Deliverable 4.8: Report describing the result of the machine learning benchmark carried out during the WP DONUT

Work Package DONUT

The project leading to this application has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 847593.

Document information

Project Acronym	**EURAD**
Project Title	**European Joint Programme on Radioactive Waste Management**
Project Type	**European Joint Programme (EJP)**
EC grant agreement No.	**847593**
Project starting / end date	**1st June 2019 – 30 May 2024**
Work Package No.	**4**
Work Package Title	**D**evelopment and Improvement **O**f **NU**merical methods and **T**ools for modelling coupled processes
Work Package Acronym	**DONUT**
Deliverable No.	**4.8**
Deliverable Title	**Report describing the result of the machine learning benchmark carried out during the WP DONUT**
Lead Beneficiary	**ANDRA**
Contractual Delivery Date	**M52**
Actual Delivery Date	**M60**
Type	**Report**
Dissemination level	**PU**
Authors	

To be cited as:

Prasianakis N.I., Laloy E., Jacques D., Meeussen J.C.L., Tournassat C., Miron G.D., Kulik D. A., Idiart A., Demirer E., Coene E., Cochepin B., Leconte M., Savino M., Samper II J., De Lucia M., Yang C., Churakov S. V., Samper J., Kolditz O., Claret F., (2024). Report describing the result of the machine learning benchmark carried out during the WP DONUT. Final version as of 29.05.2024 of deliverable D4.4 of the HORIZON 2020 project EURAD. EC Grant agreement no: 847593.

Disclaimer

Acknowledgement

This document is a deliverable of the European Joint Programme on Radioactive Waste Management (EURAD). EURAD has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 847593.

Status of deliverable		
	By	Date
Delivered (Lead Beneficiary)	Andra	30/05/2024
Verified (WP Leader)	Francis CLARET	28/05/2024
Reviewed (Reviewers)	Bernd Grambow	29/05/2024
Approved (PMO)	Bernd Grambow	29/05/2024
Submitted to EC	Andra (Coordinator)	03/06/2024

Executive Summary

Due to recent technological developments, the fields of artificial intelligence and machine learning methods (ML) are growing at a very fast pace. The DONUT scientific community has recently started using ML for a) accelerating numerical simulations, b) multiscale and multiphysics couplings, c) uncertainty quantification and sensitivity analysis. There are first evidences, which suggest an overall acceleration of calculations between one to four orders of magnitude. Within DONUT a benchmark was designed to coordinate activities and test a variety of ML techniques relevant to geochemistry and reactive transport. It aimed at benchmarking the major geochemical codes, at generating high quality data for training/validation of existing/new ML methodologies and at providing basic guidelines about the benefits and drawbacks of using ML techniques. A joined publication will be submitted in the upcoming weeks to disseminate the conducted work.

Table of content

Executive Summary 4

Table of content 5

List of figures 6

1. INTRODUCTION 7

2. BENCHMARK RELEVANT TO MACHINE LEARNING AND GEOCHEMISTRY 8

2.1 Benchmark philosophy 8

2.2 Benchmark Roadmap and main milestones 9

2.3 Involved teams 10

2.4 High quality training datasets and problem set 10

2.5 Data Management and open research 11

2.6 Benchmarking of the geochemical solvers 13

2.7 Application and accuracy of predictions of ML algorithms 13

3. OUTLOOK 17

4. REFERENCES 18

List of figures

Figure 1*: In typical reactive transport simulations, thermodynamics and chemistry consume most of the computational time. Coupling of carefully trained surrogate models provides an overall acceleration of the simualation between one to four orders of magnitude* (Laloy and Jacques, 2019; Prasianakis et al., 2020)..... 8

Figure 2*: Benchmark timeline including major milestones and workshops*..... 9

Figure 3: *Research teams and contact persons involved in the geochemistry and machine learning benchmark*..... 10

Figure 4: *(top)The major geochemical solvers are operated by experts in the field, benchmarked and used to produce consistent training datasets for the machine learning techniques. (bottom) the 6 levels of complexity of the cementitious system.*..... 11

Figure 5: *Online Benchmark Data management and SWITCH drive. The benchmark workflows, documentation, models, input files, output files and results are accessible in well structured folder system.*..... 12

Figure 6: *The workflow from system definition to creation of training files is illustrated.*..... 12

Figure 7: Verification figures for the minimal cementitious system. The three geochemical codes produce resutls in excellent agreement with each other. 13

Figure 8*: Surrogate models created by machine learning are tested against independent samples with the metrics of accuracy as shown in this figure.*..... 14

Figure 9: *Example of a cementitious system with input and output parameters (left). Surrogate model prediction (example of PSI-team using neural networks) for the amount of Ca in the solid phase after equilibration, and visual comparison with the geochemical solver GEMS output is shown; The y-axis is the molar concentration of Ca in the solid phase, while the x-axis signifies the test case under consideration. A total of 500 random 3-dimensional input test is shown. In this graph 500 random samples are compared (right)*. Accuracy and speed up of calculations is maintained at very high levels. 15

Figure 10: Benchmark results for the *simple* cementitious system*. The metric mean absolute percentage error with (MAPE) is plot for all output parameters. For the acronyms of the method see main text. (from Prasianakis et al. 2024)*..... 16

Figure 11: Benchmark results for the simple *cementitious system relevant to the computational efficiency of the methods. A significant increase in computational efficiency is observed when ML models are used (from Prasianakis et al. 2024)*..... 17

1. INTRODUCTION

In addition to the specific work which was conducted by each partner, a specific outcome of DONUT project is the definition of benchmarks that will be use both inside DONUT and outside to foster interactions. While international benchmarks initiative are existing (Bildstein et al., 2021; Birkholzer et al., 2019; Steefel et al., 2015), the goal here is to define benchmarks of methods and tools to quantify efficiency and added-value in terms of :

- increase of knowledge (e.g. better physical representation, integration of couple processes, exchange between viewpoints of different disciplines))
- accuracy, robustness, computational cost,
- robustness of scale-transition approaches
- ability to manage uncertainty and sensitivity analyses

Recently, Bildstein et al. (2021) in a guest editorial to the subsurface environmental simulation benchmarks special issue mentioned emerging benchmarking opportunities. Amongst others, machine learning was identified. Indeed it is considered as a recent disruptive technology in the field of reactive transport and will possibly unlock the next generation of simulation that require high demanding CPU time (Leal et al., 2017). The high computing cost associated with chemical equilibrium calculations is considering as the most demanding one in comparison to fluid flow or heat transfer. To circumvent this issue the use of surrogate model provides promising perspectives (Laloy and Jacques, 2019; Prasianakis et al., 2020). Therefore, having a benchmark that tackle this issue will be very useful. In that context DONUT has defined a benchmark relating to ***machine learning and geochemistry***. This latter aims at providing a point of reference for testing and addressing the challenges relevant to: (i) producing high quality training datasets, which can be used by all available ML techniques, (ii) using Deep neural network learning, Polynomial Chaos Expansion and Gaussian processes to learn from the generated data, (iii) testing the accuracy of predictions for geochemical calculations, reactive transport and uncertainty analysis. The philosophy of the benchmark is described in the paragraph 2. It is worth noticing that two geochemical system will be investigated: one related to cement based material degradation and one related to uranium sorption on clay materials. **It therefore provides a clear link to the ACED and FUTURE WPs.**

2. BENCHMARK RELEVANT TO MACHINE LEARNING AND GEOCHEMISTRY

2.1 Benchmark philosophy

Machine learning (ML) is a subset of artificial intelligence with special focus on learning from experimental/numerical data and subsequently representing the correlations of the data in multidimensional variable spaces. This is achieved by using a variety of mathematical models, which result in methodologies like deep neural network learning, polynomial chaos expansion and Gaussian processes. In the context of radioactive waste management, ML may be used to create surrogate models, which are computationally more efficient than the full physical models. For example, it can be used to accelerate the geochemical calculations used in reactive transport calculation (Laloy and Jacques, 2019; Prasianakis et al., 2020). An illustrative example is shown in figure 1. Typically, in reactive transport simulations the transport solver is much faster to compute compared to the thermodynamic and chemistry solver, with the latter being responsible for 90-99% of the overall computational time. The reason for that is that the thermodynamic solver typically involves the iterative solution of several equations until convergence for a single computational grid point, while transport equations are less demanding for the same grid point. With chemistry being the bottleneck, effort has to be spent to accelerate that part of the code. Using machine learning a surrogate model for the chemistry may be created. Once trained and coupled with the transport solver, significant speed-ups are obtained. These models require a training dataset, which is always produced by the full physical-chemical numerical code. The number of training points depends on the number of variables, which defines the input multidimensional space. Once the training dataset is available, the training phase takes place where the ML algorithm learns from the data and is able to represent complex data correlations. After the training is finished, the ML algorithm is ready to be used for predictions for a combination of parameters, which does not belong to the training dataset, which however lies within the range set from the minimum and maximum values of the input parameters existing in the training dataset. The accuracy of the predictions highly depends a) on the size and quality of the training dataset (typically the larger the better), b) on the ML algorithm that was used, and c) on the tuning of the hyper-parameters of the each ML algorithm. Hyper parameter is a parameter whose value is used to control the learning process.

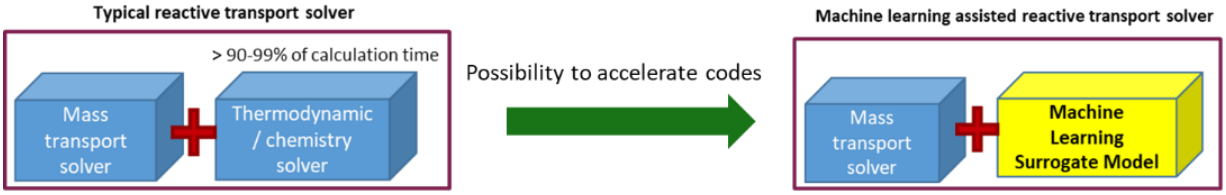

Figure 1*: In typical reactive transport simulations, thermodynamics and chemistry consume most of the computational time. Coupling of carefully trained surrogate models provides an overall acceleration of the simulation between one to four orders of magnitude* (Laloy and Jacques, 2019; Prasianakis et al., 2020)

This benchmark aims in providing a point of reference for testing and addressing the challenges relevant to:

- Streamline the production of high-quality consistent training datasets, using the major geochemical solvers. Setting the specifications such that the datasets may directly be used by all available and future ML techniques.
- Using Deep neural network learning, Polynomial Chaos Expansion and Gaussian processes and other techniques to learn from the generated data.
- Testing the accuracy of predictions for geochemical calculations in a few systems of interest.

2.2 Benchmark Roadmap and main milestones

This benchmark exercise is completed and a manuscript is available, currently under finalization. The history and timeline of the benchmark is described below and illustrated in figure 2:

- **June to December 2021**: System of interest specification and definition of the different benchmark levels of complexity. This action is done.
- **January 2022 to December 2022**: Production of the training sets. This action is done
- **May 2022 to June 2023**: Resolution of the benchmarks by the different teams
- **November 2022**: 1st Workshop to discuss geochemical systems and preliminary ML results (took place 29.11.2022)
- **November 2022 to May 2023**: Based on the November workshop some ew calculations have been performed to enriche the training set. Accuracy metrics were defined to compare the results.
- **April 2023**: 2nd Workshop to exchange information and compare ML results for different levels of complexity. Following the workshop the teams started working in improving the accuracy of their methodologies
- **8-10 November 2023**: ML session during the ACED/DONUT Workshop and parallel discussions of the participating teams
- **March 2023 to June 2024**: Wrap up of results and preparation of common publication and online model and data repository
- **June 2024:** submission of the manuscript for publication

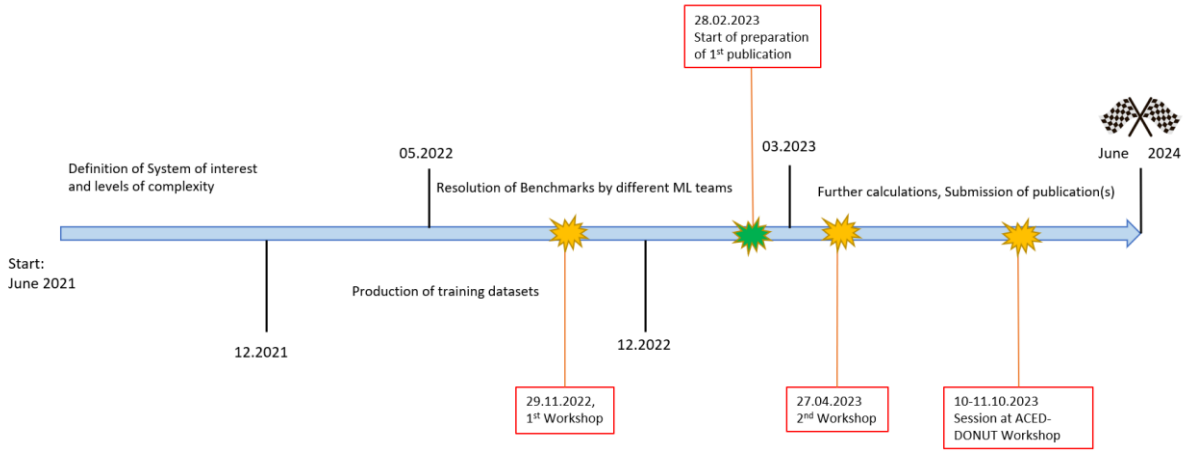

Figure 2*: Benchmark timeline including major milestones and workshops*

2.3 Involved teams

More than 10 research teams across Europe (within and outside EURAD) have joined this benchmark, both at the level of chemical system definition and production of training data, as well as for the part relevant to machine learning techniques. The main participating teams and contact persons are listed below in FIGURE 3:

Geochemical Systems / training datasets: **SCK-CEN**	D. Jacques	(*PHREEQC*)
Geochemical Systems/ training datasets: **NRG**	H. Meeussen	(*ORCHESTRA*)
Geochemical Systems/ training datasets: **PSI**	D. Kulik	(*GEMS*)
ML: **SCK-CEN** (Neural Networks)	E. Laloy / D. Jacques	
ML: **Amphos21** (Neural Networks)	A. Idiart et al.	
ML: **PSI** (Neural Networks)	N. Prasianakis	
ML: **ANDRA/ Uni Paris Saclay** (Polynomial Chaos PCE)	M. Savino, M. Leconte, B. Cochepin	
ML: **UDC/ Env. Data Tech Inc.** (Gaussian Processes and Random forests)	J. Samper, J. Samper (Jr.), C. Yang	
ML: **GFZ-Potsdam** (Geometric Method and Trees)	M. De Lucia	
ML: **BRGM/Uni Orleans/CEA**	F. Claret et al	

Figure 3: *Research teams and contact persons involved in the geochemistry and machine learning benchmark*

2.4 High quality training datasets and problem set

Two systems of interest have been identified. The first system is relevant to cement dissolution/precipitation. The chemical system includes Ca-Si and simple C-S-H models and the CEMDATA-18 thermodynamic database is used (Lothenbach et al., 2019). The used approach is based on C-S-H solid solution thermodynamic models developed by Kulik (2011). . With increasing complexity, a more complete system is addressed, including Al-Mg-S-C-(Na-K) using a structurally-consistent CASH+ sublattice solid solution model for fully hydrated C-S-H phases (Kulik et al., 2022) and its extension for the uptake of alkali metals and alkaline earth metals in C-S-H (Miron et al., 2022).

The second system is relevant to the sorption of U on claystone formation (e.g. Callovo-Oxfordian), Opalinus or BOOM clay). Some preliminary work on the used of surrogate model to decipher Cs sorption uncertainty on Callovo-Oxfordian formation has been made and has been used as starting point to define the benchmark storyboard. Uranium is a more complex system because of its complex speciation and required to build a benchmark with an increasing complexity. The model used for U(VI) sorption on montmorillonite in the absence and presence of carbonate is described in Marques et al. (2012) For the two systems of interest the major geochemical solvers PHREEQC, ORCHESTRA, GEMS (Figure 4) have been used by experts to produce data relevant to the geochemical systems of interest. The cementitious systems that are considered with increased level fo complexity are shown in Fig. 4. It is stressed that for GEMS and ORCHESTRA the lead developers have accompanied the benchmark since the beginning guaranteeing that all necessary details are adequately considered. The results of the geochemical modelling of all systems have shown that all three geochemical solvers are in very close agreement with each other even if different algorithms and theoretical approaches are used for the calculation of the geochemical equilibrium. The systems of interest, the thermodynamic databases and

the exact models along with the input and output files will be provided as supplementary material along with the open access publication. .

CEMENTITIOUS SYSTEM: 6 LEVELS OF COMPLEXITY

System	Acronym	Oxides	Hydrates	CaO	SiO_2	CO_2	Al_2O_3	SO_3	K_2O	H_20
Primitive	P	CaO SiO_2	Portlandite AmorfSi C-S-H[1]	0.1-1.8	0.2-0.7					0.05-0.15
Primitive+C	Pc	CaO SiO_2 CO_2	P + calcite	0.1-1.8	0.2-0.7	0.001-0.96				0.05-0.15
Minimal	M	CaO SiO_2 Al_2O_3	P + gibsite katoite[2] chabazite straetlingite	0.9-1.4	0.3-0.6		0.03-0.07			0.05-0.15
Minimal+C	Mc	CaO SiO_2 Al_2O_3 CO_2	M + calcite hemicarboaluminate monocarboaluminate	0.9-1.4	0.3-0.6	0.001-0.8	0.03-0.07			0.05-0.15
Simple	S	CaO SiO_2 Al_2O_3 SO_3 K_2O	M + monosulfoaluminate[3] ettringite gypsum	0.9-1.4	0.3-0.6		0.03-0.07	0.02-0.05	0.006-0.012	0.05-0.15
Simple+C	Sc	CaO SiO_2 Al_2O_3 SO_3 K_2O	Mc + monosulfoaluminate[3] gypsum thaumasite SS-AFt[4]	0.9-1.4	0.3-0.6	0.001-0.8	0.03-0.07	0.02-0.05	0.006-0.012	0.05-0.15

Figure 4: *(top)The major geochemical solvers are operated by experts in the field, benchmarked and used to produce consistent training datasets for the machine learning techniques. (bottom) the 6 levels of complexity of the cementitious system.*

2.5 Data Management and open research

Throughout the benchmark, all relevant input files, models and results are stored in an online secure cloud service (SWITCH), offered by PSI and Swiss Universities. The status and progress may be followed online at any time at the following a dedicated shared link(the drive is password protected; credentials have been provided to the interested parties and EURAD participants). It is intended that after the end of the benchmark all files will be accessible online as well as supplemental material to the journal publications. Snapshot of the folder structure may be seen in Figure 5

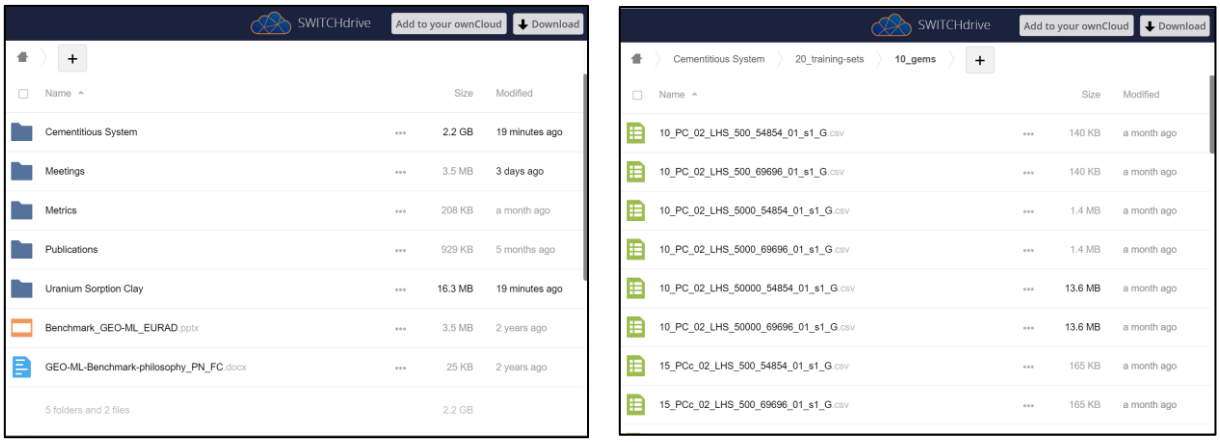

Figure 5: *Online Benchmark Data management and SWITCH drive. The benchmark workflows, documentation, models, input files, output files and results are accessible in well structured folder system.*

The specific benchmark has been data intensive and more than 2'500 files have been produced occupying more than 12GB of disk space. To date the Zenondo online open platform is selected for uploading the produced research data and models after the end of the benchmark. Zenodo is a CERN supported research data repository which provides along with the necessary space a digital object identifier (DOI) to every upload

The benchmark team has defined a specific workflow in order to standardize the process from system definition to machine learning output. In figure 6, the workflow from system definition to creation of training files is illustrated.

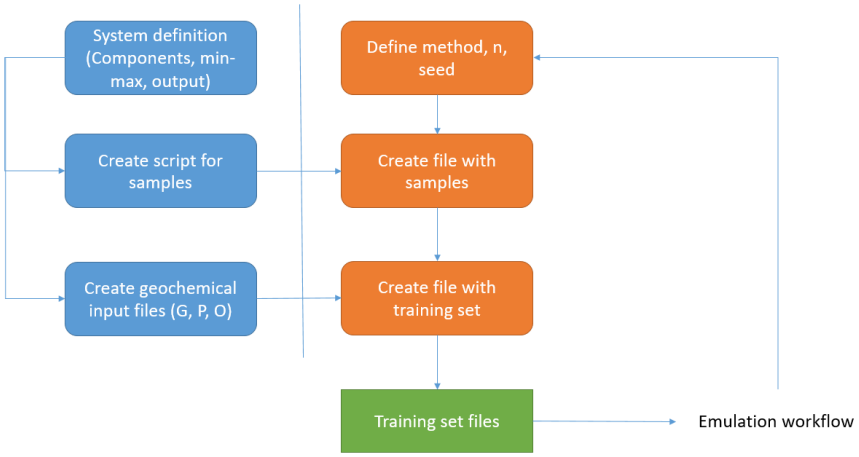

Figure 6: *The workflow from system definition to creation of training files is illustrated.*

2.6 Benchmarking of the geochemical solvers

The results of the three geochemical solvers have been compared for all systems considered. The demonstrated very close agreement can be visualized for the minimal cementitious system in Figure 7. The plot depicts the results of the geochemical codes for the same input conditions at the plane of Gel water Vs Bulk Al_2O_3. Boxes are for GEMS, circles for ORCHESTRA and corsses for PHREEQC. The results of the geochemical codes practically overlap.

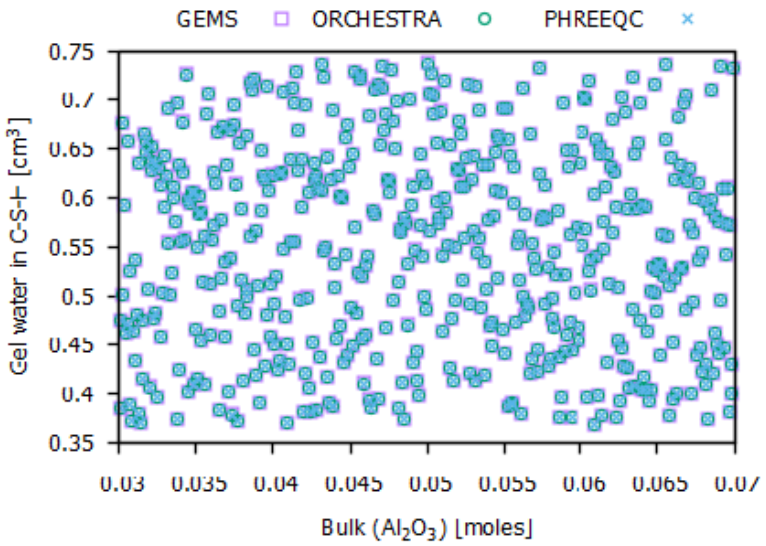

Figure 7: Verification figures for the minimal cementitious system. The three geochemical codes produce resutls in excellent agreement with each other.

2.7 Application and accuracy of predictions of ML algorithms

Several ML methods (e.g. deep neural networks technique, polynomial chaos expansion, Gaussian processes) are used to create the surrogate models. The trained surrogate models are then tested in an independent set of samples to evaluate their accuracy by using a fifteen different metrics. The teams , as shown in figure 3, have already trained their algorithms and produced the final results. For the low complexity cementitious system, very good agreement in terms of accuracy can already be demonstrated. The accuracy of the surrogate models can be measured by using an independent set of input samples and testing against the ground truth (result of geochemical solver). During the course of the benchmarking exercise it has been observed that a single metric of accuracy is not adequate to describe the accuracy of the produced models for parameters (e.g. concentrations) which range across a few orders of magnitude. The most important metrics of accuracy which are used at the moment are shown in figure 8. Several measures of accuracy are used to scrutinize the efficiency of the produced surrogate models. When a model passes all criteria e.g. mean square error (MSE), mean average error (MAE), root mean square error (RMSE) being lower than specific desired accuracy values, then it can be considered suitable for use in reactive transport simulations in the range of input data and the according learning set. The level of accuracy can be set depending on the scope of the respective application.

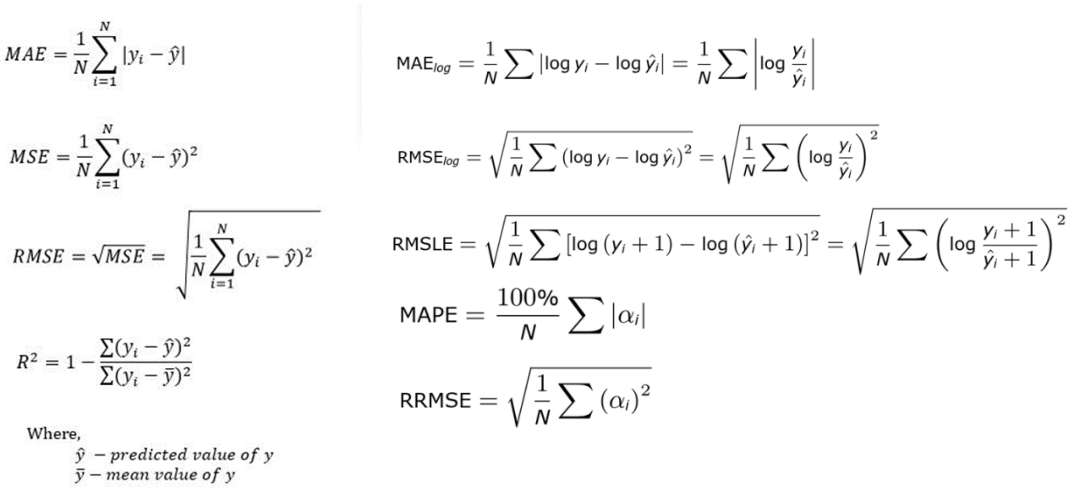

Figure 8*: Surrogate models created by machine learning are tested against independent samples with the metrics of accuracy as shown in this figure.*

The produced results on the cases tested indicate that all teams can achieve very high accuracy with their modelling techniques.. Moreover, a significant speedup of the order of two to four orders of magnitude is already demonstrated depending on the complexity of the system, when compared to the efficiency of the geochemical solvers at “cold start” conditions. “Cold start” conditions are considered when the successive geochemical calculations are not related with each other. For successive calculations which are related to each other, as for example in a reactive transport setup in two successive time steps at the same grid point, some of the geochemical solvers have shown an increased efficiency. This is relevant to the initial conditions of the iterative solvers which are very close to equilibrium thus requiring a significant smaller number of iterations to convergence. Some indicative results are presented below in figure 9. These are results from the PSI team using neural networks for creating the surrogate model. Similar results from all teams are deposited online in the Benchmark cloud directory and will be the central part of the forthcoming publication. In figure 9, (left) the primitive cementitious system with its input parameters and operational range is illustrated in detail. The machine learning model, in this case neural networks, accepts as input three variables which define the composition of the system. The output are the 17 variables as mentioned on the table. For a specific output dimension, namely the amount of Ca in the solid phase after equilibration, the model predictions (red boxes) is tested against the ground truth, which in this case is the geochemical solver GEMS (blue crosses). Both the visual interpretation and the aforementioned metrics (for all output parameters) indicate close agreement between the ML-model and the geochemical solver.

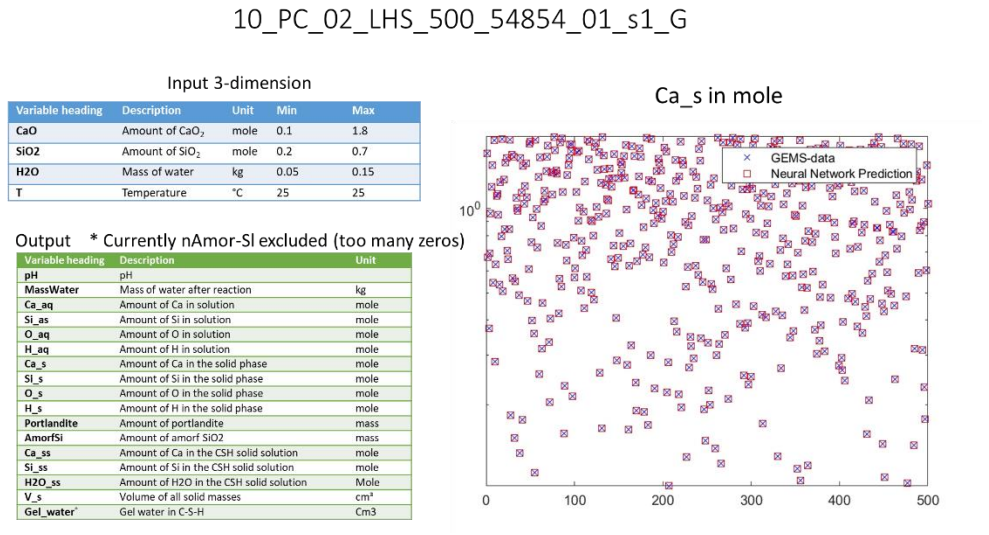

Figure 9: *Example of a cementitious system with input and output parameters (left). Surrogate model prediction (example of PSI-team using neural networks) for the amount of Ca in the solid phase after equilibration, and visual comparison with the geochemical solver GEMS output is shown; The y-axis is the molar concentration of Ca in the solid phase, while the x-axis signifies the test case under consideration. A total of 500 random 3-dimensional input test is shown. In this graph 500 random samples are compared (right)*. Accuracy and speed up of calculations is maintained at very high levels.

The results of all teams (Prasianakis et al. 2024), which participated at the cement system benchmark are plot for the simple cement system are depicted in Figure 10 for the mean average percentage error (MAPE) metric. The bars with different colors depict the difference between the two geochemical solvers GEMS and ORCHESTRA as well as the accuracy of each ML model.

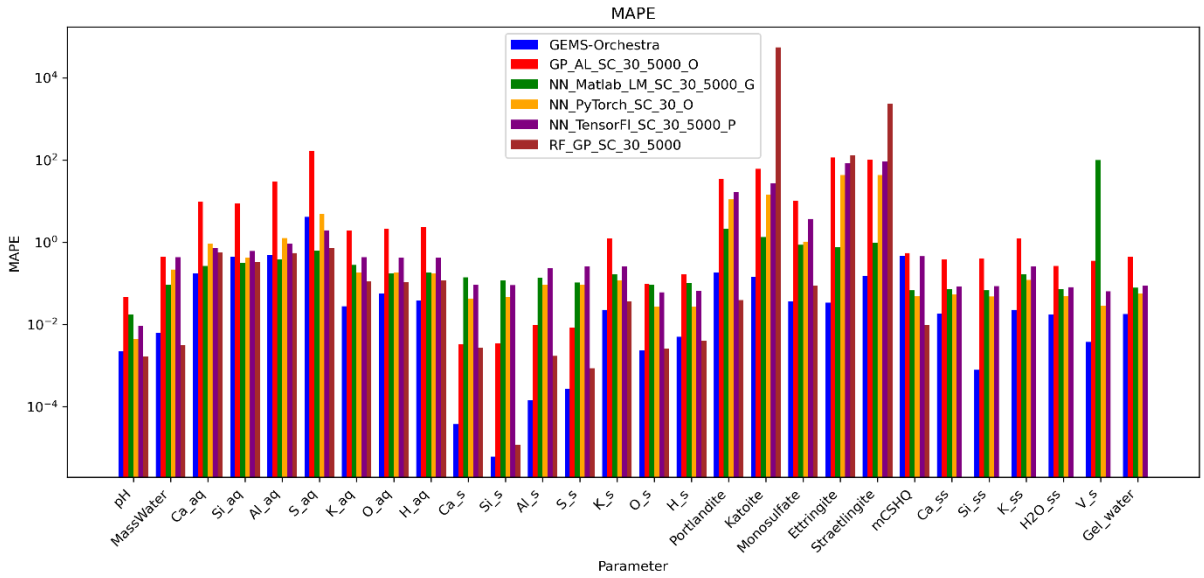

Figure 10: *Benchmark results for the simple cementitious system. The metric mean absolute percentage error with (MAPE) is plot for all output parameters. For the acronyms of the method see main text. (from Prasianakis et al. 2024). blue color: Is the difference between the output of the geochemical solvers GEMS and ORCHESTRA; red color: the accuracy of the Gausian processes active learning (GP-AL) method; green color: the accuracy of the neural network (NN_Matlab) model produced by Matlab; yellow color: the accuracy of the neural network (NN_Pytorch) model produced by PyTorch; purple color: the accuracy of the neural network (NN_Tensorflow) model produced by TensorFlow; brown color: the accuracy of the Random Forest Gausian Processes (RF-GP) model.*

All ML models have shown very good performance by balancing between training effort and accuracy.

The ML models have been also benchmarked in terms of computational efficiency on the same CPU hardware. The results are depicted in Figure 11 (Prasianakis et al. 2024). All ML models provide a significant speed-up in the geochemical calculations compared for example to the PHREEQC geochemical code which represents the typical efficiency of a geochemical solver when random input geochemical calculations are considered. However, it was also observed that during a reactive transport simulation setup the geochemical code ORCHESTRA was able to show a significant speed-up, mainly benefiting from the efficient initialization between successive time steps at the same grid point (“warm start”).

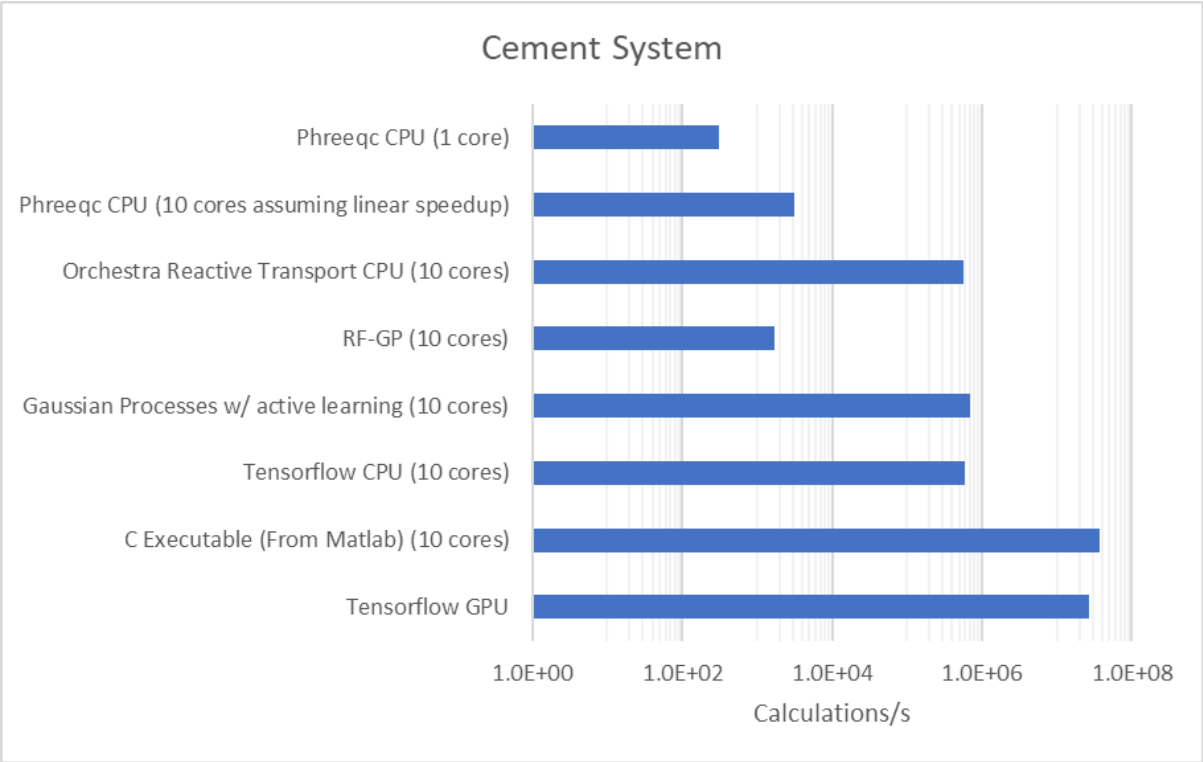

Figure 11: Benchmark results for the simple *cementitious system relevant to the computational efficiency of the methods. A significant increase in computational efficiency is observed when ML models are used (from Prasianakis et al. 2024)*

3. OUTLOOK

A detailed version of the work carried out (Prasianakis et al. 2024) will be soon submitted to a peer review journal. Through the benchmark it has been possible to analyze and address all challenges relevant to training and validation of the ML models. The development of such surrogate models has shown a great potential in accelerating the geochemical calculations and an estimate on training efforts versus expected achieved accuracy can be calculated. This allows already at the stage of designing of the numerical simulations to compute the expected gains in efficiency and to decide whether to proceed with a traditional implementation or with a ML-assisted implementation. The larger the problem and the number of simulations which need to be performed, the more the expected gains. The actual performance of ML-enhanced reactive transport simulations will be a topic to be investigated in our future works.

4. REFERENCES

Bildstein O., Claret F. and Lagneau V. (2021) Guest editorial to the special issue: subsurface environmental simulation benchmarks. *Computational Geosciences* **25**, 1281-1283.

Birkholzer J. T., Tsang C.-F., Bond A. E., Hudson J. A., Jing L. and Stephansson O. (2019) 25 years of DECOVALEX-Scientific advances and lessons learned from an international research collaboration in coupled subsurface processes. *International Journal of Rock Mechanics and Mining Sciences* **122**, 103995.

Kulik, D. A. (2011). Improving the structural consistency of CSH solid solution thermodynamic models. *Cement and Concrete Research*, *41*(5), 477-495.

Kulik D. A., Miron G. D. and Lothenbach B. (2022) A structurally-consistent CASH+ sublattice solid solution model for fully hydrated C-S-H phases: Thermodynamic basis, methods, and Ca-Si-H2O core sub-model. *Cement and Concrete Research* **151**, 106585.

Laloy E. and Jacques D. (2019) Emulation of CPU-demanding reactive transport models: a comparison of Gaussian processes, polynomial chaos expansion, and deep neural networks. *Computational Geosciences* **23**, 1193-1215.

Leal A. M., Kulik D. A. and Saar M. O. (2017) Ultra-fast reactive transport simulations when chemical reactions meet machine learning: chemical equilibrium. *arXiv preprint arXiv:1708.04825*.

Lothenbach B., Kulik D. A., Matschei T., Balonis M., Baquerizo L., Dilnesa B., Miron G. D. and Myers R. J. (2019) Cemdata18: A chemical thermodynamic database for hydrated Portland cements and alkali-activated materials. *Cement and Concrete Research* **115**, 472-506.

Marques Fernandes M., Baeyens B., Dähn R., Scheinost A. C. and Bradbury M. H. (2012) U(VI) sorption on montmorillonite in the absence and presence of carbonate: A macroscopic and microscopic study. *Geochimica et Cosmochimica Acta* **93**, 262-277.

Miron G. D., Kulik D. A., Yan Y., Tits J. and Lothenbach B. (2022) Extensions of CASH+ thermodynamic solid solution model for the uptake of alkali metals and alkaline earth metals in C-S-H. *Cement and Concrete Research* **152**, 106667.

Prasianakis N. I., Haller R., Mahrous M., Poonoosamy J., Pfingsten W. and Churakov S. V. (2020) Neural network based process coupling and parameter upscaling in reactive transport simulations. *Geochimica et Cosmochimica Acta* **291**, 126-143.

Prasianakis N.I., Laloy E., Jacques D., Meeussen J.C.L., Tournassat C., Miron G.D., Kulik D. A., Idiart A., Demirer E., Coene E., Cochepin B., Leconte M., Savino M., Samper II J., De Lucia M., Yang C., Churakov S. V., Samper J., Kolditz O., Claret F., (2024) Geochemistry and Machine Learning benchmark, (manuscript available)

Steefel C. I., Yabusaki S. B. and Mayer K. U. (2015) Reactive transport benchmarks for subsurface environmental simulation. *Computational Geosciences* **19**, 439.