

SYSTEMATIC ANALYSIS OF MULTIVARIATE SCENARIOS USING ADVANCED CLUSTERING METHODS

ERNOULT Marc¹, DAVID Sylvain¹, DOLIGEZ, Xavier¹, LIANG Jiali¹,
THIOLLIERE Nicolas², TILLARD Lea³

¹Institut de physique nucleaire dOrsay, IN2P3/CNRS University Paris-Sud
15 rue Georges CLEMENCEAU 91406 ORSAY, France

²Subatech, IMTA IN2P3/CNRS University of Nantes
Nantes, F-44307, France

³IRSN/PSN-EXP/SNC/LN
92262 Fontenay-aux-Roses, France

ernoult@inpo.in2p3.fr

ABSTRACT

The continuous improvement of fuel cycle simulators in conjunction with the increase of computing capacities have led to a new scale of scenario studies. Taking into consideration multiple variable parameters and observing their effect on multiple evaluation criteria, these scenario studies regroup several thousands of trajectories paving the different possible values for multiple operational parameters. If global methods like sensitivity analysis allow extracting useful information from these groups of trajectories, they only provide average and global values.

In this work we present a new method to analyze these groups of trajectories while keeping some localization in the information. Based on principal component analysis, clustering method have been implemented in order to mathematically extract, from the ensemble of trajectories simulated for a scenario study, subgroups of trajectories that have similar behaviors. Typical trajectories, representative of these subgroups, are then determined. The application of this new method on a sample scenario for two different output, the total amount of transuranic elements within the fuel cycle and the number of time the MOX fuel could not be built during the simulated time, is presented. The comparison of the results between the two analyses shows that the method allows good clustering for continuous and regular outputs but struggle with discrete highly non-linear ones.

KEYWORDS: fuel cycle, MOX, scenario, clustering

1. INTRODUCTION

If fuel cycle simulators have been used for several decades, the last years have seen a renewed interest and many new developments, new codes as well as major updates for older codes. In this context, fuel cycle simulators capabilities have improved and several simulators adapted themselves to parallels calculations [1–3]. For example, in CLASS [4], the use of parallel calculation

and method of simplified modeling as macro reactor [5], up to 800 trajectories can be simulated in less than one hour depending on the complexity of the simulations. The FIT Benchmark [6] were 6 different fuel cycle simulators simulate 2000 trajectories of a simple scenario specifically illustrates that such capacity to run large number of simulation has spread. Analysis in scenario studies will thus involve several hundreds or thousands of trajectories.

To understand variables of impact on one observable and analyze sensitivities, a method taking advantage of this newly possible big number of calculation have been developed within the CLASS team : the Wide Parametric Sweeping method [6]. This method is a direct extension of the Global sensitivity analysis for fuel cycle studies [3]. In this method, a design of experiment is built setting out a broad scenario and a list of inputs sampled within a predefined range of variation. After simulation, because numerous trajectories are simulated, the analyst can easily apply a filter on the inputs and see the impact on the solution space without the need of any new calculations from the fuel cycle simulator.

In the past, the result trajectories have been analyzed through global sensitivity indices like Sobol indices [3,7]. However, for more in depth analysis, those global indices are not enough. Reference trajectories are a thus chosen within the sampling space and analyzed individually [7].

2. SCENARIO USED FOR METHOD TESTING

2.1. Scenario Definition And Sampling Space

To test our new method of clustering of trajectories and selection of representative Reference Trajectories the Wide Parametric Sweeping method [6] has been applied to generate numerous trajectories. This method asks for the design of a scenario in broad terms and the choice of several controlled variables for which constraints are released. Because the objective of this work is more focused on method testing than specific fuel cycle analysis, the scenario has been chosen to only use only reactor for which the physics and the fuel cycle behavior is well understood : PWRs, with UOX and MOX fuels. The MOX one uses plutonium from spent UOX fuel and no multi-recycling of the plutonium is considered.

The scenario is built in 3 phases : an initial stable phase, a transition phase and a final stable phase. The first phase, identical in all the trajectories of the studied scenario, is a simplified representation of the current French fleet : Total thermal power of 188 GW from which 7.6% is produced by MOX fuel [7]. The simulation starts in year 2015. The final phase has its power (P_f) and MOX fraction (Fr_{MOX_f}), as well as the maximum Burn-Up for both fuels (BU_{UOX} and BU_{MOX}) and the minimum cooling before UOX spent fuel reprocessing (TC_{UOX}), as variable parameters used in the Wide Parametric Sweep. This phase is simulated until the year 2180. The starting time (T_{Start}) and ending time (T_{End}) of the transition phase are also variable parameters used in the Wide Parametric Sweep. During this phase MOX fraction and Total power change linearly. Burn-ups and cooling time change abruptly at the beginning of this phase. Selection of UOX spent fuel to be reprocessed first for MOX fuel fabrication is also a variable parameter with two possibles choices : First In / First Out (FiFo) and Last in /First Out (LiFo). It stays the same during all the trajectory. For each of these parameters used in the Wide Parametric Sweep, we set out the wide ranges presented in table 1.

	T_Start	T_End	P_f	FrMOX_f	BU_UOX	BU_MOX	TC_UOX	ManagePu
Min	2025	2035	19	0	30	30	3	1/LiFo
Max	2125	2180	282	0.3	60	60	10	2/FiFo
Units	date	date	GWth	ratio	GWd/t	GWd/t	years	N.A.
Init.	-	-	188	0.076	30	30	5	1/LiFo

Table 1: Input space considered for application for the Wide Parametric Sweeping method

2.2. Generation Of Trajectories Bundle

Once the input space has been set out, to generate the input for the fuel cycle simulator, we used a Latin Hypercube Sampling [8] (LHS), creating a set of input parameters representing 20000 possible trajectories. A cut in the LHS has been undertaken, keeping only points for which $T_Start < T_End - 10$. and reducing the number of trajectories to be simulated to 13098. Each of these 13089 trajectories has been then simulated with CLASS [4]. The total amount of data created is 1.1TB. To facilitate the analysis and the storage of data, the individual output files are thus aggregated into a single file keeping only data of interest reducing the size to 15GB.

2.3. Outputs Considered

To explore the possibilities and benefits of clustering methods in scenario analysis, two outputs have been chosen : **INC_TRU** : The total amount of transuranic (TRU) elements (Np, Pu, Am and Cm) in the fleet (in reactors, storage, cooling and fabrication plants but not in the waste) at the end of the simulation in year 2180; and **ML** : the total number of Missed Loads. Each time plutonium shortage leads to a failure in the MOX fuel fabrication, the reactor in the simulation is not loaded with any kind of fuel : this is what we call a Missed Load. The total fleet power is then reduced by the considered reactor power.

These two outputs are interesting to evaluate a TRU management strategy that may want to reduce **INC_TRU** while keeping **ML** equal to 0. They also are interesting for this work as they are very different by nature : **INC_TRU** is a continuous variable that react smoothly and mostly linearly to most change in strategies while **ML** is a discrete variable sensible to a lot of thresholds effects leading to strong non-linearity.

3. METHOD OF DATA CLUSTERING

3.1. Data Clustering From Principal Component Analysis

Using the Wide Parametric Sweeping method, numerous trajectories, which are representative of all our input space, have been obtained. To get a better grasp of behaviors existing in these trajectories, clustering using the unsupervised model-based clustering developed within the EMCluster

package [9] in R have been used. It allows a projection of the data matrix (inputs and outputs) in the principal component space and then create clusters using the Hartigan-Wong k-means algorithm [10]. The results are then presented in the initial input-output space for a better readability.

3.2. Selection Of Representative Reference Trajectories

Once clustering carried out, groups of trajectories that tile the entirety of our ensemble of trajectories is obtained and each cluster regroups trajectories that behave similarly. However, the number of trajectories is still the same, to allow more focused and in depth analysis extraction of representative trajectories is needed. These trajectories can be re-simulated with fewer approximations for verification that the conclusions of the study are not numerical artifacts creating by simplifications and reflect the physics [7].

For each cluster, a Representative Reference Trajectory is built as the trajectory generated by the inputs that are the barycenter of the trajectories of the cluster. The input parameter i of the Representative Reference Trajectory of the cluster j is the average of all the values of the input parameter i of all trajectories from cluster j . The clustering method used creates compact clusters, so the Representative Reference Trajectory is always part of the cluster it represents.

Calculation cost of simulating a trajectory is low. An additional simulation of the Representative Reference Trajectory can be performed. However, if the initial sampling is dense enough, the closest simulated trajectory can serve a similar purpose.

4. CLUSTERING OF SCENARIO AND ANALYSIS

Because the number of Missed Loads is one of our outputs of interest, trajectories with missed loads are kept in all following analyses.

4.1. Clustering Trajectories considering incycle TRU

4.1.1. Clustering

The first test of our clustering analysis method was performed considering only in-cycle transuranic elements, **INC_TRU**. To be sure that the number of missed loads have no influence on the clustering carried out in this part, the corresponding column have been deleted from the data-set. The clustering has thus been undertaken on a data-set assembling 8 inputs and 1 output variables. The ensemble of trajectories has been distributed between 4 clusters.

The pair plot presented in figure 1 allows a good visualization of which subsection of the input phase is associated with each cluster.

The 4 clusters can be described as follow : **1.** (in red) : Low **P_f**, which lead to low **INC_TRU** **2.** (in black) : intermediate **P_f** and low **FrMOX_f**, which leads to average **INC_TRU**; **3.** (in green) : high **P_f** and low **FrMOX_f**, which leads to very high **INC_TRU**; **4.** (in blue) : high **P_f** and high **FrMOX_f**, which leads to lower **INC_TRU**.

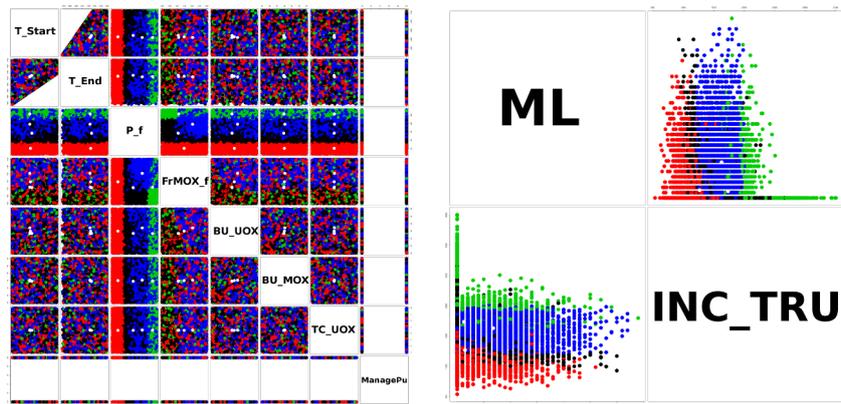


Figure 1: Clusters of trajectories from clustering on INC_TRU

From this clustering we recognize the links between power, MOX fraction and quantity of in-cycle transuranics shown in several previous works [3,7] : higher final power leads to higher **INC_TRU** and higher MOX fraction decreases **INC_TRU**. The decrease of **INC_TRU** in the blue cluster is due to effective power during simulation being lower in this cluster than what the **P_f** value indicates due to missed loads. The uniform distribution of the cluster in other input variables is a sign of a low effect of these variables on the considered output.

4.1.2. Representative Reference Trajectories

Using these clusters and the method presented in 3.2, 4 Representative Reference Trajectories have been generated. The input values of these trajectories are given in table 2. The representative trajectories (represented as white dots in figure 1) are well in the center of the considered cluster and carry manifestly the main characteristics of the cluster (from low to high **P_f** and **FrMOX_f**).

	T_Start	T_End	P_f	FrMOX_f	BU_UOX	BU_MOX	TC_UOX	ManagePu
red	2065	2127	52	0.15	45	45	6	FiFo
black	2069	2130	143	0.11	43	45	6	FiFo
green	2064	2126	257	0.11	43	45	7	FiFo
blue	2064	2125	196	0.20	47	45	7	LiFo

Table 2: Representative Reference Trajectories from clustering on INC_TRU

For inputs of low impact to the considered output, the Representative Reference Trajectories have a value close to the middle of the sampling range.

The **ManagePu** presents a specific behavior. Because it's a discrete variable each Representative

Reference Trajectory is either FiFo or LiFo. However, because it's a variable of low impact on **INC_TRU**, each cluster has almost the same number of point using FiFo or LiFo plutonium management. The fact that each cluster has a higher amount of FiFo or LiFo trajectory is random and due to the LHS sampling, this random unbalance set the **ManagePu** of the Representative Reference Trajectory.

4.2. Clustering Trajectories considering number of missed loads

4.2.1. Clustering

The second application of the method considers only the missed loads, **ML** as output of interest. Like previously, the other output, here **INC_TRU**, has been completely removed from the data-set before clustering. The clustering has thus been performed on a data-set assembling 8 inputs and 1 output variables. The ensemble of trajectories has been distributed between 4 clusters.

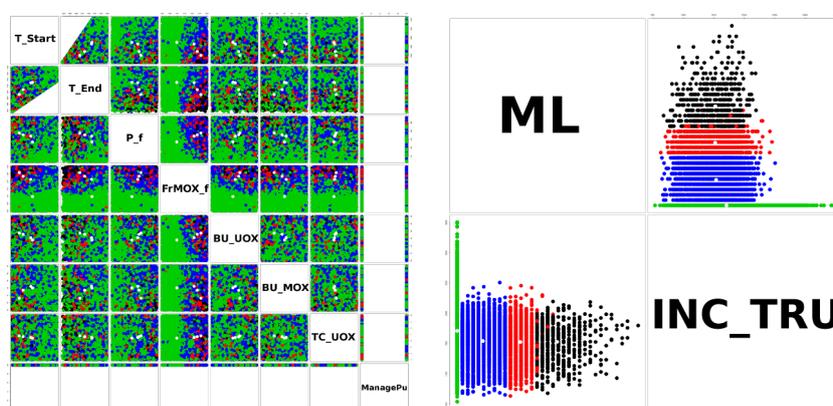


Figure 2: Clusters of trajectories from clustering on ML

The pair plot presented in figure 2 allows a good visualization of which subsection of the input phase is associated with each cluster.

The 4 clusters can be described as follows : **1.** (in green) : Trajectories without missed loads, with **P_f** not too high and **FrMOX_f** not too high; **2.** The 3 other clusters regroup trajectories with missed loads. They cut the ensemble of trajectories in slices along a almost diagonal axis relative to **P_f** and **FrMOX_f**; **a** (in blue) : low number of **ML** with not so high **P_f** and not so high **FrMOX_f**; **b** (in red) : higher number of **ML** with high **P_f** and high **FrMOX_f**; **c** (in black) : a very big number of **ML** with very high **P_f** and very high **FrMOX_f**.

The three clusters with non-zero **ML** are completely interlocked with each other at a point were they are almost indistinguishable. The cut between clusters are clear in the output space but it does not translate well in an input space tiling. This is most probably because principal component analysis that plays a key intermediate role in our clustering process is not well suited to deal neither with discrete values neither with highly non-linear variables and **ML** is in both of these unwanted categories.

From the first cluster grouping trajectories with zero **ML**, the links between power, MOX fraction and shortage of plutonium for MOX fuels shown in several previous works [3,7] is seen : if **P_f** is high the reserves of plutonium built up during the initial phase are not enough to supply the increase of MOX fuels proportional to the power during the transition. Furthermore, the more MOX fuel you need to fabricate the higher the risk of plutonium shortage is. This creates a cluster with all trajectories having **FrMOX_f** < 0.15 as well as some higher MOX fraction when **P_f** is lower than 100 GWth.

Like with the previous clustering, we see a uniform distribution of clusters in the input with low effect on the output.

Furthermore, because the value of the output is of primary importance in the clustering process, the method is also unable to create multiple clusters within the zone without missed loads.

4.2.2. Representative Reference Trajectories

The 4 Representative Reference Trajectories generated from the clusters have the input values presented in table 3. The representative trajectories (represented as white dots in figure 2) are by design in the center of the considered cluster. In these cases of almost superposed clusters, the positions of the centers make the difference between them more readable. These centers are placed along a diagonal line in the box of the pairs representing the link between **P_f** and **FrMOX_f** and helped the characterization of the cluster presented in last section.

The main difference between the Representative Reference Trajectories are in the **P_f** and **FrMOX_f**. The **ManagePu** present a similar random behavior as explained in the previous clustering.

	T_Start	T_End	P_f	FrMOX_f	BU_UOX	BU_MOX	TC_UOX	ManagePu
green	2070	2131	135	0.10	44	45	6	FiFo
blue	2066	2130	170	0.21	46	46	7	LiFo
red	2049	2108	177	0.24	48	44	7	LiFo
black	2041	2092	189	0.26	49	38	7	LiFo

Table 3: Representative Reference Trajectories from clustering on ML

5. CONCLUSIONS

A first application of a new method to analyze large number of trajectories has been presented. This method gives, with an additional calculation cost of only several seconds, an automatic and deterministic way of grouping trajectories that help to grasp the variability of analyzed output. It also provides representative reference trajectories for each of the cluster that can serve as starting point for more focused studies while assuring a good representation of the phase space.

An ensemble of trajectories exploring possibilities within a fleet of PWR using UOX and MOX fuels has been analyzed with this new method considering two very different outputs of interest : the total amount of transuranics elements within the fuel cycle (**INC_TRU**) and the number of times the MOX fuel could not be built during the simulated time (**ML**). For **INC_TRU** which behaves linearly, the method creates 4 meaningful clusters that represent 4 different behaviors of the fuel cycle and the center are good representatives of these clusters. For **ML** a discrete and highly non-linear output, the method identifies easily one meaningful cluster, the points without missed loads, but struggle for the rest, creating clusters difficult to discriminate. Because the value of the output is of primary importance in the clustering process, the method is also unable to create multiple clusters within the zone without missed loading decreasing its interest for this kind of output.

There is in theory no obstacle to apply this method considering multiple outputs at the same time. However early attempts have not been fruitful due to scale problems between outputs leading to clustering practically ignoring some considered outputs. More work can thus be carried out in this direction using scaling methods to overcome this limitation.

REFERENCES

- [1] G. Krivtchik et al. “Scenarios with COSI6: Optimization, Uncertainty and Beyond.” In *Technical Workshop On Fuel Cycle Simulation*. Columbia, SC, USA (2017).
- [2] R. Flanagan et al. “Rickshaw : a new stochastic driver for Cyclus.” In *Technical Workshop On Fuel Cycle Simulation*. Columbia, SC, USA (2017).
- [3] N. Thiolliere et al. “A methodology for performing sensitivity analysis in dynamic fuel cycle simulation studies applied to a PWR fleet simulated with the CLASS tool.” *EPJ Nuclear Sci Technol*, **volume 4** (2018).
- [4] B. Mouginot et al. “Core Library for Advanced Scenario Simulation, C.L.A.S.S. : Principle and Application.” In *PHYSOR*. Kyoto, Japan (2014).
- [5] A.-A. Zakari-Issoufou et al. “Impact of Macro Reactor Approximation on Scenario : Modeling in C.L.A.S.S.” In *Technical Workshop On Fuel Cycle Simulation*. Paris, FRANCE (2018).
- [6] N. Thiolliere et al. “The FIT Project: Improving confidence in fuel cycle model.” In *Technical Workshop On Fuel Cycle Simulation*. Urbana-Champaign, ID, USA (2019).
- [7] F. Courtin. *Etude de l’incinération du plutonium en REP MOX sur support duranium enrichi avec le code de simulation dynamique du cycle CLASS*. PhD in French, Nantes, FRANCE (2017).
- [8] K. Ye. “Orthogonal column Latin hypercubes and their application in computer experiments.” *Journal of the American Statistical Association*, **volume 93**, pp. 1430–1439 (1998).
- [9] W.-C. Chen and R. Maitra. “EMCluster: EM Algorithm for Model-Based Clustering of Finite Mixture Gaussian Distribution.” (2015). R Package, URL <http://cran.r-project.org/package=EMCluster>.
- [10] J. A. Hartigan and M. A. Wong. “A K-Means Clustering Algorithm.” *Journal of the Royal Statistical Society Series C (Applied Statistics)*, **volume 28** (1979).