Improving estimates of land protection costs in a tropical biodiversity hotspot

Christoph Nolte[1,2*], Ana Reboredo Segovia[1], Jose Manuel Ochoa-Quintero[3,4], and Jaime Burbano-Girón[3,5]

Accurate estimates of the costs of land protection are useful for understanding where biodiversity conservation goals can be achieved at the lowest cost to society. However, because of the scarcity of high-quality cost maps for tropical countries, conservation planning studies often ignore cost or rely on untested proxies, such as agricultural rent or land-use intensity. Here, we show how analysts can estimate land protection costs using original data of public land acquisitions, global predictor datasets, and simple machine-learning models. For the Colombian Andes, a global biodiversity hotspot, we found that the principal driver of the cost of land protection is urban proximity, not agricultural rent. We derived cost estimates that predict public land protection costs more accurately than available cost proxies and identified new protection priorities for 143 threatened bird species. A more systematic collection of cost records of land protection will help inform public decisions on national and global biodiversity protection priorities.

Front Ecol Environ 2024; 22(2): e2626, doi:10.1002/fee.2626

The cost associated with long-term land protection—such as the value of foregone uses and the public budgets leveraged to compensate for them—is an important variable for guiding biodiversity conservation decisions. Public agencies and nonprofit organizations often have only limited funding available to achieve ambitious land protection goals. Targets to effectively conserve 30% of the Earth's lands and waters by 2030 (the global "30 × 30" conservation initiative), as endorsed by the parties to the Convention on Biological Diversity in the Kunming-Montreal Global Biodiversity Framework, have not yet materialized in equivalent increases in public funding for new land protection. Aware of such constraints, conservation planners often seek to integrate cost information into their spatial planning (eg to identify combinations of sites where ecologically representative species' habitats and ecosystems can be protected or restored at the lowest cost; Strassburg *et al*. 2020).

To be useful for conservation planning, estimates of land protection costs need to accurately capture local differences in costs, often across large and heterogeneous landscapes. Capturing local variation is important because the cost of land protection tends to be correlated with opportunity costs from a range of alternative land uses (residential, agricultural, extractive, etc.), which in turn can vary substantially between land parcels as a function of climate, terrain, infrastructure, accessibility, demographics, property rights, and other attributes. Analysts wanting to provide guidance for real-life decision making also need to know how accurately their cost estimates can predict the actual costs that conservation actors face in practice; however, such validation of cost estimates is rare (Armsworth 2014).

To our knowledge, empirically validated maps of protection costs do not exist for most tropical regions. As a result, many proposals for large-scale biodiversity protection priorities ignore local variation in protection cost (Brum *et al*. 2017; McGowan *et al*. 2020) or rely on untested cost proxies. The most common cost proxies are: (1) agricultural rent (Naidoo and Iwamura 2007) and (2) land-use intensity, such as the global human footprint (GHF; Stralberg *et al*. 2020) or global human modification (GHM; Kennedy *et al*. 2019) indices. Empirical studies have already cast doubt on the ability of such proxies to capture key variations in the cost of public and nonprofit conservation acquisitions (eg Sutton *et al*. 2016). For instance, in the contiguous US, empirical cost estimates derived from data on private land sales offered stronger predictive power than proxies (Nolte 2020).

These findings raise important questions about the choice of protection cost estimates for tropical conservation planning. How well can empirical cost models predict the spatial variation in the cost of long-term land protection in tropical biodiversity hotspots? Do these models provide greater accuracy than cost proxies such as agricultural rent and land-use intensity? And do any differences in accuracy matter for the identification of cost-effective protection priorities?

Here, we explore these questions through examination of a case study of the Colombian Andes, a globally important biodiversity hotspot that is exceptionally rich in threatened and endemic species (CEPF 2015). Its remaining habitats are vulnerable to shrinking climate envelopes (Velásquez-Tibatá

[1]*Department of Earth & Environment, Boston University, Boston, MA* [*]*(chrnolte@bu.edu); [2]Faculty of Computing & Data Sciences, Boston University, Boston, MA; [3]Instituto de Investigación de Recursos Biológicos Alexander von Humboldt, Bogotá, Colombia; [4]Instituto de Biociencias, Universidad Federal de Mato Grosso do Sul, Campo Grande, Brasil; [5]The Nature Conservancy, Bogotá, Colombia*

et al. 2013; Báez *et al.* 2016) and new development projects following local-scale declines in civil conflict (Baptiste *et al.* 2017). Colombia is also home to the longest standing publicly financed land acquisition program for conservation in the tropics. Since 1993, a national law requires local governments to allocate 1% of their annual income to the acquisition and conservation of private properties in watersheds within their jurisdictions. This legal mandate has led to thousands of public land acquisitions over nearly three decades, offering a rich empirical context to study the predictability of spatial variation in the cost of protection actions.

Methods

We compiled a novel spatial database of public land acquisitions for conservation in Colombia. Using a combination of email requests, mailed letters, telephone surveys, and in-person visits, our research team collected data on the costs, dates, and geospatial boundaries of publicly financed land purchases for conservation purposes from regional environmental authorities (CARs), as well as from departmental and municipal governments (Reboredo Segovia *et al.* 2023). After geoprocessing and data cleaning (Appendix S1: Panel S1), we retained data for 2094 public land acquisitions in 333 municipalities from 106 distinct government entities.

We first used these data to develop spatially predictive cost models for the Colombian Andes. The predicted variable of interest was the per-hectare acquisition cost in 2020 Colombian pesos (COP, deflated, log-transformed). We considered a range of globally available datasets as predictors: nighttime lights, built area, population density, distance to roads, travel time to cities, slope, agricultural rent, forest cover, and civil conflict (Appendix S1: Panel S1). After testing several alternative modeling strategies (Appendix S1: Panel S1), we selected a simple spatial machine-learning algorithm—the extremely randomized tree regressor (Geurts *et al.* 2006)—to generate landscape-wide cost estimates for our study region at 1-km^2 resolution.

The new cost estimates were used as an input to address an illustrative conservation planning problem. Specifically, we identified the locations for protecting the potential distributions of 143 threatened bird species at the lowest overall acquisition cost. We chose to focus on birds because their potential distributions in the Colombian Andes are better understood than those of other taxa (Renjifo *et al.* 2014, 2016). To illustrate outcomes for a hypothetical 30 × 30 planning scenario, we defined "ambitious" species-specific area-based targets as a function of the overall size of each species' potential distribution and calibrated these targets such that they led to an overall protection of 30% of our study area (Appendix S1: Panel S1).

Finally, we evaluated how results would have changed if we had omitted the cost estimation step and had instead relied on commonly used cost proxies. We considered six proxies used in published large-scale conservation planning studies or national priority setting: three estimates of agricultural rent and three land-use intensity indices, with each group composed of two global estimates and one national estimate (Appendix S1: Table S1 and Panel S1). We assessed the extent to which each proxy captured variation in the observed cost data and examined how the overall cost and location of proposed protected area networks for threatened birds would have changed if we had relied on cost proxies or ignored any spatial variation in costs altogether.

Results

Cost estimation

Our data on public land acquisitions in the Colombian Andes spanned a broad range of geographic and temporal contexts: acquisitions that occurred continuously between 1993 and 2019 (median: 2007) near urban centers, in remote rural areas, and anywhere in between (Figure 1a). Sizes of acquired parcels ranged from 0.1 to 4494 ha (median: 13 ha, mean: 43 ha). Per-hectare acquisition prices were positively skewed, with a median of 5.6 million COP per hectare (~US$1173 per hectare) and a mean of COP 50.5 million per hectare (~US$15,567 per hectare). High prices were visibly clustered in and around metropolitan areas, most notably Colombia's capital and largest city, Bogotá, and second-largest city, Medellín.

Indicators of urban proximity were the strongest predictors of observed land acquisition costs in our sample. Nighttime light intensity alone accounted for >50% of overall predictor importance in our preferred model, a greater proportion than the other 13 predictors combined (Figure 1b). It remained the dominant predictor of costs across a range of alternative modeling frameworks, predictor selection strategies, and metrics of importance (Appendix S1: Figure S1 and Panel S1). Slope, travel time to cities, population density, and distance to roads were also regularly selected as predictors by the best-performing models. Although local estimates of agricultural rent, a widespread cost proxy, contributed predictive power to our preferred model as well, this factor was not always selected in spatial cross-validation and never ranked among the top two predictors in any model (Appendix S1: Figure S1).

Our preferred model captured a large percentage (75%) of the in-sample variation in log-transformed per-area acquisition cost (Figure 1d) when using "out-of-bag" predictions (ie predictions from models that have not seen the predicted observation). Although capturing much of the observed variation, predictions were not very precise. In random cross-validation, we obtained a prediction error (root mean square error or RMSE) of 0.72, which implied a 95% confidence interval (CI) of –76% to +314% around predicted per-area prices. In spatial cross-validation, the RMSE was 1.3 (95% CI: –92.3% to

+1201%). These prediction errors applied to 80% of our study area (Figure 1c; Appendix S1: Panel S1) and were likely worse in remote regions where little training data exists. The resulting cost maps illustrated that the estimated costs of land acquisition in the Colombian Andes varied by several orders of magnitude between urban areas and remote rural regions (Figure 1c).

Conservation priorities for birds

Using this cost data, we estimated that ambitious 30 × 30 protection targets for threatened birds (Figure 2a) could be achieved most cost-effectively by protecting an additional 20.4% of the study region at an estimated cost of COP 25.3 trillion (~US$7.78 billion; Figure 2b). This estimate was a sum of pixel-level cost estimates and ignored the effects of such a large acquisition program on land prices. If the cost variation were to be ignored in the prioritization, the proposed expansion of the protected network would be 2.4 times as expensive (COP 61 trillion; Figure 3c). Three-quarters (74.7%) of locations selected under an equal-cost assumption were retained by the cost-effective scenario, notably in very species-rich regions, such as the southern part of the Andes (Vélez *et al*. 2021) and along the Western Cordillera (Ocampo-Peñuela and Pimm 2014), or in regions supporting threatened species with restricted ranges such as the recently described chestnut-capped piha (*Lipaugus weberi*) in the mountains north of Medellín (Cuervo *et al*. 2001). Yet accounting for cost variation also shifted 25.3% of priority areas away from expensive cities, peri-urban regions, and valuable agricultural areas and toward more remote parts of the landscape (Figure 2c).

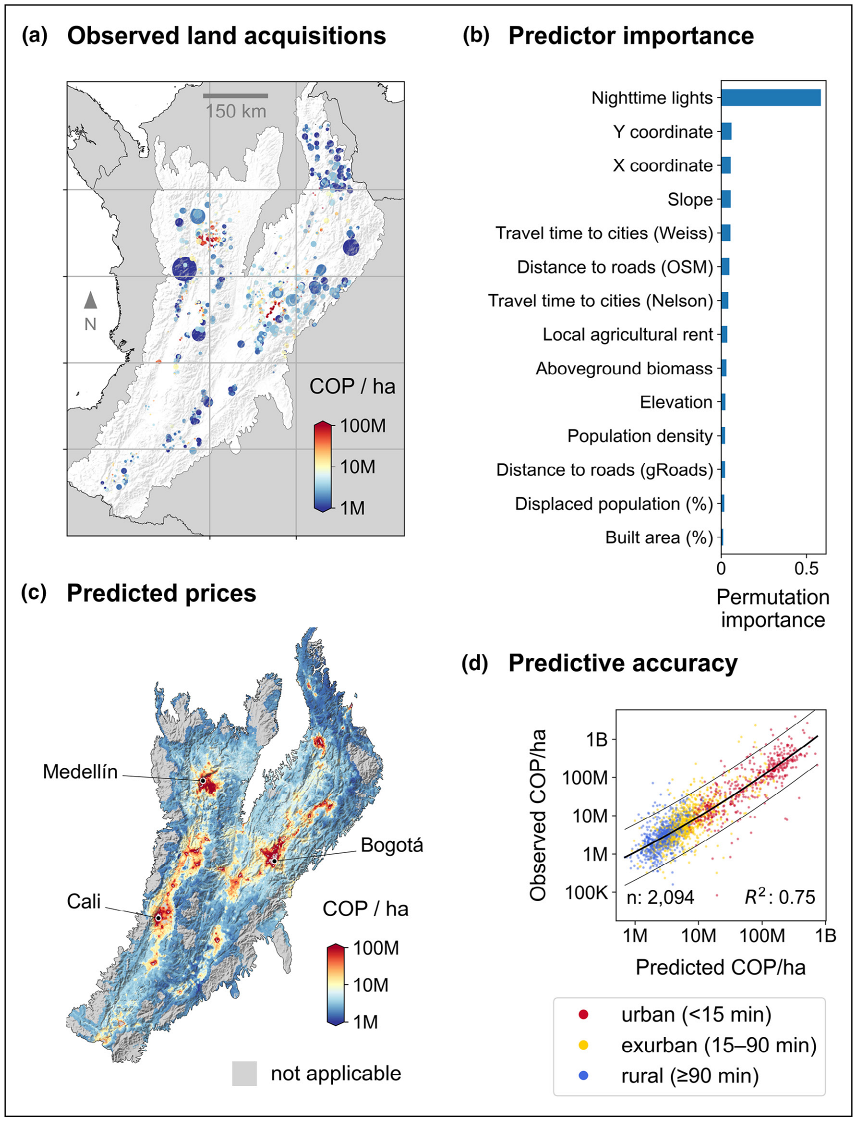

Figure 1. Prediction of land acquisition costs for conservation purposes in the Colombian Andes (main model). (a) Location, size, and cost of 2094 publicly financed land acquisitions. Marker size is proportional to area and enlarged. Grid shows spatial blocks used in cross-validation. COP = 2020 Colombian pesos. (b) Predictor importance. Weiss = Weiss *et al*. (2018), OSM = OpenStreetMap, Nelson = Nelson (2008), and gRoads = gROADSv1 (see also Appendix S1: Panel S1). (c) Landscape-wide predictions of land acquisition costs for the study region. Hillshade added. Gray areas are outside the model's area of applicability (Appendix S1: Panel S1). (d) Predictive accuracy. Colors distinguish urban (red), exurban (yellow), and rural (blue) acquisitions by travel time to cities (Appendix S1: Panel S1). Black lines represent predicted values (central bold line) and 95% prediction intervals (thin lines).

Examining the use of cost proxies

We found cost proxies to be only weakly correlated with observed acquisition costs, especially in rural areas. Agricultural rent proxies were particularly weak predictors (R^2: 0.05–0.10) and, in some cases, negatively correlated with cost (Figure 3a). Global land-use intensity indices performed comparatively well (R^2: 0.55 for GHM, 0.56 for GHF), presumably because their inclusion of indicators of urban proximity (nighttime lights, distance to roads, population density) allowed them to capture much of the observed cost variation along rural–urban gradients (Appendix S1: Figures S2 and S3). However, in remote rural areas—here defined as any parcel located farther than 90 travel minutes away from cities (Appendix S1: Panel S1)—all six proxies had almost no explanatory power (R^2: 0.00–0.04, compared to 0.27 for the empirical model; Figure 3b; Appendix S1: Figure S4). This finding casts serious doubt on the validity of all six proxies of protection cost: because a cost-minimizing optimization will tend to gravitate toward low-cost rural areas, misestimations of cost in these regions can have a large influence on spatial patterns.

Consistent with this hypothesis, we found that the spatial priorities identified by our conservation planning exercise changed, sometimes substantially, when using cost proxies

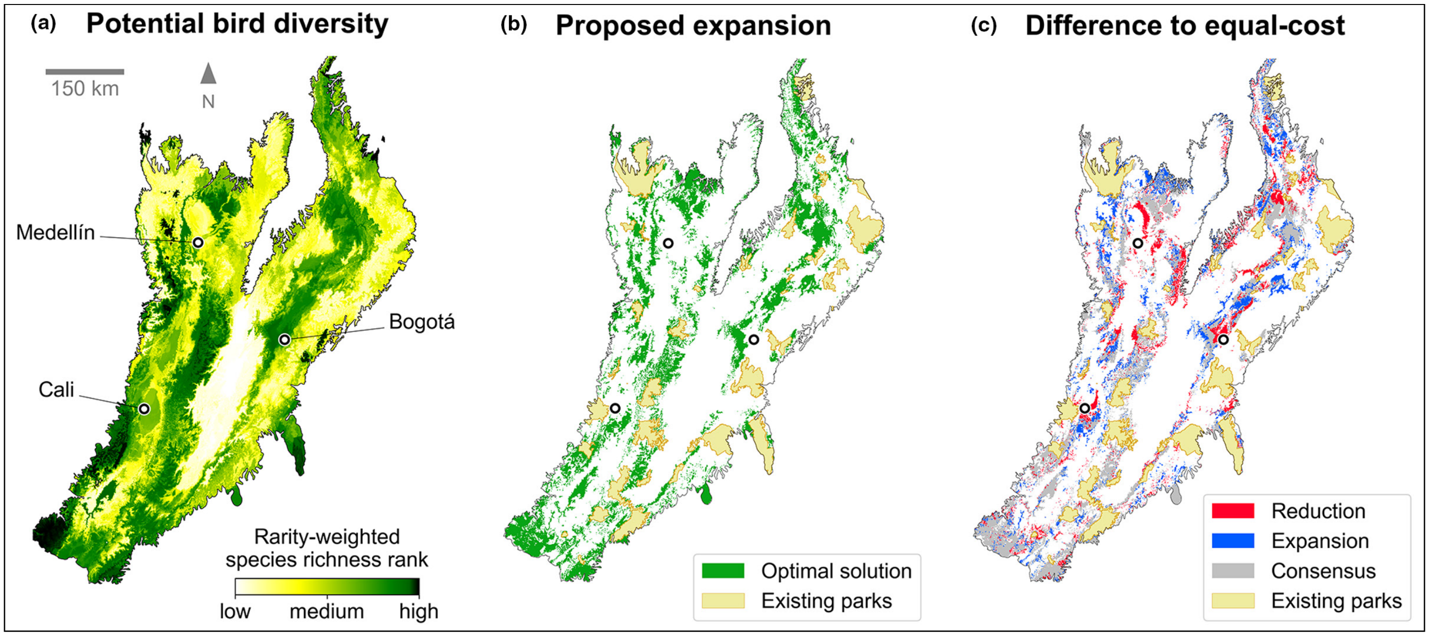

Figure 2. Illustrative conservation planning for 143 threatened bird species in the Colombian Andes, using "ambitious" bird-specific area targets (see Appendix S1: Panel S1). (a) Rank of rarity-weighted potential bird species richness (sum of all potential distributions, inversely weighted by their area). (b) Spatial locations of the most cost-effective protected area expansion. (c) Changes to spatial locations when empirical cost estimates are included, as compared to a scenario in which costs are homogeneous across the region. Small black circles show locations of Colombia's three largest metro areas (Bogotá, Medellín, and Cali).

instead of empirical models as the input to the reserve selection algorithm (Figure 3, c and d). Switching to agricultural rent estimates prompted the algorithm to propose priorities for expansion, which, based on our own cost estimates (Figure 1c), would be 2.3–3.9 times as expensive as the optimal solution (Figure 3c), largely due to the inclusion of urban areas. In contrast, global land-use intensity proxies correctly identified urban areas as expensive and instructed the algorithm to avoid them if possible, resulting in priorities whose estimated costs were markedly similar to ours (1.2 times as expensive). However, all six cost proxies led to changes in the proposed spatial priorities: between 21.8% and 32.7% of locations identified as priorities in our "ambitious" scenario (Figure 2b) were not selected when proxies were used (Figure 3d; Appendix S1: Figure S5). These shifts became even more pronounced (37.2% to 52.8%) when targets were scaled back to "modest" scenarios (Figure 3d; Appendix S1: Panel S1). Lower targets imply greater spatial flexibility in where targets are met and therefore greater responsiveness to estimates of protection cost.

Discussion

Taken together, our results suggest that (1) priorities for cost-effective conservation actions in tropical biodiversity hotspots can be sensitive to underlying protection cost estimates; (2) widespread cost proxies miss important spatial heterogeneities in the cost of long-term land protection actions, affecting the cost-effectiveness of proposed protection plans; and (3) efforts to collect and synthesize data on observed protection costs permit the development of better cost estimates. Although these insights are derived from only one tropical biodiversity hotspot and one type of protection action (land acquisition), we presume that they might also be valid for other tropical landscapes and actions.

The primary takeaway of our analysis is that conservation planners should choose empirical cost estimates over untested cost proxies (or equal-cost assumptions) when searching for cost-effective priorities for conservation actions. In particular, we caution against the practice of using agricultural rent estimates as the only proxy of long-term land protection costs in landscapes that include major human settlements. In such regions, non-agricultural benefits from land ownership (such as residential rights) can be the dominant drivers of land values and need to be included in the estimation of protection costs.

Land-use intensity indices capture a large share of the urban–rural price gradient and are therefore preferable over agricultural rent estimates in these regions. However, even land-use intensity indices should only be used with great caution, as they can fail to capture observed cost variation found within rural areas (Figure 1c), where many of the cost-effective priority areas are located. Furthermore, relationships between land-use indices and observed cost are nonlinear (Appendix S1: Figure S3; Nolte 2020) and usually not known ex ante, and therefore need to be estimated and calibrated with empirical data on observed costs.

Finally, we note that a conservation planner's preference for accurate cost estimates should be stronger in regions with tighter budget constraints for conservation programs (for example, throughout much of the tropics): as seen in Figure 3c, cost-effective priorities become more sensitive to cost estimates, and subsequently more vulnerable to errors and biases in cost estimations, if more modest conservation targets are chosen.

Land acquisitions will likely continue to be part of Colombia's efforts to conserve and protect biodiversity, water, carbon, and other ecosystem goods in the foreseeable future. Empirical cost maps enable planners to predict where conservation funds can be spent most cost-effectively. These maps can also highlight spatial discrepancies between cost-effective priorities (Figure 2) and the local availability of conservation funding, which tends to be concentrated in wealthy municipalities that are usually more expensive to protect (Reboredo Segovia *et al.* 2023). Policy options to reduce such discrepancies include, for instance, a legal change to allow for greater flexibility in where Colombia's "1%" of conservation funding can be spent (Armsworth *et al.* 2023), or the development of a national funding mechanism that concentrates its efforts on locations where cost-effective priorities exist but where local government income is insufficient to achieve protection goals.

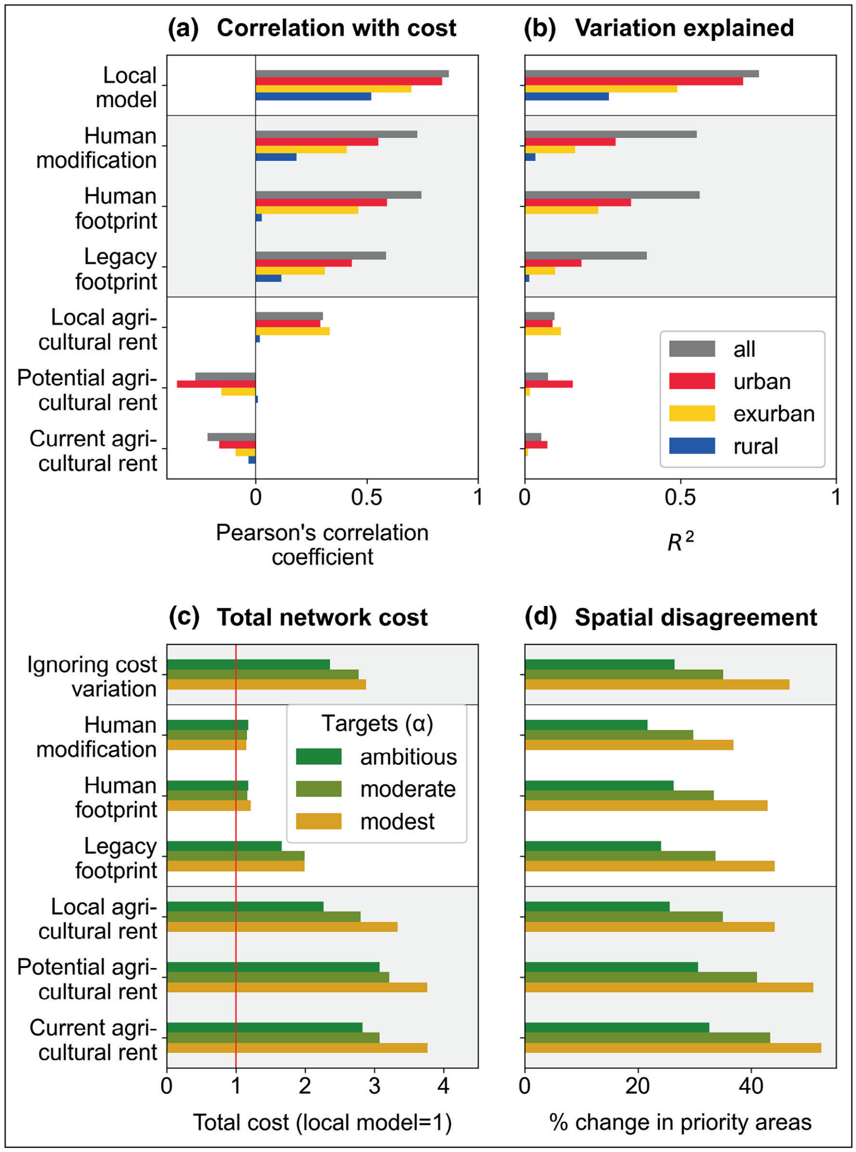

Figure 3. Predictive accuracy of conservation cost datasets and implications for using proxies when identifying priority networks. (a) Pearson's correlation of each cost dataset with the observed acquisition (in log-transformed Colombian pesos per hectare) for urban (red), exurban (yellow), and rural (blue) acquisitions (see Appendix S1: Panel S1). (b) Variation explained (R^2) by squared linear regression of observed costs in each cost dataset. (c) Increase in estimated total network cost (predictions of local model) if optimizations ignore cost variation or use cost proxies. (d) Percentage area of the most cost-effective protected area network (Figure 2b) omitted when switching cost proxies.

Collaboration as the way forward

Analysts interested in improving empirical cost estimates face lower technological barriers today than ever before. Digital cadaster datasets are widespread; predictor datasets continue to be refined due to advances in remote sensing and open data initiatives; computational costs are trending downward; and "off-the-shelf" machine-learning algorithms, such as the extremely randomized tree regressor we used here, are increasingly accessible and user-friendly. Although prediction errors remain too high for the models presented here to directly inform investment choices on individual parcels, errors can likely be reduced with improved understanding and integration of missing variables, including information on local regulations, land tenure security, landowner willingness-to-sell, and suitability for high-value crops (eg coffee [*Coffea* spp]).

However, for most conservation organizations and practitioners, the creation of empirical cost estimates will remain challenging. Access to data on observed cost and its drivers will vary from region to region. Many conservation planning applications need cost estimates for actions other than acquisitions, such as easements, temporary incentives, enforcement, restoration, and management. Spatial variation in cost types other than capital cost (eg consumables, labor, overhead, non-monetary) could also be substantial. The development and long-term maintenance of such estimates and expertise might not be cost-effective for many organizations, especially the most underfunded ones.

We therefore believe that an international collaborative strategy offers a particularly promising pathway toward a more widespread integration of empirical cost estimates into conservation planning. Future attempts to develop global conservation plans, action portfolios, and financing agreements will likely benefit from early efforts to compile, standardize, and share data on the costs of long-term land protection actions and other types of conservation transactions (Coomes *et al.* 2018; Iacona *et al.* 2018). Substantial synergies are likely possible by (1) pooling computational, statistical, ecological, and economic expertise; (2) defining shared standards, software tools, and privacy protocols to merge and share cost datasets from a range of conservation organizations; and (3) maintaining a data archive for standardized parcel-level data of costs and predictors. Such an information hub could be affiliated with an internationally recognized collaborative center (for instance, the World Conservation Monitoring Centre of the UN Environment Programme) and build on the expertise of related efforts (for example, Arizona State University's Intervention Cost Data Portal, http://web.asu.edu/conservation-cost-data). As human society faces increasingly arduous questions concerning who should give up which benefits to help other species survive and thrive, it deserves accurate and interpretable information about the trade-offs between options.

Acknowledgements

We thank P Armsworth, T Iwamura, C Kennedy, N Ocampo-Peñuela, A Pollack, BA Simmons, and participants of the PLACES lab for useful feedback on earlier versions of this manuscript. We thank D Amaya Torres, JF Gomez Candamil, V Lopez Gamboa, C Matallana, D Meza Chavez, and officials in dozens of government agencies for invaluable support with data acquisition. We thank CA Correa Ayram, T Iwamura, and O Venter for sharing cost proxy data. This research was supported by Boston University and the National Aeronautics and Space Administration (grant #80NSSC20K1486).

Data Availability Statement

Data and code are available at https://doi.org/10.5061/dryad.2bvq83brr.

References

Armsworth PR. 2014. Inclusion of costs in conservation planning depends on limited datasets and hopeful assumptions. *Ann NY Acad Sci* **1322**: 61–76.

Armsworth PR, Dilkina B, Fargione J, *et al.* 2023. Multiplying the impact of conservation funding using spatial exchange rates. *Front Ecol Environ*; doi.org/10.1002/fee.2678.

Báez S, Jaramillo L, Cuesta F, and Donoso DA. 2016. Effects of climate change on Andean biodiversity: a synthesis of studies published until 2015. *Neotrop Biodivers* **2**: 181–94.

Baptiste B, Pinedo-Vasquez M, Gutierrez-Velez VH, *et al.* 2017. Greening peace in Colombia. *Nat Ecol Evol* **1**: 102.

Brum FT, Graham CH, Costa GC, *et al.* 2017. Global priorities for conservation across multiple dimensions of mammalian diversity. *P Natl Acad Sci USA* **114**: 7641–46.

CEPF (Critical Ecosystem Partnership Fund). 2015. Ecosystem profile: tropical Andes biodiversity hotspot. Arlington, VA: CEPF.

Coomes OT, MacDonald GK, and le Polain de Waroux Y. 2018. Geospatial land price data: a public good for global change science and policy. *BioScience* **68**: 481–84.

Cuervo AM, Salaman PGW, Donegan TM, and Ochoa JM. 2001. A new species of piha (Cotingidae: *Lipaugus*) from the Cordillera Central of Colombia. *Ibis* **143**: 353–68.

Geurts P, Ernst D, and Wehenkel L. 2006. Extremely randomized trees. *Mach Learn* **63**: 3–42.

Iacona GD, Sutherland WJ, Mappin B, *et al.* 2018. Standardized reporting of the costs of management interventions for biodiversity conservation. *Conserv Biol* **32**: 979–88.

Kennedy CM, Oakleaf JR, Theobald DM, *et al.* 2019. Managing the middle: a shift in conservation priorities based on the global human modification gradient. *Glob Change Biol* **25**: 811–26.

McGowan J, Beaumont LJ, Smith RJ, *et al.* 2020. Conservation prioritization can resolve the flagship species conundrum. *Nat Commun* **11**: 994.

Naidoo R and Iwamura T. 2007. Global-scale mapping of economic benefits from agricultural lands: implications for conservation priorities. *Biol Conserv* **140**: 40–49.

Nelson A. 2008. Travel time to major cities: a global map of accessibility. Ispra, Italy: Joint Research Centre of the European Commission.

Nolte C. 2020. High-resolution land value maps reveal underestimation of conservation costs in the United States. *P Natl Acad Sci USA* **117**: 29577–83.

Ocampo-Peñuela N and Pimm SL. 2014. Setting practical conservation priorities for birds in the Western Andes of Colombia. *Conserv Biol* **28**: 1260–70.

Reboredo Segovia AL, Matallana C, Amaya Torres D, *et al.* 2023. Implementing land acquisitions for watershed services in the tropical Colombian Andes produces marginal progress for advancing area-based conservation objectives. *Conserv Lett* **16**: e12932.

Renjifo LM, Amaya-Villarreal ÁM, Burbano-Girón J, and Velásquez-Tibatá J (Eds). 2016. Libro rojo de aves de Colombia (vol II): ecosistemas abiertos, secos, insulares, acuáticos continentales, marinos, tierras altas del Darién y Sierra Nevada de Santa Marta y bosques húmedos del centro, norte y oriente del país. Bogotá, Colombia: Pontificia Universidad Javeriana e Instituto Alexander von Humboldt.

Renjifo LM, Gómez MF, Velásquez-Tibatá J, *et al.* (Eds). 2014. Libro rojo de aves de Colombia (vol I): bosques húmedos de los Andes y la costa Pacífica. Bogotá, Colombia: Pontificia Universidad Javeriana e Instituto Alexander von Humboldt.

Stralberg D, Carroll C, and Nielsen SE. 2020. Toward a climate-informed North American protected areas network: incorporating climate-change refugia and corridors in conservation planning. *Conserv Lett* **13**: e12712.

Strassburg BBN, Iribarrem A, Beyer HL, *et al.* 2020. Global priority areas for ecosystem restoration. *Nature* **586**: 724–29.

Sutton NJ, Cho S, and Armsworth PR. 2016. 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.

Velásquez-Tibatá J, Salaman P, and Graham CH. 2013. Effects of climate change on species distribution, community structure, and conservation of birds in protected areas in Colombia. *Reg Environ Change* **13**: 235–48.

Vélez D, Tamayo E, Ayerbe-Quiñones F, *et al*. 2021. Distribution of birds in Colombia. *Biodivers Data J* **9**: e59202.

Weiss DJ, Nelson A, Gibson HS, *et al*. 2018. A global map of travel time to cities to assess inequalities in accessibility in 2015. *Nature* **553**: 333–36.

Supporting Information

Additional material can be found online at http://onlinelibrary.wiley.com/doi/10.1002/fee.2626/suppinfo