Improving the Design Stage of Air Pollution Studies Based On Wind Patterns

Léo Zabrocki: Paris School of Economics and École des Hautes Etudes en Sciences Sociales, 48 Boulevard Jourdan, 75014 Paris, France. Corresponding author. Email adress: leo.zabrocki@psemail.eu

Anna Alari: Barcelona Institute for Global Health (ISGlobal), Carrer del Rosselló, 132, 08036 Barcelona, Spain

Tarik Benmarhnia: Department of Family Medicine and Public Health & Scripps Institution of Oceanography, University of California, 8622 Kennel Way, La Jolla, San Diego, CA, USA

1 Abstract

2 A growing literature in economics and epidemiology has exploited changes in wind
3 patterns as a source of exogenous variation to better measure the acute health ef-
4 fects of air pollution. Since the distribution of wind components is not randomly
5 distributed over time and related to other weather parameters, multivariate regres-
6 sion models are often used to adjust for these confounding factors. However, this
7 type of analysis can rely on strong functional form assumptions and suffer from ex-
8 trapolation bias. As an alternative to current practices, we propose to implement
9 a causal inference pipeline to better approximate an hypothetical randomized ex-
10 periment. We illustrate this approach using daily data from Paris, France, over the
11 2008-2018 period. Using the Neyman-Rubin potential outcomes framework, we
12 first define the treatment of interest as the effect of North-East winds on particulate
13 matter concentrations compared to other wind directions. We then implement a
14 matching algorithm to approximate a pairwise randomized experiment. It adjusts
15 nonparametrically for observed confounders while avoiding model extrapolation by
16 discarding treated days without similar control days. Surprisingly, we find that the
17 effective sample size for which treated and control units are comparable is relatively
18 small. Once covariates balanced was achieved, we estimate the average differences
19 in air pollutant concentrations and their sampling variability using Neymanian in-
20 ference, a mode of inference specifically designed for randomized experiments. We
21 finally carry out a quantitative bias analysis to check whether our results could be
22 altered by an unmeasured confounder. Our causal inference pipeline highlights the
23 importance of checking covariates balance and bias from unmeasured confounders
24 in air pollution studies based on wind patterns.

1 Introduction

A growing literature in economics and epidemiology has recently re-examined the short-term effects of air pollution on mortality and emergency admissions using causal inference methods. Among these techniques, instrumental variable strategies have been very popular since they can overcome the bias caused by unmeasured confounders (1, 2, 3, 4, 5, 6). Daily changes in wind patterns have been used as instrumental variables since they create plausibly exogenous variation in air pollution that could be less affected by unmeasured confounders than raw concentrations. As a case study, imagine that we are interested in estimating the influence of North-East winds on local air pollutant concentrations in Paris, France. Research in atmospheric science has shown that winds blowing from the North-East could transport particulate matters due to wood burning in the region but also from other sources in North-Western Europe (7, 8, 9).

Figure 1: Polar Plots of Air Pollutant Concentrations Predicted by Wind Components.

Notes: Each panel represents the concentrations (in $\mu\text{g}/\text{m}^3$) of an air pollutant that were predicted using a generalized additive model based on a smooth isotropic function of the two wind components u and v (10). The direction from which the wind blows is described on a 360° compass rose and wind speed (in m/s) is represented by a series of increasing circle starting from the intersection of the two cardinal directions axis where wind speed is null: the farther the circle is away from the intersection, the faster the wind speed is.

In Figure 1, we display polar plots of air pollutant concentrations that were pre-

dicted using a Generalized Additive Model (GAM) and wind components as inputs (10). We see that wind blowing from the North-East are associated with higher PM_{10} and $PM_{2.5}$ concentrations. These patterns could however be confounded by other variables such as the average temperature or a shared seasonality in air pollution and wind patterns. We must therefore take into account these confounding variables to credibly measure the influence of North-East winds on particulate matters. Multivariate linear regression or more flexible methods such as generalized additive models and machine learning algorithms can help limit confounding bias (11, 12). Yet, current practices have not taken advantage of well-established recommendations found in the causal inference literature to better analyze observational data (13, 14).

In this paper, we propose to use a causal inference pipeline as an alternative approach to illustrate its advantages for air pollution studies. Our goal is to improve the credibility of an observational study by embedding it within an hypothetical randomized experiment, also referred to as target trials (15, 13, 16). In this context, we follow the four consecutive stages proposed by (17, 18). At each step of the procedure, assumptions are explicitly stated so that researchers can easily discuss whether they are deemed plausible.

First, in a *conceptual stage*, we clearly state the causal question of interest using the Neyman-Rubin potential outcomes framework (19, 20). Our treatment of interest is the effect of North-East winds on air pollution compared to other wind directions. To estimate this effect, we need to impute for treated days with winds blowing from the North-East the concentrations that would have been observed had winds blown from other directions. The issue is that wind patterns are not randomly assigned: control days with wind blowing from other direction are not similar to treated days.

We therefore implement a *design stage* where we approximate a pairwise ran-

domized experiment using a matching algorithm recently designed for air pollution studies (21). Matching is a transparent method to adjust for confounders without making parametric assumption and looking at observed outcomes (22, 23). Given a set of chosen covariates distances, each treated day is matched to its closet control day. This method avoids model extrapolation since treated days for which no control days exist in the data are discarded from the analysis.

The third step is an *analysis stage* where we estimate the influence of North-East winds on air pollutant concentrations. We simply compute the average difference in concentrations between matched treated and control days and rely on Neymanian inference to compute an estimate of the sampling variability (14). The last and fourth step is to carry out a *sensitivity analysis*. Throughout the previous steps, we must make the strong assumption that no unmeasured variables could be related both to wind patterns and air pollutant concentrations. Quantitative bias analysis was initially proposed by (24) to assess which magnitude of hidden bias would be required to alter observed results. We follow here the method developed by (13) and (25).

With this study, we aim to contribute to two literatures. First, we show that using wind as instrumental variables requires much more caution to make convincing the assumption that wind are “as-if” randomly distributed according to observed and unobserved variables. The effective sample where treated and control units are similar on a set of observed covariates is relatively small. Our quantitative bias analysis also reveals that the estimated increase in particulate matter concentrations due to North-East winds is qualitatively robust but becomes much less informative if an unobserved confounding factor is 2 times more common among days with winds blowing from the North-East than other directions.

Second, the fact that wind patterns play a key role in the variation of air pollution concentrations is obviously not new (26, 27, 28, 29). Causal inference methods

103 have rarely been implemented by atmospheric scientists to answer questions that
104 are causal by nature, such as the influence of weather parameters on air pollution.
105 We believe that mimicking a randomized experiment corresponds to an intuitive
106 approach and could complement source apportionment and emission inventory ap-
107 proaches. While wind is non manipulable, emission sources are and our framework
108 could also serve as a stepping-stone to evaluate potential interventions to control
109 emissions—if a source is shut-down in the North-East of Paris, would wind blowing
110 from this direction influence less specific air pollutant concentrations?

111 We took great care to make our work fully reproducible to help researchers im-
112 plement but also improve and criticize our approach. Data and detailed **R** codes are
113 available at the following address: <https://osf.io/7x23u/>.

104 2 Methods

105 2.1 Data

106 We built a dataset combining daily time series of air pollutant concentrations and
107 weather parameters in Paris over the 2008-2018 period. We choose to carry out
108 an analysis at the daily level as it done in studies on the acute health effects of air
109 pollution (3, 4, 6).

110 First, we obtained hourly air quality data from AirParif, the local air quality
111 surveillance organization. Figure 2 displays the location of the selected measuring
112 stations. Using a 2.5% trimmed mean, we first averaged at the daily level the con-
113 centrations ($\mu\text{g}/\text{m}^3$) of background measuring stations for NO_2 , O_3 and PM_{10} . For
114 a given day, if more than 3 hourly readings were missing, the average daily con-
115 centration was set to missing. The proportion of missing values for stations ranged
116 from 2.8% up to 9.1%. We also computed the average daily concentrations of $\text{PM}_{2.5}$

Figure 2: Map of road network and location of air pollution measuring stations in Paris, France.

Notes: Grey lines represent the road network. The orange line is the orbital ring surrounding Paris. Blue crosses are the locations of air pollution measuring stations. NO₂ concentrations are measured at stations PA07, PA12, PA13, PA18; O₃ concentrations at PA13, PA18; PM₁₀ at PA18; PM_{2.5} at PA01H and PA04C. The map was created with the R programming language (version 4.1.0) (30), data were provided by OpenStreetMap (31) and retrieved with the osmdata package (32).

117 but 25% of the recordings were missing: the air pollutant stopped being recorded
 118 between 2009/09/22 and 2010/06/23. It is important to note that we only focus on
 119 background monitors as they are used to assess the residential exposure of a city
 120 population in epidemiological studies.

121 We then retrieved meteorological data from the single monitoring station lo-
 122 cated in the South of the city and ran by the French national meteorological service
 123 Météo-France. We extracted daily observations on wind speed (m/s), wind direction
 124 (measured on a 360° wind rose where 0° is the true North), the average temperature

125 (°C), and the rainfall duration (min). Weather parameters had very few missing
126 values (e.g., at most 2.5% of observations were missing for the rainfall duration).

127 Finally, to avoid working with a reduced sample size, we imputed missing values
128 for all variables but $\text{PM}_{2.5}$. There were no clear patterns in the missingness of NO_2 ,
129 O_3 and PM_{10} concentrations. We used the chained random forest algorithm imple-
130 mented by the **R** package `missRanger` (33). A small simulation exercise showed that
131 it had good performance for imputing NO_2 concentrations (the absolute difference
132 between observed and imputed values was equal to $3.2 \mu\text{g}/\text{m}^3$ for an average con-
133 centration of $37.6 \mu\text{g}/\text{m}^3$) but was much less effective for imputing PM_{10} concen-
134 trations (the absolute difference between observed and imputed values was equal
135 to $6.1 \mu\text{g}/\text{m}^3$ for an average concentration of $23.4 \mu\text{g}/\text{m}^3$). As a robustness check,
136 we also provide results of the proposed method we advocated in this article for a
137 complete-case analysis. Once the data were imputed, we averaged the air pollutant
138 concentrations at the city level as it the spatial level of analysis used in (3, 4, 6).

139 Further details on the data wrangling procedure and an exploratory analysis
140 of the data can be found in the replication materials available at [https://osf.io/](https://osf.io/7x23u/)
141 [7x23u/](https://osf.io/7x23u/).

142 2.2 A Causal Inference Pipeline

143 We present below the four stages of the causal inference pipeline we advocate to
144 use for improving the design of air pollution studies based on wind patterns. Its
145 implementation was done with the R programming language (version 4.1.0) (30).

146 Stage 1: Defining the Treatment of Interest

147 The first step of our causal inference approach is to clearly state the question we are
148 trying to answer: *What is the effect of North-East winds on air pollutant concentrations*
149 *in Paris over the 2008-2018 period?* Our treatment of interest is the comparison of

150 air pollutant concentrations when winds are blowing from the North-East (10° - 90°)
 151 with concentrations when wind come from other directions. We frame this question
 152 in the Rubin-Neyman causal framework (19, 20). Our units are 4,018 days indexed
 153 by i ($i=1, \dots, I$). For each day, we define our treatment indicator W_i which takes
 154 two values. It is equal to 1 if the unit is treated (the wind blows from the North-
 155 East), and 0 if the unit belongs to the control group (the wind is blowing from
 156 another direction). Under the Stable Unit Treatment Value Assumption (STUVA),
 157 we assume that each day can have two potential concentrations in $\mu\text{g}/\text{m}^3$ for an air
 158 pollutant: $Y_i(1)$ if the wind blows from the North-East and $Y_i(0)$ if the wind blows
 159 from another direction.

160 The fundamental problem of causal inference states that we can only observe
 161 for each day one of these two potential outcomes: it is a missing data problem
 162 (34, 35). The observed concentration of an air pollutant Y^{obs} is defined as $Y^{\text{obs}} =$
 163 $(1-W_i) \times Y_i(0) + W_i \times Y_i(1)$. If the unit is treated, we observe $Y_i(1)$. If it is a control,
 164 we observe $Y_i(0)$. To estimate the effect of North-East winds on air pollutant con-
 165 centrations, we therefore need to impute the missing potential outcomes of treated
 166 units—what would have been the air pollutant concentrations if the wind had blown
 167 from another direction?

168 **Stage 2: Designing the Hypothetical Randomized Experiment**

169 The second stage of our causal inference pipeline is to embed our non-randomized
 170 study within an hypothetical randomized experiment. We are dealing with an ob-
 171 servational study where North-East winds are not randomly distributed through
 172 a year and are correlated with other weather parameters influencing air pollutant
 173 concentrations. In Figure 3, we plot, for each month, the absolute standardized
 174 mean differences between treated and control units for the average temperature,
 175 relative humidity and wind speed: most differences are superior to 0.1, which is

176 often considered as a threshold to assess the imbalance of covariates.

Figure 3: Evidence of Imbalance for Weather covariates.

Notes: For each month, we compute the absolute standardized differences in a continuous weather covariates between treated and control groups. These differences are represented as blue points. The vertical orange line is the 0.1 threshold which is used in the matching literature to spot covariates imbalance. The vertical black line is at 0.

177 To better approximate a randomized experiment, we must therefore find the sub-
 178 set of treated units which are similar to control units. Formally, we want to make
 179 plausible for this subset of units the assumption that the treatment assignment is
 180 independent from the potential outcomes of units given their covariates $\mathbf{X}$: $\Pr(\mathbf{W}$
 181 $|\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = \Pr(\mathbf{W} | \mathbf{X})$. The issue is that some units' covariates are observed
 182 while others are not. Contrary to a randomized experiment where both observed
 183 and unobserved covariates will be, on average, balanced across treatment and con-
 184 trol groups, we must assume that no unobserved covariates affect the treatment
 185 assignment.

186 Matching methods are particularly convenient to design hypothetical random-

187 ized experiments. Contrary to standard regression approaches, matching is non-
188 parametric way to adjust for observed covariates while avoiding model extrapola-
189 tion since units without counterfactuals in the data are discarded from the analysis.
190 Specifically, we use a constrained matching algorithm to design a pairwise random-
191 ized experiment where, for each pair, the probability of receiving the treatment is
192 equal to 0.5 (see (21) for further details on the algorithm). Each treated unit is
193 matched to its closest unit given a set of covariates constraints which represent the
194 maximum distance, for each covariates, allowed between treated and control units.
195 We match on the two sets of covariates influencing both wind directions and air
196 pollutant concentrations.

197 First, we match on calendar variables such as the Julian date, weekend, holidays
198 and bank days indicators. A treated unit could be matched up to a control unit
199 with a maximum distance of 60 days. If we extend this distance, it would be easier
200 to match treated units to control units but the treatment effect could be biased by
201 seasonal variation in air pollutant concentrations. We match exactly treated and
202 control units for the other calendar indicators. Second, we match on weather vari-
203 ables. The average temperature between treated and control units could not differ
204 more than 5°. The difference in wind speed should be less than 0.5 m/s. The rainfall
205 duration (divided in four ordinal categories) needs to be the same and the absolute
206 difference in average humidity could be up to 12 percentage points. We also force
207 the absolute difference in PM_{10} concentrations in the previous day to be less or equal
208 to $8 \mu g/m^3$. The thresholds we set up were chosen through an iterative process were
209 we check (i) that they lead to balanced sample of treated and control units and (ii)
210 that there were enough matched pairs to draw our inference upon.

Stage 3: Analyzing the Experiment using Neymanian Inference

In the third stage, we proceed to the analysis of our hypothetical pairwise randomized experiment. Several modes of statistical inference such as Fisherian, Neymanian, Bayesian or the standard model-based approaches could be implemented (36). Here, we take a Neymanian perspective where the potential outcomes are assumed to be fixed and the treatment assignment is the basis of inference. Our goal is to measure the average causal effect for the sample of matched units. We assume that each of the two units of a matched pair j has two potential concentrations for an air pollutant. If we were able to observe these potential outcomes, we could simply measure the effect of North-East winds on air pollutant concentrations by computing the finite-sample average treatment effect for matched treated units τ_{fs} . We would first compute for each pair the mean difference in concentrations and then average the differences over the J pairs. While we only observe one potential outcome for each unit, we can nonetheless estimate τ_{fs} with the average of observed pair differences $\hat{\tau}$:

$$\hat{\tau} = \frac{1}{J} \sum_{j=1}^J (Y_{t,j}^{\text{obs}} - Y_{c,j}^{\text{obs}}) = \bar{Y}_t^{\text{obs}} - \bar{Y}_c^{\text{obs}}$$

Here, the subscripts t and c respectively indicate if the unit in a given pair is treated or not. Since there are only one treated and one control unit within each pair, it is not statistically possible to compute the exact estimator of the variance of the average of pair differences. We can however compute a conservative estimate of the variance (14):

$$\hat{V}(\hat{\tau}) = \frac{1}{J(J-1)} \sum_{j=1}^J (Y_{t,j}^{\text{obs}} - Y_{c,j}^{\text{obs}} - \hat{\tau})^2$$

We finally compute an asymptotic 95% confidence interval using a Gaussian distri-

232 bution approximation:

$$CI_{0.95}(\tau_{fs}) = \left(\hat{\tau} - 1.96 \times \sqrt{\hat{V}(\hat{\tau})}, \hat{\tau} + 1.96 \times \sqrt{\hat{V}(\hat{\tau})} \right)$$

233 This 95% confidence intervals gives the set of effect sizes compatible with our data
234 (37).

235 **Stage 4: Sensitivity Analysis**

236 The fourth step of our causal inference pipeline is to explore how sensitive our anal-
237 ysis is to violation of the assumptions it relies upon. We carry out three types of
238 robustness checks.

239 First, we make the strong assumption that the treatment assignment is ignor-
240 able: winds blowing from the North-East occur randomly conditional on a set of
241 measured covariates. Other researchers could however argue that we fail to adjust
242 for unmeasured variables influencing both the occurrence of North-East winds and
243 air pollutant concentrations. Within matched pairs, these unobserved counfounders
244 could make the treated day more likely to have wind blowing from the North-East
245 than the control day. We therefore implement the quantitative bias analysis devel-
246 oped by (13) and (25). It allows us to explore how our results would be altered
247 by the effect of an unobserved confounder on the treatment odds, denoted by Γ .
248 For instance, if an unmeasured confounder has a small effect on the odds of treat-
249 ment (i.e., for a Γ close to 1) but the resulting 95% confidence intervals become
250 completely uninformative, it would imply that our results are highly sensitive to
251 hidden bias. In a complementary manner, we also check whether unmeasured bi-
252 ases could be present by using the first daily lags of air pollutant concentrations as
253 control outcomes (38). If our matched pairs are indeed similar in terms of unob-
254 served covariates, the treatment occurring in t should not influence concentration
255 of air pollutants in $t - 1$.

256 Second, for many matched pairs, air pollutant concentrations were imputed us-
257 ing the chained random forest algorithm (33). We check whether the results are
258 sensitive to the imputation by re-running the analysis for the non-missing concen-
259 trations.

260 Third, we make sure that the treatment assignment within pair was effective to
261 increase the precision of estimates. We compare the estimate of the sampling vari-
262 ance of a pair randomized experiment to the one of a completely randomized ex-
263 periment. If the estimate of sampling variability for the pair experiment is smaller
264 than the estimate of sampling variability for a complete experiment, it means that
265 our matching procedure was successful to match similar units within pairs com-
266 pared to randomly selected units (14).

267 3 Results

268 3.1 Performance of the Matching Procedure

269 Our initial dataset consists in 4,018 daily observations, divided into 912 treated
270 units and 3,106 control units. The matching procedure results in 271 pairs of
271 matched treated-control units—only 30% of treated units could be matched to sim-
272 ilar control units given the constraints we set. In the supplementary materials (),we
273 show that the matched sample has different characteristics from the initial sample:
274 days belong more the period ranging from May to October, their average tempera-
275 ture is higher and their relative humidity lower.

276 In Figure 4, we display how the balance of continuous and categorical covariates
277 improves after the matching procedure. Blue dots represent either the absolute
278 mean differences between treated and control units for continuous variables or the
279 absolute difference in percentage points for categorical variables. For continuous

280 covariates, the average standardized mean differences between treated and control
 281 days is 0.26 before matching and reduces to 0.03 after the procedure. For categorical
 282 covariates, the average difference in percentage points diminishes from 6.2 to 1.8
 283 after matching. Our matching procedure therefore leads to a consequent reduction
 284 of our sample size but allows us to compare treated units that are more similar to
 285 control units.

Figure 4: Overall Balance Improvement in Continuous and Categorical covariates.

Notes: In Panel A, we plot, before and after matching, the absolute standardized differences in continuous covariates between treated and control groups. Each blue dot represents an absolute mean difference for a given covariates. In panel B, we plot, before and after matching, the absolute difference in percentage points for categorical covariates.

3.2 North-East Wind Effects on Air Pollutant Concentrations

For each air pollutant, we plot in [Figure 5](#) the average difference in concentration ($\mu\text{g}/\text{m}^3$) between North-East winds and other wind directions. We also display the differences for the previous day and the following day. Thick lines represent the 95% confidence intervals while thin lines are the 99% confidence intervals. The third panel of [Figure 5](#) confirms the exploratory analysis of the polar plot. When wind blows from the North-East, PM_{10} concentrations increase by $4.3 \mu\text{g}/\text{m}^3$, with the lower and upper bounds of the 95% confidence being respectively equal to an increase by $2.5 \mu\text{g}/\text{m}^3$ and $6 \mu\text{g}/\text{m}^3$. The estimated difference represents a 17% increase in the average concentration of PM_{10} . We also observe a positive difference of 25% in PM_{10} concentrations the following day (point estimate of 6.2; 95% CI: 3.9, 8.4).

North-East winds do not seem to influence NO_2 (point estimate of -0.1; 95% CI: -3.4, 3.2), and O_3 (point estimate of 0; 95% CI: -2.7, 2.7) concentrations on the current day. Yet, there could be a positive influence of North-East winds on the following day for O_3 (point estimate of 2.8; 95% CI: -1, 6.5). The point estimate for the average difference in O_3 concentrations would represent a 7% relative increase.

Regarding the effects of North-East winds on $\text{PM}_{2.5}$, we restrain our analysis to pairs without missing concentrations. For the current and following days, we respectively find an average increase of $2.3 \mu\text{g}/\text{m}^3$ (95% CI: 1, 3.6) and $5.4 \mu\text{g}/\text{m}^3$ (95% CI: 3.9, 6.9). These point estimates respectively represent a 15% and a 34% relative increases in $\text{PM}_{2.5}$ concentrations.

3.3 Sensitivity Analysis

Our quantitative bias analysis reveals that if we have failed to adjust for an unobserved confounder 2 times more common among treated days, the resulting 95%

Figure 5: Effects of North-East Winds on Air Pollutant Concentrations.

Notes: In each panel, we plots the estimated effects of North-East winds on air pollutant concentrations for the previous, current and following days. Point estimates are depicted by blue points; blue thick lines are 95% confidence intervals and thin lines are 99% confidence intervals.

confidence intervals for the estimated effects of North-East winds on PM_{10} would be equal to (0.8, 7.8) for the current day and to (0.6, 10.9) for the the following day. Confidence intervals are still consistent with positive effects but are relatively wide. Potential candidates for such unobserved confounders could be lags of covariates we did not adjust for or a weather parameter for which we have not retrieved the recordings, such as atmospheric pressure. As a complementary test for unobserved confounders, we also check that the occurrence of North-East winds on the current day do not have any effect on concentrations measured in the previous day. Reassuringly, for all fourth pollutants, 95% confidence intervals do not suggest clear negative or positive average differences in concentrations as shown in Figure 5 (for $PM_{2.5}$, the estimated average difference is $0.3 \mu g/m^3$ (95% CI: -0.4, 0.9).

In the supplementary materials, we check whether the imputation of missing air pollutant concentrations did not drive our results. For NO_2 , O_3 and PM_{10} , 27%,

18% and 19% of concentrations were imputed. We replicate our analysis on the subset of pairs without missing observations: point estimates remain very similar but confidence intervals are a bit larger due to the sample size loss. This robustness check implies that our imputation did not bias our estimates.

Finally, we confirm that the pairwise design of our hypothetical experiment does increase precision for O_3 and PM_{10} . Compared to a complete randomized experiment, the pairwise standard errors decrease by 30% for O_3 and 11% for PM_{10} . However, our matched pair design leads to a loss in precision of 40% for NO_2 : if we could have adjusted for the previous daily concentrations of NO_2 , we could perhaps have increased precision.

4 Discussion

Following our causal inference pipeline, we tried to approximate an hypothetical experiment for measuring the effects of North-East winds on daily air pollutant concentrations in Paris. In the first stage of the pipeline, we had to specify what the treatment of interest was, that is to say the difference of air pollutant concentrations between North-East winds and other wind directions. If this comparison is easy to understand, the treatment we defined is not manipulable contrary to randomized controlled trials. It might therefore lack a certain appeal to policy-makers as our estimates only indicate whether North-East winds lead to higher local air pollutant concentrations than other wind directions. Our preliminary results provide however a stepping stone to investigate the sources emitting air pollution. Our causal inference pipeline would be particularly suited to analyze quasi-natural experiments based on closures of industrial plants nearby the city or changes in transportation networks (39, 40).

In the second stage, we built the hypothetical experiment using a constrained

349 pair matching algorithm to find the subset of treated days that were similar to con-
350 trol days. Matching is a very convenient and easy-to-use method to avoid making
351 modeling assumption to adjust for observed confounders. Despite these advantages,
352 readers unfamiliar with matching methods could be rightly shocked to see that only
353 30% of treated units could be matched to control units. This is a however a cost that
354 we were willing to pay since we did not want our analysis to rely on treated units
355 which do not have any credible counterfactual control units in the data. The diffi-
356 culty to find similar treated and control units should make researchers interested in
357 the acute health effects of air pollution worry that instrumental variable strategies
358 exploiting wind patterns and based on multivariate regression models might suffer
359 from extrapolation bias (41, 22, 42). Further research is needed to better understand
360 if improving the design stage of instrument variable studies with matching actually
361 lead to different results (43, 44).

362 In the third step, we analyzed our matched data using Neymanian inference, a
363 mode of inference specifically elaborated for randomized experiments. We found
364 that compared to other wind directions, North-East winds could lead to relatively
365 large increases in PM_{10} and $PM_{2.5}$ concentrations on the current and following days.
366 In the final fourth step, we implemented a quantitative bias analysis. An unmea-
367 sured confounder twice times more common among treated days could not alter
368 our qualitative conclusion on the effect of North-East winds on PM_{10} concentra-
369 tions: we would however be much less confident on the magnitude of this effect.
370 Future works should be carried out to see if the lag of an observed covariates or a
371 variable unmeasured in our analysis such as atmospheric pressure could had such
372 an influence on the treatment odds.

373 Last but not least, our analysis showed, as one could have expected, that a par-
374 ticular wind direction can simultaneously influence different air pollutants. This
375 well-known fact is another major issue for instrumental variable studies on the

acute health effects of air pollution. Yet, for wind direction to be a valid instrument, its effects on a health outcome must be fully mediated by a single air pollutant. In the statistics and economics literature, this assumption is often called the *exclusion restriction* (45, 46, 47). As recognized by researchers, studies exploiting wind patterns are therefore likely to violate this assumption. To overcome this issue, we could instead directly estimate the effects of wind patterns on a given health outcome. Which pollutant explains the observed relationship is another question which requires much needed methodological work. For instance, (48) propose to run a multi-pollutant model where each air pollutant concentration is predicted by selecting the optimal set of instrumental variables. The authors show that results of an instrumented multi-pollutant model can be very different from those found by single-pollutant models. In a future extension of our paper, we will run fake-data simulations to determine the air pollutants co-variance structure that leads to biased dose-response.

Acknowledgments

Airparif and Météo-France helped us access the data we used in our analysis. Grégory Abbou provided valuable feedback on an early and different draft of the paper. We especially thank Stéphane Shao for developing the matching algorithm and Marie-Abèle Bind and Alice Sommers for formulating and advocating in several papers the causal inference pipeline we rely on. Map data copyrighted OpenStreetMap contributors and available from <https://www.openstreetmap.org>.

References

- [1] Wolfram Schlenker and W Reed Walker. Airports, air pollution, and contemporaneous health. *The Review of Economic Studies*, 83(2):768–809, 2016.
- [2] Eva Arceo, Rema Hanna, and Paulina Oliva. Does the effect of pollution on infant mortality differ between developing and developed countries? evidence from mexico city. *The Economic Journal*, 126(591):257–280, 2016.
- [3] Joel Schwartz, Marie-Abele Bind, and Petros Koutrakis. Estimating causal effects of local air pollution on daily deaths: effect of low levels. *Environmental health perspectives*, 125(1):23–29, 2017.
- [4] Joel Schwartz, Kelvin Fong, and Antonella Zanobetti. A national multicity analysis of the causal effect of local pollution, no 2, and pm 2.5 on mortality. *Environmental health perspectives*, 126(8):087004, 2018.
- [5] Timothy J Halliday, John Lynham, and Aureo de Paula. Vog: Using volcanic eruptions to estimate the health costs of particulates. *The Economic Journal*, 129(620):1782–1816, 2019.
- [6] Tatyana Deryugina, Garth Heutel, Nolan H Miller, David Molitor, and Julian Reif. The mortality and medical costs of air pollution: Evidence from changes in wind direction. *American Economic Review*, 109(12):4178–4219, 2019.
- [7] M Bressi, Jean Sciare, Véronique Gherzi, N Mihalopoulos, J-E Petit, JB Nicolas, Sophie Moukhtar, A Rosso, Anaïs Feron, N Bonnaire, et al. Sources and geographical origins of fine aerosols in paris (france). *Atmospheric Chemistry and Physics*, 14(16):8813–8839, 2014.
- [8] H Petetin, M Beekmann, J Sciare, M Bressi, A Rosso, O Sanchez, and V Gherzi. A novel model evaluation approach focusing on local and advected contributions to urban pm 2.5 levels—application to paris, france. *Geoscientific Model Development*, 7(4):1483–1505, 2014.

- [9] Roland Stirnberg, Jan Cermak, Simone Kotthaus, Martial Haeffelin, Hendrik Andersen, Julia Fuchs, Miae Kim, Jean-Eudes Petit, and Olivier Favez. Meteorology-driven variability of air pollution (pm 1) revealed with explainable machine learning. *Atmospheric Chemistry and Physics*, 21(5):3919–3948, 2021.
- [10] David C Carslaw and Karl Ropkins. Openair—an r package for air quality data analysis. *Environmental Modelling & Software*, 27:52–61, 2012.
- [11] Stuart K Grange, David C Carslaw, Alastair C Lewis, Eirini Boleti, and Christoph Hueglin. Random forest meteorological normalisation models for swiss pm 10 trend analysis. *Atmospheric Chemistry and Physics*, 18(9):6223–6239, 2018.
- [12] Stuart K Grange and David C Carslaw. Using meteorological normalisation to detect interventions in air quality time series. *Science of the Total Environment*, 653:578–588, 2019.
- [13] Paul R Rosenbaum, PR Rosenbaum, and Briskman. *Design of observational studies*, volume 10. Springer, 2010.
- [14] Guido W Imbens and Donald B Rubin. *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press, 2015.
- [15] Donald B Rubin. For objective causal inference, design trumps analysis. *The Annals of Applied Statistics*, 2(3):808–840, 2008.
- [16] Miguel A Hernán and James M Robins. Using big data to emulate a target trial when a randomized trial is not available. *American journal of epidemiology*, 183(8):758–764, 2016.
- [17] Marie-Abele C Bind and Donald B Rubin. Bridging observational studies and randomized experiments by embedding the former in the latter. *Statistical methods in medical research*, 28(7):1958–1978, 2019.
- [18] M-AC Bind and DB Rubin. The importance of having a conceptual stage when

- reporting non-randomized studies. *Biostatistics & Epidemiology*, pages 1–10, 2021.
- [19] Jersey Neyman. Sur les applications de la théorie des probabilités aux expériences agricoles: Essai des principes. *Roczniki Nauk Rolniczych*, 10:1–51, 1923.
- [20] Donald B Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, 66(5):688, 1974.
- [21] Alice J Sommer, Emmanuelle Leray, Young Lee, and Marie-Abèle C Bind. Assessing environmental epidemiology questions in practice with a causal inference pipeline: An investigation of the air pollution-multiple sclerosis relapses relationship. *Statistics in Medicine*, 40(6):1321–1335, 2021.
- [22] Daniel E Ho, Kosuke Imai, Gary King, and Elizabeth A Stuart. Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political analysis*, 15(3):199–236, 2007.
- [23] Elizabeth A Stuart. Matching methods for causal inference: A review and a look forward. *Statistical science: a review journal of the Institute of Mathematical Statistics*, 25(1):1, 2010.
- [24] Jerome Cornfield, William Haenszel, E Cuyler Hammond, Abraham M Lilienfeld, Michael B Shimkin, and Ernst L Wynder. Smoking and lung cancer: recent evidence and a discussion of some questions. *Journal of the National Cancer institute*, 22(1):173–203, 1959.
- [25] Colin B Fogarty. Studentized sensitivity analysis for the sample average treatment effect in paired observational studies. *Journal of the American Statistical Association*, 115(531):1518–1530, 2020.
- [26] William E Wilson and Helen H Suh. Fine particles and coarse particles: concentration relationships relevant to epidemiologic studies. *Journal of the Air & Waste Management Association*, 47(12):1238–1249, 1997.
- [27] Gerard Hoek, Rob Beelen, Kees De Hoogh, Danielle Vienneau, John Gulliver,

- Paul Fischer, and David Briggs. A review of land-use regression models to assess spatial variation of outdoor air pollution. *Atmospheric environment*, 42(33):7561–7578, 2008.
- [28] Amos PK Tai, Loretta J Mickley, and Daniel J Jacob. Correlations between fine particulate matter (pm_{2.5}) and meteorological variables in the united states: Implications for the sensitivity of pm_{2.5} to climate change. *Atmospheric environment*, 44(32):3976–3984, 2010.
- [29] Rosana Aguilera, Alexander Gershunov, Sindana D Ilango, Janin Guzman-Morales, and Tarik Benmarhnia. Santa ana winds of southern california impact pm_{2.5} with and without smoke from wildfires. *GeoHealth*, 4(1):e2019GH000225, 2020.
- [30] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2021.
- [31] OpenStreetMap contributors. Planet dump retrieved from <https://planet.osm.org> . <https://www.openstreetmap.org>, 2017.
- [32] Mark Padgham, Bob Rudis, Robin Lovelace, and Maëlle Salmon. osmdata. *The Journal of Open Source Software*, 2(14), jun 2017.
- [33] Michael Mayer. missranger: Fast imputation of missing values (2.1.0). URL <https://CRAN.R-project.org/package=missRanger>. *R package version*, 2(0), 2019.
- [34] Paul W Holland. Statistics and causal inference. *Journal of the American statistical Association*, 81(396):945–960, 1986.
- [35] Peng Ding and Fan Li. Causal inference: A missing data perspective. *Statistical Science*, 33(2):214–237, 2018.
- [36] Donald B Rubin. Practical implications of modes of statistical inference for causal effects and the critical role of the assignment mechanism. *Biometrics*, pages 1213–1234, 1991.
- [37] Valentin Amrhein, David Trafimow, and Sander Greenland. Inferential statis-

- tics as descriptive statistics: There is no replication crisis if we don't expect replication. *The American Statistician*, 73(sup1):262–270, 2019.
- [38] Paul Rosenbaum. *Observation and experiment*. Harvard University Press, 2018.
- [39] Corwin Matthew Zigler and Francesca Dominici. Point: clarifying policy evidence with potential-outcomes thinking—beyond exposure-response estimation in air pollution epidemiology. *American journal of epidemiology*, 180(12):1133–1140, 2014.
- [40] Francesca Dominici and Corwin Zigler. Best practices for gauging evidence of causality in air pollution epidemiology. *American journal of epidemiology*, 186(12):1303–1309, 2017.
- [41] Gary King and Langche Zeng. The dangers of extreme counterfactuals. *Political analysis*, 14(2):131–159, 2006.
- [42] Hyunseung Kang, Benno Kreuels, Jürgen May, and Dylan S Small. Full matching approach to instrumental variables estimation with application to the effect of malaria on stunting. *The Annals of Applied Statistics*, 10(1):335–364, 2016.
- [43] Joel Schwartz, Elena Austin, Marie-Abele Bind, Antonella Zanobetti, and Petros Koutrakis. Estimating causal associations of fine particles with daily deaths in boston. *American journal of epidemiology*, 182(7):644–650, 2015.
- [44] Michela Baccini, Alessandra Mattei, Fabrizia Mealli, Pier Alberto Bertazzi, and Michele Carugno. Assessing the short term impact of air pollution on mortality: a matching approach. *Environmental Health*, 16(1):1–12, 2017.
- [45] Joshua D Angrist and Jörn-Steffen Pischke. *Mostly harmless econometrics*. Princeton university press, 2008.
- [46] William H Crown, Henry J Henk, and David J Vanness. Some cautions on the use of instrumental variables estimators in outcomes research: how bias in instrumental variables estimators is affected by instrument strength, instrument contamination, and sample size. *Value in Health*, 14(8):1078–1084, 2011.

- 531 [47] Michael Baiocchi, Jing Cheng, and Dylan S Small. Instrumental variable meth-
532 ods for causal inference. *Statistics in medicine*, 33(13):2297–2340, 2014.
- 533 [48] Alexandre Godzinski and Milena Suarez Castillo. Disentangling the effects of
534 air pollutants with many instruments. *Journal of Environmental Economics and*
535 *Management*, page 102489, 2021.