multiplestressR: An R package to analyse factorial multiple stressor data using the additive and multiplicative null models

Benjamin J. Burgess[1]

David J. Murrell[1]

[1] Centre for Biodiversity and Environment Research, Department of Genetics, Evolution and Environment, University College London, Gower Street, London, WC1E 6BT, UK

Abstract

Globally, ecosystems are being affected by multiple simultaneous stressors (also termed drivers, factors, or perturbations). While the effects of single stressors are becoming increasingly well understood, there remains substantial uncertainty regarding how multiple stressors may interact to affect ecosystems. Accordingly, there is substantial interest in documenting how stressors combine to impact individuals through to entire communities. Indeed, understanding how stressors interact represents one of the grand challenges currently facing ecologists and conservationists.

Popular methods used to classify stressor interactions comprise multiple steps, including complex mathematical equations. Accordingly, there is the potential for errors to occur at multiple points, any of which can result in erroneous conclusions being drawn. Furthermore, there are frequently minor methodological differences between studies which may limit, or even prevent, direct comparisons of their results from being made.

Here, we introduce the *multiplestressR* R package, a statistical tool which addresses the above issues. The package allows researchers to easily conduct a rigorous analysis of their multiple stressor data and provides results which are simple to interpret. The *multiplestressR* package can implement either the additive or multiplicative null model using iterations of these tools which are commonplace within multiple stressor ecology. The *multiplestressR* package can classify interactions as being synergistic, antagonistic, reversal, or null and requires minimal experience in either R or statistics to implement. Additionally, we provide example R code which can be easily modified to analysis any given factorial multiple stressor dataset. Indeed, widespread use of this software will allow for an easier and more robust comparison of results. Ultimately, we hope that the *multiplestressR* package will provide a stronger understanding of how stressors combine to affect individuals, populations, communities, and ecosystems.

1   Introduction

Ecosystems are being subjected to a diverse array of stressors (O'Hara et al. 2021; Reid et al. 2019), *sensu* threats, perturbations, or factors (see Orr et al. 2020). Such stressors are capable of directly, or indirectly, acting across all levels of ecological organisation (Simmons et al. 2021). However, ecosystems are primarily affected by multiple co-occurring stressors, rather than single stressors in isolation (Halpern et al. 2015). Accordingly, these simultaneous stressors are capable of interacting and as such there is considerable interest in cataloguing how such stressors may interact (Côté et al. 2016). Indeed, understanding how stressors interact to affect individuals through to ecosystems is one of the grand challenges currently facing ecologists (Fleishman et al. 2011). Recent examples of such studies have investigated the combined effect of pesticides on bee mortality (Siviter et al. 2021), the joint effect of chemical stressors and nutrient enrichment of litter decomposition (Beaumelle et al. 2021), and the combined effect of CO_2 and temperature on phytoplankton growth rates (Seifert et al. 2020). Such studies are capable of analysing a wide range of response metrics, and frequently employ fully factorial experiments.

Previously, it has been assumed that the effects of stressors combine in an additive manner, with any interactions which deviate from this expectation being termed an 'ecological surprise' (Christensen et al. 2006; Paine et al. 1998; Thompson et al. 2018). However, there is now considerable evidence that this assumption of additivity does not necessarily hold, and that multiple stressors are capable of combining to produce antagonistic (Burgess et al. 2021), synergistic (Przeslawski et al. 2015), or even reversal (Jackson et al. 2016) interactions (see Orr et al., 2020). Furthermore, there is considerable debate as to whether alternatives to the additive null model are more appropriate for ecological datasets (Kerkhoff & Enquist, 2009; Schäfer & Piggott, 2018). Of the alternatives, the multiplicative null model is the most widely implemented within ecological research (e.g., Harvey et al. 2013; Gomez Isaza et al. 2020). However, by considering the multiple stressor literature, it is evident that there is little consistency in the methodologies implemented across studies. While the additive and multiplicative null models may be employed in various studies, multiple versions of these null models are implemented (e.g., Darling & Côté 2008; Lange et al. 2018; Yue et al. 2017). As such, it is unclear how these different methods compare to one another, or even if the results of two studies which utilise different statistical analyses are comparable. Here, we detail the *multiplestressR* R package, which provides a simple framework for researchers to easily analyse factorial stressor pair datasets using only a few lines of R code. At present, numerous steps are required in order to classify stressor interactions, including several complex statistical equations. Accordingly, this means that there are multiple points at which errors can occur and hence affect both the results and the subsequent ecological interpretation. However, there is currently no software to automate the classification of

multiple stressor interactions that we are aware of. The *multiplestressR* package fills this gap and allows researchers to easily, and rigorously, analyse their multiple stressor data, while bypassing potential errors.

The *multiplestressR* R package allows researchers to implement either the additive (Gurevitch et al. 2000) or multiplicative (Lajeunesse 2011) null models. Importantly, these forms of the null models have been implemented within previous multiple stressor research (e.g., Burgess et al. 2021; Harvey et al. 2013; Hillebrand 2002) and have been shown to be statistically rigorous (Burgess et al. 2022). Alongside a robust statistical analysis, the *multiplestressR* also provides several tools for researchers to interpret their results, enabling the analysis of individual experiments through to meta-analytical datasets. Each of the functions within the *multiplestressR* package is described in detail below, with example code also provided.

The *multiplestressR* package is openly available via the CRAN repository (https://CRAN.R-project.org/package=multiplestressR) and can be easily installed using the code shown in Figure 1. To install the *multiplestressR* package, R v2.1.0 or greater is required. The *multiplestressR* package also requires the *ggplot2* (Wickham 2016), *patchwork* (Pederson 2020), and *viridis* (Garnier et al. 2021) R packages.

```
install.packages("multiplestressR")
library(multiplestressR)
```

Figure 1: Code for installing and loading the *multiplestressR* package.

2   Input data

In order to implement the *multiplestressR* package, 12 variables are required (Table 1). This corresponds to the same three variables across four different experimental treatments. The required experimental treatments are as follows: i) baseline experimental conditions (*Control*); ii) only the first stressor present (*StressorA*); iii) only the second stressor present (*StressorB*); iv) both the first and second stressors present together (*StressorsAB*). Accordingly, this R package should only be used to analyse data from fully factorial experiments which consider pairs of stressors.

For each experimental treatment, the three required variables are as follows: i) the mean value of a given response metric (*Mean*); ii) the standard deviation associated with the mean response metric (*SD*); iii) the sample size of the experiment treatment (*N*). Ultimately, a failure to include any of these variables for any of the experimental treatments will severely limit the functionality of the

4

multiplestressR package. Importantly, in order to utilise this package, each experimental treatment must comprise at least two replicates per treatment. Likewise, it is imperative for correct computation, and therefore interpretation of interactions, that any users of this package ensure that all uncertainty data corresponds to standard deviation and not alternative metrics (e.g., variance, standard error, or confidence intervals). Similarly, all response metric data should correspond to means, not other averaged metrics (e.g., medians). Additionally, values for response metrics should remain untransformed as this will disrupt the true interpretation of the interactions. For example, log-transforming data and then implementing the additive model will actually implement a version of the multiplicative null model (Griffen et al. 2016). This is a particularly pertinent consideration when analysing response metrics such as specific growth rates. Finally, there are some limitations on the values that mean treatment values can take, and below we detail the affected functions.

Table 1: Variables required in order to implement the *multiplestressR* package.

Argument	Description
Control_Mean	Mean value of the response metric for the control treatment.
Control_SD	Standard deviation for *Control_Mean.*
Control_N	Sample size for the control treatment.
StressorA_Mean	Mean value of the response metric for the treatment of stressor A only.
StressorA_SD	Standard deviation for *StressorA_Mean.*
StressorA_N	Sample size for the treatment of stressor A only.
StressorB_Mean	Mean value of the response metric for the treatment of stressor B only.
StressorB_SD	Standard deviation for *StressorB_Mean.*
StressorB_N	Sample size for the treatment of stressor B only.
StressorsAB_Mean	Mean value of the response metric for the treatment of both stressors A and B together.
StressorsAB_SD	Standard deviation for *StressorsAB_Mean.*
StressorsAB_N	Sample size for the treatment of both stressors A and B together.

3   multiplestressR functions

The *multiplestressR* package comprises five functions and one example dataset. Example code is provided for each function and the code is easily adaptable so that researchers can modify the example to be used for their own analyses. A brief summary of each function and dataset is provided in Table 2, with required arguments for each function also described.

Table 2: Overview of the functions and datasets included within the *multiplestressR* package.

Function and arguments	Description
effect_size_additive	Implementation of the additive null model for factorial datasets (Equation S1; Gurevitch et al. 2000).
Data arguments (Table 1); *Significance_Level*; *Small_Sample_Correction*	The *Significance_Level* argument is used to specify the level of alpha for this analysis (default = 0.05).
	The *Small_Sample_Correction* argument is used to specify whether the bias correction factor should be incorporated within calculations (see Equation 1) of the additive effect size (TRUE or FALSE; default = TRUE).
effect_size_multiplicative	Implementation of the multiplicative null model for factorial datasets (Equation S2; Lajeunesse 2011).
Data arguments (Table 1); *Significance_Level*	The *Significance_Level* argument is used to specify the level of alpha for this analysis (default = 0.05).
classify_interactions	Assign interaction classifications (Jackson et al. 2016) to the output of either *effect_size_additive* or *effect_size_multiplicative* (Figure 2).
effect_size_dataframe; *assign_reversals*; *remove_directionality*	The *effect_size_dataframe* argument corresponds to the output of either the *effect_size_additive* or *effect_size_multiplicative* functions.
	The *assign_reversals* argument is used to specify whether, following the method of Jackson et al. (2016), reversal interactions should be distinguished from antagonisms (TRUE or FALSE; default = TRUE).
	The *remove_directionality* argument is used to specify whether, following the method of Jackson et al. (2016), directionality should be removed from the effect sizes (TRUE or FALSE; default = FALSE).
summary_plots	Generate summary figures using the output of *classify_interactions* (Figure 2).
effect_size_dataframe; *Significance_Level*; *Small_Sample_Correction*	The *effect_size_dataframe* argument corresponds to the output of the *classify_interactions* function.
	The arguments *Significance_Level* and *Small_Sample_Correction* should match the values specified for the *effect_size_additive* or *effect_size_multiplicative* functions (default values are same as above). Note that if the *effect_size_multiplicative* function was used, the *Small_Sample_Correction* argument can be ignored.
critical_effect_size_additive	Determine the critical (additive) effect size for a given experiment design (Equation S3; Burgess et al. 2022).
Control_N; *StressorA_N*; *StressorB_N*; *StressorsAB_N*; *Significance_Level*; *Small_Sample_Correction*	The *Control_N*, *StressorA_N*, *StressorB_N*, and *StressorsAB_N* are used to specify the sample sizes for each treatment of the factorially designed experiments (see Table 1).
	The *Significance_Level* argument is used to specify the level of alpha for this analysis (default = 0.05).
	The *Small_Sample_Correction* argument is used to specify whether the bias correction factor should be incorporated within calculations (see Equation 1) of the additive effect size (TRUE or FALSE; default = TRUE).
survival	A randomly computer-generated dataset comprising survival data for 250 interactions.

The functions within the *multiplestressR* package have been written in a modular manner, and users of the package need only run the functions appropriate for their analyses (i.e., there is no requirement to run the *classify_interactions* or *summary_plots* functions). Examples of potential workflows are illustrated in Figure 2. However, it is important to note that the *classify_interactions* function can only be applied to the output of either the *effect_size_additive* or *effect_size_multiplicative* functions. Likewise, the *summary_plots* function can only be applied to the output of the *classify_interactions* function. However, both the *effect_size_additive and effect_size_multiplicative* functions can be used independently of the *classify_interactions* and *summary_plots* functions. Likewise, the *critical_effect_size_additive* function is a stand-alone function.
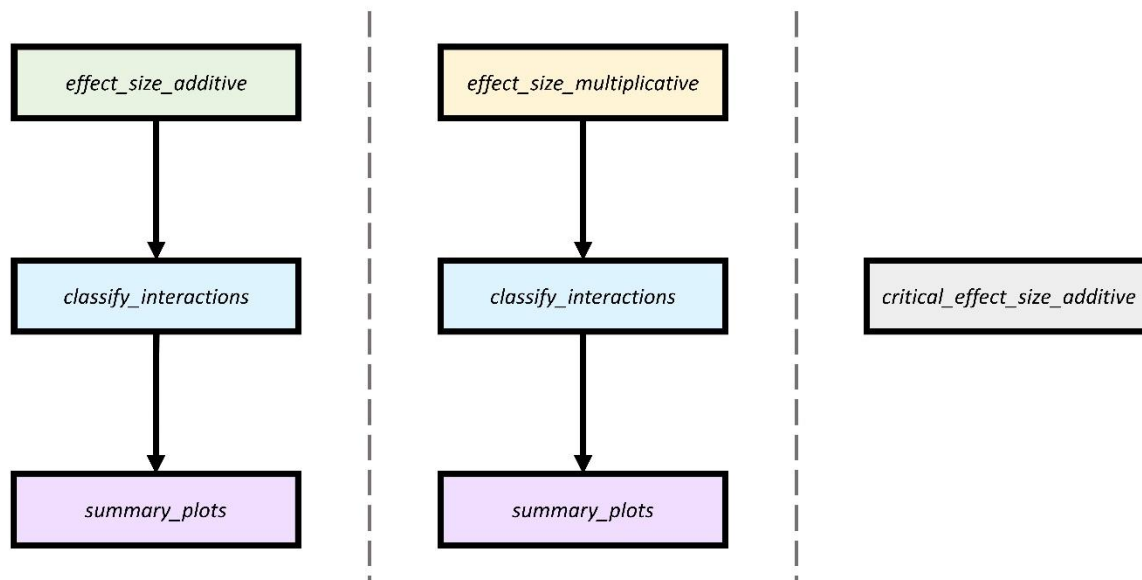

Figure 2: Example workflows for the functions within the *multiplestressR* package

3.1   *survival*

The *multiplestressR* package contains an example dataset named *survival*. This computer-generated dataset comprises simulated data for the survival rates of 250 populations (each composed of 100 individuals) exposed to the notional stressors of increased temperature and decreased pH. The dataset uses a factorial design comprising four treatments: i) Control conditions; ii) Exposed to increased temperature; iii) Exposed to decreased pH; iv) Exposed to both increased temperature and decreased pH. The data for each treatment (means, standard deviations, and sample sizes)

8

corresponds to the data described within Table 1. The main use of this dataset is to illustrate the functionality of the *multiplestressR* package, and all subsequent examples of code use this dataset. The *survival* dataset can be obtained in R using the code shown in Figure 3. The *survival* dataset also acts as a useful guide for the structure for any datasets which researchers may wish to analyse.

```
df <- multiplestressR::survival
```

Figure 3: Code demonstrating the easy extraction of the *survival* dataset from the *multiplestressR* package.

3.2   *effect_size_additive*

Both the *effect_size_additive* and *effect_size_multiplicative* functions calculate effect sizes which are in turn used to classify stressor interactions. In brief, effect sizes are a widely used statistical tool and are central to null hypothesis significance testing (Nakagawa & Cuthill 2007). Effect sizes provide a standardised approach to comparing groups, often focussing on group means. Indeed, effect sizes are calculated in many commonplace statistical tools (e.g., t-tests or analysis of variance) which compare group means.

When applying a null model in multiple stressor ecology, the primary research question of interest is whether the observed effect of two stressors differ from a prediction based upon the effects of these stressors acting independently. In other words, we are comparing two groups, an *observed* effect and a *predicted* effect based upon either the additive or multiplicative null model. Accordingly, these null models can be easily, and rigorously, implemented through the factorial forms of two established effect sizes Hedges'd (Gurevitch et al. 2000) and the response ratio (Lajeunesse 2011). For a detailed explanation of effect sizes in biological sciences see Nakagawa & Cuthill (2007), while both Burgess et al. (2022) and Schäfer & Piggott (2018) discuss effect sizes and null models in the context of multiple stressor ecology.

The additive null model is the most commonly applied null model (e.g., Crain et al. 2008; Burgess et al. 2021), although there is variation between studies in how it is applied. Here, the *effect_size_additive* function implements the additive null model through the factorial form of Hedges' d (Equation S1; Gurevitch et al. 2000), perhaps the most widely-used form of the additive null model. Previous research (Burgess et al. 2022) has shown this particular form of the additive null model to be statistically rigorous and robust. An example of R code for the *effect_size_additive* function is shown in Figure 4.

9

```
df_additive <- effect_size_additive(
        Control_N          = df$Sample_Size_Control,
        Control_SD         = df$Standard_Deviation_Control,
        Control_Mean       = df$Mean_Control,
        StressorA_N        = df$Sample_Size_Temperature,
        StressorA_SD       = df$Standard_Deviation_Temperature,
        StressorA_Mean     = df$Mean_Temperature,
        StressorB_N        = df$Sample_Size_pH,
        StressorB_SD       = df$Standard_Deviation_pH,
        StressorB_Mean     = df$Mean_pH,
        StressorsAB_N      = df$Sample_Size_Temperature_pH,
        StressorsAB_SD     = df$Standard_Deviation_Temperature_pH,
        StressorsAB_Mean   = df$Mean_Temperature_pH,
        Significance_Level = 0.05)
```

Figure 4: Code demonstrating the use of the *effect_size_additive* function.

The *effect_size_additive* function returns a data frame which contains seventeen columns. The first twelve columns correspond to the input data for the function (see Table 1), while the final five columns are as follows:

i)   *Interaction_Effect_Size*;
ii)   *Interaction_Variance*;
iii)   *Interaction_CI_Upper*;
iv)   *Interaction_CI_Lower*;
v)   *Null_Model.*

Interaction_Effect_Size corresponds to the effect size calculated by the additive null model. The interpretation of this metric is relatively simple in that larger effect sizes indicate a greater statistical ability to distinguish between observed and predicted responses. Additionally, each value of *Interaction_Effect_Size* also has an associated variance, shown by *Interaction_Variance*. The *Interaction_Variance* variable is in turn used to calculate the corresponding confidence intervals for each value of *Interaction_Effect_Size*, shown by the variables *Interaction_CI_Upper* and *Interaction_CI_Lower*. By default, 95% confidence intervals are calculated, however this can be altered by using the *Significance_Level* argument (see Table 2 and Figure 4). Finally, the variable *Null_Model* details which null model was used to calculate the effect sizes, variances, and associated confidence intervals.

10

3.3   *effect_size_multiplicative*

The multiplicative null model is becoming increasingly popular within multiple stressor research (e.g., Harvey et al. 2013; Gomez Isaza et al. 2020), and is recognised as potentially being more ecologically realistic than the additive null model (Schäfer & Piggott, 2018). Here, the *effect_size_multiplicative* function implements the multiplicative null model through the factorial form of the response ratio (Equation S2; Lajeunesse 2011), which has been shown to be statistically robust (Burgess et al. 2022). From a much simplified perspective, the multiplicative null model can be viewed as the logarithmic form of the additive null model, with this transformation resulting in fundamental differences in the assumptions of these approaches. However, it is important to note that due to the logarithmic nature of the *effect_size_multiplicative* function, mean values for a given response metric must be greater than zero.

The *effect_size_multiplicative* function requires exactly the same input parameters as the *effect_size_additive* function (Table 2), with the exception that the *Small_Sample_Correction* argument is not utilised. An example of R code for the *effect_size_multiplicative* function is shown in Figure 5.

```
df_multiplicative <- effect_size_multiplicative(
            Control_N          = df$Sample_Size_Control,
            Control_SD         = df$Standard_Deviation_Control,
            Control_Mean       = df$Mean_Control,
            StressorA_N        = df$Sample_Size_Temperature,
            StressorA_SD       = df$Standard_Deviation_Temperature,
            StressorA_Mean     = df$Mean_Temperature,
            StressorB_N        = df$Sample_Size_pH,
            StressorB_SD       = df$Standard_Deviation_pH,
            StressorB_Mean     = df$Mean_pH,
            StressorsAB_N      = df$Sample_Size_Temperature_pH,
            StressorsAB_SD     = df$Standard_Deviation_Temperature_pH,
            StressorsAB_Mean   = df$Mean_Temperature_pH,
            Significance_Level = 0.05)
```

Figure 5: Code demonstrating the use of the *effect_size_multiplicative* function.

The structure of the output of the *effect_size_multiplicative* function is identical to that of the *effect_size_additive* function. As such, refer to the description of the output for the *effect_size_additive* function above.

3.4   *classify_interactions*

There is considerable interest in documenting how stressors can interact to affect a given system, and numerous studies have described synergisms or antagonisms between various combinations of stressors (e.g., Jackson et al. 2016). Accordingly, the *classify_interactions* function can be applied to the output of either the *effect_size_*additive or *effect_size_*multiplicative functions and be used to classify each stressor interaction into one of four categories. Namely, such categories are i) synergistic, ii) antagonistic, iii) null, and iv) reversal interactions. Interaction classifications are assigned according to interaction effect sizes (*Interaction_Effect_Size*) and associated confidence intervals (*Interaction_CI_Upper* and *Interaction_CI_Lower*). For instance, if the confidence intervals for an effect size overlap zero, then this interaction will be assigned a *null* classification as there is insufficient evidence to overturn the null expectation. In contrast, if confidence intervals for an effect size do not overlap zero, then this interaction will be assigned a classification of either *antagonistic, synergistic,* or *reversal*. The method for classifying interactions within the *classify_interactions* function is adopted from the methodology outlined by Burgess et al. (2022). An example of R code for the *classify_interactions* function is shown in Figure 6.

```
df_additive_classified <- classify_interactions(
                    effect_size_dataframe = df_additive,
                    assign_reversals      = TRUE,
                    remove_directionality = TRUE)
```

Figure 6: Code demonstrating the use of the *classify_interactions* function.

To date, most studies have only considered the interaction classifications of *antagonistic, synergistic,* and *null* interactions. However, there is increasing discussion of the additional category of *reversal* interactions. While the exact nomenclature may vary (e.g., Travers-Trolet et al. 2014; Jackson et al. 2016) reversal interactions can be regarded as a specific subset of antagonisms. As such, antagonisms

and reversals interactions have frequently been combined into a single classification. A detailed description of all four interaction classifications is available elsewhere (see Burgess et al. 2022; Jackson et al. 2016; Orr et al. 2020); however, the *classify_interactions* function provides users with the option to specify whether reversal and antagonistic interactions should be differentiated through the *assign_reversals* argument (see Table 2, Figure 6).

The *classify_interactions* function contains an argument (*remove_directionality*) for removing directionality from a dataset (see Table 2). This argument corresponds to a phenomenon described by Jackson et al. (2016) and Piggott et al. (2015) where the comparability of effect sizes may be compromised. Jackson et al. (2016) first implemented a method for accounting for this limitation which has subsequently been adopted by other studies (e.g., Siviter et al. 2021). Accordingly, this approach can be easily implemented through the *remove_directionality* argument. The *remove_*directionality argument is only likely to be required under particular conditions; however, for a full discussion of this approach, see Jackson et al. (2016).

The output of this function is straightforward. The input data frame (i.e., the *effect_size_dataframe* argument) is returned with an additional column titled *Interaction_Classification*. For each effect size, the *Interaction_Classification* is assigned a value of either *Null, Antagonisic, Synergistic,* or *Reversal*.

3.5   summary_plots

Within the *multiplestressR package,* statistical analyses are conducted using the *effect_size_additive, effect_size_multiplicative*, and *classify_interactions* functions. However, the results of these analyses can be easily visualised (and hence interpreted) using the *summary_plots* function. The *summary_plots* function can only be applied to the output of the *classify_interactions* function and can be used to generate a series of figures which can aid with the interpretation of these analyses; however, it is primarily of use where multiple interactions have been analysed. An example of R code for the *summary_plots* function is shown in Figure 7, while an example output of the *summary_plots* function is illustrated by Figure 8.

```
additive_plots <- summary_plots(
                effect_size_dataframe = df_additive_classified,
                Significance_Level    = 0.05)
```

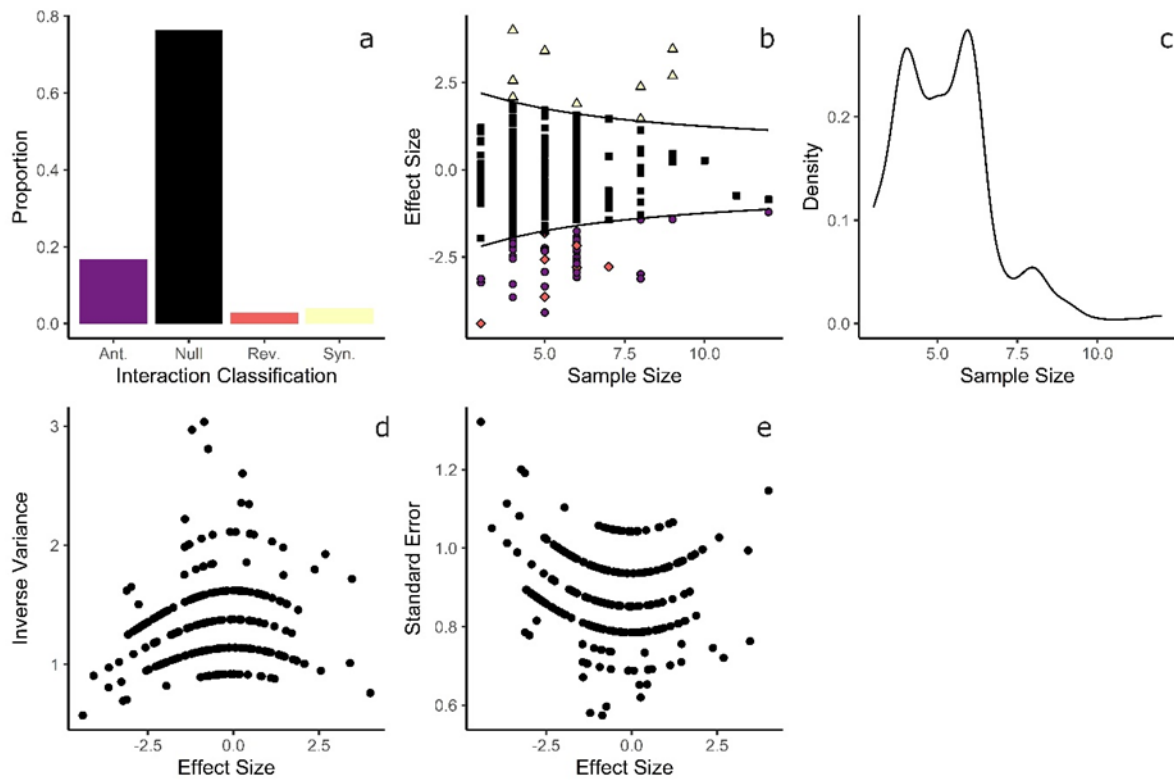Figure 7: Code demonstrating the use of the *summary_plots* function.

Figure 8: Example output of the *summary_plots* function using the example *survival* dataset. The code in Figures 3, 4, 6, and 7 can be used to generate these figures. Note that labels have been included for clarity but are not normally incorporated by the *summary_plots* function.

As shown by Figure 8, the output of the *summary_plots* function produces a figure comprising five panels. Figure 8a shows the proportions of each of the interaction classifications across the entire dataset; here it is clear to see that the majority of interactions are assigned a null classification. Figure 8b plots mean treatment sample sizes against interaction effect size. Where the additive null model has been implemented, lines are plotted which denote the smallest effect size (hereafter referred to as the critical effect size) that can lead to the rejection of the null expectation for a given mean treatment sample size (see *critical_effect_size_additive* function below). Figure 8b can therefore provide insights into whether a predominance of null interactions is likely to be due to ecological variables or simply underpowered experiments. Figure 8c supports Figure 8b by providing a density plot for mean treatment sample sizes. Figures 8d and 8e illustrate funnel plots plotting interaction effect size against inverse effect size variance and effect size standard error respectively. A detailed explanation of funnel plots is provided by Sterne et al. (2011); however, these funnel plots may be of primary use for researchers seeking to conduct a meta-analysis. Additionally, for the additive null model, it is important to note that sample sizes play a prominent role in the calculation of the effect size variance (Figure 8d) and hence effect size standard error (Figure 8e). Accordingly, under

commonplace sample sizes for multiple stressor experiments (see Burgess et al. 2022), calculations of effect size variance may be driven by sample size. This is evidenced by the various 'bands' in Figures 8d and 8e which correspond to mean treatment sample sizes of three, four, five, six, seven, and eight. Finally, these figures are generated using *ggplot2* (Wickham 2016) meaning that they are easily modified (e.g., colour scheme).

3.6   *critical_effect_size_additive*

The *critical_effect_size_additive* function is a stand-alone function which can be used to calculate the critical effect size for the additive null model (see Burgess et al. 2022) but it is not required to complete analysis and assign interaction classifications.

Broadly, the critical effect size is the minimum effect size that an experiment is capable of detecting as being significantly different to the null hypothesis (Lakens 2022; Mudge 2013). In brief, experiments with greater sample sizes are able to detect smaller effects which are significantly different to the null model. Accordingly, an experiment which has a large critical effect size (i.e., small sample sizes) may miss biologically important synergistic or antagonistic interactions (Burgess et al., 2022). Ultimately, it is up to any researcher to determine what constitutes the minimum effect size of interest for their system (Burgess et al., 2022; Lakens et al. 2022). However, it is possible to calculate the critical effect size, for the additive null model, for a given experimental design (see Equation S3). As such, the *critical_effect_size_additive* function may aid researchers in the design of their experiments. An example of R code for the *critical_effect_size_additive* function is shown in Figure 9.

```
es_critical <- critical_effect_size_additive(
          Control_N    = 4,
          StressorA_N   = 4,
          StressorB_N   = 4,
          StressorsAB_N = 4)
```

Figure 9: Code demonstrating the use of the *critical_effect_size_additive* function.

The output of the *critical_effect_size_additive* function is straightforward to understand. For each combination of sample sizes analysed a single value is provided. For example, using the code in Figure 9 the *critical_effect_size_additive* function would return a value of 1.94 (to three decimal places) and

the null model expectation could only be rejected for those stressor combinations that yield an effect size with an absolute value greater than this number.

4   Conclusion

As demonstrated, the *multiplestressR* package allows researchers to easily implement either the additive or multiplicative null models in order to infer how pairs of stressors combine in their factorial datasets. The *multiplestressR* package allows relatively complex analyses to be easily conducted using only a few lines of code. More importantly, the *multiplestressR* package will allow analyses to be standardised across multiple independent studies, and also prevent any small calculation errors from affecting into the results. The computations carried out within *multiplestressR* are also useful for the generation of effect sizes that are then used in meta-analytical studies (e.g., Burgess et al. 2021; Jackson et al. 2016). Accordingly, standardising effect sizes will also allow for better comparison of future meta-analyses and hopefully lead to more general patterns emerging within the general field of biological responses to multiple stressors.

Acknowledgements

References

Beaumelle, L., De Laender, F., & Eisenhauer, N. (2020). Biodiversity mediates the effects of stressors but not nutrients on litter decomposition. *Elife*, 9, e55659.

Burgess, B. J., Purves, D., Mace, G., & Murrell, D. J. (2021). Classifying ecosystem stressor interactions: Theory highlights the data limitations of the additive null model and the difficulty in revealing ecological surprises. *Global Change Biology*, 27(13), 3052-3065.

Burgess, B. J., Jackson, M. C., & Murrell, D. J. (2022). Are experiment sample sizes adequate to detect biologically important interactions between multiple stressors?. *bioRxiv*, 2021-07.

Christensen, M. R., Graham, M. D., Vinebrooke, R. D., Findlay, D. L., Paterson, M. J., & Turner, M. A. (2006). Multiple anthropogenic stressors cause ecological surprises in boreal lakes. *Global Change Biology*, 12(12), 2316-2322.

Côté, I. M., Darling, E. S., & Brown, C. J. (2016). Interactions among ecosystem stressors and their importance in conservation. *Proceedings of the Royal Society B: Biological Sciences*, 283(1824), 20152592.

Crain, C. M., Kroeker, K., & Halpern, B. S. (2008). Interactive and cumulative effects of multiple human stressors in marine systems. *Ecology Letters*, 11(12), 1304-1315.

Darling, E. S., & Côté, I. M. (2008). Quantifying the evidence for ecological synergies. *Ecology Letters*, 11(12), 1278-1286.

Fleishman, E., Blockstein, D. E., Hall, J. A., Mascia, M. B., Rudd, M. A., Scott, J. M., ... & Vedder, A. (2011). Top 40 priorities for science to inform US conservation and management policy. *BioScience*, 61(4), 290-300.

Garnier, S., Ross N., Rudis R., Camargo A.P., Sciaini M., and Scherer C. (2021). Rvision - Colorblind-Friendly Color Maps for R. *R package version 0.6.2*.

Gomez Isaza, D. F., Cramp, R. L., & Franklin, C. E. (2020). Living in polluted waters: A meta-analysis of the effects of nitrate and interactions with other environmental stressors on freshwater taxa. *Environmental Pollution*, 261, 114091.

Griffen, B. D., Belgrad, B. A., Cannizzo, Z. J., Knotts, E. R., & Hancock, E. R. (2016). Rethinking our approach to multiple stressor studies in marine environments. *Marine Ecology Progress Series*, 543, 273-281.

Gurevitch, J., Morrison, J.A. and Hedges, L.V. (2000). The interaction between competition and predation: a meta-analysis of field experiments. *The American Naturalist*, 155(4), 435-453.

Halpern, B. S., Frazier, M., Potapenko, J., Casey, K. S., Koenig, K., Longo, C., ... & Walbridge, S. (2015). Spatial and temporal changes in cumulative human impacts on the world's ocean. *Nature Communications*, 6(1), 1-7.

Harvey, B. P., Gwynn-Jones, D., & Moore, P. J. (2013). Meta-analysis reveals complex marine biological responses to the interactive effects of ocean acidification and warming. *Ecology and Evolution*, 3(4), 1016-1030.

Hillebrand, H. (2002). Top-down versus bottom-up control of autotrophic biomass—a meta-analysis on experiments with periphyton. *Journal of the North American Benthological Society*, 21(3), 349-369.

Jackson, M. C., Loewen, C. J., Vinebrooke, R. D., & Chimimba, C. T. (2016). Net effects of multiple stressors in freshwater ecosystems: a meta-analysis. *Global Change Biology*, 22(1), 180-189.

Kerkhoff, A. J., & Enquist, B. J. (2009). Multiplicative by nature: why logarithmic transformation is necessary in allometry. *Journal of Theoretical Biology*, 257(3), 519-521.

Lajeunesse, M.J. (2011). On the meta-analysis of response ratios for studies with correlated and multi-group designs. *Ecology*, 92(11), 2049-2055.

Lakens, D. (2022). Sample size justification. *Collabra: Psychology*, 8(1), 33267.

Lange, K., Bruder, A., Matthaei, C. D., Brodersen, J., & Paterson, R. A. (2018). Multiple-stressor effects on freshwater fish: Importance of taxonomy and life stage. *Fish and Fisheries*, 19(6), 974-983.

Mudge, J. F. (2013). Explicit consideration of critical effect sizes and costs of errors can improve decision-making in plant science. *New Phytologist*, 199(4), 876-878.

Nakagawa, S., & Cuthill, I. C. (2007). Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biological Reviews*, 82(4), 591-605.

O'Hara, C. C., Frazier, M., & Halpern, B. S. (2021). At-risk marine biodiversity faces extensive, expanding, and intensifying human impacts. Science, 372(6537), 84-87.

Orr, J. A., Vinebrooke, R. D., Jackson, M. C., Kroeker, K. J., Kordas, R. L., Mantyka-Pringle, C., ... & Piggott, J. J. (2020). Towards a unified study of multiple stressors: divisions and common goals across research disciplines. *Proceedings of the Royal Society B*, 287(1926), 20200421.

Paine, R. T., Tegner, M. J., & Johnson, E. A. (1998). Compounded perturbations yield ecological surprises. *Ecosystems*, 1(6), 535-545.

Pedersen, T.L. (2020). patchwork: The Composer of Plots. *R package version 1.1.1*. https://CRAN.R-project.org/package=patchwork

Piggott, J. J., Townsend, C. R., & Matthaei, C. D. (2015). Reconceptualizing synergism and antagonism among multiple stressors. *Ecology and Evolution*, 5(7), 1538-1547.

Przeslawski, R., Byrne, M., & Mellin, C. (2015). A review and meta-analysis of the effects of multiple abiotic stressors on marine embryos and larvae. *Global Change Biology*, 21(6), 2122-2140.

Reid, A. J., Carlson, A. K., Creed, I. F., Eliason, E. J., Gell, P. A., Johnson, P. T., ... & Cooke, S. J. (2019). Emerging threats and persistent conservation challenges for freshwater biodiversity. *Biological Reviews*, 94(3), 849-873.

Schäfer, R. B., & Piggott, J. J. (2018). Advancing understanding and prediction in multiple stressor research through a mechanistic basis for null models. *Global Change Biology*, 24(5), 1817-1826.

Seifert, M., Rost, B., Trimborn, S., & Hauck, J. (2020). Meta-analysis of multiple driver effects on marine phytoplankton highlights modulating role of pCO2. *Global Change Biology*, 26(12), 6787-6804.

Simmons, B. I., Blyth, P. S., Blanchard, J. L., Clegg, T., Delmas, E., Garnier, A., ... & Beckerman, A. P. (2021). Refocusing multiple stressor research around the targets and scales of ecological impacts. *Nature Ecology & Evolution*, 5(11), 1478-1489.

Siviter, H., Bailes, E. J., Martin, C. D., Oliver, T. R., Koricheva, J., Leadbeater, E., & Brown, M. J. (2021). Agrochemicals interact synergistically to increase bee mortality. *Nature*, 596(7872), 389-392.

Sterne, J. A., Sutton, A. J., Ioannidis, J. P., Terrin, N., Jones, D. R., Lau, J., ... & Higgins, J. P. (2011). Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ*, 343.

Thompson, P. L., MacLennan, M. M., & Vinebrooke, R. D. (2018). An improved null model for assessing the net effects of multiple stressors on communities. *Global Change Biology*, 24(1), 517-525.

Travers-Trolet, M., Shin, Y. J., Shannon, L. J., Moloney, C. L., & Field, J. G. (2014). Combined fishing and climate forcing in the southern Benguela upwelling ecosystem: an end-to-end modelling approach reveals dampened effects. *PLoS One*, 9(4), e94286.

Wickham H. ggplot2: Elegant Graphics for Data Analysis. *Springer-Verlag New York*

Yue, K., Fornara, D. A., Yang, W., Peng, Y., Li, Z., Wu, F., & Peng, C. (2017). Effects of three global change drivers on terrestrial C: N: P stoichiometry: a global synthesis. *Global Change Biology*, 23(6), 2450-2463.

Appendix

Required Variables

Table S1: Variables required in order to implement the *multiplestressR* package. The notation used within Equations S1-S3 is also shown.

Argument	Description	Notation
Control_Mean	Mean value of the response metric for the control treatment.	X_C
Control_SD	Standard deviation for *Control_Mean.*	SD_C
Control_N	Sample size for the control treatment.	N_C
StressorA_Mean	Mean value of the response metric for the treatment of stressor A only.	X_A
StressorA_SD	Standard deviation for *StressorA_Mean.*	SD_A
StressorA_N	Sample size for the treatment of stressor A only.	N_A
StressorB_Mean	Mean value of the response metric for the treatment of stressor B only.	X_B
StressorB_SD	Standard deviation for *StressorB_Mean.*	SD_B
StressorB_N	Sample size for the treatment of stressor B only.	N_B
StressorsAB_Mean	Mean value of the response metric for the treatment of both stressors A and B together.	X_I
StressorsAB_SD	Standard deviation for *StressorsAB_Mean.*	SD_I
StressorsAB_N	Sample size for the treatment of both stressors A and B together.	N_I

Additive Null Model Equations

The *multiplestressR* package implements the widely used version of the additive null model defined by Gurevitch et al. (2000). The notation for variables used throughout Equation S1 is detailed in Table S1. The calculation of the additive null model (i.e., Hedges' d, ES_{Add}) is shown by Equation S1.1, where the pooled standard deviation (s) is shown by Equation S1.2.

S1.1)
$$ES_{Add} = \frac{X_I - X_A - X_B + X_C}{s} \cdot J(m)$$

S1.2)
$$s = \sqrt{\frac{(N_I - 1) \cdot (SD_I)^2 + (N_A - 1) \cdot (SD_A)^2 + (N_B - 1) \cdot (SD_B)^2 + (N_C - 1) \cdot (SD_C)^2}{N_I + N_A + N_B + N_C - 4}}$$

For each value of ES_{Add}, there is an associated effect size variance (V_{Add}) shown by Equation S1.3. V_{Add} is in turn used to calculate the standard error (Equation S1.4) and confidence intervals (Equation S1.5) for each value of ES_{Add}. For Equation S1.5, the default value of α is 0.05 (i.e., $Z_{\alpha/2} \approx 1.96$), resulting in 95% confidence intervals being calculated. However, this can be altered using the *Significance_Level* argument.

S1.3)
$$V_{Add} = (J(m))^2 \cdot \left(\frac{1}{N_I} + \frac{1}{N_A} + \frac{1}{N_B} + \frac{1}{N_C} + \frac{(ES_{Add})^2}{2 \cdot (N_I + N_A + N_B + N_C)}\right)$$

S1.4)
$$SE_{Add} = \sqrt{V_{Add}}$$

S1.5)
$$CI_{Add} = Z_{\alpha/2} \cdot SE_{Add}$$

Hedges' d has a known bias, where sample treatment sample sizes lead to the effect size being overestimated. Accordingly, within the *effect_size_additive* function, there is the option of specifying the *Small_Sample_Correction* argument with determines whether a bias correction factor ($J(m)$) is implemented. If the *Small_Sample_Correction* is specified as *TRUE* (the default value), then $J(m)$ calculated as (Equation S1.6):

S1.6)
$$J(m) = 1 - \frac{3}{4 \cdot (N_I + N_A + N_B + N_C - 4) - 1}$$

If the *Small_Sample_Correction* is specified as *FALSE*, then $J(m)$ calculated as (Equation S1.7):

S1.7)
$$J(m) = 1$$

Multiplicative Null Model Equations

The *multiplestressR* package uses the factorial extension to the multiplicative null model derived by Lajeunesse (2011). The notation for variables used throughout Equation S2 is detailed in Table S1. The calculation of the multiplicative null model (i.e., Response Ratio, ES_{Mul}) is shown by Equation S2.1.

$S2.1)$
$$ES_{Mul} = \ln(X_I) - \ln(X_A) - \ln(X_B) + \ln(X_C)$$

For each value of ES_{Mul}, there is an associated effect size variance (V_{Mul}) shown by Equation S2.2. V_{Mul} is in turn used to calculate the standard error (Equation S2.3) and confidence intervals (Equation 2.4) for each value of ES_{Mul}. For Equation S2.4, the default value of α is 0.05 (i.e., $Z_{\alpha/2} \approx 1.96$), resulting in 95% confidence intervals being calculated. However, this can be altered using the *Significance_Level* argument.

$S2.2)$
$$V_{Mul} = \frac{(SD_I)^2}{(X_I)^2 \cdot N_I} + \frac{(SD_A)^2}{(X_A)^2 \cdot N_A} + \frac{(SD_B)^2}{(X_B)^2 \cdot N_B} + \frac{(SD_C)^2}{(X_C)^2 \cdot N_C}$$

$S2.3)$
$$SE_{Mul} = \sqrt{V_{Mul}}$$

$S2.4)$
$$CI_{Mul} = Z_{\alpha/2} \cdot SE_{Mul}$$

Critical Effect Size (Additive) Equations

Within the *multiplestressR* package, we use the critical effect size equation derived by Burgess et al. (2022) from the additive null model defined by Gurevitch et al. (2000) and defined in equations S1. This means the critical effect size computed is only applicable to this null model. The notation for variables used throughout Equation S3 is detailed in Table S1. The calculation of the critical effect size for the additive null model is shown by Equation S3.

$$S3) \quad Crit_{Add} \ = \ \sqrt{\left(\frac{1}{N_C} + \frac{1}{N_A} + \frac{1}{N_B} + \frac{1}{N_I}\right) \cdot \left(\frac{2 \cdot (Z_{\alpha/2})^2 \cdot (N_C \ + \ N_A \ + \ N_B \ + \ N_I \ ) \cdot \left(J(m)\right)^2}{2 \cdot (N_C \ + \ N_A \ + \ N_B \ + \ N_I \ ) - \ (Z_{\alpha/2})^2 \ \cdot \ \left(J(m)\right)^2}\right)}$$

The default value of α (Equation S3) is 0.05 (i.e., $Z_{\alpha/2}$ ≈ 1.96), resulting in 95% confidence intervals being calculated. Although, as with the *effect_size_additive* and *effect_size_multiplicative* functions, this can be altered using the *Significance_Level* argument.

As with the *effect_size_additive* function, the *Small_Sample_Correction* argument can be used to specify whether the bias correction factor ($J(m)$) for small sample sizes is included (see Equations S1.6-S1.7).