Interactive comment on “Understanding meteorological influences on \( \text{PM}_{2.5} \) concentrations across China: a temporal and spatial perspective” by Ziyue Chen et al.

Ziyue Chen et al.
zychen@bnu.edu.cn

Received and published: 30 October 2017

Anonymous Referee #3

R: Dear Referee, thanks so much for your comments, which helped us improve this manuscript a lot. We have fully revised this manuscript according to your general and detailed comments. Yes, as you pointed out, the principle of CCM and its advantages compared with the correlation analysis are not well-known to all scholars and thus a better explanation of the CCM method would be needed. Due to limited space, we did not add all details to the modified manuscript and some responses and explanations are given here according to your specific comments. Meanwhile, some other issues were also addressed. As your suggested that multiple year data are required, we have extended the research period from one year to three years using latest published data. Other issues are responded as follows. Thanks again for all your valuable comments and we are willing to conduct further revisions if you have further requirements.

General Comments
This is an interesting paper that applies an exciting and fairly new statistical method (convergent cross mapping) to quantify the relationship between local air quality and local meteorology (wind, temperature, precipitation, etc.). The authors argue that unlike a simple correlation analysis, this method is able to demonstrate causal relationships between variables. My understanding is that this method is quite new, and it is central to this study, so I think this paper would benefit from a clearer discussion of why it is better than correlation at determining causal relationships. Would the authors’ findings have differed significantly if they used correlation instead of CCM?

R: The CCM method was proposed by Sugihara et al. (2012). 1 Sugihara, G., May, R., Ye, H., Hsieh, C., Deyle, E., Fogarty, M., Munch, S. 2012. Detecting Causality in Complex Ecosystems. Science, 338, 496-500. Sugihara et al. (2012) pointed out that correlation analysis could extract mirage correlations, especially in complicated ecosystems. For instance, two variables A and B that have no causality may demonstrate significant correlations due to the existence of an agent variable C, which interacts with both A and B. Through a series of experiments, Sugihara et al. (2012) proved that this type of mirage correlations could be detected using the CCM method by calculating a p value of 0. The CCM method not only performed better than the correlation analysis in causality analysis by excluding the influence of other variables, but also demonstrated the advantage of detecting weak causality compared with other causality analysis method (e.g. Granger Causality), which may fail to detect weak to moderate coupling between variables. In our previous studies, we employed both the correlation and the CCM method to examine the influence of individual meteorological factors on \( \text{PM}_{2.5} \) concentrations in the Beijing-Tianjin-Hebei region and compared the performance of correlation and CCM method. 2 Chen, Z., Cai, J., Gao, B.B., Xu, B.,
Dai, S., He, B., Xie, X.M. Detecting the causality influence of individual meteorological factors on local PM2.5 concentration in the Jing-Jin-Ji region. Scientific Reports 2017. 7:407352 The comparison suggests that the causality influence of individual meteorological factors on PM2.5 concentration is better revealed using the CCM method than the correlation analysis. By comparing the correlation coefficient and value in Table 2, one can see that some correlations between meteorological factors and PM2.5 concentration may result from mirage correlations (e.g. the correlation between meanRHU and PM2.5 concentration in Hengshui in summer). Secondly, CCM analysis reveals weak or moderate coupling (e.g. the interactions between SSD and PM2.5 concentration in Cangzhou in summer) whilst correlation analysis cannot. Additionally, due to interactions between different meteorological factors, the value of correlation coefficients cannot interpret the quantitative influence of individual meteorological factors on PM2.5 concentration. Instead, the value from CCM method is designed to understand the coupling between two variables by excluding influences from other factors. Through comparison, the value of the correlation coefficient for some meteorological factors is notably different from the value for these meteorological factors. A large correlation coefficient for one meteorological factor may correspond to a much smaller value from the CCM analysis, indicating that the value of the correlation coefficient usually over-estimates the influence of individual meteorological factors on PM2.5 concentration. The previous research (Chen et al., 2017) proved that the CCM method outperform the correlation analysis in many aspects. And this research extended the study area from the Beijing-Tianjin-Hebei region to a national scale, so the CCM method, instead of the correlation analysis, remains the ideal tool for quantifying meteorological influences on PM2.5 concentration over China.

There are some aspects of CCM that were not clear to me. Does CCM account for relationships between meteorological factors? E.g. if wind is affecting both precipitation and PM, how is the affect of wind on precip on PM counted?

R: Yes, you made a very good point here. The exclusion of influences from other variables and solely focus on the interactions between two target variables are the most important advantages of the CCM. Sugihara et al. (2012) proposed the CCM method and examined its performance in removing the influence of agent variables through a diversity of experiments. The result proved that mirage correlations caused by the influence of other variables could be detected and removed by the CCM method. For instance, Sugihara (2012) suggested that the CCM method could quantify the bi-directional interactions between two individual variables without being affected other variables, which were also proved by a diversity of studies. So for your question, for each calculation, the CCM method simply examine the bi-directional coupling between two selected variables in complicated ecosystem whilst the influence of other factors was excluded. For your instance, we could calculated the coupling between PM and wind speed, and the coupling between precipitation and wind speed using the CCM method respectively, and the results for both calculations would not influence each other.

Another question I have is what produces a large value. For example, are the values higher in winter simply because there is more PM available to be effected?

R: Generally, the CCM method simply calculates the quantitative influences of individual meteorological factors on PM concentrations, whilst the mechanisms were not revealed. The mechanisms how even one individual meteorological factor influences local PM concentrations can be highly complicated. The chemical compositions, size and mass concentrations of particulate matters actually vary significantly across locations and seasons. There is no comprehensive research on how one individual meteorological factor influence PM2.5 concentrations through different mechanisms. The large value in winter may result from both the much higher PM2.5 concentrations, which may be easily influenced by meteorological factors, and unique meteorological conditions. For instance, PM2.5 induced haze weather occurred frequently in winter in the Beijing, indicating a much higher PM2.5 concentration. Meanwhile, some meteorological factors in Beijing in winter were quite different from those in other seasons. For
instance, the northwest wind prevails in winter and it has become a common scene and a popular saying that “the best solution for haze in Beijing is to wait for the wind”. Thus the value for wind speed in many cities within the Beijing-Tianjin-Hebei region was much larger in winter than in other seasons for both PM2.5 concentrations due to both high PM concentrations and strong wind speeds. Another instance is that the chemical composition of PM, which is related to photochemical reaction and solar radiations, also varied significantly in different seasons. Hence, the reason for the variation of value across seasons is highly complicated, and the high PM2.5 concentration is one of major reasons. Thanks so much for pointing this out. I believe systematic research on the influencing mechanism of individual meteorological factors should be examined in-depth by scholars from a diversity of backgrounds.

Or as another example, is precipitation more effective at removing PM along the coasts because it rains more? If there were a way to normalize by the amount of total rainfall, would precip still be more important along the coasts than in the drier interior?

R: Yes, it is highly possible that precipitation is more effective at removing PM along the coasts because it rains at a higher frequency and intensity. As we know that the PM2.5 concentrations drop significantly after a heavy rain whilst light rain may not reduce PM2.5 concentrations significantly. Meanwhile, PM2.5 concentrations may also affect the influence of precipitation. Light rains may have limited washing-off effects on high-concentration PM2.5 concentrations and may increase the relative humidity in the environment, which is favorable for the rising of PM2.5 concentrations. In the drier interior, the PM2.5 concentrations are usually much higher and the intensity and frequency of precipitation are much lower than those along coasts. These two factors may both be the reason that precipitation is more effective at removing PM along the coasts. Due to these influencing factors, precipitation may still be a less important meteorological driving force for PM2.5 concentrations in the drier interior, even if there were a way to normalize by the amount of total rainfall. In the revised manuscript, we added a new part to briefly introduce the underlying reasons for the variation of

My second major concern is that this study uses a single year of data to make general comments about PM-meteorology relationships. This gives us little sense of how much year-to-year variability may exist in these relationships and generally weakens the conclusions. If CCM is too computationally expensive to use on multiple years, perhaps a different method could be used to supplement it.

R: This is a very good point. When we were preparing this manuscript, there were only one-year’s data available. And now, we endeavored to acquire another two years’ data for multiple year analysis. Although the CCM is computationally expensive, we still believe it is more persuasive to use multiple years’ data for drawing conclusions. So thanks so much for this suggestions. We have re-calculated the CCM results using three years’ data, and the updated results have been added to the revised manuscript. By comparing the mean annual p value for eight variables for 2014 with the mean annual p value for eight variables during 2014-2016, we can see that the meteorological influences on PM2.5 concentrations at the national scale did not change significantly in the past three years.

Finally, I would like to see a deeper discussion of the scientific significance of this work. As it is written currently, this paper is almost purely descriptive. The new method is interesting, but the paper could do a better job of articulating what we are learning from it. Perhaps some discussion about why different meteorological factors are more/less important in different regions/seasons, for example, would help give the paper more depth. I would also suggest spending more time discussing the implications for modeling, especially in the introduction and conclusions, as that was what I took away as the most important implication in this paper.

R: Thanks so much for this valuable suggestion. Yes, it is important to provide more in-depth discussion for the acquired results based on the CCM. And as you suggested, some discussion, e.g. the mechanisms between meteorology factors and PM2.5 concentrations were well-known, and it is not necessary to introduce it in details. So in the revised manuscript, we have fully removed this part and left more space for the
parts you suggested. We added some introduction on the potential applications the CCM method, instead of the correlation analysis, in research concerning meteorology-included PM2.5 concentration estimation prediction. Meanwhile, the potential reason for large variations of meteorological influences on PM2.5 concentrations across China has been added to the revised version. Thanks again for this valuable comment.

Specific Comments pp 3, In 93-95 - Can you elaborate further on how your previous study showed that CCM is better than correlation for the benefit of readers who have not read that paper. It is important for your results here to make as clear as possible why CCM is a better method/provides new insights.

R: This is a very good point. We have mentioned the advantages of the CCM method, compared with the correlation analysis in the method part. However, as you pointed out, we should give more information on how the CCM method performed better than the correlation analysis. So we added some extra details concerning the CCM method in the method part (since Line 197).

As mentioned above, our previous study (Chen et al., 2017) employed the correlation analysis and the CCM method to examine the influence of individual meteorological factors on PM2.5 concentrations in the Beijing-Tianjin-Hebei region. In the paper, we demonstrated detailed results of both correlation and causality influence in two large tables. By comparing these results, we found that (1) some correlations between meteorological factors and PM2.5 concentration may result from mirage correlations. For instance, there was a correlation between meanRHU and PM2.5 concentration in Hengshui in summer whilst the causality influence of meanRHU on PM2.5 concentrations is 0, indicating a mirage correlation. (2) CCM analysis reveals weak or moderate coupling whilst correlation analysis cannot. For instance, the correlation between SSD and PM2.5 concentration in Cangzhou in summer was not significant whilst there is a weak causality influence of SSD on PM2.5 concentrations detected. (3) Due to interactions between different meteorological factors, the value of correlation coefficients cannot interpret the quantitative influence of individual meteorological factors on PM2.5 concentration. Instead, the value from CCM method is designed to understand the coupling between two variables by excluding influences from other factors. A large correlation coefficient for one meteorological factor may correspond to a much smaller value from the CCM analysis. We found that the correlation coefficient between individual meteorological factors and PM2.5 concentrations was usually much larger than the value. This indicated that the causality of individual meteorological factors on PM2.5 concentrations was generally underestimated using the correlation analysis, due to the influences from other meteorological factors. So the value is a more reliable indicator for understanding and comparing quantitative influences of different individual meteorological factors than correlation coefficient. Due to limited spaces, in the previous manuscript, we did not give all information on the comparison between correlation analysis and the CCM method. But according to your suggestion, more details have been added to the revised manuscript.

pp 8 ln 221 – Are the results sensitive to the choice of parameters?

R: The CCM detects the bi-directional coupling between two variables highly automatically and only several parameters are required to run this model. However, the CCM result mainly depends on the time series data of the two variables and the several parameters mainly influence the presentation effects of the Convergent maps. We have tested different setting of these parameters, and the extracted value was simply the same, only the presentation of convergent maps were more smooth with different settings. So the results were not sensitive to the choice of parameters, which is also the major advantage of the CCM method. Since CCM method is not sensitive to the choice of parameters and the reliability of the causality analysis result has been proved by hundreds of studies from different ecosystems, this method has been widely employed.

pp 9 ln 237 – how is the value of determined? Are you calculating the limit or taking
the value at a specific time series length?

R: This is a good question. When we determine whether a curve is convergent or not, we set a $\Delta$ to represent the variation of value along with increasing time series length. If the $\Delta$ was less than a given threshold (e.g. 0.01) for a consecutive several date until the end of the time series, we consider that the was convergent to the p value of last date.

What about cases such as PM2.5 xmap minTEM, which (at least by eye) does not appear to be converging? Wouldn’t that suggest that minTEM was not influencing PM2.5 at all?

R: The CCM method considers strict causality influence of one variable on the other one and if the curve is convergent to 0 or demonstrates no generally convergent trend, then no causality influence exists based on long-term time series analysis of two variables. And for your instance, PM2.5 xmap minTEM demonstrate a generally convergent trend to a value approximating to 0.2. You may argue that this curve is not strictly convergent. However, the CCM method does not actually conducts a limit calculation and the CCM curve simply demonstrates a general trend of convergence. According to a diversity of instances provided by Sugihara et al. (2012), the curve shape of PM2.5 xmap minTEM can be regarded as a convergent curve, indicating a detected causality. Those instances from Sugihara et al. (2012) that demonstrated a non-convergent trend are near-linear shape or totally irregular shapes, which are quite different from the PM2.5 xmap minTEM. In summary, according to the PM2.5 xmap minTEM, minTEM exerted a weak influences on PM2.5 concentrations.

pp 9 ln 244 – It’s not clear here how PM is changing wind speed

R: High PM2.5 concentrations may lead to haze episodes, which usually result in generally stable atmospheric environment. In this case, the formation of winds, especially strong winds within this atmospheric environment are influenced significantly. And this is the reason why there are few winds within the urban areas during severe haze episodes. Meanwhile, winds across regions with haze episodes are also influenced notably. Yang et al. (2015) observed four haze episodes during Oct to Nov, 2014 and during these four haze episodes in the North China plain, the very high PM2.5 concentrations all led to stagnant condition and weak high-pressure systems, which further led to slowed wind speed and disturbed wind direction. This phenomenon was also observed by Liu et al. (2014) in haze episodes in Beijing in 2013. Very high PM2.5 concentrations induced haze episodes further led to stagnant and stable high-pressure systems, which made megacities served as obstacles to significantly slowed down the wind speed (Yang et al., 2015). Therefore, the effects of aerosols, especially high-concentration PM2.5 concentrations, prevented the wind occurrence mainly through indirect mechanisms.


pp 10 ln 270 - I would guess that (out of the 189 cities) those clustered regionally would show similar maps. Is that true? I.e. are the 37 cities shown representative of the cities not shown?

R: The main reason we selected some representative cities, instead of all cities for presenting wind rose maps is that there is limited space in the map. As one can see, it is already very crowded to demonstrate 37 wind roses and we should pay special attention to the selection of representative cities. Since the provincial capital in each province is the most important city (and usually the largest city) within the province and thus the meteorological influences on PM2.5 concentrations for the provincial capital
city usually receive the most emphasis. In this case, we selected the provincial capital for 31 provinces to present meteorological influences on PM2.5 concentration in each province, which considers both the spatial locations and the importance of the representative cities. For most provinces, especially provinces with low PM2.5 concentrations, the number of monitoring cities and the variations of PM2.5 concentrations are small. For regions with heavy air pollution (e.g., the number of monitoring cities and the variations of PM2.5 concentrations are large, so we selected 6 additional cities for these regions. In that case, we believe that the selective cities can be representative of no-shown cities within the same province.

pp 13 In 292-296 – This only seems to be true in some seasons.
R: Thanks so much for this point. We also noticed that this part should be described more rigorously. So in the revised manuscript, we have fully revised this part.

pp 13 In 299-300 - Can you quantify this more rigorously? By eye, it seems like there are enough outliers to call this in to question
R: Thanks for pointing this out. We realized that this conclusion may not apply to all regions in China. Instead, this pattern "the higher PM2.5 concentrations, the stronger meteorological influences on PM2.5 concentrations" was most obvious for the North China region, which is the most polluted region in China. In the revised manuscript, we have rephrased this statement to “For the North China region”. Thanks again for this suggestions.

pp 14 In 337-339 – This seems true for winter vs. summer, but what about spring vs autumn?
R: Again, thanks for pointing this out. Generally, PM2.5 concentrations for one specific city are highest in winter and the lowest in summer. So in the paper, when we mention the season when PM2.5 concentration is high (low), we mainly mean winter (summer). Meanwhile, the characteristics of PM2.5 concentrations in spring and autumn are not obvious, so we mainly compare the characteristics of PM2.5 concentrations in winter and summer, which provides the most important reference for the management of PM2.5 induced haze episodes in winter. So thanks again for this comments and we realized that we should make the statement clearer to avoid some confusion. In the revised manuscript, this sentence has been rephrased.

pp 18 Fig 4 caption – Is there a particular argument for only including the dominant factor in each city?
R: Thanks for this comment. As explained in the text, there are more than 189 cities for this research and it requires some space in the map to place the wind rose map for each city without overlapping with each other. And as you can see, the wind rose map for 37 representative cities has already been filled with 37 wind roses with different sizes and it is impossible to present 189 wind roses on a single map without severe overlapping effects. So we employed an alternative approach to only present the dominant meteorological factor for all cities and an entire wind rose for 37 representative cities in separate maps.

Section 5.1 of the discussion feels out of place. This is more of a discussion of what we already know about aerosol-meteorology interactions than a discussion of the implications of the work done in this study. I would recommend either rewriting it so that it builds more on the results from this paper, or cut it and integrate the important information into earlier sections.
R: Thanks so much for your comments. As also suggested by another referee, this part has been well known to scholars with relevant background. So in the revised manuscript, this part has been totally cut. Meanwhile, we have added a separate paragraph to explain the underlying reasons for large variations of meteorological influences

pp 24 In 562-566 – as per my earlier comment, I am not sure that you have shown this. If Beijing were to receive the same amount of precipitation that a coastal city does, is it possible that precipitation would become more important in Beijing? Does looking at
how a specific factor changed PM in a year tell us about how effective changes in that factor would be?

R: Thanks for pointing this out. I thought it is a bit difficult to test the hypothesis using the CCM method. This is because the precipitation amount for most days in Beijing is 0 and thus it is difficult to normalize the 0 value to a value similar to that in coastal cities. As an alternative solution, as responded above, we added another two years data to conduct a multiple-year analysis and checked whether the influence of precipitation on PM2.5 concentrations in Beijing vary with longer time series.

Technical Corrections pp 1, ln 17 (and later occurrences) - “causality influence” is redundant. “Influence” already implies causality. R: All Corrected.

pp 2, ln 35-36 - “Amongst these environmental elements, . . . concerned social and ecological issues.” The wording of this sentence is unclear. R: This sentence has been rephrased.

pp 2, ln 42 - “Serious haze not only influences peoples daily life,” this wording is vague. How does haze influence peoples daily lives? R: Serious haze episodes, usually presented as very thick and heavy black fogs, cause serious negative influences on people’s daily life, especially the traffic. During a severe haze episode in Beijing in January, the extreme high fog episodes led to very low visibility and heavy traffic jam. This is one instance how haze influence people’s daily life.

pp 2, ln 57 - “controversy” should be changed to “controversial” R: Corrected.

pp 3, ln 68 - “. . . fractions of three different sizes. . .” this is unclear. Authors should indicate that they are talking about aerosols, and specify the sizes. R: Thanks for pointing this out. We have corrected this in the revised manuscript.

pp 3, ln 7 - what region is this study referring to? R: The missing “research region” content has been added to the revised manuscript.

pp 3, ln 86 - word choice: I would suggest “well studied” instead of “massively studied”

R: Corrected.


pp 5, ln 136-137 – I’m not clear on what small and large evaporation are. R: Small evaporation indicates the evaporation amount calculated using a small-diameter measurement equipment and large evaporation indicates the evaporation amount calculated using a large-diameter measurement equipment. Generally, the amount calculated using the two types of equipment is usually the same, although slightly difference may exist between measured evaporation values. And the p value for the factor evaporation is decided by the larger value of the two indicators.

pp 5, ln 144 – sunshine duration for the day is a less widely used term and should be defined here R: Thanks so much for pointing this out. sunshine duration is a climatological indicator, measuring duration of sunshine in given period (usually, a day or a year) for a given location on Earth. This definition has been added to the modified manuscript.

R: Corrected.

pp 5 ln 146 – what qualifies as extreme wind speed? R: The max wind speed indicates the max mean wind speed during any 10 minutes within a day's time. The extreme wind speed indicates the max instant (for 1s) wind speed within a day's time. Thanks for pointing this out and the definition has been added to the modified manuscript.

pp 5 ln 146 – how is max wind direction defined? R: The max wind direction indicates the dominant wind direction for the period with the max wind speed. pp 6 ln 174 – “Two time series” is unnecessary. Suggest changing the sentence to “\(X=[X(1), \ldots X(L)]\) and \(Y=[Y(1), \ldots Y(L)]\) are defined as the temporal variations of variables \(X\) and \(Y\)” R: Corrected according to your suggestions.

pp 6 ln 175 – It’s unclear what \(r\) and \(S\) are. R: Thanks for pointing this out. \(r\) is the current position in the time series and the \(S\) presents the start position in the time series.
why can E be 2 or 3? R: This parameter decides at what dimension the CCM was calculated. When E equals 3, the calculation accuracy is higher. Through experiments, we found that the results were generally the same using the value of 2 or 3. For this research, we set the value of E 3 for a theoretically optimal CCM result.

the phrasing here is unclear. R: This sentence has been rephrased.

is there a way to test if these values are significant? R: The CCM method is different from the correlation analysis and another classic causality analysis method, Granger causality, which provides readers with the significance for the coupling between two variables. The CCM did not give us a value to present the significance for the revealed causality. However, while the Granger causality mainly revealed the qualitative causality between two variables, the p value from the CCM method revealed the quantitative causality between them. And the CCM method suggests that if the significance between two variables was not significant, then the calculated p value would be 0. So the p value was a direct metric for the quantitative influence and an indirect metric for the significance.

this wording is unclear R: This part has been removed in the revised manuscript.

Wikipedia is not an appropriate source. Better to cite a scientific paper that defines SSD. R: In the revised manuscript, the PM2.5-meteorology interaction part has been removed according to your and other referees’ comments. However, here we would like specifically add the explanation here that SSD could reduce PM2.5 concentrations not only through atmospheric photolysis, but also by enhancing surface temperature and promoting upward movement of aerosols.

Temperature inversion is certainly important, but none of the metrics in this study measure it directly. R: Thanks for pointing this out. Although this part has been removed in the revised manuscript, we would like to add some other mechanisms. Actually, in addition to temperature inversion, another important mechanism is that Temperature is closely related to pollutant concentrations by affecting atmospheric turbulence and chemical reactions. The temperature is positively correlated with pollutant concentrations in the majority of cities (He et al., 2017).


what about horizontal transport (advection)? R: Yes, horizontal transport of airborne pollutants should be major reason for the variation of PM2.5 concentrations. i.e. Anticyclones (i.e., high pressure systems) induced low wind speed was not favorable for the dispersion of pollutants. On the other hand, low pressure systems may lead to large wind speeds, favorable for the dispersion of PM2.5.

change “social economic” to “socio-economic” pp 22 In 492 – change “negative causality on” to “decreases” and “positive causality on” to “increases” R: Corrected.

do you have citations for this? R: We have added more relevant references to the revised manuscript according to your comments.

can you give more details/citations about the controversy? R: Some reports concerning different effects of this project has been added to the revised manuscript.

C15

that Temperature is closely related to pollutant concentrations by affecting atmospheric turbulence and chemical reactions. The temperature is positively correlated with pollutant concentrations in the majority of cities (He et al., 2017).

C16