

University of Groningen

The structure of marine benthic food webs

van Oevelen, Johannes

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version

Publisher's PDF, also known as Version of record

Publication date:

2006

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

van Oevelen, J. (2006). *The structure of marine benthic food webs: Combining stable isotope techniques and inverse modeling*. s.n.

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Chapter 2

Reconstructing food web flows using linear inverse models

Dick van Oevelen, Karline Soetaert, Jack J. Middelburg. *In preparation*

2.1 Introduction

The food web is a pivotal concept in ecology. It is the formal expression of the exchanges of energy or matter among living and non-living compartments in an ecosystem. The structure of food webs has important implications for ecosystem stability (De Ruiter et al., 1995; McCann et al., 1998), biogeochemical cycling (Berg et al., 2001), fishery harvest (Ryther, 1969) and exposure to toxic chemicals (Parmelee et al., 1993). The complex nature of food webs renders their study a challenge, which is taken up by means of field observations, experimental manipulation and modeling.

Food webs can be studied at various levels of abstraction. Topological studies analyze and generalize the patterns and statistics of the linkages within food webs (Pimm et al., 1991). Theoretical ecologists apply simple mechanistic models to study food web functioning. Classical food web paradigms, such as the paradox of enrichment (Rosenzweig, 1971) and competitive exclusion (Armstrong and McGehee, 1980) arise from these Lotka-Volterra models. The realism of the model predictions is assessed by qualitative comparison to empirical data (e.g. Fox and Olsen (2000) and Diehl and Feissel (2001)) or statistical testing (e.g. Scheffer (2001) and Van de Koppel et al. (2001)).

Such theoretical and experimental studies of simple food webs have provided the core of ecological theory. However, there is a strong need for a more quantitative approach of natural food webs (Cohen et al., 1993; Hall and Raffaelli, 1993; Johnson and Omland, 2004). This trend is apparent in population modeling (Kendall et al., 1999; Persson et al., 2004), stability analysis (De Ruiter et al., 1995), nutrient cycling studies (Vanni, 2002) and research on food limitation (Gaedke et al., 2002). In addition, ecological network analysis, in which descriptive indices are derived from food web matrices (Ulanowicz, 1986; Bersier et al., 2002), relies on accurate estimates of food web flows.

Advances in our knowledge through quantifying food webs relies on the crucial merge of observations and models; the area of inverse data assimilation techniques. Data assimilation refers to data incorporation within a model structure and inverse means that

field observations (e.g. primary production) are used to reconstruct the underlying model parameters (e.g. food web flows). Inverse methods are highly valued in geophysical sciences (Menke, 1984; Sun, 1994; Wunsch, 1996; Kasibhatla et al., 1999; Lary, 1999; Wang et al., 2000), where data inferences can only be made indirectly. The essentials of inverse solutions are treated extensively in these texts. They comprise 1) solution existence, 2) solution identifiability, 3) solution uniqueness and 4) solution sensitivity. These properties reveal important information of the system at hand.

Advanced ecological applications of data assimilation in food web models are mainly found in the marine realm (e.g. Prunet et al. (1996), Vallino (2000) and Friedrichs (2002)) and rely on fitting model parameters to observed time series or spatial distribution patterns, after which the food web flows are recovered. Such non-linear inverse problems are solved with complex numerical techniques (e.g. simulated annealing, adjoint or genetic algorithms), which require significant computer time, are complicated to implement and do not always retrieve the optimal parameter set (Lawson et al., 1996; Athias et al., 2000; Vallino, 2000; Soetaert et al., 2002). Moreover, additional reasons limit the use of these mechanistic modeling approaches in ecology (Gaedke, 1995). First of all, environmental sciences typically deal with poorly or partly understood mechanisms, complicating appropriate process descriptions. Second, the model outcome can be sensitive to closure terms (Steele and Henderson, 1992), i.e. terms that describe phenomena not explicitly accounted for. Third, parameter values are often not known accurately enough to determine a start range from which they can be fitted. Fourth, data sets are usually too sparse to constrain all parameters of a mechanistic model.

In contrast, simple linear inverse box models (LIM) have fewer data requirements and regard the food web as a linear system of flows that interconnect the compartments. The flows are quantified by solving a mass balance description supplemented with site-specific and literature data on biomass and flows. Their simplicity renders these LIM a widely applied inverse technique, also known as inverse analysis (e.g. Klepper and Van de Kamer, 1987; Vézina and Platt, 1988; De Ruiter et al., 1995). In addition, ECOPATH models (Christensen and Pauly, 1992) are based on similar mass balance modeling. Even though data requirements are limited as compared to fully mechanistic models, LIM problems are initially under determined; the available measurements are insufficient to uniquely determine all flows. Dealing with this non-uniqueness and under determinacy poses the major challenge in quantitative food web studies. Moreover, the analysis of the food web solution has received comparatively limited attention in the ecological literature.

In this paper we first develop a framework of LIM by explicit distinction of its elements (1) ecological information sources, (2) equations and (3) minimization norms. This framework is used to show that inconsistency in ecological LIM approaches are rooted in the way these elements are combined when solving the non-uniqueness and under determinacy problem. Secondly, 5 published LIM of food webs will be analyzed in terms of solution existence, identifiability, uniqueness and sensitivity. This solution analysis shows the current limitations and pitfalls of the inverse method. Finally, we present some recommendations to develop further LIM in ecology.

2.2 Linear inverse data assimilation in food web modeling

The crux of linear inverse data assimilation is to optimally merge ecological data within a food web model. This definition identifies three distinct elements of LIM: 1) ecological information, 2) linear model equations and 3) an optimization norm (Fig. 2.1). These elements will be discussed in the remainder of this section.

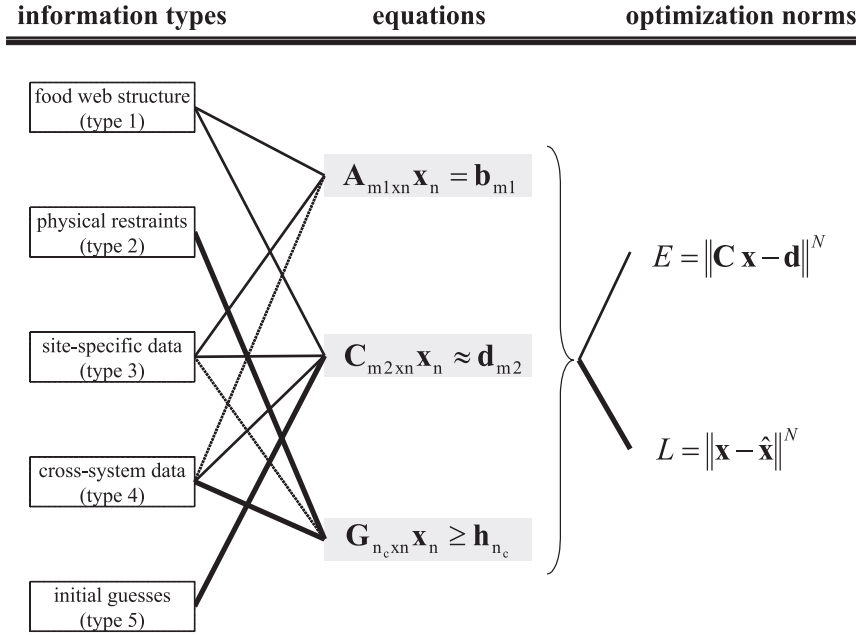


Figure 2.1: The three elements of linear inverse models in food web research: information types (see 2.2.1 Ecological information), equations (see 2.2.2 Model equations) and optimization norms (see 2.2.3 Optimization norms) and how they are combined in applications (connectors). The thickness of the connectors indicates the occurrence of a particular combination in the literature: dashed = few, solid = moderate and thick = many occurrences.

2.2.1 Ecological information

Ecological information sources relevant for food web modeling can be partitioned into 5 types, of which the first two are universal and the latter three are optional in LIM applications.

1. Food web structure consists of flows interconnecting the compartments and thereby defines a mass balance in terms of input (e.g. consumption) and output (e.g. respiration) for each compartment (Fig. 2.2).
2. Physical reality is information of physical nature, e.g. flows must be non-negative.
3. Site-specific observations are data on biomass, flows or combination of flows from the system under consideration. Typical flow measurements are production rates and community respiration.

4. Cross-system or literature data is information taken from other similar systems or from expert judgment. Typical data in this category are assimilation and production efficiencies.
5. Initial guesses are inferences about the possible values of the flows. Researchers may use measurements, literature (rate) or expert information to impose the most likely value of a flow. Initial guesses are also frequently used to find the simplest or parsimonious solution by setting them to zero. The parsimonious solution then is the solution that is closest to the initial estimates of zero.

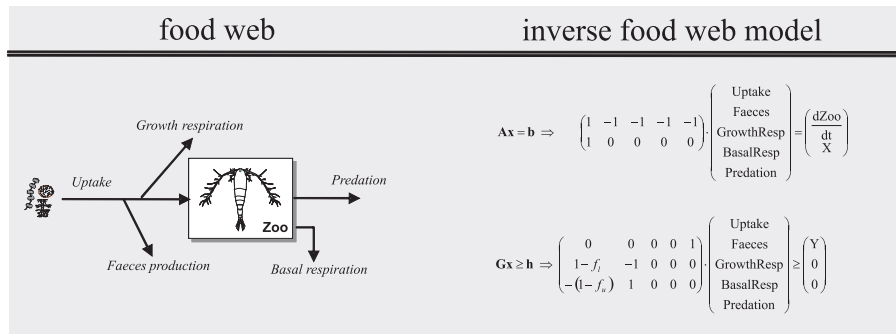


Figure 2.2: Example of a simple 'food web', implemented as a linear inverse model. Zooplankton (*Zoo*) take up algae; uptake is either lost as faeces, used for growth or growth respiration; other loss terms are predation and basal respiration. Mass balance is assured by the term accounting for the increase of biomass ($\frac{dZoo}{dt}$). X is the measured uptake by zooplankton, Y is the lower boundary imposed on the predation rate, f_l and f_u are the lower and upper bounds on the assimilation efficiency, respectively.

2.2.2 Model equations

Linear inversion allows for three types of equations: (1) a set of equations that must be satisfied exactly, (2) one that must be satisfied as closely as possible and (3) one set of inequalities (Fig. 2.1).

$$\mathbf{A}_{m1,n}\mathbf{x}_n = \mathbf{b}_m \tag{2.1}$$

$$\mathbf{C}_{m2,n}\mathbf{x}_n \approx \mathbf{d}_m \tag{2.2}$$

$$\mathbf{G}_{m_c,n}\mathbf{x}_n \geq \mathbf{h}_{m_c} \tag{2.3}$$

where the vector \mathbf{x} holds the n unknown flows, $m1$ is the number of equalities that must be satisfied exactly, $m2$ is the number of equalities that must be satisfied as closely as possible and m_c is the number of constraints. Matrices \mathbf{A} , \mathbf{C} and \mathbf{G} contain linear expressions of the flows and vectors \mathbf{b} , \mathbf{d} and \mathbf{h} hold numerical data (Fig. 2.2). The number of equations ($m1 + m2$) compared to the number of unknowns (n) determines the three possible states of a LIM: over determined ($m1 + m2 > n$), even determined ($m1 + m2 = n$) and under determined ($m1 + m2 < n$).

2.2.3 Optimization norms

In inverse theory two norms are distinguished: the prediction (or residual) norm and the solution norm. The prediction norm (E) measures the error between model predictions and observations

$$E = \| \mathbf{C}\mathbf{x} - \mathbf{d} \| ^N \quad (2.4)$$

whereas the solution norm (L) is a length measure based on the model parameters (flows)

$$L = \| \mathbf{x} - \hat{\mathbf{x}} \| ^N \quad (2.5)$$

where $\hat{\mathbf{x}}$ are the initial guesses of the flows and $\| \cdot \| ^N$ denotes some norm of power N , which generally takes the value of 1, 2 or ∞ (see below). The determinacy state principally determines which optimization norm is chosen to solve the LIM (see below); sometimes, the model solution is a compromise between both types of norms (Menke, 1984; Vézina and Platt, 1988). Vézina et al. (2004) have recently tested some optimization norms based on ecological theory (i.e. exergy and ascendancy) in a twin modeling experiment, but the results were ambiguous as no single norm was optimal under all circumstances.

Over determined model

Inevitable measurement noise implies that over determined models can never be solved with zero prediction error. Minimizing the prediction norm assures that the solution optimally agrees with all observations. The power of the prediction norm is related to the statistical distribution of the residuals. The L1 prediction norm ($N = 1$) weighs all deviations equally and is most appropriate in data sets that have a high probability of extreme data, e.g. when data are exponentially distributed (Menke, 1984). The L2 prediction norm is most commonly selected and assumes Gaussian statistics; larger deviations weigh more heavily on the norm. Successively higher powers of the norm give increasingly larger weight to the largest deviation, with the L_∞ prediction norm as most extreme case, i.e. only the equation that deviates most is optimized (Menke, 1984).

Measurement uncertainty can be used to assure that deviations from more accurate observations weigh heavier on the prediction norm. This is achieved by so-called row scaling (Wunsch, 1978; Menke, 1984; McIntosh and Rintoul, 1997)

$$\mathbf{S}_{m2 \times m2}^{1/2} \mathbf{C}_{m2 \times n} \mathbf{x}_n \approx \mathbf{S}_{m2 \times m2}^{1/2} \mathbf{d}_{m2} \quad (2.6)$$

where ideally \mathbf{S} is the error covariance matrix of the observations. As covariances are generally unknown, a diagonal matrix with the observation variance is used instead.

Even determined model

An even determined model has one unique solution that has zero prediction error.

Under determined model

An under determined model has infinite solutions with similar minimal prediction norm. Therefore, the model is solved by minimizing the prediction norm and the solution norm. When minimizing the solution norm, the power of the norm affects the weight that is given to the magnitude of the flow. Each flow weighs proportionally in the L1 solution norm, whereas large flows weigh more heavily in the L2 solution norm and with the L_∞ solution

norm only the largest flow is minimized. Following Vézina and Platt (1988), ecological applications typically solve parsimonious food webs (with initial guesses of unknown flows as zero, i.e. $\mathbf{x}_n = 0$) with the L2 solution norm. The solution is interpreted as the simplest food web satisfying mass conservation and biological constraints (Vézina and Platt, 1988). When a mixture of currencies (e.g. C and N) is used, scaling of the flows may be necessary, assuring equal weight in the solution norm (see e.g. Jackson and Eldridge, 1992). For a LIM solved with the L2 solution norm, this is achieved by (Wunsch, 1978)

$$\left(\mathbf{C}_{m2 \times m2} \mathbf{W}_{n \times n}^{-1/2} \right) \left(\mathbf{W}_{n \times n}^{1/2} \mathbf{x}_n \right) \approx \mathbf{b}_{m2} \quad (2.7)$$

which transforms to

$$\mathbf{C}'_{m2 \times n} \mathbf{x}'_n \approx \mathbf{b}_{m2} \quad (2.8)$$

from which the weighed \mathbf{x}' is solved. Subsequently, the unknowns flows are recovered with

$$\mathbf{x}_n = \mathbf{W}^{-1/2} \mathbf{x}'_n \quad (2.9)$$

Formally, \mathbf{W} is the error covariance matrix of the model parameters (Wunsch, 1978), but is usually taken to be a diagonal matrix containing column weights, for example C:N ratios to transform N flows to C equivalents (Jackson and Eldridge, 1992) or biomass such as to weigh the flows with stock size (Vézina and Savenkoff, 1999).

2.2.4 Combining the elements of LIM

The major challenge in LIM is the non-uniqueness and under determinacy of food webs. With 5 information types, 3 types of model equations and different optimization norms, there is great potential for variation in strategies to resolve these issues (Fig. 2.1). Based on this combination of LIM elements it is possible to categorize the various applications encountered in the literature.

Information type 1

Mass balances are based on the conservation of mass principle. Most applications assume that mass balances are in steady-state (i.e. input = output) over the investigated period and enforce this condition (Stone et al., 1993; Niquil et al., 1998; Diffendorfer et al., 2001; Leguerrier et al., 2003; Breed et al., 2004; Leguerrier et al., 2004; Richardson et al., 2004), whilst others allow small deviations from this condition, which are interpreted as changes in the stock over time (Vézina and Platt, 1988; Chardy et al., 1993; Vézina and Pace, 1994; Vézina and Savenkoff, 1999; Richardson et al., 2003). Other researchers start from non steady-state conditions, based on measured stock changes with time and require that the accumulation term in the mass balance is met exactly (Donali et al., 1999) or approximately (Vézina et al., 1997; Hart et al., 2000; Gaedke et al., 2002). Again others define the accumulation term as an additional unknown input or output flow to be solved in the LIM (Klepper and Van de Kamer, 1987, 1988; Jackson and Eldridge, 1992; Eldridge and Jackson, 1993; Lyche et al., 1996a,b; Vézina et al., 2000; Richardson et al., 2003).

Information type 2:

Physical constraints are always implemented in equation 2.3.

Information type 3

Site-specific observations are the most important data source for LIM as they allow discriminating the studied ecosystem from other systems. However, there is little consistency as into which equation this information is implemented. Niquil et al. (1998), Donali et al. (1999), Leguerrier et al. (2003), Breed et al. (2004), Leguerrier et al. (2004) and Richardson et al. (2004) impose that observations should be exactly reproduced in the final solution. This is in contrast with others who allow deviations from the observations (Klepper and Van de Kamer, 1987, 1988; Vézina and Platt, 1988; Jackson and Eldridge, 1992; Chardy et al., 1993; Eldridge and Jackson, 1993; Vézina and Pace, 1994; Vézina et al., 1997; Vézina and Savenkoff, 1999; Hart et al., 2000; Vézina et al., 2000; Diffendorfer et al., 2001; Gaedke et al., 2002; Richardson et al., 2003; Vézina and Pahlow, 2003). Lyche et al. (1996a,b) fitted data of a radioisotope study with a tracer model and implemented the 95 % confidence interval of each flow in the constraint equation (Eqs. 2.3). Leguerrier et al. (2003), Leguerrier et al. (2004) and Vézina et al. (2000) introduce field data also in the constraint equation (Eqs. 2.3).

Information type 4

For most processes and flows in food webs there exist data in the literature. However, the idiosyncrasy of natural food webs implies that one cannot simply assume universal validity of flow or process data across systems and seasons. Authors rate this validity and uncertainty differently. Some put this type of information on equal footing as site-specific observations and introduce these cross-system data in the same equations as their site-specific observations (Klepper and Van de Kamer, 1987, 1988; Chardy et al., 1993; Stone et al., 1993; Hart et al., 2000; Diffendorfer et al., 2001; Gaedke et al., 2002; Leguerrier et al., 2003, 2004). By doing so, the level of under determinacy decreases and the LIM may even become over determined. Others consider this type of data as rather uncertain and refer literature information to the inequality equation (Eqs. 2.3) (Vézina and Platt, 1988; Jackson and Eldridge, 1992; Eldridge and Jackson, 1993; Vézina and Pace, 1994; Lyche et al., 1996b; Vézina et al., 1997; Niquil et al., 1998; Donali et al., 1999; Vézina and Savenkoff, 1999; Vézina et al., 2000; Leguerrier et al., 2003; Richardson et al., 2003; Breed et al., 2004; Leguerrier et al., 2004; Richardson et al., 2004).

Information type 5

A suite of applications minimize the sum of squared flows (i.e. L2 solution norm) and thereby implicitly use initial guesses of zero in the solution norm (Vézina and Platt, 1988; Jackson and Eldridge, 1992; Eldridge and Jackson, 1993; Vézina and Pace, 1994; Lyche et al., 1996a,b; Vézina et al., 1997; Niquil et al., 1998; Donali et al., 1999; Vézina and Savenkoff, 1999; Vézina et al., 2000; Leguerrier et al., 2003; Richardson et al., 2003; Breed et al., 2004; Leguerrier et al., 2004; Richardson et al., 2004). Others derive initial guesses from the literature data and arrive at an over determined system, which is solved by minimizing the residual norm (Klepper and Van de Kamer, 1987, 1988; Chardy et al., 1993; Stone et al., 1993; Hart et al., 2000; Diffendorfer et al., 2001; Gaedke et al., 2002).

Power and weighting of the minimization norms

There is also some leeway in which norm is used. The L1 prediction norm has been applied by Stone et al. (1993), Hart et al. (2000), Diffendorfer et al. (2001) and Gaedke

et al. (2002) and the weighted L1 prediction norm is applied by Klepper and Van de Kamer (1987) and Klepper and Van de Kamer (1988). The L2 solution norm is applied in many studies (Lyche et al., 1996a), Niquil et al. (1998) Donali et al. (1999) Leguerrier et al. (2003) Breed et al. (2004), Leguerrier et al. (2004) and Richardson et al. (2004). The weighted L2 solution norm is used by Jackson and Eldridge (1992) and Eldridge and Jackson (1993) and the weighted L2 prediction norm is used by Chardy et al. (1993). Finally, a combination of the L2 solution and residual norm is used by Vézina and Platt (1988) and Lyche et al. (1996a) and a combination of the weighted L2 solution and prediction norm is used by Vézina and Pace (1994), Lyche et al. (1996b), Vézina et al. (1997), Vézina and Savenkoff (1999), Vézina et al. (2000) and Richardson et al. (2003).

This systematic evaluation of LIM applications shows that there is little consistency in the way LIM elements are combined (Fig. 2.1), but most prominent differences exist in

1. the amount of cross-system data, prior information and initial guesses available and how these data enter the different equations, which relates to how one overcomes the under determinacy of the model and
2. the minimization norm that is chosen to solve the resulting equations, i.e. how one singles out one solution.

2.3 Analysis of the inverse solution

The quality of parsimonious food web reconstructions and the implications of the steady-state assumption have been tested through twinning experiments in which food web flows are reconstructed from data generated by mechanistic food web models (Vézina and Pahlow, 2003). However, a rigorous analysis of the inverse food web solution is usually lacking. In what follows, a review of the literature and a systematic analysis of five published LIM according to 4 criteria is presented: solution existence, identifiability, uniqueness and sensitivity and show that such an analysis is not only informative, but also essential if we are to increase our understanding of real food web functioning. Our objective is to show that solution analysis reveals informative features of food webs and that these may aid in the evaluation of the results and illumination of potential limitations and pitfalls of the inverse method.

2.3.1 Example food webs

Five studies that applied LIM to recover carbon flows in marine (Donali et al. 1999, Niquil et al. 1998, Leguerrier et al. 2003, Chapter 3) and freshwater (Diffendorfer et al., 2001) systems were selected for analysis. The systems differed in compartment and flow resolution, number of in situ and literature flow data, number of inequalities and minimization norm (Table 2.1). The marine examples considered the benthic or pelagic food web, consisting of autotrophs, bacteria, size-class based eucaryotes (i.e. flagellates and zooplankton) and non-living compartments DOC (dissolved organic carbon) and detritus, whereas only Leguerrier et al. (2003) considered a coupled benthic-pelagic system. There were biomass estimates for most compartments and several site-specific flow measurements (e.g. primary production). Literature data were used either to increase the number of imposed flow measurements (Leguerrier et al., 2003) or to constrain respiration and production rates, assimilation and bacterial growth efficiencies and diet composition. Diffendorfer et al. (2001) restricted the analysis to a herpetological assemblage (9 groups, including salamanders, frogs and snakes), for which biomass

and diet composition data were measured or estimated. There were no direct flow measurements, but consumption and respiration rates were estimated from the literature and used to set up constraints and initial guesses. All food webs were considered to be in steady-state, except for Donali et al. (1999) who estimated rates of change through successive sampling. Van Oevelen et al. (Chapter 3) used carbon and $\delta^{13}\text{C}$ mass balances to reconstruct carbon flows in an intertidal benthic food web and two cases are considered: one where only total carbon flows were considered, one with both carbon mass balances. All food webs are implemented, solved and analyzed in the modeling environment FEMME (Soetaert et al., 2002) (see <http://www.nioo.knaw.nl/cemo/femme> for downloads).

Table 2.1: Characteristics of 5 food webs from the literature in terms of number of flows (F) and compartments (C), site-specific data, number of equations (E) and inequalities (I) and solution norm (N). References are 1) Donali et al. (1999), summer food web 2) Niquil et al. (1998), 3) Leguerrier et al. (2003), 4) Diffendorfer et al. (2001), wet prairie food web and 5) Van Oevelen et al. (Chapter 3). Abbreviations are bm = biomass pp = primary production, bp = bacterial production, zp = zooplankton production and respiration, op = oyster predation, cr = community respiration, sr = sedimentation rate, rc = rate of change, dc = diet composition, ni = natural abundance isotope data, ti = tracer isotope data.

Site characteristics ^{Reference}	F	C	Site-specific data	E	I	N
Gulf of Riga, pelagic ¹	26	7	bm, pp, bp, cr, sr, rc	14	25	L2
Takapoto Atoll, pelagic ²	32	7	bm, pp, bp, zp	15	26	L2
Marennes-Oléron Bay, tidal ³	95	16	bm, pp, zp, op	33	62	L2
Everglades, wet prairie ⁴	39	9	bm, dc	39	36	E1
Molenplaat, intertidal benthic ⁵	39	8	bm, pp, bp, cr, ni, ti	20	33	E2

2.3.2 Solution existence

Solution existence boils down to a check of the internal consistency of the various data sources and model structure. When inconsistent, the specified LIM cannot be solved; this situation might arise for example when, under steady-state, primary production is insufficient to satisfy the minimal carbon demands by herbivores. Identifying conflicts in the data set may reveal the need for better, more or alternative data. Inconsistent data sets have been dealt with in various ways. Vézina and Savenkoff (1999), Vézina et al. (1997), Vézina et al. (2000) and Richardson et al. (2003) allowed deviations from a steady state mass balance and observations. Van Oevelen et al. (Chapter 3) detected inconsistencies between two observations and adapted one to agree with the one that was regarded a better long term integrator of food web processes. Other authors change the model structure (Vézina and Pace, 1994; Richardson et al., 2003) or modify or remove some constraints from the model (Eldridge and Jackson, 1993; Vézina and Pace, 1994; Vézina and Savenkoff, 1999) in order to obtain a solution.

2.3.3 Solution identifiability

Solution identifiability deals with the question whether an inverse model contains sufficient information to identify the unknown parameters (in this case the flows). Identifiability is a quality label that allows distinguishing between flows that are well determined from those that are more uncertain. The uncertainty that surrounds inverse solutions has spurred

discussion in geophysical sciences (McIntosh and Rintoul, 1997), but is under appreciated in biological applications (but see Stone et al. (1993) and Vézina and Pahlow (2003)). There exist different measures of identifiability, here the discussion is limited to two measures that are easy to calculate and interpret: range estimation and flow resolution.

During **range estimation** an envelope around each flow is generated that depicts the feasible range that a flow can attain. The ranges are calculated by minimizing and maximizing each flow successively. Put more formally, the following is solved for each flow:

$$\text{minimize } x_i \text{ and maximize } x_i \tag{2.10}$$

subject to

$$\mathbf{Ax} = \mathbf{b} \text{ and } \mathbf{Gx} \leq \mathbf{h} \tag{2.11}$$

Note that when a flow is minimized or maximized this eliminates the need for the optimization equation (Eqs. 2.2) and data this equation does therefore not influence the outcome of the range estimation. Range estimation reveals which flows are well constrained (small range) in the LIM and which are not (large range). Thereby it also indicates which additional data can better constrain the food web. Note that flow values are not independent of each other. The linear dependency amongst flows might imply that fixing the value of one flow may fix the value of another flow or of a combination of flows. Thus, the solution space is sparse and the uncertainty is exaggerated in this type of analysis.

Klepper and Van de Kamer (1987) introduced range estimation to show that a mass-balance approach may greatly reduce the span of the initial range estimate. Stone et al. (1993) estimated ranges while changing some input data by 10 %. The potential ranges were large and they concluded that small changes in the environment may produce large changes in individual flows.

It is clear that for each of the example food webs, some of the flows are highly uncertain whereas others are much better constrained (Fig. 2.3). The food web reported by Donali et al. (1999) has the largest number of in situ measurements relative to the number of flows that have to be estimated (Table 2.1) and it is one of the best constrained food webs (Fig. 2.3A). In contrast, with 95 flows to be estimated based on only 33 equations, the food web of Leguerrier et al. (2003) is least well constrained (Fig. 2.3C). The inclusion of stable isotope data in the intertidal food web increased the number of mass balances and improved the quality of the inverse solution considerably (compare Fig. 2.3E and 2.3F, see also 2.4 Discussion).

In addition, the food web structure (the number of flows) is uncertain in all reported webs, because some flows have a lower boundary of 0 and may therefore be potentially absent from the food web. The proportion of flows with a lower boundary of 0 differed remarkably among the food webs (Table 2.2), > 30 % of the food web flows in Leguerrier et al. (2003) and Niquil et al. (1998) to < 5 % for Donali et al. (1999). In addition, the food web by Leguerrier et al. (2003) contained many flows with an infinite upper boundary (Table 2.2 en Fig. 2.3C).

2.3. Analysis of the inverse solution

Table 2.2: Summary of the range estimation of the example food webs. '0-flows' are number of flows that are zeroed in the final solution, 'Lower bound 0' is the number of flows with a lower boundary of 0 and 'Upper bound ∞ ' is the number of flows with an infinite upper boundary.

Reference	0-flows	Lower bound 0	Upper bound ∞
Donali et al. (1999)	1	8	0
Niquil et al. (1998)	4	21	0
Leguerrier et al. (2003)	26	71	19
Diffendorfer et al. (2001)	6	6	0
Van Oevelen et al. (Chapter 3)	9	16	0

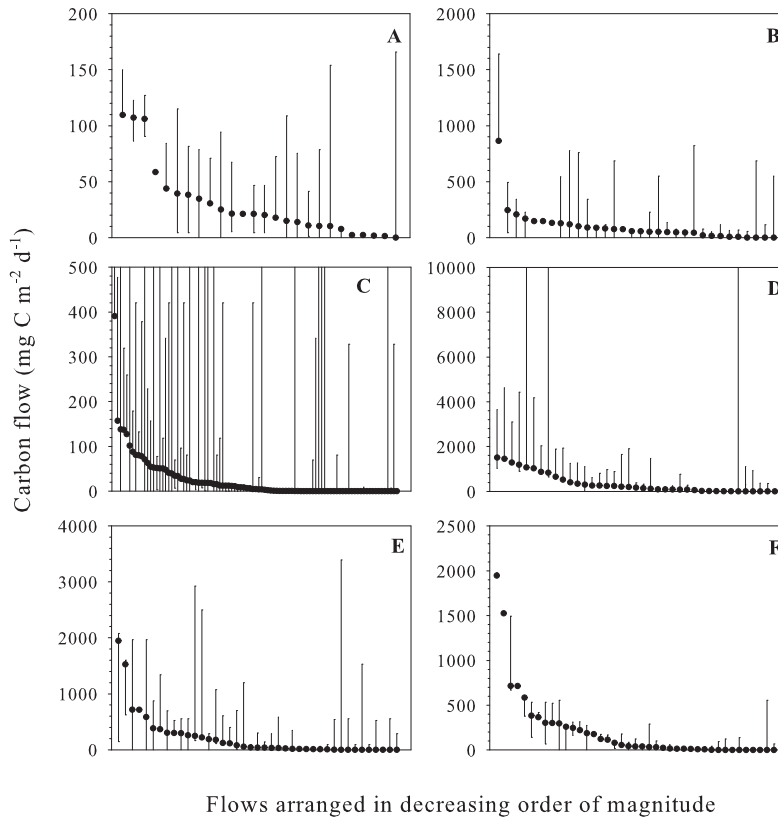


Figure 2.3: Range estimation of the example food webs A) Donali et al. (1999), B) Niquil et al. (1998), C) Leguerrier et al. (2003), flows out of scale indicate unbounded flows, D) Diffendorfer et al. (2001), flows out of scale are $> 40000 \text{ mg C m}^{-2} \text{ d}^{-1}$, E) Van Oevelen et al. (Chapter 3) without natural abundance stable isotope data and F) Van Oevelen et al. (Chapter 3) with natural abundance stable isotope data. Markers show food web flows solved with the L2 solution norm with initial estimates as zero.

Flow resolution is a measure of how well each flow is determined by information contained in the equality equation (Eqs. 2.1). The flow resolution varies between 0 and 1,

and indicates to what extent each flow is resolved independently of other flows (Menke, 1984; Vézina and Platt, 1988): 1 implies that the flow is resolved uniquely and lower values imply increasing dependence on other flows. The resolutions depend solely on the model and data structure contained in matrix \mathbf{A} and not on the numerical data in vector \mathbf{b} , hence they can be used for the optimization of sampling strategies (Menke, 1984).

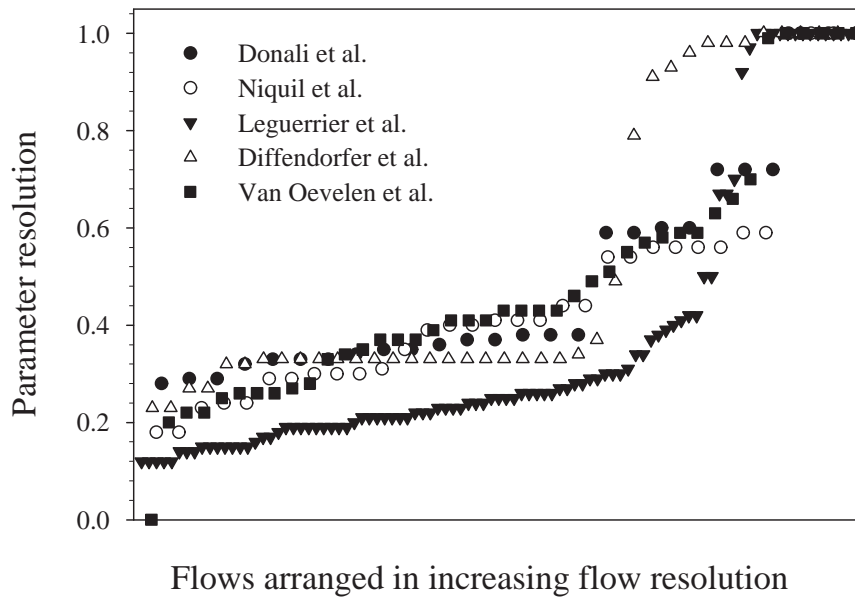


Figure 2.4: Flow resolutions of the example food webs. Flow resolutions for Van Oevelen et al. are calculated only without the stable isotope equations, because the use of stable isotope data complicates their calculation.

Reports on flow resolution are few, see Vézina and Platt (1988) and Niquil et al. (1998). The flow resolutions for all the examples show that the majority of the flows are not uniquely determined but are strongly related with other flows (i.e. most flow resolutions < 0.50 ; see Fig. 2.4). The food webs by Donali et al. (1999), Niquil et al. (1998) and Diffendorfer et al. (2001) are approximately equally resolved, whereas interdependencies are strongest in the food web by Leguerrier et al. (2003) as shown by the low resolution values. All food webs show sharp transitions in resolution patterns and these are most pronounced in Diffendorfer et al. (2001) and Leguerrier et al. (2003). In the case of Leguerrier et al. (2003) the flows related to macrobenthos are well constrained, because the relative contributions in their diet were fixed by prey biomass proportionality. Hence, it is possible to link the resolution patterns to the imposed information.

2.3.4 Solution uniqueness

Some inverse problems may have more than one optimal solution. This raises the question as to how one judges different food webs (i.e. different solutions) that equally comply with

the inverse model formulation. It appears that the question of non-uniqueness is related to the chosen norm. Quadratic minimization (L2 solution norm) should always lead to a unique solution (Lawson and Hanson, 1995), whereas linear minimization solutions are not always unique (Klepper and Van de Kamer, 1987). For example when two flows are possible, the L2 solution norm will equally divide among these possibilities, whereas when using the L1 solution norm, one combination will be arbitrarily selected.

When the example food webs were solved with linear programming techniques to find the most parsimonious solution (Eqs. 2.5) they all had a very large (probably infinite) number of optimal solutions (results not shown). Linear optimization is normally not used to calculate the most parsimonious food web, however it is commonly applied to over determined systems where the deviation from observations and initial estimates is minimized (see 2.2.4 Combining the elements of LIM) (Stone et al., 1993; Hart et al., 1997; Diffendorfer et al., 2001; Gaedke et al., 2002), and the question of non-uniqueness is usually not addressed (but see Klepper and Van de Kamer (1987)). To illustrate the issue, the example food web from Diffendorfer et al. (2001) was solved repeatedly using the L1 prediction norm with the ordering of the equations altered. Twenty-three different optimal solutions were found. Most flows retained their values (54 %), but for some their value varied considerably (Fig. 2.5). Clearly non-uniqueness should be an aspect to consider when solving an inverse model.

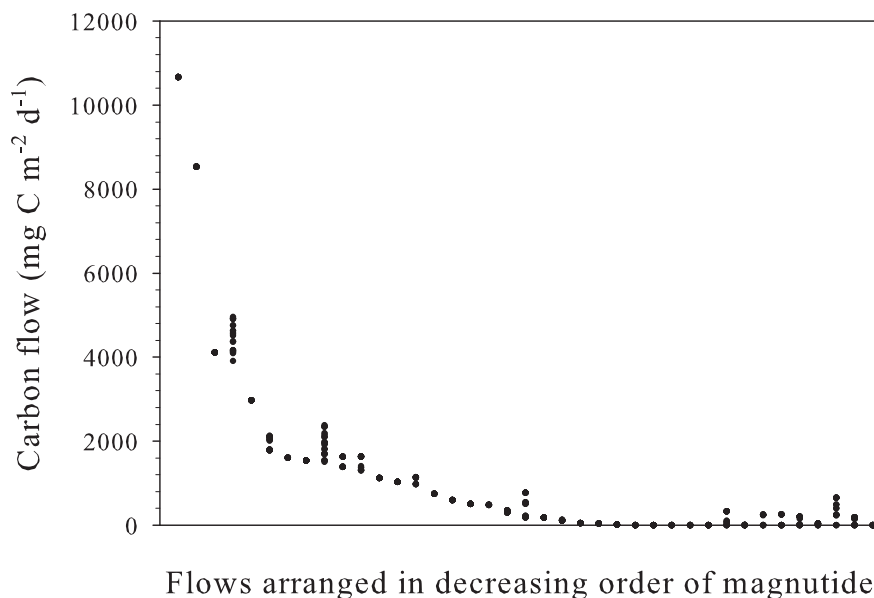


Figure 2.5: Twenty-three optimal solutions of the inverse problem by Diffendorfer et al. obtained by the L1 prediction norm.

2.3.5 Solution sensitivity

Solution sensitivity deals with the question of how (small) changes in the input propagate and change the inverse solution. A detailed ecological analysis of a food web that changes dramatically with small changes in the input data bears little relevance. Consequently, the sensitivity aspect of the solution has received considerable attention in the literature.

Systematic sensitivity analysis to test the robustness of inverse analysis range from small perturbations (10 - 20 %) of input data (Jackson and Eldridge, 1992; Niquil et al., 1998; Donali et al., 1999; Vézina and Savenkoff, 1999; Gaedke et al., 2002; Leguerrier et al., 2003; Richardson et al., 2003; Breed et al., 2004; Richardson et al., 2004) to rigorous changing or removing input values (Niquil et al., 1998; Donali et al., 1999; Vézina and Savenkoff, 1999; Hart et al., 2000; Vézina et al., 2000). Others have tested the sensitivity of the assumed food web structure (Vézina et al., 1997; Niquil et al., 1998; Richardson et al., 2003) or the imposed constraints (Leguerrier et al., 2004). The main conclusion for food webs solved with the L2 solution norm is that sensitivity to structure, input data and constraints is generally limited, although individual flows may change considerably.

Stone et al. (1993) and Diffendorfer et al. (2001) tested the sensitivity of an over determined model solved with the L1 residual norm. Stone et al. (1993) identified large changes in some individual flows as obtained by range estimation. The L2 solution norm assessment studies did not use range estimation and these ranges may be large (see 2.3.3 Solution identifiability), which complicates a direct comparison with the results of Stone et al. (1993). Diffendorfer et al. (2001) tested how much a constraint was allowed to change before a structural change in the food web was detected. The sensitivity appeared to be highly dependent on which constraint was perturbed, some did not give significant structural changes whereas others did. The potential non-uniqueness of linear programming problems (see 2.3.3 Solution identifiability) was not explicitly addressed in these L1 residual norm based sensitivity studies and this might have affected the analysis.

2.4 Discussion

There is a growing appreciation in the ecological literature for a more quantitative approach to natural food webs. Linear inverse modeling (LIM) is frequently used in food web research to quantify energy and elemental flows among ecological compartments. A systematic evaluation of inverse problems in ecology identified methodological aspects of LIM that have remained hidden or were under appreciated in the ecological literature. A general framework based on data sources, model equations and minimization norms was developed to distinguish different LIM approaches and reveal the essential choices that are to be made during the modeling process. In the literature there are major differences in the way these elements are used and combined to solve the food web model (Fig. 2.1). The alternatives imply a quality appraisal of the site-specific (type 3) and literature (type 4 - 5) data. The number of equations relative to the number of flows to be estimated is the major factor governing the quality of the inverse solutions. Researchers have therefore adopted strategies to include non-site specific, i.e. cross-system, literature (type 4) data in the equalities to increase the level of determination. Other applications select the parsimonious solution (e.g. Vézina and Platt, 1988) and in fact impose initial estimates as zero and the optimal solution is the one that meets these initial estimates as closely as possible.

An alternative approach is to include additional site-specific (type 3) data in the mass balance equalities, for instance, based on ecological stoichiometry (Sturner and Elser, 2002). Although the dynamic nature of unbalanced growth of autotrophs is not readily

implemented in a linear mass balance representation, homeostasis of consumers (bacteria, heterotrophic eukaryotes) can be used to link carbon, nitrogen and phosphorus flows in the food web (e.g. Vézina and Savenkoff, 1999; Eldridge and Jackson, 1993; Gaedke et al., 2002). Another promising approach involves the use of biomarkers, such as fatty acids (Iverson et al., 2004) or stable isotopes to quantify the contribution of different food sources in an organisms diet (Post, 2002). As these data fix relative proportions of diet sources, they increase the equations without increasing the number of unknowns and thereby narrow the uncertainty of the food web flows. Van Oevelen et al. (Chapter 3) pioneered the use of stable isotopes (both natural abundance and tracer data) in an inverse analysis of a benthic food web and show that the uncertainty in the inverse solution decreased considerably (compare Fig. 2.3E and F). Stable isotope and fatty acid signatures are becoming standard tools in food web ecology and their incorporation in LIMs is straightforward (Van Oevelen et al. (Chapter 3)).

Although LIMs are now frequently used to reconstruct element flows in food webs, there is comparatively little attention for the quality and properties of the inverse solution. Vézina and Pahlow (2003) tested the accuracy of the commonly used L2 solution norm reconstructed using twin-modeling experiments. Kones et al. (Accepted) uses factorial analysis on a large set of generated solutions of two under determined inverse models to reveal general patterns in these two sets of solutions. The factor analysis revealed that a few groups of flows, or sub-food webs, explain a large part of the variation in the set of different solutions. This means that different structures that may exist in these sub-food webs explain the uncertainty observed in the large complex food webs. This can be used to gain insight into the uncertainty of the complex food web and aid in developing a sampling strategy for further constraining the model.

In addition however, inverse theory has provided criteria such as identifiability and uniqueness to analyze inverse solutions, but these are not often applied in ecological applications. The systematic analysis of solutions of five previously quantified food webs revealed some important aspects. First of all, these inverse problems are strongly under determined and a large uncertainty surrounds most of the reconstructed food web flows (Fig. 2.3). These uncertainties should be considered when the solutions of the models are input to subsequent food web analysis, such as ecological network analysis (Ulanowicz, 1986; Bersier et al., 2002). Secondly, in all the revisited examples, most of the flows were strongly related to other flows and were not uniquely determined. Last but not least, when linear programming is used to single out an optimal solution, several alternative solutions may exist (Fig. 2.5).

