

Optimal experimental design for model selection: A partial review

López-Fidalgo J.¹ and Tommasi Ch.²

Abstract Model selection is a core topic in modern Statistics. This is a review of what has been researched on optimal experimental design for model selection. The aim is to find good designs for increasing the test power for discriminating between rival models. This topic has a special impact nowadays in the area of experimental design.

1 Introduction

Model selection is nowadays one of the hot topics in Statistics and finding optimal experimental designs for that purpose will save time, money and risk to the researchers. A joint solution to the problem of identifying the maximum information both for discriminating between rival models and for fitting the best of them is very much desirable. Obtaining a good model is crucial for prediction in a world of presence of massive data. Optimal Experimental Design (OED) theory and practice has a lot to say about that. The reason for that belief is that OED is very much concerned with the information behind the data. Traditionally this has been used for designing informative experiments in order to save data. This is still very important if we think, e.g. in experimentation with animals or humans. In the contemporary scenario with great quantities of observations, frequently with low quality data, we need tools to clean the data and discriminate which is the best model to extract the information from it.

OED searches both for better estimates and predictions as well as for optimizing the cost of experimenting. The research on OED has experienced a noteworthy increase in the last few years. Main beneficiaries of the theory

¹Universidad de Navarra, ICS, Unidad de Estadística, Campus Universitario, 31080, Pamplona, Spain fidalgo@unav.es · ²Department of Economics, Business and Statistics, University of Milano, Via Conservatorio 7, 20122-Milano, Italy chiara.tommasi@unimi.it

of optimal design are the areas of engineering, well-being in society and very much in health with clinical trials and Phase I/II experimentation.

We dare to say there are two important research directions in OED at this moment. On the one hand the most traditional interest in outlining correctly complex, but real, situations. On the other hand there is special interest nowadays in developing efficient algorithms for finding optimal designs, which is a non-standard task in this area where the aim is a probability measure in a non-Euclidean space. Both directions must go together, complementing each other, in the research.

Whenever a variable is under the control of the experimenter OED has a lot to say. The main idea is to search for the most informative experimental conditions with the optimal number of replicates at each of these experimental conditions. The sample size has to be fixed in advance and this is frequently a difficult task to be solved, frequently needing number theory. A wonderful idea was introduced by Kiefer [21] extending the concept of experimental design to any probability measure on a compact design space. Then the so called General Equivalence Theorem (GET) [22, 40] provided a fantastic tool for checking whether a particular design is optimal or not. Additionally this gives the way of computing either closed-form optimal designs or developing numerical algorithms for their computation. What one considers optimal here depends very much on the aim, say either estimating the parameters, predicting results or discriminating between models; as well as on the model itself.

Most of the optimality criteria focus on the inverse of the Fisher Information Matrix (FIM). For linear models this is proportional to the covariance matrix of the estimates of the parameters. The usual inferences and predictions are based on this matrix, so some appropriate function of it has to be optimized. For a linear model this matrix depends only on the design and thus the optimization challenge is just to look for the best design according to the corresponding criterion. If the model is nonlinear the inverse of the FIM is asymptotically proportional to the covariance matrix, but the FIM now depends on the unknown, and no yet, estimated parameters at the time of planning the experimentation. There are different ways to approach this issue,

Locally optimal designs [15], assuming some nominal values of the parameters where the inverse of the FIM is locally approximated to the covariance function.

Adaptive sequential designs, where the next design point takes into account the observed data from previous experiments. In this case the design is a stochastic process. This approach is rather popular, especially in medicine and pharmacological studies with clinical trials.

Bayesian optimal designs [36], assuming a prior distribution on the parameters and a joint utility function including both objectives at the same time, estimating the parameters of the model and finding the optimal design for that.

Another important difficulty in OED arises for correlated observations, e.g. over time or space. The world of industry is a champion on the used of experimental designs. Computational aspects of OED have always been an important issue. Note the contribution of Fisher [20], which gives to the experimental design theory the current statistical approach.

Although the OED appears in 1918 with an extensive article by Kristine Smith [32], it will not be developed at some extent until the 50's. The volume by Cox (1958) [16] contains a short and descriptive introduction to the basic ideas of the theory. Pioneering monographs were written by Fedorov [19] and Silvey [31]. A number of monographs have arrived later and can not be mentioned here for space reasons.

The most significant advances in theory and practice of Optimal Design of Experiments can be followed through the volumes of conference proceedings mODa (Model-Oriented Data Analysis and Design) apart from other relevant collections coming from workshops and conferences. The family of algorithms traditionally used in the calculation of optimal designs are not, in a general sense, the best possible. Although there is an important effort in this aspect it remains as an important challenge for the theory.

2 Different approaches to the problem

One of the main criticisms to OED is that a design has to be found for a particular model and the model has to be guessed without having the data yet. Honestly speaking we should say there are not solutions to this issue, which in fact is still present even with the data at hand. George Box used to say that "Models, of course, are never true, but fortunately it is only necessary that they be useful", e.g. [13]. On the other hand it is quite common that many models come from the experience, retrospective data or intuitions of the practitioner. Sometimes they are analytically derived, e.g. as a solution of differential equations, as happens with the majority of pharmacokinetics models [25, 30, 29]. But, this issue is more severe in Experimental Design. In a variety of situations before having the data two or more models may be potential candidates. Once the data are collected a model has to be chosen after a model selection procedure, then in a second step it has to be fitted. Thus, the optimal design in this case faces two different objectives. On the one hand, it has to be good for discriminating between models. This means to organize the experiment in such a way some distances between the fitted models from the data are as large as possible in order to make clear the differences between them [11, 12, 23, 7, 36, 14, 37]. Appropriate distances need to be chosen according to the existing statistical tests for discriminating [17]. On the other hand the design has to be good for inferences with the chosen model, either for estimating the parameters, estimating some functions of them or making predictions. Even for linear model this is not trivial at all.

There are different approaches to this issue with some controversial and therefore some unclear features of them. Optimal designs from different reasonable perspectives may be rather different even for simple models. There is the need of investigating this issue providing clear directions to model selection. One of the first attempts to tackle the discrimination paradigm from an OED point of view consisted in embedding two (or more) rival models in a more general model and designing to estimate the additional parameters [5, 10]. This is the so called D_s -optimal design, which makes much sense from an intuitive point of view. But it is not clear that this criterion increases the power of the usual discrimination tests. As a matter of fact [24] have proved that T- and D_s -optimality coincide only in the case that the optimal values of the parameters in the T-criterion are the nominal values for the D_s -criterion, which is something very artificial and may be far from reality. Otherwise, the designs could be rather different. There is nothing definitively proved about its relationship with the test power of the likelihood ratio test or any other discrimination test. [28] considered three different approaches for discriminating between models: i) augmenting a given design in an optimal way, ii) evaluating a mixture of the various criteria, and iii) optimizing an objective subject to achieving a prescribed efficiency level for the others.

Another intuitive idea is maximizing the distance between two models assuming one of them is considered as the “true” model, which is the model in the alternative hypothesis (T-optimality) [11, 12, 23]. It happens that this focuses on maximizing the non-centrality parameter of the likelihood ratio test statistic, which is a function of the test power. The traditional F discrimination test is a particular case for nested linear models. This criterion was extended to generalized linear models (GLM) [27] as well as heteroscedasticity and multiple response [38]. As a definitive extension of T-optimality [23] gave a criterion based on the Kullback-Leibler distance, KL-optimality, which accounts for the likelihood ratio test power, which is also related to the AIC. In particular, assuming just two rival models with pdf's

$$f_i(y, x, \theta_i), \quad i = 1, 2,$$

and assuming f_t is the “true” model for either $t = 1$ or $t = 2$, the Kullback-Leibler distance between them is

$$\mathcal{I} [f_t(y, x, \theta_t), f_i(y, x, \theta_i)] = \int f_t(y, x, \theta_t) \log \left[\frac{f_t(y, x, \theta_t)}{f_i(y, x, \theta_i)} \right] dy,$$

where $i \neq t$, y is the vector of responses, θ_i and θ_t are the parameters in the two models and x is the vector of experimental conditions at which the response y is observed.

Then, KL-optimality is defined by the following objective function,

$$I_{i,t}(\xi) = \min_{\theta_i \in \Omega_i} \int_{\mathcal{X}} \mathcal{I} [f_t(y, x, \theta_t), f_i(y, x, \theta_i)] \xi(dx).$$

The relationship with the celebrated AIC criterion is illustrated as follows. Let $L_i(\theta_i)$ and $L_t(\theta_t)$ be the log-likelihoods of each model. It is assumed that θ_t^* is known. Then

$$\min_{\theta_i} \mathcal{I}[f_t(y, x, \theta_t^*), f_i(y, x, \theta_i)] = E_t[L_t(\theta_t^*)] - \max_{\theta_i} E_t[L_i(\theta_i)],$$

where E_t stands for the expectation according to distribution given by f_t . The expected Akaike criterion for model i is

$$E_t[AIC_i] = 2\{m_i - E_t[L_i(\hat{\theta}_i)]\},$$

where m_i is the number of parameters of model i and $\hat{\theta}_i$ is the Maximum Likelihood Estimator (MLE) of θ_i ,

$$\hat{\theta}_i = \arg \max_{\theta_i} L_i(\theta_i).$$

Thus, if the minimum can be exchanged with the expectation then KL-optimality also minimizes the AIC of model i .

Another way to see this relationship is using the so called relative AIC of model i with respect to model t ,

$$\exp \left[\frac{AIC_t - AIC_i}{2} \right] = \exp \left[\log \frac{L_i(\hat{\theta}_i)}{L_t(\theta_t^*)} - m_i \right],$$

where the log-likelihood ratio appears in the right hand side.

[23] have proved also that T-optimality and all the mentioned existing extensions are particular cases of KL-optimality. This means the discrimination can be considered among non-Normal models or even for correlated observations [2, 9]. The criterion has been generalized in different ways for more than two rival models, essentially assuming convex combinations of the efficiencies for several models [34]. Following this idea [37] considered a max-min criterion and provided a couple of suggesting examples with different probability distributions or different mean in GLMs. Compound criteria with D-optimality have been used to search for good designs also for fitting the model [35]. The sequential [26], Bayesian [36] and Copula [1] perspectives have also been taken into account. The computational issue still needs a lot of work [18].

Summarizing all this, Model Selection is a major topic in contemporary Statistics and the OED perspective can provide a significant improvement to this problem. Finding a joint solution to the problem of identifying the maximum information, both for discriminating between rival models and for fitting the best them, is still a challenging topic.

3 Hot areas for further research

In what follows some ideas for potential further work are presented.

3.1 Robustness of KL-optimal designs from different points of view

It would be interesting to check which features in the rival models seem more sensitive for a successful discrimination using the KL-criterion. In particular, it would be useful to check how this criterion compared to others is able to detect differences between rival models in the probability distribution, the mean function, the variance structure or the dimension of the model, e.g. more variables in the mean function. Analytic results would be very much welcomed, but simulations studies will be there in any case. For generalized linear models the link function must be added to this list. Atkinson [6] worked out all the details for making inference on an extended model that includes several submodels as GLMs.

3.2 Other divergence measures

Once a deep knowledge of what the Kullback-Leibler divergence does, better solutions may be looked for. [23] proved that KL-optimality accounts for maximizing the power of the likelihood ratio test, but there are other statistical tests frequently used in practice [17, 33, 39]. The test power in these cases is likely to be connected with other divergence measures [8]. In particular, there is the need to check for the meaning of the D_s -optimality of the differing parameters from nested models. At first sight it seems that for nested models D_s - and KL-optimality should be the same. This is true just in the very particular case of linear models on the parameters when the nested model differs from the root model in just one parameter. In this case the KL-optimal design is the D_s -optimal design for estimating that parameter. For other models the optimal designs may be very different with very low relative efficiencies with respect to each other [24]. For other discrimination tests appropriate criteria must be found, taking into account other information divergence measures, such as those derived from the Rényi entropy of f-divergences.

3.3 Bayesian paradigm

As mentioned above most of the optimality criteria focus on the inverse of the information matrix. If the model is nonlinear the FIM depends on the parameters of the model. KL-optimality, and so T-optimality, does not focus on this matrix, but directly on the probability measure (the design). Nevertheless, even for linear models nominal values of the parameters of the “true” model are needed. One way for approaching this issue is the use of the Bayesian paradigm assuming a prior distribution on the parameters and a joint utility function including both objectives at the same time, estimating the parameters of the model and finding the optimal design for that. This theory has been considered in the literature as an important and necessary approach. The Kullback-Leibler divergence jointly with the Shannon information is used, in a different way than for discrimination, to develop a Bayesian D-optimality criterion. [36] introduced the Bayesian approach for discriminating between two rival models. Following-up this idea a systematization of the Bayesian theory for discriminating between models jointly with fitting purposes would be welcomed. Additionally, this will avoid the annoying assumption of the “true model”.

Utility functions focused on discrimination must be considered. This is very much related again to looking for different divergence measures. Some but not all of the classic optimality criteria have a utility based Bayesian version. Mixing utility functions may help to describe several simultaneous goals. Additionally, [37] considered a max-min criterion for more than two rival models and gave a relationship with a Bayesian criterion assuming a prior distribution of the weights of each model. This particular point may be explored using mathematical programming techniques applied here. This would be very useful for computational purposes. Max-min criteria are not easy to deal with because of the lack of differentiability. The GET is still applicable but an annoying auxiliary probability measure has to be found.

Summarizing these ideas, the Bayesian paradigm applied here brings mainly three results: i) a convincing way of dealing with the unknown parameters, ii) better justified criteria not supported on the artificial assumption of a “true” model and iii) efficient computational techniques for more than two rival models.

3.4 Correlated observations

The Big Data world is a scenario of correlations, which needs to be considered from different perspectives. In contrast with other criteria for discrimination, KL-optimality is still valid in this situation. [9] proved that a standard generalization of T-optimality to correlated observations can be done just when the covariance matrix of the observations is assumed completely known. Designs

for models with a partially unknown covariance structure have been widely studied (see e.g. [4]). Most of this work has been done for D-optimality. The approximation of the covariance matrix by the inverse of the FIM holds in this situation under some assumptions. If these conditions are not satisfied we have usually performed simulations with the designs obtained in the last steps of the algorithm in order to check the monotonicity between the determinants of the FIM and the covariance matrix. KL-optimality is not based on the FIM. Therefore, there is no problem with the mentioned approximation.

A lot of work can be done on the area of optimal designs for discrimination in presence of correlated observations. In particular, the usual time series models require a discrimination process to select the best model, e.g. the best values of p , d and q in an ARIMA(p, d, q) model. This is an area where things are not so simple from an experimental design point of view. [3] considered a rather simple time series model where the implicit covariance structure is worked out from the model. This can be done analytically in very simple models, but it needs some new results to be able to find a proper criterion both for discriminating and for estimating the models.

[4] provided theoretical results for a function to be definite positive. These results are then used to generate potential covariance structures from Bernstein polynomials. These results provide a powerful tool for spatio-temporal modeling, which is not trivial since the covariance structure needs to be such that the generated covariance matrix is nondefinite negative. Another interesting point would be a deep study of the covariance structure behind the usual time series models in order to look for appropriate optimal designs.

4 Computing optimal designs

While iterative procedures are very much needed for OED in general, they are specially needed for finding optimal information for discriminating between models. For KL-optimality some classical algorithms have been adapted, but much more work has to be done here. [23] provided a general algorithm based on the directional derivative,

(i) For a given design ξ_s let

$$\begin{aligned}\theta_{i,s} &= \arg \min_{\theta_i \in \Omega_i} \int \mathcal{I}(f_t, f_i, x, \theta_i) \xi_s(dx) \\ x_s &= \arg \max_{x \in \mathcal{X}} \mathcal{I}(f_t, f_i, x, \theta_{i,s}).\end{aligned}$$

(ii) For a chosen α_s with $0 \leq \alpha_s \leq 1$ let

$$\xi_{s+1} = (1 - \alpha_s) \xi_s + \alpha_s \xi_{x_s},$$

where ξ_{x_s} is a design with measure concentrated at the single point x_s .

Typical conditions for the sequences $\{\alpha_s\}$ are

$$\lim_{s \rightarrow \infty} \alpha_s = 0, \quad \sum_{s=0}^{\infty} \alpha_s = \infty, \quad \sum_{s=0}^{\infty} \alpha_s^2 < \infty.$$

This algorithm becomes slow after a while and needs to be combined with a finer algorithm in the last part of the procedure. [37] provided another algorithm, this time for a max-min criterion considering more than two rival models. Convenient algorithms need to be adapted to this criterion and then their performance need to be evaluated.

Other algorithms to be adapted include exact methods such as multiplicative, interior point method, active set method, sequential quadratic programming, Nelder Mead and metaheuristic algorithms such as particle swarm optimization, simulated annealing, genetic algorithm, and hybridizations of these methods.

A friendly software is critical for the actual application of these ideas. Nowadays there are web application frameworks for most of the commercial and non commercial Mathematical and Statistical software offering a product available for anyone without having that particular software in his or her computer.

5 Discussion

Although OED for discrimination between models is quite popular at this moment there is not a definitive systematization of the features and objectives in this field converging to the joint concern of discriminating and fitting the model. As described above, there are some controversial and unsolved issues around all that has been done so far. Model Selection is essentially demanded for the current need of a scientific and correct massive data treatment. KL-optimality is currently the most general justified criterion. Nevertheless, other criteria have to be developed according to some other statistical tests not based on the likelihood ratio test [17]. This includes a deep study of D_s -optimality as well as trying additional divergence measures by justifying and comparing them.

Checking the robustness of KL-optimality for different aspects is an important task in order to have a clear idea of their strengths and limitations. A helpful Bayesian approach can be made mainly in two ways. On the one hand for dealing with the unknown parameters at the time of planning an experiment. On the other hand for the case of more than two models, which introduces a rather more complex approach with a multicriteria perspective. The later is very much related to massive data consideration. Then in a parallel way there is the case of correlated data in the so called Spatio-Temporal Statistics, which nowadays is present everywhere with great quantities of

possible data collection. Implementing algorithms and software for obtaining optimal designs is very much desirable in the whole area of OED.

Model selection and so designing experiments for that purpose is very much demanded today in our world of data analytics in the gates of the so called Industry 4.0.

Acknowledgements The authors have been sponsored by Spanish Research Agency and fondos FEDER MTM2013-47879-C2-1-P and MTM2016-80539-C2-R1. Prof. López-Fidalgo wants to appreciate the invitation to write a paper for this book. Although there was not direct collaboration in the topic with Pedro, he was always a reference both in a human perspective as well as in the scientific field. His capacity for team building taking into account every one beside him and his generosity makes him a good example of what a scientific should be.

References

1. Aletti G and May C and Tommasi Ch (2016) Best estimation of functional linear models. *Journal of Multivariate Analysis* 151:54–68
2. Amo-Salas M, López-Fidalgo J and López-Ríos VI (2012) Optimal Designs for two Nested Pharmacokinetic Models with Correlated Observations. *Communications in Statistics. Simulation and Computation* 41(7):944–963
3. Amo-Salas M, López-Fidalgo J and Pedregal DJ (2015) Experimental designs for autoregressive models applied to industrial maintenance. *Reliability engineering & system safety* 133, 87–94
4. Amo-Salas M, López-Fidalgo J and Porcu E (2013) Optimal designs for some stochastic processes whose covariance is a function of the mean. *Test* 22(1), 159–181
5. Atkinson AC (1972) Planning experiments to detect inadequate regression models. *Biometrika* 59:275–293.
6. Atkinson AC (1995) Multivariate transformations, regression diagnostics and seemingly unrelated regression. In: Kitsos CP, Müller WG, editors. *MODA 4: Advances in Model-Oriented Data Analysis*. Physica Verlag: Heidelberg:181–192
7. Atkinson A (2008) DT-optimum designs for model discrimination and parameter estimation. *Journal of Statistical Planning and Inference*. 138:56–64
8. Basseville M (2013) Divergence measures for statistical data processing—An annotated bibliography. *Signal Processing* 93, 621-633
9. Campos-Barreiro S. and Lopez-Fidalgo J (2016) KL-optimal experimental design for discriminating between two growth models applied to a beef farm. *Mathematical biosciences and engineering* 13(1), 67–82
10. Atkinson A and Cox DR (1974) Planning experiments for discriminating between models. With discussion by H. P. Wynn, D. M. Titterton, P. J. Laycock, D. V. Lindley, D. H. Hill, Agnes M. Herzberg, P. A. Tukey, A. O'Hagan, V. V. Fedorov, J. Dickey, J. Kiefer and C. A. B. Smith. *Journal of the Royal Statistical Society, Series B* 36:321–348
11. Atkinson A and Fedorov VV (1975) The design of experiments for discriminating between two rival models. *Biometrika* 62:57–70
12. Atkinson A and Fedorov VV (1975) Optimal design: experiments for discriminating between several models. *Biometrika* 62:289–303
13. BOX GEP (1979) Some problems of statistics and everyday life. *Journal of the American Statistical Association* 74, 1–4

14. Braess D and Dette H (2013) Optimal discriminating designs for several competing regression models. *Annals of Statistics* 1(2):897–922
15. Chernoff H (2000) *Sequential Analysis and Optimal Design*. SIAM Society for Industrial & Applied Mathematics, Philadelphia
16. Cox, DR (1958) *Planning of Experiments*. Wiley, New York
17. Cox DR (1962) Further Results on Tests of Separate Families of Hypotheses. *Journal of the Royal Statistical Society, Series B* 24(2):406–424
18. Deldossi L and Osmetti SA and Tommasi Ch (2016) PKL-Optimality Criterion in Copula Models for Efficacy-Toxicity Response. In: *MODA 11: Advances in Model-Oriented Data Analysis*. Eds. Ch. Müller, J. Kunert and A.C. Atkinson. Physica Verlag: Heidelberg, 79–86
19. Fedorov VV (1972) *Theory of optimal experiments*. Academic Press, New York
20. Fisher, RA (1960) *The Design of Experiments*. Oliver and Boyd, 7th Edition, Edinburgh
21. Kieffer J (1959) Optimum experimental designs. *Journal Royal Statistical Society, Ser. B* 21:272–319
22. Kieffer J and Wolfowitz J (1960) The equivalence of two extremum problems. *Canad. J. Math.* 12:363–366
23. López-Fidalgo J and Tommasi Ch and Trandafir C (2007) An optimal experimental design criterion for discriminating between non-normal models. *Journal of the Royal Statistical Society, Series B* 69:231–242
24. López-Fidalgo J and Tommasi Ch and Trandafir C (2007) Optimal designs for discriminating between some extensions of the Michaelis–Menten model. *Journal of Statistical Planning and Inference* 138:3797–3804
25. López Fidalgo, J and Wong WK (2002) Design for the Michaelis-Menten Model. *J. Theoret. Biol.* 215, 1–11
26. May C and Tommasi Ch (2014) Model selection and parameter estimation in non-linear nested models. *Statistica Sinica* 24:63–82
27. Ponce de León A and Atkinson A (1992) The design of experiments to discriminate between two rival generalized linear models. In: *Lecture Notes in Statistics. Advances in GLM and Statistical Modelling*. Springer-Verlag, 159–164
28. Pukelsheim F and Rosenberger JL (1993) Experimental Designs for Model Discrimination. *Journal of the American Statistical Association* 88(442):642–649
29. Rodríguez Aragón LJ and López Fidalgo J (2007) T-, D- and c- Optimum Designs for BET and GAB Adsorption Isotherms. *Chemom. Intell. Lab. Syst.* 89(1):36–44
30. Sánchez G and López Fidalgo J (2003) Mathematical techniques for solving analytically large compartmental systems. *Health Physics* 85(2):184–193
31. Silvey (1980) *Optimal Design*. Chapman & Hall, London
32. Smith K (1918) On the standard deviations of adjusted and interpolates values of an observed polynomial functions and its constants and the guidance they give towards a proper choice of the distribution of observations. *Biometrika* 12, 1–85
33. Stewart WE, Shon Y and Box GEP (1998) Discrimination and Goodness of Fit of Multiresponse Mechanistic Models. *AIChE Journal* 44 (6), 1404–1412
34. Tommasi Ch (2007) Optimal designs for discriminating among several non-normal models. In: *Advances in Model-Oriented Design and Analysis mODa 8*, Physica-Verlag, 213–220
35. Tommasi Ch (2009) Optimal designs for both model discrimination and parameter estimation. *Journal of Statistical Planning and Inference* 139:4123–4132
36. Tommasi Ch and López-Fidalgo J (2010) Bayesian optimum designs for discriminating between models with any distribution. *Computational Statistics & Data Analysis* 54:143–150
37. Tommasi Ch and Martín-Martín R and López-Fidalgo J (2016) Max–min optimal discriminating designs for several statistical models. *Computational Statistics & Data Analysis* 26(6):1163–1172

38. Ucinski D and Bogacka B (2004) T -Optimum Designs for Multiresponse Dynamic Heteroscedastic Models. In: 7th International Workshop on Model-Oriented Design and Analysis, MODA7. Springer Verlag, 167–174
39. van de Wal M and de Jager B (2001) A Review of Methods for Input/Output Selection. *Automatica* 37, 487-510
40. Whittle P (1973) Some general points in the theory of optimal experimental design. *Journal Royal Statistical Society, Ser. B* 1, 123–130