Lee, Wang-Sheng; Tran, Trang My

Working Paper

Emissions from Military Training: Evidence from Australia

IZA Discussion Papers, No. 16889

Provided in Cooperation with:
IZA – Institute of Labor Economics

Suggested Citation: Lee, Wang-Sheng; Tran, Trang My (2024) : Emissions from Military Training: Evidence from Australia, IZA Discussion Papers, No. 16889, Institute of Labor Economics (IZA), Bonn

This Version is available at:
https://hdl.handle.net/10419/295912

I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 16889

Emissions from Military Training: Evidence from Australia

Wang-Sheng Lee
Trang My Tran

MARCH 2024

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 16889

Emissions from Military Training: Evidence from Australia

Wang-Sheng Lee
Monash University and IZA

Trang My Tran
Monash University

MARCH 2024

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Emissions from Military Training: Evidence from Australia*

Environmental research related to military activities and warfare is sparse and fragmented by discipline. Although achieving military objectives will likely continue to trump any concerns related to the environment during active conflict, military training during peacetime has environmental consequences. This research aims to quantify how much pollution is emitted during regular military exercises which has implications for climate change. Focusing on major military training exercises conducted in Australia, we assess the impact of four international exercises held within a dedicated military training area on pollution levels. Leveraging high-frequency data, we employ a machine learning algorithm in conjunction with program evaluation techniques to estimate the effects of military training activities. Our main approach involves generating counterfactual predictions and utilizing a "prediction-error" framework to estimate treatment effects by comparing a treatment area to a control area. Our findings reveal that these exercises led to a notable increase in air pollution levels, potentially reaching up to 25% relative to mean levels during peak training hours.

JEL Classification: C55, Q53, Q54

Keywords: machine learning, military emissions, military training, pollution

Corresponding author:
Trang My Tran
Department of Economics
School of Business
Monash University
47500 Subang Jaya
Selangor
Malaysia

E-mail: mytrang.tran@monash.edu

* This project was funded by an internal grant from the Faculty of Business and Economics at Monash University.

1 Introduction

Climate change and global warming are becoming increasingly important global issues. According to climate scientists, time is running out for concrete action that can be undertaken by countries around the world to reduce their carbon emissions and footprints. The 2015 Paris Agreement aims to limit global warming to below 2 degrees Celsius. Land reserved for military use accounts for approximately 6% of Earth's terrestrial surface and global defense spending is approximately 2.5% of total world GDP (Zentelis and Lindenmayer, 2015). While much recent attention in climate change forums has been placed on how we make things, how food is grown and how power is generated, all industries around the world need to play a part in reducing emissions, including the military.

Little is known how much pollution is generated during regular military exercises that are conducted by many developed and developing countries. Despite the outsized role of militaries, we know surprisingly little about their emissions.[1] Furthermore, the reports of the Intergovernmental Panel on Climate Change (IPCC), the UN's scientific advisory body on the issue, have barely mentioned the military sector. National security restrictions have considerably limited access to data for military emissions, with very few nations reporting such statistics, and many not compiling them at all. Under the 2015 Paris Agreement, rules for reporting of military and conflict-related emissions need to be developed. Unfortunately, the Russian invasion of Ukraine in February 2022, the Israel-Hamas war that started in October 2023, and continuing political developments makes it likely that the issue of military and conflict-related emissions will become worse rather than get better in the short term.

Conflict and war have been ever present aspects of human civilization. Through the expansion and contraction of various empires over time, history has shown the importance of having a strong military for defence and to act as a deterrence against possible invaders. Numerous studies have also shown that the presence of security guarantees and international peacekeeping troops correlate with longer peace duration (Rohner, 2024). For example, Hultman et al. (2014) find that greater United Nations military troop strength significantly reduced battlefield deaths during civil wars in Africa from 1992 to 2011.

Today, regular military training exercises are conducted on a daily basis around the world (Svenningsen et al., 2019). As military activities are largely exempt from environmental protection legislation in the interest of national security, environmental degradation as a consequence of military activities largely remains a neglected subject. Due to challenges associated with conducting research in areas with military activities (e.g., restricted access,

[1] https://theconversation.com/how-the-worlds-militaries-hide-their-huge-carbon-emissions-171466.

hazardous conditions), information pertaining to military impacts on the environment is relatively scarce and is often studied years after military activities have ceased and with no knowledge of baseline conditions (Lawrence et al., 2015).

Gould (2007) refers to militarization as "the single most ecologically destructive human endeavor" as wars and conflicts not only traumatize people physically and psychologically but also contaminate the environment by using chemicals, herbicides, and radiation. One reason why military production and deployment zones have been allowed to become enormous environmental and public health disasters is the secrecy with which such operations are conducted.

Military training exercises and live-firing training during peacetime not only lead to local ecological disruption, landscape alteration, vegetation destruction, soil and water contamination, but also result in considerable emissions of pollutants that have adverse consequences for the climate. Zentelis et al. (2017) conducted a desktop review of Australian and German military training areas management documentation to determine whether they contained management principles that recognized both military training and environmental values. They found that there were no specific management principles for these values.

Air pollutants and greenhouse gases often come from the same sources. Although we do not have measures of CO_2 and methane in our analysis, given that these sources of pollutants are also key contributors to climate warming, tackling air pollution from these sources will also mitigate climate change.[2] It was shown by Fishman et al. (1979) that the increase in tropospheric ozone from air pollution is an important contributor to global warming.

Furthermore, air pollution was previously thought to be just an urban or a local problem. But it is now understood that air pollution is transported across continents and ocean basins due to fast long-range transport, resulting in trans-oceanic and trans-continental plumes of atmospheric brown clouds. These atmospheric brown clouds in turn can give rise to regional cooling and warming effects (Ramanathan and Feng, 2009).

By quantifying and highlighting the amount of pollution generated by military training exercises on the environment, our research contributes to the discussion regarding the role that all sectors in an economy need to play in order to help mitigate global warming.

In this study, we analyze the effects of major military exercises in Australia on emissions. Our study is novel in being the first to focus on direct local emissions that are generated as a result of intensive military exercises. With the precise locations and dates of the military

[2]According to the World Bank, air pollution and climate change are two sides of the same coin, but they are typically addressed separately. They should be tackled jointly, with a focus on protecting peoples' health – particularly in low- and middle-income countries – to strengthen human capital and reduce poverty. See: https://www.worldbank.org/en/news/feature/2022/09/01/what-you-need-to-know-about-climate-change-and-air-pollution.

exercises known, using satellite data on pollutant emissions, our approach involves using econometric approaches on high-frequency data to estimate toxic emissions generated by a sudden period of intense military training activities. In particular, we adopt a machine learning (ML) algorithm and pair it with program evaluation tools to estimate treatment effects of military training activities. Our work follows the approach outlined in Prest et al. (2023) who show that nonexperimental treatment effects based on ML algorithms can replicate the true treatment effects from a randomized experiment. Based on analyzing high-frequency data on air pollution, we find that major military training exercises in 2015 and 2017 in Australia resulted in elevated levels of air pollution.

The rest of the paper is outlined as follows. Section 2 reviews the literature on military emissions and describes the four major military training exercises in Australia we focus on in our empirical analysis. Section 3 describes the data used for the analysis. The econometric approach we use to analyze our high-frequency data is explained in Section 4. Section 5 discusses the results and Section 6 concludes.

2 Background

The military is an important sector which is usually overlooked in discussions regarding climate policy, even though military operations are a major polluter to the environment. For example, the US Department of Defence is one of the world's worst polluters and its footprint dwarfs that of any corporation.[3] While Michaelowa and Koch (2001) raised the issue of military emissions more than two decades ago, policymakers have largely turned a blind eye to this question. Greenhouse gas (GHG) emissions of the military in peacetime and wars are highly relevant, especially as the world becomes ever more conflict-prone (Michaelowa et al., 2022).

The lion's share of direct emissions from the military is due to the consumption of liquid fuel and the operation of combat aircraft. Residential emissions from military bases and emissions of naval operations are also significant (Michaelowa et al., 2022; Belcher et al., 2020). Absent any change in military fuel use policy, the fuel consumption of the military forces will necessarily continue to generate high levels of greenhouse gases. Strikingly, despite the military having added the national security implications of climate change to its long list of national security concerns, the Pentagon does not acknowledge that its own fuel use is a major contributor to climate change (Crawford, 2019).

Several estimates of GHG emissions have been made for the military sectors of developed

[3]See the 2014 Newsweek article by Alexander Nazaryan: https://www.newsweek.com/2014/07/25/us-department-defence-one-worlds-biggest-polluters-259456.html.

countries. Parkinson (2020) focuses on calculating GHG emissions from the UK military-industrial sector, including the armed forces, arms industry and related employment. They find that the UK-based GHG emissions of the sector in the financial year 2017–18 using a production-based approach was greater than the carbon dioxide emissions of about 60 nations. Their figures do not include the GHG emissions related to impacts of weapons use on the battlefield. They state that such emissions could potentially be large, but are highly uncertain. Sparrevik and Utstøl (2020) apply an organizational perspective to assess life cycle GHG emissions for the Norwegian defense sector. They find that the estimated emissions suggest that the Norwegian defense sector is responsible for is 1.1% of the annual GHGs emitted in the Norwegian economy in 2017.

Previous research has utilized aggregate country level data to document a link between military spending and carbon dioxide emissions (Bildirici, 2017a,b,c; Bradford and Stoner, 2017) as well as military spending and economic growth (Bildirici, 2016, 2017a). The data on aggregate territorial carbon dioxide emissions are obtained from sources such as the Global Carbon Atlas and the World Bank's statistics on CO_2 emissions.[4]

There has also been research focusing on military metal pollution (Gębka et al., 2016; Lach et al., 2018; Skalny et al., 2021). Military activity is associated with environmental contamination involving heavy metals including chromium, copper, zinc, lead, and cadmium. Specifically, significant accumulation of metals has been observed in areas of battle fields, small-arm shooting ranges, artillery, mortar and rocket ranges, and grenade courts. Emission of metals into the environment from military activity occurs from gunshot residues containing high levels of metal-containing particles, as well as from use of artillery, grenades, and rockets (Barker et al., 2021).

There is very limited research documenting the impacts of military training exercises. An exception is Bobonis et al. (2020) who examine the effects of the end of bombing exercises in Puerto Rico. They find that the sudden end of bombing practices is associated with a 56–79% decrease in the incidence of congenital anomalies of nearby populations.

To our knowledge, there has been no work done that attempts to measure directly the impacts of military training on emissions.

2.1 Measuring Military Training Emissions

Air and particulate emissions are a growing concern in military activities. Although the environmental impacts of military training in ranges and training areas on soil, biomass, surface and groundwater contamination are generally well characterized at the military bases,

[4]See: http://www.globalcarbonatlas.org/en/CO2-emissions and https://data.worldbank.org/indicator/EN.ATM.CO2E.PC.

the air characterization from live firing of various weapons is not well understood and studied. In fact, there are only a few studies in the literature on gas and particle emissions from live firing of weapons (Diaz et al., 2012). Using an outdoor set-up to collect gases and particles from live firing with a 105-mm Howitzer gun, Diaz et al. (2012) find that the main combustion gases obtained after live firing are carbon monoxide, methane and ammonia.

With the development of remote sensing technology, satellite measurement is becoming one of the most effective approaches for high spatio-temporal resolution monitoring of air pollutants. Together with specific geographic knowledge of where military training is taking place, it is possible to measure the direct emissions from operations of military forces from the use of weapons and ammunition in training.

In this paper, we measure the impacts of major military training exercises undertaken by the Australian Defence Force (ADF) in 2015 and 2017 in a military training area located in Darwin, Australia.

2.2 Exercise Predator Walk 2015 (EPW2015)

The exercise is a three-week long training evolution held at the Mount Bundey Training Area, Northern Territory, Australia from 20 May to 10 June 2015. Working alongside the Australian Army were members of the US Marines to support the ground and aviation combat elements. It brought together 1,800 armed force members from the Australian Army, the US Marine Corps and a small number of soldiers from Malaysia.[5]

The US Marine Corps has been involved in the training as part of its Marine Rotation Force Darwin, announced by former prime minister Julia Gillard and US president Barrack Obama in 2011. One purpose of the bilateral training exercise is for the US Marines to become familiar with and improve their knowledge of the Australians technical and tactical procedures and standard operating procedures, which ultimately strengthens interoperability.

2.3 Talisman Sabre 2015 (TS2015)

Talisman Sabre is an Australian-United States biennial military exercise designed to improve readiness and enhance interoperability across the spectrum of conflict operations. At the time it took place, TS2015 broke new ground for scale, complexity and the diversity of participating forces. The exercise involved more than 30,000 personnel, 21 ships, 200 aircraft and three submarines. The major US elements, unsurprisingly accounting for the lion's share

[5]See: https://www.abc.net.au/news/2015-06-04/australian-army-and-us-marines-live-fire-training-in-nt-bush/6520328.

of combat power, were the Seventh Fleet's George Washington carrier strike group and the 31^{st} Marine Expeditionary Unit, both of which are based in Japan.

For the first time, the exercise also included personnel and equipment from Japan and New Zealand. The embedding of Japanese and New Zealand contingents with the US Marines and ADF respectively was the most noteworthy innovation to TS2015, lending the core bilateral format a loose quadrilateral aspect.

Hosted in both the Northern Territory and Queensland in Australia, TS2015 activities in the Northern Territory took place between 4 July and 19 July. The exercise was the sixth in the series and had a greater footprint in Northern Australia than in previous years. Our analysis focuses on activities at the Mount Bundey Training Area in the Northern Territory.[6]

2.4 Exercise Southern Jackaroo 2017 (ESJ2017)

Exercise Southern Jackaroo is the Australian Defence Force's annual trilateral exercise with elements from United States Pacific Command (US PACOM) and the Japanese Ground Self Defence Force (JGSDF) held from 18 May to 2 June 2017.

The 2017 exercise was also hosted at the Mount Bundey Training Area by the Australian 1st Brigade. It also included elements from the 5th Battalion of the Royal Australian Regiment, the Japanese Ground Self Defense Forces, and the U.S. Marine Rotational Force - Darwin. Set in the Australian outback, the ad hoc battle group coalesced to engage a fictitious enemy invasion of northern Australia. As part of Exercise Southern Jackaroo, each multinational partner participated in live-fire training exercises at the platoon and company levels.

2.5 Talisman Sabre 2017 (TS2017)

The 2017 edition of Talisman Sabre involving military personnel from Australia and the United States took place between late June and late July. Field training exercises incorporating live firing exercises were held at Mount Bundey from 21 June to 7 July 2017.

The exercise was designed primarily to maximize collective training benefits within a Combined Task Force setting, and to expose participants to a wide spectrum of military capabilities and training experiences, including live fire opportunities. An important aim of TS2017 was to improve training and interoperability between the Australian and US Armed Forces at the operational and tactical level. The land and air activities undertaken

[6]See: https://thediplomat.com/2015/07/australia-and-us-conclude-major-military-exercise-in-pacific-region/.

at Mount Bundey Training Area included parachute drops, special operations, live firing, close air support, and combat support.

In all 33,000 service personnel, 33 ships and 200 aircraft from Australia, the US, New Zealand, Japan and Canada participated in TS2017.

One key difference between TS2017 and TS2015 is that the land component of the training held in the Northern Territory (NT) was primarily held at the Mount Bundey Training Area rather than Bradshaw Field Training Area. However, there was overall a reduced presence of troops in the NT compared to TS2015.[7]

3 Data

The Mount Bundey Training Area is a large inland military training area located approximately 120km southeast of Darwin, Australia and spread over 117,300 hectares. The Mount Bundey Training Area used by the Australian government's Department of Defence is designed to support training, including mechanized battle group sized field firing and manoeuvre training and aerial bombing. It has field firing areas, high explosive impact areas, training sectors and infrastructure to support management and operational use.

Prior to the property being acquired by the Commonwealth of Australia in 1988, it was used for commercial grazing. The area was first used as a military training area in 1992. Infrastructure at the site is limited, and includes a road network, maintenance areas, a range control facility, a 200 person campsite and a number of support facilities, including purpose built ranges.

We focus on three pollutants in our empirical analysis: carbon monoxide (CO), particulate matter ($PM_{2.5}$), and sulfur dioxide (SO_2). The air pollutant and meteorological condition data we use for our analysis comes from the Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA-2).[8] The data are available at the hourly level and have a spatial resolution of 50 km x 62.5 km.

To determine the control area, we choose an area with the same latitude as the Mount Bundey Training Area as this area is more likely to share the same meteorological conditions.

[7] See the Talisman Sabre 2017 Public Environment Report.

[8] The surface CO concentration data are from the collection M2T1NXCHM. The precipitation and wind components (surface eastward wind and surface northward wind) data are from the collection M2T1NXFLX. Other meteorological data (surface pressure, specific humidity, surface air temperature) come from the collection M2I1NXLFO. As $PM_{2.5}$ data is not readily available, we follow Buchard et al. (2016) and construct $PM_{2.5}$ as follows: $PM_{2.5}=[DUST_{2.5}]+[SS_{2.5}]+[BC]+1.4\times[OC]+1.375\times[SO_4]$ where $DUST_{2.5}$ is dust surface mass concentration, $SS_{2.5}$ is sea salt surface mass concentration, BC is black carbon surface mass concentration, OC is organic carbon surface mass concentration, and SO_4 is SO_4 surface mass concentration. All of these data come from the collection M2T1NXAER.

In particular, latitude is regarded as the most important climatic control variable due to the effect it has on the amount of solar radiation reaching the earth's surface and the fact that different latitudes receive different amounts of solar radiation.[9] The chosen area is about 250 km away from the training area, so it is very unlikely to be affected by the training. In addition, this area shares the same elevation as the Mount Bundey. The centroids of the MERRA-2 cells identified for the Mount Bundey Training Area and the control area are highlighted in yellow in Figure 1.

Table 1 provides summary statistics for the training area and control area, including CO, $PM_{2.5}$ and SO_2 levels and various meteorological conditions.[10] As can be seen from the table, all meteorological conditions appear to be quite similar for both areas, suggesting that the two areas are similar. On the other hand, for the outcomes of interest, average CO, $PM_{2.5}$, and SO_2 levels over the sample period in the training area are higher compared to the control area.

4 Empirical Strategy

High frequency data allows greater flexibility in modeling the relationships between outcomes and treatment variables of interest. There is evidence that high-frequency data sets can provide causal estimates that are more reliable than those derived from lower-frequency data. Using data on hourly electricity consumption, Ghanem and Smith (2021) show that aggregating from the hourly to the monthly level can increase coefficient estimates by up to 60%.

In economics, the use of high-frequency data is most readily observable in the energy efficiency literature. The adoption of smart-meter technology increased the availability of hourly data in recent years and these new data sets have allowed the literature to provide a deeper understanding of the benefits of energy efficiency upgrades (Boomhower and Davis, 2020; Burlig et al., 2020; Novan and Smith, 2018).

High-frequency analysis is relevant in our context as our pollution data for the Mount

[9]It is also much more difficult to find a suitable control area to the north or south. The north of Mount Bundey Training Area is close to the ocean where pollution levels are often lower because pollutants are more likely to be carried away by ocean winds. As seen in Figure 1, there are also other military training sites to the south of Mount Bundey Training Area, which will have their own sources of pollution.

[10]Our statistics for wind speed and wind direction are calculated from the available data on northward wind (v wind component) and eastward wind (u wind component). In particular, wind speed is defined as the square root of the sum of v-squared and u-squared. Wind direction can be calculated using the trigonometric function: angle = arctan(v/u) as long as u is not equal to zero. Specifically, wind direction is calculated as 180 + (180/pi)*arctan(v/u), measured relative to North (360 degrees). For more details, see: https://disc.gsfc.nasa.gov/information/data-in-action?title=Derive%20Wind%20Speed%20and%20Direction%20With%20MERRA-2%20Wind%20Components.

Bundey Training Area are available at the hourly level. Our main variables of interest are hourly levels of CO, $PM_{2.5}$, and SO_2. One complication is that although the pollution data are available at the hourly level, the training dates of the military exercises are provided in terms of a range of dates. Furthermore, the exact times of the training that occurs during each of the days during the military exercise are not publicly known.

Following insights from Ghanem and Smith (2021), we decided to conduct the analysis at the hourly level as an attempt to exploit the benefits that the hourly data sets provide in terms of accounting for response heterogeneity and other features of the high-frequency outcome that low-frequency estimates are not able to capture.[11] More specifically, over the six-month period 1 April 2015 to 30 September 2015, we set a training dummy variable equal to 1 if the observation is between 12:00 a.m. to 11:59 p.m. (i.e., the whole day) during periods when training exercises took place, and zero otherwise.[12] The two military exercises we analyze which utilized the Mount Bundey Training Area in 2015 are EPW2015 (from 20 May to 10 June) and TS2015 (from 4 July to 19 July).[13] In a similar fashion, when analyzing the two military exercises which utilized the Mount Bundey Training Area in 2017 (ESJ2017 and TS2017), the training dummy variable is set to 1 if the observation is between 12:00 a.m. to 11:59 p.m. during the training period for ESJ2017 (18 May to 2 June) and TS2017 (21 June to 7 July), and takes a value of 0 otherwise.

An additional issue when it comes to the analysis is the complex and non-linear relationship between air pollution and meteorological conditions. A common practice in environmental economics literature is to include high-order polynomials of the weather variables to capture this relationship. However, as documented in the literature outside economics, this relationship could be far more complicated. To address this issue, a number of studies have suggested the use of ML approaches (e.g., random forest) to properly control for the weather effects (Grange et al., 2018; Grange and Carslaw, 2019). In data-rich settings, ML approaches can allow for more flexible functions that can capture complex interactions and nonlinearities in covariates. In this way, researchers do not need to make any assumptions about the relationship between air pollution and meteorological conditions.

[11] Another option is to aggregate the hourly pollution data and conduct the analysis of the effects of military training on pollution at a daily level or weekly level. However, such an approach will not maximize any gains from the high frequency data and can also result in biased estimated coefficients (Ghanem and Smith, 2021).

[12] The raw data is recorded in the UTC timezone. For our analysis, we convert the time stamp to GMT+9:30, which is the timezone in NT.

[13] It is possible that the Mount Bundey Training Area is utilized during other periods in the year for ADF training which is not publicly announced, which will induce measurement error into our analysis as our independent variable will not fully or accurately capture military training activities.

4.1 Program Evaluation with ML

In this paper, a ML algorithm is paired with program evaluation tools to estimate treatment effects. Such an approach is particularly valuable in settings where outcomes are cyclical or seasonal and detailed, high-frequency data are available. Using high-frequency hourly data in their analysis of energy efficiency, Burlig et al. (2020) find that their estimated panel fixed effects model is sensitive to outliers and to specification. Moreover, choosing the best set of control variables is challenging with so many possible candidate covariates.

Prest et al. (2023) show that nonexperimental treatment effects based on ML algorithms can replicate the true treatment effects from a randomized experiment in the context of electricity demand. Accordingly, we generate treatment effect estimates using three approaches following Prest et al. (2023). First, we apply standard two-way fixed effects (TWFE) regressions to the observed outcome of interest that highlight differences in observed outcomes for the treatment and control group.

Second, we use a random forest model to generate counterfactual predictions and use a "prediction-error" framework to estimate treatment effects comparing a treatment area to a control area. Here, we train our ML model using pretreatment hourly data, weather, various calendar fixed effects, and interactions among these variables from the control area (see Figure 1). Next, we make out-of-sample predictions for pollution levels during the treatment period in both the treatment and control areas. We then compare these counterfactuals to observed pollution levels in a fixed-effects regression framework. Put another way, the predicted outcomes serve as counterfactuals. The intuition behind this approach is if, conditional on fixed effects, the treated area exhibits higher pollution levels beyond what would otherwise be predicted by a ML model —that is, a positive prediction error —and that this positive prediction error is more pronounced during the treated time periods, then this provides evidence for a positive treatment effect.

Third, we use only data on the treated area for our analysis. It was shown by Prest et al. (2023) that it is possible to use ML approaches to recover experimental benchmark impacts even without the use of a comparison group. In other words, in our context, estimation will be based solely on using data on the treated area, with predicted pollution levels serving as the counterfactual for the treatment area. An important caveat in following this approach of Prest et al. (2023) in our context is that for the pre-treatment period in the treated area, Mount Bundey will still have been used as a military training site by the Australian Defence Forces. This implies that the counterfactual that is created using pre-treatment data from the treated area is not ideal. Nevertheless, we apply this approach to see if any interesting results emerge.

4.2 Prediction Error Framework

We compare the predicted counterfactuals in the treated and control areas generated using data from the control area as our main approach. In the first stage, we build the random forest (RF) model using hourly-level data from eight areas surrounding the control area (i.e., the centroid of eight neighboring areas, denoted in orange color in Figure 1). The training data comprises of data that is from two years prior to each training event. Specifically, for estimating the effects of military training in 2015, we use pretreatment data over the period 2012 - 2014. For analyzing the impacts of military training in 2017, we use pretreatment data over the period 2014 - 2016.

This "fake" control area is chosen as it is expected to be similar to the actual control area in terms of pollution levels and meteorological conditions in the absence of military training. This also helps to ensure that the counterfactual predictions are not improperly derived using treatment period data. The algorithm splits the sample data into a training set of 70 percent of the data and a test set of 30 percent of the data to test the model's performance.

The predictors used to train the RF model include the 12-hour lag of the concentration for each pollutant of interest, and all meteorological variables (current and 24-h lag of wind speed, wind direction, temperature, humidity, pressure, and precipitation). We also control for seasonality using temporal variables (month, day of the week, and hour of the day and weekend interactions) in our RF model. In the analysis without a comparison group, we use the hourly-level data of the treated area over the period 2012 - 2014 (or 2014 - 2016) to build the RF model.

For estimating the effects of training in 2015, once the RF model is built using data from the "fake" control area from 2012 - 2014, the original temporal variables and weather data of the treatment area and actual control area are used to predict the pollutant concentration levels in these areas for 2015. We then generate a "prediction error" for each hour reflecting the difference between observed pollution levels in 2015 and the counterfactual. Prediction errors for the 2017 military training events are obtained in a similar fashion using pretreatment data from 2014 - 2016.

RF models generally require little parameter tuning from users, but there are three important parameters that one must decide on: the number of trees, the number of predictors tried at each split, and the tree depth. Prest et al. (2023) find that using the default "off-the-shelf" model options in ML software packages appear to work fine.

In our algorithm, we choose 500 trees as the model error remains stable around this

number.[14] In terms of the number of variables to be tested at each split, the default value is one-third of the number of predictors in the model for RF regression. In our case, we use a value of 5 in the main specification. The package `randomForest` in R does not allow users to specify the depth of the tree directly. However, researchers could indirectly reduce tree depth using one of the three options: using more data, reducing the minimum size of the terminal nodes, and reducing the number of variables tried at each split. We use the default minimum size of the terminal nodes of 5.

In the second stage, we use the prediction error (observed values minus RF predicted values) as the dependent variable.

$$\text{Pred}\tilde{\text{E}}\text{rror}_{ihd} = \beta_0 + \beta_1 \text{Training}_{hd} + \beta_2 \text{Treated}_i + \beta_3 \text{Training}_{hd} \times \text{Treated}_i + \theta_t + \epsilon_{ihd} \quad (1)$$

where $\text{Pred}\tilde{\text{E}}\text{rror}_{ihd}$ is the prediction error from the first stage in area i at hour h of day d. The term Treated_i is an indicator taking a value of one if area i is in the training site and zero otherwise. Training_{hd} takes a value of one during the hours 12:00 a.m. to 11:59 p.m. of the day if the day falls within the training period and zero otherwise (we later vary the hours of the day where military training is most likely to have taken place). In the model, β_3 is the coefficient of interest. The term θ_t represents time fixed effects, including hour-of-sample fixed effects. We bootstrap the standard errors using 500 repetitions in the second stage in order to reflect the standard error from the first stage.

5 Results

5.1 Main Results

Figure 2 depicts the hourly variation in CO (top panel), $PM_{2.5}$ (middle panel) and SO_2 (bottom panel) in the treated and control areas during the military training days in 2015. This graph helps to reveal the within-day fluctuations in CO, $PM_{2.5}$ and SO_2 levels and also helps to indicate the hours in the day military training most likely occurred in the treatment area (the exact timing of the military exercises during each day is not publicly available). There is an indication that pollutant levels are higher after in the evening after 6 p.m. and continue past midnight till early next morning. Similarly, the hourly variation in pollutants for the treated and control areas for 2017 is shown in Figure 3.

The results of analyzing the effects of the two training exercises on CO in 2015 based on

[14] The number of trees in the model should stabilize the model error. Higher number of trees beyond this value will have marginal benefit as there is no significant decrease in the model error and it consumes more computing time.

Equation (1) are provided in the top panel of Table 2.[15] In column (1), the treatment effect is estimated by comparing the difference in observed pollutant levels between the treated and control areas. Alternatively, column (2) uses the prediction-error framework to estimate the treatment effect. In both cases, the estimated treatment effects are statistically significant and similar in magnitude. According to column (2), military training has led to an increase in CO of 10.658 parts-per-billion volume (ppbv) relative to a mean of 99.3 ppbv, which translates to an increase of approximately 10 percent.

In the middle and bottom panels of Table 2, the corresponding results of the effects of military training in 2015 for $PM_{2.5}$ and SO_2 are presented, respectively. The results in column (2) suggest that there are no significant effects on $PM_{2.5}$. This is not implausible as $PM_{2.5}$ is not a main pollutant emitted from military weapons and live firing and might only be present when certain weapons (e.g., tanks) are used. The result for SO_2 is, however, surprising as the point estimate is negative and statistically significant. This implies that military training reduced the levels of SO_2. When we investigate this result further, checking to see if we obtain parallel trends for SO_2 before the training occurs, we find that the reason this strange result arises is likely because we are not able to obtain parallel trends, unlike the case for CO and $PM_{2.5}$ (see Figures A7 and A8 in the appendix).[16]

5.2 Results Using Treated Observations

The estimated treatment effects using only treatment group observations are provided in columns (3) and (4) of Table 2. Specifically, column (3) compares pollutant levels during the training periods in the treated area with pollutant levels during non-training periods in the treated area. As discussed in Prest et al. (2023), such an estimate that is solely based on differences between treatment periods and nontreatment periods for the treated tends to perform poorly and in their study does not recover the experimental benchmark results. In our case, the estimated effect is smaller than what is observed when using the control group. In column (4), the RF approach is used only on the treated area to create the counterfactual. This approach worked out well in Prest et al. (2023). Recall that we had previously mentioned that in our context, the counterfactual that is created using pre-treatment data from the treated area is not necessarily well-constructed as military training almost surely occurred in the same location in previous years. Here, we find an estimated treatment effect of 14.29 ppbv for CO which is slightly larger than what was estimated using

[15] The top panel of Figures A1 to A6 in the appendix depicts the actual and predicted values for the treated and control areas, while the bottom panel presents the treatment-control differences using observed data and predicted data.

[16] Despite trying many different specifications and tuning parameters for the ML model, this issue stubbornly persisted.

a control group in columns (1) and (2).[17]

The corresponding results of the effects of military training in 2015 for $PM_{2.5}$ are presented in the middle panel of Table 2. Although the positive significant effects of military training on $PM_{2.5}$ levels are small in absolute magnitude and are between 1 to 1.5 $\mu g/m^3$ (columns (1) and (4)), in a low pollution environment like the northern part of Australia, they are similar in magnitude to the effects on CO when expressed in percentage terms. For example, the effect in column (1) is 1.579/12.99 = 12.1 percent.

Finally, in the bottom panel of Table 2, estimates for SO_2 are presented. The estimated effect is now positive, but this instability of results between columns (2) and (3) is likely due to our inability to obtain parallel trends for SO_2.

Table 3 provides the estimated treatment effects of military training in 2017. While there are again positive and significant effects using the approach involving ML and the use of a control group, the magnitude of the effects relative to mean levels are slightly smaller as compared to 2015 (9.8 percent for CO and 6.9 percent for $PM_{2.5}$).

5.3 Results Using Different Training Hours

While the official training periods for the military training exercises are announced in public media, it is not known when exactly during the day training activities are held. Earlier, based on Figures 2 and 3, we had already indicated that there are times during the day that the intensive military activities most likely took place. Therefore, in this section, we re-estimate our models by defining Training_{hd} to take a value of one during the hour 6:00 p.m. to 6:00 a.m. if the day falls within the training period and zero otherwise.

The results are shown in Table 4 and Table 5. As expected, the magnitudes of the treatment effects increase considerably, for example, with the estimate for CO emissions larger than 25 percent relative to their mean values in 2015.

5.4 Results Using Different Training Days

It is worth noting that the details on any preparatory activities before the official training dates are not reported. For example, spikes in CO and $PM_{2.5}$ can be seen in Appendix Figures A1 and A2. This is likely due to Mount Bundey being used for some unreported military training prior to the actual publicly announced training exercise. Any training activities that occur outside the official training window that lead to higher pollution levels

[17] Note that we do not include hour-of-sample fixed effects here in the absence of a control group due to its perfect collinearity with the treatment variables.

will lead to an underestimate of the treatment effect in our model setup due to measurement error in the treatment variable.

Unobserved factors are also likely to be at play affecting SO_2 levels in non-training periods, as reflected in Figures A7 and A8 in the appendix.

6 Conclusions

The military is a major source of air pollution, primarily due to its extensive use of fossil fuels to operate military vehicles and military jets. However, this point is often overlooked in discussions on air pollution. In addition, the emissions generated from live firing of various military weapons is not well understood and studied. In this paper, we analyse the effects of major joint military exercises held in an inland military training area in a remote part of Australia on levels of air pollution. Involving thousands of soldiers, these exercises are simulations of war activities and provide natural experiments where there are sudden bursts of intense military activities at local sites.

We utilize high-frequency fixed effects models to estimate how intense military training activities affect levels of CO, $PM_{2.5}$, and SO_2. Our causal estimates are based on making comparisons with a control area that was not subject to the military training activities located several hundred kilometers to the east and on the same latitude.

For the two training exercises examined in 2015, we find a positive and significant effect of military training on levels of CO and $PM_{2.5}$. For the two major training exercises examined in 2017, our results are also positive but more muted. We were not able to obtain reliable estimates on the effects on SO_2 levels using our estimation strategy.

In conclusion, our research sheds light on an important but often overlooked aspect of environmental impact: the emissions generated by military training activities. By quantifying the pollution released during these exercises, we contribute valuable insights to the broader conversation on combating climate change. While our study addresses emissions from training exercises, the military's environmental footprint extends beyond this and includes substantial emissions from various operational activities such as fuel consumption by vehicles, tanks, jets, and warships. Moving forward, comprehensive assessments of military-related emissions are essential for developing effective strategies to mitigate environmental degradation and advance global sustainability goals. Accordingly, acknowledging and addressing all sources of pollution, including those stemming from military operations, is essential.

References

Barker, A. J., Clausen, J. L., Douglas, T. A., Bednar, A. J., Griggs, C. S., and Martin, W. A. (2021). Environmental impact of metals resulting from military training activities: A review. *Chemosphere*, 265:129110.

Belcher, O., Bigger, P., Neimark, B., and Kennelly, C. (2020). Hidden carbon costs of the "everywhere war": Logistics, geopolitical ecology, and the carbon boot-print of the US military. *Transactions of the Institute of British Geographers*, 45(1):65–80.

Bildirici, M. E. (2016). The defense industry sector, economic growth, and energy consumption in 20 emerging countries. *The Journal of Energy and Development*, 42(1/2):147–159.

Bildirici, M. E. (2017a). The causal link among militarization, economic growth, CO2 emission, and energy consumption. *Environmental Science and Pollution Research*, 24(5):4625–4636.

Bildirici, M. E. (2017b). CO2 emissions and militarization in G7 countries: Panel cointegration and trivariate causality approaches. *Environment and Development Economics*, 22(6):771–791.

Bildirici, M. E. (2017c). The effects of militarization on biofuel consumption and CO2 emission. *Journal of Cleaner Production*, 152:420–428.

Bobonis, G. J., Stabile, M., and Tovar, L. (2020). Military training exercises, pollution, and their consequences for health. *Journal of Health Economics*, 73:102345.

Boomhower, J. and Davis, L. (2020). Do energy efficiency investments deliver at the right time? *American Economic Journal: Applied Economics*, 12(1):115–139.

Bradford, J. H. and Stoner, A. M. (2017). The treadmill of destruction in comparative perspective: A panel study of military spending and carbon emissions, 1960-2014. *Journal of World-Systems Research*, 23(2):298–325.

Buchard, V., Da Silva, A., Randles, C., Colarco, P., Ferrare, R., Hair, J., Hostetler, C., Tackett, J., and Winker, D. (2016). Evaluation of the surface PM2.5 in version 1 of the NASA MERRA aerosol reanalysis over the United States. *Atmospheric Environment*, 125:100–111.

Burlig, F., Knittel, C., Rapson, D., Reguant, M., and Wolfram, C. (2020). Machine learning from schools about energy efficiency. *Journal of the Association of Environmental and Resource Economists*, 7(6):1181–1217.

Crawford, N. C. (2019). *Pentagon fuel use, climate change, and the costs of war*. Watson Institute, Brown University.

Diaz, E., Svard, S., and Poulin, I. (2012). Air residues from live firings during military training. *Air Pollution XX*, 157:399.

Fishman, J., Ramanathan, V., Crutzen, P., and Liu, S. (1979). Tropospheric ozone and climate. *Nature*, 282(5741):818–820.

Gębka, K., Bełdowski, J., and Bełdowska, M. (2016). The impact of military activities on the concentration of mercury in soils of military training grounds and marine sediments. *Environmental Science and Pollution Research*, 23(22):23103–23113.

Ghanem, D. and Smith, A. (2021). What are the benefits of high-frequency data for fixed effects panel models? *Journal of the Association of Environmental and Resource Economists*, 8(2):199–234.

Gould, K. A. (2007). The ecological costs of militarization. *Peace Review*, 19(3):331–334.

Grange, S. K. and Carslaw, D. C. (2019). Using meteorological normalisation to detect interventions in air quality time series. *Science of The Total Environment*, 653:578–588.

Grange, S. K., Carslaw, D. C., Lewis, A., Boleti, E., and Heuglin, C. (2018). Random forest meteorological normalisation models for Swiss PM10 trend analysis. *Atmospheric Chemistry and Physics Discussions*.

Hultman, L., Kathman, J., and Shannon, M. (2014). Beyond keeping peace: United Nations effectiveness in the midst of fighting. *American Political Science Review*, 108(4):737–753.

Lach, K., Bernatíková, š., Frišhansová, L., Klouda, K., and Mička, V. (2018). Study of air contamination by heavy metals at firing ranges. *WIT Transactions on Ecology and the Environment*, 230:29–39.

Lawrence, M. J., Stemberger, H. L., Zolderdo, A. J., Struthers, D. P., and Cooke, S. J. (2015). The effects of modern war and military activities on biodiversity and the environment. *Environmental Reviews*, 23(4):443–460.

Michaelowa, A. and Koch, T. (2001). Military emissions, armed conflicts, border changes and the Kyoto Protocol. *Climatic Change*, 50(4):383–394.

Michaelowa, A., Koch, T., Charro, D., Gameros, C., Burton, D., and Lin, H.-C. (2022). Military and conflict-related emissions: To Glasgow and beyond. *Perspectives Climate Group*.

Novan, K. and Smith, A. (2018). The incentive to overinvest in energy efficiency: Evidence from hourly smart-meter data. *Journal of the Association of Environmental and Resource Economists*, 5(3):577–605.

Parkinson, S. (2020). The environmental impacts of the UK military sector. *Scientists for Global Responsibility*.

Prest, B. C., Wichman, C. J., and Palmer, K. (2023). RCTs against the machine: Can machine learning prediction methods recover experimental treatment effects? *Journal of the Association of Environmental and Resource Economists*, 10(5):1231–1264.

Ramanathan, V. and Feng, Y. (2009). Air pollution, greenhouse gases and climate change: Global and regional perspectives. *Atmospheric environment*, 43(1):37–50.

Rohner, D. (2024). Mediation, military, and money: The promises and pitfalls of outside interventions to end armed conflicts. *Journal of Economic Literature*, 62(1):155–195.

Skalny, A. V., Aschner, M., Bobrovnitsky, I. P., Chen, P., Tsatsakis, A., Paoliello, M. M., Djordevic, A. B., and Tinkov, A. A. (2021). Environmental and health hazards of military metal pollution. *Environmental Research*, 201:111568.

Sparrevik, M. and Utstøl, S. (2020). Assessing life cycle greenhouse gas emissions in the Norwegian defence sector for climate change mitigation. *Journal of Cleaner Production*, 248:119196.

Svenningsen, S. R., Levin, G., and Perner, M. L. (2019). Military land use and the impact on landscape: A study of land use history on Danish defence sites. *Land Use Policy*, 84:114–126.

Zentelis, R. and Lindenmayer, D. (2015). Bombing for biodiversity—enhancing conservation values of military training areas. *Conservation Letters*, 8(4):299–305.

Zentelis, R., Lindenmayer, D., Roberts, J. D., and Dovers, S. (2017). Principles for integrated environmental management of military training areas. *Land Use Policy*, 63:186–195.

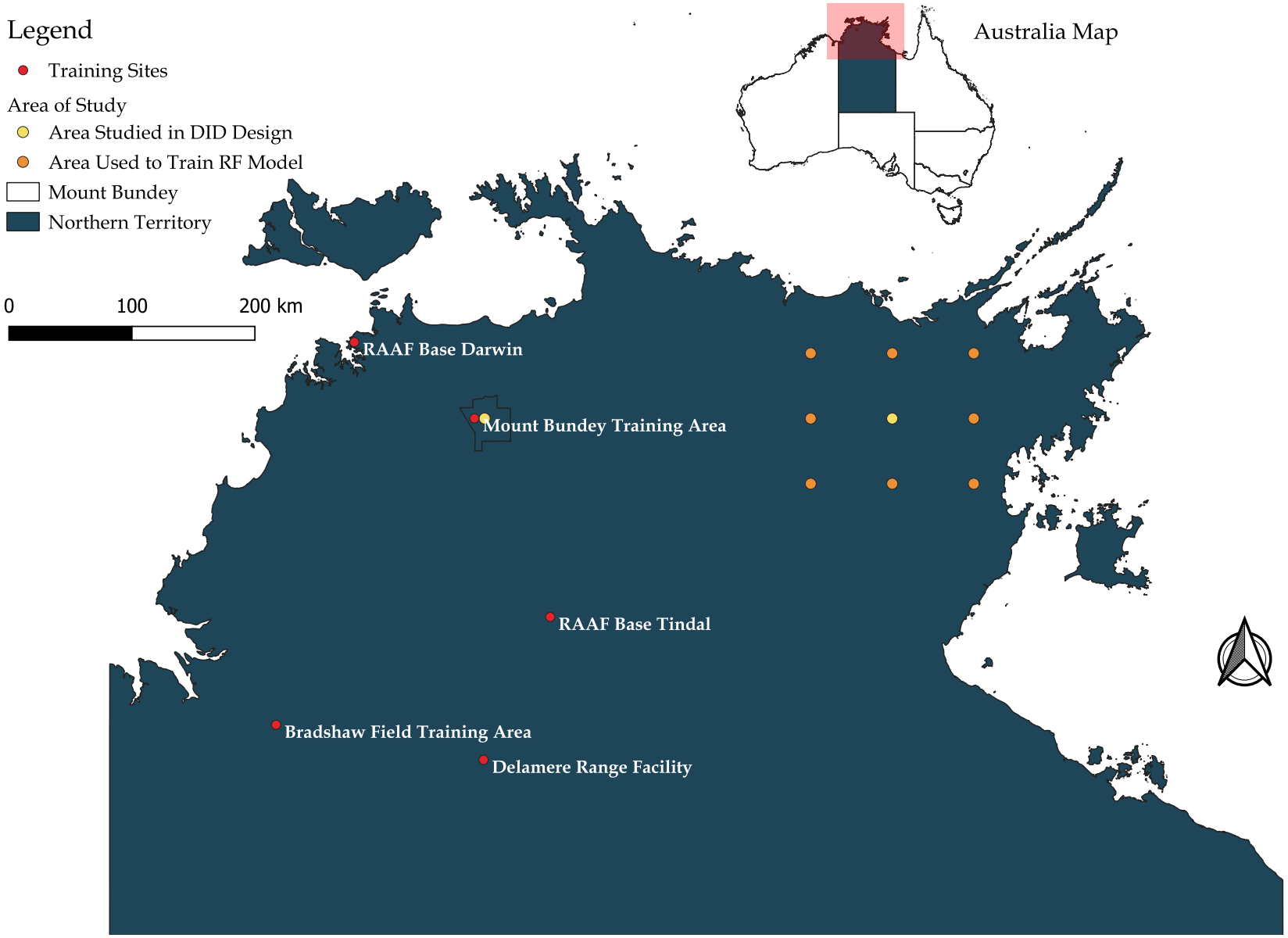


Figure 1: Training Sites in Northern Territory

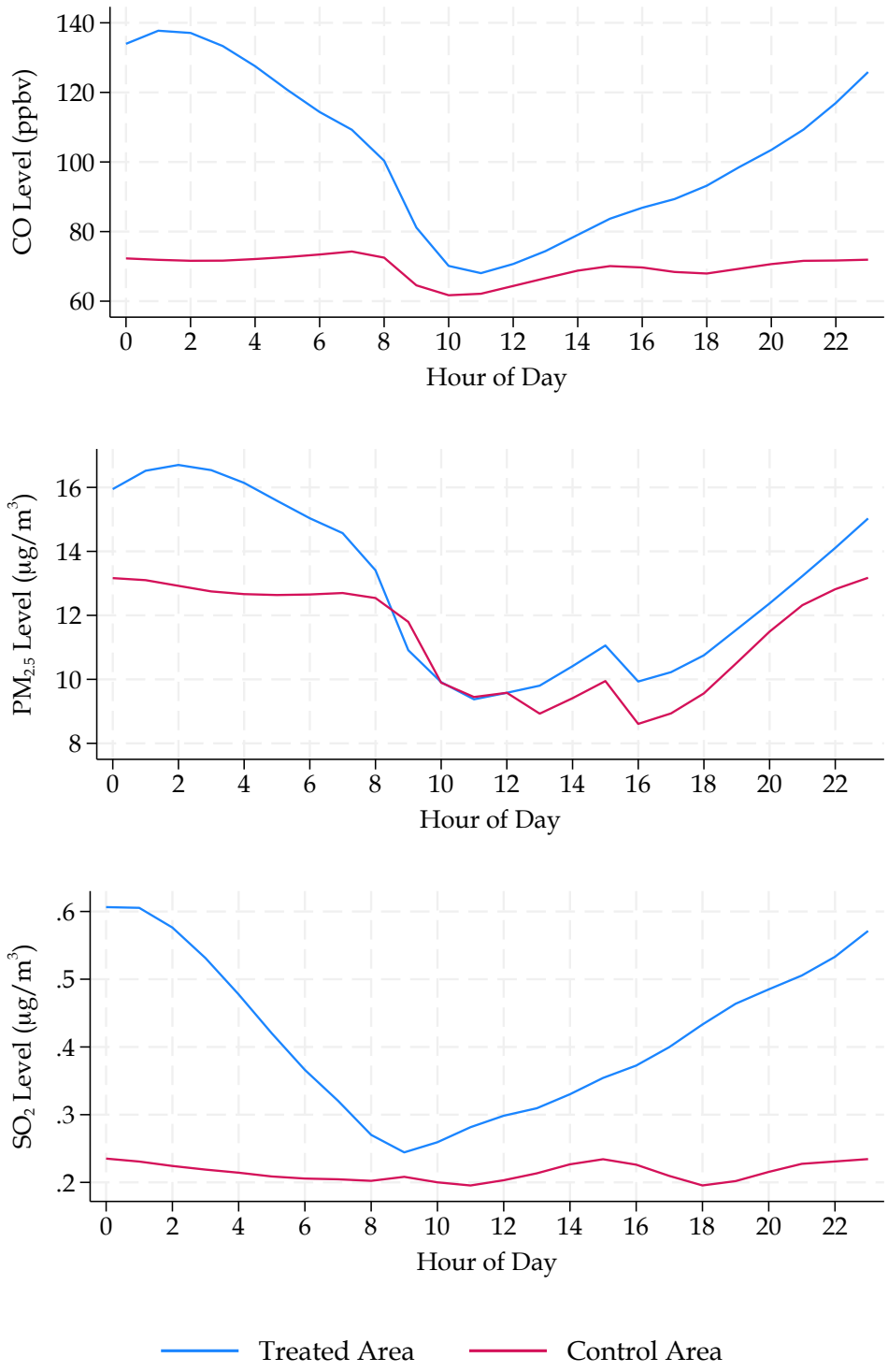


Figure 2: 2015 Training Day Average Pollution Level by Hour of Day

Notes. This plot shows average pollution level at each hour for the training area and control area on the training days (i.e., 20 May - 10 June 2015 and 4 - 19 July 2015).

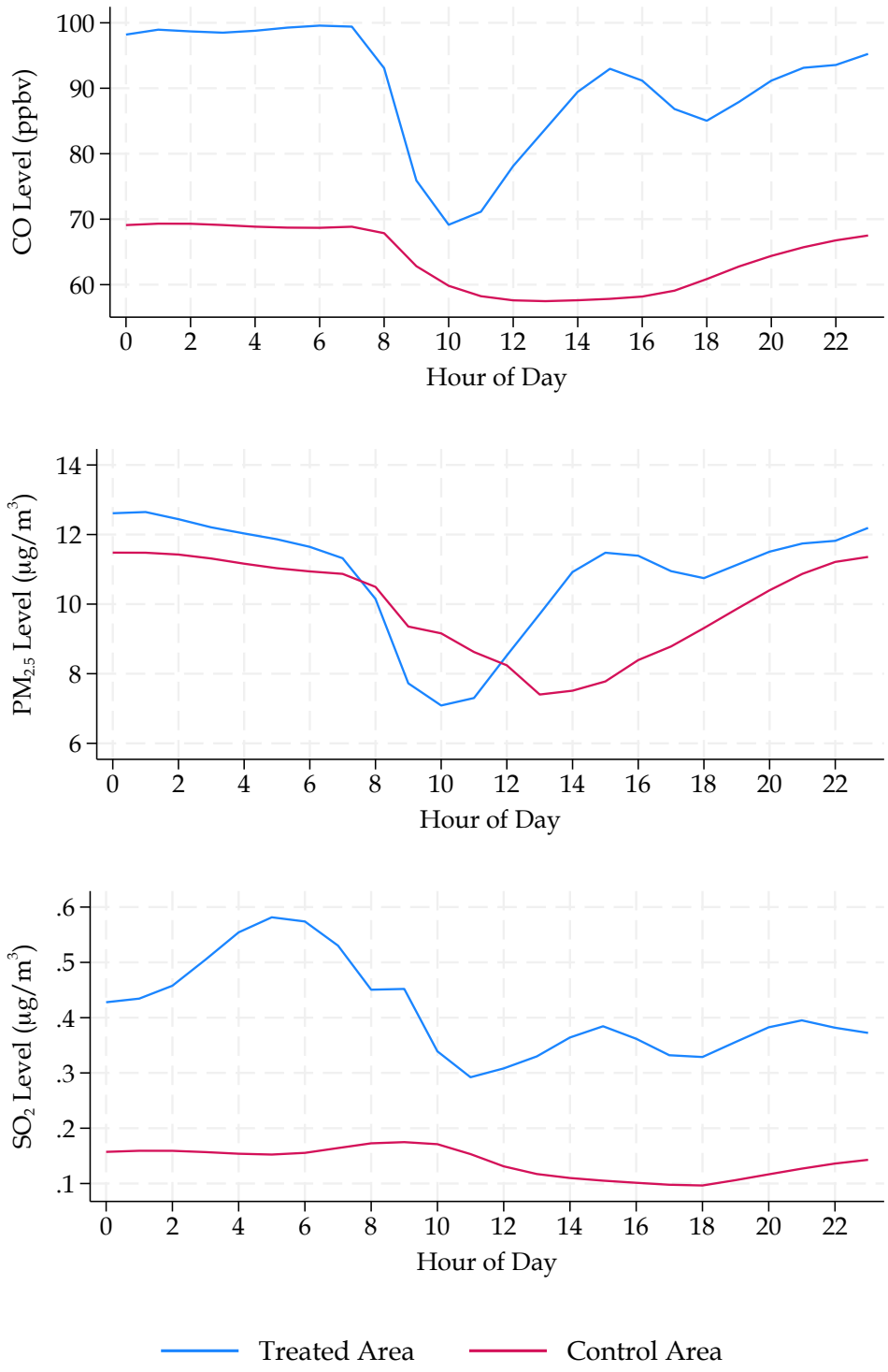


Figure 3: 2017 Training Day Average Pollution Level by Hour of Day

Notes. This plot shows average pollution level at each hour for the training area and control area on the training days (i.e., 18 May - 2 June 2017 and 21 June - 7 July 2017).

Table 1: Summary Statistics

	Treatment Area			Control Area		
VARIABLES	N	mean	sd	N	mean	sd
Predator Walk 2015 and Talisman Sabre 2015						
CO (ppbv)	4,392	99.30	42.48	4,392	74.64	20.04
$PM_{2.5}$ ($\mu g/m^3$)	4,392	12.99	6.676	4,392	11.64	5.404
SO_2 ($\mu g/m^3$)	4,392	0.553	1.207	4,392	0.224	0.225
Humidity	4,392	0.00946	0.00390	4,392	0.0120	0.00314
Temperature (K)	4,392	299.6	4.659	4,392	297.5	3.891
Wind direction (degree)	4,392	238.2	136.0	4,392	299.5	90.59
Wind speed (m/s)	4,392	5.622	2.300	4,392	7.046	1.830
Precipitation (mm/day)	4,392	0.911	10.50	4,392	0.424	2.120
Pressure (Pa)	4,392	100,721	247.3	4,392	100,407	234.2
Southern Jackaroo 2017 and Talisman Sabre 2017						
CO (ppbv)	4,392	88.67	28.35	4,392	68.05	12.48
$PM_{2.5}$ ($\mu g/m^3$)	4,392	11.58	6.191	4,392	11.17	5.444
SO_2 ($\mu g/m^3$)	4,392	0.367	0.526	4,392	0.154	0.123
Humidity	4,392	0.0107	0.00380	4,392	0.0135	0.00308
Temperature (K)	4,392	300.3	4.412	4,392	297.8	3.537
Wind direction (degree)	4,392	235.8	141.8	4,392	304.8	80.89
Wind speed (m/s)	4,392	5.831	2.237	4,392	6.894	1.733
Precipitation (mm/day)	4,392	0.926	8.526	4,392	1.544	8.150
Pressure (Pa)	4,392	100,599	230.3	4,392	100,298	208.3

Notes. Observations are at the hourly level and restricted to the period 1 April to 30 September. Wind speed and wind direction are calculated using the northward wind and eastward wind component.

Table 2: Random Forest Prediction Error TWFE Regressions for Military Training in 2015

	With Control Group		Treatment Only	
	Actual	Pred Error	Actual	Pred Error
	(1)	(2)	(3)	(4)
Panel A: CO				
Training_d x Treated_i	10.212***	10.658***	7.760***	7.184***
	(2.898)	(3.337)	(2.153)	(2.210)
Observations	8,784	8,784	4,392	4,392
R-squared	0.666	0.538	0.185	0.066
Hour-of-Sample FE	YES	YES	NO	NO
Hour-of-Day FE	NO	NO	YES	YES
Day-of-Week FE	NO	NO	YES	YES
Panel B: $PM_{2.5}$				
Training_d x Treated_i	1.579***	0.334	0.492	1.083***
	(0.394)	(0.434)	(0.332)	(0.318)
Observations	8,784	8,784	4,392	4,392
R-squared	0.727	0.593	0.184	0.032
Hour-of-Sample FE	YES	YES	NO	NO
Hour-of-Day FE	NO	NO	YES	YES
Day-of-Week FE	NO	NO	YES	YES
Panel C: SO_2				
Training_d x Treated_i	0.032	-0.162***	0.133***	0.019
	(0.035)	(0.037)	(0.027)	(0.022)
Observations	8,784	8,784	4,392	4,392
R-squared	0.640	0.507	0.321	0.017
Hour-of-Sample FE	YES	YES	NO	NO
Hour-of-Day FE	NO	NO	YES	YES
Day-of-Week FE	NO	NO	YES	YES

Notes. The variable "Training_d" takes value of one if the observation falls within training period. Columns (1) and (2) are estimated using the control group. Columns (3) and (4) are estimated using the treatment group only. For regressions using the treatment group, the reported estimates are the coefficient of variable "Training_d" only, not interaction of "Training_d" and "Treated_i." In the "Actual" columns, the dependent variables is the pollution concentration at the hourly level (CO or $PM_{2.5}$ or SO_2). All regressions under these columns include the current and 24-h lag of weather variables (i.e., temperature, humidity, pressure, precipitation, wind direction, and wind speed). In the "Pred Error" columns, the dependent variables is the residual obtained from the first stage based on the difference between the actual data and the prediction made by the RF model. In column (2), the RF model is built from the data of eight neighboring area surrounding the control area over the period of 2012 - 2014. In column (4), the RF model is built from the data of the training site over the period of 2012 - 2014. Standard errors, in parentheses, are bootstrapped 500 times. ***, **, and * Significance level at 1%, 5% and 10%.

Table 3: Random Forest Prediction Error TWFE Regressions for Military Training in 2017

	With Control Group		Treatment Only	
	Actual	Pred Error	Actual	Pred Error
	(1)	(2)	(3)	(4)
Panel A: CO				
Training_d x Treated_i	3.918**	8.680***	3.081**	-0.073
	(1.571)	(1.568)	(1.234)	(1.101)
Observations	8,784	8,784	4,392	4,392
R-squared	0.742	0.586	0.302	0.030
Hour-of-Sample FE	YES	YES	NO	NO
Hour-of-Day FE	NO	NO	YES	YES
Day-of-Week FE	NO	NO	YES	YES
Panel B: $PM_{2.5}$				
Training_d x Treated_i	0.555*	0.796***	-0.606**	0.093
	(0.292)	(0.279)	(0.268)	(0.235)
Observations	8,784	8,784	4,392	4,392
R-squared	0.792	0.651	0.215	0.024
Hour-of-Sample FE	YES	YES	NO	NO
Hour-of-Day FE	NO	NO	YES	YES
Day-of-Week FE	NO	NO	YES	YES
Panel C: SO_2				
Training_d x Treated_i	0.092**	0.069*	0.056**	0.096***
	(0.039)	(0.038)	(0.023)	(0.022)
Observations	8,784	8,784	4,392	4,392
R-squared	0.650	0.518	0.249	0.028
Hour-of-Sample FE	YES	YES	NO	NO
Hour-of-Day FE	NO	NO	YES	YES
Day-of-Week FE	NO	NO	YES	YES

Notes. The variable "Training_d" takes value of one if the observation falls within training period. Columns (1) and (2) are estimated using the control group. Columns (3) and (4) are estimated using the treatment group only. For regressions using the treatment group, the reported estimates are the coefficient of variable "Training_d" only, not interaction of "Training_d" and "Treated_i." In the "Actual" columns, the dependent variables is the pollution concentration at the hourly level (CO or $PM_{2.5}$or SO_2). All regressions under these columns include the current and 24-h lag of weather variables (i.e., temperature, humidity, pressure, precipitation, wind direction, and wind speed). In the "Pred Error" columns, the dependent variables is the residual obtained from the first stage based on the difference between the actual data and the prediction made by the RF model. In column (2), the RF model is built from the data of eight neighboring area surrounding the control area over the period of 2014 - 2016. In column (4), the RF model is built from the data of the training site over the period of 2014 - 2016. Standard errors, in parentheses, are bootstrapped 500 times. ***, **, and * Significance level at 1%, 5% and 10%.

Table 4: Random Forest Prediction Error TWFE Regressions for Military Training in 2015: Restricting Definition of Training to 6 PM to 6 AM the Following Day

	With Control Group		Treatment Only	
	Actual	Pred Error	Actual	Pred Error
	(1)	(2)	(3)	(4)
Panel A: CO				
Training_d x Treated_i	25.283***	29.391***	19.148***	20.901***
	(4.918)	(5.345)	(3.680)	(3.817)
Observations	8,784	8,784	4,392	4,392
R-squared	0.674	0.553	0.196	0.085
Hour-of-Sample FE	YES	YES	NO	NO
Hour-of-Day FE	NO	NO	YES	YES
Day-of-Week FE	NO	NO	YES	YES
Panel B: $PM_{2.5}$				
Training_d x Treated_i	2.991***	0.952	1.924***	2.929***
	(0.660)	(0.677)	(0.510)	(0.522)
Observations	8,784	8,784	4,392	4,392
R-squared	0.729	0.594	0.190	0.048
Hour-of-Sample FE	YES	YES	NO	NO
Hour-of-Day FE	NO	NO	YES	YES
Day-of-Week FE	NO	NO	YES	YES
Panel C: SO_2				
Training_d x Treated_i	0.161***	-0.046	0.286***	0.106***
	(0.051)	(0.046)	(0.047)	(0.033)
Observations	8,784	8,784	4,392	4,392
R-squared	0.641	0.506	0.324	0.018
Hour-of-Sample FE	YES	YES	NO	NO
Hour-of-Day FE	NO	NO	YES	YES
Day-of-Week FE	NO	NO	YES	YES

Notes. The variable "Training_d" takes value of one if the observation falls within training period and within the range of midnight to 6 a.m. or 6 p.m. 'till midnight. Columns (1) and (2) are estimated using the control group. Columns (3) and (4) are estimated using the treatment group only. For regressions using the treatment group, the reported estimates are the coefficient of variable "Training_d" only, not interaction of "Training_d" and "Treated_i." In the "Actual" columns, the dependent variables is the pollution concentration at the hourly level (CO or $PM_{2.5}$ or SO_2). All regressions under these columns include the current and 24-h lag of weather variables (i.e., temperature, humidity, pressure, precipitation, wind direction, and wind speed). In the "Pred Error" columns, the dependent variables is the residual obtained from the first stage based on the difference between the actual data and the prediction made by the RF model. In column (2), the RF model is built from the data of eight neighboring area surrounding the control area over the period of 2012 - 2014. In column (4), the RF model is built from the data of the training site over the period of 2012 - 2014. Standard errors, in parentheses, are bootstrapped 500 times. ***, **, and * Significance level at 1%, 5% and 10%.

Table 5: Random Forest Prediction Error TWFE Regressions for Military Training in 2017: Restricting Definition of Training to 6 PM to 6 AM the Following Day

	With Control Group		Treatment Only	
	Actual	Pred Error	Actual	Pred Error
	(1)	(2)	(3)	(4)
Panel A: CO				
Training_d x Treated_i	3.471*	10.642***	3.234**	-1.354
	(1.896)	(1.801)	(1.507)	(1.248)
Observations	8,784	8,784	4,392	4,392
R-squared	0.742	0.586	0.301	0.031
Hour-of-Sample FE	YES	YES	NO	NO
Hour-of-Day FE	NO	NO	YES	YES
Day-of-Week FE	NO	NO	YES	YES
Panel B: $PM_{2.5}$				
Training_d x Treated_i	0.455	0.788**	-0.586*	-0.025
	(0.357)	(0.307)	(0.349)	(0.302)
Observations	8,784	8,784	4,392	4,392
R-squared	0.792	0.651	0.215	0.024
Hour-of-Sample FE	YES	YES	NO	NO
Hour-of-Day FE	NO	NO	YES	YES
Day-of-Week FE	NO	NO	YES	YES
Panel C: SO_2				
Training_d x Treated_i	0.121**	0.106**	0.028	0.080***
	(0.053)	(0.052)	(0.032)	(0.030)
Observations	8,784	8,784	4,392	4,392
R-squared	0.650	0.519	0.247	0.024
Hour-of-Sample FE	YES	YES	NO	NO
Hour-of-Day FE	NO	NO	YES	YES
Day-of-Week FE	NO	NO	YES	YES

Notes. The variable "Training_d" takes value of one if the observation falls within training period and within the range of midnight to 6 a.m. or 6 p.m. 'till midnight. Columns (1) and (2) are estimated using the control group. Columns (3) and (4) are estimated using the treatment group only. For regressions using the treatment group, the reported estimates are the coefficient of variable "Training_d" only, not interaction of "Training_d" and "Treated_i." In the "Actual" columns, the dependent variables is the pollution concentration at the hourly level (CO or $PM_{2.5}$ or SO_2). All regressions under these columns include the current and 24-h lag of weather variables (i.e., temperature, humidity, pressure, precipitation, wind direction, and wind speed). In the "Pred Error" columns, the dependent variables is the residual obtained from the first stage based on the difference between the actual data and the prediction made by the RF model. In column (2), the RF model is built from the data of eight neighboring area surrounding the control area over the period of 2014 - 2016. In column (4), the RF model is built from the data of the training site over the period of 2014 - 2016. Standard errors, in parentheses, are bootstrapped 500 times. ***, **, and * Significance level at 1%, 5% and 10%.

Online Appendix
(Not for Publication)

This page is intentionally left blank.

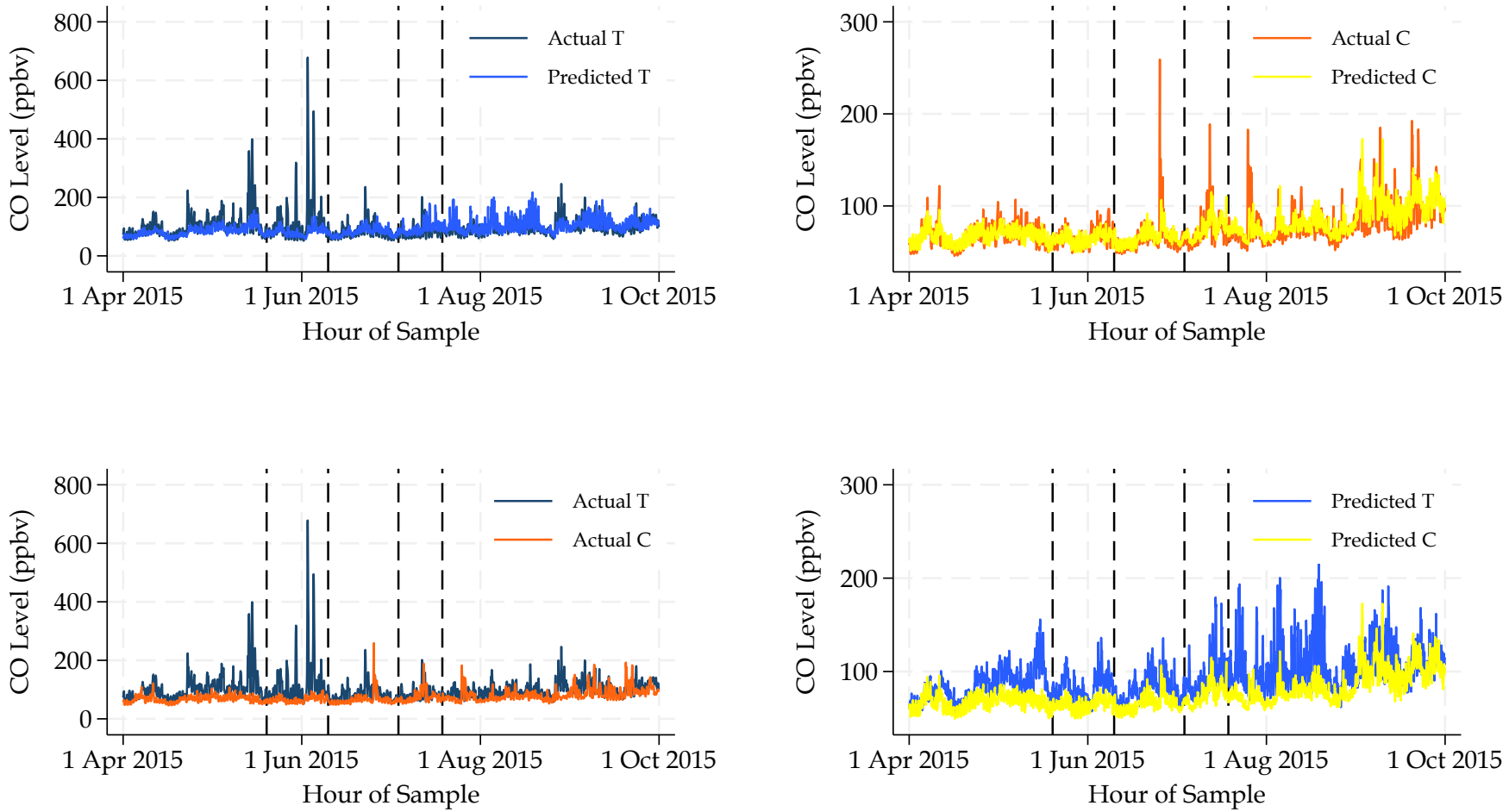


Figure A1: Actual and Predicted Levels of CO in 2015

Notes. These plots show the actual and RF-predicted concentration levels of CO over the period of 1 April to 30 September 2015. The first two dashed lines represent the Predator Walk Exercise (20 May to 10 June). The last two dashed lines denote the Talisman Sabre 2015 (4 July to 19 July). The ticks on the x-axis indicate 12 AM of the respective date.

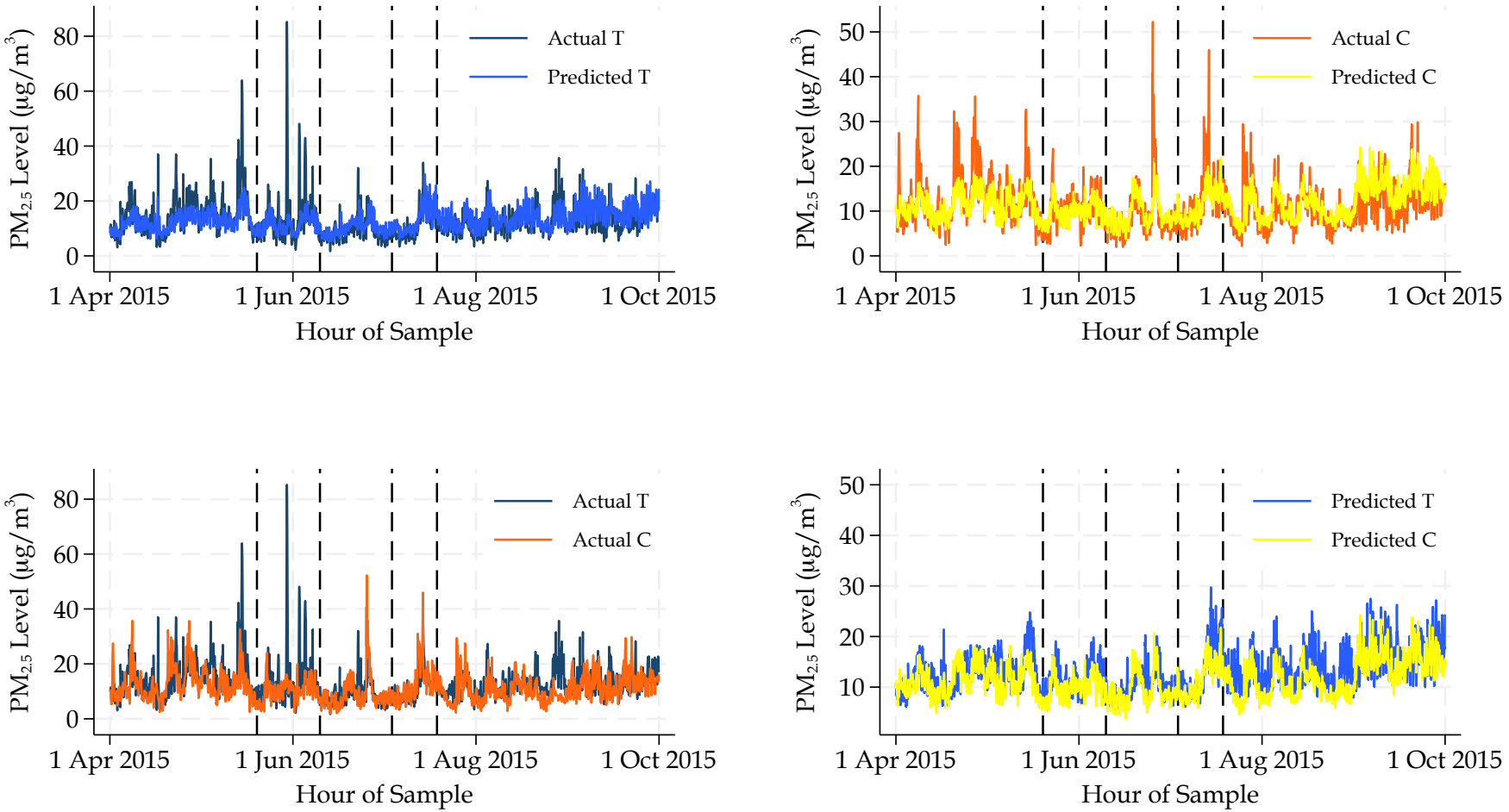


Figure A2: Actual and Predicted Levels of $PM_{2.5}$ in 2015

Notes. These plots show the actual and RF-predicted concentration levels of $PM_{2.5}$ over the period of 1 April to 30 September 2015. The first two dashed lines represent the Predator Walk Exercise (20 May to 10 June). The last two dashed lines denote the Talisman Sabre 2015 (4 July to 19 July). The ticks on the x-axis indicate 12 AM of the respective date.

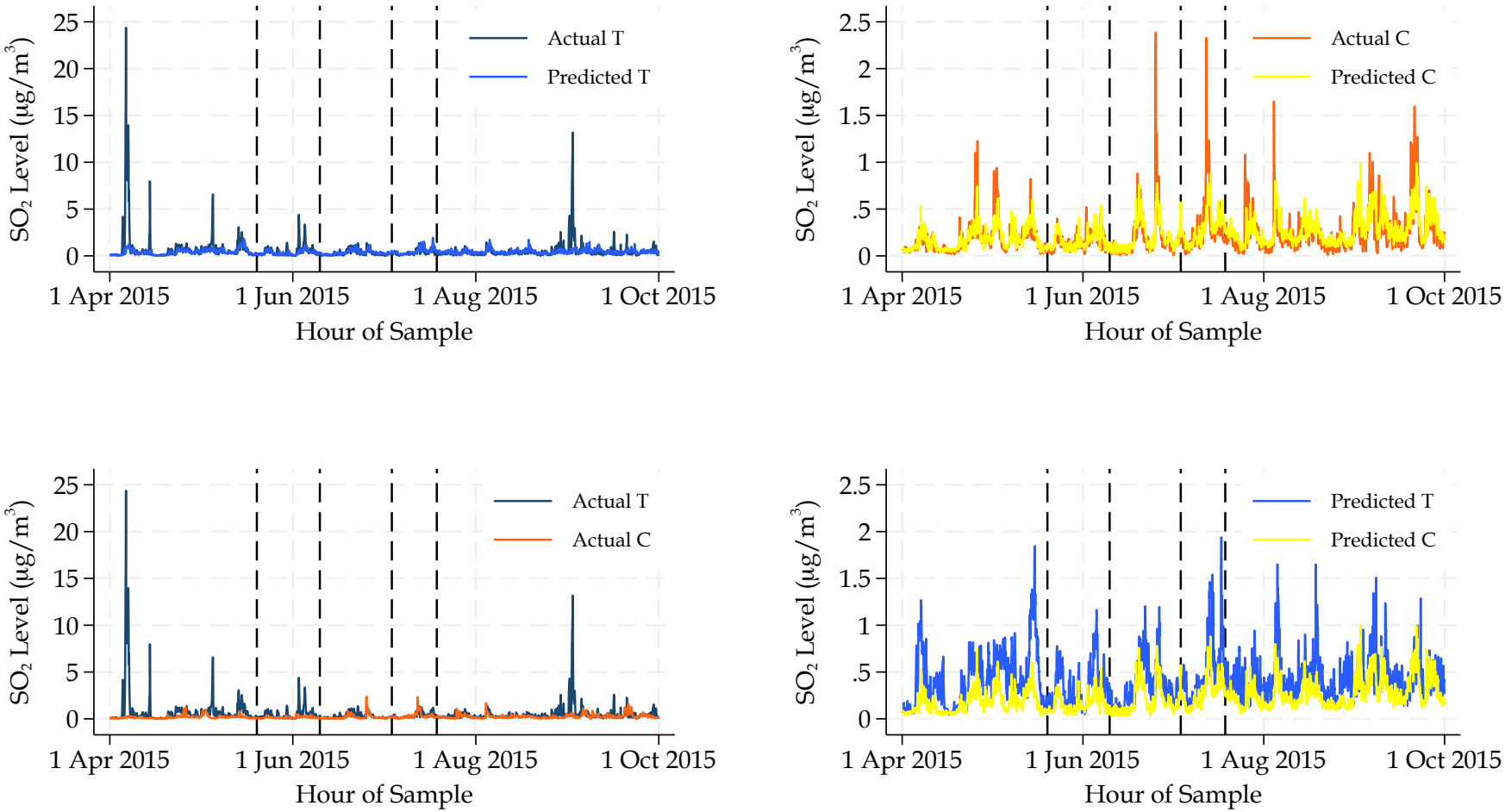


Figure A3: Actual and Predicted Levels of SO_2 in 2015

Notes. These plots show the actual and RF-predicted concentration levels of SO_2 over the period of 1 April to 30 September 2015. The first two dashed lines represent the Predator Walk Exercise (20 May to 10 June). The last two dashed lines denote the Talisman Sabre 2015 (4 July to 19 July). The ticks on the x-axis indicate 12 AM of the respective date.

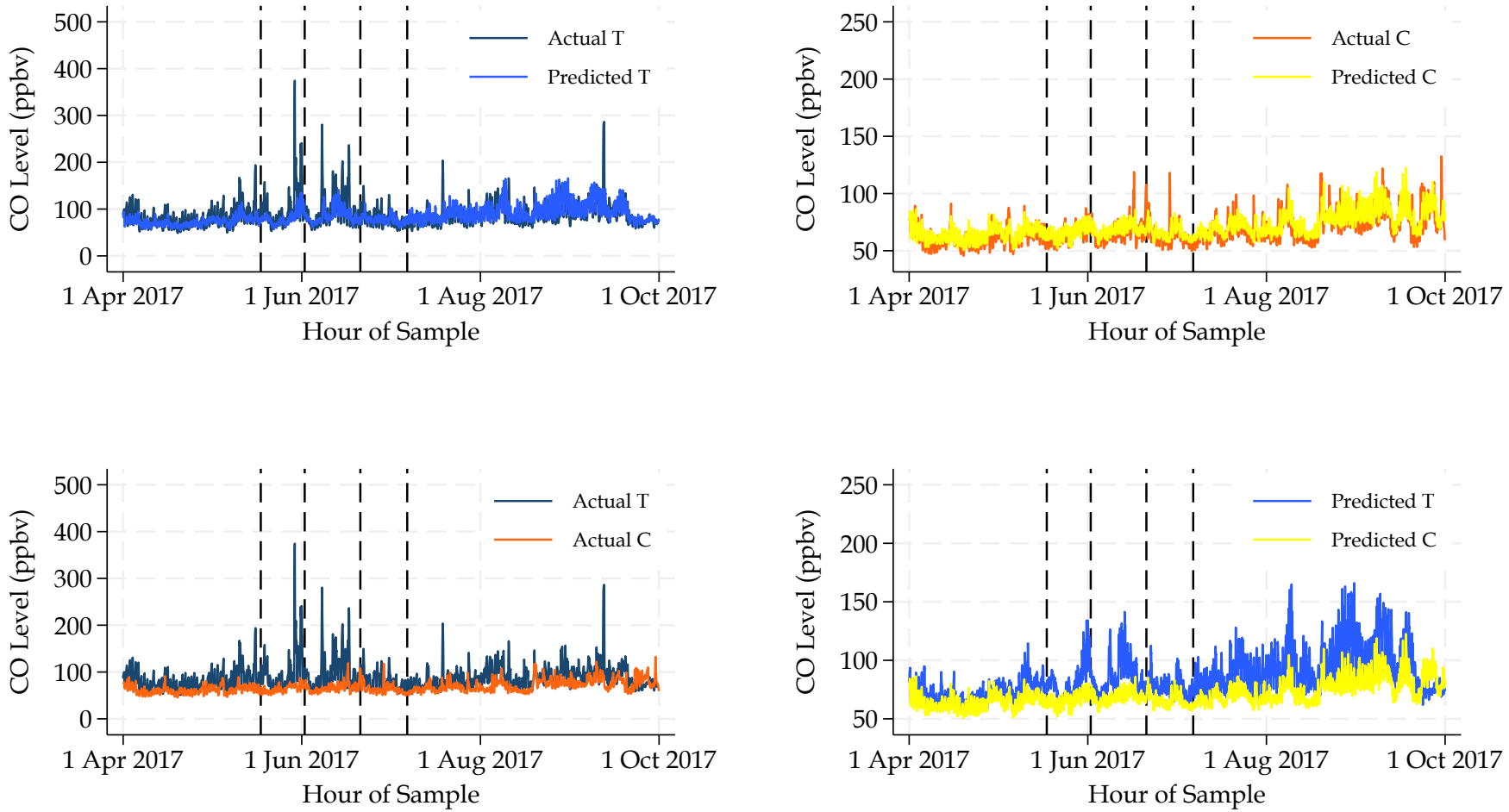


Figure A4: Actual and Predicted Levels of CO in 2017

Notes. These plots show the actual and RF-predicted concentration levels of CO over the period of 1 April to 30 September 2017. The first two dashed lines represent the Southern Jackaroo Exercise (18 May to 2 June). The last two dashed lines denote the Talisman Sabre 2017 (21 June - 7 July). The ticks on the x-axis indicate 12 AM of the respective date.

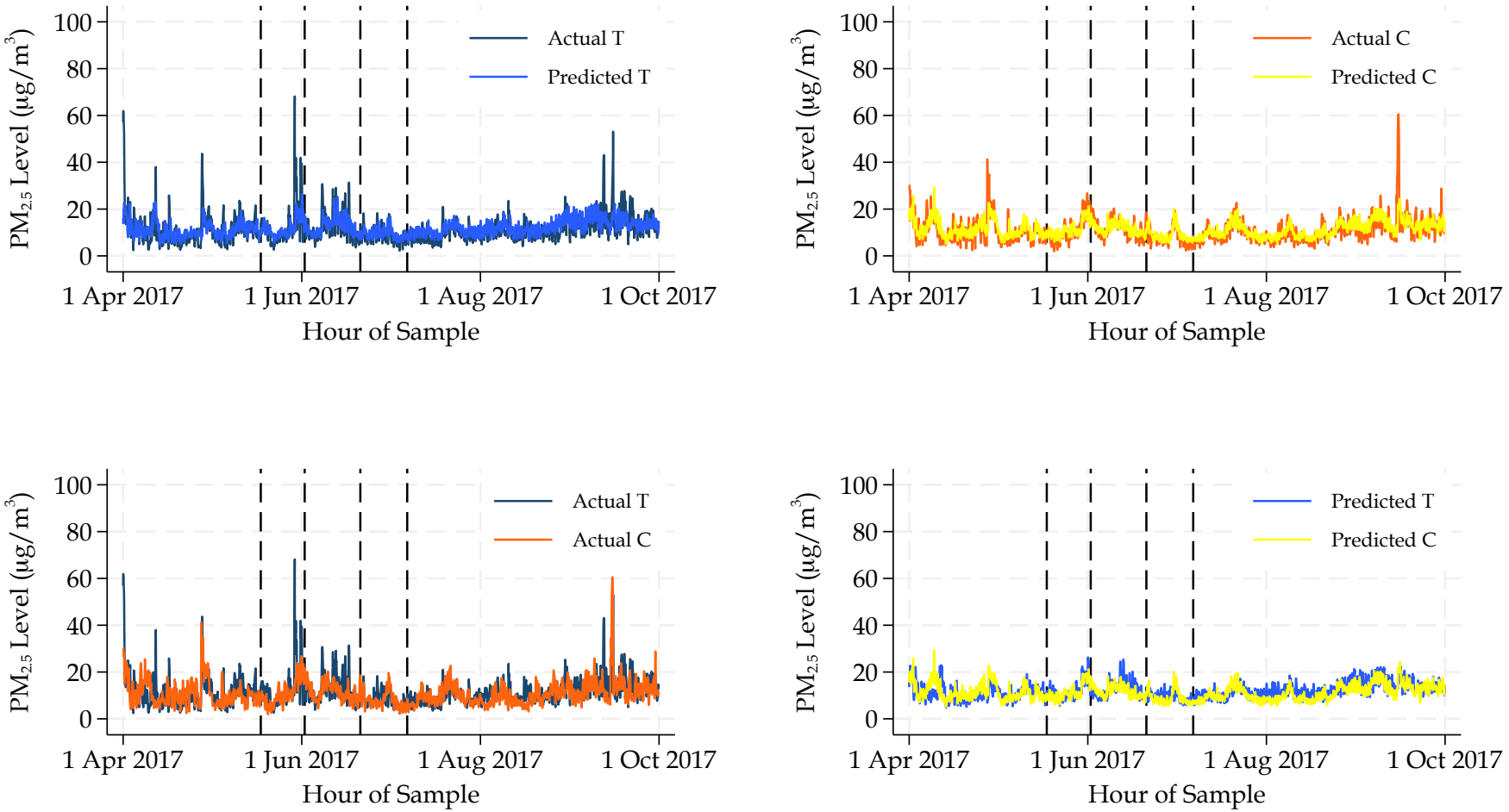
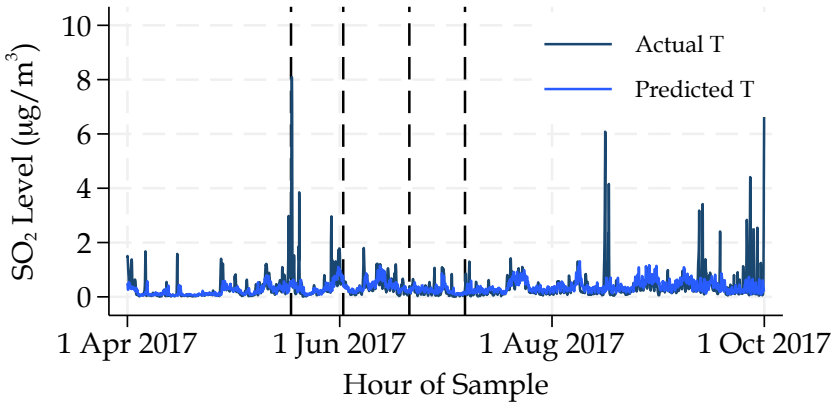
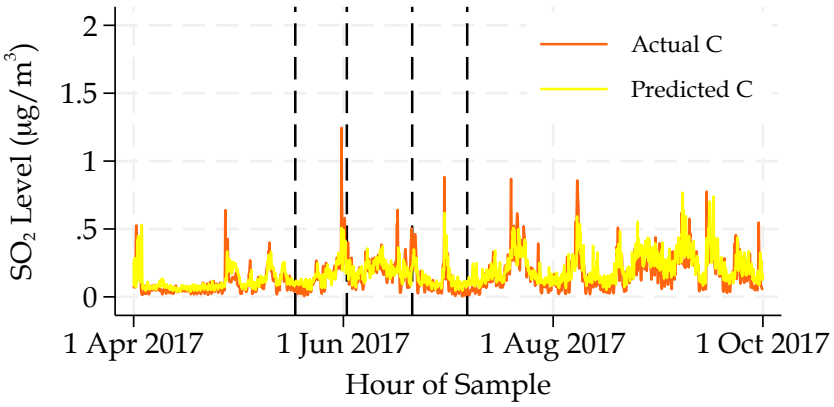
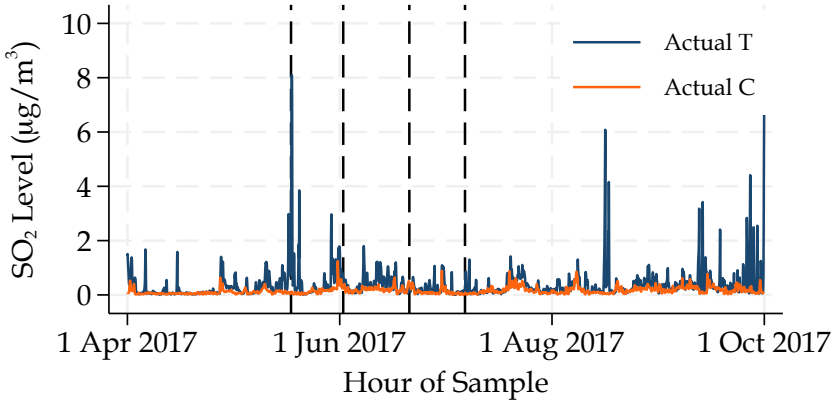


Figure A5: Actual and Predicted Levels of $PM_{2.5}$ in 2017

Notes. These plots show the actual and RF-predicted concentration levels of $PM_{2.5}$ over the period of 1 April to 30 September 2017. The first two dashed lines represent the Southern Jackaroo Exercise (18 May to 2 June). The last two dashed lines denote the Talisman Sabre 2017 (21 June - 7 July). The ticks on the x-axis indicate 12 AM of the respective date.







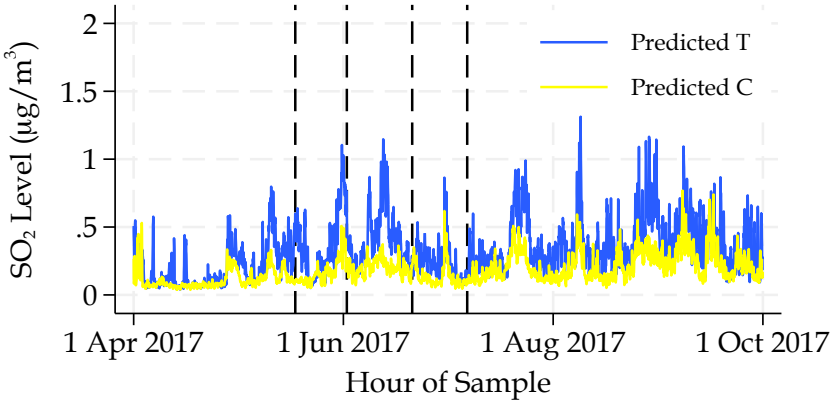


Figure A6: Actual and Predicted Levels of SO_2 in 2017

Notes. These plots show the actual and RF-predicted concentration levels of SO_2 over the period of 1 April to 30 September 2017. The first two dashed lines represent the Southern Jackaroo Exercise (18 May to 2 June). The last two dashed lines denote the Talisman Sabre 2017 (21 June - 7 July). The ticks on the x-axis indicate 12 AM of the respective date.

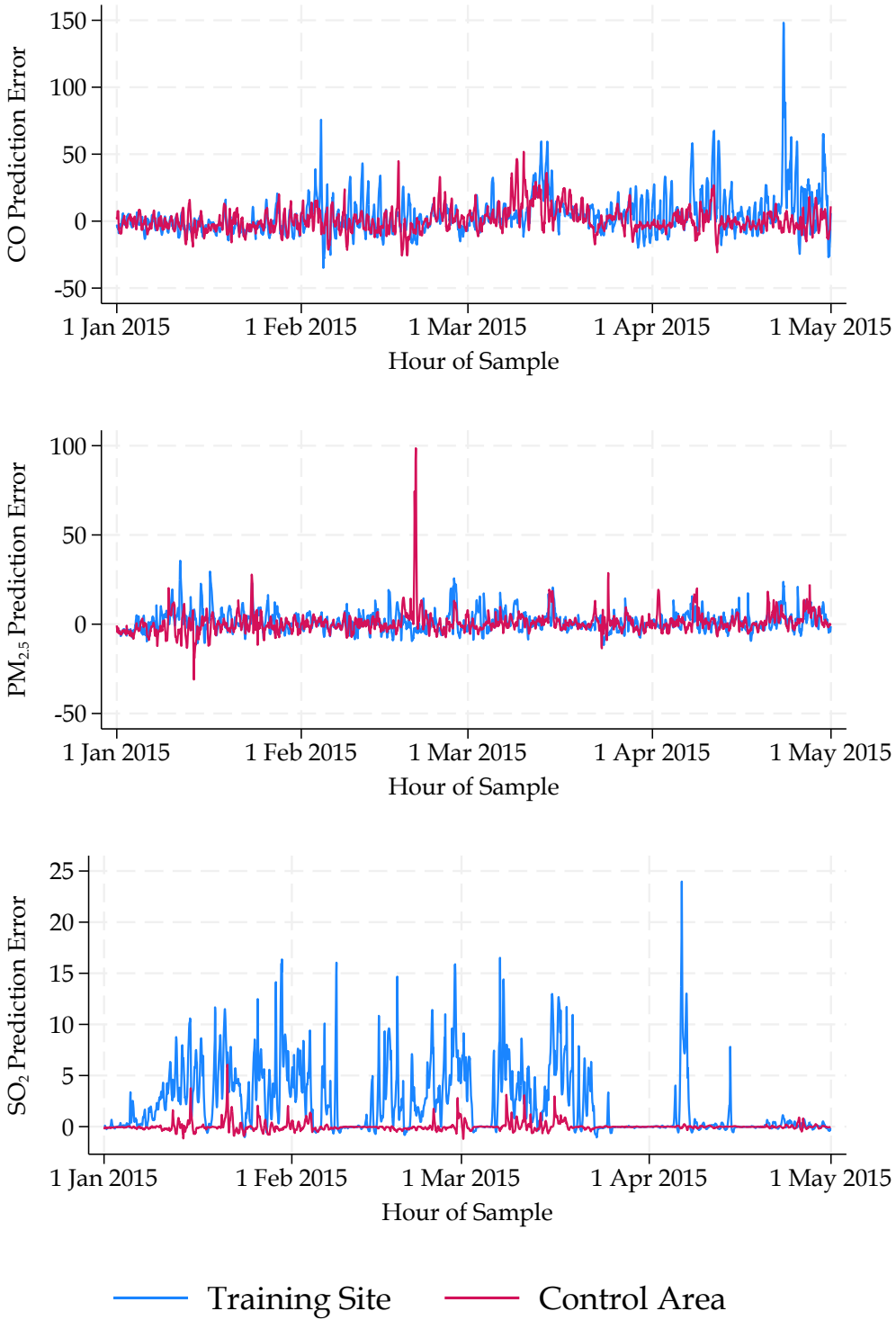


Figure A7: Prediction Error During Jan - Apr 2015

Notes. This plot shows the prediction error (difference between actual value and RF-predicted value) over the period 1 Jan - 30 Apr 2015. The ticks on the x-axis indicate 12 AM of the respective date.

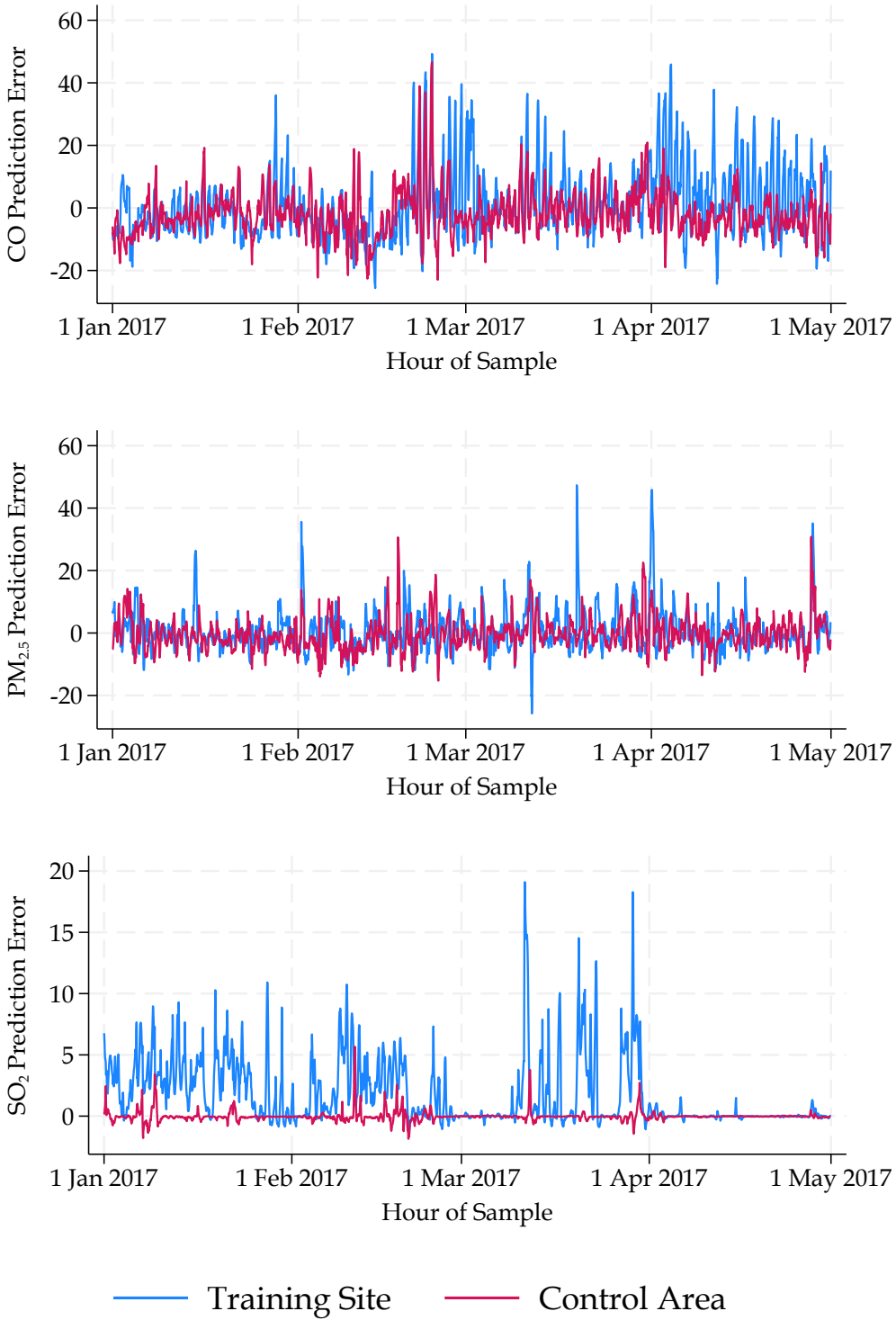


Figure A8: Prediction Error During Jan - Apr 2017

Notes. This plot shows the prediction error (difference between actual value and RF-predicted value) over the period 1 Jan - 30 Apr 2017. The ticks on the x-axis indicate 12 AM of the respective date.