Weighted composite analysis and its application: an example using ENSO and geopotential height

Zhang Xie,1,2* Anmin Duan1 and Qun Tian1,2
1State Key Laboratory of Numerical Modeling for Atmospheric Sciences and Geophysical Fluid Dynamics, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China
2University of Chinese Academy of Sciences, Beijing, China

Abstract

The arbitrariness of composite analysis due to criterion of defining positive and negative events is an unsolved problem. To deal with it, a new method based on the covariance and defined as weighted composite analysis was developed in this study. Two ideal cases reveal that the new method can eliminate noise more effectively in small sample size, and has a higher signal to noise ratio than the traditional method in some cases. And a real case on the relationship between ENSO and geopotential height over East Asia in summer indicates that the new method shows some similar features with composite analyses, but can effectually avoid some inappropriate conclusions. It therefore provides an alternative approach to investigate linkages between different variables.

Keywords: composite analysis; weighted composite analysis; statistic test

1. Introduction

Composite analysis (CA), also referred to as superposed epoch analysis and conditional sampling, is a useful tool to help understand the relationships among different phenomena. It is implemented as follows. Supposing that we want to investigate linkage between index I and field variable F, at first, we try to define some ‘events’ – they are usually defined as positive events and negative events – in terms of index I. If I satisfies a list of criteria, then the timings of those events are selected as key times. After that, we isolate the signal in variable F corresponding to I by calculating the average of F in all key times. Finally, by using a series of statistic tests, we can identify whether there is significant connection between those two variables or not. Through this sequence of steps, signals are enhanced resulting from extraction of events while noises are averaged out (Haurwitz and Brier, 1981; Laken and Čalogović, 2013). CA was originally developed by Chree (1913, 1914) in space science. And now it is widely-used in various fields of earth science (Shi et al., 1997; Kosaka and Nakamura, 2010; Fogt et al., 2011; Laken and Čalogović, 2013).

However, the criterion to define events is usually a dilemma in CA. Various criteria have been applied in previous work. Even, in some cases, researchers choose diverse criteria despite their subject being the same. Some prefer to select positive or negative events according to whether the index is greater (less) than (minus) one standard deviation (std.) from the mean (O’Reilly and Czaja, 2015), which means the events can be extreme enough to overcome the influence of noise. However, others have chosen to use criteria to produce positive, normal, and negative events equally (Lu, 2001). This means including more events into the composition to allow different types of noise to counteract each other. There is no standard answer in defining positive or negative events, leading to variable conclusions. As a result, many researchers have to compare different criteria to conform representativeness of event definition (Fogt et al., 2011). In this article, we discuss how unclear criteria lead to an apparent (seemingly reasonable) conclusion. Furthermore, we use a case study in Section 4 to demonstrate how a statistical test fails to avoid the influence of the subjective criteria.

To optimize CA, we developed an alternative approach to detect the relationship between a variable and an index with objective definition and higher signal to noise ratio. In Section 2, we discuss the problem of CA and why, and provide a new solution called weighted composite analysis (WCA). In Section 3, two ideal cases are designed to illustrate WCA is more effective in reducing noise and increasing signal to noise ratio than CA. In Section 4, the new method is compared with the traditional algorithm in real case study. Finally, the conclusions are drawn and discussed in Section 5.
positive pattern, \( P \); negative pattern, \( N \); and the difference pattern between them, \( D \), into mathematical expressions. First, to carry out CA, we require a reference index \( I \) to classify field \( F \) into a positive events set \( I_p \) and a negative events set \( I_N \). Then, based on the algorithm of CA, we can define:

$$
P = \frac{\sum X_{I_p}(I_i) \cdot F_i}{\sum X_{I_p}(I_i)}, N = \frac{\sum X_{I_N}(I_i) \cdot F_i}{\sum X_{I_N}(I_i)}, D = P - N
$$

(1)

where, \( I_i \) and \( F_i \) are the \( i \)th time point of \( I \) and \( F \), respectively, and

$$
\sum X_{I_p}(I_i) = \begin{cases} 1, & I_i \in I_p \setminus I_N \\ 0, & I_i \notin I_p \end{cases}, \sum X_{I_N}(I_i) = \begin{cases} 1, & I_i \in I_N \setminus I_p \\ 0, & I_i \notin I_N \end{cases}
$$

(2)

According to Equation (1), it is clear that the results of CA can be regarded as a weighted sum of \( F_i \) and the weight \( (C_{I_p},P \) denotes weight of the positive events set, and \( C_{I_N},N \) denotes the weight of the negative events set) at the \( i \)th point is given by

$$
C_{I_p} = \frac{X_{I_p}(I_i)}{\sum X_{I_p}(I_i)}, C_{I_N} = \frac{X_{I_N}(I_i)}{\sum X_{I_N}(I_i)}
$$

(3)

The artificiality in CA is clearly determined by the set \( I_p \) and \( I_N \). Because of even weight for every member in the set, the increasing set volume tends to abate the amplitude of harmonic information. But in the sense of noise reduction, it is imperative to gain sufficient sets to average out irrelevant fluctuations. As a result, we encountered a dilemma of criteria, and thus the result of CA is terribly determined by the user’s preference.

2.2. Definition of WCA

To develop a new method, we identify a reasonable weight in a weighted sum. Considering the issue in CA, we try to identify a reasonable weight in a weighted sum, which takes form

$$
F_w = \sum C_i \cdot F_i
$$

(4)

where \( F_w \) is our analytical result with values dependent on whether positive or negative phase, which remains constant in a specific phase but differs from one phase to another, \( C_i \) is weight for field value \( F_i \) at \( i \)th time point.

Traditionally, covariance between two variables can be explained as an indicator to represent the connection among them. It suggests that if we design a method in which its production reproduces the covariance between two events by replacing \( F_i \), the result could be a representative of relationship between those phenomena. To achieve this goal, we assume

$$
\sum \left( I_i - \bar{I} \right) \left( F_a - \bar{F} \right) = \sum \left( I_i - \bar{I} \right) \left( F_i - \bar{F} \right)
$$

(5)

where \( \bar{I} \) and \( \bar{F} \) is the mean of \( I \) and \( F \) and the other variables are defined as previously. After substituting Equations (4) into Equations (5) and reordering the calculation, the equation can be rewritten as

$$
\sum \left[ \left( I_i - \bar{I} \right) - \left( \sum \left( I_i - \bar{I} \right) \right) \right] F_i = 0
$$

(6)

Therefore, the values of \( C_i \) in the positive and negative event sections are identified easily. To obtain the value of \( C_i \) in the positive and negative event sections, the positive events set \( I_p \) contains events where the reference index \( I_i \) is above \( \bar{I} \), while the events set \( I_N \) contains the other events. In terms of the positive event section, \( C_i \), which is outside \( I_p \), can be set equal to zero. Furthermore, considering Equation (4), we can obtain weights within \( I_p \). Finally, \( C_i \) takes the form

$$
C_i = \frac{I_i}{\sum I_i}, I_i = \begin{cases} 1, & I_i - \bar{I}, I_i > \bar{I} \\ 0, & I_i \leq \bar{I} \end{cases}
$$

(7)

The sum of the weight given by Equation (5) is equal to one, and therefore this method preserves the scale of the data. The weights can also be defined using the same procedure in the negative event section. As a result, we can define a positive pattern, \( P \); and a negative pattern, \( N \); and the difference between them, \( D \), in a new way as:

$$
P = \frac{\sum I_p F_i}{\sum I_p}, N = \frac{\sum I_N F_i}{\sum I_N}, D = P - N
$$

$$
I_p = \begin{cases} 1, & I_i - \bar{I}, I_i > \bar{I} \\ 0, & I_i \leq \bar{I} \end{cases}, I_N = \begin{cases} 1, & I_i - \bar{I}, I_i < \bar{I} \\ 0, & I_i \geq \bar{I} \end{cases}
$$

(8)

It can be noticed that new weight for \( D \) in WCA resembles the regression coefficient \( \beta \) of variable \( F \) onto index \( I \):

$$
\beta = \frac{\sum \left( I_i - \bar{I} \right) \left( F_i - \bar{F} \right)}{\sum \left( I_i - \bar{I} \right)^2}
$$

(9)

Explicitly, because the progresses of both algorithms are based on the covariance represented by numerators of Equations (8) and (9), and thus reflect the connection between two phenomena, WCA resembles the regression analysis from this perspective. However, taking the individual weight as whole into account, since WCA is an estimate for variable \( F \) itself rather than variation of \( F \) with respect to index \( I \), the unit differentiates them in utilization, that WCA prepares for problem related to event phase while regression coefficient is usually applied in prediction.

Based on previous work, we selected the Monte Carlo method to test the significance of our results (Haurwitz
3. Signal to noise comparison between CA and WCA

3.1. Comparison of noise reduction

An ideal experiment was designed to test the effectiveness on noise reduction of two algorithms. We randomly created 30 samples according to the standard normal distribution as noise, and compared the amplitude of the mean noise in the results of the two approaches before counting the results of the approach that produced smaller amplitude (the weights used to carry out WCA were chosen from a standard normal distribution and the sample sizes for CA were 5, 10, and 15). After 1000 tests (Table 1), the performance of WCA with five events was statistically better than that of CA, and worse with a decrease in sample size, where the critical value was 10. We use a simple theoretical model to explain the result, and the model indicates when selected events are less than approximately 1/3 total sample size, the noise reduction of WCA is more effective than CA (Appendix S2).

Close scrutiny of the statistic features in the experiment reveals WCA is quite robust in case with small sample size, which means that the method can effectively avoid the impact from occasional event with strong noise, such as the case in Section 4, and successfully reduce the possibility to produce extremely huge noise. Therefore, in the real case where we can hardly know what the true value is, the stability of WCA can effectively avoid the cause of disastrous estimate, leading to a more reliable conclusion.

Overall, compared with CA, the increase of sample size in WCA can reduce the noise even though every member is weighted unequally.

3.2. Comparison of signal to noise ratio

It is clear that no method is suitable for all situations. To illustrate, we created a series of index x obeying the standard normal distribution with a sample size of 30, and selected a linear function $y = 10x$, which is a representative of linear function with a slight stronger WNPSH. In terms of the negative events (Figure 2(b), (d), (f), (h)), with the increase of the standard, the significant regions also increased, including area B in the western North Pacific, northeastern Japan. In addition, a negative anomaly emerged over area A in the CA with strict criteria, which was absent in WCA. Overall, all of the algorithms could reproduce the main characteristics of the ENSO’s effect on the geopotential height field at 500 hPa in East Asia during July. Conversely, significant areas in the CA tended to differ based on the different standards.

Focusing on the similarity and difference between CA and WCA, the strong signals over area A only

![Table 1. Noise elimination performance counts of CA and WCA.](https://github.com/YMI33/method-library)

| Strong events in CA | CA   | WCA  |
|--------------------|------|------|
| 5                  | 394  | 606  |
| 10                 | 501  | 499  |
| 15                 | 589  | 411  |

more effective in nearly 60% of total sample. After examining diverse parameters and functions, the result is robust for most cases with significant signal fluctuation (Appendix S3).

The unsatisfactory performance of CA occurs because large samples are enough to overcome noise, but cannot guarantee the strength of signal while few members introduce not only strong signal but also increasing noise. The dilemma between the noise reduction and signal strength is reflected in the criterion of sample size in CA, which results in a lower signal to noise ratio than WCA in this case. Therefore, at least in this case, WCA overrides CA statistically.

4. Application in an ENSO related case

We selected the summer geopotential height anomalies at 500 hPa that are associated with ENSO as a case study to examine the reliability of the new method. The geopotential data were obtained from the Japanese 55-year Reanalysis (JRA-55, 1.25°×1.25°, 37 pressure levels) (Ebita et al., 2011). The Niño3.4 index was downloaded from http://www.esrl.noaa.gov/psd/gcos-wgsp/Timeseries/Nino34/ (GCOS-AOPC/OOPC) (Rayner et al., 2003).

Analyzing the positive events (Figure 2(a), (c), (e), (g)), the circulation involved in the East Asian climate can be clearly seen, including the western North Pacific subtropical high (WNPSH) and wave activities in the westerlies, in the CA using all criteria and WCA. All the figure panels show that the WNPSH intensified with southwestward movement. However, the areas that passed the test of significant confidence varied from one algorithm to another. In CA, with the criteria increasing from 0.5 std. to 1.0 std., the significant positive anomalies over East Siberia disappeared, whereas negative anomalies emerged east of 160°E. Furthermore, the pattern of WCA resembled that of CA using 0.5 std. with a slightly stronger WNPSH. In terms of the negative events (Figure 2(b), (d), (f), (h)), with the increase of the standard, the significant regions also increased, including area B in the western North Pacific, northeast of Japan. In addition, a negative anomaly emerged over area A in the CA with strict criteria, which was absent in WCA. Overall, all of the algorithms could reproduce the main characteristics of the ENSO’s effect on the geopotential height field at 500 hPa in East Asia during July. Conversely, significant areas in the CA tended to differ based on the different standards.

Focusing on the similarity and difference between CA and WCA, the strong signals over area A only

© 2017 The Authors, Atmospheric Science Letters published by John Wiley & Sons Ltd on behalf of the Royal Meteorological Society.
existed in the negative event of the CA with strict criteria, whereas significant anomalies appeared over area B in the negative event of all algorithms (Figure 2(b), (d), (f), (h)). The CA result can usually be explained as a response to one event. For example, if there is a negative geopotential height anomaly at 500 hPa in the following summer after a negative phase ENSO using Niño3.4 index in CA, we will usually conclude that a depression appeared there after a strong La Niña event. That is to say, the result of the CA will be perceived as what usually occurs. Looking at the signals over area A after a strong La Niña event (Niño3.4 < -0.75 std., Figure 3), we found five cases of negative and four cases of positive geopotential height anomalies. In particular, a positive case occurred in 1999 when an extreme La Niña event took place, which was different from area B, where there were seven cases of negative events versus two cases of positive events. In fact, the negative anomalies originated from the impact from five extremely strong events instead of the general dataset. Admittedly, we cannot exclude the hypothesis that the signal in area A is a response to the extreme La Niña event (Niño3.4 < -1.0 std., with five cases of negative anomalies versus one case of positive anomaly). However, for the criterion of 0.75 std., although the signal in this region passes the confidence test of significance due to 5 strong events, it is inappropriate to conclude that a depression ‘usually’ (5 out of 9) occurs after a strong La Niña event. Thus, the conclusion is inconsistent with traditional understanding. In this point of view, CA therefore leads to confusion in this ENSO-related case, while this confusion does not occur in WCA. The strong signal ‘usually’ (7 out of 9) occurred in area B. Therefore, it is a reliable response to a La Niña event, which can be detected in both CA and WCA. In general, WCA filtered the noise over area A and maintained the harmonic signal over area B, while CA failed to do so.

5. Discussion and conclusions

We explored the problems related to composite analyses, and their causes, and developed a new method called WCA. The following are the main conclusions:

1. Composite analysis, due to the subjectivity of dividing an event set, may produce several divergences on a real phenomenon. To deal with the issue, a new algorithm, called WCA, was developed, which is a composite sum with weights defined according to covariance. It is a well-defined method with both high signal to noise ratio and clear reflection of covariance between two variables.

2. The ideal case study substantiates, the increasing sample size for the calculation gives WCA a better performance of noise elimination than CA in small sample size case. And the former can achieve a higher signal to noise ratio than the latter in some cases.

3. A case study of the geopotential height anomalies in summer at 500 hPa that are associated with ENSO was selected to compare the reliability of the two approaches. The results indicate CA introduced a confusing signal in the high latitudes. WCA performed better than CA in this case.

If the sample size is large enough, the results of WCA and CA are similar. However, this goal is hard to achieve. Because of the quality of data or the removal of low-frequency fluctuation, data are usually limited. Therefore, if we try to explore the relationship between field and index in different phase, the WCA may be a better choice than CA. Clearly, CA cannot be replaced in all cases. For instance, if the studied events are well-defined, it may be more appropriate to apply CA. In addition, WCA exhibits some similar features like regression and correlation analysis, because all of three approaches pertain to the covariance. However,
the difference of amplitude among the three methods suggests us to utilize them for different purposes, such as WCA for system description, regression for variable prediction and correlation for circumstance relationship. It is therefore important to decide which tools to apply based on specific research.

As with all methods, there are still several drawbacks with WCA. The main disadvantage is its sensitivity to extreme events. Because the weights vary at different times, the impact of extreme events, which take the main share of the total weight, tends to be magnified. That is, if there are some extreme events, the result will resemble the state when these extreme events occurred instead of representing the mean state of the total dataset. Furthermore, additional statistical testing of WCA will be carried out in the future. The Monte Carlo
test requires extensive computing resources. Therefore, a simpler and more effective way to test the outcomes of WCA is required when dealing with a large amount of data. Nevertheless, WCA is an approach that can help understand the linkages among data from a new perspective.

Acknowledgements
This work is supported by the National Natural Science Foundation of China (Grants 91337216). We gratefully acknowledge Zipeng Yu, Yaoxian Yang, Zhaolu Hou, and Jiaqing Xue of the Institute of Atmospheric Physics, Chinese Academy of Sciences (IAP) for comments on the original draft. We also thank Professor Yihua Lin of IAP and Professor Yuanxi Yang of the Xi’an Institute of Surveying and Mapping for helpful suggestions. And we are grateful to the comments on the paper from two reviewers.

Supporting information
The following supporting information is available:
Appendix S1 Statistic test.
Appendix S2 Theoretical model of noise reduction.
Appendix S3 Signal to noise ratio in different circumstance.

References
Chree C. 1913. Some phenomena of sunspots and of terrestrial magnetism at Kew Observatory. Philosophical Transactions of the Royal Society of London Series A: Containing Papers of a Mathematical or Physical Character 212: 75–116. https://doi.org/10.1098/rsta.1913.0003.
Chree C. 1914. Some Phenomena of Sunspots and of Terrestrial Magnetism. Part II. Philosophical Transactions of the Royal Society of London Series A: Containing Papers of a Mathematical or Physical Character 213: 245–277. https://doi.org/10.1098/rsta.1914.0006.
Ebita A, Kobayashi S, Ota Y, Moriya M, Kumabe R, Onogi K, Harada Y, Yasui S, Miyaoka K, Takahashi K, Kamahori H, Kobayashi C, Endo H, Soma M, Oikawa Y, Ishinizu T. 2011. The Japanese 55-year Reanalysis "JRA-55": an interim report. SOLA 7: 149–152. https://doi.org/10.2151/sola.2011-038.
Fogt RL, Bromwich DH, Hines KM. 2011. Understanding the SAM influence on the South Pacific ENSO teleconnection. Climate Dynamics 36: 1555–1576. https://doi.org/10.1007/s00382-010-0905-0.
Haurwitz MW, Brier GW. 1981. A critique of the superposed epoch analysis method: its application to solar-weather relations. Monthly Weather Review 109: 2074–2079. https://doi.org/10.1175/1520-0493(1981)109<2074:ACOTSE>2.0.CO;2.
Kosaka Y, Nakamura H. 2010. Mechanisms of meridional teleconnection observed between a summer monsoon system and a subtropical anticyclone. Part I: the Pacific-Japan pattern. Journal of Climate 23: 5085–5108. https://doi.org/10.1175/2010JCLI3413.1.
Laken BA, Čalogović J. 2013. Composite analysis with Monte Carlo methods: an example with cosmic rays and clouds. Journal of Space Weather and Space Climate 3: A29. https://doi.org/10.1051/swsc/2013051.
Lu R. 2001. Interannual variability of the summertime North Pacific subtropical high and its relation to atmospheric convection over the warm pool. Journal of the Meteorological Society of Japan Series II 79: 771–783. https://doi.org/10.2151/jmsj.79.771.
O’Reilly CH, Czaja A. 2015. The response of the Pacific storm track and atmospheric circulation to Kuroshio Extension variability. Quarterly Journal of the Royal Meteorological Society 141: 52–66. https://doi.org/10.1002/qj.2334.
Rayner NA, Parker DE, Horton EB, Folland CK, Alexander LV, Rowell DP, Kent EC, Kaplan A. 2003. Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. Journal of Geophysical Research 108: 4407. https://doi.org/10.1029/2002JD002670.
Shi X, Wei FY, Feng GL, Sheng TL. 1997. Monte Carlo test used in correlation and composite analysis of meteorological fields. Journal of Nanjing Institute of Meteorology (in Chinese) 20: 355–359. https://doi.org/10.13878/j.cnki.dqxxb.1997.03.011.
Singh YP, Badruddin. 2006. Statistical considerations in superposed epoch analysis and its applications in space research. Journal of Atmospheric and Solar-Terrestrial Physics 68: 803–813. https://doi.org/10.1016/j.jastp.2006.01.007.