Landscape Analysis for Surrogate Models in the Evolutionary Black-Box Context (Supplementary Material)

Zbyněk Pitra z.pitra@gmail.com
Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University
Břehová 7, 115 19 Prague, Czech Republic

Jan Koza koza@cs.cas.cz
Faculty of Information Technology, CTU in Prague,
Thákurova 9, 160 00 Prague 6, Czech Republic

Jiří Tumpach tumpach@cs.cas.cz
Faculty of Mathematics and Physics, Charles University in Prague,
Malostran. nám. 25, 118 00 Prague, Czech Republic

Martin Holeňa martin@cs.cas.cz
Institute of Computer Science, Czech Academy of Sciences,
Pod Vodárenskou věží 2, 182 07 Prague, Czech Republic

Abstract

Surrogate modeling has become a valuable technique for black-box optimization tasks with expensive evaluation of the objective function. In this supplementary material for the original article *Landscape Analysis for Surrogate Models in the Evolutionary Black-Box Context* (Pitra et al., 2022), we report additional information for investigation of the relationship between the predictive accuracy of surrogate models and features of the black-box function landscape. We also report additional information on study of properties of features for landscape analysis in the context of different transformations and ways of selecting the input data.

Keywords

Adaptive learning, Metalearning, Black-box optimization, Surrogate model, Landscape analysis

1 Introduction

In (Pitra et al., 2022) we have presented an investigation of relationships between the prediction error of the surrogate models, their settings, and features describing the fitness landscape during evolutionary optimization by the Covariance Matrix Adaptation Evolution Strategy (CMA-ES) (Hansen and Ostermeier, 1996; Hansen, 2006) state-of-the-art optimizer for expensive continuous black-box tasks. Its pseudocode is outlined in Algorithm 1. Four models in 39 different settings for the Doubly Trained Surrogate CMA-ES (DTS-CMA-ES) by Pitra et al. (2016), a surrogate-assisted version of the CMA-ES, were compared using three sets of 14 landscape features selected according to their properties, especially according to their robustness and similarity to other features. We analysed MSE and RDE dependence of various models and model settings

Algorithm 1 Pseudocode of the CMA-ES

Input: λ (population-size), $\mathbb{f}$ (original fitness function), $(w_i)_{i=1}^{\mu}$, μ_w (selection and recombination parameters), c_σ, d_σ (step-size control parameters), c_c, c_I, c_μ (covariance matrix adaptation parameters)

init $g = 0, \sigma^{(0)} > 0, \mathbf{m}^{(0)} \in \mathbb{R}^D, \mathbf{C}^{(0)} = \mathbf{I}, \mathbf{p}_\sigma^{(0)} = \mathbf{0}, \mathbf{p}_c^{(0)} = \mathbf{0}, \mu = \lfloor \lambda/2 \rfloor$
repeat
  $\mathbf{x}_k \leftarrow \mathbf{m}^{(g)} + \sigma^{(g)} \mathcal{N}(\mathbf{0}, \mathbf{C}^{(g)}) \qquad k = 1, \ldots, \lambda$
  $\mathbb{f}_k \leftarrow \mathbb{f}(\mathbf{x}_k) \qquad k = 1, \ldots, \lambda$
  $\mathbf{m}^{(g+1)} \leftarrow \sum_{i=1}^{\mu} w_i \mathbf{x}_{i:\lambda}$, where $\mathbf{x}_{i:\lambda}$ is the i-th best individual out of $\mathbf{x}_1, \ldots, \mathbf{x}_\lambda$
  $\mathbf{p}_\sigma^{(g+1)} \leftarrow (1 - c_\sigma)\mathbf{p}_\sigma^{(g)} + \sqrt{c_\sigma(2 - c_\sigma)\mu_w}\, {\mathbf{C}^{(g)}}^{-1/2} \frac{\mathbf{m}^{(g+1)} - \mathbf{m}^{(g)}}{\sigma^{(g)}}$
  $h_\sigma \leftarrow \mathbb{I}\left(\|\mathbf{p}_\sigma^{(g+1)}\| < \sqrt{1 - (1 - c_\sigma)^{2(g+1)}} \left(1.4 + \frac{2}{D+1}\right) E\|\mathcal{N}(0, 1))\| \right)$
  $\mathbf{p}_c^{(g+1)} \leftarrow (1 - c_c)\mathbf{p}_c^{(g)} + h_\sigma \sqrt{c_c(2 - c_c)\mu_w}\, \frac{\mathbf{m}^{(g+1)} - \mathbf{m}^{(g)}}{\sigma^{(g)}}$
  $\mathbf{C}_\mu \leftarrow \sum_{i=1}^{\mu} w_i {\sigma^{(g)}}^{-2} \left(\mathbf{x}_{i:\lambda}^{(g+1)} - \mathbf{m}^{(g)}\right) \left(\mathbf{x}_{i:\lambda}^{(g+1)} - \mathbf{m}^{(g)}\right)^\top$
  $\mathbf{C}^{(g+1)} \leftarrow (1 - c_1 - c_\mu)\mathbf{C}^{(g)} + c_1 \mathbf{p}_c^{(g+1)} {\mathbf{p}_c^{(g+1)}}^\top + c_\mu \mathbf{C}_\mu$
  $\sigma^{(g+1)} \leftarrow \sigma^{(g)} \exp\left(\frac{c_\sigma}{d_\sigma} \left(\frac{\|\mathbf{p}_\sigma^{(g+1)}\|}{E\|\mathcal{N}(\mathbf{0}, \mathbf{I})\|} - 1 \right) \right)$
  $g \leftarrow g + 1$
until stopping criterion is met

Output: $\mathbf{x}_{\text{res}}$ (resulting optimum)

on the features calculated using three methods selecting sample sets extracted from DTS-CMA-ES runs on noiseless benchmarks from the Comparing Continuous Optimisers (COCO) benchmark function testbed (Hansen et al., 2016). Up to our knowledge, this analysis of landscape feature properties and their connection to surrogate models in evolutionary optimization is more profound than any other published so far.

The employed training set selection (TSS) methods were:

- ◇ *TSS full*, taking all the already evaluated points, i. e., $\mathcal{T} = \mathcal{A}$,
- ◇ *TSS knn*, selecting the union of the sets of k-nearest neighbors of all points for which the fitness should be predicted, where k is user defined (e. g., in Kern et al. (2006)),
- ◇ *TSS nearest*, selecting the union of the sets of k-nearest neighbors of all points for which the fitness should be predicted, where k is maximal such that the total number of selected points does not exceed a given maximum number of points N_{max} and no point is further from current mean $\mathbf{m}^{(g)}$ than a given maximal distance r_{max} (e. g., in Bajer et al. (2019)).

We have performed all the feature investigations for each TSS method separately. Employed extracted *basic sample sets* were:

- ◇ *archive* $\mathcal{A}$ containing all points evaluated so far by the original fitness;
- ◇ *training set* $\mathcal{T}$ selected from $\mathcal{A}$ used for model training.

By combining the two basic sample sets with a population $\mathcal{P}$ consisting of the points without a known value of the original fitness to be evaluated by the surrogate model, we have obtained two new sets $\mathcal{A}_\mathcal{P} = \mathcal{A} \cup \mathcal{P}$ and $\mathcal{T}_\mathcal{P} = \mathcal{T} \cup \mathcal{P}$. We have utilized either transformed and non-transformed sets for feature calculations, resulting in 8 different sample sets (4 in case of TSS full due to $\mathcal{T} = \mathcal{A}$): $\mathcal{A}$, $\mathcal{A}^\top$, $\mathcal{A}_\mathcal{P}$, $\mathcal{A}_\mathcal{P}^\top$, $\mathcal{T}$, $\mathcal{T}^\top$, $\mathcal{T}_\mathcal{P}$, $\mathcal{T}_\mathcal{P}^\top$, where $^\top$ denotes transformation into the $\sigma^2\mathbf{C}$ basis.

The supplementary material is structured as follows. Section 2 describes landscape features utilized in (Pitra et al., 2022) in more detail. Sections 3 and 4 supply the testing and analysis, respectively, in the original paper by more detailed results.

2 Feature Definitions

Let us consider an input set $\mathbf{X}$ of N points

$$\mathbf{X} = \left\{\mathbf{x}_i \middle| \, \mathbf{x}_i \in \bigtimes_{j=1}^{D} [l_j, u_j], l_j, u_j \in \mathbb{R}, l_j < u_j \right\}_{i=1}^{N} .$$

Then we can define a *sample set* $\mathcal{S}$ of N pairs of observations

$$\mathcal{S} = \{(\mathbf{x}_i, y_i) \,|\, \mathbf{x}_i \in \mathbf{X}, y_i \in \mathbb{R} \cup \{\circ\}, i = 1, \ldots, N\} \ ,$$

where $\circ$ denotes missing y_i value (e. g., $\mathbf{x}_i$ was not evaluated yet). Then the sample set can be utilized to describe landscape properties using a *landscape feature* φ

$$\varphi : \bigcup_{N \in \mathbb{N}} \mathbb{R}^{N,D} \times (\mathbb{R} \cup \{\circ\})^{N,1} \mapsto \mathbb{R} \cup \{\pm\infty, \bullet\} \,,$$

where $\bullet$ denotes impossibility of feature computation. The y_i values can be gathered to a set $\mathbf{y} = \{y_i\}_{i=1}^{N}$.

The following features were reimplemented in Matlab according to the R-package `flacco` by Kerschke and Dagefoerde (2017). In the words of Kerschke (2017a), *"It is important to notice that, independent of the research domain, most of these features do not provide intuitively understandable numbers. Therefore, we strongly recommend not to interpret them on their own. Moreover, some of them are stochastic and hence should be evaluated multiple times on an instance and afterwards be aggregated in a reasonable manner. Nevertheless, they definitely provide information that can be of great importance to scientific models, such as machine learning algorithms in general or algorithm selectors in particular."* Following the remark about stochasticity, we have fixed random seeds which are not fixed in the original implementation to always return the same values for identical input.

2.1 Basic Features $\Phi_{\textbf{Basic}}$

Kerschke (2017b) summarized the following features providing obvious information about the input space:

- **Dimension** of the input space $\varphi_{\texttt{dim}} = D$.
- **Number of observations** $\varphi_{\texttt{obs}} = N$.
- **Minimum** and **maximum of lower bounds** $\varphi_{\texttt{lower\_min}} = \min_{i \in D} l_i$, $\varphi_{\texttt{lower\_max}} = \max_{i \in D} l_i$.
- **Minimum** and **maximum of upper bounds** $\varphi_{\texttt{upper\_min}} = \min_{i \in D} u_i$, $\varphi_{\texttt{upper\_max}} = \max_{i \in D} u_i$.
- **Minimum** and **maximum of y values** $\varphi_{\texttt{objective\_min}} = \min_{i \in N} y_i$, $\varphi_{\texttt{objective\_max}} = \max_{i \in N} y_i$.
- **Minimum** and **maximum of cell blocks per dimension** $\varphi_{\texttt{blocks\_min}}$, $\varphi_{\texttt{blocks\_max}}$.
- **Total number of cells** $\varphi_{\texttt{cells\_total}}$.
- **Number of filled cells** $\varphi_{\texttt{cells\_filled}}$.
- Binary flag stating whether the **objective function** should be **minimized** $\varphi_{\texttt{minimize\_fun}}$.

In this research, we have used only $\varphi_{\texttt{dim}}$ and $\varphi_{\texttt{obs}}$ because the remaining ones were constant on the whole $\mathcal{D}_{100}$ dataset. $\varphi_{\texttt{dim}}$ is identical regardless the sample set and $\varphi_{\texttt{obs}}$ was not computed in the $\sigma^2\mathbf{C}$ basis because it does not influence the resulting feature values.

2.2 CMA Features $\Phi_{\mathbf{CMA}}$

In each CMA-ES generation g during the fitness evaluation step (Algorithm 1, step 4), we have defined the following features in Pitra et al. (2019) for the set of points $\mathbf{X}$:

- **Generation number** $\varphi^{\text{CMA}}_{\texttt{generation}} = g$ indicates the phase of the optimization process.

- **Step-size** $\varphi^{\text{CMA}}_{\texttt{step\_size}} = \sigma^{(g)}$ provides an information about the extent of the approximated region.

- **Number of restarts** $\varphi^{\text{CMA}}_{\texttt{restart}} = n_{\text{r}}^{(g)}$ performed till generation g may indicate landscape difficulty.

- **Mahalanobis mean distance** of the CMA-ES mean $\mathbf{m}^{(g)}$ to the sample mean $\mu_{\mathbf{X}}$ of $\mathbf{X}$
$$\varphi^{\text{CMA}}_{\texttt{mean\_dist}}(\mathbf{X}) = \sqrt{(\mathbf{m}^{(g)} - \mu_{\mathbf{X}})^{\top} \mathbf{C}_{\mathbf{X}}^{-1} (\mathbf{m}^{(g)} - \mu_{\mathbf{X}})}\,, \tag{1}$$
where $\mathbf{C}_{\mathbf{X}}$ is the sample covariance of $\mathbf{X}$. This feature indicates suitability of $\mathbf{X}$ for model training from the point of view of the current state of the CMA-ES algorithm.

- Square of the $\mathbf{p}_c$ **evolution path** length $\varphi^{\text{CMA}}_{\texttt{evopath\_c\_norm}} = \left\|\mathbf{p}_c^{(g)}\right\|^2$ is the only possible non-zero eigenvalue of *rank-one update* covariance matrix $\mathbf{p}_c^{(g+1)}\mathbf{p}_c^{(g+1)\top}$ (see step 10 in Algorithm 1). That feature providing information about the correlations between consecutive CMA-ES steps indicates a similarity of function landscapes among subsequent generations.

- $\mathbf{p}_\sigma$ **evolution path** ratio, i. e., the ratio between the evolution path length $\left\|\mathbf{p}_\sigma^{(g)}\right\|$ and the expected length of a random evolution path used to update step-size. It provides a useful information about distribution changes:
$$\varphi^{\text{CMA}}_{\texttt{evopath\_s\_norm}} = \frac{\left\|\mathbf{p}_\sigma^{(g)}\right\|}{E\left\|\mathcal{N}(\mathbf{0},\mathbf{I})\right\|} = \frac{\left\|\mathbf{p}_\sigma^{(g)}\right\| \Gamma\left(\frac{D}{2}\right)}{\sqrt{2}\,\Gamma\left(\frac{D+1}{2}\right)}\,. \tag{2}$$

- **CMA similarity likelihood**. The log-likelihood of the set of points $\mathbf{X}$ with respect to the CMA-ES distribution may also serve as a measure of its suitability for training
$$\varphi^{\text{CMA}}_{\texttt{cma\_lik}}(\mathbf{X}) = -\frac{N}{2}\left(D \log 2\pi{\sigma^{(g)}}^2 + \log\det \mathbf{C}^{(g)}\right) - \frac{1}{2}\sum_{\mathbf{x}\in\mathbf{X}} \left(\frac{\mathbf{x}-\mathbf{m}^{(g)}}{\sigma^{(g)}}\right)^{\top} {\mathbf{C}^{(g)}}^{-1} \left(\frac{\mathbf{x}-\mathbf{m}^{(g)}}{\sigma^{(g)}}\right). \tag{3}$$

2.3 Dispersion Features Φ_{Dis}

The dispersion features by Lunacek and Whitley (2006) compare the dispersion among observations from $\mathcal{S}$ and among a subset of these points. The subsets are created based on tresholds using the `quantile`s 0.02, 0.05, 0.1, and 0.25 of the objective values $\mathbf{y}$.

The set of all distances within the set $\mathcal{S}$:

$$D_{\text{all}} = \{\text{dist}(\mathbf{x}_i, \mathbf{x}_j) \,|\, i, j = 1, \ldots, N\} . \tag{4}$$

The quantile subset of all distances:

$$D_{\texttt{quantile}} = \{\text{dist}(\mathbf{x}_i, \mathbf{x}_j) \,|\, y_i, y_j <= Q_{\texttt{quantile}}(\mathbf{y}), \} . \tag{5}$$

◇ **Ratio of** the quantile subset and all points **median distances**

$$\varphi^{\text{Dis}}_{\texttt{ratio\_median\_[quantile]}}(\mathcal{S}) = \frac{\text{median}\, D_{\texttt{quantile}}}{\text{median}\, D_{\text{all}}} . \tag{6}$$

◇ **Ratio of** the quantile subset and all points **mean distances**

$$\varphi^{\text{Dis}}_{\texttt{ratio\_mean\_[quantile]}}(\mathcal{S}) = \frac{\text{mean}\, D_{\texttt{quantile}}}{\text{mean}\, D_{\text{all}}} . \tag{7}$$

◇ **Difference between** the quantile subset and all points **median distances**

$$\varphi^{\text{Dis}}_{\texttt{diff\_median\_[quantile]}}(\mathcal{S}) = \text{median}\, D_{\texttt{quantile}} - \text{median}\, D_{\text{all}} . \tag{8}$$

◇ **Difference between** the quantile subset and all points **mean distances**

$$\varphi^{\text{Dis}}_{\texttt{diff\_mean\_[quantile]}}(\mathcal{S}) = \text{mean}\, D_{\texttt{quantile}} - \text{mean}\, D_{\text{all}} . \tag{9}$$

2.4 Information Content Features Φ_{Inf}

The Information Content of Fitness Sequences by Muñoz et al. (2015) approach is based on a symbol sequence $\Psi = \{\psi_1, \ldots, \psi_{N-1}\}$, where

$$\psi_i = \begin{cases} \bar{1}\,, & \text{if} \quad \frac{y_{i+1}-y_i}{\|\mathbf{x}_{i+1}-\mathbf{x}_i\|} \quad < \quad -\varepsilon\,, \\ 0\,, & \text{if} \quad \left|\frac{y_{i+1}-y_i}{\|\mathbf{x}_{i+1}-\mathbf{x}_i\|}\right| \quad \leq \quad \varepsilon\,, \\ 1\,, & \text{if} \quad \frac{y_{i+1}-y_i}{\|\mathbf{x}_{i+1}-\mathbf{x}_i\|} \quad > \quad \varepsilon\,. \end{cases} \tag{10}$$

The sequence considers observations from a sample set $\mathcal{S}$ as a random walk accross the objective landscape and depends on the information sensitivity parameter $\varepsilon > 0$.

The symbol sequence Ψ is aggregated by the information content $H(\varepsilon) = -\sum_{i \neq j} p_{ij} \log_6 p_{ij}$, where p_{ij} is the probability of having the block $\psi_i\psi_j$, where $\psi_i, \psi_j \in \{\bar{1}, 0, 1\}$, within the sequence. Note that the base of the logarithm was set to six as this equals the number of possible blocks $\psi_i\psi_j$ for which $\psi_i \neq \psi_j$, i. e., $\psi_i\psi_j \in \{\bar{1}0, 0\bar{1}, \bar{1}1, 1\bar{1}, 01, 10\}$.

Another aggregation of the information is the so-called partial information content $M(\varepsilon) = |\Psi'|/(N-1)$, where Ψ' is the symbol sequence of alternating $\bar{1}$s and 1s, which is derived from Ψ by removing all 0 and repeated symbols.

When we do consider $\circ$ as a valid state of y, the sequence Ψ consists of $\psi_i \in \{\bar{1}, 0, 1, \bar{N}\}$, where $\psi_i = \bar{N}$, if $y_{i+1} = \circ$ or $y_i = \circ$. Thus, $H(\varepsilon) = -\sum_{i \neq j} p_{ij} \log_{12} p_{ij}$ due to the increased number of possible $\psi_i\psi_j$ blocks, i. e., $\psi_i\psi_j \in \{\bar{1}0, 0\bar{1}, \bar{1}1, 1\bar{1}, 01, 10, \bar{N}\bar{1}, \bar{1}\bar{N}, \bar{N}0, 0\bar{N}, \bar{N}1, 1\bar{N}\}$. Therefore, features based on Ψ' can be utilized only when $\circ$ state is not present in $\mathbf{y}$.

Based on sequences Ψ and Ψ' the following features can be defined according to Muñoz et al. (2015):

- ◇ **Maximum information content** $\varphi^{\text{Inf}}_{\texttt{h\_max}}(\mathcal{S}) = \max_\varepsilon H(\varepsilon)$.
- ◇ **Settling sensitivity** $\varphi^{\text{Inf}}_{\texttt{eps\_s}}(\mathcal{S}) = \log_{10} \min(\varepsilon | H(\varepsilon) < s)$, where default $s = 0.05$ (see Muñoz et al. (2015)).
- ◇ **Maximum sensitivity** $\varphi^{\text{Inf}}_{\texttt{eps\_max}}(\mathcal{S}) = \arg\max_\varepsilon H(\varepsilon)$.
- ◇ **Initial partial information** $\varphi^{\text{Inf}}_{\texttt{m0}}(\mathcal{S}) = M(\varepsilon = 0)$.
- ◇ **Ratio of partial information sensitivity** $\varphi^{\text{Inf}}_{\texttt{eps\_ratio}}(\mathcal{S}) = \log_{10} \max(\varepsilon | M(\varepsilon) > r\varphi^{\text{Inf}}_{\texttt{m0}}(\mathcal{S})))$, where default $r = 0.5$ (see also Muñoz et al. (2015)).

2.5 Levelset Features $\Phi_{\mathbf{Lvl}}$

In Levelset features by Mersmann et al. (2011), the sample set $\mathcal{S}$ is split into two classes by a specific treshold calculated using the quantiles 0.1, 0.25, and 0.5. Linear, quadratic, and mixture discriminant analysis (`lda`, `qda`, and `mda`) are used to predict whether the objective values $\mathbf{y}$ fall below or exceed the calculated threshold. The extracted features are based on the distribution of the resulting cross-validated mean misclassification errors of each classifier.

- ◇ **Mean misclassification error** of appropriate discriminant analysis `method` using defined `quantile` $\varphi^{\text{Lvl}}_{\texttt{mmce\_[method]\_[quantile]}}(\mathcal{S})$.
- ◇ **Ratio between** mean misclassification **errors** of two discriminant analysis `method`s using a given `quantile` $\varphi^{\text{Lvl}}_{\texttt{[method1]\_[method2]\_[quantile]}}(\mathcal{S})$.

2.6 Metamodel Features Φ_{MM}

To calculate metamodel features by Mersmann et al. (2011), linear and quadratic regression models (`lin` and `quad`) with or without interactions are fitted to a sample set $\mathcal{S}$. The adjusted coefficient of determination R^2 and features reflecting the size relations of the model coefficients are extracted:

- **Adjusted** R^2 of a **simple model** $\varphi^{\mathrm{MM}}_{\texttt{[model]\_simple\_adj\_r2}}(\mathcal{S})$.
- **Adjusted** R^2 of a **model with interactions** $\varphi^{\mathrm{MM}}_{\texttt{[model]\_w\_interact\_adj\_r2}}(\mathcal{S})$.
- **Intercept** of a simple **linear model** $\varphi^{\mathrm{MM}}_{\texttt{lin\_simple\_intercept}}(\mathcal{S})$.
- **Minimal** absolute **value of linear model coefficients** $\varphi^{\mathrm{MM}}_{\texttt{lin\_simple\_coef\_min}}(\mathcal{S})$.
- **Maximal** absolute **value of linear model coefficients** $\varphi^{\mathrm{MM}}_{\texttt{lin\_simple\_coef\_max}}(\mathcal{S})$.
- **Ratio** of **maximal** and **minimal** abs. **value of linear model coefficients** $\varphi^{\mathrm{MM}}_{\texttt{lin\_simple\_coef\_max\_by\_min}}(\mathcal{S})$.
- **Ratio** of **maximal** and **minimal** abs. **value of quadratic model coefficients** $\varphi^{\mathrm{MM}}_{\texttt{quad\_simple\_cond}}(\mathcal{S})$.

The feature $\varphi^{\mathrm{MM}}_{\texttt{lin\_simple\_intercept}}$ was excluded because it is useless if fitness normalization is performed. The remaining features were not computed on sample sets with $\mathcal{P}$ because it does not influence the resulting feature values.

2.7 Nearest Better Clustering Features Φ_{NBC}

Nearest better clustering features by Kerschke et al. (2015) extract information based on the comparison of the sets of distances from all observations towards their nearest neighbors (D_{nn}) and their nearest better neighbors (D_{nb}).

The distance to the nearest neighbor of a search point $\mathbf{x}_i$, $i = 1, \ldots, N$ from a set of points $\mathbf{X}$:

$$d_{\text{nn}}(\mathbf{x}_i, \mathbf{X}) = \min(\operatorname{dist}(\mathbf{x}_i, \mathbf{x}_j) | \mathbf{x}_j \in \mathbf{X}, i \neq j)\,. \tag{11}$$

The distance to the nearest better neighbor:

$$d_{\text{nb}}(\mathbf{x}_i, \mathbf{X}) = \min(\operatorname{dist}(\mathbf{x}_i, \mathbf{x}_j) | y_j < y_i, \mathbf{x}_j \in \mathbf{X}))\,. \tag{12}$$

The set of all nearest neighbor distances within the set $\mathbf{X}$:

$$D_{\text{nn}} = \{d_{\text{nn}}(\mathbf{x}_i, \mathbf{X}) | i = 1, \ldots, N\}\,. \tag{13}$$

The set of all nearest better distances:

$$D_{\text{nb}} = \{d_{\text{nb}}(\mathbf{x}_i, \mathbf{X}) | i = 1, \ldots, N\}\,. \tag{14}$$

The features are defined as follows:

- **Ratio of the standard deviations** between the D_{nn} and D_{nb} $\varphi^{\text{NBC}}_{\texttt{nb\_std\_ratio}}(\mathcal{S}) = \frac{\operatorname{std} D_{\text{nn}}}{\operatorname{std} D_{\text{nb}}}$.
- **Ratio of the means** between the D_{nn} and D_{nb} $\varphi^{\text{NBC}}_{\texttt{nb\_mean\_ratio}}(\mathcal{S}) = \frac{\operatorname{mean} D_{\text{nn}}}{\operatorname{mean} D_{\text{nb}}}$.
- **Correlation between the distances** of the nearest neighbors and nearest better neighbors $\varphi^{\text{NBC}}_{\texttt{nb\_cor}}(\mathcal{S}) = \operatorname{corr}(D_{\text{nn}}, D_{\text{nb}})$.
- **Coefficient of** variation of the **distance ratios** $\varphi^{\text{NBC}}_{\texttt{dist\_ratio}}(\mathcal{S}) = \frac{\operatorname{std} \mathcal{W}}{\operatorname{mean} \mathcal{W}}$, where $\mathcal{W} = \left\{ \frac{d_{\text{nn}}(\mathbf{x}, \mathbf{X})}{d_{\text{nb}}(\mathbf{x}, \mathbf{X})} \middle| \mathbf{x} \in \mathbf{X} \right\}$.
- **Correlation between the fitness value, and the count of observations** to whom the current observation is the nearest better neighbor $\varphi^{\text{NBC}}_{\texttt{nb\_fitness\_cor}}(\mathcal{S}) = -\operatorname{corr}(\{\deg^-(\mathbf{x}_i), y_i | i = 1, \ldots, N\})$, where $\deg^-(\mathbf{x}_i)$ is the indegree of $\mathbf{x}_i$ in the nearest better graph (the number of points for which a certain point is the nearest better point).

The features from Φ_{NBC} were not computed on sample sets with $\mathcal{P}$ because it does not influence the resulting feature values.

2.8 $\mathbf{y}$-Distribution Features $\Phi_{\mathbf{y}\text{-D}}$

The $\mathbf{y}$-distribution features from Mersmann et al. (2011) compute the basic statistics of the fitness values $\mathbf{y}$.

- **Skewness** and **kurtosis** of $\mathbf{y}$ values $\varphi^{\text{y-D}}_{\texttt{skewness}}(\mathbf{y})$ and $\varphi^{\text{y-D}}_{\texttt{kurtosis}}(\mathbf{y})$.
- **Number of peaks** of the kernel-based estimation of the density of $\mathbf{y}$-distribution $\varphi^{\text{y-D}}_{\texttt{number\_of\_peaks}}(\mathbf{y})$, where the peak is defined according to Kerschke and Dagefoerde (2017).

All three features were computed only on $\mathcal{A}$ and $\mathcal{T}$ because neither $\mathcal{P}$ nor the $\sigma^2\mathbf{C}$ basis influence the resulting feature values.

3 Landscape Feature Investigation Results

First, we have investigated the impossibility of feature calculation (i. e., the feature value $\bullet$) for each feature. Considering that large amount of $\bullet$ values on the tested dataset suggests low usability of the respective feature, we have excluded features resulting in $\bullet$ in more than 25% of all measured values. Many features are difficult to calculate using low numbers of points. Therefore, for each feature we have measured the minimal number of points $N_\bullet$ in a particular combination of feature and sample set, for which the calculation resulted in $\bullet$ in at most 1% of cases. All measured values can be found in Tables 1–6.

We have tested the dependency of individual features on the dimension using feature medians from 100 samples for each distribution from the dataset. The Friedman's test rejected the hypothesis that the feature medians are independent of the dimension for all features at the family-wise significance level 0.05 using the Bonferroni-Holm correction. Moreover, for most of the features, the subsequently performed pairwise tests rejected the hypothesis of equality of feature medians for all pairs od dimensions. There were only several features for which the hypothesis was not rejected for some pairs of dimensions (see Tables 1–6).

Any analysis of the influence of multiple landscape features on the predictive error of surrogate models requires high robustness of features against random sampling of points. Therefore, we define feature *robustness* as a proportion of cases for which the difference between the 1st and 100th percentile calculated after standardization on samples from the same CMA-ES distribution is ≤ 0.05. The robustness calculated for individual features is listed in Tables 1–6.

Table 1: TSS full features ($\mathcal{A} = \mathcal{T}$ for TSS full) from feature sets Φ_{Basic}, Φ_{CMA}, Φ_{Dis}, and $\Phi_{\text{y-D}}$. Features are grouped according to their feature sets (separated by horizontal lines). Features with less than 25% of values equal to $\bullet$ and robustness greater than 0.9, are in gray. $N_\bullet$ denotes the lowest measured number of points from which at most 1% of feature calculations resulted in $\bullet$ ($N_\bullet = 0$ for sample set independent φ). The (D_i, D_j) column shows the pairs of feature dimensions for which the two-sided Wilcoxon signed rank test with the Bonferroni-Holm correction does not reject the hypothesis of equality of median feature values, at the family-wise level 0.05 for each individual feature.

	$\mathcal{A}$			$\mathcal{A}^\top$			$\mathcal{A}_\mathcal{P}$			$\mathcal{A}_\mathcal{P}^\top$		
	$N_\bullet$	rob.(%)	(D_i, D_j)	$N_\bullet$	rob.(%)	(D_i, D_j)	$N_\bullet$	rob.(%)	(D_i, D_j)	$N_\bullet$	rob.(%)	(D_i, D_j)
$\varphi_{\texttt{dim}}$	0	100.00		0	100.00		0	100.00		0	100.00	
$\varphi_{\texttt{obs}}$	1	100.00		1	100.00		13	100.00		13	100.00	
$\varphi^{\text{CMA}}_{\texttt{generation}}$	0	100.00		0	100.00		0	100.00		0	100.00	
$\varphi^{\text{CMA}}_{\texttt{step\_size}}$	0	100.00		0	100.00		0	100.00		0	100.00	
$\varphi^{\text{CMA}}_{\texttt{restart}}$	0	100.00		0	100.00		0	100.00		0	100.00	
$\varphi^{\text{CMA}}_{\texttt{cma\_mean\_dist}}$	6	70.11		6	56.83		13	63.61		13	54.12	
$\varphi^{\text{CMA}}_{\texttt{evopath\_c\_norm}}$	0	100.00		0	100.00		0	100.00		0	100.00	
$\varphi^{\text{CMA}}_{\texttt{evopath\_s\_norm}}$	0	100.00		0	100.00		0	100.00		0	100.00	
$\varphi^{\text{CMA}}_{\texttt{cma\_lik}}$	1	99.75		1	99.96		13	99.75		13	99.96	
$\varphi^{\text{Dis}}_{\texttt{ratio\_mean\_02}}$	71	57.69	$(2,5)$, $(3,10)$, $(3,20)$	71	62.33	$(2,5)$, $(5,20)$	86	57.36	$(2,5)$, $(3,20)$	86	62.16	
$\varphi^{\text{Dis}}_{\texttt{ratio\_median\_02}}$	71	65.95	$(3,5)$	71	69.28		86	65.41		86	69.60	
$\varphi^{\text{Dis}}_{\texttt{diff\_mean\_02}}$	71	47.75		71	99.52		86	48.77		86	99.52	
$\varphi^{\text{Dis}}_{\texttt{diff\_median\_02}}$	71	50.21		71	99.44		86	51.44		86	99.45	
$\varphi^{\text{Dis}}_{\texttt{ratio\_mean\_05}}$	27	55.22	$(5,20)$	27	60.59		42	54.76	$(5,20)$	42	60.32	
$\varphi^{\text{Dis}}_{\texttt{ratio\_median\_05}}$	27	59.80		27	64.43		42	59.66	$(5,20)$	42	64.57	
$\varphi^{\text{Dis}}_{\texttt{diff\_mean\_05}}$	27	47.99		27	99.49		42	49.16		42	99.49	
$\varphi^{\text{Dis}}_{\texttt{diff\_median\_05}}$	27	51.16		27	99.46		42	52.88		42	99.48	
$\varphi^{\text{Dis}}_{\texttt{ratio\_mean\_10}}$	12	54.63	$(5,20)$	12	59.33		25	54.06	$(5,20)$	25	58.71	
$\varphi^{\text{Dis}}_{\texttt{ratio\_median\_10}}$	12	58.29		12	62.38	$(2,3)$	25	57.80		25	61.82	
$\varphi^{\text{Dis}}_{\texttt{diff\_mean\_10}}$	12	49.70		12	99.52		25	50.88		25	99.52	
$\varphi^{\text{Dis}}_{\texttt{diff\_median\_10}}$	12	53.51		12	99.48		25	55.53		25	99.49	
$\varphi^{\text{Dis}}_{\texttt{ratio\_mean\_25}}$	6	54.80		6	58.86		15	53.87		15	58.06	
$\varphi^{\text{Dis}}_{\texttt{ratio\_median\_25}}$	6	56.21		6	59.74		15	54.91		15	58.26	
$\varphi^{\text{Dis}}_{\texttt{diff\_mean\_25}}$	6	53.78		6	99.55		15	55.29		15	99.55	
$\varphi^{\text{Dis}}_{\texttt{diff\_median\_25}}$	6	57.40		6	99.49		15	60.24		15	99.50	
$\varphi^{\text{y-D}}_{\texttt{skewness}}$	6	36.71		–	–	–	–	–	–	–	–	–
$\varphi^{\text{y-D}}_{\texttt{kurtosis}}$	6	63.06		–	–	–	–	–	–	–	–	–
$\varphi^{\text{y-D}}_{\texttt{number\_of\_peaks}}$	6	32.34		–	–	–	–	–	–	–	–	–

Table 2: TSS full features ($\mathcal{A} = \mathcal{T}$ for TSS full) from feature sets Φ_{Lvl}, Φ_{MM}, Φ_{Inf}, and Φ_{NBC}. Features are grouped according to their feature sets (separated by horizontal lines). Features with less than 25% of values equal to $\bullet$ and robustness greater than 0.9, are in gray. $N_\bullet$ denotes the lowest measured number of points from which at most 1% of feature calculations resulted in $\bullet$ ($N_\bullet = 0$ for sample set independent φ). The (D_i, D_j) column shows the pairs of feature dimensions for which the two-sided Wilcoxon signed rank test with the Bonferroni-Holm correction does not reject the hypothesis of equality of median feature values, at the family-wise level 0.05 for each individual feature.

	$\mathcal{A}$			$\mathcal{A}^\top$			$\mathcal{A}_\mathcal{P}$			$\mathcal{A}_\mathcal{P}^\top$		
	$N_\bullet$	rob.(%)	(D_i, D_j)	$N_\bullet$	rob.(%)	(D_i, D_j)	$N_\bullet$	rob.(%)	(D_i, D_j)	$N_\bullet$	rob.(%)	(D_i, D_j)
$\varphi^{\text{Lvl}}_{\texttt{mmce\_lda\_10}}$	6	39.47		6	39.47		13	53.01		13	53.01	
$\varphi^{\text{Lvl}}_{\texttt{mmce\_qda\_10}}$	6	69.30		6	69.88		13	65.30		13	65.91	
$\varphi^{\text{Lvl}}_{\texttt{mmce\_mda\_10}}$	6	28.07		6	28.14		13	27.38		13	27.64	
$\varphi^{\text{Lvl}}_{\texttt{lda\_qda\_10}}$	6	80.09		6	80.12		18	92.37		18	92.38	
$\varphi^{\text{Lvl}}_{\texttt{lda\_mda\_10}}$	6	44.32		6	42.71		13	46.42		13	48.00	
$\varphi^{\text{Lvl}}_{\texttt{qda\_mda\_10}}$	6	81.14		6	80.51		13	78.27		13	77.33	
$\varphi^{\text{Lvl}}_{\texttt{mmce\_lda\_25}}$	6	32.64		6	32.64		13	54.19		13	54.19	
$\varphi^{\text{Lvl}}_{\texttt{mmce\_qda\_25}}$	6	62.01		6	62.41		13	56.49		13	56.69	
$\varphi^{\text{Lvl}}_{\texttt{mmce\_mda\_25}}$	6	22.86		6	22.63		13	27.03		13	26.99	
$\varphi^{\text{Lvl}}_{\texttt{lda\_qda\_25}}$	6	76.25		6	76.29	(3, 5)	15	94.13		15	94.12	
$\varphi^{\text{Lvl}}_{\texttt{lda\_mda\_25}}$	6	32.07	(3, 20)	6	27.68	(3, 10), (10, 20)	13	22.86		13	20.39	
$\varphi^{\text{Lvl}}_{\texttt{qda\_mda\_25}}$	6	77.24		6	77.68		13	63.08		13	63.11	
$\varphi^{\text{Lvl}}_{\texttt{mmce\_lda\_50}}$	6	22.40		6	22.39		13	18.46		13	18.46	
$\varphi^{\text{Lvl}}_{\texttt{mmce\_qda\_50}}$	6	48.24		6	49.46		13	32.88		13	33.15	
$\varphi^{\text{Lvl}}_{\texttt{mmce\_mda\_50}}$	6	20.13		6	19.60		13	38.52		13	37.58	
$\varphi^{\text{Lvl}}_{\texttt{lda\_qda\_50}}$	6	69.22	(10, 20)	6	69.24	(10, 20)	15	84.45		15	84.41	
$\varphi^{\text{Lvl}}_{\texttt{lda\_mda\_50}}$	6	36.05		6	30.90		13	6.80	(5, 20)	13	4.51	
$\varphi^{\text{Lvl}}_{\texttt{qda\_mda\_50}}$	6	42.65	(3, 20)	6	40.68		13	21.96	(5, 20)	13	22.47	(3, 5), (5, 20)
$\varphi^{\text{MM}}_{\texttt{lin\_simple\_adj\_r2}}$	6	19.64		6	19.62		–	–	–	–	–	–
$\varphi^{\text{MM}}_{\texttt{lin\_simple\_coef\_min}}$	6	51.95		6	85.41		–	–	–	–	–	–
$\varphi^{\text{MM}}_{\texttt{lin\_simple\_coef\_max}}$	6	29.73		6	77.68		–	–	–	–	–	–
$\varphi^{\text{MM}}_{\texttt{lin\_simple\_coef\_max\_by\_min}}$	6	27.78		6	69.78		–	–	–	–	–	–
$\varphi^{\text{MM}}_{\texttt{lin\_w\_interact\_adj\_r2}}$	100	44.77		100	43.89		–	–	–	–	–	–
$\varphi^{\text{MM}}_{\texttt{quad\_simple\_adj\_r2}}$	6	50.23		6	52.16		–	–	–	–	–	–
$\varphi^{\text{MM}}_{\texttt{quad\_simple\_cond}}$	6	46.45		6	95.21		–	–	–	–	–	–
$\varphi^{\text{MM}}_{\texttt{quad\_w\_interact\_adj\_r2}}$	629	82.50		531	77.81		–	–	–	–	–	–
$\varphi^{\text{Inf}}_{\texttt{h\_max}}$	6	14.46	(3, 5)	6	33.83		13	35.74		13	35.42	
$\varphi^{\text{Inf}}_{\texttt{eps\_s}}$	6	6.59		6	34.65		3056	26.05		3196	54.56	
$\varphi^{\text{Inf}}_{\texttt{eps\_max}}$	6	64.93		6	84.99		13	65.17		13	84.18	
$\varphi^{\text{Inf}}_{\texttt{m0}}$	6	1.30		6	3.88		–	–	–	–	–	–
$\varphi^{\text{Inf}}_{\texttt{eps\_ratio}}$	6	27.67		6	44.91		–	–	–	–	–	–
$\varphi^{\text{NBC}}_{\texttt{nb\_std\_ratio}}$	6	31.11		6	18.27		–	–	–	–	–	–
$\varphi^{\text{NBC}}_{\texttt{nb\_mean\_ratio}}$	6	38.98		6	32.96		–	–	–	–	–	–
$\varphi^{\text{NBC}}_{\texttt{nb\_cor}}$	6	45.94		6	32.39		–	–	–	–	–	–
$\varphi^{\text{NBC}}_{\texttt{dist\_ratio}}$	6	55.23		6	50.78		–	–	–	–	–	–
$\varphi^{\text{NBC}}_{\texttt{nb\_fitness\_cor}}$	6	77.17		6	77.13		–	–	–	–	–	–

Table 3: TSS nearest features from feature sets Φ_{Basic}, Φ_{CMA}, Φ_{Dis}, and $\Phi_{\text{y-D}}$ (only $\mathcal{T}$-based and sample set indepent, $\mathcal{A}$-based are identical to TSS full in Table 1). Features are grouped according to their feature sets (separated by horizontal lines). Features with less than 25% of values equal to $\bullet$ and robustness greater than 0.9, are in gray. $N_\bullet$ denotes the lowest measured number of points from which at most 1% of feature calculations resulted in $\bullet$ ($N_\bullet = 0$ for sample set independent φ). The (D_i, D_j) column shows the pairs of feature dimensions for which the two-sided Wilcoxon signed rank test with the Bonferroni-Holm correction does not reject the hypothesis of equality of median feature values, at the family-wise level 0.05 for each individual feature.

	$\mathcal{T}$			$\mathcal{T}^\top$			$\mathcal{T}_\mathcal{P}$			$\mathcal{T}_\mathcal{P}^\top$		
	$N_\bullet$	rob.(%)	(D_i, D_j)	$N_\bullet$	rob.(%)	(D_i, D_j)	$N_\bullet$	rob.(%)	(D_i, D_j)	$N_\bullet$	rob.(%)	(D_i, D_j)
$\varphi_{\texttt{dim}}$	0	100.00		0	100.00		0	100.00		0	100.00	
$\varphi_{\texttt{obs}}$	1	100.00		1	100.00		13	100.00		13	100.00	
$\varphi^{\text{CMA}}_{\texttt{generation}}$	0	100.00		0	100.00		0	100.00		0	100.00	
$\varphi^{\text{CMA}}_{\texttt{step\_size}}$	0	100.00		0	100.00		0	100.00		0	100.00	
$\varphi^{\text{CMA}}_{\texttt{restart}}$	0	100.00		0	100.00		0	100.00		0	100.00	
$\varphi^{\text{CMA}}_{\texttt{cma\_mean\_dist}}$	6	70.11		6	56.83		13	63.61		13	54.12	
$\varphi^{\text{CMA}}_{\texttt{evopath\_c\_norm}}$	0	100.00		0	100.00		0	100.00		0	100.00	
$\varphi^{\text{CMA}}_{\texttt{evopath\_s\_norm}}$	0	100.00		0	100.00		0	100.00		0	100.00	
$\varphi^{\text{CMA}}_{\texttt{cma\_lik}}$	1	99.75		1	99.96		13	99.75		13	99.96	
$\varphi^{\text{Dis}}_{\texttt{ratio\_mean\_02}}$	71	57.69	$(2,5)$, $(3,10)$, $(3,20)$	71	62.33	$(2,5)$, $(5,20)$	86	57.36	$(2,5)$, $(3,20)$	86	62.16	
$\varphi^{\text{Dis}}_{\texttt{ratio\_median\_02}}$	71	65.95	$(3,5)$	71	69.28		86	65.41		86	69.60	
$\varphi^{\text{Dis}}_{\texttt{diff\_mean\_02}}$	71	47.75		71	99.52		86	48.77		86	99.52	
$\varphi^{\text{Dis}}_{\texttt{diff\_median\_02}}$	71	50.21		71	99.44		86	51.44		86	99.45	
$\varphi^{\text{Dis}}_{\texttt{ratio\_mean\_05}}$	27	55.22	$(5,20)$	27	60.59		42	54.76	$(5,20)$	42	60.32	
$\varphi^{\text{Dis}}_{\texttt{ratio\_median\_05}}$	27	59.80		27	64.43		42	59.66	$(5,20)$	42	64.57	
$\varphi^{\text{Dis}}_{\texttt{diff\_mean\_05}}$	27	47.99		27	99.49		42	49.16		42	99.49	
$\varphi^{\text{Dis}}_{\texttt{diff\_median\_05}}$	27	51.16		27	99.46		42	52.88		42	99.48	
$\varphi^{\text{Dis}}_{\texttt{ratio\_mean\_10}}$	12	54.63	$(5,20)$	12	59.33		25	54.06	$(5,20)$	25	58.71	
$\varphi^{\text{Dis}}_{\texttt{ratio\_median\_10}}$	12	58.29		12	62.38	$(2,3)$	25	57.80		25	61.82	
$\varphi^{\text{Dis}}_{\texttt{diff\_mean\_10}}$	12	49.70		12	99.52		25	50.88		25	99.52	
$\varphi^{\text{Dis}}_{\texttt{diff\_median\_10}}$	12	53.51		12	99.48		25	55.53		25	99.49	
$\varphi^{\text{Dis}}_{\texttt{ratio\_mean\_25}}$	6	54.80		6	58.86		15	53.87		15	58.06	
$\varphi^{\text{Dis}}_{\texttt{ratio\_median\_25}}$	6	56.21		6	59.74		15	54.91		15	58.26	
$\varphi^{\text{Dis}}_{\texttt{diff\_mean\_25}}$	6	53.78		6	99.55		15	55.29		15	99.55	
$\varphi^{\text{Dis}}_{\texttt{diff\_median\_25}}$	6	57.40		6	99.49		15	60.24		15	99.50	
$\varphi^{\text{y-D}}_{\texttt{skewness}}$	6	36.71		–	–	–	–	–	–	–	–	–
$\varphi^{\text{y-D}}_{\texttt{kurtosis}}$	6	63.06		–	–	–	–	–	–	–	–	–
$\varphi^{\text{y-D}}_{\texttt{number\_of\_peaks}}$	6	32.34		–	–	–	–	–	–	–	–	–

Table 4: TSS nearest features from feature sets Φ_{Lvl}, Φ_{MM}, Φ_{Inf}, and Φ_{NBC} (only $\mathcal{T}$-based and sample set indepent, $\mathcal{A}$-based are identical to TSS full in Table 2). Features are grouped according to their feature sets (separated by horizontal lines). Features with less than 25% of values equal to $\bullet$ and robustness greater than 0.9, are in gray. $N_\bullet$ denotes the lowest measured number of points from which at most 1% of feature calculations resulted in $\bullet$ ($N_\bullet = 0$ for sample set independent φ). The (D_i, D_j) column shows the pairs of feature dimensions for which the two-sided Wilcoxon signed rank test with the Bonferroni-Holm correction does not reject the hypothesis of equality of median feature values, at the family-wise level 0.05 for each individual feature.

	$\mathcal{T}$			$\mathcal{T}^\top$			$\mathcal{T}_\mathcal{P}$			$\mathcal{T}_\mathcal{P}^\top$		
	$N_\bullet$	rob.(%)	(D_i, D_j)	$N_\bullet$	rob.(%)	(D_i, D_j)	$N_\bullet$	rob.(%)	(D_i, D_j)	$N_\bullet$	rob.(%)	(D_i, D_j)
$\varphi^{\mathrm{Lvl}}_{\texttt{mmce\_lda\_10}}$	6	39.47		6	39.47		13	53.01		13	53.01	
$\varphi^{\mathrm{Lvl}}_{\texttt{mmce\_qda\_10}}$	6	69.30		6	69.88		13	65.30		13	65.91	
$\varphi^{\mathrm{Lvl}}_{\texttt{mmce\_mda\_10}}$	6	28.07		6	28.14		13	27.38		13	27.64	
$\varphi^{\mathrm{Lvl}}_{\texttt{lda\_qda\_10}}$	6	80.09		6	80.12		18	92.37		18	92.38	
$\varphi^{\mathrm{Lvl}}_{\texttt{lda\_mda\_10}}$	6	44.32		6	42.71		13	46.42		13	48.00	
$\varphi^{\mathrm{Lvl}}_{\texttt{qda\_mda\_10}}$	6	81.14		6	80.51		13	78.27		13	77.33	
$\varphi^{\mathrm{Lvl}}_{\texttt{mmce\_lda\_25}}$	6	32.64		6	32.64		13	54.19		13	54.19	
$\varphi^{\mathrm{Lvl}}_{\texttt{mmce\_qda\_25}}$	6	62.01		6	62.41		13	56.49		13	56.69	
$\varphi^{\mathrm{Lvl}}_{\texttt{mmce\_mda\_25}}$	6	22.86		6	22.63		13	27.03		13	26.99	
$\varphi^{\mathrm{Lvl}}_{\texttt{lda\_qda\_25}}$	6	76.25		6	76.29	(3, 5)	15	94.13		15	94.12	
$\varphi^{\mathrm{Lvl}}_{\texttt{lda\_mda\_25}}$	6	32.07	(3, 20)	6	27.68	(3, 10), (10, 20)	13	22.86		13	20.39	
$\varphi^{\mathrm{Lvl}}_{\texttt{qda\_mda\_25}}$	6	77.24		6	77.68		13	63.08		13	63.11	
$\varphi^{\mathrm{Lvl}}_{\texttt{mmce\_lda\_50}}$	6	22.40		6	22.39		13	18.46		13	18.46	
$\varphi^{\mathrm{Lvl}}_{\texttt{mmce\_qda\_50}}$	6	48.24		6	49.46		13	32.88		13	33.15	
$\varphi^{\mathrm{Lvl}}_{\texttt{mmce\_mda\_50}}$	6	20.13		6	19.60		13	38.52		13	37.58	
$\varphi^{\mathrm{Lvl}}_{\texttt{lda\_qda\_50}}$	6	69.22	(10, 20)	6	69.24	(10, 20)	15	84.45		15	84.41	
$\varphi^{\mathrm{Lvl}}_{\texttt{lda\_mda\_50}}$	6	36.05		6	30.90		13	6.80	(5, 20)	13	4.51	
$\varphi^{\mathrm{Lvl}}_{\texttt{qda\_mda\_50}}$	6	42.65	(3, 20)	6	40.68		13	21.96	(5, 20)	13	22.47	(3, 5), (5, 20)
$\varphi^{\mathrm{MM}}_{\texttt{lin\_simple\_adj\_r2}}$	6	19.64		6	19.62		–	–	–	–	–	–
$\varphi^{\mathrm{MM}}_{\texttt{lin\_simple\_coef\_min}}$	6	51.95		6	85.41		–	–	–	–	–	–
$\varphi^{\mathrm{MM}}_{\texttt{lin\_simple\_coef\_max}}$	6	29.73		6	77.68		–	–	–	–	–	–
$\varphi^{\mathrm{MM}}_{\texttt{lin\_simple\_coef\_max\_by\_min}}$	6	27.78		6	69.78		–	–	–	–	–	–
$\varphi^{\mathrm{MM}}_{\texttt{lin\_w\_interact\_adj\_r2}}$	100	44.77		100	43.89		–	–	–	–	–	–
$\varphi^{\mathrm{MM}}_{\texttt{quad\_simple\_adj\_r2}}$	6	50.23		6	52.16		–	–	–	–	–	–
$\varphi^{\mathrm{MM}}_{\texttt{quad\_simple\_cond}}$	6	46.45		6	95.21		–	–	–	–	–	–
$\varphi^{\mathrm{MM}}_{\texttt{quad\_w\_interact\_adj\_r2}}$	629	82.50		531	77.81		–	–	–	–	–	–
$\varphi^{\mathrm{Inf}}_{\texttt{h\_max}}$	6	14.46	(3, 5)	6	33.83		13	35.74		13	35.42	
$\varphi^{\mathrm{Inf}}_{\texttt{eps\_s}}$	6	6.59		6	34.65		3056	26.05		3196	54.56	
$\varphi^{\mathrm{Inf}}_{\texttt{eps\_max}}$	6	64.93		6	84.99		13	65.17		13	84.18	
$\varphi^{\mathrm{Inf}}_{\texttt{m0}}$	6	1.30		6	3.88		–	–	–	–	–	–
$\varphi^{\mathrm{Inf}}_{\texttt{eps\_ratio}}$	6	27.67		6	44.91		–	–	–	–	–	–
$\varphi^{\mathrm{NBC}}_{\texttt{nb\_std\_ratio}}$	6	31.11		6	18.27		–	–	–	–	–	–
$\varphi^{\mathrm{NBC}}_{\texttt{nb\_mean\_ratio}}$	6	38.98		6	32.96		–	–	–	–	–	–
$\varphi^{\mathrm{NBC}}_{\texttt{nb\_cor}}$	6	45.94		6	32.39		–	–	–	–	–	–
$\varphi^{\mathrm{NBC}}_{\texttt{dist\_ratio}}$	6	55.23		6	50.78		–	–	–	–	–	–
$\varphi^{\mathrm{NBC}}_{\texttt{nb\_fitness\_cor}}$	6	77.17		6	77.13		–	–	–	–	–	–

Table 5: TSS knn features from feature sets Φ_{Basic}, Φ_{CMA}, Φ_{Dis}, and $\Phi_{\text{y-D}}$ (only $\mathcal{T}$-based and sample set indepent, $\mathcal{A}$-based are identical to TSS full in Table 1). Features are grouped according to their feature sets (separated by horizontal lines). Features with less than 25% of values equal to $\bullet$ and robustness greater than 0.9, are in gray. $N_\bullet$ denotes the lowest measured number of points from which at most 1% of feature calculations resulted in $\bullet$ ($N_\bullet = 0$ for sample set independent φ). The (D_i, D_j) column shows the pairs of feature dimensions for which the two-sided Wilcoxon signed rank test with the Bonferroni-Holm correction does not reject the hypothesis of equality of median feature values, at the family-wise level 0.05 for each individual feature.

	$\mathcal{T}$			$\mathcal{T}^\top$			$\mathcal{T}_\mathcal{P}$			$\mathcal{T}_\mathcal{P}^\top$		
	$N_\bullet$	rob.(%)	(D_i, D_j)	$N_\bullet$	rob.(%)	(D_i, D_j)	$N_\bullet$	rob.(%)	(D_i, D_j)	$N_\bullet$	rob.(%)	(D_i, D_j)
$\varphi_{\texttt{dim}}$	0	100.00		0	100.00		0	100.00		0	100.00	
$\varphi_{\texttt{obs}}$	1	92.16		1	92.16		13	92.26		13	92.26	
$\varphi^{\text{CMA}}_{\texttt{generation}}$	0	100.00		0	100.00		0	100.00		0	100.00	
$\varphi^{\text{CMA}}_{\texttt{step\_size}}$	0	100.00		0	100.00		0	100.00		0	100.00	
$\varphi^{\text{CMA}}_{\texttt{restart}}$	0	100.00		0	100.00		0	100.00		0	100.00	
$\varphi^{\text{CMA}}_{\texttt{cma\_mean\_dist}}$	6	78.64		6	98.40		13	93.96		13	98.71	
$\varphi^{\text{CMA}}_{\texttt{evopath\_c\_norm}}$	0	100.00		0	100.00		0	100.00		0	100.00	
$\varphi^{\text{CMA}}_{\texttt{evopath\_s\_norm}}$	0	100.00		0	100.00		0	100.00		0	100.00	
$\varphi^{\text{CMA}}_{\texttt{cma\_lik}}$	1	98.81		1	99.95		13	98.80		13	99.96	
$\varphi^{\text{Dis}}_{\texttt{ratio\_mean\_02}}$	74	30.53		74	23.95		109	29.18		109	23.33	
$\varphi^{\text{Dis}}_{\texttt{ratio\_median\_02}}$	74	34.49	$(5, 10)$	74	25.97		109	32.08		109	24.94	
$\varphi^{\text{Dis}}_{\texttt{diff\_mean\_02}}$	74	50.17		74	96.75		109	50.99		109	96.76	
$\varphi^{\text{Dis}}_{\texttt{diff\_median\_02}}$	74	49.55		74	95.54		109	50.37		109	95.47	
$\varphi^{\text{Dis}}_{\texttt{ratio\_mean\_05}}$	30	28.43		30	21.63		53	26.56		53	21.19	
$\varphi^{\text{Dis}}_{\texttt{ratio\_median\_05}}$	30	35.85		30	24.73	$(3, 10)$	53	31.45	$(2, 20)$	53	23.58	
$\varphi^{\text{Dis}}_{\texttt{diff\_mean\_05}}$	30	55.94		30	96.80		53	56.94		53	96.85	
$\varphi^{\text{Dis}}_{\texttt{diff\_median\_05}}$	30	55.11		30	95.80		53	56.23		53	95.82	
$\varphi^{\text{Dis}}_{\texttt{ratio\_mean\_10}}$	15	25.57	$(2, 3)$, $(2, 10)$, $(3, 10)$, $(5, 10)$	15	19.13		28	24.33	$(5, 10)$	28	18.97	
$\varphi^{\text{Dis}}_{\texttt{ratio\_median\_10}}$	15	32.86		15	21.96		28	29.68		28	21.49	
$\varphi^{\text{Dis}}_{\texttt{diff\_mean\_10}}$	15	58.33		15	96.96		28	59.42		28	97.02	
$\varphi^{\text{Dis}}_{\texttt{diff\_median\_10}}$	15	57.49		15	96.09		28	58.85		28	96.16	
$\varphi^{\text{Dis}}_{\texttt{ratio\_mean\_25}}$	6	19.39		6	14.96		15	19.01		15	14.16	
$\varphi^{\text{Dis}}_{\texttt{ratio\_median\_25}}$	6	23.38		6	15.76	$(3, 5)$	15	24.08		15	16.25	
$\varphi^{\text{Dis}}_{\texttt{diff\_mean\_25}}$	6	62.31		6	97.06		15	62.72		15	97.12	
$\varphi^{\text{Dis}}_{\texttt{diff\_median\_25}}$	6	61.67		6	96.27		15	62.60		15	96.36	
$\varphi^{\text{y-D}}_{\texttt{skewness}}$	6	30.49		–	–	–	–	–	–	–	–	–
$\varphi^{\text{y-D}}_{\texttt{kurtosis}}$	6	72.61		–	–	–	–	–	–	–	–	–
$\varphi^{\text{y-D}}_{\texttt{number\_of\_peaks}}$	6	26.63		–	–	–	–	–	–	–	–	–

Table 6: TSS knn features from feature sets Φ_{Lvl}, Φ_{MM}, Φ_{Inf}, and Φ_{NBC} (only $\mathcal{T}$-based, $\mathcal{A}$-based are identical to TSS full in Table 2). Features are grouped according to their feature sets (separated by horizontal lines). Features with less than 25% of values equal to $\bullet$ and robustness greater than 0.9, are in gray. $N_\bullet$ denotes the lowest measured number of points from which at most 1% of feature calculations resulted in $\bullet$ ($N_\bullet = 0$ for sample set independent φ). The (D_i, D_j) column shows the pairs of feature dimensions for which the two-sided Wilcoxon signed rank test with the Bonferroni-Holm correction does not reject the hypothesis of equality of median feature values, at the family-wise level 0.05 for each individual feature.

	$\mathcal{T}$			$\mathcal{T}^\top$			$\mathcal{T}_\mathcal{P}$			$\mathcal{T}_\mathcal{P}^\top$		
	$N_\bullet$	rob.(%)	(D_i, D_j)	$N_\bullet$	rob.(%)	(D_i, D_j)	$N_\bullet$	rob.(%)	(D_i, D_j)	$N_\bullet$	rob.(%)	(D_i, D_j)
$\varphi^{\text{Lvl}}_{\texttt{mmce\_lda\_10}}$	6	11.39		6	11.39		13	9.63		13	9.93	
$\varphi^{\text{Lvl}}_{\texttt{mmce\_qda\_10}}$	6	67.83		6	69.40		13	69.58		13	70.79	$(3, 5)$
$\varphi^{\text{Lvl}}_{\texttt{mmce\_mda\_10}}$	6	21.15		6	20.32		13	19.93		13	19.25	
$\varphi^{\text{Lvl}}_{\texttt{lda\_qda\_10}}$	23	67.38		23	67.38		37	70.31		37	70.32	
$\varphi^{\text{Lvl}}_{\texttt{lda\_mda\_10}}$	6	15.74		6	13.69		13	22.17		13	20.80	
$\varphi^{\text{Lvl}}_{\texttt{qda\_mda\_10}}$	6	57.54		6	56.51		13	55.10		13	55.06	
$\varphi^{\text{Lvl}}_{\texttt{mmce\_lda\_25}}$	6	9.73		6	9.63		13	15.42	$(3, 20)$	13	15.65	$(3, 20)$
$\varphi^{\text{Lvl}}_{\texttt{mmce\_qda\_25}}$	6	37.27		6	38.00		13	36.70		13	37.25	
$\varphi^{\text{Lvl}}_{\texttt{mmce\_mda\_25}}$	6	17.79		6	16.60		13	15.60		13	14.90	
$\varphi^{\text{Lvl}}_{\texttt{lda\_qda\_25}}$	6	73.93	$(5, 10)$	6	73.87	$(5, 10)$	18	89.97		18	89.95	
$\varphi^{\text{Lvl}}_{\texttt{lda\_mda\_25}}$	6	8.97		6	6.33		13	4.46		13	2.95	
$\varphi^{\text{Lvl}}_{\texttt{qda\_mda\_25}}$	6	47.47		6	49.72		13	35.90		13	39.50	$(2, 10)$, $(3, 5)$
$\varphi^{\text{Lvl}}_{\texttt{mmce\_lda\_50}}$	6	17.46		6	16.62		13	7.86		13	7.88	
$\varphi^{\text{Lvl}}_{\texttt{mmce\_qda\_50}}$	6	34.21		6	34.32		13	23.54		13	27.64	
$\varphi^{\text{Lvl}}_{\texttt{mmce\_mda\_50}}$	6	24.08		6	20.56		13	20.74		13	19.07	$(2, 3)$
$\varphi^{\text{Lvl}}_{\texttt{lda\_qda\_50}}$	6	41.38		6	41.37		18	73.19		18	73.01	
$\varphi^{\text{Lvl}}_{\texttt{lda\_mda\_50}}$	6	26.79		6	32.00		13	2.70	$(5, 10)$	13	2.30	
$\varphi^{\text{Lvl}}_{\texttt{qda\_mda\_50}}$	6	17.30		6	24.22		13	5.95		13	7.69	$(5, 20)$
$\varphi^{\text{MM}}_{\texttt{lin\_simple\_adj\_r2}}$	6	15.67		6	15.63		–	–	–	–	–	–
$\varphi^{\text{MM}}_{\texttt{lin\_simple\_coef\_min}}$	6	97.40		6	63.93		–	–	–	–	–	–
$\varphi^{\text{MM}}_{\texttt{lin\_simple\_coef\_max}}$	6	98.12	$(2, 3)$	6	87.04		–	–	–	–	–	–
$\varphi^{\text{MM}}_{\texttt{lin\_simple\_coef\_max\_by\_min}}$	6	34.41		6	45.65		–	–	–	–	–	–
$\varphi^{\text{MM}}_{\texttt{lin\_w\_interact\_adj\_r2}}$	100	37.50		100	30.25		–	–	–	–	–	–
$\varphi^{\text{MM}}_{\texttt{quad\_simple\_adj\_r2}}$	6	41.50		6	37.91		–	–	–	–	–	–
$\varphi^{\text{MM}}_{\texttt{quad\_simple\_cond}}$	6	92.55		6	87.21		–	–	–	–	–	–
$\varphi^{\text{MM}}_{\texttt{quad\_w\_interact\_adj\_r2}}$	688	69.48		632	60.02		–	–	–	–	–	–
$\varphi^{\text{Inf}}_{\texttt{h\_max}}$	6	1.94		6	1.57	$(3, 5)$	13	31.85		13	38.67	
$\varphi^{\text{Inf}}_{\texttt{eps\_s}}$	6	38.77		6	3.58		2424	66.99		2485	70.18	$(5, 10)$
$\varphi^{\text{Inf}}_{\texttt{eps\_max}}$	6	97.84		6	59.04		13	97.70		13	58.95	
$\varphi^{\text{Inf}}_{\texttt{m0}}$	6	0.84		6	0.44		–	–	–	–	–	–
$\varphi^{\text{Inf}}_{\texttt{eps\_ratio}}$	6	17.19		6	5.59		–	–	–	–	–	–
$\varphi^{\text{NBC}}_{\texttt{nb\_std\_ratio}}$	6	20.10		6	14.20		–	–	–	–	–	–
$\varphi^{\text{NBC}}_{\texttt{nb\_mean\_ratio}}$	6	36.38		6	30.42		–	–	–	–	–	–
$\varphi^{\text{NBC}}_{\texttt{nb\_cor}}$	6	28.49		6	26.50		–	–	–	–	–	–
$\varphi^{\text{NBC}}_{\texttt{dist\_ratio}}$	6	41.24		6	26.53		–	–	–	–	–	–
$\varphi^{\text{NBC}}_{\texttt{nb\_fitness\_cor}}$	6	45.12		6	45.14		–	–	–	–	–	–

4 Connection of Landscape Features and Model Error Results

We have tested the statistical significance of the MSE and RDE differences for 19 surrogate model settings using TSS full and TSS nearest methods and also the lmm surrogate model utilizing TSS knn, i. e., 39 different combinations of model settings ψ and TSS methods (ψ, TSS), on all available sample sets using the non-parametric two-sided Wilcoxon signed rank test with the Holm correction for the family-wise error. To better illustrate the differences between individual settings, we also count the percentage of cases at which one setting had the error lower than the other. The pairwise score and the statistical significance of the pairwise differences are summarized in Tables 7 and 8.

To compare the convenience of individual features as descriptors of areas where the surrogate model $\mathcal{M}$ with a particular (ψ, TSS) combination has the best performance, we use the Kolmogorov-Smirnov test (KS test) testing the equality of the distribution of values of individual features calculated on the whole testing dataset and on only those sample sets from the testing dataset for which the (ψ, TSS) combination leads to the lowest error (MSE or RDE) among all tested combinations. The hypothesis of distribution equality is tested at the family-wise significance level $\alpha = 0.05$ after the Holm correction. The resulting p-values are summarized in Tables 9–13.

Table 7: A pairwise comparison of the model settings MSE in different TSS. The percentage of wins of i-th model setting against j-th model setting over all available data is given in the i-th row and j-th column. The numbers in bold mark the row model setting being significantly better than the column model setting according to the two-sided Wilcoxon signed rank test with the Holm correction at family-wise significance level $\alpha = 0.05$.

MSE		TSS	full																			nearest																			knn
		model	GP								RF									lmm	lq	GP								RF									lmm	lq	lmm
TSS	model	setting	Gibbs	LIN	Mat	NN	Q	RQ	SE+Q	SE	CART MSE	CART RDE	OC1	PAIR MSE	PAIR RDE	SCRT MSE	SCRT RDE	SUPP MSE	SUPP RDE			Gibbs	LIN	Mat	NN	Q	RQ	SE+Q	SE	CART MSE	CART RDE	OC1	PAIR MSE	PAIR RDE	SCRT MSE	SCRT RDE	SUPP MSE	SUPP RDE			
full	GP	Gibbs		**85**	**68**	**53**	43	35	38	47	**73**	**72**	**77**	**78**	**77**	**78**	**58**	**71**	**62**	**70**	**70**	42	**69**	**52**	49	35	34	34	36	**62**	**61**	**54**	**63**	**61**	**63**	**57**	**59**	**60**	**51**	**51**	**59**
		LIN	15		13	6	10	10	5	10	11	11	14	12	13	15	8	10	9	9	11	12	11	8	7	6	5	5	5	11	10	9	10	10	11	10	10	11	10	8	16
		Mat	32	**87**		24	25	19	18	22	42	41	46	45	**52**	**54**	30	40	33	27	49	29	41	24	27	19	20	20	20	34	33	30	34	32	35	32	33	34	25	25	33
		NN	47	**94**	**76**		**50.2**	34	36	47	**77**	**80**	**78**	**75**	**87**	**88**	**52**	**75**	**59**	**80**	**81**	37	**58**	42	36	25	19	20	21	**53**	**52**	46	**55**	**52**	**56**	48	**54**	**52**	34	36	38
		Q	**57**	**90**	**75**	49.8		36	38	48	**66**	**67**	**69**	**69**	**73**	**75**	**54**	**65**	**59**	**63**	**68**	48	**61**	**52**	49	35	34	35	37	**57**	**57**	**52**	**58**	**57**	**59**	**54**	**57**	**57**	45	45	48
		RQ	**65**	**90**	**81**	**66**	**64**		**58**	**68**	**78**	**81**	**81**	**78**	**83**	**84**	**62**	**77**	**67**	**79**	**78**	46	**61**	**55**	49.6	37	29	30	31	**61**	**60**	**55**	**62**	**60**	**62**	**57**	**62**	**60**	42	42	44
		SE+Q	**62**	**95**	**82**	**64**	**62**	42		**69**	**83**	**85**	**83**	**80**	**90**	**91**	**62**	**81**	**67**	**84**	**84**	47	**61**	**53**	48	35	28	28	31	**61**	**60**	**55**	**62**	**60**	**63**	**57**	**62**	**60**	41	43	43
		SE	**53**	**90**	**78**	**53**	**52**	32	31		**74**	**77**	**77**	**73**	**81**	**81**	**55**	**73**	**61**	**77**	**77**	41	**56**	47	43	29	24	24	25	**54**	**54**	48	**55**	**53**	**56**	**51**	**55**	**53**	35	36	38
	RF	CART MSE	27	**89**	**58**	23	34	22	17	26		**50.02**	**57**	**55**	**72**	**74**	23	42	33	**59**	**63**	22	42	25	19	15	12	12	13	33	32	26	35	31	36	29	34	34	25	25	30
		CART RDE	28	**89**	**59**	20	33	19	15	23	49.98		**56**	**53**	**68**	**71**	22	42	31	**59**	**63**	24	42	25	19	15	12	12	14	33	32	25	34	30	35	29	32	34	25	25	31
		OC1	23	**86**	**54**	22	31	19	17	23	43	44		44	**61**	**64**	18	39	27	**56**	**59**	20	42	21	16	13	11	11	12	26	25	20	28	24	29	22	29	27	24	24	33
		PAIR MSE	22	**88**	**55**	25	31	22	20	27	45	47	**56**		**63**	**66**	19	41	26	**57**	**60**	17	47	22	15	12	9	9	11	26	26	19	28	24	29	22	29	27	24	23	33
		PAIR RDE	23	**87**	48	13	27	17	10	19	28	32	39	37		**57**	16	24	23	47	**52**	20	35	19	14	11	9	9	10	25	25	20	27	23	28	23	26	27	21	20	26
		SCRT MSE	22	**85**	46	12	25	16	9	19	26	29	36	34	43		15	22	22	44	49.7	19	34	18	13	11	8	9	9	24	23	19	26	22	26	22	24	26	20	19	26
		SCRT RDE	42	**92**	**70**	48	46	38	38	45	**77**	**78**	**82**	**81**	**84**	**85**		**75**	**58**	**73**	**75**	34	**67**	43	34	26	22	22	24	**54**	**53**	42	**58**	53	**59**	47	**55**	**54**	40	41	45
		SUPP MSE	29	**90**	**60**	25	35	23	19	27	**58**	**58**	**61**	**59**	**76**	**78**	25		37	**62**	**65**	24	45	27	20	16	13	13	14	37	36	28	39	35	39	32	38	37	26	27	32
		SUPP RDE	38	**91**	**67**	41	41	33	33	39	**67**	**69**	**73**	**74**	**77**	**78**	42	**63**		**69**	**71**	30	**64**	38	30	22	19	19	21	47	46	35	**50.4**	45	**51**	40	47	47	37	38	43
	lmm		30	**91**	**73**	20	37	21	16	23	41	41	44	43	**53**	**56**	27	38	31		**77**	27	37	27	24	19	18	18	19	30	29	27	30	29	31	28	30	30	24	27	32
	lq		30	**89**	**51**	19	32	22	16	23	37	37	41	40	48	**50.3**	25	35	29	23		27	35	24	23	19	17	17	17	29	28	26	29	28	30	27	28	29	20	20	26
nearest	GP	Gibbs	**58**	**88**	**71**	**63**	**52**	54	**53**	**59**	**78**	**76**	**80**	**83**	**80**	**81**	**66**	**76**	**70**	**73**	**73**		**79**	**63**	**58**	45	41	39	44	**72**	**71**	**64**	**73**	**71**	**74**	**67**	**69**	**70**	**62**	**63**	**70**
		LIN	31	**89**	**59**	42	39	39	39	44	**58**	**58**	**58**	**53**	**65**	**66**	33	**55**	36	**63**	**65**	21		20	15	13	9	9	10	25	24	17	25	22	25	19	27	23	21	25	29
		Mat	48	**92**	**76**	**58**	48	45	47	53	**75**	**75**	**79**	**78**	**81**	**82**	**57**	**73**	**62**	**73**	**76**	37	**80**		41	26	22	23	25	**58**	**57**	47	**61**	**56**	**62**	**51**	**57**	**56**	37	44	47
		NN	**51**	**93**	**73**	**64**	**51**	**50.4**	**52**	**57**	**81**	**81**	**84**	**85**	**86**	**87**	**66**	**80**	**70**	**76**	**77**	42	**85**	**59**		36	24	25	28	**71**	**70**	**60**	**75**	**70**	**76**	**63**	**68**	**69**	49	**54**	**54**
		Q	**65**	**94**	**81**	**75**	**65**	**63**	**65**	**71**	**85**	**85**	**87**	**88**	**89**	**89**	**74**	**84**	**78**	**81**	**81**	**55**	**87**	**74**	**64**		41	39	49	**77**	**76**	**70**	**80**	**76**	**80**	**71**	**76**	**75**	**60**	**62**	**64**
		RQ	**66**	**95**	**80**	**81**	**66**	**71**	**72**	**76**	**88**	**88**	**89**	**91**	**91**	**92**	**78**	**87**	**81**	**82**	**83**	**59**	**91**	**78**	**76**	**59**		48	**53**	**83**	**83**	**78**	**85**	**83**	**86**	**79**	**83**	**81**	**67**	**74**	**72**
		SE+Q	**66**	**95**	**80**	**80**	**65**	**70**	**72**	**76**	**88**	**88**	**89**	**91**	**91**	**91**	**78**	**87**	**81**	**82**	**83**	**61**	**91**	**77**	**75**	**61**	**52**		**61**	**83**	**83**	**78**	**85**	**82**	**86**	**79**	**82**	**81**	**68**	**74**	**72**
		SE	**64**	**95**	**80**	**79**	**63**	**69**	**69**	**75**	**87**	**86**	**88**	**89**	**90**	**91**	**76**	**86**	**79**	**81**	**83**	**56**	**90**	**75**	**72**	**51**	47	39		**81**	**80**	**74**	**83**	**80**	**83**	**76**	**80**	**79**	**65**	**72**	**69**
	RF	CART MSE	38	**89**	**66**	47	43	39	39	46	**67**	**67**	**74**	**74**	**75**	**76**	46	**63**	53	**70**	**71**	28	**75**	42	29	23	17	17	19		46	33	**56**	45	**58**	39	49	47	37	39	45
		CART RDE	39	**90**	**67**	48	43	40	40	46	**68**	**68**	**75**	**74**	**75**	**77**	47	**64**	**54**	**71**	**72**	29	**76**	43	30	24	17	17	20	**54**		34	**58**	48	**60**	40	49.8	49	38	40	46
		OC1	46	**91**	**70**	**54**	48	45	45	**52**	**74**	**75**	**80**	**81**	**80**	**81**	**58**	**72**	**65**	**73**	**74**	36	**83**	53	40	30	22	22	26	**67**	**66**		**72**	**65**	**73**	**57**	**66**	**65**	46	47	**51**
		PAIR MSE	37	**90**	**66**	45	42	38	38	45	**65**	**66**	**72**	**72**	**73**	**74**	42	**61**	49.6	**70**	**71**	27	**75**	39	25	20	15	15	17	44	42	28		40	**52**	33	43	42	36	37	43
		PAIR RDE	39	**90**	**68**	48	43	40	40	47	**69**	**70**	**76**	**76**	**77**	**78**	47	**65**	**55**	**71**	**72**	29	**78**	44	30	24	17	18	20	**55**	**52**	35	**60**		**62**	41	**51**	**50.2**	39	40	46
		SCRT MSE	37	**89**	**65**	44	41	38	37	44	**64**	**65**	**71**	**71**	**72**	**74**	41	**61**	49	**69**	**70**	26	**75**	38	24	20	14	14	17	42	40	27	48	38		32	43	41	35	37	43
		SCRT RDE	43	**90**	**68**	**52**	46	43	43	49	**71**	**71**	**78**	**78**	**77**	**78**	**53**	**68**	**60**	**72**	**73**	33	**81**	49	37	29	21	21	24	**61**	**60**	43	**67**	**59**	**68**		**58**	**60**	42	44	49
		SUPP MSE	41	**90**	**67**	46	43	38	38	45	**66**	**68**	**71**	**71**	**74**	**76**	45	**62**	**53**	**70**	**72**	31	**73**	43	32	24	17	18	20	**51**	**50.2**	34	**57**	49	**57**	42		49.9	39	40	45
		SUPP RDE	40	**89**	**66**	48	43	40	40	47	**66**	**66**	**73**	**73**	**73**	**74**	46	**63**	**53**	**70**	**71**	30	**77**	44	31	25	19	19	21	53	**51**	35	**58**	49.8	**59**	40	**50.1**		39	40	46
	lmm		49	**90**	**75**	**66**	**55**	**58**	**59**	**65**	**75**	**75**	**76**	**76**	**79**	**80**	**60**	**74**	**63**	**76**	**80**	38	**79**	**63**	**51**	40	33	32	35	**63**	**62**	**54**	**64**	**61**	**65**	**58**	**61**	**61**		**67**	**58**
	lq		49	**92**	**75**	**64**	**55**	**58**	**57**	**64**	**75**	**75**	**76**	**77**	**80**	**81**	**59**	**73**	**62**	**73**	**80**	37	**75**	**56**	46	38	26	26	28	**61**	**60**	**53**	**63**	**60**	**63**	**56**	**60**	**60**	33		47
knn	lmm		41	**84**	**67**	**62**	**52**	**56**	**57**	**62**	**70**	**69**	**67**	**67**	**74**	**74**	**55**	**68**	**57**	**68**	**74**	30	**71**	**53**	46	36	28	28	31	55	54	49	**57**	**54**	**57**	**51**	**55**	54	42	**53**	

Table 8: A pairwise comparison of the model settings RDE in different TSS. The percentage of wins of i-th model setting against j-th model setting over all available data is given in the i-th row and j-th column. The numbers in bold mark the row model setting being significantly better than the column model setting according to the two-sided Wilcoxon signed rank test with the Holm correction at family-wise significance level $\alpha = 0.05$.

RDE	TSS		full																			nearest																			knn
	model		GP								RF									lmm	lq	GP								RF									lmm	lq	lmm
TSS	model	setting	Gibbs	LIN	Mat	NN	Q	RQ	SE+Q	SE	CART MSE	CART RDE	OC1	PAIR MSE	PAIR RDE	SCRT MSE	SCRT RDE	SUPP MSE	SUPP RDE			Gibbs	LIN	Mat	NN	Q	RQ	SE+Q	SE	CART MSE	CART RDE	OC1	PAIR MSE	PAIR RDE	SCRT MSE	SCRT RDE	SUPP MSE	SUPP RDE			
full	GP	Gibbs		**76**	**63**	**55**	47	42	44	48	**63**	**62**	**67**	**67**	**67**	**67**	**58**	**62**	**58**	**62**	**62**	42	**74**	**60**	**55**	42	39	39	41	**60**	**60**	**56**	**63**	**60**	**63**	**59**	**57**	**61**	**50.03**	**51**	**50.1**
		LIN	24		34	29	26	24	26	27	35	32	34	33	39	40	25	33	24	36	36	17	41	25	19	16	10	10	11	25	25	21	28	24	28	24	22	26	17	18	17
		Mat	37	**66**		40	33	31	32	34	48	47	**52**	**51**	**53**	**54**	42	47	41	43	47	32	**62**	40	37	27	24	24	25	43	43	40	47	43	47	42	41	45	31	33	33
		NN	45	**71**	**60**		43	37	39	42	**58**	**57**	**61**	**61**	**64**	**64**	51	**57**	**51**	**57**	**57**	35	**68**	51	42	33	23	23	25	**53**	**53**	48	**57**	**53**	**57**	**52**	**51**	**54**	34	37	34
		Q	**53**	**74**	**67**	**57**		44	46	**51**	**62**	**61**	**65**	**65**	**67**	**67**	**56**	**61**	**56**	**62**	**63**	44	**72**	**58**	**51**	39	35	34	36	**58**	**57**	**54**	**61**	**58**	**61**	**57**	**56**	**59**	45	46	44
		RQ	**58**	**76**	**69**	**63**	**56**		**54**	**57**	**65**	**64**	**68**	**68**	**70**	**70**	**60**	**64**	**59**	**65**	**65**	44	**74**	**62**	**53**	44	30	31	33	**61**	**61**	**57**	**64**	**61**	**64**	**60**	**59**	**62**	42	45	41
		SE+Q	**56**	**74**	**68**	**61**	**54**	46		**57**	**64**	**63**	**67**	**66**	**69**	**69**	**59**	**62**	**58**	**63**	**63**	43	**72**	**61**	**52**	43	30	29	31	**60**	**60**	**56**	**63**	**60**	**63**	**59**	**58**	**61**	40	43	40
		SE	**52**	**73**	**66**	**58**	49	43	43		**62**	**60**	**65**	**64**	**67**	**68**	**55**	**61**	**55**	**62**	**62**	40	**71**	**58**	49	38	28	26	27	**57**	**57**	**52**	**60**	**57**	**61**	**56**	**54**	**58**	38	41	37
	RF	CART MSE	37	**65**	**52**	42	38	35	36	38		48	**55**	**53**	**58**	**59**	40	48	41	**51**	**51**	30	**61**	42	34	29	20	21	21	44	44	38	48	43	49	42	41	46	30	33	30
		CART RDE	38	**68**	**53**	43	39	36	37	40	**52**		**57**	**55**	**60**	**60**	41	49.8	41	**53**	**53**	30	**63**	43	36	29	22	22	23	44	44	39	49	44	49.7	43	42	47	32	35	32
		OC1	33	**66**	48	39	35	32	33	35	45	43		48	**53**	**54**	35	43	34	48	49	27	**60**	39	31	26	20	20	21	37	37	33	43	37	43	36	36	39	30	32	30
		PAIR MSE	33	**67**	49	39	35	32	34	36	47	45	**52**		**55**	**55**	35	45	34	49.6	**50.2**	27	**63**	40	31	26	19	19	20	39	39	33	45	38	45	37	36	41	30	32	30
		PAIR RDE	33	**61**	47	36	33	30	31	33	42	40	47	45		**51**	33	40	33	45	46	26	**56**	37	28	24	16	16	17	36	36	31	42	35	42	35	34	39	26	28	26
		SCRT MSE	33	**60**	46	36	33	30	31	32	41	40	46	45	49		32	39	32	45	45	26	**56**	36	27	23	16	16	17	35	35	30	40	34	41	34	33	38	25	28	25
		SCRT RDE	42	**75**	**58**	49	44	40	41	45	**60**	**59**	**65**	**65**	**67**	**68**		**58**	49.8	**59**	**59**	35	**71**	50.02	42	34	28	27	29	**53**	**52**	47	**58**	**52**	**58**	51	**50.3**	**55**	39	41	39
		SUPP MSE	38	**67**	**53**	43	39	36	38	39	**52**	**50.2**	**57**	**55**	**60**	**61**	42		42	**53**	**53**	31	**63**	43	35	29	22	22	22	46	45	40	49.9	45	**50.3**	44	42	48	32	34	32
		SUPP RDE	42	**76**	**59**	49	44	41	42	45	**59**	**59**	**66**	**66**	**67**	**68**	**50.2**	**58**		**60**	**60**	35	**73**	**51**	43	34	28	28	29	**53**	**53**	47	**59**	**53**	**59**	**52**	51	**56**	40	42	40
	lmm		38	**64**	**57**	43	38	35	37	38	49	47	**52**	50.4	**55**	**55**	41	47	40		**55**	32	**60**	44	37	30	23	23	23	43	42	38	46	42	46	41	40	44	28	31	29
	lq		38	**64**	**53**	43	37	35	37	38	49	47	**51**	49.8	**54**	**55**	41	47	40	45		32	**60**	43	37	30	22	22	22	42	42	39	46	42	46	41	40	43	27	29	28
nearest	GP	Gibbs	**58**	**83**	**68**	**65**	**56**	**56**	**57**	**60**	**70**	**70**	**73**	**73**	**74**	**74**	**65**	**69**	**65**	**68**	**68**		**80**	**67**	**63**	**52**	47	46	49	**68**	**68**	**65**	**71**	**67**	**70**	**67**	**66**	**68**	**59**	**60**	**60**
		LIN	26	**59**	38	32	28	26	28	29	39	37	40	37	44	44	29	37	27	40	40	20		26	21	17	12	12	13	29	29	25	32	28	32	27	26	30	20	22	21
		Mat	40	**75**	**60**	49	42	38	39	42	**58**	**57**	**61**	**60**	**63**	**64**	49.98	**57**	49	**56**	**57**	33	**74**		42	28	24	24	26	**52**	**51**	47	**56**	**51**	**56**	50.2	49	**53**	34	37	37
		NN	45	**81**	**63**	**58**	49	47	48	**51**	**66**	**64**	**69**	**69**	**72**	**73**	**58**	**65**	**57**	**63**	**63**	37	**79**	**58**		38	26	26	29	**61**	**60**	**56**	**65**	**60**	**66**	**59**	**59**	**61**	43	45	44
		Q	**58**	**84**	**73**	**67**	**61**	**56**	**57**	**62**	**71**	**71**	**74**	**74**	**76**	**77**	**66**	**71**	**66**	**70**	**70**	48	**83**	**72**	**62**		40	36	40	**68**	**68**	**65**	**72**	**68**	**72**	**67**	**67**	**69**	**53**	**55**	53
		RQ	**61**	**90**	**76**	**77**	**65**	**70**	**70**	**72**	**80**	**78**	**80**	**81**	**84**	**84**	**72**	**78**	**72**	**77**	**78**	**53**	**88**	**76**	**74**	**60**		**47**	**52**	**75**	**75**	**72**	**78**	**75**	**78**	**74**	**74**	**75**	**63**	**67**	**63**
		SE+Q	**61**	**90**	**76**	**77**	**66**	**69**	**71**	**74**	**79**	**78**	**80**	**81**	**84**	**84**	**73**	**78**	**72**	**77**	**78**	**54**	**88**	**76**	**74**	**64**	**53**		**62**	**75**	**75**	**72**	**78**	**75**	**78**	**74**	**74**	**75**	**64**	**67**	**64**
		SE	**59**	**89**	**75**	**75**	**64**	**67**	**69**	**73**	**79**	**77**	**79**	**80**	**83**	**83**	**71**	**78**	**71**	**77**	**78**	**51**	**87**	**74**	**71**	**60**	48	38		**74**	**73**	**70**	**76**	**73**	**77**	**72**	**72**	**74**	**61**	**65**	**62**
	RF	CART MSE	40	**75**	**57**	47	42	39	40	43	**56**	**56**	**63**	**61**	**64**	**65**	47	**54**	47	**57**	**58**	32	**71**	48	39	32	25	25	26		49	43	**56**	49	**57**	48	47	**53**	37	39	38
		CART RDE	40	**75**	**57**	47	43	39	40	43	**56**	**56**	**63**	**61**	**64**	**65**	48	**55**	47	**58**	**58**	32	**71**	49	40	32	25	25	27	**51**		44	**57**	49.8	**57**	49	47	**53**	38	40	38
		OC1	44	**79**	**60**	**52**	46	43	44	48	**62**	**61**	**67**	**67**	**69**	**70**	**53**	**60**	**53**	**62**	**61**	35	**75**	**53**	44	35	28	28	30	**57**	**56**		**63**	**56**	**63**	**55**	**55**	**59**	41	43	42
		PAIR MSE	37	**72**	**53**	43	39	36	37	40	**52**	**51**	**57**	**55**	**58**	**60**	42	**50.1**	41	**54**	**54**	29	**68**	44	35	28	22	22	24	44	43	37		43	**51**	42	41	47	34	37	35
		PAIR RDE	40	**76**	**57**	47	42	39	40	43	**57**	**56**	**63**	**62**	**65**	**66**	48	**55**	47	**58**	**58**	33	**72**	49	40	32	25	25	27	**51**	50.2	44	**57**		**58**	49	47	**53**	38	40	38
		SCRT MSE	37	**72**	**53**	43	39	36	37	39	**51**	**50.3**	**57**	**55**	**58**	**59**	42	49.7	41	**54**	**54**	30	**68**	44	34	28	22	22	23	43	43	37	49	42		41	40	46	34	36	35
		SCRT RDE	41	**76**	**58**	48	43	40	41	44	**58**	**57**	**64**	**63**	**65**	**66**	49	**56**	48	**59**	**59**	33	**73**	49.8	41	33	26	26	28	**52**	**51**	45	**58**	**51**	**59**		49.5	**55**	39	41	39
		SUPP MSE	43	**78**	**59**	49	44	41	42	46	**59**	**58**	**64**	**64**	**66**	**67**	49.7	**58**	49	**60**	**60**	34	**74**	**51**	41	33	26	26	28	**53**	**53**	45	**59**	**53**	**60**	**50.5**		**55**	40	42	40
		SUPP RDE	39	**74**	**55**	46	41	38	39	42	**54**	**53**	**61**	**59**	**61**	**62**	45	**52**	44	**56**	**57**	32	**70**	47	39	31	25	25	26	47	47	41	**53**	47	**54**	45	45		37	39	38
	lmm		49.97	**83**	**69**	**66**	**55**	**58**	**60**	**62**	**70**	**68**	**70**	**70**	**74**	**75**	**61**	**68**	**60**	**72**	**73**	41	**80**	**66**	**57**	47	37	36	39	**63**	**62**	**59**	**66**	**62**	**66**	**61**	**60**	**63**		**61**	**50.1**
	lq		49	**82**	**67**	**63**	**54**	**55**	**57**	**59**	**67**	**65**	**68**	**68**	**72**	**72**	**59**	**66**	**58**	**69**	**71**	40	**78**	**63**	**55**	45	33	33	35	**61**	**60**	**57**	**63**	**60**	**64**	**59**	**58**	**61**	39		45
knn	lmm		49.9	**83**	**67**	**66**	**56**	**59**	**60**	**63**	**70**	**68**	**70**	**70**	**74**	**75**	**61**	**68**	**60**	**71**	**72**	40	**79**	**63**	**56**	47	37	36	38	**62**	**62**	**58**	**65**	**62**	**65**	**61**	**60**	**62**	49.9	**55**	

Table 9: The p-values of the Kolmogorov-Smirnov (KS) test comparing the equality of probability distributions of individual TSS full feature representatives on all data and on those data on which a particular model setting scored best in MSE. The p-values are after the Holm correction and they are shown only if the KS test rejects the equality of both distributions at the family-wise significance level $\alpha = 0.05$, non-rejecting the equality hypothesis is indicated with —. Zeros indicate p-values below the smallest double precision number.

$\mathcal{M}$	GP								RF									lmm	lq
settings	Gibbs	LIN	Mat	NN	Q	RQ	SE+Q	SE	CART MSE	CART RDE	OC1	PAIR MSE	PAIR RDE	SCRT MSE	SCRT RDE	SUPP MSE	SUPP RDE		
$\varphi_{\texttt{dim}}$	2e-101	—	3.2e-2	1.3e-3	2.9e-5	1.7e-23	2.8e-13	—	1.3e-59	1.9e-16	—	—	2.4e-2	3.9e-2	1.5e-3	1.0e-16	7.4e-3	4e-115	7.8e-46
$\varphi_{\texttt{obs}}(\mathcal{A})$	1.8e-40	—	1.0e-7	1.7e-7	7.0e-29	5.9e-36	1.6e-21	—	2.6e-43	1.7e-7	8.4e-11	—	2.3e-10	1.1e-6	6.0e-15	1.4e-10	—	3.2e-14	3e-259
$\varphi^{\text{CMA}}_{\texttt{evopath\_c\_norm}}$	3.5e-7	3.5e-8	7.9e-37	—	2.6e-69	7.6e-62	9.1e-30	2.1e-10	9.9e-20	3.7e-18	4.5e-4	—	1.2e-8	9.3e-8	7.0e-18	2.2e-10	4.3e-24	1e-216	5e-307
$\varphi^{\text{CMA}}_{\texttt{evopath\_s\_norm}}$	1.3e-88	2.1e-3	9.7e-10	1.5e-25	3.3e-48	6.0e-33	9.2e-17	7.4e-15	5.1e-18	7.0e-13	6.9e-14	2.7e-6	2.6e-5	3.9e-5	1.3e-10	2.1e-14	6.3e-11	1.3e-16	1.1e-13
$\varphi^{\text{CMA}}_{\texttt{restart}}$	7e-105	1.5e-7	9.7e-20	2.3e-61	7e-273	5e-193	3e-137	3.5e-46	3.3e-21	8.2e-18	1.7e-60	—	2.0e-4	4.1e-9	8.7e-98	5.5e-14	5.1e-37	1.9e-12	2.2e-15
$\varphi^{\text{CMA}}_{\texttt{step\_size}}$	0	5.2e-19	1.5e-62	2e-123	1e-267	9e-183	6.1e-90	5.4e-79	5.0e-53	1.5e-57	1.1e-58	2.0e-20	3.4e-48	2.2e-47	2.5e-79	5.6e-46	8.4e-85	1.3e-28	9e-194
$\varphi^{\text{CMA}}_{\texttt{cma\_lik}}(\mathcal{A}^{\top}_{\mathcal{P}})$	1.5e-37	1.8e-2	8.2e-7	2.7e-14	2.9e-30	2.8e-33	3.8e-8	6.6e-14	1.7e-9	7.8e-9	4.3e-8	2.0e-3	2.2e-6	1.6e-5	9.3e-14	9.7e-8	4.5e-12	6.0e-34	1.5e-22
$\varphi^{\text{Dis}}_{\texttt{diff\_median\_02}}(\mathcal{A}^{\top})$	0	2.7e-30	3e-103	3e-159	0	0	4e-160	2e-112	3.9e-94	7e-110	1.1e-91	6.1e-7	8.3e-50	2.3e-38	3.8e-89	5.9e-62	6e-107	3e-103	6e-250
$\varphi^{\text{Dis}}_{\texttt{diff\_median\_10}}(\mathcal{A}^{\top})$	0	1.9e-35	1e-126	9e-173	0	0	3e-166	2e-130	2.8e-89	4e-118	5.0e-93	4.7e-11	1.8e-62	5.6e-53	2.4e-90	1.4e-61	4e-125	2.5e-82	0
$\varphi^{\text{Dis}}_{\texttt{diff\_mean\_05}}(\mathcal{A}^{\top}_{\mathcal{P}})$	0	1.6e-35	5e-132	7e-188	0	0	1e-185	7e-138	6e-115	2e-133	1e-104	3.1e-9	3.0e-68	5.8e-55	7e-102	6.5e-81	5e-129	2.0e-81	0
$\varphi^{\text{Dis}}_{\texttt{diff\_mean\_25}}(\mathcal{A}^{\top}_{\mathcal{P}})$	0	6.2e-40	3e-157	6e-186	0	0	2e-181	4e-142	3e-102	8e-130	7e-102	3.9e-10	4.9e-77	7.2e-60	1.1e-99	1.7e-79	1e-128	1e-103	0
$\varphi^{\text{Lvl}}_{\texttt{lda\_qda\_10}}(\mathcal{A}_{\mathcal{P}})$	9.7e-29	1.7e-5	8.2e-18	3.4e-19	1.6e-22	4.4e-5	7.8e-25	1.5e-4	5.4e-4	5.4e-12	1.7e-10	—	1.0e-11	1.1e-7	1.7e-35	2.5e-3	1.7e-9	2.0e-56	0
$\varphi^{\text{Lvl}}_{\texttt{lda\_qda\_25}}(\mathcal{A}_{\mathcal{P}})$	6.9e-30	5.7e-33	6.2e-31	6.1e-8	3.8e-9	1.2e-12	2.6e-16	2.0e-29	7.2e-14	1.2e-23	3.1e-13	1.5e-6	1.6e-44	9.8e-33	3.9e-27	1.3e-14	1.2e-17	3.8e-14	3e-181
$\varphi^{\text{MM}}_{\texttt{quad\_simple\_cond}}(\mathcal{A}^{\top})$	7.6e-83	1.4e-9	3.5e-25	8.7e-13	2e-145	1.7e-96	2.9e-46	2.3e-34	8.3e-11	1.1e-11	3.9e-16	—	5.2e-5	5.5e-3	7.3e-8	3.0e-7	1.9e-13	7e-235	1e-127

Table 10: The p-values of the Kolmogorov-Smirnov (KS) test comparing the equality of probability distributions of individual TSS full feature representatives on all data and on those data on which a particular model setting scored best in RDE. The p-values are after the Holm correction and they are shown only if the KS test rejects the equality of both distributions at the family-wise significance level $\alpha = 0.05$, non-rejecting the equality hypothesis is indicated with —. Zeros indicate p-values below the smallest double precision number.

$\mathcal{M}$	GP								RF									lmm	lq
settings	Gibbs	LIN	Mat	NN	Q	RQ	SE+Q	SE	CART MSE	CART RDE	OC1	PAIR MSE	PAIR RDE	SCRT MSE	SCRT RDE	SUPP MSE	SUPP RDE		
$\varphi_{\texttt{dim}}$	6e-112	3e-182	4.8e-30	4.8e-18	3.0e-14	1.6e-25	2.5e-14	2.2e-8	—	2.8e-12	1.2e-10	1.1e-17	1.1e-26	1.9e-15	1.5e-22	6.6e-5	1.3e-25	1.5e-17	4.1e-72
$\varphi_{\texttt{obs}}(\mathcal{A})$	2e-286	2e-152	6e-308	5e-103	3e-232	4e-111	8e-175	7e-179	1.5e-17	2.5e-8	1.9e-2	1.4e-4	6.1e-26	4.0e-20	6.4e-5	5.5e-5	7.1e-6	4e-120	4e-280
$\varphi^{\text{CMA}}_{\texttt{evopath\_c\_norm}}$	2.6e-56	0	0	3e-131	6e-295	7e-152	1e-170	7e-140	6.9e-4	9.5e-6	2.2e-3	1.2e-4	—	—	3.9e-6	1.4e-7	1.1e-3	1e-172	8e-309
$\varphi^{\text{CMA}}_{\texttt{evopath\_s\_norm}}$	5.6e-69	8.8e-27	5.5e-4	3.1e-62	9.8e-13	6.1e-25	3.7e-16	9.0e-20	4.3e-11	7.3e-4	3.9e-2	—	1.4e-11	1.7e-5	1.3e-8	9.6e-4	3.3e-7	4.9e-27	1.7e-21
$\varphi^{\text{CMA}}_{\texttt{restart}}$	4.3e-2	2.2e-77	1.7e-59	3.5e-6	3.1e-9	1.5e-7	2.2e-25	6.6e-37	5.1e-31	1.1e-4	3.8e-2	—	2.0e-2	—	3.3e-8	1.0e-4	—	1.7e-54	7e-138
$\varphi^{\text{CMA}}_{\texttt{step\_size}}$	0	3e-234	1e-227	0	0	0	0	0	5.3e-75	2.3e-23	1.8e-4	5.5e-6	1.3e-68	2.7e-48	2.4e-23	5.0e-40	3.3e-17	8.7e-36	4.7e-87
$\varphi^{\text{CMA}}_{\texttt{cma\_lik}}(\mathcal{A}^{\top}_{\mathcal{P}})$	6.3e-65	9.4e-19	1.3e-53	1.4e-36	2.1e-74	1.3e-48	1.3e-46	5.4e-38	—	—	—	—	7.8e-4	5.3e-3	4.2e-4	4.5e-2	5.8e-9	2.1e-46	5.1e-42
$\varphi^{\text{Dis}}_{\texttt{diff\_median\_02}}(\mathcal{A}^{\top})$	0	9e-234	0	0	0	0	0	0	1.2e-88	1.7e-23	1.0e-7	—	2.7e-72	7.0e-57	9.6e-15	1.8e-37	3.3e-15	2e-104	5e-110
$\varphi^{\text{Dis}}_{\texttt{diff\_median\_10}}(\mathcal{A}^{\top})$	0	0	0	0	0	0	0	0	4.8e-88	3.6e-21	1.3e-9	7.6e-6	1.0e-93	3.3e-79	1.8e-13	1.1e-43	1.6e-12	3.8e-76	1e-130
$\varphi^{\text{Dis}}_{\texttt{diff\_mean\_05}}(\mathcal{A}^{\top}_{\mathcal{P}})$	0	0	0	0	0	0	0	0	5.2e-98	4.2e-22	1.2e-9	8.6e-4	1.9e-93	4.7e-71	1.7e-8	3.6e-47	1.1e-8	2.9e-66	1e-128
$\varphi^{\text{Dis}}_{\texttt{diff\_mean\_25}}(\mathcal{A}^{\top}_{\mathcal{P}})$	0	0	0	0	0	0	0	0	1.3e-93	1.3e-20	5.5e-11	6.1e-4	2e-103	2.8e-86	5.5e-9	2.0e-44	3.6e-8	1e-106	5e-201
$\varphi^{\text{Lvl}}_{\texttt{lda\_qda\_10}}(\mathcal{A}_{\mathcal{P}})$	6e-221	1e-268	0	1e-105	0	4e-167	2e-241	8e-223	6.8e-6	5.2e-4	1.7e-2	1.2e-7	3.5e-5	2.5e-9	5.1e-4	1.7e-2	—	4e-129	0
$\varphi^{\text{Lvl}}_{\texttt{lda\_qda\_25}}(\mathcal{A}_{\mathcal{P}})$	1.6e-54	2.0e-40	3.4e-90	7.7e-24	1.3e-74	1.4e-35	6.3e-45	1.0e-51	5.2e-5	1.2e-3	5.2e-10	1.8e-18	3.0e-36	5.1e-33	3.3e-3	7.5e-5	1.1e-7	5.5e-25	4.8e-53
$\varphi^{\text{MM}}_{\texttt{quad\_simple\_cond}}(\mathcal{A}^{\top})$	1e-190	8.5e-68	6e-220	2.1e-64	0	2e-177	3e-180	4e-148	7.1e-5	1.6e-3	—	8.8e-3	4.3e-9	6.0e-6	2.6e-4	2.0e-3	3.5e-2	0	0

Table 11: The p-values of the Kolmogorov-Smirnov (KS) test comparing the equality of probability distributions of individual TSS nearest feature representatives on all data and on those data on which a particular model setting scored best in MSE. The p-values are after the Holm correction and they are shown only if the KS test rejects the equality of both distributions at the family-wise significance level $\alpha = 0.05$, non-rejecting the equality hypothesis is indicated with —. Zeros indicate p-values below the smallest double precision number.

$\mathcal{M}$	GP								RF									lmm	lq
settings	Gibbs	LIN	Mat	NN	Q	RQ	SE+Q	SE	CART MSE	CART RDE	OC1	PAIR MSE	PAIR RDE	SCRT MSE	SCRT RDE	SUPP MSE	SUPP RDE		
$\varphi_{\texttt{dim}}$	0	—	1.9e-7	8.2e-19	3.8e-3	1.8e-9	5.5e-81	7.9e-3	9.7e-30	5.7e-14	2.9e-3	7.2e-3	2.9e-10	4.4e-7	2.5e-13	—	1.5e-2	6.7e-36	1.5e-77
$\varphi_{\texttt{obs}}(\mathcal{T})$	0	5.0e-3	6.1e-6	6.0e-20	6.2e-6	2.2e-9	3.7e-50	1.5e-6	4.5e-35	4.7e-22	2.1e-9	2.0e-5	8.2e-13	2.2e-6	2.3e-24	9.3e-3	3.4e-11	2e-124	2.7e-12
$\varphi^{\text{CMA}}_{\texttt{evopath\_c\_norm}}$	5e-131	6.1e-5	1.9e-46	1.1e-7	7.0e-32	5.2e-18	1.6e-33	1.5e-80	3.5e-26	7.9e-19	3.8e-23	5.5e-12	2.7e-19	2.2e-17	7.4e-19	4.7e-30	9.2e-13	2.6e-61	1e-183
$\varphi^{\text{CMA}}_{\texttt{evopath\_s\_norm}}$	3.2e-25	2.1e-2	1.7e-3	7.9e-10	2.8e-4	2.5e-20	6.2e-21	1.0e-4	3.2e-9	1.1e-5	2.9e-7	1.4e-5	1.6e-7	8.8e-7	1.6e-6	4.6e-6	3.9e-5	1.7e-56	8.3e-7
$\varphi^{\text{CMA}}_{\texttt{restart}}$	2.4e-69	2.5e-4	—	1.7e-2	1.2e-3	2.3e-75	3.8e-42	1.0e-45	2.6e-2	—	3.2e-2	—	3.3e-2	—	—	7.9e-4	2.5e-4	1.1e-42	7.0e-64
$\varphi^{\text{CMA}}_{\texttt{step\_size}}$	8e-311	2.5e-5	2.4e-21	1.3e-4	1.6e-27	0	0	6e-178	6.3e-13	1.6e-14	6.6e-29	3.1e-8	2.1e-16	1.9e-9	9.7e-25	1.7e-34	6.3e-29	2e-136	2.0e-78
$\varphi^{\text{CMA}}_{\texttt{cma\_lik}}(\mathcal{A}^{\top}_{\mathcal{P}})$	8.2e-60	—	6.1e-9	2.0e-4	1.1e-28	3e-182	1e-202	6e-180	1.0e-6	1.2e-4	1.2e-15	2.6e-7	1.9e-5	1.1e-6	1.4e-11	1.6e-10	2.8e-10	5.1e-83	3.7e-22
$\varphi^{\text{Dis}}_{\texttt{diff\_median\_10}}(\mathcal{A}^{\top})$	3e-146	1.9e-10	4.9e-37	7.4e-16	5.2e-57	0	0	0	3.8e-23	1.6e-22	3.8e-38	4.6e-18	5.0e-26	7.3e-23	1.3e-32	1.5e-60	5.9e-45	0	2.7e-14
$\varphi^{\text{Dis}}_{\texttt{diff\_mean\_05}}(\mathcal{A}^{\top}_{\mathcal{P}})$	5e-192	2.4e-10	4.3e-38	4.6e-15	1e-100	0	0	0	6.3e-34	1.6e-25	9.3e-52	9.9e-27	5.7e-31	5.8e-35	5.2e-43	1.7e-73	2.2e-51	0	2.3e-19
$\varphi^{\text{Dis}}_{\texttt{diff\_mean\_25}}(\mathcal{A}^{\top}_{\mathcal{P}})$	4e-197	3.7e-11	6.3e-47	5.8e-18	8e-101	0	0	0	5.8e-31	2.3e-25	5.2e-50	9.3e-26	2.0e-30	5.4e-35	1.3e-40	1.0e-79	8.4e-50	0	9.6e-43
$\varphi^{\text{Dis}}_{\texttt{diff\_median\_02}}(\mathcal{T}^{\top})$	1.2e-26	1.9e-14	5.6e-7	2.0e-33	4.8e-21	2e-208	7e-228	3e-106	—	—	6.3e-5	2.5e-3	—	—	1.6e-4	—	5.0e-9	2.7e-81	2.1e-28
$\varphi^{\text{Lvl}}_{\texttt{lda\_qda\_10}}(\mathcal{A}_{\mathcal{P}})$	4.1e-35	2.2e-2	3.8e-16	1.4e-20	—	2.4e-76	8e-143	6.8e-66	1.8e-4	1.3e-5	3.9e-5	2.0e-7	3.4e-5	2.5e-6	1.2e-2	2.1e-14	6.2e-4	1.3e-21	4e-150
$\varphi^{\text{Lvl}}_{\texttt{lda\_qda\_25}}(\mathcal{T}_{\mathcal{P}})$	9.5e-15	3.2e-11	—	8.0e-14	2.9e-5	1.4e-15	5.7e-23	2.7e-22	1.1e-4	2.6e-5	6.1e-15	8.9e-15	1.6e-4	2.4e-12	9.7e-13	4.5e-16	3.7e-12	9e-123	9.8e-57
$\varphi^{\text{MM}}_{\texttt{quad\_simple\_cond}}(\mathcal{T}^{\top})$	8.8e-96	—	6.8e-6	1.9e-10	3.3e-38	4e-111	2e-118	3.0e-90	2.2e-11	1.7e-8	5.2e-8	5.6e-4	2.7e-9	7.4e-3	9.1e-8	1.3e-4	4.4e-5	2.5e-20	3.6e-3

Table 12: The p-values of the Kolmogorov-Smirnov (KS) test comparing the equality of probability distributions of individual TSS nearest feature representatives on all data and on those data on which a particular model setting scored best in RDE. The p-values are after the Holm correction and they are shown only if the KS test rejects the equality of both distributions at the family-wise significance level $\alpha = 0.05$, non-rejecting the equality hypothesis is indicated with —. Zeros indicate p-values below the smallest double precision number.

$\mathcal{M}$	GP								RF									lmm	lq
settings	Gibbs	LIN	Mat	NN	Q	RQ	SE+Q	SE	CART MSE	CART RDE	OC1	PAIR MSE	PAIR RDE	SCRT MSE	SCRT RDE	SUPP MSE	SUPP RDE		
$\varphi_{\texttt{dim}}$	0	2e-189	1.2e-31	1.5e-28	9.2e-32	2.6e-15	7.0e-10	5.6e-44	9.5e-7	2.3e-10	2.2e-34	2.4e-26	8.3e-11	1.5e-24	1.4e-8	4.3e-59	2.0e-29	7.8e-54	2.7e-7
$\varphi_{\texttt{obs}}(\mathcal{T})$	0	9.3e-50	1.2e-11	4.7e-16	7.0e-54	1.7e-15	1.5e-12	3.1e-48	2.1e-5	8.1e-5	7.5e-20	4.4e-16	7.3e-5	5.3e-16	9.8e-5	6.6e-37	1.0e-15	4.8e-58	1.2e-28
$\varphi^{\text{CMA}}_{\texttt{evopath\_c\_norm}}$	2e-171	0	0	1e-132	6.0e-83	5e-159	4e-202	2e-284	3.3e-2	3.3e-2	2.2e-5	—	3.4e-2	1.5e-2	1.9e-3	4.2e-12	4.9e-5	3e-202	4.5e-13
$\varphi^{\text{CMA}}_{\texttt{evopath\_s\_norm}}$	1.0e-67	6.7e-34	1.0e-17	2.6e-45	4.1e-12	1.1e-9	5.8e-17	7.7e-10	5.3e-3	2.7e-3	1.4e-3	1.3e-3	2.1e-2	6.0e-6	—	7.3e-5	1.5e-4	2.2e-71	1.5e-59
$\varphi^{\text{CMA}}_{\texttt{restart}}$	4.0e-93	2.9e-91	1.9e-83	7.1e-43	1e-146	2e-280	0	2e-322	1.1e-2	—	1.6e-3	3.2e-4	3.5e-4	1.9e-4	4.4e-2	5.7e-7	3.4e-2	2e-196	3e-270
$\varphi^{\text{CMA}}_{\texttt{step\_size}}$	0	1e-236	0	2e-285	7e-234	8.6e-80	3e-139	2.8e-72	1.0e-22	5.3e-22	1.7e-19	3.5e-20	2.5e-17	9.8e-27	2.5e-16	3.9e-22	1.6e-20	2e-191	9e-138
$\varphi^{\text{CMA}}_{\texttt{cma\_lik}}(\mathcal{A}^{\top}_{\mathcal{P}})$	6.9e-64	2.1e-14	1.4e-50	2.5e-31	2.5e-86	6.3e-71	5.2e-95	1.3e-86	4.3e-8	1.9e-8	9.0e-11	6.4e-11	2.1e-8	7.3e-13	1.0e-9	2.0e-11	3.4e-7	5.2e-72	3.4e-12
$\varphi^{\text{Dis}}_{\texttt{diff\_median\_10}}(\mathcal{A}^{\top})$	1e-214	0	0	0	3e-307	0	0	0	1.2e-23	4.5e-24	1.5e-15	3.9e-27	1.2e-15	5.9e-28	3.8e-16	1.5e-15	4.6e-16	0	5e-225
$\varphi^{\text{Dis}}_{\texttt{diff\_mean\_05}}(\mathcal{A}^{\top}_{\mathcal{P}})$	1e-191	0	0	0	0	0	0	0	2.6e-22	8.3e-21	9.9e-20	1.6e-29	3.8e-15	1.9e-29	3.8e-19	9.5e-15	4.9e-14	0	0
$\varphi^{\text{Dis}}_{\texttt{diff\_mean\_25}}(\mathcal{A}^{\top}_{\mathcal{P}})$	6e-197	0	0	0	0	0	0	0	2.6e-22	2.7e-21	1.7e-20	1.3e-30	9.8e-15	9.5e-31	3.6e-19	2.7e-15	5.0e-14	0	2e-317
$\varphi^{\text{Dis}}_{\texttt{diff\_median\_02}}(\mathcal{T}^{\top})$	1.0e-22	8e-132	9.6e-62	2.9e-84	1e-107	5e-262	0	4e-256	—	1.3e-3	—	1.2e-3	1.1e-3	5.3e-3	1.7e-2	3.4e-5	9.7e-6	6e-200	1e-215
$\varphi^{\text{Lvl}}_{\texttt{lda\_qda\_10}}(\mathcal{A}_{\mathcal{P}})$	8.2e-52	3e-266	0	2e-116	9e-130	2.1e-26	1.4e-19	6.8e-31	7.1e-3	2.4e-5	3.5e-3	1.7e-5	8.8e-6	2.8e-4	—	1.7e-3	1.3e-2	1.8e-49	5e-107
$\varphi^{\text{Lvl}}_{\texttt{lda\_qda\_25}}(\mathcal{T}_{\mathcal{P}})$	6.8e-88	8.2e-12	1.0e-22	2.3e-17	1.2e-42	2e-268	9e-231	1e-186	4.0e-6	1.4e-3	5.1e-7	1.3e-10	6.1e-5	7.5e-10	2.7e-9	2.2e-11	2.5e-8	7e-146	8.2e-85
$\varphi^{\text{MM}}_{\texttt{quad\_simple\_cond}}(\mathcal{T}^{\top})$	3.4e-30	3.9e-71	4.2e-27	1.7e-22	3.3e-58	1.1e-46	9.8e-55	4.2e-55	2.4e-4	6.2e-4	—	1.2e-3	3.5e-3	6.8e-4	—	—	—	1.1e-21	6.0e-16

Table 13: The p-values of the Kolmogorov-Smirnov (KS) test comparing the equality of probability distributions of individual TSS knn feature representatives on all data and on those data on which the lmm model setting scored best in MSE and RDE. The p-values are after the Holm correction and they are shown only if the KS test rejects the equality of both distributions at the family-wise significance level $\alpha = 0.05$. Zeros indicate p-values below the smallest double precision number.

	$\varphi_{\mathtt{dim}}$	$\varphi_{\mathtt{obs}}(\mathcal{A}_{\mathcal{P}})$	$\varphi^{\text{CMA}}_{\mathtt{evopath\_s\_norm}}$	$\varphi^{\text{CMA}}_{\mathtt{restart}}$	$\varphi^{\text{CMA}}_{\mathtt{cma\_lik}}(\mathcal{A}^{\top}_{\mathcal{P}})$	$\varphi^{\text{CMA}}_{\mathtt{cma\_lik}}(\mathcal{T}_{\mathcal{P}})$	$\varphi^{\text{Dis}}_{\mathtt{diff\_mean\_10}}(\mathcal{A}^{\top})$	$\varphi^{\text{Dis}}_{\mathtt{diff\_median\_02}}(\mathcal{A}^{\top})$	$\varphi^{\text{Dis}}_{\mathtt{diff\_mean\_10}}(\mathcal{T}^{\top}_{\mathcal{P}})$	$\varphi^{\text{Dis}}_{\mathtt{diff\_mean\_05}}(\mathcal{T}^{\top})$	$\varphi^{\text{Inf}}_{\mathtt{eps\_max}}(\mathcal{T})$	$\varphi^{\text{Lvl}}_{\mathtt{lda\_qda\_10}}(\mathcal{A}_{\mathcal{P}})$	$\varphi^{\text{Lvl}}_{\mathtt{lda\_qda\_25}}(\mathcal{A}_{\mathcal{P}})$	$\varphi^{\text{MM}}_{\mathtt{quad\_simple\_cond}}(\mathcal{T})$
MSE	5.2e-32	9.2e-08	1.6e-63	1.9e-75	1.2e-78	3e-267	0	0	8.3e-126	7.4e-161	0	1.1e-23	2.8e-60	2e-25
RDE	9.5e-45	1.2e-28	2.8e-67	1.4e-196	1.6e-46	0	0	0	2.8e-188	4.9e-209	0	9.2e-14	7.4e-119	6.6e-319

References

Bajer, L., Pitra, Z., Repický, J., and Holeňa, M. (2019). Gaussian process surrogate models for the CMA Evolution Strategy. *Evolutionary Computation*, 27(4):665–697.

Hansen, N. (2006). The CMA evolution strategy: A comparing review. In *Towards a New Evolutionary Computation*, Studies in Fuzziness and Soft Comp., pages 75–102. Springer.

Hansen, N., Auger, A., Mersmann, O., Tusar, T., and Brockhoff, D. (2016). COCO: A platform for comparing continuous optimizers in a black-box setting. arXiv:1603.08785.

Hansen, N. and Ostermeier, A. (1996). Adapting arbitrary normal mutation distributions in evolution strategies: The covariance matrix adaptation. CEC '96., pages 312–317. IEEE.

Kern, S., Hansen, N., and Koumoutsakos, P. (2006). Local Meta-models for Optimization Using Evolution Strategies. PPSN '06, pages 939–948. Springer.

Kerschke, P. (2017a). *Automated and Feature-Based Problem Characterization and Algorithm Selection Through Machine Learning*. Dissertation, University of Münster. Publication status: Published.

Kerschke, P. (2017b). Comprehensive feature-based landscape analysis of continuous and constrained optimization problems using the R-package flacco. *ArXiv e-prints*.

Kerschke, P. and Dagefoerde, J. (2017). *flacco: Feature-Based Landscape Analysis of Continuous and Constraint Optimization Problems*. R-package v. 1.7.

Kerschke, P., Preuss, M., Wessing, S., and Trautmann, H. (2015). Detecting funnel structures by means of exploratory landscape analysis. GECCO '15, pages 265–272.

Lunacek, M. and Whitley, D. (2006). The dispersion metric and the CMA evolution strategy. GECCO '06, pages 477–484.

Mersmann, O., Bischl, B., Trautmann, H., Preuss, M., Weihs, C., and Rudolph, G. (2011). Exploratory landscape analysis. GECCO '11, pages 829–836.

Muñoz, M. A., Kirley, M., and Halgamuge, S. K. (2015). Exploratory landscape analysis of continuous space optimization problems using information content. *IEEE Transactions on Evolutionary Computation*, 19(1):74–87.

Pitra, Z., Bajer, L., and Holeňa, M. (2016). Doubly trained evolution control for the Surrogate CMA-ES. PPSN '16, pages 59–68. Springer.

Pitra, Z., Koza, J., Tumpach, J., and Holeňa, M. (2022). Landscape analysis for surrogate models in the evolutionary black-box context. *arXiv*, 2203.11315:1–34. Under review in journal.

Pitra, Z., Repický, J., and Holeňa, M. (2019). Landscape analysis of Gaussian process surrogates for the covariance matrix adaptation evolution strategy. GECCO '19, pages 691–699.