Complexity in hydroecological modelling: A comparison of stepwise selection and information theory

Annie Gallagher Visser | Lindsay Beevers | Sandhya Patidar

Institute for Infrastructure and Environment; School of Energy, Geoscience, Infrastructure and Society

Correspondence
Annie Gallagher Visser, Heriot-Watt University, Edinburgh, UK.
Email: av96@hw.ac.uk

Funding Information
Engineering and Physical Science Research Council

Abstract
Understanding of the hydroecological relationship is vital to maintaining the health of the river and thus its ecosystem. Stepwise selection is widely used to develop numerical models which represent these processes. Increasingly, however, there are questions over the suitability of the approach, and coupled with the increasing complexity of hydroecological modelling, there is a real need to consider alternative approaches. In this study, stepwise selection and information theory are employed to develop models which represent two realizations of the system which recognizes increasing complexity. The two approaches are assessed in terms of model structure, modelling error, and model (statistical) uncertainty. The results appear initially inconclusive, with the information theory approach leading to a reduction in modelling error but greater uncertainty. A Monte Carlo approach, used to explore this uncertainty, revealed modelling errors to be only slightly more distributed for the information theory approach. Consideration of the philosophical underpinnings of the two approaches provides greater clarity. Statistical uncertainty, as measured by information theory, will always be greater due to its consideration of two sources, parameter and model selection. Consequently, by encompassing greater information, the measure of statistical uncertainty is more realistic, making an information theory approach more reflective of the complexity in real-world applications.

KEYWORDS
complexity, ecological lag, hydroecological modelling, information theory, regression, statistical uncertainty, stepwise selection, uncertainty

1 | INTRODUCTION

The ecological role of flow is increasingly understood. Rivers are not solely dependent on low flows; they represent extremely variable and dynamic systems (Arthington, 2012; Poff et al., 1997). It is widely acknowledged that the flow regime is a major determinant of the ecological health of river ecosystems (Lake, 2013; Lytle & Poff, 2004; Poff et al., 1997; Poff & Zimmerman, 2010). The inherent complexity makes it challenging to identify and quantify hydroecological relationships.

Numerical modelling is a well-established technique for testing hydroecological hypotheses. Hydroecological models can be developed at different scales, from the single case study river model (Exley, 2006; Visser, Beevers, & Patidar, 2017) with multiple sample sites to models encompassing a given region or particular flow regime (Monk, Wood, Hannah, & Wilson, 2007; Worrall et al., 2014). Ecological data and hydrological (ecologically/biologically relevant) predictors serve as the basis for these models. The ecological component is frequently characterized by macro-invertebrates, fish, or other invertebrates (Bradley et al., 2017).
Hydroecological models are predominantly developed through statistical methods such as regression analysis, including multiple linear regression (e.g., Clarke & Dunbar, 2005, and Monk et al., 2007), and multilevel models (recent examples include Bradley et al., 2017, and Chadd et al., 2017). Algorithms are commonly employed to do the "heavy lifting" in the determination of model structure; in hydroecology specifically, stepwise multiple regression is widely used.

Examples of the use of stepwise multiple regression in hydroecological modelling include Wood, Hannah, Agnew, and Petts (2001) for the identification of hydrological indicators of importance in a groundwater stream; Wood and Armitage (2004) to determine the influence of drought and low flow variability on macro-invertebrate abundance; Knight, Brian Gregory, and Wales (2008) to establish environmental flow requirements; Monk et al. (2007) and Worrall et al. (2014) on a Principal Component Analysis (PCA)-reduced set of hydrological indices; Surridge, Bizzi, and Castelletti (2014) included stepwise selection methods in their development of the iterative input variable selection algorithm (for the development of hydroecological models); Greenwood and Booker (2015) for the identification of important indices in the case of invertebrate response to floods; and Bradley et al. (2017) to realize important terms when considering the effects of groundwater abstraction and fine sediment pressures. Additionally, the authors have previously used a stepwise-based method as part of a preliminary analysis to identify a more complex aspect of the hydroecological relationship with regard to long-term flow variability and lag in ecological response (Visser et al., 2017). Nonstatistical hydroecological modelling is also known to make use of stepwise selection, for example, Parasiewicz et al. (2013) apply stepwise methods in their application of the MesoHABSIM model.

Stepwise methods are attractive, in general, as the statistical theory and assumptions are well established (Whittingham, Stephens, Bradbury, & Freckleton, 2006). Burnham and Anderson (2002) assert that they represent a particularly straightforward and accessible method for the nonstatistician. An algorithm adds and/or subtracts variables (indices) according to identified criteria, stopping once the criterion has been met, resulting in a single, final model. The assumption is that this single model represents the “best” model with the most predictive power.

Increasingly, there is widespread recognition of the limitations of stepwise methods, which have, in the past, been overlooked (Hurvich & Tsai, 1990; Steyerberg, Eijkemans, & Habbema, 1999; Whittingham et al., 2006). A model of a system is, by nature, only ever an approximation of reality; there is no such thing as a true model (Burnham & Anderson, 2002). Coupled with the increasing complexity of hydroecological modelling, the robustness and validity of the statistical approach is critical. In applied statistics, alternative modelling approaches are increasingly favoured, particularly in the ecological sciences (Burnham & Anderson, 2014; Hegyi & Garamszegi, 2011; Stephens, Buskirk, Hayward, & Martínez Del Río, 2005; Wasserstein & Lazar, 2016; Whittingham et al., 2006). Alternate regression methodologies include partial least squares regression, an option when the predictors are not truly independent (common in hydroecological modelling); and shrinkage methods, where penalties/constraints are introduced; ridge and lasso regression can be effective when there are a large number of predictors (Dahlgren, 2010).

Since the beginning of the 21st century, three measures of statistical validity have been identified with unanimity across disciplines: effect size, levels of (statistical) uncertainty, and the weight of evidence supporting the hypothesis (Burnham & Anderson, 2002; Burnham, Anderson, & Huyvaert, 2011; Stephens et al., 2005; Wasserstein & Lazar, 2016; Whittingham et al., 2006). In asking what methods can satisfy these requirements, the field of information theory stands out as the dominant alternative (see position arguments and extensive discussion: Burnham et al., 2011 and Whittingham et al., 2006). Chadd et al. (2017) represents one of the few examples of the application of information theory in hydroecological modelling.

In this paper, the standard hydroecological approach for developing statistical models (stepwise selection) is compared with the increasingly popular information theory, now regularly utilized in applied ecology to investigate which is the most appropriate approach to model the complexities of the hydroecological relationship. Multiple regression models are developed for a groundwater-dominated catchment, where two scenarios of different levels of complexity are considered: The first features standard interannual variables, whereas the second considers lagged ecological response. The performance of each approach in each scenario is assessed.

**FIGURE 1** Left: Location of the River Nar. Right: chalk subcatchment [Colour figure can be viewed at wileyonlinelibrary.com]
2 | METHODS

Models are developed for the groundwater-fed River Nar (Norfolk, UK; Figure 1). All analysis is performed using R (Version 3.4.0), an open source software environment for statistical programming (R Core Team, 2017).

2.1 | Catchment data

The River Nar has a distinctive change at its midpoint, from chalk to fen river. The focus of this paper is the 153.3 km² chalk subcatchment (Figure 1). A reliance on groundwater and aquifer recharge (BFI 0.91) results in a highly seasonal flow regime (Sear, Newson, Old, & Hill, 2005). Aquifer recharge primarily occurs in the winter months, with a progressive rise in flow until March/April.

Daily mean flow data (1990–2014; Figure 2) was extracted from the National River Flow Archive for the Marham gauge (TF723119; Figure 1; NRFA, 2014). The derived hydrological indices describe the magnitude component of the flow regime: high/low flows (Q10/Q90), moderate high/low flows (Q25/Q75), and median flows (Q50). The hydrological indices are considered multiseasonally, with the hydrological year subdivided into the two standard hydrologic seasons, winter (October–March) and summer (April–September).

Macro-invertebrates serve as the proxy for ecological response. Response is determined using the Lotic-Invertebrate Index for Flow Evaluation (LIFE), accounting for macro-invertebrate flow velocity preferences (Extence, Balbi, & Chadd, 1999). Macro-invertebrate sampling data were provided by the Environment Agency for six sites (Figure 1; EA, 2016); the sampling methodology follows the Environment Agency’s standard semi-quantitative protocol (see Murray-Bligh (1999). Seventy-two macro-invertebrate samples, collected in the spring season (April–June, 1993–2012), were used to determine LIFE scores at the species level; see Figure 2 for the average spring LIFE scores during the study period. The ecological data were paired with the antecedent seasonal hydrologic indices.

2.2 | Modelling scenarios

The multiple linear regression modelling approaches are applied to two scenarios. In scenario A, the 10 (interannual) hydrologic indices described previously are considered. Scenario B incorporates ecological lag in response, a reflection of the inherent complexity of the hydroecological relationship. Following Visser et al. (2017), 30 hydrologic indices result from the interannual indices being time-offset up to 2 years (t-2).

2.3 | Stepwise regression

Two methods of stepwise selection are applied, backwards and bidirectional. Being unidirectional, backwards represents greater economy, performing fewer steps to select the smallest model. The algorithms are specified to remove variables which are not significant (alpha threshold = 0.05) and hence presumed unimportant to the hydroecological relationship. Bidirectional stepwise selection is applied using the function step, from the base statistical package stats, whereas the backwards algorithm is applied using the ols_step_backward function from olsrr, a package for the development of ordinary least squares regression models (Hebbali, 2017). These methods yielded the same models, therefore no further differentiation is made.

![Flow duration curve (top) and average spring LIFE scores (bottom) during the study period](image-url)
2.4 Information theory

The information theory approach provides a quantitative measure of support for candidate models. Subsequently, inference is made from multiple models through model averaging. The candidate models are evaluated with respect to the three steps detailed below; for further information, see Burnham and Anderson (2002).

Step 1. Loss of information from model $f$

Kullback-Leibler measures the amount of information lost when model $g$ is used to approximate reality, $f$. The model with the least information loss (greatest supporting evidence of the candidates) is considered the best approximation of reality.

The information loss, $I(f,g)$, is determined through computation of an information criterion. The Akaike Information Criterion (AIC) represents the standard estimate (Burnham & Anderson, 2002). In hydroecological modelling, the sample size is often small relative to the number of variables; here, a second order bias correction, AICc, is used (Burnham & Anderson, 2002).

Step 2. Evidence in support of model $g$

The value of AICc is dependent on the scale of the data; the goal is to achieve the smallest loss of information. This difference is rescaled and ranked relative to the minimum value of AICc:

$$\Delta_i = AIC_{ci} - AIC_{cmin} \text{ for } i = 1, 2, ..., R.$$  \hspace{1cm} (1)

This provides a measure of evidence, from which the likelihood that model $g_i$ is the best approximating model can be determined. This is known as the Akaike weight, $w_i$, ranging from 1 to 0, for the most and least likely models, respectively:

$$w_i = \frac{\exp\left(\frac{1}{2}\Delta_i\right)}{\sum_{r=1}^{R} \exp\left(\frac{1}{2}\Delta_i\right)}.$$  \hspace{1cm} (2)

Step 3. Multimodel inference

The best approximating model is inferred from a weighted combination of all the candidates. Parameter averages, $\hat{\theta}$, are the sum of the Akaike weights for each model containing the predictor, $\hat{\theta}$:

$$\hat{\theta} = \sum_{i=1}^{R} w_i \hat{\theta}_i.$$  \hspace{1cm} (3)

Parameter averages are ranked, such that the highest value represents the most important in the model.

2.5 Package glmulti

There are two options for the application of information theory in R: MuMIn (Barto, 2018) and glmulti (Calcagno, 2013). The application of the former centres around “dredging” (data mining) to determine the model subset (e.g., see Grueber, Nakagawa, Laws, and Jamieson (2011)). The package glmulti offers apposite functionality (see below) and has been developed and applied in a relevant discipline (see). In glmulti, information theory is applied to subsets of models selected by a genetic algorithm (GA) from which the multimodel average is derived using the function coef. A GA is a type of optimization that mimics biological evolution. The GA incorporates an immigration operator, allowing reconsideration of removed variables. Immigration increases the level of randomisation and hence the likelihood of model convergence on the global optima (the best models from the available data) rather than some local optima (Calcagno & de Mazancourt, 2010). Inference from a consensus of five replicate GA runs has been shown by Calcagno and de Mazancourt (2010) to greatly improve convergence.

2.6 Analysis

For each scenario/approach, the best approximating model is derived. The comparative assessment looks at model structure, modelling error and statistical uncertainty.

The analysis of the model structures begins with a review of the selected indices and summary statistics (adjusted $R$-squared and $P$ values). Being evidence-centric, these statistics are at odds with the underlying philosophies of information theory (revisited in the discussion). Instead, importance, the relative weight of evidence in support of each index in the model (Step 3), is considered.

Model error assesses how well the given model simulates the data, here, the observed data. Analysis centres on relative error, defined as the measure of error difference divided by observed value. These errors are presented as an observed-simulated plot. The distribution and magnitude of modelling errors is further considered through probability density functions.

Uncertainty is introduced throughout the modelling process. In this paper, the focus is on statistical uncertainty defined by Warnink et al. (2010, p.1520) as a measure of “the difference between a simulated value and an observation” and “the possible variation around the simulated and observed values,” quantified as 1.96-$\sqrt{\text{variance}}$, where 1.96 represents the 95% confidence level. Simply put the model with the least uncertainty, and hence, the most support should be the best representation of reality. In practice, statistical uncertainty dictates the usefulness of the model. Inaccurate appreciation of this uncertainty, however, prevents meaningful interpretation of the results, leading to less than optimal decision-making (Warnink et al., 2010).

The type of statistical uncertainty quoted is dependent on the modelling approach. For the stepwise approach, parameter (conditional) uncertainty, a measure of the parameter variance in the selected model, is provided. However, model selection represents a further source of statistical uncertainty (Anderson, 2007); when a model is derived from a single data set, there is a chance that other replicate data sets, of the same size and from the same process, would lead to the selection of different models. As a multimodel average, information theory provides a measure for this additional uncertainty, referred to herein as structural uncertainty.

A Monte Carlo approach (MC) is used to explore model parameter space (uncertainty at the 95% confidence interval represents the upper/lower bounds). Traditional MC methods suffer from clumping of points; this occurs because the points “know” (Caflisch, 1998)
nothing about each other. To reduce the number of simulations required, a Quasi-MC method (Sobol-sequence) is applied, where elements are correlated and more uniformly well-distributed; 200 simulations appeared sufficient. The relative error distributions (based on the observed data) are again plotted. An extract of these plots, at the 5/50/95% densities illustrates the error distribution across the simulations.

3 | RESULTS

3.1 | Scenario A

3.1.1 | Model structure

The structure of the best approximating models is detailed in Table 1 and Figure 3 (facet 1). The information theory multimodel average, features five hydrologic indices, with a focus on low flows in summer and winter (Qs90 and Qw90). The stepwise selected model is similar, except here the Qs90 index is not present, with this model favouring less extreme low flows (Qs75).

Summary statistics for the two best approximating models are detailed in Table 1, with both achieving similar adjusted R-squared values. The P value, the principal selection characteristic in stepwise approaches, is distinctly lower in the information theory model. The second and third best-performing stepwise selection models saw the removal of the Qs25 index, and then Qw10 in the final step. These models have a similar fit to the selected model, with adjusted R-squared values of 0.52 and 0.55, respectively. For the information theory model, an estimate of the relative weight of evidence in support of each index (Figure 3, facet 1) suggests that the winter hydrologic indices are the most meaningful.

3.1.2 | Model error

Modelling errors are presented in Figure 4. Overall, there appear to be minimal differences between the two approaches, with the stepwise selected model featuring marginally less error. In Figure 4a, it can be seen that the models perform slightly worse at the extremes, with the stepwise model achieving a slightly better fit overall. This is further evidenced in Figure 4b, where errors can be seen to concentrate on the left. Finally, the fitted distributions in Figure 4c feature considerable overlap, further emphasizing the similarities in model performance.

3.1.3 | Model uncertainty

The statistical uncertainty, relative to the parameter estimate, is summarized in Figure 5 (facet 1); the stepwise selected model

![Table 1: Model structures and summary statistics](attachment:table1.png)

![Figure 3: Model structure and estimates of parameter coefficients](attachment:figure3.png)
displays the least uncertainty. Differences are most notable in hydrological summer, suggesting greater confidence in the winter indices; this is in agreement with the information theory importance statistic.

Further inference regarding the implications of statistical uncertainty is made through the consideration of MC simulations (Figure 6). The cumulative density function (fitted to a normal distribution; Figure 6a,b) for each simulation provides an overview of the errors. This is further clarified in Figure 6c), where the errors at cumulative densities of 5/50/95% indicate the distribution of error across the simulations. For 5% of the data, the majority of the simulations feature 2.5% absolute error or less; this represents approximately 9% (stepwise) and 16% (information theory) of the simulations. At 50/95, the errors are similarly spread; the majority of stepwise models have approximately 20% error, whereas for information theory, this is 27.5%.

**FIGURE 4** Scenario A modelling error. Observed simulated LIFE scores (left); probability density functions (fitted to a normal distribution) of relative error (top right); absolute relative error cumulative density functions (bottom right) [Colour figure can be viewed at wileyonlinelibrary.com]

**FIGURE 5** Uncertainty (95% confidence interval) relative to parameter estimates. For scenario B, each facet indicates a time-offset [Colour figure can be viewed at wileyonlinelibrary.com]
3.2 Scenario B

3.2.1 Model structure

Here, the differences between model structures are greater than Scenario A (Table 1 and Figure 3, facets 2–4). The stepwise selected model incorporates two nonlagged and two lagged indices. The two nonlagged parameters represent summer median and moderate low flows (Qs50 and Qs75). The large coefficients of these two parameters suggest a preference for mid range flows which are not too low or high; in this, the scenario B model is broadly consistent with scenario A. However, the model takes no account of winter flows. In contrast, the information theory model structures (and measures of parameter importance) for both scenarios are similar, with the only difference being the inclusion of lagged winter high flow (t−2). Physically, this could represent the time delay of the groundwater recharge. There is no acknowledgement of this phenomenon in the stepwise selected model, whether subject to lag or otherwise. In this scenario, the summary statistics (Table 1, rows 3 and 4) associated with the stepwise model remain relatively static. However, the adjusted R-squared for the information theory model is 14% greater than the stepwise model.

Overall, the information theory model indicates a preference for variability in flow magnitude, possibly a reflection of the seasonal nature of the flow regime. Winter flows stand out as the most important facet of the flow regime. In contrast, the stepwise selected model suggests a preference for more uniform flows (that are not too low); unusually, winter flows are considered unimportant.

3.2.2 Model error

The errors associated with each model are detailed in Figure 7. At first glance, Figure 7a suggests that the models perform equally well for lower LIFE scores, whereas for higher values the information theory model provides marginally better estimates. This is reinforced in Figure 7b, where the relative errors are centred around 0% and −4% for the information theory and stepwise models, respectively. The stepwise model also has a tendency to overestimate. The extent of these differences is evident in Figure 7c. For the information theory model, 56% of the estimated data points have 5% or less absolute relative error, in fact, almost 50% of the data has less than 2.5%. This is in direct contrast to the stepwise selected model, where only 15% of the data has less than 2.5% absolute relative error; this increases to approximately 48% at 5%. The models do not converge until approximately 9.25% absolute relative error, that is, the largest errors for both models are comparable.

3.2.3 Model uncertainty

Relative to scenario A, there is an increase in the range of statistical uncertainty (Figure 5, facets 2–4), particularly for the information...
FIGURE 7  Scenario B modelling error. Observed simulated LIFE scores (left); probability density functions (fitted to a normal distribution) of relative error (top right); absolute relative error cumulative density functions (bottom right) [Colour figure can be viewed at wileyonlinelibrary.com]

FIGURE 8  Scenario B, distribution of modelling errors following MC simulation. (a,b) Cumulative density function of the absolute relative error per simulation (fitted to a normal distribution); (c) distribution of the absolute relative error for 5/50/95% of the data [Colour figure can be viewed at wileyonlinelibrary.com]
theory model. Figure 8a,b shows the information theory MC simulations to be more widely distributed than the stepwise. A snapshot of the error distributions at cumulative densities of 5/50/95% is shown in Figure 8c. It is noteworthy that, here, the range of densities on the y-axis is narrower than in scenario A. The difference is more marked at the 50% and 95% densities, where the distribution of the information theory simulations is flatter and wider, indicating a greater spread of error; in contrast, the error in the stepwise simulations tends towards the lower end.

4 | DISCUSSION

The initial focus herein is model inference and consideration of the explicit implications of the results. To gain further information on the relative strengths and weaknesses of the two approaches, it is necessary to look beneath the surface. The statistical robustness of the models produced is considered, as well as the underlying philosophies of each approach.

4.1 | Model inference

In scenario A, the principle difference in model structure is the parameterisation of summer low flows; information theory focuses on low flows (Q90) rather than moderate low flows (Q75). Consequently, the differences in modelling error is small. Consideration of statistical uncertainty and the error distributions reveals the stepwise selected model to be more balanced in terms of error distribution.

Despite demonstrated importance, as a groundwater-fed river (Sear et al., 2005), aquifer recharge is not recognized under the scenario B stepwise selected model. There is no consideration of hydrological winter and a low number of parameters overall; concerns thus emerge over the parameterisation of the stepwise selected model in this more complex scenario. In contrast, the information theory model includes seven hydrological indices, three of which reflect winter flows. The importance of these indices is further emphasized by the relatively high weight of evidence (Figure 3, facets 2–4).

Given the difference in model structure, the similarities in modelling errors are unexpected. Interestingly, with the information theory model, the shape of the error distributions is consistent across the two scenarios, it is only the magnitude of the error that varies (increasing in the lagged scenario). In contrast, the stepwise selected approach sees an increase in error.

The increased uncertainty in scenario B can be considered a direct consequence of the increased modelling complexity. Figure 5 (facets 2–4) suggests that the stepwise model is subject to less uncertainty. However, the MC simulations (Figure 8) show that the associated error distributions are similar with regard to shape. However, the information theory curve is slightly flatter, leading to errors of higher magnitude.

Based on these findings, it could be concluded that these two hydroecological modelling approaches perform at similar levels. The principal area for concern may be the increase in statistical uncertainty for the information theory model in the lagged scenario. However, it should be noted that the reasons for the increased uncertainty are multifaceted, a matter discussed further below. Despite differences in model structure and statistical uncertainty, all models, and hence both approaches, have been able to provide satisfactory predictions with comparable modelling error.

4.2 | Philosophical underpinnings

Looking to the statistical robustness of the approaches, and underlying philosophies, may serve to further elucidate which method is most appropriate in the hydroecological setting. Considered herein are model selection, the definition of evidence, and statistical uncertainty.

4.2.1 | Model selection

In scenario A, the stepwise model was selected following six steps, that is, six hydroecological models were considered. In terms of their summary statistics, the models considered in the fourth and fifth steps are remarkably similar to the selected model; despite this, under this methodology, these models are rejected. Consequently, in the baseline scenario, hydrological indices capturing summer high flows (Q10 and Q25) were rejected as model parameters. It could be argued that, by simply making this observation, that this is an elementary form of multimodel inference. In practice, however, these second and third ranking models would not be subject to analysis, and thus, such information would be left unknown. Consequently, it is inferred that the selected model is the only model fit (Burnham & Anderson, 2002). In the alternate approach, information theory considers a larger candidate set, the result being a model-average. Consequently, important variables have not been subject to rejection. This is reinforced by the index of importance (Figure 3), an indication of the relative “importance” of each parameter. By calculating and reporting this statistic, it is evident that the variables incorporated in the model are those which are most supported by the data. Consequently, more conclusive statements may be made with regard to the model. For example, in scenario A, it would not be incorrect to state that, given the data, low flows are more important than high flows in the hydroecological relationship for this case study river. Such conclusions would be pure conjecture in the case of the stepwise selected model.

4.2.2 | Evidence

The use of P values has been subject to considerable criticism in excess of 80 years (Burnham & Anderson, 2002). In the context of hydroecological modelling, the fundamental problem is with misinterpretation, where P values are interpreted as evidential. In a statistical sense, the P value is a measure of the probability that the effect seen is a product of random chance. Probability is a measure of uncertainty, not a measure of strength of evidence, which is based on likelihood (Burnham & Anderson, 2002).

This misinterpretation is not exclusive to hydroecological modelling or stepwise selection, it is prevalent in academia (for example, see Wasserstein and Lazar (2016)). As such, this is such an ingrained error that it cannot be viewed as a criticism of the hydroecological modeller. In their paper on the development of the LIFE
hydroecological index (Extence et al., 1999, p. 558), the authors fall into this trap:

“At Brigsley on the Waite Beck, for example, there are 177 separate correlation coefficients significant at $p < 0.001$, 13 at $p < 0.005$, six at $p < 0.01$, ten at $p < 0.05$ and eight correlations that are non-significant, for the period 1986–1997. From this surfeit of usable statistics, those flow variables showing the best relationships with the invertebrate fauna are proposed as being of primary importance in determining community structure in particular river systems.”

Here, the authors interpreted the $P$ value as a weight of evidence, assuming that those 177 models with the lowest $P$ values were “best.” Such explicit use of $P$ values is no longer commonplace; however, the stopping rule applied in stepwise selection does utilize $P$ values in precisely this manner, a practice which is described by Burnham and Anderson (2002, p.627) as “perhaps the worst” application of $P$ values.

The misunderstanding of the definition and purpose of the $P$ value raises concerns with its use in hydroecological modelling. Questions are therefore raised over the statistical robustness, or accuracy, in the application of stepwise methods, and thus, its ability to recognize the inherent complexities of the hydroecological relationship, as well as the selection of the final model. In this case, it could be argued that as an evidence-based methodology, information theory offers clearer, more robust statistical inference. Indeed, this is recognized by two of the authors of the 1999 LIFE paper, who have recently looked to information theory when developing and applying a new index, the Drought Effect of Habitat Loss on Invertebrates (DELHI; Chadd et al. (2017)).

4.2.3 | Statistical uncertainty

Statistical uncertainty is the principal determinant of the usefulness and validity of a model. As suggested previously, given the lower uncertainty, it might be concluded that the stepwise approach performs better overall, particularly in the case of the more complex scenario B, where the uncertainty increases further. However, as discussed under methods, these two modelling approaches report different statistical uncertainties. The stepwise model considers parameter uncertainty, whereas the information theory model also quantifies error due to model selection, thereby providing a measure of the overall structural uncertainty. The subsequent higher uncertainty simply represents a more realistic measure, as Anderson, 2007 (p. 113) points out, when only parameter uncertainty is considered, the “confidence intervals are too narrow and achieved coverage will often be substantially less than the nominal level (e.g., 95%).”

5 | CONCLUSION

As further aspects of hydroecological relationships are understood, such as ecological lag in response, the likelihood of modelling errors and statistical uncertainty is increased, commensurate with the additional complexity. It is thus vital to ensure the modelling approach is suitably robust. Here, the performance of stepwise selection, one of the standard hydroecological approaches, is considered alongside an alternative popular in applied statistics, information theory. The best approximating models are analysed comparatively. The approaches are applied to two scenarios with increasing complexity; scenario A, focusing on standard interannual variables, and scenario B, taking into account any effect of lag in ecological response.

Notable differences in the models are confined to the lagged scenario. Of foremost concern is the structure of the stepwise selected model. Aquifer recharge is fundamental to flow in ground-water-fed rivers, which is a feature of the case study river examined. In this paper, this physical property is assumed to be represented by the winter variables. In scenario A, this is accounted for through two winter low flow variables, $Q_{w75}$ and $Q_{w90}$. This is repeated in scenario B for the information theory model, plus an additional lagged high flow variable. Despite their recognized importance, the stepwise selected model includes no winter variables, leading to concerns over its ability to capture the essential physical processes in such complex scenarios.

In terms of model performance, the information theory approach resulted in fewer modelling errors but in greater statistical uncertainty. Despite this, the measure of uncertainty provided by stepwise selection is considered an underestimate (Burnham & Anderson, 2002) as the variance due to model selection is not incorporated; the estimate considers only parameter variance. It may seem contradictory to say that the model subject to greater uncertainty provides the better measure; however, the stepwise selected model inherently suffers from confidence intervals which are too narrow, with the achieved coverage being less than the standard, nominal 95% (Anderson, 2007).

From a utilitarian perspective, one might say that no approach has been demonstrated to be categorically better than the other. However, modelling is only ever an approximation of reality, and the best, true model remains an unknown. Still, in approaching the truth, we must have some criteria to adjudge success, and here, these have been identified as approaches which focus on effect size, statistical uncertainty, and the weight of evidence. Based on the results presented here, information theory satisfies these three best. In contrast, stepwise selection offers a $P$ value, an arbitrary probability measure, which is not a measure of effect size. The uncertainty it measures is an underestimate and weight of evidence is not possible under this approach.

Finally, though no approach emerged a clear “winner,” the information theory model still performed empirically better in the two scenarios considered. It presented significantly fewer modelling errors, and although the measure of statistical uncertainty is larger, and thus inconvenient, it may also be viewed as a truer representation of a complex reality, than that provided by its stepwise counterpart.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge funding from the Engineering and Physical Science Research Council through award. Further thanks go
to the Environment Agency and the Centre for Ecology and Hydrology for the provision of data.

ORCID

Annie Gallagher Visser  http://orcid.org/0000-0003-1787-6239

REFERENCES

Anderson, D. R. (2007). Model based inference in the life sciences. New York: Springer.

Arthington, A. H. (2012). Chapter 9. Introduction to Environmental Flow Methods. In: Environmental Flows: Saving Rivers in the Third Millennium. California: University of California Press.

Bartoń, K. (2018). Package ‘MuMIn’ Version 1.40.4. Retrieved from https://cran.r-project.org/web/packages/MuMIn/

Bradley, D. C., Streetly, M. J., Cadman, D., Dunscombe, M., Farren, E., & Banham, A. (2017). A hydroecological model to assess the relative effects of groundwater abstraction and fine sediment pressures on riverine macro-invertebrates. River Research and Applications., 33, 1630–1641. https://doi.org/10.1002/rra.3191

Burnham, K. P., & Anderson, D. (2002). Model selection and multi-model inference: A practical information-theoretic approach. New York: Springer.

Burnham, K. P., Anderson, D. R., & Huyvaert, K. P. (2011). AIC model selection and multimodel inference in behavioral ecology: Some background, observations, and comparisons. Behavioral Ecology and Sociobiology, 65, 23–35. https://doi.org/10.1007/s00265-010-1029-6

Burnham, K. P., & Anderson, D. R. (2014). P-values are only an index to evidence: 20th- vs. 21st-century statistical science. Ecology, 95(3), 627–630.

Calisch, R. E. (1998). Monte Carlo and quasi-Monte Carlo methods. Acta Numer., 7, 1–49. https://doi.org/10.1017/S0962492900002804

Calcagnò, V. (2013). glmulti: Model selection and multimodel inference made easy. Version 1.0.7. Retrieved from https://cran.r-project.org/package=glmulti

Calcagnò, V., & de Mazancourt, C. (2010). glmulti: An R package for easy automated model selection with (generalized) linear models. Journal of Statistical Software, 34, 1–29.

Chadd, R. P., England, J. A., Constable, D., Dunbar, M. J., Exteche, C. A., Leeming, D. J., ... Wood, P. J. (2017). An index to track the ecological effects of drought development and recovery on riverine invertebrate communities. Ecological Indicators, 82, 344–356. https://doi.org/10.1016/j.ecolind.2017.06.058

Clarke, R., & Dunbar, M. (2005). Producing generalised LIFE response curves. Bristol: Environment Agency.

Dahlgren, J. P. (2010). Alternative regression methods are not considered in Murphy (2009) or by ecologists in general. Ecology Letters, 13, E7–E9. https://doi.org/10.1111/j.1461-0248.2010.01460.x

EA. (2016). River Nar macroinvertebrate monitoring data. (Available upon request from the EA.).

Exley, K. (2006). River Itchen macro-invertebrate community relationship to river flow changes. Winchester: Environment Agency.

Exteche, C. A., Balbi, D. M., & Chadd, R. P. (1999). River flow indexing using British benthic macroinvertebrates: A framework for setting hydroecological objectives. Regulated Rivers: Research & Management, 15, 545–574. https://doi.org/10.1002/rra.545

Greenwood, M. J., & Booker, D. J. (2015). The influence of antecedent floods on aquatic invertebrate diversity, abundance and community composition. Ecohydrology, 8, 188–203. https://doi.org/10.1002/eco.1499

Gruber, C. E., Nakagawa, S., Laws, R. J., & Jamieson, I. G. (2011). Multimodel inference in ecology and evolution: Challenges and solutions. Journal of Evolutionary Biology, 24, 699–711. https://doi.org/10.1111/j.1420-9101.2010.02210.x

Grubbs, F., & Tóth, L. (2018). Choosing a multivariate distribution: A practitioner’s guide. Available: https://CRAN.R-project.org/package=multiv

Hebbali, A. (2017). oslrr: Tools for Teaching and Learning OLS Regression. Version 0.3.0. Available: https://CRAN.R-project.org/package=oslrr

Hegyi, G., & Garamszegi, L. Z. (2011). Using information theory as a substitute for stepwise regression in ecology and behavior. Behavioral Ecology and Sociobiology, 65, 69–76. https://doi.org/10.1007/s00265-010-1036-7

Hurvich, C. M., & Tsai, C.-L. (1990). The impact of model selection on inference in linear regression. The American Statistician, 44, 214–217. https://doi.org/10.2307/2685338

Knight, R. R., Brian Gregory, M., & Wales, A. K. (2008). Relating streamflow characteristics to specialized invertevores in the Tennessee River Valley: A regional approach. Ecohydrology, 1, 394–407.

Lake, P. S. (2013). Resistance, resilience and restoration. Ecological Management & Restoration, 14, 20–24. https://doi.org/10.1111/emr.12016

Lytle, D. A., & Poff, N. L. (2004). Adaptation to natural flow regimes. Trends in Ecology & Evolution, 19, 94–100. https://doi.org/10.1016/j.tree.2003.10.002

Munk, W. A., Wood, P. J., Hannah, D. M., & Wilson, D. A. (2007). Selection of river flow indices for the assessment of hydroecological change. River Research and Applications, 23, 113–122. https://doi.org/10.1002/rra.964

Murray-Blyth, J. A. (1999). Quality management systems for environmental monitoring: Biological techniques, BT001. Procedure for collecting and analysing macro-invertebrate samples. Version 2.0. Retrieved from Bristol:

NRFA. (2014). Marham gauge daily flow data. (Available upon request from the NRFA).

Parasiewicz, P., Rogers, J. N., Vezza, P., Gortazar, J., Seager, T., Pegg, M., ... Comoglio, C. (2013). Applications of the MesoHABSIM Simulation Model. In I. Maddock, A. Harby, P. Kemp, & P. Wood (Eds.), Ecohydrodynamics: An integrated approach (pp. 109–124). John Wiley & Sons, Ltd.

Poff, N. L., Allan, J. D., Bain, M. B., Karr, J. R., Prestegaard, K. L., Richter, B. D., ... Stromberg, J. C. (1997). The natural flow regime. Bioscience, 47, 769–784. https://doi.org/10.2307/1313099

Poff, N. L., & Zimmerman, J. K. H. (2010). Ecological responses to altered flow regimes: A literature review to inform the science and management of environmental flows. Freshwater Biology, 55, 194–205. https://doi.org/10.1111/j.1365-2427.2009.02272.x

R Core Team. (2017). R: A language and environment for statistical computing. Retrieved from https://www.r-project.org/

Sear, D.A., Newson, M., Old, J.C., & Hill, C. (2005). Geomorphological appraisal of the River Nar Site of Special Scientific Interest. (N684): English Nature.

Stephens, P. A., Buskirk, S. W., Hayward, G. D., & Martínez Del Rio, C. (2005). Information theory and hypothesis testing: A call for pluralism. Journal of Applied Ecology, 42, 4–12. https://doi.org/10.1111/j.1365-2427.2005.02002.x

Steyerberg, E. W., Eijkemans, M. J., & Habbema, J. D. (1999). Stepped selection in small data sets: A simulation study of bias in logistic regression analysis. Journal of Clinical Epidemiology, 52, 935–942.

Surridge, B. W. J., Bizzi, S., & Castelletti, A. (2014). A framework for coupling explanation and prediction in hydroecological modelling. Environmental Modelling & Software, 61, 274–286. https://doi.org/10.1016/j.envsoft.2014.02.012

Visser, A., Beevers, L., & Patidar, S. (2017). Macro-invertebrate community response to multi-annual hydrological indicators. River Research and Applications, 33, 707–717. https://doi.org/10.1002/rra.3125

Warnick, J. J., Janssen, J. A. E. B., Booij, M. J., & Krol, M. S. (2010). Identification and classification of uncertainties in the application of environmental models. Environmental Modelling & Software, 25, 1518–1527. https://doi.org/10.1016/j.envsoft.2010.04.011

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA’s statement on p-values: Context, process, and purpose. The American Statistician, 70, 129–133. https://doi.org/10.1080/00031305.2016.1154108
Whittingham, M. J., Stephens, P. A., Bradbury, R. B., & Freckleton, R. P. (2006). Why do we still use stepwise modelling in ecology and behaviour? *The Journal of Animal Ecology, 75*, 1182–1189. https://doi.org/10.1111/j.1365-2656.2006.01141.x

Wood, P. J., & Armitage, P. D. (2004). The response of the macroinvertebrate community to low-flow variability and supra-seasonal drought within a groundwater dominated stream. *Archiv für Hydrobiologie, 161*, 1–20. https://doi.org/10.1127/0003-9136/2004/0161-0001

Wood, P. J., Hannah, D. M., Agnew, M. D., & Petts, G. E. (2001). Scales of hydroecological variability within a groundwater-dominated stream. *Regulated Rivers: Research & Management, 17*, 347–367. https://doi.org/10.1002/rrr.658

Worrall, T. P., Dunbar, M. J., Extence, C. A., Laizé, C. L. R., Monk, W. A., & Wood, P. J. (2014). The identification of hydrological indices for the characterization of macroinvertebrate community response to flow regime variability. *Hydrological Sciences Journal, 59*, 645–658. https://doi.org/10.1080/02626667.2013.825722

**How to cite this article:** Visser AG, Beevers L, Patidar S. Complexity in hydroecological modelling: A comparison of stepwise selection and information theory. *River Res Applic*. 2018;34:1045–1056. https://doi.org/10.1002/rra.3328