



# MS8401 (2022 春)课程项目

危国锐 120034910021

(上海交通大学海洋学院, 上海 200030)

**摘要:** 在气候学研究中,经常需要将一个时间序列分解为长期趋势、季节信号、余项三部分,称这种分解为(加性)经典分解.经典分解方法有多种,各方法在不同的研究中有应用.它们的分解结果差异有多大?影响各方法间分解差异的因素是什么?我们应如何对待这些差异?为了研究这些问题,本文将常用的经典分解方法分为简单类、预估-校正类、全局模型类三个大类,每大类包含几种方法.本文通过设计理想试验,发现 a) 剧烈的长期趋势, b) 强非线性的长期趋势, c) 大的余项方差和 d) 大幅度的季节信号,是导致各方法间差异的主要原因.本文将三大类经典分解方法应用于 ERSST v5 数据,发现各方法间差异在某些区域较大,可接近  $0.3^{\circ}\text{C}$ .进一步分析造成这些差异的原因,猜想在中纬度地区,因素 d 起主导作用,各方法(主要是 M-1A 与其余三种)的输出余项差异可接近  $0.3^{\circ}\text{C}$ ;在赤道东太平洋,因素 c 起一定作用,各方法的输出余项差异可超过  $0.1^{\circ}\text{C}$ ;在北冰洋部分区域,因素 a 起的作用相对大,影响输出季节信号和余项均值,但对输出余项影响不大;在其余大部分地区,各因素作用较小,各方法输出余项差异一般不超过  $0.1^{\circ}\text{C}$ .在气候变化的背景下,这些差异在未来可能更大.通过本文的讨论,我们指出,气候态(包括长期趋势,季节分量等)应是客观的,但趋势和季节周期的定义,以及相应的分解方案及其参数选取却带有主观性,这是分解方案选取问题的根源.为此,我们建议后续研究对气候态各分量建立客观定义和相应的自适应分解方案;在此实现之前,气候学研究应关注时间序列经典分解方法误导科学结论的可能.

**关键词:** 词 1, 词 2

## Course Project

Guorui Wei 120034910021

(School of Oceanography, Shanghai Jiao Tong University, Shanghai 200030, China)

**Abstract:** The purpose of this paper is to compare several commonly used classical decomposition (seasonal adjustment) methods systematically and to provide suggestions for future research.

**Keywords:** keyword 1, keyword 2



# Table of Contents

摘要 .....	i
Abstract.....	i
Table of Contents.....	ii
<b>1 Introduction.....</b>	<b>1</b>
<b>2 Data and Method.....</b>	<b>2</b>
2.1 Data .....	2
2.1.1 NOAA Extended Reconstructed SST V5 (ERSST v5) .....	2
2.2 Classical decomposition.....	2
2.2.1 Simple method .....	2
2.2.2 Predictor-corrector method .....	4
2.2.3 Global model-fitting method.....	6
2.3 Experiment setup.....	7
2.3.1 Idealized time series .....	7
2.3.2 Real time-indices.....	8
2.4 Assessment methodology .....	8
2.4.1 Idealized time series .....	8
2.4.2 Real time-indices.....	8
<b>3 Results .....</b>	<b>9</b>
3.1 Idealized experiments.....	9
3.1.1 Classical decomposition results .....	9
3.1.2 Comparison of different methods.....	15
3.2 Classical decomposition of several real time-indices.....	30
3.2.1 ERSST v5 global monthly SST (1902-2021).....	30
3.2.2 Comparison of different methods.....	41
<b>4 Discussion.....</b>	<b>46</b>
4.1 经典分解的基本假定及合理性.....	46
4.2 对时间序列经典分解方案的物理-数学理解.....	46
4.3 造成各分解方式间差异的因素.....	46
4.4 预估-校正的收益 .....	46
4.5 时间序列分解质量评价.....	46
4.6 年际及以上变率的数学期望和样本均值.....	46
4.7 未来研究建议.....	46
<b>5 Conclusions.....</b>	<b>47</b>
<b>References .....</b>	<b>48</b>





# 1 Introduction

A time series of a climate variable can be viewed as a single realization of a stochastic process that represents the mechanism generating it. One goal of the time-series analysis is to figure out (estimate) the underlying mechanism of the time series. It is usually a daunting challenge to accomplish this goal, because of the lack of high-quality observations (limited samples), and the non-linear and non-stationary nature of the process. To alleviate this, climatologists may want to decompose the data generation mechanism into a somewhat deterministic component and a stochastic component. For example, if the time series of a variable exhibits a pattern of annual cycle superimposed on a long-term trend, one may want to do the (*additive*) *classical decomposition*, which decomposes its underlying process into a long-term trend component, a seasonal component, and a residual component. This paper will focus on the analysis of this type of variable.

From the observed time series, an estimate of the first two deterministic components can be drawn. Once the decomposition is done, the removal of the deterministic component from the original time series may facilitate the analysis of the remaining part. However, this decomposition may also produce ambiguous results, since it is not easy to distinguish long-term trends and seasonal components from the residual. Different decomposition methods may produce different results, which may have implications for the classification and analysis of climate events.

There are at least three kinds of procedures that are commonly used for the classical decomposition of a given time series. The first, which we call the *Simple method*, is to determine the long-term trend using either moving mean or locally linear regression, and then determine the annual cycle by averaging the state of the detrended time series concerning a fixed phase of the cycle, or by fitting the detrended time series to a sum of sines and cosines. The second, which we call the *Predictor-corrector method*, is to do method-1 iteratively, that is, update the estimated trend using the deseasonalized series, and then update the estimated seasonal component using the newly detrended series, and so on, until convergence criteria are met. The third, which we call the *Global model-fitting method*, is to determine the long-term trend and seasonal components simultaneously, by fitting the time series to a function containing a global linear trend and some sine and cosine terms.

Previous studies (Chen & Li, 2021; Findley et al., 1998; Narapusetty et al., 2009; Pezzulli et al., 2005; Wu et al., 2007)..., (Hua et al., 2018, 2022; Qin et al., 2020)... (文献综述待补充). However, few of those ... (研究 gap 待补充). The purpose of this paper is to compare several commonly used classical decomposition (*seasonal adjustment*) methods systematically and to provide suggestions for future research.

This paper is organized as follows. (the plan of development 待补充)

## 2 Data and Method

### 2.1 Data

#### 2.1.1 NOAA Extended Reconstructed SST V5 (ERSST v5)

“A global monthly SST analysis from 1854 to the present derived from ICOADS data with missing data filled in by statistical methods.” ()

“ERSST is suitable for long-term global and basin-wide studies, and smoothed local and short-term variations are used in the dataset.” ()

NOAA Extended Reconstructed SST V5 data provided by the NOAA PSL, Boulder, Colorado, USA, from their website at <https://psl.noaa.gov>.

Citation: Boyin Huang, Peter W. Thorne, Viva F. Banzon, Tim Boyer, Gennady Chepurin, Jay H. Lawrimore, Matthew J. Menne, Thomas M. Smith, Russell S. Vose, and Huai-Min Zhang (2017): NOAA Extended Reconstructed Sea Surface Temperature (ERSST), Version 5. [indicate subset used]. NOAA National Centers for Environmental Information. doi:10.7289/V5T72FNM. Obtain at NOAA/ESRL/PSD at their website <https://www.esrl.noaa.gov/psd/> [2022-07-23].

### 2.2 Classical decomposition

*Classical decomposition* is to decompose a time series into the following three components: 1) a long-term trend, 2) an annual cycle, or seasonal component, and 3) a residual component. In climatological studies, these three terms are usually assumed to be combined in an additive manner. The trend plus the annual cycle is sometimes referred to as the *climatological mean*.

Typically, climate data are sampled at equal time intervals. For simplicity, the following discussion will be focused on monthly data. However, the following discussions can easily be generalized to daily data, or data obtained by sampling at other time intervals.

Denote the original time series as  $x_i := x(t_i)$ ,  $i = 1, \dots, N$ , where  $t_i$  is the time value of the  $i^{\text{th}}$  month. One can simply let  $t_i := i$ .

#### 2.2.1 Simple method

1 Trend determined by locally weighted linear regression (LWLR)

For each time  $t$ , the *trend* component  $x_{\text{trend}}(t)$  of the original series is determined through the following locally weighted linear regression (LWLR) procedure.

The goal of LWLR is to find a model of the form  $\hat{x}(t) = \hat{\beta}_0 + \hat{\beta}_1 t$  that best fits the original series in the sense of least mean square loss, that is, to find  $\hat{\beta} := (\hat{\beta}_0, \hat{\beta}_1)^{\text{T}}$  that minimize

$$(\mathbf{x} - \mathbf{M}\hat{\beta})^{\text{T}} \mathbf{W}(\mathbf{x} - \mathbf{M}\hat{\beta}) = \sum_{i=1}^N w_i \left( x_i - (\hat{\beta}_0 + \hat{\beta}_1 t) \right)^2,$$

where  $\mathbf{x}$  is the *response vector*,  $\mathbf{M}$  is the *data matrix*, and  $\mathbf{W}$  is the diagonal *weight matrix*, which are defined as

$$\mathbf{x} := (x_1, \dots, x_N)^{\text{T}}, \quad \mathbf{M} := \begin{pmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_N \end{pmatrix}, \quad \mathbf{W} := \begin{pmatrix} w_1 & & \\ & \ddots & \\ & & w_N \end{pmatrix}.$$

The solution of  $\hat{\beta}$  is

$$\hat{\beta} := \operatorname{argmin}_{\beta \in \mathbb{R}^2} \{(\mathbf{x} - \mathbf{M}\beta)^T \mathbf{W}(\mathbf{x} - \mathbf{M}\beta)\} = (\mathbf{M}^T \mathbf{W} \mathbf{M})^{-1} \mathbf{M}^T \mathbf{W} \mathbf{x}. \quad (1)$$

The trend component is determined by

$$x_{\text{trend}}(t) = \hat{x}(t) = \hat{\beta}_0 + \hat{\beta}_1 t. \quad (2)$$

Note that  $\mathbf{W}$  depends on query time  $t$ , so the above LWLR procedure should be executed individually for each  $t$ .

Let  $\mathbf{x}_{\text{trend}} := (x_{\text{trend}}(t_1), \dots, x_{\text{trend}}(t_N))^T$ , the trend series can be written as

$$\mathbf{x}_{\text{trend}} := \mathbf{M}\hat{\beta}. \quad (2)'$$

## 2 Seasonal components determined by simple averaging

The seasonal component can be determined by the *simple averaging* method, in which the time series is averaged concerning a fixed phase of the cycle. Specifically, the seasonal component

$$x_{\text{season}}(t_i) := \begin{cases} \frac{1}{M} \sum_{j=1}^M x_{i+12j} - \bar{x}, & 1 \leq i \leq 12, \\ x_{\text{season}}(t_{i-12}), & N \geq i > 12, \end{cases} \quad (3)$$

where  $\bar{x} := \frac{1}{N} \sum_{i=1}^N x_i$ .

## 3 Classical decomposition by LWLR and simple averaging

Using the LWLR procedure and simple averaging, the classical decomposition of a time series can be done. We call this kind of decomposing method the *simple method*. In this paper, the following two commonly used simple methods will be evaluated.

### (a) M-1A method

**Step-1** Determine the trend component. Apply LWLR procedure (1)(2) to the original series to obtain  $x_{\text{trend}}$ , with the diagonal weight matrix  $\mathbf{W}$  in (1) defined as

$$w_i := \mathbb{I}\{t - \rho \leq t_i < t + \rho\} := \begin{cases} 1, & t - \rho \leq t_i < t + \rho, \\ 0, & \text{otherwise,} \end{cases} \quad (4)$$

which means  $x_{\text{trend}}(t)$  is determined by those data near  $t$ . When  $t$  is near boundary  $t_1$  or  $t_N$  so that  $t - \rho < t_1$  or  $t + \rho > t_N$ , less data is available to determine  $x_{\text{trend}}(t)$ . In this case, one can adapt the following *boundary adjustment* strategy: let  $w_i := \mathbb{I}\{t_1 \leq t_i < 2\rho - t_1\}$  or  $w_i := \mathbb{I}\{t_N - 2\rho \leq t_i < t_N\}$  such that the same number of data is used for determining  $x_{\text{trend}}(t)$ .

**Step-2** Determine the seasonal component. Apply simple averaging (3) to the detrended series  $x'_i := x_i - x_{\text{trend}}(t_i)$  to obtain

$$x_{\text{season}}(t_i) := \begin{cases} \frac{1}{M} \sum_{j=1}^M x'_{i+12j} - \bar{x}', & 1 \leq i \leq 12, \\ x_{\text{season}}(t_{i-12}), & N \geq i > 12, \end{cases} \quad (5)$$

where  $\bar{x}' := \frac{1}{N} \sum_{i=1}^N x'_i$ .

**Step-3** Determine the residual component. Simply let  $\mathbf{x}_{\text{residual}}(t_i) := x_i - x_{\text{trend}}(t_i) - \bar{x}' - x_{\text{season}}(t_i)$ .

The procedure above decomposes the original time series as

$$x_i = x_{\text{trend}}^{(M-1A)}(t_i) + \overline{x}^{(M-1A)} + x_{\text{season}}^{(M-1A)}(t_i) + x_{\text{residual}}^{(M-1A)}(t_i). \quad (6)$$

(b) *M-1B method*

M-1B method is like M-1A, but determine the seasonal component first, before detrending.

**Step-1** Determine the seasonal component. Apply simple averaging (3) to the original series  $\{x_i\}_{i=1}^N$  to obtain

$$x_{\text{season}}(t_i) := \begin{cases} \frac{1}{M} \sum_{j=1}^M x_{i+12j} - \bar{x}, & 1 \leq i \leq 12, \\ x_{\text{season}}(t_{i-12}), & N \geq i > 12, \end{cases} \quad (7)$$

where  $\bar{x} := \frac{1}{N} \sum_{i=1}^N x_i$ .

**Step-2** Determine the trend component. Apply LWLR procedure (1)(2) to the *deseasonalized* series  $x'_i := x_i - x_{\text{season}}(t_i)$  to obtain trend, with the diagonal weight matrix  $\mathbf{W}$  in (1) defined as (4). The same near boundary adjustment strategy as in M-1A can be adapted.

**Step-3** Determine the residual component. Simply let  $x_{\text{residual}}(t_i) := x_i - x_{\text{trend}}(t_i) - x_{\text{season}}(t_i)$ .

The M-1B method decomposes the original time series as

$$x_i = x_{\text{trend}}^{(M-1B)}(t_i) + x_{\text{season}}^{(M-1B)}(t_i) + x_{\text{residual}}^{(M-1B)}(t_i). \quad (8)$$

## 2.2.2 Predictor-corrector method

Intuitively, the trend estimator will work better if there's no seasonal component, and the seasonal estimator will work better if there's no trend. This is the motivation of the following *predictor-corrector method*.

(a) *M-2 method*

**Step-1** Initial trend and seasonal estimates

Apply M-1A method to obtain the initial estimate  $x_{\text{trend}}^{(M-1A)}(t_i)$  and  $x_{\text{season}}^{(M-1A)}(t_i)$ .

**Step-2** Revised trend and seasonal estimates

Apply the LWLR procedure (1)(2) on the deseasonalized series  $x_i^{(1)} := x_i - x_{\text{season}}^{(M-1A)}(t_i)$  to obtain a revised estimator of trend  $x'_{\text{trend}}(t_i)$ . Then apply the simple averaging method (3) on the revised detrended series  $x_i^{(2)} := x_i - x'_{\text{trend}}(t_i)$  to obtain a final estimator of seasonal components,  $x_{\text{season}}^{(M-2)}(t_i)$ .

**Step-3** Final trend and residual estimates

Apply the LWLR procedure (1)(2) on the revised deseasonalized series  $x_i^{(3)} := x_i - x_{\text{season}}^{(M-2)}(t_i)$  to obtain the final estimator of trend,  $x_{\text{trend}}^{(M-2)}(t_i)$ , and finally compute the residual components  $x_{\text{residual}}^{(M-2)}(t_i) := x_i - x_{\text{trend}}^{(M-2)}(t_i) - x_{\text{season}}^{(M-2)}(t_i)$ .

The M-2 method decompose the original time series into

$$x_i = x_{\text{trend}}^{(M-2)}(t_i) + x_{\text{season}}^{(M-2)}(t_i) + x_{\text{residual}}^{(M-2)}(t_i). \quad (9)$$

(b) *M-2A method*

M-2A method is like M-2 but determines the seasonal component first, before detrending.

**Step-1** Initial seasonal and seasonal estimates

Apply M-1B method to obtain the initial estimate  $x_{\text{season}}^{(M-1B)}(t_i)$ ,  $x_{\text{trend}}^{(M-1B)}(t_i)$ .

**Step-2** Revised seasonal and trend estimates

Apply the simple averaging method (3) on the detrended series  $x_i^{(1)} := x_i - x_{\text{trend}}^{(M-1B)}(t_i)$  to obtain a revised estimator of seasonal components,  $x'_{\text{season}}(t_i)$ , and then apply the LWLR procedure

(1)(2) on the revised deseasonalized series  $x_i^{(2)} := x_i - x'_{\text{season}}(t_i)$  to obtain a final estimator of trend, denoted as  $x_{\text{trend}}^{(M-2A)}(t_i)$ .

**Step-3** Final seasonal components and residual estimates

Apply the simple averaging method (3) on the revised detrended series  $x_i^{(3)} := x_i - x_{\text{trend}}^{(M-2A)}(t_i)$  to obtain the final estimator of seasonal components,  $x_{\text{season}}^{(M-2A)}(t_i)$  and  $\overline{x_i}^{(M-2A)} := \frac{1}{N} \sum_{i=1}^N x_i'$ .

Finally, compute the residual components  $x_{\text{residual}}^{(M-2A)}(t_i) := x_i - x_{\text{trend}}^{(M-2A)}(t_i) - \overline{x_i}^{(M-2A)} - x_{\text{season}}^{(M-2A)}(t_i)$ .

The M-2A method decompose the original time series into

$$x_i = x_{\text{trend}}^{(M-2A)}(t_i) + \overline{x_i}^{(M-2A)} + x_{\text{season}}^{(M-2A)}(t_i) + x_{\text{residual}}^{(M-2A)}(t_i). \quad (10)$$

(c) M-2S method

The M-2S method is like M-2 but uses *the spectral method* (Narapusetty et al., 2009) to determine the seasonal component. The spectral method is to fit several sines and cosines

$$x_{\text{SM}}(t) := \hat{a}_0 + \sum_{k=1}^H \hat{a}_j \cos(k\omega_0 t) + \hat{b}_j \sin(k\omega_0 t)$$

to the raw series  $\{x(t_i)\}_{i=1}^N$ , where  $\omega_0 := 2\pi/P$ ,  $P$  is the annual period. If  $t$  is in units of months, i.e.,  $t = 0, \dots, 12$  corresponds to month 0, ..., 12, then  $P$  should be taken as 12.

Let  $\mathbf{M}$  be the data matrix,  $\mathbf{x}$  be the response vector, and  $\hat{\beta}$  be the parameters vector, which are defined as

$$\mathbf{M} := \begin{pmatrix} 1 & \cos(1\omega_0 t_1) & \cdots & \cos(H\omega_0 t_1) & \sin(1\omega_0 t_1) & \cdots & \sin(H\omega_0 t_1) \\ 1 & \cos(1\omega_0 t_2) & \cdots & \cos(H\omega_0 t_2) & \sin(1\omega_0 t_2) & \cdots & \sin(H\omega_0 t_2) \\ \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ 1 & \cos(1\omega_0 t_N) & \cdots & \cos(H\omega_0 t_N) & \sin(1\omega_0 t_N) & \cdots & \sin(H\omega_0 t_N) \end{pmatrix},$$

$$\mathbf{x} := (x(t_1), \dots, x(t_N))^T, \quad \hat{\beta} := (\hat{a}_0, \dots, \hat{a}_N, \hat{b}_1, \dots, \hat{b}_N)^T. \quad (11)$$

The least-square solution of  $\hat{\beta}$  is

$$\hat{\beta} := \underset{\beta \in \mathbb{R}^2}{\text{argmin}} \{(\mathbf{x} - \mathbf{M}\beta)^T(\mathbf{x} - \mathbf{M}\beta)\} = (\mathbf{M}^T \mathbf{M})^{-1} \mathbf{M}^T \mathbf{x}, \quad (12)$$

and the seasonal series is defined as

$$x_{\text{SM}}(t_i) := \sum_{k=1}^H \hat{a}_j \cos(k\omega_0 t_i) + \hat{b}_j \sin(k\omega_0 t_i). \quad (13)$$

Let  $\mathbf{x}_{\text{SM}} := (x_{\text{SM}}(t_1), \dots, x_{\text{SM}}(t_N))^T$ , the seasonal series (13) can be written as

$$\mathbf{x}_{\text{SM}} := \mathbf{M}\hat{\beta} - \hat{a}_0 \mathbf{u}, \quad (13)'$$

where  $\mathbf{u} := (1, \dots, 1)^T \in \mathbb{R}^N$ . Since (13) contains no trend, the original series should be detrended first.

The procedure of the M-2P method is as follows.

**Step-1** Initial trend and seasonal estimates

Apply the LWLR procedure (1)(2) on the original series to obtain the initial trend estimate  $x_{\text{trend}}^{(1)}(t_i)$ , and then apply the spectral method (12)(13) on the detrended series  $x_i^{(1)} := x_i - x_{\text{trend}}^{(1)}(t_i)$  to obtain the initial seasonal series  $x_{\text{SM}}^{(1)}(t_i)$ .

**Step-2** Revised trend and seasonal estimates

Apply the LWLR procedure (1)(2) on the deseasonalized series  $x_i^{(2)} := x_i - x_{\text{SM}}^{(1)}(t_i)$  to obtain a revised estimator of trend,  $x_{\text{trend}}^{(2)}(t_i)$ . Then apply the spectral method (12)(13) on the revised

detrended series  $x_i^{(3)} := x_i - x_{\text{trend}}^{(2)}(t_i)$  to obtain a final estimator of seasonal components,  $x_{\text{season}}^{(M-2S)}(t_i)$ .

### Step-3 Final trend and residual estimates

Apply the LWLR procedure (1)(2) on the revised deseasonalized series  $x_i^{(4)} := x_i - x_{\text{season}}^{(M-2S)}(t_i)$  to obtain the final estimator of trend,  $x_{\text{trend}}^{(M-2S)}(t_i)$ , and finally compute the residual components  $x_{\text{residual}}^{(M-2S)}(t_i) := x_i - x_{\text{trend}}^{(M-2S)}(t_i) - x_{\text{season}}^{(M-2S)}(t_i)$ .

The M-2S method decompose the original time series into

$$x_i = x_{\text{trend}}^{(M-2S)}(t_i) + x_{\text{season}}^{(M-2S)}(t_i) + x_{\text{residual}}^{(M-2S)}(t_i). \quad (14)$$

## 2.2.3 Global model-fitting method

Another classical decomposition method, which we call the *global model-fitting method*, assumes that the original series is composed of trend and seasonal components with a pre-assumed form, and a residual term. Typically, the trend is assumed to be a linear or quadratic polynomial, and the seasonal component is assumed to be the sum of a few sine and cosine signals.

### (a) M-3L method

M-3L method fits a model of the form

$$\hat{x}(t) := \hat{c}_0 + \hat{c}_1 t + \sum_{k=1}^H \hat{a}_k \cos(k\omega_0 t) + \hat{b}_k \sin(k\omega_0 t),$$

to the original time series, where  $\omega_0 := 2\pi/P$ ,  $P$  is the annual period. If  $t$  is in units of months, i.e.,  $t = 0, \dots, 12$  corresponds to month 0, ..., 12, then  $P$  should be taken as 12.

Let  $\mathbf{M}$  be the data matrix,  $\mathbf{x}$  be the response vector, and  $\hat{\boldsymbol{\beta}}$  be the parameters vector, which are defined as

$$\begin{aligned} \mathbf{M} &:= (\mathbf{M}_{\text{trend}}, \mathbf{M}_{\text{season}}), \\ \mathbf{M}_{\text{trend}} &:= \begin{pmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_N \end{pmatrix}, \\ \mathbf{M}_{\text{season}} &:= \begin{pmatrix} \cos(1\omega_0 t_1) & \cdots & \cos(H\omega_0 t_1) & \sin(1\omega_0 t_1) & \cdots & \sin(H\omega_0 t_1) \\ \cos(1\omega_0 t_2) & \cdots & \cos(H\omega_0 t_2) & \sin(1\omega_0 t_2) & \cdots & \sin(H\omega_0 t_2) \\ \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ \cos(1\omega_0 t_N) & \cdots & \cos(H\omega_0 t_N) & \sin(1\omega_0 t_N) & \cdots & \sin(H\omega_0 t_N) \end{pmatrix}, \\ \mathbf{x} &:= (x(t_1), \dots, x(t_N))^T, \quad \hat{\boldsymbol{\beta}} := (\hat{\boldsymbol{\beta}}_{\text{trend}}^T, \hat{\boldsymbol{\beta}}_{\text{season}}^T)^T, \\ \hat{\boldsymbol{\beta}}_{\text{trend}} &:= (\hat{c}_0, \hat{c}_1)^T, \quad \hat{\boldsymbol{\beta}}_{\text{season}} := (\hat{a}_1, \dots, \hat{a}_H, \hat{b}_1, \dots, \hat{b}_H)^T. \end{aligned} \quad (15)$$

The least-square solution of  $\hat{\boldsymbol{\beta}}$  is

$$\hat{\boldsymbol{\beta}} := \underset{\boldsymbol{\beta} \in \mathbb{R}^2}{\text{argmin}} \{ (\mathbf{x} - \mathbf{M}\boldsymbol{\beta})^T (\mathbf{x} - \mathbf{M}\boldsymbol{\beta}) \} = (\mathbf{M}^T \mathbf{M})^{-1} \mathbf{M}^T \mathbf{x}. \quad (16)$$

The trend series  $\{x_{\text{trend}}(t_i)\}_{i=1}^N$  and the seasonal series  $\{x_{\text{SM}}(t_i)\}_{i=1}^N$  is defined to be

$$x_{\text{trend}}(t_i) := \hat{c}_0 + \hat{c}_1 t_i, \quad (17a)$$

$$x_{\text{SM}}(t_i) := \sum_{k=1}^H \hat{a}_k \cos(k\omega_0 t_i) + \hat{b}_k \sin(k\omega_0 t_i). \quad (17b)$$

Let  $\mathbf{x}_{\text{trend}} := (x_{\text{trend}}(t_1), \dots, x_{\text{trend}}(t_N))^T$ ,  $\mathbf{x}_{\text{SM}} := (x_{\text{SM}}(t_1), \dots, x_{\text{SM}}(t_N))^T$ , the trend and seasonal series (17) can be written as

$$\mathbf{x}_{\text{trend}} := \mathbf{M}_{\text{trend}} \hat{\boldsymbol{\beta}}_{\text{trend}}, \quad \mathbf{x}_{\text{SM}} := \mathbf{M}_{\text{season}} \hat{\boldsymbol{\beta}}_{\text{season}}. \quad (17)'$$

### (b) M-3Q method

The M-3Q method is like M-3L method but assumes a quadratic trend.

M-3Q method fits a model of the form

$$\hat{x}(t) := \hat{c}_0 + \hat{c}_1 t + \hat{c}_2 t^2 + \sum_{k=1}^H \hat{a}_k \cos(k\omega_0 t) + \hat{b}_k \sin(k\omega_0 t),$$

to the original time series, where  $\omega_0 := 2\pi/P$ ,  $P$  is the annual period. If  $t$  is in units of months, i.e.,  $t = 0, \dots, 12$  denotes the month  $0, \dots, 12$ , then  $P$  should be taken as 12.

Let  $\mathbf{M}$  be the data matrix,  $\mathbf{x}$  be the response vector, and  $\hat{\boldsymbol{\beta}}$  be the parameters vector, which are defined as

$$\begin{aligned} \mathbf{M} &:= (\mathbf{M}_{\text{trend}}, \mathbf{M}_{\text{season}}), \\ \mathbf{M}_{\text{trend}} &:= \begin{pmatrix} 1 & t_1 & t_1^2 \\ \vdots & \vdots & \vdots \\ 1 & t_N & t_N^2 \end{pmatrix}, \\ \mathbf{M}_{\text{season}} &:= \begin{pmatrix} \cos(1\omega_0 t_1) & \cdots & \cos(H\omega_0 t_1) & \sin(1\omega_0 t_1) & \cdots & \sin(H\omega_0 t_1) \\ \cos(1\omega_0 t_2) & \cdots & \cos(H\omega_0 t_2) & \sin(1\omega_0 t_2) & \cdots & \sin(H\omega_0 t_2) \\ \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ \cos(1\omega_0 t_N) & \cdots & \cos(H\omega_0 t_N) & \sin(1\omega_0 t_N) & \cdots & \sin(H\omega_0 t_N) \end{pmatrix}, \\ \mathbf{x} &:= (x(t_1), \dots, x(t_N))^T, \quad \hat{\boldsymbol{\beta}} := (\hat{\boldsymbol{\beta}}_{\text{trend}}^T, \hat{\boldsymbol{\beta}}_{\text{season}}^T)^T, \\ \hat{\boldsymbol{\beta}}_{\text{trend}} &:= (\hat{c}_0, \hat{c}_1, \hat{c}_2)^T, \quad \hat{\boldsymbol{\beta}}_{\text{season}} := (\hat{a}_0, \dots, \hat{a}_N, \hat{b}_1, \dots, \hat{b}_N)^T. \end{aligned} \quad (18)$$

The least-square solution of  $\hat{\boldsymbol{\beta}}$  is

$$\hat{\boldsymbol{\beta}} := \underset{\boldsymbol{\beta} \in \mathbb{R}^2}{\operatorname{argmin}} \{(\mathbf{x} - \mathbf{M}\boldsymbol{\beta})^T (\mathbf{x} - \mathbf{M}\boldsymbol{\beta})\} = (\mathbf{M}^T \mathbf{M})^{-1} \mathbf{M}^T \mathbf{x}. \quad (19)$$

The trend series  $\{x_{\text{trend}}(t_i)\}_{i=1}^N$  and the seasonal series  $\{x_{\text{SM}}(t_i)\}_{i=1}^N$  is defined to be

$$x_{\text{trend}}(t_i) := \hat{c}_0 + \hat{c}_1 t_i + \hat{c}_2 t_i^2, \quad (20a)$$

$$x_{\text{SM}}(t_i) := \sum_{k=1}^H \hat{a}_k \cos(k\omega_0 t_i) + \hat{b}_k \sin(k\omega_0 t_i). \quad (20b)$$

Let  $\mathbf{x}_{\text{trend}} := (x_{\text{trend}}(t_1), \dots, x_{\text{trend}}(t_N))^T$ ,  $\mathbf{x}_{\text{SM}} := (x_{\text{SM}}(t_1), \dots, x_{\text{SM}}(t_N))^T$ , the trend and seasonal series (20) can be written as

$$\mathbf{x}_{\text{trend}} := \mathbf{M}_{\text{trend}} \hat{\boldsymbol{\beta}}_{\text{trend}}, \quad \mathbf{x}_{\text{SM}} := \mathbf{M}_{\text{season}} \hat{\boldsymbol{\beta}}_{\text{season}}. \quad (20)'$$

## 2.3 Experiment setup

### 2.3.1 Idealized time series

理想试验，共设置三组：

1. 第 1 组（理想仿真），对多项式 trend 的次数和余项的敏感性。

(1) 线性/二次 trend + 1 年周期信号 + 噪声。

(2) 线性/二次 trend + 1 年周期信号 + 4.3 年周期信号。

(3) 线性/二次 trend + 1 年周期信号 + 4.3 年周期信号 + 噪声。

2. 第 2 组，对 trend 的增量的敏感性。

(3) 同上。

(4) 剧烈的线性/二次 trend + 1 年周期信号 + 4.3 年周期信号 + 噪声。

3. 第 3 组，对 seasonal cycle 的解释方差的敏感性。

(3) 同上。

(5) 线性/二次 trend + 小幅度的 1 年周期信号 + 4.3 年周期信号 + 噪声。



表 1 理想试验各算例中, 理想余项的样本方差.

Case Name:	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
StdVar.Residual (1e-1 °C)	7.99	7.99	14.2	14.2	16.0	16.0	16.0	16.0	16.0	16.0

### 2.3.2 Real time-indices

Global SST (Monthly), ....

## 2.4 Assessment methodology

### 2.4.1 Idealized time series

对于按经典分解方式构造的理想时间信号, “真实”分量已知, 故可将各方法的输出与“真实”误差比较.

指标 1: 输出的余项 (或, 气候平均态) 对 “真实” 余项 (或, 气候平均态) 的样本相关系数(sample Pearson correlation coefficient, CC)和样本均方误差(sample root mean square error, RMSE).

指标 2: 输出的气候平均态对 “真实” 原始序列的样本相关系数和样本均方误差.

指标 3: 气候平均态的交叉验证误差(cross-validated mean squared error, CVE). CVE 定义为: 将原始序列中的第  $i$  年数据作为测试集, 其余作为训练集, 求输出的测试集上气候平均态对真实气候平均态的平方误差, 对所有  $i$  重复执行, 对  $i$  的平均值.

### 2.4.2 Real time-indices

对于实际时间指数, “真实”分量未知. 根据气候平均态的定义, 气候平均态应最小化总体均方误差, 故合理期待方法的输出应使得样本均方误差较小.

指标 4: 输出的气候平均态对原时间序列的样本相关系数和样本均方误差.

指标 5: 输出的气候平均态对原时间序列的交叉验证误差 CVE. CVE 的定义类比指标 3.



### 3 Results

#### 3.1 Idealized experiments

##### 3.1.1 Classical decomposition results

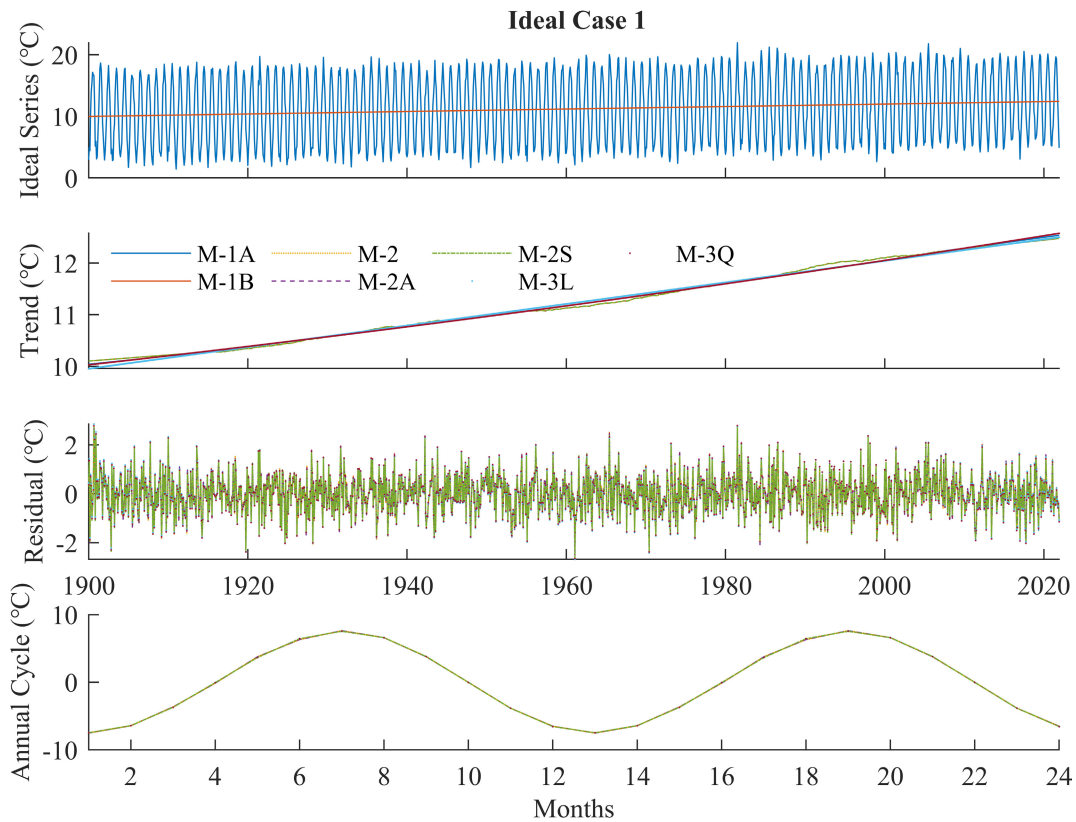


图 1 理想试验 Case 1. 理想时间序列(a)由线性趋势、以 12 个月为周期的 seasonal components 和 Gaussian noise 组成,(a)中红线展示其 trend 分量. 用三类方法分解出的 trend 分量示于(b), Residual 分量示于(c), seasonal components 示于(d).

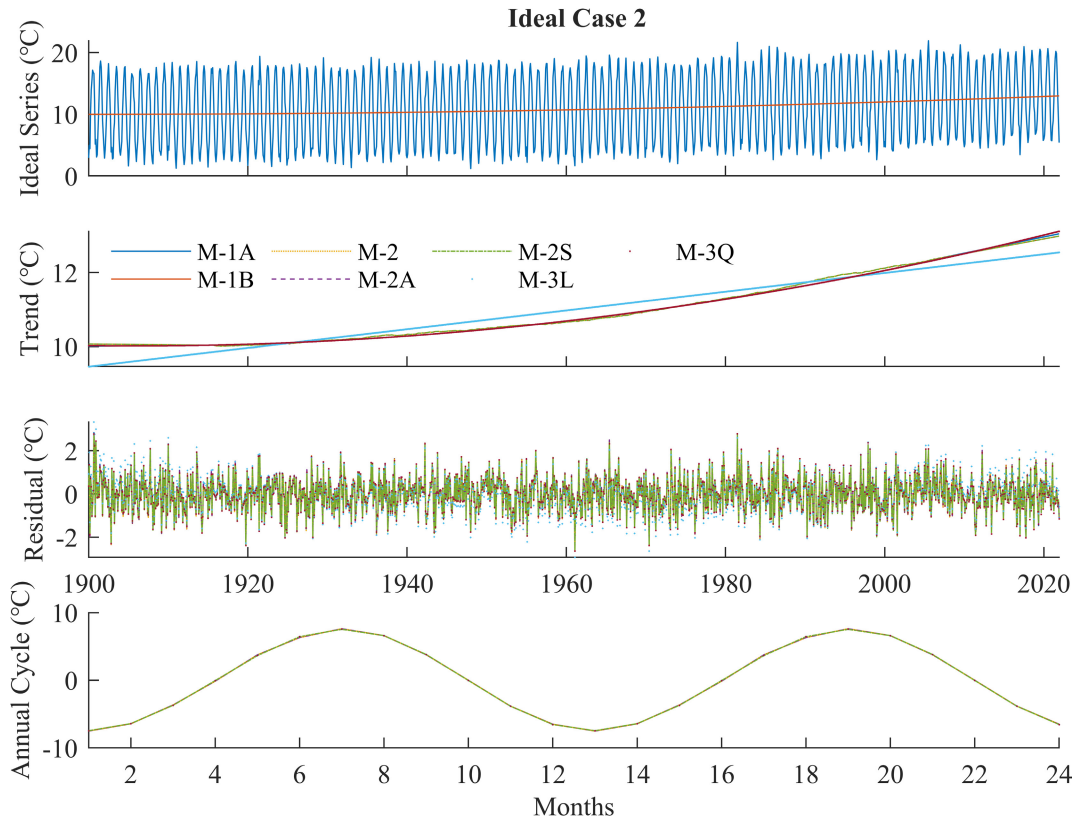


图 2 理想试验 Case 2. 理想时间序列(a)由二次多项式趋势、以 12 个月为周期的 seasonal components 和 Gaussian noise 组成, (a)中红线展示其 trend 分量. 用三类方法分解出的 trend 分量示于(b), residual 分量示于(c), seasonal components 示于(d).

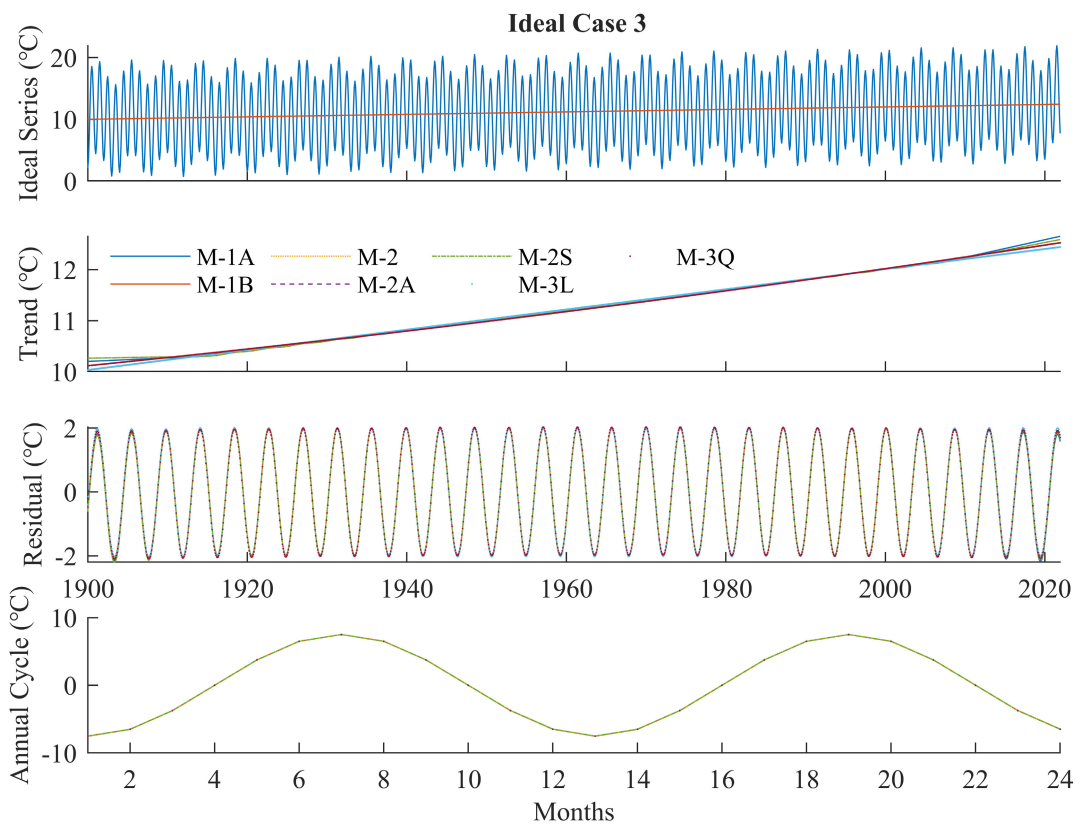


图 3 理想试验 Case 3. 理想时间序列(a)由线性趋势、以 12 个月为周期的 seasonal components 和



以 4.3 年为周期的 inter-annual 变率组成, (a) 中红线展示其 trend 分量. 用三类方法分解出的 trend 分量示于(b), residual 分量示于(c), seasonal components 示于(d).

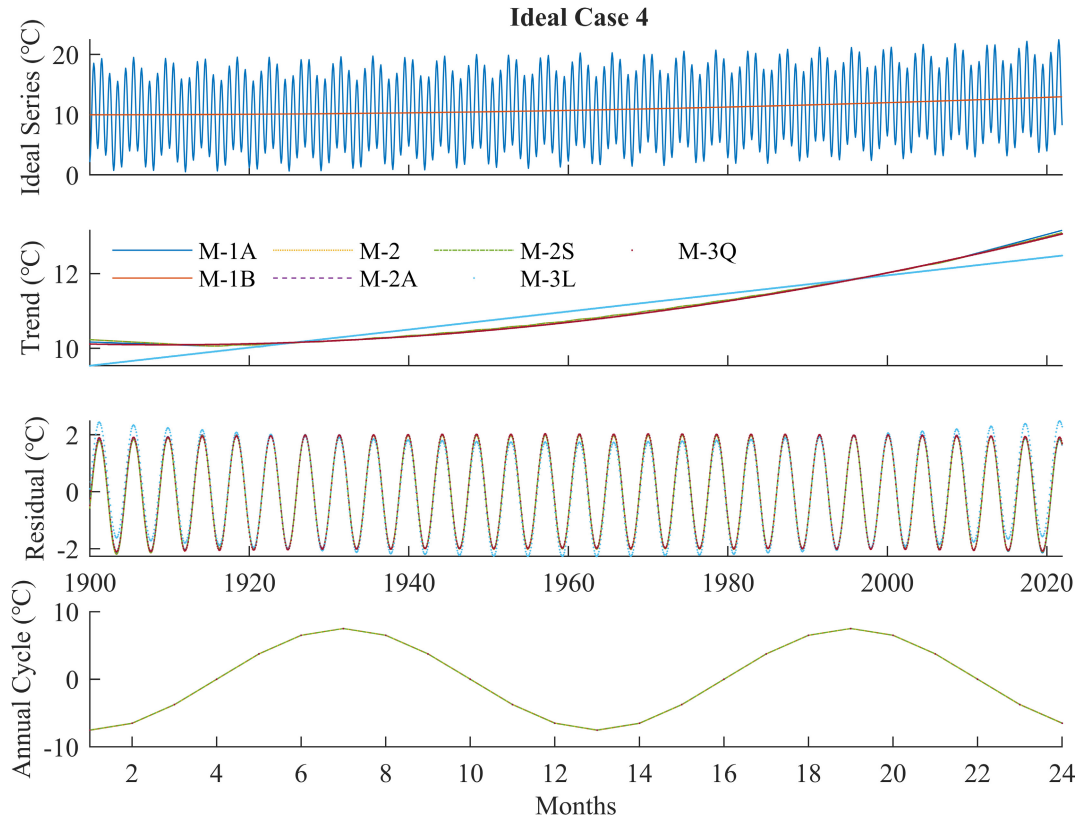


图 4 理想试验 Case 4. 理想时间序列(a)由二次多项式趋势、以 12 个月为周期的 seasonal components 和以 4.3 年为周期的 inter-annual 变率组成, (a) 中红线展示其 trend 分量. 用三类方法分解出的 trend 分量示于(b), residual 分量示于(c), seasonal components 示于(d).

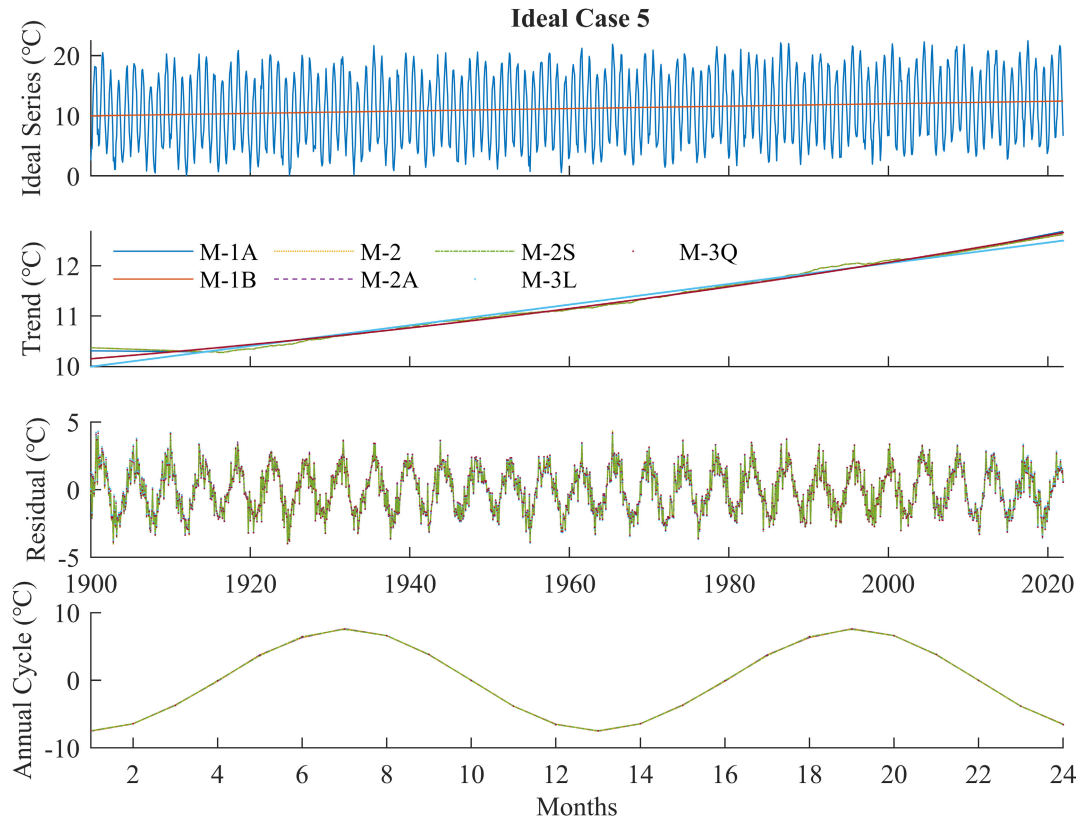


图 5 理想试验 Case 5. 理想时间序列(a)由线性趋势、以 12 个月为周期的 seasonal components、以 4.3 年为周期的 inter-annual 变率和 Gaussian noise 组成, (a)中红线展示其 trend 分量. 用三类方法分解出的 trend 分量示于(b), residual 分量示于(c), seasonal components 示于(d).

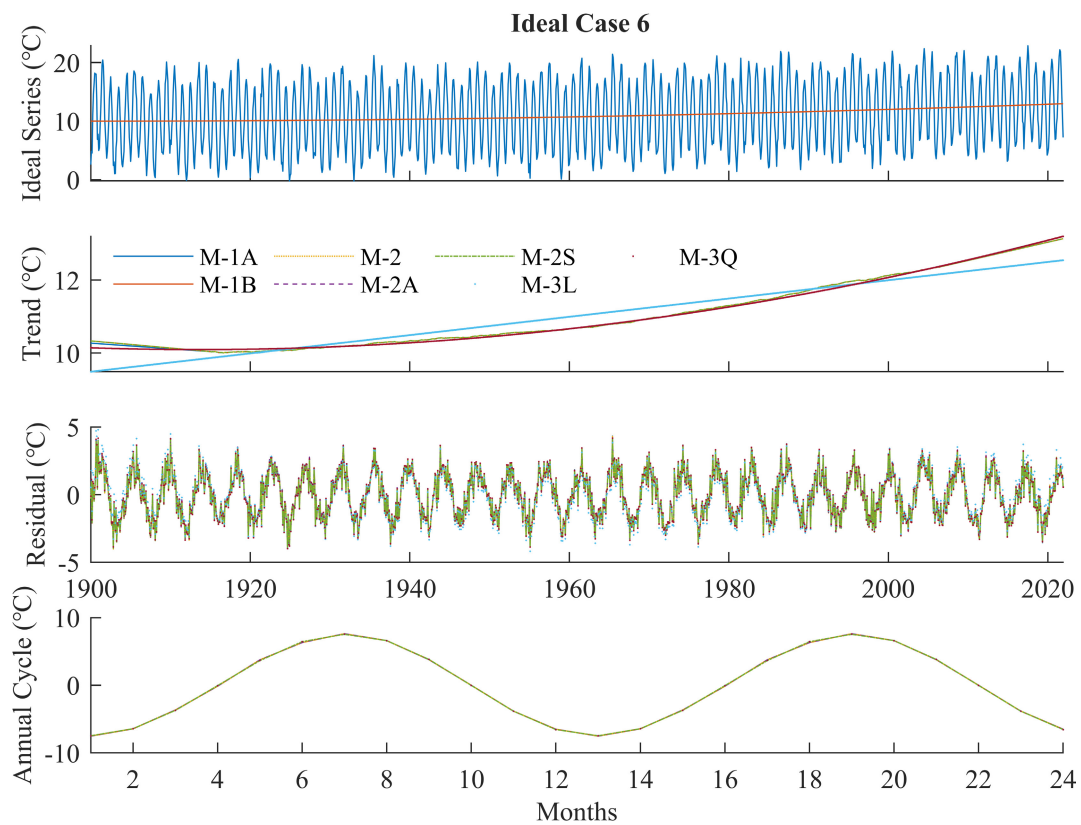


图 6 理想试验 Case 6. 理想时间序列(a)由二次多项式趋势、以 12 个月为周期的 seasonal



components、以 4.3 年为周期的 inter-annual 变率和 Gaussian noise 组成, (a)中红线展示其 trend 分量. 用三类方法分解出的 trend 分量示于(b), residual 分量示于(c), seasonal components 示于(d).

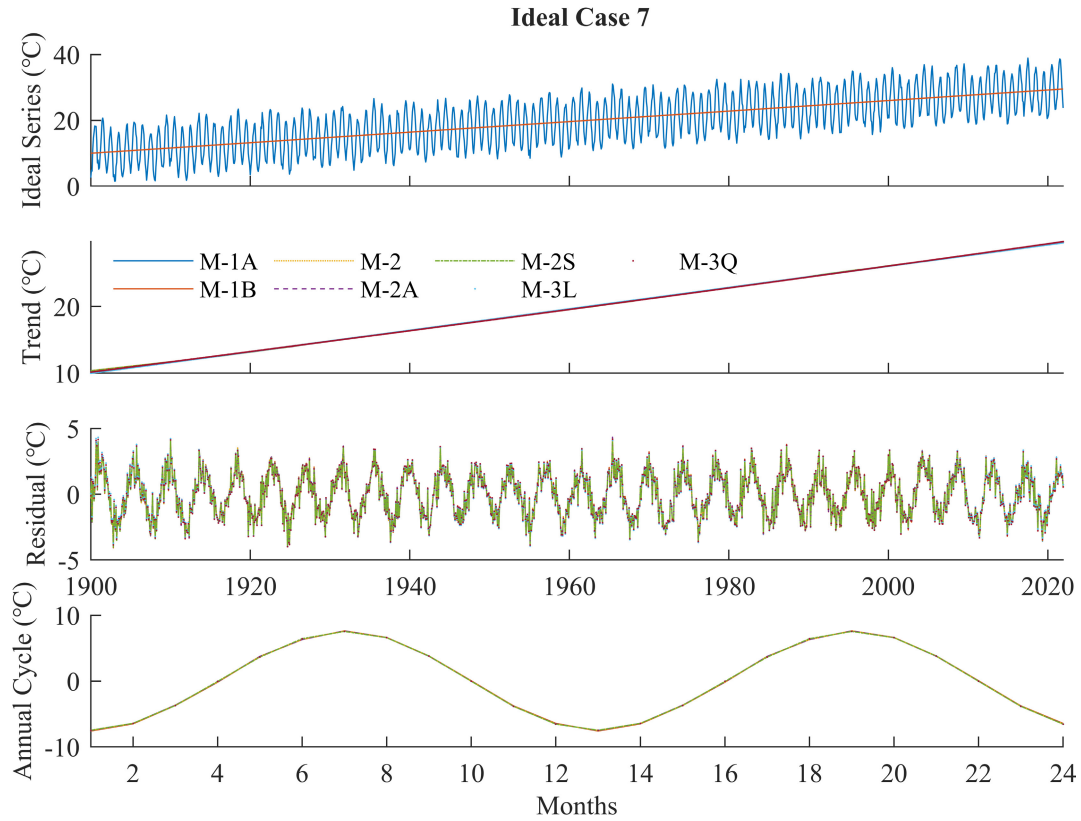


图 7 理想试验 Case 7. 理想时间序列(a)由剧烈的线性趋势、以 12 个月为周期的 seasonal components、以 4.3 年为周期的 inter-annual 变率和 Gaussian noise 组成, (a)中红线展示其 trend 分量. 用三类方法分解出的 trend 分量示于(b), residual 分量示于(c), seasonal components 示于(d).



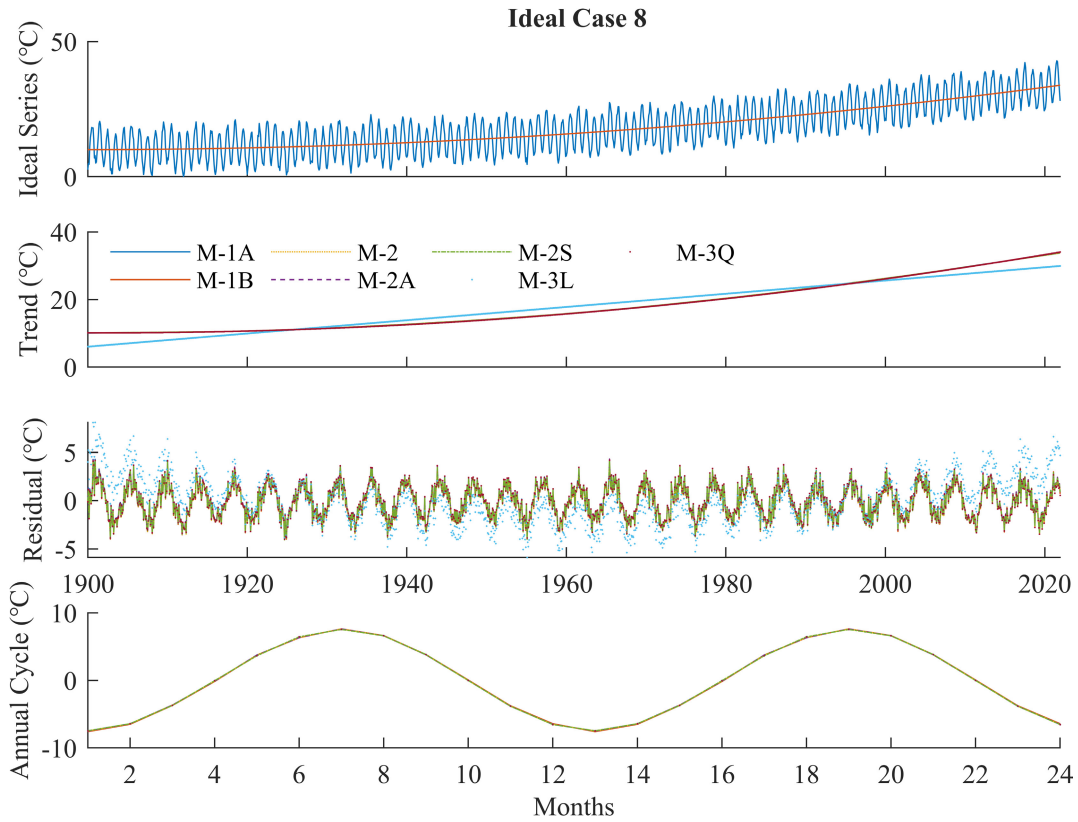


图8 理想试验 Case 8. 理想时间序列(a)由剧烈的二次多项式趋势、以 12 个月为周期的 seasonal components、以 4.3 年为周期的 inter-annual 变率和 Gaussian noise 组成, (a)中红线展示其 trend 分量. 用三类方法分解出的 trend 分量示于(b), residual 分量示于(c), seasonal components 示于(d).

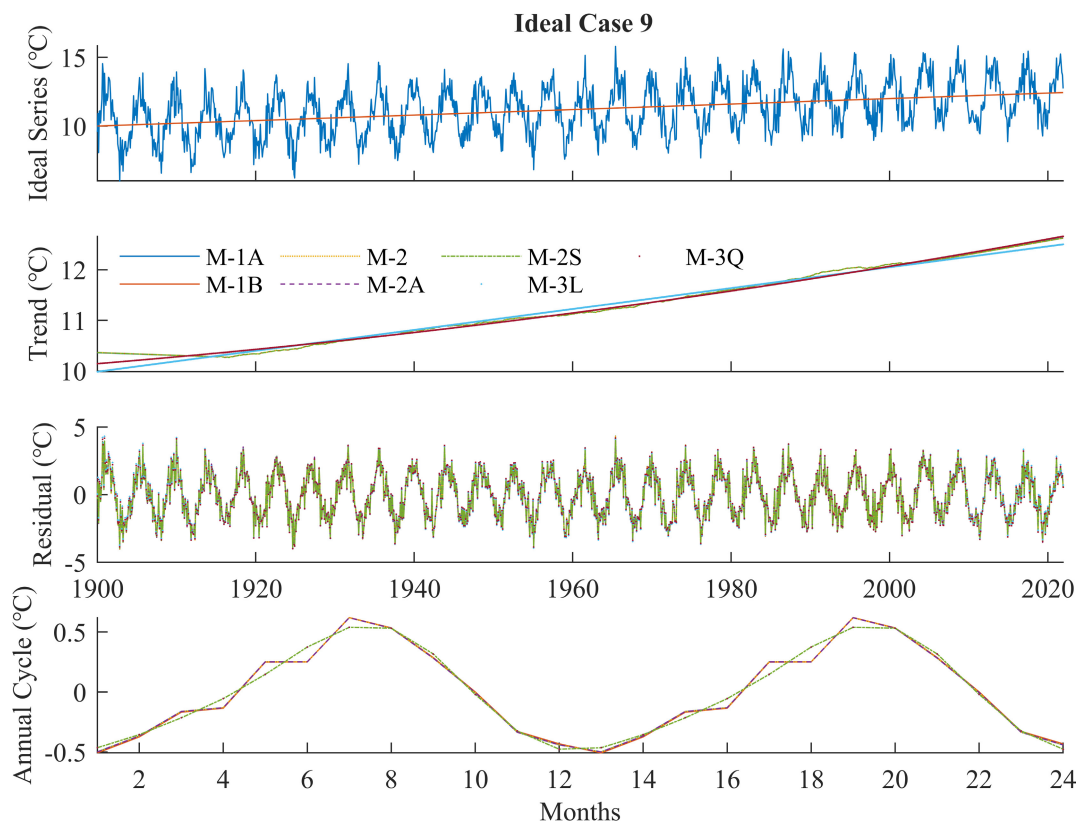


图9 理想试验 Case 9. 理想时间序列(a)由线性趋势、以 12 个月为周期的微弱 seasonal components、



以 4.3 年为周期的 inter-annual 变率和 Gaussian noise 组成, (a)中红线展示其 trend 分量. 用三类方法分解出的 trend 分量示于(b), residual 分量示于(c), seasonal components 示于(d).

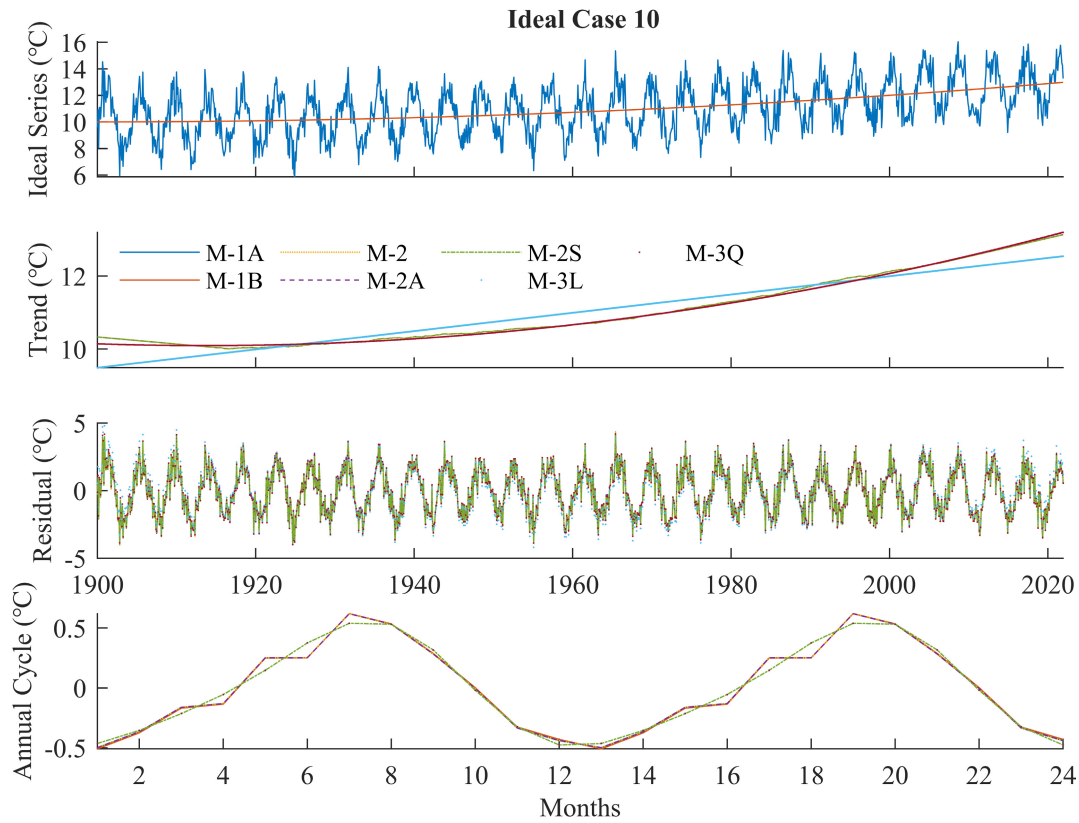


图 10 理想试验 Case 10. 理想时间序列(a)由二次多项式趋势、以 12 个月为周期的微弱 seasonal components、以 4.3 年为周期的 inter-annual 变率和 Gaussian noise 组成, (a)中红线展示其 trend 分量. 用三类方法分解出的 trend 分量示于(b), residual 分量示于(c), seasonal components 示于(d).

### 3.1.2 Comparison of different methods

M-2S 方法在有噪声 (对应算例 C1, 2, 5, 6, 7, 8, 9, 10) 时, 在特定指标 (RMSE, CVE) 上略优于其他 M-1, M-2 方法, 但不及当全局模型选择正确时的 M-3 系列方法 (图 11, 图 12, 图 13, 图 14) .

M-1A 方法在长期趋势剧烈 (对应算例 C7, C8) 或年际变率理想 (对应算例 C3, C4, 其中的年际变率信号为理想的 4.3 年周期单音信号) 时, 在特定指标上略优于 M-1B, 但不及 M-2S (图 11, 图 12, 图 13, 图 14) .

M-2 与 M-2A 的差异. 在理想试验中, M-2 和 M-2A 输出的季节分量差异小于  $10^{-6}^{\circ}\text{C}$  (表 2) . 输出余项的差异在各理想算例中分别为常数 (表 3) , 且当长期趋势的非线性强烈时, 余项差异较大. 在强非线性算例 C8 中, 这个差异超过  $0.1^{\circ}\text{C}$ ; 在其余算例中, 这个差异为  $10^{-2}^{\circ}\text{C}$  量级. 根据定义, M-2A 的输出余项均值被保证为零, 而 M-2 不是这样. 另从理想试验 Case 5 (图 16) 和 Case 6 (图 17) 的结果可见, M-2 和 M-2A 的输出 trend 的差异很小 ( $\sim 10^{-7}^{\circ}\text{C}$ ) ; 在时间序列的首、尾 15 年, 越靠近序列的两端点, M-2 和 M-2A 方法输出的 trend 与理想 trend 的差异就相对越大 (仍很小), 这是由 LWLR 的边缘效应造成的. 这些结果提示, **M-2 与 M-2A 的输出差异主要由余项的非零均值造成; 当理想趋势非线性较强、理想余项样本方差较大时, M-2 的输出余项的样本均值可能离零较远, 可造成对 M-2A 输出余项的较大差异.**



M-2 对 M-1A, M-1B 的差异. 1) 季节信号的差异. M-2 对 M-1A 的输出季节信号差异(表 4) 很小, 当理想季节信号为 $\sim 10^1$  ( $10^0$ )  $^{\circ}\text{C}$  量级时, 差异小于  $10^{-3}$  ( $10^{-4}$ )  $^{\circ}\text{C}$ . M-2 对 M-1B 的输出季节信号差异(表 7) 主要随长期趋势剧烈程度的增大而增加. 当百年增温为  $2^{\circ}\text{C}$  时(除 C7, 8 外的其他算例), 这个差异约为  $0.01^{\circ}\text{C}$ ; 当百年增温达  $20^{\circ}\text{C}$  时(算例 C7, 8), 这个差异可接近  $0.1^{\circ}\text{C}$ . 这个差异还随着长期趋势的非线性的增强而增大(非主导). 在百年增温都设定为约  $2^{\circ}\text{C}$  的前提下, 这个差异在每个非线性趋势算例(C2, 4, 6, 8, 10) 中都比相应的线性趋势算例(C1, 3, 5, 7, 9) 中的稍大(表 7), 但这可能与非线性算例的序列总增温(122 年) 稍大于相应线性算例有关. 造成 M-2 对 M-1B 的输出季节信号差异的主要原因, 是长期趋势对由简单平均估计的季节信号的污染. 例如, 非常剧烈的长期升温趋势可导致在一年内年末的温度倾向于比年初高; 每年都如此, 就会造成由简单平均决定的季节信号出现年末高于年初的倾向. 是否可以消除这种污染? 若对季节分量的简单平均估计是对已去除长期趋势的序列执行的, 如 M-1A 和 M-2 方法那样, 则上述原因将失效; 这便是 M-2 对 M-1A 的输出季节信号差异远小于对 M-1B 的原因. 2) 余项均值的差异. 在定义上, M-1A 的输出余项总是零均值的, 即使理想余项不是; 而 M-1B 和 M-2 不具有此性质(图 15). 在理想试验中, M-2 对 M-1B 的输出余项均值差异很小(表 8). M-2 对 M-1A 的输出余项均值差异(表 5) 主要随理想趋势非线性的增强而增大. 在强烈非线性趋势算例 C8 中, 这个差异超过  $0.1^{\circ}\text{C}$ . 注意 C5, C7, C9 的输出余项均值差异相同(表 5), 说明理想趋势的剧烈程度不影响这个差异, C8 中这差异的升高应归因于非线性而不是剧烈程度. 这个差异还随着理想余项方差的增加而增大. 在具有线性理想趋势的算例 C1, 3, 5, 7, 9 中, 当理想余项方差相同时, M-1A 和 M-2 输出余项的差异亦相同; 理想余项方差愈大, 则这个差异越大; 在具有相同二次多项式趋势的算例 C2, 4, 6, 10 中, 仍有前述观察; 在具有剧烈二次多项式趋势的算例 C8 中, 这个差异最大, 超过  $0.1^{\circ}\text{C}$ , 这是因为 C8 算例的趋势非线性程度和理想余项方差都是理想试验各算例中最高的. 3) 输出长期趋势的差异. M-2 对 M-1A 的输出 trend 差异(表 6) 主要随理想季节信号幅度的增加而增大, 而与理想趋势几乎无关. 在具有相同理想季节信号的理想算例 C1-8 中, 这个差异很接近, 约为  $0.06^{\circ}\text{C}$ ; 在理想季节信号幅度很小的算例 C9-10 中, 这个差异不到  $4 \times 10^{-3}^{\circ}\text{C}$ . 为进一步探究导致 M-2 对 M-1A 输出 trend 差异的原因, 我们仔细观察使得这个差异最大的理想算例 C3(图 19). 可见, 这个差异主要出现在时间序列的首、尾 15 年处, 且越接近序列两端点则差异越大. 注意到 LWLR 区间被选取为 30 年; 在首、尾 15 年(称为边缘区)内, 各点的 trend 值都是由以距离首、尾端点 15 年时刻为中心的 30 年区间上各点通过线性拟合决定的; 在边缘区外, 每点的 trend 值是由以该点为中心的 30 年区间上各点通过线性拟合决定的. 还注意到 M-1A 和 M-2 的 trend 估计法的主要区别, 是前者的 LWLR 是对带有季节信号的时间序列执行的, 而后者的是对已去除季节信号估计的时间序列执行的; 于是, 季节信号的存在与否将对 trend 的估计产生影响, 即使季节信号是长期零均值的. 在边缘区外的各点, LWLR 区间中心正位于当前点, 故这个影响很小; 而在边缘区内, 这个影响将随着对序列端点距离的减少而增大. 因此, 季节信号对线性回归的影响及其被 LWLR 的边缘效应放大, 是造成 M-2 对 M-1A 输出 trend 差异的主要原因, 而这原因在边缘区外作用很小. M-2 对 M-1B 的输出 trend 差异(表 9) 较小, 主要随理想趋势的剧烈程度增大而增加. 当理想百年增温约为  $2$  ( $20$ )  $^{\circ}\text{C}$  时, 这个差异不到  $4 \times 10^{-4}$  ( $4 \times 10^{-3}$ )  $^{\circ}\text{C}$ . 这个差异远比 M-2 对 M-1B 的输出 trend 差异要小, 是因为 M-1B 的 trend 和 M-2 一样也是从已去除季节信号的序列中提取的, 失效了导致 M-2 对 M-1A 输出 trend 差异的主要原因. 4) 输出(余项)的综合差异. 根据前面的分析, 影响 M-2 对 M-1A 和 M-1B 的输出(余项)差异的因素包括: a) 剧烈的理想趋势, 通过污染 M-1B 的季节信号估计, 利于 M-2 的输出季节信号向 M-1A 接近; b) 强非线性的理想趋势, 通过污染 M-1B 的季节信号估计, 利于 M-2 的输出季节信号向 M-1A 接近(非主导); 但利于输出余项均值向





M-1B 接近; c) 大的理想余项方差, 利于 M-2 的输出余项均值向 M-1B 接近; d) 大幅度的季节信号, 通过影响线性拟合及被 LWLR 边缘效应放大, 利于 M-2 的输出 trend 向 M-1B 接近 (仅边缘区)。这四个因素相互制衡的结果, 决定了 M-2 是更接近 M-1A 还是 M-1B。

M-2A 对 M-1A, M-1B 的差异. 根据前面对 M-2 与 M-2A 的差异分析, 和对 M-2 对 M-1A, M-1B 的差异分析, 得出影响 M-2A 对 M-1A 和 M-1B 的输出 (余项) 差异的因素包括: a) 剧烈的理想趋势, 通过污染 M-1B 的季节信号估计, 利于 M-2A 的输出季节信号向 M-1A 接近; b) 强非线性的理想趋势, 通过污染 M-1B 的季节信号估计, 利于 M-2A 的输出季节信号向 M-1A 接近 (非主导); 也利于输出余项均值向 M-1A 接近; c) 大的理想余项方差, 利于 M-2A 的输出余项均值向 M-1A 接近; d) 大幅度的季节信号, 通过影响线性拟合及被 LWLR 边缘效应放大, 利于 M-2A 的输出 trend 向 M-1B 接近 (仅边缘区)。在边缘区外, 因素 d 作用很小, 因素 a, b, c 都利于 M-2A 向 M-1A 接近; 在边缘区内, 因素 d 的作用方向与因素 a, b, c 相反, 它们的制衡结果决定了在边缘区内 M-2A 是更接近 M-1A 还是 M-1B。

M-2, 2A 与 M-1A, 1B 的差异. 以理想算例 C8 (图 8) 为例, 该算例具有剧烈的强非线性理想趋势. 现在应用上两段的结论来分析这算例. M-2 对 M-2A 的差异 (图 18) 主要表现在余项均值上, 差异超过  $0.1^{\circ}\text{C}$ , 这是由理想趋势的强非线性导致的. M-2 对 M-1A 的差异 (图 20) 主要表现在输出 trend 边缘区内, 以及输出余项均值上, 而输出季节信号差异很小. 这是由理想趋势的强非线性对余项均值的影响, 以及季节信号对边缘区 LWLR 的影响所致. 因为 M-1A 和 M-2 在决定季节信号时, 长期趋势已被去除, 故剧烈的理想趋势未使得两种方法的输出季节信号产生较大差异. M-2 对 M-1B 的差异 (图 21) 主要表现为输出季节信号的差异及其导致的输出余项差异, 而输出 trend 和余项均值的差异都很小. 这是因为, 剧烈的理想趋势污染了 M-1B 的季节信号估计, 导致其与 M-2 差异较大; 还使得两种方法的输出余项均值都偏离零且程度相近, 所以输出余项均值差异很小. 由于两种方法在估计 trend 时, 都已去除了季节信号, 不会被季节信号干扰, 故二者输出接近. M-2A 对 M-1A 的差异 (图 22) 主要体现在输出 trend 的边缘差异及其导致的输出余项差异, 而输出季节信号和余项均值差异很小. 这是因为 M-1A 在估计 trend 时未去除季节信号, 故估计的 trend 与 M-2A 的差异较大; 在估计季节信号时, 两种方法都已去除长期趋势, 输出的季节信号差异小; 在估计余项时, 两种方法都有零均值保证, 故输出余项的均值都接近零. M-2A 对 M-1B 的差异 (图 23) 主要表现为输出季节信号差异和输出余项均值差异及二者导致的输出余项差异, 而输出 trend 差异较小. 这是因为, M-1B 估计的季节信号估计被剧烈的理想趋势污染, 故与 M-2A 的差异较大; M-1B 的输出余项均值受理想趋势的强非线性影响而远离零, 故与 M-2A 的差异较大; M-1B 的 trend 估计已排除季节信号的干扰, 故与 M-2A 的差异小。

M-1A 与 M-1B 的输出差异. 前面已讨论了 M-2, M-2A 对 M-1A, M-1B 的差异, 据此, 以 M-2 或 M-2A 为中介, 可以解释 M-1A 与 M-1B 的输出差异 (表 16, 表 17, 表 18)。M-1A 与 M-1B 的输出差异主要受以下因素影响. a) 剧烈的理想趋势, 污染 M-1B 的季节信号估计, 但对 M-1A 的影响被抑制, 导致两种方法的输出季节信号差异 (表 16); b) 强非线性的理想趋势, 通过污染 M-1B 的季节信号估计, 增加 M-1A 和 M-1B 的输出差异 (非主导); 还利于增大 M-1B 的输出余项均值的绝对值, 从而增大其输出余项对具有零均值保证的 M-1A 的差异 (表 17); c) 大的理想余项方差, 利于增大 M-1B 的输出余项均值的绝对值, 从而增大其输出余项对具有零均值保证的 M-1A 的差异 (表 17); d) 大幅度的季节信号, 通过影响 M-1A 的线性拟合及被 LWLR 边缘效应放大, 利于增大 M-1A 的输出 trend 与在估计 trend 前已去除长期趋势的 M-1B 的差异 (仅边缘区, 表 18)。上述四个因素导致了 M-1A 和 M-1B 的输出差异。

预估-校正方法 (M-2 系列) 总是优于简单方法 (M-1 系列) 吗? 待补充。

长期趋势对分解结果的影响. 在本文的试验设置下, 长期趋势的线性 (对应算例 C1, 3, 5, 7, 9) 与非线性 (对应算例 C2, 4, 6, 8, 10) 对 M-1, 2 系列的影响较小 (图 11, 图 12, 图 13). 当长期趋势剧烈 (对应算例 C7, 8) 时, M-1, 2 系列的 CVE 略有增大 (图 14), M-1B 和 M-2 的余项样本均值增大 (图 15). 长期趋势的非线性对 M-3 系列的影响很大. 当长期趋势的实际形式与 M-3 模型假定相同时 (算例 C1, 3, 5, 7, 9 之于 M-3L, C2, 4, 6, 8, 10 之于 M-3Q), 方法输出与理想信号一致性好; 当长期趋势的实际形式不能被 M-3 模型包含时 (算例 C2, 4, 6, 8, 10 之于 M-3L), 方法输出与理想信号一致性不好; 当长期趋势的实际形式被 M-3 模型的退化版本包含时 (算例 C1, 3, 5, 7, 9 之于 M-3Q), 方法输出与理想信号的一致性比非退化时 (算例 C1, 3, 5, 7, 9 之于 M-3L) 略有恶化 (图 9, 图 10).

季节周期对分解结果的影响. 在理想试验中, 当季节信号的幅度或解释方差较小时 (算例 C9, 10), M-1A, 1B, 2, 2A 这四个采取简单平均提取季节信号的方法的输出几乎一样, 均存在畸变, 与 M-2S 和 M-3 系列等采取谱方法提取季节信号的方法的输出有差异.

余项和季节信号的零样本均值性. M-1A, 2A, 3L, 3Q 方法可保证 residual 的样本均值接近零 (图 15). 这是因为, M-1A, 2A 方法的 residual 是 deseason 操作的副产品, 其均值 (如果非零) 已被归入 seasonal component. **M-3L, 3Q 的 residual 零样本均值性似乎不能获得理论保证?** 当采样时间是年周期的整数倍时 (本文所有算例都满足), 所有方法输出的 seasonal component 的样本均值都接近零, 这是由各方法的定义保证的.

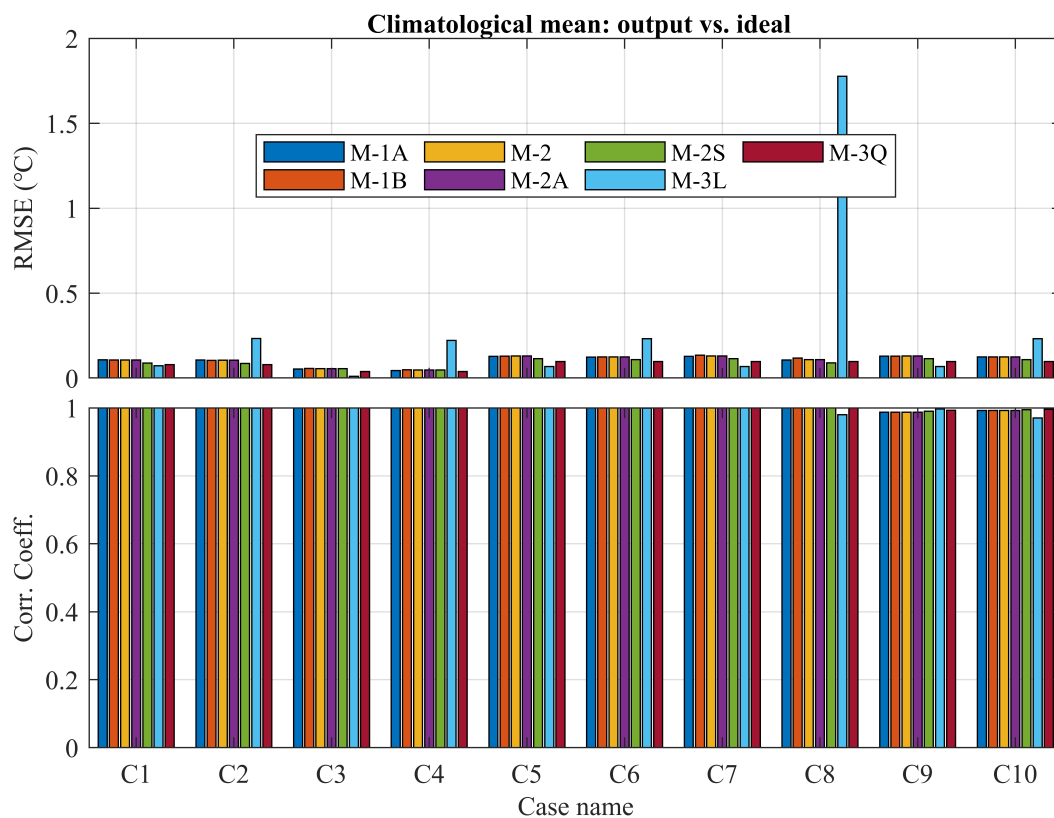


图 11 在理想试验中, 三类 classical decomposition methods 输出的气候平均态 (trend + seasonal components) 对理想时间序列中气候平均态的样本均方根误差(RMSE)和 Pearson 线性相关系数(CC).

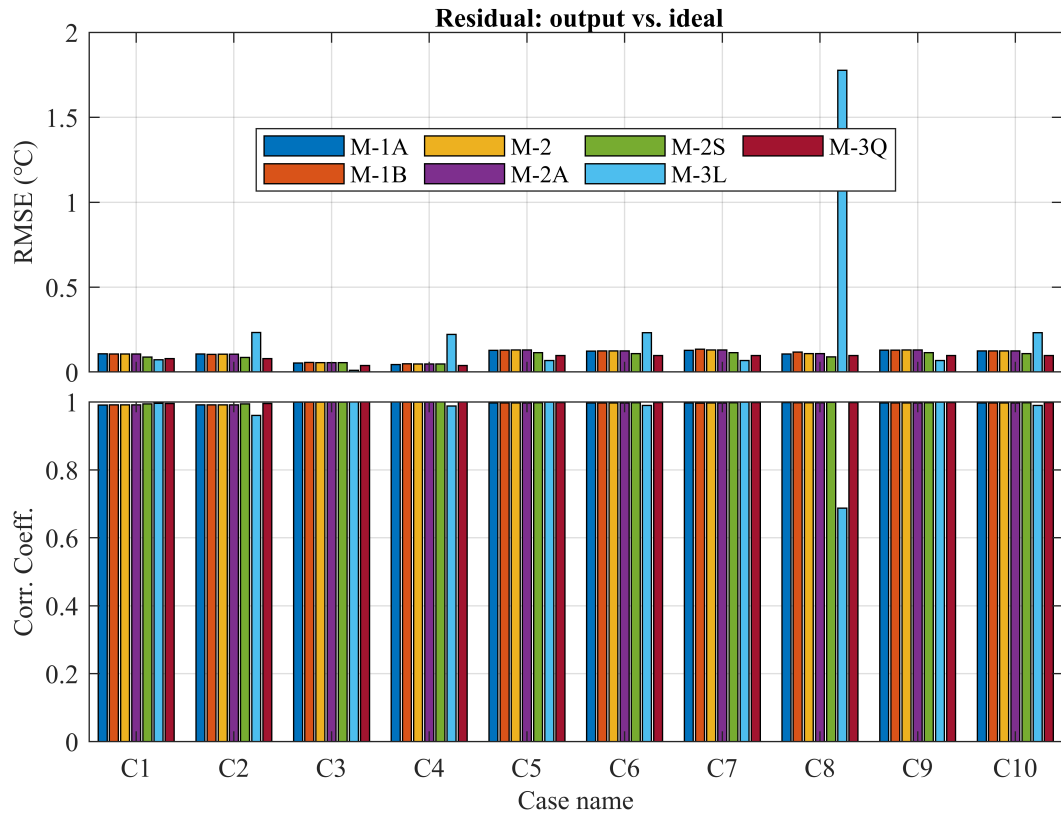


图 12 在理想试验中，三类 classical decomposition methods 输出的 residual 分量对理想时间序列中 residual 分量的样本均方根误差(RMSE)和 Pearson 线性相关系数(CC).

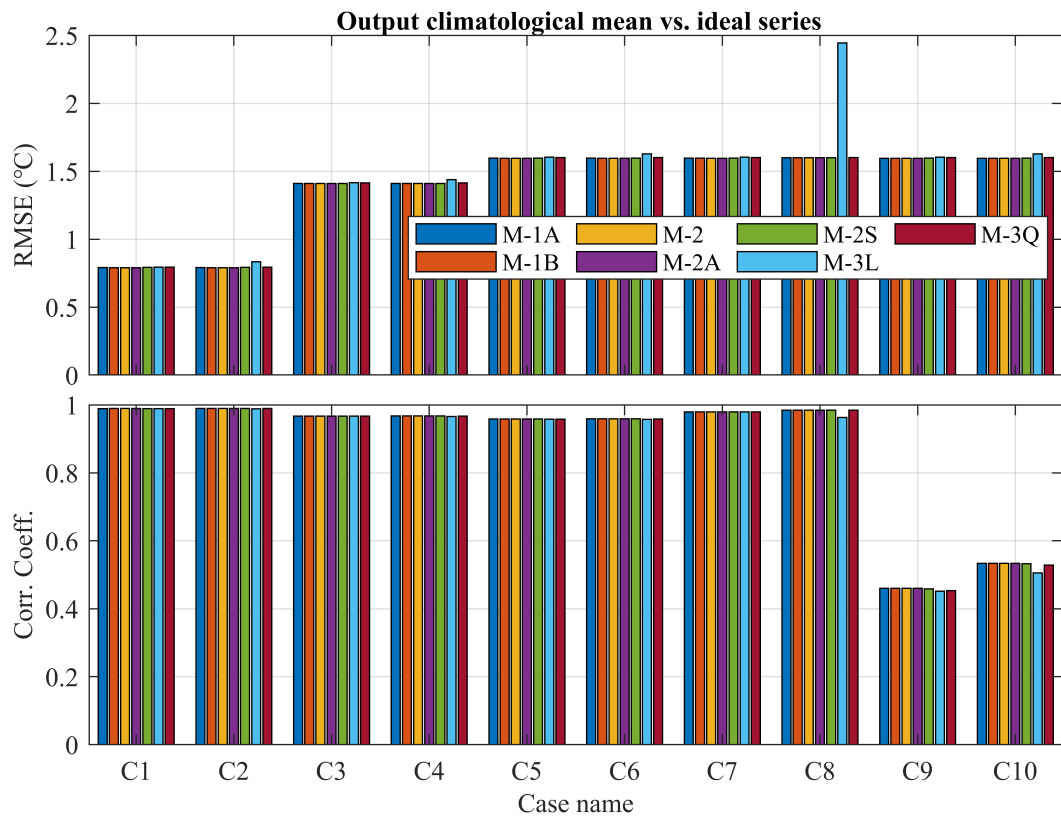


图 13 在理想试验中，三类 classical decomposition methods 输出的气候平均态 (trend + seasonal components) 对理想时间序列的样本均方根误差(RMSE)和 Pearson 线性相关系数(CC).

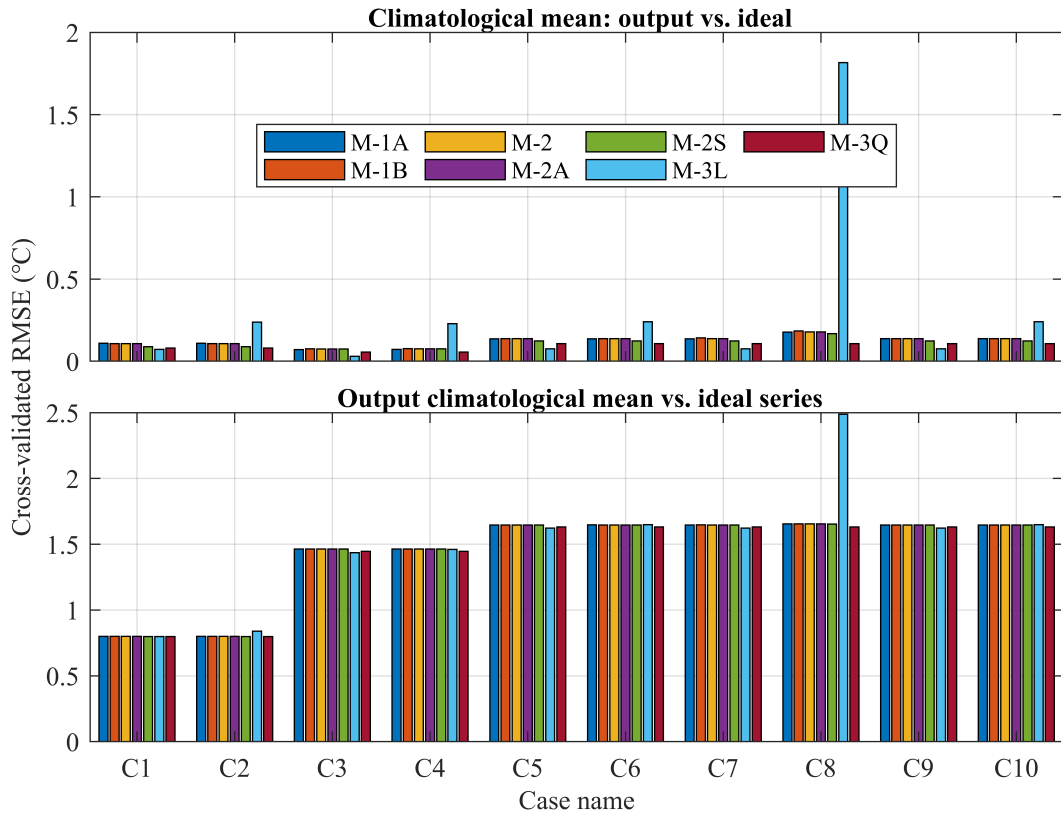


图 14 在理想试验中，三类 classical decomposition methods 输出的气候平均态 (trend + seasonal components) 对理想时间序列(b)及其气候平均态(a)的交叉验证均方根误差(CVE).

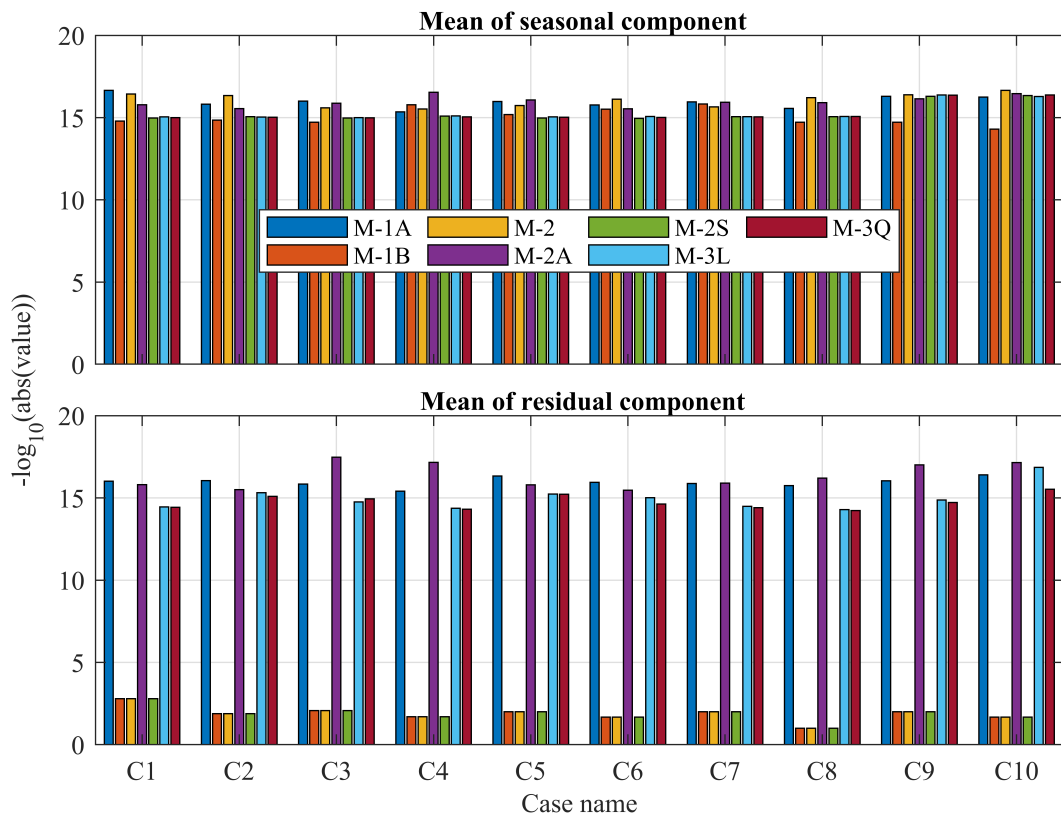


图 15 在理想试验中，三类 classical decomposition methods 输出的气候平均态 (trend + seasonal components) 对理想时间序列(b)及其气候平均态(a)的交叉验证均方根误差(CVE).



表 2 在理想试验中, M-2A minus M-2 方法输出的季节分量的差异.

Case Name:	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
Max.Err. (1e-8 °C)	26.8	26.9	27.1	27.1	26.9	26.9	27.4	27.5	1.76	1.77
(Order↓):	(8)	(7)	(4)	(3)	(6)	(5)	(2)	(1)	(10)	(9)
Min.Err. (1e-8 °C)	-16.8	-16.8	-16.9	-16.9	-16.8	-16.8	-17.1	-17.2	-1.10	-1.11
(Order↑):	(8)	(7)	(4)	(3)	(6)	(5)	(2)	(1)	(10)	(9)
Max.Abs.Err. (1e-8 °C)	26.8	26.9	27.1	27.1	26.9	26.9	27.4	27.5	1.76	1.77
(Order↓):	(8)	(7)	(4)	(3)	(6)	(5)	(2)	(1)	(10)	(9)
RMSE (1e-9 °C)	132	132	133	133	132	132	135	135	8.57	8.65
(Order↓):	(8)	(7)	(4)	(3)	(6)	(5)	(2)	(1)	(10)	(9)
Mean.Abs.Err. (1e-9 °C)	111	111	112	112	111	112	114	114	7.19	7.26
(Order↓):	(8)	(7)	(4)	(3)	(6)	(5)	(2)	(1)	(10)	(9)

表 3 在理想试验中, M-2A minus M-2 方法输出的余项的差异.

Case Name:	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
Max.Err. (1e-3 °C)	1.59	12.9	8.49	19.8	10.1	21.4	10.1	101	10.1	21.4
(Order↓):	(10)	(5)	(9)	(4)	(7)	(2)	(6)	(1)	(8)	(3)
Min.Err. (1e-3 °C)	1.59	12.9	8.49	19.8	10.1	21.4	10.1	101	10.1	21.4
(Order↑):	(1)	(6)	(2)	(7)	(4)	(8)	(3)	(10)	(5)	(9)

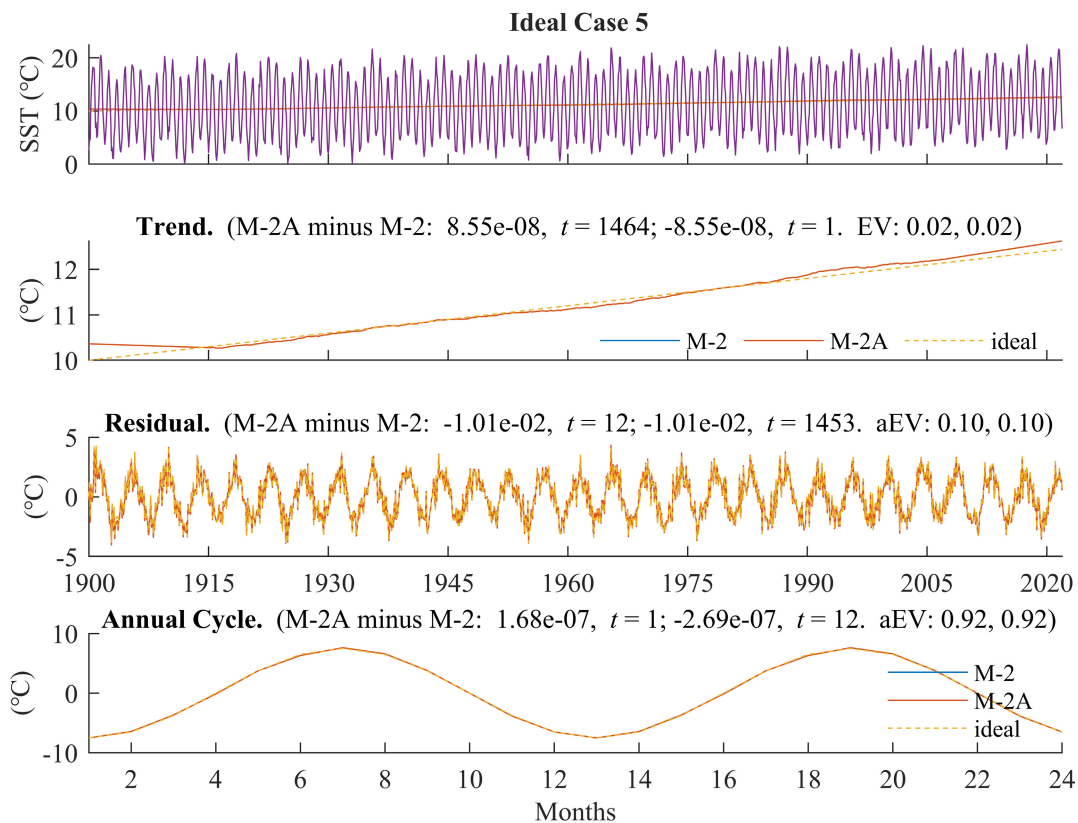


图 16 在理想试验 Case 5 中, M-2 方法与 M-2A 方法的经典分解结果.

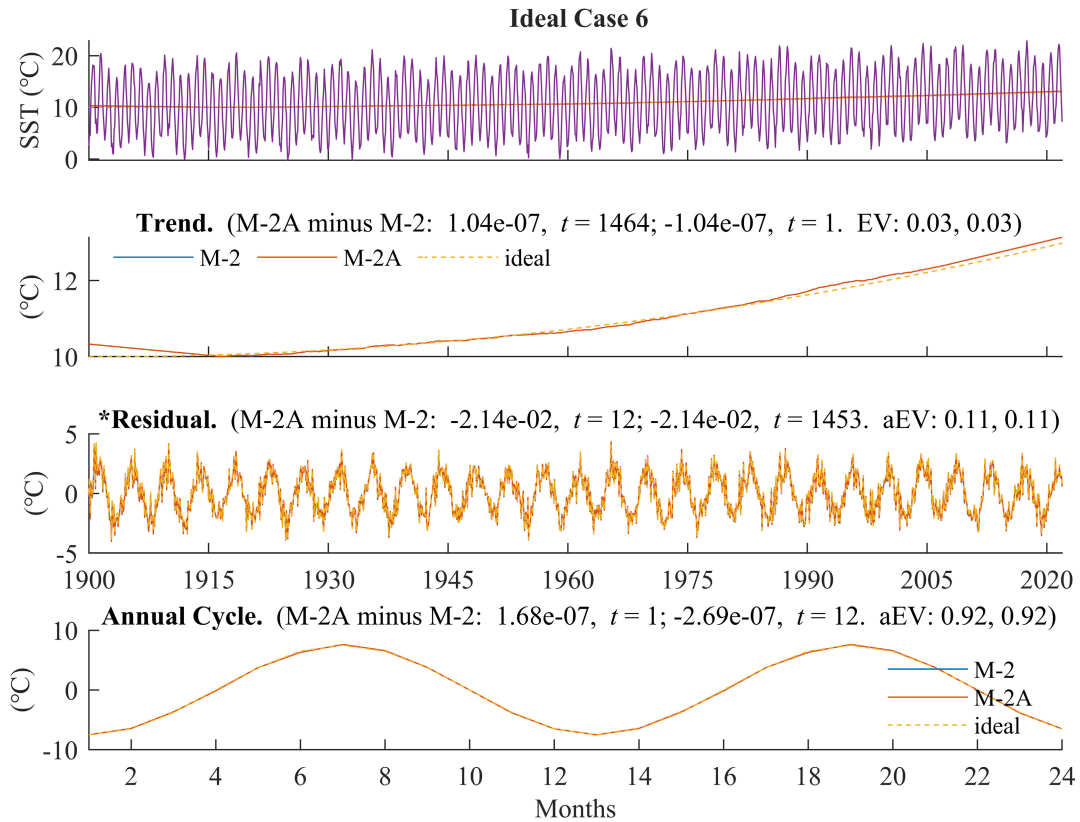


图 17 在理想试验 Case 6 中，M-2 方法与 M-2A 方法的经典分解结果.

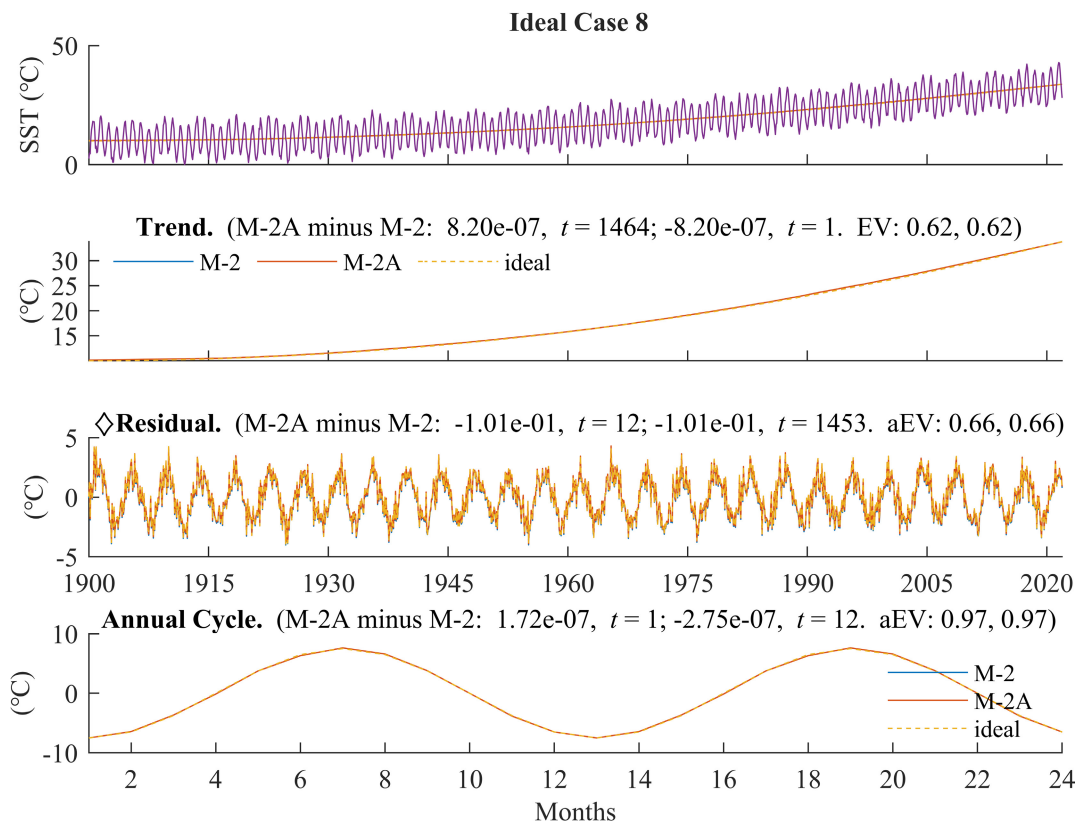


图 18 在理想试验 Case 8 中，M-2 方法与 M-2A 方法的经典分解结果.





表 4 在理想试验中, M-1A minus M-2 方法输出的季节分量的差异.

Case Name:	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
Max.Err. (1e-5 °C) (Order↓):	69.5	69.5	70.3	70.3	69.6	69.6	69.6	69.6	4.07	4.07
Min.Err. (1e-5 °C) (Order↑):	-87	-87	-87.5	-87.5	-87.2	-87.2	-87.2	-87.2	-5.74	-5.74

表 5 在理想试验中, M-1A minus M-2 方法输出的余项的差异.

Case Name:	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
Max.Err. (1e-2 °C) (Order↓):	6.28	7.41	7.08	8.21	7.16	8.29	7.16	16.2	1.37	2.50
Min.Err. (1e-3 °C) (Order↑):	-59.1	-47.8	-53.3	-42.0	-50.9	-39.6	-50.9	39.5	6.46	17.8
Max.Abs.Err. (1e-2 °C) (Order↓):	6.28	7.41	7.08	8.21	7.16	8.29	7.16	16.2	1.37	2.50
RMSE (1e-2 °C) (Order↓):	1.77	2.19	1.99	2.67	2.04	2.78	2.04	10.2	1.01	2.14
Mean.Abs.Err. (1e-3 °C) (Order↓):	8.78	17.6	14.3	23.4	15.4	24.7	15.4	101	10.1	21.4
Mean.Err. (1e-3 °C) (Order↓):	1.59	12.9	8.49	19.8	10.1	21.4	10.1	101	10.1	21.4

表 6 在理想试验中, M-1A minus M-2 方法输出的 trend 的差异.

Case Name:	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
Max.Err. (1e-3 °C) (Order↓):	61.5	61.5	62.7	62.7	61.8	61.8	61.8	61.8	3.67	3.67
Min.Err. (1e-3 °C) (Order↑):	-61.5	-61.5	-62.7	-62.7	-61.8	-61.8	-61.8	-61.8	-3.67	-3.67
RMSE (1e-3 °C) (Order↓):	17.7	17.7	18.0	18.0	17.8	17.8	17.8	17.8	1.05	1.05
Mean.Err. (1e-16 °C) (Order↓):	-53.0	33.3	-29.4	48.1	4.60	-19.4	72.0	25.9	-42.2	40.5

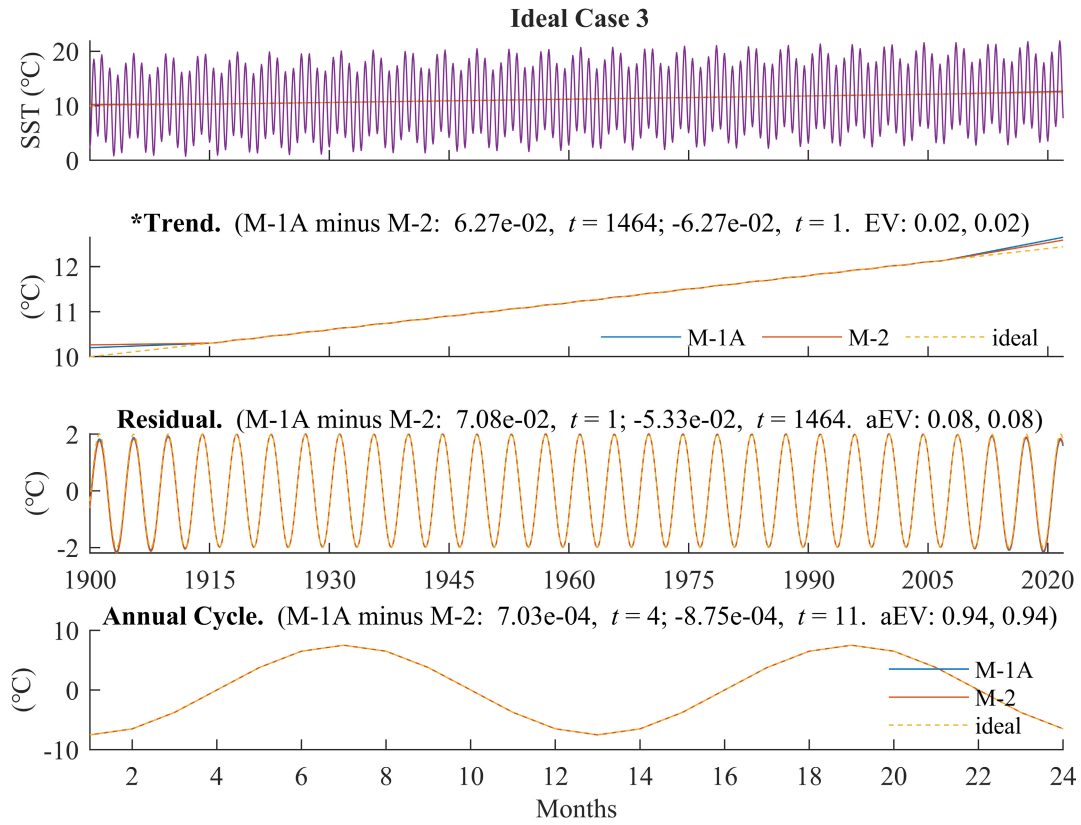


图 19 在理想试验 Case 3 中, M-1A 方法与 M-2 方法的经典分解结果.

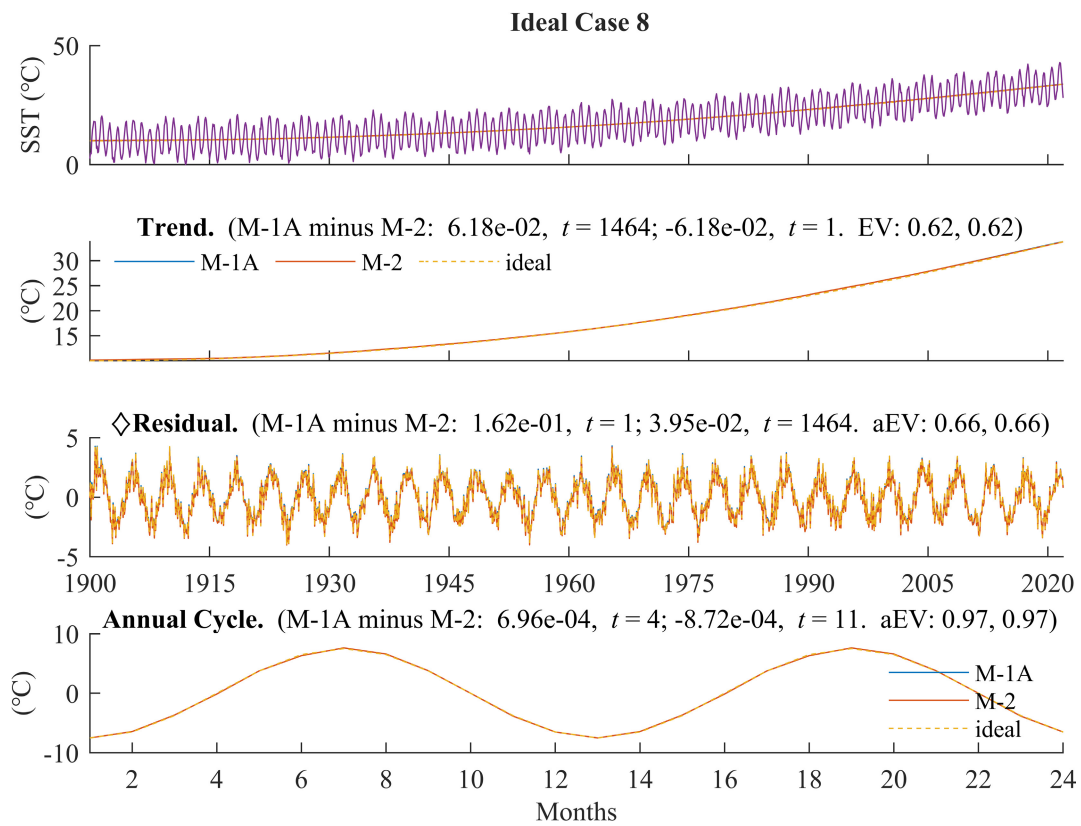


图 20 在理想试验 Case 8 中, M-1A 方法与 M-2 方法的经典分解结果.





表 7 在理想试验中, M-1B minus M-2 方法输出的季节信号的差异.

Case Name:	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
Max.Err. (1e-3 °C)	8.86	10.9	8.75	10.8	8.44	10.5	72.6	89.0	8.44	10.5
(Order↓):	(7)	(3)	(8)	(4)	(9)	(5)	(2)	(1)	(10)	(6)
Min.Err. (1e-3 °C)	-9.11	-11.1	-8.73	-10.7	-8.67	-10.7	-72.8	-88.9	-8.67	-10.7
(Order↑):	(7)	(3)	(8)	(4)	(9)	(5)	(2)	(1)	(10)	(6)
Max.Abs.Err. (1e-3 °C) (Order↓):	9.11	11.1	8.75	10.8	8.67	10.7	72.8	89.0	8.67	10.7
RMSE (1e-3 °C)	5.61	6.88	5.48	6.75	5.34	6.61	45.6	55.8	5.34	6.61
(Order↓):	(7)	(3)	(8)	(4)	(9)	(5)	(2)	(1)	(10)	(6)
Mean.Abs.Err. (1e-3 °C) (Order↓):	4.85	5.96	4.77	5.87	4.62	5.72	39.6	48.4	4.62	5.72
Mean.Err. (1e-16 °C)	16.3	14.2	17.6	-5.4	5.13	-2.87	-4.44	18.7	-19.4	-50.3
(Order↓):	(3)	(4)	(2)	(8)	(5)	(6)	(7)	(1)	(9)	(10)

表 8 在理想试验中, M-1B minus M-2 方法输出的余项的差异.

Case Name:	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
Max.Err. (1e-3 °C)	9.41	11.5	9.02	11.1	8.96	11.0	75.3	91.9	8.96	11.0
(Order↓):	(7)	(3)	(8)	(4)	(9)	(5)	(2)	(1)	(10)	(6)
Min.Err. (1e-3 °C)	-9.17	-11.3	-9.04	-11.2	-8.73	-10.8	-75.1	-92.0	-8.73	-10.8
(Order↑):	(7)	(3)	(8)	(4)	(9)	(5)	(2)	(1)	(10)	(6)
Max.Abs.Err. (1e-3 °C) (Order↓):	9.41	11.5	9.04	11.2	8.96	11.0	75.3	92.0	8.96	11.0
RMSE (1e-3 °C)	5.61	6.88	5.48	6.75	5.34	6.61	45.6	55.8	5.34	6.61
(Order↓):	(7)	(3)	(8)	(4)	(9)	(5)	(2)	(1)	(10)	(6)
Mean.Abs.Err. (1e-3 °C) (Order↓):	4.85	5.95	4.76	5.87	4.62	5.72	39.6	48.4	4.62	5.72
Mean.Err. (1e-16 °C)	2.31	24.0	57.0	-66.2	-39.6	-15.9	-178	126	29.0	3.42
(Order↓):	(6)	(4)	(2)	(9)	(8)	(7)	(10)	(1)	(3)	(5)

表 9 在理想试验中, M-1B minus M-2 方法输出的 trend 的差异.

Case Name:	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
Max.Err. (1e-4 °C)	3.22	3.95	3.15	3.88	3.06	3.79	26.2	32.0	3.06	3.79
(Order↓):	(7)	(3)	(8)	(4)	(9)	(5)	(2)	(1)	(10)	(6)
Min.Err. (1e-4 °C)	-3.22	-3.95	-3.15	-3.88	-3.06	-3.79	-26.2	-32.0	-3.06	-3.79
(Order↑):	(7)	(3)	(8)	(4)	(9)	(5)	(2)	(1)	(10)	(6)
RMSE (1e-5 °C)	9.24	11.3	9.03	11.1	8.79	10.9	75.1	91.9	8.79	10.9
(Order↓):	(7)	(3)	(8)	(4)	(9)	(5)	(2)	(1)	(10)	(6)
Mean.Err. (1e-16 °C) (Order↓):	-18.6	-38.3	-74.7	71.7	34.5	18.6	182	-145	-9.44	46.7
(Order↓):	(7)	(8)	(9)	(2)	(4)	(5)	(1)	(10)	(6)	(3)

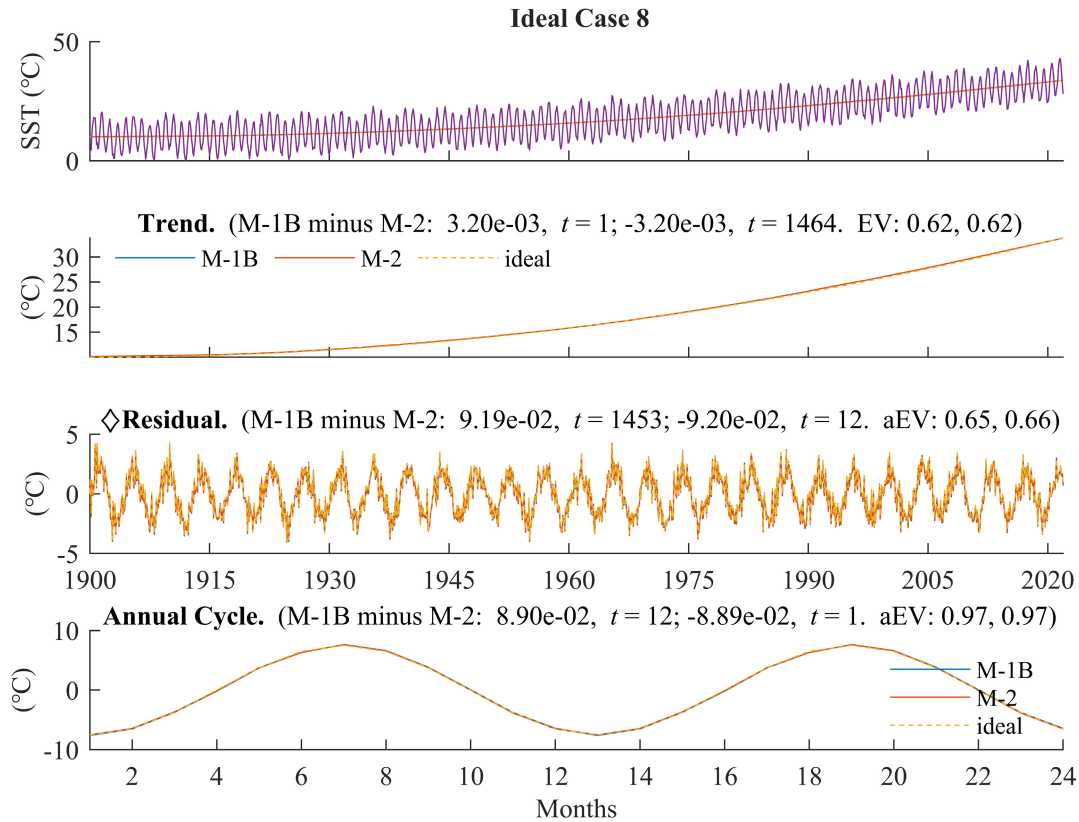


图 21 在理想试验 Case 8 中，M-1B 方法与 M-2 方法的经典分解结果.

表 10 在理想试验中，M-1A minus M-2A 方法输出的季节信号的差异.

Case Name:	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
Max.Err. (1e-5 °C) (Order↓):	69.5	69.5	70.3	70.3	69.6	69.6	69.6	69.6	4.07	4.07
Min.Err. (1e-5 °C) (Order↑):	-87.0	-87.0	-87.5	-87.5	-87.2	-87.2	-87.2	-87.2	-5.74	-5.74

表 11 在理想试验中，M-1A minus M-2A 方法输出的余项的差异.

Case Name:	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
Max.Err. (1e-3 °C) (Order↓):	61.2	61.2	62.3	62.3	61.5	61.5	61.5	61.5	3.65	3.65
Min.Err. (1e-3 °C) (Order↑):	-60.7	-60.7	-61.8	-61.8	-61.0	-61.0	-61.0	-61.0	-3.62	-3.62
Max.Abs.Err. (1e-3 °C) (Order↓):	61.2	61.2	62.3	62.3	61.5	61.5	61.5	61.5	3.65	3.65
RMSE (1e-3 °C) (Order↓):	17.7	17.7	18.0	18.0	17.7	17.7	17.7	17.7	1.05	1.05
Mean.Abs.Err. (1e-4 °C) (Order↓):	77.2	77.2	78.6	78.6	77.6	77.6	77.6	77.6	4.61	4.61
Mean.Err. (1e-18 °C) (Order↓):	177	434	140	-377	276	230	-244	129	103	-9.3



表 12 在理想试验中, M-1A minus M-2A 方法输出的 trend 的差异.

Case Name:	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
Max.Err. (1e-3 °C)	61.5	61.5	62.7	62.7	61.8	61.8	61.8	61.8	3.67	3.67
(Order↓):	(8)	(7)	(2)	(1)	(6)	(5)	(4)	(3)	(10)	(9)
Min.Err. (1e-3 °C)	-61.5	-61.5	-62.7	-62.7	-61.8	-61.8	-61.8	-61.8	-3.67	-3.67
(Order↑):	(8)	(7)	(2)	(1)	(6)	(5)	(4)	(3)	(10)	(9)
RMSE (1e-3 °C)	17.7	17.7	18.0	18.0	17.8	17.8	17.8	17.8	1.05	1.05
(Order↓):	(8)	(7)	(2)	(1)	(6)	(5)	(4)	(3)	(10)	(9)
Mean.Err. (1e-16 °C) (Order↓):	-7.62	26.1	-17.9	-18.1	-36.6	-14.9	-109	93.7	-11.7	-32.3
(Order↓):	(3)	(2)	(6)	(7)	(9)	(5)	(10)	(1)	(4)	(8)

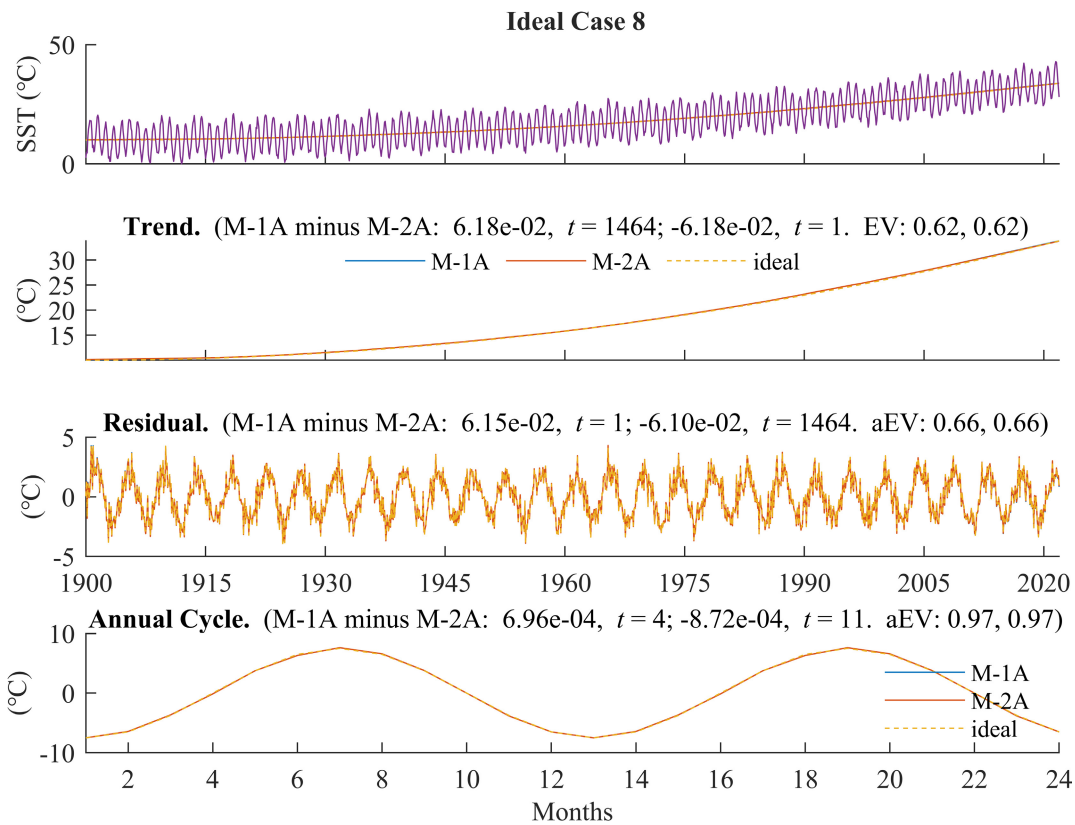


图 22 在理想试验 Case 8 中, M-1A 方法与 M-2A 方法的经典分解结果.

表 13 在理想试验中, M-1B minus M-2A 方法输出的季节信号的差异.

Case Name:	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
Max.Err. (1e-3 °C)	8.86	10.9	8.75	10.8	8.44	10.5	72.6	89.0	8.44	10.5
(Order↓):	(7)	(3)	(8)	(4)	(10)	(5)	(2)	(1)	(9)	(6)
Min.Err. (1e-3 °C)	-9.11	-11.1	-8.73	-10.7	-8.67	-10.7	-72.8	-88.9	-8.67	-10.7
(Order↑):	(7)	(3)	(8)	(4)	(10)	(6)	(2)	(1)	(9)	(5)
Max.Abs.Err. (1e-3 °C) (Order↓):	9.11	11.1	8.75	10.8	8.67	10.7	72.8	89.0	8.67	10.7
RMSE (1e-3 °C)	5.61	6.88	5.48	6.75	5.34	6.61	45.6	55.8	5.34	6.61
(Order↓):	(7)	(3)	(8)	(4)	(9)	(5)	(2)	(1)	(10)	(6)
Mean.Abs.Err. (1e-3 °C) (Order↓):	4.85	5.96	4.77	5.87	4.62	5.72	39.6	48.4	4.62	5.72
(Order↓):	(7)	(3)	(8)	(4)	(9)	(5)	(2)	(1)	(10)	(6)



Case Name:	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
Mean.Err. (1e-17 °C)	140	120	188	-17.4	57.1	-76.0	-4.19	199	-192	-504
(Order↓):	(3)	(4)	(2)	(7)	(5)	(8)	(6)	(1)	(9)	(10)

表 14 在理想试验中, M-1B minus M-2A 方法输出的余项的差异.

Case Name:	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
Max.Err. (1e-4 °C)	78.2	-14.2	5.34	-87.0	-11.2	-104	652	-86.9	-11.2	-104
(Order↓):	(2)	(6)	(3)	(8)	(5)	(10)	(1)	(7)	(4)	(9)
Min.Err. (1e-2 °C)	-1.08	-2.42	-1.75	-3.10	-1.88	-3.22	-8.51	-19.3	-1.88	-3.22
(Order↑):	(10)	(6)	(9)	(5)	(8)	(3)	(2)	(1)	(7)	(4)
Max.Abs.Err. (1e-2 °C) (Order↓):	1.08	2.42	1.75	3.10	1.88	3.22	8.51	19.3	1.88	3.22
(Order↓):	(10)	(6)	(9)	(5)	(8)	(3)	(2)	(1)	(7)	(4)
RMSE (1e-3 °C)	5.83	14.6	10.1	20.9	11.4	22.4	46.7	115	11.4	22.4
(Order↓):	(10)	(6)	(9)	(5)	(7)	(3)	(2)	(1)	(8)	(4)
Mean.Abs.Err. (1e-3 °C) (Order↓):	5.04	12.9	8.53	19.8	10.1	21.4	40.2	101	10.1	21.4
(Order↓):	(10)	(6)	(9)	(5)	(7)	(3)	(2)	(1)	(8)	(4)
Mean.Err. (1e-3 °C)	-1.59	-12.9	-8.49	-19.8	-10.1	-21.4	-10.1	-101	-10.1	-21.4
(Order↓):	(1)	(6)	(2)	(7)	(4)	(9)	(5)	(10)	(3)	(8)

表 15 在理想试验中, M-1B minus M-2A 方法输出的 trend 的差异.

Case Name:	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
Max.Err. (1e-4 °C)	3.22	3.95	3.15	3.87	3.06	3.79	26.2	32.0	3.06	3.79
(Order↓):	(7)	(3)	(8)	(4)	(10)	(5)	(2)	(1)	(9)	(6)
Min.Err. (1e-4 °C)	-3.22	-3.95	-3.15	-3.87	-3.06	-3.79	-26.2	-32.0	-3.06	-3.79
(Order↑):	(7)	(3)	(8)	(4)	(10)	(6)	(2)	(1)	(9)	(5)
RMSE (1e-5 °C)	9.24	11.3	9.03	11.1	8.79	10.9	75.1	91.9	8.79	10.9
(Order↓):	(7)	(3)	(8)	(4)	(10)	(6)	(2)	(1)	(9)	(5)
Mean.Err. (1e-16 °C) (Order↓):	26.8	-45.4	-63.1	5.52	-6.69	23.1	1.02	-77.1	21.0	-26.2
(Order↓):	(1)	(8)	(9)	(4)	(6)	(2)	(5)	(10)	(3)	(7)

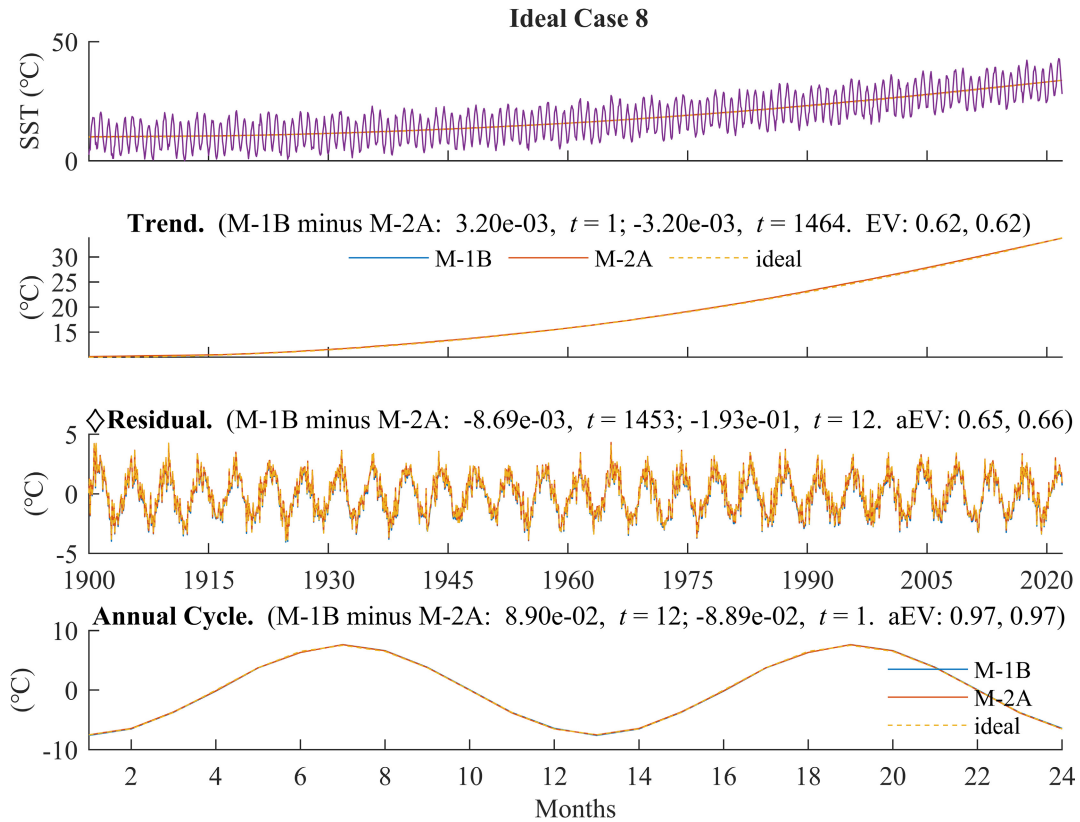


图 23 在理想试验 Case 8 中, M-1B 方法与 M-2A 方法的经典分解结果.

表 16 在理想试验中, M-1B minus M-1A 方法输出的季节信号的差异.

Case Name:	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
Max.Err. (1e-3 °C)	9.68	11.7	9.58	11.6	9.26	11.3	73.4	89.8	8.49	10.5
(Order↓):	(7)	(3)	(8)	(4)	(9)	(5)	(2)	(1)	(10)	(6)
Min.Err. (1e-3 °C)	-9.44	-11.4	-9.07	-11.1	-9.00	-11.0	-73.2	-89.2	-8.69	-10.7
(Order↑):	(7)	(3)	(8)	(4)	(9)	(5)	(2)	(1)	(10)	(6)
Max.Abs.Err. (1e-3 °C) (Order↓):	9.68	11.7	9.58	11.6	9.26	11.3	73.4	89.8	8.69	10.7
RMSE (1e-3 °C) (Order↓):	6.15	7.42	6.03	7.30	5.88	7.15	46.2	56.3	5.38	6.65
Mean.Abs.Err. (1e-3 °C) (Order↓):	5.39	6.49	5.30	6.41	5.15	6.26	40.2	49.0	4.65	5.75
Mean.Err. (1e-16 °C) (Order↓):	15.9	16.0	20.2	-5.70	7.81	-4.52	-2.02	21.1	-19.0	-50.2
(Order↓):	(4)	(3)	(2)	(8)	(5)	(7)	(6)	(1)	(9)	(10)

表 17 在理想试验中, M-1B minus M-1A 方法输出的余项的差异.

Case Name:	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
Max.Err. (1e-3 °C)	65.9	56.7	59.7	50.5	57.3	48	124	49.7	2.34	-6.89
(Order↓):	(2)	(5)	(3)	(6)	(4)	(8)	(1)	(7)	(9)	(10)
Min.Err. (1e-2 °C)	-6.93	-8.27	-7.72	-9.06	-7.77	-9.11	-14.4	-25.1	-2.23	-3.57
(Order↑):	(8)	(5)	(7)	(4)	(6)	(3)	(2)	(1)	(10)	(9)
Max.Abs.Err. (1e-2 °C) (Order↓):	6.93	8.27	7.72	9.06	7.77	9.11	14.4	25.1	2.23	3.57
(Order↓):	(8)	(5)	(7)	(4)	(6)	(3)	(2)	(1)	(10)	(9)

Case Name:	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
RMSE (1e-2 °C)	1.87	2.3	2.07	2.77	2.12	2.86	5.02	11.6	1.15	2.24
(Order↓):	(9)	(5)	(8)	(4)	(7)	(3)	(2)	(1)	(10)	(6)
Mean.Abs.Err. (1e-2 °C) (Order↓):	(9)	(6)	(8)	(4)	(7)	(3)	(2)	(1)	(10)	(5)
Mean.Err. (1e-3 °C)	-1.59	-12.9	-8.49	-19.8	-10.1	-21.4	-10.1	-101	-10.1	-21.4
(Order↓):	(1)	(6)	(2)	(7)	(4)	(9)	(5)	(10)	(3)	(8)

表 18 在理想试验中, M-1B minus M-1A 方法输出的 trend 的差异.

Case Name:	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
Max.Err. (1e-3 °C)	61.8	61.9	63.0	63.0	62.1	62.2	64.5	65.0	3.98	4.05
(Order↓):	(8)	(7)	(4)	(3)	(6)	(5)	(2)	(1)	(10)	(9)
Min.Err. (1e-3 °C)	-61.8	-61.9	-63.0	-63.0	-62.1	-62.2	-64.5	-65.0	-3.98	-4.05
(Order↑):	(8)	(7)	(4)	(3)	(6)	(5)	(2)	(1)	(10)	(9)
Max.Abs.Err. (1e-3 °C) (Order↓):	(8)	(7)	(4)	(3)	(6)	(5)	(2)	(1)	(10)	(9)
RMSE (1e-3 °C)	17.8	17.8	18.1	18.1	17.8	17.9	18.5	18.7	1.14	1.16
(Order↓):	(8)	(7)	(4)	(3)	(6)	(5)	(2)	(1)	(10)	(9)
Mean.Abs.Err. (1e-4 °C) (Order↓):	(8)	(7)	(4)	(3)	(6)	(5)	(2)	(1)	(10)	(9)
Mean.Err. (1e-16 °C)	34.4	-71.6	-45.2	23.6	29.9	38.0	110	-171	32.7	6.18
(Order↓):	(3)	(9)	(8)	(6)	(5)	(2)	(1)	(10)	(4)	(7)

## 3.2 Classical decomposition of several real time-indices

### 3.2.1 ERSST v5 global monthly SST (1902-2021)

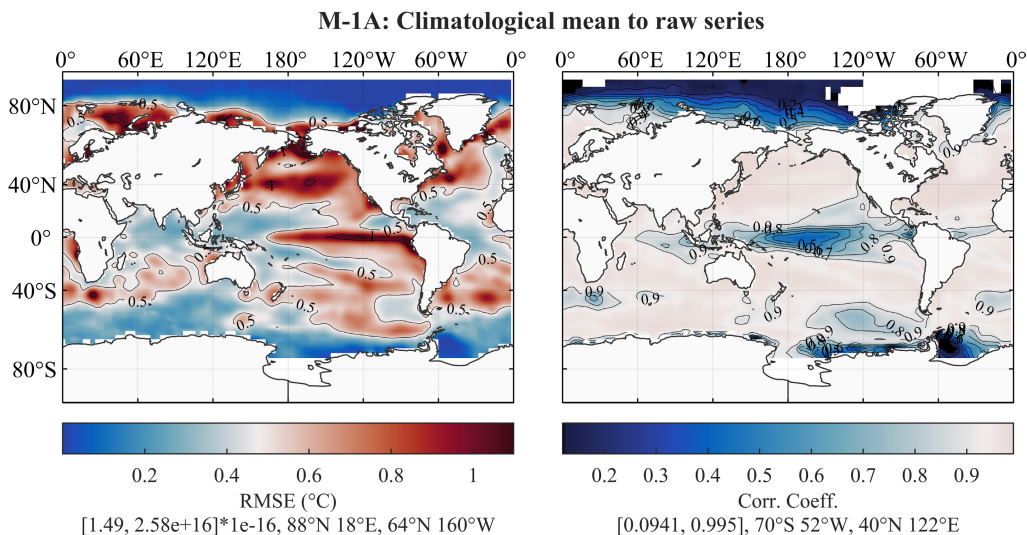


图 24 对 ERSST v5 全球 SST 数据(1902-2021), M-1A 方法输出的气候平均态 (trend + seasonal components) 对原时间序列的 root-mean-square error (RMSE) (a) and Pearson linear correlation coefficients (CC) (b).



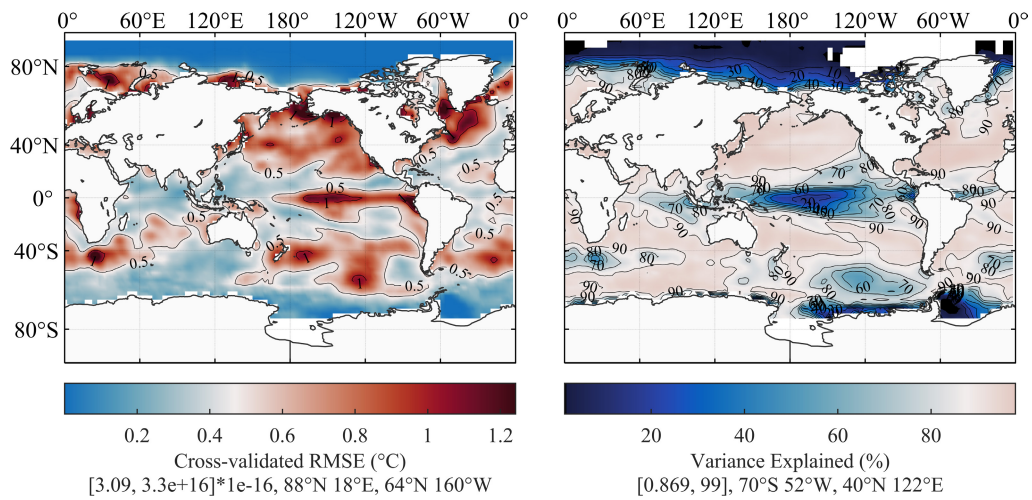
**M-1A: Climatological mean to raw series**

图 25 对 ERSST v5 全球 SST 数据(1902-2021), M-1A 方法输出的气候平均态 (trend + seasonal components) 对原时间序列的 cross-validated RMSE (CVE) (a) and explained variance (EV) (b).

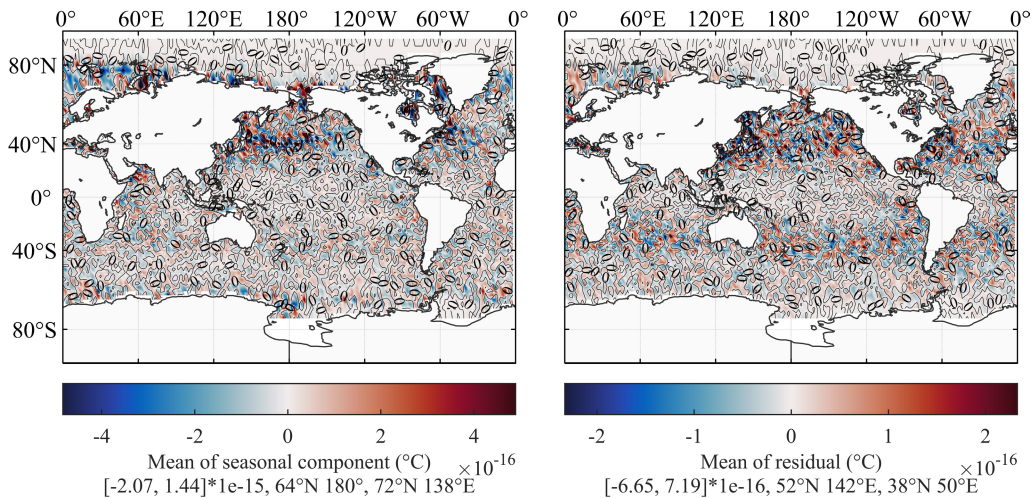
**M-1A: Mean of annual cycle (seasonal component) and residual**

图 26 对 ERSST v5 全球 SST 数据(1902-2021), M-1A 方法输出的 seasonal component (a)和 residual (b)的平均值.

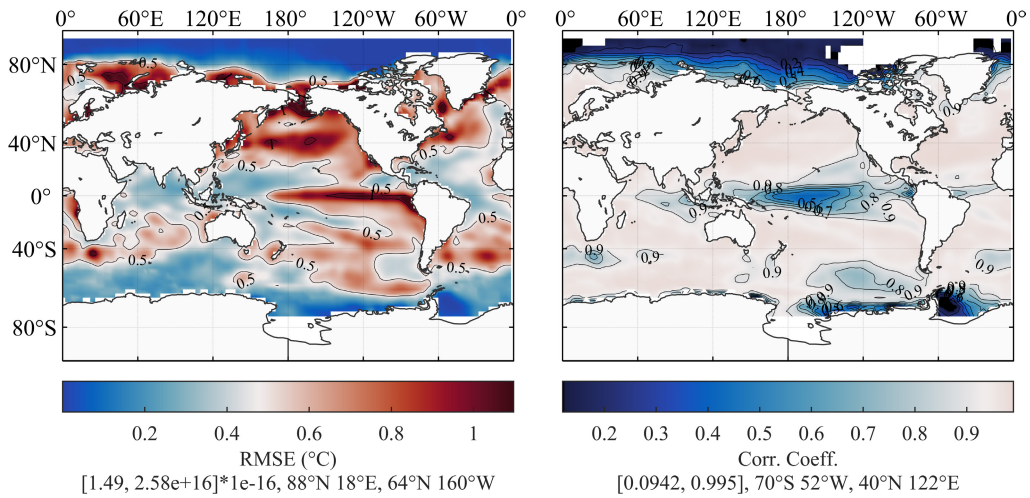
**M-1B: Climatological mean to raw series**

图 27 对 ERSST v5 全球 SST 数据(1902-2021), M-1B 方法输出的气候平均态 (trend + seasonal components) 对原时间序列的 root-mean-square error (RMSE) (a) and Pearson linear correlation coefficients (CC) (b).

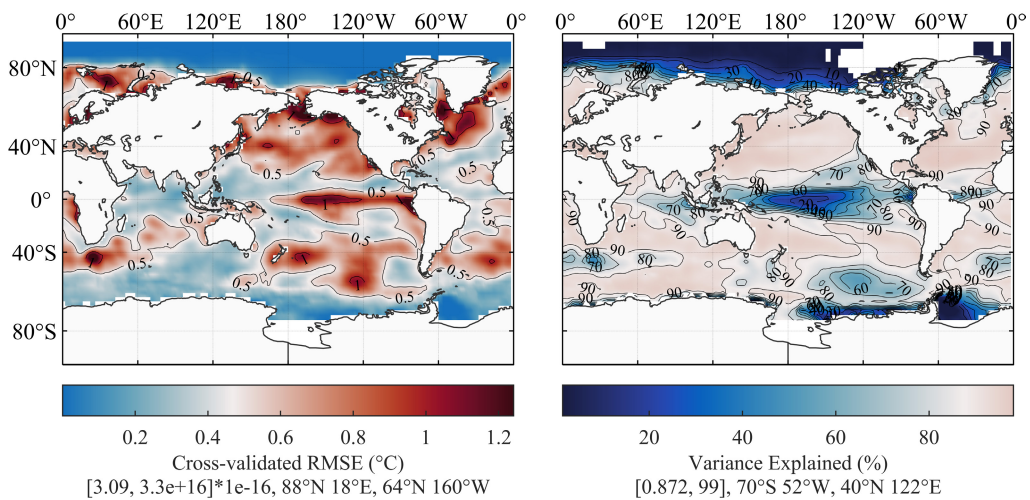
**M-1B: Climatological mean to raw series**

图 28 对 ERSST v5 全球 SST 数据(1902-2021), M-1B 方法输出的气候平均态 (trend + seasonal components) 对原时间序列的 cross-validated RMSE (CVE) (a) and explained variance (EV) (b).



**M-1B: Mean of annual cycle (seasonal component) and residual**

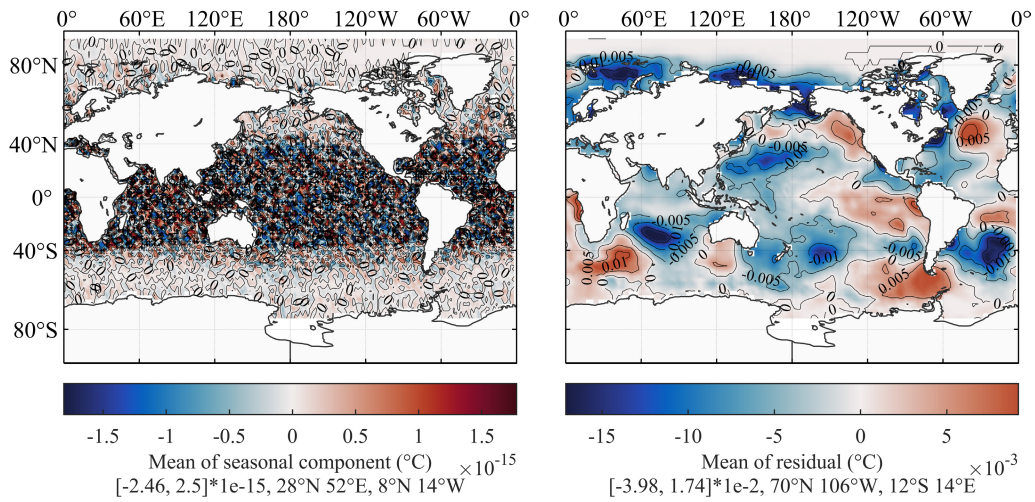


图29 对ERSST v5全球SST数据(1902-2021),M-1B方法输出的seasonal component (a)和residual (b)的平均值.

**Annual cycle: M-1B minus M-1A**

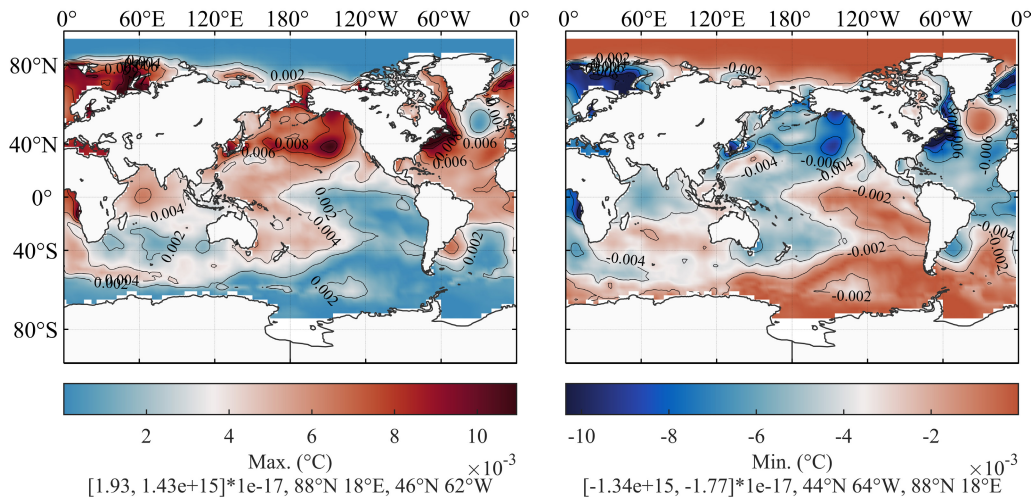


图30 对ERSST v5全球SST数据(1902-2021),M-1B方法与M-1A方法输出的seasonal component之差.

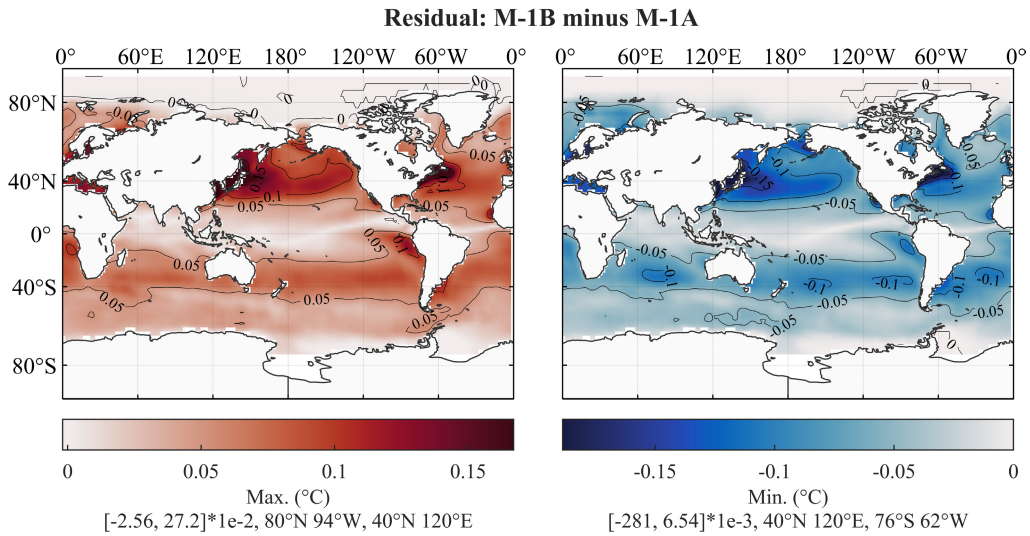


图 31 对 ERSST v5 全球 SST 数据(1902-2021), M-1B 方法与 M-1A 方法输出的 residual 之差.

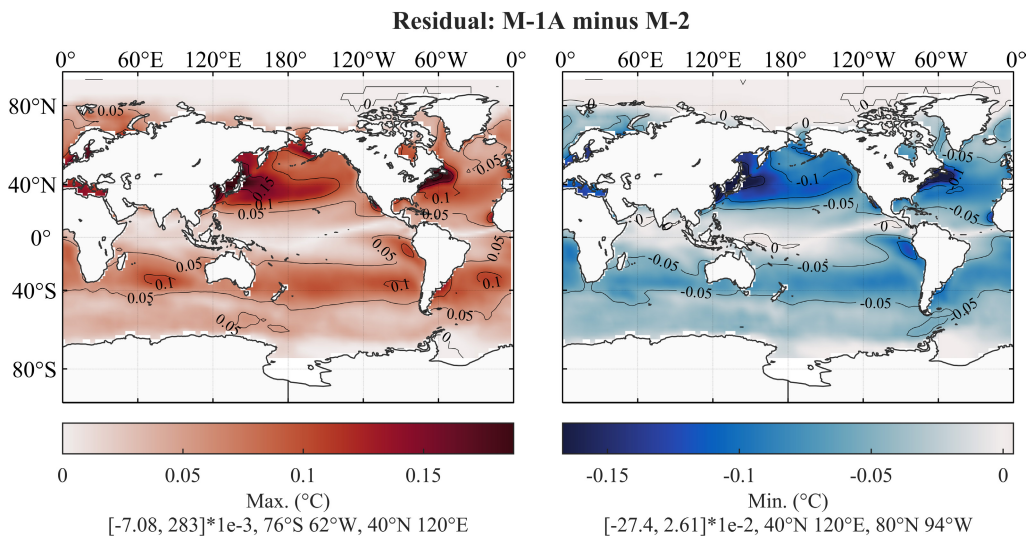


图 32 对 ERSST v5 全球 SST 数据(1902-2021), M-1A 方法与 M-2 方法输出的 residual 之差.

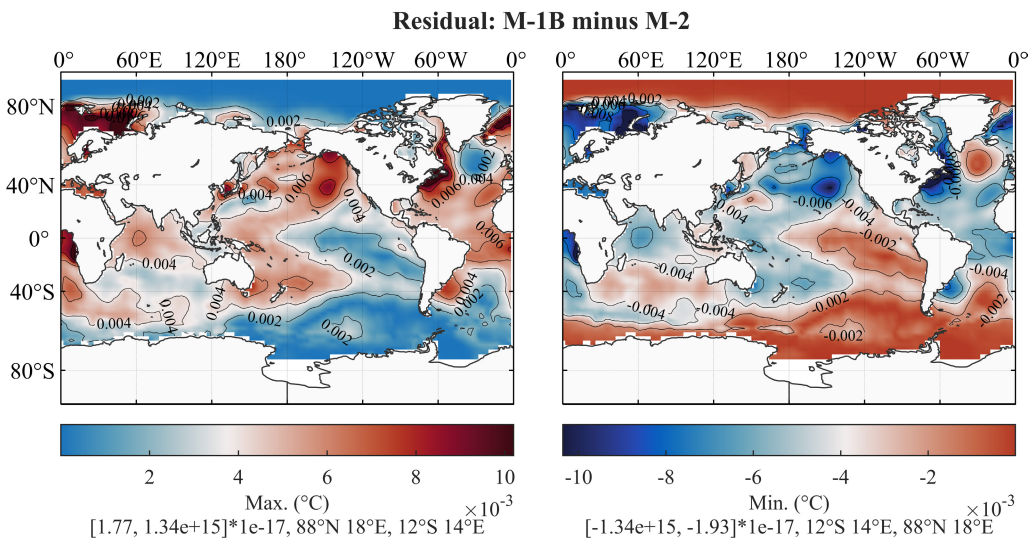


图 33 对 ERSST v5 全球 SST 数据(1902-2021), M-1B 方法与 M-2 方法输出的 residual 之差.

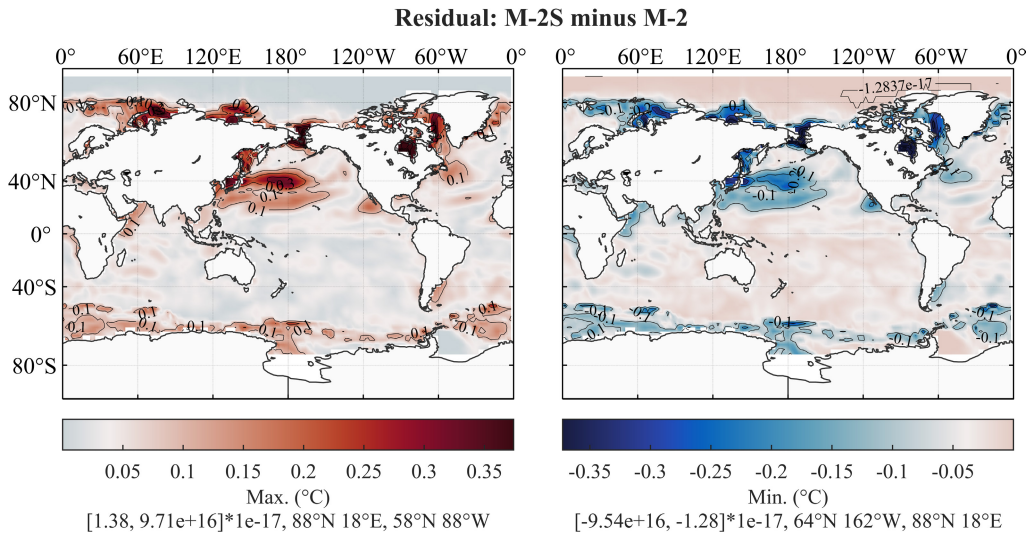


图 34 对 ERSST v5 全球 SST 数据(1902-2021), M-2S 方法与 M-2 方法输出的 residual 之差.

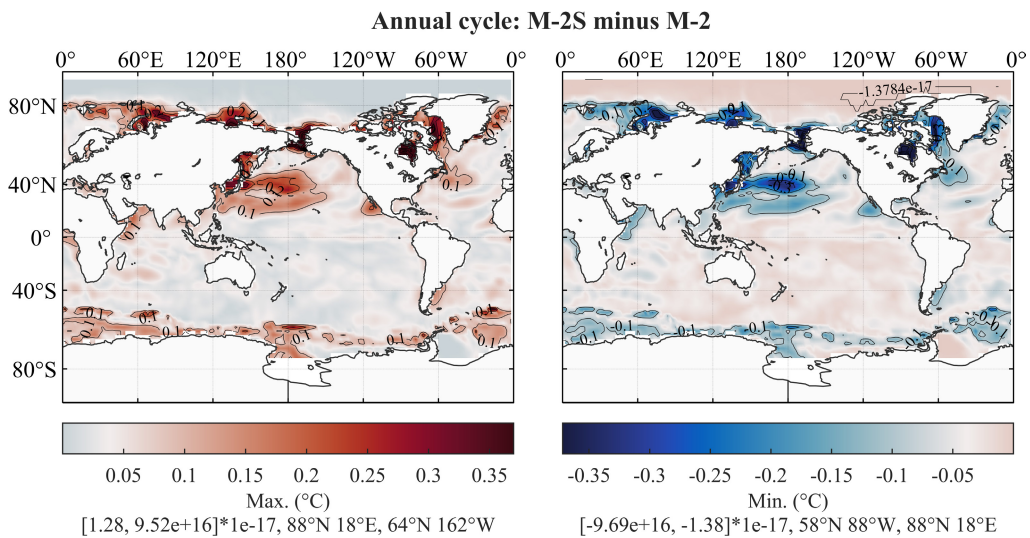


图 35 对 ERSST v5 全球 SST 数据(1902-2021), M-2S 方法与 M-2 方法输出的 seasonal component 之差.

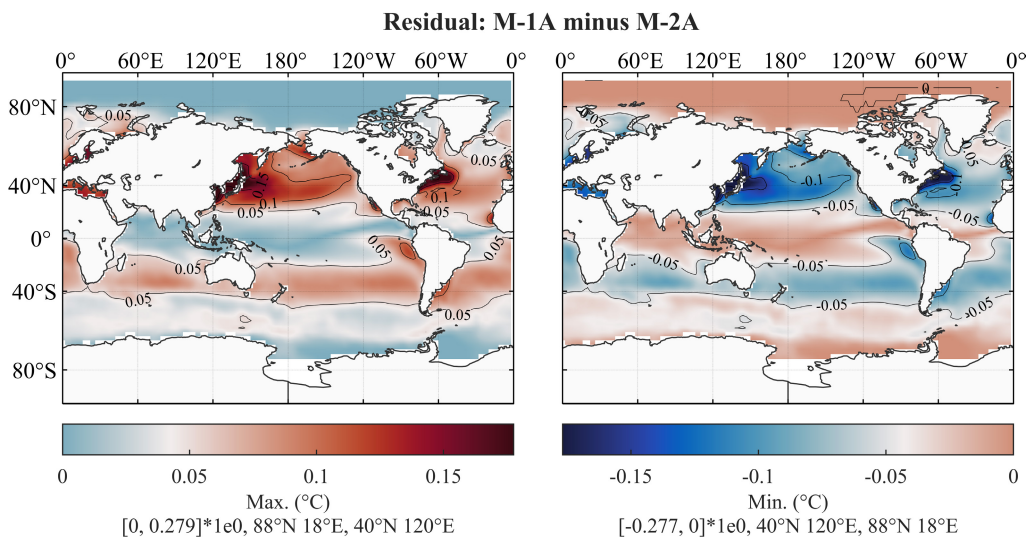


图 36 对 ERSST v5 全球 SST 数据(1902-2021), M-2A 方法与 M-1A 方法输出的 residual 之差.



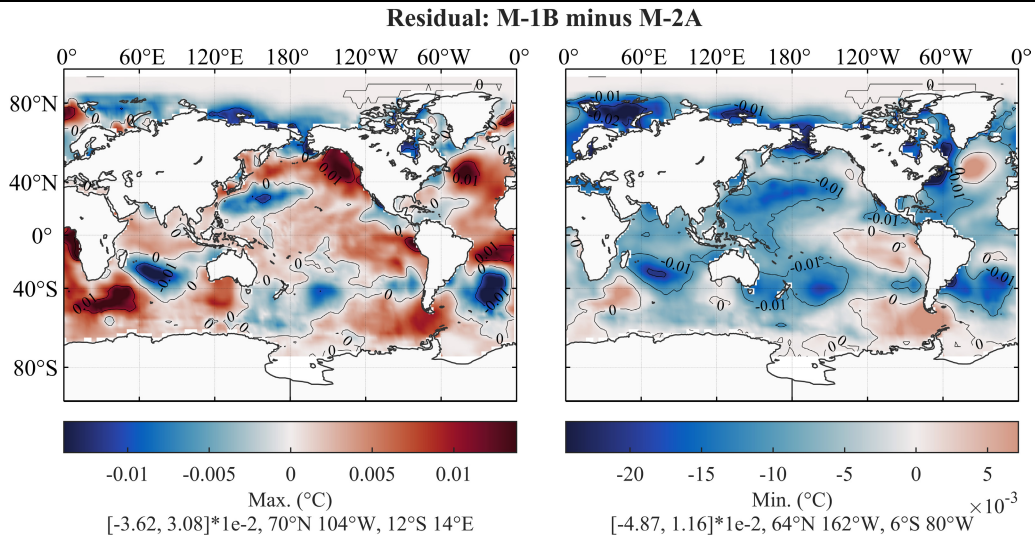


图 37 对 ERSST v5 全球 SST 数据(1902-2021), M-2A 方法与 M-1B 方法输出的 residual 之差.

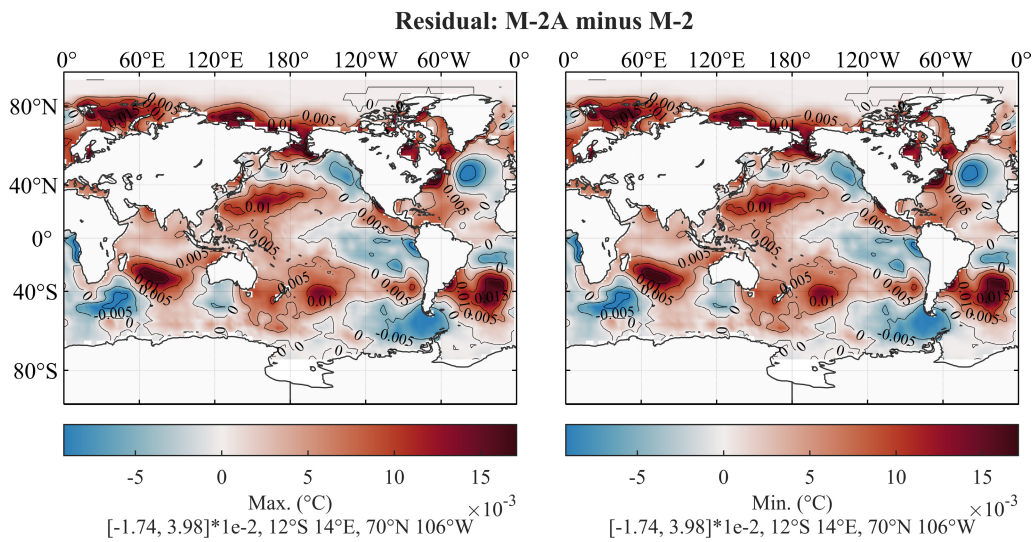


图 38 对 ERSST v5 全球 SST 数据(1902-2021), M-2A 方法与 M-2 方法输出的 residual 之差.

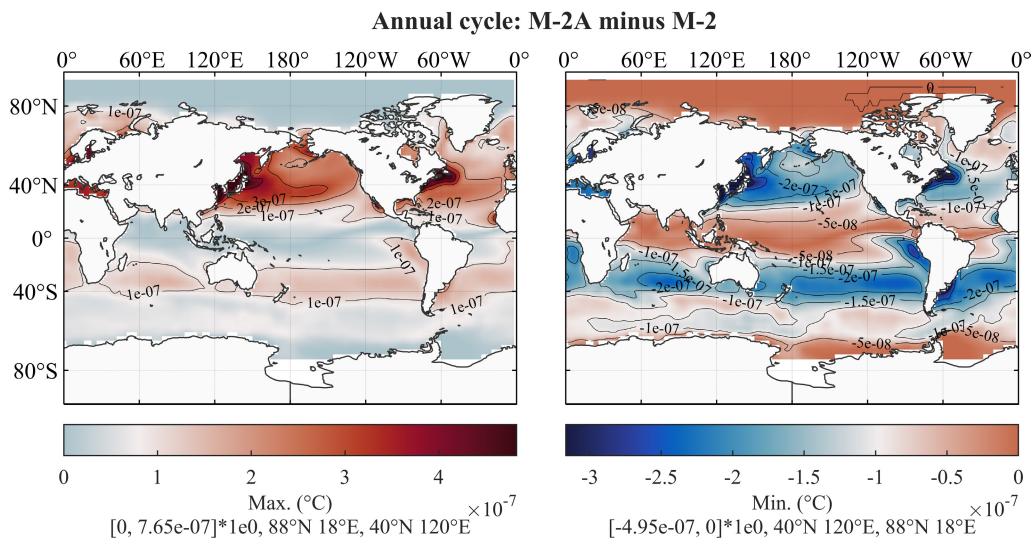


图 39 对 ERSST v5 全球 SST 数据(1902-2021), M-2A 方法与 M-2 方法输出的 seasonal component 之差.

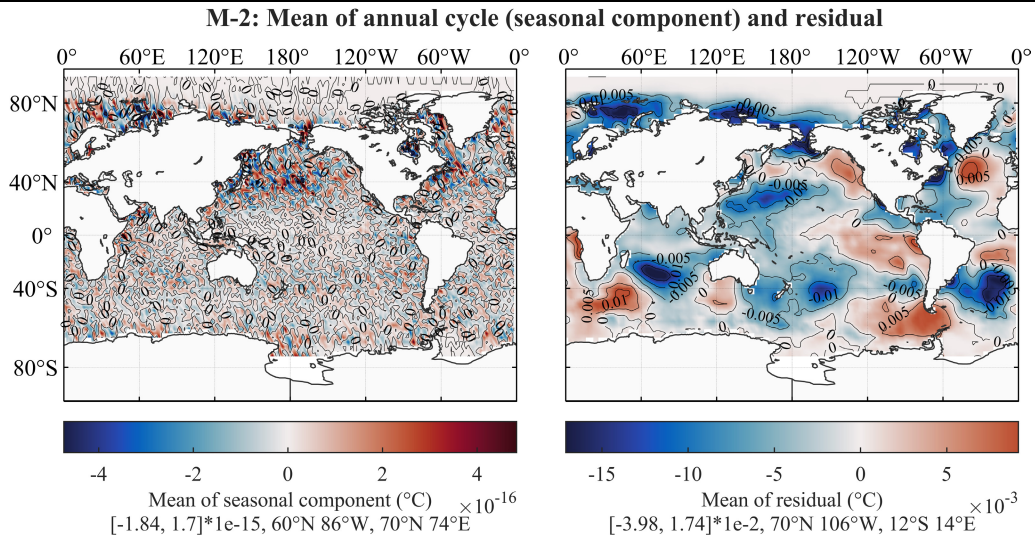


图 40 对 ERSST v5 全球 SST 数据(1902-2021), M-2 方法输出的 seasonal component (a)和 residual (b)的平均值.

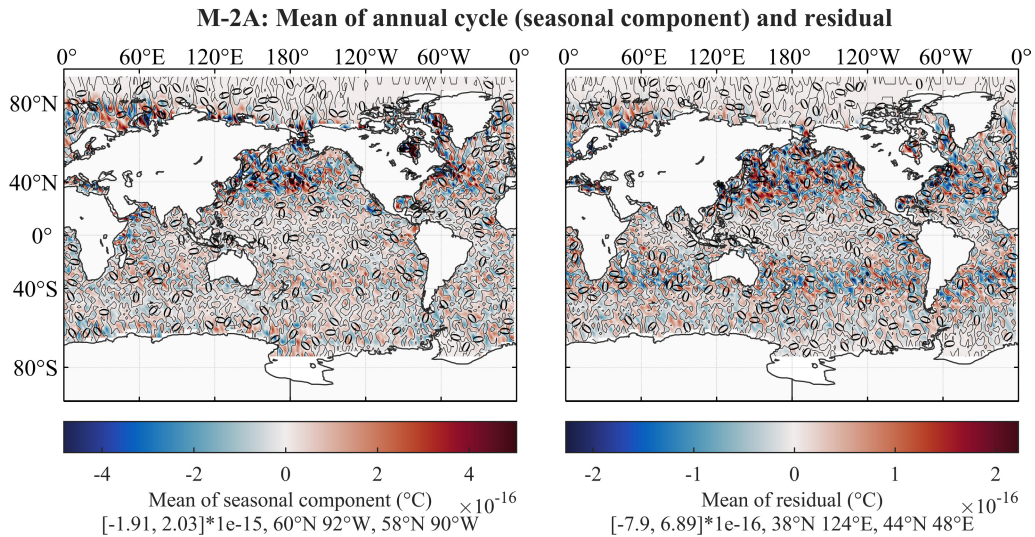


图 41 对 ERSST v5 全球 SST 数据(1902-2021), M-2A 方法输出的 seasonal component (a)和 residual (b)的平均值.

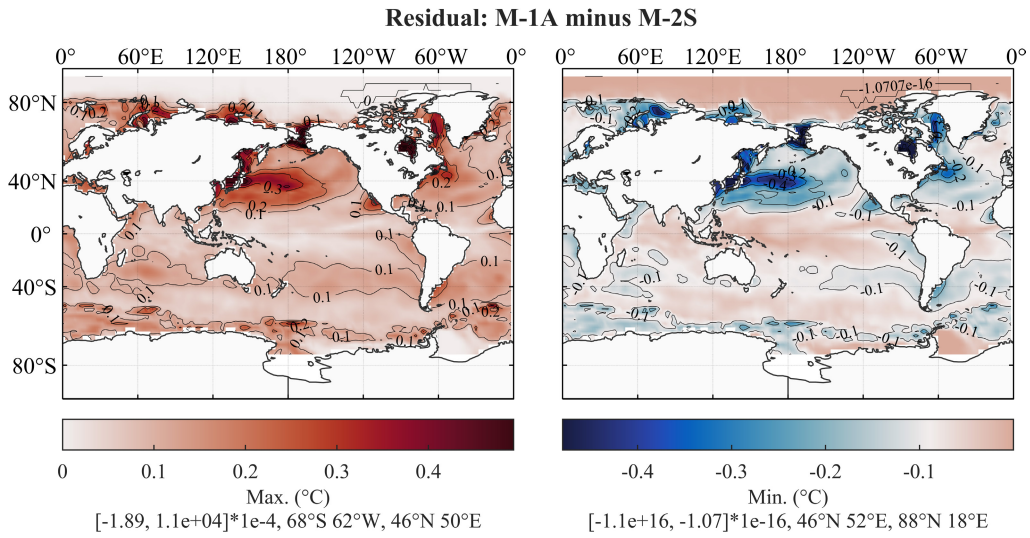


图 42 对 ERSST v5 全球 SST 数据(1902-2021), M-2S 方法与 M-1A 方法输出的 residual 之差.

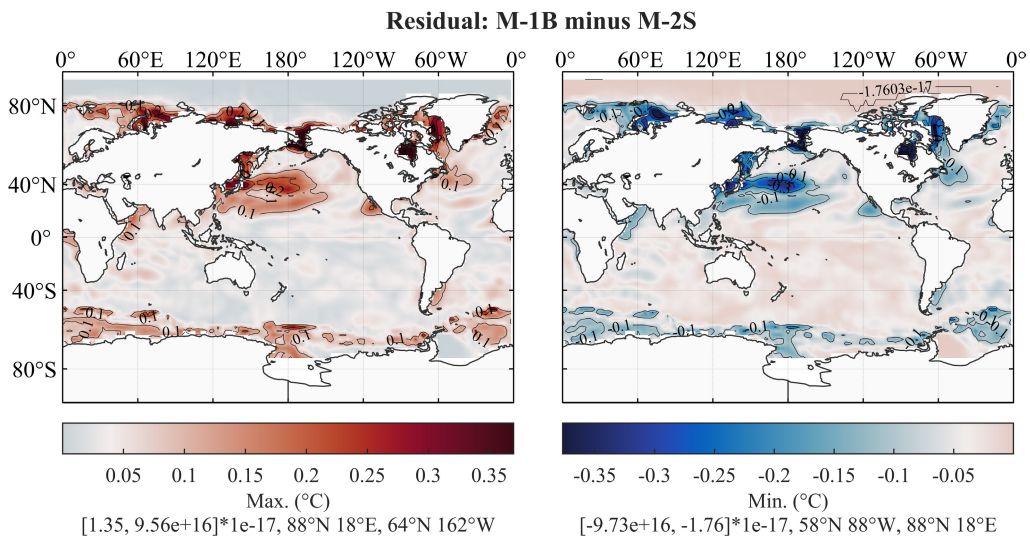


图 43 对 ERSST v5 全球 SST 数据(1902-2021), M-2S 方法与 M-1B 方法输出的 residual 之差.



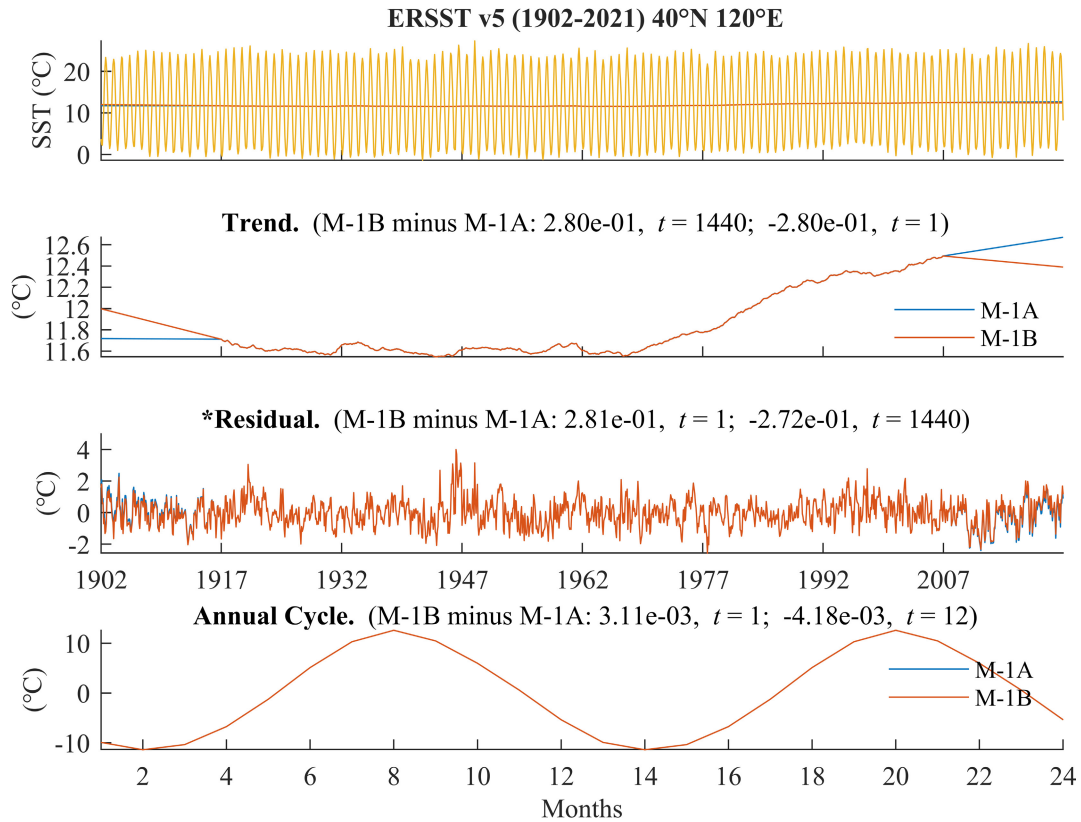


图 44 对 ERSST v5 全球 SST 数据(1902-2021)在 40°N, 120°E, M-1A 方法与 M-1B 方法的经典分解结果.

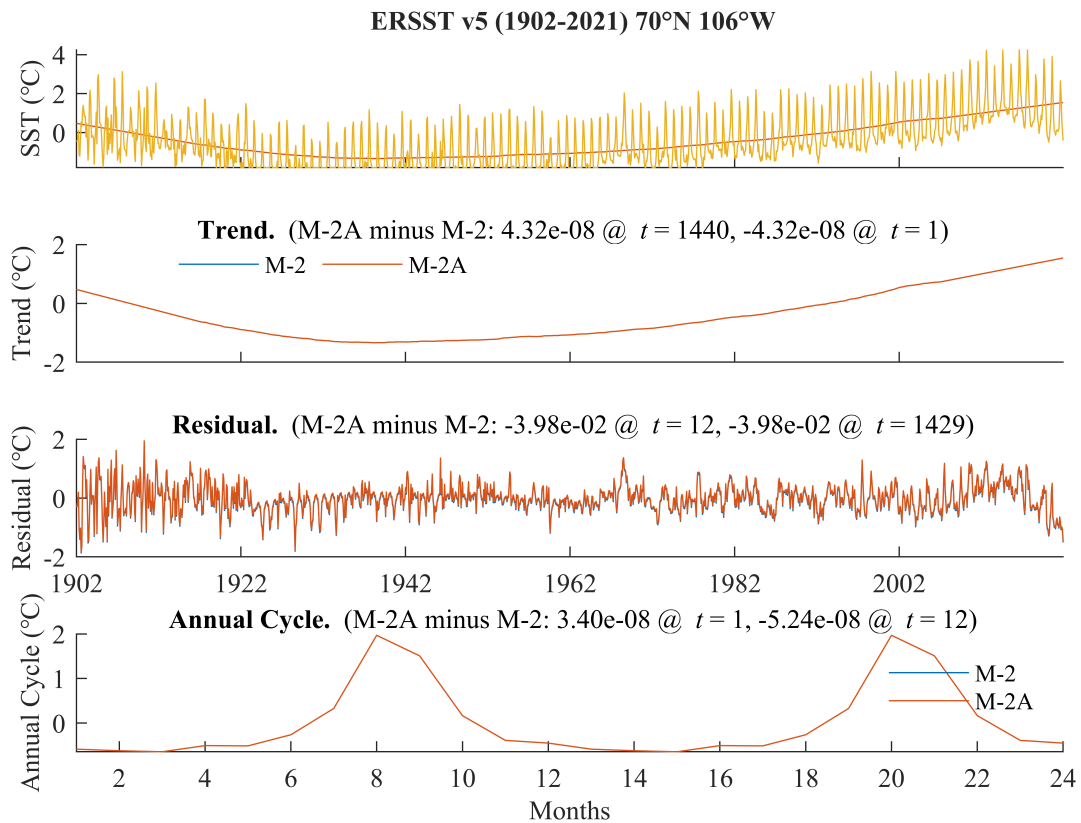


图 45 对 ERSST v5 全球 SST 数据(1902-2021)在 70°N, 106°W, M-2 方法与 M-2A 方法的经典分解结果.

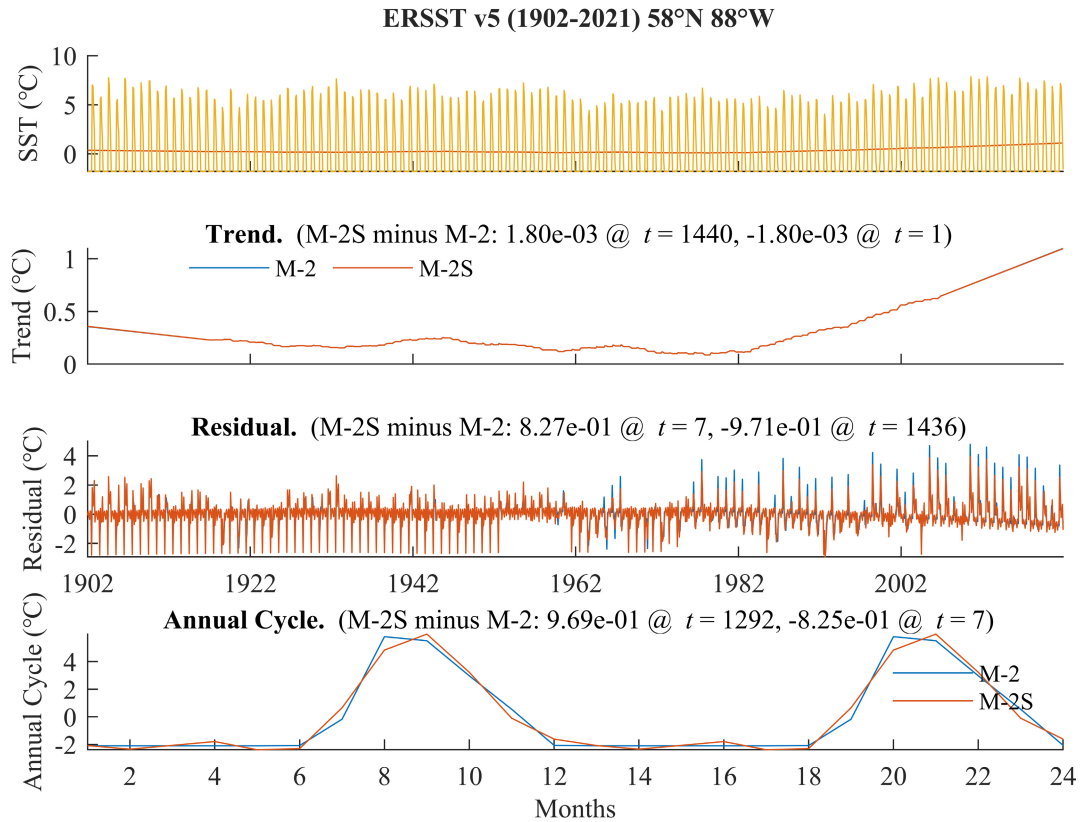


图 46 对 ERSST v5 全球 SST 数据(1902-2021)在 58°N, 88°W, M-2S 方法与 M-2 方法的经典分解结果.

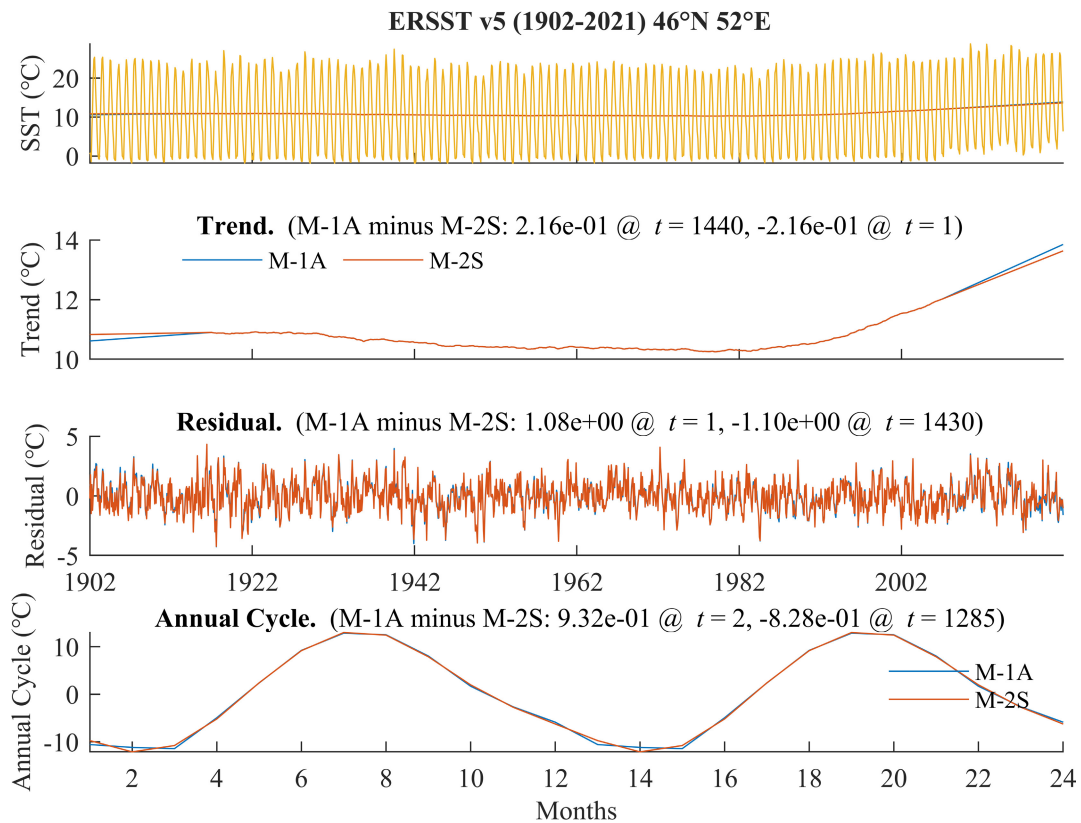


图 47 对 ERSST v5 全球 SST 数据(1902-2021)在 46°N, 52°E, M-2S 方法与 M-1A 方法的经典分解结果.



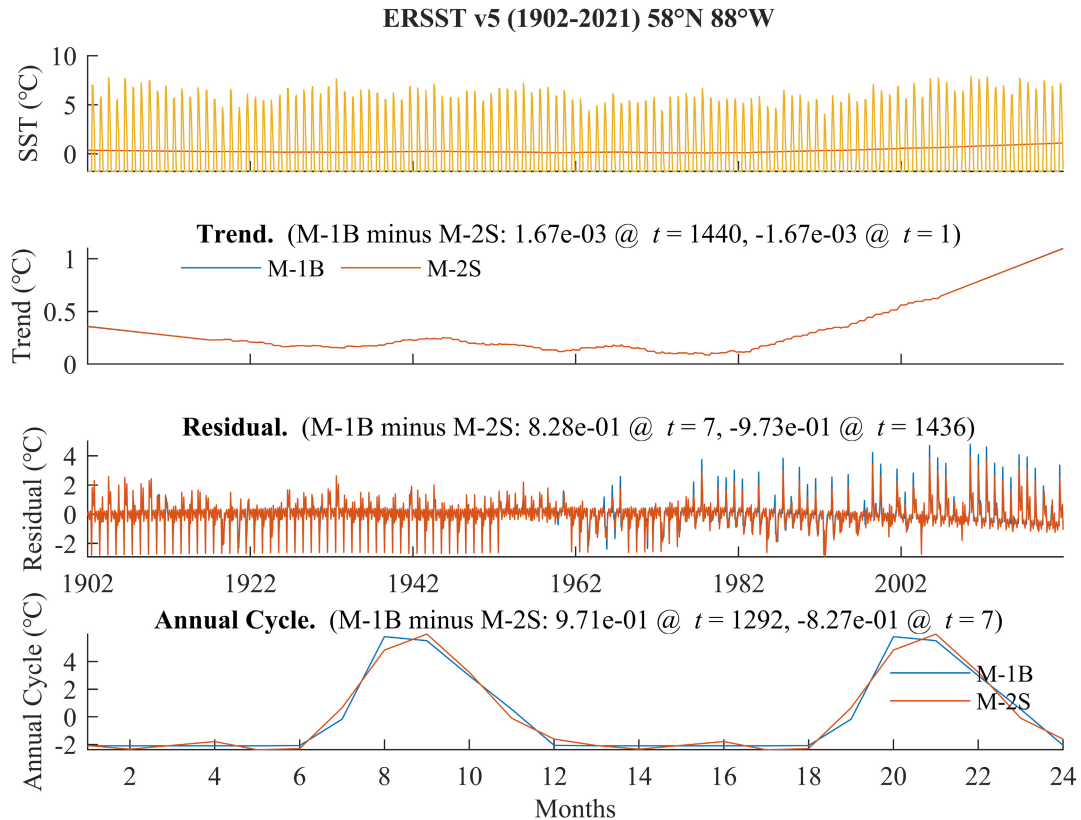


图 48 对 ERSST v5 全球 SST 数据(1902-2021)在 58°N, 88°W, M-2S 方法与 M-1B 方法的经典分解结果.

### 3.2.2 Comparison of different methods

M-1B 与 M-1A 的差异. 由于这两种分解方式的区别, 仅在于 trend 与 seasonal component 估计和去除次序的不同, 下文称这种区别导致的分解结果差异为**交换误差(swap error)**. 在 ERSST 算例中, 输出 residual 差异的最值分布 (图 31) 表明, 差异较大区域出现在北半球中纬度地区, 最大可达约 0.28°C (40°N, 120°E, 图 44); 其他多数区域, 这个差值不超过 0.1°C. 从图 44 可进一步看出, M-1B 与 M-1A 的差异主要出现在时间序列的首尾两端, 最值出现在端点处, 其原因如下: a) **LWLR 的边缘效应**. 本文采取的 LWLR 区间长度是 30 年, 即时间序列首、尾 15 年的 trend 是由以首、尾 15 年时刻为中心、30 年时间区间上的原序列值 (M-1A) 或 deseason 后的值 (M-1B) 通过线性拟合决定的, 而两种方法的拟合结果的差异在 LWLR 区间中心附近是很小的 (这便是为什么除在首尾 15 年外的时刻, M-1A 和 M-1B 的输出差异很小); 而越靠近时间序列的两端点, 就越远离 LWLR 区间中心, 从而 trend 的差别就越大. b) **季节信号对线性拟合的影响**. M-1A 在实施 LWLR 时, 时间序列中仍含有季节信号, 而 M-1B 没有. 季节信号虽然是长期零均值的, 但其去除与否仍将对线性拟合结果产生影响, 当季节周期的解释方差较大时, 这影响有可能相对大. 这影响通过 LWLR 的边缘效应获得放大, 造成在时间序列的首尾两端, M-1B 与 M-1A 的输出差异变大. 还观察到, 两种方法输出差异较大 (小) 的地方 (图 31), 季节信号的幅度往往也较大 (小) (图 50). 上述结果提示: **季节信号对线性回归的影响及其被 LWLR 的边缘效应放大, 是造成 M-1A 和 M-1B 输出差异的重要原因**. 另外, 输出季节信号差异本身的最大差异仅为 0.014°C (46°N, 62°W). c) **余项均值差异**. M-1A 的输出余项是零均值的 (图 26), M-1B 的输出余项均值的绝对值不超过 0.04°C (图 29), 对两种方法输出差异的贡献较小.



M-1A 与 M-2 的差异. 在 ERSST 算例中, M-1A 对 M-2 的输出 residual 差异的最值分布 (图 32) 类似于对 M-1B 的 (图 31); 这是因为, M-1B 对 M-2 的输出 residual 差异较小 (图 33), 最大仅约  $0.013^{\circ}\text{C}$  ( $12^{\circ}\text{S}$ ,  $14^{\circ}\text{E}$ ), 其可能的原因将在后面分析.

M-2S 与 M-2 的差异. 在 ERSST 算例中, M-2S 对 M-2 的输出 residual 差异较大 (图 34), 空间分布类似于 M-1B 和 M-1A 的输出 residual 差异 (图 31), 差异较大区域同样出现在北半球两个西边界暖流的延伸区, 但差值更大, 最大可达约  $0.83^{\circ}\text{C}$  ( $58^{\circ}\text{N}$ ,  $88^{\circ}\text{W}$ ). 还注意到, M-2S 方法提取的 seasonal components 与 M-2 的差异的最值分布 (图 35) 与 M-2S 方法与 M-2 方法输出的 residual 差异的最值分布 (图 34) 类似. 考察 M-2S 与 M-2 的输出 residual 差异最大处的经典分解结果 ( $58^{\circ}\text{N}$ ,  $88^{\circ}\text{W}$ , 图 46), 发现两个方法的输出 trend 差异很小 ( $\sim 10^{-3}^{\circ}\text{C}$ ), 而输出季节信号的差异的最值与输出 residual 的差异的最值同量级 ( $\sim 1^{\circ}\text{C}$ ). 这些观察提示, **因季节周期定义的不同导致的输出季节信号差异, 是造成同属预估-校正类方法的 M-2 和 M-2S 的经典分解结果差异的主要原因.**

M-2 与 M-1A, M-1B 的差异. 在 ERSST 算例中, M-2 的输出 residual 非常接近 M-1B 的 (图 33), 差异最大约  $0.013^{\circ}\text{C}$  ( $12^{\circ}\text{S}$ ,  $14^{\circ}\text{E}$ ); 但远离 M-1A 的 (图 32), 差异最大约  $0.28^{\circ}\text{C}$  ( $40^{\circ}\text{N}$ ,  $120^{\circ}\text{E}$ ). 这个观察提示: **相比于 M-1A, M-1B 可取得更接近预估-校正方法的效果; 或者说, 以 M-1B 为参考, 本文采取的 M-2 系列方法的校正收益不明显.**

M-2A 与 M-1A, M-1B, M-2 的差异. 上面的结果启发我们, 若在 M-2 类中, 交换 trend 和 season 的估计次序, 成为 M-2A 方法, 是否将使输出接近 M-1A? 在 ERSST 算例中, M-2A 的输出 residual 仍非常接近 M-1B 的 (图 37), 差异最大仅约  $0.025^{\circ}\text{C}$  ( $64^{\circ}\text{N}$ ,  $162^{\circ}\text{W}$ ); 但仍远离 M-1A 的 (图 36), 差异最大约  $0.28^{\circ}\text{C}$  ( $40^{\circ}\text{N}$ ,  $120^{\circ}\text{E}$ ). 与此一致的是, M-2A 的输出 residual 对 M-2 的差异较小 (图 38), 差值最大约  $0.04^{\circ}\text{C}$  ( $70^{\circ}\text{N}$ ,  $106^{\circ}\text{W}$ ). 上述试验结果表明: 与 M-1 系列相比, 在 M-2 系列中, 交换误差被抑制. 这个结果提示: **与 M-1 系列方法相比, 采取 M-2 系列方法, 可缓解由于对气候变率特别是长期趋势的物理理解的争议而引起的方案选择问题. 从这个角度, 若要采取 M-1 方法, 似乎应建议采取分解结果更接近 M-2 系列的 M-1B.**

M-1B, M-2, M-2A 三种方法的输出接近, 但相对远离 M-1A 的可能原因是什么? 前面在讨论 M-1A 和 M-1B 的差异时, 我们已经指出, M-1B 对 trend 的估计量是从去除季节信号的时间序列中通过 LWLR 得到的, 这与 M-1A 从带有季节信号的时间序列中估计 trend 的做法不同; 季节信号对线性回归的影响及其被 LWLR 的边缘效应放大, 是造成 M-1A 和 M-1B 输出差异的重要原因. 类似地, 同属于预估-校正类方法的 M-2 和 M-2A, 不论第一步是估计 trend (M-2) 还是估计 seasonal component (M-2A), 经过一步校正后, trend 的估计量都是从已去除季节信号的时间序列中通过 LWLR 得出的, 这与 M-1A 不同. 从这个角度, **M-1B, M-2, M-2A 三种方法的输出接近, 但相对远离 M-1A 的观察, 是符合预期的, 可用季节信号对 LWLR 的影响来解释.**

进一步考察导致 M-2 和 M-2A 差异的原因. 在 ERSST 算例中, M-2 和 M-2A 的输出 seasonal component 差异一般不超过  $10^{-6}^{\circ}\text{C}$  (图 39); M-2A 可保持 residual 样本均值 (图 41), 而 M-2 不可 (图 40), 且 M-2 的 residual 样本均值空间分布 (图 40) 与 M-2 和 M-2A 的输出 residual 差异的最值空间分布 (图 38) 一致. 再考察 M-2 和 M-2A 的输出 residual 的差异最大位置 ( $70^{\circ}\text{N}$ ,  $106^{\circ}\text{W}$ , 图 45) 上的经典分解结果, 发现输出的 trend 和 seasonal component 的差别很小 ( $\sim 10^{-8}^{\circ}\text{C}$ ), 而输出 residual 分量的差别近乎为一常数 ( $0.040^{\circ}\text{C}$ ). 这些结果说明, **提取的 residual 的样本均值非零, 是造成 M-2 和 M-2A 输出差异的主要原因; 因 residual 的样本均值很小, 故 M-2 与 M-2A 的输出差异很小; 若不计这种原因导致的差异, 则 M-2 和 M-2A 方法的输出是一致的. 这个结果支持了上一段的观点.**



对 M-1A, M-1B, M-2, M-2A 输出差异再讨论. 在对理想试验的分析中, 我们指出: 导致这四种方法间差异的主要原因有 a) **剧烈的长期趋势**; b) **强非线性的长期趋势**; c) **大的余项(内部气候变率)方差**; d) **大幅度的季节信号**. 在 ERSST 算例中, **猜想**: 在中纬度地区, 因素 d 起主导作用, 各方法(主要是 M-1A 与其余三种)的输出余项差异可接近  $0.3^{\circ}\text{C}$  (图 31); 在赤道东太平洋, 因素 c 起一定作用, 各方法的输出余项差异可超过  $0.1^{\circ}\text{C}$ ; 在北冰洋部分区域, 因素 a 起的作用相对大, 影响输出季节信号和余项均值, 但对输出余项影响不大; 在其余大部分地区, 各因素作用较小, 各方法输出余项差异一般不超过  $0.1^{\circ}\text{C}$ . 为了验证上述结论, 我们观察 ERSST 百年增温(图 50)、季节信号峰峰值(图 50)、以及余项样本方差和均值(图 49), 发现在 M-1A 对 M-1B 的输出余项差异较大的北半球中纬度地区(图 31), 正是季节信号幅度较大的地区(图 50); 在 M-1A 对 M-1B 的输出余项差异超过  $0.1^{\circ}\text{C}$  的赤道东太平洋地区(图 31), 季节信号的幅度较小(图 50), 而余项标准差较大(图 49); 在北冰洋百年增温较剧烈的地区(图 50), M-1A 对 M-1B 的输出季节信号差异(图 30)和 M-1B 的余项均值(图 29)都相对大, 但也不过  $0.01$  和  $0.04^{\circ}\text{C}$ , 对输出余项差异(图 31)似无明显影响. **还可选择一些可疑点画单站图. 但如何定量长期趋势的非线性? (待补充)**

余项均值是如何造成的? 观察 M-2 对 M-1A(M-2A), 以及 M-2A 对 M-1B 的输出余项均值差异最大位置  $70^{\circ}\text{N}, 106^{\circ}\text{W}$  上, M-1A 和 M-2 方法的经典分解差异(图 52). **猜想**: 长期趋势的非线性是造成余项均值非零的主要原因.

M-2S 与 M-1A, M-1B 的差异. 前面的结果启发我们, 既然 M-2 和 M-2A 都接近 M-1B 而远离 M-1A, 那么同属 M-2 系列的 M-2S 方法, 是否存在 M-1 系列近似? 在 ERSST 算例中, M-2S 对 M-1A(图 42), M-1B(图 43)的输出 residual 的差异都较大, 最大约  $1^{\circ}\text{C}$ . 考察 M-2S 对 M-1A 的输出 residual 的最大差异所在位置 ( $46^{\circ}\text{N}, 52^{\circ}\text{E}$ , 图 47) 上的经典分解结果, 可见 M-2S 和 M-1A 的输出季节信号的差异可超过  $0.9^{\circ}\text{C}$ , 这对输出余项的差异有重要贡献. 两个方法的输出 trend 的差异在时间序列的首、尾 15 年以外时刻较小, 在两个端点处取得最值(约  $0.22^{\circ}\text{C}$ ), 故在时间序列两端附近, M-2S 对 M-1A 输出 trend 的差异对 residual 的差异也有贡献. 考察 M-2S 对 M-1B 的输出 residual 的最大差异所在位置 ( $58^{\circ}\text{N}, 88^{\circ}\text{W}$ , 图 48) 上的经典分解结果, 可见 M-2S 对 M-1B 的输出 trend 的差异很小 ( $\sim 10^{-3}\text{C}$ ), 而输出季节信号的差异的最值和输出 residual 差异的最值同量级 ( $\sim 1^{\circ}\text{C}$ ), 说明 M-2S 与 M-1B 的输出 residual 差异主要是由输出季节信号的差异造成的. 同时, M-1B 对 M-2S 的输出 residual 的差异不是零均值的, 提示输出 residual 的样本均值对两种方法的输出 residual 差异亦有少量贡献. 综上, **造成 M-2S 与 M-1A, M-1B 的差异的可能原因包括: a) 因季节周期定义的不同导致的输出季节信号差异, 这差异在冰冻地区可能更大(Narapusetty et al., 2009); b) 季节信号对线性回归的影响及其被 LWLR 的边缘效应放大(M-1B 无此项); c) 提取的 residual 的样本均值非零(M-1A 无此项).**

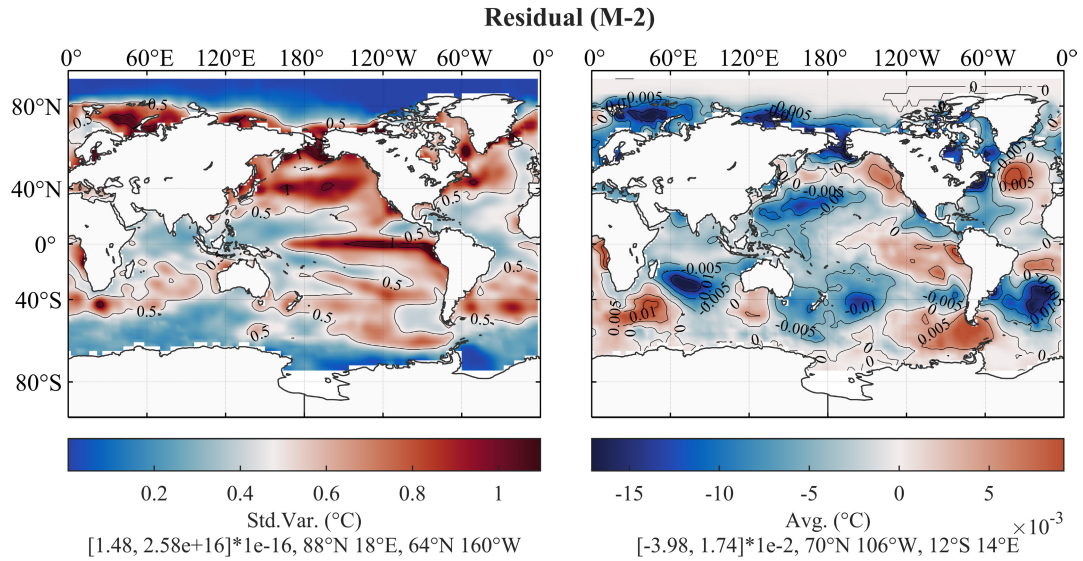


图 49 对 ERSST v5 全球 SST 数据(1902-2021), M-2 方法输出 residual 的样本标准差(a)和均值(b).

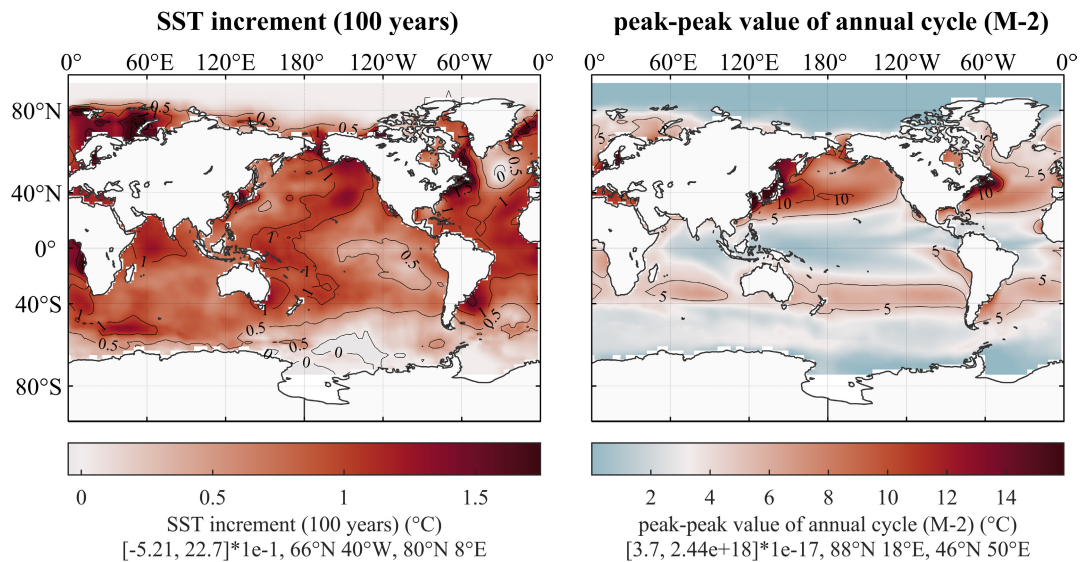


图 50 对 ERSST v5 全球 SST 数据(1902-2021), 原始数据的百年增温(a), 和 M-2 方法输出的季节信号的峰-峰值(b).



M-2: Climatological mean to raw series

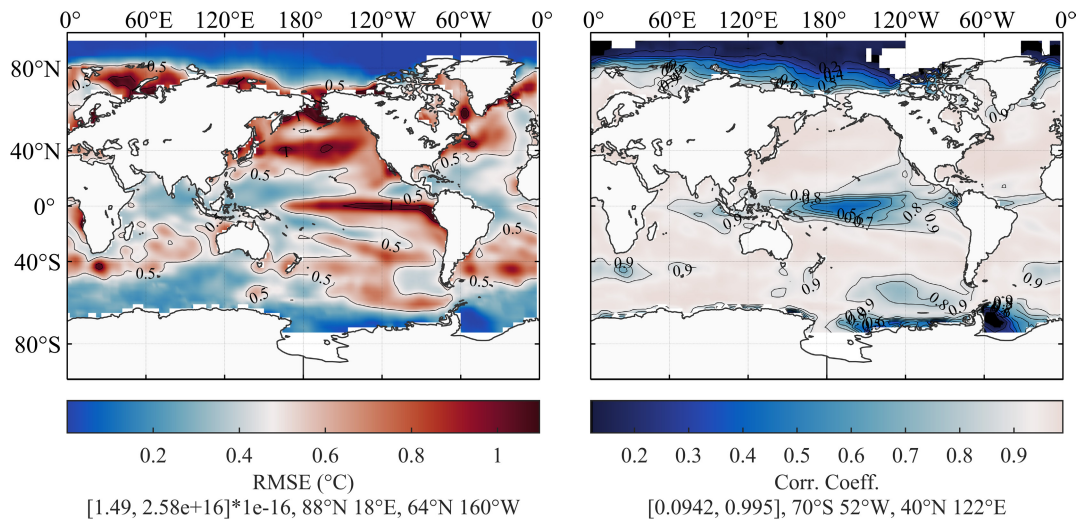


图 51 对 ERSST v5 全球 SST 数据(1902-2021), 原始数据对 M-2 方法输出气候平均态的均方根误差(a)和 Pearson 线性相关系数(b).

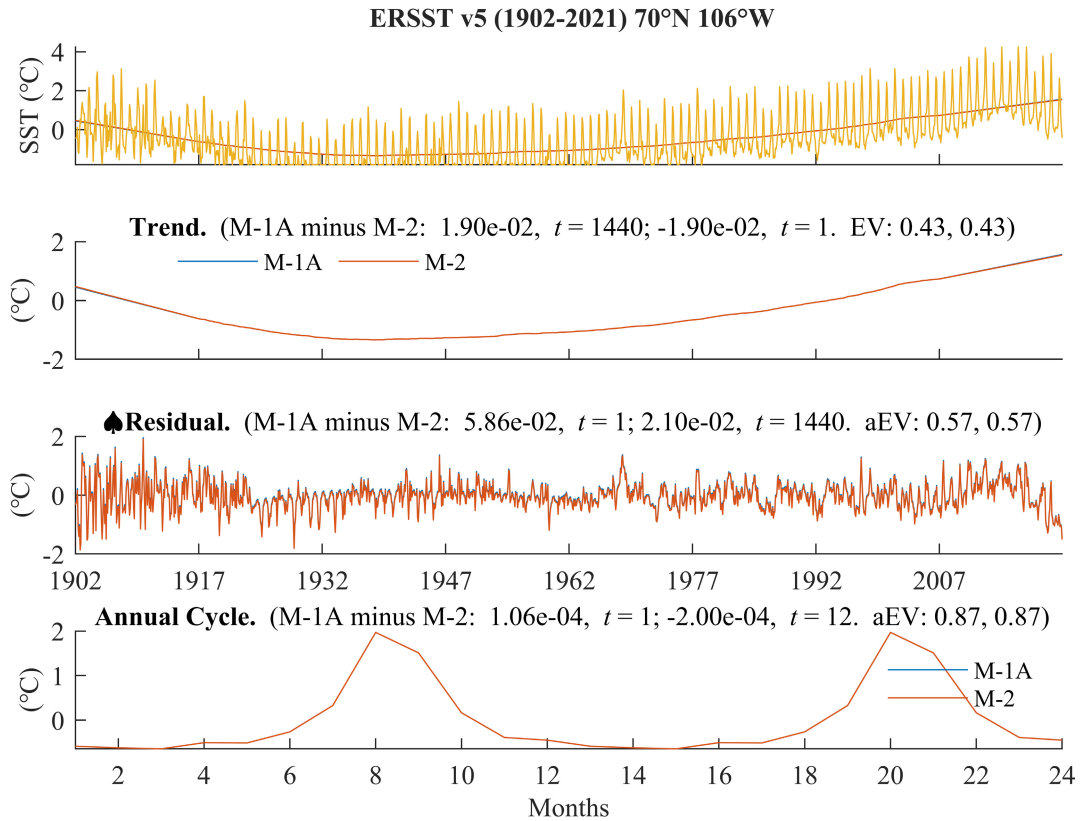


图 52 对 ERSST v5 全球 SST 数据(1902-2021)在 70°N, 106°W, M-1A 方法和 M-2 方法的经典分解结果. 此位置是 M-2 对 M-1A(M-2A), 以及 M-2A 对 M-1B 的输出余项均值差异最大位置.



## 4 Discussion

### 4.1 经典分解的基本假定及合理性

假定季节信号是单音的？（对比：日变化周期）？加性分解？empirical mode decomposition (EMD)？数据驱动的预报？

### 4.2 对时间序列经典分解方案的物理-数学理解

trend 视作气候系统本身的一种特殊内部变率，或是外加的独立强迫，将导致 M-1B, A 两种方法，引起混淆。建议：将物理和序列分析过程分开理解，因为最终都是要做经典分解。

### 4.3 造成各分解方式间差异的因素

trend 的非线性，季节信号的幅度，年际、年代际变率的幅度。M-1A, B 的差异-空间分布，和其他指标的联系。

### 4.4 预估-校正的收益

比较 M-2 系列内部，和横向比较 M-1 系列。

### 4.5 时间序列分解质量评价

因气候系统的非线性决定了总响应不是各输入独立作用的输出之和、要素时间序列采取“随机过程 -> 一次实现”的理解方式、各分量没有数学上严格的客观定义，故难以找到一种完美的、唯一的、纯客观的评价指标。这是分解方式多样的理论根源。本文采取的各指标，仅供参考，不必然表示各方法优劣。

### 4.6 年际及以上变率的数学期望和样本均值

由于观测时间有限，“长期变率可能是更长时间周期信号的一部分”。有些方法，若假定季节周期是零均值的，则可自动保证变率零均值。是否合理？

### 4.7 未来研究建议

本文研究表明，在气候变化的背景下，各分解方式确可能造成差异，有时甚至显著。故建议：后续研究应注意讨论这种差异对结论的影响；在承认非线性的情况下，建立各分量的内秉、客观、物理、数学定义，和配套的客观分解方式；和应用统计、计量经济学等部门合作。





## 5 Conclusions

本文以三类常用的经典分解方案为例,研究了不同的分解方案对结果的影响.结果表明,采取不同的分解方法,对结果会造成一定的影响.在某些情况(趋势强非线性,趋势增量,季节信号的幅度大,余项方差大)下,影响更明显.在全球气候变化的背景下,本文的结果值得学界关注.

基于本文的讨论,我们认为,气候态(包括长期趋势,季节周期分量等)应是客观的,但趋势和季节周期的定义,以及相应的分解方案及其参数的选取却带有主观性,这是分解方案选取问题的根源.因此,我们对后续研究提出以下建议:(1) 后续研究应对气候态各分量建立客观定义,并建立相应的自适应分解方案.(2) 在对气候事件的分类和讨论中,若使用了某种时间序列分解方法,则应增加讨论采取其他分解方法对结果的影响,以避免研究结论被分解方案选取带来的虚假变率误导.



## References

- Chen, X., & Li, T. (2021). An Improved Method for Defining Short-Term Climate Anomalies. *Journal of Meteorological Research*, 35(6), 1012-1022. <https://doi.org/10.1007/s13351-021-1139-2>
- Findley, D. F., Monsell, B. C., Bell, W. R., Otto, M. C., & Chen, B.-C. (1998). New Capabilities and Methods of the X-12-ARIMA Seasonal-Adjustment Program. *Journal of Business & Economic Statistics*, 16(2), 127-152. <https://doi.org/10.1080/07350015.1998.10524743>
- Hua, W., Dai, A., & Qin, M. (2018). Contributions of Internal Variability and External Forcing to the Recent Pacific Decadal Variations. *Geophysical Research Letters*, 45(14), 7084-7092. <https://doi.org/10.1029/2018GL079033>
- Hua, W., Dai, A., & Qin, M. (2022). Reconciling Roles of External Forcing and Internal Variability in Indian Ocean Decadal Variability Since 1920. *Geophysical Research Letters*, 49(9), e2021GL097198. <https://doi.org/10.1029/2021GL097198>
- Narapusetty, B., DelSole, T., & Tippet, M. K. (2009). Optimal Estimation of the Climatological Mean. *Journal of Climate*, 22(18), 4845-4859. <https://doi.org/10.1175/2009jcli2944.1>
- Pezzulli, S., Stephenson, D. B., & Hannachi, A. (2005). The Variability of Seasonality. *Journal of Climate*, 18(1), 71-88. <https://doi.org/10.1175/jcli-3256.1>
- Qin, M., Dai, A., & Hua, W. (2020). Quantifying Contributions of Internal Variability and External Forcing to Atlantic Multidecadal Variability Since 1870. *Geophysical Research Letters*, 47(22), e2020GL089504. <https://doi.org/10.1029/2020GL089504>
- Wu, Z., Huang, N. E., Long, S. R., & Peng, C.-K. (2007). On the trend, detrending, and variability of nonlinear and nonstationary time series. *Proceedings of the National Academy of Sciences*, 104(38), 14889-14894. <https://doi.org/10.1073/pnas.0701020104>