

Analysing the allometry of multiple interacting traits

Bill Shipley

Département de Biologie, Université de Sherbrooke, Canada

Received: 15 October 2003 · Revised version accepted: 23 January 2004

Abstract

Since form and function are tightly integrated in plants, and since plant attributes often scale allometrically, it follows that plant allometry is inherently multivariate. Unfortunately, traditional statistical methods for studying allometric relationships are very restrictive and do not allow one to model multivariate allometric patterns that follow realistic biological hypotheses. In this paper I describe a new statistical test ('d-sep test') that allows one to test, and potentially falsify, alternative multivariate orderings of cause-and-effect in the context of allometry.

Key words: allometry, causal models, directed acyclic graphs, graphical models, plant size, *Prunus mahaleb*, fruit production, seed dispersal

Introduction

Allometry, the study of how the attributes of organisms change with respect to one other, has a long and distinguished history in animal biology (Gould 1966; Huxley 1972; Peters 1983). Although different definitions of 'allometry' exist in the ecological and statistical literature (Muller et al. 2000), in this paper I adopt the most general meaning: proportional changes in the sizes or rates of change in organismal attributes. Plant biologists are relative newcomers, although comprehensive treatments are now available (Niklas 1994). We know that form and function are tightly integrated in organisms, and that many different attributes contribute to both form and function. Presumably, these integrated suites of attributes should change together with complicated patterns of direct and indirect relationships between them. In other words, allometry should be inherently multivariate. Curiously, although there are literally hundreds of published allometric relationships, a large majority of

these are simple bivariate relationships of the form $Y = aX^b$. Why should this be?

I doubt that the many researchers who study allometry truly believe that organismal attributes interact only two at a time. The reason, I suspect, is more Procrustean and unfortunate. We have had to force our biological hypotheses into such simple statistical models because we have not had the statistical tools needed to test and describe such multivariate processes. Certainly, there are well-developed multivariate methods of data reduction, such as principle components and related techniques, but these techniques are not designed to test and describe structural relationships of cause and effect; quite the contrary, the developers of such methods intended for them to ignore such relationships (Shipley 2000a). The purpose of this paper is to describe some new statistical methods that can be used to study multivariate allometry.

As an example, consider Fig. 1, which shows the bivariate allometric trends between five attributes of 60 individual trees of St. Lucie's Cherry (*Prunus mahaleb*)

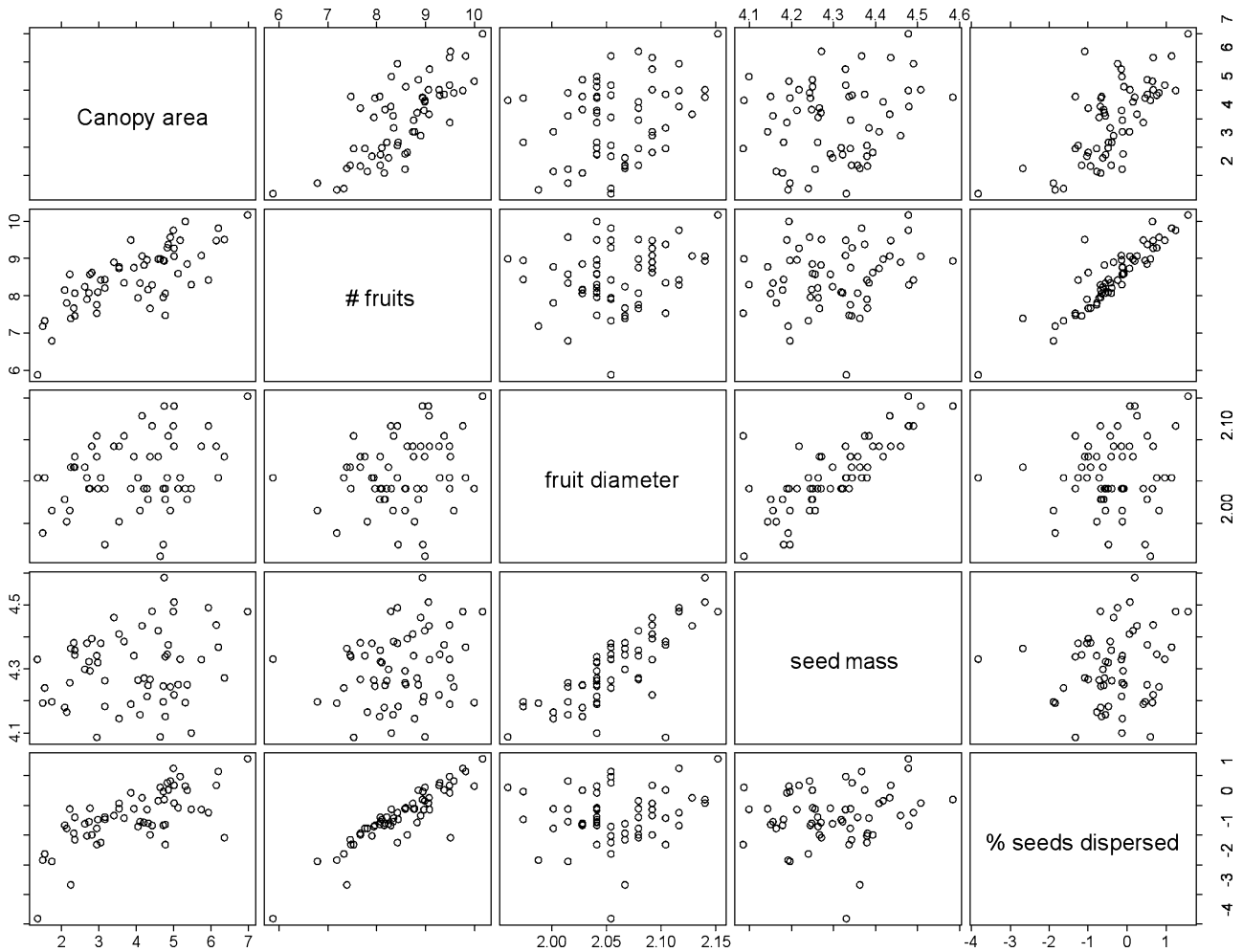


Fig. 1. Scatterplot matrix of five attributes measured on 60 plants of St. Lucie's Cherry (*Prunus mahaleb*). Data from Jordano (1995); all data are Ln-transformed.

(Jordano 1995). For each tree Pedro Jordano measured its canopy projection area, the number of fruits produced, the average fruit diameter, the average seed mass and the percentage of seeds dispersed away from the tree by birds, all attributes related to evolutionary fitness. As is typical in such data sets, there appear to be complicated patterns of correlation between these attributes. These patterns of correlation likely reflect underlying functional linkages between the attributes and these patterns show clear allometric trends. What biological processes might have generated such patterns?

As plant ecologists, we could propose different hypothetical explanations and these would lead to different cause-effect linkages. Figure 2 shows three different hypotheses; the ' ϵ_i ' variables represent those unique unknown other causes of each variable. The first (model A) is the simplest; the size of the photosynthetic capital of the tree, measured by the canopy projection area, determines each of the other attributes.

Model B also assumes that canopy projection area is the driving variable, but differs from model A by assuming that the effect of canopy projection area on the number of seeds dispersed and on the average mass of seeds is indirect. These indirect effects are mediated by the number of fruit produced and the average size of fruits, respectively. Model C proposes that the fruit diameter and seed mass are allometrically related but are independent of the allometric effects among the other three attributes. There exist methods of exploring alternative structures in the data (Shipley 1997, 1999, 2000a), but here we will assume that the three alternative orderings represent alternative biological explanations that the researcher wants to test. There are two different, but related, methods of testing (and potentially falsifying) such hypotheses. One, called a 'd-sep test' (Shipley 2000a, b, 2003), is based on newly derived relationships between causal graphs and probability distributions (Pearl 1988, 2000). Another, called 'structural equations modelling' (Shipley 2000a), is a

maximum likelihood technique which has some disadvantages with respect to the d-sep test, but also an important advantage since one can include unmeasured (latent) variables. In this paper, I concentrate on d-sep tests, but will compare the advantages and disadvantages of each class of methods in context. For a more detailed explanation of the history of these methods in biology, see Shipley (2000a).

D-sep tests

The models in Fig. 2 are called ‘directed graphs’ in the mathematical language of graph theory. Biologically, they are the type of box-and-arrow diagrams that ecologists

have used for a long time and they specify how causal effects are hypothesized to propagate through the measured variables in our trees. If we imagine these causal effects being generated in all of the trees that potentially interest us, then the distribution of all possible values of our five variables defines a multivariate probability density function. Different quantitative values for these causal effects, and different functional forms defining the equations linking the variables, will produce a family of different multivariate probability densities, but they will all have certain mathematical constraints that are common to all and that are caused by the way the variables are linked together as causes and effects. These constraints are in the form of conditional independence relationships.

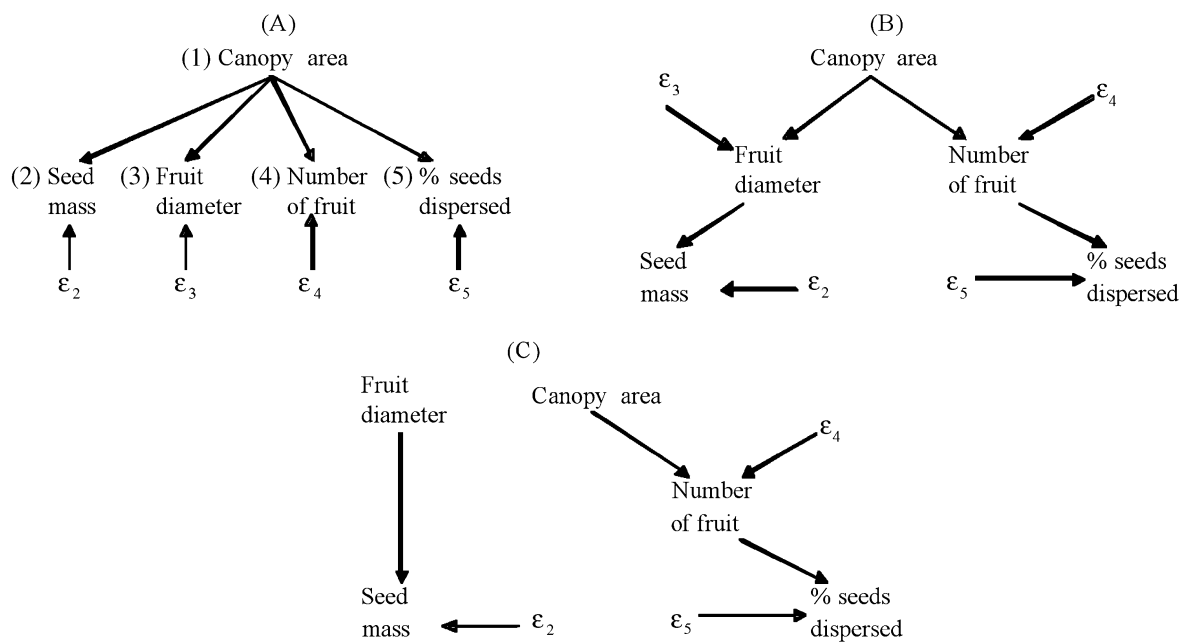


Fig. 2. Three alternative hypotheses (models A, B and C) describing the cause-effect relationships among the five attributes. Error variables are indicated by ϵ_i .

Table 1. Basis sets for the partial independence constraints implied by each of the three hypotheses shown in Fig. 1. The notation ‘(X,Y){A,B,...}’ means that variables X and Y are d-separated, and hypothesized to be probabilistically independent, conditional on the set of variables {A,B,...} and the ‘ ϕ ’ represents the null (empty) set. Pearson’s partial correlation coefficient (r), and probability assuming the null hypothesis (in parentheses), are given for each conditional independence claim; values in bold have a probability below 0.05. The overall model is tested with Fisher’s C statistic, which is distributed as a chi-squared variate if the data agree with the model.

Model A		Model B		Model C	
Basis set	r (probability)	Basis set	r (probability)	Basis set	r (probability)
(2,3){1}	0.766 (10^{-12})	(1,2){3}	-0.066 (0.322)	(1,5){4}	0.077 (0.566)
(2,4){1}	0.022 (0.870)	(2,4){1,3}	0.141 (0.292)	(1,2){3}	-0.066 (0.622)
(2,5){1}	0.107 (0.422)	(2,5){3,4}	0.005 (0.973)	(3,4){1}	0.021 (0.875)
(3,4){1}	-0.077 (0.566)	(3,4){1}	0.021 (0.875)	(3,5){4}	-0.137 (0.301)
(3,5){1}	0.021 (0.875)	(3,5){1,4}	-0.155 (0.245)	(2,4){3}	0.141 (0.292)
(4,5){1}	0.800 (10^{-14})	(1,5){4}	0.077 (0.566)	(2,5){3,4}	0.005 (0.973)
				(1,3){ ϕ }	0.301 (0.019)
12 df, C = 137.64 ($\ll 10^{-16}$)		12 df, C = 7.69 (0.809)		14 df, C = 15.16 (0.368)	

Pearl (1988; but see also Geiger et al. 1990; Geiger & Pearl 1993; Pearl 2000) has shown that all such conditional independence relationships can be predicted directly from the directed graph using a graph theoretic operation called 'directed separation' ('d-separation'), and this holds independently of the distributional details of each variable or of the functional form of the equations linking the variables together. In other words, given multivariate cause-effect hypotheses like those shown in Fig. 2, we can obtain the alternative predicted conditional independence constraints that must exist if the different hypotheses apply. If these predicted constraints do not exist in the empirical data, then we must reject the multivariate causal hypothesis that implied such constraints. Based on these results, I developed a new statistical test, which I call a d-sep test (Shipley 2000b).

The first step is to obtain the basis set (B_U). This is the smallest set of d-separation relations that, in combination, predict all other d-separation relations that exist in the directed graph using the axioms of conditional probability (Geiger et al. 1991). To obtain this basis set, simply list each pair of variables in the directed graph that do not have an arrow between them. Next, list the 'causal parents' of the pair, i.e. those variables that directly cause either variable in the pair. Each d-separation relation in the basis set states that the two variables forming the pair are probabilistically independent conditional on the causal parents of the pair. Table 1 lists the basis sets for each of the three hypotheses shown in Fig. 2.

The next step is to actually test each predicted independence claim against the empirical data. How this is done depends on the nature of the variables and the type of relationship between them. If the two variables in the pair are approximately normally distributed and the relationship between them is linear, then the claim can be tested using Pearson's partial correlation coefficient. If the two variables are at least continuous and the relationship is monotonic, then one can use Spearman's partial correlation coefficient. More complicated nonlinear relationships require nonparametric regression smoothers and permutation tests; details are in Shipley (2000a).

The final step is to combine these separate tests of independence into a combined test of the entire model. This is done using Fisher's C statistic. Obtain the exact probability level for each of the k separate d-separation claims (p_i) in the basis set, assuming conditional independence. The statistic

$$C = -2 \sum_{i=1}^k \ln(p_i) \quad \text{Eqn. 1}$$

is distributed as a chi-squared variate with $2k$ degrees of freedom if all of the d-separation claims in the model hold in the data. An unusually large value, asso-

ciated with a probability below your significance level, is evidence against the model in question.

The minimum number of observations needed depends on the statistical tests used to evaluate independence. For instance, if (conditional) independence is tested using Pearson's or Spearman's (partial) correlation coefficient, then the minimum sample size would be $N-p-1$, where N is the sample size and p is the largest number of causal parents in a single basis set. Since p rarely exceeds 2 or 3, this means that one can test models with quite small samples; of course, with such small samples the statistical power would be low. Furthermore, if the tests of independence are exact then the overall probability of the model is also exact. This emphasizes an important advantage of d-sep tests over structural equation modelling (SEM). SEM, being a maximum likelihood technique that gives only asymptotic estimates, requires a large minimum sample size. This minimum size for SEM depends, among others, on the number of estimated parameters, and this can often require hundreds of observations.

There are other advantages of the d-sep test over SEM. First, SEM assumes multivariate normality and linearity. The d-sep test does not necessarily assume either since it is possible to test (conditional) independence using non-parametric tests or even permutation methods (Shipley 2000a). Second, since SEM simultaneously estimates all free parameters (path coefficients, variances, covariances), errors anywhere in the model (including distributional violations) will spread throughout it and it is therefore not possible to unambiguously identify where such errors occur. The d-sep test, being based on mutually independent claims in the basis set, allows one to identify the source of structural errors. However, there are two important disadvantages of the d-sep test that are not shared by SEM. If your model includes a latent (i.e. unmeasured) variable that is a common cause of more than two observed variables in your model, or if you cannot completely specify the causal structure and thus include undirected correlations, then you must use SEM, although an important class of models including undirected correlations (or covariances) can be tested using the d-sep test (Shipley 2003).

Since the data shown in Fig. 1 are approximately normally distributed once Ln-transformed, and the relationships do not show any nonlinear patterns, I use Pearson's partial correlation coefficient to test the claims of conditional independence. Table 1 lists the results. Model A is rejected by the data; the probability of a causal process hypothesized by this model actually generating data like those observed by Jordano (1995) is much less than 10^{-16} . Looking at the individual claims of conditional independence implied by this model, we see that two are clearly wrong: seed mass and fruit diameter are not independent conditional on

Table 2. Likelihood ratios (probability of model in row *i*/probability of model in column *j*) of the three models shown in Fig. 1.

	Model A	Model B	Model C
Model A	1.00	$\sim 1.25 \times 10^{-16}$	$\sim 2.7 \times 10^{-16}$
Model B	8.1×10^{15}	1.00	2.2
Model C	3.7×10^{15}	0.45	1.00

canopy area, and the percentage of seeds dispersed by birds is not independent conditional on canopy area. Model B is not rejected by the data; the probability of a causal process hypothesized by this model actually generating data like those observed by Jordano (1995) is 0.809. Model C is not rejected by the data either since its probability is 0.368. This emphasizes an important aspect of this statistical test: as in every statistical test the failure to reject a model is not the same as proving that it is true. All we can say is that both models B and C are possible explanations and that model A is not. If we wish to apply a likelihood ratio approach (Royall 1997) in choosing between the three models (Table 2), then model A is again clearly rejected in favour of either models B and C, while the likelihood ratio of model B versus model C (2.2) provides only mild support in favour of model B.

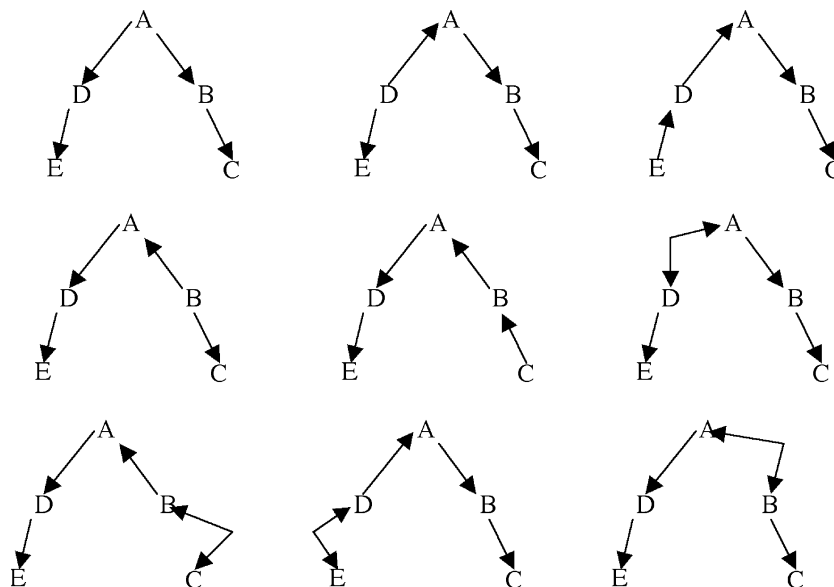
Equivalent models

There is another, and less obvious, reason why the failure to reject a model cannot be construed as strong evidence in favour of it. Given a directed graph there are

almost always statistically 'equivalent' models. Equivalent models are different directed graphs that imply the same conditional independence constraints in the resulting multivariate probability densities, and are therefore statistically indistinguishable (Shibley 2000a). Here is how to obtain such equivalent models relative to a given model *M*.

1. Convert model *M* to a new model *E* by changing all arrows in *M* to lines.
2. Look for 'unshielded colliders' in model *M*. An unshielded collider is a set of three variables (*X*, *Y*, *Z*) such that there is an arrow pointing into *Y* from *X* and an arrow pointing into *Y* from *Z*, but no arrow between *X* and *Z*. For every unshielded collider in model *M*, make the equivalent variables in model *E* form the same unshielded collider. Such a diagram is called a 'partially oriented dependency graph'.
3. For every other line in model *E*, you can orient the direction of the arrow however you want so long as (i) you do not create any unshielded collider that does not exist in model *M* and (ii) you do not create any feedback loops in model *E*. All such different possible models created in this way are statistically equivalent, and will give the same *C* value and probability when tested by a *d*-sep test.

Figure 3 shows the equivalent models relative to model B. A double-headed arrow between two variables means that there is an unmeasured (latent) variable that is the common cause of the two variables to which the arrows are pointing. These models too can be tested using a *d*-sep test since they are statistically equivalent to a model without such a latent variable

**Fig. 3.** Nine equivalent models. All of these models impose exactly the same conditional independence constraints and so cannot be distinguished statistically. Double-headed arrows indicate an unmeasured (latent) variable that is a common causal parent of the two variables at the head of the arrow.

(Shipley 2003). In order to choose between these equivalent models we cannot use statistical evidence and must appeal to biological knowledge.

Multivariate allometry

Given the three alternate models in Fig. 2, model B is best supported by the data. Looking at the equivalent models in Fig. 3, we can exclude most based on some basic biology. First, the percentage of seeds dispersed occurs after the fruits are produced, therefore we can exclude any equivalent model that has an arrow from the percentage of seeds dispersed to the number of fruits produced. Second, since leaves form before fruits, and since photosynthate flows from the leaves (i.e. the canopy) to fruits and seeds during their development, we can exclude any equivalent model that posits causal effects from fruits to the canopy. This leaves just two models, the first and sixth, in Fig. 3. Here, I will concentrate on the first model, which is model B in Fig. 2. This model specifies the ordering of dependent and independent variables, and therefore decomposes the full model into a series of four allometric relationships. Since the relationships appear linear on a log scale, these correspond to four simple bivariate regressions on the Ln-transformed data. If we had a variable with more than one arrow pointing into it then this would imply a multiple regression. We can now estimate the individual allometric relationships (\pm SE) using least-squares regression:

$$(2.1) \text{Ln}(\text{Fruit diameter}) = 2.022 (\pm 0.016) + 0.009 (\pm 0.004) \text{Ln}(\text{Canopy projection area}); r = 0.30, \text{SE}_{\text{residual}} = 0.040.$$

$$(2.2) \text{Ln}(\text{Seed mass}) = 0.003 (\pm 0.461) + 2.089 (\pm 0.224) \text{Ln}(\text{Fruit diameter}); r = 0.77, \text{SE}_{\text{residual}} = 0.072$$

$$(2.3) \text{Ln}(\text{Number of fruit}) = 6.723 (\pm 0.221) + 0.451 (\pm 0.053) \text{Ln}(\text{Canopy projection area}); r = 0.74, \text{SE}_{\text{residual}} = 0.555.$$

$$(2.4) \text{Ln}(\% \text{ seeds dispersed}) = -9.056 (\pm 0.554) + 1.023 \text{Ln}(\text{Number of fruit}); r = 0.90, \text{SE}_{\text{residual}} = 0.411.$$

The final multivariate allometric model is shown graphically in Fig. 4.

Conclusions

The very notion of a plant strategy (Grime 2001) implies suites of plant attributes that interact as a system. If such strategies exist, and if plant attributes often covary following allometric trends, then it follows that plant allometry is inherently multivariate and embedded in a web of cause-effect relationships. When these attributes can be measured with little error, and when the causal relationships between the traits are hypothesized to be expressible without recourse to unmeasured variables, then the d-sep test described here allows one to both test and then fit the resulting multivariate allometric model. For a particular class of models including latent variables the test is still applicable (Shipley

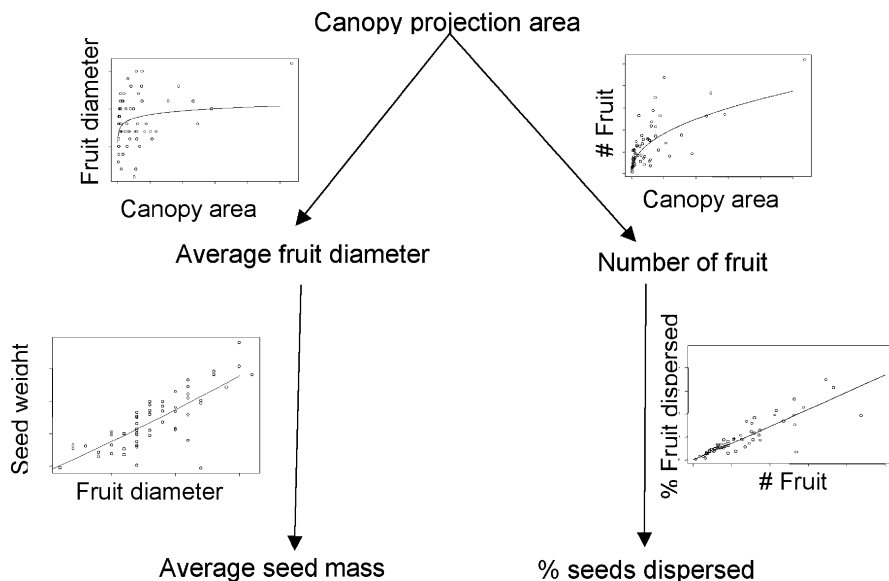


Fig. 4. The final allometric model obtained from model B of Fig. 1. Graphs show the empirical data and the predicted allometric trend.

2003) if the latent variable model is equivalent to a model without latent variables. In cases in which such an equivalent model does not exist then one can use structural equations modelling with latent variables (Shipley 2000a). These methods, by allowing plant ecologists to fit the appropriate statistical model to their hypotheses of multivariate allometry, should allow more interesting allometric hypotheses to be developed.

Acknowledgements. This research was financially supported by the Natural Sciences and Engineering Research Council of Canada. Pedro Jordano kindly provided the empirical data.

References

- Geiger, D & Pearl, J (1993) Logical and algorithmic properties of conditional independence and graphical models. *The Annals of Statistics* 21: 2001–2021.
- Geiger, D, Verma, T & Pearl, J (1990) Identifying independence in Bayesian Networks. *Networks* 20: 507–534.
- Geiger, D, Paz, A & Pearl, J (1991) Axioms and algorithms for inferences involving probabilistic independence. *Information and Computation* 91: 128–141.
- Gould, SJ (1966) Allometry and size in ontogeny and phylogeny. *Biological Reviews of the Cambridge Philosophical Society* 41: 587–640.
- Grime, JP (2001) *Plant Strategies, Vegetation Processes, and Ecosystem Properties*. John Wiley, New York.
- Huxley, JS (1972) *Problems of Relative Growth*. Dover, New York.
- Jordano, P (1995) Frugivore-mediated selection on fruit and seed size: birds and St. Lucie's cherry. *Ecology* 76: 2627–2639.
- Muller, I, Schmid, B & Weiner, J (2000) The effect of nutrient availability on biomass allocation patterns in 27 species of herbaceous plants. *Perspectives in Plant Ecology, Evolution and Systematics* 3: 115–127.
- Niklas, KJ (1994) *Plant Allometry. The Scaling of Form and Process*. University of Chicago Press, Chicago.
- Pearl, J (1988) *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Mateo.
- Pearl, J (2000) *Causality*. Cambridge University Press, Cambridge.
- Peters, RH (1983) *The Ecological Implications of Body Size*. Cambridge University Press, Cambridge.
- Royall, R (1997) *Statistical Evidence. A Likelihood Paradigm*. Chapman & Hall, London.
- Shipley, B (1997) Exploratory path analysis with applications in ecology and evolution. *The American Naturalist* 149: 1113–1138.
- Shipley, B (1999) Exploring hypothesis space: examples from organismal biology. *Computation, Causation and Discovery* (eds. C Glymour & GF Cooper), pp. 441–452. MIT/AAAI Press, Menlo Park.
- Shipley, B (2000a) *Cause and Correlation in Biology: A User's Guide to Path Analysis, Structural Equations, and Causal Inference*. Oxford University Press, Oxford.
- Shipley, B (2000b) A new inferential test for path models based on directed acyclic graphs. *Structural Equation Modeling* 7: 206–218.
- Shipley, B (2003) Testing recursive path models with correlated errors using d-separation. *Structural Equation Modeling* 10: 214–221.