

# Forecasting South China Sea monsoon onset using insight from theory

Geen, Ruth

DOI:

[10.1029/2020GL091444](https://doi.org/10.1029/2020GL091444)

License:

Creative Commons: Attribution (CC BY)

*Document Version*

Publisher's PDF, also known as Version of record

*Citation for published version (Harvard):*

Geen, R 2021, 'Forecasting South China Sea monsoon onset using insight from theory', *Geophysical Research Letters*, vol. 48, no. 6, e2020GL091444. <https://doi.org/10.1029/2020GL091444>

[Link to publication on Research at Birmingham portal](#)

## General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

## Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact [UBIRA@lists.bham.ac.uk](mailto:UBIRA@lists.bham.ac.uk) providing details and we will remove access to the work immediately and investigate.

# Geophysical Research Letters



## RESEARCH LETTER

10.1029/2020GL091444

### Key Points:

- Recent theoretical studies have revealed feedbacks at work during monsoon onset
- If the spring atmosphere is in a state where these feedbacks are easily triggered, onset is earlier
- This insight is used to produce a simple forecast of South China Sea Monsoon onset timing

### Supporting Information:

- Supporting Information S1

### Correspondence to:

R. Geen,  
[r.geen@exeter.ac.uk](mailto:r.geen@exeter.ac.uk)

### Citation:

Geen, R. (2021). Forecasting South China Sea Monsoon onset using insight from theory. *Geophysical Research Letters*, 48, e2020GL091444. <https://doi.org/10.1029/2020GL091444>

Received 29 OCT 2020

Accepted 12 FEB 2021

## Forecasting South China Sea Monsoon Onset Using Insight From Theory

Ruth Geen<sup>1</sup> 

<sup>1</sup>College of Engineering, Mathematics and Physical Sciences, University of Exeter, Exeter, UK

**Abstract** Monsoon onset over the South China Sea occurs in April–May, marking the start of the wet season over East Asia. Skillful prediction of onset timing remains an open challenge. Recently, theoretical studies using idealized models have revealed feedbacks at work during the seasonal transitions of the Hadley cells and have shown that these are relevant to monsoon onset over Asia. Here, I hypothesize that monsoon onset occurs earlier in years when the atmosphere over the South China Sea is already in a state where these feedbacks are more easily triggered. I find that local anomalies in lower-level moist static energy in the preceding January–March are well correlated with South China Sea Monsoon onset timing. This relationship remains relatively consistent on decadal timescales, while correlations with other teleconnections vary, and is used to develop a simple forecast model for onset timing that shows skill competitive with that of more complex models.

**Plain Language Summary** Arable land in China is estimated to feed 20% of the world's population on 7% of the world's farmlands, with much of this region watered by rainfall from the East Asian Summer Monsoon. Forecasting the arrival of the monsoon, which occurs first over the South China Sea in April–May, is therefore helpful for agricultural planning. However, producing a reliable forecast remains an open challenge. Recently, simple climate models, for example, models including seasons but with no land, have been used along with observations to investigate the most basic processes controlling climatological monsoon onset over Asia. In this paper, I suggest that year-to-year variations in monsoon onset timing are controlled by the same processes and, based on these, I suggest climate variables that may help in predicting when the monsoon will begin. By using this approach, I find that the “Moist Static Energy” (a quantity that combines temperature and humidity) over the South China Sea in January–March is strongly correlated with monsoon onset timing. This insight can be used to predict future onset timing competitively compared with more complex forecast models.

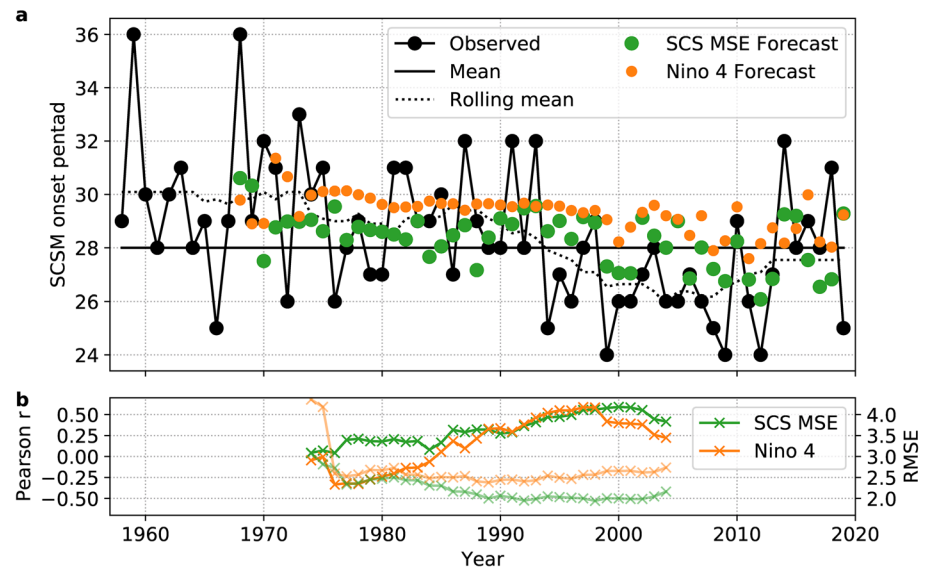
## 1. Introduction

The East Asian Summer Monsoon (EASM) affects up to one third of the world's population (Chang, 2004), and the arable land in China is estimated to feed 20% of the Earth's population on 7% of the world's farmlands (Cui & Shoemaker, 2018). The reliance of agriculture on EASM rainfall means that there is a strong drive to predict when rain will arrive.

In East Asia, the wet season begins with the abrupt reversal of the lower-level zonal wind over the South China Sea (SCS), accompanied by an off-equatorial shift in the tropical rainband over this area, known as the South China Sea Monsoon (SCSM; B. Wang & LinHo, 2002). This is followed by the development of an east-west oriented front of precipitation, known as Meiyu in China, which brings intense rainfall to the Yangtze River valley from mid-June to -July (Ding & Wang, 2005; B. Wang & LinHo, 2002). SCSM onset timing is considered a precursor to the broader-scale onset of the monsoon over East Asia (Lau & Yang, 1997; B. Wang et al., 2004) and additionally appears correlated to monsoon rainfall intensity over China, with an early onset associated with a drier summer (He & Zhu, 2015; Huang et al., 2006). The monsoon onset over the South China Sea varies interannually from late-April to mid-June (e.g., B. Wang et al., 2004 and see Figure 1), giving over a month's uncertainty. In this paper, I propose a novel approach to improving predictability of the SCSM onset by exploring whether recent theoretical advances in monsoon dynamics can provide insight into controls on SCSM onset timing and inform seasonal forecasting efforts.

© 2021. The Authors.

This is an open access article under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits use, distribution and reproduction in any medium, provided the original work is properly cited.



**Figure 1.** (a) Onset timing of the SCSM in pentads (5-day means; black circles). The dotted line shows the 11-year rolling mean, which I use to distinguish longer term trends from interannual variability. The overall mean onset timing is pentad 28 (solid line). Green (orange) circles show predicted onset pentads for each year based on MSE over  $5^{\circ}$ – $15^{\circ}$ N,  $110$ – $125^{\circ}$ E (SST over  $-5^{\circ}$ – $5^{\circ}$ N,  $160$ – $210^{\circ}$ E). Details of how these forecasts are generated is given in Section 4. (b) Rolling mean skill statistics for the forecasts in (a) to show how skill varies in time. Darker lines show Pearson  $r$ , lighter lines show root mean square error, both based on centered 31-year rolling windows. An equivalent plot for CERA-20C data is given in Figure S3.

SCSM onset timing shows both interannual variability and slower, interdecadal trends (Kajikawa et al., 2012; Kajikawa & Wang, 2012; Figure 1). The El Niño Southern Oscillation (ENSO) has been identified as one clear source of interannual variability, with La Niña (El Niño) events linked to early (late) monsoon onset via their influence on the Western North Pacific subtropical high (Zhou & Chan, 2007). However, the strength of the relationship between the SCSM and ENSO varies dramatically on decadal timescales, likely influenced by the the Pacific Decadal Oscillation (PDO; Chan & Zhou, 2005) and Atlantic Multidecadal Oscillation (AMO; Fan et al., 2018). In addition to ENSO, SCSM onset variability has been related to a range of factors in the preceding winter and spring: thermal and mechanical forcing over the Tibetan Plateau (G. Wu & Zhang, 1998); temperature contrasts between the South China Sea and Western North Pacific and the land surface temperature to the north (P. Liu et al., 2009); and the cross-equatorial flow over the South China Sea (Hu et al., 2018; Lin et al., 2017).

Physical-Empirical models have been developed to predict SCSM onset timing and intensity based on correlations with sea surface temperature (SST), sea level pressure and temperature tendency anomalies in the preceding months. These models can show high forecast correlation skill over the time-periods analyzed (e.g.,  $r = 0.72$ , Zhu & Li, 2017). Recently, dynamical seasonal forecasting ensembles have also been found to give skillful prediction of the SCSM onset in hindcasts (Fan et al., 2016; Martin et al., 2019). However, the skill of both types of model stems from teleconnections such as ENSO, whose correlation with the SCSM vary decadal.

Here, I try a different approach. First, I identify the processes found to be most important to monsoon onset in idealized modeling studies (e.g., Bordoni & Schneider, 2008; Geen et al., 2018). I then explore whether these insights can help to identify direct, local precursors to the monsoon. My aim is to find common pathways via which multiple teleconnections affect onset timing. In Section 2, I motivate the precursors that I hypothesize to be relevant and detail the datasets used in the paper. In Section 3, I then examine the correlation between SCSM onset and these precursors over different time periods. In Section 4, I develop a simple forecast of SCSM onset and test its predictive skill. Section 5 concludes.

## 2. Methods

### 2.1. Hypothesized Precursors

To select potential mechanistic pathways, I apply results from idealized modeling studies that compare the seasonal behavior of the Hadley cell in aquaplanets with Earth's monsoons (Geen et al., 2020; Hill, 2019). One important expectation from this theoretical work is that, if the tropical atmosphere is near convective quasi-equilibrium (CQE) and the influence of extra-tropical eddies on the Hadley cell is weak, then the 0-streamfunction line separating the Hadley cells is collocated with the maximum in subcloud moist static energy (MSE). If this maximum is off the Equator, as is the case during the monsoon season, then the maximum ascent and associated rainfall will lie just equatorward of this (Nie et al., 2010; Privé & Plumb, 2007). MSE is defined

$$h \equiv c_p T + Lq + gz. \quad (1)$$

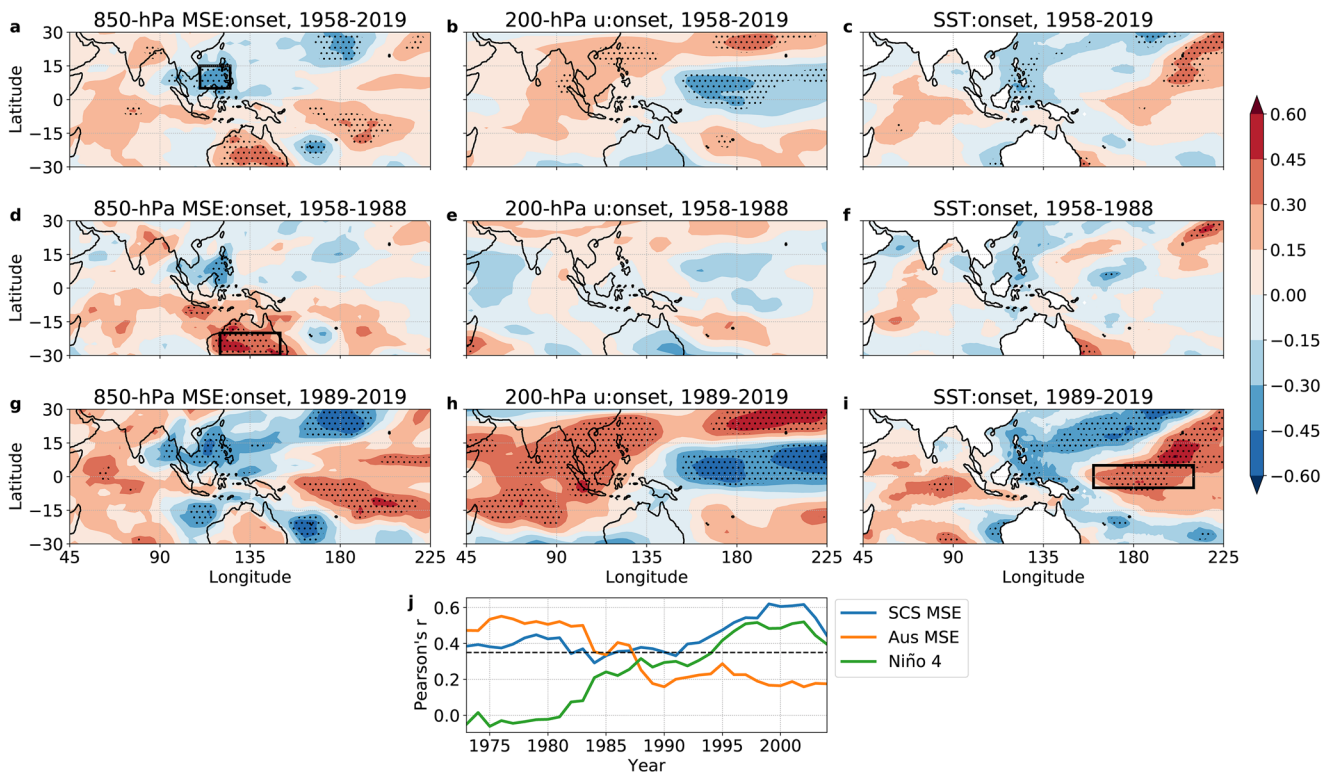
where  $c_p = 1,004.6 \text{ J K}^{-1} \text{ kg}^{-1}$  is the heat capacity at constant pressure,  $T$  is temperature,  $L = 2.507 \times 10^6 \text{ J kg}^{-1}$  is the latent heat of vapourization of water,  $q$  is specific humidity,  $g = 9.80665 \text{ m s}^{-2}$  is the gravitational constant, and  $z$  is geopotential height. Values for constants are those used in the JRA-55 reanalysis.

The connection between the overturning circulation and MSE distribution results in two feedbacks occurring during monsoon onset. First, diabatic heating by the insolation warms the summer hemisphere. In response, the ITCZ shifts into the summer hemisphere and the winter-hemisphere Hadley cell becomes cross-equatorial. This cross-equatorial cell advects cooler, drier air up the MSE gradient, while diabatic processes increase MSE poleward. As a result, the MSE maximum shifts farther poleward and the cell becomes more cross-equatorial. The result is a positive feedback between the circulation and the thermal forcing, so that the convergence zone jumps abruptly into the summer hemisphere (Bordoni & Schneider, 2008). The second feedback relates to the tropical upper-level easterlies generated by a cross-equatorial Hadley cell. These limit the propagation of eddies to lower latitudes. As a result, the cell becomes primarily thermally driven, rather than eddy-driven, and responds strongly to changes in the MSE distribution, strengthening the cell, and so further enhancing the easterlies (Geen et al., 2019; Schneider & Bordoni, 2008).

Although these ideas have been developed in an idealized framework, they appear to apply to both the climatology (Bordoni & Schneider, 2008; Geen et al., 2018; Ma et al., 2019; Nie et al., 2010) and interannual variability (Hurley & Boos, 2013) of the Asian monsoons in reanalysis data. In this study, I further hypothesize that in the months prior to monsoon onset, both local and remote influences may cause the atmosphere to be in a state where these feedback cycles will more readily begin, so that onset may then occur earlier in the season. Based on this, I suggest that early SCSM onset will be associated with positive 850-hPa MSE and negative 200-hPa zonal wind speed anomalies in the SCSM region. The present study is motivated by an interest in prediction of climate over China. However, some initial exploration of the correlations with monsoon onset over the Bay of Bengal and India are shown in Figure S1.

### 2.2. Data and Metrics

Results are presented for the JRA-55 reanalysis data set (Japan Meteorological Agency/Japan, 2013; Kobayashi et al., 2015) for years 1958–2019, with SSTs taken from the COBE SST data set (Japan Meteorological Agency, 2006; Japan Meteorological Agency, Ongoing). In addition, results are shown in supporting material (Figures S2 and S3) for the CERA-20C data set (Laloyaux et al., 2016), confirming that similar relationships are seen over a longer record. Daily mean 850-hPa zonal wind data were used to establish the SCSM onset pentad using the criteria developed by B. Wang et al. (2004). SCSM onset is defined as the first pentad after April 25th (pentad 24) where the average zonal wind speed over  $5^\circ\text{--}15^\circ\text{N}$ ,  $110^\circ\text{--}120^\circ\text{E}$ ,  $U_{SCS}$ , is westerly, and where  $U_{SCS}$  is positive in at least three of the four subsequent pentads (including the onset pentad) and the accumulative 4-pentad mean of  $U_{SCS} > 1 \text{ m s}^{-1}$ . The onset pentads identified in the JRA-55 data set are shown by the black circles in Figure 1. These are broadly consistent with dates evaluated in previous studies using the NCEP/NCAR (Kalnay et al., 1996; B. Wang et al., 2004) or ERA-Interim reanalyses (Martin et al., 2019; Uppala et al., 2005). JRA-55 data are presented here due to the long data record and use of 4Dvar data assimilation, but correlations were also checked using the ERA-Interim, NCEP/NCAR and NCEP/DOE-R2 (Kanamitsu et al., 2002) data sets, with similar conclusions obtained overall (not shown).

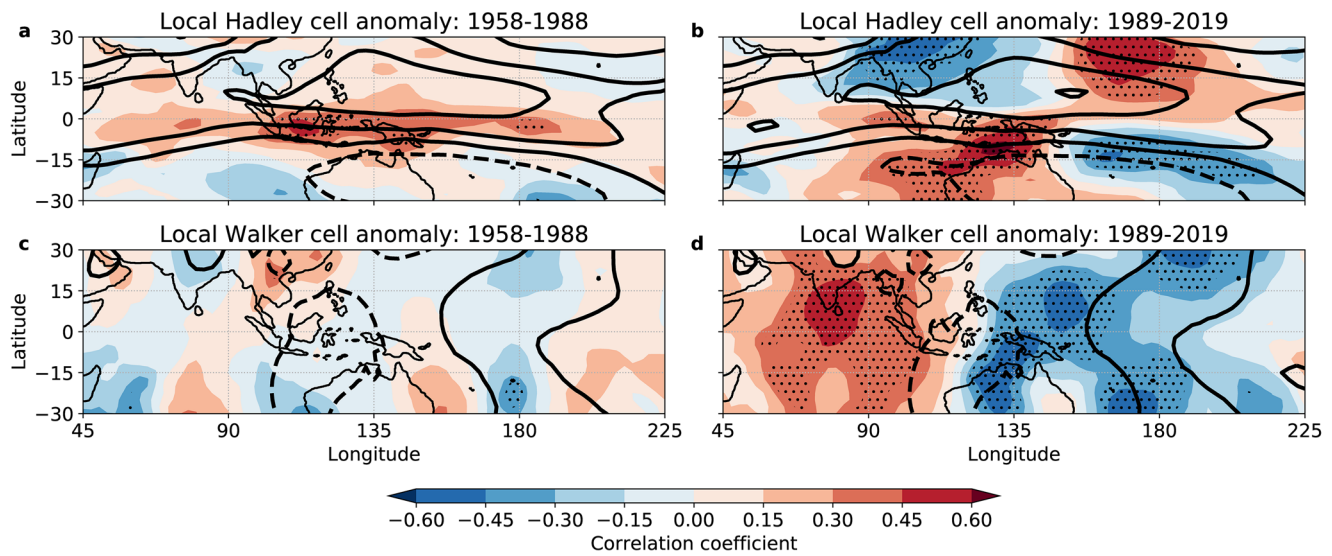


**Figure 2.** Correlations between potential predictors and the SCSM onset pentad. Panels ((a)–(h)) show maps of Pearson correlation coefficient between SCSM onset pentad and January–March averaged 850-hPa MSE (left column), 200-hPa zonal wind (center column), and SST (rightmost column). The top row shows the correlations for the full data set, second and third rows show correlations for early and late periods respectively (see titles). Bottom center (j) shows running correlations of SCSM onset pentad with: MSE averaged over 5°–15°N, 110–125°E (blue); MSE averaged over -30° to -20°N, 120–150°E (orange); SST averaged over the Niño 4 region (green). Averaging areas for MSE are indicated by boxes in (a) and (d), and for SST by the box in (i). Stippling on maps and the dashed line in (j) indicate the threshold where the correlation is significantly different from 0 ( $p < 0.05$ ). For all panels, data has been detrended relative to an 11-year rolling mean.

In this study, I aim to explore precursors for interannual variability, but slower decadal trends are also present in the data. When investigating correlations data were therefore detrended with an 11-year rolling mean, which is illustrated for the onset dates by the dashed line in Figure 1. This ensures that the correlations presented relate to interannual variability, rather than to coincident trends in variables due to, for example, global warming or variations in multi-decadal modes. Note that the initial and final 5 years are detrended using the mean of the initial and final 11 years to allow these to be included.

### 3. Hypothesis Testing

Figure 2 shows correlations between the hypothesized predictors, averaged from January to March, and the SCSM onset pentad, with both detrended with an 11-year rolling mean. A negative correlation indicates that a positive anomaly is associated with earlier monsoon onset. Looking first over the full reanalysis record, I see correlations that are consistent with the hypothesized relationships. Spring 850-hPa MSE over the South China Sea is negatively correlated with SCSM onset timing (Figure 2a), while 200-hPa zonal wind over East Asia is weakly positively correlated with monsoon onset (Figure 2b). MSE anomalies can be expected to relate to anomalies in SST and MSE advection. Over the full record, SCSM onset is correlated with SST over the Philippines to the east (Figure 2c), but the correlation is weaker than that with MSE. This suggests that the MSE pattern is partially, but not completely, related to local SST anomalies. Breaking down MSE into its contributions from internal, latent, and potential energy, the majority of the correlation was found to come from the latent heat (not shown). The seasonal evolution of the MSE correlation with SCSM onset is shown in Figure S1. A dipole is seen in April and May, with enhanced MSE over the SCS and reduced MSE to the south associated with earlier onset. By June onset has occurred in most years, but a significant



**Figure 3.** Correlations of January–March averaged local Hadley (top) and Walker (bottom) cell overturning (Schwendike et al., 2014) with the SCSM onset pentad. Left and right columns show correlations for 1958–1988 and 1989–2019, respectively. Thick dark lines show 500-hPa cross sections of the time-mean Hadley and Walker overturning circulations over the two periods, with contour interval  $3,000 \text{ kg s}^{-1} \text{ m}^{-1}$ . Stippling indicates where the correlation is significantly different from 0 ( $p < 0.05$ ). For all panels, data have been detrended relative to an 11-year rolling mean.

correlation remains to the south indicating that late onset is linked to enhanced MSE over and to the south of Indonesia.

To investigate whether these relationships are consistent throughout the reanalysis record, I also divided the data into two 31-year sections, an early period spanning 1958–1988 (Figures 2d–2f) and a later period spanning 1989–2019 (Figures 2f–2i). A statistically significant negative correlation between January–March 850-hPa MSE is present over the South China Sea for both of these time periods. In contrast, for upper-level zonal wind, it becomes clear that the correlation is dominated by the later period, with the earlier period showing no statistically significant relationship over East Asia. The mean SCSM onset pentad became earlier in 1994, but I find that dividing the data as 1958–1993 and 1994–2019 gives similar results (not shown).

It has been noted that the ENSO-SCSM relationship appears to have strengthened from the late 1970s onwards (B. Wang, Huang, et al., 2009) and that this provides a strong source of predictability for SCSM onset (Martin et al., 2019). The correlations over the later period are consistent with the ENSO teleconnection influence exerting a strong influence on SCSM onset. The MSE correlation pattern shows a clear East-West asymmetry across the Pacific Basin (Figure 2g), while the zonal wind correlation resembles the upper branch of the Walker cell (Figure 2h) and the SST correlation shows a clear ENSO pattern (Figure 2i). The lack of correlation of 200-hPa zonal wind with SCSM onset in the earlier period suggests two possibilities. First, upper-level easterly anomalies in the preceding spring may not causally connect to an earlier Hadley cell regime change as hypothesized in Section 2.1, and may instead simply be coincident with SST anomalies that result in warm, humid air converging over the SCS region. Alternatively, these upper-level easterlies may contribute causally to an earlier transition, but may only be steady if SST anomalies are present to support zonal flow anomalies. Systematic model simulations would be needed to distinguish these possibilities. Overall, I conclude that upper-level zonal wind does not provide a steady predictor for SCSM onset.

In contrast, in the earlier period there is no clear connection between ENSO and SCSM onset. Instead, the strongest correlation is found to be with MSE over Australia (Figure 2d). This correlation is not captured by looking purely at SSTs (Figure 2f) and is predominantly due to the latent heat (not shown). Specifically, in this period, higher MSE over Australia from January–March was associated with later onset of the SCSM. The dipole around the Equator in Figure 2d suggests that meridional thermal gradients, and their influence on the Hadley circulation, were more important precursors for monsoon onset in this period, compared with the role of zonal thermal gradients and the Walker circulation in the later period. This is supported by Figure 3, which shows the correlation of onset with the local Hadley and Walker cells, as defined by

Schwendike et al. (2014). In the earlier period, Walker circulation anomalies show little correlation with SCSM onset timing, but a weakening of the Northern Hemisphere Hadley cell in January–March is associated with earlier monsoon onset. In the later period, a strong correlation can be seen between both the Walker and Hadley circulations and SCSM onset, reflecting the influence of ENSO on the circulation.

Figure 2j shows running correlations with a 31-year mean, indicating how the teleconnections vary in time. The blue and orange lines show MSE averaged over the boxes shown in panels (a) and (d), while the green line shows the SST averaged over the Niño 4 region (panel (i)). For SCS MSE, a statistically significant negative correlation is found to be present over almost the entire record, although the strength of the correlation does vary in time. The correlation of onset with Australian MSE is in fact stronger than that with SCS MSE, but begins to drop off after 1983, while the relationship with Niño 4 SST strengthens at this point. SCS MSE approximately follows the correlation with both connections but is consistent across the record. Previous studies (Zhu & Li, 2017) identified different precursors to those presented here; I find that rolling correlations of these also show strong interdecadal variations (Figure S4).

Figure S2 confirms that similar behavior is seen in the CERA-20C data set. Here, the correlation does dip below the 95% confidence level, but nonetheless remains more steady than other teleconnections. I note that all correlations are low in the very early years of the data set (prior to ~1930). This might result from either a lack of predictability or sparse observations in this period; an in depth analysis of the available observations would be needed to explore this issue.

#### 4. A Simple Forecast

The correlations in Figure 2 suggest that the area mean MSE over 5°–15°N, 110°–125°E could provide a useful predictor of SCSM onset. To assess the predictive skill, I test how well data from previous years can be used to predict the onset timing for the next year. To mimic a plausible operational forecast, an expanding window method is applied to the un-detrended data. The prediction for a given year is estimated by using least squares regression to fit a linear model between the observed onset dates and SCS MSE from all previous years. Each year the window used for generating the forecast expands as new observations are incorporated. This approach was repeated using the SST averaged over the Niño 4 region to confirm the skill is not purely a result of the correlation with ENSO and to provide a model for comparison in periods not covered by previous studies (specifically Martin et al., 2019; Zhu & Li, 2017).

The forecasts produced are shown in Figure 1a from 1968 onwards, providing 10 years of training data for the first forecast plotted. Figure 1b shows rolling skill metrics evaluated with a 31-year window (cf. Figure 2j). Note that skill is expected to be initially low, as the early years are forecast based on a limited amount of data. An equivalent forecast for CERA-20C is given in Figure S3, helping to distinguish whether lower skill results from a lack of training data or reduced correlation with MSE. In general, for both JRA-55 and CERA-20C, the SCS MSE based forecast has a higher correlation with the observed onset pentads and a lower RMSE than the forecast using the Niño 4 SST, although there are short windows where the Niño 4 forecast shows higher correlation with the observed dates. Skill is low for both models prior to 1930, reflecting the weak correlations seen in Figure 2j in this period.

Averaging over a subset of years allows comparison with previous efforts with more complex models. Zhu and Li (2017) applied three predictors to model the SCSM onset dates in the NCEP/DOE-R2 data set, achieving a correlation over their test period, 2005–2014, of 0.72 (RMSE 2.08 pentads). Martin et al. (2019) found that the Met Office GloSea5 ensemble (MacLachlan et al., 2015; Williams & Coauthors, 2015) could predict SCSM onset with a correlation of 0.5 over a study period from 1993 to 2015. Over these periods, the predictions using SCS MSE show correlations with observed onset dates of 0.70 ( $p = 0.03$ ; RMSE 1.96 pentads) and 0.67 ( $p = 0.0005$ ; RMSE 2.00 pentads), respectively. The Niño 4 forecasts show lower skill: 0.23 ( $p = 0.5$ ; RMSE 2.93 pentads) and 0.27 ( $p = 0.2$ ; RMSE 2.88 pentads). It is worth noting that the correlations and forecast skill vary interdecadally (Figures 1, 2, S2 and S3), and that correlations were relatively strong throughout 1993–2015. The stronger correlation with ENSO in the later period may have given optimistic estimates of forecast skill in previous studies. Where possible, it would be helpful to produce longer hindcasts to give a more complete a picture of model skill.

Repeating the predictions using only MSE averaged in January–February or January, I find the correlations also remain high when only earlier data are used. For the 2005–2014 and 1993–2015 periods, January–February averaged MSE gives correlations of 0.69 (RMSE 1.92 pentads) and 0.65 (RMSE 1.98 pentads), respectively, while January-mean MSE gives a correlations of 0.55 (RMSE 2.13 pentads) and 0.60 (RMSE 2.06 pentads). Using only a single precursor, this simple model is able to give an initial estimate of SCSM onset timing with roughly 3–4 months lead time (given that onset occurs in May on average), and shows correlations competitive with previous models.

Last, I note that the SCS MSE-based forecast in Figure 1 appears reasonably well correlated with the observed variability, but does not capture the extremes in onset timing, which might have the highest impact on agriculture. I find that, for the JRA-55 data, this issue appears to be improved by applying exponential smoothing to account for non-stationary statistics in the forecasting model. For example, with exponential smoothing applied RMSE decreases for both the GloSea5 and Zhu study periods (Figure S5). However, this improvement was not reproduced when forecasting the longer CERA-20C record. An explanation of the method and figures showing the forecast generated are included in the supporting information (Text S1 and Figure S5) for interested readers.

## 5. Discussion

Based on theoretical insights into controls on the meridional overturning circulation that have been developed in aquaplanets, I set out with two hypothesized predictors for SCSM onset: lower-level MSE and upper-level zonal wind in the preceding January–March. While the latter does not correlate well with SCSM onset timing, I find that MSE is a useful predictor of interannual variability in SCSM onset. Although the strength of individual teleconnections to the South China Sea varies in time, looking at local MSE allows us to take a step farther along the mechanistic chain from a remote forcing to a local impact on the monsoon onset. Multiple processes can produce local MSE anomalies, but these anomalies are consistently correlated with SCSM onset timing across the JRA-55 record (Figure 2) and additionally show relatively steady correlations throughout twentieth century in the CERA-20C data set (Figure S2). The correlation strength was also tested in other reanalysis data sets (not shown). In ERA-Interim, similar correlation patterns and strengths are seen to those shown in Figure 2. Correlations in NCEP/NCAR and NCEP/DOE-R2 were also found to follow similar patterns, although these are weaker in magnitude. These findings are thus robust across data sets.

Generating a simple linear regression model based on this single predictor, I produced predictions of onset timing from 1993 onwards via an expanding-window approach. I find predictive skill comparable with ensemble forecast results from as early as January. I conclude that local MSE in the months preceding SCSM onset is a useful source of predictability and would merit further exploration both for use in Physical-Empirical forecast models and for guiding development of dynamical forecasting ensembles.

Despite this favorable comparison in skill, onset in several years was poorly predicted. Subseasonal factors such as intraseasonal oscillations (R. Wu, 2010; Zhou & Chan, 2005) and tropical cyclones (B. Liu & Zhu, 2020; Mao & Wu, 2008) have been found to trigger onset in individual years, limiting seasonal predictability. Although MSE anomalies do appear to precondition the region for early or late onset, it is therefore also highly important to account for sub-seasonal systems that may cause deviation from the seasonal forecast. Recent results suggest that these problems of seasonal and sub-seasonal prediction may be best approached in parallel: the sub-seasonal character of SCSM onset varies between early and late onsets, so that an improved sub-seasonal prediction of SCSM onset is expected to benefit from the consideration of interannual variability (H. Wang et al., 2018). I also note that although the correlation indicates skill, the predictions in Figure 1 do not capture the extremes in onset well. Exponential smoothing was explored as a method to address this issue (Text S1 and Figure S5).

The present study focuses on the SCSM onset, motivated by an interest in seasonal prediction of Chinese climate. However, it is also interesting to explore if this relationship with MSE applies elsewhere. Figure S1 shows how the correlation with MSE evolves over Spring and Summer for the SCSM, Bay of Bengal monsoon (BOBM), and Indian summer monsoon (ISM). For consistency, onset pentads for the BOBM and ISM are evaluated using 850-hPa wind based metrics (Mao & Wu, 2007; B. Wang, Ding, et al., 2009), with average



pentads 24 and 30, respectively. A negative correlation with MSE over the Bay of Bengal is not seen for BOBM onset. Instead the dominant signals are positive correlations over the Indian Ocean and central Pacific. These indicate that the BOBM onset is strongly related to ENSO, but that delayed onset is also associated with enhanced MSE to the south, physically consistent with the aquaplanet physics. However, I note that in the BOBM region the reversal of the meridional MSE gradient does not appear to be concurrent with onset and that the 850-hPa wind reversal used to define onset is here associated with the development of a shallow, rather than deep, overturning circulation (not shown). MSE reversal occurs later in approximately pentad 28, at which time the cross-equatorial cell deepens. The aquaplanet dynamics may therefore not be a suitable model for BOBM onset.

For the ISM, a weak correlation with MSE over the Tibetan Plateau is seen in January–March, but does not persist into April. A strong, negative correlation over India develops in May, which develops into a dipole in June, similar to that seen for the SCSM (left column). ISM onset occurs in pentad 30 on average, corresponding to the end of May, with a standard deviation of 1.7 pentads. The correlation seen in May might, therefore, indicate some late-stage predictability from MSE or could simply reflect the increase in MSE that occurs during monsoon onset.

This theory-motivated approach has successfully identified a predictor whose correlation with SCSM onset persists across the record and highlights the direct benefits of better understanding the controls on the large-scale circulation. Consistent behavior is observed in other regions but with reduced lead time. Further work is needed to examine where the aquaplanet is an appropriate model for monsoon onset and where MSE might be used to predict regional monsoon onset. However, the second proposed predictor, 200-hPa zonal wind, did not correlate well with monsoon onset. This could relate to a lack of memory in the system for upper-level wind anomalies or could indicate issues with applying ideas from highly idealized models to Earth. The theoretical foundation for understanding the climatological monsoon still shows some key gaps, in particular the role of zonal asymmetries and transient weather systems in the seasonality of the Hadley cells (Geen et al., 2020). The results presented here suggest that bridging these gaps may provide further opportunities for improved seasonal forecasts.

### Data Availability Statement

Data sets for this research are available in these in-text data citation references: Japan Meteorological Agency/Japan (2013); Japan Meteorological Agency (Ongoing); ECMWF (2016).

### Acknowledgments

The work was supported by the UK-China Research and Innovation Partnership Fund, through the Met Office Climate Science for Service Partnership (CSSP) China, as part of the Newton Fund. The author thanks Phil Sansom, Adam Scaife, and Gill Martin for useful discussion on forecasting and cross-validation approaches, and two anonymous reviewers for their helpful comments.

### References

Bordoni, S., & Schneider, T. (2008). Monsoons as eddy-mediated regime transitions of the tropical overturning circulation. *Nature Geoscience*, 1(8), 515–519. <https://doi.org/10.1038/ngeo248>

Chang, C.-P. (2004). East Asian Monsoon (Vol. 2). World Scientific. <https://doi.org/10.1142/5482>

Chan, J. C. L., & Zhou, W. (2005). PDO, ENSO and the early summer monsoon rainfall over south China. *Geophysical Research Letters*, 32(8). <https://doi.org/10.1029/2004GL022015>

Cui, K., & Shoemaker, S. P. (2018). A look at food security in China. *Npj Science of Food*, 2(4). <https://doi.org/10.1038/s41538-018-0012-x>

Ding, Q., & Wang, B. (2005). Circumglobal teleconnection in the Northern Hemisphere summer. *Journal of Climate*, 18(17), 3483–3505. <https://doi.org/10.1175/JCLI3473.1>

ECMWF. (2016). *CERA-20C*. Reading, UK: European Centre for Medium-Range Weather Forecasts. Retrieved from <https://apps.ecmwf.int/datasets/data/cera20c>

Fan, Y., Fan, K., & Tian, B. (2016). Has the prediction of the South China Sea summer monsoon improved since the late 1970s? *Journal of Meteorological Research*, 30(6), 833–852. <https://doi.org/10.1007/s13351-016-6052-8>

Fan, Y., Fan, K., Xu, Z., & Li, S. (2018). ENSO-South China Sea summer monsoon interaction modulated by the Atlantic multidecadal oscillation. *Journal of Climate*, 31(8), 3061–3076. <https://doi.org/10.1175/JCLI-D-17-0448.1>

Geen, R., Bordoni, S., Battisti, D. S., & Hui, K. (2020). Monsoons, ITCZs and the concept of the global monsoon. *Reviews of Geophysics*, 58(4), e2020RG000700. <https://doi.org/10.1029/2020RG000700>

Geen, R., Lambert, F. H., & Vallis, G. K. (2018). Regime change behavior during Asian monsoon onset. *Journal of Climate*, 31, 3327–3348. <https://doi.org/10.1175/JCLI-D-17-0118.1>

Geen, R., Lambert, F. H., & Vallis, G. K. (2019). Processes and timescales in onset and withdrawal of ‘aquaplanet monsoons’. *Journal of the Atmospheric Sciences*, 76, 2357–2373. <https://doi.org/10.1175/JAS-D-18-0214.1>

He, J., & Zhu, Z. (2015). The relation of South China Sea monsoon onset with the subsequent rainfall over the subtropical East Asia. *International Journal of Climatology*, 35(15), 4547–4556. <https://doi.org/10.1002/joc.4305>

Hill, S. A. (2019). Theories for past and future monsoon rainfall changes. *Current Climate Change Reports*, 5, 160–171. <https://doi.org/10.1007/s40641-019-00137-8>

- Huang, R., Gu, L., Zhou, L., & Wu, S. (2006). Impact of the thermal state of the tropical western pacific on onset date and process of the South China Sea summer monsoon. *Advances in Atmospheric Sciences*, 23, 909–924. <https://doi.org/10.1007/s00376-006-0909-1>
- Hu, P., Chen, W., Huang, R., & Nath, D. (2018). On the weakening relationship between the South China Sea summer monsoon onset and cross-equatorial flow after the late 1990s. *International Journal of Climatology*, 38, 3202–3208. <https://doi.org/10.1002/joc.5472>
- Hurley, J. V., & Boos, W. R. (2013). Interannual variability of monsoon precipitation and local subcloud equivalent potential temperature. *Journal of Climate*, 26(23), 9507–9527. <https://doi.org/10.1175/JCLI-D-12-00229.1>
- Japan Meteorological Agency/Japan. (2013). *JRA-55: Japanese 55-year reanalysis, monthly means and variances*. Boulder, CO: Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory. Retrieved from <https://doi.org/10.5065/D60G3H5B>
- Japan Meteorological Agency. (2006). Characteristics of Global Sea Surface Temperature Analysis Data (COBE-SST) for climate use. *Monthly Report on Climate System Separated*, Volume, 12, 116.
- Japan Meteorological Agency (Ongoing). *COBE-SST*. Author. Retrieved from <http://ds.data.jma.go.jp/tcc/tcc/products/elnino/cobesst/cobe-sst.html>
- Kajikawa, Y., & Wang, B. (2012). Interdecadal change of the South China Sea monsoon onset. *Journal of Climate*, 25, 3207–3218. <https://doi.org/10.1175/JCLI-D-11-00207.1>
- Kajikawa, Y., Yasunari, T., Yoshida, S., & Fujinami, H. (2012). Advanced Asian summer monsoon onset in recent decades. *Geophysical Research Letters*, 39(3), 1–5. <https://doi.org/10.1029/2011GL050540>
- Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., et al. (1996). The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, 77(3), 437–472. [https://doi.org/10.1175/1520-0477\(1996\)077<0437:TNYRP>2.0.CO;2](https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2)
- Kanamitsu, M., Ebisuzaki, W., Woollen, J., Yang, S.-K., Hnilo, J. J., Fiorino, M., & Potter, G. L. (2002). NCEP–DOE AMIP–II reanalysis (R–2). *Bulletin of the American Meteorological Society*, 83(11), 1631–1644. <https://doi.org/10.1175/BAMS-83-11-1631>
- Kobayashi, S., Ota, Y., Harada, Y., Ebata, A., Moriya, M., Onoda, H., et al. (2015). The JRA-55 reanalysis: General specifications and basic characteristics. *Journal of the Meteorological Society of Japan*, 93(1), 5–48. <https://doi.org/10.2151/jmsj.2015-001>
- Lalouaux, P., Balmaseda, M., Dee, D., Mogensen, K., & Janssen, P. (2016). A coupled data assimilation system for climate reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 142(694), 65–78. <https://doi.org/10.1002/qj.2629>
- Lau, K. M., & Yang, S. (1997). Climatology and interannual variability of the southeast Asian summer monsoon. *Advances in Atmospheric Sciences*, 14, 141–162. <https://doi.org/10.1007/s00376-997-0016-y>
- Lin, A., Zhang, R., & He, C. (2017). The relation of cross-equatorial flow during winter and spring with South China Sea summer monsoon onset. *International Journal of Climatology*, 37, 4576–4585. <https://doi.org/10.1002/joc.5106>
- Liu, B., & Zhu, C. (2020). Boosting effect of tropical cyclone “fani” on the onset of the South China Sea summer monsoon in 2019. *Journal of Geophysical Research: Atmospheres*, 125, e2019JD031891. <https://doi.org/10.1029/2019JD031891>
- Liu, P., Qian, Y., & Huang, A. (2009). Impacts of land surface and sea surface temperatures on the onset date of the South China Sea summer monsoon. *Advances in Atmospheric Sciences*, 26, 493–502. <https://doi.org/10.1007/s00376-009-0493-2>
- MacLachlan, C., Arribas, A., Peterson, K. A., Maidens, A., Fereday, D., Scaife, A. A., et al. (2015). Global seasonal forecast system version 5 (glosea5): A high-resolution seasonal forecast system. *Quarterly Journal of the Royal Meteorological Society*, 141, 1072–1084. <https://doi.org/10.1002/qj.2396>
- Ma, D., Sobel, A. H., Kuang, Z., Singh, M. S., & Nie, J. (2019). A moist entropy budget view of the south Asian summer monsoon onset. *Geophysical Research Letters*, 46(8), 4476–4484. <https://doi.org/10.1029/2019GL082089>
- Mao, J., & Wu, G. (2007). Interannual variability in the onset of the summer monsoon over the Eastern Bay of Bengal. *Theoretical and Applied Climatology*, 89, 155–170. <https://doi.org/10.1007/s00704-006-0265-1>
- Mao, J., & Wu, G. (2008). Influences of Typhoon Chanchu on the 2006 South China Sea summer monsoon onset. *Geophysical Research Letters*, 35, L12809. <https://doi.org/10.1029/2008GL033810>
- Martin, G. M., Chevuturi, A., Comer, R. E., Dunstone, N. J., Scaife, A. A., & Zhang, D. (2019). Predictability of South China Sea summer monsoon onset. *Advances in Atmospheric Sciences*, 36(3), 253–260. <https://doi.org/10.1007/s00376-018-8100-z>
- Nie, J., Boos, W. R., & Kuang, Z. (2010). Observational evaluation of a convective quasi-equilibrium view of monsoons. *Journal of Climate*, 23(16), 4416–4428. <https://doi.org/10.1175/2010JCLI3505.1>
- Privé, N. C., & Plumb, R. A. (2007). Monsoon dynamics with interactive forcing. Part I: Axisymmetric studies. *Journal of the Atmospheric Sciences*, 64(5), 1417–1430. <https://doi.org/10.1175/JAS3916.1>
- Schneider, T., & Bordoni, S. (2008). Eddy-Mediated regime transitions in the seasonal cycle of a hadley circulation and implications for monsoon dynamics. *Journal of the Atmospheric Sciences*, 65(1), 915–934. <https://doi.org/10.1175/2007JAS2415.1>
- Schwendike, J., Govekar, P., Reeder, M. J., Wardle, R., Berry, G. J., & Jakob, C. (2014). Local partitioning of the overturning circulation in the tropics and the connection to the Hadley and Walker circulations. *Journal of Geophysical Research*, 119(3), 1322–1339. <https://doi.org/10.1002/2013JD020742>
- Uppala, S. M., Kållberg, P., Simmons, A., Andrae, U., Bechtold, V. D. C., Fiorino, M., et al. (2005). The ERA-40 re-analysis. *Quarterly Journal of the Royal Meteorological Society*, 131, 2961–3012. <https://doi.org/10.1256/qj.04.176>
- Wang, B., Ding, Q., & Joseph, P. V. (2009). Objective definition of the Indian summer monsoon onset. *Journal of Climate*, 22(12), 3303–3316. <https://doi.org/10.1175/2008JCLI2675.1>
- Wang, B., Huang, F., Wu, Z., Yang, J., Fu, X., & Kikuchi, K. (2009). Multi-scale climate variability of the South China Sea monsoon: A review. *Dynamics of Atmospheres and Oceans*, 47(1), 15–37. <https://doi.org/10.1016/j.dynatmoce.2008.09.004>
- Wang, B., & LinHo (2002). Rainy season of the Asian-Pacific summer monsoon. *Journal of Climate*, 15, 386–398. [https://doi.org/10.1175/1520-0442\(2002\)015%3C0386:RSOTAP%3E2.0.CO;2](https://doi.org/10.1175/1520-0442(2002)015%3C0386:RSOTAP%3E2.0.CO;2)
- Wang, B., LinHo, Zhang, Y., & Lu, M. M. (2004). Definition of South China Sea monsoon onset and commencement of the East Asian summer monsoon. *Journal of Climate*, 17(4), 699–710. <https://doi.org/10.1175/2932.1>
- Wang, H., Liu, F., Wang, B., & Li, T. (2018). Effects of intraseasonal oscillation on South China Sea summer monsoon onset. *Climate Dynamics*, 51, 2543–2558. <https://doi.org/10.1007/s00382-017-4027-9>
- Williams, K. D., & Coauthors (2015). The Met Office Global Coupled model 2.0 (GC2) configuration. *Geoscientific Model Development*, 8(5), 1509–1524. <https://doi.org/10.5194/gmd-8-1509-2015>
- Wu, G., & Zhang, Y. (1998). Tibetan plateau forcing and the timing of the monsoon onset over South Asia and the South China Sea. *Monthly Weather Review*, 126(4), 913–927. [https://doi.org/10.1175/1520-0493\(1998\)126<0913:TPFATT>2.0.CO;2](https://doi.org/10.1175/1520-0493(1998)126<0913:TPFATT>2.0.CO;2)
- Wu, R. (2010). Subseasonal variability during the South China Sea summer monsoon onset. *Climate Dynamics*, 34, 629–642. <https://doi.org/10.1007/s00382-009-0679-4>

- Zhou, W., & Chan, J. C. L. (2005). Intraseasonal oscillations and the South China Sea summer monsoon onset. *International Journal of Climatology*, 25, 1585–1609. <https://doi.org/10.1002/joc.1209>
- Zhou, W., & Chan, J. C. L. (2007). ENSO and the South China Sea summer monsoon onset. *International Journal of Climatology*, 27, 157–167. <https://doi.org/10.1002/joc.1380>
- Zhu, Z., & Li, T. (2017). Empirical prediction of the onset dates of South China Sea summer monsoon. *Climate Dynamics*, 48(5–6), 1633–1645. <https://doi.org/10.1007/s00382-016-3164-x>

### References From the Supporting Information

- Agnew, R. A. (1982). Econometric forecasting via discounted least squares. *Naval Research Logistics Quarterly*, 29(2), 291–302. <https://doi.org/10.1002/nav.3800290210>
- Brown, R. G., & Meyer, R. F. (1961). The fundamental theorem of exponential smoothing. *Operations Research*, 9(5), 673–685. <https://doi.org/10.1287/opre.9.5.673>
- Johnstone, R. M., Johnson, C. R., Jr., Bitmead, R. R., & Anderson, B. D. O. (1982). Exponential convergence of recursive least squares with exponential forgetting factor. *Systems & Control Letters*, 2(2), 77–82. [https://doi.org/10.1016/S0167-6911\(82\)80014-5](https://doi.org/10.1016/S0167-6911(82)80014-5)
- Young, P., & Ord, J. K. (1985). The use of discounted least squares in technological forecasting. *Technological Forecasting and Social Change*, 28(3), 263–274. [https://doi.org/10.1016/0040-1625\(85\)90048-4](https://doi.org/10.1016/0040-1625(85)90048-4)