Marine Biodiversity Records

Converting SACFOR data for statistical analysis: validation, demonstration and further possibilities

James Asa Strong[1*] and Magnus Johnson[2]

Abstract

Background: the context and purpose of the study: Semi-quantitative scales are often used for the rapid assessment of species composition and abundance during time-limited surveys. The semi-quantitative SACFOR abundance scale was developed to support the observation of marine habitats, communities and species and is widely used in the UK. As such, there is now a vast accumulation of SACFOR data. However, there several acknowledged limitations associated with its format that prevent re-analysis.

Methods: how the study was performed and statistical tests used: A conversion process is proposed here that allows: (i) the merging of taxa within counts or cover data sub-sets; (ii) observations, based on either counts and cover, to be unified into one matrix; (iii) counts and cover data to have an equal weighting in the final matrix; and (iv) the removal of the influence of body size and growth form from the final values. To achieve this, it is only possible to preserve the ordinal structure of the data set.

Results: the main findings: Simulations verified that the SACFOR conversion process (i) converted random cover and counts data whilst maintaining the majority of the ordinal structure and (ii) aligned abundance values regardless of whether it was recorded as a cover or count. A case study is presented, that uses real SACFOR observations, to demonstrate the conversion process and the application of statistical analyses routinely used in ecological assessments.

Conclusions: brief summary and potential implications: It is hoped that the SACFOR conversion process proposed here facilitates: (i) the quantitative re-analysis of the burgeoning SACFOR data repository; and (ii) initiates a debate on alternative methods for the conversion of SACFOR data into analysable end products.

Keywords: SACFOR scale, Conversion process, Semi-quantitative data analysis, Marine data reanalysis, Ordinal data, Benthic ecology

Introduction

The full quantification assessment of the seabed communities is often not possible or necessary. Investigations of marine habitats are often severely limited by the availability of survey time. For example, periodic tidal exposure, high ship costs and the limited bottom time of diving operations, all constrain the time available for the collection of information. This constraint is particularly acute when undertaking descriptive or inventory surveys of marine habitats, which requires the recording of numerous physical and biological variables (e.g. the identity and abundance of the common species present) across large areas of seabed. Habitats that are also highly heterogeneous or hard to sample (e.g. boulder-strewn shores) are also harder to assess quantitatively (Hawkins and Jones, 1992). Effective sampling using standard quantitative techniques, such as quadrats, is further hampered by, among others, a number of unknowns such as aversion of mobile species to sampling equipment, differential abilities to escape nets/traps, taxonomic uncertainty, cryptic species, differences in deployment of equipment between operators and visibility (Millier and Ambrose, 2000;

* Correspondence: james.strong@noc.ac.uk
[1]National Oceanography Centre, University of Southampton Waterfront Campus, European Way, Southampton SO14 3ZH, UK
Full list of author information is available at the end of the article

Guisan et al., 2006). Thus it could be argued that even "fully quantitative" techniques are in reality often actually semi-quantitative.

When standard quantitative sampling that results in counts of individuals or measurements of cover is not practical, biologists have developed various semi-quantitative scales, also called abundance scales, for the rapid assessment of abundance and cover (e.g. the Semi-Quantitative Macroinvertebrate Community Index (Stark, 1998) and EPOS ANTARKTIS Scale (Arnaud et al., 1990). Although these scales typically contain 5 to 7 broad categories and therefore lack the precision of quantitative methods, they do allow the coarse assessment of abundance both accurately and quickly (Hawkins and Jones, 1992). These scales were originally developed for terrestrial applications, such as the six point Braun-Blanquet cover-abundance scale (Braun-Blanquet 1932, 1964), which has been used extensively in Europe. Semi-quantitative scales remain the mainstay of terrestrial vegetational surveys. For example, the Domin scale of cover and abundance (Dahl and Hadac, 1941) remains at the heart of the UK's National Vegetation Survey (Rodwell et al., 2006).

Fischer-Piette (1936), an early pioneer of semi-quantitative scales, used a selection of similar scales to assess the biogeographic range of intertidal organisms. Southward and Crisp (1954) initially developed a log-base abundance scale for rapidly assessing marine communities at a varied of geographic locations. It is likely that this later developed, by Crisp and Southward (1958), into the ACFOR scales ('Abundant Common Frequent Occasional and Rare' - which also included a 'Not Found' class), which was used extensively for mapping the geographical distribution of marine species around British and European Coasts (pers. comm. S.J.Hawkins following discussions with both Crisp and Southward). The ACFOR scale was subsequently used for other studies of vertical and horizontal patterns (Nelson-Smith, 1967) and biologically-derived wave exposure scales (Ballantine, 1961) on rocky shores. More recently, the ACFOR scale was again adopted to resurvey sites assessed in the 1950s using ACFOR under the MARCLIM (Herbert et al., 2003, 2007; Simkanin et al., 2005; and specifically Mieszkowska et al., 2006a, 2006b). Hawkins and Jones (1992) provide a table that illustrates the relationship between ACFOR and abundance scales with as many as eight categories. They lament the fact that adding more categories spoil the semi-logarithmic progression of the original scales and may create an impression of spurious accuracy.

The ACFOR scales were ultimately used as the basis for the SACFOR (Superabundant, Abundant, Common, Frequent, Occasional and Rare) abundance scales – a system developed to support the Marine Nature Conservation Review (Hiscock, 1990) in its aim to survey and describe the marine habitats, communities and species around Great Britain. The SACFOR scale was originally developed as a standardised, semi-quantitative, methodology for experienced biologists undertaking roving surveying techniques such as diving, rapid intertidal surveys and subtidal video collection (Hiscock, 1998). The SAC-FOR scale records species in terms of percentage cover or counts (Table 1). The assessment based on cover is modified according to the growth form of the species (i.e. 'crust/meadow' or 'massive/turf') and the counts scale is modified by body size (< 1 cm; 1–3 cm; 5–15 cm; > 15 cm). The counts and cover scales use the same six classes, namely 'Superabundant', 'Abundant', 'Common', 'Frequent', 'Occasional', 'Rare', and 'Less than rare'.

The cover classes are separated by a base-2 logarithmic scale, i.e. the cover doubles between increasing classes. The counts codes are on a base-10 logarithmic scale, i.e. density changes 10-fold between classes. The growth form and body size 'block-shift' the appropriate SAC-FOR scale class for a particular growth form or body size. For example, large solitary ascidians are likely to fall into the 3–15 cm high category. For such species, a density of 1–9 per 100 m^2 would be classed as 'Occasional', while species over 15 cm high, such as a large anemone, occurring at this density would be classified as 'Frequent'. Example body size classes and growth forms for common British marine species are provided, with the SACFOR scale, in Table 1. Logarithms are commonly applied to raw, quantitative data to reduce the signal-to-noise ratio or to balance the influence of differences in relative abundance in some approaches – this process also reduces the numerical range of the data in a manner comparable to those used in many semi-quantitative scales. Raw, continuous data can be summed, divided and multiplied before having a logarithm applied. Furthermore, raw data that is log transformed is still continuous data. Scales, such as SACFOR, also have a greatly reduced range but cannot be initially changed through basic arithmetic operations.

The SACFOR scale has also beenused to define the representative communities for the biotopes listed in the UK's Marine Habitat Classification for Britain and Ireland (JNCC, 2015). As such, the SACFOR scale is now firmly established in the UK, being routinely used for undergraduate teaching (Hawkins and Jones, 1992; Gray and Elliott, 2009; Wheater et al., 2011) with the majority of the surveys relying on roving or remotely collected survey techniques. As of March 2017, a national database of marine survey data (UK Marine Recorder 'snap-shot' available from the Joint Nature Conservation Committee[1]) listed 1874 surveys using the SACFOR scale, which has collectively generated well over 1 million SACFOR observations in this database alone.

---

[1]http://jncc.defra.gov.uk/page-1599

Table 1 The SACFOR abundance scales for cover and counts observations (Hiscock, [1990]). SACFOR codes are: *S* Superabundant, *A* Abundant, *C* Common, *F* Frequent, *O* Occasional, *R* Rare, and *L* Less than rare indicated by extrapolation (which is no longer used in the modern SACFOR scale)

SACFOR cover scale			SACFOR counts scale		Size of individuals or colonies			
Percentage cover	Growth form		Counts (various spatial unit)	Minimum density at 1000 m^2				
	Crust/Meadow	Massive/Turf			< 1 cm	1–3 cm	3–15 cm	> 15 cm
> 80%	S		> 1 / 0.001 m^2 (1 × 1 cm)	10,000,000	S			
40–79%	A	S	1–9 / 0.001 m^2	1,000,000	A	S		
20–39%	C	A	1–9 / 0.01 m^2 (10 × 10 cm)	100,000	C	A	S	
10–19%	F	C	1–9 / 0.1 m^2	10,000	F	C	A	S
5–9%	O	F	1–9 / m^2	1000	O	F	C	A
1–5%	R	O	1–9 / 10 m^2 (3.16 × 3.16 m)	100	R	O	F	C
< 1%	L	R	1–9 / 100 m^2 (10 × 10 m)	10	L	R	O	F
		L	1–9 / 1000 m^2 (31.6 × 31.6 m)	1		L	R	O
			< 1 / 1000 m^2 (100 × 100 m)	0.1			L	R
			< 1 / 10,000 m^2 (1 km^2)	0.01				L
Example taxa	*Halichondria*	*Pachymatisma*			*Spirorbis*	*Grantia*	*Chaetopterus*	*Eunicella*
	Corynactis	*Alcyonium*			*Nucula*	*Nephtys*	*Arenicola*	*Funiculina*
	Fucoids	Foliose algae			Amphipods	*Pomatoceros*	*Sabella*	*Pachycerianthus*
	Kelp						*Zostera*	*Nephrops*

Although widely used in the UK, the SACFOR scale has several advantages as well as some acknowledged limitations associated with both data collection and analysis. The advantages of SACFOR include:

1. the rapid assessment of relative community composition, especially across expansive or rugose environments that may not be compatible with the use of more time-consuming or focused methods such as quadrats.
2. The simultaneous assessment of species enumerated as either cover or density (counts) using the same set of scales.
3. As semi-quantitative scales can be applied to larger areas, they are better suited for the detection of rare species that might over wise not be detected by less extensive methods.
4. The SACFOR scale can be used without additional equipment, hence making it a suitable method for diver-based seabed surveys.
5. Although the broad cover and count classes lack precision, their breadth ensures a high level of accuracy and repeatability between users – this design feature underpins its consistent application between users and across a variety of habitats.

These benefits confirm obvious and understandable limitations associated with the collection and processing of SACFOR data, which include:

1. Although supported by quantitative thresholds, SACFOR classifications are often applied in a subjective manner leading to intra and inter-observer variability over space and time - this can be reduced substantially with experience, training and predefined field methods.
2. The incremental changes between classes are large. Although the semi-logarithmetic progression of the classes is large, the size of the increments was carefully considered to reflect the natural abundance patterns of species, and thereby aid the survey in rapidly recording and reflecting the abundance patterns present (Hawkins and Jones, [1992]), i.e. the development of the ACFOR scale (Crisp and Southward, [1958]), which may have built on the earlier work of Fischer-Piette ([1936]) and Preston ([1948]).
3. Encoded SACFOR classes cannot easily be assessed directly with quantitative statistical methods, although many sophisticated statistical assessments can be undertaken on ordinal data.
4. Converting SACFOR codes into a corresponding number within the class value range still does not render the entire observation suitable for quantitative analysis – this is due to the presence of 'count' and 'cover' assessments within the same set of observations that operate over different value ranges. For example, counts range from 0 to abundances in excess of 1000,000 (increasing on a

base 10 logarithmic scale), whereas cover ranges from 0 to approximately 100 (increasing on a base 2 logarithmic scale). Direct conversion of mixed count and cover classes to numbers within the same sample will therefore lead to species assessed with counts to dominate the variance within the data. However, if one choses to accept that SACFOR cover and count classes are broadly aligned, it is possible to merge these observations into one ordinal outputs – this approach is the basis of the conversion process below.

Due to the inadmissibility of ordinal data for arithmetic operations, many common statistical operations are not suitable for ordinal data sets (Podani, 2006). As such, most SACFOR datasets are typically used once for descriptive purposes only (e.g. habitat classification). Hoever, some statistical methods are compatible for the analysis of ordinal data and include Mann-Whitney U tests (for comparisons differences between two independent groups) and Kruskal–Wallis H test (for comparisons between two or more independent groups). Multivariate techniques are less prevalent but include clustering methods (e.g. Ordinal Cluster Analysis described by Podani, 2006), non-metric multidimensional scaling (Digby and Kempton, 1987) and any tests allowing the similarity of objects to be based on rank values only (e.g. rank correlation, Legendre and Legendre, 2012). However, the conversion of ordinal data into continuous data, as provided by the SACFOR scale table, would greatly improve the availability of tests.

The SACFOR scale has now been in use for over 27 years and has generated a substantial quantity of observations – if the processing limitations can be overcome, this information could be suitable for other forms of analysis. This study describes a process for converting SACFOR encoded information into an ordinal scale that can be used in statistical analysis (i.e. ordinal values indicate an order or ranking between categories, but the actual distance between these orderings does not have any meaning). The conversion process (i) can combine SACFOR counts and cover information within one, data set, (ii) supports the merging of species (counts or cover) or observations during the production of the aligned data set, and (iii) allows a wide selection of quantitative statistics to be applied to the aligned data set, e.g. descriptive statistics, hypothesis testing, and multivariate analysis. A simulation study has been included to validate the conversion process and confirm the fidelity of the data during processing. The conversion has also been applied to a typical SACFOR data set to demonstrate some of the statistical methods that can be applied. SACFOR was originally designed for rapid biogeographic surveys and has been widely used since for a variety of purposes over

many decades of a wide variety of marine habitats; we propose a conversion process that provides a route for exploiting this wealth of data for a wider range of analyses.

It is acknowledged that data analysts regularly replace categorical and ordinal names and numbers with appropriate numbers to facilitate analysis. The value of these substitution techniques is that they are consistently applied to the same scale across studies to allow comparisons to be made. Despite the vast amount of SACFOR data available, there are no peer-reviewed published studies that have numerically converted this data for reuse (although see Burrows et al. (2008) for an example of the use of SACFOR data in an unconverted format). This study hope to highlight the subtle yet important changes that can occur within the converted dataset that can occur in what seems to be a deceptively simple process but is significantly complicated by body size and the combination of counts and cover observations. This study also hopes to provide a standardised approach for the conversion of SACFOR data that can be accessed by other scientists, thereby allowing the consistent conversion and analysis of this valuable data set between studies.

The specific objectives of this analysis are:

1) To present a conversion process that translates SACFOR codes into numerical values, which allows observations to be merged (counts with other count data and cover with other cover data only).
2) To assess the fidelity of conversion for SACFOR <u>count</u> codes converted to values.
3) To assess the fidelity of conversion for SACFOR <u>cover</u> codes converted to values.
4) Validate the alignment of converted cover and counts observations within a single, ordinal data set.
5) Present a validated conversion pathway for SACFOR information and recommend statistical analyses that are suitable for converted and aligned data sets.

Materials and methods
The first section describes the development of the conversion process. The second section details the final process used to convert SACFOR classes (counts and cover) into an aligned, numerical dataset. The third section describes the simulation tests (random data) and case study (real data) used to validate the conversion process.

Development of the conversion process
The desired attributes for the conversion process were as follows:

1. The conversion merges the observations, based on counts and cover, into one, unified community matrix;
2. The influence of body size and growth form are removed from the data set so that changes in absolute abundance (as measured as counts or cover) is the only factor generating change in the data set;
3. Where possible, as much relative information between classes should be maintained in the final matrix;
4. The final expression of the counts and cover observations must be on the same value range; and
5. Converted values are distributed in a similar pattern across the value range regardless of source (counts or cover).

Attributes 4 and 5 were considered particularly important to prevent the type of observation (counts or cover) weighting or biasing the final matrix i.e., the larger value range for species assessed with the counts scale translates to a greater influence within the community matrix when examined with univariate and multivariate statistical analyses. Without alignment, the results from these analyses will, in part, be driven by changes in the proportion of species assessed with either the counts or cover scales rather than underlying changes in abundance. As such, it was necessary to fit both counts and cover observations onto the same value range.

The primary requirement to prevent artefacts appearing in the unified community matrix required that most of the relative information between classes was removed, thereby compromising point 4. As the SACFOR count scale has increments based on a power of 10 but the cover scale is based on a power of 2, it was not possible to maintain this relative information without introducing artefacts into the community matrix (and compromising point 6). During the development of the conversion process, several other methods were examined – these included:

1) Processes that used body size to estimate the area occupied by individuals and thereby derive cover for taxa enumerated with the counts scale. This system allowed us to understand the relationships between abundance and cover for different body sizes. However, the resulting value range for cover values converted using body size and counts, was very different to the existing cover value range. Attempts to align the existing cover values with them compromised the counts data. The conversion process posited here conversely aligns cover data to values derived from the counts scale.
2) Processes that retained the power of 10 and 2 increments for the counts and cover data
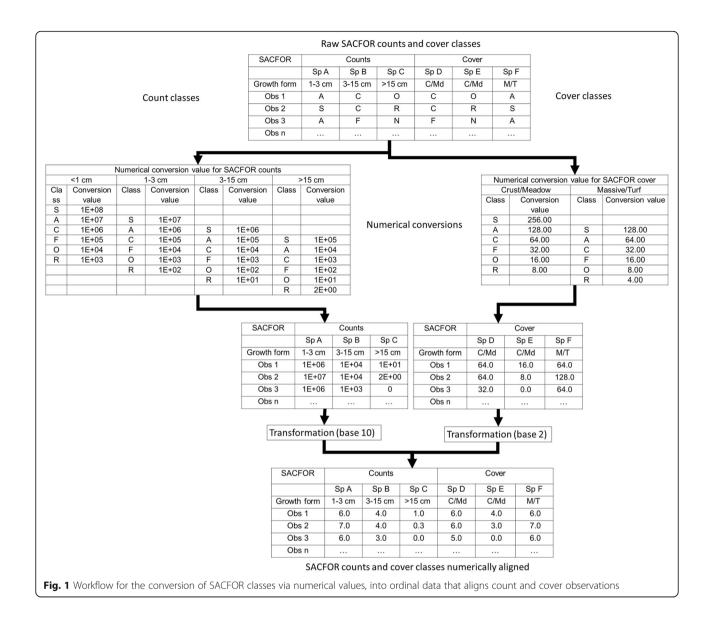
respectively. However, attempts to keep the relative information for the counts and cover classes within one value range resulted in count data over-powering the variance within the unified data set. As stated earlier, discrepancies in the final representation of counts and cover observations in the community matrix compromised subsequent analyses, i.e. differences between communities could be driven simply by the ratio of counts and cover observations in within a data set.
3) Standardized conversions that attempted to align counts and cover yet maintain the different relative step changes for counts and cover were all unable to prevent significant artefacts appearing in the final community matrix.

Ultimately, the objective for the conversion process is to allow some basic statistical analysis of count and cover data merged into one data set. As such, the conversion process selected for use here removes the majority of the relative information and aligns the count and cover observations within an unified, ordinal value range. Based on the incompatibility of the original units used for cover and counts (i.e. density verses percentages), it is not possible to merge the two types of data into a completely ordered set. However, if the ordering of merged count and cover observations purely relies on the merging of information at the categorical level (i.e. 'Common' refers to the same level of abundance regardless of whether it is derived from counts or cover) information, then the creation of a totally ordered set it possible. It is acknowledged that this represents a significant simplification of the data. However, the benefit of being able to perform statistical analyses on a larger, unified dataset representing the entire community, potentially outweighs the loss of information inherent in the original cover and counts units. This conversion meets all but one (point 4) of the desired attributes, and provides a reliable and unified community matrix for subsequent analysis. Certain statistical limitations are imposed through the use of ordinal data – these are described in more detail in the discussion. Alternative methods were examined that convert cover to counts based on the average body size (and areal footprint estimated) of 'cover' species. Unfortunately, the body size/areal footprint was not available for all of the species assessed using cover, hence it could not be implemented here. The authors are continuing to collate information on body size in the hope that it can be incorporated into a more robust merging of cover and counts in future iterations of this process.

Process for the numerical conversion of SACFOR data

Step 1) Attribution of observations with species body size (counts) and growth form (cover)

Count classes — **Numerical conversions** — **Cover classes**

Raw SACFOR counts and cover classes

SACFOR	Counts			Cover		
	Sp A	Sp B	Sp C	Sp D	Sp E	Sp F
Growth form	1-3 cm	3-15 cm	>15 cm	C/Md	C/Md	M/T
Obs 1	A	C	O	C	O	A
Obs 2	S	C	R	C	R	S
Obs 3	A	F	N	F	N	A
Obs n	…	…	…	…	…	…

Numerical conversion value for SACFOR counts

<1 cm		1-3 cm		3-15 cm		>15 cm	
Class	Conversion value	Class	Conversion value	Class	Conversion value	Class	Conversion value
S	1E+08						
A	1E+07	S	1E+07				
C	1E+06	A	1E+06	S	1E+06		
F	1E+05	C	1E+05	A	1E+05	S	1E+05
O	1E+04	F	1E+04	C	1E+04	A	1E+04
R	1E+03	O	1E+03	F	1E+03	C	1E+03
		R	1E+02	O	1E+02	F	1E+02
				R	1E+01	O	1E+01
						R	2E+00

Numerical conversion value for SACFOR cover

Crust/Meadow		Massive/Turf	
Class	Conversion value	Class	Conversion value
S	256.00		
A	128.00	S	128.00
C	64.00	A	64.00
F	32.00	C	32.00
O	16.00	F	16.00
R	8.00	O	8.00
		R	4.00

SACFOR	Counts		
	Sp A	Sp B	Sp C
Growth form	1-3 cm	3-15 cm	>15 cm
Obs 1	1E+06	1E+04	1E+01
Obs 2	1E+07	1E+04	2E+00
Obs 3	1E+06	1E+03	0
Obs n	…	…	…

SACFOR	Cover		
	Sp D	Sp E	Sp F
Growth form	C/Md	C/Md	M/T
Obs 1	64.0	16.0	64.0
Obs 2	64.0	8.0	128.0
Obs 3	32.0	0.0	64.0
Obs n	…	…	…

Transformation (base 10) — Transformation (base 2)

SACFOR	Counts			Cover		
	Sp A	Sp B	Sp C	Sp D	Sp E	Sp F
Growth form	1-3 cm	3-15 cm	>15 cm	C/Md	C/Md	M/T
Obs 1	6.0	4.0	1.0	6.0	4.0	6.0
Obs 2	7.0	4.0	0.3	6.0	3.0	7.0
Obs 3	6.0	3.0	0.0	5.0	0.0	6.0
Obs n	…	…	…	…	…	…

SACFOR counts and cover classes numerically aligned

Fig. 1 Workflow for the conversion of SACFOR classes via numerical values, into ordinal data that aligns count and cover observations

Each species observation must be attributed according to whether it has been assessed according to cover or counts. Species encoded with the counts scale must be attributed according to the body size scale used. Species using the cover scale must also be attributed according to the growth form scale used. The growth form and body size information is usually provided as survey metadata or can be estimated using biological information from online sources e.g. BIOTIC - Biological Traits Information Catalogue.[2] An overview of the conversion process is provided in Fig. 1.

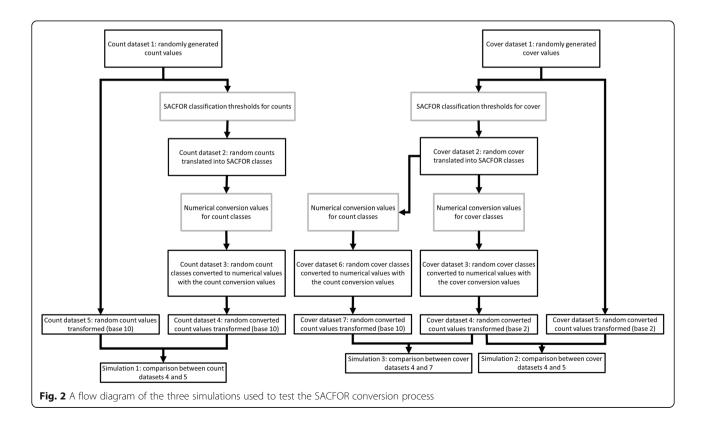Step 2) Numeric conversion of counts and cover

The conversion values for the counts are based on the lowest possible density for each class. A constant of 0.1 was added to each conversion value to ensure that all of the values can be log transformed correctly (i.e. to avoid the log transformation of 1 returning 0) – the resulting values are the 'numerical conversion values for counts'. The lowest possible density was selected to numerically represent each class because the mid and upper values cannot be defined for the superabundance class of any size class. The numerical conversion values for the cover classes are based on the conversion value for count classes. To derive the conversion values for the cover classes, the numerical conversion values for the counts were log transformed (base 10) before being antilog transformed (base 2). All of the final conversion values for counts and cover are shown in Table 2.

[2]http://www.marlin.ac.uk/biotic/

Table 2 Numerical conversion values for SACFOR cover and counts classes. SACFOR classes are *S* Superabundant, *A* Abundant, *C* Common, *F* Frequent, *O* Occasional, *R* Rare; and *L* Less than rare indicated by extrapolation (which is no longer used in the modern SACFOR scale)

Numerical conversion value for SACFOR count classes for each of body sizes								Numerical conversion value for SACFOR cover classes for each growth forms			
< 1 cm		1–3 cm		3–15 cm		> 15 cm		Crust/Meadow		Massive/Turf	
Class	Conversion value	Class	Conversion value	Class	Conversion value	Class	Conversion value	Class	Conversion value	Class	Conversion value
S	1E+ 08							S	256		
A	1E+ 07	S	1E+ 07					A	128	S	128
C	1E+ 06	A	1E+ 06	S	1E+ 06			C	64	A	64
F	1E+ 05	C	1E+ 05	A	1E+ 05	S	1E+ 05	F	32	C	32
O	1E+ 04	F	1E+ 04	C	1E+ 04	A	1E+ 04	O	16	F	16
R	1E+ 03	O	1E+ 03	F	1E+ 03	C	1E+ 03	R	8	O	8
L	NA	R	1E+ 02	O	1E+ 02	F	1E+ 02	L	NA	R	4
		L	NA	R	1E+ 01	O	1E+ 01			L	NA
				L	NA	R	1E+ 00				
						L	NA				

To convert SACFOR counts information, each class should be substituted with the corresponding 'numerical conversion values for counts' - each body size has a specific set of numerical conversion values (Table 2). To covert SACFOR cover information, each class should be substituted with the corresponding 'numerical conversion values for cover' - once again, each growth form has a specific set of numerical conversion values for cover (Table 2). These conversion values should not be interpreted as abundances or cover values - they are conversion numbers that will align with the converted cover and count values onto an ordinal scale <u>after</u> transformation (step 3).

Fig. 2 A flow diagram of the three simulations used to test the SACFOR conversion process

Table 3 Data source and composition for the two sites containing sublittoral rock

Survey area	Depth (m)	Distance from shore (km)	Substratum	Year of survey	Observations (stations)	Region
East of Haig Fras SAC/MPA	~ 97	64.30	Bedrock	2015	1053 (132)	Celtic Sea beyond the coastal margin
Wyville Thomson Ridge SAC/MPA	~ 562	157.42	Bedrock	2012	345 (25)	Atlantic Ocean at the northern end of the Rockall Trough

Step 3) Alignment of the numerically converted counts and cover through transformation

The final step aligns the numerical count and cover values along an ordinal value range. To achieve this, the conversion values for counts are log transformed (base 10). The conversion values for the cover information are log transformed (base 2). This step unifies the count and cover information within a single range of values, i.e. the transformed value for a species assessed as 'Common' using counts is the same as another species assessed as 'Common' using cover. The final values are: (i) adjusted to remove the influence of body size and growth form; (ii) merged with similar taxonomic/morphological entries when required; (iii) numerically aligned to prevent offsets between those measured with counts and those as a cover; and (iv) log transformed (appropriate for observations spanning multiple orders of magnitude). As mentioned earlier, it was not possible to maintain the relative information separating classes – as such, the aligned values are ordinal in nature.

Validation of the process for the conversion of SACFOR data

Simulations using randomly generated data were used to test the fidelity of the conversion process. In addition, a case study converted real SACFOR data to demonstrate the validated conversion process and the potential analyses that can be applied. The three simulations and the case study used R (R Core Team, 2013) - the scripts are available within the supplementary information. The linkages between the simulations are shown in Fig. 2. The simulation and demonstration steps are:

- Simulation 1 - assess the fidelity of the conversion of a random count-based SACFOR data set into numerical values and comparison with a basic rank value conversion;
- Simulation 2 - assess the fidelity of the conversion of a random cover-based SACFOR data set into numerical values and comparison with a basic rank value conversion;
- Simulation 3 - assess the alignment of numerical count and cover values within an unified ordinal data set; and
- Case study 1 - demonstrate the conversion of a real data set, containing both count and cover observations, as well as some standard statistics for the detection of changes between sites
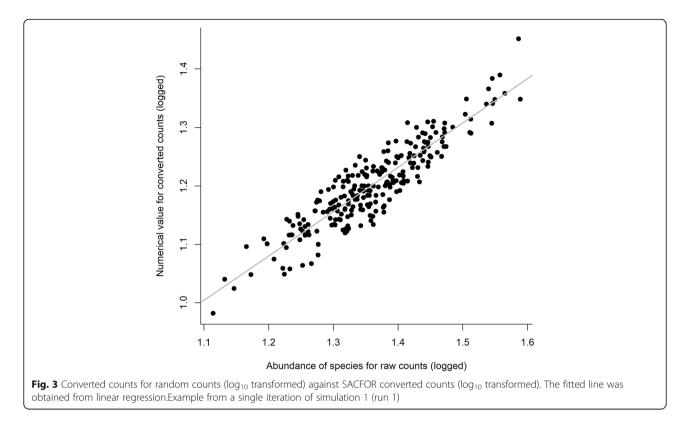
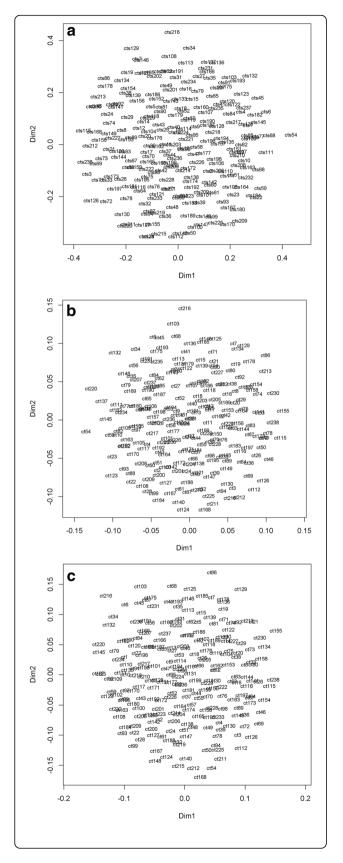Simulation 1: comparison between random counts values with numerically converted and transformed count values.

Hypothesis: there is no appreciable difference between randomly generated count data and the numerically converted, $\log_{10}$ transformed, counts data.

A random set of count data was generated using R. The 'rnorm' function in R generated random values using a multivariate lognormal distribution (mean = 0, variance = 2). The random count data set was designed to reflect a typical SACFOR data set. The Marine Recorder database contains the majority of the UK's SACFOR surveys. Microsoft Access was used to establish the average number of observations collected by a survey using SACFOR (a mean of 560 observations based on 1874 surveys) and the average number of species encoded within a survey (a mean of 119 species based on 1874 surveys). The data frame dimensions were therefore 119 species variables (columns) and 560 observations (rows). The L code (Less than rare indicated by

Table 4 Descriptive statistics and correlation coefficients for randomly generated count data sets (i) before and (ii) after the SACFOR encoding and numeric conversion. Both before and after data sets are transformed ($\log_{10}$). Values are based on 10 iterations of the simulation

	Number of species per observation		Density/numerical conversion value	
	Raw density	Converted value	Raw density	Converted value
Mean (standard deviation)	106.88 (± 3.15)	106.88 (± 3.15)	1.39 (± 0.08)	1.22 (± 0.07)
Paired values (standard deviation)	99.63% (± 0.34)	99.63% (± 0.34)	0.01% (± 0.00)	7.30% (± 2.64)
Spearman rank test *rho* statistic (standard deviation)	1.00 (± 0.00)		0.89 (± 0.02)	

Fig. 3 Converted counts for random counts (log₁₀ transformed) against SACFOR converted counts (log₁₀ transformed). The fitted line was obtained from linear regression.Example from a single iteration of simulation 1 (run 1)

extrapolation) was not used as it not included in the vast majority of marine data sets.

The random count observations were then classified into SACFOR classes using the standard SACFOR thresholds provided in Table 1 (based on a body size of 1–3 cm). The SACFOR classes were then substituted with the 'numerical conversion values for counts' appropriate for each SACFOR class (Table 2). As a comparison, SACFOR values were also substituted with their ranked values, i.e. S = 6, A = 5, C = 4, F = 3, O = 2, R = 1, absent = 0. Finally, the random count data set and numerically converted count data set were both log transformed (based 10). Tests conducted on the two data sets were correlation between paired samples using

Spearman rank *rho*. PERMANOVA, using default options in the ADONIS function in the 'vegan' R package, was used on both the numerical conversion values and the rank value substitutions. Simulation 1 was repeated ten times and the mean of each statistic was reported with the standard deviation.

Simulation 2: comparison between random cover values with numerically converted and transformed count values.

Hypothesis: there is no appreciable difference between randomly generated cover data and the numerically converted, $\log_2$ transformed, cover data.
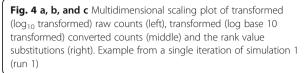
A random cover data set was generated using R. For each observations, a random species is selected and given a random cover value from between 0 and 100. A

Table 5 PERMANOVA comparison of randomly generated count data sets (i) before and (ii) after the SACFOR encoding and numeric conversion. Both before and after data sets are transformed (log₁₀) before analysis and converted to: (i) relative abundances; and (ii) rank of abundances. Values are based on 10 iterations of the simulation. Number of permutations: 999

Logged values	Df	Sequential sums of squares	Mean squares	F statistic	Partial R-squared	p value
Factor: before/after counts conversion	1	2.448 (± 0.149)	2.448 (± 0.149)	54.669 (± 6.166)	0.103 (± 0.010)	0.001 (± 0.000)
Residuals	474	21.355 (± 1.495)	0.045 (± 0.003)		0.897 (± 0.010)	
Total	475	23.803 (± 1.429)			1.000	
Rank values						
Factor: before/after counts conversion	1	2.314 (± 0.052)	2.314 (± 0.052)	47.050 (± 3.826)	0.090 (± 0.007)	0.001 (± 0.000)
Residuals	474	23.432 (± 1.608)	0.049 (± 0.003)		0.910 (± 0.007)	
Total	475	25.746 (± 1.568)			1.000	

Df degrees of freedom

Fig. 4 a, b, and c Multidimensional scaling plot of transformed (log$_{10}$ transformed) raw counts (left), transformed (log base 10 transformed) converted counts (middle) and the rank value substitutions (right). Example from a single iteration of simulation 1 (run 1)
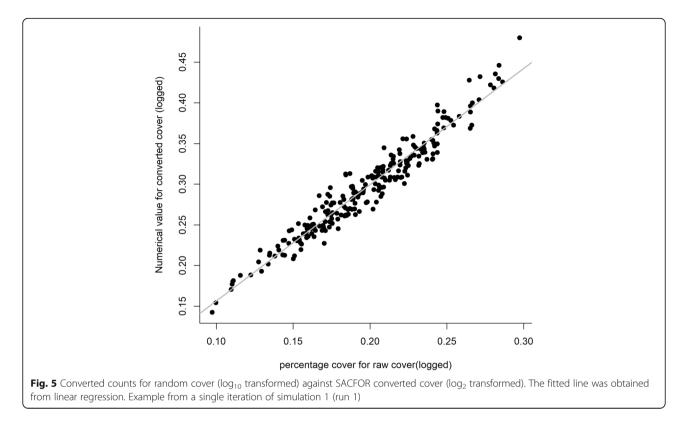
loop is used to: (i) then calculate the remaining area; (ii) randomly select a species not already allocated a cover value; and (iii) randomly allocate a cover value within the remaining range of available cover – this continues until there is no remaining cover within an observations. Once again, the 'less than rare indicated by extrapolation' L code was not used.

The data frame dimensions were 119 species variables (columns) and 560 observations (rows). The random cover observations were then classified into SACFOR classes using the standard SACFOR thresholds provided in Table 1. The SACFOR classes were then substituted with the 'numerical conversion values for cover' appropriate for each class (Table 2). As a comparison, SACFOR values were also substituted with their ranked values, i.e. S = 6, A = 5, C = 4, F = 3, O = 2, R = 1, absent = 0. Finally, both the random cover values and the numerically converted cover values were $\log_2$ transformed. Tests conducted on the two data sets were correlation between paired samples using Spearman rank *rho* and PERMANOVA, using default options in the ADONIS function in the 'vegan' R package, was used on both the numerical conversion values and the rank value substitutions. Simulation 2 was repeated ten times and the mean of each statistic was reported with the standard deviation.

Simulation 3: assessment of the alignment of cover and counts values on an ordinal scale following the numerical conversion and transformation process.

Hypothesis: for a randomly generated data set of SACFOR classes, there is no appreciable difference between the final ordinal values regardless of whether the counts or cover conversion processing route is followed.

The SACFOR scale for count-based scale is structured on base 10 increments. The SACFOR cover scale has base 2 increments. Real SACFOR data is always a mix of both count and cover observations. A primary objective of the numerical conversion process is that the conversion should result in the same transformed value for each class, regardless of whether it was recorded as cover or counts, i.e. an 'Abundant' count should have the same value as an 'Abundant' cover after transformation. Simulation 3 used the SACFOR classes generated from the randomly generated counts (Simulation 1). These classes were then converted with the cover conversion process. The converted counts values from the counts conversion route (simulation 1) and counts values from the cover conversion route (simulation 3) were compared statistically with PERMANOVA (relative

Fig. 5 Converted counts for random cover (log$_{10}$ transformed) against SACFOR converted cover (log$_2$ transformed). The fitted line was obtained from linear regression. Example from a single iteration of simulation 1 (run 1)

abundances and using default adonis options). Simulation 3 was iterated ten times and the mean of each statistic was reported with a standard deviation. It was not necessary to run Simulation 3 to confirm that the rank value substitution method would align counts and cover observations.

Case study 1: community comparison between two sublittoral rock sites using real SACFOR data (containing a mix of both count and cover values) after applying the conversion and transformation process.

Hypothesis: a significant community different is apparent between two sublittoral sites and this can be detected following the conversion and transformation of SACFOR classes.

SACFOR data for two sublittoral rock outcrops (East of Haig Fras SAC and Wyville Thomson Ridge SAC - Table 3) were extracted from Marine Recorder.[3] SACFOR observations were obtained from drop-down camera observations (comparable equipment used on both surveys). Both sites are in UK waters and contain sublittoral rock substrata dominated by epifaunal species. Survey data from both sites were merged into one species matrix. Different taxonomic levels and labels had been used for many of the species and groups. After numerical conversion, taxa were merged into a higher, unifying taxonomic identifiers, e.g. records for (i) *Caryophyllia smithii* (ii)

Caryophyllia sp., and (iii) *Caryophyllia*, were merged into 'Caryophyllia' to improve the consistency between sites for these species. Taxonomic entries higher than a family were removed from the matrix, e.g. porifera.

The SACFOR classes were converted numerically using the numerical conversion values for counts and cover. Log transformed using base 10 and base 2 were used to align the counts and cover data sets respectively. Multivariate statistics suitable for ordinal data was used to test for (i) differences between the communities at the two sites using PERMANOVA (relative abundance and using default adonis options) and (ii) the influence of environmental variables on the communities using Correspondence Analysis and Redundancy Analysis (vegan package) in R. Although initially controversial (Sullivan & Artino, 2013), it is now accepted that both parametric (requiring an adequate sample size and data that are normally distributed) and non-parametric tests are appropriate for the analysis of ordinal (i.e. Likert scales) data dependent variables (Norman, 2010). Descriptive statistics should use the median as a measure of central tendency rather than means (Jamieson, 2004).

Results

Simulation 1: the fidelity of the conversion process for SACFOR count classes converted to numerical values

Simulation 1 generates a random counts dataset, encodes using the SACFOR scale, and then applies the

---

[3]http://jncc.defra.gov.uk/page-1599

Table 6 Descriptive statistics and correlation coefficients for randomly generated cover data sets (i) before and (ii) after the SACFOR encoding and numeric conversion. Both before and after data sets are transformed (log_2) and converted to relative abundances. Values are based on 10 iterations of the simulation

	Number of species per observation		Density/numerical conversion value	
	Raw density	Converted value	Raw density	Converted value
Mean (standard deviation)	9.63 (± 2.02)	9.63 (± 2.02)	0.20 (± 0.04)	0.29 (± 0.06)
Paired values (standard deviation)	99.37% (± 0.43)	99.37% (± 0.43)	2.61% (± 1.11)	19.20% (± 2.76)
Spearman rank test *rho* statistic (standard deviation)	1.00 (± 0.00)		0.97 (± 0.01)	

numerical conversion process to these codes. For a comparison, a basic ranked value has also been used to substitute the SACFOR codes. Statistical testing was used subsequently to detect relative changes between: (i) the original random dataset (log transformed) and the converted values (log transformed); and (ii) the original random dataset (log transformed) and the ranked values.

There was a significant difference between the transformed (mean) abundance before the conversion process and the numerical values used to represent abundance after conversion (Table 4 and Fig. 3). This difference was also apparent for the rank value substitution. The numerical value is substantially smaller than the original abundance. However, the conversion process, and the numerical conversion values used, are not designed to provide an absolute match with the abundances but rather to capture the relative differences between classes. As such, both the descriptive statistics indicate a substantial difference (also tested with a Wilcoxon rank sum test but not shown).

The Spearman rank test has been included to examine maintenance of relative sorting before and after the conversion process. This indicates that the majority of the relative order has been maintained during the conversion process. The process of classifying the abundance using SACFOR removes a large amount of quantitative information (i.e. the full value range is reduced to just six classes). This simplification of the data is highlighted by the increase in tied values post-conversion. Tied values disrupt the ranking process

and may explain some of the decline in rho statistic from an ideal value of 1.

Analysis using PERMANOVA found that there was a significant difference between the numerically converted data set and the original as well as between the ranked values and the original dataset (Table 5). Multidimensional scaling plots for the raw, converted count observations and rank value substitution are provided in Fig. 4. Comparisons of the transformed data sets (the final product of the conversion process) generates no patterns to artefact structures within the plots, suggesting that the entire conversion process does not impart any structure or artefacts within the data. Equally no artefacts were observed in the MDS plot for the rank value substitution (Fig. 4c).

Simulation 2: comparison between random (raw) cover values and converted SACFOR cover values

Simulation 2 generates a random cover dataset, encodes using the SACFOR scale, and then applies the conversion process to these codes. Statistical testing was subsequently used to detect relative changes between the original raw dataset and the converted values. The descriptive statistics indicated significant differences between the raw (random) and converted cover values for the converted/numerical values (Fig. 5) but not for the number species (Table 6). Once again, it is expected that the pre-conversion 'cover' and post-conversion 'numerical conversion value' does not match - the conversion process, and the numerical conversion values used, are not designed to provide an

Table 7 PERMANOVA comparison of randomly generated cover data sets (i) before and (ii) after the SACFOR encoding and numeric conversion. Both before and after data sets are transformed (log_2) before analysis and converted to: (i) relative abundances; and (ii) rank of abundances. Values are based on 10 iterations of the simulation. Number of permutations: 999

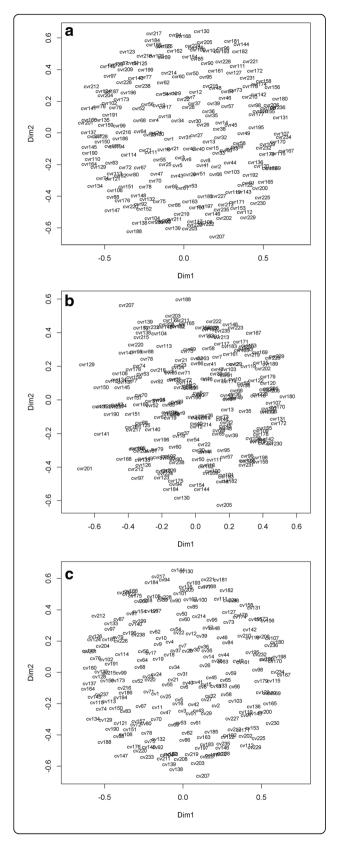| Logged values	Df	Sequential sums of squares	Mean squares	F statistic	Partial R-squared	p value
Factor: before/after cover conversion	1	0.832 (± 0.000)	0.832 (± 0.000)	1.874 (± 0.000)	0.004 (± 0.000)	0.001 (± 0.000)
Residuals	474	210.493 (± 0.000)	0.444 (± 0.000)		0.996 (± 0.000)	
Total	475	211.271 (± 0.000)			1.000 (± 0.000)	
Rank values	Df	Sequential sums of squares	Mean squares	F statistic	Partial R-squared	p value
Factor: before/after cover conversion	1	0.050 (± 0.003)	0.050 (± 0.003)	0.112 (± 0.007)	0.000 (± 0.000)	0.001 (± 0.000)
Residuals	474	212.591 (± 0.388)	0.449 (± 0.001)		1.000 (± 0.000)	
Total	475	212.41 (± 0.381)			1.000	

Df degrees of freedom

Fig. 6 a, b, and c Multidimensional scaling plot of transformed (log$_2$ transformed) raw cover (left), transformed (log base 2 transformed) converted cover (middle) and the rank value substitutions (right). Example from a single iteration of simulation 1 (run 1)

absolute match with the abundances but rather to capture the relative differences between classes. Despite a change in the absolute values, the relative ordering of the observations appears, as captured by the Spearman rank tests, are similar before and after the conversion process. Any changes in the ordering may be related to the increase in frequency of tied values following the encoding of values with the SACFOR scale (paired cover values increases from 2.6 to 19.2% during the encoding phase.

Analysis using PERMANOVA found that there was a significant difference between the numerically converted data set and the original as well as between the ranked values and the original dataset (Table 7). Multidimensional scaling plots for the raw, numerically converted cover and the rank value substitutions observations (both untransformed and transformed) are provided in Fig. 6. The figures are all similar and plot the observations in a loose circle. This structure is similar both before (Fig. 6) and after numerical conversion and transformation (Fig. 6b) as well as in the plot for the rank value substation (Fig. 6c).

Simulation 3: confirmation of the alignment of cover values and counts after transformation

Simulation 3 converted randomly generated SACFOR class (letters not values) data sets (10 iterations) using both the cover and then counts conversion processes. The converted data sets from both processes were then compared statistically, using PERMANOVA, to confirm the similarity, and hence alignment, of the cover and counts conversion processes. The small F statistic and a p value greater than 0.05 suggest that the transformed values produced by the cover and counts conversion processes are the same (Table 8).

Case study: detection of difference between two sites, within a real SACFOR dataset, containing a mix of both count and cover values (transformed), after conversion

The case study is based on two real SACFOR surveys (both containing a typical mixture of cover and count observations, as well as a range of body sizes and growth forms). The conversion processes have been applied to both surveys to demonstrate its application for real data and that a typical suite of statistical tests can be applied.

Species richness was similar between sites (Table 9). The F and p values returned by the PERMANOVA indicate a large and significant difference between the relative abundances of the epifaunal communities at the two

Table 8 PERMANOVA comparison of randomly generated SACFOR cover classes converted using (i) the <u>counts</u> conversion and transformation (log_{10}) process and (ii) the <u>cover</u> conversion and transformation (log_2) process and converted to relative abundances. Values are based on 10 iterations of the simulation. Number of permutations: 999

	Df	Sequential sums of squares	Mean squares	F statistic	Partial R-squared	p value
Factor: counts or cover process	1	0.103 (± 0.01)	0.103 (± 0.01)	0.233 (± 0.01)	0.001 (± 0.00)	1.000 (± 0.00)
Residuals	474	209.08 (± 0.56)	0.441 (± 0.01)		0.999 (± 0.00)	
Total	475	209.19 (± 0.56)			1.000 (± 0.00)	

Df degrees of freedom

sites (Table 10). This difference is apparent as a low level of overlap between the site point clouds displayed in the Multi-dimensional plot below (Fig. 7). Correspondence analysis has been used to highlight environmental variables that co-vary with the epifaunal community. The suite of environmental variables included did not explain much of the variance (inertia) present in the epifaunal data (Table 11). Co-varying environmental variables include depth, mud/boulder content and surface rugosity (Fig. 8).

Discussion

The SACFOR conversion process advocated here allows: (i) the merging of taxa within counts or cover data subsets; (ii) observations, based on either counts and cover, to be unified into one matrix; (iii) counts and cover data to have an equal weighting in the final matrix; and (iv) the removal of the influence of body size and growth form from the final values. To achieve this, it is only possible to preserve the ordinal structure of the data set i.e., while the order of the variable has been retained, the spacing of the original classes (base 2 for cover and base 10 for counts) variable has been removed. At no point within the conversion process do the numerical values attempt to correspond to the cover or abundance values presented by the SACFOR scale. Once transformed, the relative differences between classes for counts and cover are effectively lost. If it is more important for the user to analyse relative change, it is advised that step 3 (transformation) is not undertaken and the counts and cover observations are not merged but analysed separately. Equally, it is likely that comparisons made within SACFOR data are likely to be more powerful when factors that introduce variance, such data sets containing both cover and counts or those comprising multiple body sizes, are minimised. It is likely that more power might be obtained by extracting and using data sets confined to a single growth form.

Simulations 1 and 2 verified that the SACFOR conversion process can convert random cover and counts data to numerical values (allowing the merging of taxa) and then to transformed values whilst maintaining the majority of the ordinal structure. A small loss of relative sorting associated with simulations 1 and 2 were associated with paired values, which themselves are a product of the full value range present in the random data sets being reduced to 7 classes during the SACFOR encoding phase i.e., a step within the data collection phased and not the numerical conversion process itself. The agreement between the scale classes and the numerical equivalents is an obvious reflection of the careful structuring and design of the SACFOR, and also its precursor, the ACFOR scale. Interestingly, a similar result was obtained by simply substituting SACFOR codes with a rank value. Despite this, the numerical conversion provides two important advantages of the rank value substitution method. These advantages are firstly the ability to merge observations together because the numerical conversion process as an intermediate step that approximates the absolute abundance values (i.e. the ability to merge taxa into higher taxonomic levels) and secondly the ability to incorporate quantitative observations with the converted SACFOR observations. The latter step is also possible, in a coarser manner, with the rank value substitution method. Simulation 3 confirmed that the numerical alignment of abundance values regardless of whether it was recorded as a cover or count, and also means that data sets containing both types of information can be safely analysed as one combined package of observations. Clearly the use of the same rank value scale for counts and cover will also allow the alignment of the two different abundance types.

A case study has been presented that uses real SACFOR observations, i.e. a matrix including species encoded according to counts (multiple body sizes) and cover (both growth forms). The real SACFOR observations within the case study were converted and presented as one species matrix. Common tests, such as PERMANOVA and

Table 9 Descriptive statistics for East of Haig Fras and Wyville Thompson Ridge SACFOR epifaunal data (relative abundance) following SACFOR conversion and log transformation

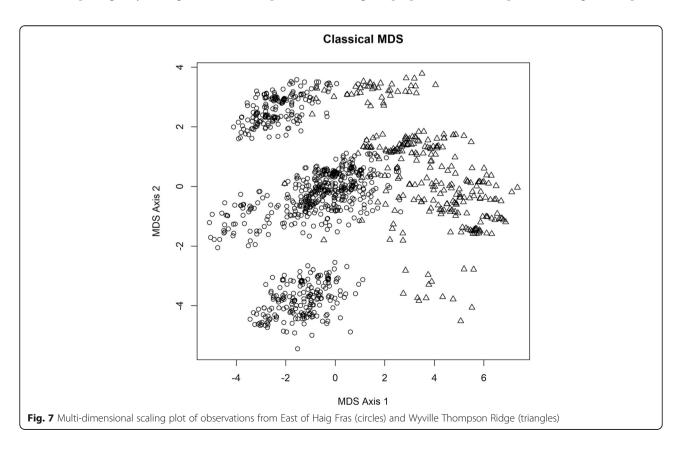| Site	Mean number of species (standard deviation)	Converted and transformed ordinal 'abundance'
East of Haig Fras	4.46 (± 1.97)	0.32 (± 0.14)
Wyville Thompson Ridge	4.73 (± 1.78)	0.33 (± 0.14)

Table 10 PERMANOVA comparison of SACFOR epifaunal data (relative abundance) from: (i) East of Haig Fras; and (ii) Wyville Thompson Ridge, following SACFOR conversion and transformation. Number of permutations: 999

	Df	Sequential sums of squares	Mean squares	*F* statistic	Partial R-squared	*p* value
Factor: sites	1	63.383	63.379	270.652	0.162	0.001
Residuals	1339	328.785	0.234		0.838	
Total	1340	392.164			1.000	

Df degrees of freedom

Canonical correspondence analysis, were used to demonstrate that the converted data are compatible with statistical analyses routinely used in ecological assessments. Indeed, it is recognised that semi-quantitative data such as SACFOR are compatible with a broad suite of non-parametric statistical methods including simple (e.g. difference tests, correlation and concordance, and ANOVA analogues) and complex (multidimensional scaling and Permanova) techniques (Legendre & Legendre, 2012). Most non-parametric tests are as powerful as their parametric equivalents, and if there is any doubt about equality of variances or divergence from normal distributions, then this small advantage provided by parametric approaches breaks down quickly (Field et al., 2012).

If the objective of the analysis is to assess the response of the whole community in relation to treatments or environmental variables, multivariate approaches (e.g. the *mvabund* package by Wang et al., 2012) provide an alternative to the conversion process suggested here. For example, the mvabund package (Wang et al., 2012) fits individual generalised linear models to species in a multispecies data set but summarise the models collectively to make conclusions on the influence of treatments and variables. The benefit of this approach is that each model can be based on differing scales and units of 'abundance' for each species, hence allowing the simultaneous utilisation of cover and counts class data sets without an initial merging step (as required in the process proposed here).

In order to better reflect reality it may be necessary to refine the method used in this study for each particular situation. It is acknowledged that the method to generate the random data set used in the simulations assumes that the distribution of abundance for each species is both identical and independent of all other species. However, actual marine communities have relatively few common species and a higher proportion of rare species, leading to a species
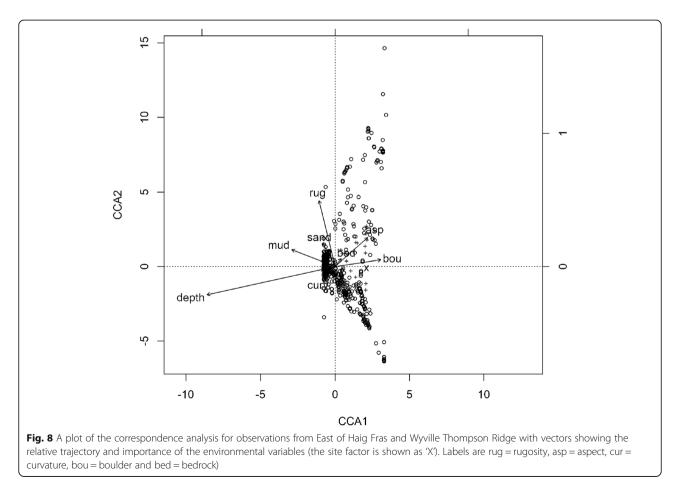

Fig. 7 Multi-dimensional scaling plot of observations from East of Haig Fras (circles) and Wyville Thompson Ridge (triangles)

Table 11 Correspondence analysis of SACFOR epifaunal data (relative abundance) from East of Haig Fras and Wyville Thompson Ridge with depth, seabed rugosity, seabed aspect, seabed curvature and substrata (mud, sand, boulder and bedrock)

	Variance	Proportion	Rank
Total	15.23	1.00	
Constrained	1.00	0.07	9
Unconstrained	14.23	0.93	47

abundance distribution following a lognormal distribution (Connolly et al., 2014). Equally, biotic processes can be linked to the abundance of co-occurring species thereby tempering the assumption of independence used here. Inclusion of a log normal function to better structure the random abundances between species could provide a more realistic representation of a typical marine community. Furthermore, the use of a more realistic community structure, through the inclusion of a log normal distribution across the simulated species, could highlight other characteristics intrinsic to the SACFOR scale, such as how important levels of information are captured between common and rare species at the point of classification.

Much of the variance within the biological data could not be explained by the environmental data - it is possible that aspects of the SACFOR coding (reducing abundance to a seven-point scale), taxonomic aggregation and possible variations in the survey design, apparatus or conditions experienced (e.g. visibility) between sites and stations introduced variation that obfuscates the environmental variables included in the analysis. Categorical data of species abundance has also been used to produce species distribution models, e.g. Mieszkowska et al. (2013) use of ACFOR observations to produce predicted species distributions for the trochid gastropods *Phorcus lineatus* and *Gibbula umbilicalis* at several points in time.

The SACFOR scale purposely lacks precision in order to provide accuracy for rapid surveys where species identification, access and time are issues. It is also better suited when the investigator is more interested in documenting rarer species (i.e. inventory surveys) rather than the quantitative analysis of commoner species, which is often conducted

Fig. 8 A plot of the correspondence analysis for observations from East of Haig Fras and Wyville Thompson Ridge with vectors showing the relative trajectory and importance of the environmental variables (the site factor is shown as 'X'). Labels are rug = rugosity, asp = aspect, cur = curvature, bou = boulder and bed = bedrock)

with quadrats through a stratified random approach, but more likely to miss rarer species unless heavily replicated. Eleftheriou & McIntyre (2005) suggest that SACFOR is inappropriate as a tool for monitoring as it is not sufficiently quantitative. However, we suggest that SACFOR data, that includes information on multiple taxa, in well replicated surveys from large areas of marine habitat, provides sufficient power that these data sets should be considered useful for monitoring studies in areas lacking quantitative observations. Despite this, the conversion of data generated from descriptive to analysable ordinal scales does not improve its precision, and its accuracy remains the same. Bearing this in mind we suggest that anyone adopting our methodology (or similar) should resist the temptation to over-analyse the data that it makes available and be mindful of the inherent limitations of the underlying data collection methodology. Indeed, as most marine ecological data collection techniques are only semi-quantitative, we should be ever mindful of the limitations of all data collected and wary of attributing unjustifiable accuracy when interpreting imprecise data.

Conclusions

It is hoped that the SACFOR conversion process proposed here facilitates: (i) the quantitative re-analysis of the burgeoning SACFOR data repository; and (ii) initiates a debate on alternative methods for the conversion of SACFOR data into analysable end products. The repository of existing SACFOR observations is vast and generally under-utilised. Equally, this repository contains repeated observations for several locations and an extensive array of habitat types and geographic locations. It is hoped that the conversion of historical SACFOR data into a format available for statistical analysis opens up a plethora of new re-analysis possibilities including temporal analysis, broad-scale spatial analysis as well as modelling and regression analyses. The objectives and content of this paper are simple and intuitive, i.e. that ordinal data can be substituted with numerical values. It is hoped that this study highlights the basic operations required to access and analyse a wealth of biological information that has accumulated over 27 years of survey work. The conversion presented here, if repeated, also provided consistent and objective conversion of SACFOR data, thereby allowing comparisons between studies and over time.

Authors' contributions
JS and MJ contributed equally to all sections of the analysis and manuscript preparation. Both authors read and approved the final manuscript.

Availability of data and materials
Please contact author for access to R scripts and the data used in the analysis derived from Marine Recorder.

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]National Oceanography Centre, University of Southampton Waterfront Campus, European Way, Southampton SO14 3ZH, UK. [2]Biology and Marine Science, University of Hull, Cottingham Road, Hull HU6 7RX, UK.

References
Arnaud PM, Galeron J, Arntz W, Petersen GH. Semi-quantitative study of macrobenthic assemblages on the Weddell Sea shelf and slope using trawl catch subsamples. In: Arntz W, Ernst W, Hempel I, editors. The expedition ANTARKTIS VII/4 (EPOS leg 3) and VII/5 of RV Polarstern in 1989, *Ber Polarforsch*, vol. 68; 1990. p. 98–104.

Ballantine WJ. A biologically-defined exposure scale for the comparative description of rocky shores; 1961. p. 1–19.

Braun-Blanquet J. *Plant Socoilogy*. Mcgraw-Hill book company, Inc; New York; London; 1932.

Braun-Blanquet J. Pflanzensociologie: Grundzüge der Vegetationskunde. 3rd ed. Springer Verlag: Vienna; 1964.

Burrows MT, Harvey R, Robb L. Wave exposure indices from digital coastlines and the prediction of rocky shore community structure. Marine Ecology Progress Series. 2008;353:1–12.

Connolly SR, MacNeil MA, Caley MJ, Knowlton N, Cripps E, Hisano M, Thibaut LM, Bhattacharya BD, Benedetti-Cecchi L, Brainard RE, Brandt A. Commonness and rarity in the marine biosphere. Proc Natl Acad Sci. 2014;111:8524–9.

Crisp DJ, Southward AJ. The distribution of intertidal organisms along the coasts of the English Channel. J Mar Biol Assoc U K. 1958;37(1):157–203.

Dahl E, Hadac E. Strandgesellschaften der Insel Ostøy im Oslofjord. Eine pflanzensoziologische studie. Nytt Magasin for Naturvidenskapene B. 1941;82: 251–312.

Digby PGN, Kempton RA. Multivariate analysis of ecological communities. London: Chapman and Hall; 1987.

Eleftheriou A, McIntyre A. Methods for the study of marine benthos: Blackwell Publishing; 2005.

Fischer-Piette J (1936) Etudes sur la biogeographie intercotidale des deux rives de la Manche, J. Linnean Society (Zoology), XL: 181-272.

Field A, Miles J, Field Z. Discovering statistics using R: Sage, London; 2012. p. 958.

Gray, J.S. & Elliott, M. (2009) Ecology of marine sediments: from science to management. OUP Oxford.

Guisan A, Broennimann O, Engler R, Vust M, Yoccoz NG, Lehmann A, Zimmermann NE. Using niche-based models to improve the sampling of rare species. Conserv Biol. 2006;20:501–11.

Hawkins SJ, Jones HD. Rocky shores (Vol. 1). Sea challengers; 1992.

Herbert RJH, Hawkins SJ, Sheader M, Southward AJ. Range extension and reproduction of the barnacle Balanus perforatus in the eastern English Channel. J Mar Biol Assoc U K. 2003;83:73–82.

Herbert RJH, Southward AJ, Sheader M, Hawkins SJ. Influence of recruitment and temperature on distribution of intertidal barnacles in the English Channel. J Mar Biol Assoc U K. 2007;87:487–99.

Hiscock K. In situ survey of intertidal biotopes using abundance scales and checklists at exact locations (ACE surveys). Version 1 of 23 March 1998. In Biological monitoring of marine Special Areas of Conservation: a hand book of methods for detecting change. Part 2. Procedural guidelines (ed. K Hiscock), 3 pp. Peterborough, Joint Nature Conservation Committee. 1998.

Hiscock K. Marine Nature Conservation Review: methods. Joint Nature Conservation Committee, Peterborough, Nature Conservancy Council, CSD Report, No. 1072. (Marine Nature Conservation Review Report, No. MNCR/OR/5). 1990.

Jamieson S. Likert scales: how to (ab) use them. Med Educ. 2004;38:1217–8.

JNCC. The Marine Habitat Classification for Britain and Ireland Version 15.03. [15/01/2020]. 2015. https://mhc.jncc.gov.uk/

Legendre P, Legendre LFJ. *Numerical Ecology*: Elsevier; 2012.

Mieszkowska N, Kendall MA, Hawkins SJ, Leaper R, Williamson P, Hardman-Mountford NJ, Southward AJ. Changes in the range of some common rocky shore species in Britain—a response to climate change? *Marine Biodiversity*. Dordrecht: Springer; 2006a. p. 241–51.

Mieszkowska N, Leaper R, Moore P, Kendall MA, Burrows MT, Lear D, Poloczanska ES, Hiscock K, Moschella P, Thompson RC, Herbert RJH. Marine biodiversity and climate change: assessing and predicting the influence of climatic change using intertidal rocky shore biota. Occasional Publication of the Marine Biological Association. 2006b;20.

Mieszkowska N, Milligan G, Burrows MT, Freckleton R, Spencer M. Dynamic species distribution models from categorical survey data. J Anim Ecol. 2013;82:1215–26.

Miller AW, Ambrose RF. Sampling patchy distributions: comparison of sampling designs in rocky intertidal habitats. Mar Ecol Prog Ser. 2000;196:1–14.

Nelson-Smith A. Marine biology of Milford Haven: the physical environment: Field studies council; 1967.

Norman G. Likert scales, levels of measurement and the "laws" of statistics. Adv Health Sci Educ. 2010;15:625–32.

Podani J. Braun-Blanquet's legacy and data analysis in vegetation science. J Veg Sci. 2006;17:113–7.

Preston FW. The commonness, and rarity, of species. Ecology. 1948;29:254–83.

R Core Team. *R: A language and environment for statistical computing*. Vienna: *R Foundation for Statistical Computing*; 2013. URL http://www.R-project.org/

Rodwell, J.S. & Joint Nature Conservation Committee (2006). National vegetation classification: Users' handbook. Peterborough: Joint nature conservation committee. Available at: http://archive.jncc.gov.uk/pdf/pub06_NVCusershandbook2006.pdf (last Accessed 24/09/2019).

Stark JD. SQMCI: A biotic index for freshwater macroinvertebrate coded-abundance data. N Z J Mar Freshw Res. 1998;32:55–66.

Simkanin C, Power AM, Myers A, McGrath D, Southward A, Mieszkowska N, Leaper R, O'Riordan R. Using historical data to detect temporal changes in the abundances of intertidal species on Irish shores. J Mar Biol Assoc U K. 2005;85:1329–40.

Southward AJ, Crisp DJ. The distribution of certain intertidal animals around the Irish coast. In: *In Proceedings of the Royal Irish Academy. Section B: Biological, Geological, and Chemical Science*, vol. 57: Royal Irish Academy; 1954. p. 1–29.

Sullivan GM, Artino AR Jr. Analyzing and interpreting data from Likert-type scales. J Grad Med Educ. 2013;5:541–2.

Wang YI, Naumann U, Wright ST, Warton DI. mvabund–an R package for model-based analysis of multivariate abundance data. Methods in Ecology and Evolution. 2012;3(3):471–4.

Wheater CP, Bell JR, Cook PA. Practical Field ecology: A project guide: Wiley; 2011.

Publisher's Note