Received: 25 April 2019 | Accepted: 23 May 2019

DOI: 10.1111/2041-210X.13238

RESEARCH ARTICLE

Methods in Ecology and Evolution BRITISH ECOLOGICAL SOCIETY

# Measuring individual identity information in animal signals: Overview and performance of available identity metrics

**Pavel Linhart**[1] | **Tomasz S. Osiejuk**[1] | **Michał Budka**[1] | **Martin Šálek**[2,3] | **Marek Špinka**[4,5] | **Richard Policht**[4,6] | **Michaela Syrová**[4,7] | **Daniel T. Blumstein**[8,9]

[1]Department of Behavioural Ecology, Adam Mickiewicz University, Poznań, Poland; [2]The Czech Academy of Sciences, Institute of Vertebrate Biology, Brno, Czech Republic; [3]Faculty of Environmental Sciences, Czech University of Life Sciences Prague, Prague, Czech Republic; [4]Department of Ethology, Institute of Animal Science, Prague, Czech Republic; [5]Department of Ethology and Companion Animal Science, Faculty of Agrobiology, Food and Natural Resources, Czech University of Life Sciences Prague, Prague, Czech Republic; [6]Department of Game Management and Wildlife Biology, Faculty of Forestry and Wood Sciences, Czech University of Life Sciences Prague, Prague, Czech Republic; [7]Department of Zoology, Faculty of Sciences, University of South Bohemia, České Budějovice, Czech Republic; [8]Department of Ecology and Evolutionary Biology, University of California, Los Angeles, California and [9]Rocky Mountain Biological Laboratory, Crested Butte, Colorado

**Correspondence**
Pavel Linhart
Email: pavel.linhart83@gmail.com

**Funding information**
Narodowe Centrum Nauki, Grant/Award Number: UMO-2015/19/P/NZ8/02507; Ministerstvo Zemědělství, Grant/Award Number: MZE-RO0718; H2020 Marie Skłodowska-Curie Actions, Grant/Award Number: 665778; Akademie Věd České Republiky, Grant/Award Number: RVO 68081766; National Science Foundation; Grantová agentura České republiky, Grant/Award Number: GA14-27925S

Handling Editor: Andres Lopez-Sepulcre

## Abstract

1. Identity signals have been studied for over 50 years but, and somewhat remarkably, there is no consensus as to how to quantify individuality in animal signals. While there is a variety of different metrics to quantify individuality, these methods remain un-validated and the relationships between them unclear.
2. We contrasted three univariate and four multivariate identity metrics (and their different computational variants) and evaluated their performance on simulated and empirical datasets.
3. Of the metrics examined, Beecher's information statistic ($H_S$) performed closest to theoretical expectations and requirements for an ideal identity metric. It could be also easily and reliably converted into the commonly used discrimination score (and vice versa). Although Beecher's information statistic is not entirely independent of study sampling, this problem can be considerably lessened by reducing the number of parameters or by increasing the number of individuals in the analysis.
4. Because it is easily calculated, has superior performance, can be used to quantify identity information in single variable or in a complete signal and because it indicates the number of individuals who can be discriminated given a set of measurements, we recommend that individuality should be quantified using Beecher's information statistic in future studies. Consistent use of Beecher's information statistic could enable meaningful comparisons and integration of results across different studies of individual identity signals.

KEYWORDS

acoustic discrimination, acoustic identification, Beecher's information statistic, discriminant analysis, identity signal, individual recognition, social behaviour, vocal individuality

## 1 | INTRODUCTION

The fact that individuals differ in consistent ways is a both a central attribute of life and one that underlies a number of theoretically important questions such as explaining cooperative behaviour or understanding the evolution of sociality (Bradbury & Vehrencamp, 1998; Crowley et al., 1996; Tibbetts, 2004). Such individuality can also be important in wildlife conservation as well when it is used to help census or monitor individuals based on individually distinctive traits (Blumstein et al., 2011; Terry & McGregor, 2002). And, because, animals may base their decisions on the identity of the individual with whom they interact or respond to (Godard, 1991; Wilkinson, 1984), there may be selection to both produce individually distinctive signals, and selection to discriminate among them (Tibbetts & Dale, 2007; Wiley, 2013).

Quantification of individual identity (individuality) requires the assessment of variation in one or more traits between at least two individuals. For identity signals to function properly, they should maximize the between-individual variation and minimize the within-individual variation (Beecher, 1982, 1989). A variety of identity metrics have proliferated because of recognized biases (e.g. it is more likely to find similar individuals in larger populations and, hence, it will be more difficult to discriminate individuals in large populations or studies involving more individuals). These biases make the comparison of results among studies unreliable (Beecher, 1989; Mathevon, Koralek, Weldele, Glickman, & Theunissen, 2010). Additionally, some existing metrics were considered unsuitable for a particular signal type (Searby & Jouventin, 2004). Nevertheless, new alternatives were not always thoroughly tested and were not shown to be superior to the metrics they attempted to replace. Furthermore, there are methodological problems that result from the calculation of particular identity metrics, and some studies have used different equations to calculate the same identity metric. Thus, somewhat remarkably given its importance, there is no consensus about how to properly measure identity. As a result, researchers have generally avoided quantitative comparisons between studies (Insley, Phillips, & Charrier, 2003). In a few cases, researchers tried to overcome problems with identity metrics in comparative analyses by using exactly the same methods across involved species (Beecher, Medvin, Stoddard, & Loesche, 1986; Lengagne, Lauga, & Jouventin, 1997; Pollard & Blumstein, 2011). Thus, hundreds of isolated studies have been published on individuality in animal signals but because they used different metrics there is limited prospect that we can benefit from the cumulative evidence of these studies. The lack of a commonly used identity metric is a major impediment towards understanding the evolution of identity signalling and indeed, the evolution of individuality.

Here, we review previously developed univariate (quantifying individuality within a single trait) and multivariate metrics (quantifying individuality across multiple traits) that have been used to quantify individual identity information in signals and we test their performance on simulated and empirical datasets. In particular, we examine the following metrics: *F*-value, Potential of individual coding PIC, Beecher's information statistic $H_S$, Information capacity $H_M$, and Mutual information MI. We further evaluate the different computational variants found in the literature in case of PIC and $H_S$ (see Table 1 and Supplement 1 for a detailed overview of metrics and their variants).

We compare the performance of metrics to hypothetical ideal identity information metric. The main principle of measuring individual identity in continuous traits is to quantify the ratio of between- and within-individual variation (Beecher, 1982, 1989; Robisson, Aubin, & Bremond, 1993; Searby & Jouventin, 2004). Thus, an ideal individual identity metric should be expressed on a ratio scale with a meaningful zero value, equivalent to the situation when there is no between-individual variation. Further, there is no expected upper limit for individuality. High between- to within-individual variation ratio indicates easy discrimination of individuals.

The datasets for the assessment of individual identity in different species vary in properties such as the number of individuals, the number of samples per individual, the number of variables measured (i.e. number of individualistic traits) and the covariance between the multiple variables measured. Hence, we further propose that an ideal identity metric should be robust or respond predictably to these dataset parameters to allow meaningful comparisons between studies. Therefore, an ideal identity metric: (a) should not be systematically biased by the sampling effort, that is there should be no systematic effects of number of individuals and number of observations per individual in a study on individuality estimate, and the sampling should ideally only impact on precision of individuality estimate; and (b) in the multivariate case, it should well capture the intrinsic multidimensionality of identity signals. In particular, it should rise with number of meaningful variables because each of the uncorrelated variables can encode another level of individual variation. In addition, it should also decrease with covariance between the variables because increasing covariance between the variables essentially decreases the number of independent variables. For our comparison, we gave the same weight to all criteria because these are very basic requirements and an ideal metric should fulfil all of them. In addition, we will list other potential pros and cons of each metric to provide a comprehensive evaluation of existing metrics.

We also wished to see if each of two commonly used metrics (Beecher's information statistic $H_S$ and discrimination score DS) could be converted to the other metric. We focused only on $H_S$ and DS metrics. DS has been used in the vast majority of past studies and DS has been found to correlate well with potentially unbiased $H_S$ in a previous study (Beecher, 1989). However, the previous study only tested the relationship between $H_S$ and DS on datasets with equal number of individuals and observations per individual, thus, ignoring the known biases associated with DS. Reliable conversion of DS into potentially unbiased $H_S$ could facilitate comparative analyses of results reported in past and future studies.

**TABLE 1** Overview of the identity metrics and their variants

| Metric | Metric variant and equation | Description | Reference | IDmeasurer function |
|---|---|---|---|---|
| Univariate Metrics | | | | |
| ANOVA *F*-value (*F*); | $F = \frac{MS_b}{MS_w}$ | *F* from one-way ANOVA where the individual is treated as independent variable and trait as dependent variable; $MS_b$ = between group mean squares; $MS_w$ = within group mean squares | e.g. Miller (1978) | calcF |
| Potential of individual coding (PIC); | $PIC_{beetweentot} = \frac{CV_{beetweentot}}{CV_w}$ | $CV_{betweentot}$ = between-individual coefficient of variation calculated from all data points; $CV_w$ = within-individual coefficient of variation | e.g. Robisson et al. (1993) | calcPICbetweentot |
| | $PIC_{beetweenmeans} = \frac{CV_{beetweenmeans}}{CV_w}$ | $CV_{betweenmeans}$ = between-individual coefficient of variation calculated with means from each individual; $CV_w$ = within-individual coefficient of variation | e.g. Lein (2008) | calcPICbetweenmeans |
| Beecher's information statistic ($H_S$) | $H_{Sntot} = \log_2\sqrt{\frac{F + n_{tot} - 1}{n_{tot}}}$ | *F* = ANOVA *F*-value; $n_{tot}$ = total sample size | Possible variant from Beecher (1989) | calcHSntot |
| | $H_{Sngroups} = \log_2\sqrt{\frac{F + n_{groups} - 1}{n_{groups}}}$ | *F* = ANOVA *F*-value; $n_{groups}$ = number of groups (individuals) | possible variant from Beecher (1989); e.g. Pollard, Blumstein, & Griffin (2010) | calcHSngroups |
| | $H_{Snpergroup} = \log_2\sqrt{\frac{F + n_{pergroup} - 1}{n_{pergroup}}}$ | *F* = ANOVA *F*-value; $n_{pergroup}$ = number of samples in each group (observations per individual) | possible variant from Beecher (1989) | calcHSnpergroup |
| | $H_{Svarcomp} = \log_2\frac{\sigma_T}{\sigma_W}$ | $\sigma_T$ = total variance in mixed model; $\sigma_W$ = residual variance associated with random factor in mixed model | Beecher (1989); Carter, et al. (2012) | calcHSvarcomp |
| Multivariate Metrics | | | | |
| Beecher's information statistic ($H_S$) | $H_{Snpergroup} = \log_2\sqrt{\frac{F + n_{pergroup} - 1}{n_{pergroup}}}$ | *F* = ANOVA *F*-value; $n_{pergroup}$ = number of samples in each group (observations per individual); original variables are subjected to PCA to get uncorrelated components and $H_S$ is calculated and summed over each independent component | Beecher (1989) | calcHSnpergroup |
| Information capacity ($H_M$) | $H_M = \log_2\sqrt{\frac{F_M + n - 1}{n}}$<br>$F_M = \frac{n-1}{g-1} * \frac{dist_t - g * dist_w}{dist_w}$ | $dist_t$ = sum of distances of all samples from their centroid; $dist_w$ = sum of distances of samples within individual to its centroid; *n* = number of observations; *g* = number of groups; | Searby and Jouventin (2004) | calcHM |
| Discrimination score (DS) | $DS = \frac{C}{N}$ | C = samples correctly classified by Discriminant analysis; *N* = total number of samples | e.g. Hafner et al. (1979) | calcDS |
| Mutual information (MI) | $MI = \sum_{i,j} \log_2 \frac{p(i,j)}{p(i) * p(j)}$ | *p*(*i*) = probability of predicted individual; *p*(*j*) = probability of actual individual; *p*(*i*,*j*) = probability of match between predicted and actual individual | Mathevon et al. (2010) | calcMI |

## 2 | MATERIALS AND METHODS

We used R for simulations and statistical analysis (R Core Team, 2012). Functions to calculate identity metrics, associated functions and datasets are available within an IDMEASURER package. This package is available on CRAN (https://cran.r-project.org/web/packages/IDmeasurer/index.html) and GitHub (https://github.com/pygmy83/IDmeasurer).

### 2.1 | Datasets

#### 2.1.1 | Simulated datasets

Datasets were constructed to mimic typical data on individuality (Figure 1). Parameters of datasets vary among studies. In particular, there are different numbers of individuals, observations (samples) per individual, variables and different covariances among variables. Effect of these parameters was simulated along with individuality within datasets. The level of individual identity in data was modified by changing the ratio of between- and within-individual variance in accordance with theoretical assumptions of individual identity signals and previous studies (Beecher, 1989; Searby & Jouventin, 2004). We developed R scripts involving "rnorm" and MASS package (Venables & Ripley, 2002) "mvrnorm" function to generate the datasets. These functions generate random values with a given standard deviation around pre-specified mean and, in "mvrnorm", with pre-specified covariance.

We constructed datasets with univariate and multivariate normal distributions with parameters covering a wide range of values, specifically, five values for individuality (id = 0.01, 1, 2.5, 5 and 10), five values for number of observations per individual (o = 4, 8, 12, 16 and 20), eight values for number of individuals (i = 5, 10, 15, 20, 25, 30, 35 and 40). Additionally, for multivariate datasets, five values for covariance among variables (cov = 0, 0.25, 0.5, 0.75 and 1) and five values for number of variables (p = 2, 4, 6, 8 and 10). Thus, 200 and 5,000 unique parameter combinations were possible in case of univariate and multivariate datasets respectively. Individuality (id) represents the ratio of standard deviations between and within individuals (id = $SD_{between}/SD_{within}$; $SD_{between}$ was calculated from means for each individual and $SD_{within}$ was set to be $SD_{between}$/id) (Figure 1). A single covariance (cov) value was used in the variance–covariance matrix to define covariances between all pairs of variables. For univariate datasets, we first generated individual means for a pre-defined number of individuals (normal distribution, "rnorm" function, $M$ = 1,000, $SD_{between}$ = 1) and then we generated a pre-defined number of random observations "o" around each individual mean (normal distribution, "rnorm" function, mean = individual mean, $SD_{within}$ = $SD_{between}$/individuality "id"). In the multivariate case, we first created a matrix representing mean individual values of variables for each of the individuals (multivariate normal distribution, "mvrnorm" function, mean for each variable = 0, variance-covariance matrix). Variances on the diagonal of the covariance matrix were set equal to 1 (hence $SD_{between}$ = 1) and all covariances between variable pairs were set equal to the pre-defined covariance "cov". Then, we generated a pre-defined number of random observations "o" around each individual and a variable mean ("rnorm" function, mean = individual mean, $SD_{within}$ = $SD_{between}$/individuality "id").

We asked how dataset parameters (i, o, p, cov, id) influenced the value of each identity metric. To explore this, 20 randomization cycles were run for each unique combination of parameter values. For example, in the multivariate case, 20 * 5,000 = 100,000 independent datasets were generated (datasets 1–20: i = 5, o = 4, p = 2, cov = 0, id = 0.01; datasets 21–40: i = 10, o = 4, p = 2, cov = 0, id = 0.01; …; datasets 99,981–100,000: i = 40, o = 20, p = 10, cov = 1, id = 10). Identity metrics were calculated for each dataset.

#### 2.1.2 | Empirical datasets

While the general performance of identity metrics was evaluated on simulated datasets, empirical datasets were used to evaluate the consistency of DS and $H_S$ metrics and reliability of $H_S$ and DS conversion on real data. We used six empirical datasets from four different species: little owls *Athene noctua* (ANmodulation, ANspec) (Linhart & Šálek, 2017), corncrake *Crex crex* (CCformants, CCspec) (Budka & Osiejuk, 2013), yellow-breasted boubous *Laniarius atroflavus*

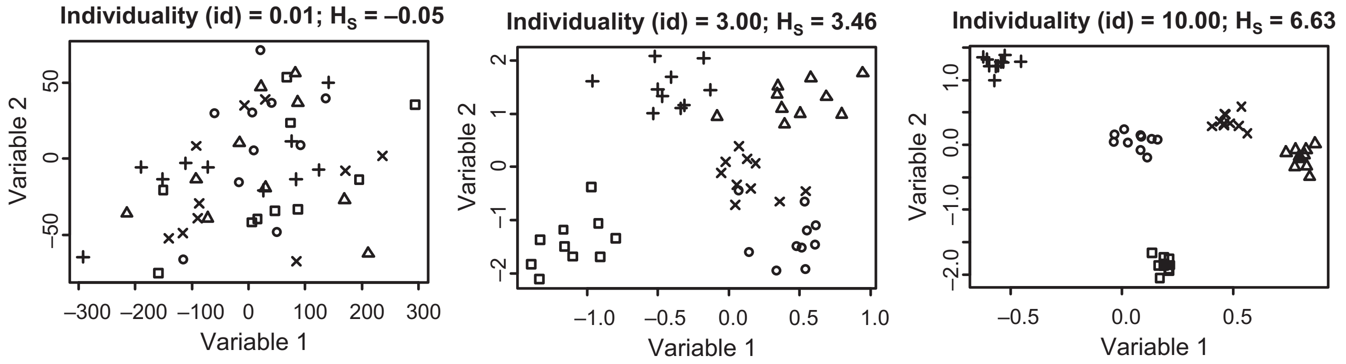

**FIGURE 1** Illustration of three artificial multivariate datasets that differ only in the individuality used to generate datasets. Settings for the function generating these datasets: i = 5, o = 10, p = 2, cov = 0, id = 0.01, 3 and 10

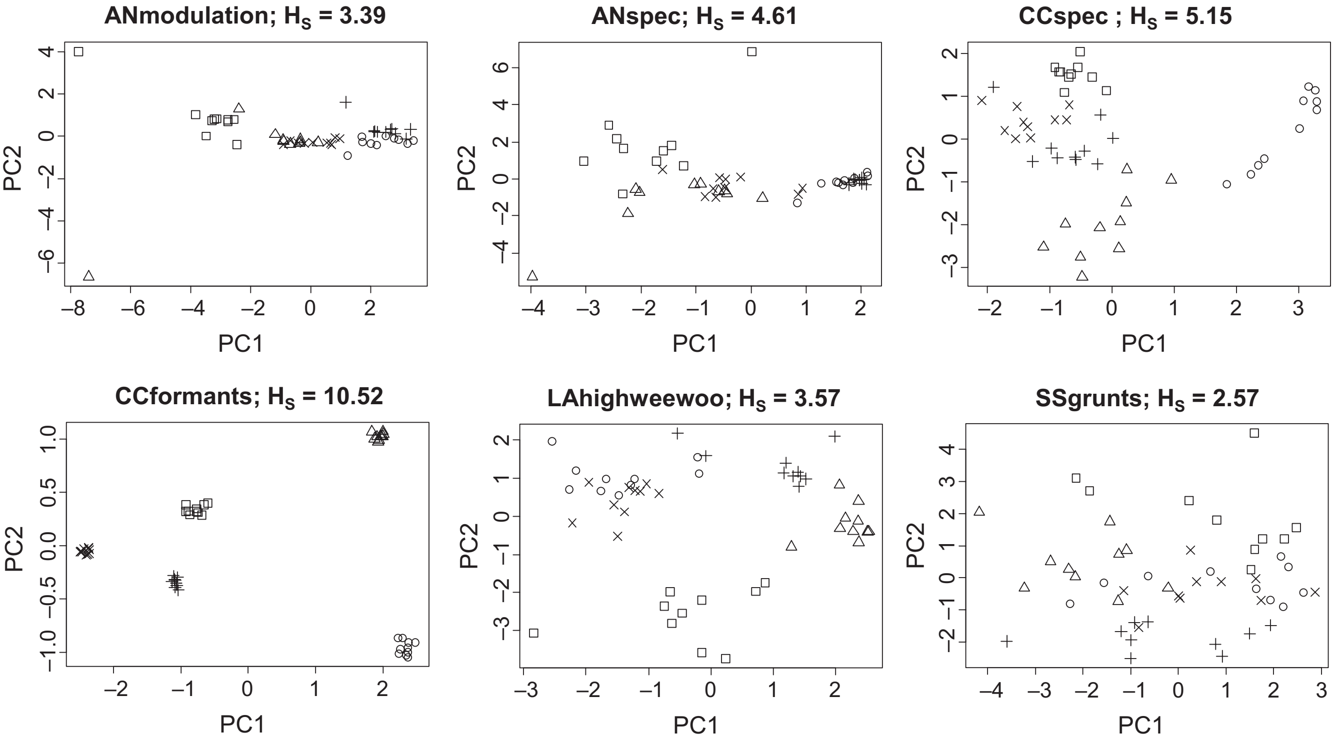

**FIGURE 2** Illustration of empirical datasets. Five individuals were randomly sampled from each dataset of 33 individuals and all 10 calls per individual were selected. $H_S$ for a full dataset is shown. Data were centred and scaled and subjected to PCA. The first two Principal Components are plotted

(LAhighweewoo) (Osiejuk et al. unpubl. data) and domestic pigs *Sus scrofa* (SSgrunts) (Syrová, Policht, Linhart, & Špinka, 2017) (Figure 2). In two species – corncrakes and little owls – calls were described by two different sets of variables. In little owls, we described calls by frequency modulation by measuring fundamental frequency at 10 measuring points evenly spread through the duration of the call (ANmodulation) or parameters describing the distribution of the frequency spectrum such as peak frequency, minimum and maximum frequencies and frequencies dividing spectrum by energy content (ANspec). In corncrakes, we used formants (CCformants) and parameters describing the distribution of the frequency spectrum (CCspec) (see the Supplement 2 for detail description of empirical datasets). Because datasets varied with respect to the number of individuals (33–100) and the number of calls per individual available (10–20), we scaled all datasets down to lowest common denominator by randomly selecting individuals and calls from bigger datasets. Eventually, each dataset had 33 individuals and 10 calls per individual. Each dataset also used different numbers of variables to describe the calls' acoustic structure (ANmodulation = 11, ANspec = 7, CCformants = 4, CCspec = 7, LAhighweewoo = 7, SSgrunts = 10). In all these empirical datasets, assumptions of multivariate normality were tested (Korkmaz, Goksuluk, & Zararsiz, 2014), but not met. We found various issues on the level of univariate variables and the whole dataset. For instance, there were issues with outliers, skewness, kurtosis and multimodal distributions (see Supplement 2 for univariate histograms and multivariate Chi-square Q-Q plots). Normality issues are common for research studies on acoustic individual identity. Authors deal with normality issues by eliminating problematic variables (e.g. Couchoux & Dabelsteen, 2015), using nonparametric classification methods (e.g. Mielke & Zuberbuehler, 2013), or by relying on robustness of cross-validated DFA and Principle Component Analysis (PCA) towards relaxed assumptions (e.g. Mathevon et al., 2010). We used the last approach. If the assumptions of discriminant analysis that are not met the results should be less stable when using different sampling and hence our results should be viewed as conservative.

## 2.2 | Statistical analysis

The relationship between a given identity metric and each of the parameters was assessed graphically by plotting the mean value and the 95% confidence intervals of an identity metric against all of the modelled data parameters separately. We then used a one-way ANOVA to test whether an identity metric was constant across all levels of a parameter. One-way ANOVA along with graphical evaluation of relationships between metrics and model parameters was preferred over multivariate regression because it simply, but adequately, addresses our main question (i.e. does the metric change in response to model parameter?) without the need to specify and compare many different multivariate regression models. If we found significant differences, we followed up these with post-hoc Tukey tests to identify which parameter levels differed. Due to the large number of comparisons, we only reported comparisons of neighbouring parameter levels. We used linear and nonparametric loess regression to convert $H_S$ to DS and vice versa.

Loess regression identifies a function that best describes complex data by fitting simple models to sequential subsets of data. Its main advantage is that it does not require specifications of the function and, hence, it is suitable for modelling of complex relationships. Loess regression included the number of individuals and the number of observations per individual as additional predictors. We used Spearman correlation coefficients to quantify between-metric consistency of ranking individuality in datasets. Pearson correlations were used to assess consistency within identity metrics in full and partial datasets. We then used Friedman tests, followed by a series of Wilcoxon tests (for post-hoc comparison of differences between levels), to compare correlation coefficients obtained for each pair of the metrics.

## 3 | RESULTS

The comparison of available univariate and multivariate metrics to an ideal metric is shown in Table 2.

### 3.1 | Univariate metrics

All explored univariate metrics increased with increasing individuality in the data. However, only $PIC_{betweentot}$, $PIC_{betweenmeans}$, $H_{Snpergroup}$ and $H_{Svarcomp}$ estimates were independent of the number of observations and the number of individuals used to calculate the metric (Figure 3). These general patterns were qualitatively identical when all simulated data were pooled or if only one of the parameters (number of observations, number of individuals, individuality) was changed at a time and the others were kept constant at the middle value (see Supplement 3 for detailed results including ANOVA tests).

All four sampling-independent metrics ($PIC_{betweentot}$, $PIC_{betweenmeans}$, $H_{Snpergroup}$ and $H_{Svarcomp}$) were highly correlated (Spearman correlation, all $r > 0.99$). $H_{Snpergroup}$ and $H_{Svarcomp}$ correctly converged to 0 in the case when individuality was set to be negligible (id = 0.01), while $PIC_{betweentot}$ and $PIC_{betweenmeans}$ converged to higher values (1.01 and 0.32 respectively). $PIC_{betweentot}$ reflects the number of potential individual signatures within a population in same way as $2^{H_S}$ does (Beecher, 1989), and, both, $PIC_{betweentot}$ and $2^{H_{Snpergroup}}$ reflect the ratio of between- to within-individual variation. Hence, convergence of $PIC_{betweentot}$ to 1 could be also seen as desirable quality and meaningful value for a signal with no individuality. $H_{Svarcomp}$ was equal to 2 * $H_{Snpergroup}$ (see Supplement 4 for details). We further considered only the $H_{Snpergroup}$ variant in multivariate analyses.

### 3.2 | Multivariate metrics

The performance of multivariate identity metrics is illustrated in Figure 4. All metrics increased with increasing individuality. DS, $H_S$ and MI increased with increasing number of variables available and decreased with increasing covariance between variables. Only $H_M$ did not change in response to increasing the number of individuals. $H_S$ and $H_M$ did not change in response to increasing the number of observations per individual. These general patterns were qualitatively identical when all simulated data were pooled or if only one dataset parameter was changed at a time and others were kept constant at the middle value (see Supplement 5 for detailed results including ANOVA tests).

Despite the different response of metrics to some of the simulated parameters, there was still moderate to high agreement among metrics about identity content in the data (Spearman correlations, mean $r \pm SD = 0.82 \pm 0.07$; minimum $r = 0.71$ for correlation between DS and MI; maximum $r = 0.95$ for correlation between DS and $H_S$). $H_S$ had the greatest correlations with other metrics (average $R = 0.88$). We found no advantage to using $H_M$ over $H_S$ as previously suggested. Instead, $H_M$ was equal to $H_S$ per variable ($H_M = H_S/p$) in data with zero covariance between variables. (Supplement 6).

Thus, our simulations show that $H_S$ matched the characteristics of the ideal metric in 6/7 cases, followed by $H_M$ (5/7), DS (4/7) and MI (both 3/7) (Table 1).

### 3.3 | Potential for removing bias in $H_S$

We observed no significant association between $H_S$ and the number of individuals in the univariate case, so we investigated the origin of the sampling bias in the multivariate case. This bias was only present when data were subjected to PCA. However, PCA is required to create uncorrelated components for $H_S$ calculation.

It is possible that the more variables measured, the more individuals need to be sampled in order to reduce this bias. We therefore fixed the number of variables to 5, 10 and 20 (p = 5, 10, 20) and varied the ratio of the number of individuals to the number of variables "i to p ratio" from 0.5 to 5 ("i to p ratio" = 0.5, 1, 1.5, 2, 3, 5) by using different numbers of individuals in our simulations (i = 3, 5, 8, 10, 15, 20, 25, 30, 40, 50, 60, 100 depending on number of variables and "i to p ratio"). The number of observations per individual was set to 10. Individuality and covariance were both chosen randomly in each iteration from pre-defined intervals used in the earlier simulations (covariance range = [0, 0.25, 0.5, 0.75, 1]; individuality range = [0.01, 1, 2.5, 5, 10]). We used 100 iterations for each "i to p ratio". $H_S$ did not rise significantly after the number of individuals reached at least the number of parameters (One-way ANOVA $F_{5,1794} = 7.68$, $p < 0.001$; no significant differences between levels if "i to p" ≥ 1, all $p > 0.132$) (Figure 5).

### 3.4 | Converting DS to $H_S$ and vice versa

We used simple linear regression and non-parametric loess regression to estimate $H_S$ based on DS and vice versa. There was a previously suggested linear relationship that had a limit of $H_S = 8$ where the DS values were 100% correct discrimination (Beecher, 1989). Because the $H_S$ values in our original simulated datasets far exceeded 8, we generated a new set of simulated datasets with individuality ranging between 0.1 and 2 (id = 0.1, 0.25, 0.5, 0.75, 1, 1.33,

**TABLE 2** The comparison of available univariate and multivariate metrics to a hypothetical ideal metric and summary of their pros and cons. We summed the number of matches (points) to compare different metrics to the ideal metric

| | Zero | Limit | id | cov | p | o | i | Points | Pros | Cons |
|---|---|---|---|---|---|---|---|---|---|---|
| Univariate metrics | | | | | | | | | | |
| Ideal | y | n | + | | | ns | ns | 5/5 | | |
| *F* | y | n | + | | | + | ns | 4/5 | | Sample dependent |
| $\mathrm{PIC}_{\mathrm{betweentot}}$ | y | n | + | | | ns | ns | 5/5 | Intuitive and straightforward calculation; allows separate assessment of within- and between-individual variation | Not meaningful for variables with positive and negative values; cannot be summed or averaged over different variables = univariate only |
| $\mathrm{PIC}_{\mathrm{betweenmeans}}$ | n | n | + | | | ns | ns | 4/5 | | Converges to non-meaningful value for no individuality in data |
| $\mathrm{H}_{\mathrm{Sntot}}$ | y | n | + | | | ns | – | 4/5 | | Sample dependent; incorrect HS variant |
| $\mathrm{H}_{\mathrm{Snpergroup}}$ | y | n | + | | | ns | ns | 5/5 | Standard variant of HS; univariate and multivariate | |
| $\mathrm{H}_{\mathrm{Sngroups}}$ | y | n | + | | | + | – | 3/5 | | Sample dependent; incorrect HS variant |
| $\mathrm{H}_{\mathrm{Svarcomp}}$ | y | n | + | | | ns | ns | 5/5 | Allows including various covariates in mixed models | Values twice as big as in case of standard HSnpergroup |
| Multivariate metrics | | | | | | | | | | |
| Ideal | y | n | + | – | + | ns | ns | 7/7 | | |
| DS | n | y | + | – | + | + | – | 3/7 | Population and individual metric; the most commonly used metric | Sample dependent; not suitable for high individuality signals because values are limited from the top |
| $\mathrm{H}_{\mathrm{S}}$ | y | n | + | – | + | ns | + | 6/7 | Univariate and multivariate; partial sample dependance is introduced by PCA but can be to large extent eliminated; biologically meaningful - provides number of unique individual signatures within population; good theoretical framework for both discrete and continuous individuality traits | Partially sample dependent |
| $\mathrm{H}_{\mathrm{M}}$ | y | n | + | ns | ns | ns | ns | 5/7 | Sample independent; various types of similarity metrics can be potentially used (euclidean distances, Jaccard similarity, string edit distance, dynamic time warping, etc.) | Number of independent variables needs to be known to calculate total identity information |
| MI | n | y | + | – | + | – | + | 3/7 | Could be applied with various classification methods | Sample dependent; not suitable for high individuality signals because values are limited from the top |

Abbreviations: zero, metric has a meaningful zero; limit, metric is limited from the top by an asymptote; id, change in response to increasing identity information in data; cov, response to increasing covariance between variables; p, response to increasing number of variables; o, response to increasing number of observations per individual; I, response to increasing number of individuals; y, yes; n, no; +, increase; −, decrease; ns, not significant, does not change with a parameter.

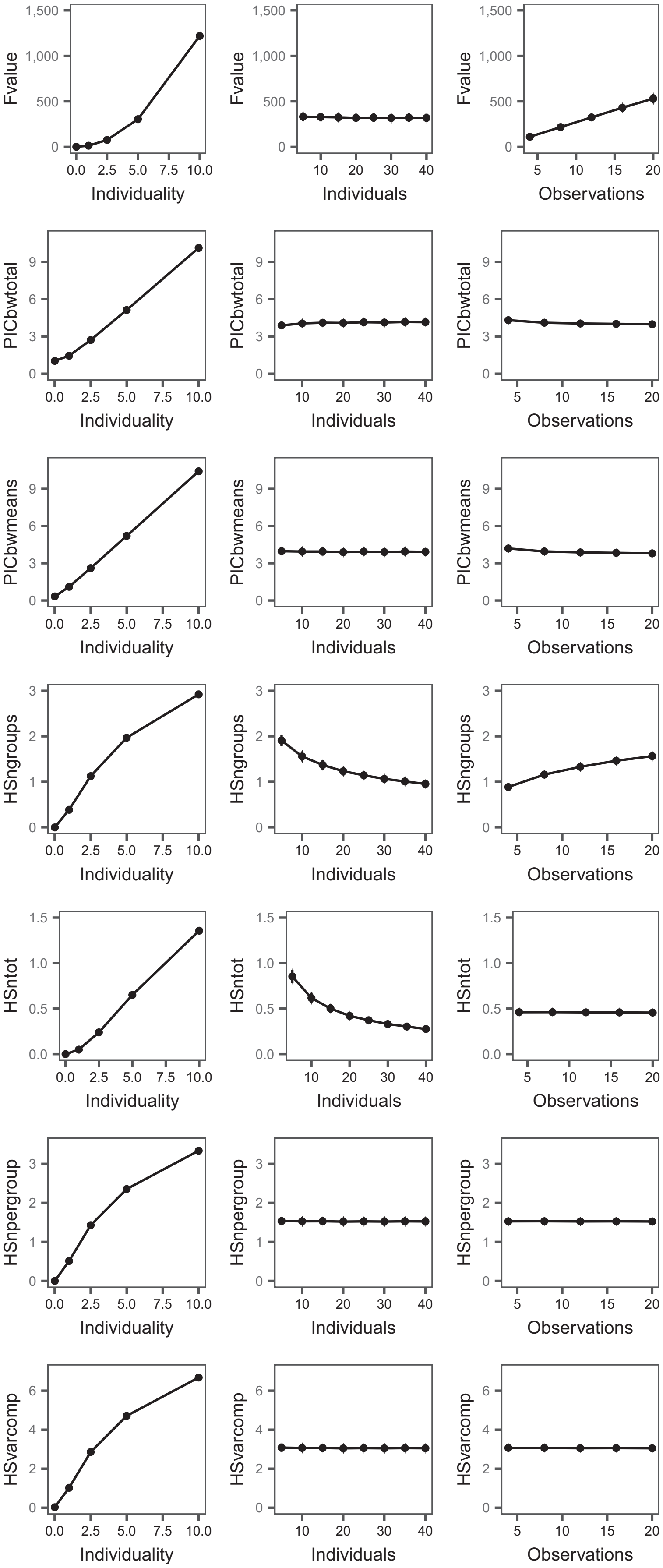

**FIGURE 3** Variation in univariate identity metrics in response to simulated dataset parameters: individuality, number of observations per individual and number of individuals. Means and 95% confidence intervals are shown. Graphs were plotted using all simulated univariate data pooled together. For the graphs with only a single parameter changing at a time see Supplement 3

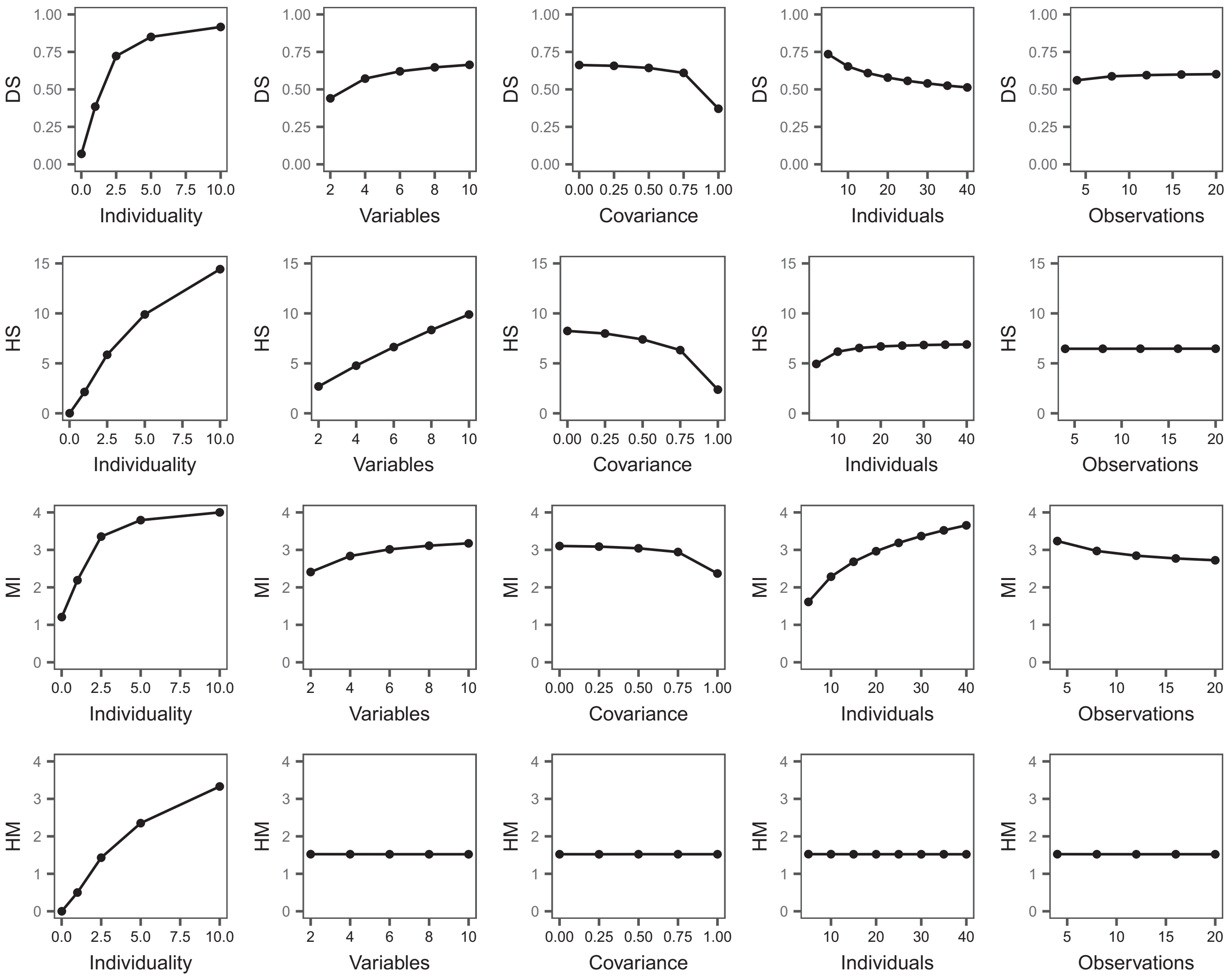

**FIGURE 4** Multivariate identity metrics in response to simulated dataset parameters: individuality, covariance between variables, number of variables, number of observations per individual and number of individuals. Means and 95% confidence intervals are shown. Graphs were plotted using all simulated multivariate data pooled together. For the graphs with only a single parameter changing at a time see Supplement 4

1.66, 2), covariance set to zero (cov = 0), number of iterations was reduced to 10 (it = 10), and other parameters were set as in previous models (p = 2, 4, 6, 8, 10; i = 5, 10, 15, 20, 25, 30, 35, 40; o = 4, 8, 12, 16, 20). These settings led to $H_S$ values up to about 13 for data used for model building, and $H_S$ values up to about 14 in the case of data used for model testing. These values are much closer to 8 and also much closer to $H_S$ values reported from nature.

Loess models took into account the number of observations per individual and the number of individuals. We compared the loess conversion and linear conversion models of DS and $H_S$. In general, loess estimates were closer to the ideal prediction (intercept = 0, beta = 1) and the loess model reduced the error of both DS and $H_S$ estimates to about a half compared to linear estimates (Figure 6). Both $H_S$ estimates were underestimated for high values of $H_S$. The ceiling value is clearly apparent for linear estimates of $H_S$. It is still visible in the case of loess estimates but loess predictions remain reasonably good up to about $H_S = 10$.

### 3.5 | Correlations between calculated and estimated metrics

We were further interested in how $H_{Sest}$ and $DS_{est}$ might represent $H_S$ and DS of a particular sample of individuals or $H_{Sfull}$ and $DS_{full}$ of the whole population. For this purpose, we first generated 50 full datasets with different identity levels representing 50 hypothetical populations of different species. Each dataset comprised 40 individuals, 20 calls per individual and 10 parameters. For these datasets, individuality was set randomly ranging between 0.2 and 2 (0.1 increments), and the covariance was set randomly ranging between 0.2 and 0.8 (0.1 increments). These settings generated datasets with $H_{Sfull}$ values that ranged from 0.22 to 9.89 ($M \pm SD$: 4.72 ± 2.95). Then, we repeatedly subsampled full datasets to get partial datasets which simulated different sampling of the population. We subsampled 5–40 individuals and 4–20 calls per individual per dataset in

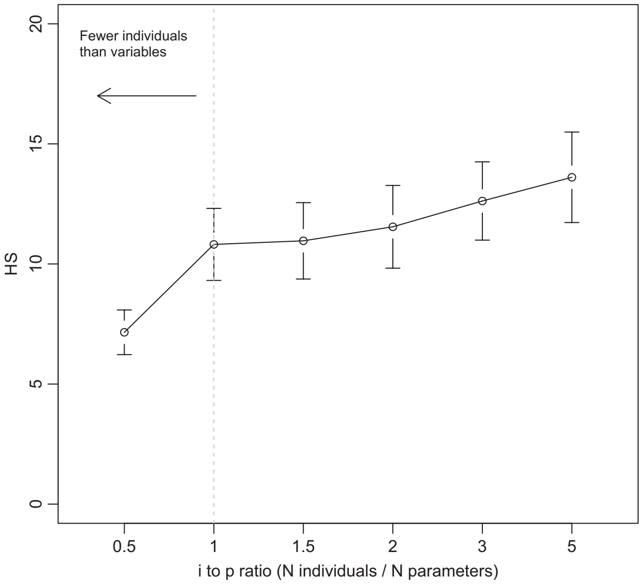

**FIGURE 5** $H_S$ and "i to p ratio" (number of individuals/ number of variables). $H_S$ was understimated if there were fewer individuals than variables. Means and 95% confidence intervals are shown

each of total 20 iterations. We also repeatedly subsampled our empirical datasets. We subsampled 5–33 individuals and 4–10 calls per individual per dataset in each of total 20 iterations. The number of parameters was not randomized – we always kept the original number of variables.

In simulated datasets, $H_S$ and $H_{Sest}$ were correlated almost perfectly with each other and with $H_{Sfull}$ (all average Pearson $r > 0.97$). There was no difference among the correlation coefficients from correlations between $H_{Sfull}$, $H_S$ and $H_{Sest}$ (Friedman Chi Square = 3.6, $p = 0.165$). In empirical datasets, $H_S$ calculated on partial datasets still reflected the $H_{Sfull}$ almost perfectly (average Pearson $r = 0.99$). While $H_{Sest}$ reflected $H_S$ of partial dataset (average Pearson $r = 0.90$) and $H_{Sfull}$ (average Pearson $r = 0.88$) slightly worse, it remained a reasonable fit. However, $H_{Sest}$ did not reflect $H_{Sfull}$ as precisely as it did $H_S$ (Friedman Chi Square = 33.6, $p < 0.001$, post-hoc test: $H_S$ - $H_{Sfull}$ vs. $H_{Sest}$ - $H_{Sfull}$, $p < 0.001$).

DS in simulated datasets was almost perfectly correlated with $DS_{est}$ (average Pearson $r = 0.99$). Although the relationship between DS in full datasets ($DS_{full}$) and DS and $DS_{est}$ was significantly worse (Friedman Chi Square = 40.0, $p < 0.001$; both post-hoc tests: $p < 0.005$), these associations remained strong ($DS_{full}$ and DS: average Pearson $r = 0.95$; $DS_{full}$ and $DS_{est}$: average Pearson $r = 0.96$). In empirical datasets, the correlation between DS and $DS_{est}$ was lower than in case of artificial datasets (average Pearson $r = 0.91$). DS and $DS_{est}$ of partial datasets had comparable correlations to $DS_{full}$ ($DS_{full}$ and DS: average Pearson $r = 0.88$; $DS_{full}$ and $DS_{est}$: average Pearson $r = 0.86$). Thus, the performance of DS and $DS_{est}$ to reflect each other or $DS_{full}$ did not differ (Friedman Chi Square = 0.9, $p = 0.638$).

## 4 | DISCUSSION

We provided an overview of the metrics used to quantify individual identity in animal signals in order to identify the best method for reporting individuality in animal signals. Biases associated with some of the commonly used metrics, and the use of different metrics across studies, make it difficult to compare results and integrate the accumulated knowledge from the numerous published studies on individual identity in animal signals.

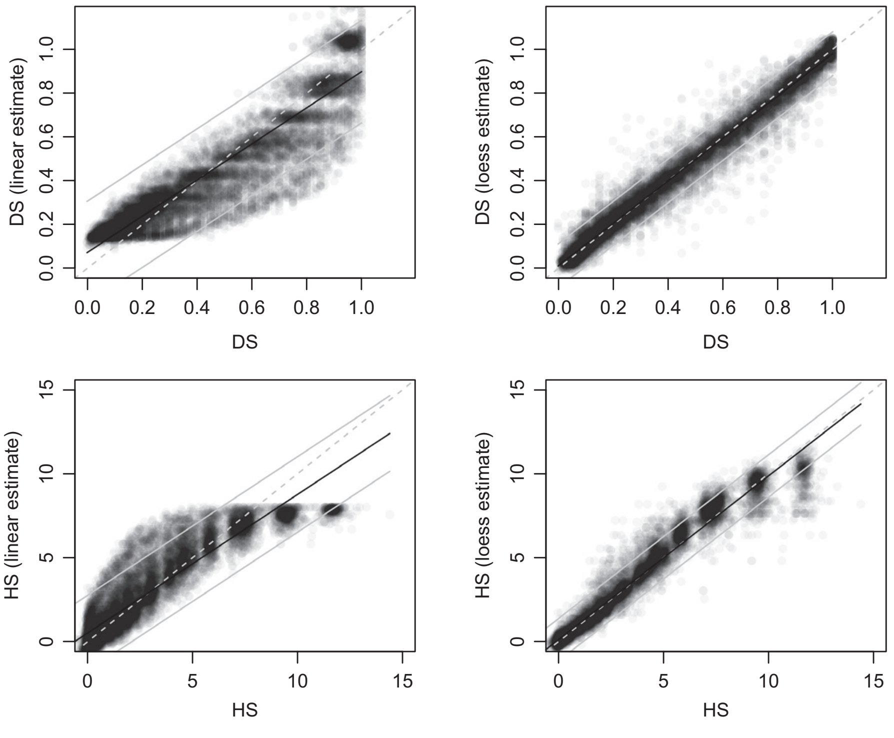

**FIGURE 6** Estimation of $H_S$ and DS based on linear and loess transformation of DS and $H_S$ respectively for datasets with $H_S$ up to 14.4. Linear DS estimation: Intercept = 0.07, Beta = 0.83, $R^2 = 0.83$, Standard Error of Estimate (SEE) = 0.12, 95% Prediction interval = predicted value ± 0.23; DS loess estimation: Intercept = 0.01, Beta = 0.98, $R^2 = 0.97$, Standard Error of Estimate (SEE) = 0.05, 95% Prediction interval = predicted value ± 0.10. Linear $H_S$ estimation: Intercept = 0.51, Beta = 0.83, $R^2 = 0.83$, Standard Error of Estimate (SEE) = 1.14, 95% Prediction interval = predicted value ± 2.24; HS loess estimation: Intercept = 0.11, Beta = 0.98, $R^2 = 0.95$, Standard Error of Estimate (SEE) = 0.64, 95% Prediction interval = predicted value ± 1.26

We show that the assessment of individual identity is relatively straightforward when considering a single trait (univariate case). Both, PIC ($PIC_{betweentot}$) and $H_S$ ($H_{Snpergroup}$) performed according to expectations. Multivariate identity metrics based on direct quantification of between- to within-individual variation ratios ($H_S$, $H_M$) performed better than the metrics derived from discrimination of individuals (DS, MI). We confirmed sampling-associated biases where they were reported previously (DS), but we found them even in metrics that had been developed to overcome these biases ($H_S$, MI). We also described yet unrecognized issues (the need to assess dimensionality for $H_M$ to quantify the total individuality of a signal). We further found that some metrics created values that were so close that they could be viewed as redundant (PIC and $H_S$; $H_M$ and $H_S$) and using them simultaneously brings unnecessary confusion to the field.

Based on our review and systematic analysis, we suggest $H_S$ should be routinely reported as the standard individual identity metric because it performed closest to an ideal identity metric in the univariate, as well as in the multivariate case. The partial bias in $H_S$ caused by the number of individuals in a study could be removed by having at least the same number of individuals as the number of variables. $H_S$ was the most consistent metric and correlated the best with DS and other identity metrics. Further, $H_S$ could be converted reliably into DS if needed.

The robustness of $H_M$ towards sampling bias (number of individuals, number of observations, as well as the number of variables and covariance) is an attractive feature. However, as we show, $H_M$ quantifies identity information per variable and not the identity information of the entire signal. It is necessary to know the effective number of variables to calculate the total identity information of a signal (i.e. if there is perfect covariance between the variables, the effective number of variables is 1 no matter how many variables are used), which may be difficult to assess. On the other hand, $H_M$ uses distances (similarity scores) of samples to calculate individuality and, hence, it could be potentially used not just with Euclidean distances (Searby & Jouventin, 2004, this study) but also together with other various methods assessing similarity (e.g. cross-correlation, dynamic time warping or string edit distances).

Mutual information (MI) is derived from a confusion matrix of discrimination analysis and we show it has similar shortcomings as discrimination scores. Our results that found systematic biases in MI are in line with previous studies that investigated measures of clustering for various machine learning purposes where potentially unbiased variants of MI are constantly searched for (e.g. Amelio & Pizzuti, 2017).

## 4.1 | Identity metrics in comparative analyses

We show that biases associated with DS (the most often used metric) and $H_S$ (the best metric) are not necessarily fatal for comparisons of different published studies because $H_S$ and DS values that are based on an entire population or subsamples from a population were well correlated in both simulated and empirical datasets. Additionally, the conversion of sample-biased DS values into less-biased $H_S$ values could allow better comparisons between studies. Both $H_S$ and $H_M$ values were previously found to correlate well with DS (Beecher, 1989; Searby & Jouventin, 2004). We extend previous findings for $H_S$ (Beecher, 1989) to situations with unequal sampling and we show it is possible to convert between $H_S$ and DS with an acceptable amount of error even when datasets differ in the number of individuals and observations per individual, and have important issues associated with multivariate normality (Supplement 2). Discriminant analysis (DA) and Principal component analysis (PCA) used for DS and $H_S$ calculations both assume multivariate normality for optimal results. While using these methods with non-normal data cannot be, in general, recommended, relatively high correlations between our metrics in empirical datasets suggest that DA and PCA scores were quite robust to these normality issues. Discrimination and dimensionality reduction analytical techniques that are able to handle normal and non-normal data definitely need to be considered in future individual identity studies.

## 4.2 | Future individual identity metrics

We hope that our study will stimulate further discussions about how individual identity should be properly measured. Although we suggest that $H_S$ should be generally used to quantify individuality, different metrics or more complex approaches might be required for particular interesting questions. For example, $H_S$ can only provide a population estimate of individual identity. Researchers might be interested in whether distinctiveness of individuals increases during ontogeny (Syrová et al., 2017). In this case, discrimination scores can be reported for each individual, thus making statistical evaluation possible. Furthermore, separate assessments of within- and between-individual variations when calculating PIC might be useful to test hypotheses about which of the two has been selected for. Within-individual variation could be reduced by, for example, ritualized behaviour while between-individual variation could be increased through, for example, morphological variation in structures producing or carrying the signal (e.g. Sheehan & Nachman, 2014). The dimensionality of identity signals might be an important factor for recognition processes (Trunk, 1979) and evolution could favour low dimensional signals. Paralleling the distribution of individuals in space (territoriality, living in colonies), individual signatures within a population, too, could have random, clumped, or regular distributions depending on the mechanisms behind individual distinctiveness and the degree of plasticity of identity signals.

We evaluated the efficacy of all metrics within the acoustic modality only. It is increasingly recognized that signals employ multiple modalities (Partan, 2013; Partan & Marler, 1999). All of the identity metrics discussed here could be, in principal, used in visual or chemical domains as well. $H_S$ has an advantage that it could be used both for discrete traits, such as colour variants, presence of particular alleles or chemicals, and for continuous traits such as size of visual patterns, duration of calls, etc. (Beecher, 1982, 1989). However, identity information outside the acoustic domain has been rarely quantified and meaningful comparison of individual identity across modalities remains a challenge for the future.

It is likely that automatic data collection and analysis techniques will be increasingly applied for various recognition tasks, including individual recognition (Elie & Theunissen, 2018; Stowell, Petrusková, Šálek, & Linhart, 2019). While these methods will allow studying individual identity signalling on unprecedented scales and sample sizes, the resulting classification accuracy scores will be analogous to the discrimination score, with similar positives and drawbacks. However, many different feature sets, pre-defined or automatically derived from data, as well as many different classification methods could be combined to test for the robustness of identity signals and/or to mimic and test for different alternatives of possible real recognition processes (Elie & Theunissen, 2018).

## 5 | CONCLUSION

We suggest that, at the current state of knowledge and methodology development, $H_S$ should be generally reported as the "golden standard" individual identity metric to allow the best comparison of individuality in signals across different studies. Given that $H_S$ may not be sufficient in all cases, we encourage further research to develop new metrics to quantify identity information in signals. However, new metrics should always be appropriately assessed and their performance directly compared to the best existing metrics. We provide datasets and scripts that should help to assess individual identity information in animal signals and benchmark the future metrics.

## ACKNOWLEDGEMENTS

P.L. received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 665778 administered by the National Science Centre, Poland (UMO-2015/19/P/NZ8/02507). D.T.B. is supported by the NSF. M.Šp., M.S. and R.P. were supported by the Czech Science Foundation (GA14-27925S) and the Czech Ministry of Agriculture (MZE-RO0718). M.Šá. work was supported by the research aim of the Czech Academy of Sciences (RVO 68081766).

## AUTHORS' CONTRIBUTIONS

P.L. and D.T.B. conceived the ideas, designed methodology and led the writing of the manuscript; P.L., T.O., M.B., M.Šá., M.Šp., M.S. and R.P. collected the data; P.L. analysed the data; All the authors contributed critically to the drafts and gave final approval for publication.

## DATA ACCESSIBILITY STATEMENT

Data and code used for this article are available within IDMEASURER R package currently available on CRAN (https://cran.r-project.org/web/packages/IDmeasurer/index.html) and GitHub (https://github.com/pygmy83/IDmeasurer).

## ORCID

*Pavel Linhart* https://orcid.org/0000-0002-6097-5892
*Tomasz S. Osiejuk* https://orcid.org/0000-0001-5980-7421
*Michał Budka* https://orcid.org/0000-0002-1244-013X
*Martin Šálek* https://orcid.org/0000-0003-2639-9745
*Marek Špinka* https://orcid.org/0000-0002-2506-9552
*Daniel T. Blumstein* https://orcid.org/0000-0001-5793-9244

## REFERENCES

Amelio, A., & Pizzuti, C. (2017). Correction for closeness: Adjusting normalized mutual information measure for clustering comparison. *Computational Intelligence*, *33*(3), 579–601. https://doi.org/10.1111/coin.12100

Beecher, M. D. (1982). Signature systems and kin recognition. *American Zoologist*, *22*(3), 477–490. https://doi.org/10.1093/icb/22.3.477

Beecher, M. D. (1989). Signaling systems for individual recognition - An information-theory approach. *Animal Behaviour*, *38*, 248–261. https://doi.org/10.1016/S0003-3472(89)80087-9

Beecher, M. D., Medvin, M. B., Stoddard, P. K., & Loesche, P. (1986). Acoustic adaptations for parent-offspring recognition in swallows. *Experimental Biology*, *45*, 179–193.

Blumstein, D. T., Mennill, D. J., Clemins, P., Girod, L., Yao, K., Patricelli, G., … Kirschel, A. N. G. (2011). Acoustic monitoring in terrestrial environments using microphone arrays: Applications, technological considerations and prospectus. *Journal of Applied Ecology*, *48*(3), 758–767. https://doi.org/10.1111/j.1365-2664.2011.01993.x

Bradbury, J. W., & Vehrencamp, S. L. (1998). *Principles of animal communication* (1st ed.). Sunderland, MA: Sinauer Associates.

Budka, M., & Osiejuk, T. S. (2013). Formant frequencies are acoustic cues to caller discrimination and are a weak indicator of the body size of corncrake males. *Ethology*, *119*(11), 960–969. https://doi.org/10.1111/eth.12141

Carter, G. G., Logsdon, R., Arnold, B. D., Menchaca, A., Medellin, R. A., & Brigham, R. M. (2012). Adult vampire bats produce contact calls when isolated: acoustic variation by species, population, colony, and individual. *PLoS ONE*, *27*(6), e38791.

Couchoux, C., & Dabelsteen, T. (2015). Acoustic cues to individual identity in the rattle calls of common blackbirds: A potential for individual recognition through multi-syllabic vocalisations emitted in both territorial and alarm contexts. *Behaviour*, *152*(1), 57–82. https://doi.org/10.1163/1568539X-00003232

Crowley, P. H., Provencher, L., Sloane, S., Dugatkin, L. A., Spohn, B., Rogers, L., & Alfieri, M. (1996). Evolving cooperation: The role of individual recognition. *Biosystems*, *37*(1), 49–66. https://doi.org/10.1016/0303-2647(95)01546-9

Elie, J. E., & Theunissen, F. E. (2018). Zebra finches identify individuals using vocal signatures unique to each call type. *Nature Communications*, *9*(1), 4026. https://doi.org/10.1038/s41467-018-06394-9

Godard, R. (1991). Long-term memory of individual neighbors in a migratory songbird. *Nature*, *350*(6315), 228–229.

Hafner, G. W., Hamilton, C. L., Steiner, W. W., Thompson, T. J., & Winn, H. E. (1979). Signature information in the song of the humpback whale. *The Journal of the Acoustical Society of America*, *66*(1), 1–6.

Insley, S. J., Phillips, A., & Charrier, I. (2003). A review of social recognition in pinnipeds. *Aquatic Mammals*, *29*, 181–201. https://doi.org/10.1578/016754203101024149

Korkmaz, S., Goksuluk, D., & Zararsiz, G. (2014). MVN: An R package for assessing multivariate normality. *The R Journal*, *6*(2), 151–162. https://doi.org/10.32614/RJ-2014-031

Lein, M. R. (2008). song variation in buff-breasted flycatchers (Empidonax fulvifrons). *The Wilson Journal of Ornithology*, *120*(2), 256–267.

Lengagne, T., Lauga, J., & Jouventin, P. (1997). A method of independent time and frequency decomposition of bioacoustic signals: Inter-individual recognition in four species of penguins. *Comptes Rendus De L Academie Des Sciences Serie Iii-Sciences De La Vie-Life Sciences*, *320*, 885–891. https://doi.org/10.1016/s0764-4469(97)80873-6

Linhart, P., & Šálek, M. (2017). The assessment of biases in the acoustic discrimination of individuals. *PLoS ONE*, *12*(5), e0177206. https://doi.org/10.1371/journal.pone.0177206

Mathevon, N., Koralek, A., Weldele, M., Glickman, S. E., & Theunissen, F. E. (2010). What the hyena's laugh tells: Sex, age, dominance and individual signature in the giggling call of *Crocuta crocuta*. *BMC Ecology*, *10*(1), 9–10. https://doi.org/10.1186/1472-6785-10-9.

Mielke, A., & Zuberbuehler, K. (2013). A method for automated individual, species and call type recognition in free-ranging animals. *Animal Behaviour*, *86*(2), 475–482. https://doi.org/10.1016/j.anbehav.2013.04.017

Miller, D. B. (1978). Species-typical and Individually distinctive acoustic features of crow calls of red jungle fowl. *Zeitschrift für Tierpsychologie*, *47*(2), 182–193.

Partan, S. R. (2013). Ten unanswered questions in multimodal communication. *Behavioral Ecology and Sociobiology*, *67*(9), 1523–1539. https://doi.org/10.1007/s00265-013-1565-y

Partan, S., & Marler, P. (1999). Communication goes multimodal. *Science*, *283*(5406), 1272–1273. https://doi.org/10.1126/science.283.5406.1272

Pollard, K. A., Blumstein, D. T., & Griffin, S. C. (2010). Pre-screening acoustic and other natural signatures for use in noninvasive individual identification. *Journal of Applied Ecology*, *47*(5), 1103–1109.

Pollard, K. A., & Blumstein, D. T. (2011). Social group size predicts the evolution of individuality. *Current Biology*, *21*(5), 413–417. https://doi.org/10.1016/j.cub.2011.01.051

R Core Team. (2012). *R: A Language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Robisson, P., Aubin, T., & Bremond, J. (1993). Individuality in the voice of the emperor penguin *Aptenodytes-Forsteri* - Adaptation to a noisy environment. *Ethology*, *94*(4), 279–290. https://doi.org/10.1111/j.1439-0310.1993.tb00445.x

Searby, A., & Jouventin, P. (2004). How to measure information carried by a modulated vocal signature? *Journal of the Acoustical Society of America*, *116*, 3192–3198. https://doi.org/10.1121/1.1775271

Sheehan, M. J., & Nachman, M. W. (2014). Morphological and population genomic evidence that human faces have evolved to signal individual identity. *Nature Communications*, *5*, 4800. https://doi.org/10.1038/ncomms5800

Stowell, D., Petrusková, T., Šálek, M., & Linhart, P. (2019). Automatic acoustic identification of individuals in multiple species: Improving identification across recording conditions. *Journal of the Royal Society Interface*, *16*(153), 20180940. https://doi.org/10.1098/rsif.2018.0940

Syrová, M., Policht, R., Linhart, P., & Špinka, M. (2017). Ontogeny of individual and litter identity signaling in grunts of piglets. *The Journal of the Acoustical Society of America*, *142*(5), 3116–3121. https://doi.org/10.1121/1.5010330

Terry, A. M. R., & McGregor, P. K. (2002). Census and monitoring based on individually identifiable vocalizations: The role of neural networks. *Animal Conservation*, *5*, 103–111. https://doi.org/10.1017/s1367943002002147

Tibbetts, E. A. (2004). Complex social behaviour can select for variability in visual features: A case study in *Polistes wasps*. *Proceedings of the Royal Society of London B: Biological Sciences*, *271*(1551), 1955–1960. https://doi.org/10.1098/rspb.2004.2784

Tibbetts, E. A., & Dale, J. (2007). Individual recognition: It is good to be different. *Trends in Ecology & Evolution*, *22*(10), 529–537. https://doi.org/10.1016/j.tree.2007.09.001

Trunk, G. V. (1979). A problem of dimensionality: A simple example. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *1*(3), 306–307. https://doi.org/10.1109/TPAMI.1979.4766926

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S (Fourth)*. New York: Springer. Retrieved from http://www.stats.ox.ac.uk/pub/MASS4

Wiley, R. H. (2013). Specificity and multiplicity in the recognition of individuals: Implications for the evolution of social behaviour. *Biological Reviews*, *88*(1), 179–195. https://doi.org/10.1111/j.1469-185X.2012.00246.x

Wilkinson, G. S. (1984). Reciprocal food sharing in the vampire bat. *Nature*, *308*(5955), 181–184. https://doi.org/10.1038/308181a0

## SUPPORTING INFORMATION

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

**How to cite this article:** Linhart P, Osiejuk T, Budka M, et al. Measuring individual identity information in animal signals: Overview and performance of available identity metrics. *Methods Ecol Evol*. 2019;10:1558–1570. https://doi.org/10.1111/2041-210X.13238