



## Data Article

# One-hectare fine-scale dataset of a fynbos plant community in the Cape Floristic Region

Jan-Hendrik Keet<sup>a,b,\*</sup>, Cang Hui<sup>b,c,d</sup><sup>a</sup> EcoFloristix Specialist Botanical Surveys, Hillcrest Road, Somerset West 7130, South Africa<sup>b</sup> Centre for Invasion Biology, Department of Mathematical Sciences, Stellenbosch University, Stellenbosch 7602, South Africa<sup>c</sup> Biodiversity Informatics Unit, African Institute for Mathematical Sciences, Cape Town 7945, South Africa<sup>d</sup> National Institute for Theoretical and Computational Sciences, Stellenbosch 7602, South Africa

## ARTICLE INFO

*Article history:*

Received 28 November 2024

Revised 15 January 2025

Accepted 21 January 2025

Available online 27 January 2025

Dataset link: [One-hectare fine-scale dataset of a fynbos plant community in the Cape Floristic Region \(Original data\)](#)*Keywords:*

Species composition

Spatial distribution

Plant communities

Biodiversity hotspot

Vegetation surveys

## ABSTRACT

Cape fynbos, which forms part of the Cape Floristic Region (CFR) of South Africa, a global biodiversity hotspot, is renowned for its high levels of plant species endemism and diversity. This extraordinary ecosystem, characterised by nutrient-poor soils and fire-adapted vegetation, is a treasure trove of endemic flora. However, this fragile system faces increasing threats from habitat loss, climate change, and invasive species. Pristine fynbos, naturally high in plant diversity and which forms a large part of the CFR, presents an ideal opportunity to gather fine-scale data on community assembly patterns. Most fynbos vegetation surveys use a plot size of about 100 m<sup>2</sup>, with no spatial structures within plots to demarcate individual subplots. Here, a groundbreaking dataset is presented that fully covers 1-hectare of pristine fynbos, systematically gridded into 50 × 50 subplots, each measuring 2 × 2 m, arranged evenly within a square-shaped survey site. Each plot was assigned a unique Y–X coordinate combination. For each plot, all plant species present were recorded, along with their total percentage covers and maximum height values. Total percentage covers were also recorded for bare soil, rock, and termite mounds. This dataset provides a valuable contribution to the field of fynbos ecology, as well as plant community ecology in general, and establishes a benchmark for future one-hectare surveys of sim-

\* Corresponding author at: EcoFloristix Specialist Botanical Surveys, Hillcrest Road, Somerset West 7130, South Africa.  
E-mail address: [keetjanhendrik@gmail.com](mailto:keetjanhendrik@gmail.com) (J.-H. Keet).

ilar fynbos vegetation types, delineating the fine-scale composition and structure of fynbos in the CFR. The dataset will be useful for a wide audience, including community and spatial ecologists, plant and environmental scientists, and biodiversity informaticians and statistical ecologists, offering ideal data for testing new metrics of diversity and compositional turnover.

© 2025 The Authors. Published by Elsevier Inc.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

## Specifications Table

Subject	Bioinformatics
Specific subject area	Plant community ecology, with a specific emphasis on fine-scale community turnover over a large contiguous area
Type of data	Tables and Images
Data collection	Raw A one-hectare pristine fynbos site on the property of Vergelegen Wines near Somerset West (Western Cape, South Africa) was selected. The hectare was divided into 2500 contiguous $2 \times 2$ m plots. Plots were surveyed row-wise (with rows laid out from west to east), starting from the northwestern corner, and progressing eastwards to the southeastern corner. A $2 \times 2$ m plot template, composed of PVC pipes marked in major (0.5 m) and minor (0.1 m) intervals to facilitate percentage cover estimation, was used during surveying. In each plot, all plant species present were recorded, together with their respective total percentage covers and maximum height values. Furthermore, total percentage covers were also recorded for bare soil, rock, and termite mounds. A photo for each plot was taken using a Samsung Galaxy S10e SM-G970F.
Data source location	Region: Somerset West Country: South Africa Latitude and longitude of the site (centre of the survey hectare): 34.05487°S, 18.92879°E
Data accessibility	Repository name: Zenodo Data identification number: <a href="https://zenodo.org/records/14230083">10.5281/zenodo.14230083</a> Direct URL to data: <a href="https://zenodo.org/records/14230083">https://zenodo.org/records/14230083</a>
Related research article	None

## 1. Value of the Data

- The dataset is of interest to plant and environmental scientists as it provides extensive information on species assemblages in a pristine environment. It is equally valuable to biodiversity informaticians and statistical ecologists, offering ideal data for testing new metrics of diversity and compositional turnover.
- The high value of the dataset lies in its fine-scale resolution and large size, targeting a key gap as similar datasets for global biodiversity hotspots remain limited.
- The dataset serves as a benchmark for comparing similar surveys for restoration. It also helps to assess the health of similar ecosystems and inform sustainable land-use practices.
- Finally, this dataset provides a baseline against which to measure the impacts of future climate change on vegetation composition. The same site could be resurveyed in the future, offering additional insights into temporal variation.

## 2. Background

Biodiversity represents variation in species across time and space. Various frameworks for understanding biodiversity have been developed [1–5], with compositional diversity and turnover at the heart of community ecology [1].

While large-scale trait-based plant community assembly patterns have been extensively studied, fine-scale datasets remain scarce, limiting our understanding of fine-scale trait-based plant community assembly patterns. This is especially pronounced in biodiversity hotspots such as the Cape Floristic Region, or CFR [6,7], and more specifically in Cape fynbos.

The urgency of such data collection is heightened by the growing threats to global biodiversity, such as habitat loss and invasive species [8]. Within fynbos, various plot sizes have been used over the past couple of decades to survey vegetation, some being as small as 4 m<sup>2</sup> [9] and up to 100 m<sup>2</sup> [10,11], with general recommendations being between 100 and 250 m<sup>2</sup> [12]. However, no dataset exists yet in which a large contiguous area of one hectare, divided into fine-scale plots (such as 2 × 2 m gridded subplots), has been fully surveyed.

This study addresses this need by conducting a comprehensive, fine-scale vegetation survey in a specific region of the Fynbos biome, contributing data to conserving one of the world's most unique ecosystems.

## 3. Data Description

The dataset contains a total of 56,897 records.

The accompanying dataset (Excel workbook) comprises seven sheets, namely Cover (provides relevant information regarding the project), Site Map (includes a map of the survey site; the same as Fig. 1), Methodology (summarises the methodology used for data collection), Header Descriptions (explains the relevant fields within the dataset), Plot Metadata (contains all relevant metadata for each plot), Plot Species Data (lists species data for each plot), and Plant Species Info (relevant metadata regarding the plant species).

More specifically, the dataset contains a total of 56,897 records in the Plot Species Data sheet, and represents the data, in the form of plant species names and respective maximum heights and percentage cover values, collected from a total of 2500 survey plots. Together with this is accompanying metadata regarding the plots and species. The headers for the Plot Metadata, Plot Species Data, and Plant Species Info are given in Table 1.

The dataset also includes 2500 photos, namely a representative photo for each survey plot.

## 4. Experimental Design, Materials and Methods

A one-hectare pristine fynbos site on the property of Vergelegen Wines near Somerset West (Western Cape, South Africa) was selected for this study, with approximate coordinates of the four vertices, NE: 34.05444°S, 18.92935°E; NW: 34.05442°S, 18.92827°E; SE: 34.05534°S, 18.92933°E; and SW: 34.05533°S, 18.92825°E.

The site has a long history of invasive alien plant species clearing. Specifically, a large-scale clearing effort was initiated in the late 1990s, during which time approximately 140 hectares of the estate were cleared of alien species, including the survey site. Species that were cleared included *Acacia mearnsii*, *A. saligna*, *Pinus pinaster*, and various species of *Eucalyptus* and *Hakea*. Since 2006, the site has been maintained in pristine condition through annual alien plant species removal. Furthermore, two fires have impacted the site, one in 2009, and another in January 2017 which completely burned all vegetation. Thus, the age of the site when surveying commenced (September 2022) was approximately 5 years.

The 1-hectare site was divided into a 50 × 50 grid of 2 × 2 m plots (totalling 2500 plots; Fig. 1). Plots were surveyed row-wise, starting from the northwestern corner (Y01X01), and pro-

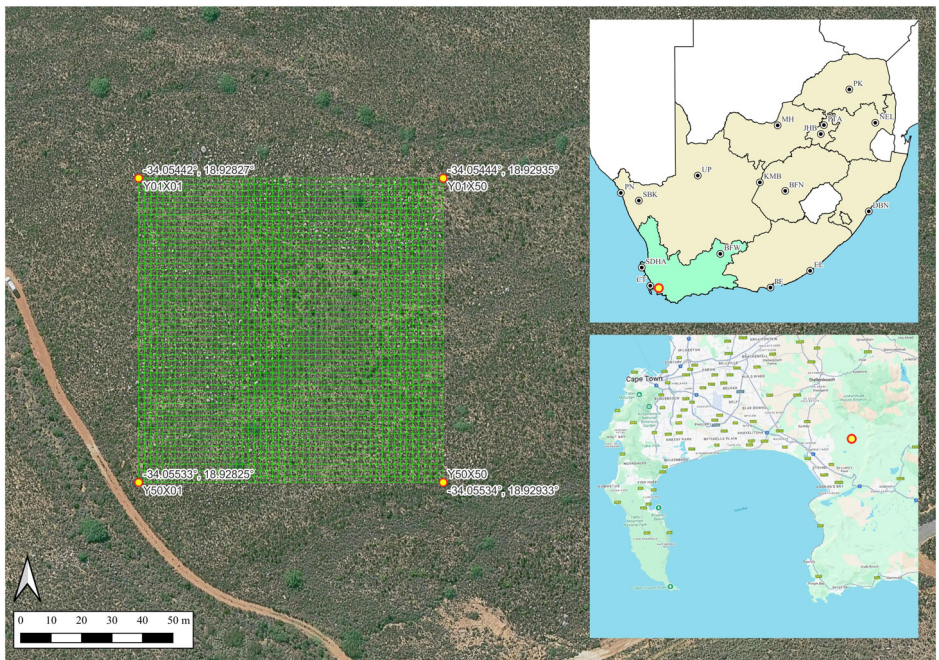
**Table 1**  
Overview of the accompanying plant species survey dataset.

Sheet	Header	Description
Plot Metadata	PlotID	The unique Y–X plot identifier label.
	Cover_Termite	Total percentage cover of termite mounds within the plot.
	Cover_Rock	Total percentage cover of rocks within the plot.
	Cover_BareSoil	Total percentage cover of bare soil within the plot.
	Ycenter_WGS84	Y-coordinate of the plot centre in WGS84 coordinates (EPSG: 4326).
	Xcenter_WGS84	X-coordinate of the plot centre in WGS84 coordinates (EPSG: 4326).
	Ycenter_UTM34S	Y-coordinate of the plot centre in WGS / UTM Zone 34S coordinates (EPSG: 32734).
	Xcenter_UTM34S	X-coordinate of the plot centre in WGS / UTM Zone 34S coordinates (EPSG: 32734).
	DateTimeOriginal	The date-time values of when the specific plot photo was taken. Values are in South Africa Standard Time (SAST). Note that in a few instances the timestamps (which were extracted from the plot photos) do not correspond exactly to the times that the plots were surveyed due to errors in field labelling; consequently, they were recaptured and/or relabelled to correctly reflect the corresponding plots. These errors were generally rectified within less than a month, and they only form a small part of the overall photo collection.
	FileName_Photo	The name of the corresponding plot photo. Filenames are the unique Y–X plot identifier labels. Note that for some photos, the labels given in the photo itself (written on a small whiteboard) are inaccurate; these are indicated in the filenames (example: “Y19X23_NOT X22.jpg”, which indicates that the plot is “Y19X23” but incorrectly labelled in the photo itself as “Y19X22”). See main text for more details.
Plot Species Data	PlotID	The unique Y–X plot identifier label.
	Species	The scientific name of the respective plant, including subspecific designations, but excluding author names. (this is same as in “Plant Species Info”)
	MaxHeight	The maximum height for the species in the plot. Values are in centimetres and in multiples of 5. The value “f” (short for “flat”) indicates a ground-hugging species for which height could not easily be estimated while surveying at a reasonable pace; these species are usually less than 1 cm in height. In a few instances height values were accidentally not recorded in the field; these values were estimated afterwards based on corresponding plot photos and are formatted as “NA[estimated value]”. See main text for details.
	SpeciesCover	Total percentage cover for the species within the plot. The value “R” (short for “Rare”) indicates that the total cover was too little to be estimated accurately. These values therefore represent anything below approximately 1 % of the total plot cover. In a few instances cover values were accidentally not recorded in the field; these values were estimated afterwards based on corresponding plot photos and are formatted as “NA[estimated value]”. See main text for details.
	DriedHeight	The ephemeral (seasonal) nature of many geophytic species (i.e., those that survive as bulbs, corms, tubers, and rhizomes below the soil surface during the dormant season) means that their aboveground parts (stems, leaves, and/or inflorescences) die back during certain times of the year. These parts often remain aboveground as dried-out structures, thus alluding to the presence of the species on site. However, the heights of these structures might not necessarily accurately reflect their true height during the growing season. Nevertheless, height values were still recorded where possible, and the species recorded since they are still present within the plot (even if only belowground). Thus, a value of “Yes” for “DriedHeight” indicates that the MaxHeight value was measured for such dried-out structures.

(continued on next page)

**Table 1** (continued)

Sheet	Header	Description
Plant Species Info	Species	The scientific name of the respective plant, including subspecific designations, but excluding author names. (this is same as in "Plot Species Data")
	SpeciesAuthor	The full scientific name of the species, including author(s) who formally described it.
	Family	The botanical family of the species. Family names were assigned based on the Red List of South African Plants ( <a href="http://redlist.sanbi.org/index.php">http://redlist.sanbi.org/index.php</a> ).
	Seasonality	This indicates whether the species is seasonal or not. Non-seasonal species were specifically defined in this context to be those that do not enter an annual period of dormancy (e.g., perennial shrubs, graminoids, trees, etc.); thus, these species remain observable throughout the year. Seasonal species (e.g., perennial geophytes or annuals) were differentiated based on those that completely die back during dormancy and become unobservable/undetectable, and those that retain some dried-out aboveground structures during dormancy, making them still observable/detectable and recordable as being present on site. Thus, species were classified in one of three categories here: "Non-seasonal", "Seasonal – observable", and "Seasonal – unobservable". However, note that in some instances, seasonal species that are usually unobservable might leave detectable aboveground traces, at least for a short period after entering dormancy. Similarly, some seasonal species that usually remain observable might completely disappear in terms of their aboveground features during dormancy and thus become unobservable/undetectable. This can occur, for example, when high winds blow away any dried-out aboveground structures. As such, the classification of being "observable/ unobservable" was made based on general trends that were experienced during surveying, and are not rigid.



**Fig. 1.** Locality of the sampling site and hectare grid. The two inset maps show the site locality within South Africa (top inset) and within a local context (red and yellow circle) near Somerset West, Western Cape Province (bottom inset). Coordinates indicated are in latitude and longitude (WGS84).

gressing eastwards to the southeastern corner (Y50X50). Within each plot, the following data were collected:

- All plant species present were recorded, along with their estimated percentage cover within the plot.
- The maximum height of each species was measured. The ephemeral (seasonal) nature of many geophytic species (i.e., those that survive as bulbs, corms, tubers, and rhizomes below the soil surface during the dormant season) mean that their aboveground parts (stems, leaves, and/or inflorescences) die back during certain times of the year. These parts often remain aboveground as dried-out structures, thus alluding to the presence of the species on site. However, the heights of these structures might not necessarily accurately reflect their true height during the growing season. Nevertheless, height values were still recorded where possible (and indicated as dried heights where applicable), and the species recorded since they are still present within the plot (even if only belowground).
- Percentage cover of bare soil, rock, and termite mounds was also recorded.

One photo was also taken for each survey plot. The photos are labelled with the unique Y-X coordinates of the respective plot. In a few instances, photos were incorrectly labelled in the field (Y-X coordinates written on a small white board as an identifier). These were corrected afterwards and are instead correctly labelled by their corresponding photo filenames (an example is “Y19X23\_NOT X22.jpg,” which indicates that the label on the white board in the photo itself reads Y19X22, but is incorrect, and the plot is instead Y19X23). All photos were taken with a Samsung Galaxy S10e SM-G970F [camera details: 10 MP, f/1.9, 26 mm (wide), 1/3”, 1.22 µm, dual pixel PDAF] at an approximate height of 1.88 m above ground level.

To minimize surveyor bias, a single individual conducted all surveys. The survey period extended from September 2022 to May 2023.

Three plant species could not be identified to species level since they did not flower during the survey period, and were also rare – only one instance of each species was observed – which meant that other specimens could not be used for potential identification. These entries are named “MorphoSpecies A”, “MorphoSpecies B”, and “MorphoSpecies C”.

In some instances, the species cover was too small to be estimated accurately; in such cases the letter “R” was given to indicate its percentage cover. Similarly, certain species were too small for their heights to be estimated (especially ground-hugging species such as certain *Oxalis* species); these species were assigned the value “f” for their maximum height.

## Limitations

Due to the extended field collection period required by the sole data collector to minimize trampling and damage to sensitive species and a sensitive ecosystem, as well as various intermittent environmental conditions preventing surveying (e.g., strong winds that could bias height measurements, or heavy rains that impaired visibility), minor seasonal variations in species composition between the initial and final plots may have occurred. These variations are more likely to affect geophytic and annual species, but are unlikely for perennial species. Thus, these effects are expected to be minimal, given the meticulous and thorough survey effort applied to each plot, and also given that the bulk of the records in the dataset are from perennial species.

During the survey, photos of six plots were accidentally missed (“Y13X42.jpg”, “Y23X44.jpg”, “Y27X41.jpg”, “Y31X07.jpg”, “Y33X15.jpg”, and “Y47X17.jpg”); these were recaptured on 09 October 2024 to ensure a complete photo dataset. While these six photos might present slightly different environmental conditions, their inclusion ensures dataset completeness. Furthermore, in a few instances photos and/or timestamps (which were extracted from the plot photos) do not correspond exactly to the times that the plots were surveyed due to errors in field labelling; consequently, they were recaptured and/or relabelled to correctly reflect the corresponding plots. Apart from the aforementioned six photos that were captured much later, these labelling errors were generally rectified within less than a month, so they do not present an actual limitation

(though this caveat should be noted), and they only form a small part of the overall photo collection.

Finally, there were instances where species cover and/or height values were accidentally not recorded in the field, and these appear as “NA” values in the dataset. Firstly, these species records are still retained within the dataset since they remain invaluable for presence/absence analyses. Secondly, to ensure completeness, these missing values were estimated afterwards based on the corresponding plot photos; these estimated values are provided in brackets and are formatted as “NA[estimated value]”.

## Ethics Statement

The work did not involve the use of human subjects, animal experiments, or data collected from social media platforms.

## CRediT Author Statement

**Jan-Hendrik Keet:** Conceptualization, Methodology, Data collection and curation, Writing, Original draft preparation, Reviewing and Editing. **Cang Hui:** Conceptualization, Supervision, Validation, Writing- Reviewing and Editing.

## Data Availability

[One-hectare fine-scale dataset of a fynbos plant community in the Cape Floristic Region \(Original data\)](#) (Zenodo).

## Acknowledgments

We are grateful to the support from the [National Research Foundation of South Africa](#) (NRF grant [89967](#)), the National Institute for Theoretical and Computational Sciences (NITheCS Focus Area: Complexity in Biological Systems), and the European Union’s Horizon Europe Research and Innovation Programme (ID No. [101059592](#); B3 – Biodiversity Building Blocks for policy). Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the funders; no funders can be held responsible for them. We also wish to acknowledge Vergelegen Wines who generously allowed us to survey on their property.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References

- [1] M.J. Anderson, T.O. Crist, J.M. Chase, M. Vellend, B.D. Inouye, A.L. Freestone, N.J. Sanders, H.V. Cornell, L.S. Comita, K.F. Davies, S.P. Harrison, N.J.B. Kraft, J.C. Stegen, N.G. Swenson, Navigating the multiple meanings of  $\beta$  diversity: a roadmap for the practicing ecologist, *Ecol. Lett.* 14 (2011) 19–28, doi:[10.1111/j.1461-0248.2010.01552.x](#).
- [2] A. Baselga, Partitioning the turnover and nestedness components of beta diversity, *Glob. Ecol. Biogeogr.* 19 (2010) 134–143, doi:[10.1111/j.1466-8238.2009.00490.x](#).

- [3] C. Hui, M.A. McGeoch, Zeta diversity as a concept and metric that unifies incidence-based biodiversity patterns, *Am. Nat.* 184 (2014) 684–694, doi:[10.1086/678125](https://doi.org/10.1086/678125).
- [4] L. Jost, Partitioning diversity into independent alpha and beta components, *Ecology* 88 (2007) 2427–2439.
- [5] G. Latombe, P. Boittiaux, C. Hui, M.A. McGeoch, A kernel integral method to remove biases in estimating trait turnover, *Methods Ecol. Evol.* 2024 (2024) 682–700, doi:[10.1111/2041-210X.14246](https://doi.org/10.1111/2041-210X.14246).
- [6] N. Myers, Biodiversity hotspots revisited, *Bioscience* 53 (2003) 916–917.
- [7] J.C. Manning, P. Goldblatt, *Plants of The Greater Cape Floristic Region 1: The Core Cape Flora*, South African National Biodiversity Institute, Pretoria, 2012.
- [8] J.J. Le Roux, C. Hui, M.L. Castillo, J.M. Iriondo, J.-H. Keet, A.A. Khapugin, F. Médail, M. Rejmánek, G. Theron, F.A. Yáñez-Nelli, H. Hirsch, Recent anthropogenic plant extinctions differ in biodiversity hotspots and coldspots, *Curr. Biol.* 29 (2019) 1–7, doi:[10.1016/j.cub.2019.07.063](https://doi.org/10.1016/j.cub.2019.07.063).
- [9] M.J.A. Werger, Species-area relationship and plot size: with some examples from South African vegetation, *Bothalia* 10 (1972) 583–594, doi:[10.4102/abc.v10i4.1568](https://doi.org/10.4102/abc.v10i4.1568).
- [10] Ş. Procheş, R.M. Cowling, L. Mucina, Species-area curves based on relevé data for the Cape Floristic Region, *S. Afr. J. Sci.* 99 (2003) 474–476.
- [11] R.M. Cowling, B.M. Campbell, P. Mustart, D.J. McDonald, M.L. Jarman, E.J. Moll, Vegetation classification in a floristically complex area: the Agulhas Plain, *S. Afr. J. Bot.* 54 (1988) 290–300, doi:[10.1016/s0254-6299\(16\)31329-1](https://doi.org/10.1016/s0254-6299(16)31329-1).
- [12] E. van der Maarel, Vegetation ecology - an overview, in: E. van der Maarel (Ed.), *Veg. Ecol.*, Blackwell Science Ltd., Malden, MA, USA, 2006: p. 395.