Mind the Gap: Improving Water Markets with Field-Level Remote Sensing*

Katherine Wright† Andrew Ayres‡ Bryan Leonard§

PRELIMINARY DRAFT. PLEASE DO NOT CIRCULATE WITHOUT PERMISSION.

Abstract

Alternative Transfer Mechanisms (ATMs) are temporary arrangements between farmers and municipalities that are designed to support urban water demand while avoiding permanent transfers of water out of agricultural communities. These arrangements entail incomplete contracting over land use (e.g., fallowing) rather than a transfer of water rights, thereby avoiding many of the transaction costs associated with water markets. The success of ATMs hinges on the ability to enforce contracts and accurately measure water savings from changes in land use. Yet, existing programs use a spatially coarse back-of-the-envelope approach to estimate water savings. This paper combines high-resolution satellite estimates of evapotranspiration with administrative data to estimate the causal effect of rotational fallowing on water use at the parcel level. We study the PVID-MWD Fallowing and Forbearance Program, the largest and longest-running ATM, which transfers water from rural California to the largest municipal water provider in the Western U.S. We find that actual water savings from 2005 to 2021 were 29% lower than reported and that "fallowing" only reduces water use on treated parcels by 55% on average. We asses the extent to which these differences are driven by selection into the program, non-additionality, and residual water use on fallowed parcels and find evidence that non-additionality and hydrologic spillovers each account for 51% of the shortfall.

*For helpful comments we thank Josh Abbott, Anna Boser, Tamma Carleton, Eric Edwards, Robert Heilmayr, Kelsey Jack, Kailin Kroetz, Morgan Levy, Buzz Thompson, and Will Rafey, as well as seminar participants at the Workshop on Advancing Sustainable Water Management at Stanford University, the Bren School of Environmental Science and Management at UC Santa Barbara, the Environmental and Resource Economics Series at UC San Diego, the Land, Environment, Economics and Policy Institute at the University of Exeter, the Economics Department at the Colorado School of Mines, the Department of Agricultural Economics & Economics at Montana State University, and the UC Riverside Water Dialogues. Talia Greco, Danielle Moon, and Mithran Mathiraj provided excellent research assistance. We have also benefited greatly from data sharing and discussions with staff at Metropolitan Water Distrct.

†Hillsdale College kwright1@hillsdale.edu

‡University of Nevada, Reno andrew.ayres@unr.edu

§University of Wyoming bryan.leonard@uwyo.edu

1 Introduction

Many regions worldwide are already experiencing long-run declines in surface water availability due to climate change (Rodell et al., 2018). In arid regions, climate change is predicted to increase both water scarcity and the variability of freshwater supplies due to changes in precipitation, snow pack, and the timing of runoff (Elliott et al., 2014; Gordon et al., 2024), threatening the viability of irrigated agriculture in many of these regions. The Western United States is a prime example—the Colorado River supplies water for 40 million people and irrigates more than five million acres of crop land annually, and demands on the river exceed supply for 16 of the years between 2000 and 2020 (Richter et al., 2024).

In arid regions like the Western United States, ensuring that water is allocated to its highest valued use is critical for reducing the societal costs of climate change and increasing scarcity of water supplies more broadly. Recent evidence suggests that the current distribution of water use is far from efficient. Currently, agriculture accounts for as much as 80% of freshwater use throughout the Colorado River Basin and in California in particular (Richter et al., 2020; Medellín-Azuara et al., 2024), despite the fact that the value of water for urban uses often far exceeds the market value of water in agriculture (Brewer et al., 2008). Within agriculture, there is mounting evidence that the observed distribution of water use is inefficient, especially during times of drought and curtailment (Arellano-Gonzalez et al., 2021; Rafey, 2025).

Scholars have suggested that more active water markets could help allocate water more efficiently, reducing the costs of meeting new environmental and urban demands and grappling with dwindling supplies (Brewer et al., 2008; Bruno and Jessoe, 2021; Rafey, 2023, 2025). More recently, markets have emerged as a critical component of climate adaptation, particularly in arid regions facing water stress (Bruno and Jessoe, 2024). The expansion of water markets has been hindered by a variety of transaction costs related to unclear property rights, administrative approvals, physical conveyance, and objections from third parties and agricultural communities (Bretsen and Hill, 2008; Leonard et al., 2019; Hagerty, 2023). Many of these costs arise from the fundamental difficulty of estimating field-scale consumptive water use to determine how much of their legal diversion right a farmer is entitled to transfer without impairing other users or the environment.

Due to these barriers to formal transfers of water rights, various buyers of water including municipal water utilities, NGOs, and the U.S. Bureau of Reclamation increasingly utilize "alternative transfer mechanisms" (ATMs) to secure additional water supplies on a temporary basis. ATMs vary in their design, but typically involve payments to farmers in exchange for specific practices—usually crop switching or fallowing—with the understanding that water savings are made available to the buyer. Examples include the Upper Colorado River Basin's System Conservation Pilot Program, the Lower Colorado River Basin's drought contingency actions with the Central Arizona Water Conservation District, the Nature Conservancy's Georgia Flow Incentive Trust, and efforts by the Great Salt Lake Commissioner's office to lease water from Utah farmers.

ATMs avoid the nexus of transaction costs associated with traditional water markets because, technically, no formal water rights are exchanged. Rather, parties come to an agreement about an short-run change in *practices* and approximate the implied water savings, usually with simple back-of-the-envelope calculations based on estimates of average water use by crop or by irrigation district. There has been limited peer-reviewed evaluation of ATMs' effectiveness, largely due to the same measurement challenges that have historically stymied the development of water markets.[1] Rigorously estimating the water savings from such a program requires field-scale estimates of consumptive water use before and after program implementation, as well as detailed administrative data on program participation that can be matched to water use estimates.

In this paper, we combine recently developed satellite data on evapotranspiration from OpenET Volk et al. (2024) with administrative data on fallowing "calls" to study the largest, longest-running ATM in the world: the PVID/MWD Forbearance and Fallowing Program. The Program, which began in 2005 and runs through 2040, is an arrangement between the Metropolitan Water District (the largest conglomerate of municipal water utilities in the Western United States) and the Palo Verde Irrigation District, which holds some California's most senior rights to Colorado River Water. Farmers are paid on a per-acre basis to fallow up to 1/3 of their land when MWD issues a "call," and the resulting savings are made available to MWD the following year.

We use remote sensing data from OpenET to develop monthly, field-scale estimates of consumptive water use over 2000 to 2021 and combine these with GIS data on irrigated lands and administrative records of monthly field-level fallowing calls, which we hand-digitized. We estimate average monthly water savings from fallowing in PVID using modern difference-in-difference techniques (De Chaisemartin and d'Haultfoeuille, 2020). We aggregate these monthly, field-scale estimates into implied annual water savings due to the program and compare to MWD's own internal estimates and explore the mechanisms that explain why our causal, field-level estimates differ from MWD's estimates, including field-level heterogeneity in water use, residual water use on fallowed lands, and non-additionality of some fallowed fields.

2 Results

2.1 Average Water Savings from Fallowing

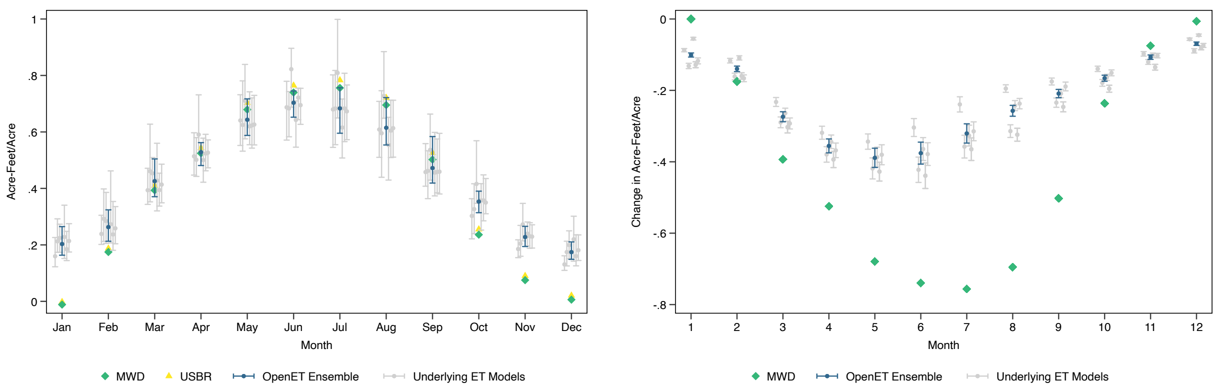
Our chosen ET metric from OpenET is their Ensemble Mean, which is an average of six distinct underlying satellite-derived ET estimates that provides the most reliable metric by smoothing out outliers in any of the constituent models (Volk et al., 2024). Using the 30-meter resolution ET estimates yields the ability to measure use at the field level, but we begin by characterizing per-acre water use for Palo Verde Irrigation District (PVID) as a whole—the level at which the U.S. Bureau of Reclamation (USBR) and the Metropolitan Water District (MWD) measure and track consumptive water use in PVID.

Panel (a) of Figure 1 depicts average monthly water use and 95% confidence intervals over 2000 to 2021 as measured by the Ensemble (blue), MWD (green), USBR (yellow), and the six underlying

[1]The main exception is Harris (2024), who studies the Central Arizona Water Conservation District's fallowing program. Wong et al. (2021) uses ET to evaluation agricultural efficiency enhancements, but not fallowing per se.

ET models from OpenET (grey). Broadly, there is close alignment between the ET estimates and the administrative records, which is consistent with previous studies that have used ET to quantify water use in PVID (Senay et al., 2013; Senay, 2018; Wobus et al., 2024). Regressing MWD's average water use estimates on the Ensemble produces a slope (forced through the origin) of 0.987 with a r-squared of 0.926 and a root mean squared error of 0.134, with similarly tight alignment across the individual ET Models (Table S1). Field-level estimates of water use based on the Ensemble also align with expectations for crop-specific water demand (Fig. S1). The close alignment between raw ET and traditional measures of consumptive water use may be somewhat unique to our setting, where effective precipitation typically accounts for less than 2% of total ET (Fig. S2a and S2b) and is inversely correlated with the irrigation season (Fig. S2c).

Figure 1: Water Use and Water Savings



(a) Consumptive Use

(b) Estimated Effect of Fallowing

Notes: Average field-scale monthly water use (panel a) and estimated changes in water use on fallowed parcels (panel b). In panel a, we depict average monthly ET (and its 95% confidence interval) from OpenET's Ensemble model in blue, with the six individual models underlying the ensemble in grey. For comparison, we include MWD's estimate of district-wide average monthly water use in green and USBR's district-wide average in yellow. In panel b, we depict the regression coefficients from our difference-in-difference model that estimates the change in ET due to fallowing on "called" parcels for each month (along with the 95% confidence interval), with estimates with the preferred OpenET Ensemble in blue and estimates from the component models in grey. In green, we also depict the average of MWD's estimate of monthly savings, which is equivalent to their estimate of average monthly water use per acre.

Panel (b) of Figure 1 depicts several sets of estimates of the per-acre changes in water use associated with a fallowing "call" in PVID. The blue coefficients represent our causal estimates of the per-acre changes in water use associated with a fallowing "call" using doubly-robust difference-in-difference techniques (De Chaisemartin and d'Haultfoeuille, 2020) to compare changes in Ensemble ET on called vs. uncalled fields over time (grey coefficients represent similar estimates with each of the six ET models making up the ensemble). Importantly, this difference-in-difference design accounts for any time-constant differences in ET between fields and any monthly factors that affect all parcels simultaneously, which should ameliorate any level differences in measures of consumptive use between the Ensemble and MWD's administrative records—while baseline levels of water use may differ somewhat in panel (a) of Figure 1, panel (b) focuses on *changes* that net out these differences. MWD's estimates based on district-wide average water use and averaged across 17 years of annual verification reports are depicted in green.

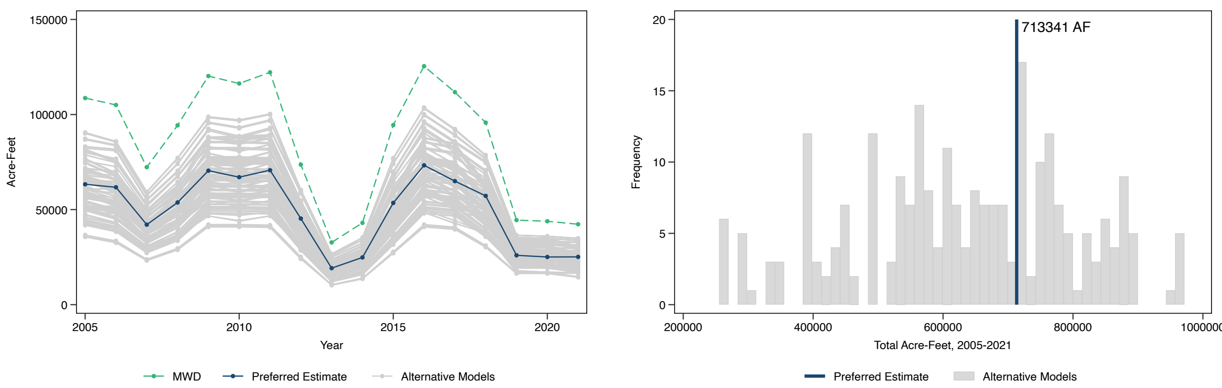
The field-scale causal estimates diverge significantly from MWD's district-wide approach in the majority of months. Farmers in PVID produce crops nearly year-round thanks to the warm, sunny conditions, but most irrigation occurs between May and and September. In these months, our causal estimates indicate that water savings are roughly 1/2 of what MWD and USBR estimate them to be. In contrast, we find that savings may be slightly larger than records suggest in the off-peak winter months (but the magnitudes are much smaller). Aggregating over the course of a year, the causal estimates suggest an average annual savings of 2.63 acre-feet per acre (AFA) with a 95% confidence interval of 2.49 to 2.75, which is 29% lower than the average of MWD's annual estimates over 2005 to 2021. We discuss examine this difference in more detail in next section.

The validity of our core finding—that realized water savings are substantially lower than the estimates produced by MWD and USBR—is supported by a series of diagnostic tests of the De Chaisemartin and d'Haultfoeuille (2020) estimator and a large battery of robustness tests across different sample restrictions, estimation strategies, and variable coding decisions (Materials and Methods). Placebo tests show limited evidence of adverse selection into the treatment group (Fig. S8) or differential trends on called vs. uncalled fields prior to treatment (Fig. S6) and the core finding is robust across roughly 250 different specifications that vary how treatment is coded, which fields are included in the sample, which ET model is used, and which difference-in-difference estimator is used (Materials and Methods; Fig. S7). Focusing on the valid doubly-robust DiD approach, estimates range from 1.58 to 3.03 AFA, representing an 18% to 58% reduction relative to administrative estimates.

2.2 Differences Relative to MWD "Historical Use Method"

To see how differences in average per-acre savings this affect aggregate water savings of the program, we multiply our monthly average water savings depicted in Figure 1 by the number of fallowed acres reported by MWD each month of the sample, and then aggregate to the annual level. Panel (a) of Figure 2 depicts these estimates in blue, with administrative estimates depicted in green and our 250 alternative specifications depicted in grey. Differences between our estimates and administrative records vary over time and scale roughly in proportion to the number of acres fallowed. Importantly, we find significantly lower water savings across all 250 specifications. Panel (b) of Figure 2 integrates under the curves depicted in Panel (a) to produce cumulative estimates of the implied differences in water savings over 2005–2021, with our preferred estimate indicated in blue and the other specifications indicated in grey. Cumulative differences range from 255,000 AF to 972,000 AF over the first 17 years of the program. Our preferred estimates imply that administrative estimates overstate cumulative program savings by 713,000 AF. For context, PVID as a whole typically consumes 350,000 to 420,000 AF per year.

Figure 2: Aggregate Water Use Implications



(a) Annualized Water Savings Estimates

(b) Cumulative Differences Relative to MWD

Notes: This figure depicts annual estimated water savings and cumulative differences between our estimates and MWD's. In panel a, we depict estimated annual water savings using our preferred methodology in blue, along with over 250 alternative estimates in grey that vary regression specifications, the ET models, sample restrictions, "call" variable definitions, and the procedure used to translate monthly field-scale coefficient estimates into annual district-wide savings. For comparison, we also plot MWD's annual water savings estimate in green. In panel b, we plot a histogram of the the cumulative difference between MWD's estimates (the green line in panel a) and each of our estimates, summed up over 2005 to 2011. The solid blue line represents our preferred estimate, which implies an aggregate difference of 509,940 acre-feet over 2005 to 2021.

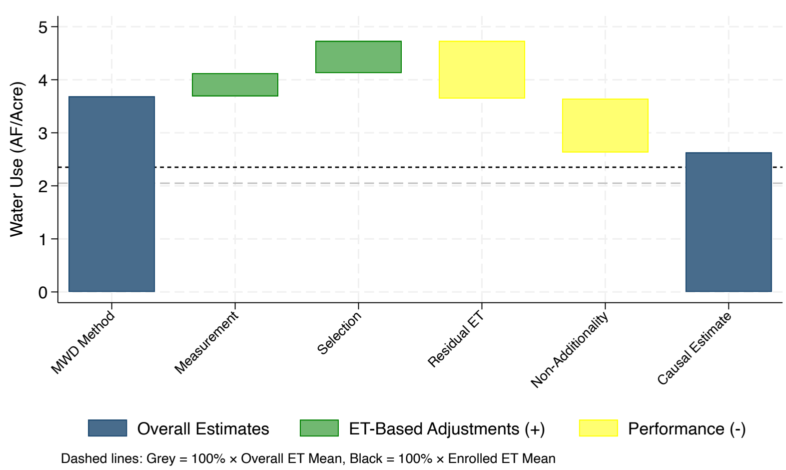
2.3 Assessing Mechanisms

We provide a detailed breakdown of the core differences in our methodology vs. MWD/USBR's approach in the Materials and Methods, but summarize the core intuition here. MWD/USBR's approach for estimating savings is to simply multiply district-wide average water use per acre by the number of acres called to fallow each month. This approach implicitly makes the following assumptions: i) average water use does not differ across enrolled vs. unenrolled fields, ii) fallowing is 100% additional (no fields fallowed in response to the call would have been fallowed anyway), and iii) water use goes to zero on fallowed lands. Our empirical approach is specifically designed to address all three of these concerns and to provide a causal estimate of how much water is truly saved by a "call." An additional source of divergence is the (small) level differences in water use estimates between the Ensemble and administrative records. Figure 3 provides a visual and mathematical decomposition of the overall difference into each of these factors.

Consistent with Figure 1, district-wide average water use is slightly higher than administrative records when measured with ET, suggesting savings could be as high as 4.12 AFA with perfect compliance and additionality. Moreover, high water-use fields differentially selected into the program (Fig. S8), implying that savings could have been as high as 4.73 AFA. However, residual water use on fallowed fields is still 1.09 AFA, reducing potential savings to 3.65 AFA. The remaining gap between this number and the causal estimate of 2.63 AFA is explained by non-additionality—fields that were fallowed as part of a "call" but would have been fallowed anyway. The dashed lines in Figure 3 help explain how much of the overall difference in estimates is explained by level differences vs. the water savings gap—the fact that ET only falls by about 55% on called fields. Applying this 55% reduction to MWD's raw annual use estimates or a re-scaled version would actually *reduce* estimated program savings further, given the direction of the measurement differences.

Figure 3 illustrates two striking facts about the program: many fallowed fields are non-additional, and water use on (verified) fallowed fields is far from zero (roughly 23% of baseline use on enrolled fields). Figure S11 provides evidence consistent with the non-additionality of many fallowed fields: enrolled fields that were more likely to have been fallowed prior to the start of the program are selected for fallowing more frequently on a rotational basis during the life of the program. Fields that were fallowed 60% of the time over 2000-2004 are roughly twice as likely to be selected for a "call" than fields with no history of fallowing. What is more striking is the finding that fallowed lands exhibit non-trivial ET Figure S9. This fact was previously noted by Wobus et al. (2024), who attributed it to errors in satellite classification of fallowed lands. Such misclassification is not an issue here, as MWD verifies that called field are in fact fallow. What, then, explains water use on these fields? Our final set of results answers this question.

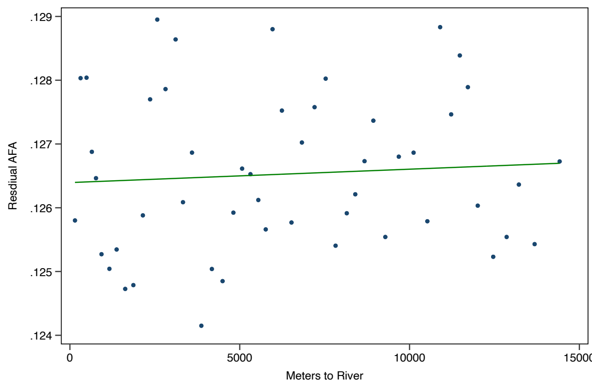
Figure 3: Decomposition of Estimated Gap in Water Savings



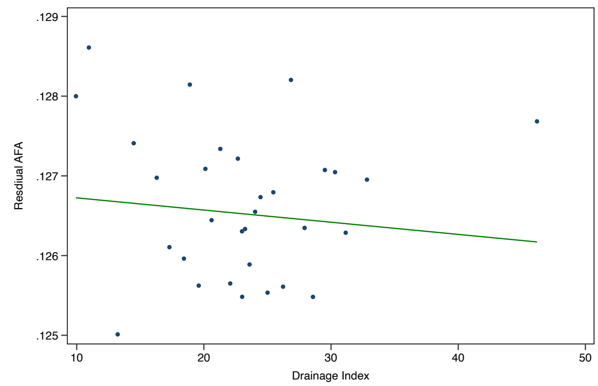
Component	Step Value	Balance	Interpretation
MWD Estimate	–	3.69	Administrative Benchmark
1. Measurement Effect	0.44	4.12	Adjustment to MWD Baseline
2. Enrollment Selection	0.61	4.73	Adjustment to Enrolled Baseline
3. Residual ET	-1.09	3.65	Residual use on fallowed land
4. Non-Additionality	-1.02	2.63	Causal counterfactual gap
Actual Savings	–	**2.63**	**Final Causal Estimate**
Comparative Estimates:			
Hybrid (Naive)	–	2.05	MWD Mean × 45% Causal δ
Hybrid (Rescaled)	–	2.35	(MWD × Ratio) × 45%

Notes: This figure depicts our decomposition of the differences between our causal estimate and the MWD/USBR verification methodology. We account for i) level differences in water use measurement between ET vs. administrative records, ii) compositional differences in average water use on enrolled vs. unenrolled files, and iii) residual water water use on fallowed fields. We attribute to remaining difference to non-additionality (Materials and Methods).

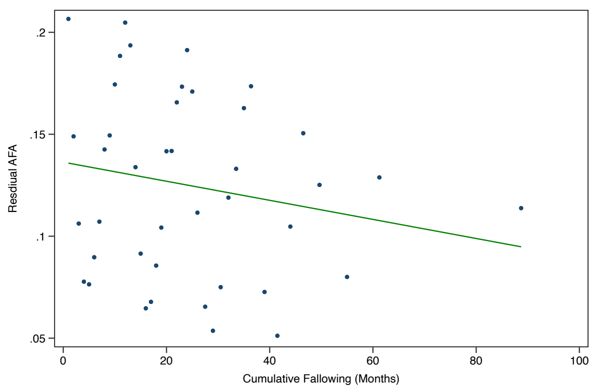
Figure 4: Explaining Residual Water Use



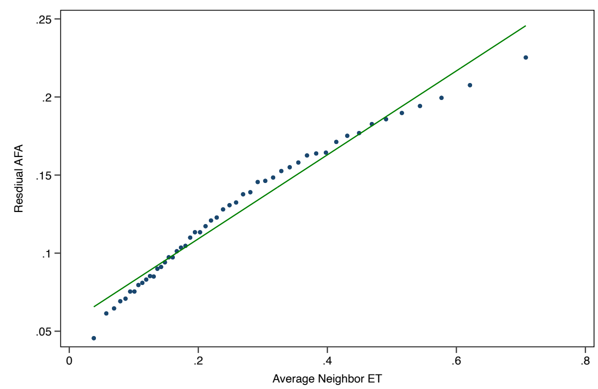
(a) Distance to Colorado River



(b) Soil Drainage



(c) Cumulative Fallowing



(d) Neighbors' Water Use

Notes: This figure depicts binned scatter plots of residualized (unexplained) ET on *called* fields against several variables thought to potentially explain this variation. In each panel, the underlying data have been binned into 50 quantiles of the y variable, and the mean of each quantile if plotted, along with the best linear fit through these means. Panels a and b residualize ET based on month-of-sample fixed effects and plot the variation in ET not explained by these fixed effects against distance to the Colorado River (panel a) and Schaetzl (1986)'s Soil Drainage Index (panel b). Panels c and d residualize ET based on both month-of-sample and field fixed effects, and plot the unexplained variation in ET against the cumulative number of months that a field has been "called" to fallow (panel c) and the average ET of all of a field's neighbors in each month (pane d).

Moreover, there is no groundwater pumping in PVID—all supplemental irrigation is surface water diverted directly from the Colorado River. Could capillary effects from shallow groundwater explain the residual ET depicted in Figure S9? If so, then parcels that are likely to have greater near-surface groundwater tables should exhibit greater ET, but we find that this is not the case. Figure 4 indicates that neither proximity to the Colorado River (panel a) nor structure soil drainage characteristics (panel b) are correlated with water use. Rather, the evidence suggests that residual ET is driven by spillovers—over both time and space—from irrigation on uncalled fields. Panel (c) shows that residual ET falls as the length of a "fallowing spell" increases, suggesting that hold-over water from previous irrigation seasons can take a period of many months to fully "dry out," and raising important questions about the rototational nature of fallowing. This relationship is quite flat, however: residual ET on parcels that are fallowed for up to 5 years is only about .05

AFA lower than on a just-fallowed field.

Panel (d) of Figure 4 suggests that the primary culprit behind the residual water use on fallowed fields is seepage for neighboring fields that are still being irrigated. Average per-acre water use of a fallowed field's neighbors is positively related to own-field ET with a highly significant slope of roughly 0.44, and about 12% of the variation in fallow ET (Fig. S13). We confirm that this relationship reflects ground-level spillovers and not atmospheric contamination of of the ET signature from nearby fields by comparing the effect of upwind vs. downwind neighbors and finding that the two do not differ (Materials and Methods; Fig. S12 and S13). The upshot is that the flood irrigation techniques used in PVID likely result in significant cross-field spillovers that complicate managers' use of simple heuristics when estimating the water savings from fallowing.

3 Discussion

Our results indicate that reductions in consumptive agricultural water use due to the PVID-MWD Forbearance and Fallowing Program are substantially (29%) lower than estimates associated with the verification methodology employed by MWD and USBR. Crucially, the MWD/USBR method applies district-wide average water use per acre and assumes 100% reductions in consumptive use when estimating savings. Using remote sensing to develop field-scale estimates of ET allows us to improve on this methodology in two key ways.

First, field-scale consumptive use estimates provide a more accurate benchmark for potential savings under the assumption of 100% reductions. Whereas USBR's approach quantifies consumptive use as a residual difference between measured diversions and mostly unmeasured return flows, remote sensing of ET directly targets the quantity of interest: evaporation and transpiration that leave the hydrologic system and are consumptively used. In our setting where effective precipitation is millimeters per year and soils are naturally dry, ET may be the most reliable metric of consumptive use. Moreover, field-scale estimation facilitates measurement of average water use on the subset of lands enrolled in the program, providing a more accurate measure of potential savings that can be re-scaled to be consistent with USBR's district wide estimates.

Second, and crucially, field-scale water use estimates enable researchers to use causal inference techniques to address the performance of the program itself. In this study, reductions in water use causally related to the program are just just 55% of baseline use, as opposed to the assumed 100%. Half of this gap is likely explained by the strategic selection by farmers of which field to fallow in response to calls—many of which would have been fallowed anyway and so are not "additional." The other half is explained by residual ET on lands that have been verified as fallow. Much of this ET is driven by spillovers in soil moisture over time (previous irrigation) and space (neighbors' irrigation).

These findings relate to a long-standing debate in water policy and hydrology about how to properly conceptualize the potential savings from changes in agricultural practices including both efficiency enhancements and fallowing (Frederiksen and Allen, 2011; Gleick et al., 2011; Frederiksen et al., 2012). Ultimately, the question is whether the consumptive water use found on fallowed fields should be deducted from savings estimates or not, which relates to whether it is considered "beneficial" or "non-beneficial" consumptive use. The answer to that question depends, in part, on whether fallowing causes changes in water use on neighboring fields, a thorny question that runs afoul of inherent difficulties associated with scaling changes in field-scale water use to district or basin-level water savings (Lankford et al., 2020). However, regardless of how one conceptualizes this number, our results indicate that half of the water savings gap is explained by non-additionality. Roughly speaking, this factor alone would account for over 350,000 AF of water over the life of the program.

The program we study here is designed to free up water for urban customers of MWD. If they have over-estimated true water savings by 350,000 to 750,000 AF, who is absorbing the difference? The answer to that question turns out to be far from straightforward. The MWD/USBR savings estimates contained in this paper are published annually in a Verification Report for the Program, but these estimates do not directly alter MWD's diversion rights to Colorado River Water.[2] Instead, the USBR engages in a complex process of accounting at the end of each year that takes months to complete (Wobus et al., 2024), and MWD is the junior-most user of Colorado River Water in this process. Ultimately, MWD benefits from the fallowing program to the extent that it "moves the needle" enough in USBR's river accounting to allow MWD to diver more water. Whether the savings generated in PVID are realized by MWD depends on how water use changes for every other user of Lower Colorado River Basin water.

A complete water accounting for the LCRB is beyond the scope of this paper, but we *can* place some useful bounds on the implications of our findings for the cost effectiveness of the fallow program. Specifically, our estimate of the difference in saved water is the maximum amount that MWD might've paid for but not received if the mis-estimation affected their realized water savings on a one-for-one basis. Through 2021, MWD spent roughly $210 million on the program. According to their calculations, this would imply a price of roughly $145 per acre-foot. If the results here reduced their actual deliveries on a one-for-one basis, their cost per acre-foot was actually $286. Hence, accurately accounting for how much water is actually saved could change the cost effectiveness of supply augmentation via fallowing contracts relative to various demand-reduction strategies available to MWD's member utilities (Table S2).

Our results underscore how the design and evaluation of various water transfer mechanisms—and perhaps the enforcement of appropriative water rights themselves—can be improved with field-scale estimates of consumptive water use made possible by remote sensing. Field-scale water use estimation allows managers to more accurately forecast program savings by accounting for baseline differences in water use across enrolled vs. unenrolled fields, where they were previously

[2]These estimates are also used to calculate Intentionally Created Surplus (ICS) credits that MWD can virtually bank in Lake Mead and withdraw later. Hence, the estimation errors revealed by our analysis could inflate MWD's ICS balance and ultimately be associated with excessive drawdown of Lake Mead. However, MWD engages in a litany of programs to produce ICS credits had have hit the annual cap on credit creation almost every year, suggesting that it is relatively unlikely that errors in the savings estimates directly led to changes in the level of Lake Mead.

blind to these differences. Perhaps more crucially, field-scale estimates enable the use of causal program evaluation to reliably estimate water savings ex post, facilitating program tweaks to address non-additionality, compliance gaps, and hydrologic spillovers.

Our study has several key limitations. First, we do not observe the relationship between individual fields and larger farm operations. Doing so would allow us to analyze the strategic selection of which fields to fallow in a more robust way by focusing on within-farmer selection. Unfortunately, these data are not available in PVID. Second, our study area likely represents a "best case scenario" for using total ET to measure consumptive agricultural use, because effective precipitation and natural ET are negligible in PVID. In other settings such as the Central Valley of California, techniques to separate natural ET from applied ET are more essential (Boser et al., 2024). Finally, our setting benefits from a lack of groundwater pumping, making it straightforward to relate change in ET to changes in surface water use. In many other settings, it would prove more difficult to relate changes in ET to the supply of irrigation water, which may limit the utility of ET in facilitating transfers of water rights tied to specific sources.

4 Materials and Methods

4.1 Field-Level Water Use Estimation

We combine two data sets to develop field-scale estimates of water use across The first is monthly data on evapotranspiration from OpenET (Volk et al., 2024). These data are monthly raster files with measures of total evapotranspiration at a 30-meter resolution. We utilize the OpenET Ensemble mean as well as the six individual ET models that make up the mean. The constituent models include: ALEXI/DisALEXVI v 0.0.32, eeMETRIC v 0.20.26, geeSEBAL v 0.2.2, PT-JPL v0.2.1, SIMS v 0.1.0, and SSEBop v 0.2.6.

We obtained polygons of irrigated fields from the Metropolitan Water District (MWD). These polygons were created by PVID to track "water toll acres" that receive deliveries of Colorado River Water. We used zonal statistics to calculate monthly mean ET for each field over 2000 to 2021. Monthly means are depicted in Panel (a) of Figure 1.

4.2 Fallowing Calls

We obtained our data on which fields were selected for compliance with fallowing "calls" each year from annual Verification Reports published by MWD. These reports contain information about MWD/USBR's monthly estimates of total water use and total fallowed acreage for each year of the program, as well as maps of fields that were selected for fallowing by farmers in response to that year's call. An example map is depicted in Figure S3. PVID created these maps by hand with a printed aerial map the district and an overhead transparency—neither MWD nor PVID possessed statial data describing fallowed fields.

We geo-referenced each map from all available verification reports and overlaid these images with the irrigated fields shapefile obtained from MWD. We tasked three separate research assistants to independently code "called" fields from each image. We also had the research assistants create indicator variables for instances where the shaded area on the map represented a partial subset of a polygon in the shapefile, or when image quality made it difficult to determine whether a specific polygon was part of call.

From these underlying data, we create indicators for each field in each map to express whether one, two, or all three research assistants indicated that a field was i) called, ii) partial, and iii) of dubious quality. E.g., we have three separate definitions of treatment, three separate notions of which fields were only partially fallowed, and three separate notions of image quality.

The final step in creating a month-year panel of fallowing calls is to interpolate from the images in the verification reports to the intervening months. Typically, each annual verification report contains two images—one in July and one in January/December. From these snapshots, we must infer how to assign treatment to parcels in February-July and August-November. This is made possible by the structure of the program: MWD issues a fallowing call near the end of each calendar year that takes effect the following August and runs for two full years. E.g., a fallowing call issued in December 2012 would correspond to required fallowing from August 2013 through July 2025. With this logic, we can construct "fallowing spells" and interpolate forward and backward from a given image. For each image, which determine candidate 24-month windows that it could be a member of, and then assign treatment within the associated window.

This method is imperfect because for any given month, there are multiple candidate spells. However, the relative frequency of the snapshots allows us to discipline potential over-coding by forcing our assignment to agree with previous and subsequent images. We do this for each version of treatment described above. Figure S4 depicts month called acreage implied by our approach, in comparison to reported fallowing by MWD. Across most moths, we obtain close alignment with MWD's official records.

4.3 Other Data

We use several other data supplemental data sets in our analysis. First, we download monthly data on total precipitation at a 400-meter resolution from PRSIM. We calculate total precipitation and "effective precipitation"—defined as 0.65×precipitation (Wobus et al., 2024) for each parcel in each month of the sample. We use these data to illustrate that effective precipitation is small percentage (< 2) of total ET and that it varies inversely with the irrigation season. We also decompose the variation in precipitation into its cross-sectional and time-series components. Roughly 80% of the variation in precipitation is monthly, with 20% being parcel-specific. Less than 2% of the overall variation in precipitation is unexplained by parcel and month-of-sample fixed effects, suggesting it is unlikely to affect our results (Fig. S2).

We also estimate distance to the Colorado River as well as the mean of Schaetzl (1986)'s soil drainage index for each field in the sample. We overlay fields with PLSS sections–square-mile grids that comprise the land survey system) to enable us to cluster at a coarser spatial unit than individual fields.

Finally, we obtain hourly wind speeds from the Iowa Environmental Mesonet. We represent these windspeeds as vectors and take the vector average over each month of the sample to obtain a prevailing wind direction for each month. This approach is depicted in Panel (a) of Figure S12.

4.4 Difference-in-Difference Estimation

We estimate the causal impact of a fallowing "call" using a difference-in-difference framework that estimates changes in ET on called fields before vs. during a call compared to changes in uncalled fields over the same time period. Recent advances have demonstrated that classic two-way fixed effects regressions can result in biased difference-in-difference (DiD) estimates when treatment is staggered (different units are treated at different times) or when treatment effects change over time. A variety of estimators have been proposed to address this problem. We employ the estimator proposed by De Chaisemartin and d'Haultfoeuille (2020), which constructs a weighted average of all "valid" comparisons of switchers to non-switchers in the data.

We prefer the De Chaisemartin and d'Haultfoeuille (2020) approach to the prevailing alternatives for several reasons. First and foremost, to our knowledge, this is the only modern DiD estimator that is able to deliver causal estimates when treatment can switch on and off, which it does frequently in our setting of rotational fallowing. Second, the double-robust implementation of the De Chaisemartin and d'Haultfoeuille estimator constructs a propensity-score weighted average of comparisons that is explicitly designed to address potential selection into treatment–a major concern in our setting.

The identifying assumption behind the De Chaisemartin and d'Haultfoeuille is that there are parallel trends in untreated outcomes between the treated an untreated groups. While the potential untreated outcome for the group that ultimately receives treatment cannot be observed, De Chaisemartin and d'Haultfoeuille argue that placebo test to rule out differential pre-trends provides evidence in support of the identifying assumption of the estimator. After discussing our implementation of the estimator, we present the results of these tests.

Ultimately, we are interested in comparing *annual* water savings to MWD's estimates. However, because fallowing calls start in August and end in July two years later, treatment status can vary within a calendar year. We therefore estimate the model at monthly level (the finest level of measurement in the data) and at a seasonal level, where each year is divided into a January–July season and August-December season. The monthly estimates are useful for some of our policy calculations because MWD reports fallowed acres by month. The seasonal regressions are more statistically convenient for two reasons. First, aggregating ET estimates over time tends to smooth out any noise in the data (Volk et al., 2024). Second, each model run is computationally intensive, and obtaining a confidence interval for the annual estimate requires a bootstrap procedure with 500 iterations that estimates the sub-annual coefficients and sums them within each iteration—using the seasonal approach cuts the time down by a factor of six. Figure S5 illustrates that we obtain very similar annual savings estimates when aggregating monthly vs. seasonal coefficients.

Figure S6 depicts the results of the season-level estimates using the De Chaisemartin and d'Haultfoeuille (2020) approach, along with placebo test for differential pre-treatment trends in the five periods leading up to treatment. Our approach is to subset the data and run regressions on each season separately. E.g., for panel (a) the estimating sample contains one observation per field per year corresponding to the total early-season ET in that year. This means that the pre-trends compare ET across the five previous *years* as opposed to five previous *seasons*. For both the early and the late season, pre-treatment differences across treated and untreated parcels are not statistically different from zero, with the exception of a small dip in period $(t-1)$.

The dip in period $(t-1)$ is mechanical in nature. Moving from production to complete fallowing entails a major shift in agricultural operations, and unless this occurs on the final day of a given month or season, there is likely to be some dip in ET in the period just before fallowing it set to begin. E.g, if a farmer is set to begin fallowing on August 1, unless the wait and clear their field literally overnight on July 31, there will be some decline in ET associated with the foregone production in preparation for fallowing. With this caveat in mind, Figure S6 provides support for the parallel trends assumption and supports a causal interpretation of the drop in ET depicted in period 1 (roughly 1.6 AFA for the early season and 0.8 AFA for the late season).

4.5 Robustness

We run roughly 250 alternative speciations to illustrate the robustness of our results and present the results in Figure S7. We vary the choice of estimator (TWFE vs. De Chaisemartin and d'Haultfoeuille), whether or not we include a pure control group of unenrolled parcels, the choice of treatment coding (3 versions), sample restrictions based on RA quality flags (3 versions) and ET models (7 versions), yielding 252 specifications. The results are broadly similar across specifications. We note that the largest estimates that approach MWD's own figure all correspond with TWFE estimator that is likely biased in this setting. Restricting attention to the De Chaisemartin and d'Haultfoeuille estimates suggests that water savings are no more than 3 AFA, at most.

4.6 Comparison to MWD Estimates and Decomposition of Differences

The difference between the MWD/USBR approach and our approach can be be written as:

$$\text{Difference} = S_{\text{MWD}} - S_{\text{Actual}} = (\Omega_{\text{Fallow}} + \Omega_{\text{Add}}) - (\Delta_{\text{Meas}} + \Delta_{\text{Select}})$$

$$S_{\text{Actual}} = S_{\text{MWD}} + \underbrace{\Delta_{\text{Meas}} + \Delta_{\text{Select}}}_{\text{Spatial Resolution}} - \underbrace{\Omega_{\text{Fallow}} - \Omega_{\text{Add}}}_{\text{Performance Audit}}$$

Where the components are defined as:

- $\Delta_{\text{Meas}} = ET_{\text{avg}} - S_{\text{MWD}}$ (Measurement correction)
- $\Delta_{\text{Select}} = ET_{\text{enrolled}} - ET_{\text{avg}}$ (Selection bonus)

- $\Omega_{\text{Fallow}} = ET_{\text{residual}}$ (Residual water use)
- $\Omega_{\text{Add}} = (ET_{\text{enrolled}} - \Omega_{\text{Fallow}}) - S_{\text{Actual}}$ (Additionality gap)

With the exception of Ω_{Add}, these can all be calculated directly from the data. Therefore, we estimate the extent of non-additonality in the program as a residual that is unexplained by measurement differences in residual ET on fallowed lands. The results are presented in Figure 3. Δ_{Meas} accounts for level difference in ET vs. administrative measures of consumptive use. Δ_{Select} accounts for within-district differences in water use on enrolled vs. unenrolled lands that are invisible with the district-wide averages. Ω_{Fallow} is observed ET on fallowed lands.

4.7 Mechanisms

We perform several exercises to better understand the two main sources of less-than-100% savings due to fallowing depicted in Figure 3 and Figure S9. First, we present evidence consistent with the finding that many of the fields that fallow in response to fallowing calls are "non-additional" in the sense that they had a higher-than-average probability of fallowing in the absence of a call. Second, we examine potential determinants of residual water use. For both exercises, we focus our attention on enrolled parcels only.

To assess the plausibility of the non-additonality channel, we examine rates of fallowing on enrolled lands before the program began. We assume that any land with less than 2 AFA of annual water use was fallow, but our results are not sensitive to the choice of cutoff. Figure S10 depicts the probability of fallowing (simply how many years out of five a field was fallowed) for enrolled fields. The vast majority of enrolled fields were *never* fallowed prior to the program, however a non-trivial number we fallowed 1, 2, 3, or 4 times. Given this heterogeneity, were some fields subsequently selected for rotational fallowing after the program began? Figure S11 suggests that the answer was yes. It depicts a binned scatter plot that shows average rates of post-program fallowing for each bin of pre-program fallowing propensity. Fields that had never been fallowed still have a 40% change of being selected for a fallowing call, but the probability of selection increases as the pre-enrollment fallowing propensity increases. For instance, fields that were fallow 3 of 5 years prior to the program are roughly twice as likely to be selected for fallowing during the program than fields with no pre-program fallowing. This is consistent with the finding in Figure 3 that non-additionality accounts for roughly have of the water savings gap observed on fallowed fields.

The remainder of the water savings gap is attributable to the fact that average water use on fallowed fields is still roughly 1 AFA (Fig. S9. We rule out soil drainage and distance to the Colorado River as plausible explanations in Figure 4. Binned scatter plots show and linear regressions confirm that there is no detectable relationship between these variables and ET on fallowed lands.

Finally, we ask whether spillovers from irrigation itself could explain residual water use. One possibility is that it takes fields and extended period of time to dry out, such that ET continues after fallowing begins. We do find some evidence of this, depicted in panel (c) if Figure 4, which shows that residual ET is declining as function of how many cumulative months a parcel has been called. This relationship is statistically significant but the slope is quite flat, suggesting that this phenomenon cannot fully explain residual ET.

We also consider the role of irrigation on neighboring parcels, which may seep onto fallowed lands, especially given the relatively crude flood irrigation techniques being employed here. We estimate average ET across each fallowed parcel's neighbors on a monthly basis and provide a binned scatter plot in panel (d) of Figure 4. There is a strong positive relationship between neighbor ET and own ET, with a slope of about 0.44, suggesting that surface spillovers and seepage are a major driver of residual ET.

It is possible that the neighboring ET effect depicted in panel (d) of Figure 4 reflects measurement error in ET itself, rather than ground-level spillovers. ET models work by using variation in surface temperature to infer evaporation and transpiration associated with evaporative cooling. It may be that there are low-level atmospheric spillovers associated with evaporative cooling, such that the air about a fallow field is cooler when its neighbors are evaporating more, leading to a false ET signature.

We assess the plausibility of this alternative mechanism by exploiting variation in wind direction. The intuition is that evaporative/atmospheric spillovers should be stronger from upwind neighbors than from downwind neighbors. If neighbor effects are not correlated with wind direction, it suggests the effects are indeed ground-level spillovers. We calculate prevailing winds for each month in the sample by obtaining hourly wind speeds and directions and constructing a vector average for each month. We then construct a parcel's average upwind vs. downwind ET by taking the cosine of the difference between the angle between two parcel's centroids and the angle of the prevailing wind for a given month. The assigned weight is then the maximum of the cosine and zero. This means that a parcel directly downwind of its neighbor would receive a weight of one, whereas a parcel at exactly 90 degrees would obtain a weigh of zero. Panel (b) of Figure S12 illustrates.

Figure S13 demonstrates that neighbor spillovers are not correlated with wind direction. This can be seen visually in the figure, which plots a fallow parcel's own ET against its upwind vs. downwind neighbors. Regression results presented below the figure confirm this by regression fallow ET on total neighbor ET and upwind vs. downwind neighbor ET. The p-value on the hypothesis test that upwind and downwind ET matter equally ranges from 0.57 to 0.70. Moreover, the sum of upwind and downwind effects is roughly equal to the total effect.

References

Arellano-Gonzalez, J., A. AghaKouchak, M. C. Levy, Y. Qin, J. Burney, S. J. Davis, and F. C. Moore (2021). The adaptive benefits of agricultural water markets in california. *Environmental Research Letters 16*(4), 044036.

Boser, A., K. Caylor, A. Larsen, M. Pascolini-Campbell, J. T. Reager, and T. Carleton (2024). Field-scale crop water consumption estimates reveal potential water savings in california agriculture. *15*(1), 2366.

Bretsen, S. N. and P. J. Hill (2008). Water markets as a tragedy of the anticommons. *Wm. & Mary Envtl. L. & Pol'y Rev. 33*, 723.

Brewer, J., R. Glennon, A. Ker, and G. Libecap (2008). 2006 presidential address water markets in the west: prices, trading, and contractual forms. *Economic Inquiry 46*(2), 91–112.

Bruno, E. M. and K. Jessoe (2021). Missing markets: Evidence on agricultural groundwater demand from volumetric pricing. *Journal of Public Economics 196*, 104374.

Bruno, E. M. and K. Jessoe (2024). Designing water markets for climate change adaptation. *Nature Climate Change 14*(4), 331–339.

De Chaisemartin, C. and X. d'Haultfoeuille (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American economic review 110*(9), 2964–2996.

Elliott, J., D. Deryng, C. Müller, K. Frieler, M. Konzmann, D. Gerten, M. Glotter, M. Flörke, Y. Wada, N. Best, et al. (2014). Constraints and potentials of future irrigation water availability on agricultural production under climate change. *Proceedings of the National Academy of Sciences 111*(9), 3239–3244.

Frederiksen, H. D. and R. G. Allen (2011). A common basis for analysis, evaluation and comparison of offstream water uses. *Water International 36*(3), 266–282.

Frederiksen, H. D., R. G. Allen, C. M. Burt, and C. Perry (2012). Responses to gleick et al., which was itself a response to frederiksen and allen. *Water International 37*(2), 183–197.

Gleick, P. H., J. Christian-Smith, and H. Cooley (2011). Water-use efficiency and productivity: rethinking the basin approach. *Water International 36*(7), 784–798.

Gordon, B. L., G. F. Boisrame, R. W. Carroll, N. K. Ajami, B. Leonard, C. Albano, N. Mizukami, M. A. Andrade, E. Koebele, M. H. Taylor, et al. (2024). The essential role of local context in shaping risk and risk reduction strategies for snowmelt-dependent irrigated agriculture. *Earth's Future 12*(6), e2024EF004577.

Hagerty, N. (2023, 23 February). What holds back water markets? Transaction costs and the gains from trade. *Working Paper, Montana State University*.

Harris, L. (2024). Farmer response to policy induced water reductions: Evidence from the colorado river. *Journal of Environmental Economics and Management 125*, 102986.

Lankford, B., A. Closas, J. Dalton, E. L. Gunn, T. Hess, J. W. Knox, S. Van der Kooij, J. Lautze, D. Molden, S. Orr, et al. (2020). A scale-based framework to understand the promises, pitfalls and paradoxes of irrigation efficiency to meet major water challenges. *Global Environmental Change 65*, 102182.

Leonard, B., C. Costello, and G. Libecap (2019, Winter). Expanding water markets in the western United States: Barriers and lessons from other natural resource markets. *Review of Environmental Economics and Policy 13*(1), 43–61.

Medellín-Azuara, J., A. Escriva-Bou, A. C. Gaudin, K. A. Schwabe, and D. A. Sumner (2024). Cultivating climate resilience in california agriculture: Adaptations to an increasingly volatile water future. *Proceedings of the National Academy of Sciences 121*(32), e2310079121.

Rafey, W. (2023, February). Droughts, deluges, and (river) diversions: Valuing market-based water reallocation. *American Economic Review 113*(2).

Rafey, W. (2025, October). Measuring water misallocation in California. *Working Paper, UCLA Economics*.

Richter, B. D., D. Bartak, P. Caldwell, K. F. Davis, P. Debaere, A. Y. Hoekstra, T. Li, L. Marston, R. McManamay, M. M. Mekonnen, et al. (2020). Water scarcity and fish imperilment driven by beef production. *Nature Sustainability 3*(4), 319–328.

Richter, B. D., G. Lamsal, L. Marston, S. Dhakal, L. S. Sangha, R. R. Rushforth, D. Wei, B. L. Ruddell, K. F. Davis, A. Hernandez-Cruz, et al. (2024). New water accounting reveals why the colorado river no longer reaches the sea. *Communications Earth & Environment 5*(1), 134.

Rodell, M., J. S. Famiglietti, D. N. Wiese, J. Reager, H. K. Beaudoing, F. W. Landerer, and M.-H. Lo (2018). Emerging trends in global freshwater availability. *Nature 557*(7707), 651–659.

Schaetzl, R. J. (1986). Soilscape analysis of contrasting glacial terrains in wisconsin. *Annals of the Association of American Geographers 76*(3), 414–425.

Senay, G. B. (2018). Satellite psychrometric formulation of the operational simplified surface energy balance (SSEBop) model for quantifying and mapping evapotranspiration. *34*(3), 555–566.

Senay, G. B., S. Bohms, R. K. Singh, P. H. Gowda, N. M. Velpuri, H. Alemu, and J. P. Verdin (2013). Operational evapotranspiration mapping using remote sensing and weather datasets: A new parameterization for the SSEB approach. *49*(3), 577–591.

Volk, J. M., J. L. Huntington, F. S. Melton, R. Allen, M. Anderson, J. B. Fisher, A. Kilic, A. Ruhoff, G. B. Senay, B. Minor, et al. (2024). Assessing the accuracy of openet satellite-based evapotranspiration data to support water resource and land management applications. *Nature Water 2*(2), 193–205.

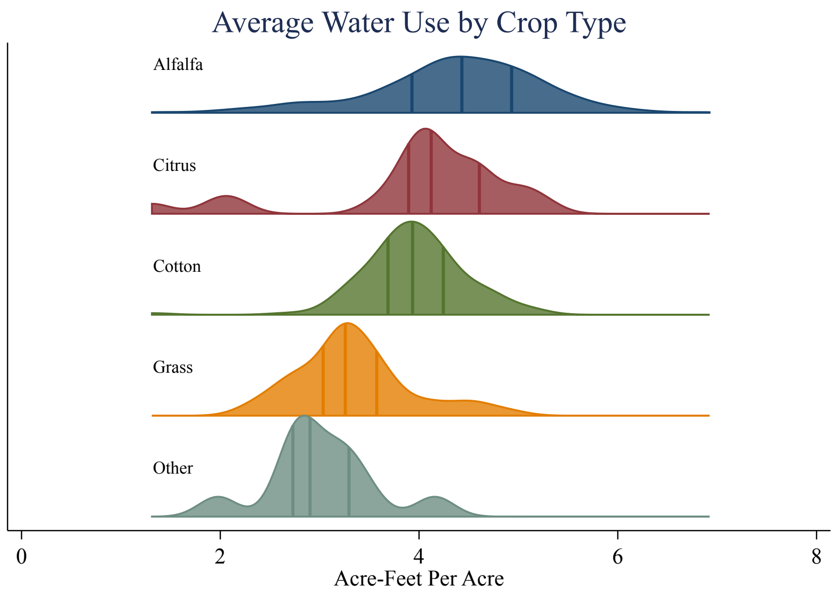
Wobus, C., C. Nash, P. Culp, M. Kelly, and K. Kennedy (2024). Simplified agricultural water use accounting in the colorado river basin using openet. *Environmental Research Letters 20*(1), 014020.

Wong, A. J., Y. Jin, J. Medellín-Azuara, K. T. Paw, E. R. Kent, J. M. Clay, F. Gao, J. B. Fisher, G. Rivera, C. M. Lee, K. S. Hemes, E. Eichelmann, D. D. Baldocchi, and S. J. Hook (2021, September). Multiscale assessment of agricultural consumptive water use in California's Central Valley. *Water Resources Research 57*(9).

Supplementary Material

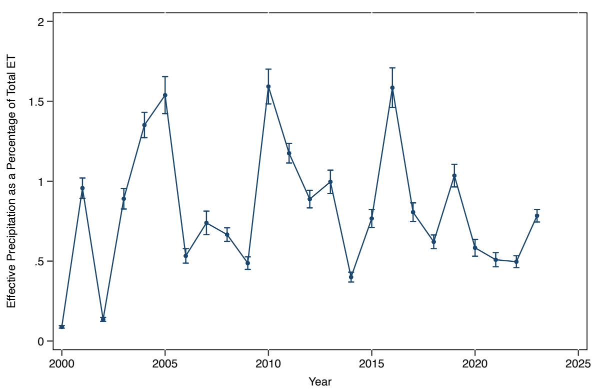
Figures

Figure S1: Example Verification Report Map

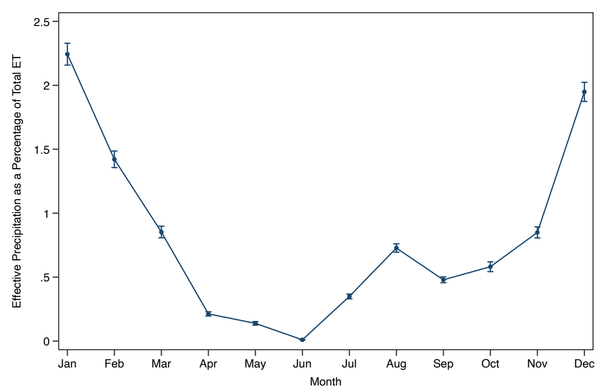


Notes: This figure depicts the distribution of per-acre water use on unenrolled fields for which we could determine a dominant crop using the Cropland Data Layer beginning in 2008.

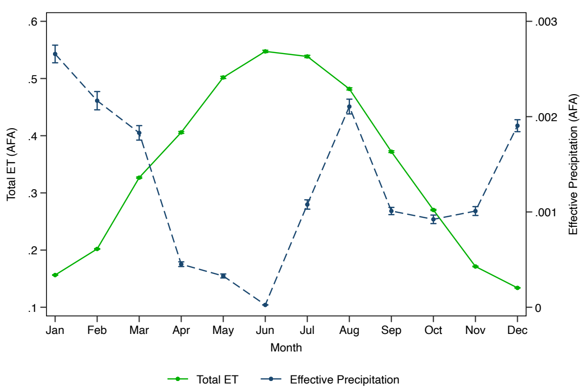
Figure S2: Effective Precipitation



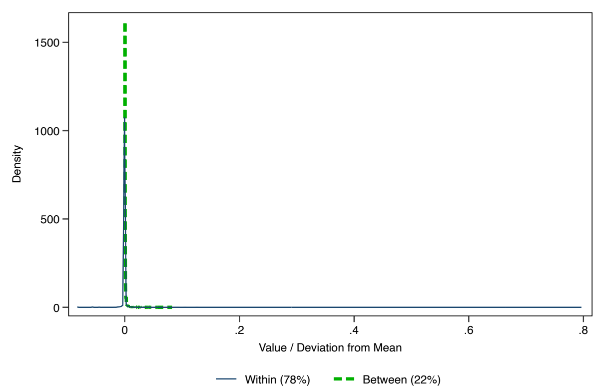
(a) Annual Effective Precipitation



(b) Monthly Effective Precipitation



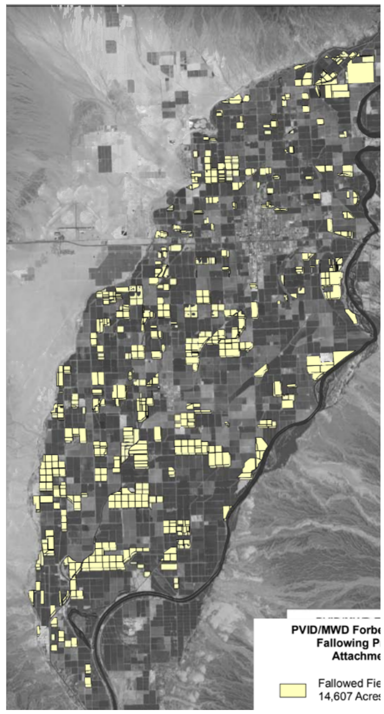
(c) Effective Precipitation vs. Total ET by Month



(d) Decomposition of Variation in Effective Precipitation

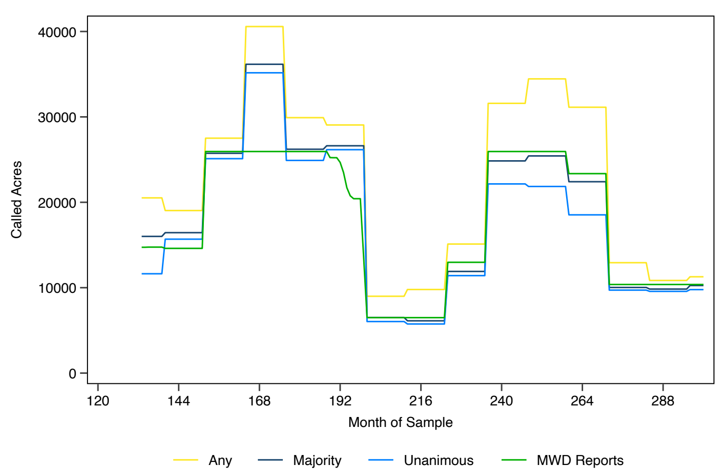
Notes: This figure characterizes effective precipitation in PVID across our study period. Panel (a) effective precipitation as a percentage of total ET in each year, averaged across parcels in the district with 95% confidence intervals. Panel (b) is similar to panel (a) but averages by month rather than year. Panel (c) illustrates how total ET (left axis) varies inversely with effective precipitation (right axis) across the year due to supplemental irrigation peaking in the summer. Panel (d) decomposes the variation in effective precipitation to illustrate that most of the variation is over time, rather than specific to individual parcels.

Figure S3: Example Verification Report Map



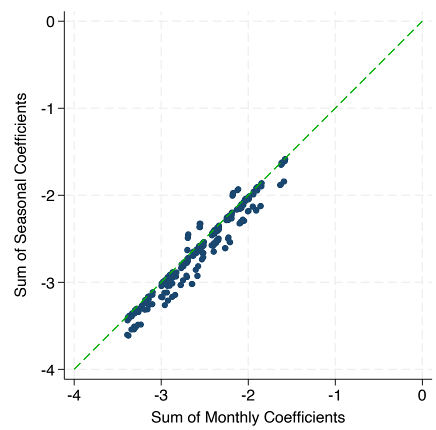
Notes: This figure depicts an example of the fallowing maps contained in MWD's annual Verification Report. Highlighted fields were identified as participating in the fallowing "call" in the month in question.

Figure S4: Monthly Fallow Acreage Estimates



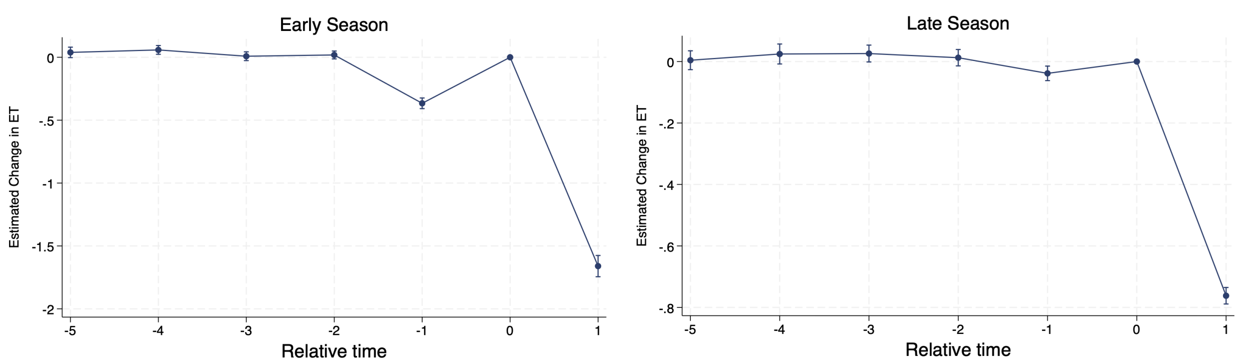
Notes: This figure depicts the alignment between total called acres implied by our interpolation of verification report images over time and reported fallowed acres from the verification reports. Our preferred approach is depicted in dark blue, where we code a parcel as "called" if at least 2 (out of 3) RA's coded it as such.

Figure S5: Implied Annual Savings from Monthly vs. Seasonal Estimates



Notes: This figure compares implied annual savings from summing across monthly coefficients vs. seasonal coefficients.

Figure S6: Pre-Trends in Seasonal Regression Estimates

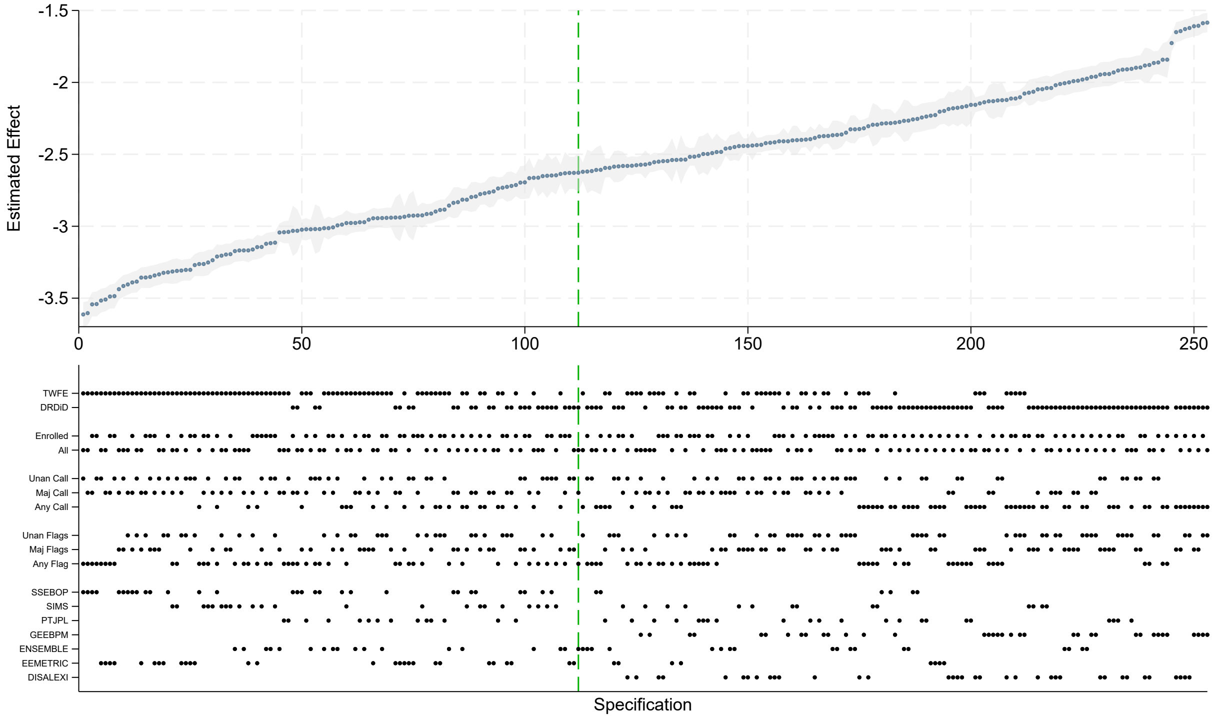


(a) January–July

(b) August–December

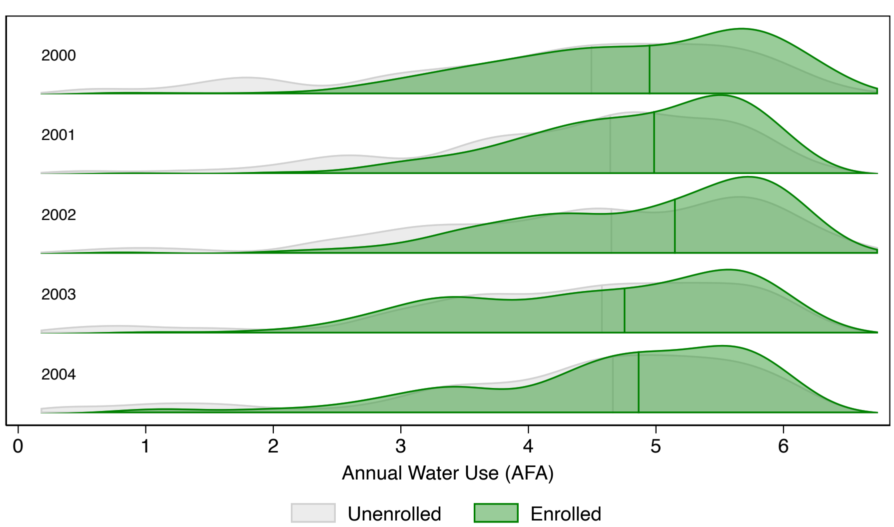
Notes: This figure depicts pre-treatment trends across treated vs. untreated parcels using the placebo approach proposed by De Chaisemartin and d'Haultfoeuille (2020). Here, we aggregate monthly data to 2 sub-annual seasons spanning January to July and August to December. The slight dip in t-1 is a mechanical result of the fact that farmers must harvest crops and stop growing the month prior to fallowing. There will always be such a dip unless they harvest in the middle of the night prior to the first fallowing date.

Figure S7: Robustness of Aggregated Annual Estimates to Alternative Specifications



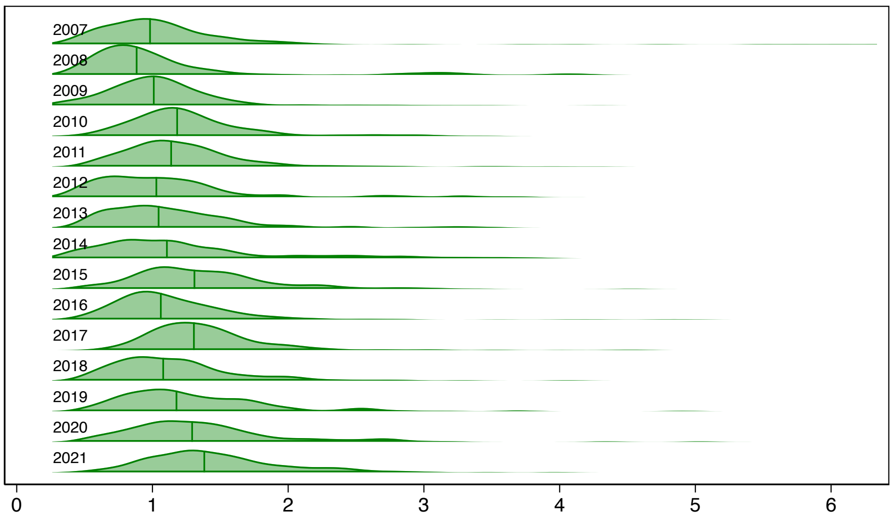
Notes: This figure depicts the robustness of our results across 252 specifications that vary: i) the estimator, ii) the inclusion of a pure control group, iii) the aggregating role for individual RA treatment coding, iv) sample restrictions based on RA-generated quality control flags in the map data, and iv) ET models.

Figure S8: Enrollment Selection



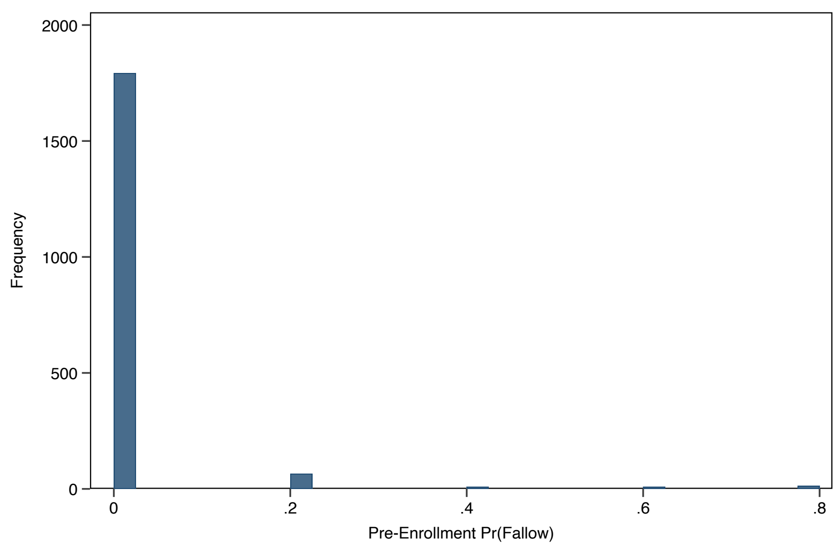
Notes: This figure depicts the distribution of ET on fields that eventually enrolled in the fallowing program (green) vs. those that did not (gray) prior to the beginning of the program.

Figure S9: Residual ET On Fallow Fields



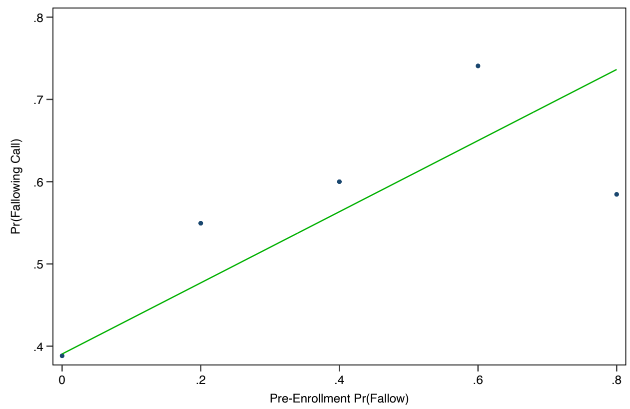
Notes: This figure depicts the distribution of water use on fields that were called to fallow in each year during the program.

Figure S10: Pre-Enrollment Fallowing Rates (Enrolled Fields)



Notes: This figure depicts raters of fallowing prior to the program on the subset of fields that ultimately enrolled in the program. There are five years of pre-program data, so the probabilities translate directly into counts (e.g., 0.8 indicates a field was fallowed 4 of 5 years).

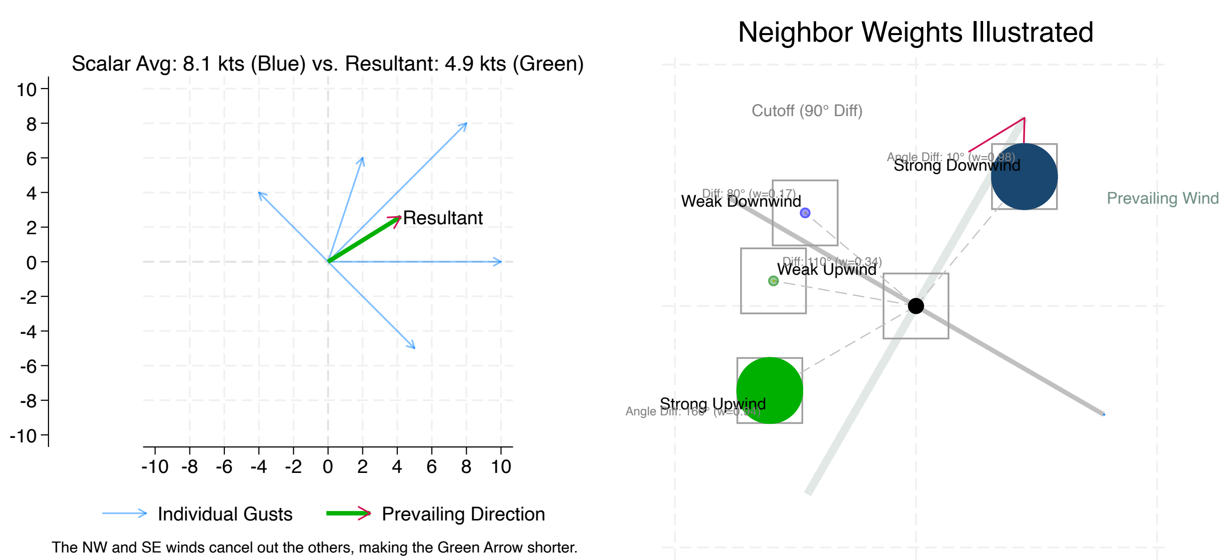
Figure S11: Pre-Program Fallowing vs. Call Frequency



	$Y = Pr(Called)$
0.2	0.161***
	(0.0323)
0.4	0.212**
	(0.102)
0.6	0.352***
	(0.0661)
0.8	0.196**
	(0.0995)
Constant	0.388***
	(0.00513)
adj. R^2	0.035

Notes: This figure depicts the relationship between the frequency of fallowing prior to the program (2000-2004) and fallowing rates during the observed program years. Each point represents the average fallowing rate during the program for parcels with a baseline fallowing rate of 0, 0.2, 0.4, etc. The table reports the differences in these fallowing rates for each bin relative to parcels that were never fallowed prior to the start of the program.

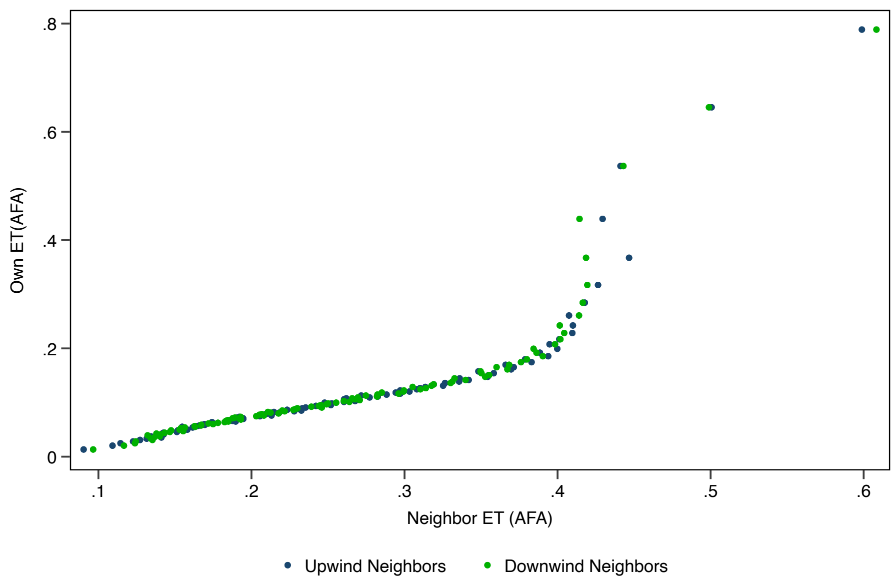
Figure S12: Constructing Upwind vs. Downwind Neighbors



(a) Estimating Prevailing Winds (b) Parsing Neighbors

Notes: This figure depicts our strategy for determining monthly prevailing winds (panel a) in PVID and constructing average upwind vs. downwind neighbor ET for each parcel.

Figure S13: Upwind vs. Downwind Neighbor Effects



	$Y = OwnET$			
All Neighbors' ET Mean		0.444***		0.325***
		(0.0115)		(0.0217)
Upwind Neighbors' ET Mean			0.201***	0.0601***
			(0.00690)	(0.0101)
Downwind Neigbors' ET Mean			0.204***	0.0650***
			(0.00651)	(0.0105)
N	100611	100611	100611	100611
adj. R^2	0.308	0.424	0.415	0.426
p-val (Upwind = Downwind)			0.706	0.574

Notes: This figure depicts differences in the relationship between a fallow parcel's own ET and upwind vs. downwind neighbors' ET. The reported regression results are based on linear regressions with field and month-of-sample fixed effects on a sample that includes fallow parcels only. The table reports the p-val associated with the null hypothesis that upwind and downwind ET have the same effect on a fallow parcel's ET.

Tables

Table S1: District-Wide Consumptive Use Estimate: ET vs. Traditional Method

	ENSEMBLE	PTJPL	SIMS	SSEBOP	GEEBPM	EEMETRIC	DISALEXI
Slope	0.987***	1.014***	0.995***	1.003***	0.991***	0.859***	1.028***
	(0.0157)	(0.0222)	(0.0168)	(0.0146)	(0.0192)	(0.0132)	(0.0147)
adj. R^2	0.926	0.887	0.919	0.933	0.910	0.928	0.941
RMSE	0.134	0.165	0.140	0.127	0.148	0.132	0.120
N	204	204	204	204	204	204	204

Notes: This table depicts the correlation between ET estimates of water use and MWD's estimates at the district-month-of-sample level. In each regression MWD's consumptive use estimate is the dependent variable and the indicated ET measure is the sole regressor. Constant terms are omitted so that the slopes are forced to go through zero.

Table S2: Cost Effectiveness Comparison

Water Savings Method	**Price per AF**	**Source**
Social Nudges (low end)	$78	Bernedo, Ferraro, and Price (2014)
Low Flow Technology (low end)	$129	Chermak, and Felardo (2014)
MWD Estimate	**$145**	MWD
Social Nudges (high end)	$189	Ferraro and Price (2013)
DIDm + OpenET	**$286**	This Paper
Xeriscape	$613	Brelsford and Abbott (2021)
Low Flow Technology (high end)	$3,535	Beneear et al. (2013)

Notes: This table depicts estimates of the cost of either reducing demand or enhancing supply for urban water use from a variety of studies in the literature, alongside our estimates from this paper.