Unsupervised learning for nuclear fuel cycle applications

A.V. Skarbeli, F. Álvarez-Velarde

CIEMAT

3rd Workshop of Spanish Users on Nuclear Data

Machine Learning in Nuclear Science and Technology Applications

May 27, 2021

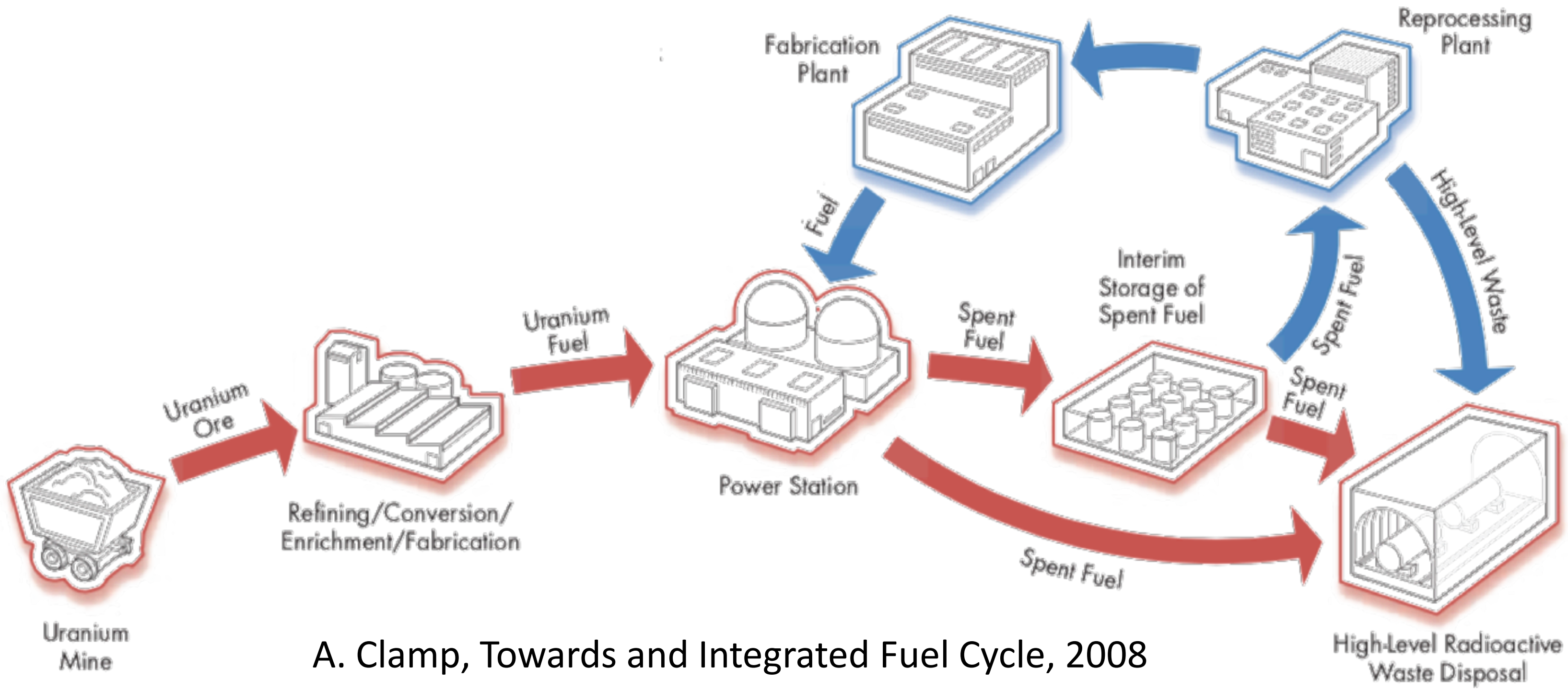
GOBIERNO DE ESPAÑA

MINISTERIO DE CIENCIA E INNOVACIÓN

Ciemat Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas

The nuclear fuel cycle

- Nuclear fuel cycle studies cover all the processes involved in the production of energy from nuclear materials: from mining to final disposal
- They can be broadly grouped in two categories:
  - Open fuel cycles, with no recycling of the materials (*e.g.* Spain)
  - Closed fuel cycles, where the materials are recycled/multirecycled in the reactors before their final disposal



A. Clamp, Towards and Integrated Fuel Cycle, 2008

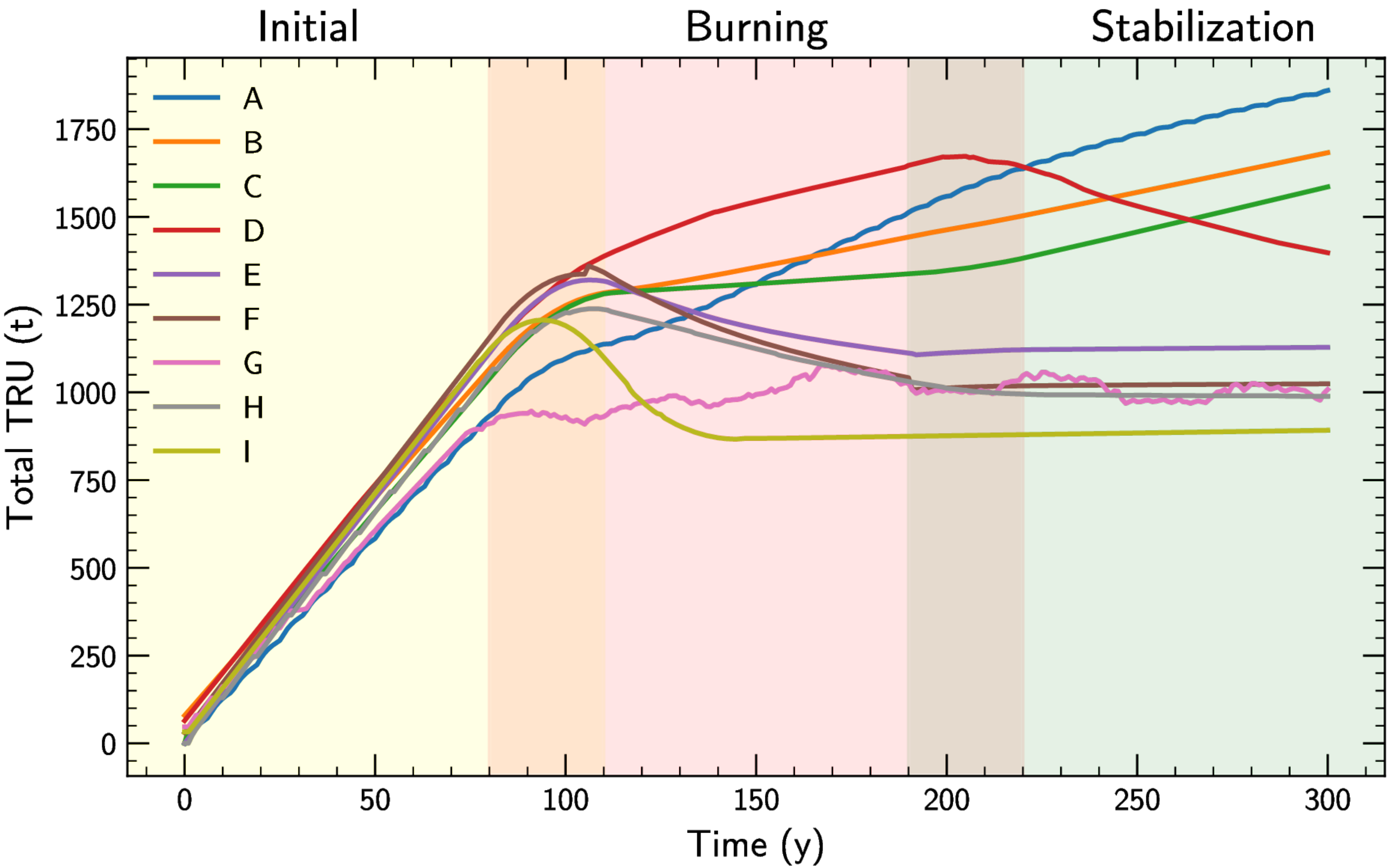
The nuclear fuel cycle

- There are multiple approaches for closing the nuclear fuel cycle
- Should we use
  - LWR MOX?
  - PWR MOX UE IMR?
  - FR?
    - GFR?
    - SFR?
      - Burn MA?
      - Breeding ratio?
  - ADS?
  - A combination?
  - …
- Can machine learning techniques help decision makers and experts in the selection of a particular technology?

Dataset description

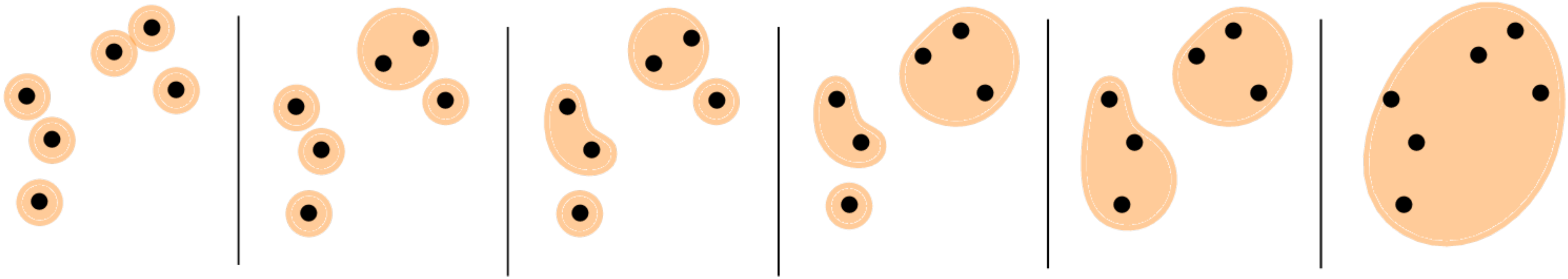
- Scenario objectives: starting with a PWR(UOX) fleet, we want to minimize and stabilize the TRU (Pu + MA) inventories
  - Three stages:
    - Initial phase with only PWR(UOX) fleet
    - Burning phase where the inventories are reduced as much as possible
    - Stabilization phase where the inventories are stabilized and thus the scenario can be perpetuated indefinitely
- Simulate the scenario with different configuration of technologies
  - Anonymized dataset from NEA/EGAFCS (Expert Group on Advanced Fuel Cycle Scenarios)
    - Different technologies: ADS, GFR, breeder SFR, breakeven SFR, MOX(RMA),...

Dataset description



Hierarchical clustering

- For a given objective, we would like to find which selection of technologies provide similar results
- We have multiple simulations
- And we want to group them based on their similarity, *i.e.* how close they are
- This defines an unsupervised learning (clustering) problem: *find the objects that are in the same group*
  - We will use hierarchical clustering given that the number of clusters is unknown
  - 2D-space example:



Hierarchical clustering

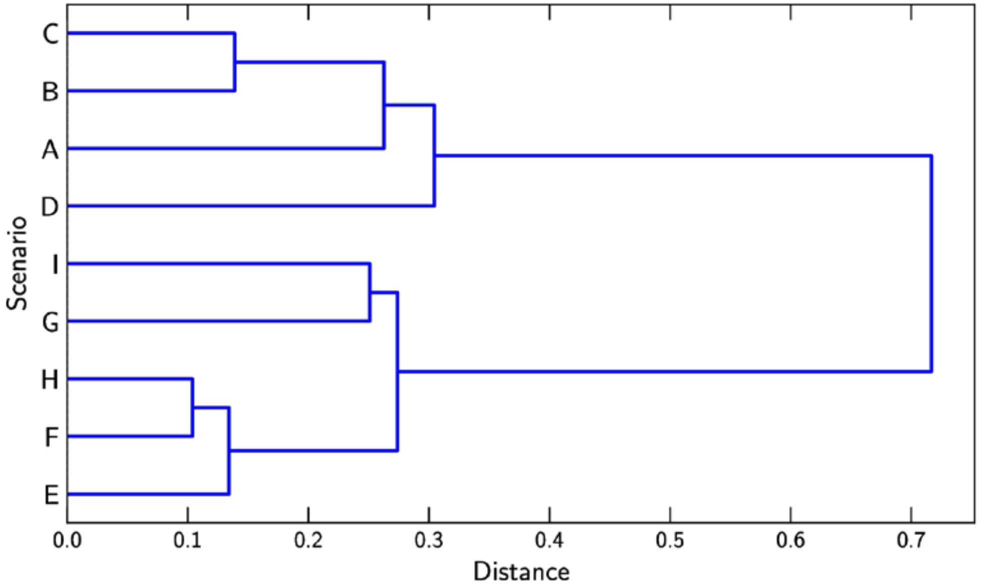
- For a given output, the scenario results is a *n*-dimensional vector where each component corresponds to the value of the observable at year *i*

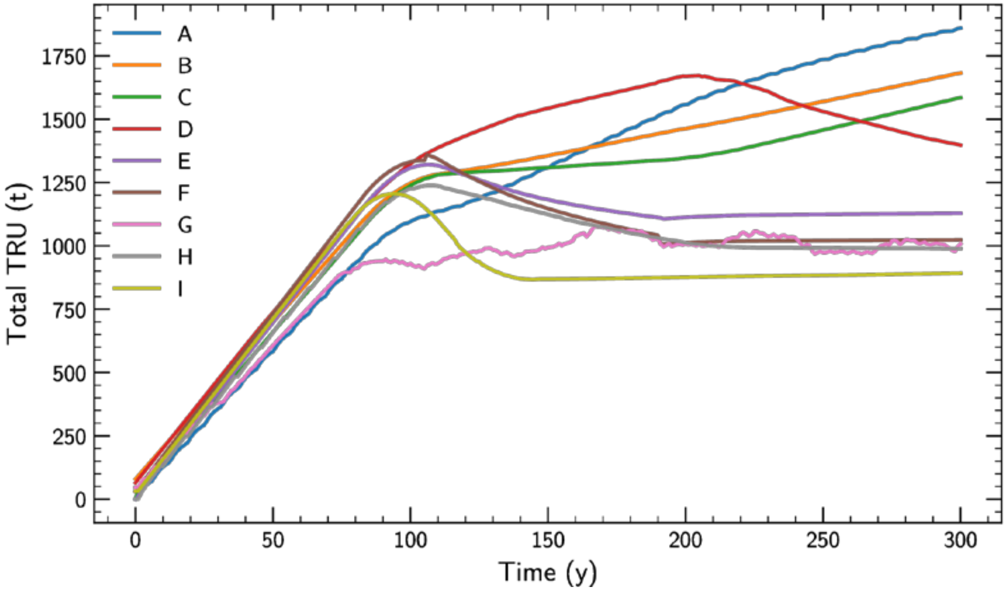
$$S_i = \{x_1, x_2, \ldots, x_n\}$$

- Two assumptions required:
  - Metric for measuring the distance between two scenarios S_1 and S_2: Euclidean distance
  - Metric for measuring the distance between two clusters (linkage): WPGMC algorithm
    - When two clusters are combined, the average of the centroids of the clusters are combined in the new centroid
- We will use SciPy Python library

(There are multiple alternatives but this presentation will only focus on the underlying ideas!)

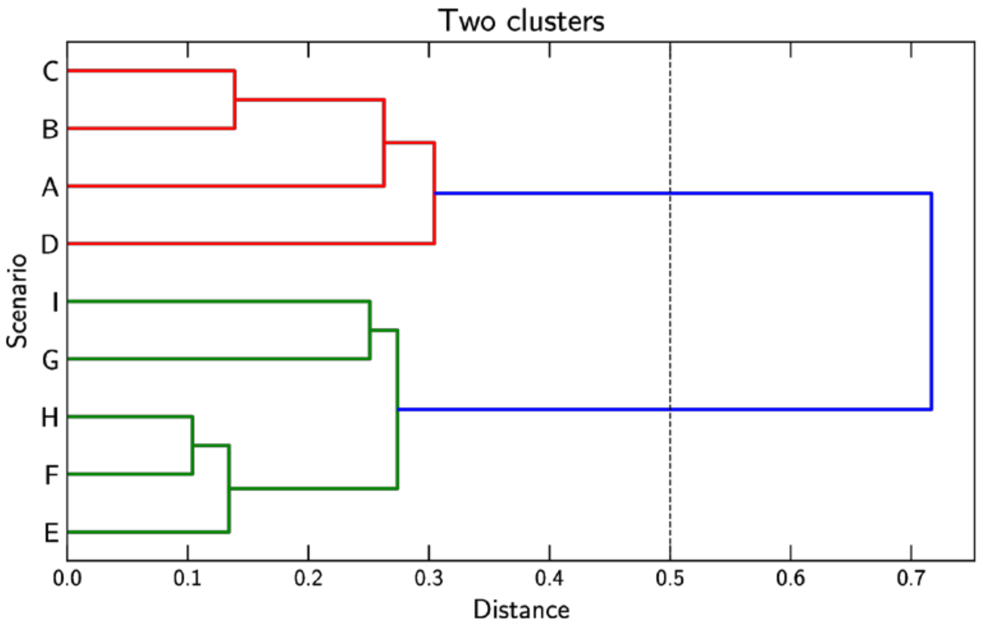
Clustering results

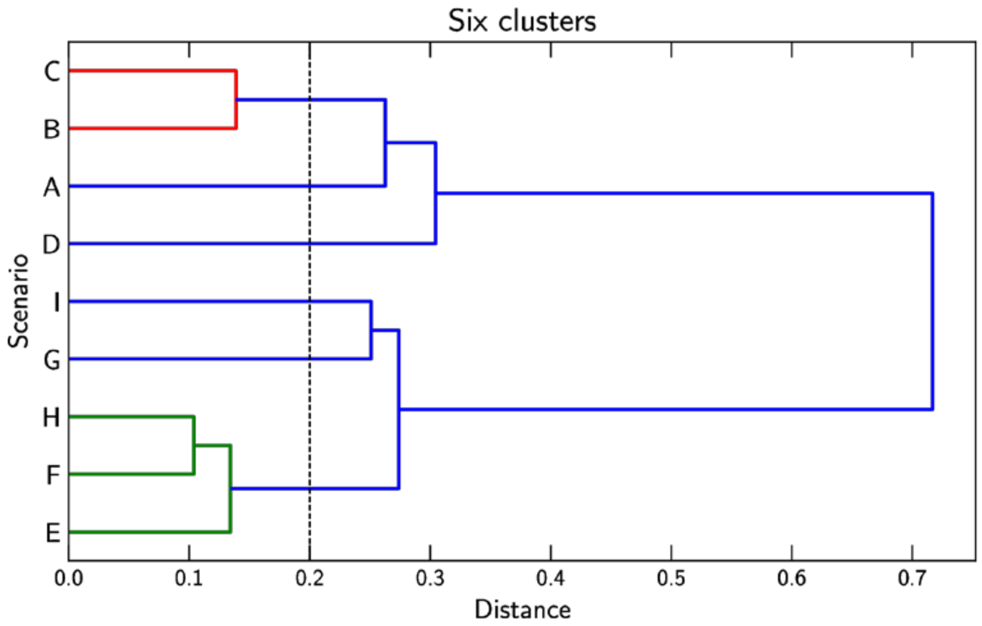


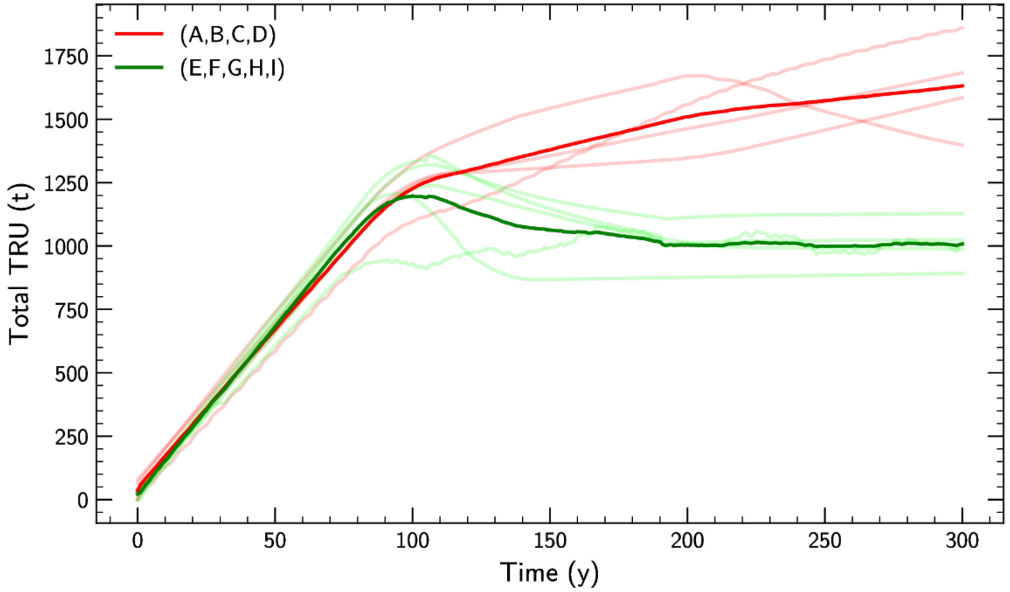


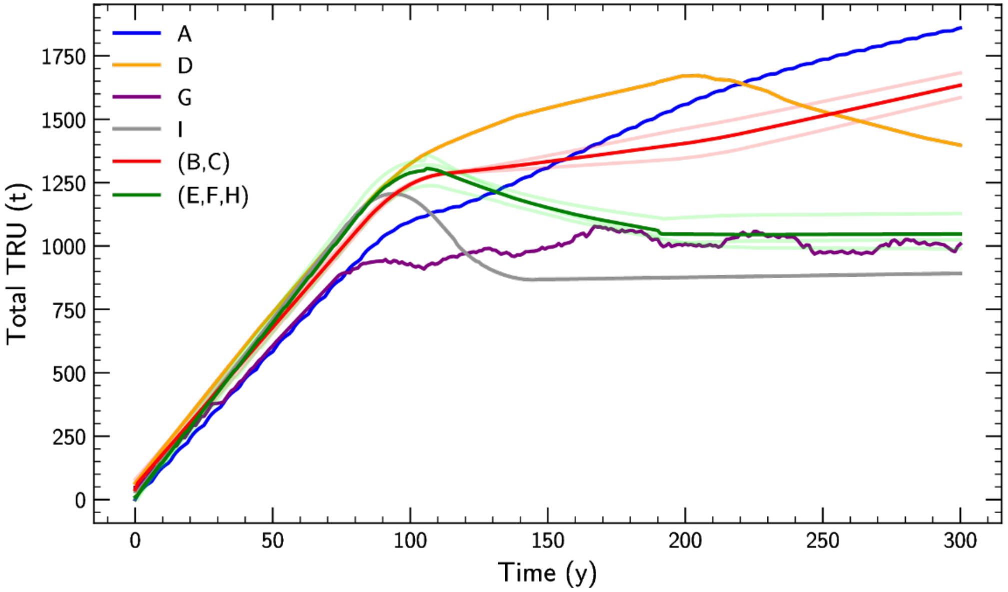
- The dendogram represents the distance (they have been normalized) at which the clusters are merged
  - (H,F) are the closest scenarios
  - Then the cluster (H,F) and (E)
  - Then (C,B)
  - ….
- Cutting the dendogram at distance *d* produces *k* clusters

Clustering results

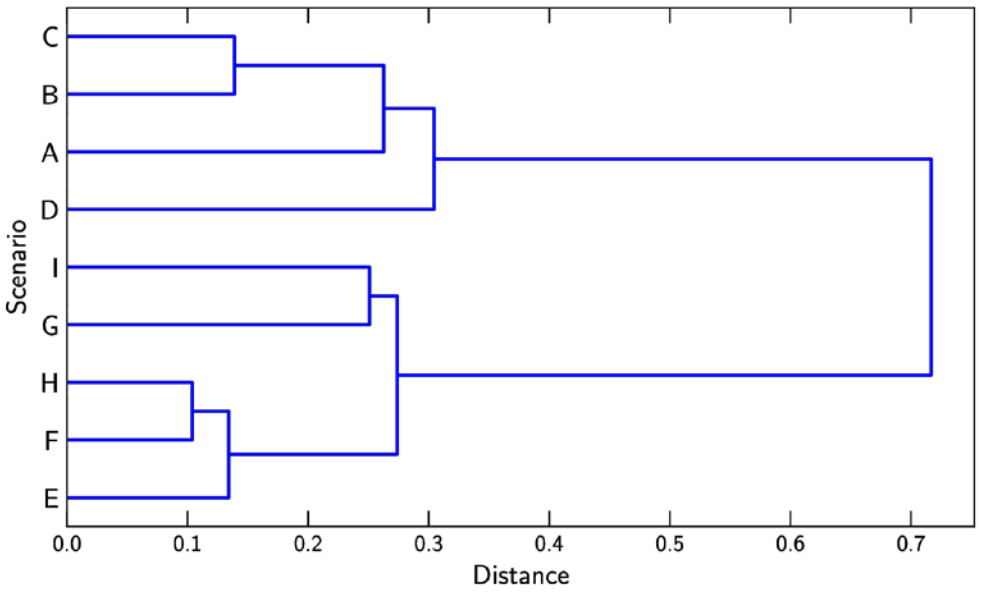


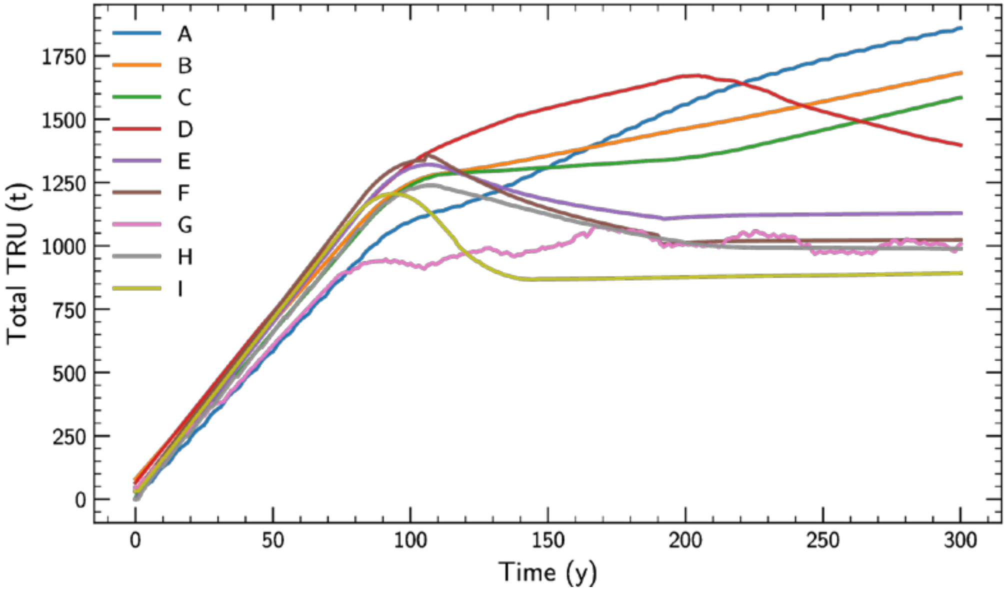






Clustering results





- Scenarios (H) and (F) use ADS, the rest of them (FR)
- It would be reasonable to expect that all FR scenarios will be in the same group
- But the clustering shows that (E,H,F) are the most similar ones despite they use different technologies

Conclusions

- The hierarchical clustering provides a methodology for grouping scenarios based on their similarity
- It can be used for selecting those technologies that are more suitable for given objectives
- It can be easily extended to large datasets
- No assumptions about how the data is generated
- It only required defining the distance metrics
  - They can have a strong dependence in the results
  - But not a problem if it is chosen according to the desired outcomes