Identifying climate-vulnerable populations using satellite imagery

Jonathan Proctor[1†], Luke Sherman[2], Jeanette Tseng[2], and Solomon Hsiang[2†]

[1]Food and Resource Economics, University of British Columbia

[2]Global Policy Lab, Stanford University

[†]correspondence: jon.proctor@ubc.ca, solhsiang@stanford.edu

Feburary 17, 2026

Abstract

Understanding how climate hazards like extreme heat, drought, and flooding influence human wellbeing is critical for preparing societies for extreme weather and climate change. Yet empirical estimates of how vulnerability to climate hazards differs across space is limited by both a lack of knowledge about and data on determinants of climate vulnerability. Here, we leverage the rich and diverse information within high-resolution optical satellite imagery to predict local climate vulnerability in the U.S. and globally. Including imagery information into empirical models of how temperature and moisture influence agricultural productivity, how temperature influences human mortality and how flood depth damages buildings, reveals substantial local differences in climate vulnerability and resilience and improves predictions of historic and future damages from climate hazards.

Introduction

Quantifying how the climate influences human societies and ecosystems is critical for managing current climate hazards, adapting to future climate changes, and designing climate policy (Carleton and Hsiang, 2016; Carleton and Greenstone, 2021). While substantial progress has been made in understanding how a diversity of social outcomes respond, on average, to changes in climatic conditions, it it remains less understood how responses differ from one location to another. Better quantification of local vulnerability to climate hazards would improve predictions of climate-related damages and their distribution across populations. This would, in turn, benefit climate change adaptation and disaster response efforts, including improving insurance design and other risk management strategies.

Here, we develop an approach to estimate local climate vulnerability using satellite imagery and apply it to a diverse set of social outcomes. We use satellite imagery because it is a rich and universally available source of information (Rolf et al., 2021; Burke et al., 2021; Jean et al., 2016; Head et al., 2017; Khachiyan et al., 2022). Using our approach, the determinants of climate vulnerability – such as whether an agricultural field is irrigated – can be learned and measured directly from the imagery, social outcome data and weather data, rather than having to be specified and collected by the researcher. This allows models to capture additional determinants of vulnerability, including those that are unknown or unmeasured but visible in the imagery. Application of this approach reveals substantial differences in local climate vulnerability across space for agriculture, human mortality, and buildings. The approach identifies spatial heterogeneity in vulnerability not captured by traditional approaches, leading to substantially improved understanding of which communities are relatively sensitive and resilient to weather shocks and climate change.

Modern empirical approaches for quantifying the influence of climate variables on social systems at scale use high performance computing to combine statistical methods for causal inference with large datasets on human outcomes and climate variables (Carleton et al., 2024). A primary product of this body of work are dose-response functions, which measure how the social outcome (e.g. human mortality rate) responds to changes in the climate variable (e.g. temperature). The steepness of the response curve indicates the sensitivity of the social outcome to the climate variable, with steeper response curves indicating greater vulnerability (Fig. 1A). A key challenge in the empirical climate impacts literature is the difficulty resolving differences in climate vulnerability across populations, and studies tend to estimate a single average response for the entire population (e.g., Schlenker and Roberts (2009); Hsiang, Meng and Cane (2011); Burke, Hsiang and Miguel (2015); Missirian and Schlenker (2017); Carleton (2017); Proctor et al. (2018); Proctor (2021); Opper, Park and

Husted (2023)).

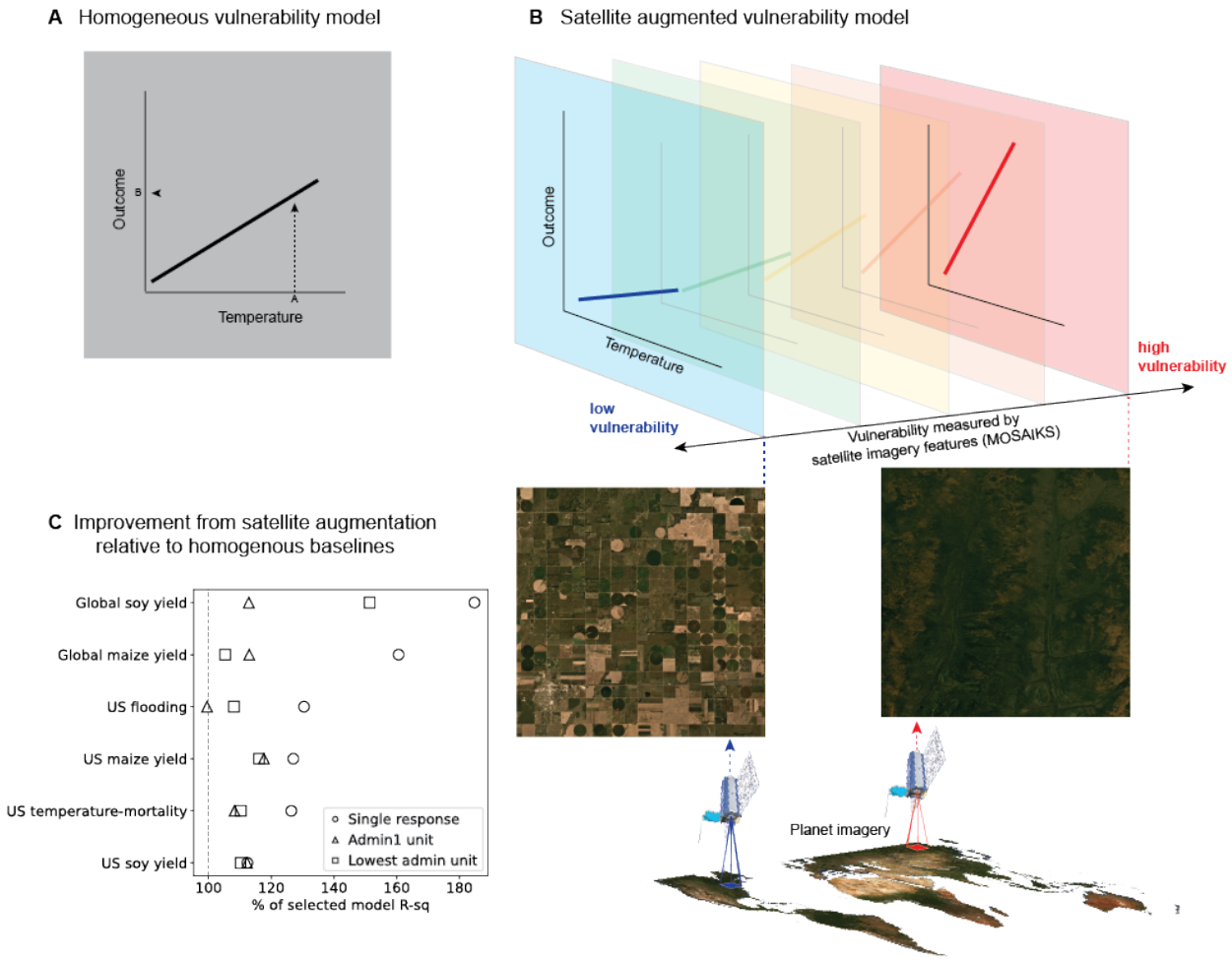


Figure 1: **Predicting vulnerability to climate hazards using satellite imagery** A) Empirical approaches combine social outcome data and weather data to estimate climate-outcome dose-response functions. Steeper response functions indicate greater vulnerability and flatter response functions indicate lesser vulnerability. B) Our approach uses satellite imagery to estimate local climate vulnerability – i.e. local dose-response function between social outcomes and climate variables. C) When applied to a variety of social outcomes our approach improves predictive performance relative to models using a single homogeneous effect (circles), spatial features such as the county, state, or country of an observation (triangles), and climatological features such as the average temperature of an observation (squares). For spatial features, here we compare to an indicator for the panel unit of observation; performance using coarser spatial units is shown in Tables S1,S2,S3,S4.

Two main approaches have been developed to estimate local climate vulnerability. The first estimates how vulnerability differs across populations using observed determinants of climate vulnerability or proxies thereof, such as irrigation (Zaveri and B. Lobell, 2019; Proctor et al., 2022), air conditioning (Park et al., 2020), income (Carleton et al., 2022), or previous

exposure to heat and cold (Heutel, Miller and Molitor, 2021; Auffhammer, 2022; Baylis, 2020; Rode et al., 2021; Hultgren et al., 2025). Using observed determinants, however, is limited by the need for *a priori* knowledge of and data on the drivers of vulnerability, which can be difficult to obtain, and using proxies for determinants can have relatively limited predictive power. The second approach estimates local vulnerability using spatial information – for example, estimating a separate dose response curve in each county or state (e.g., Butler and Huybers (2013); Rising and Devineni (2020); Baylis (2020); Auffhammer (2022)). Such models, however, are unable to pool the influence of climate in areas that are spatially distant but similar in adaptive characteristics, leading to relatively high parameter uncertainty and low predictive performance, as shown below. Further, they are unable to make predictions of climate vulnerability outside the areas where they were trained.

Our approach predicts local climate vulnerability based on what an area looks like in high-resolution optical satellite imagery (Fig. 1B). Instead of having to specify and obtain data on the determinants of climate vulnerability, it learns and measures these determinants directly, incorporating all drivers of vulnerability visible in the imagery. The use of satellite imagery also enables the approach to pool information from locations that look similarly vulnerable, even if they are far from each other, which provides greater statistical power and predictive performance than independent estimations of vulnerability at each location. This approach complements existing approaches, as information from satellite imagery, known determinants, and spatial features can all be used together to explain local climate vulnerability.

Our approach is designed to be both generalizable and accessible. It can be used without modification to measure the local climate vulnerability of a diverse set of social outcomes. It is relatively easily to use because it has a structure that builds on prior approaches for estimating climate vulnerability, and because it incorporates information from satellite imagery using a set of global pre-computed satellite-image features (Carleton, Chong, Druckenmiller, Noda, Proctor, Rolf and Hsiang, 2022; Rolf et al., 2021). This means that the approach can be used for new variables or in new locations of interest without a researcher having to manipulate any raw imagery.

While we apply this approach to estimate local climate vulnerability, the same tools could in principle be used to estimate local treatment effects in any panel data setting where the determinants of effect heterogeneity are visible in the imagery. This contributes to the literature on estimation of conditional average treatment effects (e.g., Künzel et al., 2019; Wager and Athey, 2018) by allowing heterogeneity in treatment effects to be estimated using satellite imagery. Incorporating imagery information into existing approaches for estimating conditional average treatment effects is challenging because it is high-dimensional and lacks

structure. Jerzak, Johansson and Daoud (2023) address this using a Bayesian convolutional neural network and apply their method to estimate imagery-based treatment effects in a randomized control trial. Methodologically, we complement and extend this work by developing an approach that is: compatible with panel fixed-effect regression research designs; compatible with irregularly shaped imagery like that of countries, provinces, or municipalities; relatively simple to understand and easy to implement; computationally efficient to facilitate uncertainty quantification through bootstrapping; and amenable to methods for interpreting the potential drivers of effect heterogeneity.

Predicting local climate vulnerability using satellite imagery

The vulnerability of social outcomes, Y to changes in climate variables, C, is commonly estimated using a panel fixed effect linear regression framework (Carleton and Hsiang, 2016; Hsiang, 2016):

$$Y_{it} = \beta C_{it} + \text{controls} + \epsilon_{it} \tag{1}$$

where i indexes location and t indexes time. Differences in climate vulnerability, β, across populations due to known determinants of vulnerability such as irrigation capacity, or proxies thereof such as historical climate or income, are commonly estimated using interaction terms in the regression model (e.g., Carleton et al. (2022); Rode et al. (2021); Proctor et al. (2022)):

$$\beta_i = \beta_0 + \gamma_1 \text{irrigation}_i + \gamma_2 \text{climate}_i + \gamma_3 \text{income}_i + ... + \gamma_n \text{covariate}_i^n \tag{2}$$

A similar approach is used to model vulnerability as a function of location: $\beta_i = \alpha_i \lambda_i$, where λ_i is a vector of indicator variables for each location (e.g., Butler and Huybers (2013); Rising and Devineni (2020); Baylis (2020)).

Here, we use a similar regression model structure to estimate local climate vulnerability using the rich information within satellite imagery. We focus our analysis on differences in vulnerability across space, but not over time. To extract the information on climate vulnerability contained in the imagery, we use the Multi-task Observation using Satellite Imagery and Kitchen Sinks (MOSAIKS) approach. MOSAIKS image features, $\mathbf{X}$, have been shown to predict a wide range of social and environmental outcomes with skill competitive with more expensive computer vision algorithms like convolutional neural networks (Rolf et al., 2021) and with other generalizable satellite embeddings (Proctor et al., 2025). Here, we use these satellite image features to capture information on the determinants of climate

vulnerability:

$$\begin{aligned} \text{climate}_i &= \alpha_1 \mathbf{X} + \epsilon_1 \\ \text{income}_i &= \alpha_2 \mathbf{X} + \epsilon_2 \\ \text{irrigation}_i &= \alpha_3 \mathbf{X} + \epsilon_3 \\ &\vdots \\ \text{covariate}_i^n &= \alpha_n \mathbf{X} + \epsilon_n. \end{aligned} \tag{3}$$

Substituting Equations 3 into Equation 2 shows how predicting climate vulnerability using image information can be conducted in a single step:

$$\begin{aligned} \beta_i &= \beta_0 + \gamma_1(\alpha_1 \mathbf{X}_i) + \gamma_2(\alpha_2 \mathbf{X}_i) + \gamma_3(\alpha_3 \mathbf{X}_i) + ...\gamma_n(\alpha_n \mathbf{X}_i) \\ &= \beta_0 + \underbrace{(\gamma_1\alpha_1 + \gamma_2\alpha_2 + \gamma_3\alpha_3 + ...\gamma_n\alpha_n)}_{\psi} \mathbf{X}_i \\ &= \beta_0 + \psi \mathbf{X}_i \end{aligned}$$

Substituting into Equation 1, gives the general specification of our regression model that we apply to each social outcome:

$$Y_{it} = \underbrace{\beta_0 C_{it}}_{\text{homogeneous response}} + \underbrace{\psi \mathbf{X}_i C_{it}}_{\text{satellite-based heterogeneity}} + \text{controls} + \epsilon_{it} \tag{4}$$

A benefit of this approach is that it predicts vulnerability based on all determinants of vulnerability that are visible in the imagery, including those that may not be known *a priori*, or for which traditional sources of data do not exist.

Intuition for the information on the determinants of climate vulnerability contained in the imagery can be gleaned from visual analysis. Figure S1 shows three examples of maize growing counties across the U.S. with differing agricultural practices and natural endowments that are visible in the imagery and could influence climate vulnerability. These include farm infrastructure such as roads, buildings and plot size; the presence and type of surrounding vegetation, which could inform climatologic and soil conditions as well as pest pressure or control; and hydrologic features like rivers and lakes which could indicate a high water table or surface water irrigation. Information on these features are extracted from the imagery and contained in $\mathbf{X}$. Model estimation learns from the data what aspects of the imagery indicate greater vulnerability or resilience (ψ).

Model training and evaluation By combining methods from econometrics, computer vision, and machine learning our approach delivers causal estimates of local vulnerability with high predictive performance. We evaluate whether using satellite imagery improves estimates of climate vulnerability by testing whether 1) it accurately recovers prescribed local climate vulnerability in simulation experiments and 2) it increases model predictive performance in empirical applications. Predictive performance is evaluated on a held-out test set using the coefficient of determination (R^2) (Fig. S2). To flexibly control model complexity and avoid over-fitting to the training data, models of local climate vulnerability are estimated using ridge regression rather than ordinary least squares regression. To obtain an unbiased average treatment effect but allow for heterogeneity in effect to the degree it improves performance, we impose a ridge penalty on the coefficients for the satellite features (ψ) and on the coefficients for other drivers of effect heterogeneity, but not the coefficients for the average treatment effect (β_0). We tune model hyperparameters using five-fold cross-validation on 80% of the data and test on the remaining 20%. For model tuning and evaluation we split the data randomly, by spatial groups, and by temporal groups to test performance under different types of extrapolation. Model variables are residualized with respect to controls prior to model training and evaluation to isolate plausibly random variation in climate variables that is used to estimate effects on the social outcome that can be interpreted as causal (Proctor et al., 2022; Benveniste, Huybers and Proctor, 2025; Hsiang, 2016). For example, yield and temperature data are demeaned and detrended by spatial and temporal fixed effects in our application to U.S. agriculture; this means that effects of temperature on yields are estimated by comparing variations in temperature within a location over time – that are arguably as good as randomly assigned – to variations in yield. Uncertainty of local climate vulnerability is estimated by calculating the sampling distribution of estimates using a block bootstrap procedure (Fig. S3). MOSAIKS satellite image features are calculated from a global cross-section of 2019 Planet Labs optical (i.e., red, green, and blue) imagery with a pixel resolution of 4.77m×4.77m at the equator. The image features are time-invariant and thus predict climate vulnerability over space (β_i), but do not directly predict the temporal variation in the social outcome (Y) that is used for model evaluation. Thus, improvements in model performance from including imagery information stem solely from improvements in the estimates of climate vulnerability. The linear model structure enables the model to be trained on and make predictions for irregularly shaped spatial units (Sherman et al., 2023). Model training and evaluation is described in greater detail in the methods and supplement.

Results

We estimate the local vulnerability of agricultural productivity to temperature and soil moisture shocks, human mortality to temperature shocks and buildings to flooding. In each setting, we first replicate prior estimates of the average climate vulnerability across the entire population. We then, holding other aspects of the model fixed, extend the model to predict local climate vulnerability using information from the satellite imagery. We compare these satellite-based estimates to predictions of local climate vulnerability using conventionally modeled determinants of vulnerability including average temperature, income, irrigation, or location. Across settings, allowing the local climate vulnerability to vary according to information in the satellite imagery improves model predictive performance relative to the homogeneous model by 15-180% and generally outperforms conventional approaches (Fig. 1C). We detail results from each setting below, beginning with a set of simulation experiments.

Simulation experiments Before analyzing real social outcome data – where true climate vulnerability is unknown – we first conduct a simulation experiment in which we simulate social outcome data using a prescribed local climate vulnerability, and then test the ability of our approach to recover the known local vulnerability. Local vulnerability is prescribed based on measured variables such as irrigation infrastructure and population density, and this is combined with observed weather data to simulate the social outcome data. These simulations indicate that: (1) the imagery-based approach can explain a substantial fraction of local climate vulnerability, (2) the approach predicts local climate vulnerability roughly as well as the satellite features it uses predict the determinants of vulnerability, and (3) performance predicting local vulnerability correlates with performance predicting the outcome, but is generally substantially higher. See Supplementary Information A for details, as well as for additional results regarding model uncertainty quantification and vulnerability downscaling performance.

U.S. Agriculture Analyzing county-level maize yields from 2000 to 2019 using an experimental design based on Schlenker and Roberts (2009) (Eqn. 5), we find that allowing the temperature response of yields to differ across counties based on the satellite imagery of each county increases the amount of interannual variability in yield explained by temperature and precipitation by 26% for maize and 11% for soybeans, relative to the standard model with a single homogeneous effect (Fig. 2 A-B, Table S1). Specifically, model R^2 predicting annual county yield anomalies from trend increases from 0.34 to 0.42 for maize and 0.23 to 0.26 for soybeans ($p<0.01$ for both crops). For maize, this performance increase

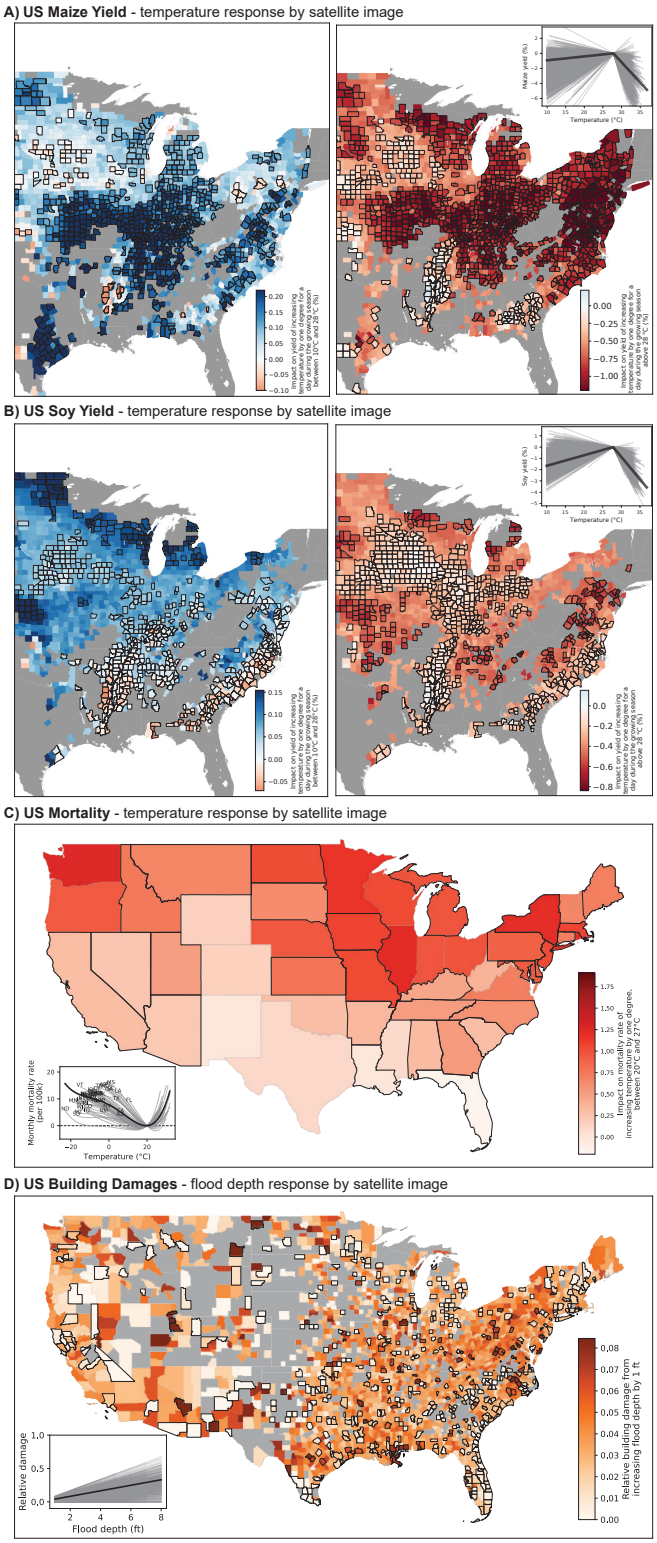


Figure 2: **Local vulnerability of U.S. crop yields to temperature, human mortality to temperature, and buildings to flooding predicted using satellite imagery.** Each response function indicates the vulnerability for a given county (for agriculture and buildings) or state (for mortality). Response functions are shown across the historic range of local exposure for each administrative unit. Maps show how the predicted benefits of warming and damages from extreme heat, as well as flooding, differ across space. Gray areas are not estimated.

stems from better capturing differences in temperature vulnerability across space, with the largest improvements in prediction accuracy from farms bordering the Mississippi river in Arkansas, Mississippi and Louisiana, as well as farms in the central Eastern Seaboard and in south-western Texas (Fig. S4). Predicting local vulnerability using satellite imagery performs better than using irrigation, average temperature, or spatial features, such as the state or county of the observation, for maize yields (Table S1, Fig. S5). For soybean yields, modeling local vulnerability by state outperforms the satellite-based model (Fig. S6). For both crops, the best performing model predicts local vulnerability using a combination of satellite imagery, spatial features, and irrigation, and explains 43% of the year-to-year variance of maize yields and 27% of soybean yields.

The value of using ridge regression to flexibly adjust model complexity to optimize performance is shown by the improved performance relative to models that do not impose a ridge penalty when estimating local climate vulnerability using spatial features (R^2 is 0.35 for maize and 0.23 for soybeans with the ridge penalty and 0.28 and 0.09 without when using county-level spatial features, Table S1). This unpenalized approach is equivalent to estimating a separate regression for each spatial unit, as in (e.g., Butler and Huybers, 2013), which can lead to coefficients that are poorly constrained due to the relatively low number of observations used to estimated each coefficient. Penalizing the coefficients on the spatial features encourages the model to allow a spatial unit's effect to differ from the average treatment effect only if that difference is supported by the data. This leads the spread of coefficients around the average treatment effect to be smaller, but each coefficient to be better constrained, which leads to better predictive performance because model error is composed of both model bias and model variance (Fig. S7) (Hastie, Tibshirani and Friedman, 2009). Including satellite imagery in the model allows the model to share information between spatial units that have similar characteristics, which improves performance by allowing the data to support a greater spread of estimated local vulnerability.

Agricultural climate vulnerability differs substantially across the U.S. (Fig. 2 A-B). Across counties, 18% have an estimated yield sensitivity to extreme heat (i.e., increases in temperature above 28°C) that is more than twice the average sensitivity, and 9% have an estimated sensitivity that is less than half the average sensitivity. This lack of symmetry can be understood by the fact that the estimated average sensitivity is a variance-weighted average of the local sensitivity (Goodman-Bacon, 2021) and that more sensitive counties tend to have lower variance in extreme heat. The local vulnerability to extreme heat differs significantly from the homogeneous effect for 87% of counties for maize, and 77% for soybean (Fig. 2 and S3, significance evaluated as $p<0.05$).

Consistent with previous results, irrigated farmlands – such as those in Mississippi,

Louisiana, and Arkansas irrigated by the Mississippi river, and in Nebraska irrigated by the Ogallala aquifer – are estimated to be less vulnerable to extreme heat ($p<0.01$, Fig. S8), despite the fact that irrigation was not directly specified as a determinant of vulnerability in the satellite-based model. Farmlands with higher average temperatures are also, on average, estimated to be less sensitive to extreme heat ($p<0.01$, Fig. S8). Importantly, there are also substantial differences in climate vulnerability predicted by the satellite model that are not explained by either of these previously studied determinants of agricultural climate vulnerability; average temperature and irrigation together explain only 21% of the variation in county-level climate vulnerability predicted by the satellite-based model for maize, and 7% for soybeans. And maps of local vulnerability estimated using satellite imagery differ substantially from maps of estimated sensitivity using average temperature, irrigation, or spatial features (Figs. S5, S6). These findings are robust to using alternative functional forms for the temperature response (Fig. S9). The improved predictive performance and additional identification of local vulnerability compared with previous approaches demonstrates the value of using satellite imagery to predict local agricultural climate vulnerability.

The above results show model performance when splitting the data randomly into training, validation, and test sets. Evaluation using random splits is useful when one is interested in understanding historical climate vulnerability within the same counties and years on which the model is trained. A potential concern when evaluating performance using random splits, however, is that correlation between observations in the training and tests sets could allow even non-informative features to improve performance by fitting to noise. To test for this possibility we conduct a placebo test where we randomly shuffle the satellite image features, and find substantially reduced performance very similar to that of the homogeneous effect (Table S1 model 5s). This indicates that, rather than fitting to noise, the satellite image features contain substantial visual information about climate vulnerability that improves model performance.

To evaluate the ability of the model to perform the more challenging tasks of extrapolating over space or time, we hold out observations in groups by county, state, or years. Such extrapolation reduces the predictive performance of all models, and tends to reduce the predictive performance of imagery-based model more than that of simpler models, due to their greater complexity. In all extrapolation exercises, however, performance of the satellite-based model remains higher than the homogeneous model. Qualitative results are the same when extrapolating to randomly held out observations and to observations held-out by county. When extrapolating to held out states, models based on irrigation and average temperature have higher performance than the satellite-based model. When extrapolating to held-out 2-year time periods, satellite-based models and location-based models perform

equally well for maize; for soy, irrigation-based models perform best in this setting. Estimates of local climate vulnerability using these different cross-validation approaches are similar (Fig. S10, S11). With few exceptions across all evaluation approaches, models using the combination of satellite imagery, location, and irrigation perform the best.

U.S. Human Mortality We estimate the local vulnerability of human mortality to temperature shocks across the U.S. using state-by-month observations of mortality rates and temperature from 1930 to 2015, building on the approach of Young and Hsiang (2024) (Eqn. 11). Allowing the influence of temperature on U.S. state-level mortality rates to differ based on imagery increases the amount of month-to-month variability in mortality explained by 26%, relative to the homogeneous model (R^2 increases from 0.034 to 0.043, $p<0.01$, Table S2). Predicting local vulnerability using imagery performs better than using average temperature, median income or spatial features. And these findings hold when observations are held-out randomly, spatially or temporally during cross-validation. A placebo test where the satellite image features are randomly shuffled shows substantially reduced performance (Table S2), again indicating that the satellite image features contain substantial information about climate vulnerability that improves model performance. Interestingly, the model using the shuffled image features performs close to as well as the model using spatial features, indicating that the shuffled satellite image features are able to act as a mathematical indicator for the state of interest. Accordingly, performance using the shuffled satellite features drops substantially when the model is evaluated on data from states not used for model training, rather than on randomly held out observations.

Estimated response curves and maps of mortality sensitivity to cold and heat indicate that the increase in performance when using the imagery-based model is driven by substantial local climate vulnerability and resilience predicted by the image features (Fig. 2C) that is not fully recovered using average temperature, median income or spatial features (Fig. S12). Consistent with prior findings (Heutel, Miller and Molitor, 2021; Carleton et al., 2022), mortality has a U-shape response to monthly average temperature, with hotter places generally less sensitive to heat waves but more sensitive to cold spells, and colder places generally less sensitive to cold spells but more sensitive to extreme heat. The temperature leading to the lowest mortality rates is substantially higher in the South than the North, potentially indicating adaptive responses in these areas to their hotter climatology (Fig. S12). Importantly, there is also substantial variation in local vulnerability estimated using the satellite imagery that improves model performance, but is not explained by temperature or income, which together explain only 69% of the variation in predicted local sensitivity to cold spells (i.e., decreasing monthly temperature under 20°C) and 41% to heat waves

(i.e., increasing monthly temperature over 20°C). For example, New Mexico, Colorado and Wyoming are estimated to be relatively resilient to extreme heat using satellite imagery, whereas estimates of vulnerability using average temperature indicate the opposite, and have lower predictive performance. Across states, 41% have an estimated sensitivity to heat waves that is more than 50% greater than the average sensitivity, and one quarter have an estimated sensitivity that is less than half the average sensitivity to heat. With respect to cold spells, 14% of states have an estimated sensitivity that is over 50% larger than the average, and the same fraction have an estimated sensitivity that is less than half the average. Collectively, these results indicate that human mortality vulnerability to changes in temperature differ substantially across U.S. states, and that satellite imagery can be used to estimate this vulnerability at a local level, with performance exceeding that of commonly used approaches.

U.S. Flooding We estimate how floods of differing depths damage structures using millions of claims from the National Flood Insurance Program in the United States from 1970 to 2020, building on the approach in Wing et al. (2020) (Eqn. 13). Allowing the influence of flood depth to differ based on imagery increases the amount of temporal variability in building damage explained by 31%, relative to the homogeneous model (Table S3, R^2 increases from 0.075 to 0.098, $p<0.01$). Like agriculture and mortality, satellite imagery reveals substantial differences in local vulnerability, with 8% of counties having 1.5 times the average damages from an increased foot of flooding, and 30% of counties having half the average damages.

In the case of building damages from flooding, while satellite imagery improves performance relative to the homogeneous model, other variables are stronger predictors of local vulnerability than satellite imagery (Table S3, Fig. S13). For randomly held-out evaluation sets, county-level spatial features deliver the highest performance of all models ($R^2 = 0.099$). And for evaluation sets held-out by county , state, or years, models based on historical average flood depth exposure outperform satellite and other features. The difficulty of satellite models to predict and extrapolate local vulnerability in this setting could be because many of the determinants of vulnerability, such the building materials or the presence of below-ground living spaces, are not directly visible in the imagery, and so the satellite-based model must instead use visible proxies. Placebo tests using shuffled satellite image features lose little performance relative to the unshuffled features. In combination with the high performance of the spatial features, this indicates that the satellite image features may be largely capturing spatial information, which indirectly determines vulnerability, rather than directly capturing the determinants of vulnerability themselves. Though imagery is not the

strongest predictor of local vulnerability to flooding it does still complement models using other features – models using both historical flood exposure and satellite features have the best performance out of all models for both spatial and temporal extrapolation. Collectively, these results indicate that there are substantial differences across space in local vulnerability to flooding, and that this vulnerability can be partially estimated using satellite imagery, but that other variables are stronger predictors and can be used in combination with satellite imagery for the best performance.

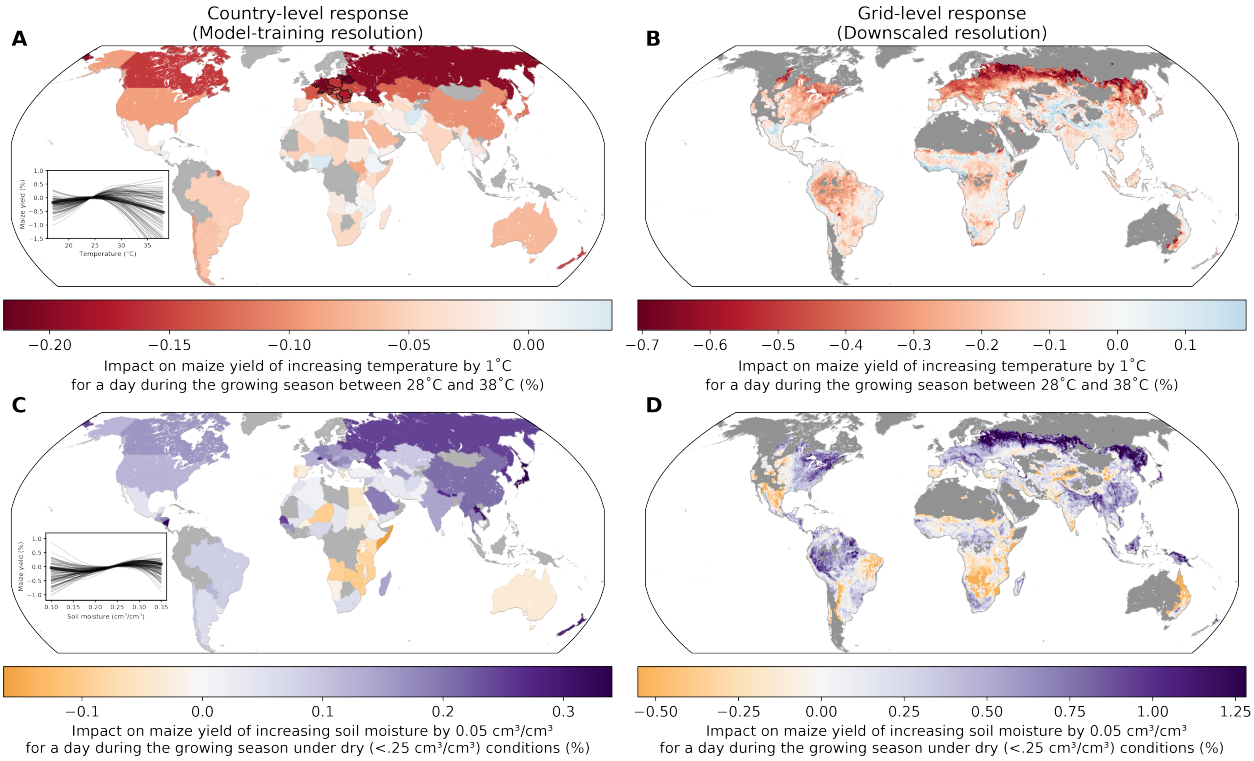


Figure 3: **Local sensitivity of maize yields to temperature and soil moisture globally predicted by satellite imagery.** Panels A and B show estimated maize yield sensitivity to extreme heat using satellite imagery at the national and subnational resolution. Panels C and D show estimated maize yield sensitivity to dry soils at national and subnational resolution.

Global Agriculture A strength of using satellite imagery to predict climate vulnerability is that it can be used anywhere, even where traditional sources of data on the determinants of vulnerability are inaccurate or unavailable. We estimate the local vulnerability of maize and soybean yields globally to temperature and soil moisture shocks using country-by-year observations of crop yields from 2007-2018, building on the analysis in Proctor et al. (2022) (Eqn. 9). We find that allowing the effects of temperature and soil moisture on agricultural productivity to differ across space according to satellite imagery increases the amount of

interannual variability in yield explained by 66% for maize and 85% for soybeans (Table S4, $p<0.01$ for both). The R^2 predicting yield anomalies with temperature and soil moisture increases from 0.09 to 0.15 for maize and 0.11 to 0.19 for soybeans.

For both crops, predicting local climate vulnerability using satellite imagery outperforms the use of spatial features, irrigation, or climatological temperature and soil moisture when evaluating on randomly held-out test sets. When extrapolating to new countries, the performance of the imagery-based model remains highest for both crops. When extrapolating to new years of data, the satellite-based model remains the best performing for maize, but for soybeans, modeling vulnerability using climatological temperature and soil moisture has the highest performance. Placebo tests using shuffled image features have substantially lowered performance for all crops and evaluation settings with performance generally similar to the homogeneous response. This indicates that the satellite image features are contributing substantially to model performance. The one exception is temporal extrapolation for soybeans where the shuffled model has a relatively high, but also highly uncertain, performance that is not statistically distinguishable from the unshuffled performance. The ability of satellite imagery to complement directly specified determinants of vulnerability is again indicated by the increased performance of using imagery and irrigation together, compared with each individually. In many but not all evaluation settings this is the best performing model.

The increase in predictive performance when using the satellite image features is driven by substantial differences in local vulnerability to temperature and soil moisture shocks across countries (Fig 3 A,C). For maize, 35% of countries have a sensitivity to extreme heat that is greater than twice the population average and 32% have one that is less than half the homogeneous response; 66% have a sensitivity to drought that is twice the population average and 25% have a sensitivity that is half the population average; and 73% show an impact of sodden conditions that is negative, whereas the population average response is positive. The impacts of heat, drought and sodden conditions on soybean yields similarly differ strongly across global agricultural lands (Fig. S16, S17).

We find a latitudinal gradient in vulnerability to extreme heat for both crops (Fig 3A, S14, S16) that is driven in part by hotter areas having lower sensitivity to extreme heat ($p< 0.01$ for both crops), consistent with prior estimates (Hultgren et al., 2025). We also find greater sensitivity to oversupply of water in areas equipped for irrigation than not ($p< 0.1$ for both crops), again consistent with prior work (Proctor et al., 2022). Importantly, however, the satellite-recovered local vulnerability cannot be fully explained by these commonly used determinants of local vulnerability (Fig. S14, S15, S16, S17). For maize, 47% of the variability in local vulnerability to extreme heat, 19% of local vulnerability to drought, and 52% of local vulnerability to sodden conditions can be explained by average temperature, average

soil moisture, and irrigation; the remaining estimated differences in vulnerability may be explained by differences in agricultural practices, infrastructure, or environmental services across farmlands that are visible in the imagery. For soy, average temperature, average soil moisture, and irrigation can explain 37%, 47%, and 38% of the local vulnerability to extreme heat, drought, and sodden conditions measured by the imagery. Collectively, these results demonstrate that satellite imagery can be used to predict local climate vulnerability at global scale. Given that poor regions tend to be both data scarce and particularly vulnerable to climate change, it is encouraging that satellite imagery delivers the largest increase in performance for predictions of climate vulnerability when applied outside of the just the U.S. at global scale. This result is also intuitive, given the diversity of agricultural practices across global croplands that could lead to different levels of climate vulnerability.

Predictions of vulnerability at fine resolution Many use cases would benefit from predictions of vulnerability at finer resolution than the social outcome data used for training. Here, we build from prior theoretical and empirical demonstrations of the ability of the MOSAIKS approach to make predictions at finer resolution than the training data (Rolf et al., 2021; Sherman et al., 2023) to make predictions of climate vulnerability at sub-national resolution globally. The MOSAIKS approach can make predictions at various spatial resolutions because it captures information about small sub-regions of the imagery and then averages that information to the image scale. This, combined with the linearity of the regression model used to relate these image features to climate vulnerability, means that the coefficients learned at the image scale can also theoretically be used to predict vulnerability at the sub-image scale. We evaluate the ability of the MOSAIKS approach to make predictions of climate vulnerability at finer resolution than the training data both empirically (Fig. S18) and in simulation (Fig. S19) and find that it is able to capture substantial information on vulnerability, though not all, within the spatial units on which it was trained.

We use our trained model to make predictions of maize yield vulnerability across global cropped areas at $0.25^\circ \times 0.25^\circ$ resolution (Fig. 3). In contrast with the aggregation of these estimates at the national level, the fine-resolution local predictions of climate vulnerability to extreme heat, drought, and flooding vary more smoothly across space. These results indicate that while country-level vulnerability assessments improve substantially upon global average estimates, there is also substantial variation in vulnerability within countries. Repeating this analysis with sub-national yield data, rather than national yield data, could improve upon these estimates in future work.

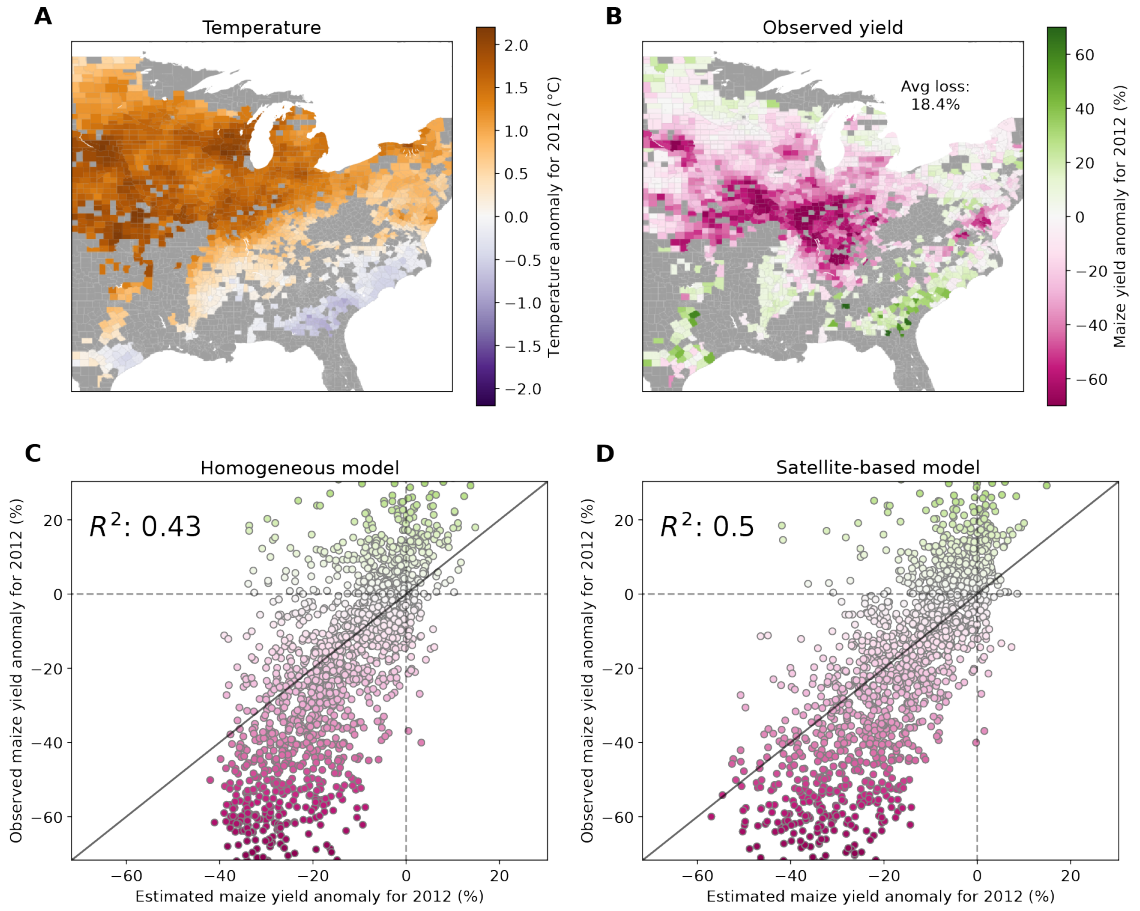


Figure 4: **Improved prediction of yield losses from the 2012 North American heatwave using satellite-based predictions of local vulnerability.** A) The maize growing season temperature anomaly. B) The observed yield anomaly. C) Observed and predicted yield anomalies; each point represents values for a county and year. Model results come from the 2 Year OOS model (Table S1), where observations from 2012 and 2013 are not included in model training.

Example applications Improved predictions of climate vulnerability could be used to improve management of climate hazards, inform investment in adaptation to climate change, and design climate policy. One specific way to improve hazard management is through better prediction of damages, such as those of the 2012 North American heat-wave. This heat-wave increased average growing season temperatures by over 1°C and decreased observed maize yields by roughly a fifth on average across counties east of the 100th meridian (Fig. 4 A-B). We find that predictions of yield losses in 2012 that incorporate satellite imagery to predict local temperature vulnerability perform 16% better than models assuming a homogeneous effect (Fig. 4 C-D, R^2 increases from 0.43 to 0.50; model not trained on data from 2012). Importantly, satellite-based predictions of local vulnerability improve predicted

2012 crop losses by improving estimates for those who incurred the largest yield losses; the maximum predicted county-level yield loss is 42% under the homogeneous model, 58% under the satellite-based model, and 62% in observations.

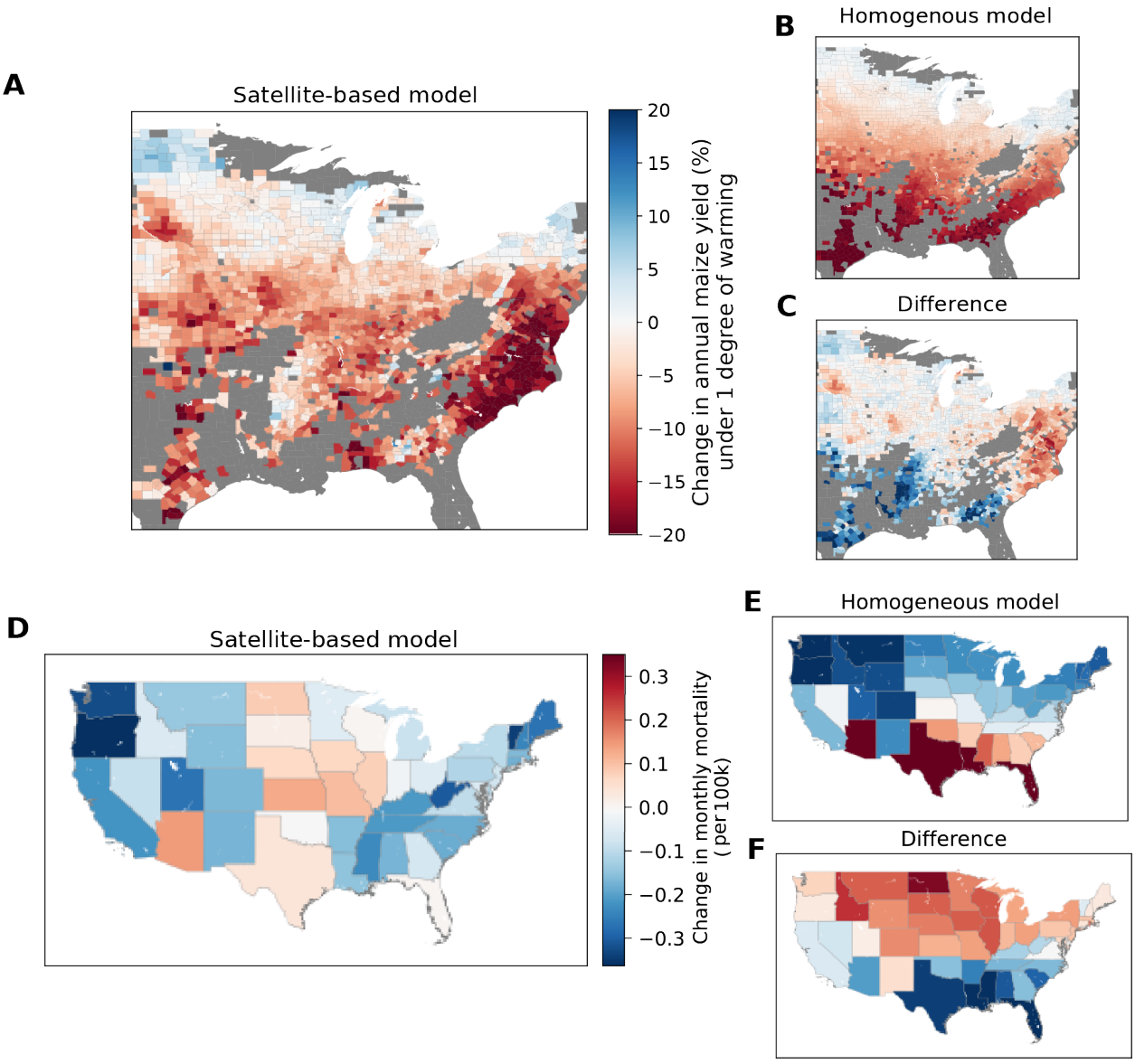


Figure 5: **Imagery-based predictions of local vulnerability indicate different projected effects of 1 °C of warming on U.S. annual maize yields and mortality rates.** Projected effects of 1°C warming using the satellite-based model (A,D), the homogeneous model (B,E), and their difference (C,F) for maize yields and human mortality rates.

Another application of local climate vulnerability estimates is to project how social outcomes might be affected by climate change. Projections of yield impacts from 1 °C of warming on U.S. maize yields substantially differ when using the traditional homogeneous model compared with the imagery-based model (Figure 5A). Projected damages are overestimated

using the traditional model, compared with the satellite-based model, in the South, particularly in irrigated areas around the Mississippi River, and underestimated in the Eastern U.S. from Virginia to Georgia. Our results thus suggest a substantial spatial re-distribution of projected climate damages to U.S. agriculture. While average projected yields are relatively similar using the homogeneous and satellite-based models for 1°C of warming, at higher levels of warming the satellite-based model projects relatively larger yield losses due to its greater estimated climate vulnerability in historically cooler locations (Figure S20).

Projected impacts of 1 °C of warming on U.S. mortality rates also substantially differ when using the homogeneous model compared with the imagery-based model. Instead of substantial reductions in mortality rates in the North and increases in the South, the imagery-based model generally projects relatively moderated effects of climate change. The satellite-based model suggests that the central U.S. will have increased mortality rates with warming, but that the coasts will see reductions in mortality.

Discussion and Conclusion

Across the U.S. and globally, fields, populations, and buildings differ substantially in their vulnerability to climate hazards. Satellite imagery offers a generalizable and universally available source of information to predict this vulnerability. Models of climate vulnerability using satellite imagery generally outperform traditional approaches using hand-selected covariates or spatial features, and their combination offers the best estimates. Improved predictions of local climate vulnerability are useful for better estimating damages from extreme weather events and climate change, especially for the most vulnerable populations. The additional information provided by satellite imagery is likely to be of most use in traditionally data poor regions, where information on the determinants of climate vulnerability from traditional sources are lacking. The approach to estimate local vulnerability using satellite imagery developed here is compatible with existing approaches for estimating climate vulnerability and can likely be applied to a broad range of panel fixed effects regression settings where the determinants of effect heterogeneity are visible in satellite imagery. To estimate satellite-based local effects, a user would need to download publicly available MOSAIKS features from `mosaiks.org`, spatially merge them with with their panel dataset, add an interaction term between the satellite features and the treatment effect of interest, and re-run their analysis using ridge regression rather than ordinary least squares regression. The structure of the approach builds upon existing panel fixed effect regression models, familiar to many researchers in economics and the natural sciences, and the approach can be implemented without downloading or processing any satellite imagery.

In addition to obtaining skillful estimates of local climate vulnerability, it would be useful to understand how the model is making these estimates and what specific variables are responsible. One interpretation of how the model works is that it effectively groups together locations that look similar in the imagery, and then estimates separate effects in each of these groups. Consistent with this interpretation, grouping U.S. counties by their satellite-estimated vulnerability of maize yields to extreme heat and then re-estimating vulnerability within each group using traditional ordinary least squares regression – with no penalization – gives similar estimates (Pearson's ρ^2=0.94, Figure S21). These groups share visual similarities. For example, counties around the Mississippi River in Arkansas, Tennessee, Mississippi and Louisiana share a characteristic look of rivers winding through farmlands and are collectively less sensitive to extreme heat (Fig. S22). In contrast, Northeastern Mississippi appears more mountainous and woody and, like much of visually similar Appalachia, has much higher sensitivity to extreme heat. Southwestern Georgia and southeastern Alabama are also particularly resistant to extreme heat and share visibly distinctive central pivot irrigation systems that are fed by the Floridian Aquifer (Fig. S23). To further understand the drivers of climate vulnerability represented in the imagery we can visually inspect fine-resolution predictions of vulnerability made using the same approach as in Fig. 3 B,D. In Morehouse County, Lousiana we can see that, within the county, the areas surrounding the winding rivers and streams with low forest cover – characteristic of the land bordering the Mississippi River – are predicted to have low maize-yield sensitivity to extreme heat, whereas non-riverine areas with thicker background vegetation are predicted to be more vulnerable (Fig. S24). While it is not possible to fully interpret what variables are driving predicted vulnerability, these approaches provide a rich entry-point for understanding determinants of vulnerability beyond what are able to be measured using traditional data sources.

There are several important limitations of our approach. First, while our approach estimates plausibly causal effects of climate variables on social outcomes, we treat the task of predicting heterogeneity in climate vulnerability across space as predictive, rather than causal. This is consistent with how heterogeneity in climate vulnerability with respect to variables such as irrigation, income, or average temperature has been estimated in prior work (e.g., Proctor et al. (2022); Hultgren et al. (2025); Carleton et al. (2022)). Our approach, however, is limited to a greater degree than this prior work with respect to causality because, while there are experimental setups where, e.g., irrigation is randomly assigned (e.g., Tack, Barkley and Hendricks (2017)), it would be challenging to find or interpret a setting where satellite imagery – and all the determinants of climate vulnerability it measures, is as good as randomly assigned.

Third, our approach will be more effective for some social outcomes than others. Results

from simulation experiments indicate that our approach can estimate local vulnerability to roughly the degree to which the determinants of local vulnerability are visible in the satellite imagery (Supplementary Information A). Information on what can be measured with satellite imagery (e.g., Proctor et al. (2025)) can help inform whether this approach may be appropriate or not for certain settings, and why. Ultimately, the most direct way to understand whether satellite imagery can predict additional variation in local vulnerability for a given setting is to try it out, which is facilitated by it's relatively low cost of implementation.

Fourth, our approach uses information from only high-resolution imagery with red, green and blue bands. It's possible that use of additional spectral bands or different ways of extracting information from the imagery could improve upon the performance found here. To test this, we repeat our estimation of local vulnerability for U.S. maize and soybeans using the recently released AlphaEarth Foundations (AEF) features, which were created using optical, thermal and C-band synthetic aperture radar information from Sentinel 1 and 2 and Landsat 8 and 9 satellite imagery (Brown et al., 2025). We find that the AEF feature give generally similar performance to that of MOSAIKS features, often with slightly higher performance but sometimes lower performance. This indicates that the high resolution optical MOSAIKS features we use may deliver a majority of the benefit that satellite imagery of differing resolutions and spectra can provide in this setting. It also highlights the flexibility of our approach to use information from different satellite sources and embedding methodologies.

Lastly, our approach predicts differences in climate vulnerability only across space, predicting differences in vulnerability across time using satellite imagery is an important area for future work. We note, however, that recent estimates suggest that changes in vulnerability over time may be relatively small (Burke et al., 2024), whereas we find large differences in vulnerability across space. Our hope is that better measurements of local vulnerability can facilitate both improved understanding of its determinants and more efficient investments in increased climate change resilience and disaster response.

References

Auffhammer, Maximilian. 2022. "Climate Adaptive Response Estimation: Short and long run impacts of climate change on residential electricity and natural gas consumption." *Journal of Environmental Economics and Management* 114:102669.

Barreca, Alan, Karen Clay, Olivier Deschenes, Michael Greenstone and Joseph S. Shapiro. 2016. "Adapting to Climate Change: The Remarkable Decline in the US Temperature-Mortality Relationship over the Twentieth Century." *Journal of Political Economy* 124(1).

Baylis, Patrick. 2020. "Temperature and temperament: Evidence from Twitter." *Journal of Public Economics* 184:104161.

Benveniste, Hélène, Peter Huybers and Jonathan Proctor. 2025. "Global climate migration is a story of who and not just how many." *Nature Communications* 16(1):7752.

Brown, Christopher F, Michal R Kazmierski, Valerie J Pasquarella, William J Rucklidge, Masha Samsikova, Chenhui Zhang, Evan Shelhamer, Estefania Lahera, Olivia Wiles, Simon Ilyushchenko et al. 2025. "Alphaearth foundations: An embedding field model for accurate and efficient global mapping from sparse label data." *arXiv preprint arXiv:2507.22291* .

Burke, Marshall, Anne Driscoll, David B Lobell and Stefano Ermon. 2021. "Using satellite imagery to understand and promote sustainable development." *Science* 371(6535):eabe8628.

Burke, Marshall, Mustafa Zahid, Mariana CM Martins, Christopher W Callahan, Richard Lee, Tumenkhusel Avirmed, Sam Heft-Neal, Mathew Kiang, Solomon M Hsiang and David Lobell. 2024. Are We Adapting to Climate Change? Technical report National Bureau of Economic Research.

Burke, Marshall, Solomon M Hsiang and Edward Miguel. 2015. "Global non-linear effect of temperature on economic production." *Nature* 527(7577):235–239.

Butler, Ethan E and Peter Huybers. 2013. "Adaptation of US maize to temperature variations." *Nature Climate Change* 3(1):68–72.

Carleton, Chong, Druckenmiller, Noda, Proctor, Rolf and Hsiang. 2022. "Multi-Task Observation Using Satellite Imagery and Kitchen Sinks (MOSAIKS) API." `https://mosaiks.org`.

Carleton, Tamma A. 2017. "Crop-damaging temperatures increase suicide rates in India." *Proceedings of the National Academy of Sciences* 114(33):8746–8751.

Carleton, Tamma A and Solomon M Hsiang. 2016. "Social and economic impacts of climate." *Science* 353(6304):aad9837.

Carleton, Tamma, Amir Jina, Michael Delgado, Michael Greenstone, Trevor Houser, Solomon Hsiang, Andrew Hultgren, Robert E Kopp, Kelly E McCusker, Ishan Nath et al. 2022. "Valuing the global mortality consequences of climate change accounting for adaptation costs and benefits." *The Quarterly Journal of Economics* 137(4):2037–2105.

Carleton, Tamma, Esther Duflo, B Kelsey Jack and Guglielmo Zappalà. 2024. Adaptation to climate change. In *Handbook of the Economics of Climate Change*. Vol. 1 Elsevier pp. 143–248.

Carleton, Tamma and Michael Greenstone. 2021. "Updating the United States Government's Social Cost of Carbon." *University of Chicago, Becker Friedman Institute for Economics Working Paper* (2021-04).

Center for International Earth Science Information Network (CIESIN) Columbia, University. 2018. "Gridded Population of the World, Version 4 (GPWv4): Population Density, Revision 10.".

CPC Global Daily Temperature. N.d. `https://psl.noaa.gov/data/gridded/data.cpc.globaltemp.html`.
URL: *https://psl.noaa.gov/data/gridded/data.cpc.globaltemp.html*

Daly, Christopher, GH Taylor and WP Gibson. 1997. The PRISM approach to mapping precipitation and temperature. In *Proc., 10th AMS Conf. on Applied Climatology*. pp. 20–23.

Dorie, Vincent, Jennifer Hill, Uri Shalit, Marc Scott and Dan Cervone. 2018. "Automated versus do-it-yourself methods for causal inference: Lessons learned from a data analysis competition.".
URL: *https://arxiv.org/abs/1707.02641*

Dorigo, Wouter, Wolfgang Wagner, Clement Albergel, Franziska Albrecht, Gianpaolo Balsamo, Luca Brocca, Daniel Chung, Martin Ertl, Matthias Forkel, Alexander Gruber, Eva Haas, Paul D. Hamer, Martin Hirschi, Jaakko Ikonen, Richard de Jeu, Richard Kidd, William Lahoz, Yi Y. Liu, Diego Miralles, Thomas Mistelbauer, Nadine Nicolai-Shaw,

Robert Parinussa, Chiara Pratola, Christoph Reimer, Robin van der Schalie, Sonia I. Seneviratne, Tuomo Smolander and Pascal Lecomte. 2017. "ESA CCI Soil Moisture for improved Earth system understanding: State-of-the art and future directions." *Remote Sensing of Environment* 203:185–215.

EarthStat. N.d.
URL: *http://www.earthstat.org/*

Efron, Bradley and Robert J Tibshirani. 1994. *An introduction to the bootstrap*. Chapman and Hall/CRC.

FAOSTAT Crops and Livestock Products. N.d. `https://www.fao.org/faostat`.
URL: *https://www.fao.org/faostat*

Giesselmann, Marco and Alexander W. Schmidt-Catran. 2022. "Interactions in Fixed Effects Regression Models." *Sociological Methods & Research* 51(3):1100–1127.
URL: *https://doi.org/10.1177/0049124120914934*

Goodman-Bacon, Andrew. 2021. "Difference-in-differences with variation in treatment timing." *Journal of econometrics* 225(2):254–277.

Gruber, Alexander, Tracy Scanlon, Robin Van Der Schalie, Wolfgang Wagner and Wouter Dorigo. 2019. "Evolution of the ESA CCI Soil Moisture climate data records and their underlying merging methodology." *Earth System Science Data* 11(2):717–739.

Hastie, Trevor, Robert Tibshirani and Jerome H Friedman. 2009. *The elements of statistical learning: data mining, inference, and prediction*. Vol. 2 New York: .

Head, Andrew, Mélanie Manguin, Nhat Tran and Joshua E Blumenstock. 2017. "Can human development be measured with satellite imagery?" *Ictd* 17:16–19.

Heutel, Garth, Nolan H Miller and David Molitor. 2021. "Adaptation and the mortality effects of temperature across US climate regions." *Review of Economics and Statistics* 103(4):740–753.

Hsiang, Solomon. 2016. "Climate econometrics." *Annual Review of Resource Economics* 8:43–75.

Hsiang, Solomon M, Kyle C Meng and Mark A Cane. 2011. "Civil conflicts are associated with the global climate." *Nature* 476(7361):438–441.

Hultgren, Andrew, Tamma Carleton, Michael Delgado, Diana R Gergel, Michael Greenstone, Trevor Houser, Solomon Hsiang, Amir Jina, Robert E Kopp, Steven B Malevich et al. 2025. "Impacts of climate change on global agriculture accounting for adaptation." *Nature* 642(8068):644–652.

Jean, Neal, Marshall Burke, Michael Xie, W Matthew Davis, David B Lobell and Stefano Ermon. 2016. "Combining satellite imagery and machine learning to predict poverty." *Science* 353(6301):790–794.

Jerzak, Connor T, Fredrik Johansson and Adel Daoud. 2023. "Image-based Treatment Effect Heterogeneity." *Proceedings of Machine Learning Research vol* 213:1–22.

Khachiyan, Arman, Anthony Thomas, Huye Zhou, Gordon Hanson, Alex Cloninger, Tajana Rosing and Amit K Khandelwal. 2022. "Using Neural Networks to Predict Microspatial Economic Growth." *American Economic Review: Insights* 4(4):491–506.

Künzel, Sören R, Jasjeet S Sekhon, Peter J Bickel and Bin Yu. 2019. "Metalearners for estimating heterogeneous treatment effects using machine learning." *Proceedings of the national academy of sciences* 116(10):4156–4165.

Missirian, Anouch and Wolfram Schlenker. 2017. "Asylum applications respond to temperature fluctuations." *Science* 358(6370):1610–1614.

Monfreda, Chad, Navin Ramankutty and Jonathan A. Foley. 2008. "Farming the planet: 2. Geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000." *Global Biogeochemical Cycles* 22(1):1–19.

National Agricultural Statistcis Service - Quick Stats. N.d.
URL: *https://quickstats.nass.usda.gov/*

of Agriculture National Agriculture Statistics Service, US Department. 2010. "Field Crops Usual Planting and Harvesting Dates (October 2010).".

Opper, Isaac M, R Jisung Park and Lucas Husted. 2023. "The effect of natural disasters on human capital in the United States." *Nature human behaviour* 7(9):1442–1453.

Park, R Jisung, Joshua Goodman, Michael Hurwitz and Jonathan Smith. 2020. "Heat and learning." *American Economic Journal: Economic Policy* 12(2):306–39.

Planet Team. 2017. "Planet Application Program Interface: In Space for Life on Earth.".
URL: *https://api.planet.com*

Proctor, Jonathan. 2021. "Atmospheric opacity has a nonlinear effect on global crop yields." *Nature Food* 2(3):166–173.

Proctor, Jonathan, Angela Rigden, Duo Chan and Peter Huybers. 2022. "More accurate specification of water supply shows its importance for global crop production." *Nature Food* 3(9):753–763.

Proctor, Jonathan, Solomon Hsiang, Jennifer Burney, Marshall Burke and Wolfram Schlenker. 2018. "Estimating global agricultural effects of geoengineering using volcanic eruptions." *Nature* 560(7719):480–483.

Proctor, Jonathan, Tamma Carleton, Trinetta Chong, Taryn Fransen, Simon Greenhill, Jessica Katz, Hikari Murayama, Luke Sherman, Jeanette Tseng, Hannah Druckenmiller and Solomon Hsiang. 2025. What Can Satellite Imagery and Machine Learning Measure? Technical report National Bureau of Economic Research.

Rising, James and Naresh Devineni. 2020. "Crop switching reduces agricultural losses from climate change in the United States by half under RCP 8.5." *Nature communications* 11(1):4991.

Rode, Ashwin, Tamma Carleton, Michael Delgado, Michael Greenstone, Trevor Houser, Solomon Hsiang, Andrew Hultgren, Amir Jina, Robert E Kopp, Kelly E McCusker et al. 2021. "Estimating a social cost of carbon for global energy consumption." *Nature* 598(7880):308–314.

Rolf, Esther, Jonathan Proctor, Tamma Carleton, Ian Bolliger, Vaishaal Shankar, Miyabi Ishihara, Benjamin Recht and Solomon Hsiang. 2021. "A generalizable and accessible approach to machine learning with global satellite imagery." *Nature communications* 12(1):4392.

Sacks, William J., Delphine Deryng, Jonathan A. Foley and Navin Ramankutty. 2010. "Crop planting dates: An analysis of global patterns." *Global Ecology and Biogeography* 19(5):607–620.

Schlenker, Wolfram and Michael J Roberts. 2009. "Nonlinear temperature effects indicate severe damages to US crop yields under climate change." *Proceedings of the National Academy of sciences* 106(37):15594–15598.

Sherman, Luke, Jonathan Proctor, Hannah Druckenmiller, Heriberto Tapia and Solomon M Hsiang. 2023. Global High-Resolution Estimates of the United Nations Human Devel-

opment Index Using Satellite Imagery and Machine-learning. Technical report National Bureau of Economic Research.

Stefan, Siebert, Henrich Verena, Frenken Karen and Jacob Burke. 2013. "Global Map of Irrigation Areas version 5." *Food and Agriculture , AQUASTAT* .

Tack, Jesse, Andrew Barkley and Nathan Hendricks. 2017. "Irrigation offsets wheat yield reductions from warming temperatures." *Environmental Research Letters* 12(11):114027.

Wager, Stefan and Susan Athey. 2018. "Estimation and inference of heterogeneous treatment effects using random forests." *Journal of the American Statistical Association* 113(523):1228–1242.

Wing, Oliver EJ, Nicholas Pinter, Paul D Bates and Carolyn Kousky. 2020. "New insights into US flood vulnerability revealed from flood insurance big data." *Nature communications* 11(1):1444.

Wooldridge, Jeffrey M. 2010. *Econometric analysis of cross section and panel data*. MIT press.

Young, Rachel and Solomon Hsiang. 2024. "Mortality caused by tropical cyclones in the United States." *Nature* pp. 1–8.

Zaveri, Esha and David B. Lobell. 2019. "The role of irrigation in changing wheat yields and heat sensitivity in India." *Nature communications* 10(1):4144.

Methods

1.1 Data and model specifications

1.1.1 Satellite image features

We use pre-computed MOSAIKS satellite image features produced from Planet's Surface Reflectance Basemaps product from 2019, which has a pixel resolution of 4.77m x 4.77m at the equator (Carleton, Chong, Druckenmiller, Noda, Proctor, Rolf and Hsiang, 2022). The features are produced using the random convolutional features method described in Rolf et al. (2021). Features are available on a global 0.01° x 0.01° equal-angle grid for all land tiles with available imagery, amounting to ≈ 151 million feature vectors in total. Features are often unavailable above 60° latitude, due to a lack of available imagery. We use 4000 task-agnostic daytime image features. Half of these features use a patch size of 4 x 4 x 3 pixels, and the other half use a patch size of 6 x 6 x 3 pixels. The third dimension of the patch size refers to the color bands (i.e., red, green, and blue) available in Planet imagery. Daytime quarterly image mosaics are processed by Planet to minimize cloud cover, balance color across seasons, and remove seams from images (Planet Team, 2017; Carleton, Chong, Druckenmiller, Noda, Proctor, Rolf and Hsiang, 2022). Data from quarter 3 is used to produce the features because it corresponds with less ice coverage in the northern hemisphere and less cloud cover in the tropics.

Aggregating satellite image features to administrative polygons To connect the satellite image features to the tabular data on crop yields, mortality rates, and flooding damage, we aggregate MOSAIKS features from their 0.01° x 0.01° source resolution to the administrative polygon level (i.e., county, state, country) by taking a weighted average of the feature values, following Sherman et al. (2023). Flood and mortality data are aggregated using population weights from the Gridded Population of the World (Center for International Earth Science Information Network , CIESIN). U.S. and global agricultural yield data are aggregated using cropped area weights from EarthStat (*EarthStat*, N.d.).

1.1.2 U.S. agriculture: model specification and data

Following Schlenker and Roberts (2009), we estimate the homogeneous nonlinear influence of temperature on crop yields by regressing the logarithm of yields, Y, in a given county, i, and year, t, on features representing the distribution of temperature and rainfall during the

growing season:

$$Y_{it} = \underbrace{\beta_1 GDD_{it} + \beta_2 KDD_{it}}_{\text{temperature response}} + \underbrace{\beta_3 P + \beta_4 P^2 + \alpha_i + \delta_s t}_{\text{controls}} + \epsilon_{it} \tag{5}$$

Effects are estimated separately for maize and soybeans. To estimate the nonlinear effect of daily temperature on yields we calculate the number of growing degree days (GDD) and killing degree days (KDD) across the growing season from the daily temperature data in each county over cropped areas, following e.g., Schlenker and Roberts (2009); Butler and Huybers (2013). We use June, July and August as the growing season; results are consistent using state-specific usual planting and harvesting dates (of Agriculture National Agriculture Statistics Service, 2010). We model precipitation, P, quadratically using as features the growing season average of linear and squared daily precipitation values. The model includes flexible spatio-temporal controls: county-specific fixed effects, α_i, and state-specific linear time trends $\delta_s \cdot t$. These account for any time-invariant differences between counties such as soil type, and for county-specific trending variables such as technological adoption. β_1 and β_2 are the coefficients of interest for this homogeneous model and form the response functions shown in black lines in Fig. 2 A-B.

To estimate local nonlinear response functions using satellite imagery, we add interaction terms to the model, following Equation 4:

$$\begin{aligned} Y_{it} = &\underbrace{\beta_1 GDD_{it} + \beta_2 KDD_{it}}_{\text{homogeneous response}} + \underbrace{\psi_{GDD}\mathbf{X_i} GDD_{it} + \psi_{KDD}\mathbf{X_i} KDD_{it}}_{\text{satellite-based heterogeneity}} \\ &+ \underbrace{\beta_3 P + \beta_4 P^2 + \alpha_i + \delta_{i \in state} \cdot t}_{\text{controls}} + \epsilon_{it} \end{aligned} \tag{6}$$

where $\mathbf{X}_i$ represents the satellite image features for county i. Note that ψ_{GDD} and ψ_{GDD} are 4000-dimensional vectors, not scalars, with a coefficient for each dimension of the image features ($\mathbf{X}$). β_1 and β_2 give the average response across all counties and $\psi_{GDD}\mathbf{X_i}$ and $\psi_{KDD}\mathbf{X_i}$ tell how the response for county i differs from that average response. Thus, the response curve for each county is the determined by β_1, β_2, $\psi_{GDD}\mathbf{X_i}$, and $\psi_{KDD}\mathbf{X_i}$.

To estimate local response functions using spatial features, $\mathbf{S_i}$, we similarly interact these with the temperature features:

$$Y_{it} = \underbrace{\beta_1 GDD_{it} + \beta_2 KDD_{it}}_{\text{homogeneous response}} + \underbrace{\kappa_{GDD}\mathbf{S_i} GDD_{it} + \kappa_{KDD}\mathbf{S_i} KDD_{it}}_{\text{spatial-based heterogeneity}} \tag{7}$$

$$+ \underbrace{\beta_3 P + \beta_4 P^2 + \alpha_i + \delta_{i \in state} \cdot t}_{\text{controls}} + \epsilon_{it} \tag{8}$$

where, like a county fixed effect, S_i is a set of indicators for whether county i is in a given state or county, depending on the set of spatial features used. To estimate heterogeneity in response with respect to irrigation or average temperature, these variables are similarly added as interaction terms with the temperature features, following, for example, Proctor et al. (2022).

Yield and irrigation data Yield data come from the United States Department of Agriculture National Agricultural Statistics Service (USDA NASS) (*National Agricultural Statistcis Service - Quick Stats*, N.d.). We use county-level annual yield observations from 2000 to 2020 in this analysis. We analyze counties with a centroid east of the 100th meridian to exclude the qualitatively distinct arid portions of the western United States, following e.g., Schlenker and Roberts (2009); Butler and Huybers (2013). The panel contains 29,405 county by year observations for maize from 1,684 counties and 26,723 observations for soybeans from 1,496 counties. Irrigation data are also from USDA NASS. We take irrigation data from 2012 because it is roughly the middle of our panel; we assume that irrigation does not change over time. The irrigation feature we use in the model is the fraction of total harvested acres that is equipped for irrigation; this feature is not crop-specific.

Temperature and precipitation data Surface (2m) air temperature data for the U.S. at 4km by 4km resolution are from the Parameter-elevation Regressions on Independent Slopes Model (PRISM) (Daly, Taylor and Gibson, 1997). Following Schlenker and Roberts (2009) we construct GDD and KDD variables using a daily sinusoidal temperature approximation from the observed daily minimum and maximum temperature using a kink point of 28°C. These GDDs and KDDs are summed over the length of the growing season, which we assume to be June 1 to Aug 30 for all US counties. Note that use of restricted cubic spline features, in place of GDD-KDD features, give similar results, indicating that the specific choice of kink point does not substantially influence our results (Fig. S9).

1.1.3 Global agriculture: model specification and data

Following Proctor et al. (2022), we estimate the influence of temperature, T, and soil moisture, S, on crop yields by regressing the logarithm of yields, Y, in a given country, i, and year, t, on features representing the distribution of temperature and soil moisture during the

growing season:

$$Y_{it} = \underbrace{\sum_{k=1}^{3} \beta_k^T R_k(T)_{it}}_{\text{temperature response}} + \underbrace{\sum_{k=1}^{3} \beta_k^S R_k(S)_{it}}_{\text{soil moisture response}} + \underbrace{\alpha_i + \delta_i t}_{\text{controls}} + \epsilon_{it} \tag{9}$$

The climate variable features are restricted cubic spline transformations of daily maximum surface air temperature ($R_k(T)_{it}$) and surface soil moisture ($R_k(S)_{it}$) over the growing season. Like GDD and KDD, which form a piecewise linear spline, restricted cubic splines allow for a nonlinear response of yield to daily temperatures; see Proctor et al. (2022) for details. We also include country fixed effects (α_i) linear time trends for each country $\delta_i t$. The homogeneous response function for temperature is determined by β_1^T, β_2^T, and β_3^T.

To estimate local response functions using satellite imagery, we add interaction terms to the model:

$$\begin{aligned} Y_{it} = & \underbrace{\sum_{k=1}^{3} \beta_k^T R_k(T)_{it}}_{\text{homogeneous temperature response}} + \underbrace{\sum_{k=1}^{3} \psi_k^T \mathbf{X_i} R_k(T)_{it}}_{\text{satellite-based heterogeneity}} + \\ & \underbrace{\sum_{k=1}^{3} \beta_k^S R_k(S)_{it}}_{\text{homogeneous soil moisture response}} + \underbrace{\sum_{k=1}^{3} \psi_k^S \mathbf{X_i} R_k(S)_{it}}_{\text{satellite-based heterogeneity}} + \\ & \underbrace{\alpha_i + \delta_i t + \epsilon_{it}\alpha_i + \delta_i t}_{\text{controls}} + \epsilon_{it} \end{aligned} \tag{10}$$

where $\mathbf{X_i}$ represents the satellite image features for country i, which is interacted with the features for temperature to allow the nonlinear response of temperature to differ across countries according to the imagery. Interactions with spatial features, irrigation, and climatology are specified similarly to those for U.S. agriculture.

Yield, temperature, soil moisture and irrigation data Data are obtained from Proctor et al. (2022) – see the data description therein for details. In brief, we use national level annual yield data for maize and soybeans from the Food and Agricultural Organization of the United Nations from 2007-2018 (*FAOSTAT Crops and Livestock Products*, N.d.). Daily temperature data comes from the Climate Prediction Center (*CPC Global Daily Temperature*, N.d.) and daily soil moisture data comes from the European Spaces Agency Climate Change Initiative Plus Soil Moisture Project v.06.1 dataset (Dorigo et al., 2017; Gruber et al., 2019). To match the national yield data to the distribution of daily temperature

and soil moisture during the growing season, we take the cropped-fraction-weighted mean of each variable's restricted cubic spline features over year 2000 cropped area (Monfreda, Ramankutty and Foley, 2008) and the growing season (Sacks et al., 2010). The sample contains 1,047 country-year observations for maize and 693 for soybeans. Irrigation is specified as the fraction of crop-specific cropped area in each country equipped for irrigation. Irrigation data is for the year 2005 from the Global Map of Irrigation Areas v.5 (Stefan et al., 2013).

1.1.4 US Mortality: model specification and data

Following Young and Hsiang (2024), but not including tropical cyclone effects, we estimate the influence of temperature, T, on the mortality rate (deaths per 100k individuals per month), Y, in a given state, i, and month-of-sample, t using the following specification:

$$Y_{it} = \underbrace{\sum_{k=1}^{4} \beta_k T_{it}^k}_{\text{temperature response}} + \underbrace{\mu_1 m_{it} + \sum_{n=1}^{8} (\eta_n s_i t^n) + \mu_2 m_{it} t + \mu_3 h_t}_{\text{controls}} + \epsilon_{it}. \tag{11}$$

The temperature response is modeled using a fourth-order polynomial in monthly average surface air temperature. We include the same spatiotemporal controls as Young and Hsiang (2024): m_{it} are state-by-month-of-year fixed effects, $\sum_{n=1}^{8}(s_i \cdot t^n)$ are state-specific eighth-order polynomial time trends, $m_{it} \cdot t$ are state-by-month-of-year fixed effects interacted with a linear trend in month-of-sample, and h_t are national month-of-sample fixed effects.

To estimate local temperature response functions using satellite imagery we add interaction terms to the regression model:

$$\begin{aligned} Y_{it} = & \underbrace{\sum_{k=1}^{4} \beta_k T_{it}^k}_{\text{homogeneous response}} + \underbrace{\sum_{k=1}^{4} \psi_k \mathbf{X}_i T_{it}^k}_{\text{satellite-based response}} + \\ & \underbrace{\mu_1 \cdot m_{it} + \sum_{n=1}^{8} (\eta_n \cdot s_i \cdot t^n) + \mu_2 \cdot m_{it} \cdot t + \mu_3 \cdot h_t}_{\text{controls}} + \epsilon_{it} \end{aligned} \tag{12}$$

where $\mathbf{X_i}$ represents the satellite image features for state i. Interactions with spatial features, income, and climatology are specified similarly to those for U.S. agriculture.

Mortality, temperature, and income data We use U.S. mortality and temperature data from Young and Hsiang (2024). Their dataset provides N = 37,044 observations of temperature and mortality at the state-year-month level from 1950-2012. Young and Hsiang

(2024) compile all-cause mortality data from Barreca et al. (2016) and the public Center for Disease Control and Prevention Underlying Cause of Death database. The authors spatially aggregate average monthly temperature data from Berkeley Earth Surface Temperatures (BEST) to the state-level; see Young and Hsiang (2024) for details. For estimating heterogeneous effects of temperature by income, we use median household income data by state from the 2020 American Community Survey.

1.1.5 US Flood Damages: model specification and data

We estimate the homogeneous influence of flood depth, D, on the fraction of property value damaged, Y as follows:

$$Y_{cit} = \underbrace{\beta_1 D_{it}}_{\text{flood depth response}} + \underbrace{\alpha_i + \delta_s t}_{\text{controls}} + \epsilon_{it} \tag{13}$$

where c indexes each NFIP claim, i indexes the county for each claim, t indexes the year of each claim's date of loss, and s indexes state. The outcome, Y, is relative damage, i.e., building damages from flooding divided by the property value. The county fixed effects, α_i, and state-specific linear time trends, $\delta_s t$, account for potential unobserved confounding factors.

To estimate local building vulnerability to flooding we add interaction terms to the model:

$$Y_{cit} = \underbrace{\beta_1 D_{it}}_{\text{homogeneous response}} + \underbrace{\psi \mathbf{X}_i D_{it}}_{\text{satellite-based heterogeneity}} + \underbrace{\alpha_i + \delta_s t}_{\text{controls}} + \epsilon_{it} \tag{14}$$

where $\mathbf{X}_i$ represents the satellite image features for county i.

Flood depth and building damage data We obtain National Flood Insurance Program (NFIP) claims data from the U.S. Federal Emergency Management Agency (FEMA) for the years 1970 through 2020. These include both the flood depth data and the damages data. We follow the data cleaning steps outlined in Wing et al. (2020), focusing on claims for one-story residential buildings without a basement and with flood depths between one and eight feet. Of these, our analysis includes only claims with corresponding county information. The full, cleaned dataset contains 433,679 claims from 1970 to 2020.

1.2 Model training and evaluation

1.2.1 Estimating causal effects using panel fixed effects regression models

Following the climate econometrics literature, the regression models used here aim to estimate causal effects of climate variables on social outcomes. They do so by isolating variation in the climate variable that is arguably as good as randomly assigned using a set of model "fixed effects." These spatiotemporal controls generally account for average differences across space and time in the social outcome and climate variable. After accounting for these controls, causal effects are estimated by analyzing anomalies in the climate variable with respect to anomalies in the social outcome variable. For a detailed discussion of these methods see Hsiang (2016).

1.2.2 Residualizing data to estimate causal effects

The estimation of these panel fixed effects regression models can be viewed in two steps (Proctor et al., 2022; Hultgren et al., 2025). In the first step, the social outcome variable and climate variable are "residualized" with respect to the model controls, generating anomalies (Y' and C'). This step sets the experimental design by isolating variation that is plausibly as good as randomly assigned enabling estimated effects to be interpreted as causal. This step is implemented by regressing both the outcome variable and the climate variable on the controls and then computing the the residuals. In the second step, these anomalies (i.e., residuals) are used to identify the response function of interest by regressing the anomalies of the social outcome variable on the anomalies of the climate variable. Importantly, this two step process of residualizing and then regressing gives the same estimated response curve as a single regression of the social outcome variable on the climate variable plus the controls, according to the Frisch-Waugh-Lovell Theorem (Wooldridge (2010)). Here, the two step procedure is useful because it allows us to conduct model evaluation using only the identifying variation of interest (i.e., Y' and C'). The fraction of the identifying variation in the social outcome explained by the identifying variation in the climate variable is called the "within-R^2".

When residualizing the interation terms in the model specification (e.g., $\psi \mathbf{X_i} C_{it}$ from Equation 4) we implement a double-residualizing procedure following Giesselmann and Schmidt-Catran (2022). Specifically, we residualize the treatment variable (e.g., temperature) with respect to the model fixed effects prior to creating the interaction terms and then we also residualize the interaction terms. We do not also residualize the satellite features in the first step because they are time constant and could have no variation remaining after being residualized by location fixed effects.

1.2.3 Including a ridge regression penalty to avoid overfitting

Homogeneous models are trained using ordinary least squares regression meaning that the coefficients are chosen to minimize the squared difference between predicted and observed values. Models of local vulnerability that include heterogeneous effects using satellite, spatial, or other features are trained using ridge regression. A ridge penalty term is useful in this setting for controlling model complexity, which could otherwise be too high given the large number of satellite (k=4,000) or spatial (e.g., k=1,684 U.S. counties) features. A ridge penalty is applied only to these terms that drive heterogeneity (i.e., ψ in Equation 4). This means that the coefficients are chosen to minimize the squared difference between the predicted and observed values plus a penalty term related to the magnitude of ψ. The magnitude of this penalty is chosen to maximize out-of-sample predictive performance as measured by cross validation in the training and validation set, discussed below. If there is no useful heterogeneity in response captured by the satellite features ($\mathbf{X}$), then ψ will be shrunk to zero by the penalty term. And if there is plenty of useful heterogeneity in response captured by the satellite features, then ψ will be relatively large and contribute to model predictive performance (i.e., reducing the difference between predicted and observed values despite incurring a penalty for its size).

In some models we combine multiple types of features when estimating local climate vulnerability. When we do so, we include and tune a separate ridge penalty for each set of features. For example, in Model (7) of Table S1 we include a separate penalty value for the state, satellite, and irrigation interaction terms. These multiple hyperparameters are tuned sequentially. In the global agriculture models, we allow for separate hyperparameters for the soil moisture and temperature response curves. This means, for example, that in Model (5) of Table S4 we tune the hyperparameter for the temperature $\times$ satellite image interactions and then tune a second hyperparameter for the soil moisture $\times$ satellite image interactions.

1.2.4 Cross-validation for model selection and out-of-sample performance estimation

Models are trained and evaluated using a two-stage cross-validation procedure. The social outcome variable and climate variable are residualized with respect to the model controls before cross-validation to isolate the identifying variation. During the two-stage cross-validation procedure the data is first split into five folds. The first fold (20% of the data) is assigned to be the test set and the remaining four folds (80% of the data) is assigned to be the training and validation set (see Fig. S2). The model is trained in the training and validation set and then used to make predictions in the held-out test set. For the models

with a ridge penalty, the ridge penalty is selected by conducting an inner round of 5-fold cross-validation in the training and validation set. The ridge model is then re-estimated using the selected penalty value and full training and validation set before being applied to the test set. After this first round of training and testing is complete, the second fold is assigned to be the test set and the remaining four folds are assigned to be the training and validation sets. The process repeats until all five folds have been used as a test set. The model performance is the within-R^2 calculated on the predicted and true values for the five combined tests sets.

1.2.5 Data splitting for cross-validation

Generally, the type of extrapolation evaluated in cross-validation should match the way the model is used in practice. We use three types of data splitting when conducting cross validation to understand model performance when extrapolating in different contexts.

Uniform at random splits Evaluation using random splits is useful when one is interested in understanding model performance within the same context in which the model was trained, and is common practice in the climate impacts literature (Proctor et al., 2022; Hultgren et al., 2025). This is useful, for example, for understanding how well a model trained on national yield data globally could predict damages from a drought in a given country and year. And for understanding, historically, what the relationship between extreme heat and crop yields has been.

Spatial splits We conduct spatial cross-validation by splitting data into folds by administrative units (i.e., county, state, or country). This forces models to make prediction on, for example, data from states on which the model was never trained. This is useful for testing the degree to which a model can extrapolate well to new regions on which it was not trained.

Temporal splits We conduct temporal cross-validation by splitting data into folds by consecutive two-year periods. Note that rather than 5-folds, this generates a number of folds equal to the number of years in the dataset divided by two. Testing is still conducted on a single fold and training on the remaining folds.

1.3 Uncertainty estimation

We use a block bootstrapping approach to estimate sampling uncertainty of model performance, response functions, and related values (Fig. S2) (Efron and Tibshirani, 1994). To

implement this block bootstrapping approach, we resample with replacement from the training dataset 50 times to generate 50 new training datasets of equal size. We repeat this for each held-out test set. We then train models in each of these resampled datasets in the same way as done using the non-bootstrapped data. For computational feasibility, we use the same ridge penalty across all bootstrapped datasets for a given outcome variable, setting the value to that optimized using the non-bootstrapped data. We then evaluate model performance by comparing predicted values to those in the held-out test set.

When resampling from the training and validation dataset we sample data using the same type of splits (i.e., uniform, spatial, temporal) as used for cross-validation. For example, when resampling with replacement to estimate uncertainty for a model evaluated using spatial cross-validation by states, samples of entire state time series are taken.

Uncertainty of model performance We use this approach to estimate the uncertainty of model performance (e.g., Table S1). For each held-out test set we generate 50 bootstrapped samples from the training data, train models, and make predictions on the test set. After concatenating the test sets across folds, we calculate 50 R^2 values. These values indicate the sampling distribution of model performance. Tables show the standard deviation of these 50 R^2 values. When calculating 95% confidence intervals for model performance, we take the R^2 estimated using the non-bootstrapped data and add/subtract 1.96 times this standard deviation of the R^2, following a normal approximation of the distribution.

Uncertainty of estimated climate vulnerability To estimate the uncertainty of estimated response curves (e.g., Fig. S3), we use a similar approach. For each held-out test set, we generate 50 bootstrapped samples from the training data and train models using these samples. Note that when calculating the distribution of response functions for a given location, we use all models that were used to make predictions for that location. When using spatial cross-validation, this is the set of 50 bootstrapped models trained when the location is in the held-out test set. For uniform at random and temporal splits, this is all of the 50 boostraps $\times$ N-folds models.

We also use a similar approach to estimate whether satellite-based estimates of local climate vulnerability differ significantly from homogeneous estimates (Fig. 2 and 3). Specifically, for each held-out test set and for each bootstrap sample from the respective training data, we train both homogeneous and satellite-based models. Then, for a given location of interest, we calculate the local climate vulnerability using each of the two model types (Methods 1.5.1) and take their difference. We then calculate the 2.5^{th} and 97.5^{th} percentiles of this distribution of differences and check whether this interval includes zero. If zero is

not within this 95% interval then the null that the two models have the same local climate vulnerability (i.e., that their difference is zero) is rejected using a 5% confidence level. Note that when calculating the distribution of differences in local climate vulnerability between the homogeneous and satellite-based model for a given location, we use all models that were used to make predictions for that location. Further, because we use only 50 bootstrap samples to limit computation, we use a normal approximation to estimate the 2.5^{th} and 97.5^{th} percentiles of the distribution of vulnerability differences between the homogeneous and satellite-based models; we do so by taking the mean plus/minus 1.96 times the standard deviation of the values.

1.4 Model performance evaluation estimating local vulnerability

The primary goal of this analysis is to estimate local climate vulnerability ($\beta_0 + \psi\mathbf{X_i}$ in Equation 4). A challenge with evaluation of estimates of local climate vulnerability, however, is that the true local climate vulnerability is unobserved. Nonetheless we use three approaches to understand the ability of our estimates to recover local climate vulnerability. First, we evaluate the degree to which estimating local vulnerability using satellite image features improves model performance. The satellite image features are time-constant, and we evaluate the performance of models to predict variation in social outcomes over time, so the image features can improve performance only by better predicting local vulnerability. It is not possible for the features to directly predict temporal variation in the social outcome (Y_{it}). Second, we evaluate the performance of a model using shuffled features, which removes all informational content from the image features. Decreases in performance using shuffled features indicate that the imagery contains information useful for estimating local vulnerability. We note, however, that model performance using shuffled satellite features should not necessarily return to the performance of the homogeneous model when evaluated using uniform at random cross-validation because the shuffled satellite features can collectively act similarly to a location-specific intercept and thus allow for useful predictive heterogeneity in effect across spatial units. In accordance with this interpretation, however, models that use shuffled satellite features should have performance similar to the homogeneous model when evaluated using spatial cross-validation, where the the location-specific effect cannot be identified for locations not included in model training. Third, we conduct a simulation experiment where we prescribe local climate vulnerability as a function of observable variables and then estimate the degree to which our approach using satellite imagery can recover the prescribed values. See Supplementary Information A for a description of these experiments and their results.

1.5 Analysis of results

1.5.1 Calculating average slopes from response curves

To depict the estimated local climate vulnerability in maps (e.g., Fig. 2,3) we calculate average response function slopes over different ranges of the climate variable. We do so by taking the difference of the response function value at the beginning and end of the range and dividing by the length of the range. For the piecewise linear response functions used for U.S. agriculture, this is identical to the slopes on the GDD and KDD features. For the quartic response function used for U.S. mortality and the restricted cublic spline response function used for global agriculture this is the average slope across the given range.

Summarizing slopes across ranges requires designation of ranges. For both U.S. and global yield responses to temperature, we use 17°C to 28°C to quantify the benefits of warming and 28°C to 38°C to quantify the damages of extreme heat. The values of 17°C and 38°C were chosen based on the minima and maxima of growing season average values for the global panel. The value of 28°C was selected to be between the estimated optimal temperatures found previously in the U.S. and globally (Schlenker and Roberts, 2009; Proctor et al., 2022). For global yield responses to soil moisture we use 0.1 to 0.25 to calculate benefits of moisture and 0.25 to 0.35 to calculate damages from sodden conditions; the cuttoff was chosen based on the observed median value. For human mortality responses to temperature, we use -7°C to 20°C to quantify benefits of warming and 20°C to 27°C to quantify damages from extreme heat; the cutoff of 20°C was chosen based on the optimum of the homogeneous response curve, and -7 and 27 are the 2.5^{th} and 97.5^{th} percentiles of the state-by-month temperature distribution in our sample. For building damages from flooding, the response is linear and thus no calculation is required.

Maps show the average of slopes calculated for each model used to predict a given spatial unit. For example, for the model of U.S. maize yields trained using uniform at random splits and 5-fold cross validation, the mapped KDD slope is the average of the KDD slope used to predict each of the 5 held out tests sets. And for the model trained using county splits the mapped slope is the value from the model trained to predict that county. We take this approach – rather than re-estimating the model using the full dataset – because re-estimation would require definition of a single ridge penalty parameter, which varies across the folds. Re-estimation using the geometric mean of the fold hyperparameters gives consistent results.

1.5.2 Predictions of local climate vulnerability at the level of the training data and finer

The outcome-climate response function for any spatial unit is calculated as the product of the estimated model parameters and that location's features. For the satellite based model, the response function for any location, i, is given by $\hat{\beta}_0 + \hat{\psi}\mathbf{X}_i$ (where β_0 and ψ have additional dimensions if the response is nonlinear).

Using the global maize yield model, we use this approach to make predictions at the national and 0.1 x 0.1 degree resolutions (Fig. 3). To generate the sub-national estimates, we aggregate the satellite features to a $0.1^\circ \times 0.1^\circ$ resolution global grid using cropped area weights from Earthstat *EarthStat* (N.d.); these are the same weights used to aggregate features to the national level for model training. We make predictions only for grid cells that have at least 10% of their area cropped. Predictions are made using the model trained with uniform at random splits because the model is not being used to extrapolate to countries or time periods not used in model training.

Evaluation of local climate vulnerability predicted at finer resolution than the model was trained We evaluate the ability of our approach to make fine resolution predictions of climate vulnerability in two ways. First, we compare local vulnerability to extreme heat (28°C to 38°C) predicted using the model trained on U.S. county-level data (Fig. 2A) to that predicted using the model trained on global national-level data (Fig. 3). For this comparison, we update the model specification of the global national-level model to control for precipitation, rather than soil moisture, when estimating temperature impacts to match the specification of the U.S. county-level model. The findings of this evaluation are discussed in the main text. A challenge for this evaluation procedure is that the true county-level vulnerability is unknown, and so the evaluation is comparing two estimated values, each of which is likely to have error. Further, this form of evaluation evaluates the global model in only a single country.

Our second form of evaluation uses synthetic data to evaluate the ability of the approach to make predictions of climate vulnerability at finer resolution than the model was trained. This approach complements the first form evaluation because in this setting the true local climate vulnerability is known, and evaluation can be conducted at global scale. This approach and its findings are discussed in the main text and in Supplementary Information A.

1.5.3 Warming experiments for U.S. maize yields and mortality rates

To estimate the influence of 1°C of warming on U.S. maize yields (Fig. 5) we first add 1°C to the county-level daily maximum and minimum temperature values in our sample, which consists of data from the years 2000-2020. We then calculate growing season GDD and KDD values from the warmed data, identically to our calculation of these values for the observed data. We then predict yield for each growing season and take the average over the 21 years of data for both the observed and warmed data. We do so using both a homogeneous model (Equation 5) and the satellite-based model (Equation 7) trained using uniform at random splits. We calculate the effect of warming by differencing the average predicted yields between the warmed and observed data. We estimate effects of greater degrees of warming using an identical procedure (Fig. S20).

To estimate the influence of warming on U.S. mortality rates we use a similar approach. We add 1 °C to the temperature for each state-by-month observation, compute the quartic terms used to estimate the model, and then predict mortality rates using the trained model. We then compare average mortality rates predicted using the warmed and observed data.

Supplementary Information for

Identifying climate-vulnerable populations using satellite imagery

Jonathan Proctor, Luke Sherman, Jeanette Tseng, Solomon Hsiang

Contents

A Supplementary Discussion 1: Simulation Experiments **43**

B Supplementary Tables **49**

C Supplementary Figures **53**

A Supplementary Discussion 1: Simulation Experiments

Here, we conduct synthetic data experiments where we prescribe the local vulnerability (β_i), and then evaluate the ability of our approach to recover the prescribed value. This complements evaluation of our approach based on its ability to improve predictions of the outcome variable (Y_{it}) using observed data, where the true local climate vulnerability is unknown (Methods 1.4).

These synthetic data experiments explore the following five questions: (1) what is the performance of our approach predicting local climate vulnerability, β_i? (2) How does performance predicting local climate vulnerability, β_i, relate to performance predicting the determinants of vulnerability? (3) How does performance predicting local climate vulnerability, β_i, relate to performance predicting the outcome, Y_{it}? (4) How often do estimated 95% confidence intervals for predicted local climate vulnerability contain the true value? (5) What is the performance of our approach predicting local climate vulnerability at finer resolution than the training data?

We find that: (1) the approach using satellite features explains a substantial fraction of local climate vulnerability, (2) the approach predicts local climate vulnerability roughly as well as the satellite features it uses predict the determinants of vulnerability, (3) performance predicting local vulnerability correlates with performance predicting the outcome, but is generally higher, (4) estimated 95% confidence intervals for the satellite-based treatment effect contain the true local climate vulnerability less than 95% of the time, largely because the satellite features do not capture all of the variance in the determinants of vulnerability; though this does not indicate that sampling uncertainty is incorrectly estimated, and (5) our approach can explain substantial variation in local vulnerability at resolutions finer than the data on which it was trained, though performance is lower than at the original resolution.

A.1 Synthetic data experiments: experimental design

We conduct synthetic data experiments for U.S. maize yields, global maize yields, and U.S. mortality to span a range of outcomes and spatial scales. For each dataset we keep the features (X variables) as they are and create a synthetic outcome (Y variable) that is a function of the features and local climate vulnerability. The panel structure and number of observations are the same as in the empirical analyses. We prescribe a pattern of local vulnerability by setting vulnerability, β, in a given location, i, equal to a linear function of observed features values for that location. For U.S and global agriculture, we prescribe vulnerability as a function of the: observed fraction of cropped area equipped for irrigation, average growing season precipitation, average growing season temperature, and population

density.

$$\beta_i = \beta_0 + \gamma_1 \text{irrigation}_i + \gamma_2 \text{precipitation}_i + \gamma_3 \text{temperature}_i + \gamma_4 \text{population}_i$$

For U.S. mortality, we prescribe vulnerability as a function of average income, average temperature, average precipitation, and average population density. Prescribed vulnerability does not change within a location over time.

Using these prescribed values of local climate vulnerability, we then simulate the outcome of interest, Y^s, using the observed weather for each spatial unit i and time period t:

$$Y^s_{it} = \beta_i T_{it} + \text{controls} + e_{it}$$

where e_{it} represents variation in yields not due to weather, which we assume has a normal distribution with mean zero and standard deviation σ. For simplicity, we assume a linear response function. To simulate the fact that different locations have different means, we add a constant ("controls") to the simulated values in each location based on the location's irrigation, precipitation, temperature and population; these values, however, are all removed by the fixed effects used in model estimation.

Once the outcome values are simulated and matched with the observed weather and satellite features, we then apply our satellite-based approach to estimate local vulnerability. Specifically, we assume that we cannot directly observe the determinants of climate vulnerability and instead use our satellite image features ($\mathbf{X_i}$) in place of these prescribed determinants. We estimate the model:

$$Y^s_{it} = \beta_0 T_{it} + \psi \mathbf{X_i} T_{it} + \text{controls} + \epsilon_{it}$$

where the controls are the same set of fixed effects used in the estimation using observations (Methods 1.1). We train and evaluate models in the same way as when using observations. We use spatial splits for cross-validation so that local climate vulnerability for each location is estimated using data not from that location.

When simulating outcome values, we set prescribed coefficients and the variance of the error term such that the predictive performance of the homogeneous model and satellite-based model match what was found in observations. We implement this by tuning β_0 and σ while setting $\gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 = 1$ and standardizing (i.e., subtracting the mean and dividing by the standard deviation) the determinants of local vulnerability: area equipped for irrigation, average precipitation, average temperature, and population density. With standardized variables and the γs all equal to 1, each covariate contributes relatively evenly

to the spatial differences in simulated vulnerability.

A.2 Synthetic data experiments: results

Approach using satellite features predicts substantial variation in local climate vulnerability. Consistent with our tuning procedure, the satellite-based approach improves predictive performance relative to the homogeneous model in this simulated data (e.g., R^2 for synthetic county-by-year U.S. maize yield increases from 0.38 to 0.48). Unlike when we use observed outcome variables, in this experiment we are able to compare the estimated $\hat{\beta}_i$ to the true β_i. We find that our satellite-based approach is able to explain 70% (R^2=.70, ρ^2=0.70) of the variation in simulated maize yield vulnerability to temperature shocks across U.S. counties and 6% (R^2=.06, ρ^2=0.26) across global countries, as well as 50% (R^2=.50, ρ^2=0.51) of the variation in simulated mortality vulnerability to temperature shocks across U.S. states (Fig. S25).

Approach using satellite features predicts local climate vulnerability to the degree that the satellite features predict the determinants of vulnerability. We next test how the ability of the satellite image features to predict the determinants of vulnerability affects the ability of our approach to predict vulnerability by repeating the experiment above for U.S. maize four times. In each iteration, we prescribe U.S. maize vulnerability to be a function of just one of: irrigation, precipitation, temperature, or population. We see that, our approach can predict local vulnerability to the extent that it can predict the local drivers of vulnerability (Table S5). For example, the satellite-based features can predict 69% of the variation in county-level average precipitation across U.S. counties, and 65% of the variation in simulated vulnerability, when differences in simulated vulnerability are fully driven by precipitation. And the satellite-based features can predict 49% of the variation in irrigated area, and 41% of the variation in simulated vulnerability, when simulated vulnerability is driven by only irrigation.

Performance predicting local vulnerability can be substantially higher than performance predicting the social outcome. We next test how performance predicting local vulnerability, β, relates to performance predicting the outcome, Y^s. To do so, we re-estimate the initial simulation experiment many times with different levels of added noise (i.e., e_{it} with different magnitudes of σ). Higher levels of noise make it more difficult to recover the signal within the simulated data, and thus makes it more difficult for the approach to accurately estimate local vulnerability, as well as the outcome. Performance predicting local vulnerability generally falls more slowly with added noise than performance predicting

the outcome (Fig. S26A). Thus, the ability of the model to accurately recover variation in local vulnerability across space can be substantially higher than the ability of the model to predict variation in the outcome. For simulated U.S. maize yields, model performance predicting vulnerability remains relatively high and stable (R^2 = 0.65 to 0.75) for model performances predicting the outcome between R^2 = 0.2 and 1 (R^2 in the observations was 0.42), but the ability to predict vulnerability falls quickly for R^2s predicting the outcome below 0.2 (Figure S26A-B).

For simulated mortality, we find similar qualitative results with performance predicting vulnerability falling slowly with increased noise and reduced model performance predicting the outcome, though here the ability to explain variation in vulnerability remains relatively high even at relatively low model performance predicting the outcome (Figure S26D). For simulated global maize, we also find similar qualitative results, though performance is generally lower and falls more quickly with added noise. Better ability of the approach to separate signal from noise in the U.S. mortality and maize settings, relative to the global maize setting, may be because of the relatively longer time series for each panel unit for U.S. mortality (N = 756 monthly observations per state) and maize (N = 21 annual observations per county) compared with global agriculture (N = 10 annual observations per country), which provides more information to estimate the local vulnerability at each location.

Uncertainty quantification of satellite-based local vulnerability estimates We estimate the sampling distribution of the satellite-based local climate vulnerability using a block bootstrap approach, as described in Methods 1.3. This approach calculates the distribution of vulnerabilities that would be estimated if we were to repeat the experiment – collecting and analyzing the social outcome data, weather data and satellite data – many times. This allows us, for example, to calculate a 95% confidence interval for the satellite-based local vulnerability, or to test whether a local vulnerability estimate differs significantly from the average treatment effect, or from zero (Efron and Tibshirani, 1994).

Another relevant, though distinct, question is the frequency with which the estimated 95% confidence interval for the satellite-based effect contains the true local vulnerability. Note that we should not necessarily expect the estimated 95% confidence interval for the satellite-based effect to contain the true local vulnerability 95% of the time because the satellite imagery does not contain perfect information on all the determinants of vulnerability. Likewise, the 95% confidence interval for the average treatment effect, or for the average effect conditional on a single covariate like irrigation, would also likely not contain the true local vulnerability 95% of the time. With this in mind, it is nonetheless informative to evaluate the frequency with which our confidence intervals for the satellite-based estimates of local

climate vulnerability contain the true prescribed local vulnerability.

We find that the estimated 95% confidence interval for the satellite-based effect contains the prescribed true value 48-71% of the time across social outcomes (Fig. S27). For context, the 95% confidence intervals for the average treatment effect contain the true values XX-YY% of the time. A primary reason for the satellite-based confidence intervals sometimes not containing the true effect is that the satellite image features do not perfectly measure the determinants of vulnerability (Table S5), and thus do not perfectly estimate the true local vulnerability (Fig. S25). This causes the confidence intervals to be centered on the wrong value.

To test this, we re-simulate the outcome, Y, such that information on the determinants of vulnerability is fully contained in the satellite-based features. We implement this by simulating vulnerability as a function of the image features themselves: $\beta_i = \gamma \mathbf{X}_i$, rather than as a function of observed irrigation, temperature, precipitation, ...etc. We find that when vulnerability is simulated as a function of the imagery, our approach is able to explain substantially more of the variance in β_i and obtain confidence intervals that contain the true effect 71-85% of the time (Fig. S27). This experimental design is more comparable with that of other studies, which similarly provide algorithms to estimate conditional average treatment effects with the same information used to simulate the treatment effects and outcomes of interest (e.g., Wager and Athey (2018); Künzel et al. (2019)). Accordingly, the performance of our approach in this setting is consistent with prior results, though direct comparison is complicated by differences in simulation design (Wager and Athey, 2018; Künzel et al., 2019; Dorie et al., 2018). It would be valuable for future research to develop methods that can reliably construct 95% confidence intervals that contain the true value 95% of the time – particularly for cases where the drivers of heterogeneity are high-dimensional and not directly observed, as in our setting of estimating local climate vulnerability.

Performance predicting vulnerability at fine resolution We next evaluate the ability of our model to make predictions of local vulnerability at fine resolution when trained on coarse data using a similarly structured synthetic data experiment and taking crop yields as a case study. We again simulate local vulnerability as a function of irrigation, precipitation, temperature and population, but now at 0.25 degree resolution over global cropped areas. We then aggregate this vulnerability, along with temperature data, to the national level weighting by cropped area, and then simulate yields as in the previous experiment. We then apply our approach to estimate local vulnerability using satellite imagery.

For this downscaling experiment, we no longer evaluate country performance out-of-sample, but instead use a pooled model (e.g., UAR in Table S4). We find that at the national

level we are able to recover 6% (R^2=.06, ρ^2=0.26) of the variance in local vulnerability when the level of simulated noise is set such that the model R^2 predicting simulated yields equals that of actual yields, as described above (Fig. S19). When predicting vulnerability at 0.25 degree resolution performance we get R^2<0 but are able to explain 19% of the variation in general (ρ^2=0.19) and 26% of the variation in rank order for 0.25 degree grid tiles (Spearman $r_s^2 = 0.26$). These results, along with those of Fig S18, indicate that predictions of vulnerability made at finer resolutions than the training data have some skill predicting relative vulnerabilities but are not very skillful at evaluating the exact level of vulnerability.

B Supplementary Tables

US Maize		UAR	County OOS	State OOS	2 Year OOS
(1)	**Single temperature response**	0.335	0.335	0.328	0.314
		(0.0)	(0.0)	(0.007)	(0.011)
(2)	**Local temperature response by state**	0.410	0.409	-	0.351
		(0.001)	(0.001)	-	(0.016)
(2B)	**Local temperature response by county (penalized)**	0.356	-	-	0.351
		(0.003)	-	-	(0.016)
(2Bnp)	**Local temperature response by country (not penalized)**	0.284	-	-	<0
		-	-	-	-
(3)	**Local temperature response by irrigation**	0.365	0.366	0.358	0.345
		(0.0)	(0.0)	(0.005)	(0.011)
(4)	**Local temperature response by avg temperature**	0.366	0.367	0.359	0.345
		(0.0)	(0.0)	(0.007)	(0.012)
(5)	**Local temperature response by satellite image**	0.422	0.419	0.342	0.351
		(0.002)	(0.003)	(0.02)	(0.018)
(5s)	**Local temperature response by scrambled satellite image**	0.343	0.335	0.328	0.346
		(0.003)	(0.0)	(0.007)	(0.015)
(6)	**Local temperature response by satellite image & irrigation**	0.423	0.420	0.341	0.355
		(0.002)	(0.002)	(0.021)	(0.014)
(7)	**Local temperature response by state, satellite image & irrigation**	0.429	0.430	-	0.361
		(0.001)	(0.001)	-	(0.014)

US Soy		UAR	County OOS	State OOS	2 Year OOS
(1)	**Single temperature response**	0.233	0.233	0.229	0.207
		(0.0)	(0.0)	(0.004)	(0.015)
(2)	**Local temperature response by state**	0.263	0.264	-	0.209
		(0.001)	(0.001)	-	(0.016)
(2B)	**Local temperature response by county (penalized**	0.233	-	-	0.213
		(0.001)	-	-	(0.017)
(2Bnp)	**Local temperature response by county (not penalized)**	0.091	-	-	0.126
		-	-	-	-
(3)	**Local temperature response by irrigation**	0.24	0.241	0.238	0.214
		(0.0)	(0.0)	(0.004)	(0.015)
(4)	**Local temperature response by avg temperature**	0.238	0.238	0.232	0.211
		(0.0)	(0.0)	(0.004)	(0.015)
(5)	**Local temperature response by satellite image**	0.259	0.259	0.241	0.211
		(0.001)	(0.002)	(0.005)	(0.018)
(5s)	**Local temperature response by shuffled satellite image**	0.233	0.233	0.229	0.206
		(0.0)	(0.0)	(0.04)	(0.015)
(6)	**Local temperature response by satellite image & irrigation**	0.26	0.26	0.242	0.214
		(0.001)	(0.002)	(0.004)	(0.017)
(7)	**Local temperature response by state, satellite image & irrigation**	0.268	0.269	-	0.213
		(0.001)	(0.001)	-	(0.017)

Table S1: **Allowing the influence of temperature on yields to differ locally according to satellite imagery increases model performance predicting U.S. maize and soybean yield anomalies**. Models are evaluated on their ability to predict yield shocks – i.e. anomalies from local yield trends. Model performance is measured using the coefficient of determination, R^2. Models are evaluated out-of-sample (OOS) using uniform at random (UAR) cross-validation, as well as cross-validation splitting training and validation sets by county, state, and 2-year periods.

US Mortality		UAR	State OOS	2 Year OOS
(1)	**Single temperature response**	0.0341	0.0333	0.0327
		(0.0002)	(0.0007)	(0.001)
(2)	**Local temperature response by state**	0.0398	-	0.0390
		(0.0008)	-	(0.0013)
(2B)	**Local temperature response by state (not penalized)**	0.0370	-	0.0372
		(0.0013)	-	(0.002)
(3)	**Local temperature response by med household income**	0.0360	0.0350	0.0343
		(0.0003)	(0.0007)	(0.0011)
(4)	**Local temperature response by avg temperature**	0.0391	0.0384	0.0376
		(0.0003)	(0.0006)	(0.0011)
(5)	**Local temperature response by satellite image**	0.0431	0.0413	0.0402
		(0.0006)	(0.0012)	(0.0015)
(5s)	**Local temperature response by shuffled satellite image**	0.0390	0.0176	0.0376
		(0.0012)	(0.0041)	(0.0015)
(6)	**Local temperature response by state & satellite image**	0.0419	-	0.0405
		(0.0009)	-	(0.0015)

Table S2: **Allowing the influence of temperature on human mortality rates to differ locally according to satellite imagery increases model performance.** Models are evaluated on their ability to predict mortality rate anomalies. Models are evaluated out-of-sample (OOS) using uniform at random (UAR) cross-validation, as well as cross-validation splitting training and validation sets by state and 2-year periods.

US Building Damages		UAR	County OOS	State OOS	2 Year OOS
(1)	**Single flood depth response**	0.0754	0.0725	0.0691	0.0635
		(0.0000)	(0.0026)	(0.0036)	(0.0019)
(2)	**Local flood depth response by state**	0.0832	0.0754	-	0.0704
		(0.0000)	(0.0047)	-	(0.0028)
(2B)	**Local flood depth response by county**	0.0989	-	-	0.0755
		(0.0001)	-	-	(0.0021)
(2C)	**Local flood depth response by county (not penalized)**	0.0978	-	-	0.0723
		(0.0002)	-	-	(0.0028)
(3)	**Local flood depth response by avg flood depth**	0.0910	0.0908	0.0910	0.0761
		(0.0000)	(0.0005)	(0.0018)	(0.0016)
(4)	**Local flood depth response by satellite image**	0.0984	0.0735	0.0694	0.0711
		(0.0001)	(0.0084)	(0.0023)	(0.0019)
(4S)	**Local flood depth response by shuffled satellite image**	0.0981	0.0712	0.0658	0.0726
		(0.0002)	(0.0047)	(0.0039)	(0.0022)
(5)	**Local flood depth response by state & satellite image**	0.0984	0.0756	-	0.0705
		(0.0001)	(0.0052)	-	(0.0019)
(5B)	**Local flood depth response by county & satellite image**	0.0989	-	-	0.0756
		(0.0001)	-	-	(0.0021)
(5C)	**Local flood depth response by avg depth & satellite image**	0.0983	0.0915	0.0913	0.0766
		(0.0001)	(0.0017)	(0.0014)	(0.0018)

Table S3: **Allowing the impact of flood depth on relative building damages to differ locally according to satellite imagery increases model performance.** Models are evaluated on their ability to predict damage anomalies from local property value trends. Models are evaluated out-of-sample (OOS) using uniform at random (UAR) cross-validation, as well as cross-validation splitting training and validation sets by state and 2-year periods.

Global Maize		UAR	Country OOS	2 Year OOS
(1)	**Single temperature & soil moisture response**	0.088	0.083	0.081
		(0.009)	(0.009)	(0.008)
(2)	**Local temperature & soil moisture response by country (penalized**	0.126	-	0.101
		(0.017)	-	(0.019)
(2np)	**Local temperature & soil moisture response by country (not penalized)**	<0	-	<0
		-	-	-
(3)	**Local temperature & soil moisture response by irrigation**	0.1	0.088	0.086
		(0.01)	(0.01)	(0.011)
(4)	**Local temperature & soil moisture response by avg temperature & avg soil moisture**	0.139	0.099	0.127
		(0.011)	(0.018)	(0.011)
(5)	**Local temperature & soil moisture response by satellite image**	0.146	0.123	0.154
		(0.029)	(0.03)	(0.021)
(5s)	**Local temperature & soil moisture response by shuffled satellite image**	0.068	0.081	0.074
		(0.01)	(0.009)	(0.009)
(6)	**Local temperature & soil moisture response by satellite image & irrigation**	0.161	0.144	0.138
		(0.028)	(0.016)	(0.02)
(7)	**Local temperature & soil moisture response by country, satellite image & irrigation**	0.158	-	0.123
		(0.027)	-	(0.026)

Global Soy		UAR	Country OOS	2 Year OOS
(1)	**Single temperature & soil moisture response**	0.105	0.097	0.133
		(0.013)	(0.023)	(0.009)
(2)	**Local temperature & soil moisture response by country**	0.172	-	0.154
		(0.026)	-	(0.034)
(2np)	**Local temperature & soil moisture response by country (not penalized)**	<0	-	<0
		-	-	-
(3)	**Local temperature & soil moisture response by irrigation**	0.127	0.12	0.152
		(0.015)	(0.023)	(0.01)
(4)	**Local temperature & soil moisture response by avg temperature & avg soil moisture**	0.128	0.096	0.155
		(0.015)	(0.024)	(0.011)
(5)	**Local temperature & soil moisture response by satellite image**	0.194	0.143	0.114
		(0.026)	(0.043)	(0.047)
(5s)	**Local temperature & soil moisture response by shuffled satellite image**	0.145	0.075	0.161
		(0.096)	(0.019)	(0.111)
(6)	**Local temperature & soil moisture response by satellite image & irrigation**	0.192	0.149	0.092
		(0.022)	(0.04)	(0.051)
(7)	**Local temperature & soil moisture response by country, satellite image & irrigation**	0.192	-	0.135
		(0.027)	-	(0.053)

Table S4: **Allowing the influence of temperature and soil moisture on yields to differ locally according to satellite imagery increases model performance predicting maize and soybean yield shocks globally**. Models are evaluated on their ability to predict yield shocks – i.e. anomalies from local yield trends. Models are evaluated out-of-sample (OOS) using uniform at random (UAR) cross-validation, as well as cross-validation splitting training and validation sets by country and 2-year periods.

	R^2 predicting covariate	R^2 predicting $\hat{\beta}$
Irrigated area	0.49	0.41
Avg precipitation	0.69	0.65
Avg temperature	0.97	0.95
Population	0.67	0.62

Table S5: **Performance predicting both the determinants of simulated vulnerability and simulated vulnerability.** The first column shows the ability of the satellite features to predict each of four determinants of vulnerability. The second column shows the ability of satellite features to predict simulated temperature vulnerability driven by the individual determinants. Observations are at the county level and the panel data used for the $\hat{\beta}$ estimates are structured to mirror that of the US maize data. Both the R^2 for the covariate and the R^2 from the $\hat{\beta}$ are estimated county out-of-sample. For each simulation, the total amount of prescribed heterogeneity from each single covariate is rescaled so that it is approximately consistent with the simulation using all four covariates.

C Supplementary Figures

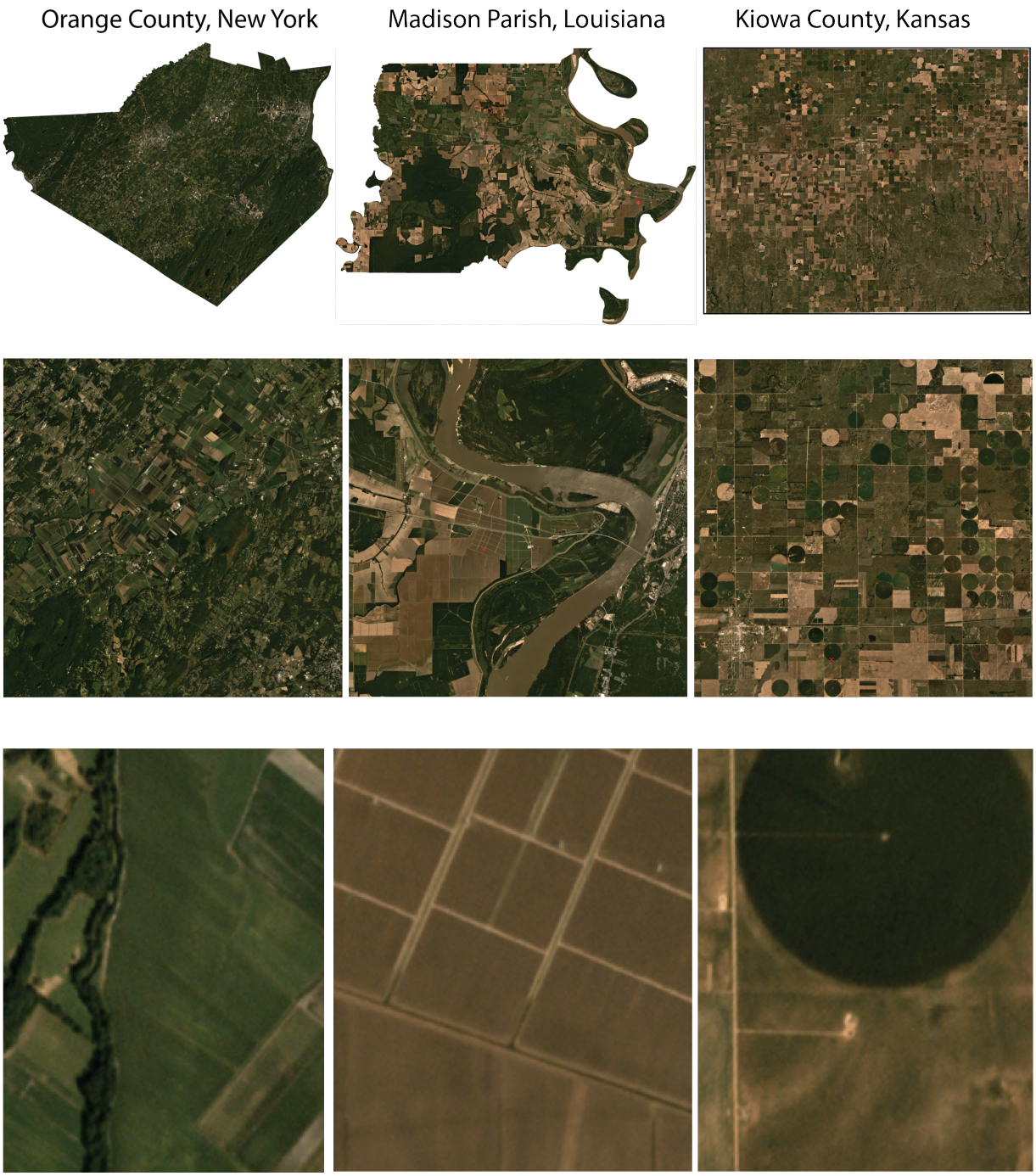


Figure S1: **Three U.S. counties that grow maize with differing agricultural practices and endowments visible in the satellite imagery.** The first row shows county-scale imagery used to predict local climate vulnerability of maize yields. The second and third rows zoom into the imagery to show the resolution and examples of the type of agricultural practices that are visible. Imagery is 2019 Planet Labs optical (i.e., red, green, and blue) imagery with a pixel resolution of 4.77m×4.77m at the equator.

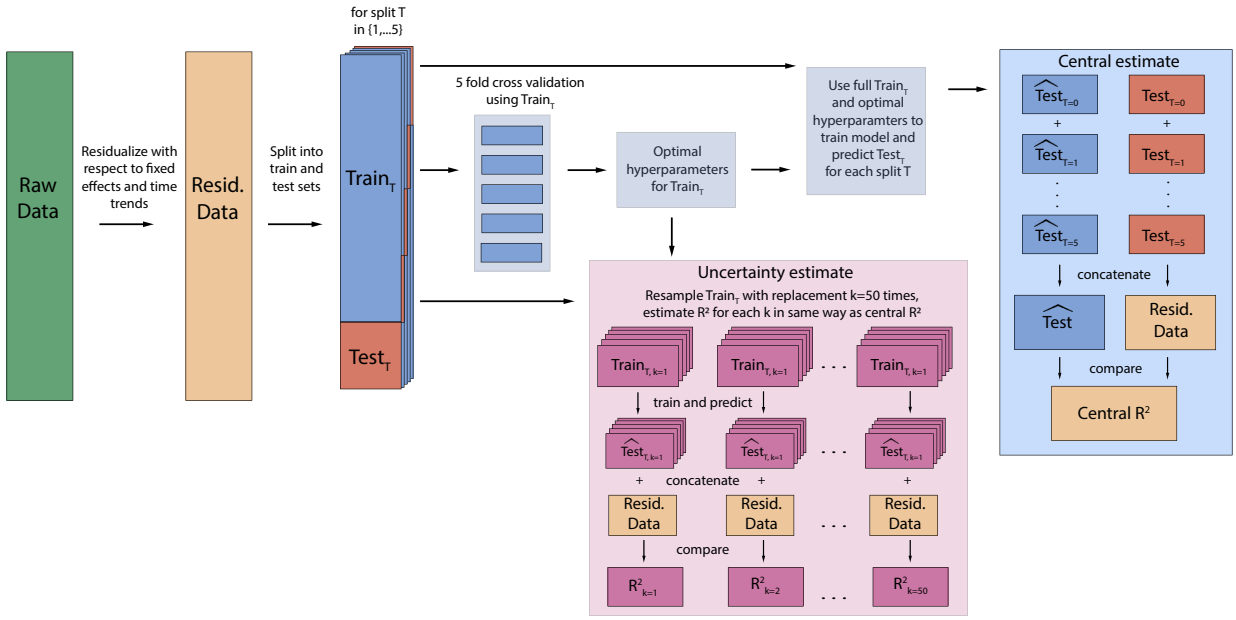


Figure S2: **A depiction of model training, evaluation, and uncertainty quantification.** See Methods 1.2 for details.

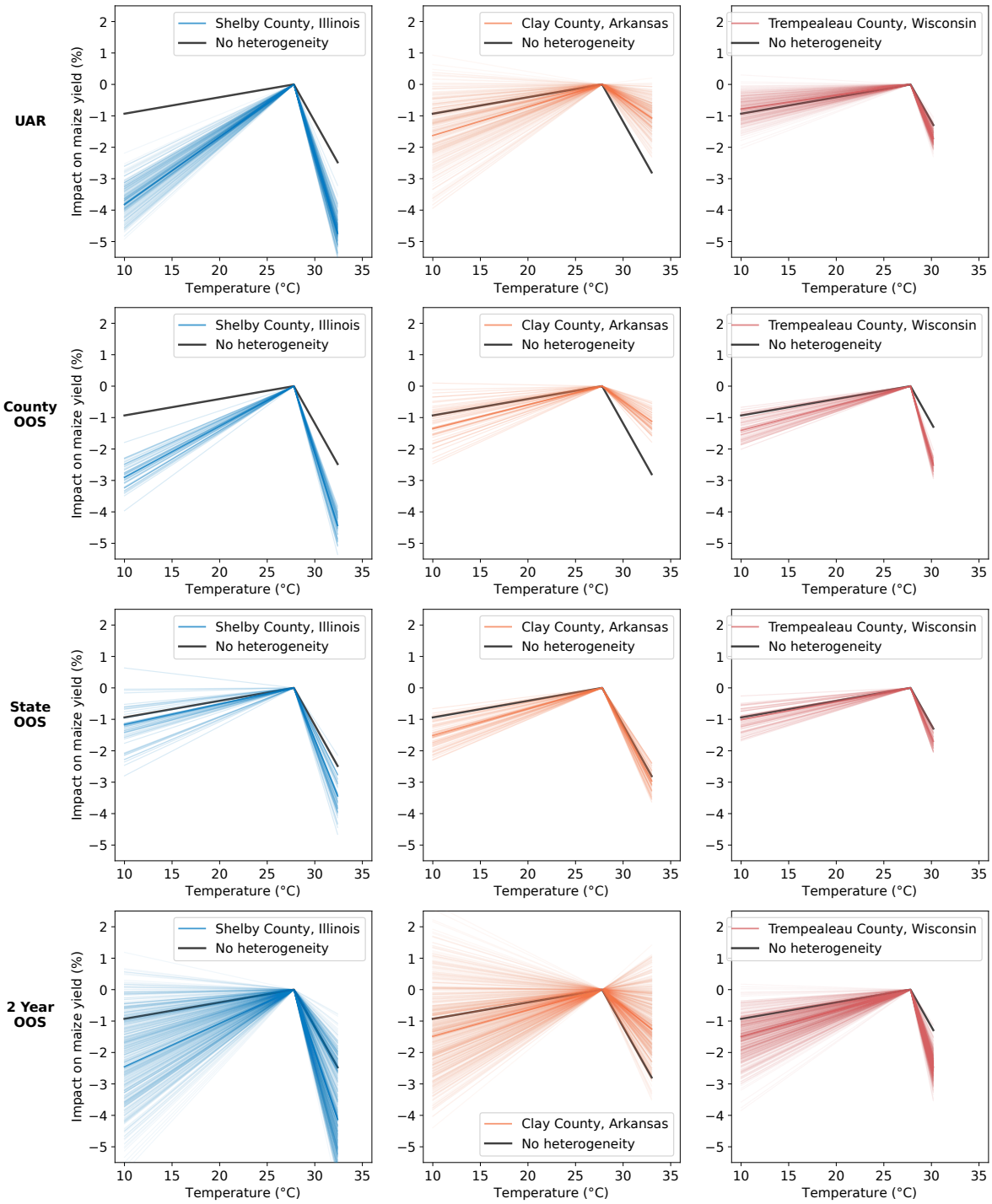


Figure S3: **Estimating uncertainty: the response function sampling distribution for US maize by cross-validation type.** Each of the three columns shows the response function for a different US county. Each row shows models trained using a different cross-validation type. The thick colored line shows the estimated response function for that county. The thin colored lines show response functions estimated using bootstrapped samples. The thick black line shows the estimated homogeneous response across the entire U.S. for reference.

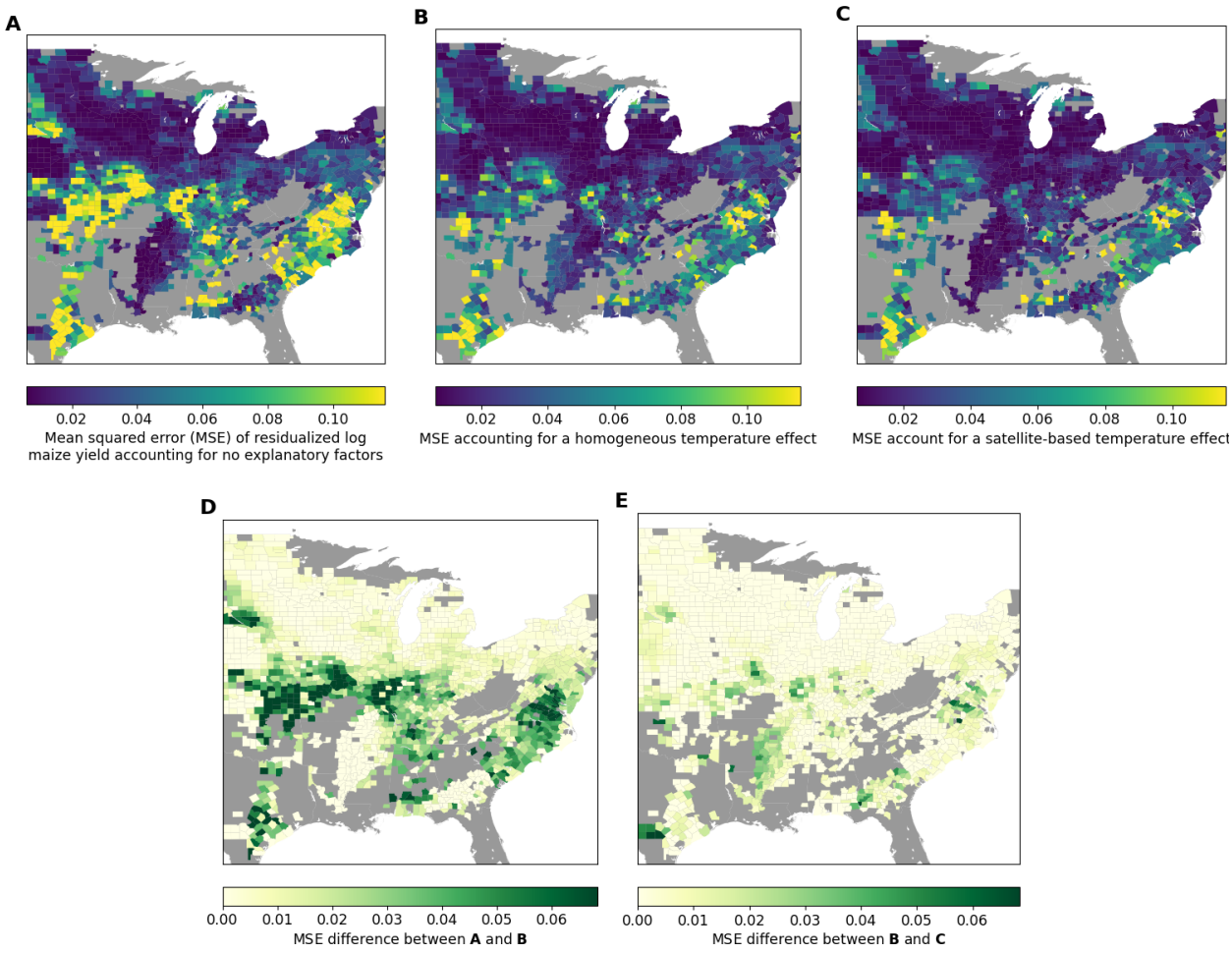


Figure S4: **For U.S. maize we examine where inter-annual variance in yield occurs and where the homogeneous and image-based models are able to explain that variance.** Panel A shows the variance of maize yields after the data has been demeaned and detrended. This is the identifying variation. Panel B shows the variation in yield left that is not explained by temperature and rainfall using a homogeneous model. Panel C shows the variation in yield left that is not explained by temperature and rainfall using a satellite-based model. The first map on the bottom shows the difference between A and B – i.e., the variation explained by a homogeneous temperature effect. And panel E shows the difference between B and C – i.e., the additional variation explained by allowing for heterogeneity in the temperature effect. Models are trained using uniform-at-random cross-validation and models accounting for temperature effects also include precipitation.

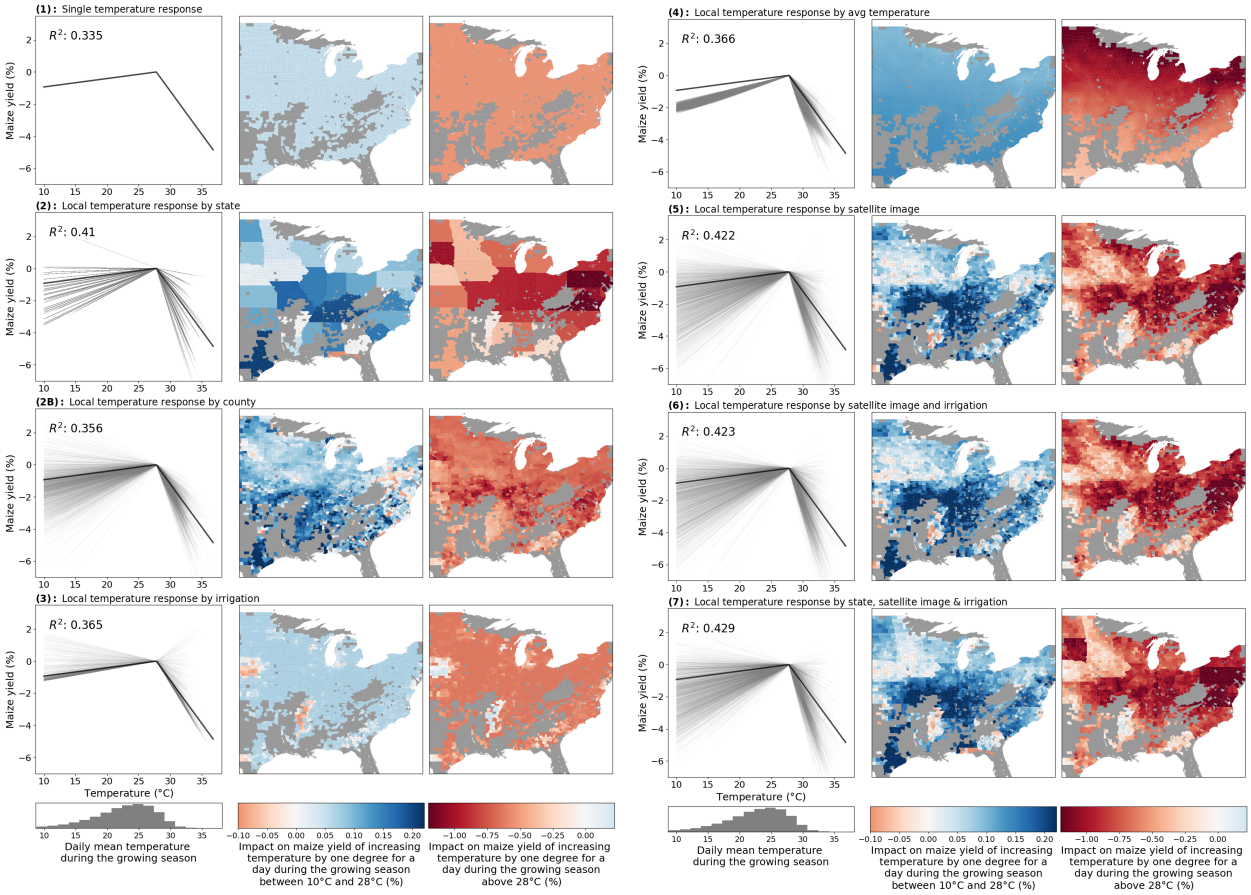


Figure S5: **Local vulnerability of U.S. maize yields to temperature estimated using traditional and satellite-based approaches.** Response curves show the estimated influence of temperature on yields. The thick line shows the estimated homogeneous effect and the light grey lines show the local response estimated for each county. Responses are visually clipped to the range of daily temperature values experienced in each county over the time of our study. The map on the left shows benefits of warming (the slope of the response function to the left of the kink) and the map on the right shows damages from extreme heat (the slope to the right of the kink). Darker blues indicate greater benefits of warming and darker reds indicate greater damages of extreme heat. Results are shown for 8 different models, with performance evaluated with UAR cross validation indicated in text. Performance using other cross-validation strategies is given in Table S1.

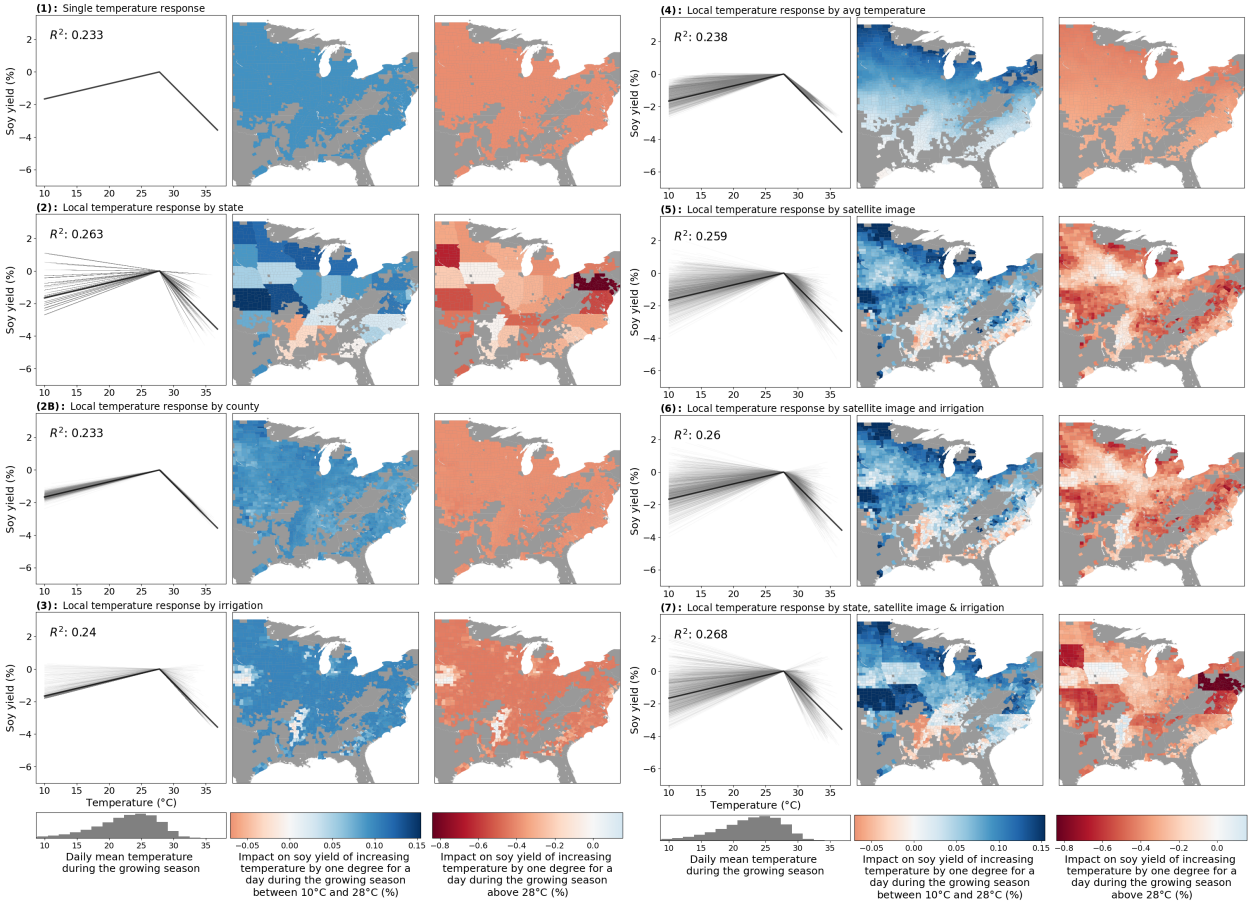


Figure S6: **Local vulnerability of U.S. soybean yields to temperature estimated using traditional and satellite-based approaches.** Same as Fig. S5.

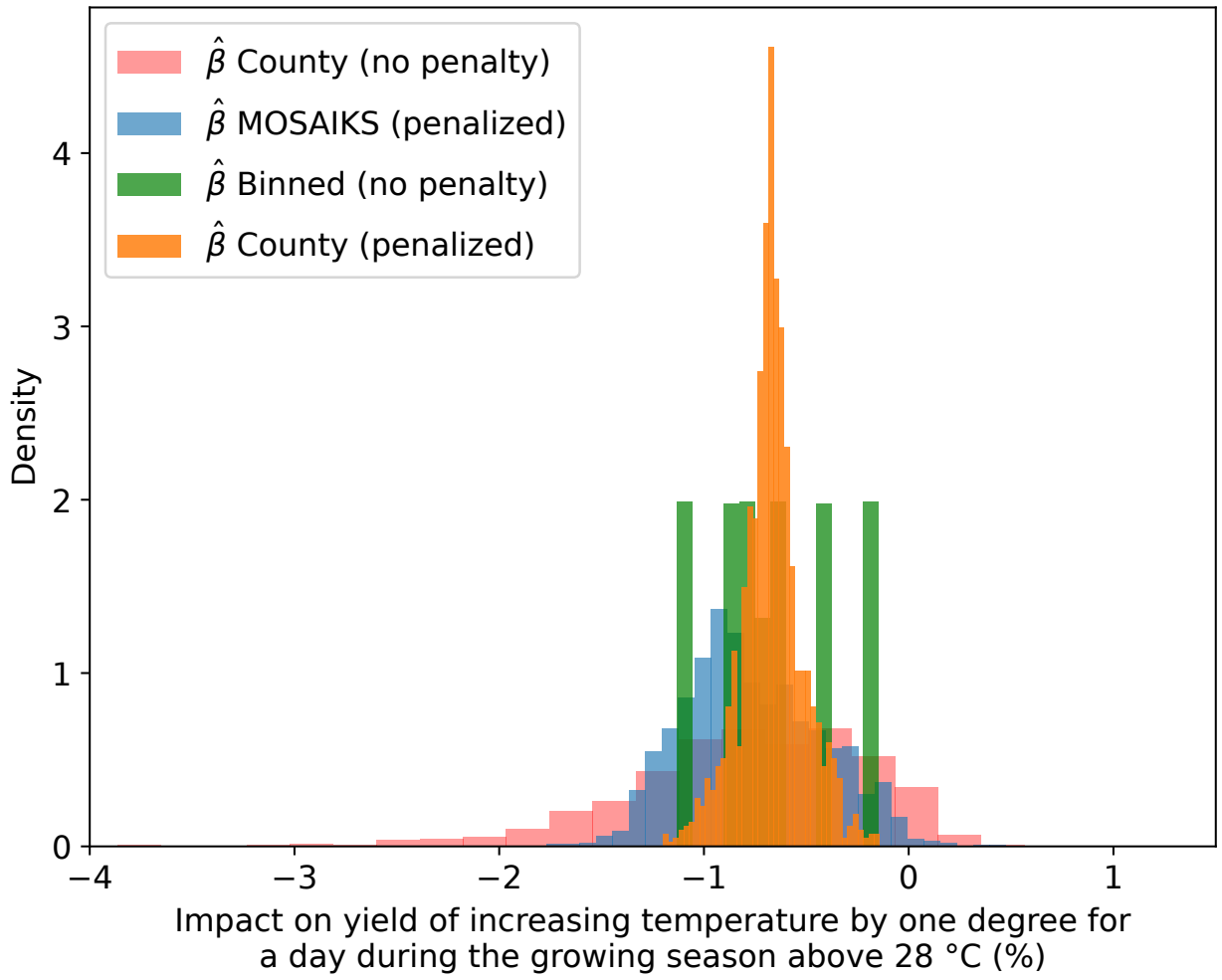


Figure S7: **The distribution of U.S. maize vulnerability to extreme heat estimated using four different approaches.** The blue histogram shows values estimated using the satellite-based approach estimated using ridge regression. The pink histogram shows estimates using county-level spatial features without any ridge penalty applied (i.e., standard OLS), and the orange values show estimates from the same specification but with an optimal ridge penalty applied. The green histogram shows estimates from the approach described in Fig. S21 where estimates are grouped according to the satellite-based estimates and then vulnerability is re-estimated using standard OLS with no ridge penalty.

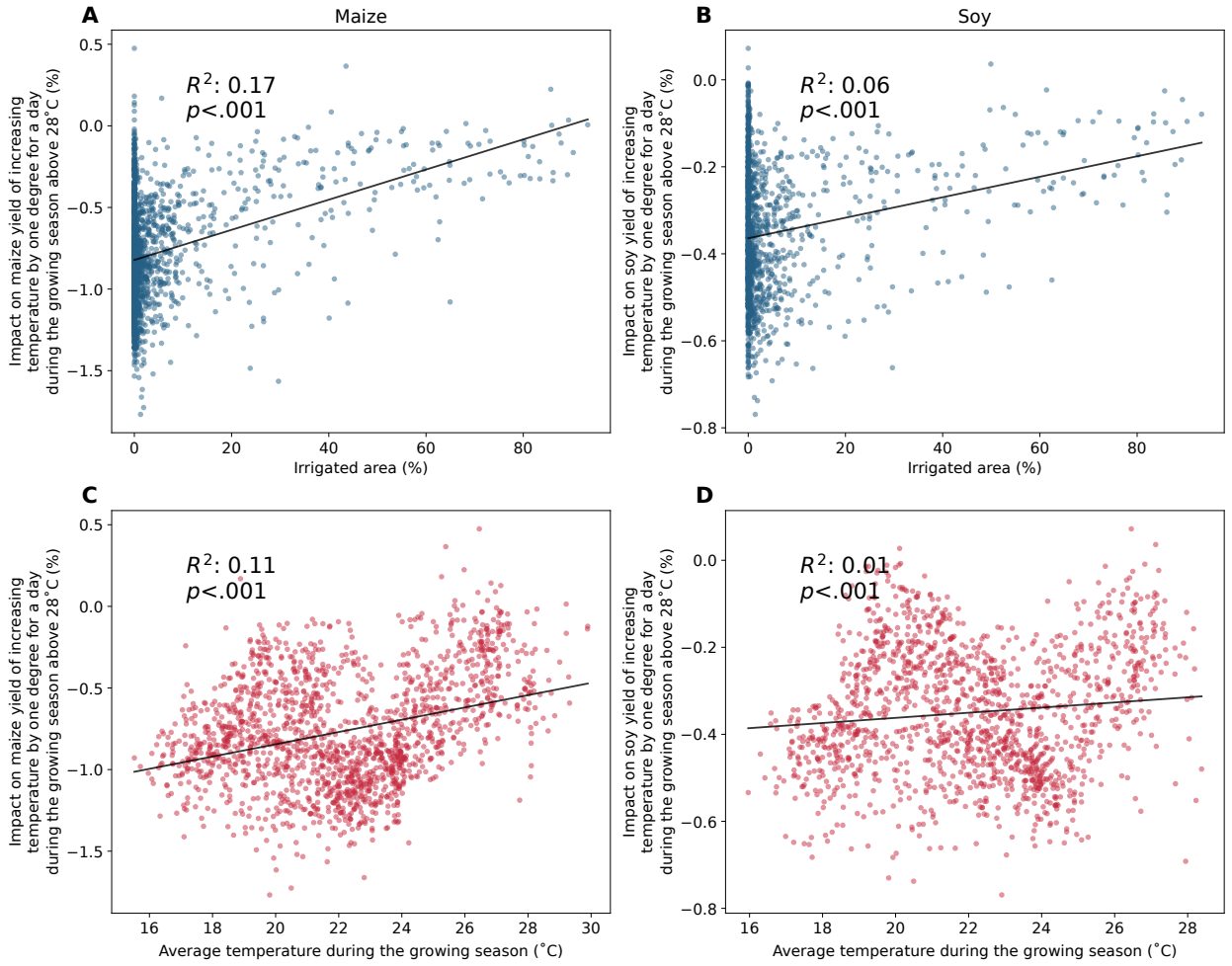


Figure S8: **Irrigation and average temperature explain some but not all of the differences in local U.S. maize and soy vulnerability to extreme heat estimated using satellite imagery.** Climate vulnerability estimated using satellite imagery (y-axis) plotted against area equipped for irrigation (x-axis) (top) and average temperature (x-axis) (bottom) for each analyzed U.S. county for maize (left) and soybeans (right).

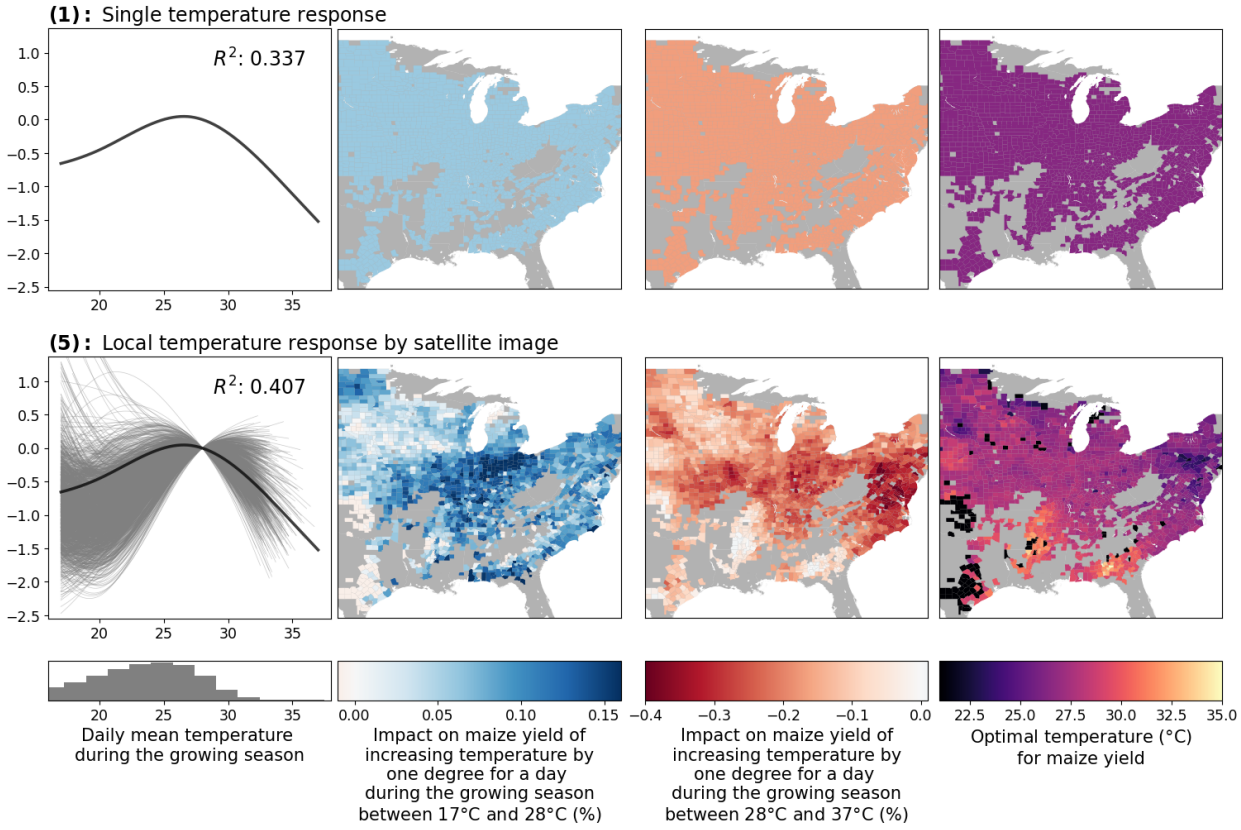


Figure S9: **Results for U.S. maize are consistent when using a restricted cubic spline feature expansion of temperature rather than than growing degree and killing degree days.** Response curves and maps are similar to those in Fig. S5.

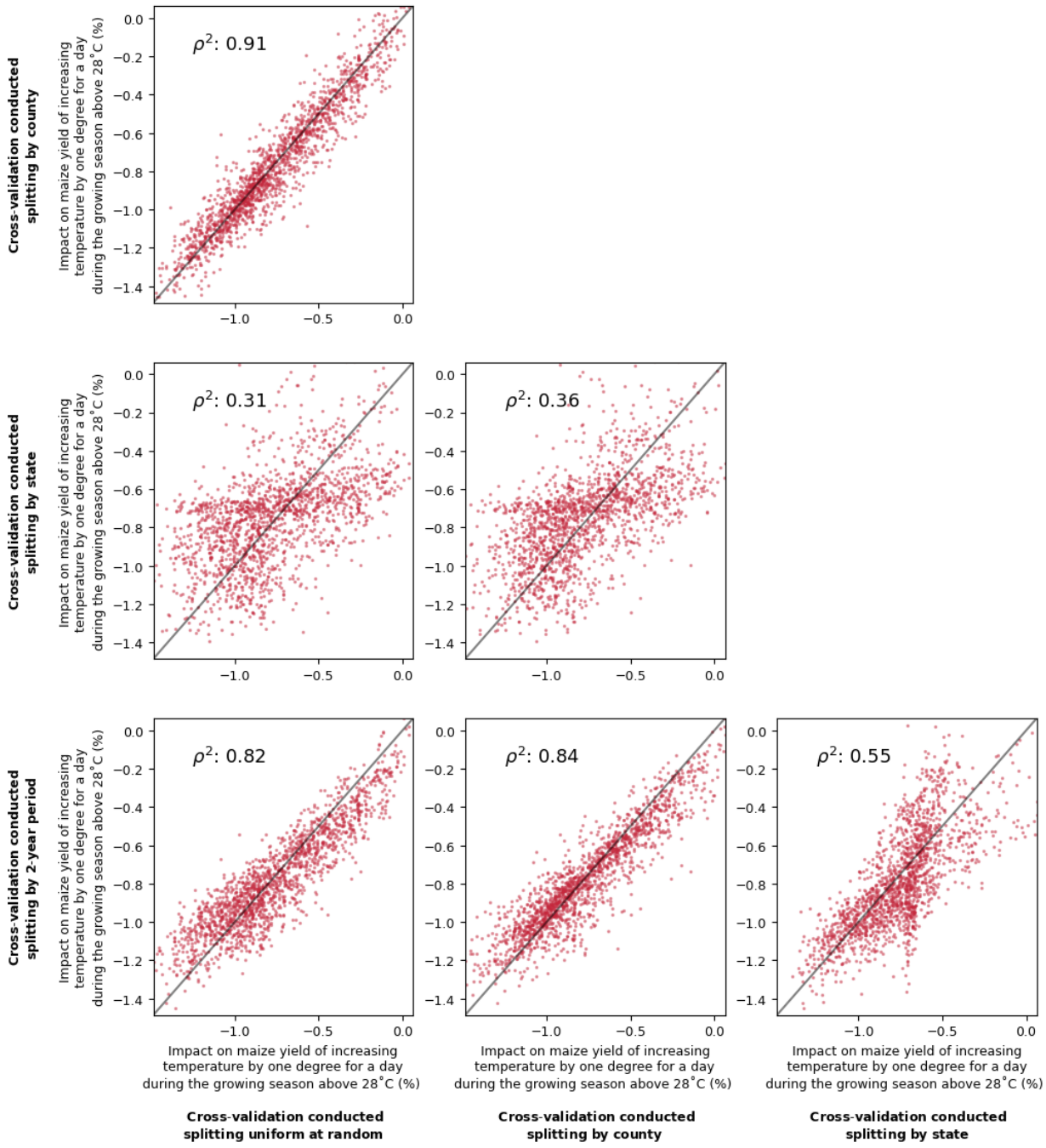


Figure S10: **Comparing local vulnerability of U.S. maize yields to extreme heat estimated using varying sampling strategies.** Each point is the estimated vulnerability for a given U.S. county.

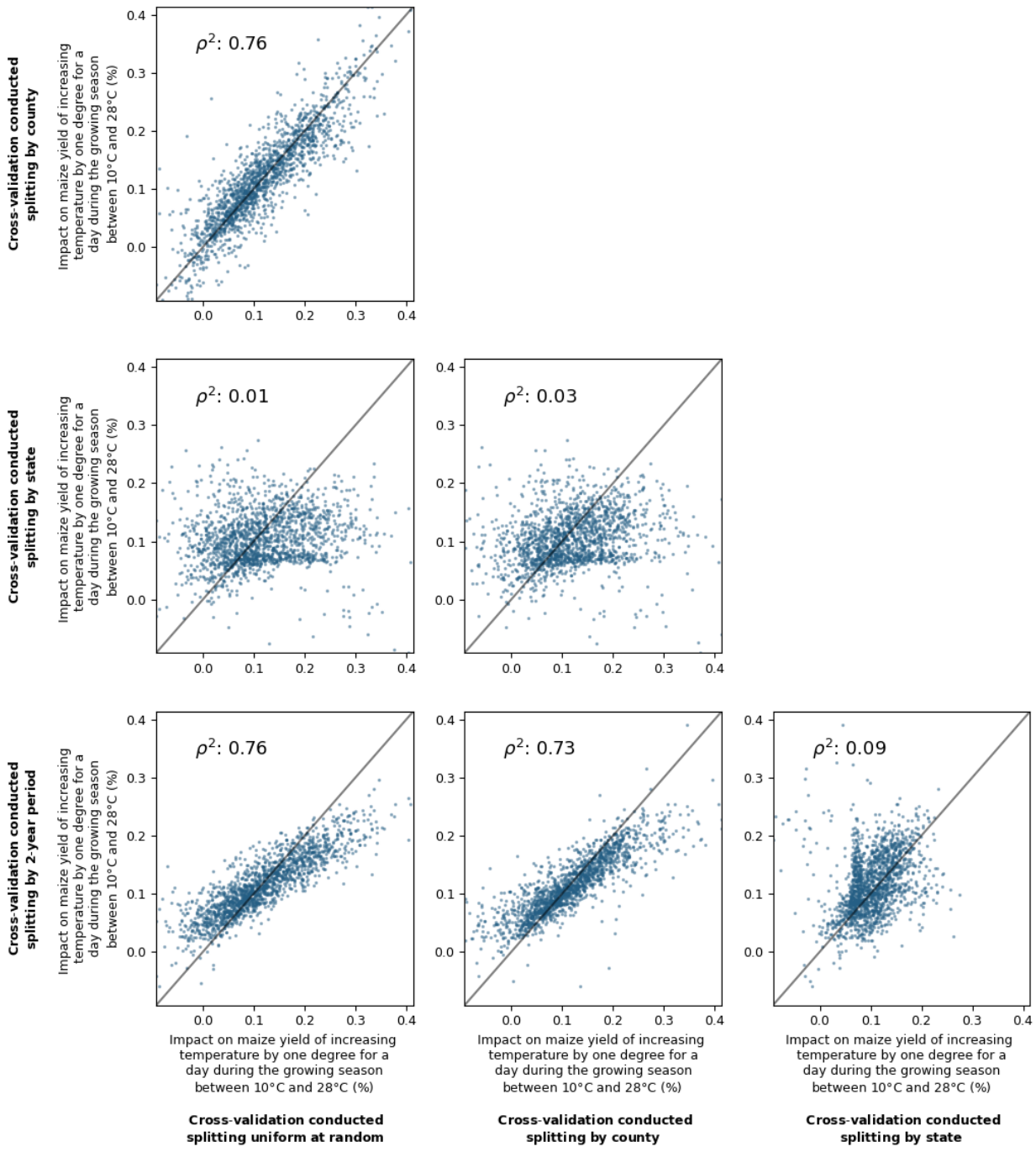


Figure S11: **Comparing local sensitivity of U.S. maize yields to warming estimated using varying sampling strategies.** Each point is the estimated vulnerability for a given U.S. county.

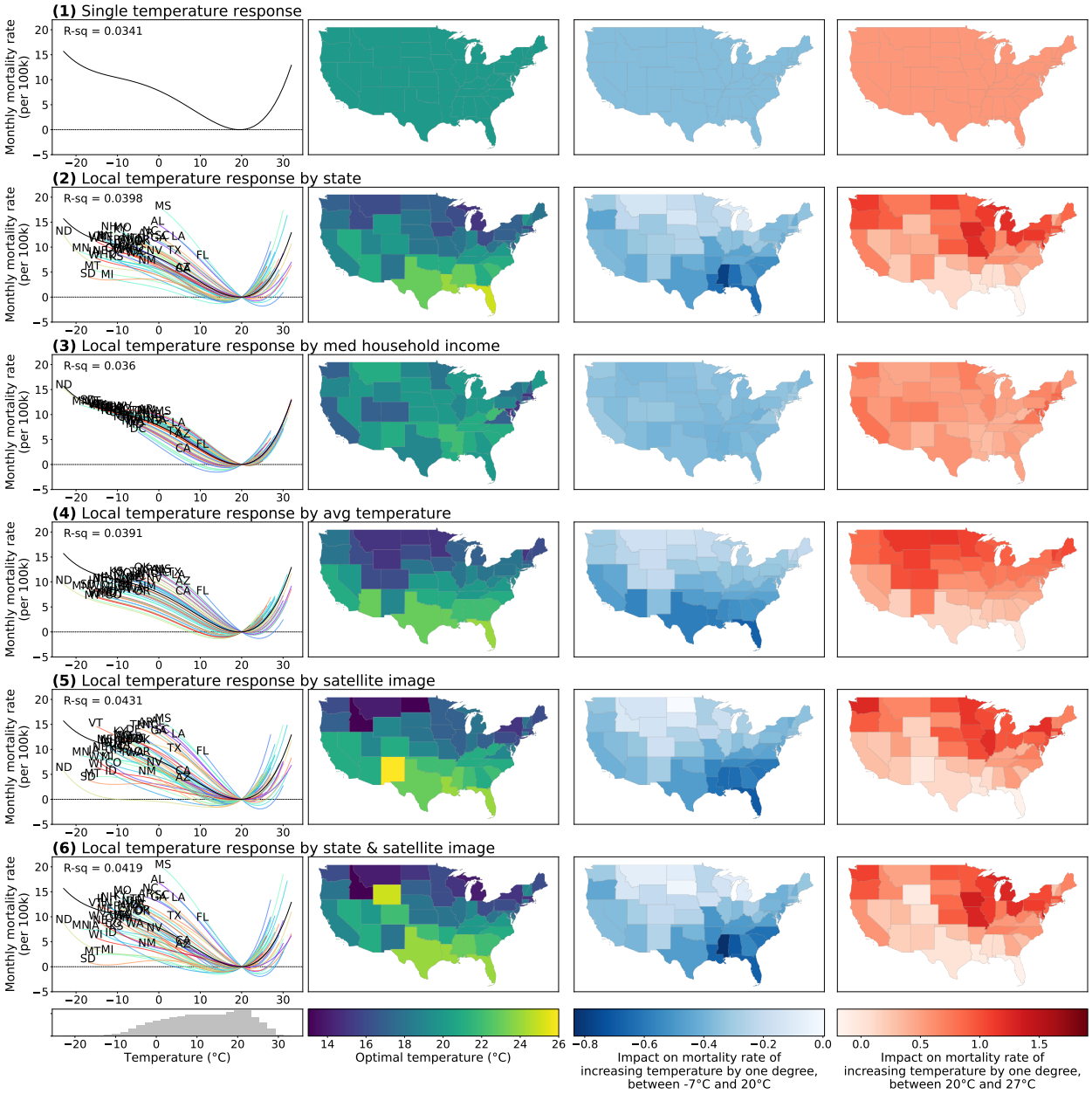


Figure S12: **Local vulnerability of human mortality to temperature estimated using traditional and satellite-based approaches.** Response curves show the estimated influence of temperature on all cause mortality rate per 100k. The thick line shows the estimated homogeneous effect and the light colored lines show the local response estimated for each state. Responses are visually clipped to the range of monthly temperature values experienced in each state over the time of our study. The map on the left shows the estimated mortality-minimizing temperature. The map in the middle shows the benefits of warming (the slope of the response function from -7°C to 20°C) and the map on the right shows damages from extreme heat (the slope from 20°C to 27°C). Results are shown for 6 different models, with performance evaluated with UAR cross validation indicated. Performance using other cross-validation strategies is given in Table S2.

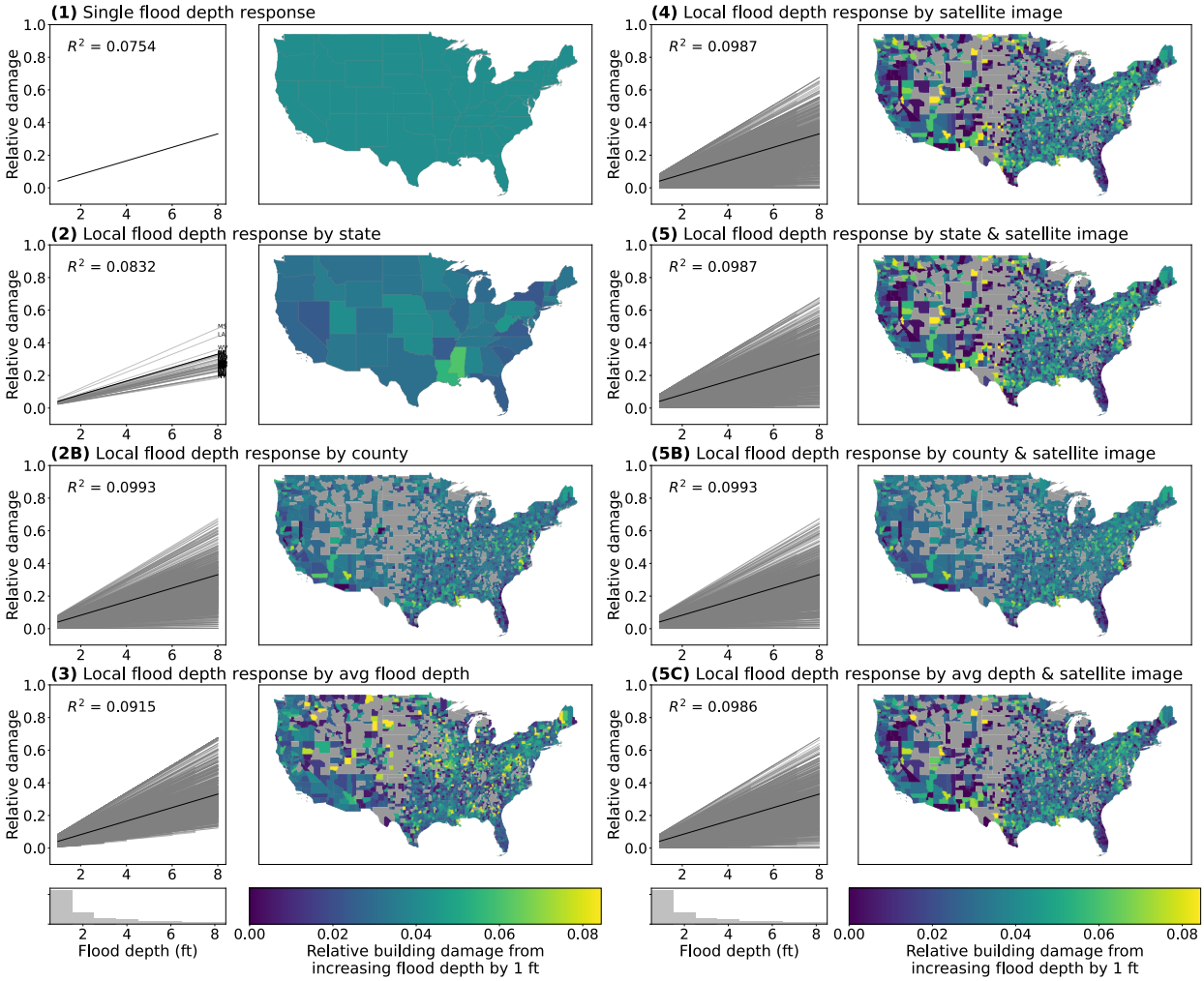


Figure S13: **Local vulnerability of structures to flooding estimated using traditional and satellite-based approaches.** Response lines show the estimated influence of flood depth on relative building damage (0.1 is 10% of building value damaged). The thick line shows the estimated homogeneous effect and the light grey lines show the local response estimated for each county. Responses are visually clipped to the range of flood depth values experienced in each county over the time of our study. The map shows the estimated local vulnerability to flooding (the slope of the response line). Results are shown for 8 different models, with performance evaluated with UAR cross validation indicated. Performance using other cross-validation strategies is given in Table S3.

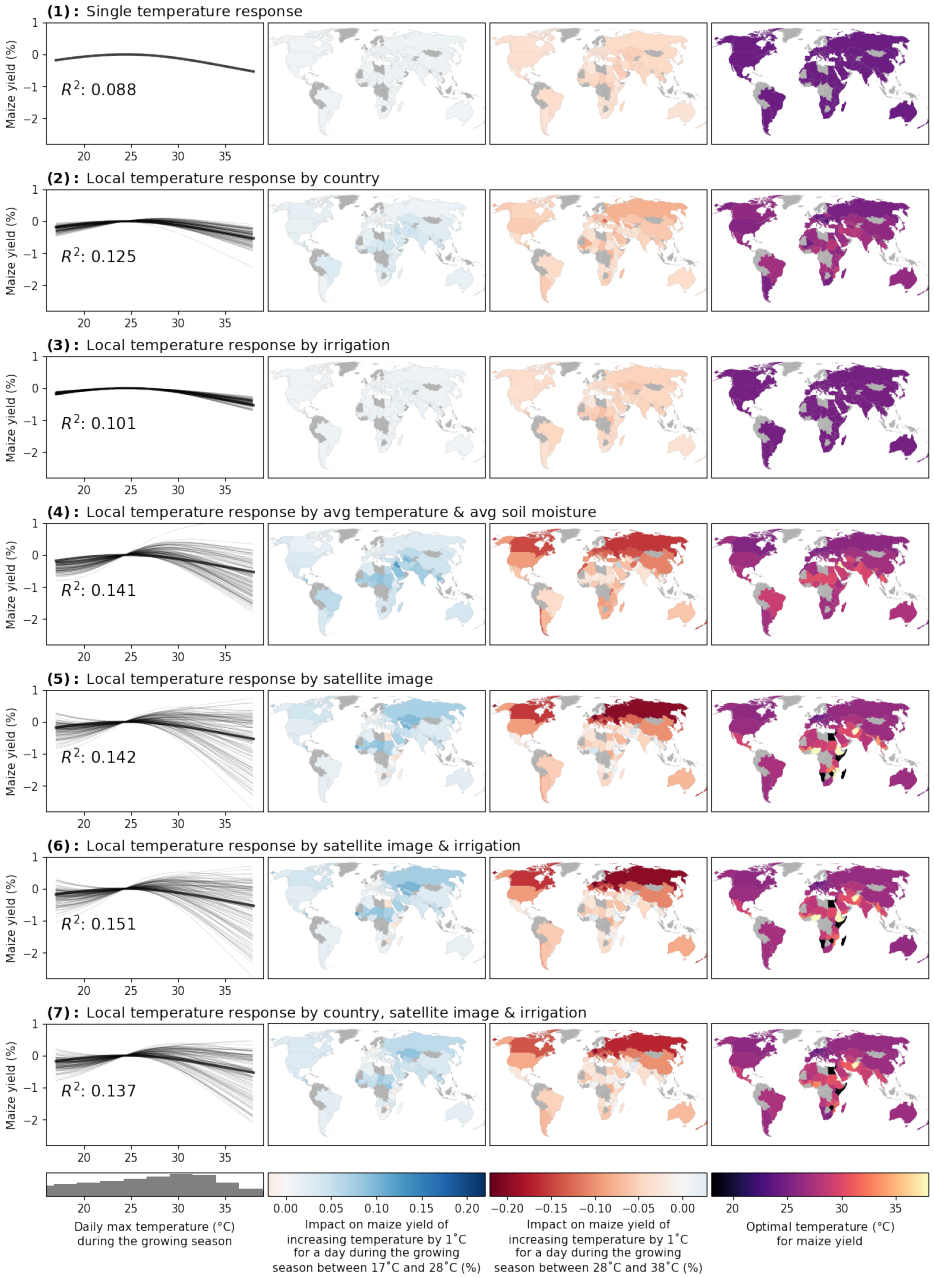


Figure S14: **Local vulnerability of global maize yields to temperature estimated using traditional and satellite-based approaches.** Response curves show the estimated influence of temperature on yields. The thick line shows the estimated homogeneous effect and the light grey lines show the local response estimated for each country. Responses are visually clipped to the range of daily temperature values experienced in each country over the time of our study. The map on the left shows benefits of warming (the slope of the response function to the left of 28°C) and the map in the middle shows damages from extreme heat (the slope to the right of 28°C). Darker blues indicate greater benefits of warming and darker reds indicate greater damages of extreme heat. The map on the right shows the estimated yield-maximizing temperature. Results are shown for 7 different models, with performance evaluated with UAR cross validation indicated in text. Performance using other cross-validation strategies is given in Table S4.

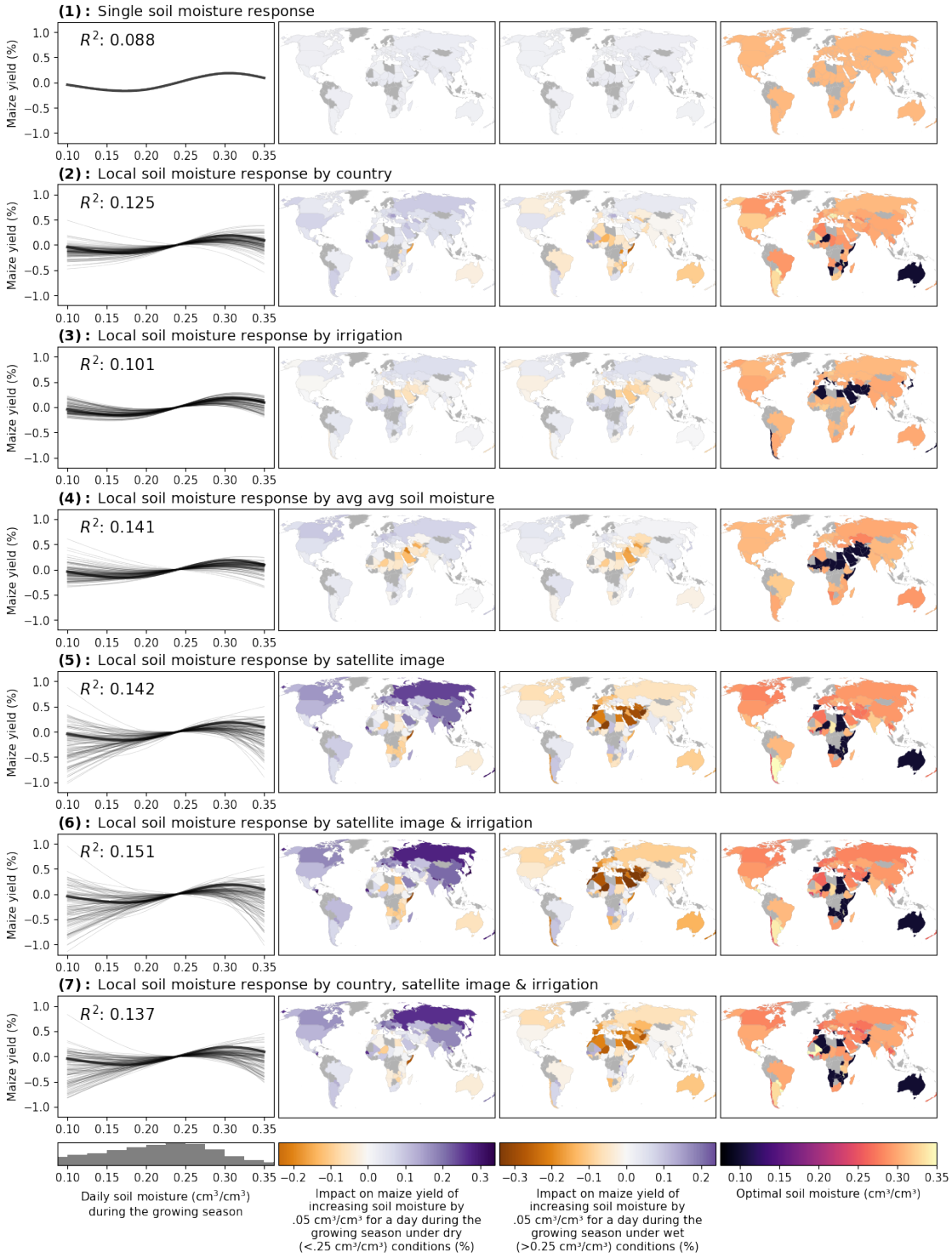


Figure S15: **Local vulnerability of global maize yields to soil moisture estimated using traditional and satellite-based approaches.** Like Fig. S14, but for soil moisture rather than temperature.

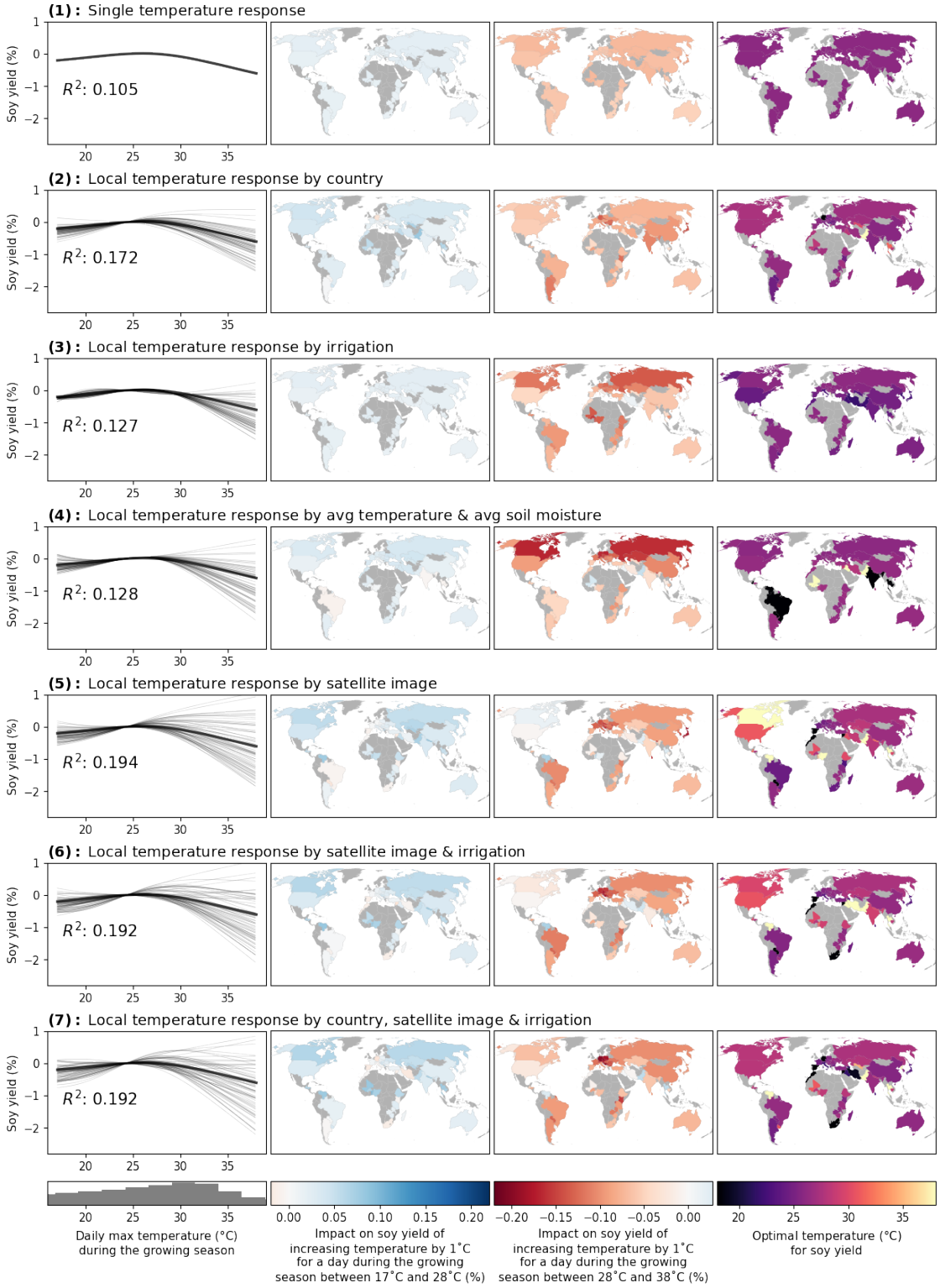


Figure S16: **Local vulnerability of global soy yields to temperature estimated using traditional and satellite-based approaches.** Like Fig. S14, but for soybeans rather than maize.

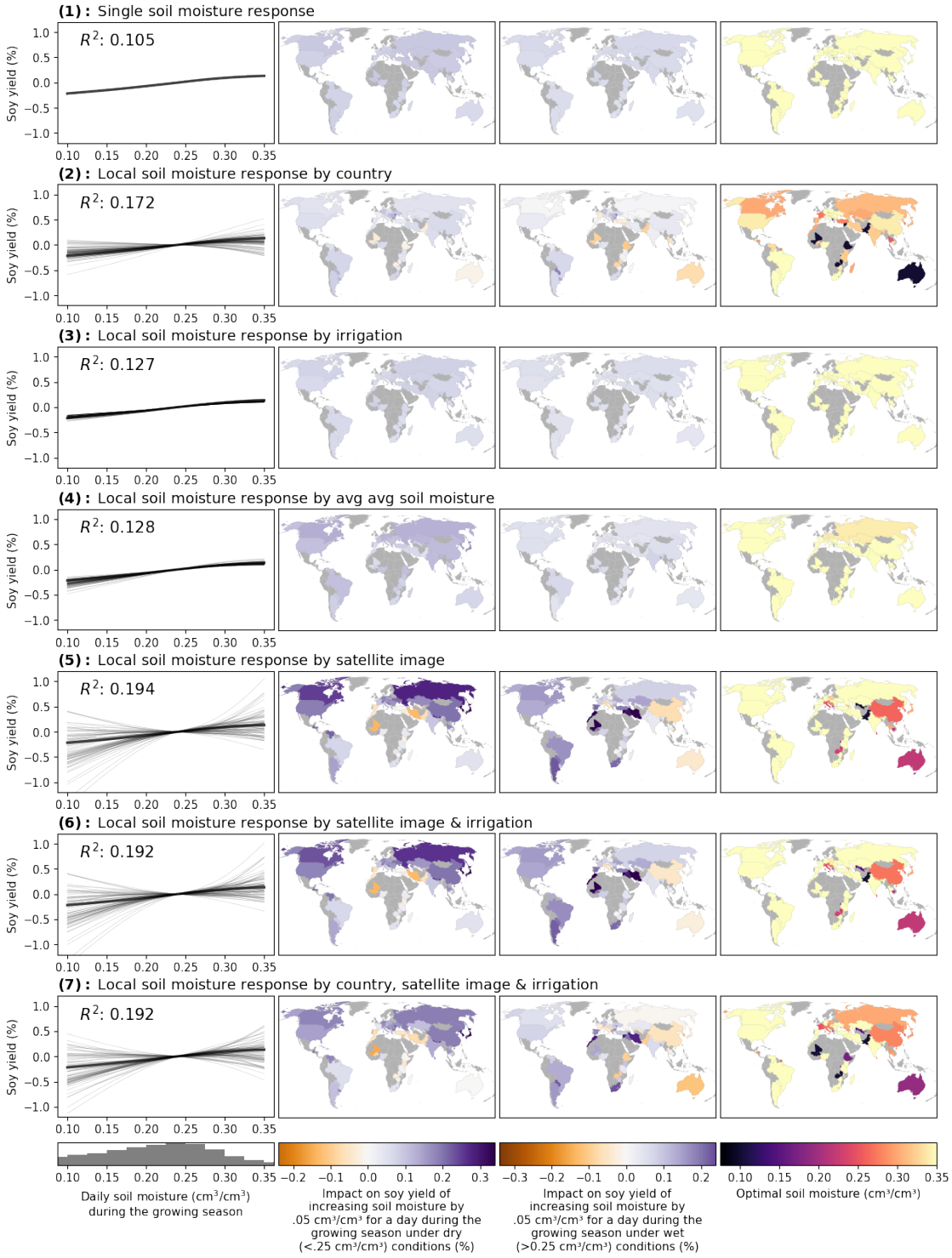


Figure S17: **Local vulnerability of global soybean yields to soil moisture estimated using traditional and satellite-based approaches.** Like Fig. S14, but for soybeans rather than maize and for soil moisture rather than temperature.

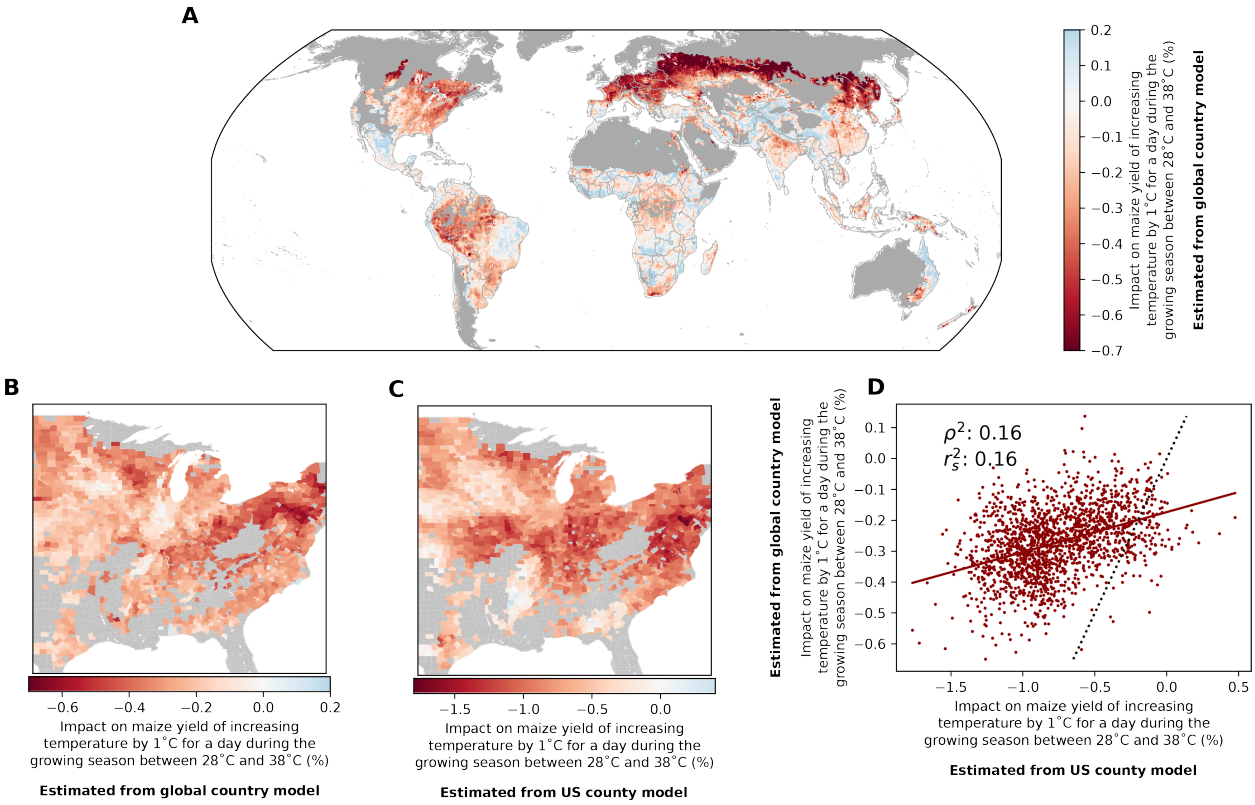


Figure S18: **Evaluating predictions of local vulnerability at resolution finer than the data on which it was trained.** A) Predicted sensitivity of maize yields to extreme heat globally at 10km^2 resolution. B) Predicted sensitivity at county resolution using a model trained on national yield observations globally. C) Predicted sensitivity at county resolution using a model trained on county yield observations. D) a scatter of the values from B) and C). The solid black line shows the line of best fit and the dotted line shows the 1:1 line.

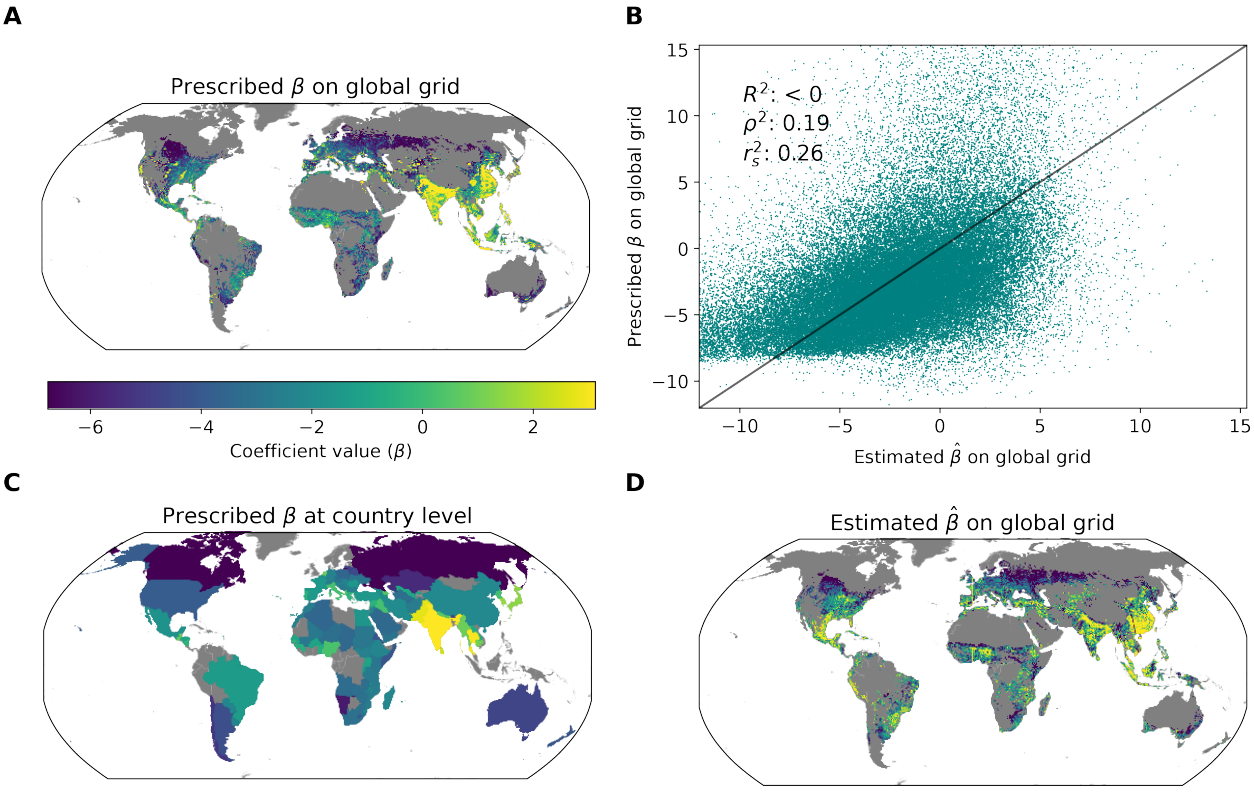


Figure S19: **Simulated data experiment to test the ability of the approach to make predictions of vulnerability at finer resolution than the model was trained.** Panel A shows a map of prescribed vulnerability at national resolution used when training the model. Panels C and D show maps of prescribed vulnerability and estimated vulnerability at fine resolution. The scatter in panel D compares these two.

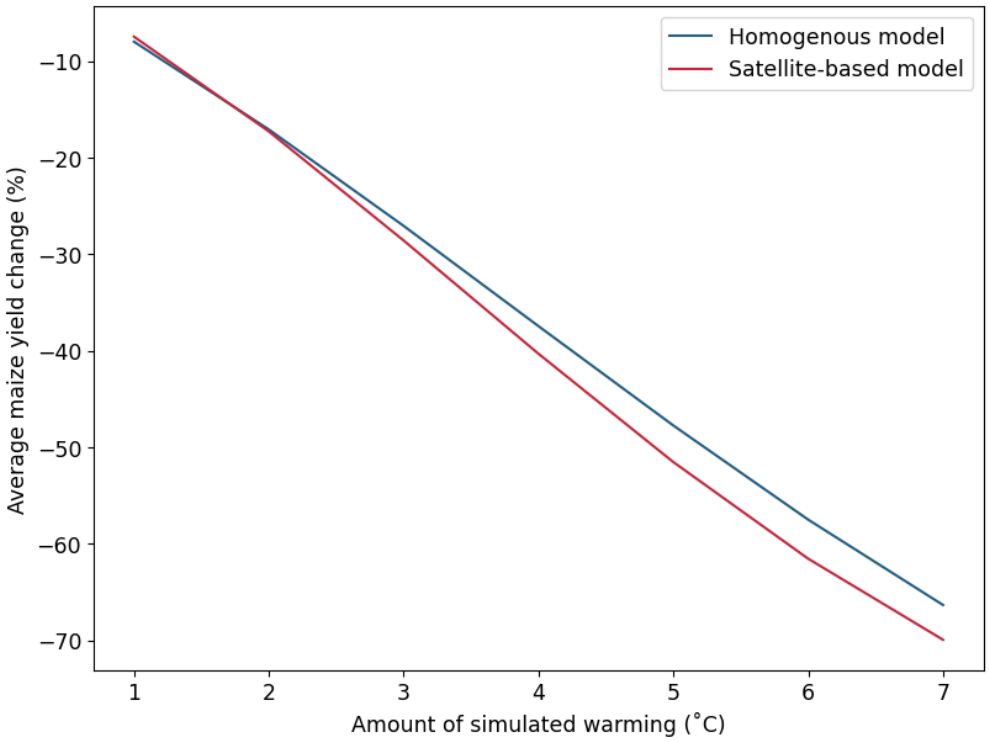


Figure S20: **Projected average yield damages from increased growing season temperature over U.S. croplands East of the 100th meridian.** A stylized climate change simulation where growing season temperature values are increased by one to seven degrees each day of the growing season. The y-axis shows average yield losses due to this warming.

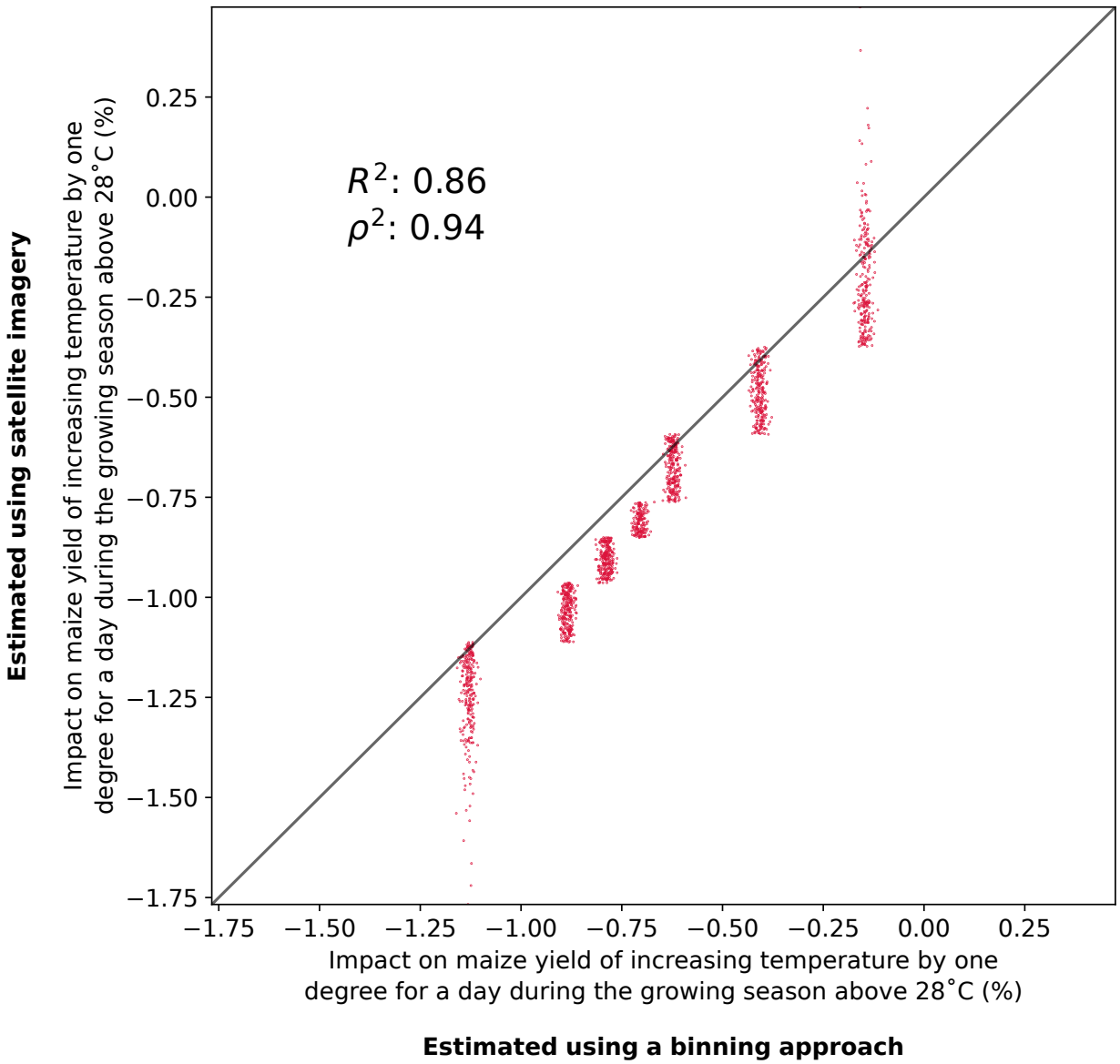


Figure S21: **Comparison of local maize vulnerability to extreme heat estimated in two different ways.** Each dot shows the estimated vulnerability of U.S. county-level maize yields to extreme heat. The value on the y-axis is the local vulnerability estimated using the satellite-based approach. The value on the x-axis is the local vulnerability estimated by (1) binning counties into seven bins according to their satellite-estimated local vulnerability and then (2) re-estimating vulnerability to extreme heat using a standard OLS regression in each bin with no ridge penalty or satellite features.

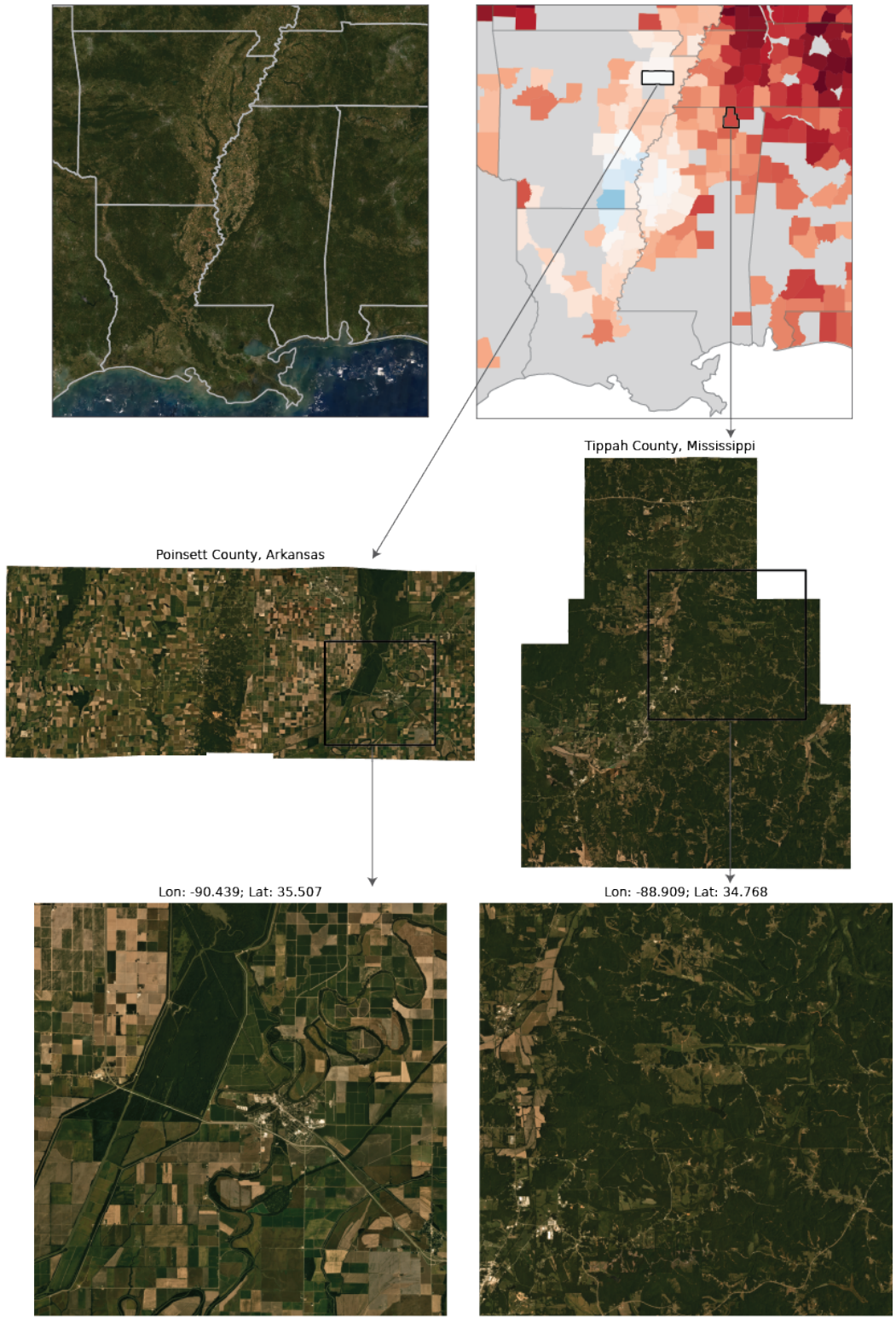


Figure S22: **Imagery and predicted climate vulnerability for maize croplands surrounding the Mississippi river.**

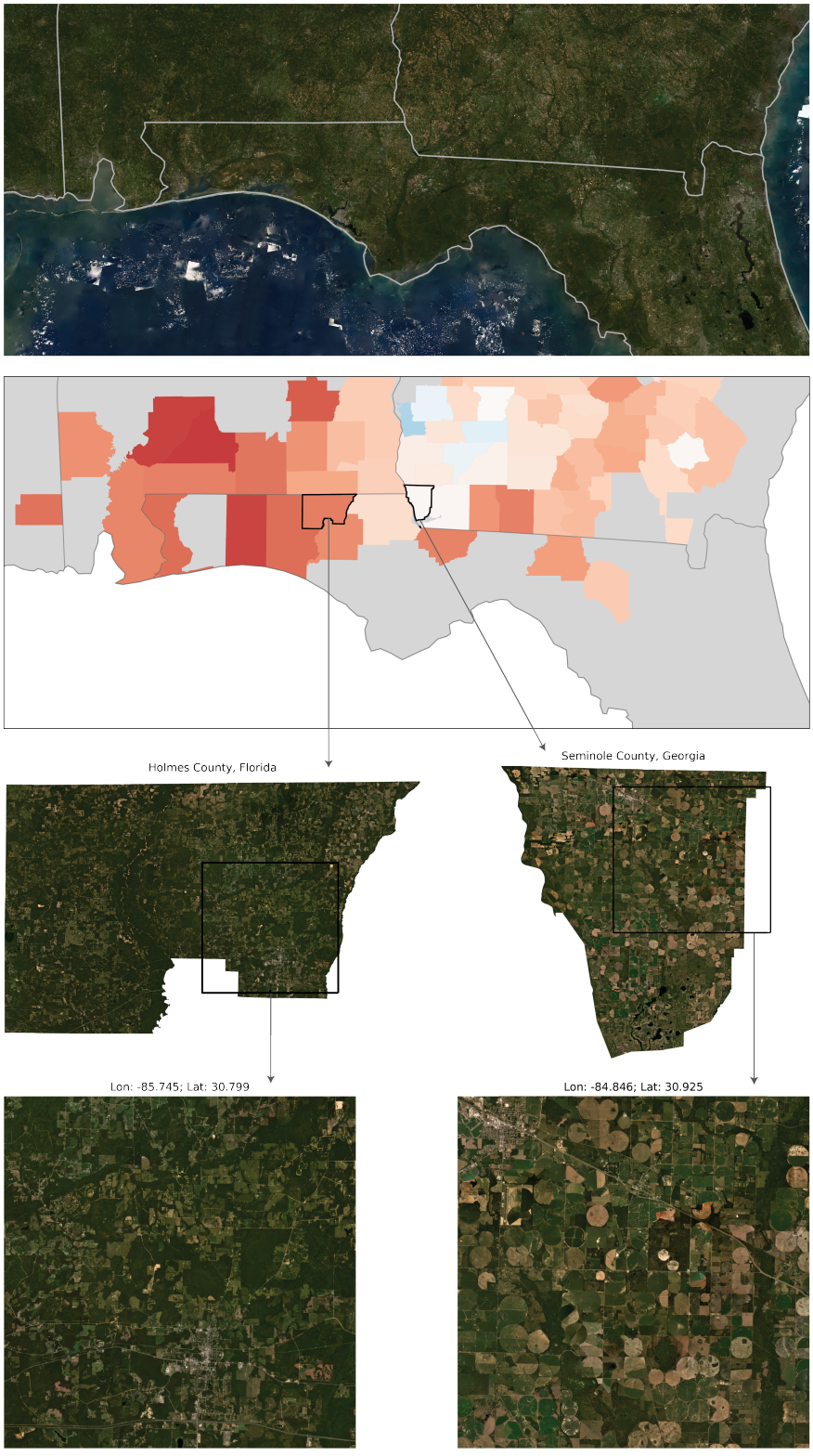


Figure S23: **Imagery and predicted climate vulnerability for maize croplands surrounding the Floridian Aquifer.**

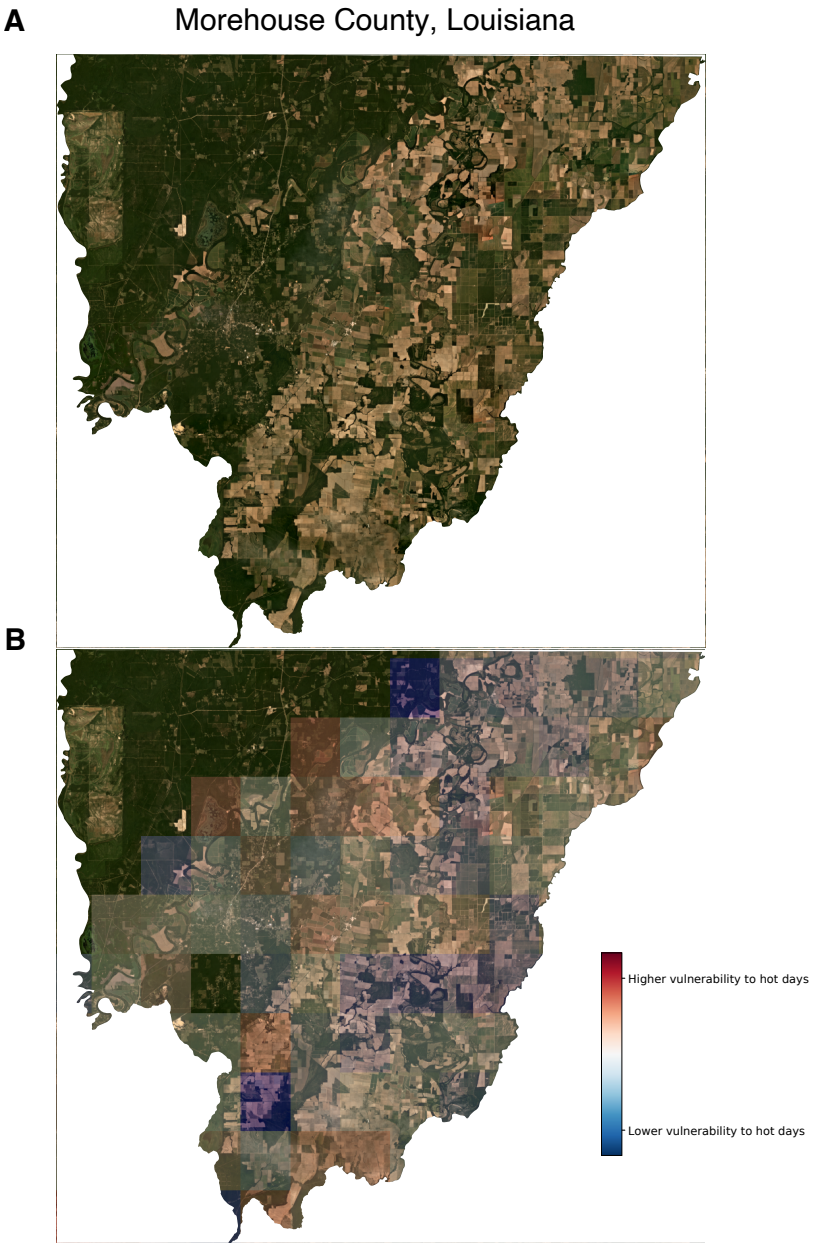


Figure S24: **Imagery and predicted maize climate vulnerability for Morehouse County, Louisiana.** Panel A shows Planet imagery. Panel B shows the same imagery with predictions of climate vulnerability overlaid. Lighter purples indicate greater resilience to extreme heat and darker reds indicate greater vulnerability to extreme heat.

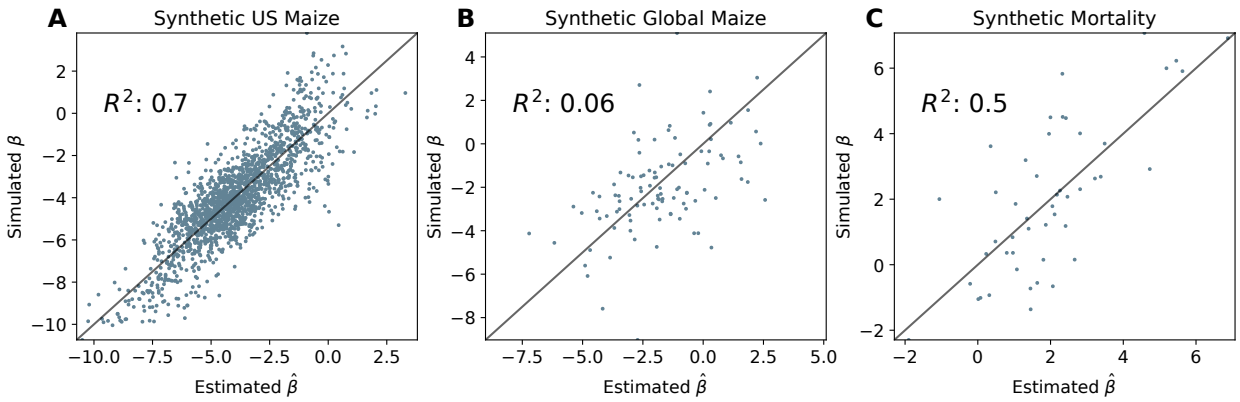


Figure S25: **Plot of simulated β against estimated $\hat{\beta}$ in the simulation experiments.** Each point represents a county.

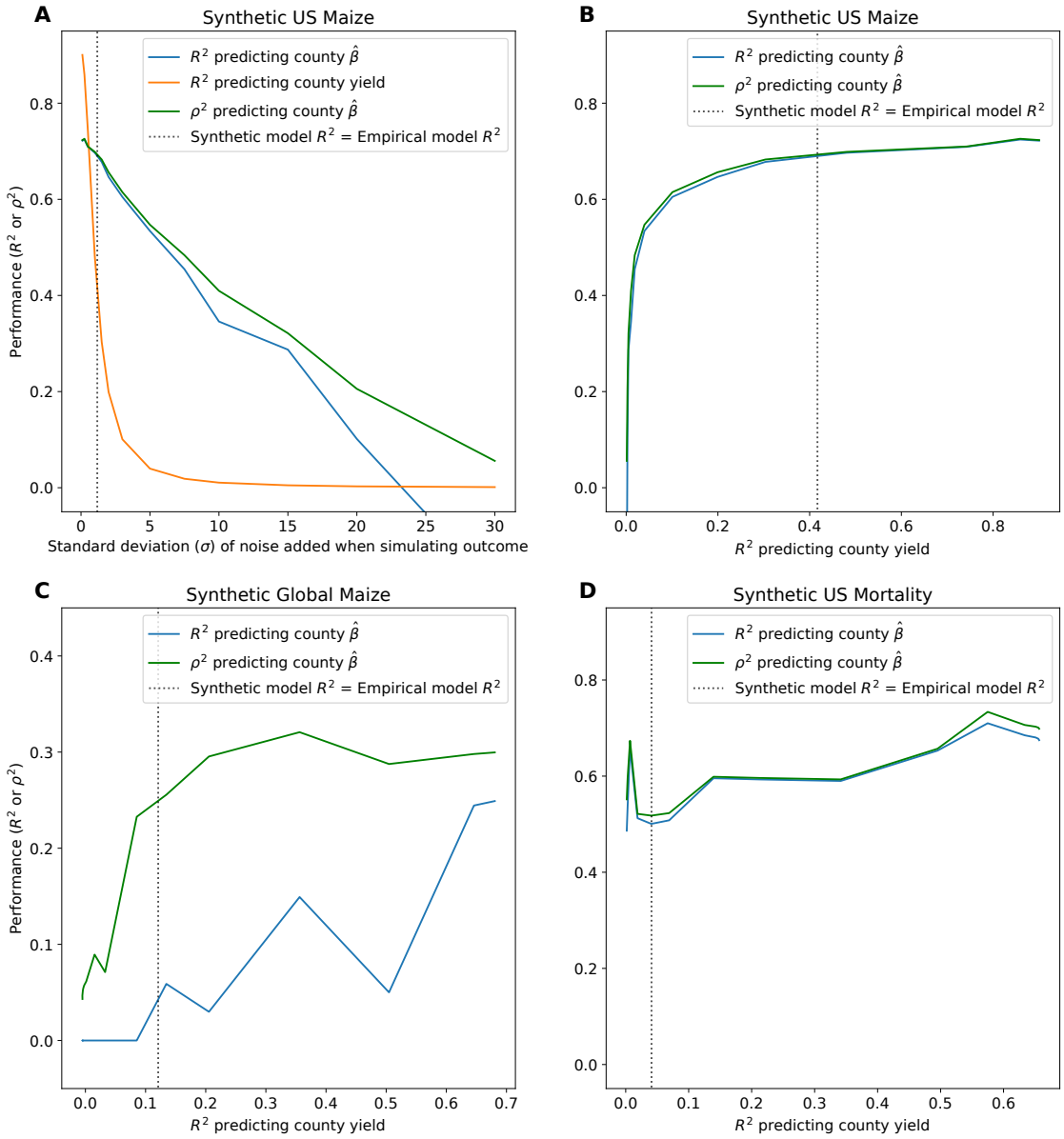


Figure S26: **Model performance estimating local vulnerability, β, as a function of model performance predicting the outcome variable, Y.** (**A**) Results from a synthetic data experiment for U.S. maize yield. As more Gaussian noise is added to simulated yield, model performance predicting simulated yield declines quickly. However, the performance predicting vulnerability declines more slowly. (**B**) The same information as **A** but presented so that the performance predicting vulnerability is a function of model R^2 predicting the outcome. (**C**) The same as **B** but for synthetic global maize yield. (**D**) The same as **B** but for synthetic U.S. human mortality rate.

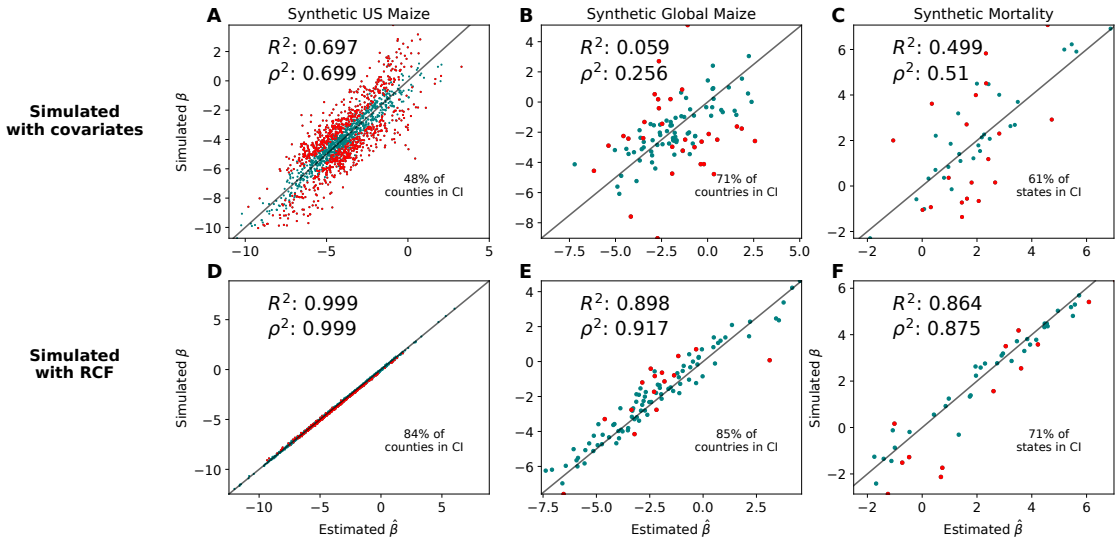


Figure S27: **The ability to recover simulated climate vulnerability (β) for US maize yields, global maize yields, and US mortality rates.** In the first row, vulnerability is set as a function of unobserved covariates. In the second row, vulnerability is generated from the image features, meaning that the prescribed determinants of vulnerability are fully represented in the imagery. Vulnerability for units in red fall outside the estimated confidence interval for the local satellite-based estimate while units colored green fall within the estimated confidence interval.