

**Spatially explicit mapping of ecological similarity  
across a large region using matching techniques.**

**Evan Muise<sup>1</sup>, Nicholas Coops<sup>1</sup>**

<sup>1</sup>Department of Forest Resource Management, University of British Columbia, Vancouver, 2424 Main  
Mall, British Columbia, Canada,

---

Corresponding author: Evan Muise, [evanmuis@student.ubc.ca](mailto:evanmuis@student.ubc.ca)

**Abstract****ABSTRACT****Plain Language Summary**

Probably am not doing a plain language summary

**0.1 Introduction**

A global biodiversity crisis is currently underway, driven by anthropogenic changes (Dirzo & Raven, 2003). Pressures such as climate change, land-use changes, and invasive species, are leading to species extinctions (Thomas et al., 2004; Urban, 2015) and the homogenization of biological communities (McGill et al., 2015). The Kunming-Montreal Global Biodiversity Framework (GBF) has been adopted as of December 2022, with the goal of restoring global biodiversity (*Report of the conference of the parties to the convention on biological diversity on the second part of its fifteenth meeting*, 2023). Targets within this framework include restoring 30% of all degraded ecosystems, and protecting 30% of of Earth’s terrestrial, inland water, and marine areas (*Report of the conference of the parties to the convention on biological diversity on the second part of its fifteenth meeting*, 2023). In the terrestrial environment, forest biomes have been shown to harbour the largest amount of biodiversity (Cardinale et al., 2012; Myers, 1988; Pimm & Raven, 2000), and provide key ecosystem services (Thompson et al., 2009). To provide these services, it is integral that these forest ecosystems are in good ecological condition, as represented by natural or near-natural levels of forest structure, forest function, and forest composition (Marín et al., 2021).

While understanding forest condition is a key aspect of understanding biodiversity and the provision of ecosystem services due to their inherent linkages (Cardinale et al., 2012; Marín et al., 2021), it is difficult to acquire suitable field-derived data across wide expanses of land due to the extreme financial and temporal costs associated with field campaigns. Remote sensing data, however, can provide a time-efficient, cost-effective alternative to field data, providing access to new spatially explicit and exhaustive datasets that can be linked to ecological condition, with additional metrics being proposed at a rapid pace (Pereira et al., 2013; Volker C. Radeloff et al., 2024; Skidmore et al., 2021). Advances in lidar technologies and modelling methods are allowing wall-to-wall estimates of forest stand structure to be generated across entire countries (Becker et al., 2023; Matasci, Hermosilla, Wulder, White, Coops, Hobart, Bolton, et al., 2018; Matasci, Hermosilla, Wulder, White, Coops, Hobart, & Zald, 2018), serving as a finer scale indicator of ecosystem structure than the often previously used landscape fragmentation metrics (Andrew et al., 2012). Productivity metrics have been used as a proxy for ecosystem function for many years (Pettorelli et al., 2005, 2018), with new Landsat derived datasets providing integrative annual estimates of energy availability at a 30 m spatial resolution (V. C. Radeloff et al., 2019; Volker C. Radeloff et al., 2024; Razenkova et al., n.d.). Remote sensing is quickly providing access to a plethora of datasets suitable monitoring the various facets of biodiversity and ecological condition (Noss, 1990). Using these datasets in conjunction with information on the location of known or assumed natural forests can allow researchers to identify similar forest stands or ecosystems, thus locating high quality forests across entire jurisdictions, even when they are facing anthropogenic pressures.

However, identifying natural forests is not a simple task. There are many definitions of what constitutes ‘naturalness’ in forest environments (Winter, 2012), and these difficulties are compounded by shifting baseline syndrome (Pinnegar & Engelhard, 2008) and a lack of comprehensive historical data to generate reference states (Baloguer et al., 2014). Other proposed methods to delineate baseline conditions include protected areas (Arcese & Sinclair, 1997), and empirical estimates of the reference

state, generated by modelling outcomes (oftentimes species abundances and occurrence) in the absence of anthropogenic disturbance (Nielsen et al., 2007).

In this study, we seek to identify natural forests of good ecological condition at a 30 m spatial resolution across British Columbia, Canada. We first identify reference states for ecosystems across the province as protected areas with little-to-no anthropogenic disturbances. We then use an expanded coarsened exact matching (CEM) approach to compare the entire forested landscape of British Columbia to a reference state which accounts for forest type as well as topographic and climatic information. We expand the CEM technique to allow for the the spatially explicit reconstruction of natural ecosystem similarity, even when a perfect match is not available. We further examine how the coverage of natural forest areas varies between protected and unprotected areas in British Columbia. These methods will allow managers to identify regions in need of restoration, or that could be suitable for protection, and could be expanded to other regions.

## 0.2 Methods

### 0.2.1 Data

#### 0.2.1.1 Natural forests

We defined natural forests as forests under formal protection with little-to-no anthropogenic pressure. We obtained boundaries for the protected areas of British Columbia using the **wdpar** R package (Hanson, 2022 version 1.3.7), and filtered these protected areas larger than 100 ha and in the IUCN categories Ia, Ib, II, and IV Muise et al. (2022). We use the Canadian Human Footprint generated by Hirsh-Pearson et al. (2022), a 300 m estimating of anthropogenic pressures across the entirety of Canada, which accounts for pressures included in the global human footprint (Venter et al., 2016), as well as pressures specific to the Canadian environment. The Canadian Human Footprint considers built environments, crop land, pasture land, human population density, nighttime lights, railways, roads, navigable waterways, dams and associated reservoirs, mining activity, oil and gas, and forestry pressures (Hirsh-Pearson et al., 2022). We consider any pixel within our protected area criteria with a Canadian Human Footprint score below four as representative of natural ecosystems, which balances the tradeoff between having a suitable number of high-quality pixels, and low anthropogenic pressures being represented in the natural areas.

#### 0.2.1.2 Forest Ecosystem Structure

Wall-to-wall, 30 m forest structure metrics (canopy height, canopy cover, structural complexity, and aboveground biomass) were generated by Matasci, Hermosilla, Wulder, White, Coops, Hobart, Bolton, et al. (2018) for 2015 across the forested landscape of British Columbia. This data was generated by using a random forest-kNN approach, imputing airborne laser scanning derived forest structural attributes across the entirety of Canada using Landsat-derived best-available-pixel (BAP) composites (Hermosilla et al., 2016; White et al., 2014) and topographic information (Matasci, Hermosilla, Wulder, White, Coops, Hobart, Bolton, et al., 2018; Matasci, Hermosilla, Wulder, White, Coops, Hobart, & Zald, 2018). The BAP composites were derived by selecting optical observations from the Landsat archive (including Landsat-5 Thematic Mapper, Landsat-7 Enhanced Thematic Mapper Plus, and Landsat-8 Operational Land Imager imagery) over the course of the growing season, considering atmospheric effects (haze, clouds, cloud shadows) and distance from the desired composite date (in this case, August 31st). Details on the pixel scoring method can be found in White et al. (2014). The BAP composites were further refined by using a spectral trend analysis on the normalized burn ratio to remove noise, and missing pixels are infilled using temporally-interpolated values. This procedure results in gap-free, surface-reflectance image composites (Hermosilla et al., 2015).

### 0.2.1.3 Forest Ecosystem Function

Landsat DHIs

### 0.2.1.4 Auxiliary Datasets

In our matching procedure (see Section ) we use two core datasets as our covariates. Firstly, we use a 30 m digital elevation model and derived slope dataset from the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) Version 2 GDEM product (Tachikawa et al., 2011). We also match on four climate variables; mean annual precipitation (MAP), mean annual temperature (MAT), mean warmest month temperature (MWMT), and mean coldest month temperature (MCMT) calculated from 1990-2020 climate normals using the ClimateNA software package at a 1 km spatial resolution, and downsampled to 30 m using cubic spline resampling in the **terra** R package (Hijmans, 2024 version 1.7-71) in R (R Core Team, 2024 version 4.4.0).

### 0.2.2 Matching

To identify natural forests across the province, we implement an expanded coarsened exact matching (CEM) technique (Iacus et al., 2012). This allows us to assess the similarity of all forested pixels in the province to known natural forests, while controlling for covariates such as climate and topography. CEM creates comparable groups of observation among covariates by first coarsening the covariates into meaningful bins. In our case, we coarsen all six of our covariates into 10 quantile-based bins. As an example, elevation observations in the 0-10th percentile would be grouped into elevation bin one, the 10th-20th percentile of observations into bin two, and so on. Following the binning procedure, CEM performs exact matching on the coarsened covariates, and discards any unmatched samples. We diverge from this method by not dropping unmatched samples. Instead, we group our unmatched strata with their nearest neighbour in quartile space, up to a minimum of 30 samples, while minimizing the overall nearest neighbour distance. This method combines the advantages of CEM (no bias checking, simple computation Iacus et al., 2012), with the ability to generate wall-to-wall comparisons of our treated cells to all other cells in the province.

### 0.2.3 Analysis

## 0.3 Results

## 0.4 Discussion

## 0.5 Conclusion

## References

- Andrew, M. E., Wulder, M. A., & Coops, N. C. (2012). Identification of *de facto* protected areas in boreal Canada. *Biological Conservation*, 146(1), 97–107. <https://doi.org/10.1016/j.biocon.2011.11.029>
- Arcese, P., & Sinclair, A. R. E. (1997). The role of protected areas as ecological baselines. *The Journal of Wildlife Management*, 61(3), 587–602. <https://doi.org/10.2307/3802167>
- Balaguer, L., Escudero, A., Martín-Duque, J. F., Mola, I., & Aronson, J. (2014). The historical reference in restoration ecology: Re-defining a cornerstone concept. *Biological Conservation*, 176, 12–20. <https://doi.org/10.1016/j.biocon.2014.05.007>
- Becker, A., Russo, S., Puliti, S., Lang, N., Schindler, K., & Wegner, J. D. (2023). Country-wide retrieval of forest structure from optical and SAR satellite imagery with deep ensembles. *ISPRS Journal of Photogrammetry and Remote Sensing*, 195, 269–286. <https://doi.org/10.1016/j.isprsjprs.2022.11.011>
- Bolton, D. K., Coops, N. C., Hermosilla, T., Wulder, M. A., White, J. C., & Ferster, C. J. (2019). Uncovering regional variability in disturbance trends between parks

- and greater park ecosystems across canada (1985–2015). *Scientific Reports*, 9(1), 1323. <https://doi.org/10.1038/s41598-018-37265-4>
- Cardinale, B. J., Duffy, J. E., Gonzalez, A., Hooper, D. U., Perrings, C., Venail, P., et al. (2012). Biodiversity loss and its impact on humanity. *Nature*, 486(7401), 59–67. <https://doi.org/10.1038/nature11148>
- Dirzo, R., & Raven, P. H. (2003). Global State of Biodiversity and Loss. *Annual Review of Environment and Resources*, 28(Volume 28, 2003), 137–167. <https://doi.org/10.1146/annurev.energy.28.050302.105532>
- Hanson, J. O. (2022). Wdpar: Interface to the world database on protected areas. *Journal of Open Source Software*, 7, 4594. <https://doi.org/10.21105/joss.04594>
- Hermosilla, T., Wulder, M. A., White, J. C., Coops, N. C., & Hobart, G. W. (2015). Regional detection, characterization, and attribution of annual forest change from 1984 to 2012 using landsat-derived time-series metrics. *Remote Sensing of Environment*, 170, 121132. <https://doi.org/10.1016/j.rse.2015.09.004>
- Hermosilla, T., Wulder, M. A., White, J. C., Coops, N. C., Hobart, G. W., & Campbell, L. B. (2016). Mass data processing of time series landsat imagery: Pixels to data products for forest monitoring. *International Journal of Digital Earth*, 9(11), 10351054. <https://doi.org/10.1080/17538947.2016.1187673>
- Hijmans, R. J. (2024). *Terra: Spatial data analysis*. Retrieved from <https://rspatial.org/>
- Hirsh-Pearson, K., Johnson, C. J., Schuster, R., Wheate, R. D., & Venter, O. (2022). Canada’s human footprint reveals large intact areas juxtaposed against areas under immense anthropogenic pressure. *FACETS*, 7, 398–419. <https://doi.org/10.1139/facets-2021-0063>
- Iacus, S. M., King, G., & Porro, G. (2012). Causal Inference without Balance Checking: Coarsened Exact Matching. *Political Analysis*, 20(1), 1–24. <https://doi.org/10.1093/pan/mpr013>
- Marín, A. I., Abdul Malak, D., Bastrup-Birk, A., Chirici, G., Barbati, A., & Kleeschulte, S. (2021). Mapping forest condition in europe: Methodological developments in support to forest biodiversity assessments. *Ecological Indicators*, 128, 107839. <https://doi.org/10.1016/j.ecolind.2021.107839>
- Matasci, G., Hermosilla, T., Wulder, M. A., White, J. C., Coops, N. C., Hobart, G. W., & Zald, H. S. J. (2018). Large-area mapping of Canadian boreal forest cover, height, biomass and other structural attributes using Landsat composites and lidar plots. *Remote Sensing of Environment*, 209, 90–106. <https://doi.org/10.1016/j.rse.2017.12.020>
- Matasci, G., Hermosilla, T., Wulder, M. A., White, J. C., Coops, N. C., Hobart, G. W., Bolton, D. K., et al. (2018). Three decades of forest structural dynamics over canada’s forested ecosystems using landsat time-series and lidar plots. *Remote Sensing of Environment*, 216, 697714. <https://doi.org/10.1016/j.rse.2018.07.024>
- McGill, B. J., Dornelas, M., Gotelli, N. J., & Magurran, A. E. (2015). Fifteen forms of biodiversity trend in the Anthropocene. *Trends in Ecology & Evolution*, 30(2), 104–113. <https://doi.org/10.1016/j.tree.2014.11.006>
- Muise, E. R., Coops, N. C., Hermosilla, T., & Ban, S. S. (2022). Assessing representation of remote sensing derived forest structure and land cover across a network of protected areas. *Ecological Applications*, 32(5), e2603. <https://doi.org/10.1002/eap.2603>
- Myers, N. (1988). Threatened biotas: “Hot spots” in tropical forests. *Environmentalist*, 8(3), 187–208. <https://doi.org/10.1007/BF02240252>
- Nielsen, S. E., Bayne, E. M., Schieck, J., Herbers, J., & Boutin, S. (2007). A new method to estimate species and biodiversity intactness using empirically derived reference conditions. *Biological Conservation*, 137(3), 403–414. <https://doi.org/10.1016/j.biocon.2007.02.024>

- Noss, R. F. (1990). Indicators for Monitoring Biodiversity: A Hierarchical Approach. *Conservation Biology*, 4(4), 355–364. <https://doi.org/10.1111/j.1523-1739.1990.tb00309.x>
- Pereira, H. M., Ferrier, S., Walters, M., Geller, G. N., Jongman, R. H. G., Scholes, R. J., et al. (2013). Essential Biodiversity Variables. *Science*, 339(6117), 277–278. <https://doi.org/10.1126/science.1229931>
- Pettorelli, N., Vik, J. O., Mysterud, A., Gaillard, J.-M., Tucker, C. J., & Stenseth, N. Chr. (2005). Using the satellite-derived NDVI to assess ecological responses to environmental change. *Trends in Ecology & Evolution*, 20(9), 503–510. <https://doi.org/10.1016/j.tree.2005.05.011>
- Pettorelli, N., Schulte to Bühne, H., Tulloch, A., Dubois, G., Macinnis-Ng, C., Queirós, A. M., et al. (2018). Satellite remote sensing of ecosystem functions: opportunities, challenges and way forward. *Remote Sensing in Ecology and Conservation*, 4(2), 71–93. <https://doi.org/10.1002/rse2.59>
- Pimm, S. L., & Raven, P. (2000). Extinction by numbers. *Nature*, 403(6772), 843–845. <https://doi.org/10.1038/35002708>
- Pinnegar, J. K., & Engelhard, G. H. (2008). The ‘shifting baseline’ phenomenon: a global perspective. *Reviews in Fish Biology and Fisheries*, 18(1), 1–16. <https://doi.org/10.1007/s11160-007-9058-6>
- R Core Team. (2024). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>
- Radeloff, V. C., Dubinin, M., Coops, N. C., Allen, A. M., Brooks, T. M., Clayton, M. K., et al. (2019). The Dynamic Habitat Indices (DHIs) from MODIS and global biodiversity. *Remote Sensing of Environment*, 222, 204–214. <https://doi.org/10.1016/j.rse.2018.12.009>
- Radeloff, Volker C., Roy, D. P., Wulder, M. A., Anderson, M., Cook, B., Crawford, C. J., et al. (2024). Need and vision for global medium-resolution landsat and sentinel-2 data products. *Remote Sensing of Environment*, 300, 113918. <https://doi.org/10.1016/j.rse.2023.113918>
- Razenkova, E., Lewińska, K. E., Yin, H., Farwell, L. S., Pidgeon, A. M., Hostert, P., et al. (n.d.). Medium-resolution dynamic habitat indices from landsat satellite imagery. *Remote Sensing of Environment*.
- Report of the conference of the parties to the convention on biological diversity on the second part of its fifteenth meeting.* (2023).
- Skidmore, A. K., Coops, N. C., Neinavaz, E., Ali, A., Schaepman, M. E., Paganini, M., et al. (2021). Priority list of biodiversity metrics to observe from space. *Nature Ecology & Evolution*. <https://doi.org/10.1038/s41559-021-01451-x>
- Tachikawa, T., Kaku, M., Iwasaki, A., Gesch, D. B., Oimoen, M. J., Zhang, Z., et al. (2011). *ASTER global digital elevation model version 2 - summary of validation results* (p. 27). Retrieved from <http://pubs.er.usgs.gov/publication/70005960>
- Thomas, C. D., Cameron, A., Green, R. E., Bakkenes, M., Beaumont, L. J., Collingham, Y. C., et al. (2004). Extinction risk from climate change. *Nature*, 427, 5.
- Thompson, I. D., Mackey, B., McNulty, S., & Mosseler, A. (2009). *Forest resilience, biodiversity, and climate change: a synthesis of the biodiversity / resiliende / stability relationship in forest ecosystems*. Montreal: Secretariat of the Convention on Biological Diversity.
- Urban, M. C. (2015). Accelerating extinction risk from climate change. *Science*, 348(6234), 571–573. <https://doi.org/10.1126/science.aaa4984>
- Venter, O., Sanderson, E. W., Magrach, A., Allan, J. R., Beher, J., Jones, K. R., et al. (2016). Sixteen years of change in the global terrestrial human footprint and implications for biodiversity conservation. *Nature Communications*, 7(1), 12558. <https://doi.org/10.1038/ncomms12558>

- 271 White, Joanne. C., Wulder, M. A., Hobart, G. W., Luther, J. E., Hermosilla, T.,  
272 Griffiths, P., et al. (2014). Pixel-based image compositing for large-area dense  
273 time series applications and science. *Canadian Journal of Remote Sensing*, 40(3),  
274 192212. <https://doi.org/10.1080/07038992.2014.945827>  
275 Winter, S. (2012). Forest naturalness assessment as a component of biodiversity  
276 monitoring and conservation management. *Forestry*, 85(2), 293–304. [https://](https://doi.org/10.1093/forestry/cps004)  
277 [doi.org/10.1093/forestry/cps004](https://doi.org/10.1093/forestry/cps004)