COMPLIMENTARY/POSTER SESSION PAPER

Detecting Mismatch in Functional Narratives of Animal Morphology: A Test Case with Fossils

Nicholas Hebdon*, P.David Polly†, David Joseph Peterman† and Kathleen A. Ritterbush‡

*Department of Biological Sciences, Chapman University, Keck Center, 450 North Center Street, Orange, CA 92866, USA; †Departments of Earth & Atmospheric Science, Biology, and Anthropology, Indiana University, Bloomington, IN 47405, USA; ‡Department of Geology and Geophysics, University of Utah, Frederick Albert Sutton Building, 115 S 1460 E, Salt Lake City, UT 84112-0102, USA

1E-mail: nicholas.hebdon@gmail.com

Synopsis  A boom in technological advancements over the last two decades has driven a surge in both the diversity and power of analytical tools available to biomechanical and functional morphology research. However, in order to adequately investigate each of these dense datasets, one must often consider only one functional narrative at a time. There is more to each organism than any one of these form–function relationships. Joint performance landscapes determined by maximum likelihood are a valuable tool that can be used to synthesize our understanding of these multiple functional hypotheses to further explore an organism’s ecology. We present an example framework for applying these tools to such a problem using the morphological transition of ammonoids from the Middle Triassic to the Early Jurassic. Across this time interval, morphospace occupation shifts from a broad occupation across Westermann Morphospace to a dense occupation of a region emphasizing an exposed umbilicus and modest frontal profile. The hydrodynamic capacities and limitations of the shell have seen intense scrutiny as a likely explanation of this transition. However, conflicting interpretations of hydrodynamic performance remain despite this scrutiny, with scant offerings of alternative explanations. Our analysis finds that hydrodynamic measures of performance do little to explain the shift in morphological occupation, highlighting a need for a more robust investigation of alternative functional hypotheses that are often intellectually set aside. With this we show a framework for consolidating the current understanding of the form–function relationships in an organism, and assess when they are insufficiently characterizing the dynamics those data are being used to explain. We aim to encourage the broader adoption of this framework and these ideas as a foundation to bring the field close to comprehensive synthesis and reconstruction of organisms.

Introduction

Biomechanical and functional morphological analyses of extinct organisms seek to reconstruct its ecology using an understanding of how its form interacted with its environment. Recent advances in the capabilities of our technological tools for approaching these questions have paved the way for rapid growth in our understanding of functional performance in animal systems. Computational approaches leveraging machine learning, computational fluid dynamics (CFD), and finite element analysis (FEA) studies all help to thoroughly explore study systems, generating vast amounts of data across many potential variables of interest (Anderson et al. 2012; Hebdon et al. 2020b; Peterman and Ritterbush 2022). Meanwhile hardware advancements, such as increasingly small and sophisticated components for measurement and robotics, help to establish and explore new and increasingly intricate hypotheses (Nyakatura et al. 2019; Johnson et al. 2021; Peterman and Ritterbush 2022). However, these tools are often best equipped to analyze one structure-function relationship at a time, which makes it difficult to evaluate competing hypotheses about which function played a more important role in the evolution of the organism’s form or whether there were trade-offs between two or more function. These piecewise comparisons are insufficient to interrogate
macroevolutionary trends where morphological fitness is a moving target compromising between performance trade-offs and an ever-changing environment.

**Ammonite functional studies navigate a murky environment**

The limitations of prior approaches are exemplified in the many competing hypotheses of what drove the morphological shift in planispiral ammonoid cephalopods following the End Triassic mass extinction, the relative strengths of which are difficult to evaluate (Dommergues et al. 2002; Ritterbush and Bottjer 2012; Smith et al. 2014; Longridge and Smith 2015). We focus this case study on these shelled cephalopods for two reasons: first, they have one of the best-studied fossil records of any ocean macrofauna; and, second, their architecture presents a finite number of potentially functional traits. Externally shelled cephalopods today are represented only by a few species of nautiloids (most recognizable is *Nautilus pompilius*, the Chambered Nautilus) that persist in the Indo-Pacific. But fossils show us that externally shelled cephalopods flourished across the globe in the time of dinosaurs and earlier. Nautiloids and their sister clade, the ammonoids, rose to great abundance in the mid Paleozoic, and remained major constituents of ocean ecosystems until the ammonites’ complete extinction at the end of the Mesozoic (due to the Cretaceous–Paleogene mass extinction, which also extinguished all non-avian dinosaurs). Ammonoid fossils are particularly abundant and valuable to science; they are present on every continent, are quarried commercially, and provide vital index fossils that help geologists correlate strata across the globe (Monnet et al. 2015; Abdelhady et al. 2019). In addition to detailed documentation of their diversity and biogeography, ammonoids’ conch morphology has been a persistent source of speculation for clues to the animals’ life mode.

Intense interest in the life habits of externally shelled cephalopods has persisted for nearly a century due to their deceptively simple architecture and dynamic fossil record of morphology. The conch itself is straightforward in architecture: A tube that expands at the aperture is divided by internal walls into a series of compartments for buoyancy (the phragmocone) and the remaining open-ended portion housing the animal’s soft tissues (the body chamber). The tube can grow as a straight, slender cone; as a planispiral (spiral in a single plane, without snail-like torsion); or as any combination of helixes, bends, and hooks. Planispiral conchs are the most common as fossils from the Devonian through Cretaceous, and have long gained attention for shape characteristics that different clades seem to mix-and-match across macroevolutionary timescales. The complexity of interior walls that divide the phragmocone have been the basis of much taxonomy, biostratigraphy, and phylogeny. Functions ascribed to these frilled structures include resistance to hydrostatic pressure (refuted; Lemanis 2020; Lemanis et al. 2020); resistance to predation (Lemanis et al. 2016); and manipulation of internal fluid (Peterman et al. 2021). The overall external shape of planispiral conchs can range from an extremely compressed disk to a fully inflated orb, for which evolutionary and developmental plasticity is conceived as a set of independently variable growth modes (Tendler et al. 2015; Parent et al. 2020). Different conch shapes are linked to varied propulsive efficiency and maneuverability (Bayer 1982; Jacobs 1992). Finally, external ornaments (ribs, keels, spines, etc.) are interpreted to add predation resistance, or hydrodynamic stability (Bayer 1982; Westermann 1996; Peterman and Ritterbush, 2022). Recent advances in biomechanics methods applied to the fossil record have targeted cephalopods due in part to their prominent history in paleoecological speculation.

Cephalopods are well-suited to hydrodynamic analysis methods development because the conch can be evaluated as a single rigid structure that forms the interface between animal and fluid. After the first computer models of planispiral conch growth were presented in the 1960s, experiments to distinguish the hydrodynamic consequences of conch shape continued throughout the twentieth century (Raup 1966, 1967). In the current century, increased accessibility of three-dimensional (3D) modeling software and fabrication equipment has advanced experimentation methods and computer simulations alike. Experimental specimens can be made to represent replicas of specific fossils, idealized conch morphologies, or hybrids of the two. Production of such replicas has even been made tremendously easier and cheaper with the surge in 3D printing technology (Johnson and Carter 2019). Improved instrumentation and machine learning capture unprecedented detail of experiments in flume tanks, static water chambers, and open water settings (Neil and Askew 2018; Peterman and Ritterbush 2022). Computer simulations, meanwhile, can query the aquatic performance of the very same 3D models used in experimental set-ups via computational fluid dynamics (Hebdon et al. 2020a, 2020b; Peterman et al. 2020). Modern experiments and simulations allow rapid and iterative hypothesis testing, and refining which attributes attributed to conch shape hold up under scrutiny. But the wealth of quantitative results, and the disparate methods used, make it hard to arrive at a comprehensive synthesis of conch shape impacts on locomotion.
Cephalopod conch shapes vary radically across geologic time and biogeography, and trends in collective or lineage-specific morphology are increasingly interpreted as signals of paleoecological change. Cephalopod diversity produces dramatic boom-and-bust cycles, with regional and global mass extinction obliterating lineages piecemeal or wholesale. Intense scrutiny has focused on the conch shapes that ammonoids produce—and their ecological significance—in the wake of regional biogeographic changes (Yacobucci 2004a, 2004b), widespread extinction pulses (Dera et al. 2010), and global mass extinction events (McGowan and Smith 2007; Smith et al. 2014). We select as a case study the contrast in conch morphology that ammonoids expressed well before, and right after, the global mass extinction that distinguishes the Triassic from Jurassic Periods, using the shell ensembles of Ritterbush and Bottjer (2012). As a “pre-extinction” case, this includes a variety of ammonoid conchs from the Anisian Stage of the Middle Triassic, which is long enough after the Permo–Triassic extinction, yet not affected by the distantly approaching Triassic–Jurassic mass extinction. As the “post-extinction” case, the focus is on ammonoids from the Hettangian Stage of the Early Jurassic. Over this interval ammonoids shift from a broad occupation of Westermann morphospace (which characterizes planispiral ammonoid morphologies specifically) to occupying a very tight zone of the morphospace, emphasizing relatively high umbilical exposure and moderate lateral compression (Ritterbush and Bottjer 2012; Smith et al. 2014).

This transition has been intensely scrutinized by paleobiological researchers (Dommergues et al. 2002; Ritterbush and Bottjer 2012; Smith et al. 2014; Ritterbush 2015). Some studies have concluded that based on drag coefficient estimates of these morphologies that swimming must have been de-emphasized as a performance factor because of the relatively high umbilical exposure morphotypes of the Jurassic have previously been associated with high drag coefficients. More recent studies that have leveraged CFD and computing have concluded that in actuality the morphospace occupation appears to strongly emphasize a balance of between two different swimming goals; distance achieved relative to effort and acceleration. Some have also claimed that some non-swimming performance target is driving the emphasis on high umbilical exposure morphs. These many competing hypotheses all leverage trade-offs in performance and how different types of performance may be balanced and emphasized. Previously, pareto optimality indices were applied to these questions more broadly to weigh in on these different lines of inquiry (Tendler et al. 2015). Since then, however, new and more nuanced data from CFD analyses have offered updated hydrodynamic performance data with more detail and nuance, which have overturned previous relationships and assumptions of hydrodynamic performance in these groups (Hebden et al. 2020a, 2022).

**Likelihood: a light through the fog**

In this study, we aim to show how the competing functional consequences of evolutionary change in ammonoid conch morphology can be reconciled by adapting the maximum-likelihood methods of Polly et al. (2016) to a Westermann shell shape space. This method uses individual data sets of performance data to construct “surfaces” that model how performance varies across a defined morphospace as a function of the initial data set. These performance surfaces are then aggregated into a joint performance surface with each individual performance possessing a weighting coefficient to describe its relative importance in the system. Maximum-likelihood methods are then used to find the weight coefficients that maximize the value of the joint performance surface for a chosen set of points in the morphospace.

By applying these methods to our ammonoid data set, we can determine the relative importance of any number of performance metrics in shaping how the morphospace is occupied between any two intervals, in this case the Triassic versus the Jurassic. Changes in how different performance models are weighted in the joint performance surface can then be interpreted more readily.

An advantage of this approach is that it can also be used to demonstrate when the available functional data are insufficient for explaining the observed occupation of the morphospace, an important question that was often relegated to an afterthought in the discussion sections many previous studies. In an ideal study, one would identify a suite of performance factors that jointly explain the range of observed morphologies of a group, weighted such that they produce a combined performance peak near the center of the observed morphologies in a morphospace that is broad enough to encompass them all. But it is possible that the performances under consideration cannot be combined in a way that encompasses the realized morphologies, suggesting that some as yet unidentified factor is necessary to explain their evolution. This approach can pinpoint which functional hypotheses require more clarity, or even whether a particular function is appropriate to explain the observations.

**Methods**

**Measuring performance with CFD**

We use 19 idealized 3D model ammonoid morphologies in this study that are widely distributed through-
For each shell, three quantities were calculated or measured as proxies for different functional scenarios: (1) meters traveled per newton of thrust exerted at a diameter of 20 cm ($\Delta m/N$; henceforth referred to as coasting efficiency for simplicity), (2) volume accretion per unit surface area accreted ($\Delta V/\Delta SA$), which we will call volume accumulation, and (3) diameter accretion per unit of surface area ($\Delta D/\Delta SA$), which we will call diameter accumulation. Coasting efficiency is derived from the hydrodynamics data and methods of Hebdon et al. (2022). Coasting efficiency is not well described by any form of linear model and is therefore not easily summarized across all sizes for any one morphotype. Therefore, in this study, we chose to use the coasting efficiency of each shell at a size of diameter 20 cm, which encompasses most of the largest ammonoids from our interval of interest and maximizes the possible differences in this parameter between two morphologies. This coasting parameter is closely related to the more commonly considered coefficient of drag, which essentially describes an object’s streamlining with a single dimensionless number. Low coefficients of drag are associated with highly streamlined objects that meet less resistive forces traveling through fluid and, in this case, travel farther per unit of thrust generated (Hebdon et al. 2022; Ritterbush and Hebdon 2022). Volume accretion per unit surface area is one way to assess a fundamental return on investment: Biomineralization requires energy expense, and the animal must be housed. Increased volume allows for more soft tissue which in turn, for these simplified models, directly governs the animals’ ability to generate thrust and acceleration via the water ejected and the musculature with which to force it through a narrow siphon (Hebdon et al. 2022). In this capacity, it also functions as a close to direct trade-off with our coasting efficiency metric (though this metric is size independent). This parameter is more closely tied to the updated ideas and hypotheses presented in Hebdon et al. (2022) that streamlining is unrelated to acceleration and swimming speed, which are often conflated in traditional discussions of ammonoids which attribute an animals general “swimming” performance exclusively to coefficient of drag and, by extension, streamlining. Lastly, a somewhat arbitrary alternate hypothesis is included in the form of diameter accrued per unit of surface area. This would represent emphasis on the selection for a larger apparent size, particularly in a profile view, a trait that is generally believed to help deter predation. The idea that a non-swimming function may better explain the dominance of serpenticones in the Jurassic has been proposed on more than one occasion, typically tied to the idea that their swimming capabili-
ties were inferior. These explanations rarely come with no specific hypotheses or predictions of what that alternate function might be (Ritterbush and Bottjer 2012; Smith et al. 2014, Ritterbush 2015). We include diameter accrual here as an example non-swimming factor, which is readily retrievable for the same set of model shells. Other performance data, such as hydrostatics or FEA results, could just as readily be integrated, but at this time there is no such dataset that could match up with the hydrodynamics data being included.

**Estimating performance surfaces in Westermann morphospace**

We adapted the maximum-likelihood optimization developed by Polly et al. (2016) to determine the balance between the three performances that best explains the observed morphology of one or more shell shapes in Westermann space. The original method has since been adapted for specific applications in several different studies (Stayton 2018, 2019a, 2019b; Stayton et al. 2018; Dickson and Pierce 2019; Polly 2020; Dickson et al. 2021), and this is another variant that is appropriate for the reduced dimensionality of Westermann morphospace.

First, we transformed our morphospace into a simple x–y coordinate system. Westermann space is constructed from the proportions of three coiling parameters: the umbilical ratio \( U \), the whorl expansion rate \( W \), and the thickness ratio \( Th \) (Ritterbush and Bottjer 2012). Because \( U \), \( W \), and \( Th \) are expressed as proportions, the space only has two mathematical degrees of freedom and is therefore normally represented with a ternary plot. To project our data into the appropriate coordinate system, we divided the raw parameters for each training shell and each real fossil by the sum of their three parameters to turn them into proportions, which lie on a bounded triangular subspace of the original \( U \), \( W \), \( Th \) coordinate space (Fig. 1A). We then centered and rotated the 3D data points to a plane using singular value decomposition, discarded the third dimension, and rotated the configuration into the standard Westermann orientation (Fig. 1B).

Next, we estimated performance surfaces for coasting, as estimated from the training data. To do this, we fit 2D second-order quadratic equations to values for coasting, volume accumulation, and diameter accumulation. The performance surface is a model for how each of these three performances is expected to vary across the morphospace. Each surface was constructed by fitting 2D second-order quadratic equation \( z_i = x^2 + xy + y^2 + x + y + c \) to the performance values of the 19 training shells, where \( x \) and \( y \) are coordinates in Westermann space and \( z_i \) is the performance value for function \( i \). (Fig. 2). We chose a second-order quadratic model over others (e.g., spline fitting, kernel density fitting) because it is closely related to the bi-variate normal adaptive peaks that are widely used as models for selection in evolutionary biology (e.g., Lande 1976; Phillips and Arnold 1989; Arnold 2003; Svensson and Calsbeek 2012) and because we expect these performance metrics to be approximately monotonic in Westermann space without multiple performance peaks in different locations.

Because the three performances are measured in different units, we equated them by scaling each performance variable so that minimum value, representing the poorest performance, was 0.0 and the highest performance value was 1.0 before quadratic fitting, and further standardizing by translating the fitted surface so that its maximum value within the Westermann space is zero (cf., Polly et al. 2016; Polly 2020). This ensures that the maximum and minimum values for each of the three performances are mathematically comparable.

**Maximum-likelihood estimation of weighting coefficients**

We then used the maximum-likelihood method of Polly et al. (2016) to estimate the combined contribution of the three performances to the morphology to a given shell or group of shells. This approach assumes that functional performance and reproductive fitness are linked such that natural selection will favor shells with better performance if everything else is equal (e.g., Arnold 2003). Evolution often involves trade-offs between optimizing morphology for one performance over another, such as speed versus maneuverability. Such a trade-off implies two different performance peaks in morphospace, one for speed and another for maneuverability. Realized shell morphologies could lie anywhere between those peaks depending on the balance between the two performances for long-term fitness (Polly 2020). The realized fitness for any point in morphospace is therefore the weighted sum of the performance values from two or more factors at a given point in morphospace, where the weighting coefficients represent the proportional contribution of each of the three factors (Polly et al. 2016):

\[
F (x, y) \propto \omega_{{\text{COAST}}} \times P_{{\text{COAST}}}(x, y) + \omega_{{\Delta D/\text{ASA}}} \times P_{{\Delta D/\text{ASA}}}(x, y) + \omega_{{\Delta V/\text{ASA}}} \times P_{{\Delta V/\text{ASA}}}(x, y),
\]

where \( F \) is fitness, \( \omega_i \) is the weight given to performance surface \( i \), and \( P_i \) is the value of performance surface \( i \) at location \( x, y \) in morphospace. If we make the assumption that fitness is indeed proportional to the weighted
Fig. 2 Performance surfaces. (A) Location of the training points for which performances were estimated with CFD. Quadratic performance surfaces for coasting (B; $R^2 = 0.79$), diameter accumulation (C; $R^2 = 0.82$), and volume accumulation (D; $R^2 = 0.89$). Training points and performance surface are both colored by their performance values (white = maximum, black = minimum). Vector field shows the direction of morphological evolution of selection if that factor alone was optimized. X indicates the peak of the performance surface.

Contributions of each of these performance factors, then the likelihood $L$ of a given shape $x, y$ is maximized with the combination of weight coefficients that yield the highest possible fitness value at point $x, y$ in morphospace:

$$L (x, y | \omega_{\text{COAST}}, \omega_{\Delta D/\Delta SA}, \omega_{\Delta V/\Delta SA}) \propto \omega_{\text{COAST}} \times P_{\text{COAST}} (x, y) + \omega_{\Delta D/\Delta SA} \times P_{\Delta D/\Delta SA} (x, y) + \omega_{\Delta V/\Delta SA} \times P_{\Delta V/\Delta SA} (x, y). \quad (2)$$

The combination of performance factors that best explain a given shell $x, y$ are the weights $\omega_i$ that maximize equation (2). Note that the maximum height of surface $F$ is higher at the three performance peaks than it is at intermediate $x, y$ points so the peak must continually be rescaled to 0.0 at each iteration of the maximization routine (Polly et al. 2016; Polly 2020). Note also that the weight coefficients must sum to 1.0, which means that only two weights need to be estimated because $\omega_{\Delta V/\Delta SA} = 1 - (\omega_{\Delta D/\Delta SA} + \omega_{\Delta V/\Delta SA})$. The combi-
nation of performance factors that best explain a group of shells can be found by maximizing their summed likelihoods (Polly et al. 2016).

Given the standardization described above, the highest value that the likelihood can ever assume is 0.0. If even the maximized weights produce a negative likelihood value for a given shell or collection, then its evolution cannot be explained solely by the three performance factors under consideration here. The more negative that value, the less the observations are explained by the chosen performance factors. The likelihood analysis can thus be used to transpose not only find what combination of weights between the factors best explains the evolution of an ammonoid shape, but also can indicate whether unidentified factors would have been involved.

Results

The performance surface for each metric used in this analysis is presented independently along with their adjusted $R^2$ values in Fig. 2B–D. Coasting performance and volume accumulation are close to direct trade-offs with each other, the lowest point of coasting being associated with slightly lower values of $W$ than the peak of volume accumulation (Fig. 2B and D). Diameter accumulation shows a similar general shape to that of coasting, but has a much lower gradient and is very slightly rotated, such that morphologies with high $W$ are more strongly emphasized than they are for coasting performance (Fig. 2C).

Triassic and Jurassic ammonite overlapped in Westermann space, but disparity was higher in the Triassic with more forms that approached the oxycone—spherocone axis than in the Jurassic (Fig. 3).

Surprisingly perhaps, the best weightings for our three performance surfaces are identical for the Triassic and Jurassic assemblages (Fig. 3A and C). In both cases, the most likely weights for the three performance factors are $\omega_{\text{COAST}} = 0.5$, $\omega_{\Delta D/\Delta S} = 0.0$, and $\omega_{\Delta V/\Delta S} = 0.5$. Note that there is some ambiguity to the relative weights of $\omega_{\text{COAST}}$ and $\omega_{\Delta D/\Delta S}$. These factors have very similar performance surfaces with their peaks near the oxycone corner and their lowest point along the serpenticone—spherocone axis (Fig. 2B and C). Any combination of weights that sum to 0.5 are almost equally likely, as shown by the dotted line marking a ridge across the likelihood surfaces connecting the point $\omega_{\text{COAST}} = 0.5$ & $\omega_{\Delta D/\Delta S} = 0.0$ with the point $\omega_{\text{COAST}} = 0.0$ & $\omega_{\Delta D/\Delta S} = 0.5$.

Visual inspection of the best-fit peak indicates that it does not explain either the Triassic or the Jurassic distribution of forms very well and the mean likelihood of both assemblages is low: $L_{\text{Triassic}} = -0.108$ and $L_{\text{Jurassic}} = -0.053$, indicating that average morphology has a low likelihood of being observed if the three functional performances were jointly responsible for their evolution.

Interestingly, the lowest likelihood areas of the Jurassic joint performance surface are densely populated with shells, whereas its peak is not (Fig. 3C). This result would seem to suggest that these three performance factors are insufficient to explain the evolution of the observed diversity of Jurassic ammonite form. But perhaps there was not a single combination of weights that applied to all Jurassic species, perhaps some lived in environments that favored oxyconic shapes and others lived where serpenticonic shape was optimal. To test this, we searched for the weights that would maximize the performance of the serpentonic Pseudotentomoceras doetzkirchneri, which occupies the region of low likelihood on the combined performance surface that is optimized to explain the entire Jurassic assemblage (Fig. 4). However, no combination of weighting of these three performances are sufficient to explain a serpenticone form. Even individually optimized, its likelihood remained low ($L = -0.18$) under its own best-fit weights ($\omega_{\text{COAST}} = 0.24$, $\omega_{\Delta D/\Delta S} = 0.33$, and $\omega_{\Delta V/\Delta S} = 0.43$). This set of weights is the one that maximizes serpenticone shape given these three functions, but even this joint performance surface has a peak near the oxycone corner of Westermann space, similar in its contours to the surface that best explains the Jurassic assemblage as a whole (Fig. 3A). Note that this combination of weights worsens the likelihood of the fauna as a whole ($L = -0.122$). While this is not entirely unexpected considering that the Jurassic assemblage is not made exclusively of strongly serpenticone animals, it does, however, highlight that the occurrence of serpenticones is almost entirely divorced from their swimming performance. It also contrasts the zeitgeist of the Jurassic being shaped by these serpenticones. Instead, the performance surface of the Jurassic is seemingly tailored more, which is actually what should be expected considering that very soon after the earliest parts of the Jurassic ammonites being to repopulate back into the previously vacated regions of the morphospace rather than continuing to solely emphasize the serpenticone shell plan.

Discussion

These results paint an interesting picture of the morphological shift in ammonoids at the End Triassic, one that runs counter to many of the longstanding narratives surrounding this transition: Even though there is an overall shift of forms from the oxycone—spherocone axis toward the serpenticone corner, the trade-offs between coasting, volume accumulation, and diameter accumulation that best explain them are identical in the
Fig. 3 Results for Triassic and Jurassic ammonoids (A and C). Likelihood surfaces from the parameter estimations are also shown (B and D). Colors and symbols in A and C are the same as in Fig. 2. Colors on the likelihood surface are white for the highest likelihoods and orange for the lowest. The sum and mean likelihood values for all ammonites in each set are provided alongside their likelihood surface. The more negative a likelihood value is, the less effective the performance variables considered are at explaining the assemblages morphospace occupation with 0.00 representing something perfectly explained.

two eras. This suggests that whatever caused this shift, it was not a change in the average relative importance of these three factors. Furthermore, these three factors are insufficient to explain the evolution of serpenticone forms in either era. Why?

The performance factors we chose for this analysis are, intentionally, heavily biased towards hydrodynamic functional relationships. We chose these factors because so many previous studies have considered swimming effectiveness, or lack thereof, to be a likely expla-
Mismatches in functional narratives of morphology

Fig. 4 Results of optimizing the performance surface weights to fit the serpenticone *P. doetzkirchneri*. Colors and symbols as in Fig. 2. The likelihood value of *P. doetzkirchneri* for this surface, $L$, is $-0.18$, reflecting that performance parameters do a relatively poor job explaining the occurrence of a highly serpenticonic ammonoid like this.

Fig. 4 Results of optimizing the performance surface weights to fit the serpenticone *P. doetzkirchneri*. Colors and symbols as in Fig. 2. The likelihood value of *P. doetzkirchneri* for this surface, $L$, is $-0.18$, reflecting that performance parameters do a relatively poor job explaining the occurrence of a highly serpenticonic ammonoid like this.

nation for the evolution of ammonoid shells, including
the major shift in forms between the Triassic and Jurassic (Ritterbush and Bottjer 2012; Smith et al. 2014; Longridge and Smith 2015; Whalen et al. 2020). Even hypotheses that proposed that ammonoids of the Jurassic are likely explained by preference for some non-swimming function have posited so on the grounds that swimming and hydrodynamics were changing across this regime (Ritterbush and Bottjer 2012; Smith et al. 2014).

Hydrodynamic components of Triassic and Jurassic shell shape fitness appear to have been constant

Our analysis recovers something quite different. What we see is a hydrodynamic stability from Triassic through Jurassic in that the combination of hydrodynamic performances that best explain the observed shell shapes does not change, despite a contraction in the occupation of Westermann space. In each time period, the joint performance surface is evenly weighted between coasting efficiency and volume accumulation, which represent the two hydrodynamically relevant performance measurements of this analysis. This alone contradicts the idea that the change across the End Triassic was likely driven by some change in the prioritization of hydrodynamics.

Releasing this constraint slightly by expanding the model to include shells with lower values of whorl expansion does change the joint performance surface in the Jurassic slightly. The shift toward more oxyconic forms results from the incorporation of diameter accumulation. Coasting performance is interestingly de-emphasized slightly here despite coasting efficiency being moderate to high in the region of the Jurassic morphotypes. This may, however, be due to the fact that overall, the performance surfaces for diameter accumulation and coasting are very similar and the weighting of these two surfaces may be somewhat arbitrary. Regardless, the resulting peak of this surface, and the others for that matter, are not centered close to the actual data to which they are supposedly maximizing. This trend implies that, on their own, the three performance surfaces used in this analysis do not provide a strong explanation for the functions being balanced and emphasized by ammonoids, particularly strongly serpenticonic ammonoids like those of the Jurassic.

Serpenticones cannot be explained by coasting, $\Delta V/\Delta SA$, and $\Delta D/\Delta SA$

In fact, coasting, volume accumulation, and diameter accumulation are incapable of explaining the evolution of serpenticone shells because under any combination these factors favor non-serpenticone shapes. The inability of these factors to predict serpenticonic ammonoids is demonstrated by the results of estimating the best weights for just the strongly serpenticonic *Pseudaeotomoceras*. The peak is similar to that of the unconstrained Jurassic joint surface, although serpenticone form is marginally more favored when diameter accumulation receives slightly stronger weighting. Particularly, this relationship could be indicating a slightly decreased aptitude for distance swimming (which is counter to the conclusions of Hebdon et al. 2022). However, the peak of this joint surface remains quite far from the region itself, suggesting that the functional hypotheses represented here are insufficient to characterize serpenticones. Some unknown additional factor is necessary to create a performance peak in the serpenticone regions of Westermann space.

Overall, our results suggest that significantly more attention in ammonite research should be paid to alternative functional hypotheses and perhaps even non-functional hypotheses. While hydrodynamic performance does appear to be important, it serves as the foundation upon which other functional considerations trade on rather than the functional trade-off that alone dictates every morphological transition of ammonoids.
**Broader application**

The implementation of maximum likelihood that we have implemented affords a way to build upon the foundations of decades of collected data on how taxa perform in their environment. Building these form–function relationships, however, forms another major hurdle along the path to synthesis. For some taxa, there may be substantial morphological data but quite limited functional data or vice versa and thus the distinct form–function links required for the maximum-likelihood method are not readily available.

In the case where morphological data is limited, relative to performance data techniques such as, geometrics–morphometrics can be used to build out morphological data sets. The morphological variables need not even be that intricate. Uncovering the correlations in first order of shape such as limb lengths or thicknesses, and ratios in proportions of limbs are incredibly powerful as initial explorations. These types of measurement also lend themselves well to the construction of theoretical morphotypes like those used in this study.

Theoretical morphotypes can also help to rectify the difficulties that might stem from lacking performance data to characterize a space. With only realized specimens, you may face a variety of limitations such as patchy coverage of the morphospace or being limited in experiments that can be performed because some samples are more rare or difficult to work with than others. Theoretical morphotypes can exist both virtually and physically thanks to 3D modeling and printing allowing measures of performance of those shapes to be acquired in a variety of possible applications, both experimental and simulated. Their nature as purely theoretical also affords the ability to populate an entire morphospace readily and with exceptional control over the nature of that population. From there, the performance surfaces are generated on the theoretical data, meaning that only shape data is required for the realized specimens opening the door to more possible. Their flexible nature of theoretical morphology assemblages also makes them ideal for exploring how one might test and measure performance. In many cases, as we explored here, linking performance to a simple measurable quantity may not be straightforward and iterative experimentation may be needed to develop that process. Theoretical morphologies help to mitigate the difficulties of specimen availability and potentially even cost in the case where simulations can be leveraged allowing such problems to be explored.

**Conclusion**

Broadly, the study we present here serves as an example framework for bringing investigations of functional morphology and biomechanics to the next level. Functional morphology and biomechanics researchers are at a turning point. We now have tools that allow us to intricately explore any form–function dynamic that catches our interest with incredible depth. Each time, we crystallize what we know about some core aspect of a taxon’s interaction and we add a new puzzle piece to our collection. The next critical step is to use these individual pieces to build an understanding of our target taxa and its interactions with the world and how those change through time. An individual one of these pieces is insufficient to explain the entire ecology of any taxa particularly throughout the entirety of its evolutionary history. But the process of evaluating the relationships of each piece to another one by one is difficult and often imprecise. What we have lacked are robust tools and approaches for synthesizing this knowledge, putting the pieces of the puzzle together and critically seeing how much of the puzzle is missing to get a sense of what and where we need to explore further.

The maximum-likelihood tool and analysis framework we present here offers a flexible means to begin making progress in this direction. It is highly scalable, allowing for any number of metrics of performance to be implemented in a single analysis provided that model data exist to fit a surface for each metric across the region of the morphospace you aim to visualize (Stayton et al. 2018; Dickson et al. 2021). Furthermore, as this paper shows, it can be adapted for morphospaces that do not have traditional orthogonal dimensions such as Westermann space. It can be used to understand the trade-offs between functional performance factors at one time, between times, or in different locations (or environments). This allows it to be used to test competing hypotheses about the functional drivers of evolutionary, extinction, and environmental sorting.

Perhaps the most valuable aspect of this approach is that it challenges whether the current state of knowledge sufficiently explains the selection of particular taxa/morphotypes during particular environmental/ecological transitions. Here, we have presented a prime example of when the explanations dominating a space appeared to have characterized the problem in a vacuum, but brought into context through this analysis they tell a different story. Instead, what we find is a gap in our broader knowledge of the form–function relationships in planispiral ammonoids, which even extends beyond the Triassic–Jurassic transition. As in this analysis, using these tools to find and call attention to gaps in our performance data can help us to redirect our efforts towards more fully characterizing our taxa of interest and developing a more complete picture of the dynamic interactions with their corresponding ecosystems and environments through time.
Acknowledgments

We would like to thank our session organizers L. Ferry and T. Higham for organizing an engaging set of sessions. We would also like to thank our reviewers and editor for their patience and time in bringing this manuscript to completion. We also thank Y. Choi and M. Heberer for many long discussions about ammonites, why we even bother doing research, and pursuing the things we really find interesting.

Funding

This work was supported by the National Science Foundation [grant number 1945597].

References

Abdelhady AA, Seuss B, Hassan HF. 2019. Stratigraphic ranking [grantnumber1945597]. This work was supported by the National Science Foundation. Funding the things we really find interesting. monstrouss, why we even bother doing research, and pursuings the things we really find interesting.


