Clim Dyn
DOI 10.1007/s00382-013-1967-6

What spatial scales are believable for climate model projections of sea surface temperature?

Lester Kwiatkowski · Paul R. Halloran · Peter J. Mumby · David B. Stephenson

Received: 13 May 2013 / Accepted: 8 October 2013

Abstract Earth system models (ESMs) provide high resolution simulations of variables such as sea surface temperature (SST) that are often used in off-line biological impact models. Coral reef modellers have used such model outputs extensively to project both regional and global changes to coral growth and bleaching frequency. We assess model skill at capturing sub-regional climatologies and patterns of historical warming. This study uses an established wavelet-based spatial comparison technique to assess the skill of the coupled model intercomparison project phase 5 models to capture spatial SST patterns in coral regions. We show that models typically have medium to high skill at capturing climatological spatial patterns of SSTs within key coral regions, with model skill typically improving at larger spatial scales ($\geq 4°$). However models have much lower skill at modelling historical warming patters and are shown to often perform no better than chance at regional scales (e.g. Southeast Asian) and worse than chance at finer scales ($<8°$). Our findings suggest that output from current generation ESMs is not yet suitable for making sub-regional projections of change in coral bleaching frequency and other marine processes linked to SST warming.

Keywords CMIP5 · Coral bleaching · Ocean resolution · Spatial scales · Haar wavelets · Regional modelling

1 Introduction

Accurately simulating the coastal zones represents a significant challenge for ESMs due to the complex local physics, biogeochemical and biophysical interactions in these regions, driven by strong bathymetric constraints on circulation, and the impacts of terrestrial and sedimentary geochemical fluxes (Holt et al. 2009; Allen et al. 2010).

Earth system models (ESMs) are global circulation models (GCMs) coupled to submodels representing, for example, the ocean and terrestrial carbon cycle, global vegetation dynamics and atmospheric chemistry (e.g. Collins et al. 2011). Over the past 10 years there has been a proliferation of papers that have used GCM and more recently ESM outputs to make historical and future projections of coral reefs (e.g. Donner et al. 2005; Frieler et al. 2013; Kwiatkowski et al. 2013). Coral reefs have become a focal point of climate science because experimental and observational work has demonstrated high sensitivity to anomalous SSTs (e.g. Hoegh-Guldberg 1999; Eakin et al. 2010) and future projections have indicated that globally these ecosystems could be impacted even under the most conservative climate scenarios (Frieler et al. 2013).

Corals form an intimate association with dinoflagellate algae called zooxanthellae that live and photosynthesize in the coral's tissue. When sunlight is high it overwhelms the

Electronic supplementary material The online version of this article (doi:10.1007/s00382-013-1967-6) contains supplementary material, which is available to authorized users.

L. Kwiatkowski (✉) · P. R. Halloran · D. B. Stephenson
College of Engineering, Mathematics and Physical Sciences, University of Exeter, Exeter EX4 4QF, UK
e-mail: lk244@exeter.ac.uk

P. R. Halloran
Hadley Centre, Met Office, Exeter EX1 3PB, UK

P. J. Mumby
School of Biological Sciences, University of Queensland, St. Lucia, Brisbane, QLD 4072, Australia

Published online: 18 October 2013

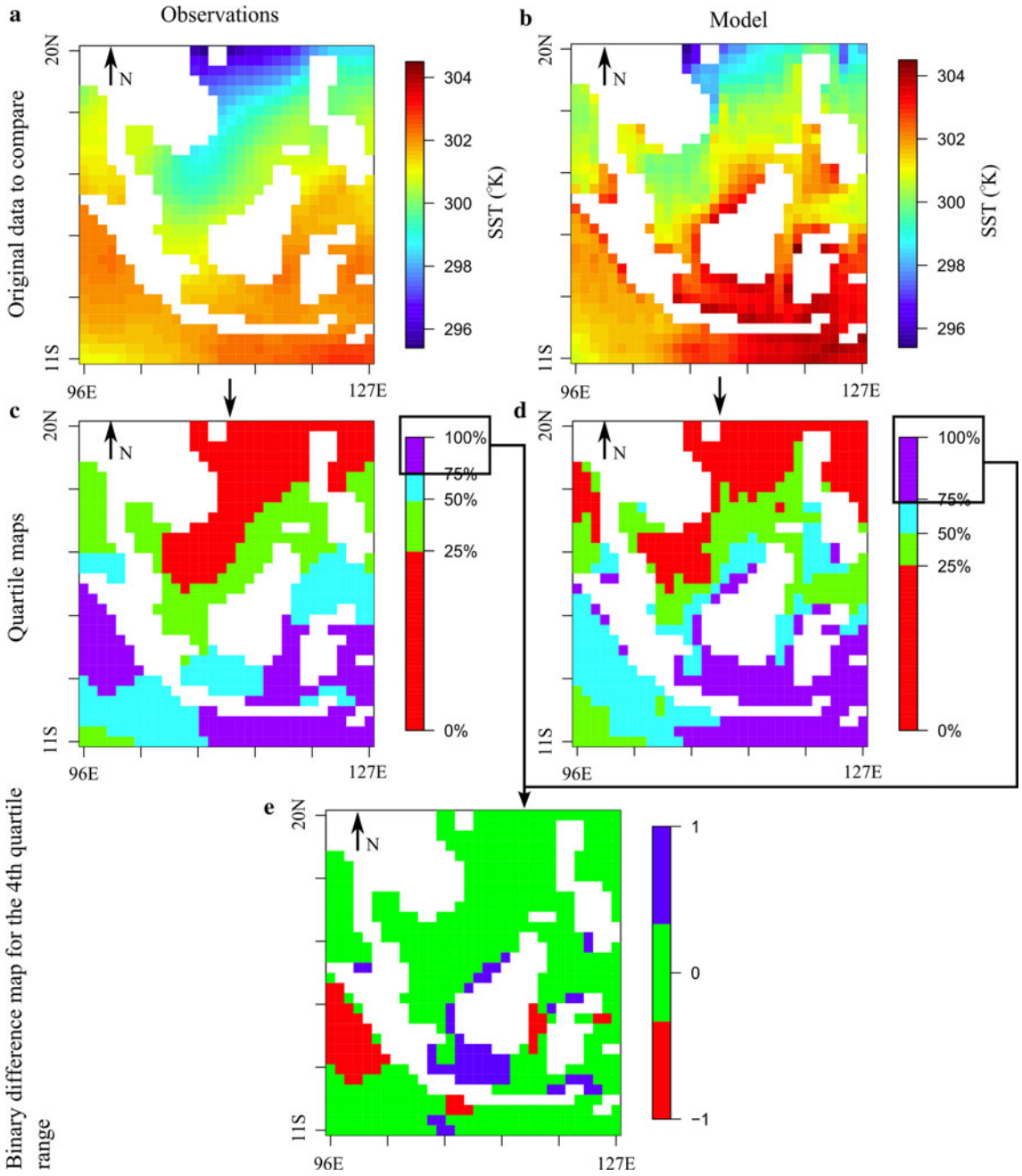

Fig. 1 Binary difference map creation. At the *top* **a** HadISST and **b** regridded HadGEM2-ES SSTs for July from the respective 1985–2000 climatologies. In the *centre* quartile range maps of the same fields for **c** HadISST and **d** HadGEM2-ES. On the *bottom* **e** binary difference map for the uppermost quartile range. *Green* areas in the binary difference map represent areas of agreement between the uppermost quartile range maps. *Blue* and *red* areas represent areas of disagreement. Figure adapted from Saux-Picart et al. (2012)

capacity of the zooxanthellae to photosynthesize, causing photoinhibition. A number of biological systems are available to cope with the excess excitation of the photosystem (Brown 1997). However, when sea temperatures become exceptionally warm, often by only a matter of 0.5–1 °C (Edwards et al. 2001), less light is needed to over-excite the photosystem (Enríquez et al. 2005) and it becomes more likely that the over-excitation will overwhelm the mitigative processes and result in the release of dangerous oxygen free radicals which can damage the symbiont complex and result in a loss of zooxanthellae. The loss of zooxanthellae turns the coral white (hence termed "coral bleaching"), and can result in mass coral mortality within a period of weeks and at regional to global scales. The SSTs associated with the 1998 El Niño event caused massive coral mortality in the

Indian Ocean (Edwards et al. 2001), Pacific Ocean (Mumby et al. 2001) and tropical West Atlantic (Kramer et al. 2003).

Most model studies have used SST simulations to project coral bleaching in both individual coral regions like the Indian Ocean (Sheppard 2003) and Hawaii (Hoeke et al. 2011), as well as at the global scale (e.g. Donner et al. 2005; Donner 2009; Frieler et al. 2013). These studies have used global model outputs in shallow coastal areas (the location of most coral regions) at, or close to, the model's maximum spatial resolution (typically one degree of latitude by one degree of longitude). They have also concluded that such outputs provide high resolution evidence of coral vulnerability to bleaching within a region (e.g. Van Hooidonk et al. 2013 conclude that central French Polynesia will be less susceptible to coral bleaching than other parts of the Polynesian region). There is, however, a potential spatial mismatch between the small spatial scales at which coral projections are made and the large spatial scales at which GCM/ESM outputs are likely to be most reliable. We explore the notion of believable scales (Lander and Hoskins 1997) in order to determine the spatial scales at which we have the greatest confidence in using such climate models to make coral projections.

Table 1 The CMIP5 models analysed, their institutions and their original ocean resolution

Model name	Model institution	Ocean resolution (latitude° × longitude°)
CCSM4	National Center for Atmospheric Research, USA	0.25–0.5 × 1.125
CNRM-CM5	Centre National de Recherches Météorologiques/ Centre Europeen de Recherche et Formation vancees en Calcul Scientifique, France	0.2–0.7 × 1
CSIRO-Mk3.6.0	Commonwealth Scientific and Industrial Research Organization/ Queensland Climate Change Centre of Excellence, Australia	0.93–0.94 × 1.875
GFDL-ESM-2G	Geophysical Fluid Dynamics Laboratory, USA	0.3–1 × 1
HadGEM2-ES	Met Office Hadley Centre, UK	0.3–1 × 1
INM-CM4	Institute for Numerical Mathematics, Russia	0.3–0.5 × 0.6–1
IPSL-CM5A-LR	Institut Pierre Simon Laplace, France	0.2–1 × 1–2
IPSL-CM5A-LR	Institut Pierre Simon Laplace, France	0.2–1 × 1–2
MIROC5	Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute, National Institute for Environmental Studies, Japan	0.5–0.8 × 0.3–1.3
MPI-ESM-LR	Max Planck Institute for Meteorology, Germany	0.2–1.4 × 0.2–0.6
MRI-CGCM3	Meteorological Research Institute, Japan	0.5 × 1
NorESM1-M	Norwegian Climate Centre, Norway	0.25–0.5 × 1.125

Projections of coral bleaching typically involve the calculation of "degree heating months" (DHM). A DHM is equal to 1 month of SST that is 1 °C greater than the maximum monthly mean (MMM) SST taken from a historical climatology for a given grid cell. The annual accumulation of DHM for a given year is then calculated as the maximum 4 consecutive month accumulation of DHM in a given year (Frieler et al. 2013). Although there have been attempts to quantify the skill of bleaching algorithms (e.g. van Hooidonk and Huber 2009) the skill of model SST outputs, in coral regions, at the resolution of the latest CMIP models, has not been adequately assessed. One of the advantages of assessing SSTs instead of directly assessing bleaching projections is the availability of a far greater observational resource against which models can be tested. We use a local wavelet technique, developed by Casati et al. (2004) for verifying spatial precipitation forecasts, to assess the skill of CMIP5 models in capturing the SST features critical to coral bleaching. The analysis is based on multi-model means, which invariably perform better than individual models (Palmer et al. 2005). The analysis presented here informs the spatial-resolution at which CMIP5 and earlier generation climate models should be used to project coral futures, and indeed whether global models are adequate tools to address the questions being asked of coral scientists by policy makers.

2 Methods

2.1 Conceptual overview

Wavelet theory involves representing a signal, for example a sound or an image, in terms of simpler fixed "building blocks" at different scales and positions (Jawerth and Sweldens 1994). These "building blocks" or wavelets can

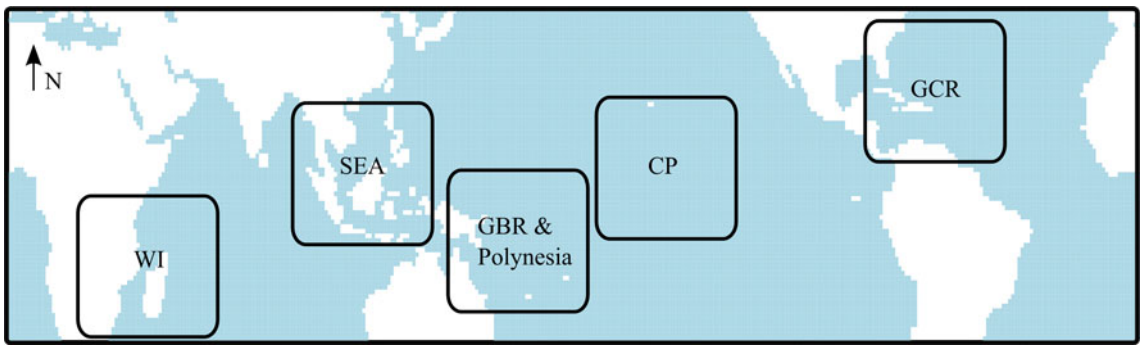

Fig. 2 Map of the coral regions analysed in this study, the Western Indian Ocean (WI), Southeast Asia (SEA), the Great Barrier Reef (GBR) and Polynesia, the Central Pacific (CP) and the Greater Caribbean Region (GCR)

then be analysed in isolation to better understand a process and how it varies over different scales. Unlike other methods such as Fourier decomposition, wavelets have the advantage of being local in space and therefore provide a simple approach for dealing with signals that are highly non-smooth (Walker 1997). Throughout this study we use Haar wavelets. Haar wavelets are the simplest possible wavelets and are essentially a series of square shaped functions.

To assess the believability of the climate model simulations, we use the wavelet intensity-scale method introduced by Casati et al. (2004) that has been successfully used in various diverse applications (De Sales and Xue 2010, Shutler et al. 2011, Saux-Picart et al. 2012). The method first involves making binary maps for the model simulation and for the observations based on whether or not the variable at each grid point exceeds a prescribed intensity threshold. The binary error map is then decomposed into the sum of l different spatial scales by using Haar wavelets. This decomposition allows us to write the mean of the squares of the binary error field (a performance score) as the sum of the mean squared errors on each of the different spatial scales. These scores are then compared to what one would expect for random unrelated fields (no skill) thereby providing a way of assessing believability in the model-simulated spatial scales for each intensity threshold.

2.2 Wavelet decomposition method

The wavelet based spatial comparison technique initially derived and developed by Casati et al. (2004, 2010) has been previously used to assess the skill of GCM precipitation outputs over South America (De Sales and Xue 2010) as well as hydrodynamic ecosystem models on the North West European shelf (Shutler et al. 2011). This wavelet-based technique has recently been extended to allow its generic application to a range of continuous and discontinuous geophysical data fields (Saux-Picart et al. 2012).

The methodology is based on binary difference maps (error). The initial conversion of 2D fields containing continuous observed and modelled values to a binary map (i.e. a map composed of 0 and 1 s) is a crucial step in the method, defining the patterns in the datasets that are later compared. As implemented by Saux-Picart et al. (2012) our thresholds are determined based on the empirical quartiles of the input fields. This study uses the three quartiles $X_{0.25}$, $X_{0.5}$ and $X_{0.75}$ to threshold the input fields but other percentiles could have been used. An example of the process of creating a binary difference map is shown in Fig. 1 for the Hadley Centre sea surface temperature observation-based product (HadISST) (Rayner et al. 2003) and the latest generation Hadley Centre ESM (HadGEM2-ES) (Collins et al. 2011). In this example SST fields for July in the respective 1985–2000 averaged climatologies are converted into a binary difference map. Prior to the process, the HadGEM2-ES fields are regridded onto the same $1° \times 1°$ spatial grid as HadISST using the CDO (Climate Data Operators) bilinear interpolation programme and both fields are given the same land mask (the combined land masks of the original model and observation fields). Note that the ocean resolution of HadGEM2-ES is one degree or higher everywhere, so we interpolate to the lower-resolution grid. The SST fields are converted into maps of quartile ranges, before the binary difference map is taken for the uppermost quartile range ($X > X_{0.75}$) in this example (Fig. 1). The subsequent decomposition of the binary difference map and assessment of model skill requires that both observation and model grids have the same dimensions. That is, they are squares with dimensions that are $2^l \times 2^l$ (Casati et al. 2004). Throughout our analysis we use regions that have dimensions $32° \times 32°$ corresponding to approximately 3,200 km $\times$ 3,200 km near the equator.

The methodology requires the initial model and observation maps to have an identical land mask. There are small differences in the land masks of CMIP5 models and HadISST due to varying resolutions prior to regridding and

Table 2 The $X_{0.25}$, $X_{0.5}$ and X_{075} quartiles of monthly HadISST climatologies (°K) for each coral region

Month	Greater Caribbean Region ($X_{0.25}$, $X_{0.5}$, X_{075})	Central Pacific ($X_{0.25}$, $X_{0.5}$, X_{075})	Great Barrier Reef and Polynesia ($X_{0.25}$, $X_{0.5}$, X_{075})	Southeast Asia ($X_{0.25}$, $X_{0.5}$, X_{075})	Western Indian Ocean ($X_{0.25}$, $X_{0.5}$, X_{075})
Jan	294.2, 297.9, 299.7	299.6, 300.6, 301.5	301.8, 302.2, 302.6	300.0, 301.3, 301.9	299.6, 301.0, 301.6
Feb	293.6, 297.5, 299.4	299.2, 300.5, 301.3	301.8, 302.2, 302.5	300.0, 301.4, 301.8	299.9, 301.3, 301.8
Mar	293.4, 297.4, 299.5	299.4, 300.7, 301.4	301.8, 302.2, 302.5	300.9, 301.8, 302.2	299.8, 301.6, 302.2
Apr	294.1, 298.0, 300.0	299.5, 300.9, 301.8	301.0, 302.1, 302.5	302.1, 302.4, 302.6	298.8, 301.2, 302.4
May	295.9, 299.2, 300.6	299.9, 301.3, 302.0	300.2, 302.2, 302.6	302.1, 302.7, 303.1	297.7, 300.2, 301.3
June	298.5, 300.5, 301.1	300.4, 301.4, 301.9	299.1, 301.8, 302.6	301.9, 302.5, 302.8	296.4, 299.0, 299.8
July	300.5, 301.1, 301.5	300.7, 301.4, 301.8	298.4, 301.4, 302.4	301.2, 302.2, 302.5	295.6, 297.9, 298.8
Aug	301.0, 301.5, 301.8	300.9, 301.4, 301.8	298.2, 301.2, 302.4	300.8, 302.0, 302.3	295.2, 297.4, 298.4
Sep	300.6, 301.7, 302.0	301.0, 301.6, 302.0	298.7, 301.4, 302.5	300.8, 302.0, 302.2	295.5, 297.7, 298.7
Oct	299.0, 301.2, 301.9	301.0, 301.6, 302.1	299.3, 301.9, 302.6	301.4, 301.9, 302.2	296.0, 298.3, 299.7
Nov	297.1, 300.0, 301.4	300.7, 301.4, 302.0	300.4, 302.3, 302.8	301.2, 301.8, 302.4	297.1, 299.5, 300.9
Dec	295.4, 298.8, 300.5	300.0, 300.9, 301.7	301.2, 302.3, 302.7	300.7, 301.5, 302.0	298.4, 300.7, 301.6

numerous small islands in coral regions such as Southeast Asia. We account for this by combining the land mask of a CMIP5 model with that of HadISST before assessing that given model. The statistical process of evaluating model skill from the wavelet decompositions of binary difference maps is described in full in Saux-Picart et al. (2012).

2.3 Application to coral reefs

The projection of thermally induced coral bleaching can be broken down into a number of discrete components. With respect to GCMs/ESMs, two of the most important components are: (1) producing an accurate climatology and (2) capturing long-term trends in annual mean temperature (van Hooidonk and Huber 2012). We assess the skill of the CMIP5 models at capturing these features at varying spatial scales using the wavelet based spatial comparison technique discussed above.

The analysis presented here is based-on output from the climate modelling groups associated with CMIP5—the models prepared for the IPCC 5th Assessment Report (AR5). We used 12 of the CMIP5 models (Table 1) with model selection based on those with the highest ocean field spatial resolution, permitting more levels of wavelet decomposition. The wavelet technique is applied to each individual CMIP5 model and quoted skill values for a given month, quartile range and spatial scale are averages of this CMIP5 multi-model ensemble. Skill was calculated as the mean square error relative to the mean square error of a random no skill projection. Model climatology skill was assessed by applying the wavelet based spatial comparison technique to the monthly climatologies calculated by averaging data across the years 1985–2000 for each model, and the HadISST observational record. Within each month of the climatology, skill was calculated for each quartile range. The climatological period was chosen due to its importance in calculating "Degree Heating Months" (DHM) bleaching thresholds. These are the monthly SST thresholds above which accumulated temperature anomalies typically result in coral bleaching. In studies that make future projections of coral bleaching these thresholds have been taken as the maximum monthly temperature of each grid cell in the 1985–2000 model climatologies (Donner 2009).

The ability of the models to capture historical warming was assessed by calculating warming anomalies. Anomalies were calculated for each grid cell in each model by subtracting the mean annual SST for 1960–1980 from the mean annual SST for 1985–2005. Anomalies were then compared to those of the HadISST observations using the wavelet based comparison technique. Anomalies were calculated for this period to overlap with the more recent HadISST record in which there is far more confidence at high spatial resolution. This is due to greater in situ sampling of SSTs and the incorporation of satellite measurements into the HadISST record in the early 1980s (Rayner et al. 2003). The use of 20 years averaging periods was chosen to minimise the influence of inter-annual variability, which freely running models (i.e. those starting from their internal equilibrium, rather than the observed ocean/atmosphere state) cannot be expected to reproduce in phase with that which is observed in reality, whilst attempting to preserve the long-term trends in warming anomalies. Model skill was assessed across five coral regions: the Greater Caribbean Region, the Central Pacific, the Great Barrier Reef and Polynesia, Southeast Asia and the western Indian Ocean (Fig. 2). Post 1960, in situ SST observational coverage in the regions of interest is essentially complete at least at the 5° × 5° scale (Kennedy et al. 2011a, b), with the exception of the Great

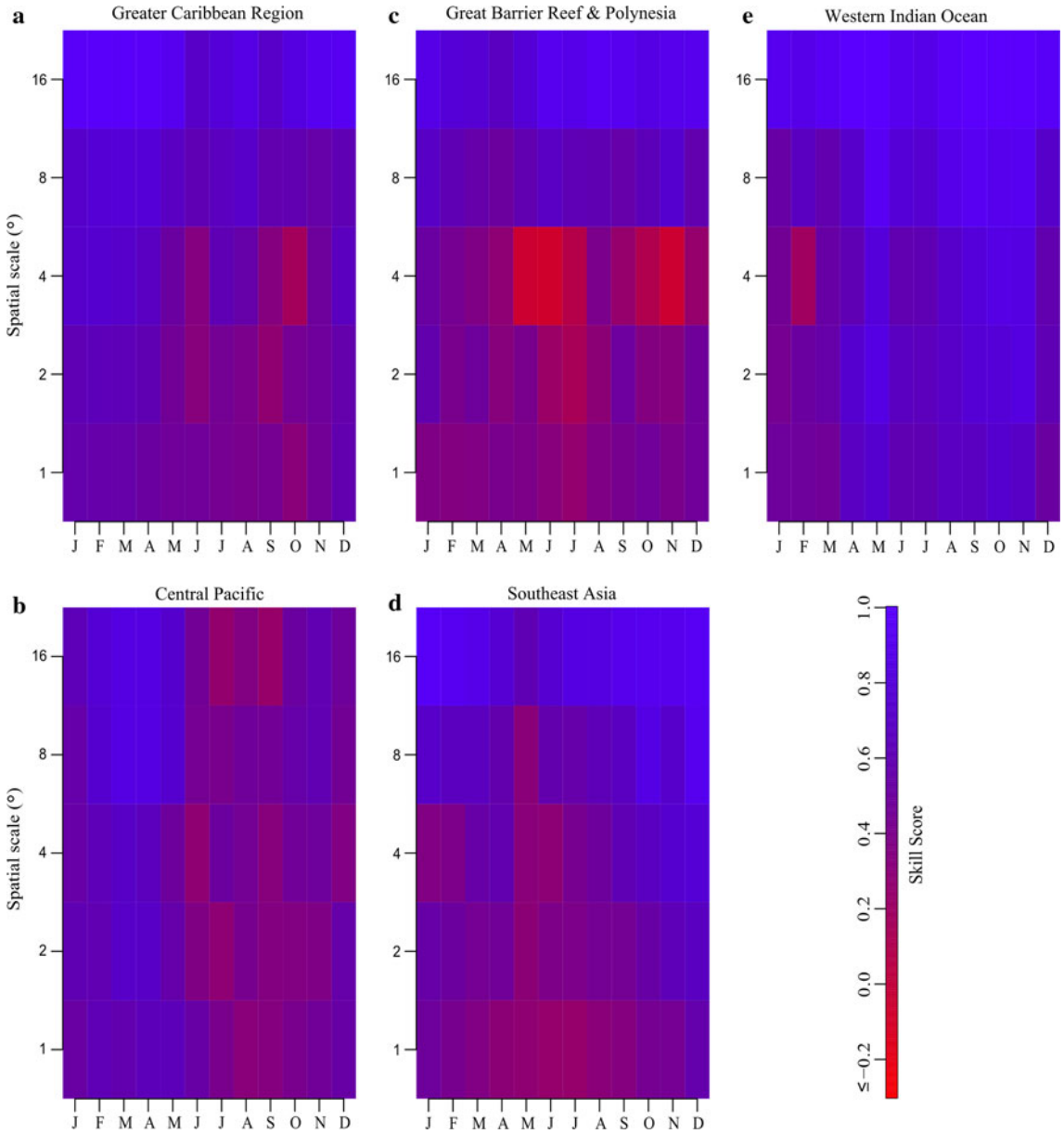

Fig. 3 Spatial scale versus time for the 4th quartile range ($X > X_{0.75}$) of each month. Multi-model skill shown for spatial scale against month for the **a** Greater Caribbean Region, **b** Central Pacific, **c** Great Barrier Reef and Polynesia, **d** Southeast Asia and **e** Western Indian Ocean. Skill is for the 1985–2000 climatology. Skill is calculated as the mean square error relative to the mean square error of a random no skill projection, in which 1 equates to a perfect simulation, 0 represents model skill that is equal to chance and values below 0 correspond to skill that is worse than chance alone

Barrier Reef and Polynesia region. In the Great Barrier Reef and Polynesia region, near complete in situ SST observation coverage is not achieved in all months until the late 1960s. Whilst our 20 years averaging period will diminish the effect of the small number of unobserved grid-cells on the 1960–1980 mean pattern, it is possible that the move to complete coverage in the latter period (1985–2005) will have introduced a minor spurious pattern change in this region, and therefore we recommend against strict interpretation of the results from the Great Barrier Reef and Polynesia region.

3 Results

3.1 Coastal tropical climatologies

The approximate quartiles ($X_{0.25}$, $X_{0.5}$ and $X_{0.75}$) for each month of the climatologies in each coral region are given in Table 2 with further details given in supplementary table S1. These quartiles are based on HadISST and will differ somewhat for each CMIP5 model. The CMIP5 models are shown to have skill at capturing the uppermost quartile range ($X > X_{0.75}$), that is the distribution of the warmest waters, of

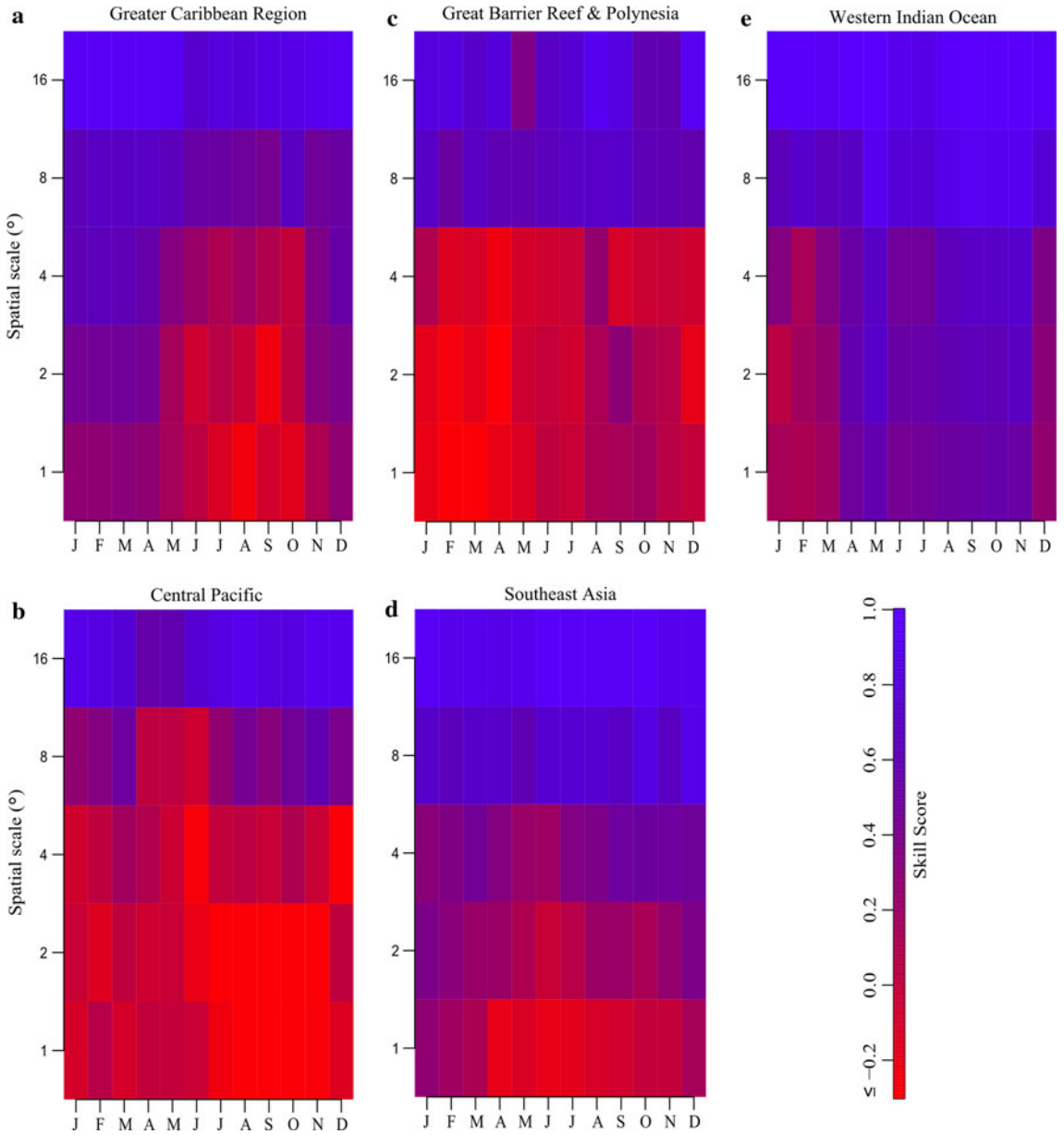

Fig. 4 Spatial scale versus time for the 3rd quartile range ($X_{0.50} < X \leq X_{0.75}$) of each month. Multi-model skill shown for spatial scale against month for the **a** Greater Caribbean Region, **b** Central Pacific, **c** Great Barrier Reef and Polynesia, **d** Southeast Asia and **e** Western Indian Ocean. Skill is for the 1985–2000 climatology. Skill is calculated as the mean square error relative to the mean square error of a random no skill projection, in which 1 equates to a perfect simulation, 0 represents model skill that is equal to chance and values below 0 correspond to skill that is worse than chance alone

each month, of the 1985–2000 climatology, even at low spatial scales (Fig. 3). At 1° spatial scales multi-model ensemble mean skill is better than chance (>0.0) across all coral regions and each month; although within each region skill is shown to vary through the months of the climatology (Fig. 3). With increasing spatial scale there is a general improvement in skill across all regions, with the highest skill observed at 8°–16° spatial scales. However, an exception to this is seen in the GBR (Great Barrier Reef) and Polynesia region where there is a consistent patch of low skill observed at a spatial scale of 4° between May and November. This pattern is not apparent in the other regions. Across regions, multi-model ensemble mean skill for the third ($X_{0.50} < X \leq X_{0.75}$) quartile range of each month is also highest for spatial scales of 8°–16° and generally lower for the Central Pacific and GBR and Polynesia than for other regions (Fig. 4). This is also seen in the second quartile range ($X_{0.25} < X \leq X_{0.50}$) (Fig. 5) and to a lesser extent for the first quartile range ($X \leq X_{0.25}$) where model skill is far higher across all regions and all spatial scales (Fig. 6).

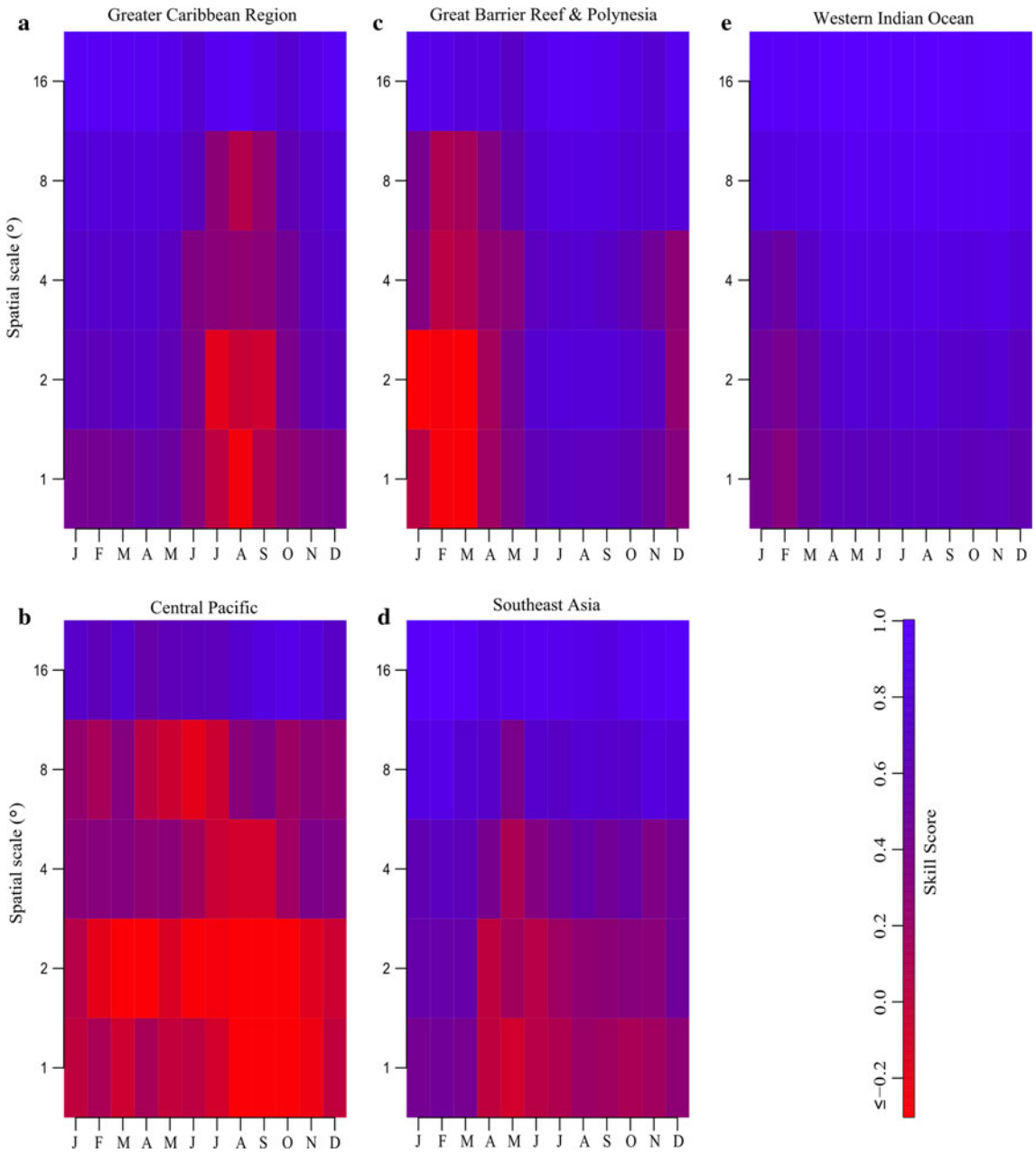

Fig. 5 Spatial scale versus time for the 2nd quartile range ($X_{0.25} < X \leq X_{0.50}$) of each month. Multi-model skill shown for spatial scale against month for the **a** Greater Caribbean Region, **b** Central Pacific, **c** Great Barrier Reef and Polynesia, **d** Southeast Asia and **e** Western Indian Ocean. Skill is for the 1985–2000 climatology. Skill is calculated as the mean square error relative to the mean square error of a random no skill projection, in which 1 equates to a perfect simulation, 0 represents model skill that is equal to chance and values below 0 correspond to skill that is worse than chance alone

When comparing the skill of the CMIP5 models across other quartile ranges (Figs. 4, 5, 6), there are both consistencies and some emergent patterns. The general improvement in skill at larger spatial scales is evident for all regions across all quartile ranges. In addition, within certain regions, periods of the year with lower skill are consistent across quartile ranges. In the Greater Caribbean Region for example, skill is typically lower between June and October, coinciding with the Hurricane season and the time of the year that SSTs are highest—the period when most bleaching occurs. A similar pattern is seen for the Central Pacific region across quartile ranges. In the western Indian Ocean and the GBR and Polynesia regions skill is typically lower between January and March, also coinciding with the period of highest annual SSTs. Across quartile ranges the region with the consistently lowest relative skill is the Central Pacific. Conversely the region with typically the highest skill across quartile ranges is the Western

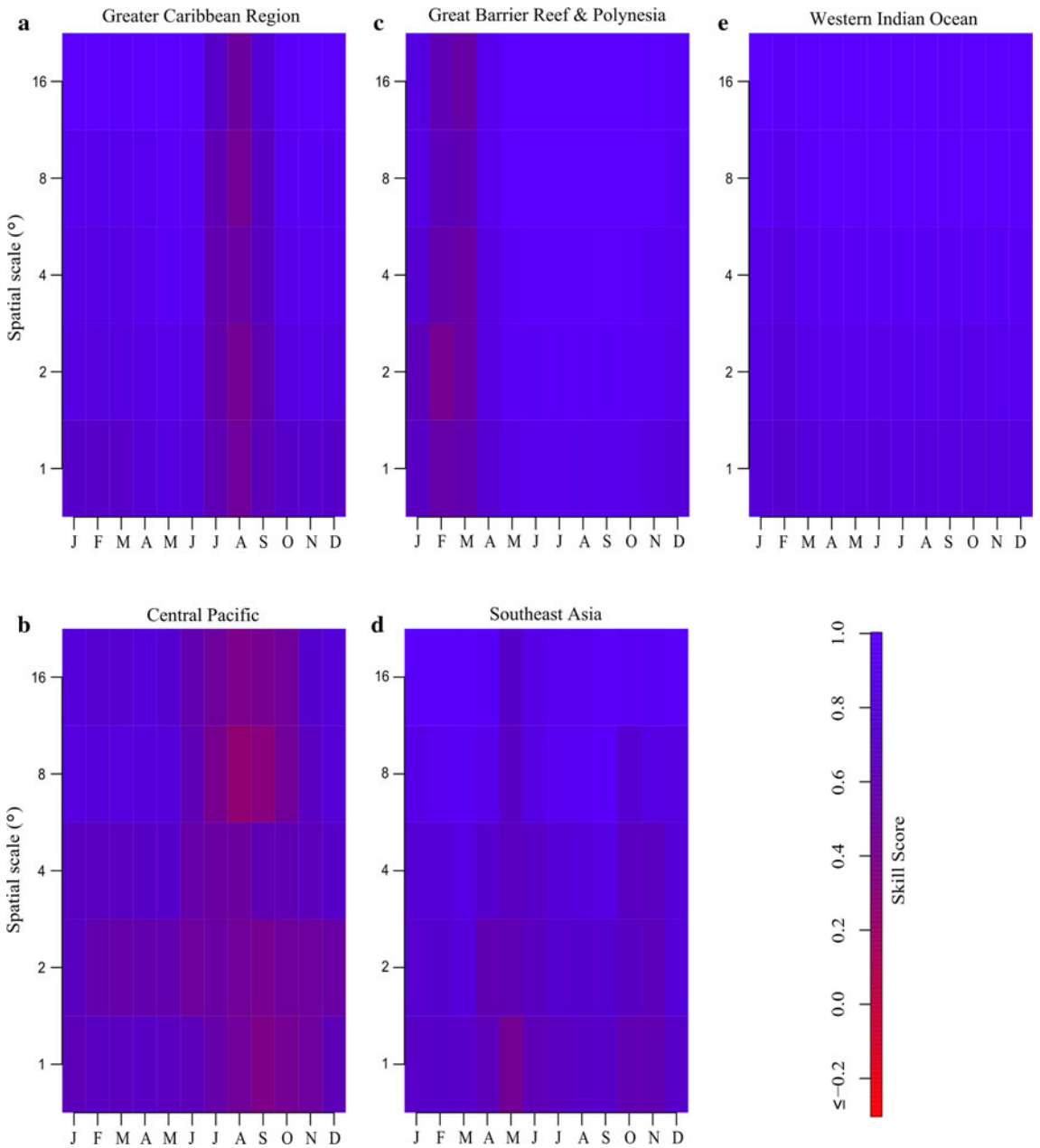

Fig. 6 Spatial scale versus time for the 1st quartile range ($X \leq X_{0.25}$) of each month. Multi-model skill shown for spatial scale against month for the **a** Greater Caribbean Region, **b** Central Pacific, **c** Great Barrier Reef and Polynesia, **d** Southeast Asia and **e** Western Indian Ocean. Skill is for the 1985–2000 climatology. Skill is calculated as the mean square error relative to the mean square error of a random no skill projection, in which 1 equates to a perfect simulation, 0 represents model skill that is equal to chance and values below 0 correspond to skill that is worse than chance alone

Indian Ocean. Across all regions skill is typically lower for the second ($X_{0.25} < X \leq X_{0.50}$) and third ($X_{0.50} < X \leq X_{0.75}$) quartile ranges than for the first ($X \leq X_{0.25}$) and fourth ($X > X_{0.75}$) quartile ranges (Figs. 3, 4, 5, 6).

3.2 Historical spatial SST warming patterns

In Fig. 7 the regional mean warming anomalies between 1960–1980 and 1985–2005 are compared against HadISST for each of the CMIP5 models. The range of model values encompasses HadISST results in all regions except the Greater Caribbean Region (GCR). In the GCR models overestimate the relatively low regional warming observed most probably due to the influence of the Atlantic Multi-decadal Oscillation.

Maps of the HadISST regional warming anomalies observed between 1960–1980 and 1985–2005 are shown in Fig. 8. The patterns of warming show some discernible

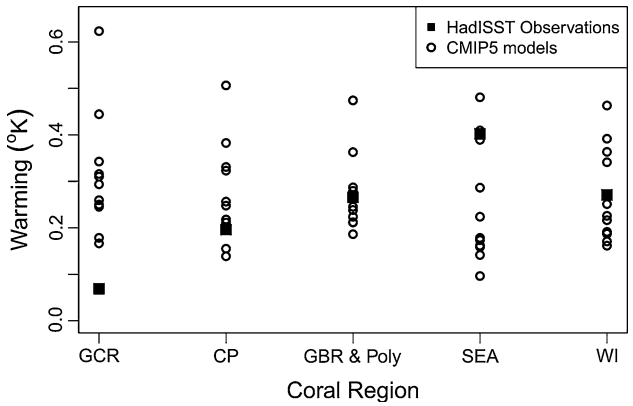

Fig. 7 Comparison of the mean historical warming between 1960–1980 and 1985–2005 hindcast by each model averaged across the Greater Caribbean Region (GCR), Central Pacific (CP), Great Barrier Reef (GBR) and Polynesia (Poly), Southeast Asia (SEA) and Western Indian Ocean (WI). HadISST observations are plotted as closed squares, individual ESMs/GCMs as open circles

features. For example in the western Indian Ocean there appears to be greater warming in the waters between the continental land mass and Madagascar. A similar pattern is apparent for Southeast Asia where the greatest warming is observed between the continental land mass and Borneo. Such features may be due to bathymetry, with the relatively shallow ocean basins in these regions warming at a faster rate than surrounding deeper waters because of less deep mixing. Other features in these regional SST warming plots are far harder to explain. It is apparent from our assessment of multi-model skill (Fig. 9) that these spatial patterns of warming anomalies are not consistently found in CMIP5 models.

Multi-model skill for the warming anomalies is shown across quartile ranges in Fig. 9. At 1° spatial scales the multi-model ensemble skill is typically no better than chance and often considerably worse than chance, contrasting significantly with the skill observed for climatologies (e.g. Fig. 3). Moreover, even at larger spatial scales of 8°−16° skill is still not consistently better than chance across all quartile ranges for all regions. Skill is, however, generally better at larger spatial scales, and better for Southeast Asia and the western Indian Ocean than for other regions. Additionally we observe that the second and third SST quartile ranges typically have lower skill than the other quartile ranges at smaller spatial scales however show higher relative skill than the other quartile ranges at larger spatial scales. The standard error of all multi-model skill score means is given in the supplementary material.

4 Discussion

Multi-model mean CMIP5 model skill is shown to vary considerably with spatial scale in terms of capturing both climatological periods and historical changes in mean annual SSTs between 1960–1980 and 1985–2005. With regard to climatologies, the finding of typically lower skill for the second and third quartile ranges than for the outer quartile ranges (Figs. 3, 4, 5, 6) is most likely due to these outer quartiles covering a far larger range in absolute SSTs. This is consistently shown across all regions and all months of the year in supplementary table S1. The CMIP5 models are therefore more likely to contain the physical processes required to simulate the spatial distribution of these upper and lower quartile ranges. When assessing spatial warming patterns, the finding of higher skill at larger spatial scales for Southeast Asia and the western Indian Ocean than for the other regions was also interesting as these are the two regions in which we suggest that bathymetry may have had a greater role in determining warming anomalies (Fig. 8).

Skill values for climatologies are typically far higher than those for historical changes in mean annual SSTs, across all spatial scales. In many ways this might be expected given that spatial patterns of climatologies are dominated by meridional SST gradients that show low inter-annual variability and are therefore relatively well modelled by GCMs/ESMs, especially at spatial scales of ≥4°. In contrast, patterns of historical warming in coastal areas are influenced by a number of complex coastal processes as well as strong inter-annual variability and at present contain only a weak climate-change signal. A consequence of this is typically very low model skill that is often worse than chance (Fig. 9). Moreover model skill only shows minimal improvements with increasing spatial scale and not across all quartile ranges (Fig. 9). It should be noted that as climate change progresses and anthropogenically-induced warming increasingly dominates spatial patterns of SST warming anomalies, we might expect model validation to show skill at increasingly small spatial scales. This is because the signal to noise ratio will increase as the influence of the anthropogenic forcing (signal) increases relative to the presumably relatively stable influence of natural and stochastic variability (noise). Even a perfect model would not consistently capture the natural variability in the system without being initialised from observations. However, the rotation of the Earth and the heterogeneity of the planet's surface alone mean that the anthropogenically-forced component of the change will result in a non-uniform pattern of warming, and it is this, that the ESM experiments can capture. Given a stronger anthropogenic forcing, we would therefore anticipate that the ESM skill, as judged in this study, would improve. At this stage however, it is not possible to robustly quantify how much skill models might have.

In terms of modelling coral bleaching, the analysis presented here has a number of important implications. The maximum monthly mean (MMM) as used in the

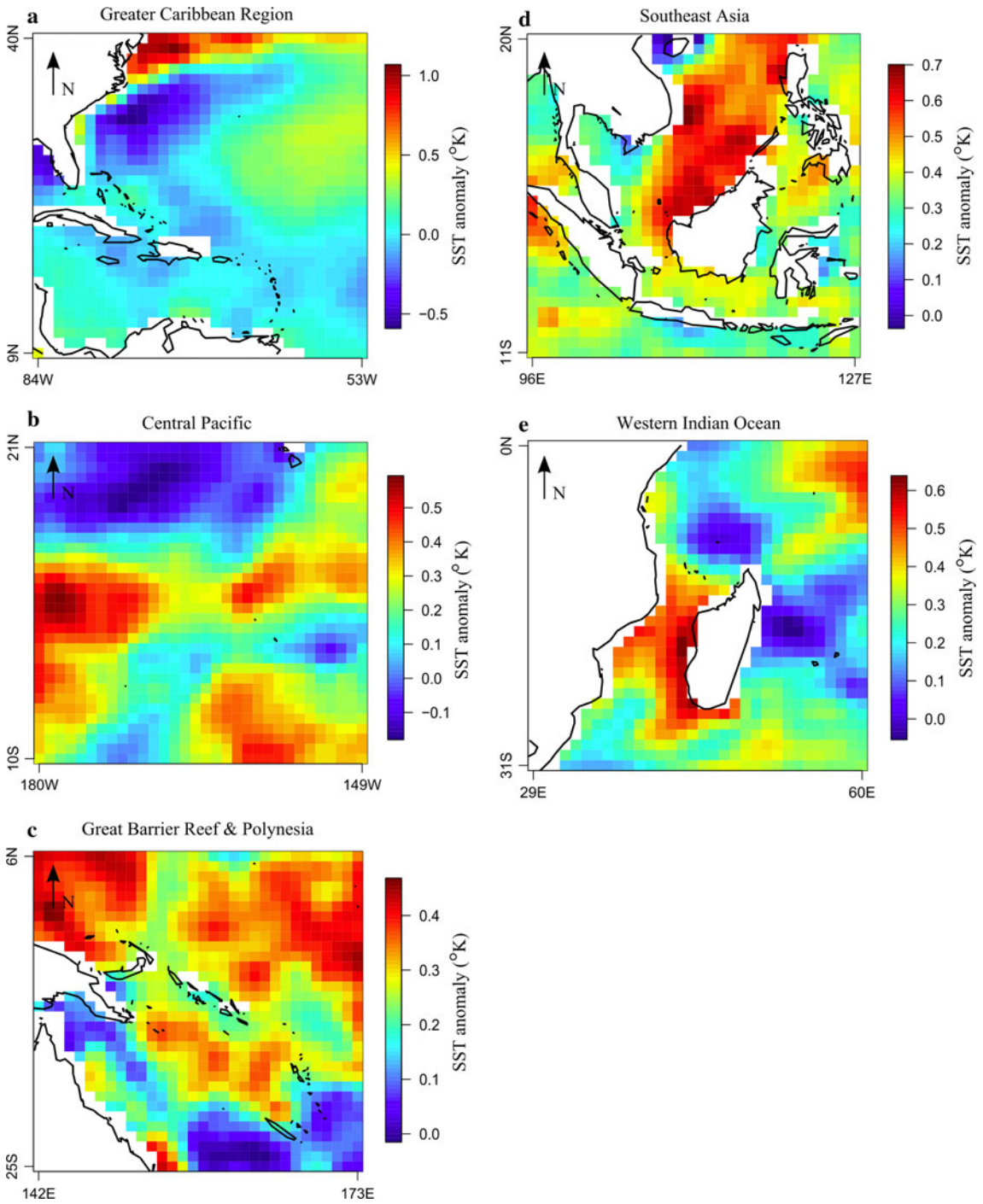

Fig. 8 HadISST mean warming anomalies (°K) for 1985–2005 relative to mean 1960–1980 SSTs for the **a** Greater Caribbean Region, **b** Central Pacific, **c** Great Barrier Reef and Polynesia, **d** Southeast Asia and **e** Western Indian Ocean

calculation of "Degree Heating Months" (DHM) is typically going to be the SST of the hottest month in a historical climatology, e.g. for a grid cell on the Great Barrier Reef it may be the mean 1985–2000 February SST. However, as our analysis shows, model skill is typically lower in relation to the spatial patterns of climatological SSTs during the warmest months of the year and lower at finer spatial scales. For example, across quartile ranges in the Caribbean, skill is shown to be less between June and October. Consequently, models will potentially have low skill at producing patterns of MMM values at small spatial scales.

Furthermore, the CMIP5 models show poor skill in relation to spatial patterns of historical warming (Fig. 9).

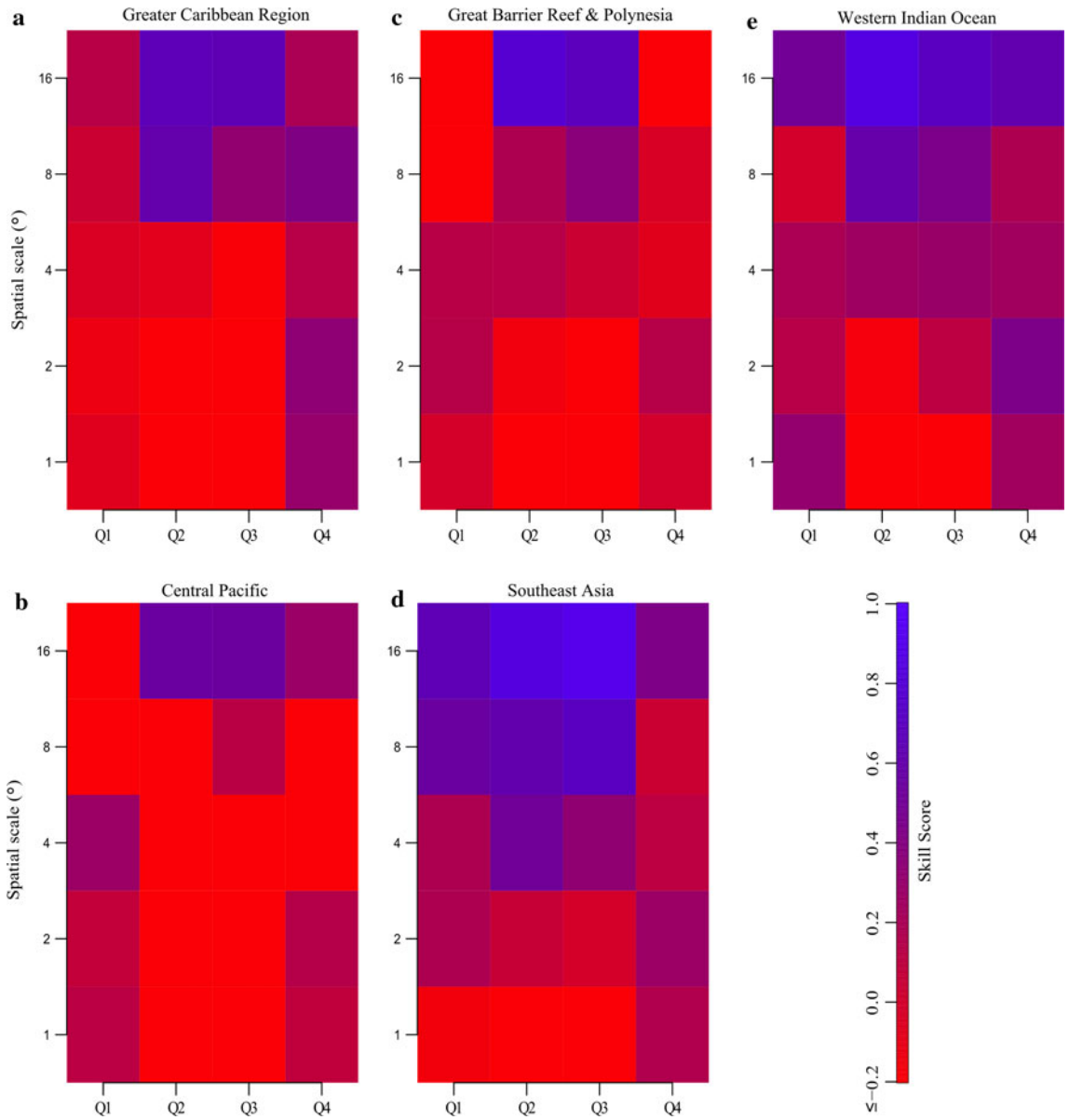

Fig. 9 Skill for the SST warming anomalies between 1960–1980 and 1985–2005 calculated as annual average values. Multi-model skill shown for spatial scale against quartile range for the **a** Greater Caribbean Region, **b** Central Pacific, **c** Great Barrier Reef and Polynesia, **d** Southeast Asia and **e** Western Indian Ocean

Although such values show slight improvements at larger spatial scales they still remain close to 0 for certain quartile ranges in certain regions across all scales investigated here. The implications of this are that CMIP5 models typically do not contain the necessary processes to consistently model patterns of historical coastal SST warming at spatial scales of $\leq 16°$. They are therefore unlikely to skilfully project patterns of future warming at small spatial scales. We would anticipate skill over the 21st Century to improve as anthropogenically-induced warming increasingly drives patterns of SST warming anomalies and the relative importance of natural variability diminishes. However, global model skill is unlikely to be as high as it is for climatological patterns and therefore one should avoid interpreting SST outputs at $<8°$. As long as poor skill at small spatial scales does not introduce any systemic bias, the conclusions of global scale bleaching projection studies (e.g. Frieler et al. 2013) should remain robust. However, if future coral bleaching is projected at sub-regional spatial resolution (i.e. $<16°$), then the resulting heterogeneity within a region should be interpreted with caution, potentially casting doubt on some of the projections of coral refugia presented in van Hooidonk et al. (2013). It would be more robust to use all grid cells within a coral region to

produce mean regional projections and avoid making projections at sub-regional scales (e.g. Kwiatkowski et al. 2013).

There is a desire to move towards providing long-term regional high resolution bleaching projections. Such projections are potentially very valuable and could aid regional decision making on marine protected areas (MPAs), fisheries policy and coastal development (Mumby et al. 2011). However, given that we show that bleaching projections based on current generation ESM SST outputs are likely to have very poor skill at smaller spatial scales, an alternative approach is required. Where observational spatial warming patterns can be shown to be non-time dependent, there may be scope to separate regional SST warming projections into a fixed spatial pattern derived from historical observations and a spatially averaged time dependent function derived from models (Huntingford and Cox 2000). Another potential solution requiring further research is the use of carefully validated regional coastal-shelf models to down-scale global model results.

5 Conclusions

In this paper the wavelet intensity-scale method was used to assess the skill of CMIP5 models at capturing the spatial patterns of SST features in five coral regions. The models were assessed for their ability to capture the patterns of monthly SSTs in a historical climatology and the patterns of SST warming anomalies between 1960–1980 and 1985–2005. Our key findings are:

- The spatial patterns of monthly climatological SSTs are generally well produced by the CMIP5 models in the coral regions we analysed.
- Patterns of monthly climatological SSTs are best produced by the CMIP5 models at spatial scales >4°.
- Across spatial scales the skill of CMIP5 models to capture spatial patterns of monthly climatological SSTs is generally lower during the warmest months of the year in a given coral region.
- CMIP5 models have typically very poor skill and often perform worse than chance at capturing spatial patterns of SST warming anomalies between 1960–1980 and 1985–2005 in the coral regions we analysed.
- The skill of the CMIP5 models at capturing sub-regional patterns of SST warming anomalies does not consistently improve at larger spatial scales of up to 16°.

In future work, techniques that could potentially increase the skill of CMIP5 models to project SSTs at small spatial scales will be explored. Of particular interest is the effectiveness of using pattern scaling techniques and coastal-shelf models to down-scale global model outputs.

Acknowledgments We thank Barbara Casati for providing useful guidance on how to code wavelet-based spatial comparison techniques in R. We thank David Long for assistance downloading CMIP5 data. The study was funded by a NERC grant to Peter J. Mumby and Peter M. Cox, the University of Exeter, the EU FORCE project and was supported by the Joint DECC/Defra Met Office Hadley Centre Climate Programme.

References

Allen JI, Aiken J, Anderson T, Buitenhuis E, Cornell S, Geider R, Haines K, Hirata T, Holt J, Le Quéré C, Hardman-Mountford N, Ross ON, Sinha B, While J (2010) Marine ecosystem models for earth systems applications: the MarQUEST experience. J Mar Syst 81:19–33

Brown BE (1997) Adaptations of reef corals to physical environmental stress. Adv Mar Biol 31:221–299

Casati B (2010) New developments of the intensity-scale technique within the Spatial Verification Methods Intercomparison Project. Weather Forecast 25:113–143

Casati B, Ross G, Stephenson D (2004) A new intensity-scale approach for the verification of spatial precipitation forecasts. Meteorol Appl 11:141–154

Collins WJ, Bellouin N, Doutriaux-Boucher M, Gedney N, Halloran P, Hinton T, Hughes J, Jones CD, Joshi M, Liddicoat S, Martin G, O'Conner F, Rae J, Senior C, Sitch S, Totterdell I, Wiltshire A, Woodward S (2011) Development and evaluation of an earth-system model–hadgem2. Geosci Model Dev 4:1051–1075

De Sales F, Xue Y (2010) Assessing the dynamic-downscaling ability over South America using the intensity-scale verification technique. Int J Climatol 31:1205–1221

Donner S (2009) Coping with commitment: projected thermal stress on coral reefs under different future scenarios. PLoS One 4:e5712

Donner S, Skirving W, Little C, Oppenheimer M, Hoegh-Guldberg O (2005) Global assessment of coral bleaching and required rates of adaptation under climate change. Glob Chang Biol 11:2251–2265

Eakin CM, Morgan JA, Heron SF et al (2010) Caribbean corals in crisis: record thermal stress, bleaching, and mortality in 2005. PLoS One 5:e13969

Edwards AJ, Clark S, Zahir H, Rajasuriya A, Naseer A, Rubens J (2001) Coral bleaching and mortality on artificial and natural reefs in Maldives in 1998: sea surface temperature anomalies and initial recovery. Mar Pollut Bull 42:7–15

Enríquez S, Méndez ER, Iglesias-Prieto R (2005) Multiple scattering on coral skeletons enhances light absorption by symbiotic algae. Limnol Oceanogr 50:1025–1032

Frieler K, Meinshausen M, Golly A, Mengel M, Lebek K, Donner S, Hoegh-Guldberg O (2013) Limiting global warming to 2 °C is unlikely to save most coral reefs. Nature Clim Chang 3:165–170

Hoegh-Guldberg O (1999) Climate change, coral bleaching and the future of the world's coral reefs. Mar Freshw Res 50:839–866

Hoeke R, Jokiel P, Buddemeier R, Brainard R (2011) Projected changes to growth and mortality of Hawaiian corals over the next 100 years. PLoS One 6:e18038

Holt J, Harle J, Proctor R, Michel S, Ashworth M, Batstone C, Allen I, Holmes R, Smyth T, Haines K, Bretherton D, Smith G (2009) Modelling the global coastal ocean. Philos T Roy Soc A 367:939–951

Huntingford C, Cox PM (2000) An analogue model to derive additional climate change scenarios from existing GCM simulations. Clim Dyn 16:575–586

Jawerth B, Sweldens W (1994) An overview of wavelet based multiresolution analyses. SIAM Rev 36:377–412

Kennedy JJ, Rayner NA, Smith RO, Saunby M, Parker DE (2011a) Reassessing biases and other uncertainties in sea-surface temperature observations since 1850 part 1: measurement and sampling errors. J Geophys Res 116:D14103. doi:10.1029/2010JD015218

Kennedy JJ, Rayner NA, Smith RO, Saunby M, Parker DE (2011b) Reassessing biases and other uncertainties in sea-surface temperature observations since 1850 part 2: biases and homogenisation. J Geophys Res 116:D14104. doi:10.1029/2010JD015220

Kramer PA, Kramer PR, Ginsburg RN (2003) Assessment of the Andros Island reef system, Bahamas (Part 1: stony corals and algae). Atoll Res Bull 496:76–99

Kwiatkowski L, Cox PM, Economou T, Halloran PR, Mumby PJ, Booth BBB, Carilli J, Guzman HM (2013) Caribbean coral growth influenced by anthropogenic aerosols. Nat Geosci 6:362–366

Lander J, Hoskins BJ (1997) Believable scales and parameterizations in a spectral transform model. Mon Weather Rev 125:292–303

Mumby PJ, Chisholm JRM, Clark CD, Hedley JD, Jaubert J (2001) Spectrographic imaging: a bird's-eye view of the health of coral reefs. Nature 413:36

Mumby PJ, Elliott IA, Eakin MC, Skirving W, Paris CB, Edwards HJ, Enriquez S, Iglesias Prieto R, Cherubin LM, Stevens JR (2011) Reserve design for uncertain responses of coral reefs to climate change. Ecol Lett 14:132–140

Palmer TN, Doblas-Reyes FJ, Hagedorn R, Weisheimer A (2005) Probabilistic prediction of climate using multi-model ensembles: from basics to applications. Philos T Roy Soc B 360:1991–1998

Rayner NA, Parker DE, Horton EB, Folland CK, Alexander LV, Rowell DP, Kent EC, Kaplan A (2003) Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. J Geophys Res 108:4407

Saux-Picart S, Butenschön M, Shutler JD (2012) Wavelet-based spatial comparison technique for analysing and evaluating two-dimensional geophysical model fields. Geosci Model Dev 5:223–230

Sheppard CR (2003) Predicted recurrences of mass coral mortality in the Indian Ocean. Nature 425:294–297

Shutler JD, Smyth TJ, Saux-Picart S, Wakelin SL, Hyder P, Orekhov P, Grant MG, Tilstone GH, Allen JI (2011) Evaluating the ability of a hydrodynamic ecosystem model to capture inter-and intra-annual spatial characteristics of chlorophyll-a in the north east Atlantic. J Mar Syst 88:169–182

van Hooidonk R, Huber M (2009) Quantifying the quality of coral bleaching predictions. Coral Reefs 28:579–587

van Hooidonk R, Huber M (2012) Effects of modeled tropical sea surface temperature variability on coral reef bleaching predictions. Coral Reefs 31:121–131

van Hooidonk R, Maynard JA, Planes S (2013) Temporary refugia for coral reefs in a warming world. Nature Clim Chang 3:508–511

Walker JS (1997) Fourier analysis and wavelet analysis. Notices AMS 44:658–670