

b3gbi: Standardized Biodiversity Indicators from Spatiotemporal Occurrence Cubes

Shawn Dove ¹

¹ Department of Animal Ecology & Systematics, Justus Liebig University Giessen, Giessen, Germany

DOI: [10.21105/joss.10963](https://doi.org/10.21105/joss.10963)

Software

- [Review](#) 
- [Repository](#) 
- [Archive](#) 

Editor: [Matthias Grenié](#) 

Reviewers:

- [@equitable-equations](#)
- [@Aariq](#)

Submitted: 27 June 2026

Published: 23 September 2026

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](#)).

Summary

Standardized and reproducible biodiversity indicators are essential for tracking ecological trends, assessing conservation policies, and reporting on international biodiversity frameworks. However, raw species occurrence records, such as those hosted by the Global Biodiversity Information Facility (GBIF), are typically voluminous, unevenly sampled, and noisy. To address this, the **b3gbi** (B-Cubed General Biodiversity Indicators) R package provides a standardized, automated, and reproducible pipeline to calculate spatial and temporal biodiversity indicators directly from pre-processed spatiotemporal species occurrence cubes ([Desmet et al., 2023](#)). The package supports a suite of indicators, including species richness, evenness, rarity, completeness, and turnover, complete with decoupled, robust uncertainty estimation via bootstrapping.

Statement of Need

A major bottleneck in using global biodiversity datasets for policy reporting is the lack of standardized workflows for converting raw occurrence records into reliable indicators. Occurrence data cubes aggregate occurrences into specific spatial grid systems (e.g., EEA, MGRS, EQDGC, or ISEA3H grids) and temporal units (e.g., years), which significantly reduces data volume and mitigates some spatial and temporal biases ([Desmet et al., 2023](#)).

While packages like `rgbif` ([Chamberlain et al., 2020](#)) interface with GBIF data, and packages like `iNEXT` ([Hsieh et al., 2016](#)) calculate specific diversity metrics (such as Hill numbers), there is a distinct gap for a comprehensive, user-friendly framework that takes raw data cubes and automates the entire process of spatial boundary clipping, multi-scale grid aggregation, indicator calculation, and visualization. Furthermore, assessing the reliability of these indicators is critical. **b3gbi** addresses this need by integrating seamlessly with the `dubicube` package ([Langerhaert & Van Daele, 2026](#)) to perform cube-level bootstrapping, allowing users to calculate robust confidence intervals that capture the underlying sampling uncertainty of the original occurrence data. While researchers frequently build ad-hoc scripts to process occurrence cubes, **b3gbi** provides a unified, optimized framework that eliminates inconsistent, manual data-handling pipelines, ensuring cross-project reproducibility for international policy reporting.

State of the Field

Calculating biodiversity metrics from raw species observations is a common workflow in ecology, supported by established packages such as `vegan` ([Oksanen et al., 2026](#)) and `iNEXT` ([Hsieh et al., 2016](#)). However, these packages operate on community ecology matrices (e.g., site-by-species matrices) and do not support spatial coordinate grids, boundaries, or spatiotemporal occurrence cubes natively.

- **vegan** ([Oksanen et al., 2026](#)) is the standard R package for community ecology diversity

metrics (e.g., Shannon and Simpson diversity indices, Pielou evenness). However, it does not handle spatial grid mapping (EEA, MGRS, ISEA3H), spatial clipping, multi-scale grid aggregation, or spatiotemporal trend mapping.

- **iNEXT** (Hsieh et al., 2016) specializes in Hill numbers (rarefaction/extrapolation). While **b3gbi** integrates **iNEXT**'s core routines to compute Hill numbers and their associated confidence intervals, **iNEXT** on its own does not provide wrapper functions to ingest spatiotemporal data cubes, clip boundaries, aggregate grid resolutions, or compute other non-Hill metrics (e.g., area-based rarity, species range, turnover).
- **ebvcube** (Quoss et al., 2024) is designed to access, query, and visualize Essential Biodiversity Variables (EBVs) stored in netCDF structures, but it does not process Darwin Core-aligned spatiotemporal occurrence data cubes (as retrieved from GBIF) or calculate taxonomic, evenness, and turnover indicators from raw occurrence lists.
- **b3verse siblings**: Sibling packages within the **b3verse** ecosystem cover complementary indicator types: **pdindicator** focuses on phylogenetic diversity indicators, and **impIndicator** focuses on alien species impacts. **b3gbi** serves as the core, general-purpose library for taxonomic diversity, evenness, rarity, and occupancy indicators.

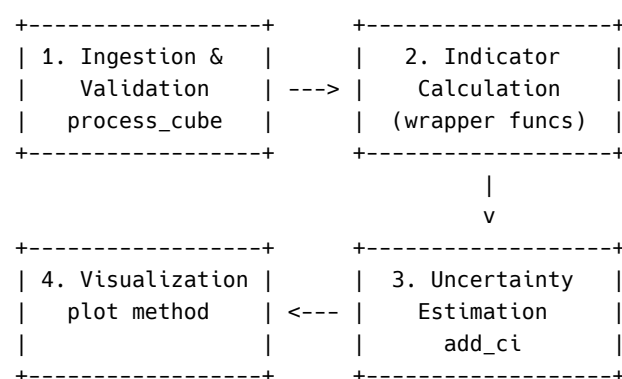
Build vs. Contribute Justification

Contributing these features directly to **vegan** or **iNEXT** would be inappropriate, as those packages are mathematically focused libraries for community ordination and Hill number estimation, respectively. Extending them to parse raw spatiotemporal occurrence cubes, manage geographical coordinate transformations (e.g., converting EEA codes to coordinates), and handle spatial geometries would expand their scopes excessively.

Designing **b3gbi** as a standalone, modular package allows first-class support for the “species occurrence cube” data standard (Desmet et al., 2023) developed by the B-Cubed project. It also enables seamless, decoupled integration with sibling packages like **dubicube** (Langerhaert & Van Daele, 2026) for cube-level bootstrapping (estimating indicator uncertainty) and **gcube** for synthetic cube simulations.

Software Design and Workflow

The **b3gbi** package is designed around a clean, four-step pipeline:



- 1. Ingestion & Validation:** The `process_cube()` function imports spatiotemporal species occurrence cubes (as CSV files), validates the required fields, auto-detects grid types (including gridded coordinate columns), and filters the data by year. The resulting output is structured into a `processed_cube` class.
- 2. Indicator Calculation:** A series of wrapper functions calculate indicators over space (`*_map()`) or time (`*_ts()`). Supported indicators include:
 - **Richness:** Observed richness (`obs_richness`), cumulative richness (`cum_richness`), species richness density (`spec_richness_density`), and estimated Hill numbers (`q`

- = 0, 1, 2) (Chao et al., 2014; Hsieh et al., 2016).
 - **Evenness:** Pielou evenness (`pielou_evenness`) and Williams evenness (`williams_evenness`).
 - **Completeness & Rarity:** Sample completeness/coverage (`completeness`), abundance-based rarity (`ab_rarity`), and area-based rarity (`area_rarity`).
 - **Occurrence & Density:** Total occurrences (`total_occ`) and occurrence density (`occ_density`).
 - **Occupancy & Dynamics:** Species occurrences (`spec_occ`), species range (`spec_range`), relative occupancy (`relative_occupancy`), species turnover (`occ_turnover`), and mean year of occurrence (`newness`).
3. **Uncertainty Estimation:** Uncertainty can be calculated by passing an indicator object to `add_ci()`. The package supports both cube-level bootstrapping (default, utilizing `dubicube`) and indicator-level bootstrapping. Decoupling the intensive bootstrapping process from initial indicator calculations allows for quick data exploration before executing heavy simulations.
 4. **Visualization:** Calling the generic `plot()` function on an indicator object automatically generates clean, publishable map or time-series plots. The package includes a minimalist, internal spatiotemporal occurrence cube dataset, allowing users to execute, visualize, and verify the entire pipeline out of the box.

Research Impact Statement

The `b3gbi` package provides evidence of both realized impact and credible near-term significance for researchers and policy reporting:

- **Core of the B-Cubed Ecosystem:** `b3gbi` is the central indicator package of the European Union's Horizon Europe B-Cubed project (Grant Agreement No. 101059592), which aims to standardize biodiversity workflows for policy reporting.
- **Policy Reporting Workflows:** The package computes taxonomic, evenness, rarity, and occupancy indicators that map directly to international policy frameworks, including the Kunming-Montreal Global Biodiversity Framework (GBF) and the EU Biodiversity Strategy for 2030.
- **FAIR Compliance:** `b3gbi` outputs are mapped to standard vocabularies (Darwin Core MeasurementOrFact extension, ISO 8601, and GBIF Data Cube specifications) as documented in `FAIR_MAPPING.md`. This ensures that generated indicators comply with the Research Data Alliance (RDA) FAIR Data Maturity Model standards (specifically RDA-I2-01D and RDA-I3-03M), allowing for seamless sharing and aggregation in global repositories.
- **Robust & Flexible Uncertainty Estimation:** A major limitation in existing biodiversity reporting is the lack of uncertainty quantification. By integrating with `dubicube`, `b3gbi` enables researchers to calculate robust confidence intervals via cube-level bootstrapping, while also offering faster, internal bootstrapping at the indicator level for rapid, resource-efficient data exploration.
- **Community Adoption and Case Studies:** The package has been integrated directly into the analytical workflows of multiple case studies within the B-Cubed project. Additionally, community adoption has been supported by developer-led training webinars to help researchers transition from raw occurrence data cubes to reproducible policy metrics. To lower the barrier to entry for non-programmer stakeholders, a Shiny-based graphical user interface (GUI) was initiated during a project hackathon and is maintained in the [b-cubed-eu/b3gbi-gui](https://github.com/b-cubed-eu/b3gbi-gui) repository.

Uncertainty Estimation & Statistical Limitations

When calculating uncertainty via `add_ci()`, users can choose between cube-level bootstrapping (resampling original occurrences) and indicator-level bootstrapping (resampling calculated values) (Langerhaert & Van Daele, 2026). Cube-level bootstrapping is the default and most robust level, and `add_ci()` automatically customizes the bootstrap settings based on the indicator type:

- **Hill Numbers:** Because Hill numbers are calculated via iNEXT's internal rarefaction/extrapolation engine, `add_ci()` delegates directly to iNEXT to avoid double-bootstrapping.
- **Bounded Indicators:** Evenness indicators automatically use logit transformation to keep confidence intervals within the valid $[0, 1]$ bounds.

Importantly, several indicators are excluded from post-hoc bootstrapping (`noci_list`):

- **Observed Richness and Richness Density:** Resampling from observed occurrences can never discover new/unobserved species, meaning bootstrapped richness estimates are strictly less than or equal to the observed richness ($S_{boot} \leq S_{obs}$). This forces the confidence intervals to lie entirely at or below the observed richness, which is statistically incorrect.
- **Sample Completeness:** Sample coverage (completeness) is calculated as a deterministic sample statistic using iNEXT methods, and confidence interval estimation is not supported.
- **Relative Occupancy:** Standard bootstrapping is not supported because resampling occurrences strips essential spatial metadata (like `total_num_cells` or `total_area_sqkm`). These spatial attributes cannot be regenerated from occurrences alone because occurrences contain no information about empty grid cells (zero occurrences) in the study area.
- **Cumulative Richness, Species Turnover, and Taxonomic Distinctness:** Standard bootstrapping is inappropriate due to the inherently sequential nature of cumulative counts, the high instability introduced in consecutive species lists, and extreme sensitivity to exact taxonomic compositions.

AI Usage Disclosure

While the overall software architecture, design, and the majority of the codebase and documentation of `b3gbi` were human-created, artificial intelligence was utilized to write, modify, and document portions of the package's recent iterations. Specifically, the Gemini 3.1 Flash and Gemini 3.5 Flash models, along with the Antigravity agentic coding assistant, were used for code generation, bug fixing, test coverage enhancement, documentation updates, and drafting JOSS manuscript sections. All design decisions and final code integrations were directed by a human author, and all AI-generated code and documentation were reviewed, edited, and approved by a human.

Acknowledgements

This package was developed as part of the "Biodiversity Building Blocks for Policy" (B-Cubed) project, which has received funding from the European Union's Horizon Europe Research and Innovation Programme under Grant Agreement No. 101059592. I thank Ward Langerhaert, Peter Desmet, Sandra MacFadyen, Damiano Oldoni, and Jasmijn Hillaert for providing code review and valuable feedback and suggestions during development. I also thank Hanno Seebens and Yanina Sica for providing valuable scientific input and support.

References

- Chamberlain, S., Noesgaard, D., & Altamirano, C. (2020). *rgbif: Interface to the Global Biodiversity Information Facility API*. <https://doi.org/10.32614/cran.package.rgbif>
- Chao, A., Gotelli, N. J., Hsieh, T. C., Sander, E. L., Ma, K. H., Colwell, R. K., & Ellison, A. M. (2014). Rarefaction and extrapolation with Hill numbers: A framework for sampling and estimation in species diversity studies. *Ecological Monographs*, 84(1), 45–67. <https://doi.org/10.1890/13-0133.1>
- Desmet, P., Oldoni, D., Blissett, M., & Robertson, T. (2023). *Specification for species occurrence cubes and their production*. B-Cubed Project. <https://docs.b-cubed.eu/guides/occurrence-cube/>
- Hsieh, T. C., Ma, K. H., & Chao, A. (2016). iNEXT: An R package for rarefaction and extrapolation of species diversity (Hill numbers). *Methods in Ecology and Evolution*, 7(12), 1451–1456. <https://doi.org/10.1111/2041-210X.12613>
- Langerart, W., & Van Daele, T. (2026). *dubicube: Calculation and interpretation of data cube indicator uncertainty*. <https://doi.org/10.5281/zenodo.14850237>
- Oksanen, J., Simpson, G. L., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., Solymos, P., Stevens, M. H. H., Szoecs, E., Wagner, H., Bedward, M., Bolker, B., Borcard, D., Carvalho, G., Caceres, M. D., Durand, S., Evangelista, H., Hannigan, G. F., ... Weedon, J. (2026). *vegan: Community ecology package*. <https://doi.org/10.32614/CRAN.package.vegan>
- Quoss, L., Fernandez, N., Langer, C., Valdez, J., & Pereira, H. M. (2024). *ebvcube: Working with netCDF for Essential Biodiversity Variables*. German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig. <https://doi.org/10.32614/CRAN.package.ebvcube>