Università Commerciale "Luigi Bocconi"

PhD School

PhD program in: Economics and Finance

Cycle: 35th

Disciplinary Field (code): SECS-P/06

Essays on Environmental and Health Economics

Advisor: Stefano Fiorin

Co-Advisor (if any): Valentina Bosetti

PhD Thesis by

Jacopo Lunghi

ID number: 3109067

Year: 2024

Abstract

This thesis unravels some aspects the complex relationship between anthropogenic activities inducing environmental externalities and human health under an economic perspective.

The first chapter analyses the causal impact of manure spreading on fine particulate matter concentrations, as well as respiratory and cardiovascular hospitalisations, mortality rate at discharge, and treatment costs using air quality and hospital discharge data from the Lombardy region in Italy. Exogenous variation in spreading prohibitions is used to design a repeated event-study framework. It is estimated an increase of in $PM_{2.5}$ concentrations in the five days following a ban lift, paired with a spike in urgent hospitalisations, and higher hospital mortality rate during spreading events, which impose a financial burden of 30.9 to 67.7 million euros per year.

The second chapter investigates deeper the linkage between livestock farming and atmospheric pollution by studying the impact of bovine and swine farming on the concentration levels of ammonia (NH3) and coarse particulate matter (PM10) in Lombardy's atmosphere. Our econometric model of cattle and swine variation and simulations indicate that bovine bovine and swine farming could account for up to 25% of local pollution exposure, emphasizing the need for targeted mitigation strategies.

The third chapter studies how soil aridity impacts child wellbeing in Sub-Saharan Africa. Among the numerous implications of climate change, water access for people worldwide remains a key concern. we find that infants born in arid areas are comparatively more likely to die under the age of 5 and be systematically underweight at birth. In addition, we show how aridity reduces the effect of rainfall on child wellbeing and drives substantial heterogeneity in the estimated response to increasing precipitation.

Acknowledgements

I am grateful to my advisor Stefano Fiorin and my co-advisor Valentina Bosetti for supporting me throughout my graduate journey, guiding me with patience and precious expertise, and encouraging me to explore and think out of the box in the field of environmental economics.

I thank Marco Percoco, Maurizio Malpede, Lara Reis, Roberto Vincenzi, Pamela Giustinelli, Francesco Decarolis, Michela Braga, Gaurav Khanna, and Kate Ricke for contributing to shape my research interests and skills as coauthors, advisors, and teaching supervisors. Special mention goes to the team members of the INHALE project and my coworkers at EIEE: without their support, this thesis would not have been the same.

I thank Teevrat Gaarg for hosting me at UC San Diego and for encouraging support. I thank all the fellow students I met at Bocconi and UCSD. Special appreciation goes to Lorenzo, for sharing this experience with me from day one.

I am grateful to my family for their unwavering support and unconditional acceptance throughout the journey of completing this thesis. Their love and encouragement have been the cornerstone of my strength and resilience. I wish to thank all my friends, in particular the members of the community "The true friendship".

Contents

Page No.

1 A Spreading Malaise: Manure Management, Air Pollution, and Hospital Admissions in Italy 1

1.1 Online Appendix 53

2 An assessment of the contribution of livestock intensity on particulate matter concentration in Italy 77

2.1 Online Appendix 119

3 Soil Aridification, Precipitations, and Infant Health: Evidence from Africa 118

3.1 Online Appendix 160

A Spreading Malaise: Manure Management, Air Pollution, and Health Outcomes in Italy

Jacopo Lunghi[†]

Bocconi University, Milan, Italy

RFF-CMCC European Institute for Economics and the Environment

This version: May 15, 2024

Abstract

Manure application is a widespread soil fertilization practice in agriculture which may constitute an environmental and public health hazard. Using air quality and hospital discharge data from the Lombardy region in Italy, in this paper I estimate the causal effect of manure spreading on fine particulate matter concentrations, and study how short-term spikes in air pollution affect respiratory and cardiovascular hospitalisations, mortality rate at discharge, and medical treatment costs. To isolate the impact of manure application, I exploit exogenous variation in spreading prohibitions as a repeated event-study design. I estimate an increase of around 27% in $PM_{2.5}$ concentrations in the five days following a ban lift, paired with an increase in urgent hospitalisations by a factor of 1.04 to 1.145, and higher hospital mortality rate (0.7 to 1 percentage points) during spreading events. Conversely, it is found no significant difference in the cost per hospitalisation. I estimate the financial burden limited to this health threat to range between 30.9 and 67.7 million euros per year. Finally, I simulate the impact of including particulate matter targets in the current regulatory framework and find that, despite reductions in concentrations could be achieved, refined policy may not be sufficient to curb air pollution.

Keywords: Air Pollution; Pollution Control; Livestock Farming; Cost Benefit; Environmental Health and Safety

JEL Classification: Q00; Q51; Q53; K32; I18

[†]Contact: Jacopo Lunghi, Bocconi University, Via Rontgen 1, 20136 Milano.
E-mail: jacopo.lunghi@phd.unibocconi.it. Web: jacopolunghi.github.io.

Access to the National Hospital Discharges database was authorized by the Ministry of Health, Planning Department, under confidential disclosure agreement. The evidence produced in this paper respects data confidentiality and was formerly approved for disclosure by the Ministry of Health. Any error and inaccuracy are the author's responsibility.

1 Introduction

Manure application is a common practice in agriculture to fertilize and condition the soil. Although a potentially valuable source of plant nutrients, the disposal of animal organic waste may pose important environmental concerns, some of which seem notably overlooked. In particular, the contribution of manure application to coarse (PM_{10}) and fine ($PM_{2.5}$) particulate matter concentrations is still one of the most poorly characterized sources of air pollution and, as such, often poorly addressed in most regulatory frameworks worldwide (Cambra-López et al. 2010). PM emissions from farming activities are primarily the result of secondary inorganic aerosol (SIA) in the atmosphere from its precursors: ammonia (NH_3), nitrogen oxides (NO_x), and sulfur dioxide (SO_2) among the most relevant. NO_x and SO_2 originate mostly from combustion of coal, oil, and fossil fuels, whereas the vast majority of ammonia emissions are produced by agriculture. More specifically, NH_3 almost entirely originates from soil application and disposal of livestock waste and fertilizers.

Research has shown that the agricultural industry can be a key source of directly emitted PM (*primary emissions*), mostly from storage and handling of agricultural products, agricultural waste burning, land preparation and harvesting (Garg et al. 2023, He et al. 2020, Nian 2023). As such, emissions from agriculture are associated with thousands of annual deaths (Domingo et al. 2021, Lelieveld et al. 2015). Much less is known about the potential harm induced by secondary PM formation and manure management. First, it is often challenging to isolate the contribution of farming activities from other sources of airborne pollution in aggregate settings. Second, even isolating and quantifying PM mass from livestock activities may not fully characterize the hazard posed by animal breeding. Indeed, toxicological studies have questioned the existence of a specific association between SIA particles and risk to health at ambient concentrations, despite the information being still limited (Cassee et al. 2013). Lack of fully conclusive evidence has led the WHO to maintain a cautious stand, assuming equal toxicity of all chemical particles contribution to PM mass, including SIA (WHO 2021).

In this paper, I investigate the environmental and health risks posed by manure spreading activities in the Lombardy region in Italy. I use the unique features of Italy's regulation in matter of manure spreading prohibitions as a quasi-natural experimental framework to estimate the short-term causal effect of manure spreading on PM concentrations, hospital access, and patient's conditions for respiratory and cardiovascular acute episodes. To this aim, I collect geolocated patient-level data from hospital discharge forms for the years 2016 through 2019, matched with high-frequency data on pollutants' concentration and weather conditions at municipal level. To capture the true causal effect of manure spreading, I build a repeated event-study framework, where each event is determined by the conditionally exogenous opening of a "spreading window", i.e. a period of consecutive days of spreading ban followed by consecutive days of lifted prohibitions. This allows me to estimate the portion of PM mass originating from manure management activities as well as overcome endogeneity concerns that arise from selective exposure of individuals to

air pollution based on unobservable characteristics correlated with individual health. My identifying assumptions rely on the absence of systematic differences between hospitals and patients around the window cutoff, and the absence of endogenous selection of patients on either side of the threshold. I argue that, conditional on observables, the short time interval considered and the limited knowledge of farming regulation of individuals allow to pinpoint causality between spreading of manure, air pollution, and subsequent health externalities. Defining symmetric short-term windows around a spreading event also helps to cope with staggered treatment concerns (De Chaisemartin & d'Haultfoeuille 2020, Sun & Abraham 2021, Callaway & Sant'Anna 2021), as municipalities receive treatment at the same time within each window. To complement my preferred specification, I show the persistence of the results to a difference-in-difference strategy that exploits geographical discrepancies in prohibitions across the region.

I estimate that spreading windows induce a spike in $PM_{2.5}$ concentrations, roughly 27% increase, peaking at more than 40% at the third consecutive day of lifted prohibitions. The effect is stronger in areas more with higher concentration of farming animals, stressing the key role of husbandry activities as a driver to the observed spike in PM. Municipalities in the highest two deciles of the distribution in terms of farming animal stock experience an increase in $PM_{2.5}$ 7% to 10% higher than those in the three lowest deciles. The results are robust to different combinations of weather controls and fixed effects, and corroborated by both a placebo test using rainfall alone as a discriminant for window opening, and by detecting no comparable trends in pollutants that are not directly caused by livestock emissions, such as Ozone and SO_2. I then analyse the evolution hospital admissions, mortality at discharge, and treatment costs of inpatients with respiratory and cardiovascular diseases (R&CD) during spreading-induced PM peaks. I find that spreading events are associated with one additional daily urgent admission every 25 patients: this effect is stronger when excluding days close to the cutoff days and peaks around the fourth day of prohibitions lift, up to a maximum of one every 7 patients, suggesting partial lag in the response to PM spikes. Moreover, I find that individuals admitted during a spreading event experience between 0.7% and 1% higher rate of mortality at discharge. Finally, it is found no premium in the cost per hospitalisation during PM peaks: patients do not appear to stay longer in hospital and do not require more expensive procedures compared to the baseline, although the effect could be confounded by limitations in the cost measure available and the endogenous response from medical facilities. I find no effect on the same indicators in inpatients hospitalized for treatment more remotely related to air quality, such as digestive, genitourinary, and musculoskeletal system diseases.

I use these results to estimate the financial burden implied by the observed health externalities associated with spreading activities. I simulate the incremental admissions due to manure application and calculate the monetary equivalent total years of life lost due to the increase in hospital mortality assuming a value of statistical life year (VSLY) of €158,488 (Cots et al. 2011). I obtain a total cost from increased usage of health care facilities between 16.7 and 48.1 million euros, in addition to a total VSLY quantified between 13.8 to 19.5 million euros. These figures likely represent a lower bound to the true

impact of manure application, as they abstracts from potential morbidity and co-morbidity effects, as well as long-term implications to PM exposure.

Finally, I simulate the gains of incorporating targets related to PM concentrations in the region. Hence, I assume a regulator who aims at minimizing the days for which PM levels exceed the threshold considered unhealthy for sensitive groups (set at 35 $\mu g/m^{-3}$), leaving other restrictions to spreading activities in the current regulation unaltered, and deducing counterfactual PM concentrations considering manure application as an additive increase. I design a constrained finite recursive optimization problem under uncertainty to define a new allocation of spreading days, and benchmark the results with the best allocation achievable under perfect forecasting (i.e. retrospective information on PM levels and weather conditions). The results show the limitations of introducing PM targets in manure management regulation without any complementary policy to sustain such targets. With a potential maximum reduction in days of hazard concentrations around 10% estimated under perfect forecasting, introducing uncertainty cuts the estimated gain to less than 1%, due to low precision in PM and weather predictions.

This paper contributes to the current literature in environmental and health economics that estimates the causal effects of air pollution on human health. There exist many examples of studies that investigate the impact of air quality, many of which focus on one specific health issues such as adult and infant mortality (Chay et al. 2003, Chay & Greenstone 2003, Currie & Neidell 2005, Jayachandran 2009, Currie & Walker 2011, Chen et al. 2013), infant health (Friedman et al. 2001, Neidell 2004, Knittel et al. 2016), mental health (Chen, Oliva & Zhang 2018). Only a subset of them focus specifically on particulate matter, even less so on $PM_{2.5}$ (Pope III et al. 1999, Schwartz et al. 2017, Deryugina et al. 2019). Importantly, only few papers are able to quantify healthcare costs and to look in detail at medical procedures and cost per admission, thus separating the effect on the extensive and intensive margin. Schlenker & Walker (2016) estimate hospitalization rates and costs in California, but focus on carbon monoxide exposure. A paper by Deryugina et al. (2019) uses elderly Medicare recipients data to estimate the effect of $PM_{2.5}$ exposure on elderly mortality, health care use, and medical costs. However, the authors provide an estimate for overall inpatient spending, without assessing whether the effect may originate from patients requiring more cost-intensive treatment during periods of high PM exposure. There are instead fewer examples of quantification of healthcare costs outside the US setting are found in the literature, given scarce data availability (Xia et al. 2022). In this paper, I look closely at different cost components in a novel setting.

Furthermore, this study is, to the best of the author's knowledge, the first attempt to estimate the causal impact of manure management in particular on air pollution, health, and healthcare spending. Some studies have tried to impute air-quality related deaths to agricultural activities, including manure management (Jerrett 2015, Lelieveld et al. 2015, Giannakis et al. 2019, Domingo et al. 2021), but they have not yet fully established a direct causal link to pinpoint the share of PM concentrations and the health concerns for which manure management is responsible. I believe this is the first paper that uses winter spreading prohibitions as conditionally exogenous local variation to study the effect of

manure management on air pollution levels and respiratory and cardiovascular diseases. This constitutes a novel approach to directly identify the implications of manure application activities on a large scale. Amid lack of reliable data on PM decomposition and, as such, the impossibility to observe specifically SIA particles, by isolating livestock as primary emission source, the paper indirectly adds to the current knowledge on the hazard to human health posed by different airborne particles (Kelly & Fussell 2016, 2020). Finally, this paper contributes to the regulatory debate around the most suitable policy response to reduce emissions from manure application and increase welfare.

The remainder of this paper is organized as follows. Section 2 provides background information about the current legislation on manure spreading activities in Lombardy, that justifies my strategy. Section 3 presents the empirical strategy employed, and Section 4 describes the data. Section 5 presents main results and assesses their robustness. Section 6 simulates the healthcare spending costs of manure spreading, and Section 7 describes a counterfactual scenario of fully-flexible spreading prohibitions and estimates its benefit in terms of pollution reduction. Finally, Section 8 concludes.

2 Background: Spreading Policies in Lombardy and Worldwide

2.1 The environmental concerns of manure application

Manure application provides nutrients that improve soil quality and increase crop production, but carries consistent environmental risk and, consequently, serious potential harm to human health (Bouwman et al. 2013).

First, manure management and farming activities contribute significantly to emissions of greenhouse gases (GHGs)[1], which are the main responsible of rising temperature and climate change. The phases of storage, treatment and spreading of manure have all been associated to GHG emissions (Chadwick et al. 2011), and manure management alone is estimated to contribute between 1.6% and 13.6% of total GHGs emissions across Europe[2]. This is particularly true for Lombardy as well, where GHGs concentrations, particularly with regards to N_2O and CH_4, spatially correlate with farming intensity (see Figure E.1 in Appendix). While GHGs can potentially pose direct risk to human health, this occurs primarily at very high concentrations and in constrained environments. High concentrations of ambient CO_2 (at 20,000 ppm) have been found to induce physiological responses in blood pressure, hearth and respiratory rate (Maniscalco et al. 2022), while lower atmospheric concentrations ($<$ 5,000 ppm) are thought to represent a direct health risk only through chronic exposure (Jacobson et al. 2019) and long-term modifications of global climate.

Second, the large quantities of nitrogen (N) released into the soil by slurry redepositing can enter groundwater, and later be transported into freshwater, through infiltration and surface runoff (Ongley 1996). This, in turn, poses several environmental concerns

[1] Most notably, farming activities are associated to higher concentrations of N_2O (nitrous oxide), CH_4 (methane), and CO_2 (carbon dioxide).

[2] Source: EEA - Approximated greenhouse gas inventories.

such as eutrophication, algal blooms, loss of biodiversity, and fish stock depletion, in addition to direct threats to human health through the consumption of contaminated water. Evidence of water-linked diseases has been retrieved especially interesting the digestive and endocrine system (Majumdar & Gupta 2000, Azizullah et al. 2011). There is instead no consistent evidence that links exposure to polluted water and respiratory and cardiovascular conditions, which could however be linked to manure application through a different channel. Indeed, organic droppings from livestock are a primary source of ammonia emissions (NH_3), which is a key component for SIA formation of fine particulate matter ($PM_{2.5}$)[3]. Despite the reaction time of precursors forming $PM_{2.5}$ is difficult to determine with precision, recent studies suggest that, while part of the reaction may happen locally within few minutes, second stages of the reaction may continue overnight and in the following days after ammonia is released (Kim et al. 2022). In Italy, roughly 95.7% of ammonia emissions between 2008 and 2019 are associated to fertilizers and livestock, which is marginally below the EU average (96.4%)[4]. Exposure to $PM_{2.5}$, in turn, has been largely associated in the literature with serious health concerns. Importantly, even short-term exposure to polluted air (Wong et al. 2008, Larrieu et al. 2007) and high $PM_{2.5}$ concentrations (Guaita et al. 2011) have been found to induce respiratory-, natural-, and cardiovascular-cause mortality. If the effects of PM have been largely studied, much less is known about the exact toxicity of particles with different chemical compositions. Finding refined guidelines on the potentially heterogeneous hazard level of airborne particles has been in the agenda for almost 20 years (Council et al. 2004), but the results are still far from conclusive. Toxicological research has found little to no proof of a biological association with ammonium nitrates and sulphates and human health, but strong epidemiological association leaves open the possibility of alternative underlying mechanisms, such as cations associated with these compounds and absorbed components (Schlesinger & Cassee 2003, Reiss et al. 2007, Cassee et al. 2013). Recent lab evidence suggests increased airflow obstruction and respiratory following chronic exposure to SIA in mice (Zhang et al. 2022, 2021). According to the guidelines of the WHO, all chemical particles constituting PM mass are to be considered having equal toxicity (WHO 2021), but it is yet to be determined with precision whether a peak in fine PM concentrations induced by organics waste from livestock should have health consequences on individuals and, if so, to what extent. Scarce empirical evidence on the topic is potentially one of the reasons why PM emissions emerge a second-order concern in the regulation of manure application.

2.2 Manure Application Policies in Lombardy

After a first attempt to legislate the utilization of fertilizers in 1984, following the new wave of European legislation on environmental protection, Italy has firstly introduced

[3]Secondary PM formation takes place through the chemical reactions between ammonia and gaseous precursors such as sulfates (SO_x) and nitrates (NO_x), forming crystalline solid compounds (ammonium nitrates and ammonium sulphates) that take part in the composition of $PM_{2.5}$ (Squizzato et al. 2013).

[4]Source: Eurostat, Air emissions accounts by NACE Rev. 2 activity.

comprehensive regulation about manure management and usage of fertilizers in 2006[5]. The new regulatory framework introduced spatial and temporal prohibitions to the ground application of droppings, in the interest of maintaining proper sanitary conditions of urban and extra-urban areas and, most importantly, preventing pollution of groundwater bodies, which are particularly threatened by nitrates originating from fertilizers (Galloway et al. 2008, Sebilo et al. 2013). Among the restrictions imposed, which vary on the basis of soil composition, average slope, and distance from metropolitan areas, limitations in the timing of sludge application are of particular interest to this paper. Indeed, the law established a minimum ban of 90 days[6] during the winter season, identified as "*typically elapsing between 1st November and 28th February*". This regulatory action was justified with winter posing the highest risk of frozen ground and, consequently, nitrates runoff and pollution of surface waters during snowmelt (Young & Mutchler 1976). For similar reasons, the law prohibited manure spreading during a rainy day, and the day immediately after. Regional entities were then responsible to introduce further legislation to regulate how exact dates for the constrained period were defined. In Lombardy, these have been regulated by a series of action plans (*Piano d'Azione Nitrati*, henceforth PdA), updated every three years. Starting and ending dates of the ban period have been variable from one year to the other and communicated during spring for the upcoming winter by regional decree.

As manure storage is often a costly activity, especially due to capacity constraints of farmers, to promote cost reduction and prevent overload, in 2016 Italy increased regulation flexibility. While maintaining a 90-days prohibition, the fixed ban period was shortened to 62 days (between 1st December and 31st January), while the remaining 28 out of 58 days were once again set by regional governments. In Lombardy, spreading prohibitions and permits started to be regulated through bulletins (*Bollettini Nitrati*) issued every two to five days throughout the period subjected to the authority's discretionality. The bulletin reports the manure spreading potential (either "allowed" or "not allowed") of six different climate zones within the region (*Zone pedoclimatiche*) for the upcoming days. Restrictions are imposed at the climate zone level: discrepancies between zones may occur in case of significantly different weather conditions across the region. [7] Bulletins are easily accessible through a dedicated app as well as regularly posted online in a dedicated webpage[8]. Failure to comply with the restrictions can result in an administrative fee up to 5.000 euros, which is substantial given that the economic size class of more than 70% of farms in Lombardy does not exceed 50.000 euros [9][10], and even a criminal liability

[5]MD 7 Apr 2006. A summary of the most salient regulatory advancements in Italy is reported in Appendix (Section C).

[6]The limitation was extended to 120 days for specific categories of poultry manure.

[7]Climate zone areas are: *Alps* (provinces: SO); *Central plain* (provinces: BG, BS, CR); *West plain* (provinces: LO, MI, PV); *East plain* (provinces: MN); *West Prealps* (provinces: BG, CO, LC, MB); *East Prealps* (provinces: BG, BS). See Figure E.2 in Appendix.

[8]Figure E.4 in Appendix shows the content of the app and online documents available to farmers to gather information about day-to-day restrictions to manure application.

[9]Economic size classes are defined at EU level (Commission Regulation No 1242/2008).

[10]Source: Agricultural Census (ISTAT). See Figure in Appendix.

depending on the degree of environmental damage. Figure 1 summarizes the regulatory framework, which remained in place between 2016 and the beginning of 2021.

From the farmers' perspective, there is a strong incentive to apply manure when winter restrictions are temporarily lifted, given uncertainty about future spreading opportunities. Indeed, initial investment and maintenance of storage containment structures, such as tanks or pits, is a remarkable cost for farmers which, in turn, tend to have limited resources to stock manure during long periods of prohibition and face the risk of storehouse overcapacity. Moreover, while application is not necessarily tied to an individual's farmland (farmers who exhausted storing capacity can rent parts of land from other farms to apply the excess manure), rental fees and high transportation costs lower the incentive to dispose of manure far from the farmstead. Data from 2010 ISTAT Agricultural Census (Figure 2) indicate that more than 50% of farms apply more than 75% of their manure production to their own land. Additionally, considering cattle- and pigs-specialized entities, there exist a negative correlation between farm size and amount of manure disposed outside, meaning that comparatively smaller producers tend to rely on rental farmland, reducing geographical discrepancy between a farm's location and where the manure produced is applied.

Despite the regulatory framework defined a broad set of rules to ensure transparency and control over manure application activities, it did not include any consideration about air pollution in determining spreading prohibitions prior to 2021, which justifies the research design employed in this paper. In fact, only with PdA 2020-2023, a first attempt to bring PM concentrations into the picture was made, in the form of an additional criterion to spreading. Specifically, manure application was forbidden in municipalities where containment measures to reduce air pollution (e.g. traffic stops) are in place.[11]

subsectionOther Examples of Manure Application Regulation Italy and the Lombardy region are not an isolated example of the attempt to cope with the environmental challenges posed by manure management. Other European countries, following the guidelines outlined in the Nitrates Directive (91/676/EEC), have imposed similar restrictions, including prohibitions to manure application during winter months. For example, in Germany spreading is not allowed between 1 November and 31 January on arable land and between 15 November and 31 January on grassland[12]. In the Netherlands, prohibitions span between 1 September to 31 January[13], while the longest period of restrictions in the UK, depending on soil type and land use, can be as long as 7 months, from 1 August to the end of February[14]. Possibly the regulation closer to the Italian case is observed France,

[11]Pollution control measures can be applied in Lombardy between October 1st and March 31st. They are triggered by PM_{10} concentrations exceeding $50\mu/m^3$ for at least four consecutive days, and stay in place until favorable weather conditions to the reduction of airborne pollutants (e.g. rain) or at least two consecutive days below $50\mu/m^3$ PM_{10} concentrations. Temporary measures include restrictions on bonfires, manure application (since 2021), indoor temperatures (i.e. heating systems control), and traffic. They are implemented at province level, with the exception of traffic limitations, which are applied at municipal level and only for municipalities with more than 30,000 inhabitants.

[12]Fertilizer Act; Ordinance on Fertilizer Application (Düngeverordnung. Latest amendment 2007)

[13]Decision Using Fertilisers (*Besluit gebruik meststoffen*), 1997.

[14]The Reduction and Prevention of Agricultural Diffuse Pollution (England) Regulations 2018.

where a manure application calendar is in force in winter months. This similarly imposes flexible ban periods depending on soil vulnerability and land use, plus some (less stringent compared to Italy) weather-based restrictions.[15]

Outside Europe, the regulation on manure spreading is relatively laxer: (Liu et al. 2018, (Fig. 2)) provide a summary of existing restrictions. For instance, in countries like Canada, Australia, and the US, where the regulation is applied at sub-national level, with the exception of some Canadian provinces where winter prohibitions are in place, the policy action ranges from application of partial weather restrictions, to submission by farmers of voluntary manure management plans, to even complete absence of guidelines. To the best of the author's knowledge, there exist no similar regulation in large Asian countries, such as India and China, as well as in developing economies.

The attention, amid considerable heterogeneity, to manure management governance reinforce the relevance of the findings presented in this paper outside of the Italian context, as well as the broad scope of its policy implications. Furthermore, while Lombardy displays high concentration of farming activities, making it potentially more susceptible to the environmental and health threats of manure application (Lombardy is in the 96^{th} percentile for animals total headcounts in the EU)[16], it is not an isolated case. For example, other regions in countries like Spain, France, Germany, and the UK rely heavily on farming industry, displaying comparable figures in terms of livestock presence, as well as similar weather and geographical characteristics.[17]

3 Methodology

3.1 Sample definition

To estimate the relationship between manure application, PM concentrations, and health outcomes, I exploit the variation induced by the spreading prohibition framework in Lombardy. To cope with unobservable confounders that may drive the level of pollutants in parallel with manure spreading, I restrict the sample to periods of "spreading windows" (W), i.e. consecutive days of spreading ban (W^-) followed or preceded by consecutive days in which spreading is allowed (W^+). A window is considered closed whenever prohibitions are in place, open otherwise. This implies that windows that are in temporal order either open-closed or closed-open will be considered the same event. The choice of closing and opening intervals lies on two major considerations. First, under the assumption that spreading happens relatively close to the cut-off day[18], the window should be as such so to allow most of the chemical reaction of gaseous precursors to take place and, in turn, the formation of PM due to the release of ammonia. Second, a window should be short enough

[15] Interministerial decree 19 December 2011, 23 October 2013, 11 October 2016.

[16] Source: Eurostat. Animal Production Statistics.

[17] See Figure E.7 in Appendix.

[18] As previously argued, farmers have the incentive to apply manure when a spreading ban is about to be imposed (in order to maximize storing capacity in face of restrictions) and or when this is lifted (to free up storage space).

to ensure as much comparability as possible, conditional on observables, between days before and after the opening/closing cut-off. As such, the choice of the threshold needs to balance the need to let spreading and SIA particles formation occur in the days without prohibitions, as well as to minimize the possibility of confounders, ensuring comparability before and after. I set the windows to be open and close for five consecutive days each, obtaining windows of 10-day duration in total. This number, which serves both goals of particles formation and short time span, is obtained using a data-driven bandwidth selection approach implemented through a regression-discontinuity (RD) design proposed by Calonico et al. (2017). Using data on prohibitions over the entire winter period in my sample years, the RD estimator provides a strategy for bandwith selection obtained by investigating in the data the persistence of a perturbation induced by a change in treatment status, i.e. relieved spreading prohibitions. Details on the RD estimation and the bandwith selection method can be found in Appendix A.1.

I consider positive exposure (i.e. open window) as the treatment received by a municipality. Panel [A] in Figure 3 plots the calendar distribution of window days. Spreading windows usually take place around the start (October-November) and the end (February-March) of the regulated winter period. One window exceptionally occurs in December: due to pressure of farmers and impossibility to allocate enough spreading days in the previous months, in 2019 the regulator agreed to a temporary suspension of bans to allow farmers to get rid of accumulated waste[19]. During bulletin-regulated periods, prohibitions may be relieved or imposed heterogeneously across climate zones. For instance, manure application can be delayed in some areas of a few days compared to the rest of the region because of unfavorable weather conditions. As such, in the same 10-day span, some climate zones may exhibit a window with two balanced periods of prohibition and permission, while others may experience longer or shorter periods of window closures (e.g., imagine a climate zone where manure application is restricted for seven consecutive days, and allowed for three). To avoid imperfect compliance with the treatment status of allowed manure application in my estimates, After identifying the days in which windows take place, in my preferred specification I restrict the sample to climate zones where the window is fully realized (Figure 3, Panel [B]). In most cases (seven out of nine windows), this implies excluding a maximum of two climate zones. The geographical discontinuities in the realization of spreading windows is then used to test the robustness of the results.

After identifying days in which spreading windows took place, using information available through hospital discharge forms, I then select all patients whose hospitalisation occurred on one spreading window day. I focus on patients that have been diagnosed with a respiratory and/or cardiovascular disease. Exposure to high concentrations of particulate matter is associated with additional medical conditions such as skin diseases, psychological disorders, and depression (Kim et al. 2016, Braithwaite et al. 2019). Yet, R&CDs are the two major and more extensively analyzed effects in the literature, in addition to being comparatively more likely to prompt urgent hospital admissions in case of acute symptoms.

[19]Source: *Bollettini Nitrati*, December 2019.

3.2 Empirical strategy

My objective is to estimate the effect of short-run exposure to manure spreading events PM concentrations, hospital admissions, hospital mortality, health care use and spending, net of potentially confounding factors. I propose an event-study model summarized by the following equation:

Model 1:

$$Y_{ihmt} = \sum_{k \neq -1 \in g} \eta_k \mathbb{1}\{H_{mt} = k\} + X'_{ihmt}\Gamma + \alpha_d + \alpha_{m\tilde{m}} + \alpha_{\tilde{m}y} + \alpha_w + \varepsilon_{ihmt} \tag{1}$$

where $H_{mt} = t - E_m$ indicates the period relative, to time t, to the event of manure application being allowed in municipality m (E_m). Periods relative to the spreading event are limited to set $g = \Big[k : k \in [-T, T-1]\Big]$, where $T = \frac{\text{card}(W)}{2}$ equal to five days. η_k coefficients are estimated pooling together all window events in the sample, with $k < 0$ characterizing pre-trend coefficients and $k \geq 0$ capturing treatment at k days from the window opening.

Y_{ihmt} is the outcome observed in municipality m at time t. In the case of healthcare measures, these are also defined at the individual i and the hospital h level. X represent a matrix of of controls, which varies depending on the outcome studied. It always includes six weather controls: temperature, rainfall, humidity, radiance, wind direction (16 indicators), wind speed, average boundary layer height, measured up to the third lag, and interacted amongst each other in current and lagged periods. For health outcomes, when the variable is identified at the individual level (e.g. mortality at discharge, cost of admission), I control for a patient's age class, gender, citizenship, and type of disease (respiratory or cardiovascular) and admission (surgical or medical). α_d, $\alpha_{m\tilde{m}}$, $\alpha_{\tilde{m}y}$, α_w, respectively identify day-level controls (day-of-the-week and holiday fixed effects), province-by-month, month-by-year, and window fixed effects.

Restricting the sample to climate zones that fully experience a spreading window implies that my framework does not include never-treated unit. Thus, the causal identification of the impact of manure application relies crucially on the full comparability of patients, municipalities and hospitals to the left and to the right of the spreading window. In other words short time intervals around spreading windows entails that a municipality's counterfactual is given by the municipality itself few days before. This is true under the assumption of existence of parallel trends in the baseline outcome, conditionally on observables[20]. The assumption rests on multiple considerations. First, spreading limitations are independent from air quality in the regulatory framework, since no PM control objective

[20]Formally, let $Y^{\infty}_{m,t}$ be the potential outcome of municipality m which never receives the treatment. Parallel trends require that:

$$\mathbb{E}\left[Y^{\infty}_{m,t} - Y^{\infty}_{m,s} | E_m = k, X_{mt}, \alpha\right] = \bar{\nu}, \quad \forall\ s \neq t \text{ and } \forall\ k \in supp(E_m)$$

, where in my framework the absence of never-treated unit is represented by $\infty \notin supp\,(E_m)$

exist in the regulator's agenda. Even more so, manure application is independent from the level and severity of hospital admissions in the region in the eyes of the policymaker. Second, windows are driven by the combination of exogenous events, such as the structure of temporal winter prohibitions (exogenous cut-off days, 90-day fixed prohibition period) and weather conditions (i.e. temperature, rainfall) in the region. Restricting the attention to a short time-span around the cut-off also reduces importantly the probability of capturing underlying trends induced by events correlated with manure application.

Moreover, causal identification of the impact of manure spreading implies no pre-treatment anticipatory behaviour able to affect outcome in absence of treatment[21]. For what concerns PM concentrations, once again the assumption is reinforced by the nature of the spreading control policy implemented in the region, seemingly unrelated to air pollution. Furthermore, manure spreading prohibitions are quite far from the knowledge of the general public and health management structures, and receive minimal news coverageYet, being correlated with spikes in airborne pollutants, individuals may have knowledge of high levels of PM in the region potentially induced by manure spreading even without knowing its exact source. In turn, they may take preventive action to avoid exposure: other studies have found that individuals can exhibit shielding behaviour when air quality decreases (Zivin & Neidell 2009, Moretti & Neidell 2011, Neidell 2009). Even under this assumption, however, observing a negative impact on the health of individuals would imply that the real effect of manure management activities could be even higher in absence of prevention from citizens.[22]

Using closed spreading windows as suitable counterfactual is my preferred choice to capture the true impact of manure application. However, the existence of discrepancies in spreading restrictions across climate zones may lead to consider municipalities where spreading is still prohibited compared to the rest of the region as effectively non-treated units, representing a more robust counterfactual and allowing for a difference-in-difference (DiD) kind of estimator. There are two major limitations to this alternative strategy. First, manure application is a source of secondary $PM_{2.5}$ generating through the reaction of gaseous precursors. Due to air transportation of pollutants, spillovers, in the form of SIA particles reaching areas where spreading is still restricted, are to be expected and quite complex to quantify. As such, a DiD estimator would likely be biased downward. Second, discrepancies between climate zones are often short-lived: given the 30-day prohibitions lift period universally applying to all climate zones, the regulator has incentive to avoid differences in terms of cumulative restriction days imposed growing large across the region.

[21] Formally, this entails:

$$\mathbb{E}\left[Y^{k}_{m,k+l} - Y^{\infty}_{m,k+l} | E_m = k, X_{mt}, \alpha\right] = 0 \quad \forall k \in supp(E_m) \text{ and } l < 0$$

[22] It is important to notice that the treatment such defined does not suffer from staggered adoption the canonical sense (De Chaisemartin & d'Haultfoeuille 2020, Sun & Abraham 2021, Callaway & Sant'Anna 2021). Indeed, within each window, units enter and exit the window period at the same time, while receiving treatment homogeneously in the same time period. Demeaning by a window identifier allows to cope with window-specific differences in the effect of spreading.

As such, municipalities stay untreated for rather short time, which makes it cumbersome to isolate the effect of manure application. In addition, my main strategy allows to capture causality while preserving a larger number of data points, implying higher statistical power to estimate an impact of manure application rather small in magnitude, as it is expected in particular for health outcomes. These constraints drive the choice to adopt Model **??** as preferred specification. Using a refined sample definition, I still investigate the robustness of the results to a DiD strategy that considers municipalities where prohibitions are still to be relieved as control group. I detail the empirical model in Appendix, Section A.2.

In the spirit of Borusyak & Jaravel (2017), I refer to Equation 1 as *semi-dynamic*. In addition, I estimate the same model in *static* form, i.e.:

Model 2:

$$Y_{ihmt} = \eta_0 D_{mt} + X'_{ihmt}\Gamma + \alpha_d + \alpha_{m\tilde{m}} + \alpha_{\tilde{m}y} + \alpha_w + \varepsilon_{ihmt} \tag{2}$$

where $D_{mt} = \mathbb{1}\{t \geq E_m\}$. I use the static model to explore spatial treatment heterogeneity based on livestock concentration. I interact D_{mt} with a set of indicators obtained by binning the number of livestock units at municipal level:

Model 2b:

$$\begin{aligned} Y_{ihmt} =& \eta_0 D_{mt} + \sum_{b=1}^{9} \gamma_{0b} L_{b0mt} + \sum_{b=1}^{9} \rho_b \left[E_{mt} \times L_{bmt}\right] + \\ &+ X'_{ihmt}\Gamma + \alpha_d + \alpha_{m\tilde{m}} + \alpha_{\tilde{m}y} + \alpha_w + \varepsilon_{ihmt} \end{aligned} \tag{3}$$

Finally, when my dependent variable has a count nature, such as for daily hospital admissions, I account for non-negative and discrete nature of hospital admissions data with possibly many zeros through a pseudo-Poisson regression model with high dimensional fixed effects (Correia et al. 2020):

Model 3:

$$\mathbb{E}\left[Y_{ihmt} | D_{mt}, X_{ihmt}, \alpha\right] = \alpha \exp\left[\eta_0 D_{mt} + X'_{ihmt}\Gamma\right] \tag{4}$$

The model thus specified allows the coefficient to differ all across the dimensions accounted by the fixed effects.

4 Data Description

4.1 Air Quality and Weather

I access high frequency data on major air pollutants between 2016 and early 2020 from ARPA Lombardia.[23] Hourly concentrations of four pollutants, $PM_{2.5}$, PM_{10}, ozone, and sulphur dioxide (SO_2), are recorded through a grid of monitoring stations (Figure E.5, Panel [A]). Data on NO_2 are also collected, but not employed in the analysis. ARPA also provides daily estimates of corresponding municipality-level values are calculated by model interpolation[24], and they are used to compute daily average concentrations. ARPA is also responsible for collecting data on principal weather variables: humidity, radiance, rainfall, temperature, wind direction and speed. These measures, all potentially correlated with PM levels[25], are provided uniquely at station level: to match the dimension of air quality measures, municipality-level daily conditions (average temperature, wind direction and speed, humidity, radiance, total rain) are derived taking the value registered at the closest station, as the grid of weather stations provides extensive coverage over the region (Figure E.5, Panel [B]). Alongside weather data, I collect information on Planetary Boundary Layer Height (PBLH) through the ERA5 Reanalysis provided by ECMWF. PBLH is the height above the surface of the ground of the lowest part of the atmosphere, and it is important in affecting weather patterns and the rate of exchange of particulate matter (PM) between the boundary layer and the free atmosphere above it, thus influencing concentrations at ground level (Seidel et al. 2010). The measure is computed on a $0.25° \times 0.25°$ grid, hence its value derived at municipal level is imputed using the corresponding value at higher geospatial level. Each of these variables directly impacts airborne pollutants concentrations.

The initial sample of winter days (November to March) comprises 2046920 day- municipality observations. Subsequent restriction to windows of spreading discontinuity reduces the sample size 111050 data points. Table 1 describes the dataset, including summary statistics for the full sample, winter months and spreading windows days only. Looking at levels of pollutants (Panel [A]), they all exhibit consistent variation both across municipalities and across time. Focusing on levels of particulate matter, as predictable given colder

[23] Regional Agency for Environmental Protection.

[24] Data on concentration of air pollutants are calculated by a private ARPA contractor (AriaNet) using a specific Chemical Transport Model (CTM) called Eulerian FARM-type (Flexible Air Quality Regional Model) (Silibello et al. 2008). FARM-type models account for the transport, chemical conversion and deposition of atmospheric pollutants. Detailed information about this modelling approach is described in Calori et al. (2008). AriaNet also provides an extensive description of the model. See Arianet R2016.12 - FARM (Flexible Air quality Regional Model) formulation and user's Manual (Version 4.11).

[25] Warmer temperatures are usually associated with lower concentrations, given higher thermal dispersion. Positively correlated with temperature, PBLH constitutes an even more cogent measure for vertical dispersion: higher PBL implies increased dispersion capacity and is associated with lower pollutants concentrations. Similarly, increased level of rainfall reduces PM concentrations through "wet deposition". As previously noted, wind speed and direction can affect the presence of pollutants in an area by dispersing pollution plums. With increased humidity, moisture particles grow in size to the point of "dry deposition", reducing PM10 concentrations. Finally, radiance can impact PM levels especially through photochemical reactions.

temperatures and thermal inversion taking place at much lower height,winter months and spreading windows days display consistently higher levels of both PM_{10} and $PM_{2.5}$. Interestingly, the table highlights the extremely poor air quality of Lombardy which, in turn, is matched by a more tolerant policy about warning levels and concentration targets. Indeed, Italy sets the targets for annual average levels for PM_{10} and $PM_{2.5}$ at 40 μ/m^3 and 25 μ/m^3 respectively[26], while the WHO recommendations for the same parameters are respectively 15 μ/m^3 and 5 μ/m^3 (Organization & for Environment 2021). Despite more permissive limitations, warning levels are exceeded frequently, especially during the winter season. Between November and February, the probability of exceeding annual averages is 57% for PM_{10} and 48% for $PM_{2.5}$. Moreover surpassing of control levels happens with similar probability in days when spreading is prohibited against days when it is allowed[27]. Looking at other pollutants, it is important to underline the relatively higher presence of precursors (here only limited to SO_2 and NO_2) to secondary aerosol formation.

Turning to weather variables (Panel [B]), the table shows comparable climate conditions between winter months and window days (temperature, radiance, humidity, wind speed and direction tend to display similar averages and standard deviations). Rainfall levels differ instead between winder and spreading windows, with the latter taking place during relatively drier periods. Assessing the correlation between rainfall and spreading window days by regressing rainfall level on a set of dummies for each $t \in W$, controlling for fixed effects and other weather conditions (Figure E.8 in Appendix), I notice that there is indeed negative correlation when $t \in W^+$, (per regulation, barring spillovers and unexpected weather conditions, spreading is prohibited during rainy days). Yet, the same is true for days right before the opening threshold, and with the size of the coefficients being particularly low, manure application does not strike as a simple proxy for absence of rainfall.

4.2 Livestock concentration and land use

Data on livestock concentrations are accessed through the national zootechnics registry (*Anagrafe Nazionale Zootecnica*) database. The database contains information about number of livestock units and farms for four main breeding animals (chicken, cattle, sheep and goats, swine), calculated at municipal level through a census repeated twice a year (June and December). Summary statistics are reported in Table 2. The table highlights chicken and pig rearing tends to be more concentrated in the region, with a much higher number of animals per farm compared to cow and beef rearing, which is instead more dispersed. The same is true for sheep and goat husbandry, which is however generally carried out in smaller numbers.

Figure 4 shows the presence of farming animals in the region, with the South-West end of the Po Valley emerging as the area with the highest concentrations. As stocks of farming animals are rather stable through time, the number of livestock units is used as a

[26] Legislative Decree 155/2010

[27] Figure E.6 shows the cumulative distribution of daily average concentrations of PM under presence and absence of manure application prohibitions.

cross-sectional proxy for intensity of farming activities in a municipality. I apply two main transformations to the data. First, the livestock units in a municipality are recomputed as the average number of livestock within a 5km-radius buffer. This is motivated as census data may account for large farms by assigning all animals to a single municipality. Yet, in this case animals will not be accounted for in neighbouring municipalities which, however, would be similarly exposed to livestock presence due to geographic proximity. Figure E.9 in Appendix show the result of the buffering process.[28] Secondly, to account for differences among farming animals in terms of organic waste impact, the number of units is recalculated using weights given by relative manure daily production, obtained from Hillel & Hatfield (2005). Weights used are reported in Table 2. Weighting does not induce dramatic changes in the computation of animal stocks, thus it affects only marginally the results and does not influence the main conclusions of the analysis. All additional data on the farming industry in the region, i.e. economic class size of firms in the farming industry, disposal outside of farm territory, are computed as time-invariant measures and provided by the 2010 ISTAT Agricultural Census.

4.3 Hospital admissions and treatment outcomes

Data on hospital admission, prognosis, and discharge are obtained from the National Hospital Discharges database. Access to the full dataset is protected, and was authorized by the Ministry of Health, Planning Department. Data is avaialable between 2016 and the end of 2019. The dataset contains main patient-level characteristics, including gender, age, nationality, marital status. Each record also contains detailed information about the urgency of the hospital admission (i.e. either scheduled or unscheduled), date of admission and discharge, municipality of residence and hospital location. For the purpose of this study, I focus on patients resident and hospitalised in Lombardy. A patient is assigned a principal diagnosis and up to five secondary diagnoses that identify co-existing medical conditions. Similarly, the treatment received by each individual is divided into main procedure and up to five secondary procedures performed during the stay. Moreover, it is included the patient's condition at discharge (death, transfer to another structure, regular discharge).

The dataset does not include a cost measure of the hospitalisation. Medical costs of hospital admissions are usually obtained through billing data, but often patients who benefit from NHS coverage do not sustain the true cost of the treatment received thanks to subsidized healthcare provided by the mostly publicly financed NHS. Indeed, the Italian NHS is a decentralized potential payment system (PPS), where regions developed a regional DRG fee-schedule to classify admissions and to identify DRG tariffs, which are designed to cover most of hospital costs (Fattore & Torbica 2006). The Lombardy region provides a detailed tariff scheme that lists monetary values of different DRG points, with further distinction among multiple categories of hospital services (e.g. ordinary admissions, long-term hospital care etc.). I compensate for the lack of a direct cost measure

[28] Results are comparable using 2km and 10km buffering radius. See the replication material.

using regional tariffs as a proxy for the actual cost of the hospital stay of an individual. The cost calculation depends on the length of the hospital stay and the treatment received by the individual. A detailed explanation of how the measure is computed using information available is included in Section B. Fee-schedule tariffs are likely less sensitive at the margin, as hospitalisations are assigned a fixed reimbursement whenever the duration of the stay does not exceed the DRG-specific threshold, after which the reimbursement is quantified daily. That is, as long as patients are not hospitalised for longer than the threshold assigned to their diagnosis group, hospitalisations of different duration will be compensated with the same amount, which may not reflect precisely the marginal cost faced by the health facility. Yet, the measure still proxies for the true financial burden of the admission on the public system.

I use the information available to define the following variables. First, I compute an aggregate measure at municipality level which contains the number of hospital admissions of patients from the given municipality on each day. Second, I define a set of admission-specific variables. I include an indicator for the hospitalisation terminating with a patient's demise versus regular discharge, and calculate the length of a hospital stay in days. I also compute the cost of hospitalisation using DRG tariffs according to Equation A2, and calculate the ratio between admission cost and number of medical procedures executed on a patient. Procedures count alone importantly abstracts from the complexity of treatment received and is not fully representative of the burden sustained by the healthcare facility. Conversely, the ratio defined aims to capture a monetary form of complexity associated with the hospitalisation: patients exhibiting higher cost per procedure are assumed to require more cost-intensive treatment. For brevity, I refer to this measure as "cost-to-procedure". Finally, I use another proxy for severity derived from a measure provided by the SDO database to gauge the complexity of the hospitalisation. This measure is expressed by a relative weight assigned to DRG codes, relating the average costs of treating patients within one DRG to the average costs of treating all patients included in the DRG system in Italy (Cots et al. 2011).[29] The variable takes value one when the score exactly coincides with the national average. Relative weights of 2 and above are considered serious instances.

To study the impact of manure application through secondary PM emissions, I focus on patients hospitalized for respiratory and cardiovascular medical conditions. The SDO database is organized according to the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) 2007. I select all patients which display as either primary or secondary diagnosis a disease of the respiratory system (ICD 460-519), a disease of the circulatory system (ICD 390-459), or a malignant neoplasm of respiratory and intrathoracic organs (ICD 160-165). This includes 173,405 patients between 2016 and 2019, 83,675 of which are hospitalised during spreading windows. I differentiate between urgent and non-urgent (regular) patients. Despite regular admissions being scheduled in advance, usually in accordance with patient's and practitioner's needs, the health facility maintains the right to reschedule an hospitalisation in order to prevent overcrowding. As

[29] The relative score assigned to DRG categories in Italy is listed in the Ministerial Decree 18/12/2008.

such, fluctuations in the number of urgent admissions required could reflect in variability of regular hospitalisations, and vice-versa.

Table 3 summarizes the main characteristics of the healthcare data. R&CD admissions during spreading windows constitute slightly more than 10% of the total in winter months, and around 3.7% of all admissions in the years considered. Spreading window days show negligible aggregate differences in terms of patients average age, sex, and hospital mortality rate. Also financial variables show little difference between the selected sample and the overall population of patients. An average weight of 1.3 shows how the average cost of R&CDs, calculated between 16.5 and 17 thousands euros, is generally higher than the average over all DRG codes. Figure 5 shows instead the distribution of hospitalisations across healthcare facilities in the region. Each bubble pinpoints a facility, with size being proportional to the number of patients admitted for R&CD, and blue shade indicating generally larger hospitals in terms of treated patients. With few exceptions, the map shows possibly unsurprising positive correlation between overall capacity of the facility and hospitalisations for respiratory and cardiovascular conditions. Furthermore, while the largest number of admissions is once again registered in the Milan area, several large facilities also exist in the area with higher livestock concentration.

5 Results

5.1 Manure Spreading on Air Quality

Panel [A] of Figure 7 reports the estimated pre- and post-treatment coefficients from Equation 1, with the natural logarithm of $PM_{2.5}$ concentrations as outcome of interest. Coefficients for days before the window opening period are reported to visually inspect the absence of clear pre-trends. Even after conditioning for weather conditions and their interaction up to the third lag, a clear jump in $PM_{2.5}$ levels emerges when spreading windows are opened. Pre-trend coefficients are considerably lower in magnitude, and lose significance when standard errors are clustered at station level. I perform an F-test in the spirit of Borusyak & Jaravel (2017), who suggest to drop any two pre-trend terms (usually, the omitted categories as far apart as possible, in my case four and one day before the event) and perform an F-test on the remaining ones. The p-value for the test (reported in Panel [A]) barely rejects absence overall significance at 5% level, and fails to reject at 1% level. Conversely, coefficients for the first 3 days remain strongly significant even after conservative clustering and the addition of municipality monthly trends.

The magnitude of the effect of manure application in days of relaxed prohibitions is estimated under the assumption of absence of pre-trends, setting the coefficients for days prior to the opening window day to zero. I report the estimates of static (Columns 1 to 3) and semi-dynamic models (Columns 4 to 6) in Table 4. Static model estimates show an increase between roughly 26.9% and 27.4% in $PM_{2.5}$ concentrations, extremely stable across specifications. This correspond to around 7.24 to 7.30 $\mu g/m^3$.[30] Semi-dynamic

[30] Results for the same models estimated without log transformation of $PM_{2.5}$ concentrations are reported

estimates highlight a monotonic increase in the impact of manure application on PM up to the third day after a window opens, with a peak at more than 40% after three consecutive spreading days. The impact also appears to fade out quickly, decreasing to roughly a 8.2% increase on day five, with once again relative stability across specifications.

Panel [B] in Figure 7 reports the estimates for the interaction terms in Model 2b. The estimated coefficients follow an increasing path, with the the effect of a manure application expected to be larger for municipalities where livestock is more concentrated. When a spreading window is open, a municipality in the highest decile of the distribution of livestock units, $PM_{2.5}$ concentrations is expected to experience an increase between 7% and almost 10% higher than a municipality in the lowest decile. The effect is also substantially stable to weighting livestock units by animal-specific manure production.

The results of estimating the same models using log concentrations of pollutants other than $PM_{2.5}$ as outcome are listed in Appendix. PM_{10} exhibits a similar trend (Figure E.10), possibly unsurprising thinking that this measure also includes $PM_{2.5}$ particles. Pre-trend coefficients tend to oscillate more, but this could be due by the concurrence of a higher number of sources in the formation of PM_{10} compared to fine particulate matter. Conversely, it is found no impact of spreading windows on the levels of pollutants likely unrelated to farming activities, such as ozone and sulphur dioxide (Figure E.12). The coefficients of interest follow no specific patterns, with scattered significance: positive peaks both to the right and to the left of the opening window period. The increasing pattern of the interactions with livestock concentration indicators also disappears.

As the correlation with rainy events and temperatures implied by the regulatory framework on manure application, despite extensive controls, may pose some concerns regarding the presence of omitted confounders, I further assess the robustness of the results using a placebo test which looks at rain events disregarding manure spreading prohibitions. The regulation imposes no spreading during rainy days, and the day immediately after. Assuming that rainfall is the primary driver of the effect on PM concentrations, one may expect to observe a similar increment every time rainfall follows the same pattern underlying the presence of spreading windows. Hence, I restrict the sample to what I define as "rain windows", rainy days followed by seven days without precipitations. These, per regulation, are viable over which a spreading window as defined in my sample may take place. To make sure that rain windows are not perfectly overlapping with spreading windows, which would nullify the scope of this exercise, I focus on the months not regulated by bulletins (October, December, January and March), where spreading windows are less likely to occur. I then newly estimate Equation 1, where the discontinuity is now set two days away from the rainy event. The results reported in Appendix (Figure E.11) show no similar pattern to the one observed for spreading windows, as coefficients appear not significant, fluctuating in magnitude.

in Table D.2 in Appendix. Log transformation does not impact significantly the magnitude and sign of the estimates.

5.2 Manure Spreading on Health Outcomes

I explore three sets of results that highlight how of spreading-induced air pollution spikes constitute an active threat to individuals. Table 5 reports the estimated coefficients of the static model, where the outcome investigated is now the number of urgent hospital admissions. The estimation comes with the challenge of municipalities (i.e. patients resident in a municipality) recording zero admissions in several periods, with the fitting zero-inflated models through linear regression often results in unreasonable fit (Greene 1994). To overcome this concern, I first estimate a linear model using day-municipality observations with strictly positive hospital admissions (Columns 1 to 3). Second, I compare these results to the same coefficients estimated on the overall sample through the Poisson model in Equation 4 (Columns 5 to 8). The Poisson model requires minimal assumptions on the data distribution, and can account for the many zeros entries in the number of hospitalisations at day-municipality level. Controls and fixed effects are the unchanged from previous regressions. Columns 1 and 2 show how daily admissions during positive PM spikes are expected to be higher by roughly 0.04 units, around 2% increase with respect to the average number of daily admissions. Since PM concentrations are found to peak around the third consecutive day of manure application, the effect of the spike may be stronger towards the end of opening period. To explore this possibility, in Columns 3 and 4 I restrict the sample to observations that are at least two days apart from the cutoff date.[31]. The effect of manure application is even stronger comparing days further apart, since daily admissions now increase by around 0.1 units, with the coefficient significant. The existence of a partial lag in the response of R&CD admissions is also confirmed by estimating the semi-dynamic model (Figure 8). Despite reduced statistical power induced by adding more indicators to the regression, the highest impact is observed on the fourth days into a spreading event, with the coefficient peaking at around 1.08 (Panel A). The same trend is not observed for non-urgent hospitalisations, which are not expected to react to the spike, with coefficients oscillating around zero, never meeting significance at 5% level (Panel B). The pseudo-Poisson estimates show comparable results: lifted prohibitions are associated with an increase in daily urgent hospitalisations by a factor of around 1.046 for the entire sample, and between 1.132 and 1.145 then restricting to days away from the cutoff.

Table 6 illustrates the second set of results concerning mortality at discharge. Model 2 is estimated using the probability of hospital mortality as outcome measure. To enhance interpretability of the coefficients, I use a linear probability model, which also allows to control for a more comprehensive set of fixed effects more easily[32]. Columns 1 to 3 show a 0.7% to 1% increase in the probability of hospital mortality for patients admitted

[31] The reason for excluding days close to the threshold symmetrically must be traced in the definition of spreading window. By considering a spreading event as the five days preceding or following the start of a prohibition period, a lag in the impact of PM spike could confound the effect both in days preceding and following the observed cutoff.

[32] The estimated effect is comparable in magnitude to the one obtained at the mean of all continuous controls in a fixed-effects logit model employing the estimation routine on the pseudo demeaning algorithm developed by Stammann et al. (2016).

during spreading events. This could be the result of the increased number of urgent patients hospitalised as spreading windows open, that is a change in the composition of hospitalisations toward patients in more severe conditions. Still, focusing on urgent patients only in Columns 4 to 6 leads to comparable coefficients, with slightly reduced statistical significance due to losing roughly half of the observations. Hence, even urgent hospitalisations appear more likely to result in a patient's demise when the individual is admitted during a PM spike.

Third, Table 7 shows the estimated impact of manure application on treatment costs and severity. I again separate between the entire sample of patients (Panel A) and urgent admissions only (Panel B)[33]. First, I investigate the effect on the duration of hospital stay in days (Columns 1 and 2). While the negative sign may arise from the highlighted increased mortality rate during hospitalisation, the coefficient for the threshold indicator is never significant and rapidly decreases in magnitude when a more extensive structure of fixed effects is imposed. A comparable patterns is observed for urgent patients. Looking at the "cost-to-procedure" indicator in Columns 3 and 4, there seem to be no significant difference in the average pricing per procedure between patients hospitalised before and after spreading events, although the effect estimated for urgent patients is now stronger in magnitude. Similar considerations apply for the total cost of hospitalisation: the coefficients, while mostly positive and higher for urgent patients, are never significant at 5% level (Column 5 and 6). Lastly, in line with the previous results, the relative financial weight calculated displays positive but non-significant coefficients (Column 7 and 8). The absence of evidence for differential estimated treatment costs per patient during spreading events may have multiple explanations. Firstly, the use of tariffs as an approximation of the cost faced by the healthcare facility may not capture relatively small fluctuations at the margin. Indeed, in a scenario in which the impact of spikes in PM concentrations following manure spreading is not substantial enough to increase the duration of hospital admissions over the DRG-specific threshold, heterogeneous resource utilization spurring from difference in health conditions of inpatients will not result in an increased compensation for the hospital. Moreover, again under the assumptions that higher concentrations of airborne pollutants may not induce severe complications in the majority of patients, healthcare facilities may discharge patients in non-critical conditions also on the basis of budgetary decisions, and the data at hand would not capture general health status at discharge. Being specific to a DRG code and not a single patients, even relative weights may fail to capture any effect on healthcare expenditure per patient, especially under the assumption that increasing PM concentrations will not shift hospital admissions towards diseases with more expensive estimated cost of treatment. While not able to dispel any doubt, these results point towards the existence of no detectable difference in the cost per hospital admissions between outside and during spreading events.

To further ensure my strategy is not capturing underlying trends in hospital activity potentially correlated with winter spreading activities, I propose a placebo test that repeats

[33]The results are mostly unchanged when excluding hospitalisations terminated with the patient's decease.

the same estimation using diseases that are not directly associated with fine particulate matter. I include digestive system diseases (ICD 520-579), genitourinary diseases (ICD 580-629), and musculoskeletal diseases (ICD 710-740). Whereas these pathologies could still be tangentially correlated to fine PM concentrations,[34] their sensitivity is expected to be much lower, and a comparable effect would rather indicate the presence of underlying omitted factors. The results are condensed in Table D.1. Most of the coefficients are not significant and tend to oscillate between positive and negative. The only recognizable pattern is an observed decrease in hospital admissions, which is however considerably lower in magnitude compared to the effect estimated for R&CD. This may be due to the endogenous response of medical facilities, which may choose to delay or relocate admissions in other departments due to resource constraints (e.g. medical staff, operating rooms, etc.) and limited capacity, although this result is not robust. No upward trend in admissions, as well as no increased mortality from patients in other departments is detected, which reinforces the belief that PM spikes are indeed responsible for the deterioration in R&CD patients' conditions.

6 Medical Costs Associated with Manure Application

The results have shown the role of manure application in indirectly influencing the hospital activity and patients' wellbeing through PM emissions. This, in turn, implies a cost for the Italian NHS. I propose a simple back-of-the-envelope calculation to quantify the monetary implications short-term impact of winter spreading activities. This analysis does not encompass the entire cost of manure application borne by Lombardy's population, as it does not account for the potential long-term impact of PM exposure, as well as the presence of symptoms and morbidity not captured by my data as not severe enough to require hospital assistance. As such, this is to be considered an attempt to identify a lower bound to the overall economic cost of manure application in the region. Further research may expand this result with additional evidence on the health concerns of livestock-related emissions other than hospitalisation.

I proceed in two steps. First, the estimates obtained suggest that daily urgent R&CD hospital admissions at the municipal level increase by a factor of 1.04 to 1.145 in response to spreading events. In the 4-year period considered, Lombardy has recorded 355,611 urgent admissions during winter months. Given that spreading activities are allowed 30 out of 120 days, i.e. 25% of winter days, I estimate the average yearly R&CD hospitalisations induced by manure spreading through a simple linear calculation. I find that spreading is responsible for 977.4 to 2814.6 yearly R&CD hospitalisations. This approach rests on the assumption that the effect obtained using a 5-day cutoff is unchanged for windows of different temporal durations. Under the hypothesis that PM may have a more severe and non-linear impact on individuals as exposure to high concentrations is prolonged through time, the impact may vary depending on how prolonged and/or isolated

[34] It is argued in the literature that PM could have an immunosuppression function and create systemic inflammation (Marín-Palma et al. 2023).

manure application opportunities are. Yet, given the tendency of the regulator to lift spreading prohibitions for consecutive days, taking advantage of favorable weather conditions observed in the data, this assumption is deemed suitable for the purpose of my cost analysis. As my estimates show no sign of systematic cost differences of hospitalisations during spreading events, I calculate the additional healthcare expenditure resulting from increased hospital activity using the average cost per urgent R&CD admission, set at 17.1 thousand euros. This leads to a 16.7 to 48.1 million euros expenditure increase per year.

Second, I provide a monetary value to the increase in hospital mortality induced by spreading activities. Given the impossibility of following patients after their discharge in the data, pinpointing with precision the average life expectancy of an individual previously hospitalized for a R&CD disease requires some assumptions. INdeed, I rely on life expectancy at birth, set at 78.9 years for men and 83.9 years for women[35], as a counterfactual to estimate the total of years of life lost in my sample due to spreading activities. Between 2016 and 2019, a total of 20,096 male and 17,633 female casualties upon R&CD hospitalisation took place in winter months, averaging respectively 5024 and 4408 per year. Using the estimated effect in Table 6 and assuming a mortality rate increase by 0.7% to 1% for 30 out of 120 days per year, I calculate a total of 8.8 to 12.6 male and 7.7 to 11 female yearly casualties imputable to manure application. These numbers alone, however, cannot inform on the actual loss in terms of years of life. To derive a total of years of life lost (YLL), I use the gender-specific probability of death upon R&CD hospitalisation for one-year population age groups, computed as the ratio between casualties per age group over total hospital mortality in winter months[36]. I then calculate years of life lost according to the following expression:

$$\text{YLL}_g = \frac{1}{4}\sum_{k=0}^{90} \min\left[\text{Exp}_g - k, 0\right] \cdot \overline{\text{Dth}}_g \cdot \frac{\text{Dth}_{g,k}}{\sum_{k=0}^{90} \text{Dth}_{g,k}} \tag{5}$$

where $\text{Dth}_{g,k}$ is the number of casualties in age group k of patients of gender g and $\overline{\text{Dth}}_g$ the average yearly casualties in R&CD departments. The number of inferred casualties in each age group is multiplied by the corresponding expected life span, given life expectancy at birth (Exp_g).[37]

The calculation of years of life lost leads to a yearly total between 86.3 and 123.3. I multiply this number by the value of a statistical life year (VSLY). Conventionally assumed in the literature at 100,000 US dollars (Cutler 2005), a more recent review of European

[35] Source: ISTAT, Statistical survey of "Deaths of resident population" (Istat/P.5)

[36] The weights distribution can be found in Figure 9.

[37] Using life expectancy at birth of the general population may overestimate the years of life lost imputed to young cohorts, under the assumption that history of R&CD hospitalisation may shorten the life expectancy of individuals. Conversely, this strategy necessarily underestimates the years of life lost of cohorts older than the life expectancy threshold, as these individuals will be mechanically imputed a loss of zero. In the absence of any post-discharge information, and given the heterogeneous set of diseases considered, this calculation can still constitute a reasonable approximation, also considering that the inverted U-shaped relationship with age exhibited by the value of statistical life (VSL) in the literature (Aldy & Viscusi 2008).

studies by Schlander et al. (2017) sets this number at 158,448 euros for continental Europe. I thus obtain an estimated cost of 13.8 to 19.5 million.

Altogether, the back-of-the-envelope calculation performed in this section shows an immediate burden on Lombardy's healthcare system spurring from short-term exposure to high PM concentrations, ranging between 30.9 and 67.7 million euros, which sets around 0.8% to 1.8% of the yearly gross value added by the agriculture industry in the region.[38]

7 Counterfactual Simulation: Air Pollution with PM targets

7.1 Theoretical Framework

The evidence presented in this paper highlights the importance of combining water quality preservation with curbing the impact particulate of matter emissions generated by manure management activities. In 2021, the regional regulator partially acknowledged this concern by adding manure spreading among the prohibited economic activities when temporary measures of air pollution control are in place. As argued throughout the paper, while recognizing spreading as an additional source of air pollution, this policy does not elevate PM concentrations as a driver of manure management activities, but rather as an additional and hardly stringent constraint. However, it is not obvious if, and how, different air pollution targets for farmers should be introduced and what could be the potential gains of bringing PM concentrations into the picture without neither limiting farming activities or improving the current technology for livestock waste disposal.

In this Section, I propose and test the performance of an alternative prohibition scheme that combines the objectives of water- and air-quality preservation, aiming to improve the externalities of management of manure spreading without limiting farmers' activity. The objective of this simple exercise is to minimize the number of days exceeding levels of $PM_{2.5}$ unhealthy for sensitive groups, set at 35 $\mu g/m^3$.[39] The key idea behind this procedure is that a more flexible and comprehensive decision rule of spreading prohibitions could optimally target days of foreseen higher PM concentrations, relieving restrictions when pollutants' levels are expected to be low, conditional on suitable weather conditions. As such, the problem entails a decision rule under constraints, amid imperfect predictability of PM levels in future weather conditions. To this aim, the winter prohibition scheme implemented in the Lombardy region, and in other countries in a similar fashion, appears unnecessarily rigid. The long two-month prohibition interval between December and January reduces flexibility and limits the ability of authorities to allocate spreading windows according to multiple types of goals. I abstract from this constraint in my simulated exercise.

I approach the problem in two steps. First, I evaluate the reduction in days character-

[38] Total value added calculated at chain linked prices (reference year 2015). Source: ISTAT - National Accounts regional main aggregates

[39] This value is defined according to the standards for the Air Quality Index (AQI) set by the US Environmental Protection Agency (EPA). The threshold was adopted after the revision of the Clean Air Act in 2006. See 71 FR 61144 Oct 17, 2006.

ized by hazardous levels of PM (henceforth, HL days) using a decision rule under perfect forecasting. That is, I use data on pollutant concentration, rainfall, and temperature over my sample period under retrospective allocation of prohibitions days by the regulator. This exercise is an abstraction from the true decision-making process of the regulator, which inevitably can only have a reasonable guess on what will be weather conditions and PM levels in the future. Yet, the strategy is useful to define the maximum reduction achievable through redistribution: the optimum achieved under imperfect forecasting can only approach (but likely never reach) the optimum determined under perfect forecasting. In turn, this helps determine the gains to be expected as precision in forecasting increases. The allocation problem can be solved through a standard linear programming allocation algorithm, regulated by the following set of equations:

$$\min_{D} \quad \sum_{i}^{T} f(D_i) = \quad \sum_{i}^{T} (L1_i + L2_i \cdot D_i) \tag{6a}$$

$$\text{s.t.} \qquad \sum_{i} D_i = \frac{T}{4} = W, \tag{6b}$$

$$D_i <= (1 - \mathbb{1}(R_i > 0)) - \varepsilon, \tag{6c}$$

$$D_i <= (1 - \mathbb{1}(R_{i-1} > 0)) + \varepsilon, \tag{6d}$$

$$D_i <= \mathbb{1}(T_i > 0) \cdot \varepsilon + 1 \tag{6e}$$

where D_i is our decision binary variable taking value one when spreading is allowed, zero otherwise. $L1_i$ is an indicator taking value one whenever PM levels exceed hazardous levels in the absence of spreading. This entails calculating counterfactual concentrations in days when spreading is observed in the sample. Formally, $L1_i$ on day i is calculated as follows:

$$L1_i = \mathbb{1}\left\{PM_i - \hat{\eta}_0 PM_i \cdot D_i > \bar{\kappa}\right\}$$

Conversely, $L2_i$ is an indicator taking value one whenever spreading induces PM concentrations to exceed HL under our counterfactual scenario. Formally:

$$L2_i = \mathbb{1}\left\{PM_i + \hat{\eta}_0 PM_i \cdot D_i > \bar{\kappa} > PM_i - \hat{\eta}_0 PM_i \cdot D_i\right\}$$

$\mathbb{1}(R_i > 0)$ and $\mathbb{1}(T_i > 0)$ are two indicators taking value one when respectively positive rainfall and average daily temperature below 0°C are observed on day i. Given $\varepsilon \in \mathbb{R}_0^+$ such that $\varepsilon \to 0$, conditions 6c to 6e ensure that spreading cannot be allocated in rainy days and the day immediately after, and in days when the air temperature entails the risk of frozen ground. Condition 6b are integrates in the problem the 90-days winter prohibition period already imposed by the regulator. Finally, $\hat{\beta}$ represents the counterfactual average effect of spreading activities on PM levels, which is retrieved through the estimation of Equation 2, at 7.8 $\mu g/m^3$.[40]

[40]The estimated $\hat{\eta}_0$ value is reported in Table D.2 in Appendix.

In the day-to-day reality of spreading management, counterfactual PM concentrations cannot be possibly measured retrospectively. A more realistic representation of the decision-making process faced by the regulator implies introducing uncertainty in the model. Specifically, to represent the allocation problem of a forward-looking policymaker, the model needs to incorporate forecasts about future levels of PM. There exist multiple examples of predictive models for PM levels in the literature, each presenting a series of advantages and drawbacks (Gianquintieri et al. 2024). In this paper, I employ a class of Bayesian spatio-temporal models, which exploit the geographically referenced and temporally correlated nature of pollution data, allowing for a minimal set of context-specific assumptions in terms of regional geography (Banerjee et al. 2003, Sahu 2022). Bayesian modeling assumes PM concentrations in site s_i and time t to depend linearly on a set of regressors, accommodating for time-varying coefficients, as well as a random process ($\omega(s_i, t)$) and a pure error term ($\epsilon(s_i, t)$) independent across time and space.

$$\text{PM}(s_i, t) = \mathbf{X}'(s_i, t)\beta(s_i, t) + \omega(s_i, t) + \epsilon(s_i, t), \quad i = 1, \dots, n \quad t = 1, \dots, T \tag{7}$$

Given a moderately large number of locations (i.e. municipalities) in my data, model fitting involves large dimensional matrices. As such, I estimate the model using a Gaussian Predictive Process (GPP) proposed by Sahu & Bakar (2012), which defines an autoregressive structure of the random effects at a smaller number of locations (knots), and then predicts those random effects at the data and prediction locations via interpolation (Kriging method). Methodological details on the GPP model choice and the forecasting methodology can be found in Appendix (Section D.1).

As observable regressors, the model employs the same set of weather covariates used in other specifications. To forecast PM levels, it is assumed that expectations of weather conditions coincide with their true value up to three days in the future. This assumption simplifies the reality of weather forecasting, which is usually estimated with partial error[41]. Yet, given the absence of systematic bias in weather forecasts, the prediction error on weather variables is considered a zero-mean random variable, i.e.

$$\begin{aligned} \tilde{T}_{i+1} &= T_{i+1} + u_{i+1} \\ \mathbb{E}\left[\tilde{T}_{i+1}\right] &= \mathbb{E}\left[T_{i+1} + u_{i+1}\right] \\ &= T_{i+1} + \mathbb{E}\left[u_{i+1}\right] = T_{i+1} \end{aligned} \tag{8}$$

This assumption allows to importantly simplify the algorithm by taking the observed future values of weather regressors as a proxy for forecast values. On the other hand, it necessarily abstracts from inaccuracies in weather forecasts, with implications that will be discussed in the next paragraph.

[41] The Italian Air Force Meteorological Service estimates that temperatures are normally predictable through weather forecasts within a confidence interval of 4 degrees Celsius up to 72 hours in the future. Wind speed and the probability of rainfall are estimated respectively with a 5 m/s confidence interval and an accuracy of 90%. See https://www.meteoam.it/it/rmsc---general-information.

The forecasting of PM levels comes with considerable error.This is related to the idiosyncratic shocks affecting air quality that are not captured by simple weather conditions. As such, my spatio-temporal model does not possess enough precision to simply apply the static algorithm in Equations 6a to 6e substituting actual with predicted values. Instead, I propose a refined dynamic optimization algorithm that compares the predicted daily levels of PM with a reasonable benchmark, identified with the monthly average of the previous year. The rationale behind this approach is to recognize days in which weather conditions are more favorable to limit emissions from spreading and prioritize allocation to those days. Formally, let a location s_0, and let the forecast level of PM at time $i+k$ given by

$$\widehat{\text{PM}}_{i+k} = \Phi\left(\mathbf{X}'_{i+k}(s_{j\neq i}), \omega_{i+k}(s_{j\neq i})\right) \tag{9}$$

In addition, let $\overline{\text{PM}}_i$ be the corresponding monthly average PM concentrations at time i in the previous year. Allocation is determined by solving the following dynamic optimization problem:

$$\min_{D} \quad V(D_i) = \sum_{i}^{T} g(D_i) = \sum_{i}^{T} (D_i + w_i) \exp\left[\overline{\text{PM}}_i - \widehat{\text{PM}}_i\right] \tag{10a}$$

$$\text{s.t.} \quad \sum_{i}^{T} D_i = \frac{T}{4} = W, \tag{10b}$$

$$D_i <= \left(1 - \mathbb{1}\left(\mathbb{E}\left[\tilde{R}_i\right] > 0\right)\right) - \varepsilon, \tag{10c}$$

$$D_i <= (1 - \mathbb{1}\left(\mathbb{E}\left[\tilde{R}_{i-1}\right] > 0\right)) + \varepsilon, \tag{10d}$$

$$D_i <= \mathbb{1}\left(\mathbb{E}\left[\tilde{T}_i\right] > 0\right) \cdot \varepsilon + 1 \tag{10e}$$

where w_i is a time-varying penalty score that depends on the number of days allocated up to period i, i.e.:

$$w_i = \begin{cases} -1 & \text{if } T - i \leq W - \sum_i D_i \\ 1 - \frac{T/W\left(\sum_i W - D_i\right)}{T+1-i-\left(\sum_i W - D_i\right)} & \text{if } T - i > W - \sum_i D_i \end{cases} \tag{11}$$

Despite adding complexity to the problem, this formulation of w_i ensures that the payoff of allowing a spreading day increases when the number of days still to be allocated increases relative to the days remaining till the end of the period, up to the point of forcing allocation regardless of predicted levels of PM when the end period is approaching and constraint 10b has yet to be met[42].

The algorithm is solved via dynamic programming. The three-day assumption on forecast precision of weather covariates conveniently mimics the frequency of bulletins issued by the authority (between 3 and 4 days). As such, the problem is solved iteratively in intervals of three days. After solving the algorithm, I then compare the number of HL

[42]Penalty weights are plotted in Section D.

days obtained under the optimized allocation with the actual number between 2016 and 2020.

7.2 Simulation Results

The results of the allocation algorithms under perfect forecasting and prediction uncertainty are summarized in Table 8. Supplemental material and additional results of the simulations are detailed in Appendix (Section D). The exercise is repeated separately for each winter excluding 2019-20, year in which the number of actual prohibition days was reduced since spreading was exceptionally allowed in December as an emergency response to farmers exceeding storage capacity. In turn, this complicates the comparison with the simulation algorithm. The Table reports the number of days exceeding hazard levels (*HL*) and days allocated to spreading (*Spread*) by climate zone under the current scenario and compares them to the new allocation under perfect and imperfect forecasting. The number of days allocated using the algorithm is higher in both simulations. In the case of perfect information, the discrepancy originates from imperfect compliance with actual regulation in the sample, i.e. prohibitions being extended over 90 days. With imperfect information, this is also due to the tendency of the algorithm to overshoot by one to two days, which is related to the discrete nature of recurring intervals over which optimization is achieved, and the strong penalty assigned to missing out on the allocation of at least 30 days.[43] In those instances where the algorithm allows more spreading days, I reconcile the number of HL days to be expected with fewer prohibitions involved by using the average yearly probability of a spreading day exceeding hazard levels. For the current allocation, the table also reports violations (Column 4), that is days where the weather constraints are not met (either rainfall during the current day or the day before or temperature below the zero mark), but manure application was allowed. I exclude from the calculation municipalities where PM levels never exceed hazard levels throughout the year. Moreover, I exclude municipalities where weather conditions are such that the restrictions imposed on temperature and rainfall impose more than 90 days of prohibitions regardless of air pollution levels.

First, looking at the results for the allocation under perfect forecasting (Columns 4 to 7), it emerges how targeting PM concentrations alongside other regulatory objectives could potentially reduce HL days in the region by around 10.1 to 14.1 percentage points every year. This only represents an upper bound to the maximum abatement achievable and does not inform on how a prohibitions redistribution scheme may perform without the luxury of knowing PM concentrations in advance. In fact, the results under imperfect forecasting (Columns 8 to 11) are far less stunning in terms of HL days reduction. The percentage variation after accounting for partial overshooting (Column 11) oscillates between positive and negative across climate zones. A modest net gain (i.e. a reduction in HL days) is achieved in winter 2016 and 2018 (around 0.8% in both years), whereas in the winter

[43] For instance, if the number of days allocated by the last 3-days forecast window is still 29, the algorithm will consider all possible remaining days as suitable for spreading, allowing up to 32 days. See Appendix for more details.

season of 2017 the algorithm even induces an increase in the HL days, although this is partially justified by that winter having significantly more violations, induced by a wetter and colder season.

Overall, it seems that the refined spreading scheme is not particularly more suitable for keeping PM concentrations within safe levels through simple redistribution, despite the theoretical existence of at least some more room for reduction demonstrated by the static minimization results. This, in turn, gives rise to two considerations. First, the discrepancy between perfect forecasting and allocation of prohibitions under uncertainty implies some room for further improvement in the design of environmental policies regulating manure application. To bridge the gap between the two algorithms, more sophisticated models could be considered to predict PM levels with increased precision. Second, there is only so much that including PM targets can do without any complementary policy action in the farming industry. A perfectly informed management of spreading prohibitions could reduce HL days by around 10%, and even in this scenario, one out of four days would still exceed the danger threshold set for PM concentrations. Alternative measures exist for policymakers to address the concern of secondary emission from manure application, including subsidizing the adoption of best available technologies (e.g. manure injector units) and other mitigation practices (e.g. feeding formulation adjustment) that encompass not only the application phase, but also production and storage (Yan et al. 2024).

8 Discussion and Conclusion

Understanding the implications of manure management activities for air quality is essential for crafting adequate environmental conservation policies and protecting the population from unsafe exposure to pollutants. Yet, it is difficult to perform a residual analysis able to isolate the impact of animal waste disposal, given the presence of multiple confounders and simultaneous sources of airborne pollution. Moreover, uncertainty surrounding the toxicity of different PM chemical species makes it even more cumbersome to infer the hazard imposed by organic fertilizers on human health.

My study exploits the unique regulatory framework of the Lombardy region in Italy to isolate the causal effect of manure spreading on PM concentrations. I find a substantial effect on PM levels, stronger in areas where farming animals are densely concentrated. I also find spreading activities negatively affect health, increasing by a small yet relevant amount of urgent respiratory- and cardiovascular-disease-related hospitalisations and on the mortality rate of patients at discharge. I calculate this effect to account for an immediate financial burden on health care facilities and individual of as much as 37.14 million euros per year. This effect does not take into account the economic cost of long-term repeated exposure to PM spikes, in addition to potentially submerged costs from individual morbidity and sickness not resulting in hospital admission. In this regard, the findings in this paper are in line with the evidence of epidemiological studies showing association between inorganic aerosol particles, respiratory and cardiovascular diseases and mortality (Dockery & Pope 1994, Lippmann & Thurston 1996, Chen, Xu, He, Wang, Du, Du, Qian,

Ji & Li 2018, Joshi et al. 2022).

I conclude my analysis by investigating the potential gains of a more flexible spreading scheme, which incorporates the minimization of winter days with $PM_{2.5}$ concentration exceeding the hazard levels into the utility function of a recursively optimizing regulator. Under the assumption of perfect forecasting, the algorithm achieves an average reduction in hazardous days by around 10%. However, using value function iteration by forecasting future weather conditions and PM levels, the new spreading scheme achieves only negligible reductions in days when PM is in dangerous concentrations.

The paper highlights an important, yet rather specific type of externality implied by manure management. Future research should focus on quantifying other aspects of airborne pollution from farming, such as milder forms of individual sickness. Simulating the impact of a more flexible regulation including PM targets in the objective function of the policymaker calls for more research expanding the results and improving formulas to regulate spreading and reduce environmental concerns. Finally, the paper did not consider how different policies may influence farmers' perception and productivity. Hence, understanding the shape and determinants of the demand curve for environmental regulation in the farming sector constitutes a possible avenue for future research.

References

Aldy, J. E. & Viscusi, W. K. (2008), 'Adjusting the value of a statistical life for age and cohort effects', *The Review of Economics and Statistics* **90**(3), 573–581.

Azizullah, A., Khattak, M. N. K., Richter, P. & Häder, D.-P. (2011), 'Water pollution in pakistan and its impact on public health—a review', *Environment international* **37**(2), 479–497.

Banerjee, S., Carlin, B. P. & Gelfand, A. E. (2003), *Hierarchical modeling and analysis for spatial data*, Chapman and Hall/CRC.

Borusyak, K. & Jaravel, X. (2017), 'Revisiting event study designs', *Available at SSRN 2826228* .

Bouwman, L., Goldewijk, K. K., Van Der Hoek, K. W., Beusen, A. H., Van Vuuren, D. P., Willems, J., Rufino, M. C. & Stehfest, E. (2013), 'Exploring global changes in nitrogen and phosphorus cycles in agriculture induced by livestock production over the 1900–2050 period', *Proceedings of the National Academy of Sciences* **110**(52), 20882–20887.

Braithwaite, I., Zhang, S., Kirkbride, J. B., Osborn, D. P. & Hayes, J. F. (2019), 'Air pollution (particulate matter) exposure and associations with depression, anxiety, bipolar, psychosis and suicide risk: a systematic review and meta-analysis', *Environmental health perspectives* **127**(12), 126002.

Callaway, B. & Sant'Anna, P. H. (2021), 'Difference-in-differences with multiple time periods', *Journal of Econometrics* **225**(2), 200–230.

Calonico, S., Cattaneo, M. D., Farrell, M. H. & Titiunik, R. (2017), 'rdrobust: Software for regression-discontinuity designs', *The Stata Journal* **17**(2), 372–404.

Calori, G., Finardi, S., Nanni, A., Radice, P., Riccardo, S., Bertello, A. & Pavone, F. (2008), 'Long-term air quality assessment: modeling sources contribution and scenarios in ivrea and torino areas', *Environmental Modeling & Assessment* **13**(3), 329–335.

Cambra-López, M., Aarnink, A. J., Zhao, Y., Calvet, S. & Torres, A. G. (2010), 'Airborne particulate matter from livestock production systems: A review of an air pollution problem', *Environmental pollution* **158**(1), 1–17.

Cassee, F. R., Héroux, M.-E., Gerlofs-Nijland, M. E. & Kelly, F. J. (2013), 'Particulate matter beyond mass: recent health evidence on the role of fractions, chemical constituents and sources of emission', *Inhalation toxicology* **25**(14), 802–812.

Chadwick, D., Sommer, S., Thorman, R., Fangueiro, D., Cardenas, L., Amon, B. & Misselbrook, T. (2011), 'Manure management: Implications for greenhouse gas emissions', *Animal Feed Science and Technology* **166**, 514–531.

Chay, K., Dobkin, C. & Greenstone, M. (2003), 'The clean air act of 1970 and adult mortality', *Journal of risk and uncertainty* **27**(3), 279–300.

Chay, K. Y. & Greenstone, M. (2003), 'The impact of air pollution on infant mortality: evidence from geographic variation in pollution shocks induced by a recession', *The quarterly journal of economics* **118**(3), 1121–1167.

Chen, C., Xu, D., He, M. Z., Wang, Y., Du, Z., Du, Y., Qian, Y., Ji, D. & Li, T.

(2018), 'Fine particle constituents and mortality: a time-series study in beijing, china', *Environmental science & technology* **52**(19), 11378–11386.

Chen, S., Oliva, P. & Zhang, P. (2018), Air pollution and mental health: evidence from china, Technical report, National Bureau of Economic Research.

Chen, Y., Ebenstein, A., Greenstone, M. & Li, H. (2013), 'Evidence on the impact of sustained exposure to air pollution on life expectancy from china's huai river policy', *Proceedings of the National Academy of Sciences* **110**(32), 12936–12941.

Correia, S., Guimarães, P. & Zylkin, T. (2020), 'Fast poisson estimation with high-dimensional fixed effects', *The Stata Journal* **20**(1), 95–115.

Cots, F., Chiarello, P., Salvador, X., Castells, X. & Quentin, W. (2011), 'Drg-based hospital payment: Intended and unintended consequences', *Diagnosis-Related Groups in Europe: Moving towards transparency, efficiency and quality in hospitals* pp. 75–92.

Council, N. R. et al. (2004), 'Research priorities for airborne particulate matter: Iv. continuing research progress'.

Currie, J. & Neidell, M. (2005), 'Air pollution and infant health: what can we learn from california's recent experience?', *The Quarterly Journal of Economics* **120**(3), 1003–1030.

Currie, J. & Walker, R. (2011), 'Traffic congestion and infant health: Evidence from e-zpass', *American Economic Journal: Applied Economics* **3**(1), 65–90.

Cutler, D. M. (2005), *Your money or your life: Strong medicine for America's health care system*, Oxford University Press.

De Chaisemartin, C. & d'Haultfoeuille, X. (2020), 'Two-way fixed effects estimators with heterogeneous treatment effects', *American Economic Review* **110**(9), 2964–96.

Deryugina, T., Heutel, G., Miller, N. H., Molitor, D. & Reif, J. (2019), 'The mortality and medical costs of air pollution: Evidence from changes in wind direction', *American Economic Review* **109**(12), 4178–4219.

Dockery, D. W. & Pope, C. A. (1994), 'Acute respiratory effects of particulate air pollution', *Annual review of public health* **15**(1), 107–132.

Domingo, N. G., Balasubramanian, S., Thakrar, S. K., Clark, M. A., Adams, P. J., Marshall, J. D., Muller, N. Z., Pandis, S. N., Polasky, S., Robinson, A. L. et al. (2021), 'Air quality–related health damages of food', *Proceedings of the National Academy of Sciences* **118**(20).

Fattore, G. & Torbica, A. (2006), 'Inpatient reimbursement system in italy: how do tariffs relate to costs?', *Health care management science* **9**(3), 251–258.

Friedman, M. S., Powell, K. E., Hutwagner, L., Graham, L. M. & Teague, W. G. (2001), 'Impact of changes in transportation and commuting behaviors during the 1996 summer olympic games in atlanta on air quality and childhood asthma', *Jama* **285**(7), 897–905.

Galloway, J. N., Townsend, A. R., Erisman, J. W., Bekunda, M., Cai, Z., Freney, J. R., Martinelli, L. A., Seitzinger, S. P. & Sutton, M. A. (2008), 'Transformation of the nitrogen cycle: recent trends, questions, and potential solutions', *Science* **320**(5878), 889–892.

Garg, T., Jagnani, M. & Pullabhotla, H. K. (2023), 'Rural roads, farm labor exits, and

crop fires', *American Economic Journal: Economic Policy* .

Giannakis, E., Kushta, J., Giannadaki, D., Georgiou, G. K., Bruggeman, A. & Lelieveld, J. (2019), 'Exploring the economy-wide effects of agriculture on air quality and health: Evidence from europe', *Science of The Total Environment* **663**, 889–900.

Gianquintieri, L., Oxoli, D., Caiani, E. G. & Brovelli, M. A. (2024), 'State-of-art in modelling particulate matter (pm) concentration: a scoping review of aims and methods', *Environment, Development and Sustainability* pp. 1–23.

Greene, W. H. (1994), 'Accounting for excess zeros and sample selection in poisson and negative binomial regression models'.

Guaita, R., Pichiule, M., Maté, T., Linares, C. & Díaz, J. (2011), 'Short-term impact of particulate matter (pm2.5) on respiratory mortality in madrid', *International Journal of Environmental Health Research* **21**(4), 260–274. PMID: 21644129.
URL: *https://doi.org/10.1080/09603123.2010.544033*

Hausman, C. & Rapson, D. S. (2018), 'Regression discontinuity in time: Considerations for empirical applications', *Annual Review of Resource Economics* **10**, 533–552.

He, G., Liu, T. & Zhou, M. (2020), 'Straw burning, pm2.5, and death: Evidence from china', *Journal of Development Economics* **145**, 102468.
URL: *https://www.sciencedirect.com/science/article/pii/S0304387820300432*

Hillel, D. & Hatfield, J. L. (2005), *Encyclopedia of Soils in the Environment*, Vol. 3, Elsevier Amsterdam.

Imbens, G. & Kalyanaraman, K. (2011), 'Optimal Bandwidth Choice for the Regression Discontinuity Estimator', *The Review of Economic Studies* **79**(3), 933–959.
URL: *https://doi.org/10.1093/restud/rdr043*

Jacobson, T. A., Kler, J. S., Hernke, M. T., Braun, R. K., Meyer, K. C. & Funk, W. E. (2019), 'Direct human health risks of increased atmospheric carbon dioxide', *Nature Sustainability* **2**(8), 691–701.

Jayachandran, S. (2009), 'Air quality and early-life mortality evidence from indonesia's wildfires', *Journal of Human resources* **44**(4), 916–954.

Jerrett, M. (2015), 'The death toll from air-pollution sources', *Nature* **525**(7569), 330–331.

Joshi, P., Dey, S., Ghosh, S., Jain, S. & Sharma, S. K. (2022), 'Association between acute exposure to pm2. 5 chemical species and mortality in megacity delhi, india', *Environmental Science & Technology* **56**(11), 7275–7287.

Kelly, F. J. & Fussell, J. C. (2016), 'Health effects of airborne particles in relation to composition, size and source', *Airborne Particulate Matter: Sources, atmospheric processes and health. London: Royal Society of Chemistry* pp. 344–82.

Kelly, F. J. & Fussell, J. C. (2020), 'Toxicity of airborne particles—established evidence, knowledge gaps and emerging areas of importance', *Philosophical Transactions of the Royal Society A* **378**(2183), 20190322.

Kim, D.-y., de Foy, B. & Kim, H. (2022), 'The investigations on organic sources and inorganic formation processes and their implications on haze during late winter in seoul, korea', *Environmental Research* **212**, 113174.

Kim, K. E., Cho, D. & Park, H. J. (2016), 'Air pollution and skin diseases: Adverse effects of airborne particulate matter on various skin diseases', *Life sciences* **152**, 126–134.

Knittel, C. R., Miller, D. L. & Sanders, N. J. (2016), 'Caution, drivers! children present: Traffic, pollution, and infant health', *Review of Economics and Statistics* **98**(2), 350–366.

Larrieu, S., Jusot, J.-F., Blanchard, M., Prouvost, H., Declercq, C., Fabre, P., Pascal, L., Le Tertre, A., Wagner, V., Rivière, S. et al. (2007), 'Short term effects of air pollution on hospitalizations for cardiovascular diseases in eight french cities: the psas program', *Science of the Total Environment* **387**(1-3), 105–112.

Lelieveld, J., Evans, J. S., Fnais, M., Giannadaki, D. & Pozzer, A. (2015), 'The contribution of outdoor air pollution sources to premature mortality on a global scale', *Nature* **525**(7569), 367–371.

Lippmann, M. & Thurston, G. D. (1996), 'Sulfate concentrations as an indicator of ambient particulate matter air pollution for health risk evaluations.', *Journal of exposure analysis and environmental epidemiology* **6**(2), 123–146.

Liu, J., Kleinman, P. J., Aronsson, H., Flaten, D., McDowell, R. W., Bechmann, M., Beegle, D. B., Robinson, T. P., Bryant, R. B., Liu, H. et al. (2018), 'A review of regulations and guidelines related to winter manure application', *Ambio* **47**(6), 657–670.

Majumdar, D. & Gupta, N. (2000), 'Nitrate pollution of groundwater and associated human health disorders', *Indian journal of environmental health* **42**(1), 28–39.

Maniscalco, J., Hoffmeyer, F., Monsé, C., Jettkant, B., Marek, E., Brüning, T., Bünger, J. & Sucker, K. (2022), 'Physiological responses, self-reported health effects, and cognitive performance during exposure to carbon dioxide at 20 000 ppm', *Indoor air* **32**(1), e12939.

Marín-Palma, D., Fernandez, G. J., Ruiz-Saenz, J., Taborda, N. A., Rugeles, M. T. & Hernandez, J. C. (2023), 'Particulate matter impairs immune system function by up-regulating inflammatory pathways and decreasing pathogen response gene expression', *Scientific Reports* **13**(1), 12773.

Moretti, E. & Neidell, M. (2011), 'Pollution, health, and avoidance behavior evidence from the ports of los angeles', *Journal of human Resources* **46**(1), 154–175.

Neidell, M. (2009), 'Information, avoidance behavior, and health the effect of ozone on asthma hospitalizations', *Journal of Human resources* **44**(2), 450–478.

Neidell, M. J. (2004), 'Air pollution, health, and socio-economic status: the effect of outdoor air quality on childhood asthma', *Journal of health economics* **23**(6), 1209–1236.

Nian, Y. (2023), 'Incentives, penalties, and rural air pollution: Evidence from satellite data', *Journal of Development Economics* **161**, 103049.

Ongley, E. D. (1996), *Control of water pollution from agriculture*, Vol. 55, Food & Agriculture Org.

Organization, W. H. & for Environment, E. C. (2021), *WHO global air quality guidelines: particulate matter (PM2. 5 and PM10), ozone, nitrogen dioxide, sulfur dioxide and*

carbon monoxide, World Health Organization.

Pope III, C. A., Verrier, R. L., Lovett, E. G., Larson, A. C., Raizenne, M. E., Kanner, R. E., Schwartz, J., Villegas, G. M., Gold, D. R. & Dockery, D. W. (1999), 'Heart rate variability associated with particulate air pollution', *American heart journal* **138**(5), 890–899.

Reiss, R., Anderson, E. L., Cross, C. E., Hidy, G., Hoel, D., McClellan, R. & Moolgavkar, S. (2007), 'Evidence of health impacts of sulfate-and nitrate-containing particles in ambient air', *Inhalation toxicology* **19**(5), 419–449.

Sahu, S. (2022), *Bayesian modeling of spatio-temporal data with R*, CRC Press.

Sahu, S. K. & Bakar, K. S. (2012), 'Hierarchical bayesian autoregressive models for large space–time data with applications to ozone concentration modelling', *Applied Stochastic Models in Business and Industry* **28**(5), 395–415.

Schlander, M., Schaefer, R. & Schwarz, O. (2017), 'Empirical studies on the economic value of a statistical life year (vsly) in europe: what do they tell us?', *Value in Health* **20**(9), A666.

Schlenker, W. & Walker, W. R. (2016), 'Airports, air pollution, and contemporaneous health', *The Review of Economic Studies* **83**(2), 768–809.

Schlesinger, R. B. & Cassee, F. (2003), 'Atmospheric secondary inorganic particulate matter: the toxicological perspective as a basis for health effects risk assessment', *Inhalation toxicology* **15**(3), 197–235.

Schwartz, J., Bind, M.-A. & Koutrakis, P. (2017), 'Estimating causal effects of local air pollution on daily deaths: effect of low levels', *Environmental health perspectives* **125**(1), 23–29.

Sebilo, M., Mayer, B., Nicolardot, B., Pinay, G. & Mariotti, A. (2013), 'Long-term fate of nitrate fertilizer in agricultural soils', *Proceedings of the National Academy of Sciences* **110**(45), 18185–18189.

Seidel, D. J., Ao, C. O. & Li, K. (2010), 'Estimating climatological planetary boundary layer heights from radiosonde observations: Comparison of methods and uncertainty analysis', *Journal of Geophysical Research: Atmospheres* **115**(D16).

Silibello, C., Calori, G., Brusasca, G., Giudici, A., Angelino, E., Fossati, G., Peroni, E. & Buganza, E. (2008), 'Modelling of pm10 concentrations over milano urban area using two aerosol modules', *Environmental Modelling & Software* **23**(3), 333–343.

Squizzato, S., Masiol, M., Brunelli, A., Pistollato, S., Tarabotti, E., Rampazzo, G. & Pavoni, B. (2013), 'Factors determining the formation of secondary inorganic aerosol: a case study in the po valley (italy)', *Atmospheric chemistry and physics* **13**(4), 1927–1939.

Stammann, A., Heiss, F. & McFadden, D. (2016), 'Estimating fixed effects logit models with large panel data'.

Sun, L. & Abraham, S. (2021), 'Estimating dynamic treatment effects in event studies with heterogeneous treatment effects', *Journal of Econometrics* **225**(2), 175–199.

WHO (2021), *WHO global air quality guidelines: particulate matter (PM2.5 and PM10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide*, World Health Organiza-

tion.

Wong, C.-M., Vichit-Vadakan, N., Kan, H. & Qian, Z. (2008), 'Public health and air pollution in asia (papa): a multicity study of short-term effects of air pollution on mortality', *Environmental health perspectives* **116**(9), 1195–1202.

Xia, F., Xing, J., Xu, J. & Pan, X. (2022), 'The short-term impact of air pollution on medical expenditures: Evidence from beijing', *Journal of Environmental Economics and Management* **114**, 102680.
URL: *https://www.sciencedirect.com/science/article/pii/S0095069622000523*

Yan, X., Ying, Y., Li, K., Zhang, Q. & Wang, K. (2024), 'A review of mitigation technologies and management strategies for greenhouse gas and air pollutant emissions in livestock production', *Journal of Environmental Management* **352**, 120028.
URL: *https://www.sciencedirect.com/science/article/pii/S0301479724000148*

Young, R. & Mutchler, C. (1976), Pollution potential of manure spread on frozen ground, Technical report, Wiley Online Library.

Zhang, J., Cheng, H., Di Narzo, A., Zhu, Y., Shan, M., Zhang, Z., Shao, X., Chen, J., Wang, C. & Hao, K. (2022), 'Within- and cross-tissue gene regulations were disrupted by pm2.5 nitrate exposure and associated with respiratory functions', *Science of The Total Environment* **850**, 157977.
URL: *https://www.sciencedirect.com/science/article/pii/S0048969722050768*

Zhang, J., Cheng, H., Wang, D., Zhu, Y., Yang, C., Shen, Y., Yu, J., Li, Y., Xu, S., Zhang, S. et al. (2021), 'Chronic exposure to pm2. 5 nitrate, sulfate, and ammonium causes respiratory system impairments in mice', *Environmental Science & Technology* **55**(5), 3081–3090.

Zivin, J. G. & Neidell, M. (2009), 'Days of haze: Environmental information disclosure and intertemporal avoidance behavior', *Journal of Environmental Economics and Management* **58**(2), 119–128.

Tables

Table 1: Summary statistics - Pollutants and Weather Measures

	Full year			November - March			Spreading windows		
	Overall	Within	Between	Overall	Within	Between	Overall	Within	Between
Panel A - Pollutants									
PM_{10}	22.61			30.49			34.02		
($\mu g/m^3$)	(16.10)	(14.57)	(6.85)	(18.89)	(16.10)	(9.88)	(21.70)	(19.19)	(10.27)
	[0.2; 205.8]	[-9.3; 200.1]	[4.7; 36.1]	[0.2; 201.3]	[-16.2; 187.6]	[3.7 ; 53.5]	[0.7; 198.0]	[-11.2; 187.8]	[4.9; 54.7]
$PM_{2.5}$	18.41			25.75			24.34		
($\mu g/m^3$)	(13.34)	(12.42)	(4.87)	(15.50)	(13.40)	(7.79)	(14.75)	(13.27)	(6.61)
	[0.2; 167.3]	[-8.9; 161.9]	[4.2; 30.3]	[0.2 ; 167.3]	[-16.9; 158.8]	[3.3; 50.4]	[0.4; 137.9]	[-9.5; 134.7]	[3.9; 45.1]
NO_2	23.21			33.06			40.11		
($\mu g/m^3$)	(16.59)	(13.30)	(9.92)	(17.19)	(11.98)	(12.34)	(21.71)	(17.54)	(12.95)
	[0.0; 174.5]	[-19.5; 157.7]	[1.1; 56.4]	[0.0; 148.3]	[-16.9; 119.6]	[1.8; 68.8]	[0.0; 157.6]	[-14.5; 139.7]	[7.3; 76.7]
Ozone	98.38			61.08			57.57		
($\mu g/m^3$)	(43.87)	(43.56)	(5.21)	(28.84)	(27.11)	(9.84)	(18.92)	(16.22)	(9.80)
	[1.9; 370.8]	[-3.0; 376.0]	[79.2 ; 112.8]	[1.9; 359.5]	[-6.2; 359.2]	[38.7; 90.5]	[6. ; 263.3]	[3.2; 252.1]	[34.9; 86.8]
SO_2*	2.51			2.90			2.86		
($\mu g/m^3$)	(2.34)	(2.10)	(1.03)	(2.54)	(2.33)	(1.01)	(2.45)	(2.17)	(1.14)
	[0.0; 204.5]	[-3.0; 201.7]	[1.4; 5.6]	[0.0; 204.5]	[-2.8; 201.8]	[1.8 ; 5.7]	[0.0; 64.3]	[-3.2; 61.1]	[1.4; 6.0]
Panel B - Weather									
Temperature	13.15			5.35			7.13		
(° C)	(8.32)	(8.04)	(2.16)	(4.25)	(3.90)	(1.70)	(4.05)	(3.77)	(1.53)
	[-18.7 ; 37.3]	[-13.3 ; 42.0]	[0.5 ; 15.9]	[-18.7 ; 21.3]	[-14.6 ; 20.9]	[-6.2 ; 8.3]	[-16.5 ; 23.3]	[-6.7 ; 23.0]	[-3.3 ; 9.8]
Rainfall	0.15			0.12			0.05		
(mm)	(11.29)	(11.28)	(0.27)	(11.00)	(10.99)	(0.43)	(0.74)	(0.74)	(0.08)
	[0.0; 3,250.8]	[-2.1; 3,248.7]	[0.0; 2.3]	[0.0; 2,812.8]	[-4.5 ; 2,808.3]	[0.0; 4.7]	[0.0; 255.5]	[-2.5; 253.1]	[0.0; 2.5]
Wind Speed	2.25			2.13			1.96		
(m/s)	(1.34)	(1.12)	(0.73)	(1.51)	(1.28)	(0.80)	(1.43)	(1.26)	(0.68)
	[0.0; 26.3]	[-3.7; 26.7]	[1.0; 6.5]	[0.0; 26.3]	[-4.5 ; 26.5]	[0.8; 7.1]	[0.0; 20.5]	[-2.8 ; 19.2]	[0.7; 6.2]
Wind Direction	178.87			189.17			187.97		
(Degree (°))	(108.81)	(103.42)	(33.95)	(110.65)	(102.94)	(40.63)	(110.89)	(105.63)	(33.80)
	[0.0; 360.0]	[-86.5; 484.9]	[53.9; 266.4]	[0.0; 360.0]	[-101.5; 506.7]	[42.2; 292.6]	[0.0; 360.0]	[-96.4; 473.1]	[73.3; 298.0]
Radiance	158.96			80.63			71.90		
(W/m^2)	(103.09)	(102.74)	(8.39)	(59.85)	(59.54)	(6.09)	(44.90)	(44.46)	(6.33)
	[0.0; 1,156.7]	[-22.8; 1,147.4]	[115.4; 181.7]	[0.0; 545.9]	[-19.3; 544.6]	[40.3; 102.4]	[0.0; 321.8]	[-22.4; 317.3]	[32.9; 94.5]
Humidity	73.90			80.47			81.90		
(%)	(16.62)	(16.00)	(4.50)	(18.60)	(16.79)	(8.02)	(16.43)	(15.20)	(6.28)
	[0; 100]	[-8; 111]	[63; 82]	[0; 100]	[-12; 123]	[56; 93]	[0; 100]	[-6; 122]	[57; 91]
PBLH	1.65			0.38			0.41		
(km)	(1.44)	(1.44)	(0.11)	(0.35)	(0.34)	(0.07)	(0.38)	(0.37)	(0.09)
	[0.0; 5.7]	[-0.2; 5.8]	[1.3; 1.8]	[0.0; 3.1]	[-0.1; 3.2]	[0.2; 0.5]	[0.0; 3.1]	[-0.1; 3.0]	[0.2 ; 0.6]

Notes: summary statistics report mean values, standard deviation (in parentheses), minimum and maximum value (in brackets). Statistics are reporting including all municipality-day observations from 2016 to 2019 (2046920 obs.), months from November till March (848808 obs.), and only spreading window days (111050 obs.).

* Data on SO_2 are only provided by ARPA at station level. Municipality-level estimates have been computed through interpolation, i.e. distance-weighted average of 4 closest stations.

Table 2: Summary statistics - Livestock

Animal	Livestock Units	Farms	Weight
Chicken	45,985.8	1.6	0.34
	(116,685.7)	(2.5)	
	[0;3,616,334]	[0;21]	
Cow and Beef	1,040.2	13.4	74.77
	(2,402.3)	(16.4)	
	[0;31,517]	[0;225]	
Sheep and Goat	134.3	10.3	40
	(347.0)	(13.8)	
	[0;6,267]	[0;192]	
Pigs	2,916.5	1.9	76.73
	(8,953.3)	(3.9)	
	[0;94,935]	[0;46]	

Notes: summary statistics report mean values, standard deviation (in parentheses), minimum and maximum value (in brackets). Column 3 reports statistics for the number of farms per municipality .

* Weights are computed using relative manure daily production in kg, obtained from Hillel & Hatfield (2005).

Table 3: Summary statistics - Hospital Admissions

	Full year (2016-19)		Nov - Feb (2016-19)		Spreading windows (2016-19)	
	R&CD	Urgent R&CD	R&CD	Urgent R&CD	R&CD	Urgent R&CD
N. admissions	1,831,792	968,667	641,880	355,611	67,979	37,315
Admissions by hospital*	2,026.00 (1,507.95)	1,294.67 (821.63)	714.35 (519.27)	478.85 (292.73)	77.90 (57.83)	54.07 (36.14)
Admissions by municipality*	3,302.49 (7,457.50)	1,918.51 (4,184.33)	1,143.87 (2,588.08)	682.68 (1,498.93)	116.91 (257.81)	62.63 (137.29)
Age	66.08 (24.68)	68.64 (23.72)	65.37 (27.16)	67.07 (25.91)	65.87 (25.15)	68.15 (24.71)
Female (%)	0.43	0.45	0.43	0.46	0.43	0.45
Italian (%)	0.96	0.95	0.95	0.95	0.96	0.95
Mortality rate**						
0-10	3.56	3.63	2.96	2.71	2.83	3.34
10-20	5.72	12.28	5.90	11.68	2.58	6.79
20-30	7.10	22.14	7.52	22.73	9.51	32.93
30-40	10.54	28.43	10.74	28.19	9.49	21.85
40-50	16.51	37.20	16.91	38.75	16.18	37.88
50-60	24.43	49.65	25.20	49.82	25.88	52.96
60-70	32.26	62.90	34.87	64.68	34.10	64.17
70-80	48.78	83.59	54.11	90.02	47.86	81.27
80+	105.31	137.59	115.35	147.27	105.66	136.73
Total					54.84	89.54
Length (days)	10.82 (13.26)	10.86 (10.39)	10.77 (13.19)	10.77 (10.17)	10.50 (12.16)	10.78 (9.80)
N. Procedures	3.35 (1.72)	3.55 (1.65)	3.34 (1.72)	3.52 (1.65)	3.32 (1.72)	3.54 (1.64)
Cost (thousands euros)	17.12 (122.82)	17.29 (139.46)	16.81 (121.25)	16.85 (137.15)	16.55 (108.48)	17.13 (125.98)
Cost-to-procedure (thousands euros)	1.35 (2.20)	0.99 (2.09)	1.32 (2.18)	0.97 (2.07)	1.38 (2.95)	1.08 (2.12)
Severity (weight)	1.36 (1.30)	1.38 (1.30)	1.35 (1.29)	1.35 (1.29)	1.33 (1.21)	1.37 (1.24)

Notes: summary statistics report mean values, standard deviation (in parentheses), minimum and maximum value (in brackets) for all four years (2016-19), winter months and days of spreading windows only, including respiratory and cardiovascular diseases (R&CD).

* Admissions per million of inhabitants.

** Number of casualties per 1000 admissions.

Table 4: Effect of spreading windows on $PM_{2.5}$

	Static Model			Semi-dynamic model		
log($PM_{2.5}$)	(1)	(2)	(3)	(4)	(5)	(6)
Window	0.269***	0.274***	0.274***			
	(0.004)	(0.004)	(0.014)			
Day 1				0.221***	0.225***	0.225***
				(0.006)	(0.006)	(0.020)
Day 2				0.352***	0.358***	0.358***
				(0.004)	(0.004)	(0.014)
Day 3				0.399***	0.406***	0.406***
				(0.003)	(0.003)	(0.015)
Day 4				0.137***	0.134***	0.134***
				(0.002)	(0.002)	(0.008)
Day 5				0.079***	0.057***	0.057***
				(0.003)	(0.003)	(0.017)
Obs	112213	112207	112207	112213	112207	112207
Adj. R^2	0.768216	0.802428	0.799565	0.774324	0.809452	0.806691
Weather (extended)	✓	✓	✓	✓	✓	✓
Municipality FE	✓	✓	✓	✓	✓	✓
Municipality-by-month FE		✓	✓		✓	✓
Month-by-year FE	✓	✓	✓	✓	✓	✓
DoW & Holiday FE		✓	✓		✓	✓
Standard errors	Rob	Rob	Clust	Rob	Rob	Clust

Notes: the table reports the estimated η_0 from Equation 2 (Columns 1 to 3) and the estimated η_k coefficients from Equation 1, where pre-trend coefficients are set to zero. Weather controls include temperature, wind direction, wind speed, rainfall, radiance, humidity, and average planetary boundary layer height, interacted with each other up to three lags. Robust (Columns 1-2 and 4-5) and clustered at sensor level (Columns 3 and 6) standard errors, are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 5: Effect of spreading windows on hospital admissions

	Linear model				Poisson			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Window	0.040**	0.041*			0.046**	0.048**		
	(0.020)	(0.021)			(0.021)	(0.021)		
Window†			0.104**	0.102**			0.132***	0.145***
			(0.051)	(0.052)			(0.041)	(0.041)
Obs	19,689	19,689	11,587	11,587	80225	80225	46210	46210
Adj. R^2	0.94	0.94	0.94	0.95				
Dep. Var mean					0.42	0.42	0.42	0.42
Dep. Var mean (>0)	1.81	1.81	1.81	1.81				
Weather Contr.	✓	✓	✓	✓	✓	✓	✓	✓
Distant Days			✓	✓			✓	✓
Month-by-year FE	✓	✓	✓	✓	✓	✓	✓	✓
Municipality FE	✓	✓	✓	✓	✓	✓	✓	✓
DoW & Holiday FE	✓	✓	✓	✓	✓	✓	✓	✓
Month-by-prov FE		✓		✓		✓		✓
Window FE		✓		✓		✓		✓

Notes: the table reports the estimated impact of a spreading window on daily urgent R&CD hospital admissions registered in the five days before and five days after an opening event. Weather controls include temperature, wind direction, wind speed, rainfall, radiance and humidity, interacted with each other at up to the third lag. Columns 1 to 4 include only strictly positive observations for hospital admissions, and are estimated through a linear model. Columns 5 and 8 include observations with no admissions, and are estimated through pseudo-Poisson ML fixed effects model.
†: only days at least two days away from the cutoff date are considered to account for potential lag in the effect of PM spikes.
Standard errors, clustered at municipal level, are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 6: Effect of spreading windows on hospital mortality

	All patients			Urgent patients		
	(1)	(2)	(3)	(4)	(5)	(6)
Window	0.010***	0.008***	0.007***	0.009**	0.011**	0.011**
	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)
Dep. Var mean	0.063	0.063	0.063	0.108	0.108	0.108
Obs	57902	57900	57844	30090	30090	30038
Adj. R^2	0.002	0.063	0.069	0.002	0.057	0.060
Pseudo R^2						
Weather contr.	✓	✓	✓	✓	✓	✓
Indiv. contr.		✓	✓		✓	✓
Municipality FE	✓	✓	✓	✓	✓	✓
Month-by-year FE	✓	✓	✓	✓	✓	✓
Month-by-prov FE			✓			✓
Hospital FE			✓			✓
Window FE			✓			✓

Notes: The table reports the estimated impact of a spreading window on hospital mortality. Weather controls include temperature, wind direction, wind speed, rainfall, radiance, humidity, and average planetary boundary layer height, interacted with each other at time t. Individual controls include sex, age class, disease. Standard errors, clustered at municipal level, are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 7: Effect of spreading windows on length and complexity of stays

	Length		Cost-to-procedure		Total cost		Severity (weight)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	All patients							
Window	-0.148	-0.040	0.011	0.022	-0.417	0.123	0.018	0.006
	(0.113)	(0.105)	(0.288)	(0.276)	(1.022)	(0.991)	(0.013)	(0.011)
Dep. Var mean	10.061	10.061	4.255	4.255	13.989	13.989	1.335	1.335
Obs	66,771	66,706	55,955	55,900	66,684	66,619	66,772	66,707
R^2	0.03	0.23	0.02	0.04	0.01	0.03	0.02	0.28
	Urgent patients							
Window	-0.141	-0.088	0.371	0.513	0.673	1.426	0.015	0.020
	(0.125)	(0.143)	(0.547)	(0.580)	(1.764)	(1.907)	(0.017)	(0.015)
Dep. Var mean	10.458	10.458	4.482	4.482	14.084	14.084	1.356	1.356
Obs	36,175	36,117	30,058	30,007	36,167	36,109	36,175	36,117
R^2	0.05	0.14	0.04	0.07	0.03	0.06	0.04	0.39
Weather contr.	✓	✓	✓	✓	✓	✓	✓	✓
Indiv. contr.		✓		✓		✓		✓
Municipality FE	✓	✓	✓	✓	✓	✓	✓	✓
Month-by-year FE	✓	✓	✓	✓	✓	✓	✓	✓
Month-by-prov FE		✓		✓		✓		✓
Hospital FE		✓		✓		✓		✓
Window FE		✓		✓		✓		✓
DoW & Holiday FE		✓		✓		✓		✓

Notes: The table estimates the impact of a spreading window on indicators of healthcare utilization and expenditure. Weather controls include temperature, wind direction, wind speed, rainfall, radiance, humidity, and average planetary boundary layer height, interacted with each other at time t. Individual controls include sex, age class, disease. Cost-to-procedure and total cost of hospitalisation are derived from average pricing and reimbursement tables determined at regional level.

Table 8: Optimized spreading scheme - summary

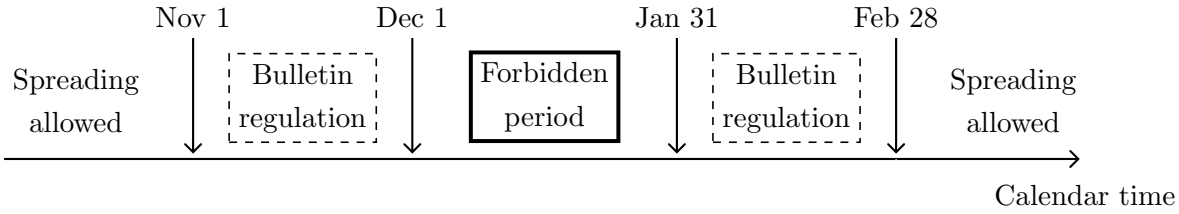
Climate zone	Current allocation			Perfect information				Imperfect information			
	(1) HL	(2) Spread	(3) Violations	(4) HL	(5) HL*	(6) Spread	(7) % change	(8) HL	(9) HL*	(10) Spread	(11) % change
2016-2017											
Alps	972	1050	10	854	854	1050	-12.14	1016	1002	1072	3.09
Central plain	16704	8250	0	15129	15129	8250	-9.43	16722	16630	8396	-0.44
West plain	17134	9780	4	15039	15039	9780	-12.23	16955	16833	9974	-1.76
East plain	2672	1410	0	2450	2450	1410	-8.31	2750	2732	1439	2.25
West Prealps	17693	9750	8	16197	16197	9750	-8.46	17708	17616	9897	-0.44
East Prealps	9297	5880	4	8226	8226	5880	-11.52	9197	9130	5986	-1.8
Total	64472	36120	26	57895	57895	36120	-10.2	64348	63943	36764	-0.82
2017-2018											
Alps	818	1440	23	778	778	1440	-4.89	931	927	1449	13.33
Central plain	9914	6168	76	8974	8247	7674	-16.81	10651	9831	7868	-0.84
West plain	10314	10260	16	9119	9119	10260	-11.59	11022	10952	10406	6.19
East plain	2085	1296	0	1911	1758	1614	-15.68	2284	2133	1609	2.3
West Prealps	13111	9750	273	11719	11719	9750	-10.62	13124	13058	9886	-0.4
East Prealps	6168	5886	449	5647	5481	6231	-11.14	6430	6067	6639	-1.64
Total	42410	34800	837	38148	37102	36969	-12.52	44442	42967	37857	1.31
2018-2019											
Alps	303	630	8	302	302	630	-0.33	411	406	640	26.28
Central plain	10899	8400	0	9670	9670	8400	-11.28	10448	10414	8468	-4.32
West plain	13428	11280	0	12220	12220	11280	-9	13263	13193	11417	-1.24
East plain	1902	1530	0	1729	1729	1530	-9.1	1848	1834	1560	-2.92
West Prealps	7691	9850	19	6925	6925	9850	-9.96	7932	7797	10116	3.04
East Prealps	4154	5310	9	3790	3790	5310	-8.76	4148	4099	5414	-0.14
Total	38377	37000	36	34636	34636	37000	-9.75	38050	37742	37615	-0.86

Notes: The table shows the variation in days exceeding hazard levels of PM by climate zone, together with the number of spreading days allocated. The penalty scheme induces the algorithm to prioritize allocating more rather than less days than the imposed target of 30. Winter 2019-20 is excluded due to imperfect compliance with the fixed prohibition period by the regulator (spreading allowed in mid-December), which makes comparison more cumbersome.

* Days exceeding hazard levels are recalculated accounting for fewer prohibitions than what imposed by the algorithm, using the average yearly probability of a spreading day exceeding hazard levels.

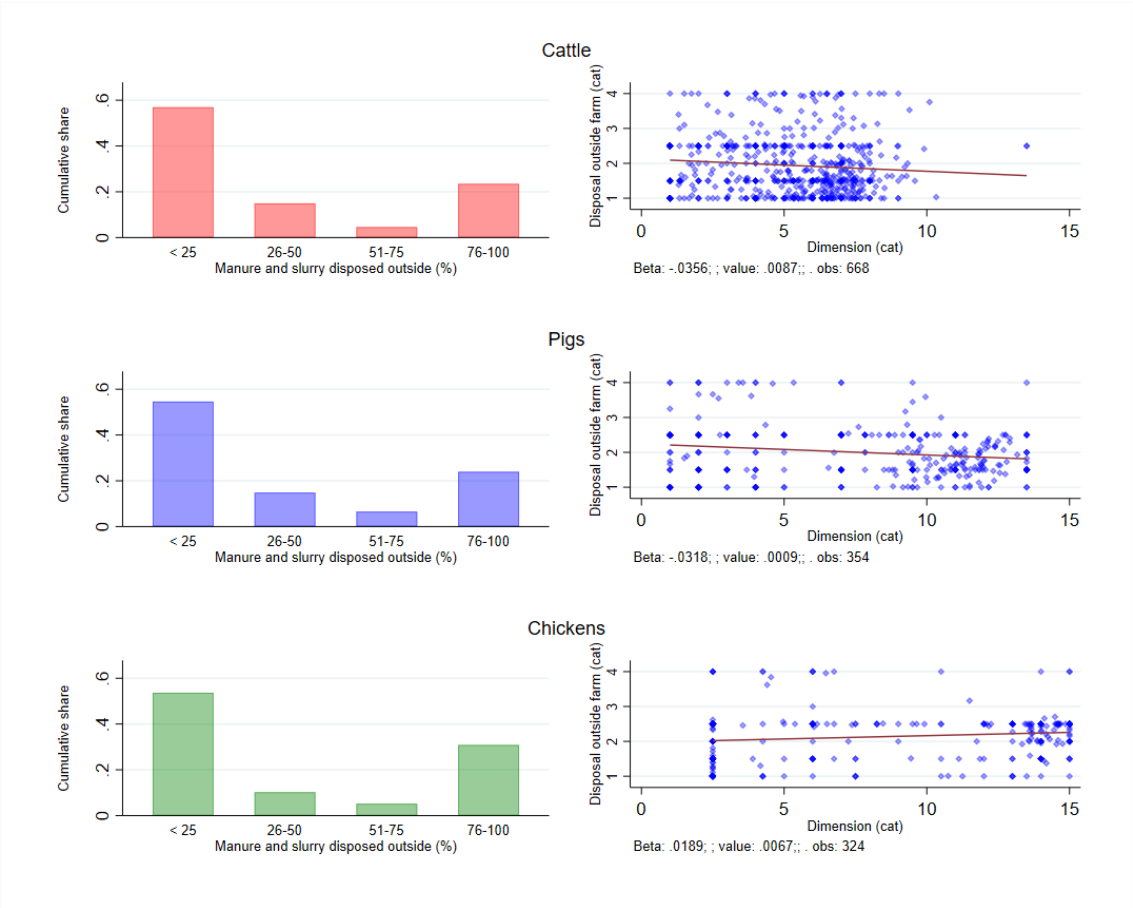
Figures

Figure 1: Prohibition periods structure in Lombardy



Notes: the figure summarizes the periods in which spreading is conditionally or entirely forbidden. This prohibition scheme has been in force in Lombardy between 2016 and 2021.

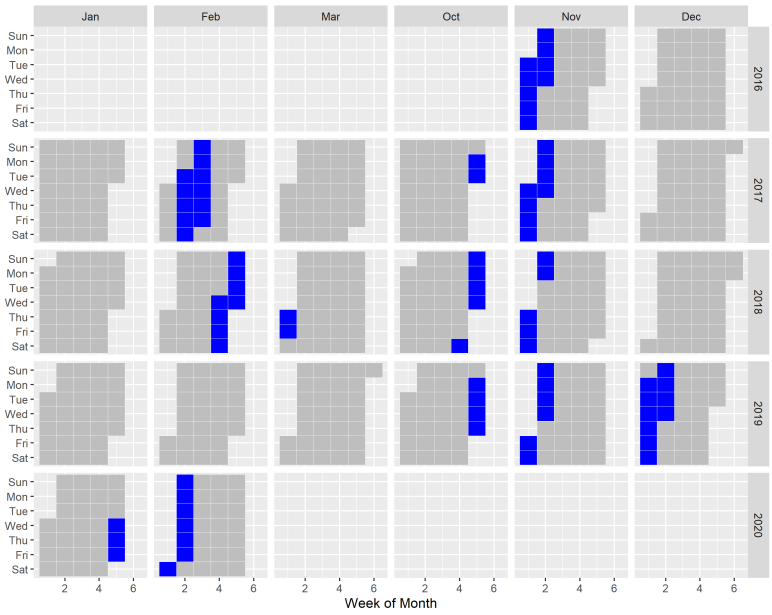
Figure 2: Frequency of manure disposal outside farm and firm size



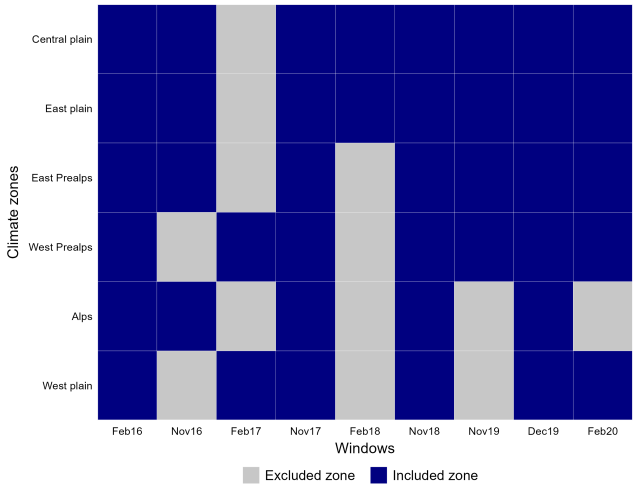
Notes: to the left, the figure shows the average cumulative share at municipal level of manure and slurry disposed outside of the farm, divided in 4 categories. To the right, the figure plots the relationship between disposal outside the farm and average farm dimension at municipality level. The variable on the Y-axis is calculated at municipal level as the mean of outside disposal categories 1 (<25) to 4 (76-100) of each farm within the municipality weighted by the relative concentration of animals of the farm. The same weights are applied to the X-axis variables, which is calculated at municipal level as the weighted mean of dimension categories 1 (1-2) to 15 (50000 +) of each farm within the municipality. Coefficient, robust p-values and number of observations of the fitted regression line are reported. Data is not available for goats and sheeps. Source: 2010 Agricultural Census, ISTAT.

Figure 3: Window days calendar distribution and included climate zones

[A] Window days distribution

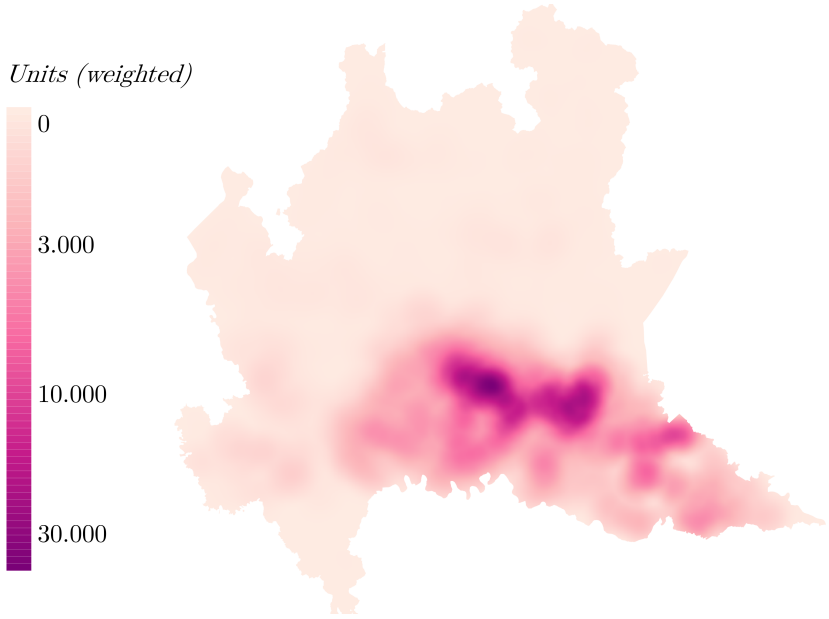


[B] Treated municipalities



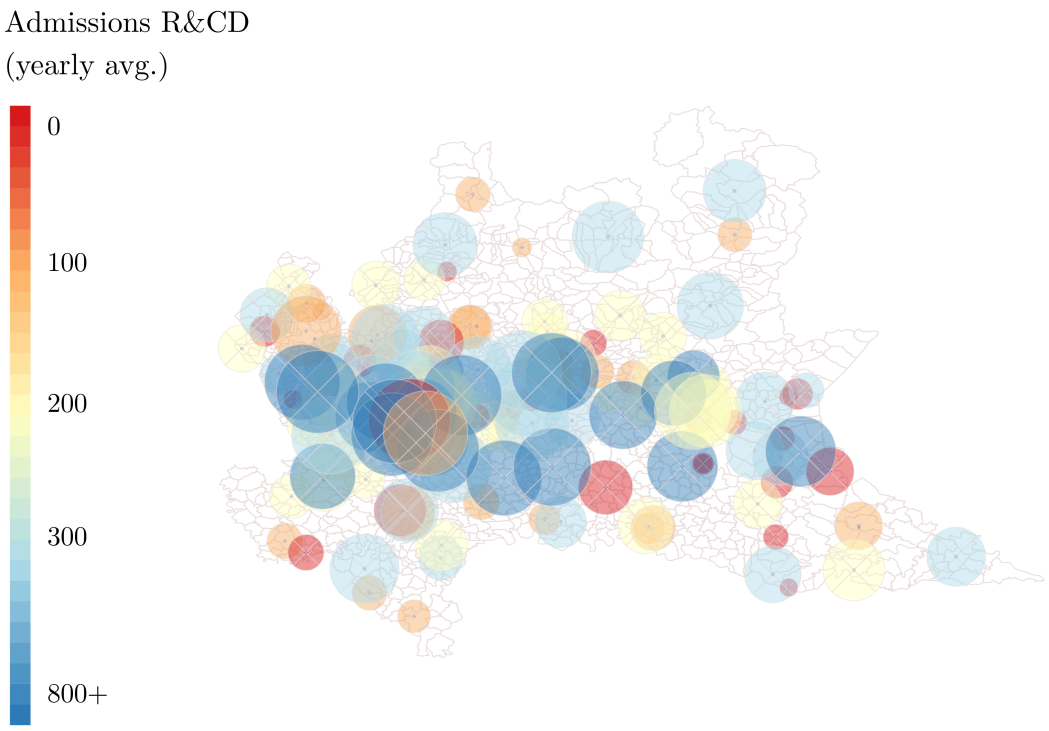
Notes: Panel [A] shows the calendar distribution of window days since 2016. One window is located during the yearly fixed prohibition period (December 2019) due to the result of extraordinary amend to regulation (Source: *Bollettini Nitrati*).
Panel [B] presents the climate zones included in the sample for each window. Municipalities are aggregated according to six climate zones, as per current regulation on spreading bans, see Figure E.2. Four out of nine windows set equal restrictions for all climate zones, with seven windows set equal restrictions for more than 50% of the sample municipalities.

Figure 4: Livestock concentration in the Lombardy region



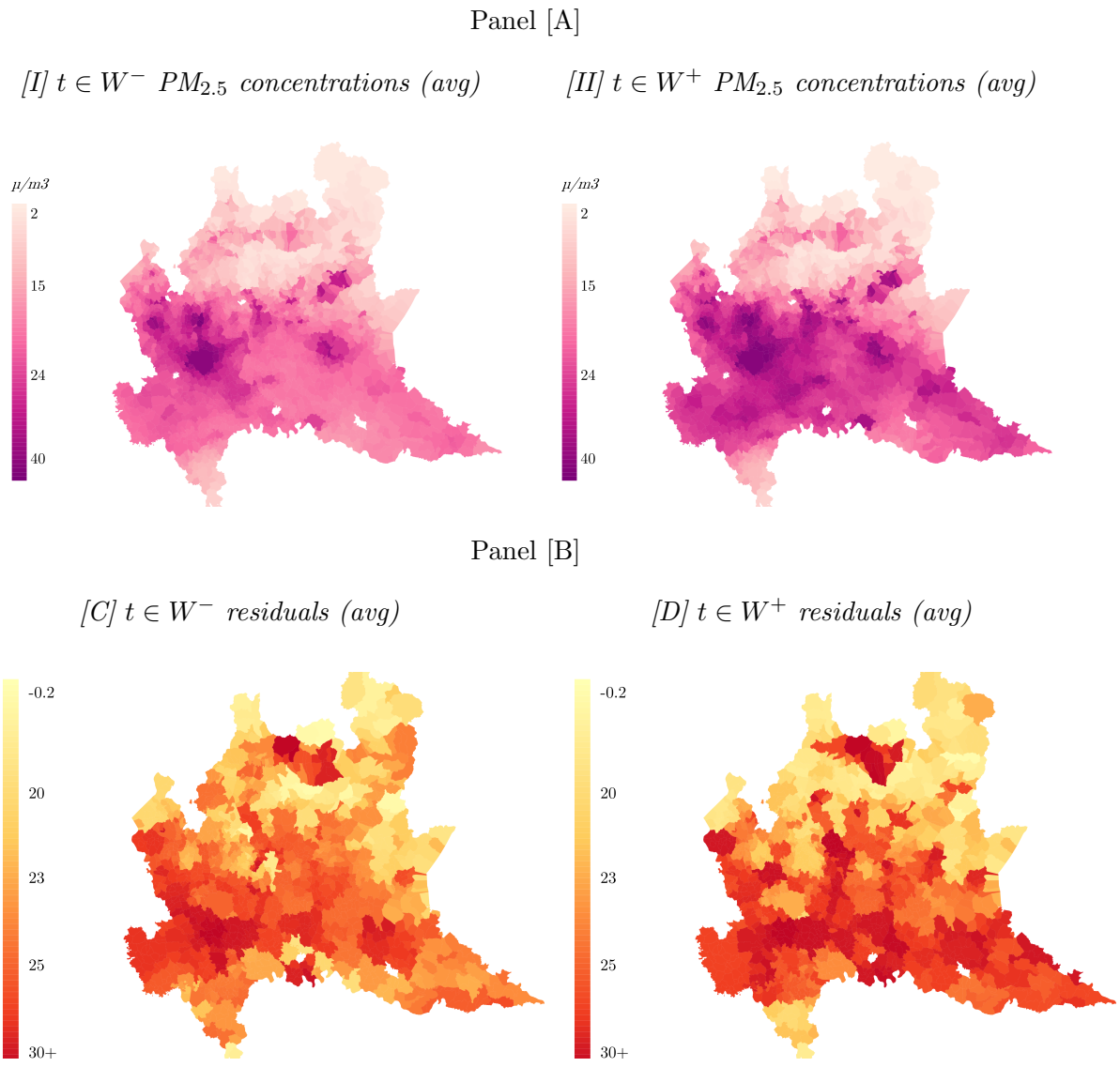
Notes: the figure shows the concentration of farming animals (chicken, cow and beef, sheep and goat, pigs) in the Lombardy region. In computing the total headcount, each animal is weighted by their expected daily production of manure. Weights are obtained from Hillel & Hatfield (2005) and are reported in Table 2.

Figure 5: Admissions by Hospital



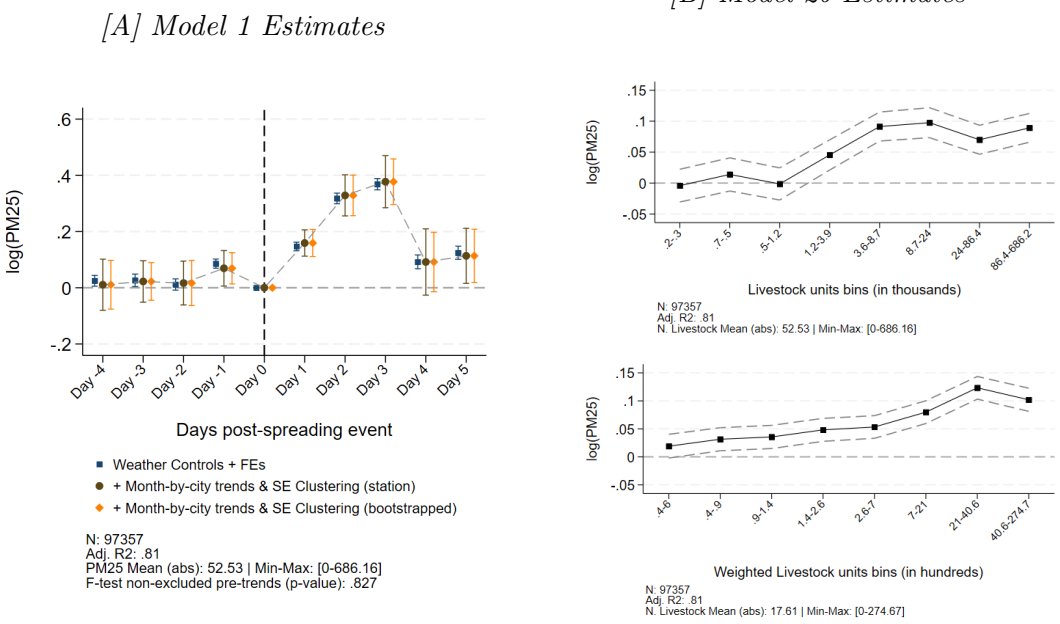
Notes: each buffered geopoint represents a hospital. The buffer radius is proportional to the yearly average number of daily hospital admissions across all departments: geopoints of similar size indicate hospitals of comparable overall capacity. The color heatmap represents the yearly average number of daily hospital admissions limited to R&CD: geopoints of similar color indicate hospitals with comparable reception capacity of R&CD patients.

Figure 6: $PM_{2.5}$ concentrations around spreading windows



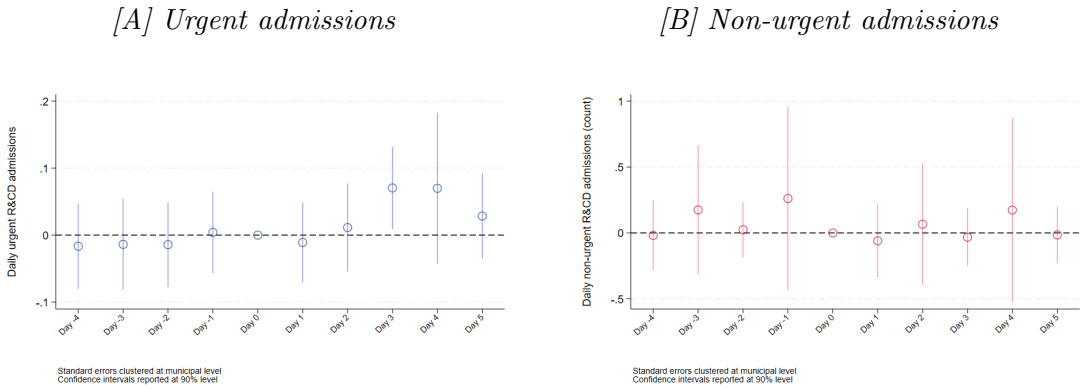
Notes: the figure depicts air quality conditions around spreading windows (W). In Panel [A], average $PM_{2.5}$ concentrations for days before ($t \in W^-$) and after ($t \in W^+$) the window allowing spreading opening are plotted at municipality level. When spreading is allowed, systematically higher levels of $PM_{2.5}$ appear in the Po Valley area, where the majority of farming activities is concentrated. In Panel [B], $PM_{2.5}$ levels in days before and after the window opening are regressed on a set of environmental covariates: humidity, radiance, rainfall, temperature, wind speed, wind direction, included singularly and interacted with each up to the third lag. Predicted residuals are then plotted at municipality level. When spreading is allowed, systematically higher residual variation appears in the Po Valley.

Figure 7: Effect of manure spreading on $PM_{2.5}$ concentrations (baseline)



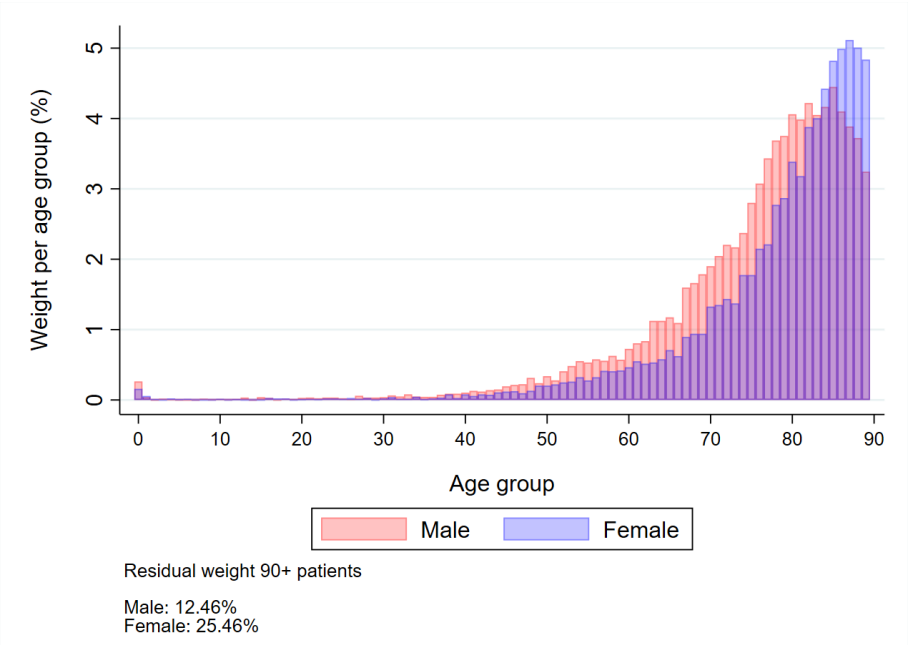
Notes: Panel [A] plots the effect of manure spreading windows opening on log $PM_{2.5}$ concentrations, depicting the estimates for η_k coefficients in Equation 1. Controls include weather conditions up to the third lag and interacted in each period, and FEs include municipality, month-by-year, day of the week, holiday. Bootstrapped standard errors are sampled with 100 iterations. Coefficients on last prohibition day (day 0) are set to zero. Confidence intervals are plotted at 95% level.
Panel [B] provides the estimates of the ρ_b coefficients in the static model 2b. Weighted livestock units at municipality level are computed using the weights in Table 2.

Figure 8: Effect of manure spreading on $PM_{2.5}$ concentrations (baseline)



Notes: the figure shows the effect of manure spreading windows opening on hospital admissions, plotting the estimates for η_k coefficients in Equation 1, where k is restricted to observations at least two days apart from the window opening cutoff. Controls include weather conditions up to the third lag and interacted in each period, and FEs include municipality, month-by-year, day of the week, holiday, month-by-province and window. Standard errors are plotted at municipality level. Confidence intervals plotted at 95% level.
Coefficients in Panel [A] and [B] refer to urgent and non-urgent R&CD hospital admissions respectively.

Figure 9: Computing total VSLY lost to manure spreading - Age group weights distribution



Notes: the figure reports the distribution of weights assigned to one-year age groups in the computation of total VSLY lost in response to manure spreading activities. The weights are calculated as the share of deaths at discharge per age group over the total during the sample period. Weights are employed to redistribute the estimated number total casualties across age groups, and derive a counterfactual simulation of years of life lost by individuals in Lombardy.

Online Appendix

A Supplementary Models

A.1 Regression-Discontinuity Design

In a typical RD design, it is observed an outcome $Y_i = Y_i(0) \times (1 - T_i) + Y_i(1) \times T_i$, where $Y_i(0)$, $Y_i(1)$ represent the outcome in the absence and in the presence of treatment respectively, and T_i is a determinant of treatment status. Treatment is determined by a running variable X_i influencing the treatment status whenever X_i exceeds a cutoff c, such that $T_i = \mathbb{1}(X_i \geq \bar{x})$. My setting can be considered a regression discontinuity in time (Hausman & Rapson 2018), so the unit is considered treated (spreading allowed) past the date of prohibition lift ($t_i > c$). The parameter of interest is identified at the cutoff as:

$$\tau = \tau(c) = \mathbb{E}\left[Y_i(1) - Y_i(0) | t_i = c, \mathbf{Z_i}\right]$$

where $\mathbf{Z_i}$ is a set of covariates. For the derivation of the RD estimator $\tilde{\tau}$, I refer to Calonico et al. (2017).

In this framework, estimating the effect of manure spreading using an RD design could pose several challenges, given the volatile and repeated nature of prohibitions, which give rise to multiple discontinuities within a short time span. Yet, the process of neighbor selection implied in the RD estimation can provide a useful benchmark to capture the persistence of the effect of a spreading event through time. For bandwidth selection, it is employed a mean squared error (MSE) optimal criterion, obtained by fitting a curve over a sub-portion of the support of the data, subsequently minimizing errors (Imbens & Kalyanaraman 2011). Let h be the bandwidth such that $\tilde{\tau}$ is estimated within the interval $t_i \in [c - h, c + h]$, it can be shown that the typical asymptotic MSE expansion is such that:

$$\text{MSE}\left(\tilde{\tau}(h)\right) \approx h^{2p+2} B + \frac{1}{nh} V$$

where B and V denote the squared bias and the variance of $\tilde{\tau}$, and p is the polynomial order of fit. As such, optimal bandwidth choice can be derived as

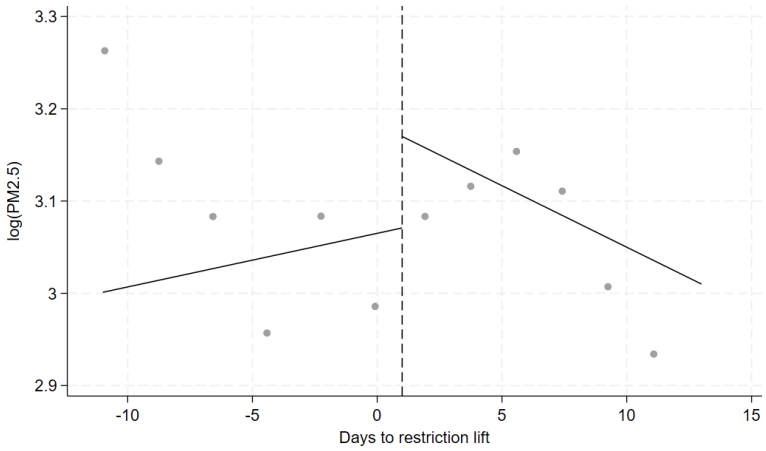
$$h_{mse} = \left\{ \frac{\hat{V}_j / n}{2(1+p)\hat{B}_j} \right\}^{\frac{1}{3+2p}}$$

an alternative version of this estimate allows for the bandwidth to vary at the right ($h_{mse,r}$) and at the left ($h_{mse,l}$) of the cutoff, allowing for side-specific $\hat{B}$ and $\hat{V}$ estimated parameters.

By defining a running variable as a time counter resetting every time a window is opened (or closed), bandwidth selection can be employed to investigate in the data the persistency of a perturbation induced by a prolonged policy change, in the form of lifted (or reintroduced) spreading prohibitions. I estimate the quantities $h_{mse,r}$ and $h_{mse,l}$ using the package `rdrobust` in Stata version 18.0. I consider a first-order polynomial ($p = 1$)

and control for the same structure of fixed effects explained in Section 3.2, as well as weather covariates. I focus only on observations between November and March, to align with the period of spreading restrictions.

Figure A.1: Spreading-induced discontinuity in PM levels



Notes: the figure plots binned sample means for outcome variable $\log(PM_{2.5})$ conditional on a running variable computed as relative distance to the day of spreading restrictions being lifted (set to 1), weather controls and fixed effects. Bins including two days around the threshold. The domain is limited to 12 days around the cutoff for visualization convenience. Solid lines represent the conditional mean functions computed with a polynomial of degree 1.

Table A.1: Estimated RD treatment effect and bandwidth span

	(1)	(2)	(3)	(4)
	log(PM2.5)	log(PM2.5)	log(PM2.5)	log(PM2.5)
RD Estimate	0.0779***	0.0580***	0.0836***	0.0538***
	(0.0146)	(0.0183)	(0.0179)	(0.0144)
Observations	1,225,279	1,156,051	1,084,346	1,002,643
Weather controls	N	N	Y	Y
Fixed effects	Y	Y	Y	Y
Bandwidth (L)	5.174	4.262	5.132	4.173
Bandwidth (R)	5.174	4.643	5.132	4.9879
BW bias (L)	12.19	12.23	12.02	12.56
BW bias (R)	12.19	12.86	12.02	12.99
BW selector	One	Two	One	Two

Notes: the table reports the estimated RD coefficient $\tilde{\tau}$, together with left- and right-side bandwidth, and bias estimator. In Columns (1) and (3), one common MSE-optimal bandwidth selector for the RD treatment effect estimator is allowed, whereas two separate selectors are specified in Column (2) and (4). Weather controls include temperature, wind direction, wind speed, rainfall, radiance, humidity, and average planetary boundary layer height, up to three lags. Fixed effects include municipality, month-by-year, day of the week, and holiday. Standard errors clustered at PM sensor level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

As previously noted, the RD estimator implies a lower effect of spreading on $PM_{2.5}$ concentrations at the cutoff (roughly 25% of the effect estimated through Equation 2). This is explained by considering that in this framework, it is relaxed the requirement of windows exhibiting consecutive days of prohibitions alternated to consecutive days of ban lift. This, in turn, casts part of the effect on either side of the threshold, resulting in a much lower point estimate. Importantly, the estimated bandwidth for discontinuity around the cutoff oscillates between a value of 5.17 and 4.26, which drives my choice of adopting $T = \frac{\text{card}(W)}{2} = 5$ in Equation 1.

A.2 Difference-in-Difference Estimator

I propose here an alternative specification that exploits short-lived discrepancies in spreading prohibitions across climate zones. Differently from Models 1 and 2, the goal is to maintain a set of never-treated observations to construct a more classic double-difference estimator. To obtain a reasonable counterfactual to PM concentrations in the absence of spreading, I focus again on short time windows, this time of five days in total. Compared with my preferred strategy, the reduced length is mostly forced by data availability. Instances where spreading is allowed in only part of the region are relatively rare and often involve only a few days of fragmented prohibitions. I consider as a control group municipalities where spreading is not allowed throughout all five days. I then define three levels of spreading-induced treatment, depending on whether spreading is allowed on the fifth day only (level one), on the fourth and fifth day (level two), or starting from the third day onward (level three).[†] Table A.2 lists the days suitable for this strategy and identifies the treated climate zones.

Table A.2: Inclusion of climate zones in DiD strategy

	Day-municipality observations									
	Level three		Level two				Level one			
	Control	Treat	Control	Treat	Control	Treat	Control	Treat	Control	Treat
Alps	385	0	0	385	0	385	0	385	0	385
Central plain	1485	0	0	1485	1485	0	1485	0	1485	0
West plain	0	1900	1900	0	0	1900	0	0	0	1900
East plain	320	0	0	0	320	0	320	0	320	0
West Prealps	0	2135	2135	0	0	2135	0	0	0	2135
East Prealps	1315	0	0	1315	1315	0	0	1315	1315	0
Start date	5-Nov-16		8-Feb-17		31-Jan-18		6-Nov-16		30-Jan-18	

Notes: the table reports the number of day-municipality observations identified as treatment and control sample in each event, by climate zone. It is reported start date of the five-day period over which the DiD model is estimated.

I then estimate the following difference-in-difference (DiD) model:

[†]Differently from my main specification, I do not allow a window to be defined backward in time. This prevents confusion in the definition of pre-treatment periods and allows for an easier interpretation of time relative to the relaxation of prohibitions.

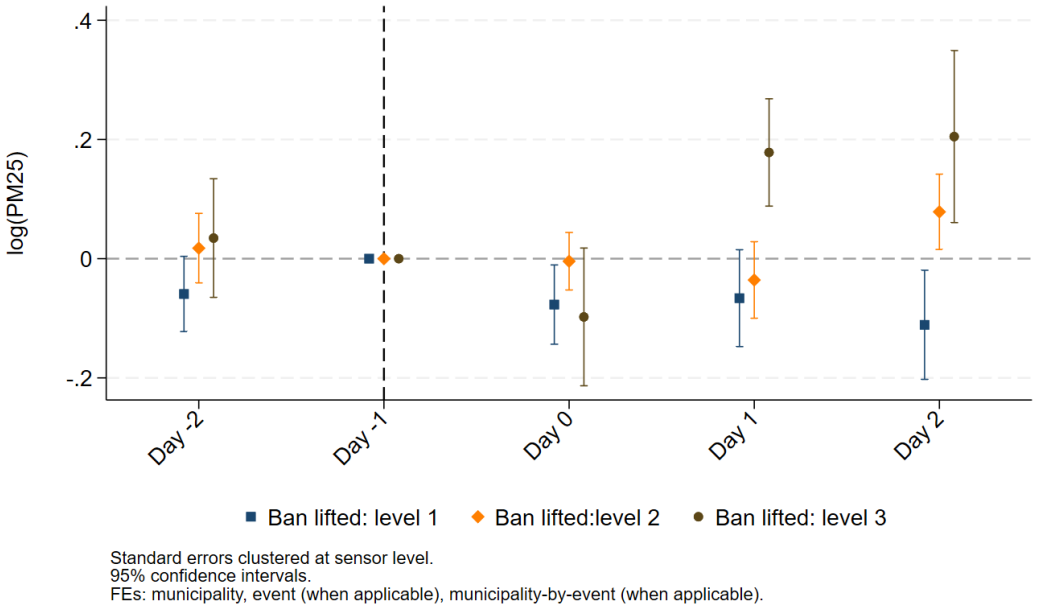
Model 1b:

$$Y_{mt} = \lambda_0 Treat_m + \sum_{k=-2, k\neq -1}^{k=3} \lambda_{1k} \mathbb{1}\{H_{mt} = k\} + \sum_{k=-2, k\neq -1}^{k=3} \lambda_{2k} \mathbb{1}\{H_{mt} = k\} \times Treat_m + \\ + X'_{mt}\Gamma + \alpha_m + \alpha_w + \alpha_{wm} + \varepsilon_{mt} \tag{A1}$$

where indices i and h are omitted for conciseness as the model is used uniquely to study concentrations of pollutants. The variable $Treat$ indicates whether spreading is either partially or fully allowed in the municipality starting from the third day. Other variables are defined as in Models 1 and 2, with the exception of fixed effects, for which the reduced number of days considered imposes a different structure to avoid perfect separation. I include municipality, window (when more than one event is present), and municipality-by-window. The DiD coefficients of interest λ_{2k} capture the difference in PM concentrations in climate zones where spreading is allowed compared to the rest of the region. Importantly, one must notice the possibility of treatment spillover induced by the transportation of pollutants. Especially in municipalities bordering climate zones, the effect of spreading in a neighboring area is expected to impact the concentrations of airborne pollutants, albeit to a reduced extent. As such, looking only at DiD coefficients is likely underestimating the true impact of spreading on air quality. Yet, by re-introducing never-treated units in the framework, this strategy reinforces the conclusions presented in the paper.

In the sample I construct, for some municipalities, it is possible to define multiple levels of treatment, depending on whether the start of the observation period is set on a day or the day immediately after. To avoid overlapping definitions of the treatment, I estimate separately Equation A1 for each level of treatment. I then combine the estimated values for λ_{2k} into the same graph (Figure A.2).

Figure A.2: Effect of manure application on $PM_{2.5}$ concentrations - DiD estimation



Notes: the figure plots the estimated for λ_{2k} coefficients in Equation A1. Controls include weather conditions up to the third lag and interacted in each period, and FEs include municipality, event (when applicable), municipality-by-event (when applicable). Standard errors are clustered at sensor level. Coefficients on last prohibition day (day -1) are set to zero. Confidence intervals are plotted at 95% level. It is reported for each day the coefficient of three levels of treatment: restrictions lifted on day five (level one), restrictions lifted on day four (level two), restrictions fully lifted (level three).

The results exhibit dynamics similar to what already obtained through the main specification in Section 5. Focusing first on level three treatment, it is estimated a sizeable increase in $PM_{2.5}$ concentrations in municipalities with early lift in spreading prohibitions starting from the second day. The maximum is observed on day three, with estimated PM concentrations around 20% higher in treated municipalities. As previously argued, the effect is lower than the one reported in Figure 7, likely due to the presence of spillover in the form of circulation of pollutants emitted in treated climate zones being air transported to control zones. The results for municipalities with other degrees of treatment point toward a similar scenario. For treatment of level two, the effect of spreading emerges, even though with a smaller magnitude, on the last day of the selected time window, when spreading prohibitions have been lifted for two consecutive days in treated municipalities. Since the spike in PM concentrations is expected to emerge with a one-day lag, for treatment of level one it is estimated no increase in treated municipalities, with even a slight decrease on the last day of the window. By considering a different combination of days and municipalities compared to the sample included in the estimation of Model 1, these results, paired with the rest of the evidence presented in this paper, strongly support the presence of a true effect of spreading on $PM_{2.5}$ pollution.

B Variables Description

B.1 Procedure Reimbursement and Cost of Hospitalisation

Similarly to other countries, the Italian National Healthcare System is a prospective payment system (PPSs). Reimbursement for each hospitalisation is based on predetermined, fixed amounts, based on Diagnosis Related Group (DRG) Codes. Being the Italian system subject to considerable decentralisation, tariff lists are determined at regional level. In the case of Lombardy, these were lastly updated in 2015. The table exemplifies the nature of tariff lists.

DRG	MDC	Description	OrdTariff	DHTariff	Threshold	OverTariff
078	04	Pulmonary Embolism	4,466	214	33	155

where columns left to right report DRG code and description, Major Diagnostic Category (MDC), the corresponding daily tariffs for respectively an ordinary hospitalisation lasting more than one day, and a day-hospital regime. In addition, it is set a threshold value for each hospitalisation, after which a second tariff regime (usually with reduced amounts) is applied. In addition to threshold value specific to each DRG, an hospital admission is considered long-term care after 60 days, to which it is applied a fixed reimbursement of 154 euros per day.

As my dataset allows to retrieve the length of the hospital stay and the DRG codes of all procedures undergone by a patient, my measure of cost of hospital stay is calculated according to the following formula:

$$\begin{aligned} Cost = & \mathbb{1}\{t = 1\} \times \text{DHTariff} + \mathbb{1}\{t \in (1, \text{Threshold}]\} \times \text{OrdTariff} + \\ & \mathbb{1}\{t \in (\text{Threshold}, 60]\} \times (t - \text{Threshold}) \times \text{OverTariff} + \\ & \mathbb{1}\{t > 60\} \times (t - 60) \times 154 \end{aligned} \tag{A2}$$

C Manure Spreading Regulation in Italy

Following the Council Directive 91/676/EEC of 1991 concerning the protection of waters against pollution caused by nitrates from agricultural sources, Italy has started regulating the application and storage of manure and slurry. In 2006, the regulatory authority was decentralized from the central to the regional governments (Ministerial Decree 7 April 2006), *de facto* initiating the stream of reforms in the governance of pollutants from farming activities. The table below summarizes the evolution of the regulatory framework in Lombardy since then.

Period	Length prohibitions	Start	End	Scheme	Bullettin	PM attention
2007-2011	90 days	1st Dec (fixed)	28th Feb	Continuous	No	No
2012-2015	90 days to 120*	1st Nov (variable)	28th Feb	Continuous	No	No
2016-2020	**90 days (62 continuous)**	**1st Nov (1st Dec)**	**28th Feb (31st Jan)**	**Mixed**	**Yes**	**No**
2021-Present	90 days (32 continuous)	1st Nov (15th Dec)	28th Feb (15th Jan)	Mixed	Yes	Yes**

* Restrictions varied depending on the animal source (e.g. poultry) and nature of animal waste (e.g. liquid slurry).
** Attention to PM is limited to additional restrictions at municipal level, imposed when PM control measures (e.g traffic restrictions) are implemented by a municipality.

D Flexible Spreading Scheme Simulation: Dynamic Optimization

I hereby present in detail the underlying theory and the most salient results of the intermediate steps involved in the dynamic optimization process.

D.1 Spatio-temporal models of PM concentrations

Spatio-temporal models have been systematically employed in environmental studies related to airborne pollutants, given the geographically referenced and temporally correlated nature of pollution data. In this paper, I focus on the class known as Bayesian spatio-temporal models[†]. Bayesian modeling firstly assumes the independent variable to be described as

$$Y(s_i,t) = \mu(s_i,t) + e(s_i,t), \quad i = 1,\dots,n \quad t = 1,\dots,T$$

where s_i and t identify site and time, $\mu(s_i,t)$ is a space-time process modelled as a regression model of the form

$$\mu(s_i,t) = \mathbf{X}'(s_i,t)\beta(s_i,t)$$

and $e(s_i,t)$ is a zero-mean space-time process. The model is flexible in accommodating spatio-temporally varying regression coefficients ($\beta(s_i,t)$). It is common to model this overall error term as the some of two independent zero-mean process, a random process $\omega(s_i,t)$ and a pure error term $\epsilon(s_i,t)$, assumed to be independent for all spatial locations at all time points, i.e.

$$Y(s_i,t) = \mathbf{X}'(s_i,t)\beta(s_i,t) + \omega(s_i,t) + \epsilon(s_i,t), \quad i = 1,\dots,n \quad t = 1,\dots,T$$

Starting from this general formulation, I test the performance of multiple model specifications. Firstly, I consider a model which assumes a temporally independent GP for $\omega(s_i,t)$ and time independent and spatially varying regression coefficients such that

$$\mathbf{X}'(s_i,t)\beta(s_i,t) = \sum_{j=1}^{p} x_j(s_i,t)\beta_j(s_i,t) \qquad \text{(Dynamic GP)}$$

and

$$\beta_j(s_i,t) = \beta_{j0} + \beta_j(s_i)$$

where each $\beta_j(s)$ is assumed to follow an independent GP.

Secondly, I consider an autoregressive model (AR) that reintroduces times series dependence. The model specifies autoregression on a centered random effect O_t instead of

[†] Banerjee et al. (2003) and Sahu (2022) explain thoroughly the nature and properties of this class of models.

directly on the data, and limits autoregression to a single period. The model can be formulated as

$$\begin{aligned} Y(s_i,t) =& O(s_i,t) + \epsilon(s_i,t) \\ O(s_i,t) =& \rho O(s_i,t-1) + \mathbf{X}'(s_i,t)\beta + \omega(s_i,t) \end{aligned} \tag{AR}$$

Part of the reason for imposing independence of one process is the dimensionality implied in fitting the model, especially when the number of locations increases. As a third model, I implement a model of Gaussian Predictive Process (GPP) proposed by Sahu & Bakar (2012), which defines random effects at a smaller number of locations (knots), and assumes an AR model only for the random effects at the knot locations. The model is specified as follows:

$$\begin{aligned} \mathbf{Y}_t =& \mathbf{X}_t\beta + A\mathbf{O}_t + \epsilon_t \\ \mathbf{O}_t =& \rho\mathbf{O}_{t-1} + \mathbf{w}_t \end{aligned} \tag{GPP}$$

where

$$A = CS_{\omega^*}^{-1}$$

with C the cross-correlation matrix between random effects at locations $(s_1,\ldots,s_n)$ and knots $(s_1^*,\ldots,s_m^*)$ and S_{ω^*} the correlation matrix of the random effects $\mathbf{o}_t$, depending on the distance between sites.

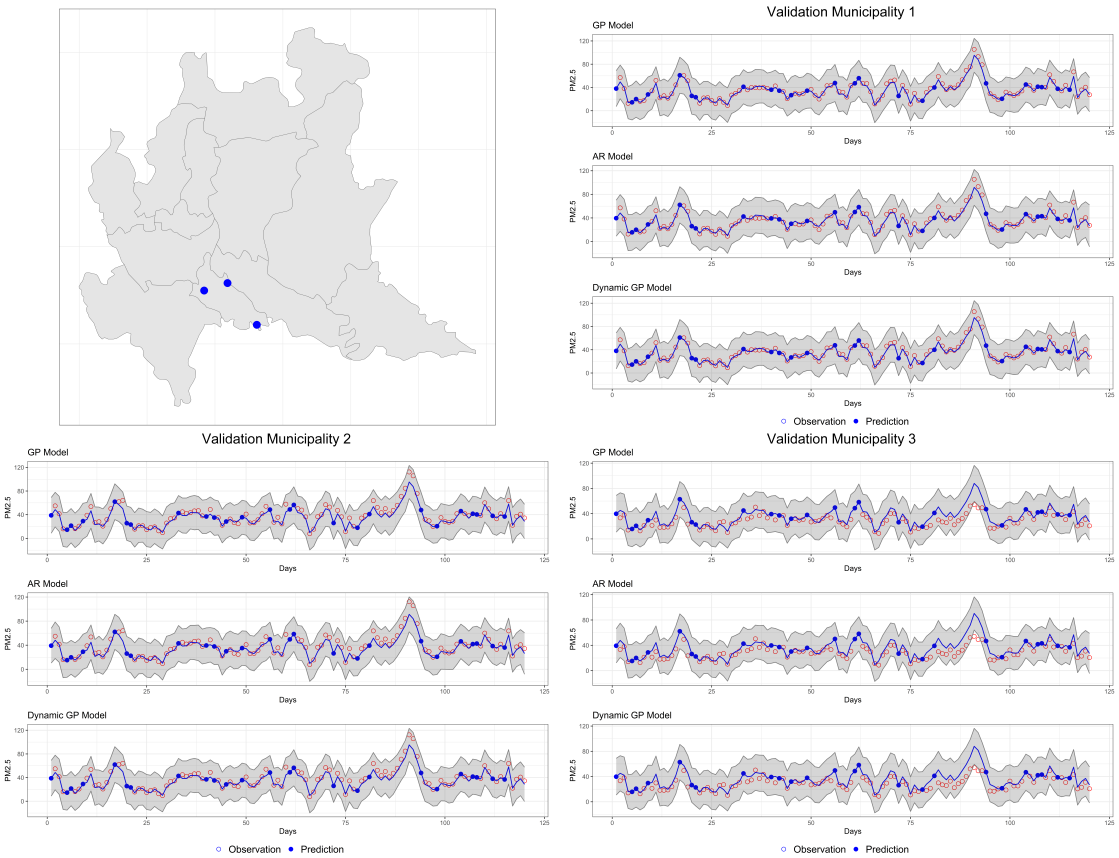
I fit the models separately for each province and use them to forecast PM levels at municipal level and winter months between 2016 and the beginning of 2020. I use as covariates all weather conditions available (temperature, rainfall, radiance, humidity, wind direction and wind speed). All models are implemented `spTimer` in R. I report here the results for the Lodi province as an example. Results for other provinces and values of PM predicted concentrations can be found in the replication code of this paper. To compare the models, I look in Table A.3 primarily the Posterior Predictive Model Choice Criteria (PMCC), which is the sum of a model's goodness-of-fit (GOF) and penalty terms. According to PMCC, the best model must balance goodness-of-fit and predictive penalty, minimizing the sum of the two. I compare all models with a simple linear alternative as benchmark. In this case, AR strikes as the most suitable model, but all three performs much better than the linear model. Secondly, I compare statistics of model validation through out of sampling predictions. This is done by randomly selecting three municipalities and using half of the observations as a training set to predict the remaining half through the estimated parameters of the models. The results are illustrated in Figure A.3. Visually, all models display very high predictive power. To confirm this intuition, I report commonly used validation criteria, which indeed show little discrepancy between the three models. Given that models score comparatively in validation tests, I opt for the GPP which also minimizes computational burden having to operate with large matrices.

Table A.3: Selection criteria and validations statistics for PM spatio-temporal models

Model	GOF	Penalty	PMCC	RMSE	MAE	CRPS	CVG
Linear	1445979	1453800	2899779	14.7435	11.0604	7.862227	94.44
GPP	1930.16	19820.52	21750.68	7.090132	5.315292	18.50795	100
AR	1928.81	7396.39	9325.2	7.067555	5.200269	7.903886	100
Dyn GP	1727.86	8506.37	10234.23	7.299482	5.519052	8.722338	100

The table reports model selection and validation criteria for three different spatio-temporal models (Gaussian Predictive Process, Autoregressive, Dynamic Gaussian Process) compared to a simple linear alternative. For selection, it is reported the Posterior Predictive Model Choice Criteria (PMCC), obtained as the sum between goodness-of-fit (GOF) and penalty. For validation, Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Continuous Ranked Probability Score (CRPS), and coverage (CVG) are reported. Best-fitting models are expected to minimize all selection and validation criteria while maximizing coverage.

Figure A.3: Validation of spatio-temporal models at three random locations



The figure plots observation predictions and 95% intervals for three municipalities (top-left corner) and three spatio-temporal models (GPP, AR, Dynamic GP). Fitted values are plotted with a solid line. The (true) observations in the training set are plotted as open circles and the (true) observations set aside for validation are plotted as filled circles.

Note that Figure A.3 does not inform on the ability of spatio-temporal models to

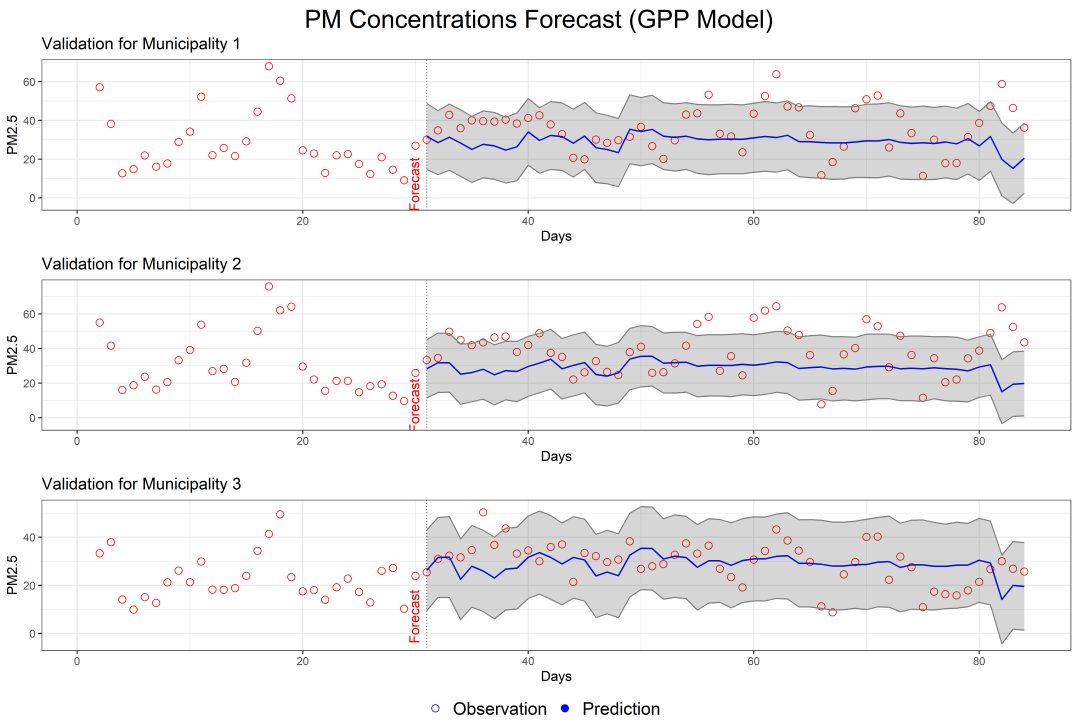
forecast PM concentrations, which constitutes a far more demanding task. This validation technique considers concentrations in neighbouring municipalities as known quantities, and thus exploits tight geospatial correlation in PM distribution to achieve close-to-perfect fit. However, this does not reflect the decision-making process of a regulator who needs to lift spreading prohibitions using incomplete information about future PM levels. Hence, I use the GPP model to forecast PM concentrations three days into the future.

Forecasting implies estimating random effects for each location into the future. Let an unobserved location s_0. The k-steps ahead forecasting at such location is performed by assuming independently normally distributed errors ($\omega_t \sim N(\mathbf{0}, \sigma_\omega S_{\omega*})$) and random effects ($\mathbf{O}_0 \sim N(\mathbf{0}, \sigma_w^2 S_0)$) where S_0 is a correlation matrix obtained via the exponential correlation function with a decay parameter ϕ_0. The predictive distribution of $Y(s_0, T+k)$ is obtained by simply advancing the top-level model equation by k periods:

$$Y\left(s_0, T+k\right) = \mathbf{x}'\left(s_0, T+k\right)\beta + \mathbf{a}_0' S_{\omega*}^{-1} \mathbf{O}_{T+k} + \epsilon(s_0, T+k)$$

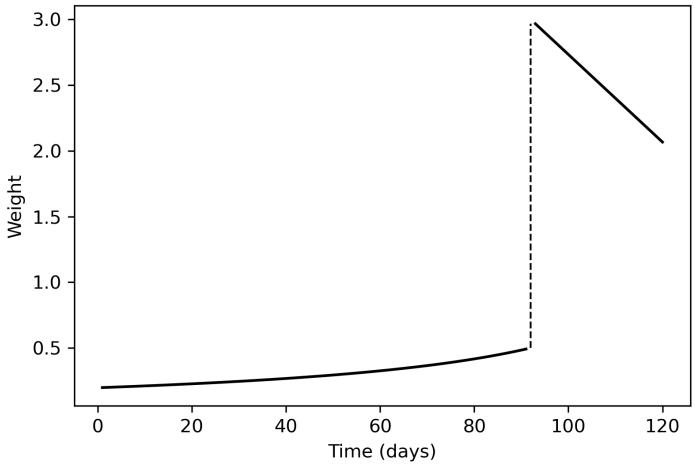
where $\mathbf{a}_0'$ is a vector of covariance functions that depend on the distance between s_0 and site s_j. The forecast value for $Y\left(s_0, T+k\right)$ is iteratively drawn from the obtained distribution via Markov chain Monte Carlo sampling. The approach can be then generalized for each s_n. The forecasting is performed iteratively: the model is estimated using the 30 days prior to the starting date of forecasting. Then, PM levels are observed and the true values are added to the training set before the model is estimated again and forecasts are computed for the following three days. The results for the same three validation municipalities are reported in Figure A.4.

Figure A.4: Forecast PM levels at three random locations



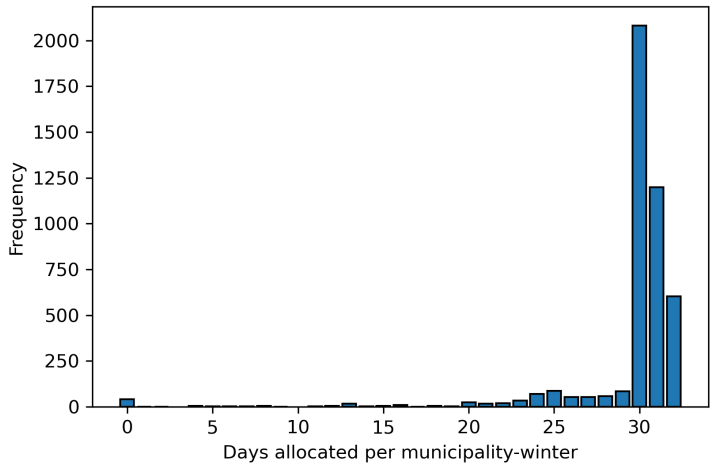
D.2 Dynamic Optimization Simulation: Additional Results

Figure A.5: Dynamic optimization algorithm - weighting



Notes: the figure plots the weighting function ω under the assumption of having still to allocate 30 days over the entire length T. The discontinuity is set such that, as the algorithm reaches 30 days left till the end of T with still 30 days to allocate, a strong penalty is imposed for not selecting the following days In turn, this forces the algorithm to allocate all the following days to spreading as long as weather constraints are met.

Figure A.6: Dynamic optimization algorithm - distribution of allocated days



Notes: the figure plots the distribution of spreading days allocated by the dynamic algorithm for each municipality-season combination.

D Appendix Tables

Table D.1: Placebo test - Effect of spreading prohibitions on other diseases

	Pooled		Digestive		Genitourinary		Muscoskeletal	
	All patients	Urgent	All patients	Urgent	All patients	Urgent	All patients	Urgent
Admission	0.184	-0.187	0.178	-0.180*	-0.280	-0.162	-0.391	-0.280
(linear)	(0.461)	(0.140)	(0.531)	(0.105)	(0.236)	(0.159)	(0.259)	(0.190)
Obs	69982	47144	29115	20699	25278	18917	21081	11723
R^2	0.93	0.97	0.92	0.97	0.92	0.97	0.94	0.97
Admissions	-0.012	-0.016**	-0.009	-0.012	-0.011	-0.017**	-0.016*	-0.024*
(Poisson)	(0.009)	(0.008)	(0.009)	(0.007)	(0.012)	(0.008)	(0.009)	(0.013)
Obs	69982	69733	29115	28859	25278	25131	21081	19686
R^2								
Mortality	-0.001	0.000	0.002	0.010	0.007	0.019	0.001	0.088
	(0.003)	(0.009)	(0.006)	(0.018)	(0.005)	(0.011)	(0.003)	(0.120)
Obs	25928	8298	10835	3816	9975	4646	6770	339
R^2	0.12	0.20	0.20	0.32	0.18	0.27	0.28	0.96
Length	-0.151	0.137	0.087	0.883*	-0.111	0.167	0.059	-2.777
	(0.123)	(0.295)	(0.232)	(0.479)	(0.251)	(0.437)	(0.199)	(1.758)
Obs	28239	8787	11303	4121	10204	4779	8619	506
R^2	0.33	0.27	0.40	0.39	0.30	0.34	0.58	0.87
Cost-to-procedure	-284.311*	225.257	120.116	1,759.920	-621.581**	-458.110	-198.179	763.017
	(168.007)	(584.765)	(369.545)	(1,392.755)	(289.127)	(623.723)	(154.683)	(1,526.433)
Obs	25820	7472	10183	3417	9276	4107	7840	374
R^2	0.08	0.22	0.15	0.32	0.19	0.32	0.26	0.93
Total cost	-605.278	1,668.843	707.800	6,653.459	-1,634.978	206.293	-333.029	-3,412.881
	(784.199)	(2,751.575)	(1,852.196)	(6,658.251)	(1,076.116)	(2,145.908)	(448.733)	(5,401.949)
Obs	28193	8782	11264	4119	10189	4779	8591	503
R^2	0.07	0.17	0.14	0.29	0.18	0.30	0.33	0.85
Severity (weight)	-0.011	0.018	0.014	0.113***	-0.040	-0.029	0.012	-0.144
	(0.014)	(0.036)	(0.020)	(0.043)	(0.029)	(0.054)	(0.020)	(0.155)
Obs	28242	8787	11305	4121	10205	4779	8619	506
R^2	0.19	0.34	0.26	0.47	0.23	0.39	0.47	0.89

Notes: the table reports the estimated impact of a spreading window on daily urgent admissions, mortality, length of stay, cost-to-procedure, total cost of stay, and relative severity weight in three class of diseases (digestive, genitourinary, muskoskeletal) observed in the five days before and after an opening event. Weather controls include temperature, wind direction, wind speed, rainfall, radiance and humidity, interacted with each other at up to the third lag. When estimated through a linear model, only strictly positive observations for hospital admissions are included. Observations with no admissions when the model is estimated through fixed-effects pseudo-Poisson ML.
Standard errors, clustered at municipal level, are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

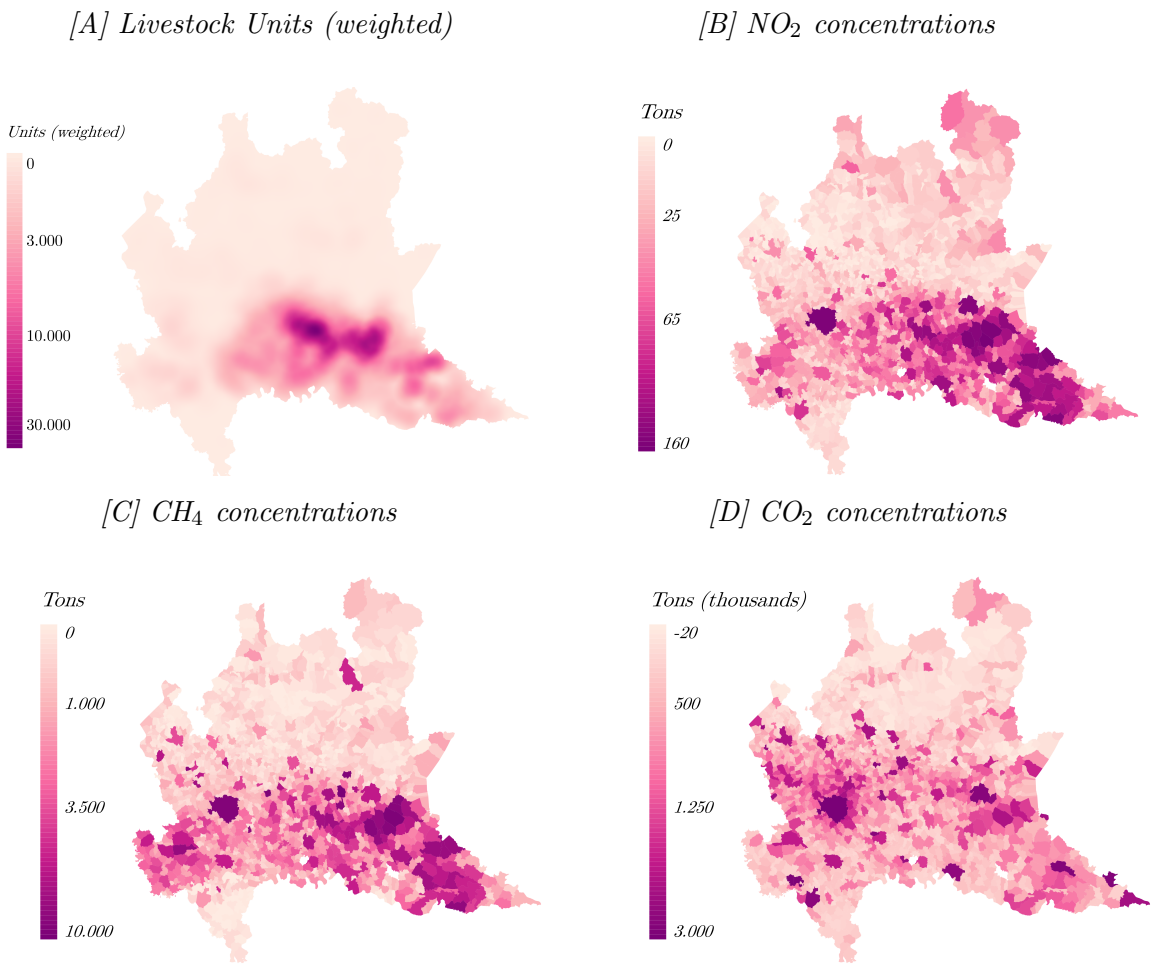
Table D.2: Effect of spreading windows on $PM_{2.5}$ concentrations (untransformed)

	Static Model			Semi-dynamic model		
$PM_{2.5}$	(1)	(2)	(3)	(6)	(7)	(8)
Window	7.297***	7.299***	7.243***			
	(0.095)	(0.091)	(0.416)			
Day 1				6.422***	6.411***	6.417***
				(0.131)	(0.122)	(0.436)
Day 2				10.030***	10.088***	10.013***
				(0.078)	(0.075)	(0.421)
Day 3				10.487***	10.378***	10.264***
				(0.063)	(0.061)	(0.372)
Day 4				3.898***	3.589***	3.473***
				(0.042)	(0.040)	(0.314)
Day 5				0.865***	0.205***	0.137
				(0.062)	(0.057)	(0.084)
Obs	112213	112207	105932	112213	112207	105932
Adj. R^2	0.739401289	0.778952049	0.775829476	0.749539	0.790209	0.787153
Weather (extended)	✓	✓	✓	✓	✓	✓
Municipality FE	✓	✓	✓	✓	✓	✓
Municipality-by-month FE		✓	✓		✓	✓
Month-by-year FE	✓	✓	✓	✓	✓	✓
DoW & Holiday FE		✓	✓		✓	✓
Standard errors	Rob	Rob	Clust	Rob	Rob	Clust

Notes: the table reports the estimates η_0 from Equation 2 (Columns 1 to 3) and the estimates η_k coefficients from Equation 1, where pre-trend coefficients are set to zero. Weather controls include temperature, wind direction, wind speed, rainfall, radiance, humidity, and average planetary boundary layer height, interacted with each other up to three lags. Robust (Columns 1-2 and 4-5) and clustered at sensor level (Columns 3 and 6) standard errors, are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

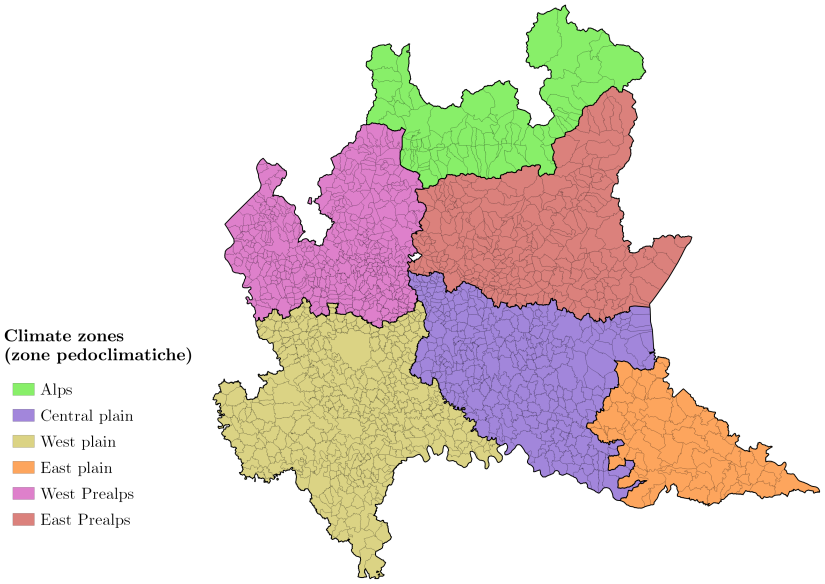
E Supplementary Figures

Figure E.1: Spatial correlation betwwn livestock presence and GHGs concentrations in Lombardy



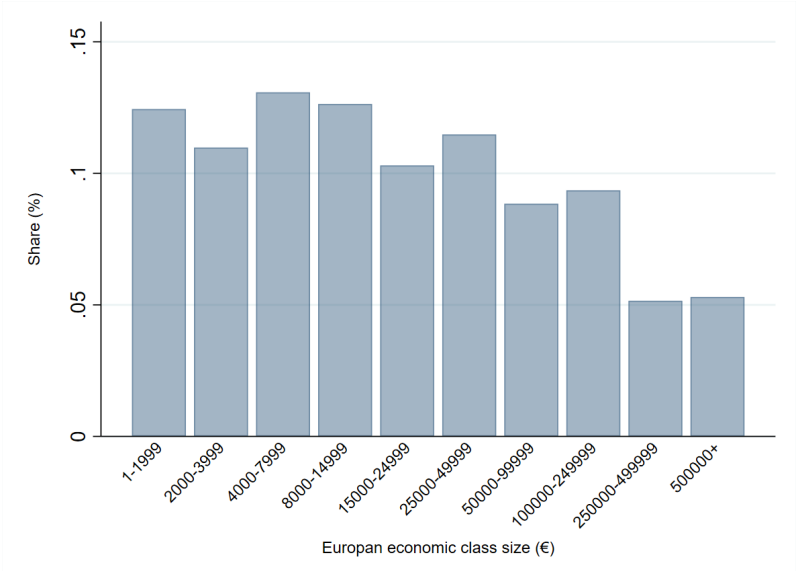
Notes: the figure shows high correlation between NO_2 and CH_4 concentrations and livestock presence in the region. Livestock units represent the weighted average number of units between 2016 and 2019; weighting follows the strategy explained in Section 4. Concentrations of pollutants are expressed as the yearly average (tons) in year 2017 (Source: INEMAR). Livestock units Spearman indeces: 0.8541 (NO_2); 0.5538 (CH_4); 0.0261 (CO_2).

Figure E.2: Lombardy region divided into manure application climate zones



Notes: the figure shows climate zones borders for the purpose of manure application prohibitions. Climate zone areas are: *Alps* (provinces: SO); *Central plain* (provinces: BG, BS, CR); *West plain* (provinces: LO, MI, PV); *East plain* (provinces: MN); *West Prealps* (provinces: BG, CO, LC, MB); *East Prealps* (provinces: BG, BS).

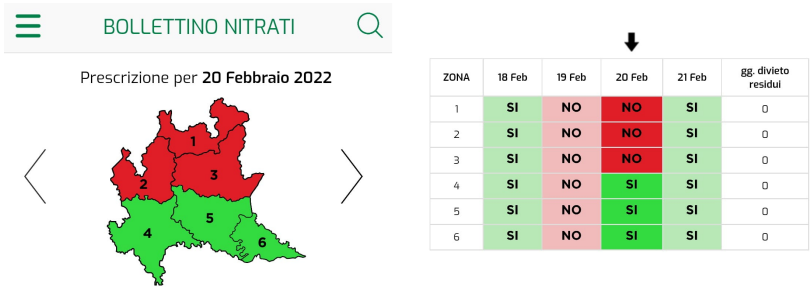
Figure E.3: European economic size class of farms in Lombardy



Notes: The figure plots the share of farms in Lombardy over economic size class. The measure of economic size is defined at European level (Commission Regulation (EC) No 1242/2008). Source: 2010 Agricultural Census, ISTAT.

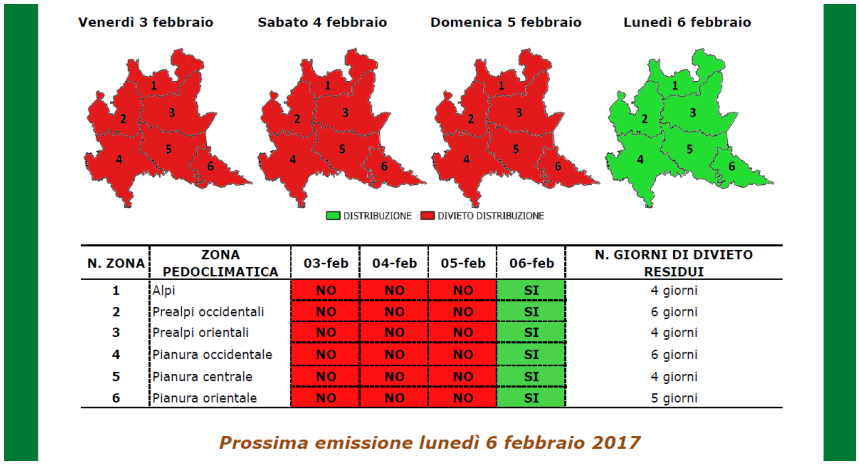
Figure E.4: Spreading bullettins layout

Panel [A] - App version



ZONA	18 Feb	19 Feb	20 Feb	21 Feb	gg. divieto residui
1	SI	NO	NO	SI	0
2	SI	NO	NO	SI	0
3	SI	NO	NO	SI	0
4	SI	NO	SI	SI	0
5	SI	NO	SI	SI	0
6	SI	NO	SI	SI	0

Panel [B] - PDF version



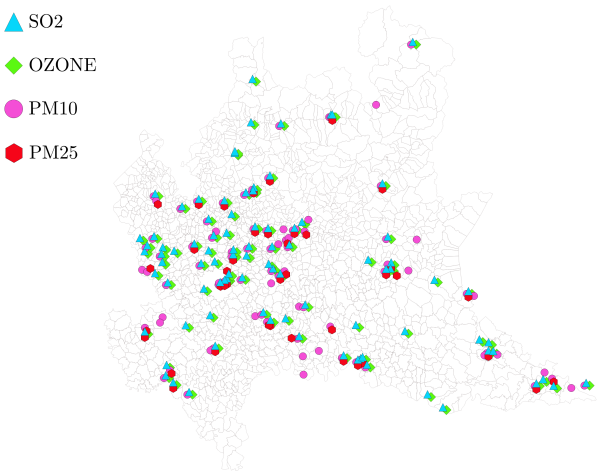
N. ZONA	ZONA PEDOCLIMATICA	03-feb	04-feb	05-feb	06-feb	N. GIORNI DI DIVIETO RESIDUI
1	Alpi	NO	NO	NO	SI	4 giorni
2	Prealpi occidentali	NO	NO	NO	SI	6 giorni
3	Prealpi orientali	NO	NO	NO	SI	4 giorni
4	Pianura occidentale	NO	NO	NO	SI	6 giorni
5	Pianura centrale	NO	NO	NO	SI	4 giorni
6	Pianura orientale	NO	NO	NO	SI	5 giorni

Prossima emissione lunedì 6 febbraio 2017

Notes: the figure shows the App and PDF document layout of the bulletins available to farmers to learn the status of prohibitions between November and March. Provider: ERSAF Lombardia. App: *Nitrati*.

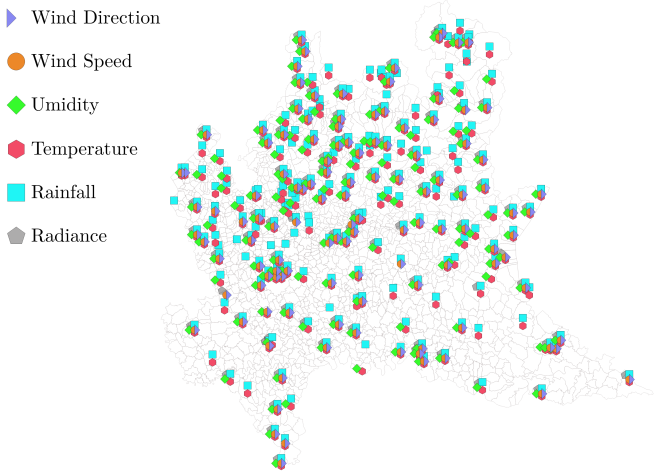
Figure E.5: Stations Geolocations

Panel [A]



Stations count	
Ozone	25
PM_{10}	39
$PM_{2.5}$	41
SO_2	36

Panel [B]

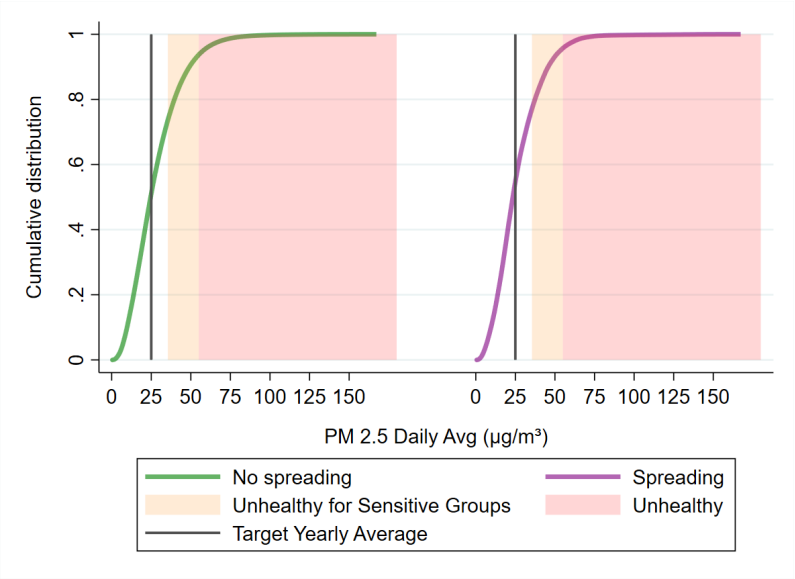


Stations count	
Humidity	94
Radiance	82
Rainfall	154
Temperature	154
Wind Direction	112
Wind Speed	112

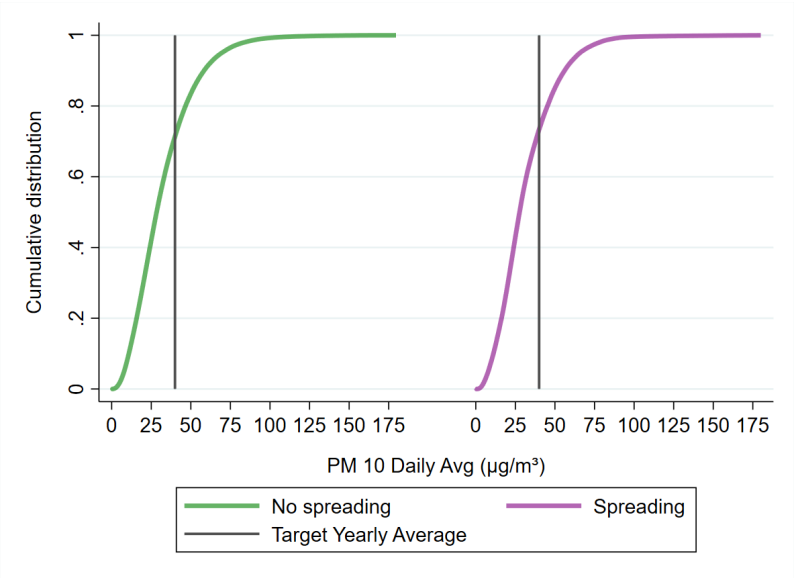
Notes: The figure reports geolocation and counting of stations for pollutants concentrations (Panel A) and weather factors (Panel B) in the Lombardy region. Only stations active continually between 2016 and 2019 are considered.

Figure E.6: Daily average distributions (November to February) - $PM_{2.5}$ and PM_{10}

[A] $PM_{2.5}$

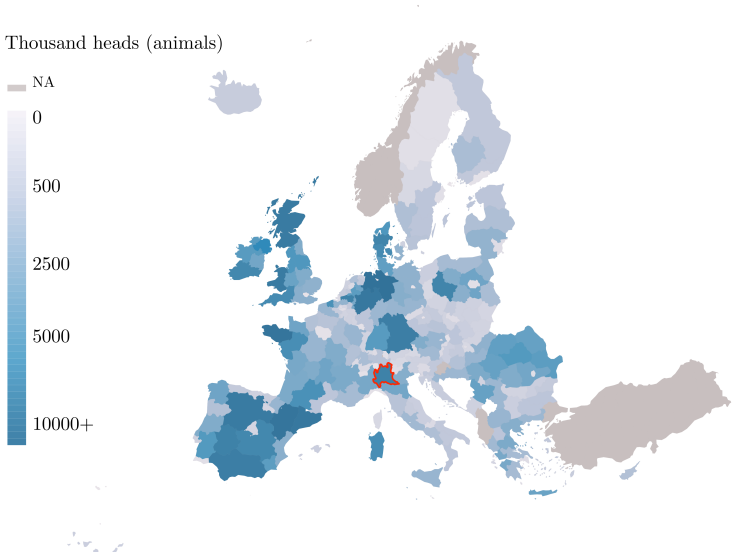


[B] PM_{10}



Notes: the figure plots the cumulative distribution of daily average concentrations of $PM_{2.5}$ (Panel [A]) and PM_{10} (Panel [B]), when spreading prohibitions are enforced (left) and when they are lifted (right). Vertical lines mark Italy's target values for yearly average concentrations. For $PM_{2.5}$, levels unhealthy for sensitive groups (35 $\mu g/m^{-3}$) and generally unhealthy (55 $\mu g/m^{-3}$) are reported. Reference values are defined according the standards for the Air Quality Index (AQI) set by the US Environmental Protection Agency (EPA).

Figure E.7: Farming animals concentration - EU NUTS-2 level (2010-2021 average)



Notes: the figure shows the average total headcount of four livestock animals (bovine, sheeps, goats, suine) between 2010 and 2021. Data is aggregated at NUTS-2 Eurostat classification level (Germany and UK data is only available at NUTS-1 level).
EU average (SD): 1300.8 (1817.715); EU max: 11267.27; Lombardy average: 5714.293.
Source: Eurostat: Animal Production Statistics.

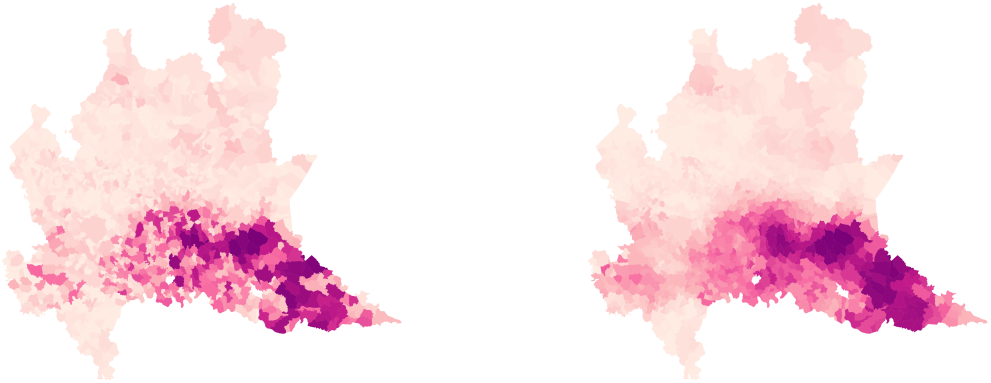
Figure E.8: Rainfall and window days correlation

Notes: the figure plots the coefficient of a regression of rainfall levels on a set of indicators for relative time from the window opening, weather controls and FEs. Weather controls include temperature, humidity, radiance, wind speed and direction, average PBLH, up to a third lag and interacted within each other. FEs comprise day-of-the-week-by-year, holiday, municipality-by-month, month-by-year, and window. Clustering of standard errors happens at rainfall station level.

Figure E.9: Rainfall and window days correlation

[I] Unbuffered

[II] 5km-radius buffering

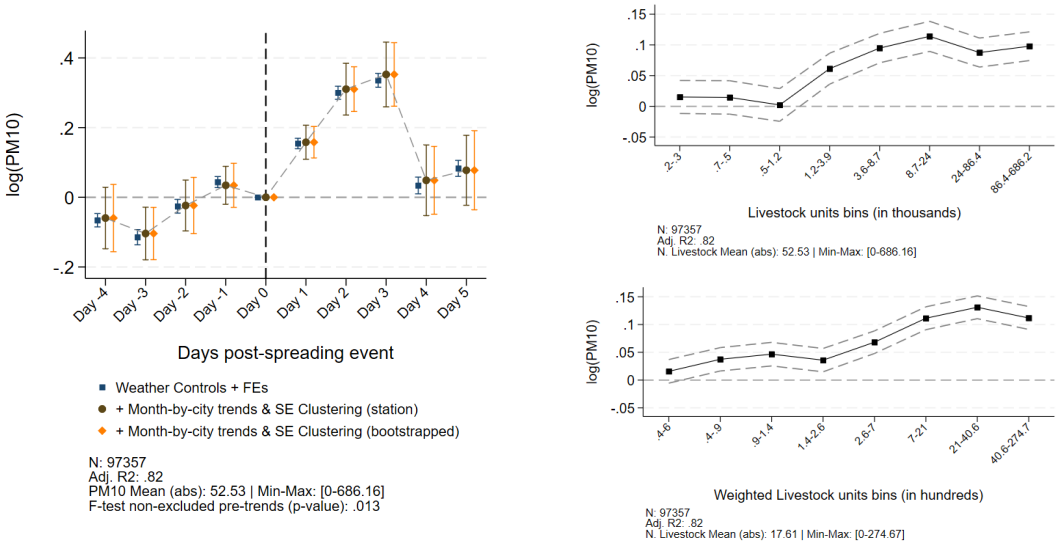


Notes: the figure exemplifies a visual representation of how livestock concentration is calculated using buffering. The stock of animals at municipal levels is recomputed as the average of animal stocks in neighbouring municipalities within a 5km radius. Using 2km and 10km radius leads to comparable results.

Figure E.10: Effect of manure spreading on PM_{10} concentrations (baseline)

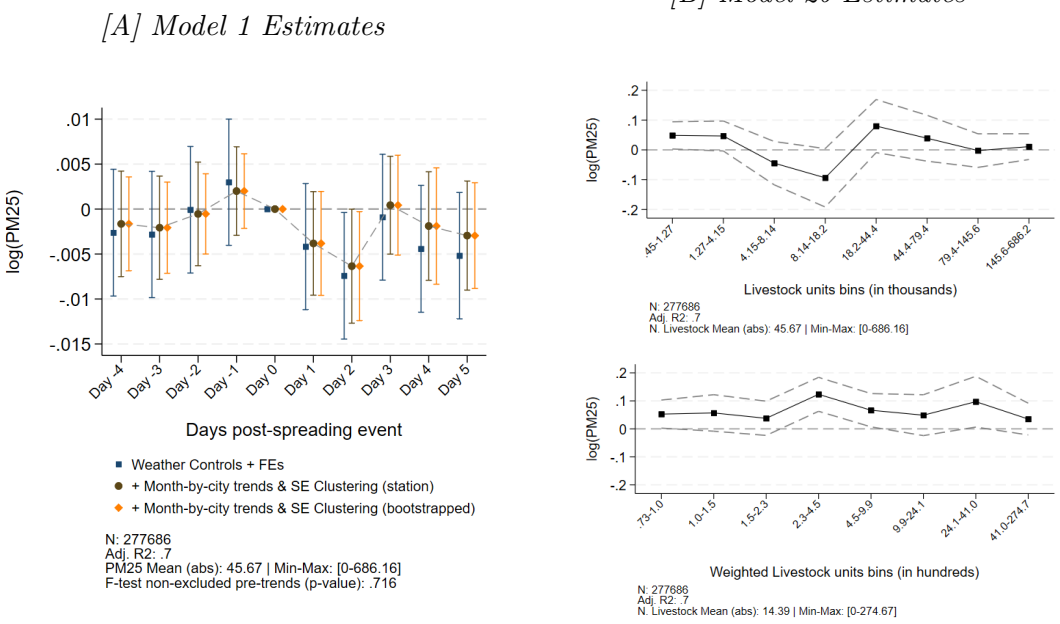
[A] Model 1 Estimates

[B] Model 2b Estimates



Notes: Panel [A] plots the effect of manure spreading windows opening on log PM_{10} as in Figure 7. Controls include weather conditions up to the third lag and interacted in each period, and FEs include municipality, month-by-year, day of the week, holiday, window. Bootstrapped standard errors are sampled with 100 iterations. Coefficients on last prohibition day (day 0) are set to zero. Confidence level plotted at 95% level.
Panel [B] provides the estimates of the ρ_b coefficients in the static model 2b.

Figure E.11: Placebo test - effect of rain windows on $PM_{2.5}$ concentrations

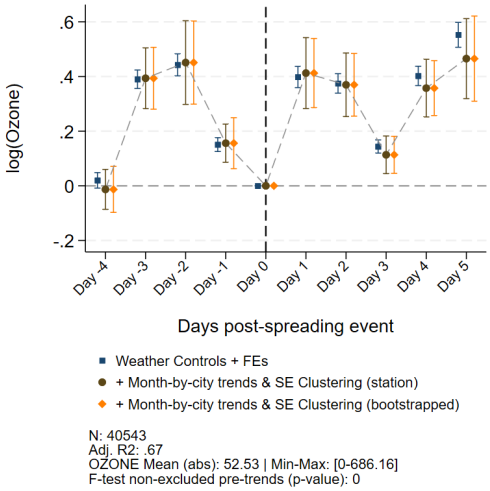


Notes: Panel [A] plots the effect of placebo rain windows opening on log $PM_{2.5}$ concentrations. Controls include weather conditions up to the third lag and interacted in each period, and FEs include municipality, month-by-year, day of the week, holiday, window. Bootstrapped standard errors are sampled with 100 iterations. Coefficients two days after the rainy event (day 0) are set to zero. Confidence level plotted at 95% level.
Panel [B] provides the estimates of the ρ_b coefficients in the static model 2b, with placebo rain window opening. Weighted livestock units at municipality level are computed using the weights in Table 2.

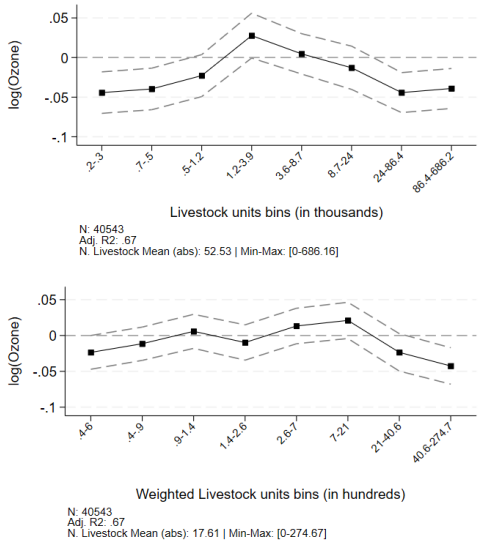
Figure E.12: Effect of manure spreading on O_3 and SO_2 concentrations

Panel [I - Ozone]

[A] Model 1 Estimates

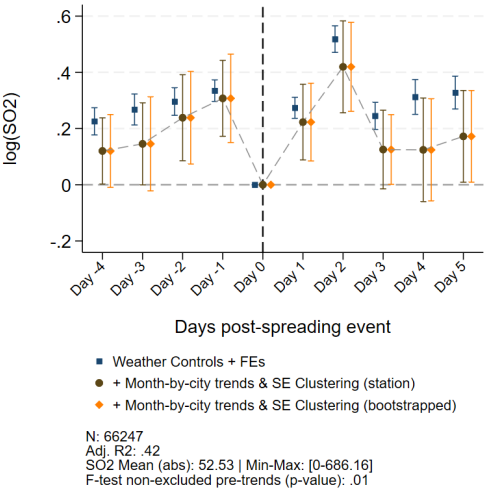


[B] Model 2b Estimates

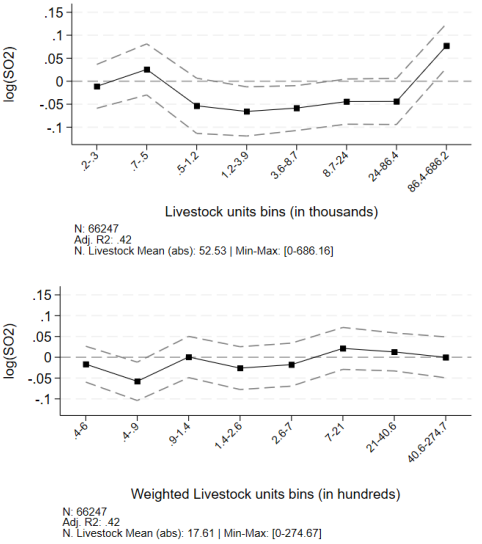


Panel [II - SO_2]

[A] Model 1 Estimates



[B] Model 2b Estimates



Notes: the figure plots the effect of manure spreading windows opening on log concentrations of other pollutants, plotting the estimates for η_k [A] and ρ_b [B] coefficients from Equation 1 and 2 where the outcome variable is $\log(SO_2)$ and $\log(O_3)$ concentrations. Controls include weather conditions up to the third lag and interacted in each period, and FEs include municipality, month-by-year, day of the week, holiday, window. Bootstrapped standard errors are sampled with 100 iterations. Coefficients on last prohibition day (day 0) are set to zero. Confidence level plotted at 95% level. Weighted livestock units at municipality level are computed using the weights in Table 2. Panel [I] reports the coefficients for O_3, Panel [II] those for SO_2.

*

The published version of this paper is available at:
https://doi.org/10.1016/j.eiar.2024.107456

*

Exploring the Impact of Livestock on Air Quality: A Deep Dive into Ammonia and Particulate Matter in Lombardy

Abstract

The linkage between agricultural activities, particularly livestock farming, and atmospheric pollution is broadly acknowledged, and its magnitude is widely analyzed. Lombardy, one of Europe's most critical areas with regard to air pollution, has significantly large contributions from the farming industry. Although studies aimed at informing policy reflect uncertain and moderate pollution reduction even under simulated stringent policy scenarios, granular causal evidence at a sub-sector level remains insufficient to inform local and regional policies effectively. In this study, we employ a spatially and temporally indexed econometric model to investigate the specific impact of bovine and swine farming on the concentration levels of ammonia (NH_3) and coarse particulate matter (PM_{10}) in Lombardy's atmosphere. Our findings indicate that an increase of 1000 units in livestock, equating to roughly a 1% and 0.3% rise in the average per-quadrant bovine and swine populations, respectively—triggers a corresponding daily increase in NH_3 and PM_{10} concentrations. These increases are quantified as 0.26 [0.22; 0.33] and 0.29 [0.27; 0.41] $\mu g/m^3$ for bovines (about 2% and 1% of the respective daily averages) and 0.01 [0.01; 0.05] and 0.04 [0.004; 0.16] $\mu g/m^3$ for swine. Notably, these impacts are intensified under northerly upwind conditions, minimizing the potential for concurrent pollution sources and reinforcing the robustness of our estimated impacts. Finally, we employ our findings to extrapolate the potential environmental implications of reducing livestock emissions. Our analysis suggests that bovine and swine farming could account for up to 25% of local pollution exposure, emphasizing the need for targeted mitigation strategies.

Keywords: Ammonia; Particulate Matter; Air Pollution; Livestock; Farming;

Acknowledgments

This work has received funding from the Fondazione CARIPLO under the project: “INHALE - Impact on humaN Health of Agriculture and Livestock Emissions”.

1 Introduction

Atmospheric particulate matter (PM) ranks as a major environmental health threat (Burnett, et al., 2018), and the fourth mortality risk factor worldwide: in 2019, 1 in 9 death worldwide were caused by fine particulate matter ($PM_{2.5}$) and ozone (O_3) air pollution[1], with the former contributing to such outcome by more than 94% (Murray, et al., 2020). By threatening human welfare through poor air quality, PM also implies a large morbidity burden on individuals: exposure to high PM levels has been associated with increased incidence of respiratory and cardiovascular diseases, such as asthma, pneumonia, hyper- tension, and diabetes (Dominici, et al., 2006; Feng, et al., 2016; Mannucci, Harari, & Franchini, 2019).

While there exists a large amount of literature focusing on the effects of industrial activities and motor-vehicle traffic on air pollution and health, the empirical evidence about the effects of farming on the concentration of human-threatening pollutants is relatively scarcer (Anenberg, et al., 2019; Gibson & Carnovale, 2015; He, Liu, & Salvo, 2019). Indeed, livestock farms are a key contributor to PM emissions (Pue et al., 2019). Animal husbandry operations are responsible for large releases of ammonia (NH_3), a gaseous alkaline compound that serves as a precursor in secondary particle formation, from reactions with other compounds, such as sulfur oxides (SO_x) and nitrogen oxides (NO_x), ammonia contributes to a major part of the inorganic composition of $PM_{2.5}$. This explains why air pollution from livestock farms is associated with airway obstruction diseases and severe pneumonia (Borlée, et al., 2017; Kalkowska, et al., 2018).

In the case of Lombardy, farming constitutes almost the only source of ammonia releases: the emission inventory of the Lombardy environmental agency (INEMAR, consulted on 16/09/2020) estimates that as much as 97% of all its emissions originate from farming activities in the Italian Po-valley region. Since 2005, Italy has successfully reduced NO_x and SO_2 emissions from major sources (Marco, et al., 2019). NO_x and SO_2 emissions have decreased by 41% and 70% respectively, since 2005 and 2016, primarily thanks to policies

[1] Source: The Institute for Health Metrics and Evaluation (IHME) data. Available at: http://ghdx.healthdata.org/gbd-2019

tackling emissions from road traffic, residential heating, and industry (Marco, et al., 2019). Conversely, ammonia decreased only by 10% and $PM_{2.5}$ emissions show positive and negative variations from 2005 to 2016 resulting in a 7% reduction between 2005-2016. Thus, ammonia remains a concern, as actions in the agriculture sector have been less consistent, and PM levels remain high compared to the rest of Europe, especially in Lombardy. The detrimental role of livestock in the absence of efficient air pollution control practices is well recognized in the literature (McDuffie, et al., 2021). Yet, the marginal contribution of the different species of farming animals to ammonia and, in turn, PM concentrations is still poorly understood. The emission factor of a farming animal can vary considerably, depending, among others, on species, animal characteristics, facility type, and manure removal system. As such, different measurement methodologies and experimental settings have resulted in a vast range of possible emission factors attributable to a single unit. By reviewing multiple approaches and studies, Hristov et al. (2011) find emission factors from cows varying from 0.82 to 250 g ammonia per day. In a similar effort, Philippe et al. (2011) reported the same value for swine, which was between 0.38 and 27.2 g per day. However, there have been limited efforts to measure the impact of animals on ammonia and PM levels on a significant scale. Roman et al. (2021), which looked at particulate emissions from animal farming rather than concentrations, find higher values in rural areas compared to urban areas and that the contribution of animal farming to PM emissions varied significantly across different regions in Poland. Spencer and Van Heyst (2018) provide a review of the literature on PM emissions resulting from different sources in Canadian agricultural and rural areas. The study found that PM emissions from agricultural and rural sources, including animal farming, can contribute to elevated PM concentrations in these areas and negatively impact human health. Livestock intensity changes can be attributed to concentration, which has a direct impact on human exposure and health, unlike emissions-specific factors. In this paper, we approach the problem of quantifying livestock-originating concentration from a broader perspective.

A wide variety of source apportionment techniques are available (Thunis, et al., 2019). Some of these techniques employ bottom-up models that perturb source emissions (Thunis, et al., 2019), while others utilize inverse modeling (Carozzi, et al., 2013) or tagged

trajectories (Kranenburg, et al., 2013). Specifically, for PM, numerous methods rely on monitored chemical composition of particles to identify the sources contributing to the overall PM mass (Giardi, et al., 2022). Here we employ a fixed-effects model with spatially and temporally indexed data that builds on exogenous high-frequency variation in wind direction and detailed data on farming animals' movements across the Lombardy region in Italy. We estimate the marginal impact of two animal kinds (cattle and swine) on ammonia and PM_{10} levels. Lombardy offers a particularly suitable setting for the analysis: in addition to providing publicly available high-frequency information on pollutants and weather conditions through a granular network of sensors, it is one of the most farming-intensive regions in Europe, with more than 1 million live cattle and 4 million live swine head (see Figure 1). This, in turn, results in frequent movements of animals in and out. We take advantage of this variation to accurately identify the impact of farming on the concentration of pollutants. We access daily observations from 12 ammonia monitoring stations and 75 PM_{10} measuring points. For three stations, we obtain PM chemical decomposition data that allows us to isolate the share of ammonium sulfates (AS) and ammonium nitrates (AN), two inorganic salts that are part of the secondary PM share and are directly associated with the NH_3 precursor.

We combine this information with daily weather conditions and monthly fluctuations in livestock units. We use variation in animal heads occurring in the upwind quadrant of a given sensor (the 90-degree portion of a circular area around the sensor) to estimate the marginal impact of farming animals on the levels of ammonia and PM_{10} recorded at the station level. Using variation in wind direction allows our specification to cope with potentially endogenous movements in livestock units induced by air pollutants. Indeed, conditional on observables and fixed effects, in order to identify the causal impact of farming animals on pollutant levels, our specification crucially rests on the assumption of orthogonality between livestock allocation decisions, weather conditions, and air quality considerations.

Using our estimates, we then simulate average daily levels of PM_{10} under the counterfactual stylized scenario of removing all livestock units around a sensor. We find a simulated local percentage reduction in daily concentrations of up to 25%. While

negatively correlated with average daily levels of particulate matter, the drop in concentrations under our counterfactual simulation still emerges as a sizeable improvement in daily air quality for many densely populated areas.

Our results differ in nature from those retrieved in micro-level studies. Our study aims to quantify the average relative contribution of livestock animals to station-level recorded concentrations of pollutants rather than pinpointing emissions in terms of mass and differentiating for animal characteristics. Previous studies on the impacts of livestock on air quality focus mainly on emissions (Hristov A. N., 2011; INEMAR, consulted on 16/09/2020; Kabelitz et al., 2020; Roman, Roman, Roman, 2021), and those going beyond emissions used averaged emission factors derived from the emissions studies (Pue, Bral, Buysse, 2019; Rao et al., 2017). At a later stage, the most completed ones would then use source-receptor models or chemical transport models to derive concentrations and, ultimately, exposure (Lelieveld et al. 2015; McDuffie et al. 2021). As such, this research adds to the existing literature by estimating the contribution of different animal species on the levels of harmful pollutants in a highly polluted and livestock-dense area of Europe, a topic often overlooked in comparison to the livestock contribution to greenhouse gas emissions (Kipling et al. 2019*a*,*b*, Garnett 2009). The paper contributes by establishing a necessary step to evaluate the nature of the direct correlation between changes in livestock levels and the impact on human health due to air pollution. The use of causal inference methods is a novel approach to this type of analysis, and our findings are functional to policymakers' informed decisions regarding farming practices and air pollution control measures.

The remainder of the paper is organized as follows. Section 2 details the empirical strategy employed. Section 3 describes the data, and Section 4 reports the main estimation results. Section 5 explores effect heterogeneity, while Section 6 presents a counterfactual calculation of pollutant concentrations and policy considerations following the evidence at hand. Finally, Section 7 concludes.

2 Methods

We calculate the incremental impact of a unit of animal, per species, on ammonia and

PM_{10} levels. It's important to understand that our calculation is based on the number of animals within a fixed area. As a result, the impact can be interpreted as a variation in animal density, since the area remains constant. For this reason, we refer to 'intensity' as the quantity of animals per unit area. To estimate the marginal contribution to ammonia and PM_{10} concentrations specific to each farming animal at the aggregate level, we estimate the following regression through OLS:

$$Y_{s,t} = \beta_0 + \sum_{j \in \mathcal{B} \cap \mathcal{G}} \sum_{a \in A} \beta_a \Delta L_{a,j,t} \times \mathbf{\Omega} + \mathbf{X}'_{s,t}\mathbf{\Gamma} + \delta_{s,q} + \delta_m + \delta_y + \varepsilon_{s,t} \qquad (1)$$

where Y represent the dependent variable of interest, ΔL is our main regressor, $\mathbf{\Omega}$ is a weighting matrix, X is a matrix of covariates with the respective coefficients ($\mathbf{\Gamma}$), and δs captures the fixed effects of our model. β_a represents the main coefficient of our study. As our study focuses on pinpointing the causal effect of livestock on pollutant concentrations, OLS has the advantage, under our set of assumptions, to provide an unbiased and easily interpretable estimator for β.

The outcomes of interest (Y) are ammonia concentrations (NH_3), overall particulate matter (PM_{10}), and mass concentration of ammonium sulfates and nitrate (PM_{10}^{ASN}) measured daily by station s at time t. ΔL is the net sum of inflows and outflows for animal a (both within the region and from and to other regions and countries), births, and slaughters at the municipal level. The set $\mathcal{B}$ characterises a municipality j as:

$$\{ j \in \mathcal{B} : d_{i,j} < \bar{r} \}$$

hence containing municipalities within $\bar{r}$ distance from municipality i. We alternatively consider 50km and 60km centroid-distance as the two values of $\bar{r}$.[2] The set $\mathcal{G}$ is instead

[2] There exists no universal rule to assess the distance potentially traveled by pollutants, as this is closely dependent on the area's morphology, wind conditions, and the nature of airborne particles. As such, we set the boundaries of circular areas around sensors employing a data-driven approach. In Figure A2.2 in the Appendix, we show the sensitivity of the estimated β_a from Equation 1 to gradually expanding circular areas from a radius of 10km. For both animals and pollutants, the coefficient of interest converges to an asymptotic value between 40km and 60km, which leads us to exclude variations in livestock headcount taking place outside of this range. Furthermore, as the area radius increases, so does the probability of sensors

defined as:

$$\{ j \in \mathcal{G} : \angle ij_t \in WD_{i,t} \}$$

and includes all municipalities that are in the same quadrant of the direction from which wind originates as measured in municipality i at time t ($\text{WD}_{i,t}$). We use the concept of geometric angle ($\angle$) to indicate that municipalities are assigned to quadrants depending on the angle between the station and the municipality. We consider four quadrants: North (315 - 45), East (45 - 135), South (135 - 225), and West (225 - 315). Thus, for each station, we obtain a time-specific total variation in the number of livestock units (ΔL), calculated as the sum of variations at the municipal level for all municipalities that are located in the quadrant of wind direction at time t and within distance $\bar{r}$ from the station. To visualize the quadrant-wind direction variation strategy implemented, we provide a graphical illustration in Figure 2. A is instead a set of two farming animals, bovine and swine, for which monthly variation in headcount is available.

This variation in livestock units is only available at the monthly level, while ammonia levels and weather conditions are measured daily. Given the impossibility of exactly pinpointing the day of the variation in farming animals' headcount, we test the robustness of the results by applying a set of analytic weights to magnify the weight of observations occurring toward the end of the month. By defining analytic weights as the probability of a given variation in animal headcount has realized (assuming a probability increasing linearly and monotonically), we want to impose that data points at the beginning of the month could be estimating the marginal effect captured by our β less precisely. Analytic weights are equivalent to assuming observation j belongs to a sub-population with variance $\frac{\sigma^2}{w_j}$, where σ^2 is a common variance and w_j is the weight of the observation j.[3] This is justified by thinking that, during the last days of each month, the movements depicted

relatively far from each other showing overlapping circular areas, which may induce noise in our estimates. The average intra-sensor distance is 75.4km and 74.5km for ammonia and PM_{10} sensors respectively. Hence, we deem the $60km$ threshold to be an adequate upper bound to consider variation in livestock units relevant to a given sensor.

[3] Weights are implemented in Stata. For further, refer to Stata Technical Bulletin, issue 20, July 1994.

with monthly frequency in the data are more likely to be fully realized. Specifically, observations on the first day of the month are assigned a weight of 1/30, while observations on the last day of each month are assigned a weight of 1, with other observations in between weighted with a monotonic linear increment of 1/30.

$\boldsymbol{\Omega}$ is a diagonal matrix of weights based on the distance between sensors and municipalities. This is motivated by assuming that the impact on ammonia levels of animals that are in closer proximity to the station will be stronger than that of animals further away, as dispersion of emissions during transportation will be less likely to occur. As such, $\boldsymbol{\Omega}$ partially discounts the variation happening further away from each station. Considering a total of M municipalities around a given sensor, in our baseline specification, $\boldsymbol{\Omega}$ is an identity matrix of size $M \times M$ (i.e., no discounting based on distance implied). We then test the robustness of our results by populating $\boldsymbol{\Omega}$ with linear and Gaussian distance weights[4]. To better understand how we compute variation in livestock units under different weighting schemes, a numerical example is reported in the Appendix (Section A3.1).

X is a matrix of weather controls, including temperature, rainfall, radiance, wind speed, humidity, and boundary layer height, up to the third lag and interacted with each other, and $\boldsymbol{\Gamma}$ is a matrix of coefficients. $\delta_{s,q}$, δ_m , and δ_y represent a set of sensor-by-quadrant, month, and year-fixed effects[5]. Sensor-by-quadrant allows for different time-invariant intrinsic characteristics not only across sensors, but also around a sensor and, as such, it

[4] Linear weights are computed as in Equation 2:

$$w_{ij} = 1 - \frac{d_{ij}}{\bar{r}} \tag{2}$$

while Gaussian weights obey to Equation 3:

$$w_{ij} = \begin{cases} 0 \text{ if } d_{ij} \geq \bar{r} \\ \exp\left(\frac{1}{2}\left(\frac{d_{ij}}{\bar{r}}\right)^2\right) \text{ if } \frac{\bar{r}}{\sqrt{2}}\ln\left(2\pi^{-\frac{1}{2}}\right) < d_{ij} < \bar{r} \\ 1 \text{ otherwise} \end{cases} \tag{3}$$

[5] The use of lags and interactions, as well as the choice of fixed effects, follows the strategy adopted by Deryugina, et al (2019). The results are robust to less conservative structure of weather covariates, excluding lags and interaction terms.

is deemed as the most conservative approach.[6] Our model estimates 66 weather control parameters and 62 fixed effects parameters overall, in addition to our coefficients of interest. Finally, $\varepsilon_{s,t}$ is an error term, assumed to be normally distributed. We allow for variance in error to be dependent on our regressors, estimating heteroskedasticity-robust standard errors[7].

The marginal effect of a livestock unit for animal a is captured by our main coefficient of interest, β_a. To be able to identify, the variation in the number of animals at the municipal level should be independent of ammonia levels and PM levels. If farmers were to time their buying, selling, and slaughtering decisions based on air quality, this could induce a reverse causality bias in our estimates. Despite the absence, in the current regulatory framework of Lombardy, of policies aimed to curb livestock presence as a function of pollution levels, even assuming that part of farmers' decision concerning animal net flows is indirectly correlated with air quality, the use of wind direction to mediate the source of variation in livestock units allows us to restore exogeneity. Indeed, in our specification, it is enough to assume non-adapting behavior from farmers to wind flows, i.e., animal stock decisions being independent of observed and expected wind flows. In addition, the presence of station-, quadrant-, and time-fixed effects allows differentiating part of the confounding variation that may be related to more polluted areas with relatively more frequent animal displacement[8].

Conditional on observables and fixed effects, we also argue against the likelihood ofour results being driven by the presence of omitted variable bias. While we can look separately at the fluctuations in the concentration of the two among the most important farming animals in terms of pollutant contribution, the absence of data available on the movements

[6] The results are mostly unchanged when only sensor fixed effects are included in the regression.

[7] The choice of heteroskedasticity-robust standard errors is motivated by our model using a relatively small number of sensors and a large number of temporal observations (days), hence likely inducing serial correlation in the error. We refrain from clustering standard errors at the sensor level given the limited number of clusters available, especially with regards to ammonia stations (Cameron & Miller, 2015: Abadie, et al., 2023). To assess the presence of residual correlation in the model, we plot model residuals against the fitted values in Figure A2.3.

[8] To this aim, it is also important how, with specific reference to ammonia, more than 95% of total emissions are ascribable to livestock. This importantly reduces the concern of unobservables spatially correlated with sensor proximity (e.g. other agricultural activities) inducing bias in the estimates. PM concentrations are more susceptible to confounding emission sources, which, however, are less likely to be spatially correlated with proximity to a measuring station, such as traffic or industrial activities.

of other animals, particularly poultry, may be especially concerning, being this the third major specie in terms of air pollutants contribution.[9] This would be particularly concerning under the hypothesis that variations in our measure of animal headcount at the municipal level may co-vary with the unobservable variation in the number of other farming animals (especially in the case of multi-breed farms) whose effect, in turn, would be wrongly imputed to variation in cattle and swine units alone, biasing the estimator.

However, we notice in Figure A2.4 that the share farms specializing in more than one animal are relatively small. Farms whose production includes at least two out of three species (cattle, chickens, and swine) are less than 1% of all breeders, with the share of farms breeding all three animals being less than 0.2%. This is partially confirmed by observing that variation in the number of cattle and swine units at the municipal level does not correlate[10]. Moreover, poultry farming appears to be more concentrated, with a relative density of more than 19,000 animals per farm, the same figure being 103 for cattle and 544 for swine. Thus, while it is not possible to fully rule out the possibility of noise induced by the absence of comprehensive data on all farming animals, the low likelihood of correlated shocks significantly reduces the concern of omitted variable bias, reinforcing our assumption. Finally, our model implies linearity in the effect of livestock change. This assumption simplifies the intricate process of PM formation through secondary aerosol via chemical reactions with ammonia, which can lead to non-linear effects at varying concentrations. Amid this simplifying assumption, our model serves as a valuable reference point, as it enables us to analyze the overall contribution of livestock under minimal computing and modeling requirements.

3 Data description

A flowchart describing the data used in this paper can be found in Figure 3. We access

[9] While the emission factor of hen is importantly lower than cattle and swine, data from INEMAR quantify poultry total particulate matter emissions in the Lombardy region at 438.9 tons, with the same number for cattle and swine being respectively 358.6 and 739.4 tons. The contribution of other animals (ovine, equine) is marginal. No disaggregated data on ammonia emissions are currently available.
10 Pearson's product-moment correlation coefficient: 0.02

publicly available daily data on NH_3 and PM_{10} concentration levels and weather conditions in the Lombardy region from ARPA Lombardia[11]. We focus on the years between 2015 and 2020 to match the frequency of livestock data. Some stations have been active for a short amount of time during those years (as the measurement activity ceased or started at the extremes of our sample period)[12]. As such, we restrict the sample to stations for which daily concentration data is available for at least 365 days between 2015 and 2020, obtaining 12 NH_3 stations and 75 PM_{10} stations. For a subset of stations (Schivenoglia, Milano Pascal, Milano Senato), for a total of 3299 sensor-day combinations available, we obtain information on the mass concentration of ammonium nitrates and ammonium sulfates, two compounds that enter the composition of PM_{10} and require ammonia to form. In Lombardy, the share of ammonium salts on the total PM mass can be higher than 50% (Lanzani, et al., 2020). We obtain a final dataset of 16577 day-station-wind direction observations for NH_3, 109663 observations for PM_{10}, and 3299 observations for decomposed AS and AN. Summary statistics on pollutants are reported in Table 1, Panel A. Especially for PM concentrations, variation within the same sensor appears to be larger, given natural seasonal fluctuations. Yet, sizeable differences across stations can be observed, particularly in the case of ammonia concentrations.

Each station is imputed weather conditions recorded at the respectively closest weather stations. We collect daily data on temperatures (°C), rainfall (mm), wind direction (degrees) and speed (m/s), humidity (%), and radiance (W/m^2). In addition, we collect hourly data on Planetary Boundary Layer Height (PBLH) through the ERA5 Reanalysis provided by ECMWF[13] and compute average daily values. Each of these variables directly impacts airborne pollutant concentrations. Warmer temperatures are usually associated with lower concentrations, given higher thermal dispersion. Positively correlated with temperature, PBLH constitutes an even more cogent measure for vertical dispersion: higher

[11] ARPA: Regional Agency for Environmental Protection. The agency collects hourly data on NH_3 and PM_{10} concentrations but disseminates the information as daily averages.
[12] For NH_3 sensors, only two stations are not active throughout the entire period (Cremona Borghi, inactive since January 2017, and Piadena, inactive between March 2014 and June 2016). For PM_{10}, 11 stations cease measurement only between 2017 and 2018.
[13] The measure is provided at $0.25^\circ \times 0.25^\circ$ grid level.

PBL implies increased dispersion capacity and is associated with lower pollutant concentrations (Seidel, Ao, & Li, 2010). Similarly, increased level of rainfall reduces PM concentrations through "wet deposition".

As previously noted, wind speed and direction can affect the presence of pollutants in an area by dispersing pollution plums. With increased humidity, moisture particles grow in size to the point of "dry deposition", reducing PM_{10} concentrations. Finally, radiance can impact PM levels, especially through photochemical reactions. These variables are summarized in Table 1, Panel B.

To visualize the correlation between wind direction and pollutants in the region, we look at the polar plots reported in Figure 4. Lombardy's morphological territory implies lower levels of pollutants are recorded when winds flow from the Alpine arch in the Northern part of the region. In general, wind in the Po Valley plays an important role in dispersing pollutants and leads to lower average concentrations than the winds that flow longitudinally within the region. However, the relative frequencies of wind flowing from each quadrant indicate significant variation across NH_3 stations. For instance, South-East stations are more susceptible to West and North winds, while North-West areas receive more wind from the South. Similarly, PM stations show a prevalence of West winds in the region's central plains, but South-East and South-West areas experience a higher probability of winds flowing respectively from the North and the East. Despite some patterns, considerable variability at the station level is observed. This is particularly relevant for our strategy: observing wind consistently blowing from the same direction throughout the month would imply that our fixed effects structure, which controls both for month and sensor-by-quadrant time-invariant characteristics, would absorb most of the effect of the change in livestock units. In this case, our coefficients of interest would capture noisy residual variation. Significant variation in wind direction, both within the same sensor and across sensors, mitigates this concern. It is worth noting that the Po Valley, particularly Lombardy, is surrounded by mountains on three sides, which limits outward air circulation and can lead to very low winds and stable conditions, especially in winter. This condition creates the perfect environment for air pollution accumulation, making the region a pollution hotspot.

Data on livestock presence and movements are available through the National Zootechnics Registry (*Anagrafe Nazionale Zootecnica*, ANZ) database. The registry provides monthly municipal-level data on inflows and outflows of livestock (either transferred within municipalities or acquired from and sold abroad), animal slaughtering, and births. Given insufficient data on other farming species, our study focuses on two animals, cattle and swine.[14] These two breeds are the primary contributors to ammonia emissions. Data on newborns for swine are incorporated into monthly inflow data, thus resulting indivisible from positive variation originating from other activities. Conversely, they can be computed separately for cattle.[15] Our analysis is concentrated on Lombardy and its three adjacent regions, namely Piemonte, Veneto, and Emilia Romagna, which includes stations situated near the borders of Lombardy. This helps us consider the presence of animals in close proximity to a station while formally being located across the region's borders. To supplement our data, we compute the stock of animals registered in each municipality, which is available twice a year. This measure enables us to differentiate between areas with high livestock density and those with relatively scarce farming activities.

The municipalities surrounding Lombardy's sensors exhibit the high prevalence of livestock animals typical of the Lombardy region, with an average of more than 1000 cattle units and 2,500 swine units per municipality. Both cattle and swine numbers appear to be decreasing, although the variation is still a relatively small share of the existing stock (Table 2). Figure 5 shows instead how the majority of animal husbandry activities are concentrated in the South-East area of the Po Valley, both in terms of cattle and swine breeding. This is reflected both in average monthly outflows and inflows, which tend to be larger in numbers in areas more populated by farming animals (Figure A2.1) and in consistently higher concentrations of ammonia located within areas of high livestock density

[14] Cattle identifies all bovine farming species, including Italian Mediterranean buffalos. Data on swine is only available starting in 2016.

[15] As we are not able to separate between adult animals and calves for all species in the dataset, in the headcount, we assign to all animals a unit weight. This assumption neglects the difference in emission factors between adults and calves. We deem this strategy viable in our setting in light of the objective to quantify an aggregated impact of livestock movements on airborne pollutants in the region. In addition, given the existence of a positive correlation between adult animals and calves, this distinction is unlikely to induce bias in our estimates.

(Figure 6, Panel A). Conversely, due to the more heterogeneous composition of airborne particulate matter, the spatial correlation between farming animals' presence and PM_{10} is instead blurred. Thus, we employ our empirical strategy to explore the existence and magnitude of a causal relationship between animal husbandry and air pollutants and present our findings in the next Section.

4 Results

The results of estimating Equation 1 are reported in Table 3. At the baseline, we look at variations in the number of animals not discounted by distance from the station. To enhance intuition, we present our estimates in two separate forms.

In Panel A, coefficients have been re-scaled to capture a 1000 livestock units variation at the quadrant level, which is approximately a 1% change in bovines and 0.3% change in swine with respect to the overall average quadrant-level animal density. We report the results separately for the different pollutants considered: NH_3 (Columns 1 to 3), PM_{10} (Columns 4 to 6), and ammonium compounds share of PM (Columns 7 to 9). For each outcome variable, the first two columns show the estimates of βa, respectively, when including only the variation in cattle units and swine units. The third column includes the two variations as separate variables and estimates the marginal contributions when the two regressors are included together. In Panel B, we instead present standardized coefficients of the same estimated relationship. We center the variation around the mean and standard deviation of livestock units present in the neighboring quadrants. As such, one standard deviation increase represents a sizeable shock in animal heads, given the high concentration at the quadrant level.

When looking at concentrations of ammonia, all coefficients are significant, at least at a 5% level across different specifications. The inclusion of variations in both species in the equation has only a minor impact on the respective coefficients. A 1000-unit increase in the number of cattle upwind (Panel A) raises ammonia levels between 0.286 and 0.332 $\mu g/m^3$, resulting in a 1.8% variation from the average ammonia concentrations during the sample period. The effect of a positive variation of 1000 units in swine headcount is more modest,

at around 0.04, or about 0.26% relative to the average concentrations. This can be attributed not only to lower emission factors of swine but also to the fact that swine are almost four times more prevalent in the region than cattle. The standardized coefficients reported in Panel B confirm the relatively sizeable impact of livestock variation for both species: one standard deviation increase in cattle in an upwind quadrant leads to a 1.63 to 1.51 standard deviation spike in ammonia concentration. A similar increase in swine results in a 0.85 standard deviation spike.

Looking at the same estimated effect for PM_{10}, despite PM mass concentration being almost double in size compared to ammonia, the marginal impact estimated is comparable in magnitude to the one previously obtained. Indeed, upwind 1000-units increases in cattle and swine units are expected to increase PM concentrations by respectively 0.247 to 0.289 $\mu g/m^3$ and 0.01 to 0.04 $\mu g/m^3$ (which are respectively around 0.8% and 0.03% deviations from mean concentrations). On the one hand, this evidence supports the validity of our empirical strategy: if our estimates had been affected by confounding factors, the impact on PM and ammonia concentrations would not necessarily be equal, as these are present in the atmosphere with varying levels of mass concentrations. On the other hand, similarity in the coefficients shows how positive variation in livestock units induces a comparable increase in NH_3 and PM_{10} concentrations and, as such, supports the belief that PM mass concentrations observed when livestock increases are indeed the result of secondary aerosol formation through ammonia.

While we would expect the observed increase in PM_{10} to be attributable to ammonium nitrates and ammonium sulfates particles spurring from ammonia gaseous emissions, the relatively different and not significant coefficients observed in Columns 7 to 9 can be explained by data on $PM_{10}{}^{ASN}$ being available only for three stations, which implies around 3% of the entire station-day level sample for PM. Furthermore, two sensors are located in the Milan area, where pollutants from other sources are present in the highest concentration. Even when the assumptions of our empirical model are satisfied, a sizeable reduction in the sample size may violate the asymptotic properties of our estimator, implying less precise and potentially biased estimates. With these caveats in mind, it is still meaningful to notice that the main coefficients remain positive and deviate by a small amount, with respect to

sample average concentrations, when compared to their counterpart estimated for ammonia and overall PM concentrations.

We then proceed to explore the robustness of our results, addressing two main concerns with our empirical design. First, the variation in livestock units cannot be identified with daily frequency. As such, we repeat the estimations, placing more weight on the observations of air pollutant concentrations occurring towards the end of the month, where the shift in the animal count is more likely to be fully realized. The results obtained are comparable in magnitude and significance to our baseline estimates (Table A1.1 in Appendix). Second, as we argued that animals further away from the sensor location may contribute differently to pollutant measurement than those located in close proximity to it, we apply different specifications of $\mathbf{\Omega}$, i.e. varying the distance discounting weights to the variation in livestock units. In this case, coefficients are not directly comparable to the ones obtained before, as the weighting implies a rescaling of our main regressor (ΔL) and, in other words, inevitably inflates the magnitude of $\hat{\beta}_a$ by magnifying the relevance of a one-unit increase. To compare our estimates, we iteratively simulate a 1000-unit increase in a quadrant and use the derived values of $\Delta L \times \mathbf{\Omega}$ in the given quadrant to rescale the estimated coefficients.[16] We summarize the results in Figure 7[17]. Each weighting method calculates a corresponding distribution of the estimated coefficient by multiplying the point estimates and the simulated 1000-unit variation distribution. The median result is then marked and compared to the point estimates of the non-weighted strategy. While linear discounting affects the estimated coefficients by a more sizeable amount with respect to Gaussian weighting, different specifications of $\mathbf{\Omega}$ lead to comparable results. The marginal effect of 1000 cattle units oscillates between 0.22 and 0.33 $\mu g/m^3$ of NH_3 and 0.27 and 0.41 $\mu g/m^3$ of PM_{10}. The same variation in terms of swine units provides estimates fluctuating

[16] To clarify this aspect implied by our weighting strategy further, assume a 1000-units positive variation taking place around a station. Livestock units are located at a random distance $\tilde{d}$ from the sensor, where $\tilde{d}$ is drawn form a uniform distribution $\tilde{d} \sim U(0, \bar{r})$. Each unit is then assigned a distance-based weight according to our different weighting strategies. It is then computed the corresponding ΔL (refer to the numerical example in Section A3.1). By randomly simulating the distance of each unit, we are actively randomizing the weight received by each unit. This, in turn, implies a different computed value of $\Delta L \times \mathbf{\Omega}$ depending on the outcome of the randomization. To show it, we iteratively simulate (10,000 iterations) a 1000-units positive variation around a station and apply the corresponding weighting to each unit. We then plot the corresponding value of $\Delta L \times \mathbf{\Omega}$ in Figure A2.5.

[17] Estimates of the weighted variation strategy are reported in Appendix, Tables A1.2, A1.3, A1.4.

between around 0.02 and 0.05 $\mu g/m^3$ of NH_3 and 0.004 and 0.16 $\mu g/m^3$ of PM_{10}.

To get a better understanding of our results, we need to stress how the impact on concentrations is fundamentally different from that on emissions, which causes our results to be inherently separate from emission factors more commonly found in the literature[18]. Concentrations are influenced by the specific geographical, meteorological, and chemical conditions of the region where the emissions occur. This is why we can only draw a partial analogy between our estimated impact on ammonia concentrations and the ammonia emission factors from the regional emission inventory (INEMAR, consulted on 16/09/2020), which would otherwise constitute a natural benchmark, at least in terms of geographic region. Comparing these, we observe a similar order of magnitude difference between cattle and swine emission factors as the one identified in our estimates, with cows showing emissions one order of magnitude higher. We cannot, however, make the same comparison with PM_{10} concentrations and the corresponding emission factors. In fact, the latter pertain to direct emissions, whereas our estimates also include secondary PM_{10} concentrations.

There, our results provide a robust and new perspective on the aggregate impact of animal husbandry on concentrations of air pollutants in a region with a high density of livestock, such as Lombardy. This evidence can help guide the cost-benefit analysis of expansions and reduction of livestock intensity from a policymaking perspective. To this aim, we explore heterogeneity in effect retrieved that may result in better-informed policy considerations.

5 Heterogeneity and Sensitivity

We test the sensitivity and heterogeneity of our results in two ways. First, we account for potential differential effects of livestock variation depending on the quadrant of the source. To this aim, we add a set of interactions to Equation 1, letting the marginal impact of farming

[18] For instance, Hristov et al. (2011) find an average ammonia emission factor 59 g per cow per day. Philippe et al. (2011) provide a summary of swine emission factors under different waste management systems, between 0.38 and 27.2 g/day. However, it is not straightforward to determine how this would translate into ammonia concentrations at aggregate level in the context of their studies.

animal variation vary through the source quadrant. Analytically, Equation 1 is expanded as follows:

$$Y_{s,t} = \beta_0 + \sum_{j \in \mathcal{B} \cap \mathcal{G}} \sum_{a \in A} \beta_a \Delta L_{a,j,t} + \sum_{j \in \mathcal{B} \cap \mathcal{G}} \sum_{a \in A} \sum_{q \in Q} \eta_a \Delta L_{a,j,t} \times D_q + \mathbf{X}'_{s,t} \mathbf{\Gamma} + \delta_{s,q} + \delta_m + \delta_y + \varepsilon_{s,t} \quad (4)$$

where, for simplicity, we consider the absence of weighting ($\mathbf{\Omega}$=I), and D_q is an indicator assuming value 1 when variation originates from quadrant $q \in \mathcal{Q}$ (the set including the four quadrants), zero otherwise. Note that our fixed effects structure naturally absorbs the differential intercept for each quadrant. The results are presented graphically in Figure 8. We take as reference group livestock headcount variation happening in Southern quadrants. The results highlight how movements in farming animals tend to have a larger impact on pollutant concentrations at the sensor level when they occur to the North of a station. This finding appears in line with the evidence presented in Figure 4: North winds are usually associated with lower levels of pollutants, which reduces the extent of confounding variation, particularly with respect to particulate matter, and makes fluctuations in the livestock units more crucial in driving up and down the concentrations of airborne pollutants. The effect appears instead to be homogeneous across other quadrants, with smaller and primarily non-significant coefficients associated with the interaction terms.

Second, we investigate whether the effect retrieved is driven by using only a limited number of sensors. This is particularly of concern when considering ammonia concentrations measured on a relatively smaller network of stations. The presence of one or few sensors driving the results may cast doubt over the accuracy and generalizability of our results. To this aim, we iteratively repeat the estimation, dropping one sensor at each iteration. The new coefficients obtained for ammonia through this methodology are plotted in Figure 9. On the horizontal axis is reported the name of the dropped station. Stations are sorted from left to right according to the number of animal units within the defined $\bar{r}$ radius circular area. The coefficients remain relatively stable with some minor fluctuations, and most instances show significance at a 95% level. In Panel B, we also notice that only one sensor offers a noticeable fluctuation in the effect retrieved, which is located in the Corte de Cortesi municipal area. This can be attributed to the proximity of a large swine farm

near the station.[19] This station was purposely placed next to a large-scale swine livestock facility in order to monitor emissions from swine husbandry. Similarly, the Bertonico station is located next to a large-scale cattle husbandry area to closely monitor concentrations in the farming area.[20] In turn, local fluctuations in ammonia levels originating from daily farming activities of different natures may overcast the movements in animal units taking place further away from the station, hence inducing particular noise in the estimates retrieved through our empirical strategy.[21] Nonetheless, while the coefficient decreases in magnitude when excluding the sensor from the sample, it remains positive and comparable in size.

Since the sample available for PM_{10} includes a considerably larger number of sensors, dropping a single sensor has a more marginal impact on the overall sample. Hence, to assess the presence of sensors in critical areas driving the results, we repeat the above procedure but drop all stations in a province (Figure 10).[22] The results again show minor fluctuations around the average estimated effect, proving the relative stability of the effect of farming animals across the region.

6 Policy Considerations

Assessing the agricultural sector's impact on ammonia and particulate matter (PM) concentrations is crucial for policymaking in Lombardy. The region is susceptible to environmental and health threats due to its dense population, intense farming, and low wind conditions caused by its orographic features. To comprehend the implications of our findings, we propose a straightforward calculation to determine the toll that farming takes on air

[19] The sensor is located within 100 meters from the breeding facility. The exact location of the farm is excluded for data privacy.

[20] The sensor is located between two facilities placed within 1 and 1.5 km. The exact location of the farms is excluded for data privacy.

[21] For instance, ammonia levels can fluctuate due to manure management practices, such as storage and disposal, or even due to the application of nitrogen-based fertilizers, which can release ammonia gas into the air. This can lead to the release of ammonia into the air, affecting local air quality. The use of litter and manure management practices can also contribute to fluctuations in local ammonia levels in poultry farming operations. Finally, the handling of dairy waste, such as urine and manure, can also lead to local fluctuations in ammonia levels.

[22] The Lombardy region is divided into 12 provinces. In brackets, the number of PM_{10} sensors per province is reported: BG (9); BS (6); CO (3); CR (6); LC (5); LO (7); MB (4); MI (11); MN (8); PV (7); SO (4); VA (5).

quality and, consequently, public health.

Our objective is to establish the impact of farming animals on air pollution levels in the area surrounding a station. Using data from ISTAT[23], we calculate the resident population within a 50km radius of the station and couple it with information on the number of livestock units within each circular area. We then simulate a hypothetical scenario where we remove all farming animals from each circular area, *all else equal* (i.e., keeping all other observable and unobservable factors constant, including weather conditions), leveraging the coefficients we obtained from a 1000-unit variation analysis to estimate the corresponding reduction in concentrations of air pollutants.[24][25] This exercise does not aim to explore a viable policy action to improve air quality in the region (i.e., the complete dismantlement of the farming industry) but rather to provide an estimate of the contribution of livestock to daily pollutant concentrations. Given the linearity of our approach, the expected results of a less sizeable reduction in livestock units can be easily inferred from our analysis. Moreover, provided that adverse health effects are associated with PM rather than gaseous ammonia alone, which instead acts as a precursor to the particulate formation, in this part of the paper, we only focus on PM_{10} concentrations.

Panel A in Figure 11 shows the results of this exercise by plotting the reduction in daily PM_{10} concentrations over twenty sensor bins, with the latter calculated conditioning on yearly average concentrations. Panel B plots the same reduction paired with the total resident population in each bin. Two main considerations are in place. First, it appears that the areas with lower average daily concentrations of PM_{10} are more severely affected by the threat to air quality posed by livestock (Panel A). The largest reduction (approximately 25%) observed in the simulation is in sensors with an average yearly

[23] Source: Resident Population on 1st January.

[24] This strategy once again simplifies by assuming the effect to be linear and unsusceptible to the number of livestock units already present in the area. While this may constitute a limitation to our approach, we still deem this procedure informative to approximate the true impact of the farming industry on air pollution in the region.

[25] Population and livestock headcount data are available at the municipality level. To avoid double- counting, whenever a municipality lies within a 50km radius of multiple stations, its population is imputed to different circular areas in equal shares. The potential noise in the calculations induced by this strategy is tapered by counterfactual concentrations being computed as the mean across stations in the same decile of the distribution of yearly average concentrations. As stations in close proximity are likely to register similar yearly levels of pollutants, the population in the area is likely to be imputed the same counterfactual exposure levels regardless of whether individuals are assigned to one station or the other.

concentration of less than 30 $\mu g/m^3$. This can be attributed to the fact that areas with more farming activity generally have a lower degree of urbanization and a reduced incidence of emission factors from other industries like transportation, construction, and manufacturing. However, this also means that less urbanized areas are disproportionately burdened by the presence of livestock and are unable to fully benefit from high air quality.

Second, looking at Panel B, the areas touched more heavily by air pollution from livestock sources, despite lower urbanization, display considerably high population density: nearly 7 out of 14 million people reside within 50km of those stations that would benefit from a counterfactual level of PM_{10} concentrations below $30\mu g/m^3$ in our simulation. Furthermore, circular areas around stations that would experience the highest percentage reduction (more than 20%, peaking at roughly 25%) appear surrounded by almost 2 million inhabitants.[26] These findings highlight how the estimated deterioration in air quality is likely to affect a significant proportion of the population rather than being limited to sparsely populated rural municipalities.

Our simulations advocate for integrated policies in the agricultural sector, particularly in densely populated regions with high livestock density, like Lombardy, where the secondary formation of ammonium nitrates often reaches more than 50% of the total PM mass (Tao et al. 2016, Wu et al. 2020). It is particularly important to target concentration reduction that can effectively minimize the effects of agricultural activities. These may include the use of BATs (best available technologies, e.g., injector systems and genetic engineering) in agriculture and farming practices, improved integrated management of farming activities (such as improved animal diet, efficient disposal of slurry and manure, and efficiency in the production system), and livestock intensity (Ammann et al. 2022, OECD 2019).

[26] In this calculation, we do not factor in individuals residing outside the 60km circular areas used to obtain our estimates, as this would require a more comprehensive analysis of how pollutants are transported across the region, which is beyond the scope of this paper.

7 Conclusion

This paper estimated the marginal impact of cattle and swine farming on the levels of ammonia and PM_{10} in the Lombardy region. We used daily observations from 12 ammonia monitoring stations and 75 PM_{10} measuring points and combined them with monthly fluctuations in livestock units and daily weather conditions.

The results showed that an increase in upwind cattle and swine presence by 1000 units respectively raised ammonia levels by 0.332 μg/m3 (around 1.8% variation from mean concentrations) and 0.04 μg/m3 (around 0.26% with respect to mean concentrations), and PM_{10} levels by 0.289 μg/m3 and 0.04 μg/m3 respectively. The results are robust to different weighting schemes and provide information on the average relative contribution of livestock to station-level recorded concentrations of pollutants. Our simulation showed that livestock presence is expected to cause sensitive deterioration in air quality for a sizeable share of the region's population. Hence, the study provides insights into the potential impact of changing livestock in the Lombardy region and highlights the need for further research to understand the role of livestock in air pollution. In particular, future research should focus on carefully evaluating the cost-benefit tradeoff involved by technology and organizational practices available in the industry to prevent harmful effects on individual health and guide the evolution of the industry onto a more sustainable path.

References

Abadie, A., Athey, S., Imbens, G. W., & Wooldridge, J. M. (2023). When should you adjust standard errors for clustering? *The Quarterly Journal of Economics, 138*, 1–35.

Ammann, J., Umstätter, C., & El Benni, N. (2022). The adoption of precision agriculture enabling technologies in Swiss outdoor vegetable production: a Delphi study. *Precision Agriculture, 23*, 1354–1374.

Anenberg, S. C., Achakulwisut, P., Brauer, M., Moran, D., Apte, J. S., & Henze, D. K. (2019). Particulate matter-attributable mortality and relationships with carbon dioxide in 250 urban areas worldwide. *Scientific reports, 9*, 11552.

Borlée, F., Yzermans, C. J., Aalders, B., Rooijackers, J., Krop, E., Maassen, C. B., . . . Smit, L. A. (2017). Air pollution from livestock farms is associated with airway obstruction in neighboring residents. *American journal of respiratory and critical care medicine, 196*, 1152–1161.

Bouwman, L., Goldewijk, K. K., Van Der Hoek, K. W., Beusen, A. H., Van Vuuren, D. P., Willems, J., . . . Stehfest, E. (2013). Exploring global changes in nitrogen and phosphorus cycles in agriculture induced by livestock production over the 1900–2050 period. *Proceedings of the National Academy of Sciences, 110*, 20882–20887.

Burnett, R., Chen, H., Szyszkowicz, M., Fann, N., Hubbell, B., Pope Iii, C. A., . . . others. (2018). Global estimates of mortality associated with long-term exposure to outdoor fine particulate matter. *Proceedings of the National Academy of Sciences, 115*, 9592–9597.

Calori, G., Finardi, S., Nanni, A., Radice, P., Riccardo, S., Bertello, A., & Pavone, F. (2008). Long-term air quality assessment: modeling sources contribution and scenarios in Ivrea and Torino areas. *Environmental Modeling & Assessment, 13*, 329–335.

Cameron, A. C., & Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. *Journal of human resources, 50*, 317–372.

Carozzi, M., Loubet, B., Acutis, M., Rana, G., & Ferrara, R. M. (2013). Inverse dispersion modelling highlights the efficiency of slurry injection to reduce ammonia losses by agriculture in the Po Valley (Italy). *Agricultural and Forest Meteorology, 171*, 306–318.

Carslaw, D. C., Beevers, S. D., Ropkins, K., & Bell, M. (2006). Detecting and quantifying aircraft and other on-airport contributions to ambient nitrogen oxides in the vicinity of a large international airport. *Atmospheric Environment*, 5424-5434.

Deryugina, T., Heutel, G., Miller, N. H., Molitor, D., & Reif, J. (2019). The mortality and medical costs of air pollution: Evidence from changes in wind direction. *American Economic Review, 109*, 4178–4219.

Dominici, F., Peng, R. D., Bell, M. L., Pham, L., McDermott, A., Zeger, S. L., & Samet, J. M. (2006). Fine particulate air pollution and hospital admission for cardiovascular and respiratory diseases. *Jama, 295*, 1127–1134.

Feng, S., Gao, D., Liao, F., Zhou, F., & Wang, X. (2016). The health effects of ambient PM2. 5 and potential mechanisms. *Ecotoxicology and environmental safety, 128*, 67–74.

Galloway, J. N., Townsend, A. R., Erisman, J. W., Bekunda, M., Cai, Z., Freney, J. R., . . . Sutton, M. A. (2008). Transformation of the nitrogen cycle: recent trends, questions,

and potential solutions. *Science, 320*, 889–892.

Garnett, T. (2009). Livestock-related greenhouse gas emissions: impacts and options for policy makers. *Environmental Science & Policy, 12*, 491-503. doi:https://doi.org/10.1016/j.envsci.2009.01.006

Giardi, F., Nava, S., Calzolai, G., Pazzi, G., Chiari, M., Faggi, A., . . . others. (2022). PM10 variation, composition, and source analysis in Tuscany (Italy) following the COVID-19 lockdown restrictions. *Atmospheric Chemistry and Physics, 22*, 9987–10005.

Gibson, M., & Carnovale, M. (2015). The effects of road pricing on driver behavior and air pollution. *Journal of Urban Economics, 89*, 62–73.

He, J., Liu, H., & Salvo, A. (2019). Severe air pollution and labor productivity: Evidence from industrial towns in China. *American Economic Journal: Applied Economics, 11*, 173–201.

Hristov, A. N. (2011, June). Technical note: Contribution of ammonia emitted from livestock to atmospheric fine particulate matter (PM2.5) in the United States. *Journal of Dairy Science, 94*, 3130–3136. doi:10.3168/jds.2010-3681

Hristov, A. N., Hanigan, M., Cole, A., Todd, R., McAllister, T. A., Ndegwa, P. M., & Rotz, A. (2011). Ammonia emissions from dairy farms and beef feedlots. *Canadian journal of animal science, 91*, 1–35.

INEMAR. (consulted on 16/09/2020). *Principali risultati Inventario 2017 (revisione pubblica).* Tech. rep., INEMAR. Retrieved from https://www.inemar.eu

Kabelitz, T., Ammon, C., Funk, R., Münch, S., Biniasch, O., Nübel, U., . . . Amon, T. (2020, October). Functional relationship of particulate matter (PM) emissions, animal species, and moisture content during manure application. *Environment International, 143*, 105577. doi:10.1016/j.envint.2020.105577

Kalkowska, D. A., Boender, G. J., Smit, L. A., Baliatsas, C., Yzermans, J., Heederik, D. J., & Hagenaars, T. J. (2018). Associations between pneumonia and residential distance to livestock farms over a five-year period in a large population-based study. *PLoS One, 13*, e0200813.

Kipling, R. P., Taft, H. E., Chadwick, D. R., Styles, D., & Moorby, J. (2019). Challenges to implementing greenhouse gas mitigation measures in livestock agriculture: A conceptual framework for policymakers. *Environmental Science & Policy, 92*, 107-115. doi:https://doi.org/10.1016/j.envsci.2018.11.013

Kipling, R. P., Taft, H. E., Chadwick, D. R., Styles, D., & Moorby, J. (2019). Implementation solutions for greenhouse gas mitigation measures in livestock agriculture: A framework for coherent strategy. *Environmental Science & Policy, 101*, 232-244. doi:https://doi.org/10.1016/j.envsci.2019.08.015

Kranenburg, R., Segers, A. J., Hendriks, C., & Schaap, M. (2013). Source apportionment using LOTOS-EUROS: module description and evaluation. *Geoscientific Model Development, 6*, 721–733.

Lanzani, G., D'Angelo, L., Cuccia, E., Corbella, L., Santo, U. D., Colombi, C., . . . Deserti, M. (2020). Ammoniaca e formazione di particolato secondario. *Ecoscienza*.

Lelieveld, J., Evans, J. S., Fnais, M., Giannadaki, D., & Pozzer, A. (2015, September 16). The contribution of outdoor air pollution sources to premature mortality on a global scale. *Nature, 525*, 367—371. doi:10.1038/nature15371

Mannucci, P. M., Harari, S., & Franchini, M. (2019). Novel evidence for a greater burden of

ambient air pollution on cardiovascular disease. *Haematologica, 104*, 2349.

Marco, A. D., Proietti, C., Anav, A., Ciancarella, L., D\textquotesingleElia, I., Fares, S., . . . Leonardi, C. (2019, April). Impacts of air pollution on human and ecosystem health, and implications for the National Emission Ceilings Directive: Insights from Italy. *Environment International, 125*, 320–333. doi:10.1016/j.envint.2019.01.064

McDuffie, E. E., Martin, R. V., Spadaro, J. V., Burnett, R., Smith, S. J., O'Rourke, P., . . . Brauer, M. (2021, June). Source sector and fuel contributions to ambient PM2.5 and attributable mortality across multiple spatial scales. *Nature Communications, 12*. doi:10.1038/s41467-021-23853-y

Murray, C. J., Aravkin, A. Y., Zheng, P., Abbafati, C., Abbas, K. M., Abbasi-Kangevari, M., . . . others. (2020). Global burden of 87 risk factors in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *The lancet, 396*, 1223–1249.

OECD. (2019). *Digital opportunities for better agricultural policies.* OECD.

Philippe, F.-X., Cabaraux, J.-F., & Nicks, B. (2011). Ammonia emissions from pig houses: Influencing factors and mitigation techniques. *Agriculture, ecosystems & environment, 141*, 245–260.

Pue, D. D., & Buysse, J. (2020). Safeguarding Natura 2000 habitats from nitrogen deposition by tackling ammonia emissions from livestock facilities. *Environmental Science & Policy, 111*, 74-82. doi:https://doi.org/10.1016/j.envsci.2020.05.004

Pue, D. D., Bral, A., & Buysse, J. (2019). Abatement of ammonia emissions from livestock housing fine-tuned according to impact on protected habitats. *Agricultural Systems, 176*, 102667. doi:10.1016/j.agsy.2019.102667

Rao, S., Klimont, Z., Smith, S. J., Dingenen, R. V., Dentener, F., Bouwman, L., . . . Tavoni, M. (2017, January). Future air pollution in the Shared Socio-economic Pathways. *Global Environmental Change, 42*, 346–358. doi:10.1016/j.gloenvcha.2016.05.012

Roman, M., Roman, K., & Roman, M. (2021). Spatial variation in particulate emission resulting from animal farming in Poland. *Agriculture, 11*, 168.

Seidel, D. J., Ao, C. O., & Li, K. (2010). Estimating climatological planetary boundary layer heights from radiosonde observations: Comparison of methods and uncertainty analysis. *Journal of Geophysical Research: Atmospheres, 115*.

Spencer, J., & Van Heyst, B. (2018). A review of particulate matter emissions and impacts on human health: A focus on Canadian agricultural and rural emission sources. *Canadian Biosystems Engineering/Le génie des biosystèmes au Canada, 60*, 6–9.

Tao, Y., Ye, X., Ma, Z., Xie, Y., Wang, R., Chen, J., . . . Jiang, S. (2016). Insights into different nitrate formation mechanisms from seasonal variations of secondary inorganic aerosols in Shanghai. *Atmospheric Environment, 145*, 1–9.

Thunis, P., Clappier, A., Tarrasón, L., Cuvelier, C., Monteiro, A., Pisoni, E., . . . others. (2019). Source apportionment to support air quality planning: Strengths and weaknesses of existing approaches. *Environment International, 130*, 104825.

Thunis, P., Pisoni, E., Zauli Sajani, S., Monforti-Ferrario, F., Bessagnet, B., Vignati, E., & De Meij, A. (2023). *Urban PM2.5 Atlas, Air Quality in European Cities, 2023 Report.* Luxembourg (Luxembourg): Publications Office of the European Union. doi:10.2760/63641 (online),10.2760/200848 (print)

Westmoreland, E. J., Carslaw, N., Carslaw, D. C., Gillah, A., & Bates, E. (2007). Analysis of

air quality within a street canyon using statistical and dispersion modelling techniques. *Atmospheric Environment*, 9195-9205.

Wu, C., Zhang, S., Wang, G., Lv, S., Li, D., Liu, L., . . . others. (2020). Efficient heterogeneous formation of ammonium nitrate on the saline mineral particle surface in the atmosphere of East Asia during dust storm periods. *Environmental science & technology, 54*, 15622–15630.

Tables

Table 1: Descriptive statistics - Pollutants and Weather

	Overall	Within	Between
Panel A - Pollutants			
NH_3	15.74		
($\mu g/m^3$)	(19.97)	(14.06)	(12.84)
	[0.0 ; 430.6]	[-29.1 ; 429.2]	[3.0 ; 45.4]
PM_{10}	30.61		
($\mu g/m^3$)	(20.70)	(19.99)	(5.15)
	[0.0 ; 264.0]	[-10.0 ; 264.4]	[13.0 ; 41.6]
PM_{10} (AS + AN)*	10.95		
($\mu g/m^3$)	(10.78)	(10.77)	(0.57)
	[0.0 ; 58.3]	[0.0 ; 58.7]	[10.5 ; 11.7]
Panel B - Weather			
Temperature	13.87		
(° C)	(8.25)	(8.12)	(1.41)
	[-11.3 ; 32.7]	[-7.1 ; 31.4]	[9.7 ; 15.3]
Rainfall	0.05		
(mm)	(2.19)	(2.19)	(0.04)
	[0.0 ; 256.8]	[-0.1 ; 256.8]	[0.0 ; 0.1]
Wind Speed	1.97		
(m/s)	(0.95)	(0.92)	(0.30)
	[0.0 ; 26.3]	[-0.4 ; 26.4]	[1.5 ; 2.6]
Wind Direction	176.01		
(Degree)	(97.61)	(95.62)	(21.82)
	[0.1 ; 360.0]	[-28.8 ; 404.2]	[131.8 ; 205.0]
Radiance	161.25		
(W/m^2)	(103.94)	(103.64)	(8.12)
	[0.0 ; 517.6]	[-18.4 ; 528.7]	[150.2 ; 179.6]
Humidity	73.22		
(%)	(16.83)	(16.17)	(4.85)
	[0.0 ; 100.0]	[-2.1 ; 107.1]	[65.5 ; 79.9]
PBLH	1,654.82		
(m)	(1,415.71)	(1,412.01)	(100.24)
	[11.4 ; 5,553.5]	[-127.5 ; 5,543.8]	[1,439.3 ; 1,803.9]

Notes: the table reports summary statistics for pollutants (A) and weather variables (B). Mean values are presented first, both within the same sensor across time and between the sensor and the overall mean. Parentheses include standard deviations. Brackets report minimum and maximum values. Within and between statistics are computed through the command xtsum in Stata.
Source: ARPA Lombardia, ECMWF.

Table 2: Descriptive statistics - Livestock

	Cattle			Swines		
	Overall	Within	Between	Overall	Within	Between
Inflow*	13.84			456.75		
(monthly)	(59.43)	(57.55)	(16.11)	(1,429.43)	(1,348.00)	(565.28)
	[0.0 ; 1,663.0]	[-37.0 ; 1,641.0]	[0.8 ; 50.9]	[0.0 ; 23,932.0]	[-967.8 ; 23,342.1]	[1.9 ; 1,424.5]
Births**	43.39			-		
(monthly)	(88.57)	(80.46)	(42.84)	-	-	-
	[0.0 ; 1,379.0]	[-74.4 ; 1,362.2]	[8.9 ; 117.8]	-	-	-
Outflow	-6.61			-450.06		
(monthly)	(30.35)	(29.91)	(6.02)	(1,650.23)	(1,583.29)	(547.31)
	[-1,201.0 ; 0.0]	[-1,197.3 ; 10.2]	[-16.8 ; -0.5]	[-20,431.0 ; 0.0]	[-19,957.8 ; 994.5]	[-1,444.6 ; -0.7]
Slaughters	-57.16			-370.86		
(monthly)	(190.71)	(179.90)	(74.50)	(1,044.20)	(980.95)	(412.65)
	[-3,643.0 ; 0.0]	[-3,511.9 ; 137.0]	[-194.2 ; -5.2]	[-14,064.0 ; 0.0]	[-13,575.4 ; 631.7]	[-1,002.6 ; -1.5]
Net variation	-262.63			-415.14		
(monthly)	(4,719.15)	(3,977.39)	(3,035.25)	(433.52)	(263.28)	(384.11)
	[-20,706.0 ; 9,072.0]	[-16,007.1 ; 13,482.5]	[-7,324.1 ; 2,386.3]	[-18,574.10 ; 10,742.0]	[-14,710.72 ; 53,331.7]	[-10,574.50 ; -11.2]
Tot animals	137,984			320,928		
(quadrant)	(139,002)	(91,009)	(117,959)	(430,301)	(320,167)	(291,739)
	[2,204.25 ; 497,245]	[-170,965.74 ; 351,963]	[13,925.67 ; 326,303]	[0.00 ; 1,642,738]	[-552,364.85 ; 1,448,462]	[1,117.73 ; 873,293]
Tot animals	1,088			2,533		
(municipality)	(2,382)	(197)	(2,318)	(7,655)	(1,102)	(7,189)
	[1.00 ; 35,915]	[-5,255.98 ; 4,957]	[1.00 ; 34,079]	[0.00 ; 94,944]	[-16,040.20 ; 34,015]	[0.00 ; 85,873]

Notes: the table reports summary statistics for livestock variables. Mean values are presented first, both within the same sensor across time and between the sensor and the overall mean. Parentheses include standard deviations. Brackets report minimum and maximum values.

* Inflow and outflow variables include animal movements taking place between facilities within and outside the region.

** Data on newborns for swine are incorporated into the provided measure for monthly inflow by the data provider and cannot be accessed separately.

Source: National Zootechnics Registry.

Table 3: Baseline Estimates

	NH_3			PM_{10}			PM_{10}^{ASN}		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A/ - $\Delta 10^3$-units									
Δ - Cattle	0.332***		0.286**	0.247***		0.289***	0.118		0.150
	(0.106)		(0.112)	(0.052)		(0.052)	(0.13)		(0.14)
Δ - Swine		0.040**	0.0403***		0.004	0.0099***		0.014	0.0147
		(0.016)	(0.016)		(0.003)	(0.003)		(0.02)	(0.02)
Panel B/									
Δ - Cattle	1.63***		1.51**	1.38***		1.62***	1.01		1.28
	(0.66)		(0.69)	(0.29)		(0.29)	(1.12)		(1.18)
Δ - Swine		0.84**	0.85***		0.08	0.2123***		0.30	0.3230
		(0.36)	(0.36)		(0.06)	(0.06)		(0.42)	(0.41)
Observations	16579	13919	13919	109202	109650	109650	3299	2790	2790
Adj R^2	0.5767	0.5694	0.5698	0.5144	0.5143	0.5146	0.5061	0.5109	0.5114
Dep. Var. Mean	15.53	15.53	15.53	30.42	30.42	30.42	10.68	10.68	10.68
Weather Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y
Month FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Sensor-by-quadrant FE	Y	Y	Y	Y	Y	Y	Y	Y	Y

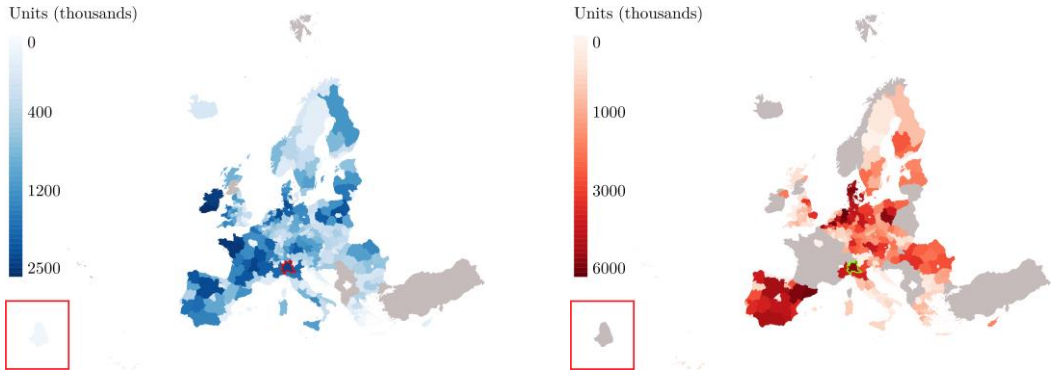
Notes: the table reports the estimates of β_a from Equation 1, where $\mathbf{\Omega}$ is an identity matrix (absence of distance weighting). Weather controls include temperature, wind direction, wind speed, rainfall, radiance, humidity, and average PBLH, interacted with each other up to three lags. Robust standard errors are reported in parentheses.

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Figures

Figure 1: Livestock presence - Eurostat NUTS2 level

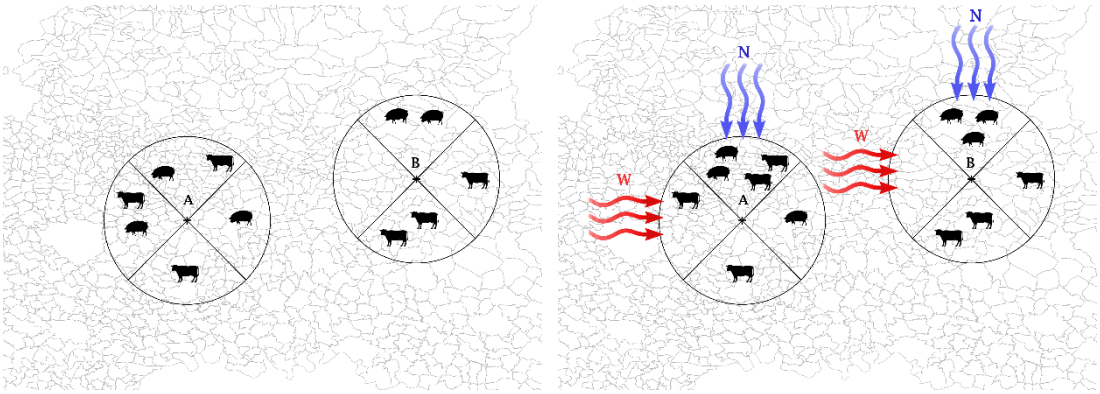
[A] - Cattle [B] - Swine



Notes: the figure reports live cattle (Panel A) and swine animals (Panel B) across European NUTS2 regions (the French Guiana region is relocated in the bottom left of the map). The Lombardy region (framed) is the 14th area in terms of absolute units of bovine in Europe, 8th in terms of swine absolute units. Units are reported to the most recent data point available (2020 for bovine, 2016 for swine). Grey areas have no data available on livestock presence.
Source: Eurostat.

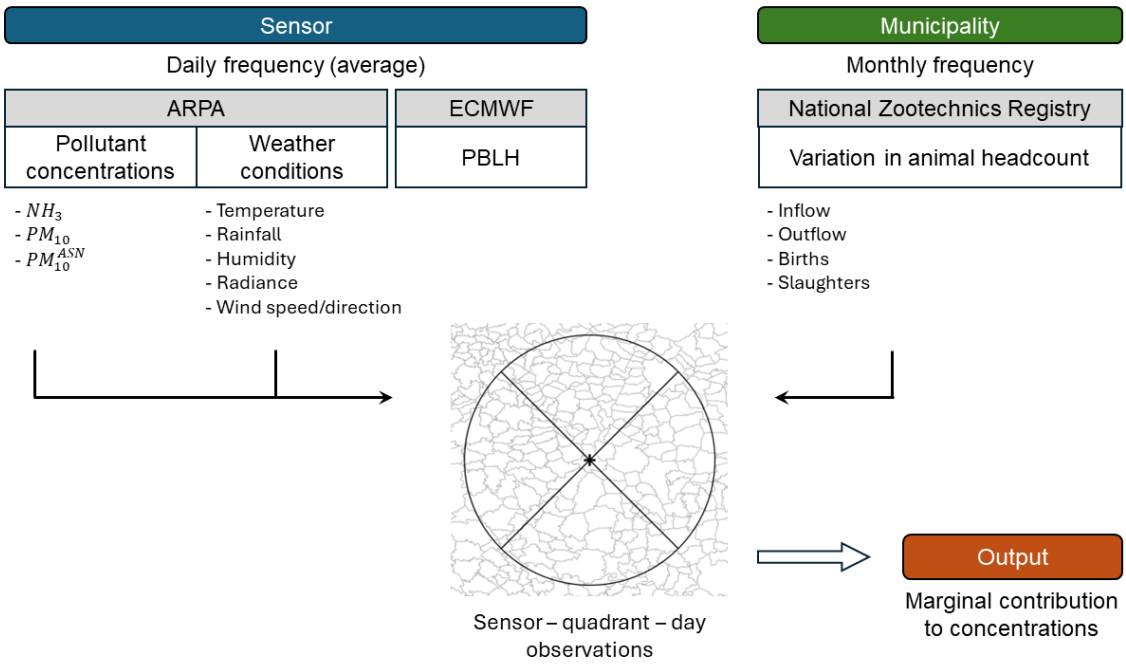
Figure 2: Station quadrants and wind direction

[I] – Time t [II] – Time $t+1$



Notes: the figure provides an example of quadrant-specific variation conditional on wind direction. Consider the stations A and B at time t (Panel I). The existing stock of swine and cattle within r kilometers from the sensor is divided across four quadrants, i.e., 90-degree portions oriented along main cardinal directions (North, East, South, West). Consider the variation in animal headcount from time t in time $t+1$. In our specification, this variation is expected to influence concentrations of pollutants only as long as it takes place upwind from the sensor. For instance, on days when West wind is blowing (red arrows), station A will be imputed a reduction in swine stock, while station B will exhibit no change in animal presence around the station. Conversely, on days of North wind, station A will be imputed a positive change in the stock of both swine and cattle, while in station B the increase will be observed only in swine headcount.

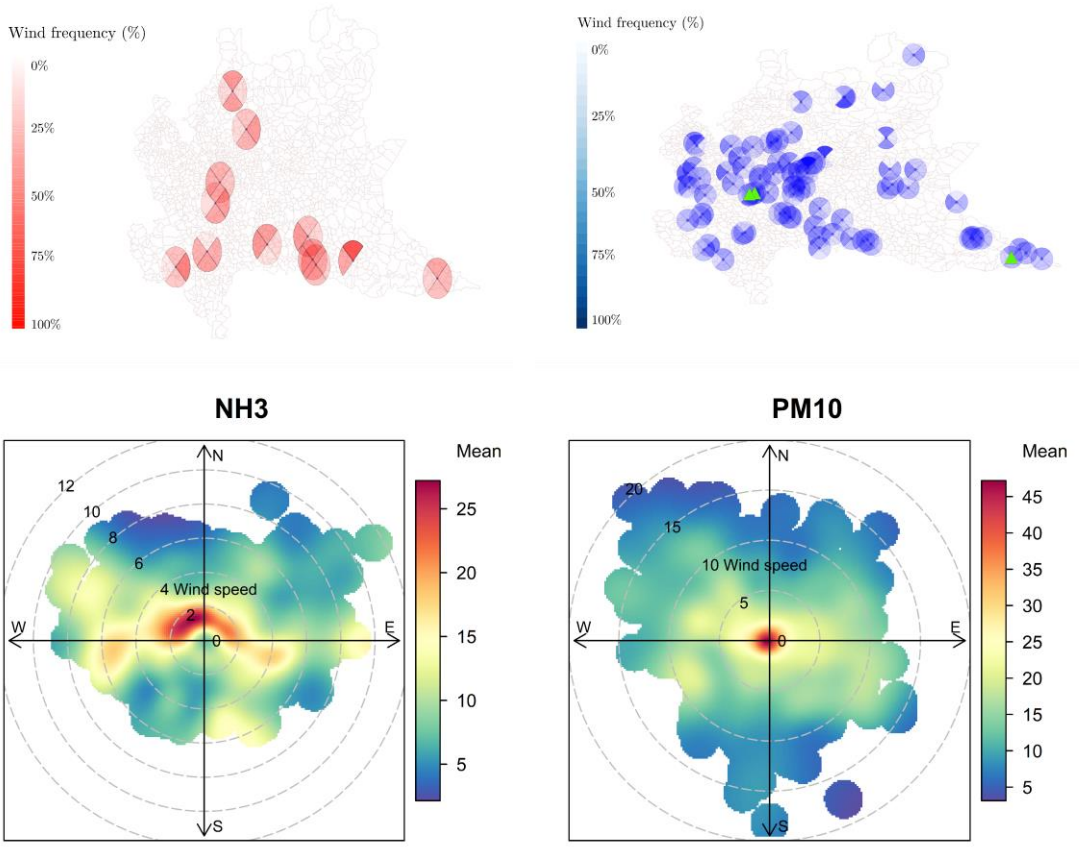
Figure 3: Data sources - flowchart



Notes: the figure plots a flowchart of the data used in our study, specifying the geographic and temporal level of the information available and the corresponding sources.

Figure 4: Frequency of wind directions - regional and sensor values

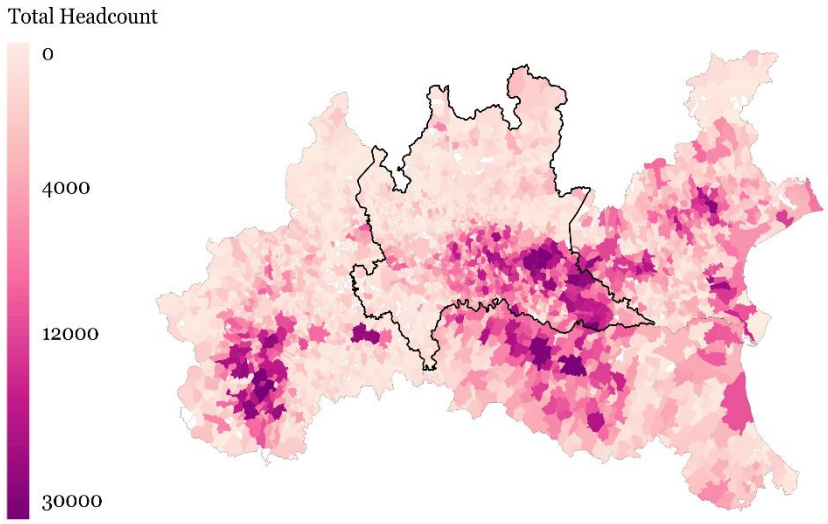
[A] - NH_3 Circular areas [B] - PM_{10} Circular areas



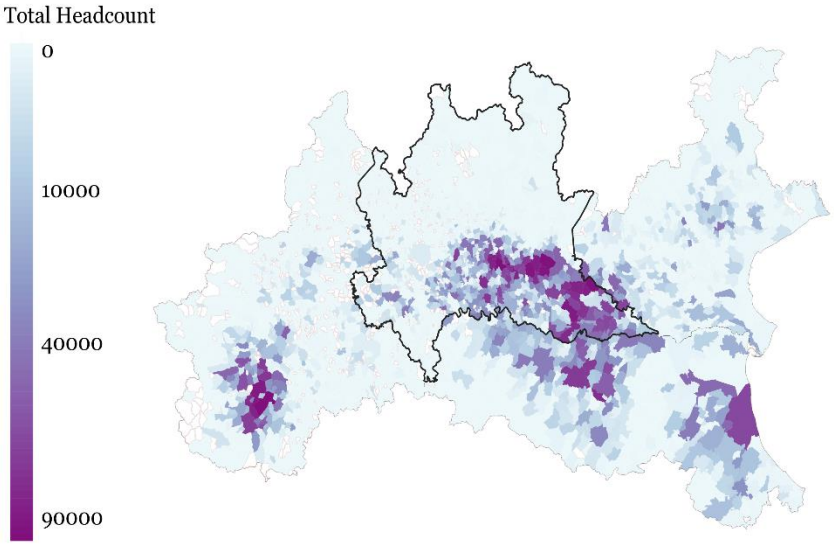
Notes: the figure reports quadrant-specific wind frequency at the station level, calculated as the number of days recording wind flowing from a given quadrant over the entire sample period (2015-2020). Triangles mark sensors that provide decomposed data on ammonium nitrates and ammonium sulfates. Panel A plots wind frequencies for ammonia stations, while Panel B plots the same statistics for PM_{10} stations. Polar plots are reported at the bottom to visualize the mean concentrations of each pollutant of each combination of wind direction and speed at the regional level. These plots are obtained using the function polarPlot in R. Computational details for calculating the concentration surface can be found in Carslaw et al. (2006) and Westmoreland et al. (2007).

Figure 5: Animals total headcount

[A] - Cattle



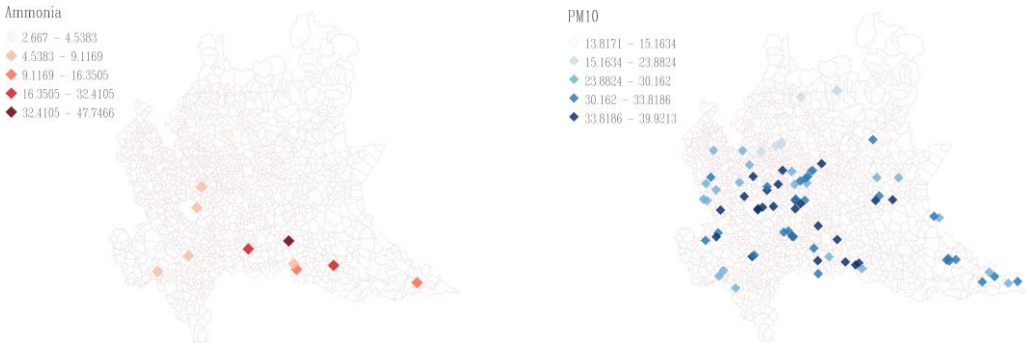
[B] - Swine



Notes: the figure reports the time average total headcount of cattle (Panel A) and swine (Panel B) at the municipal level across four regions: Lombardy (borders in bold), Piedmont, Emilia-Romagna, and Veneto. The region's area covers all municipalities located within a 60km radius of at least one NH3 or PM station.

Figure 6: Pollutants concentration - sensor sample average

[A] NH_3 [B] PM_{10}

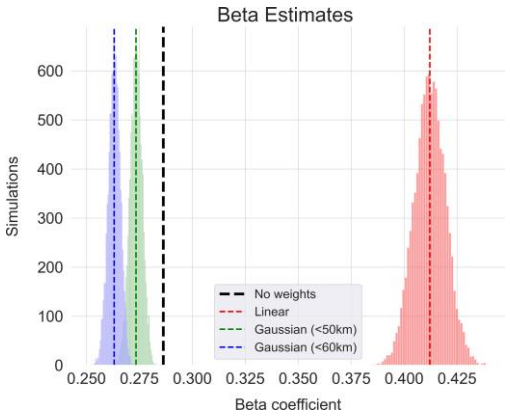


Notes: the figure plots ammonia (Panel A) and PM stations in Lombardy. Sensor color is determined by average daily concentration ($\mu g/m^3$) throughout the year at the sensor level. Max - Min values: [2.7; 47.7] Panel A; [13.8; 39.9] Panel B.

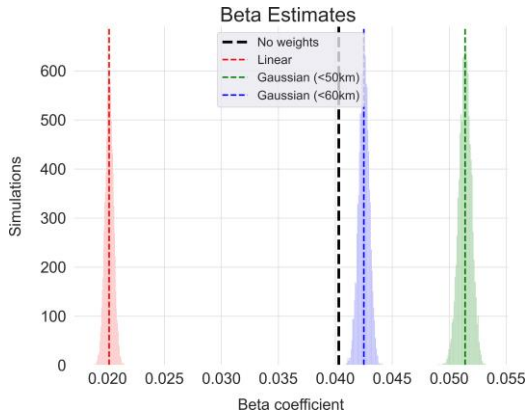
Figure 7: Distributions of simulated weighted variation in livestock units (quadrant)

[I] NH_3

[A] Cattle

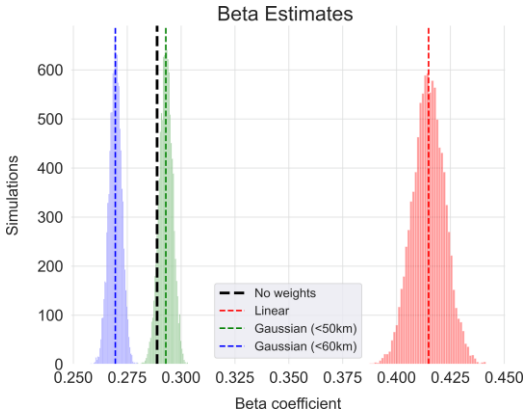


[B] Swine

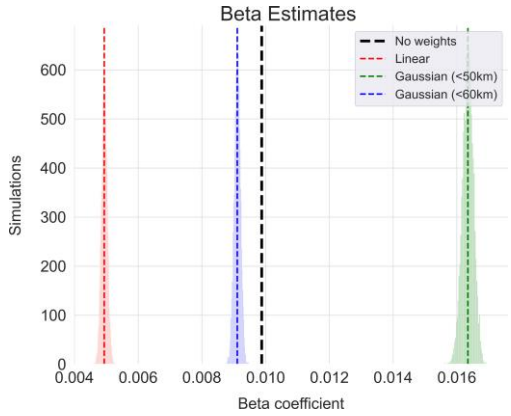


[I] PM_{10}

[A] Cattle

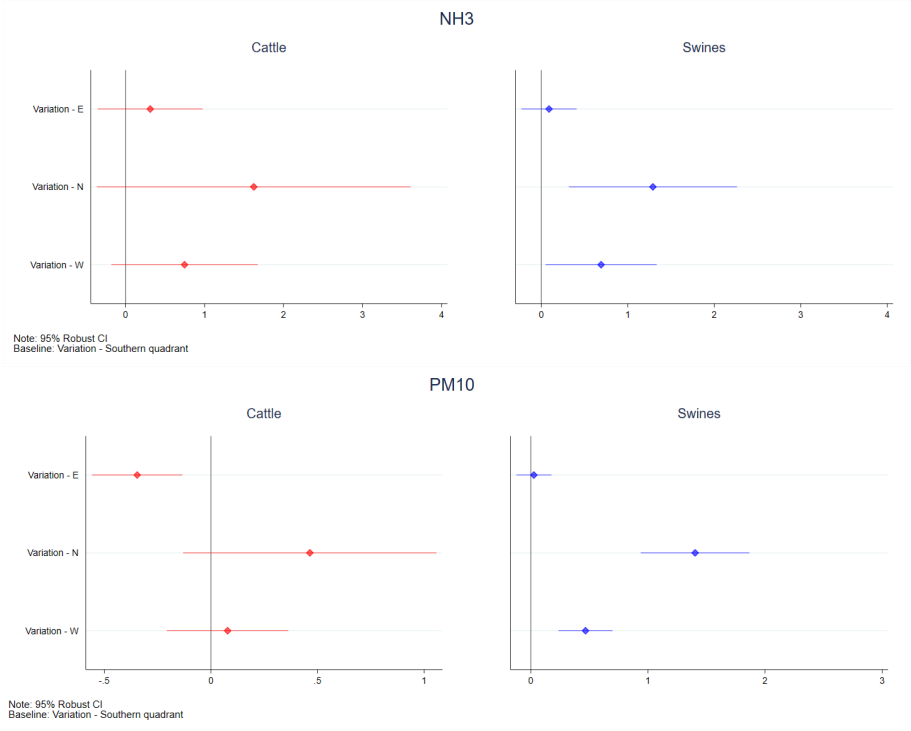


[B] Swine



Notes: the figure compares the marginal contribution of a 1000-unit positive variation estimated without distance discounting weighting with that obtained through different specifications of $\mathbf{\Omega}$. Estimates are presented separately by pollutant (Panels I and II) and farming animal (Panels A and B). Coefficients are estimated according to Equation 1, while $\mathbf{\Omega}$ weights are computed according to Equations 2 and 3. The resulting effect plotted in the graph is obtained by multiplying point estimates (See Appendix, Tables A1.1 through A1.4) and the simulated 1000-unit variation distribution.

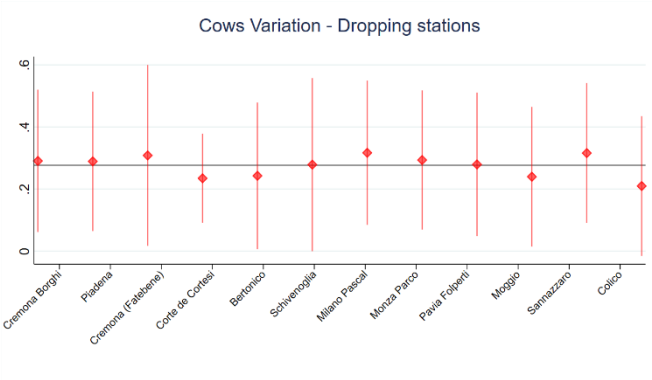
Figure 8: Effect heterogeneity - Wind direction



Notes: The table reports the estimates of η_a coefficients from Equation 4. The control group is the variation in livestock units taking place in the quadrant South of each sensor. Weather controls (temperature, wind direction, wind speed, rainfall, radiance, humidity, average PBLH, interacted with each other up to three lags) and month, year, and station-by-quadrant fixed effects are included. Robust confidence intervals at 95% are plotted.

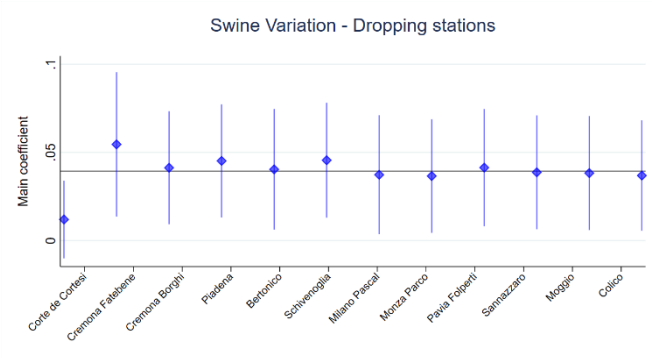
Figure 9: Effect heterogeneity - Dropping NH_3 stations

[A] – Cattle



Station	*Cattle (Area avg.)*
Cremona Borghi	1297819
Piadena	1265482
Cremona Fatebene	1053679
Corte de Cortesi	1030262
Bertonico	877854
Schivenoglia	712969
Milano Pascal	402840
Monza Parco	395167
Pavia Folperti	365971
Moggio	203124

[B] – Swine

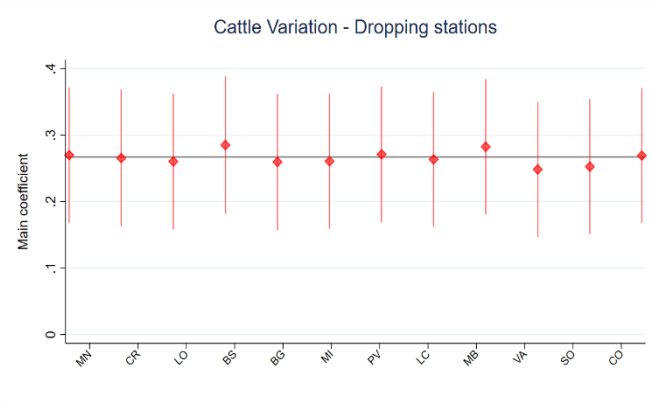


Station	*Swine (Area avg.)*
Corte de Cortesi	3954152
Cremona Fatebene	3779251
Cremona Borghi	3690717
Piadena	3633510
Bertonico	2946117
Schivenoglia	1967174
Milano Pascal	989238
Monza Parco	929888
Pavia Folperti	828486
Sannazzaro	433970

Notes: The figure plots the estimates of β_a from Equation 1, with $\mathbf{\Omega} = \mathrm{I}$, when observations from the sensor reported on the horizontal axis are excluded from the sample. Horizontal lines in Panel A and B correspond to the coefficients estimated in Table 3, Column 3. In the table, the sample average number of animals per station circular area is reported.

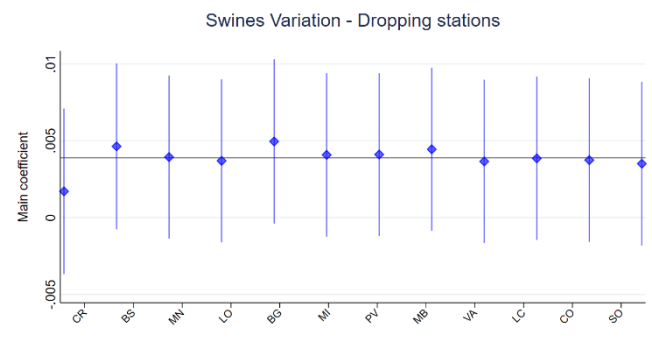
Figure 10: Effect heterogeneity - Dropping PM_{10} stations

[A] – Cattle



Station	*Cattle (Area avg.)*
MN	49498.51
CR	38685.26
LO	30789.42
BS	27381.61
BG	23418.84
MI	12662.75
PV	8613.302
LC	4849.589
MB	4355.362
VA	4065.111

[B] – Swine



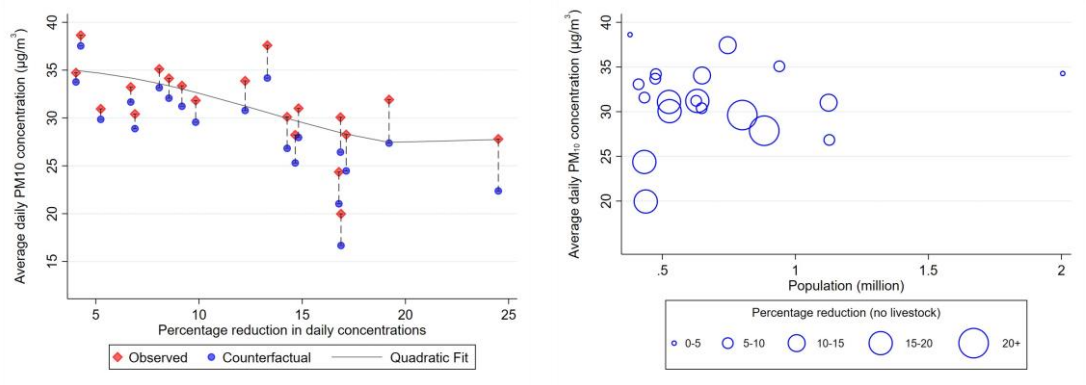
Station	*Swine (Area avg.)*
MN	181736.6
CR	120776.5
BS	109817.4
LO	101199.3
BG	80343.15
MI	33902.45
PV	20898.13
MB	10134.75
LC	8952.52
VA	6564.874

Notes: The figure plots the estimates of β_a from Equation 1, with $\mathbf{\Omega = I}$, where observations in the province reported on the horizontal axis are excluded from the sample. Horizontal lines in Panel A and B correspond to the coefficients estimated in Table 3, Column 3. In the table, the sample average number of animals per station circular area is reported.

Figure 11: Counterfactual PM levels and population exposure

[A] Reduction in PM levels

[B] Population affected



Notes: the figure shows the counterfactual scenario simulating the absence of bovine and swine livestock units. For visual purposes, sensors are grouped into twenty bins, calculated conditioning on yearly average concentrations. Panel A shows the relationship between average daily concentration and the corresponding percentage daily average reduction in each bin in the absence of swine and cattle. Panel B relates reduction under the counterfactual scenario with the population residing within a 50km radius of a station. Marker's size varies with the calculated percentage reduction in PM_{10} in the absence of livestock units.

Appendix

A1.Appendix Tables

Table A1.1: Estimates Robustness - Probability weighting

	NH_3			PM_{10}			PM_{10}^{ASN}		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Δ - Cattle	0.289**		0.201	0.351***		0.392**	0.069		0.128
	(0.126)		(0.133)	(0.060)		(0.060)	(0.144)		(0.150)
Δ - Swine		0.060***	0.0596***		0.002	0.0089***		0.020	0.0218
		(0.018)	(0.018)		(0.003)	(0.003)		(0.020)	(0.020)
Observations	16579	13919	13919	109202	109650	109650	3299	2790	2790
Adj R^2	0.5690	0.5656	0.5657	0.5215	0.5213	0.5217	0.5228	0.5367	0.5370
Dep. Var. Mean	15.53	15.53	15.53	30.42	30.42	30.42	10.68	10.68	10.68
Weather Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y
Month FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Sensor-by-quadrant FE	Y	Y	Y	Y	Y	Y	Y	Y	Y

Notes: the table reports the estimates of β_a from Equation 1, where $\mathbf{\Omega}$ is an identity matrix (absence of distance weighting). Analytical weighting assigning greater importance to sensor-day observations toward the end of each month is applied. Weather controls include temperature, wind direction, wind speed, rainfall, radiance, humidity, and average PBLH, interacted with each other up to three lags. Robust standard errors are reported in parentheses.

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table A1.2: Estimates Robustness - Linear Distance Weighting

	NH_3			PM_{10}			PM_{10}^{ASN}		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Δ - Cattle	0.975***		0.824**	0.738***		0.830***	0.38		0.345
	(0.306)		(0.322)	(0.128)		(0.128)	(0.42)		(0.45)
Δ - Swine		0.062*	0.0667*		0.062*	0.0212***		0.062*	0.0290
		(0.037)	(0.037)		(0.037)	(0.037)		(0.04)	(0.04)
Observations	16579	13919	13919	109202	109650	109650	3299	2790	2790
Adj R^2	0.5768	0.5693	0.5698	0.5145	0.5143	0.5146	0.5061	0.5109	0.5113
Dep. Var. Mean	15.53	15.53	15.53	30.42	30.42	30.42	10.68	10.68	10.68
Weather Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y
Month FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Sensor-by-quadrant FE	Y	Y	Y	Y	Y	Y	Y	Y	Y

Notes: the table reports the estimates of β_a from Equation 1, where $\mathbf{\Omega}$ is populated using linear weights. Weather controls include temperature, wind direction, wind speed, rainfall, radiance, humidity, and average PBLH, interacted with each other up to three lags. Robust standard errors are reported in parentheses.

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table A1.3: Estimates Robustness - Gaussian (<50) Distance Weighting

	NH_3			PM_{10}			PM_{10}^{ASN}		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Δ - Cattle	0.418***		0.355**	0.328***		0.380***	0.15		0.174
	(0.137)		(0.144)	(0.066)		(0.066)	(0.17)		(0.18)
Δ - Swine		0.054***	0.0552***		0.004	0.0118***		0.018	0.0188
		(0.020)	(0.020)		(0.003)	(0.004)		(0.02)	(0.02)
Observations	16579	13919	13919	109202	109650	13919	3299	2790	13919
Adj R^2	0.5767	0.5694	0.5698	0.5145	0.5143	0.5146	0.5061	0.5110	0.5114
Dep. Var. Mean	15.53	15.53	15.53	30.42	30.42	30.42	10.68	10.68	10.68
Weather Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y
Month FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Sensor-by-quadrant FE	Y	Y	Y	Y	Y	Y	Y	Y	Y

Notes: the table reports the estimates of β_a from Equation 1, where $\mathbf{\Omega}$ is populated using Gaussian weights, with maximum radius 50km. Weather controls include temperature, wind direction, wind speed, rainfall, radiance, humidity, and average PBLH, interacted with each other up to three lags. Robust standard errors are reported in parentheses.

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table A1.4: Estimates Robustness - Gaussian (<60) Distance Weighting

	NH_3			PM_{10}			PM_{10}^{ASN}		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Δ - Cattle	0.397***		0.341**	0.301***		0.350***	0.14		0.171
	(0.127)		(0.133)	(0.060)		(0.060)	(0.16)		(0.17)
Δ - Swine		0.046**	0.0464***		0.004	0.0113***		0.017	0.0180
		(0.019)	(0.019)		(0.003)	(0.003)		(0.02)	(0.02)
Observations	16579	13919	13919	109202	109650	109650	3299	2790	2790
Adj R^2	0.5767	0.5694	0.5699	0.5145	0.5143	0.5146	0.5061	0.5110	0.5114
Dep. Var. Mean	15.53	15.53	15.53	30.42	30.42	30.42	10.68	10.68	10.68
Weather Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y
Month FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Sensor-by-quadrant FE	Y	Y	Y	Y	Y	Y	Y	Y	Y

Notes: the table reports the estimates of β_a from Equation 1, where $\mathbf{\Omega}$ is populated using Gaussian weights, with maximum radius 60km. Weather controls include temperature, wind direction, wind speed, rainfall, radiance, humidity, and average PBLH, interacted with each other up to three lags. Robust standard errors are reported in parentheses.

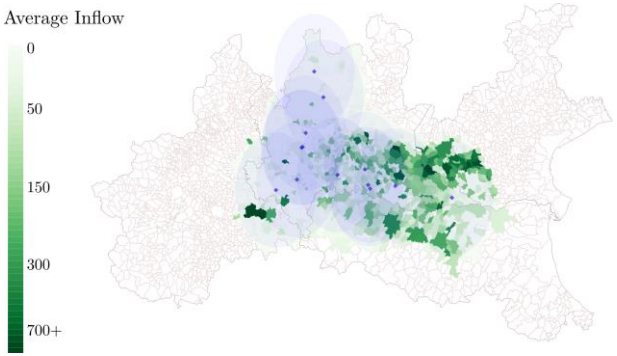
$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

A2. Appendix Figures

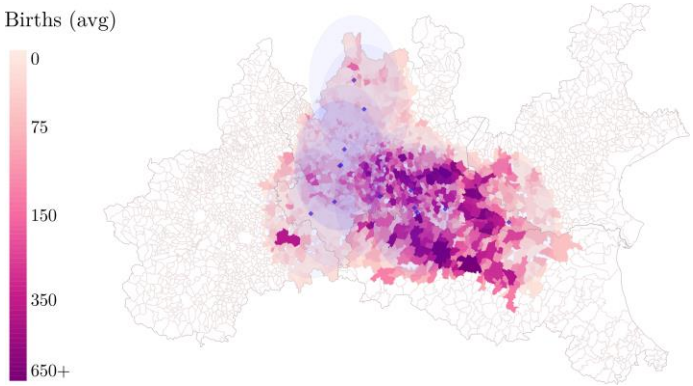
Figure A2.1: Livestock movements around stations - Sample averages (2015-2020)

[I] Cattle

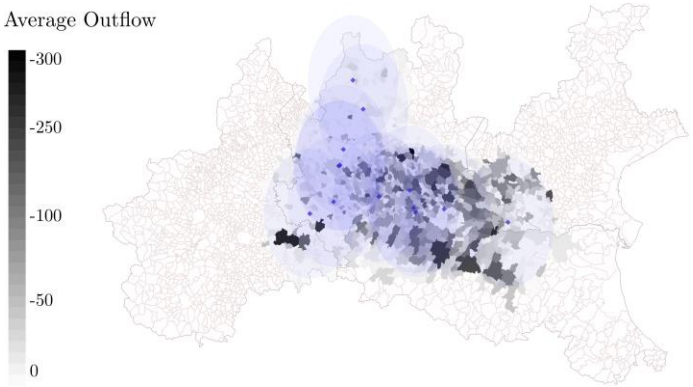
[A] Inflow



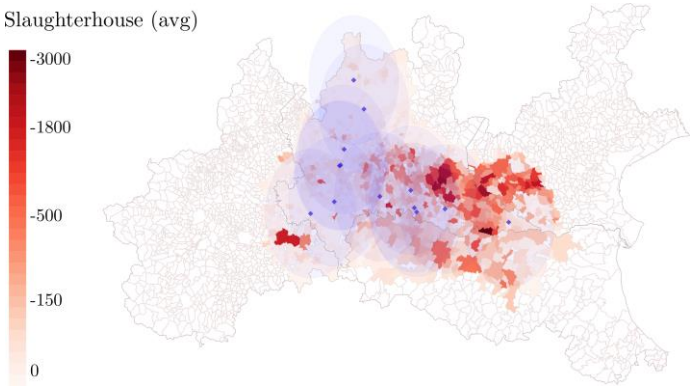
[B] Births



[C] Outflow (alive)

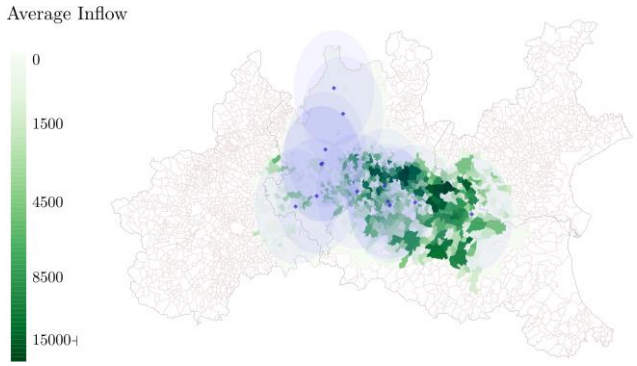


[D] Outflow (slaughter)

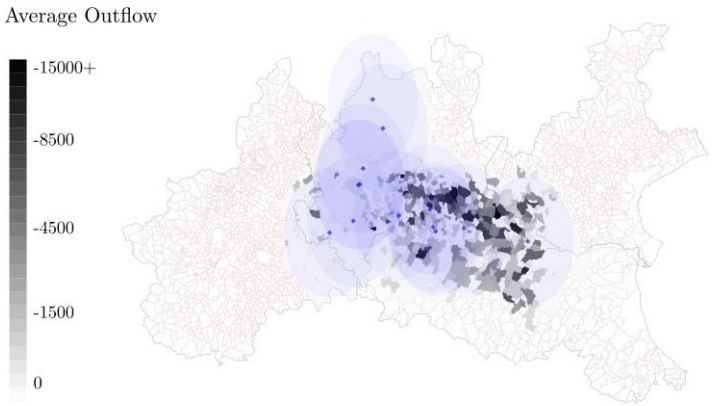


[II] Swine

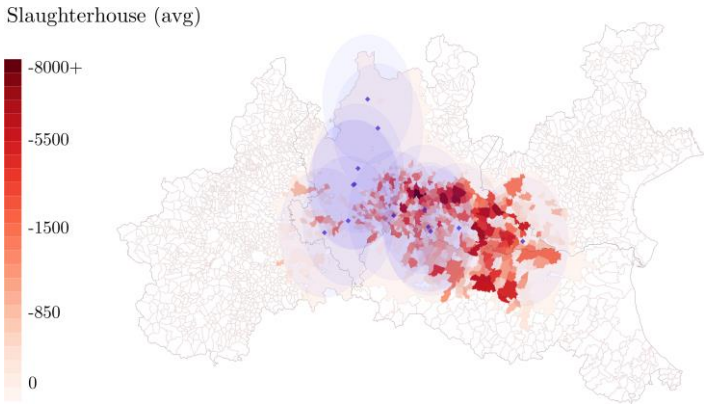
[A] Inflow



[C] Outflow (alive)



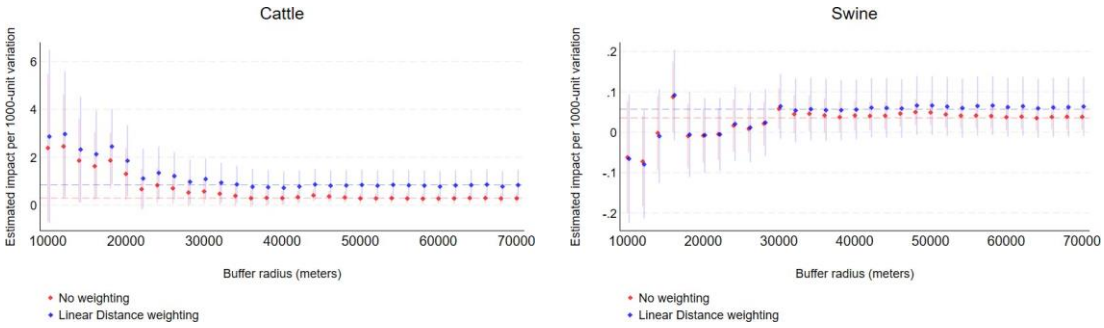
[D] Outflow (slaughter)



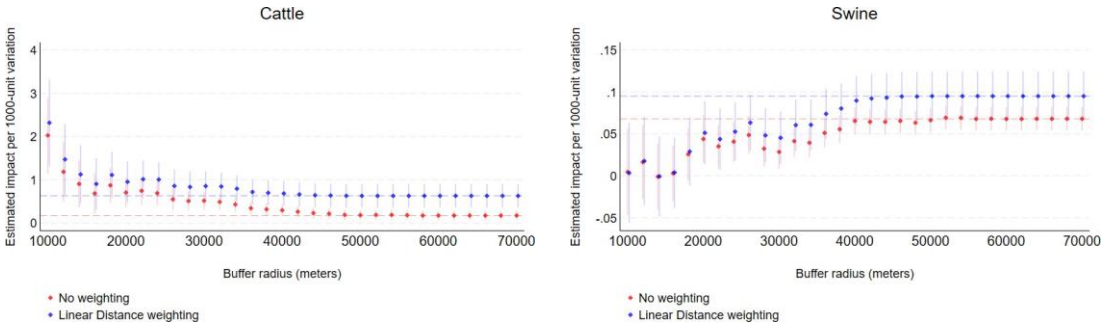
Notes: the figure plots the average monthly inflows (A), births (B), outflows (C), and slaughtered units (D) of cattle [I] and swine [II] throughout the sample period. Birth data is not separable from the overall inflow of swine. Animals displaced for slaughtering purposes are considered as an immediate depletion of the municipality's stock. Blue circles represent circular areas around sensors which includes municipalities within a 50km radius of each station. For brevity, the figure is only plotted for NH_3 stations.

Figure A2.2: Estimates sensitivity to expanding circular areas

[A] NH_3



[B] PM_{10}

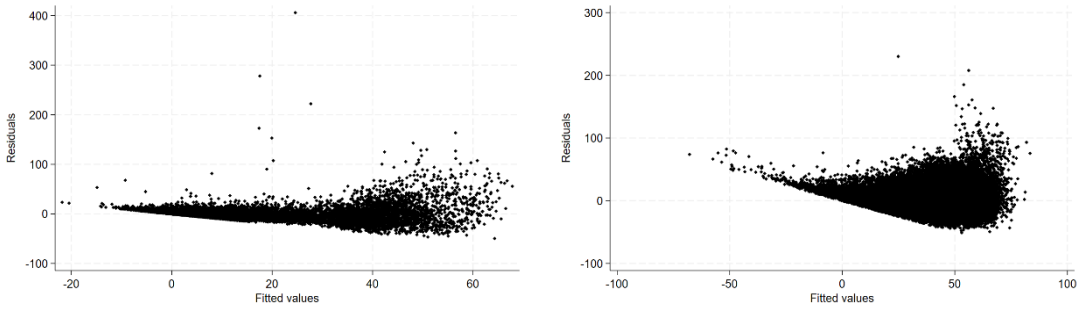


Notes: the figure shows the estimates of β_a from Equation 1, estimated using increasing values of $\bar{r}$, at 2km increment. Hence, only the variation in livestock units within $\bar{r}$ distance from the sensor is used to explain variation in pollutant concentrations. Weather controls (temperature, wind direction, wind speed, rainfall, radiance, humidity, average PBLH, interacted with each other and up to three lags), month, year, and sensor-by-quadrant fixed effects are included. Robust standard errors are reported.

Figure A2.3: Model residuals and fitted values

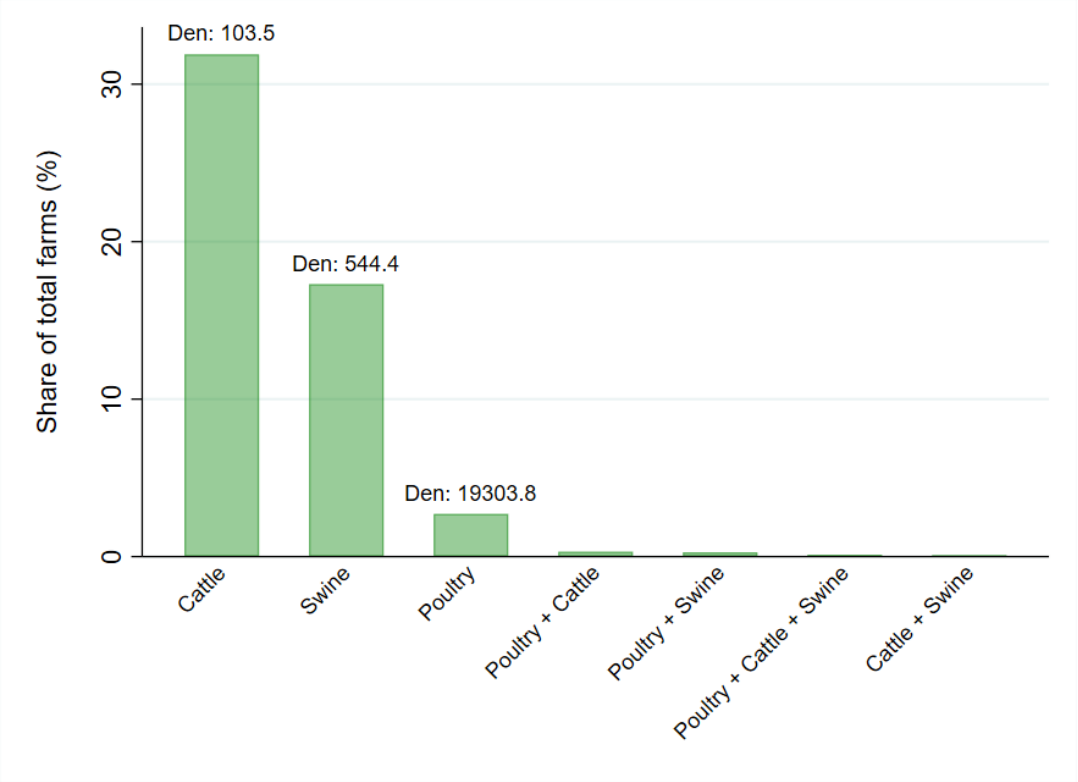
[I] – Ammonia

[II] – PM_{10}



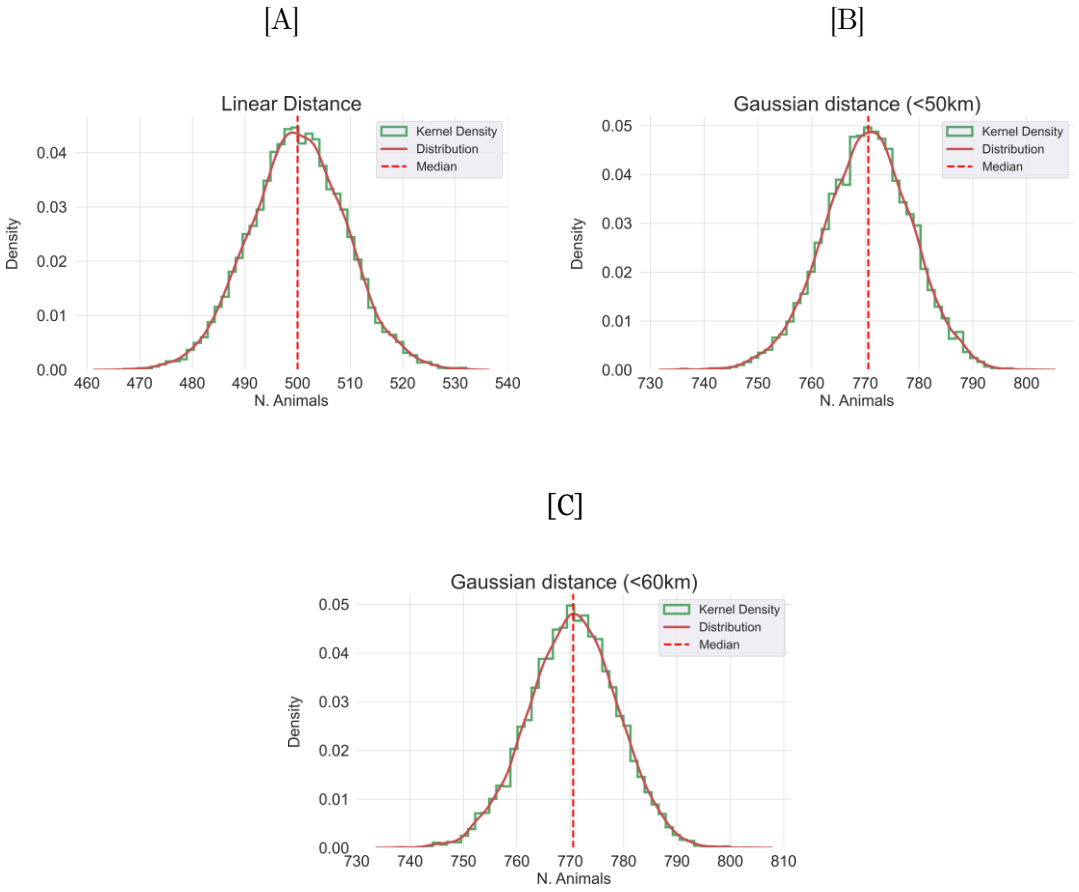
Notes: the figure plots the residuals of our model versus fitted values. Both variation in cattle ande swine units is included in absence of weighting (corresponding to columns 3 and 6 in Table 3), and the plot is reported for ammonia (Panel I) and PM_{10} (Panel II). The plots show a rather linear trend in residuals, with increased variance towards the right end of the fitted values distribution, supporting the correction for heteroskedasticity in our standard errors.

Figure A2.4: Multi-animal farming incidence



Notes: the figure reports the share of farms in the Lombardy region specializing in each combination of the most prevalent farming animals: cattle, pigs, and chicken.
Source: ISTAT, 2010 Agricultural Census.

Figure A2.5: Distributions of simulated weighted variation in livestock units (quadrant)



Notes: the figure reports the resulting distribution of a 10,000 iterations simulation of $\Delta L \times \mathbf{\Omega}$, where ΔL is a 1000-unit positive variation around a station. A unit is located at random distance $\tilde{d} \sim U(0, \bar{r})$. It is then weighted through $\mathbf{\Omega}$ according to three different specifications: linear (A), Gaussian <50km (B), Gaussian <60km (C). The resulting headcount distribution, corresponding kernel density, and median outcome are plotted.

A3. Supplementary Material

A3.1. Computing variation in livestock units: numerical example

We present in this Section a numerical example aimed at clarifying how the variation in livestock units, in case of either distance or unit weighting, is computed in each period. Consider a simple example with only one station and $Q = 2$ quadrants (North and South). Let there be one animal of interest such that A is a singleton and β_a is a singleton matrix. In addition, assume that quadrants only contains two municipalities each, and for simplicity, assume a symmetric scenario in which municipalities respectively at 25km and 10km from the sensor in each quadrant. Consider a maximum radius for the area surrounding a sensor of $r = 50$ km.
Let ΔL be a $Q \times M$ matrix, where M is the number of municipalities within distance r from the sensor, defined at time t as follows:

$$\Delta L = \begin{bmatrix} +3 & -4 & 0 & 0 \\ 0 & 0 & +1 & -2 \end{bmatrix}$$

where positive entries indicate an increase in the headcount of farming animals in municipality j in quadrant i. First, consider the case of absence of distance weighting: variation in animals within distance r from the sensor contributes equally to the concentrations of pollutants recorded at sensor level. In this case, the matrix $\mathbf{\Omega}$ will be an $M \times M$ identity matrix, which leaves ΔL unchanged. Second, consider linear weighting of variation in livestock units (the case of Gaussian weights is easily derived).

$$w_j = 1 - \frac{d_j}{\bar{r}}$$

This implies weights of respectively .5 and .8 for municipalities 25 km and 10 km away from the sensor. In this case, $\mathbf{\Omega}$ is a diagonal matrix expressed as follows:

$$\mathbf{\Omega} = \begin{bmatrix} .5 & 0 & 0 & 0 \\ 0 & .8 & 0 & 0 \\ 0 & 0 & .5 & 0 \\ 0 & 0 & 0 & .8 \end{bmatrix}$$

Which leads to a weighted variation matrix of the form:

$$\Delta L \times \boldsymbol{\Omega} = \begin{bmatrix} +1.5 & -3.2 & 0 & 0 \\ 0 & 0 & +.5 & -1.6 \end{bmatrix}$$

Assume that time t is characterized by North wind, such that the set $\mathcal{B}$ only includes municipalities in the first row of ΔL. It descends that the sensor will be imputed a variation in number of livestock units at time t of:

$$\sum_{j \in \mathcal{B} \cap \mathcal{G}} \Delta L_j \times \boldsymbol{\Omega} = +1.5 - 3.2 = -1.7$$

Starting from this example, it is possible to generalize for N stations, A animals, and Q quadrants.

Soil Aridification, Precipitations, and Infant Health: Evidence from Africa

Jacopo Lunghi[†], Maurizio Malpede[‡], and Marco Percoco[⋆]

[†] Bocconi University, Milan, Italy

[‡] University of Verona and GREEN

[⋆]Bocconi University and GREEN

This version: May 15, 2024

Abstract

While research has studied the effects of climate variations on child nutrition, how soil aridification impacts child wellbeing is relatively understudied. Using climate and infant health data combined with an original measure of soil aridity, we show that infants born in arid areas are comparatively more likely to die under the age of 5 and be systematically underweight at birth. We also find that aridification deteriorates the positive association between rainfall and child health. Our findings emphasize the importance of accounting for aridity alongside precipitations when assessing the economic impact of climate.

Keywords: Rainfall, climate change, potential evapotranspiration, child mortality, infant health

JEL Classification: J1, J13, I15, Q54, Q56, O15

1 Introduction

The effects of climate change and global warming have been central issues in the recent scientific and economic literature. Studies have found strong evidence of a close connection between the evolution of weather factors and economic welfare, both at the micro and macro level (Dell et al. 2008, 2012, Burke et al. 2015, Zhang et al. 2017, Peri & Sasahara 2019).

Among the numerous implications of climate change, water access for people worldwide remains a key concern. A reliable water source is crucial for developing sustainable human settlements and agricultural and economic systems, and water availability is strongly influenced by precipitation and its intensity. According to current predictions, global rainfall is expected to rise in the next few decades due to climate change. Panel [A] in Figure 1 illustrates the upcoming trend in rainfall using five of the most commonly used earth system models (ESM) for the periods 2040–2079 and 2060–2099: all the models predict a consistent increase in global precipitation levels, roughly between 5.7% and 12.2%.

Most of the economic literature has examined the effects of precipitation and temperatures as key determinants of the amount of water available to individuals in developing countries. For instance, Maccini & Yang (2009) use early-life precipitations in Indonesia to show how fluctuations in agricultural output have a long-lasting impact on women's health, schooling, and socio-economic status. Kudamatsu et al. (2012) used rainfall and temperature to determine how a period of relative drought affects infant mortality in Africa, finding that in-utero drought exposure increases babies' risk of death. In both studies, nutrition and malnutrition (experienced either by a mother or an infant during early life) constitute the primary channel through which climate conditions affect individuals. Conversely, Jayachandran (2006) used rainfall shocks in underdeveloped rural India to explain the impact of productivity shocks to agriculture on wages, migration, and credit constraints.

Thiede & Strube (2020) found that high temperatures and lower precipitation are associated with reductions in child weight and an increase in the risk of wasting. Similar results were obtained using the En Nino Southern Oscillations (Anttila-Hughes et al. 2021). More recently, Blom et al. (2022) used hour-degree bins of temperature exposure to assess

the effects of extreme heat on early child nutrition, finding that a $2°C$ rise in temperature increases the prevalence of stunting by more than seven percentage points.

Based on these findings, rainfall is viewed to have a positive impact on human development in rural settings, primarily through its beneficial effect on the agricultural sector, while temperature negatively impacts human health and development (McMahon & Gray 2021). Nonetheless, precipitation and temperature alone do not capture actual soil water availability, which also depends on concurring factors, such as land quality, solar radiation, and wind speed. A direct consequence of these interactions is that as climate change accelerates and temperatures rise globally, the stability of the relationship between precipitations and water availability may change dramatically, with substantial effects on human wellbeing. Young (2021) argues that water access and the actual water availability of the soil are key for more effective food and nutrition policies in Africa.

In this paper, we aim to explore the capacity of the soil to retain water and its effect on child health in Africa. To this end, we combine data on temperature and rainfall defined on 0.5° latitude x 0.5° longitude grids (ca. 56 km x 56 km at the equator) with an original measure of soil aridification, namely the potential evapotranspiration (henceforth PET) index. PET considers the combination of two sources of soil water loss, namely the soil surface evaporation (i.e., the process whereby liquid water is converted to water vapor and removed from the evaporating surface) and reference crop transpiration (the vaporization of liquid water contained in plant tissues and the vapor removal to the atmosphere; Allen et al. 1998). Following environmental research on the topic, we consider the PET as an indicator of the aridity of an area (Rind et al. 1990, Cherlet et al. 2018) and study the impact of our measures on infant wellbeing in Africa (Figure 1, Panel [B]).

Consistent with previous literature, this study focuses on the core indicators of infant health as a proxy for economic wellbeing (Aber et al. 1997, Benshaul-Tolonen 2018). We use geolocalized data from four waves of the Demographic and Health Surveys (DHS) to assemble information on women and children in 34 countries between 1992 and 2018, with a primary focus on child mortality. Based on a sub-sample of individuals, the study also documents the joint effect of rainfall and PET on body mass index (BMI) at the time of the interviews and in terms of their size at birth in children under five.

We find that precipitation alone is associated with higher child health measures. How-

ever, while this result is consistent with previous studies, our findings suggest that soil aridification counteracts the beneficial effects of higher rainfall levels: children in more arid regions are more likely to die before the age of five, tend to be smaller at birth, and have systematically lower BMI values by the time the interviews were conducted. The results are robust for redefining the primary outcome variables with comparable measures, such as one-year mortality and weight-to-height ratios. Moreover, the positive effect of rainfall on child health is shown to be susceptible to aridity and progressive land aridification: children in dry climates and those living in areas that have experienced a stronger increase in PET levels in the last decades benefit disproportionally less from years of abundant rainfall.

Finally, we address the role of migration and assess the heterogenous impacts of soil aridity in rural and urban communities. Those findings show that individuals living in urban areas are still susceptible to negative effects of soil aridification.

Our research contributes to the debate on the relationship between precipitations and human wellbeing, adding to the discussion on the effect of rainfall on health during childbearing and early life (Kudamatsu et al. 2012, Le & Nguyen 2021, Sivadasan & Xu 2021, Ponnusamy 2022). Moreover, with rainfall being associated with a variety of socioeconomic outcomes (Rose 1999, Barrios et al. 2010, Brückner & Ciccone 2011, Björkman-Nyqvist 2013, Rocha & Soares 2015), the paper offers a new perspective to the literature about heterogeneity in the response to rainfall shocks (Sarsons 2015, Damania 2020, Mary 2022) and, more in general, unexpected weather events (Burgess & Donaldson 2010, Dell et al. 2014).

The paper also contributes to the discussion about the present and future impact of climate change on low-income countries, providing new evidence on the relationship between climatic factors, soil aridity, and child health (Rabassa et al. 2014, Bharadwaj et al. 2020, Emediegwu et al. 2022, Randell et al. 2021).

The remainder of the paper is organized as follows: Section 2 summarizes the debate on the future of precipitations and presents the data. In Section 3, we propose a research design to study the impact of weather conditions on child health. In Section 4, we discuss our main findings, and in Section 5, we complement the evidence by testing for potential urban-rural heterogeneity and assessing robustness to potentially endogenous migration.

Finally, Section 6 concludes with a simple back-of-the-envelope calculation to quantify the expected toll of aridification on children's wellbeing.

2 Background and Data Description

2.1 Child Health and Soil Aridity

Climate change has emerged as a significant factor affecting both the economy and society, as evidenced by recent studies (Lanfredi et al. 2022, Cattaneo & Foreman 2023). However, it is children, particularly those living in low- and middle-income countries, who are among the most vulnerable (Burke et al. 2015). Newborns are uniquely susceptible to the health implications of climate change due to their weakened immune systems, higher metabolic rates, and increased exposure to environmental hazards. Moreover, their survival and well-being rely on their caregivers, who may be adversely affected by the disruptions and stresses brought about by climate change (Chersich et al. 2018).

Among the many aspects of climate change, the issue of aridification is a substantial challenge that the world faces today, as reported by the Intergovernmental Panel on Climate Change (IPCC) in 2019. Soil aridification can affect food security and nutrition through agriculture, pushing food prices and disrupting food distribution, leading to chronic undernutrition and micronutrient deficiencies among children, which can impede their growth, development, and cognitive abilities. In related literature, soil evapotranspiration has been shown to influence crop productivity and lead to a sensitive loss in agricultural yield, depending on the characteristics of the cultivated land (Mendelsohn & Dinar 2003, Mendelsohn 2009, Malpede & Percoco 2023). In the African continent, where a comparatively high share of the population lives in rural areas, agriculture has historically accounted for more than 50% of the gross domestic product (GDP) until the beginning of the 1990s (Diao et al. 2007). Even in 2020, in 34 African countries, agriculture, forestry, and fishing accounted for an average of 21.4% of the GDP, with some countries even exceeding 50%.[1] As such, in an area where agriculture is vital, aridity could affect human development by impacting agricultural productivity.

[1]Source: World Development Indicators, The World Bank Group: Agriculture, forestry, and fishing, value added (% of GDP). Data for Comoros, Cote d'Ivoire, and Mozambique are available from 2019. Data for Zimbabwe are available from 2018.

Finally, aridification could indirectly impact the physical well-being of children. Periods of climatic fluctuation and global environmental change have been associated with social phenomena such as conflict (Harari & Ferrara 2018) and migration (Black et al. 2011). Migratory flows, trauma, and violence resulting from climatic shocks can lead to anxiety, depression, post-traumatic stress disorder, or other psychological disorders in children and their caregivers (Crandon et al. 2022). Children may also experience reduced access to education, social support, and recreational opportunities (Vergunst & Berry 2022).

The link between desertification and the well-being of children is intricate and multifaceted. This study aims to highlight the existence of a robust causal connection between desertification and newborns' health and to reconsider the role of rainfall vis-à-vis changing climate while developing the ground for further research on the specific mechanisms through which these relationships unravel.

Furthermore, soil aridification can induce health concerns for individuals in Africa to the extent that less humid conditions can reduce natural habitats for vectors and their predators, altering the transmission patterns and geographic range of vector-borne diseases such as malaria, dengue, and yellow fever, which are already endemic in many parts of Africa. Soil aridification can also increase the incidence and severity of waterborne diseases such as cholera, typhoid, and dysentery, which are linked to poor sanitation and hygiene (Caminade et al. 2019).

2.2 Measures of Soil Aridity

The role of precipitation within the general pattern of climate change is still disputed. It was initially posited that climate change could lead to a wet-get-wetter and dry-get-drier pattern due to atmospheric moisture convergence and divergence (Held & Soden 2006). Nonetheless, due to a lack of consistent empirical evidence, this hypothesis has seemingly been replaced by a contrasting view in which mean precipitations are expected to increase at high- and mid-latitudes. Still, they will likely not decrease in subtropical regions (Kirtman et al. 2013, Donat et al. 2016). In turn, average rainfall at the global level is expected to increase in the coming years (Cherlet et al. 2018).

The evidence found in the existing literature suggests that higher precipitation levels

could help prevent drought and boost agricultural productivity (Gornall et al. 2010). As a result, it is possible that communities relying on agricultural output as a primary source of income and food supply may potentially benefit from foreseen weather variations. In this scenario, climate change could positively affect rural households in developing countries in the upcoming decades. In this paper, we study the direct effect of rainfall and soil evapotranspiration on a set of infant health measures. To perform this empirical assessment, we use the CRU TS4.04 dataset, which contains time-series data on month-by-month variations in climate over the period 1901–2019, provided on high-resolution (0.5° longitude x 0.5° latitude) grids.[2]

We focus on an area covering almost 40% of the entire African continent. Moreover, this paper focuses on weather conditions between 1951 and 2019 and considers three measures of climate variation: PET (mm/month), precipitations (mm/month), and monthly mean temperature (°C). The PET represents the amount of water lost from a cropped reference surface that is not short of water (a hypothetical grass reference crop with specific characteristics). As such, this measure estimates the evaporative demand of the atmosphere independently of crop type, crop development, and management practices. The PET estimates are calculated using a variant of the Penman-Monteith method, briefly summarized in the Appendix.

We report the evolution of our environmental measures over the sample period in Table 1. The yearly averages of each variable are included in time windows of 15 years each. The predictions regarding increasing rainfall cannot be confirmed using historical data and a limited sample period; however, the study's series points in that direction. Indeed, the level of average precipitations appears to follow a reverting trend, decreasing at first and then rising again in the early 2000s. Conversely, average PET levels and temperatures are on a stable, increasing path. Additionally, the process of aridification and rising temperatures seem to have accelerated since the 1980s.

Table 1 also reports correlations between the climate variables used in this research. To interpret cross-sectional variability across the grid, the average yearly precipitations and PET levels throughout the sample period are plotted in Figure 2. While the PET

[2]As is usually the case for model-computed weather data, the decision to use the CRU database comes with partial concerns regarding the quality of data. We justify the suitability of this dataset for our purpose in Section A.1 in the Appendix.

shows steady but little volatility over time, the sample retains substantial cross-sectional variation, ranging from areas with almost no evapotranspiration to cells where this measure exceeds the average precipitations.[3]

2.3 Measures of Infant Health

This study combines data on soil aridity with information on child health from the Demographic and Health Surveys (DHS). This program collects nationally representative data on health and population in developing countries, compensating for the lack of high-quality infant health statistics, particularly in Sub-Saharan Africa. In the present study, we construct a dataset using survey waves conducted between 1992 and 2018 for 34 African countries. The survey is stratified into clusters, localized with displacements of up to 2 km for urban and 10 km for rural points. The geographical distribution of clusters available in at least one wave is plotted in Figure 3. Data from the Individual Recode dataset and the Child Recode dataset are utilized. The former contains data on every eligible woman, including individual socio-economic characteristics, birth history, pregnancy, and postnatal care; the latter comprises some core child health indicators for children under five years old and their mothers. Here, we restricted the sample to women who experienced at least one completed pregnancy.

Table 2 summarizes the descriptive statistics on mothers and infant mortality across the sample period.[4] We obtained information for 468,873 women aged 13 to 50, corresponding to more than 2 million births between 1955 and 2019. However, wealth and education characteristics and records on size at birth and BMI were only available for a sub-sample of individuals, leaving us with a final sample of roughly 1.7 million births and over 400,000 children under five. Given the large number of countries and the variety of developing settings considered, the women in the sample are heterogeneous in terms of education and economic wealth scores. The average years of schooling are relatively low: the women in our sample attended school for less than 4 years, with almost 50% declaring having received no formal education. Moreover, roughly 70% to 75% of the sample comprises women living

[3] When the PET exceeds the actual precipitation, it indicates that the soil may eventually dry out unless irrigation is used to offset the loss. However, the effective amount of water dispersed also depends on the type of plants cultivated on the land.

[4] All the variables relevant to our study are described in Section A.1 in the Appendix.

in rural households. As extensively highlighted in previous literature, fertility appears to be high in the sample, with an average of more than five births per woman. These measures tend to remain homogeneous across different survey waves.

The first outcome variable is child mortality, retrieved from a woman's reported completed pregnancies. The variable is a binary indicator that takes a value of 1 if a woman reports that the child died within 60 months of the date of birth. This methodology of computing infant mortality has already been adopted in the literature employing DHS data (Kudamatsu et al. 2012, Benshaul-Tolonen 2018). As Table 2 indicates, despite a steady decrease, the average probability of a pregnancy resulting in a child's death is still very high, ranging from 8.9% to 12% across the survey waves. However, infant mortality is highly volatile across the sample, with large standard deviations common to all the rounds. Moreover, there seems to be only a negligible difference between the mortality rates of male and female children in the sample.

In addition to child mortality, we consider other infant health indicators to provide further evidence of the relationship discussed in this paper. Specifically, we consider the Body Mass Index (BMI), calculated using the new Child Growth Standards (CGS) from the World Health Organization (WHO) at the time of the interviews, and a categorical variable indicating whether the child's size at birth was below or above average.[5] These measures allow us to better understand whether aridity could also affect the health of those who survive. The BMI is provided for all newborns and children under five years old.

From Table 2, we notice that the infants in our sample consistently display below-average weight/height ratios (the average child falls in the left tail of the weight/height distribution, around the 40^{th} percentile). At the same time, regarding BMI, this only appears to occur towards the last waves of survey collection.

[5]We check the robustness of our findings using a slight variation in our outcome measures. We focus on five-year mortality and the ratio of weight over height, expressed in standard deviations and again calculated using the CGS method.

3 Methodological Approach

We start with a linear fixed-effects panel data model, exploiting annual variations in weather conditions across cells to identify the effects of rainfall and PET on the infant health indicators. We assembled an unbalanced panel dataset using the DHS to compare the impact of the PET and precipitations. The following equation is estimated:

$$Y_{i,m,q,c,t} = \sum_{r=0}^{1} \beta_{1,r} \mathrm{PRE}^{g}_{c,t^*-r} + \sum_{r=0}^{1} \beta_{2,r} \mathrm{PET}^{g}_{c,t^*-r} + \rho X_{i,m,c} + \sigma_{trend} + \phi_{q(t^*)} + \epsilon_{i,m,q,c,t} \tag{1}$$

where Y_{i,m,q,c,t^*} represents our measure of infant health for child i born from mother m in cluster q in cell c. $t \in \{t^*, \tilde{t}\}$ can represent either a child's year of birth (t^*) or the year of the interview ($\tilde{t}$). As the main regressand, we consider a measure of infant mortality (mortality at five years), the BMI in standard deviations recorded at the time of the interview, and the size at birth. We also assess the robustness of our results on a different yet comparable set of outcome variables (mortality at one year and the weight/height ratio).

$\mathrm{PRE}_{c,t}$ and $\mathrm{PET}_{c,t}$ are, respectively, the levels of precipitations and potential evapotranspiration at a child's year of birth (t^*) in cell c. In a variation of our specification, precipitation and PET levels are measured only using growing season months[6]. One-year lagged terms are included. Indeed, precipitation and PET levels may impact infant health by impinging on a mother's physical well-being or ability to take care of her child effectively. Thus, the study controls for weather conditions one year before childbirth, which more clearly isolates a potential effect on the mother that is not directly linked to the child's health.

In our baseline specification, the reference model for the binary outcome variable of infant mortality is a linear probability model (LPM). Using a linear estimator eases the interpretation of the coefficients and allows for immediate comparisons between the differ-

[6]Data on the growing season in each cell are available through the Global Agro-Ecological Zones (GAEZ) v4 database, provided by the Food and Agriculture Organization (FAO). The growing season is calculated using the beginning date of the earliest growing period and the total number of growing period days.

ent specifications. Moreover, this choice is preferable given the ample set of fixed effects included in the regression.[7]

$X_{i,m,c}$ includes a set of covariates at the birth, woman, and grid level. We control for characteristics that could be correlated to a child's status at the time of the survey while also being associated with PET. As such, we control for a woman's education, the wealth index of the household, the kid's month of birth, and the main source of water declared by the household[8]. We also control for cell temperature, to ensure that our measure of PET is not simply a proxy for temperatures. This approach allows for seasonality in precipitations and the PET to be smoothened out while accounting for infant health, which may be influenced by exposure to weather conditions both in utero and immediately after childbirth. Moreover, the study controls for a mother's education and wealth index, the household's main source of water supply, and, again, for temperature in the grid at time t for the same reason as above. In addition, $\phi_{c(t)}$ includes two sets of time and cluster fixed effects.[9] Lastly, the term σ_{trend} captures country trends.

Based on previous findings and the existing literature, precipitation is expected to be positively associated with child health. However, what happens when we include our soil aridity measure is an empirical question. To explore this question, we focus on the effect of precipitation *vis-à-vis* steadily increasing levels of soil water dispersion. To this end, we estimate the following regression:

$$Y_{i,m,q,c,t} = \sum_{r=0}^{1}\sum_{k=1}^{10} \gamma_{r,k} \mathrm{PRE}^{g}_{c,t^*-r} \times \mathbb{1}\left[\Delta\mathrm{PET}_{dec} = k\right]_c + \rho X_{i,m,c} + \sigma_{trend} + \phi_{q(t^*)} + \epsilon_{i,m,q,c,t} \tag{2}$$

where the term $\mathbb{1}\left[\Delta\mathrm{PET}_{dec} = k\right]$ indicates a binary indicator for the k-th decile in the sample distribution of ΔPET. The latter is a cell-specific measure that aims to capture the long-term variation in evapotranspiration. It is therefore calculated as the growth rate of

[7]While it is possible to estimate non-linear models through iterative algorithms, such as iteratively reweighted least squares (IRLS), some theoretical challenges exist regarding the existence of a solution to the algorithm or whether relevant parameters are identifiable.

[8]Controls are justified as follows. More educated and wealthier households may influence a child's wellbeing while also being present in areas with better soil conditions. Also, access to a water source may be correlated with soil quality and potentially affect childcare. Lastly, children born in different months may be susceptible to the effect of aridification in different phases of their early growth.

[9]Adding cluster FEs constitutes a more conservative approach than cell FEs, as a cell in our grid contains at least one cluster.

the 10-year average PET between 1950 and 1959 and 2010 and 2019. The grid distribution of the long-term variation in the PET (Figure 4) shows how systematically arid cells do not necessarily identify areas in which the increase in the PET has been relatively more pronounced in the last decades, with the south-east regions, in particular, experiencing a steep increase in soil aridity. As such, Equation 2 more precisely captures the role of soil aridification, and not simply the overall soil quality, in driving a heterogeneous response of individuals to rainfall.

4 Results

Table 3 reports the standardized estimates of the impacts of precipitations and the PET from Equation 1[10]. As in the previous sections, in Panel [A], the weather variables are computed over the entire year of reference, while in Panel [B], they are calculated based on the corresponding growing season. The PET and precipitation variables are standardized to ease comparison. Column (1) in Panel [A] shows the effect of precipitations on infant mortality within five years from birth. One standard deviation increase in yearly rainfall levels reduces mortality by 0.48 percentage points. The same increase in the year before a child's birth is also expected to reduce the probability of child mortality by 0.33 percentage points, suggesting that part of the effect may be realized during pregnancy. All the coefficients are significant at the 99% level.

When adding the PET into the regression [Column (2)], the effect of precipitations drops to 0.26% in the year of birth and 0.13% one year before, with the latter coefficient losing significance. Conversely, one standard deviation increase in the (lagged) PET increases the expected mortality by (0.33%) 0.53%, hinting at a detrimental effect of increased PET on child mortality, although this is never statistically significant. When only focusing on the growing season months (Panel [B]), the coefficients estimated are similar in magnitude, and the same dynamics in the impact of rainfall are detected. The effect of the PET in the year of birth is now significant at the 95% level.

A similar tendency is found in the coefficients estimated for BMI, expressed in 100 standard deviations [Columns (4) and (5)]. When the precipitations alone are included,

[10]Unstandardized coefficients are reported in Table A.4.

one standard deviation increase raises the BMI by 0.0746 standard deviations in the year of birth and 0.0331 standard deviations in the previous year. However, when the PET is brought into the regression, the combined effect of the two terms is almost halved. Conversely, PET appears to reduce BMI: one standard deviation increase in PET decreases BMI by 0.0961 standard deviations in the year of birth and by 0.1356 standard deviations in the previous year. The effect in the growing season months is relatively lower for both regressors, but signs and significance remain considerably stable. One standard deviation increase in rainfall one year before and in the year of birth is estimated to boost the size at birth (between a 0.0046 and 0.036 probability of scoring one category higher across Panels [A] and [B]), while an increase in PET has an opposite effect, which in most cases is stronger in magnitude[11].

Second, we turn to the results of Equation 2. We plot the $\gamma_{r,k}$ estimates in Figure 5. Existing trends in the coefficients of our decile interactions can inform about the evolution of rainfall's effect when considering diverging land aridification levels. In Panel [A], we notice how precipitations in the year of birth and one year before tend to reduce child mortality, especially in areas where the PET has been stable in the last decades. In contrast, the effect gets rapidly offset in cells with severe aridification. The downward trend in interaction coefficients for BMI depicts a similar scenario: the positive effect of precipitation on children's body mass tends to dissipate in cells where the PET has been rising more significantly in the last 60 years. The results for size at birth are less transparent: while a similar initial drop is observed in the coefficients, these exhibit some mild spikes toward the end of the PET distribution.

Overall, this evidence supports multiple considerations regarding the relationship between precipitation and the PET. First, the effects of the two climate factors often move in opposite directions. Second, whenever the PET is introduced in the analysis, the explanatory power of rainfall systematically decreases, which may hint that the role of rainfall may vary depending on the soil water retention ability. This is confirmed by noticing how the beneficial effect of increased rainfall before a child's birth dissipates significantly when

[11] In Table A.2, we show how the results observed for infant health are robust when substituting the outcome variables adopted with similar measures of health and mortality. Indeed, repeating the estimation with mortality within one year and the weight-to-height ratio in standard deviations returns remarkably similar coefficients, both in terms of magnitude and significance.

looking at areas that have experienced stronger aridification. Multiple mechanisms may coexist in generating this effect, including agricultural yields and income deriving from farming activities[12], climate-induced migration, and direct effects of aridity on health. We explore the presence of heterogeneity in our results based on some of these potential mechanisms.

With the agricultural sector a candidate channel directly or indirectly connecting it to individual wellbeing through climate, the adverse effects of soil aridification could be more directly relevant to individuals whose first income and food resources lie in the primary sector. As a result, the effect estimated with our empirical strategy could be strongly heterogeneous concerning this aspect, with relevant implications for future social and environmental policies. In fact, as individuals are expected to move away from rural areas (Cohen 2006, Güneralp et al. 2020), strong differences in the impact of soil aridification in urbanized cities could justify prioritizing measures of urban development over soil quality conservation in the short term.

To explore this scenario, we use the information provided by the DHS dataset, which classifies clusters as urban and rural. This classification is country-specific, and the assessment can consider multiple criteria, such as the availability of electricity or piped water and access to healthcare, schools, and transportation. We estimate Equation 1 separately for urban and rural clusters.[13] The results are reported in Table A.3. Despite the sample being almost equally divided between rural (60%) and urban clusters (40%), the higher number of children per woman in rural areas leads to more than 70% of births occurring in rural households. Looking at the coefficients for precipitation alone in the two panels, the effect of rainfall appears systematically stronger for individuals living in rural clusters. Still, it persists in sizable magnitude in the urban sample, with the only exception of size at birth.

In both the urban and rural clusters, soil aridification systematically counteracts the beneficial effects of higher rainfall despite having reduced statistical significance. These results reveal how individuals in urbanized areas are still susceptible to similar trends

[12]In a recent study, Malpede & Percoco (2023) show that the aridification process between 1990 and 2015 contributed to a decline in agricultural output of between 0.4 and 1.1 tons per hectare of cultivated cropland in Africa. Amare & Balana (2023) found similar effects focusing on Nigeria.

[13]For this exercise, we focus only on yearly values of precipitation and the PET. Using growing season definitions of our climate variables leads to comparable results.

due to aridification and reinforce the importance of tapering future soil quality loss. In rural regions, indeed, aridification can lead to reduced agricultural productivity due to soil degradation and water scarcity, which in turn affects food security and livelihoods (Vermeulen et al. 2012, Wheeler & Von Braun 2013). Urban areas, although they may have more infrastructure to deal with water scarcity, still face challenges such as increased demand for water resources and potential conflicts over these resources (Unfried et al. 2022). Furthermore, Masih et al. (2014) highlights how drought phenomena exacerbate existing social and economic inequalities, particularly affecting subsistence farmers.

Changes in climate conditions, including aridity and rainfall, may also induce people to move toward other regions based on some non-random characteristics. Migration, in turn, may have mixed implications. Migratory flows may drive the relocation of those suffering from a worse quality of life, creating selection of more resilient individuals into the sample of long-settled households. This, in turn, would result in an observed improvement in the levels of child health. The opposite might also be true: more dynamic and wealthy individuals may find it easier to relocate as opposed to those living in poorer conditions, inducing negative selection concerning the wellbeing of children from stayers households. To explore the presence of climate-induced migration underlying the effect observed in our estimates, we drop from the sample those women reporting with certainty having lived in the same place of residence for less than 15 years at the time of the interviews (*long-settled*). Observing significant discrepancies with the study's baseline estimates would support the hypothesis that individuals tend to react to climate change by relocating endogenously. Estimates of Equation 1 are then repeated for the sub-sample of long-settled individuals, who are assumed to be fully affected by the impact of aridification. The results reported in the Appendix (Table A.2) seem to alleviate this concern. The reduction in sample size translates into only a marginal loss of significance in the estimates, while the order of magnitude and the direction of the effect appears to be roughly unchanged. This strategy does not rule out the presence of climate-induced migration but rather suggests that more factors are at play and selective migration alone does not strike as a main channel to the health effects of aridity.

5 Conclusion

In this paper, we have studied the combined effect of rainfall and of the capacity of the soil to retain water on child health in Sub-Saharan Africa, finding that the evapotranspiration of the soil negatively affects infant health, reducing the size of children at birth and their BMIs. This result is particularly important when we consider that, since 1980, PET levels in our sample have increased by an average of 0.928 mm per year. Furthermore, most ESM projections foresee a peak in emissions, which will fuel aridification even further in the upcoming years: different representative pathway scenarios (RCPs) depict an increase in the yearly PET from a minimum of 30 mm/year to a maximum of over 300 mm/year.[14]

The results presented in this study call for policy actions to tackle soil aridification, which may generate benefits not only in terms of land productivity but also in terms of child health. New avenues for investigation may focus on detailing and thoroughly analyzing the potential transmission mechanism between rainfall, soil quality, and human development, which may help drive systematic and setting-specific interventions.

[14]Time series at $0.5^\circ \times 0.5^\circ$ grid resolutions are available for five ESMs: GFDL-ESM2M, HadGEM2-ES, IPSL-CM54-LR, MIROC-ESM-CHEM, and NorESM1-M. These are all part of the Coupled Model Intercomparison Project Phase 5 (CMIP5). We report the average projections limited to our grid in Table A.5. More information about the nature and specificities of these models can be found in Noce et al. (2020).

References

Aber, J. L., Bennett, N. G., Conley, D. C. & Li, J. (1997), 'The effects of poverty on child health and development', *Annual Review of Public Health* **18**(1), 463–483.

Allen, R. G., Pereira, L. S., Raes, D., Smith, M. et al. (1998), 'Crop evapotranspiration-guidelines for computing crop water requirements – FAO irrigation and drainage paper 56', *FAO, Rome* **300**(9), D05109.

Allen, R., Smith, M., Perrier, A. & Pereira, L. S. (1994), 'An update for the definition of reference evapotranspiration', *ICID bulletin* **43**(2), 1–34.

Amare, M. & Balana, B. (2023), 'Climate change, income sources, crop mix, and input use decisions: Evidence from nigeria', *Ecological Economics* **211**, 107892.

Anttila-Hughes, J. K., Jina, A. S. & McCord, G. C. (2021), 'Enso impacts child undernutrition in the global tropics', *Nature communications* **12**(1), 5785.

Barrios, S., Bertinelli, L. & Strobl, E. (2010), 'Trends in rainfall and economic growth in Africa: A neglected cause of the African growth tragedy', *The Review of Economics and Statistics* **92**(2), 350–366.

Benshaul-Tolonen, A. (2018), 'Local industrial shocks and infant mortality', *The Economic Journal* **129**(620), 1561–1592.
URL: *https://doi.org/10.1111/ecoj.12625*

Bharadwaj, P., Fenske, J., Kala, N. & Mirza, R. A. (2020), 'The green revolution and infant mortality in India', *Journal of Health Economics* **71**, 102314.

Björkman-Nyqvist, M. (2013), 'Income shocks and gender gaps in education: Evidence from Uganda', *Journal of Development Economics* **105**, 237–253.
URL: *https://www.sciencedirect.com/science/article/pii/S0304387813001120*

Black, R., Bennett, S. R., Thomas, S. M. & Beddington, J. R. (2011), 'Migration as adaptation', *Nature* **478**(7370), 447–449.

Blom, S., Ortiz-Bobea, A. & Hoddinott, J. (2022), 'Heat exposure and child nutrition: Evidence from west africa', *Journal of Environmental Economics and Management* **115**, 102698.

Brückner, M. & Ciccone, A. (2011), 'Rain and the democratic window of opportunity', *Econometrica* **79**(3), 923–947.

Burgess, R. & Donaldson, D. (2010), 'Can openness mitigate the effects of weather shocks? evidence from india's famine era', *American Economic Review* **100**(2), 449–453.

Burke, M., Hsiang, S. M. & Miguel, E. (2015), 'Global non-linear effect of temperature on economic production', *Nature* **527**(7577), 235–239.

Caminade, C., McIntyre, K. M. & Jones, A. E. (2019), 'Impact of recent and future climate change on vector-borne diseases', *Annals of the New York Academy of Sciences* **1436**(1), 157–173.

Cattaneo, C. & Foreman, T. (2023), 'Climate change, international migration, and interstate conflicts', *Ecological Economics* **211**, 107890.

Cherlet, M., Hutchinson, C., Reynolds, J., Hill, J., Sommer, S. & Von Maltitz, G. (2018), *World atlas of desertification: Rethinking land degradation and sustainable land management*, Publications Office of the European Union.

Chersich, M. F., Wright, C. Y., Venter, F., Rees, H., Scorgie, F. & Erasmus, B. (2018), 'Impacts of climate change on health and wellbeing in south africa', *International journal of environmental research and public health* **15**(9), 1884.

Cohen, B. (2006), 'Urbanization in developing countries: Current trends, future projections, and key challenges for sustainability', *Technology in Society* **28**(1–2), 63–80.

Couttenier, M. & Soubeyran, R. (2014), 'Drought and civil war in sub-saharan africa', *The Economic Journal* **124**(575), 201–244.

Crandon, T. J., Dey, C., Scott, J. G., Thomas, H. J., Ali, S. & Charlson, F. J. (2022), 'The clinical implications of climate change for mental health', *Nature Human Behaviour* **6**(11), 1474–1481.

Damania, R. (2020), 'The economics of water scarcity and variability', *Oxford Review of Economic Policy* **36**(1), 24–44.

Dell, M., Jones, B. F. & Olken, B. A. (2008), Climate change and economic growth: Evidence from the last half century, Technical report, National Bureau of Economic Research.

Dell, M., Jones, B. F. & Olken, B. A. (2012), 'Temperature shocks and economic growth: Evidence from the last half century', *American Economic Journal: Macroeconomics* **4**(3), 66–95.

Dell, M., Jones, B. F. & Olken, B. A. (2014), 'What do we learn from the weather? The new climate-economy literature', *Journal of Economic literature* **52**(3), 740–798.

Diao, X., Hazell, P. B., Resnick, D. & Thurlow, J. (2007), *The role of agriculture in development: Implications for Sub-Saharan Africa*, Vol. 153, Intl Food Policy Res Inst.

Donat, M. G., Lowry, A. L., Alexander, L. V., O'Gorman, P. A. & Maher, N. (2016), 'More extreme precipitation in the world's dry and wet regions', *Nature Climate Change* **6**(5), 508–513.

Ekström, M., Jones, P. D., Fowler, H. J., Lenderink, G., Buishand, T. A. & Conway, D. (2007), 'Regional climate model data used within the SWURVE project: Projected changes in seasonal patterns and estimation of pet', *Hydrology and Earth System Sciences* **11**(3), 1069–1083.

URL: *https://hess.copernicus.org/articles/11/1069/2007/*

Emediegwu, L. E., Wossink, A. & Hall, A. (2022), 'The impacts of climate change on agriculture in Sub-Saharan Africa: a spatial panel data approach', *World Development* **158**, 105967.

Fischer, G., Nachtergaele, F. O., van Velthuizen, H., Chiozza, F., Francheschini, G., Henry, M., Muchoney, D. & Tramberend, S. (2021), 'Global agro-ecological zones (GAEZ v4)–model documentation'.

Gornall, J., Betts, R., Burke, E., Clark, R., Camp, J., Willett, K. & Wiltshire, A. (2010), 'Implications of climate change for agricultural productivity in the early twenty-first century', *Philosophical Transactions of the Royal Society B: Biological Sciences* **365**(1554), 2973–2989.

Güneralp, B., Reba, M., Hales, B. U., Wentz, E. A. & Seto, K. C. (2020), 'Trends in urban land expansion, density, and land transitions from 1970 to 2010: A global synthesis', *Environmental Research Letters* **15**(4), 044015.

Harari, M. & Ferrara, E. L. (2018), 'Conflict, climate, and cells: A disaggregated analysis', *Review of Economics and Statistics* **100**(4), 594–608.

Harris, I., Osborn, T. J., Jones, P. & Lister, D. (2020), 'Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset', *Scientific data* **7**(1), 1–18.

Held, I. M. & Soden, B. J. (2006), 'Robust responses of the hydrological cycle to global

warming', *Journal of climate* **19**(21), 5686–5699.

Jayachandran, S. (2006), 'Selling labor low: Wage responses to productivity shocks in developing countries', *Journal of Political Economy* **114**(3), 538–575.

Kirtman, B., Power, S. B., Adedoyin, A. J., Boer, G. J., Bojariu, R., Camilloni, I., Doblas-Reyes, F., Fiore, A. M., Kimoto, M., Meehl, G. et al. (2013), 'Near-term climate change: Projections and predictability'.

Kudamatsu, M., Persson, T. & Strömberg, D. (2012), Weather and infant mortality in Africa, Technical report, CEPR Discussion Papers N. DP9222. Available at: `https://ssrn.com/abstract=2210191`.

Lanfredi, M., Egidi, G., Bianchini, L. & Salvati, L. (2022), 'One size does not fit all: A tale of polycentric development and land degradation in italy', *Ecological Economics* **192**, 107256.

Le, K. & Nguyen, M. (2021), 'In-utero exposure to rainfall variability and early childhood health', *World Development* **144**, 105485.

Maccini, S. & Yang, D. (2009), 'Under the weather: Health, schooling, and economic consequences of early-life rainfall', *American Economic Review* **99**(3), 1006–26.

Malpede, M. & Percoco, M. (2023), 'Aridification, precipitations and crop productivity: evidence from the aridity index', *European Review of Agricultural Economics* **50**(3), 978–1012.

Mary, S. (2022), 'Dams mitigate the effect of rainfall shocks on hindus-muslims riots', *World Development* **150**, 105731.

Masih, I., Maskey, S., Mussá, F. & Trambauer, P. (2014), 'A review of droughts on the african continent: a geospatial and long-term perspective', *Hydrology and earth system sciences* **18**(9), 3635–3649.

McMahon, K. & Gray, C. (2021), 'Climate change, social vulnerability and child nutrition in south asia', *Global Environmental Change* **71**, 102414.

Mendelsohn, R. (2009), 'The impact of climate change on agriculture in developing countries', *Journal of Natural Resources Policy Research* **1**(1), 5–19.

Mendelsohn, R. & Dinar, A. (2003), 'Climate, water, and agriculture', *Land economics* **79**(3), 328–341.

Noce, S., Caporaso, L. & Santini, M. (2020), 'A new global dataset of bioclimatic indicators', *Scientific data* **7**(1), 1–12.

Peri, G. & Sasahara, A. (2019), The impact of global warming on rural-urban migrations: Evidence from global big data, Technical report, National Bureau of Economic Research.

Ponnusamy, S. (2022), 'Rainfall shocks, child mortality, and water infrastructure', *Health Economics* **31**(7), 1317–1338.

Rabassa, M., Skoufias, E. & Jacoby, H. (2014), 'Weather and child health in rural Nigeria', *Journal of African Economies* **23**(4), 464–492.

Randell, H., Jiang, C., Liang, X.-Z., Murtugudde, R. & Sapkota, A. (2021), 'Food insecurity and compound environmental shocks in Nepal: Implications for a changing climate', *World development* **145**, 105511.

Rind, D., Goldberg, R., Hansen, J., Rosenzweig, C. & Ruedy, R. (1990), 'Potential evapotranspiration and the likelihood of future drought', *Journal of Geophysical Research: Atmospheres* **95**(D7), 9983–10004.

Rocha, R. & Soares, R. R. (2015), 'Water scarcity and birth outcomes in the Brazilian semiarid', *Journal of Development Economics* **112**, 72–91.
URL: *https://www.sciencedirect.com/science/article/pii/S0304387814001096*

Rose, E. (1999), 'Consumption smoothing and excess female mortality in rural India', *The Review of Economics and Statistics* **81**(1), 41–49.
URL: *http://www.jstor.org/stable/2646784*

Sarsons, H. (2015), 'Rainfall and conflict: A cautionary tale', *Journal of development Economics* **115**, 62–72.

Sivadasan, J. & Xu, W. (2021), 'Missing women in india: Gender-specific effects of early-life rainfall shocks', *World Development* **148**, 105652.

Thiede, B. C. & Strube, J. (2020), 'Climate variability and child nutrition: Findings from sub-saharan africa', *Global Environmental Change* **65**, 102192.

Unfried, K., Kis-Katos, K. & Poser, T. (2022), 'Water scarcity and social conflict', *Journal of Environmental Economics and Management* **113**, 102633.

Vergunst, F. & Berry, H. L. (2022), 'Climate change and children's mental health: a developmental perspective', *Clinical Psychological Science* **10**(4), 767–785.

Vermeulen, S. J., Campbell, B. M. & Ingram, J. S. (2012), 'Climate change and food systems', *Annual review of environment and resources* **37**, 195–222.

Vicente-Serrano, S. M., Beguería, S. & López-Moreno, J. I. (2010), 'A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index', *Journal of climate* **23**(7), 1696–1718.

Wheeler, T. & Von Braun, J. (2013), 'Climate change impacts on global food security', *Science* **341**(6145), 508–513.

Young, S. L. (2021), 'The measurement of water access and use is key for more effective food and nutrition policy', *Food Policy* **104**, 102138.

Zhang, P., Zhang, J. & Chen, M. (2017), 'Economic impacts of climate change on agriculture: The importance of additional climatic variables other than temperature and precipitation', *Journal of Environmental Economics and Management* **83**, 8–31.

Tables

Table 1: Summary statistics – Climate variables

	1951–1965	1966–1980	1981–1995	1996–2010	2010–2019	Avg growth
Prec	1064.9	1032.9	965.6	1006.1	1016.9	-0.3%
	(545.99)	(527.67)	(523.01)	(520.64)	(516.55)	
	[1.67;3,194.77]	[2.44;3,110.26]	[2.09;3,107.39]	[2.07;3,102.73]	[0.89;2,959.36]	
PET	1402.1	1396.6	1412.0	1431.9	1435.7	0.07%
	(315.89)	(314.28)	(310.26)	(318.03)	(310.12)	
	[809.00;2,664.20]	[811.60;2,680.00]	[823.20;2,704.00]	[822.20;2,794.80]	[836.67;2,713.33]	
Temp	23.231	23.243	23.555	23.905	23.993	0.07%
	(3.58)	(3.61)	(3.58)	(3.58)	(3.56)	
	[8.76;29.68]	[8.90;29.¡93]	[9.35;30.15]	[9.51;30.65]	[9.99;30.68]	

Correlations

	Prec	PET	Prec	PET	Prec	PET	Prec	PET	Prec	PET
PET	-0.58	.	-0.62	.	-0.63	.	-0.63	.	-0.61	.
Tmp	0.21	0.39	0.13	0.41	0.13	0.41	0.13	0.42	0.13	0.40

Notes: summary statistics are shown for a sample of 4,052 grid cells. Precipitations and PET show the total millimetres of rain and water lost by the soil in the year, averaged throughout the indicated period, while temperature is a yearly average (°C). Standard deviations are reported in parentheses; minimum and maximum values are reported in brackets. The average growth column is calculated as $\Delta = 1/T \sum_t^T (X_{t+1} - X_t)/X_t$, where T is the entire sample period (68 years).

Table 2: Summary statistics – Mothers and children

	All sample	Wave III	Wave IV	Wave V	Wave VI	Wave VII	Min	Max
*Sample characteristics**								
N. countries	32	3	14	13	28	12		
N. clusters	36842	859	7885	6433	16010	5655		
Mothers	468873	11277	120289	110359	186949	89867		
Births	2358849	46343	511967	501581	979807	319151		
Births u5	697710	12237	132073	149913	296656	106831		
	Mean (sd)							
Mother's characteristics								
Age	35.254 (8.07)	35.630 (8.02)	36.041 (8.02)	35.083 (8.15)	35.061 (8.03)	34.796 (8.05)	15	50
Education	3.451 (4.18)	2.469 (3.99)	2.852 (4.12)	3.806 (4.00)	3.558 (4.23)	3.669 (4.29)	0	27
Wealth**	2.796 (1.40)	. .	2.822 (1.42)	2.804 (1.39)	2.790 (1.40)	2.781 (1.40)	1	5
Rural household	0.279 (0.45)	0.263 (0.44)	0.266 (0.44)	0.230 (0.42)	0.296 (0.46)	0.330 (0.47)	0	1
Births	5.402 (2.75)	5.724 (2.67)	5.673 (2.78)	5.694 (2.81)	5.226 (2.70)	5.004 (2.65)	1	19
Age at first birth	18.784 (3.64)	18.225 (3.52)	18.915 (3.57)	18.402 (3.51)	18.883 (3.72)	18.976 (3.66)	3	47
Children under 5	1.550 (1.32)	1.694 (1.39)	1.386 (1.28)	1.546 (1.22)	1.623 (1.40)	1.576 (1.26)	0	24
N. of living children	4.560 (2.24)	4.494 (2.19)	4.584 (2.24)	4.652 (2.29)	4.543 (2.22)	4.438 (2.20)	0	16
Infant mortality								
1 year	0.084 (0.28)	0.120 (0.33)	0.103 (0.30)	0.097 (0.30)	0.073 (0.26)	0.062 (0.24)	0	1
5 years	0.120 (0.33)	0.172 (0.38)	0.144 (0.35)	0.140 (0.35)	0.105 (0.31)	0.089 (0.28)	0	1
1 year – boys	0.090 (0.29)	0.126 (0.33)	0.110 (0.31)	0.104 (0.30)	0.079 (0.27)	0.068 (0.25)	0	1
5 years – boys	0.127 (0.33)	0.177 (0.38)	0.151 (0.36)	0.148 (0.35)	0.111 (0.31)	0.096 (0.30)	0	1
Infant nutrition–health								
BMI (SD)***	0.57 (142.94)	. .	. .	1.82 (156.37)	1.22 (144.29)	-2.72 (119.50)	-500	500
Weight/Height (SD)***	-12.38 (140.44)	.	.	-12.88 (152.18)	-11.83 (140.38)	-13.20 (124.27)	0	99.8
Size at birth (cat) ****	2.22 (0.99)	2.45 (0.99)	2.08 (0.95)	2.32 (0.98)	2.23 (0.99)	2.25 (1.00)	0	4

Notes: the DHS surveys employed were conducted between 1992 and 2019.

* The number of countries, clusters, mothers and generic births refers to the data available from the Individual Recode DHS survey. The number of births under 5 refers to the data available from the Child Recode DHS survey.

** Wealth is a categorical index ranging from 1 (very poor) to 5 (very rich).

*** The measures are presented with two implied decimal places; the actual value is obtained by dividing the variable by 100.

**** Size at birth is a categorical variable ranging from 0 (very small) to 4 (very large).

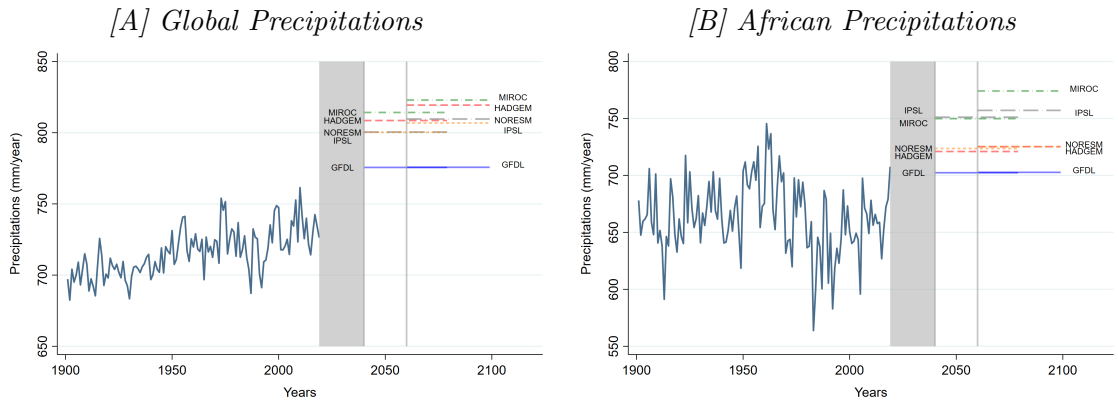
Table 3: Impact of precipitations and the PET on infant health

	Mort (5y)		**BMI (SD)**		**Size at birth**	
	(1)	(2)	(3)	(4)	(5)	(6)
			[A] Yearly values			
Prec_t	-0.0048***	-0.0026***	7.4652***	5.1982***	0.0363***	0.0222***
	(0.0009)	(0.0010)	(1.2166)	(1.2941)	(0.0064)	(0.0067)
$\text{Prec}_{(t-1)}$	-0.0033***	-0.0013	3.3141***	1.2999	0.0308***	0.0172***
	(0.0009)	(0.0010)	(1.1741)	(1.2384)	(0.0059)	(0.0063)
PET_t		0.0053		-9.6136*		0.0139
		(0.0045)		(5.1651)		(0.0266)
$\text{PET}_{(t-1)}$		0.0033		-1.3564		-0.0754***
		(0.0044)		(5.1662)		(0.0266)
Temperature_t	0.0211	0.0161	-2.4049	-1.2553	0.0190	0.0174
	(0.0244)	(0.0253)	(2.4183)	(2.4610)	(0.0129)	(0.0132)
$\text{Temperature}_{(t-1)}$	-0.0130	-0.0204	-1.2065	-0.9917	-0.0239*	-0.0148
	(0.0244)	(0.0254)	(2.4268)	(2.4632)	(0.0129)	(0.0131)
Observations	2,059,690	2,059,690	266,469	266,469	518,591	518,591
R^2	0.0601	0.0601	0.1999	0.2000	0.1331	0.1332
			[B] Growing season			
Prec GS $_t$	-0.0027***	-0.0017*	3.4957***	2.4391**	0.0173***	0.0080
	(0.0009)	(0.0009)	(1.0977)	(1.1367)	(0.0059)	(0.0062)
Prec GS $_{(t-1)}$	-0.0034***	-0.0027***	3.2944***	2.2290**	0.0117**	0.0046
	(0.0008)	(0.0009)	(1.0955)	(1.1369)	(0.0058)	(0.0059)
PET GS $_t$		0.0053**		-4.5811*		-0.0416***
		(0.0025)		(2.7278)		(0.0147)
PET GS $_{(t-1)}$		0.0006		-4.5907*		-0.0137
		(0.0024)		(2.7214)		(0.0146)
Temperature GS_t	0.0014	0.0002	0.9829	1.6357	-0.0122*	-0.0038
	(0.0010)	(0.0012)	(1.5990)	(1.7208)	(0.0072)	(0.0082)
Temperature GS_{t-1}	-0.0001	-0.0001	-5.0490***	-3.9712**	0.0031	0.0055
	(0.0010)	(0.0012)	(1.5980)	(1.7220)	(0.0072)	(0.0082)
Observations	1,721,384	1,721,384	227,786	227,786	432,729	432,729
R^2	0.0598	0.0598	0.1833	0.1835	0.1287	0.1289
Controls	Long	Long	Long	Long	Long	Long
Year FE	Y	Y	Y	Y	Y	Y
Cluster FE	Y	Y	Y	Y	Y	Y
Country Trends	Y	Y	Y	Y	Y	Y

Notes: the table presents the estimates of vectors $\beta_{\mathbf{1}}$ and $\beta_{\mathbf{2}}$. Long controls include the woman's education in single years, wealth index of the household, kid's month of birth, main source of water, and cell temperature at times t and $t-1$. Columns (1) and (2) are based on a sample of 33 countries and 22,757 clusters. Columns (3) to (6) are based on a sample of 34 countries and 22,909 clusters. In Panel [A], precipitations and the PET are computed over the entire year; in Panel [B], only the growing season months are considered. Robust standard errors are clustered at the DHS cluster level, with significance levels at 10, 5, and 1 percent.

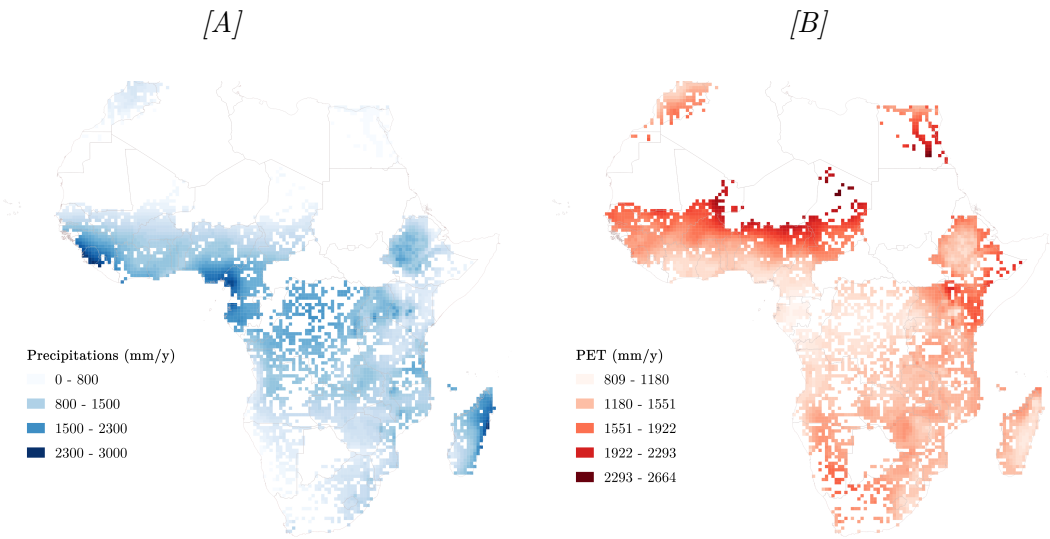
Figures

Figure 1: Historical trend and future projections in yearly precipitation, 1901–2099



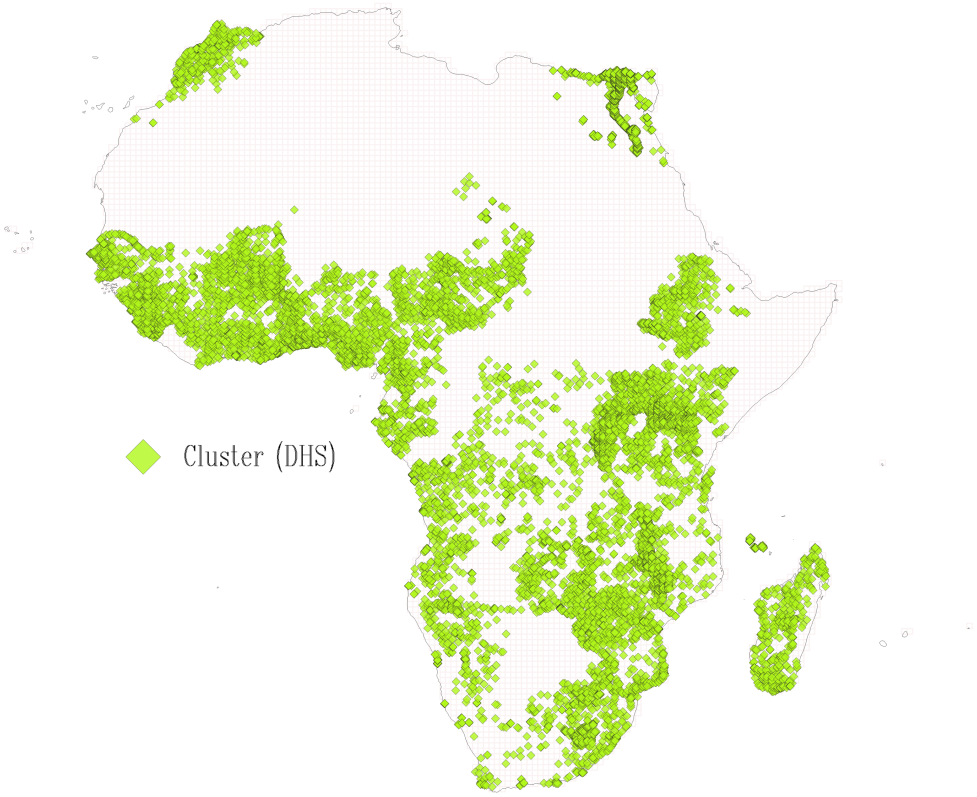
Notes: the figure depicts the trend in yearly precipitations (mm/year) starting form 1901. The series between 1901 and 2019 is computed from the CRU TS4.04 dataset. Projections for two time horizons (2040–2079 and 2060–2099) are accessed from five commonly employed earth system models (ESMs): GFDL-ESM2M, HadGEM2-ES, IPSL-CM54-LR, MIROC-ESM-CHEM, and NorESM1-M. Bias-corrected projections are plotted under the representative concentration pathway (RCP) 4.5, a greenhouse gas concentration trajectory, which possibly constitutes the most probable baseline scenario by taking into account the exhaustible character of non-renewable fuels. Panel [A] plots average global precipitations levels; Panel [B] focuses on the African continent. Source: CMCC-BioClimInd dataset.

Figure 2: Geographical variation in precipitations and the PET (1951-2018)



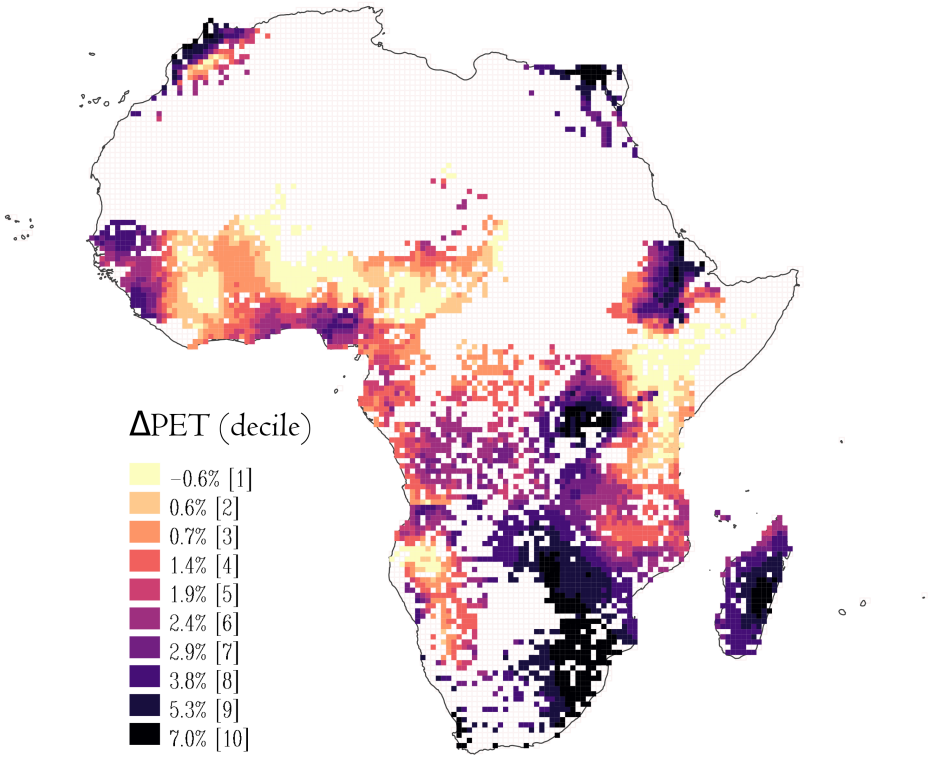
Notes: Panel [A] depicts precipitations (mm/year) over the sample grid. Panel [B] shows the PET (mm/year) for the same cells. In Panel [A], lighter cells identify areas of scarce precipitations. In Panel [B], darker cells identify arid regions.

Figure 3: Sampled DHS clusters



Notes: the image reports all the clusters appearing at least once in the sample. The sample includes up to 22,756 unique clusters, with an average (standard deviation) number of mothers per cluster-survey year of 53.3 (34.1). Geolocation displacement can be up to 2 km for urban and 10 km for rural clusters.

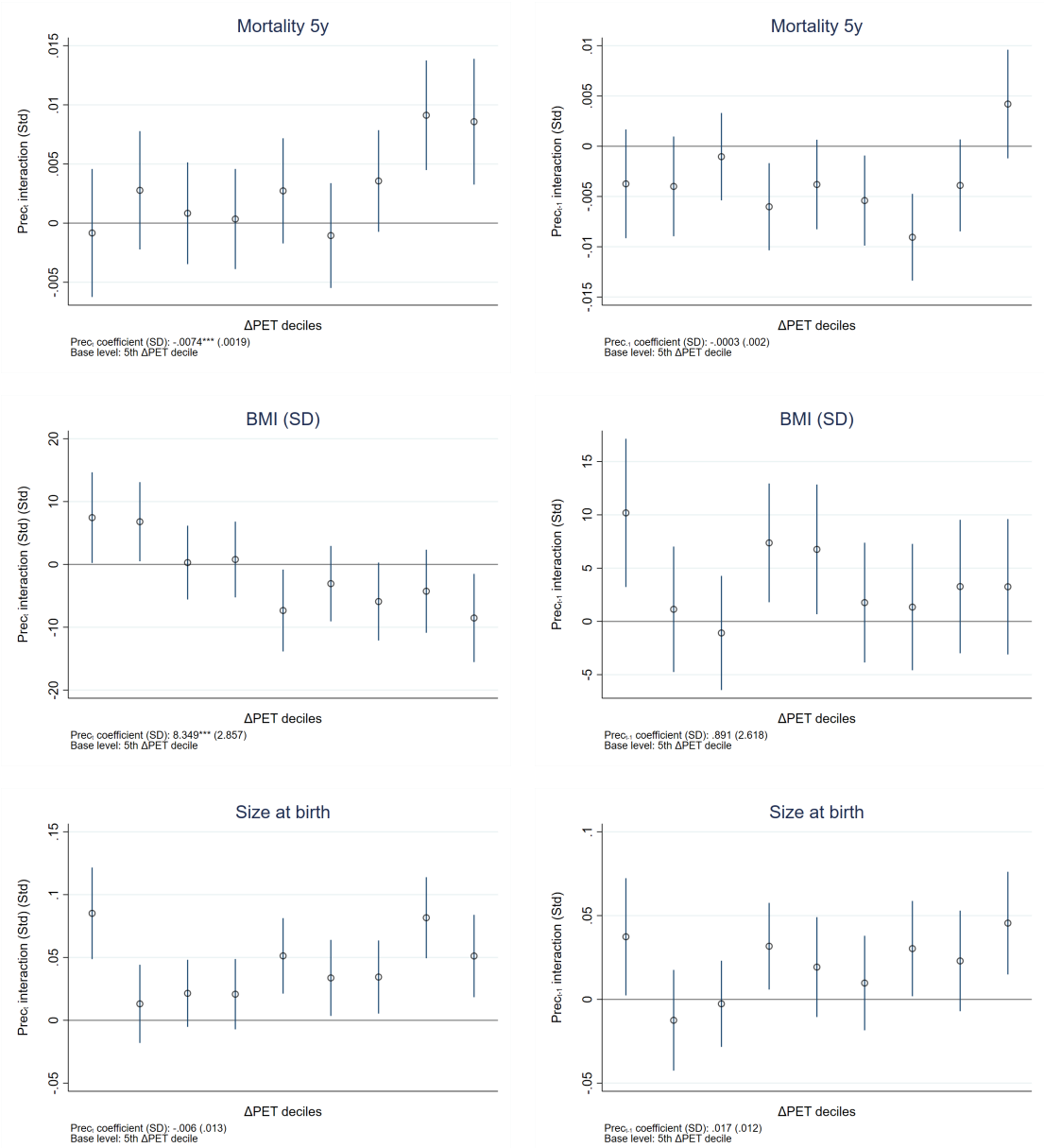
Figure 4: PET variation deciles–Cell distribution



Notes: the image plots the sample grid classified by deciles in the distribution of PET long-term variation. Long-term variation is calculated as the growth rate of the 10-year average PET in the decades between 1950 and 1959 and 2010 and 2019. The legend reports the average percentage variation in the PET per decile (in parentheses).

Figure 5: Effect of rainfall by deciles of land aridification

[A] $\sum_k \gamma_{0,k}$ coefficients (PRE_t) [B] $\sum_k \gamma_{1,k}$ coefficients (PRE_{t-1})



Notes: the image reports $\gamma_{r,k}$ (with $r \in \{0,1\}$ and $k \in \mathbb{N} \cap [1,10]$) from Equation 2. Panel [A] reports interaction terms with precipitation at time t. In Panel [B], the same decile indicators are interacted with lagged precipitation. The sample distribution of ΔPET is found in Figure 4. 95% confidence intervals clustered at DHS cluster level are reported.

A Online Appendix

A.1 Computation of the Potential Evapotranspiration

We access a measure for the potential evapotranspiration from the CRU TS.04 dataset. This is calculated using a modeling scheme based on climate simulations developed by the Hadley Centre (HadRM3H). A full description of the relevant regional climate models can be found in Ekström et al. (2007). Here, we report a short summary explaining the computation of the PET.

The estimates for the PET are provided using a variant of the Penman-Monteith method, as proposed by the FAO. This indicator is addressed as *potential* since it employs a grass reference crop. [15] The PET is computed according to the following equation:

$$PET = \frac{0.408\Delta\left(R_n - G\right) + \gamma + \frac{900}{T+273.16}U_2\left(e_a - e_d\right)}{\Delta + \gamma\left(1 + 0.34U_2\right)} \tag{3}$$

where R_n represents the net radiation at the crop surface (MJ m^{-2} per day), G is soil heat flux (MJ m^{-2} per day), T is the mean temperature, U is wind speed ($\frac{m}{s}$), $(e_a - e_d)$ and Δ are the vapour pressure deficit and the relative slope of the vapour pressure curve ($\frac{kPa}{\circ C}$) respectively, and γ is a psychrometric constant. While wind speed and temperature are direct outputs from the HadRM3H, the other constants in the formula are calculated using the model data.

As Equation 3 suggests, while the temperature is indeed relevant in the computation of the PET, which justify an average positive correlation of around 40% between PET and temperature, it is only part of the story. As such, by controlling for the yearly average temperature in our main specifications, we are able to isolate the effect of soil water availability without capturing potential noise coming from heat volatility.

A.2 Datasets and Variables Description

Precipitations (PRE): total, mm/year

Potential evapotranspiration (PET): total, mm/year. See Section A.1.2 for the computational details

Temperature: °C, average monthly value at 2 m altitude.

Growing Season: growing season months are calculated using the start date of the earliest growing period (day of year) for the time period 1981–2010 and the total number of growing period days.

[15] For the original contribution on this computation, see Allen et al. (1994).

Both measures are accessed through the GAEZ v4 dataset, which employs the climate data source HadGEM2-ES. More information can be found in Fischer et al. (2021).

Infant mortality (1–5 years): binary indicator computed using the mother-reported time of death after birth (in months).

BMI (SD): Body mass index, defined as the weight in kilograms divided by the square of the height in meters (W/H^2), and then expressed in standard deviations.

Weight/height: Weight for height standard deviations from the reference median based on the DHS reference standard.

Education: Highest year of education to give the years of education completed.

Wealth index: a composite measure of a household's cumulative living standard, calculated using easy-to-collect data on a household's ownership of selected assets, such as televisions and bicycles and materials used for housing construction. Generated by principal components analysis, the index separates individual households into five wealth quintiles.

Source of drinking water: Main source of drinking water for members of the household (major categories).

A.3 CRU TS.04 Choice and Comfront with Alternative Datasets

When it comes to environmental studies, researchers in need of high-frequency data on weather and climate conditions have more than one alternative. Researchers have compared and highlighted the peculiarities of these different sources. However, no rule of thumb exists to guide them through the adoption of one particular dataset.

In their paper on weather shocks, malaria, and child mortality, Kudamatsu et al. (2012) access observations on monthly rainfall through the 45-Year European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis (ERA-40) data archive. The authors justify their choice by claiming its superiority over the more well-known CRU dataset. Their main argument relies on the fact that rainfall gauge data in Africa lack the necessary quality and show bias in arid and semi-arid areas, where departures from standard seasonal fluctuations are more pronounced. A similar argument is provided by Harari & Ferrara (2018) to justify the adoption of the ERA-40. Other authors have instead deemed gauge data suitable for the purpose of their studies, and have,

thus, turned at the CRU dataset, usually in its previous versions (Vicente-Serrano et al. 2010, Couttenier & Soubeyran 2014).

While the concerns about gauge data are surely legitimate, significant drawbacks are also implied by the choice of reanalysis data. First, the ERA-40 dataset is provided at more than twice the resolution of the CRU TS.04 one, at 1.25 $\times$ 1.25 degress (roughly 139 km $\times$ 139 km), which is a significant loss in terms of spatial variation. As our sample is an unbalanced panel, variation across grid cells is of great importance, and, as such, this could impinge on the detection of an effect of precipitations and the PET on agricultural productivity and infant health. Aanother alternative available to researchers is the ERA-5 dataset, in which near-surface meteorological variables have been re-gridded to a half-degree resolution. Yet, in addition to using monthly-scale bias corrections still based on CRU data, this dataset is only available from 1980. Second, reanalysis relies on a variety of sources, including weather stations, ships, aircraft, and satellites. To provide the corresponding weather measures, recorded data are analyzed through an atmospheric circulation model (IFS CY23r4). Compared to gauge data, this augments the risk of measurement error.

It is indeed true that CRU stations are partially dispersed in Sub-Saharan Africa, and, since they cannot provide full direct coverage, the resulting data rely on interpolation. However, the data on station location and resulting cover contained in Harris et al. (2020), in which stations are included if they contribute at least 75% of observations in the decade, show that the problem of loss of variability due to interpolation may be more of a concern in the areas of scarce coverage (in this case, historical data would have a greater role in filling in for missing observations). In Figure A.1, we plot the spatial distribution of the clusters available from the DHS together with the spatial coverage of the CRU stations in the decades 1970-79 and 2000-09. Coverage in the CRU dataset is defined as an area that experiences the direct measurement of at least 75% of all potential observations in the decade. We notice that most of the clusters in the sample are actually within the declared coverage, which provides reassurance, at least regarding the probability of errors generated by the stations. As we use the area identified by the DHS cluster to run the analysis on crop productivity, a similar reasoning applies to crop yield observations. This evidence ultimately helps reduce the concern that our data may not properly capture the variability in precipitations and PET. This, in turn, is essential to our research strategy and contributes to justifying the choice of the CRU TS.04 dataset.

A.4 Appendix Tables

Table A.1: Impact of precipitations and the PET on infant health – alternative measures

	Mort (1y)		**W/H ratio (SD)**	
	(1)	(2)	(3)	(4)
	[A] Yearly values			
Prec_t	-0.0025***	-0.0020**	7.9419***	5.4943***
	(0.0008)	(0.0008)	(1.1803)	(1.2505)
$\text{Prec}_{(t-1)}$	-0.0010	-0.0005	3.1379***	0.8670
	(0.0008)	(0.0008)	(1.1443)	(1.2045)
PET_t		0.0022		-6.9669
		(0.0039)		(5.0209)
$\text{PET}_{(t-1)}$		-0.0001		-5.1389
		(0.0038)		(4.9873)
Temperature_t	0.0245	0.0218	-1.2422	-0.3121
	(0.0212)	(0.0220)	(2.3459)	(2.3840)
$\text{Temperature}_{(t-1)}$	-0.0209	-0.0212	-2.2272	-1.6508
	(0.0213)	(0.0220)	(2.3541)	(2.3837)
Observations	2,059,690	2,059,690	267,083	267,083
R^2	0.0391	0.0391	0.1964	0.1966
	[B] Growing season			
Prec GS $_t$	-0.0014*	-0.0012*	3.0697***	1.9820*
	(0.0007)	(0.0007)	(1.0663)	(1.1028)
Prec GS $_{(t-1)}$	-0.0017**	-0.0017**	2.9896***	1.7785
	(0.0007)	(0.0007)	(1.0694)	(1.1072)
PET GS $_t$		0.0014		-4.1160
		(0.0021)		(2.7023)
PET GS $_{(t-1)}$		-0.0005		-5.8206**
		(0.0021)		(2.6913)
Temperature GS_t	0.0009	0.0005	0.2902	0.7620
	(0.0009)	(0.0010)	(1.5493)	(1.6715)
$\text{Temperature GS}_{(t-1)}$	-0.0003	-0.0001	-4.2276***	-2.8285*
	(0.0009)	(0.0010)	(1.5515)	(1.6744)
Observations	1,721,384	1,721,384	228,387	228,387
R^2	0.0386	0.0386	0.1811	0.1812
Controls	Long	Long	Long	Long
Year FE	Y	Y	Y	Y
Cluster FE	Y	Y	Y	Y
Country Trends	Y	Y	Y	Y

Notes: the table presents the estimates of $\beta_{\mathbf{1}}$ and $\beta_{\mathbf{2}}$ for comparable outcome variables (infant mortality and weight/height ratio). Long controls include the woman's education in single years, wealth index of the household, child's month of birth, main source of water, and cell raw temperature in t and $t-1$. In Panel [A], precipitations and PET are computed over the entire year; in Panel [B], only the growing season months are considered. Robust standard errors are clustered at the DHS cluster level, with significance levels at 10, 5, and 1 percent.

Table A.2: Impact of precipitations and the PET on infant health - long-settled sample

	Mort (5y)		**BMI (SD)**		**Size at birth**	
	(1)	(2)	(3)	(4)	(5)	(6)
			[A] Yearly values			
Prec_t	-0.0047***	-0.0023**	7.3400***	5.3897***	0.0336***	0.0169**
	(0.0010)	(0.0011)	(1.4110)	(1.5072)	(0.0077)	(0.0083)
$\text{Prec}_{(t-1)}$	-0.0037***	-0.0014	4.0424***	2.2763	0.0321***	0.0166**
	(0.0010)	(0.0011)	(1.3594)	(1.4348)	(0.0071)	(0.0076)
PET_t		0.0050		-7.4546		0.0280
		(0.0050)		(6.0538)		(0.0315)
$\text{PET}_{(t-1)}$		0.0048		-2.4066		-0.0998***
		(0.0050)		(6.0461)		(0.0315)
Temperature_t	0.0182	0.0111	-16.8108***	-15.6758***	0.0280	0.0210
	(0.0426)	(0.0448)	(5.3342)	(5.4628)	(0.0267)	(0.0274)
$\text{Temperature}_{(t-1)}$	0.0129	0.0135	13.3098**	12.0212**	-0.0327	-0.0235
	(0.0425)	(0.0448)	(5.3425)	(5.4646)	(0.0267)	(0.0274)
Observations	1,673,718	1,673,718	199,674	199,674	378,859	378,859
R^2	0.0647	0.0647	0.2291	0.2292	0.1527	0.1529
			[B] Growing season			
Prec GS $_t$	-0.0003	0.0007	3.8901***	2.8326**	0.0159**	0.0061
	(0.0009)	(0.0009)	(1.3120)	(1.3546)	(0.0074)	(0.0077)
Prec GS $_{(t-1)}$	-0.0013	-0.0005	3.6749***	2.7837**	0.0059	-0.0020
	(0.0009)	(0.0009)	(1.2832)	(1.3257)	(0.0070)	(0.0072)
PET GS $_t$		0.0036		-5.3823*		-0.0393**
		(0.0027)		(3.1412)		(0.0172)
PET GS $_{(t-1)}$		0.0021		-3.1989		-0.0229
		(0.0027)		(3.1077)		(0.0172)
Temperature GS_t	0.0008	-0.0013	0.8861	2.5824	-0.0080	0.0009
	(0.0017)	(0.0021)	(3.2192)	(3.6601)	(0.0124)	(0.0151)
Temperature $\text{GS}_{(t-1)}$	-0.0004	-0.0003	-3.8953	-3.8184	-0.0019	-0.0034
	(0.0017)	(0.0021)	(3.1864)	(3.6175)	(0.0124)	(0.0148)
Observations	1,384,393	1,384,393	166,138	166,138	311,944	311,944
R^2	0.0386	0.0387	0.2100	0.2101	0.1465	0.1467
Controls	Long	Long	Long	Long	Long	Long
Year FE	Y	Y	Y	Y	Y	Y
Cluster FE	Y	Y	Y	Y	Y	Y
Country Trends	Y	Y	Y	Y	Y	Y

Notes: the table presents the estimates of $\beta_{\mathbf{1}}$ and $\beta_{\mathbf{2}}$ for the sample of women who have been living in the same place of residence for at least 15 years. Long controls include the woman's education in single years, wealth index of the household, child's month of birth, main source of water, and cell temperature in t and $t-1$. Columns (1) to (2) are based on a sample of 34 countries and 22,900 clusters. Columns (3) to (6) are based on a sample of 33 countries and 22,757 clusters. In Panel [A], precipitations and PET are computer over the entire year; in Panel [B], only the growing season months are considered. Robust standard errors are clustered at the DHS cluster level, with significance levels at 10, 5, and 1 percent.

Table A.3: Impact of precipitations and the PET on infant health – Urban vs rural areas

	Mort (5y)		**BMI (SD)**		**Size at birth**	
	(1)	(2)	(3)	(4)	(5)	(6)
			[A] Rural Sample			
Prec_t	-0.0043***	-0.0027**	9.1989***	6.9904***	0.0488***	0.0319***
	(0.0012)	(0.0012)	(1.6058)	(1.6906)	(0.0088)	(0.0092)
$\text{Prec}_{(t-1)}$	-0.0021*	-0.0006	2.9441*	0.9855	0.0408***	0.0249***
	(0.0012)	(0.0012)	(1.5243)	(1.5885)	(0.0080)	(0.0083)
PET_t		0.0016		-8.9526		0.0163
		(0.0055)		(6.3722)		(0.0326)
$\text{PET}_{(t-1)}$		0.0052		-3.4781		-0.0932***
		(0.0055)		(6.3143)		(0.0322)
Temperature_t	-0.0111	-0.0213	-2.3204	-0.8837	0.0364	0.0362
	(0.0422)	(0.0445)	(4.7090)	(4.8426)	(0.0237)	(0.0243)
$\text{Temperature}_{(t-1)}$	0.0150	0.0120	-0.7348	-0.3710	-0.0475**	-0.0412*
	(0.0425)	(0.0448)	(4.7395)	(4.8602)	(0.0237)	(0.0243)
Observations	1,477,984	1,477,984	186,831	186,831	367,939	367,939
R^2	0.0634	0.0634	0.2093	0.2094	0.1408	0.1409
			[B] Urban Sample			
Prec_t	-0.0030**	-0.0014	6.9626***	5.0379**	0.0052	-0.0029
	(0.0015)	(0.0016)	(2.3029)	(2.4781)	(0.0107)	(0.0116)
$\text{Prec}_{(t-1)}$	-0.0035**	-0.0019	6.5940***	4.8267**	0.0074	-0.0005
	(0.0015)	(0.0016)	(2.3223)	(2.4551)	(0.0105)	(0.0114)
PET_t		0.0071		-9.9607		0.0082
		(0.0078)		(9.8640)		(0.0493)
$\text{PET}_{(t-1)}$		0.0012		-2.5634		-0.0508
		(0.0079)		(9.8445)		(0.0495)
Temperature_t	0.0231	0.0229	-3.0286	-1.7841	0.0086	0.0073
	(0.0302)	(0.0312)	(2.8960)	(2.9471)	(0.0156)	(0.0160)
$\text{Temperature}_{(t-1)}$	-0.0141	-0.0240	-0.2935	0.0294	-0.0126	-0.0022
	(0.0302)	(0.0313)	(2.9163)	(2.9487)	(0.0155)	(0.0158)
Observations	581,679	581,679	79,505	79,505	150,620	150,620
R^2	0.0602	0.0602	0.2276	0.2277	0.1563	0.1563
Controls	Long	Long	Long	Long	Long	Long
Country FE	N	N	N	N	N	N
Year FE	Y	Y	Y	Y	Y	Y
Cluster FE	Y	Y	Y	Y	Y	Y
Country Trends	Y	Y	Y	Y	Y	Y

Notes: the table presents the estimates of vectors $\beta_{\mathbf{1}}$ and $\beta_{\mathbf{2}}$ separately for rural (Panel [A]) and urban (Panel [B]) clusters. Climate variables are measured using the entire years t and $t-1$. Long controls include the woman's education in single years, wealth index of the household, child's month of birth, main source of water, and cell temperature at times t and $t-1$. Columns (1) and (2) are based on a sample of 16,319 rural and 11,292 urban clusters. Columns (3) to (6) are based on a sample of 16,560 rural and 11,368 urban clusters. Robust standard errors are clustered at the DHS cluster level, with significance levels at 10, 5, and 1 percent.

Table A.4: Precipitations, the PET and child health – Unstandardized coefficients

	Mort (5y)		**BMI (SD)**		**Size at birth**	
	(1)	(2)	(3)	(4)	(5)	(6)
			[A] Yearly values			
Prec_t	-0.0008***	-0.0004***	1.2536***	0.8729***	0.0061***	0.0037***
	(0.0002)	(0.0002)	(0.2043)	(0.2173)	(0.0011)	(0.0011)
$\text{Prec}_{(t-1)}$	-0.0006***	-0.0002	0.5567***	0.2184	0.0052***	0.0029***
	(0.0002)	(0.0002)	(0.1972)	(0.2080)	(0.0010)	(0.0011)
PET_t		0.0017		-3.0913*		0.0045
		(0.0014)		(1.6609)		(0.0086)
$\text{PET}_{(t-1)}$		0.0010		-0.4352		-0.0242***
		(0.0014)		(1.6575)		(0.0085)
Temperature_t	0.0211	0.0161	-2.4049	-1.2553	0.0190	0.0174
	(0.0244)	(0.0253)	(2.4183)	(2.4610)	(0.0129)	(0.0132)
$\text{Temperature}_{(t-1)}$	-0.0130	-0.0204	-1.2065	-0.9917	-0.0239*	-0.0148
	(0.0244)	(0.0254)	(2.4268)	(2.4632)	(0.0129)	(0.0131)
Observations	2,059,690	2,059,690	266,469	266,469	518,591	518,591
R^2	0.0601	0.0601	0.1999	0.2000	0.1331	0.1332
			[B] Growing season			
Prec GS_t	-0.0001***	-0.0000*	0.0700***	0.0489**	0.0003***	0.0002
	(0.0000)	(0.0000)	(0.0220)	(0.0228)	(0.0001)	(0.0001)
Prec GS_{t-1}	-0.0001***	-0.0001***	0.0658***	0.0445**	0.0002**	0.0001
	(0.0000)	(0.0000)	(0.0219)	(0.0227)	(0.0001)	(0.0001)
PET GS_t		0.0083**		-7.2105*		-0.0655***
		(0.0038)		(4.2934)		(0.0231)
$\text{PET GS}_{(t-1)}$		0.0010		-7.2090*		-0.0216
		(0.0038)		(4.2735)		(0.0230)
Temperature GS_t	0.0014	0.0002	0.9829	1.6357	-0.0122*	-0.0038
	(0.0010)	(0.0012)	(1.5990)	(1.7208)	(0.0072)	(0.0082)
$\text{Temperature GS}_{(t-1)}$	-0.0001	-0.0001	-5.0490***	-3.9712**	0.0031	0.0055
	(0.0010)	(0.0012)	(1.5980)	(1.7220)	(0.0072)	(0.0082)
Observations	1,721,384	1,721,384	227,786	227,786	432,729	432,729
R^2	0.0598	0.0598	0.1833	0.1835	0.1287	0.1289
Controls	Long	Long	Long	Long	Long	Long
Year FE	Y	Y	Y	Y	Y	Y
Cluster FE	Y	Y	Y	Y	Y	Y
Country Trends	Y	Y	Y	Y	Y	Y

Notes: the table presents unstandardized estimates of $\beta_{\mathbf{1}}$ and $\beta_{\mathbf{2}}$. Rainfall and PET are rescaled to represent a 100 mm/month variation. Long controls include the woman's education in single years, wealth index of the household, child's month of birth, main source of water, and cell temperature in t and $t-1$. In Panel [A], precipitations and PET are computer over the entire year; in Panel [B], only growing season months are considered. Robust standard errors are clustered at the DHS cluster level, with significance levels at 10, 5, and 1 percent.

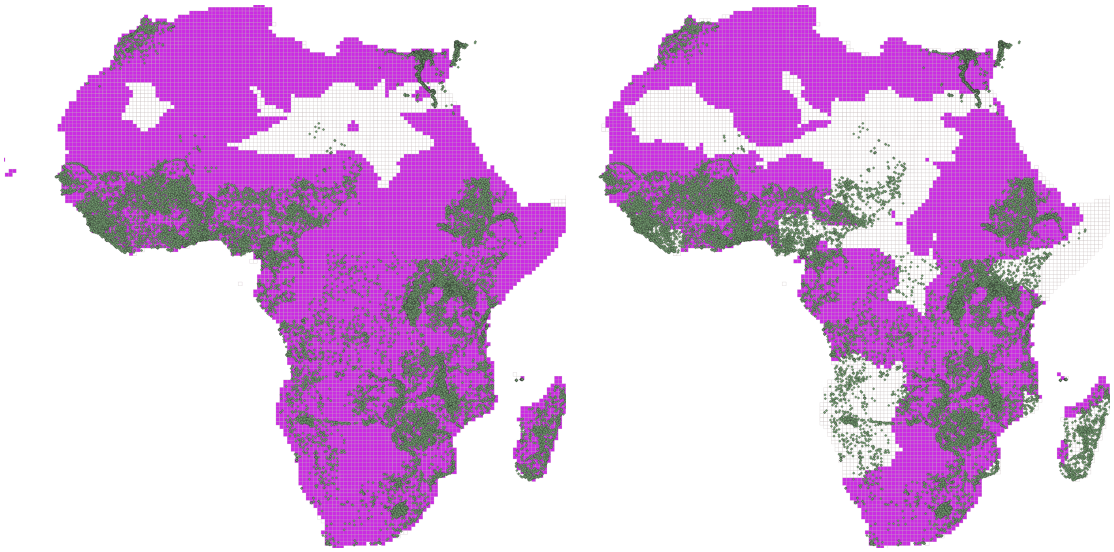
Table A.5: Precipitations and the PET projections – Sample averages

	RCP 4.5				**RCP 8.5**			
	Rainfall		PET		Rainfall		PET	
	2040–2079	2060–2099	2040–2079	2060–2099	2040–2079	2060–2099	2040–2079	2060–2099
GFDL	1094.3	1098.4	1437.5	1458.7	1099.7	1105.2	1520.6	1600.6
HadGEM2	1117.4	1123.7	1524.2	1566.7	1116.7	1114.1	1627.9	1743.6
IPSL	1176.0	1189.0	1500.6	1533.6	1222.3	1255.6	1608.7	1730.0
MIROC	1150.9	1180.3	1458.6	1488.6	1183.7	1224.0	1547.1	1547.1
NorESM1	1107.0	1113.4	1419.4	1441.1	1111.9	1128.0	1496.6	1578.9
Sample	1021.1	1021.1	1412.9	1412.9	1021.1	1021.1	1412.9	1412.9

Notes: the table presents the projections for yearly precipitations and the PET (mm/year) averaged over our sample grid from five ESMs (GFDL-ESM2M, HadGEM2-ES, IPSL-CM54-LR, MIROC-ESM-CHEM, NorESM1-M) for the time intervals 2040–2079 and 2060–2099. Data are displayed for two RCP scenarios: 4.5 (decreasing "intermediate" emission levels by 2100) and 8.5 (non-decreasing "worst-case" emission levels by 2100). Source: CMCC-BioClimInd dataset.

A.5 Appendix Figures

Figure A.1: DHS clusters and CRU TS.04 sensor coverage



[1970–1979] [2000–2009]

Notes: green dots represent the spatial distribution of DHS clusters in our sample. The purple-shaded area identifies the grid cell station coverage of the CRU dataset in the decades 1970–2079 and 2000–2009.
Source: Harris et al. (2020).