Erratum of Step 1

by Gauthier limpens (gauthier.limpens@uclouvain.be)

1 Context

The *EnergyScope Typical days* model relies on typical days selected in a pre-processing step. This step, called **STEP 1** in [1], clusters the day of the years and select typical days. In the model presentation, three clustering methods were compared with a reference one: k-medoids.

Different criteria were used to assess the performances of the clustering methods. The one proposed by Domínguez-Muñoz et al. [2], which is a revisited k-medoids method, had the best performances. Thus, it has been implemented in the code.

As a reminder, this step is independent to the second one, which is the whole-energy system model. Thus, another method could be used to select the typical days and sequence of days.

2 Error

The error was made on the distance matrix. The latter being the distance for two days between their time series. The correct formulation for a norm n is given in Eq. 1.

$$distance(i,j) = \sqrt[\mathbf{n}]{\sum_{k \in DIMENSION} (Ndata(i,k) - Ndata(j,k))^{\mathbf{n}}} \; \forall i \in DAYS \; , j \in DAYS \quad (1)$$

Where, DAYS is a set containing the day's number of the year; Ndata is a matrix containing the hourly value of each time series for each day i or j. In this case, there are 5 time series accounted: electricity, heat, wind onshore, wind offshore and photovoltaic. Thus, the k dimension is 24 hours times 5 : 120 data.

In the current implementation, there were an error in the distance matrix (*distance*) defined in Eq. 2; the square-root has been forgotten.

$$distance(i,j) = \sum_{k \in DIMENSION} (Ndata(i,k) - Ndata(j,k))^2 \; \forall i \in DAYS \; , j \in DAYS \quad (2)$$

3 Impact on the typical days

To assess the impact of the error, we performed a double analysis in order to estimate: (i) the impact of forgetting the square-root (hereafter called *wrong formulation*), and (ii) the impact of choosing another norm instead of the Euclidean one (i.e. norm 2).

Figure 1 shows the impact of the analysis based on two criteria: the clustering error and the duration curve error. Details are described in Appendix B.2. of [1]. The norm 1 has the best performances.

4 Impact on the energy model

The author find out negligible impact on the energy model results. This was expected as the typical days have a little impact on the key performance indicators. These latter being the design strategy, the resources used and the overall emissions of the system.

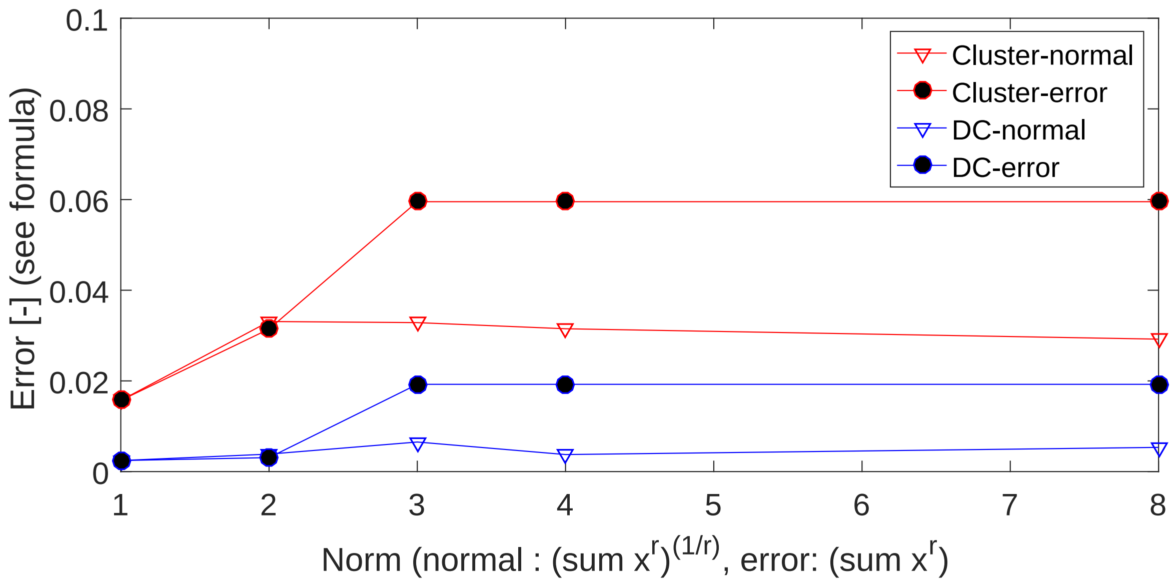

Figure 1: Cluster (red) and duration curves (DC) (blue) errors for 5 different norm (x-axis). The triangles refer to the correct formulation of the distance (Eq. 1) and dotted curve for the wrong formulation (Eq. 2). Except for an Euclidean norm, the accurate formulation provides better results. Moreover, the norm-1 gives the best results.

5 Conclusion

Based on previous analysis, the norm 1 seems the most relevant. The latter is shown in Eq. 3, where the absolute value (abs) of the distance is taken.

$$distance(i,j) = \sum_{k \in DIMENSION} \text{abs}(Ndata(i,k) - Ndata(j,k))^2 \ \forall i \in DAYS\ , j \in DAYS \quad (3)$$

This new formulation has been adopted in the STEP 1.

References

[1] G. Limpens et al. "EnergyScope TD: A novel open-source model for regional energy systems". In: *Applied Energy* 255 (2019), p. 113729.

[2] F. Domínguez-Muñoz et al. "Selection of typical demand days for CHP optimization". In: *Energy and buildings* 43.11 (2011), pp. 3036–3043.