Inferring competitive outcomes, ranks and intransitivity from empirical data: A comparison of different methods

Yanhao Feng1,2,3 | Santiago Soliveres4,5 | Eric Allan5 | Benjamin Rosenbaum2,6 | Cameron Wagg7 | Andrea Tabi7 | Enrica De Luca7 | Nico Eisenhauer2,8 | Bernhard Schmid7 | Alexandra Weigelt2,8 | Wolfgang W. Weisser9 | Christiane Roscher1,2 | Markus Fischer5

1Department of Physiological Diversity, Helmholtz Centre for Environmental Research, UFZ, Leipzig, Germany; 2German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany; 3State Key Laboratory of Grassland Agro-ecosystems, College of Pastoral Agriculture Science and Technology, Lanzhou University, Lanzhou, China; 4Department of Ecology, University of Alicante, Alicante, Spain; 5Institute of Plant Sciences, University of Bern, Bern, Switzerland; 6Institute of Biodiversity, University of Jena, Jena, Germany; 7Department of Evolutionary Biology and Environmental Studies, University of Zurich, Zurich, Switzerland; 8Institute of Biology, University of Leipzig, Leipzig, Germany and 9Terrestrial Ecology Research Group, Department of Ecology and Ecosystem Management, School of Life Sciences Weihenstephan, Technical University of Munich, Freising, Germany

Correspondence
Yanhao Feng
Email: yanhao.feng@outlook.com

Funding information
Deutsche Forschungsgemeinschaft, Grant/Award Number: FOR 456/1451 and RO2397/8; Fundamental Research Funds for the Central Universities, Grant/ Award Number: lzujbky-2019-32; Spanish Government, Grant/Award Number: RYC-2016-20604

Handling Editor: Satu Ramula

Abstract

1. The inference of pairwise competitive outcomes (PCO) and multispecies competitive ranks and intransitivity from empirical data is essential to evaluate how competition shapes plant communities. Three categories of methods, differing in theoretical background and data requirements, have been used: (a) theoretically sound coexistence theory-based methods, (b) index-based methods, and (c) ‘process-from-pattern’ methods. However, how they are related is largely unknown.

2. In this study, we explored the relations between the three categories by explicitly comparing three representatives of them: (a) relative fitness difference (RFD), (b) relative yield (RY), and (c) a reverse-engineering approach (RE). Specifically, we first conducted theoretical analyses with Lotka–Volterra competition models to explore their theoretical linkages. Second, we used data from a long-term field experiment and a short-term greenhouse experiment with eight herbaceous perennials to validate the theoretical findings.

3. The theoretical analyses showed that RY or RE applied with equilibrium data indicated equivalent, or very similar, PCO to RFD, but these relations became weaker or absent with data further from equilibrium. In line with this, both RY and RE converged with RFD in indicating PCO over time in the field experiment as the communities became closer to equilibrium. Moreover, the greenhouse PCO (far from equilibrium) were only similar to the field PCO of earlier rather than later years. Intransitivity was more challenging to infer because it could be reshuffled by even a small competitive shift among similar competitors. For example, the field intransitivity inferred by three methods differed greatly: no intransitivity was detected with...
RFD; intransitivity detected with RY and RE was poorly correlated, changed substantially over time (even after equilibrium) and failed to explain coexistence.

4. Our findings greatly help the comparison and generalization of studies using different methods. For future studies, if equilibrium data are available, one can infer PCO and multispecies competitive ranks with RY or RE. If not, one should apply RFD with density gradient or time-series data. Equilibria could be evaluated with T tests or standard deviations. To reliably infer intransitivity, one needs high quality data for a given method to first accurately infer PCO, especially among similar competitors.

**KEYWORDS**
reverse-engineering approach, Chesson's coexistence theory, competitive outcomes, competitive ranks, intransitivity, niche differences, relative fitness differences, relative yield

---

1 | **INTRODUCTION**

The inference of long-term competitive outcomes from empirical data is essential to understand how competition shapes the structure, dynamics and functioning of plant communities (Aschehoug, Brooker, Atwater, Maron, & Callaway, 2016). For instance, to understand and predict plant coexistence, we need to explicitly estimate long-term competitive outcomes between plant species (Hart, Freckleton, & Levine, 2018), rather than merely the intensity of competition or overall competitive effects and responses. In theory, a more competitive species will ultimately exclude a less competitive one over the long term unless sufficient niche differences exist between the two in space and time (Chesson, 2000; Gause, 1934). Therefore, we need to focus on the long-term competitive outcomes at equilibrium, rather than the short-term outcomes that do not easily extrapolate to the effects of competition on population dynamics. Based on pairwise competitive outcomes (PCO), multispecies competitive ranks can be constructed (Keddy & Shipley, 1989). If the ranks are not hierarchical, intransitive competition (i.e. intransitivity) can emerge as in the ‘paper-rock-scissors’ game: species A beats B which beats C which in turn beats A (i.e. A > B > C > A) (Gallien, 2017). The existence of such intransitivity could promote multispecies coexistence and affect ecosystem functioning (Levine, Bascompte, Adler, & Allesina, 2017; Maynard, Crowther, & Bradford, 2017).

A number of methods have been developed thus far to infer PCO and thereby multispecies competitive ranks and intransitivity from empirical data (e.g. de Wit, 1960; Chesson, 2000; Ulrich, Soliveres, Kryszewski, Maestre, & Gotelli, 2014; see Table 1 for more examples). Given their differences in theoretical background and data requirements, we think these methods fall into three categories (Table 1). The first category of methods are rooted in coexistence theory, and they infer PCO from parameters in dynamic models that describe long-term competition between species (Chesson, 2000; MacArthur & Levins, 1967; Tilman, 1982; see Table 1). These methods provide a theoretically sound measure of PCO, because they indicate which species, in the absence of niche differences, is the competitive winner at equilibrium (Hart et al., 2018). Among them, relative fitness difference (RFD) derived from Chesson’s coexistence theory (Chesson, 2000) is the commonly used one in recent studies (e.g. Chu & Adler, 2015; Godoy, Kraft, & Levine, 2014). In theory, RFD is largely consistent with other measures from classical coexistence theories, e.g. MacArthur and Levins (1967)’s $K^*$, May (1974)’s $R^*$ and Tilman (1982)’s $R^*$ (Carroll, Cardinale, & Nisbet, 2011; Letten, Ke, & Fukami, 2017). However, these methods are challenging to apply, because they require a large amount of density gradient or time-series data (often from field experiments) to parameterize dynamic competition models. The second category comprises traditional indices that are often calculated by comparing the performance of a plant of a given species grown in monoculture (or alone) vs. mixture (Table 1; for a full review, see Weigelt & Jolliffe, 2003). Relative yield (RY), i.e. the ratio of a species’ yield in mixture to its yield in monoculture (de Wit, 1960), is the commonly used among them. The application of RY and other indices (e.g. relative mixture response, relative competition intensity) in this category requires much less data (often from greenhouse experiments). However, these indices have been frequently criticized as they may only indicate short-term rather than long-term competitive outcomes (Freckleton & Watkinson, 2000; Hart et al., 2018). The third category refers to the ‘process-from-pattern’ methods, which often use statistical tools to infer competition from observed species abundance data based on classic community assembly rules (Diamond, 1975), e.g. a reverse-engineering approach (RE), C-scores, multivariate logistic regression (Table 1). The RE is a recently developed method that applies Markov chain models to estimate competition matrices that best fit the observed species abundances (Ulrich et al., 2014). Within this category, the RE is the only one that can explicitly estimate PCO and has been frequently applied in recent studies (e.g. Maynard, Bradford, et al., 2017; Soliveres et al., 2015), but how it relates to methods in the other two categories is not well known (but see Soliveres et al., 2018). In short, the three categories of methods not only have different theoretical backgrounds, but also often use data from different sources (e.g. field experiments, greenhouse experiments, observations) which to some extent characterize different conditions of study systems (from equilibrium to non-equilibrium). However, it is largely unknown how different methods are related, which hinders the appropriate interpretation and application of them.
TABLE 1  The three categories of methods and their representatives to infer pairwise competitive outcomes and multispecies competitive ranks and intransitivity

| Category                        | Theory, representative, and related methods                                                                 | Data requirement                        | Merits and limitations                                                                 | Applications                                                                 |
|---------------------------------|-------------------------------------------------------------------------------------------------------------|----------------------------------------|----------------------------------------------------------------------------------------|----------------------------------------------------------------------------|
| Coexistence theory-based methods | Theory: coexistence theory                                                                                  | Density gradient or time-series data   | Merits: theoretically sound; Limitations: data requirement; model parameterization     | Relative fitness difference (e.g. Godoy et al., 2014; Godoy et al., 2017); Related methods (applications of R* reviewed in Miller et al., 2005) |
|                                 | Representative: relative fitness difference (Chesson, 2000)                                                |                                        |                                         |                                                                            |
|                                 | Related methods: R* (Tilman, 1982), $\frac{K_i}{K_j}$                                                      |                                        |                                         |                                                                            |
|                                 | (MacArthur & Levins, 1967), $\frac{a_i}{a_j}$ (May, 1974)                                                |                                        |                                         |                                                                            |
| Index-based methods              | Theory: theory implicit                                                                                    | Monoculture (or grown-alone) and mixture data | Merits: easy to apply; Limitations: theory implicit; not explicitly separating competitive differences from niche differences | Relative yield (e.g. Keddy & Shipley, 1989; Perkins, Holmes, & Weltzin, 2007; Walck, Baskin, & Baskin, 1999); Related methods (see examples in Weigelt & Jolliffe, 2003) |
|                                 | Representative: relative yield (de Wit, 1960)                                                                |                                        |                                         |                                                                            |
|                                 | Related methods: e.g. relative mixture response, relative interaction index, relative competition intensity, and many other indices (reviewed in Weigelt & Jolliffe, 2003) |                                        |                                         |                                                                            |
| ’Process-from-pattern’ methods   | Theory: community assembly rules (Diamond, 1975)                                                            | Temporal, spatial or temporal × spatial species abundance data | Merits: easy to apply (especially with multispecies communities); Limitations: not explicitly separating competitive differences from niche differences | The reverse-engineering approach (e.g. Soliveres et al., 2015; Ulrich et al., 2016); Related methods (Burns, 2007; Gotelli & Graves, 1996; Ovaskainen et al., 2010) |
|                                 | Representative: the reverse-engineering approach (Ulrich et al., 2014)                                       |                                        |                                         |                                                                            |
|                                 | Related methods: C-scores (Stone & Roberts, 1990), the null model approach (Gotelli & Graves, 1996), multivariate logistic regression (Ovaskainen, Hottola, & Siltanen, 2010) |                                        |                                         |                                                                            |

In this study, we chose three commonly used methods: RFD, RY and RE, as representatives of the three categories, respectively, and thoroughly compared them. As it is impossible to comprehensively compare the large number of existing methods (Table 1), and as different methods within each category share similarities, we chose specific methods as representatives of different categories to perform a full comparison. First, we used analytical approaches and simulations with Lotka-Volterra competition models to theoretically analyse how the three methods are related to infer PCO and multispecies competitive ranks and intransitivity, when using data from equilibrium to non-equilibrium conditions. Second, we used data from a long-term biodiversity experiment in the field and a short-term microcosm experiment in the greenhouse to validate the theoretical findings. Based on our results, we discuss other methods in the three categories and previous studies using different methods to draw general conclusions, and also provide recommendations for future studies to choose an appropriate method for a particular objective and situation.

2 | MATERIALS AND METHODS

2.1 | Theoretical background

2.1.1 | Three representative methods

The RFD between two species i and j is often defined as $f_i/f_j = \sqrt{\frac{a_i}{a_j}}$ with Lotka-Volterra competition models (Equation 1) (Chesson, 2013). Note, however, equivalent measures can also be derived from other phenomenological or mechanistic competition models (e.g. Carroll et al., 2011; Hart et al., 2018; Saavedra et al., 2017).

$$
\begin{align*}
N_{i,t+1} &= N_i e^{(1-\alpha_i N_i - \alpha_j N_j)} \\
N_{j,t+1} &= N_j e^{(1-\alpha_j N_i - \alpha_i N_j)},
\end{align*}
$$

(1)

where $N_{i,t+1}$ and $N_{j,t+1}$ and $N_{i,t}$ and $N_{j,t}$ are biomasses of species i and j at time $t+1$ and $t$, respectively, and $r_i$ and $r_j$ are intrinsic growth rates of species i and j, respectively, and alphas ($\alpha_i$, $\alpha_j$, $\alpha_{ij}$, $\alpha_{ji}$) are intraspecific and interspecific competition coefficients.

If $f_i/f_j > 1$, species i will outcompete species j in the absence of niche differences $\left(1-\sqrt{\frac{a_i}{a_j}}\right)$, because species j is overall more sensitive to competition than species i. To estimate the parameters in Equation 1, data on performance of species at very low densities (to estimate $r_i$ and $r_j$) and with different densities of conspecific competitors (to estimate $\alpha_i$ and $\alpha_j$) and heterospecific ones (to estimate $\alpha_{ij}$ and $\alpha_{ji}$) are required. The required data can be obtained by experimentally creating density gradients (Godoy et al., 2014) or using time-series data of population dynamics with sufficient variation in densities (Chu & Adler, 2015).

The RY is calculated as a species’ yield in mixture divided by its yield in monoculture, often using data from the design of replacement series where total density is held constant (sensu de Wit, 1960). This does not specify whether the data should be from study systems at equilibrium or not, and in fact RY is often calculated with ‘non-equilibrium’ data from short-term greenhouse experiments...
Species is supposed to outcompete species j if RY of species i is greater than that of species j (de Wit, 1960), but the theoretical basis for this is rather elusive. Interestingly, RY and related indices are also used to indicate niche differentiation between species, since they appear to reflect the relative strength of intraspecific vs. interspecific competition (by comparing monoculture vs. mixture treatments) (Jolliffe, 2000).

The RE estimates PCO by using Markov chain models to (a) randomly generate a large number (100,000) of competition matrices whose elements $C_{ij}$ specify the probability that species $i$ outcompetes species $j$, and $C_{ii} = 1 - C_{ij}$ (if $C_{ij} > C_{ji}$, species $i$ outcompetes species $j$), (b) transform the competition matrix to patch transition matrices from which species abundances can be predicted, and (c) select the competition matrix that best fits the observed abundances of species (for more details, see Ulrich et al., 2014). If the goodness-of-fit of the best fitting matrix ($r_i$) is low, it indicates that forces other than competition (e.g. niche-based processes) may also play important roles in determining species abundances. The RE can be applied to temporal, spatial, or temporal × spatial abundance data, and allows one to infer PCO for many species without doing a large number of pairwise competition experiments (e.g. Soliveres et al., 2015).

### 2.1.2 Multispecies competitive ranks and intransitivity

Using PCOs inferred by any of the methods above, multispecies competitive ranks and intransitivity can be constructed. The competitive ranks can simply be constructed by counting the number of wins each species has. To measure intransitivity, six indices have been proposed to capture different elements of topological variation in an intransitive network (Laird & Schamp, 2018) (see Appendix S1 for details); Slater’s and Petraitis’s $i$ and Ulrich’s $I$ (overall degree of intransitivity), Kendall and Babington Smith’s $d_i$ (proportion of three-species intransitive loops), Bezembinder’s $\delta$ (proportion of pairs in intransitive relationships), unbeatability ($u$; whether there is an ‘unbeatable’ species, 1, or not, 0), and always-beatability ($a$; whether there is an ‘always beatable’ species, 1, or not, 0).

### 2.2 Theoretical test

First, we used the analytical approach with the two-species Lotka–Volterra model (Equation 1) to derive the relations between RFD, RY and RE in indicating PCO using data at equilibrium and before equilibrium. To seek generality, we also derived the relations using other competition models (e.g. Beverton–Holt model, annual plant model; see Table S1 in Appendix S2). For the ‘before-equilibrium’ case where it turned out analytical solutions are not possible, however, we used simulations to explore the relations (see below). The details of these theoretical analyses are in Appendix S2.

Second, we used numerical simulations with the Lotka–Volterra model to explore how the relations changed as the systems move away from equilibrium (Appendix S2). In the simulations, we also had eight species and 28 two-species pairs as in the experimental test (see below). Model parameters were randomly drawn from the range of parameters in the Lotka–Volterra models fitted with the field experiment (see below), i.e. with intrinsic growth rates, intra- and inter-specific competition coefficients randomly drawn from 0.2–2, 0.03–0.1 and 0.001–0.05 respectively. We used the drawn competition coefficients to directly calculate RFD. For RY and RE, we first ran the simulations of monocultures and two-species mixtures at three continuous steps around year 1 (a before-equilibrium point) and around year 10 (an at-equilibrium point) to calculate RY. For example, for year 1 we used biomasses at steps 0.9, 1 and 1.1 years. Similarly, we applied RE to the time-series (i.e. the three steps) of species relative abundances in two-species mixtures (biomass of each species/ total biomass) also around year 1 and 10. Note, although the RE was initially developed for multispecies cases (Ulrich et al., 2014), we applied it to two-species cases to make it comparable with the other two methods. The criteria of three methods to indicate if species $i$ outcompetes species $j$ are as follows: $\sqrt{\frac{R_Y}{i}} > 1, R_Y > R_Y, C_{ij} > C_{ji}$, respectively. All the simulations were performed with the ‘deSolve’ package (Soetaert, Petzoldt, & Setzer, 2010) in R 3.4.3 (R Core Team, 2018).

Based on PCO inferred by each method, we then constructed multispecies competitive ranks for the eight species. For intransitivity, we first theoretically constructed 30 multispecies communities (species richness >2) as in the experimental test (see below), and then calculated six intransitivity indices for each community based on the PCO. We also counted the number of three-species intransitive loops among the eight species as another measure of intransitivity.

### 2.3 Experimental test

We used data from a long-term field experiment and a short-term greenhouse experiment to validate the findings in the theoretical test. Below, we briefly describe the two experiments (for details, see Appendix S3).

#### 2.3.1 Field experiment

The ‘dominance’ experiment is a long-term grassland biodiversity experiment conducted from 2002 to 2015 in Jena (Roscher et al., 2004). The experiment includes nine potentially dominant perennial species: two forbs (Geranium pratense, Anthriscus sylvestris), two legumes (Trifolium pratense, T. repens), and five grasses (Alopecurus pratensis, Arrhenatherum elatius, Daucylis glomerata, Phleum pratense...
Our analyses, however, excluded A. sylvestris due to its absence in the greenhouse experiment. Thus, we used data from 132 plots (1 × 1 m²), representing 66 different compositions of communities across five diversity levels (1, 2, 3, 4 and 6 species, with 8, 28, 16, 10 and 4 compositions, respectively), with each composition replicated twice. Species-level annual above-ground biomass was determined each year.

2.3.2 | Greenhouse experiment

A microcosm experiment with the eight species was conducted from April to September in 2011, and repeated in 2012, in a greenhouse in Zurich. An individual of a species either grew alone (mono1), with a conspecific individual (mono2), or with a heterospecific individual (mix), in pots (11 × 11 × 21 cm³). Each treatment was replicated twice, resulting in 88 pots in total. Species-level above-ground biomass was determined in June and September each year.

2.3.3 | PCO and multispecies competitive ranks and intransitivity

As in the theoretical test, we applied each method to the field and greenhouse experiments to calculate PCO and multispecies competitive ranks and intransitivity, and count the number of three-species intransitive loops.

RFD

For the field experiment, we first fitted exponential growth models with annual monoculture biomass data from 2002 to 2003, during which all species appeared to experience exponential growth, to estimate their intrinsic growth rates. We then fitted a joint two-species Lotka–Volterra model with time-series biomass data (2002–2015) of all two-species mixtures to simultaneously estimate all intra- and inter-specific competition coefficients. Similarly, for the greenhouse experiment, we estimated intrinsic growth rates with mono1 biomass data in June and September, and competition coefficients with mono2 and mix data in June and September. After the models for the two experiments were fitted, we used posteriors of competition coefficients to calculate posterior RFD and take their median as an estimate of RFD. To assess the statistical uncertainty of the fits, we calculated posterior RFD and take their median as an estimate of RFD. To assess the statistical uncertainty of the fits, we calculated posterior RFD and take their median as an estimate of RFD. To assess the statistical uncertainty of the fits, we calculated posterior RFD and take their median as an estimate of RFD. To assess the statistical uncertainty of the fits, we calculated posterior RFD and take their median as an estimate of RFD. To assess the statistical uncertainty of the fits, we calculated posterior RFD and take their median as an estimate of RFD. To assess the statistical uncertainty of the fits, we calculated posterior RFD and take their median as an estimate of RFD.

RY

For the field experiment, we calculated RY using biomass data from monocultures and two-species mixtures at each year (2002–2015), but in the main text, we only presented the results of 2015 where communities appeared to be closest to equilibrium (see Appendix S5 for the results of other years, and equilibrium evaluation below in Data Analysis). As the monoculture biomass of P. pratense in 2015 was zero, we used its monoculture data from 2014. For the greenhouse experiment, we calculated RY using mono1 and mix data by averaging the biomasses in June and September.

RE

For the field experiment, we applied RE with data of species relative abundance (biomass of each species divided by total biomass) of two ‘spatial’ replicates of each two-species mixture at each year (2002–2015), but similarly here we only presented the results of 2015 (see Appendix S6 for those of other years). For the greenhouse experiment, we applied RE with data of two replicates of each two-species mixture.

2.4 | Data analysis

For both theoretical and experimental tests, we tested how PCO (Phi correlations), multispecies competitive ranks (Wilcoxon signed-rank tests) or values of each intransitivity index (Phi or Pearson correlations depending whether the variable is binary or not) inferred by the three methods were related. Moreover, we checked whether three-species intransitive loops based on different methods differed (Appendix S7). We also used Pearson correlations to test how each intransitivity index is related to extinction rates (i.e. number of species lost by 2015 divided by total number of species; lower extinction rates indicate greater coexistence) of 30 multispecies communities in the field experiment (Appendix S8). To evaluate whether and when the equilibrium is achieved in the field experiment, we used pairwise T tests to test whether the biomasses of species (in monocultures or two-species mixtures) in the current year were significantly different from those in the previous year (Appendix S9). A significant difference should indicate the system is not at equilibrium. We also used standard deviations to evaluate the equilibrium but the results were similar (for the details, see Appendix S9). We illustrated all the correlations using the ‘ggpairs’ function in the GGally R package (Schloerke et al., 2018), and plotted competitive networks of eight species using iGraph R package (Csardi & Nepusz, 2006). All the analyses were performed in R 3.4.3 (R Core Team, 2018).

3 | RESULTS

3.1 | PCO

Our analytical and simulation tests demonstrated that, PCO inferred by RY or RE with data from study systems at equilibrium were mathematically equivalent or very similar respectively to those inferred by RFD, regardless of competition models used (Figure 1a; for details see Appendix S2). This is because $R_{i}^{−} - R_{i}^{−} > 0$ and mostly $C_{ij} − C_{ij} > 0$ if $f_{i}/f_{j} > 1$ (Appendix S2). However, the simulations showed that these relationships became weaker the further the system was from equilibrium, where $R_{i}^{−} - R_{i}^{−}$ and
both were affected not only by competition coefficients (as in the equilibrium case) but also by species’ intrinsic growth rates and previous biomass or density (Figure 1a and Appendix S2).

In line with the theoretical findings, our tests with the field experiment showed that PCO inferred by RE (using data from 2015) were strongly positively correlated with those inferred by RFD (Figure 1b), but for RY its correlation with RFD peaked at 2010 (Figure S6 in Appendix S5). Moreover, the correlations between PCO inferred by RY or RE and those by RFD became stronger over time (Figures S6 and S9 in Appendix S6), as communities became closer to equilibrium (Appendix S9). Note the equilibrium in monocultures and two-species mixtures was apparently achieved around 2005 and 2008 respectively (Appendix S9). The PCO inferred by the three methods were not correlated in the greenhouse experiment (Figure 1a), and overall the greenhouse PCO were only positively correlated with those in the field experiment of earlier (2003–2005) rather than later years (Figures S6 and S9). The reliability of the field RFD (mean = 0.90 and most >0.80) was on average much greater than that of the greenhouse RFD (mean = 0.79, with many lower than 0.75) (Figures S4 and S5 in Appendix S4).

3.2 Multispecies competitive ranks and intransitivity

Multispecies competitive ranks inferred by three methods showed equilibrium vs. non-equilibrium patterns similar to those of PCO (Table 2 and Figures 2,3; also see Table S4 in Appendix S5 and Table S6 in Appendix S6). In the simulations, we found positive correlations between the three methods with equilibrium data in inferring most of intransitivity indices (Figure 4) (also see Appendix S7), but the correlations were overall weaker than those for PCO with equilibrium data, and also tended to get much weaker when the system was far from equilibrium (Figure 4). The tests with the field experiment showed that there was no intransitivity inferred by RFD (Figure 5). The RY and RE detected intransitivity but their intransitivity values (and three-species intransitive loops) were poorly correlated and also changed substantially over time (Appendix S5 and S6). Moreover, the three methods overall inferred very different values of six intransitivity indices in the greenhouse experiment (Figure 5), and the values were also poorly correlated with the field ones (Figure 5; Figure S8 in Appendix S5 and Figure S11 in Appendix S6). In addition, all the intransitivity indices were poorly or counter-intuitively correlated with extinction rates (Appendix S8).

4 DISCUSSION

Among the three categories, only the coexistence theory-based methods explicitly separate the forces determining competitive outcomes from those causing niche differences, through the strict analysis of population dynamics (Chesson, 2000; Tilman, 1982). Therefore, they provide a theoretically robust measure of PCO (Hart...
TABLE 2  Results of Wilcoxon signed-rank tests of competitive ranks of eight species based on the three different methods (relative fitness difference: RFD, relative yield: RY, the reverse-engineering approach: RE), in (a) the simulation test and (b) the experimental test. In the simulation test, each method was applied to equilibrium (EQM) and before-equilibrium (bef-EQM) conditions. In the experimental test, each method was applied to the field experiment (field) and the greenhouse experiment (GH). For the field experiment, both RY and RE used data from 2015. Numbers are the p-values of Wilcoxon signed-rank tests.

| (a) | RFD (EQM) | RY (EQM) | RE (EQM) | RY (bef-EQM) | RE (bef-EQM) |
|-----|------------|-----------|-----------|--------------|--------------|
| RFD (EQM) | 1.000      | 0.850     | 0.430     | 0.730        |
| RY (EQM)  | 0.850      | 0.430     | 0.730     |              |
| RE (EQM)  | 0.345      | 0.526     | 1.000     |              |
| RY (bef-EQM) |              |           |           |              |
| RE (bef-EQM) |              |           |           |              |

| (b) | RFD (field) | RY (field) | RE (field) | RFD (GH) | RY (GH) | RE (GH) |
|-----|-------------|------------|------------|----------|---------|---------|
| RFD (field) | 0.281      | 0.168     | 0.393     | 0.292    | 0.149   |
| RY (field)  | 0.546      | 0.784     | 1.000     | 0.386    |         |
| RE (field)  | 0.525      | 0.671     | 0.230     |          |         |
| RFD (GH)   |            |           |           | 1.000    | 0.393   |
| RY (GH)    |            |           |           |          | 0.202   |
| RE (GH)    |            |           |           |          |         |

FIGURE 2  Competitive ranks and networks of eight simulated species based on the three different methods (relative fitness difference: RFD; relative yield: RY; the reverse-engineering approach: RE), in the simulation test. Each method was applied to (a–c) equilibrium (EQM) and (d–e) before-equilibrium (bef-EQM) conditions. In the networks, node sizes are proportional to species’ competitive ranks (i.e. nodes are larger if ranks are higher), and arrows point to the winner in competition. Different species are indicated with different colours.
et al., 2018), in comparison to the other two categories whose inferences are mainly based on species biomasses or abundances (state variables). Interestingly, however, our theoretical analyses of the three representative methods showed that, RY or RE applied to data at equilibrium, give equivalent or very similar estimates of PCO respectively to RFD. Note, this also implies RY is irrelevant to niche differences in such case, which questions whether RY could, as often claimed (Jolliffe, 2000), indicate niche differentiation between species. As different methods within each category share similarities, our findings could to some extent generalize beyond the three chosen representative methods. For example, as RFD can also be derived from a wide range of other competition models (e.g. resource competition model), our findings should therefore also apply when other coexistence theory-based measures (e.g. Tilman’s $R^*$) are used to infer PCO. Moreover, most of the other index-based methods also compare monoculture vs. mixture treatments as RY (e.g. relative mixture response, relative interaction index), while some compare grown-alone vs. mixture treatments (e.g. relative competition intensity). As the formulas of these indices deviate from that of RY, our theoretical analyses showed that with equilibrium data they could indicate PCO either same to (relative mixture response) or more frequently different from (relative interaction index, relative competition intensity) those inferred by RY and RFD (for details, see Appendix S2). Lastly, among the ‘process-from-pattern’ methods to tease apart the relative importance of competition vs. other forces (such as niche-based processes, stochasticity) in community assembly, the RE is the only one which can explicitly give competitive outcomes between species. However, our findings of the close link between RE and RFD may also help the other ‘process-from-pattern’ methods (e.g. C-scores, the null model approach) to build their basis of inferring community assembly on species coexistence theory.

In support of the theoretical results, our test with the field experiment showed that both RY and RE converged with RFD in indicating PCO over time as communities got closer to equilibrium. However, for RY, the convergence peaked in 2010 rather than 2015, which may be because of the deterioration of monoculture performance for some species after 2010 (Appendix S9), possibly due to the accumulation of pathogens over time (Cortois, Schröder-Georgi, Weigelt, Putten, & Deyn, 2016). In line with our findings, a former study found that RY using long-term, presumably equilibrual, data was positively correlated with Tilman’s $R^*$ (Fargione & Tilman, 2006). Moreover, some recent studies also
showed that RY at or close to equilibrium could reasonably measure the degree to which Lotka–Volterra models parameterized with monoculture and biculture data capture species abundances of polycultures (Fort, 2018; Halty, Valdés, Tejera, Picasso, & Fort, 2017). Our findings also suggest that the former applications of the RE with equilibrium data (e.g. Soliveres et al., 2015; Ulrich et al., 2016) should reliably indicate PCO. In short, when only the equilibrium data are available, both RY and RE could be a reasonably good proxy of RFD to infer PCO.

In our study, to estimate RFD, we fitted Lotka–Volterra models under a Bayesian framework. For all competition coefficients in a model, we had 10,000 posterior values to calculate 10,000 posterior RFD $\frac{R_{ij} - R_{ji}}{\text{SD}}$. If RFD between species $i$ and $j > 1$, species $i$ is more competitive than species $j$, and vice versa. Our 'reliability' analyses showed that the estimates of the field RFD were much more reliable than the greenhouse ones, because for the former on average most of their posterior RFD (90%) agreed with the estimates (either $>1$ or $<1$), a lot greater than that for the latter (79%) (Appendix S4). Moreover, compared to the greenhouse one, the posteriors of competition coefficients in the field experiment yielded much smaller SD relative to means (Appendix S4). These results are likely because the field experiment consists of long time-series data with substantial density variations. However, the data from the greenhouse experiment (and also from most of other greenhouse experiments) have very limited density gradients, and this may have led to less accurate estimation of model parameters.

In contrast to the equilibrium case, both RY and RE only poorly predicted PCOs in non-equilibrium systems. This is because the RY and RE calculated from non-equilibrium data are affected not only by competition coefficients (as in the equilibrium cases) but also by species’ growth rates, previous biomass or density (Appendix S2) (Freckleton & Watkinson, 2000). This suggests that the use of RY with short-term greenhouse data in many former studies (Table 1) probably fail to correctly indicate PCO. In line with this, our experimental test showed that the greenhouse PCO was only positively correlated with the field PCO from earlier rather than later years. In the non-equilibrium cases, therefore, PCO could only be reliably inferred from RFD by fitting competition models with density gradient or time-series data. However, if the required

FIGURE 4  Pearson correlations of each of intransitivity indices (a-f) (Slater’s and Petraitis’s $i$, Ulrich’s $I$, Kendall and Babington Smith’s $d$, Bezembinder’s $\delta$, unbeatability $u$ and always-beatability $a$) derived from the pairwise competitive outcomes based on each of the three different methods (relative fitness difference: RFD; relative yield: RY; the reverse-engineering approach: RE), in the simulation test. Each method was applied to equilibrium (EQM) and before-equilibrium (before-EQM) conditions. Asterisks indicate significance of correlations as *$p < .05$; **$p < .01$; ***$p < .001$. The lines in the lower diagonal are based on linear regressions, and the grey area indicates the 95% confidence interval. Data points are slightly jittered to improve readability.
Multispecies competitive ranks inferred by three methods showed similar equilibrium vs. non-equilibrium patterns as PCO. However, competitive relationships among multiple species are more subtle when their competitive ranks are not hierarchical, as intransitivity indices were much more sensitive to the methods used and the condition of study systems (at equilibrium or not). This is because intransitivity tended to occur mainly among similar competitors (‘weak intransitivity’) (see Figure S12 in Appendix S7; Soliveres & Allan, 2018). Therefore, even a small shift in competitive outcomes could reshuffle intransitive relationships. In theory, RY or RE with equilibrium data should indicate the same or similar intransitivity as RFD, but our experimental tests showed that the three methods detected very different values of intransitivity. For example, no intransitivity was detected at all with RFD in the field experiment. Both RY and RE detected intransitivity but their values of intransitivity (across all the six indices capturing different topologies of an intransitive network) were poorly correlated and also changed substantially over time (Appendix S5 and S6). Moreover, the detected intransitivity failed to explain species coexistence (Appendix S8), which may imply the estimates of intransitivity are not very reliable. These results suggest that the intransitivity inferred by RY or related indices using short-term (often greenhouse) data in former studies (e.g. Grace, Guntenspergen, & Keough, 1993; Keddy & Shipley, 1989; Shipley, 1993) are probably not reliable. Moreover, our findings also call for stricter evaluation when inferring intransitivity by RE with equilibrium data (e.g. Soliveres et al., 2015) and by RFD with density gradient data (e.g. Godoy, Stouffer, Kraft, & Levine, 2017). In short, all this makes it generally challenging to accurately infer intransitivity regardless of the methods, unless there are high quality data to allow a given method to accurately infer competitive outcomes among similar competitors.
4.1 | Choosing an appropriate method

Based on our findings, we recommend future studies to choose appropriate methods depending on questions of interest and data availability.

1. PCO. If data of two-species mixtures and monocultures from study systems at or close to equilibrium are available, one can choose RY or RE (or similar methods) to infer PCO, because in such cases both RY and RE could reasonably approximate RFD. The application of RE does not need monocultures which may often be unavailable or of low quality and therefore the RE can be more widely applied. Such ‘equilibrium’ data may be available in long-term experiments (e.g. long-term biodiversity experiments) or observations of natural communities. Equilibrium of study systems could be evaluated with methods such as T tests or SD by comparing temporal or spatial samples (Appendix S9). If ‘equilibrium’ data are not available (where RY or RE could not properly indicate PCO), however, one should apply RFD (or similar methods) to density gradient or time-series data with sufficient variation in population densities, to accurately estimate PCO.

2. Multispecies competitive ranks and intransitivity. Our recommendations for inferring multispecies competitive ranks are similar to those for PCO. However, for inferring intransitivity, which is also derived from PCO but very sensitive to even a small shift of PCO among similar competitors, our recommendations are stricter. In other words, to reliably infer intransitivity, one needs high quality data for a given method (e.g. unambiguous equilibrium data for RY and RE, and a large amount of density gradient or time-series data with sufficient density variations for RFD) to first accurately infer PCO especially among similar competitors.

ACKNOWLEDGEMENT

This study has been supported by the German Science Foundation (RO2397/8) in the framework of the Jena Experiment (FOR 456/1451). We thank Anne Ebeling for coordination of the field site and contribution of data, as well as the gardeners and numerous students for their help with plot maintenance and during harvest. Y.H.F. was also supported by the Fundamental Research Funds for the Central Universities (lzujbky-2019-32). S.S. was supported by the Spanish Government under a Ramón y Cajal contract (RYC-2016-20604).

AUTHORS’ CONTRIBUTIONS

Y.H.F. and C.R., S.S., E.A. and M.F. designed the research; C.W., A.T., N.E., B.S., A.W., W.W.W., C.R. and M.F. contributed the data of the field experiment, and E.D.L. and B.S. contributed the data of the greenhouse experiment; B.R. parametrized the Lotka–Volterra models with input from Y.H.F. Y.H.F. analysed the data; Y.H.F. drafted the manuscript, and all other authors contributed substantially to the improvement of the manuscript.

DATA AVAILABILITY STATEMENT

The data of the field and greenhouse experiments associated with this study are archived respectively in the Pangaea (https://doi.org/10.1594/PANGAEA.866358) and Dryad (De Luca & Schmid 2019).

ORCID

Yanhao Feng https://orcid.org/0000-0003-0460-4883
Benjamin Rosenbaum https://orcid.org/0000-0002-2815-0874
Cameron Wagg https://orcid.org/0000-0002-9738-6901
Bernhard Schmid https://orcid.org/0000-0002-8430-3214
Christiane Roscher https://orcid.org/0000-0001-9301-7909

REFERENCES

Aschehoug, E. T., Brooker, R., Atwater, D. Z., Maron, J. L., & Callaway, R. M. (2016). The mechanisms and consequences of interspecific competition among plants. Annual Review of Ecology, Evolution, and Systematics, 47, 263–281. https://doi.org/10.1146/annurev-ecolsys.121415-032123
Burns, K. (2007). Patterns in the assembly of an island plant community. Journal of Biogeography, 34, 760–768. https://doi.org/10.1111/j.1365-2699.2006.01625.x
Carroll, I. T., Cardinale, B. J., & Nisbet, R. M. (2011). Niche and fitness differences relate the maintenance of diversity to ecosystem function. Ecology, 92, 1157–1165. https://doi.org/10.1890/10-0302.1
Chesson, P. (2000). Mechanisms of maintenance of species diversity. Annual Review of Ecology and Systematics, 31, 343–366. https://doi.org/10.1146/annurev.ecolsys.31.1.343
Chesson, P. (2013). Species competition and predation. In R. Leemans (Ed.), Ecological systems: Selected entries from the encyclopedia of sustainability science and technology (pp. 223–256). New York, NY: Springer.
Chu, C., & Adler, P. B. (2015). Large niche differences emerge at the recruitment stage to stabilize grassland coexistence. Ecological Monographs, 85, 373–392. https://doi.org/10.1890/14-1741.1
Cortois, R., Schröder-Georgi, T., Weigtel, A., Putten, W. H., & De Deyn, G. B. (2016). Plant-soil feedbacks: Role of plant functional group and plant traits. Journal of Ecology, 104, 1608–1617. https://doi.org/10.1111/1365-2745.12643
Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. InterJournal, Complex Systems, 1695, 1–9.
De Luca, E., & Schmid, B. (2019). Data from: Inferring competitive outcomes. ranks and intransitivity from empirical data: A comparison of different methods. v3, Dryad Dataset. https://doi.org/10.5061/dryad.mgkq98vz
de Wit, C. T. (1960). On competition. Wageningen: Pudoc.
Diamond, J. (1975). Assembly of species communities. In M. L. Cody & J. Diamond (Eds.), Ecology and evolution of communities (pp. 342–444). Cambridge, MA: Harvard University Press.
Fargione, J., & Tilman, D. (2006). Plant species traits and capacity for resource reduction predict yield and abundance under competition in nitrogen-limited grassland. Functional Ecology, 20, 533–540. https://doi.org/10.1111/j.1365-2435.2006.01116.x
Freedman, H. (2018). On predicting species yields in multispecies communities: Quantifying the accuracy of the linear Lotka-Volterra generalized model. Ecological Modelling, 387, 154–162. https://doi.org/10.1016/j.ecolmodel.2018.09.009
Freckleton, R., & Watkinson, A. (2000). Designs for greenhouse studies of interactions between plants: An analytical perspective. Journal of Ecology, 88, 386–391. https://doi.org/10.1046/j.1365-2745.2000.00467.x
Gallien, L. (2017). Intransitive competition and its effects on community functional diversity. Oikos, 126, 615–623. https://doi.org/10.1111/oik.04033
Gause, G. F. (1934). The struggle for existence. Baltimore, MD: The Williams & Wilkins Company.
