

# Towards a Benchmark Framework for Model Order Reduction in the Mathematical Research Data Initiative (MaRDI)

Peter Benner<sup>1</sup>   Kathryn Lund<sup>1</sup>   Jens Saak<sup>1</sup>

<sup>1</sup>*Max Planck Institute for Dynamics of Complex Technical Systems, Sandtorstr. 1, 39106 Magdeburg*  
Email: {benner, lund, saak}@mpi-magdeburg.mpg.de; ORCID: 0000-0003-3362-4103,  
ORCID: 0000-0001-9851-6061, ORCID: 0000-0001-5567-9637

**Abstract:** The race for the most efficient, accurate, and universal algorithm in scientific computing drives innovation. At the same time, this healthy competition is only beneficial if the research output is actually comparable to prior results. Fairly comparing algorithms can be a complex endeavor, as the implementation, configuration, compute environment, and test problems need to be well-defined. Due to the increase in computer-based experiments, new infrastructure for facilitating the exchange and comparison of new algorithms is also needed. To this end, we propose a benchmark framework, as a set of generic specifications for comparing implementations of algorithms using test cases native to a community. Its value lies in its ability to fairly compare and validate existing methods for new applications, as well as compare newly developed methods with existing ones. As a prototype for a more general framework, we have begun building a benchmark tool for the model order reduction (MOR) community.

The data basis of the tool is the collection of the Model Order Reduction Wiki (MOR-Wiki). The wiki features three main categories: benchmarks, methods, and software. An editorial board curates submissions and patrols edited entries. Data sets for linear and parametric-linear models are already well represented in the existing collection. Data sets for non-linear or procedural models, for which only evaluation data, or codes / algorithmic descriptions, rather than equations, are available, are being added and extended. Properties and interesting characteristics used for benchmark selection and later assessments are recorded in the model metadata.

Our tool, the Model Order Reduction Benchmark (MORB) is under active development for linear time-invariant systems and solvers. An ontology (MORBO) and knowledge graph are being developed in parallel. They catalog benchmark problem sets and their metadata and will also be integrated into the Mathematical Research Data Initiative (MaRDI) Portal<sup>1</sup>, to help improve the findability of such data sets. MORB faces a number of technical and field-specific challenges, and we hope to recruit community input and feedback while presenting some initial results.

**Keywords:** model order reduction, benchmark framework, scientific computing, research data management

**AMS subject classifications:** 65-04, 93-04, 93-11, 93C05

---

<sup>1</sup><https://www.mardi4nfdi.de/>

**Novelty statement:** We present a preliminary software tool for benchmarking model order reduction software on linear, time-invariant, first-order systems.

## 1 Introduction

The Mathematical Research Data Initiative (MaRDI)<sup>2</sup> is a consortium of the National Research Data Initiative (NFDI)<sup>3</sup>, whose overarching goal is to improve and promote responsible research data management practices in the German scientific landscape and beyond. MaRDI itself concentrates on several fields of mathematics featured as content-specific task areas (TA):

- (TA1) Computer Algebra, (exact data)
- (TA2) Scientific Computing, (inexact data)
- (TA3) Statistics and Machine Learning, and (uncertain data)
- (TA4) Cooperation with Other Disciplines. (mixed data and interoperability)

Within TA2, we focus primarily on numerical algorithms and their software implementations, which are used to compute approximate solutions and simulations of scientific models. Via collaboration between the Max Planck Institute for Dynamics of Complex Technical Systems and the University of Münster, TA2 addresses the following measures (M):

- (M1) Knowledge graph of numerical algorithms;
- (M2) Open interfaces for scientific computing;
- (M3) Benchmark framework; and
- (M4) Description and design of Findable, Accessible, Interoperable, and Reproducible (FAIR)<sup>4</sup> workflows for computational science and engineering (CSE).

For this report, we will concentrate on progress in (M3).

## 2 MaRDIMark

We begin with a domain-independent specification of a benchmark workflow, MaRDIMark. MaRDIMark is broken into five modules, as displayed in Figure 1. The content and function of each module is detailed in the following sections.

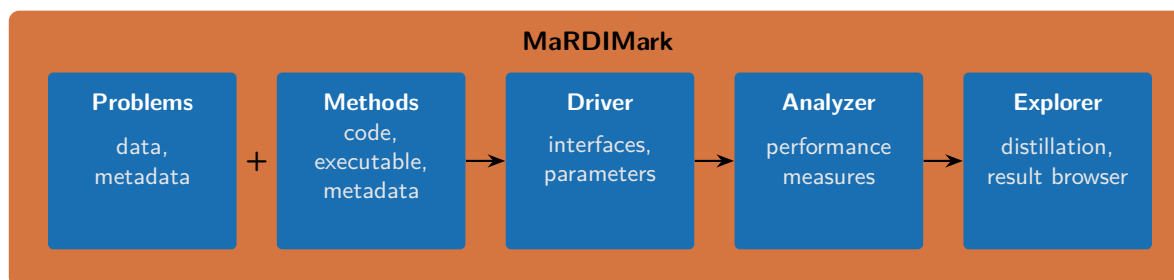


Figure 1: Overview of MaRDIMark.

<sup>2</sup>[www.mardi4nfdi.de](http://www.mardi4nfdi.de)

<sup>3</sup>[www.nfdi.de](http://www.nfdi.de)

<sup>4</sup>[www.go-fair.org/fair-principles](http://www.go-fair.org/fair-principles)

## 2.1 Problems

This module consists of the benchmark problems themselves, both as mathematical abstractions and as datasets stored on a computer. Note that datasets might take the form of both assembly scripts and static floating-point numbers. A good benchmark collection should minimally comprise a database of unique problem IDs, download links, mathematical descriptions, fields referring to domain-specific mathematical and numerical properties, file formats, license data, creator and editor identifiers (ideally ORCID<sup>5</sup>), and any other information that helps with data provenance, findability, and accessibility. Ontologies and knowledge graphs like AlgoData<sup>6</sup> can add value to such a database, as they allow users to query complex relations, instead of just look up values in a table. For example, a user may want to know which methods can be applied to a given benchmark, which brings us to the next module.

## 2.2 Methods

The Methods module is similar to the Problems module, except that the objects of interest are instead mathematical procedures and algorithms for solving a class of problems. The definition of a method for a particular field can be nontrivial, especially when all implementation details are taken into account. For example, most methods in numerical mathematics are initially presented as pseudocode, which a practitioner must translate into actual code to use on a computer. Minimally, the practitioner must choose hardware (i.e., a physical machine), an operating system (and version), and a programming language (and version). Within that language, there are multiple ways to execute any mathematical operation, and each may have slightly different floating-point behavior, which further depends on what libraries the language and operating system are running on the backend. We loosely refer to a method, its pseudocode, and its implementation choices as an *algorithm isotope*, in analogy to elemental isotopes. Each community must decide on its own how fine-grained their algorithm isotopes should be, and these definitions will necessarily vary from application to application, depending on what aspects of a class of methods are being studied.

## 2.3 Driver

The linchpin of a benchmark workflow is a routine (or set of routines) that executes specified methods on specified problems, i.e., the Driver module. In practice, a benchmark workflow may be written in a general-purpose language but need to call routines written in another language, without introducing unnecessary overhead. Here open interfaces developed by (M2) can be applied, which facilitate communication between different software packages and programming languages.

## 2.4 Analyzer

Benchmarking by definition is concerned with the comparison of methods with respect to different measures, such as, time to solution, accuracy, quality, optimized cost function, and so forth. We refer to the collection and calculation of these measures as the Analysis module. A community must determine on its own which measures are relevant and how to implement them in an unbiased manner, so that no method is artificially penalized by the procedure used to calculate a measure.

## 2.5 Explorer

Finally, it is important that practitioners are able to visualize the results and share them with colleagues. The Explorer module thus comprises graphs, plots, tables, and reports that MaRDIMark synthesizes. Convenient formats for many mathematical fields might be encoded in TeX, HTML, or Markdown, or as a Jupyter notebook.

---

<sup>5</sup>[www.orcid.org](http://www.orcid.org)

<sup>6</sup>A deliverable of (M1). See <https://algodata.mardi4nfdi.de/>.

### 3 Model Order Reduction Benchmark (MORB)

To demonstrate MaRDIMark in action, we have begun developing a benchmark tool for the field of MOR, relying on the curated benchmark database hosted in the MORWiki [13].

#### 3.1 MORWiki

MORWiki was launched in 2012 as a platform for the MOR community. It hosts established benchmark collections like the Oberwolfach Collection [11] and SLICOT [14, 15] and new datasets. Currently, approximately 100 distinct datasets have been documented for a variety of linear, nonlinear, and parametric real-world and test examples. Various methods have also been documented, along with a comparison table for MOR software and a BibTeX file for MOR literature. Since the start of MaRDI, a searchable database for problem metadata has been compiled, along with an initial ontology that will be integrated into AlgoData and the MaRDIPortal<sup>7</sup>. Eventually, a search tool will be published that allows users to look up benchmarks with desired properties.

A number of challenging tasks remain open for MORWiki. In particular, licenses are missing for older datasets like SLICOT and the Oberwolfach, despite the fact that they have been treated as if they were part of the public domain for nearly 20 years. This poses a unique legal situation, especially if we upload updated versions of the data, to the MORWIKI Zenodo community<sup>8</sup> that conform to the file format standards of MORB (cf. Section 3.2). Furthermore, it is important that such a database is built by the larger community, even if the overall editorial work is maintained by a small group. Incentivizing researchers to not only publish their data, but also prepare it in a specific format to improve interoperability is difficult. A key selling point is that papers with public code and data accumulate higher citations over time [4]. Finally, much work remains regarding the mathematical metadata for MORWiki benchmarks. Systems theory enjoys a plethora of problem varieties, which must often be determined by hand or via numerical eigensolvers, which can be prohibitively expensive for large systems.

#### 3.2 Design specifications

For the initial prototype of MORB, we have aimed to keep things as simple as possible. We have focused on linear, time-invariant, first-order systems (LTI-FOS) of the form

$$\begin{aligned} E\dot{x}(t) &= Ax(t) + Bu(t), \\ y(t) &= Cx(t) + Du(t), \end{aligned}$$

where  $A, E \in \mathbb{R}^{n \times n}$ ,  $B \in \mathbb{R}^{n \times m}$ ,  $C \in \mathbb{R}^{q \times n}$ , and  $D \in \mathbb{R}^{q \times m}$ . When  $E$  and  $C$  are not explicitly specified, the identity matrix is assumed for both; when  $D$  is not specified, it is assumed to be zero. For interoperability, we require that all the matrices for a given benchmark are stored as `.mat` files, version 7, which is supported by MATLAB<sup>®</sup><sup>9</sup> and SciPy<sup>10</sup>, as well as GNU Octave<sup>11</sup>, Julia<sup>12</sup>, and can be linked into most compiled languages via libmatio<sup>13</sup>. We have ensured that all LTI-FOS benchmarks in MORWiki are converted to such a format, and will be made public<sup>7</sup>, as soon as the open licensing questions have been resolved.

A key component of MORB is the configuration file. It is written in JSON<sup>14</sup>, a format compatible with most widely used programming languages, and is relatively straightforward to set up and modify.

<sup>7</sup><https://portal.mardi4nfdi.de/>

<sup>8</sup><https://zenodo.org/communities/morwiki/>

<sup>9</sup>[https://de.mathworks.com/help/matlab/import\\_export/mat-file-versions.html](https://de.mathworks.com/help/matlab/import_export/mat-file-versions.html)

<sup>10</sup><https://docs.scipy.org/doc/scipy/reference/generated/scipy.io.loadmat.html>

<sup>11</sup><https://octave.org/>

<sup>12</sup><https://julialang.org>

<sup>13</sup><https://github.com/telehan/libmatio>

<sup>14</sup><https://www.json.org/>

The highest level must correspond to the filename (i.e., benchmark ID). The secondary level should minimally have the "alg\_iso" field and optional fields

- "meas\_opt" (options for Analyzer module);
- "bode\_opt" (options for Bode plots; part of Explorer module);
- "plot\_opt" (general options for plots; part of Explorer module); and
- "report\_opt" (options for final report; part of Explorer module).

See Figure 2 for an example configuration file. In this example, the solver `bt` (Balanced Truncation) is being configured for the packages `cst` (Control Systems Toolbox)<sup>15</sup>, `emgr` (Empirical Grammian Framework) [5], `mess` (Matrix Equation Sparse Solver) [1], `morlab` (Model Order Reduction Laboratory) [2], and `pymor` (Model Order Reduction with Python) [10]. Depending on the package, different parameters like `tol` (Hankel singular value tolerance) and `max_order` (maximum order of the reduced model) may be set. Note that actually two algorithm isotopes are specified for both the `cst` and `mess` packages, because two different `tol` options have been provided. Because `emgr` is set to `null`, six total algorithm isotopes will be run for `bt`.

For this version of MORB, the encasing workflow has been written in MATLAB. This choice has allowed us to minimize interface issues, as most MOR software for this class of models is written in MATLAB<sup>16</sup>. Furthermore, the MATLAB package MORLAB contains well documented routines for computing standard errors in the  $L_0$ ,  $L_1$ ,  $L_2$ ,  $L_\infty$  and  $\mathcal{H}^2$  norms, as well as Bode, Frobenius, and Sigma plots, which we have relied on for the Analyzer module.

### 3.3 Early prototype and sample reports

A prototype of MORB that works on the steel profile benchmark [12] can be found on Zenodo [8]. Furthermore, sample reports generated from other datasets are included there that demonstrate what the Explorer module looks like. Details about the run itself— a timestamp, the benchmark ID, operating system, software version, and problem size— are reported, along with a table and graph of the run times and a table of error norms, as well as error, Bode, Sigma, and Frobenius plots. See Figure 3 and Table 1 for a preview of the full report. In this version of MORB, reports are auto-generated as TeX files, which can then be compiled as PDFs. Additional details about the run are stored in the comments of the TeX file, generated by the MATLAB command `system('systeminfo')`<sup>17</sup>.

<sup>15</sup>[www.mathworks.com/products/control.html](http://www.mathworks.com/products/control.html)

<sup>16</sup>[https://morwiki.mpi-magdeburg.mpg.de/morwiki/index.php/Comparison\\_of\\_Software](https://morwiki.mpi-magdeburg.mpg.de/morwiki/index.php/Comparison_of_Software)

<sup>17</sup>This command only works on Windows. <https://www.mathworks.com/help/matlab/ref/system.html>

```

{
  "newEngland_n66m1q1":
  {
    "alg_iso":
    {
      "bt":
      {
        "cst":
        [
          {
            "tol": 1E-6
          },
          {
            "tol": 1E-12
          }
        ],
        "emgr": null,
        "mess":
        [
          {
            "max_order": 100,
            "tol": 1E-6
          },
          {
            "tol": 1E-12
          }
        ],
        "morlab":
        {
          "tol": 1E-6
        },
        "pymor":
        {
          "tol": 1E-6
        }
      },
      "bg":
      {
        "emgr": null,
        "mess": null
      }
    },
    "meas_opt":
    {
      "norm_id": ["l0", "l1", "l2", "linf", "h2"],
      "time_points": 250,
      "h2_method": "lyap",
      "ml_bodemag":
      {
        "FreqRange": [-8,8],
        "ShowPlot": 0,
        "MaxPoints": 500
      },
      "ml_sigmaplot":
      {
        "FreqRange": [-8,8],
        "ShowPlot": 0,
        "MaxPoints": 500
      },
      "ml_frobeniusplot":
      {
        "FreqRange": [-8,8],
        "ShowPlot": 0,
        "MaxPoints": 500
      }
    },
    "bode_opt":
    {
      "FreqRange": [-8,8],
      "ShowPlot": 0,
      "MaxPoints": 500
    },
    "plot_opt":
    {
      "save_eps": true,
      "save_fig": true
    }
  }
}

```

Figure 2: Example of a configuration file for the benchmark problem `newEngland_n66m1q1`.

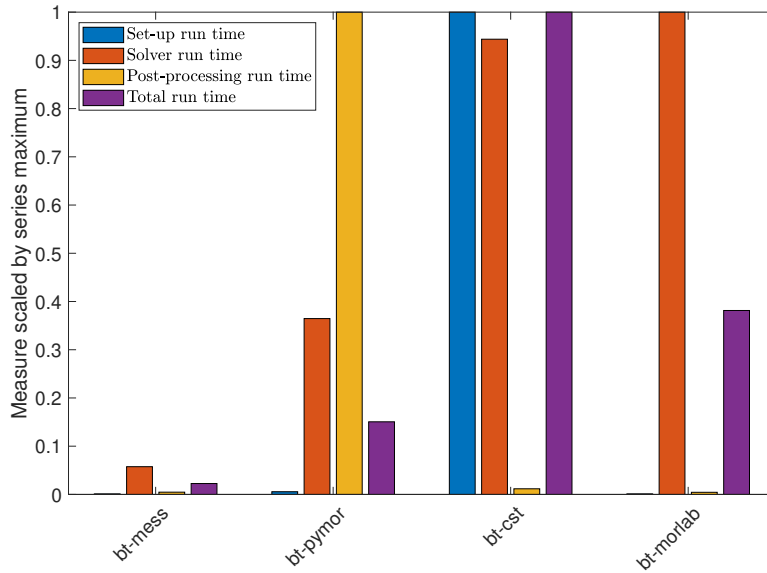


Figure 3: Sample run time bar chart for `steelProfile_n1357m7q6`. Note that timings are scaled by the maximum in the series.

Algorithm Isotope	$L_2$	$L_\infty$	$H_2$
bt-mess	2.23e-06	2.83e-06	4.66e-06
bt-pymor	1.56e-05	1.78e-05	2.92e-05
bt-cst	2.45e-04	1.95e-04	2.90e-03
bt-morlab	2.86e-06	3.00e-06	5.55e-06

Table 1: Sample error table for `steelProfile_n1357m7q6`.

## 4 Conclusions and Outlook

We have demonstrated the versatility of MaRDIMark on a curated database of MOR benchmarks. The resulting tool, MORB, is an initial attempt to compare solvers written in different languages with different syntax and parameter-parsing standards for a large class of benchmark examples. There is still much room for improvement. First of all, we plan to refactor it in Python, a free, open-source language with a large, active user-base. This will ensure we comply with FAIR principles, especially accessibility and reproducibility, which the MATLAB version currently violates due to the requirement for paid licenses. Secondly, the configuration file as described in Section 3.2 will be updated to remove redundancies in the hierarchy, to reflect better the modular design of the tool, and to enable parallelization for large numbers of algorithm isotopes. In addition, we plan to integrate more algorithms from the field to allow for benchmarks on a wider variety of problems, particularly nonlinear and parametric examples. Finally, the report itself will be updated to allow for a wider variety of formats (e.g., HTML and Markdown) and to incorporate progress from (M4) and MaRDIFlow [16]

Principles from MaRDIMark can already be seen in other benchmark tools. One of the co-authors has actively incorporated the modular design in the stability analysis of communication-avoiding Gram-Schmidt and Krylov subspace methods [6, 7]. Within MaRDI, TA3 has developed `m1r3`, which provides a unified interface for machine learning in the R language [3]. The toolkit includes a benchmark fea-

ture<sup>18</sup>, which generates a grid of learner configurations, runs them on a set of tasks, and returns performance results that can be explored as a table or box plot. Another well-developed open-source tool is `FitBenchmarking` [9], which compares fitting and minimizer frameworks from software that is callable from Python. The primary output is an interactive table, allowing users to quickly navigate results for interesting solver and benchmark combinations. A key feature of both `mlr3` and `FitBenchmarking` is that they have been developed long-term by large teams of domain experts and software engineers. Furthermore, both rely on curated benchmark databases established and maintained by their respective academic communities, and they both provide pathways for users to contribute new software to be benchmarked.

## Acknowledgments

We would like to thank Christian Himpe and Tim Mitchell for conversations that laid the foundation of the benchmark tool, as well as Alexander Stage and Malte Speidel, for work on the MORB ontology and search tool. We further thank Martin Köhler for helping debug MORB.

## References

- [1] P. Benner, M. Köhler, and J. Saak. Matrix equations, sparse solvers: M-M.E.S.S.-2.0.1 – philosophy, features and application for (parametric) model order reduction. In P. Benner, T. Breiten, H. Faßbender, M. Hinze, T. Stykel, and R. Zimmermann, editors, *Model Reduction of Complex Dynamical Systems*, volume 171 of *International Series of Numerical Mathematics*, pages 369–392. Birkhäuser, Cham, 2021. doi:10.1007/978-3-030-72983-7\_18.
- [2] P. Benner and S. W. R. Werner. Frequency- and time-limited balanced truncation for large-scale second-order systems. *Linear Algebra Appl.*, 623:68–103, 2021. doi:10.1016/j.laa.2020.06.024.
- [3] B. Bischl, R. Sonabend, L. Kotthoff, and M. Lang. Mlr3: Machine Learning in R. <https://mlr3.ml-org.com/>, 2023.
- [4] G. Colavizza, I. Hrynaszkiewicz, I. Staden, K. Whitaker, and B. McGillivray. The citation advantage of linking publications to research data. *PLOS ONE*, 15(4):e0230416, 2020. doi:10.1371/journal.pone.0230416.
- [5] C. Himpe. emgr – EMpirical GRamian framework (version 5.99). <https://gramian.de>, 2022. doi:10.5281/zenodo.6457616.
- [6] K. Lund. LowSyncBlockArnoldi. <https://gitlab.mpi-magdeburg.mpg.de/lund/low-sync-block-arnoldi/-/releases/v1.0>, 2022.
- [7] K. Lund, E. Carson, and E. Oktay. BlockStab. <https://github.com/katlund/BlockStab/releases/tag/v2.2023-beta>, 2023.
- [8] K. Lund, J. Saak, and P. Benner. mini-MORB (Model Order Reduction Benchmark). <https://zenodo.org/record/8093833>, 2023. doi:10.5281/zenodo.8093833.
- [9] A. Markvardsen, T. Rees, M. Wathen, A. Lister, J. Farmer, R. Farooq, J. Kitenge, J. Wilkins, S. Hanrahan, B. Peters, R. Applin, J. Fowkes, P. Odagiu, A. Anuchitanukul, T. Farmer, A. Lim, F. M. Pouzols, T. Snow, and A. McCluskey. FitBenchmarking: An open source Python package comparing data fitting software. <https://github.com/fitbenchmarking/fitbenchmarking>, 2022.

<sup>18</sup>[https://mlr3book.ml-org.com/chapters/chapter3/evaluation\\_and\\_benchmarking.html](https://mlr3book.ml-org.com/chapters/chapter3/evaluation_and_benchmarking.html)



- 
- [10] R. Milk, S. Rave, and F. Schindler. pyMOR – generic algorithms and interfaces for model order reduction. *SIAM J. Sci. Comput.*, 38(5):S194–S216, 2016. doi:10.1137/15M1026614.
- [11] Oberwolfach model reduction benchmark collection. available at <http://portal.uni-freiburg.de/imteksimulation/downloads/benchmark>.
- [12] J. Saak and M. Behr. The Oberwolfach Steel-Profile benchmark revisited. <https://zenodo.org/record/5113560>, 2021. doi:10.5281/zenodo.5113560.
- [13] The MORwiki Community. MORwiki - Model Order Reduction Wiki. <http://modelreduction.org>.
- [14] A. Varga. Task II. B.1 – selection of software for controller reduction. SLICOT Working Note 1999–18, The Working Group on Software (WGS), Dec. 1999. available from [www.slicot.org](http://www.slicot.org).
- [15] A. Varga. Model reduction software in the SLICOT library. In B. N. Datta, editor, *Applied and Computational Control, Signals, and Circuits*, volume 629 of *The Kluwer International Series in Engineering and Computer Science*, pages 239–282. Kluwer Academic Publishers, Boston, MA, 2001.
- [16] P. Veluvali, J. Heiland, and P. Benner. MaRDIFlow: A Workflow Framework for Documentation and Integration of FAIR Computational Experiments. <https://zenodo.org/record/7863520>, 2023. doi:10.5281/zenodo.7863520.