
PhD project - A system identification framework for globally optimal models

A Data Management Plan created using DMPonline.be

Creators: Sibren Lagauw, Aldona Niemirow-Sznajder

Affiliation: KU Leuven (KUL)

Funder: Fonds voor Wetenschappelijk Onderzoek - Research Foundation Flanders (FWO)

Template: FWO DMP (Flemish Standard DMP)

Data Manager: Aldona Niemirow-Sznajder

Grant number / URL: FWO - 11K5623N

ID: 193491

Start date: 01-10-2021

End date: 30-09-2024

Project abstract:

System identification constructs dynamical models from measured data. These mathematical models are used for simulation, prediction, monitoring, classification, or control tasks. State-of-the-art identification methods, such as prediction-error methods (PEMs) and subspace methods are suboptimal: PEMs rely on numerical optimization algorithms for the parameter identification, where many heuristics prevail, while subspace identification is based on projections and is not provably optimal. In a new unifying misfit versus latency (MvL) framework, we deal with inexact data by modifying the data (in least-squares sense) with misfits and adding unobservable, latent inputs. The MvL identification problem is a multivariate polynomial optimization problem, that can be solved without any heuristics by means of a multiparameter eigenvalue problem (MEVP), resulting in globally optimal models. The main research objectives (ROs) of this proposal are: (RO1) deepening our mathematical understanding of the MvL framework and studying how the problem is related to (and may benefit from) operator theory, algebraic geometry and (behavioral) systems theory, (RO2) extending the MvL approach to the multi-input multi-output (MIMO) case, and (RO3) developing tailored numerical methods to perform the involved structured matrix computations.

Last modified: 07-02-2023

PhD project - A system identification framework for globally optimal models

FWO DMP (Flemish Standard DMP)

1. Research Data Summary

List and describe all datasets or research materials that you plan to generate/collect or reuse during your research project. For each dataset or data type (observational, experimental etc.), provide a short name & description (sufficient for yourself to know what data it is about), indicate whether the data are newly generated/collected or reused, digital or physical, also indicate the type of the data (the kind of content), its technical format (file extension), and an estimate of the upper limit of the volume of the data.

				Only for digital data	Only for digital data	Only for digital data	Only for physical data
Dataset Name	Description	New or reused	Digital or Physical	Digital Data Type	Digital Data format	Digital data volume (MB/GB/TB)	Physical volume
		<i>Please choose from the following options:</i> <ul style="list-style-type: none"> • Generate new data • Reuse existing data 	<i>Please choose from the following options:</i> <ul style="list-style-type: none"> • Digital • Physical 	<i>Please choose from the following options:</i> <ul style="list-style-type: none"> • Observational • Experimental • Compiled/aggregated data • Simulation data • Software • Other • NA 	<i>Please choose from the following options:</i> <ul style="list-style-type: none"> • .por, .xml, .tab, .csv, .pdf, .txt, .rtf, .dwg, .gml, ... • NA 	<i>Please choose from the following options:</i> <ul style="list-style-type: none"> • <100MB • <1GB • <100GB • <1TB • <5TB • <10TB • <50TB • >50TB • NA 	
Synthetic numerical data	Periodic signals (sin-waves) perturbed with noise (random number generator), signals generated using system theoretic (mathematical) models.	Generate new data	Digital	Simulation data, software	.mat, .dat	<100MB	

If you reuse existing data, please specify the source, preferably by using a persistent identifier (e.g. DOI, Handle, URL etc.) per dataset or data type:

No existing data will be reused.

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? Describe these issues in the comment section. Please refer to specific datasets or data types when appropriate.

- No

N.A.

Will you process personal data? If so, briefly describe the kind of personal data you will use in the comment section. Please refer to specific datasets or data types when appropriate.

- No

Does your work have potential for commercial valorization (e.g. tech transfer, for example spin-offs, commercial exploitation, ...)? If so, please comment per dataset or data type where appropriate.

- No

Do existing 3rd party agreements restrict exploitation or dissemination of the data you (re)use (e.g. Material/Data transfer agreements/ research collaboration agreements)? If so, please explain in the comment section to what data they relate and what restrictions are in place.

- No

Are there any other legal issues, such as intellectual property rights and ownership, to be managed related to the data you (re)use? If so, please explain in the comment section to what data they relate and which restrictions will be asserted.

- No

2. Documentation and Metadata

Clearly describe what approach will be followed to capture the accompanying information necessary to keep data understandable and usable, for yourself and others, now and in the future (e.g., in terms of documentation levels and types required, procedures used, Electronic Lab Notebooks, README.txt files, Codebook.tsv etc. where this information is recorded).

All files will be documented *in-line*, so that the published numerical examples can be reproduced while making sure the end-user knows what is happening in the code. Each file will contain a reference to the publication that describes the methodology. The overall repository contains a README.txt such guiding the user through the folders/files.

Will a metadata standard be used to make it easier to find and reuse the data? If so, please specify (where appropriate per dataset or data type) which metadata standard will be used. If not, please specify (where appropriate per dataset or data type) which metadata will be created to make the data easier to find and reuse.

- No

3. Data storage & back-up during the research project

Where will the data be stored?

The algorithms and codes will be kept on Gitlab repository.

How will the data be backed up?

When a work package is finished, the dataset moves to the STADIUS Dataset Server which is regularly backed up.

**Is there currently sufficient storage & backup capacity during the project? If yes, specify concisely.
If no or insufficient storage or backup capacities are available, then explain how this will be taken care of.**

- Yes

The storage requirements are small as no raw-data has to be saved. The STADIUS dataset server has sufficient storage capacities.

How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

The STADIUS Dataset Server is maintained carefully and *write access* to the Gitlab repository is kept privately.

What are the expected costs for data storage and backup during the research project? How will these costs be covered?

The costs will be negligible.

4. Data preservation after the end of the research project

Which data will be retained for at least five years (or longer, in agreement with other retention policies that are applicable) after the end of the project? In case some data cannot be preserved, clearly state the reasons for this (e.g. legal or contractual restrictions, storage/budget issues, institutional policies...).

We will not deviate from the principle of data preservation. All files will be preserved for at least five years.

Where will these data be archived (stored and curated for the long-term)?

The STADIUS Dataset Server which is regularly backed up.

What are the expected costs for data preservation during the expected retention period? How will these costs be covered?

These costs for storage on a personal device are negligible, the costs involved with the maintenance of the STADIUS dataset server are covered by STADIUS.

5. Data sharing and reuse

Will the data (or part of the data) be made available for reuse after/during the project? In the comment section please explain per dataset or data type which data will be made available.

- Yes, in an Open Access repository

We will make files and code available in an open access repository (public Gitlab repository) so that all published numerical examples can be reproduced.

If access is restricted, please specify who will be able to access the data and under what conditions.

Data corresponding to *work-in-progress* and/or unpublished results will be kept private, only accessible for researchers of our group.

Are there any factors that restrict or prevent the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)? Please explain in the comment section per dataset or data type where appropriate.

- No

N.A.

Where will the data be made available? If already known, please provide a repository per dataset or data type.

Publicly available Gitlab repository, see e.g., <https://gitlab.esat.kuleuven.be/Christof.Vermeersch/macaulaylab-public>

When will the data be made available?

Once the results are published and/or the PhD-project is finished.

Which data usage licenses are you going to provide? If none, please explain why.

N.A.

Do you intend to add a PID/DOI/accession number to your dataset(s)? If already available, you have the option to provide it in the comment section.

- No

N.A.

What are the expected costs for data sharing? How will these costs be covered?

The costs of maintaining the Gitlab repository will be covered by STADIUS.

6. Responsibilities

Who will manage data documentation and metadata during the research project?

Myself (Sibren Lagauw)

Who will manage data storage and backup during the research project?

Myself (Sibren Lagauw)

Who will manage data preservation and sharing?

Myself (Sibren Lagauw), in collaboration with the STADIUS data manager.

Who will update and implement this DMP?

Myself (Sibren Lagauw)

