



Multi-objective Optimization in High-Dimensional Molecular Systems

Debora Slanzi^{1,2}(✉), Valentina Mameli¹, Marina Khoroshiltseva¹,
and Irene Poli^{1,2}

¹ European Centre for Living Technology, S. Marco 2940, 30124 Venice, Italy
{debora.slanzi, valentina.mameli, marina.khoroshiltseva, irenepoli}@unive.it

² Department of Environmental Sciences, Informatics and Statistics,
Ca' Foscari University of Venice, Cannaregio 873, 30121 Venice, Italy

Abstract. The paper proposes a methodological approach to design complex experiments for multi-objective optimization. The strategy is based on evolutionary statistical inference to search for the optimal values in high-dimensional experimental spaces. We developed this approach to study a particular molecular system and discover the best molecules to be proposed as candidate drugs.

Keywords: Multi-objective optimization · Evolutionary strategies
Statistical models

1 Introduction

In the complex process of developing a new drug a major challenge concerns the construction of small molecules that interacting with a pharmacological target of interest can have a therapeutic effect on a particular disease. Current drug design practices involve the screening of large chemical libraries, composed of thousands or millions of compounds, with the aim of identifying a candidate molecule with suitable characteristics, known as a lead molecule. This lead molecule may not fulfil the properties needed to become a drug, such as Absorption, Distribution, Metabolism and Excretion (ADME) [1]. To achieve these properties, while retaining the interaction capacity with the target protein, the molecule must be modified for optimizing a set of variables. A simultaneous optimization is then required and the problem is framed as a multi-objective optimization problem with several conflicting objectives; see for example [2,3]. This field of research in drug discovery has been developed using different approaches, mainly based on the evolutionary principle. Among the most relevant contributions we mention the studies in *De novo Designs* [4], in *Molecular docking* [5] and in *Quantitative structure-activity relationships* [6]. These approaches have been successful in detecting the relevant information for discovering the molecule optimal values using computer based algorithms.

The optimization problem that we address in this study is concerned with a molecular system where each molecule is described by a very large number of features determining the high-dimensionality of the system. In order to discover the optimal molecular structure we intend to develop an evolutionary approach based on statistical models constructed to extract information in high-dimensional systems. In particular, we consider the Lasso model [7], Neural Networks [8], Stepwise Regression and Boosting models [9]. The main purpose of this paper is to develop an efficient approach that is able to find the best candidate molecules, testing a very small number of experimental compositions. This approach is then built on both the evolutionary paradigm and statistical models for high-dimensional experimental spaces. The paper extends the Evolutionary data Design for Optimization (*EDO*) proposed in [10, 11] allowing several objective functions to be optimized simultaneously. The approach, called *m-EDO* (multi-objective Evolutionary data Design for Optimization), drives the evolution towards the target by estimating and combining predictions from different models and different objective functions. We evaluate the method on the molecular library provided by [12] as a test set, investigated by [11, 13] and more recently by [14, 15].

The structure of the paper is as follows. In Sect. 2, we describe the main aspects of a multi-objective optimization problem. In Sect. 3, we briefly introduce the Model-Based Evolutionary Design for Optimization and the statistical models proposed for high-dimensional experimental spaces. In Sect. 4, we present the results achieved with the procedure for the data set provided by [12] and in Sect. 5, we present some concluding remarks.

2 The Multi-objective Optimization Problem

Discovering optimal values in high-dimensional systems can be a very difficult problem, in particular when the number of experimental tests (or observations) is small. Moreover, the optimal values can involve different properties of the system elements, introducing multiple (and possible conflicting) objective functions to be optimized simultaneously. This framing of the problem can make the search of the optimal values pretty hard.

In general, a multi-objective optimization problem can be described in the following way:

Consider a vector valued objective function $f : C \rightarrow \mathbb{R}^k$ with $C \subseteq \mathbb{R}^d$, where d is the dimension of each element of C and $f(c) = (f_1(c), \dots, f_k(c))$; search the element $c_0 \in C$ such that $f(c_0) \leq f(c)$ for all $c \in C$ (minimization) or such that $f(c_0) \geq f(c)$ for all $c \in C$ (maximization).

Frequently, in multi-objective optimization problems, there does not exist a solution, c_0 , which minimizes (or maximizes) all objective functions simultaneously. Therefore, the goal is to identify the Pareto optimal solutions which are the solutions that cannot be improved in any of the objectives without degrading at least one of the other objectives [2, 3].

Formally in a minimization problem, a point $c^* \in C$ is a Pareto optimal solution if for every $c \in C$ and $I = \{1, 2, \dots, k\}$ either,

$$\forall_{i \in I} f_i(c) = f_i(c^*) \quad (2.1)$$

or, there is at least one $i \in I$ such that

$$f_i(c) > f_i(c^*). \quad (2.2)$$

The set of optimal points is called the Pareto optimal set \mathcal{P}^* and the Pareto front \mathcal{F}^* , composed of all the values of the function f at the optimal points, is defined as:

$$\mathcal{F}^* := \{f(c) = (f_1(c), \dots, f_k(c)) \mid c \in \mathcal{P}^*\}. \quad (2.3)$$

In this research we will introduce a methodological approach to address multi-objective optimization in the context above described and related to a molecular system of interest for drug discovery. In particular, we propose a strategy which may aid the process of lead molecule optimisation.

3 Methods

3.1 The Model-Based Evolutionary Design for Optimization

In the drug discovery research field evolutionary algorithms can provide efficient and effective experimental designs; see for instance [16] and references therein. These evolutionary procedures, as described in [10, 11], enable to explore the whole experimental space meanwhile exploiting the capacity of statistical models to uncover relevant information. The evolutionary design for optimization, namely *EDO*-design, randomly generates a first initial population of experimental points. This initial set of experimental points is then tested in laboratory in order to derive the experimental response values. This data set (test composition and responses) are then used to build a class of statistical predictive models. The information gathered from these models is then used to drive the evolution towards the optimal value. In fact, with this information we can identify the next generation of experimental points which evolves from one generation to the next. The process continues until a pre-defined amount of objective-function evaluations is conducted. The *EDO* enables to capture the characteristics of the data and discover the optimum value by testing a very small set of points. This method has been developed for both single and multi-objective optimization.

Multi-objective optimization is based on the idea of guiding the evolution towards the target of the experimentation by building predictive statistical models for the different objective functions and using a linear combination of the best predicted values. The weights in the linear combination can be decided *a priori* with respect to the relevance of the objective functions. In order to distinguish between the single and multi-objective optimization, we indicate just by *EDO* the Evolutionary data Design for single Optimization and by *m-EDO* the Evolutionary data Design for multi-objective Optimization.

3.2 Models for Prediction

The Evolutionary data Design for Optimization, drives the evolution towards the target by estimating and combining predictions with different stochastic models for high dimensional settings, namely Lasso Regression, Stepwise Regression, Boosting, Neural Networks; see [15] and references therein. Herein, we briefly survey some features of these statistical tools.

Modelling data with a multiple linear regression we can write

$$y_i = X_i\beta + \epsilon_i, \quad i = 1, \dots, n \quad (3.1)$$

where y_i is the response variable, $X_i = (x_{i1}, \dots, x_{ip})$ is a p -dimensional vector of predictors (or covariates), $\beta = (\beta_1, \dots, \beta_p)^T$ is the regression vector of p unknown parameters, ϵ_i is the error term, and n is the number of observations. When the number of predictors is much larger than the number of observations ($p \gg n$) estimating regression models can be a very hard task. Under this setup, penalized regression procedures offer powerful methods to simultaneously estimate models and perform variable selection. Among these procedures, the most common is the least absolute shrinkage selection operator (Lasso); see [7].

According to Lasso model we estimate the parameter vector β by minimizing the following Lagrangian objective function

$$Q(\beta) = \frac{1}{2n} \sum_{i=1}^n (y_i - \sum_{j=1}^p \beta_j x_{ij})^2 + \lambda \sum_{j=1}^p |\beta_j|; \quad (3.2)$$

where λ is a tuning parameter which should be assessed by cross validation or information criteria (see [17]). The objective function in Eq. (3.2) is composed of two terms, the former is the least square loss function, the latter is the Lasso penalty, which imposes a constraint on the components of the vector β .

A different strategy for selecting the relevant variables in a regression framework are the step-wise selection methods. These are iterative techniques which allow to identify redundant variables by successively adding or removing variables on the basis of statistical significance criteria. Alternative to the linear regressions the Neural Networks (NN) models can be considered for this structure of data. Neural Networks, in fact, are suitable models for dealing with data characterized by complex non-linear relationships and have become a popular tool for many applications in a wide range of fields including drug discovery research [18]. From a statistical perspective of Neural Network models, we refer the reader to [8, 19]. The topology of a Neural Network can be described as a collection of nodes, namely neurons, which are arranged into ordered layers. A Neural Network usually contain input, hidden and output layers. Considering a Feed Forward NN the input layer communicates information to one or more hidden layers linked to the output layer of this net. With the single hidden layer, the dynamics of the information can be summarized by the following expression

$$y = f(\phi(X, w)) \quad (3.3)$$

where y is the output neuron which can be univariate or multivariate, X represents a set of covariates, w are the weight connections among the neurons directly connected. The function f is called the activation function whose form depends on the problem under consideration. This function can take different forms, which include the linear, the sigmoidal, the logistic or the Gaussian forms. Frequently, the function ϕ is linear and is called the propagation function, it represents the relationship among a neuron and their predecessors. There are several training algorithms for estimating Neural Networks parameters, such as the classical Back-propagation, the scaled conjugate gradient Back-propagation and the Bayesian regularization Back-propagation methods; see [20–22]. To analyse complex structure in the data also the family of boosting algorithms can be considered. Boosting is a class of ensemble techniques that construct multiple estimates or predictions by using combined and averaged estimates or predictions [9]. More formally, the aim of boosting algorithms is to construct or estimate a complex relationship between a set of covariates and a response variable, i.e. $y = f(X)$. This goal will be achieved by minimizing a loss function which measures the discrepancy among y and f by estimating M times a particular model by using weighted fitting and at the end of M iterations we consider a weighted sum of the estimate founds. The specificity of the algorithms is related to different loss functions which depend on the characteristics of the response variable. In our approach we consider the least square loss function as in Lasso framework for estimating a linear regression model; see [9] for an introduction on the boosting algorithm from a statistical perspective.

4 Lead Optimization in a Molecular System

4.1 Data Description

In this research we address the lead optimization of MMP-12 Inhibitors, using the library of molecules made publicly available by [12]. This library consists of 2500 molecules described by the presence of 22272 fragments. In our approach fragments take the role of predictors and are represented as binary variables indicating the presence or absence of each fragment into the molecule. The analysis of these data showed the presence of linear dependence among the predictors (fragments) leading then to a reduction of the total amount of fragments to 4059. Given the high-dimensionality of the system for the very high number of fragments that characterizes each molecule, we adopted the Formal Concept Analysis and reduced the number of fragments to 175; see [14]. Each molecule is then characterized by a set of properties: the pharmacological activity at the target protein; physicochemical properties, such as the solubility; the toxicity property; and structural properties, such as the lipophilicity and the molecular weight. In particular, the pharmacological activity at the target protein, defined as the capacity to produce physiological or chemical effects by the binding of a compound to the therapeutic target, will be denoted by *Activity*. The solubility of a compound is the capacity to dissolve in a liquid and it will be denoted by

Solubility. The toxicity refers to the capacity of any chemicals to produce undesirable effects, and it will be called *Safety*. The lipophilicity is the capacity of the compound to dissolve into lipid structures, and it will be denoted by *cLogP*. The molecular weight relates to the compound size and will be denoted by *MW*. For a more detailed explanation of these biological concepts we refer to the book [1], which describes the influence of each of compound property on ADME and toxicity. Therefore, for this system the experimental response variables (molecular properties) that we considered are: *Activity*; *Solubility*; *Safety*; *cLogP* and *MW*. These response variables represent the target of our optimization study, and some summary statistics of the molecular library made available are reported in the following:

- *Activity*, y_1 : the maximum value is **8**, which corresponds to the optimal value. The **99-th percentile** of the response variable distribution is **7.5**. The target is the **maximization** of y_1 .
- *Solubility*, y_2 : the maximum value is **-1.766**, which corresponds to the optimal value. The **99-th percentile** of the response variable distribution is **-2.415**. The target is the **maximization** of y_2 .
- *Safety*, y_3 : the maximum value is **3.6262**, which corresponds to the optimal value. The **99-th percentile** of the response variable distribution is **3.2309**. The target is the **maximization** of y_3 .
- *cLogP*, y_4 : the minimum value is **-2.505**, which corresponds to the optimal value. The **1-th percentile** of the response variable distribution is **0.033**. The target is the **minimization** of y_4 .
- *MW*, y_5 : the minimum value is **291.3**, which corresponds to the optimal value. The **1-th percentile** of the response variable distribution is **339.3**. The target is the **minimization** of y_5 .

Figure 1 depicts the box-plot of the distribution of the five properties that characterize the set of 2500 molecule, and the blue stars represent the optimum value of each response variable.

The aim of this study is to develop a multi-objective optimization procedure based on experimental data (no simulation), and involving a very small number of experimental tests, to avoid unnecessary waste of research resources. Knowing the whole experimental space (complete library) allowed us to evaluate the performance of the approach in searching the best response values repeating the procedure 1000 times. In order to obtain drug candidates with suitable properties, some constraints are imposed on the molecular properties. In particular, we consider molecules with: *Activity* values $y_1 > 6$, *Solubility* values $y_2 > -3$, *Safety* values $y_3 > 2.57$, *cLogP* values $y_4 < 3$ and *MW* values $y_5 < 450$. Then the goal of the multi-objective optimization is to discover the molecules (three molecules in the library) that satisfy the constraints of the problem and reach their best response values. These molecules are described (in red) in Fig. 2, and represent the molecules belonging to the Pareto Front of *Solubility* and *Safety* with the constraints above described on the other properties. Moreover, the dashed lines represent the constraints values on *Solubility* and *Safety*, respectively. The response values of the three best molecules, goal of our study, are

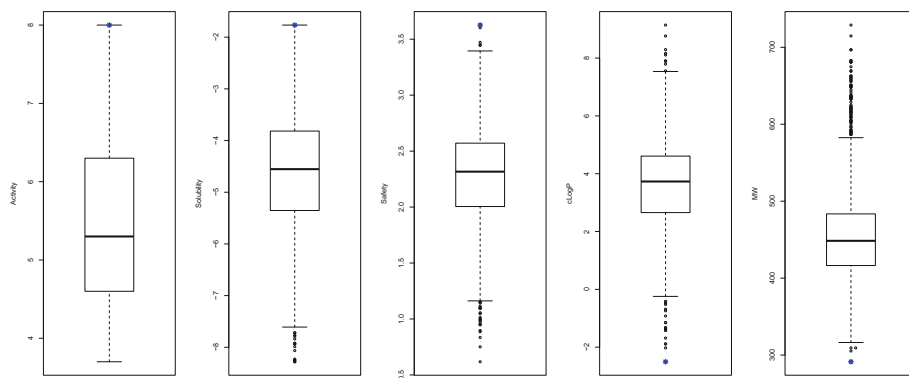


Fig. 1. Box-plot of the response values distribution, from left to right *Activity*, *Solubility*, *Safety*, *cLogP*, and *MW*. (Color figure online)

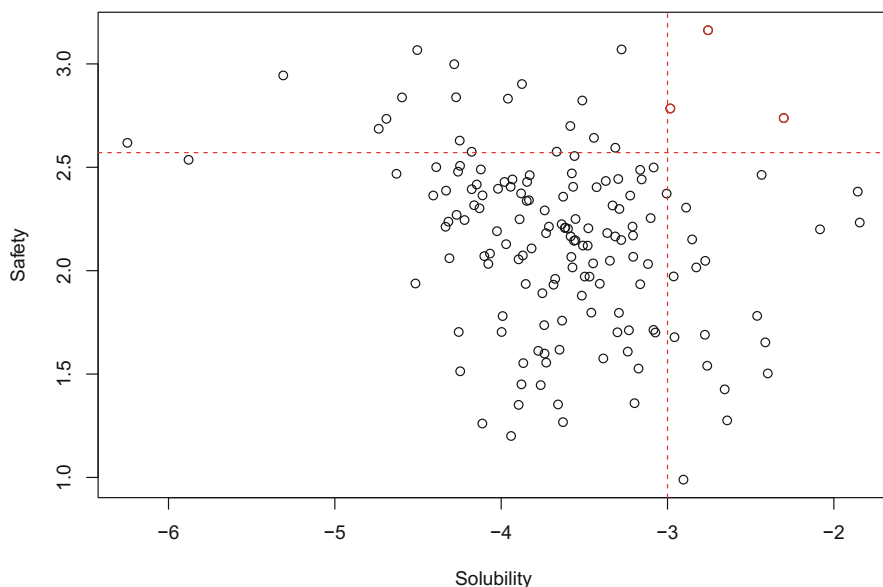


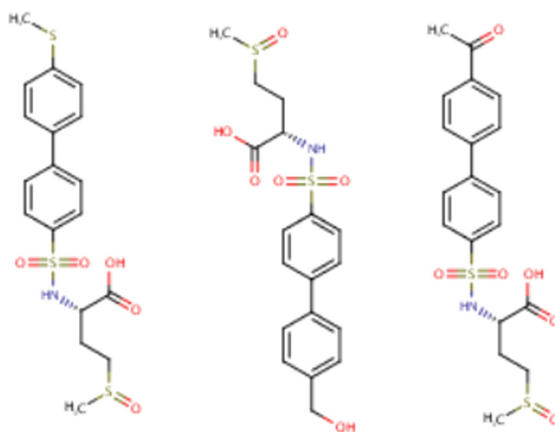
Fig. 2. The molecule values of *Solubility* and *Safety*, respecting the defined constraints for *Activity*, *cLogP*, and *MW* and in red the molecules belonging to the Pareto Front. The dashed lines represent the constraints values on *Solubility* and *Safety*, respectively. (Color figure online)

presented in Table 1. In this study we would like to discover these three molecules by conducting the minimum possible set of experimental tests.

The chemical representation of these molecules as reported in Fig.3 has been obtained by using the **SwissADME** web tool freely available at <http://www.swissadme.ch/>; see [23].

Table 1. Values of the five response variables assumed by the three best molecules.

	<i>Activity</i>	<i>Solubility</i>	<i>Safety</i>	<i>cLogP</i>	<i>MW</i>
Molecule 1	6.90	-2.98	2.78	1.39	427.56
Molecule 2	6.20	-2.76	3.16	0.81	425.48
Molecule 3	6.50	-2.30	2.74	0.38	423.50

**Fig. 3.** Chemical structure of the three target molecules.

4.2 The Evolutionary Procedure and the Optimization Results

At each generation the population size of the evolutionary algorithm is 20 and the algorithm is run for 7 generations, with the constraint that all individuals tested must be different. In this study the maximum number of tests considered is 140 on the total of 2500 candidate compositions. The process is iteratively repeated, generation after generation, maintaining the same size in each population of experimental points and ends when the maximum total number of experimental points is reached. The structure of evolutionary approach consists in randomly selecting an initial small population, in this study 20 molecules, and then evaluating the response variables values of each molecule. In the evolutionary algorithm, the next population of experiments is then built by selecting the 20 molecules with the best predicted response values.

At first, to better understand the performance of the approach, we developed the procedure of single objective optimization for each response variable. The evolution has been driven by the information achieved with the Lasso model, Stepwise regression, Boosting, Neural Networks, and the mixture of these three linear models (hereafter Mixture of linear Models) [14]. The architecture of the Neural Network used for this single-objective problem consists of 175 input neurons, one hidden layer with 7 neurons and one output layer with one neuron.

The activation function is the sigmoidal function, and the Neural Network has been fitted by using Bayesian regularization Back propagation; see [22].

We study the robustness of the procedure with respect to the change of the initial population by repeating the entire process 1000 times. The good performance of the procedure is evaluated in its capacity to uncover the optimum value and the set of values in the region of optimality (the 1% best values of the distribution), conducting a very limited set of experimental tests (5.6% of the whole experimental space). The results achieved for the single optimization process are represented in Table 2.

Table 2. Single objective optimization: number of runs (out of 1000 runs) in which *EDO* uncovers the optimum value and values in the region of optimality (1% best values of the distribution).

Objective		Lasso	Stepwise	Boosting	Mixture of linear models	NN
<i>Activity</i>	Optimum	844	782	665	916	660
	Region of optimality	1000	995	998	1000	990
<i>Solubility</i>	Optimum	875	745	872	912	556
	Region of optimality	995	998	1000	1000	996
<i>Safety</i>	Optimum	387	358	278	467	228
	Region of optimality	1000	1000	1000	1000	999
<i>cLogP</i>	Optimum	848	821	917	918	760
	Region of optimality	950	946	981	1000	945
<i>MW</i>	Optimum	738	822	751	887	346
	Region of optimality	905	966	956	1000	780

From these results we can learn that *EDO* procedure is able to achieve the best response values in a very high proportion of the 1000 runs, showing also a better performance of the Mixture of linear Models with respect to the single model optimization. Concerning the response values in the region of optimality (1% best values of the distribution) we observe that the Mixture of linear Models is able to achieve these values in all the 1000 runs and for all the variables.

We then address the problem of the multi-objective optimization by extending the *EDO* approach which involves the evolution driven by the information achieved with the Lasso model, Neural Network, and the Mixture of the Lasso and the Neural Network models [14]. The architecture of the Neural Network in the multi-objective optimization has the same topology proposed for the single-objective optimization except that the output layer consists of 3 neurons. In fact, we consider 3 response variables *Activity*, *Solubility*, *Safety*. The variables *cLogP* and *MW* are not taken into consideration in the multi-objective procedure because the *cLogP* is highly correlated with the *Solubility*, and the *MW* could be easily predicted on the basis of its amino acids composition; see [24]. In particular for the multi-objective procedure, we selected in a random way 20

Table 3. Multi-objective optimization: number of runs (out of 1000 runs) in which *m*-EDO uncovers the best molecules.

Number of best molecules	Lasso	NN	Mixture of models
0	130	161	92
1	43	59	51
2	320	288	384
3	507	492	473
At least one	870	839	908

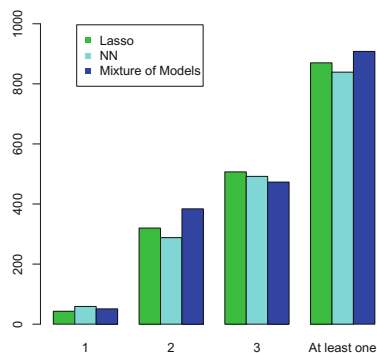


Fig. 4. Multi-objective optimization: best molecules found in 1000 runs.

experimental points from the whole population of 2500 molecules, then we built the model on these 20 data and predict the response variables values by using an estimated model, as in the single optimization procedure. We then associated a weight at each objective value and derived the linear combination of objectives. The molecules with the best estimated linear combination of the objective values are then selected for the next generation of experimental points. In the following table (Table 3) we present the results for the multi-objective optimization achieved with the Lasso model, the Neural Network model (NN) and the Mixture of Lasso and Neural Network models (Mixture of Models) and in Fig. 4 we depict these results. The three ways chosen to optimize give similar results in discovering the three best molecules. We notice that Mixture of models outperforms the alternatives in discovering at least one molecule of the three in more than 90% of 1000 runs.

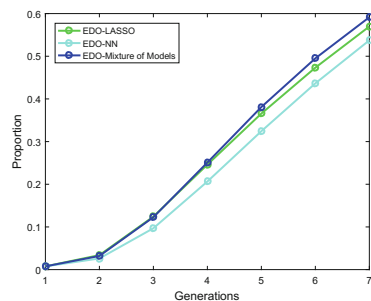
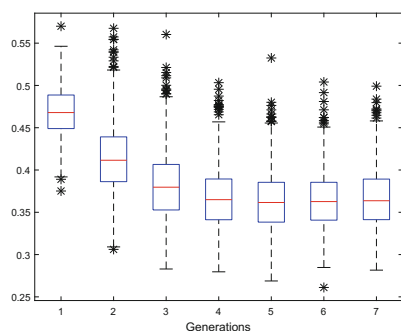


Fig. 5. Evolution through generations. Left: box-plot of the mean of the results achieved in 1000 runs by using the Mixture of Models. Right: proportion (average on 1000 runs) of the results found in the 1% best values of the distribution.

From the generation results, as presented in Fig. 5, we notice the relevance of the evolutionary principle in the search process: from the first generation there is a clear tendency of the procedure to converge towards the optimal value. From the left-hand panel of Fig. 5, we can notice that the mean of the objective response value is decreasing in each generation and get closer to the optimal solution which is identified by the value 0. In fact, we transform our variable values to lie in the interval $[0, 1]$ and the multi-objective optimization becomes just a minimization of the linear combination of the objective response values. Moreover, we also notice in Fig. 5 that the median of the distribution decreases generation after generation and becomes stable after the fourth generation. From the right-hand panel of Fig. 5, we can notice that the proportion of the new objective response values, i.e. the linear combination of the objective response values, found in the region of optimality (1% best values of the distribution) increases generation after generation.

5 Concluding Remarks

The purpose of this research was the development of a methodological strategy able to address the multi-objective optimization problem for complex experimentation conducting a very small number of tests. The procedure proposed is a model-based evolutionary strategy for designing experiments, which involves the construction and the estimation of predictive linear and non-linear models. The study of a particular molecular system for drug discovery problems shows the very good performance of the approach that we propose. Moreover, we would like to stress that we achieve these results by conducting an extremely small number of generations (7 generations) that usually is regarded too small to even approach convergence of the algorithm.

Acknowledgements. The authors would like to acknowledge the fruitful collaboration with Darren Green and his Molecular Design group at GlaxoSmithKline (GSK), Medicines Research Centre, Stevenage (UK).

References

1. Kerns, E.H., Di, L.: *Drug-Like Properties: Concepts, Structure Design and Methods: From ADME to Toxicity Optimization*. Academic Press, San Diego (2008)
2. Coello, C.A., Lamont, G.B., Van Veldhuizen, D.A.: *Evolutionary Algorithms for Solving Multi-Objective Problems*. Genetic and Evolutionary Computation. Springer, New York (2006). <https://doi.org/10.1007/978-0-387-36797-2>
3. Lobato, F.S., Steffen, V.: *Multi-objective Optimization Problem*. SpringerBriefs in Mathematics. Springer International Publishing, Cham (2017). <https://doi.org/10.1007/978-3-319-58565-9>
4. Ekins, S., et al.: Evolving molecules using multi-objective optimization: applying to ADME/Tox. *Drug Discov. Today* **15**, 410–451 (2010)
5. Li, H., et al.: An effective docking strategy for virtual screening based on multi-objective optimization algorithm. *BMC Bioinform.* **10**(58), 1–12 (2009)

6. Soto, A.J., et al.: Multi-objective feature selection in QSAR using a machine learning approach. *QSAR Comb. Sci.* **28**, 1509–1523 (2009)
7. Tibshirani, R.: Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **58**(1), 267–288 (1996)
8. Cheng, B., Titterton, D.M.: Neural networks: a review from a statistical perspective. *Stat. Sci.* **9**(1), 2–30 (1994)
9. Bühlmann, P., Hothorn, T.: Boosting algorithms: regularization, prediction and model fitting. *Stat. Sci.* **22**(4), 477–505 (2007)
10. Baragona, R., Battaglia, F., Poli, I.: *Evolutionary Statistical Procedures: An Evolutionary Computation Approach to Statistical Procedures Designs and Applications*. Springer, Heidelberg (2013). <https://doi.org/10.1007/978-3-642-16218-3>
11. Borrotti, M., De March, D., Slanzi, D., Poli, I.: Designing lead optimization of MMP-12 inhibitors. *Comput. Math. Methods Med.* **2014**, 8 (2014). Article ID 258627
12. Pickett, S.D., Green, D.V.S., Hunt, D.L., Pardoe, D.A., Hughes, I.: Automated lead optimization of MMP-12 inhibitors using a genetic algorithm. *ACS Med. Chem. Lett.* **2**(1), 28–33 (2011)
13. Brown, P.J., Ridout, M.S.: Level-screening designs for factors with many levels. *Ann. Appl. Stat.* **10**(2), 864–883 (2016)
14. Giovannelli, A., Slanzi, D., Khoroshiltseva, M., Poli, I.: Model-based lead molecule design. In: Rossi, F., Piotto, S., Concilio, S. (eds.) *WIVACE 2016*. CCIS, vol. 708, pp. 103–113. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-57711-1_9
15. Mameli, V., Lunardon, N., Khoroshiltseva, M., Slanzi, D., Poli, I.: Reducing dimensionality in molecular systems: a Bayesian non-parametric approach. In: Rossi, F., Piotto, S., Concilio, S. (eds.) *WIVACE 2016*. CCIS, vol. 708, pp. 114–125. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-57711-1_10
16. Lameijer, E.-W., Bäck, T., Kok, J.N., Ijzerman, A.D.P.: Evolutionary algorithms in drug design. *Nat. Comput.* **4**(3), 177–243 (2005)
17. Akaike, H.: Information theory and an extension of the maximum likelihood principle. In: Petrov, B.N., Csaki, F. (eds.) *2nd International Symposium on Information Theory*, pp. 267–281. Akademiai Kiado, Budapest (1973)
18. Livingstone, D.: *Artificial Neural Networks: Methods and Applications*. Humana Press, Totowa (2008)
19. Poli, I., Jones, R.D.: A neural net model for prediction. *J. Am. Stat. Assoc.* **89**(425), 117–121 (1994)
20. Foresee, F.D., Hagan, M.T.: Gauss-Newton approximation to Bayesian learning. In: *Proceedings of the International Joint Conference on Neural Networks* (1997)
21. Moller, M.F.: A scaled conjugate gradient algorithm for fast supervised learning. *Neural Netw.* **6**, 525–533 (1993)
22. MacKay, D.J.C.: Bayesian interpolation. *Neural Comput.* **4**(3), 415–447 (1992)
23. Daina, A., Michielin, O., Zoete, V.: SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci. Rep.* **7**, Article No. 42717 (2017)
24. Guan, Y., Zhu, Q., Huang, D., Zhao, S., Jan, L., Peng, J.: An equation to estimate the difference between theoretically predicted and SDS PAGE-displayed molecular weights for an acidic peptide. *Sci. Rep.* **5**, Article No. 13370 (2015)