2023 IEEE Symposium Series on Computational Intelligence (SSCI)
Mexico City, Mexico. December 5-8, 2023

Exploring the Uncertainty of Approximated Fitness Landscapes via Gaussian Process Realisations

Melike D. Karatas
Department of Computer Science
University of Exeter
Exeter, United Kingdom
M.D.Karatas@exeter.ac.uk

Marc Goodfellow
Department of Mathematics
University of Exeter
Exeter, United Kingdom
M.Goodfellow@exeter.ac.uk

Jonathan E. Fieldsend
Department of Computer Science
University of Exeter
Exeter, United Kingdom
J.E.Fieldsend@exeter.ac.uk

***Abstract*—Gaussian processes (GPs) serve as powerful surrogate models in optimisation by providing a flexible data-driven framework for representing complex fitness landscapes. We provide an analysis of realisations drawn from GP models of fitness landscapes—which represent alternative coherent fits to the data—and use a network-based approach to investigate their induced landscape consistency. We consider the variation of constructed local optima networks (LONs: which provide a condensed representation of landscapes), analyse the fitness landscapes of GP realisations, and delve into the uncertainty associated with graph metrics of LONs. Our findings contribute to the understanding and practical application of GPs in optimisation and landscape analysis. Particularly that landscape consistency between GP realisations can vary considerably dependent on the model fit and underlying landscape complexity of the optimisation problem.**

***Index Terms*—Gaussian Processes, Uncertainty Quantification, Fitness Landscapes, Local Optima Networks**

I. Introduction

The fundamental goal of optimisation is to discover the most favourable solution within a specified domain, often subject to certain constraints. Many real-world problems require optimising objective functions that are complex, exhibit multiple modes, and are computationally demanding to evaluate. Such factors intensify the demand for surrogate-based optimisation strategies [1], [2]. Within this context, Gaussian Processes (GPs) play a significant role [3], [4]. As a probabilistic model, GPs offer a means of approximating the objective function in a manner that quantifies uncertainty with a particularly powerful feature: their ability to produce a distribution of possible outputs, or realisations, rather than a single mean prediction (reflecting their uncertainty).

Here we present a detailed analysis of GP realisations drawn from fits to a set of benchmark functions. Instead of relying on the mean GP fit, we adopt a more comprehensive approach, considering the entire set of potential functions represented by the GP posterior. This enables us to capture the inherent uncertainty in the model and provides a more complete picture of the possible relationships present in the data. We further consider the structure of these GP realisations through the construction of Local Optima Networks (LONs) [5]. Each LON is calculated for a multitude of GP fits and for each fit, we generate a set of realisations. We apply graph theory techniques to these LONs, computing network metrics such as degree centrality, closeness centrality, and the number of nodes and edges. When calculating a network measure on the mean Gaussian process (GP), we obtain an estimated measure with an associated level of uncertainty, reflecting our confidence in the estimate based on the specific GP model. Conversely, by calculating the same measure on an ensemble of realisations, we utilise multiple models to estimate the measure and, consequently, assess the variance of the measure across the ensemble.

Working with GP realisations allows us to comprehend how uncertainties in the GP model translate into uncertainties in the network metrics. See [6] for a general study on uncertainty of graph measures. This research aims to enhance our understanding of GPs as surrogates of fitness landscapes by exploring their capabilities in the context of varying sample sizes and diverse benchmark functions. By specifically focusing on GP realisations and employing rigorous landscape analysis techniques using LONs, we seek to establish a foundation for a more nuanced and comprehensive utilisation of GPs in complex predictive modelling tasks.

The rest of the paper is organised as follows. Section II reviews the literature on GPs and LONs. Section III provides details of the experimental setup, Section IV presents the results obtained in this study and finally Section V concludes the paper with a discussion on the results and future work.

II. Background

A. Gaussian Process

A GP is a probabilistic model widely employed in machine learning, statistics, and optimisation [7]. It serves as a powerful tool for modelling and analysing functions by defining a distribution over them. A GP represents a set of random variables, with any finite subset of these variables following a joint multivariate Gaussian distribution. This enables a GP to describe the distribution of functions, treating each function as a random variable.

A GP is characterised by a mean function and a covariance function, with mean $\mu(x)$ and variance $\sigma^2(x)$ respectively, i.e. $p(f|x, D, \theta) = \mathcal{N}(\mu(x), \sigma^2(x))$ where the mean and

This work was supported by the Engineering and Physical Sciences Research Council [grant number EP/N017846/1].

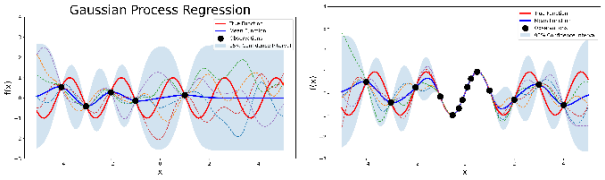

Fig. 1: GP fit for Sine Function (shown in red) including the mean GP (shown in blue) and some realisations (dashed lines). The shaded area around the mean GP represents the uncertainty, which usually shrinks as more observations are incorporated (unless the length scale changes).

variance are $\mu(x) = \kappa(x, X)K^{-1}f$, and $\sigma^2(x) = \kappa(x, x) - \kappa(x, X)K^{-1}\kappa(X, x)$. Where X is a matrix of size $M \times d$ representing the design locations, while f is a corresponding vector of size M that contains the true function evaluations. Thus, the dataset D can be expressed as $\{(X, f)\}$. The covariance matrix K of size $M \times M$ represents the covariance function $\kappa(x, x'; \theta)$ evaluated for each pair of observations. $\kappa(x, X)$ is a vector of size M representing the covariances between x and each of the observations. θ is the kernel hyper-parameters.

Kernel functions define the covariance between different points in the input space, reflecting the similarity or correlation between function values at those points. Various types of kernel functions exist, each with its own characteristics and suitability for different types of functions. We use in this paper are the RBF, Matérn 3/2 and Linear kernels in this work.

When data are observed, a GP model updates its prior distribution to form a posterior distribution. As depicted in Figure 1, as more data are observed, the GP realisations converge closer to the mean function. The posterior distribution represents the updated beliefs regarding the function values at unobserved locations based on the observed data.[1] This process obtains the posterior GP mean and covariance. Subsequently, samples are drawn from a multivariate Gaussian distribution with this mean and covariance, representing possible function values that align with the observed data and the assumptions of the GP model. Prior research [9], [10] explored efficient posterior distribution sampling. Insights for handling GP hyper-parameters in Bayesian optimisation were studied [11]. Earlier work [12] emphasised graph theory and GPs for comparing posterior distributions of finite population variance.

B. Local Optima Networks

A fitness landscape refers to the relationship between the performance or fitness of solutions and their positions in the problem's search space. By analysing the characteristics of the fitness landscape, such as its ruggedness, smoothness, or presence of local optima, optimisation algorithms can adapt their search strategies to efficiently navigate and find optimal or near-optimal solutions.

[1]We use the GPy package for our GP fit and to draw realisations [8].

LONs are graph representations that capture the relationships between local optima in a fitness landscape [5] [13]. In LONs, nodes represent individual local optima, while edges denote the connections between these optima. The presence of an edge between two nodes signifies the feasibility of transitioning from one local optimum to another under some neighbourhood function and transition mechanism. Weighted edges offer additional information about the strength or significance of the connections.

LONs have been extensively studied in various domains. Initially, they were investigated for combinatorial optimisation problems [14]–[16]. Subsequently, their application was extended to continuous optimisation problems [17], [18], where the construction of LONs involved the use of "basin-hopping" [19] and Nelder-Mead algorithms [20]. Additionally, search trajectory networks were proposed for population-based algorithms in continuous spaces [21] and their properties have been studied by graph theory [17], [22], [23]. LONs have been extensively studied for example, landscape-aware algorithm configuration, exploring neutral and rugged landscapes [24], characterising constrained continuous optimisation problems [25], visualising multi-objective landscapes [26].

Local Optima (L). The set of solutions of an optimisation problem, whose fitness is superior to all other solutions in their neighbourhood. For example, in a minimisation problem, L is a local optimum if its fitness value is not worse than all other solutions in its neighbourhood: $\forall x \in N(L) : f(x) \geq f(L)$, where $N(L)$ is the neighbourhood of L.

Edge (E). Search transitions among local optima are represented by directed and weighted edges. The weight w_{ij} of an edge from an optimum L_i to an optimum L_j represents the probability of the transition.

Basin of attraction. The basin of attraction B_i of a local optimum L_i in the search space X, is the set $B_i = \{ \mathbf{x} \in X \mid optimiser(\mathbf{x}) = L_i \}$. The cardinality $|B_i|$ of this set gives the size of the basin of attraction of L_i.

Local Optima Network (LON). A directed and weighted graph $LON = (L, E)$, which compacts a fitness landscape by taking local optima in the search space as nodes and connecting these nodes with edges based on their transition as a result of search operators.

III. EXPERIMENTAL SETUP

To ensure the uncertainty in our final results is not unduly influenced by variation in the LON generation process, we conducted experiments on known functions with varying sample sizes. LONs were created on the original function with sample sizes ranging from 100 to 10,000 (see Figure 3). Once convergence was achieved, the random walks were saved for generating LONs in subsequent experiments. Our goal was to examine the uncertainty stemming from different GP fits and realisations. We performed 30 GP fits, details of which can be found in Section III-B. The results of these GP fits are presented in Table II. Each realisation trajectory was then used to construct a corresponding LON, resulting in a total of 30 LONs. Graph measures were calculated for these

TABLE I: Optimisation Benchmark Functions with Boundaries and Optima.

Function	Optima	
	# Local Optima	Global Optimum $x, n = 2$
Schwefel	49	(420.9687, 420.9687)
Levy	40	(1, 1)
Griewank	5	(0, 0)
Styblinski-Tang	4	(-2.9035, -2.9035)

LONs, as described in Section III-D. Our objective was to compare the 'true' graph measures obtained from the actual functions to those from the approximated lanscapes. Latin Hypercube Sampling (LHS) was employed for random walks in generating LONs, and the same method was used for fitting the GPs.

A. Optimisation Functions

The functions used in this paper, sourced from [27], are listed in Table I.

B. GP Hyperparameters

We have used three different kernel functions. Our goal was to identify the best fit for the given data rather than specifically analysing the effects of individual kernel functions. For each iteration, all kernels were employed. To evaluate the performance of each kernel function, cross-validation was conducted. The data was partitioned into training and validation sets with 70% for training and 30% for testing. The kernel function that achieved the highest performance on the validation set was chosen. To determine the optimal values for the variance and lengthscale of each kernel, we used an approach based on [28] to systematically explore hyperparameter combinations and find the settings that maximised the model's performance using cross-validation.

C. Graph Construction

LONs were constructed following the method described in [18], specifically referring to Algorithm 1 and Algorithm 2. Regarding the Nelder-Mead (NM) hyperparameters, the initial simplex size was set to 0.05. If any coordinates had a value of zero, the initial simplex size was adjusted to 0.025, following the default value for initial simplex generation in NM [29]. To update the discovered nodes, a mean distance measure was utilized. Similarly, for edges, the algorithm employed the mean distances of the initially found nodes using a pivoting approach.

D. Graph Theory Metrics & Statistical Analysis

Network analysis relies on key metrics to understand network structure and behavior. Assortativity measures node connections based on attributes, while closeness centrality indicates information flow speed. Degree centrality quantifies node connections and density reflects network compactness. In-degree and out-degree centrality assess node attractiveness and accessibility. Node and edge count reveal size and complexity, while PageRank identifies influential nodes and information flow patterns.

IV. RESULTS

Note that the results are averages and uncertainties derived from 30 separate Gaussian process fits, conducted with different sample sizes to fit GPs. Additionally, 30 realisations were obtained from each GP to generate violin plots. By calculating the average of these we obtain 30×30 distinct network measures.

We first present the network measures we acquired from LONs for the actual functions. Our primary objective is to determine the true values of these network measures and identify the sample sizes at which convergence occurs. This is to assure ourselves (as much as possible) that any uncertainty observed in subsequent experiments is not derived from the domain sampling process to generate the LONs. Boxplots of 2D functions in Figure 3 show that a maximum of 10,000 random walks effectively covered the landscapes across all the functions. While Schwefel started with a high range in the measures calculated for the smallest sample size (100) the measures converge around a size of 10,000 random walks, Levy and Griewank trends converged around 5,000 walks, and for Styblinski-Tang, even 100 walks sufficed. The error metrics in Table II across 30 GP iterations demonstrates their utility in quantifying model discrepancy from actual values. Discrepancies for Schwefel and Levy function reduced dramatically from sample sizes of 200, while Griewank required more samples, and Styblinski-Tang needed only 50 for a good GP fit. Figure 4 illustrates the correlation length over time, showing it converges with increasing sample size without overfitting or underfitting. The analysis of node degree assortativity reveals a consistent trend across functions. However, it is important to note that the GP fit for the Griewank function was based on a relatively small sample size of 50. This limited sample size introduces some uncertainties in the results obtained for the Griewank function. Despite this limitation, it is noteworthy that the uncertainties observed across all fits and realisations are not significant. Closeness centrality varied across functions, as did degree centrality and density measures, which were more volatile for GP fits with smaller observation sets. Meanwhile, in-degree and out-degree centrality reflected the average number of connections per node, and PageRank helped track changes in node influence over time. We see the most variability for Griewank function where the values for example for density bicenteres around two points 0.6 and 0.45. This can be explained on the table II for 400 sample size to fit GP the R^2 value is 0.679 whereas for the other functions this value if around 0.9.

V. CONCLUSIONS & FUTURE WORK

1 This research focused on the intricate interplay between GP fits and the ensuing LONs derived from four different functions. We systematically conducted 30 GP fits per function, drew realisations, and evaluated the similarities/discrepancies between the resultant LONs via graph measures. The key discovery was the distinct variation in LON measures with GP sample size changes. Significant impacts were observed

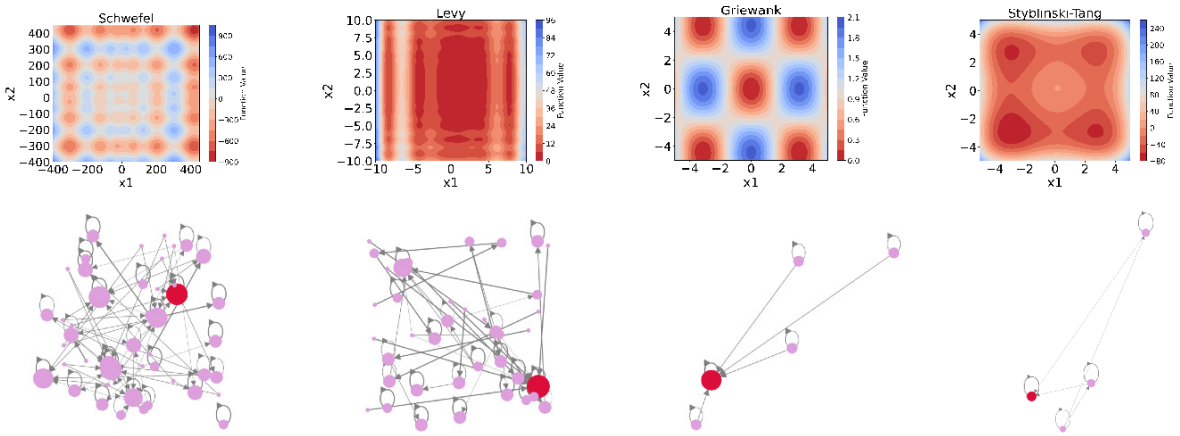

Fig. 2: *Top row:* Contour-plots of the benchmark functions *Bottom row:* Corresponding LONs. Red nodes indicate global optima, while blue nodes show local optima. The size of each node is proportional to the size of its basin of attraction. Edges are weighted based on the probabilities of transition among optima.

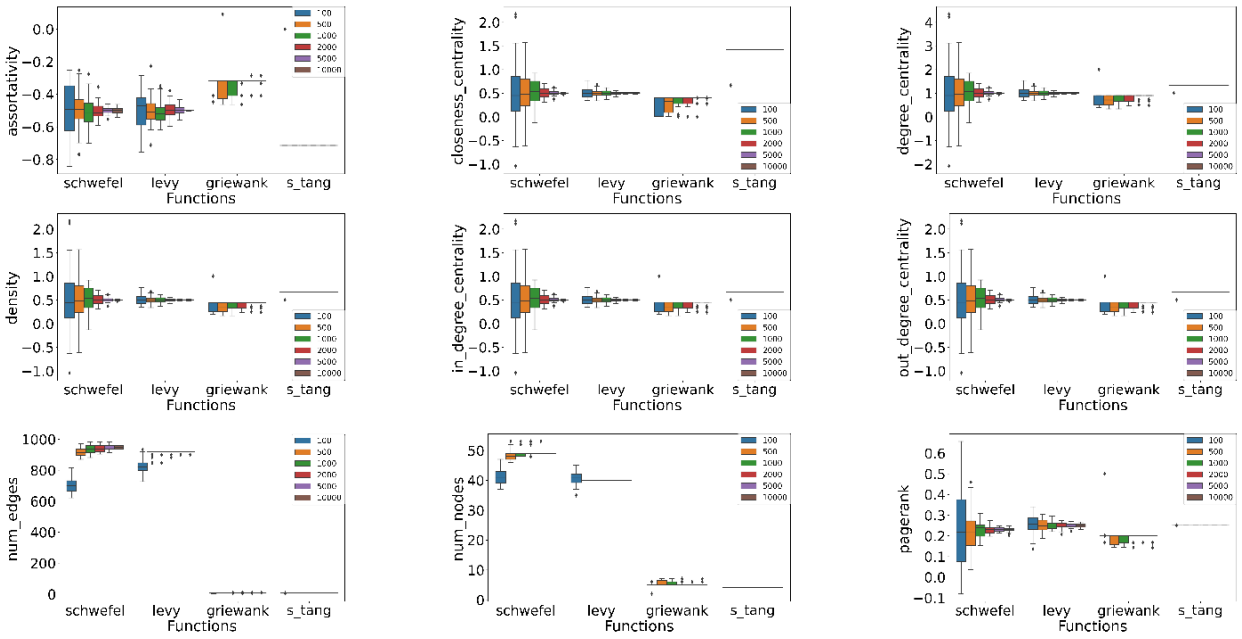

Fig. 3: Box Plots of LON Measures in 2D with Varying Sample Sizes. Each measure was obtained from 30 iterations using sample sizes ranging from 100 to 10,000.

on the number of nodes, edges, and closeness centrality—measures intrinsically tied to network structure. On the contrary, assortativity, PageRank, and degree centrality, which rely more on overall structure rather than precise network size, remained largely stable. Such variations suggest an opportunity for optimising LON creation through strategic GP sample size selection, offering efficiency implications in practical GP fit applications. Further exploration could determine optimal sample sizes for specific functions or probe the effects of diverse GPs on LONs. In summary, our work advances the understanding of GP fit and LON relationships, emphasising the importance of fitness landscape analysis. Additionally, this sheds light on how well GP fits serve as reliable surrogate models for optimisation and holds considerable potential for future research in fields applying GP fits.

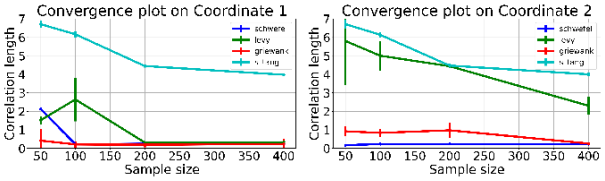

Fig. 4: Convergence Plots of correlation lengths. Since the benchmark functions are two-dimensional, each figure showcases the correlation length in a separate dimension for GP models fitted with varying sample sizes. Each measure was averaged across 30 iterations, showing the behaviour of the correlation length as the sample sizes change.

TABLE II: Error measures and their standard deviations from 30 GP fits for all benchmark functions on varying random walk sizes.

Schwefel Function					
Sample Size	RMSE	Rel MSE	Abs percentage	MAE	R^2
50	$203.020_{7.656}$	$0.859_{0.032}$	$199.256_{9.659}$	$154.662_{0.792}$	$0.262_{0.055}$
100	$162.989_{23.071}$	$0.689_{0.098}$	$177.037_{35.014}$	$124.458_{23.876}$	$0.515_{0.135}$
200	$62.303_{7.242}$	$0.263_{0.031}$	$58.790_{15.032}$	$38.548_{4.581}$	$0.930_{0.016}$
400	$34.519_{4.248}$	$0.146_{0.018}$	$20.311_{3.592}$	$16.898_{2.157}$	$0.978_{0.005}$
Levy Function					
Sample Size	RMSE	Rel MSE	Abs percentage	MAE	R^2
50	$10.893_{0.042}$	$0.684_{0.003}$	$142.843_{8.810}$	$7.825_{0.018}$	$0.532_{0.004}$
100	$3.793_{0.929}$	$0.151_{0.002}$	$46.900_{5.913}$	$3.169_{0.490}$	$0.593_{0.014}$
200	$1.093_{0.107}$	$0.069_{0.007}$	$16.641_{2.245}$	$0.760_{0.052}$	$0.995_{0.001}$
400	$1.158_{0.035}$	$0.073_{0.002}$	$11.143_{1.296}$	$0.750_{0.017}$	$0.995_{0.000}$
Griewank Function					
Sample Size	RMSE	Rel MSE	Abs percentage	MAE	R^2
50	$2.759_{0.064}$	$0.957_{0.022}$	$581.006_{212.015}$	$1.224_{0.120}$	$0.084_{0.043}$
100	$2.419_{3.232}$	$0.901_{0.203}$	$228.278_{183.349}$	$5.789_{2.321}$	$0.169_{0.217}$
200	$2.022_{0.085}$	$0.701_{0.029}$	$166.985_{67.361}$	$0.878_{0.044}$	$0.507_{0.043}$
400	$2.598_{0.396}$	$0.529_{0.137}$	$81.221_{32.020}$	$1.121_{0.200}$	$0.679_{0.153}$
STYBLINSKI-TANG Function					
Sample Size	RMSE	Rel MSE	Abs percentage	MAE	R^2
50	$0.019_{0.006}$	$0.007_{0.000}$	$0.049_{0.028}$	$0.009_{0.003}$	$0.921_{0.02}$
100	$0.009_{0.003}$	$0.001_{0.000}$	$0.031_{0.018}$	$0.005_{0.003}$	$0.965_{0.039}$
200	$0.002_{0.001}$	$0.001_{0.000}$	$0.008_{0.003}$	$0.001_{0.000}$	$1.000_{0.003}$
400	$0.001_{0.000}$	$0.000_{0.000}$	$0.002_{0.001}$	$0.000_{0.000}$	$1.000_{0.000}$

REFERENCES

[1] A. I. J. Forrester and A. J. Keane. Recent advances in surrogate-based optimization. *Progress in aerospace sciences*, 45(1-3):50–79, 2009.

[2] Z. Zhou, Y. S. Ong, M. H. Nguyen, and D. Lim. A study on polynomial regression and gaussian process global surrogate model in hierarchical surrogate-assisted evolutionary algorithm. In *2005 IEEE congress on evolutionary computation*, volume 3, pages 2832–2839. IEEE, 2005.

[3] L. Bajer, Z. Pitra, J. Repickỳ, and M. Holeňa. Gaussian process surrogate models for the cma evolution strategy. *Evolutionary computation*, 27(4):665–697, 2019.

[4] B. Liu, Q. Zhang, and G. G. E. Gielen. A gaussian process surrogate model assisted evolutionary algorithm for medium scale expensive optimization problems. *IEEE Transactions on Evolutionary Computation*, 18(2):180–192, 2013.

[5] G. Ochoa, M. Tomassini, S. Vérel, and C. Darabos. A study of nk landscapes' basins and local optima networks. In *Proc. GECCO 2008*, pages 555–562, 2008.

[6] G. Kossinets. Effects of missing data in social networks. *Social networks*, 28(3):247–268, 2006.

[7] C. E. Rasmussen, C. K. I. Williams, et al. *Gaussian processes for machine learning*, volume 1. Springer, 2006.

[8] GPy. GPy: A gaussian process framework in python. http://github.com/SheffieldML/GPy, since 2012.

[9] J. Wilson, V. Borovitskiy, A. Terenin, P. Mostowsky, and M. Deisenroth. Efficiently sampling functions from gaussian process posteriors. In *International Conference on Machine Learning*, pages 10292–10302. PMLR, 2020.

[10] G. Pleiss, J. Gardner, K. Weinberger, and A. G. Wilson. Constant-time predictive distributions for gaussian processes. In *International Conference on Machine Learning*, pages 4114–4123. PMLR, 2018.

[11] G. De Ath, R. M. Everson, and J. E. Fieldsend. How bayesian should bayesian optimisation be? In *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, pages 1860–1869, 2021.

[12] Cari Kaufman and Stephan Sain. Bayesian functional anova modeling using gaussian process prior distributions. *Bayesian Analysis*, 5, 2010.

[13] M. Tomassini, S. Vérel, and G. Ochoa. Complex network analysis of combinatorial spaces: The n-k landscape case. *Phys. Rev. E*, 78(6):066114, 2008.

[14] G. Ochoa, S. Vérel, F. Daolio, and M. Tomassini. Local optima networks: A new model of combinatorial fitness landscapes. *Recent advances in the theory and application of fitness landscapes*, pages 233–262, 2014.

[15] F. Daolio, M. Tomassini, S. Vérel, and G. Ochoa. Communities of minima in local optima networks of combinatorial spaces. *Physica A: Statistical Mechanics and its Applications*, 390(9):1684–1694, 2011.

[16] S. Vérel, F. Daolio, G. Ochoa, and M. Tomassini. Sampling local optima networks of large combinatorial search spaces: The qap case. In *Parallel Problem Solving from Nature–PPSN XV: 15th International Conference, Coimbra, Portugal, September 8–12, 2018, Proceedings, Part II 15*, pages 257–268. Springer, 2018.

[17] J. Adair, G. Ochoa, and K. M. Malan. Local optima networks for continuous fitness landscapes. In *Proc. GECCO 2019*, pages 1407–1414, 2019.

[18] M. D. Karatas, O. E. Akman, and J. E. Fieldsend. Towards population-based fitness landscape analysis using local optima networks. In *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, pages 1674–1682, 2021.

[19] D. J. Wales and J. P. K. Doye. Global optimization by basin-hopping and the lowest energy structures of lennard-jones clusters containing up to 110 atoms. *The Journal of Physical Chemistry A*, 101(28):5111–5116, 1997.

[20] J. A. Nelder and R. Mead. A simplex method for function minimization. *Comput. J.*, 7(4):308–313, 1965.

[21] G. Ochoa, K. M. Malan, and C. Blum. Search trajectory networks of population-based algorithms in continuous spaces. In *EvoApplications 2020*, pages 70–85, 2020.

[22] F. Daolio, S. Vérel, G. Ochoa, and M. Tomassini. Local optima networks and the performance of iterated local search. In *Proceedings of the 14th annual conference on Genetic and evolutionary computation*, pages 369–376, 2012.

[23] F. Chicano, F. Daolio, G. Ochoa, S. Vérel, M. Tomassini, and E. Alba. Local optima networks, landscape autocorrelation and heuristic search performance. In *Parallel Problem Solving from Nature-PPSN XII: 12th International Conference, Taormina, Italy, September 1-5, 2012, Proceedings, Part II 12*, pages 337–347. Springer, 2012.

[24] A. Liefooghe, B. Derbel, S. Verel, H. Aguirre, and K. Tanaka. Towards landscape-aware automatic algorithm configuration: preliminary experiments on neutral and rugged landscapes. In *Proc. EvoCOP 2017*, pages 215–232. Springer, 2017.

[25] K. M. Malan, J. F. Oberholzer, and A. P. Engelbrecht. Characterising constrained continuous optimisation problems. In *Proc. CEC 2015*, pages 1351–1358. IEEE, 2015.

[26] J. E. Fieldsend and K. Alyahya. Visualising the landscape of multi-objective problems using local optima networks. In *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, pages 1421–1429, 2019.

[27] X. Li, K. Tang, M. N. Omidvar, Z. Yang, K. Qin, and H. China. Benchmark functions for the cec 2013 special session and competition on large-scale global optimization. *Gene*, 7(33):8, 2013.

[28] J. Bergstra and Y. Bengio. Random search for hyper-parameter optimization. *Journal of machine learning research*, 13(2), 2012.

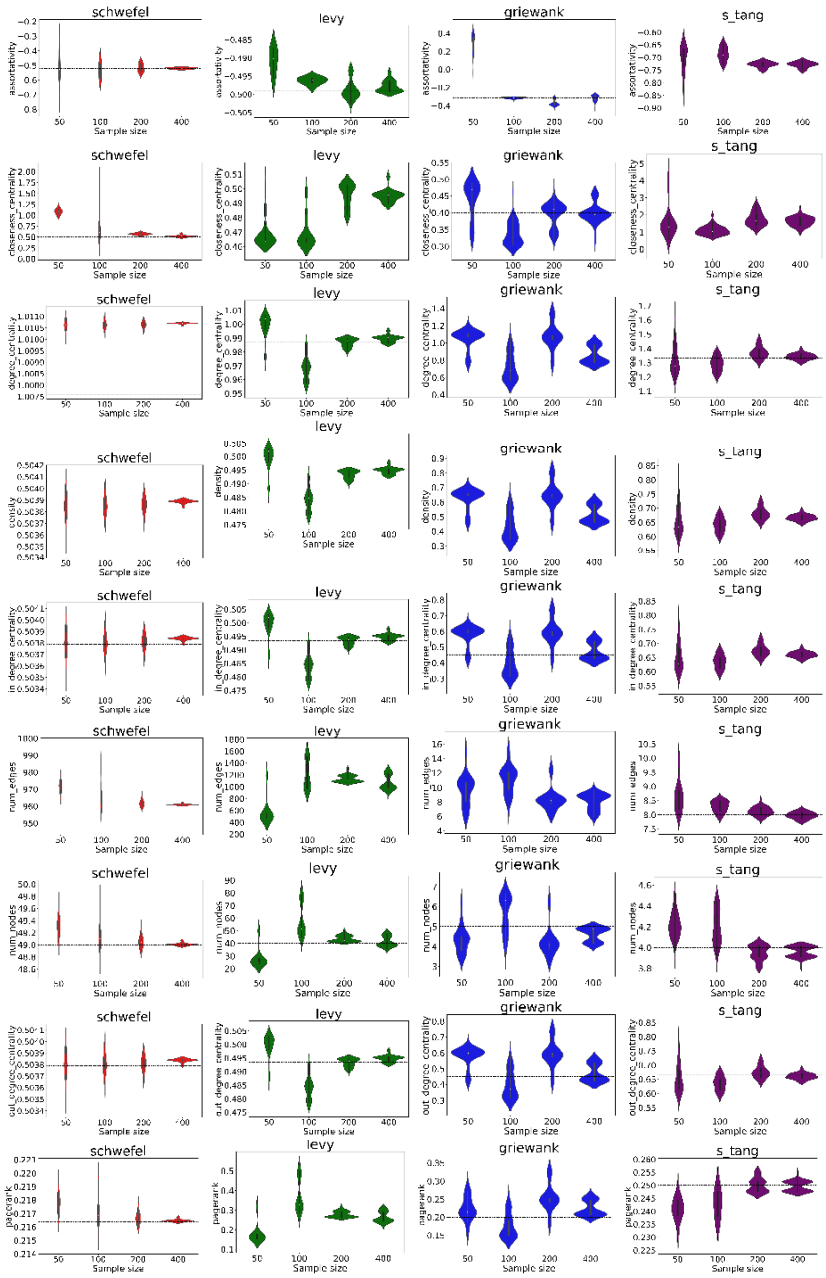

Fig. 5: Violin plots showing the distribution of network metrics for each benchmark function. Each violin represents the network metrics calculated from LONs, which were created 30 times using 30 realisations drawn from 30 different GP fits. The black dashed line indicates the true measure obtained from the true function.

[29] P. C. Wang and T. E. Shoup. Parameter sensitivity study of the nelder–mead simplex method. *Adv. Eng. Softw.*, 42(7):529–533, 2011.