

---

## Towards validated mean-field descriptions of anomalous transport in the tokamak plasma edge for reactor-relevant regimes

*A Data Management Plan created using DMPonline.be*

**Creator:** Olivier Renders

**Affiliation:** KU Leuven (KUL)

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

**Template:** FWO DMP (Flemish Standard DMP)

**Grant number / URL:** 1SHII24N

**ID:** 204374

**Start date:** 01-11-2023

**End date:** 31-10-2027

### Project abstract:

While the cross-field transport in the plasma edge is dominated by turbulent processes, mean-field transport codes are still the workhorses for the design of the exhaust system of tokamaks due to the computational cost of turbulence simulations and the physical processes they do not yet contain. In the mean-field codes the average effect of the turbulent transport is mostly modeled through ad-hoc diffusive relations in which the diffusion coefficients are specified by the user. This practice severely limits the interpretive and predictive capabilities of these codes. Recent research has taken inspiration from the Reynolds-Averaged-Navier-Stokes approach used in hydrodynamic turbulence modeling to propose self-consistent mean-field models for the cross-field transport. Thus far this research has however focused on the sheath limited regime and did not consider plasma-neutral interactions. This PhD project aims to extend the validity of the models to reactor-relevant conditions by treating also the conduction limited, high recycling regime and including the neutrals, as a prerequisite towards detached conditions. A self-consistent mean-field model will be derived by time-averaging of the turbulence equations and remaining closure terms will be modeled. Unknown coefficients in these closures will be determined by an in-house Bayesian Parameter Estimation tool. The models will be implemented in SOLPS-ITER and validated against 3D turbulence simulations and experimental data.

**Last modified:** 30-04-2024

**Towards validated mean-field descriptions of anomalous transport in the tokamak plasma edge for reactor-relevant regimes**  
**Application DMP**

---

**Questionnaire**

**Describe the datatypes (surveys, sequences, manuscripts, objects ... ) the research will collect and/or generate and /or (re)use. (use up to 700 characters)**

Question not answered.

**Specify in which way the following provisions are in place in order to preserve the data during and at least 5 years after the end of the research? Motivate your answer. (use up to 700 characters)**

Question not answered.

**What's the reason why you wish to deviate from the principle of preservation of data and of the minimum preservation term of 5 years? (max. 700 characters)**

Question not answered.

**Are there issues concerning research data indicated in the ethics questionnaire of this application form? Which specific security measures do those data require? (use up to 700 characters)**

Question not answered.

**Which other issues related to the data management are relevant to mention? (use up to 700 characters)**

Question not answered.

**Towards validated mean-field descriptions of anomalous transport in the tokamak plasma edge for reactor-relevant regimes**

**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"> <li>• Generate new data</li> <li>• Reuse existing data</li> </ul>	<i>Please choose from the following options:</i> <ul style="list-style-type: none"> <li>• Digital</li> <li>• Physical</li> </ul>	<i>Please choose from the following options:</i> <ul style="list-style-type: none"> <li>• Observational</li> <li>• Experimental</li> <li>• Compiled/aggregated data</li> <li>• Simulation data</li> <li>• Software</li> <li>• Other</li> <li>• NA</li> </ul>	<i>Please choose from the following options:</i> <ul style="list-style-type: none"> <li>• .por, .xml, .tab, .csv, .pdf, .txt, .rtf, .dwg, .gml, ...</li> <li>• NA</li> </ul>	<i>Please choose from the following options:</i> <ul style="list-style-type: none"> <li>• &lt;100MB</li> <li>• &lt;1GB</li> <li>• &lt;100GB</li> <li>• &lt;1TB</li> <li>• &lt;5TB</li> <li>• &lt;10TB</li> <li>• &lt;50TB</li> <li>• &gt;50TB</li> <li>• NA</li> </ul>	
(1a) TOKAM2D source code	source files for TOKAM2D to extend model	Generate new data	Digital	Software	.f90	<100 MB	
(1b) TOKAM2D postprocessing scripts	MATLAB functions to post-process the raw T2D-data and extract requested mean-field	Generate new data	Digital	Software	.m (matlab postprocessing)	< 100 MB	
(1c) TOKAM2D visualization scripts	MATLAB scripts for visualizing results from (1b)	Generate new data	Digital	Software	.m	< 100 MB	
(1d) TOKAM2D raw data	raw output from TOKAM2D (containing snapshot of fields)	Generate new data	Digital	Simulation data	HDF5	< 5TB	

(1e) TOKAM2D averaged data	Output from (1b) containing mean-field results per T2D-run	Generate new data	Digital	Compiled/aggregated data	.mat	< 10 GB	
(1f) TOKAM2D input files	input files for running a TOKAM2D simulation	Generate new data	Digital	Simulation data	NA (linux)	< 100 MB	
(2a) SOLEDGE3X-post-processing scripts	Script(s) to calculate mean-field quantities / evaluate balances from SOLEDGE3X simulations, required for benchmarking model	Generate new data	Digital	Software	.py (python postprocessing) .HDF5	< 100 MB	
(2b) SOLEDGE3X visualization scripts	MATLAB or python scripts for visualizing results from (2a)	Generate new data	Digital	Software	.py or .m (to be determined)	< 100 MB	
(2c) SOLEDGE3X-reference-data	Turbulence data from SOLEDGE3X to benchmark model	Reuse existing data	Digital	Simulation data	.HDF5	> 50 TB	
(2d) SOLEDGE3X averaged data	Output from (2c) containing mean-field results per S3X-run	Generate new data	Digital	Compiled/aggregated data	.HDF5	< 100 GB	
(3a) SOLPS-ITER source code	Adaptations to SOLPS-ITER to include (to be) proposed RANS-models	Generate new data	Digital	Software	.f77	< 100MB	
(3b) SOLPS-ITER postprocessing functions + visualization scripts	Visualization scripts + functions used within it to analyze simulations	Generate new data	Digital	Software	.m	< 100 MB	
(3c) SOLPS-ITER input files	Input files to run SOLPS-ITER simulations	Generate new data	Digital	Simulation data	NA (multiple extensions, linux)	< 100 MB	
(3d) SOLPS-ITER raw data	raw data from SOLPS-ITER	Generate new data	Digital	Simulation data	NA	< 10 GB	
(4) Experimental data	Experimental data to validate the proposed models	Reuse existing data	Digital	Observational	<At the moment unknown: likely: .txt, .csv>	< 100 GB	

(5a) DivOptLight (DoL) Source code	source files for DoL to include extensions of the model	Generate new data	Digital	Software	.m	< 100 MB	
(5b) DoL visualization scripts	Visualization of DoL results, used for benchmarking proposed models in 1D context	Generate new data	Digital	Software	.m	< 100 MB	
(5c) DoL Input files & raw data	Input and raw output data of DoL	Generate new data	Digital	Simulation data	.mat	< 1 GB	
(6) Bayesian parameter estimation and model selection tool: source code	source code to extend the Bayesian tool (f.i. treatment of 2D iso 1D averaged profiles)	Generate new data	Digital	Software	.m	< 100 MB	

**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:**

SOLEEDGE3X raw simulation output data will be provided to me by Patrick Tamain from the research center CEA, Cadarache, France. The data (currently) does not at the moment have a persistent identifier and is not registered in any database.

The experimental data that will be provided to us, consists of both publicly available data from the TCV-X21 case [<https://github.com/SPCData/TCV-X21.git>] (and will also be for TCV-X23-case) and non-publicly available data from research centers, (likely) such as IPP-Garching from for instance Dominik Brida (who provided us with experimental data when I did my master thesis).

**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

The mean-field models that will be developed during this work, will be made available to the community by implementing them in SOLPS-ITER and through the publishing of papers. Those models could in principle be used by commercial parties to (adapt the) design (of) certain components of their machines. (However, they then require to have a valid license to use the SOLPS-ITER code, or need to implement the models themselves.)

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

The raw turbulence data from SOLEDGE3X was/is only shared with us to benchmark the (to be) proposed models against (i.e. personal use). We are therefore not allowed to further distribute that data.

Likewise, the non-publicly available experimental data that we will obtain from our partners cannot be further distributed by us.

Lastly, the TOKAM2D code to which I will add the model extensions, was provided to us also by Patrick Tamain and hence cannot be made available by us when publishing results.

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.

- Yes

The involved people from CEA have the IPR on the 3D turbulence data. However, they only required us to refer to their publication(s) in which the dataset(s) was/were used.

The experimental data provided to us is also considered to be part of the IP of the institution which delivered us the data.

## 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).

The source code (1a,3a,5a,6) and associated post-processing scripts in matlab/python (1b,2a,3b) form the basis of obtaining reproducible results, which is a precursor for understanding the data. Both modifications to the source codes and post-processing scripts will be version controlled by the use of GITLAB, which is hosted by KU Leuven. Furthermore the raw TOKAM2D-data also contains the githash of the compiled code-version. Through the use of comments in the code / post-processing scripts, the code should be understandable (after the usual familiarization with the code and structure).

To organize the simulation data from TOKAM2D (1d, 1e, 1f), we will create one main-folder per topic which will contain the subdirectories: 'raw\_data', 'processed\_data' and 'post-processing-scripts'. In the 'raw\_data' folder the different simulation cases will each have their own run-folder. We don't plan on letting the run-folders represent the changed parameters in a parameter-scan, but add a ReadMe\_parameters.txt file to clarify that mapping. Additionally a readMe.txt will be provided, which will additionally contain the git-hashes of the used versions (of the TOKAM2D-code and postprocessing script used) and a description of the simulation set-up. As a last resort, there is still the parameter-file (1f) present in those run-folders in case a next-user would still have some doubts.

For the raw SOLEDGE3X data (2c) we will create a latex-document based on the notes made in oneNote, which will describe the different stored fields (with units) and additionally, through the introduction of comments in the post-processing script(2a), the raw data could in principle be used by other people. The post-processed data (2d) will also receive a git-hash-input field and I will also foresee a document describing the meaning of the different output-fields. The visualization scripts (2b, 1c, 3b, 5b) will be provided with comments, which should help with the understanding of them.

For the experimental data(4) that we still need to receive, we will add a readMe.txt file to it to describe the obtained data (+ details on it's received data-analysis), when we obtained them and from which institution.

The SOLPS-data will be structured in a similar way as the TOKAM2D data.

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

We don't plan on adopting a metadata standard to keep it simple. Since only a fraction of the raw/processed data will be stored long-term, the reuse of data in a 'far-future' will require the regeneration of it. The stored ReadMe.txt file will contain the needed extra information for this

regeneration, while ReadMe\_parameters.txt is not strictly necessary. (Since for each simulation case, we will minimally store the input-file and the relevant restart-file.)

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

#### Where will the data be stored?

The source code and postprocessing scripts will (initially) be stored on Onedrive or the DATA-drive of the VSC, and will finally be stored on the gitlab-server of KU Leuven (after pushing).

The visualization scripts will (initially) be stored on the DATA-drive of the VSC and will additionally be stored in the ARCHIVE-drive of the VSC in case they were used for a paper/presentation/reporting.

Relevant simulation data will be stored on the archive drive of the VSC. The postprocessed data will initially only be stored on the SCRATCH-drive, since it is easy and fast to reproduce from the raw-data.

The raw data of SOLEDGE3X will be stored on a HPC-system of EUROfusion (at the moment Marconi) and the postprocessed data will both be stored there (temporary) and a selection of it will be moved to the ARCHIVE of the VSC-cluster.

Experimental data will both be stored in the DATA-drive and the ARCHIVE-drive of the VSC.

#### How will the data be backed up?

As mentioned in the previous question, a back-up of the data (except from SOLEDGE3X) is made through storing it on the ARCHIVE-drive of the VSC. (The data from SOLEDGE3X is backed up by CEA.)

The scripts stored on the DATA-drive of the VSC / OneDrive are automatically backed up (on an at least daily basis).

**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

Sufficient data storage is available, since I won't back up all simulation data from a TOKAM2D-run but only a subset of the snapshots (evenly spaced). The not stored snapshots can then easily be regenerated (in parallel!) from the stored ones.

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

While in the project no sensitive data will be collected, the secure storage of it is warranted through the permission-properties of the files.

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

We will use, as mentioned above, the ARCHIVE-drive of the VSC to back-up the data. Therefore, we expect that for most of the project ~ 3 TB will be sufficient to store the data (belonging to the then actual and the next subWP). (The data of the next but the next subWP will already be acquired and stored when finishing a subWP, to prevent needing to wait on the TOKAM2D simulations.)

Therefore we expect a cost of 210 euro/year for back-up of the data, which will be covered by the bench fee belonging to the FWO-grant.

### 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...).**

First, at KU Leuven we will need to store our research data for 10 years. However, we will only store a selection of our data at the end of the project (due to the storage/budget issue of storing the raw TOKAM2D data [while we can just reproduce it] and not owning the SOLEDGE3X data, which will be maintained by CEA).

Namely for the TOKAM2D simulations we will store for each subWP:

- the meta-data (i.e. the readMe-files),
- the parameter-, and a restart-file(s)\* for each simulation in the subWP,
- the raw data of the reference simulation of the subWP,
- the post-processed data will be stored.
- Note: (\*) the restart-file will be the raw data file from the moment we started time averaging the snapshots, and hence will allow to regenerate the relevant data without needing to simulate first the 'initial transient' of the turbulence simulation. Additionally, by storing a couple of restart files, we can reduce the time needed to regenerate the data. (Similar to how we foresee to handle the back-upping of the simulation data during the project.)

For the SOLPS-ITER simulations, we will store the input and relevant output files of each simulation, together with the metadata (the readMe-files). (I.e. We won't store the raw batch\_averages to calculate the statistical error, only the result of the statistical error-script.)

For SOLEDGE3X, the post-processed files will be stored, but not any raw input data, as was stated in the '3rd party arrangements'.

The experimental data, as already mentioned, will be stored including the meta-data.

(The source code is automatically stored for this period through the use of gitLab.)

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

The data will be stored on the ARCHIVE-drive of the VSC.

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

The expected cost for the data storage is relatively limited, through not storing all the raw TOKAM2D data. Making an educated guess for the storage requirement, we obtain:

50 MB / snapshot \* 1500 snapshots / reference sim \* 10 reference sim = 750 GB

So that ~ 1 TB should suffice, which equates to 350(/700) euro for the period of 5(/10) years. Part of the cost will be covered by the bench fee belonging to the FWO-grant (i.e. the 5 out of 10 year storage).

#### **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.**

- Other, please specify:
- Yes, in an Open Access repository

Upon request by email (and after a waiting period for regenerating the data), we can share the raw data which was used in a paper / PhD-thesis / ... However, the raw data won't be stored in an repository (due to its sheer size) and will need to be shared via other means. (Such as for instance Box.)

Additionally, the post-processed data used for a paper (and associated scripts) will be made available through KU Leuven's RDR (research data repository).

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

The raw data will be made accessible after we regenerate it after having received a request by email. When this happens, a license will need to be determined for which we will contact KU Leuven Research & Development.



**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, Other

The raw turbulence data from SOLEDGE3X cannot be shared by us, and this would also apply to the (to be) received experimental data that would not be publicly available. When publishing a paper involving the time-averaged turbulence data we will contact CEA to ask whether we can share the averaged data as is or need to limit it to the data used within the paper.

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

The data (and scripts) supporting the papers will be published in KU Leuven's RDR, as each researcher gets 50 GB/year for free.

**When will the data be made available?**

The raw and post-processed data will become available upon publication of the research results. (However, the raw data only upon request by mail.)

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

Question not answered.

**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

Datasets published in RDR get an DOI.

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

The expected cost for publishing the data underpinning the research papers is zero, as we get 50 GB/year for free at KU Leuven.

The cost of sharing the raw data is expected to be limited since at KU Leuven we have Belnet filesender (<https://admin.kuleuven.be/icts/e-mune/filessharing>) which allows us to transfer files up to (combined) 6 TB.

## 6. Responsibilities

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

Olivier Renders

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

Olivier Renders

**Who will manage data preservation and sharing?**

Olivier Renders and Wouter Dekeyser (for long term preservation / sharing in the long term)

**Who will update and implement this DMP?**

Olivier Renders