for

Table 1 Averages and standard deviations of SST_{JP} for August 1981 to 2010 and correlation coefficients of SST_{JP} with COBE-SST2 for August over the 30 years. Numbers in parentheses represent the number of samples for the 30 years

|	COBE-SST2 (30)	CAssm (300)	CHist (1500)
Average [K]	298.46	298.14	297.47
Standard deviation [K]	0.452	0.420	0.400
Correlation coefficient with COBE-SST2	1.000	0.741	0.299

Table 2 Averages and standard deviations of SAT$_{JP}$ for August 1981–2010 and correlation coefficients of SAT$_{JP}$ with JRA-55 for August over the 30 years. Numbers in parentheses represent the number of samples for the 30 years

	JRA-55 (30)	AMIP (600)	CAssm (300)	CHist (1500)
Average [K]	297.29	297.42	297.42	297.01
Standard deviation [K]	0.919	0.635	0.633	0.639
Correlation coefficient with JRA-55	1.000	0.649	0.271	0.105

August 1981–2010 was approximately 0.92 K in JRA-55 and 0.63 K in AMIP, CAssm, and CHist. Therefore, MIROC6 tended to underestimate the temperature variability than the reanalysis around Japan in August, regardless of the atmosphere–ocean coupling. When comparing the simulation results with the observation or reanalysis, the anomaly from the climatological average

was used to avoid climatological errors in this study. The bias of the interannual variance was not corrected.

Next, we investigated the relationship of the interannual variability of SAT$_{JP}$ in August with the Silk Road and PJ patterns in the AGCM and CGCM simulations. Regression patterns for August 200 hPa geopotential height onto August SAT$_{JP}$, obtained from the 350 year
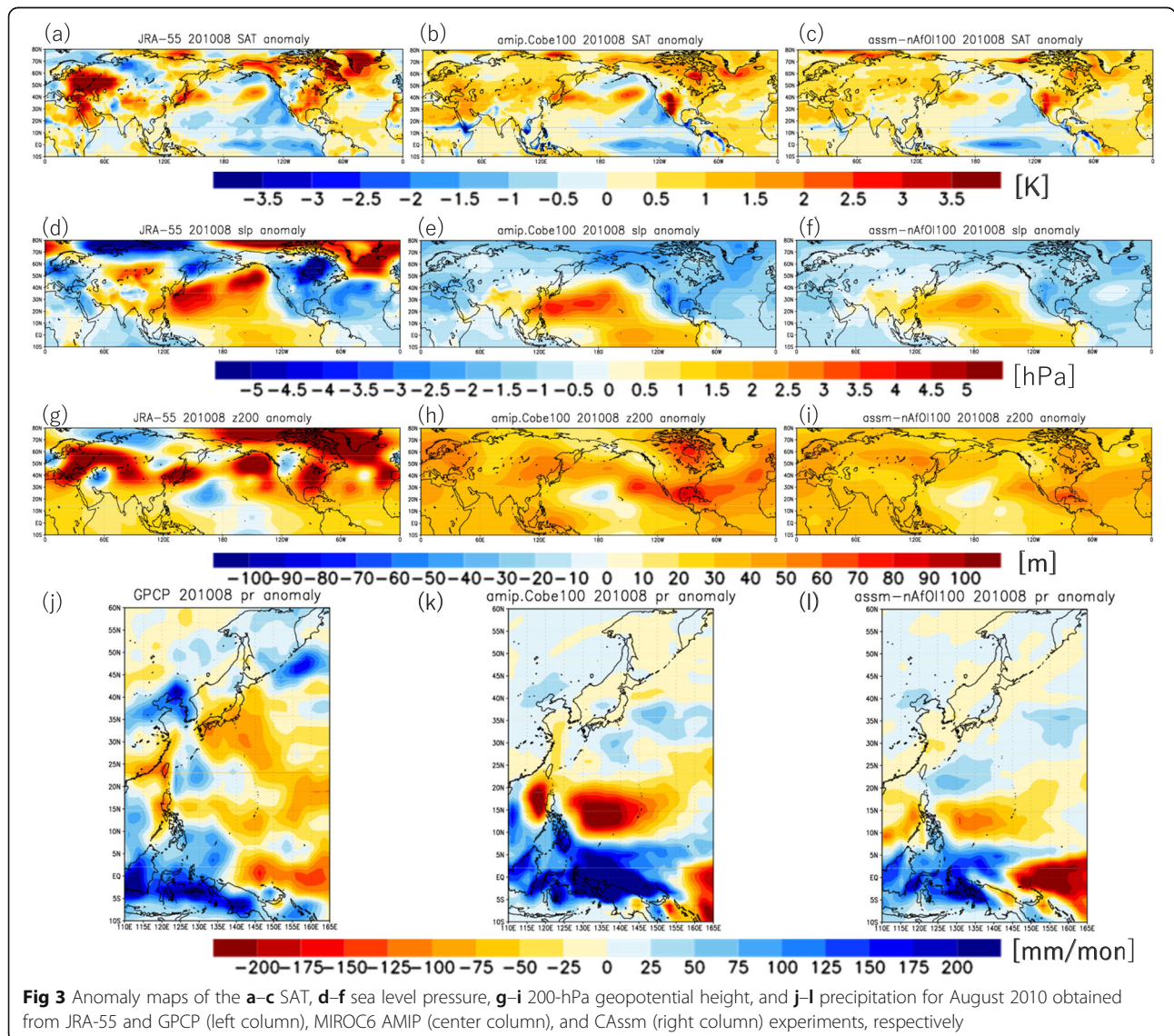
Fig 3 Anomaly maps of the **a–c** SAT, **d–f** sea level pressure, **g–i** 200-hPa geopotential height, and **j–l** precipitation for August 2010 obtained from JRA-55 and GPCP (left column), MIROC6 AMIP (center column), and CAssm (right column) experiments, respectively
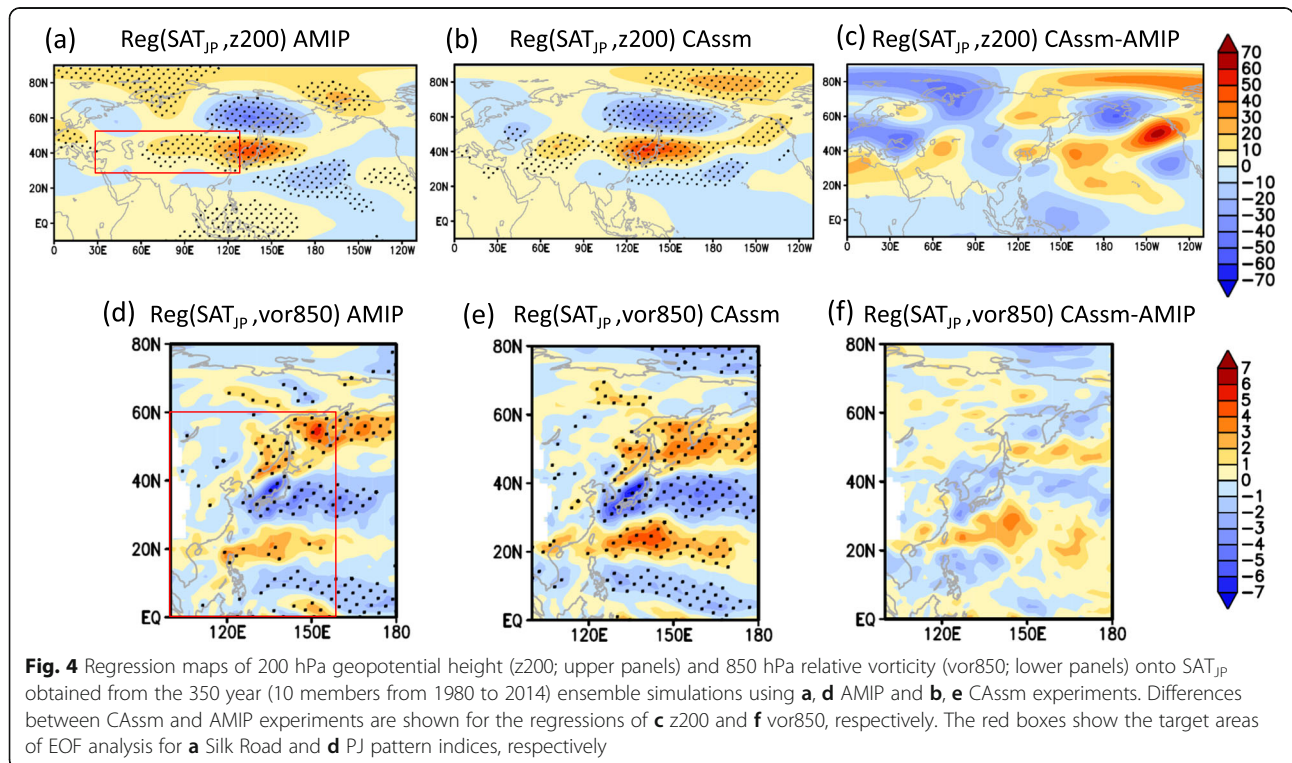
samples (10-member ensemble for 35 years for 1980–2014), showed wave trains over Eurasia forming a positive anomaly over Japan, with a 98% significance level for both AMIP and CAssm, as shown in Fig. 4a and b. It is known that the Tibetan high extends eastward and brings heat waves over Japan, associated with the Silk Road teleconnection (Imada et al. 2019). The difference between Fig. 4a and b shows that the Silk Road pattern is emphasized in CAssm compared to AMIP (Fig. 4c).

By contrast, regression patterns for August 850 hPa relative vorticity (vor850) onto August SAT_{JP} obtained from the 350 year samples showed a negative (anticyclonic) anomaly over Japan and positive (cyclonic) anomalies to the north and south of Japan with a 98% significance level for both AMIP and CAssm, as shown in Fig. 4d and e. This is consistent with the PJ pattern, which is often observed during Japanese high-temperature events (Imada et al. 2019).
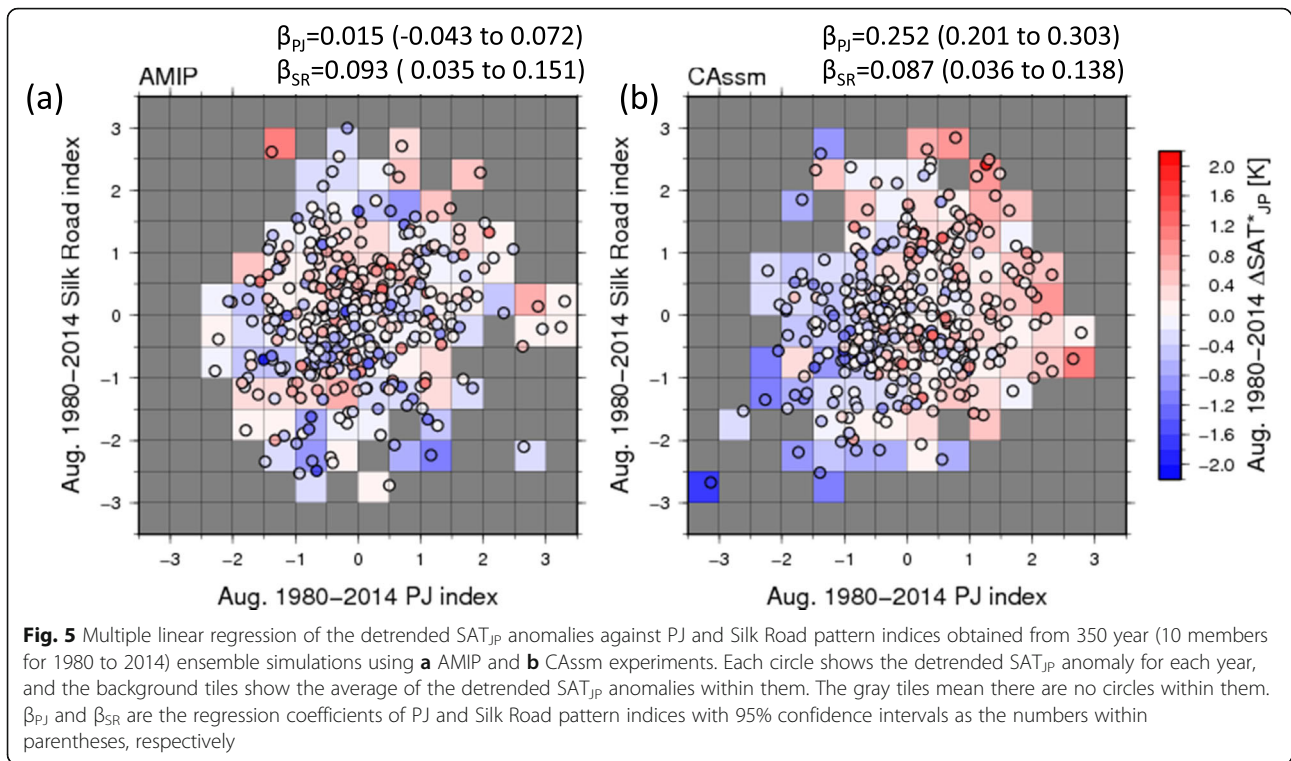
Figure 5 shows the multiple linear regression of the detrended SAT_{JP} (SAT^{*}_{JP}) anomaly in August for the 350 year ensemble simulations of AMIP and CAssm. The horizontal and vertical axes indicate the PJ and Silk Road indices, respectively. The positive (negative) values of horizontal and vertical axes of Fig. 5 mean that the anticyclonic (cyclonic) anomalies cover the Japanese islands, associated with the PJ and Silk Road patterns, respectively. The PJ and Silk Road pattern indices in Fig. 5 are normalized by each standard deviation. The 350

scattered circles show the SAT^{*}_{JP} anomaly colored with a temperature level for each experiment. The background tiles show the average of SAT^{*}_{JP} anomalies within them. Figure 5a shows that SAT^{*}_{JP} is less correlated to the PJ and Silk Road pattern teleconnections in the 350 year ensemble simulations of AMIP. By contrast, hot (cold) SAT^{*}_{JP} cases are located at the upper-right (lower-left) side of the diagram in CAssm (Fig. 5b). Thus, the anticyclonic (cyclonic) anomaly over the Japanese islands, associated with the positive (negative) values of the PJ and Silk Road pattern indices, makes the detrended SAT_{JP} hotter (colder) in August for 1980 to 2014 in CAssm, as we expect in the real world. The multiple regression coefficients β_{PJ} and β_{SR} are listed in Fig. 5 with the 95% confidence intervals as the numbers within parentheses. Although β_{PJ} is not significant in AMIP, it is significantly positive in CAssm for August 1980–2014. The detrended SAT over the Japanese islands is more sensitive to the PJ pattern in CAssm than in AMIP. Nonetheless, β_{SR} is significant but small (~ 0.09) in AMIP and CAssm. The contribution of the Silk Road pattern to SAT^{*}_{JP} is similar regardless of the air–sea coupling.

We also calculated the multiple linear regression of SAT^{*}_{JP} anomaly against both the teleconnection indices in August 1980–2014 using JRA-55. The multiple regression coefficients β_{PJ} and β_{SR} are not significant in JRA-55 dataset. As the number of samples is quite

Fig. 4 Regression maps of 200 hPa geopotential height (z200; upper panels) and 850 hPa relative vorticity (vor850; lower panels) onto SAT_{JP} obtained from the 350 year (10 members from 1980 to 2014) ensemble simulations using **a**, **d** AMIP and **b**, **e** CAssm experiments. Differences between CAssm and AMIP experiments are shown for the regressions of **c** z200 and **f** vor850, respectively. The red boxes show the target areas of EOF analysis for **a** Silk Road and **d** PJ pattern indices, respectively

Fig. 5 Multiple linear regression of the detrended SAT$_{JP}$ anomalies against PJ and Silk Road pattern indices obtained from 350 year (10 members for 1980 to 2014) ensemble simulations using **a** AMIP and **b** CAssm experiments. Each circle shows the detrended SAT$_{JP}$ anomaly for each year, and the background tiles show the average of the detrended SAT$_{JP}$ anomalies within them. The gray tiles mean there are no circles within them. β_{PJ} and β_{SR} are the regression coefficients of PJ and Silk Road pattern indices with 95% confidence intervals as the numbers within parentheses, respectively

smaller in the reanalysis than in large ensemble simulations, the relationship between SAT$^{*}_{JP}$ and the PJ and Silk Road teleconnection patterns is not clear in JRA-55 (figures not shown). The PJ and Silk Road pattern indices are approximately – 0.59 and 1.07 in August 2010 in JRA-55, respectively. Thus, the positive contribution of the Silk Road teleconnection to SAT$_{JP}$ is partly canceled by the negative effect of the PJ pattern in August 2010 in the reanalysis dataset.

3.2 Large ensemble simulations for August 2010
Figure 6a and b shows the PDFs of SAT$_{JP}$ and SST$_{JP}$ anomalies in August 2010 for each experiment, respectively, estimated by the kernel method (Silverman 1986; Kimoto and Ghil 1993).

Table 3 shows the comparison of SST$_{JP}$ for August 2010 between COBE-SST2 and the 100-member CAssm experiments. The 100-member ensemble average of SST$_{JP}$ in CAssm was approximately 0.76 K smaller than the SST$_{JP}$ in COBE-SST2 for August 2010. In CAssm, we can estimate the possible SST variance induced by the atmospheric intrinsic variability using the 100-member ensemble, whereas it is not possible with a single realization of the observation. The standard deviation of SST$_{JP}$ in August 2010 is approximately 0.22 in CAssm. Figure 6b shows the PDFs of SST$_{JP}$ in August 2010 for COBE-SST2, AMIP, and CAssm. The black and blue dashed lines are fixed at approximately 0.86 K because COBE-SST2 is used as a fixed boundary

condition for 100-member AMIP simulations. The exceedance probability of SST$_{JP}$ in August 2010 was approximately 9.1% in CAssm.

Table 4 shows that the 100-member ensemble averages of SAT$_{JP}$ in AMIP and CAssm were approximately 0.79 and 1.26 K smaller than those in JRA-55 for August 2010, respectively. One reason for this is the systematic biases of SAT$_{JP}$ in MIROC6, as shown in Table 2. The other reason is the difference between the observation (including the stochastic noise) and ensemble mean values (without the stochastic noise). The ensemble average of SAT$_{JP}$ in AMIP is higher than that in CAssm in August 2010 because the AMIP simulations are strongly constrained by the observed SST. Table 4 also shows that the standard deviation of SAT$_{JP}$ among the 100-member ensemble was larger in CAssm (approximately 0.49 K) than in AMIP (approximately 0.37 K).

Figure 6a shows the PDFs of SAT$_{JP}$ anomalies in August 2010 for JRA-55, AMIP, and CAssm. The black line at 1.94 K indicates the SAT$_{JP}$ anomaly in JRA-55. The exceedance probabilities over the threshold of 1.94 K were estimated for each experiment (Fig. 6a and Table 4). The exceedance probability was approximately 1.1% and 0.18% in AMIP and CAssm, respectively. The difference in the shape of the PDFs between AMIP (narrow and tall) and CAssm (wide and short) was similar to that expected from Fig. 1. Figure 6a also shows a large difference in the PDF peaks between AMIP and CAssm. These differences imply that the AGCM framework used
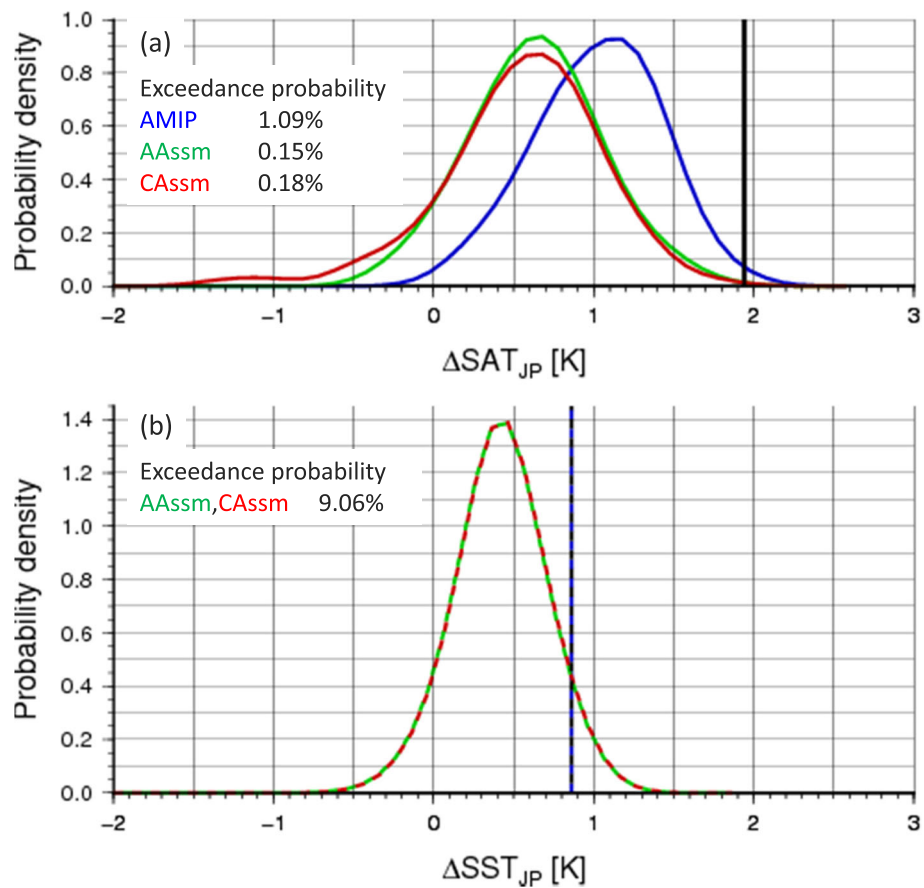
Fig. 6 Probability density functions of **a** SAT$_{JP}$ and **b** SST$_{JP}$ anomalies for August 2010 obtained from CAssm (red), AAssm (green), and AMIP (blue) experiments, estimated by the kernel method (Silverman 1986; Kimoto and Ghil 1993). Vertical lines at approximately **a** 1.94 K and **b** 0.86 K show the SAT$_{JP}$ and SST$_{JP}$ anomalies obtained from JRA-55 and COBE-SST2, respectively

in the conventional EA approach might underestimate the variance of PDF and be strongly constrained by the observed value, resulting in overestimation of the probability of occurrence of an extreme event compared to the CAssm framework, because of the absence of the internal variability associated with the air–sea coupling.

4 Discussion

We could not detect a pure effect of atmosphere–ocean interaction by comparing AMIP and CAssm because the SST boundary condition of AMIP is a

Table 3 Averages and standard deviations of SST$_{JP}$ for August 2010, and exceedance probabilities over ΔSST$_{JP}$ in COBE-SST2 for August 2010 in CAssm. Numbers in parentheses represent the number of samples for August 2010

	COBE-SST2 (1)	CAssm (100)
Average [K]	299.32	298.56
Standard deviation [K]	–	0.223
Exceedance probability [%]	–	9.06

single realization among the possible perturbed SST patterns used in CAssm. The differences in the PDFs between AMIP and CAssm could be strongly affected by different SST patterns (one-way direct impact from SST). To examine the effect of pure air–sea coupling (two-way interaction), we conducted additional 100-member ensemble experiments using AGCM, obtaining the daily boundary conditions of SST and sea ice concentration from the 100-member outputs of CAssm. The additional 100-member experiment is denoted as AAssm in this study.

We compared the PDFs of SAT$_{JP}$ between CAssm and AAssm in Fig. 6. The anomalies in AAssm were defined by the deviation from their climatological averages in CAssm. In Fig. 6a, the shape of the SAT$_{JP}$ PDF curve for CAssm is wider and shorter than that of AAssm. Figure 6a and Table 4 show that the exceedance probability was slightly smaller in AAssm as a result of smaller variance due to the absence of air–sea interaction. Furthermore, the PDF of CAssm showed negative skewness compared to that of AAssm.

Table 4 Averages and standard deviations of SAT$_{JP}$ for August 2010, and exceedance probabilities over ΔSAT$_{JP}$ in JRA-55 for August 2010 in AMIP, AAssm, and CAssm experiments. Numbers in parentheses represent the number of samples for August 2010

	JRA-55 (1)	AMIP (100)	AAssm (100)	CAssm (100)
Average [K]	299.23	298.45	298.06	297.97
Standard deviation [K]	–	0.369	0.390	0.488
Exceedance probability [%]	–	1.09	0.147	0.183

This can be understood from the relation with the PJ and Silk Road pattern indices. Figure 7 shows the multiple linear regression of the SAT$_{JP}$ anomalies from the 100-member ensemble averages (δSAT$_{JP}$) against the PJ and Silk Road pattern indices for August 2010 of AMIP, AAssm, and CAssm, as shown in Fig. 5. The PJ and Silk Road pattern indices for both AAssm and CAssm in Fig. 7 are commonly normalized by the standard deviation for the 350 year ensemble simulations of CAssm, while those for AMIP in Fig. 7a is normalized by the standard deviation for the 350 year ensemble simulations of AMIP.

Figure 7a and b shows that the SAT$_{JP}$ anomaly fluctuates in AMIP and AAssm regardless of the pressure variability over Japan due to PJ and Silk Road pattern variabilities in August 2010, which are consistent with Fig. 5a. Both regression coefficients β_{PJ} and β_{SR} are not significant to the 100-member AAssm experiment in August 2010, as shown in Fig. 7b. By contrast, Fig. 7c shows that SAT$_{JP}$ becomes higher (lower) under the anticyclonic (cyclonic) anomaly over the Japanese islands associated with the positive (negative) teleconnection indices in August 2010, which is consistent with Fig. 5b. Both regression coefficients β_{PJ} and β_{SR} are positive with 95% significance level in August 2010 in CAssm, as shown in Fig. 7c. SAT$_{JP}$ variability associated with the PJ and Silk Road teleconnection patterns is reasonably reproduced in CGCM simulations.

The SAT$_{JP}$ variance among the 100-member ensemble was larger in CAssm than in AAssm for August 2010. The standard deviations are approximately 0.39 and 0.49 for AAssm and CAssm, respectively, as shown in Table 4. Figure 6a shows that the enhanced variance of SAT$_{JP}$ in CAssm is mainly due to the lower SAT$_{JP}$ cases, which are located at the lower-left side in Fig. 7c. Therefore, the SAT variance over Japan could be increased in association with the increase of negative skewness induced by atmospheric teleconnection patterns and atmosphere–ocean coupling. Because anticyclonic anomalies change the surface temperature through the adiabatic heating and less cloudiness, the nonlinear variation of cloud covers is one of the reasons of the negative skewness.

5 Conclusions

This study investigated the impact of air–sea interaction on the probability of occurrence of extreme events in mid-latitude small islands surrounded by the ocean. We compared the event probability, estimated by AGCM and CGCM large-ensemble experiments, of the heat waves that occurred in Japan in August 2010. The observed ocean temperature, salinity, and sea ice were assimilated into 100-member CGCM experiments. We compared the CGCM assimilation ensemble (CAssm) for August 2010 with two types of 100-member AGCM

Fig. 7 Multiple linear regression of SAT$_{JP}$ anomalies against PJ and Silk Road pattern indices for August 2010 from **a** AMIP, **b** AAssm, and **c** CAssm experiments. β_{PJ} and β_{SR} are the regression coefficients of PJ and Silk Road pattern indices with 95% confidence intervals as the numbers within parentheses, respectively
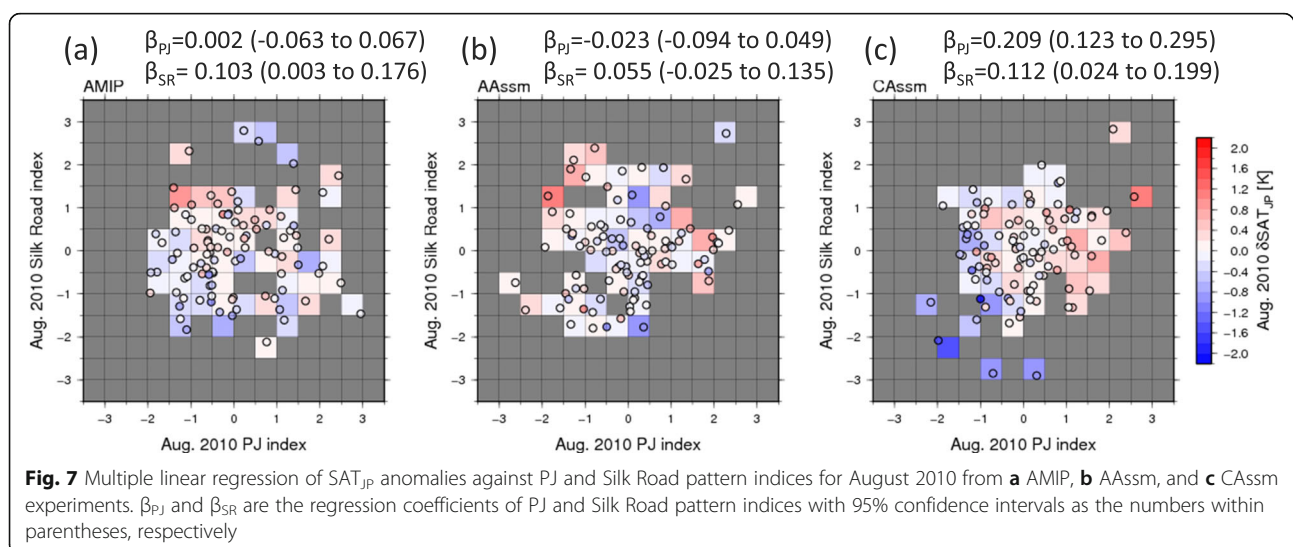
experiments: the AGCM ensemble with the boundary conditions from the CGCM assimilation experiment (AAssm), and the AMIP-type ensemble with the single boundary condition from the observation (AMIP). The AMIP-type ensemble has generally been used in the study of event attribution.

We can interpret the difference of SAT anomaly around Japan between AMIP and CAssm experiments, illustrated in Fig. 1c and d, by decomposing the difference into two parts: one between AMIP and AAssm, and the other between AAssm and CAssm experiments.

One difference between AMIP and AAssm is the ensemble average shift of the SAT anomalies due to the SST distribution difference between the observation and assimilated fields. The ensemble-averaged intensity of the Japanese heat wave in 2010 in AAssm becomes smaller than that in AMIP, as shown in Fig. 6a. The other difference between AMIP and AAssm is the ensemble variance of the SAT due to the different variance of the SST field. The SAT ensemble spread under the 100-kind boundary conditions in AAssm becomes larger than that under the single boundary condition in AMIP. These are the main sources of the difference in the SAT PDFs between the two types of AGCM experiments in Fig. 6, and both differences depend on the assimilation intensity in the CGCM.

The difference between AAssm and CAssm is induced by the pure effect of the air–sea coupling. The assimilated CGCM experiment can reproduce the SAT variability due to the pressure variability related with the Silk Road and PJ pattern teleconnections, as expected in the real world. Such a mechanism could expand the ensemble spread of the SAT over Japan under the air–sea coupled condition. However, the absence of air–sea coupling in AAssm could distort the response mechanism of the SAT over Japan to such atmospheric internal variability. Even if the SST fields are common between AGCM and CGCM ensembles, the ensemble spread of the SAT over Japan in August 2010 is reduced in the AGCM experiment. Note that we did not investigate whether the air–sea coupling modified the PJ and Silk Road pattern teleconnections themselves in this study.

The results showed that the ensemble spread of the SAT over Japan in the CGCM experiment was larger than those in the two types of AGCM experiments for August 2010. If the ensemble average of the SAT anomaly is equal in CGCM and AGCM ensembles, the probability of occurrence of the heat wave over Japan in August 2010 could be estimated to be smaller by AMIP ensemble, compared to CAssm. However, as the ensemble average of the SAT anomaly was large in AMIP, the probability of occurrence of the heat wave was estimated to be smaller in CAssm ensemble, compared to AMIP.

Further analysis suggested that the SAT anomaly over Japan was well related to the pressure variability due to the Silk Road and PJ pattern indices in the CGCM assimilation experiment, as reported for the real world. By contrast, the simulated SAT over Japan by AGCM was less sensitive to these atmospheric internal variabilities, and the ensemble spread became narrower in the AGCM experiment.

In many extreme EA studies, the probability of an event has been estimated using large ensemble simulations performed by AGCMs. This study raised the issue of the absence of SST response to atmospheric variability in AGCMs, which can critically impact the estimation of extreme event probability, particularly in mid-latitude islands, such as Japan. The new framework using the CGCM assimilation system proposed in this study conceptually provides a more realistic probability distribution. Our next step is to produce a counterfactual ensemble without anthropogenic climate change by applying the CGCM assimilation system and verify the impact of air–sea coupling on the estimated probability ratio from the two ensembles.

Authors' contributions
AH performed the experimental study and analyzed the data. YI proposed and designed the study and analyzed the data. HS proposed and designed the study. MM prepared the MIROC6 AGCM system and related datasets. HT created the full-field assimilation system of MIROC6 and the long-term assimilation simulation. MW assisted in its interpretation. All authors have read and approved the final manuscript.

Authors' information
MW serves as the area representative of Area Theme A "Prediction and Projection of Large-Scale Climate Changes Based on Advanced Model Development" under TOUGOU program (Integrated Research Program for Advancing Climate Models). HT serves as the representative of subject (i)-a "Near-future climate change predictions and promotion of CMIP6 experiments" in Area Theme A of TOUGOU program. MW also serves as the representative of subject (ii)-b "Analysis of factors in past climate changes and unusual weather and future projections" in Area Theme A. YI, HS, MM, and AH have been studying event attribution under the subject (ii)-b of Area Theme A of TOUGOU program.

Author details
[1]Atmosphere and Ocean Research Institute, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8564, Japan. [2]Meteorological Research Institute, Japan Meteorological Agency, Tsukuba, Ibaraki, Japan. [3]Center for Global Environmental Research, National Institute for Environmental Studies, Tsukuba, Ibaraki, Japan. [4]Research Institute for Applied Mechanics, Kyushu University, Fukuoka, Japan. [5]Research Center for Environmental Modeling and Application, Japan Agency for Marine-Earth Science and Technology, Yokohama, Kanagawa, Japan.

References
Adler RF, Huffman GJ, Chang A, Ferraro R, Xie P-P, Janowiak J, Rudolf B, Schneider U, Curtis S, Bolvin D, Gruber A, Susskind J, Arkin P, Nelkin E (2003) The Version-2 Global Precipitation Climatology Project (GPCP) Monthly Precipitation Analysis (1979-Present). J Hydrometeorol 4:1147–1167

Allen MR (2003) Liability for climate change. Nature 421:891–892

Easterling DR, Kunkel KE, Wehner MF, Sun L (2016) Detection and attribution of climate extremes in the observed record. Weather Clim Extremes 11:17–27. https://doi.org/10.1016/j.wace.2016.01.001

Enomoto T (2004) Interannual variability of the Bonin high associated with the propagation of Rossby waves along the Asian jet. J Meteor Soc Jpn 82:1019–1034. https://doi.org/10.2151/jmsj.2004.1019

Enomoto T, Hoskins BJ, Matsuda Y (2003) The formation mechanism of the Bonin high in August. Quart J Roy Meteor Soc 129:157–178. https://doi.org/10.1256/qj.01.211

Eyring V, Bony S, Meehl GA, Senior CA, Stevens B, Stouffer RJ, Taylor KE (2016) Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. Geosci Model Dev 9:1937–58. https://doi.org/10.5194/gmd-9-1937-2016

Fischer EM, Beyerle U, Schleussner CF, King AD, Knutti R (2018) Biased estimates of changes in climate extremes from prescribed SST simulations. Geophys Res Lett 45:8500–8509. https://doi.org/10.1029/2018GL079176

Hirahara S, Ishii M, Fukuda Y (2014) Centennial-scale sea surface temperature analysis and its uncertainty. J Climate 27:57–75. https://doi.org/10.1175/jcli-d-12-00837.1

Imada Y, Watanabe M, Kawase H, Shiogama H, Arai M (2019) The July 2018 high temperature event in Japan could not have happened without human-induced global warming. SOLA 15A:8–12. https://doi.org/10.2151/sola.15A-002

Kataoka T, Tatebe H, Koyama H, Mochizuki T, Ogochi K, Naoe H, Imada Y, Shiogama H, Kimoto M, Watanabe M (2020) Seasonal to decadal predictionswith MIROC6: Description and basic evaluation. J Adv Modeling Earth Systems. https://doi.org/10.1029/2019MS002035

Kimoto M, Ghil M (1993) Multiple flow regimes in the Northern Hemisphere winter. Part I: Methodology and hemispheric regimes. J Atmos Sci 50:2625–2643

Kitoh A, Arakawa O (1999) On overestimation of tropical precipitation by an atmospheric GCM with prescribed SST. Geophys Res Lett 26:2965–2968

Kobayashi S, Ota Y, Harada Y, Ebita A, Moriya M, Onoda H, Onogi K, Kamahori H, Kobayashi C, Endo H, Miyaoka K, Takahashi K (2015) The JRA-55 reanalysis: General specifications and basic characteristics. J Meteor Soc Jpn 93:5–48. https://doi.org/10.2151/jmsj.2015-001

Kosaka Y, Nakamura H (2010) Mechanisms of meridional teleconnection observed between a summer monsoon system and a subtropical anticyclone. Part I: the Pacific-Japan pattern. J Climate 23:5085–5108. https://doi.org/10.1175/2010JCLI3413.1

Nitta T (1987) Convective activities in the tropical western Pacific and their impact on the Northern Hemisphere summer circulation. J Meteor Soc Jpn 65:373–390. https://doi.org/10.2151/jmsj1965.65.3_373

Otto FEL (2017) Attribution of weather and climate events. Annu Rev Env Resour 42:627–646. https://doi.org/10.1146/annurev-environ-102016-060847

Shepherd TG (2016) A common framework for approaches to extreme event attribution. Curr Climate Change Rep 2:28–38. https://doi.org/10.1007/s40641-016-0033-y

Silverman BW (1986) Density estimation for statistics and data analysis. Monographs on Statistics and Applied Probability, Vol. 26, Chapman andHall, London. p 175

Stott PA, Christidis N, Otto FEL, Sun Y, Vanderlinden J-P, van Oldenborgh GJ, Vautard R, von Storch H, Walton P, Yiou P, Zwiers FW (2016) Attribution of extreme weather and climate-related events. WIREs Clim Change 7:23–41. https://doi.org/10.1002/wcc.380

Tatebe H, Ogura T, Nitta T, Komuro Y, Ogochi K, Takemura T, Sudo K, Sekiguchi M, Abe M, Saito F, Chikira M, Watanabe S, Mori M, Hirota N, Kawatani Y, Mochizuki T, Yoshimura K, Takata K, O'ishi R, Yamazaki D, Suzuki T, Kurogi M, Kataoka T, Watanabe M, Kimoto M (2019) Description and basic evaluation of simulated mean state, internal variability, and climate sensitivity in MIROC6. Geosci Model Dev 12:2727–2765. https://doi.org/10.5194/gmd-12-2727-2019

Uhe P, Otto FEL, Haustein K, van Oldenborgh GJ, King AD, Wallom DCH, Allen MR, Cullen H (2016) Comparison of methods: Attributing the 2014 record European temperatures to human influences. Geophys Res Lett 43:8685–8693. https://doi.org/10.1002/2016GL069568

Wakabayashi S, Kawamura R (2004) Extraction of major teleconnection patterns possibly associated with the anomalous summer climate in Japan. J Meteor Soc Jpn 82:1577–1588. https://doi.org/10.2151/jmsj.82.1577

Wang B, Ding Q, Fu X, Kang I-S, Jin K, Shukla J, Doblas-Reyes F (2005) Fundamental challenge in simulation and prediction of summer monsoon rainfall. Geophys Res Lett 32:L15711. https://doi.org/10.1029/2005GL022734

Yasunaka S, Hanawa K (2006) Interannual summer temperature variations over Japan and their relation to large-scale atmospheric circulation field. J Meteor Soc Jpn 84:641–652. https://doi.org/10.2151/jmsj.84.641

Yu J-Y, Mechoso CR (1999) A discussion on the errors in the surface heat fluxes simulated by a coupled GCM. J Climate 12:416–426