



Journal of Statistical Software

August 2014, Volume 59, Code Snippet 3.

<http://www.jstatsoft.org/>

Nestedness for Dummies (NeD): A User-Friendly Web Interface for Exploratory Nestedness Analysis

Giovanni Strona
European Commission

Paolo Galli
Milano-Bicocca

Davide Seveso
Milano-Bicocca

Simone Montano
Milano-Bicocca

Simone Fattorini
Azorean Biodiversity Group

Abstract

Recent theoretical advances in nestedness analysis have led to the introduction of several alternative metrics to overcome most of the problems biasing the use of matrix ‘temperature’ calculated by Atmar’s **Nestedness Temperature Calculator**. However, all of the currently available programs for nestedness analysis lack the user friendly appeal that has made the **Nestedness Temperature Calculator** one of the most popular community ecology programs. The software package **NeD** is an intuitive open source application for nestedness analysis that can be used online or locally under different operating systems. **NeD** is able to automatically handle different matrix formats, has batch functionalities and produces an output that can be easily copied and pasted to a spreadsheet. In addition to numerical results, **NeD** provides a graphic representation of the matrix under examination and of the corresponding maximally packed matrix. **NeD** allows users to select among the most used nestedness metrics, and to combine them with different null models. Integrating easiness of use with the recent theoretical advances in the field, **NeD** provides researchers not directly involved in theoretical debates with a simple yet robust statistical tool for a more conscious performance of nestedness analysis. **NeD** can be accessed at <http://purl.oclc.org/ned>.

Keywords: matrix temperature, NODF, null model, Python, **Django**.

1. Introduction

The word ‘nestedness’ is used in ecology to indicate inclusive distribution patterns. The idea of nestedness was introduced to describe the differences in species assemblages among different

islands (Patterson and Atmar 1986). In a perfectly nested distribution, species occurring at the site of interest are always present in a more species-rich site, whereas species absent from the site of interest never occur in a less species-rich one (Atmar and Patterson 1993). Thus, if occurrences are arranged in a species-area matrix, and the rows and columns of that matrix are ordered according to their respective totals, the matrix is nested if most of its presences are concentrated in the upper left triangle. Comparing an ideal perfect nested matrix with the one under examination makes it possible to quantify the extent of matrix nestedness on the basis of unexpected presences or absences (Ulrich, Almeida-Neto, and Gotelli 2009).

The software package **Nestedness Temperature Calculator** (Atmar and Patterson 1993) was the first program constructed to detect nested patterns in presence-absence matrices. This software uses a nestedness metric called matrix temperature (MT), which is intuitively appealing because it measures the order (entropy) of the matrix: the more nested (ordered) is a matrix, the lower is its temperature. Thanks to its ease of use and to its ability to produce graphical representations of matrices, the **Nestedness Temperature Calculator** has achieved great popularity among ecologists, promoting further exploration of nestedness (Ulrich and Gotelli 2007). However, the **Nestedness Temperature Calculator** has been severely criticized for many aspects and, particularly, for the way it calculates temperature, for the way it reshuffles the matrix to maximally pack it, for the way it estimates probability levels, and for the appropriateness of temperature as an adequate measure of matrix nestedness (Rodríguez-Gironés and Santamaría 2006; Almeida-Neto, Guimaraes, Guimaraes, Lodola, and Ulrich 2008; Ulrich *et al.* 2009). In the **Nestedness Temperature Calculator** the significance of the computed temperature is expressed as the probability of finding a random matrix colder than the observed one, with an estimate of the time required (impressively varying from a few seconds to millions of years).

There is now some agreement that probability levels can be assessed only using Z scores, i.e., by comparing the observed nestedness value with the mean of a series of values obtained by reshuffling the original matrix to produce a number of random matrices according to a certain null model. Additional software tools have been developed to automate nestedness computation with different nestedness measures and null models (Ulrich *et al.* 2009). The most popular and recommended metrics are BR (Brualdi and Sanderson metric) and NODF (nestedness measure based on overlap and decreasing fills) (Almeida-Neto *et al.* 2008). BR is a count of the number of discrepancies (absences or presences) that must be ‘corrected’ to produce a perfectly nested matrix (Brualdi and Sanderson 1999), whereas NODF quantifies whether depauperated assemblages constitute subsets of progressively richer ones and whether less frequent species are found in subsets of the sites where the most widespread species occur (Almeida-Neto *et al.* 2008). However, which combination of metrics and null models should be used in each particular circumstance is a matter of debate (Ulrich and Gotelli 2007; Ulrich *et al.* 2009). Although most of the currently available programs for nestedness analysis are highly efficient and provide many computation utilities, they all lack user friendliness. Table 1 summarizes functionalities, pros and cons of the most popular available programs for nestedness analysis.

In general, these programs are quite restrictive in input requirements and do not provide an easy-to-read output. For example, **Aninhado** (Guimarães and Guimarães 2006; Almeida-Neto *et al.* 2008) is an excellent program in terms of computation speed, but cannot handle row and column names, and does not provide automatic calculation of Z scores. Also, this software can produce up to 14 different output files for each analyzed matrix. Thus, a batch

	NTC	Aninhado	Binmatnest	Nestedness	NODF	NeD
<i>Metrics</i>						
BR				×	×	×
NODF		×		×	×	×
MT	×	×	×	×	×	×
<i>Null Models</i>						
FF		×	×	×	×	×
EE	×	×	×	×	×	×
FE		×		×	×	×
EF		×	×	×	×	×
<i>Packing type</i>						
Matrix richness		×		×	×	×
Matrix temperature	×		×	×		
<i>Software details</i>						
Graphic interface	×					×
Output graphics	×					×
Input format flexibility						×
Batch capability		×		×	×	×

Table 1: Functionalities, pros and cons of the most popular available software packages for nestedness analysis. The **Nestedness Temperature Calculator** is denoted by NTC.

run of 100 matrices would produce up to 1,400 different output files (note that these files contain nestedness measures of the null matrices and are therefore necessary to compute Z values). As an alternative to **Aninhado**, researchers interested in computing NODF can use the homonymous software **NODF**, by Almeida-Neto and Ulrich (2010). **NODF** saves results for a set of matrices in the same file, but its output includes a lot of very technical information which is of little interest for most users, cannot be readily imported to electronic spreadsheets, and therefore requires heavy editing for further examinations. Moreover, although it can handle rows and column names, **NODF** only accepts a specific file format, which can be a problem when users have to examine a large number of files not satisfying that requirement.

With the exception of the **Nestedness Temperature Calculator**, none of the available programs has a graphical user interface, neither provides graphical outputs showing the original matrix and the resulting maximally packed ones. In addition, their source code is not publicly available, and thus cannot be modified or integrated in other programs. Finally, all of them are Windows based software, and do not run on Linux machines.

It should be noted that some of the most commonly used nestedness indices can be computed using a few dedicated R (R Core Team 2014) functions from the packages **vegan** (Oksanen *et al.* 2013) and **bipartite** (Dormann, Gruber, and Fründ 2008). Yet, their use and possible integration with null models require a good knowledge of the R programming language.

The software package **NeD** (nestedness for dummies; Strona and Fattorini 2014) is aimed at solving some of these problems by integrating the easiness of use of the **Nestedness Temperature Calculator** with the recent theoretical advances in the field of nestedness analysis. **NeD** is primarily implemented as a dynamic web system using the open source scripting lan-

guage Python (van Rossum and de Boer 1991) and **Django** web framework (Holovaty and Kaplan-Moss 2007). In addition, a static version of **NeD** is freely available for download.

2. Technical details

2.1. Input and output

Data can be uploaded from a single text file, or directly pasted in a text box. Batch functionalities are fully implemented in the static version, allowing users to perform multiple analyses in a very easy way. **NeD** is able to automatically recognize a wide range of matrix formats (including all those required by the available nestedness programs). Users can therefore separate their presence-absence data with commas, spaces or tabulators, or do not separate them at all. **NeD** is also able to automatically recognize row and column names. The only obvious requirement is that the number of entries in each row should be maintained across the whole matrix. If a folder contains some matrix files in different formats and some other files (with any extensions) not containing a matrix, **NeD** will identify and analyze the first ones, disregarding the rest. In both online and static versions, results are displayed in a simple table that makes it possible to easily copy and paste them to an electronic spreadsheet. The output page includes useful information about the uploaded matrix (file name, number of rows, columns and occurrences), nestedness indices and (if calculated) Z values, relative nestedness values, and the corresponding probability levels. In addition, **NeD** will display graphics of both the original (submitted) matrix and the maximally packed matrix.

2.2. Matrix packing

The way a matrix is packed can significantly affect nestedness results. **NeD** uses the most straightforward packing procedure, reordering the original matrix according to row and column totals, which is the default method in most of the available nestedness programs. Different from other available nestedness software, **NeD** allows users to decide whether to keep or exclude empty rows and columns.

2.3. Available metrics

There are at least eight different metrics that have been proposed to compute nestedness (Ulrich and Gotelli 2007; Ulrich *et al.* 2009). Yet, most of them have been severely criticized, and only MT, BR, and NODF are recommended and generally used (Ulrich and Gotelli 2007; Ulrich *et al.* 2009). As the main purpose of **NeD** is to keep things simple, and to provide users with a simple interface for exploratory analyses, **NeD** implements only the aforementioned three nestedness metrics:

- MT (matrix temperature):

MT uses the Euclidean distances of unexpected empty or filled cells from the isocline that separates presences from absences in a perfectly nested matrix. The sum of these distances is rescaled relative to the maximum possible value for a given matrix size and fill (Rodríguez-Gironés and Santamaría 2006).

- BR (Brualdi and Sanderson discrepancy):

BR is the count of the minimum number of discrepancies, i.e., the number of absences or presences that must be modified to produce a perfectly nested matrix (Brualdi and Sanderson 1999).

- NODF (nestedness measure based on overlap and decreasing fills):

NODF is the percentage of presences in inferior rows and in right columns that are in the same position (column or row) of the presences in, respectively, upper rows and left columns with higher marginal totals for all pairs of columns and rows (Almeida-Neto *et al.* 2008; Ulrich *et al.* 2009).

2.4. Available null models

In order to assess the significance of the measured nestedness, **NeD** computes Z values as:

$$Z = \frac{(NIR - \overline{NIS})}{\sigma(NIS)}, \quad (1)$$

where NIR is the nestedness index (for the selected metric) of the matrix under examination, NIS is the set of index values for the null matrices generated by the program, and $\sigma(\cdot)$ is the standard deviation. For the NODF metric, Z values > 1.64 indicate significance at $p = 0.05$, while for BR and MT Z values < -1.64 indicate significance at $p = 0.05$.

In addition, **NeD** is the only available software that computes values of relative nestedness (RN; Bascompte, Jordano, Melián, and Olesen 2003) as:

$$RN = \frac{(NIR - \overline{NIS})}{(\overline{NIS})} \quad (2)$$

NeD provides five null models to assess the significance of the selected metric(s) (Ulrich and Gotelli 2007):

- 1) EE (equiprobable row totals, equiprobable column totals) maintains the total number of species occurrences in the matrix, but allows both row and column totals to vary freely.
- 2) CE (proportional row totals, proportional column totals) assigns to each matrix cell a probability to be occupied proportional to the corresponding row and column totals. The probability of a cell belonging to the i th row and to the j th column to be occupied is computed as:

$$P_{ij} = \frac{\frac{TotR_i}{C} + \frac{TotC_j}{R}}{2}, \quad (3)$$

where $TotR_i$ is the number of presences in the row i , $TotC_j$ is the number of presences in the column j , C is the number of matrix columns and R is the number of matrix rows.

- 3) EF (equiprobable row totals, fixed column totals) maintains observed column totals (i.e., species richness per site) but allows row totals (i.e., species occurrence frequencies) to vary randomly.

- 4) FE (fixed row totals, equiprobable column totals) maintains observed row totals (i.e., species occurrence frequencies) but allows column totals (i.e., species richness per site) to vary randomly.
- 5) FF (fixed-fixed) maintains both observed row and column totals.

2.5. NeD testing

To evaluate robustness of **NeD** results, we analyzed 270 matrices taken from the 294 matrices provided by [Atmar and Patterson \(1995\)](#) with **NeD** and other programs (**NODF**, [Almeida-Neto and Ulrich 2010](#), for **NODF**, and **Nestedness**, [Ulrich 2006](#), for **BR** and **MT**). We excluded 24 matrices from the original set because they were too small or exceeded the maximum number of columns allowed by the program **Nestedness**. Outputs were very consistent, with values usually differing for the second or third decimal digit. Pearson correlation coefficients between the nestedness values obtained with **NeD** and those produced by other packages were $r = 0.9999$ for **NODF** [$y = 0.9990 (\pm 0.0010) x + 0.0248 (\pm 0.0557)$, errors refer to standard errors], $r = 0.9613$ for **MT** [$y = 0.8962 (\pm 0.0157) x - 1.1502 (\pm 0.4436)$], and $r = 0.9988$ for **BR** [$y = 0.9950 (\pm 0.0029) x - 0.9129 (\pm 0.2441)$], with slopes of regression lines close to 1 and intercepts close to 0.

2.6. Limitations

In order to not overuse hosting server resources, the online version of **NeD** accepts matrices of maximum 22,500 cells, whereas the static version has no explicit matrix size limits. However, you should be aware that nestedness computation (especially for **NODF** and **MT**) for large matrices could take a long time. **NeD** has been coded in **Python**, which is a high-level programming language that offers great code readability. The other side of the coin is that code execution in **Python** is generally slower than in other low-level languages. To overcome this limitation, we have converted (in both the online and the static version) part of **NeD** **Python** source code in **C** extensions using **Cython** ([Behnel, Bradshaw, Citro, Dalcin, Seljebotn, and Smith 2011](#)). Thus, for medium sized matrices (say smaller than 10,000 cells), **NeD** shows performances comparable to those of the other available nestedness software. In addition, the complete source code of **NeD** is freely available in order to encourage improvements from researchers and developers.

3. Conclusions

Most of the available literature on nestedness focuses on technical and statistical issues. The current debate about properties of nestedness metrics, choice of null models and methods to pack the matrix may discourage broad applications of nestedness analysis, which, paradoxically, could be much helpful to better understand the behavior of nestedness metrics. Similarly, available software programs do not make things easier for occasional users. A nestedness beginner would probably have a hard time in discriminating among the 78 potential different combinations of settings (and the respective Z values) obtainable with **Nestedness** ([Ulrich 2006](#)). By contrast, **NeD** deliberately offers a low number of possible settings, selected among the most commonly used and recommended ([Ulrich and Gotelli 2007](#)). Noteworthy, the combinations of nestedness metrics and null models provided by **NeD** allow users to perform

analyses with the most robust parameter set-ups currently available (Ulrich *et al.* 2009). In addition, **NeD** requires no effort in input formatting or output processing and interpretation. Thanks to these features, **NeD** has the potential to promote the use of nestedness analyses also among researchers not directly involved in theoretical debates. Moreover, because binary matrices and bipartite networks are interchangeable representations for identical structures, NeD might be a useful tool in all fields where bipartite networks are commonly used, such as biology and medicine, economics, technology, or social sciences (Bascompte *et al.* 2003; Newman 2003; Newman and Park 2003; Doreian, Batagelj, and Ferligoj 2005; Goh, Cusick, Valle, Childs, Vidal, and Barabasi 2007).

Acknowledgments

We thank the anonymous reviewer for the useful comments on both the manuscript and the code. We are very grateful to Francesco Boccacci for his help in debugging the static version of **NeD**. The views expressed are purely those of the writers and may not in any circumstances be regarded as stating an official position of the European Commission.

References

- Almeida-Neto M, Guimaraes P, Guimaraes PR, Lodola RD, Ulrich W (2008). “A Consistent Metric for Nestedness Analysis in Ecological Systems: Reconciling Concept and Measurement.” *Oikos*, **117**(8), 1227–1239.
- Almeida-Neto M, Ulrich W (2010). “A Straightforward Computational Approach for Quantifying Nestedness Using Abundance Data.” *Environmental Modelling & Software*, **26**(2), 173–178.
- Atmar W, Patterson BD (1993). “The Measure of Order and Disorder in the Distribution of Species in Fragmented Habitat.” *Oecologia*, **96**(3), 373–382.
- Atmar W, Patterson BD (1995). *The Nestedness Temperature Calculator: A Visual Basic Program, Including 294 Presence-Absence Matrices*. AICS Research Inc., University Park, NM, USA and The Field Museum, Chicago, IL, USA. URL <http://aics-research.com/nestedness/tempcalc.html>.
- Bascompte J, Jordano P, Melián C, Olesen JM (2003). “The Nested Assembly of Plant-Animal Mutualistic Networks.” *Proceedings of the National Academy of Sciences of the United States of America*, **100**(16), 9383–9387.
- Behnel S, Bradshaw R, Citro C, Dalcin L, Seljebotn DS, Smith K (2011). “**Cython**: The Best of Both Worlds.” *Computing in Science Engineering*, **13**(2), 31–39.
- Brualdi RA, Sanderson JG (1999). “Nested Species Subsets, Gaps, and Discrepancy.” *Oecologia*, **119**(2), 256–264.
- Doreian P, Batagelj V, Ferligoj A (2005). *Generalized Blockmodeling*. Cambridge University Press.

- Dormann CF, Gruber B, Fründ J (2008). “Introducing the **bipartite** Package: Analysing Ecological Networks.” *R News*, **8**(2), 8–11. URL <http://CRAN.R-project.org/doc/Rnews/>.
- Goh KL, Cusick ME, Valle D, Childs B, Vidal M, Barabasi AL (2007). “The Human Disease Network.” *Proceedings of the National Academy of Sciences United States of America*, **104**, 8685–8690.
- Guimarães PRJ, Guimarães P (2006). “Improving the Analyses of Nestedness for Large Sets of Matrices.” *Environmental Modelling & Software*, **21**(10), 1512–1513.
- Holovaty A, Kaplan-Moss J (2007). *The Definitive Guide to Django: Web Development Done Right*. Apress, New York.
- Newman MEJ (2003). “The Structure and Function of Complex Networks.” *SIAM Review*, **45**(2), 167–256.
- Newman MEJ, Park J (2003). “Why Social Networks Are Different from other Types of Networks.” *Physical Reviews*, **E68**, 036122.
- Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O’Hara RB, Simpson GL, Solyomos P, Stevens MHH, Wagner H (2013). *vegan: Community Ecology Package*. R package version 2.0-10, URL <http://CRAN.R-project.org/package=vegan>.
- Patterson BD, Atmar W (1986). “Nested Subsets and the Structure of Insular Mammalian Faunas and Archipelagos.” *Biological Journal of the Linnean Society*, **28**(1), 65–82.
- R Core Team (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- Rodríguez-Gironés MA, Santamaría L (2006). “A New Algorithm to Calculate the Nestedness Temperature of Presence-Absence Matrices.” *Journal of Biogeography*, **33**(5), 924–935.
- Strona G, Fattorini S (2014). *NeD – Nestedness for Dummies*. URL <http://purl.oclc.org/ned>.
- Ulrich W (2006). *Nestedness – A Fortran Program for Calculating Ecological Matrix Temperatures*. URL <http://www.keib.umk.pl/nestedness/>.
- Ulrich W, Almeida-Neto M, Gotelli NG (2009). “A Consumer’s Guide to Nestedness Analysis.” *Oikos*, **118**(1), 3–17.
- Ulrich W, Gotelli NJ (2007). “Null Model Analysis of Species Nestedness Patterns.” *Ecology*, **88**(7), 1824–1831.
- van Rossum G, de Boer J (1991). “Interactively Testing Remote Servers Using the Python Programming Language.” *CWI Quarterly*, **4**(4), 283–303.

Affiliation:

Giovanni Strona
European Commission
Joint Research Centre
Institute for Environment and Sustainability
Ispra, I-21027, Italy
E-mail: giovanni.strona@jrc.ec.europa.eu