Sensitivity-based uncertainty quantification for partially stochastic plasma edge simulations of nuclear fusion devices

A Data Management Plan created using DMPonline.be

Creator: Niels Horsten

Affiliation: KU Leuven (KUL)

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

Template: FWO DMP (Flemish Standard DMP)

Grant number / URL: 12AES24N

ID: 206249

Start date: 01-11-2023

End date: 31-10-2026

Project abstract:

Plasma edge simulations are extensively used for the interpretation of exhaust scenarios in current nuclear fusion devices and the design of future reactors. Multiple uncertainties in the input data and model parameters lead to errors on the output quantities of interest. Hence, there is a need for a thorough uncertainty quantification (UQ). The computational cost of plasma edge simulations prevents the use of standard UQ methods such as basic Monte Carlo (MC) and stochastic expansion methods. Past research has shown that sensitivity-based UQ techniques can drastically reduce the number of forward simulations. Therefore, I will develop sensitivity-based UQ methods for plasma edge modeling.

A major challenge is the efficient and accurate calculation of the sensitivities with respect to the input uncertainties in the presence of statistical noise originating from the kinetic MC simulation of the neutral particles. I will exploit algorithmic differentiation (AD) tools for a semi-automatic, correlation-preserving differentiation of the complex stochastic model in combination with averaging techniques to minimize the statistical error. AD in *adjoint* mode is of main interest as its computational cost for sensitivity calculation is independent of the number of input uncertainties, but existing memory
reduction techniques need to be adapted for partially stochastic plasma edge simulations. Finally, I will accelerate the simulations with a hybrid fluid-kinetic neutral approach.

Last modified: 29-04-2024

Sensitivity-based uncertainty quantification for partially stochastic plasma edge simulations of nuclear fusion devices

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
		Please choose from the following options: • Generate new data • Reuse existing data	*Please choose from the following options:* • Digital • Physical	*Please choose from the following options:* • Observational • Experimental • Compiled/aggregated data • Simulation data • Software • Other • NA	*Please choose from the following options:* • .por, .xml, .tab, .csv,.pdf, .txt, .rtf, .dwg, .gml, … • NA	*Please choose from the following options:* • <100MB • <1GB • <100GB • <1TB • <5TB • <10TB • <50TB • >50TB • NA	
SOLPS-ITER code	Fortran source code for simulations	• Generate new data • Reuse existing data	Digital	Software	Text files in syntax of programming language (typically .F, .f, .F90 or .f90)	<100MB	
SOLPS-ITER code documentation	Latex and Word files describing the physics in the code and the code options	• Generate new data • Reuse existing data	Digital	Other	Latex .tex source files and corresponding pdf files Word documents for tutorials	<100MB	
PESDT code	Python code for postprocessing	• Generate new data • Reuse existing data	Digital	Software	Text files in syntax of programming language (typically .py)	<100MB	
Postprocessing code linked to SOLPS-ITER	Matlab tools for postprocessing	• Generate new data • Reuse existing data	Digital	Software	Text files in syntax of programming language (typically .m)	<100MB	

Simulation data	Computed solutions from the simulation software with corresponding input files	• Generate new data	Digital	Simulation data	Text and binary files specific to SOLPS-ITER and common to all SOLPS-ITER users	<1TB	
Experimental data	Measurement data from the JET tokamak	• Reuse existing data	Digital	Experimental	Easily interpretable ASCII format (.dat extension)	<1GB	
Postprocessing tools for experimental data	Matlab and Python scripts to plot the data	• Generate new data	Digital	Software	Text files in syntax of programming language (typically .m or .py)	<100MB	
Research output	All kinds of scientific output: papers, presentations and additional documents with corresponding figures	• Generate new data	Digital	Other	Latex and Word files (.tex and .docx) with corresponding .pdf files, powerpoint (.ppt) files for presentations	<1GB	

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:

- The SOLPS-ITER simulation software, including documentation and postprocessing tools are all part of the ITER GIT repository: https://git.iter.org/projects/BND/repos/solps-iter/browse.
- The PESDT code is available at Github: https://github.com/lomanowski/PESDT.
- The JET experimental data is available at the JDC (JET data centre) servers. There, the validated data can be written out to ASCII files.

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

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.

- Yes

- SOLPS-ITER code agreement with ITER does not allow dissemination/redistribution of the code developed.
- Experimental data from JET cannot be redistributed without permission.

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).

- SOLPS-ITER code: version control on ITER GIT repository. There are different parts of the code that will be used:
  - Top SOLPS-ITER repository: https://git.iter.org/projects/BND/repos/solps-iter/browse
  - B2.5 plasma code: https://git.iter.org/projects/BND/repos/b2.5/browse
  - EIRENE kinetic neutral code: https://git.iter.org/projects/BND/repos/eirene/browse
  - DivGeo to set up cases: https://git.iter.org/projects/BND/repos/divgeo/browse
  - Carre mesh generator: https://git.iter.org/projects/BND/repos/carre/browse
- SOLPS-ITER code documentation: part of the SOLPS-ITER top GIT repository
- PESDT code: version control on Github repository: https://github.com/lomanowski/PESDT
- Postprocessing code linked to SOLPS-ITER: part of the top SOLPS-ITER GIT repository
- Simulation data: documentation of the simulations with GIT hash of the code version, the required input files and a description of how to obtain the results. To this end, standardized storage scripts for archiving of important data will be used. Store on KU Leuven RDR repository inputs and relevant outputs for reproducibility of data linked to publications.
- Experimental data: the ASCII experimental data files will be stored on the JET data servers. If permitted, the data used for publications and presentations will be placed on KU Leuven RDR repository.
- Postprocessing tools for experimental data: they will be added to the RDR repository
- Research output: there will be made an entry on RDR for each publication/presentation linking the simulation and experimental data and the postprocessing scripts to the corresponding paper.

Within our group, we already have an archive space for long-term storage on the VSC cluster, but the idea is to start exploiting "ManGO" for active data storage and management.

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

For KU Leuven RDR and ManGO the format of the metadata will be tailored to our specific needs of computational experiments.

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

Where will the data be stored?

- File synchronization (devices/collaborators) and backup: for version-controlled files, we use the KU Leuven Gitlab service or ITER GIT repository. For other files, we use KU Leuven OneDrive.
- Large simulation results are stored on a dedicated archive on the VSC cluster.
- When possible, all relevant data will be made publicly available with the paper. For this, we will use a coupling between ManGO and the KU Leuven RDR system that is designed for this purpose.
- Manuscripts and all corresponding data are stored on an archive drive in the research group and/or on an archive drive at the VSC depending on the nature of the research data.

How will the data be backed up?

All solutions described above are equipped with automated back-up procedures that are provided by the central university services. Personal laptops of researchers are synced with their home directory on the departmental servers. These servers are backed up as per the above.

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

KU Leuven OneDrive services and dedicated archive storage on the VSC (presently 4TB archive storage available) suffice to store the data. If needed, the VSC archive space can be extended.

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

All systems described above are configured to prohibit access to unauthorized persons. All data of computational experiments stored in ManGO will be read-only to avoid accidental changes by authorized persons.

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

We expect the data management cost can be covered by the working budget of this FWO grant and other research projects within the group.

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...).

All simulation code, simulation data, scripts and figures/tables that are needed to reproduce published results will be available in open access repositories and findable via a link in the publication. All other data (e.g. on the archive or ManGO) will be retained as long as needed, and at least for five years after the end of the project.

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

- Simulation inputs and relevant outputs, postprocessing tools will be published on the KU Leuven RDR repository in open access
- Manuscripts and all corresponding data are stored on an archive drive in the research group and/or on an archive drive at the VSC depending on the nature of the research data.

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

- Publishing on RDR requires no costs.
- We expect data management costs to be eligible costs from the working budget of this or other research projects within the research group.
- Archive storage at the VSC currently costs 100 EUR/TB/year (with 4TB available to the project).

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 a restricted access repository (after approval, institutional access only, …)
- Yes, in an Open Access repository

- SOLPS-ITER code, including documentation and postprocessing tools: ITER repository has restricted access
- PESDT code: open access GIT repository
- Simulation data: open access repository
- Experimental data: restricted access repository
- Research output: restricted access repository

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

- SOLPS-ITER code, including documentation and postprocessing tools: accessible to all institutions that have signed the agreement with the ITER Organization, with contact person Xavier Bonnin
- Experimental data: all people that have a JDC account. The JDC account can be requested by all instituted linked to EUROfusion.
- Research output: directly accessible by other members of the KU Leuven TFE research 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.

- Yes, Intellectual Property Rights

- SOLPS-ITER code: software is subject to IMAS agreement (contact Simon Pinches or Xavier Bonnin at ITER)
- JET data is in general not shareable

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

- SOLPS-ITER code, including documentation and postprocessing tools: ITER GIT repository, access to be granted by the institution

When will the data be made available?

Upon publication of research results

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

Datasets: https://creativecommons.org/licenses/by-nc-sa/4.0/
Code: https://www.gnu.org/licenses/agpl-3.0.en.html

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.

- Yes

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

Open access on RDR is free of charge for KU Leuven affiliates.
We expect other data management costs to be eligible costs from the working budget of research projects.

6. Responsibilities

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

Niels Horsten

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

Niels Horsten

Who will manage data preservation and sharing?

Niels Horsten, followed up by the responsible for data management in the group (Wouter Dekeyser for now), if/when Niels Horsten will leave the research group.

Who will update and implement this DMP?

Niels Horsten