Optimal Targeting of Interventions to Reduce Crop Residue Burning

Martin Kosík

February 2023

Abstract

I apply an air pollution transport model to compute the average population exposure to pollution caused by crop residue burning across different locations in northern India. Using these results together with satellite data on cropland fires, I propose how to optimally target interventions to reduce crop residue burning in order to minimize the total harm caused by the created air pollution subject to a given budget constraint. I develop a structural model of farmers' decisions, which I estimate using the experimental results of Jack et al. (2022), to predict the effects and take-up of an intervention across villages. My preliminary results suggest that the same mass of emissions can cause in expectation twice as many infant deaths depending on the location even if we restrict ourselves only to areas where cropland was burned in the past. Furthermore, I compute how the expected reduction in infant deaths varies with the budget size of a policymaker.

1 Introduction

Biomass burning is responsible for up to 15.4% of infant deaths at the country-level due to the air pollution it creates and this share has been on the rise globally from 2004 to 2018 (Pullabhotla et al., 2022). In addition to the heath effects, adverse impacts of air pollution on productivity (Chang et al., 2016) and human capital (Graff Zivin et al., 2020) have been documented.

Many of the biomass fires are intentionally caused by humans. First, fire is often used to destroy forests in order to transform the land for agriculture (Balboni et al., 2021) . Second, crop residue burning (which is the main focus of this paper) is practiced by some farmers to quickly clear-up the land after a harvest and prepare it for the next sowing (Abdurrahman et al., 2020). The problem with crop residue burning is especially severe in northern India where it has been an important contributor to air pollution during the late fall months (Kulkarni et al., 2020). There has been various interventions proposed to incentivize farmers not to burn their land (Abdurrahman et al., 2020). Most notably, Jack et al. (2022) conducted a RCT to evaluate a simple but potentially highly scalable intervention: offering the farmers a payment conditional on not burning their land.

However, with crop residue burning being fairly widespread (which is the case for for northern India, especially Punjab), a policymaker might not be able to implement a desired intervention in all locations. In this paper, I will calculate how the policymaker should target the interventions aimed to reduce crop residue burning across different locations to achieve the greatest reduction in harm with the limited budget. The local weather patterns and unequal distribution of population in space imply that the number of people exposed to the pollution might vary across the source locations. I will use atmospheric models of air pollution transport to compute average exposure to smoke emitted from different locations in northern India. Combining these results with population density maps enables me to calculate the average number people affected by pollution depending on the source location. Second, I will use a structural model of farmers decisions on crop residue burning using the results of the RCT by Jack et al. (2022), in which the farmers received a payment conditional on not burning their land, for calibration of key parameters. Together with satellite data on fires on agricultural land, this model will give us estimates of the expected reduction of stubble burning achieved by an intervention in a given location (which could be village or other unit

of aggregation depending on the context).

2 Literature review

The impact of air pollution on economic, health, psychological, and political outcomes has been studied extensively. Specifically, particulate matter pollution[1] ($PM_{2.5}$ and PM_{10}) have been shown to have adverse effects on health (Deryugina et al., 2019), productivity (Chang et al., 2016), human capital (Graff Zivin et al., 2020), and academic achievement (Gilraine and Zheng, 2022).

The issue of crop residue burning in particular have also received increased attention in recent years (see Abdurrahman et al., 2020 for a review). In atmospheric sciences, the main focus of the work has been on quantifying the contribution of crop residue burning to overall air pollution (Nair et al., 2020) and studying the chemical, physical and optical properties of the aerosols (Mishra and Shibata, 2012; Ram et al., 2016). In economics and public health, the impact of stubble burning on various outcomes has been examined. The air pollution caused by agricultural fires has been shown to increase both infant and older-age mortality (He et al., 2020; Pullabhotla et al., 2022) and decrease birthweight, gestational length, and in utero survival (Rangel and Vogl, 2019). Graff Zivin et al. (2020) report that crop residue burning reduces performance at collage entrance exams in China. Economic literature on intervention to reduce crop residue burning is much more limited. Most relevant to this paper is the RCT by Jack et al. (2022) conducted in Punjab, India in which payments were made to paddy farmers conditional on them not burning their fields. I describe the RCT in greater detail in subsection 4.1 as its results will be used as an input for my analysis.

In contrast to the studies mentioned above, this paper is focused on optimal targeting of interventions to reduce crop-residue burning, which relate it more to the literature on optimal policies in environmental economics. Assunção et al. (2019) estimate ex-post optimal assignment to "Priority Lists" which were municipalities in Brazil subject to more intense environmental monitoring and enforcement to combat deforestation. The policy of Assunção et al. (2019) aims to minimize total deforestation but in contrast to this paper they do not consider the effect of air pollution and

[1] Particulates (particulate matter) are microscopic particles suspended in the air (Seinfeld and Pandis, 2006, p. 97). PM_{10} to particles that have a diameter smaller or equal to 10 μm whereas $PM_{2.5}$ refers to particles with a diameter smaller or equal to 2.5 μm. These particles are inhalable and can enter the blood stream and brain (especially $PM_{2.5}$ since they tend to be not filtered out in lungs due to their small size) which can cause serious health problems.

heterogeneity in its dispersion.

3 Modelling air pollution transport

I will use HYSPLIT average dispersion (HyADS) approach introduced by Henneman et al. (2019) for modelling the spatial impact of air pollution. The HyADS approach is based on averaging the pollution concentrations predicted by the HYSPLIT[2] model, which is an air pollution transport model developed and maintained by NOAA (Draxler and Hess, 1998; Stein et al., 2015) that computes air parcel[3] trajectories to determine the dispersion of pollution. In contrast to Chemical transport models (CTMs), HYSPLIT abstracts away from simulating complex chemical processes in the atmosphere. While this might lead to less accurate predictions, it also makes HYSPLIT substantially less computationally demanding. HYSPLIT model has been extensively used in the atmospheric sciences. Its applications include tracking and forecasting the release of wildfire smoke (Rolph et al., 2009), wind-blown dust (Ashrafi et al., 2014), volcanic ash (Stunder et al., 2007), and crop residue burning (Liu et al., 2018). Henneman et al. (2021) shows that HyADS have normalized mean errors between 20 and 28% in comparison with the CTMs predictions of PM2.5 source impacts of coal power plants in the US.

For illustration, figure A3 shows the predicted concentrations by HYSPLIT for an emissions event in which 2500 air parcels are released on October 6, 2019 at 9:00 AM. We can see that the wind direction strongly influences the dispersion of the pollution (in this case dispersing it in the south-east direction).

Main output of interest for our analysis is the source-receptor matrix (denoted as SRM_{ij}), which describes the concentration of air pollution (in mass units per m^3) in location j for pollutant emitted from source location i under average weather conditions. Average weather conditions in this context means those that are typical when crop residue burning occurs, which in northern India tends to be October and November (for post-monsoon crop residue burning which is the main focus of this paper). To obtain the source-receptor matrix for a typical crop residue burning event, I simply run

[2]The acronym HYSPLIT stands for Hybrid Single-Particle Lagrangian Integrated Trajectory.

[3]An air parcel is an imaginary body of air which can be assigned basic dynamic and thermodynamic properties of atmospheric air.

the simulations for many emission events with different starting dates and then average the results. An emission events consists of releasing the same amount of emissions at the same time from many source locations (for my preliminary results I use 121 source locations) spread over a regular grid. The dispersion of the pollutants is then computed with the weather data for the given time period while keeping track of the source location of the pollutants so that the source-receptor matrix can be estimated. The technical details of the HYSPLIT simulations used for the preliminary results are described in subsection A.1 in the appendix. Naturally, many simulations need to be run for an accurate estimate of the source-receptor matrix which underscores the computational demands of this project.

Once we have the source-receptor matrix SRM_{ij}, we can estimate the impact of emitting pollution from different locations. Let E_i be the total air pollution emitted from location i and P_i be the total air pollution concentration in i. By definition, we have that $P_j = \sum_i SRM_{ij}E_i$, i.e., the total pollution concentration in j is a sum of the emission from all locations i weighted by the corresponding source-receptor matrix entry. Let us further define $L_j = f(P_j)$ to be loss (harm) to a single person from being exposed to air pollution of concentration P_j. The total loss TL then simply is the sum of the losses across all locations weighted by their population N_j, i.e., $TL = \sum_j L_j \cdot N_j$.[4]

The impact of small change in emissions from i on total loss can be expressed as

$$\frac{\partial TL}{\partial E_i} = \sum_j \frac{\partial L_j}{\partial E_i} N_j = \sum_j \frac{\partial f(P_j)}{\partial P_j} \frac{\partial P_j}{\partial E_i} N_j = \sum_j \frac{\partial f(P_j)}{\partial P_j} SRM_{ij} N_j.$$

Clearly, if a given intervention can achieve the same reduction in emissions at the same costs in all locations then it is optimal to target the locations with the highest $\frac{\partial TL}{\partial E_i}$. Of course, this assumption of uniform effects is not realistic and therefore I will introduce the the full model that takes into account the heterogeneity of costs across locations in section 4.

We can also consider an even simpler case when the loss function is linear ($f(P) = \psi_0 + \psi \cdot P$).

[4]If there are heterogeneous effects of air pollution across individuals (due to e.g., richer households having financial resources to invest into air purifiers), it might be desirable to take these into account when aggregating the loss function. I will consider this in the future work. Nevertheless, the existing studies (Heft-Neal et al., 2018; Heft-Neal et al., 2020) do not show large differences in the effects sizes by wealth levels.

Then $\frac{\partial TL}{\partial E_i}$ simplifies to

$$\frac{\partial TL}{\partial E_i} = \psi \sum_j SRM_{ij} N_j = \psi \cdot \alpha_i,$$

where the term $\alpha_i \equiv \sum_j SRM_{ij} N_j$ is sometimes referred to as the source impact of i in the literature (Henneman et al., 2021). We can see that under those circumstances, the source-receptor matrix and population in each location become sufficient for determining the optimal allocation. This extreme simplicity allows me to present some very preliminary estimates of the distribution of α_i in section 5.

There is some evidence to support linear effects of $PM_{2.5}$ concentrations at least in the case of infant mortality (at least in the policy relevant ranges). In particular, Heft-Neal et al. (2018) shows that higher order polynomials for the effect of post-birth exposure to $PM_{2.5}$ concentrations on the infant mortality are not statistically significant implying linear response function (nevertheless they do find that the quadratic term for in utero exposure is statistically significant).

4 Model

4.1 Farmers responses to intervention

Now, I proceed to modelling the decision of farmers to understand how they would respond to interventions incentivizing them not to burn their land. The main purpose of this model is to obtain reasonable counterfactual estimates of the share of burned land with and without an intervention for every location. I will use the results of the RCT by Jack et al. (2022) which involved offering farmers a payment conditional on not burning their land to estimate the demand for burning using a discrete choice model (in a spirit of Souza-Rodrigues, 2019).

Consider a farmer with a plot of land p used for winter cropping in a location i (which could be a village or a square on a grid). There is a continuum of such plots in every location and a farmer makes a separate decision whether or not to clear the crop residues by burning them. Let Π^B be the value that farmer attains by clearing the land by burning and and Π^N be the value of the alternative

(not burning). I will assume the following functional form

$$\Pi_{pi}^{B} - \Pi_{pi}^{N} = \beta s_i + x_i'\gamma + \epsilon_{pi}$$

where s_i is amount of payment offered conditional on not burning and x_i is a vector of location-level characteristics including soil quality, share of farmers with tractors (obtained from 2011 Socioeconomic and caste census), and the share of land burnt in a previous years. Finally, ϵ_{pi} is a idiosyncratic shock that captures unobserved plot-level factors and is assumed to follow the type 1 extreme value distribution. A plot will be cleared by burning (which I will denote by an indicator variable B_{pi}) if the benefits exceed the costs, i.e., $B_{pi} = \mathbb{1}\left(\Pi_{pi}^{B} > \Pi_{pi}^{N}\right)$. It follows from the above that the share of land cleared by burning in location in i, denoted as b_i, can be expressed as

$$\log\left(\frac{b_i}{1-b_i}\right) = \beta s_i + x_i'\gamma \tag{1}$$

The coefficients in this equation can be estimated using OLS as has been typically done in the literature (e.g., Pfaff, 1999; Souza-Rodrigues, 2019). For this, we will use the data from Jack et al. (2022) who conducted a RCT in 171 villages in Punjab, India, in which they offered farmers varying amounts of payments conditional on farmers not burning their fields. In addition to conditional payment only, Jack et al. (2022) also included a treatment variant in which the farmers received upfront (unconditional) payment upon accepting the contract to help alleviate potential liquidity constraints and build trust. In this paper I do not model the difference between these two treatment variants but it is an interesting area for future research. The full individual-level data from this RCT are not yet available as of October 2022 but I use the aggregated results in a simplified model as described in section 5.

Besides the treatment effect, the enrollment into the program and the nature of the self-selection are important determinants of the cost-effectiveness of an intervention (Jack and Jayachandran, 2019). The higher the enrollment of farmers who would never burn their fields, the lower the per enrollee benefits of the program are. In the extreme case, if the vast majority of farmers who enroll in the program are those that would refrain from burning their fields even in the absence of the intervention, then the cost-effectiveness of the program would be very small.

To parsimoniously capture this effect, the model has only two groups that differ in their enrollment rate. First, the plots that would be burned in the absence of the program have enrollment rate $\omega^B = P(R_{ip} = 1|B_{pi} = 1, s_i = 0)$ where R_{ip} is an indicator for enrollment and $s_i = 0$ corresponds to absence of the program in the location i. Second, the plots that would not be burned in the absence of the program have enrollment rate $\omega^N = P(R_{ip} = 1|B_{pi} = 0, s_i = 0)$. Note that $\omega^B > \omega^N$ implies that there is a positive correlation between the cost of enrollment and the cost of reducing burning and hence lower cost-effectiveness. It simply follows from the above that the location-level enrollment rate, r_i, can then be expresses as

$$r_i = \omega^B b(s_i = 0, x_i) + \omega^N \left(1 - b(s_i = 0, x_i)\right)$$

We can estimate ω^B and ω^N from experimental microdata of Jack et al. (2022) using a method inspired by Jack and Jayachandran (2019). We first use plot-level data from the control group to fit a flexible logit model to obtain estimates of the probability of burning conditional on pre-treatment covariates $\hat{b}_{pi}$ (i.e., the propensity to burn). Due to the random assignment of treatment with respect to location, we should expect the the distribution of the propensities to burn in the control and treatment locations to be the same (ignoring the sampling error). The comparison of the density of $\hat{b}_{pi}$ in the control locations to the corresponding density of those enrolled in the program in the treated locations allows us identify ω^B and ω^N. Notice that for any convex interval $\mathcal{B}$

$$\begin{aligned} P\left(b_{pi} \in \mathcal{B}|R_{pi} = 1, s_i \neq 0\right) =& \omega^B \, \mathbb{E}\left[b_{pi}|b_{pi} \in \mathcal{B}, s_i = 0\right] P\left(b_{pi} \in \mathcal{B}|s_i = 0\right) \\ &+ \omega^N \left(1 - \mathbb{E}\left[b_{pi}|b_{pi} \in \mathcal{B}, s_i = 0\right]\right) P\left(b_{pi} \in \mathcal{B}|s_i = 0\right) \end{aligned}$$

All of the terms above (except ω^B and ω^N) can be estimated from the data. $P\left(b_{pi} \in \mathcal{B}|R_{pi} = 1, s_i \neq 0\right)$ is the density of propensities to burn in a given interval for plots enrolled in the program, $P\left(b_{pi} \in \mathcal{B}|s_i = 0\right)$ is the corresponding density for all plots in the control, and $\mathbb{E}\left[b_{pi}|b_{pi} \in \mathcal{B}, s_i = 0\right]$ is the expected propensity to burn in the given interval in the control. Choosing only two disjoint intervals $\mathcal{B}$ is sufficient for identification of ω^B and ω^N, as it leads to a system of 2 linear equations with 2 unknowns. However, with a finite number of observations, the estimates in the equations are subject to sampling error and therefore it might be preferable to choose a higher number partitions and find

the solution that minimizes the sum of least squared errors.

4.2 Problem formulation

We finally proceed to formulating the problem. I will consider a policymaker with a budget M who chooses the level of the conditional payments s_i offered to farmers for not burning their fields in each location i to minimize the total population-weighted loss caused by air pollution. In my formulation mainly for computational reasons, the policymaker is choosing from a finite set of payment levels, $s_i \in \{\bar{s}_1, \ldots, \bar{s}_J\}$ expressed as money units per hectare.[5] Naturally, no intervention (i.e., $\bar{s}_1 = 0$) is always included in this set.

The policymaker minimizes the total loss (TL), which is a sum of losses (L_j) for all locations j weighted by their population N_j

$$\min_{s_i \in \{\bar{s}_1, \ldots, \bar{s}_J\}} \text{TL} = \min_{s_i \in \{\bar{s}_1, \ldots, \bar{s}_J\}} \sum_j L_j N_j, \tag{2}$$

subject to the budget constraint(where l_i is the total area of eligible land in i and r_i is the enrollment rate in the program, and F are the fixed costs of implementing the intervention)

$$\sum_i r_i s_i l_i + \mathbb{1}\left(s_i > 0\right) F \leq M, \tag{3}$$

the pollution loss function which specifies the harm by air pollution concentration P_j (which is a sum of pollution concentration due to crop-residue burning p_j^b and other sources p_j^0)

$$L_j = f(P_j) = f(p_j^b + p_j^0), \tag{4}$$

the equation for source-receptor matrix decomposition of air pollution

$$p_j^b = \sum_i SRM_{ij} E_i, \tag{5}$$

[5] In the RCT of Jack et al. (2022), the levels of the payments were $\{\bar{s}_1, \bar{s}_2, \bar{s}_3\} = \{0, 800, 1600\}$ denominated in INR per acre. Hence if a policymaker does not want to extrapolate the effect of the intervention beyond those actually implemented in the experiment, he or she should only consider this set of values.

the equation relating the emissions due to crop residue burning (E_i) to the predicted share of land burned ($b(s_i, x_i)$) and eligible land area l_i

$$E_i = \phi b(s_i, x_i) \cdot l_i, \tag{6}$$

the predicted share of land burned given the conditional payment amount (s_i) and the covariates (x_i)

$$b(s_i, x_i) = \frac{\exp(\beta s_i + x_i'\gamma)}{1 + \exp(\beta s_i + x_i'\gamma)}, \tag{7}$$

and the equation for enrollment rate into the program (discussed in subsection 4.1)

$$r_i = \omega^B b(s = 0, x_i) + \omega^N (1 - b(s = 0, x_i)). \tag{8}$$

There are several key parameters and functional form assumptions that need to be specified. First, the policymaker has to set the maximum budget size M. Second, the loss function for air pollution needs to be defined. One of the main concerns with regard to crop residue burning is its effect on mortality via increasing $PM_{2.5}$ concentrations. In my main analysis, I will therefore focus only on the effects on mortality.[6] Regarding the functional form, linear loss ($f(P) = \psi_0 + \psi \cdot P$) substantially reduces the computational complexity of the optimization and allows greater robustness in specifying some of the parameters (i.e., ψ_0, ψ, p_j^0, ϕ have no influence on the optimal allocation of s_i). Moreover, as discussed at the end of section 3, there is some evidence for a linear effect of $PM_{2.5}$ concentration on infant mortality (Heft-Neal et al., 2018). Nevertheless, alternative parametrizations for the effects on mortality have been proposed in the public health literature. In particular, Burnett et al. (2014) suggests the following function form for the effect of $PM_{2.5}$ concentration C on relative risk of mortality RR:

$$RR(C) = \begin{cases} 1 + \alpha \left[1 - \exp\left(-\gamma (C - C_0)^{\delta}\right)\right] & \text{for } C > C_0 \\ 1 & \text{for } C \leq C_0 \end{cases} \tag{9}$$

[6] Other relevant effects of air pollution have been studied and documented (e.g., on productivity or human capital). Nevertheless, these effects tend to be more context specific and less precisely estimated than the impact on mortality and therefore I only consider mortality.

C_0 represents a minimum concentration above which there is evidence indicating health benefits of PM2.5 exposure reductions. This parametrization has been used by several other studies as well (e.g., Apte et al., 2015). Therefore I will consider both of these functional forms to assess the robustness of the final allocations (although I might solve the model with the non-linear loss only for a smaller number of aggregated locations if the computational demands are too high).

Third, the source-receptor matrix needs to be estimated. In this paper, I use the HyADS approach, which I described in greater details in section 3. Fourth, regarding the relationship between a unit of area burned (in our case hectars) and the mass of pollutant emitted (measured in grams), I will again rely on the existing literature in atmospheric science. In particular, I can apply decompositions used by Jain et al. (2014) and Liu et al. (2020) to express ϕ as

$$\phi = CY \times RC \times f_{DM} \times f_{CC} \times EF \tag{10}$$

where CY is the crop yield (produced weight per a unit of area), RC is the residue-to-crop weight ratio, f_{DM} is the dry matter fraction of the crop, f_{CC} is the combustion completeness (fraction of the dry matter burned), and EF is the emission factor for the pollutant. Using the estimates of these parameters from the literature (provided in table A1) for rice paddy as the crop and $PM_{2.5}$ as the pollutant, we get $\phi \approx 23772$. Finally, I already described above how I would obtain the predictions $\hat{b}_i(s_i, x_i)$ in the previous subsection.

4.3 Solving the model

With potentially large number locations and a non-convex objective function, solving the model might be computationally challenging. However, in the case of linear loss and binary intervention (i.e., each location can either receive or not receive an intervention), we can achieve substantial speed-ups by reformulating the model as a knapsack problem[7] (this is demonstrated in subsection A.2 in the appendix). The knapsack problem is NP-Complete and the existing algorithms may, in the worst case, take exponential time (Kellerer et al., 2004, p. 491). However, there is a fully

[7]The knapsack problem is a canonical problem in combinatorial optimization in which the goal is to determine which items (each with its weight and value) to include in order to maximize the value of the included items while not exceeding a given weight limit (the capacity constraint). For more details see Kellerer et al. (2004).

polynomial-time approximation scheme based on dynamic programming that achieves polynomial time in the number of items while controlling the desired approximation error (Kellerer et al., 2004, p. 37). Moreover, the knapsack solvers are efficiently implemented in OR-tools library by Google.[8] The runtime of the OR-tools knapsack solvers I used in section 5 on problems with 8 756 items (villages) and rounding of values to 6 decimal places was in the order of second.

Furthermore, even in the case of linear loss and finite discrete interventions (in which we can assign different levels of payment in each location chosen from a finite set of options), we can still attain significant efficiency gains since the model can be formulated as a multiple-choice knapsack problem. In the multiple-choice knapsack problem, the set of items is partitioned into classes and only one item within a set can be chosen (Kellerer et al., 2004, p. 317). In our case the classes are different types of the interventions (including no intervention) and the constraint enforces that only one type of the intervention will be chosen in a given location. While the multiple-choice knapsack problem is NP-hard, there again exists a fully polynomial-time approximation scheme (Kellerer et al., 2004, p. 338).

Finally, the case of non-linear loss function poses even greater challenge. Nevertheless, there are several heuristic algorithms that have been successfully applied to various optimization problems to find approximate global maximum in a large search space most notably simulated annealing and genetic algorithms (Mitchell, 1998). Simulated annealing methods tend to explore the space very widely in early stages but tend to become more greedy in time (i.e., it tends to select points that are close to currently best solution). Genetic algorithms keep a whole population of solution which are probabilistically mutated, recombined, and discarded. In both cases, the constraint can be included as a penalty into the objective function. Furthermore, regardless of the optimization algorithm selected, the complexity of the problem can always be reduced simply by aggregating some of the locations together and thus reducing the size of the search space.

[8] `https://developers.google.com/optimization`

5 Preliminary results

Since the individual-level data from Jack et al. (2022) were not yet available as of November 2022, I simplify the proposed model so that I can obtain some preliminary results even in the absence of these data. First, I will only consider the case linear loss function and binary intervention. As the intervention, I will consider 3 treatments[9] evaluated by Jack et al. (2022): 1) the conditional payment of INR 800 per acre with 25% of it paid out upfront, 2) the payment of INR 800 per acre with 25% upfront 3) the payment of INR 1600 per acre with 0% upfront. Second, instead of logit model, I will assume that each treatment will reduce the share of burned area by β across all locations. Thus we can then express the share of burned area as

$$b(s_i) = (1 - \beta s_i) b_i^0$$

where s_i is an indicator for location i receiving intervention and b_i^0 is the burned area share in the absence of the intervention. As an estimate for b_i^0, I will use the satellite data estimates for the year prior to the intervention. The estimate of β is calculates as the ratio of the treatment effect to the average share of farmers burning land in the year prior to intervention. The estimated effect of each treatment on burned area is taken from Table A4 in Jack et al. (2022) for the "Not burned, Balanced Accuracy" outcome and the average share of farmers burning land is computed separately for each location based on the MODIS satellite data estimates of cropland and burned land.

I then compute the expected costs of treatment type t in each location i as

$$\text{costs}_{it} = \left(u_t r_t + (1 - u_t)\kappa_i\right) s_t l_i$$

where s_t is the payment offered, u_t is the share paid upfront, r_t is the enrollment rate, l_i is the total eligible land, and κ_i is the share of land belonging to farmers who complied with the program (and therefore will receive the full payment) on the total eligible land. I took the estimates of r_t from Figure A4 in Jack et al. (2022). For calculation of κ_i I assumed the following formula $\kappa_i = \beta_t b_i^0 + 0.06333$ (0.06333 was used to match the relationship between $\beta_t b_i^0$ and *kappa* in Figure

[9] I do not consider the conditional payment of INR 800 per acre with 0% upfront as its treatment effect (for the case of burned land predictions with balanced accuracy) was estimated to be negative and not statistically significant.

A4 of Jack et al. (2022)). All cropland in location i (estimated using the MODIS land use data) is assumed to be eligible for the program. Finally, I abstracted away from any monitoring or fixed costs of the intervention.

The region of interest in this analysis is northwestern India (in particular states of Punjab and Haryana) as crop residue burning is common practice here. Due to the computational demands of the HYSPLIT simulations, I ran them only for 483 source location on a grid covering northwestern India and parts of Pakistan. Figure A2 shows the spatial extent of the grid together with the the cropland and burned area using the predictions from the MODIS satellite data for 2016 at 500m resolution. The expected infant deaths per hectare of burned land are plotted in figure A3a. I then applied bicubic interpolation to increase the resolution of the grid as depicted in figure A3b. Figure A3c shows these interpolated values of the expected infant deaths only for the pixels with burned land in 2019. Finally, figure A3d shows a histograms of these pixel values. These results imply that burning land in locations close to large population centers of Delhi or Lahore can lead to more than twice as many infant deaths as burning in more remote regions even we restrict ourselves only to locations where crop residue burning occurred in the past. The wind patterns also seem to play a role as the areas north-west from Delhi tend to have higher values than locations with similar distance south of Delhi.

In policymaker's problem described in section 4, the interventions can be assigned only at the level of a "location" as the policymaker cannot offer different contracts to farmers within the same location. I define locations in this setting as villages as delineated in the 2001 census map data from Meiyappan et al. (2018). I restrict the analysis to villages in Indian states of Punjab and Haryana as crop residue burning is common there.

Figure A4 depicts the total negative loss (i.e., the total number of infant deaths averted) of optimal allocations of interventions for various levels of the budget constraint and 3 different types of treatment evaluated by Jack et al. (2022). The marginal change in infant lives saved per additional USD 114,975 is shown in figure A5. The marginal impact for all treatments falls sharply for the first several millions of dollars and then continues to decline approximately linearly. This suggests that there could be high gains from optimal targeting for very small budgets

Note that the results are only preliminary and incomplete largely due to the current unavailability

if the individual-level data from Jack et al. (2022). Once these that becomes available, I extend this analysis by relaxing the assumption of uniform costs of the intervention using the full model proposed in section 4.

6 Reduced-form approach

Our main results presented above relied heavily on the accuracy of the air pollution transport modeling. While this has the advantage in allowing us to estimate the impact of air pollution in every location, it requires more intermediate steps (i.e., modelling each step in the chain: area burned $\rightarrow$ pollutant mass emitted $\rightarrow$ $PM_{2.5}$ concentration $\rightarrow$ infant mortality). In this section, I will apply an alternative approach that directly uses the empirical estimates of the effect of area burned on infant mortality. Specifically, this more reduced-form approach is based on the results of Pullabhotla et al. (2022) for the effect of an additional km^2 burned area within a 30 km radius on infant mortality in the downwind direction. The upwind and downwind areas are defined by bisecting the disk with 30 km radius and the center in the birth location along the lines of longitude and latitude into 4 sections of equal area (see figure A6). As expected, Pullabhotla et al. (2022) find significant positive effects on infant mortality only for the burned area in upwind direction of the birth location and no significant effects of the downwind direction.

We can use these results to compute the expected increase in infant deaths per an additional area burned for each location in northwestern India. First, I form a grid of source locations with a resolution of 0.1 latitude degree and compute the population within each quarter of a disk centered at the source with radius 30 km.[10] Then for each source location on the grid, the expected increase in infant deaths per an additional burned hectare (L_i) in location i is calculated as

$$L_i = \sum_d N_{id} \cdot c \cdot f_{id} \cdot \beta_{\text{BA}} \tag{11}$$

where L_i is the expected increase in infant deaths per an additional burned hectare in location i, f_{id} is the relative frequency of historical monthly wind directions, N_{id} is population living in a quarter

[10]The population raster used for this computation is taken from CIESIN (2018) at 0.008333 degree resolution ($\approx$ 1 km).

in the direction d within the 30 km radius of location i, c is the population share of infants (defined as individuals under the age of one), and β_{BA} is the effect of an additional hectare burned on infant mortality.

The data source for the historical monthly wind direction is the NCEP reanalysis (Kalnay et al., 1996) (using only the values for October and November from 1948 to 2022).[11] The estimate of β_{BA} is taken from Pullabhotla et al. (2022, Extended Data Table 2, column (1), Upwind exposure, post-birth) to be 1.06×10^{-5} per hectare (which is 0.00106 per km^2). The value of c was set to $\frac{26}{1210}$ based on the the estimated 26 million new births annually[12] and the total population of 1,210 million from 2011 population census[13].

Figures A7 and A7 show the map and histogram, respectively, of the expected infant deaths per hectare of burned land for both approaches. The reduced-form approach clearly has higher mean and variance which is likely a consequence of the fact that in the reduced-form approach the air pollution does not effect locations farther than than 30 km from the source whereas the HYSPLIT average dispersion approach allow for more realistic dispersion. Nevertheless, it is somewhat reassuring that the mean effects for both approaches are roughly of the same order of magnitude despite the differences in the methods.

7 Conclusion

In this paper, I proposed how to optimally target interventions to reduce crop residue burning in order to minimize the total harm caused by the created air pollution. The preliminary results I presented suggest that there could potentially be meaningful efficiency gains from targeting the interventions optimally, especially for relatively small budgets. Nevertheless, there remains much work to be done in future research. Firstly and most importantly, the full analysis proposed in this could not be executed since the individual-level data from Jack et al. (2022) were not yet published. Secondly, there are rather purely technical improvements that could strengthen the credibility of the results. This includes increasing the number of source locations, simulating more emission events

[11] In particular, the variables `vwnd.10m.mon.mean.nc` and `uwnd.10m.mon.mean.nc` were downloaded from `https://downloads.psl.noaa.gov/Datasets/ncep.reanalysis/Monthlies/surface_gauss/`

[12] `https://nhm.gov.in/index1.php?lang=1&level=2&sublinkid=819&lid=219`

[13] https://www.census2011.co.in/

and using meteorological data with greater resolution. Finally, there are possible new directions in which to extend this project could be extended. There are other sources of biomass burning such as forest fires and slash-and-burn agriculture, which are at least partially caused by human activity (Balboni et al., 2021). The modelling framework I developed could be applied with some modifications to these problems as well to better understand the costs and benefits of possible interventions in different regions of the world.

References

Abdurrahman, Muhammad Isa, Chaki, Sukalpaa and Saini, Gaurav (2020), 'Stubble burning: Effects on health & environment, regulations and management practices', *Environmental Advances* 2, p. 100011, DOI: 10.1016/j.envadv.2020.100011.

Akagi, S. K., Yokelson, R. J., Wiedinmyer, C., Alvarado, M. J., Reid, J. S., Karl, T., Crounse, J. D. and Wennberg, P. O. (2011), 'Emission factors for open and domestic biomass burning for use in atmospheric models', *Atmospheric Chemistry and Physics* 11 (9), pp. 4039–4072, DOI: 10.5194/acp-11-4039-2011.

Apte, Joshua S., Marshall, Julian D., Cohen, Aaron J. and Brauer, Michael (2015), 'Addressing Global Mortality from Ambient PM2.5', *Environmental Science & Technology* 49 (13), pp. 8057–8066, DOI: 10.1021/acs.est.5b01236.

Ashrafi, Khosro, Shafiepour-Motlagh, Majid, Aslemand, Alireza and Ghader, Sarmad (2014), 'Dust storm simulation over Iran using HYSPLIT', *Journal of Environmental Health Science and Engineering* 12 (1), p. 9, DOI: 10.1186/2052-336X-12-9.

Assunção, Juliano, McMillan, Robert, Murphy, Joshua and Souza-Rodrigues, Eduardo (2019), *Optimal environmental targeting in the amazon rainforest*, tech. rep., National Bureau of Economic Research.

Balboni, Clare, Burgess, Robin and Olken, Benjamin A. (2021), *The Origins and Control of Forest Fires in the Tropics*, tech. rep., Tech. Rep.

Burnett, Richard T., Pope, C. Arden, Ezzati, Majid, Olives, Casey, Lim, Stephen S., Mehta, Sumi, Shin, Hwashin H., Singh, Gitanjali, Hubbell, Bryan, Brauer, Michael, Anderson, H. Ross, Smith, Kirk R., Balmes, John R., Bruce, Nigel G., Kan, Haidong, Laden, Francine, Pr, üss-Ustün Annette, Turner, Michelle C., Gapstur, Susan M., Diver, W. Ryan and Cohen, Aaron (2014), 'An Integrated Risk Function for Estimating the Global Burden of Disease Attributable to Ambient Fine Particulate Matter Exposure', *Environmental Health Perspectives* 122 (4), pp. 397–403, DOI: 10.1289/ehp.1307049.

Chang, Tom, Graff Zivin, Joshua, Gross, Tal and Neidell, Matthew (2016), *The Effect of Pollution on Worker Productivity: Evidence from Call-Center Workers in China*, Working Paper 22328, Series: Working Paper Series, National Bureau of Economic Research, DOI: 10.3386/w22328.

CIESIN, - Center for International Earth Science Information Network (2018), *Gridded Population of the World, Version 4 (GPWv4): Population Count, Revision 11*, Palisades, New York, DOI: 10.7927/H4JW8BX5.

Deryugina, Tatyana, Heutel, Garth, Miller, Nolan H., Molitor, David and Reif, Julian (2019), 'The Mortality and Medical Costs of Air Pollution: Evidence from Changes in Wind Direction', *American Economic Review* 109 (12), pp. 4178–4219, DOI: 10.1257/aer.20180279.

Draxler, Roland R. and Hess, G. D. (1998), 'An overview of the HYSPLIT_4 modelling system for trajectories', *Australian meteorological magazine* 47 (4), pp. 295–308.

Gilraine, Michael and Zheng, Angela (2022), *Air Pollution and Student Performance in the U.S.* Working Paper 30061, Series: Working Paper Series, National Bureau of Economic Research, DOI: 10.3386/w30061.

GOI, DAC (2018), 'Agricultural Statistics at a glance 2018', *Government of India, Ministry of Agriculture & Farmers Welfare, Department of Agriculture, Cooperation & Farmers Welfare, Directorate of Economics and Statistics*.

Graff Zivin, Joshua, Liu, Tong, Song, Yingquan, Tang, Qu and Zhang, Peng (2020), 'The unintended impacts of agricultural fires: Human capital in China', *Journal of Development Economics* 147, p. 102560, DOI: 10.1016/j.jdeveco.2020.102560.

He, Guojun, Liu, Tong and Zhou, Maigeng (2020), 'Straw burning, PM2.5, and death: Evidence from China', *Journal of Development Economics* 145, p. 102468, DOI: 10.1016/j.jdeveco.2020.102468.

Heft-Neal, Sam, Burney, Jennifer, Bendavid, Eran and Burke, Marshall (2018), 'Robust relationship between air quality and infant mortality in Africa', *Nature* 559 (7713), pp. 254–258, DOI: 10.1038/s41586-018-0263-3.

Heft-Neal, Sam, Burney, Jennifer, Bendavid, Eran, Voss, Kara K. and Burke, Marshall (2020), 'Dust pollution from the Sahara and African infant mortality', *Nature Sustainability* 3 (10), pp. 863–871, DOI: 10.1038/s41893-020-0562-1.

Henneman, Lucas R. F., Choirat, Christine, Ivey, Cesunica, Cummiskey, Kevin and Zigler, Corwin M. (2019), 'Characterizing population exposure to coal emissions sources in the United States using the HyADS model', *Atmospheric Environment* 203, pp. 271–280, DOI: 10.1016/j.atmosenv.2019.01.043.

Henneman, Lucas R. F., Dedoussi, Irene C., Casey, Joan A., Choirat, Christine, Barrett, Steven R. H. and Zigler, Corwin M. (2021), 'Comparisons of simple and complex methods for quantifying exposure to individual point source air pollution emissions', *Journal of Exposure Science & Environmental Epidemiology* 31 (4), pp. 654–663, DOI: 10.1038/s41370-020-0219-1.

Jack, B. Kelsey and Jayachandran, Seema (2019), 'Self-selection into payments for ecosystem services programs', *Proceedings of the National Academy of Sciences* 116 (12), pp. 5326–5333, DOI: 10.1073/pnas.1802868115.

Jack, B. Kelsey, Jayachandran, Seema, Namrata, Kala and Rohini, Pande (2022), *Money (Not) to Burn: Payments for Ecosystem Services to Reduce Crop Residue Burning*, Working Paper.

Jain, Niveta, Bhatia, Arti and Pathak, Himanshu (2014), 'Emission of air pollutants from crop residue burning in India', *Aerosol and Air Quality Research* 14 (1), pp. 422–430.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne, Roy and Joseph, Dennis (1996), 'The NCEP/NCAR 40-Year Reanalysis Project', *Bulletin of the American Meteorological Society* 77 (3), Publisher: American Meteorological Society Section: Bulletin of the American Meteorological Society, pp. 437–472, DOI: 10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2.

Kellerer, Hans, Pferschy, Ulrich and Pisinger, David (2004), *Knapsack Problems*, Berlin: Springer.

Kulkarni, Santosh H., Ghude, Sachin D., Jena, Chinmay, Karumuri, Rama K., Sinha, Baerbel, Sinha, V., Kumar, Rajesh, Soni, V. K. and Khare, Manoj (2020), 'How Much Does Large-Scale Crop Residue Burning Affect the Air Quality in Delhi?', *Environmental Science & Technology* 54 (8), pp. 4790–4799, DOI: 10.1021/acs.est.0c00329.

Lasko, Kristofer and Vadrevu, Krishna (2018), 'Improved rice residue burning emissions estimates: Accounting for practice-specific emission factors in air pollution assessments of Vietnam', *Environmental Pollution* 236, pp. 795–806, DOI: 10.1016/j.envpol.2018.01.098.

Liu, Tianjia, Marlier, Miriam E., DeFries, Ruth S., Westervelt, Daniel M., Xia, Karen R., Fiore, Arlene M., Mickley, Loretta J., Cusworth, Daniel H. and Milly, George (2018), 'Seasonal impact of regional outdoor biomass burning on air pollution in three Indian cities: Delhi, Bengaluru, and Pune', *Atmospheric Environment* 172, pp. 83–92, DOI: 10.1016/j.atmosenv.2017.10.024.

Liu, Tianjia, Mickley, Loretta J., Singh, Sukhwinder, Jain, Meha, DeFries, Ruth S. and Marlier, Miriam E. (2020), 'Crop residue burning practices across north India inferred from household survey data: Bridging gaps in satellite observations', *Atmospheric Environment: X* 8, p. 100091, DOI: 10.1016/j.aeaoa.2020.100091.

Meiyappan, P., Roy, P. S., Soliman, A., Li, T., Mondal, P., Wang, S. and Jain, A. K. (2018), *India Village-Level Geospatial Socio-Economic Data Set: 1991, 2001*, Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC).

Mishra, Amit Kumar and Shibata, Takashi (2012), 'Synergistic analyses of optical and microphysical properties of agricultural crop residue burning aerosols over the Indo-Gangetic Basin (IGB)', *Atmospheric Environment* 57, pp. 205–218, DOI: 10.1016/j.atmosenv.2012.04.025.

Mitchell, Melanie (1998), *An Introduction to Genetic Algorithms*, Reprint edition, Cambridge, Mass.: MIT Press.

Nair, Moorthy, Bherwani, Hemant, Kumar, Suman, Gulia, Sunil, Goyal, Sanjeev and Kumar, Rakesh (2020), 'Assessment of contribution of agricultural residue burning on air quality of Delhi using remote sensing and modelling tools', *Atmospheric Environment* 230, p. 117504, DOI: 10.1016/j.atmosenv.2020.117504.

Pan, X. L., Kanaya, Y., Wang, Z. F., Komazaki, Y., Taketani, F., Akimoto, H. and Pochanart, P. (2013), 'Variations of carbonaceous aerosols from open crop residue burning with transport and its implication to estimate their lifetimes', *Atmospheric Environment* 74, pp. 301–310, DOI: 10.1016/j.atmosenv.2013.03.048.

Pfaff, Alexander S. P. (1999), 'What Drives Deforestation in the Brazilian Amazon?: Evidence from Satellite and Socioeconomic Data', *Journal of Environmental Economics and Management* 37 (1), pp. 26–43, DOI: 10.1006/jeem.1998.1056.

Pullabhotla, Hemant Kumar, Zahid, Mustafa, Heft-Neal, Sam, Rathi, Vaibhav and Burke, Marshall (2022), 'Global biomass fires and infant mortality', Publisher: EarthArXiv.

Ram, Kirpa, Singh, Sunita, Sarin, M. M., Srivastava, A. K. and Tripathi, S. N. (2016), 'Variability in aerosol optical properties over an urban site, Kanpur, in the Indo-Gangetic Plain: A case study of haze and dust events', *Atmospheric Research* 174-175, pp. 52–61, DOI: 10.1016/j.atmosres.2016.01.014.

Rangel, Marcos A. and Vogl, Tom S. (2019), 'Agricultural Fires and Health at Birth', *The Review of Economics and Statistics* 101 (4), pp. 616–630, DOI: 10.1162/rest_a_00806.

Rolph, Glenn D., Draxler, Roland R., Stein, Ariel F., Taylor, Albion, Ruminski, Mark G., Kondragunta, Shobha, Zeng, Jian, Huang, Ho-Chun, Manikin, Geoffrey, McQueen, Jeffery T. and Davidson, Paula M. (2009), 'Description and Verification of the NOAA Smoke Forecasting System: The 2007 Fire Season', *Weather and Forecasting* 24 (2), pp. 361–378, DOI: 10.1175/2008WAF2222165.1.

Seinfeld, John H. and Pandis, Spyros N. (2006), *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change*, 2nd edition, Hoboken, N.J: Wiley-Interscience.

Souza-Rodrigues, Eduardo (2019), 'Deforestation in the Amazon: A Unified Framework for Estimation and Policy Analysis', *The Review of Economic Studies* 86 (6), pp. 2713–2744, DOI: 10.1093/restud/rdy070.

Stein, A. F., Draxler, R. R., Rolph, G. D., Stunder, B. J. B., Cohen, M. D. and Ngan, F. (2015), 'NOAA's HYSPLIT Atmospheric Transport and Dispersion Modeling System', *Bulletin of the American Meteorological Society* 96 (12), Publisher: American Meteorological Society Section: Bulletin of the American Meteorological Society, pp. 2059–2077, DOI: 10.1175/BAMS-D-14-00110.1.

Stunder, Barbara J. B., Heffter, Jerome L. and Draxler, Roland R. (2007), 'Airborne Volcanic Ash Forecast Area Reliability', *Weather and Forecasting* 22 (5), pp. 1132–1139, DOI: 10.1175/WAF1042.1.

A Appendix

A.1 Description of HYSPLIT simulations

In general, I tend to follow the setup and the parameters of Henneman et al. (2019), nevertheless I deviate from their approach in certain aspects, which I will describe below, due the differences in goals (Henneman et al., 2019 aim to estimate exposure to actual historical pollution from coal power plants whereas I am interested in counterfactual exposure) and context (coal power plants in the US vs. crop residue burning in India).

As mentioned in section 3, I estimate the source-receptor matrix by averaging the source-receptor concentration estimates across emission events. An emission event proceed as follows: An unit mass of pollutant is released from each source at the height 7.5 meters. The 121 source locations were spread over a regular rectangular grid with latitude ranging from 40° to 45° and longitude from 78° to 70°. The HYSPLIT model then tracks the dispersion of 2500 air parcels for each sources (this number was chosen to balance the fidelity and computational demands of the simulations) for 4 full days (96 hours). The emission event length of 4 days was chosen based on the results of Pan et al. (2013) who measured that the approximate atmospheric lifetime of carbonaceous aerosols from crop residue burning is 1 to 6 days. The source-specific concentrations were calculated for each location on a grid with a resolution 0.05° of latitude and longitude. The concentration grids were averaged over the emission events to produce a single grid of the mean concentrations for each source location. These resulting grids formed our final estimate of the source-receptor matrix.

The start dates for the emission events were either 9 AM on October 1 or 9 AM on October 20 both for years 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2015, 2016, and 2017 (hence 20 emission events in total). I used the NCEP reanalysis (Kalnay et al., 1996) for historical meteorological data mainly due to its global coverage, long temporal coverage, and decent resolution.

A.2 Reformulation to knapsack problem

In case of linear loss function and binary intervention, we can reformulate the model as a canonical knapsack problem. First, let us first denote

$$b_i^0 = \hat{b}_i(s_i = 0, x_i)$$

and

$$b_i^1 = \hat{b}_i(s_i = \bar{s}^T, x_i)$$

therefore we can express $b_i(s_i, x_i)$ as

$$b_i(s_i, x_i) = b_i^1 \cdot \mathbb{1}\left(s_i = s^T\right) + b_i^0.$$

By plugging this into the objective function and dropping the linear terms that do not depend on s_i and all the scaling terms, we get

$$\underset{s_i \in \{0, \bar{s}^T\}}{\arg\min} = \sum_j \sum_i N_j SRM_{ij} l_i b_i^1 \cdot \mathbb{1}\left(s_i = s^T\right)$$

Since maximization of the negative of a function is equivalent to minimzation of the original function, we can write

$$\underset{s_i \in \{0, \bar{s}^T\}}{\arg\max} = \sum_i \sum_j (-1) N_j SRM_{ij} l_i b_i^1 \cdot \mathbb{1}\left(s_i = s^T\right)$$

If we denote the values to be $v_i = \sum_j (-1) N_j SRM_{ij} l_i b_i^1$, the weights to be $w_i = \bar{s}^T l_i$ and binary treatment to be $t_i = \mathbb{1}\left(s_i = s^T\right)$, we arrive at the canonical formulation of the knapsack problem:

$$\underset{s_i \in \{0, \bar{s}^T\}}{\arg\max} = \sum_i v_i \cdot t_i$$

subject to

$$\sum_i w_i \cdot t_i \leq M$$

For linear loss and finite discrete interventions, it can be easily shown that the multiple knapsack

formulation can be obtained by simply including the additional interventions as new items and by imposing new constraints requiring only up to one intervention to be chosen in each location.

Table A1: Emission conversion parameters for rice paddy and $PM_{2.5}$

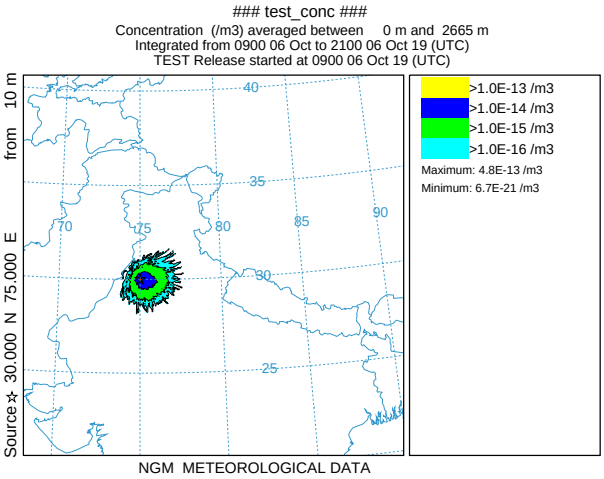
Parameter	Description	Units	Value	Source
CY	Crop yield (production per area)	$\frac{kg}{ha}$	3774	GOI (2018, p. 150) [14]
RC	Residue-to-crop ratio	unitless	1.5	Jain et al. (2014)
f_{DM}	Dry matter(DM) fraction of the crop	unitless	0.86	Jain et al. (2014)
f_{CC}	Combustion completeness	unitless	0.78	Lasko and Vadrevu (2018) [15]
EF	Emission factor for $PM_{2.5}$	$\frac{g}{kg}$	6.26	Akagi et al. (2011)

[14] The average of crop yield for Punjab (4366) and Haryana (3181) in 2017-18 season was used since two states cover the most of our area of interest.

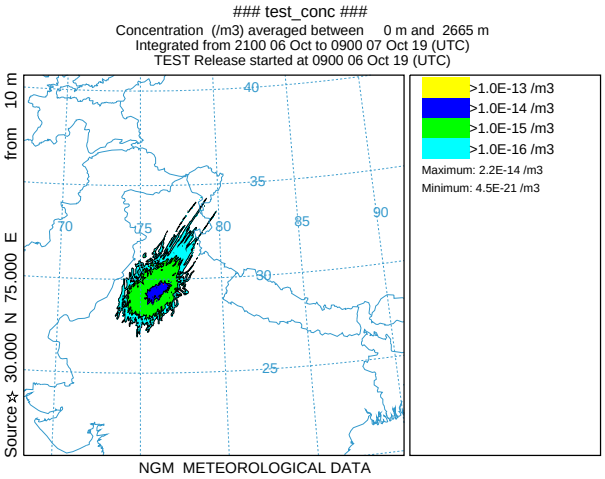
[15] The average of values for complete burn (0.89) and partial burn (0.67) was used.

Figure A1: $PM_{2.5}$ dispersion from an illustrative emission event

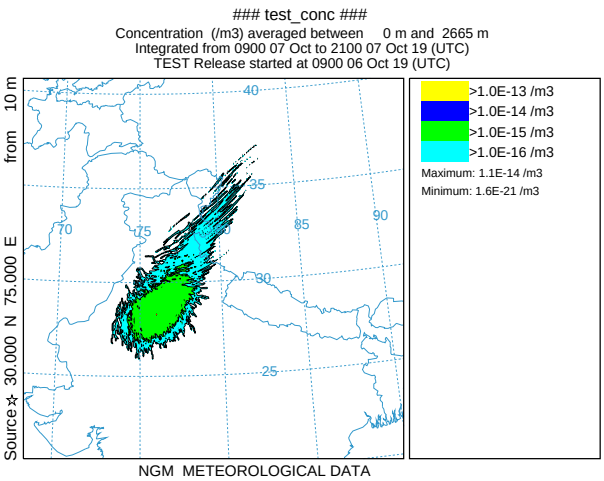
(a) 0 to 12 hours after release



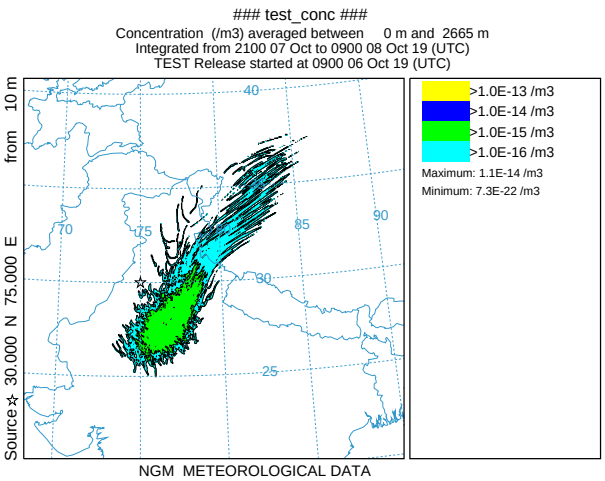
(b) 12 to 24 hours after release



(c) 24 to 36 hours after release



(d) 36 to 48 hours after release



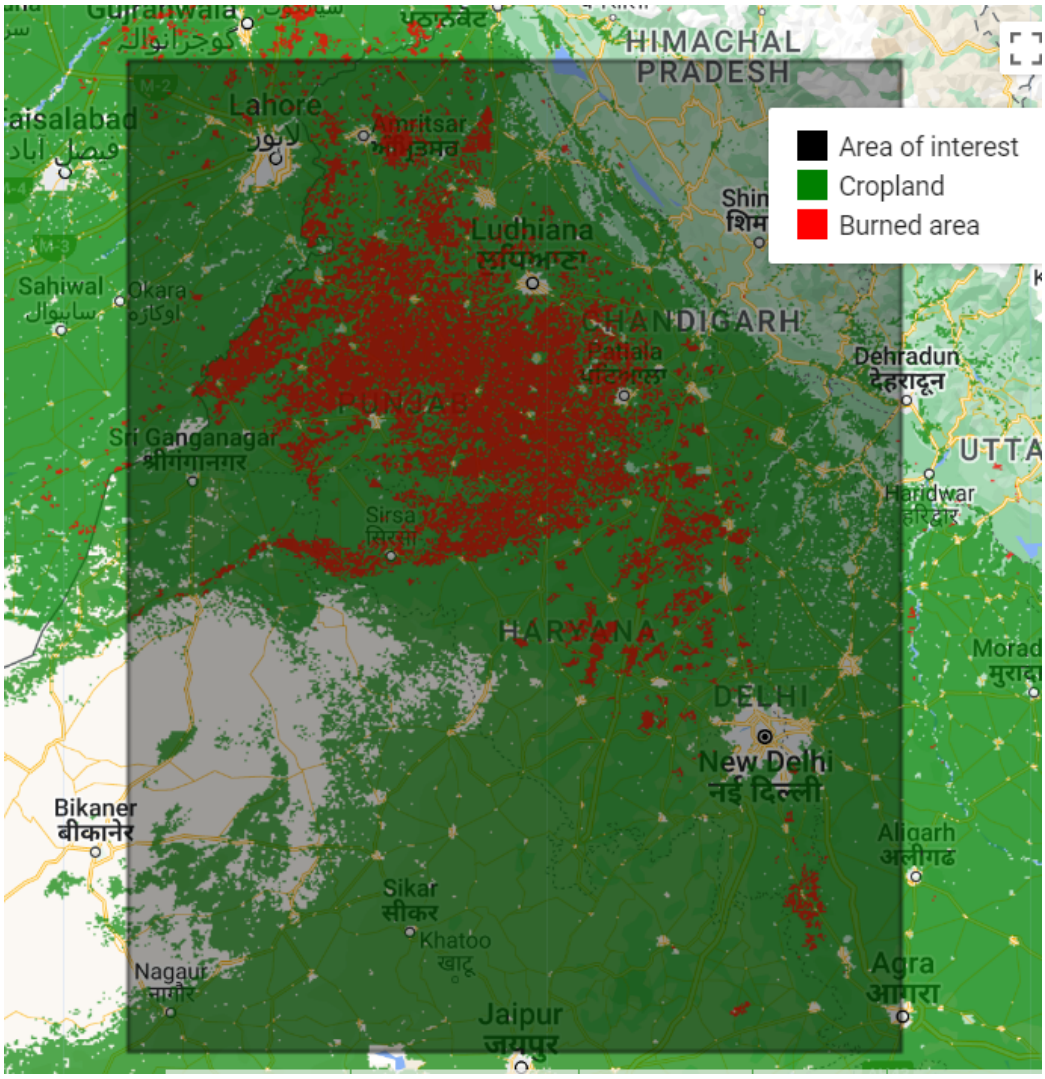
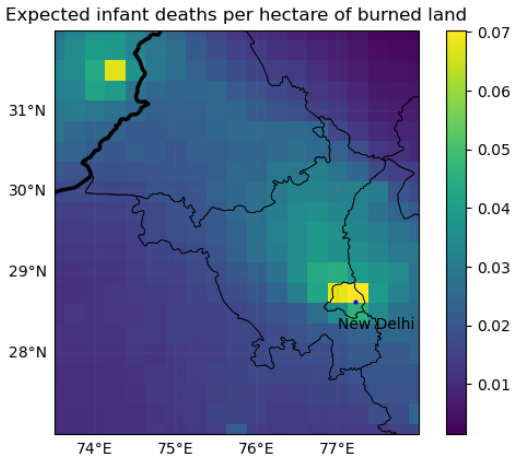


Figure A2: Cropland and burned area in 2016

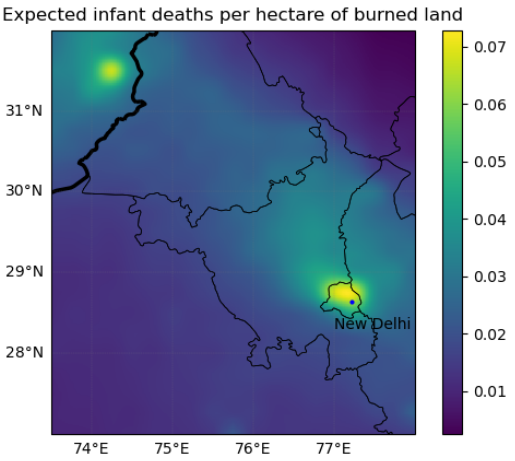
Note: The source of the burned area raster is MODIS MCD64A1 product and the source of the cropland raster is MODIS MCD12Q1 product both at 500m resolution

Figure A3: Expected infant deaths per hectare of burned land

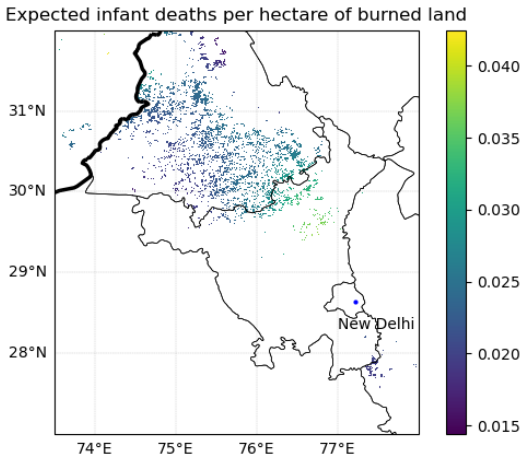
(a) No interpolation



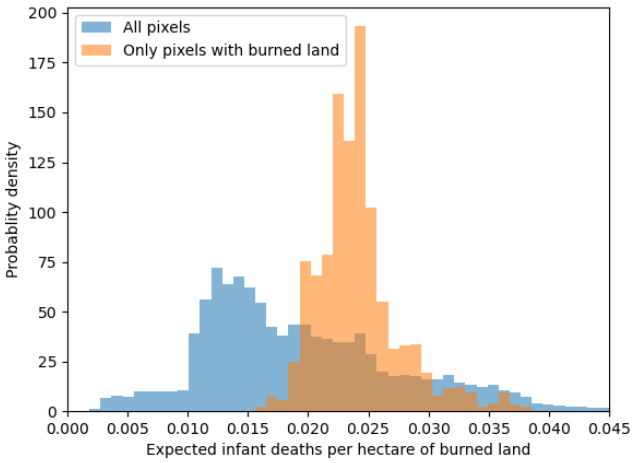
(b) Bicubic interpolation



(c) Only burned areas in 2019 (with bicubic interp.)



(d) Histogram of pixel values



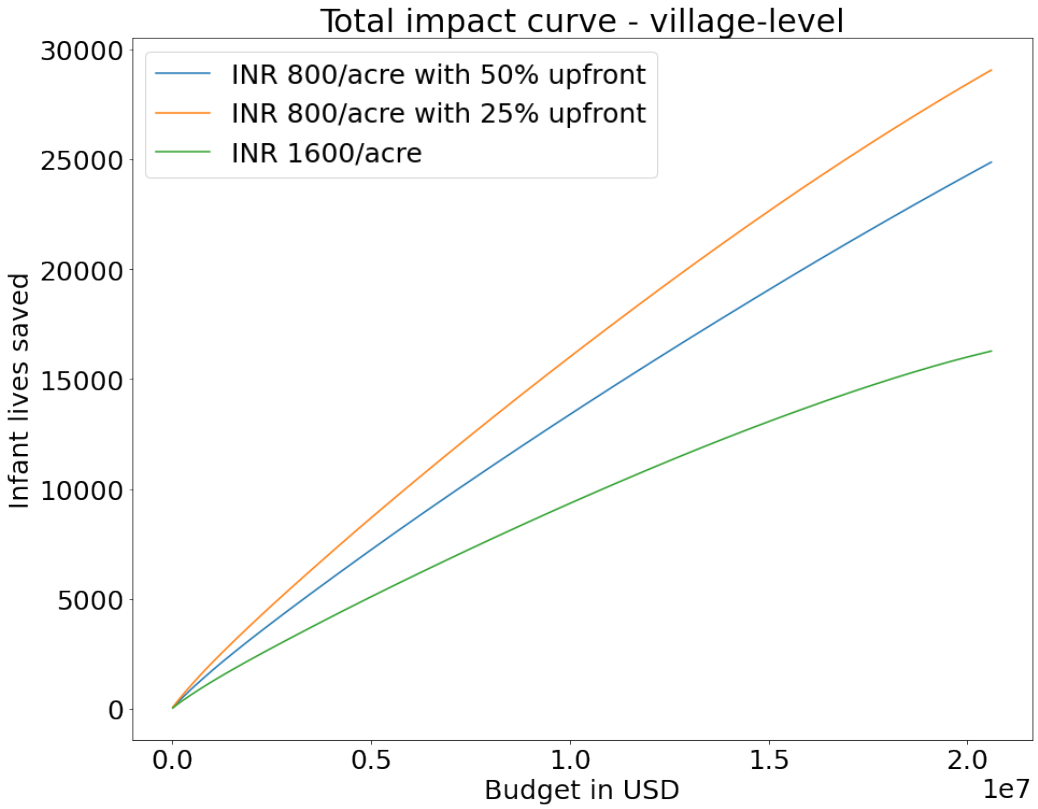


Figure A4: Total impact curve under linear loss (village-level)

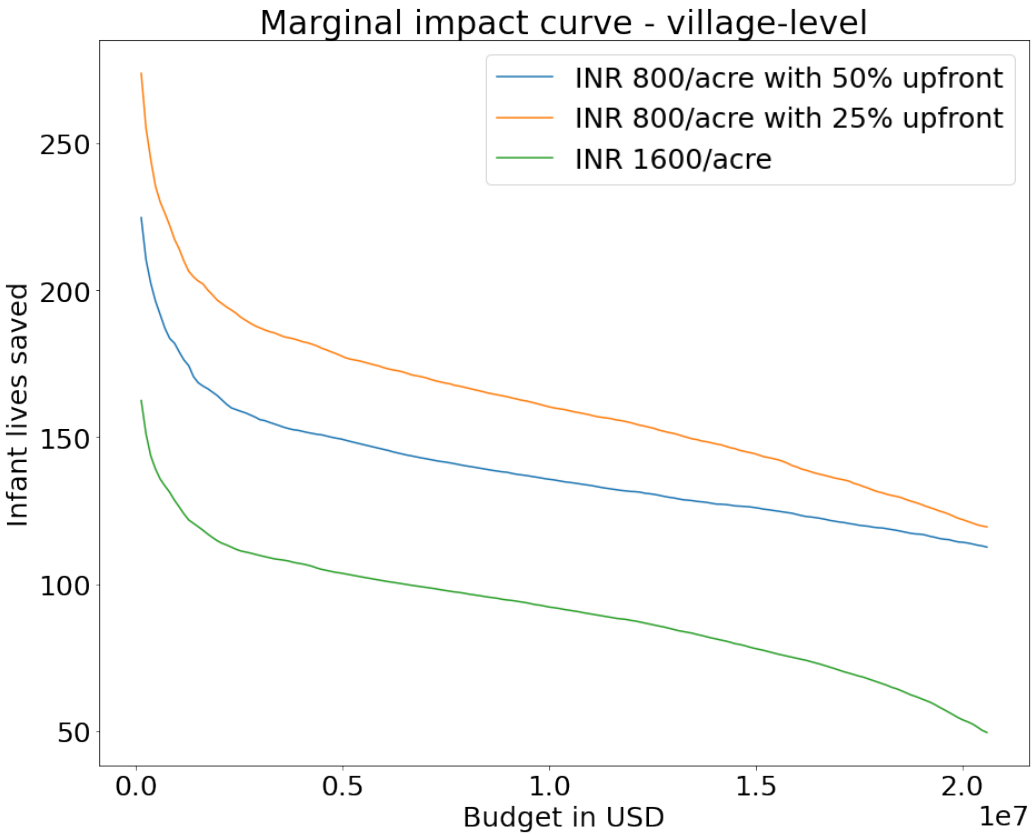


Figure A5: Marginal impact curve under linear loss (village-level)

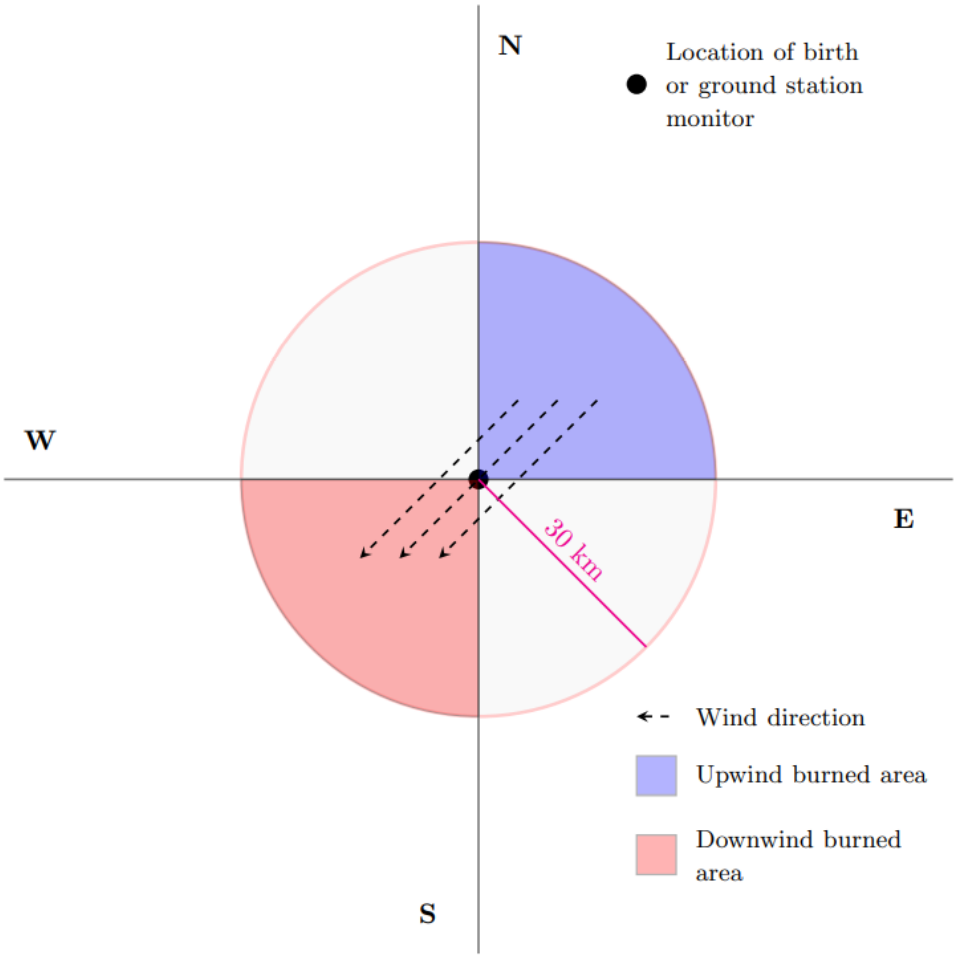
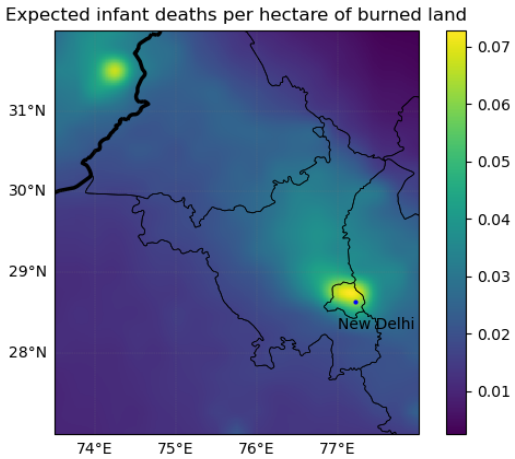


Figure A6: Schematic showing definition of up and downwind burned areas in Pullabhotla et al. (2022)

Source: Extended Data Fig 2. in Pullabhotla et al. (2022)

Figure A7: Comparison of HyADS and reduced-form approaches - map

(a) HYSPLIT average dispersion approach (HyADS)



(b) Reduced-form approach

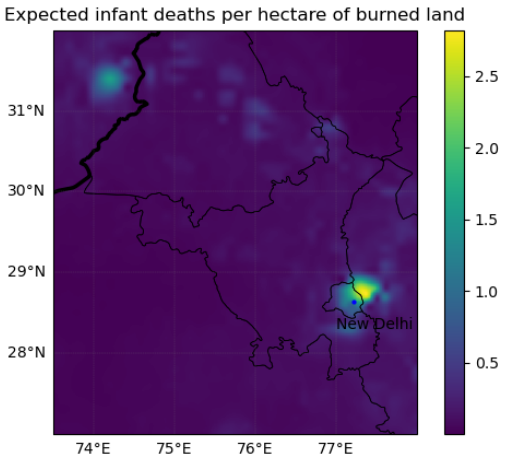
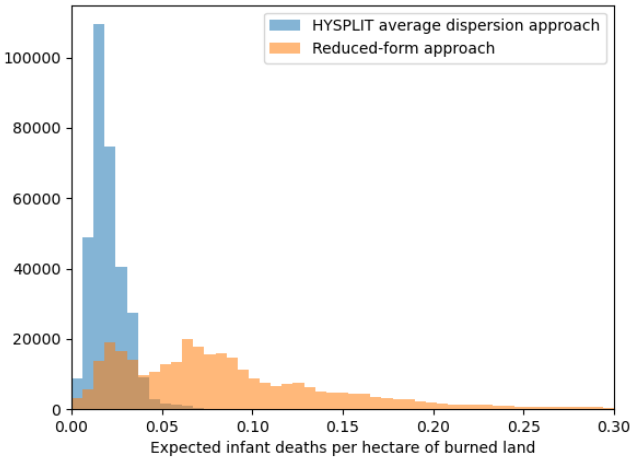


Figure A8: Comparison of HyADS and reduced-form approaches - histogram

(a) All pixels



(b) Only pixels with burned land in 2019