## Appendix: Case Study

We demonstrate the use of synthetic control for inferring management intervention effects without *a priori* controls in the context of a brush-clearing treatment which occurred in southeastern Utah, USA in 2009. The Shay Mesa Restoration Project was designed to reduce fuel loads and improve wildlife habitat by removing Pinion (*Pinus edulis*) and Juniper (*Juniperus osteosperma*) trees over a 750 ha treatment area (details in Karl et al. 2014 and Gillan et al. 2016). There is some contention around the long-term effectiveness of such treatments, as well as potential erosion risks from increased exposure of bare ground following treatment (Archer et al. 2011, Gillan et al. 2016).

Within a designated section of the broader treated area, three types of brush-clearing methods were applied in distinct zones (fig 5): (1) Mechanical tree mastication, leaving debris scattered throughout, (2) Lopping followed by burning piled debris and (3) Lopping followed by broadcast burn of scattered debris. A fourth area was used as a control and monitored for pre- and post-treatment surface cover as with the other areas (Karl et al. 2014). We obtained rough outlines of the treated and control areas from the Utah watershed Restoration Initiative dataset (https://wri.utah.gov/wri/).

We assessed treatment effects based on the Soil Adjusted Total Vegetative Index (SATVI ; Marsett et al. 2006), which has been shown to accurately reflect total vegetative cover in the region of the case study (Poitras et al. 2018) . We calculated SATVI as:

using a timeseries of images from the Landsat archive from 1984 to 2018. Since single sensors do not span the entire timeseries, we used Landsat 5 for years between 1984 and 2011, Landsat 7 for 2012, and Landsat 8 for 2013 to 2018. As the synthetic control method in theory automatically accounts for satellite-derived noise shared among pixels, we were interested in performance of methods without recalibrating different Landsat products to a standard reflectance or subjecting images to cloud-masking algorithms. We used tier-1 surface reflectance products from all satellites, compiled using google earth engine (Gorelick et al. 2017).

For each pixel in the target areas, we identified a set of 100 control pixels, adapting methods from Nauman and Duniway (2016). Briefly, within a search radius of 3 km surrounding the perimeter of the treated area, we first performed a ‘masking’ operation, removing from consideration any pixels known to be part of another treatment or within 90 m of the focal treatment boundary, those disturbed by infrastructure (roads, oil and gas development), or those belonging to a non-analogous landscape cover-class (e.g. agriculture, urban, water) according to the National Landcover Dataset (NLCD 2011). We then narrowed candidate pixels to those with similar salinity (+- 5%) measured as saturated paste soil electrical conductivity and particle size in the control section classification (Soil Survery Staff 2010) to the focal treated pixel (Nauman et al. 2019). Restricting candidate reference locations by salinity and soil textural class was found to be an important step in this process, given the outsized role these variables play in determining ecological potential in these arid contexts (Nauman and Duniway 2016). From this subset, we selected the 100 most-similar pixels in the pool of control pixels, using Gower’s distance (van der Loo 2019) based on a suite of topo-edaphic variables (Table A1). We estimated treatment effects using the same methods outlined in the simulation model exercise, for each pixel, considering all observations after June 1, 2009 as post-treatment.

## Results

For the brush-clearing case study, estimated treatment effects were similar among all methods which included controls (Fig 6), all of which providing greater discrimination among treatment types than BFAST. The pixel-level estimates were heterogeneous within treatment areas (Fig 5, panel D; Fig. 6), with some pixels having greater treatment effects than others. This can be visualized in certain regions of the mastication treatment (Fig 5 panel D), where small stands of brush were left intact or in peripheral rocky areas with little brush to begin with.

Remote sensing estimates for treatment effects using SATVI, a proxy for ground cover, followed the same general ordering as ground cover change reported in Karl et al. (2014, fig. 6), with broadcast burn (B) having the greatest overall drop in SATVI, followed by pile burn (P) and then mastication (M). However, the increase in ground cover for the mastication treatment (M) observed by Karl et al. (2014) was not detected in this exercise, perhaps indicating that SATVI was not sensitive to the increased litter derived from slash debris or that debris did not compensate for reduced canopy cover. The broadcast burn area was also associated with greater wind and water erosion than the pile burn and control areas, as reported in Gillan et al. (2016), indicating that simple satellite-derived assessments may be useful for indicating relative functional treatment effects.

Discussion

However, in the brush clearing case study, where the landscape is dominated by perennial tree and shrub species, use of the CausalImpact method helped discriminate slight variations in treatment effects, compared to other methods, suggesting that at least some confounding noise was removed (Fig 6). Patterns among treatments in particular become more apparent when visualizing cumulative values (Fig. 7).