Geophysical Research Letters

RESEARCH LETTER
10.1029/2020GL088548

The Influence of Internal Climate Variability on Projections of Synoptically Driven Beijing Haze

Christopher W. Callahan[1,2] and **Justin S. Mankin**[2,3,4]

[1]Program in Ecology, Evolution, Environment and Society, Dartmouth College, Hanover, NH, USA, [2]Department of Geography, Dartmouth College, Hanover, NH, USA, [3]Department of Earth Sciences, Dartmouth College, Hanover, NH, USA, [4]Ocean and Climate Physics, Lamont-Doherty Earth Observatory of Columbia University, New York, NY, USA

Key Points:
- Poor air quality in Beijing is linked to stagnant meteorology at the surface and anticyclonic circulation that promotes such conditions
- Climate models project increases in atmospheric conditions favorable to pollutant accumulation throughout the 21st century
- Internal variability strongly affects such projections, implying that previous projections have undersampled initial-condition uncertainty

Supporting Information:
- Supporting Information S1

Correspondence to:
C. W. Callahan,
christopher.w.callahan.gr@dartmouth.edu

Citation:
Callahan, C. W., & Mankin, J. S. (2020). The influence of internal climate variability on projections of synoptically driven Beijing haze. *Geophysical Research Letters*, *46*, e2020GL088548. https://doi.org/10.1029/2020GL088548

Received 27 APR 2020
Accepted 9 MAY 2020
Accepted article online 16 MAY 2020

Abstract Beijing has suffered a series of poor air quality ("haze") episodes which have damaged human health and economic growth. Beijing's haze is generally driven by pollutant-trapping meteorological conditions, but climate model projections differ on how these conditions will evolve under forcing. Such differences are driven in part by (1) disagreements over which meteorological conditions matter most and (2) an undersampling of internal variability in projections. To resolve these differences, we show that Beijing haze is associated with anticyclonic circulation, thereby linking multiple meteorological ingredients to a single process well simulated by models. We use this to inform our assessment of future haze risks in both a multimodel ensemble and a single-model large ensemble, partitioning model uncertainty and internal variability. We estimate that forcing increases haze-favorable conditions, but internal variability significantly affects these projections, emphasizing the importance of fully sampling sources of uncertainty to ensure accurate projections of haze risks.

Plain Language Summary Episodes of poor air quality ("haze") in Beijing, China have been linked to cardiovascular and respiratory disease, damage to crops, and decreases in economic growth. These episodes tend to occur when moist, still air lets pollutants accumulate. Human-caused climate change may affect the weather conditions that cause haze, but that effect is uncertain due to how scientists characterize haze-favorable weather, how such weather links to actual haze occurrence, and how those features change in climate model experiments that neither simulate pollutants explicitly nor represent the full distribution of climate variability ("internal variability"). Here we show that different weather conditions that promote haze in Beijing are linked to the same large-scale wind patterns, putting those conditions in the context of a single physical process simulated by climate models. When evaluated against this observed benchmark, we can confidently use multiple climate model ensembles to estimate how climate change affects haze-favorable conditions while also addressing uncertainties due to model choices and internal variability. We show that climate change will make haze-favorable conditions over Beijing more frequent, but that internal variability can generate large uncertainties in these projections, demonstrating the importance of fully sampling climate variability to constrain the risks of climate-induced health impacts.

1. Introduction

Poor outdoor air quality caused one in 13 deaths in 2016 globally (WHO, 2018). Events like Beijing's 2013 "airpocalypse" create a disease burden on society, reducing human well-being and economic vitality (Gao et al., 2015). In Beijing, "haze" occurs primarily during winter, when pollutant emissions combine with stagnant meteorology such as increased humidity and reduced ventilation. Pollutant emissions in the North China Plain have been relatively stable (Zheng et al., 2015), so daily variability in Beijing's air quality has been dominated by weather variation. These features of Beijing's air quality—its human toll and sensitivity to meteorology—mean that understanding how climate change may affect stagnation in Beijing is necessary to position health risk management and inform pollutant regulations. Previous research indicates that warming may increase haze-favorable meteorology *globally* (Dawson et al., 2014; Horton et al., 2014; Jacob & Winner, 2009; Mickley et al., 2004). Yet two uncertainties complicate this picture for Beijing: First, there is disagreement about the best way to measure haze-favorable meteorology, leading to

divergent estimates of its future change (cf. Cai et al., 2017; Shen et al., 2018); and second, internal climate variability likely plays an important, but as yet uncharacterized, role in shaping haze-favorable weather under forcing.

Many indices are used to describe the meteorology driving Beijing's haze episodes (e.g., Cai et al., 2017; Pei et al., 2018; Shen et al., 2018), with consequences for haze risk projections. Despite the agreement that haze is associated with a weakened East Asian Winter Monsoon (EAWM, Jia et al., 2015; Liu et al., 2017; Zhong et al., 2019), there is less agreement on which weather quantities create the most skillful index (cf. Cai et al., 2017; Wang et al., 2018; Zou et al., 2017). Cai et al. (2017) and Zou et al. (2017) suggest that climate change will weaken regional ventilation, Wang et al. (2015) suggest that Arctic sea ice decline will shift the storm track away from Beijing, and Chen et al. (2019) project decreases in ventilation and light precipitation; all four studies thus suggest an increase in haze. In contrast to these assessments, Shen et al. (2018) and Pendergrass et al. (2019) argue that there will be a negligible change to the key northwesterly winds that ventilate Beijing, and including relative humidity results in a projected decrease in haze (Pendergrass et al., 2019). So while there is some agreement about the factors that generate haze presently, there is less agreement about how these factors will shape haze risk under forcing.

Complicating haze projections further are poor constraints on the influence of internal climate variability (Garcia-Menendez et al., 2017; Pienkosz et al., 2019). Internal variability generates a range of weather trends consistent with the same model and emissions assumptions (Deser et al., 2012, 2014) and is undersampled in the multimodel ensembles that are typically used in climate projections (Kay et al., 2015; Mankin et al., 2017). Such limitations confound assessments of whether differences in haze projections are because of index choice, model structure, or poorly characterized internal variability. Moreover, haze indices differ widely in initial-condition ensembles (Callahan et al., 2019), but previous literature on climate and haze in Beijing has generally relied on multimodel ensembles without examining their sensitivity to internal variability (Cai et al., 2017; Jia et al., 2018; Shen et al., 2018; Wang et al., 2015). As a result, adaptation decisions like emissions limits may be made without proper benchmarks for the distribution of air quality conditions that are consistent with the same climate policy.

We pursue a three-step strategy to constrain the air quality risks of anthropogenic forcing in Beijing while accounting for internal variability: First, we assess the observed connections between haze, its driving meteorology, and synoptic circulation. Linking haze and its associated meteorology to regional circulation places disparate variables in the context of a single process that can be compared to climate models, which is our second step. Most global climate models do not explicitly simulate pollutants, but they provide information about climate variability unavailable from chemistry-focused models. Hence, we use observed synoptic drivers of haze-favorable meteorology for model validation. Finally, we assess how anthropogenic forcing affects haze-favorable conditions in Beijing, partitioning internal variability and structural uncertainty by using both single-model and multimodel ensembles. This approach clarifies how researchers could come to different conclusions about future haze risks in Beijing, reconciling divergent projections while providing a sound and transparent approach to using climate models to assess future air quality risks.

2. Data and Methods

2.1. Data

We quantify air quality using observations of particulate matter with a diameter less than 2.5 μm ($PM_{2.5}$) from the U.S. embassy in Beijing from 2008 through February 2020 (San Martini et al., 2015). We use meteorology data from the Integrated Global Radiosonde Archive (IGRA) at the Beijing station (Durre et al., 2016) from 1981 through February 2020 and 500 mb geopotential height (GPH) data from NCEP/DOE R2 reanalysis (Kanamitsu et al., 2002), with similar results using ERA-Interim GPH (Dee et al., 2011).

Climate model data are drawn from the CESM1 Large Ensemble (CESM1-LE, Kay et al., 2015) and 10 models with available data from the fifth phase of the Coupled Model Intercomparison Project (CMIP5, Taylor et al., 2012). CESM1-LE is a single-model microinitialized large ensemble: Each realization is initialized using roundoff-error-level differences (~10^{-14}) in atmospheric temperature (Kay et al., 2015), providing an estimate of CESM1's representation of atmosphere-induced internal variability. We use 35 realizations from CESM1-LE, with a historical experiment spanning 1979–2005 and an RCP8.5 experiment (Riahi et al., 2011)

from 2006 to 2100. We use the same experiments from CMIP5, the data from which we regrid to a common 2° × 2° resolution (Table S1 in the supporting information).

We conduct all analysis over boreal winter (December–February), as pollutant concentrations in the North China Plain are generally highest in winter and the relevant synoptic features are seasonally dependent (Chen & Wang, 2015).

2.2. Analysis

Our analytical framework is threefold: (1) quantify the observed linkages between synoptic circulation, local meteorology, and haze in Beijing; (2) compare circulation-meteorology linkages in observations and models; and (3) examine projections of those patterns to quantify how anthropogenic forcing and internal variability shape future haze risk.

To characterize regional circulation, we cluster 500 mb GPH anomalies into typical circulation patterns ("nodes") using self-organizing maps (SOM, Kohonen, 2001). SOM clustering has been widely used to analyze atmospheric circulation (Cassano et al., 2007; Gibson et al., 2017; Horton et al., 2015); further details are provided in Text S1 and Figure S1. Previous analyses examining circulation on haze days in Beijing (Chen & Wang, 2015; Li et al., 2018) have not objectively analyzed typical circulation patterns over all days, and SOM clustering provides an independent metric to assess multiple circulation patterns and how they affect air quality. We independently classify SOM nodes from R2 reanalysis, CESM1-LE, and CMIP5 over 1979–2016 and extend the analysis to 2017–2100 by matching projected GPH anomaly days with the most similar historically model-derived node (see Text S1 and Li et al., 2013; Singh et al., 2016; MacQueen, 1967). We can use these patterns to examine the associated states of meteorological variables ("ingredients") and air quality.

We select three ingredients from the IGRA data and each climate model to represent haze-favorable meteorology: specific humidity (*Q*), since water vapor promotes hydroscopic formation of pollutants (Zhong et al., 2018), 850 mb southerly wind anomalies (*V850*), representing the slowing of dry northwesterlies that ventilate Beijing (Pei et al., 2018), and positive environmental lapse rate anomalies (*ELR*), representing stable atmospheric conditions that prevent vertical ventilation, defined as the vertical temperature gradient between 850 and 1,000 mb (Zou et al., 2017). In the models, we average each value over the nine grid cells closest to Beijing (40°N, 116.5°E). We standardize each time series by subtracting the mean and normalizing by the standard deviation (SD) over 1981–2010.

These ingredients are correlated (Table S2), so we use principal components analysis (PCA) to produce orthogonal linear combinations of the ingredients, controlling for collinearity that would complicate multivariate regression. We focus on the first principal component ("PC1"), though we examine the other components in section 3.1. To estimate uncertainty in the PCs, we use a block bootstrap ($n = 10{,}000$), with the block size determined by the maximum autocorrelation length among the ingredients. We bootstrap all three ingredients together, preserving their co-occurrence, before recalculating the PCA within each iteration. We then associate PC1 magnitudes with $PM_{2.5}$ loadings and circulation pattern occurrence, linking circulation to meteorology to haze risk (Text S2). We use 3-day running means of the PCs and $PM_{2.5}$ to reduce random variation.

As a basis for observation-model comparison, we examine the joint occurrence of each SOM node and each of four PC1 bins (where "PC1 bin" refers to days with the following PC1 magnitudes: <−1, between −1 and 0, between 0 and 1, and >1 standard deviations) and then use the model projections of these joint occurrences to construct future haze risks. The combination of four nodes and four PC1 bins results in 16 total bins. The probability of haze in any period can be decomposed as follows:

$$P(Haze) = \sum_{i=1}^{n} HazeFraction_i * BinProbability_i$$

$HazeFraction_i$ refers to the fraction of haze days falling into each SOM node/PC1 bin pair, $BinProbability_i$ refers to the overall fraction of days that fall into each pair, and n refers to the total number of bins. We first compare observed and modeled bin probabilities over 1981–2010 to assess model representation of the circulation-meteorology linkage. To project haze under anthropogenic forcing, we assume that haze

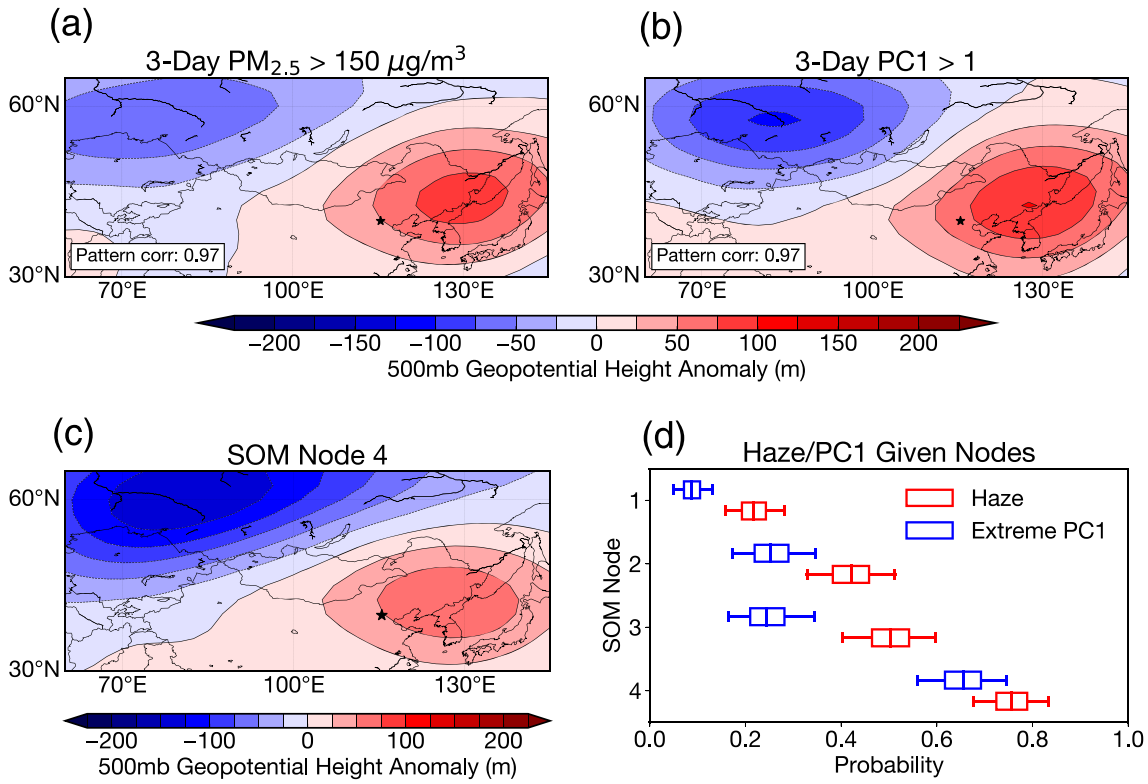

Figure 1. (a, b) 500 mb GPH composites on (a) high $PM_{2.5}$ days (>150 μg/m^3) and days (b) when PC1 > 1 SD, 2008–2020. Inset text denotes the pattern correlation between the composites in (a) and (b). (c) Node 4 from a four-node SOM configuration of 500 mb GPH from 1979–2016 (full SOM array shown in Figure S6). (d) Probability of haze ($PM_{2.5}$ > 100 μg/m^3) and elevated PC1 (PC1 > 0.5 SD) given the occurrence of each SOM node, calculated using Bayes' theorem (Text S2). Boxplots span the 5th–95th bootstrap percentiles. Probabilities are slightly reduced with higher thresholds (e.g., 150 μg/m^3 and 1 SD). In all maps, star denotes Beijing (40°N, 116.5°E) and domain is 30–65°N and 60–145°E.

probability given the occurrence of a certain bin ($HazeFraction_i$) is stationary and allow $BinProbability_i$ to vary according to projections of SOM node and PC1 change. To constrain the models with our observed probabilities, we calculate the projected change in bin probability relative to 1981–2010 in each model and add that change to the observed probability, allowing us to reconstruct projected haze probabilities.

We evaluate these projections at global mean temperature targets of 1.5°C, 2°C, and 3°C to evaluate the effects of climate policy on Beijing haze (Text S3). Global mean temperature depends on cumulative emissions but not emissions pathway (Allen et al., 2009), so this strategy reduces the sensitivity of our results to the RCP8.5 emissions scenario. Targets are defined as anomalies relative to the 1861–1900 mean (Figure S2), and we define the time period of each target as the 5 years before and after the first year in which temperature permanently exceeds the target (Text S3).

We quantify signal-to-noise ratio with the ratio of the ensemble mean to the ensemble standard deviation (Hawkins & Sutton, 2009; Mankin et al., 2017), with values above 1 considered statistically significant. To control for ensemble size (10 CMIP5 models vs. 35 CESM1-LE realizations) in ensemble comparisons, we sample 10 random realizations from CESM1-LE 100 times and present the median and standard deviation.

3. Results

3.1. Observed Haze-Meteorology-Circulation Linkages

There is a strong relationship between $PM_{2.5}$ concentrations and the first principal component of the meteorological ingredients ("PC1," Figure S3), consistent with research that haze is supported by increased humidity and reduced ventilation (Huang et al., 2017; Shen et al., 2018; Wang et al., 2018; Zou et al., 2017). The ingredients are strongly related, with covariance values greater than 0.4 (Table S2), supporting the use of PCA to control for multicollinearity and our hypothesis that such meteorology is driven by a shared

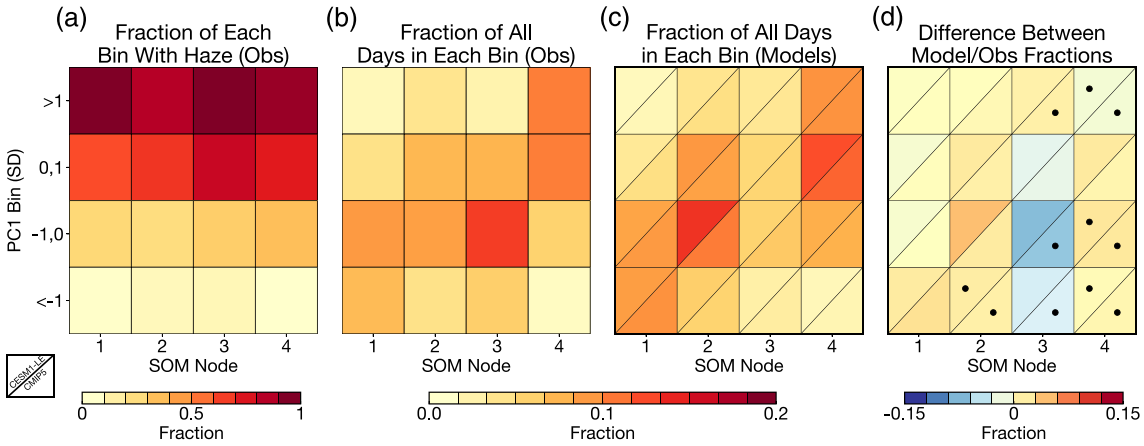

Figure 2. (a) Observed fraction (bootstrap median) of each node/PC1 bin with poor air quality ($PM_{2.5}$ > 100 μg/m^3), 2008–2020. (b) Observed fraction (bootstrap median) of all days falling into each bin, 1981–2010. (c) Ensemble mean fraction of all days falling into each bin, model years 1981–2010. Left triangles refer to CESM1-LE and right triangles to CMIP5 (see inset). (d) Difference between modeled and observed fractions. Dots represent statistically insignificant ($p > 0.005$) differences, determined by a Kolmogorov–Smirnov (K-S) test between the observed bootstrap distribution and each ensemble. CESM1-LE dots refer to the bootstrap median ($n = 100$) of K-S tests on 10 random realizations.

synoptic cause. PC1 explains ~65% of the variance among the ingredients (Figure S4) and is the only one of the three PCs strongly related to $PM_{2.5}$ (Figure S3), so we do not analyze the others further.

Anticyclones contribute to each ingredient and therefore haze formation by (1) creating subsidence inversions and warm air advection, increasing ELR (Chen & Wang, 2015; Zhao et al., 2013; Zhong et al., 2019); (2) indicating a weakened EAWM, reduced northwesterlies, and enhanced anomalous southerlies, increasing V850 (Chen & Wang, 2015; Pei et al., 2018; Zhong et al., 2019); and (3) allowing moisture to accumulate in Beijing due to weakened dry northwesterlies, increasing humidity (Jia et al., 2015; Zheng et al., 2015). Circulation composites on days with extreme $PM_{2.5}$ concentrations and PC1 values are consistent with these expectations (Figures 1a and 1b). In both composites, an anticyclonic anomaly along the Chinese coast and a cyclonic anomaly over Russia emerge (Figures 1c and 1d), with similar patterns appearing independently for each ingredient (Figure S5). This indicates that similar synoptic structures are present when pollutants actually accumulate (Figure 1a) and during pollutant-favorable meteorology (Figures 1b and S5).

SOM classification reveals a similar cyclonic-anticyclonic dipole (Figures 1c and S6) to the compositing analysis (Figures 1a and 1b), and the probability of haze is highest on days with this pattern ("node 4," Figure 1d). Poor air quality is more than three times more likely on node 4 days than node 1 days and ~50% more likely than on days associated with nodes 2 and 3 (Figure 1d). Node 4 is also more than twice as likely to generate extreme PC1 values than any other node, emphasizing that anticyclonic circulation generates local haze-favorable meteorology (Figure 1d). This independent metric of circulation provides a firm grounding for evaluation of circulation-meteorology linkages in climate models.

3.2. Modeled Circulation-Meteorology Linkages

Coupled global climate models generally do not interactively simulate pollutants, but they do simulate atmospheric circulation and the associated meteorology. We can therefore use observed relationships among air quality, meteorology, and circulation to infer future haze risks using the models, assuming that (1) pollutant emissions do not change and (2) anthropogenic forcing does not alter the relationship between air quality and meteorology.

The circulation-meteorology linkage can be expressed by grouping days into joint SOM node/PC1 bins (Figure 2). While it appears that any node can generate haze as long as PC1 is elevated (Figure 2a, red colored bins), the extreme PC1 anomalies that select for haze are more likely under the synoptic conditions associated with node 4 than any other node (Figures 1d and 2b). Both ensembles reproduce the pattern that greater PC1 value are more likely under node 4 and lesser PC1 values are more likely under nodes 1 and 2 (Figure 2c). The models have a slightly less stark partitioning of PC1 across the nodes (Figure 2d); both ensembles underestimate the association between node 1 and negative PC1 values and slightly

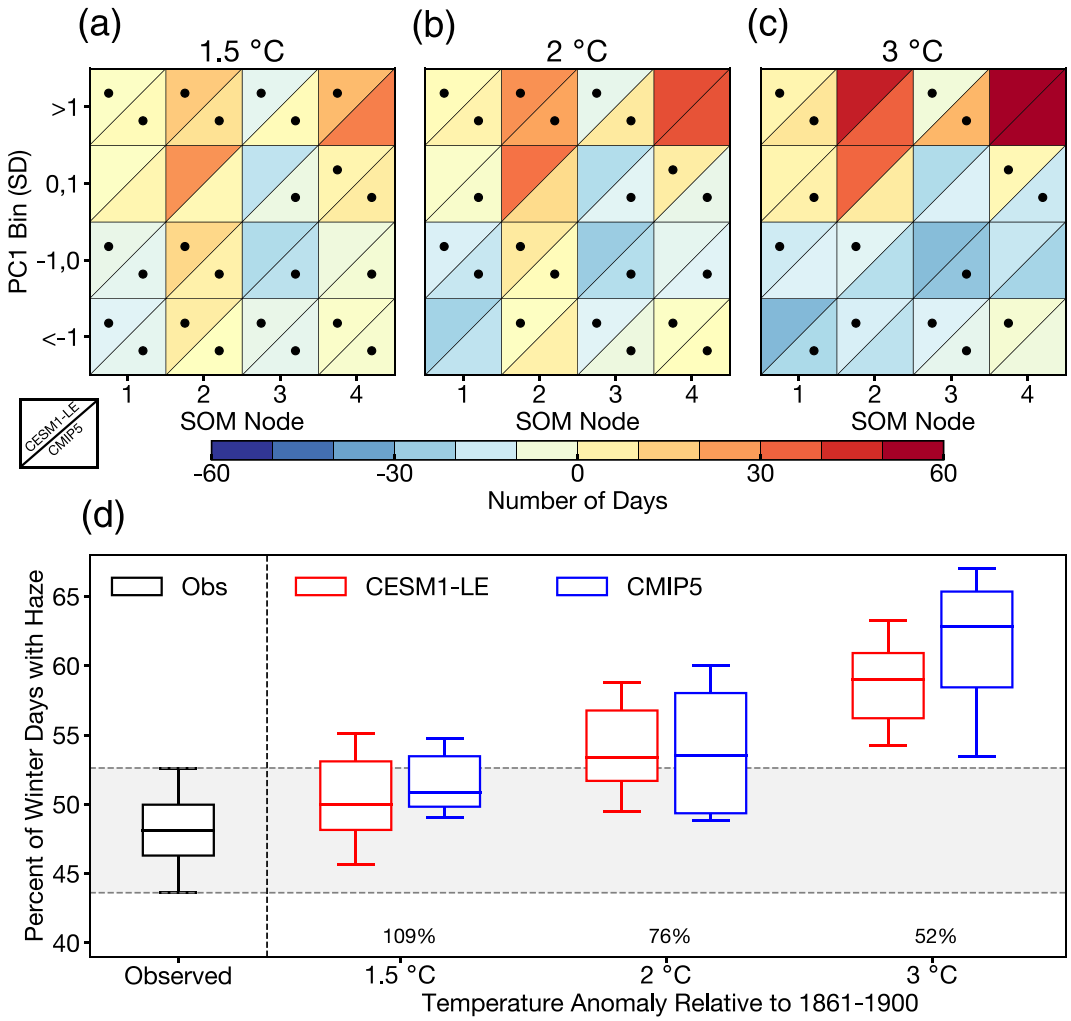

Figure 3. (a–c) Projected change in number of days allocated to each node/bin pair at each temperature target. Dots represent signal-to-noise ratios less than 1 (insignificant change; see section 2). (d) Projected haze probabilities at three temperature targets for CESM1-LE (red) and CMIP5 (blue). Gray shading shows the bootstrap 5th–95th percentiles, 1981–2010. Model boxplots span the 5th–95th percentiles of each ensemble. Inset numbers denote the percent of the CMIP5 range spanned by CESM1-LE.

overestimate the association between node 4 and PC1 values between −1 and 1. These differences are often statistically significant, though neither ensemble is statistically different than the observations for the node 4/extreme PC1 bin that is most favorable to haze (Figure 2d).

Our projection strategy involves calculating changes in modeled probabilities (Figure 2c) at three temperature targets and adding them to the observed probabilities (see section 2). This strategy constrains projections of haze-favorable conditions under anthropogenic forcing. We then multiply the projected probability of each pair by the probability of haze associated with that pair (Figure 2a), which we assume is stationary, thus reconstructing haze probabilities at each temperature benchmark.

3.3. Assessing Projections and the Role of Internal Variability

Anthropogenic warming increases the frequency of days with haze-favorable circulation and meteorology (Figures 3a–3c). In both the CESM1-LE and CMIP5 ensemble means, the node 4/extreme PC1 bin co-occurs on >40 additional days at 3°C than it does at 1.5°C (Figures 3a and 3c), an increase of 40–50%. Decreases are apparent in days allocated to nodes 1 and 2 and low PC1 values (Figure 3c), indicating an increase in haze-favorable days at the expense of days that would limit pollutant accumulation. However,

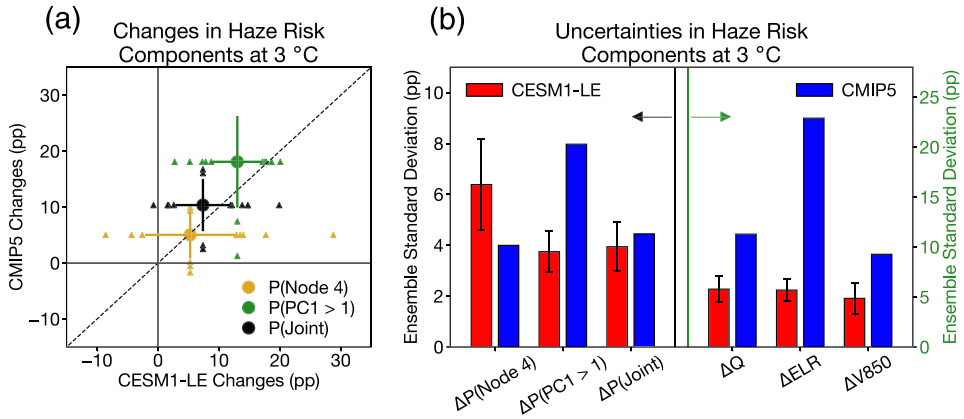

Figure 4. (a) Projected changes in the probabilities of node 4 occurrence (gold), extreme PC1 values (green), and the joint occurrence of node 4 and PC1 > 1 SD (black) at 3°C temperature change. The central dots represent the ensemble mean, bars denote SDs, and triangles denote ensemble members more than 1 SD from the mean. (b) Ensemble SDs in the three quantities in panel (a) (left *y*-axis), as well as the three meteorological ingredients (right *y*-axis). CESM1-LE bar height and errors bars correspond to the median and SD of 100 random samples of 10 realizations, matching the 10 CMIP5 models.

signal-to-noise ratios are small for many node/bin pairs at each temperature target; even with warming exceeding Paris Agreement targets, both internal variability and model structural uncertainty are large enough to mask forced changes in haze-favorable conditions (Figures 3a–3c).

Increases in haze-favorable circulation and meteorology translate into monotonic increases in haze probabilities (Figure 3d). With warming of 3°C, both ensembles project between 60% and 65% of wintertime days in Beijing to have poor air quality, 10–15 percentage points (pp) more than the present. Internal climate variability, however, plays a significant role in these projections. CESM1-LE spans more uncertainty than CMIP5 at 1.5°C and almost as much at 2°C, consistent with findings that internal variability is greater earlier in the century (Hawkins & Sutton, 2009) and indicating that previous air quality projections may have undersampled initial-condition uncertainty and overestimated the precision of their results. Moreover, there is overlap between the CESM1-LE distributions at the three temperature targets: almost 15% of CESM1-LE realizations have *greater* haze probability at 1.5°C than 2°C and more than 10% have greater haze probability at 2°C than 3°C. Hence, it is geophysically plausible for climate policy that is successful at limiting global temperature rise to in fact worsen air quality in Beijing, entirely due to unforced climate variability, despite the forced signal being a worsening of haze risk.

The drivers of haze risk change are presented in Figure 4. In both ensembles, anthropogenic forcing increases the probability of extreme PC1 values (P(PC1 > 1), Figure 4a), slightly more in CMIP5 than CESM1-LE. This PC1 change in both ensembles is due almost entirely to a secular, thermodynamic increase in specific humidity with warming (Figure S7). Neither ensemble projects large changes in ELR or V850 (Figure S7), consistent with Shen et al. (2018), though there is considerable ensemble variability in ELR. Furthermore, while both ensembles predict mean increases in haze-favorable node 4 occurrence (Figures 4a and S7), their representation of future circulation variability is large and spans the possibility for both multidecadal increases *and* decreases in node 4 occurrence. Uncertainty in node 4 changes in the CESM1-LE stems entirely from CESM's representation of internal variability and spans some 40 pp, from a decrease of 10 pp to an increase of 30 pp, all consistent with the same forcing. This range of circulation changes is more than three times the uncertainty in the multimodel CMIP5 (P(Node 4), Figure 4a). Our projection of increased haze risk is therefore driven primarily by a thermodynamic increase in atmospheric moisture along with a modest dynamical increase in anticyclone occurrence.

Using both single-model and multimodel ensembles allows us to explore the relative uncertainties associated with each component of haze risk. While changes in the joint probability of SOM node 4 and extreme PC1 values are similarly uncertain in both ensembles (ΔP(Joint), Figure 4b), internal variability substantially exceeds structural uncertainty in node 4 frequency projections (ΔP(Node 4), Figures 4a and 4b). However, structural uncertainty remains crucial in determining the extreme meteorology underpinning

future haze risks, being twice as large as that arising from internal variability ($\Delta P(PC1 > 1)$, Figure 4b). Structural uncertainty in PC1 is dominated by uncertainty in ELR (Figure 4b), which is the only ingredient that decreases under forcing in CMIP5 (Figure S7).

These results are a physically cogent partitioning of structural uncertainty and internal variability. For example, thermodynamic changes in the lapse rate likely depend more on model physics than model expressions of modes of variability (Pithan et al., 2014), suggesting a dominant role for intermodel differences in determining ELR changes. At the same time, it is physically plausible for internal variability to dominate projections of circulation change (Deser et al., 2014). Several mechanisms have been proposed to explain changes to anticyclone occurrence near Beijing, such as circulation anomalies induced by the loss of Arctic sea ice (Chen & Wang, 2015) or teleconnections to El Niño-Southern Oscillation (ENSO, Liu et al., 2017). Internal variability has been highlighted as a significant factor in historical and near-future sea ice loss (England et al., 2019), and it is becoming clear that changes to ENSO are difficult to constrain without many ensemble members for a single model (Maher et al., 2018). These mechanisms may partially explain the significant role of internal variability in CESM1-LE circulation projections.

4. Conclusions and Discussion

Our work addresses two uncertainties in projections of air quality and climate in Beijing: (1) the role of correlated ingredients that influence pollutant concentrations and (2) the influence of internal variability on projections of these conditions. Anticyclonic circulation explains co-occurring increased humidity and reduced ventilation, and cluster analysis can independently characterize this circulation, including the anticyclonic mode most favorable to haze in Beijing (Figure 1). This analysis provides an objective metric to compare representations of circulation in observations and models (Figure 2). Climate models indicate that anthropogenic forcing is likely to increase pollutant-favorable circulation and weather (Figures 3 and 4), but internal climate variability plays a significant role in shaping these projections (Figures 3 and 4). Benchmarking these projections to climate mitigation targets demonstrates the value of large ensembles as a policy evaluation tool, as we can isolate the role of climate variability in influencing mitigation policy success or failure.

Our projection of increased haze is consistent with some existing studies (Cai et al., 2017; Chen et al., 2019; Pei et al., 2018; Wang et al., 2015), but our work highlights the critical role of internal climate variability in these projections, which may have been undersampled in assessments making the opposite claim. Our results also reconcile other divergent projections of Beijing's air quality (cf. Cai et al., 2017; Shen et al., 2018). Unifying multiple ingredients with a single physical explanation allows us to account for changes in each ingredient and their driving circulation. Additionally, projections that use different indices and combinations of models may be confounded by internal variability, since uncertainty due to internal variability can surpass intermodel uncertainty in the early 21st century (Figure 3). Accounting for internal variability is therefore necessary to enable apples-to-apples comparisons between studies such as Cai et al. (2017) and Shen et al. (2018). More generally, our findings caution against relying on multimodel ensembles for projections of Beijing's air quality without incorporating the influence of internal variability, and we recommend studies examining future air quality use both multimodel and large ensembles.

The "climate penalty" on air quality may be one of anthropogenic warming's most direct influences on human health, and Beijing has seen calls for emissions controls to offset increases in meteorologically driven haze (Yin & Zhang, 2020). Our results indicate that in order to fully quantify the costs and benefits of decisions such as emissions regulations, decision makers must incorporate a robust treatment of internal variability. Incorporating this source of uncertainty may result in a wider spread of possible outcomes, but these outcomes will represent a more comprehensive account of Beijing's air quality future and enable robust decision making to counter this major health hazard.

Acknowledgments

We thank Dr. Daniel Horton and Dr. Jordan Schnell for discussions that improved the analysis. We thank the CESM1 Large Ensemble Community Project and supercomputing resources provided by NSF/CISL/Yellowstone for developing and hosting CESM1-LE. We thank Dartmouth Research Computing for high-performance computing resources. Processed data and analysis code are available in the GitHub repository *Callahan-Mankin_GRL_2020* at the following URL: doi.org/10.5281/zenodo.3820502.

References

Allen, M. R., Frame, D. J., Huntingford, C., Jones, C. D., Lowe, J. A., Meinshausen, M., & Meinshausen, N. (2009). Warming caused by cumulative carbon emissions towards the trillionth tonne. *Nature*, *458*(7242), 1163–1166. https://doi.org/10.1038/nature08019

Cai, W., Li, K., Liao, H., Wang, H., & Wu, L. (2017). Weather conditions conducive to Beijing severe haze more frequent under climate change. *Nature Climate Change*, *7*(4), 257–262. https://doi.org/10.1038/nclimate3249

Callahan, C. W., Schnell, J. L., & Horton, D. E. (2019). Multi-index attribution of extreme winter air quality in Beijing, China. *Journal of Geophysical Research: Atmospheres*, *124*, 4567–4583. https://doi.org/10.1029/2018JD029738

Cassano, J. J., Uotila, P., Lynch, A. H., & Cassano, E. N. (2007). Predicted changes in synoptic forcing of net precipitation in large arctic river basins during the 21st century. *Journal of Geophysical Research*, *112*, 63. https://doi.org/10.1029/2006JG000332

Chen, H., & Wang, H. (2015). Haze days in North China and the associated atmospheric circulations based on daily visible data from 1960 to 2012. *Journal of Geophysical Research: Atmospheres*, *120*, 5895–5909. https://doi.org/10.1002/2015JD023225

Chen, S., Guo, J., Song, L., Li, J., Liu, L., & Cohen, J. B. (2019). Inter-annual variation of the spring haze pollution over the North China Plain: Roles of atmospheric circulation and sea surface temperature. *International Journal of Climatology*, *39*(2), 783–798. https://doi.org/10.1002/joc.5842

Dawson, J., Bloomer, B., Winner, D., & Weaver, C. (2014). Understanding the meteorological drivers of U.S. particulate matter concentrations in a changing climate. *Bulletin of the American Meteorological Society, April*, *2014*, 521–532.

Dee, D. P., Uppala, S., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., et al. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, *137*(656), 553–597. https://doi.org/10.1002/qj.828

Deser, C., Knutti, R., Solomon, S., & Phillips, A. S. (2012). Communication of the role of natural variability in future North American climate. *Nature Climate Change*, *2*(11), 775–779. https://doi.org/10.1038/nclimate1562

Deser, C., Phillips, A. S., Alexander, M. A., & Smoliak, B. V. (2014). Projecting North American climate over the next 50 years: Uncertainty due to internal variability. *Journal of Climate*, *27*(6), 2271–2296. https://doi.org/10.1175/JCLI-D-13-00451.1

Durre, I., Vose, R., Yin, X., Applequist, S., & Arneld, J. (2016). Integrated Global Radiosonde Archive (IGRA) version 2. NOAA National Centers for Environmental Information, 10, V5X63X0Q.

England, M., Jahn, A., & Polvani, L. (2019). Nonuniform contribution of internal variability to recent arctic sea ice loss. *Journal of Climate*, *32*(13), 4039–4053. https://doi.org/10.1175/JCLI-D-18-0864.1

Gao, M., Guttikunda, S. K., Carmichael, G. R., Wang, Y., Liu, Z., Stanier, C. O., et al. (2015). Health impacts and economic losses assessment of the 2013 severe haze event in Beijing area. *Science of the Total Environment*, *511*, 553–561. https://doi.org/10.1016/j.scitotenv.2015.01.005

Garcia-Menendez, F., Monier, E., & Selin, N. E. (2017). The role of natural variability in projections of climate change impacts on us ozone pollution. *Geophysical Research Letters*, *44*, 2911–2921. https://doi.org/10.1002/2016GL071565

Gibson, P. B., Perkins-Kirkpatrick, S. E., Uotila, P., Pepler, A. S., & Alexander, L. V. (2017). On the use of self-organizing maps for studying climate extremes. *Journal of Geophysical Research: Atmospheres*, *122*, 3891–3903. https://doi.org/10.1002/2016JD026256

Hawkins, E., & Sutton, R. (2009). The potential to narrow uncertainty in regional climate predictions. *Bulletin of the American Meteorological Society*, *90*(8), 1095–1108. https://doi.org/10.1175/2009BAMS2607.1

Horton, D. E., Johnson, N. C., Singh, D., Swain, D. L., Rajaratnam, B., & Diffenbaugh, N. S. (2015). Contribution of changes in atmospheric circulation patterns to extreme temperature trends. *Nature*, *522*(7557), 465–469. https://doi.org/10.1038/nature14550

Horton, D. E., Skinner, C. B., Singh, D., & Diffenbaugh, N. S. (2014). Occurrence and persistence of future atmospheric stagnation events. *Nature Climate Change*, *4*(8), 698–703. https://doi.org/10.1038/nclimate2272

Huang, Q., Cai, X., Song, Y., & Zhu, T. (2017). Air stagnation in China (1985-2014): Climatological mean features and trends. *Atmospheric Chemistry and Physics*, *17*(12), 7793–7805. https://doi.org/10.5194/acp-17-7793-2017

Jacob, D., & Winner, D. (2009). Effect of climate change on air quality. *Atmospheric Environment*, *43*(1), 51–63. https://doi.org/10.1016/j.atmosenv.2008.09.051

Jia, B., Wang, Y., Huang, S., Nan, Y., & Zhou, X. (2018). Variations of Siberian High Position under climate change: Impacts on winter pollution over north China. *Atmospheric Environment*, *189*, 227–234. https://doi.org/10.1016/j.atmosenv.2018.06.045

Jia, B., Wang, Y., Yao, Y., & Xie, Y. (2015). A new indicator on the impact of large-scale circulation on wintertime particulate matter pollution over China. *Atmospheric Chemistry and Physics*, *15*(20), 11919–11929. https://doi.org/10.5194/acp-15-11919-2015

Kanamitsu, M., Ebisuzaki, W., Woollen, J., Yang, S.-K., Hnilo, J., Fiorino, M., & Potter, G. (2002). NCEP-DOE AMIP-II reanalysis (r-2). *Bulletin of the American Meteorological Society*, *83*(11), 1631–1644. https://doi.org/10.1175/BAMS-83-11-1631

Kay, J. E., Deser, C., Phillips, A., Mai, A., Hannay, C., Strand, G., et al. (2015). The Community Earth System Model (CESM) large ensemble project: A community resource for studying climate change in the presence of internal climate variability. *Bulletin of the American Meteorological Society*, *96*(8), 1333–1349. https://doi.org/10.1175/BAMS-D-13-00255.1

Kohonen, T. (2001). *Self-organizing maps* (Vol. 30). Berlin Heidelberg: Springer-Verlag.

Li, K., Liao, H., Cai, W., & Yang, Y. (2018). Attribution of anthropogenic influence on atmospheric patterns conducive to recent most severe haze over eastern China. *Geophysical Research Letters*, *45*, 2072–2081. https://doi.org/10.1002/2017GL076570

Li, L., Li, W., & Deng, Y. (2013). Summer rainfall variability over the Southeastern United States and its intensification in the 21st century as assessed by CMIP5 models. *Journal of Geophysical Research: Atmospheres*, *118*, 340–354. https://doi.org/10.1002/jgrd.50136

Liu, T., Gong, S., He, J., Yu, M., Wang, Q., Li, H., et al. (2017). Attributions of meteorological and emission factors to the 2015 winter severe haze pollution episodes in China's Jing-Jin-Ji area. *Atmospheric Chemistry and Physics*, *17*(4), 2971–2980. https://doi.org/10.5194/acp-17-2971-2017

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. *In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, *1*(14), 281–297.

Maher, N., Matei, D., Milinski, S., & Marotzke, J. (2018). ENSO change in climate projections: Forced response or internal variability? *Geophysical Research Letters*, *45*, 11,390–11,398. https://doi.org/10.1029/2018GL079764

Mankin, J. S., Viviroli, D., Mekonnen, M. M., Hoekstra, A. Y., Horton, R. M., Smerdon, J. E., & Diffenbaugh, N. S. (2017). Influence of internal variability on population exposure to hydroclimatic changes. *Environmental Research*, *12*(4), 044007.

Mickley, L. J., Jacob, D. J., Field, B., & Rind, D. (2004). Effects of future climate change on regional air pollution episodes in the United States. *Geophysical Research Letters*, *31*, L24103. https://doi.org/10.1029/2004GL021216

Pei, L., Yan, Z., Sun, Z., Miao, S., & Yao, Y. (2018). Increasing persistent haze in Beijing: Potential impacts of weakening East Asian winter monsoons associated with northwestern Pacific sea surface temperature trends. *Atmospheric Chemistry and Physics*, *18*(5), 3173–3183. https://doi.org/10.5194/acp-18-3173-2018

Pendergrass, D., Shen, L., Jacob, D., & Mickley, L. (2019). Predicting the impact of climate change on severe wintertime particulate pollution events in Beijing using extreme value theory. *Geophysical Research Letters*, *46*, 1824–1830. https://doi.org/10.1029/2018GL080102

Pienkosz, B., Saari, R. K., Monier, E., & Garcia-Menendez, F. (2019). Natural variability in projections of climate change impacts on fine particulate matter pollution. *Earth's Future*, *7*(7), 762–770. https://doi.org/10.1029/2019ef001195

Pithan, F., Medeiros, B., & Mauritsen, T. (2014). Mixed-phase clouds cause climate model biases in Arctic wintertime temperature inversions. *Climate Dynamics*, *43*(1–2), 289–303. https://doi.org/10.1007/s00382-013-1964-9

Riahi, K., Rao, S., Krey, V., Cho, C., Chirkov, V., Fischer, G., et al. (2011). RCP 8.5—A scenario of comparatively high greenhouse gas emissions. *Climatic Change*, *109*(1–2), 33–57. https://doi.org/10.1007/s10584-011-0149-y

San Martini, F., Hasenkopf, C., & Roberts, D. (2015). Statistical analysis of PM2.5 observations from diplomatic facilities in China. *Atmospheric Environment*, *110*, 174–185. https://doi.org/10.1016/j.atmosenv.2015.03.060

Shen, L., Jacob, D., Mickley, L., Wang, Y., & Zhang, Q. (2018). Insignificant effect of climate change on winter haze pollution in Beijing. *Atmospheric Chemistry and Physics*, *18*(23), 17489–17496. https://doi.org/10.5194/acp-18-17489-2018

Singh, D., Swain, D. L., Mankin, J. S., Horton, D. E., Thomas, L. N., Rajaratnam, B., & Diffenbaugh, N. S. (2016). Recent amplification of the North American winter temperature dipole. *Journal of Geophysical Research: Atmospheres*, *121*, 9911–9928. https://doi.org/10.1002/2016JD025116

Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, *93*(4), 485–498. https://doi.org/10.1175/BAMS-D-11-00094.1

Wang, H.-J., Chen, H.-P., & Liu, J. (2015). Arctic sea ice decline intensified haze pollution in eastern China. *Atmospheric and Oceanic Science Letters*, *8*, 1–9.

Wang, X., Dickinson, R., Su, L., Zhou, C., & Wang, K. (2018). PM2.5 pollution in China and how it has been exacerbated by terrain and meteorological conditions. *Bulletin of the American Meteorological Society, January*, *2018*, 105–119. https://doi.org/10.1175/bams-d-16-0301.1

WHO (2018). World Health Statistics 2018: Monitoring health for the SDGs, sustainable development goals. World Health Organization.

Yin, Z., & Zhang, Y. (2020). Climate anomalies contributed to the rebound of PM2.5 in winter 2018 under intensified regional air pollution preventions. *Science of the Total Environment*, *726*, 138514. https://doi.org/10.1016/j.scitotenv.2020.138514

Zhao, X., Zhao, P., Xu, J., Meng, W., Pu, W., Dong, F., et al. (2013). Analysis of a winter regional haze event and its formation mechanism in the North China Plain. *Atmospheric Chemistry and Physics*, *13*(11), 5685–5696. https://doi.org/10.5194/acp-13-5685-2013

Zheng, G., Duan, F., Su, H., Ma, Y., Cheng, Y., Zheng, B., et al. (2015). Exploring the severe winter haze in Beijing: The impact of synoptic weather, regional transport and heterogenous reactions. *Atmospheric Chemistry and Physics*, *15*(6), 2969–2983. https://doi.org/10.5194/acp-15-2969-2015

Zhong, J., Zhang, X., Dong, Y., Wang, Y., Liu, C., Wang, J., et al. (2018). Feedback effects of boundary-layer meteorological factors on cumulative explosive growth of PM2.5 during winter heavy pollution episodes in Beijing from 2013 to 2016. *Atmospheric Chemistry and Physics*, *18*(1), 247–258. https://doi.org/10.5194/acp-18-247-2018

Zhong, W., Yin, Z., & Wang, H. (2019). The relationship between the anticyclonic anomalies in Northeast Asia and severe haze in the Beijing-Tianjin-Hebei region. *Atmospheric Chemistry and Physics*, *19*(9), 5941–5957. https://doi.org/10.5194/acp-19-5941-2019

Zou, Y., Wang, Y., Zhang, Y., & Koo, J.-H. (2017). Arctic sea ice, Eurasia snow, and extreme winter haze in China. *Science Advances*, *3*(3), e1602751. https://doi.org/10.1126/sciadv.1602751