

# Untitled project: Test 0 (pre-experiment)

研-MS8401-44000-M01-海洋环境数据分析 (2022 Spring)

- [课程主页](#)
- [个人主页](#)
- Untitled project: Test 0 (pre-experiment)
  - [Introduction](#)
    - Two ways to define climatology and variability
    - Objectives of this program
  - [Methods](#)
  - [Results](#)
    - Case 1- slow linear trend, without noise
    - Case 2- slow quadratic trend, without noise
    - Case 3- fast linear trend, without noise
    - Case 4- fast quadratic trend, without noise
    - Case 5- slow linear trend, with noise added
  - [Conclusion and discussion](#)
    - About climatology (or season)
    - About variability
  - [Resources](#)
  - [Contact information](#)

## Introduction

The original time series  $y(t)$  is made up of four components:  $y := y_{\text{trend}} + y_{\text{climatology}} + y_{\text{var}} + y_{\text{noise}}$ , where  $y_{\text{climatology}} = y_0 + y_{\text{season}}$ , such that `mean(y_season) = 0`, following the convention in [CDT:: season ::Seasons vs Climatology](#).  $y_{\text{trend}}$  is a polynomial, which is often assumed to be linear.  $y_{\text{climatology}}$  and  $y_{\text{trend}}$  has a one-year cycle.  $y_{\text{var}}$  has a cycle of several years.  $y_{\text{noise}}$  is a high frequency signal and sometimes difficult to resolve.

Generally, there exist two ways to define (and solve) the term *climatology* and *variability*.

## Two ways to define climatology and variability

- **Algorithm B** ("climatology" determined *before* detrending)  
E-mail: [weiguorui@sjtu.edu.cn](mailto:weiguorui@sjtu.edu.cn)

- **Step-1** Determines the "climatology" and "anomaly" directly, without detrending.
  - Note: The term "climatology" is defined as the average values of the *orginal* data for each of the 366 days (or 12 months) of the year.
  - The definition of "climatology" here is different from that in Algorithm A.
  - Note: The term "anomaly" here is defined as the difference between the *original* data and "climatology". The definition of "anomaly" here is different from that in Algorithm A.
  - Note: It is expected that "anomaly" = "trend" + "variability" + "noise".
  - Remark: The above definition leads to the fact that "anomaly" (expected to be "trend" + "variability" + "noise") has an arithmetic mean of 0.
- **Step-2** Determines the "trend" of "anomaly".
  - Note: It is expected that, removing the "trend" from "anomaly" results in "variability" + "noise".
  - Note: We can choose a linear (or other forms of polynomial, of course) least squares trend.
- **Step-3** Determines the "variability" (with "noise", expected) by removing the "trend" of "anomaly".
  - Note[Ref]: By default, anomalies (of "anomaly", here) are calculated after removing the linear least squares trend, but if, for example, warming is strongly nonlinear, you may prefer the 'quadratic' option. Default is 'linear'.
  - Remark: If a linear least squares trend is chosen, the "variability"(with "noise") also has zero arithmetic mean, since "anomaly" is zero arithmetic mean, and the fact that linear regression model pass the mean point of sample.
- Question: What would happen if a non-linear trend is chosen?
- **Algorithm A** ("climatology" determined *after* detrending)
  - **Step-1** Determines the "trend" of the orginal data.
    - Note: It is expected that removing the "trend" from the orginal data results in "climatology" + "variability" + "noise".
  - **Step-2** Determines the "climatology" and "anomaly" ("variability") of the *detrended* data.
    - It is expected that this "anomaly" represents "variability" + "noise".
    - Note: The term "climatology" is defined as the average values of the *detrended* data for each of the 366 days (or 12 months) of the year. The definition of "climatology" here is different from that in Algorithm B.
    - Note: The term "anomaly" here is defined as the difference between the *detrended* data and "climatology". The definition of "anomaly" here is different from that in Algorithm B.
    - Note: It is expected that "anomaly" = "variability" + "noise".
    - Remark: The above definition leads to the fact that "anomaly" (expected to be "variability" + "noise") has an arithmetic mean of 0.

- **Question:** Is the "variability" obtained by Algorithm A the same as that obtained by Algorithm B? If not, what are the factors that determine the difference between them? Can a quantitative conclusion be drawn?

## Objectives of this program

**Scientific problem:** Is the "variability" obtained by Algorithm B, where "climatology" is determined *before* detrending, the same as that obtained by Algorithm A, where "climatology" is determined *after* detrending? If not, what are the factors that determine the difference between them? Can a quantitative conclusion be drawn?

## Methods

Create a time series consisting of four components, i.e.  $y := y_{\text{trend}} + y_{\text{climatology}} + y_{\text{var}} + y_{\text{noise}}$ , as mentioned above, as the input of Algorithm A and B implemented in Matlab R2022b. The output signal is considered as a superposition of three, not four, components, since both Algorithms could not tell high-frequency *noise* from several-year-period *variability*. Each output components is compared with its original counterpart. It is expected that each pairs are almost identical in the sense of differing by a constant. Specially, the output variability should share a common periods and amplitude with the orginal ones. Repeat the experiments under controlled conditions, including 1) degree or coefficient of the trend polynomial, 2) periods or amplitude of variability, 3) periods or amplitude of noise.

## Results

### Case 1- slow linear trend, without noise

- **variability**
  - shape (periods, amplitude):
    - A: good
    - B: period is good, but suffers from small burr
  - difference to original:
    - A: changes through successive and uniform small amplitude (~0.2%) staircases, with a linear trend from -5% to 5%
    - B: relatively large amplitude (~6%) sawtooth wave, with a linear trend from -5% to 5%
- **season (or climatology)**
  - shape (periods, amplitude):
    - A: good

- B: good
- difference to original:
  - A: changes through successive and small-amplitude ( $\sim 0.4\%$ ) sawtooth wave, around zero
  - B: relatively large-amplitude ( $\sim 1\%$ ) sawtooth wave, around zero

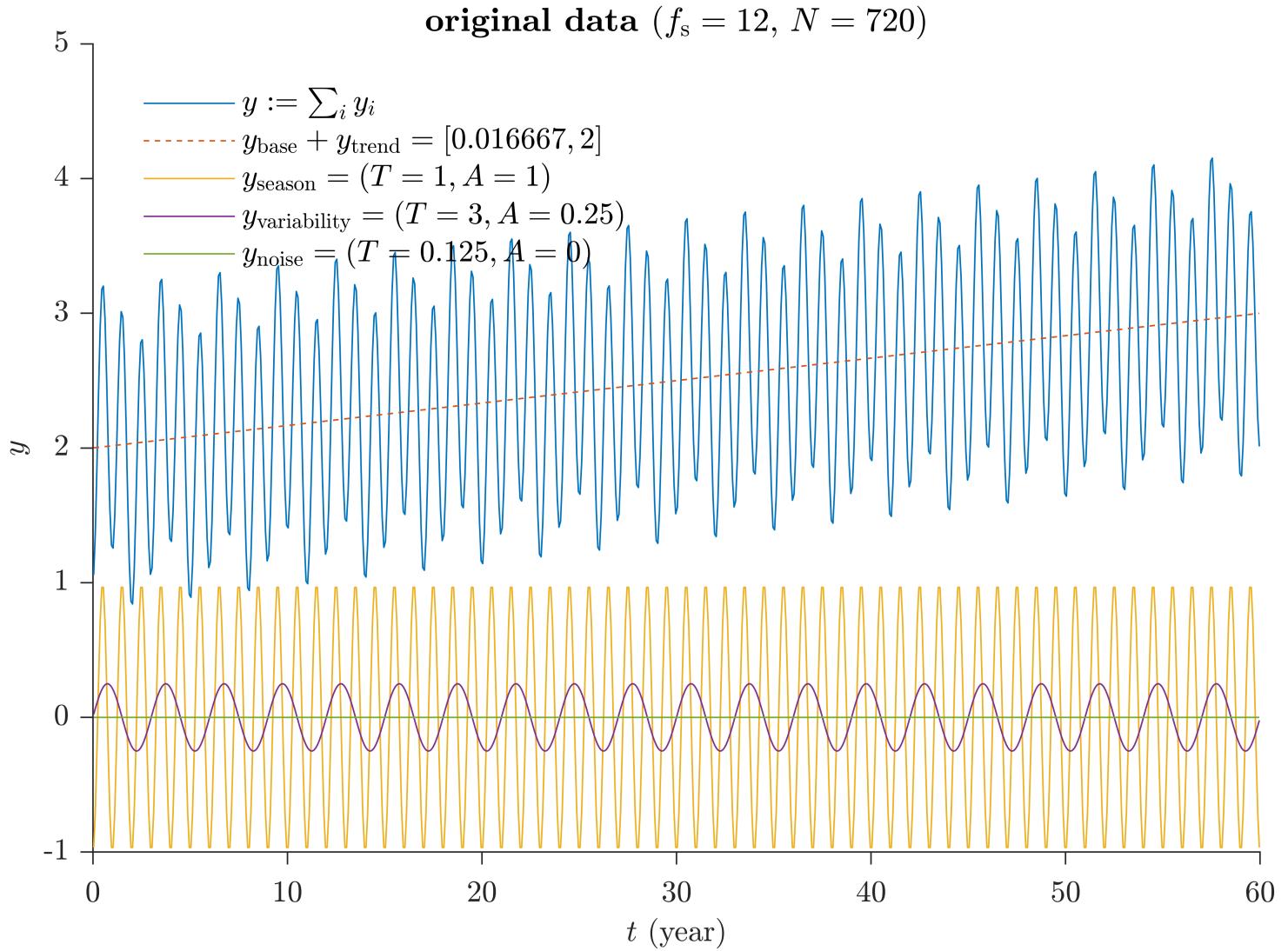


Figure 1.1 original data

### Algorithm A: climatology determined *after* detrending

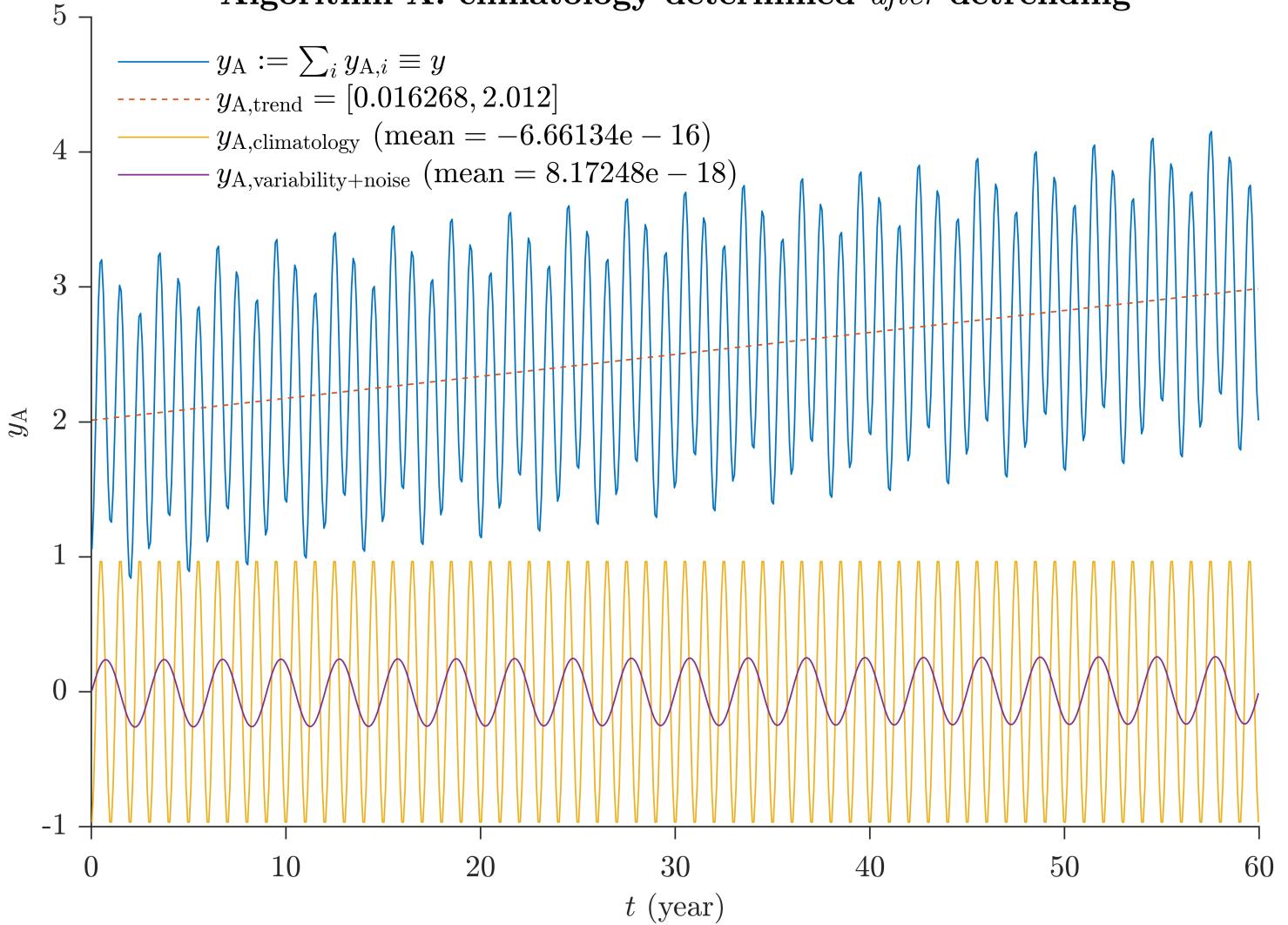


Figure 1.2(a) Algorithm A: "climatology" determined *after* detrending

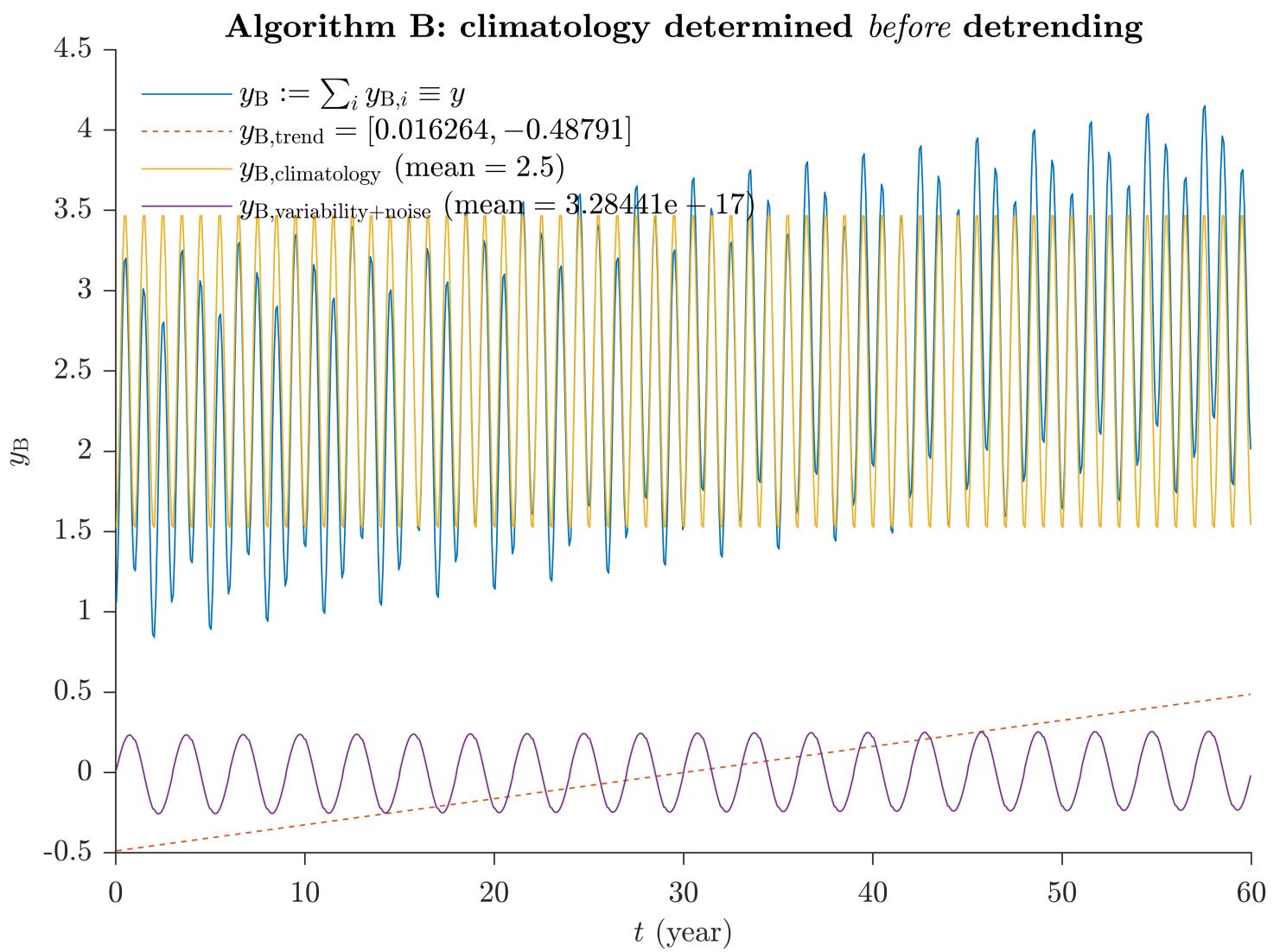


Figure 1.2(b) Algorithm B: "climatology" determined *before* detrending

## difference between output and original

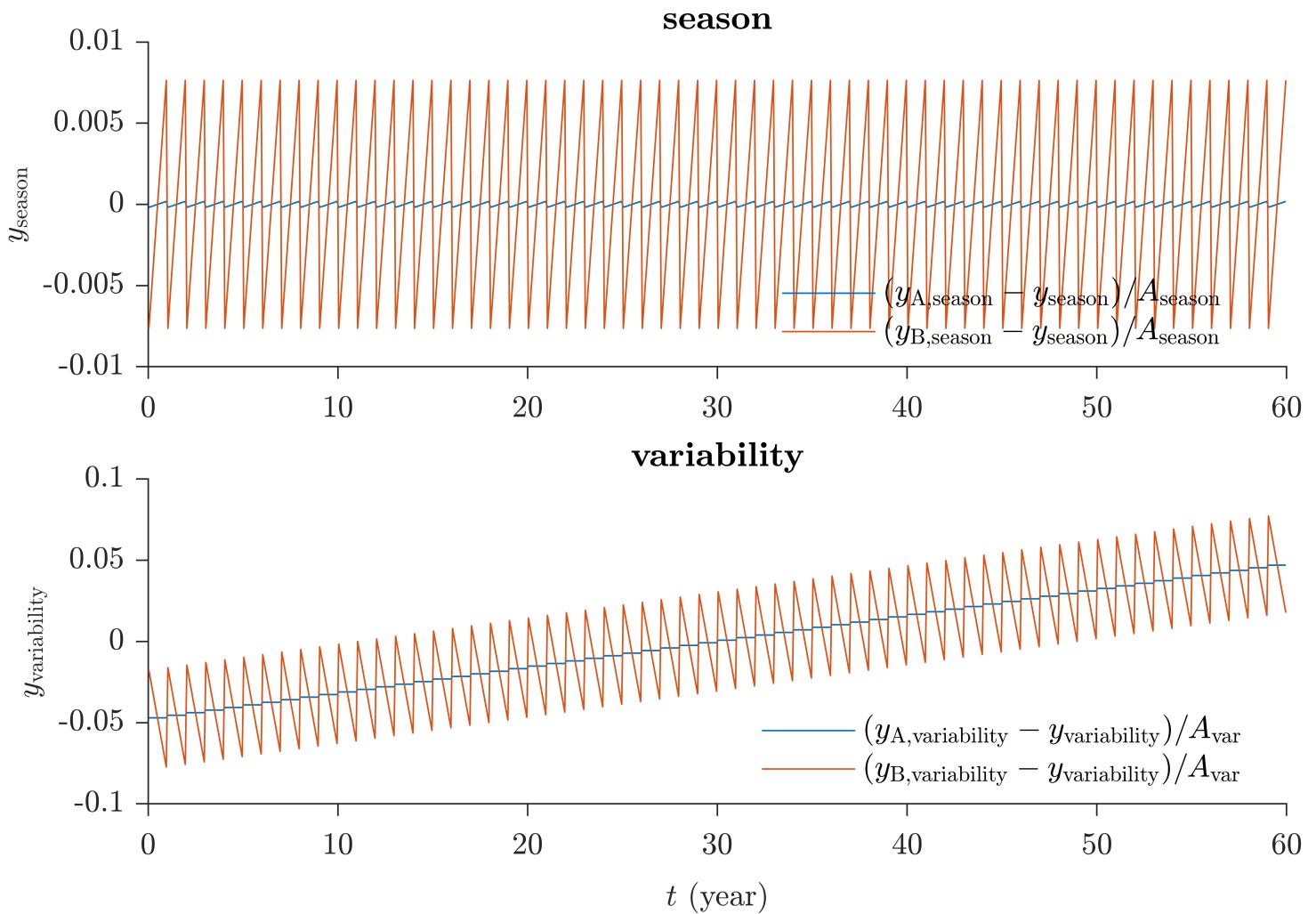


Figure 1.3 difference between output and original

## Case 2- slow quadratic trend, without noise

- **variability**
  - shape (periods, amplitude):
    - A: good
    - B: period is good, but suffers from small burr
  - difference to original:
    - A: changes through successive and uniform small amplitude (~0.2%) staircases, with a linear trend from -5% to 5%
    - B: large amplitude (~15%) sawtooth wave, with a linear trend from -5% to 5%
- **season (or climatology)**
  - shape (periods, amplitude):
    - A: good

- B: good
- difference to original:
  - A: changes through successive and small-amplitude ( $\sim 0.4\%$ ) sawtooth wave, around zero
  - B: large-amplitude ( $\sim 3\%$ ) sawtooth wave, around zero

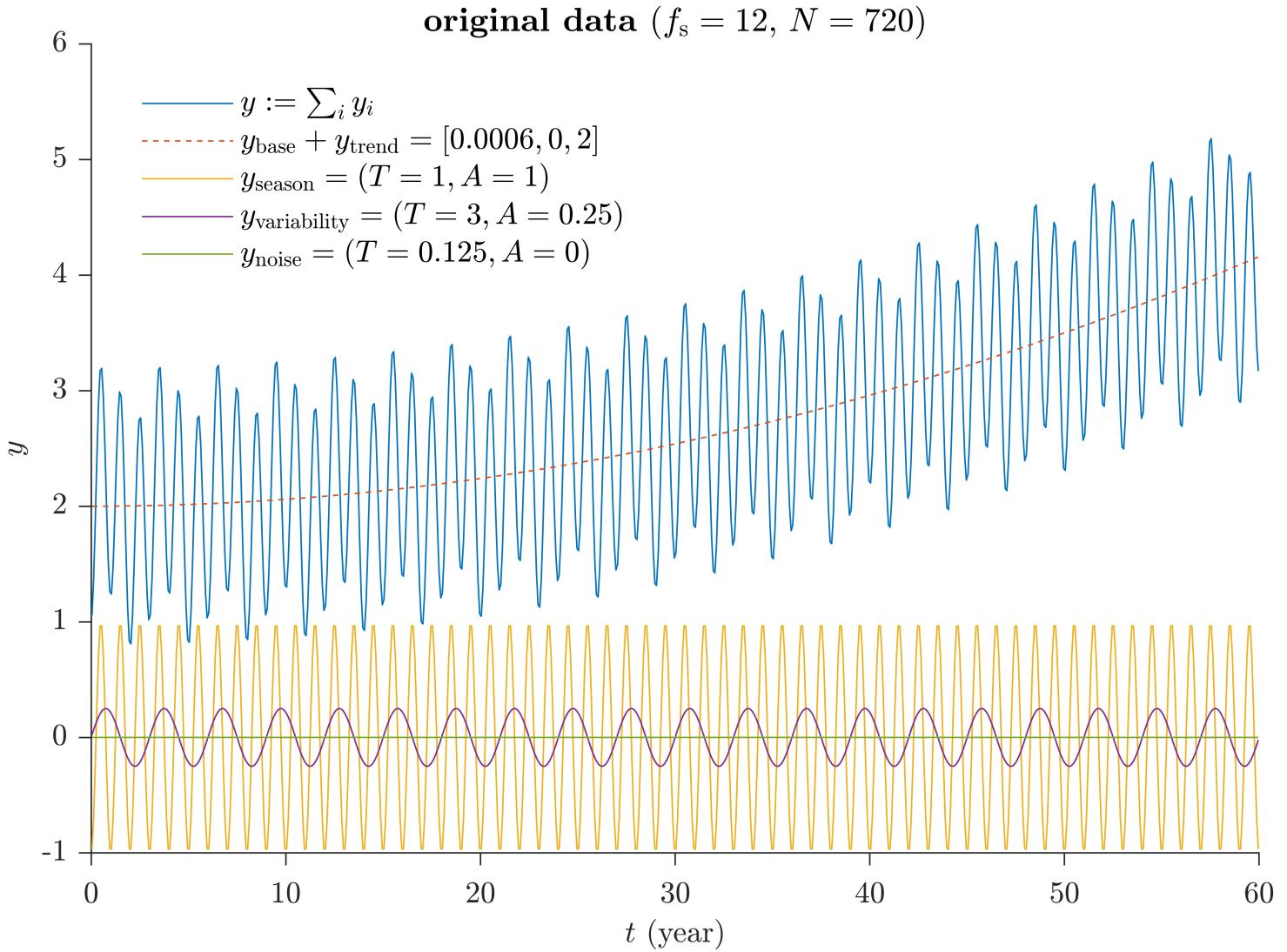


Figure 2.1 original data

### Algorithm A: climatology determined *after* detrending

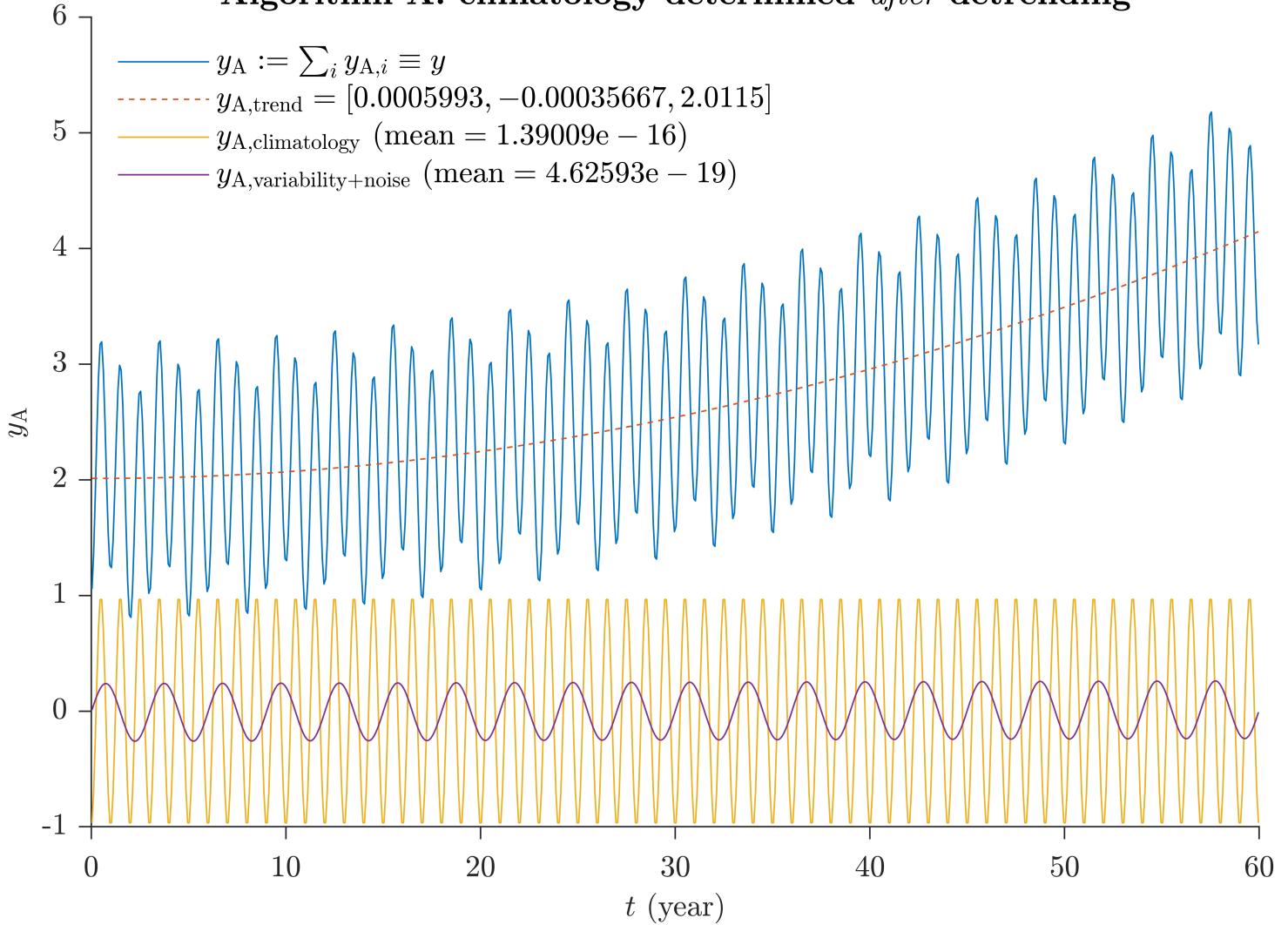


Figure 2.2(a) Algorithm A: "climatology" determined *after* detrending

### Algorithm B: climatology determined *before* detrending

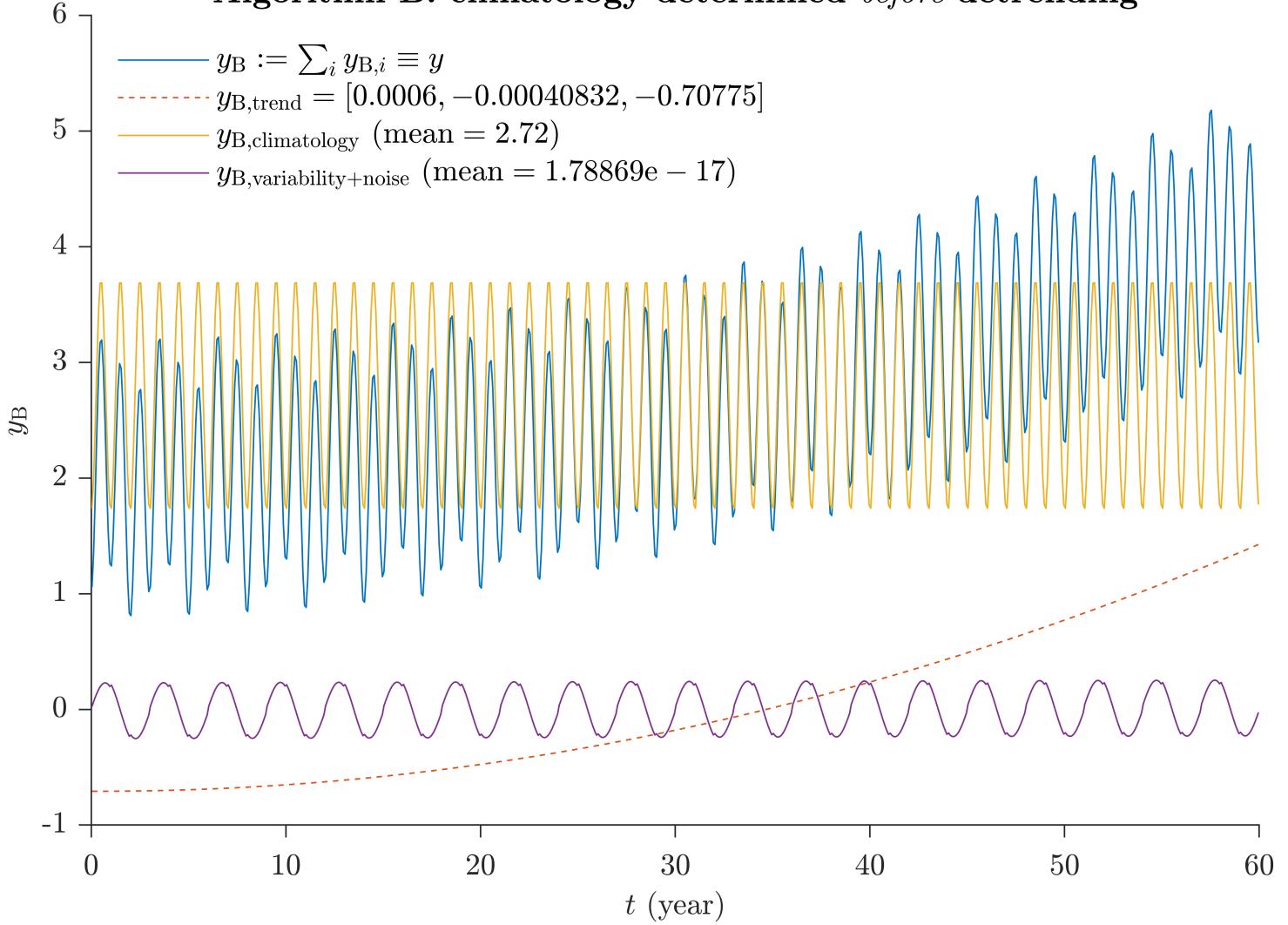


Figure 2.2(b) Algorithm B: "climatology" determined *before* detrending

## difference between output and original

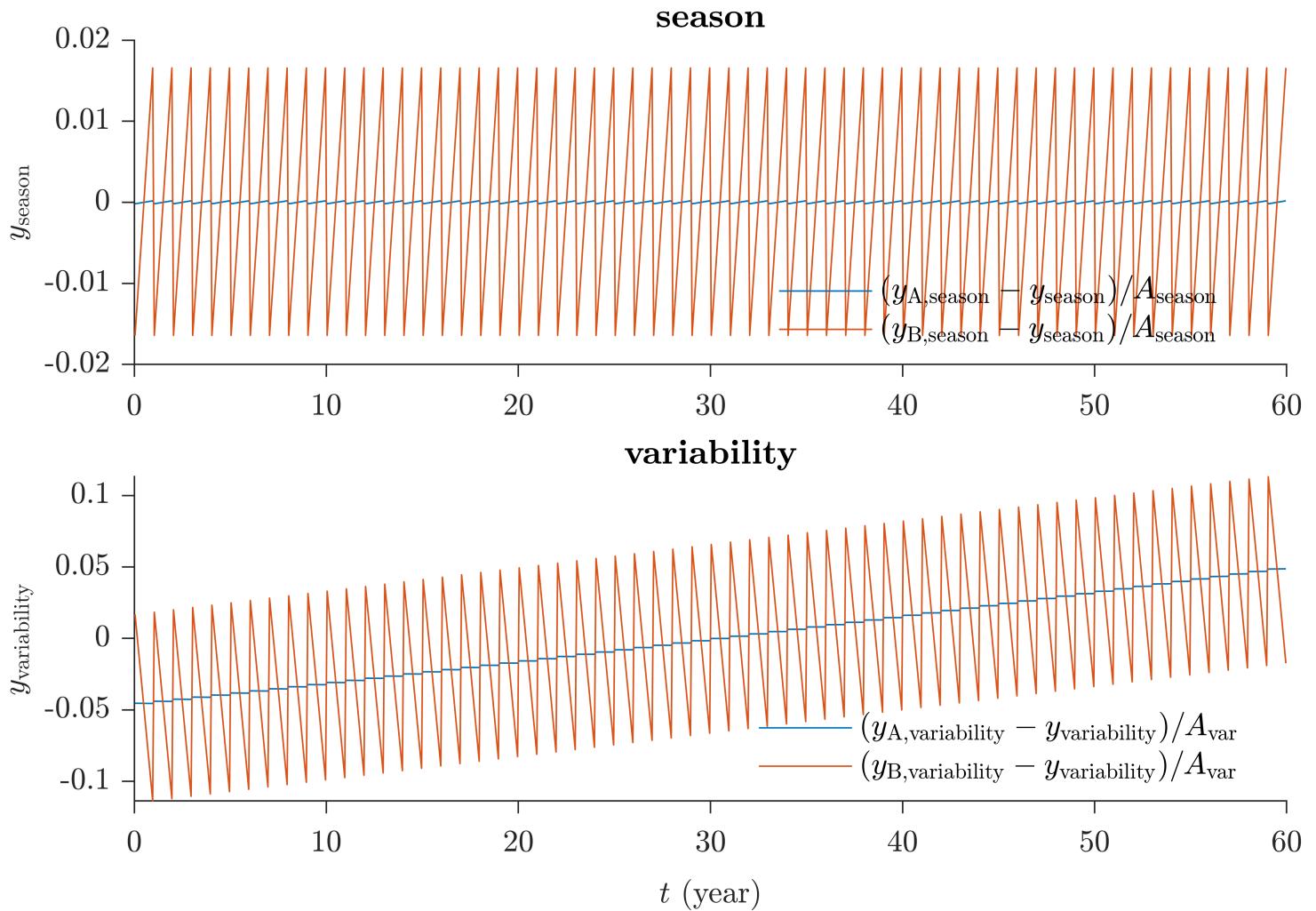


Figure 2.3 difference between output and original

### Case 3- fast linear trend, without noise

- **variability**
  - shape (periods, amplitude):
    - A: good
    - B: period is good, but distorts severely
  - difference to original:
    - A: changes through successive and uniform small amplitude (~0.2%) staircases, with a linear trend from -5% to 5%
    - B: very large amplitude (~90%) sawtooth wave, with a linear trend from -5% to 5%
- **season (or climatology)**
  - shape (periods, amplitude):
    - A: good

- B: good
- difference to original:
  - A: changes through successive and small-amplitude ( $\sim 0.4\%$ ) sawtooth wave, around zero
  - B: large-amplitude ( $\sim 23\%$ ) sawtooth wave, around zero

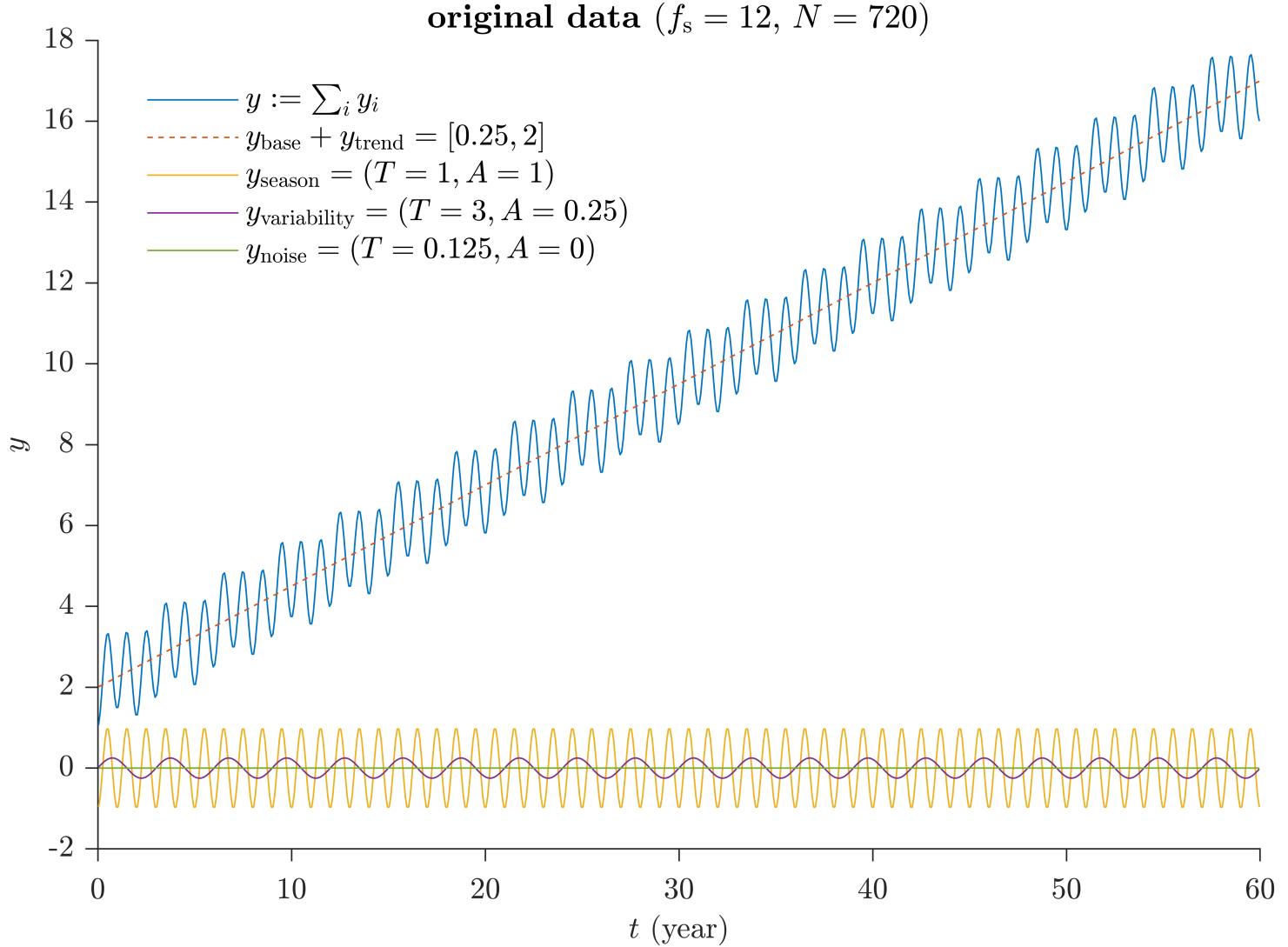


Figure 3.1 original data

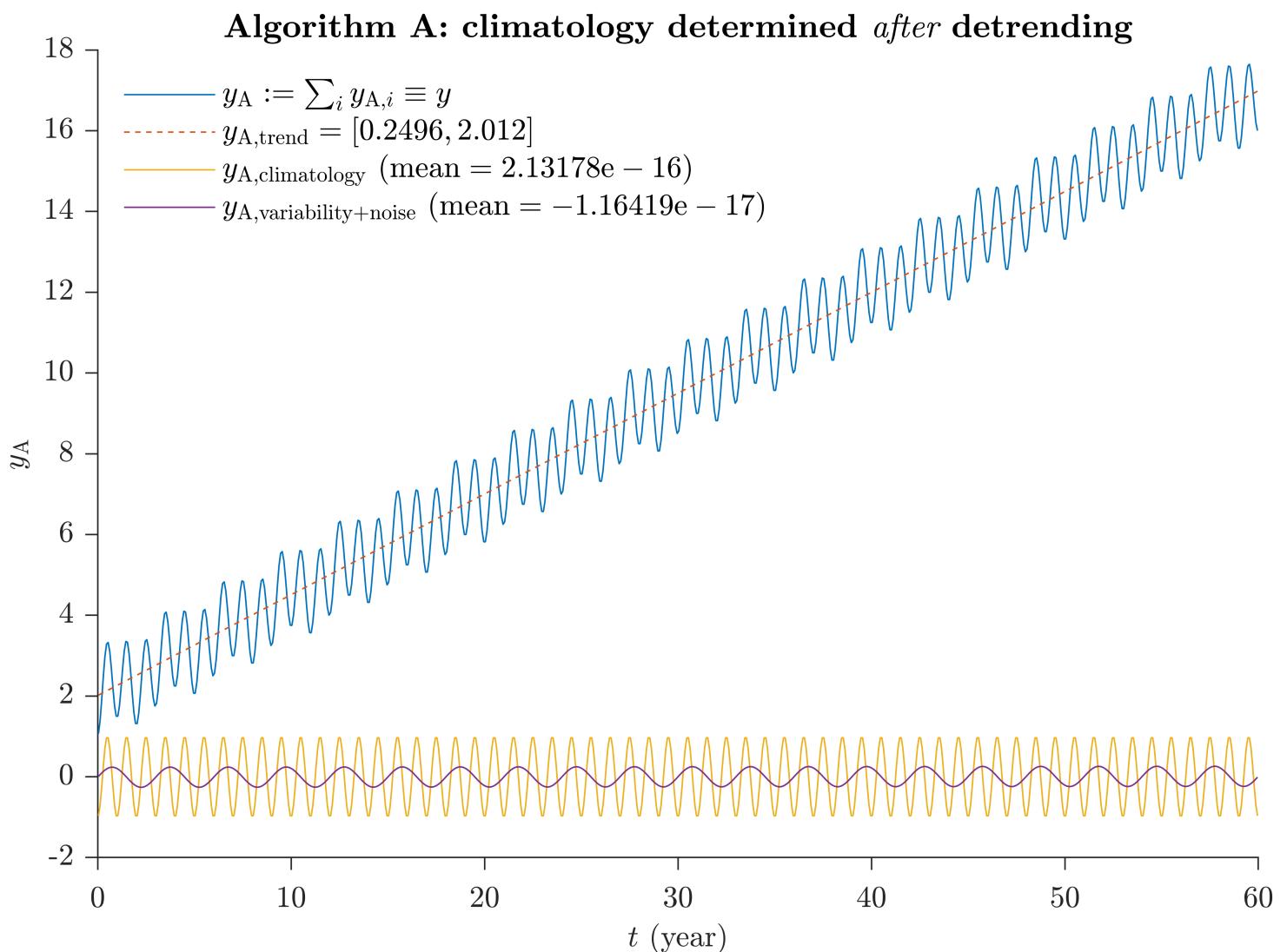


Figure 3.2(a) Algorithm A: "climatology" determined *after* detrending

## Algorithm B: climatology determined *before* detrending

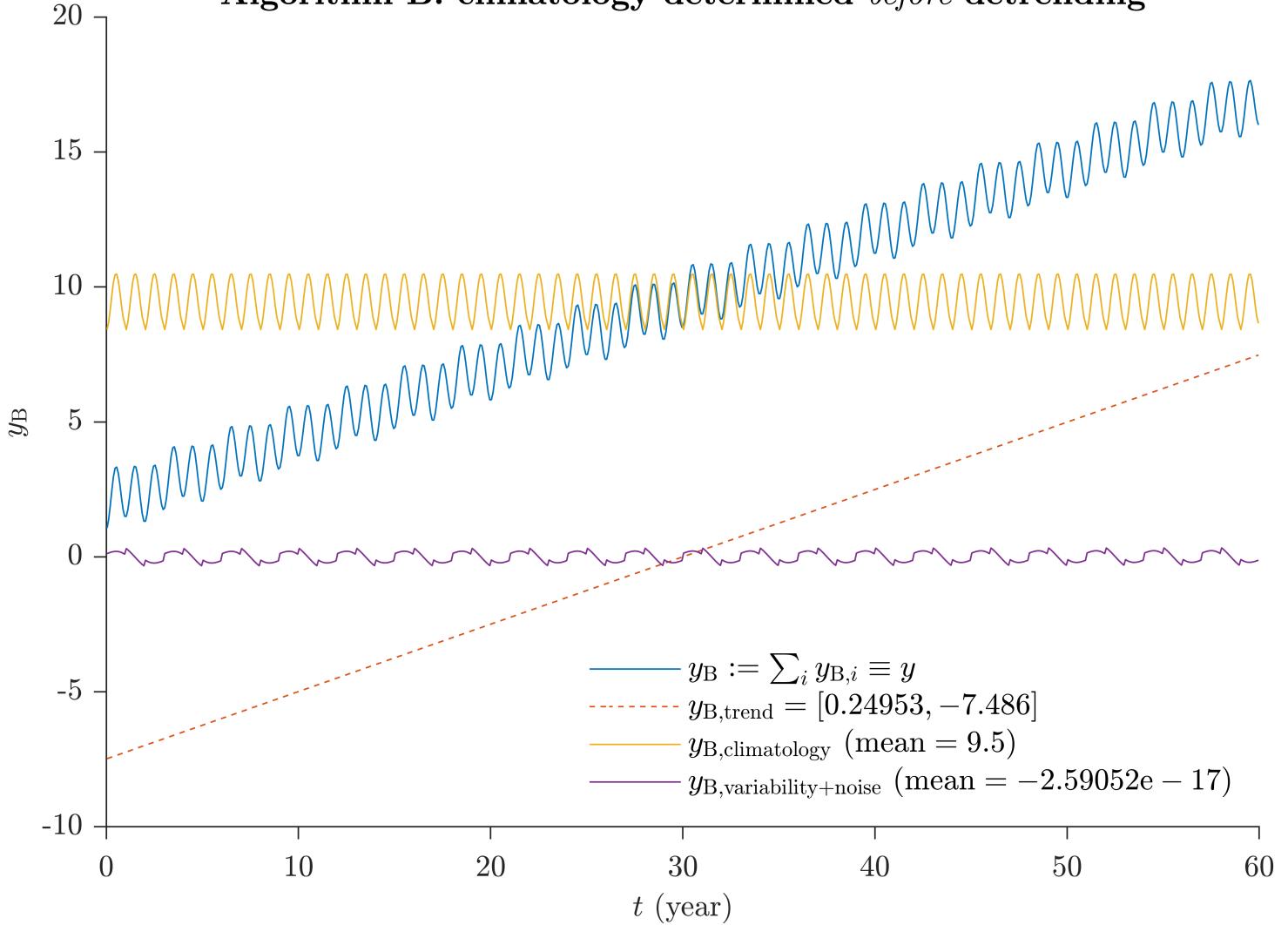


Figure 3.2(b) Algorithm B: "climatology" determined *before* detrending

## difference between output and original

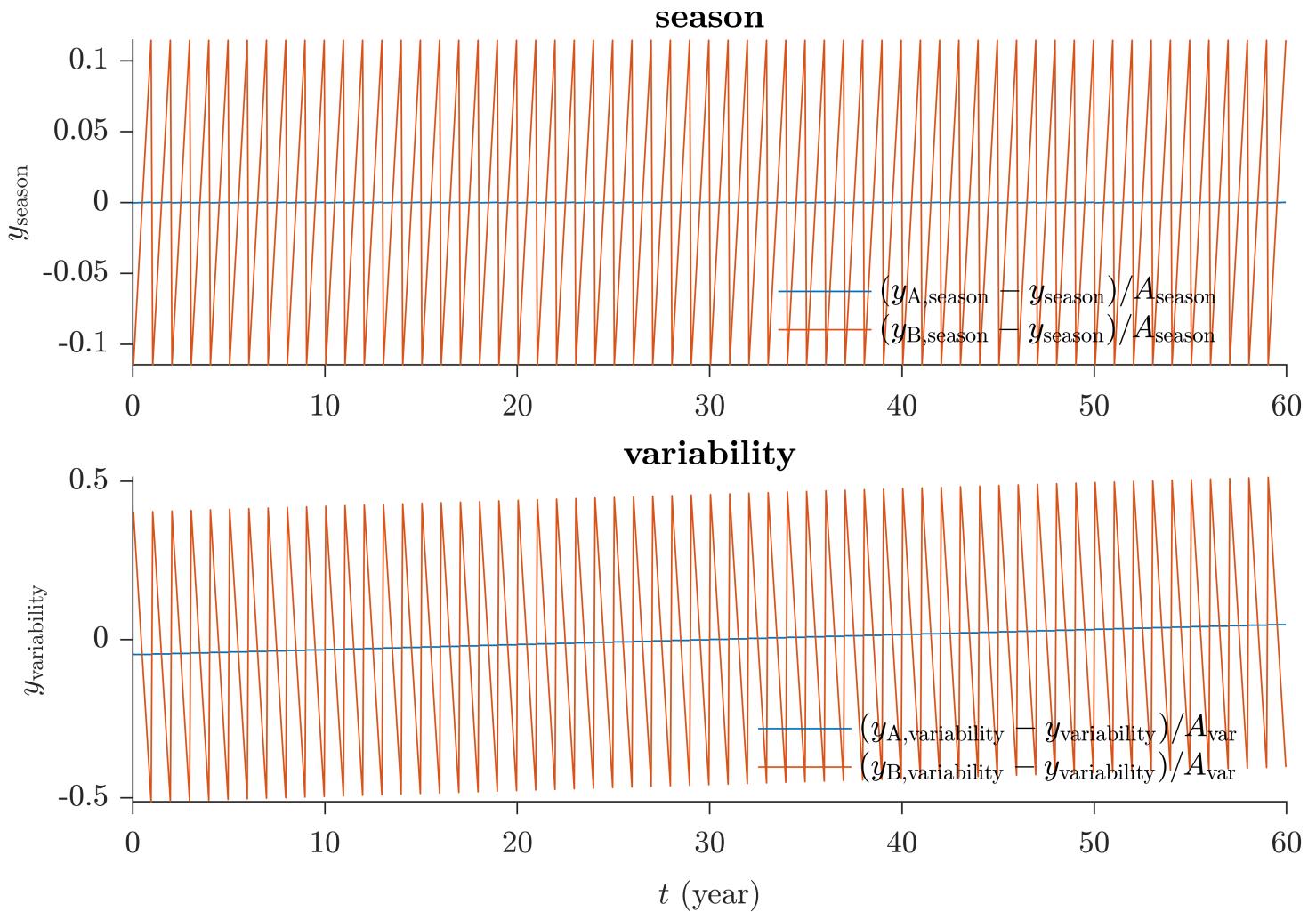


Figure 3.3 difference between output and original

## Case 4- fast quadratic trend, without noise

- **variability**
  - shape (periods, amplitude):
    - A: good
    - B: period is good, but distorts severely
  - difference to original:
    - A: changes through successive and uniform small amplitude (~0.2%) staircases, with a linear trend from -5% to 5%
    - B: huge amplitude (~110%) sawtooth wave, with a linear trend from -5% to 5%
- **season (or climatology)**
  - shape (periods, amplitude):
    - A: good

- B: good
- difference to original:
  - A: changes through successive and small-amplitude ( $\sim 0.4\%$ ) sawtooth wave, around zero
  - B: large-amplitude ( $\sim 28\%$ ) sawtooth wave, around zero

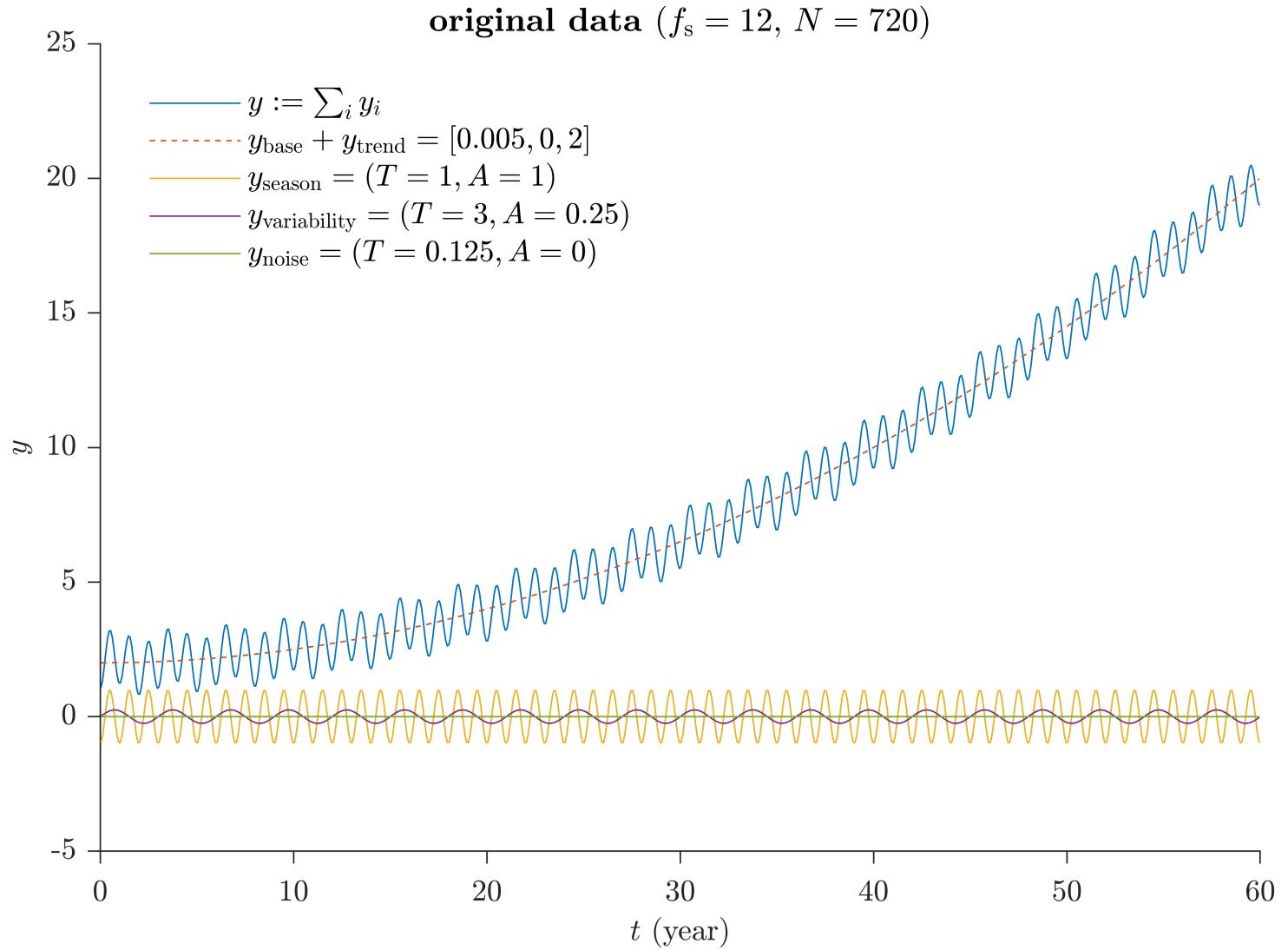


Figure 4.1 original data

### Algorithm A: climatology determined *after* detrending

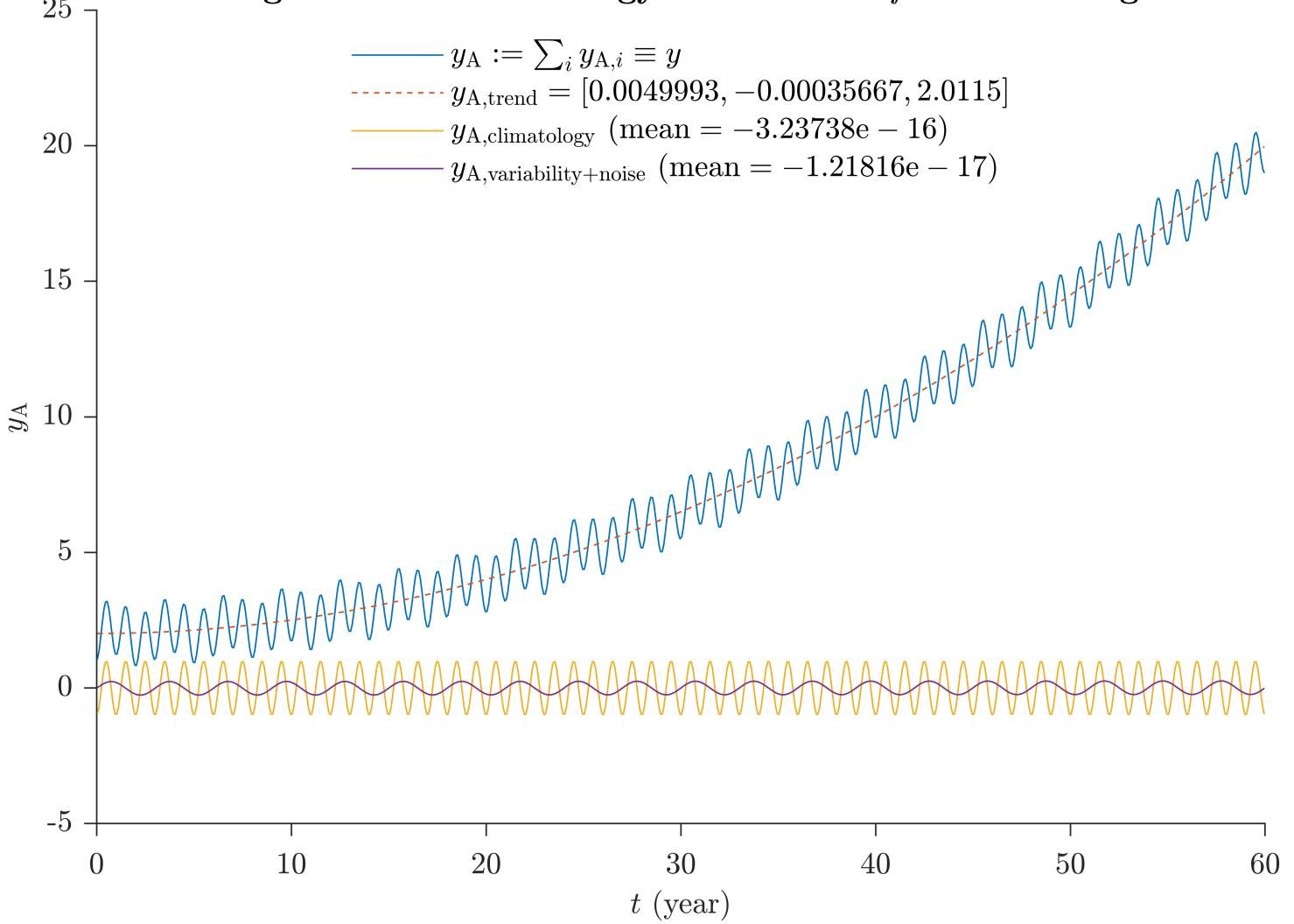


Figure 4.2(a) Algorithm A: "climatology" determined *after* detrending

### Algorithm B: climatology determined *before* detrending

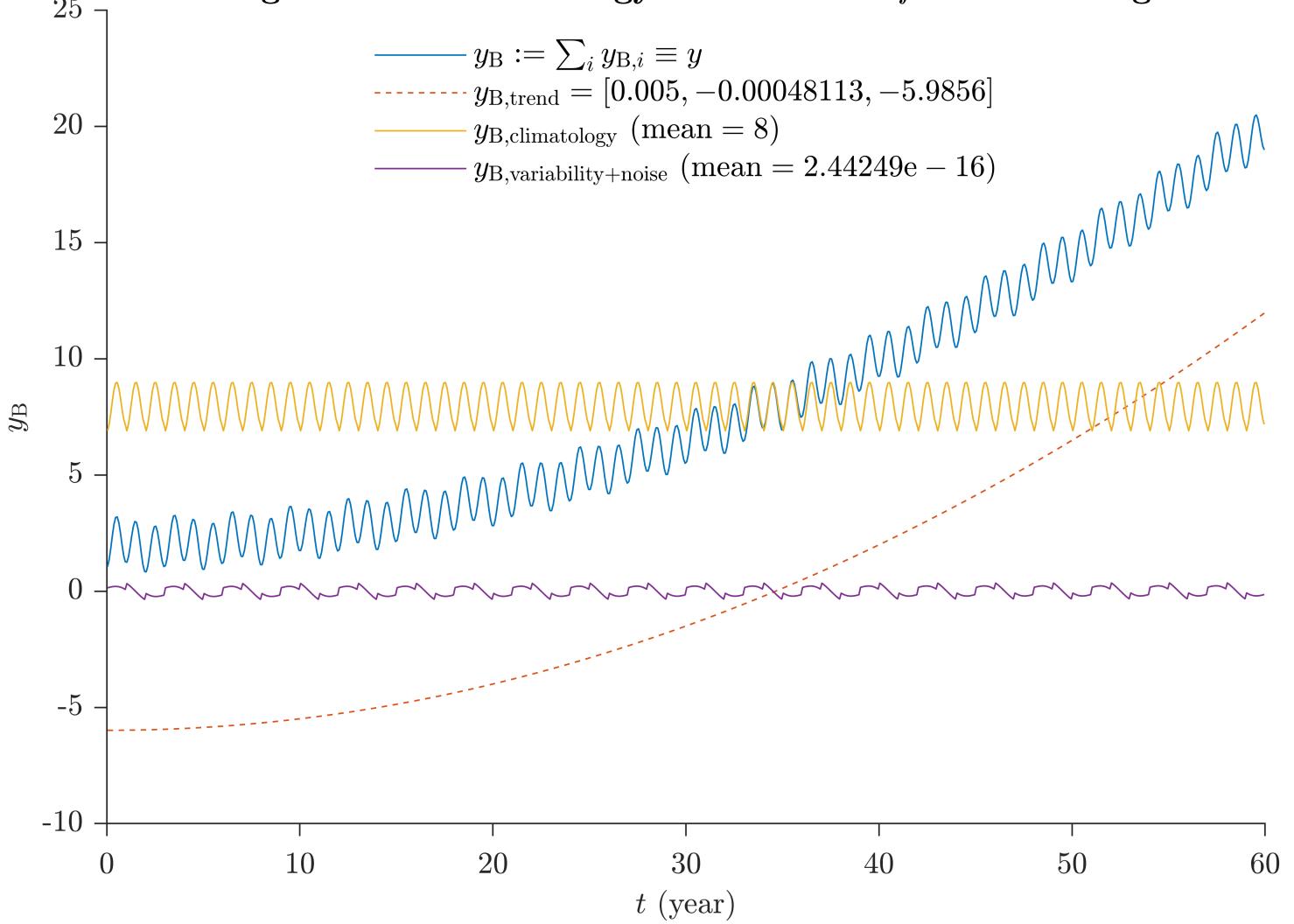


Figure 4.2(b) Algorithm B: "climatology" determined *before* detrending

## difference between output and original

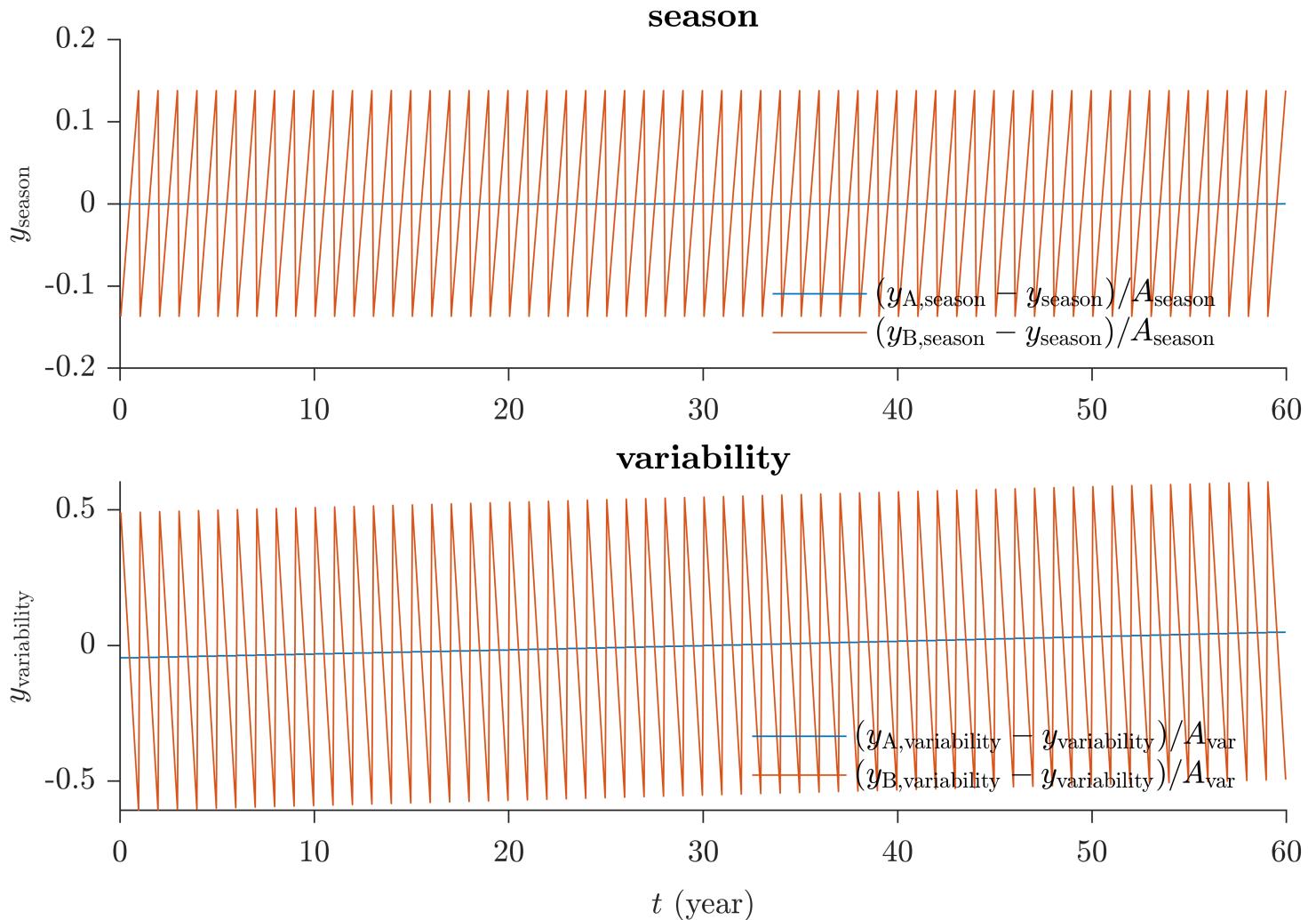


Figure 4.3 difference between output and original

## Case 5- slow linear trend, with noise added

- **variability**
  - shape (periods, amplitude):
    - A: good
    - B: period is good, but suffers from small burr
  - difference to original:
    - A: changes through successive and uniform small amplitude (~0.2%) staircases, with a linear trend from -5% to 5%
    - B: relatively large amplitude (~6%) sawtooth wave, with a linear trend from -5% to 5%
- **season (or climatology)**
  - shape (periods, amplitude):
    - A: bad

- B: bad
- difference to original:
  - A: large-amplitude ( $\sim 70\%$ ) sawtooth wave, around zero
  - B: Changes almost simultaneously with A

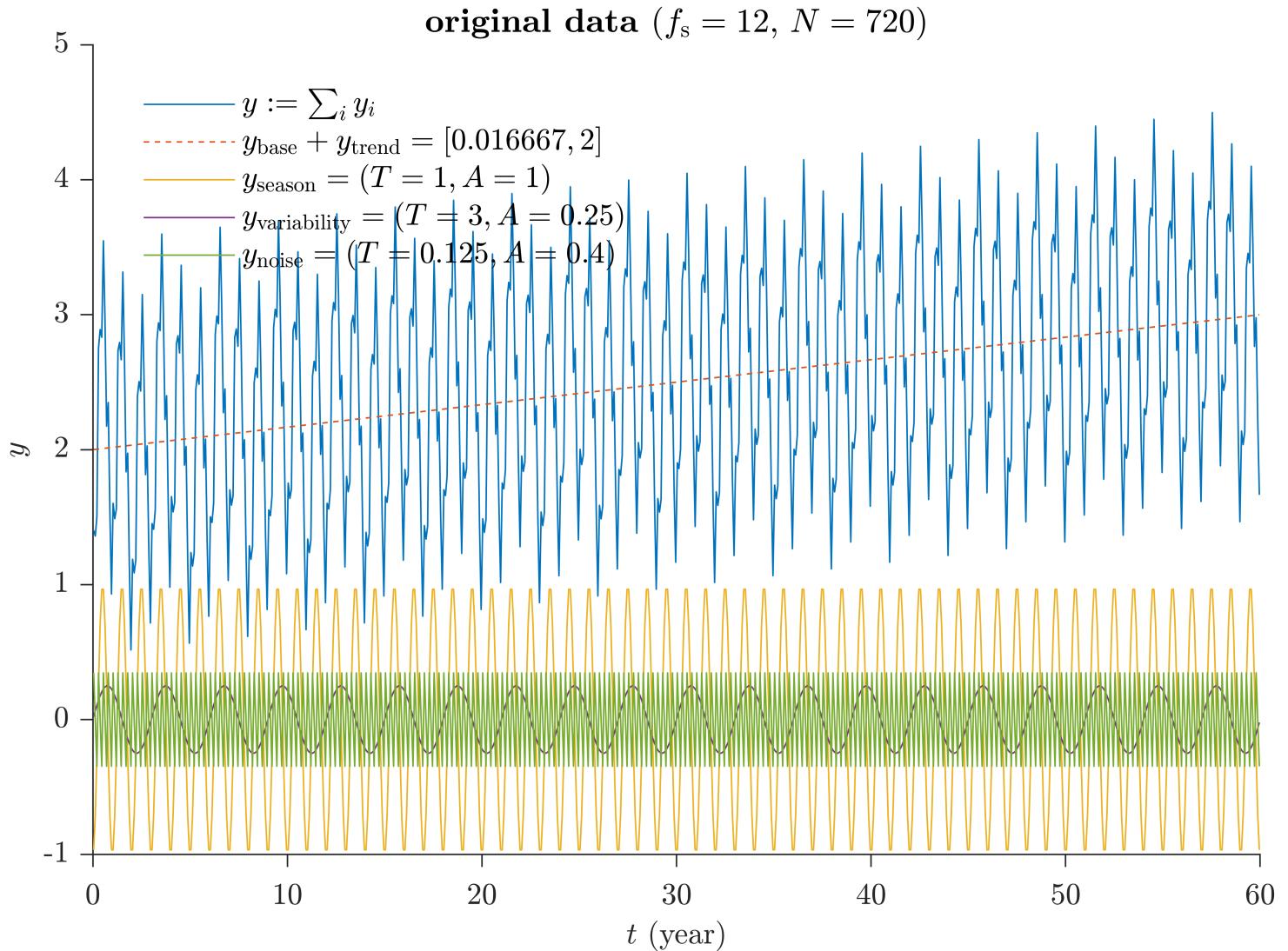


Figure 5.1 original data

### Algorithm A: climatology determined *after* detrending

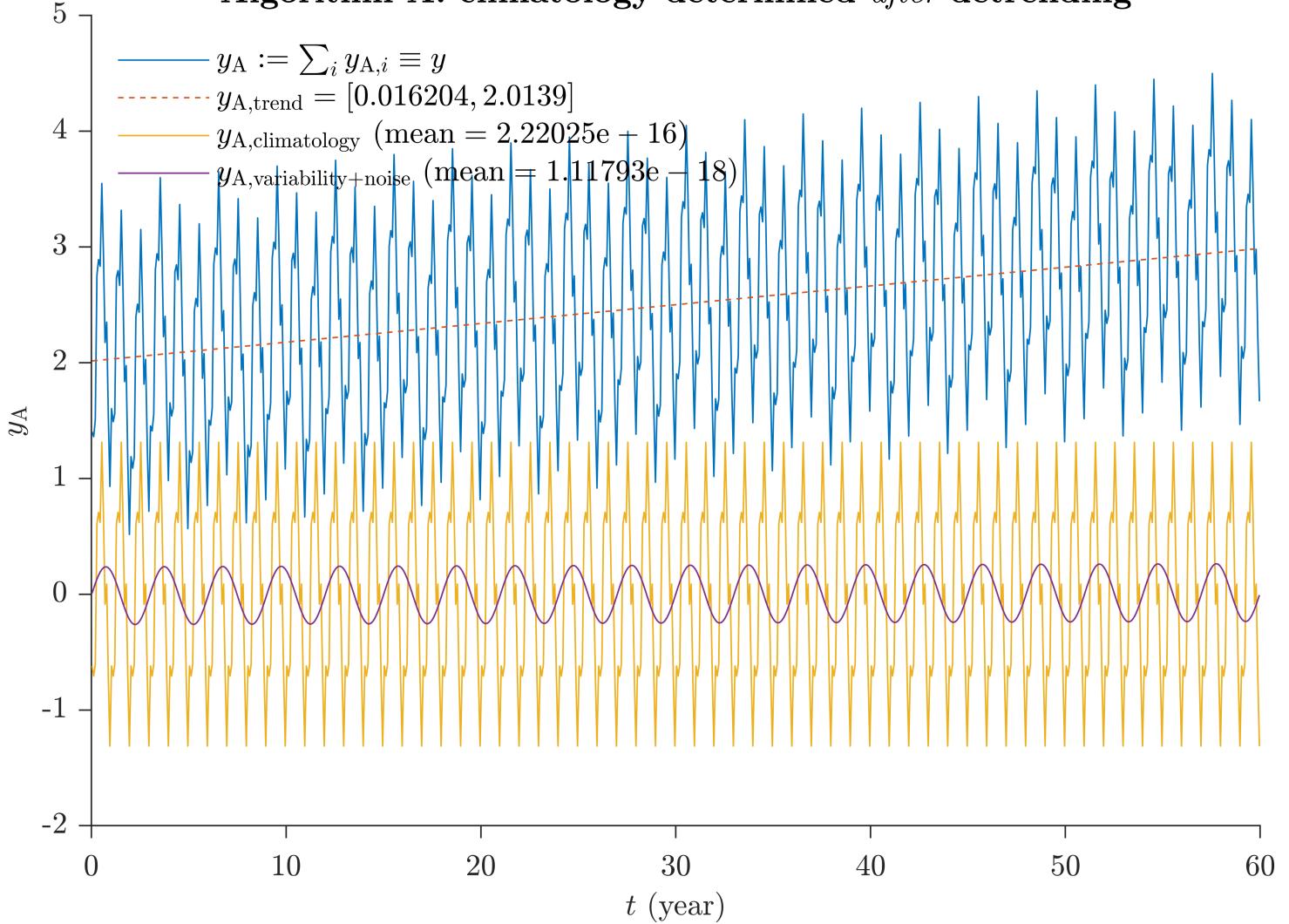


Figure 5.2(a) Algorithm A: "climatology" determined *after* detrending

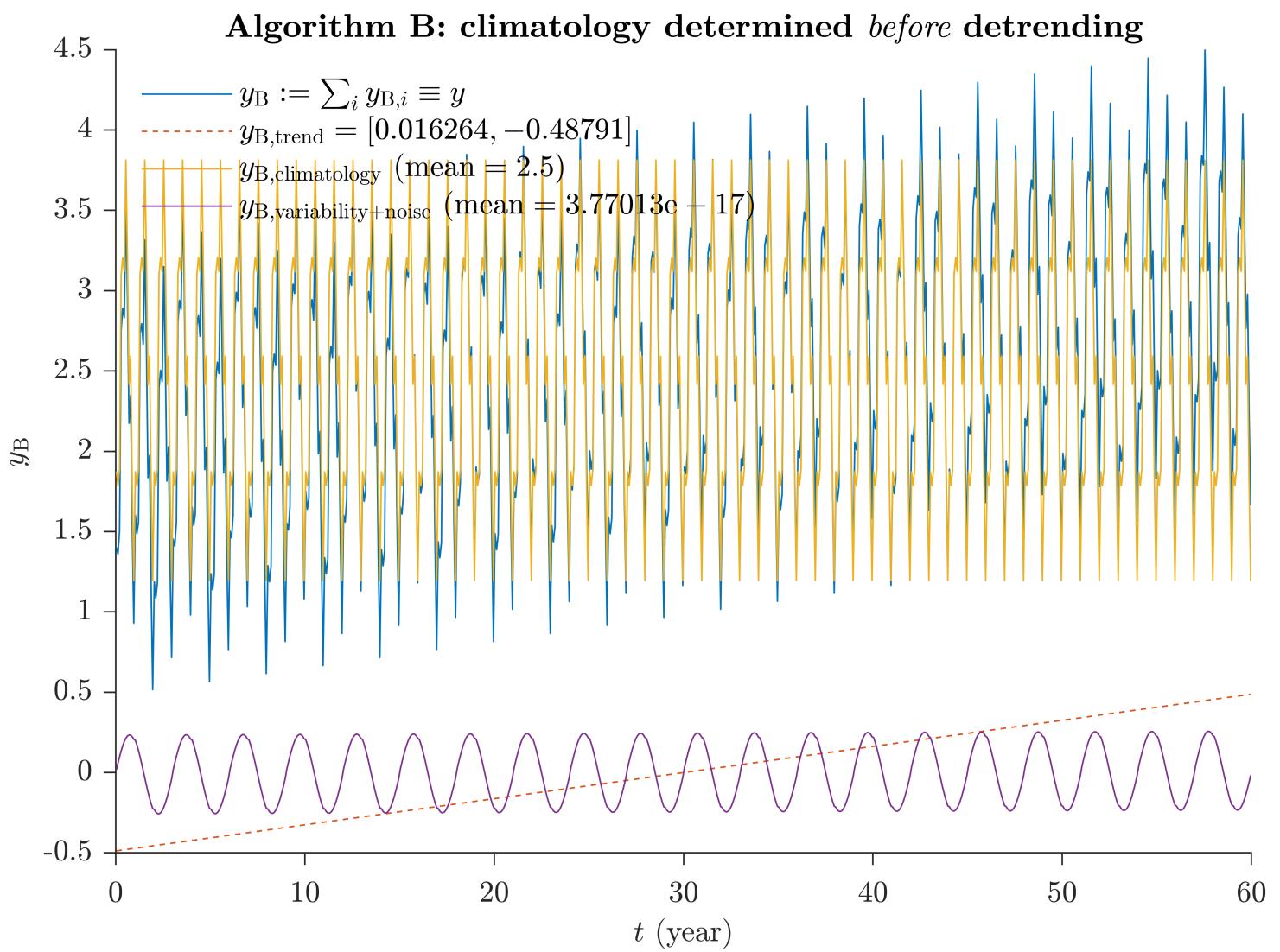


Figure 5.2(b) Algorithm B: "climatology" determined *before* detrending

## difference between output and original

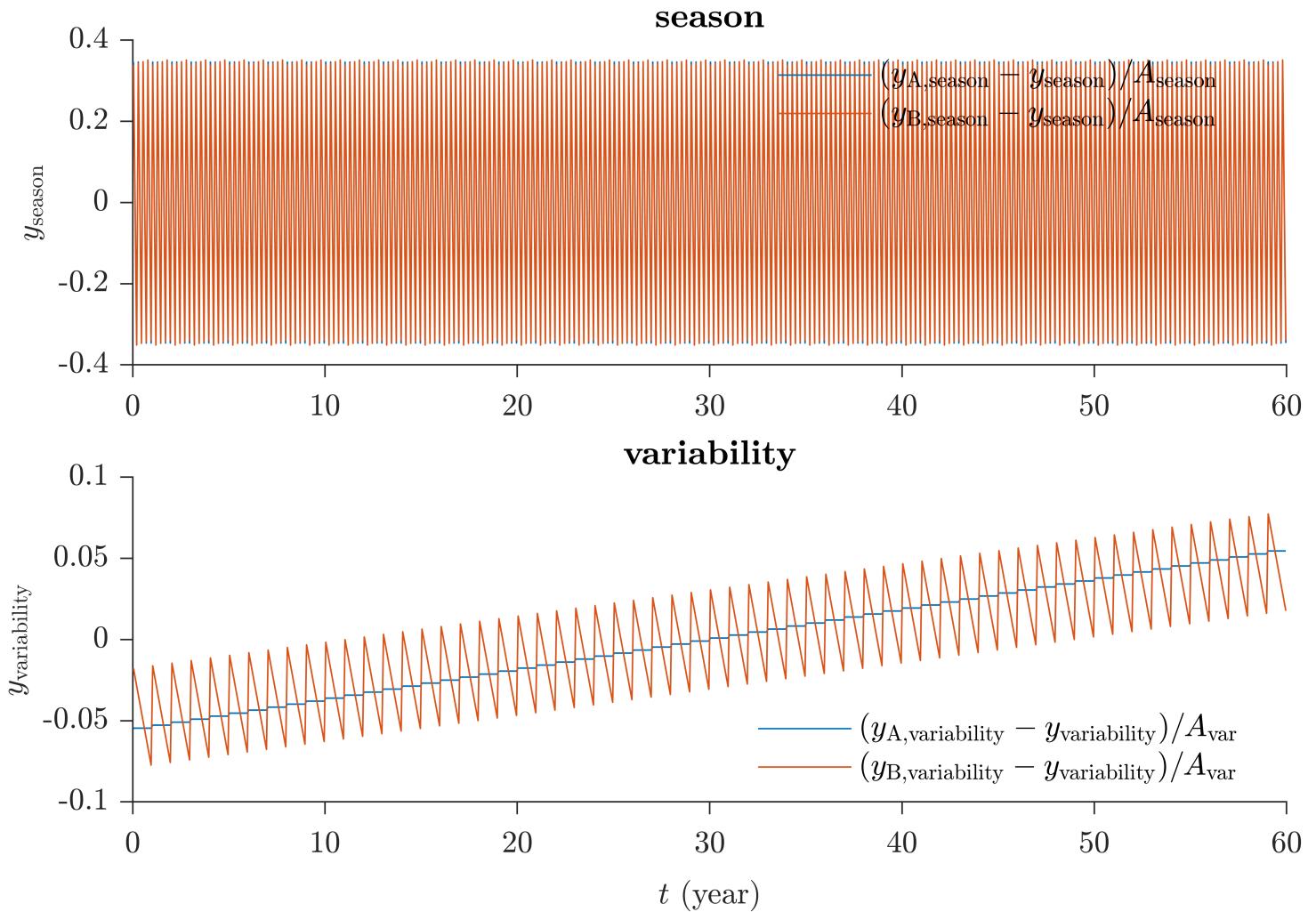


Figure 5.3 difference between output and original

## Conclusion and discussion

Several observations can be drawn from the results of this experiment.

### About climatology (or season)

- **Observation 1:** The output of climatology (or season) components of **Algorithm A** (*climatology determined after detrending*) is not sensitive to the degree or coefficients of the trend polynomial, while it is very sensitive in **Algorithm B** (*climatology determined before detrending*).
  - Note: In Algorithm B, the more intense the trend, the larger the amplitude of difference of output season components from its original counterpart.
  - Guess: In **Algorithm A**, trend is removed before climatology is determined, and therefore the output climatology is almost unaffected by the trend. In **Algorithm B**, climatology is

determined while the trend preserves. Intense long-term trend is equated to significant inter-monthly variability, so the shape of the trend is reflected to climatology. (The converse is *not* true!)

- **Observation 2:** The output of climatology (or season) components of **both** Algorithm is very sensitive to noise.
  - Guess: The sample rate 12 / year in this experiment is not sufficient to distinguish high frequency noise with seasonal anomaly signal.
  - Note: In this experiment, I have assumed that noise is added to the *variability* components instead of the *season* components of the output signals, which seems not true as suggested by this observation.

## About variability

- **Observation 3:** The output of *variability* components of **Algorithm A** (*climatology* determined *after* detrending) is not sensitive to the degree or coefficients of the trend polynomial, while it is very sensitive in **Algorithm B** (*climatology* determined *before* detrending).
  - Note: When free of noise (Case 1 to 4), the magnitudes of difference from its original counterpart of *climatology* and *variability* vary in the same direction, but are of opposite signs.
  - Guess 1: In **Algorithm A**, the effect of trend is eliminated by performing detrend before other operations are performed.
  - Guess 2: In **Algorithm B**, *trend* is determined after *climatology* is removed. Although *trend* has a significant effect on *climatology*, The removal of *climatology* has little on *trend*, as the former always has a 1-year cycle no matter what shape does it actually takes. Since the sum of *trend*, *climatology* and *variability* must be the *original* data when free of noise, the difference of *climatology* from the original one is mainly compensated by *variability*, which explains the above *note*. On the other hand, we have the observation present in previous section: More intense the trend, the larger the amplitude of difference of output season components from its original counterpart. Together they explained this observation.
- **Observation 4:** The output of *variability* components of **both** Algorithm is not sensitive to noise.
  - Guess: Under the sample rate 12 / year in this experiment, the high frequency noise signal is indistinguishable and is incorporated into climatology (or season) signals. The *variability* components should not be affected by noise, since it is the last components to be determined in **both** Algorithms.
- **Observation 5:** The output *variability* components in **both** Algorithms seems to share a common linear trend, which is unexpectedly consistent in all the cases performed here. It's really confusing.
  - Guess: Matlab function `polyfit` tends to underestimate the trend? That's impossible, since polynomial curve fitting methods has been well established.

# Resources

- [SEA-MAT: Matlab Tools for Oceanographic Analysis](#)
- [M\\_Map: A mapping package for Matlab](#)
- [Climate Data Toolbox for Matlab](#)

# Contact information

- Guorui Wei (危国锐)
- E-mail: [313017602@qq.com](mailto:313017602@qq.com); [weiguorui@sjtu.edu.cn](mailto:weiguorui@sjtu.edu.cn)
- Website: <https://grwei.github.io/>