PLOS Computational Biology

SOFTWARE

PEtab Select: Specification standard and supporting software for automated model selection

Dilan Pathirana[1,2*], **Frank T. Bergmann**[3], **Domagoj Doresic**[1,2], **Polina Lakrisenko**[1,2,4,5], **Sebastian Persson**[6], **Niklas Neubrand**[7], **Jens Timmer**[8,9,10], **Clemens Kreutz**[7,9,10], **Harald Binder**[7,9,10], **Marija Cvijovic**[6], **Daniel Weindl**[1,2,4], **Jan Hasenauer**[1,2,4,5*], **with the PEtab Select community**[¶]

1 Bonn Center for Mathematical Life Sciences, University of Bonn, Bonn, Germany, **2** Life and Medical Sciences (LIMES) Institute, University of Bonn, Bonn, Germany, **3** BioQUANT, Heidelberg University, Heidelberg, Germany, **4** Computational Health Center, Helmholtz Zentrum München Deutsches Forschungszentrum für Gesundheit und Umwelt (GmbH), Neuherberg, Germany, **5** School of Life Sciences, Technische Universität München, Munich, Germany, **6** Department of Mathematical Sciences, Chalmers University of Technology, and University of Gothenburg, Gothenburg, Sweden, **7** Institute of Medical Biometry and Statistics (IMBI), Medical Faculty and Medical Center, University of Freiburg, Freiburg im Breisgau, Baden-Württemberg, Germany, **8** Institute of Physics, University of Freiburg, Freiburg im Breisgau, Baden-Württemberg, Germany, **9** CIBSS, Centre for Integrative Biological Signalling Studies, University of Freiburg, Freiburg im Breisgau, Baden-Württemberg, Germany, **10** Freiburg Center for Data Analysis and Modelling, University of Freiburg, Freiburg im Breisgau, Baden-Württemberg, Germany

¶ Membership list can be found in the Acknowledgments section.
* dilan.pathirana@uni-bonn.de (DP); jan.hasenauer@uni-bonn.de (JH)

Check for updates

OPEN ACCESS

Citation: Pathirana D, Bergmann FT, Doresic D, Lakrisenko P, Persson S, Neubrand N, et al. (2026) PEtab Select: Specification standard and supporting software for automated model selection. PLoS Comput Biol 22(9): e1014774. https://doi.org/10.1371/journal.pcbi.1014774

Editor: Christopher E Miles, University of Utah, UNITED STATES OF AMERICA

Received: January 30, 2026

Accepted: August 28, 2026

Published: September 30, 2026

Copyright: © 2026 Pathirana et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data availability statement: Code, data, and specific package versions for reproducing the reported results is available at Zenodo (https://doi.org/10.5281/zenodo.17185040). Code for the PEtab Select package is maintained at GitHub (https://github.com/PEtab-dev/petab_select/), and documentation is hosted at

Abstract

A central question in mathematical modeling of biological systems is determining which processes are most relevant and how they can be described. There are often competing hypotheses, which yield different models. Model comparison requires parameter optimization and sampling methods. Yet, standards for the specification of model selection problems and the swift evaluation of a broad spectrum of approaches are not available. PEtab Select addresses this challenge by providing a concise, standardized specification of model selection and its associated calibration problems through a new file format standard and software package. The standard facilitates the compact representation of even very large model selection problems; in one example, billions of model alternatives. PEtab Select builds on the PEtab standard for the specification of parameter estimation problems, and enables the use of state-of-the-art modelling and calibration workflows utilizing COPASI, Data2Dynamics, PEtab.jl, and pyPESTO. PEtab Select supports common model selection criteria (e.g., Akaike and Bayesian information criteria) and can be easily extended to use others. To ensure flexibility, PEtab Select implements several model space exploration approaches, including basic brute-force, forward, and backward selection, and also advanced, flexible selection methods. PEtab Select introduces the first standardization of model selection tasks, filling a critical gap in existing computational

ReadTheDocs (https://petab-select.readthedocs.io/).

Funding: This work was supported by the TRA Modelling (University of Bonn) as part of the Excellence Strategy of the federal and state governments in Germany, the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy (project IDs 390685813 - EXC 2047 and 390873048 - EXC 2151. DP; JH; DW; DD; PL) and through Metaflammation (project ID 432325352 – SFB 1454. JH; DP; PL), the German Federal Ministry of Education and Research (BMBF) within the e:Med funding scheme (junior research alliance PeriNAA, grant no. 01ZX1916A. DW; PL), the Horizon Europe - ERC Consolidator Grant 2023 (grant agreement no. 101126146. JH), the Federal Ministry of Education and Research (grant no. 031L0293C. JH), and by the University of Bonn via the Schlegel professorship to JH. FTB was supported by LIBIS, and de.NBI, the German Network for Bioinformatics Infrastructure (W-de.NBI-016). MC and SP were supported by the Swedish Research Council (VR2017-05117 and VR2023-04319) and the Swedish Foundation for Strategic Research (FFL15-0238). The work of NN, JT, CK, and HB has been funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 499552394 – SFB 1597 Small Data. The funder data are: - Deutsche Forschungsgemeinschaft (DFG, German Research Foundation). https://www.dfg.de/en - German Federal Ministry of Education and Research (BMBF) (now BMFTR). https://www.bmftr.bund.de/EN/Home/home_node.html - European Research Council (ERC). https://erc.europa.eu/homepage - University of Bonn. https://www.uni-bonn.de/en - Baden-Württemberg Institute for Bioinformatics Infrastructure (LIBIS). https://libis.uni-heidelberg.de/ - German Network for Bioinformatics Infrastructure – de.NBI. https://www.denbi.de/ - Swedish Research Council. https://www.vr.se/english.html - Swedish Foundation for Strategic Research. https://strategiska.se/en/ The sponsors or funders did not play any role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have no competing interests to declare.

pipelines. It constitutes an essential contribution to FAIR research software in systems biology by promoting interoperability and reusability in model selection.

Author summary

Model selection is a crucial step in mathematical modeling, guiding the choice of components to include in a model. PEtab Select automates this process across diverse modeling software frameworks and programming languages via (1) a new interoperable, language-agnostic standard for specifying large-scale model selection problems, and (2) a comprehensive software package that implements these selection methods. Developed through a community effort, PEtab Select has been integrated into multiple modeling frameworks and is accessible to users of COPASI, Julia, MATLAB, and Python.

Introduction

In systems biology, mechanistic models explain and predict dynamic processes, guide experimental design, and support hypothesis testing [1,2]. Because models encode biological assumptions in mathematical form, uncertainty arises both from incomplete hypotheses and from imperfect mathematical implementations. Model selection helps resolve these ambiguities by quantifying how well competing model formulations explain observed data [3].

The model-building cycle is increasingly supported by community standards. In this work, we primarily focus on ordinary differential equation (ODE) models; however, our work is designed to be adapted to other model classes as well. Models can be encoded in SBML [4], while simulations and parameterization workflows can be described using SED-ML [5] and PEtab [6]. Indeed, PEtab already allows one to specify measurement data, experimental conditions, observation models, and parameters, and thereby enables statistically-rigorous parameter estimation with noisy data via likelihood- or posterior-based objective functions. PEtab additionally enables interoperable calibration and analysis pipelines across tools and programming languages. Building on this, several software tools also offer model selection capabilities, including information criteria (AIC, AICc, BIC) [7–9], Bayes factors [10], likelihood ratio tests [11], and cross-validation [12]. Related functionality includes feature selection and model reduction to improve identifiability [13–18], as well as multi-model inference [19]. Widely used and versatile tools include Data2Dynamics (D2D) [20], dMod [21], and pyPESTO [22]. However, while model selection is possible using these tools, model selection problems are typically specified in tool-specific or ad hoc ways, limiting reproducibility and interoperability. Furthermore, support for different model selection approaches remains fragmented.

To address this gap, we introduce the PEtab Select standard for encoding model selection problems in systems biology. Developed in a community effort, the standard extends SBML [4] and PEtab [6] to provide a compact, interoperable

description of model spaces and associated calibration problems, enabling large-scale comparative analyses without requiring new toolchains. Alongside the standard, we provide a reference PEtab Select software package that integrates into established calibration workflows. The package interfaces with PEtab-compatible tools such as Data2Dynamics [20], PEtab.jl [23], pyPESTO [22], and COPASI [24] via BasiCO [25]. It supports multiple exploration strategies—including brute force, forward, backward, and FAMoS [26]—while delegating model calibration to the selected backend tool. The following sections describe the standard and package and how they combine with existing workflows to streamline model selection (Fig 1).

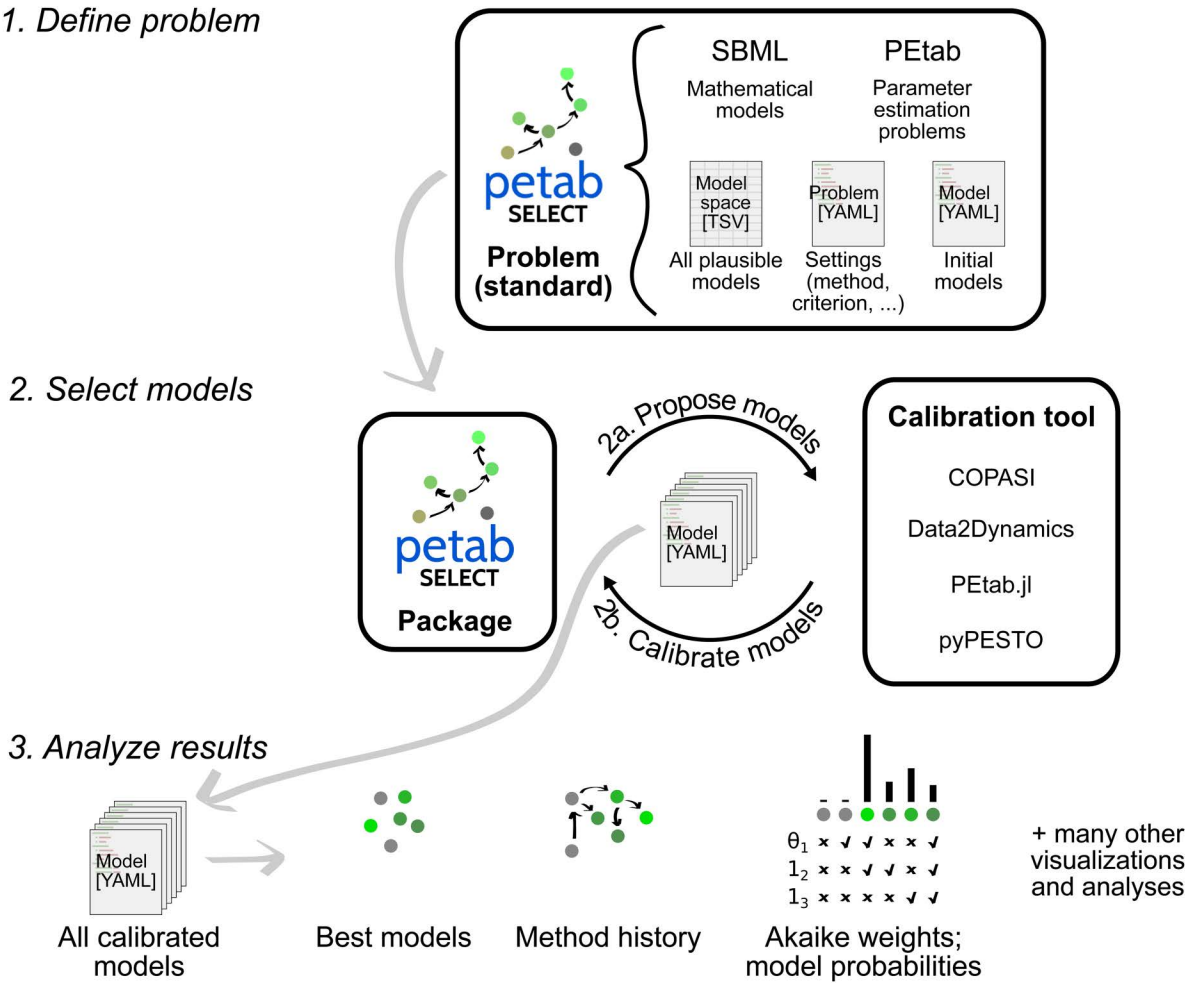


Fig 1. Model selection workflow with PEtab Select. (1) The user defines the model selection **problem**, including the model space and the mathematical description of the set of possible process components, in the **standard**. (2) The user runs model selection via a **calibration tool** by sending the PEtab Select problem to the PEtab Select **package**, which initiates iterations of model selection. (2a) The package provides model candidates to the calibration tool based on the problem specification and the available calibration results. (2b) The calibration tool calibrates the proposed candidate models and sends the calibration results to the package. Iterations of (2a) and (2b) continue until the package sends a termination signal to indicate the end of model selection. The results are provided in the PEtab Select model format and include all calibrated models with their calibration results. (3) The package can be used to perform subsequent analyses and visualizations of the results.

https://doi.org/10.1371/journal.pcbi.1014774.g001

Design and implementation

Scope

PEtab Select is designed to make model selection tasks portable across tools by separating (i) a standardized, tool-agnostic description of the model selection problem from (ii) the numerical calibration engine that evaluates candidate models. Concretely, PEtab Select provides (1) a concise specification (standard) that describes a model selection task in terms of candidate model definitions and associated calibration problems, (2) a reference software package that orchestrates model space exploration and model comparison, and (3) standardized output that records calibrated models and selection results for downstream analysis (Fig 1). Calibration itself is performed by an external PEtab-compatible tool, allowing users to retain their preferred solvers, optimizers, and workflow-specific features.

The limitations of PEtab and the calibration tools are mostly inherited by PEtab Select. For example, although there is good support for ordinary differential equation (ODE) models, there is little support for partial differential equation (PDE) models. Furthermore, although the supported model format SBML supports stochastic models in addition to ODEs, PEtab Select provides no direct support for stochastic models – one reason is that model selection becomes complicated due to the (in general) intractable likelihood. However, we provide an example of model selection with a non-ODE model (with a user-supplied calibration tool) in the documentation.

The PEtab Select standard

The standard extends **PEtab** using the same TSV/YAML-based structure while adding model-selection-specific configuration. PEtab enables maximum likelihood computation that is required for model selection with PEtab Select. The likelihood information is encoded in the PEtab file format, which comprises tabular (TSV) data that specify experimental data and conditions, the observation and noise models, and parameters to be estimated, as well as the mathematical (structural) model.

PEtab Select builds upon PEtab with a layer for defining the candidate model set and selection settings (e.g., exploration method and selection criterion). This ensures existing PEtab workflows remain valid, while enabling large and complex model selection problems to be specified compactly and exchanged between tools.

Model space (TSV) The model space specifies which candidate models are admissible for selection. In PEtab Select, model spaces are represented as a combination of (i) explicitly listed single models and (ii) *superset models* from which multiple candidates are derived by varying designated components through a table of plausible parameter values. This permits compact specifications of large combinatorial spaces. In the current scope, candidate models are derived by parameter-level switches and value assignments (e.g., "estimate" versus fixed values), which covers common use cases such as including/excluding processes, reactions, or effects via parameterization.

An example of defining a model space based on two superset models is given in Fig 2. Each superset model has an associated mathematical model and parameter estimation problem specified in the `model_subspace_petab_yaml` column. The set of submodels is constructed by customizing the superset model with all combinations of *plausible parameter values* in the additional table columns; for this specific example, subspaces `supersetA_models` and `supersetB_models` encode 16 and 8 models (Fig 2c'), respectively. Note that for the example superset models in Fig 2b and 2c, processes are turned "off" by setting parameters to 0, which is not always the case. For example, a fold-change effect parameter might be turned off by setting it to 1.

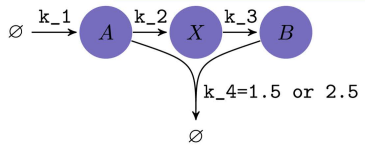
An additional feature is "metaparameters", which we designed for model selection hypotheses that require multiple parameters to be turned on or off simultaneously. For example, a hypothesis involving the Michaelis-Menten kinetic, which involves the `Vmax` and `Km` parameters, requires that either both or neither of the parameters are in the model, to toggle the hypothesis on or off, respectively. This example is provided, with additional details, in the online documentation.

Model files (YAML) Given mathematical models, parameter estimation problems, and the model space, the package can run model selection methods. As the methods require calibrated models, the package sends the uncalibrated models

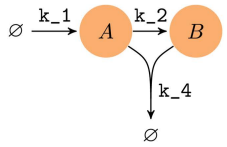
(a) The PEtab Select model space table.

Model subspace metadata		*Plausible parameter values*			
model_subspace_id	model_subspace_petab_yaml	k_1	k_2	k_3	k_4
supersetA_models	problem1.yaml	0;estimate	0;estimate	0;estimate	1.5;2.5
supersetB_models	problem2.yaml	0;estimate	0;estimate	0	0;estimate

(b) The superset model for the subspace **supersetA_models**.



(c) The superset model for the subspace **supersetB_models**.



(c') All models in the subspace **supersetB_models**.

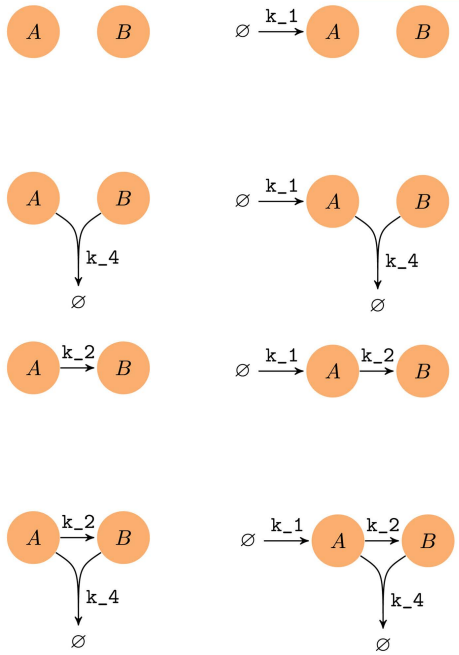


Fig 2. Example model space. (a) The model space table encodes model subspaces as rows. Each model subspace has an ID and an associated superset PEtab problem, which includes the mathematical model. This superset problem is customized into single models according to the remaining columns, which contain semicolon-delimited lists of values that parameters (column headers) can take in the subspace. The next subfigures represent models as graphs. Nodes (colored circles) are variables (e.g., biochemical species), and arrows are processes, e.g., conversion (• → •), import (∅ →), or export (→ ∅) of the variables. The arrows are annotated with their associated parameters, e.g., `k_4` is the rate constant associated with export of variables *A* and *B* due to an interaction between *A* and ***B*. (b,c)** Each subspace is associated with a superset model, where all possible parameters are estimated. Based on the superset models, candidate models are encoded by turning (estimation of unknown) parameters on or off. The total number of models in this example is 16 in subspace `supersetA_models`, plus (c') 8 in subspace `supersetB_models`.

https://doi.org/10.1371/journal.pcbi.1014774.g002

to a calibration tool, which calibrates the models and then sends the results back to the package (step 2 in Fig 1). The standard model file format is used to transmit all individual model information at all steps of the workflow (Fig 1). These files are generated automatically by the package, but can also be supplied manually, for example to specify initial models (as done in test case 0009 of the PEtab Select test suite).

Model selection problem file (YAML) As in PEtab, a single YAML-based "problem" file bundles the complete model selection task so compatible tools can load it from one entry point. The problem file links the relevant inputs (e.g., model space and PEtab problems) and specifies the selection configuration, including the exploration method and the selection criterion.

The PEtab Select package

The PEtab Select package extends calibration workflows by providing an API that enables model selection while leaving model calibration to the backend tool. Users provide PEtab/PEtab Select files, initiate model selection, and obtain

standardized output in the PEtab Select model format for analysis and visualization (Fig 1). Basic usage is typically with the calibration environment directly, rather than with the package. However, the pipelines are flexible, and more fine-grained interaction with the calibration tool and the package directly is supported during each iteration of model selection.

Calibration tools To support PEtab Select, a calibration tool must be able to calibrate PEtab problems and interface with the package API. The package is written in Python 3 and can be accessed via a Python API or via the command line for non-Python tools. A dedicated test suite supports validation of tool integrations. Model selection with PEtab Select can currently be performed by COPASI [24] via BasiCO [25], Data2Dynamics [1,20], PEtab.jl [23], and pyPESTO [22].

Model selection methods

The PEtab Select package supports global and local model selection methods to explore a user-defined model space. In each iteration, the package proposes one or more candidate models, obtains their calibration results from the backend tool, evaluates the chosen selection criterion, and decides which model(s) to consider next. Bold terms below denote methods and features implemented in the package.

Global methods For exhaustive exploration, the package provides a **brute-force** mode that calibrates and scores every model in the specified space. This mode is useful for small or moderate model spaces and for benchmarking, but becomes impractical for combinatorial spaces (e.g., when n components can be switched on/off, yielding up to 2^n candidates).

Local methods For larger spaces, the package implements stepwise local exploration. The **forward** method iteratively adds components (or releases fixed parameters for estimation), the **backward** method removes them, and the **lateral** method performs swap moves that add and remove in the same step. These local methods operate from an initial model, which can be provided by the user or generated automatically by the package.

The package also supports **FAMoS** [26], a metaheuristic local strategy that switches between forward, backward, and lateral moves according to a customizable switching scheme. In addition, FAMoS supports a "**jump to most distant**" move to escape local neighborhoods [26].

In general, if brute-force is computationally feasible, then we recommend it. Otherwise, we recommend the FAMoS method. This is because more of the model space is reachable, than with classical backward or forward searches. However, FAMoS itself is primarily a sequence of local searches, which can result in longer runtimes in practice, compared to forward or backward searches. As the lateral method cannot reach models that differ in size to the initial model, we recommend only using it via FAMoS.

Model selection criteria All methods rely on a model selection criterion provided in the PEtab Select problem specification. The package supports the popular information criteria **AIC** [7], **AICc** [8], and **BIC** [9]. Criteria are used consistently across candidate models by evaluating the calibrated fit while accounting for model complexity, enabling automated ranking and decision-making during exploration [3].

Other features The package provides several other features to enable flexible model selection. This includes automated summaries of the model selection run, custom or automated selection of initial models, limits on the number of models calibrated in any iteration, restarting of model selection runs with previous model selection results, and object-oriented programming such that other methods and selection criteria can be easily implemented.

FAIR data principles

The FAIR principles emphasize machine-actionable interoperability and reuse [27]. PEtab Select supports these goals by: using machine-readable TSV/YAML formats; defining a documented controlled vocabulary; providing schemas (JSON Schema) for YAML-based files; and, enabling end-to-end reproducibility of model selection tasks across tools by building on community standards (SBML and PEtab). Results are stored in the same standardized format, enabling consistent downstream analysis independent of the calibration backend (e.g., via package-provided visualization utilities).

Results

PEtab Select enables model selection in existing workflows, each with unique tool-specific features

To assess the interoperability of PEtab Select with different software suites, we tested its functionality in four PEtab-compatible tools: BasiCO (built on COPASI), Data2Dynamics, PEtab.jl, and pyPESTO. Each tool successfully passed the tests in the test suite, confirming that they can parse and execute model selection tasks defined in the PEtab Select standard. While the software tools have a broad spectrum of overlapping features, there are a few distinctive characteristics.

BasiCO (COPASI) provides the possibility of custom evaluation functions in place of the standard information criteria. Moreover, its programmable interface enables model-specific optimizer settings, so callbacks or post-processing steps can be incorporated into the model selection workflow. This flexibility allows users to adjust parameters on a per-model basis or to visualize results automatically after each calibration step.

Data2Dynamics supports transferring parameter estimates from one calibrated model to another, potentially reducing computation time when new models are introduced. Additionally, users can define custom calibration routines to adjust the workflow before, during, or after parameter fitting. For example, the number of multi-starts in an optimizer can be dynamically set based on the size of the parameter space, or diagnostic plots can be automatically generated to assess calibration quality.

PEtab.jl integrates with the Julia SciML ecosystem, tapping into highly efficient ODE solvers provided by `DifferentialEquations.jl` [28] and automatic differentiation routines [29]. These capabilities enable rapid prototyping of novel approaches and runtime-efficient parameter estimation [23].

pyPESTO supports ensemble-based approaches, for instance, using Akaike-weighted ensembles to combine the results from multiple candidate models. Its ready-to-use post-processing tools also facilitate structured output of calibration results and automated generation of diagnostic figures, streamlining the analysis of potentially large model spaces.

PEtab Select enables reproducible model selection across workflows

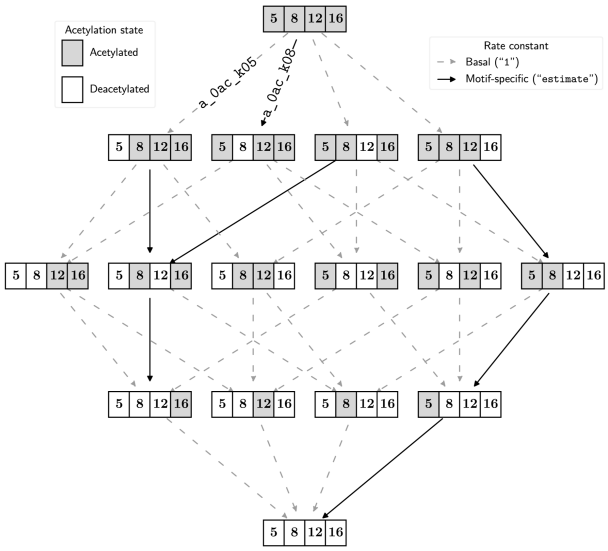
To assess the comparability of results across workflows, we consider an exemplary model selection problem from the work of Blasi et al. (2016) [13] that studied the acetylation of the histone H4 protein, which can be acetylated independently at 4 positions, yielding 16 possible acetylation patterns (motifs). The mathematical model is an ordinary differential equation (ODE) model that describes the acetylation and deacetylation reactions that convert one histone H4 motif into another. The ODE has 16 state variables that describe the abundance of the different histone motifs. The experimental data are taken from the original publication. Briefly, the experiment begins with an initial amount of the unacetylated histone H4 motif, and measurement data are the relative concentrations of 15 motifs at a single later time point. As the acetylation and deacetylation reactions are assumed to be fast compared to the experiment timescale, the measurements are assumed to be at steady state. Further details on the mathematical model and experimental data are provided in the original publication [13] and in the PEtab Select documentation.

A visualization of all 16 motifs, and the 32 acetylation reactions between them, is provided in Fig 3a. Deacetylation reactions are not depicted and are assumed to occur with the same rate constant. The same assumption serves as the null hypothesis for the acetylation reactions. The model selection problem is therefore to identify those acetylation reactions that require motif-specific, i.e., reaction-specific, rate constant parameters.

This model selection problem differs from the original publication in that we omit the additional 11 “site-specific” models in the original publication, which add a negligible computational cost and are not amongst the best models. Our model selection problem of 32 plausible motif-specific parameters therefore involves a model space of $2^{32} \approx 4.3$ billion plausible models.

The corresponding PEtab Select model space table is shown in Fig 3b. The problem files and all technical details of the experiments in this section are provided in the Zenodo repository (see Availability and future directions).

(a) Published best model, taken from Blasi et al. (2016).



(b) PEtab Select model space table for the Blasi et al. (2016) model selection problem.

Model subspace metadata		*Plausible parameter values*		
`model_subspace_id`	`model_subspace_petab_yaml`	`a_0ac_k05`	`a_0ac_k08`	...
`M`	`petab_problem.yaml`	`1;estimate`	`1;estimate`	...

Fig 3. The Blasi et al. (2016) [13] model selection problem in PEtab Select. (a) The reaction network for the published best model, represented as a graph similarly to Figs 2b-2c'. The histone chemical species can undergo acetylation reactions at four different positions (5, 8, 12, and 16). Nodes are acetylation states (motifs) of the histone, and arrows are acetylation reactions. The model selection problem involves two different possibilities for the rate constant of each reaction; it is either: the shared basal rate constant (the value is fixed to 1), or an estimated motif-specific rate constant (the value is estimated). The model selection problem is to identify the reactions with motif-specific rate constants, and the published best model contains seven (solid arrows). **(b)** There are 32 reactions, and therefore 32 independent hypotheses for motif-specific rate constants. Only two parameterized hypotheses (`a_0ac_k05` and `a_0ac_k08`) are shown here for brevity, with the remaining 30 parameters represented by the ellipses. This two-row, 34-column table encodes the full model space of ≈ 4.3 billion models.

https://doi.org/10.1371/journal.pcbi.1014774.g003

In the original publication, all ≈ 4.3 billion models were calibrated then compared, and the best model (Table S1 in the Supplemental Information of [13]) selected 7 reactions with motif-specific rate constants (Fig 3a, solid black arrows).

To solve our implementation of the model selection problem, we ran the PEtab Select implementation of the FAMoS method [26] 100 times, and each run was initialized at a randomly chosen model in the model space that had on average 16 motif-specific parameters (Fig 4b). This first experiment with 100 FAMoS runs was to determine whether PEtab Select can find the same published best model, and was conducted using only pyPESTO as the calibration tool. The best model across the 100 searches matched the published best model, and the FAMoS algorithm found it three times in the 100 runs (Fig 4c). In total, there were 83,125 (≈ 0.00194% of 2^{32}; 58,858 unique) model calibrations across the 100 runs, which took 655.8 CPU hours.

Extrapolating to the full model space, this saved $\frac{2^{32}}{83125} \cdot \frac{655.8}{24 \cdot 365} = 3868$ CPU years; hence, FAMoS rendered the standardized PEtab Select approach computationally feasible, meaning that tailoring the simulation and optimization routines to the problem, as was done in the original publication, is not necessary.

One of the three successful runs involved 339 model calibrations with pyPESTO, and its initial model is given in Fig 4a. To assess reproducibility of model selection results generated with PEtab Select across all calibration tools, we repeated this FAMoS run with each tool.

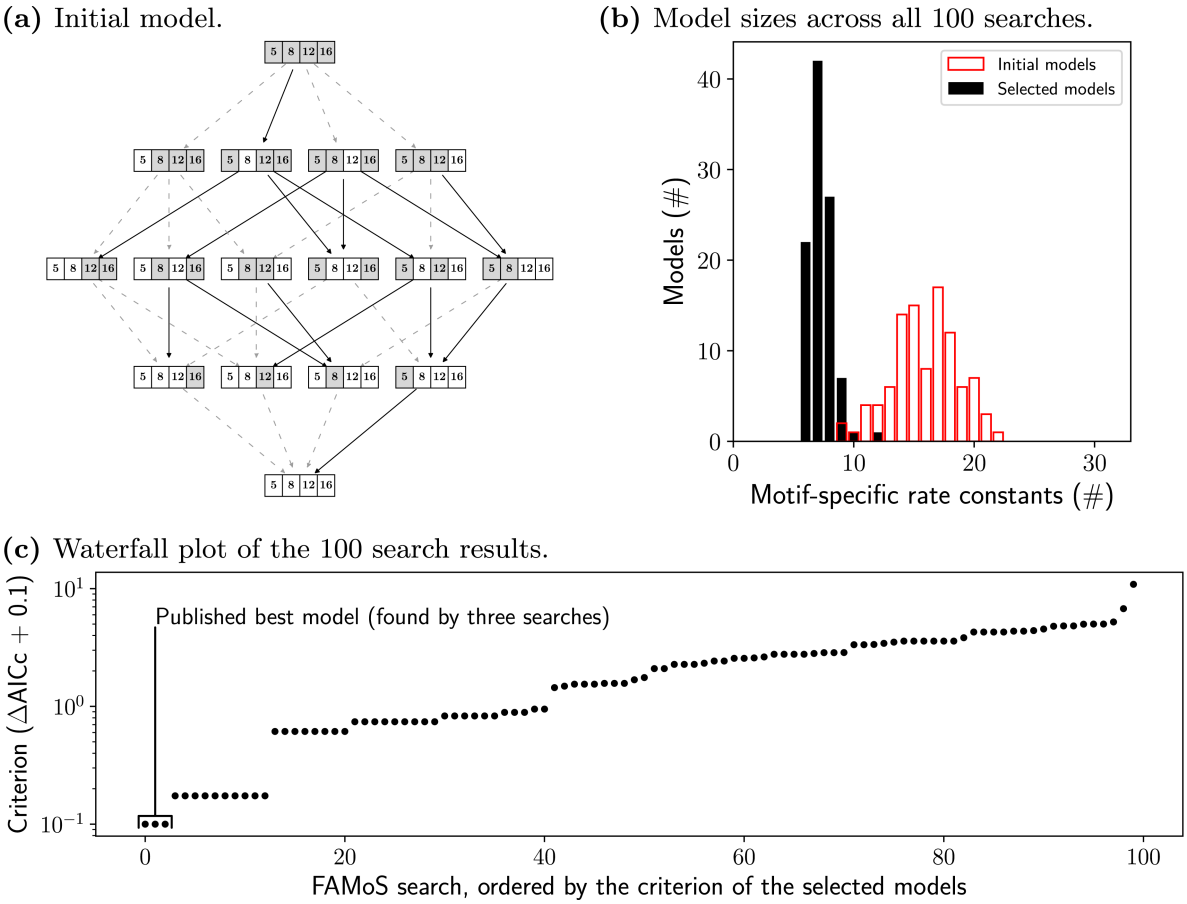


Fig 4. Multi-start FAMoS searches solve the Blasi et al. (2016) model selection problem. (a) An initial model, visualized as in Fig 3a, that was used to initialize one of the FAMoS searches. This is the initial model that was used later to assess reproducibility with all calibration tools. **(b)** The sizes of models before and after model selection. The published best model had seven motif-specific rate constants. **(c)** The criterion of the selected model from each of the 100 FAMoS searches.

https://doi.org/10.1371/journal.pcbi.1014774.g004

All runs, i.e., with COPASI (BasiCO), Data2Dynamics and PEtab.jl, reproduced the same sequence of local search methods and calibrated similar sets of 339 models. Importantly, the same best model (Fig 3a) was found with each of the calibration tools.

Model selection for the considered problem is inherently challenging. The search space is vast and populated by many candidate models whose information criterion values differ only within numerical precision, so minute floating-point variations render the ranking effectively non-deterministic across machines. Although the reproduced single FAMoS search (with 339 model calibrations) takes slightly different trajectories through the model space on repeated executions within and between calibration tools (Fig 5, calibration tool rows), each trajectory delivers the same overall improvement within simulator and optimizer tolerances (Fig 5, row "Criterion (ΔAICc)"), with negligible differences (< 0.01) in criterion across calibration tools.

Indeed, classical forward and backward selection mostly failed to arrive at the optimal solution. By contrast, the FAMoS algorithm [26] reproducibly converged on models with markedly better criterion scores. Overall, this example highlights that, in large model spaces where brute-force is intractable, multi-start local searches can be a useful alternative. Although no guarantees can be made about multi-start local searches, they can be performed similarly to multi-start optimization in

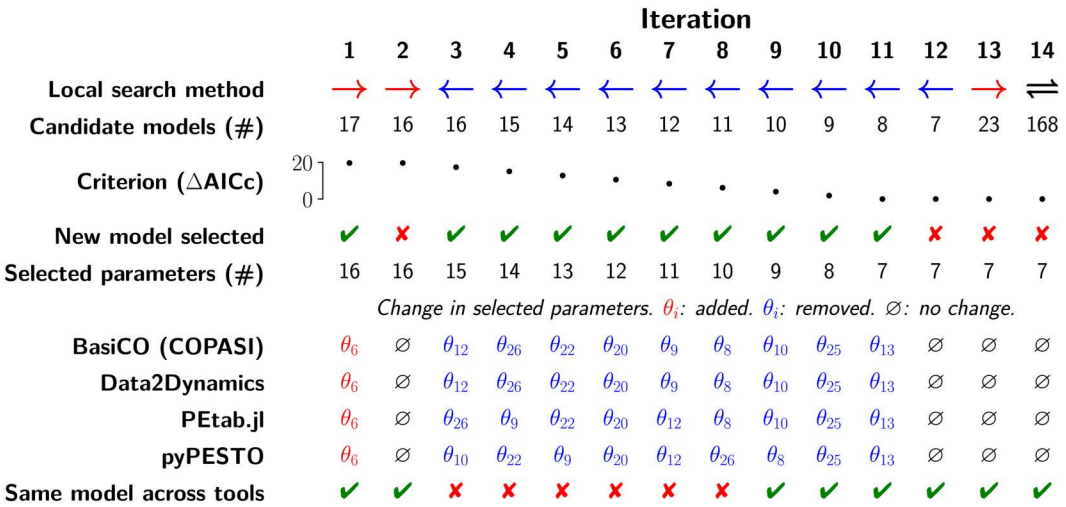


Fig 5. Reproducibility of model selection across calibration tools. Local search methods are the forward (→), backward (←), and lateral (⇌) methods. All calibration tools reproduced the same criterion value in each iteration, within 0.01 ΔAICc. The only difference between tools in each iteration is in the change in selected parameters. θ_i is the ith motif-specific parameter; see the Zenodo repository for details.

https://doi.org/10.1371/journal.pcbi.1014774.g005

order to increase the chance of finding the best model(s). We recommend running local searches until (i) the goodness-of-fit (and other standard checks) of the selected model(s) are acceptable [2], and (ii) those selected model(s) have been found by multiple independent local searches.

Availability and future directions

Code, data, and specific package versions for reproducing the reported results is available at Zenodo (https://doi.org/10.5281/zenodo.17185040). Code for the PEtab Select package is maintained at GitHub (https://github.com/PEtab-dev/petab_select/), and documentation is hosted at ReadTheDocs (https://petab-select.readthedocs.io/).

There are several useful directions in which PEtab Select can be improved. Currently, PEtab Select only provides support for the SBML mathematical model format, as this is the only model format supported by PEtab. However, PEtab Select has no direct dependency on the model format. Hence, this is expected to improve with the next version of PEtab, which will support non-SBML model formats. An important issue in model selection is parameter identifiability, which is required for many of the popular model selection criteria including the AIC. Accurate identifiability analysis is computationally costly and therefore infeasible in large model spaces; hence, a potential improvement for PEtab Select is automated identifiability analysis, e.g., for a subset of models as a heuristic to ensure model selection is accurate. Further implementation improvements include shifting calibration-tool specific features into the package directly, such that those features can be used with any calibration tool. One example would be the termination of a model selection iteration as soon as a better model is found.

Several opportunities for the improvement of model selection methodology were identified when working with the problem published by Blasi et al. (2016) [13]. On one hand, the analysis revealed that reproducibility could be boosted by developing tailored strategies for handling the case that multiple models are equivalent in best criterion value (up to numerical noise). If performing a backward search, one possibility is to branch the backward search into multiple backward searches, one per equivalent model. On the other hand, a further improvement of search algorithms would be beneficial, in particular for large model spaces, e.g., an automatic tuning of metaheuristics in the FAMoS method across runs.

In summary, PEtab Select establishes, for the first time, a unified standard and reference implementation for large-scale, reproducible model selection. By decoupling the description of a model selection task from the calibration engine that solves it, the framework enables researchers to tackle search spaces of billions of candidates in whichever environment—COPASI, Data2Dynamics, pyPESTO, PEtab.jl, or future tools—best suits their workflows. This interoperability not only accelerates day-to-day practice, but also lowers the barrier for methodological innovation, allowing new inference algorithms to be benchmarked immediately on realistic problems. We therefore expect PEtab Select to become a cornerstone for transparent, comparable, and scalable model selection across the life and physical sciences.

Acknowledgments

The authors are grateful to the many people who have provided feedback and contributions to the standard, package, and other aspects of PEtab Select, including participants of the HARMONY and COMBINE conferences (https://co.mbine.org).

PEtab Select community Severin Bang, Adrian Hauber, Svenja Kemmer, Daniel Lill, Lukas Refisch, Marcus Rosenblatt, Christian Tönsing, Jakob Vanhoefer, Franz-Georg Wieland.

Author contributions

Conceptualization: Dilan Pathirana, Frank T. Bergmann, Polina Lakrisenko, Sebastian Persson, Daniel Weindl, Jan Hasenauer.

Data curation: Dilan Pathirana, Polina Lakrisenko.

Formal analysis: Dilan Pathirana, Frank T. Bergmann, Sebastian Persson, Niklas Neubrand.

Funding acquisition: Dilan Pathirana, Jens Timmer, Clemens Kreutz, Harald Binder, Marija Cvijovic, Daniel Weindl, Jan Hasenauer.

Investigation: Dilan Pathirana, Frank T. Bergmann, Sebastian Persson, Niklas Neubrand.

Methodology: Dilan Pathirana, Domagoj Doresic, Polina Lakrisenko, Sebastian Persson, Jan Hasenauer.

Project administration: Dilan Pathirana, Jan Hasenauer.

Software: Dilan Pathirana, Frank T. Bergmann, Domagoj Doresic, Polina Lakrisenko, Sebastian Persson, Niklas Neubrand, Daniel Weindl.

Supervision: Dilan Pathirana, Jens Timmer, Clemens Kreutz, Harald Binder, Marija Cvijovic, Daniel Weindl, Jan Hasenauer.

Validation: Dilan Pathirana, Frank T. Bergmann, Sebastian Persson, Niklas Neubrand.

Visualization: Dilan Pathirana, Polina Lakrisenko, Daniel Weindl, Jan Hasenauer.

Writing – original draft: Dilan Pathirana, Frank T. Bergmann, Sebastian Persson, Niklas Neubrand, Daniel Weindl, Jan Hasenauer.

Writing – review & editing: Dilan Pathirana, Frank T. Bergmann, Domagoj Doresic, Polina Lakrisenko, Sebastian Persson, Niklas Neubrand, Jens Timmer, Clemens Kreutz, Harald Binder, Marija Cvijovic, Daniel Weindl, Jan Hasenauer.

References

1. Raue A, Schilling M, Bachmann J, Matteson A, Schelker M, Kaschek D, et al. Lessons learned from quantitative dynamical modeling in systems biology. PLoS One. 2013;8(9):e74335. https://doi.org/10.1371/journal.pone.0074335 PMID: 24098642
2. Villaverde AF, Pathirana D, Fröhlich F, Hasenauer J, Banga JR. A protocol for dynamic model calibration. Brief Bioinform. 2022;23(1):bbab387. https://doi.org/10.1093/bib/bbab387 PMID: 34619769

3. Burnham KP, Anderson DR. Model Selection and Multimodel Inference. New York, NY: Springer. 2004.
4. Hucka M, Finney A, Sauro HM, Bolouri H, Doyle JC, Kitano H, et al. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. Bioinformatics. 2003;19(4):524–31. https://doi.org/10.1093/bioinformatics/btg015 PMID: 12611808
5. Waltemath D, Adams R, Bergmann FT, Hucka M, Kolpakov F, Miller AK, et al. Reproducible computational biology experiments with SED-ML--the Simulation Experiment Description Markup Language. BMC Syst Biol. 2011;5:198. https://doi.org/10.1186/1752-0509-5-198 PMID: 22172142
6. Schmiester L, Schälte Y, Bergmann FT, Camba T, Dudkin E, Egert J, et al. PEtab-Interoperable specification of parameter estimation problems in systems biology. PLoS Comput Biol. 2021;17(1):e1008646. https://doi.org/10.1371/journal.pcbi.1008646 PMID: 33497393
7. Akaike H. A new look at the statistical model identification. IEEE Trans Automat Contr. 1974;19(6):716–23. https://doi.org/10.1109/tac.1974.1100705
8. Sugiura N. Further analysis of the data by Akaike's information criterion and the finite corrections. Communications in Statistics - Theory and Methods. 1978;7(1):13–26. https://doi.org/10.1080/03610927808827599
9. Schwarz G. Estimating the Dimension of a Model. Ann Statist. 1978;6(2). https://doi.org/10.1214/aos/1176344136
10. Kass RE, Raftery AE. Bayes Factors. Journal of the American Statistical Association. 1995;90(430):773–95. https://doi.org/10.1080/01621459.1995.10476572
11. Tönsing C, Steiert B, Timmer J, Kreutz C. Likelihood-ratio test statistic for the finite-sample case in nonlinear ordinary differential equation models. PLoS Comput Biol. 2023;19(9):e1011417. https://doi.org/10.1371/journal.pcbi.1011417 PMID: 37738254
12. Arlot S, Celisse A. A survey of cross-validation procedures for model selection. Statist Surv. 2010;4(none). https://doi.org/10.1214/09-ss054
13. Blasi T, Feller C, Feigelman J, Hasenauer J, Imhof A, Theis FJ, et al. Combinatorial Histone Acetylation Patterns Are Generated by Motif-Specific Reactions. Cell Syst. 2016;2(1):49–58. https://doi.org/10.1016/j.cels.2016.01.002 PMID: 27136689
14. Milias-Argeitis A, Oliveira AP, Gerosa L, Falter L, Sauer U, Lygeros J. Elucidation of Genetic Interactions in the Yeast GATA-Factor Network Using Bayesian Model Selection. PLoS Comput Biol. 2016;12(3):e1004784. https://doi.org/10.1371/journal.pcbi.1004784 PMID: 26967983
15. Henriques D, Alonso-Del-Real J, Querol A, Balsa-Canto E. Saccharomyces cerevisiae and S. kudriavzevii Synthetic Wine Fermentation Performance Dissected by Predictive Modeling. Front Microbiol. 2018;9:88. https://doi.org/10.3389/fmicb.2018.00088 PMID: 29456524
16. Mosbacher M, Lee SS, Yaakov G, Nadal-Ribelles M, de Nadal E, van Drogen F, et al. Positive feedback induces switch between distributive and processive phosphorylation of Hog1. Nat Commun. 2023;14(1):2477. https://doi.org/10.1038/s41467-023-37430-y PMID: 37120434
17. Gupta S, Lee REC, Faeder JR. Parallel Tempering with Lasso for model reduction in systems biology. PLoS Comput Biol. 2020;16(3):e1007669. https://doi.org/10.1371/journal.pcbi.1007669 PMID: 32150537
18. Maiwald T, Hass H, Steiert B, Vanlier J, Engesser R, Raue A, et al. Driving the Model to Its Limit: Profile Likelihood Based Model Reduction. PLoS One. 2016;11(9):e0162366. https://doi.org/10.1371/journal.pone.0162366 PMID: 27588423
19. Theorell A, Nöh K. Reversible jump MCMC for multi-model inference in Metabolic Flux Analysis. Bioinformatics. 2020;36(1):232–40. https://doi.org/10.1093/bioinformatics/btz500 PMID: 31214716
20. Raue A, Steiert B, Schelker M, Kreutz C, Maiwald T, Hass H, et al. Data2Dynamics: a modeling environment tailored to parameter estimation in dynamical systems. Bioinformatics. 2015;31(21):3558–60. https://doi.org/10.1093/bioinformatics/btv405 PMID: 26142188
21. Kaschek D, Mader W, Fehling-Kaschek M, Rosenblatt M, Timmer J. Dynamic modeling, parameter estimation, and uncertainty analysis in R. Journal of Statistical Software. 2019;88:1–32. https://doi.org/10.18637/jss.v088.i10
22. Schälte Y, Fröhlich F, Jost PJ, Vanhoefer J, Pathirana D, Stapor P, et al. pyPESTO: a modular and scalable tool for parameter estimation for dynamic models. Bioinformatics. 2023;39(11):btad711. https://doi.org/10.1093/bioinformatics/btad711 PMID: 37995297
23. Persson S, Fröhlich F, Grein S, Loman T, Ognissanti D, Hasselgren V. PEtab.jl: Advancing the Efficiency and Utility of Dynamic Modelling. bioRxiv. 2025;:2025.04.30.651378. https://doi.org/10.1101/2025.04.30.651378
24. Hoops S, Sahle S, Gauges R, Lee C, Pahle J, Simus N, et al. COPASI--a COmplex PAthway SImulator. Bioinformatics. 2006;22(24):3067–74. https://doi.org/10.1093/bioinformatics/btl485 PMID: 17032683
25. Bergmann FT. BASICO: A simplified Python interface to COPASI. JOSS. 2023;8(90):5553. https://doi.org/10.21105/joss.05553
26. Gabel M, Hohl T, Imle A, Fackler OT, Graw F. FAMoS: A Flexible and dynamic Algorithm for Model Selection to analyse complex systems dynamics. PLoS Comput Biol. 2019;15(8):e1007230. https://doi.org/10.1371/journal.pcbi.1007230 PMID: 31419221
27. Wilkinson MD, Dumontier M, Aalbersberg IJJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data. 2016;3:160018. https://doi.org/10.1038/sdata.2016.18 PMID: 26978244
28. Rackauckas C, Nie Q. DifferentialEquations.jl – A performant and feature-rich ecosystem for solving differential equations in Julia. Journal of Open Research Software. 2017;5(1):15. https://doi.org/10.5334/jors.151
29. Revels J, Lubin M, Papamarkou T. Forward-Mode Automatic Differentiation in Julia. arXiv. 2016. https://doi.org/10.48550/arXiv.1607.07892