UNIVERSIDAD PONTIFICIA ICAI ICADE COMILLAS MADRID

Data Analytics for Transmission Expansion Planning
Estadística II

Andrés Ramos

January 2018

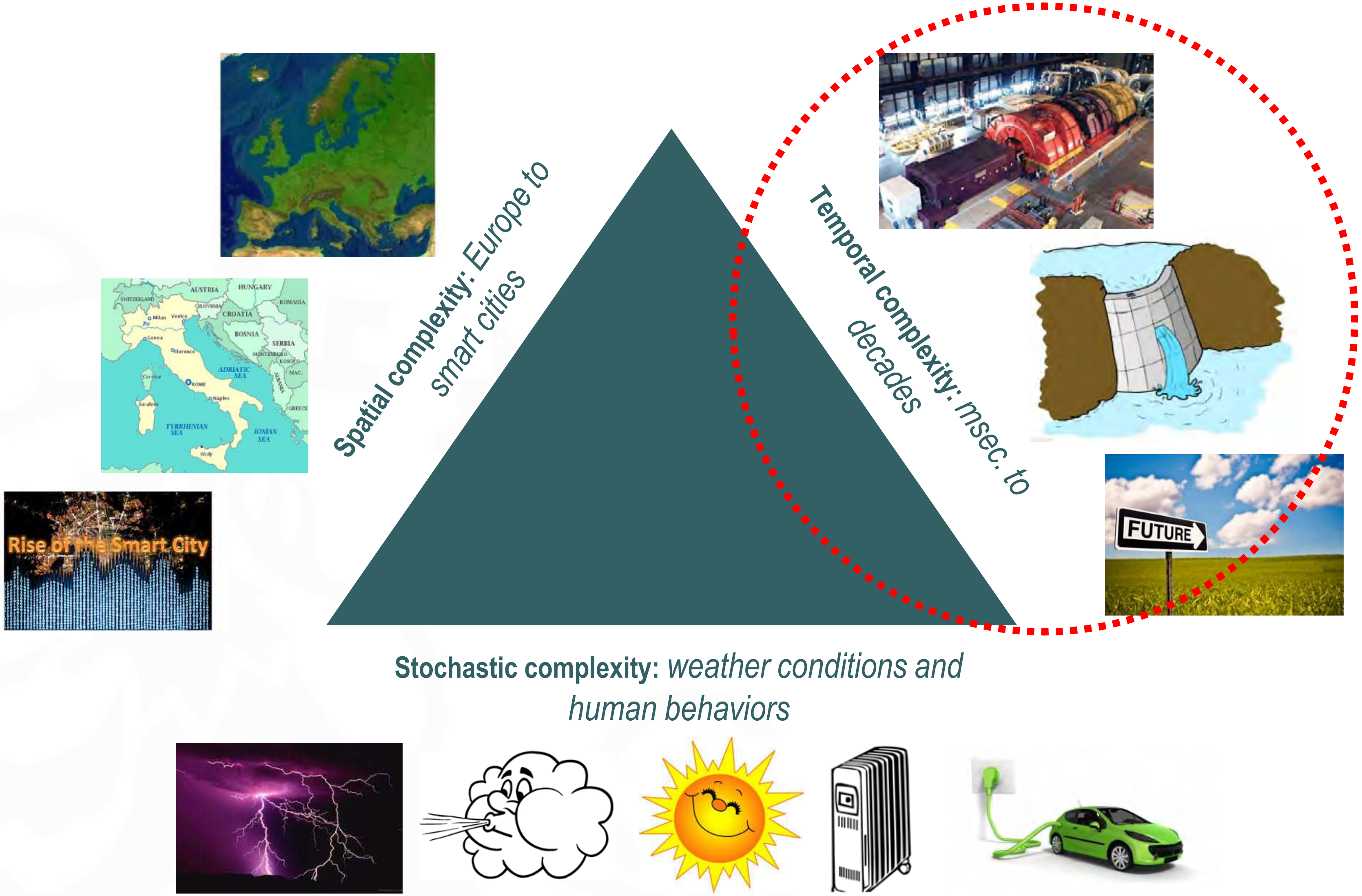
1

Introduction

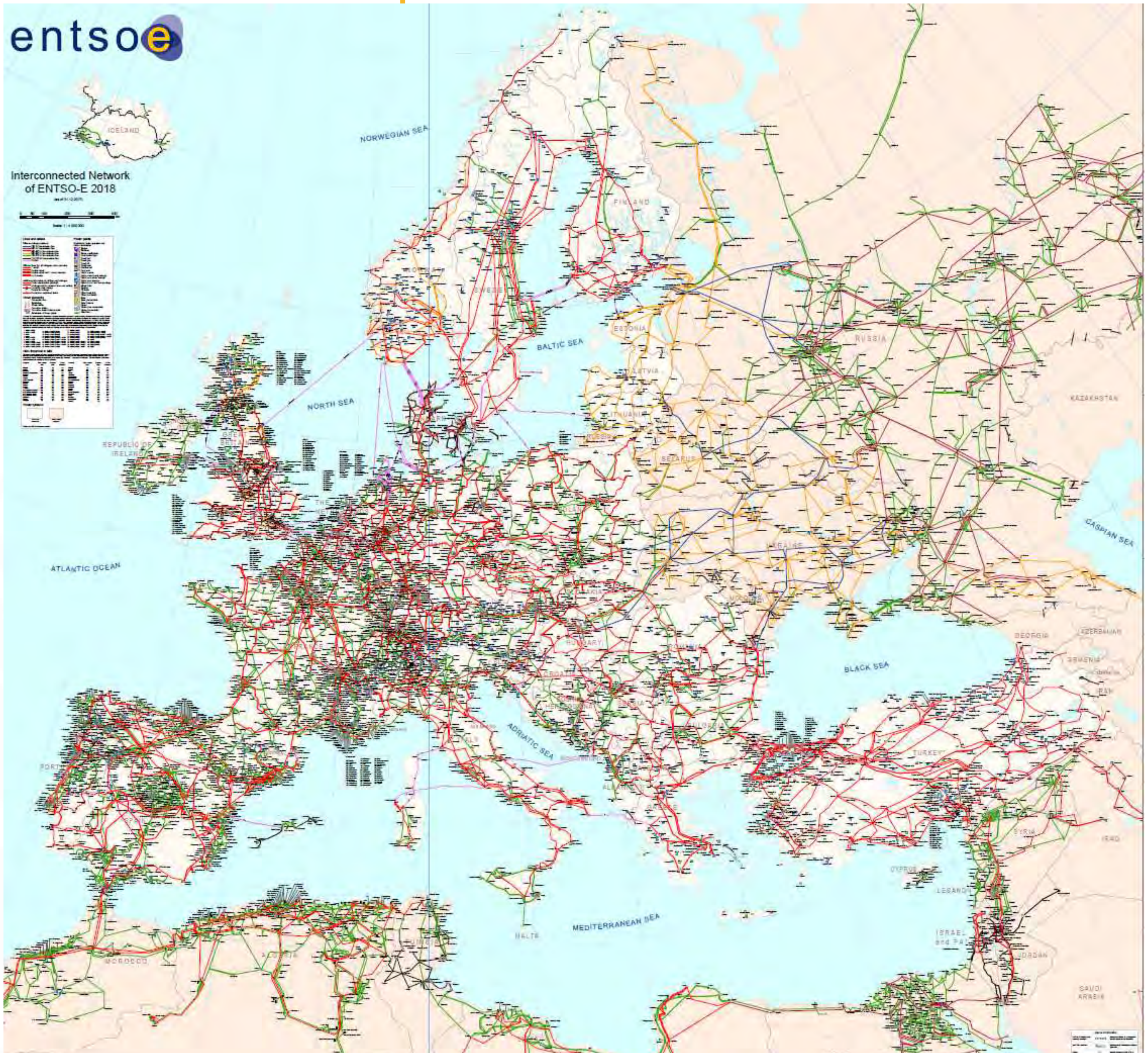
Definition

- Determine which lines and transformers and when to build optimizing total investment and operation costs

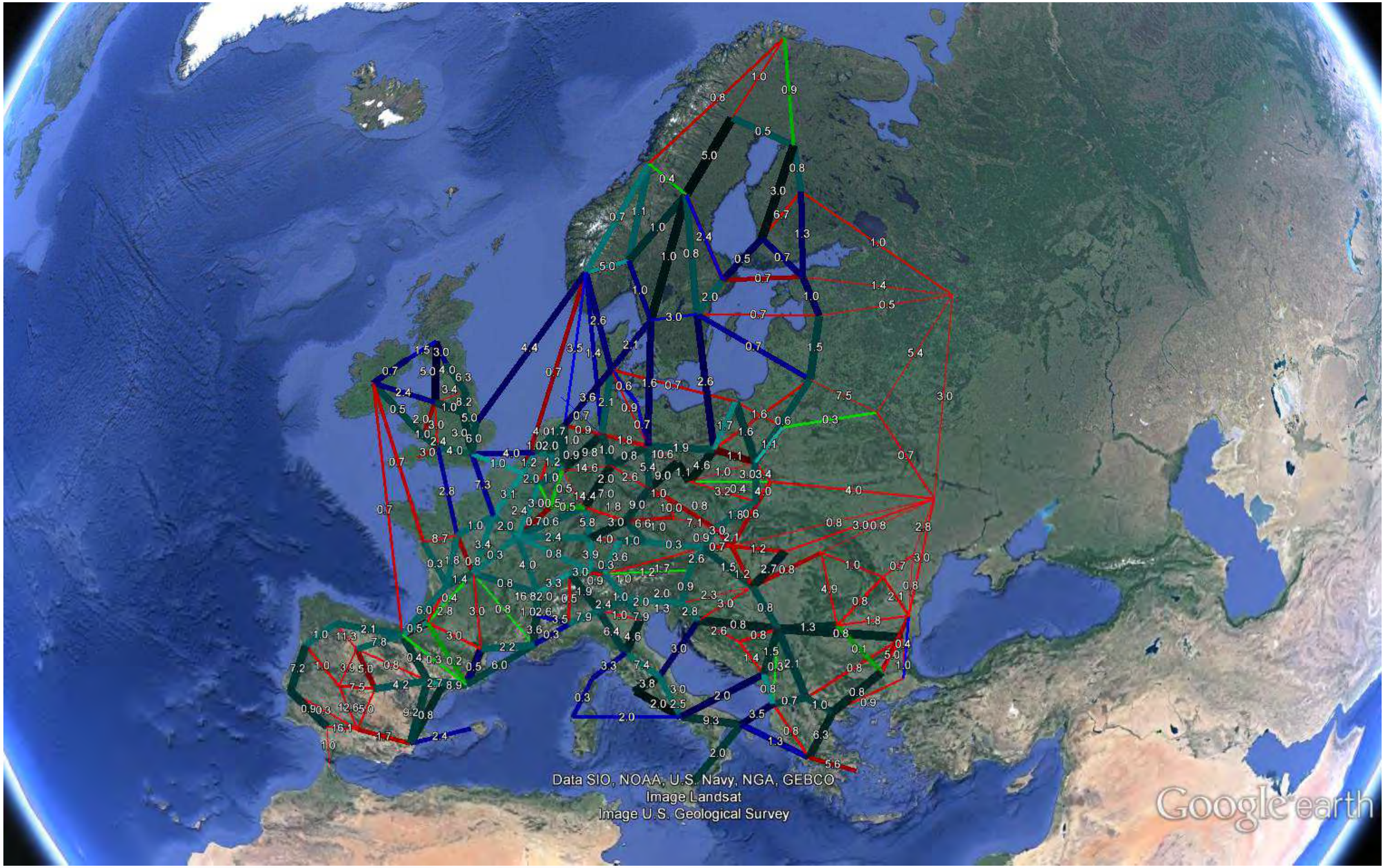
Challenges for TEP



ENTSO-E Grid Map 2018



TEP for Europe



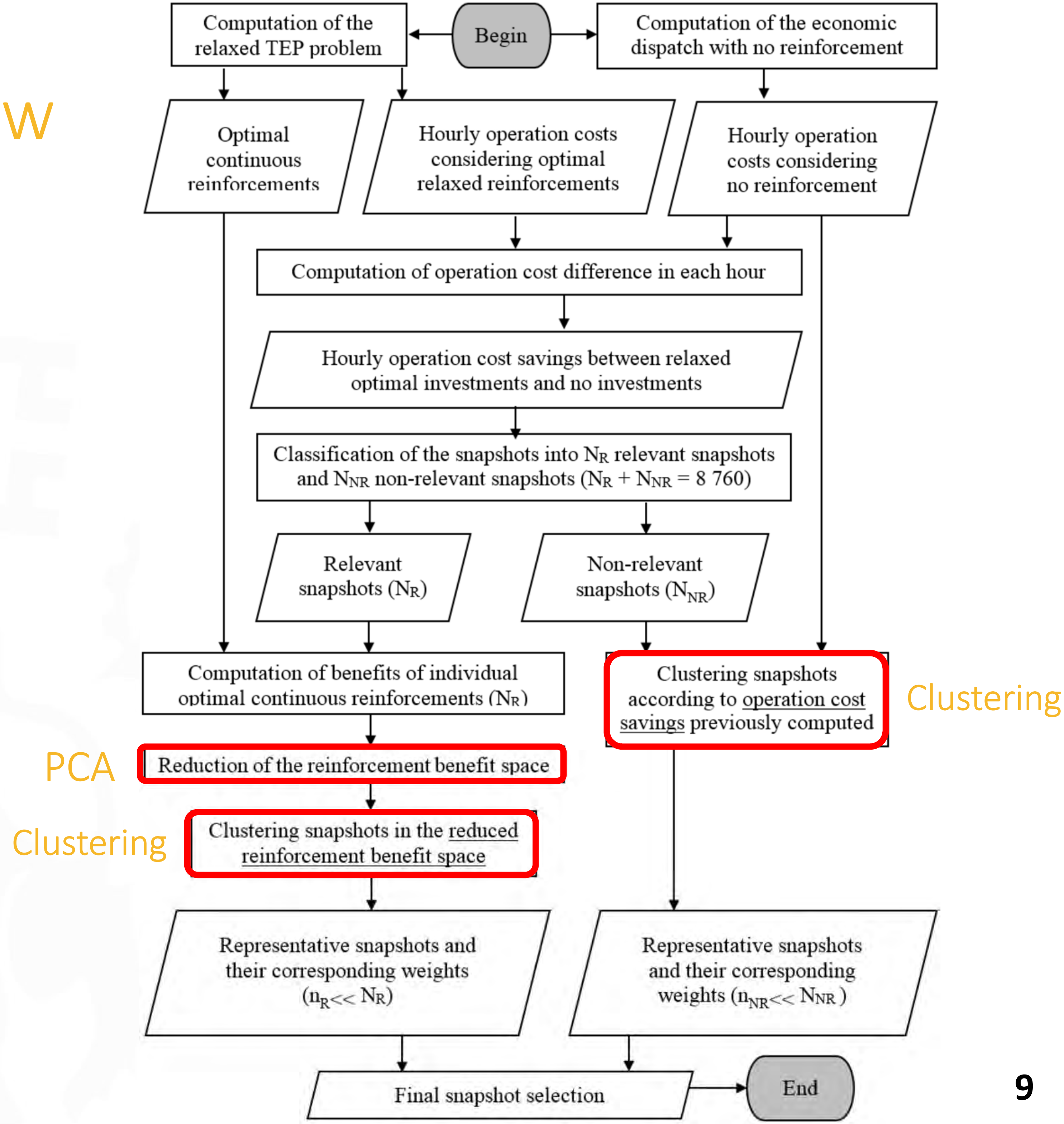
2

Data Analytics for TEP

Reference

- Q. Ploussard, L. Olmos and A. Ramos *An operational state aggregation technique for transmission expansion planning based on line benefits* IEEE Transactions on Power Systems 32 (4): 2744-2755 Oct 2017 10.1109/TPWRS.2016.2614368

Overview



Begin
Computation of the relaxed TEP problem
Computation of the economic dispatch with no reinforcement
Optimal continuous reinforcements
Hourly operation costs considering optimal relaxed reinforcements
Hourly operation costs considering no reinforcement
Computation of operation cost difference in each hour
Hourly operation cost savings between relaxed optimal investments and no investments
Classification of the snapshots into NR relevant snapshots and NNR non-relevant snapshots (NR + NNR = 8 760)
Relevant snapshots (NR)
Non-relevant snapshots (NNR)
Computation of benefits of individual optimal continuous reinforcements (NR)
Clustering snapshots according to operation cost savings previously computed
Clustering
PCA
Reduction of the reinforcement benefit space
Clustering
Clustering snapshots in the reduced reinforcement benefit space
Representative snapshots and their corresponding weights (nR<< NR)
Representative snapshots and their corresponding weights (nNR<< NNR)
Final snapshot selection
End

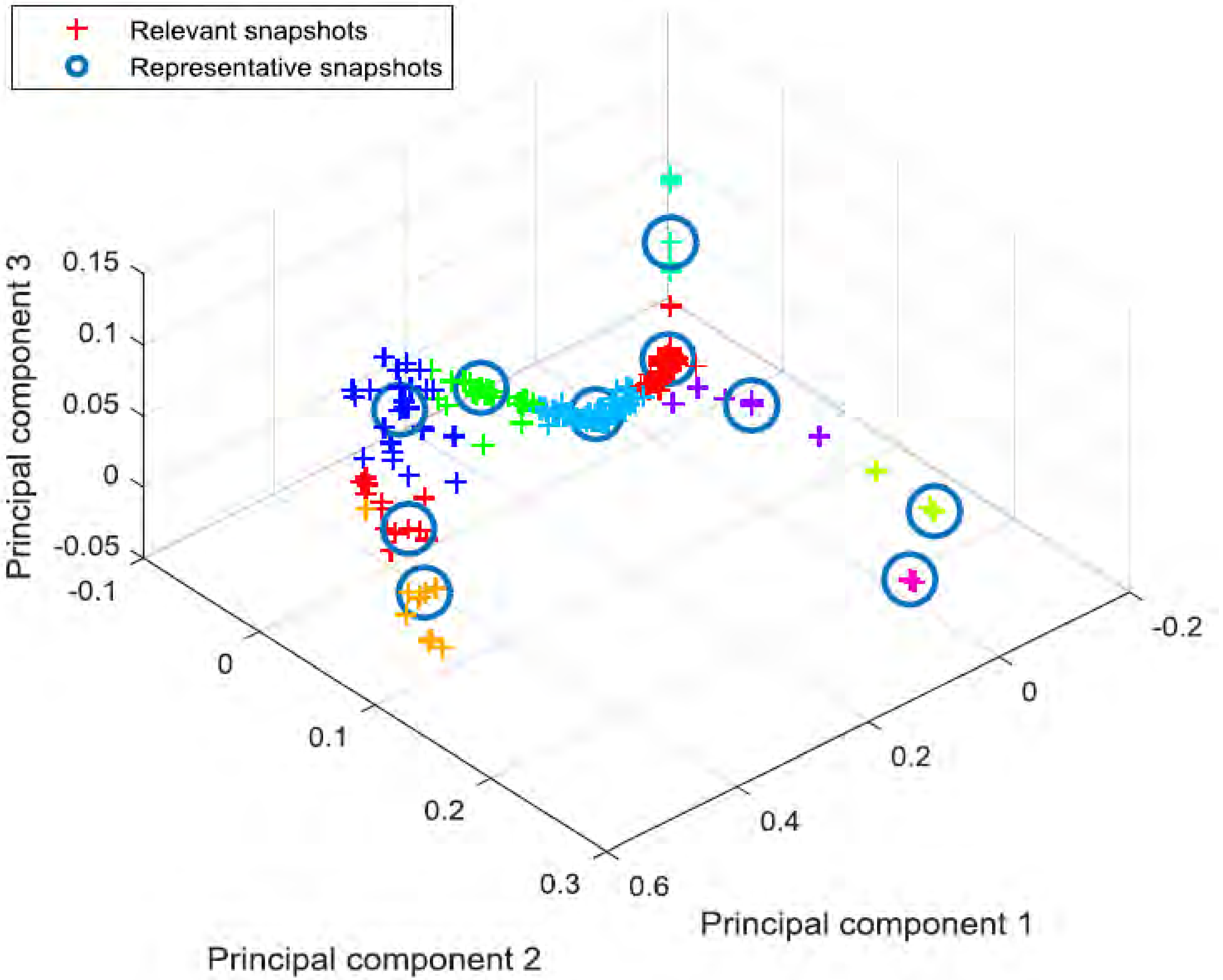
Principal Component Analysis (PCA)

- In order to reduce the dimension of the benefit space, Principal Component Analysis (PCA) is applied to original reinforcement benefits.
- Transforms a dataset in a new one that provides the same information as the original one but in a more compact way. The transformation of the original dataset is defined in such a way that the first principal component of the transformed dataset has the largest possible variance (that is, accounts for as much of the variability in the dataset as possible), and each succeeding component has, in turn, the highest variance possible under the constraint that it is orthogonal to the preceding components. The major part of the variability across operation situations existing in the initial reinforcement benefit space is captured using the first Principal Components.
- Thanks to this process, the dimension of the reinforcement benefit space can be drastically reduced without incurring an important loss of relevant information about the operation situations to be clustered.

Clustering

- The K-means algorithm is applied to cluster operation situations within the reduced reinforcement benefit space (comprising the principal components of the original reinforcement benefit space). The Euclidean distance is used to compare operation situations. There are different possible criteria to measure the goodness of a clustering of samples (operation situations) for a given number of clusters K. The one used aims to minimize the sum of distances of samples to their associated centroid (center of their cluster), i.e., the intra-cluster distance.
- Most clustering algorithms start from an initial set of K samples, defined as the initial set of K centroids, and iteratively try to increase the goodness of the clustering of operation situations by modifying the set of centroids. Improving the initial choice of centroids leads to computing a good partition of snapshots in fewer iterations and avoiding trivial partitions (those where some clusters contain only one snapshot). The K-means++ algorithm addresses this concern by increasing the spread of the initial set of centroids. A first centroid is chosen randomly, considering a uniform distribution for the probability of choosing any snapshot as the first centroid. Then, each subsequent centroid is chosen among the remaining snapshots when each of these snapshots is assigned a probability of being chosen proportional to the square distance in the reduced benefit space from this snapshot to the closest snapshot already chosen as a centroid. This process ends when the initial set of K centroids are chosen.

Clustering relevant snapshots considering 10 clusters



Thank you for your attention

Andrés Ramos