High-resolution land value maps reveal underestimation of conservation costs in the United States

Christoph Nolte[a,1]

[a]Department of Earth & Environment, Boston University, Boston, MA 02215

The justification and targeting of conservation policy rests on reliable measures of public and private benefits from competing land uses. Advances in Earth system observation and modeling permit the mapping of public ecosystem services at unprecedented scales and resolutions, prompting new proposals for land protection policies and priorities. Data on private benefits from land use are not available at similar scales and resolutions, resulting in a data mismatch with unknown consequences. Here I show that private benefits from land can be quantified at large scales and high resolutions, and that doing so can have important implications for conservation policy models. I developed high-resolution estimates of fair market value of private lands in the contiguous United States by training tree-based ensemble models on 6 million land sales. The resulting estimates predict conservation cost with up to 8.5 times greater accuracy than earlier proxies. Studies using coarser cost proxies underestimate conservation costs, especially at the expensive tail of the distribution. This has led to underestimations of policy budgets by factors of up to 37.5 in recent work. More accurate cost accounting will help policy makers acknowledge the full magnitude of contemporary conservation challenges and can help improve the targeting of public ecosystem service investments.

conservation cost | land value | machine learning | conservation planning

Trade-offs between public and private benefits from land use are at the heart of many policy debates. Land use decisions that focus on private benefits, such as development and resource extraction, can negatively affect public benefits, including carbon storage, biodiversity conservation, water provision, flood control, and recreational open space. Arguments for policy responses often rely on the idea that in the absence of intervention, the loss of public benefits from land use will exceed private gains. Scientists have a long history of informing such policy choices through the quantification of public and private benefits and the identification of cost-effective spatial targeting (1–4).

Advances in Earth system observation and modeling have transformed the ability of scientists to quantify natural features associated with public benefits at high spatial resolutions and large scales. Recent examples of this include carbon storage (5), species distributions (6), habitat connectivity (7), and flood risk (8). The release of new datasets often prompts new arguments in favor of long-term land protection and new proposals for spatial priorities (2). In the United States, where public environmental data are readily accessible, scholars have recently proposed nationwide priorities for species conservation (9, 10), habitat connectivity (11, 12), and floodplain acquisitions (13).

The quantification of private benefits from land—and its flipside, the private cost of conservation—has not kept up, however. High-resolution, large-scale estimates of conservation costs are rarely available for public use and are almost never empirically validated. In the United States, the aggregation of nationwide public records on land sales and valuation is largely the domain of commercial endeavors. Because such large-scale data have long been inaccessible to the public, many conservation planning studies have paid only limited attention to the costs that conservation organizations actually face (14). Nationwide prioritizations either ignore cost (15, 16) or rely on untested proxies, such as estimated returns from extractive uses (9, 11), farmer-reported agricultural land values (10, 17), county-level models trained on observed acquisition costs (13), and remotely sensed human footprints (12), among others. Studies developing high-resolution, parcel-level estimates of land value are constrained to small spatial extents (18, 19), limiting their utility for land use prioritization at large scales.

Here I show that land values in the contiguous United States can be estimated at the parcel level and near-continental scales with an accuracy that surpasses previous proxies by factors of up to 8.5. The estimation strategy uses parcel-level data on ownership, sales, buildings footprints, terrain, accessibility, land cover, hydrography, flood risk, demographics, and protection, synthesized from aggregated public records and open-access sources for 140.9 million properties in 3,055 counties using a novel data synthesis platform (PLACES). A tree-based bagging algorithm (20), parallelized at the county level, is trained on 6 million sales of properties larger than 1 acre (0.4 ha) occurring between 2000 and 2019. Trained models are used to predict expected sale prices for all properties in the conterminous United States in 2010. The results can be interpreted as parcel-level estimates of fair market value (FMV) and constitute a plausible cross-sectional proxy for the cost of conservation strategies that fully extinguish private property rights (Fig. 1).

Significance

This paper presents high-resolution maps of the estimated value of private lands in the contiguous United States. The estimates permit the prediction of the cost of conservation interventions at a much higher accuracy than proxies used in previous nationwide conservation planning studies. A close look at two recent studies on US-wide floodplain protection and species conservation planning for climate change suggests that the use of previous cost proxies led to the underestimation of policy budgets necessary to achieve environmental goals by a factor of 2 and a factor of 37.5, respectively. This can undermine the validity of findings. Future analyses of conservation policies can and should use high-resolution cost estimates in their justification and spatial prioritization of interventions.

Results

Compared with verified prices of 4,128 publicly funded land acquisitions for conservation in 659 counties, the new estimates explain between 67% and 72% of the variance in logged per-area prices, in contrast to 8% to 32% achieved by proxies used in previous US-wide studies (Fig. 2 and *SI Appendix*, Fig. S1). A large share of this difference in predictive power is attributable to within-county heterogeneities in land values, which exist throughout the country and can reach extreme values in locations containing both expensive urban areas and land of very low market value, such as deserts and wetlands (*SI Appendix*, Fig. S2). County-level averages of PLACES FMV estimates only explain 26% to 27% of the variance in acquisition cost (*SI Appendix*, Fig. S3). The ranking of the predictive performance of county-level proxies is sensitive to the selection of the validation sample, but high-resolution estimates consistently outperform county-level proxies (Fig. 2 and *SI Appendix*, Fig. S1).

Validation points toward an underestimation of conservation costs by several previous proxies: unadjusted, a proxy used in two studies (9), underestimates observed acquisition cost by a factor of 3.5 on average; another (13) by a factor of 2.2 (*SI Appendix*, Table S1 and Fig. S4). County-level proxies tend to miss the expensive tail of the land value distribution associated with urban proximity, which contains important conservation priorities, and attracts a significant share of public and private conservation investments (Fig. 2 and *SI Appendix*, Fig. S5).

Property tax assessments offer an alternative source of parcel-level property value estimates. Tax assessors estimate the FMV of each property in their jurisdiction as a precursor for the determination of "assessed values," which form the basis of property taxation. Because tax assessors are in a privileged position to consider local drivers of value, such as zoning, urban services, and amenities, one would expect their FMV estimates to outperform models based on nationwide datasets. Yet for verified land acquisitions where FMV estimates from tax assessments were available ($n = 2,903$), these estimates explain less of the variance in costs than PLACES FMV (56% vs. 72%) (Fig. 2 and *SI Appendix*, Fig. S1). Assessed values, which can be subject to locally specific adjustments, perform worse (44%). Both sources of tax assessor data underestimate conservation cost by a factor of 2.1 or greater on average (*SI Appendix*, Table S1 and Fig. S4). The utility of tax assessor data as a proxy of conservation cost is further limited by missing data. To the best of my knowledge, there is currently no property-level tax value dataset available that seamlessly covers the contiguous United States.

High-resolution land value estimates can help predict the cost of conservation strategies other than land acquisitions. In the United States, conservation easements (CEs) are a widely used conservation instrument, with 167,721 recorded transactions (25). CEs extinguish a subset of property rights (e.g., development rights) in perpetuity, allowing landowners to retain ownership. Because the subset of transacted rights varies, predicting the costs of CEs is more complex than predicting the costs of land acquisitions. Validation datasets of CE costs are scarce, as neither easement presence nor easement costs are regularly recorded by tax assessors or registries. Using a dataset of 335 public CE transactions from one state program (Great Outdoors Colorado [GOCO]), with each transaction validated by a private and a public appraiser, I find that PLACES FMV estimates explain 73% of the variance in appraised easement value (*SI Appendix*, Fig. S6).

The quality of cost estimates can have important implications for insights from prospective modeling of conservation policies. Consider two examples: flood damage prevention and species conservation in the face of climate change. Flooding is the deadliest and most costly natural disaster in the United States.
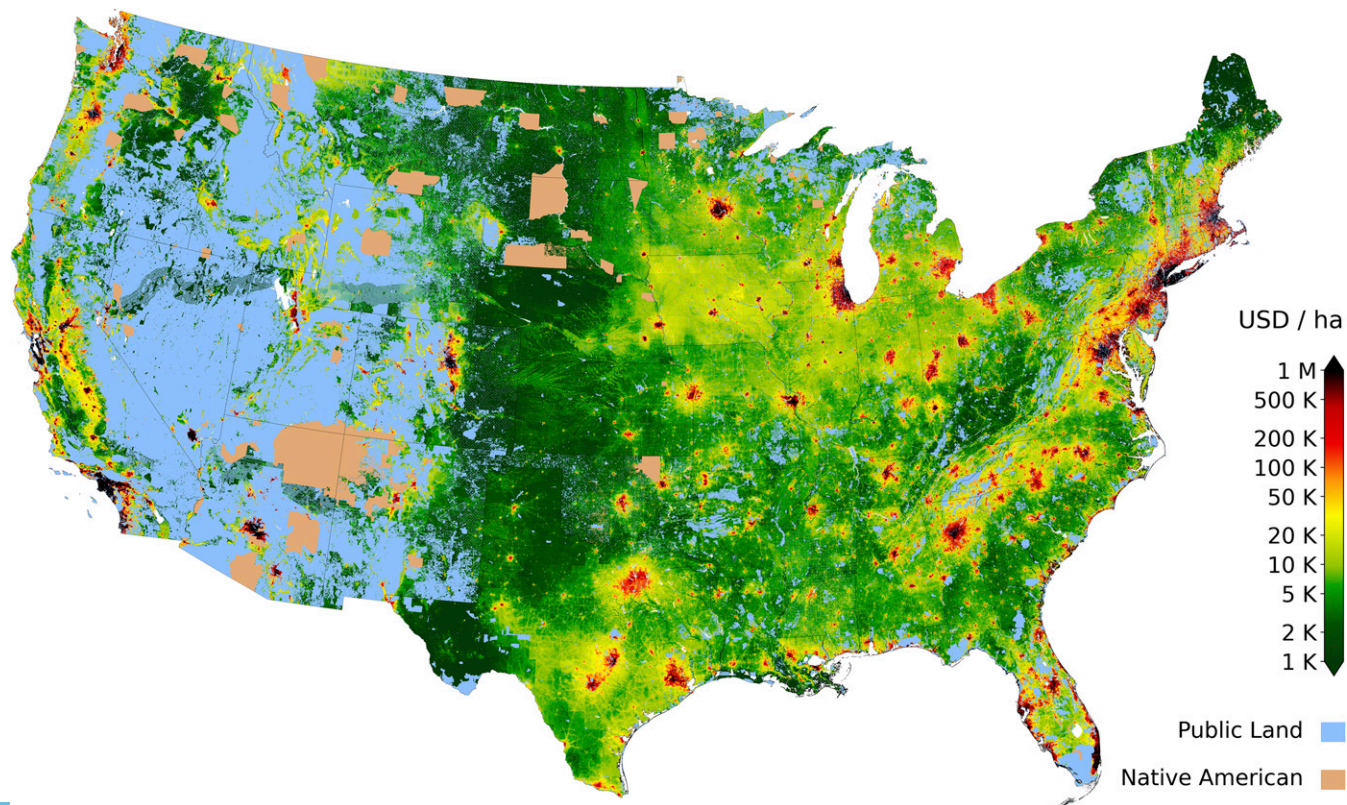

Fig. 1. Estimated FMV of private properties in the United States.
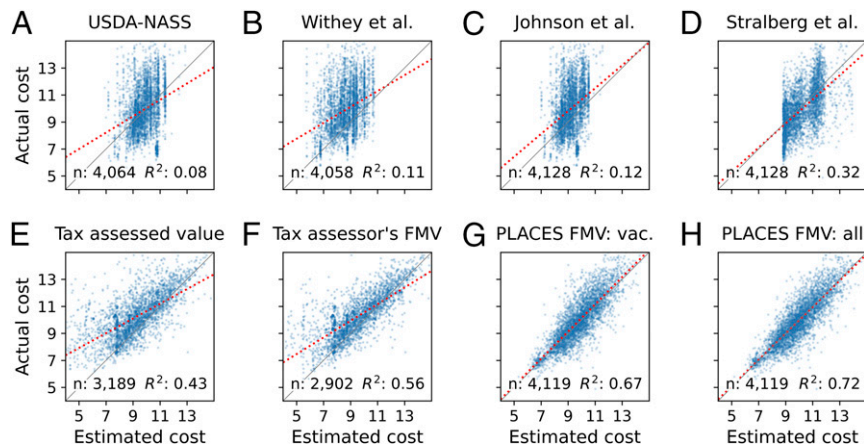
www.manaraa.com

Fig. 2. Predictive accuracy of proxies for the conservation cost of 4,128 publicly funded land acquisitions. Cost units are logged 2017 USD per hectare. Solid gray line shows diagonal. Dotted red line shows fitted regression. (*A*) County-level agricultural land values (21), used in (10, 17). (*B*) County-level returns from extractive uses (22), used in (9, 11). (*C*) County-level predictions of land acquisition costs, developed and used in (13). (*D*) Unitless proxy derived from global human footprint (23), used in (12), adjusted to fit graph. (*E*) Tax assessor valuation (24). (*F*) Tax assessor FMV (24). (*G*) PLACES FMV trained on sales of vacant land only. (*H*) PLACES FMV trained on sales of both vacant and developed properties.

In 2020, Johnson et al. (13) published the first US-wide benefit-cost analysis of acquiring undeveloped floodplains to prevent future development and avoid associated private damages. As a proxy for acquisition cost, the analysis used county-level cost predictions from a model trained on 1,405 land purchases by The Nature Conservancy, a large conservation nonprofit (Fig. 2*C*). The study found that the benefits (in terms of avoided private damages) of acquiring all natural, undeveloped floodplains within a 100-y flood return period outweigh acquisition costs by a factor of 1.94.

PLACES FMV estimates for vacant land suggest that Johnson et al. underestimate the costs of floodplain acquisitions by a factor of 2.06 (Fig. 3). This difference is largest near cities, where most development and associated flooding damages are likely to occur. Within 25 min of a city center, estimates of floodplain costs differ by a factor of 6.6 on average. Thus, costs of floodplain acquisitions might outweigh the added private benefits to landowners in more locations than Johnson et al. suggest. However, public benefits from public ownership, such as recreation, ecosystem services, species conservation, and avoided flood rescue cost, could make up for the difference and should be included in future cost-benefit analyses of floodplain acquisitions.

High-resolution cost estimates also lead to major shifts in the estimated cost and spatial patterns of species conservation policies in the face of climate change. Lawler et al. (11) have explored the effects of incorporating future species distributions, climate refugia, and ecological corridors into conservation planning in the contiguous United States. As a proxy for protection cost, the study uses county-level estimates of land prices based on estimated returns from different land uses (9, 22) (Fig. 2*B*). The analysis finds that addressing climate change in conservation plans has significant implications for the spatial configuration of cost-effective protected area networks, and that the cost of doing so relative to planning without addressing climate change might be relatively modest. Although those conclusions hold true in relative terms when better cost data are used, the total cost of the proposed conservation strategies rises dramatically, by factors ranging from 31.8 to 37.5 (Fig. 4). The difference is largely driven by planning units in expensive locations near cities, such as San Francisco, Los Angeles, Miami, Washington, and New York, which are considered irreplaceable (*SI Appendix*, Fig. S8). Incorporating high-resolution land value estimates into the optimization algorithm changes spatial priorities by 23% to 36% (Fig. 4 and *SI Appendix*, Table S2 and Fig. S7).

Discussion

The quality and resolution of conservation cost estimates can have important implications for arguments about the benefits, costs, and spatial priorities of land policies. Advances in predictive algorithms, data quality, and data access are bringing high-resolution cost estimates within the reach of conservation planners. The analysis presented here finds that the use of coarse and unvalidated cost proxies has led to an underestimation of budgets required to attain proposed goals. To put the results in perspective, on July 22, 2020, the US Congress passed a bipartisan bill that will fully fund the Land and Water Conservation
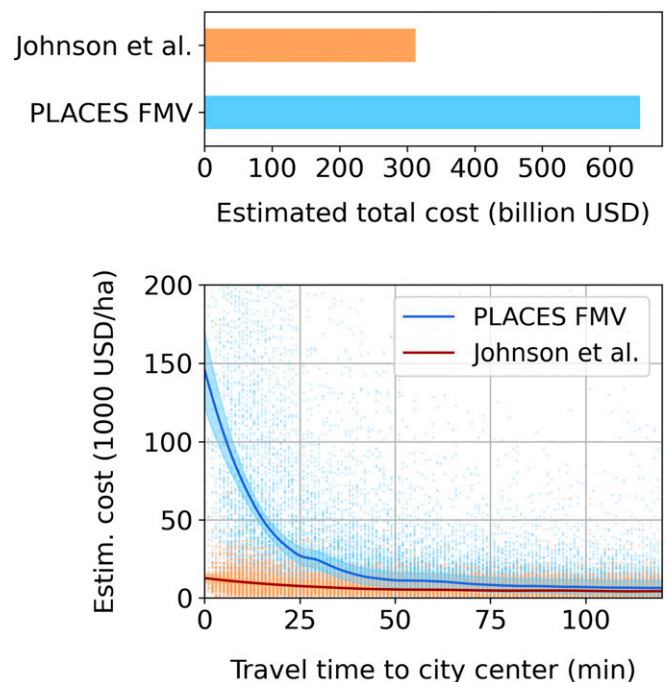

Fig. 3. Differences in cost estimates for the acquisition of natural, undeveloped, privately owned lands within 100-y floodplains. Cost estimates from Johnson et al. (13) and PLACES FMV for vacant land. Travel time to city center from Nelson et al. (26).
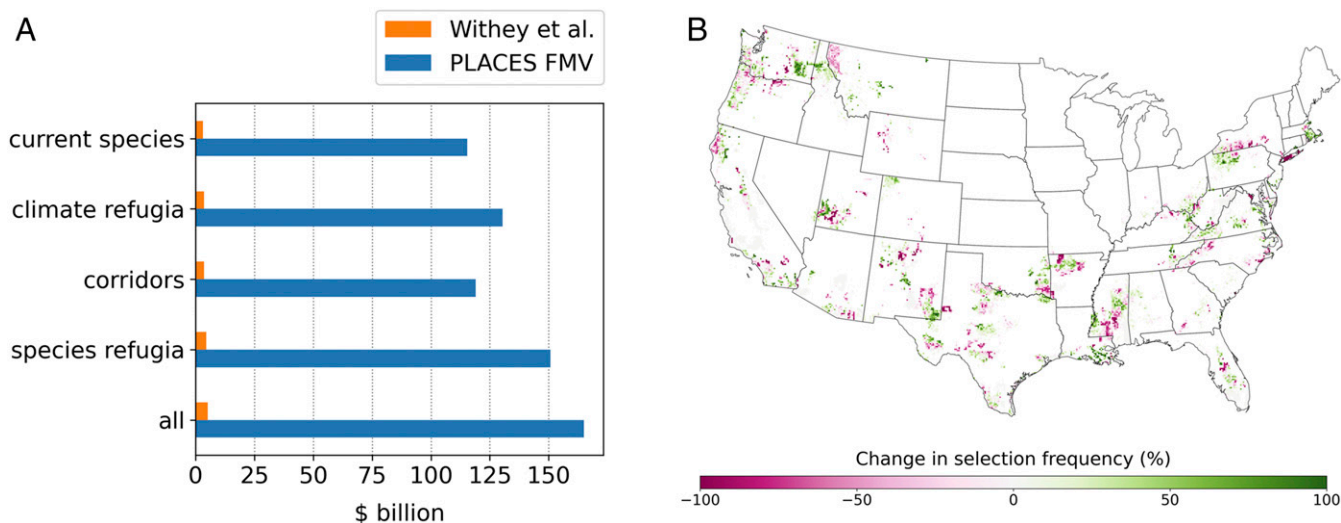
Nolte

www.manaraa.com

Fig. 4. Differences in cost estimates and spatial priorities for species conservation in the face of climate change. (*A*) Protected area network costs using different cost estimates during optimization and cost estimation. Network definitions are as in Lawler et al. (11): 1) current species (CS), 2) CS + climate refugia (CR), 3) CSD + corridors, 4) species refugia (SR), and 5) SR + CR + corridors. (*B*) Changes to the spatial configuration of cost-effective reserve networks induced using PLACES FMV cost estimates in the CS network case. *SI Appendix*, Fig. S7 provides maps of changes to other networks.

Fund, a major federal funding mechanism, for the first time since its creation in 1964 (27). This decision creates a budget of approximately $4.5 billion to secure additional lands for conservation. If this funding were entirely dedicated to species conservation, it could fully cover several of Lawler et al.'s proposed land protection scenarios (11). Revisiting the analysis with high-resolution estimates suggests that even such an unprecedented budgetary decision covers <5% of what would be needed to reach the proposed policy targets. Earlier findings on the cost-effectiveness of floodplain acquisitions also might need to be revisited in the light of new cost estimates.

To facilitate the replication and reexamination of findings from previous conservation planning studies and to encourage new research in this area, I have published rasterized maps of the property-level land value estimates alongside this article (*Data Availability*). When using these estimates and interpreting results, analysts need to be cognizant of several key limitations. First, the quality of the value estimates will vary as a function of the density and representativeness of training observations, as well as the extent to which observed predictors capture the most important drivers of value in a given locality. Attention to known spatial heterogeneities in training data density (*SI Appendix*, Fig. S9) and prediction error (*SI Appendix*, Fig. S10) will allow analysts to exercise caution in the interpretation of future findings based on these estimates. Second, while these estimates have been validated against a convenience sample of publicly funded land conservation transactions, further systematic research is needed to examine their relationship with other types of conservation investments, such as private acquisitions, donated easements, land use regulations, and time-bound conservation contracts, among others. Third, the land value estimates are cross-sectional and not predictive of the future. Land values and conservation costs can and will change in response to a wide range of factors, including natural disasters, economic crises, agricultural subsidies, regulatory change, and path-dependent development patterns, among others (28, 29). The development of models to predict future property values remains an active area of research in machine learning (30).

Limited access to high-resolution land cost data remains a barrier for the advancement of global change science and policy (31). The public disclosure of land transaction records and spatialized property data, in combination with remote sensing and machine learning techniques, offers new opportunities for resolving this bottleneck. Future investments in public access to property-level information, alongside further research into the strengths and limitations of spatial-temporal land value modeling and prediction, will help improve the empirical foundation for proposing and targeting conservation interventions across large and heterogeneous landscapes.

Materials and Methods

Training Data. The main unit of analysis in this study is the tax assessor parcel. All datasets were synthesized to the parcel-level using PLACES, a parallelized data pipeline deployed on Boston University's shared computing cluster at the Massachusetts Green High-Performance Computing Center (www. placeslab.org/places). This study uses digital maps of tax assessor parcels for 3,055 out of 3,108 counties in the contiguous United States (*SI Appendix*, Fig. S11). For 913 counties (29.4% of counties with parcel data), these maps were obtained from open-access sources; the remainder (*n* = 2,195; 70.4%) were licensed from three commercial providers (Loveland, Boundary Solutions, and CoreLogic). Parcel boundaries reflect conditions observed recently, mostly between 2016 and 2019. In areas where digital parcel boundaries are unavailable, PLACES creates wall-to-wall hexagonal dummy parcels with an area of 25 ha (average parcel size across counties). The following set of variables is computed for all parcels:

- Building footprints of 125.2 million buildings were obtained from Microsoft's open-source building footprint dataset (32) and used to compute the number of buildings on each parcel, the percentage area of the parcel covered by buildings, and the density of building footprints within the vicinity of each parcel (as a proxy for nearby development).
- Distance to paved roads is computed based on TIGER/Line shapefiles from the US Census Bureau (33).
- Travel time to major cities is extracted from a global map developed by the Joint Research Center of the European Commission (26).
- Demographic indicators (here median household income at the census block level) are imported from the National Historical Geographic Information System (34).
- Data on the long-term protection of parcels comes from the Protected Area Database of the United States (PAD-US 2.0) (35) for fee ownership and from the National Conservation Easement Database (25) for conservation easements. The exceptions are: New England, where superior coverage is offered by the New England Protected Open Space database (36), and Colorado, where superior coverage is provided by the Colorado Ownership, Management, and Protection (COMaP) database (37). To account

www.manaraa.com

for neighborhood effects, I compute the presence of nearby protection (% area) within different radii and years.

- Average slope and average elevation are computed from the US Geological Service (USGS) National Elevation Dataset (1/3 arc-second) (38).
- Wetland coverage (% of parcel area) is based on the US Fish and Wildlife Service National Wetlands Inventory (39).
- Water frontage to rivers and lakes is based on the USGS National Hydrography Dataset (40) and computed using polygon buffering and intersections.
- Flood risk for pluvial and fluvial flooding, measured as average meters of inundation depth within the 1% annual exceedance probability floodplain, comes from the US-wide flood hazard layers developed by Wing et al. (8).
- Proximity to coast is computed as the percentage of coastal waters within a short-distance radius (50 m, to flag properties with beach and boating access) and a long-distance radius (2,500 m, capturing both distance to coast as well as the added value of properties surrounded by coastal waters on several sides, such as islands, peninsulas, etc.), and based on Esri's North American water polygons (41).
- Land cover estimates are obtained from the 2011 National Land Cover Database (42) and used to compute percent land cover for the following classes: forest, crops, pasture, grassland, shrub, and barren.

Property sales data come from the Zillow Transaction and Assessment Database (ZTRAX, version: 9 October 2019) (24). ZTRAX contains tax assessor data (parcel numbers, owner names, geographic coordinates, assessed values, FMV estimates, and last sale information) and transaction data (parcel numbers, sale dates and prices, interfamily transfer flags). ZTRAX does not provide digital parcel maps. The provided geographic coordinates are often incomplete and not precise enough to identify parcels correctly. Thus, PLACES uses assessor parcel numbers to link ZTRAX data to parcel boundaries using county and town-specific string pattern matching and geographic quality controls. The algorithm identifies >1,000 unique combinations of syntaxes and links digital parcel boundaries from 2,951 counties to ZTRAX data with a median county-level success rate of 98.2% and a mean of 95.5% (measured as the percentage of the number of parcel boundaries matched to a tax assessor record). In counties for which the algorithm fails to identify the correct link (*SI Appendix*, Fig. S12), PLACES creates hexagonal dummy parcels based on available geographic coordinates and parcel area and computes all parcel-level variables for these dummies. Sales information is extracted from both ZTRAX datasets (transaction data and last sale information in tax assessor data). Multiparcel sales are aggregated to single observations.

Validation Data. Validation data for the cost of land acquisitions for conservation purposes ("fee transactions") comes from the Conservation Almanac (CA) (43), a US-wide dataset of 67,187 publicly funded land acquisitions. A large share of the CA data are not usable for this analysis. Only 21,026 (31.3%) fee transactions 1) have spatial information (polygons); 2) contain information on spending; 3) are larger than 1 acre (0.4 ha), the smallest size of parcels considered here; and 4) pass additional spatial quality checks, namely the absence of overlaps between polygons and satisfactory spatial matches between CA polygons and tax assessor parcels (66% minimum in both directions). Furthermore, spending data in the CA are not always accurate and do not always capture the full cost of land acquisitions (e.g., in the case of multiple contributions and partial donations). For an increased confidence in the validation data, I compare spending amounts for CA fee transactions with prices of parcel sales recorded in digital parcel maps or ZTRAX for the corresponding parcels and year (± 3 y), and flag prices of fee acquisitions as "verified" if single-parcel prices are within $\pm 20\%$ of each other. A total of 4,883 fee transactions (7.3%) pass this quality check. Because these transactions are clustered in space, I cap the density of validation data per county at 0.05 per km² and select a random sample of transactions in counties that surpass this threshold. This leads to a final validation dataset containing 4,128 fee transactions in 659 counties (*SI Appendix*, Fig. S12).

Validation data for the cost of conservation easements come from GOCO, a large state-level conservation program funded by Colorado's state lottery. The dataset contains estimated FMVs of easements supported with GOCO funding. Each transaction was independently appraised by two certified appraisers, one selected by the involved land trust and the other selected by GOCO. Colorado's COMaP database (see above) contains the ID of GOCO-supported easements, which permits establishment of the link between GOCO transactions and PLACES parcels. The dataset contains 335 easement transactions occurring between 1996 and 2017.

Cost Proxies. Data on US-wide proxies for conservation cost were retrieved from published sources and from the lead author of published articles:

- The National Agricultural Statistics Services (NASS) of the US Department of Agriculture (USDA) publishes county-level estimates of the average market value of agricultural land for all US counties, based on farmers' responses to the Census of Agriculture (21). USDA-NASS land value estimates have been used by several US-wide conservation planning analyses as cost proxies (10, 17); however, a recent study casts doubt on the utility of these estimates as proxies for conservation cost (44).
- Withey et al. (9) develop county-level estimates of land cost based on estimates of annual returns for six different land uses developed by Lubowski et al. (22, 45). They use these estimates as proxies for conservation cost to identify cost-effective investment strategies for US-wide species protection. Lawler et al. (11) subsequently use these estimates to identify cost-effective protected area networks under climate change in the United States.
- Johnson et al. (13) estimate the benefits and costs of floodplain land acquisition for flood damage reduction. Their county-level, US-wide predictions for conservation costs are based on a statistical model that regresses the cost of 1,405 land acquisitions by The Nature Conservancy on parcel size, USDA-NASS estimates for agricultural land value, and the price of land under single-family residences (46).
- Stralberg et al. (12) develop a proxy for land cost from the 2009 Global Human Footprint (GHF) raster layer (23), using the following formula: $\text{cost} = \text{GHF}^2/100 + 1$. The resulting values are unitless. To compare this proxy against actual conservation cost in Fig. 2, I convert it into 2017 USD by multiplying it with a constant derived from PLACES estimates (average vacant PLACES FMV/average raw cost proxy).

Tax assessor estimates were obtained directly from ZTRAX. FMV estimates are meant to reflect the value that a buyer would be willing to pay for a property on the open market with no undue influence. Tax assessors estimate FMV through a process that considers home sales, location, and inspections. Tax assessed value (TAV) identifies the value of a property for tax purposes. It is used as a basis for the calculation of property taxes, and computed by tax assessors. TAV might not reflect FMV for a range of reasons, including locally specific adjustments; for instance, many state or local governments have preferential tax assessments for rural land (e.g., current use programs).

Model Fitting. I train tree-based ensemble learning methods in the prediction of per-hectare sales prices to estimate parcel-level FMVs. The estimation process is parallelized at the county level: sample selection, fitting, and prediction occur for each county ("target county") independently. Predictive models for each target county are based on sales of properties of at least 1 acre (0.4 ha) that occurred between 2000 and 2019 in the target county and its adjacent counties. Adjacent counties are defined as counties that intersect with a 10-km buffer around the target county boundary. In counties where sales data are scarce, sales from nearby counties are added (in order of the distance between the centroids of target county and nearby counties) until the pool of training data contains at least 1,000 sales of vacant properties larger than 1 ha. Sales duplicates across data sources (transaction data vs. assessment data, linked parcels vs. dummies) are identified based on parcel number and year and removed. In addition, I exclude:

- Sales that were likely non–arms-length sales, i.e., sales in which data indicate that buyers and sellers might have been related, based on the presence of an interfamily transfer indicator (provided by ZTRAX) or similarity of seller, buyer, and owner names.
- Sales likely involving public buyers or sellers, as estimated from seller, buyer, and owner names.
- Sales of properties encumbered by easements.
- Sales with prices below $1,000.
- Sales with prices that deviate from assessor's FMV estimates by a factor of 100.

Vacant sales are identified as sales without a building footprint, without a land use code indicating the presence of a building (ZTRAX), and without a positive assessment value or FMV for buildings in the tax assessor data (ZTRAX). To prevent developed sales from overwhelming the model fitting procedure in urbanized locations, the ratio of developed sales vs. vacant sales is capped at 2:1. In locations where this ratio is found to be larger, the cap is enforced by drawing a random sample from the pool of developed sales.

The full training sample contains 6.01 million nonduplicate land sales, distributed unevenly across the contiguous United States (*SI Appendix*, Fig. S9). The training data for each target county is used to fit extremely randomized

Nolte

www.manaraa.com

trees models, a tree-based ensemble method for supervised regression (20). As the dependent variable, I use logged per-hectare sales price, inflated to 2017 USD using the monthly unadjusted Consumer Price Index for urban consumers (47). The main model ("PLACES FMV: all") uses all sales and 27 predictor variables (features) (*SI Appendix*, Table S3). I also train a model on vacant land sales only ("PLACES FMV: vacant"), omitting parcel-level building density and building area as predictors. Both models use 500 base learners (decision trees) and require three samples per leaf (*SI Appendix* provides alternative specifications).

Model performance is evaluated based on the mean squared error (MSE) of out-of-bag (OOB) predictions. OOB predictions are model predictions that use only the subset of base learners (decision trees) that were not trained on the observation whose value they predict. Across all 6.01 million sales observations, the average OOB MSE is 0.98; 95% of counties have an MSE <1.93. Variables associated with the presence and size of buildings are the most important drivers of FMV in the main model, followed by variables representing variation across time and space and those representing wetland presence (*SI Appendix*, Fig. S13). These predictors alone account for 60% of overall feature importance. Fitted county-level models are then used to predict expected sales prices (FMV) for all parcels within a county. To counter overfitting, OOB predictions are used for properties that were part of the training sample.

Validation and Robustness Checks. Conservation proxies are compared against observed conservation costs based on simple linear models that regress observed conservation cost on an intercept and the corresponding cost proxy (Fig. 2). Before fitting, all proxies and observed cost estimates are converted to the same unit (log 2017 USD per hectare). Predictive power is reported as the R^2 value of the linear models. Bias in cost proxies is reported as the estimated slope of linear regressions (1 = no bias) and average differences between actual and predicted costs (0 = no bias) (*SI Appendix*, Table S1).

I perform a suite of robustness checks to test for differences in the predictive accuracy and bias of alternative parameterizations of the model fitting procedure. *SI Appendix* provides detailed descriptions of these checks, and *SI Appendix*, Figs. S4 and S14 present results. Key observations include the following:

- The omission of any set of predictors reduces predictive power for both the training data (all property sales) and the validation data (conservation acquisitions).
- The addition of parcel size as a predictor improves predictive power in the training data but leads to an overestimate of conservation cost in the validation data, especially for large parcels, without an improvement in predictive power (R^2). This is likely the result of a systematic difference in training data (skewed toward smaller, more urban parcels) and conservation acquisitions (which include a larger share of large and undeveloped parcels). It points to the need for further study of the circumstances under which parcel size is a suitable predictor in conservation cost modeling (13).
- Any tested changes to the parameterization of the extremely random tree regressor either decrease predictive power or increase bias in the validation sample.

Replication of Results from Previous Studies. To reproduce the floodplain acquisition cost estimates from Johnson et al. (13), I follow their methodology and identify natural, private, undeveloped floodplains as the intersection of natural land uses in the National Land Cover Database (42), 100-y flood return layers (8), and private ownership (PLACES). Cost estimates are based on the model that only includes vacant sales ("PLACES FMV: vacant").

I reproduce the cost estimates for climate-sensitive protection strategies for species from Lawler et al. (11) using the datasets made available via Dryad (48) and the MARXAN conservation planning software (49). PLACES cost estimates are based on the model for vacant sales ("PLACES FMV: vacant"). In line with the published maps, I use targets of 100 planning units for each species occurring in ≥100 planning units and all planning units for species with <100 planning units. Runs with species refugia include 100% of all refugia. For each scenario, I identify the lowest species penalty factors that ensure that all conservation targets are reached. Each scenario is run with 10 iterations. I quantify the magnitude of changes to spatial priorities induced by PLACES FMV estimates using three indicators (*SI Appendix*, Table S2): 1) mean absolute difference in selection frequency, including all planning units selected at least once in any scenario; 2) percentage loss of top spatial priorities, defined as the percentage of planning units frequently selected in scenarios of Lawler et al. (>50% selection frequency) that were not identified as priorities in the scenarios with updated cost data (i.e., whose selection frequency was reduced by an absolute value of at least 50%); and 3) percentage gain of top spatial priorities, defined as the percentage of planning units frequently selected in the scenarios using PLACES FMV cost data (>50% selection frequency) that had not been identified as priorities in the scenarios of Lawler et al. (i.e., that gained at least 50% in selection frequency).

Data Availability. Rasterized cost estimates, validation datasets, and data used in the published figures are available from Dryad at https://datadryad.org/stash/dataset/doi:10.5061/dryad.np5hqbzq9 (50). Improvements, updates, and extensions to the US land value estimates will be hosted at placeslab.org/fmv_usa.

1. K. A. Wilson, M. F. McBride, M. Bode, H. P. Possingham, Prioritizing global conservation efforts. *Nature* **440**, 337–340 (2006).
2. F. Montesino Pouzols *et al.*, Global protected area expansion is compromised by projected land-use and parochialism. *Nature* **516**, 383–386 (2014).
3. C. Banks-Leite *et al.*, Using ecological thresholds to evaluate the costs and benefits of set-asides in a biodiversity hotspot. *Science* **345**, 1041–1045 (2014).
4. R. A. Fuller *et al.*, Replacing underperforming protected areas achieves better conservation outcomes. *Nature* **466**, 365–367 (2010).
5. A. Baccini *et al.*, Tropical forests are a net carbon source based on aboveground measurements of gain and loss. *Science* **358**, 230–234 (2017).
6. A. Guisan *et al.*, Predicting species distributions for conservation decisions. *Ecol. Lett.* **16**, 1424–1435 (2013).
7. S. A. Parks, C. Carroll, S. Z. Dobrowski, B. W. Allred, Human land uses reduce climate connectivity across North America. *Glob. Chang. Biol.* **26**, 2944–2955 (2020).
8. O. E. J. Wing *et al.*, Estimates of present and future flood risk in the conterminous United States. *Environ. Res. Lett.* **13**, 034023 (2018).
9. J. C. Withey *et al.*, Maximising return on conservation investment in the conterminous USA. *Ecol. Lett.* **15**, 1249–1256 (2012).
10. P. R. Armsworth *et al.*, Allocating resources for land protection using continuous optimization: Biodiversity conservation in the United States. *Ecol. Appl.* **30**, e02118 (2020).
11. J. J. Lawler *et al.*, Planning for climate change through additions to a national protected area network: implications for cost and configuration. *Philos. Trans. R Soc. B Biol. Sci.* **375**, 20190117 (2020).
12. D. Stralberg, C. Carroll, S. E. Nielsen, Toward a climate-informed North American protected areas network: Incorporating climate-change refugia and corridors in conservation planning. *Conserv. Lett.* **13**, e12712 (2020).
13. K. A. Johnson *et al.*, A benefit–cost analysis of floodplain land acquisition for U.S. flood damage reduction. *Nat. Sustain.* **3**, 56–62 (2020).
14. P. R. Armsworth, Inclusion of costs in conservation planning depends on limited datasets and hopeful assumptions. *Ann. N. Y. Acad. Sci.* **1322**, 61–76 (2014).
15. C. N. Jenkins, K. S. Van Houtan, S. L. Pimm, J. O. Sexton, U.S. protected lands mismatch biodiversity priorities. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 5081–5086 (2015).
16. C. Carroll *et al.*, Scale-dependent complementarity of climatic velocity and environmental diversity for identifying priority areas for conservation under climate change. *Glob. Change Biol.* **23**, 4508–4520 (2017).
17. A. Ando, J. Camm, S. Polasky, A. Solow, Species distributions, land values, and efficient conservation. *Science* **279**, 2126–2128 (1998).
18. D. A. Newburn, P. Berck, A. M. Merenlender, Habitat and open space at risk of land-use conversion: Targeting strategies for land conservation. *Am. J. Agric. Econ.* **88**, 28–42 (2006).
19. S. Polasky *et al.*, Where to put things? Spatial land management to sustain biodiversity and economic returns. *Biol. Conserv.* **141**, 1505–1524 (2008).
20. P. Geurts, D. Ernst, L. Wehenkel, Extremely randomized trees. *Mach. Learn.* **63**, 3–42 (2006).
21. U.S. Department of Agriculture, Agricultural land values and cash rents—final estimates. https://usda.library.cornell.edu/concern/publications/5425k968s?locale=en. Accessed 26 January 2019.

22. R. N. Lubowski, A. J. Plantinga, R. N. Stavins, Land-use change and carbon sinks: Econometric estimation of the carbon sequestration supply function. *J. Environ. Econ. Manage.* **51**, 135–152 (2006).

23. O. Venter *et al*., Global terrestrial human footprint maps for 1993 and 2009. *Sci. Data* **3**, 160067 (2016).

24. Zillow, ZTRAX: Zillow Transaction and Assessor Dataset, 2019-Q4. (2019). www.zillow.com/ztrax. Accessed 21 August 2019.

25. The National Conservation Easement Database. https://www.conservationeasement.us/. Accessed 10 April 2020.

26. A. Nelson, *Travel Time to Major Cities: A Global Map of Accessibility* (Global Environment Monitoring Unit, Joint Research Centre of the European Commission, 2008).

27. K. Hildreth, Congress Passes Great American Outdoors Act with bipartisan support. https://www.ncsl.org/ncsl-in-dc/publications-and-resources/house-passes-great-american-outdoors-act-magazine2020.aspx . Accessed 8 October 2020.

28. J. P. Cohen, C. C. Coughlin, Spatial hedonic models of airport noise, proximity, and housing prices. *J. Reg. Sci.* **48**, 859–878 (2008).

29. J. Wu, E. G. Irwin, Optimal land development with endogenous environmental amenities. *Am. J. Agric. Econ.* **90**, 232–248 (2008).

30. Zillow, Zillow Prize Kaggle competition. (2020). https://www.zillow.com/z/info/zillow-prize/. Accessed 23 September 2020.

31. O. T. Coomes, G. K. Macdonald, Y. Le, P. D. E. Waroux, Geospatial land price data : A public good for global change science and policy. *BioScience* **68**, 481–484 (2018).

32. Microsoft, U.S. building footprints. (2018). https://github.com/microsoft/USBuildingFootprints. Accessed 2 February 2020.

33. U.S. Census Bureau, 2019 TIGER/Line Shapefiles (machine-readable data files). (2019). https://www.census.gov/cgi-bin/geo/shapefiles/index.php. Accessed 19 May 2020.

34. S. Manson, J. Schroeder, D. Van Riper, S. Ruggles, IPUMS National Historical Geographical Information System, version 13.0, 2018; doi:10.18128/D050.V13.0.

35. U.S. Geological Survey, Gap Analysis Project (GAP), 2018, Protected Areas Database of the United States (PAD-US): U.S. Geological Survey data release, https://doi.org/10.5066/P955KPLE.

36. Harvard Forest (2020) New England protected open space. https://zenodo.org/record/3606763#.X4ys0dR7mM8/. Accessed 16 October 2020.

37. Colorado Natural Heritage Program, The Colorado Ownership, Management, and Protection Map (COMaP). https://cnhp.colostate.edu/projects/comap/. Accessed 29 August 2019.

38. U.S. Geological Survey, National Elevation Dataset. (2017). https://lta.cr.usgs.gov/NED. Accessed 6 December 2018.

39. U.S. Fish & Wildlife Service, National Wetlands Inventory: Seamless Wetlands Data. https://www.fws.gov/wetlands/data/data-download.html. Accessed 6 December 2018.

40. U.S. Geological Survey, National Hydrography Dataset USGS national map. (2017). https://nhd.usgs.gov/NHD_High_Resolution.html. Accessed 29 August 2019.

41. Esri, North America Water polygons. (2009). https://bit.ly/2vFchmY. Accessed 6 December 2018.

42. C. Homer *et al*., Completion of the 2011 National Land Cover Database for the conterminous United States—Representing a decade of land cover change information. *Photogramm. Eng. Remote Sensing* **81**, 345–354 (2015).

43. The Trust for Public Land, *Conservation Almanac* (Boston, MA, 2019).

44. N. J. Sutton, S. Cho, P. R. Armsworth, A reliance on agricultural land values in conservation planning alters the spatial distribution of priorities and overestimates the acquisition costs of protected areas. *Biol. Conserv.* **194**, 2–10 (2016).

45. A. J. Plantinga, R. N. Lubowski, R. N. Stavins, The effects of potential land development on agricultural land prices. *J. Urban Econ.* **52**, 561–581 (2002).

46. M. A. Davis, W. D. Larson, S. D. Oliner, J. Shui, (2019) The price of residential land for counties, ZIP codes, and census tracts in the United States. Federal Housing Finance Agency Working Paper 19-01. https://www.fhfa.gov/PolicyProgramsResearch/Research/PaperDocuments/wp1901.pdf. Accessed 16 October 2020.

47. U.S. Bureau of Labor Statistics, Consumer Price Index for all urban consumers: all items in U.S. city average. (2019). https://fred.stlouisfed.org/series/CPIAUCSL. Accessed 15 December 2019.

48. J. J. Lawler *et al*., Data from: Planning for climate change through additions to a national protected area network: Implications for cost and configuration. https://datadryad.org/stash/dataset/doi:10.5061/dryad.jm63xsj6d. Accessed 10 April 2020.

49. I. R. Ball, H. P. Possingham, M. E. Watts, "Marxan and relatives: Software for spatial conservation prioritization" in *Spatial Conservation Prioritization Quantitative Methods and Computational Tools*, A. Moilanen, K. A. Wilson, H. P. Possingham, Eds. (Oxford University Press, 2009), pp. 185–195.

50. C. Nolte, Data for: High-resolution land value maps reveal underestimation of conservation costs in the United States. *Dryad*. https://datadryad.org/stash/dataset/doi:10.5061/dryad.np5hqbzq95. Deposited 8 October 2020.

SUSTAINABILITY SCIENCE

Nolte

www.manaraa.com