Method Article

Quantification of ecological complexity and resilience from multivariate biological metrics datasets using singular value decomposition entropy

Antoni Ginebreda[a,*], Laia Sabater-Liesa[a], Damià Barceló[a,b]

[a] *Department of Environmental Chemistry, Institute of Environmental Assessment and Water Research (IDAEA-CSIC), Jordi Girona 18-26, 08034, Barcelona, Spain*
[b] *Catalan Institute for Water Research (ICRA). Carrer Emili Grahit 101, 17003, Girona, Spain*

A B S T R A C T

The concept of resilience has become popular in many disciplines far beyond its original use in the field of ecology. Despite of its wide use, it has received different definitions not always coincident. Such ambiguity is still more evident in its quantitative characterization. Most of the available methods are heavily context dependent and often difficult to apply in the practice. Here, we propose to define and calculate resilience starting from the data matrices resulting from multivariate measurements of different biological metrics.

- The resilience between two field scenarios (each one characterized by their corresponding datasets) can be conveniently captured as the difference between its respective data complexities.
- Complexity is quantified by means of the entropy associated to the spectral distribution of the singular values of each data matrix.
- The method proposed has been illustrated with a case study in which the resilience of a river (Ebro River, NE Spain) is calculated comparing six biological metrics associated to the phytoplankton, upstream and downstream to a series of large reservoirs that alter the natural river flow regime.

---

* Corresponding author.
  E-mail addresses: agmqam@cid.csic.es (A. Ginebreda), laia.sabater@idaea.csic.es (L. Sabater-Liesa), Dbcqam@cid.csic.es (D. Barceló).

Specifications Table

Subject Area:	Environmental Science
More specific subject area:	Freshwater Ecology
Method name:	Singular Value Decomposition Entropy
Name and reference of original method:	Singular Value Decomposition Entropy
Resource availability:	The method has been evaluated using experimental data from Sabater-Liesa et al., 2019 [20]

Method details

The term 'resilience', appeared for the first time in the ecological science in 1973 [1], rapidly influenced other scientific domains such as engineering, economics, medicine or social sciences. Since that time many alternative definitions of resilience have been proposed [2–7]. For the purposes of this article, we will follow Holling's seminal concept [1], which refers to the capacity of an ecosystem to cope with changing external conditions without losing its structural and functional characteristics. Despite its broad use, the definitions and interpretations of resilience are still the matter of deep discussion in the literature [2–7], particularly when they need to be applied to specific case studies. Such difficulties are particularly challenging when resilience has to be quantified. Although there are many methods reported in the literature, in general, they all tend to be strongly context-dependent so that their application is only feasible for specific experiments or scenarios [2,3]. Therefore, there is a need for general methods of resilience quantification capable of broad application and suitable to be used in the common practice of field ecology.

Data gathered from environmental biological field monitoring typically consists of measurements of different variables spanning on space and time, that are conveniently organized in the form of data matrices. Extracting information from such data matrices is a problem usually addressed from multivariate statistics [8]. Among the plethora of techniques available, here we specifically focus on the singular value decomposition (SVD) technique (see details below), which is underlying in many of the existing methods broadly used in multivariate data analysis. Furthermore, SVD has been successfully applied in a large variety of scientific and technical domains ranging from signal and image processing, genomic analysis, weather forecast, chemometrics, disease surveillance or big-data analysis [9–15].

Here we are specifically interested in the characterization of the data complexity (organized in appropriate matrices of empirical measurements or derived metrics) which is assumed to quantitatively reflect the system's own complexity. In turn, the extent of changes in system complexity between two situations or scenarios of a given system is proposed as a general quantitative empirical metric of resilience [16]. To that end, we make use of the so-called SVD entropy, which captures how is the distribution of the singular values (SVs) of the data matrix analyzed (see Fig. 1). SVD entropy has found applications in a variety of areas like econometrics [9–11], genome expression data processing [12], image processing [13] or medical sciences [14,15].

Singular Value Decomposition of a data matrix (SVD)

Briefly, this technique consists of the decomposition of any A $(m \times n; m \geq n)$ matrix into a product of three matrices as:

$$A = U \cdot \Sigma \cdot V^T \tag{1}$$

where U is an $(m \times m)$ unitary matrix, Σ is an $(m \times n)$ rectangular diagonal matrix with non-negative real numbers on the diagonal, and V is an $(n \times n)$ real or complex unitary matrix and V^T denotes its transpose. The diagonal entries λ_i of Σ are known as the singular values of A. The columns of U and the columns of V are respectively called the left-singular vectors and right-singular vectors of A.
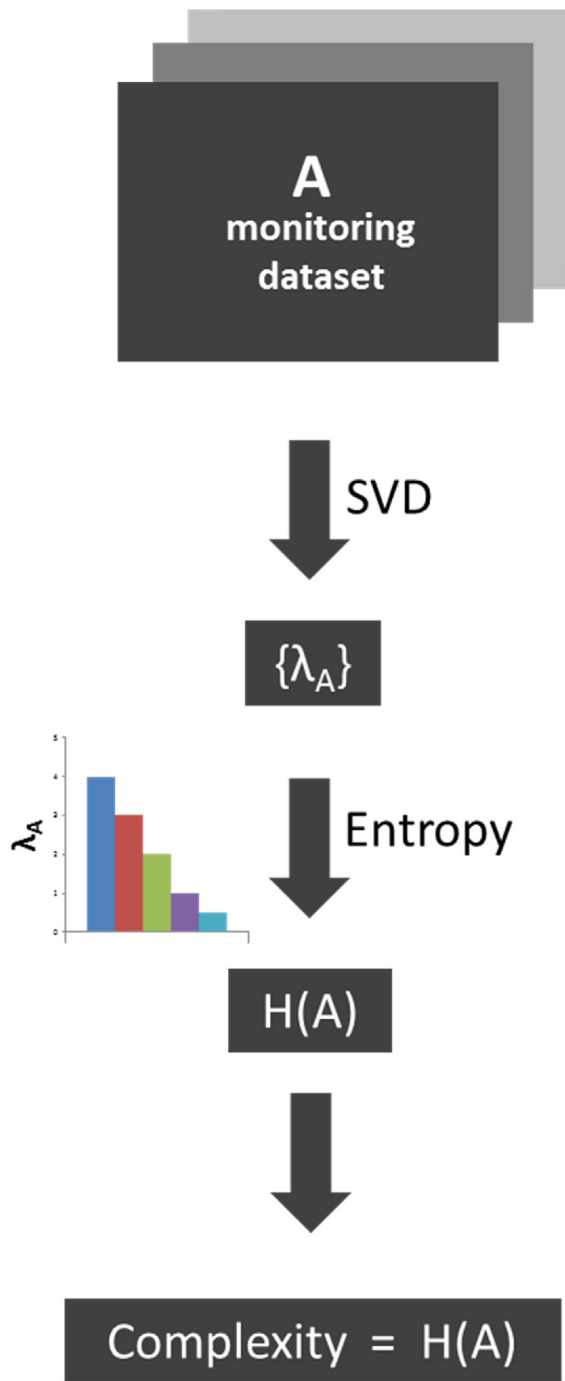
Fig. 1. Flow-chart overview of the method showing the connection between a dataset of measurements with its complexity quantification in terms of SVD entropy. The workflow includes the following steps: (a) SVD decomposition of the monitoring dataset matrix *A* which allows obtaining the set of singular values {λ_A} (Eqs. (1) and (2)); (b) Calculation of the SVD entropy *H(A)* (Eq. (4)) that is assimilated to the dataset complexity.

SVD entropy

Following [10] it is possible to define a complexity measure of the dataset contained in matrix *A*, using the set of singular values $(\lambda_i)_{i=1,n}$ by means of a suitable 'Shannon type entropy' [17] (Fig. 1). To do so, we first arrange the singular values $(\lambda_i)_{i=1,n}$ in decreasing order and normalize them so that:

$$\overline{\lambda}_i = \frac{\lambda_i}{\sum_i \lambda_i} \tag{2}$$

with $\Sigma_i \overline{\lambda}_i = 1$ \hfill (3)

The SVD Entropy of *A* denoted as *H(A)* is thus defined as:

$$H(A) = -\sum_{i=1}^{n} \overline{\lambda}_i \cdot \ln(\overline{\lambda}_i) \tag{4}$$

For comparison purposes between matrices having different dimensions, *H(A)* is conveniently normalized dividing by the factor ln*(n)* which corresponds to the maximum value attainable by *H(A)*. In this way, *H(A)* is bounded between 0 and 1:

$$H(A) = \frac{-1}{\ln(n)} \sum_{i=1}^{n} \overline{\lambda}_i \cdot \ln(\overline{\lambda}_i) \tag{5}$$

Fig. 2 shows two hypothetical examples of singular values distribution with their respective entropies calculated using Eq. (5).

SVD entropy and resilience

For a given variable, the resilience quantification proposed here involves comparing two related scenarios, each one characterized by its corresponding data matrix, using SVD entropy in terms of increase/decrease of the dataset complexity. i.e., a lower entropy reflects a non-uniform distribution of the singular values λ_i thus corresponding to low-complexity of the underlying data; conversely, higher SVD entropy denotes that the set of λ_i is more evenly distributed (Fig. 2).
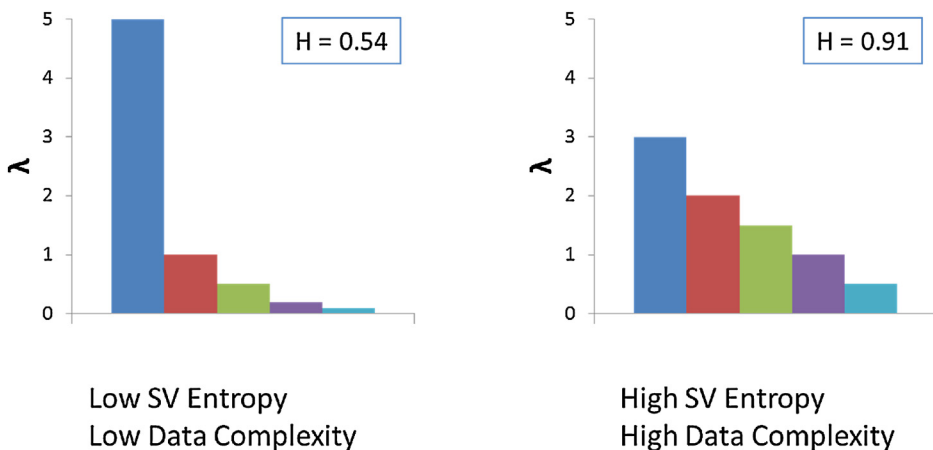

Fig. 2. Two hypothetical distributions of singular values (arbitrary scale) highlighting a low and high entropy (complexity) profiles and their respective entropies (calculated using Eq. (5)).

A system that is able to maintain its complexity after a perturbation will be qualified as 'resilient', while the opposite behavior would be indicative of a lack of resilience. Let us considered a system in two states A and B, each one characterized by the corresponding matrices of measurements or metrics of their respective variables. The difference in complexity between states A and B of such a system (expressed as the corresponding difference on SVD entropies) can be related to the system's resilience. Since high resilience is associated with low changes in data complexity, a suitable and general measure of resilience can be conveniently captured by the following equation:

$$Resilience\ (A, B) = 1 - |H(A) - H(B)| = 1 - |\Delta H| \tag{6}$$

Since H is always comprised between 0 and 1, this resilience index is comprised between 0 and 1 too. Resilience equals 1 if $H(A) = H(B)$ corresponding to a lack of change in complexity between A and B scenarios, and thus to a maximum resilience. Conversely, if $H(A) = 1$ and $H(B) = 0$ (or the opposite) then resilience becomes 0 thus reflecting a maximum change in complexity between the scenarios compared. The whole process is summarized in Fig. 3.
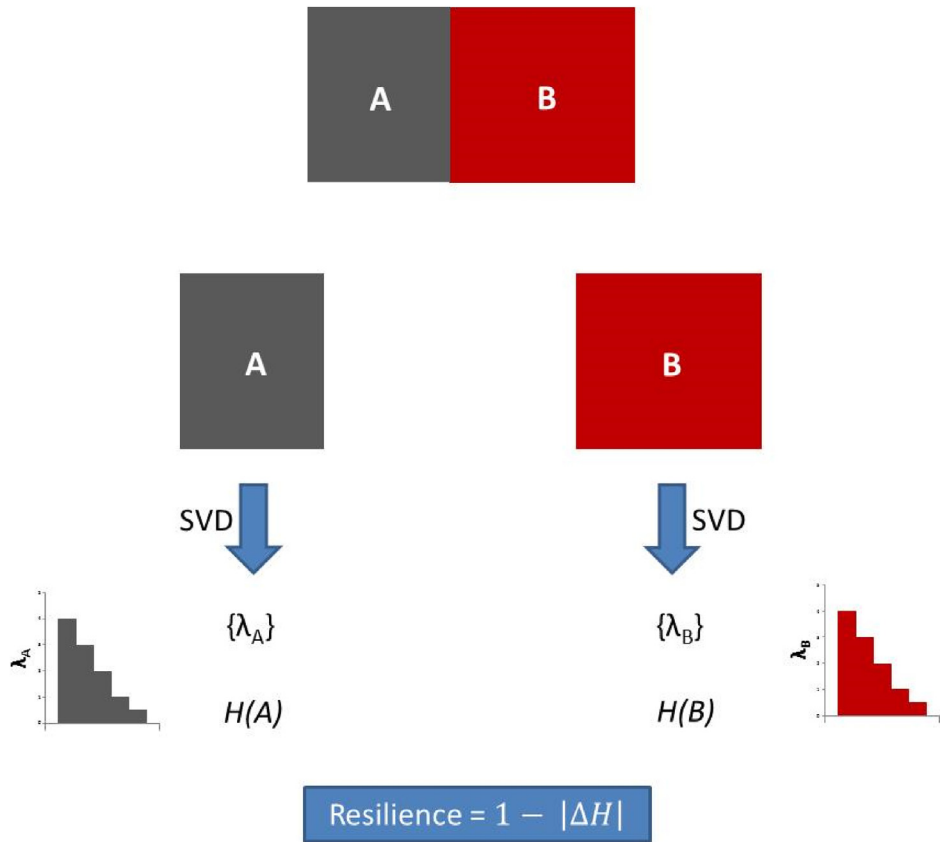

Fig. 3. Flowchart for the calculation of resilience, comparing the variation of entropy (complexity) between two related data matrices. The workflow includes the following steps: (a) SVD decomposition of the two monitoring dataset matrices A and B, which allows obtaining the sets of singular values $\{\lambda_A\}$ and $\{\lambda_B\}$ (Eqs. (1) and (2)); (b) Calculation of the SVD entropies $H(A)$ and $H(B)$ (Eq. (4)) that are assimilated to the respective datasets complexities; (c) Calculation of the system's resilience (Eq. (6)).

Method validation using a case study

The foregoing method was tested in a stretch of the Ebro River basin (NE Spain). The Ebro basin is located in the Northeastern part of the Iberian Peninsula occupying a total surface of 85362 km^2. The main river is 910 km length and flows from the Cantabrian Mountains to the Mediterranean Sea. In terms of water flow the Ebro River is the largest one in the Iberian Peninsula (mean annual discharge 435 m^3 s^{-1}). The middle course of Ebro mainstream is affected by three consecutive large reservoirs, Mequinenza (1500 Hm^3), Riba-roja (210 Hm^3) and Flix (11 Hm^3) [18,19], causing major changes in the hydromorphological dynamics (flood peaks alteration, retention of sediments, etc.) that are reflected on the ecological status of the river. The purpose of our exercise aimed at quantifying the system resilience comparing the data measured upstream and downstream to the reservoirs.

Biological data used in the present study were published elsewhere [20–23]. Twelve sites located in the mid-lower course from Zaragoza to the proximity of the river mouth were selected (Fig. 4). The first six sites were located upstream to the reservoirs, while the remaining were downstream. Six biological variables related to the phytoplankton were considered. They included metrics related to the algal community structure (Shannon-Wiener diversity, number of species, cell density, biovolume, and chlorophyll-*a* concentration) and function (alkaline phosphatase activity, APA). Datasets used can be found in [23].

We constructed two dataset matrices for every measured biological variable, for the upstream and downstream sites respectively. Every matrix is constituted by a table of *sites* $\times$ *time*. These are handled as rectangular matrices of *m* columns (*m*: number of spatial sites) and *n* rows (*n*: number of campaigns). The method outlined above was applied to each of the six metrics considered. The main results are summarized in Table 1 and Figs. 5 and 6.
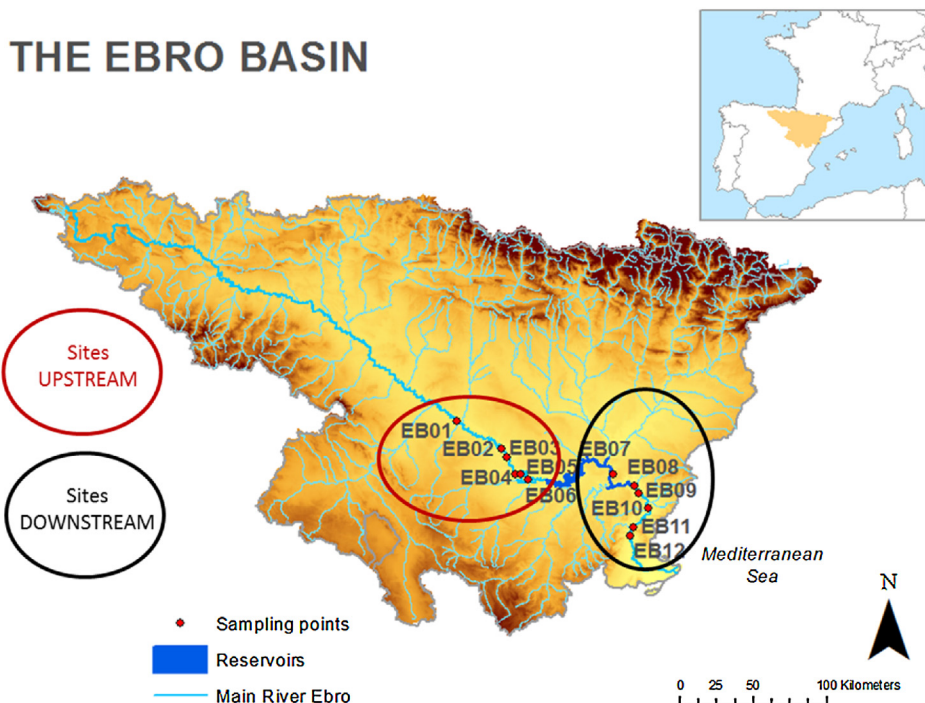
Fig. 4. Area of study: The Ebro river middle course, showing the sampling points located upstream and downstream to the reservoirs.

Table 1

Results of complexity (singular value entropy), resilience (complexity maintenance), for the biological metrics considered upstream (UP) and downstream (DOWN) to the reservoirs.

METRICS	SVD-Entropy (Complexity)		Resilience
	H (UP)	H (DOWN)	$1 - |\Delta H|$
APA	0.638	0.389	0.751
Biovolume	0.685	0.538	0.853
Cell Density	0.521	0.457	0.936
Chlorophyll-a	0.525	0.723	0.802
Diversity	0.596	0.562	0.967
Number of Species	0.520	0.590	0.930
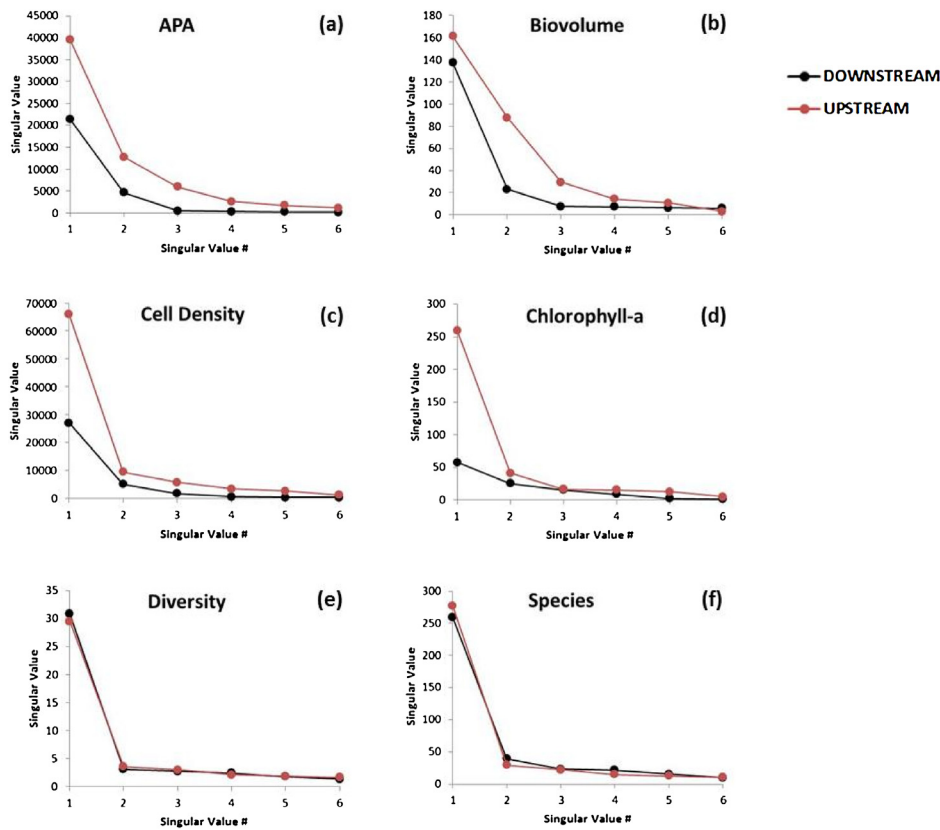
Fig. 5. Distribution of the singular values for the biological metrics considered, upstream and downstream to the reservoirs. Note that the scales of the vertical axis are different for each biological metrics.

The spectra or distribution of the singular values used in the calculation of the entropies of the biological metrics considered is shown in Fig. 5. The entropies, calculated using Eq. (5), had values in the range 0.38–0.72 (i.e., 38% to 72% of its maximum value). Four out of the six variables measured (all except chlorophyll-a and, in a less extent, the number of species) exhibited higher entropy (complexity) in the sites located upstream to the reservoirs (and thus subjected to a more natural hydrologic regime) than those located downstream (regulated regime) (Table 1, Fig. 5). Resilience was
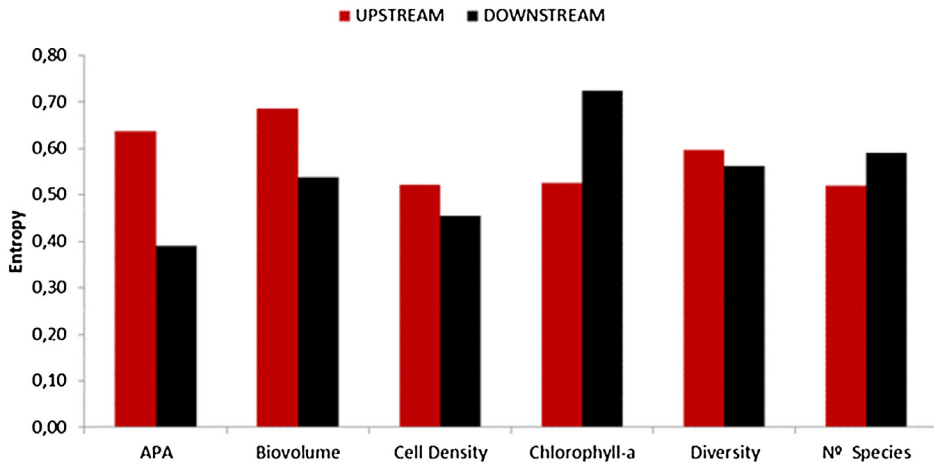
Fig. 6. Comparison of entropy (complexity) values between upstream and downstream sites for the different biological metrics studied.

quantified using Eq. (6) for the six biological metrics studied. Values obtained were in the range 0.75–0.97, that correspond to APA and diversity respectively. The high resilience values obtained for diversity (0.967) and the number of species (0.930) is also perceptible from the tight closeness of the singular values distribution for the upstream and downstream as shown in Fig. 5e and f. Altogether, the medium to high resilience values quantified indicates that the system is likely capable to recover its complexity after the perturbation caused by the reservoirs, at least for the six variables examined. A deeper discussion and interpretation of the results can be found in [23].

In summary, the foregoing example highlights the generality and broad applicability of the proposed method of resilience quantification consisting of comparing the complexity of two data blocks (matrices) in terms of their respective singular value entropy.

Acknowledgments

References

[1] C.S. Holling, Resilience and stability of ecological systems, Annu. Rev. Ecol. Syst. 4 (1) (1973) 1–23.
[2] I.H. Myers-Smith, S.A. Trefry, V.J. Swarbrick, Resilience: easy to use but hard to define, Ideas Ecol. Evol. 5 (2012).
[3] L. Todman, et al., Defining and quantifying the resilience of responses to disturbance: a conceptual and modelling approach from soil science, Sci. Rep. 6 (2016) 28426.
[4] E. Desjardins, et al., Promoting resilience, Q. Rev. Biol. 90 (2) (2015) 147–165.
[5] L. Xu, D. Marinova, X. Guo, Resilience thinking: a renewed system approach for sustainability science, Sustain. Sci. 10 (1) (2015) 123–138.
[6] D. Hodgson, J.L. McDonald, D.J. Hosken, What do you mean,'resilient'? Trends Ecol. Evol. 30 (9) (2015) 503–506.
[7] P.J. Mumby, et al., Ecological resilience, robustness and vulnerability: how do these concepts benefit ecosystem management? Curr. Opin. Environ. Sustain. 7 (2014) 22–27.
[8] P. Legendre, L. Legendre, Numerical Ecology, 2nd english edition, Elsevier, Amsterdam, 1998.
[9] P. Caraiani, The predictive power of singular value decomposition entropy for stock market dynamics, Phys. A Stat. Mech. Appl. 393 (2014) 571–578.
[10] R. Gu, Y. Shao, How long the singular value decomposed entropy predicts the stock market?—Evidence from the Dow Jones industrial Average Index, Phys. A Stat. Mech. Appl. 453 (2016) 150–161.
[11] R. Gu, W. Xiong, X. Li, Does the singular value decomposition entropy have predictive power for stock market?—Evidence from the Shenzhen stock market, Phys. A Stat. Mech. Appl. 439 (2015) 103–113.

[12] O. Alter, P.O. Brown, D. Botstein, Singular value decomposition for genome-wide expression data processing and modeling, Proc. Natl. Acad. Sci. 97 (18) (2000) 10101–10106.
[13] R.A. Sadek, SVD based image processing applications: state of the art, contributions and research challenges, arXiv 3 (7) (2012) 26–34 preprint arXiv:1211.7102.
[14] S.-y. Li, et al., Analysis of heart rate variability based on singular value decomposition entropy, J. Shanghai Univ. (Engl. Ed.) 12 (5) (2008) 433.
[15] A. Sabatini, Analysis of postural sway using entropy measures of signal complexity, Med. Biol. Eng. Comput. 38 (6) (2000) 617–624.
[16] L. Parrott, Measuring ecological complexity, Ecol. Indic. 10 (6) (2010) 1069–1076.
[17] C.E. Shannon, A mathematical theory of communication, Bell Syst. Tech. J. 27 (3) (1948) 379–423.
[18] J. Prats, R. Val, J. Armengol, J. Dolz, Temporal variability in the thermal regime of the lower Ebro River (Spain) and alteration due to anthropogenic factors, J. Hydrol. 387 (2010) 105–118.
[19] A.M. Romaní, S. Sabater, I. Muñoz, The physical framework and historic human influences in the Ebro River, in: D. Barceló, M. Petrovic (Eds.), The Ebro River Basin, Springer, Berlin Heidelberg, Berlin, Heidelberg, 2011, pp. 1–20.
[20] J. Artigas, et al., Phosphorus use by planktonic communities in a large regulated Mediterranean river, Sci. Total Environ. 426 (2012) 180–187.
[21] S. Sabater, et al., Longitudinal development of chlorophyll and phytoplankton assemblages in a regulated large river (the Ebro River), Sci. Total Environ. 404 (1) (2008) 196–206.
[22] E. Tornes, et al., Reservoirs override seasonal variability of phytoplankton communities in a regulated Mediterranean river, Sci. Total Environ. 475 (2014) 225–233.
[23] L. Sabater-Liesa, A. Ginebreda, D. Barceló, Shifts of environmental and phytoplankton variables in a regulated river: a spatial-driven analysis, Sci. Total Environ. 642 (2018) 968–978.