

1 **metan: an R package for multi-environment**
2 **trial analysis**

3 *Tiago Olivoto^{1,*}*

4 *Alessandro Dal'Col Lúcio¹*

5 ¹ Departament of Crop Science, Federal University of Santa Maria, 1000 Roraima Avenue,
6 Santa Maria, Rio Grande do Sul, Brazil

7 * Correspondence: [Tiago Olivoto <tiagoolivoto@gmail.com>](mailto:tiagoolivoto@gmail.com)

8 **Type:** Application

9 **Running headline:** analysis of multi-environment trials

10 Abstract

- 11 1. Multi-environment trials (MET) are crucial steps in plant breeding programs that aim
12 increasing crop productivity to ensure global food security. The analysis of MET data
13 requires the combination of several approaches including data manipulation, visualization,
14 and modeling. As new methods are proposed, analyzing MET data correctly and
15 completely remains a challenge, often intractable with existing tools.
- 16 2. Here we describe the `metan` R package, a collection of functions that implement a
17 workflow-based approach to (a) check, manipulate and summarise typical MET data; (b)
18 analyze individual environments using both fixed and mixed-effect models; (c) compute
19 parametric and non-parametric stability statistics; (c) implement biometrical models
20 widely used in MET analysis; and (d) plot typical MET data quickly.
- 21 3. In this paper, we present a summary of the functions implemented in `metan` and how
22 they integrate into a workflow to explore and analyze MET data. We guide the user
23 along a gentle learning curve and show how adding only a few commands or options at
24 a time, powerfull analyzes can be implemented.
- 25 4. `metan` offers a flexible, intuitive, and richly documented working environment with tools
26 that will facilitate the implementation of a complete analysis of MET data sets.

27 **Key-words:** AMMI, biometry, genotype–environment interaction, GGE biplot, multi-
28 environment trials, R software, stability, statistics

29 1 Introduction

30 In 50 years (1967-2017) the world average of cereal yields has increased by 64%, from 1.68 to
31 2.76 t ha^{-1} . In the same period, the total production of cereals has raised from 1.305×10^9 to
32 3.6×10^9 t, an increase of 175%, while the cultivated area increased by only 7.9% in the same
33 period (FAOSTAT, 2019). These unparallel increases have been possible due to the improved
34 cultivation techniques in combination with superior cultivars. For maize, for example, 50%
35 of the increase in yield was due to breeding (Duvick, 2005). Plant breeding programs have
36 been developing new cultivars for adaptation to new locations, management practices, or
37 growing conditions, in a clear and crucial example of exploitation of genotype-vs-environment
38 interaction (GEI).

39 The breeders' desire to modeling the GEI appropriately has led to the development of
40 the so-called stability analyses, which includes ANOVA-based methods (Yates & Cochran,
41 1938; Wricke, 1965; Shukla, 1972; Annicchiarico, 1992); regression-based methods (Eberhart
42 & Russell, 1966); non-parametric methods (Huehn, 1979; Lin & Binns, 1988; Fox, Skovmand,
43 Thompson, Braun, & Cormier, 1990; Thennarasu, 1995) and some methods that combines
44 different statistical techniques, such as the Additive Main Effect and Multiplicative Interaction,
45 AMMI, (Gauch, 2013), and Genotype plus Genotype-vs-Environment interaction, GGE, (Yan
46 & Kang, 2003). Then, it is no surprise that scientific production related to multi-environment
47 trial analysis has been growing fast in the last decades. A bibliometric survey in the SCOPUS
48 database revealed that in the last half-century (1969–2019) 6590 documents were published
49 in 902 sources (Journals, books, etc.) by 19.351 authors. In this period, the number of
50 publications has been increased on average by $11.22\% \text{ year}^{-1}$ but were in the last ten years
51 that the largest amount (~64%) of the documents were published (See Appendix S1, item 1
52 for more details).

53 Linear Mixed-effect Models (LMM) has been more frequently used to analyze MET data.
54 For example, between 2013 and 2015, the larger number of papers proposing methods to deal

55 with GEI were related to the Best Linear Unbiased Prediction (BLUP) in LMMs (Eeuwijk,
56 Bustos-Korts, & Malosetti, 2016). Recent advances in this field showed that BLUP is more
57 predictively accurate than AMMI and that the main advantages of these methods can be
58 combined to help researchers to select or recommend stable and high productive genotypes
59 (T. Olivoto, Lúcio, et al., 2019). Thus, the rapid spread of these methods to users around the
60 world can be facilitated if these procedures are implemented in specific software.

61 In most cases, analyzing MET data involves manual checking of the data subset(s) to
62 identify possible outliers, using some biometrical model to explore the relationships between
63 traits(or groups of traits), computing a within-environment ANOVA, computing a joint-
64 ANOVA, and, in case of a significant GEI, applying some stability method to explore it.
65 While a spreadsheet program (e.g. Microsoft Excel) may be used to perform a visual check
66 for outliers, an integrated development environment (IDE, e.g. R, SAS, or Matlab) is often
67 required to process the complex matrix operations required in some stability methods. IDEs,
68 however, require a certain degree of expertise to use and have steep learning curves, which
69 sometimes prevents that a coding layman implements certain methods. In this sense, R (Team,
70 2019) packages have been making easier the life of hundreds of thousands of researchers by
71 providing freely collections of functions developed by the community.

72 Some open-source R software packages that are designed –or are suitable– for analyzing
73 MET data are available. The `stability` package ([https://CRAN.R-project.org/package=](https://CRAN.R-project.org/package=stability)
74 [stability](https://CRAN.R-project.org/package=stability)) contains a collection of functions to perform stability analysis. The `ammistability`
75 package (<https://CRAN.R-project.org/package=ammistability>) computes multiple AMMI-
76 based stability parameters. The `gge` (<https://CRAN.R-project.org/package=gge>) and
77 `GGEbiplots` (<https://CRAN.R-project.org/package=GGEbiplots>) packages may be used to
78 perform a GGE model. The R packages `agricolae` ([https://CRAN.R-project.org/package=](https://CRAN.R-project.org/package=agricolae)
79 [agricolae](https://CRAN.R-project.org/package=agricolae)) and `plantbreeding` (<http://plantbreeding.r-forge.r-project.org/>), while not specif-
80 ically coded for MET analysis provides useful functions for computing parametric and
81 nonparametric stability statistics. Although useful, these packages do not offer options to

82 perform a complete analysis of MET data, i.e., to provide tools for all steps of the analysis
83 (check, manipulation, analysis, and visualization of data). For example, `GGEbiplots` requires
84 as input data a two-way table containing genotype by environment means with genotypes
85 in rows and environments in columns, but doesn't provide any function to create quickly
86 such table from data that often is in a "long" format in R. In addition, several studies often
87 compare different stability methods (e.g., Woyann et al., 2018; Scapim et al., 2010; Bornhofen
88 et al., 2017; Freiria et al., 2018; Shahbazi, 2019; Teodoro et al., 2019). This requires a
89 range of different packages to be used, making it the coding tedious and difficult to follow.
90 Thus, it seems to be value the creation of an R package that presents an easy workflow, and
91 incorporates the most used stability statistics, as well as recent introduced stability methods
92 (T. Olivoto, Lúcio, et al., 2019; T. Olivoto et al., 2019) in addition to options for computing
93 cross-validation (Piepho, 1994) and BLUP-based stability statistics (Colombari Filho et al.,
94 2013), features frequently used but not yet implemented in any other R package for MET
95 analysis.

96 Here, we describe the `metan` (**m**ulti-**e**nvironment **t**rial **a**nalysis) package, an open-source
97 R package designed to provide an efficient and reproducible workflow for the analysis of MET
98 data. Our main aim in this paper is to describe the features of `metan` and how this collection
99 of functions can be useful for an intuitive and complete analysis of MET data.

100 2 The `metan` package

101 The conceptual focus of `metan` is centered on five components (Fig. 1): (a) check, manipulate
102 and summarise typical MET data; (b) performs within-environment analysis of variance; (c)
103 compute parametric and non-parametric stability analysis; (d) compute biometrical models
104 widely used in plant MET analysis of plant breeding trials; and (e) quickly create typical
105 plots for two-way data considering any combination of qualitative and quantitative factors.

106 The development version of `metan` is available on Github (<https://github.com/>

107 `TiagoOlivoto/metan`) and can be installed directly via the R console using `devtools`:

```
# install.packages("devtools") uncomment to run  
devtools::install_github("TiagoOlivoto/metan")  
library(metan)
```

108 To illustrate the main features of the package, six example datasets (`data_alpha`, `data_g`,
109 `data_ge`, `data_ge2`, `int.effects`, and `meansGxE`) are distributed with `metan`. Comprehen-
110 sive details and examples of the functionality of `metan` are available in our online documenta-
111 tion (<https://tiagoolivoto.github.io/metan/>). Indeed, we strongly encourage readers to refer
112 to the vignettes as the primary source for information on `metan`'s functionality since they are
113 updated with every package release.

114 The `metan` package is constructed on an object-oriented approach, which allows for
115 -among other things- the reliable use of S3 generic functions such as `plot()`, `predict()` and
116 `print()`. These functions can be called any time to inspect and visualize a model. All
117 functions in `metan` have a non-standard evaluation, where the expressions are evaluated in
118 the specified data frame rather than in the current or global environments, thus avoiding
119 ambiguity in input data. This makes it possible to evaluate code in non-standard ways.
120 Basically, we can pass the argument as an expression rather than a value, reducing the
121 amount of typing.

122 In `metan`, all functions have as first argument the input data. So, all of them work
123 naturally with the forward-pipe operator `%>%` (Bache & Wickham, 2014), which makes the
124 typing cleaner and more logical. Most of MET analyze more than one trait in each genotype.
125 Thus, when possible, functions in `metan` analyze a vector of variables and return the results
126 into a list, saving a lot of time and code when several variables need to be analyzed. In
127 `metan`, if we want to compute the AMMI stability value (Purchase, Hatting, & Deventer,
128 2000) for several traits, we can combine the functions `performs_ammis()`, `AMMI_indexes()`,
129 and `get_model_data()` with `%>%` to get a two-way table with the statistic for each genotype

130 and traits (see an example in Appendix S1, item 8.5.4). To our current knowledge, no other
131 package designed for MET analysis presents these features.

132 Sometimes in MET, a certain analysis needs to be run for each level of a factor, e.g.,
133 compute a path analysis or check outliers for each environment of the trial. The R base
134 function `subset()` could be useful, but worryingly tedious if a large number of levels need to
135 be evaluated. Users of `metan` can count with the function `split_factors()`, which split the
136 original data into n subsets according to the grouping variable(s), where n is the total number
137 of combinations of the factors used. The object of class `split_factors` can be passed on
138 to several functions `%>%`. If a function recognizes such class of data them it will take care
139 of details and compute what is required for each one of the n levels (See an example in
140 Appendix S1, item 6.3).

141 2.1 Checking data

142 It is assumed that MET data has the following structure (columns): **ENV**, a factor with e
143 levels, being e the number of environments; **GEN** a factor with g levels, being g the number
144 of genotypes; **REP** a factor with r levels, being r the number of replicates within each
145 environment; and at least one numeric variable, e.g., grain yield. The expected number of
146 rows in a typical MET data is then $e \times g \times r$.

147 The function `inspect()` scans all columns of a data frame object for errors that may
148 affect the use of functions in `metan` and return a warning if (i) the data has less than three
149 columns as factor; (ii) the data has less than the expected number of rows based on the levels
150 of factor variables; (iii) any variable has missing values; (iv) any possible outliers is detected.
151 Running `inspect()` is an optional and exploratory step that flags potential issues before
152 analysis. Error check results are summarised in the R console as warnings while a plot (Fig.
153 1a) can also be created by using the argument `plot = TRUE` in the function (See more details
154 in Appendix S1, item 6.1).

155 Outliers may violate the assumption of identically distributed errors in ANOVA models.
156 Anomalous values tend to increase the estimate of sample variance, thus lowering the chance
157 of rejecting the null hypothesis. In this regard, we strongly recommend checking for outliers,
158 especially if the function `inspect()` returned a warning about them. Users of `metan` can use
159 the function `find_outliers()` to check for possible outliers in a numeric variable, returning
160 a summary in the console (Appendix S1, item 6.2) and a plot (Fig. 1b) if `plots = TRUE` is
161 used.

162 Descriptive statistics help researchers to describe and understand the structure of a MET
163 data. The function `desc_stat()` computes a total of 30 statistics and when combined with
164 `split_factors()` can be used to implement a descriptive analysis for each level of a factor,
165 e.g., for each genotype (See more details in Appendix S1, item 6.3).

166 Frequently in MET analysis two-way tables (e.g., genotypes in rows and environments
167 in columns) need to be created to serve as data input in some procedure, for example, in the
168 R package `GGEbiplots`. The function `make_mat()` can be used to create such a table. You
169 inform the data frame in the “*long*” format, the two variables to be mapped to rows and
170 columns and one numeric variable from which the values will fill the table and `make_mat()`
171 take care of the details. Conversely, `make_long()` can be used to quickly convert a “*wide*”
172 table to a “*long*” data frame (See an example in Appendix S1, item 6.4).

173 **2.2 Analyzing individual environments**

174 Individual analysis performed within each environment gives to researchers important in-
175 formation regarding the performance of genotypes in such environments. Provided that a
176 typical MET data is available, the function `anova_ind()` can be used to compute, for each
177 environment, a fixed-effect ANOVA considering a Randomized Complete Block design. The
178 function returns the significance of factors, coefficient of variation, heritability, and accuracy
179 of selection (See a numeric example in Appendix S1, item 7).

180 The function `gamem()` is used to specifically analyze genotypes using a mixed-effect model
181 considering both a randomized complete block design or an alpha-lattice design (Patterson
182 & Williams, 1976). The function `get_model_data()` can be used to extract the model
183 information such as variance components, genetic parameters, and *P*-values for the Likelihood
184 ratio test for random effects. By using the function `plot_blup()` with an object of class
185 `gamem` the plot in Fig. 1c is produced.

186 2.3 Stability analysis

187 After inspecting data, checking for outliers and possibly analyzing individual environments, a
188 quick visual inspection of the genotype–environment interaction can be performed with the
189 function `ge_plot()`, which will generate the plots in Fig. 1m-n. Statistically, GEI can be
190 checked in a joint analysis of variance performed with the function `anova_joint()` (Appendix
191 S1, item 7). If GEI is significant, then it is reasonable to proceed with some stability analysis
192 to explore such interaction. `metan` provides a collection of functions to implement widely
193 used methods for stability analysis in the evaluation of multi-environment trials (Table 1).

194 After fitting a model, users can obtain custom plots to interpret the GEI. By invoking
195 `plot()` in an object of class `performs_amm` residual plots (Fig. 1d) can be obtained. In
196 AMMI analysis, biplots (Fig. 1f) are produced with the function `plot_scores()`, provided
197 that an object of class `performs_amm`, `waas` or `waasb` is available in the Global Environment
198 (See Appendix S1, item 8.5.3 for more details). In GGE models, fitted with the function
199 `gge()`, 10 types of biplots (Yan & Kang, 2003) can be created. Fig. 1g shows the biplot
200 type 8, used for ranking genotypes. All plots are produced with package `ggplot2` (Wickham,
201 2016). So, users of `metan` can count on the high level of personalization provided by `ggplot2`
202 to change any non-data elements of your plot (See an example in Appendix S1, item 7.5.3).

203 Users who research the associations between stability indexes (e.g., Woyann et al.,
204 2018; Bornhofen et al., 2017; Freiria et al., 2018; Shahbazi, 2019) often find difficulties in

205 computing the set of statistics and binding them into a “*ready-to-read*” file. `metan` provides
206 an efficient solution for doing that. The function `ge_stats()` is a wrapper function and can
207 be used to compute all the stability methods shown in Table 1 at once. Then, users can use
208 `get_model_data()` to extract either the statistics or the ranks related to each genotype in
209 each index and variable –if multiple variables are used in `ge_stat()`–, or `corr_stab_ind()`,
210 to compute a Spearman’s rank correlation matrix between the computed stability indexes
211 (See Appendix S1, item 8.9 for more details).

212 **2.4 Biometrical models**

213 Multi-environment trials often generate data on several traits, and this data should be
214 exploited. In breeding trials (as well as in many other areas), indirect selection helps geneticists
215 and breeders to select superior genotypes (Meira et al., 2017; T. Olivoto, Nardino, et al.,
216 2017; T. Olivoto, Souza, et al., 2017; Ferrari et al., 2018; Santos et al., 2018; Fonseca, Lima,
217 Dardengo, Silva, & Xavier, 2019; Gediya et al., 2019; Lopes Costa, Melo, & Oliveira Mano,
218 2019); thus, any tool that facilitates this work is welcome. `metan` provides useful functions
219 for implementing biometrical models easily. This includes the functions `corr_coef()` for
220 computing Pearson product-moment correlation with P -values, `lpcor()` for computing partial
221 correlation coefficients; `covcor_design()` for computing phenotypic, genotypic, and residual
222 (co)variance/correlation matrices based on designed experiments; `can_cor()` for computing
223 canonical correlation analysis; `path_coeff()` for computing path coefficients; `corr_ss()` for
224 sample size planning; `corr_plot()` for a mixed (text and plot) visualization of a correlation
225 matrix (Fig. 1j); `corr_ci()` for computing nonparametric confidence intervals of Pearson’s
226 correlation (Fig. 1k); and `clustering()` for clustering analysis (Fig. 1l).

227 Since `metan` was conceived for multi-environment trial analysis, the function
228 `split_factors()` can be used to pass grouped data allowing, for example, that a path
229 analysis or a canonical correlation be computed within each level of a factor, as shown in
230 Santos et al. (2018). For more details, please, refer to Appendix S1, item 7.

231 **2.5 Data visualization**

232 `metan` provides useful functions for creating quickly typical plots of two-way data, such as
233 those observed in MET data. The function `ge_plot()` can be used for a visual inspection
234 of the GEI (Fig. 1m-n). The function `plot_factbars()` is used to create bar plots with
235 two factor variables (Fig. 1o). `plot_factbars()` has as mandatory arguments only the
236 data, the factors 1 and 2, and the response variable. Similarly, line plots with options for
237 fitting different polynomial degrees can be made with the function `plot_factlines()`. In
238 an experiment with two quantitative factors, the function `resp_surf()` can be used to fit a
239 response surface model; Then a surface plot (Fig. 1p) can be created with `plot()` (See more
240 details in Appendix S1, item 10).

241 **3 Concluding remarks and future improvements**

242 The package `metan` was designed to facilitate the analysis of multi-environment trials, allowing
243 for more effective and less time-consuming handling and processing of MET datasets that have
244 been increasing rapidly in the last years. Users will find in `metan` a complete framework to
245 implement the most used parametric and non-parametric stability statistics for MET analysis.
246 The package implements stability methods not available in any other R package, including the
247 estimation of BLUP-based stability statistics (Colombari Filho et al., 2013), newer stability
248 methods such as the weighted average of absolute scores from the (T. Olivoto, Lúcio, et
249 al., 2019), the multi-trait stability index (Olivoto et al., 2019), and the implementation of
250 cross-validation procedures for AMMI and BLUP models (Piepho, 1994). `metan` can also be
251 useful for to a lot of other researchers since it provides options for implementing worldwide
252 used multivariate statistics, e.g., path analysis, linear, partial and canonical correlations, thus
253 allowing exploiting the maximum of (good or bad) information that a data set can offer. The
254 estimation of stability indexes for several variables at once and the estimation of biometrical
255 models for each level of a factor makes `metan` to outperform already published R packages

256 for MET analysis. These features will reduce the amount of coding and save the precious
257 time of the researchers when running their analyzes. The `metan` package is (and will always
258 be) extensively documented online, with transparent and fully reproducible examples. `metan`
259 is currently under active development; so, new functions will be implemented in the near
260 future. Our next efforts will be focused on implementing cross-validation procedures for GGE
261 models, allowing cross-validation to run in parallel, and increasing the number of stability
262 methods available.

263 **4 Acknowledgment**

264 We thank the National Council for Scientific and Technological Development (CNPq) and
265 Coordination for the Improvement of Higher Education Personnel (CAPES) for fellowships
266 and grants to the authors. The authors have no conflicts of interest to declare.

267 **5 Author's contributions**

268 T.O conceived the ideas, authored the software and manuscript; A.D.L assisted in the
269 implementation of methods, and critically revised the manuscript; both authors gave final
270 approval for publication.

271 **6 Data accessibility**

272 The `metan` R package is open-source and available on GitHub (<https://github.com/TiagoOlivoto/metan>).
273 Package vignettes are also open-source, accessible at
274 <https://tiagoolivoto.github.io/metan/>. Installing and loading `metan` will automatically
275 load all example data used in this paper. Since the package is updated regularly,
276 all code, data, and documentation used in this manuscript have been archived at
277 <https://doi.org/10.5281/zenodo.3548917> as `metan` version 1.1.0.

278 7 Supporting information

279 Additional supporting information may be found online in the Supporting Information section
280 at the end of the article.

281 8 References

- 282 Annicchiarico, P. (1992). Cultivar adaptation and recommendation from alfalfa trials in
283 Northern Italy. *Journal of Genetics and Breeding*, *46*, 269–278.
- 284 Bache, S. M., & Wickham, H. (2014). magrittr: a forward-pipe operator for R. Retrieved
285 from <https://cran.r-project.org/package=magrittr>
- 286 Bornhofen, E., Benin, G., Storck, L., Woyann, L. G., Duarte, T., Stoco, M. G., & Marchioro,
287 S. V. (2017). Statistical methods to study adaptability and stability of wheat genotypes.
288 *Bragantia*, *76*(1), 1–10. doi:[10.1590/1678-4499.557](https://doi.org/10.1590/1678-4499.557)
- 289 Colombari Filho, J. M., Resende, M. D. V., Morais, O. P., Castro, A. P., Guimarães, É. P.,
290 Pereira, J. A., ... Breseghello, F. (2013). Upland rice breeding in Brazil: a simultaneous
291 genotypic evaluation of stability, adaptability and grain yield. *Euphytica*, *192*(1), 117–129.
292 doi:[10.1007/s10681-013-0922-2](https://doi.org/10.1007/s10681-013-0922-2)
- 293 Duvick, D. N. (2005). The contribution of breeding to yield advances in maize (*Zea mays* L.).
294 *Advances in Agronomy*, *86*, 83–145. doi:[10.1016/S0065-2113\(05\)86002-X](https://doi.org/10.1016/S0065-2113(05)86002-X)
- 295 Eberhart, S. A., & Russell, W. A. (1966). Stability parameters for comparing varieties. *Crop*
296 *Science*, *6*(1), 36–40. doi:[10.2135/cropsci1966.0011183X000600010011x](https://doi.org/10.2135/cropsci1966.0011183X000600010011x)
- 297 Eeuwijk, F. A. van, Bustos-Korts, D. V., & Malosetti, M. (2016). What should students in
298 plant breeding know about the statistical aspects of genotype × environment interactions?
299 *Crop Science*, *56*(5), 2119–2140. doi:[10.2135/cropsci2015.06.0375](https://doi.org/10.2135/cropsci2015.06.0375)
- 300 FAOSTAT. (2019). Food and Agriculture Organization of the United Nations. Retrieved
301 from <http://www.fao.org/faostat/en/%7B/#%7Ddata/QC>

- 302 Ferrari, M., Carvalho, I. R., Pelegrin, A. J. de, Nardino, M., Szareski, V. J., Olivoto, T.,
303 ... Souza, V. Q. de. (2018). Path analysis and phenotypic correlation among yield
304 components of soybean using environmental stratification methods. *Australian Journal*
305 *of Crop Science*, 12(2). doi:[10.21475/ajcs.18.12.02.pne488](https://doi.org/10.21475/ajcs.18.12.02.pne488)
- 306 Fonseca, A. S. da, Lima, J. S. de S., Dardengo, M. C. J. D., Silva, S. D. A., &
307 Xavier, A. C. (2019). Canonical correlation between soil attributes and foliar of
308 conilon coffee trees. *Journal of Experimental Agriculture International*, 35(3), 1–14.
309 doi:[10.9734/jeai/2019/v35i330208](https://doi.org/10.9734/jeai/2019/v35i330208)
- 310 Fox, P. N., Skovmand, B., Thompson, B. K., Braun, H. J., & Cormier, R. (1990).
311 Yield and adaptation of hexaploid spring triticale. *Euphytica*, 47(1), 57–64.
312 doi:[10.1007/BF00040364](https://doi.org/10.1007/BF00040364)
- 313 Freiria, G. H., Gonçalves, L. S., Furlan, F. F., Fonseca Junior, N. S., Lima, W. F., & Prete,
314 C. E. (2018). Statistical methods to study adaptability and stability in breeding lines of
315 food-type soybeans. *Bragantia*, 77(2), 253–264. doi:[10.1590/1678-4499.2017076](https://doi.org/10.1590/1678-4499.2017076)
- 316 Gauch, H. G. (2013). A simple protocol for AMMI analysis of yield trials. *Crop Science*,
317 53(5), 1860–1869. doi:[10.2135/cropsci2013.04.0241](https://doi.org/10.2135/cropsci2013.04.0241)
- 318 Gediya, N. N., Patel, D. A., Kumar, S., Kumar, D., Parmar, D. J., & Patel, S. S. (2019).
319 Phenotypic variability, path analysis and molecular diversity analysis in chickpea (*Cicer*
320 *arietinum* L.). *Vegetos*, 32(2), 167–180. doi:[10.1007/s42535-019-00020-9](https://doi.org/10.1007/s42535-019-00020-9)
- 321 Huehn, M. (1979). Beitrage zur erfassung der phanotypischen stabilitat. *EDV in Medizin*
322 *Und Biologie*, 10, 112–117. Retrieved from [https://www.scopus.com/record/display.uri?](https://www.scopus.com/record/display.uri?eid=2-s2.0-0344931259%7B/&%7Dorigin=inward)
323 [eid=2-s2.0-0344931259%7B/&%7Dorigin=inward](https://www.scopus.com/record/display.uri?eid=2-s2.0-0344931259%7B/&%7Dorigin=inward)
- 324 Lin, C. S., & Binns, M. R. (1988). A superiority measure of cultivar performance for cultivar
325 x location data. *Canadian Journal of Plant Science*, 68(1), 193–198. doi:[10.4141/cjps88-](https://doi.org/10.4141/cjps88-018)
326 [018](https://doi.org/10.4141/cjps88-018)
- 327 Lopes Costa, F. C., Melo, M. F., & Oliveira Mano, A. R. (2019). Path analysis in cowpea

- 328 reveals protein accumulation dependency of mineral nutrients accumulation. *Journal of*
329 *Experimental Agriculture International*, 32(5), 1–10. doi:[10.9734/jeai/2019/v32i530113](https://doi.org/10.9734/jeai/2019/v32i530113)
- 330 Meira, D., Meier, C., Olivoto, T., Nardino, M., Rigatti, A., Lunkes, A., ... Souza, V. Q.
331 (2017). Physiological traits and their relationships in black oat populations. *Genetics*
332 *and Molecular Research*, 16(4), gmr16039814. doi:[10.4238/gmr16039814](https://doi.org/10.4238/gmr16039814)
- 333 Murakami, D. M., & Cruz, C. D. (2004). Proposal of methodologies for environment stratifi-
334 cation and analysis of genotype adaptability. *Crop Breeding and Applied Biotechnology*,
335 4(1), 7–11. doi:[10.12702/1984-7033.v04n01a02](https://doi.org/10.12702/1984-7033.v04n01a02)
- 336 Olivoto, T., Lúcio, A. D. C., Da silva, J. A. G., Marchioro, V. S., Souza, V. Q. de, & Jost, E.
337 (2019). Mean performance and stability in multi-environment trials I: Combining features
338 of AMMI and BLUP techniques. *Agronomy Journal*. doi:[10.2134/agronj2019.03.0220](https://doi.org/10.2134/agronj2019.03.0220)
- 339 Olivoto, T., Lúcio, A. D. C., Da silva, J. A. G., Sari, B. G., & Diel, M. I. (2019). Mean
340 performance and stability in multi-environment trials II: Selection based on multiple
341 traits. *Agronomy Journal*. doi:[10.2134/agronj2019.03.0221](https://doi.org/10.2134/agronj2019.03.0221)
- 342 Olivoto, T., Nardino, M., Carvalho, I. R., Follmann, D. N., Ferrari, M., Szarecki, V. J.,
343 ... Souza, V. Q. de. (2017). REML/BLUP and sequential path analysis in estimating
344 genotypic values and interrelationships among simple maize grain yield-related traits.
345 *Genetics and Molecular Research*, 16(1), gmr16019525. doi:[10.4238/gmr16019525](https://doi.org/10.4238/gmr16019525)
- 346 Olivoto, T., Souza, V. Q., Nardino, M., Carvalho, I. R., Ferrari, M., Pelegrin, A. J., ...
347 Schmidt, D. (2017). Multicollinearity in path analysis: a simple method to reduce its
348 effects. *Agronomy Journal*, 109(1), 131–142. doi:[10.2134/agronj2016.04.0196](https://doi.org/10.2134/agronj2016.04.0196)
- 349 Patterson, H. D., & Williams, E. R. (1976). A new class of resolvable incomplete block
350 designs. Oxford University Press/Biometrika Trust. doi:[10.2307/2335087](https://doi.org/10.2307/2335087)
- 351 Piepho, H.-P. (1994). Best Linear Unbiased Prediction (BLUP) for regional yield trials:
352 a comparison to additive main effects and multiplicative interaction (AMMI) analysis.
353 *Theoretical and Applied Genetics*, 89(5), 647–654. doi:[10.1007/BF00222462](https://doi.org/10.1007/BF00222462)

- 354 Purchase, J. L., Hatting, H., & Deventer, C. S. van. (2000). Genotype \times environment
355 interaction of winter wheat (*Triticum aestivum* L.) in South Africa: II. Stability anal-
356 ysis of yield performance. *South African Journal of Plant and Soil*, 17(3), 101–107.
357 doi:[10.1080/02571862.2000.10634878](https://doi.org/10.1080/02571862.2000.10634878)
- 358 Santos, E. A., Almeida, A.-A., Branco, M. C., Santos, I. C., Ahnert, D., Baligar, V. C.,
359 & Valle, R. R. (2018). Path analysis of phenotypic traits in young cacao plants under
360 drought conditions. *PLOS ONE*, 13(2), e0191847. doi:[10.1371/journal.pone.0191847](https://doi.org/10.1371/journal.pone.0191847)
- 361 Scapim, C. A., Pacheco, C. A. P., do Amaral Júnior, A. T., Vieira, R. A., Pinto, R. J. B., &
362 Conrado, T. V. (2010). Correlations between the stability and adaptability statistics of
363 popcorn cultivars. *Euphytica*, 174(2), 209–218. doi:[10.1007/s10681-010-0118-y](https://doi.org/10.1007/s10681-010-0118-y)
- 364 Shahbazi, E. (2019). Genotype selection and stability analysis for seed yield of *Nigella sativa*
365 using parametric and non-parametric statistics. *Scientia Horticulturae*, 253, 172–179.
366 doi:[10.1016/j.scienta.2019.04.047](https://doi.org/10.1016/j.scienta.2019.04.047)
- 367 Shukla, G. K. (1972). Some statistical aspects of partitioning genotype-environmental
368 components of variability. *Heredity*, 29(2), 238–245. doi:[10.1038/hdy.1972.87](https://doi.org/10.1038/hdy.1972.87)
- 369 Team, R. C. (2019). R: a language and environment for statistical computing. Vienna, Austria:
370 R Foundation for Statistical Computing. Retrieved from <https://www.r-project.org/>
- 371 Teodoro, P. E., Farias, F. J. C., Carvalho, L. P. de, Ribeiro, L. P., Nascimento, M.,
372 Azevedo, C. F., ... Bhering, L. L. (2019). Adaptability and stability of cotton
373 genotypes regarding fiber yield and quality traits. *Crop Science*, 59(2), 518–524.
374 doi:[10.2135/cropsci2018.04.0250](https://doi.org/10.2135/cropsci2018.04.0250)
- 375 Thennarasu, K. (1995). *On certain non-parametric procedures for studying genotype-*
376 *environment interactions* (Thesis). IARI. Retrieved from [https://krishikosh.egranth.](https://krishikosh.egranth.ac.in/displaybitstream?handle=1/2035304)
377 [ac.in/displaybitstream?handle=1/2035304](https://krishikosh.egranth.ac.in/displaybitstream?handle=1/2035304)
- 378 Wickham, H. (2016). *Ggplot2: Elegant graphics for data analysis*. Houston, TX: Springer.
379 doi:[10.1007/978-3-319-24277-4](https://doi.org/10.1007/978-3-319-24277-4)

- 380 Woyann, L. G., Milioli, A. S., Bozi, A. H., Dalló, S. C., Matei, G., Storck, L., & Benin,
381 G. (2018). Repeatability of associations between analytical methods of adaptability,
382 stability, and productivity in soybean. *Pesquisa Agropecuaria Brasileira*, 53(1), 63–73.
383 doi:[10.1590/S0100-204X2018000100007](https://doi.org/10.1590/S0100-204X2018000100007)
- 384 Wricke, G. (1965). Zur berechnung der okovalenz bei sommerweizen und hafer. *Z. Pflanzen-*
385 *zuchtg*, 52, 127–138.
- 386 Yan, W., & Kang, M. S. (2003). *GGE biplot analysis: a graphical tool for breeders, geneticists,*
387 *and agronomists* (p. 271). CRC Press.
- 388 Yates, F., & Cochran, W. G. (1938). The analysis of groups of experiments. *The Journal of*
389 *Agricultural Science*, 28(04), 556–580. doi:[10.1017/S0021859600050978](https://doi.org/10.1017/S0021859600050978)

390 **Figure caption**

391 Fig. 1. Diagram showing steps in a typical workflow in the analysis of multi-environment
392 trial data using `metan`. (a) inspect plot made with `inspect()`; (b) outlier check plot
393 made with `find_outliers()`; (c) blups for genotypes made with `plot_blup()`; (d)
394 model diagnostic made with `plot.*()`; (e) radar plot showing the multi-trait stability
395 index made with `plot.mtsi()`; (f) a gge biplot made with `plot.gge()`; (g-h) an AMMI2
396 biplot and a nominal yield plot, respectively, made with `plot_scores()`; (i) results for a
397 cross validation procedure made with `plot.cv_ammif()`; (j-k) visualization of correlation
398 matrices with `corr_plot()` and `plot.corr_coef()`, respectively; (l) nonparametric
399 confidence intervals for correlation made with `plot.corr_ci()`; (m-n) genotype-vs-
400 environment plot made with `ge_plot()`; (o) a barplot created with `plot_factbars()`;
401 (p) a contour plot created with `plot.resp_surf()`.

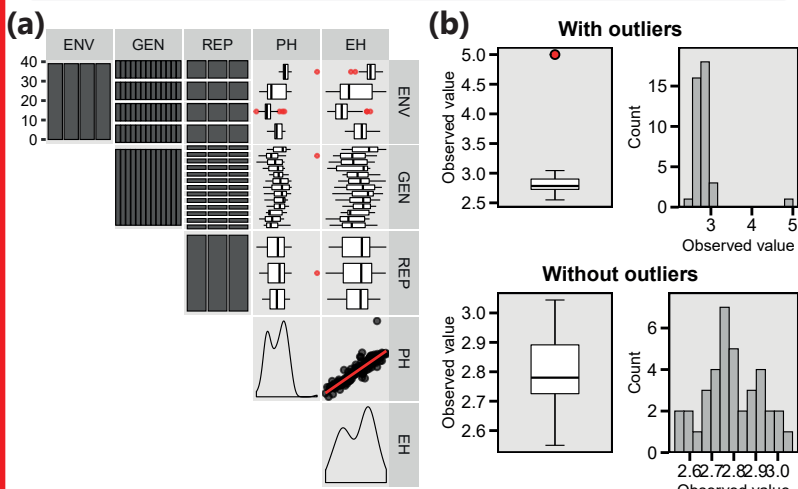
402 Table 1. Functions available in **metan** version 1.1.0 for computing stability analysis

Function	Method	Reference
Parametric		
<code>Annicchiarico()</code>	Genotypic confidence index	Annicchiarico (1992)
<code>ecovalence()</code>	Wricke's ecovalence	Wricke (1965)
<code>gai()</code>	Geometric adaptability index	Shahbazi (2019)
<code>ge_factanal()</code>	Environment stratification	Murakami & Cruz (2004)
<code>ge_reg()</code>	Joint Regression Analysis	Eberhart & Russell (1966)
<code>ge_stats()</code>	Wrapper function	NA
<code>gge()</code>	GGE biplot method	Yan & Kang (2003)
<code>mtsi()</code>	Multi-trait stability index	T. Olivoto, Lúcio, et al. (2019)
<code>performs_ammii()</code>	AMMI method	Gauch (2013)
<code>Resende_indexes()</code>	BLUP-based stability statistics	Colombari Filho et al. (2013)
<code>Shukla()</code>	Shukla's stability variance	Shukla (1972)
<code>waas()</code> , <code>waasb()</code>	Weighted average of absolute scores	T. Olivoto, Lúcio, et al. (2019)
<code>wsmpl()</code>	Stability and mean performance	T. Olivoto, Lúcio, et al. (2019)
Non-parametric		
<code>Fox()</code>	The 'top third' method	Fox et al. (1990)
<code>Huehn()</code>	Huehn's stability statistics	Huehn (1979)
<code>Superiority()</code>	Lin and Binns' superiority measure	Lin & Binns (1988)
<code>Thennarasu()</code>	Thennarasu's stability statistics	Thennarasu (1995)

MET data
ENV GEN REP/BLOCK TRAIT(S)

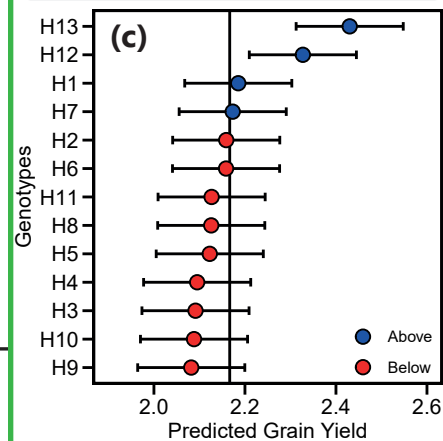
Check/manipulate/summarise data

```
desc_stat() find_outliers() inspect()
make_long() make_mat() split_factors() to_factor()
```



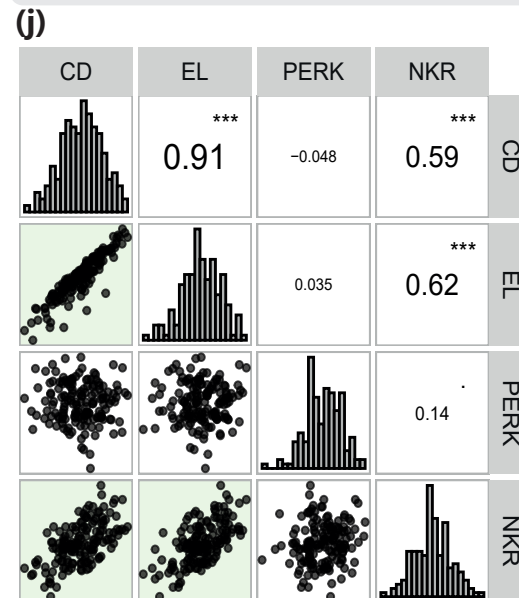
Individual environments

```
anova_ind()
anova_joint()
gamem()
```



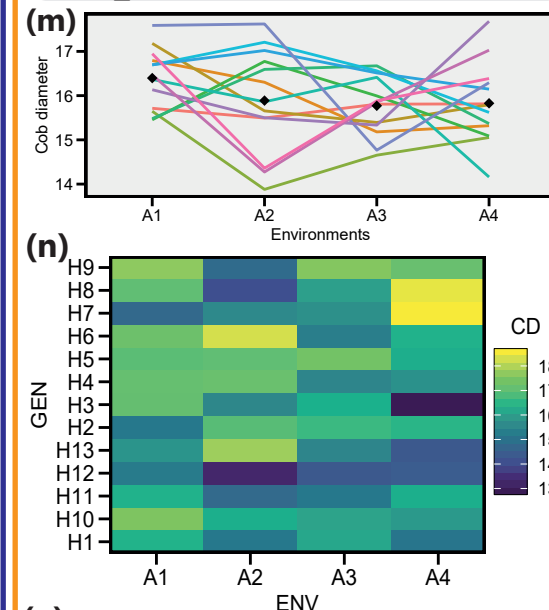
Biometrical models

```
can_cor() corr_ss()
clustering() lpcorr()
colinddiag() mahala()
corr_ci() pairs_mantel()
corr_plot() path_coeff()
```



Data visualization

```
ge_plot() plot_lines()
plot_factlines() resp_surf()
plot_factbars()
```



Results and diagnostic

```
autoplot()
get_model_data()
plot_*()
```

```
plot()
predict()
print()
```

Stability analysis

```
Annicciarico() ge_stats()
ecovalence() gg_e()
cv_amm() Huehn()
cv_blup() mtsi()
Fox() performs_amm()
fai_blup() Resende_indexes()
gai() Shukla()
ge_cluster() superiority()
ge_effects() Thennasaru()
ge_factanal() waas()
ge_reg() waasb()
```

