# No Effect of China's National Forest City Program on Urban Heat or Forest Cover: A Comment on Ma et al. (2026)

Matthew Turner*

August 2026

**Abstract**

Ma et al. (2026) use a staggered adoption difference-in-differences design to argue that China's National Forest City Program (NFCP) reduces the annual number of extreme heat days by approximately three per year in designated cities. We reassess this finding using MODIS satellite land surface temperature, NDVI, and tree cover data; Landsat-derived land cover data; and Global Historical Climatology Network daily temperature records. We apply three heterogeneity robust staggered adoption estimators and find no statistically or economically significant cooling effect of NFCP designation. We also find that three satellite based vegetation measures show no post-designation increase. Indeed, Landsat-derived CLCD forest cover declines marginally near city centers. Thus, NFCP designation does not lead to a decrease in temperature nor to the increase in tree cover that could cause such a temperature decrease.

**Keywords:** National Forest City Program; urban heat island; difference-in-differences; staggered adoption; MODIS; land surface temperature; NDVI; forest cover

*Brown University, Department of Economics, Box B, Brown University, Providence, RI 02912. email: Matthew_Turner@Brown.edu. Also affiliated with PERC, IGC, NBER, PSTC.

# 1 Introduction

Ma et al. (2026) use a staggered difference-in-differences design on city-year panel data from 243 Chinese prefecture-level cities (2000–2023) to argue that designation under China's National Forest City Program (NFCP) reduces the annual count of extreme heat days ($\geq$ 35°C) by approximately three per year.

There are four reasons to regard this conclusion with skepticism. First, Ma et al. give no evidence that the NFCP increased urban green cover. If NFCP designation does not increase vegetation or tree cover, then the mechanism through which the program could reduce heat does not operate. Second, their main event-study plot is noisy: a straight line fits within the confidence intervals of all pre- and post-designation coefficients. Thus, while their estimates do not reject a treatment effect, they also do not appear to reject a pre-existing trend difference and no NFCP effect. Third, the main estimates use a TWFE regression known to be biased under heterogeneous treatment effects (Goodman-Bacon, 2021) — the bias that the Borusyak et al. (2024) estimator, used in their own event study, corrects. Fourth, the main regression conditions on a long list of covariates, per capita GDP among them. Because per capita GDP is plausibly correlated with the propensity to apply for NFCP designation and temperature trends (e.g. from industrial or commercial development) treatment effect estimates may be confounded.

We therefore repeat the Ma et al. (2026) evaluation of the NFCP using the BJS imputation estimator with covariate adjustment (Borusyak et al., 2024), the Callaway–Sant'Anna estimator (Callaway and Sant'Anna, 2021), and its doubly robust variant (Sant'Anna and Zhao, 2020).[1] We make two main findings. First, there is no evidence that the NFCP increased forest cover. Indeed, satellite-measured vegetation near city centers appears to decline after NFCP designation. Second, there is no statistically or economically significant temperature decrease in any specification that passes pre-trend tests. Repeating these estimations at the weather station-year level (rather city-year) corroborates these results and

[1] BJS without covariate adjustment does not satisfy the parallel pre-trends condition and we exclude it.

allows us to consider exactly the same hot days outcome as Ma et al. (2026), rather the satellite derived measure of summer temperature that we use in our main results.

Figure 1 illustrates the central finding. The four panels show BJS with controls event studies using outcomes averaged over 20 km radius disks centered on the city center. This geography is similar to the one on which Ma et al. (2026) relies. The top-left panel (summer LST) shows point estimates that scatter around zero from the start of the pre-period through $t = 10$, with confidence intervals that contain zero throughout. The remaining three panels show no coherent post-designation change in MODIS NDVI, MODIS tree cover fraction, or Landsat-derived CLCD forest cover. These plots support the claim that all of the outcomes were noisy and not much affected by NFCP designation, one way or the other. A possible exception is the CLCD based forest cover plot in the bottom right. Although noisy, pre-trends look parallel and there is a weak decline in tree cover post treatment.

Section 2 discusses our concerns about the identification strategy in Ma et al. (2026). Section 3 describes our data and estimators. Section 4 reports results. Section 5 discusses the sources of discrepancy. Section 6 concludes.

## 2 Comment on Identification Strategy

The main estimates in Ma et al. (2026) use a standard OLS-TWFE regression with a treated $\times$ post indicator. Goodman-Bacon (2021) shows that the TWFE coefficient in a staggered adoption design decomposes into a weighted average of $2\times2$ DiD comparisons, some of which use already-treated units as controls for later cohorts. When treatment effects are heterogeneous or grow over time, these "forbidden comparisons" can bias the estimate in magnitude or sign. Borusyak et al. (2024) develop an imputation estimator that avoids forbidden comparisons by fitting TWFE on untreated observations only and imputing counterfactuals for treated ones. Ma et al. (2026) apply this estimator for their event study but not their main results table. Applying the BJS estimator throughout, as we do below, the temperature

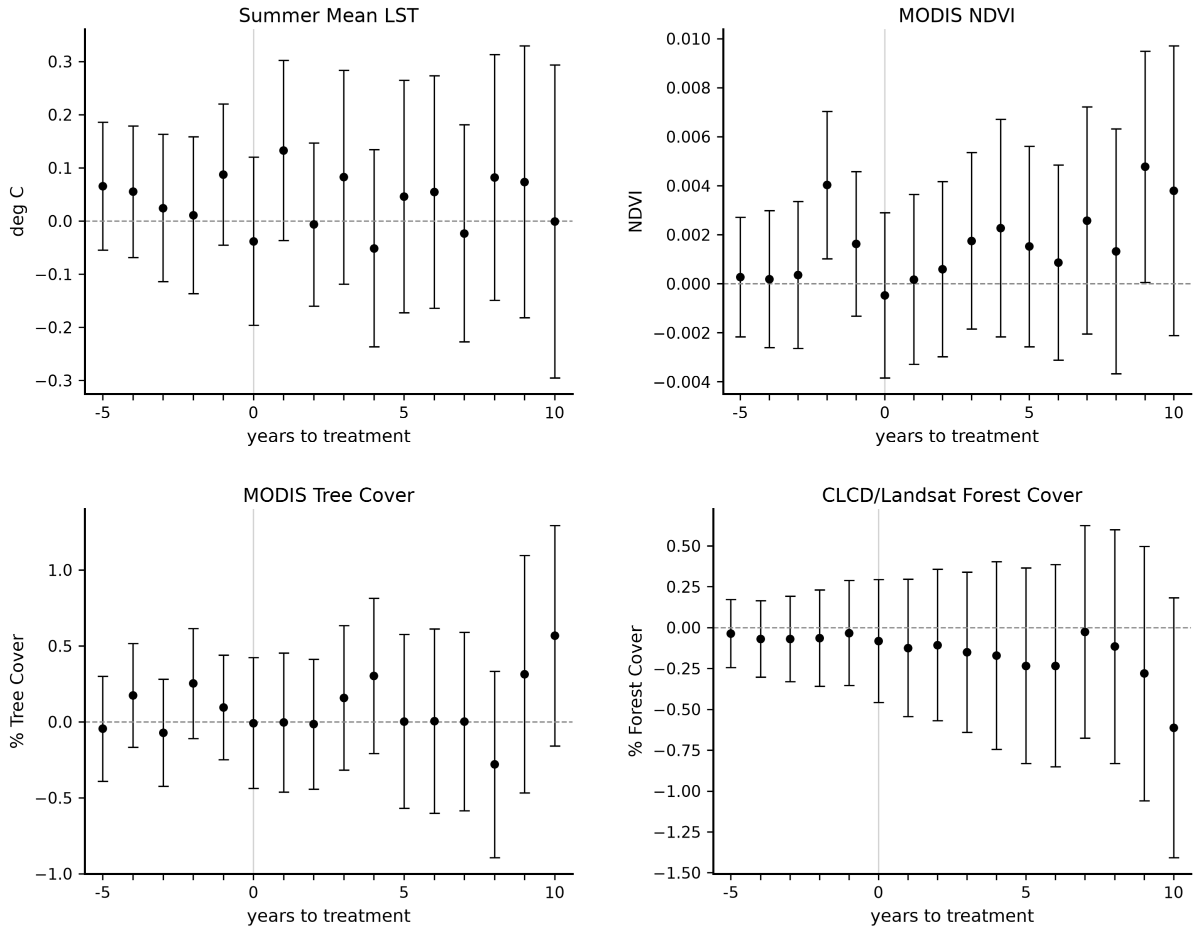


Figure 1: BJS with controls event-study coefficients for 20 km buffer. Each panel plots estimated treatment effects of NFCP designation on a satellite-derived outcome, averaged over pixels within 20 km of the city centroid. The estimator is BJS imputation (Borusyak et al., 2024) with covariate × year interactions for city latitude, longitude, and log pre-treatment population. The horizontal axis shows years to treatment ($t = 0$). Dots are point estimates; vertical whiskers show the 95% confidence interval clustered at the city level. *Top left*: summer mean LST (deg C). *Top right*: MODIS NDVI. *Bottom left*: MODIS tree cover (% tree cover). *Bottom right*: % forest cover from the China Land Cover Dataset (CLCD).

result disappears.

The main regression in Ma et al. (2026) conditions on a set of time-varying city-level controls, including per capita GDP. Per capita GDP is plausibly correlated with both the propensity to apply for NFCP designation and with temperature trends independent of the program. Therefore, the effects of this control may be confounded with the effects of the treatment.

In a difference-in-differences design, the identifying assumption is parallel trends: absent designation, treated and control cities would have followed the same outcome path. Researchers include covariates to restore conditional parallel trends when the unconditional version fails (or to improve precision when it holds). We find that BJS without covariate adjustment shows pre-designation LST coefficients that diverge from zero, indicating that NFCP and control cities are on different pre-designation temperature trajectories. Conditioning on pre-determined city characteristics (latitude, longitude, and log pre-treatment population) absorbs this imbalance.

## 3 Data and Methods

*Data.* We draw our city-level panel from the United Nations World Urbanization Prospects (WUP 2018) dataset (United Nations, Department of Economic and Social Affairs, Population Division, 2018), which covers urban agglomerations with populations of at least 300,000. WUP centroids anchor all satellite data extractions. Treatment is a binary indicator equal to one for city $i$ in all years at or after its first NFCP designation, constructed from annual lists published by the National Forestry and Grassland Administration. The panel covers 275 prefecture-level cities observed from 2000 to 2025 (7,150 city-years); 149 hold NFCP designation across 17 cohorts (2004–2022) and 126 remain never-treated. We extract all outcomes at buffer radii of 5, 10, and 20 km from the city centroid. Main estimates use the 20 km buffer; 5 km results are reported in Table 1 and in Appendix A1.

China has approximately 337 prefecture-level administrative units. Our sample of 275 is smaller for two reasons. First, the WUP includes only agglomerations exceeding its 300,000-resident threshold; smaller prefecture-level units fall below it. Second, we exclude county-level cities that appear in the WUP because the NFCP programme targets prefecture-level units exclusively.

To measure temperature, we use MODIS MOD11A2 8-day composite LST at 1 km resolution (Wan et al., 2021) (downloaded, NASA EOSDIS, June 2026.). For each city-year we average June–August pixels within each city-year buffer around the city center, excluding fill values. Note that the deviation from Ma et al. (2026). The outcome variable in Ma et al. (2026) is a weather station derived measure of hot days per year as opposed to our satellite based measure of average summer temperature. We later verify our results using weather station based outcomes to match Ma et al. (2026) more closely.

We consider three satellite based measures of urban vegetation and tree cover. First, we use MODIS MOD13A3 monthly NDVI composites at 1 km resolution (Didan, 2021)(downloaded, NASA EOSDIS, June 2026). We compute mean July NDVI for each city-year; July is the peak growing-season month. Second, the MODIS Vegetation Continuous Fields product (MOD44B) provides annual tree canopy cover at 250 m resolution (DiMiceli et al., 2021) (downloaded, NASA EOSDIS, June 2026). We compute mean canopy fraction within each city-year buffer. Finally, the CLCD provides annual Landsat-based land cover classifications at 30 m resolution (Yang and Huang, 2021). We use the extended archive (Zenodo record 18180184, 1985–2025, downloaded June 2026) and compute the forest pixel share within each buffer. At 30 m the CLCD resolves land cover changes not visible in MODIS.

The panel is largely complete. Among the 7,150 city-years, MODIS LST at 5 km has 286 missing observations (4.0%), concentrated in persistently cloudy cities; LST at 20 km, NDVI, and tree cover have fewer than 30 missing observations combined. CLCD is complete for all 7,150 city-years.

*Methods.* The *BJS imputation estimator* (Borusyak et al., 2024) fits TWFE on untreated observations (never- and not-yet-treated units), imputes counterfactual outcomes for treated units, and identifies treatment effects as observed minus imputed. This avoids forbidden comparisons by construction. BJS without covariate adjustment fails the parallel pre-trends condition (pre-designation LST coefficients diverge from zero); we exclude it from Table 1.

*BJS with controls* augments the first stage TWFE with covariate $\times$ time interactions for city latitude, longitude, and log pre-treatment population, allowing cities in different settings or of different sizes to follow different time trends. These covariates are pre-determined and time-invariant. Because they are time-invariant, entering them in levels would be collinear with city fixed effects; we therefore interact each with five-year period-bin dummies, creating regressors that vary within city over time. The *Callaway–Sant'Anna estimator* (CS; Callaway and Sant'Anna, 2021) compares each treatment cohort to never-treated cities at each event-time horizon without covariate adjustment. The *doubly robust Callaway–Sant'Anna estimator* (DR; Sant'Anna and Zhao, 2020) is a variant of CS that allows for covariates. DR conditions on the same pre-determined variables (latitude, longitude, and log pre-treatment population), entering them in levels. For each estimator we aggregate to an average post-period ATT as the simple mean of event-study coefficients at $t = 0, 1, \ldots, 10$, with the mean of period-level clustered standard errors as a conservative SE approximation.[2]

The code that organized the data and performed the estimations in this comment was written by a Claude agent. To validate the data processing pipeline, we first confirmed that all outcome values lay in plausible ranges. Second, for each raw data layer, we constructed a synthetic layer for which the final city-year level outcomes were known and verified that these values obtained. To validate the estimation routines, we first conducted all estimations twice, once using python estimation libraries, and once using stata estimation routines. We next gave a separate Claude agent, with no knowledge of the project, the final data set and a draft of this comment, and confirmed that the agent could reproduce results from these sources alone.

[2] The period-level standard errors for BJS with controls specifications (columns 1–2) are obtained from a stage-3 OLS of $\hat{\tau}_{it}$ on event-time dummies with standard errors clustered at the city level. This treats the imputed counterfactuals as known rather than estimated, and therefore does not implement the efficient influence-function variance derived in Borusyak et al. (2024) (Section 4 of the published version). Because all BJS-with-controls temperature estimates are far from any conventional significance threshold ($|z| \leq 0.3$), this limitation does not affect any substantive conclusion. Standard errors for DR (column 3) and CS (column 4) are from Stata's `csdid`, which implements the influence-function variance of Sant'Anna and Zhao (2020) and Callaway and Sant'Anna (2021) respectively.

Table 1: Average post-period treatment effects ($t = 0$ to $t = 10$). Each cell is the simple mean of event-study coefficients at $t = 0, 1, \ldots, 10$. Standard errors in parentheses are the mean of period-level clustered standard errors, a conservative approximation of the aggregate SE. Columns: (1) BJS imputation with covariate adjustment (BJS with controls) at 20 km; (2) BJS with controls at 5 km; (3) doubly robust Callaway–Sant'Anna estimator (DR) at 20 km; (4) Callaway–Sant'Anna without covariate adjustment (CS) at 20 km.

| | (1) BJS w/ctrl, 20 km | (2) BJS w/ctrl, 5 km | (3) DR 20 km | (4) CS 20 km |
|---|---|---|---|---|
| LST (°C) | +0.032 | +0.016 | +0.016 | +0.067 |
| | (0.106) | (0.121) | (0.097) | (0.095) |
| NDVI (index) | +0.002 | −0.001 | −0.000 | +0.000 |
| | (0.002) | (0.003) | (0.002) | (0.002) |
| Tree cover (%) | +0.095 | +0.047 | −0.103 | −0.165 |
| | (0.287) | (0.380) | (0.256) | (0.263) |
| CLCD forest (pp) | −0.194 | −0.585* | −0.089 | −0.291 |
| | (0.301) | (0.319) | (0.219) | (0.212) |

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered at the city level. Significance stars applied to $z = |\text{ATT}|/\text{SE}$ where SE is the mean of period-level clustered SEs.

# 4 Results

Table 1 reports average post-period ATTs. The Ma et al. (2026) temperature result does not appear. BJS with controls gives +0.032°C ($SE = 0.106$) at 20 km and +0.016°C ($SE = 0.121$) at 5 km, both indistinguishable from zero. DR gives +0.016°C ($SE = 0.097$) and CS gives +0.067°C ($SE = 0.095$). No temperature estimate is statistically significant.

MODIS NDVI, tree cover, and CLCD forest cover show no post-designation increase at any buffer radius. In fact, CLCD forest share declines near city centers. BJS with controls at 5 km gives −0.585 pp ($SE = 0.319$, $p < 0.10$); CS and DR estimates are statistically null but uniformly negative. These results indicate not merely that the NFCP fails to increase forest cover, but that cities entering the program experience vegetation losses near their centers.

Figure 2 confirms the null for the DR estimator and shows that pre-period LST coefficients are individually near zero and insignificant, validating the conditional parallel trends assumption underlying the covariate adjustment in BJS with controls. As for figure 1, the

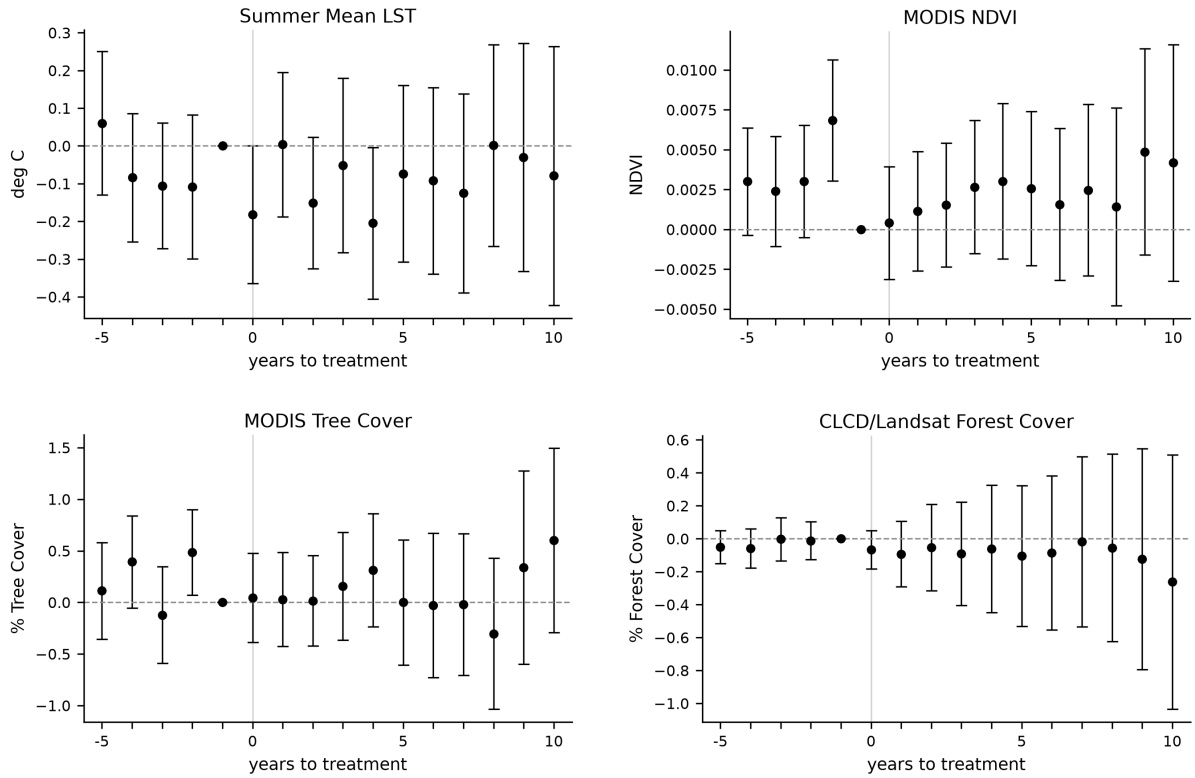


Figure 2: Doubly robust Callaway–Sant'Anna (DR) event studies at 20 km conditioning on pre-determined city characteristics (latitude, longitude, and log pre-treatment population), entered in levels. The mean post-period LST ATT is +0.016°C ($SE = 0.097$). Pre-designation coefficients are near zero and individually insignificant, confirming that the conditional parallel trends assumption holds under covariate adjustment. The horizontal axis shows years to treatment ($t = 0$ is the designation year, $t = -1$ is the reference period). Dots are point estimates; vertical whiskers show the 95% confidence interval from Stata's `csdid` (Callaway and Sant'Anna, 2021; Sant'Anna and Zhao, 2020). Panel outcomes are identical to Figure 1.

obvious conclusion to draw from figure 2 is that NFCP designation is not important for any of the outcomes, although this conclusion is tempered by the probable pre-trends violation for the MODIS NDVI outcome (upper right) and the suggestion of decline in CLCD measured forest cover post-treatment.

We supplement our satellite based city-year level analysis with a weather station-year based analysis. We use GHCN-Daily records (Menne et al., 2012), the same data source as Ma et al. (2026). We retain stations within 20 km of a city centroid with at least 30 valid observations per year and average outcomes where multiple stations fall within this buffer.

The station sample covers 52 treated cities (16 cohorts) and 47 never-treated cities.[3] For each city-year we compute summer mean daily maximum temperature (Tmax, June–August) and the count of days with Tmax $\geq$ 35°C. We also extract MODIS NDVI and tree cover for pixel closest to each station.

Figure 3 shows Callaway–Sant'Anna event studies for four station-level outcomes. Post-designation NDVI estimates scatter around zero. The station Tmax ATT is similar. The bottom row shows tree cover (bottom left) and annual hot-day counts (bottom right). Tree cover likely fails pre-trend tests: tree cover pre-designation estimates at $t = -2$ through $t = -4$ are 2.9–4.0 pp above zero ($z = 3.6$–$4.7$). As for our city-year satellite based event studies, the station based event studies suggest that NFCP was not important for the evolution of temperature or forest cover. Also similar to our other results, this conclusion is tempered by the likely pre-trend violation for the MODIS tree cover data.

# 5 Discussion

NFCP designation has no temperature effect under any estimator that satisfies the parallel pre-trends condition. BJS with controls gives +0.032°C ($SE = 0.106$) at 20 km, DR gives +0.016°C ($SE = 0.097$), and CS gives +0.067°C ($SE = 0.095$). Across all three specifications, the temperature estimates are statistically insignificant and positive. The Ma et al. (2026) temperature result does not appear in our analyses.

Covariate adjustment is motivated by the pre-trend failure of BJS without controls: NFCP and control cities are on different pre-designation temperature trajectories. Predetermined city characteristics (latitude, longitude, log pre-treatment population) absorb this imbalance. DR's pre-period coefficients (Figure 2) are near zero and individually insignificant, confirming that the conditional parallel trends assumption holds under covariate adjustment.

[3] With only 47 never-treated cities across 16 treatment cohorts, we report CS without covariate adjustment. Covariate-adjusted station estimates (DR, not-yet-treated control group) show comparable or larger positive ATTs for tree cover and hot days.

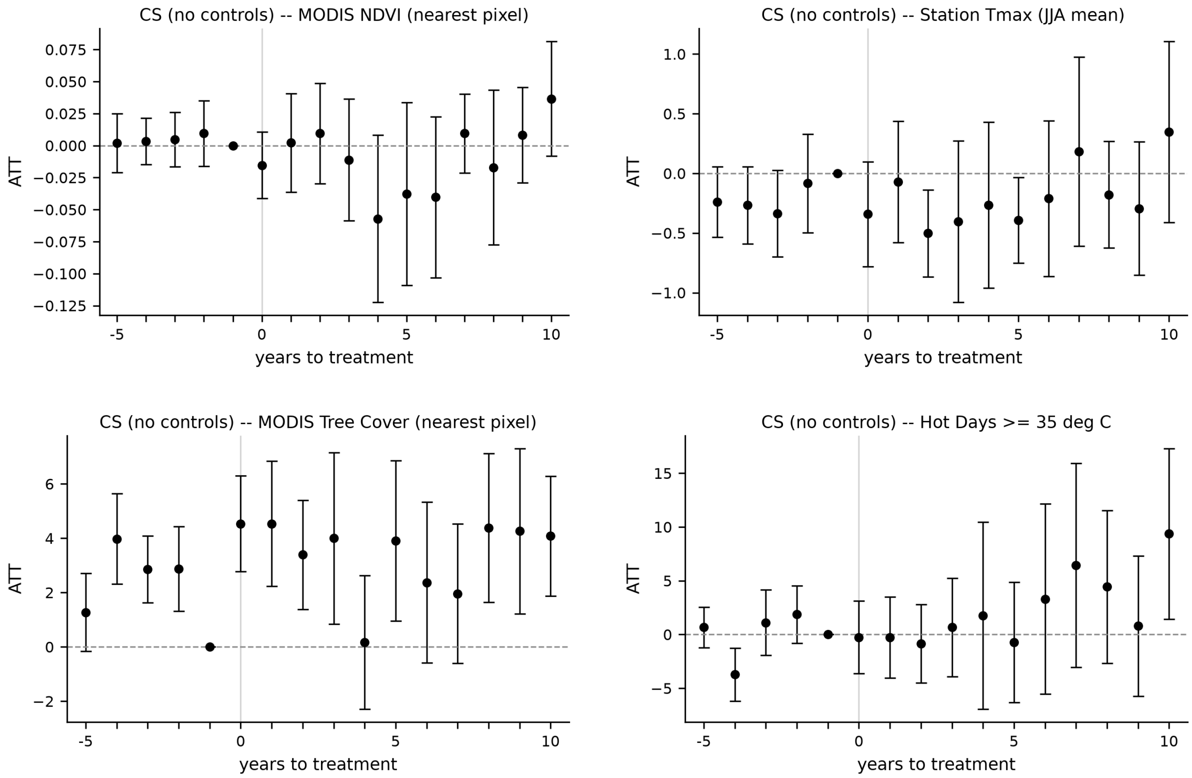


Figure 3: Station-level Callaway–Sant'Anna event studies (CS, never-treated controls, no covariate adjustment). The horizontal axis shows years to treatment ($t = 0$ is the designation year, $t = -1$ is the reference period). Dots show point estimates; vertical whiskers show 95% confidence intervals clustered at the city level. *Top left*: MODIS NDVI at the nearest pixel to each weather station, from MOD13A3 (Didan, 2021). *Top right*: summer mean daily maximum temperature (Tmax, June–August), from GHCN-Daily (Menne et al., 2012). *Bottom left*: MODIS tree cover fraction at the nearest pixel, from MOD44B (DiMiceli et al., 2021); pre-designation estimates at $t = -2$ through $t = -4$ are 2.9–4.0 pp above zero ($z = 3.6$–$4.7$). *Bottom right*: annual hot-day count (daily Tmax $\geq 35$°C), from GHCN-Daily.

The vegetation results add an independent argument against the proposed mechanism. MODIS NDVI, MODIS tree cover, and CLCD forest cover show no post-designation increase at any buffer radius. CLCD forest share declines marginally at 5 km: BJS with controls gives $-0.585$ pp ($p < 0.10$, column 2). Without a greening signal at any radius, there is no plausible mechanism by which the NFCP could reduce urban heat.

# 6 Conclusion

The main result in Ma et al. (2026) does not survive scrutiny. Applying BJS with controls, DR, and CS, we find no evidence that NFCP designation reduces summer LST, air temperature, or extreme heat days. Temperature estimates under all three specifications are small, positive, and statistically insignificant. Three independent satellite vegetation measures show no post-designation increase in forest cover at any scale, and CLCD forest cover declines marginally at 5 km ($p < 0.10$). Cities entering the NFCP experience vegetation losses near their cores. Station NDVI corroborates the vegetation null. Overall, our results suggest that NFCP designation did not have a measurable effect on either tree cover or temperature for treated cities.

# References

Kirill Borusyak, Xavier Jaravel, and Jann Spiess. Revisiting event-study designs: Robust and efficient estimation. *Review of Economic Studies*, 91(6):3253–3285, 2024. doi: 10.1093/restud/rdae007.

Brantly Callaway and Pedro H. C. Sant'Anna. Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2):200–230, 2021. doi: 10.1016/j.jeconom.2020.12.001.

Kamel Didan. MOD13A3 MODIS/Terra Vegetation Indices Monthly L3 Global 1km SIN Grid V061, 2021.

Compton DiMiceli, Mark Carroll, Robert Sohlberg, Do Hyung Kim, Maggi Kelly, and John R. G. Townshend. MOD44B MODIS/Terra Vegetation Continuous Fields Yearly L3 Global 250m SIN Grid V061, 2021.

Andrew Goodman-Bacon. Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2):254–277, 2021. doi: 10.1016/j.jeconom.2021.03.014.

Yun Ma, Yan Geng, and Shuai Zhong. Urban forest development and extreme heat mitigation: The climate adaptation effects of China's National Forest City Policy. *Forests*, 17 (1):79, 2026. doi: 10.3390/f17010079.

Matthew J. Menne, Imke Durre, Russell S. Vose, Byron E. Gleason, and Tamara G. Houston. An overview of the Global Historical Climatology Network-Daily database. *Journal of Atmospheric and Oceanic Technology*, 29(7):897–910, 2012. doi: 10.1175/JTECH-D-11-00103.1.

Pedro H. C. Sant'Anna and Jun Zhao. Doubly robust difference-in-differences estimators. *Journal of Econometrics*, 219(1):101–122, 2020. doi: 10.1016/j.jeconom.2020.06.003.

United Nations, Department of Economic and Social Affairs, Population Division. World urbanization prospects: The 2018 revision, file 22 – population of urban agglomerations with 300,000 inhabitants or more. Technical report, United Nations, New York, 2018.

Zhengming Wan, Simon Hook, and Glynn Hulley. MOD11A2 MODIS/Terra Land Surface Temperature/Emissivity 8-Day L3 Global 1km SIN Grid V061, 2021.

Jie Yang and Xin Huang. The 30 m annual land cover dataset and its dynamics in China from 1990 to 2019. *Earth System Science Data*, 13(8):3907–3925, 2021. doi: 10.5194/essd-13-3907-2021.

# Appendix

## A1. Buffer Robustness

Figure A1 reports BJS with controls event-study coefficients at the 5 km buffer for satellite LST, NDVI, and MODIS tree cover. LST at 5 km is +0.016°C ($SE = 0.121$), consistent with the 20 km null in Figure 1. NDVI and tree cover are likewise null.

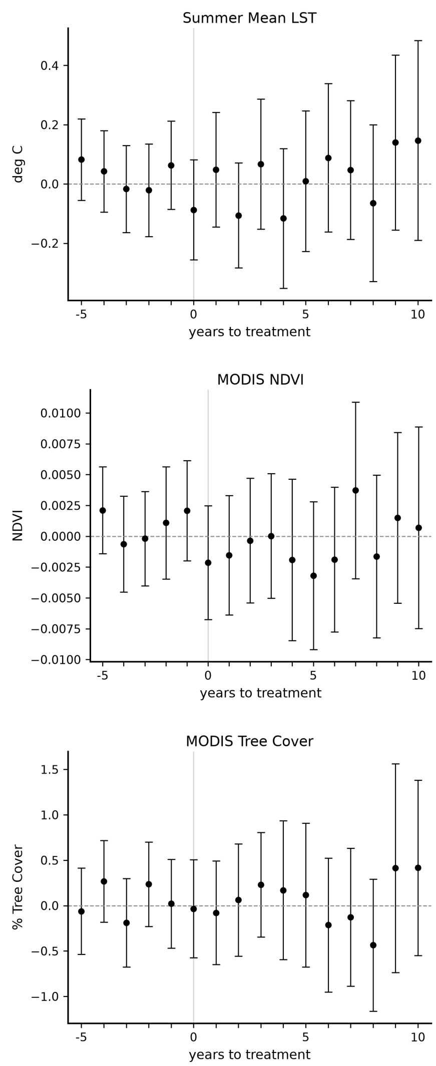


Figure A1: BJS with controls event-study coefficients at 5 km for satellite LST (top), MODIS NDVI (middle), and MODIS tree cover (bottom). The horizontal axis shows years to treatment ($t = 0$ is the designation year, $t = -1$ is the reference period). Dots are point estimates; vertical whiskers show the 95% confidence interval clustered at the city level.