University of Reading

Faculty of Sciences

School of Biological Sciences

METHODS TO ANALYSE NUTRITION IN HIGH DIMENSIONS

Juliano Morimoto

Thesis submitted in fulfilment of the requirements for

the degree of Doctor of Philosophy

Abstract

Nutrition is fundamental to all living systems. In animals, morphological, physiological, and behavioral adaptations for acquiring and digesting food shape their responses to the environment. While experimental frameworks have advanced our understanding of animal nutrition, methodological tools to address its complexity remain limited. This thesis introduces pioneering analytical methods to explore nutritional trade-offs and optimal diet balances using Geometric Framework for Nutrition (GF) performance landscapes. Chapter 1 traces the historical development of GF, highlighting its independent origins in agricultural sciences and evolutionary ecology. Chapter 2 presents the Vector of Positions approach—the only method capable of analysing n-dimensional performance landscapes—validated using two landmark datasets. Chapter 3, the first study in the Nutrigonometry series, introduces a computationally efficient algorithm to estimate nutritional trade-offs in three-dimensional performance landscapes—the most common in the field. Using the Pythagorean theorem, I demonstrate its robustness across multiple datasets, showing that ordinary linear regression outperforms machine learning models for these calculations. Chapter 4 provides the first systematic test of the GF experimental design, proposing improved methods for high-dimensional nutritional experiments to enhance accuracy in estimating optimal diets and performance landscapes. Chapter 5 applies differential geometry to measure curvature in performance landscapes and introduces the Hausdorff distance for comparing performance landscapes. Chapter 6 integrates previous analytical methods to examine the evolution of optimal diets between sexes in insects. My findings reveal that the protein-to-carbohydrate ratio is more similar among closely related species and that sexual conflict over nutrition is evolutionarily conserved. This study represents the first empirical work in precision comparative nutrition. Chapter 7 applies Thales' theorem of inscribed triangles to analyse how animals cope with imbalanced diets. By triangulating food intake under suboptimal conditions, I quantify deviations from Thales' theorem, revealing underlying nutritional strategies. This thesis advances our understanding of nutritional constraints in optimizing fitness. Through a range of methodological advancements, my work provides new insights into how animal nutrition can be studied in its full multidimensional complexity. These methods and applications have broad implications, and I hope they will be extended to other areas of biology to further our understanding of how living organisms interact with their diet and how nutritional responses evolve across the tree of life.

Contents

List of Figures

List of Tables

Acknowledgments

I have divided this section into two parts. The first, written in English, acknowledges those who have had a significant influence on my career. The second, written in my native language, Brazilian Portuguese, is dedicated to those who have supported me throughout my life.

I would like to thank my supervisors for embarking on this endeavour with me—Prof. Mark Pagel and Prof. Tom Oliver, who kindly agreed to act as my supervisors, and Prof. Richard Sibly, who has always been incredibly kind and supportive of my work and played a pivotal role in enabling this doctorate to begin and be completed. I could never have imagined that I would have the opportunity to interact with, let alone be supervised by, such influential figures in my field. I am deeply honoured to have had this experience, which I will cherish for a lifetime.

I would also like to express my gratitude to Dr. Stuart Wigby and Prof. Tommaso Pizzari, the supervisors of my first doctorate. I continue to learn so much from you, both personally and professionally, and I am forever grateful for the truly transformational opportunity you gave me in 2013. This thesis – and all the work I do – is profoundly influenced by the culture, environment, and inspiration you and so many others have created.

I extend my sincere thanks to Prof. Stephen J. Simpson and Prof. David Raubenheimer for their continued support and mentorship. You have been inspiring figures, not only because of your outstanding contributions to the field but also because of the intellectual exchanges we have had over the years. The time I spent in Australia has shaped my thinking and will continue to influence my work. Words cannot fully express how kind you have been to me and my career. This thesis, like much of my research, builds upon the Geometric Framework for Nutrition that you pioneered, and serves as a tribute to your remarkable legacy.

I am also deeply grateful to Prof. Phil Taylor, Dr. Toni Chapman, and Dr. Fleur Ponton for granting me the opportunity of a lifetime at Macquarie University. I sincerely regard this period of my career as one of the most intellectually stimulating, offering me unparalleled opportunities to grow and flourish.

I would like to express my gratitude to the many mentors, past and present, who have helped shape me into the scientist I am today and the one I aspire to become. This includes Prof. Robert May, Prof. Richard Shine, Prof. John Speakman, Rg. Prof. Mirela Delibegovic, Prof. Michelle Pinard, Prof. Pete

Smith, Prof. Luis Fernando Favaro, Prof. Elaine Benelli, Prof. Marcio Pie, Prof. Mauricio Moura (Free), Dr. Kara Layton, Dr. Marius Wenzel, Dr. Davina Derous, Dr. Sharon Mitchell, Dr. Rebecca von Hellfeld and Dr. Irem Sepil. More generally, I would like to thank many of my colleagues with whom I have had the privilege to collaborate over the years. There are too many outstanding individuals to name, and I fear that listing them all would inevitably lead to unintended omissions.

My appreciation also extends to Prof. John Hunt, Dr. Alistair Senior, Dr. Mathieu Lihoreau, Dr Christen Mirth, Prof Carla Sgrò, Dr Matt Piper and Dr Carlos Vilhafan who continue to refine and whenever necessary, challenge my work. A special thanks to Prof Kwang Lee and his group – you have always been so kind to me and I can safely state that my career would not have been as prolific without your support (and data!). The field – and science in general – needs more of the kindness you provide.

Last but not least, a special thanks to my students, who make this job rewarding even when everything else seems uncertain, and to the children whose curiosity and enthusiasm give my work greater meaning through the outreach and public engagement projects we develop.

***

Não gosto de agradecimentos tradicionais, apesar de já ter escrito um no meu primeiro doutorado. Uma lista seca, sem vida e sem alma, de pessoas que um dia ninguém saberá quem são. Então gosto de escrever e contar histórias. Mas para evitar que o agradecimento seja mais longo que a tese, me contive e escrevi um misto de histórias e listas.

Agradeço aos meus antepassados que viajaram de tão longe ao Brasil. De uma família de imigrantes, muitos que não sabiam falar português ao desembarcarem no Brasil, como meu avô Minoru Morimoto, formamos uma família de orgulho, com diversas profissões e, contando com os "agregados", tendo um total de seis PhDs (incluindo esse, que espero ter sucesso). Agradeço aos meus pais pela melhor vida que me deram e que me possibilitaram chegar aonde cheguei. Hoje sou pai também, e tenho muito orgulho e alegria de saber que conheceram meu filho, mesmo que até o momento de maneira virtual. Agradeço a todos da minha família de nascença e da família de escolha, principalmente à minha melhor amiga Gabriela Kinaske, pelo apoio e confiança na minha capacidade e nas minhas escolhas. É muito difícil viver em exílio de vocês. Agradeço a todos que, sabendo ou não, me ajudaram a ser quem sou e me proporcionaram uma vida com valores cada vez mais raros nessa nossa sociedade.

Tenho uma gratidão sem igual à minha esposa – Dra. Zuzanna Pietras – pelo amor, carinho e apoio durante muitos anos. Você é uma super-heroína e sem você, nada do que recebo seria possível. Obrigado por dividir a vida comigo e mais do que tudo, obrigado por me dar o maior presente da minha vida, nosso

filho. Falando nele, o resto desse agradecimento é inteiramente dedicado ao meu filho que me transformou desde sua chegada.

Desde muito jovem, eu tenho a mania, quase um vício, de conversar com meu eu do passado, presente e futuro. Também gosto de escrever, apesar de não necessariamente ser bom nisso. Mas na minha adolescência, decidi escrever cartas para mim mesmo para que fossem lidas quando mais velho.

Recentemente, li uma dessas cartas que escrevi quando tinha 25 anos. Lembro do entorno da carta, que escrevi em uma loja de açaí na tigela em uma venda que parecia um boteco, mas sem as bebidas e os bêbados que me acostumei a ver durante a infância. O dia estava fresco, com uma brisa, mas perfeito para um dia relaxante na praia. Como de costume, havia tirado folga do trabalho para ter um dia meu, de reflexão e planejamento. Tenho vivido na mente quando me sentei, tirei a caneta preta da mochila, e comecei a rabiscar no papel. Entretanto, depois de quase 10 anos, não me lembrava de nenhuma palavra que havia escrito. Não me recordava mais de quem eu era, por dentro ou por fora, apesar de parecer o mesmo, mesmo não sendo. De certa maneira, a carta era nova para mim, se não pela escrita, pelo conteúdo.

A carta era longa, com seis páginas. Mas um trecho da carta me comoveu mais que qualquer outro, dessa carta ou de outras que já escrevi. Nele, me perguntei se já havia tido filhos, e questionei que se não os tivera, era porque havia falhado de maneira irreparável a vida. De repente, o sentimento que tinha quando escrevi o trecho veio à tona. Na época, eu não acreditava que um dia teria a bênção de ser pai. Senti um nó na garganta, um frio na barriga como se estivesse prestes a flutuar. Lágrimas vieram aos meus olhos sem serem convidadas, e saíram como quem foge do perigo. Relembrava os momentos que um ano antes, haviam me feito pai. Lembrava de cada momento do nascimento do meu filho, que agora tinha um ano e meio, e trazia tanta alegria, amor e felicidade para a nossa casa. Talvez tivesse sido tomado por saudade, pois meu filho viajava quando li a carta. Mas parte da minha emoção era resultado da lembrança do seu nascimento, que não foi fácil e, como pais de primeira viagem, nos trouxe momentos de muita tensão, ansiedade e medo. Portanto, gostaria de escrever esse agradecimento para meu filho, em uma carta pessoal, para que ele um dia, quando eu não estiver mais aqui, tenha uma carta para carregar com ele.

Meu filho,

Você é o maior milagre da minha vida. Não importa quantos títulos, prêmios ou reconhecimentos eu receba, nada se compara ao dia em que te dei colo pela primeira vez. Eu fui seu primeiro colo na vida e no mundo. Eu te segurei enquanto dormias, exausto, pelo trabalho de vir a esse mundo, entre os minutos que pareciam uma eternidade quando ficamos sozinhos, sem nenhuma informação, na sala enquanto sua mãe foi encaminhada para a sala de cirurgia. Eu te vi respirar pela primeira vez. Eu vi você bocejar e espirrar pela primeira vez. Te segurei enquanto tomava suas vacinas e quando quiseram tirar seu sangue para exames, chorando mais que você pela dor que você não se lembra, mas que eu sim. Te vi cair pela primeira vez e te ouvi falar pela primeira vez. Cortei seu cabelo, te dei banho. Te levei para nadar e te ensinei a dar cambalhotas. Chorei quando você estava doente e chorei quando você sorriu pela primeira vez.

Quando somos novos, pensamos que ter filho muda nossa vida para sempre pelas responsabilidades que temos. Não podemos viajar, sair, dormir tarde. Não podemos isso ou aquilo. Tudo isso é verdade, mas não é a maior transformação que tive quando você nasceu.

Você chegou para purificar a minha vida. Por toda a bagagem que carreguei, pelos meus erros, erros dos meus pais, e dos pais dos meus pais, e de outras tantas gerações que me levaram a ser quem eu era. Como no encontro entre o rio e o mar, eu era um turbilhão de sentimentos crus, fortes, porém opacos. Quando chegaste, mudou tudo em um segundo. Decantou tudo que criava confusão e me trouxe clareza, como águas cristalinas, que me batizou e perdoou tudo o que havia sido.

Quando li minha carta para mim mesmo, percebi que não somente tinha tido um filho. Mas que meu filho, sem saber, tinha me curado da dor e da angústia que carregava. Como pode ser tão pequeno, mas tão poderoso?

Poderia escrever um livro de tudo que me deste desde que chegou. Talvez o escreva deveras. Mas nunca conseguirei transmitir o bem que você me fez ao me deixar ser seu pai.

Muito obrigado por dar sentido à minha vida. Eu sempre vou te amar, meu filho.

Declaration and author contribution

I confirm that this is my own work and has not been submitted for any other degree or qualification at any institution. The papers presented here, with the exception of the general introduction (Chapter 1), were peer-reviewed and published as a requirement for the title of PhD by Published Works at the University of Reading. Below, I detail my contributions to each of the papers that reflect the chapters of this thesis.

Chapter 2

Morimoto, J. and Lihoreau, M., (2019). Quantifying nutritional trade-offs across multidimensional performance landscapes. *The American Naturalist*, 193(6), pp.E168-E181.

Contributions: I am the first and corresponding author of this paper. I was responsible for the development of the proposed analytical framework, the computations, data visualisation and the writing of the manuscript.

Chapter 3

Morimoto, J., Conceição, P., Mirth, C. and Lihoreau, M., (2023). Nutrigonometry I: using right-angle triangles to quantify nutritional trade-offs in performance landscapes. *The American Naturalist*, 201(5), pp.725-740.

Contributions: I am the first and corresponding author of this paper. I was responsible for the development of the proposed analytical framework, the computations, data visualisation and the writing of the manuscript.

Chapter 4

Morimoto, J., (2022). Nutrigonometry II: Experimental strategies to maximize nutritional information in multidimensional performance landscapes. *Ecology and Evolution*, 12(8), p.e9174.

Contributions: I am the sole author of this paper and therefore, responsible for all of the work

which is presented.

Chapter 5

Morimoto, J., Conceição, P. and Smoczyk, K., (2022). Nutrigonometry III: Curvature, area and differences between performance landscapes. *Royal Society Open Science*, 9(11), p.221326.

Contributions: I am the first and corresponding author of this paper. I was responsible for the development of the proposed analytical framework, the computations, data visualisation and the writing of the manuscript.

Chapter 6

Morimoto, J., (2024). Optimum ratio of dietary protein and carbohydrate that maximises lifespan is shared among related insect species. *Aging Cell*, 23(3), p.e14067.

Contributions: I am the sole author of this paper and therefore, responsible for all of the work which is presented.

Chapter 7

Morimoto, J., (2023). Nutrigonometry IV: Thales' theorem to measure the rules of dietary compromise in animals. *Scientific Reports*, 13(1), p.7466.

Contributions: I am the sole author of this paper and therefore, responsible for all of the work which is presented.

Chapter 1

General introduction

Malnutrition is one of the most significant global challenges of our societies. Malnutrition is an umbrella term which includes the opposing consequences of both overnutrition (e.g., overweight and obesity) and undernutrition (e.g., stunting). These contrasting nutritional burdens to individuals and populations worldwide have been alluded to as the 'double burden of malnutrition' (DBM) and is pervasive (Nugent et al. 2020). More than 2.9 billion people corresponding to ca. 37% of the global population suffer from malnutrition (World Health Organization 2024). Children are also at risk, with over 186 million of children under the age of five experiencing malnutrition (World Health Organization 2024). Therefore, there is an urgent need to better understand how nutrition affects health, with insights that can help us tackle the DBM and the global malnutrition crisis.

Fundamental research using animal models is a clear asset in this context because it helps advance both theoretical and empirical research into the DBM. All living organisms need nutrients to grow, survive and reproduce (Simpson & Raubenheimer 2012, White & Brown 2010, Gude & Taga 2020). Animals, including humans, must balance their nutrient intake to survive and reproduce (Raubenheimer & Simpson 2020), which poses an opportunity to use animal models to better understand the constraints and consequences of their nutritional choices to health, and whenever appropriate, draw parallels to humans (Boone et al. 1990, Le Couteur et al. 2024, Green 2024). This approach not only uncovers how animals interact with each other and their environment, but also provides fertile ground for new discoveries about the effects of nutrients on physiological processes across the tree of life, which in turn have the potential for translational and clinical breakthroughs (Raubenheimer & Simpson 2020).

1.1 Nutrition is complex and multidimensional

Nutrition is complex and multifaceted, characteristics which have inspired and challenged theoretical and experimental frameworks (Jacobs Jr et al. 2009, Simpson & Raubenheimer 2012, Simpson et al. 2017). Evolutionary ecology is particularly interesting in this regard because theoretical developments to study nutrition follow a pattern of increasing complexity, from uni-dimensional approaches such as optimal foraging theory (Krebs et al. 1977, Charnov 1976, Stephens & Krebs 1986) to more recent multidimensional frameworks such as nutritional stoichiometry (Anderson et al. 2004) and the Geometric Framework (GF) (Simpson & Raubenheimer 1993, Raubenheimer & Simpson 1993). They cover all aspects of animal nutrition, from motivation (e.g. Sibly & McFarland 1974), behaviour (e.g. Krebs et al. 1977) to physiology and long-term (inclusive) fitness (e.g. Stearns 1998, Simpson & Raubenheimer 2012,

Anderson et al. 2004).

Among these frameworks, GF is arguably the most powerful because nutrients do not need to be transformed or converted into other currencies of analysis as in uni-dimensional frameworks that often use the intake of elementary atoms as units of investigation (Simpson & Raubenheimer 2012). The GF by contrast is *nutrient explicit*, an important requirement to maintain interpretability of models dealing with nutrients and complex diet mixtures. Another advantage of the GF relative to other frameworks relates to experiments. GF enables experiments which simultaneously manipulates unlimited numbers of interacting nutrients, thereby addressing nutrient-nutrient interactions (Jacobs Jr et al. 2009). Thus, GF is perhaps the most important and powerful framework to enable large-scale experiments in *precision nutrition*, a field which aims to understand and predict how individuals will respond to personalised diets and design these personalised diets to ensure individuals' health (Rodgers & Collins 2020). If successful, personalised nutrition promises to address the malnutrition crisis, mitigate the DBM, and guarantee long-term health for individuals and populations worldwide, although such promises may be more challenging to achieve than previously thought (Denny & Collins 2021, Bray & Ryan 2021). Nevertheless, GF remains an important tool to advance precision nutrition research and application (Simpson et al. 2017).

This thesis presents a collection of analytical methods tailored to GF. These methods were developed to address the lack of formal models which can facilitate insights from GF data, especially within the context of personalised nutrition. The analytical methods presented here apply mathematical concepts and theorems to multidimensional nutrition data, extracting insights into how animals eat, the consequences of their feeding to life histories, and the trade-offs that animals might encounter when feeding on diets with varying nutrient compositions. These are not the only methods that have been developed, and a historical overview of the main methodological advances is provided before contextualising this thesis' chapters within the broader field.

1.2 A historical overview of the applications of geometry to nutrition

1.2.1 Conceptualising the geometry of nutrition

The earliest published record of a study formally linking geometric concepts and nutrition dates back to 1973 when Parks (1973) published a paper titled *"Diet space and response surfaces"*. In this paper,

Parks conceptualised key concepts that are still used today. For instance, Parks proposed the use of an equilateral triangle coordinate system as the nutrient space wherein diet mixtures could be represented as ratios of macronutrients (Fig 1.1a). Although this representation of diet mixtures had been used in previous studies (Evans & Miller 1968), Parks discovered that such representation allows for the identification of "lines of constant energy" that represent diet mixtures with the same total energy but with varying nutrient compositions (Fig 1.1b). Perhaps more striking was Parks' formalisation of the response surface modelling approaches, which he defined as *"lines of length R, representing the response of an animal to each experimental diet, plotted orthogonal to the plane of the diet space"* (Fig 1.1c). Growth rate was the primary response mapped onto Parks' diet space but as we will see later in this chapter, other physiological, behavioural, and life history traits have been used since.

Response surface models applied to nutrition had been used in the late 1960s (Chandler et al. 1967, 1968, Gardiner et al. 1967) but Parks' formalism provided a solid foundation for the applications of this approach more widely. We shall return to response surface models below. Parks' work provided a detailed account of much of the principles of how geometry and nutrition can be integrated (see also Table 1.1).

One aspect that Parks' original 1973 paper did not cover was the explicit use of vectors to model diet mixtures. This was proposed independently by Moon & Spencer (1974) in their own version of the geometry of nutrition. They used vectors to define nutrient mixtures and applied their then newly developed concept of "holor" to model diets and nutrient blends in Euclidean space. Specifically, in their approach, diets and nutrients can be perceived as *vectors* with "magnitude and character" and plotted within right-angle triangles or other triangular shapes in Euclidean space (Moon & Spencer 1974) (Fig 1.2). Moon & Spencer (1974) were the first to propose the use of right-angle triangles as nutrient spaces to locate diet mixtures. This innovative conceptualisation of nutrition and nutrient composition of diet mixture was the first to explicitly represent diets as vectors. However, one limitation of this approach was that Moon & Spencer (1974) conceptualised nutrition using *affine geometry*, lacking a metric and vector multiplication operations. This meant that distances between diets were not defined and the approach was primarily descriptive. Both Parks (1973) and Moon & Spencer (1974) papers have remained largely overlooked from the recent literature on the geometry of nutrition.

In 1982, Parks published a book titled *'A Theory of Feeding and Growth of Animals'*, where he proposed a new way to represent diet mixtures as vectors Parks (1982) (Fig 1.3a-b). Like Moon & Spencer, Parks also explained the nutrient space as a triangular region and formalises in more detail how vectors can be used to explain, model, and interpret dietary mixtures and blends. For example, Parks

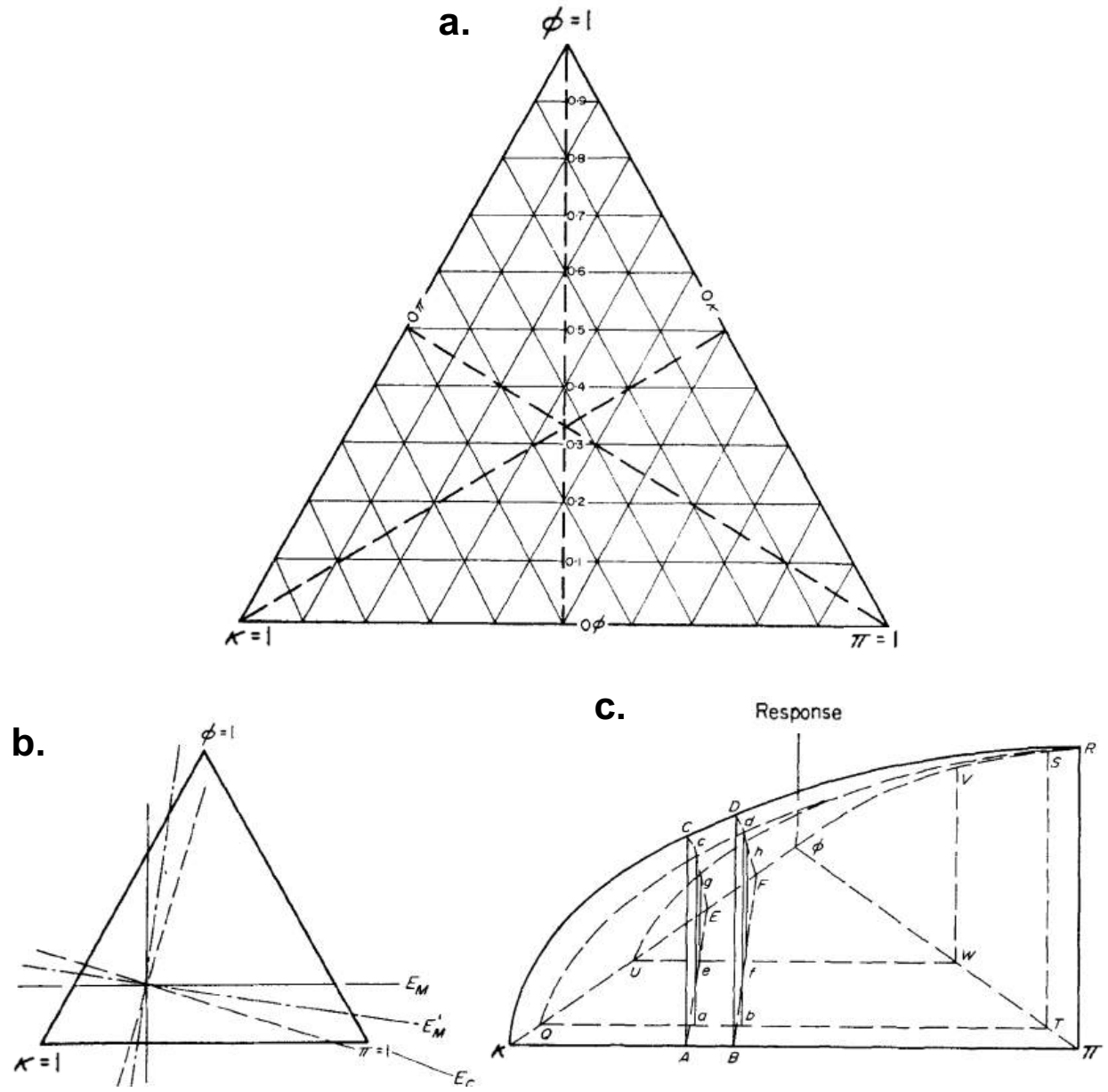


Figure 1.1: **The geometry of nutrition by Parks (1973)** (a) The equilateral triangle as coordinate system of the nutrient space where diet mixtures could be represented as ratios of nutrients. The axis represented by ϕ corresponds to fat, π corresponds to protein and κ corresponds to carbohydrates. (b) Parks (1973) also proposed the identification of lines of equal energy, including E_M and E_C as metabolisable and combustion energies, respectively. (c) A representation of a response surface over the diet space. Adapted and reproduced with permission from the publisher.

(1982) showed how diet mixture vectors could uncover the "movement" of diets in nutrient space following nutrient supplementation, as is the case of wheat that becomes progressively more similar to the diet mixture vector of beef muscle with the supplementation of lysine up to 400 mg (Fig 1.3c). Importantly, Park's advanced Moon & Spencer's original approach by incorporating in addition to vector addition and scalar multiplication, vector inner and outer products and a *metric* (i.e., a function that measure distances between two points in space) (Parks 1982). This meant that Parks' approach enabled the calculation of distances between diets, enabling studies comparing diets and their physiological effects.

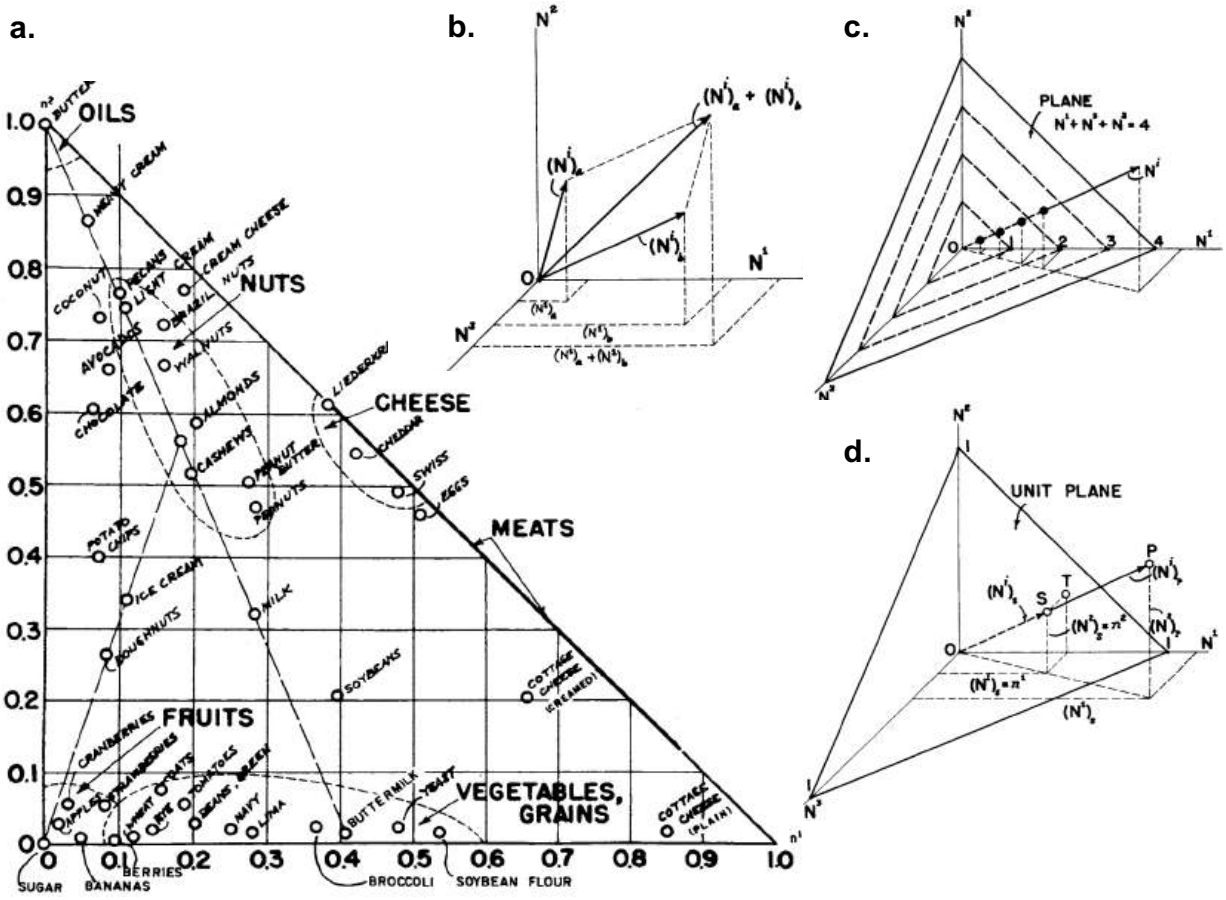


Figure 1.2: **The Geometry of Nutrition by Moon & Spencer (1974)** (a) The nutrient triangle proposed by Moon & Spencer (1974) where diet mixtures are represented in this two dimensional space. The horizontal axis represents the proportion of protein and the vertical axis, the proportion of lipids. (b) Three dimensional nutrient space with diets represented as points in such space. Each axis represent an arbitrary nutrient N^i. (c) Moon & Spencer (1974) subdivided the nutrient plane into what they referred to as "planes of constant magnitude" where foods along a vector could be differentiated. Each axis represent an arbitrary nutrient N^i. (d) Moon & Spencer (1974) also correlate the planes of constant magnitude with the nutrient triangle. Each axis represent an arbitrary nutrient N^i.

The combination of early work applying response surface models to nutrition and Parks' formalism galvanised a wave of studies using response surface models in nutritional sciences. Roush et al. (1979) used response surface modelling to estimate growth across diets in the Japanese quail. In this paper, the authors put forward the foundations for how response surface models should be visualised, thereby simplifying the original approach by Parks (1973) and setting the standards that are essentially used in the field until today. Meanwhile, in a series of publications starting in 1982, Toyomizu et al. (1982) used response surface modelling to study the body composition of chicks, late applying this methodology to other animals (Toyomizu et al. 1988, 1991, 1993). These applications of response surface modelling during the 1980s and early 1990s garnered attention to this methodological approach which became an invaluable tool in nutrition (Table 1.1).

By early 1990s, the use of response surface models and geometry to represent dietary mixtures were

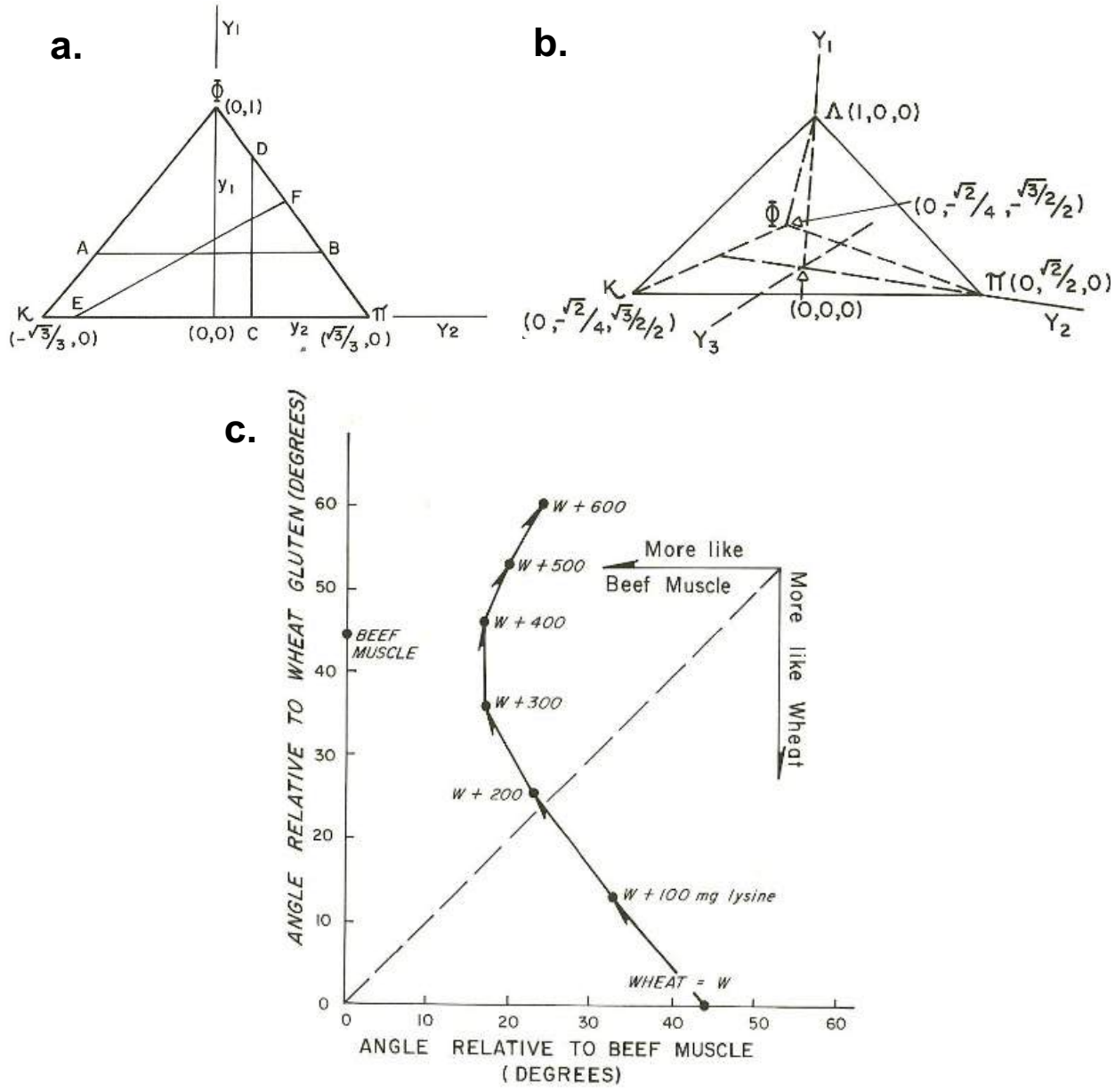


Figure 1.3: **The geometry of nutrition vector spaces by Parks (1982).** The three (a) and four (b) dimensional nutrient space. The axis represented by ϕ corresponds to fat, π corresponds to protein and κ corresponds to carbohydrates. In (b), Λ represents an arbitrary nutrient. In (a) the nutrient space is represented by an equilateral triangle of unit height whereas in (b), the nutrient space is represented as a tetrahedron of unit height with each of the faces of this tetrahedron representing a pure diet mixture of a single nutrient. (c) It is possible to estimate the angle θ between mixture vectors as proposed by Parks (1982). In this example, the composition of wheat (W) following the addition of the amino acid lysine in steps of 100 mg is shown. With the addition of lysine, the wheat mixture vector shifts in the two dimensional nutrient space and become closest to the vector for beef muscle at the mixture of W + 300 mg and W + 400 mg of lysine before diverging again for higher lysine supplementations. Reproduced with permission from the publisher.

discussed more widely. For example, Emmans (1991) published a short article in the *Proceedings of the Nutrition Society* following a meeting of the Nutrition Society where the author alluded to the idea of modelling animal dietary selection using the equilateral triangle of macronutrients (i.e. fat, protein, carbohydrate), similar to the equilateral triangle first conceptualised by Parks (1973). The article was too

short to present any formalism and omitted the crucial theoretical foundations that had been developed in nutritional sciences by Parks (1982) and Moon & Spencer (1974) as well as the parallel work from the field of ecology and evolution.

1.2.2 State-space models in ecology and evolution

Agricultural and nutritional scientists were not the only ones developing the idea of diets as mixtures that can be represented in multidimensional space (see Table 1.1). Evolutionary ecologists also developed similar frameworks when studying animal motivation. This was spearheaded by two landmark publications by David McFarland and Richard Sibly (McFarland & Sibly 1972, Sibly & McFarland 1974) which approached nutrition from the perspective of animal motivation and drives as opposed to diet mixtures as in Parks (1973) and Moon & Spencer (1974). McFarland and Sibly later expanded their framework in subsequent works (Sibly & McFarland 1974, McFarland & Sibly 1975, Sibly & McFarland 1976, Sibly & Calow 1983) which collectively, defines much of the conceptual and theoretical underpinnings of the Geometric Framework of Nutrition (GF).

It is interesting to note the independent convergence of concepts which were developed simultaneously in ecology and evolution and nutritional and agricultural sciences. For instance,McFarland & Sibly (1972) studied the concept of "drives" related to animal motivation. They formalised the use of vectors to represent non-unitary drives (as opposed to scalar unitary drives) which could be used to model animal motivation and its consequences. As a case study, McFarland & Sibly (1972) linked their multidimensional approach to animal feeding. In modelling feeding "commands", which can be thought of as directions and magnitudes of commands that act on behavioural mechanisms to elicit an observed behaviour, McFarland & Sibly (1972) conceptualised a three dimensional Euclidean space for proteins, carbohydrates, and fat required by a feeding animal (Fig 1.4). This is an analogous conceptualisation of diet as mixtures represented by vectors in multidimensional space by Parks (1973), Moon & Spencer (1974).

McFarland & Sibly (1972) also defined the dimensionality of the nutrient space (which they called the "consequence space") and discussed cases where the realised dimensionality of such space are reduced when animals feed in a single diet. This is conceptually similar to the concept of "lines of equal energy" put forth by Parks (1973). As we will see below, this is also conceptually analogous to the concept of "nutritional rails" used in the GF (Fig 1.4b).

An important insight proposed by McFarland & Sibly (1972) is the idea of "realizable consequences", which represents the subspace of achievable outcomes when animals feed on diets of varying composition.

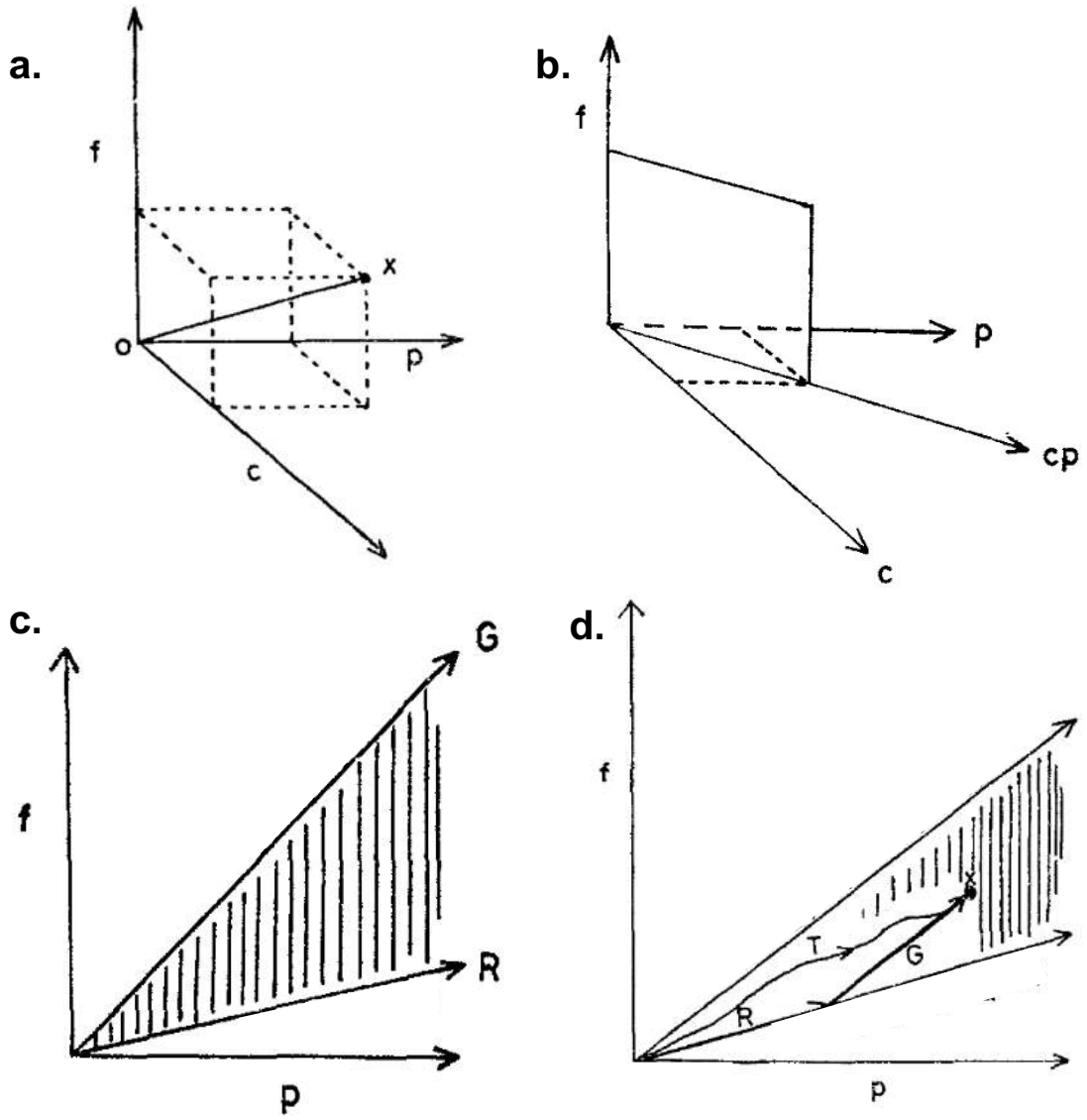


Figure 1.4: **The concept of non-unitary drives by McFarland & Sibly (1972) and its similarities with the Geometric Framework of Nutrition (GF).** (a) McFarland & Sibly (1972) conceptualised a nutrient space when developing a framework to study animal motivation and drives. The authors define the command space as the three dimensional space composed of protein (p), carbohydrates (c) and fat (f) axes. (b) When animals are kept in a diet with fixed nutrient composition, this creates a plane which couples, in this example, protein and carbohydrates (pc), along which animals can move but cannot move anywhere else in space. (c) When given a choice between two diets (in this example, G and R), animals can reach any point within the sub-space defined by the vectors $\vec{G}$ and $\vec{R}$ that represent the mixture of nutrients for each of the diets. Any point in the shaded region can be realised by eating amounts of R and G, as exemplified in panel (d). In panel (d), T represents any transient state before animals reach their desired state x. McFarland & Sibly (1972) referred to the shaded area as realisable consequences, which conceptually underpins the methodological approach of *choice* experiments to identify the intake target (IT) in the Geometric Framework (GF). Adapted and reproduced with permission from the publisher.

Specifically, these realizable consequences are the subspace which lies within two vectors, $\vec{R}$ and $\vec{G}$ that represent diet mixtures and are the boundaries of a realizable subspace outside with animals cannot reach (Fig 1.4c). Any space bounded by the two vectors is achievable if animals eat a mixture of $\vec{R}$ and $\vec{G}$, whereas animals walk along the boundaries of this subspace by feeding only on $\vec{R}$ or $\vec{G}$. (McFarland

& Sibly 1972) (Fig 1.4c-d). This insight was not, to my knowledge, conceptualised in nutritional and agricultural sciences, although Parks (1973) discuss "movements" of diet mixtures when diet supplements are added to mixtures, alluding to the idea that diets could be bounded within a subspace of the nutrient space (see Chapter 8 and 9 in Parks 1982).

The 1970s marked the simultaneous development of multidimensional nutrition with the explicit application of geometric concepts and linear algebra to model diet mixtures (Parks 1973, Moon & Spencer 1974) and animal motivations (McFarland & Sibly 1972, Sibly & McFarland 1974). Along with theoretical developments, empirical work using response surface models, particularly in relation to animal nutrition in agricultural settings, also emerged (Chandler et al. 1967, Gardiner et al. 1967, Chandler et al. 1968, Roush et al. 1979, Toyomizu et al. 1982, 1988, 1993). These developments prepared the foundations for and were paralleled to development of the Geometric Framework of nutrition.

1.2.3 The Geometric Framework of Nutrition (GF)

In 1993, Stephen J. Simpson and David Raubenheimer published two landmark papers formalising what is now known as the Geometric Framework (GF) of nutrition (Simpson & Raubenheimer 1993, Raubenheimer & Simpson 1993). These two papers have had a major impact in evolutionary ecology and, more broadly, biomedical and nutritional sciences. The GF incorporated the mathematical concepts of metric, ratios, and realisable consequences while building upon the authors' foundations on the statistical analysis of nutrient ratios (Raubenheimer & Simpson 1992). Not all previous concepts were incorporated, such as for example the use of vectors to estimate properties of GF data. This is relevant to the contributions presented in this thesis and discussed below. Nonetheless, the GF rapidly grew to become one of the most powerful methods to study animal and subsequently human nutrition. To fully understand the significance of the GF to the field as well as the importance of the analytical methods proposed in this thesis, we must first review the fundamental principles of the GF which makes this framework one of the most widely used to study the complexity of nutrition.

The GF is a state-space model designed to uncover how the intake and balance of nutrients affect the physiology, behaviour and fitness (Simpson & Raubenheimer 1995, 2012). The GF is subdivided into two sets of experiments: (1) animals are force-fed diets with a wide range of nutrient ratios and concentrations (*no-choice* experiments) (Fig 1.6a) and (2) animals ware offered diets with complementary nutrient ratios and are allowed to self-select their preferred optimum nutrient blend (i.e., their *"intake target"*) (*choice* experiments) (reviewed in Simpson & Raubenheimer 2012) (Fig 1.6b). No-choice experiments allows us

to map how different, often extreme dietary mixtures, affect physiology and fitness. For example, one can map how lifespan in the model organism *Drosophila melanogaster* changes with the intake of diets varying in protein and carbohydrate as done in the landmark paper by Lee et al. (2008) (Fig 1.6c). In this case, the GF reveals the dietary limitations which organisms have, by bringing to light diet mixtures that are prohibitively unbalanced for homeostasis and fitness realisation. On the other hand, choice experiments give insights into the nutrient balance that animals strive to achieve and defend when allowed to navigate among diets. This concept maps well onto the concept of "realizable consequences" by McFarland & Sibly (1972) whereby animals navigate a subspace of the nutrient space (bounded by their dietary options) to achieve their intake targets (Fig 1.6d). When integrated, these two sets of GF experiments provide an almost complete picture of how animals eat and the consequences of nutrition to their life-histories, and how animals respond to eating imbalanced diets relative to their self-selected optimum, which is known as "rules of compromise" (details below). Recent studies often employ one or the other set of experiments, partly because of the nature of the information that is given but also due to challenges of ascertaining dietary choices for some model systems (e.g. mice) (Solon-Biet et al. 2014, 2015). Nonetheless, the combination of the two sets of GF experiments continue to uncover important biological insights into how animals and humans respond to diets (Simpson & Raubenheimer 2012, Raubenheimer & Simpson 2020).

The GF led to the development of approaches to analyse diets in high dimensions. For instance, in another landmark paper, Raubenheimer (2011) formalised the use of right-angle mixture triangles (RMT) to analyse dietary mixtures from proportions. The RMT emerge to allow field ecologists to utilise the multidimensionality approach and visualisation power of the GF using dietary data collected in the field (Fig 1.5). At first, RMT resembles the concepts first proposed by Moon & Spencer (1974) and Emmans (1991). However, Raubenheimer (2011) formalised the application of the RMT to analyse the nutritional ecology of organisms from field data, providing a robust framework that has underpinned insights into animal and human nutrition Raubenheimer & Simpson (2020). Moon & Spencer (1974) and Emmans (1991) papers focused primarily on the abstract potential of using right-angle and equilateral triangles to represent dietary mixtures and therefore were more difficult to translate to applications. Moon & Spencer (1974) did propose a similar concept of partitioning nutrient space into sub-spaces that are interlinked with the nutrient triangle as did Raubenheimer (2011), but their development lacked details and formalism. The methodological developments geared towards the RMT lie outside the scope of the contributions presented in this thesis.

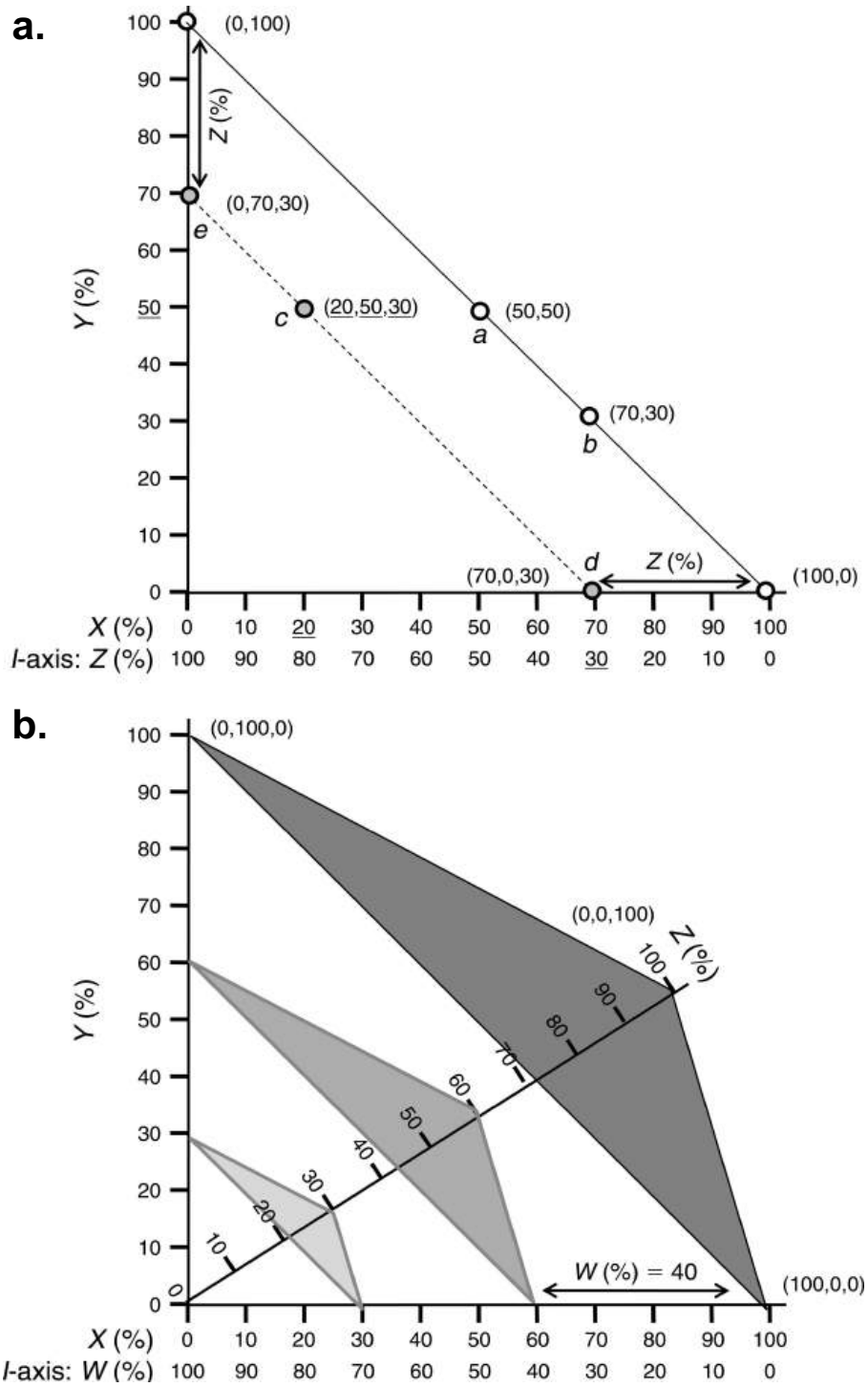


Figure 1.5: **The right-angle triangle mixture (RMT) by Raubenheimer (2011).** (a) A right-angled mixture triangle (RMT) with three components: focal axes X and Y, and an implicit axis Z (denoted generically as axis I). Mixtures containing only X and Y (where $Z = 0\%$) lie along the concentration isoline defined by $Y = 100 - X$, represented by the solid diagonal line in the figure (e.g., points a and b, satisfying $X + Y = 100\%$). When Z is introduced at 30%, the possible mixtures shift to a new isoline, $X + Y = 70\%$, shown as a dashed line (e.g., points e, c, and d). More generally, the $X + Y$ isoline for any given Z value can be plotted as a line connecting the point $(100\% - Z)$ on the X-axis to the same value on the Y-axis. (b) A four-component RMT with focal axes X, Y, and Z, and an implicit axis W. The same logic applies as outlined for the three-component mixture in (a), except that the $X + Y$ isoline is replaced by an equilateral triangle with apices at X, Y, and Z. This approach resembles that of "planes of constant magnitude" from Moon & Spencer (1974) (see e.g., Fig 1.2c above). Adapted from Raubenheimer (2011).

1.2.4 The GF experimental design and data collection

There are several commonalities in the design of *choice* and *no-choice* GF experiments and for brevity, these are explained jointly in this section. Specific aspects of the experimental design of each set of experiments are detailed below. The aim of this section is to review the fundamental details which are relevant to the contributions presented in this thesis. A users' guide is available in the literature (Simpson & Raubenheimer 1995) along with an extensive review of the general GF approach (Simpson & Raubenheimer 2012).

The GF was designed to be flexible and accommodate experiments investigating the effects of an arbitrary (i.e., n) number of nutrients. Because visualisation of high-dimensional data (i.e. four dimensions or higher) are difficult or impossible, this section will focus on the simplified case of $n = 2$ nutrients. Nonetheless, the approach discussed here are readily applicable to studies of more than these two nutrients. Suppose we are interested in studying how the intake of protein (P) and carbohydrates (C) influences the expression of a fitness-related trait (e.g. lifespan) in a given species. The intake of each nutrient can be plotted as orthogonal axis in Euclidean space, forming a region known as the "nutrient space" (Fig 1.6a). Each point p_i in this space is defined by the amounts of P and C intakes, that is, p_i has coordinates (P, C). Drawing a line from the origin through the point p_i we can define a line, known as "nutritional rail", which has slope $\frac{P}{C}$ and quantifies the amount of P and C intakes when animals feed on a diet with $\frac{P}{C}$ ratio of nutrients (henceforth "PC ratio") (Fig 1.6a).

There are infinitely nutritional rails in such nutrient space. Instead, the GF experiments partitions the nutrient space into pre-defined nutritional rails with known PC ratios. While helpful, this approach still leaves the researcher with infinitely many possibilities because there are infinitely many points in each nutritional rail. As a result, the researcher also needs to discretise each nutritional rail by concentrations of P and C to form the diets that are be given to individuals of the target species in the study. At the end of this discretisation process, the nutrient space is subdivided into "anchor points", each representing diets with different PC ratio and combined PC concentrations (Morimoto 2022). These anchor points represent diets with specific PC ratio and nutrient concentrations which are then used in experiments (Fig 1.6a). Originally, the nutrient intake on these diets was carefully measured (Lee et al. 2008, Fanson et al. 2009) (intake experiments) but over time, this practice were replaced and the anchor points became the reference points upon which traits are mapped directly (fixed ratios experiments) in *no–choice* experiments (e.g., Silva-Soares et al. 2017, Kutz et al. 2019). The methods presented in this thesis have been validated for

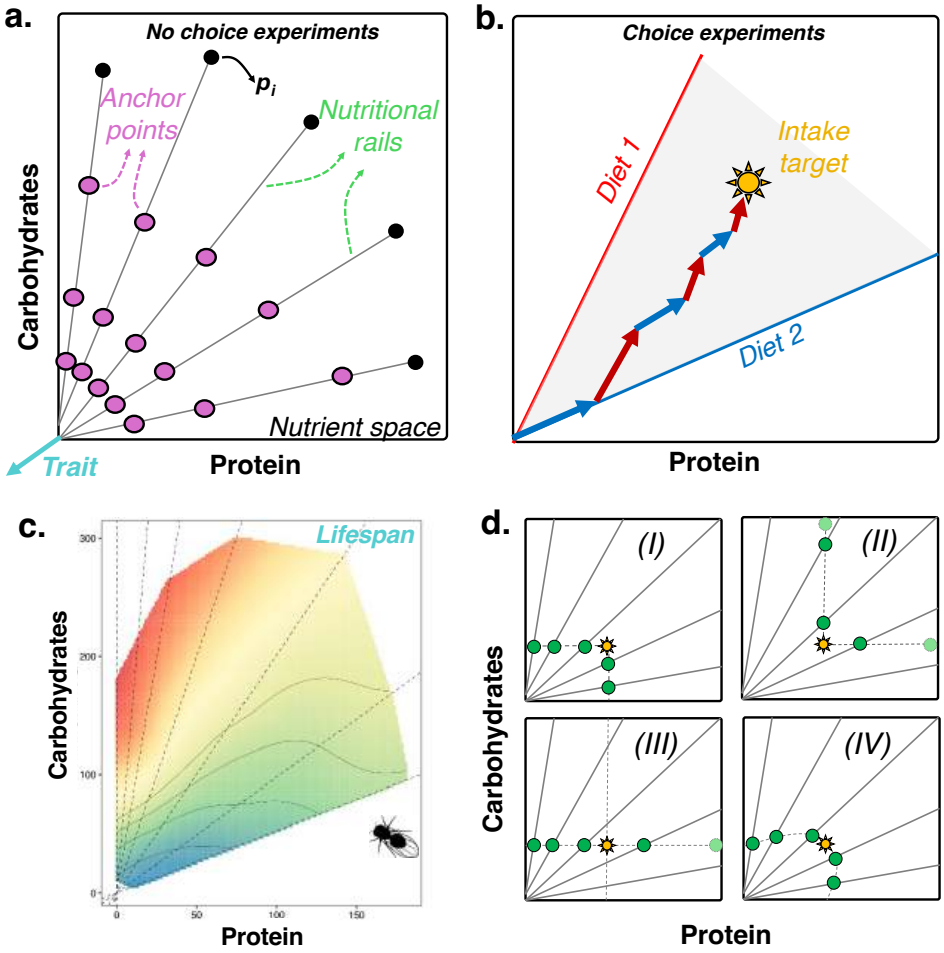


Figure 1.6: **The Geometric Framework of Nutrition (GF) by Simpson & Raubenheimer (1993) and Raubenheimer & Simpson (1993).** (a) The $n = 2$ dimensional nutrient space which is formed when the intake of two nutrients, in this example protein (x-axis) and carbohydrates (y-axis) are plotted in the Euclidean space. Discretising the nutrient space into nutritional rails is needed for experimentation, as well as discretising each nutritional rail into diets with known ratio and concentration of nutrients – the "anchor point". In *no-choice* experiments, a trait of interest is mapped onto the nutrient space, creating a $n + 1$ dimensional performance landscape. (b) Example of a three-dimensional performance landscape mapping lifespan to the intake of protein and carbohydrate in *Drosophila melanogaster*. Adapted from Lee et al. (2008) and reproduced from Morimoto et al. (2023). (c) In *choice* experiments, animals are given a choice between two complementary diets (red and blue) and can navigate a region (shaded) to achieve their self-selected intake targets (IT; yellow star) assuming this IT lies within the region between diets. Note that movements are parallel to the slope of each of the diets, and animals need to navigate accordingly to ensure that IT is reached. Adapted from Simpson & Raubenheimer (2012). (d) Example of four different "rules of compromise" when animals feed in imbalanced diets. Rule I: Eat until the intake of one of the nutrients is reached, then stop eating; Rule II: eat until the intake of both nutrients are reached, irrespective of surpluses in the intake of other nutrients; Rule III: eat until the intake of one nutrient (e.g., protein, horizontal; shown) is met, irrespective of surpluses or deficits in the intake of other nutrients; Rule IV: minimise the distance between the intake of nutrients in the imbalanced diet and the IT (aka "closest distance optimisation"). Adapted from Raubenheimer & Simpson (1993).

both experiment types. It is important to mention, however, that intake experiments are still mandatory for *choice* experiments within the context of the GF, as discussed below.

1.2.5 Choice experiments

Animals are given two complementary choices of diet and are allowed to forage freely. The intake of each of the diets and their nutrients is monitored so that their self-regulated dietary intake, also referred to as "intake target (IT)" can be measured (Fig 1.6b). To confirm that the IT is that self-selected by animals, it is often appropriate to challenge the animal by offering a wide range of complementary diets and measuring whether or not the intake target converges to the same IT (see Simpson & Raubenheimer 2012). Once the IT is defined, we can map the average intake of nutrients in each of the nutritional rails used in the *no–choice* experiment. Together, the average intake across nutritional rails and the IT form a *nutrient array*, which is thought to contain information about the physiological and behavioural rules that govern how animals feed on imbalanced diets relative to their self-regulated IT (Raubenheimer & Simpson 1993, Simpson & Raubenheimer 1993). The shape of the nutrient array uncovers how the animal cope with imbalanced diets, by demonstrating the rules that underpin their food intake when feeding at various distances from their self-selected IT (Fig 1.6c).

1.2.6 No-choice experiments

Animals are fed diets corresponding to "anchor points" (see above). A trait or traits are then measured in for these animals as a proxy of animals performance (fitness) on each of the diets. Common traits used in these experiments include lifespan and egg production, for instance (Simpson & Raubenheimer 2012). The values of the traits are assigned ("mapped") onto each of the anchor points and an interpolation method, most commonly using a thin-plate spline, is conducted to create what is known as *performance landscapes*. These landscapes resemble a traditional fitness landscape (Wright 1932, Gavrilets 2004) although in this case it relates diet intake or composition to a fitness-related performance trait(s) (Fig 1.6c). The interpretation of the performance landscapes are intuitive: peak regions (often represented in red) are combination of nutrients which maximise the trait or traits under consideration while valley regions (often represented in blue or green) are combination of nutrients which minimise the trait or traits. Simpson et al. (2004) was the first to formalise how performance landscapes in GF work are constructed. The authors assumed that the costs of nutrient surplus or deficit increased "continuously and smoothly" as nutrient intake distanced from the intake target (as in Tilman 1982) and built a response surface using a Taylor series expansion (Simpson et al. 2004). Using *Spodoptera littoralis* larvae fed in chemically defined diets, they showed that the performance landscape for fitness (measured as a combination of

larval growth and survival) (Fig 1.7). An important point to highlight in this work is that with their definitions underpinning the construction of the landscape, the authors used the shape of the performance landscape to infer the type of costs associated with nutrient surplus and deficit. In the specific example of *S. littoralis* whose performance landscape approximated a "tilted ellipse", the authors concluded that the costs associated with nutrient imbalances were quadratic (Simpson et al. 2004).

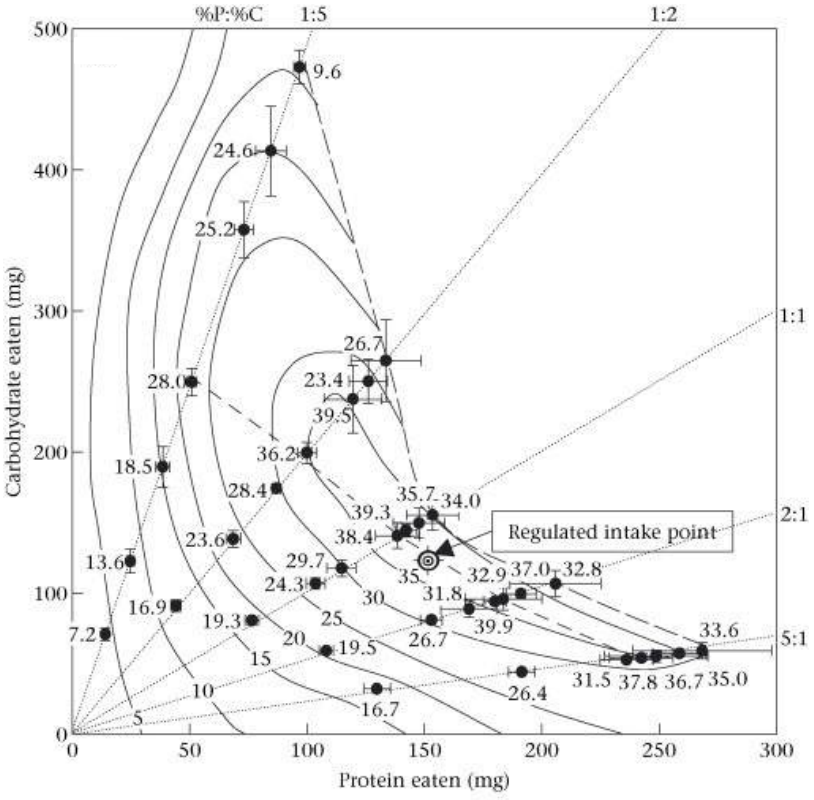


Figure 1.7: **GF performance landscape of an experiment in *Spodoptera littoralis* parameterised as Taylor series approximation by Simpson et al. (2004).** (a) Results from the experiment where *Spodoptera littoralis* larvae were fed chemically defined diets during their final larval stage. Thirty-five treatment groups received a single food type, restricting their nutrient intake to a fixed protein-to-carbohydrate ratio in the diet. In contrast, five treatment groups were allowed to choose between complementary food pairings. Refer to the main text for detailed descriptions of the treatment regimes. Each point on the graph represents the mean nutrient intake (bicoordinate mean ± SE for 10 caterpillars), and the number beside each point indicates performance, calculated as the mean wet mass growth rate (mg grown/day) multiplied by the percentage of individuals surviving to pupation. The point marked as the "regulated intake point" represents the combined mean for 58 caterpillars (with $N = 11$ or 12 per treatment) that had access to one of the five choice treatments in their experiment, that is, it is the estimated intake target (IT). The regulated intake point of self-selecting larvae is located within the summit region of the fitness surface. The short-dashed line represents treatments where the total protein and carbohydrate concentration was 42%, corresponding to near-maximal performance at various protein-to-carbohydrate ratios. The long-dashed line shows the upper limit of nutrient intake, beyond which larvae did not feed. Adapted from Simpson et al. (2004).

It is important to note the increase in dimensionality of the GF data generated by no-choice experiments. For instance, when investigating the effects of two nutrients ($n = 2$), then the performance landscapes will be of $n + 1$ dimensions. This means that visualization of the performance landscapes becomes easily intractable for a relatively low number (i.e. $n = 3$) of nutrients. Therefore, models which

can accommodate this high-dimensionality are paramount to advance insights and push the field into more robust and replicable approaches.

No-choice experiments may or may not be accompanied by the choice experiment, although the joint information they provide can help interpret the topology of the performance landscapes. A classic example is the work by Lee et al. (2008) which showed that self-selected IT maximise lifetime egg production (a proxy of fitness) as opposed to other life-histories like lifespan and reproductive rate. More recently, GF work has focused on interpreting the properties of the performance landscapes in their entirety, and for this purpose the knowledge of the IT is desirable. This thesis proposes new methods that enable researchers to find and compare peaks and valleys in GF performance landscapes within and between species, as well as analyse the geometry of nutrient arrays to better understand the rules of animals compromise.

1.3 Method development tailored to GF data

The GF is powerful but methodological developments to interpret the full range of data obtained from GF choice and no-choice experiments has lagged behind the incredible growth in applications of the GF in empirical studies. The specific details and history of these methodological developments are given in detail at the introduction of each of the chapters to provide the appropriate background against which the contributions presented have emerged. Here, it suffices to summarise the main limitations of the field leading to the contributions presented in the upcoming chapters.

There have been only few analytical developments to study the rules of compromise from GF choice experiments. Raubenheimer & Simpson (1992) first proposed a descriptive model of the rules of compromise for which they later formalised through the concept of "summary plots" (Raubenheimer & Simpson 1997). Cheng et al. (2008) then built on these models and developed an analytical approach to calculate what they referred as the "regulatory scaling factor" which was a scalar capable of estimating how organisms dealt with nutrient excesses and deficits. A more extensive analytical framework was then developed by Simpson et al. (2004) which aimed to integrate the full nutrient array and performance landscapes data. This approach was successful in that it expanded the concept of multidimensional nutrition to that of the ecological niche (e.g., Tilman 1982, Hutchinson 1957, Chase & Leibold 2009), which later underpinned the formal conceptualisation of what is known as the multidimensional dietary niche (Machovsky-Capuska et al. 2016).

Developments to measure properties in performance landscapes have also been limited. In fact, two research teams have independently focused on analytical method development for GF performance landscapes. First, John Hunt's team were the first to propose analytical methods to systematise the analysis of performance landscapes (Bunning et al. 2015, 2016, Rapkin et al. 2018, del Castillo et al. 2016). The second team belongs to the author, with contributions which compose this thesis and are presented below.

1.4 Thesis outline

This thesis presents a collection of approaches to analyse GF data. **Chapter 2** addresses a limitation in the framework proposed by Rapkin et al. (2018). In doing so, this framework becomes the first to accurately estimate differences in peaks and valleys in performance landscapes unbounded by the dimensionality of the data. While powerful, the analytical method from Chapter 2 was conceptually difficult and computationally expensive. As most GF experiments focus on the simple cases where two nutrients are investigated and three-dimensional performance landscapes are analysed, **Chapter 3** proposes a simpler conceptual approach, which is also computationally effective, to estimate differences in peaks and valleys in performance landscapes – the "Nutrigonometry" models. The approach in this chapter uses the Pythagoras theorem and trigonometric functions to achieve its purpose, and was later expanded to incorporate other theorems and applications. **Chapter 4** uses the analytical framework from Chapter 3 to propose new ways in which GF experiments could be designed to maximise the information obtained from performance landscapes. In particular, it proposes a new way in which anchor points should be designed to cover larger regions of the nutrient space. Failing to do so can significantly curtail our ability to make inferences about peak regions in performance landscapes, particularly for traits with peaks that depend on the interaction of nutrients. **Chapter 5** expands the analytical framework proposed in Chapter 3 to estimate local curvature of performance landscapes. This provides for the first time a way in which additional properties of performance landscapes are estimated, helping not only to describe and compare performance landscapes but also to potentially build predictive models. **Chapter 6** developed for the first time a semi-automated algorithm to reconstruct performance landscapes from unavailable published data, and compared the optimal diet that maximise lifespan across insect species. This was the first time that multidimensional nutritional data is mapped onto evolutionary relationship between species, opening up future directions in comparative precision nutrition. Lastly, **Chapter 7** then uses

the trigonometric approach to analyse performance landscapes and adapt it to nutrient arrays. Using the Thales' theorem of inscribed triangles, the chapter proposes a conceptually simple analytical approach to test for the rules of dietary compromise.

1.5 Critical evaluation of contributions to the field

1.5.1 Chapter 2

As described above, there have been multiple parallel developments that incorporated geometric concepts into nutrition. However, from 1967 until 2018 (over half a century), the use of geometry was primarily focused on aiding the description of biological processes modulated by nutrition, such as, for example, the expression of life-history traits. In 2018, Rapkin et al. (2018) developed the first targeted approach to quantify and compare how different diet mixtures modulated the expression of seemingly competing traits. However, there was an inconsistency in their methodology, which resulted in inaccurate estimates of the extent to which two (or more) competing traits could not be maximised with a single diet—that is, the strength of nutritional trade-offs.

Chapter 2 provides the accurate methodology needed to circumvent this limitation, as well as to formalize an analytical framework for measuring nutritional trade-offs within and between species in a n-dimensional nutritional space. The Vector of Positions approach introduced in Chapter 2 remains, until the submission of this thesis, the only method capable of estimating nutritional trade-offs in high-dimensional nutritional data from GF experiments.

However, the Vector of Positions approach has limitations. One such limitation is the computational cost associated with the machine learning models, particularly the support vector machine (SVM), used to delineate peak (or valley) regions. Another is that, due to noise and stochasticity, the algorithm relies on an arbitrary threshold to reduce noise and maximize signal from the SVM, thus improving the delineation of peak or valley regions in GF landscapes. While these limitations do not invalidate the methodology, they introduce a certain degree of subjectivity to the analysis of the data, which I then address in Chapter 3.

Despite these constraints, the Vector of Positions method still represents a significant advancement in the ability to quantify nutritional trade-offs in complex, high-dimensional dietary environments.

1.5.2 Chapter 3

To address the computational and conceptual limitations of Chapter 2 – particularly for 3D datasets, which are the most common GF datasets in the literature – I developed the Nutrigonometry models. Chapter 3 presents the first and most extensive of these, where I tested, for the first time, the performance of multiple statistical models across different dataset types in identifying peak and valley regions in GF landscapes. To accelerate computations, I implemented an extension of the Pythagorean theorem – a well-known mathematical principle. Through a series of validations, my work demonstrated that off-the-shelf machine learning models were outperformed by traditional models such as (Bayesian) linear regressions and generalised additive models (GAMs), although the latter still imposed substantial, and in some cases greater, computational costs than any other models.

In addition, Chapter 3 presents an extensive set of validations – including the pioneering use of topological data analysis – to demonstrate that the Nutrigonometry I model can accurately describe the properties of GF landscapes (such as peaks and valleys). A key improvement in this algorithm is the use of a systematic method for delineating peak or valley regions in the performance landscape, relying on the distribution of the data and quantiles, rather than the arbitrary threshold used in Chapter 2.

However, limitations remain with this algorithm, some of which have yet to be fully identified or resolved. For example, the algorithm is based on the expansion developed by Simpson et al. (2004) to parametrise the performance landscape. This parametrisation can "smooth out" multiple peaks, making the Nutrigonometry I model less effective at detecting multiple disjoint peaks within a given landscape. While disjoint peaks remain rare in the GF literature – observed only sporadically, such as in certain gene expression landscapes (e.g., Post & Tatar 2016) – the growing application of GF to diverse datasets means that the Nutrigonometry I model should ideally be capable of identifying such features when they are present.

Despite these limitations, the Nutrigonometry I model has been well received in the field and is now being implemented in a range of studies investigating other aspects of animal nutritional ecology, such as for example the effect of local adaptation and temperature on nutritional responses (e.g., Zanco et al. 2025, Morimoto 2025*a*, Rho & Lee 2025).

1.5.3 Chapter 4

Chapter 4 began with a simple question: while the Vector of Positions approach and Nutrigonometry I are accurate in identifying peaks and valleys in performance landscapes, are these GF performance landscapes being built using the best possible experimental designs to maximize information?

Chapter 4 computationally tested different experimental designs for GF experiments and assessed how well peaks and valleys in performance landscapes are identified and compared. By testing common experimental designs used in ecology and evolutionary biology, Chapter 4 showed that standard GF designs lead to accurate detection of nutrient ratios but truncated identification of nutritional peaks. As a result, we lack a reliable way to determine the full extent of nutrient combinations that maximize a given performance trait. Chapter 4 then argues that alternative designs, such as random sampling, can explore a broader portion of the nutritional space and, in turn, uncover novel and unprecedented ways in which animals respond to nutrition.

A key limitation of this work is the lack of a sensitivity analysis to determine the minimum number of sampling points (i.e., "anchor points") needed to empirically maximize the information content of a true GF landscape. Such insight could be transformative, not only for refining Nutrigonometry methods but also for advancing GF as an empirical framework. Power analyses and feasibility assessments of experimental designs could help optimise experiments in both lab and field settings, widening access to GF as a go-to method of investigation in nutritional ecology.

To date, Nutrigonometry II remains the only study to challenge the standard GF experimental design and demonstrate that alternatives can unlock new avenues of discovery.

1.5.4 Chapter 5

Since 1967, when geometry and response surface models were first incorporated into the study of nutrition, it has been assumed that animal nutrition operates within Euclidean space. While this assumption is simple and useful, it overlooks a broader range of characteristics that can be derived from performance landscapes to better understand how animals respond to their nutrition.

In this context, Chapter 5 is the first to incorporate concepts from differential geometry into the study of animal nutrition. Its main innovation lies in the calculation of local (intrinsic and extrinsic) curvature of performance landscapes. This allows animal responses to nutrition to be represented not only in terms of the absolute magnitude of performance trait expression under a given diet, but also in terms of how

rapidly that expression changes with varying diets – and in which direction those dietary changes occur.

With the growing interest in applying machine learning to biology, calculating curvature becomes a valuable feature. This is because this approach introduces new variables that can improve the predictions of machine learning models and the accuracy of classification, an approach known as *feature engineering*. Although this was not implemented in the current chapter, it remains a promising direction for future development of the Nutrigonometry models.

Another important contribution of Chapter 5 is the use of the Hausdorff distance as a metric to quantify differences between a performance landscape and a null hypothesis (i.e., a flat landscape over the same domain). The Hausdorff distance is a flexible way to estimate the distance between two sets, and this chapter formalized its application to GF datasets for the first time.

The main limitation of this work is the dimensionality of the data. I used the fundamental forms to calculate Gaussian and mean curvatures, which is only applicable to three-dimensional datasets. For higher-dimensional landscapes, alternative methods are needed. However, as most GF landscapes to date are three-dimensional, the current model can be applied to virtually all existing data (note that 4D landscapes can be represented as series of 3D ones). Future work will aim to generalize this approach to calculate curvatures of performance landscapes in n-dimensional datasets.

1.5.5 Chapter 6

Having developed the computational methods to estimate peaks, valleys, and curvatures, I moved on to integrating GF data across multiple species to gain broader insights into nutritional responses across the tree of life. Do closely related species share similar optimum nutrient ratios that maximize traits?

To answer this question in Chapter 6, I used the methodologies developed in Chapters 2 to 5 to perform the first comparative analysis of GF performance landscapes, in what is now known as the field of *comparative precision nutrition*. However, a major caveat that initially prevented such data integration was the limited accessibility of raw GF data. I circumvented this issue by creating a semi-automated algorithm that allowed me to reconstruct performance landscapes with similar features (e.g., peaks, valleys, etc.) to the published ones –without requiring access to the original data. This represents a major advance, as no prior algorithms (or even manual methods) had been able to extract meaningful data from 3D performance landscape figures.

The results were insightful: related insect species share optimum protein-to-carbohydrate (P:C) ratios that maximize lifespan. Moreover, I found that males and females have different P:C optima for maximiz-

ing lifespan, and that this difference is conserved across species. This suggests that sexual conflict over nutrition may be evolutionarily conserved. This is particularly intriguing given that males and females share a genome, implying that their responses to potential nutritional conflict may also be conserved. Uncovering the molecular mechanisms underlying this phenomenon could help us understand how evolution resolves sexual conflict through shared physiological and molecular adaptations.

The main limitation of this study is its small sample size. GF research has focused primarily on insects, particularly model organisms. As a result, we currently lack the taxonomic breadth necessary to confirm whether the patterns observed in Chapter 6 represent a broader phenomenon across the tree of life. Addressing this will require several key initiatives. First, GF studies must be conducted in a wider range of species across diverse taxonomic groups – including other insects, invertebrates, and vertebrates beyond mice and humans. While some work has been done in cats and dogs, it remains too fragmented to support robust comparative analyses. Second, we need better phylogenetic trees. The tree I used was reconstructed from barcode genes, and while it reflects expected species relationships, it is neither ultrametric nor dated. Accurate modelling of the evolution of nutritional responses across the tree of life requires high-quality phylogenies – built from taxonomic, genomic, and paleontological data.

Despite these limitations, Chapter 6 shows the powerful insights that can be gained by integrating GF with comparative methods and remains a foundational contribution to the field of comparative precision nutrition.

1.5.6 Chapter 7

GF has two experimental branches: one where animals are given a choice of diets (choice experiments), and one where they are not (no-choice experiments). All analytical methods developed in Chapters 2 to 6 apply to no-choice experiments. In contrast, Chapter 7 introduced a novel analytical method to quantify how animals respond to dietary imbalances – specifically, the rules of dietary compromise – in GF choice experiments. While the Geometric Framework (GF) has long enabled researchers to visualise and interpret animal feeding behaviour, methods for statistically analysing rules of compromise have lagged behind. Existing approaches often relied on pairwise distance calculations and/or complex summary plots, which at times obscured rather than clarified biological interpretation.

To address this, Chapter 7 applied a classic geometric principle – Thales' theorem of the inscribed triangle – to quantify deviations from the commonly observed "closest distance optimisation" (CDO) pattern in nutrient arrays. This provided a rigorous yet intuitive framework for testing whether nutrient

intake on imbalanced diets adheres to expected behavioural rules. This remains the first formal model for testing CDO and making prediciton *a priori*, offering a clear ground truth against which researchers can determine whether animals do, or do not, minimise the distance to their intake target – and if not, which nutrients drive this deviation and why.

This method was validated using three landmark GF datasets of increasing complexity: from *Drosophila melanogaster* (a model generalist), to two locust species with different dietary breadths, and finally to generalist and specialist moths (*Spodoptera littoralis* and *S. exempta*). The results revealed striking patterns: dietary specialists tended to adhere more strictly to CDO, whereas generalists showed greater flexibility in nutrient intake under imbalance. Interestingly, *D. melanogaster* exhibited strong underconsumption of high-carbohydrate diets – a response more typical of specialists – suggesting possible physiological or evolutionary constraints even in generalist species.

The main limitation of this study is that Thales' theorem applies only to two-dimensional nutrient arrays, restricting direct application to higher-dimensional landscapes. As with curvature analysis, entirely new frameworks would be required to generalise the approach to higher dimensions. However, to date, all GF choice experiments have been conducted in two-dimensional nutrient spaces, making the method widely applicable to existing datasets. An adaptation of the Vector of Positions method may be possible, though it has not yet been formalised (work in preparation).

A further limitation – highlighted in a comment by Senior et al. (2025) –is the lack of a formal maximum likelihood estimator for the angle β within this framework. This gap remains the main bottleneck for broader statistical inference on rules of compromise in animal nutrition. Despite these limitations, Chapter 7 represents a significant analytical advance and opens the door for comparative and evolutionary studies of dietary compromise across species and ecological contexts.

1.6 Concluding remarks

Nutrition is complex and multifaceted. With the growing societal impact of the double burden of malnutrition (DBM), we need ways to overcome this complexity and high-dimensionality of nutritional data to advance our prospects of healthier and more sustainable lives (Nugent et al. 2020). Analytical frameworks are paramount to make sense of complex data to inform experiments, and uncover insights. There has been a long history of applications of geometric concepts to represent diet mixtures and their impacts on animals and humans, but this has been disjoint and lacking in analytical capacity (Table 1.1). The work

presented in this thesis is situated in this context and represents a collection of analytical methods aimed at spearheading the field into new research directions (Fig 1.8).

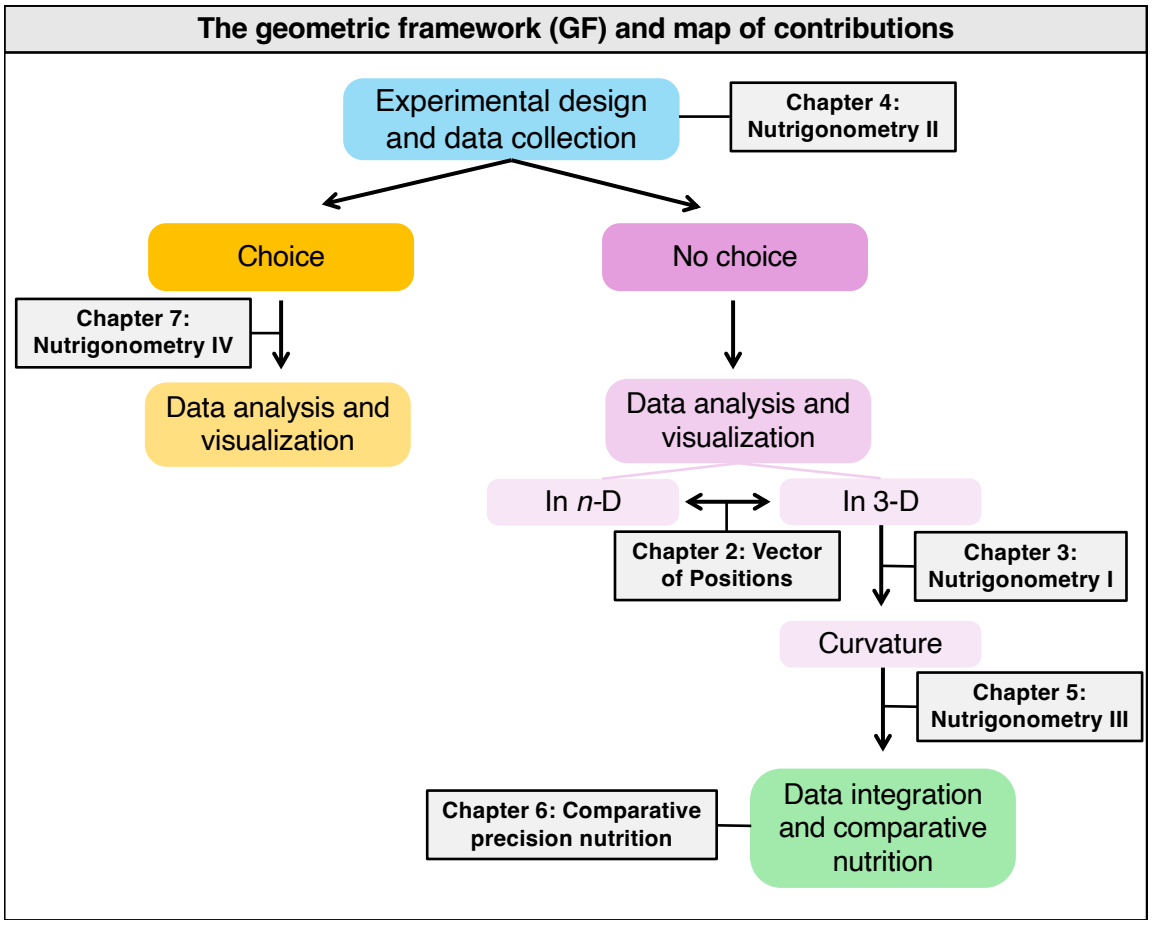


Figure 1.8: **Map of the contributions of each chapter of this thesis for the advancement of the geometry of nutrition.** From experimental design to data integration, the contributions of this thesis represent significant analytical and conceptual breakthroughs to help quantify nutritional trade-offs in high dimensional nutritional data.

Table 1.1: Historical overview of geometric concepts applied to diet and nutrition. The overview also includes pioneering empirical work using the response surface models.

Year	Authors	Overview of contributions
1967	Chandler et al. (1967), Gardiner et al. (1967)	Pioneered the use of response surface models in nutritional sciences.
1968	Evans & Miller (1968)	Pioneered the use of equilateral triangle to represent diet mixtures within the context of comparative nutrition.
1972	McFarland & Sibly (1972)	Introduced the concepts of non-unitary drives and presented a framework to study animal motivation which later underpinned the developments of the Geometric Framework in the 1990s.
1973	Parks (1973)	Introduced geometric concepts to represent diet mixtures and proposed the use of response surface models to map dietary effects on animal growth. It also included the use of equilateral triangles to represent mixtures of macronutrients.
1974	Moon & Spencer (1974)	Used affine geometry to represent diet mixtures.
1974	Sibly & McFarland (1974)	Defined state-space models of animal motivation building upon their concepts of non-unitary drives from 1972.
1979	Roush et al. (1979)	Simplified Parks' 1972 approach and applied the concept of response surface models to study growth in the Japanese quail.
1982	Parks (1982)	Built upon Parks' 1972 approach and formalised the use of vectors to represent diet mixtures; it expanded the use of affine geometry to include the concept of metric into the vector space of nutrients.
1982–1993	Toyomizu et al. (1982, 1988, 1991, 1993)	Applied geometric concepts and response surface models to study animal nutrition in a series of pioneering empirical work.
1991	Emmans (1991)	Proposed the use of equilateral triangles to represent diet mixtures in a verbal non-formal fashion.
1993	Simpson & Raubenheimer (1993), Raubenheimer & Simpson (1993)	Formalised the Geometric Framework of nutrition as is widely used today. They also proposed the use of geometric concepts to study animals' rules of compromise when feeding in imbalanced diets.
2004	Simpson et al. (2004)	Formalised the construction of performance landscapes in the Geometric Framework of nutrition using Taylor series approximations. They also derived a way to infer the costs of nutrient surplus and deficit based on the geometry of the performance landscape.
2011	Raubenheimer (2011)	Formalised the use of right-angle triangle mixtures (RTM) to study nutrition, facilitating the applications of the Geometric Framework to field data.

Chapter 2

A vector-based approach to measure nutritional trade-offs within and between species

Publication reference

Morimoto, J. and Lihoreau, M., (2019). Quantifying nutritional trade-offs across multidimensional performance landscapes. *The American Naturalist*, 193(6), pp.E168-E181.

2.1 Abstract

Animals make feeding decisions in order to simultaneously maximise fitness traits that often require different nutritional optima. Current quantitative methods to characterise these nutritional trade-offs from performance landscapes on which traits are mapped on a multi-dimensional nutrient space can lead to inaccurate estimates, which greatly limits the possibility to explore scenarios with more than two nutrients or to compare strategies across species. Here we propose a new analytical approach – the Vector of Position Approach – that is based on vectors of position, and uses 2- and 3-D vector operations combined with machine learning algorithms to accurately measure nutritional trade-offs from performance landscapes for within- and between-species comparisons. Using landmark published datasets, we show that the current method can overestimate the strength of nutritional trade-offs by up to 65%. We then demonstrate that our approach gives accurate quantifications of nutritional trade-offs, applies to scenarios with more than two nutritional dimensions, and provides new insights into the underlying nutritional differences in trait expression between species. The Vector of Positions Approach provides a unique generalised framework for investigating nutritional differences in life-history traits expression within and between species, which allows future developments in comparative research on the evolution of animal nutritional strategies.

Keywords: Nutritional Geometry, nutritional trade-off, performance landscapes, fitness.

2.2 Introduction

Nutrition underpins complex life-history traits that determine individual fitness and impact on evolutionary processes (Simpson & Raubenheimer 2012). Recent advances in nutritional ecology show how an animals' diet can differently influence the expression of key life-history traits, leading animals to trade-off between optimising multiple traits simultaneously (e.g., Bunning et al. 2015, 2016, Rapkin et al. 2018, Morimoto & Wigby 2016). In insects, for instance, lifespan is typically enhanced on high carbohydrate diets whereas reproduction is maximised on high protein diets [e.g., fruit flies (Lee et al. 2008, Fanson et al. 2009, Reddiex et al. 2013, Jensen et al. 2015, Semaniuk et al. 2018) and crickets (Maklakov et al. 2008)]. Immunity and reproduction also display complex species-specific differences in nutritional requirements [e.g., moths (Cotter et al. 2011)], fruit flies (Ponton et al. 2015) and crickets (Maklakov et al. 2008), reviewed by Ponton et al. (2011)]. Even traits related to different stages of reproduction can have specific nutritional requirements that are not attainable simultaneously [fruit flies: (Reddiex et al. 2013, Jensen et al. 2015, Morimoto & Wigby 2016), cockroaches (Bunning et al. 2015), beetles (House et al. 2016)]. Measuring these nutritional trade-offs is challenging because it necessitates to separate the potential confounding effects of energy intake and nutrient balance on the expression of traits to identify the contribution of specific food components to the expression of one trait over another (Stearns 1989, Roff 2002, Hunt et al. 2004).

Over recent years, concepts of nutritional geometry (also known as the "Geometric Framework for nutrition") have been increasingly used to experimentally generate performance landscapes on which the expression of a fitness trait can be mapped onto the nutrient space, allowing the effects of nutrient and energy intakes on the optimal expression of traits to be disentangled (Simpson & Raubenheimer 1993, Raubenheimer & Simpson 1993, Lee et al. 2008, Maklakov et al. 2008, Fanson et al. 2009, Simpson & Raubenheimer 2012). Nutritional geometry has provided key insights into the nutritional factors underpinning a wide variety of physiological and behavioural phenomena across taxonomic groups, feeding guilds and ecological contexts (e.g., Trumper & Simpson 1993, Simpson et al. 2010, Lihoreau et al. 2015, Raubenheimer & Simpson 1993), including human societies (Gosby et al. 2014). In most cases, however, measures of nutritional trade-offs are based on the visual inspection of the performance landscapes. Developing a standard quantitative method for analysing these nutritional data has become a key issue for comparative research and investigate the role of nutrition in the evolution of physiology and behaviour (Lihoreau et al. 2015).

Bunning et al. (2015) and Rapkin et al. (2018) recently proposed a method for calculating the angle and the distance between the peaks of fitness-related traits in nutrient space of nutritional geometry studies (henceforth referred as the "Angular Method"; see description in this chapters' Materials and Methods and (Figure 2.1a-g). This method is based on the representation of performance landscapes as vectors of slopes and has been applied to measure nutritional trade-offs between immunity and reproduction, as well as between different reproductive traits (Bunning et al. 2015, 2016, Rapkin et al. 2016, 2018). Although informative, the Angular Method, as currently framed, can lead to misleading results, particularly when the estimates of the angle separating the nutritional peaks are interpreted along with the performance landscapes. The risk of confusion may rapidly become more important as the analyses get more complex, for instance in the context of comparative studies, when multiple performance landscapes must be compared, or in studies using complex (high dimension) nutritional spaces, when the effect of more than two food components are investigated (e.g., Solon-Biet et al. 2014, Gosby et al. 2014, Solon-Biet et al. 2015).

To address these problems, here we introduce a new, vector-based approach to accurately measure and compare nutritional trade-offs in performance landscapes: the "Vector of Position Approach". First, we show that the Vector of Position Approach provides an accurate measure of the angle between the peaks of the traits and an in-depth description of the nutritional trade-offs using a landmark dataset on lifespan-reproduction trade-offs on *Drosophila* (Lee et al. 2008). Second, we use simulated data to illustrate how our approach can be generalised to more complex high-dimension nutrient spaces, allowing for the comparison of the trade-offs in studies with three or more nutrients. Third, we show how this novel approach can be applied for between-species comparisons, using published datasets on lifespan-reproduction trade-offs in two fruit fly species (Lee et al. 2008, Fanson et al. 2009).

2.3 Material and Methods

2.3.1 Overview of the two methods

The Angular Method

The Angular Method uses vectors to describe the relationship between nutrient intake and the expression of traits, and measures the angle between the vectors to infer the presence and strength of nutritional trade-offs (Bunning et al. 2015). The first step is to fit a linear regression of the trait onto the intake of each nutrient of interest, forcing the intercept to 0. If we consider two nutrients, e.g. protein (P) and

carbohydrate (C), we obtain:

$$trait_i \sim \beta_{1,i}\mathrm{P} + \beta_{2,i}\mathrm{C} - 1 \tag{1}$$

From (1), the slopes $\beta_{1,i}$ and $\beta_{2,i}$ – the linear slopes of P and C – are used as the components of a 2D vector $\vec{a_i}$, in which i is the trait of interest,

$$\vec{a_i} = \begin{bmatrix} \beta_{1,i} & \beta_{2,i} \end{bmatrix} \tag{2}$$

In a simple example with two traits – traits A and B – with peaks in different regions of the nutrient space (Figure 2.1a), the vectors of slopes are defined by:

$$\vec{a_A} = \begin{bmatrix} \beta_{1,A} & \beta_{2,A} \end{bmatrix} \tag{3a}$$

$$\vec{a_B} = \begin{bmatrix} \beta_{1,B} & \beta_{2,B} \end{bmatrix} \tag{3b}$$

It is then possible to calculate the angle θ' formed between a $\vec{a_A}$ and a$\vec{a_B}$ vectors (Figure 2.1b-c) as:

$$\theta' = \cos^{-1}\left(\frac{\vec{a_A} \cdot \vec{a_B}}{||\vec{a_A}|| \cdot ||\vec{a_B}||}\right) \tag{4}$$

Note that in the original proposition, the angle between these vectors of slopes was called θ instead of θ' (Rapkin et al. 2018).

The value of θ' is the angle formed between diet vectors representing the peaks in trait expressions in the performance landscapes, and has been suggested as a useful measure of the nutritional trade-off between traits (Rapkin et al. 2018). θ' can take values between 0° and 180°, in which the closer θ' is from

0°, the more aligned the peaks in the nutrient space are expected to be, whereas the closer θ' is from 180°, the farther away the peaks are located in relation to each other (Rapkin et al. 2018). Therefore, larger values of the angle θ' indicate stronger nutritional trade-offs on the expression of two traits. Since the coordinates of the vectors of slope a $\vec{a_i}$ are obtained by the linear slopes β_1 and β_2 of P and C intakes onto the expression of the trait (Figure 2.1a-c), the angle θ' between a $\vec{a_i}$ vectors measures how different the slopes of P and C intakes are between the traits. θ' thus informs about whether the slopes of either P or C intakes, or both, have a positive, negative, or zero effect on the expression of the trait (Figure 2.1d). This is interesting because the Cartesian quadrant that a given a $\vec{a_i}$ vector yields information about the sign of the slope of the nutrient on the expression of the trait. For example, if a $\vec{a_i}$ lies on the second quadrant, the slope of C intake is positive whereas the slope of the P intake is negative (Figure 2.1d). This property will be explored to estimate the strength of the nutritional trade-offs using the Vector of Position Approach (see this chapters' results).

It is important to note that β can assume any value in the real range, that is, $\beta_j \in \mathbb{R}\ \forall j = 1, 2, ..., j$, where j is the number of linear slopes, which means that the vectors $\vec{a_i} = \begin{bmatrix} \beta_{1,i} & \beta_{2,i} \end{bmatrix}$ can be defined in any quadrant of the Cartesian plane. However, in nutritional geometry, the nutrient space (δ) is defined only in the range of real positive numbers ($\delta_n \in \mathbb{R}^+\ \forall n = 1, 2, ..., n$), where n is the number of nutrients being investigated (Simpson & Raubenheimer 2012). Therefore, vectors with negative coordinates are meaningless from the point of view of the nutrient space, and any measure of distance between vectors with negative components cannot represent the true angle between the peaks observed in the performance landscapes.

In addition to θ', Rapkin et al. (2018) extended the Angular Method to include estimates of the Euclidean distances d' between the peaks of two traits in the nutrient spaces (Figure 2.1e), calculated as:

$$d' = \sqrt{\left(P_A^* - P_B^*\right)^2 + \left(C_A^* - C_B^*\right)^2} \quad \text{units} \tag{5a}$$

or alternatively, according to geometrical laws Heath (1956):

$$d' = \sqrt{\vec{a_A}^2 + \vec{a_B}^2 - 2\vec{a_A}\vec{a_B}\cos(\theta')} \quad \text{units} \tag{5b}$$

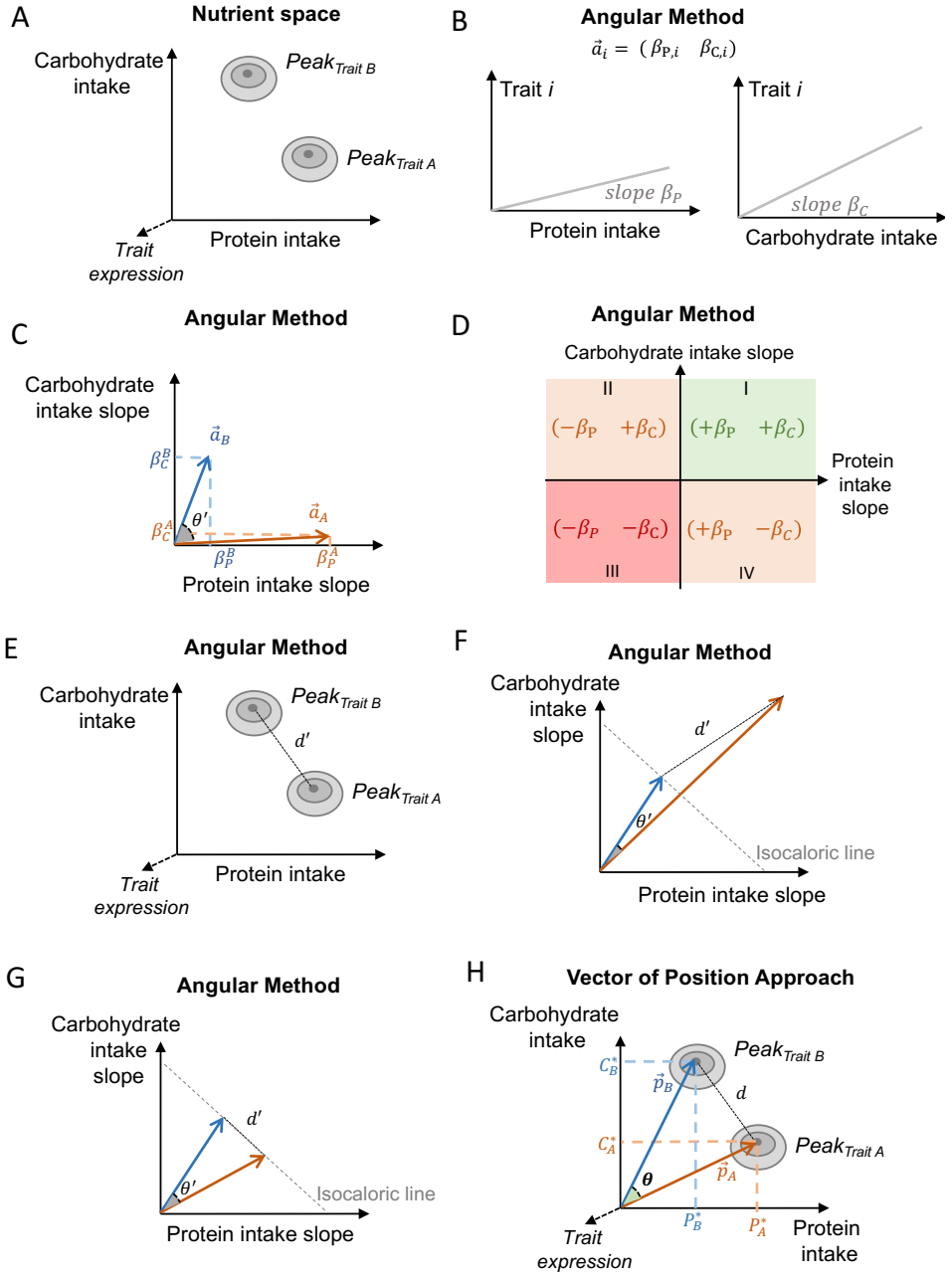


Figure 2.1: **Overview of Angular Method and Vector of Position Approach.** (a) Hypothetical performance landscape with peaks of traits A and B (grey regions). (b) Linear slopes of protein and carbohydrate intakes onto the trait i that composes the vector $\vec{a}_i$. (c) Angle θ' that separates the vectors $\vec{a}_A$ and $\vec{a}_B$. (d) Information of the slopes within slope vectors $\vec{a}_i$. Note that the direction of vectors $\vec{a}_i$ in the Cartesian plane contains information about the sign of the slopes of protein and carbohydrate intakes onto the trait i. (e) Euclidean distance d' between the peaks of traits A and B. (f) Schematic representation of the relationship between θ' and d' for nutritional peaks in different isocaloric lines. (g) Schematic representation of the relationship between θ' and d' for nutritional peaks in the same isocaloric line. (h) Position vectors $\vec{p}_i$ that describe the relationship between the peaks of traits A and B in the performance landscape.

where P_i^* and C_i^* are the P and C coordinates of the peak, $\vec{a}_A$ and $\vec{a}_B$ are the slope vectors, and θ' is the angle between the slope vectors. Units are given in milligrams (mg), grams (g), or any other unit used when measuring nutrient intakes. We discuss the validity of equation (5b) in the Angular Method in the following sections. Rapkin et al. (2018) suggested that if the peaks are located in different isocaloric

lines (i.e., lines with slope -1) in the nutrient space, d' becomes a more reliable measure of the nutritional trade-off than θ', because θ' can underestimate the true distance between the peaks (Rapkin et al. 2018) (Figure 2.1f). However, if the peaks lie in the same isocaloric lines, large θ' would be associated with large d and vice versa (Figure 2.1g).

The Vector of Position Approach

The *Vector of Position Approach* is an extension of the Angular Method that builds upon the representation of performance landscapes through vectors, while removing the confusion regarding the interpretations of the angle θ' by using vectors of positions rather than vectors of slopes. Instead of using the slopes of the linear regression as the components of the vector $\vec{a}_i$, we obtain a point estimate of the peak in the performance landscape [(see e.g., 'OptimaRegion package del Castillo et al. 2016, for this purpose)].

From (5), the point estimate of the peak in trait expression has coordinates (P^*, C^*), in which P^*, C^* are the P and C intakes, respectively, that maximize the trait, and can be represented by a vector $\vec{p}_i$.

$$\vec{p}_i = \begin{pmatrix} P_i^* & C_i^* \end{pmatrix} \tag{6}$$

In the example used above, the vectors for traits A and B are:

$$\vec{p}_A = \begin{pmatrix} P_A^* & C_A^* \end{pmatrix} \tag{7a}$$

$$\vec{p}_B = \begin{pmatrix} P_B^* & C_B^* \end{pmatrix} \tag{7b}$$

It is then possible to calculate the observed degree of separation – the angle θ – between the vectors $\vec{p}_A$ and $\vec{p}_B$ as:

$$\theta = \cos^{-1}\left(\frac{\vec{p}_A \cdot \vec{p}_B}{\|\vec{p}_A\|\|\vec{p}_B\|}\right) \tag{8}$$

Using the vectors of positions guarantees that $\vec{p}_A$ and $\vec{p}_B$ always lie in the defined region of the performance landscape (i.e., the first quadrant of the Cartesian plane). In this case, an angle $\theta = 90°$ would represent trait performance peaks that are located at the exact opposite ends of the nutrient space, but still within the first quadrant. This is important when the intake of a nutrient is inversely correlated with the expression of the trait being analysed, since in these cases, the slope of the nutrient on the trait

expression $\beta_{1,i}$ is negative, and the vector of slope $\vec{a}$ would lie in the second, third, or fourth quadrant of the Cartesian plane, which is outside the defined region of the nutrient space.

Once θ is calculated, we can calculate the Euclidean distance d as:

$$d = \sqrt{\vec{p}_A^2 + \vec{p}_B^2 - 2\vec{p}_A \cdot \vec{p}_B \cos(\theta)} \tag{9}$$

The Law of Cosines (Heath 1956) formalizes the relationship between θ and d. Note, however, that the Law of Cosines cannot be used to calculate d' using the θ' (see equation 5b) because θ' does not measure the true angle between the peaks in the performance landscape (see Results).

2.3.2 Data and statistical analyses

The data for the *D. melanogaster* study was obtained from Lee et al. (2008). The data for the Queensland fruit fly *B. tryoni* study was obtained from Fanson et al. (2009). Briefly, both studies used nutritional geometry (Simpson & Raubenheimer 2012) to investigate the effects of protein and carbohydrate intake on lifespan, lifetime reproduction (total number of eggs laid), and reproductive rate (number of eggs laid per day) (see Supplementary Methods at the end of this Chapter 2 for details).

Performance landscapes were plotted using the *Tps()* function of the 'fields' package (Nychka et al. 2017). To obtain the point estimates for the performance landscapes in the absolute scale, we extracted the coordinates of the maximum fitted values of the function Tps. To obtain the slopes of P and C intake on lifespan, lifetime reproductive success, and reproductive rate, we used the function *MCMCglmm()* from the 'MCMC' package (Hadfield 2010), with uninformative prior $nu = 0.2$ expected covariances $V = 1$, 400,000 iterations, with burn-in of 2000 and thinning parameter of 25 [as in (Bunning et al. 2015)]. For the comparison between the two species, we standardized the P and C intakes, as well as the trait values by mean-subtraction as described in the main text ('Results' section of this chapter, equations (13) and (14)). To estimate the optimum region in the standardized performance landscapes, we used a support-vector machine (SVM) regression model with the 'caret' package (Kuhn 2017), from which we could select the coordinates of the simulated P and C intakes that resulted in the predictions that matched with the optimum region in the performance landscapes (see Supplementary Methods of this chapter for details). All analyses were performed in R version 3.4.0 (Team 2017).

2.4 Results

2.4.1 The Angular Method overestimates the true degree of separation between peaks in the performance landscape

To illustrate the advantage of the *Vector of Position Approach* relative to the Angular method, we analysed and compared data from (Lee et al. 2008) with both methods. This study explored nutritional trade-offs between lifespan, lifetime reproductive success (LRS), and reproductive rate (RR) in *D. melanogaster*. Visual inspection of the performance landscapes indicates that lifespan is maximised in diets with low P:C ratio of ∼1:16, LRS is maximised at a P:C ∼1:4, whereas RR is maximised at P:C ∼1:2 (Figure 2.2).

Table 2.1: Matrix of estimates of the angles between vectors of positions (θ) and slopes (θ') (in degrees). Note that nutritional trade-offs between lifespan and reproduction in *D. melanogaster* (data from Lee et al. 2008) as calculated using the Angular Method (θ') (above the diagonal of the matrix) and the Vector of Position Approach (θ) (below the diagonal of the matrix). Estimates (95% credible intervals) are given in degrees. LRS: lifetime reproductive success. RR: reproductive rate.

	Lifespan	**LRS**	**RR**
Lifespan	–	$\theta' = 15.082$ (12.022, 18.123)	$\theta' = 63.210$ (59.876, 66.438)
LRS	$\theta = 14.073$ (10.857, 19.889)	–	$\theta' = 48.112$ (43.945, 52.149)
RR	$\theta = 26.390$ (25.471, 27.385)	$\theta = 12.322$ (7.496, 14.613)	–

Using the Angular Method, we estimated that $\vec{a}_{RR}$ and $\vec{a}_{lifespan}$ were separated by an angle $\theta' = 63.21°$ (95% CI: 59.87°, 66.43°). $\vec{a}_{lifespan}$ and $\vec{a}_{LRS}$ were separated by $\theta' = 15.08°$ (95% CI: 12.02°, 18.12°) and the vectors $\vec{a}_{LRS}$ and $\vec{a}_{RR}$ were separated by $\theta' = 48.11°$ (95% CI: 43.94°, 52.14°). These results confirm the expectation of a greater potential for a nutritional trade-off between lifespan–RR compared with lifespan–LRS and LRS–RR (Lee et al. 2008). The estimates of d' between lifespan–RR was 294.89 mg, between lifespan–LRS was 155.66 mg, and between LRS–RR was 261.91 mg.

Using the *Vector of Positions Approach*, we determined the coordinates of the vectors of position $\vec{p}_{lifespan} : \langle 0, 176.717 \rangle$, $\vec{p}_{LRS} : \langle 75.001, 299.174 \rangle$, and $\vec{p}_{RR} : \langle 141.198, 284.488 \rangle$, which corresponds, as expected, to a P:C ratio of ∼1:16, ∼1:4, and ∼1:2 as described above and in Lee et al. (2008) (Figure 2.2a-c), and calculated the angles θ that separated the vectors $\vec{p}_i$ (Table 2.1). The separation between the nutritional peaks of lifespan and RR was greater than the nutritional peaks of lifespan and LRS, or LRS and RR. These estimates of θ were confirmed using values from a standard table for θ based on the most commonly used P:C ratios in nutritional geometry studies on *Drosophila* (Table 2.2).

We then calculated d according to $\vec{p}_i$ and θ. As expected, the estimates of d were longer between

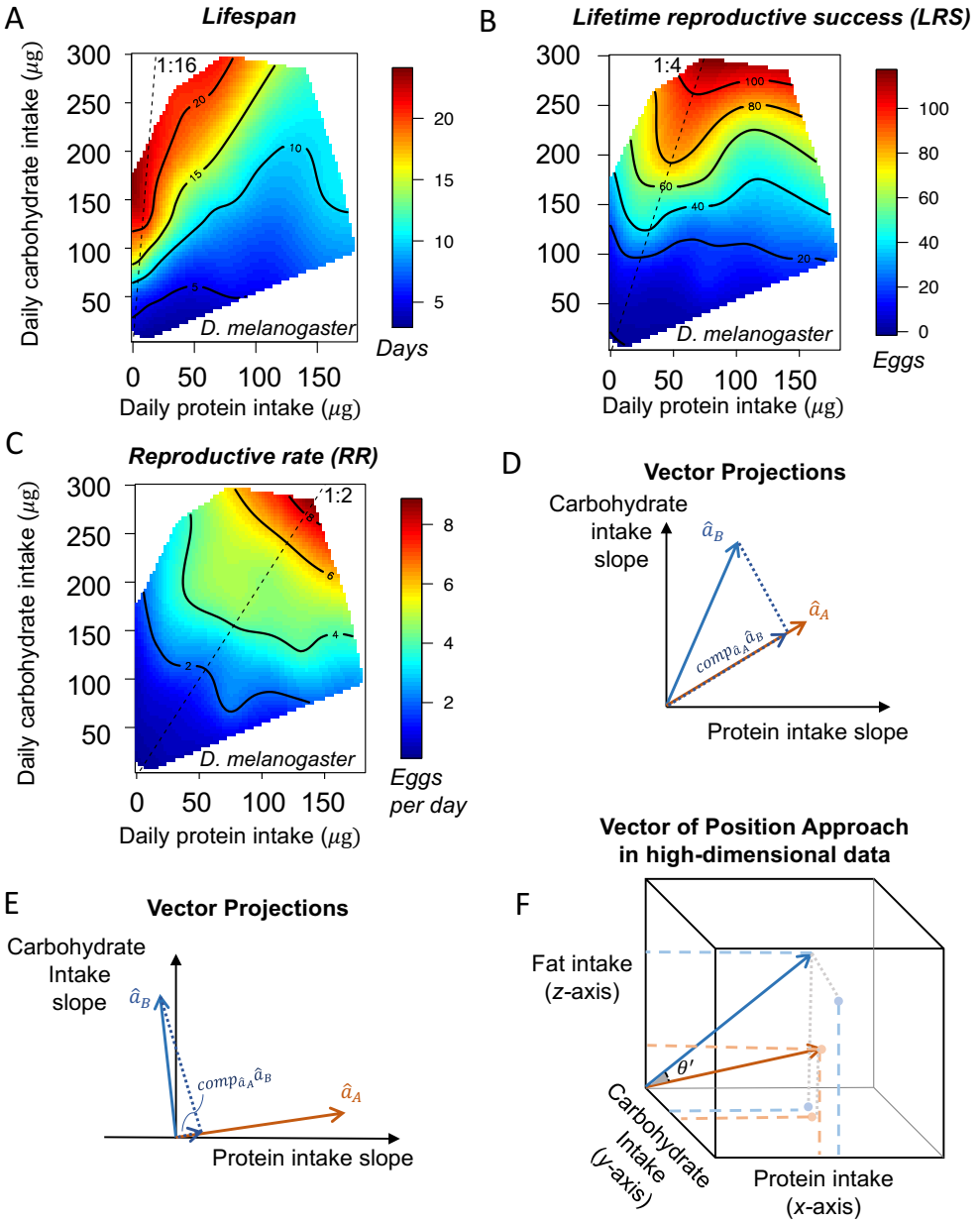


Figure 2.2: **Performance landscapes and vector projections.** (a-c) Performance landscapes of lifespan (a), lifetime reproductive success (LRS) (b), and reproductive rate (RR) (c) of *D. melanogaster* females (data from Lee et al. 2008). (d-e) Vector projection to calculate the component of the unit vector $\hat{a}_B$ onto the unit vector $\hat{a}_A$ (i.e., $\text{comp}_{(\hat{a}_A)}\hat{a}_B$). (d) Example of a large component of $\hat{a}_B$ into $\hat{a}_A$. **E)** Example of a small component of $\hat{a}_B$ into $\hat{a}_A$. (f) Schematic representation of a 3D vector in a performance landscape of protein, carbohydrate, and fat intakes.

$\vec{p}_{lifespan}$ and $\vec{p}_{RR}$ ($d = 205.02$ mg), compared to between $\vec{p}_{lifespan}$ and $\vec{p}_{LRS}$ ($d = 154.25$ mg) and $\vec{p}_{RR}$ and $\vec{p}_{LRS}$ ($d = 94.90$ mg), thereby indicating a larger separation between the peaks of lifespan and RR. Together, these results confirm the nutritional trade-offs between lifespan, LRS, and RR observed by Lee et al. (2008) [and more recently by Reddiex et al. (2013) and Jensen et al. (2015)], and suggest that nutritional trade-offs are stronger between female RR and lifespan.

The comparison between the two approaches shows that the Angular Method can lead to misleading

Table 2.2: Protein-to-carbohydrate (P:C) ratios of the valleys in performance landscapes. Note that estimated P:C ratios of the valleys in the landscapes for lifespan, lifetime reproductive success (LRS), and reproductive rate (RR) as calculated by the Vector of Position Approach (data from Lee et al. (2008)). Data shown as Estimate and 95% credible intervals (CI).

Life-history trait	Estimate	Lower CI	Upper CI
Lifespan	1:1.87	1:1.57	1:2.17
LRS	1:13.64	1:12.41	1:14.86
RR	1:13.64	1:13.55	1:13.72

results and misinterpretation of the distance between peaks in the performance landscape. For instance, our results reveal a 2-fold overestimation of the angle separating the peaks of lifespan and RR when calculating θ' compared to θ, and a 4-fold overestimation of the angle separating peaks of RR and LRS. Such overestimation of the distance between peaks could lead to an erroneous interpretation of the P:C ratios diets upon which the peaks are maximised because a separation of 63.21° that was given for the vectors of lifespan and RR is equivalent to a separation between P:C ratios of 2:1 (high-protein) and 0:1 (pure sugar), which are considerably more distant in the performance landscape than the true P:C ratios of 1:16 to 1:2 that maximised lifespan and reproductive rates (see Table 2 for angles between commonly used P:C ratios). Furthermore, the inaccuracy in the estimation to an overestimation of the Euclidean distance between peaks was of approximately 30.4% for LS and RR, approximately 1% for LS and LRS, and 64.7% for LRS and RR.

2.4.2 Vector projections can be used to extract information from vectors of slope

Because of the way in which the coordinates of the vectors of slopes $\vec{a}_i$ are obtained, θ' contains important information on whether the slopes of either P or C intakes, or both, have a positive, negative, or zero effect on the expression of the trait. Using the concepts of vector projection, the $\vec{a}_i$ can be used to quantify the influence of each nutrient on the expression of traits. For instance, consider the scalar projection ("component") of the unit vector $\hat{\vec{a}}_B$ onto the unit vector $\hat{\vec{a}}_A$ (Figure 2.2d-e):

$$\text{comp}_{\hat{\vec{a}}_A} \hat{\vec{a}}_B = \frac{\hat{\vec{a}}_A \cdot \hat{\vec{a}}_B}{\|\hat{\vec{a}}_A\|} \tag{10}$$

where $\hat{\vec{a}}_i$ is the unit vector with the same direction as $\vec{a}_i$, defined as:

$$\hat{\vec{a}}_i = \frac{\vec{a}_i}{\|\vec{a}_i\|} \tag{11}$$

The scalar projection $\text{comp}_{\hat{\vec{a}}_A}\hat{\vec{a}}_B$ measures how much of the vector $\hat{\vec{a}}_B$ is present in vector $\hat{\vec{a}}_A$ (hence the saying "component of" $\hat{\vec{a}}_B$ onto $\hat{\vec{a}}_A$). Large values of the $\text{comp}_{\hat{\vec{a}}_A}\hat{\vec{a}}_B$ indicate that $\hat{\vec{a}}_B$ shares a large component in the direction of $\hat{\vec{a}}_A$, and there exists only a weak potential for a nutritional trade-off between traits (e.g., Figure 2.2d). Conversely, small values of $\text{comp}_{\hat{\vec{a}}_A}\hat{\vec{a}}_B$ indicate that $\hat{\vec{a}}_B$ shares only a small component in the direction of $\hat{\vec{a}}_A$, and therefore there exists a strong potential for a nutritional trade-off between traits A and B (e.g., Figure 2.2e).

We applied the Vector of Position Approach to calculate the scalar projection of the unit vector $\hat{\vec{a}}_{\text{lifespan}}$ onto the unit vectors $\hat{\vec{a}}_{\text{LRS}}$ and $\hat{\vec{a}}_{\text{RR}}$ on the *D. melanogaste*r data (Lee et al. 2008). The scalar projection of $\vec{a}_{\text{lifespan}}$ measures the magnitude of the vector $\hat{\vec{a}}_{\text{lifespan}}$ in the direction of $\hat{\vec{a}}_{\text{RR}}$ and $\hat{\vec{a}}_{\text{LRS}}$; the larger the scalar projection, the greater the component of $\hat{\vec{a}}_{\text{lifespan}}$ in the direction of $\hat{\vec{a}}_{\text{RR}}$ and $\hat{\vec{a}}_{\text{LRS}}$, and thus the lower the potential for a nutritional trade-off between these traits. The results confirmed the stronger nutritional trade-off between lifespan and RR as the scalar projection between the unit vectors $\hat{\vec{a}}_{\text{lifespan}}$ and $\hat{\vec{a}}_{\text{RR}}$ was 0.053 (95% CI: 0.047, 0.058), whereas the scalar projection between unit vectors $\hat{\vec{a}}_{\text{lifespan}}$ and $\hat{\vec{a}}_{\text{LRS}}$ and $\hat{\vec{a}}_{\text{LRS}}$ and $\hat{\vec{a}}_{\text{RR}}$ were 1.716 (95% CI: 1.632, 1.799) and 10.039 (95% CI: 9.188, 10.899), respectively.

Finally, we investigated what nutrient was likely to have the most influence in the trade-off between lifespan and reproductive rate. To do this, we inspected the direction of $\vec{a}_{\text{lifespan}}$ and $\vec{a}_{\text{RR}}$ on the Cartesian plane. While $\vec{a}_{\text{RR}}$ lies in the first quadrant with positive slopes for both P and C intake on RR, $\vec{a}_{\text{lifespan}}$ is located in the second quadrant, with a negative slope of P intake and a positive slope of C intake on lifespan. Thus, the nutritional trade-off between lifespan and RR is likely driven by the opposite effects of P intake on lifespan and reproduction. Although these results demonstrate the Vector of Positions approach applied to two nutrients, the Vector of Positions approach can, in theory, be applied to n nutrients (see Figure 2.4 and supplementary files of this chapter).

2.4.3 The Vector of Position Approach can be used for comparing species

In addition to comparing different traits between classes of individuals in the same species (e.g. males and females), our approach can be useful to compare traits or trade-offs across species. In these cases, the angle that separates the vectors of P and C intake is not sufficient because we are now interested in both nutrient intake as well as the expression level of the trait between species. For simplicity, let's return to our 2D-vectors of P and C, and this time, we shall include the 3rd dimension to the vectors that describe the expression level of the trait (Figure 2.3a). This step is needed because the trait can be

maximised in the same location in the nutrient space, but with different expression levels (i.e. difference may be on the z-axis). Consider the vector of positions $\vec{p}_{\text{sp i}}$ as in (6). To work with different species, it is necessary to standardize the coordinates of the vector so that the vectors $\vec{p}_i$ are now defined as:

$$\hat{\vec{p}}_{\text{sp1}} = \left(P^*_{\text{sp1}}, C^*_{\text{sp1}}, w^*_{\text{sp1}}\right) \tag{13a}$$

and

$$\hat{\vec{p}}_{\text{sp2}} = \left(P^*_{\text{sp2}}, C^*_{\text{sp2}}, w^*_{\text{sp2}}\right) \tag{13b}$$

where

$$\hat{P}_{\text{sp1}} = \frac{P}{\overline{P}}, \quad \hat{C}_{\text{sp1}} = \frac{C}{\overline{C}}, \quad \hat{w}_{\text{sp1}} = \frac{w}{\overline{w}} \tag{14}$$

$\hat{P}, \hat{C}, \hat{w}$ are the standardized values of the protein (P) and carbohydrate (C) intake, and the trait of interest (w). Having defined the standardized position vectors $\hat{\vec{p}}_{\text{sp1}}$ and $\hat{\vec{p}}_{\text{sp2}}$, we can calculate the angle θ that separates these two vectors. Here, θ also includes the differences in the expression of traits (on the z-axis), which, in biological terms, can be interpreted as the difference in the maximum standardized expression of the trait of interest when species are at the peak in the performance landscape. It is also possible to calculate the directional angles α, β, γ separating the vector and the x-, y-, and z-axes, respectively, which provides information regarding the differences in standardized P and C intake, as well as trait expression between species (Figure 2.3b). α, β, γ are calculated as follows:

$$\alpha = \cos^{-1}\left(\frac{P^*_{\text{sp1}}}{\sqrt{P^{*2}_{\text{sp1}} + C^{*2}_{\text{sp1}} + w^{*2}_{\text{sp1}}}}\right) \tag{15a}$$

$$\beta = \cos^{-1}\left(\frac{C^*_{\text{sp1}}}{\sqrt{P^{*2}_{\text{sp1}} + C^{*2}_{\text{sp1}} + w^{*2}_{\text{sp1}}}}\right) \tag{15b}$$

$$\gamma = \cos^{-1}\left(\frac{w^*_{\text{sp1}}}{\sqrt{P^{*2}_{\text{sp1}} + C^{*2}_{\text{sp1}} + w^{*2}_{\text{sp1}}}}\right) \tag{15c}$$

If the peaks are located in the same region of the standardized performance landscapes, the angles α and β are likely to be similar between species 1 and 2, as the vectors $\hat{\vec{p_{sp_i}}}$ point in the same direction in space (Figure 2.3b). This property could be used to infer the mechanisms underlying the species-specific

effects of nutrition on trait expression. For example, consider the x-axis as being P intake. If the angles α_{sp1} and α_{sp2} measure the distance between vectors and the x-axis, the angles α_{sp1} and α_{sp2} indicate how strongly the standardized optimal P intake influences the expression of traits for each species. If the difference between α_{sp1} and α_{sp2} is large, the optimal standardized P intake between species are located far from each other, and thus P intake is likely a key modulator of the differences in trait expression between the species.

Table 2.3: Comparative application of the Vector of Positions Approach. Note that comparisons between the standardized performance landscapes of reproductive rate between two fruit fly species: *D. melanogaster* (data Lee et al. 2008) and *B. tryoni* (data from Fanson et al. (2009)). Estimates and their 95% credible intervals of the direction angles α, β, and γ for the vectors $\vec{a}_{\text{Drosophila}}$ and $\vec{a}_{\text{Bactrocera}}$. All angles are given in degrees.

Species	**Estimate**	**Lower CI**	**Upper CI**
Angle α			
Drosophila	50.02053	48.22801	52.16193
Bactrocera	29.34615	25.38286	38.79428
Angle β			
Drosophila	57.76124	56.38002	59.38668
Bactrocera	67.18635	66.7606	69.50664
Angle γ			
Drosophila	56.62615	55.7261	56.98359
Bactrocera	72.55797	58.69577	80.35599

To illustrate this approach, we compared the nutritional requirements for the reproductive rate of two fruit fly species: *D. melanogaster* (Lee et al. 2008) and *B. tryoni* (Fanson et al. 2009) (Figure 2.3c-e). We standardized the performance landscapes of the two species according to equation (14). Using the previously described machine learning algorithm (see *Statistical analyses* section in this chapter), we estimated the region of the standardized peak in reproductive rate in the standardized performance landscape for both species (Figure 2.3c-e), and used these values as the coordinates of the standardized vectors of position $\vec{p}_{\text{Drosophila}}$ and $\vec{p}_{\text{Bactrocera}}$ according to (13). We then calculated the average, the minimum, and maximum values of the θ angle separating $\vec{p}_{\text{Drosophila}}$ and $\vec{p}_{\text{Bactrocera}}$. On average, we found that $\theta = 21.28°$, with minimum value of $11.30°$ and maximum value of $29.88°$, suggesting a relatively large degree of separation between the expression of the trait between the two species.

We then investigated this separation in terms of P and C intake and expression level of the trait by calculating the angle α, β, and γ between the vectors $\vec{p}_{\text{Drosophila}}$ and $\vec{p}_{\text{Bactrocera}}$ and the standardized x-, y-, and z-axes, respectively (Table 2.3). We found a marked difference in the angles α and γ (ca. $21°$ and $16°$, respectively), whereas there was a small difference for the angle β (ca. $9°$), suggesting that the difference in nutritional requirements for the reproductive rate of females of the two species is mainly

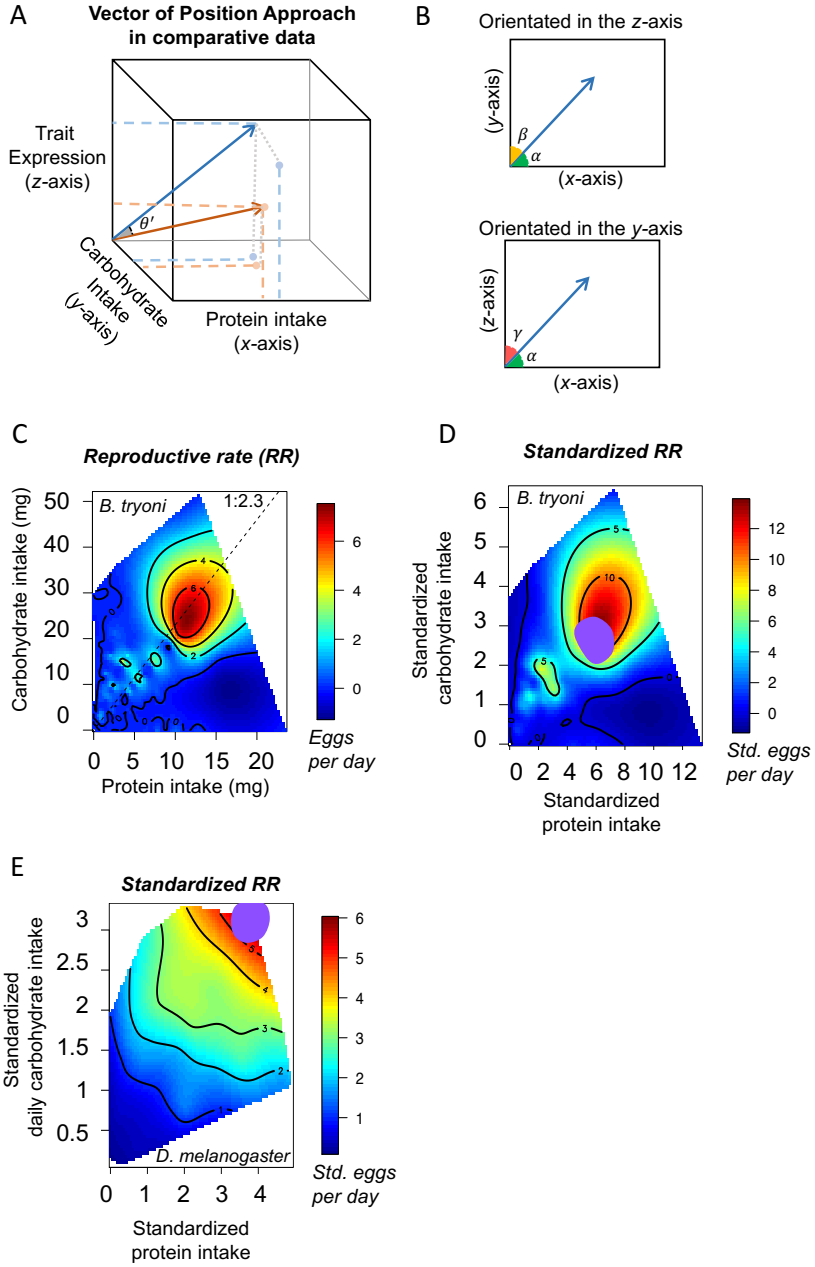


Figure 2.3: **Between-species comparison.** (a) Schematic representation of a 3D vector in a performance landscape of protein, carbohydrate, and trait expression. (b) Directional angles α, β, γ. Top panel: figure oriented according to the z-axis (perpendicular to the plane of the paper). Bottom panel: figure oriented according to the y-axis (perpendicular to the plane of the paper). (c) Performance landscapes of reproductive rate (RR) of *B. tryoni* females [data from Fanson et al. (2009)]. The dashed line highlights the P:C ratio that maximises the trait. (d-e) Mean-standardized performance landscapes of the reproductive rate of D. *D. melanogaster* [data from Lee et al. (2008) and *B. tryoni* from Fanson et al. (2009)]. Blue regions highlight the region estimated as the peak in the performance landscape from the SVM machine learning model.

driven by differences in species-specific standardized P intake and expression level of the trait.

2.5 Discussion

We introduced a vector-based approach to quantify the strength of nutritional trade-offs in nutritional geometry data and compare them across studies. Previous methods involved visual inspection of nutrient spaces or calculations of vectors of slopes that can lead to misinterpretations. Our approach strictly considers vectors of slopes in the real positive region of the nutrient space and calculates the true separation between nutritional peaks and the strength of nutritional trade-offs, which facilitates quantitative analyses of nutritional trade-offs in a broad range of studies, including complex experimental designs.

Our approach is not constrained to two dimensions but can be generalized to high-dimension nutritional data. Such generalization is key when dealing with three or more nutrients (or any other relevant food components), in which data visualisation can be difficult (if possible), and conducting analysis with the Angular Method with negative vector coordinates can lead to misleading results. Although initially developed for analyses with two nutrients (e.g., typically the macronutrients carbohydrates and protein) (Simpson & Raubenheimer 1993), nutritional geometry studies are increasingly utilising high-dimension nutrient spaces, as illustrated by recent data on the importance of carbohydrates, protein, and fat in reproduction, lifespan, and other aspects of animal and human health (Hewson-Hughes et al. 2011, 2012, Solon-Biet et al. 2014, 2015, Gosby et al. 2014). Studies on insects have also begun to look at the effects of individual amino acids (rather than full proteins) on the expression of fitness traits (e.g., Grandison et al. 2009, Arganda et al. 2017, Piper et al. 2017). Future systematic quantifications of the effects of individual amino acids in foods on physiology and behaviour may therefore require analyses in nutrient spaces with as many as 22 dimensions, for which visual inspections of nutrient spaces or inaccurate estimates of distance between performance peaks will not suffice. Many other studies have started to consider other food components such as individual fatty acids (Arien et al. 2015), mineral salts (Simpson et al. 2006), or even water (Fanson et al. 2009). Our approach therefore opens new opportunities to quantify complex nutritional trade-offs and provides the first quantitative method to quantify nutritional trade-offs in high-dimensional nutritional data.

The Vector of Positions Approach also provides a quantitative framework through which the nutritional effects of the expression of the same trait can be compared within species and between species. By reanalysing two published datasets in fruit flies, we quantified species-specific differences in female reproductive rate and identified that this difference is driven by differences in species-specific standardized protein intake and expression level of the trait. A potential explanation for this result is that Lee

et al. (2008) used mated females of *D. melanogaster*, while Fanson et al. (2009) used unmated females of *B. tryoni*. In *D. melanogaster*, mating increases protein intake and egg production rate in response to seminal fluid proteins transferred in the male ejaculate (Yapici et al. 2008, Ribeiro & Dickson 2010, Gligorov et al. 2013, Perry et al. 2013, Walker et al. 2015). Thus, differences in experimental design and in the biological state of the organisms used in the study could underpin the effects found between species, calling for future studies using standardised protocols as well.

Comparative studies of nutritional traits across species can yield fundamental insights into the role of nutrition on the evolution of physiological and behavioural strategies as well as important ecological processes (Behmer & Joern 2008). In primates, comparative research shows that nutrient availability in ancestral diets has shaped appetites and rules of compromises that greatly vary among species (Felton et al. 2009, Rothman et al. 2011, Johnson et al. 2013), yielding key insights for our understanding of the human obesity epidemic (Simpson & Raubenheimer 2014). Comparing species based on their nutritional properties is a powerful means to study the role of nutrition in social evolution (Lihoreau et al. 2015, 2017), for instance to explain social transitions from populations of isolated animals to aggregations of several million individuals (Simpson et al. 2006). All these comparisons are currently made by eye. Future developments of quantitative approaches such as the one proposed here therefore yield considerable promise to understand the role of nutrition in the evolution of animal physiology, behaviour, and ecology.

2.6 Supplementary Information

2.6.1 Supplementary Figures and Tables

Table 2.4: Convergence and robustness of the Bayesian models to different V and nu parameters. The low marginal standard deviations of the posterior mean indicate that the models converged and were robust across all variations of the parameters

Prior Robustness													
		nu = 0.0											
		V = 0.5			V = 1*			V = 2			V = 3			Std dev. of post. mean
Trait	Factor	post.mean	l-95%	u-95%	post.mean	l-95%	u-95%	post.mean	l-95%	u-95%	post.mean	l-95%	u-95%	(across V parameters)
Lifespan	P	-0.05843	-0.06865	-0.04788	-0.05843	-0.06896	-0.04819	-0.05844	-0.06894	-0.04862	-0.05837	-0.06852	-0.04831	3.2016E-05
	C	0.13417	0.12873	0.13957	0.13417	0.12894	0.13968	0.13415	0.12882	0.13965	0.13415	0.12869	0.13941	1.1547E-05
LRS	P	-0.03823	-0.07710	0.00185	-0.03810	-0.07787	0.00028	-0.03808	-0.07689	0.00098	-0.03847	-0.07675	0.00152	1.8007E-04
	C	0.25707	0.23641	0.27709	0.25702	0.23720	0.27824	0.25700	0.23691	0.27742	0.25712	0.23604	0.27698	5.4721E-05
RR	P	0.01103	0.00837	0.01371	0.01104	0.00838	0.01357	0.01104	0.00832	0.01363	0.01103	0.00843	0.01376	4.8990E-06
	C	0.01328	0.01187	0.01464	0.01328	0.01195	0.01471	0.01328	0.01189	0.01468	0.01328	0.01189	0.01467	1.9149E-06
		nu = 0.02*												
		V = 0.5			V = 1*			V = 2			V = 3			Std dev. of post. mean
Trait	Factor	post.mean	l-95%	u-95%	post.mean	l-95%	u-95%	post.mean	l-95%	u-95%	post.mean	l-95%	u-95%	(across V parameters)
Lifespan	P	-0.05847	-0.06895	-0.04832	-0.05849	-0.06878	-0.04814	-0.05837	-0.06845	-0.04788	-0.05845	-0.06879	-0.04832	5.2599E-05
	C	0.13415	0.12887	0.13971	0.13417	0.12878	0.13959	0.13412	0.12841	0.13919	0.13416	0.12877	0.13950	2.1602E-05
LRS	P	-0.03843	-0.07688	0.00343	-0.03802	-0.07841	0.00018	-0.03808	-0.07819	0.00052	-0.03820	-0.07799	0.00151	1.8419E-04
	C	0.25710	0.23687	0.27788	0.25706	0.23646	0.27768	0.25719	0.23733	0.27940	0.25709	0.23686	0.27833	5.5390E-05
RR	P	0.01103	0.00831	0.01364	0.01103	0.00844	0.01371	0.01104	0.00832	0.01369	0.01104	0.00840	0.01361	4.3970E-06
	C	0.01328	0.01191	0.01468	0.01328	0.01186	0.01462	0.01328	0.01192	0.01471	0.01328	0.01186	0.01464	2.5000E-06
		nu = 0.05												
		V = 0.5			V = 1*			V = 2			V = 3			Std dev. of post. mean
Trait	Factor	post.mean	l-95%	u-95%	post.mean	l-95%	u-95%	post.mean	l-95%	u-95%	post.mean	l-95%	u-95%	(across V parameters)
Lifespan	P	-0.05844	-0.06864	-0.04825	-0.05834	-0.06871	-0.04777	-0.05844	-0.06891	-0.04827	-0.05846	-0.06877	-0.04818	5.4160E-05
	C	0.13415	0.12901	0.13960	0.13410	0.12885	0.13964	0.13416	0.12882	0.13965	0.13419	0.12881	0.13957	3.7417E-05
LRS	P	-0.03813	-0.07813	0.00101	-0.03825	-0.07786	-0.00055	-0.03818	-0.07855	0.00028	-0.03852	-0.07931	-0.00101	1.7483E-04
	C	0.25705	0.23620	0.27801	0.25715	0.23657	0.27728	0.25703	0.23646	0.27775	0.25721	0.23608	0.27711	8.4185E-05
RR	P	0.01105	0.00841	0.01369	0.01102	0.00843	0.01376	0.01105	0.00846	0.01374	0.01102	0.00838	0.01370	1.7385E-05
	C	0.01328	0.01189	0.01466	0.01329	0.01191	0.01470	0.01328	0.01191	0.01468	0.01329	0.01189	0.01467	6.2383E-06
		nu = 0.1												
		V = 0.5			V = 1*			V = 2			V = 3			Std dev. of post. mean
Trait	Factor	post.mean	l-95%	u-95%	post.mean	l-95%	u-95%	post.mean	l-95%	u-95%	post.mean	l-95%	u-95%	(across V parameters)
Lifespan	P	-0.05845	-0.06850	-0.04821	-0.05837	-0.06880	-0.04843	-0.05837	-0.06852	-0.04805	-0.05845	-0.06865	-0.04817	4.6188E-05
	C	0.13419	0.12882	0.13954	0.13414	0.12871	0.13951	0.13414	0.12879	0.13945	0.13416	0.12868	0.13955	2.3629E-05
LRS	P	-0.03831	-0.07797	0.00067	-0.03831	-0.07808	0.00074	-0.03803	-0.07755	0.00149	-0.03860	-0.07755	0.00063	2.3343E-04
	C	0.25712	0.23624	0.27795	0.25709	0.23650	0.27774	0.25703	0.23641	0.27779	0.25726	0.23746	0.27817	9.5630E-05
RR	P	0.01103	0.00846	0.01373	0.01104	0.00840	0.01361	0.01104	0.00838	0.01365	0.01105	0.00847	0.01371	7.6158E-06
	C	0.01328	0.01187	0.01464	0.01329	0.01189	0.01468	0.01329	0.01189	0.01467	0.01328	0.01192	0.01468	5.9090E-06
Std dev. of posterior mean (across nu parameters)														
Lifespan	P	1.7078E-05			6.6521E-05			4.0415E-05			4.1932E-05			
	C	1.9149E-05			3.3166E-05			1.7078E-05			1.7321E-05			
LRS	P	1.2975E-04			1.3240E-04			6.3157E-05			1.7337E-04			
	C	3.0156E-05			5.3194E-05			8.5926E-05			7.5253E-05			
RR	P	1.1587E-05			8.5829E-06			5.0662E-06			1.2897E-05			
	C	1.7321E-06			3.4034E-06			6.6521E-06			3.5940E-06			

Table 2.5: Table with the angles θ separating the most typically used protein to carbohydrate (P:C) ratios in studies using nutritional geometry in *Drosophila*, as calculated using the Vector of Position Approach. All angles are given in degrees.

P:C ratio	(0:1)	(1:16)	(1:8)	(1:4)	(1:2)	(1:1)	(2:1)	(1:0)
(0:1)	-							
(1:16)	3.548	-						
(1:8)	7.125	3.548	-					
(1:4)	14.036	10.459	6.911	-				
(1:2)	26.565	22.988	19.440	12.528	-			
(1:1)	45.000	41.423	37.875	30.963	18.435	-		
(2:1)	63.439	59.858	56.309	49.398	36.870	18.435	-	
(1:0)	90.000	86.423	82.875	75.963	63.439	45.000	26.565	-

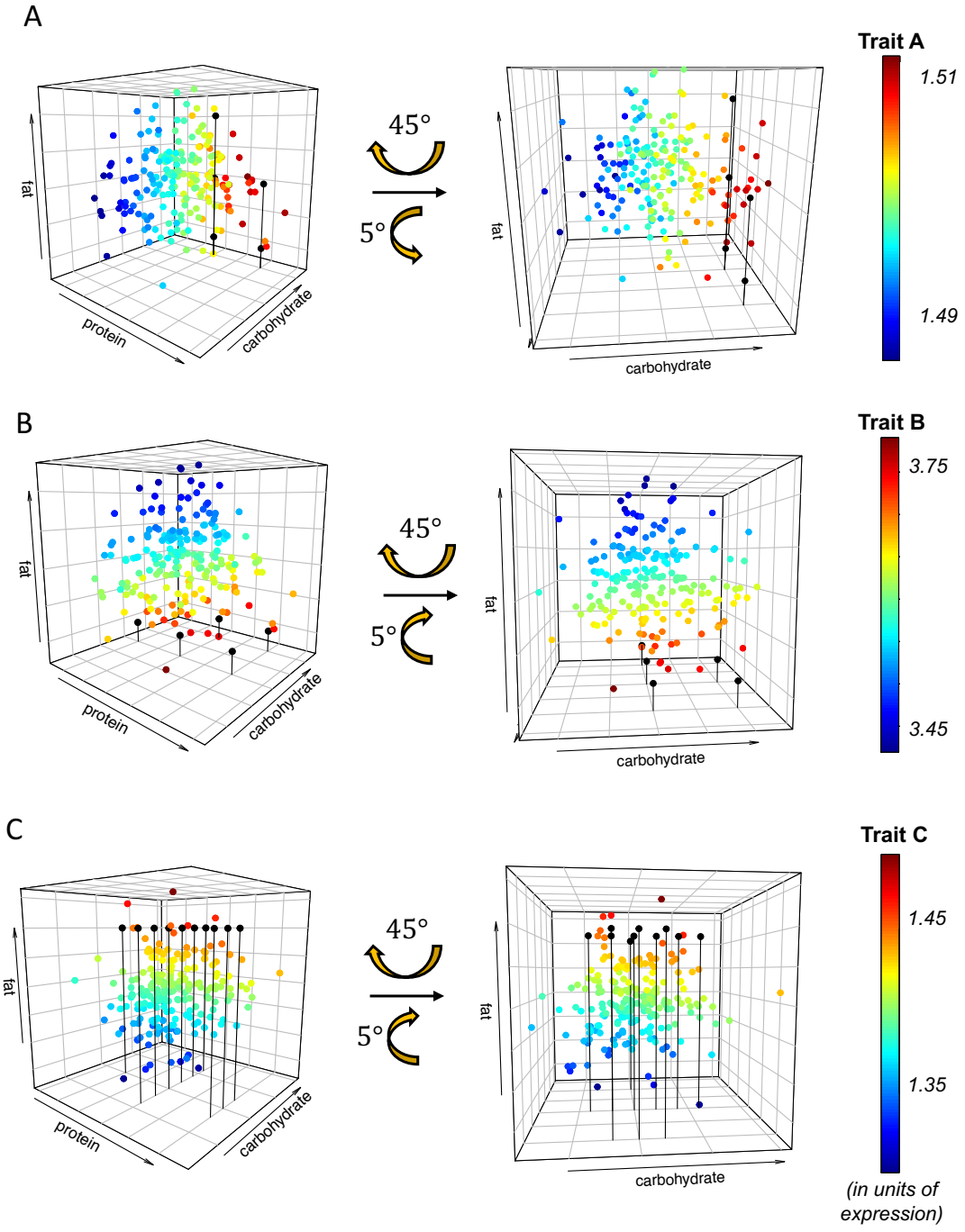


Figure 2.4: Analyses in high-dimension nutrient spaces. (a-c) 3D plots showing the simulated relationship between protein, carbohydrate and fat intake ($x-$ $y-$ and $z-$ axes, respectively) on the expression of traits A (panel A), B (panel B) and C (panel C) (in units of trait expression), represented as the colour variable. Black circles show the estimates of region of the peak in the performance landscape from the SVM machine learning model.

2.6.2 Appendix: Text S1 and Text S2 with examples of R code.

The American Naturalist

Supplementary Information: "Quantifying nutritional trade-offs across multidimensional performance landscapes"

Authors: Juliano Morimoto[1,*], Mathieu Lihoreau[2]

Affiliations:

1 - Department of Biological Sciences, Macquarie University, NSW 2109, Australia

2 - Research Centre on Animal Cognition (CRCA), Center for Integrative Biology (CBI); CNRS, University Paul Sabatier, Toulouse, France

* Correspondence: juliano.morimoto@mq.edu.au

Text S1 – Comparison between the Angular Method and the Vector of Positions Approach in a 2-dimensional nutrient space

Loading the packages

First, let's load the packages required for the data manipulation and analyses. Please use 'install.packages( )' if you don't have the packages installed.

```
library(ggplot2)
library(dplyr)
library(MCMCglmm) # for the Bayesian regression
library(fields) # for the surface plots and the thin-plate spline (Tps) funct
ion
library(tidyr)
```

Data

For 2-dimensional data, we used two published data sets. The first data set is from the study of Lee *et al* 2008 in *Drosophila melanogaster*, published in *PNAS*. The second data set is from a study by Fanson *et al* 2009 in *Bactrocera tryoni*, published in *Ageing Cell* [^1]. [^1]: The authors ob both studies have kindly agreed to share their data.

We will work in two parts. In the first part we will use Lee *et al* data to compare nutritional trade-offs of different traits, using both the Angular method and the Vector of Positions Approach.

Loading Lee *et al* data

First we load and clean data.

```
leedt <- read.csv("LEetetal-DataSet-survival.csv", na.strings = ".", header =
TRUE, stringsAsFactors = TRUE, strip.white = TRUE)
leedt$Ratio <- gsub("[\\(\\)]", "", leedt$Ratio) ## Removing the brackets aro
und the Ratio ###
leedt <-na.omit(leedt) ## removing NA's
```

and prepare the data frame for the Bayesian regressions

```
angle.data1 <- data.frame(P = leedt$proteinday, C = leedt$carboday, lifespan
= leedt$lifespan, lifetimeegg = leedt$lifetimeegg, dailyeggs = leedt$dailyegg
s)
```

Bayesian models (the Angular Method)

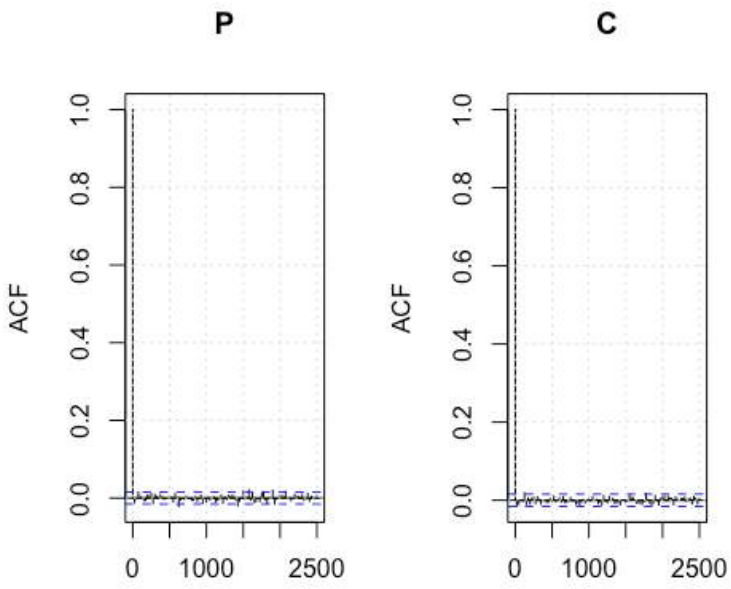
Once with the data formatted, we can start implementing the Angular Method (see main text). Following the approach by Bunning *et al* 2015, we use Bayesian Regressions to obtain slopes of Protein and Carbohydrate intake onto the expression of the life-hisotry trait. Note that this cunk of code can take some time to complete. Let's first set an informative prior

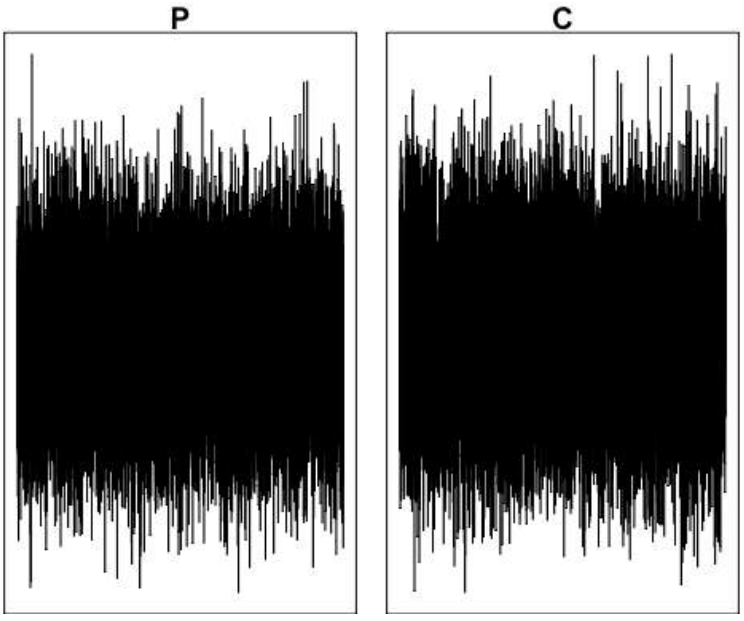
```
prior<-list(R=list(V=1,nu=0.02))
```

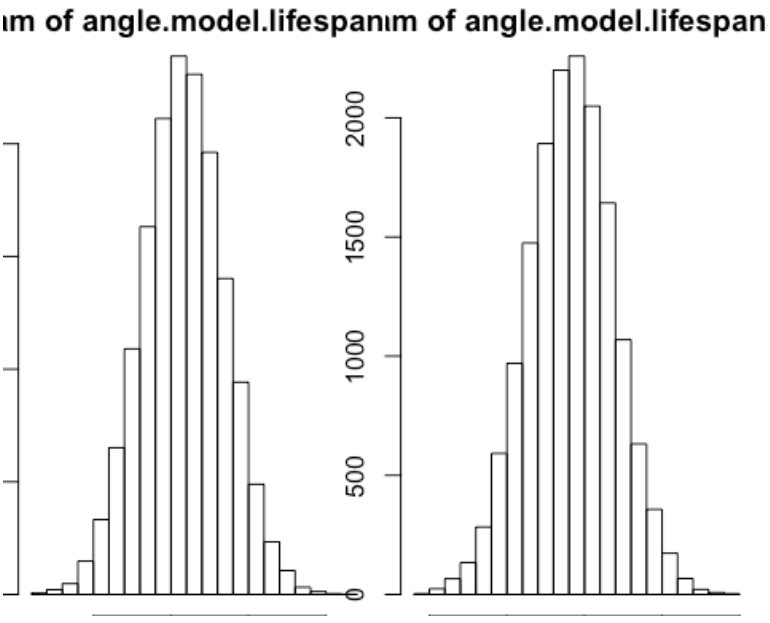
and then the models

```
angle.model.lifespan <-MCMCglmm(lifespan ~ P + C - 1,data=angle.data1, prior
= prior, nitt=400000,burnin=2000,thin=25)

angle.model.lifetimeegg <-MCMCglmm(lifetimeegg ~ P + C - 1,data=angle.data1,
prior = prior, nitt=400000,burnin=2000,thin=25)

angle.model.dailyeggs <-MCMCglmm(dailyeggs ~ P + C -1,data=angle.data1, prior
= prior, nitt=400000,burnin=2000,thin=25)
```

We have tuned the model based on diagnostic plots such as


P
C
ACF
ACF
0.0
0.2
0.4
0.6
0.8
1.0
0
1000
2500


P
C



Calculating the angle θ'

According to Bunning *et al* 2015, let's calculate the angle θ' that separate the vectors of slope for the life history traits.

Let's start with the angle between lifespan (LS) and reproductive rate (RR)

```
angles_LS_RR <-numeric(15200)
for(i in 1:15200){
  b.lifespan <- angle.model.lifespan$Sol[i,1:2]
  b.dailyeggs <- angle.model.dailyeggs$Sol[i,1:2] # creates a vector of beta
estimates for each variable for each row of the posterior distribution (and t
he loop runs through all rows)
  angles_LS_RR[i]<- acos((t(b.dailyeggs) %*% b.lifespan) / ((sqrt(t(b.dailyeg
gs) %*% b.dailyeggs)) * (sqrt(t(b.lifespan) %*% b.lifespan)))) * (180/pi) }
summary(angles_LS_RR)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   44.77   59.83   63.21   63.07   66.41   79.88
```

And the same for lifetime reproductive success (LRS)

```
angles_LRS_RR <-numeric(15200)
for(i in 1:15200){
```

```
  b.lifetimeegg <- angle.model.lifetimeegg$Sol[i,1:2]
  b.dailyeggs <- angle.model.dailyeggs$Sol[i,1:2]
  angles_LRS_RR[i]<- acos((t(b.dailyeggs) %*% b.lifetimeegg) / ((sqrt(t(b.dai
lyeggs) %*% b.dailyeggs)) * (sqrt(t(b.lifetimeegg) %*% b.lifetimeegg)))) * (1
80/pi) }
summary(angles_LRS_RR)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   20.30   43.73   48.02   47.91   52.21   69.17
```

and finally,

```
angles_LS_LRS <-numeric(15200)
for(i in 1:15200){
  b.lifetimeegg <- angle.model.lifetimeegg$Sol[i,1:2]
  b.lifespan <- angle.model.lifespan$Sol[i,1:2]
  angles_LS_LRS[i]<- acos((t(b.lifespan) %*% b.lifetimeegg) / ((sqrt(t(b.life
span) %*% b.lifespan)) * (sqrt(t(b.lifetimeegg) %*% b.lifetimeegg)))) * (180/
pi) }
summary(angles_LS_LRS)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.213  12.060  15.047  15.161  18.126  31.996
```

Vector of Positions Approach

Let's use the **Vector of Positions Approach** to calculate the angles. We first define a function to estimate the optimum region based on the thin-plate spline extrapolation

```
intervalpoint <- function(data, x1, x2, y, level = 0.95, lambda = 0.01){
  dat <- na.omit(data)
  out <-  Tps(cbind(dat[,x1], dat[,x2]), dat[,y], lambda = lambda)
  maxima <- with(out, subset(data.frame(fitted.values, x), fitted.values == m
ax(fitted.values)))

  z_critical <- qnorm(level)
  pop_stdev <- sd(out$residuals)
  sample_size <- length(dat[,x1])
  margin_of_error <- z_critical * (pop_stdev / sqrt(sample_size))

  if(maxima$X1/maxima$X2 >= 1) {
    sample_mean <- mean(maxima$X1/maxima$X2)
    confidence_interval  <- data.frame(mean = sample_mean, lower = sample_mea
n - margin_of_error,
                                       upper = sample_mean + margin_of_error)
    message("Result given as Protein-to-Carbohydrate ratio")
    return(confidence_interval)
  } else if(maxima$X1 == round(0)){
    maxima$X1 <- 1
    sample_mean <- mean(maxima$X2/maxima$X1)
    confidence_interval  <- data.frame(protein = 0, mean = sample_mean, lower
```

```
= sample_mean - margin_of_error,
                                        upper = sample_mean + margin_of_error)
    message("Result given as Carbohydrate-to-protein ratio and estimate of pr
otein intake is approximated to 0")
    return(confidence_interval)
  } else {
    sample_mean <- mean(maxima$X2/maxima$X1)
    confidence_interval  <- data.frame(mean = sample_mean, lower = sample_mea
n - margin_of_error,
                                        upper = sample_mean + margin_of_error)
    message("Result given as Carbohydrate-to-protein ratio")
    return(confidence_interval)

  }
}
```

We then find the point estimates according to the proposed approach, and create the vectors representing the performance landscape

```
point.lifespan <- as.data.frame(intervalpoint(leedt, x1 = "proteinday", x2 =
"carboday", y = "lifespan"))

point.dailyeggs <- as.data.frame(intervalpoint(leedt, x1 = "proteinday", x2 =
"carboday", y = "dailyeggs"))

point.lifetimeegg <- as.data.frame(intervalpoint(leedt, x1 = "proteinday", x2
= "carboday", y = "lifetimeegg"))

vector.lifespan <- c(point.lifespan$protein, point.lifespan$mean)
vector.dailyeggs <- c(1, point.dailyeggs$mean)
vector.LRS <- c(1, point.lifetimeegg$mean)


lower.lifespan <- c(point.lifespan$protein, point.lifespan$upper)
lower.dailyeggs <- c(1, point.dailyeggs$upper)
lower.LRS <- c(1, point.lifetimeegg$upper)


upper.lifespan <-  c(point.lifespan$protein, point.lifespan$lower)
upper.dailyeggs <- c(1, point.dailyeggs$lower)
upper.LRS <- c(1, point.lifetimeegg$lower)
```

Calculating the angle θ

We can then calculate the angle θ and its lower and upper confidence bound for all life-history traits

Let's start with the difference between lifespan and RR

```
position.angle <- acos((t(vector.dailyeggs) %*% vector.lifespan) / ((sqrt(t(v
ector.dailyeggs ) %*% vector.dailyeggs )) * (sqrt(t(vector.lifespan) %*% vect
or.lifespan)))) * (180/pi)
lower.angle <- acos((t(lower.dailyeggs ) %*% lower.lifespan) / ((sqrt(t(lower
.dailyeggs ) %*% lower.dailyeggs )) * (sqrt(t(lower.lifespan) %*% lower.lifes
pan)))) * (180/pi)
upper.angle <- acos((t(upper.dailyeggs ) %*% upper.lifespan) / ((sqrt(t(upper
.dailyeggs ) %*% upper.dailyeggs )) * (sqrt(t(upper.lifespan) %*% upper.lifes
pan)))) * (180/pi)
position.final <- data.frame(theta = position.angle, upper.margin = upper.ang
le, lower.margin = lower.angle)
position.final
```

```
##      theta upper.margin lower.margin
## 1 26.39629     27.38559     25.47133
```

The same for lifespan and LRS

```
position.angle2 <- acos((t(vector.LRS) %*% vector.lifespan) / ((sqrt(t(vector
.LRS ) %*% vector.LRS )) * (sqrt(t(vector.lifespan) %*% vector.lifespan)))) *
(180/pi)

lower.angle2 <- acos((t(lower.LRS) %*% lower.lifespan) / ((sqrt(t(lower.LRS)
%*% lower.LRS)) * (sqrt(t(lower.lifespan) %*% lower.lifespan)))) * (180/pi)

upper.angle2 <- acos((t(upper.LRS) %*% upper.lifespan) / ((sqrt(t(upper.LRS)
%*% upper.LRS)) * (sqrt(t(upper.lifespan) %*% upper.lifespan)))) * (180/pi)

position.final2 <- data.frame(theta = position.angle2, upper.margin = upper.a
ngle2, lower.margin = lower.angle2)
position.final2
```

```
##      theta upper.margin lower.margin
## 1 14.07379     19.88953     10.85754
```

anf finally, for LRS and RR

```
position.angle3 <- acos((t(vector.LRS) %*% vector.dailyeggs) / ((sqrt(t(vecto
r.LRS ) %*% vector.LRS )) * (sqrt(t(vector.dailyeggs) %*% vector.dailyeggs)))
) * (180/pi)

lower.angle3 <- acos((t(lower.LRS) %*% lower.dailyeggs) / ((sqrt(t(lower.LRS)
%*% lower.LRS)) * (sqrt(t(lower.dailyeggs) %*% lower.dailyeggs)))) * (180/pi)

upper.angle3 <- acos((t(upper.LRS) %*% upper.dailyeggs) / ((sqrt(t(upper.LRS)
%*% upper.LRS)) * (sqrt(t(upper.dailyeggs) %*% upper.dailyeggs)))) * (180/pi)

position.final3 <- data.frame(theta = position.angle3, upper.margin = upper.a
ngle3, lower.margin = lower.angle3)
position.final3
```

```
##     theta upper.margin lower.margin
## 1 12.3225     7.496069     14.61379
```

Calculating the Euclidean distance

We can then apply the Law of Cosines (see main text) to calculate the Euclidean distance between the traits. To do that, we extract the point estimate as follows:

```
### Extracting the absolute values for the point estimate of Lifespan###
out.surface_LS <- Tps(cbind(leedt$proteinday, leedt$carboday), leedt$lifespan
, lambda = 0.01)
maxima.LS <- with(out.surface_LS, subset(data.frame(fitted.values, x), fitted
.values == max(fitted.values)))
lifespan.scaled <- c(maxima.LS$X1, maxima.LS$X2)

### Extracting the absolute values for the point estimate of RR###
out.surface_RR <- Tps(cbind(leedt$proteinday, leedt$carboday), leedt$dailyegg
s, lambda = 0.01)
maxima <- with(out.surface_RR , subset(data.frame(fitted.values, x), fitted.v
alues == max(fitted.values)))
dailyeggs.scaled <- c(maxima$X1, maxima$X2)

### Extracting the absolute values for the point estimate of LRS###
out.surface_LRS <- Tps(cbind(leedt$proteinday, leedt$carboday), leedt$lifetim
eegg, lambda = 0.01)
maxima_LRS <- with(out.surface_LRS, subset(data.frame(fitted.values, x), fitt
ed.values == max(fitted.values)))
lifetimeegg.scaled <- c(maxima_LRS$X1, maxima_LRS$X2)
```

We can then use these vectors to calculate the Euclidean distance *d'* (Angular Method) and *d* (Vector of Positions approach). Let's start with the Angular Method

Euclidean distance using the Angular Method

Applying the Law of Cosines, we have

```
# LS- RR
distance.d.tprime1 <- sqrt((t(lifespan.scaled) %*% lifespan.scaled) + (t(dail
yeggs.scaled) %*% dailyeggs.scaled) - (2 * (t(lifespan.scaled) %*% dailyeggs.
scaled) * cos(median(angles_LS_RR)*(pi/180))))
# LS - LRS
distance.d.tprime2 <- sqrt((t(lifespan.scaled) %*% lifespan.scaled) + (t(life
timeegg.scaled) %*% lifetimeegg.scaled) - (2 * (t(lifespan.scaled) %*% lifeti
meegg.scaled) * cos(median(angles_LS_LRS)*(pi/180))))
# LRS-RR
distance.d.tprime3 <- sqrt((t(lifetimeegg.scaled) %*% lifetimeegg.scaled) + (
t(dailyeggs.scaled) %*% dailyeggs.scaled) - (2 * (t(lifetimeegg.scaled) %*% d
ailyeggs.scaled) * cos(median(angles_LRS_RR)*(pi/180))))
```

Euclidean distance using the Vector of Positions approach

Using the Law of Cosines, we have

```
# LS- RR
distance.d.theta1 <- sqrt((t(lifespan.scaled) %*% lifespan.scaled) + (t(daily
eggs.scaled) %*% dailyeggs.scaled) - (2 * (t(lifespan.scaled) %*% dailyeggs.s
caled) * cos((position.final$theta*(pi/180)))))

# LS - LRS
distance.d.theta2 <- sqrt((t(lifespan.scaled) %*% lifespan.scaled) + (t(lifet
imeegg.scaled) %*% lifetimeegg.scaled) - (2 * (t(lifespan.scaled) %*% lifetim
eegg.scaled) * cos((position.final2$theta*(pi/180)))))

# LRS-RR
distance.d.theta3 <- sqrt((t(dailyeggs.scaled) %*% dailyeggs.scaled) + (t(lif
etimeegg.scaled) %*% lifetimeegg.scaled) - (2 * (t(dailyeggs.scaled) %*% life
timeegg.scaled) * cos((position.final3$theta*(pi/180)))))
```

Finally, we can compare the estimates of the two methods

```
# % overstimation of the Angular method
perc.estimation1 <- data.frame(Percentage_difference = paste(round(100 - dist
ance.d.theta1/distance.d.tprime1 * 100, digits = 2), "%"))
perc.estimation2 <- data.frame(Percentage_difference = paste(round(100 - dist
ance.d.theta2/distance.d.tprime2 * 100, digits = 2), "%"))
perc.estimation3 <- data.frame(Percentage_difference = paste(round(100 - dist
ance.d.theta3/distance.d.tprime3 * 100, digits = 2), "%"))

data.frame(perc.estimation1, perc.estimation2, perc.estimation3)
```

```
##   Percentage_difference Percentage_difference.1 Percentage_difference.2
## 1                30.4 %                  0.94 %                  63.6 %
```

Vector Projections

The Vector of Positions approach extends the use of the Angular Method to include vector projections to estimate the strength of the nutritional trade-off in life-history traits. Let's calculate the vector projections for the data mentioned above. The approach will be similar to the Angular method described above, although the calculations and interpretations are different (refer to the main text). To do this, we will use the models fitted above with the Bayesian regression: `angle.model.lifespan`, `angle.model.dailyeggs`, `angle.model.lifetimeegg`.

Calculating the projections of lifespan onto RR

```
scalingfactor_LS_RR <-c(numeric(15200))

for(i in 1:15200){
  vector.lifespan <- angle.model.lifespan$Sol[i,1:2]
  unit.lifespan <- vector.lifespan/ (t(vector.lifespan) %*% vector.lifespan)
```

```
  vector.dailyeggs <- angle.model.dailyeggs$Sol[i,1:2]
  unit.dailyeggs <- vector.dailyeggs/ (t(vector.dailyeggs) %*% vector.dailyeg
gs)
  scalingfactor_LS_RR[i] <- (t(unit.dailyeggs) %*% unit.lifespan / (t(unit.da
ilyeggs) %*% unit.dailyeggs))}
summary(scalingfactor_LS_RR)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.02210 0.04745 0.05327 0.05333 0.05910 0.08862
```

Calculating the projections of lifespan onto LRS

```
scalingfactor_LS_LRS <-c(numeric(15200))

for(i in 1:15200){
  vector.lifespan <- angle.model.lifespan$Sol[i,1:2]
  unit.lifespan <- vector.lifespan/ (t(vector.lifespan) %*% vector.lifespan)
  vector.lifetimeegg <- angle.model.lifetimeegg$Sol[i,1:2]
  unit.lifetimeegg <- vector.lifetimeegg/ (t(vector.lifetimeegg) %*% vector.l
ifetimeegg)
  scalingfactor_LS_LRS[i] <- (t(unit.lifetimeegg) %*% unit.lifespan / (t(unit
.lifetimeegg) %*% unit.lifetimeegg))  }
summary(scalingfactor_LS_LRS)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.279   1.632   1.717   1.717   1.802   2.274
```

Calculating the projections of LRS onto RR

```
scalingfactor_RR_LRS <-c(numeric(15200))

for(i in 1:15200){
  vector.dailyeggs <- angle.model.dailyeggs$Sol[i,1:2]
  unit.dailyeggs <- vector.dailyeggs/ (t(vector.dailyeggs) %*% vector.dailyeg
gs)
  vector.lifetimeegg <- angle.model.lifetimeegg$Sol[i,1:2]
  unit.lifetimeegg <- vector.lifetimeegg/ (t(vector.lifetimeegg) %*% vector.l
ifetimeegg)
  scalingfactor_RR_LRS[i] <- (t(unit.lifetimeegg) %*% unit.dailyeggs / (t(uni
t.lifetimeegg) %*% unit.lifetimeegg)) }
summary(scalingfactor_RR_LRS)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   4.977   9.177  10.041  10.034  10.898  14.277
```

Comparative studies

As mentioned in the main text that accompany this RMarkDown, the Vector of Positions allow for the comparison of the expression of life-history traits between species. In this section, we will compare the nutritional differences in reproductive rates of *Drosophila*

melanogaster and *Bactrocera tryoni* (aka 'Queensland fruit fly) from two published studies (see Data section above).

The section calculates the differences in the reproductive rate of these two species in the performance landscape using machine learning models. Let's first load the required packages for this analyses (in addition to those loaded in the previous section)

```
library(caret)
```

Then let's load and tide up Fanson *et al* 2009 data.

Loading and cleaning Fanson *et al* data

```
fansondt <- read.csv("Fansonetal2008AgeingCell-NoChoiceData.csv", strip.white
= TRUE, stringsAsFactors = TRUE, header = TRUE)


#adjusting the Ratio levels

fansondt$Ratio <- ifelse(fansondt$Ratio == "0:01", "0:1",
                         ifelse(fansondt$Ratio == "1:00", "1:0",
                                ifelse(fansondt$Ratio == "1:01", "1:1",
                                       ifelse(fansondt$Ratio == "1:02", "1:2"
,
                                              ifelse(fansondt$Ratio == "1:04"
, "1:4",
                                                     ifelse(fansondt$Ratio ==
"1:08", "1:8", "1:16"))))))

### adjusting the Food columns
fansondt <- separate(fansondt, Food, into = c("Food", "unit"), sep = " ", rem
ove = TRUE)
fansondt$Food <- as.integer(fansondt$Food)


### Calculating the P and C intake of each observation
fansondt$Ceaten <- ifelse(fansondt$Ratio == "0:1", fansondt$totaleaten*fanson
dt$Food,
                          ifelse(fansondt$Ratio == "1:0", 0,
                                 ifelse(fansondt$Ratio == "1:1", (fansondt$to
taleaten/2)*fansondt$Food,
                                        ifelse(fansondt$Ratio == "1:2", (fans
ondt$totaleaten/3)*2*fansondt$Food,
                                               ifelse(fansondt$Ratio == "1:4"
, (fansondt$totaleaten/5)*4*fansondt$Food,
                                                      ifelse(fansondt$Ratio =
= "1:8",(fansondt$totaleaten/9)*8*fansondt$Food, (fansondt$totaleaten/17)*16*
fansondt$Food))))))
```

```
fansondt$Peaten <- (fansondt$totaleaten*fansondt$Food) - fansondt$Ceaten

### Getting in mg of macronutrients
fansondt$Ceaten <- fansondt$Ceaten/1000
fansondt$Peaten <- fansondt$Peaten/1000

### Cleaning the data
fansondt <- subset(fansondt, Ceaten >= 0 & Peaten >= 0)
fansondt$dailyeggs <- fansondt$eggs/fansondt$lifespan
```

Standardising the data

Because we are working with two species, the data must be standardized (refer to the main text)

```
## the standardizing function ##
stdfunc <- function(x) x/mean(x, na.rm = TRUE)

## standardizing the variable
leedt[c(5:6,9)] <- lapply(leedt[c(5:6,9)], stdfunc)
fansondt[c(14:16)] <- lapply(fansondt[c(14:16)], stdfunc)
```

Machine Learning model

Once the data has been standardized, we can now use the machine learning model to find the optimum region in the performance landscape. We start with the Lee *et al* data. For the Machine learning, we first split our data set into the training and test sets. These two data sets are, in a nutshell, the data that your model will use to find patterns (training set) and evaluate its performance against data that the model has never seen before (test set). The proportion of data allocated to the training and test sets vary according to the aim of the model. In this case, we would like accurate predictions for the nutritional peak. As such, the model needs to be trained with as much data as possible. We will use 95% of the original data as the training set, and the remaining 5% as the test test. We can achieve this as following:

```
validation_index_lee <- createDataPartition(leedt$treatment, p=0.95, list=FAL
SE) ## randomly selecting observations

validationlee <- leedt[-validation_index_lee,]

training_set_lee <- leedt[validation_index_lee,]
```

Before fitting the model, we set up the parameters

```
# Run algorithms using 10-fold cross validation
control <- trainControl(method="cv", number=10)
metric <- "RMSE" # RMSE since we are using regression models (not classifiers
)
```

Now we are ready to fit the machine learning model in our data. We will use a Radial-Kernel Support Vector Machine (RadialSVM) as, from previous trials, is the best approach for predicting the optima in the performance landscapes (in fact, other regression models such as Random Forest does a remarkably poor job!). What we will do here is to first fit the model, then validate it, and then use in a new grid. The predictions from these grids will be subsetted based on the maximum region (which is the region of the optimum in the performance landscape).

```
# SVM
set.seed(7)
fit.svm_lee <- train(dailyeggs~proteinday + carboday, data=training_set_lee,
method="svmRadial", metric=metric, trControl=control) # the machine learning
model

predictionslee <- data.frame(validationlee, dailyeggs_pred = predict(fit.svm_
lee, validationlee)) # validating

newdatalee <- expand.grid(proteinday = seq(0, max(leedt$proteinday), 0.01), c
arboday = seq(0, max(leedt$carboday), 0.01)) # creating the grid

estimatelee <- data.frame(newdatalee, dailyeggs_pred = predict(fit.svm_lee, n
ewdatalee)) # predicting in the new grid

regionoptima_lee <- subset(estimatelee, dailyeggs_pred >= quantile(dailyeggs_
pred,probs=0.95) & carboday >= 2.5 & proteinday >= 2.5) # subsetting the pred
icted optima
```

We now do the same for the Fanson *et al* data.

```
#subsetting the data frame
validation_index_fanson <- createDataPartition(fansondt$DietID, p=0.95, list=
FALSE)

validationfanson <- fansondt[-validation_index_fanson,]

training_set_fanson <- fansondt[validation_index_fanson,]

# Machine Learning Model
set.seed(7)
fit.svm_fanson <- train(dailyeggs~Peaten + Ceaten , data=training_set_fanson,
method="svmRadial", metric=metric, trControl=control)

predictionsfanson <- data.frame(validationfanson, dailyeggs_pred = predict(fi
t.svm_fanson, validationfanson))

newdatafanson <- expand.grid(Peaten = seq(0, max(fansondt$Peaten), 0.01), Cea
ten = seq(0, max(fansondt$Ceaten), 0.01))

estimatefanson <- data.frame(newdatafanson, dailyeggs_pred = predict(fit.svm_
```

```
fanson, newdatafanson))

regionoptima_fanson <- subset(estimatefanson, dailyeggs_pred >= quantile(dail
yeggs_pred,probs=0.95) & Ceaten >= 1.5 & Ceaten < 4 & Peaten >= 4.5)
```

Comparative studies of the nutritional landscape

Having defined the optima regions (i.e., `regionoptima_fanson` and `regionoptima_lee`), we can now estimate the angle and the confidence bound. Let's first define the function to calculate the angle

```
## calculating the angle between the 3-dimensional vectors ##
angle <- function(x,y){
  dot.prod <- x%*%y
  norm.x <- norm(x,type="2")
  norm.y <- norm(y,type="2")
  theta <- acos(dot.prod / (norm.x * norm.y))
  as.numeric(theta)*57.2958
}
```

and then

```
### Extrating the mean angle and the upper and lower bound of its estimates
regionoptima_fanson_mean <- regionoptima_fanson %>% summarise(meanP = mean(Pe
aten), meanC = mean(Ceaten), meanpred= mean(dailyeggs_pred))
regionoptima_lee_mean <- regionoptima_lee %>% summarise(meanP = mean(proteind
ay), meanC = mean(carboday), meanpred= mean(dailyeggs_pred))



regionoptima_fanson_upper <- regionoptima_fanson %>% summarise(lowerP = min(P
eaten), lowerC = min(Ceaten), lowerpred= min(dailyeggs_pred))
regionoptima_lee_upper <- regionoptima_lee %>% summarise(lowerP = min(protein
day), lowerC = min(carboday), lowerpred= min(dailyeggs_pred))

regionoptima_fanson_lower <- regionoptima_fanson %>% summarise(upperP = max(P
eaten), upperC = max(Ceaten), upperpred= max(dailyeggs_pred))
regionoptima_lee_lower <- regionoptima_lee %>% summarise(upperP = max(protein
day), upperC = max(carboday), upperpred= max(dailyeggs_pred))



### calculating the angle theta between vectors ###
anglesspecies_mean <- angle(as.matrix(regionoptima_fanson_mean), as.matrix(t(
regionoptima_lee_mean)))

anglesspecies_upper <- angle(as.matrix(regionoptima_fanson_upper), as.matrix(
t(regionoptima_lee_upper)))

anglesspecies_lower <- angle(as.matrix(regionoptima_fanson_lower), as.matrix(
t(regionoptima_lee_lower)))
```

```
### printing the result ###
angle_species <- data.frame(lower_bound = anglesspecies_lower, mean_theta = a
nglesspecies_mean, upper_bound = anglesspecies_upper)
angle_species
```

```
##   lower_bound mean_theta upper_bound
## 1    13.48714   17.58226    24.58529
```

Calculating the direction angles α, β and γ

With the Vector of Positions Approach, we also propose that we can use the direction angles α, β and γ to obtain a detailed description of the nutritional trade-offs between species.

```
##### Calculating the angles of each component ####
regionoptima_lee_mean
```

```
##     meanP    meanC meanpred
## 1 3.77762 3.118889 2.562604
```

```
regionoptima_fanson_mean
```

```
##    meanP    meanC meanpred
## 1 5.8917 2.620724 2.025965
```

```
#### the ANGLE ALPHa (against the x-axis (protein axis))
dirangles_alpha <- rbind(data.frame(mean_alpha = atan2(sqrt(regionoptima_lee_
mean$meanC^2 +
    regionoptima_lee_mean$meanpred^2), regionoptima_lee_mean$meanP) * (180/pi
),
    lower_alpha = atan2(sqrt(regionoptima_lee_lower$upperC^2 + regionoptima_l
ee_lower$upperpred^2),
        regionoptima_lee_lower$upperP) * (180/pi), upper_alpha = atan2(sqrt(r
egionoptima_lee_upper$lowerC^2 +
        regionoptima_lee_upper$lowerpred^2), regionoptima_lee_upper$lowerP) *
        (180/pi)), data.frame(mean_alpha = atan2(sqrt(regionoptima_fanson_mea
n$meanC^2 +
    regionoptima_fanson_mean$meanpred^2), regionoptima_fanson_mean$meanP) *
    (180/pi), upper_alpha = atan2(sqrt(regionoptima_fanson_lower$upperC^2 +
    regionoptima_fanson_lower$upperpred^2), regionoptima_fanson_lower$upperP)
*
    (180/pi), lower_alpha = atan2(sqrt(regionoptima_fanson_upper$lowerC^2 +
    regionoptima_fanson_upper$lowerpred^2), regionoptima_fanson_upper$lowerP)
*
    (180/pi)))

row.names(dirangles_alpha) <- c("Drosophila", "Bactrocera")
```

```
### The angle BETA (against the y-axis (carbohydrate axis))
dirangles_beta <- rbind(data.frame(mean_beta = atan2(sqrt(regionoptima_lee_me
an$meanP^2 +
    regionoptima_lee_mean$meanpred^2), regionoptima_lee_mean$meanC) * (180/pi
),
    lower_beta = atan2(sqrt(regionoptima_lee_lower$upperP^2 + regionoptima_le
e_lower$upperpred^2),
        regionoptima_lee_lower$upperC) * (180/pi), upper_beta = atan2(sqrt(re
gionoptima_lee_upper$lowerP^2 +
        regionoptima_lee_upper$lowerpred^2), regionoptima_lee_upper$lowerC) *
        (180/pi)), data.frame(mean_beta = atan2(sqrt(regionoptima_fanson_mean
$meanP^2 +
    regionoptima_fanson_mean$meanpred^2), regionoptima_fanson_mean$meanC) *
    (180/pi), upper_beta = atan2(sqrt(regionoptima_fanson_lower$upperP^2 + re
gionoptima_fanson_lower$upperpred^2),
    regionoptima_fanson_lower$upperC) * (180/pi), lower_beta = atan2(sqrt(reg
ionoptima_fanson_upper$lowerP^2 +
    regionoptima_fanson_upper$lowerpred^2), regionoptima_fanson_upper$lowerC)
*
    (180/pi)))

row.names(dirangles_beta) <- c("Drosophila", "Bactrocera")


### The angle GAMMA (against the z-axis (trait axis))
dirangles_gamma <- rbind(data.frame(mean_gamma = atan2(sqrt(regionoptima_lee_
mean$meanP^2 +
    regionoptima_lee_mean$meanC^2), regionoptima_lee_mean$meanpred) * (180/pi
),
    lower_gamma = atan2(sqrt(regionoptima_lee_lower$upperP^2 + regionoptima_l
ee_lower$upperC^2),
        regionoptima_lee_lower$upperpred) * (180/pi), upper_gamma = atan2(sqr
t(regionoptima_lee_upper$lowerP^2 +
        regionoptima_lee_upper$lowerC^2), regionoptima_lee_upper$lowerpred) *
        (180/pi)), data.frame(mean_gamma = atan2(sqrt(regionoptima_fanson_mea
n$meanP^2 +
    regionoptima_fanson_mean$meanC^2), regionoptima_fanson_mean$meanpred) *
    (180/pi), upper_gamma = atan2(sqrt(regionoptima_fanson_lower$upperP^2 +
    regionoptima_fanson_lower$upperC^2), regionoptima_fanson_lower$upperpred)
*
    (180/pi), lower_gamma = atan2(sqrt(regionoptima_fanson_upper$lowerP^2 +
    regionoptima_fanson_upper$lowerC^2), regionoptima_fanson_upper$lowerpred)
*
    (180/pi)))

row.names(dirangles_gamma) <- c("Drosophila", "Bactrocera")
```

```r
dirangles <- list(alpha = dirangles_alpha, beta = dirangles_beta, gamma = dir
angles_gamma)
dirangles
```

```
## $alpha
##            mean_alpha lower_alpha upper_alpha
## Drosophila   46.89844    45.90927    47.99309
## Bactrocera   29.34615    25.38286    38.79428
##
## $beta
##            mean_beta lower_beta upper_beta
## Drosophila  55.65722   56.03449   55.26419
## Bactrocera  67.18635   66.76060   69.50664
##
## $gamma
##            mean_gamma lower_gamma upper_gamma
## Drosophila   62.38538    63.16885    61.51371
## Bactrocera   72.55797    80.35599    58.69577
```

The American Naturalist

Supplementary Information: "Quantifying nutritional trade-offs across multidimensional performance landscapes"

Authors: Juliano Morimoto[1,*], Mathieu Lihoreau[2]

Affiliations:

1 - Department of Biological Sciences, Macquarie University, NSW 2109, Australia

2 - Research Centre on Animal Cognition (CRCA), Center for Integrative Biology (CBI); CNRS, University Paul Sabatier, Toulouse, France

* Correspondence: juliano.morimoto@mq.edu.au

Text S2 – High-dimensional plots and data analyses

Loading the packages

```
library(ggplot2)
library(plot3D)
library(fields) # for the surface plots and the thin-plate spline (Tps)
function
library(caret)
```

Simulating data and modelling

We simulate data for the intake of Protein, Carbohydrate, and Fat. First we simulate traits that are close in the expression of the hypothetical life-hisotry trait. Let's first focus on simulating data where traits are in similar locationg in the nutrient space.

```
set.seed(121212)
testdt <- data.frame(protein = rnorm(200, 50, 0.1), carbohydrate = rnorm(200,
30, 0.1), fat = rnorm(200, 5, 0.1))

testdt$trait1 <- with(testdt, ((carbohydrate)*protein) + rnorm(100, 0,0.1))
testdt$trait2 <- with(testdt,  (12*(carbohydrate)*protein)/fat  + rnorm(100,
0, 0.1))
testdt$treatment <- as.factor(c("1:16", "1:8", "1:4", "1:2"))
```

Fitting the model

```
newtestdt <- expand.grid(protein = seq(0, max(testdt$protein), 0.2),
carbohydrate = seq(0, max(testdt$carbohydrate), 0.2), fat = seq(0,
max(testdt$fat), 0.2))
```

Modelling

First we will focus on calculating the difference in expression of traits 1 and 2.

Model for trait 1

```
validation_index_lee <- createDataPartition(testdt$treatment, p=0.95,
list=FALSE)
validation <- testdt[-validation_index_lee,]

training_set <- testdt[validation_index_lee,]


# Run algorithms using 10-fold cross validation
control <- trainControl(method="cv", number=10, search = "grid")
metric <- "RMSE"

# SVM
set.seed(8)
fit.svm_trait1 <- train(trait1~protein + carbohydrate + fat,
```

```
data=training_set, method="svmRadial",
                        preProcess = c("center","scale"), metric=metric,
trControl=control)
```

Model for trait 2

```
fit.svm_trait2 <- train(trait2~protein + carbohydrate + fat,
data=training_set, method="svmRadial",
                        preProcess = c("center","scale"), metric=metric,
trControl=control)
```

Validation of the models

```
validation_test <- data.frame(validation, value_pred1 =
predict(fit.svm_trait1, validation), value_pred2 = predict(fit.svm_trait2,
validation))
prediction_test <- data.frame(newtestdt, value_pred1 =
predict(fit.svm_trait1, newtestdt), value_pred2 = predict(fit.svm_trait2,
newtestdt))
```

Obtaining theoptimum

```
regionoptima_trait1 <- subset(prediction_test, value_pred1 >=
quantile(value_pred1,probs=0.999996)) ## selecting the optimum region. Tune
the "probs" argument for more/less tight predictions
regionoptima_trait2 <- subset(prediction_test, value_pred2 >=
quantile(value_pred2,probs=0.999996))
```

Calculating the angle between the optimum vectors

```
angle <- if(dim(regionoptima_trait1)[1] >= dim(regionoptima_trait2)[1]){
  angle <- numeric(dim(regionoptima_trait2)[1])
} else {angle <- numeric(dim(regionoptima_trait1)[1])}

for(i in 1:length(angle)){
  btrait1 <- as.matrix(regionoptima_trait1[i,c(1:3)])
  btrait2 <- as.matrix(regionoptima_trait2[i,c(1:3)])
  norm1 <- norm(btrait1, type = "2")
  norm2 <- norm(btrait2, type = "2")
  value <- as.numeric(btrait1 %*% t(btrait2)/ (norm1 * norm2))
  angle[i] <- acos(value) * (180/pi)
}

summary(angle)
```

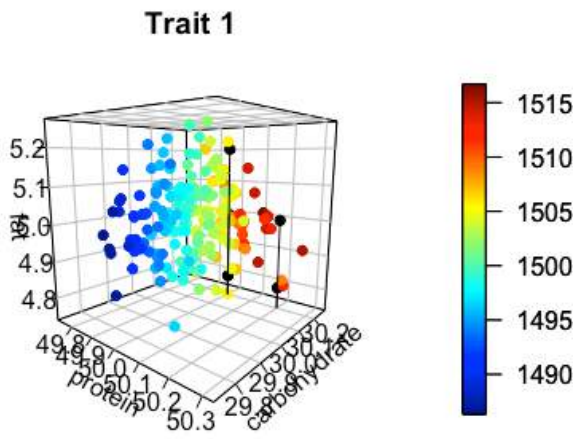
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.07015 0.07038 0.20679 0.21951 0.33464 0.41557
```

Plotting figures

Trait 1

```
#### For trait 1 ####
x <- testdt$protein
```

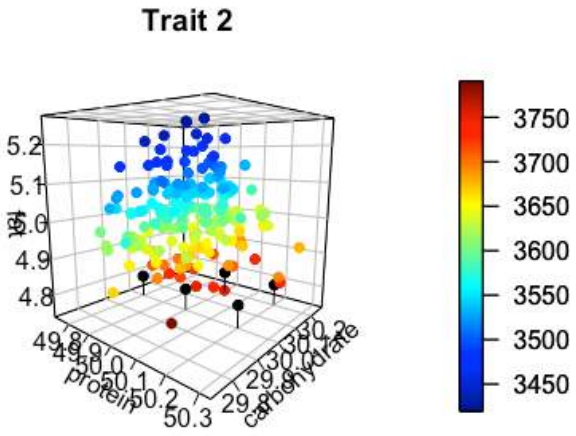
```
y <- testdt$carbohydrate
z <- testdt$fat
w <- testdt$trait1

scatter3D(x = testdt[,1], y = testdt[,2], z = testdt[, 3], colvar =
testdt[,4], pch = 20, cex = 1.2,
          theta = 40, phi = 10,ticktype = "detailed",
          xlab = "protein", ylab = "carbohydrate", zlab = "fat", main =
"Trait 1", bty = "b2")
points3D( x = regionoptima_trait1[,1], y = regionoptima_trait1[,2], z =
regionoptima_trait1[,3], col = "black", add = TRUE,
          pch = 20, cex = 1.3, type = "h") ## adding points predicted by the
model (black)
```



Trait 2

```
#### For trait 1 ####
x <- testdt$protein
y <- testdt$carbohydrate
z <- testdt$fat
w <- testdt$trait2

scatter3D(x = testdt[,1], y = testdt[,2], z = testdt[, 3], colvar =
testdt[,5], pch = 20, cex = 1.2,
```

```
          theta = 40, phi = 10,ticktype = "detailed",
          xlab = "protein", ylab = "carbohydrate", zlab = "fat", main =
"Trait 2", bty = "b2")

points3D( x = regionoptima_trait2[,1], y = regionoptima_trait2[,2], z =
regionoptima_trait2[,3], col = "black", add = TRUE,
          pch = 20, cex = 1.3, type = "h") ## adding points predicted by the
model (black)
```



Modeling (distant traits)

Now we can focus on calculating the differences between the expression of traits 1 and 3 which lie in distant locations in nutrient space. As before, let's simulate traits 1 and 3. Trait 1 is simulated with the same parameters as trait 1 before (I will repeat it here to facilitate comprehension).

```
set.seed(121212)
testdt1 <- data.frame(protein1 = rnorm(200, 50,0.1), carbohydrate1 =
rnorm(200, 30, 0.1), fat1 = rnorm(200, 5, 0.1))
testdt1$trait1 <- with(testdt1, (carbohydrate1*protein1) + rnorm(100, 0,0.1))

testdt2 <- data.frame(protein2 = rnorm(200, 50,0.1), carbohydrate2 =
rnorm(200, 70, 0.3), fat2 = rnorm(200, 5, 0.1))
```

```
testdt2$trait3 <- with(testdt2,  ((carbohydrate2)*8*protein2)*fat2  +
rnorm(100, 0, 0.1))

testdt1$treatment <- as.factor(c("1:16", "1:8", "1:4", "1:2"))
testdt2$treatment <- as.factor(c("1:16", "1:8", "1:4", "1:2"))
```

We can then fit the model

```
newtestdt1 <- expand.grid(protein1 = seq(0, max(testdt1$protein1), 0.2),
carbohydrate1 = seq(0, max(testdt1$carbohydrate1), 0.2), fat1 = seq(0,
max(testdt1$fat1), 0.2))
head(newtestdt1)
```

```
##   protein1 carbohydrate1 fat1
## 1      0.0             0    0
## 2      0.2             0    0
## 3      0.4             0    0
## 4      0.6             0    0
## 5      0.8             0    0
## 6      1.0             0    0
```

```
newtestdt2 <- expand.grid(protein2 = seq(0, max(testdt2$protein2), 0.2),
carbohydrate2 = seq(0, max(testdt2$carbohydrate2), 0.2), fat2 = seq(0,
max(testdt2$fat2), 0.2))
head(newtestdt2)
```

```
##   protein2 carbohydrate2 fat2
## 1      0.0             0    0
## 2      0.2             0    0
## 3      0.4             0    0
## 4      0.6             0    0
## 5      0.8             0    0
## 6      1.0             0    0
```

```
## Models ###

validation_index1 <- createDataPartition(testdt1$treatment, p=0.95,
list=FALSE)
validation_index2 <- createDataPartition(testdt2$treatment, p=0.95,
list=FALSE)
# select 20% of the data for validation
validation1 <- testdt1[-validation_index1,]
validation2 <- testdt2[-validation_index2,]
# use the remaining 80% of data to training and testing the models
training_set1 <- testdt1[validation_index1,]
training_set2 <- testdt2[validation_index2,]


# Run algorithms using 10-fold cross validation
control <- trainControl(method="cv", number=10, search = "grid")
metric <- "RMSE"
```

```
# SVM
set.seed(8)
fit.svm_trait1 <- train(trait1~protein1 + carbohydrate1 + fat1,
data=training_set1, method="svmRadial",
                        preProcess = c("center","scale"), metric=metric,
trControl=control)

fit.svm_trait3 <- train(trait3~protein2 + carbohydrate2 + fat2,
data=training_set2, method="svmRadial",
                        preProcess = c("center","scale"), metric=metric,
trControl=control)
```

We can then estimate the position of the peak for traits 1 and 3

```
prediction_test1 <- data.frame(newtestdt1, value_pred1 =
predict(fit.svm_trait1, newtestdt1))
prediction_test2 <- data.frame(newtestdt2, value_pred2 =
predict(fit.svm_trait3, newtestdt2))

regionoptima_trait1 <- subset(prediction_test1, value_pred1 >=
quantile(value_pred1,probs=0.999996))
regionoptima_trait3 <- subset(prediction_test2, value_pred2 >=
quantile(value_pred2,probs=0.999996))
```

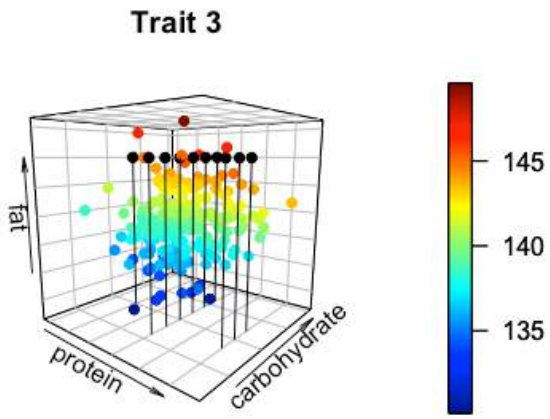
And finally calculate the angle theta

```
angle <- if(dim(regionoptima_trait1)[1] >= dim(regionoptima_trait3)[1]){
  angle <- numeric(dim(regionoptima_trait3)[1])
} else {angle <- numeric(dim(regionoptima_trait1)[1])}

for(i in 1:length(angle)){
  btrait1 <- as.matrix(regionoptima_trait1[i,c(1:3)])
  btrait3 <- as.matrix(regionoptima_trait3[i,c(1:3)])
  norm1 <- norm(btrait1, type = "2")
  norm3 <- norm(btrait3, type = "2")
  value <- as.numeric(btrait1 %*% t(btrait3)/ (norm1 * norm3))
  angle[i] <- acos(value) * (180/pi)
}

summary(angle)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   23.13   23.15   23.32   23.29   23.33   23.52
```

We finish by plotting trait 3, as we did for traits 1 and 2

```
x <- testdt2$protein2
y <- testdt2$carbohydrate2
z <- testdt2$fat2
w <- testdt2$trait3
```

```
scatter3D(x = testdt2[,1], y = testdt2[,2], z = testdt2[, 3], colvar =
testdt2[,4]/1000, pch = 20, cex = 1.2,
          theta = 40, phi = 10,
          xlab = "protein", ylab = "carbohydrate", zlab = "fat", main =
"Trait 3", bty = "b2")

points3D( x = regionoptima_trait3[,1], y = regionoptima_trait3[,2], z =
regionoptima_trait3[,3], col = "black", add = TRUE,
          pch = 20, cex = 1.3, type = "h")
```



Chapter 3

Nutrigonometry I: using right-angle triangles to quantify nutritional trade-offs in performance landscapes

Publication reference

Morimoto, J., Conceição, P., Mirth, C. and Lihoreau, M., (2023). Nutrigonometry I: using right-angle triangles to quantify nutritional trade-offs in performance landscapes. *The American Naturalist*, 201(5), pp.725-740.

3.1 Abstract

Animals regulate their food intake to maximise the expression of fitness traits, but are forced to trade-off optimal expression of some fitness traits due to differences in nutrient requirements of each trait (*'nutritional trade-offs'*). Nutritional trade-offs have been experimentally uncovered using the Geometric Framework for Nutrition (GF). However, current analytical methods to measure such responses rely on either visual inspection or complex models of vector calculations applied to multidimensional performance landscapes, making these approaches subjective, or conceptually difficult, computationally expensive, and in some cases inaccurate. Here, we present a simple trigonometric model to measure nutritional trade-offs in multidimensional landscapes (Nutrigonometry), which relies on the trigonometric relationships of right-angle triangles and thus, is both conceptually and computationally easier to understand and use than previous quantitative approaches. We apply Nutrigonometry to a landmark GF dataset for the comparison of several standard statistical models to assess model performance in finding regions in the performance landscapes. This revealed that polynomial (Bayesian) regressions can be used for precise and accurate predictions of peaks and valleys in performance landscapes, irrespective of the underlying structure of the data (i.e., individual food intakes vs fixed diet ratios). We then identified the known nutritional trade-off between lifespan and reproductive rate both in terms of nutrient balance and concentration for validation of the model. This shows Nutrigonometry enables a fast, reliable, and reproducible quantification of nutritional trade-offs in multidimensional performance landscapes, thereby broadening the potential for future developments in comparative research on the evolution of animal nutrition.

Keywords: Fitness landscapes, health, physiology, costs.

3.2 Introduction

Animals often require different nutrient blends to maximize concurrent life-history traits, creating the potential for a conflict for optimum nutrition (Simpson & Raubenheimer 2012, Raubenheimer & Simpson 2020). When the optimum nutrition for two traits cannot be achieved simultaneously, animals must make a compromise in their feeding decisions to support the optimal expression of one trait over another ('nutritional trade-off') (Lee et al. 2008, Maklakov et al. 2008, Fanson et al. 2009). Previous research has identified nutritional trade-offs between lifespan and reproduction or between immunity and reproduction across many different taxa including *Drosophila melanogaster* (Ponton et al. 2020), tephritid fruit flies (Fanson, Yap & Taylor 2012, Fanson & Taylor 2012), crickets (Harrison et al. 2014, Treidel et al. 2021, Guo et al. 2022), and mice (Solon-Biet et al. 2014) [see also reviews by (Ponton et al. 2011, Schwenke et al. 2016)]. Even traits related to different aspects of the same life-history can vary in nutritional requirements during the lifetime of an animal, as seen for instance in pre- and post-mating traits related to reproduction of many insect species such as sperm number and viability (Bunning et al. 2015), fertilization success across sperm competitive contexts Morimoto & Wigby (2016), cuticular hydrocarbons, courtship song and sperm viability Ng et al. (2019), as well as size and numbers of eupyrene and apyrene sperms Gage & Cook (1994). Thus, nutritional trade-offs are likely ubiquitous and impose significant constraints on the feeding choices of individuals.

Measuring nutritional trade-offs is critical to understand the evolution of animal feeding behaviour. However, it is highly challenging because of the interactive effects of nutrient ratios and concentrations on the expression of life-histories Stearns (1989), Roff (2002), Hunt et al. (2004), Simpson & Raubenheimer (2012). In the last decades, however, a method known as the Geometric Framework for Nutrition (GF) has emerged as a powerful unifying framework capable of disentangling the multidimensional effects of nutrients (both ratios and concentrations) on life-history traits and fitness Raubenheimer & Simpson (1993). The GF has been applied to a diverse range of nutritional studies across species such as flies Lee et al. (2008), Fanson, Yap & Taylor (2012), Ponton et al. (2015), Barragan-Fonseca et al. (2018), crickets Maklakov et al. (2008), Ng et al. (2018), Rapkin et al. (2018), cockroaches Bunning et al. (2015), domestic cats and dogs Hewson-Hughes et al. (2011, 2012), and mice Solon-Biet et al. (2014), Morimoto et al. (2019), being paramount for advancing our understanding of complex physiological and behavioural processes across ecological environments and even human health Simpson et al. (2017). With the growing applications of GF in the study of animal and human nutrition, the development of simple, intuitive, and

accurate quantitative methods for identifying optimal diets in performance landscapes and quantifying nutritional trade-offs has become a key issue for research in comparative nutrition. Simple methodologies that are friendly to biologists and ecologists will facilitate new insights into the complex nutritional decision-making that animals have to undergo in order to navigate physiological and behavioural constraints of different life-history traits Morimoto & Lihoreau (2020).

Recent approaches to analyse GF data have been complex to navigate, and many studies continue to use a combination of approaches (including visual inspection) or potentially inaccurate methods to quantify the strength of nutritional trade-offs in GF landscapes (Polak et al. 2017, Ng et al. 2018, 2019, Rapkin et al. 2018, Kutz et al. 2019, Morimoto & Lihoreau 2019, Ma et al. 2022, Barragan-Fonseca et al. 2021). There are three major gaps in our approaches to analyse GF data which contribute to inconsistencies in the analysis in the field of nutritional ecology: (i) there has been no investigation on the suitability and accuracy of different statistical models in identifying regions of interest (e.g., peaks, valleys) in multidimensional performance landscapes (see e.g., Rapkin et al. 2018, Morimoto & Lihoreau 2019). The reason for this is because (ii) current methods to analyse multidimensional GF performance landscapes are complex and difficult to implement in a comparative fashion, precluding the exploration of the full range of statistical methods commonly used in ecology and evolution (e.g., linear regressions, additive models etc). Moreover, (iii) the structure of the data that GF empirical studies generate is also inconsistent. For instance, some GF studies measure both the amount of food ingested and the trait values on each diet across a broad range of dietary compositions, while others consider diets as treatments and examine the effect of a wide range of diets on trait values without measuring food intake [e.g., (e.g., Maklakov et al. 2008, Kutz et al. 2019, respectively). This difference in experimental approach leads to distortions in the generated performance landscapes which could influence the estimates of the strength of nutritional trade-offs. There has been no study to investigate how current analytical methods perform in identifying peaks and valleys across different data structures. Collectively, these gaps in analytical methods used to estimate nutritional trade-offs limit our ability to draw conclusions between studies and, more broadly, preclude our understanding of the evolution of nutritional responses in comparative analysis. Thus, there is a need for the development of consistent analyses which derive precise estimates of peak (and valley) regions from GF landscapes, thereby enabling large-scale comparative analyses necessary to gain insights into the evolution of animal nutrition.

Here, we address these gaps by proposing a novel analytical model (Nutrigonometry) that is simple, precise, and can accommodate different data structures. We then use this model to perform the

first comparative analysis of different statistical methods and their performance in identifying regions of interest (e.g., peaks or valleys) in GF performance landscapes. Nutrigonometry is different from our previous Vector of Positions approach (Morimoto & Lihoreau 2019) because the latter used a single statistical model (i.e., a support-vector machine (SVM) model) which required an arbitrary input threshold to identify peak regions. Moreover, the Vector of Positions could not identify and delineate valley regions as accurately and the overall approach was computationally expensive to apply to multiple landscapes. Nutrigonometry uses trigonometric relationships (i.e., Pythagoras theorem) to estimate the difference in angles of right-angle triangles formed in performance traits (Figure 3.1a), after which the difference between these angles can be estimated as a proxy for the strength of nutritional trade-offs (i.e., wider angles indicate stronger trade-offs) (Figure 3.1b).

Below, we first describe the mathematical foundation underlying the Nutrigonometry method. Next, we apply Nutrigonometry to a landmark GF dataset in *Drosophila melanogaster* with known nutritional trade-offs between lifespan and reproductive rate (Lee et al. 2008, Morimoto & Lihoreau 2019). This dataset provided an important ground-truth to apply, test, and validate the precision of Nutrigonometry in identifying the known nutritional trade-off. Moreover, the use of this dataset allowed us to demonstrate the consistency of the Nutrigonometry model in identifying nutritional trade-offs when the structure of the input dataset varied (i.e., individual intake vs fixed ratio experiments). Finally, we leveraged the simplistic (and computationally cheap) approach provided by Nutrigonometry to conduct the first comparative analyses of a range of 'off-the-shelf' statistical models (including machine learning) in estimating the peak region in the performance landscapes, which is an essential component for proper quantification of nutritional trade-offs to perform a comparative analysis.

3.3 Material and Methods

3.3.1 Nutrigonometry

GF studies consider a 'nutritional space' in which animals can eat foods and navigate their nutritional state (for a review, see Raubenheimer & Simpson 2020). This nutritional space is defined by the food components (typically macro-nutrients) under investigation. Foods are represented in the nutritional space as nutritional rails (i.e., imaginary lines that pass through the origin with a given positive slope) characterized by different ratios of the food components. For example, in studies where protein and

carbohydrate effects are investigated, there is a 2D nutritional space (one dimension for each nutrient) onto which the performance landscape of the trait is mapped. This rationale can be extended to n number of nutrients (Simpson & Raubenheimer 1993), albeit to date, studies with two nutrients are the most common (Morimoto & Lihoreau 2020). If we consider this 2D nutritional space as a rectangular space in which an infinite number of nutritional rails (i.e., foods) exist that divides the space in right-angle triangles, then it is possible to use simple trigonometric functions to estimate the angle α_i and the hypothenuse of the triangle, for all fitness traits mapped onto the nutritional space. The angle α_i is the angle of the nutritional rail, relative from the x-axis, that passes through the peak in the landscape for the trait i, and the hypothenuse h_i of the triangle shows how far from the origin the peak in the landscape sits for the trait i (Figure 3.1a). α_i and h_i can be calculated using the Pythagorean theorem and the relationship between the angle and the sides of right-angle triangles (i.e., sines and cosines), as shown in (Figure 3.1a).

Once α_i and α_j are known, we can estimate the angle θ [as in (Morimoto & Lihoreau 2019)] which is the difference in the angle between nutritional rails that maximize two traits, i and j, and provides a measure of the strength of the nutritional trade-off that exists between traits i and j (Figure 3.1b). The larger the angle $\theta_{(i,j)}$, the stronger the nutritional trade-off in terms of nutrient balance (and potentially nutritional compromise) between traits. Likewise, we can compare the difference $h_{(i,j)}$ in the estimates of the hypothenuse h_j and h_i to quantify nutritional trade-offs in relation to nutrient concentration (Figure 3.1b). These metrics allowed us to disentangle the following theoretical scenarios in which nutritional trade-off can occur: (i) when $\theta_{(i,j)}$ is large but $h_{(i,j)}$ is small ('Strong nutritional trade-off in terms of nutrient balance'), (ii) when $\theta_{(i,j)}$ is small but $h_{(i,j)}$ is large ('Strong nutritional trade-off in terms of nutrient concentration'), (iii) when $\theta_{(i,j)}$ and $h_{(i,j)}$ are large ('Strong nutritional trade-off in terms of both nutrient balance and concentration'), and (iv) when $\theta_{(i,j)}$ and $h_{(i,j)}$ are small ('Weak or no nutritional trade-off') (Figure 3.1c).

Here when applying this model for empirical datasets (see below), inferences on the strength of nutritional trade-offs were made using confidence intervals for $h_{(i,j)}$ and $\theta_{(i,j)}$, whereby nutritional trade-offs were stronger when confidence intervals did not overlap zero and the magnitude of the difference was large. Estimates are presented in the units of the nutrient space in which the data was collected (e.g., mg), while angles are presented in degrees. Confidence intervals for both $h_{(i,j)}$ and $\theta_{(i,j)}$ were calculated using the significance threshold of 0.05 and the quartiles of a t-distribution. All analyses and plots were done in R version 3.6.2 (Team 2019).

3.3.2 Datasets used for model application: diet intake and fixed ratios

Nutrigonometry was designed to work on diverse types of GF data, including those generated from experimental approaches measuring diet intakes or fixed ratio diets (see 'Introduction'). To validate this, we applied the Nutrigonometry model to a landmark dataset which measured the nutritional requirements for lifespan and reproduction in *Drosophila melanogaster* (Lee et al. 2008). This dataset was used to test previous approaches and therefore has important benchmark status in the field, providing the ground-truth data for the identification of the now known nutritional trade-offs between lifespan and reproductive rate (Morimoto & Lihoreau 2019). This data has the benefit of allowing us to implement the Nutrigonometry model to measure the trade-off in data of both intake and fixed ratio structures. This is because the original data carefully estimated individual intake across diets and provided the ratio of nutrients of all diets, enabling us to transform the data structure from intake to fixed ratio. We also demonstrated the application of Nutrigonometry to a GF dataset with fixed ratios from (Kutz et al. 2019), which studied how temperature modulates nutritional responses of larval development and adult fitness in *D. melanogaster*, as an auxiliary demonstration of the application of our model (see Figure S1 in the supplementary material for Chapter 3 below).

3.3.3 Analytical approach

The implementation outline of Nutrigonometry is as follows:

Predicting peak (or valley) location and size. We designed algorithms that fitted a range of statistical models (details of the models below) to the empirical data in order to test their performance in predicting peaks (or valleys; see Supplementary Material of this chapter) in the performance landscape for lifespan and reproductive rate from our landmark dataset. This comparative statistical approach enabled us to provide the first comparative assessment of models to identify peak regions in GF performance landscapes. This was done for both the dataset with nutrient intake and with fixed ratios. This enabled us to predict peak regions in the performance landscapes for which further analysis was possible.

Model fit. We then tested the fit of all models using several quantitative parameters such as (i) root-mean-square-error (RMSE), (ii) peak area and spread, and (iii) homogeneity of points within the predicted peak region. These provide information on the error in the predictive estimates of the performance trait in the z-axis, the spread that the models predicted the peak in the landscape to comprise, and the clustering of points that delineate the predicted peak region, respectively.

Nutrient balance. We then compared the nutrient ratio of the predicted peak regions with the nutrient ratio that animals chose when given the freedom to self-balance their diet (aka 'nutrient intake target'). This allowed the model to infer whether or not animals are maximising any particular performance trait (i.e., the ratio of intake target coincides with ratio of predicted peak). Below, we provide the details of each of the steps outlined above.

Predicting peak (or valley) location and size. As with previous approaches, our model depends on accurate estimates of the coordinates for the peak in the multidimensional performance landscape. Without this, estimates of $h_{(i,j)}$ and θ are inaccurate, which in turn affects the ability of the model to estimate the strength of nutritional trade-offs. To overcome this, the basic algorithms underpinning the identification of peak regions in performance landscapes were designed as follows: (a) Empirical data was split into training (75%) and test (25%) datasets; (b) The statistical model was fitted to the training set using 10-fold cross-validation, with the fitness trait as the dependent variable and the nutrient intakes (or fixed ratios) as independent variables. The model included main and interactive effects of protein and carbohydrate, as well as quadratic effects of each nutrient (for non-linear relationships); (c) A set of 500 random points corresponding to (protein, carbohydrate) coordinates was generated, and the model of step (b) was used to predict the value of the performance trait at these points and selected the points with the highest 5% predicted performance trait values; (d) Step (c) was repeated 100 times.

Peak area was then estimated as the area of the convex hull incorporating all of the predicted points from the above algorithm. 95% confidence intervals of peak area were estimated using the 'ci' function of the 'Rmisc' package (Hope et al. 2013).

We compared the performance of several statistical models (including commonly used statistical GF literature), namely, Bayesian linear regression (Bayes), general linear regression (LM), k-nearest neighbours (KNN), Gradient boost (GBoost), random forest (RF), support vector machine (SVM) with radial basis function as well as generalized additive models (GAMs) with both smooth term or tensor product terms. Note that SVM models were used as the underlying model in the Vector of Positions approach in (Morimoto & Lihoreau 2020) and thus, provides the grounds where the two methodologies can be compared. With the exception of GAMs that were fitted using the 'mgcv' package (Wood & Wood 2015), all other models were fitted using the 'tidymodels' package of the tidyverse (Wickham et al. 2016). Performance landscapes were estimated using the 'Tps' function of the 'fields' package (Nychka et al. 2017). We fitted the majority of models with default parameters, as these are the most likely approach from a beginner starting to work with GF data. Automated parameter tuning built into the 'tidymodels' pack-

age was done for the 'mtry' argument (RF and GBoost), the 'cost' argument (SVM), and the 'neighbors' and 'weight_func' (KNN). For the Bayesian regression (Bayes), we fitted a weakly informative Cauchy prior using the 'rstan::cauchy(...)' function (Goodrich et al. 2020). All plots were constructed using the 'ggplot2' package (Wickham & Wickham 2016). We also demonstrated how the best performing models in our peak analyses can be used to predict valley regions in GF datasets (Figures S2 and S3 in the supplementary material of Chapter 3). The R code with the functions used for this paper is available as supplementary information.

Model fit

RMSE. The root-mean-square-error (RMSE) was estimated as the difference between the predicted (generated by the above algorithm) and observed (from the empirical dataset) values for the performance trait. Note that RMSE values do not interfere with accuracy of estimates of $h_{(i,j)}$ and θ, and thus the estimates of nutritional trade-offs, because the z-axis is not used in the calculation of angles and hypothenuses (Figure 3.1d). A model can have high RMSE and still be the best predictive model as long as the predicted peak correctly matches with the observed peak in the landscape.

Peak area. In addition to RMSE, we estimated the area (in squared units in which the data is collected) of the polygon delimited by the estimated predicted peak region ('Area') and the horizontal (protein) and vertical (carbohydrate) spread of the datapoints of the predicted peak region ('Nutrient spread') as proxies for the precision of our predicted peak regions (Figure 3.1d). The smaller the area and nutrient spread, the more compact the prediction of the peak region in the nutritional space.

Homogeneity of points within the predicted peak region. Even in cases where RMSE, area and nutrient spread of the predicted peak regions are small, it is important to have evenly-spaced datapoints within the predicted peak region. This is because predictions of regions which contain holes can lead to mis-estimation of the strength of nutritional trade-offs by potentially adding noise to the set of protein and carbohydrate coordinates used to calculate $h_{(i,j)}$ and the angle θ. We measured the topological structure of the predicted peak region using the concept of persistence homology (PH), which in simple terms, allows us to investigate the overall structural organization of the data (see Text S1 in the supplementary material for Chapter 3) and (Zomorodian & Carlsson 2004, Weinberger 2011) for details of the concept] (Figure 3.1d). PH was estimated using the 'TDAstats' package (Wadhwa et al. 2018). Together, the estimates of RMSE, area, nutrient spread, and PH provided a comprehensive suite of metrics to assess the quality of model predictions for the peak region in fitness landscapes.

Nutrient balance

Drosophila melanogaster adults balance their nutrient intake to a protein to carbohydrate (P:C) ratio of 1:4 when given the possibility to self-select multiple nutritionally complementary foods (see results in the original study by Lee et al. 2008). We then used the peak predictions of the Nutrigonometry framework to test whether the observed P:C ratios that maximized lifespan and reproductive rates coincided with the P:C ratio of 1:4 reached by flies in choice situations. To achieve this, we calculated the 95% confidence interval as described for the peak area but in this case, for the P:C ratio of each trait. Whenever the confidence interval overlapped 1:4, we inferred that the estimate of peak ratio did not statistically differ from the intake target of 1:4.

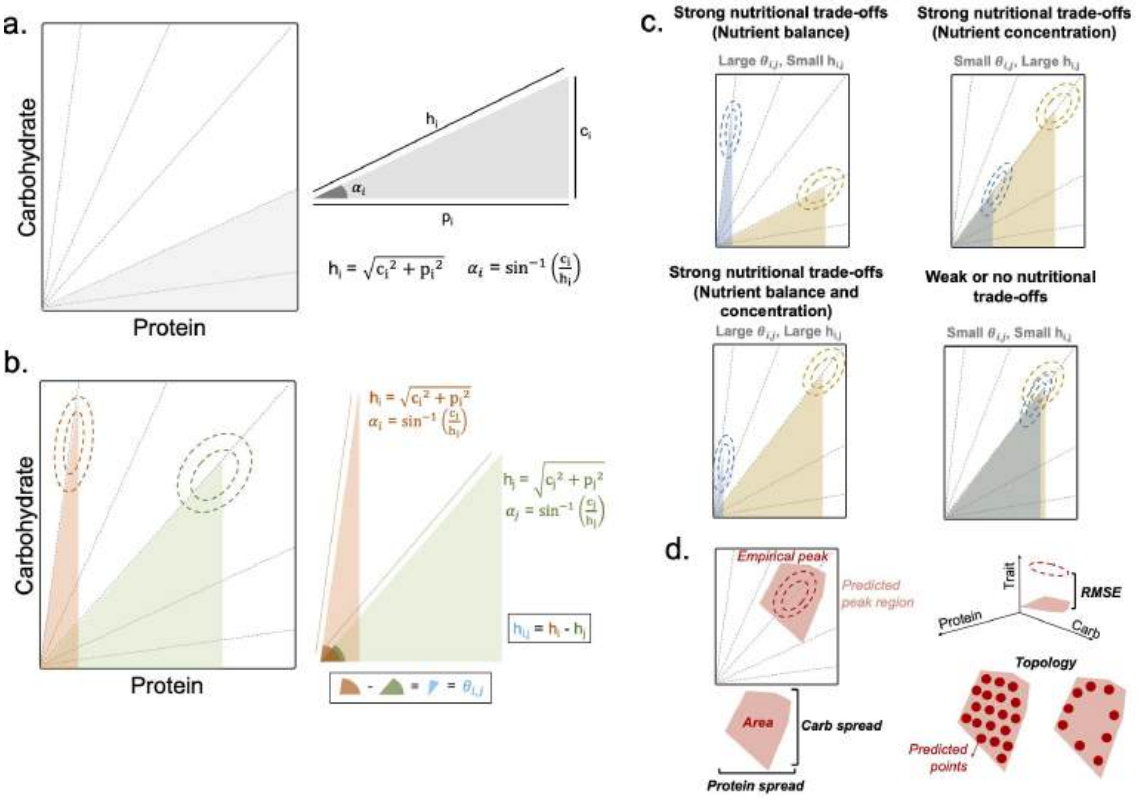


Figure 3.1: **The Nutrigonometry model.** (a) Considering an infinite number of nutritional rails that divide the nutritional space into right-angle triangles, the angle α_i and the hypothenuse h_i can be calculated from trigonometric relationships. (b) Nutrigonometry allows for the estimates of the strength of nutritional trade-offs in terms of nutrient balance (angle $\theta_{i,j}$)) and nutrient concentration (the difference $h_{i,j}$), given in absolute terms). (c) Scenarios for the estimates of the strength of nutritional trade-offs with respect to the angle $\theta_{i,j}$) and the length $h_{i,j}$). (d) Metrics used to the peak prediction in the 3D landscape (see Methods within this chapter for details).

3.4 Results

Simple (Bayesian) linear regressions outshine machine learning models when predicting peak region in multidimensional landscapes

First, we applied the Nutrigonometry model with several underlying statistical models to test the precision of the framework in finding peak regions in the performance landscapes. All models generated predictions of peak region in nutritional landscapes irrespective of data structure although the accuracy and topology of the predicted regions varied (Figures 3.2, 3.3, and 3.4). In general, GAMs with tensor product and smooth function as well as Bayes and LM linear models generated peak predictions for both lifespan and reproductive rate that were significantly more precise (narrower) than other models when the structure of the data was composed of food intakes (Figure 3.2 and Tables S1 and S2 in the supplementary material for Chapter 3).

When the data structure changed to fixed ratios, LM, GAM with tensor product, Bayes, and KNN predicted peaks with smaller area for lifespan and all but KNN perform within similar scales for the peak prediction of reproductive rate (Figure 3.3 and Tables S1 and S2 in the supplementary material for Chapter 3). In comparison, GAM smooth did not perform well in predicting peak region that was homogenous and precise in the performance landscapes.

The performance of the models was independent of the estimates of RMSE and nutrient spread which showed no clear pattern of performance. The exceptions were that LM and Bayes displayed consistently lower spread when the structure of the data were intakes (Figures 3.4 and 3.5, Tables S1-3 in the supplementary material for Chapter 3).

Interestingly, machine learning models consistently underperformed, predicting peak regions that were wider and less precise (Figures 3.2, 3.3, 3.4, Tables S1 and S2 in the supplementary material for Chapter 3). The underlying reason for this is unclear, but similar patterns were observed when predicting the peak region of (Kutz et al. 2019) dataset (Figure S1 and Table S3 in the supplementary material for Chapter 3). This is important because the previous Vector of Position approaches uses SVM models which, as shown here, do not perform as well as other simpler models (Morimoto & Lihoreau 2020). Bayes, GAMs (both smooth and tensor product), and LM also performed well when predicting valley regions (Figure S2 and Figure S3 in the supplementary material for Chapter 3). These results indicate that simple (Bayesian) linear regression consistently provides precise estimates of peak regions in performance

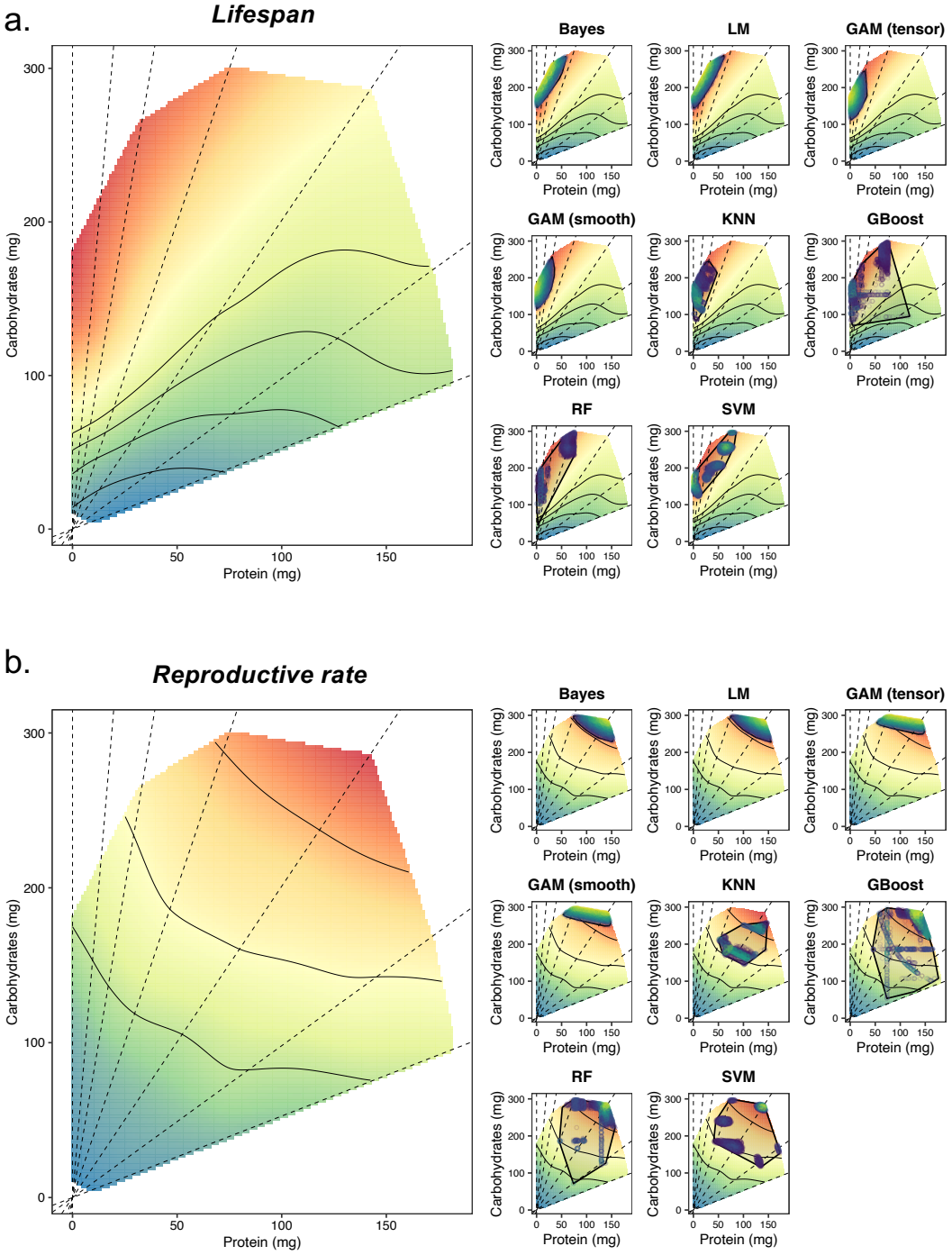


Figure 3.2: **Predictions of peak region in lifespan and reproductive rate landscape with intake data.** (a) Lifespan landscape with the overlaid predicted peak regions (left small panels). (b) Reproductive rate landscape with the overlaid predicted peak regions. Red represents peaks while light green represents valleys. For the predicted region, dark blue represents points with lower predicted $z-$values whereas bright yellow represents points with higher predicted $z-$values. Shaded polygon added to help visualisation. Dotted lines represent the nutritional rails (i.e., foods with fixed P:C ratios) upon which animals were allow to eat.

landscapes irrespective of the structure of the data, and that GAMs with tensor product (and to a smaller extent, smooth function) can also be used when the data structure is of individual intakes.

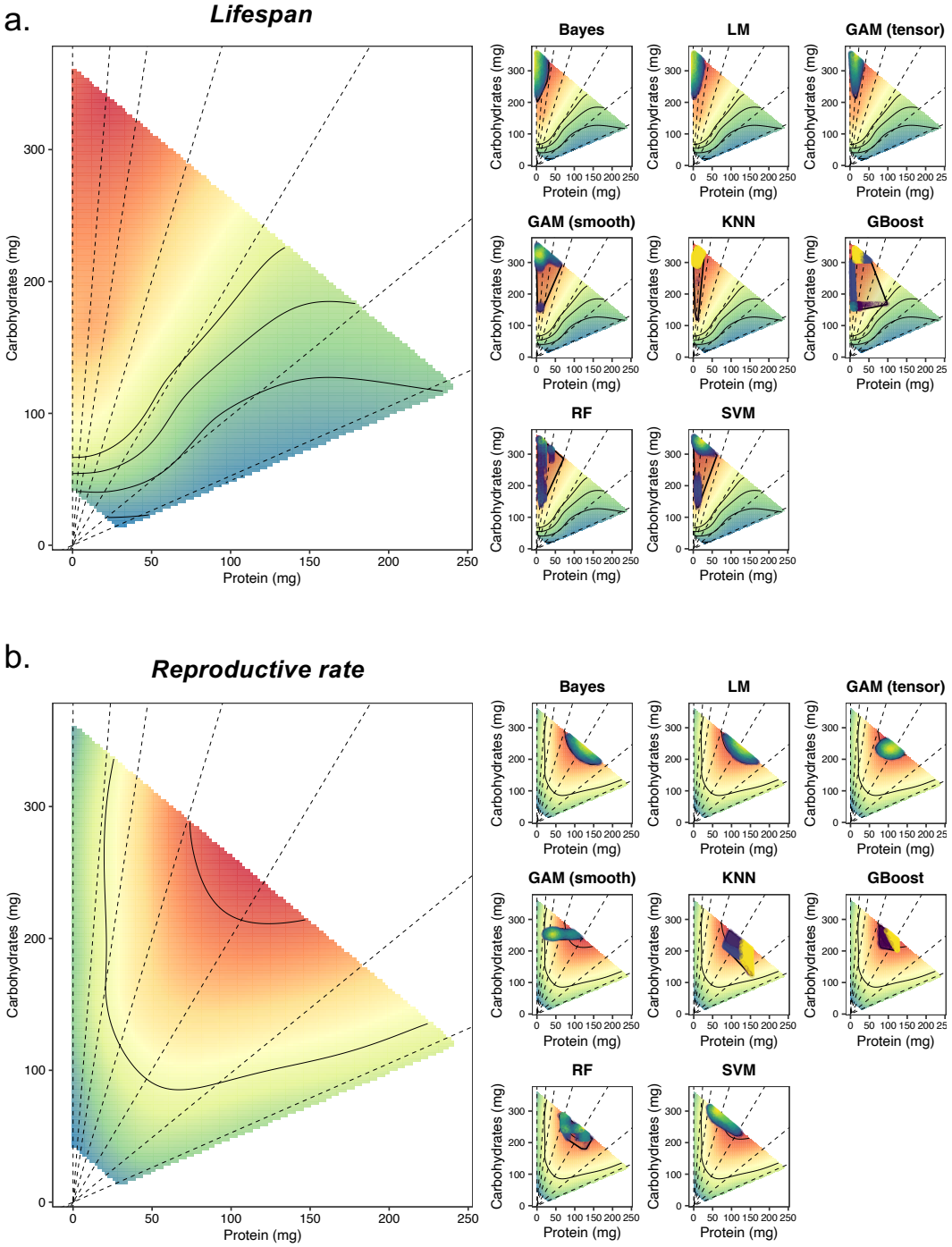


Figure 3.3: **Predictions of peak region in lifespan and reproductive rate landscape with fixed ratio data.** (a) Lifespan landscape with the overlaid predicted peak regions (left small panels). (b) Reproductive rate landscapes with the overlaid predicted peak regions. For the landscapes, red represents peaks while light green represents valleys. For the predicted region, dark blue represents points with lower predicted $z-$values whereas bright yellow represents points with higher predicted $z-$values. The shaded polygon was added to facilitate visualisation of the predicted peak region and the homogeneity of points within the predicted peak.

3.4.1 Precision in estimates of known nutritional trade-offs in performance landscapes

Next, we tested whether Nutrigonometry was capable of reliably quantifying known nutritional trade-offs in the dataset. The dataset used here is known to contain a strong trade-off between lifespan

and reproductive rate (Lee et al. 2008, Morimoto & Lihoreau 2019). We therefore tested how different statistical models in Nutrigonometry performed when estimating this nutritional trade-off. GAMs (both smooth and tensor product), Bayes, LM, and KNN models were the only models that correctly identified the known nutritional trade-off in the dataset, measured by angle θ between lifespan and reproductive rate for data with individual intakes (Table 3.1). Given the variability in the area, spread, and topology of the predicted region, estimates of $h_{i,j}$ and θ were more precise (narrower confidence intervals) for GAMs (smooth and tensor product), Bayes and LM compared with KNN. GAMs, Bayes, and LM were the only models that identified a trade-off on the hypothenuse estimate $h_{i,j}$ for data of individual intakes, while KNN was the only model that identified this trade-off in data with fixed ratio (Table 3.1). Thus, overall, simpler models are more suitable to generate peak predictions that accurately describe nutritional trade-offs in multidimensional performance landscapes for data of different structures.

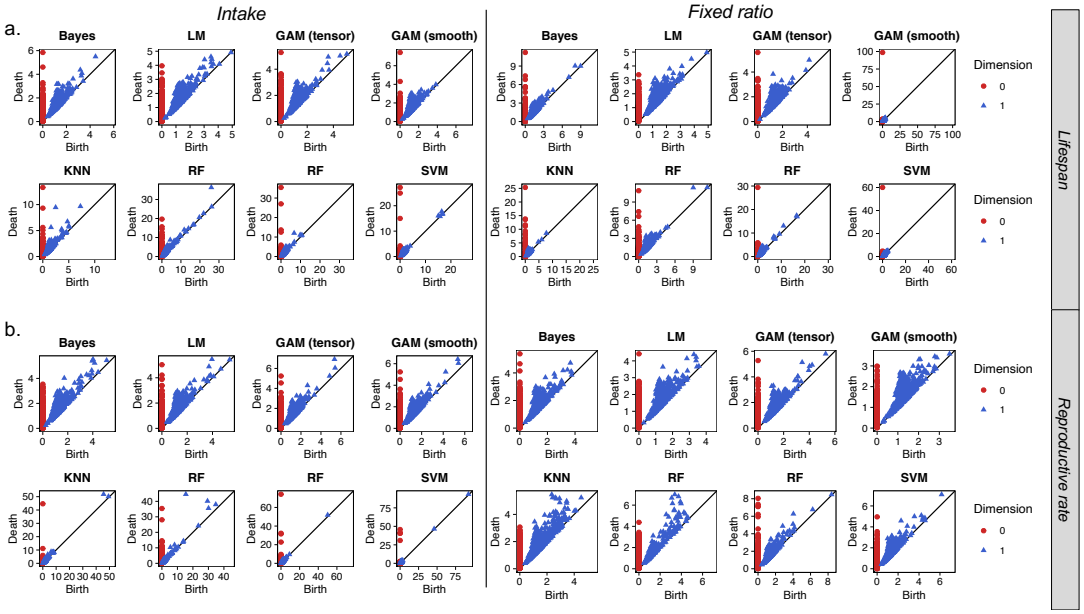


Figure 3.4: **Persistence Homology (PH) for the topological structure of the predicted peak region.** (a) PH plots of the predicted peak region in lifespan of data containing the structure of individual intake (top left) and fixed intake data (top right). (b) PH plots of the predicted peak region in reproductive rate with data of structure containing individual intake (bottom left) and fixed intake data (bottom right). Homogenous predicted peaks have red (dimension 0) and blue (dimension 1) points that are closer, as opposed to more heterogeneous predicted peaks upon which (some) points can be farther from each other.

3.4.2 Comparing trait optimum with intake target

When given a choice of imbalanced diets, animals balance their nutrient intake to ratios that maximize the expression of some fitness traits at the expense of others. In the landmark dataset used here, *D. melanogaster* flies balanced their P:C to a ratio of 1:4, which does not coincide with the P:C ratios that

maximize lifespan nor reproductive rate (Lee et al. 2008). We tested whether Nutrigonometry could identify such known nutritional trade-off in the data. All models predicted a significantly lower optimum P:C ratio that maximizes reproductive rate relative to lifespan as expected from the original visual comparison of landscapes (around 1:2 for reproductive rate and >1:9 for lifespan) (Table 3.2). However, none of the estimates overlapped the P:C 1:4. This confirms the findings from the original study that *D. melanogaster* females do not balance their nutrient intake to maximize lifespan or reproductive rate, but instead balance their P:C to 1:4 in order to maximize lifetime egg production (Lee et al. 2008). This validates the power of Nutrigonometry in analysing nutritional behaviour in GF studies (Lee et al. 2008).

Table 3.1: Quantification of nutritional trade-offs between lifespan and reproduction. Estimates of $\theta_{i,j}$ (in degrees) and $h_{i,j}$ (in mg) are provided for the nutritional trade-off between lifespan and reproductive rate. The analysis is based on the data presented in Lee et al. (2008). Confidence intervals overlapping zero imply no significant differences in the peaks. The magnitude of the estimates reflects the strength of nutritional trade-offs, with larger magnitudes indicating stronger trade-offs. Note that $\theta_{i,j}$ is constrained between 0 and 90 degrees (i.e., 0 and $\frac{\pi}{2}$).

	Parameter	**Model**	**Estimate**	**Std**	**LwrCI**	**UprCI**
Trade-off (intakes)	$\theta_{i,j}$	SVM	14.456	10.728	-6.574	35.485
		RF	14.508	8.109	-1.388	30.404
		GAM_tensor	16.128	4.984	6.358	25.897
		GAM_smooth	16.166	4.962	6.438	25.893
		GBoost	17.063	9.575	-1.706	35.831
		LM	17.940	4.826	8.479	27.400
		Bayes	18.205	4.709	8.974	27.436
		KNN	21.203	6.181	9.088	33.318
	$h_{i,j}$	SVM	16.792	65.723	-112.038	145.622
		KNN	50.015	48.137	-44.343	144.373
		GBoost	52.851	75.218	-94.591	200.293
		RF	58.561	66.066	-70.943	188.064
		LM	75.870	35.142	6.984	144.757
		Bayes	76.729	34.444	9.211	144.247
		GAM_smooth	120.245	29.406	62.604	177.886
		GAM_tensor	124.533	27.930	69.784	179.282
Trade-off (fixed)	$\theta_{i,j}$	GAM_smooth	9.645	5.897	-1.916	21.205
		SVM	11.840	5.649	0.767	22.913
		RF	17.368	5.848	5.906	28.831
		GBoost	20.177	5.057	10.264	30.090
		GAM_tensor	21.177	3.872	13.588	28.766
		Bayes	26.454	5.876	14.935	37.973
		LM	26.499	5.903	14.928	38.070
		KNN	31.428	7.186	17.342	45.513
	$h_{i,j}$	SVM	2.381	68.888	-132.653	137.416
		RF	4.819	64.841	-122.283	131.921
		GBoost	9.377	65.605	-119.222	137.975
		Bayes	41.461	34.912	-26.974	109.896
		LM	42.305	34.429	-25.182	109.791
		GAM_smooth	46.635	40.358	-32.475	125.745
		GAM_tensor	49.009	32.855	-15.394	113.412
		KNN	82.516	30.388	22.949	142.083

3.5 Discussion

We proposed a new simple analytical framework to analyse nutritional trade-offs in multidimensional fitness landscapes. Nutrigonometry uses trigonometric relationships from right-angle triangles to identify and compare peaks (or valleys) in 3D performance landscapes. Using a landmark GF dataset, we demonstrated the precision and performance of standard statistical (machine learning) models in finding the peak regions in performance landscapes and subsequently quantifying the strength of nutritional trade-offs between traits. In the first comparative analysis of statistical methods in GF datasets, we showed that simpler general linear models outshined machine learning models in the prediction of peak and valley regions in the performance landscapes. Thus, Nutrigonometry is a powerful yet easy-to-implement methodology to determine the strength of nutritional trade-offs in fitness studies in the field of animal nutrition.

Multidimensional studies of nutrition through the GF have been increasingly used to gain insight into animal and human nutrition (Lee et al. 2008, Behmer & Joern 2008, Felton et al. 2009, Simpson & Raubenheimer 2012, Hewson-Hughes et al. 2012, Gosby et al. 2014, Solon-Biet et al. 2014). Likewise, the complexity of the applications has also increased, ranging from studies with few nutrients (e.g., protein and carbohydrates, salts) through to high-dimensional studies investigating individual fatty acids and amino acids (Simpson et al. 2006, Grandison et al. 2009, Arien et al. 2015, Arganda et al. 2017, Piper et al. 2017). This means that analytical frameworks that are simple and robust must be developed to support the development of the field.

In this study, we proposed such a framework. Previous methods were either computationally expensive (and imprecise) or could, in some cases, overestimate the strength of nutritional trade-offs. For instance, Rapkin et al. (2018) proposed a model that used regression slopes of the nutrients onto the performance trait as coordinates of a vector $\vec{v}_i$, for performance trait i. From these vectors, the angle θ' between vectors $\vec{v}_i$ and $\vec{v}_j$ for traits i and j, respectively, could be calculated as the estimate of the strength of the nutritional trade-off. This method has the limitation that regression slopes can be positive, zero, or negative, whereas performance landscapes can only exist in the positive numbers (i.e., animals cannot eat negative amounts of nutrients). This can lead to overestimation of nutritional trade-offs (Morimoto & Lihoreau 2019). We then proposed a method, known as the Vector of Positions approach, that addressed this limitation and used vectors with coordinates that matched the location of peaks in the domain of positive numbers (Morimoto & Lihoreau 2019). However, this approach was limited because (i) we used

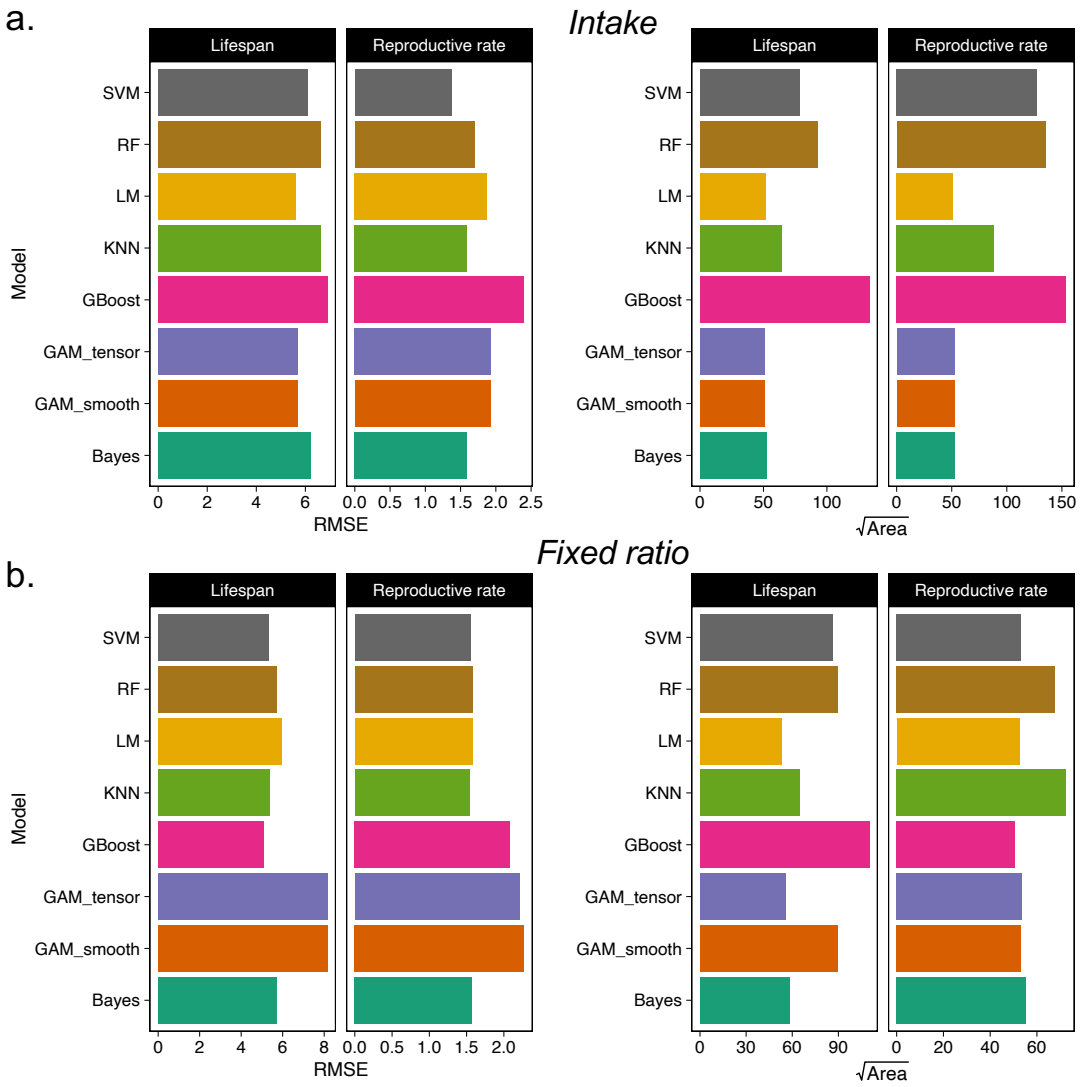


Figure 3.5: **RMSE and peak area estimates in peak region predictions.** (a) RMSE and predicted peak area (i.e., area of the shaded polygon from the predicted region for lifespan and reproductive rate data), with structure containing individual intakes. (b) RMSE and predicted peak area (i.e., area of the shaded polygon from the predicted region for lifespan and reproductive rate data), with structure containing fixed ratios. Note that models with high RMSE can still be the best predictors of peak region.

an SVM model which required an arbitrary input threshold to identify peaks and (ii) was computationally expensive to apply to multiple landscapes. Nutrigonometry resolves the limitations from the previous models as it (1) estimates nutritional trade-offs using the domain of positive numbers, (2) demonstrates, in the first comparative study of statistical methods, that simpler (Bayesian) linear regressions are often more precise, and (3) is computationally cheap to run as it relies on simple trigonometric relationships. Thus, Nutrigonometry enables reliable large-scale studies on the fitness consequences of animal foraging decisions that will shed light into the evolution of physiological and behavioural responses to nutrition.

We provided the first comparative assessment of the power and precision of common statistical mod-

Table 3.2: Estimates of nutritional compromises. Estimates of optimal intake that maximises lifespan and reproductive rate based on the predicted peak region. Comparison made with the visual peak ratio from Lee et al. (2008) Note that all but one model (i.e., GAM smooth for 29 fixed ratio reproductive rate data) predicted peak region 1:4, which is the ratio that individuals balance when given the ability to balance their diet. Other models suggest that a P:C ratio of 1:4 is lower than the ratio needed to maximise lifespan but higher than that for reproductive rate.

Data	Trait	Model	Mean	Upr CI	Lwr CI	Target (Visual)
Peak Ratio (intakes)	Lifespan	GBoost	5.235	5.205	5.265	16
		RF	5.533	5.508	5.557	
		SVM	5.864	5.836	5.892	
		LM	9.084	9.048	9.120	
		Bayes	9.154	9.118	9.190	
		KNN	12.075	12.015	12.135	
		GAM_smooth	13.055	12.997	13.114	
		GAM_tensor	13.108	13.049	13.168	
	Reproductive rate	GBoost	1.858	1.853	1.864	2
		KNN	2.041	2.037	2.045	
		RF	2.138	2.133	2.144	
		SVM	2.147	2.138	2.156	
		Bayes	2.194	2.191	2.197	
		LM	2.215	2.212	2.219	
		GAM_smooth	2.644	2.639	2.650	
		GAM_tensor	2.661	2.656	2.667	
Peak Ratio (fixed)	Lifespan	GBoost	13.078	12.987	13.170	16
		SVM	14.946	14.873	15.019	
		RF	15.107	15.045	15.169	
		GAM_smooth	16.977	16.859	17.097	
		GAM_tensor	20.623	20.536	20.710	
		KNN	26.050	25.929	26.173	
		Bayes	32.426	32.237	32.617	
		LM	32.933	32.744	33.125	
	Reproductive rate	KNN	1.467	1.463	1.470	2
		LM	1.836	1.832	1.839	
		Bayes	1.841	1.837	1.845	
		GBoost	2.171	2.168	2.174	
		GAM_tensor	2.244	2.241	2.248	
		RF	2.545	2.539	2.552	
		SVM	3.527	3.516	3.538	
		GAM_smooth	4.280	4.265	4.296	

els in identifying regions of interest (e.g., peaks or valleys) in multidimensional performance landscapes. More importantly, we showed that Nutrigonometry provides a clear, concise, and simpler foundation for analysing GF data, by demonstrating the best approach to investigate nutritional trade-offs in 3D fitness landscapes. Because Nutrigonometry uses trigonometric relationships of right-angle triangles, it is applicable to n dimensions. However, given the often counter-intuitive geometrical effects of high dimensionality (Milman 1998, Watanabe 2022), such expansion to higher dimensions requires further investigation as the topic of future developments. Nevertheless, given the broad use of 3D fitness landscapes in GF studies (Morimoto & Lihoreau 2020), Nutrigonometry readily enables important quantifications of nutritional trade-offs that were otherwise absent or cumbersome to produce.

For instance, using a range of models, Nutrigonometry uses right-angle triangles to compare the ratio of nutrients that maximize lifespan and reproductive rate along with the strength of nutritional trade-offs between these traits in a landmark paper in the field (Lee et al. 2008). Moreover, Nutrigonometry is capable of comparing the nutrient ratio that maximizes lifespan and reproductive rate with the nutrient ratio balanced by individuals when given a choice, providing important insights into the dietary choices underpinning nutritional compromises. Such quantification can bring new fundamental insights into our understanding of nutritional trade-offs such as the strength and the direction of the trade-offs (e.g., nutrient balance vs. concentration, see Figure 3.1), as well as how much animals actually resolve these trade-offs when they have the opportunity to do so and whether, for instance, they favour one trait over another (distance between optimal trade-off and observed nutrient intake target, see Table 3.2).

An important trend in the field of multidimensional nutrition is the study of nutritional effects across physiological pathways and across levels of biological organization (Lihoreau et al. 2015, Simpson et al. 2015). These studies generate multiple performance landscapes that are often compared visually, without rigorous analytical methods to measure nutritional trade-offs. This limits our ability to identify diet balances that maximize (or minimize) gene expression of a particular pathway. For example, regions of eleven performance landscapes of the expression of genes involved in the Insulin/IGF pathway were visually compared to provide insights into how a key endocrine pathway is regulated based on nutrient intake, and how gene expression can underlie the expression of life histories (Post & Tatar 2016, McDonald et al. 2021). Likewise, regions of twelve performance landscapes with gut microbial diversity or abundance were visually compared to better understand how nutrient composition can modulate host-microbe interactions (Ng et al. 2019). Similar visual comparisons have been made to understand the effects of nutrition on host-endosymbiont relationships (Ponton et al. 2015).

Of course, the goal of these molecular studies may not have been the identification of nutrient ratios that optimize (or minimize) gene expression but instead, the relative contribution of specific nutrients to changes in gene expression profiles. Nevertheless, identifying peaks and valleys in GF gene expression landscapes is a useful concept with wider application to veterinary and medical sciences (Simpson & Raubenheimer 2009). Nutrigonometry will allow researchers to move beyond visual comparisons to quantitatively assess how landscapes differ using a rigorous and reproducible framework, providing additional tools for better understanding the properties of multidimensional performance landscapes. As a result, Nutrigonometry yields considerable advances to the status quo in the field, enabling a deeper understanding of the role of nutrition in physiology, behaviour, and ecology.

3.6 Conclusion

We propose a model that is simple and robust to analyze performance landscapes from GF studies. Contrary to previous methods (Rapkin et al. 2018, Morimoto & Lihoreau 2019), Nutrigonometry does not rely on vector calculations but instead harnesses the trigonometric relationships of right-angle triangles to estimate nutritional trade-offs. This is a major advance of the model as it considerably simplifies the framework both in conceptual and computational terms. Nutrigonometry thus significantly advances our ability to generate reliable and reproducible estimates of nutritional trade-offs within and between species, facilitating quantitative studies of animal nutrition. These advances in our approach open new avenues of research in multidimensional nutrition and allow for physiological and comparative studies to be performed in a consistent and reproducible way from which insights into the evolution of animal nutrition can be gained across the tree of life.

3.7 Acknowledgements

JM is supported by the BBSRC (BB/V015249/1). PC is supported by the EPSRC (EP/P025072/) and from École Polytechnique Fédérale de Lausanne via a collaboration agreement with the University of Aberdeen. ML receives support from the CNRS, the French Research Agency (ANR 3DNaviBee: ANR-19-CE37-0024), the French Environment and Energy Management Agency (ADEME LOTAPIS), the European Regional Development Fund (FEDER ECONECT: MP0021763), and the European Research Council (ERC-CoG BEE-MOVE: GA101002644). The authors would like to thank Prof. Kwang Lee, Dr. Teresa Kutz, and Prof. Carla Sgrò for kindly sharing the data used to demonstrate the use of our model. We would also like to thank Gordon M. Hay for translating our abstract into Doric.

3.8 Data and Code Accessibility

The Kutz et al. (2019) data is available at: `https://doi.org/10.26180/5cfe1ddaaafac`. The Lee et al. (2008) dataset is accessible via Dryad: `https://doi.org/10.5061/dryad.tp7519s`. R scripts with functions for the implementation of the Nutrigonometry framework are available on Dryad/Zenodo (Morimoto et al., 2022): `https://datadryad.org/stash/share/zYqmh4j58HKRPm-Grf3F-8BiINThiworl_jVdsVi6us`.

3.8.1 Appendix: Supplementary information.

The supplementary information for Chapter 3 includes:

- Text S1
- Text S2
- References specific to the supplementary material for Chapter 3
- Figure S1
- Figure S2
- Figure S3
- Table S1
- Table S2
- Table S3

American Naturalist

Supplementary Material: "Nutrigonometry I: using right-angle triangles to quantify nutritional trade-offs in performance landscapes"

Authors: Juliano Morimoto[1,2,3*], Pedro Conceição[3], Christen Mirth[4], Mathieu Lihoreau[5]

Authors' affiliations:

[1] School of Biological Sciences, University of Aberdeen, Zoology Building, Tillydrone Ave, Aberdeen AB24 2TZ

[2]Programa de Pós-graduação em Ecologia e Conservação, Universidade Federal do Paraná, Curitiba, 82590-300, Brazil

[3] Institute of Mathematics, University of Aberdeen, King's College, Aberdeen AB24 3FX

[4] School of Biological Sciences, Monash, University, Melbourne, Victoria, Australia

[5] Research Center on Animal Cognition (CRCA), Center for Integrative Biology (CBI); CNRS, University Paul Sabatier – Toulouse III, France.

* Correspondence:

Dr Juliano Morimoto

juliano.morimoto@abdn.ac.uk

Running-title: Trigonometry to measure fitness trade-offs

Keywords: Nutritional Geometry, nutritional trade-off, performance landscapes

Text S1 – What is Persistence Homology (PH)? A brief introduction

The text below was adapted from (Conceição & Morimoto, 2022) where more details and applications are described.

Topological data analysis and persistence homology

Topological data analysis (TDA) is a relative new subject that lies at the intersection of algebraic topology, data science, statistics and computer science (Chazal & Michel, 2021). Algebraic topology is the part of mathematics that studies (topological) spaces by associating algebraic objects to it that do not change under 'well-behaved' deformations called *homotopies*, e.g., affine transformations – scaling, rotations, translations – and more generally 'squeezing and stretching the space without tearing it apart'. Such algebraic objects that are invariants to these deformations are commonly called *topological invariants*. In a way, one could encapsulate the goal of the field of TDA with the common slogan: *data has shape* and TDA aims to measure it.

One of the branches of TDA called *persistence homology* (PH) has provided new insights and methods to understand higher dimensional data and extra qualitative features of data using techniques of algebraic topology. In the last years, PH has been extensively studied and, by now, is supported by a solid theoretical framework to study complex data structures. Moreover, PH application is supported by a range of well-established softwares designed for PH analysis pipeline, making PH more accessible to different scientific communities. The aim of PH is to extract qualitative information of the data via the computation of a topological invariant called *homology*. For the purpose of the Nutrigonometry, it is enough to think of homology as an algebraic tool that records the number of holes on each dimension of a given space. However, what is a n-dimensional hole? A 0-dimensional hole is defined to be a connected component (e.g., points), a 1-dimensional hole is a cycle/loop that are formed when connected components are linked together (for instance, a circumference has a 1-dimensional hole), a 2-dimensional hole is the void space enclosed by a surface (a sphere, i.e., the boundary of a ball, has a 2-dimensional hole) and so on. Note that holes in a space are invariants under homotopies, in other words, a hole cannot be created or erased in homotopical deformation. This is important because it shows that the topological structure of the data and the information obtained from these n-dimensional holes are invariant. As a result, homology is then an algebraic structure associated to the input data robust under small perturbations (deformations).

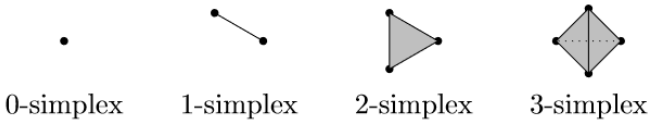
The overall structure of a PH pipeline is as following:

We move from the original cloud of datapoints, upon which we implement a filtering function that allows us to extract the n-dimensional holes described above (a). We then display the information of the n-dimensional holes with standard plots (b) from which inferences on the topological structure of the data can be made (c). For the purpose of this paper, we will focus on providing a user-guide explanation on how to read and interpret persistence diagrams. For a detailed survey and introduction on PH, refer to (Otter, Porter, Tillmann, Grindrod, &

Harrington, 2017) and (Ghrist, 2008). For an introduction on algebraic topology and precise mathematical definitions of the concepts defined here, consult (Ghrist, 2014; Hatcher, 2002).

(a)From point cloud data to a realized space: filtered simplicial complex

Topological spaces are generalizations of geometric objects and are scattered all around – from Euclidean space to fractals. It can be hard to compute the homology of a general space. However, for most of the purpose of PH and most of real-life applications the realized space is a topological space called *simplicial complex.* A simplicial complex is built out of pieces called simplices which occur in any dimension *n*, where *n* is a non-negative integer: the 0-simplices are points, the 1-simplices are edges, the 2-simplices are triangles, the 3-simplices are tetrahedrons and so on. More precisely, a n-simplex represent a convex hull of n+1 points in the n dimensional Euclidean space that are affinely independent, that is, they are not all on the same n-1 dimensional hyperplane.



A simplex is determined by its vertex and has special subsets called *boundary faces,* which are simplices of dimension one below their own. For instance, a 1-simplex has two 0-simplices as boundary faces, a 2-simplex has three 1-simplices as boundary faces and, more generally, a n-simplex has n+1 (n-1)-simplices as boundary faces. Boundary faces are important because it allows us to build a general simplicial complex out of simplices by gluing them together according to one single rule: two simplices of any dimension can be glued along a common boundary faces of the same dimension. This surprisingly naive definition has led to important developments in mathematics.

The first step of PH is to construct a simplicial complex out of the point cloud data. One of the most natural ways to do it is via the *Vietoris-Rips complex*. Recall that our data is embedded in the Euclidean space and, thus, it allows us to talk about (Euclidean) distance between points. Let ε be greater or equal than 0. The Vietoris-Rips complex for ε is the simplicial complex whose k-simplices are the k+1 data point that are pairwise ε distant. A common way to visualize it is performing the following visual experiment: at each data point we draw a ball of diameter ε, if k+1 balls intersect there is a k-simplex there. For instance, when ε is very small no ball intersect any other ball and the associated Vietoris-Rips complex is a discrete set of point (the data point themselves), whereas for very large ε the Vietoris-Rips complex is a unique n-simplex, since the balls centered at each of the n+1 data points intersect. In other words, we construct a filtration by varying the value of the scale parameter ε: it starts with ε equals to zero until there is no more change in the simplicial complex, that is, the Vietoris-Rips complex becomes a simplex of a given dimension. More precisely, denote $VR = (VR_i)_n$ a sequence of Vietoris-Rips complexes associated to data set for an increasing sequence of scale parameter ε_i, and build a sequence of inclusion of topological spaces

$$VR_0 \to VR_1 \to \ldots \to VR_n.$$

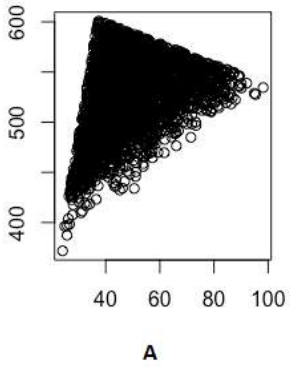
(b)Persistence diagrams: a topological signature of the data

The next step is to compute the homology of the space $VR = (VR_i)_n$ at each step of the filtration according to scale parameter ε. The name persistence homology comes from the fact that we observe which parts of the homology (for each possible dimension) *persists* as the scale parameter ε_i increases. There are two standard and equivalent ways of visualising the homological calculation: via persistence *diagrams* or via *barcodes*. One can think of a persistence diagram/barcode as a representation of a topological signature of the data. For sake of completeness, we present both.

A persistence diagram is a two-dimension plot where each point there represents an apparition of the topological feature in question (e.g., a hole) in the data set. The value horizontal coordinate (x-axis) indicates the **birth** time of a hole and the vertical coordinate (y-axis) is the **death** time of it according to ε_i. Each hole always dies after its birth, and thus, all the points in the persistence diagram must lie above the diagonal line. Note that points close to the diagonal line are features (holes) with a really short 'lifespan'.

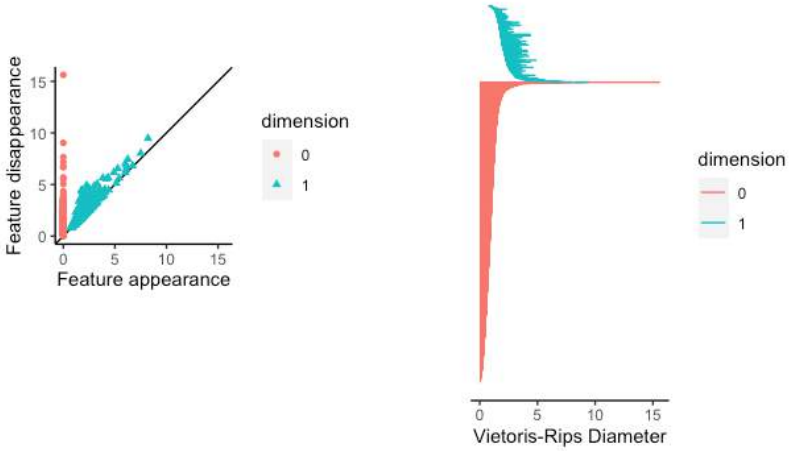
An alternative way to interpret the same information is through the barcode plot. Each bar represents an apparition of a hole in the dataset, with the beginning of the bar showing its **birth** and the end of the bar its **death** as the scale parameter varies in the filtration VR. Short bars in the barcode representation are equivalent to points that lie in the diagonal line in the persistence diagram, whereas long bars are point far from the diagonal line.

As an example, consider the point cloud data **A** below



A

The persistence diagram (on the left) and the barcode representation (on the right) of **A** are given below



Note that all the red points (i.e., connected components or 0-dimension simplex) are aggregated and have a really short lifespan as the scale parameter grows. Moreover, the blue points (i.e., cycles or 1-dimension hole) also have a relative short life space. *This usually indicates that all the data points lie in a unique cluster with points really close to each other and relatively homogenous, since there is no space to form a cycle within a persistent enough interval with respect to the scale parameter.* This is the property used to measure homogeneity in peak (or valley) region prediction in using the Nutrigonometry (see Main Text).

(c)Interpretation

As a direct and easy consequence, one is able to spot when two data sets are different by looking at their persistence diagrams/barcodes. Moreover, a persistence diagram gives *a global analysis of the data*. Higher points in persistence diagrams (long bars in barcodes representation) correspond to more persistent features of the data and potentially more informative, whereas points close to the diagonal (point with short life-span, short bars in barcodes representation) are regarded as noise or small perturbation. However, recent studies have shown that such short-life points may contain information about *the local geometry properties* of data (Adams & Moy, 2021) and, therefore, their significance will depend on the problem/data in consideration. One current research topic is to obtain a detailed statistical inference out of persistence diagrams/barcodes, as outlined in (Otter et al., 2017) and there are still challenges and open questions to be answered. Finally, PH has been used to machine learning applications as a way of providing feature vectors and enhancing machine learning algorithms.

References cited

Adams, H., & Moy, M. (2021). Topology Applied to Machine Learning: From Global to Local. *Frontiers in Artificial Intelligence*, *4*, 54.

Arganda, S., Bouchebti, S., Bazazi, S., Le Hesran, S., Puga, C., Latil, G., … Dussutour, A. (2017). Parsing the life-shortening effects of dietary protein: effects of individual amino acids. *Proceedings of the Royal Society B: Biological Sciences*, *284*(1846), 20162052.

Chazal, F., & Michel, B. (2021). An introduction to topological data analysis: fundamental and practical aspects for data scientists. *Frontiers in Artificial Intelligence*, *4*.

Conceição, P., & Morimoto, J. (2022). 'Holey'niche! Finding holes in niche hypervolumes using persistence homology. *Journal of Mathematical Biology*, *84*(58), 1–12.

Ghrist, R. (2008). Barcodes: the persistent topology of data. *Bulletin of the American Mathematical Society*, *45*(1), 61–75.

Ghrist, R. (2014). *Elementary applied topology* (Vol. 1). Createspace Seattle, WA.
Gray, L. J., Simpson, S. J., & Polak, M. (2018). Fruit flies may face a nutrient-dependent life-history trade-off between secondary sexual trait quality, survival and developmental rate. *Journal of Insect Physiology*, *104*, 60–70. doi: https://doi.org/10.1016/j.jinsphys.2017.11.010
Hatcher, A. (2002). Algebraic topology, Cambridge Univ. *Press, Cambridge*.
Jang, T., & Lee, K. P. (2018). Comparing the impacts of macronutrients on life-history traits in larval and adult Drosophila melanogaster: The use of nutritional geometry and chemically defined diets. *Journal of Experimental Biology*, *221*(21). doi: 10.1242/jeb.181115
Kutz, T. C., Sgrò, C. M., & Mirth, C. K. (2019). Interacting with change: Diet mediates how larvae respond to their thermal environment. *Functional Ecology*, *33*(10), 1940–1951.
Otter, N., Porter, M. A., Tillmann, U., Grindrod, P., & Harrington, H. A. (2017). A roadmap for the computation of persistent homology. *EPJ Data Science*, *6*, 1–38.
Rodrigues, M. A., Martins, N. E., Balancé, L. F., Broom, L. N., Dias, A. J. S., Fernandes, A. S. D., … Mirth, C. K. (2015). Drosophila melanogaster larvae make nutritional choices that minimize developmental time. *Journal of Insect Physiology*, *81*, 69–80. doi: 10.1016/j.jinsphys.2015.07.002
Runagall-McNaull, A., Bonduriansky, R., & Crean, A. J. (2015). Dietary protein and lifespan across the metamorphic boundary: protein-restricted larvae develop into short-lived adults. *Scientific Reports*, *5*(1), 11783. doi: 10.1038/srep11783
Simpson, S. J., & Raubenheimer, D. (1993). A multi-level analysis of feeding behaviour: the geometry of nutritional decisions. *Philosophical Transactions of the Royal Society B-Biological Sciences*, *342*(1302), 381–402.

Supplementary Figures

In GF studies, data can be divided into two structures: intake data and fixed ratio data. Intake data is ideal in GF studies because it allows for exploration of *realized* nutritional effects, that is, nutritional effects exerted upon traits given by the amount of nutrient eaten (Simpson & Raubenheimer, 1993). However, collecting intake data can be cumbersome or challenging, and recent approaches have adapted GF experiments to draw landscapes of traits based on the fixed ratio of the nutrients in the diets (Kutz, Sgrò, & Mirth, 2019). This approach has been successful in generating important biological insights. For instance, fixed ratio diets were used to show that high-energy, high-sugar larval diets minimize fluctuating asymmetry in adult male sex combs but at a cost of reduced growth and survival in *D. melanogaster* (Gray, Simpson, & Polak, 2018). Similarly, fixed ratio diets have shown that larvae dietary choices minimize development time (Rodrigues et al., 2015) and, in combination with chemically-defined diet formulations, fixed ratio diets revealed that each fitness-related trait such as survival, growth, reproduction, and reproductive rate have different optimal diets (Jang & Lee, 2018). Moreover, fixed ratio diets were used to show that larvae fed on protein-restricted diets resulted in adults with shorter lifespan in the neriid fly *Teleostylinus angusticollis,* suggesting a carry-over effect from larval nutrition to adult life history (Runagall-McNaull, Bonduriansky, & Crean, 2015). Despite this, however, we still do not know how the adaptation of using fixed ratio diets influence estimates of nutritional trade-offs in multidimensional performance landscapes.

Nutrigonometry allowed us to test whether the structure of the data is important for model predictions. We used Lee et al., (2008) dataset with individual intakes ('intakes') as well as with fixed ratios ('fixed') (see Main Text). We also used a second dataset from (Kutz et al., 2019), which studied how temperature modulates nutritional responses of larval development and adult fitness in *D. melanogaster*, as an additional auxiliary example of our model in fixed ratio datasets (see Fig S1 below), as we anticipate this experimental design to become progressively more common in the literature given the use of chemically-defined diets and increasing nutrient axis being analysed simultaneously [e.g., (Arganda et al., 2017)].

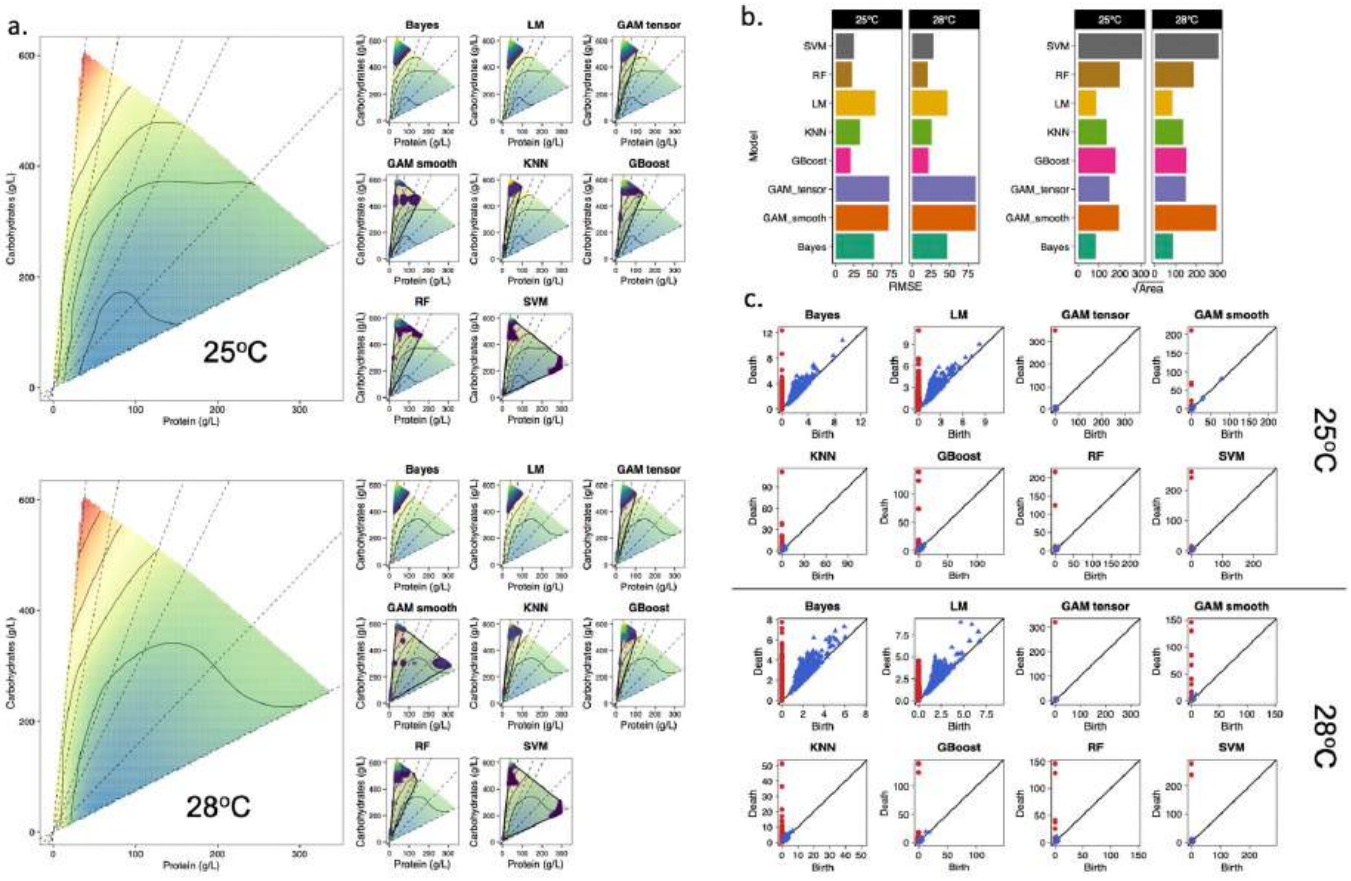


Figure S1. Model application to fixed ratio data. (a) 3D landscape for developmental time at 25ºC (top left) and 28ºC (bottom left) (from Kutz et al., 2019) with the overlaid predicted peak regions. For the landscape, red represents peaks while light green represents valleys. For

the predicted region, dark blue points represent points with lower predicted z-values whereas bright yellow represents points with higher predicted z-values. The shaded polygon was added to facilitate visualisation of the predicted peak region and the homogeneity of points within the predicted peak. (b) RMSE and predicted peak area (i.e., area of the shaded polygon in panel *a*) for the models of developmental time at 25ºC (top right) and 28ºC (bottom right) values of each model. Note that models with high RMSE can still be the best predictors of peak region. (c) Persistence homology (PH) plots for the topological analysis of the predicted peak region of the 3D landscape for developmental time at 25ºC (top panel) and 28ºC (bottom panel) (from Kutz et al., 2019). x and y- axes represent birth and death, respectively, of topological structures. The diagonal line represents the line in which the birth and death co-occur. Homogenous predicted peaks have red (dimension 0) and blue (dimension 1) points that are closer, as opposed to more heterogeneous predicted peaks upon which points are farther from each other.

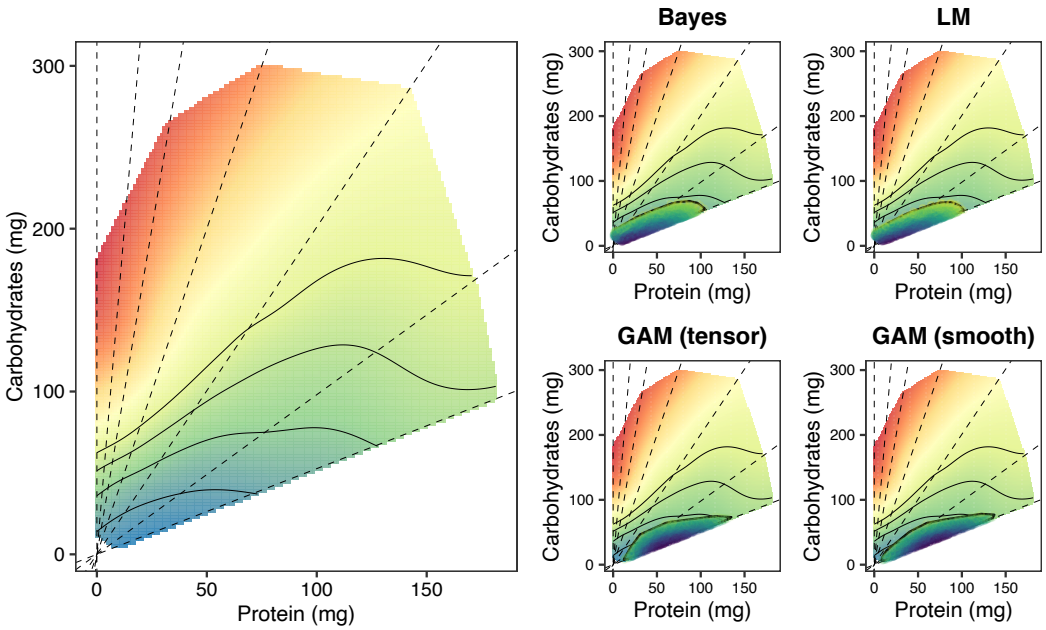


Figure S2. Prediction of the valley regions for lifespan using individual intake data from Lee et al., 2008. Note that we used the best performing models for the peak region (see Main text).

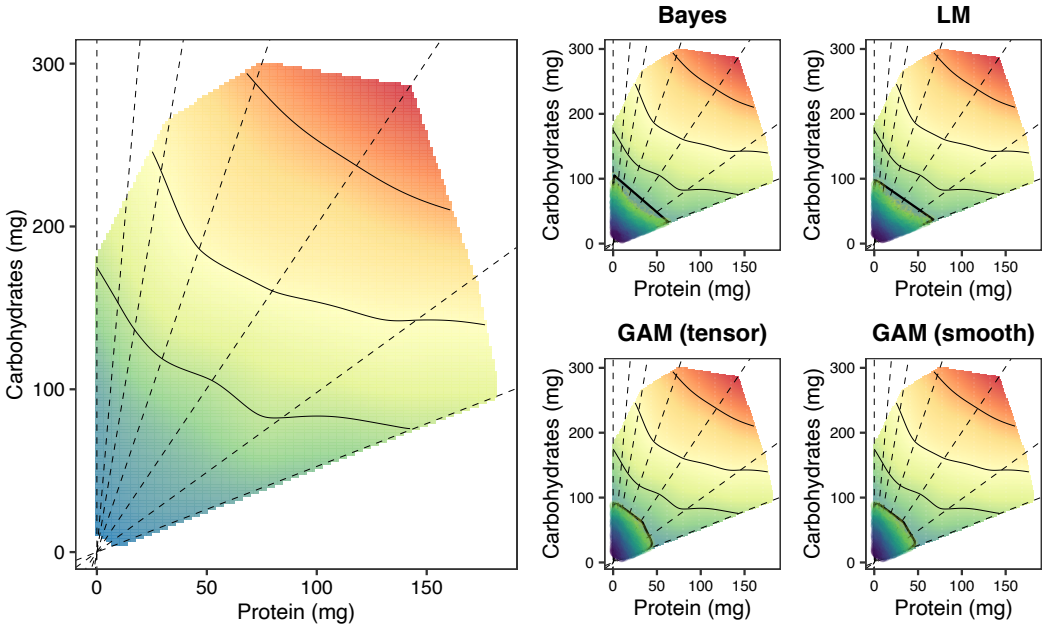


Fig S3. Prediction of the valley regions for reproductive rate using individual intake data from Lee et al. (2008). Note that we used the best performing models for the peak region (see Main text).

Supplementary Tables

Table S1. Area of the predicted peak region for all models. All values are given in unit squared of nutrient intake or diet composition (for fixed ratios).

Data	Trait	Model	Mean	Upr CI	Lwr CI
Peak Area (intakes)	Lifespan	GAM_smooth	1912.72	1922.91	1902.53
		GAM_tensor	1912.84	1923.67	1902.01
		Bayes	1954.67	1967.07	1942.26
		LM	2069.88	2081.61	2058.14
		KNN	3126.88	3145.94	3107.83
		SVM	4800.91	4824.54	4777.28
		RF	5978.74	6024.63	5932.85
		GBoost	10919.99	11148.43	10691.55
	Reproductive rate	LM	1932.41	1944.72	1920.10
		GAM_smooth	2022.98	2037.21	2008.76
		GAM_tensor	2032.59	2046.05	2019.13
		Bayes	2041.78	2055.65	2027.91
		KNN	6181.58	6216.59	6146.57
		RF	11001.14	11158.15	10844.12
		SVM	14614.46	14674.16	14554.76
		GBoost	16358.27	16555.54	16161.00
Peak Area (fixed)	Lifespan	LM	2201.23	2213.00	2189.47
		GAM_tensor	2208.55	2220.72	2196.38
		Bayes	2292.38	2310.97	2273.80
		KNN	2505.34	2570.34	2440.33
		RF	4909.45	4954.50	4864.39
		SVM	5848.48	5888.34	5808.61
		GAM_smooth	6532.01	6587.10	6476.93
		GBoost	9156.89	9302.30	9011.47
	Reproductive rate	GAM_smooth	2086.55	2098.31	2074.78
		SVM	2101.93	2113.36	2090.50
		GAM_tensor	2102.28	2112.33	2092.24
		GBoost	2145.32	2156.60	2134.04
		LM	2162.05	2171.65	2152.45
		Bayes	2211.30	2224.25	2198.35
		RF	2994.72	3025.90	2963.55
		KNN	4357.13	4377.04	4337.22

Table S2. Nutrient spread of the predicted peak region for all models. All values are given in units of nutrient intake or diet composition (for fixed ratios).

Data	Trait	Model	Protein (median)	Protein (std)	Carb (median)	Carb (std)
Nutrient spread (intakes)	Lifespan	LM	0.5333	0.0858	0.1461	0.0194
		Bayes	0.5354	0.0963	0.1446	0.0155
		GAM_tensor	0.5576	0.0772	0.1504	0.0218
		GAM_smooth	0.5598	0.0752	0.1566	0.0208
		RF	0.5884	0.1184	0.2530	0.0342
		KNN	0.6239	0.0863	0.1920	0.0253
		SVM	0.6629	0.1153	0.2202	0.0245
		GBoost	0.6721	0.1259	0.2723	0.0339
	Reproductive rate	LM	0.1348	0.0214	0.0513	0.0088
		Bayes	0.1360	0.0207	0.0512	0.0083
		GBoost	0.2006	0.0337	0.1893	0.0412
		GAM_smooth	0.2152	0.0260	0.0303	0.0049
		GAM_tensor	0.2172	0.0257	0.0290	0.0051
		RF	0.2255	0.0379	0.1280	0.0317
		KNN	0.2768	0.0287	0.1548	0.0140
		SVM	0.3541	0.0342	0.2288	0.0240
Nutrient spread (fixed)	Lifespan	GAM_tensor	0.4752	0.0980	0.0731	0.0162
		RF	0.5004	0.2436	0.0812	0.0261
		KNN	0.5376	0.0487	0.1072	0.0505
		SVM	0.6408	0.2370	0.0795	0.0461
		Bayes	0.6416	0.1078	0.0866	0.0146
		LM	0.6502	0.1075	0.0910	0.0159
		GBoost	0.6599	0.2394	0.1313	0.0454
		GAM_smooth	0.8015	0.1148	0.1080	0.0595
	Reproductive rate	GBoost	0.0807	0.0662	0.0383	0.0143
		GAM_tensor	0.1413	0.0554	0.0229	0.0088
		KNN	0.1538	0.1273	0.0202	0.0214
		Bayes	0.1610	0.0897	0.0228	0.0128
		LM	0.1620	0.0902	0.0206	0.0113
		RF	0.2238	0.0768	0.0231	0.0122
		SVM	0.2804	0.0765	0.0333	0.0106
		GAM_smooth	0.4162	0.0293	0.0444	0.0048

Table S3. Quantification of nutritional trade-offs in developmental time between two developmental temperatures. Estimates of $\theta_{i,j}$ (in degrees) and $h_{i,j}$ (in g/L). Analysis from the data presented in Kutz et al., 2019. Confidence intervals overlapping zero implies no difference in the peaks. Magnitude of the estimates indicate the strength of nutritional trade-offs (i.e., larger magnitudes indicate stronger nutritional trade-offs). Note that $\theta_{i,j}$ is bound between 0 and 90 degrees (i.e., 0 and $\frac{\pi}{2}$).

Parameter	Model	Mean	Sd	lwrCI	uprCI
$\theta_{i,j}$	LM	-0.0228	2.1983	-4.332	4.286
	Bayes	-0.0707	2.2355	-4.452	4.311
	RF	0.74892	3.1543	-5.434	6.932
	Gboost	-11.541	15.598	-42.11	19.03
	SVM	-0.0604	30.284	-59.42	59.30
	KNN	-0.0084	2.2102	-4.341	4.324
$h_{i,j}$	LM	2.0297	51.035	-98.00	102.06
	Bayes	0.3822	50.295	-98.20	98.971
	RF	17.283	125.803	-229.31	263.88
	Gboost	102.95	179.903	-249.68	455.60
	SVM	0.6422	124.636	-243.67	244.95
	KNN	5.8109	175.197	-337.61	349.23

Chapter 4

Nutrigonometry II: Experimental strategies to maximize nutritional information in multidimensional performance landscapes

Publication reference

Morimoto, J., (2022). Nutrigonometry II: Experimental strategies to maximize nutritional information in multidimensional performance landscapes. *Ecology and Evolution*, 12(8), p.e9174.

4.1 Abstract

Animals regulate their nutrient consumption to maximise the expression of fitness traits with competing nutritional needs ('nutritional trade-offs'). Nutritional trade-offs have been studied using a response surface modelling approach known as the Geometric Framework for nutrition (GF). Current experimental design in GF studies does not explore the entire area of the nutritional space resulting in performance landscapes that may be incomplete. This hampers our ability to understand the properties of the performance landscape (e.g., peak shape) from which meaningful biological insights can be obtained. Here, I tested alternative experimental designs to explore the full range of the performance landscape in GF studies. I compared the performance of the standard GF design strategy with three alternatives: hexagonal, square, and random points grid strategies with respect to their accuracy in reconstructing baseline performance landscapes from a landmark GF dataset. I showed that standard GF design did not reconstruct the properties of baseline performance landscape appropriately particularly for traits that respond strongly to the interaction between nutrients. Moreover, the peak estimates in the reconstructed performance landscape using standard GF design were accurate in terms of the nutrient ratio but incomplete in terms of peak shape. All other grid designs provided more accurate reconstructions of the baseline performance landscape while also providing accurate estimates of nutrient ratio and peak shape. Thus, alternative experimental designs can maximise information from performance landscapes in GF studies, enabling reliable biological insights into nutritional trade-offs and physiological limits within and across species.

Keywords: Nutritional Geometry, trigonometry, lifespan-reproduction trade-off, fitness maps, Drosophila melanogaster.

4.2 Introduction

Animals often balance their diet to maximize life-history traits with diverging nutritional needs (Simpson & Raubenheimer 2012, Raubenheimer & Simpson 2020). This creates the potential for trade-offs in the balance and allocation of nutrients needed for optimum fitness, aka 'nutritional trade-off' (Lee et al. 2008, Maklakov et al. 2008, Morimoto & Lihoreau 2019). Nutritional trade-offs have been described across taxa. For example, nutritional trade-offs between lifespan and reproduction and between immunity and reproduction have been described in *Drosophila melanogaster* (Lee et al. 2008, Ponton et al. 2015), tephritid fruit flies and neriid flies (Fanson, Yap & Taylor 2012, Fanson & Taylor 2012, Adler et al. 2013, Pascacio-Villafán et al. 2022), crickets (Maklakov et al. 2008, Rapkin et al. 2018, Treidel et al. 2021, Hawkes et al. 2022, Guo et al. 2022), and mice (Solon-Biet et al. 2014) (see also reviews by Ponton et al. 2011, Schwenke et al. 2016). Nutritional trade-offs have also been described between reproductive traits in *D. melanogaster* (Morimoto & Wigby 2016) and neriid flies (Sentinella et al. 2013), cockroaches (Bunning et al. 2015), crickets (Ng et al. 2018), and butterflies (Gage & Cook 1994). Thus, nutritional trade-offs appear to be ubiquitous.

A method known as the Geometric Framework of Nutrition (GF) has emerged as a powerful unifying framework capable of disentangling the multidimensional effects of nutrients (both ratios and concentrations) on life-history traits and fitness (Simpson & Raubenheimer 1993, Raubenheimer & Simpson 1993), thereby enabling accurate estimates of nutritional trade-offs. The GF framework has been used across taxa and became a cornerstone design for advancing our understanding of complex physiological and behavioural responses to nutrition, including human health (Raubenheimer et al. 2009, Simpson & Raubenheimer 2012, Simpson et al. 2017). In essence, GF is an application of the response surface modeling (RSM) approach (Box & Wilson 1951), where a n-dimensional Euclidean space is used to investigate the response of the animal to the dietary intake of various ratios and concentrations of n nutrients. The resulting $n + 1$ surface (known as 'performance landscape') maps the level of the chosen trait across the different dietary ratios and concentrations. However, contrary to standard applications of RSM, GF is not only interested in optimization (i.e., finding the 'peak' in the performance landscape). This is because the entire landscape contains valuable biological information about diet-dependent expression of traits and thus, are meaningful to biologists and ecologists. For example, both peaks and valleys can be important indicators of the overall nutritional responses and comparisons between the positions of these properties within a performance landscape can be useful to determine the degree of changes in life-histories with

small dietary changes as well as quantifying obligate nutritional trade-offs between traits (Rapkin et al. 2018, Morimoto & Lihoreau 2019, Kutz et al. 2019, Morimoto et al. 2023). However, common design of experiments used in RSM such as full factorial or fractional designs and central composite designs (CCD) (Myers et al. 2016) are not necessarily sufficient or efficient to reveal the characteristics of the entire performance landscapes (Ruohonen et al. 2001). Therefore, an optimum GF experimental design is a trade-off between the number of diets and replicates per diet to maximise resolution of the performance landscape and the costs and feasibility risks associated with geometrically increasing sample sizes.

Traditionally, GF studies have been of two types: those which measure individual diet intake as in (Lee et al. 2008, Maklakov et al. 2008, Fanson, Yap & Taylor 2012) and those that provide diets with fixed ratios and do not measure intake as in (Kutz et al. 2019, Alton et al. 2020). Both of these types share the fundamental GF design of experiments, which is as follows: (i) the standard design of experiments in GF studies divides the nutritional space (i.e., Cartesian plane with nutrients as axes) into several 'nutritional rails', which are diets with fixed nutrient ratios (Figure 4.1a), (ii) each nutritional rail is subdivided into different diet concentrations, and (iii) each combination of diet ratio and concentration (red dots in Figure 4.1a) are the 'dietary treatments,' which are given to replicate animals or groups of animals, from which the measure of the traits is taken (Simpson & Raubenheimer 2012); here I will refer to the dietary treatment points as 'anchor points'.

The difference between the two types of studies using GF is that in one type, experimenters measure individuals' (or groups') food intakes ('intake data'), whereas in the other type, individuals are given a fixed ratio of the diet without measurements of food intake ('fixed ratio data'). The anchor points (diets) are the points that contain data for the performance traits and therefore act as data-driven points (or 'anchors') for the reconstruction of the performance landscape, which is commonly done using thin-plate spline interpolation (see e.g., Ponton et al. 2015, Morimoto & Lihoreau 2019). Anchor points are directly used for interpolation in the fixed ratio data but only work as guidelines for the experimental design for intake data, since the interpolation is done using the final nutrient intake of each individual in each diet. The performance landscape has depth determined by the variance in food intake (for intake data) or the range of diet concentrations (for fixed ratio data) (Figure 4.1a).

Importantly though, both types of GF approaches are insufficient to generate anchor points that cover the entire area of the nutrient space, requiring interpolation while making performance landscapes incomplete. As a result, a large area of the nutrient space remains unexplored or in need of extrapolation for areas without anchor points (Figure 4.1a). While this may not necessarily affect our approximations

of the regions in which peaks and valleys are located, it certainly precludes us from extracting meaningful biological information across the entire domain of the nutritional space of animals. For instance, by limiting the range of the nutrient space that is explored, GF makes an underlying assumption about the *a priori* knowledge of the physiological limits that a species has or evolved in terms of diet, although this information is rarely known.

More recent GF studies have used ecological and fieldwork data to design GF diets that are ecologically relevant and guide experimental design, incorporating not only natural dietary information from natural populations (see e.g., Rothman et al. 2011, Wilder et al. 2013, Vaudo et al. 2016) but also genetics (Deans et al. 2016), environmental stability (Lawton et al. 2021), and land use (Le Gall et al. 2020). Despite this, GF studies are still inductive and do not explore the full range of the nutrient space. To date, there has been no systematic investigation as to how the standard GF experimental design can influence the resolution of the reconstructed performance landscapes, nor whether alternative experimental designs could provide more complete estimates of performance landscapes across the entire nutritional space.

Here, I investigated the performance of different sampling strategies when reconstructing performance landscapes, using a landmark dataset in the field of nutritional ecology (Lee et al. 2008). I used the pioneering Nutrigonometry framework to identify and compare the peaks in the reconstructed performance landscapes and how congruent these estimates are across sampling strategies (Morimoto et al. 2023). I tested four different sampling strategies: standard GF, hexagonal, squared, and random points sampling grids (Figure 4.1a). As a proof-of-concept, I developed the main arguments using fixed ratio datasets, as this type of GF approach is conceptually easier to explain and allows for the understanding of the foundations of my argument. I then expanded the applications of the argument for GF studies with intake datasets in the discussion section. Overall, this is the first investigation of the foundations of GF experimental design, which can have important long-term implications to the quality of data collected in field of nutritional ecology. Expanding the coverage of performance landscapes will open up possibilities to extract biological information that is currently inaccessible, allowing for more complete studies on the nutritional trade-offs that animals have evolved to circumvent physiological and nutritional constraints.

4.3 Materials and Methods

4.3.1 Terminology and sampling designs

Throughout the text, I used the term 'anchor point' to refer to a diet of given nutrient ratios and 'resolution' as the total number of different diets (anchor points) in an experiment. Anchor points in the performance landscapes were generated in three resolutions: 30, 50, and 250 anchor points (see Figure 4.6 for examples).

4.3.2 Standard GF

The standard GF sampling grid was used with nutritional rails and diet concentrations as in the original dataset (Lee et al. 2008) (Figure 4.1a).

4.3.3 Hexagonal grid

The first alternative sampling strategy was the hexagonal grid (Figure 4.1b). Consider that knowledge (or 'certainty') η_i about the estimate of the performance trait for trait i at the anchor point follows a (symmetric) Gaussian density function such that:

$$\eta_i = C \cdot e^{-\left[\frac{(x-x_0)^2}{2\sigma_X^2} + \frac{(y-y_0)^2}{2\sigma_Y^2}\right]}$$

where C is the amplitude of the distribution (e.g., determined by trait values), x_0 and y_0 are coordinates of the anchor points where the Gaussian is centred, and σ_X^2 and σ_Y^2 correspond to the uncertainty around the anchor point (Figure 4.7a). Note that for the purpose of this argument, I assume a correlation of zero between x and y and thus, a symmetric (circular) Gaussian.

Then, the performance landscape can be seen as an analogous problem of circle packing in geometry, where the hexagonal grid is the densest circle packing in 2D Euclidean space (Chang & Wang 2010) (see Figure 4.7b). In fact, the distance between any two anchor points i and j is equal to $2r$, where r is the apothem of the hexagons containing the anchor points (see Figure 4.7c). I hypothesized that a hexagonal grid with anchor points at the centre of hexagons could maximise performance landscape reconstruction in the nutritional space while minimizing the number of anchor points and replicates

4.3.4 Square grid

The second sampling strategy was the square grid (Figure 4.1b). The underlying rationale for the square grid is similar to that of the hexagonal grid above, where I divided the nutritional space into adjacent squares, with anchor points at the centre of each square. The distance between any two anchor points i and j is equal to $2r$ if the squares lie in the same column or row and $2r\sqrt{2}$ if the anchor points lie in diagonal squares, where r is the apothem of the squares containing the anchor points (Figure 4.7c).

Random Points grid

Lastly, I also investigated the accuracy of randomly probing the nutritional space (Figure 4.1b).

Dataset and Computation

Dataset

I used a landmark dataset that contains *Drosophila melanogaster* individual diet intakes and diet fixed ratios, and the consequences of diet on lifespan and reproduction (Lee et al. 2008). Two nutrients were investigated – protein and carbohydrate – such that performance landscapes have three dimensions. This dataset was previously used in my conceptualization of the Vector of Position approach and Nutrigonometry, having important benchmark status in the field (Morimoto & Lihoreau 2019, Morimoto et al. 2023).

Briefly, the Vector of Positions approach was developed to generate n-dimensional performance landscapes from GF experiments as vectors from which the strength of nutritional trade-offs between traits can be estimated via the angle θ between vectors of two traits (Morimoto & Lihoreau 2019). This approach uses a machine learning model to identify the peak region. More recently, I developed a conceptually simpler and computationally cheaper model to estimate peak regions and nutritional trade-offs in GF studies using trigonometric relationships ('Nutrigonometry'), which enabled the comparison of different statistical methods to estimate nutritional trade-offs and opened up new ways in which properties of performance landscapes can be estimated (Morimoto et al. 2023). The dataset used here was fundamental for the validation of these methods and is therefore used here.

Generating the Baseline Performance Landscape

The baseline performance landscape is the true performance landscape for the response of a trait throughout the nutritional space (Figure 4.1c). In experiments, this true performance landscape is unknown, and

the GF framework aims to approximate a reconstructed performance landscape to the baseline landscape empirically. There are no available datasets in the literature that explore the entire nutritional space, and thus, no true performance landscapes have yet been estimated experimentally.

Consequently, to obtain baseline performance landscapes, I created a high-resolution grid (of 4 units distance between points) that covered the entire nutritional landscape, including regions beyond the original boundaries of the nutritional space sampled by (Lee et al. 2008), and predicted the value of the trait at each of the grid points using a machine learning approach based on the empirical values of the sampled regions obtained in (Lee et al. 2008). This allowed me to create high-resolution (predicted) baseline performance landscapes for lifespan, lifetime eggs, and daily eggs that can be compared with the reconstructed performance landscapes for the same traits using different sampling approaches (see below). The baseline landscapes are shown in Figure 4.1c.

Approach

I simulated a real-world experiment as follows:

- I sampled the anchor points in the nutritional space according to the four grid sampling approaches and three resolutions tested here (see above).
- I used a polynomial regression with the linear and quadratic effects of protein and carbohydrate (and their interactions) fitted to the baseline performance landscape to assign a value of the trait to each of the anchor points. This is equivalent to running an experiment with the anchor points and obtaining the estimates for the performance trait at each anchor point, based on the true performance landscape (which is unknown in real-world problems).
- I reconstructed the performance landscape for each sampling grid and resolution using the thin-plate spline method.
- I applied the Nutrigonometry model to estimate the peak region for each sampling grid and resolution on the reconstructed landscape, which included the calculation of the protein-to-carbohydrate (P:C) ratios of the estimated Nutrigonometry peak.
- I overlaid the identified peak region with the true performance landscape as well as all the estimates of the peak region across all resolutions and grid sampling strategies.

- I calculated the area of the estimated peak by approximating the area to an ellipse with coordinates determined by the peak estimates using Nutrigonometry.

Reconstruction Accuracy

I estimated the accuracy of the reconstruction vs. baseline landscape by generating a 2D profile of the 3D landscape using the predicted value of the landscape as the y-axis and the multiplication of protein and carbohydrate content as the x-axis for all points in the landscape. This approach allowed for dimensionality reduction while providing topological information about the structure of the data. I then binned the x-axis ($n = 100$) and calculated the average and standard deviation of the Euclidean distance between the points from the baseline performance landscape and the reconstructed performance landscape. For this analysis, I used the reconstructed landscapes with the highest resolution (i.e., 250) because of the higher density of anchor points (and hence, expected accuracy) used to generate these performance landscapes.

Software and Packages

All simulations were performed in R version 3.6.2 (Team 2019). The 'tidymodels version 0.1.0', 'stringr version 1.4.0', 'tidyr version 1.1.0', 'purrr version 0.3.4', and 'dplyr version 0.8.5' packages of the tidyverse were used for data wrangling, as well as to generate the baseline performance landscape and manipulate data for visualization (Wickham et al. 2016, 2019). Performance landscapes were reconstructed using the 'Tps' function of the 'fields version 10.3' package with the lambda argument set to 0.05 in all models (Nychka et al. 2017). I also used the 'raster version 3.1-5', 'rgeos version 0.5-3', and 'sp version 1.4-2' packages for data manipulations for visualization and sampling of the nutritional space, the latter being used for the functionalities in the 'spsample' function (Bivand et al. 2017, **?**, Pebesma & Bivand 2005). All plots were created using the 'ggplot2 version 3.3.1' package (Wickham et al. 2016). The 'ggnewscale version 0.4.5' package was used to enhance the data visualization (Campitelli 2020).

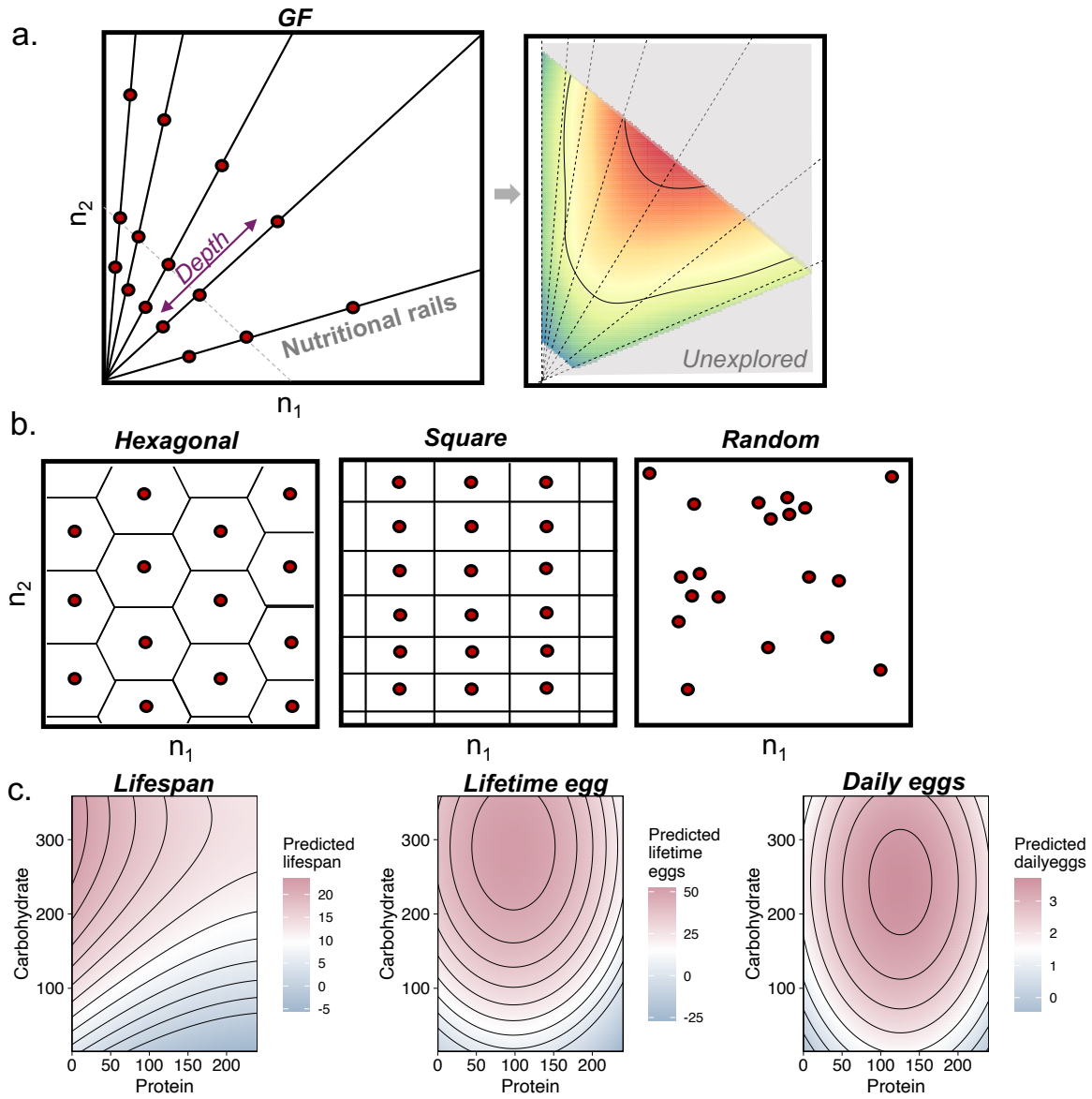


Figure 4.1: **Exploration of the nutrient space using alternative sampling strategies.** (a) The standard experimental design of a GF study (left) and a performance landscape generated form a fixed ratio dataset (reconstructed from Kutz et al. 2019) (right). Note the unexplored region in the nutritional space (shaded area). (b) The three alternative sampling strategies tested here: hexagonal, square, and random points grids. Red dots indicate anchor points (see Main Text). (c) The baseline performance landscapes for lifespan, lifetime egg production and daily egg production. These landscapes were generated with the purpose of acting as the true performance landscape of the trait, which are unknown in GF experiments. These baselines landscapes are the standard upon which the reconstructed landscapes with alternative methods were compared against the GF in this study (see Methods section).

Results

Standard GF sampling strategy finds the correct ratio of nutrients but often with inaccurate peak shape estimates

The data showed that all sampling strategies provide reasonably accurate estimates of the ratio in which the peak in the performance landscape is found (Figure 4.2). Note that estimates of peak P:C ratio were

more variable for lifespan (log10-transformed in Figure 4.2 to aid visualization) because the peak lies near the boundary of the performance landscape (i.e., P:C ∼0:1). Nonetheless, the visualization of the predicted peak region shows that all methods find peak regions in the correct area of the performance landscape (Figure 4.3a).

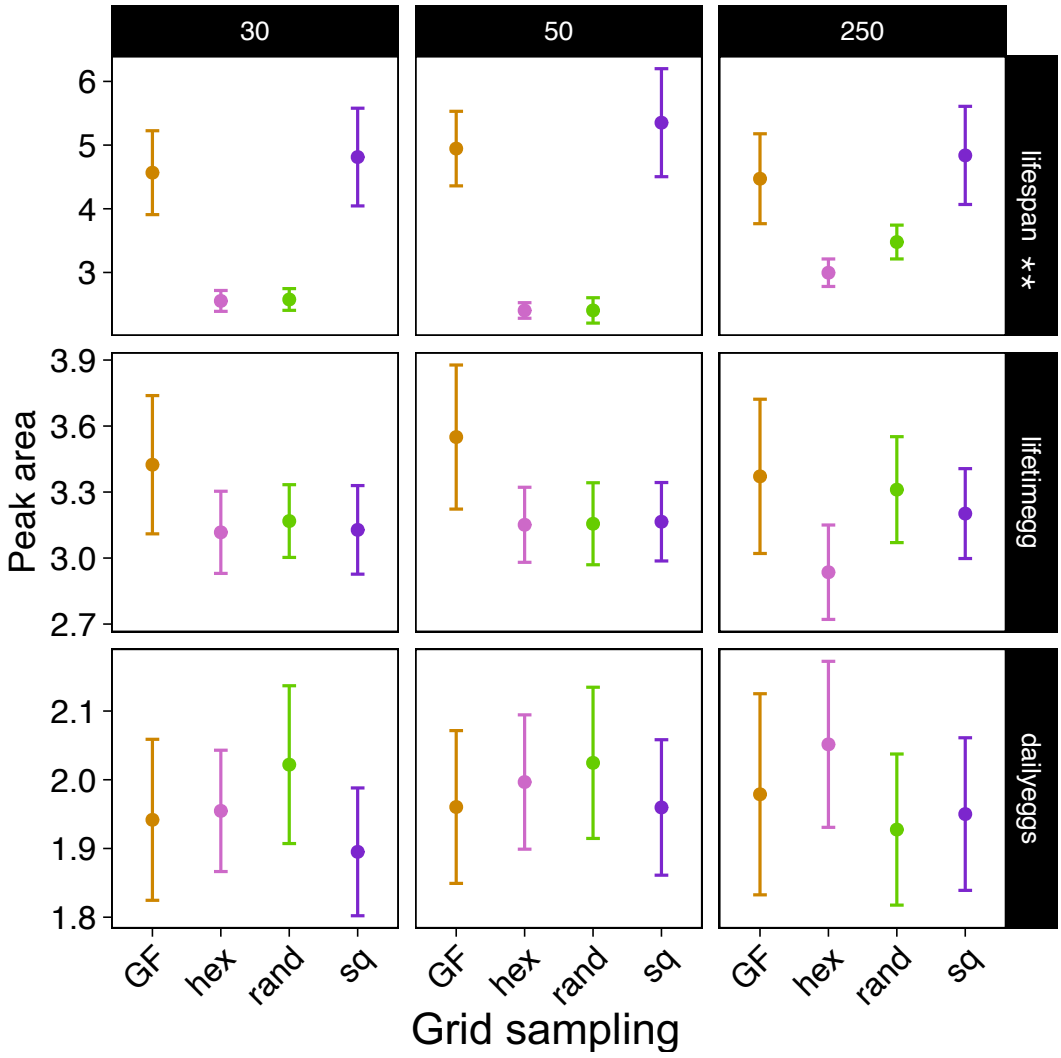


Figure 4.2: **P:C ratios of the estimated peak in the reconstructed performance landscape across the grid sampling strategies.** Note that the y-axis of the lifespan plots was log-transformed to aid data visualisation (see also Figure 4.3a). Such differences in scale for lifespan emerged from the fact that the peak lies near the boundary of the nutritional landscape, in a region of P:C 0:1. Hex: hexagonal sampling; rand: random points sampling; sq: square sampling.

Despite this, striking differences between standard GF and other sampling strategies were found in the shape of the predicted peak. For lifespan, where the peak in the performance landscape lies near the boundary of the nutritional space, the predictions of all sampling strategies were similar in shape (Figure 4.3a-b). Conversely, the shapes of the GF peaks for lifetime egg production and daily eggs, which are in the middle of the performance landscape (indicating that the trait responds to the interaction between

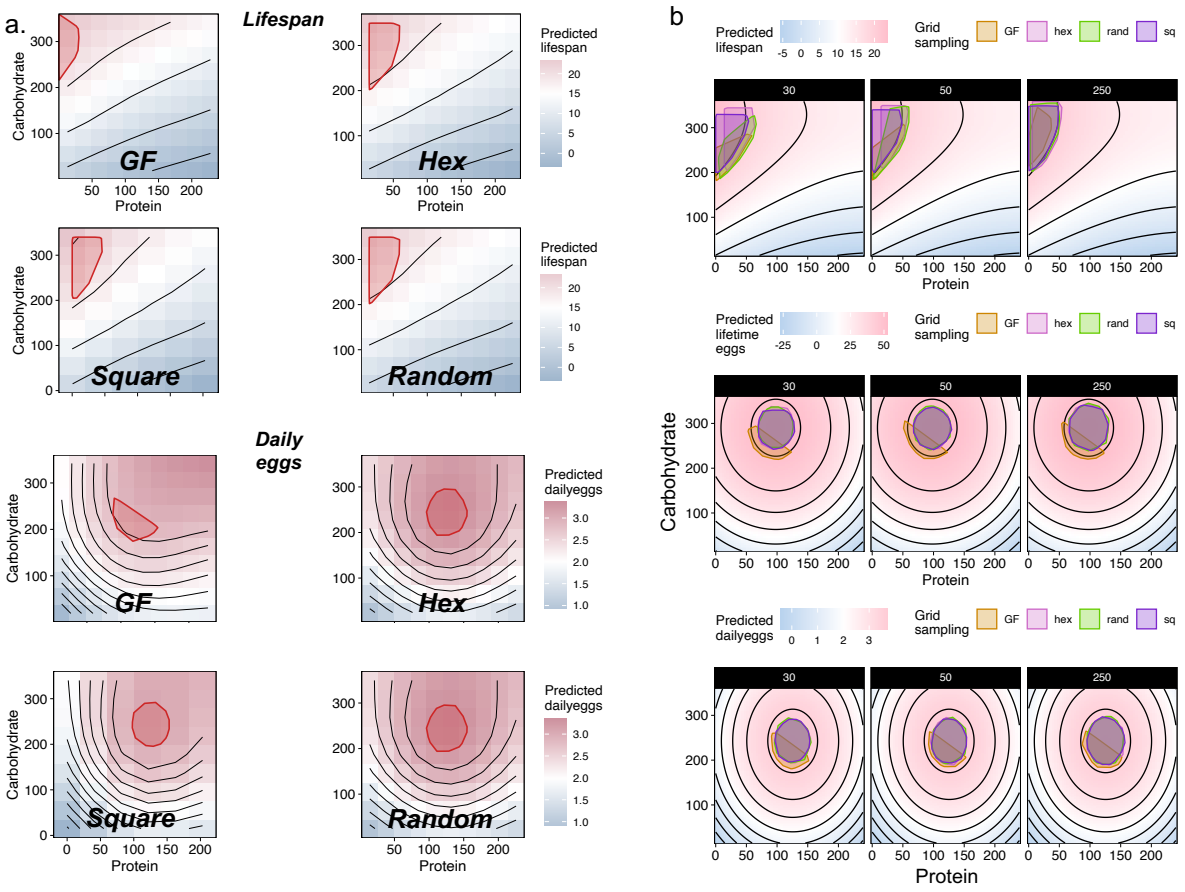


Figure 4.3: **Predicted peak region and shape across sampling strategies.** (b) Predicted peak in the performance landscape of lifespan (top) and daily eggs (bottom) (see also Figure 4.6a for lifetime egg peak predictions). Performance landscapes reconstructed from resolution equal to 50. (b) Overlaid peak predictions mapped onto the baseline performance landscapes of lifespan, lifetime eggs, and daily eggs across the sampling strategies. Note that GF sampling (orange) generates incomplete peak shape predictions for traits that respond to the interaction of nutrients. Hex: hexagonal sampling; rand: random points sampling; sq: square sampling.

protein and carbohydrate), differed substantially from those of other sampling strategies: standard GF peaks are wider and semi-circular, while all other sampling strategies find a defined circular peak covering the appropriate region of the baseline performance landscape (Figure 4.3a-b and Figure 4.6d).

The overlaid visualization of the peak estimates and the underlying baseline performance landscape clearly showed that estimates of the peak region from GF sampling were incapable of reflecting the true peak region of the baseline performance landscape relative to the other methods (Figure 4.3b). As a result, these differences are also reflected in the peak area estimates, where the peak area using standard GF was smaller relative to other sampling strategies (Figure 4.4). In other words, standard GF can only provide a partial estimate of peak area, especially for traits that are affected by the interaction between nutrients.

Reconstructed performance landscapes from standard GF sampling are more inaccurate in regions that capture responses to nutrient interactions

The topological profile of the reconstructed landscapes showed that standard GF sampling generates reconstructed performance landscapes that are more dissimilar (measured as the Euclidean distance) to the true baseline performance landscape in regions that capture the interaction of nutrients on the performance trait (e.g., high protein and high carbohydrate values) (Figure 4.4b). Importantly, the inaccuracy is less accentuated for traits that have peaks near the boundary of the nutritional space (i.e., lifespan), but progressively more pronounced for traits with peaks in the middle of the nutrient space, which indicates strong responses to the interaction of nutrients (Figure 4.4b).

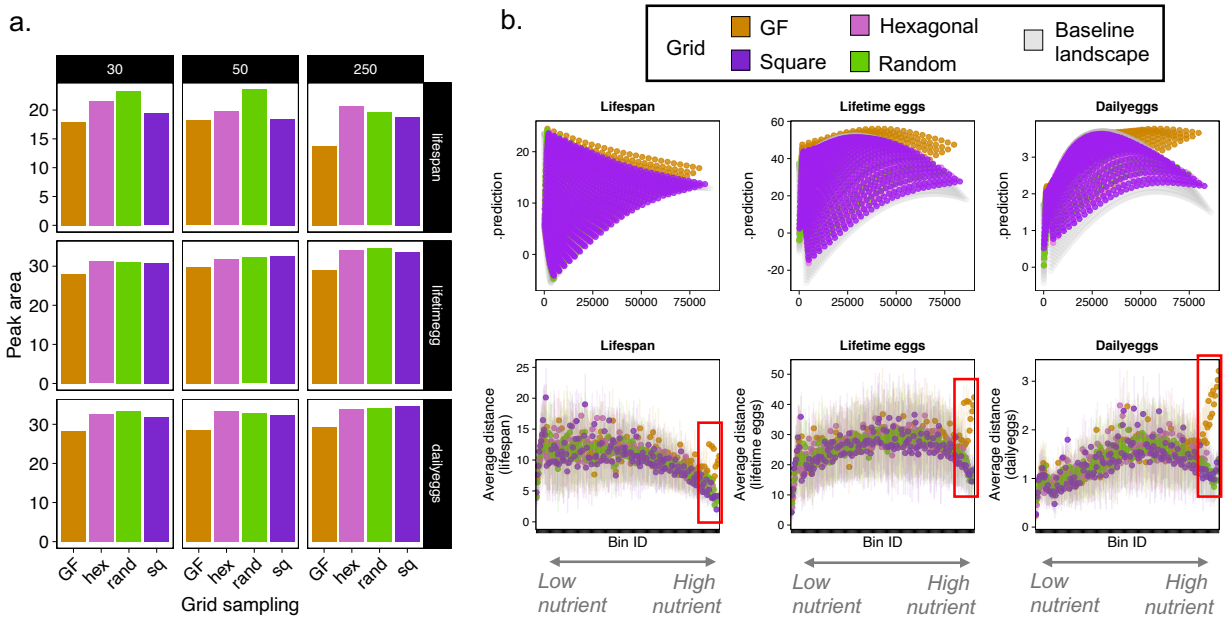


Figure 4.4: **Peak area and performance landscape topology.** (a) Predicted peak in the performance landscape of lifespan (top) and daily eggs (bottom) (see also Figure 4.6d) for lifetime egg peak predictions). Performance landscapes reconstructed from resolution equal to 50. (b) Overlaid peak predictions mapped onto the baseline performance landscapes of lifespan, lifetime eggs, and daily eggs across the sampling strategies. Note that GF sampling (orange) generates incomplete peak shape predictions for traits that respond to the interaction of nutrients. Hex: hexagonal sampling; rand: random points sampling; sq: square sampling.

For instance, in regions of high protein and carbohydrate, the average Euclidean distance between the reconstructed and true baseline performance landscapes increases rapidly for standard GF sampling relative to other sampling strategies, particularly for the landscapes of lifetime eggs (with peak at P:C ∼1:3) and daily eggs (P:C ∼1:2) (see highlighted red region in Figure 4.4b).

Discussion

I report an investigation of alternative sampling strategies of the nutritional space for GF studies. This is necessary so that GF landscapes can be made more robust, from which properties can be estimated and biological insights gained. This goes above and beyond current efforts that integrated ecological information into the design of traditional GF studies, as those suffer from similar limitations that underpinned this work (i.e., regions of the nutrient space with a lack of sampling) (Rothman et al. 2011, Wilder et al. 2013, Vaudo et al. 2016). I tested three alternative grid sampling strategies: hexagonal, square, and random point grids. Using a landmark dataset coupled with the pioneering Nutrigonometry method, I showed that all sampling strategies are able to provide reasonable estimates of the nutrient ratios where the peak in nutritional landscape is found. However, GF sampling provides incomplete estimates of peak region. This can have knock-on consequences for biological inferences when, for example, peak area is relevant to understand the nutritional conditions that maximize the expression of a trait. Importantly, GF sampling also provides inaccurate estimates of the performance landscape shape for performance traits that respond to the interaction between nutrients, highlighting additional limitations of the standard GF experimental design for biological insight using the properties of the performance landscapes. Overall, this study shows that to build performance landscapes with reliable shapes for biological inferences, alternative strategies of experimental design are needed in GF studies.

Why does the GF sampling find the correct information of nutrient ratios but not on the shape of the peaks in the landscape? Figure 4.1a (right panel) shows that the GF sampling strategy explores only a subset of the nutritional space. For fixed ratio datasets, this is usually a triangular region, whereas for intake datasets, the shape can vary but never covers the entire nutrient space. As a result, the interpolation for the construction of the performance landscape becomes an extrapolation beyond the regions upon which the anchor points exist, which can be mathematically and computationally difficult to achieve even with more complete datasets (Campagna & Perracchione 2021). As a result, the standard thin-plate spline interpolation and subsequent algorithms to estimate peak position truncate the peak estimates on the boundary of the performance landscape that can be estimated based on the anchor points. In doing so, the shape and area of the peak is also truncated, losing important biological information (Figure 4.3). The alternative methods tested here circumvent this limitation by sampling a wider range of the nutrient space, with anchor points that support a more accurate estimate of the peak shape and area.

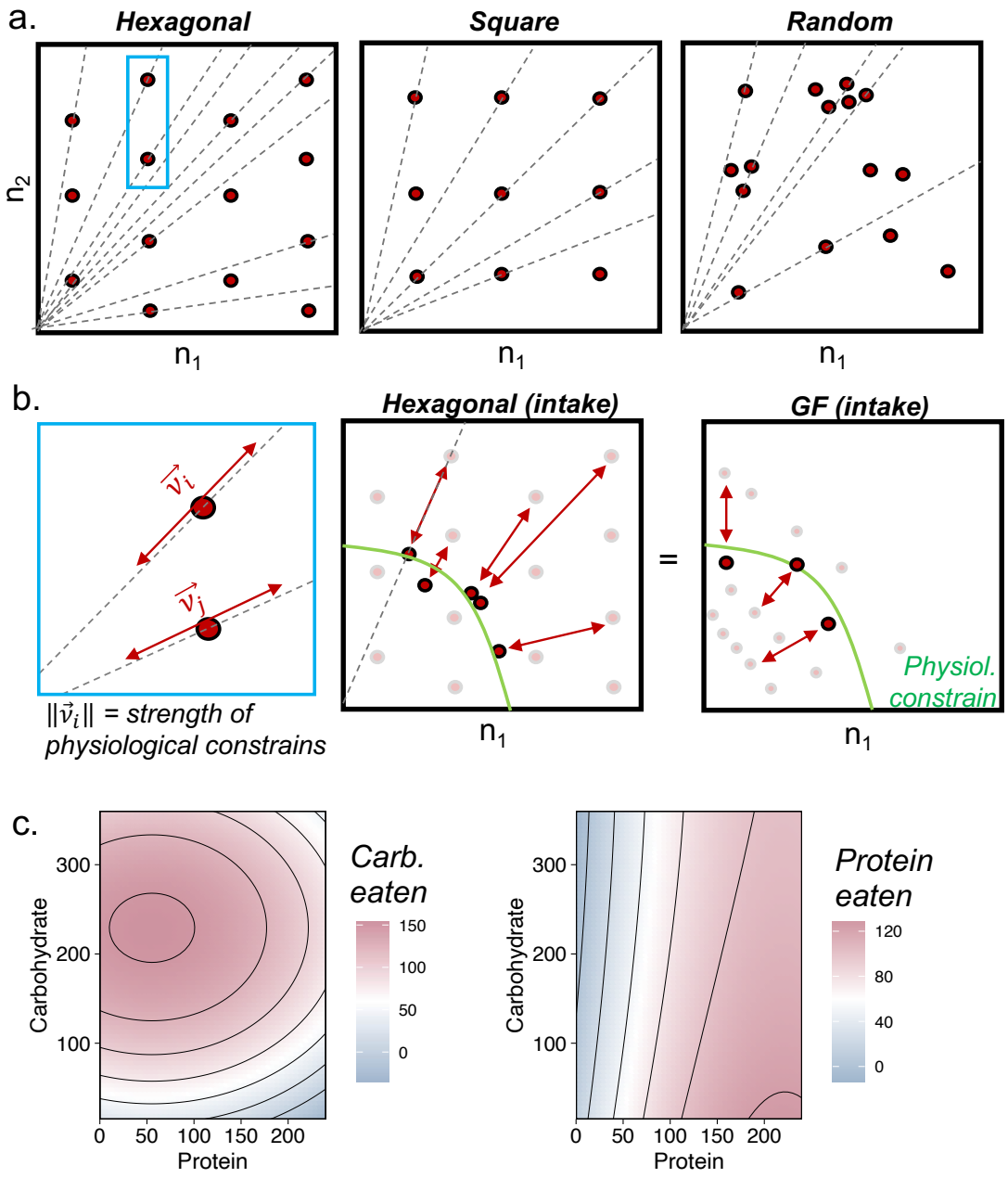


Figure 4.5: **Alternative sampling strategies used for intake datasets.** (a) Each anchor point could be seen as a point in a nutritional rail (as defined in standard GF design). This is true for all alternative sampling strategies tested in this study. (b) (left) Zoom of a specific region of the nutritional space from the hexagonal grid strategy (in a). When measuring intake, the anchor points move along nutritional rails represented by a vector $\vec{\nu}_i$ for the ith anchor point. The magnitude of the vector, $||\vec{\nu}_i||$, provides a measure of the strength of the physiological constraint experienced by the animals across diets as this metric shows the distance travelled by the anchor point along the nutritional rail (centre and right panels). Green line represents a hypothetical demand imposed by physiological constraint. (c) Performance landscapes of protein and carbohydrate intake (from Lee et al. 2008) to illustrate how intake can be used as the third dimension in performance landscapes. This can assist the inferences of rules of compromise which determine the amount of food and the quantity of each nutrient that individuals are capable of over- or under-consume in order to minimise distance between current food intake and self-balanced food intake (Simpson & Raubenheimer 1993, Raubenheimer & Simpson 1993). Rules of compromise are not dealt with in this study as it lies beyond the study's main scope, and is part of a next manuscript of this series.

Why does the GF sampling lead to more inaccurate landscapes in regions of nutrient interactions? The first reason lies in the previous point: GF only covers a subset of the nutritional space. Often, the diagonal region of the nutrient space has less 'covered area' relative to the empty nutrient space (see, e.g., Figure 4.1a). Consequently, a larger area of the performance landscape is missing and needs to be extrapolated, which can result in higher error. The second reason is likely related to the curvature of the performance landscape. I showed that the inaccuracies increase in performance landscapes for traits with peaks in the middle of the nutrient space, which indicates that the trait responds to the interaction between nutrients rather than an additive effect. For instance, the inaccuracies were almost absent for the landscape of lifespan but progressively more accentuated for the landscapes of lifetime eggs and daily eggs, respectively (Figure 4.4b). The absence of anchor points (i.e., diets) covering the full diagonal region likely precludes an adequate estimate of the curvature of the performance landscape in regions of nutrient interactions. The alternative methods tested here circumvent both of these limitations of GF sampling by covering a wider region of the nutrient space, including in the diagonal region. Note, however, that although the alternative methods perform better than standard GF sampling, they still introduce inaccuracies in the performance landscapes in the regions of nutrient interactions, providing an important area for future theoretical, computational, and empirical work to understand the underlying reasons.

In this study, I used fixed ratio datasets as a proof-of-concept, which is the structure that has been used recently in studies of GF focused on development (Silva-Soares et al. 2017, Alton et al. 2020, Kutz et al. 2019, Ma et al. 2022). However, GF sampling primarily covers datasets with individual nutrient intakes (e.g., Lee et al. 2008, Maklakov et al. 2008, Hawkes et al. 2022). Individuals' nutrient intakes are constrained by animal physiology and are difficult or impossible to overcome (e.g., individuals often die on overly unbalanced diets). Consequently, animals will unlikely eat sufficient quantities of food to explore the entire nutrient space, particularly in diets that are highly unbalanced relative to physiological constraints. As a result, the anchor points will be shifted in the direction of the physiological constraint, which can be represented by a vector $\vec{\nu}_i$ (Figure 4.5a-b). Note that each anchor point can be represented as a point in a nutritional rail that determines the direction of the vector $\vec{\nu}_i$ (Figure 4.5b).

In this case, the anchor points for any performance landscape of the alternative sampling strategies tested here, if plotted using intakes, will yield a similar performance landscape to that generated by GF sampling because individuals will shift their intakes to match the physiological constraints (Figure 4.5b). In other words, the performance landscapes from all methods will tend to converge. This is

important because nutrient intake data can reveal physiological constraints as well as compensatory feeding strategies underpinning rules of compromise, where individuals modulate the intake of more (or less) concentrated diets to achieve similar P:C ratios and total nutrient intake (Raubenheimer & Simpson 1993). This information is unavailable in fixed ratio data where intake is not measured.

Several questions could be raised, for instance: (i) How can the limitations of alternative methods in terms of representing nutrient intakes to derive rules of compromise be resolved? Or (ii) Why then, use alternative methods, if they either fail to provide intake datasets or converge towards the standard GF sampling strategy? To answer the first question, it is important to notice that it is not mandatory to use nutrient intakes to define the anchor points when generating landscapes (e.g., Kutz et al. 2019, Alton et al. 2020). It is true that, in general, GF studies have used individual nutrient intake as an input variable upon which the performance trait was mapped and the landscape built (see Simpson & Raubenheimer 2012) for a comprehensive review. However, intake estimates can and have been used as the output (performance) variable in GF studies, opening up the possibility of using fixed ratios for the design of experiments and nutrient intake as performance traits (rather than input variables). For example, a GF study showed that yeast-rich diets induce higher water intakes in *D. melanogaster* (Fanson, Yap & Taylor 2012). That said, it is possible to use the alternative methods presented here as fixed ratios upon which nutrient intake and performance traits can be mapped (Figure 4.5c).

I conjecture that this approach will enable us to extract the same rules of compromise and insights into physiological constraints as the original GF approach. The formalization of this conjecture requires an extensive argument that lies beyond the scope of this paper, as it involves introducing new concepts, e.g., intake targets (Simpson & Raubenheimer 2012), but is part of a follow-on manuscript being conceived. Importantly, the conjecture must be valid under the assumption that performance landscapes of alternative methods and GF sampling converge (as in Figure 4.5b). This leads to the answer of the second question: Why then use alternative methods? Alternative methods allow for more complete exploration and accuracy in the representation of performance landscapes, as shown here. This opens up the possibility to use properties of the performance landscapes as new proxies for biological insights. For instance, peak area and shape could provide insights into the nutritional resilience of the animal in maximizing a trait under varying nutritional conditions (e.g., the wider the peak, the more nutritionally resilient the animal). The use of the properties of the performance landscape cannot be achieved unless performance landscapes explore the entire nutritional space. Thus, alternative sampling methods expand the scope of GF methods and can unlock new measurements that can provide unique insights into compensatory

feeding strategies with biological significance and, more broadly, the evolution of nutritional trade-offs (Raubenheimer & Simpson 1993, Fanson et al. 2009, Fanson, Yap & Taylor 2012).

4.4 Conclusion

Despite the growing integration of ecological information into experimental design, current GF studies use a design aimed at sampling the nutrient space to construct performance landscapes that had not been scrutinized (Rothman et al. 2011, Wilder et al. 2013, Deans et al. 2016, Vaudo et al. 2016, Le Gall et al. 2020, Lawton et al. 2021). I tested alternative sampling strategies and show that their performances in reconstructing landscape properties are superior. Among these alternative strategies, the hexagonal design seems the most obvious choice for empirical testing as it allows for anchor points to be distributed such that more area is covered in the nutrient space per anchor point.

Future studies will illuminate how other standard metrics calculated in GF studies (i.e., rules of compromise) can be estimated and calculated from fixed ratio data with hexagonal (or other sampling strategy) design. This includes, for instance, regions in which the combination of nutrients is potentially lethal, generating holes in the performance landscapes (Blonder 2016, Conceição & Morimoto 2022). Overall, the findings presented represent an advance in current GF experimental design methodology. This has important consequences for the field because GF enables a multidimensional approach in nutrition where performance landscapes can provide important biological insights into the evolution of animal nutrition and life histories.

4.5 Acknowledgements

The author would like to thank the Riemann Center for Geometry and Physics for the award of the Riemann Fellowship.

4.6 Data accessibility

The dataset from Lee et al. (2008) accompanies the R code and is also available on Dryad: `doi:10.5061/dryad.tp7519s`. The R script containing functions for implementing the analysis will be made available as supplementary material.

4.6.1 Appendix: Supplementary figures.

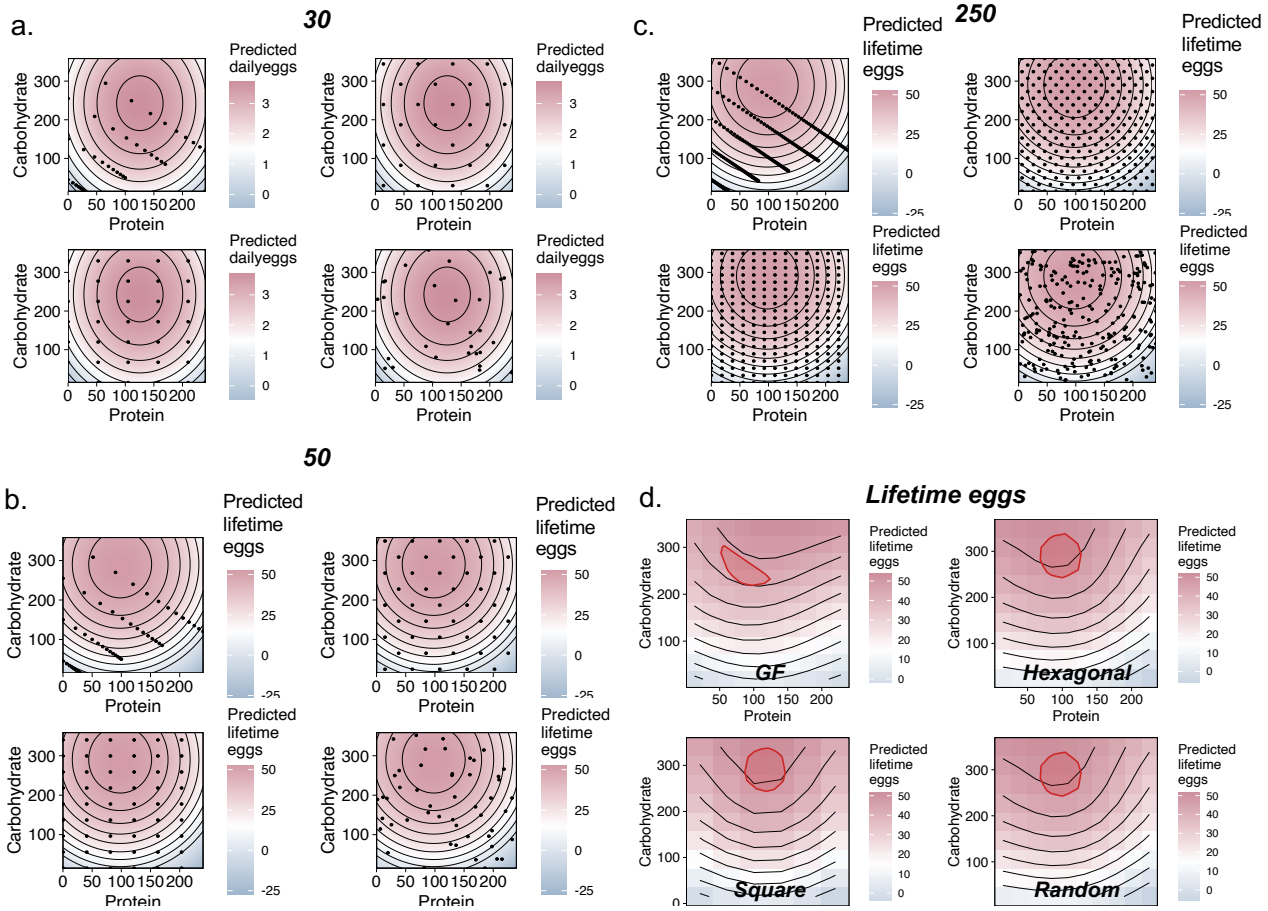


Figure 4.6: (a-c) Examples of the sampling strategies for resolutions of 30, 50, and 250, respectively. (d) Prediction of the peak regions for lifetime egg across sampling strategies (data from Lee et al. 2008).

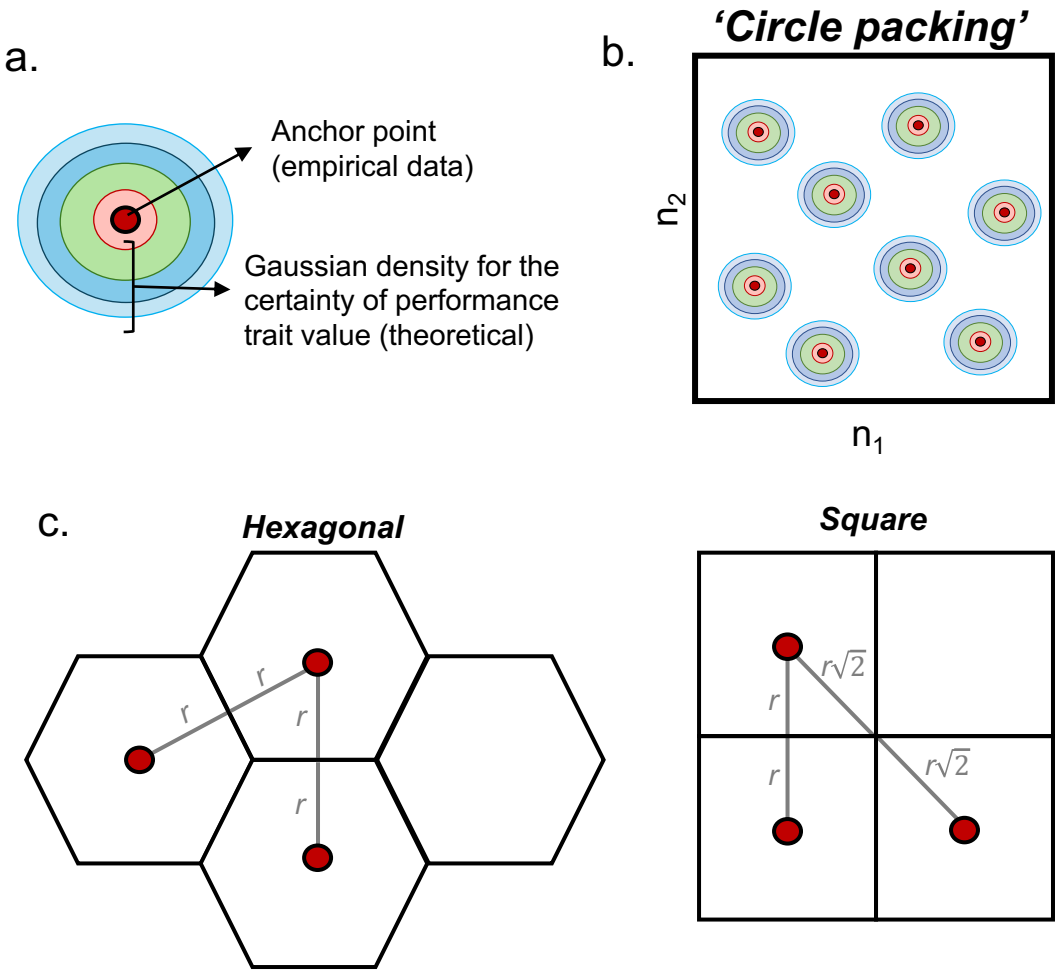


Figure 4.7: (a-b) Analogy to the circle packing problem and 2D Gaussian function as a way to determine certainty of performance trait information. (c) The distances between the anchor points in a hexagonal and square grid.

Chapter 5

Nutrigonometry III: Curvature, area and differences between performance landscapes

Publication reference

Morimoto, J., Conceição, P. and Smoczyk, K., (2022). Nutrigonometry III: Curvature, area and differences between performance landscapes. *Royal Society Open Science*, 9(11), p.221326.

5.1 Abstract

Nutrition is one of the underlying factors necessary for the expression of life-histories and fitness across the tree of life. In recent decades, the Geometric Framework (GF) has become a powerful framework to obtain biological insights through the construction of multidimensional performance landscapes. However, to date, many properties of these multidimensional landscapes have remained inaccessible due to our lack of mathematical and statistical frameworks for GF analysis. This has limited our ability to understand, describe, and estimate parameters which may contain useful biological information from GF multidimensional performance landscapes. Here, we propose a new model to investigate the curvature of GF multidimensional landscapes by calculating the parameters from differential geometry known as Gaussian and mean curvatures. We also estimate the surface-area of multidimensional performance landscapes as a way to measure landscape deviations from flat. We applied the models to a landmark dataset in the field, where we also validate the assumptions required for the calculations of curvature. In particular, we showed that linear models perform as well as other models used in GF data, enabling for landscapes to be approximated by quadratic polynomials. We then introduced Hausdorff distance as a metric to compare the similarity of multidimensional landscapes.

Keywords: Ecological specialisation, Grinnelian niche, diet; climate change; persistence homology.

5.2 Introduction

Animals can often balance their nutrient intake to maximise fitness (Raubenheimer & Simpson 1997, Raubenheimer et al. 2009). This creates the potential for nutritional trade-offs to emerge when animals cannot simultaneously maximise the nutrient balance for the expression of two competing fitness-related traits. (Simpson et al. 2004, Simpson & Raubenheimer 2012, Raubenheimer & Simpson 2020). Nutritional trade-offs are ubiquitous and have been described across the animal kingdom, from invertebrates such as flies (Lee et al. 2008, Fanson et al. 2009, Morimoto & Wigby 2016) and crickets (Maklakov et al. 2008) to vertebrates (Raubenheimer et al. 2015, Solon-Biet et al. 2014, 2016), including humans (Raubenheimer & Simpson 2016, Simpson et al. 2003). Nutritional trade-offs shape individual's fitness and can have important implications for responses to unbalanced diets as well as adaptations to novel environments (Behmer & Joern 2008, Behmer 2009).

Nutrition is complex and the ability of individuals to navigate nutritional trade-offs and properly balance their nutrient intake depend both on the amount of – and the (synergistic and antagonistic) interactions between –nutrients (Raubenheimer & Simpson 2020). A recent model known as the Geometric Framework (GF) enables the complexity of nutrition to be studied in relatively simple experimental designs, where both the quantity and the interactions between nutrients can be investigated simultaneously (Raubenheimer & Simpson 1993, Simpson & Raubenheimer 1993). GF has gained central stage in studies of nutrition both in animals and humans and has underpinned major insights into the evolution of nutritional responses (e.g., protein leverage hypothesis Simpson & Raubenheimer 2005) (Raubenheimer & Simpson 2020, Simpson & Raubenheimer 2012). Despite this, the development of analytical frameworks to analyse properties of GF multidimensional performance landscapes has lagged behind, and many studies have relied on visual interpretations to draw conclusions (see e.g., Kutz et al. 2019, Ma et al. 2020, Barragan-Fonseca et al. 2019, 2021) (but see also Pascacio-Villafán et al. 2022). While the visual approach can be useful for simple inferences, it is neither objective nor reproducible. More importantly, it overlooks the nuanced properties of multidimensional landscapes which might contain useful biological information about the responses of animals to nutrition.

Recent models have been developed to automate and standardise the analysis of GF performance landscapes. For example, (Rapkin et al. 2018) proposed the use of the coefficients of regression models, obtained by regressing the performance trait i to the intake of nutrients (say P and C for protein and carbohydrate, respectively) as components of vectors $\vec{v_i}$ for each performance trait i. This enables the

comparison between the angle $\theta_{i,j}$ of vectors $\vec{v_i}$ and $\vec{v_j}$ as proxy of the strength of nutritional trade-offs. We showed that the use of coefficients could lead to inaccurate estimates of the angle $\theta_{i,j}$ (Morimoto & Lihoreau 2019). Instead, we proposed the use of the positions of the coordinates representing the region of interest in the performance landscapes of traits i and j as the components of the (position) vector $\vec{v_i}$ and $\vec{v_j}$, from which the nutritional trade-off can be more accurately estimated from the angle $\theta_{i,j}$. Other models have been proposed to find peaks and valleys in multidimensional performance landscapes, either using bootstrapping (del Castillo et al. 2022) or Machine Learning models (Morimoto & Lihoreau 2019). More recently, we also proposed a novel way to define the peak and valleys of multidimensional performance landscapes for comparison of strengths of nutritional trade-offs using the angle $\theta_{i,j}$ which strictly represents performance landscapes as right-angle triangles and uses trigonometry for estimates of nutritional trade-offs (Morimoto et al. 2023). However, these models focus on obtaining information of either peak or valley regions (or both) of the multidimensional performance landscapes, overlooking other properties of the landscapes with potential biological significance.

Here, we explored this and proposed a model to calculate both the surface-area and local curvatures of multidimensional performance landscapes. These two properties of performance landscapes can be extracted and compared against the expected value of a flat landscape, and thus can provide invaluable information as to the overall profile of the nutritional responses not only in the regions of optimal (peaks) and minimal (valleys) responses, but across the entire sampling space of nutrients. We refer to performance landscape sensu (Morimoto & Lihoreau 2019) in that the landscape represents the possible values of the performance trait (e.g., lifespan) across a range of (x, y) values. Importantly, animals do not necessarily 'walk' onto the performance landscapes. Hence, performance landscapes can be thought of as a 'blueprint' for the expression of a performance trait. This means that the definition of performance landscapes is different than the definition of fitness landscapes sensu stricto, the latter which incorporates fitness, phenotype and genotype (Fragata et al. 2019). We first explain the mathematics underpinning the estimates of curvatures in multidimensional landscapes. For this, we assumed that multidimensional performance landscapes can be approximated by a quadratic polynomial of the form $ax^2+by^2+cx+dy+exy$, where a, b, c, d, e are the coefficients of a general linear model and x, y correspond to protein and carbohydrate intakes, respectively. We then describe how the surface-area of multidimensional performance landscapes can be estimated and compared against the area of a flat landscape of same region, which provides a proxy of how wiggly the landscape is relative to a flat landscape. Next, we demonstrated the application of the model in canonical datasets (flat landscape and a saddle landscape) and to a land-

mark GF dataset which investigates nutritional responses in terms of lifespan, lifetime egg production ('lifetime eggs'), and daily egg production ('daily eggs') in *Drosophila melanogaster*. To do this, we first tested the assumption that GF landscapes can be approximated by a quadratic polynomial by comparing the performance of a quadratic polynomial regression (LM), a generalised additive model (GAM) and a thin-plate spline (TPS) model, the latter two which are commonly used to analyse GF data. We then calculated surface-area and the Gauss and mean curvatures of the performance landscapes for lifespan, lifetime eggs and daily eggs. Finally, we compared the performance landscapes against a flat landscape as well as against each other using a metric known as the Hausdorff distance, which enabled us to compare two multidimensional performance landscapes of $n-$dimensions. This has the potential to expand the tools in which two landscapes can be directly compared, broadening our ability to make inferences about nutritional trade-offs when two performance landscapes are substantially different. Overall, the model proposed here advances our ability to study the properties of multidimensional performance landscapes and can underpin important biological insights from multidimensional studies in nutritional ecology and evolutionary biology.

5.3 The model

The model was designed to estimate properties of performance landscapes that have so far been overlooked. This is because GF studies and models have primarily focused on identifying regions of maxima or minima in the performance landscape, that is, peaks and valleys. While useful, this approach might neglect other properties of the performance landscapes which may contain important information to characterise the landscapes and gain insights into the biological responses to varying nutrient balances. To discuss the model, we first present the Taylor's theorem and the use of polynomial approximations to describe the performance landscapes. Next, we provide an intuition of curvature and present the Theorem Egregium. We then focus on the application of these two concepts to performance landscapes of empirical work to demonstrate how the model presented here can be used. For the purpose of the discussion and for the target audience of the paper (e.g., biologists), we denote an affine subspace with zero curvature as a 'flat surface'.

5.3.1 Taylor's theorem and polynomial approximation of performance landscapes

Given any smooth function f (or at least k-differentiable), it is possible to apply Taylor expansion to obtain a polynomial approximation of f of any degree k around a given point – this is known as the *Taylor's theorem* (Taylor 1715). The resulting polynomial is called the kth-order Taylor polynomial. It is important to note that to compute curvature, the smooth function f must be at least 2-differentiable (i.e., differentiable at least up to the second degree). This is because the first derivative f' of the smooth function f provides the slope of the tangent line or plane while the second derivative f'' provides information about the concavity of the tangent line or plane, and hence, its curvature. Taylor's theorem can be applied to more than one variable, providing a way to approximate functions with n variables using k degree polynomials. The approximated function is $k-$differentiable and can be used to estimate curvature of the landscape when $k \geq 2$. For the purpose of this study and the available empirical data in the literature, we will focus on functions with two variables (x, y) and approximations using $k = 2$ (i.e., quadratic approximations). Nonetheless, the arguments used here are applicable to functions with n variables and k degree polynomials.

5.3.2 Intrinsic and extrinsic curvature and the Theorem Egregium

As mentioned above, let's consider only smooth functions and in particular, performance landscapes approximated by a polynomial of degree 2. Curvature at each point p is defined via *principal curvatures*. The two principal curvatures κ_1 and κ_2 represent, respectively, the maximum and minimum amount of 'bending' in the landscape in each orthogonal direction of movement in the landscape along the tangent plane (i.e., a plane tangent to a point p on the landscape). The principal curvatures are algebraically defined as follows: κ_1 and κ_2 (at a point p) are the eigenvalues of a linear operator (i.e., can be seen as a matrix) called the *shape operator*. The shape operator at point p is defined as

$$\mathcal{S}(v) = \nabla_v(n)$$

where v is a vector tangent to p, ∇_v is the directional derivative and n is the unit normal vector field of a surface M. As $\nabla_v(n)$ is still a tangent vector at p, the shape operator is a linear operator

$$\mathcal{S}_p : T_pM \to T_pM,$$

where T_pM is the tangent space at a point p of M, that is, a vector space that encapsulates all the possible directions which pass tangentially through p. This makes sense because the eigenvectors measure the distortion of a linear operator, in this case, the eigenvalues of the shape operator tell us the how much the landscape bends.

Principal curvatures bear relationship with *extrinsic* and *intrinsic* curvatures. Extrinsic curvature is a property of the landscape which depends on the space in which the landscape exists. In other words, extrinsic cuvature are properties that depend on the *embedding* of the landscape. Conversely, intrinsic curvature is a property that only depends on the landscape itself, that is, the intrinsic curvature depends only on the *metric space*.

The relationships between principal curvatures and extrinsic and intrinsic curvatures are simple and useful for analysing landscapes. For instance, the *mean curvature*, which is an extrinsic curvature, is the arithmetic mean of the principal curvatures, that is, $\frac{\kappa_1+\kappa_2}{2}$. The *Gauss curvature*, an intrinsic curvature, is the product of the principal curvatures $\kappa_1\kappa_2$. Therefore, the Gauss curvature is the determinant of the shape operator and the mean curvature is the arithmetic mean of the trace of the shape operator. From linear algebra, the determinant of a matrix is invariant under change of bases (and row columns operations) - that is the intrinsic property - whereas the sum mean of the trace of the shape operator does depend on the embedding of the surface and, therefore, an extrinsic property.

Importantly, the Gauss curvature (but *not* the mean curvature), as it is an intrinsic invariant, does not change value if the surface bends without stretching, that is, invariant under local isometries – this is known as Gauss' *Theorema Egregium*. A consequence of this theorem illustrates the local to global interplay of geometry and topology. The theorem provides information as to when two spaces are different, for instance, a sphere of radius r has Gauss curvature $\frac{1}{r^2}$ and a plane has Gauss curvature 0. One way of developing an intuition about Gauss curvature in particular is to play with triangles. From Euclidean geometry, we know that the sum of the internal angles in any triangle sums up to π, that is, 180 degrees. Geometric objects that satisfy this condition are called planar. But what if this conditions fails? For instance, imagine that, instead of drawing a triangle in a flat piece of paper, we draw the triangle in the

surface of a sphere. Note that the triangle is not allowed to leave the surface, that is, it must be drawn onto the surface itself. In this case, the sum of the internal angles of the triangle will be *greater* than π. If we were to compare the triangles drawn in a piece of paper and in the surface of a sphere, the latter is more *curved* than the former. Sphere-like spaces are known as *convex* surface or *(locally) positively curved spaces*, whereas flat-like space are referred to as spaces with *zero curvature*. There are also situations where the sum of the internal angles of a triangle is *less than* π, for example, a triangle embedded in a saddle surface. These are known as (locally) hyperbolic spaces or *(locally) negatively curved*. But what does this have to do with Gauss curvature? The sign of the Gauss curvature provides a way to identify the characteristics of the surface: a surface with zero Gauss curvature can be classified as flat. Surfaces with Gauss curvature positive and negative are classified as convex or hyperbolic, respectively (see Figure 5.1). For the landscapes in nutritional ecology (and evolutionary biology, more generally), Gauss curvature can thus be an important parameter to characterise the local properties of the space that the landscape generates.

We can also obtain information of the local magnitude of the curvature of a given space that is positively or negatively curved. For example, as mentioned above, a flat plane has Gaussian curvature 0 across its entire domain, whereas a sphere of radius r has Gaussian curvature $\frac{1}{r^2}$ (Do Carmo 2016). This means that, for a sphere, the larger its radius, the smaller the local curvature at a given point in the surface of that sphere. In other words, if you were to stand on top and walk along a perfectly spherical ball (in any direction, since this is a perfect sphere), it is easy to imagine that this is a much harder task than walking on the surface of the Earth. This is because the radius r_s of the football is substantially smaller than the radius r_E of the Earth, and thus, the local curvature of the former is greater than the local curvature of the latter (ask the Flat Earth Society!).

For practical applications, however, the Gauss curvature has limitations. For instance, imagine rolling a sheet of paper to form a cylinder. The flat sheet of paper has Gauss curvature equals to zero and, because there were no deformations such as stretching of the sheet of paper when rolling, the Gauss curvature of a cylinder is also zero. This happens because the cylinder is an isometric transformation of a flat plane (i.e., the transformation does not tear, stretch, or shrink the flat surface). As a result, the Gauss curvature alone is unable to differentiate between the two forms. Yet, they are different, and we need additional metrics to differentiate them. This leads to the definition of other types of curvature which depend on proprieties of the object as well as the space in which the object exist (*extrinsic* properties of a surface). In this instance, we can resort to the estimates of the *mean curvature* which is non-zero for

the cylinder but zero for a flat surface (see also Do Carmo 2016, for formal definitions). Together, Gauss and mean curvatures provide important properties to characterise surfaces and, in our case, landscapes.

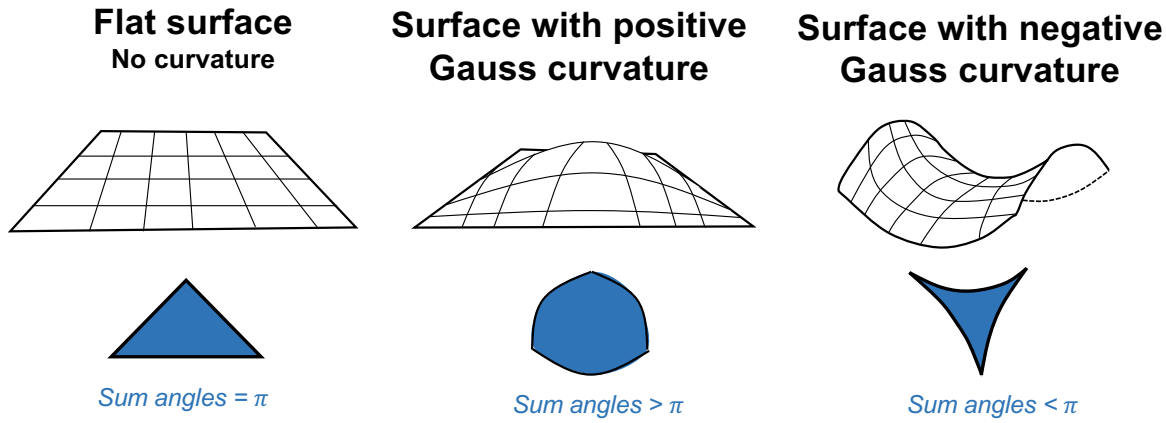


Figure 5.1: **Schematic representation of the concept of curvature** Flat surfaces had no curvature such that, if we were to draw a triangle embedded on the surface, the sum of the internal angles would equal π. Surfaces or sub-sections of surfaces with positive Gauss curvature are curved in a way that, if we were to draw a triangle embedded in the surface, the sum of the internal angles would exceed π. This could be equivalent to sub-sections of the surface corresponding to the surface peak regions. Surfaces with negative Gauss curvature are curved in a way that, if we were to draw a triangle embedded on the surface, the sum of the internal angles would be less than π. This is equivalent to sub-sections of the surface corresponding to mountain pass regions. Note that, for the purpose of this paper, we use 'flat surface' to denote an affine subspace with zero curvature.

5.3.3 Curvature and nutrition

In nutritional ecology and GF experiments, the idea of curvature can be useful to understand how animals are expected to navigate nutrient imbalances given the shape of the performance landscape. For instance, animals navigating a performance landscape with small or zero curvature likely benefit (or pay costs) of nutrient imbalances in equal magnitude throughout the landscape, as they move from valley to peak regions. It is analogous to walking uphill (or around the hill, depending on the direction). This suggests that an imbalanced diet has the same relative cost to performance in two arbitrary regions of the landscape. Conversely, animals navigating a performance landscape with large curvature likely pay a greater relative cost to performance in certain regions of the landscape (e.g., on the edges) where the negative effect of nutrient imbalance is accelerated. Thus, understanding the curvature of a performance landscape can enable us to understand and predict the costs associated with nutrient imbalance.

Following the idea of curvature, we also propose to use the idea of *surface area* of the performance landscape as an additional metric for landscape characterisation. The concept of area is easily understood

and thus, we will not delve into analogies. One important point worth mentioning here is that surface-area allows us to estimate how 'wiggly' (rugged) a performance landscape is. This is because the wiggliness of a landscape increases its surface-area (analogous to e.g., cell microvilli) relative to a flat landscape with the same domain (i.e., values of x and y). In nutritional ecology and GF experiments, the wiggliness of the performance landscape can indicate how resilient the animal is to small changes in nutrient balance, and how plastic the animal is in maximising performance traits in response to varying combinations of diet (e.g., the term we call 'nutritional plasticity'). Performance landscapes with large surface-area likely represent cases where small deviations in nutrient balance of the diet creates large differences in the expression of a performance trait (i.e., the landscape is wiggly). This could indicate that the performance trait responds rapidly to changes in nutrition and hence, is nutritionally plastic, but not resilient. On the other hand, performance traits with landscapes that are less wiggly indicate high resilience against deviations from nutrient balance, but potentially low nutritional plasticity.

With this intuition, we now present the formal mathematics underpinning the calculations of curvature and surface-area, with focus on the performance landscapes of interest to this study. Next, we apply the formulations to canonical and real-world datasets to demonstrate their use and interpretation (see also Text S1 and Fig 5.5).

5.3.4 Curvature of performance landscapes

Suppose S is a performance landscape parameterised as $(x, y) \mapsto (x, y, f(x, y))$ where $f(x, y)$ is given by a quadratic polynomial of the form $ax^2 + by^2 + cx + dy + exy$ and a, b, c, d, e are coefficients of a general linear regression. The variables x and y can represent intakes of protein and carbohydrate, respectively. The domain of $f(x, y)$ lies within $[0, x^*], [0, y^*]$, where (x^*, y^*) are the maximum intake of nutrients x and y nutrients. This domain is identical for any $S_1, S_2, \ldots, S_i$, multidimensional performance landscapes obtained from the same experiment. For reasons that we discuss below, it is necessary to ensure that the domains of the multidimensional landscapes coincide. Therefore, to compare multidimensional performance landscapes from different experimental designs, the domains can be min-max standardised as $\frac{x-\min(x)}{\max(x)-\min(x)}$ (and similarly for y) such that the domains of the landscapes are $x = [0, 1]$ and $y = [0, 1]$. We can then estimate the Gauss curvature K_{S_i} and mean curvature H_{S_i} of a performance landscape S_i. To do this, we first estimate the gradient ∇S_i in the x and y directions:

$$\nabla S_x = \frac{\partial S_i}{\partial x} = \begin{pmatrix} 1 \\ 0 \\ 2ax + c + ey \end{pmatrix}; \quad \nabla S_y = \frac{\partial S_i}{\partial y} = \begin{pmatrix} 0 \\ 1 \\ 2by + d + ex \end{pmatrix} \tag{5.1}$$

Note that these partial derivatives provide the coordinates of tangent vectors to a point within the landscape in the (x, y) directions. As a result, we can also calculate the unit normal vector $\vec{n}$ perpendicular to the landscape at a given point as the cross product of the two partial tangent vectors in the directions of x and y such that

$$\vec{n} = \frac{\frac{\partial S_i}{\partial x} \times \frac{\partial S_i}{\partial y}}{\| \frac{\partial S_i}{\partial x} \times \frac{\partial S_i}{\partial y} \|} \tag{5.2}$$

Next, we calculate the second partial derivatives of $\nabla^2 S_i$ as

$$\nabla^2 S_x = \frac{\partial^2 S_i}{\partial x^2} = \begin{pmatrix} 0 \\ 0 \\ 2a \end{pmatrix}; \nabla^2 S_y = \frac{\partial^2 S_i}{\partial y^2} = \begin{pmatrix} 0 \\ 0 \\ 2b \end{pmatrix} \tag{5.3}$$

We can now use the I and II fundamental forms to estimate Gauss and mean curvature at all points in the landscape. A simplified geometric intuition for the I and II fundamental forms can be obtained as follows: (i) the I fundamental form can be thought of as providing information of the curvature in the landscape for any direction of travel, starting at a given point in the performance landscape. (ii) the II fundamental form provides information of how much the landscape curves away (or deviates) from a flat tangent plane at a given point in the landscape. Note that the I fundamental form can be estimated solely using properties of the landscapes itself (and hence is related to 'intrinsic' curvature of the landscape) while the II fundamental form requires an additional parameter (i.e., the normal vector $\vec{n}$) and hence provides 'extrinsic' curvature of the landscape. We calculate the I fundamental form as:

$$E = \frac{\partial S_i}{\partial x} \cdot \frac{\partial S_i}{\partial x}; \; F = \frac{\partial S_i}{\partial x} \cdot \frac{\partial S_i}{\partial y}; \; G = \frac{\partial S_i}{\partial y} \cdot \frac{\partial S_i}{\partial y} \tag{5.4}$$

$$I = \begin{pmatrix} E & F \\ F & G \end{pmatrix} \tag{5.5}$$

Likewise, we calculate the II fundamental form as:

$$L = \frac{\partial^2 S_i}{\partial x^2} \cdot \vec{n}; M = \frac{\partial^2 S_i}{\partial x \partial y} \cdot \vec{n}; N = \frac{\partial^2 S_i}{\partial y^2} \cdot \vec{n}; \tag{5.6}$$

$$II = \begin{pmatrix} L & M \\ M & N \end{pmatrix} \tag{5.7}$$

The Gauss curvature K_{S_i} and the mean curvature H_{S_i} can then be calculated as

$$K_{S_i} = \frac{LN - M^2}{EG - F^2} \tag{5.8}$$

$$H_{S_i} = \frac{GL - 2FM + EN}{2(EG - F^2)} \tag{5.9}$$

K_{S_i} and H_{S_i} provide information about the local curvature of the multidimensional performance landscapes. For biological purposes, this curvature contains information on the shape of the landscape as well as the local changes in shape of the landscape across valleys and peaks (see below). Importantly, these quantities provide information regarding additional properties of the GF multidimensional performance landscapes which, combined with previous methods, can collectively describe the regions as well as the overall shapes of the landscapes.

5.3.5 Area of performance landscapes

Multidimensional performance landscapes are not always flat. Instead, performance landscapes can have oscillations in the z-direction such that the surface area A_{S_i} varies according to the magnitude of the oscillations. For performance landscapes that contains high degrees of oscillation (i.e., high *'wiggliness'*), the surface area is expected to be higher than that of flat performance landscapes if the two landscapes have the same domain. We then calculated the surface area of performance landscapes as a proxy of how 'wiggly' a landscape is relative to the area of a flat landscape over the same domain. To do this, we can estimate the area of the flat surface A_0 as the area of a rectangle with sides x, y. To estimate the surface area A_{S_i} of the performance landscapes, we can use the formula for the surface integral

$$A_{S_i} = \iint_{\Omega} \sqrt{1 + (\nabla S_i)^2}\, dx\, dy = \iint_{\Omega} \sqrt{1 + \frac{\partial S_i}{\partial x}^2 + \frac{\partial S_i}{\partial y}^2}\, dx\, dy \tag{5.10}$$

The double-integral is then evaluated over the intervals $[0, y]$ and $[0, x]$, respectively, to return the

surface area of the landscapes over its domain Ω. For performance landscapes in nutritional geometry, where we can consider x as protein intake and y as carbohydrate intake, we evaluate the surface integral from 0 to the maximum intake of x and y ($\max x, y$) so that the equation above becomes

$$A_{S_i} = \int_0^{\max y} \int_0^{\max x} \sqrt{1 + \frac{\partial S_i}{\partial x}^2 + \frac{\partial S_i}{\partial y}^2}\, dx\, dy \tag{5.11}$$

Alone, this metric is of little use. For a more informative estimate, we can calculate the ratio of the surface-area of the performance landscape over the surface area of a flat landscape S_0 with the same domain as the performance landscape. In doing so, we obtain the *surface-area ratio* of performance landscape i as $A_i = \frac{A_{S_i}}{A_0}$ where $A_i = [1, \infty)$. The greater the value of A_i, the greater the surface-area of the performance landscape relative to a flat landscape of same domain. Conversely, performance landscapes with A_i closer to 1 are nearly flat. Thus, the surface-area ratio could be interpreted as the magnitude of wiggliness of a landscape, provided that the ratio informs how many times the area of the performance landscape is greater than that of a similar flat landscape. Note that for estimating surface-area, for which the ratio is performed against a flat landscape with similar domain (as opposed to direct comparisons between two performance landscape), the z value was maintained in its original scale to represent the deviations from a flat surface in the z dimension (but see also section on Hausdorff distance below).

5.4 Material and Methods

5.4.1 Statistical analyses

All analyses were conducted in R version 4.1.3 (Team 2019). Data handling was conducted using the tidyverse packages 'dplyr' and 'tidyr' (Wickham et al. 2019). Performance landscape plots and mean curvature plots were done using the 'ggplot2' package (Wickham et al. 2016). Closed solutions for the surface integrals and the function to estimate the Hausdorff distance were obtained from the 'pracma' package (Borchers 2022). We estimated confidence intervals for the Hausdorff distance using the'boot' function from the package of the same name (Canty 2002).

We first validated our assumption that performance landscapes could be approximated using a quadratic polynomial. To do this, we split the data into training and testing sets in the proportion 60-40%. We then generated new datasets by sampling with replacement from the training set, and fitted

one of three models: (1) general linear model (LM) with trait value as response variable, the linear and quadratic effects of protein and cabrohydrate intakes, and the linear interaction between protein and carbohydrate intakes; (2) generalised linear model (GAM) using the 'gam' package (Hastie & Hastie 2015) with similar structure as the model above and with smoothing function 's' with default parameters; and (3) a thin-plate spline model (TPS) using the 'fields' package (Nychka et al. 2017) with trait value as response variable and protein and carbohydrate intakes as independent variables. We estimated model performance using the square-root of the sum of the squared residuals (i.e., root-mean-square-error or RMSE) for both the training and testing datasets.

We then developed the algorithms to estimate curvature and surface-area. The underlying algorithm used in all analyses is as following:

1. We fitted a general linear model for each performance trait, with the trait value (simulated or empirical) as a dependent variable and a polynomial with the main linear and quadratic effects of protein and carbohydrate intakes and the linear interaction between protein and carbohydrate intakes.

2. We created a square grid from 0 to max protein and carbohydrate intakes, from which the predicted value of a point x, y corresponding to the intakes of protein and carbohydrate, respectively, could be predicted using the general linear models from above. This approach has two benefits:

    - It creates an interpolation similar to that obtained with *splines* but with the advantage of conforming to a generalised equation $ax^2 + by^2 + cx + dy + exy$. This polynomial facilitates integration and estimates of partial derivatives for calculation of curvature.
    - It ensures that the predicted value for all performance traits are obtained for similar regions within the domain of the performance landscape. This is because the models predict the expected value of each performance landscape using the same underlying x, y grid.

3. In addition to the predicted values for all traits, we included the values for a flat landscape of the form $(x, y, 1)$. Note that the choice of 1 was arbitrary and does not affect curvature or surface-area.

4. We then used the grid containing the predicted values for each trait to estimate curvature, surface area, and Hausdsorff distances.

The algorithm above is needed because standard GF design only explores the nutritional space through rails, which are lines that subdivide the nutrient space. This means that a large portion of the space

remains unexplored, and the approach above is needed to cover these unexplored spaces as shown in the second study of the Nutrigonometry series (see (Morimoto et al. 2023)). A full coverage of the nutrient space is needed for a global analysis of the properties of the performance landscapes (see also 'Discussion' section for more on this topic).

5.4.2 Difference between performance landscapes

We used the Hausdorff distance $d_H(S_m, S_n)$ to estimate the difference between two performance landscapes S_m and S_n. Note that the two performance landscapes can be of two traits or the landscape of a trait and the flat landscape with same domain. We used the Hausdorff distance because it measures the distance between two sets of non-empty compact ('closed') subsets of a given metric space and thus, provide a way to measure the overall differences between two subsets. For performance landscapes in 3D, the metric space is $\mathbb{R}^3$, and the non-empty subset within the metric space are the performance landscapes S_m and S_n. Formally, the Hausdorff distance $d_H(S_m, S_n)$ is defined as

$$d_H(S_m, S_n) = \inf\{\epsilon \geqslant 0;\ S_m \subseteq S_{n_\epsilon} and\, S_n \subseteq S_{m_\epsilon}\} \tag{5.12}$$

whereby ϵ is the distance which necessary for the subset S_m to contain the subset S_n and vice versa (note that the definition is symmetric). One can think of ϵ as a the smallest quantity needed to expand the subset S_m such that it contains set S_n *and* subset S_n to contain S_m. The bidirectionality of the definition of the Hausdorff distance ensures that the value of $d_H(S_m, S_n)$ equals zero if and only if the two sets (or in this case, the two landscapes) are the same. This implies that the distance between two sets may differ depending on the way the calculation is conducted. For example, consider two sets $A = \{1, 2\}$ and $B = \{1, 2, 3, 4\}$. If the Hausdorff distance was unidirectional, then the distance between sets A and B would equal zero, since $A \subseteq B$ (i.e., the numbers 1 and 2 in set A are also present in set B). Now, considering the bidirectionality of the definition of Hausdorff distance given above, we also need to calculate the distance between the set B and set A, which is equal to 2 (i.e., the maximum distance between the elements of B and A so that $B \subseteq A$ is $4_B - 2_A = 3_B - 1_A = 2$, where the subscripts represent the sets from which the number belongs in this example). Thus, performance landscapes that are identical have Hausdorff distances $d_H(S_m, S_n)$ equal to zero. To avoid scale effects when estimating ϵ, we mean-standardized the $z-$axis (i.e., divided by the variables' mean) to ensure that the estimates of $d_H(S_m, S_n)$ were obtained on the same scale for two performance landscapes.

5.4.3 Data sets and model application

To demonstrate the functionality of our model we first implemented the model in two canonical datasets: a flat surface $(x, y, 1)$ and a saddle surface (x, y, xy). Next, we applied the model to empirical performance landscapes from a landmark GF dataset, to estimate properties of the performance landscapes for lifespan, daily egg production (daily eggs), and lifetime egg production in *Drosophila melanogaster* (Lee et al. 2008). Gauss and mean curvatures, as well as the surface-area ratio of the canonical and empirical performance landscapes were calculated as explained in the previous section (see 'The Model' section).

5.5 Results

5.5.1 Evaluating model performance for landscape construction

The fundamental assumption for our model is that performance landscapes could be approximated using a polynomial regression model of the form $ax^2 + by^2 + cx + dy + exy$. Yet, in GF studies, other models are commonly used to analyse the data, such as generalised additive models (GAMs) and thin-plate splines (TPS). These models cannot be represented with closed quadratic polynomial equations but could result in better fit of the data and consequently, better approximations of the performance landscapes. To test this, we compared the model performance (using root-mean-square error or RMSE) of the quadratic polynomial linear regression (LM), GAM, and TPS models across the three traits of the dataset. The results show that the LM model has similar performance in both training and test datasets for all traits (Fig 5.2). Therefore, quadratic LM models are a valid approximation for analysing GF data. More importantly, this enables us to represent the regression model in the quadratic polynomial form, which facilitate the calculations of curvature and surface-area ratios.

5.5.2 Model application to canonical datasets

We first demonstrate the use of our model by applying it to canonical datasets. We calculated the surface-area ratio of the canonical landscapes, where we expected the ratio to equal 1 for the *flat* landscapes (as a flat landscape has the same area as itself) and greater than 1 for the *saddle* landscape, due to the additional surface-area resulting from its curvature. Indeed, surface-area ratio of the the *flat* landscape was equal to 1 while the surface-area ratio for the *saddle* landscape was equal to 3.982, revealing that the *saddle* landscape had approximately 4-times more surface-area than a flat landscape with the same

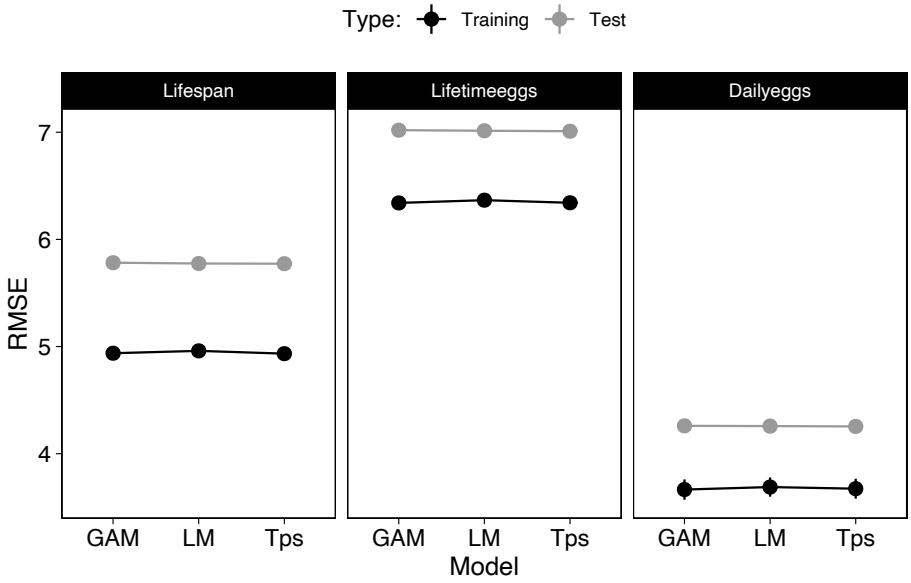


Figure 5.2: **Performance of commonly used models to analyse GF data.** RMSE: Root mean square error. Training: training dataset. Test: testing dataset.

domain. This shows that the saddle canonical landscape used here was ca. 4-fold more wiggly than a flat landscape. Importantly, aside from rounding error, this corroborates analytical expectation of the ratio between a flat landscape and a saddle landscape given by $\frac{\int_0^1 \int_0^1 xy\, dx\, dy}{\int_0^1 \int_0^1 1\, dx\, dy} = 4$.

Next, we calculated curvature. By definition, *flat* landscapes do not have curvature and, as expected, we found that the model gives both Gauss and mean curvatures equal to zero for all points (Fig 5.3). Conversely, for the *saddle* landscape we expected a point in which the landscape resembles a sphere at $(x, y) = (0, 0)$. At this point, the Gauss curvature is expected to be -1 and mean curvature, 0, provided that principal curvatures are equal to 1 and -1 (Do Carmo 2016). Mean curvature, on the other hand, was expected to be positive for regions in the saddle where the landscape bends downwards and negative for regions that bends upwards. This prediction emerges from theory on the behaviour of the normal unit vector $\vec{n}$ in saddle regions (Do Carmo 2016). As expected, our model found that both Gauss and mean curvatures were zero for a *flat* landscape. Moreover, the model found that Gauss curvature equals -1 at the point $(x, y) = (0, 0)$ in the *saddle* landscape, and that mean curvature was positive and negative for regions bending downwards and upwards, respectively (Fig 5.3). Together, these results demonstrate that our model can accurately estimate the curvature and surface-area ratio as a proxy of landscape wiggliness in canonical datasets.

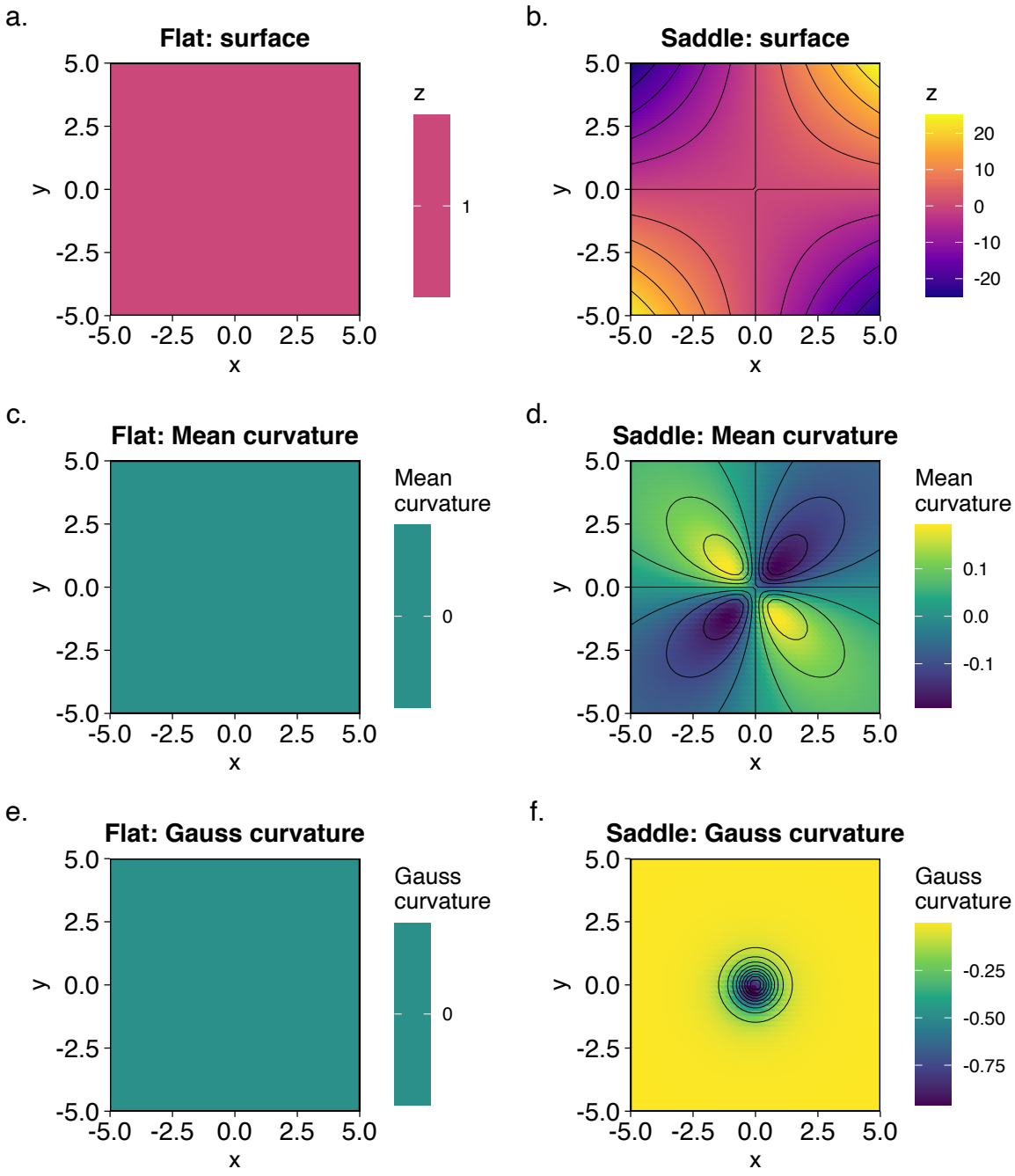


Figure 5.3: **Gauss and mean curvatures of two canonical datasets.** (a) Simulated flat and (b) saddle landscapes. (b-c) Mean curvatures of the flat (c) and (d) saddle landscapes, respectively. (e-f) Gauss and mean curvatures of the (e) the flat and (f) the saddle landscapes, respectively.

5.5.3 Model application to empirical datasets

Next, we applied the model to an empirical landmark GF dataset to estimate curvature and surface-area (see Lee et al. 2008). As in the previous section, we first estimated the surface-area ratio for the performance landscape for the three performance traits measured in the dataset: lifespan, lifetime egg production (lifetime eggs) and daily egg production (daily eggs). Our model found that the performance

landscape for lifespan had the largest deviations from a flat landscape (*surface-area ratio:* 51.125), followed by the the performance landscape of lifetime eggs (*surface-area ratio:* 49.587) and daily eggs (*surface-area ratio:* 4.194), respectively. These differences in the overall conformation of the performance landscapes were corroborated by the estimates of Hausdorff distances $d_H(S_m, S_n)$. All performance landscapes differed significantly from flat, but the magnitude of this difference was greater for lifespan, lifetime eggs, and daily eggs, respectively. This suggests that, the performance landscape for lifespan is the most different from a flat landscape, while the performance landscape for daily eggs is the closest (relatively speaking). We also compared the differences between landscapes using the Hausdorff distances and found that, as expected, the landscapes from lifespan and daily eggs are the most different, followed by the landscapes of lifespan and lifetime eggs and lifetime eggs and daily eggs, respectively (Table 5.1).

Table 5.1: Hausdorff distance estimates for the performance landscapes.

	$d_H(S_m, S_n)$			
Comparison	Mean	SD	lwr 95% CI	upr 95% CI
Lifespan-flat	2.411	0.270	1.426	2.534
Lifetime eggs-flat	1.940	0.110	1.765	2.186
Daily eggs-flat	1.783	0.134	1.375	1.967
Lifespan-Lifetime eggs	2.118	0.177	1.616	2.315
Lifespan-Daily eggs	2.615	0.197	2.589	3.311
Lifetime eggs-Daily eggs	1.661	0.212	1.418	2.214

We then investigated the curvature of the performance landscapes. Overall, the majority of the landscape had zero Gauss and mean curvatures, suggesting that these regions were equivalent to flat inclined planes. For instance, lifespan, which was the landscape with the greatest surface-area, had small regions of negative Gauss curvature and positive mean curvature at low nutrient intakes (i.e., closer to the origin), which disappeared for other regions of the landscape (see Fig 5.4a-c). Likewise, the performance landscape for lifetime egg had virtually zero Gauss and mean curvatures throughout, suggesting that the landscape as a whole was an inclined plane (see Fig 5.4d-f). The performance landscape for daily egg production was the only landscape that showed higher curvature estimates in regions close to the origin (i.e., low nutrient intake) (see Fig 5.4g-h). These results suggest that the performance landscape for daily eggs, which responds strongly to the interaction between protein and carbohydrate intakes, had relatively more positive curvature in the region of low nutrient intake. This is important because it can suggests that, in regions of low nutrient intake, animals can pay larger costs from small nutritional imbalances. This can be interpreted as analogous to walking along a ridge: small deviations from the path have potentially large implications for the position in the ridge (e.g., falling either side of the crest). Overall though, the results show that estimating surface-area and curvature of performance landscapes

can reveal important properties of the landscapes with potential biological significance (see Discussion for hypothesis generated from curvature).

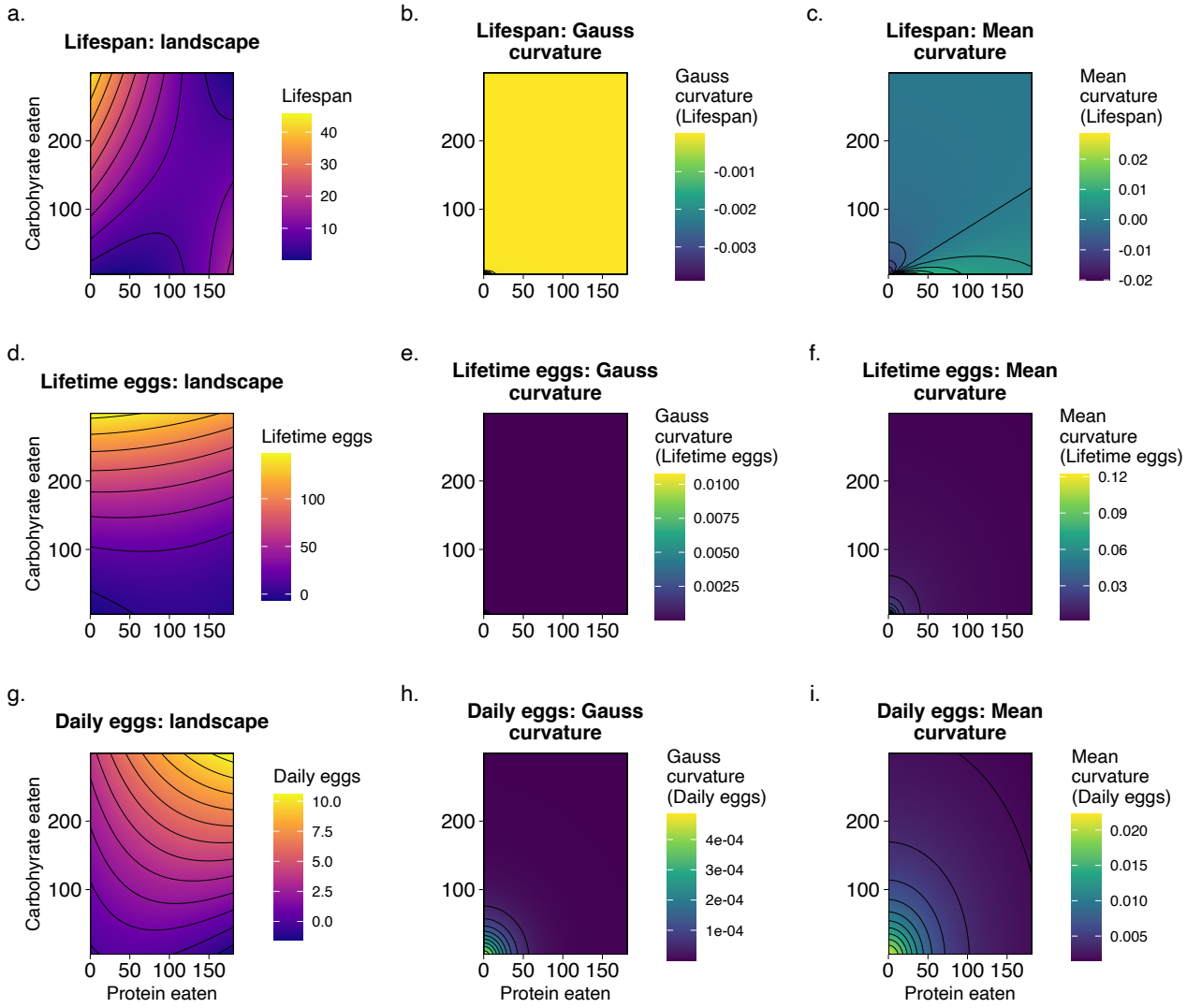


Figure 5.4: **Gauss and mean curvatures of performance traits.** (a-c) Landscape (a), Gauss curvature (b) and mean curvature (c) for lifespan. (d-f) Landscape (d), Gauss curvature (e) and mean curvature (f) for lifetime egg production. (g-i) Landscape (g), Gauss curvature (h) and mean curvature (i) for daily egg production (daily eggs).

5.6 Discussion

We proposed a new model to measure properties of performance landscapes related to surface-area and curvature. Moreover, we introduced the Hausdorff distance which can be used to compare the similarity between two $n-$dimensional performance landscapes. Our analysis first corroborate the underlying assumption that performance landscape could be approximated by a quadratic polynomial of the form $ax^2 + by^2 + cx + dy + exy$. In fact, we showed that this polynomial fit the data as well as other commonly

used models to analyse and interpolate GF data (i.e., GAM and TPS). These approximations are useful because *exact* curves and landscapes are never accessible in real-world biological data due to for example, sampling limitations and variability in responses. Therefore, the approximations used here and in previous studies provide a way in which landscapes can be constructed from a relatively sparse grid of known biological responses (Simpson & Raubenheimer 1993, Simpson et al. 2004). We used canonical simulated datasets and an empirical dataset (Lee et al. 2008), which measured the performance landscapes for lifespan and daily eggs, to demonstrate the application of our model. Together, the model proposed here is the first to estimate properties of performance landscapes other than peaks and valleys. This enables new insights from studies of nutrition in high dimensions. Currently, we do not know how (or whether) animals 'navigate' performance landscapes in similar ways to fitness landscapes (e.g., Fragata et al. 2019). Recent studies integrating the concept of performance landscapes and genetics have been carried out (see Hawkes et al. 2022) and simulations have provided insights into the genetic variance in response to diet for lifespan (Senior 2023). This remains a subject of future theoretical development. Despite this, from the perspective of curvature of multiple performance landscapes, one question emerges: how are the different performance landscapes integrated so that they influence animal nutritional resource allocation based on nutritional trade-offs? In other words, how do animals allocate nutrients to competing resources with different performance landscapes? The curvatures of the performance landscapes, if important, suggest that the cost-benefit functions are non-linear for different traits. Therefore, animals have to allocate resources to competing traits taking into account the non-linearity of the performance landscapes of each trait i, for all traits. More formally, we can state that for all fitness traits w with landscape S_w, animals will evolve to allocate resources such that the animal optimises some weighted average of all performance landscapes S_w. This implies that animal nutrition is in Pareto optimality, whereby any deviations on the nutrient intake or allocation to a trait w_i would result in a cost to other traits w_j (Debreu 1954). Pareto optimality in nutritional decisions have been demonstrated in computer algorithms (Wang & Bai 2006, Pei & Liu 2009) but to our knowledge, has not yet been shown in empirical data. Currently, we rely on the growing use of GF to generate landscapes to test Pareto optimality on nutritional trade-offs and feeding behaviour. A formal derivation of the Pareto optimal model is outside the scope of this paper and will be subject of future work.

Previous analytical models using GF data focused on identifying peaks and valleys in performance landscapes in order estimate the extent to which animals had to compromise in their nutrition for optimal trait expression. For instance, Rapkin et al. (2018) proposed a model which relied on slopes of linear

regressions as components of a vector to compare the strength of nutritional trade-offs between traits, estimated as the angle θ between vectors. Likewise, Morimoto & Lihoreau (2019) used a similar idea but instead of slopes, used the position coordinates of the regions of interest (i.e., peaks or valleys) as components of a vector and estimation of θ. The coordinates for the region of interest was found using Machine Learning support-vector-machine (SVM) with radial kernel. del Castillo et al. (2022) used bootstrapping approaches to identify the region within the performance landscape that represented the landscape optimal, which could then be compared with the similar region in other traits using confidence regions generated by the method (see also Hawkes et al. 2022). This approach had the advantage of not relying on any particular parametric distribution for generating estimates. More recently, Pascacio-Villafán et al. (2022) implemented a standard optimisation algorithm from response surface modelling approaches to identify and compared regions of interest in the landscape, primarily focusing on the comparison between peak regions. While these previous analytical models are useful, they have focused on identifying and comparing either peaks or valleys in performance landscapes, an approach that might have overlooked other properties of the landscapes which can contain biologically relevant information. The model proposed here addresses this limitation, as it estimates curvature properties from the performance landscapes. Importantly, we showed how the framework can be used to estimate surface-area and curvature of performance landscapes, which can aid interpretation and generate important predictions of animal responses to nutrition. For instance, as explained above, surface-area can be an important indicator of wiggliness in performance landscapes, as landscapes with more oscillations also have higher surface-area relative to flat landscapes (e.g., increased surface-area of cells due to microvilli). By estimating surface-area, it might be possible to predict how animals might respond physiologically and behaviourally to various nutrient imbalances. Similarly, curvature (both Gaussian and mean curvatures) can also enable further characterisation and prediction of responses to nutritional imbalances. Landscapes with regions of high (or low) curvature can lead to terrains in which animals are expected to experience higher costs of nutrient imbalances than regions in landscapes with low curvature (e.g., inclined plane). In our analysis of the empirical dataset of life-histories in *D. melanogaster*, we found that only the landscape of daily egg production showed relatively high local curvature at low nutrient intakes. This coincided with the valley region in the performance landscape, and confirms the non-linear costs of low nutrient intake in the expression of this trait. More empirical studies as well as more data for performance landscapes are needed to develop a more intuitive relationship between analytical properties of the performance landscapes (e.g., local curvature) and the broader patterns of animal nutrition (Morimoto & Lihoreau

2020). Nevertheless, we showed that our approach can be a powerful ally to characterise properties of performance landscapes that can aid biological insight.

Studies using GF have used a range of standard statistical models to analyse the data, but to date, the performance of these models had not been properly scrutinised. In particular, linear models (LM), generalised additive model (GAM), and thin-plate splines (TPS) have been the most commonly used approaches in GF studies (see e.g., Polak et al. 2017, Ponton et al. 2015, Lee et al. 2008, Maklakov et al. 2008, Kutz et al. 2019, Ma et al. 2020, Fanson et al. 2009, Morimoto & Wigby 2016, Ng et al. 2018, Solon-Biet et al. 2014, 2016, and references therein) (see also Ruohonen et al. 2007) for cubic splines. In this study, we had to assume that performance landscapes could be approximated using a general linear model of the form $ax^2 + by^2 + cx + dy + exy$ in order to calculate surface-area and curvature. We tested this assumption by measuring the performance of LM, GAM, and TPS onto GF data, and showed that indeed, a linear model can be as good of a model to GF data as the more complex GAM and TPS. This is important because the equation $ax^2 + by^2 + cx + dy + exy$ is differentiable and enables easy calculations of surface integrals and gradients for the estimates of curvatures. We have recently analysed the performance of several statistical (machine learning) models and their performance in identifying peak and valley regions in performance landscapes (Morimoto et al. 2023). The good quadratic approximation using LM presented in this study agrees with our extensive comparison of model performance in GF datasets (see also Simpson et al. 2004). Therefore, the results suggest that linear models are, in principle, reasonable approximations for the analysis of GF datasets for characterisation of regions in the performance landscape. With the growing use of GF in nutritional studies, it is crucial that the raw data to construct the performance landscapes are made open access as this will enable us to test whether the quadratic approximation presented here are suitable for other performance traits and species.

It is important to mention that the model proposed here, and all previous models developed in the literature which rely on - or estimates properties from - performance landscape assume that the landscape itself can be estimated accurately. This may not necessarily be the case for a standard GF design which relies on nutritional rails and explores a subset of all possible regions in space, leaving large parts of the space unexplored (particularly in regions that correspond to the interactions between nutrients, that is, along the diagonal of the nutrient space) (Morimoto 2022). The rationale for empirically testing a subset of diets and hence, regions of the nutrient space, is that results are reliable only if ecologically relevant ranges of diets are tested (see Point 4 in Raubenheimer & Simpson 2019). More studies are needed to

understand the consequences of imbalanced diets beyond those in which animals have evolved. We argue that exploring more combinations of diets - and hence, regions of the space - can enable more accurate approximations of performance landscapes and consequently, better predictions of landscape properties. This may come with higher practical cost (e.g., higher number of experimental diets needed) which is justified if better performance landscapes are generated as outputs. Optimal experimental designs in nutritional studies is an active research area (Ruohonen et al. 2001) (see also Ruohonen et al. 2007, Raubenheimer & Simpson 1993, Simpson & Raubenheimer 1993, 2012).

In conclusion, we proposed a novel model to measure curvature properties of performance landscapes using the Geometric Framework for Nutrition (GF). The model estimates surface-area and curvature of performance landscapes and, for the first time, estimates properties of performance landscapes other than peaks and valleys. This opens the range of parameters that can be estimated from performance landscapes, which may prove important for biological insights. We also introduce the use of the Hausdorff distance as metric for pairwise comparison of performance landscapes. The methodology proposed here can be employed to landscapes in general, for comparisons within and between species (for any fitness-related trait) and thus, can play an important role in our understanding of the responses to nutrition across the animal kingdom (Raubenheimer & Simpson 2020).

5.7 Acknowledgements

We are greateful to Prof Kwang Lee for providing the raw data used for the empirical demonstration of our model in this manuscript. The authors are grateful to Dr Dawn Shewring who made invaluable comments on the grammar of the main text.

5.8 Funding

JM is supported by the BBSRC (BB/V015249/1), a Royal Society Research Grant (RGS-R2-202220) and a Riemann Fellowship. PC is supported by the EPSRC (EP/P025072/) and from Ecole Polytechnique Federale de Lausanne via a collaboration agreement with the University of Aberdeen.

5.9 Competing interests statement

The authors have no conflict of interest to declare.

5.10 Data availability statement

The data used in the paper is available online at: `https://datadryad.org/stash/dataset/doi:10.5061\%2Fdryad.tp7519s`.

5.11 Appendix: Supplementary material.

Text S1. Primer on curvatures. Geometry has its foundation on the study of angles and distances on a space, which goes back when Euclid wrote his influential *Elements* around 300 B.C. Topology, on the other hand, is the study of spaces without dealing with angles and distances. The notion of curvature has a geometric nature (as we have seen, it can be described using triangles and the plane, spherical and hyperbolic model spaces) and represents one of the many interactions of geometry and topology. Whereas the latter deals mostly with global properties of space, the former focuses on local properties. In many situations is possible to extract global from local information, as we shall outline here. We will treat two concepts of curvature: the Gauss and mean curvature. Moreover, we will deal only with surfaces. The idea of the supplementary material is to briefly explain the intuition behind the *alphabet soup* kind of formulas used in the main text (without getting lost in details) and to motivate the reader to explore the world of geometry and topology. For a more rigorous account, please consult (Do Carmo 2016).

In continuous surfaces, curvature at each point p is defined via two *principal curvatures*. The principal curvatures can be geometric visualized as follows. One can choose a (unit) normal vector at a given point p and trace a normal place at p, that is, a plane that contains this chosen normal vector. This determines an unique tangent direction to the surface and a plane curve called *normal section*. This curvature will have different *curvatures*. By curvature here we mean the reciprocal of the radius of osculating circle, that is, the tangent circle at the point p that approaches the curve the most. The principal curvatures κ_1 and κ_2 are the maximum and the minimum values of this curvature, respectively.

One can also define the principal curvatures algebraically. It turns out that κ_1 and κ_2 at a point p are the eigenvalues of a linear operator (i.e., can be seen as a matrix) called the *shape operator*.The shape operator at point p is defined as

$$\mathcal{S}(v) = \nabla_v(n)$$

where v is a vector tangent to p, ∇_v is the directional derivative and n is the unit normal vector field of

a surface M. As $\nabla_v(n)$ is still a tangent vector at p, the shape operator is a linear operator

$$\mathcal{S}_p : T_pM \to T_pM,$$

where T_pM is the tangent space at a point p of M, that is, a vector space that encapsulates all the possible directions which pass tangentially through p. If we recall linear algebra classes, it makes sense because the eigenvectors measure the distortion of a linear operator, in our case, the eigenvalues of the shape operator tell us the how much the surface bends.

Finally, we define the **Gauss curvature** to be the product of the eigenvalues $\kappa_1\kappa_2$ and the **mean curvature** to be $\frac{\kappa_1+\kappa_2}{2}$, the sum mean of the eigenvalues (Fig 5.5). Recall that the product of eigenvalues is the determinant of matrix and the sum of eigenvalues is the trace. Hence, we can say that the Gauss curvature is the determinant of the shape operator and the mean curvature is the sum mean of the trace of the shape operator.

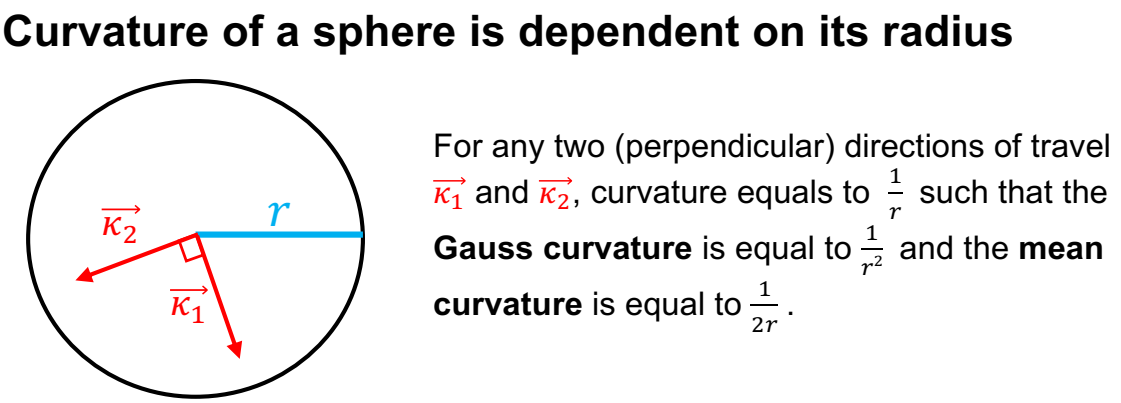


Figure 5.5: **Curvature of a sphere is dependent on its radius.** For any two (perpendicular) direction of travel, curvature of a sphere (here projected in 2D) is $\frac{1}{r}$ and the Gauss and mean curvatures are equal to $\frac{1}{r^2}$ and $\frac{1}{2r}$

Recall again from linear algebra, that the determinant of a matrix is invariant under change of bases (and row columns operations). Therefore, the Gaussian curvature should also inherit this invariance! Indeed, the famous *Theorema Egregrium* proved by Gauss tells us that the Gauss curvature does not depend on the embedding of the surface in the 3-dimensional Euclidean space. Hence, we say that the Gauss curvature is an *intrinsic* property of a surface. Moreover, the theorem tells us that the Gauss curvature is invariant under local isometries, that is, it does not change if one bends the surface without stretching it. A consequence of this theorem illustrates the local to global interplay of geometry and topology. It can tell us when two spaces are different, a sphere of radius r has Gauss curvature $\frac{1}{r}$ and a

plane has Gauss curvature 0. However, it can not tell when two spaces are equal: a plane is isometrically transformed into a cylinder, hence both have Gauss curvature 0 even though are clearly different surfaces.

Due to this pitfall of Gauss curvature, it is desired to compute also the mean curvature of a surface while analysing our nutritional landscapes. As it is defined as the sum mean of the trace of the shape operator, it does depend on the embedding of the surface and, therefore, receives the name of *extrinsic* measure of curvature of surface. It does tell us interesting properties of a surface. For instance, a surface is called *minimal surface* if and only if it has mean curvature 0 at every point. There are many equivalent definitions of minimal surfaces but, roughly, a minimal surface is a surface that minimizes the total area subject to some constraint, for instance, some surfaces formed by soap films. The theory of minimal surfaces is vast and has interesting application inside and outside mathematics.

Another important theorem connecting geometry and topology is the *Gauss-Bonnet theorem* connects the Gauss curvature with another important topological invariant, the *Euler characteristics*. The precise statement is not of importance here but its message, which tell us how to compute a global invariant (Euler characteristic) via local quantity (Gauss curvature). Moreover, it says that if you deform your surface without tearing it apart, the global invariant (i.e., Euler characteristic) remains the same.

After this conceptual detour, let's return to the aim of our work. We are interested in computing the curvature for the nutritional landscapes. For that, there is a way of deriving formulas for the Gauss and mean curvature for surfaces which are more suitable for computations and especially coding, with less of a conceptual flavour. Firstly, we need to compute the so-called *first and second fundamental forms*. We defined and explained them in the main text. We opt to use those formulas in particular because our surfaces are modelled by polynomials and we can derive the first and second fundamental form via a routine method that does not require a lot of computational power and, consequently, the Gauss and mean curvature.

Chapter 6

Optimum ratio of dietary protein and carbohydrate that maximises lifespan is shared among related insect species

Publication reference

Morimoto, J., (2024). Optimum ratio of dietary protein and carbohydrate that maximises lifespan is shared among related insect species. *Aging Cell*, 23(3), p.e14067.

6.1 Abstract

Animals often regulate the intake and quantity of nutrients to maximise fitness through life-history traits such lifespan, but we still lack a proper understanding of how specific nutrients influence these traits. Here, I developed an algorithm which allowed me to create a nutrient-specific database from literature data, and investigated how the requirements of protein (P) and carbohydrate (C) needed to maximise lifespan evolved across nine insect species. I found moderate evidence of a phylogenetic signal on the optimal ratio of protein to carbohydrate (PC ratio) that maximised lifespan, suggesting that optimal PC ratio for lifespan could have evolved non-independently among related species. I also found evidence for weak-to-strong sex-specific optimal PC ratios for lifespan, suggesting that sex-specific nutritional needs to maximise lifespan can emerge and persist in some species. Although limited in the number of species, the approach adopted here is portable to experiments with n number of nutrients and, thus, can be used in complex comparative precision nutrition studies for insights into the evolution of animal nutrition.

Keywords: Ecological specialization; Aging; precision nutrition;

6.2 Background and findings

Different animals have different nutritional needs (Wu 2017, Sauvant et al. 2008). This appears obvious at a coarse level since animals feed on different food types, but how nutritional needs differ among species at the nutrient level remains an open question. Nutrient-specific needs are not always evident from the food type of a species. For instance, animals that seemingly feed on different food types can converge towards similar nutrient profiles, as is the case of the giant panda which, although classified as a strict herbivorous species, has evolved a protein-biased diet akin to that of carnivores (Nie et al. 2019). Thus, it is important to look beyond food types and into nutrient-specific patterns to gain insights into the evolution of nutritional needs across the tree of life (Machovsky-Capuska et al. 2016, Gonzalez et al. 2017).

Comparative nutrition has long been recognised as a powerful approach to uncover how nutritional needs have evolved (Mitchell et al. 1962, Baker & Czarnecki-Maulden 1991, Payne & Wheeler 1968). It was used to provide insights into the effects of both macro- and micro-nutrients on health (see Baker & Czarnecki-Maulden 1991, Lei et al. 2022, references therein), including the effects of protein on morphological and reproductive traits (Talal et al. 2023, Swanson et al. 2016). Comparative nutrition studies are few because it is difficult to obtain a single dataset that maps the effects of specific nutrients to life-history traits, such as lifespan or reproduction. A more parsimonious approach is to identify multiple but comparable studies on different species which allow for comparative studies. A recent study by Swanson et al. (2016) showed the effects of nitrogen, phosphorus and sodium across 96 butterfly species, and found that species feeding on high nitrogen diets are more fecund but have smaller eggs, potentially highlighting the evolution of a nutritional trade-off. A limitation in this study is that dietary nitrogen may be used as proxy for protein intake and protein needs, but these are not necessarily equivalent (Mattson Jr 1980). Nakagawa et al. (2012) used the comparative approach to study lifespan extension via dietary restriction across 36 species, and showed the proportion of protein intake had greater influence on lifespan extension than the degree of caloric restriction. These studies highlight the need for studies using nutrient-specific approaches.

The Geometric Framework for Nutrition (GF) is an experimental framework which allows for the study of nutrient-specific effects on life-history traits in animals and humans (Raubenheimer & Simpson 1993, Simpson & Raubenheimer 1993). GF is portable to any species, which makes GF an attractive framework for insights into the evolution of nutrition (Raubenheimer & Simpson 2020, Simpson & Raubenheimer

2012) as well as a potential key for precision nutrition experiments and its future applications (Simpson et al. 2017). GF has gained popularity and has underpinned major insights into animal and human nutritional ecology over the last decades, primarily from studies in insect species but alse more recently in vertebrate models and humans (see e.g., Barragan-Fonseca et al. 2019, Pascacio-Villafán et al. 2022, Simpson et al. 2017, Behmer 2009, Lee et al. 2008, Maklakov et al. 2008, Rapkin et al. 2018, Ponton et al. 2015, Polak et al. 2017, Solon-Biet et al. 2014, Simpson & Raubenheimer 2005, Simpson et al. 2003, Behmer 2009, Ng et al. 2018, Bradbury et al. 2014, and references therein). Thus, GF datasets are unique in their potential to reveal broad-scale patterns of nutrition, making GF the ideal candidate for large-scale comparative nutrition studies. However, open data practices in the field remains poor, preventing broad-scale comparative studies (Morimoto & Lihoreau 2020). We therefore lack the means to study how optimum nutrient intakes evolved using a systematic and nutrient-explicit comparative approach.

I conducted a systematic literature search to assess the availability of raw data in the GF literature (see 'Experimental procedures'), I found that out of 32 studies using the GF framework to study the ratio of protein (P) and carbohydrate (C) in the diet and its effects on lifespan in male and/or female insects, only five (29.4%) provided direct access to the raw data (see Fig 6.1a). This is in line with my previously report on the open practices in the field which showed that only ca. 39% of general (i.e. not limited to a particular trait) GF studies made their raw data available (Morimoto & Lihoreau 2020). Among the identified studies, there was a notable sex-bias: all (100%) studies investigated female lifespan in response to diet, while only eight studies (25%) also investigated male lifespan in response to diet.

Open data practices are widely recognised as an imperative for research breakthroughs (Lowndes et al. 2017, Murray-Rust 2008, Reichman et al. 2011). Thus, innovative ways that open GF datasets for comparative studies are urgently needed to advance our understanding of animal nutrition. To overcome the lack of data availability, I developed a semi-automated algorithm which enabled me to reconstruct published performance landscapes from GF studies that do not make their raw data available (see Fig 6.1b and 'Experimental procedures' section for details). While this does not replace the need for better open data practices, it allowed me to create the first systematic precision nutrition database for comparative nutrition insights (Morimoto & Lihoreau 2020). For practical and financial reasons, GF has been most commonly used in insect studies and this study and database focused on this group, although it is portable to any animal group. The algorithm allowed me to create a database with optimal PC ratios for lifespan across nine insect species for females and five species in males.

I then tested the hypothesis that the optimal PC ratio that maximise lifespan is shared among related

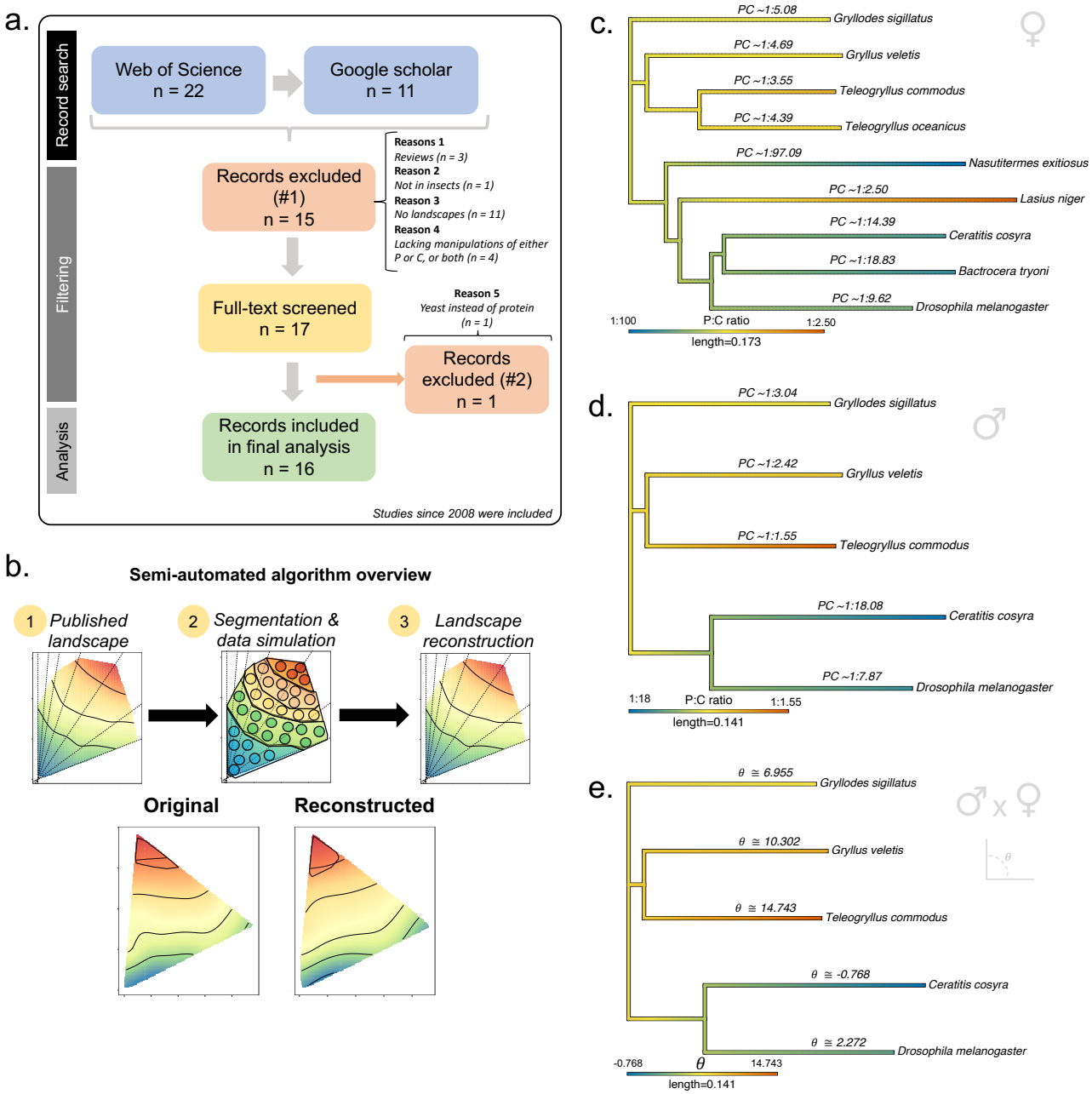


Figure 6.1: **Optimum PC ratios across related insect species** (a) Flowchart for the systematic literature search of experimental studies in insects that used GF to investigate the PC ratio effects on lifespan. (b) Overview of the semi-automated algorithm to extract information of performance landscapes for lifespan from the published GF literature. (b) Phylogeny mapping the PC ratio that maximises lifespan across insect species for which GF data is available in the published literature (females). (c) Phylogeny mapping the PC ratio that maximises lifespan across insect species for which GF data is available in the published literature (males) (d) Phylogeny mapping the angle θ, which is a metric of nutritional trade-off, for PC ratios that maximise lifespan across insect species in the database (note that the phylogeny is limited by the sex with the least amount of information. i.e. males). Negative values for the angle θ indicates that female optimal PC ratio was lower than the optimal PC ratio of males.

species, that is, optimal PC ratio is non-independent among species with shared evolutionary history. I measured phylogenetic signal, which is a measure of non-independent evolution of a continuous trait (in this case, optimal PC ratio) using two common metrics: (a) Pagel's λ, which is well-defined within the range of 0 (no signal) and 1 (strong signal) and (b) Blomberg's K which varies between 0 and ∞ and is

proportional to the strength of the phylogenetic signal, whereby $K = 1$ indicates no difference between the trait evolution as expected under Brownian motion. These metrics measure different aspects of the phylogenetic signal, the former measuring a scaling parameter relative to expectation under Brownian motion and the former, the normalised ratio of variances within and between clades over the expected variances under Brownian motion (see e.g. Revell et al. 2008, Revell & Harmon 2022, Munkemuller et al. 2012, for more details). Because of the female-bias representation in the database, the mapping of the optimal PC ratio for lifespan onto the phylogeny was more extensive for females than males, and was done separately, giving less precise (i.e. higher variance) estimates of the posterior distribution of the phylogenetic signal for males. I used the Nutrigonometry model to the reconstructed performance landscapes to compute the optimal PC ratio region in the performance landscapes that maximised lifespan in each species (Morimoto et al. 2023, 2022, Morimoto & Lihoreau 2019) (see 'Experimental procedures' section for details).

The results revealed a moderate phylogenetic signal for the PC ratio that maximised female lifespan (Blomberg's K: 1.265; p = 0.021; MCMCglmm Pagel's λ: 0.580, HPD interval: [0.222, 0.938]) and male lifespan (Blomberg's K: 1.228; p = 0.032; MCMCglmm Pagel's λ: 0.445, HPD interval: [0.021, 0.901]). In females, orthopterans had extended lifespan when fed at intermediate PC ratios. Within holometabola, termites and dipterans had extended lifespan in low PC ratios while the ant *L. niger* showed the opposite effect, with extended lifespan at high PC ratios (Fig 6.1c). In males, the patterns were similar despite the lack of hymenopterans (see Fig 6.1d). These results must be interpreted with caution though, due to the relatively small number of species used in the comparative model to estimate phylogenetic signal. This led to relatively wide credible intervals for the estimates of Pagel's λ as well as a noticeable sensitivity to prior specifications on the estimates of Pagel's λ (see Dialog S1 in the Appendix of this chapter). With more data being generated in the field, these estimates will inevitably become less variable. Nonetheless, these results show for the first time that the optimal PC ratio for lifespan may be shared among closely related insect species.

Species that live longer in diets with higher or lower PC ratio may also feed on diets with similar PC ratios when given a choice. However, I found no evidence of a positive relationship between optimal PC ratio that maximise lifespan and self-regulated PC ratios in both sexes in the database (*Females:* Mean posterior[95% CI]: 0.405 [-0.452, 1.323]; *Males:* -0.531 [-4.041, 4.024]). This does not imply that the relationship between optimal PC ratio and self-regulated PC ratio is absent: multiple studies in single insect species have shown that, when given a choice, self-regulated PC ratios often match more closely

the optimal PC ratio for reproductive traits (see e.g., Jensen et al. 2015, Lee et al. 2008, Maklakov et al. 2008, Hawkes et al. 2022, and references therein). Hosking et al. (2019) showed through simulations that self-regulated PC ratio that optimises reproduction should be prioritised under certain evolutionary contexts. Thus, it is possible that the relationship between optimal PC ratio and self-regulated PC ratio exist for reproductive traits, although this remains to be tested.

Optimal PC ratios may also vary within species between sexes, which creates the potential for sexual conflict over optimal nutrition (Jensen et al. 2015, Carey et al. 2022). This can lead to long-term 'tug-of-war' between sexes nutrient intake or alternatively, lead to the evolution of physiological and behavioural adaptations to resolve – or mitigate – the negative effects of the conflict when both sexes cannot achieve their optimal nutrition (Hawkes et al. 2022, Simpson & Raubenheimer 1993, 1999). Using the database, I tested whether there was evidence for a phylogenetic signal in the sex-specific optimal PC ratio for lifespan in insects. I used the recently developed Nutrigonometry model to compare nutritional trade-offs in the PC ratio (angle θ) and the quantity (the hypothenuse) of the diets that extended lifespan. There was evidence of moderate phylogenetic signal for the angle θ but not for the hypothenuse, captured by Blomberg's K but not Pagel's λ (θ: Blomberg's K: 1.264; p = 0.027; MCMCglmm Pagel's λ: 0.0564, HPD interval: [0.00, 0.548]; *Hypothenuse*: Blomberg's K: 0.900; p = 0.792; MCMCglmm Pagel's λ: 0.0002, HPD interval: [0.000, 0.001]. The nutritional trade-off measured by θ appeared to be stronger in Orthopterans, particularly *T. commodus*, compared to Dipterans (see Fig 6.1e). These results show that sex-specific nutritional needs are likely related to the nutrient balance rather than nutrient quantity.

Physiological and behavioural traits that regulate optimal PC ratio intakes are likely under selection. The findings that optimal PC ratio is shared among species also suggest that the underlying physiological machinery to cope with diet balances can also be shared. The database created here did not allow me to investigate shared molecular pathway activities underpinning the effects of PC ratios on lifespan. However, it is likely that the nutrient-sensing mTOR pathway is involved on the evolution of optimal PC ratio intake on lifespan, considering taxa-specific and comparative evidence for the role of this pathway in diet-dependent lifespan extension (e.g., Garratt et al. 2016, Ma & Gladyshev 2017). Furthermore, a recent comparative transcriptome study across 14 Drosophila species also suggests that lifespan regulation involves a complex orchestra of system-level gene expression and trade-offs (Ma et al. 2018), suggesting that physiological trade-offs emerging from responses to diet may also be shared among closely related species. Future studies should test whether mTOR expression level is linked to the PC ratios that extend lifespan across species and whether mTOR expression – and the expression and trade-offs across

nutrient-sensing pathways – evolves non-independently along with PC ratios. Diet-dependent comparative transcriptome analysis will shed light onto the pathways and trade-offs that are conserved in the responses to diet. It is also worth mentioning that some traits are plastic and the location of peaks in GF landscapes might vary depending on environmental conditions such as temperature (e.g., Kutz et al. 2019, Lee & Roh 2010, Chakraborty et al. 2020). This has been shown for developmental traits such as developmental time and growth rate but we do not know if similar patterns are seen for lifespan. This remains subject of future studies.

Genetic constrains may prevent rapid evolution of dietary choice as well as the physiological machinery to resolve sex-specific optimal diet intake. For example, Hawkes et al. (2022) showed that positive genetic correlations under diet choice can accelerate (males) or constraint (females) evolutionary responses to PC ratio intake in *G. sigillatus*. In *D. melanogaster*, both sexes display considerable heritability on the intake of both protein and carbohydrates (Reddiex et al. 2013). However, it is unknown how selection in response to diet acts between sexes across species, making it difficult to hypothesise the underlying factors driving the maintenance or resolution of sexual conflict over nutrition. A recent nutrient-explicit model showed that selection acting on macronutrient appetite can modulate PC ratio intakes and influence the prioritisation of life-history trait optimisation under different ecological conditions in an asexual population (Hosking et al. 2019). A generalisation of this model to include how intake targets evolve in sexual populations with sex-specific nutritional needs for optimal expression of competing life-history traits will provide important insights into the evolution and maintenance of sexual conflict over nutrition.

The main limitation of this study is the relatively low number of species for which GF lifespan landscapes are available. It was possible to retrieve GF landscapes only for nine species of insects in the female database and five species in the male database. With the low sample size, it is not surprising that the MCMC models to estimate phylogenetic signal displayed sensitivity to prior specifications (see Dialog S1). Nonetheless, in all cases, even non-informative priors found on average a moderate estimate for the Pagel's λ as measure of phylogenetic signal (Dialog S1). Is in unquestionable that more studies, in different species, are urgently needed, as they will allow for more robust estimates of phylogenetic signals in the optimal PC ratio maximising lifespan and other life-histories. It is also worth noting that the diet formulations of the studies identified here varied, with studies in orthopterans relying on powdered diets derived from (Simpson & Abisgold 1985) while studies for other groups using primarily gel-based diets (but see Carey et al. 2022). The impacts of the diet formulation on physiological responses and lifespan were assumed to be negligible, but this remains a key factor to be tested empirically in the field.

In conclusion, I started off from the common knowledge that different animals eat different foods and asked the question: is there one ratio of protein and carbohydrate that maximises lifespan across species? Using insects as case study, I showed that there may exist such shared optimal PC ratio for lifespan among related species. This is likely to be true for other life-histories, making the comparative approach proposed here an essential part of future precision nutrition predictive models. Although relatively modest in the number of species compared to other non-GF comparative nutritional studies (Swanson et al. 2016), this study adopts a direct nutrient-specific approach and lays the foundation for the use of GF nutritional data to gain insights into the evolution of nutrition needs across taxa.

6.3 Experimental procedures

Literature search in the Web of Science database were conducted to identify all studies since 2008 that used GF to study lifespan (Search terms: Geometric Framework, nutrition, lifespan OR survival [All fields]). The search was supplemented with new entries from Google Scholar search with same terms (Fig 6.1a). In total, 32 studies were identified from which 17 complied with the inclusion criteria, which were:

1. Presents experimental data;
2. Conducted in an insect species;
3. Displayed GF performance landscapes for lifespan
4. Presents results for diet manipulations of protein and carbohydrate ratios (PC ratios)

Criteria number (2) was adopted because there is a general lack of controlled experimental studies using GF to assess lifespan in non-insect groups ((but see e.g. Solon-Biet et al. 2014). Methods and data availability statements for each of the included studies were read and raw data was downloaded when available for model validation. The female database included nine different species across three insect orders (Orthoptera, Hymenoptera, and Diptera). The male database included eight studies, from which five species and two orders (Orthoptera and Diptera) were represented. Two studies presented multiple landscapes for lifespan and a decision was made as to which landscape to include in the present study. For instance, Fanson & Taylor (2012) presented lifespan performance landscapes for mated and virgin females, while Poissonnier et al. (2018) presented two lifespan performance landscapes for worker and soldier termites. In these cases, I selected the performance landscapes for virgins and workers, respectively, as representative of these studies and species. One study in *D. melanogaster* was excluded from the analysis

because the performance landscape for lifespan presented in the original study was given as yeast-to-sugar, rather than protein-to-carbohydrate ratio (Skorupa et al. 2008). The list of all studies used in this analysis is provided as supplementary information (Table 6.1).

Data extraction and analyses were conducted in R version 4.1.3 (Team 2019) (see Fig 6.1b). The semi-automated algorithm to extract simulated data to reconstruct the published lifespan performance landscapes worked as following:

1. The algorithm segmented the performance landscapes with respect to the z-axis using the 'recolorize' 0.1.0 package (Weller et al. 2022). In this study, the z-axis corresponds to lifespan values.

2. Each z-layer was manually assigned a value for lifespan based on the contour lines depicted in the image on the published literature. This ensured that the values and ranges of lifespan matched that of the original landscape.

3. This approach 'categorises' the values for lifespan, leading to a z-axis that was ladder-like. This format hinders the estimates of the performance landscape using thin-plate splines from the 'fields' package (Nychka et al. 2017). To overcome this, I added random error to the categorised lifespan values by sampling from a normal distribution using the 'rnorm' function, with mean 0 and standard deviation 0.25. Note that, without the raw data, it is currently impossible to estimate the true variance underlying the data used to build performance landscapes (see comment below).

4. A filtering function was applied to remove noise introduced during the segmentation step which could lead to simulated points existing outside the true region of the original performance landscape.

5. The reconstructed landscapes were drawn using a custom-made function which is a wrapper for the 'Tps' function from the 'fields' package (Nychka et al. 2017).

Note that without the raw data, it is virtually impossible to estimate the variance in the z-variable. By simulating the standard error from a normal distribution with mean 0 and standard deviation 0.25 throughout, I assumed that variance in lifespan across species and diets with varying PC ratios were constant. This might not reflect reality, but it is nonetheless an essential assumption for data interpolation and identification of optimal PC ratios in reconstructed performance landscapes. Future studies to reconstruct the true variance from published images are needed. Landscapes were plotted using the 'ggplot2' 3.4.2 package (Wickham et al. 2016). Data handling was conducted using the tidyverse packages 'dplyr' 1.1.2 and 'tidyr' 1.2.0 (Wickham et al. 2019).

Raw data for seven studies were available and allowed me to validate the algorithm above. Validation was done using the Nutrigonometry model to the reconstructed and original lifespan performance landscapes and measuring whether the angle θ and the hypothenuse of the identified peak region differed between the original and reconstructed landscapes. Average estimates along with 95% confidence intervals for the estimates were calculated from the model. Zero-overlapping estimates, which were found for all seven studies, showed that the reconstructed landscapes displayed its peak region in a comparable location to the landscapes made using the original data (Fig 6.2 amd 6.3. Validation was done using the female database due to accessibility to a higher number of raw data for female landscapes. Given the same methodology was implemented to extract and reconstruct both male and female landscapes, and male and female landscapes do not vary dramatically, this validation approach is suitable. Reconstructed landscapes and their estimated peak regions used in this study are given in Figure 6.4.

Phylogenetic relationship between the species in the database was reconstructed using the whole or partial mitochondrial genome (i.e. the barcoding CO1 gene sequence) obtained from the NCBI database, using the packages 'ape' 5.6-2 (Paradis & Schliep 2019), 'phytools' 1.2.0 (Revell 2012) and 'phyloseq' 1.38.0 (McMurdie & Holmes 2013). Multiple sequence alignment was done using the 'msa' 1.26.0 package (Bodenhofer et al. 2015). Blomberg's K and its statistical significance were estimated using the 'phylosig' function in the 'phytools' package; phylogenetic tree with continuous trait mapping were made using the 'contMap' function from the same package. Pagel's λ was estimated using the 'MCMCglmm' package and the function with same name (Hadfield 2010). Because of the unequal number of species, phylogenetic signals were estimated for each sex separately.

MCMCglmm models with gaussian family error distribution were fitted using optimal PC ratio as response variable with a model with fixed intercept and study and species as random variables, accounting for the non-independent variance-covariance matrix from the reconstructed phylogenetic relationships among species. Fixed effect prior parameters were set as V = 1 and nu = 1 while random effect priors were set at V = 1 and nu = 0.02. Burn-in was set at 2000, thin parameter at 20, and a total of 20000 iterations for all models. Convergence was analysed using autocorrelation for fixed and random effects as well as Gelman-Rubin diagnostics for among-chain convergence. Prior sensitivity tests were conducted for optimal PC ratio models using a range of values for the parameter nu (i.e. $10^{-}6$, 0.002, 0.02 and 0.2) (see Dialog S1 and Table 6.9). Sex-specific optimal PC ratio where compared using the angle θ and the hypothenuse, obtained using the Nutrigonometry model (Morimoto et al. 2023). θ and the hypothenuse were calculated within each study and species and the averages of these estimates per species were used

for the analysis of phylogenetic signal as described above. Phylogenetic signal for sex-specific optimal PC ratio could only be calculated for the species in which both male and female data were available and therefore, was limited to only a subset of the female database (see above). The MCMCglmm models to estimate Pagel's λ had either the angle θ or the hypothenuse as response variable and a fixed intercept, with study and species as random effects as described above. I used default MCMCglmm priors for these models. The relationship between optimal PC ratio for lifespan and self-regulated PC ratio was analysed using a MCMCglmm model with optimal PC ratio as response variable and self-regulated PC ratio as independent variable, controlling for the random effects of study, species, and observation (to account for repeated measures in the same study and species). I used the same values of V and nu for priors as in the above models with the addition of priors for the random effect of observation with parameters $V = 1$ and $nu = 0.002$. The highest posterior density intervals were calculated using the 'HPDposterior' function of the 'MCMCglmm' package for all MCMC models.

6.4 Acknowledgements

I am grateful to Prof Kwang Lee (Seoul National University) and Distinguished Prof Phil Taylor (Macquarie University) for the access to raw data of two of their articles in the past.

6.5 Funding

JM is supported by the Biotechnology and Biological Sciences Research Council [BBSRC: BB/V015249/1].

6.6 Declaration of interests

The author has no competing interests to declare.

6.7 Author contributions

J.M is the sole author of this manuscript and was responsible for conceptualising the model, coding, data analysis, manuscript preparation, and submission.

6.8 Data accessibility

Data on optimal PC ratio that maximise lifespan is available as supplementary material. Code for the semi-automated algorithm is not made available because the author, who developed the algorithm, is still using the algorithm to create and analyse other databases with similar approach to that described here.

Table 6.1: List of articles and associated species for which the performance landscape for lifespan was extracted using the semi-automated algorithm developed in this study.

Species	Year	Journal	Reference
Drosophila melanogaster	2022	BIOGERONTOLOGY	Carey et al. (2022)
Drosophila melanogaster	2015	AGING CELL	Jensen et al. (2015)
Teleogryllus oceanicus	2019	JOURNAL OF EVOLUTIONARY BIOLOGY	Ng et al. (2019)
Bactrocera tryoni	2009	AGING CELL	Fanson et al. (2009)
Ceratitis cosyra	2017	ECOLOGY AND EVOLUTION	Malod et al. (2017)
Teleogryllus commodus	2015	ANTIOXIDANTS	Archer et al. (2015)
Bactrocera tryoni	2012	PROCEEDINGS OF THE ROYAL SOCIETY B	Fanson, Fanson & Taylor (2012)
Bactrocera tryoni	2012	AGE	Fanson & Taylor (2012)
Teleogryllus commodus	2008	CURRENT BIOLOGY	Maklakov et al. (2008)
Drosophila melanogaster	2008	PNAS	Lee et al. (2008)
Drosophila melanogaster	2008	AGING CELL†	Skorupa et al. (2008)
Gryllus veletis	2014	PROCEEDINGS OF THE ROYAL SOCIETY B	Harrison et al. (2014)
Lasius niger	2012	PROCEEDINGS OF THE ROYAL SOCIETY B	Dussutour & Simpson (2012)
Gryllodes sigillatus	2022	FUNCTIONAL ECOLOGY	Hawkes et al. (2022)
Teleogryllus commodus	2017	EVOLUTION	Rapkin et al. (2017)
Drosophila melanogaster	2018	ENTOMOLOGIA EXPERIMENTALIS ET APPLICATA	Semaniuk et al. (2018)
Nasutitermes exitiosus	2018	FUNCTIONAL ECOLOGY	Poissonnier et al. (2018)

†: Study excluded from the analysis because performance landscapes were drawn using yeast:sugar ratios rather than protein:carbohydrate ratios.

6.9 Appendix: Supplementary material.

Dialog S1. MCMCglmm models. MCMCglmm models were used to estimate phylogenetic signal via Pagel's lambda. I conducted a prior sensitivity analysis where I ran models with priors of varying levels of confidence (i.e. V = 1, nu ranging from 0.2 to 10-6; see main text). These priors expanded on those tested in Nakagawa et al. (2012). Results suggest that estimates of Pagel's lambda were somewhat sensitive to prior specifications for both sexes (but more so for males which had smaller number of species), although the posterior mean for all models, irrespective of priors, found a moderate strength of phylogenetic signal (with wide credible interval estimates; see also main text for Blomberg's K). Because of my a priori belief, based on the plots of the optimal PC ratio onto the phylogeny during the exploratory data analysis stage, I selected model 2 ($V = 1$, $nu = 0.02$), which was a resonable balance between uninformative prior and a prior with more confidence, when presenting the results, while adding a cautionary tale for the interpretations of the results given prior sensitivity (see Chapter 6). Convergence of the final models were analysed using autocorrelation for fixed and random effects (which were close to or lower than 0.1) as well as Gelman-Rubin diagnostics for among-chain convergence, whereby I ran the

MCMCglmm with same specifications three independent times and compared the Gelman-Rubin score, which were 1.01 (upr: 1.02 and 1.01) for both male and female models, respectively. MCMCglmm models for theta and the hypothenuse, which used default priors, also converged (Gelman-Rubin: 1.01; upr: 1.01-1.02, Theta autocorrelation: intercept = 0.015, study = 0.02, species 0.00, units = 0.016; Hypothenuse autocorrelation: intercept = 0.03, study = 0.17, species 0.05, units = 0.004). Similarly, MCMCglmm models for the relationship between optimal and self-regulated PC ratios converged (Gelman-Rubin: 1.06; upr: 1.06, autocorrelation intercept = 0.019, self-regulated PC ratio: 0.021, study = 0.003, species = 0.051, units = 0.037).

Table 6.2: Prior sensitivity analysis for the phylogenetic signal on optimal PC ratio.

Sex	**Model**	**V**	**Nu**	**Sample size**	**Pagel's lambda**	**Lwr**	**Upr**
	1	1	0.002	900	0.332	0.0135	0.766
Females	2	1	0.02	900	0.58	0.2200	0.938
	3	1	0.2	900	0.808	0.5650	0.977
	4	1	10^{-6}	900	0.377	0.0000	0.888
	1	1	0.002	900	0.35	0.0000	0.899
Males	2	1	0.02	900	0.445	0.0210	0.901
	3	1	0.2	900	0.666	0.2680	0.992
	4	1	10^{-6}	900	0.18	0.0000	0.858

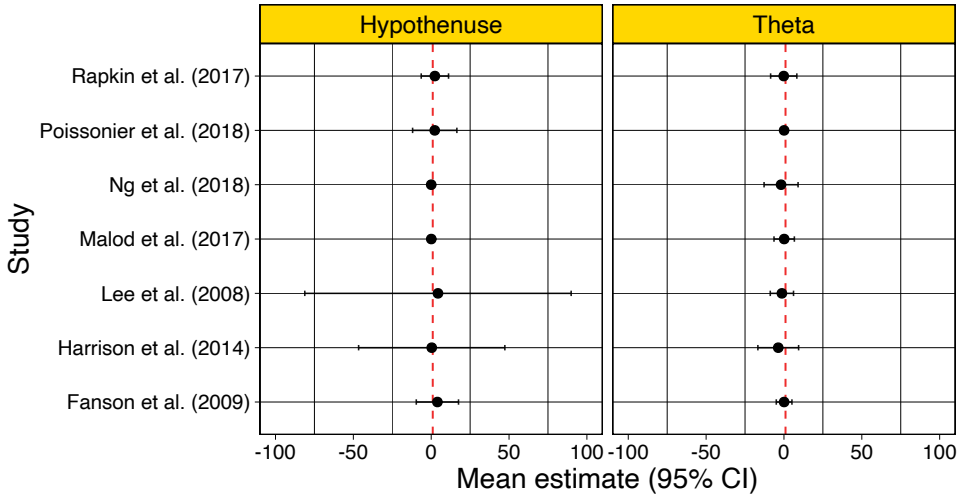


Figure 6.2: **Estimates of error between the peak regions of original (raw data) and reconstructed performance landscapes for lifespan.** Estimates of peak region were done using the Nutrigonometry model (Morimoto et al. 2023) Note that the credible intervals are relatively wider for the hypothenuse in Lee et al. (2008), Harrison et al. (2014) because of the quantity of nutrients eaten by the experimental animals in these studies were higher than the others. Importantly, the hypothenuse measures trade-offs over quantity and, therefore, has to be calculated in the same scale as the data was collected (i.e. cannot be standardised).

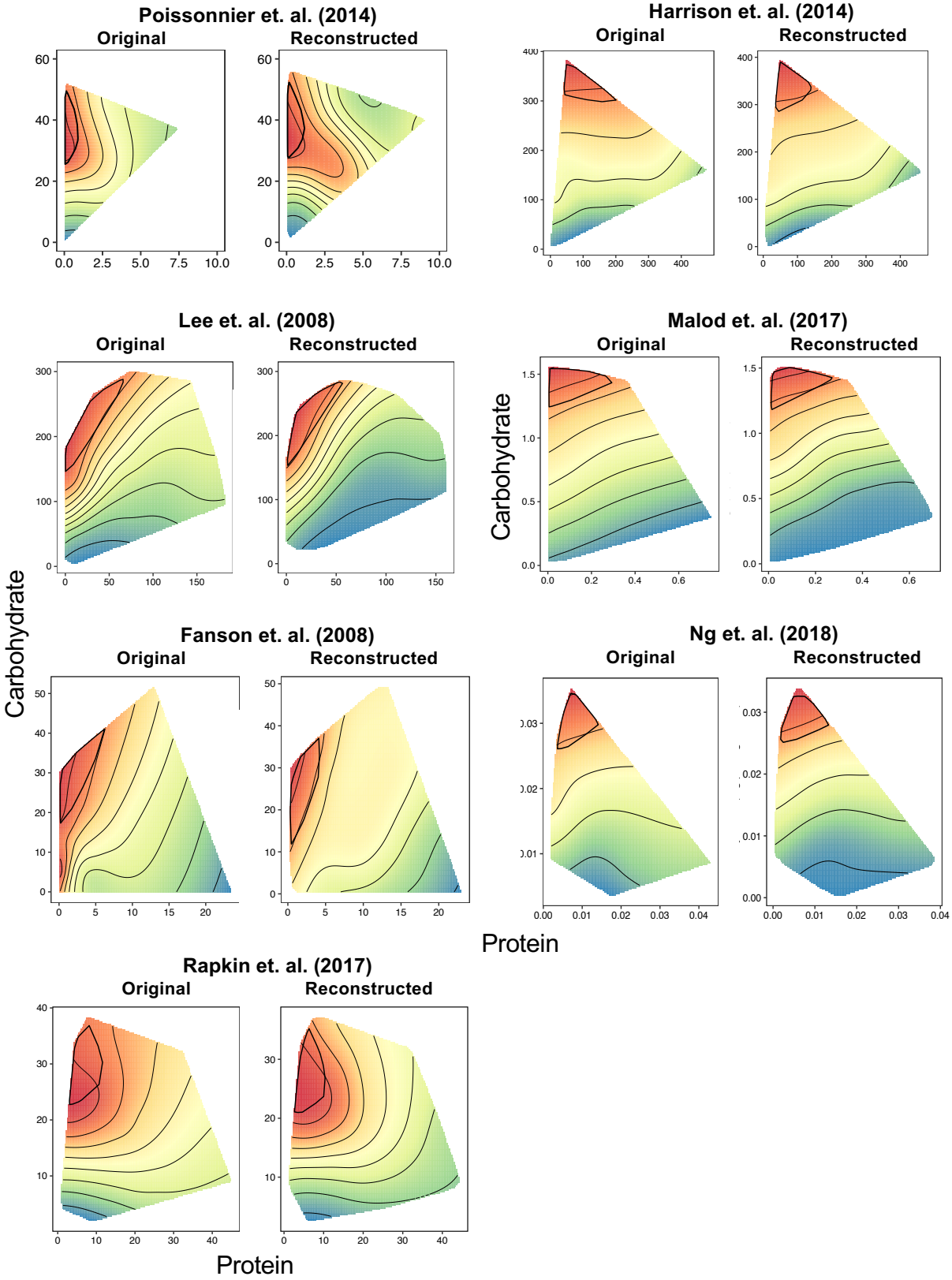


Figure 6.3: **Original and reconstructed landscapes. Red polygon represents estimated peak region (Nutrigonometry) (Morimoto et al. 2023).**

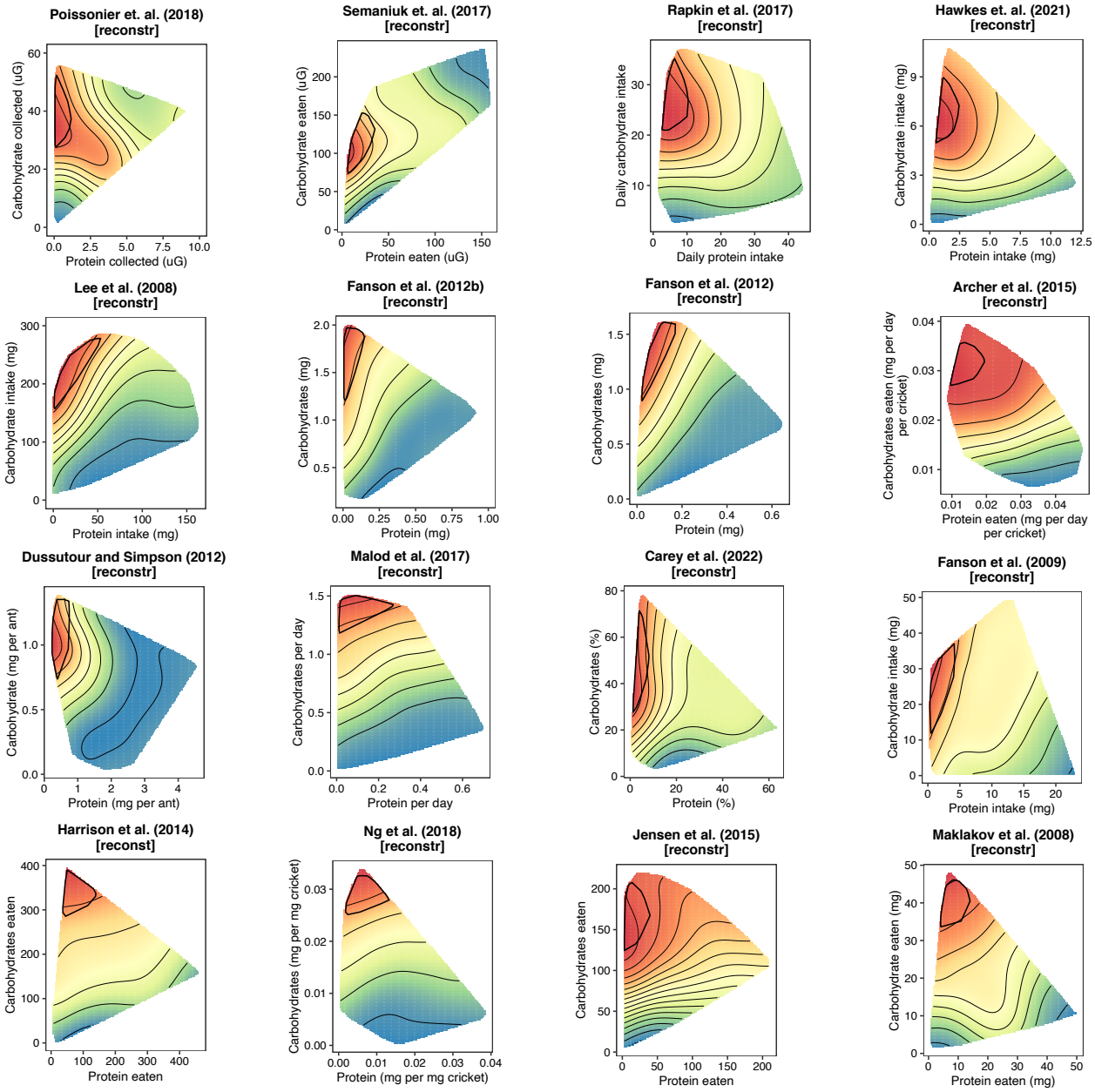


Figure 6.4: **Reconstructed landscapes used in this study. Red polygon represents estimated peak region (Nutrigonometry) (Morimoto et al. 2023).**

Chapter 7

Nutrigonometry IV: Thales' theorem to measure the rules of dietary compromise in animals

Publication reference

Morimoto, J., (2023). Nutrigonometry IV: Thales' theorem to measure the rules of dietary compromise in animals. *Scientific Reports*, 13(1), p.7466.

7.1 Abstract

Diet specialists and generalists face a common challenge: they must regulate the intake and balance of nutrients to achieve a target diet for optimum nutrition. When optimum nutrition is unattainable, organisms must cope with dietary imbalances and trade-off surplus and deficits of nutrients that ensue. Animals achieve this through compensatory rules that dictate how to cope with nutrient imbalances, known as 'rules of compromise'. Understanding the patterns of the rules of compromise can provide invaluable insights into animal physiology and behaviour, and shed light into the evolution of diet specialisation. However, we lack an analytical method for quantitative comparisons of the rules of compromise within and between species. Here, I present a new analytical method that uses Thales' theorem as foundation, and that enables fast comparisons of the rules of compromise within and between species. I then apply the method on three landmark datasets to show how the method enables us to gain insights into how animals with different diet specialisation cope with nutrient imbalances. The method opens new avenues of research to understand how animals cope with nutrient imbalances in comparative nutrition.

Keywords: Ecological specialisation, decision-making, diet choices, foraging

7.2 Introduction

Evolution has shaped animals to integrate internal (physiological) and external (behavioural, social, ecological) cues to balance their nutrition (Simpson et al. 2015, Chan et al. 2021). Whether diet specialists or generalists, all animals must regulate the intake and balance of nutrients in their diet to maximise growth and fitness (Raubenheimer & Simpson 1997, Raubenheimer et al. 2009). Despite this, optimum nutrition (henceforth referred to as the 'intake target') is not always attainable, and animals must cope with nutrient imbalances the best way possible. The rules that animals follow in order to compromise their nutrient intake remains subject of extensive interest in nutritional sciences research because they shape animal decision-making and long-term fitness (Raubenheimer & Simpson 2020).

Animals can regulate nutrient imbalances physiologically, through post-digestive processes (see e.g., Simpson & Raubenheimer 2012, Cavigliasso et al. 2020, Simpson & Raubenheimer 1999, Behmer et al. 1999) and behaviourally, by choosing the amount and types of foods to eat and hence, controlling the magnitude of nutrient surpluses and deficits (Raubenheimer & Simpson 1997) (see also Behmer 2009, for excellent review on the topic). Both of these processes are expected to follow rules (aka 'rules of compromise') that evolved under natural selection which enable animals to tolerate nutrient imbalances (Raubenheimer & Simpson 1997). These rules of compromise aim to minimise the costs of surpluses and deficits of nutrient intake in imbalanced diets (Raubenheimer & Simpson 1993). Marked differences in the rules of compromise have been described in closely related species with different dietary needs (e.g., Lee et al. 2006, Simpson & Raubenheimer 1999, Behmer 2009) but this is a new field for which large-scale comparative studies remain a fertile field of investigation. A recent framework has enabled us to unravel such rules of compromise: the framework known as the Geometric Framework for nutrition (GF). GF accommodates the complexity of nutrition through a clever experimental design where the additive and interactive effects of nutrients can be investigated simultaneously (Raubenheimer & Simpson 1993, Simpson & Raubenheimer 1993).Two concepts in the GF framework are key: *nutritional rails* and *intake target*. Nutritional rails are diets with fixed ratio of nutrients that animals are fed, and can be balanced or imbalanced. Animals can move along these nutritional rails by modulating the quantity of food they eat, but cannot move across rails because the ratio of nutrients is fixed (hence the name 'rails'), unless a choice experiment is performed (see Fig 7.1a). In standard GF experiments, the number of nutritional rails can vary, but is typically between five and ten (e.g., Simpson & Raubenheimer 1993, Lee et al. 2008, Barragan-Fonseca et al. 2021, Jang & Lee 2018, Lihoreau et al. 2016, Rapkin et al. 2018, Morimoto &

Wigby 2016, Carey et al. 2022, Bunning et al. 2015, Ponton et al. 2015, Polak et al. 2017). These rails are essential to generate the nutrient array, which is the collection of average food intake of animals in each of the nutritional rails (Raubenheimer & Simpson 1993) (Fig 7.1a). The intake target is the balance of nutrients that animals *actively* seek to achieve when allowed to feed freely (Simpson & Raubenheimer 2012, 1993, Raubenheimer & Simpson 1993). The intake target is the closest measure of the optimum nutrient balance of animals in terms of food consumption (although other targets may exist, e.g., for growth) (see Simpson & Raubenheimer 2012, Raubenheimer et al. 2022, for thorough discussion). The rules of compromise can thus be inferred from the way animals feed on the nutritional rails relative to their intake target (i.e. the shape of the nutrient array). This is precisely why nutrient arrays can be used as the fingerprint of the underlying rules of compromise guiding animal feeding (Raubenheimer & Simpson 1997). Rules of compromise are generally assumed to impose constraints on how animals feed when the available diet differs (by a little or a lot) from the optimal diet. A more detailed overview of the GF framework can be found in the literature (see e.g., Simpson & Raubenheimer 1993, 2012, Raubenheimer & Simpson 1993, Raubenheimer et al. 2009, Morimoto & Lihoreau 2020, del Castillo et al. 2016, Rapkin et al. 2018, and others). Importantly, GF is a framework that can be applied across taxa, making GF an attractive framework to reveal general patterns and responses in animal nutrition (Simpson & Raubenheimer 1993, Raubenheimer & Simpson 1993). Because of this, GF has gained popularity in studies of nutritional ecology, particularly when the focus is on nutritional trade-offs and life-history traits (see e.g., Barragan-Fonseca et al. 2019, Pascacio-Villafán et al. 2022, Simpson et al. 2017, Behmer 2009, Lee et al. 2008, Maklakov et al. 2008, Rapkin et al. 2018, Ponton et al. 2015, Polak et al. 2017, and references therein). Although GF experiments are expensive and time-consuming, and broader data sharing remains poor (Morimoto & Lihoreau 2020), GF has enabled unprecedented insights into animal and human nutrition (see e.g., Raubenheimer & Simpson 2020, Simpson & Raubenheimer 2012, 2005, Simpson et al. 2003, Behmer 2009, Solon-Biet et al. 2016, Ng et al. 2018, Bradbury et al. 2014).

An important roadblock for the wider use of GF has been the relative delay in the development of specific analytical frameworks that match the experimental complexity of GF studies. Until recently, studies have relied at least partly on visual interpretations of multidimensional performance landscapes obtained with GF to draw biological conclusions (see e.g., Kutz et al. 2019, Ma et al. 2020, Barragan-Fonseca et al. 2019, 2021) (but see also Pascacio-Villafán et al. 2022), making objective comparisons and comparative studies of nutrition using GF difficult. Recent models have been developed to address this, and this has been a fertile ground for methodological advances (see for instance Rapkin et al. 2018,

Morimoto & Lihoreau 2019, Morimoto et al. 2022, 2023, Pascacio-Villafán et al. 2022, del Castillo et al. 2016, Hosking et al. 2019, Ruohonen et al. 2001). However, the same level of methodological development has not been seen for studies on the rules of compromise, which has lagged behind and remains in need of analytical breakthroughs. Analytical methods are crucial for the advancement of this field because they provide the methodology for accurate and reproducible analyses of the rules of compromise. This can help uncover insights into the eco-evolutionary processes underpinning diet specialisation. For example, diet specialists, but not diet generalists, display a peculiar rule of compromise known as the 'closest distance optimisation' (CDO)(Raubenheimer & Simpson 1993) (see Simpson & Raubenheimer 2012, for a review) (see Fig 7.1a), which was empirically observed in dietary specialist locust and moth species (Lee et al. 2002, 2006, 2003, Simpson & Raubenheimer 1999) (see also Simpson et al. 2002). Furthermore, CDO was observed in the solitary but not the gregarious stage of the swarming locust *Schistocerca gregaria*, suggesting that solitary individuals might have more specialised diets as opposed to the swarming gregarious counterparts (Simpson & Raubenheimer 1999).

There have been few specific studies developing theoretical methods for quantitative analysis of the rules of compromise. For example, Raubenheimer & Simpson (1993) has provided conceptual overview of the nutrient intake arrays that animals display in GF studies, which can be used to infer the rules of compromise. Later, Raubenheimer & Simpson (1997) provided guidelines to study and interpret nutrient arrays and the associated rules of compromise. At the time, the proposed approach relied on Euclidean distances between the amount of food eaten between the imbalanced and optimal diets, which were plotted in 2D spaces to generate what is called 'summary plots' (e.g., Figs 4 and 5 in Raubenheimer & Simpson 1997). However, summary plots estimate the Euclidean distances for each nutrient in the data separately, resulting in a plot with two (or more) curves with different patterns that can be challenging to interpret. In general, the shape of these curves has been interpreted individually and, depending on their linearity and non-linearity, inferences on the rules of animal compromise were derived (Raubenheimer & Simpson 1997). This can be problematic due to some degree of subjectivity. A more complex model was later developed which involved the mapping of the nutrient arrays onto performance landscapes to compare the overall shape of the performance landscape relative to the shape of the nutrient arrays (Simpson et al. 2004). A geometric model was proposed by Cheng et al. (2008) which also relied on Euclidean distances between points to find what was called the 'regulatory scaling factor', which estimated how organisms cope with nutrient surpluses and deficits (Cheng et al. 2008). These models also relied on the Euclidean distances between two average points (e.g., the average intake in imbalanced and the target diets) and

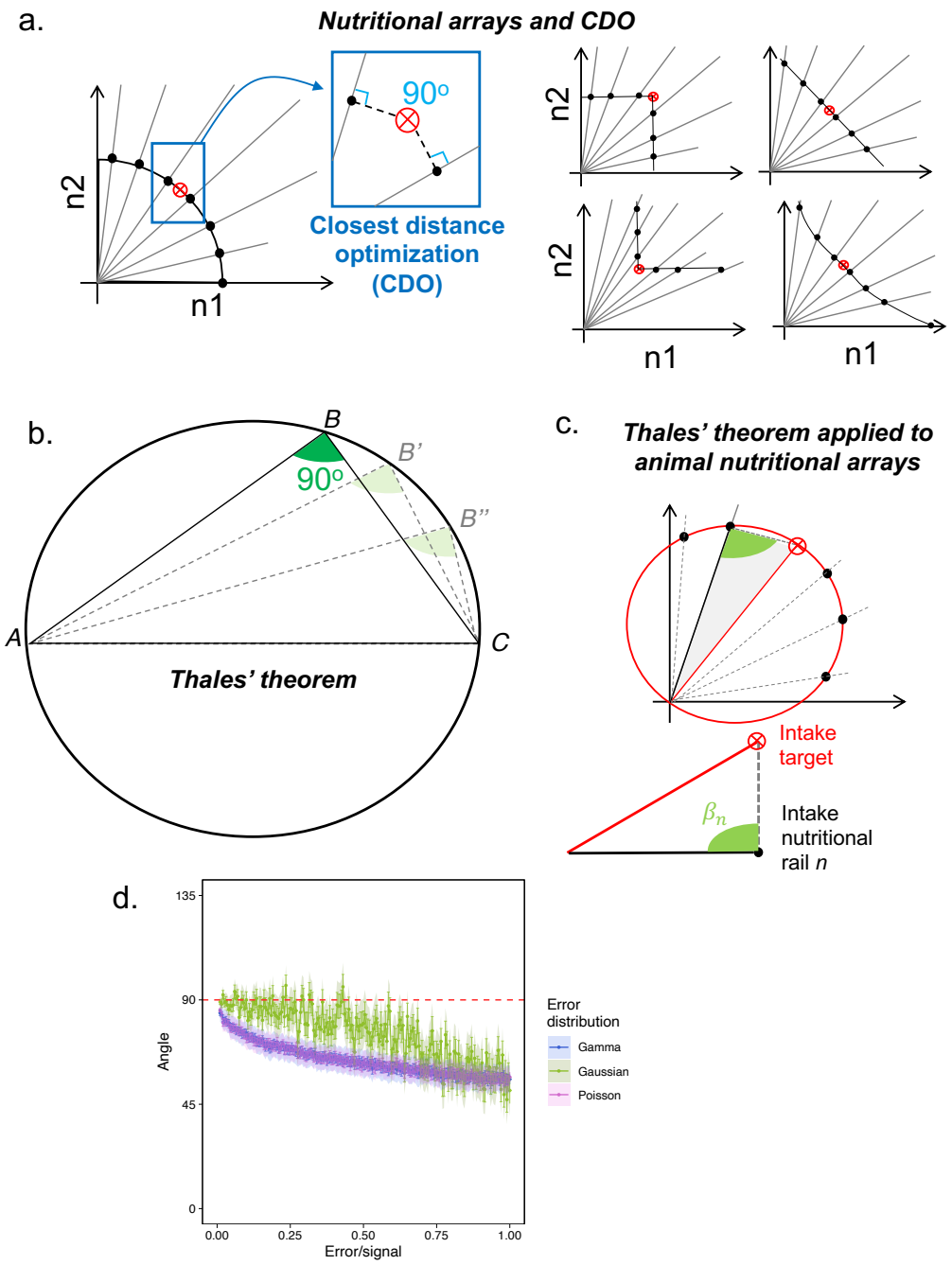


Figure 7.1: **Nutritional arrays, closest distance optimisation (CDO), and the Thales' theorem.** (a) An example of a hypothetical GF nutritional array for the intake of nutrients $n1$ and $n2$. CDO is an array in which the distance between the intake in an imbalanced nutritional rails are minimised relative to the intake target (red crossed point). This implies that the angle between the nutritional rail intake and the intake target is 90° (see zoomed blue box). Nutritional arrays do not have to adopt CDO, and can have a wide range of configurations (left panels) such as a square array, equal distance array, inverted square array, and a concave array (read small panels clockwise). This is reviewed in details in (Raubenheimer & Simpson 1993, Simpson & Raubenheimer 2012). (b) Thales' theorem states that an inscribed triangle will have angle $\beta = 90°$ when the vertices lie on the circumference and the side $\overline{AC}$ is the diameter. Note the angle remains the same as long as these conditions are fulfilled (see faded triangles with vertices β' and β'' (c) One can apply Thales' theorem to investigate whether or not nutritional arrays matches the predicted conditions for CDO, or how much and where the nutritional array deviates from CDO. (d) Normal distribution of errors provides more stability for the estimates of the angle β. X-axis represent the proportion of 'noise' (effect size over error and the y-axis is the estimate of the angle β following the Thales approach. Shaded region represents the 95% confidence intervals of the simulations.

lacked error estimates (see e.g., Simpson & Raubenheimer 1999, Lee et al. 2002, Cheng et al. 2008). Other models have been developed to analyse the trade-off between energy intake and the leverage that each nutrient has in shaping animal nutrient arrays, which have been applied to human nutrition to gain insights into obesity (Hall 2019, Raubenheimer & Simpson 2019). However, these past methods often relied on pairwise distances between average diet intakes in balanced (intake target) and imbalanced diets (nutritional rails). Thus, an integrative method that enables clear visualisation and rapid and intuitive computation of CDO will benefit our interpretation and understanding of animal feeding rules, advancing the field of nutritional ecology.

Here, I propose an analytical method to address this gap. The method builds upon the interpretation of distance summary plots from Raubenheimer & Simpson (1997), Simpson et al. (2004) but enables direct statistical tests of the patterns of nutrient arrays against the predictions from CDO. This is achieved by integrating an ancient mathematical theorem known as the *Thales' theorem* to estimate deviations of the patterns of nutrient arrays from CDO (see Fig 7.1b-c), which I show here by validating the method to a simulation (Fig 7.1d) and applying the method to three landmark datasets of increasing complexity: (1) the data for the nutrient array of a single species, *Drosophila melanogaster*, (2) the data for locusts presented in Simpson & Raubenheimer (1999), and (3) the data for generalists and specialists *Spodoptera* moths from Lee et al. (2002, 2003). I used these datasets of nutrient arrays that emerged from behavioural regulations of food intake (as opposed to post-digestive processes), but the method presented here is also applicable to physiological and molecular data from GF. Overall, the method proposed here advances our ability to make statistical inferences on the rules of compromises within and between species. This opens up new possibilities to study how the rules of compromise evolved across the animal kingdom, and how the behaviour, ecology and physiology of species can influence their ability to cope with nutrient imbalances.

7.3 Results

7.3.1 The method: Thales' theorem and CDO

Thales' theorem states that if an inscribed triangle has points A, B and C on the circumference, where the side $\overline{AC}$ is the diameter of the circumference, then the angle $\angle ABC$ equals to $\frac{\pi}{2}$ (i.e., 90°) (see Fig 7.1b). But why is this theorem useful in the context of the rules of compromise?

A common and informative rule of compromise is the CDO (see 'Introduction'), which states that animals should minimise the distance between the average intake of the imbalanced diet relative to the intake target. This leads to a semi-circle configuration of the nutrient array (Raubenheimer & Simpson 1997) (see Fig 7.1a). The CDO configuration emerges because in a flat plane, such as the Cartesian plane (or more generally, in $\mathbb{R}^2$) the closest distance between two points is a straight line. This means that the closest distance from the intake target and a point in a nutritional rail is a straight line with 90°angle between the rail that crosses the intake target and the imbalanced nutritional rail i, where i is the number of imbalanced nutritional rails used in the study (Raubenheimer & Simpson 1993). Recall that the Thales' theorem states that the angle $\beta = \angle ABC$ equals 90°if and only if the three points of an inscribing triangle lie in the circumference and $\overline{AC}$ is the diameter. Adapting this theorem, we can draw a circumference with diameter equal to the distance between the origin and the intake target. We can then triangulate the origin, the intake target, and the point in the imbalanced nutritional rail such that if the angle β equals to 90°for all nutritional rails, then the nutrient array is that of a CDO rule of compromise (see Fig 7.1c). Moreover, if the nutrient array does not match that of CDO rule of compromise, the angle β can nevertheless provide useful insights to determine the relative importance of each nutrient in determining animals' feeding priorities, such as e.g., which nutrients are more or less tightly prioritised. For example, if the angle β is greater than 90°, then the point in the nutritional rail lies *inside* the inscribing circle, which suggests stronger feeding constrain to avoid surpluses of a nutrient. Conversely, if the angle β is smaller than 90°, the point in the nutritional rail lies *outside* the inscribing circle, suggesting that the surplus of the nutrient is well tolerated. Interestingly, the simulations of error structure underpinning nutrient array data showed that the Thales approach to estimate the angle β is more stable when the error distribution in the nutritional rails is derived from a Gaussian distribution (see Fig 7.1d).

7.3.2 *Drosophila* responds to dietary imbalances with underconsumption of carbohydrate but not of protein

Firstly, I applied the method to gain insights into the feeding behaviour of *Drosophila melanogaster*. When given a choice, *Drosophila* regulates the intake of both protein and carbohydrate to reach a P:C ratio of 1:4, which is the ratio that maximises lifetime egg production (fitness) (Lee et al. 2008). Moreover, when given a choice between two complementary diets of varying concentrations, flies also choose to overeat

protein when the concentration of carbohydrate is low in the counterpart diet (Lee et al. 2008). Using the method proposed here, I confirmed that flies are able to regulate both protein and carbohydrate intakes ($F_{7,966}$: 15.585, p < 0.001, Table 7.1) , as the shape of the nutrient array for diets with P:C ratio 1:4, 1:2, 1:1 and 2:1 were according to the predictions of the CDO (see Fig 7.2a-b). However, as the nutrient imbalances in the diet increased towards high-carbohydrate contents (i.e., P:C 1:8, 1:16, 0:1), dietary intake of the nutritional rails progressively decreased, becoming statistically significantly different than the predictions from CDO for diets with P:C 1:16 and 0:1 (see Fig 7.2a-b). These results suggest that flies display remarkable underconsumption of carbohydrate-biased, but not protein-biased diets when facing strong nutrient imbalances (Table 7.1).

Table 7.1: Estimates of angle β in the nutritional array in *D. melanogaster* relative to CDO. Note that ratios in which the angle β overlaps zero implies no differences from CDO.

	D. melanogaster		
P:C ratio	Mean β	lwr 95% CI	upr 95% CI
0:1	43.526	34.643	52.410
1:16	14.596	5.676	23.508
1:8	8.587	-0.460	17.634
1:4	-6.503	-15.355	2.348
1:2	1.849	-6.908	10.607
1:1	-0.093	-9.041	8.854
2:1	-3.469	-12.258	5.320

Diet specialisation leads to different nutrient arrays in response to dietary imbalances in two locust species

Next, I applied the method to the dataset of *Locusta* and *Schistocerca* locusts species first presented in (Simpson & Raubenheimer 1999). The original study shows the difference in the nutrient array, whereby *Locusta*, a diet specialist, displayed nutrient array as predicted by CDO whereas *Schistocerca*, a diet generalist, displayed a more linear nutrient array which was more tolerant of overconsumption of both protein and carbohydrate (see Fig 4 in Simpson & Raubenheimer 1999). The method presented here provides a clear framework to distinguish between the two responses (see Fig 7.2c-d). The method corroborates the findings presented in (Simpson & Raubenheimer 1999) by showing that the nutrient array for *Locusta* fits well to the predictions for CDO, while the response for *Schistocerca* diverged substantially. This translated into different patterns in the plots of the angle β for both species. For instance, the angle β for *Locusta* fluctuated closely to 90° as expected from the Thales' theorem predictions. Meanwhile, the angle β for *Schistocerca* was often smaller than 90° in nutritional rails with more extreme nutrient

imbalances, and progressively converged to 90° as the P:C ratio of the nutritional rail approximated the optimum P:C ratio of the intake target. This led to the plot of the angle β to resemble a parabola (see Fig 7.2d). Thus, the method proposed here provides an analytical framework to clearly differentiate differences in nutrient arrays and deviations from CDO.

7.3.3 Nutrient-specific effects of diet specialisation levels on the responses to nutrient imbalances in *Spodoptera* species

Next, I used the method to gain insights into the patterns of the nutrient arrays of two *Spodoptera* species with different diet specialisation levels. There was substantial difference in the overall consumption of diets (irrespective of their P:C ratios) between the two species, with *S. littoralis* consuming greater amounts of all diets (see Fig 7.2e-f). More importantly, the method revealed interesting differences in the responses to nutrient imbalances between the two species (see Fig 7.2e). There was a statistically significant interaction between species and the nutritional rail on the estimates of the angle β (*Ratio * Species*: $F_{4,626}$: 13.668, $p < 0.001$), suggesting that the shape of nutrient arrays differ between species.

Spodoptera littoralis, a diet generalist species, displayed a clear pattern of overconsumption of the most abundant nutrient in order to minimise the underconsumption of the least abundance nutrient. This was true for all four days of feeding data collection and resembled the responses of *Schistocerca* (see above). This means that, if we were to connect the average intakes across all nutritional rails to form the nutrient array of *S. littoralis* (or *Schistocerca* previously), forming a parabola on the plot of the angle β (see Fig 7.2e, Table 7.2).

Spodoptera exempta, a diet specialist, displayed a somewhat similar nutrient array as *S. littoralis* for nutritional rails that were carbohydrate-biased relative to the intake target P:C ratio. However, for diets that were similar or protein-biased compared to the P:C ratio of the intake target, *S. exempta* displayed an imperfect resemblance to CDO. This mixture of responses showed in the plot of the angle β, in which values of β progressively increased towards 90° as the P:C ratio increased (i.e., more protein) up until the nutritional rail with similar P:C ratio to the intake target, where the intake in the nutritional rail decreased (and hence, the angle β was $\geq$ 90° (see Fig 7.2f). Furthermore, contrary to the response observed in *S. littoralis*, the angle β did not decrease as the nutritional rails became protein-biased, suggesting that *S. exempta* held the intake of imbalanced diets closer to the expected intake for CDO (see Fig 7.2f, Table 7.2). Together, the results from the method suggest that *S. littoralis* can cope with

Table 7.2: Estimates of angle β in the nutritional array in locusts and moths, relative to CDO. Note that ratios in which the angle β overlaps zero implies no differences from CDO.

Species	P:C ratio	Mean β	lwr 95% CI	upr 95% CI
L. migratoria	1:5	2.640	-0.550	5.828
	1:2	5.535	2.345	8.724
	1:1	7.702	4.512	10.891
	2:1	0.843	-2.346	4.032
	5:1	-9.628	-12.817	-6.438
S. gregaria	1:5	-23.895	-27.294	-20.494
	1:2	-17.878	-21.278	-14.478
	1:1	7.401	4.001	10.800
	2:1	-22.993	-26.392	-19.592
	5:1	-30.953	-34.353	-27.553
S. exempta	1:5	-41.504	-53.956	-29.052
	1:2	-35.560	-53.53	-17.585
	1:1	0.475	-11.977	12.927
	2:1	-6.064	-24.038	11.908
	5:1	-8.017	-20.470	4.434
S. littoralis	1:5	-23.037	-43.644	-2.430
	1:2	-9.955	-30.562	10.651
	1:1	-1.565	-22.171	19.042
	2:1	-13.086	-33.693	7.520
	5:1	-29.878	-50.484	-9.270

surpluses of both carbohydrates and proteins equally well, whereas *S. exempta* can cope well with surplus of carbohydrate but tightly regulate the intake of nutrients in protein-biased diets.

7.4 Discussion

Animals often need to regulate the intake of nutrients, a challenging task when animals feed on imbalanced diets. Rules evolved which enable animals to balance the costs and benefits of under- and over-consumption of nutrients in these situations ('rules of compromise') imposing important constrains on how animals eat (Raubenheimer & Simpson 2020, Simpson & Raubenheimer 2012). Using the Geometric Framework for nutrition, studies have generated a rich collection of datasets that allow for these rules of compromise to be studied in details. However, the development of analytical methods for statistical inferences lagged behind (Simpson & Raubenheimer 1999, Lee et al. 2002, Cheng et al. 2008). In this study, I proposed an analytical method to study rules of compromise, which was validated using three landmark datasets of increasing complexity. This method provides two main contributions to the field, namely, (1) an intuitive framework for data visualisation of animal feeding and (2) a simple method to describe and test the rules of compromise in animal nutrition. This will help advance our understanding of how animals compromise the intake of nutrients when feeding in imbalanced diets.

Diet specialists are seemingly constrained in their ability to cope with overconsumption of nutrients

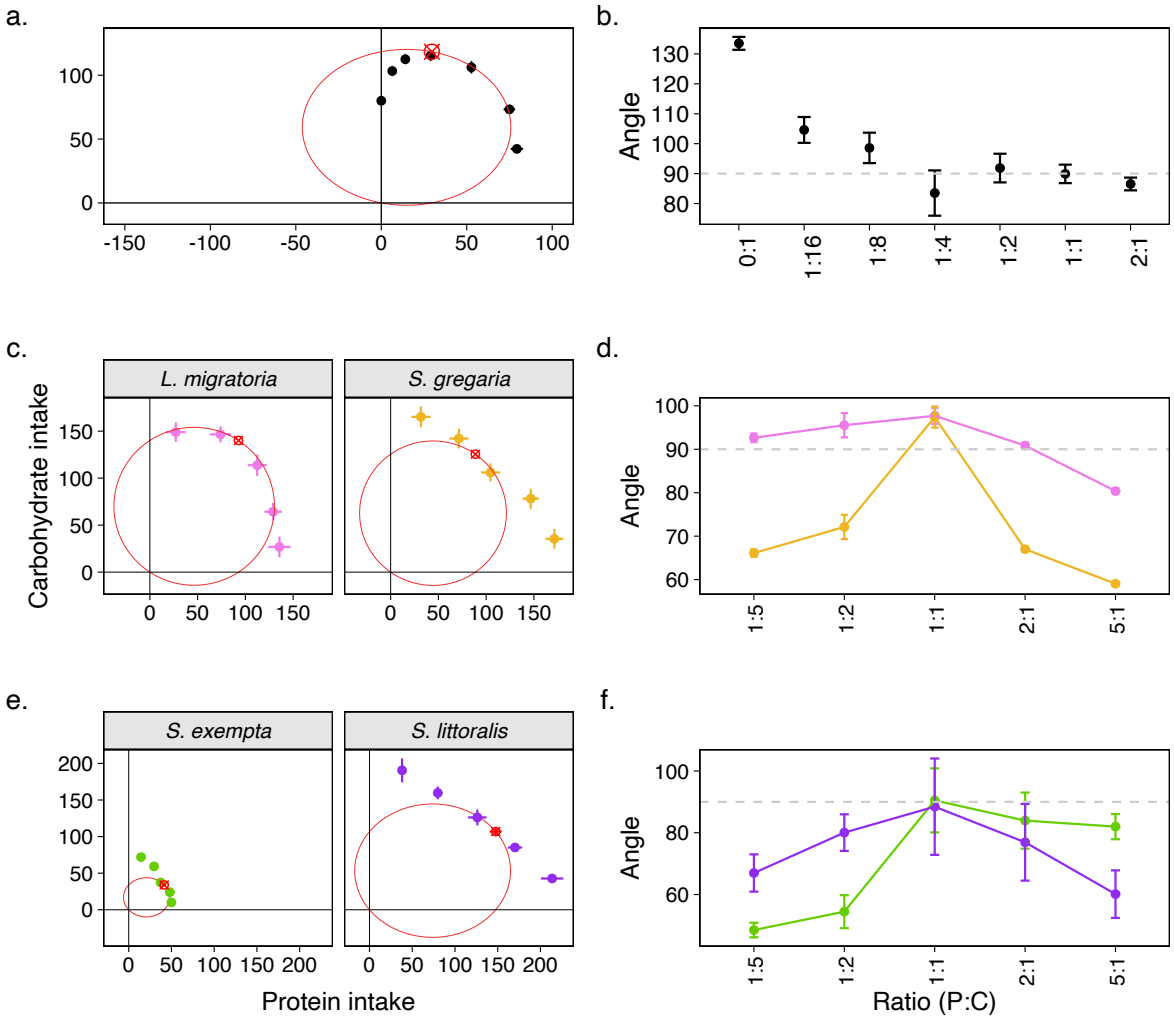


Figure 7.2: **Thales theorem applied to empirical nutritional arrays** (a) Nutritional array in *Drosophila melanogaster*, with mean diet intake for each imbalanced diets (with varying P:C ratios) from (Lee et al. 2008). (b) Summary plot of the angle β of the nutritional rails relative to the intake target. A 90° angle suggests that the nutritional array matches the prediction of CDO for a given rail. (c) Nutritional array in *L. migratoria* and *S. gregaria*, with mean diet intake for each imbalanced diets (with varying P:C ratios) extracted from (Simpson & Raubenheimer 1999). (d) Summary plot of the angle β of the nutritional rails relative to the intake target for the two species. (e) Nutritional array in *S. exempta* and *S. littoralis*, with mean diet intake for each imbalanced diets (with varying P:C ratios) extracted from (Lee et al. 2002, 2003). (b) Summary plot of the angle β of the nutritional rails relative to the intake target for the two species. Red circle: Thales' circle with diameter equals to the intake target.

and compromise on the intake of both nutrients tested (in this case, protein and carbohydrate), in accordance with the strategy of closest distance optimisation (CDO). This was evident in the array of *L. migratoria* and partly evident in the nutritional array of *S. exempta* (see Fig 7.2c-f). CDO has been hypothesised as a general pattern in diet specialists, where the consumption of multiple nutrients are tightly regulated to ensure fitness (Raubenheimer & Simpson 1999, Simpson & Raubenheimer 2001). Interestingly, in the cockroach *Blattella germanica*, in which some populations have evolved dietary specialisation to avoid glucose, rules of compromise did not comply with CDO (see Shik et al. 2014). Glucose-avoidance specialisation appears to be hardwired and individuals may be unable to compensate

for underconsumption of nutrients via digestive processes (Shik et al. 2014). One caveat is that Shik et al. (2014) only used three nutritional rails to construct the nutritional arrays and thus, the experimental design is not broad enough to allow for a proper test for CDO. Nonetheless, the model proposed here can be used in future studies to verify whether or not the nutrient arrays of diet specialists and generalists adhere to CDO rule of compromise.

The nutrient array of *D. melanogaster* was also constrained for overconsumption of imabalanced diets, particularly those with high carbohydrate, and resembled in many ways the array of a diet specialist even though *D. melanogaster* is widely defined as being diet generalist. Could the nutrient array be revealing that *D. melanogaster* is in truth a specialist species? This is unlikely, although not completely implausible scenario. The laboratory population used in the study by Lee et al. (2008) was inbred and could have behaved as a diet specialist. Moreover, a recent study using the Drosophila Genetic Reference Panel (DGRP) lines has shown considerable variability in *D. melanogaster* survival across different diets, in particular diets with high carbohydrate levels (Havula et al. 2022). This aligns with the insights gained here using the Thales' method: *Drosophila* nutrient array diverges more strongly from CDO as the concentration of carbohydrate (but not of protein) increases. Contrary to this, however, previous studies have shown that in high-sugar diets, flies have reduced responses to sweet taste which leads to overconsumption of the diet, a response that is mediated by the release of dopamine and the expression of the enzyme O-linked N-Acetylglucosamine transferase (OGT) in sweet-sensing neurons (Inagaki et al. 2012, May et al. 2019). Moreover, sugar consumption is directly linked to female fecundity (Carvalho-Santos et al. 2020) and male fertility is maximised at a relatively higher proportion of sugar consumption (compared with females) (Morimoto & Wigby 2016, Lee et al. 2008). This highlights the importance of sugar-appetite to overall fitness. Thus, the seemingly divergent findings of the nutritional array studies and studies on the consumption of sugar in *Drosophila* literature remains subject of further molecular and physiological studies.

I have shown that the method proposed here can help the interpretation of more complex nutritional arrays which display mixed responses to nutrient imbalances. To be applied in empirical datasets, the method proposed here requires a GF choice experiment to determine the coordinates of the intake target, from which an accurate Thales' circle can be drawn to analyse the nutritional array. It is also worth mentioning that the method presented here has some limitations. Firstly, I calculated the angle β for individual datapoints in each nutritional rail relative to the average intake target. As a result, the estimates might suffer from uncertainty propagation (i.e. when the uncertainty in random variable are

propagated when variables are combined). This is particularly important when variables are correlated, as failing to account for propagation of uncertainty can lead to underestimation of combined error. In the datasets used to validate the approach presented here, the errors associated with nutritional rails and intake targets were collected independently and are assumed to be uncorrelated, as most of the studies conducted in the field (e.g., Raubenheimer & Simpson 1993, Simpson et al. 2004). In the datasets analysed here, the combined error is smaller than the individual errors in nutrient intake and thus, the Thales' approach presented here shows a conservative estimate of statistical significance (Table 7.3). Recent studies have started to consider uncertainty propagation when measuring GF data within the context of the protein leverage hypothesis (Senior 2023) but this investigation lies beyond the scope of this study. Secondly, the Thales' theorem applies in two, but not higher dimensions. This 2D-nutritional-arrays approach have been proposed as the best way to analyse rules of compromise (Simpson & Raubenheimer 2001) and the method proposed in this study complies with this recommendation. Should the number of dimensions of the nutritional array increase, however, a new method has to be devised *or* the data has to be 'sliced' into lower dimension subsections (either by pairwise comparisons or through dimensionality reduction such as e.g., Principal Component Analysis), although such approach can transform the data in ways that might complicate analysis of the rules of compromise. High-dimensional data are considerably more difficult to interpret, and whether high-dimensional nutritional arrays are in themselves informative remains to be studied. So far, studies that investigated the rules of compromise have been of 2D, for which the method proposed here is perfectly suitable (e.g., Lee et al. 2002, 2003, 2006, 2008, Pascacio-Villafán et al. 2022, Barragan-Fonseca et al. 2021, Bunning et al. 2015, Morimoto & Wigby 2016, Polak et al. 2017, Ponton et al. 2015, Hosking et al. 2019, and others).

Using an ancient theorem known as the Thales' theorem, I have developed an intuitive and reproducible analytical method to study the feeding patterns of animal in response to nutrient imbalances. This method advances our previous approaches and enables statistical analysis and interpretation of complex patterns in nutrient arrays. This opens up routes for the application of this method to the broader field of nutritional ecology, including recent translational studies using GF to investigate nutrition in health and (metabolic) diseases (e.g. Solon-Biet et al. 2014, 2016, Simpson et al. 2017).

7.5 Methods

7.5.1 Datasets

We validated the method using three landmark datasets in the field of nutritional ecology:

1. The first was the data from Lee et al. (2008) on the nutritional responses in *Drosophila melanogaster*. This is a landmark paper because it was the first to demonstrate the nutritional trade-offs between lifespan, reproductive rate, and lifetime egg production (i.e., fitness), and that when given a choice, individuals feed on diets with nutrient ratios that maximise lifetime egg production. The experiments focused on the manipulation of the ratio of protein and carbohydrate (P:C ratio) of the diets. Flies were given seven P:C ratios (i.e., nutritional rails), namely, 0:1, 1:16, 1:8, 1:4, 1:2, 1:1, and 1.9 (Lee et al. 2008). This data has been extensively used for GF method development and thus, has gained a important status as a ground-truth in the field (Morimoto & Lihoreau 2019, Morimoto et al. 2022, 2023).

2. The second dataset was for two locust species originally presented in Figure 4 of Simpson & Raubenheimer (1999). The two species for which the nutrient array were extracted were *Locusta migratoria* (specialist, gregarious) and *Schistocerca gregaria* (generalist, solitary or gregarious). For the purpose of this paper, where we used the data for validation, I compared the nutrient arrays of both species but opted to omit the comparison for the solitary vs gregarious stage of *S. gregaria* presented in Simpson & Raubenheimer (1999). This is because my aim was not to replicate the original study, but to demonstrate the power of the method proposed here in identifying different shapes of nutrient arrays. This data also contained five nutritional rails with P:C ratios 7:35 (1:5), 14:28 (1:2), 21:21 (1:1), 28:14 (2:1) or 35:7(5:1).

3. The third dataset was for two Lepidopteran species of the *Spodoptera* genus: *S. littoralis* and *S. exempta*, the former a diet generalist and the latter, a diet specialist. Moths were given five P:C ratios (nutritional rails), namely, 35:7 (5:1), 28:14 (2:1), 21:21 (1:1), 14:28 (1:2) and 7:35 (1:5) (Lee et al. 2002, 2003).

More details of the findings of the above studies can be found in the original publications.

7.5.2 Statistical analyses

All analyses were conducted in R version 4.1.3 (Team 2019). Data handling was conducted using the tidyverse packages 'dplyr 1.0-10' and 'tidyr 1.2.0' (Wickham et al. 2019). Data visualisation plots were done using the 'ggplot2 3.4.0' package (Wickham et al. 2016). I studied the stability of the probability distribution of errors on the estimates of the angle β using a nutrient array with known angle (i.e., 90 °) and added error using the 'rnorm', 'rpois' and 'rgamma' functions in R, with increasing values of the the parameters related to the standard deviation of the distributions (i.e., 'sd', 'lambda', and 'shape' parameters, respectively). Simulation sample size was equal to $n = 100$. I ran 100 simulations, each with standard deviation of the data for each distribution increasing from 0.01 (virtually no error) to 100 (error equals the sample size) in steps of 0.5, totalling 59700 simulated observations of the angle estimates. I estimate the extent of the effects of increasing errors given the proportion of effect size over the error, represented by the proportion of error relative to the data (Fig 7.1d). Next, I extracted average intake for each nutritional rail from (Simpson & Raubenheimer 1999) manually using WebPlotDigitizer 4.2 (Rohatgi 2019). Errors in the intake of carbohydrate and protein for the dataset from Simpson & Raubenheimer (1999) were simulated from a normal distribution using the 'rnorm' function in R with the parameter 'mean' equal to the mean protein or carbohydrate observed in the data and standard deviation equals to 10. Because the errors were simulated, I did conduct any statistical inference on this dataset. The simulation of errors was necessary because the raw estimate of errors were not available in the original dataset. In all datasets, the angle β was estimated for each individual data point in the nutritional rails against the average coordinates of the intake target. This allowed me to estimate the 95% confidence intervals of the angle β, which were calculated using the 'confint' in-built function.

7.6 Acknowledgements

I am grateful to Prof Kwang Lee for providing the raw data used for the empirical demonstration of the method in this manuscript, and to three anonymous reviewers for comments that helped improve the manuscript.

7.7 Supplementary information

7.7.1 Supplementary tables

Table 7.3: Individual versus combined errors in nutrient intake across diets.

Species	Experiment	Nutrient	SD	SD (Combined)
Heliothis subflexa	Target	Carbohydrate	35.2	
	Rails		56.4	66.48
Heliothis virescens	Target		28.5	
	Rails		30.1	41.45
Heliothis subflexa	Target	Protein	33.6	
	Rails		41.0	53.00
Heliothis virescens	Target		38.5	
	Rails		43.8	58.31
Spodoptera exempta	Target	Carbohydrate	11.5	
	Rails		23.7	26.34
Spodoptera littoralis	Target		41.6	
	Rails		66.0	78.01
Spodoptera exempta	Target	Protein	13.7	
	Rails		16.4	21.36
Spodoptera littoralis	Target		58.8	
	Rails		68.3	90.12

7.8 Supplementary text: Matters Arising

After the publication of the manuscript presented in Chapter 3, Senior et al. (2025) wrote a commentary on the approach I used to make statistical inferences using the angle β. In particular, they argue for more theoretical work, and warned against using trigonometric functions to make statistical inferences in nutritional geometry data. My response to their critique is reproduced in full below and published in Morimoto (2025*b*).

7.8.1 Reply to: A caveat about the use of trigonometric functions in statistical tests of Nutritional Geometry models

Over the past decade, analytical models leveraging the experimental power of the Geometric Framework for nutrition (GF) have advanced significantly in extracting insights from multidimensional data (Morimoto & Lihoreau 2019, Rapkin et al. 2018, del Castillo et al. 2022, Morimoto et al. 2023, 2022, Pascacio-Villafán et al. 2022). In a recent study, I introduced the use of Thales' theorem to analyse experimental data on appetite and nutrient intake (Morimoto 2023). While supportive of the concept, Senior et al. (2025) critiqued my model, highlighting potential misestimations of the standard error and

confidence intervals for the angle β. This angle was central to statistically testing whether nutrient arrays align with the closest distance optimization (CDO) rule of dietary compromise (see Morimoto 2023, Simpson & Raubenheimer 1993, Raubenheimer & Simpson 1993, for details).

Using simulations and a reanalysis of the *Drosophila melanogaster* data from my study, Senior et al. (2025) demonstrated two points: (1) heteroscedastic models are required instead of the homoscedastic models I used and (2) estimates of standard error and confidence intervals depend on the proximity to the intake target (IT), that is, the nutrient balance animals select when given a choice.

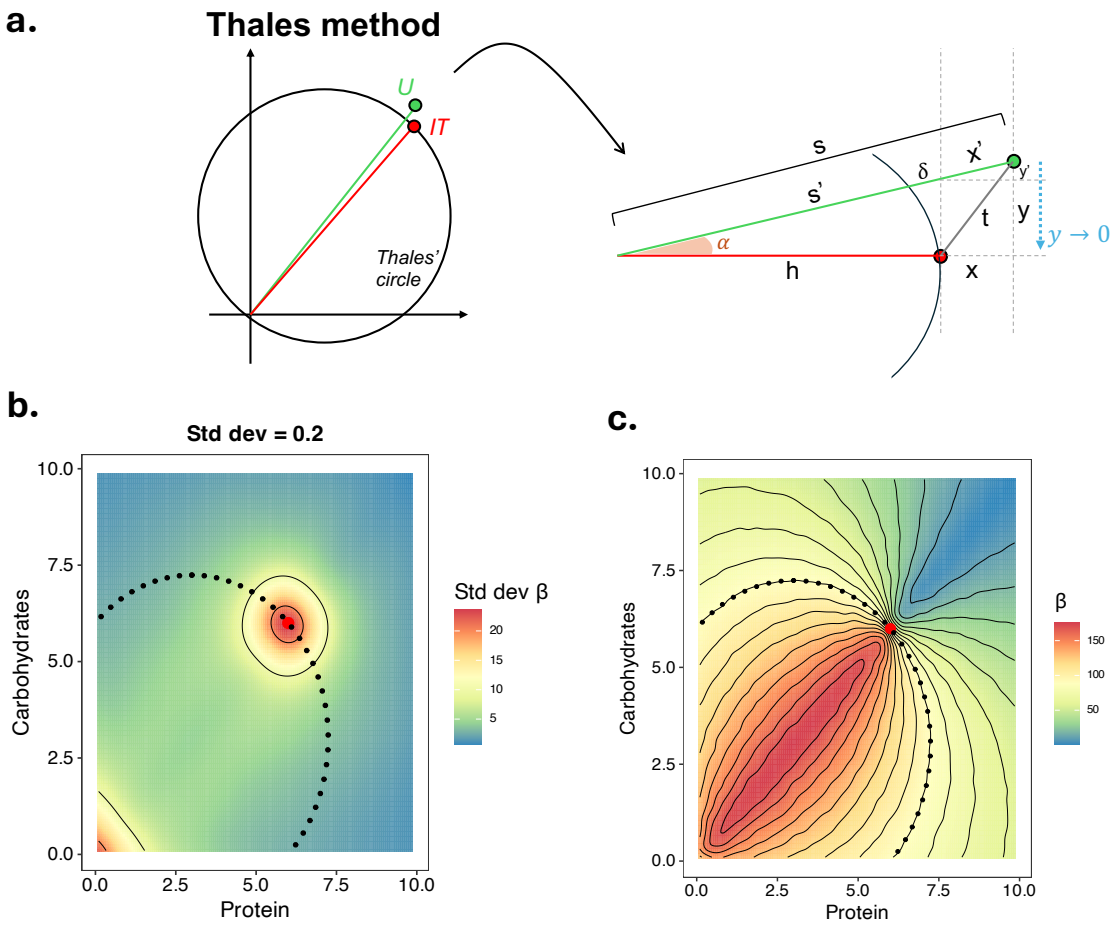


Figure 7.3: **Thales' theorem applied to nutrition.** (a) Geometric construction to show that the estimates of the angle β become unstable as U approaches the IT. (b) Landscape of the estimated standard deviation of a simulated mesh of nutrient intakes with known normally distributed noise with mean = 0 and standard deviation = 0.2. This illustrates that the standard deviation becomes unstable in a symmetric way as U approaches the IT. (c) The estimates of the angle β in the same noisy landscape as in panel (b) to highlight that, despite instabilities in the estimates of the standard deviation, error, and confidence intervals, the angle β is nevertheless a useful proxy for the CDO rule of dietary compromise.

Senior et al. (2025) offered heuristic explanations for these findings and suggested that the delta method could improve accuracy when using trigonometric functions. They also emphasised the need for further theoretical development to address these issues comprehensively.

I welcome the engagement of Senior et al. (2025) and endorse their call for more theoretical work in

nutritional ecology. I was delighted by the delta method's ability to accurately estimate standard errors and confidence intervals, enabling correct statistical inferences with my proposed model. This is because Senior et al. (2025) simulations showed that these estimates depend on the distance to the IT. However, an important question remains: why are the angle β estimates influenced by the distance to the IT? While Senior et al. (2025) heuristic argument offered valuable insights, I believe further formalisation is needed, and I aim to address that here.

Recall from Morimoto (2023) and the law of cosines that the angle β is calculated as:

$$\beta = \cos^{-1}\left(\frac{s^2 + t^2 - h^2}{2st}\right) \tag{7.1}$$

where s is the length between the origin and the average intake of an unbalanced diet U, h is the length between the origin and the IT (also the diameter of Thales' circle), and t is the Euclidean distance between the IT and the intake in the unbalanced diet. Senior et al. (2025) claimed that standard errors are wider as the intake of an unbalanced diet U approaches the IT (Figure Figure 7.3a). In the GF framework, this happens if diet U was designed (purposefully or serendipitously) and used in no-choice experiments with PC ratios near that of the IT (Raubenheimer & Simpson 1993, Simpson & Raubenheimer 1993).

The inverse cosine function $\cos^{-1}(x)$ used above has domain $[-1, 1]$. Notice that this function is non-linear in its domain, and the absolute magnitude of its gradient increases as the input value x approaches the domain boundaries:

$$\frac{d}{dx}\cos^{-1}(x) = -\frac{1}{\sqrt{1 - x^2}} \tag{7.2}$$

This means that estimates of the angle β will vary proportionally more as x approaches the boundaries, i.e., as $x \to \pm 1$. For the sake of argument, let's focus on the case where $x \to 1$, although the logic is the same when $x \to -1$. As the domain interval is closed, let $x = 1$, which means that in Equation 1, $\cos^{-1}(1)$ when $s^2 + t^2 - h^2 = 2st$. Simplifying this, we find that $\cos^{-1}(1)$ when $h = s - t$ (see Appendix I).

Let's consider a diet U that is some distance y to the IT as shown in Figure 7.3a. With some geometric constructions and projections, we have the following relationships:

$$t^2 = x^2 + y^2 \quad \text{using Pythagoras theorem} \tag{7.3}$$

$$x'^2 = x^2 + y'^2 \quad \text{using Pythagoras theorem} \tag{7.4}$$

$$\sin(\alpha) = \frac{y'}{x} \quad \text{using similarity of triangles} \tag{7.5}$$

$$s = s' + \delta + x' \quad \text{by construction} \tag{7.6}$$

$$(s' + \delta)^2 = h^2 + y^2 \quad \text{using Pythagoras theorem} \tag{7.7}$$

What happens to these equations when $U \rightarrow$ IT? Another way of stating this is: 'what happens when $y \rightarrow 0$?' (Figure 7.3a). This means that:

1. $t^2 \rightarrow x^2$, which also implies that $t \rightarrow x$ (using 7.3 and 7.4)
2. $\alpha \rightarrow 0$ and $\sin(\alpha) \rightarrow 0$, such that $x' \rightarrow \sqrt{0 + x^2}$ and $x' \rightarrow x$ (7.4 and 7.5)
3. $\delta \rightarrow 0$, as the distance between the horizontal line y becomes tangent to the Thales' circle at the IT. This also means that $s \rightarrow s' + x'$ in 7.6.
4. Note that $x' \rightarrow x$ (from step 2) and $t \rightarrow x$ (from step 1), then $s \rightarrow s' + x'$ is equivalent to $s' \rightarrow s + t$.
5. As both δ (from step above) and y (by definition) $\rightarrow 0$, 7.7 then leads us to $h \rightarrow s'$.
6. Substituting this into step 4, we have that as $y \rightarrow 0$, then $h \rightarrow s - t$, which is what we needed to show: that estimates of the angle β become more variable as the intake of diet $U \rightarrow$ IT.

To show how this looks within the GF context, I used the software R version 4.3.2 (Team 2019) to run simulations showing how the standard deviation (which is related to the standard error) of the angle β changes across a GF landscape. For the simulation, I created a mesh that simulated nutrient intakes across a range of 0 to 10 for both protein (x-axis) and carbohydrates (y-axis). For each point, I added random noise from a normal distribution using the `rnorm` function parameters of mean = 0 and standard deviation = 0.2. The IT was set at the coordinates (Protein = 6, Carbohydrate = 6) and, as in Senior et al. (2025), was assumed to be known with perfect knowledge. Figure 7.3b shows that the standard deviation of the angle β indeed increases as the estimates approach the intake target. It also increases next to the origin for the same reasons mentioned above.

Despite these criticisms, the Thales' method remains valid, even if as a qualitative tool for understanding dietary compromise rules. This is because animals following the CDO rule will have average

angle β estimates close to 90°. Figure 7.3c illustrates angle β estimates within the same noisy landscape, demonstrating the method's utility. As expected, average angle β values increase when the intake of an unbalanced diet (U) falls within the Thales' circle, while estimates near the circle approach 90°. Thus, the Thales' method continues to be a valuable approach for evaluating nutrient intake patterns.

An encouraging finding from Senior et al. (2025) is their reanalysis of the Drosophila dataset from Lee et al. (2008). Using Welch's heteroscedastic models, they concluded that neither equal distance nor CDO fully explains macronutrient regulation in females. This aligns with my original conclusions using the Thales' theorem with a homoscedastic model, where I found that flies follow the CDO rule when diets are protein-rich but deviate from it as diets become carbohydrate-rich (see Figure 7.4a-b in the original paper)6. While this does not suggest fitting an incorrect model, it underscores the overarching value of the Thales' method.

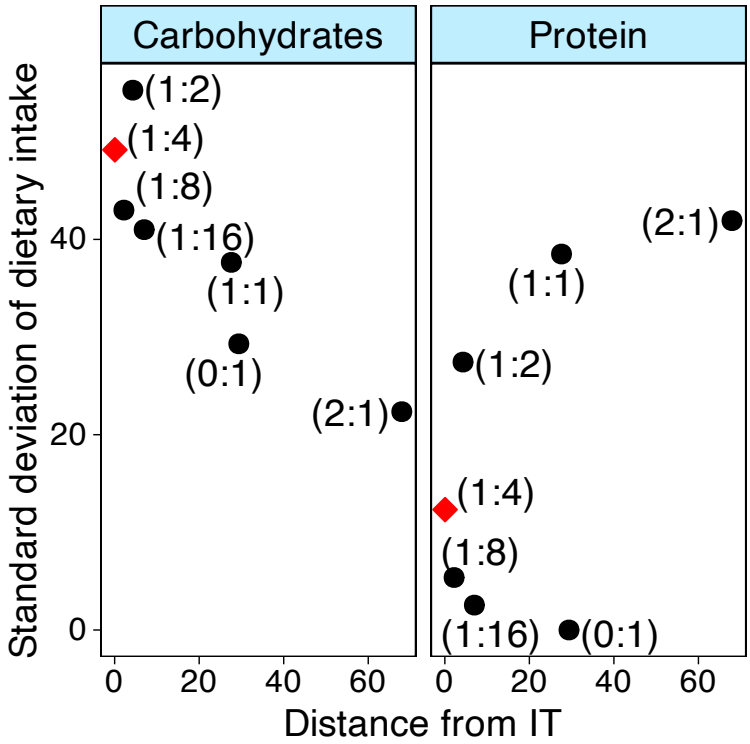


Figure 7.4: **Standard deviation of dietary intake in nutritional rails of varying distances from the IT in Drosophila.** Red diamond highlights the standard deviation in the intake of the rail associated with the IT. For carbohydrates, there is almost a linear decline in standard deviation with the distance of the nutritional rail from the IT. For proteins, a similar pattern is found, although standard deviation was higher for diets with increasing protein content, which highlights interesting physiological differences between the regulation of protein and carbohydrates in *Drosophila*. Distance from IT was estimated as the Euclidean distance from the average dietary intake for the nutritional rail and the IT, divided by the total intake in the nutritional rail. Dividing by the total intake in the nutritional rail did not change the patterns shown. Standard deviation was estimated using the 'sd()' function in R from the raw data. Data form Lee et al. (2008).

Increased intake variance near the IT may not merely result from statistical artifacts but could reflect

underlying biological mechanisms. GF studies suggest that performance landscapes for traits across diets often form 'plateaus' of maximum expression (see e.g., Lee et al. 2008, Rapkin et al. 2018, Pascacio-Villafán et al. 2022, Rodrigues et al. 2015, Morimoto 2024, Ng et al. 2018). This implies that individuals might tolerate deviations from their IT without significant fitness costs as long as these deviations remain within the plateau (Figure 7.3a). In contrast, larger deviations from IT, taking individuals off the plateau, could act as filters that reduce variance in nutrient intake. Animals feeding closer to the IT might exhibit greater variability in nutrient intake without incurring fitness costs.

This idea aligns with Simpson et al. (2004) assumptions that fitness costs increase "continuously and smoothly" with distance from IT. Under these assumptions, animals would optimize their intake to minimize this distance, shaping their nutrient array and performance landscape geometry. Deviations from IT on costly diets (i.e., off the IT) would be penalised, possibly non-linearly (Simpson et al. 2004), resulting in narrower feeding ranges for animals further from IT. Simpson et al. (2004) Figure 5a visually supports this, showing increased errors when dietary intakes neared the plateau. I reanalysed data from Lee et al. (2008) which also supports this pattern, particularly for carbohydrates and, to a lesser extent, proteins (Figure 7.4). This raises intriguing questions for future research and will certainly uncover new biological patterns, provided the field adopts stronger open-data policies to enhance experimental power and collaboration (Morimoto & Lihoreau 2020, Morimoto 2024).

In sum, I agree with Senior et al. (2025) that further theoretical work is needed to formalise the sampling distributions and statistical inferences using angles and confidence intervals. Advancing our understanding of the hidden patterns in animal nutritional ecology will require a concerted effort combining new theoretical frameworks, analytical models, and empirical data.

Acknowledgements

I would like to thank Dr Irakli Patchkoria for comments that help clarify the arguments presented here.

Appendix I

$$s^2 + t^2 - h^2 = 2st \tag{7.8}$$

$$s^2 - 2st + t^2 = h^2 \tag{7.9}$$

Recognising the left-hand side as a perfect square:

$$(s-t)^2 = h^2 \tag{7.10}$$

Bibliography

Adler, M. I., Cassidy, E. J., Fricke, C. & Bonduriansky, R. (2013), 'The lifespan-reproduction trade-off under dietary restriction is sex-specific and context-dependent', *Experimental gerontology* **48**(6), 539–548.

Alton, L. A., Kutz, T. C., Bywater, C. L., Beaman, J. E., Arnold, P. A., Mirth, C. K., Sgrò, C. M. & White, C. R. (2020), 'Developmental nutrition modulates metabolic responses to projected climate change', *Functional Ecology* **34**(12), 2488–2502.

Anderson, T. R., Boersma, M. & Raubenheimer, D. (2004), 'Stoichiometry: linking elements to biochemicals', *Ecology* **85**(5), 1193–1202.

Archer, C. R., Hempenstall, S., Royle, N. J., Selman, C., Willis, S., Rapkin, J., Blount, J. D. & Hunt, J. (2015), 'Testing the effects of dl-alpha-tocopherol supplementation on oxidative damage, total antioxidant protection and the sex-specific responses of reproductive effort and lifespan to dietary manipulation in australian field crickets (teleogryllus commodus)', *Antioxidants* **4**(4), 768–792.

Arganda, S., Bouchebti, S., Bazazi, S., Le Hesran, S., Puga, C., Latil, G., Simpson, S. J. & Dussutour, A. (2017), 'Parsing the life-shortening effects of dietary protein: effects of individual amino acids', *Proceedings of the Royal Society B: Biological Sciences* **284**(1846), 20162052.

Arien, Y., Dag, A., Zarchin, S., Masci, T. & Shafir, S. (2015), 'Omega-3 deficiency impairs honey bee learning', *Proceedings of the National Academy of Sciences of the United States of America* **112**(51), 15761–15766.

Baker, D. & Czarnecki-Maulden, G. (1991), 'Comparative nutrition of cats and dogs', *Annual review of nutrition* **11**(1), 239–263.

Barragan-Fonseca, K. B., Dicke, M. & van Loon, J. J. (2018), 'Influence of larval density and dietary nutrient concentration on performance, body protein, and fat contents of black soldier fly larvae (hermetia illucens)', *Entomologia experimentalis et applicata* **166**(9), 761–770.

Barragan-Fonseca, K. B., Gort, G., Dicke, M. & van Loon, J. J. (2019), 'Effects of dietary protein and

carbohydrate on life-history traits and body protein and fat contents of the black soldier fly hermetia illucens', *Physiological Entomology* **44**(2), 148–159.

Barragan-Fonseca, K., Gort, G., Dicke, M. & Van Loon, J. (2021), 'Nutritional plasticity of the black soldier fly (hermetia illucens) in response to artificial diets varying in protein and carbohydrate concentrations', *Journal of Insects as Food and Feed* **7**(1), 51–61.

Behmer, S. T. (2009), 'Insect herbivore nutrient regulation', *Annual review of entomology* **54**, 165–187.

Behmer, S. T., Elias, D. O. & Bernays, E. A. (1999), 'Post-ingestive feedbacks and associative learning regulate the intake of unsuitable sterols in a generalist grasshopper', *Journal of Experimental Biology* **202**(6), 739–748.

Behmer, S. T. & Joern, A. (2008), 'Coexisting generalist herbivores occupy unique nutritional feeding niches', *Proceedings of the National Academy of Sciences of the United States of America* **105**(6), 1977–1982.

Bivand, R., Rundel, C., Pebesma, E., Stuetz, R., Hufthammer, K. O. & Bivand, M. R. (2017), 'Package 'rgeos'', *The comprehensive R archive network (CRAN)* .

Blonder, B. (2016), 'Do hypervolumes have holes?', *The American Naturalist* **187**(4), E93–E105. PMID: 27028084.

Bodenhofer, U., Bonatesta, E., Horejs-Kainrath, C. & Hochreiter, S. (2015), 'msa: an r package for multiple sequence alignment', *Bioinformatics* **31**(24), 3997–3999.

Boone, C. W., Kelloff, G. J. & Malone, W. E. (1990), 'Identification of candidate cancer chemopreventive agents and their evaluation in animal models and human clinical trials: a review', *Cancer Research* **50**(1), 2–9.

Borchers, H. W. (2022), *pracma: Practical Numerical Math Functions*. R package version 2.3.8. **URL:** *https://CRAN.R-project.org/package=pracma*

Box, G. E. P. & Wilson, K. B. (1951), 'On the experimental attainment of optimum conditions', *Journal of the Royal Statistical Society: Series B (Methodological)* **13**(1), 1–45.

Bradbury, E., Wilkinson, S., Cronin, G., Thomson, P., Bedford, M. & Cowieson, A. (2014), 'Nutritional geometry of calcium and phosphorus nutrition in broiler chicks. growth performance, skeletal health and intake arrays', *animal* **8**(7), 1071–1079.

Bray, G. A. & Ryan, D. H. (2021), 'Evidence-based weight loss interventions: individualized treatment options to maximize patient outcomes', *Diabetes, Obesity and Metabolism* **23**, 50–62.

Bunning, H., Bassett, L., Clowser, C., Rapkin, J., Jensen, K., House, C. M., Archer, C. R. & Hunt, J. (2016), 'Dietary choice for a balanced nutrient intake increases the mean and reduces the variance in the reproductive performance of male and female cockroaches', *Ecology and Evolution* **6**(14), 4711–4730.

Bunning, H., Rapkin, J., Belcher, L., Archer, C. R., Jensen, K. & Hunt, J. (2015), 'Protein and carbohydrate intake influence sperm number and fertility in male cockroaches, but not sperm viability', *Proceedings of the Royal Society B: Biological Sciences* **282**(1802), 20142144.

Campagna, R. & Perracchione, E. (2021), 'Data-driven extrapolation via feature augmentation based on variably scaled thin plate splines', *Journal of Scientific Computing* **88**(1), 15.

Campitelli, E. (2020), 'ggnewscale: Multiple fill and colour scales in'ggplot2", *R package version 0.4* **1**.

Canty, A. J. (2002), 'Resampling methods in r: the boot package', *The Newsletter of the R Project Volume* **2**(3), 2–7.

Carey, M. R., Archer, C. R., Rapkin, J., Castledine, M., Jensen, K., House, C. M., Hosken, D. J. & Hunt, J. (2022), 'Mapping sex differences in the effects of protein and carbohydrates on lifespan and reproduction in drosophila melanogaster: is measuring nutrient intake essential?', *Biogerontology* **23**(1), 129–144.

Carvalho-Santos, Z., Cardoso-Figueiredo, R., Elias, A. P., Tastekin, I., Baltazar, C. & Ribeiro, C. (2020), 'Cellular metabolic reprogramming controls sugar appetite in drosophila', *Nature Metabolism* **2**(9), 958–973.

Cavigliasso, F., Dupuis, C., Savary, L., Spangenberg, J. E. & Kawecki, T. J. (2020), 'Experimental evolution of post-ingestive nutritional compensation in response to a nutrient-poor diet', *Proceedings of the Royal Society B* **287**(1940), 20202684.

Chakraborty, A., Sgro, C. M. & Mirth, C. K. (2020), 'Does local adaptation along a latitudinal cline shape plastic responses to combined thermal and nutritional stress?', *Evolution* **74**(9), 2073–2087.

Chan, L., Vasilevsky, N., Thessen, A., McMurry, J. & Haendel, M. (2021), 'The landscape of nutri-informatics: a review of current resources and challenges for integrative nutrition research', *Database* **2021**.

Chandler, P., Cragle, R. & Gardiner, D. (1967), An investigation of the nutrition of the young dairy calf by response surface techniques. i. responses of holstein calves to varying dietary levels of calcium, phosphorus, and vitamin d3, Technical Report 429, University of Tennessee Agricultural Experiment Station, Knoxville.

Chandler, P., Kesler, E., McCarthy, R. & Johnston Jr, R. (1968), 'Effects of dietary lipid and protein on growth and nutrient utilization by dairy calves at ages 8 to 18 weeks', *The Journal of Nutrition* **95**(3), 452–460.

Chang, H.-C. & Wang, L.-C. (2010), 'A simple proof of thue's theorem on circle packing', *arXiv preprint arXiv:1009.4322* .

Charnov, E. L. (1976), 'Optimal foraging, the marginal value theorem', *Theoretical population biology* **9**(2), 129–136.

Chase, J. M. & Leibold, M. A. (2009), *Ecological niches: linking classical and contemporary approaches*, University of Chicago Press.

Cheng, K., Simpson, S. J. & Raubenheimer, D. (2008), 'A geometry of regulatory scaling', *The American Naturalist* **172**(5), 681–693.

Conceição, P. & Morimoto, J. (2022), "holey'niche! finding holes in niche hypervolumes using persistence homology', *Journal of Mathematical Biology* **84**(7), 58.

Cotter, S. C., Simpson, S. J., Raubenheimer, D. & Wilson, K. (2011), 'Macronutrient balance mediates trade-offs between immune function and life history traits', *Functional Ecology* **25**(1), 186–198.

Deans, C. A., Behmer, S. T., Fiene, J. & Sword, G. A. (2016), 'Spatio-temporal, genotypic, and environmental effects on plant soluble protein and digestible carbohydrate content: implications for insect herbivores with cotton as an exemplar', *Journal of chemical ecology* **42**, 1151–1163.

Debreu, G. (1954), 'Valuation equilibrium and pareto optimum', *Proceedings of the National Academy of Sciences of the United States of America* **40**(7), 588.

del Castillo, E., Chen, P., Meyers, A., Hunt, J. & Rapkin, J. (2022), 'Confidence regions for the location of response surface optima: the r package optimaregion', *Communications in Statistics-Simulation and Computation* **51**(12), 7074–7094.

del Castillo, E., Hunt, J. & Rapkin, J. (2016), 'Optimaregion: confidence regions for optima', CRAN: R package version 0.2.

Denny, J. C. & Collins, F. S. (2021), 'Precision medicine in 2030—seven ways to transform healthcare', *Cell* **184**(6), 1415–1419.

Do Carmo, M. P. (2016), *Differential geometry of curves and surfaces: revised and updated second edition*, Courier Dover Publications.

Dussutour, A. & Simpson, S. J. (2012), 'Ant workers die young and colonies collapse when fed a high-protein diet', *Proceedings of the Royal Society B: Biological Sciences* **279**(1737), 2402–2408.

Emmans, G. (1991), 'Diet selection by animals: theory and experimental design', *Proceedings of the Nutrition Society* **50**(1), 59–64.

Evans, E. & Miller, D. (1968), 'Comparative nutrition, growth and longevity', *Proceedings of the Nutrition Society* **27**(2), 121–129.

Fanson, B. G., Fanson, K. V. & Taylor, P. W. (2012), 'Cost of reproduction in the queensland fruit fly: Y-model versus lethal protein hypothesis', *Proceedings of the Royal Society B: Biological Sciences* **279**(1749), 4893–4900.

Fanson, B. G. & Taylor, P. W. (2012), 'Protein: carbohydrate ratios explain life span patterns found in queensland fruit fly on diets varying in yeast: sugar ratios', *Age* **34**, 1361–1368.

Fanson, B. G., Weldon, C. W., Pérez-Staples, D., Simpson, S. J. & Taylor, P. W. (2009), 'Nutrients, not caloric restriction, extend lifespan in queensland fruit flies (bactrocera tryoni)', *Aging cell* **8**(5), 514–523.

Fanson, B. G., Yap, S. & Taylor, P. W. (2012), 'Geometry of compensatory feeding and water consumption in drosophila melanogaster', *Journal of Experimental Biology* **215**(5), 766–773.

Felton, A. M., Felton, A., Raubenheimer, D., Simpson, S. J., Foley, W. J., Wood, J. T., Wallis, I. R. & Lindenmayer, D. B. (2009), 'Protein content of diets dictates the daily energy intake of a free-ranging primate', *Behavioral Ecology* **20**(4), 685–690.

Fragata, I., Blanckaert, A., Louro, M. A. D., Liberles, D. A. & Bank, C. (2019), 'Evolution in the light of fitness landscape theory', *Trends in ecology & evolution* **34**(1), 69–82.

Gage, M. & Cook, P. (1994), 'Sperm size or numbers? effects of nutritional stress upon eupyrene and apyrene sperm production strategies in the moth plodia interpunctella (lepidoptera: Pyralidea)', *Functional Ecology* pp. 594–599.

Gardiner, D., Cragle, R. & Chandler, P. (1967), An investigation of the nutrition of the young dairy calf by response surface techniques. ii. the response surface method as a biological research tool, Technical Report 429, University of Tennessee Agricultural Experiment Station, Knoxville.

Garratt, M., Nakagawa, S. & Simons, M. J. (2016), 'Comparative idiosyncrasies in life extension by reduced mtor signalling and its distinctiveness from dietary restriction', *Aging Cell* **15**(4), 737–743.

Gavrilets, S. (2004), *Fitness landscapes and the origin of species (MPB-41)*, Princeton University Press.

Gligorov, D., Sitnik, J. L., Maeda, R. K., Wolfner, M. F. & Karch, F. (2013), 'A novel function for the hox gene abd-b in the male accessory gland regulates the long-term female post-mating response in drosophila', *PLoS Genetics* **9**(3), e1003395.

Gonzalez, A. L., Dezerald, O., Marquet, P. A., Romero, G. Q. & Srivastava, D. S. (2017), 'The multidimensional stoichiometric niche', *Frontiers in Ecology and Evolution* **5**, 110.

Goodrich, B., Gabry, J., Ali, I. & Brilleman, S. (2020), 'rstanarm: Bayesian applied regression modeling via stan', *R package version* **2**(1).

Gosby, A. K., Conigrave, A. D., Raubenheimer, D. & Simpson, S. J. (2014), 'Protein leverage and energy intake', *Obesity Reviews* **15**(3), 183–191.

Grandison, R. C., Piper, M. D. W. & Partridge, L. (2009), 'Amino-acid imbalance explains extension of lifespan by dietary restriction in drosophila', *Nature* **462**(7276), 1061–1064.

Green, S. (2024), *Animal Models of Human Disease*, Cambridge University Press.

Gude, S. & Taga, M. E. (2020), 'Multi-faceted approaches to discovering and predicting microbial nutritional interactions', *Current opinion in biotechnology* **62**, 58–64.

Guo, J.-W., Cui, Y., Lin, P.-J., Zhai, B.-P., Lu, Z.-X., Chapman, J. W. & Hu, G. (2022), 'Male nutritional status does not impact the reproductive potential of female cnaphalocrocis medinalis moths under conditions of nutrient shortage', *Insect Science* **29**(2), 467–477.

Hadfield, J. D. (2010), 'Mcmc methods for multi-response generalized linear mixed models: the mcmcglmm r package', *Journal of statistical software* **33**, 1–22.

Hall, K. D. (2019), 'The potential role of protein leverage in the us obesity epidemic', *Obesity* **27**(8), 1222–1224.

Harrison, S. J., Raubenheimer, D., Simpson, S. J., Godin, J.-G. J. & Bertram, S. M. (2014), 'Towards a synthesis of frameworks in nutritional ecology: interacting effects of protein, carbohydrate and phosphorus on field cricket fitness', *Proceedings of the Royal Society B: Biological Sciences* **281**(1792), 20140539.

Hastie, T. & Hastie, M. T. (2015), 'Package 'gam'', *R package version* pp. 90124–3.

Havula, E., Ghazanfar, S., Lamichane, N., Francis, D., Hasygar, K., Liu, Y., Alton, L., Johnstone, J., Needham, E., Pulpitel, T. et al. (2022), 'Genetic variation of macronutrient tolerance in drosophila melanogaster', *Nature communications* **13**(1), 1–16.

Hawkes, M., Lane, S. M., Rapkin, J., Jensen, K., House, C. M., Sakaluk, S. K. & Hunt, J. (2022), 'Intralocus sexual conflict over optimal nutrient intake and the evolution of sex differences in life span and reproduction', *Functional Ecology* **36**(4), 865–881.

Heath, T. L. (1956), *The thirteen books of Euclid's Elements*, Courier Corporation.

Hewson-Hughes, A. K., Hewson-Hughes, V. L., Colyer, A., Miller, A. T., McGrane, S. J., Hall, S. R., Butterwick, R. F., Simpson, S. J. & Raubenheimer, D. (2012), 'Geometric analysis of macronutrient selection in breeds of the domestic dog, canis lupus familiaris', *Behavioral Ecology* **24**(1), 293–304.

Hewson-Hughes, A. K., Hewson-Hughes, V. L., Miller, A. T., Hall, S. R., Simpson, S. J. & Raubenheimer, D. (2011), 'Geometric analysis of macronutrient selection in the adult domestic cat, felis catus', *Journal of Experimental Biology* **214**(6), 1039–1051.

Hope, R. M., Hope, M. R. M. & Collate'CI, R. (2013), 'Package 'rmisc'', *group* **101**, 2.

Hosking, C. J., Raubenheimer, D., Charleston, M. A., Simpson, S. J. & Senior, A. M. (2019), 'Macronutrient intakes and the lifespan-fecundity trade-off: a geometric framework agent-based model', *Journal of the Royal Society Interface* **16**(151), 20180733.

House, C. M., Jensen, K., Rapkin, J., Lane, S., Okada, K., Hosken, D. J. & Hunt, J. (2016), 'Macronutrient balance mediates the growth of sexually selected weapons but not genitalia in male broad-horned beetles', *Functional Ecology* **30**(5), 769–779.

Hunt, J., Brooks, R., Jennions, M. D., Smith, M. J., Bentsen, C. L. & Bussiere, L. F. (2004), 'High-quality male field crickets invest heavily in sexual display but die young', *Nature* **432**(7020), 1024–1027.

Hutchinson, G. E. (1957), Population studies-animal ecology and demography-concluding remarks, *in* 'Cold Spring Harbor symposia on quantitative biology', Vol. 22, CSH Press, Bungtown Road, Cold Spring, pp. 415–427.

Inagaki, H. K., de Leon, S. B.-T., Wong, A. M., Jagadish, S., Ishimoto, H., Barnea, G., Kitamoto, T., Axel, R. & Anderson, D. J. (2012), 'Visualizing neuromodulation in vivo: Tango-mapping of dopamine signaling reveals appetite control of sugar sensing', *Cell* **148**(3), 583–595.

Jacobs Jr, D. R., Gross, M. D. & Tapsell, L. C. (2009), 'Food synergy: an operational concept for understanding nutrition', *The American journal of clinical nutrition* **89**(5), 1543S–1548S.

Jang, T. & Lee, K. P. (2018), 'Comparing the impacts of macronutrients on life-history traits in larval and adult drosophila melanogaster: the use of nutritional geometry and chemically defined diets', *Journal of Experimental Biology* **221**(21), jeb181115.

Jensen, K., McClure, C., Priest, N. K. & Hunt, J. (2015), 'Sex-specific effects of protein and carbohydrate intake on reproduction but not lifespan in drosophila melanogaster', *Aging Cell* **14**(4), 605–615.

Johnson, C. A., Raubenheimer, D., Rothman, J. M., Clarke, D. & Swedell, L. (2013), '30 days in the life: daily nutrient balancing in a wild chacma baboon', *PLoS One* **8**(7), e70383.

Krebs, J. R., Erichsen, J. T., Webber, M. I. & Charnov, E. L. (1977), 'Optimal prey selection in the great tit (parus major)', *Animal Behaviour* **25**, 30–38.

Kuhn, M. (2017), 'Caret: classification and regression training', R package version 6.0-78 (Contributions from Jed Wing, Steve Weston, Andre Williams, Chris Keefer, Allan Engelhardt, Tony Cooper, Zachary Mayer, Brenton Kenkel, the R Core Team, Michael Benesty, Reynald Lescarbeau, Andrew Ziem, Luca Scrucca, Yuan Tang, Can Candan, and Tyler Hunt).

Kutz, T. C., Sgrò, C. M. & Mirth, C. K. (2019), 'Interacting with change: Diet mediates how larvae respond to their thermal environment', *Functional Ecology* **33**(10), 1940–1951.

Lawton, D., Le Gall, M., Waters, C. & Cease, A. J. (2021), 'Mismatched diets: defining the nutritional landscape of grasshopper communities in a variable environment', *Ecosphere* **12**(3), e03409.

Le Couteur, D. G., Raubenheimer, D., Solon-Biet, S., de Cabo, R. & Simpson, S. J. (2024), 'Does diet influence aging? evidence from animal studies', *Journal of Internal Medicine* **295**(4), 400–415.

Le Gall, M., Word, M. L., Thompson, N., Manneh, B., Beye, A. & Cease, A. J. (2020), 'Linking land use and the nutritional ecology of herbivores: A case study with the senegalese locust', *Functional Ecology* **34**(1), 167–181.

Lee, K., Behmer, S., Simpson, S. & Raubenheimer, D. (2002), 'A geometric analysis of nutrient regulation in the generalist caterpillar spodoptera littoralis (boisduval)', *Journal of insect physiology* **48**(6), 655–665.

Lee, K. P., Behmer, S. T. & Simpson, S. J. (2006), 'Nutrient regulation in relation to diet breadth: a comparison of heliothis sister species and a hybrid', *Journal of Experimental Biology* **209**(11), 2076–2084.

Lee, K. P., Raubenheimer, D., Behmer, S. T. & Simpson, S. J. (2003), 'A correlation between macronutrient balancing and insect host-plant range: evidence from the specialist caterpillar spodoptera exempta (walker)', *Journal of Insect Physiology* **49**(12), 1161–1171.

Lee, K. P. & Roh, C. (2010), 'Temperature-by-nutrient interactions affecting growth rate in an insect ectotherm', *Entomologia Experimentalis et Applicata* **136**(2), 151–163.

Lee, K. P., Simpson, S. J., Clissold, F. J., Brooks, R., Ballard, J. W. O., Taylor, P. W., Soran, N. & Raubenheimer, D. (2008), 'Lifespan and reproduction in drosophila: new insights from nutritional geometry', *Proceedings of the National Academy of Sciences* **105**(7), 2498–2503.

Lei, X. G., Combs Jr, G. F., Sunde, R. A., Caton, J. S., Arthington, J. D. & Vatamaniuk, M. Z. (2022), 'Dietary selenium across species', *Annual Review of Nutrition* **42**, 337–375.

Lihoreau, M., Buhl, J., Charleston, M. A., Sword, G. A., Raubenheimer, D. & Simpson, S. J. (2015), 'Nutritional ecology beyond the individual: a conceptual framework for integrating nutrition and social interactions', *Ecology Letters* **18**(3), 273–286.

Lihoreau, M., Charleston, M. A., Senior, A. M., Clissold, F. J., Raubenheimer, D., Simpson, S. J. & Buhl, J. (2017), 'Collective foraging in spatially complex nutritional environments', *Philosophical Transactions of the Royal Society B: Biological Sciences* **372**(1727), 20160238.

Lihoreau, M., Poissonnier, L.-A., Isabel, G. & Dussutour, A. (2016), 'Drosophila females trade off good nutrition with high-quality oviposition sites when choosing foods', *Journal of Experimental Biology* **219**(16), 2514–2524.

Lowndes, J. S. S., Best, B. D., Scarborough, C., Afflerbach, J. C., Frazier, M. R., O'Hara, C. C., Jiang, N. & Halpern, B. S. (2017), 'Our path to better science in less time using open data science tools', *Nature ecology & evolution* **1**(6), 0160.

Ma, C., Mirth, C. K., Hall, M. D. & Piper, M. D. (2020), 'Amino acid quality modifies the quantitative availability of protein for reproduction in drosophila melanogaster', *Journal of Insect Physiology* p. 104050.

Ma, C., Mirth, C. K., Hall, M. D. & Piper, M. D. (2022), 'Amino acid quality modifies the quantitative availability of protein for reproduction in drosophila melanogaster', *Journal of insect physiology* **139**, 104050.

Ma, S., Avanesov, A. S., Porter, E., Lee, B. C., Mariotti, M., Zemskaya, N., Guigo, R., Moskalev, A. A. & Gladyshev, V. N. (2018), 'Comparative transcriptomics across 14 drosophila species reveals signatures of longevity', *Aging cell* **17**(4), e12740.

Ma, S. & Gladyshev, V. N. (2017), Molecular signatures of longevity: Insights from cross-species comparative studies, *in* 'Seminars in Cell & Developmental Biology', Vol. 70, Elsevier, pp. 190–203.

Machovsky-Capuska, G. E., Senior, A. M., Simpson, S. J. & Raubenheimer, D. (2016), 'The multidimensional nutritional niche', *Trends in Ecology & Evolution* **31**(5), 355–365.

Maklakov, A. A., Simpson, S. J., Zajitschek, F., Hall, M. D., Dessmann, J., Clissold, F., Raubenheimer, D., Bonduriansky, R. & Brooks, R. C. (2008), 'Sex-specific fitness effects of nutrient intake on reproduction and lifespan', *Current Biology* **18**(14), 1062–1066.

Malod, K., Archer, C. R., Hunt, J., Nicolson, S. W. & Weldon, C. W. (2017), 'Effects of macronutrient intake on the lifespan and fecundity of the marula fruit fly, ceratitis cosyra (tephritidae): Extreme lifespan in a host specialist', *Ecology and Evolution* **7**(22), 9808–9817.

Mattson Jr, W. J. (1980), 'Herbivory in relation to plant nitrogen content', *Annual review of ecology and systematics* **11**(1), 119–161.

May, C. E., Vaziri, A., Lin, Y. Q., Grushko, O., Khabiri, M., Wang, Q.-P., Holme, K. J., Pletcher, S. D., Freddolino, P. L., Neely, G. G. et al. (2019), 'High dietary sugar reshapes sweet taste to promote feeding behavior in drosophila melanogaster', *Cell reports* **27**(6), 1675–1685.

McDonald, J. M., Nabili, P., Thorsen, L., Jeon, S. & Shingleton, A. W. (2021), 'Sex-specific plasticity and the nutritional geometry of insulin-signaling gene expression in drosophila melanogaster', *EvoDevo* **12**(1), 6.

McFarland, D. & Sibly, R. (1972), "unitary drives' revisited', *Animal Behaviour* **20**(3), 548–563.

McFarland, D. & Sibly, R. (1975), 'The behavioural final common path', *Philosophical Transactions of the Royal Society of London. B, Biological Sciences* **270**(907), 265–293.

McMurdie, P. J. & Holmes, S. (2013), 'phyloseq: an r package for reproducible interactive analysis and graphics of microbiome census data', *PloS one* **8**(4), e61217.

Milman, V. (1998), Surprising geometric phenomena in high-dimensional convexity theory, *in* 'European Congress of Mathematics: Budapest, July 22–26, 1996 Volume II', Springer, pp. 73–91.

Mitchell, H. H. et al. (1962), 'Comparative nutrition of man and domestic animals. volume 1.', *Comparative nutrition of man and domestic animals. Volume 1.* **1**.

Moon, P. & Spencer, D. E. (1974), 'A geometry of nutrition', *The Journal of nutrition* **104**(12), 1535–1542.

Morimoto, J. (2022), 'Nutrigonometry ii: Experimental strategies to maximize nutritional information in multidimensional performance landscapes', *Ecology and Evolution* **12**(8), e9174.

Morimoto, J. (2023), 'Nutrigonometry iv: Thales' theorem to measure the rules of dietary compromise in animals', *Scientific Reports* **13**(1), 7466.

Morimoto, J. (2024), 'Optimum ratio of dietary protein and carbohydrate that maximises lifespan is shared among related insect species', *Aging Cell* **23**(3), e14067.

Morimoto, J. (2025*a*), 'Nutritional trade-offs in drosophila melanogaster', *Biology* **14**(4), 384.

Morimoto, J. (2025*b*), 'A reply to: A caveat about the use of trigonometric functions in statistical tests of nutritional geometry models', *Scientific Reports* **15**, 8322.

Morimoto, J., Conceição, P., Mirth, C. & Lihoreau, M. (2023), 'Nutrigonometry i: using right-angle triangles to quantify nutritional trade-offs in performance landscapes', *The American Naturalist* **201**(5), 725–740.

Morimoto, J., Conceição, P. & Smoczyk, K. (2022), 'Nutrigonometry iii: curvature, area and differences between performance landscapes', *Royal Society Open Science* **9**(11), 221326.

Morimoto, J. & Lihoreau, M. (2019), 'Quantifying nutritional trade-offs across multidimensional performance landscapes', *The American Naturalist* **193**(6), E168–E181.

Morimoto, J. & Lihoreau, M. (2020), 'Open data for open questions in comparative nutrition', *Insects* **11**(4), 236.

Morimoto, J., Senior, A., Ruiz, K., Wali, J. A., Pulpitel, T., Solon-Biet, S. M., Cogger, V. C., Raubenheimer, D., Le Couteur, D. G., Simpson, S. J. et al. (2019), 'Sucrose and starch intake contribute to reduced alveolar bone height in a rodent model of naturally occurring periodontitis', *Plos one* **14**(3), e0212796.

Morimoto, J. & Wigby, S. (2016), 'Differential effects of male nutrient balance on pre-and post-copulatory traits, and consequences for female reproduction in drosophila melanogaster', *Scientific Reports* **6**, 27673.

Munkemuller, T., Lavergne, S., Bzeznik, B., Dray, S., Jombart, T., Schiffers, K. & Thuiller, W. (2012), 'How to measure and test phylogenetic signal', *Methods in Ecology and Evolution* **3**(4), 743–756.

Murray-Rust, P. (2008), 'Open data in science', *Nature Precedings* pp. 1–1.

Myers, R. H., Montgomery, D. C. & Anderson-Cook, C. M. (2016), *Response surface methodology: process and product optimization using designed experiments*, John Wiley & Sons.

Nakagawa, S., Lagisz, M., Hector, K. L. & Spencer, H. G. (2012), 'Comparative and meta-analytic insights into life extension via dietary restriction', *Aging cell* **11**(3), 401–409.

Ng, S. H., Simpson, S. J. & Simmons, L. W. (2018), 'Macronutrients and micronutrients drive trade-offs between male pre-and postmating sexual traits', *Functional Ecology* **32**(10), 2380–2394.

Ng, S. H., Stat, M., Bunce, M., Simpson, S. J. & Simmons, L. W. (2019), 'Protein and carbohydrate intakes alter gut microbial community structure in crickets: a geometric framework approach', *FEMS Microbiology Ecology* **95**(8), fiz106.

Nie, Y., Wei, F., Zhou, W., Hu, Y., Senior, A. M., Wu, Q., Yan, L. & Raubenheimer, D. (2019), 'Giant pandas are macronutritional carnivores', *Current Biology* **29**(10), 1677–1682.

Nugent, R., Levin, C., Hale, J. & Hutchinson, B. (2020), 'Economic effects of the double burden of malnutrition', *The Lancet* **395**(10218), 156–164.

Nychka, D., Furrer, R., Paige, J. & Sain, S. (2017), 'fields: Tools for spatial data', R package version 9.6.

Paradis, E. & Schliep, K. (2019), 'ape 5.0: an environment for modern phylogenetics and evolutionary analyses in r', *Bioinformatics* **35**(3), 526–528.

Parks, J. R. (1973), 'Diet space and response surfaces', *Journal of Theoretical Biology* **42**(2), 349–358.

Parks, J. R. (1982), 'A general euclidean vector representation of mixtures', *A Theory of Feeding and Growth of Animals* pp. 158–180.

Pascacio-Villafán, C., Righini, N., Nestel, D., Birke, A., Guillén, L. & Aluja, M. (2022), 'Diet quality and conspecific larval density predict functional trait variation and performance in a polyphagous frugivorous fly', *Functional Ecology* **36**(5), 1163–1176.

Payne, P. & Wheeler, E. F. (1968), 'Comparative nutrition in pregnancy and lactation', *Proceedings of the Nutrition Society* **27**(2), 129–138.

Pebesma, E. & Bivand, R. S. (2005), 'Classes and methods for spatial data: the sp package', *R news* **5**(2), 9–13.

Pei, Z. & Liu, Z. (2009), Nutritional diet decision using multi-objective difference evolutionary algorithm, *in* '2009 International Conference on Computational Intelligence and Natural Computing', Vol. 1, IEEE, pp. 77–80.

Perry, J. C., Sirot, L. & Wigby, S. (2013), 'The seminal symphony: how to compose an ejaculate', *Trends in Ecology & Evolution* **28**(7), 414–422.

Piper, M. D., Soultoukis, G. A., Blanc, E., Mesaros, A., Herbert, S. L., Juricic, P., He, X., Atanassov, I., Salmonowicz, H. & Yang, M. (2017), 'Matching dietary amino acid balance to the in silico-translated exome optimizes growth and reproduction without cost to lifespan', *Cell Metabolism* **25**(3), 610–621.

Poissonnier, L.-A., Arganda, S., Simpson, S. J., Dussutour, A. & Buhl, J. (2018), 'Nutrition in extreme food specialists: An illustration using termites', *Functional ecology* **32**(11), 2531–2541.

Polak, M., Simmons, L. W., Benoit, J. B., Ruohonen, K., Simpson, S. J. & Solon-Biet, S. M. (2017), 'Nutritional geometry of paternal effects on embryo mortality', *Proceedings of the Royal Society B: Biological Sciences* **284**(1864), 20171492.

Ponton, F., Morimoto, J., Robinson, K., Kumar, S. S., Cotter, S. C., Wilson, K. & Simpson, S. J. (2020), 'Macronutrients modulate survival to infection and immunity in drosophila', *Journal of Animal Ecology* **89**(2), 460–470.

Ponton, F., Wilson, K., Cotter, S. C., Raubenheimer, D. & Simpson, S. J. (2011), 'Nutritional immunology: A multi-dimensional approach', *PLoS Pathogens* **7**(12), e1002223.

Ponton, F., Wilson, K., Holmes, A., Raubenheimer, D., Robinson, K. L. & Simpson, S. J. (2015), 'Macronutrients mediate the functional relationship between drosophila and wolbachia', *Proceedings of the Royal Society B: Biological Sciences* **282**(1800), 20142029.

Post, S. & Tatar, M. (2016), 'Nutritional geometric profiles of insulin/igf expression in drosophila melanogaster', *PloS one* **11**(5), e0155628.

Rapkin, J., Archer, C. R., Grant, C. E., Jensen, K., House, C. M., Wilson, A. J. & Hunt, J. (2017), 'Little evidence for intralocus sexual conflict over the optimal intake of nutrients for life span and reproduction in the black field cricket teleogryllus commodus', *Evolution* **71**(9), 2159–2177.

Rapkin, J., Jensen, K., Archer, C. R., House, C. M., Sakaluk, S. K., Del Castillo, E. & Hunt, J. (2018), 'The geometry of nutrient space–based life-history trade-offs: sex-specific effects of macronutrient intake on the trade-off between encapsulation ability and reproductive effort in decorated crickets', *The American Naturalist* .

Rapkin, J., Jensen, K., Lane, S. M., M, C., Sakaluk, S. K. & Hunt, J. (2016), 'Macronutrient intake regulates sexual conflict in decorated crickets', *Journal of Evolutionary Biology* **29**(2), 395–406.

Raubenheimer, D. (2011), 'Toward a quantitative nutritional ecology: the right-angled mixture triangle', *Ecological monographs* **81**(3), 407–427.

Raubenheimer, D., Machovsky-Capuska, G. E., Gosby, A. K. & Simpson, S. (2015), 'Nutritional ecology of obesity: from humans to companion animals', *British Journal of Nutrition* **113**(S1), S26–S39.

Raubenheimer, D., Senior, A. M., Mirth, C., Cui, Z., Hou, R., Le Couteur, D. G., Solon-Biet, S. M.,

Léopold, P. & Simpson, S. J. (2022), 'An integrative approach to dietary balance across the life course', *iScience* **25**(5).

Raubenheimer, D. & Simpson, S. (2020), *Eat Like the Animals: what nature teaches us about the science of healthy eating*, Houghton Mifflin.

Raubenheimer, D. & Simpson, S. J. (1992), 'Analysis of covariance: an alternative to nutritional indices', *Entomologia experimentalis et applicata* **62**(3), 221–231.

Raubenheimer, D. & Simpson, S. J. (1993), 'The geometry of compensatory feeding in the locust', *Animal behaviour* **45**(5), 953–964.

Raubenheimer, D. & Simpson, S. J. (1997), 'Integrative models of nutrient balancing: application to insects and vertebrates', *Nutrition research reviews* **10**(1), 151–179.

Raubenheimer, D. & Simpson, S. J. (1999), Integrating nutrition: a geometrical approach, *in* 'Proceedings of the 10th International Symposium on Insect-Plant Relationships', Springer, pp. 67–82.

Raubenheimer, D. & Simpson, S. J. (2016), 'Nutritional ecology and human health', *Annual review of nutrition* **36**, 603–626.

Raubenheimer, D. & Simpson, S. J. (2019), 'Protein leverage: theoretical foundations and ten points of clarification', *Obesity* **27**(8), 1225–1238.

Raubenheimer, D., Simpson, S. J. & Mayntz, D. (2009), 'Nutrition, ecology and nutritional ecology: toward an integrated framework', *Functional ecology* **23**(1), 4–16.

Reddiex, A. J., Gosden, T. P., Bonduriansky, R. & Chenoweth, S. F. (2013), 'Sex-specific fitness consequences of nutrient intake and the evolvability of diet preferences', *American Naturalist* **182**(1), 91–102.

Reichman, O. J., Jones, M. B. & Schildhauer, M. P. (2011), 'Challenges and opportunities of open data in ecology', *Science* **331**(6018), 703–705.

Revell, L. J. (2012), 'phytools: an r package for phylogenetic comparative biology (and other things)', *Methods in ecology and evolution* (2), 217–223.

Revell, L. J. & Harmon, L. J. (2022), *Phylogenetic comparative methods in R*, Princeton University Press.

Revell, L. J., Harmon, L. J. & Collar, D. C. (2008), 'Phylogenetic signal, evolutionary process, and rate', *Systematic biology* **57**(4), 591–601.

Rho, M. S. & Lee, K. P. (2025), 'Macronutrient balance dictates lifespan and reproduction in a beetle, tenebrio molitor', *Journal of Experimental Biology* **228**(12), jeb250281.

Ribeiro, C. & Dickson, B. J. (2010), 'Sex peptide receptor and neuronal tor/s6k signaling modulate nutrient balancing in drosophila', *Current Biology* **20**(11), 1000–1005.

Rodgers, G. P. & Collins, F. S. (2020), 'Precision nutrition—the answer to "what to eat to stay healthy"', *Jama* **324**(8), 735–736.

Rodrigues, M. A., Martins, N. E., Balancé, L. F., Broom, L. N., Dias, A. J., Fernandes, A. S. D., Rodrigues, F., Sucena, É. & Mirth, C. K. (2015), 'Drosophila melanogaster larvae make nutritional choices that minimize developmental time', *Journal of insect physiology* **81**, 69–80.

Roff, D. A. (2002), *Life history evolution*, Vol. 7, Sinauer Associates Sunderland.

Rohatgi, A. (2019), 'Webplotdigitizer 4.2: Html5 based online tool to extract numerical data from plot images'.

Rothman, J. M., Raubenheimer, D. & Chapman, C. A. (2011), 'Nutritional geometry: gorillas prioritize non-protein energy while consuming surplus protein', *Biology Letters* . rsbl20110321.

Roush, W., Petersen, R. G. & Arscott, G. H. (1979), 'An application of response surface methodology to research in poultry nutrition', *Poultry Science* **58**(6), 1504–1513.

Ruohonen, K., Kettunen, J. & King, J. (2001), 'Experimental design in feeding experiments', *Food intake in fish* pp. 88–107.

Ruohonen, K., Simpson, S. J. & Raubenheimer, D. (2007), 'A new approach to diet optimisation: a re-analysis using european whitefish (coregonus lavaretus)', *Aquaculture* **267**(1-4), 147–156.

Sauvant, D., Schmidely, P., Daudin, J.-J. & St-Pierre, N. R. (2008), 'Meta-analyses of experimental data in animal nutrition', *animal* **2**(8), 1203–1214.

Schwenke, R. A., Lazzaro, B. P. & Wolfner, M. F. (2016), 'Reproduction–immunity trade-offs in insects', *Annual Review of Entomology* **61**, 239–256.

Semaniuk, U., Feden'ko, K., Yurkevych, I. S., Storey, K. B., Simpson, S. J. & Lushchak, O. (2018), 'Within-diet variation in rates of macronutrient consumption and reproduction does not accompany

changes in lifespan in drosophila melanogaster', *Entomologia Experimentalis et Applicata* **166**(1), 74–80.

Senior, A. M. (2023), 'Estimating genetic variance in life-span response to diet: Insights from statistical simulation', *The Journals of Gerontology: Series A* **78**(3), 392–396.

Senior, A. M., Andrews, C. J., Simpson, S. J. & Raubenheimer, D. (2025), 'A caveat about the use of trigonometric functions in statistical tests of nutritional geometry models', *Scientific Reports* **15**, 8294.

Sentinella, A. T., Crean, A. J. & Bonduriansky, R. (2013), 'Dietary protein mediates a trade-off between larval survival and the development of male secondary sexual traits', *Functional Ecology* **27**(5), 1134–1144.

Shik, J. Z., Schal, C. & Silverman, J. (2014), 'Diet specialization in an extreme omnivore: nutritional regulation in glucose-averse german cockroaches', *Journal of Evolutionary Biology* **27**(10), 2096–2105.

Sibly, R. & Calow, P. (1983), 'An integrated approach to life-cycle evolution using selective landscapes', *Journal of Theoretical Biology* **102**(4), 527–547.

Sibly, R. & McFarland, D. (1974), 'State space approach to motivation, motivational control system analysis'.

Sibly, R. & McFarland, D. (1976), 'On the fitness of behavior sequences', *The American Naturalist* **110**(974), 601–617.

Silva-Soares, N. F., Nogueira-Alves, A., Beldade, P. & Mirth, C. K. (2017), 'Adaptation to new nutritional environments: larval performance, foraging decisions, and adult oviposition choices in drosophila suzukii', *BMC ecology* **17**, 1–13.

Simpson, S. & Abisgold, J. (1985), 'Compensation by locusts for changes in dietary nutrients: behavioural mechanisms', *Physiological Entomology* **10**(4), 443–452.

Simpson, S. J., Batley, R. & Raubenheimer, D. (2003), 'Geometric analysis of macronutrient intake in humans: the power of protein?', *Appetite* **41**(2), 123–140.

Simpson, S. J., Clissold, F. J., Lihoreau, M., Ponton, F., Wilder, S. M. & Raubenheimer, D. (2015), 'Recent advances in the integrative nutrition of arthropods', *Annual Review of Entomology* **60**(1), 293–311.

Simpson, S. J., Le Couteur, D. G., James, D. E., George, J., Gunton, J. E., Solon-Biet, S. M. & Raubenheimer, D. (2017), 'The geometric framework for nutrition as a tool in precision medicine', *Nutrition and healthy aging* **4**(3), 217–226.

Simpson, S. J. & Raubenheimer, D. (1993), 'A multi-level analysis of feeding behaviour: the geometry of nutritional decisions', *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **342**(1302), 381–402.

Simpson, S. J. & Raubenheimer, D. (1999), 'Assuaging nutritional complexity: a geometrical approach', *Proceedings of the Nutrition Society* **58**(4), 779–789.

Simpson, S. J. & Raubenheimer, D. (2005), 'Obesity: the protein leverage hypothesis', *obesity reviews* **6**(2), 133–142.

Simpson, S. J. & Raubenheimer, D. (2009), 'Macronutrient balance and lifespan', *Aging (Albany NY)* **1**(10), 875.

Simpson, S. J. & Raubenheimer, D. (2012), *The nature of nutrition*, Princeton university press.

Simpson, S. J. & Raubenheimer, D. (2014), 'Perspective: tricks of the trade', *Nature* **508**(7496), S66–S66.

Simpson, S. J., Raubenheimer, D., Charleston, M. A. & Clissold, F. J. (2010), 'Modelling nutritional interactions: from individuals to communities', *Trends in Ecology & Evolution* **25**(1), 53–60.

Simpson, S. J., Sibly, R. M., Lee, K. P., Behmer, S. T. & Raubenheimer, D. (2004), 'Optimal foraging when regulating intake of multiple nutrients', *Animal behaviour* **68**(6), 1299–1311.

Simpson, S. J., Sword, G. A., Lorch, P. D. & Couzin, I. D. (2006), 'Cannibal crickets on a forced march for protein and salt', *Proceedings of the National Academy of Sciences* **103**(11), 4152–4156.

Simpson, S. & Raubenheimer, D. (1995), 'The geometric analysis of feeding and nutrition: a user's guide', *Journal of Insect Physiology* **41**(7), 545–553.

Simpson, S. & Raubenheimer, D. (2001), 'A framework for the study of macronutrient intake in fish', *Aquaculture research* **32**(6), 421–432.

Simpson, S., Raubenheimer, D., Behmer, S., Whitworth, A. & Wright, G. (2002), 'A comparison of nutritional regulation in solitarious-and gregarious-phase nymphs of the desert locust schistocerca gregaria', *Journal of Experimental Biology* **205**(1), 121–129.

Skorupa, D. A., Dervisefendic, A., Zwiener, J. & Pletcher, S. D. (2008), 'Dietary composition specifies consumption, obesity, and lifespan in drosophila melanogaster', *Aging cell* **7**(4), 478–490.

Solon-Biet, S. M., Cogger, V. C., Pulpitel, T., Heblinski, M., Wahl, D., McMahon, A. C., Warren, A., Durrant-Whyte, J., Walters, K. A., Krycer, J. R. et al. (2016), 'Defining the nutritional and metabolic context of fgf21 using the geometric framework', *Cell metabolism* **24**(4), 555–565.

Solon-Biet, S. M., McMahon, A. C., Ballard, J. W. O., Ruohonen, K., Wu, L. E., Cogger, V. C., Warren, A., Huang, X., Pichaud, N., Melvin, R. G. et al. (2014), 'The ratio of macronutrients, not caloric intake, dictates cardiometabolic health, aging, and longevity in ad libitum-fed mice', *Cell metabolism* **19**(3), 418–430.

Solon-Biet, S. M., Walters, K. A., Simanainen, U. K., McMahon, A. C., Ruohonen, K., Ballard, J. W. O., Raubenheimer, D., Handelsman, D. J., Le Couteur, D. G. & Simpson, S. J. (2015), 'Macronutrient balance, reproductive function, and lifespan in aging mice', *Proceedings of the national academy of sciences* **112**(11), 3481–3486.

Stearns, S. C. (1989), 'Trade-offs in life-history evolution', *Functional Ecology* **3**(3), 259–268.

Stearns, S. C. (1998), *The evolution of life histories*, Oxford university press.

Stephens, D. W. & Krebs, J. R. (1986), *Foraging theory*, Vol. 1, Princeton university press.

Swanson, E. M., Espeset, A., Mikati, I., Bolduc, I., Kulhanek, R., White, W. A., Kenzie, S. & Snell-Rood, E. C. (2016), 'Nutrition shapes life-history evolution across species', *Proceedings of the Royal Society B: Biological Sciences* **283**(1834), 20152764.

Talal, S., Harrison, J. F., Farington, R., Youngblood, J. P., Medina, H. E., Overson, R. & Cease, A. (2023), 'Body mass and growth rates predict protein intake across animals', *bioRxiv* pp. 2023–06.

Taylor, B. (1715), *Methodus incrementorum directa & inversa. Auctore Brook Taylor, LL. D. & Regiae Societatis Secretario*, typis Pearsonianis: prostant apud Gul. Innys ad Insignia Principis in ....

Team, R. D. C. (2017), *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria.
URL: *http://www.r-project.org/*

Team, R. D. C. (2019), *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria.
URL: *http://www.r-project.org/*

Tilman, D. (1982), *Resource competition and community structure*, Princeton university press.

Toyomizu, M., Akiba, Y., Horiguchi, M. & Matsumoto, T. (1982), 'Multiple regression and response surface analyses of the effects of dietary protein, fat and carbohydrate on the body protein and fat gains in growing chicks', *The Journal of nutrition* **112**(5), 886–896.

Toyomizu, M., Hayashi, K., Yamashita, K. & Tomita, Y. (1988), 'Response surface analyses of the effects of dietary protein on feeding and growth patterns in mice from weaning to maturity', *The Journal of nutrition* **118**(1), 86–92.

Toyomizu, M., Kimura, S. & Tomita, Y. (1993), 'Response surface analyses of the effects of dietary protein, fat and carbohydrate on feeding and growth pattern in mice from weaning to maturity', *Animal Science* **56**(2), 251–259.

Toyomizu, M., Matsukubo, M., Hayashi, K. & Tomita, Y. (1991), 'Response surface analyses of the effects of dietary fat on feeding and growth pattern in mice from weaning to maturity', *Animal Science* **52**(1), 207–214.

Treidel, L. A., Clark, R. M., Lopez, M. T. & Williams, C. M. (2021), 'Physiological demands and nutrient intake modulate a trade-off between dispersal and reproduction based on age and sex of field crickets', *Journal of Experimental Biology* **224**(7), jeb237834.

Trumper, S. & Simpson, S. J. (1993), 'Regulation of salt intake by nymphs of locusta migratoria', *Journal of Insect Physiology* **39**(10), 857–864.

Vaudo, A. D., Patch, H. M., Mortensen, D. A., Tooker, J. F. & Grozinger, C. M. (2016), 'Macronutrient ratios in pollen shape bumble bee (bombus impatiens) foraging strategies and floral preferences', *Proceedings of the National Academy of Sciences* **113**(28), E4035–E4042.

Wadhwa, R. R., Williamson, D. F., Dhawan, A. & Scott, J. G. (2018), 'Tdastats: R pipeline for computing persistent homology in topological data analysis', *Journal of open source software* **3**(28), 860.

Walker, S. J., Corrales-Carvajal, V. M. & Ribeiro, C. (2015), 'Postmating circuitry modulates salt taste processing to increase reproductive output in drosophila', *Current Biology* **25**(20), 2621–2630.

Wang, G. & Bai, L. (2006), Game model based co-evolutionary algorithm and its application for multiobjective nutrition decision making optimization problems, *in* 'International Conference on Computational and Information Science', Springer, pp. 177–183.

Watanabe, J. (2022), 'Detecting (non) parallel evolution in multidimensional spaces: angles, correlations and eigenanalysis', *Biology Letters* **18**(2), 20210638.

Weinberger, S. (2011), 'What is... persistent homology', *Notices of the AMS* **58**(1), 36–39.

Weller, H. I., Van Belleghem, S. M., Hiller, A. E. & Lord, N. P. (2022), 'Flexible color segmentation of biological images with the r package recolorize', *bioRxiv* pp. 2022–04.

White, P. J. & Brown, P. (2010), 'Plant nutrition for sustainable development and global health', *Annals of botany* **105**(7), 1073–1080.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J. et al. (2019), 'Welcome to the tidyverse', *Journal of open source software* **4**(43), 1686.

Wickham, H., Chang, W. & Wickham, M. H. (2016), 'Package 'ggplot2", *Create Elegant Data Visualisations Using the Grammar of Graphics. Version* **2**(1), 1–189.

Wickham, H. & Wickham, H. (2016), *Data analysis*, Springer.

Wilder, S. M., Le Couteur, D. G. & Simpson, S. J. (2013), 'Diet mediates the relationship between longevity and reproduction in mammals', *Age* **35**, 921–927.

Wood, S. & Wood, M. (2015), 'Package 'mgcv. r package version, 1 (29), 729'.

World Health Organization (2024), 'Malnutrition', `https://www.who.int/news-room/fact-sheets/detail/malnutrition`. Accessed: 11 October 2024.

Wright, S. (1932), The roles of mutation, inbreeding, crossbreeding and selection in evolution, *in* 'Sixth Int. Cong. Genet', Vol. 1, p. 356.

Wu, G. (2017), *Principles of animal nutrition*, crc Press.

Yapici, N., Kim, Y. J., Ribeiro, C. & Dickson, B. J. (2008), 'A receptor that mediates the post-mating switch in drosophila reproductive behaviour', *Nature* **451**(7174), 33–U1.

Zanco, B., Morimoto, J., Cockerell, F., Mirth, C. & Sgrò, C. M. (2025), 'Nutritional optima for life-history traits vary with temperature and across locally-adapted populations', *Journal of Insect Physiology* p. 104815.

Zomorodian, A. & Carlsson, G. (2004), Computing persistent homology, *in* 'Proceedings of the twentieth annual symposium on Computational geometry', pp. 347–356.