Research

Author for correspondence:
Helge Aufderheide
e-mail: he.adh@biond.org

Royal Society **Publishing**
Informing the science of the future

How to predict community responses to perturbations in the face of imperfect knowledge and network complexity

Helge Aufderheide[1,2], Lars Rudolf[2], Thilo Gross[2] and Kevin D. Lafferty[3]

[1]Max Planck Institute for the Physics of Complex Systems, Dresden, Germany
[2]Merchant Venturers School of Engineering, University of Bristol, Bristol, UK
[3]US Geological Survey, Western Ecological Research Center, Marine Science Institute, University of California, Santa Barbara, CA 93106, USA

Recent attempts to predict the response of large food webs to perturbations have revealed that in larger systems increasingly precise information on the elements of the system is required. Thus, the effort needed for good predictions grows quickly with the system's complexity. Here, we show that not all elements need to be measured equally well, suggesting that a more efficient allocation of effort is possible. We develop an iterative technique for determining an efficient measurement strategy. In model food webs, we find that it is most important to precisely measure the mortality and predation rates of long-lived, generalist, top predators. Prioritizing the study of such species will make it easier to understand the response of complex food webs to perturbations.

1. Introduction

Predicting the result of environmental perturbations, such as the arrival of new species, is a major goal in ecology [1–3]. What makes this challenging is the complex interconnected nature of ecological systems. In any densely connected system, a perturbation of one element can percolate across the network of interactions. This is particularly true for the complex food webs that form the backbones of most ecosystems [4–7]. Even perturbations acting on a small subset of species may propagate through the network and lead to serious systemic changes [3,6,8,9] such as the destruction of native fish populations following the introduction of a new species [10].

A central factor determining the response of a food web to perturbations is its topology, the precise map of predator–prey interactions. It has been shown that topological properties affect local and global dynamical stability [11–15]. Moreover, the food web topology gives an indication of the relative importance of species when studying notions of robustness such as the likelihood of secondary extinctions [16–19].

Topology alone, however, is not sufficient for reliable predictions of perturbation effects [4,20–22]. Taking into account biomass flows between the different species leads to better results [23]. For more detailed predictions, numerical simulation of mathematical models is often used [21,22].

An accurate prediction of the impact of a perturbation requires information about underlying biomass flows and the control coefficients characterizing the nonlinearity of processes. Such parameters require extensive measurements, and errors in their estimation quickly reduce the accuracy of predictions [4,24]. A lack of precise information on biomass flows and control coefficients thus limits the predictability of responses. For instance, for systems of more than 25 species, predictions are practically impossible with current methods, unless very detailed information is available [25].

Here, we ask whether food web responses to a perturbation are more sensitive to particular species or parameters [17,26,27]. We investigate the predictability of responses to perturbations in a broad class of food web models. Our results show that all parameters and species do not need to be measured with the same

accuracy. We use analytical calculations and numerical demonstrations to show that it is possible to assign to each species a value that indicates the importance of precise knowledge about this species for the quality of the prediction. Finally, we demonstrate that this importance can be estimated reasonably well from imprecise information, and explore correlations of a species' importance and its ecological properties.

The paper is structured as follows: we start in §2 by introducing a method for predicting the impact of given perturbations in a broad class of food web models. The method is illustrated in §3 with two examples. In §4, we then derive measures for species' influence on others and for their sensitivity to perturbations. In §5, we test these predictions in a series of numerical experiments. The numerical results illustrate a feasible strategy for field studies, where mathematical analysis and experimental measurements are used to iteratively improve predictions about the response of food webs to disturbance. In §6, we use computer experiments and statistical association to determine which parameters and types of species are most important to measure.

2. Impact evaluation

(a) Derivation of the perturbation impact

Consider a biological system described by a set of state variables $X_1, \ldots, X_N$ denoting, for instance, the abundances of established species in a food web. The system is now subjected to a perturbation characterized by another set of variables $Y_1, \ldots, Y_M$, for instance the abundances of newly arriving species.

We assume that, in the absence of the perturbation, the variables $X_1, \ldots, X_N$ are governed by a set of ordinary differential equations of the form

$$\frac{\mathrm{d}}{\mathrm{d}t} X_i = A_i(X_1, \ldots, X_N), \tag{2.1}$$

where A_i is a function representing the *right-hand side* of the differential equations. The generalized model for food webs [28] used here describes the dynamics of the populations $X_1, \ldots, X_N$ by N differential equations of the form $\mathrm{d}X_i/\mathrm{d}t = G_i(X) + S_i(X_i) - L_i(X) - M_i(X_i)$, where G_i, L_i, M_i and S_i are unspecified functions describing, respectively, the gain by predation (G_i), the loss by predation (L_i), the loss owing to natural mortality (M_i) and the gain by primary production (S_i) of the focal species.

Following Novak *et al.* [25], we consider the case where the unperturbed system resides in a stable equilibrium X^* and where the perturbation is characterized by a small and constant Y^*, such as new species persisting at a low constant abundance in the ecosystem owing to initially positive growth or constant influx. These new species affect the right-hand side of (2.1).

Because the stationary abundance, X_i^*, of a given established species i is dependent on the new species Y^*, we can regard it as a function $X_i^* = X_i^*(Y^*)$. We then define the impact $I_{i,j}$ of a perturbation variable Y_j^* on a resident species abundance X_i^* as the change of X_i^* per unit Y_j^*, i.e.

$$I_{i,j} = \left.\frac{\partial X_i^*}{\partial Y_j^*}\right|_0, \tag{2.2}$$

where $|_0$ indicates that the derivative is evaluated in the limit of vanishing densities of the arriving species Y_j^*. In other words, the entries of the impact matrix $I_{i,j}$ state the change

of units of the established species i per unit of arriving species j that enters the system.

In small systems, the impact can be computed by first defining the model functions, then solving (2.1) for the stationary solution $X_i^*(Y^*)$, and subsequently computing the derivative in (2.2). However, for more than three species, the analytical computation typically becomes prohibitively difficult.

Computing the stationary solution can be avoided by recognizing that the stationary density of a resident species X_i^* can be considered as an implicit function that is defined as the solution of the stationarity condition $0 = A_i(X^*, Y^*)$. Using a corollary to the implicit function theorem [29], we can then write the impact matrix as

$$\mathbf{I} = -\mathbf{J}^{-1}\mathbf{K}, \tag{2.3}$$

where the superscript -1 indicates the matrix inverse. The matrix $\mathbf{J}$ is the so-called Jacobian, which is defined as the derivatives of A_i with respect to the abundances of established species [30], i.e. $J_{i,j} = \partial A_i/\partial X_j|_*$, where $|_*$ indicates that the derivative is evaluated in the equilibrium under consideration. The matrix $\mathbf{K}$ is defined by $K_{i,j} = \partial A_i/\partial Y_j|_*$, which captures the direct impact of an arriving species j on an established species i. For instance, this direct impact may occur due to a reduction in production, an increase in mortality or an increase in predation gain of the established species per unit of the arriving species. To establish $\mathbf{K}$ prior to the arrival therefore requires information about the resident species with whom the arriving species is likely to interact.

We note that, beyond the example of perturbations caused by an arriving species, (2.3) applies to press perturbations on an established community in general. For instance, to estimate the impact of a slight drought, the entry K_i of the perturbation vector indicates the direct impact of the drought on species i; mathematically, we can approximate $K_i = \Delta A_i(X^*)$, where $\Delta A_i(X^*)$ denotes the absolute change of A_i owing to the drought.

In summary, (2.3) establishes a relationship between the direct proximal impact of a press perturbation, and the indirect ultimate impact $\mathbf{I}$. As a note of caution, we remark that (2.3) holds up to linear order. The impact-approximation therefore remains valid only as long as the perturbation caused is reasonably small.

(b) Parametrization of J

In the equations shown above, we refer to the steady state of the system, which seems to imply that information about this state is required. However, relationship (2.3) remains valid independently of the specific steady state under consideration. When the matrices are evaluated, the steady state appears only in the Jacobian, which contains elements of the form $J_{i,j} = \partial A_i/\partial X_j|_*$. For instance, in the generalized food web model (see the electronic supplementary material for a review), this leads to expressions such as $\partial G_i/\partial X_i|_*$ [28]. Because we cannot evaluate this expression without further assumptions, it is an unknown quantity. However, we note that for any specific system the expression is simply a number. In other words, this means that the unknown derivatives appearing in the Jacobian constitute unknown parameters of the model.

So far, we have recognized that the unknown derivatives can be formally treated as unknown parameters. However, as such, these parameters are hard to interpret and are thus not suitable for an ecological discussion. We solve this problem by using a slightly different parametrization, which is obtained

Table 1. Generalized model parameters as defined in Gross & Feudel [28]. For numerical simulations, the turnover rates α_i are scaled by the species' fitness n_i corresponding to the niche value used to generate the underlying topology [35], while the other parameters are drawn from the indicated ranges.

name	interpretation	value(s)
scale parameters: defining the biomass flows in the steady state		
α_i	rate of biomass turnover in species i	$\alpha_i = 0.08^{n_i}$
$\beta_{i,j}$	contribution of predation by i	0 (no feeding)
	to loss rate of species j	[0.1,1] (else)
$\chi_{i,j}$	contribution of species i	0 (no feeding)
	to the prey of species j	[0.1,1] (else)
ρ_i	fraction of growth in species i	0 (producers)
	gained by predation, not primary production	1 (consumers)
σ_i	fraction of mortality in species i	0 (top pred.)
	resulting from predation, not mortality	[0.5,1] (others)
elasticities: sensitivities of interactions to state variables		
γ_i	sensitivity of predation in species i to is prey density	[0.5,1.5]
$\lambda_{i,j}$	exponent of prey switching	1 (passive)
μ_i	exponent of closure in species i	[1,2]
ϕ_i	sensitivity of primary production in species i to the density of species i	[0,1]
ψ_i	sensitivity of predation in species i to the density of predators	[0.5,1.5]

either by a special normalization procedure [28] or directly by the identity $\partial G_i/\partial X_i|_* = G_i^*/X_i^* \, \partial(\log G_i)/\partial(\log X_i)|_*$, which is true for $G_i^*, X_i^* > 0$ (a condition that is generally met by definition; see Kuehn *et al.* [31] for $X_i^* = 0$).

The factors G_i^*/X_i^* and $\partial(\log G_i)/\partial(\log X_i)|_*$ have a direct interpretation in most applications. The first is a *per capita* rate. Such rates have the dimension of inverse time and can be directly interpreted as characteristic turnover rates, i.e. the *per capita* growth rate of the members of species i by predation on other species.

The second factor is a logarithmic derivative. Such derivatives are also called elasticities and have been proposed in economic theory [32], metabolic control theory [33] and ecology [34]. They can be estimated well from observational data and interpreted straightforwardly. For every power-law, $f(x) = Ax^p$, the logarithmic derivative is $\partial \log f/\partial \log x = p$, independently of A or x. Thus, for instance, any linear function has an elasticity of one, regardless of the slope. For functions that are not power-laws, the elasticity still provides an intuitive nonlinear measure of the sensitivity in the steady state.

Each partial derivative of the process in the steady state is thus the product of two constant factors, describing the *per capita* rate and the sensitivity of the process, respectively. These factors are well-defined ecological parameters in their own right, which can be understood and discussed even if the steady state of the system is unknown. For food webs, this parametrization leads to a Jacobian matrix shown in the electronic supplementary material [28] and parameters given in table 1. Using (2.3), the Jacobian that is thus parametrized can then be used to relate a perturbation to its eventual impact.

(c) Remarks

We note that the approach in this paper is closely related to that of Novak *et al.* [25]. Our main methodological contribution is to apply this approach to generalized models. The advantage of generalized modelling is its high numerical efficiency, which enables a detailed and statistically sound numerical exploration. For the practical application to real-world food webs, generalized models offer additional advantages. In contrast to half-maximum concentrations and maximal growth rates used in conventional models, all parameters of the generalized model are defined in the state observed in nature. The parameters can therefore be measured directly without requiring a fitting procedure. Furthermore, the parameters are defined in such a way that their estimation from noisy data converges maximally fast [32,36].

The formulation of the generalized model is straightforward (see the electronic supplementary material for a short review). Based on Gross & Feudel [28], or with a parametrization algorithm [37], even large models with tens to hundreds of species can be set up in few hours. The impact of different species can then be computed in seconds on a small laptop, using a simple algorithm [37]. Once the model has been set up, integration of new data requires entering new numerical values. The computation of impact and importance therefore presents only a small additional effort to the fieldwork needed to measure parameters.

3. Examples

For illustration, we consider two examples: a simple predator–prey system and the real-world food web of Gatun Lake shown in figure 1. In the predator–prey system, a predator of abundance X_1 consumes a producer of abundance X_2. A detailed treatment and discussion of the stability of this system in terms of the generalized model parameters can be found in references [28,38].

The Jacobian matrix of this system near the steady state is

$$\mathbf{J} = \begin{pmatrix} \alpha_1(\psi - \mu) & \alpha_1 \gamma \\ -\alpha_2 \sigma \psi & \alpha_2(\phi - \sigma\gamma - (1-\sigma)\mu) \end{pmatrix}, \tag{3.1}$$
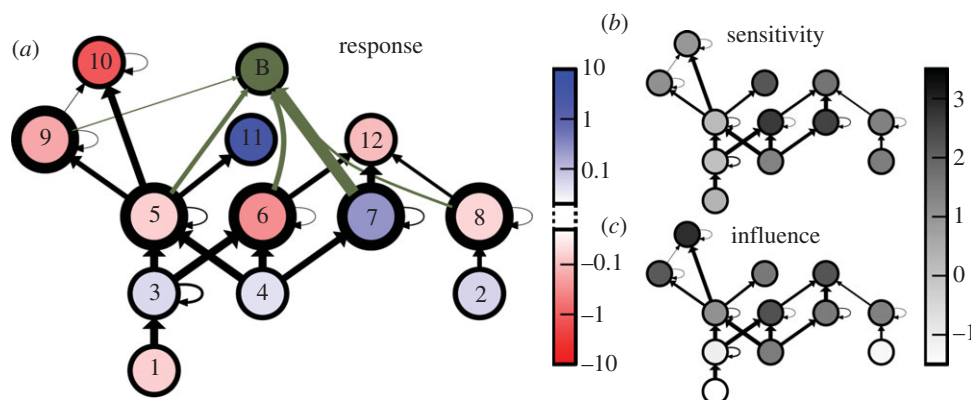
Figure 1. Predicted response of the Gatun Lake food web [10] to the perturbation caused by the introduction of the peacock bass. Shown are species (circles) and predator–prey relationships (arrows, in the direction of the biomass flow). In (a), the peacock bass population (B) is shown, together with its feeding relations. The colour code denotes the response of each species to this perturbation for a specific set of generalized parameter values (see the electronic supplementary material for details). In (b) and (c), the grey scale denotes the sensitivity and the influence of each species, respectively, approximating its average propensity to be impacted by other species or to impact to other species (details are given in §4 and the electronic supplementary material). (Online version in colour.)

where α_i represents each species' turnover rate, σ the relative loss of the producer owing to predation (instead of natural mortality), ϕ the elasticity (i.e. sensitivity) of primary production to the producer abundance, γ the elasticity of predation to primary producer abundance, ψ the elasticity of predation to predator abundance and μ the elasticity of natural mortality to a species' own abundance.

We now consider the impact of the arrival of a competing predator. It can be assumed that the new predator has a direct negative effect on the primary producer but no *direct* effect on the established predator, such that the perturbation matrix **K** contains the entries $K_{1,1} = 0$ and $K_{2,1} < 0$. As shown in detail in the electronic supplementary material, the impact on the established predator and producer is

$$I_1 = \frac{\alpha_1}{\det \mathbf{J}} \gamma K_2 \quad \text{and} \quad I_2 = \frac{\alpha_1}{\det \mathbf{J}} (\mu - \psi) K_2, \tag{3.2}$$

where $\det \mathbf{J} \geq 0$ is the determinant of **J**.

We see that generally the impact on the established predator is negative. This result is intuitive as the established predator is now in exploitative competition with the arriving predator. Of particular interest is the case where the established predator suffers from linear loss ($\mu = 1$) and also affects the producer linearly ($\psi = 1$, i.e. there is no interference between the predators). In this limit, $\det \mathbf{J} = \alpha_1 \alpha_2 \sigma \gamma$. The impact on the producer, I_2, is zero, because $\psi - \mu = 0$. However, the impact on the established predator, $I_1 = K_2/(\alpha_2 \sigma)$, is negative and corresponds directly to the fraction of producer biomass, consumed by the new predator. This is a manifestation of the well-known competitive-exclusion principle, which generally precludes the coexistence of the predators in this case [39,40].

The assessment of impact in larger food webs can be carried out analogously, but requires numerical computations in which the generalized parameters are set to specific values. Here, we illustrate this assessment by the example of the Gatun Lake food web [10] (figure 1) for a simple set of such parameters. We focus on the impact of a predatory fish, such as the peacock bass *Cichla ocellaris* (figure 1a). A step-by-step breakdown of the parametrization is presented in the electronic supplementary material.

The proposed approach predicts that peacock bass have a strong and generally negative impact on the secondary consumers on which they feed, a generally positive impact on the consumer's prey, and a generally negative impact on other top predators. These observations are consistent with basic ecological reasoning. Counterintuitive results are found for species 11 (black tern), which benefits from the strong decrease of its competing predators 9 (bigmouth sleeper) and 10 (tarpon) and for 7 (sailfin molly, mosquito fish), which benefits from the reduction of 6 (tetras) with whom 7 is both in exploitative and apparent competition.

Real-world observations of the Gatun Lake showed [10] that the introduction of the peacock bass strongly decreased the secondary consumers (5–7) and their predators (9–12), but increased the consumers prey (3–4) and, counterintuitively, species 8 (blackbelt cichlid). Comparing the predictions of the model with the real data, we note that the model correctly predicted the change in the producers 1–4, the decrease in the consumers 5 and 6 and the decrease in predators 9, 10 and 12.

The model predictions do not agree with the observed decrease in species 11 (black tern), and 7 (sailfin molly, mosquito fish) and with the increase in species 8 (blackbelt cichlid). A likely reason for the discrepancy between predicted and observed results is the simple parametrization used here, which we based on allometric scaling relationships rather than on direct observations. Reviewing the results, we can guess that the incorrect prediction regarding species 11 is due to an overestimation of the sensitivity of its competing top predators (9–10) to the decrease in their common prey 5 (silverside). Further, the incorrect prediction of the responses of 7 and 8 might result from incorrectly estimating the sensitivity of 12 (herons and kingfishers) to its prey (6–8). This exercise shows how food webs often follow the basic logic of direct and indirect effects, but that the details can be sensitive to incorrect parametrization.

4. Sensitive and influential species

To obtain better theoretical predictions for real-world systems, such as Gatun Lake, a more precise parametrization of the model is highly desirable. However, not all species are of equal importance in the food web and thus also precise information is more important for certain parameters than for

others. To gain insights into the importance of various species in the web, we now identify *sensitive species* which are easily perturbed by disturbances propagating through the web and *influential species* which have a strong effect on other species, when directly perturbed.

Close to a steady state, the dynamical properties of a system are characterized by its dynamical modes that consist of the eigenvectors and eigenvalues of the Jacobian matrix [30]. For each eigenvalue λ_k of a given matrix, there is generally a corresponding right eigenvector $v_{(k)}$ and left eigenvector $w_{(k)}$ [41].

One can visualize dynamical modes as vibrations travelling through a drum when it is struck. Here, the different modes correspond to different notes that are played on the drum. The right eigenvectors characterize the pattern of vibration when a specific note is played. Specifically, the elements of the right eigenvector describe how strongly the respective area of the drum vibrates in that note. The same is true for the food web. In a stable steady state that is hit by a short (pulse) perturbation the right eigenvectors govern how the system returns to the steady state after the perturbation.

Drummers know how to play different notes by striking different parts of the drum. This is captured by left eigenvectors. Specifically, the elements of the left eigenvector for a given dynamical mode describe how strongly the specific mode is excited when the drum is struck in a given area. Similarly, in the food web, the left eigenvectors characterize the strength of a specific dynamical response when a given species is perturbed.

Intuitively, one can think of each dynamical mode as a possible response of the system to a perturbation. The right eigenvector denotes the impact of response (which species 'feel the vibrations'), whereas the corresponding left eigenvector denotes the type of perturbation that can trigger a particular response (which species needs to be perturbed to 'play a given note'). For instance, consider the pair of a right eigenvector $v = (1,2)$ and a left eigenvector $w = (1,0)$. This mode can be excited only through perturbation of the first species, but when it is excited the second species feels the system's response twice as strong as the first.

The strength of a mode's response is determined by its *excitability*. The excitability is, for a given mode, given by the corresponding eigenvalue $-1/\lambda_k$ of the inverse Jacobian matrix $\mathbf{J}^{-1}$, which has the same eigenvectors as $\mathbf{J}$. Intuitively, the (negative) eigenvalue λ_k of a dynamical mode indicates a system's resistance to a particular perturbation. The impact of such a perturbation is therefore inversely proportional to this resistance (see the electronic supplementary material for more details).

In the case presented here, we consider that a perturbation continuously excites the same dynamical modes (press perturbation). The impact is therefore the combined continuous excitation resulting from the perturbation of these dynamical modes.

The potential impact that a species experiences owing to a given dynamical mode is the product of the mode's excitability and the component of the right eigenvector on this species. Furthermore, the potential impact from all modes is the sum over the contributions from the individual modes k. Taking the logarithm of this sum to bring the numerical values into a more manageable range, we therefore define the *sensitivity* of a species, $\mathrm{Se}_i = \log(-\sum_k |v_i^{(k)}|/\lambda_k)$, where $|v_i^{(k)}|$ is the absolute value of the entry v_i of the right
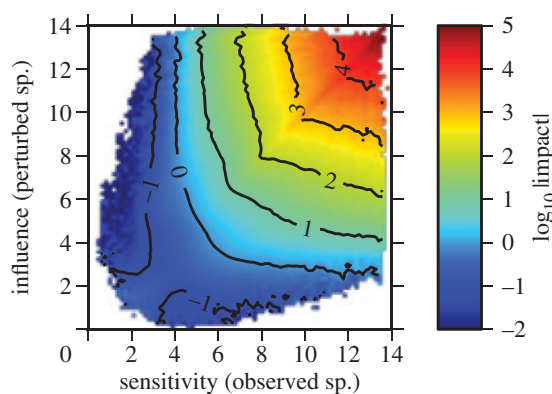

Figure 2. The average absolute impact on a species with given sensitivity if a species of given influence is perturbed. Generally, the more sensitive the species being tracked, and the more influential the species that is increasing, the greater the change (impact) to the observed species. Other parameters $N = 50$, $C = 0.04$, see the electronic supplementary material for $C = 0.08$. (Online version in colour.)

eigenvector corresponding to mode k. For a more formal derivation, see the electronic supplementary material.

The potential impact that a species causes by exciting a given dynamical mode is the product of the mode's excitability and the component of the left eigenvector on this species. Analogous to the sensitivity, we therefore define the *influence* $\mathrm{In}_i = \log(-\sum_k |w_i^{(k)}|/\lambda_k)$.

The Gatun Lake example shows that the general results for influence and sensitivity are largely consistent with the predicted response to the introduction of peacock bass (figure 1). Because the perturbation affects several influential nodes (e.g. species 6 and 9), most of the species of the food web are affected. In particular, sensitive nodes (e.g. species 6, 7, 11) respond strongly to the perturbation (see the electronic supplementary material for more details).

To further confirm the relationship between importance, sensitivity and actual impact of a perturbation, we consider an ensemble of 10^6 randomly generated food webs with 50 species and average connectance 0.04 generated as described in Gross *et al*. [14] and as reviewed in the electronic supplementary material. The topology of these food webs is generated using the niche model [35]. Then, for the numerical computations, the generalized parameters characterizing the species are drawn uniformly and independently from the ranges given in table 1. For these food webs, the colour code in figure 2 indicates the average impact that a focal species of given sensitivity experiences when a species of given importance is perturbed. This reveals a strong correlation of the impact with both the sensitivity of the focal species and the importance of the perturbed species.

Considering the ensemble of model food webs again, we observe that the number of the very influential and very sensitive species in each food web is small. For instance, we find that on average for each web only 15.4% of all species have a sensitivity value in the upper 30% of the sensitivity range for this web, and only 18.6% have an importance value in the upper 30% of the influence range for that web.

Summarizing the above, knowledge of the Jacobian of a specific food web enables us to predict the impact of specific perturbations and also allows us to gain a more general understanding of the species' sensitivity and influence with regard to perturbations of the network. The main challenge for impact assessment is thus to collect the necessary data for constructing the system's Jacobian. As hinted previously,

and shown below, precise measurement is required only for some species.

5. Iterative parameter estimation

It is intuitive to assume that accurate predictions hinge on precise measurements of the parameters of the most influential and sensitive species. However, our approach to identifying these species builds on analysis of the Jacobian and thus itself requires the same type of information. It is thus not possible to determine *a priori* which species are important or influential. To address this dilemma, we now propose an iterative strategy in which existing preliminary information is used to estimate the impact and sensitivity of species. This assessment can be used to improve parameter estimates on seemingly important species. Once additional data on these species becomes available, they can be used to further improve the estimates of the impact and sensitivity of species, refining the process. Thus, a cycle is formed in which the necessary information for precise impact predictions is iteratively assembled.

We explore the quality of impact predictions in a series of numerical experiments. In each experiment, we predict the impact of a random perturbation to a food web that is generated according to the procedure described in §4. This procedure determines values of the generalized parameters of the *true* Jacobian of the food web by drawing them uniformly from the ranges indicated in table 1. Based on this true Jacobian, we additionally generate an *estimated* Jacobian with slightly different generalized parameter values to simulate measurement errors. More precisely, we draw each generalized parameter value used in the estimated Jacobian from a lognormal distribution centred on the corresponding parameter value used in the true Jacobian; the lognormal distribution is chosen to allow large errors while keeping the sign of parameters consistent. We then compute the true impact of the random perturbation, I based on the true Jacobian, and the estimated impact $\tilde{I}$ based on the estimated Jacobian. The quality Q of the impact estimation is then evaluated as the cosine of the angle between the true and the estimated impact vectors (see the electronic supplementary material for details), i.e. $-1 \leq Q \leq 1$.

Now, we introduce a numerical implementation of the iterative strategy described earlier. We consider numerical experiments in which the knowledge of the Jacobian is initially poor, such that the generalized parameters are drawn from a lognormal distribution with a standard deviation of 10% of the true value. We furthermore assume that additional empirical work can be carried out on specific species that reduces the error in all parameters of the respective species to 2%. Our aim is to carry out the precise measurements in the order that leads to the most rapid increase in the quality of impact prediction.

For the purpose of demonstration, we consider four different protocols: (i) precise measurements are carried out in random order, (ii) species are measured in the order of decreasing influence, (iii) species are measured in the order of decreasing sensitivity and (iv) species are measured in the order of the decreasing sum of sensitivity and influence. The choice of species to measure next is always based on the *estimated* Jacobian that is available at the time. Thus, the information used is only what researchers would observe under real-world conditions.
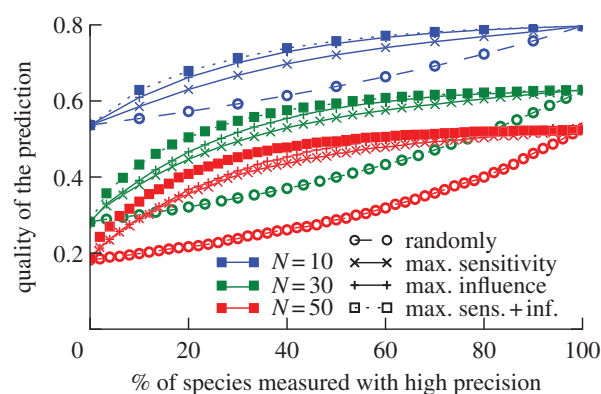

Figure 3. Focusing on sensitive and important species can reduce the measurement effort when species are measured successively with higher precision. Starting on the left, only low-precision information is available for all food web species. Advancing to the right, the measurement error is reduced for one species at a time until all nodes have been measured with high precision. The different curves refer to different strategies for selecting which species to measure precisely. For the dashed line (empty symbols), the species are selected randomly. For the solid curves, we first evaluate the sensitivity or influence of each species based on current knowledge and then select the species with the highest value of either sensitivity or influence. For the dotted curves we select the species with the highest product of influence and sensitivity ($Se_i + In_i$). Error-carrying parameters of each node: α, μ, ψ, ϕ, γ, σ, β, χ. Other parameters: initial error of each parameter 10%, final error 2%, connectance of $C = 0.04$. Higher connectance (see the electronic supplementary material) results in similar graphs, but the overall prediction quality decreases. (Online version in colour.)

The results shown in figure 3 demonstrate that estimating influence and sensitivity of the species prior to each measurement strongly increases the accuracy of predictions. This is particularly pronounced if measurements focus on the species with the highest sum of sensitivity and influence. For instance, after measuring 20% of all species according to this protocol, we attain a quality of prediction comparable to measuring 60–80% of all species when species are chosen randomly. Using the estimation of influence and sensitivity to focus observational or experimental efforts can thus significantly reduce the amount of empirical work that is needed to achieve a given prediction quality.

6. Most important parameters and species to measure

The iterative refinement procedure proposed above needs some initial information on the system as a starting point. Here, we therefore explore what types of parameters and what types of species are most important to measure.

To get an intuition of the importance of different parameters for impact prediction, we consider a situation where the estimated Jacobian is identical to the true Jacobian except for a single parameter that carries an error. The quality of the estimated impact decreases with increasing influence and sensitivity of the species affected independently by the varied parameter (figure 4). Furthermore, figure 4 shows that the decrease in quality for sensitive and influential species depends on the parameter under consideration; precision in the elasticity of the mortality μ, and of the
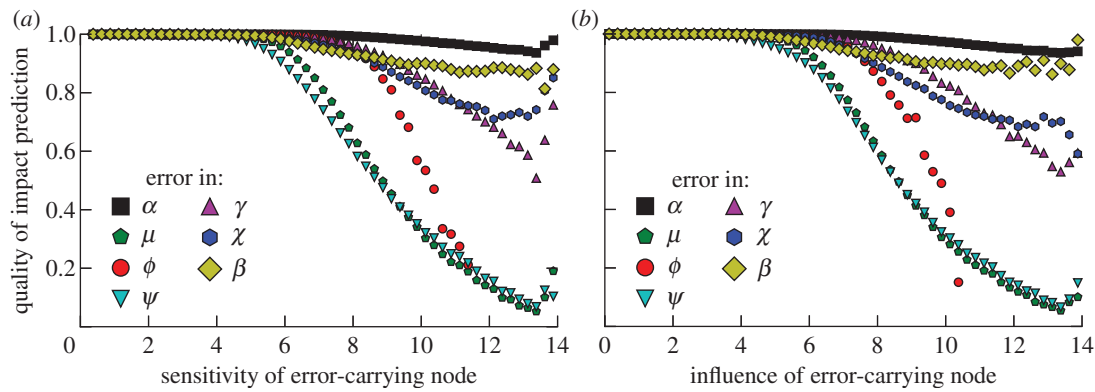
Figure 4. Average quality of an impact prediction in the presence of measurement errors. The average quality of an impact prediction if one node with the specified sensitivity (*a*) or influence (*b*) in a food web is subjected to a measurement error. The different datasets refer to errors in different parameters (cf. table 1). Other parameters are system size $N = 50$, connectance $C = 0.04$ and the standard deviation of the relative measurement error 10%. For higher C, the effect of χ and β increases (see the electronic supplementary material). Noise at high importance and sensitivity values is due to the relatively rare occurrence of these values in the numerical experiments. (Online version in colour.)

elasticity of predation ψ with respect to predator abundance were the most important.

To determine which species are most important to measure in the absence of knowledge about the Jacobian, we explore the correlations between sensitivity or influence and species properties (see below) in a set of 10^6 model food webs. In the analysis, we consider the correlations (not causal effects) of sensitivity and influence with the following potential biological indicators:

— generality, or the number of prey species of the focal species (network in-degree);
— vulnerability, or the number of predators of the focal species (network out-degree);
— a binary value that is 1 if the focal species is a primary producer and 0 otherwise;
— the trophic level (TL), which we calculate by solving a set of linear equations, such that $\mathrm{TL}_i = 1$ for primary producers and $\mathrm{TL}_i = 1 + \mathrm{mean}(\mathrm{TL}_{\mathrm{prey}})$ for consumers, where $\mathrm{mean}(\mathrm{TL}_{\mathrm{prey}})$ denotes the mean TL of *i*s prey;
— the biomass turnover rate (generalized model parameter α), indicating the amount of biomass an individual consumes in comparison with its own mass; and
— the weighted topological importance of a species (WI^s) introduced in Jordán *et al.* [42]. It estimates the indirect effect on a species, based on topology and biomass flows. Parameter *s* indicates the maximum number of direct interactions, through which indirect effects are perceived.

If for a given web, x_i indicates one of these properties for species *i* in this web, and y_i its sensitivity or influence, then we denoted the correlation coefficient between x and y as $R = \overline{(x_i - \bar{x})(y_i - \bar{y})}/(s_x s_y)$, where $\bar{x}$ denotes the mean of x_i over all species *i*, and where s_x denotes the standard deviation of the x_i.

The correlation analysis (figure 5) shows that high TLs and low biomass turnover rates (e.g. long lifespan) correlate strongly with sensitivity and influence. This result suggests that top predators and other large and long-lived species, despite their typically small total biomass, play a disproportionate role in the systems' response to perturbations. This corresponds well with observations of real-world systems [10,27,43]. When no specific information on biomass
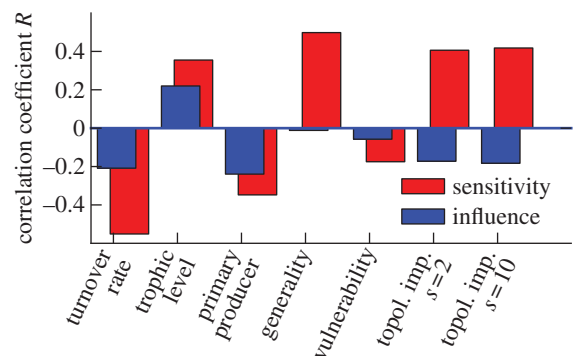

Figure 5. Correlations of species' properties with their sensitivity and influence. The pattern of correlations is consistent with large predators playing an important role for a system's response to perturbations. Parameters $N = 50$, $C = 0.04$, results are similar for higher connectance (see the electronic supplementary material). (Online version in colour.)

flows is available, these species should be targeted for initial parametrization.

The sensitivity of a species is highly correlated with its generality, whereas its influence appears to be independent of its generality. Intuitively, this can be interpreted as generalist species being sensitive to all of their prey species, while having relatively little impact on those species.

The weighted topological importance correlates strongly with sensitivity and weakly with influence. This suggests that indirect effects from proximal species in the web play generally a large role in the response, whereas indirect effects from distal species are relatively minor.

To improve perturbation assessments in the Gatun Lake ecosystem, fieldwork should prioritize better measurements of species 6 (tetras), 7 (sailfin molly, mosquito fish), and, to a lesser degree, the large predators 10–12 that have the highest combined sensitivity and influence (figure 1). Even, if the explicit sensitivity and influence values were not available, the correlation would suggest to prioritize these large predators, because of their low turnover rates, and their high generality and TLs. Furthermore, the most significant improvements should be obtained by focusing in the measurement process on the elasticities of mortality μ and of predation ψ of these species.

7. Conclusion

Previous work has suggested that without near-perfect information on a range of parameters, it may be intractable to predict the effects of perturbations to large, complex systems [25]. In this paper, we proposed a method to predict the impact of perturbations on complex systems more efficiently. We used this method to investigate the relative importance of different species in food webs. We found that there are typically a small number of species that are highly important, because they are sensitive to perturbations, have a strong influence on others, or both.

The proposed method is based on the linear stability of steady states in food webs. Strictly speaking, it therefore describes the consequences of small perturbations to systems close to a state of stable species coexistence. However, linear stability is often found to agree well with other stability criteria for food webs, such as robustness against noise [44], or permanence which measures the boundedness of a trajectory in a plausible part of the state space [45]. One can therefore expect that our methods give at least some indications about perturbation consequences and key species in systems that are not in a steady state.

While we have focused exclusively on food webs, we note that the same approach can likewise be applied to other networks of nonlinear interactions that are found in metabolism [46], gene regulation [47] and cellular population dynamics [48].

The presented results suggest that the potential impact of environmental perturbations on food webs can be predicted with reasonable accuracy if the most relevant parameters for only a small number of important species in the web are measured well.

Finding the important species in a food web appears as the key to a good and efficient impact assessment. We propose to find these important species by (i) pre-selecting species based on their biological properties, and (ii) applying an iterative refinement method once some initial information is available.

Our correlation analysis suggests that it is most important to obtain precise parameter estimates for long-lived, generalist consumers at high trophic levels. Furthermore, our analysis suggests that for these species, it is most important to precisely estimate the dependence of their mortality and predation on their abundance.

References

1. Sakai AK et al. 2001 The population biology of invasive species. Annu. Rev. Ecol. Syst. 32, 305–332. (doi:10.1146/annurev.ecolsys.32.081501.114037)

2. Abrams PA. 1996 Evolution and the consequences of species introductions and deletions. Ecology 77, 1321–1328. (doi:10.2307/2265529)

3. Parker I et al. 1999 Impact: toward a framework for understanding the ecological effects of invaders. Biol. Invasions 1, 3–19. (doi:10.1023/A:1010034312781)

4. Montoya JM, Woodward G, Emmerson MC, Solé RV. 2009 Press perturbations and indirect effects in real food webs. Ecology 90, 2426–2433. (doi:10.1890/08-0657.1)

5. Mooney HA, Cleland EE. 2001 The evolutionary impact of invasive species. Proc. Natl Acad. Sci. USA 98, 5446–5451. (doi:10.1073/pnas.091093398)

6. Mack MC, Antonio CMD. 1998 Impacts of biological invasions on disturbance regimes. Trends Ecol. Evol. 13, 195–198. (doi:10.1016/S0169-5347(97)01286-X)

7. Pascual M, Dunne JA. 2005 Ecological networks: linking structure to dynamics in food webs. New York: Oxford University Press.

8. Clavero M, Garcia-Berthou E. 2005 Invasive species are a leading cause of animal extinctions. Trends Ecol. Evol. 20, 110. (doi:10.1016/j.tree.2005.01.003)

9. Menge BA. 1995 Indirect effects in marine rocky intertidal interaction webs: patterns and importance. Ecol. Monogr. 65, 21–74. (doi:10.2307/2937158)

10. Zaret TM, Paine RT. 1973 Species introduction in a tropical lake: a newly introduced piscivore can produce population changes in a wide range of trophic levels. Science 182, 449–455. (doi:10.1126/science.182.4111.449)

11. May RM. 1973 Qualitative stability in model ecosystems. Ecology 54, 638–641. (doi:10.2307/1935352)

12. McCann KS. 2000 The diversity–stability debate. Nature 405, 228–233. (doi:10.1038/35012234)

13. Martinez ND. 1991 Artifacts or attributes? Effects of resolution on the little rock lake food web. Ecol. Monogr. 61, 367–392. (doi:10.2307/2937047)

14. Gross T, Rudolf L, Levin SA, Dieckmann U. 2009 Generalized models reveal stabilizing factors in food webs. Science 325, 747–750. (doi:10.1126/science.1173536)

15. Dambacher JM, Li HW, Rossignol PA. 2002 Relevance of community structure in assessing indeterminacy of ecological predictions. Ecology 83, 1372–1385. (doi:10.1890/0012-9658(2002)083[1372:ROCSIA]2.0.CO;2)

16. Dunne JA, Williams CV, Martinez ND. 2002 Network structure and biodiversity loss in food webs: robustness increases with connectance. Ecol. Lett. 5, 558–567. (doi:10.1046/j.1461-0248.2002.00354.x)

17. Jordán F. 2009 Keystone species and food webs. Phil. Trans. R. Soc. B 364, 1733–1741. (doi:10.1098/rstb.2008.0335)

18. Allesina S, Bodini A. 2004 Who dominates whom in the ecosystem? Energy flow bottlenecks and cascading extinctions. J. Theor. Biol. 230, 351–358. (doi:10.1016/j.jtbi.2004.05.009)

19. Libralato S, Christensen V, Pauly D. 2006 A method for identifying keystone species in food web models. Ecol. Model. 195, 153–171. (doi:10.1016/j.ecolmodel.2005.11.029)

20. Mills LS, Soulé ME, Doak DF. 1993 The keystone-species concept in ecology and conservation. Bioscience 43, 219–224. (doi:10.2307/1312122)

21. Berlow EL, Dunne JA, Martinez ND, Stark PB, Williams RJ, Brose U. 2009 Simple prediction of interaction strengths in complex food webs. Proc. Natl Acad. Sci. USA 106, 187–191. (doi:10.1073/pnas.0806823106)

22. Stouffer DB, Bascompte J. 2011 Compartmentalization increases food-web persistence. Proc. Natl. Acad. Sci. USA 108, 3648–3652. (doi:10.1073/pnas.1014353108)

23. Jordán F, Okey TA, Bauer B, Libralato S. 2008 Identifying important species: linking structure and function in ecological networks. Ecol. Model. 216, 75–80. (doi:10.1016/j.ecolmodel.2008.04.009)

24. Yodzis P. 1988 The indeterminacy of ecological interactions as perceived through perturbation experiments. Ecology 69, 508–515. (doi:10.2307/1940449)

25. Novak M, Wootton JT, Doak DF, Emmerson M, Estes JA, Tinker MT. 2011 Predicting community responses to perturbations in the face of imperfect knowledge and network complexity. Ecology 92, 836–846. (doi:10.1890/10-1354.1)

26. Klemm K, Serrano MÁ, Eguíluz VM, Miguel MS. 2012 A measure of individual role in collective dynamics. Sci. Rep. 2, 292. (doi:10.1038/srep00292)

27. Berg S, Christianou M, Jonsson T, Ebenman B. 2011 Using sensitivity analysis to identify keystone species and keystone links in size-based food webs. Oikos 120, 510–519. (doi:10.1111/j.1600-0706.2010.18864.x)

28. Gross T, Feudel U. 2006 Generalized models as a universal approach to the analysis of nonlinear dynamical systems. *Phys. Rev. E* **73**, 16205. (doi:10.1103/PhysRevE.73.016205)

29. Khalil H, Grizzle J. 1992 *Nonlinear systems*, 2nd edn. New York, NY: Macmillan Publishing Company.

30. Kuznétsòv YA. 1998 *Elements of applied bifurcation theory*, volume 112 of *applied mathematical sciences*. Berlin, Germany: Springer.

31. Kuehn C, Siegmund S, Gross T. 2013 Dynamical analysis of evolution equations in generalized models. *IMA J. Appl. Math.* **78**, 1051–1077. (doi:10.1093/imamat/hxs008)

32. Houthakker HS, Magee SP. 1969 Income and price elasticities in world trade. *Rev. Econ. Stat.* **51**, 111–125. (doi:10.2307/1926720)

33. Fell DA. 1992 Metabolic control analysis: a survey of its theoretical and experimental development. *Biochem. J.* **286**, 313–330.

34. Yeakel J, Stiefs D, Novak M, Gross T. 2011 Generalized modeling of ecological population dynamics. *Theor. Ecol.* **4**, 179–194. (doi:10.1007/s12080-011-0112-6)

35. Williams RJ, Martinez ND. 2000 Simple rules yield complex food webs. *Nature* **404**, 180–183. (doi:10.1038/35004572)

36. Lade SJ, Gross T. 2012 Early warning signals for critical transitions: a generalized modeling approach. *PLoS Comput. Biol.* **8**, e1002360. (doi:10.1371/journal.pcbi.1002360)

37. Aufderheide H. 2013 CIP - Calculator for the impact of perturbations and parametrizer for generalized models of food webs. See http://www.biond.org/content/programming-tools.

38. Gross T, Ebenhöh W, Feudel U. 2004 Enrichment and foodchain stability: the impact of different forms of predator–prey interaction. *J. Theor. Biol.* **227**, 349–358. (doi:10.1016/j.jtbi.2003.09.020)

39. Hardin G. 1960 The competitive exclusion principle. *Science* **131**, 1292–1297.

40. Gross T, Edwards AM, Feudel U. 2009 The invisible niche: weakly density-dependent mortality and the coexistence of species. *J. Theor. Biol.* **258**, 148–155. (doi:10.1016/j.jtbi.2009.01.018)

41. Wong SSM. 1997 *Computational methods in physics and engineering*, 2nd edn. Singapore: World Scientific.

42. Jordán F, Liu W-C, Davis AJ. 2006 Topological keystone species: measures of positional importance in food webs. *Oikos* **112**, 535–546. (doi:10.1111/j.0030-1299.2006.13724.x)

43. Moyle PB, Light T. 1996 Fish invasions in California: Do abiotic factors determine success? *Ecology* **77**, 1666–1670. (doi:10.2307/2265770)

44. Pimm SL. 2002 *Food webs*. Chicago, IL: University of Chicago Press.

45. Chen X, Cohen JE. 2001 Global stability, local stability and permanence in model food webs. *J. Theor. Biol.* **212**, 223–235. (doi:10.1006/jtbi.2001.2370)

46. Steuer R, Gross T, Selbig J, Blasius B. 2006 Structural kinetic modeling of metabolic networks. *Proc. Natl Acad. Sci. USA* **103**, 11 868–11 873. (doi:10.1073/pnas.0600013103)

47. Gehrmann E, Drossel B. 2010 Boolean versus continuous dynamics on simple two-gene modules. *Phys. Rev. E* **82**, 046120. (doi:10.1103/PhysRevE.82.046120)

48. Zumsande M, Stiefs D, Siegmund S, Gross T. 2011 General analysis of mathematical models for bone remodeling. *Bone* **48**, 910–917. (doi:10.1016/j.bone.2010.12.010)