

LETTER • OPEN ACCESS

## Cleaner air has contributed one-fifth of US maize and soybean yield gains since 1999

To cite this article: David B Lobell and Jennifer A Burney 2021 *Environ. Res. Lett.* **16** 074049

View the [article online](#) for updates and enhancements.

You may also like

- [Corrigendum: Satellite detection of cover crops and their effects on crop yield in the Midwestern United States \(2018 \*Environ. Res. Lett.\* 13 064033\)](#)  
Christopher A Selfert, George Azzari and David B Lobell
- [Combined heat and drought suppress rainfed maize and soybean yields and modify irrigation benefits in the USA](#)  
Xiangyu Luan, Riccardo Bommarco, Anna Scaini et al.
- [Drought effects on US maize and soybean production: spatiotemporal patterns and historical changes](#)  
Samuel C Zipper, Jiangxiao Qiu and Christopher J Kucharik

ENVIRONMENTAL RESEARCH  
LETTERS

## LETTER

## OPEN ACCESS

RECEIVED  
27 April 2021REVISED  
14 June 2021ACCEPTED FOR PUBLICATION  
29 June 2021PUBLISHED  
14 July 2021

Original content from  
this work may be used  
under the terms of the  
[Creative Commons  
Attribution 4.0 licence](#).

Any further distribution  
of this work must  
maintain attribution to  
the author(s) and the title  
of the work, journal  
citation and DOI.

Cleaner air has contributed one-fifth of US maize and soybean  
yield gains since 1999David B Lobell<sup>1,\*</sup> and Jennifer A Burney<sup>2</sup> <sup>1</sup> Department of Earth System Science and the Center on Food Security and the Environment, Stanford University, Stanford, CA, United States of America<sup>2</sup> School of Global Policy and Strategy, University of California, San Diego, La Jolla, CA, United States of America

\* Author to whom any correspondence should be addressed.

E-mail: [dlobell@stanford.edu](mailto:dlobell@stanford.edu)**Keywords:** agriculture, air pollution, remote sensing, yield trendsSupplementary material for this article is available [online](#)

## Abstract

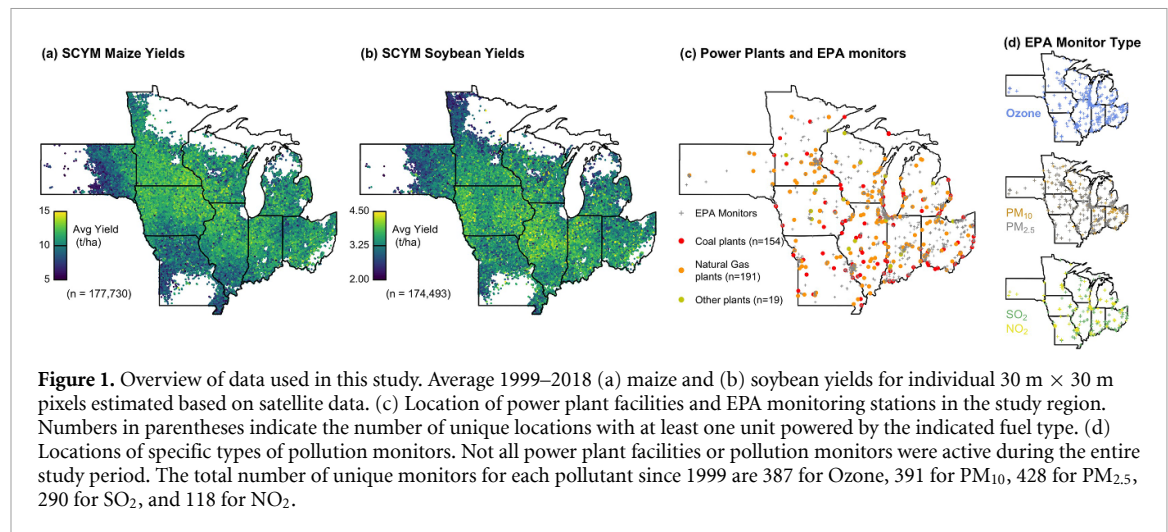
Crop productivity is potentially affected by several air pollutants, although these are usually studied in isolation. A significant challenge to understanding the effects of multiple pollutants in many regions is the dearth of air quality data near agricultural fields. Here we empirically estimate the effect of four key pollutants (ozone (O<sub>3</sub>), particulate matter (PM), sulfur dioxide (SO<sub>2</sub>), and nitrogen dioxide (NO<sub>2</sub>)) on maize and soybean yields in the United States using a combination of administrative data and satellite-derived yield estimates. We identify clear negative effects of exposure to O<sub>3</sub>, PM, and SO<sub>2</sub> in both crops, using yields measured in the vicinity of monitoring stations. We also show that while stations measuring NO<sub>2</sub> are too sparse to reliably estimate a yield effect, the strong gradient of NO<sub>2</sub> concentrations near power plants allows us to more precisely estimate NO<sub>2</sub> effects using satellite measured yield gradients. The presence of some powerplants that turn on and others that shut down during the study period are particularly useful for attributing yield gradients to pollution. We estimate that total yield losses from these pollutants averaged roughly 5% for both maize and soybean over the past two decades. While all four pollutants have statistically significant effects, PM and NO<sub>2</sub> appear more damaging to crops at current levels than O<sub>3</sub> and SO<sub>2</sub>. Finally, we find that the significant improvement in air quality since 1999 has halved the impact of poor air quality on major crops and contributed to yield increases that represent roughly 20% of overall yield gains over that period.

## 1. Introduction

A growing literature has demonstrated negative effects of poor air quality on agricultural outcomes, using a variety of experimental [1, 2], observational [3, 4], and modeling approaches [5, 6]. These effects are deemed largest in countries such as India and China that have persistent poor air quality [4, 7, 8], although yield losses have also been reported in other regions including Europe and the United States [3, 9].

Significant uncertainties remain regarding both the effect of individual pollutants as well as the overall effect of pollution on agriculture. These uncertainties limit our ability to identify the potential benefits for agriculture from air quality improvements associated with technology or policy changes [10],

as well as which pollutants deserve the most scrutiny. Numerous studies have focused on effects of O<sub>3</sub>, reflecting the well-documented effects in greenhouse experiments as well as the availability of data on O<sub>3</sub> levels from air monitoring stations. While other pollutants such as NO<sub>2</sub> and SO<sub>2</sub> are also known to be phytotoxic [11, 12], they have been subject to less empirical study [13], presumably because of a relative paucity of data. Study of these two compounds is also complicated by the fact that they react chemically in the atmosphere and contribute to formation of both O<sub>3</sub> and aerosol particulate matter (PM). PM, which is often monitored because of well-known human health effects, may also impact agricultural productivity through physical and socioeconomic pathways, including dimming and scattering



of solar radiation [7] and reduction in farm worker productivity [14].

Some studies have attempted to estimate the net impact of multiple pollutants, either by comparing yields directly to emission sources [8, 15] or by combining estimates from different studies [16]. These analyses indicate a significant role for multiple pollutants, for instance with Fischer [16] estimating 21% yield loss from O<sub>3</sub> and 9% from dimming related to PM in India wheat.

Here we consider effects of multiple pollutants on maize and soybean in the central United States, a region that is a major global producer of these crops and where air pollution has declined in recent decades subsequent to the Clean Air Amendments in 1990 (figure S1 (available online at [stacks.iop.org/ERL/16/074049/mmedia](https://stacks.iop.org/ERL/16/074049/mmedia))). The main objectives of the current study are to empirically estimate the sensitivity of maize and soybean yields to four key pollutants (O<sub>3</sub>, PM, SO<sub>2</sub>, and NO<sub>2</sub>), and to assess the role that pollution reductions have played in yield progress over the past two decades. To do this, we exploit a few useful features of the study region. First, comprehensive public datasets exist for air quality monitors, power plant emissions, and county-level crop yields (figure 1). Second, despite overlap in emissions sources (e.g. vehicles, power plants) we observe relatively low correlation between pollutants at specific locations (table S1), which allows us to estimate the effects of each. Third, recent progress in satellite analysis has resulted in fine-resolution (30 m) yield maps for both maize and soybean for a nine-state region comprising the largest rainfed maize and soybean producing states since 1999 [17, 18] (figure 1). These data enable us to tailor our data sample to cropped areas near monitors (versus, for example, interpolating pollution data) and enable detailed inspection of yields across pollution gradients. In particular, we find that yield gradients near power plants—stationary sources where emissions of SO<sub>2</sub> and NO<sub>2</sub> are continuously monitored and vary dramatically by

feedstock type and unit technology—provide a very useful constraint on the overall sensitivity of yields to pollution.

## 2. Data and methods

### 2.1. Study region and period

The United States is a major producer of both maize and soybean, constituting roughly one-third of global production for each crop [19]. In the current study, we focus on a nine-state region (Illinois, Indiana, Iowa, Michigan, Minnesota, Missouri, Ohio, South Dakota and Wisconsin) that produces roughly two-thirds of national output and for which detailed satellite-based yield estimates exist (see section 2.5). We consider the 1999–2019 period, again based on the availability of satellite-based yield records for both crops.

### 2.2. Ground-based air pollution data

We obtained hourly O<sub>3</sub> and daily PM, NO<sub>2</sub>, and SO<sub>2</sub> measurements for 1999–2019 from [https://aqs.epa.gov/aqsweb/airdata/download\\_files.html](https://aqs.epa.gov/aqsweb/airdata/download_files.html). For O<sub>3</sub>, we computed the accumulated ozone exposure over a threshold (AOT) during the main growing season (June–August for both maize and soybean) for different thresholds (thresh) from 40 to 80 ppb in 10 ppb increments:

$$\text{AOT}_{\text{thresh}} = \sum_{h=0}^H C_h \text{ where } C_h = \begin{cases} (O_h - \text{thresh}) & \text{if } O_h > \text{thresh} \\ 0 & \text{if } O_h \leq \text{thresh} \end{cases} \quad (1)$$

where  $O_h$  is hourly ozone values,  $h$  goes from 0 to  $H$  (the total number of hours with measurements in June–August) in increments of one hour. Although a threshold of 40 ppb is commonly used in the literature (i.e. AOT<sub>40</sub>) [5, 6, 20], some authors have argued for a higher threshold [21], and therefore we examined

a range of possible values. Most stations have measurements for all hours, although exceedances generally occur during daytime. A small number of station-years ( $\sim 2\%$ ) had more than 20% of days missing and were excluded from analysis. The remaining site-years were missing an average of less than one day of measurement during the 92 d June–August period.

For all pollutants other than  $O_3$  ( $PM_{10}$ ,  $PM_{2.5}$ ,  $NO_2$ , and  $SO_2$ ) we calculated the average of daily maximum observations for June–August as reported in the daily EPA files.  $PM_{2.5}$  refers to all PM below 2.5 mm in size, whereas  $PM_{10}$  includes particulates below 10 mm. Slight changes in the growing season definition, for instance adding April–May, did not qualitatively affect results.

### 2.3. Power plant data

We used emissions data from US electric power generation facilities provided in the EPA's Air Markets Program Database (<https://ampd.epa.gov/>). The AMPD includes most generation facilities in the country, and provides facility location, technology, operation, and emissions information. We aggregated emissions and fuel information to the facility level (one facility might contain many generation units), and coded facilities according to their primary fuel type with 'coal' including any coal or coke use, 'natural gas' including all variants of natural gas, and 'other' encompassing most residual oil, biomass, and diesel fuel facilities.

### 2.4. Satellite-based pollution data

For evaluating  $NO_2$  gradients near power plants and identifying facilities with high gradients, we utilized tropospheric column  $NO_2$  data from the Sentinel 5 TROPOMI sensor. TROPOMI  $NO_2$  estimates are available at  $0.01^\circ \times 0.01^\circ$  ( $\sim 1$  km) resolution starting in late June 2018. Here we use the July–August 2018 average of daily tropospheric column vertical density from the L3 OFFL product, obtained through Google Earth Engine [22].

### 2.5. Crop yield data

Annual maize and soybean yields at the county level were obtained from the USDA National Agricultural Statistics Service (<https://quickstats.nass.usda.gov/>). NASS yield estimates are based on phone surveys of farmers throughout the region, and generally cover all counties producing maize and soybean. The number of counties with data varies slightly by year because of varying survey response rates, with a notable decline in recent years [23].

Yields at 30 m spatial resolution were obtained for each year in the study period from the scalable crop yield mapper (SCYM) [17] applied to Landsat satellite imagery. Briefly, SCYM uses crop simulations to develop regressions relating final maize (or soybean) yield to two measures of canopy greenness (greenness at the peak of season and 30 d after the peak)

and four measures of weather relating to temperature, rainfall, and radiation. Multiple Landsat observations are then used to estimate the greenness measures for each pixel classified as maize or soybean in the USDA's Cropland Data Layer [24]. SCYM estimates have been validated for both crops against county and field level data in the region [17, 18]. These studies document strong agreement between SCYM and ground-measured yields, as well as the ability of SCYM to accurately estimate the yield response to management or soil factors at the field level [25]. For the current study, SCYM yields were averaged for a 20 km radius around each EPA station in each year. In addition, a sample of pixel-level estimates were taken for 20 000 random pixels in each state to examine yield variation with distance from power plant locations.

### 2.6. Weather data

Daily values of minimum temperature, maximum temperature, and precipitation were obtained from the daily PRISM 2.5 arc minutes ( $\sim 5$  km) resolution weather dataset [26] for each location with yield estimates. Growing season weather conditions were then summarized by calculating four common measures known to explain variation in maize and soybean yields [27]: growing degree days between  $8^\circ C$  and  $30^\circ C$  during the growing season (GDD), extreme degree days above  $30^\circ C$  (EDD), total precipitation between April 1 and August 31 (Pr), and the square of precipitation (Prsq).

In addition, for testing whether EPA monitoring stations (and, by inference, agricultural fields) upwind and downwind of power plants experienced significantly different pollution levels, we obtained daily wind speeds for each power plant facility from the University of Idaho Gridded Surface Meteorological Dataset [28]. Average wind direction over the June–August period was calculated for each facility in each year, and the angle between wind direction and EPA monitor direction was calculated for every monitor that was closest to that facility. Monitors were considered downwind if the cosine of the angle was greater than zero, and upwind if the cosine was less than zero, although results were not sensitive to using stricter definitions of downwind (e.g. cosine greater than 0.5).

### 2.7. Yield regressions

To relate yields to pollution, we estimated a panel model with site and year fixed effects:

$$\text{yield}_{i,t} = \beta_w W_{i,t} + \beta_p P_{i,t} + \gamma_i + \tau_t + \varepsilon_{i,t} \quad (2)$$

where  $\text{yield}_{i,t}$  is yield in site  $i$  in year  $t$ ,  $W$  is a vector of controls for weather variation,  $P$  is the measure of pollution exposure,  $\gamma_i$  is a site intercept or fixed effect, and  $\tau_t$  is a year intercept. We alternatively define yield as the average of all SCYM estimates within a 20 km radius around the EPA station

location, or the reported NASS yield for the county in which the station is located. When using NASS yields, stations located in the same county were assigned identical yields. Weather controls were the aforementioned four weather variables (GDD, EDD, Pr, and Prsq) averaged over the same 20 km radius. Equation (2) was estimated using a fixed effect ordinary least squares regression in the *fixest* R package [29], with standard errors clustered at the site level. The regression analysis was conducted separately for maize and soybean. We did not attempt to estimate interaction terms between weather and pollution given the limited number of monitoring sites (figure 1); hence the effects of pollutants were assumed to be constant across different weather conditions.

The fixed effects for site and year indicate that each location and year is given its own unique intercept, which is a common practice to control for unobserved factors (e.g. fertilizer use, soil quality) that might both affect the outcome of interest (i.e. yield) and be correlated with spatial or temporal variations in the causal factor of interest (i.e. pollution), leading to bias in the estimated coefficients. By using fixed effects, identification of  $\tau_W$  and  $\tau_P$  comes only from deviations from the expected value for a given site and year, reducing the chance for bias from omitted variables. In addition, the use of fixed effects helps reduce the influence of collinearity between pollutants, since pollutants are less correlated after accounting for fixed effects (table S1). Figure S3 presents a simple simulation to illustrate the value of fixed effects when two variables are highly correlated across locations. Similarly, removing fixed effects helps to reduce the collinearity between pollutants and weather (table S1).

## 2.8. Yield gradients near power plants

In addition to the panel analysis that uses yields in the vicinity of monitoring stations, we also examine yield gradients near power plants. As described in the Results, these yield gradients are particularly useful for constraining the effect of NO<sub>2</sub>, given the relatively weak constraints provided by the panel analysis (owing to a low number of NO<sub>2</sub> monitors), and the strong gradients of NO<sub>2</sub> near power plants.

We define  $P_i(d)$  as the expected value of a pollutant  $i$  at distance  $d$  from a power plant, estimated from a LOWESS (locally weighted scatterplot smoothing) regression of pollutant values vs distance (figure 3). Similarly, we define  $Y(d)$  as the expected value of the yield residual at distance  $d$ , where the residual is from a regression of yields against weather and soil. For the yield regression we use the same weather variables as in equation (2), while for soil we use the National Commodity Crop Productivity Index (NCCPI) produced by the USDA [30].

For a given distance, we then define

$$\delta P_i(d) = P_i(d) - P_i(1 \text{ km}) \quad (3)$$

$$\delta Y(d) = Y(d) - Y(1 \text{ km}). \quad (4)$$

Which we refer to as the gradients in pollutant and yield levels, respectively, compared to values at 1 km distance from the power plants. We do not divide these differences by the distance, so the units of pollution gradients are in the original units (e.g. ppb) and the units of yield gradients are  $\text{t ha}^{-1}$ .

We then utilize the estimated coefficients from the panel analysis (equation (2)) to estimate the expected net yield impact of gradients in O<sub>3</sub>, PM, and SO<sub>2</sub> near the power plants:

$$\delta \hat{Y}(d) = \beta_{O_3} \times \delta O_3(d) + \beta_{PM} \times \delta PM(d) + \beta_{SO_2} \times \delta SO_2(d). \quad (5)$$

The effect of NO<sub>2</sub> on yields is then estimated as:

$$\beta_{NO_2} = \frac{(\delta Y(d) - \delta \hat{Y}(d))}{\delta NO_2(d)}. \quad (6)$$

Thus, the amount of yield gradient not explained by the other three pollutants, or by local weather and soil gradients, is attributed to the gradient in NO<sub>2</sub>. To assess the uncertainty of  $\tau_{NO_2}$  estimated in equation (6), we re-estimate the equation 100 times after resampling the  $\tau_s$  for O<sub>3</sub>, PM, and SO<sub>2</sub> based on standard errors from equation (2), and resampling estimates of  $P_i(d)$  for each pollutant based on resampling of the sites used in the LOWESS fit (figure 3). We calculate equation (6) for distances of 10 km–50 km in 10 km increments, finding that estimates are most stable for 30 km–50 km (see section 3). We use  $d = 30$  km for estimates reported in the main text, as the pollution gradients are less reliable beyond 30 km because of sparse station coverage at these distances (figure S2).

## 2.9. Estimation of total pollution effects and contribution to yield trends

For each year in the dataset, we multiply the inferred yield sensitivity to pollutants from equations (2) and (6) by the estimated average level of each pollutant at crop locations. The latter is estimated as:

$$P_{i,c,t} = P_{i,s,t} \times S_i \quad (7)$$

where  $P_{i,c,t}$  is the average level of pollutant  $i$  over croplands at time  $t$ ,  $P_{i,s,t}$  is the average level at EPA station locations in year  $t$ , and  $S_i$  is a scaling factor that relates the average value at croplands to the average value at stations. We utilize scaling factors, rather than the simple average of observed concentrations at monitoring stations, since monitor locations are not representative of agricultural regions (figure S2). These scaling factors were derived in two ways. The first was to calculate a weighted average of the curve describing the dependence of  $P_i$  on distance from the nearest power plants (e.g. figure 3(b)) using weights that correspond to the distribution of crop distances to power plants (figure S2(a)). These weighted averages were calculated for each year and then divided



by the simple average of station values for that year. Table S3 (in SI appendix) reports the average value of this ratio for each pollutant. In a second approach, we averaged satellite-based measures of pollutants over crop and EPA station locations, and calculate the ratio of these two averages. These satellite-based scaling factors were only possible for pollutants with accurate satellite measures of near surface concentrations, namely PM and NO<sub>2</sub>, and served mainly to corroborate the estimates derived from the station data (table S3). Main results reported in the paper correspond to the station-based scaling factors.

The total effect of all pollutants in each year was calculated by summing across the individual estimates for the four pollutants. This assumes that the effects of each are additive, and for instance the effect of PM does not depend on the level of O<sub>3</sub>. Other studies have made similar assumptions [12, 16, 31]. It is possible that effects could be multiplicative, for instance with damage from NO<sub>2</sub> exacerbated at high levels of SO<sub>2</sub> [32], in which case our total estimates of damage would be conservative.

### 3. Results

#### 3.1. Crop yields are significantly inhibited by all four pollutants

We first statistically relate crop yields measured near air quality monitors to the levels of pollutants measured at those monitors during the growing season using panel regression models. These regressions provide average estimates, across the study region and period, of the associations between variations in pollutant concentrations and crop yields while controlling for weather, common time variation (through inclusion of linear time trends or more flexible year fixed effects), and time-invariant site-specific drivers of yields such as soil differences (through inclusion of location fixed-effects; see Methods). Because separate monitoring networks exist for each pollutant, their locations do not typically overlap (table S2). For example, less than one-fifth of the O<sub>3</sub> observations and less than one-quarter of PM<sub>10</sub> observations over the study period were coincident with SO<sub>2</sub> observations. We therefore conduct regressions separately by pollutant, with sensitivity tests used to gauge the potential effect of omitted pollution measures. We estimate these relationships using both the high-resolution satellite-based yields (SCYM) and coarser county-level crop yield data from the USDA (NASS) [33].

We find broadly negative effects of O<sub>3</sub>, PM, and SO<sub>2</sub> on both maize and soybean yields (figure 2). For maize, effects are significantly negative ( $p < 0.05$ ) whether using satellite-based estimates of yield averaged within 20 km of the EPA station or using NASS reported yields for the entire county containing the station. For soybean, effects are significantly negative except for the estimate for PM<sub>10</sub> when using

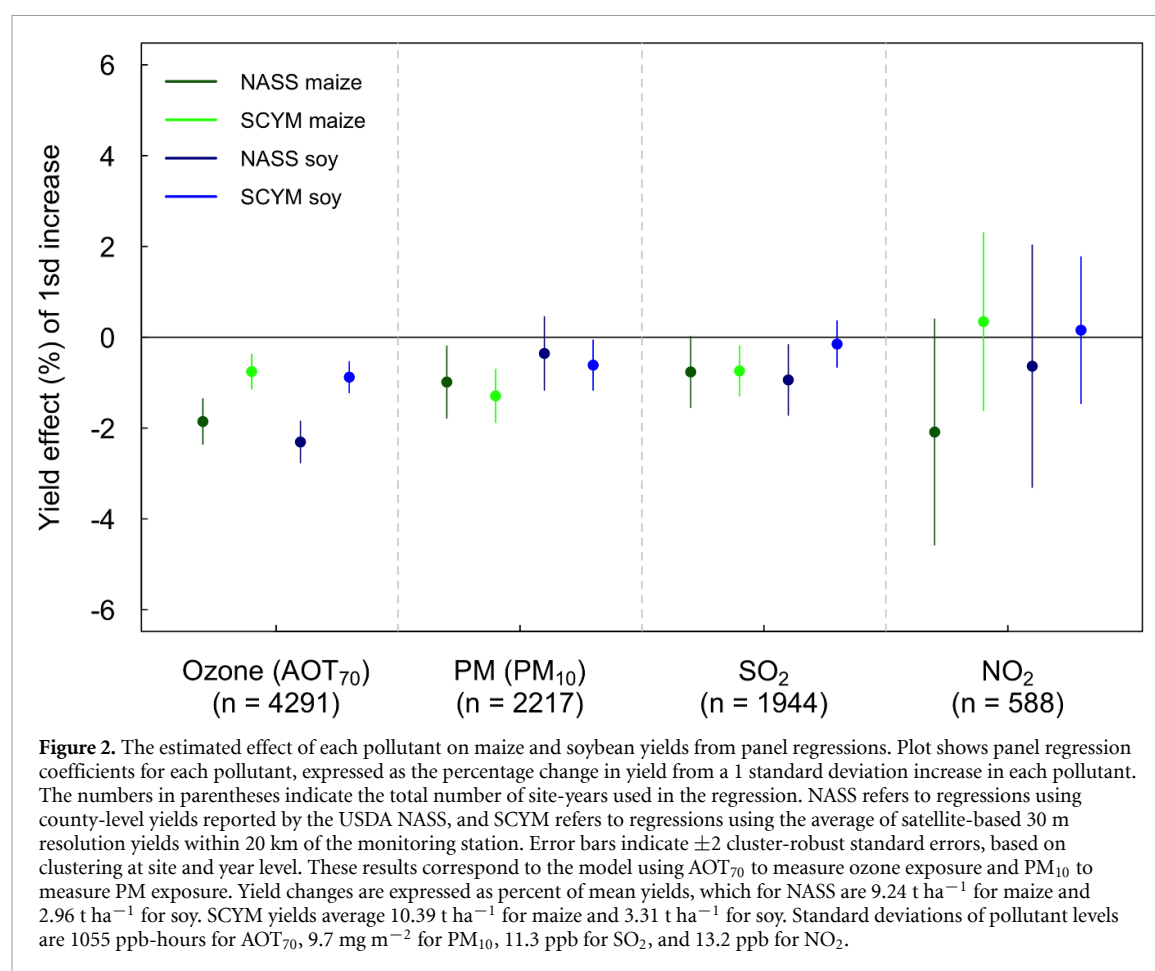
NASS yields, and the estimate for SO<sub>2</sub> using satellite yields.

To assess whether collinearity with other pollutants would significantly alter results, we repeat the regressions above for the subset of site-years with multiple pollution measures (figure S4). As expected, the much lower sample sizes lead to substantially larger error bars. However, we find that inclusion of other pollutants has relatively small effects on the estimated coefficients. This is consistent with the low correlations after removing site and year fixed effects (table S1), which indicates that while levels of pollutants may be correlated when examined in cross-section (e.g. cities versus rural areas), the fluctuations of individual pollutants within a given location are largely uncorrelated. Similarly, a recent study of O<sub>3</sub> effects in China found that O<sub>3</sub> coefficients were only slightly altered when accounting for PM [34].

The magnitudes of uncertainties in figure 2 reflect differences in the density of monitoring networks for each pollutant. Ozone has been monitored at 387 unique stations in the region, with a median coverage of 15 growing seasons of data for 1999–2018, resulting in over 4000 site-year observations. Roughly half this number was available for PM and SO<sub>2</sub>, resulting in slightly broader confidence intervals for the estimates. In contrast, fewer than 600 site-year observations were available for NO<sub>2</sub>, and as a result the panel estimates for effects of this pollutant were much less precise, and not significantly different than zero.

To further constrain the yield effects of NO<sub>2</sub>, we turn to investigation of yields in the vicinity of power plant facilities. We observe a clear increase in average satellite-based yields as distance increases from the nearest power plant facility (figure 3). The yield increases are most rapid within the first 30 km but continue to exhibit gradual increases for both maize and soybean up to at least 50 km.

Ideally, one would be able to compare yield and pollution gradients for both down- and up-wind locations, as wind is a common source of exogenous variation to aid identification [35]. However, given relatively low wind speeds and variable wind directions in the region, we found very little dependence of pollution or yields on direction (figure S5). Nonetheless, other lines of evidence suggest that these yield gradients near power plants arise primarily from air pollution, rather than other variables correlated with power plant locations such as soil quality or farmer behavior. First and foremost, we observe that the gradients increase significantly after power plants turn on, and decrease significantly after they shut down. That is, using the subset of facilities that either turn on ( $n = 111$ ) or shut down ( $n = 44$ ) during the study period, we perform a regression to estimate the effect of a facility turning on or off on the local yield gradient (the average yield at 29 km–31 km distance minus yield at 0 km–2 km distance), using fixed effects for year and location to isolate the effect of

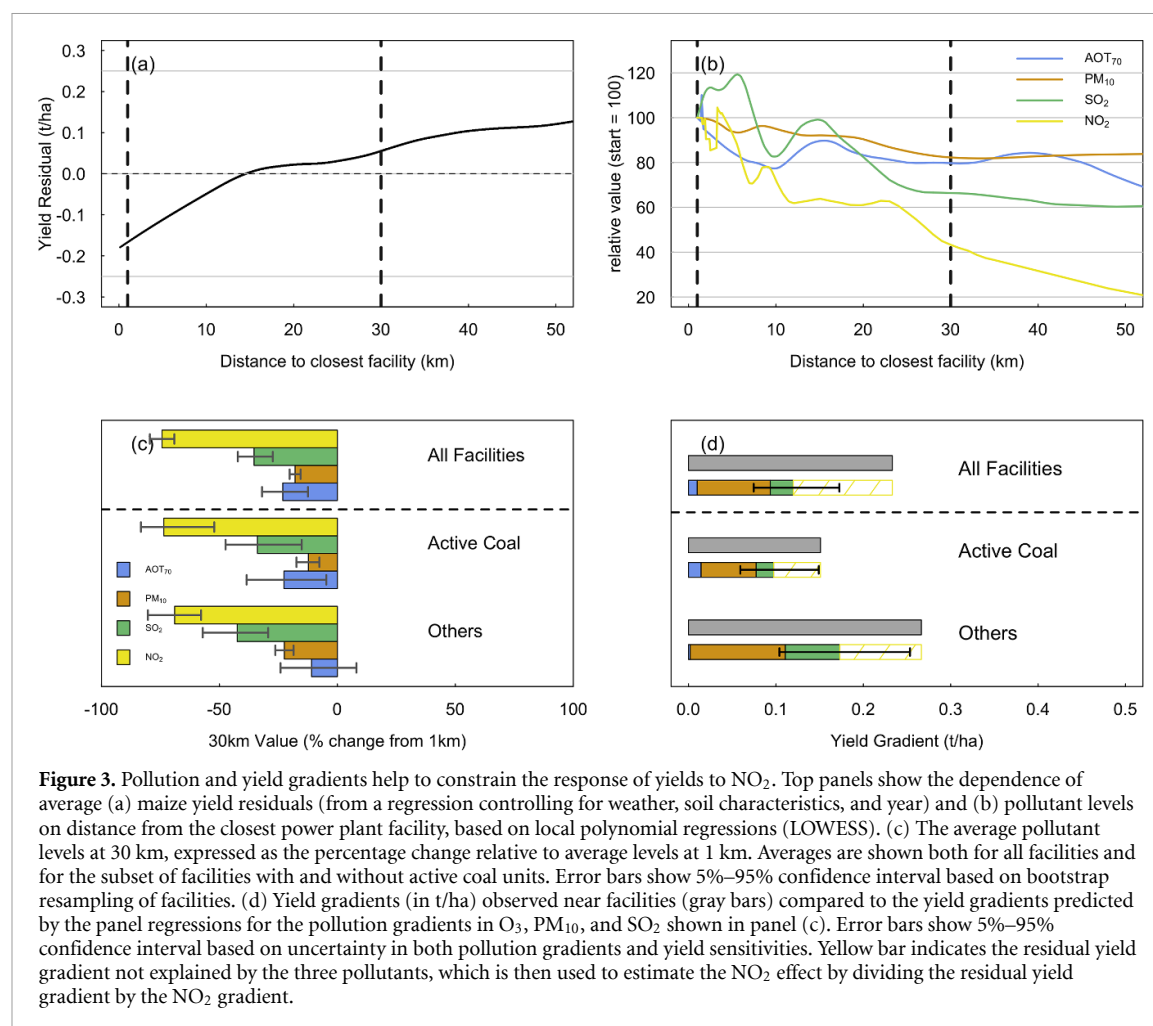


whether the facility is active (table 1). We find a significant increase in the yield gradient after facilities turn on, as well as a significant decline in yield gradients after facilities turn off. Second, we observe only small gradients in observed soil or weather variables near powerplants, with yields predicted based on observed weather and soil (described in section 2.8) exhibiting gradients of less than  $0.1 \text{ t ha}^{-1}$ . While unmeasured aspects of soil quality could be important, we note that the NCCPI variable we use is a common summary of soil quality that captures many physical and chemical aspects of soil believed to influence yield.

Moreover, several lines of evidence suggest that among the pollutants leading to these yield gradients,  $NO_2$  is likely the most important factor: (a) yield gradients are significantly larger near facilities where gradients in  $NO_2$ , as measured in recent years by the Sentinel 5 TROPOMI sensor, are larger (figure S6); (b) yield gradients are also higher near facilities with higher overall  $NO_x$  emissions (figure S6); (c) most importantly, while all pollutants generally decline with distance from power plant,  $NO_2$  exhibits a stronger gradient than the others, likely related to its short atmospheric lifetime (figures 3(b) and (c)). The lack of a rapid decline in  $O_3$  reflects the complex chemistry of  $O_3$  formation and destruction, and is consistent with studies documenting that peaks

in  $O_3$  often occur far from the primary source of  $NO_2$  [13, 36, 37]. Similarly, the gradual PM decline reflects the variety of PM sources and their broad-scale transport outward from facilities. On average,  $SO_2$  initially exhibits an increase with distance from power plants (figure 3(b)), a pattern that is driven entirely by facilities using coal as feedstock (as coal contains much more sulfur and creates much larger quantities of  $SO_2$  during combustion than natural gas).

By combining our estimates of  $O_3$ , PM, and  $SO_2$  effects from the panel estimates (figure 2) with observed relationships between their concentrations and distance to power plant (figure 3(c)), we estimate the expected yield change associated with these pollutants as one moves away from power plants (figure 3(d)). We then estimate the effect of  $NO_2$  based on the residual yield gradient that is not explained by the other pollutants, finding that these gradients provide a much better constraint on  $NO_2$  effects than the panel approach (figure S7). While we use the average gradient for all facilities to estimate  $NO_2$  effects, we also find notable differences when limiting the analysis to only facilities with (or without) active coal units (figures 3(c)–(d)). On average, larger yield gradients are observed near non-coal facilities. This finding is consistent with the larger declines in PM and  $SO_2$  as one moves away from these as compared to facilities with coal units. In both



**Table 1.** Coefficients for regression models that estimate the effect of power plant facilities (1) starting or (2) stopping operation during the study period. The dependent variable in both cases is the yield gradient near the facility, defined as the difference in average yield 29 km–31 km away from the facility and 0 km–2 km away. Values in parentheses show the standard error of the estimate, clustered by year. Asterisks indicate that coefficients are significant at  $p < 0.05$ .

	(1) Start-up model	(2) Shut-down model
PoststartTRUE	0.125* (0.060)	
PostshutTRUE		−0.339* (0.136)
Fixed-Effects:	—	—
Year	Yes	Yes
Facility ID	Yes	Yes
Observations	1928	1326
R <sup>2</sup>	0.58	0.60

cases, the amount of yield gradient attributed to NO<sub>2</sub> (i.e. not explained by the combination of ozone, PM, and SO<sub>2</sub>) is similar, as is the average NO<sub>2</sub> gradient (figure 3(d)).

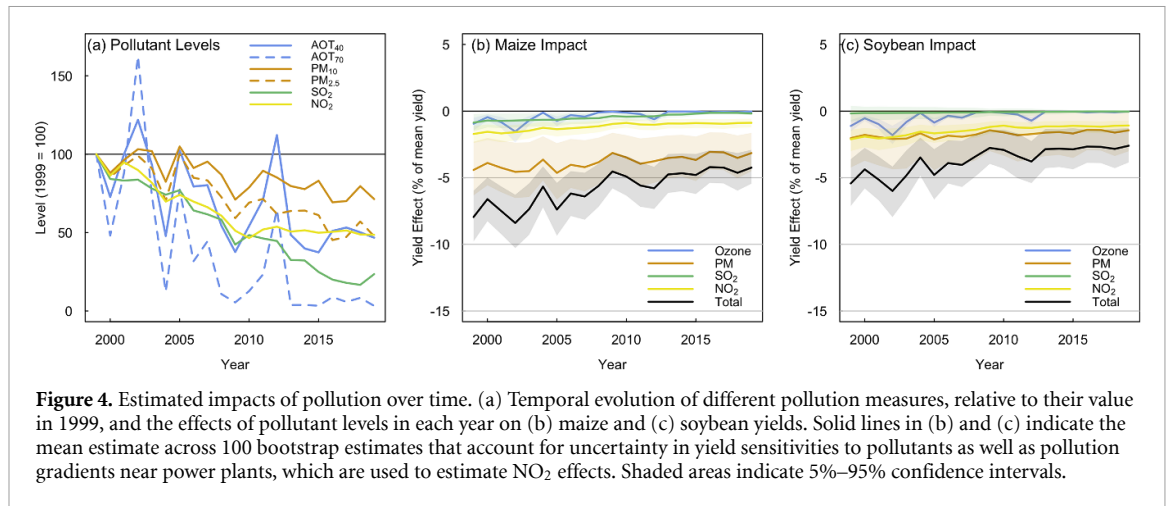
As an additional check on our primary NO<sub>2</sub> estimate we separately perform a cross-sectional regression of yields on satellite-derived NO<sub>2</sub> levels in 2018. These regressions, which control for county fixed-effects as well as weather and soil covariates, also

find significant negative associations between yields and NO<sub>2</sub> (figure S8).

### 3.2. Total pollution impacts on yields are significant but declining

Using the inferred sensitivity of yield to each pollutant, we find the total yield burden in the region for each year by multiplying the derived coefficients by the average exposure across the cropped region. For our baseline specification, we estimate that the total yield losses from pollutants averaged 5.8% for maize and 3.8% for soybean for the 1999–2019 period in the nine-state region (figure 4). Across all specifications, the median estimate was of slightly smaller magnitude than the baseline for maize (5.0%) and larger magnitude for soybean (6.6%). These losses were considerably larger at the start of the period, and gradually declined as air quality improved. On average, we estimate that the reduction in pollutants caused an increase of 4.0% for maize and 3.0% for soybean over the 1999–2019 period for the baseline specification, with a median of 5.0% increase for maize and 6.2% increase for soybean across all specifications. As a percentage of the overall yield gains for each crop over the time period, these values represent 19% for maize and 23% for soybean. Thus, we estimate that the benefits





of cleaner air comprised roughly one-fifth of yield gains since 1999 for both crops.

### 3.3. Results are robust across many plausible model specifications

The results in figure 4 correspond to our baseline specification which uses high resolution satellite-based crop yield estimates, characterizes O<sub>3</sub> exposure as cumulative hours spent above 70 ppb and PM exposure as average daily maximum value of PM<sub>10</sub>, and uses year fixed effects that gives each year its own intercept. These baseline measures for O<sub>3</sub> and PM were based on analysis of out-of-sample prediction error across the various measures (figure S9). Nonetheless, we find that while each of these assumptions has some effect on results, the overall impact of pollutants is fairly robust to each of these (figures S10–S13).

For most specifications, we find that O<sub>3</sub> and SO<sub>2</sub> have smaller impacts in a typical year than PM and NO<sub>2</sub>, with the relative importance of the latter two varying across models. In particular, effects of PM are sensitive to whether effects are estimated using PM<sub>10</sub> or PM<sub>2.5</sub>, and whether using averaging satellite-based yields in a 20 km radius around the monitoring site or using county level yields from NASS. For maize, PM effects are smaller in magnitude, and sometimes even positive in sign, when using county level data. Results for soybean are more consistent across specifications than for maize, with PM again being the most variable.

The consistency in estimates of overall impacts of pollution despite the uncertainty in PM effects results from the use of yield gradients in the vicinity of power plants to constrain estimates of NO<sub>2</sub> effects. Since all pollutant levels are lower at 30 km than 1 km distance from power plants (figure 3), a larger estimated effect for O<sub>3</sub>, PM, or SO<sub>2</sub> implies a smaller effect for NO<sub>2</sub>. Specifically, in cases where the PM effects from the panel regression were strongly negative, a large fraction of the yield gradient near power plants can be explained by PM, and a smaller effect is thus

estimated for NO<sub>2</sub>. In contrast, specifications that result in small effects of PM result in a much smaller fraction of the yield gradient explained by pollutants other than NO<sub>2</sub>. Thus, while PM uncertainty imparts significant uncertainty in NO<sub>2</sub> estimates, their combined effect is constrained by the observed yield gradient across known gradients of pollution.

A similar story emerges for the trend in pollution impacts over time. Total benefits of pollution reductions are fairly similar across all specifications, with the baseline specification ranking near the middle of the various plausible alternatives. In some specifications, including the baseline one, the biggest benefits are estimated to have accrued from reductions in NO<sub>2</sub>, whereas in others the biggest benefits are from reductions in PM or O<sub>3</sub>. O<sub>3</sub> reductions are particularly important in specifications that use higher thresholds for O<sub>3</sub> exposure (i.e. >60 ppb), because these high exposures have been declining faster than exposure to more moderate levels of O<sub>3</sub> (i.e. 40 ppb) or levels of other pollutants such as NO<sub>2</sub> (figure 4(a)).

## 4. Discussion and conclusions

Given the emphasis of much of the literature on O<sub>3</sub> impacts [3, 5, 6, 34], it is somewhat surprising that both NO<sub>2</sub> and PM emerge as more important than O<sub>3</sub> in most of our specifications. We view three factors as potentially important in explaining this apparent discrepancy. First is the relatively low levels of high-O<sub>3</sub> exposure for the study period and region compared to much of the empirical literature, which includes Asia [7, 16] as well as earlier decades and more eastern counties in the United States [3]. Second, much of the empirical literature does not adequately control for the effects of weather when estimating O<sub>3</sub> sensitivity. Among the four major pollutants studied here, O<sub>3</sub> levels are most strongly correlated with the occurrence of extreme heat (table S1), which is known to cause significant yield loss in both maize and soybean. In fact, if we omit extreme heat exposure

(days over 30 °C, or EDD) as a control variable, our estimate of O<sub>3</sub> sensitivity increases by 56%–157%, depending on the crop and source of yield data. Similar increases were observed if we replace EDD with a cruder measure of average daytime maximum temperature, as done in [3]. Thus, while [3] estimate ~10% and ~5% yield losses in the United States circa 2000 for maize and soybean, respectively, whereas we estimate a roughly 3% loss in 2000 (with a range of roughly 0%–6% loss for both crops across different specifications), some of this difference likely results from their incomplete accounting for extreme heat. Although one could argue that some of the effects attributed to extreme heat are in fact from O<sub>3</sub>, several lines of evidence indicate a strong impact of extreme heat even in the absence of O<sub>3</sub>: extreme heat has similar effects in recent decades with much lower O<sub>3</sub> than prior decades [38]; crop model simulations can reproduce the empirical sensitivity to extreme heat even though they do not include O<sub>3</sub> effects [39]; and the model's explanatory power is substantially improved by including extreme heat compared to including only O<sub>3</sub>.

Third, some empirical studies rely on O<sub>3</sub> exposure estimates based on values interpolated between stations. In a sensitivity test, we find that if these interpolations are overly smooth, such as by using inverse distance weighting of all stations [21], estimates of yield sensitivity to O<sub>3</sub> are inflated by as much as a factor of 10 (table S4). This suggests that empirical strategies that use high resolution outcome data and samples near monitors, or that leverage new sources of ground station-based monitoring, may be preferable to interpolation. These findings also suggest the need for research to untangle interpolation and aggregation effects in spatial analyses.

We are not aware of other empirical studies that have estimated effects of NO<sub>2</sub> levels on crop yields, although our results are consistent with many fumigation or observational studies that document phytotoxic effects at levels found in our study region. For example, a transect study for a wild plant species found significant reductions in plant dry weight above 18 ppb NO<sub>2</sub> [11], a level that was frequently exceeded in our region (figure S1). These prior studies also help to elucidate mechanisms, including reduction of photosynthetic rates and increased dark respiration [40, 41], in a way that our empirical approach cannot. Similarly, our study is able to assess impacts in commercial farmers' fields in a way that experimental approaches cannot. Altogether we find consistent evidence that ambient NO<sub>2</sub> causes yield loss for both maize and soybean, although the national aggregate impact is currently modest (~1%), perhaps not surprising given that NO<sub>2</sub> is short-lived and declines rapidly as one moves away from major sources.

We conclude that air pollution has had a significant negative impact on both maize and soybean

yields in the US Corn Belt. In contrast to prior work, we estimate the effects of multiple pollutants, and in doing so identify a relatively large role for PM and NO<sub>2</sub>. We also find that, because of considerable progress in air quality, the yield impact of air pollution is only half of what it was two decades ago. The estimated yield gains associated with the reduction in air pollution has been roughly 20% of total observed gains, indicating that yield growth would have been 20% slower without air quality improvements. Since US maize and soy production is currently valued at ~\$100B/year, in very coarse terms (i.e. not accounting for other factors that might have changed in response to reduced US yields in a counterfactual scenario without reduced pollution) these air quality-driven yield gains correspond to ~\$5B/year.

Finally, our results also highlight the considerable power of using satellite-based yield estimates to understand pollution impacts, as the fine spatial resolution of satellite estimates enables us to exploit pollution gradients near power plants in a manner not possible with county-level administrative yield data. These gradients provide a useful constraint on pollution effects, particularly for NO<sub>2</sub> given the sparse coverage of ground monitoring stations. In addition, the ability to restrict regressions using satellite-based yields to areas near monitoring stations provides an important validation for county-level results. The value of satellite estimates could be even higher in other countries, which typically have both fewer pollution monitors and less administrative yield data than the United States.

## Data availability statement

The data that support the findings of this study are available upon reasonable request from the authors.

## Acknowledgments

We thank Stefania Di Tommaso for help extracting data from Google Earth Engine, Ariel Ortiz-Bobea for providing the scripts to generate specification charts, and Lisa Ainsworth, Marshall Burke, and Tony Fischer for helpful comments. This work was supported by the NASA Harvest Consortium (NASA Applied 787 Sciences Grant No. 80NSSC17K0652, sub-award 54308-Z6059203 to DBL) and NSF #1715557 and NSF/USDA NIFA #1639318.

## ORCID iDs

David B Lobell  <https://orcid.org/0000-0002-5969-3476>

Jennifer A Burney  <https://orcid.org/0000-0003-3532-2934>

## References

- [1] Morgan P B, Ainsworth E A and Long S P 2003 How does elevated ozone impact soybean? A meta-analysis of photosynthesis, growth and yield *Plant Cell Environ.* **26** 1317–28
- [2] Feng Z, Kobayashi K and Ainsworth E A 2008 Impact of elevated ozone concentration on growth, physiology, and yield of wheat (*Triticum aestivum* L.): a meta-analysis *Glob. Change Biol.* **14** 2696–708
- [3] McGrath J M, Betzelberger A M, Wang S, Shook E, Zhu X-G, Long S P and Ainsworth E A 2015 An analysis of ozone damage to historical maize and soybean yields in the United States *Proc. Natl Acad. Sci.* **12** 201509777
- [4] Yi F, Jiang F, Zhong F, Zhou X and Ding A 2016 The impacts of surface ozone pollution on winter wheat productivity in China—an econometric approach *Environ. Pollut.* **208** 326–35
- [5] Avnery S, Mauzerall D L, Liu J and Horowitz L W 2011 Global crop yield reductions due to surface ozone exposure: 1. Year 2000 crop production losses and economic damage *Atmos. Environ.* **45** 2284–96
- [6] Van Dingenen R, Dentener F J, Raes F, Krol M C, Emberson L and Cofala J 2009 The global impact of ozone on agricultural crop yields under current and future air quality legislation *Atmos. Environ.* **43** 604–18
- [7] Gupta R, Somanathan E and Dey S 2017 Global warming and local air pollution have reduced wheat yields in India *Clim. Change* **140** 593–604
- [8] Burney J and Ramanathan V 2014 Recent climate and air pollution impacts on Indian agriculture *Proc. Natl. Acad. Sci.* **111** 16319–24
- [9] Mills G, Hayes F, Simpson D, Emberson L, Norris D, Harmens H and B  ker P 2011 Evidence of widespread effects of ozone on crops and (semi-)natural vegetation in Europe (1990–2006) in relation to AOT40- and flux-based risk maps *Glob. Change Biol.* **17** 592–613
- [10] Shindell D *et al* 2012 Simultaneously mitigating near-term climate change and improving human health and food security *Science* **335** 183–9
- [11] Bell J N B, Honour S L and Power S A 2011 Effects of vehicle exhaust emissions on urban wild plant species *Environ. Pollut.* **159** 1984–90
- [12] WHO 2000 Effects of nitrogen containing air pollutants: critical levels *Air Qual. Guidel. Eur.* 2nd edn (Copenhagen: World Health Organisation) pp 1–28
- [13] Ashmore M R, Bell J N B and Mimmack A 1988 Crop growth along a gradient of ambient air pollution *Environ. Pollut.* **53** 99–121
- [14] Chang T, Zivin J G, Gross T and Neidell M 2016 Particulate pollution and the productivity of pear packers *Am. Econ. J. Econ. Policy* **8** 141–69
- [15] Burney J A 2020 The downstream air pollution impacts of the transition from coal to natural gas in the United States *Nat. Sustain.* **3** 152–60
- [16] Fischer T 2019 Wheat yield losses in India due to ozone and aerosol pollution and their alleviation: a critical review *Outlook Agric.* **48** 181–9
- [17] Lobell D B, Thau D, Seifert C, Engle E and Little B 2015 A scalable satellite-based crop yield mapper *Remote Sens. Environ.* **164** 324–33
- [18] Jin Z, Azzari G and Lobell D B 2017 Improving the accuracy of satellite-based high-resolution yield estimation: a test of multiple scalable approaches *Agric. For. Meteorol.* **247** 207–20
- [19] FAO, Food and Agriculture Organization of the United Nations (FAO), FAO Statistical Databases 2021 (available at: <http://faostat.fao.org>) (Accessed February 2021)
- [20] Mauzerall D L and Wang X 2001 Protecting agricultural crops from the effects of tropospheric ozone exposure: reconciling science and standard setting in the United States, Europe, and Asia *Annu. Rev. Energy Environ.* **26** 237–68
- [21] Boone C D A, Schlenker W and Siikam  ki J 2019 Ground-level ozone and corn yields in the United States
- [22] Gorelick N, Hancher M, Dixon M, Ilyushchenko S, Thau D and Moore R 2017 Google Earth Engine: planetary-scale geospatial analysis for everyone *Remote Sens. Environ.* **202** 18–27
- [23] Johansson R, Effland A and Coble K 2017 Falling response rates to USDA crop surveys: why it matters *Farmdoc Dly.* (7)9, *Dep. Agric. Consum. Econ. Univ. Illinois Urbana-Champaign* (available at: <https://farmdocdaily.illinois.edu/2017/01/falling-response-rates-to-usda-crop-surveys.html>) (19 January 2017)
- [24] Boryan C, Yang Z, Mueller R and Craig M 2011 Monitoring US agriculture: the US department of agriculture, national agricultural statistics service, cropland data layer program *Geocarto Int.* **26** 341–58
- [25] Deines J M, Patel R, Liang S-Z, Dado W and Lobell D B 2020 A million kernels of truth: insights into scalable satellite maize yield mapping and yield gap analysis from an extensive ground dataset in the US corn belt *Remote Sens. Environ.* **112174**
- [26] Daly C, Halbleib M, Smith J I, Gibson W P, Doggett M K, Taylor G H, Curtis J and Pasteris P P 2008 Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States *Int. J. Climatol.* **28** 2031–64
- [27] Lobell D B, Deines J M and Di Tommaso S 2020 Changes in the drought sensitivity of US maize yields *Nat. Food* **1** 729–35
- [28] Abatzoglou J T 2013 Development of gridded surface meteorological data for ecological applications and modelling *Int. J. Climatol.* **33** 121–31
- [29] Berg   L 2018 Efficient estimation of maximum likelihood models with multiple fixed-effects: the {R} package {FENmlm} *CREA Discuss. Pap.* **13** 1–38
- [30] National Resource Conservation Service 2014 Gridded soil survey geographic (gSSURGO) database user guide p 85
- [31] Schiferl L D and Heald C L 2018 Particulate matter air pollution may offset ozone damage to global crop production *Atmos. Chem. Phys.* **18** 5953–66
- [32] Wellburn A R, Higginson C, Robinson D and Walmsley C 1981 Biochemical explanations of more than additive inhibitory effects of low atmospheric levels of sulphur dioxide plus nitrogen dioxide upon plants *New Phytol.* **88** 223–37
- [33] NASS, Quickstats, National Agricultural Statistics Service, US Dept. of Agriculture 2019 (available at: <http://quickstats.nass.usda.gov>)
- [34] Yi F J, Feng J A, Wang Y J and Jiang F 2020 Influence of surface ozone on crop yield of maize in China *J. Integr. Agric.* **19** 578–89
- [35] Deryugina T, Heutel G, Miller N H, Molitor D and Reif J 2019 The mortality and medical costs of air pollution: evidence from changes in wind direction *Am. Econ. Rev.* **109** 4178–219
- [36] Ryerson T B *et al* 2001 Observations of ozone formation in power plant plumes and implications for ozone control strategies *Science* **292** 719–23
- [37] Goldberg D L, Lu Z, Streets D G, De Foy B, Griffin D, McLinden C A, Lamsal L N, Krotkov N A and Eskes H 2019 Enhanced capabilities of TROPOMI NO<sub>2</sub>: estimating NOX from North American cities and power plants *Environ. Sci. Technol.* **53** 12594–601
- [38] Roberts M J and Schlenker W 2011 *The Economics of Climate Change: Adaptations past and Present* ed R H Steckel (Chicago, IL: University of Chicago Press) pp 225–51
- [39] Lobell D B, Hammer G L, McLean G, Messina C, Roberts M J and Schlenker W 2013 The critical role of extreme heat for maize production in the United States *Nat. Clim. Change* **3** 497–501
- [40] Mansfield T A and Freer-Smith P H 1981 Effects of urban air pollution on plant growth *Biol. Rev. Camb. Phil. Soc.* **56** 343–68
- [41] Agrawal M, Singh B, Rajput M, Marshall F and Bell J N B 2003 Effect of air pollution on peri-urban agriculture: a case study *Environ. Pollut.* **126** 323–9