



# **MODELLING OF FLOW-BASED DOMAINS IN ANTARES FOR ADEQUACY STUDIES**

24/10/2017

*Marion LI:* [marion.li@rte-france.com](mailto:marion.li@rte-france.com)  
*Baptiste SEGUINOT:* [baptiste.seguinot@rte-france.com](mailto:baptiste.seguinot@rte-france.com)

<https://antares.rte-france.com>

## Summary

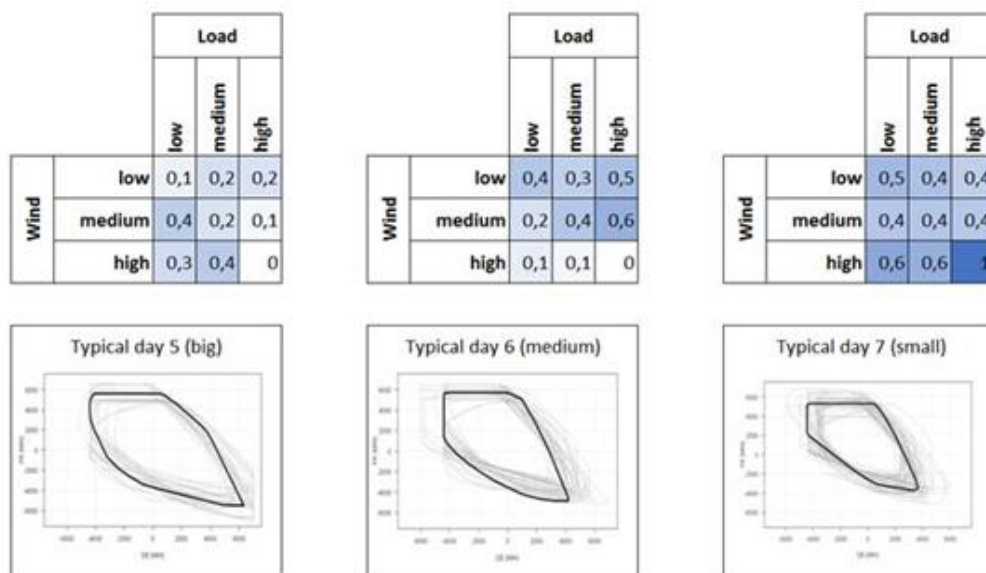
The flow-based method is used to calculate cross-borders capacities since May 2015 in the CWE area (Belgium, Germany, France, Luxemburg and the Netherlands). Two phenomenon have been observed: this method has been on overall increasing the exchanges (in capacity and volumes) while the winter 2016-2017 has proven that the increase is not necessarily correlated to tensed situations during which the French adequacy of supply is threatened. Being able to better represent the exchange capacities offered by the flow-based methodology is therefore a stake in adequacy studies.

RTE's note *Typical Flow-Based Days Selection* described how representative typical days are chosen out of a history, this note focuses on the modelling of these domains in ANTARES. Two main parts are developed: the extrapolation of historical flow-based domain to a medium term horizon and the addition of flow-based domains in ANTARES.

The first part of the modelling consists in creating scenarios of flow-based domains for medium term studies.

To create coherent scenarios, correlations have been researched between the historical shapes of the domains and ANTARES' inputs. Probability matrices have then been calculated: they link each historical day defined by a level of load, wind and solar production, to the probability of being best represented by each possible typical days. An example is displayed in Figure 1: for instance, during winter working days, when there is a high wind production for instance, it is more probable to have the "small" typical day.

Figure 1: Example of a winter probability matrix for 3 typical working days in winter

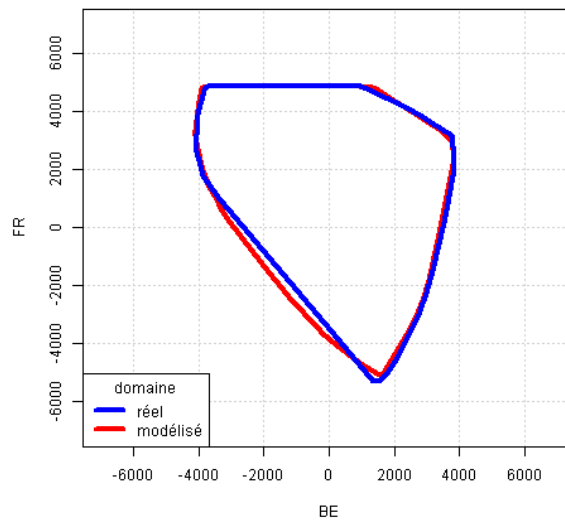


This method enables to highlight tendencies but not to accurately explain with those three variables the flow-based domains. This uncertainty is however a reminder that other factors, especially grid data, intervene in the building of the real domains, but are not available as inputs of ANTARES.

The second part focuses on the modelling of domains itself. ANTARES enables the writing of binding constraints, which are similar to the flow-based's PTDF, except that they are constant over time.

An optimization algorithm creates domains which are compatible with Antares format, while approximating the real domains with a 5% error in volume and a perfect accuracy on maximum import capacity for Belgium, France and Germany (export capacity as well for this country).

Figure 2: Real (blue) and modelled (red) domains (22/12/2015 4:00) projected on Belgium and



To deal with situations of generation shortages, an adequacy patch has been added to the operational market-coupling algorithm. The adequacy patch ensures that the Energy Not Supplied (ENS) is fairly distributed among the CWE countries and avoid counterintuitive results where (e.g.) countries with ENS are exporting. This patch has also been modelled within the methodology presented in this document. It consists in a post-processing of the Antares results which corrects the number of shortage hours, the ENS volume and the net position of each CWE country.

The steps of these methodologies have been implemented in two open source R packages (called flowBasedClustering and antaresFlowBased).

Currently, this method is used for several medium term adequacy studies in Antares, including the French Adequacy Report 2017 and the Winter Outlook 2017-2018 (RTE), the Belgian Adequacy Report (Elia) and the PentaLateral Energy Forum – the PLEF.

## SOMMAIRE

1.	Introduction.....	4
2.	Reminder: Flow-based domains .....	6
2.1	How is a flow-based domain built?.....	6
2.2	Display of flow-based domains .....	7
2.3	Typical days .....	8
3.	Factors identification.....	8
3.1	Relevant factors.....	9
3.2	Limits of a deterministic approach .....	10
3.3	Probabilistic approach.....	12
3.4	Creation of domains time series.....	15
4.	Modelling in Antares .....	17
4.1	Structure of flow-based modelling in Antares.....	17
4.2	Modelling format.....	19
4.2.1	Measure indicator.....	19
4.2.2	Construction of domain.....	20
4.2.3	Presentation of the modelling algorithm .....	21
4.3	Results.....	22
4.4	Implementation in Antares.....	24
5.	Adequacy patch.....	25
5.1	Definition and importance of the adequacy patch.....	25
5.2	The adequacy patch in Antares .....	27
5.2.1	Principle .....	27
5.2.2	Example 1: Creation of ENS.....	27
5.2.3	Example 2: Redistribution of ENS.....	29
5.3	The case of existing strategic reserves .....	30
6.	Current users and future prospects.....	31
6.1	Users .....	31
6.2	Additional specification to typical days .....	31
6.3	Adaptability to the input flow-based domain .....	31
6.4	Comparison between two domains .....	32
6.5	Additional countries.....	32
7.	Annexes .....	33
7.1.1	Annex 1: Description of the modelling optimisation problem.....	33

## 1. Introduction

**Flow-based is the name of a method calculating cross-border exchange capacities.**

Its role is to provide the market (D-1) with hourly sets of possible exchanges which ensure that the physical limitations of the grid are respected: these sets are called flow-based domains. It has been applied in the CWE area (France, Belgium, Germany, Luxembourg and the Netherlands) since May 2015.

Figure 3: Evolution of the French exchanged energy (extracted from the French Adequacy Report 2016)

Figure 4.26: NTC-based capacity offered until 20 May 2015 and exchanges with the CWE Region in 2015

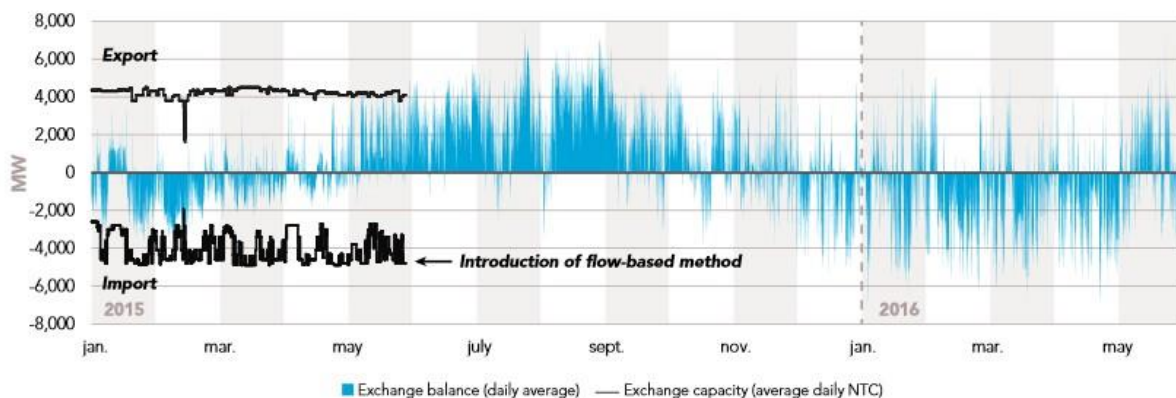


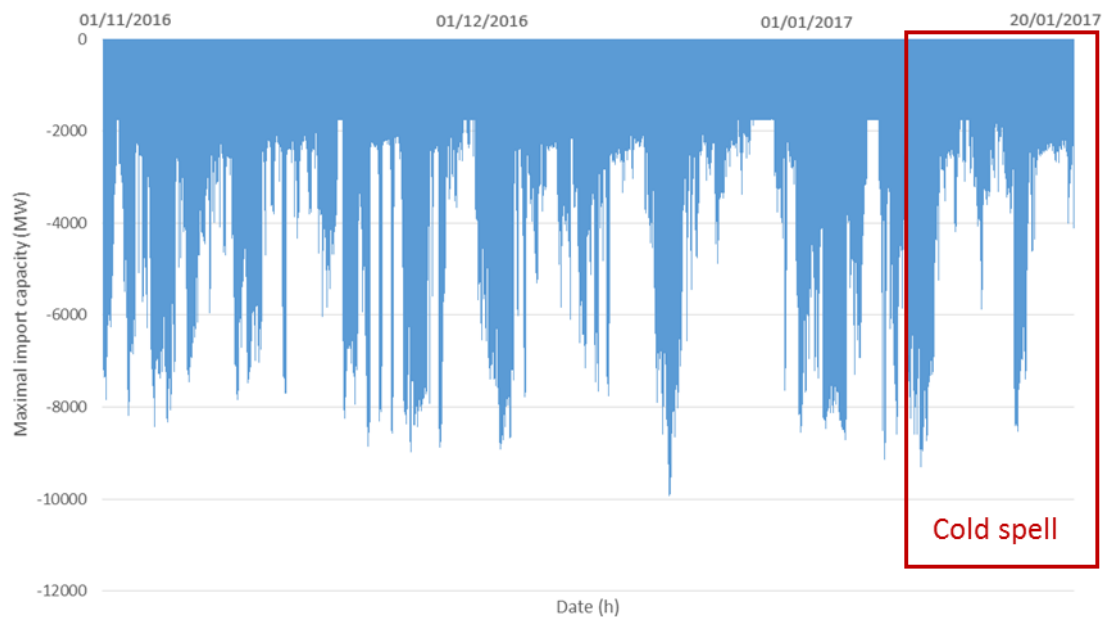
Figure 3 presents an extract of the French Adequacy Report of 2016<sup>1</sup>, displaying the French export and import, before and after the introduction of the flow-based method. On the represented period, the exchanges in the area have increased both in volume and in amplitude. The former maximal capacities, the Net Transfer Capacities (NTC), are drawn in black on the diagram: their maximal value is regularly exceeded since the flow-based method is used, over 6 GW are reached in export (summer 2015) and import (spring 2016).

**The flow-based method has therefore a significant impact on the French import capacity and consequently on the balance between demand and supply:** modelling flow-based domains in adequacy studies has become necessary to ensure the real capacities of cross-border exchanges are taken into account.

However, the flow-based method has been in place for over two years, and **the initial assumption that these domains always bring increased exchanges capacities has been contradicted by the winter 2016-2017**. The maximum cumulated import of France and Belgium over a large part of the winter is presented for each time step (hour) in Figure 4. During this specific period, two tense events for the adequacy of the system happened: numerous nuclear plants in France were unavailable (over the entire represented period) and in January, a cold spell caused a high consumption. Both these events theoretically concur to an increased need for import in France.

<sup>1</sup> [http://www.rte-france.com/sites/default/files/bp2016\\_complet\\_va.pdf](http://www.rte-france.com/sites/default/files/bp2016_complet_va.pdf)

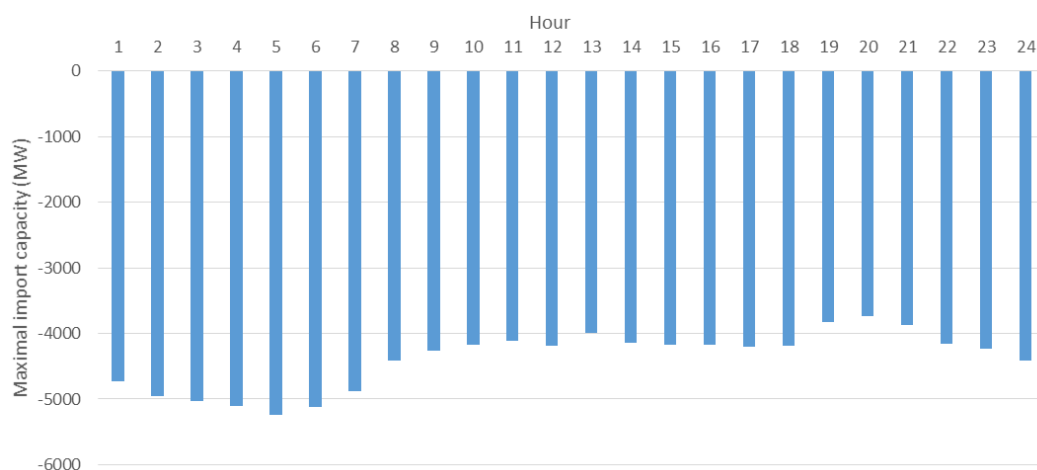
**Figure 4: Maximum hourly cumulated import capacity of France + Belgium from the 1st of November 2016 to the 20th of January 2017**



The first observation deduced from this figure is the variability of the domains over this season, offering from 10 GW to only 2 GW importable by France and Belgium together. Moreover, the two above mentioned periods' domains reveal that an increased need of imports is not necessarily translated into increased capacities: the higher capacities are reached during the weekends (over 8 GW) while during the cold spell especially, most domains did not allow over 3 GW imported simultaneously by France and Belgium.

A similar observation can be made by zooming in on each day of the winter: within the day, the peak hours are the ones where a maximum import capacity would be needed. Figure 5 presents the mean maximal cumulated import capacity per hour over winter: the difference between peak hours such as hour 19 and 20 (18:00 and 19:00) and off peak hours such hour 5 (4:00) is over 1 GW.

**Figure 5: Mean maximal cumulated import capacity of France + Belgium for each hour over the winter 2016-2017**



What can be concluded is that the flow-based method enables to increase the maximal reachable exchange capacity, but not permanently and not necessarily when it is most needed. The first question then raised is: **during strained situations for the demand/supply balance, what is the flow-based domain actually offering in terms of capacities, and what is its contribution to the security of supply?**

To this first interrogation is added a second one: **how can the shape of flow-based domains be predicted?** The real time domains are calculated at D-1, with high knowledge (or prevision) on the grid state, the production plan and the market estimated direction. There are however no indicator currently enabling one to predict an approximate shape of the domain at a further horizon.

The following study presents a beginning of an answer to these questions. It is divided into two documents dealing with the following topics:

- The selection of typical days, aiming to represent the variability of flow-based shapes over a season and the daily variations within.
- The addition of a flow-based simulation for adequacy studies with the software Antares

The following document will present the modelling of flow-based domains in Antares. Based on a set of typical days (see RTE's note *Typical Flow-based Days Selection*), this modelling aims to better represent the exchange capacities in adequacy studies. Two parts are identified to complete this aim:

- The **identification of possible correlation between flow-based shapes and production/consumption patterns**, knowing that multiple factors impact the shape of the domains, but only a few are actually available in Antares studies.
- The **creation of a flow-based model adapted to Antares simulations**, it must be compatible with Antares format of data

This document will first present a brief reminder of flow-based domains and the chosen typical days. The following parts will focus on the search for a climatic pattern explaining the flow-based domains and the modelling of domains in an Antares study, the adequacy patch implementation will finally be discussed.

R packages, named `flowBasedClustering` and `antaresFlowBased`, have been developed accordingly to these methods. The matching R functions of each step of the methodology presented in this document will be written in boxes like this one.

## 2. Reminder: Flow-based domains

### 2.1 How is a flow-based domain built?

The NTC calculation, today used for all but CWE borders, consists in setting for each border an export and import upper limit, on which the two grid operators agree. Having a flow-based approach aims to take into account the interdependencies between the flows crossing the

different borders and proposing a new methodology, shared by all the countries in the CWE area, to maximise the achievable net position, meaning the difference between its exports and its imports of each country.

**A flow-based domain is calculated from an initial situation called base case**, in which critical branches (elements in the grid that may limit the exchanges) are identified. Two kinds of coefficients are calculated from this case: GSK (Generation Shifts Keys) representing how additional exchanges would impact the production units and coefficients measuring how the resulting flows would be distributed among the critical branches. Finally, **a matrix of PTDF** (Power Transmission Distribution Factors) is established and **links to the variation of net position of each country (additional export or import) a corresponding flow variation in every critical branch**. For the exchange to be safely executed, the flow variation must stay below the available margin of the critical branches.

The domain is geometrically defined by, for each critical branch  $i$ :

$$\Delta F_i = ptdf_{FR,i} \times NP_{FR} + ptdf_{BE,i} \times NP_{BE} + ptdf_{DE,i} \times NP_{DE} + ptdf_{NL,i} \times NP_{NL}$$

$$\Delta F_i \leq \text{margin}(i)$$

With  $\Delta F_i$  the flow variation on the branch  $i$ ,  $ptdf_{X,i}$  the PTDF coefficient on line  $i$  for the country  $X$  and  $NP_X$  the net position variation of  $X$ .

## 2.2 Display of flow-based domains

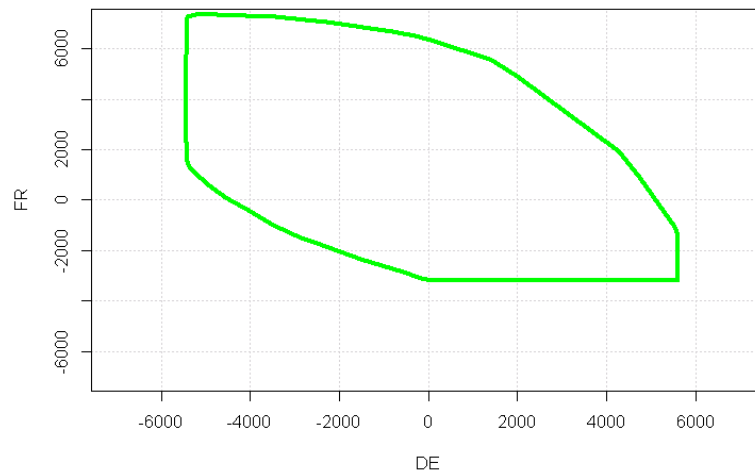
Flow-based domains are described according to the net positions of the four countries in the CWE area. A flow-based domain should then be represented by a four dimensional polyhedron. But within the area, exports and imports must match: what is exported by a country must be imported by another. The balance of net positions is therefore null and the position of a country can be expressed out of the positions of the three others.

**A flow-based domain is finally represented as a three dimensional polyhedron**: each critical branch defines a plane and the most restrictive branches constitute the sides of the polyhedron (an example is depicted in Figure 14, later presented in this document).

In this document, the examples presented will often be projections in two dimensions, which can be read as the possible net positions of two chosen countries. The perimeter is defined by the projection of the 3d-polyhedron on a 2d-plane: to simplify, each straight line delimiting the 2d domain will also be a projected critical branch.



Figure 6: Example of projection of a flow-based domain on France and Germany axis (15/07/2015 at hour 15)



In the case presented above, the grid is able to bear safely all the net positions inside the green area. A positive value corresponds to an exporting net position while a negative value means an importing net position. However, as the third dimension of the flow-based domain is not depicted in this figure, it has to be reminded that some positions can only be reached on condition that some other constraints are respected for example in case of Figure 6 the net position of Belgium.

## 2.3 Typical days

This part is detailed in RTE's note *Typical Flow-based Days Selection*.

The modelling presented in this study will be applied on **a selection of typical days**. They have been selected from a history (01/11/2015 – 20/01/2017) divided in 6 categories: winter, summer and interseason with for each, a category for working days and another for weekends. The clustering is based on an optimization algorithm which calculates a distance between two days by comparing the three-dimensional shape of each hour's domains and adding the differences on the 24 hours of the day.

The chosen typical days are quoted in Table 1.

Table 1: Typical days

Summer		Winter		Interseason	
16/05/2016 - 30/09/2016		01/11/2015 - 15/03/2016 01/11/2016 - 20/01/2017		16/03/2016 - 15/05/2016 01/10/2016 - 31/10/2016	
Weekday	Weekend	Weekday	Weekend	Weekday	Weekend
1 19/07/2016	4 13/08/2016	5 03/02/2016	8 06/02/2016	9 22/03/2016	12 09/04/2016
2 22/09/2016		6 08/02/2016		10 05/10/2016	
3 22/06/2016		7 10/12/2015		11 13/10/2016	

## 3. Factors identification

The clustering from which typical days are ensued is based on resemblance of geometrical shapes. The shape of a flow-based domain is defined by the topology of the grid, the margins of the critical branches and the way power injections are distributed on the grid at a specific

time. This means that to attribute typical days to new situations (i.e. to use them in simulations other than replay the past), one must either:

- Have all grid and market data available to build the real domain and choose the closest typical day in terms of distance: in this case, the real domain is known in details, the association to a typical day is only a means to later reduce the simulation time
- Find some empirical factors, such as climatic conditions or plants availability which characterise a cluster and its typical day: the aim is to *predict* the typical day that would better fit the chosen simulated day without calculating the domain from scratch

In software simulating the balance between demand and supply such as Antares, the grid is not represented in details. There is not enough data to calculate flow-based domains.

The aim of this part is to identify a link between typical days' clusters and external factors, enabling to assign an historical typical day to a future studied day in Antares. The factors must then be available in Antares studies; the objective of this part is to determine how supply-demand balance indicators can categorise typical flow-based days.

### 3.1 Relevant factors

Numerous factors impact the shape of flow-based domains and should be taken into account but the limitations are set on the availability of data:

- **Grid topology: the availability and layout of the elements of the grid** are essential to identify the critical branches and the extension of their margins. However, Antares has no detailed information on the grid, because each country is usually represented by a single node. These data are then not available for the characterisation of flow-based domains, but the problem may be bypassed by considering that two kinds of topological changes can happen:
  - o Controlled topological actions such as the position of phase-shifter transformers or parades: they will be considered in the study as well represented in the chosen history. One will consider that the same situation in a future horizon will lead to the same topological actions.
  - o Uncontrolled actions (from the daily operators point of view), such as an infrastructure consignment, are unpredictable and non-reproducible, they will be considered as an uncertainty on the domain
- **Location of the production:** the location of production is essential to calculate the PTDF coefficients from the GSK data. It depends from multiple factors, such as the location of the production units, the unit's availability and the market price at each hour. However, a country being already a node, there is no data locating the unit and enabling to identify the parts of the grid affected by its behaviour. This will be considered as an uncertainty on the chosen domains.
- **Load and renewable energy production are variables which are highly dependent upon the climatic hazards.** Their impact is double. First, they affect the load level of the lines on the entire grid, as their increase impacts the injection in the different nodes of the grid. Secondly, they influence the need of exchanges between countries. For example, a high wind energy production in Germany will mean the country will tend to export, while high consumption in winter in France will be translated by a need of imports. The topological actions led by the grid operators will lead in the

first case to extend the German export part of the domain and in the second example to enlarge the French import part in the flow-based shape. These variables are represented in Antares via correlated climatic scenarios.

### Three climatic factors are then selected:

- The French load level: highly influent on the country being exporter or importer
- The German wind production: highly correlated to the level of export than can be achieved
- The German solar production: same reason as above

The Antares data are hourly, as are the domains, but the clustering is performed on days. What is studied will be the mean value over the day of each factor.

## 3.2 Limits of a deterministic approach

The aim of this part was initially to establish a scheme explaining, for each day in the history, why they are represented by a specific typical day. Each typical day being the representative of a cluster, it amounts to find common characteristics within each cluster of days. The expected result would be for example to identify one typical day as representing high wind production in Germany during winter. Such a correlation however – if it exists – is highly dependent upon the history.

**The first qualitative approach consists in gathering for each cluster the climatic data of the days it contains.** The French load and German wind and solar productions are hourly values extracted from the platform Transparency (ENTSO-E database). Each day will be represented by its mean values, calculated in load factors<sup>2</sup>.

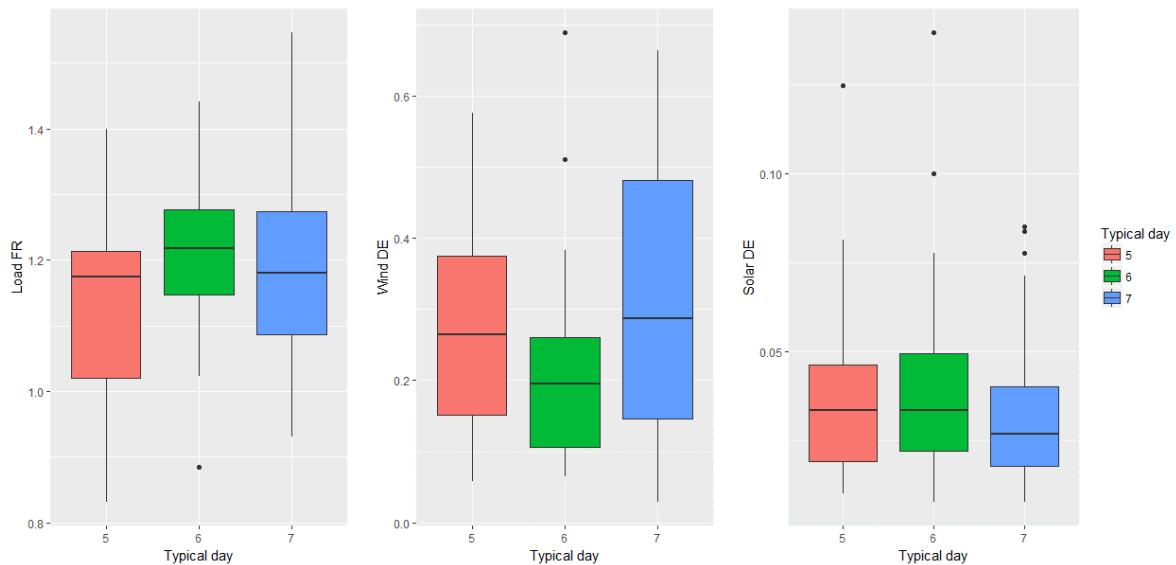
This part presents the study on winter days. Table 2 and Figure 7 present a summary of each cluster's characteristics, underlining the fact that the ranges of values reached in each are overlapping.

Table 2: Characteristics of the winter typical days

Season	Factor	ID	mean	max	min	Standard deviation
winter	Fr load	5	1,13	1,40	0,83	0,13
		6	1,21	1,44	0,88	0,11
		7	1,19	1,55	0,93	0,15
	De wind load factor	5	0,27	0,58	0,06	0,15
		6	0,20	0,69	0,06	0,12
		7	0,33	0,66	0,03	0,19
	De solar load factor	5	0,04	0,12	0,01	0,03
		6	0,04	0,14	0,01	0,03
		7	0,03	0,09	0,01	0,02

<sup>2</sup> Load factors are used for wind power and solar generation (the load factor is the equal to the generation divided by the installed capacity and ranges from 0 to 1). A normalized value is used for the load, equal to the load divided by the average yearly load.

Figure 7: box plot of normalised FR load, DE wind and solar load factors



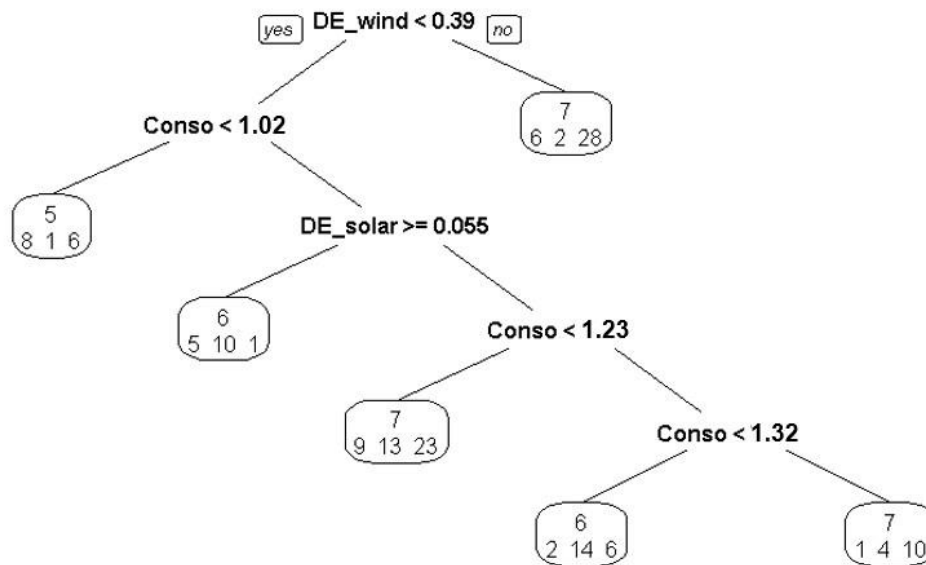
The normalised load values displayed are calculated based on the mean French electricity consumption on the entire year. These are winter days so all clusters are composed of over at least 80% of days during which the load is higher than the yearly mean value. Nevertheless, all the clusters seem to describe the entire scope, with no distinctive difference on the load enabling to distinguish them.

Unsurprisingly, the solar load factors do not allow to distinguish the three different clusters (the three boxes are similar).

Wind load factors are the most decisive data. Even if all the three clusters cover the same range of load factor, from 0 to 0.7, a clear distinction can be made between the spans of the cluster 6 (low wind) and cluster 7 (high wind).

This intuition is verified through a probability tree, drawn in Figure 8, which enables a more thorough approach by combining the factors. Each box contains the days matching all the conditions above, with in first line the typical day they would be affected and in second line, the number of days actually belonging to respectively cluster 5, 6 and 7.

Figure 8: Winter probability tree for typical days 5, 6 and 7



It can be observed on the tree that 93 days are well sorted: it represents 62 % of the season.

The conclusion on this method is that typical days cannot be efficiently sorted using a combination of deterministic factors: none of the combination enable to fit the real distribution of the domains in the clusters.

What can be deduced is that these factors are not influent enough over the shape of the flow-based domain to entirely explain it by themselves. Given that climatic factors are only one of the many variables taken into account in the calculation of domains, a deterministic approach is not the most robust option.

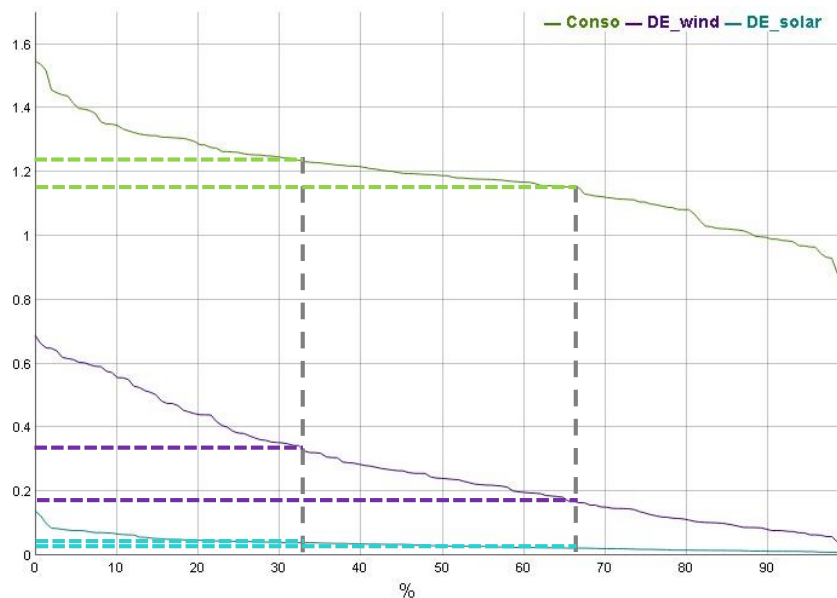
### 3.3 Probabilistic approach

Some correlation has been found between the climatic factors and the clusters distribution, but not enough to support a deterministic approach, especially as the learning period is not so long (in average 30 days per cluster), and **the correlations might be over fitted**.

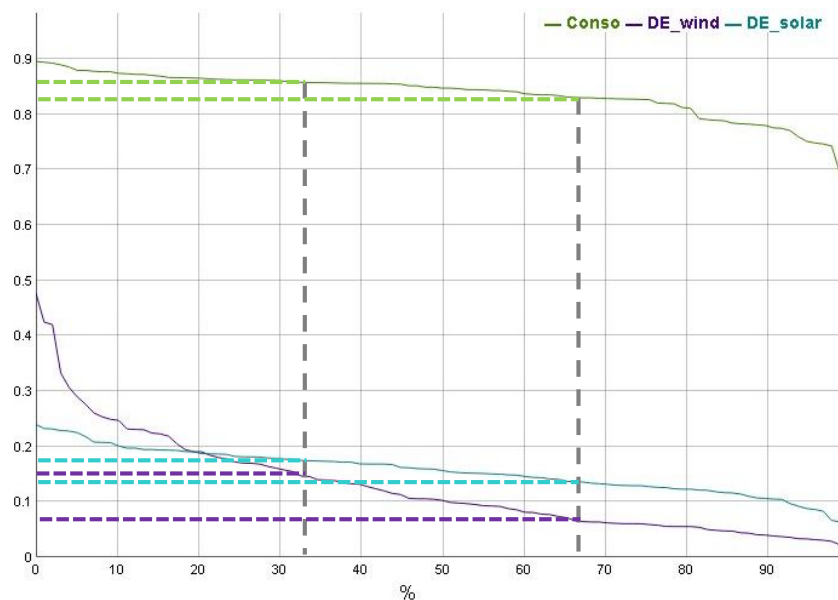
**Having a probabilistic approach enables to reveal the tendencies based on climatic factors, while still having a certain level of uncertainty.** This uncertainty is even welcome, as it is also coming from factors that have not been considered so far (topology, production location).

The idea is to **identify three levels in the variables: low, medium and high**, calculated on the normalised values to be able to be applied on different time frames. The normalised values are, for the load, the relation between the value and the historical mean load, and for renewables, the load factors. The levels are differently set depending on the season, for example load values are higher during winter; the low level of this season may be higher than the high level of summer.

**Figure 9: Top down winter normalised/load factors**  
The limits of each level are signalled with dotted points



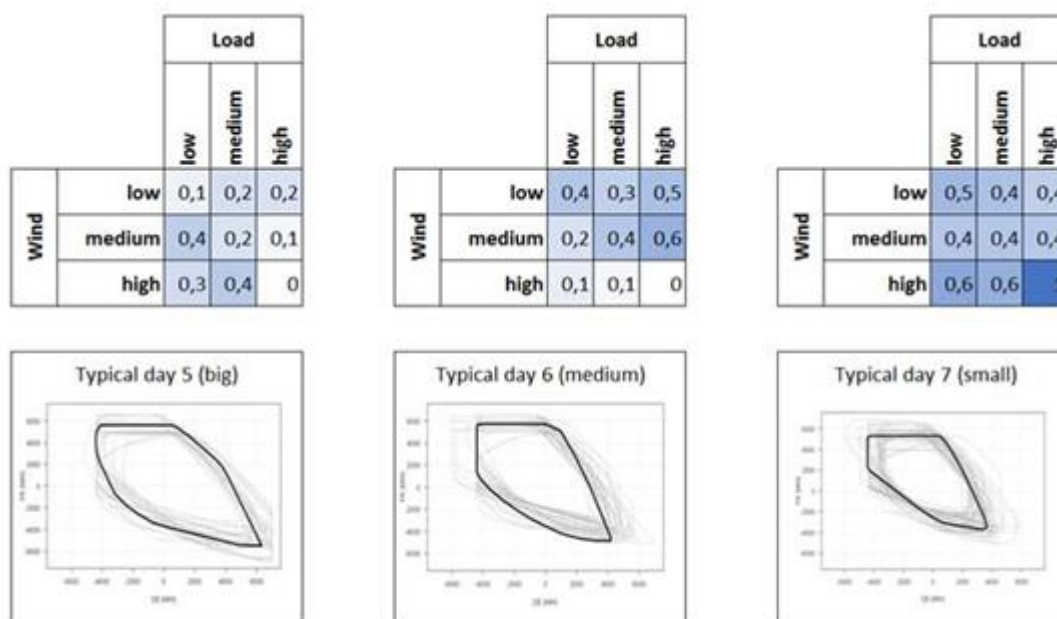
**Figure 10: Top down summer normalised/load factors**  
The limits of each level are signalled with dotted points



In Figure 9 and Figure 10, all winter weekdays are represented without differentiating the clusters, from this graphs are the levels low to medium and medium to high set. In this case the curves have a quite constant slope, the level limits can be set at 33 and 66 %. It can be useful to be able to move the limits, for example to isolate a plateau or extreme (high or low) values, as it exists on summer wind.

Combining them creates 27 categories describing the entire season. Depending on their climatic conditions, all days in the history are attributed to one of these categories. Then for each category, the number of days belonging to each cluster is accounted and from it is deduced a probability of occurrence. The results are presented as probability matrices. Considering the evolution of the solar load factor over winter in Figure 7 and Figure 9, this variable can be considered as non-influent for this season, which reduces the number of winter cases to 9.

Table 3 : Winter probability matrix



**This method enables to highlight schemes that the deterministic method did not identify:**

- The big domain (typical day 5) is unlikely to occur when the French load is high. In those situations the medium domain (typical day 6) or the small one (typical day 7) are the most represented. Moreover, in historical situations where the German wind and the French load were both high, only flow-based domains of cluster 7 appeared.
- The typical day 6 represents mostly situations with low to medium wind generation factors.
- The typical day 7 is the one which has the highest probability of occurrence. It can appear in any situation.
- In situations with high wind, the small domain (typical day 7) is most likely to occur.

The history being short, there might be some empty categories, in which no days fit. The probability in those cases is calculated as the mean probabilities of the surrounding categories, or –if the surrounding categories are themselves empty – as equally probable

The limitations of this method are mainly due to:

- The short history, making it hard to be able to obtain consistent results for 27 cases in summer and interseason

- The description of the exogenous factors with one average daily value, without considering the variations which can occur within the day
- Other possible influent factors that are not considered

Additional studies will be led by the R&D in order to investigate the two last limitations and notably analyse other correlations (e.g. with the nuclear capacity in maintenance) which could explain the shape of the flow-based domains.

R package: `flowBasedClustering`

The function `getProbability` carries out the calculation of occurrence probability for each climatic situation. It takes in input a table of daily climatic data (in this example `Fr` normalized load, `De wind` load factor and `De solar` load factor), the set of typical days and their attributes and a list of limits defining the levels (low-medium-high). The default levels are at 0.33 and 0.66, but it is possible to choose different values for each factor. A boolean indicates whether cases without values are or not to be automatically completed.

### 3.4 Creation of domains time series

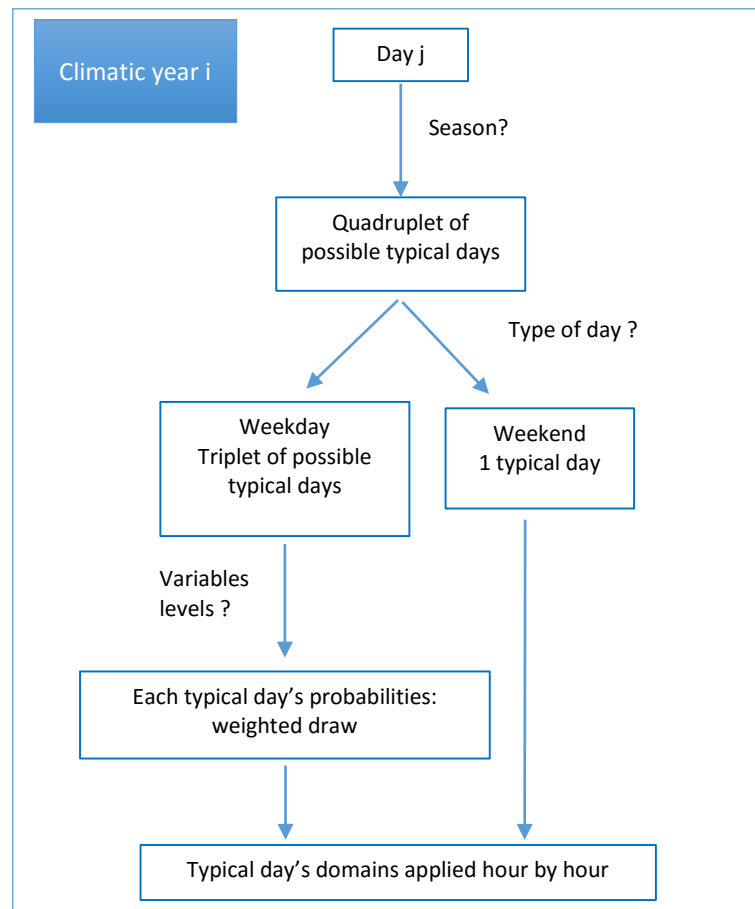
For all countries modelled in Antares, load and renewable energies production belong to the inputs of the study. Load, wind production and solar production are calculated beforehand as hourly time series, based on climatic scenarios coming from Meteo France (for the French Adequacy Report especially). There are 200 scenarios establishing temperature, sunshine, cloudiness, from which are derived these time series: they are correlated between variables (for example the wind generation and the load at a certain time are derived from climatically consistent temperature and wind speed scenarios) and in space as they span over all Europe. The probability matrices calculated before need only these data to characterize typical days, the idea is to associate to the 200 climatic scenarios consistent flow-based domains time series.

**For each climatic scenario, each day is attributed a typical day:**

- Depending on the type of day (weekend or working day), only one typical day can be possible (weekend) or the choice must be made among three (working day)
- If it is a week day, the normalized French load, the German wind and solar load factors are calculated (as mean values for the day) and the result enables to match the day to a specific entry in the probability matrix. Using the indicated probabilities, a probabilistic draw is made to determine the typical day.
- Once the typical day is chosen, the domains are attributed hour by hour, by a method detailed in section 4.



Figure 11 : Attribution of typical days



All this calculation is then done before the Antares simulation; the probability matrices are deduced from the typical days set and can be considered as fixed for any study with these days.

Flow-based time series, as explained above, depend on a probabilistic draw, it might therefore be possible to try several time-series for the same climatic scenario. However, any time series calculation must be carried out before launching the simulation, and is considered as an input to the study. The next step will be, for each year simulated in Antares, to read the flow-based time series and to model the matching domains in Antares.

R package: `antaresFlowBased`

The domains time series are produced by the function `createFBTS`. It requires the study's input (access to climatic data), and a calendar to determine the season delimitations and the weekend positions.

## 4. Modelling in Antares

### 4.1 Structure of flow-based modelling in Antares

The flow-based model is based on the binding constraints in Antares, which are constraints limiting the capacities of the links between the nodes. They are written as equations such as:

$$\beta_{fr \rightarrow be} L_{fr \rightarrow be} + \beta_{be \rightarrow de} L_{be \rightarrow de} + \dots \begin{matrix} < \\ > \\ = \end{matrix} b \quad \text{with} \quad \begin{cases} \beta_{i \rightarrow j} \text{ weight on the link } i \rightarrow j \\ L_{i \rightarrow j} \text{ flux on the link } i \rightarrow j \\ b \text{ second member, margin} \end{cases}$$

**The form of these constraints is very close to the definition of flow-based domain**, as they can be equivalently expressed as linear combinations of net positions or flows of exchanges between them (the net positions being the sum of all in and out flows). In this way,  $\beta$  represent the PTDF coefficients and  $b$  the margin on the critical branch.

In the most recent version of Antares (v6), **it is possible to write and use a huge number of binding constraints, but the coefficients (weights)  $\beta$  are fixed for the whole simulation**. Figure 12 presents a screenshot of the table enabling to set these coefficients. Each column is the name of a binding constraints and each line represents a link between two nodes (described in alphabetical order).

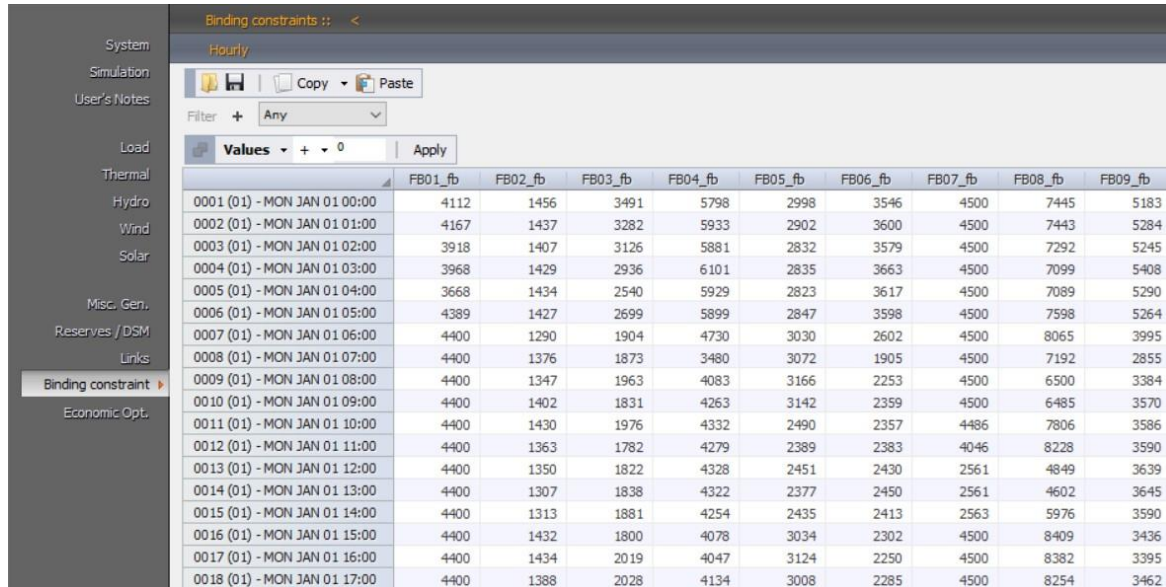
Figure 12: Antares' interface, weights of the binding constraints

Binding constraints :: Weights									
<div> <div>System</div> <div>Simulation</div> <div>User's Notes</div> <div>Load</div> <div>Thermal</div> <div>Hydro</div> <div>Wind</div> <div>Solar</div> <div>Misc. Gen.</div> <div>Reserves / DSM</div> <div>Links</div> <div>Binding constraint ▶</div> <div>Economic Opt.</div> </div> <div> <div>Filter + Any</div> <div>Values + - 0 Apply</div> </div>									
	FB01_fb	FB02_fb	FB03_fb	FB04_fb	FB05_fb	FB06_fb	FB07_fb	FB08_fb	FB09_fb
Bounds	<	<	<	<	<	<	<	<	<
Type	hourly	hourly	hourly	hourly	hourly	hourly	hourly	hourly	hourly
Links	3 links	5 links	5 links	4 links	5 links	5 links	3 links	2 links	5 links
On / Off	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
at / ch	0	0	0	0	0	0	0	0	0
at / de	0	0	0	0	0	0	0	0	0
at / it	0	0	0	0	0	0	0	0	0
at / pump_d	0	0	0	0	0	0	0	0	0
at / pump_w	0	0	0	0	0	0	0	0	0
at / turb_d	0	0	0	0	0	0	0	0	0
at / turb_w	0	0	0	0	0	0	0	0	0
be / de	1	0.16	-0.53	0.37	0.42	-0.02	-1	0	0.16
be / fr	0	0.09	-0.24	0.93	0.18	0.34	-1	-1	0.68
be / lu_be	0	0	0	0	0	0	0	0	0
be / nl	0	-0.49	0.28	0	0.75	-0.44	-1	0	-0.22
be / pump_d	0	0	0	0	0	0	0	0	0
be / turb_d	0	0	0	0	0	0	0	0	0
ch / de	0	0	0	0	0	0	0	0	0
ch / fr	0	0	0	0	0	0	0	0	0
ch / it	0	0	0	0	0	0	0	0	0
ch / pump_w	0	0	0	0	0	0	0	0	0
ch / turb_w	0	0	0	0	0	0	0	0	0
de / fr	-1	-0.07	0.29	0.56	-0.24	0.36	0	-1	0.52
de / lu_de	0	0	0	0	0	0	0	0	0
de / nl	-1	-0.65	0.81	-0.37	0.33	-0.42	0	0	-0.38
de / pump_d	0	0	0	0	0	0	0	0	0

**The variability of the binding constraints over time depends upon the second member**, called  $b$  in the equation above. For each binding constraints, the second member is

described as an hourly time series of length a year. Figure 13 presents a screenshot of the mentioned table, each column being the time series attributed to a binding constraints. In this example the table has for symbol "<", meaning that the values is an upper bound of the constraint.

Figure 13: Antares' interface, second member of the binding constraints' time series



	FB01_fb	FB02_fb	FB03_fb	FB04_fb	FB05_fb	FB06_fb	FB07_fb	FB08_fb	FB09_fb
0001 (01) - MON JAN 01 00:00	4112	1456	3491	5798	2998	3546	4500	7445	5183
0002 (01) - MON JAN 01 01:00	4167	1437	3282	5933	2902	3600	4500	7443	5284
0003 (01) - MON JAN 01 02:00	3918	1407	3126	5881	2832	3579	4500	7292	5245
0004 (01) - MON JAN 01 03:00	3968	1429	2936	6101	2835	3663	4500	7099	5408
0005 (01) - MON JAN 01 04:00	3668	1434	2540	5929	2823	3617	4500	7089	5290
0006 (01) - MON JAN 01 05:00	4389	1427	2699	5899	2847	3598	4500	7598	5264
0007 (01) - MON JAN 01 06:00	4400	1290	1904	4730	3030	2602	4500	8065	3995
0008 (01) - MON JAN 01 07:00	4400	1376	1873	3480	3072	1905	4500	7192	2855
0009 (01) - MON JAN 01 08:00	4400	1347	1963	4083	3166	2253	4500	6500	3384
0010 (01) - MON JAN 01 09:00	4400	1402	1831	4263	3142	2359	4500	6485	3570
0011 (01) - MON JAN 01 10:00	4400	1430	1976	4332	2490	2357	4486	7806	3586
0012 (01) - MON JAN 01 11:00	4400	1363	1782	4279	2389	2383	4046	8228	3590
0013 (01) - MON JAN 01 12:00	4400	1350	1822	4328	2451	2430	2561	4849	3639
0014 (01) - MON JAN 01 13:00	4400	1307	1838	4322	2377	2450	2561	4602	3645
0015 (01) - MON JAN 01 14:00	4400	1313	1881	4254	2435	2413	2563	5976	3590
0016 (01) - MON JAN 01 15:00	4400	1432	1800	4078	3034	2302	4500	8409	3436
0017 (01) - MON JAN 01 16:00	4400	1434	2019	4047	3124	2250	4500	8382	3395
0018 (01) - MON JAN 01 17:00	4400	1388	2028	4134	3008	2285	4500	8254	3462

As the coefficient  $b$  browses a time series, **it is possible to set it at a very high value (representing infinity) to artificially relax one of these constraints**. This enables to remove the constraints at specific times, while using it at others. However, a "removed" constraint will still appear in Antares' optimisation problem and its being fixed to a high value contributes to extend computation time. It is then theoretically possible to model numerous flow-based domains by using and combining sets of binding constraints, but this would be carried out to the cost of computation time.

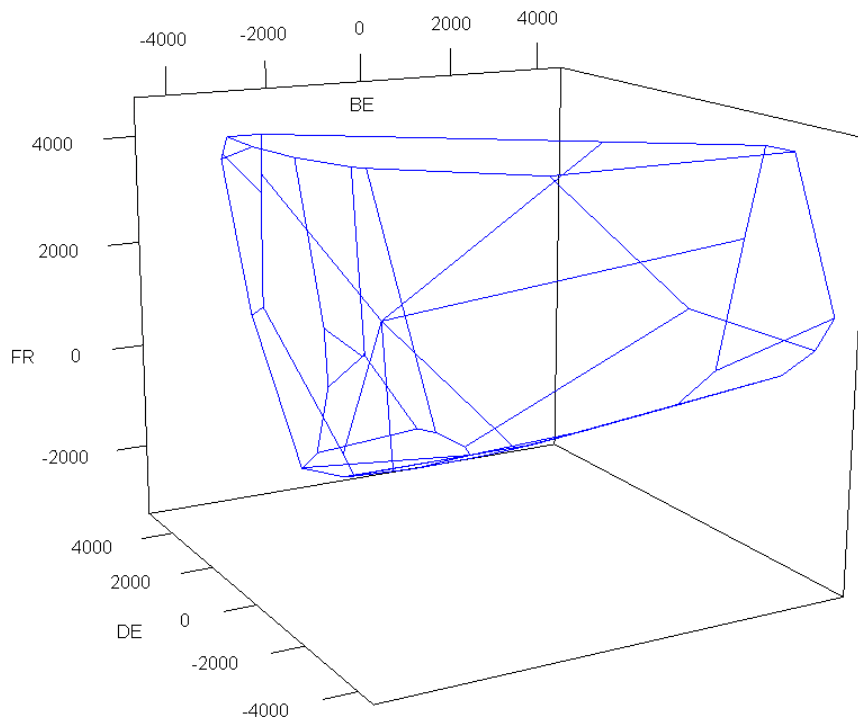
The shape of flow-based domains is determined by the critical branches taking place at that time. However, because of the above mentioned reason and despite the variability of these shapes, **flow-based domains must be represented by a reasonable and finite number of equations, defined by identical weights**. The principle of the modelling is then: flow-based domains are complex polyhedrons, with multiple different sides, the aim is to simplify them to turn them into more regular shapes, while still respecting their own characteristics enabling to differentiate each domain from its following.

As mentioned in 2.2, **a flow-based domain can be represented in a three dimensional space** (each dimension being one country's net position). The four countries (Luxembourg's energy market being attached to Belgium's and Germany's) belonging to the CWE area must in the flow-based calculation get balance between one another: if one country is exporting to the area, the consequent power must be imported by at least one other country. Knowing three countries' net position is then enough to know the net position of the fourth, calculated by:

$$NP_{NL} = -NP_{FR} - NP_{BE} - NP_{DE}$$

With NP the net position of the country.

Figure 14: 3d flow-based domain representation (7th May 2017, hour 1)



Using this representation, the coefficients (weights) defining the equations of the constraints are vectors (normal to the surfaces of the domain) and the second member is the location of the domain limit on the axis defined by each vector.

## 4.2 Modelling format

The modelling relies on two aspects: the choice of a defined number of directions (ultimately the surfaces' inclinations) and the position of the limits on these axis. To build this model, an indicator to measure the resemblance between the model and the actual domain, two polyhedron, must be chosen.

### 4.2.1 Measure indicator

**To sort the models and estimate their resemblance to the original domain, two kinds of error are measured:**

- The first error (called "inf error") makes the count of points representing exchanges which are possible in reality but forgotten in the modelled domain
- The second error (called "sup error") calculates the number of points representing exchanges which are out of the domain but are accepted by the model

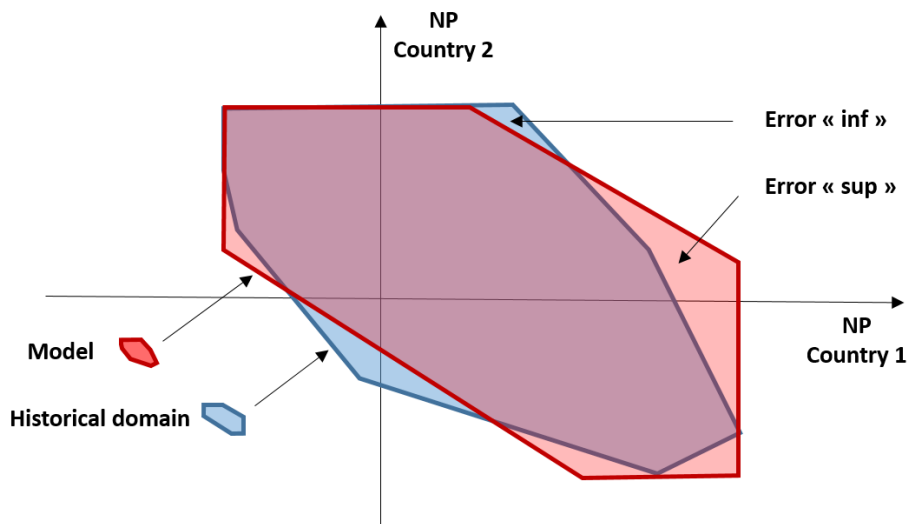
**To estimate these errors, a billion points in 3 dimensions (BE, DE, FR) are randomly drawn to represent numerous exchanges situations.** The Netherlands net position is then deduced from the three others. Each point is then tested:

- The points belonging to both domains are well represented

- The points belonging to only one domain are used to calculate the two volumetric errors (inf or sup, depending on the domain they do not belong to), estimated in %

The Figure 15 represents two domains, the real being drawn in blue and its model in red. The surfaces coloured in blue only are forgotten points and red ones are the invented ones.

Figure 15: Definition of two errors in two dimensions



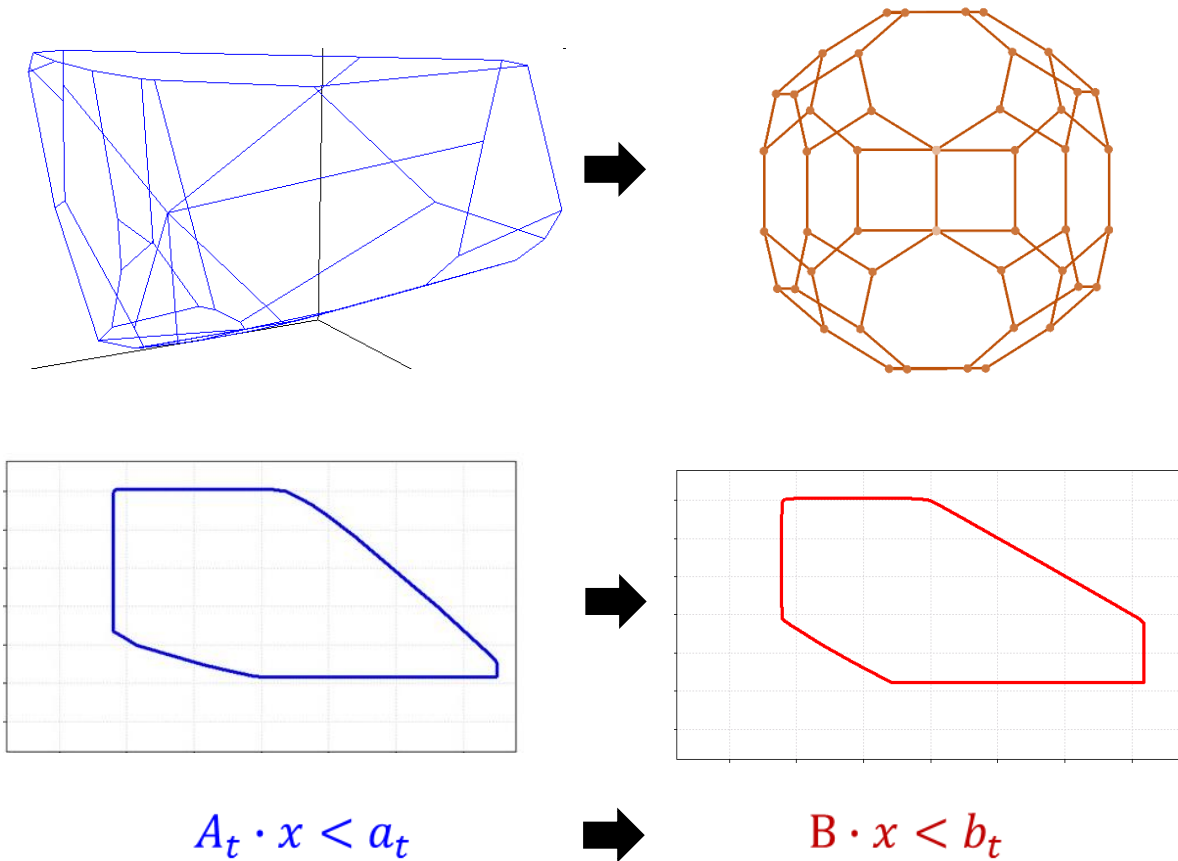
**The nature of the two errors implies their complementarity:** if a limit is lowered to decrease the number of forgotten exchanges, the number of impossible exchanges will increase and inversely. An arbitration between them must be decided, knowing that it would be possible to favour one error over the other.

**The chosen indicator to evaluate the models' accuracy will then be a combination of the two volumetric errors described above.** However, it must be reminded that it is a good evaluator of the shape of the model, but the domain actually represents a set of possible exchanges. The real validation comes from the Antares results: if the forgotten points are exchanges which would not be of interest for the market simulation, the error will not have any impact on the outputs, such as the loss of load. While the volumetric error remains the primary criteria, other aspects, such as the level of maximum imports, must be taken into account.

#### 4.2.2 Construction of domain

The flow-based model must respect the Antares format: as described in 4.1, a domain is represented by a set of equations with fixed coefficients and second members which vary over time. This problematic is illustrated by Figure 16: fixed coefficients (matrix  $B$ ) implies that the number and the inclination of the sides must be standardized. The flow-based shapes can be very different, the aim of the model is to choose the appropriate set of sides: in Figure 16, the standard shape is regular, but it is not mandatory. The second member, i.e. the position of the sides, is however function of time: the second part of the modelling is to distort the standard shape in order to meet at best the real shape of the domain.

Figure 16: Different ways of considering the real (on the left) and modelled (on the right) flow-based domains



Once the directions (matrix  $B$ ) are defined, the second members  $b_t$  are set by an optimization algorithm, enabled to move all sides at the same time while adjusting them to one another.

### 4.2.3 Presentation of the modelling algorithm

#### 4.2.3.1 Standard shape

**The  $B$  matrix is build based on the 288  $A_t$  matrices of the typical domains.** To do so, a clustering algorithm is performed on the PTDF coefficients (lines of  $A_t$  matrices) so as to be able to cover most of the surfaces' inclinations observed in these typical days. Finally, 36 sides' inclinations are deduced to represent the general shape of the entire set of domains, they correspond to the coefficients of the matrix  $B$ .

It must be noticed that with another set of typical days, the results will be different, the standard shape is then directly linked to its typical days.

#### 4.2.3.2 Optimization of the sides position

**The determination of the modelled domain is the result of an optimization algorithm,** whose variables are the vector  $b_t$ . **Its objective is to minimize a weighted sum of inf and sup errors** (defined in 4.2.1).

This entire optimization problem is detailed in Annex 1.

The final domain is defined so that the two errors are similar. Moreover, the algorithm formulation ensures that some initial proprieties of the domain are respected, namely the maximum import of France, Belgium and Germany and the maximum export of Germany.

R package: `antaresFlowBased`

The two steps of the modelling (selection of sides and second member calculation) are carried out by a function called `computeFB`. It takes in input a csv file containing the PTDF constraints of the domains to model. The number of sides, automatically set to 36, can be changed. This function can generate html report for each typical day.

The function `generateReportFb` can afterwards generate said reports.

Once the computation has been made once, the modelled domains are saved in an R object and constitute a "model". Several models are available in the package, the function `getAvailableModel` lists them.

The function `setFlowbasedPath` enables to choose the model to be applied in the next Antares simulation. It can be a model already in the package or any model created by the user and indicated by its path.

### 4.3 Results

Several variants have been tested, with a different number of directions (sides) and with or without respecting the import/export proprieties.

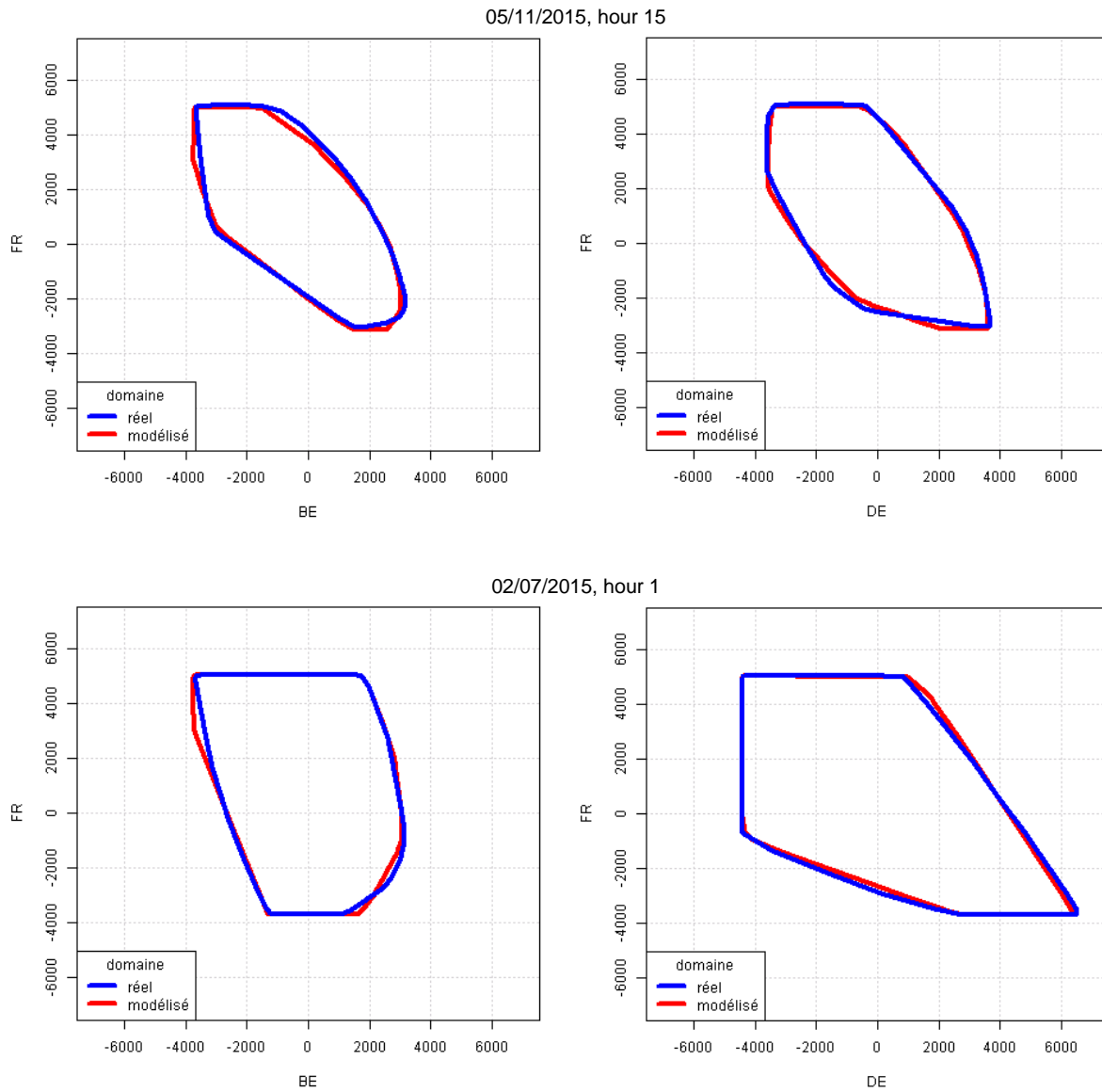
**Using 36 clustered directions, fixed French, Belgian and German imports and German export constraints, the algorithm achieves the creation of domains with high accuracy.** It has been tested on three sets of 12 typical days, so 288 domains, and the maximal error reached (one for each set) is around 10%, most domains are below 6%.

Table 4: Example of "inf" and "sup" volumetric errors

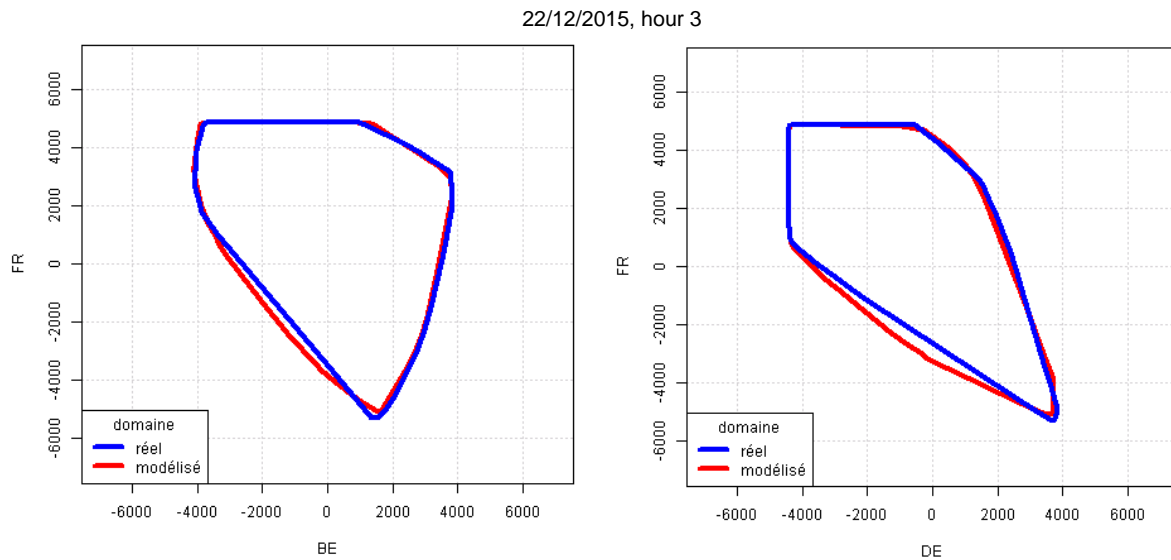
day	hour	inf	sup
05/11/2015	15	5,29 %	5,23 %
02/07/2015	1	3,51 %	3,43 %
22/12/2015	3	5,61 %	5,51 %

Figure 17 presents domains with different shapes, some being for example very sharp in one direction: the model is able to follow a diversity of shapes with minimal errors. Without optimization algorithm, the previous empirical approach was returning volumetric error varying from 35 to 40%.

Figure 17: Examples of real domains and their models, projected on the plane BE-FR (left) and DE-FR (right)







R package: `antaresFlowBased`

A Shiny application is created by the function `runAppError`, it displays the values of inf and sup errors for each domain of the set of typical days (of the chosen model) and diagrams of real and modelled domains.


#### 4.4 Implementation in Antares

This part follows the creation of domains time series (3.4), based on the matching climatic scenario. To be able to model flow-based domains, two additional steps are required.

The binding constraints shaping the domains must be created. The coefficients/weights of the equations remain the same, whatever the hour or the year, so the updating of these weights only takes place once at the beginning of the study. A small manoeuvre is needed, because the calculated coefficients are based on the net positions of the countries, and binding constraints apply on flows between nodes. Net positions can be decomposed into the sum of all their flows (for example,  $DE = -BE.DE + DE.FR + DE.NL$ )

Figure 18: Example of conversion from Net Positions coefficients into flows coefficients

Name	BE	DE	FR
1	-0.78	-0.31	-0.54
2	0.94	0.1	0.3
3	0.82	0.22	0.52
4	0	1	0



Name	BE.FR	DE.FR	DE.NL	BE.NL	BE.DE
FB1	-0.24	0.23	-0.31	-0.78	-0.47
FB2	0.64	-0.2	0.1	0.94	0.84
FB3	0.3	-0.3	0.22	0.82	0.6
FB4	0	1	1	0	-1

The domain time series is used to associate each day to the 24 consecutive matching flow-based domains. Each hourly domain brings its 36 second members. For the simulation of one Monte Carlo year, the matching domain time series is hence turned into 36 second member files.

R package: `antaresFlowBased`

To run a flow-based simulation in Antares, two functions are necessary.

The first one is `initFB` which initialise the study: it creates the binding constraints, links the domains time series to the study and to the other input time series and loads the chosen typical days set.

The function `runsimulationFB` enables then to run the Antares simulation: for each Monte Carlo year, it reads the matching domain time series, writes the second members files (based on the domain time series) and launches Antares. It also aggregates the results at the end of the simulation.

## 5. Adequacy patch

### 5.1 Definition and importance of the adequacy patch

The adequacy patch is a component of the operational flow-based coupling algorithm<sup>3</sup>.

The patch intervenes only if at least one of the 4 countries experiences a shortage. Its function is to ensure that in this case, the solution of the algorithm splits the effort in a "fair" way, often by making it less optimal.

<sup>3</sup> Annex 16\_20 Adequacy Mitigation :

<http://jao.eu/support/resourcecenter/overview?parameters=%7B%22IsCWEFBMC%22%3A%22True%22%7D>

What is the impact of the patch? Antares simulates a solution minimising the entire system costs (among others, generation and Energy Not Supplied - ENS - costs): without adequacy patch, a case like the one presented in Figure 19 could happen.

Figure 19: An output of Antares: the need for an adequacy patch

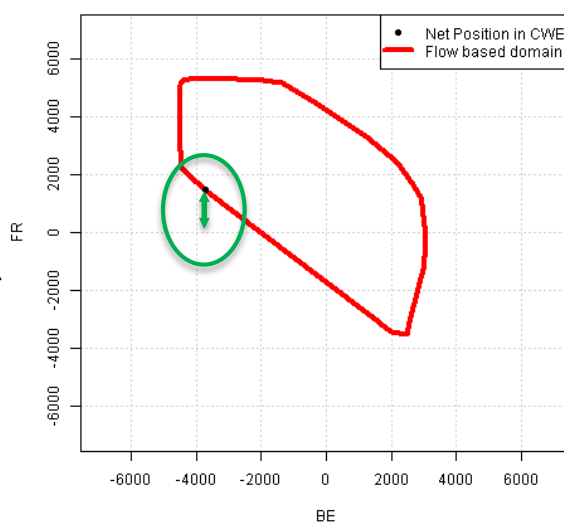
(To notice: the examples depicted here are simulated, there has been no real loss of load situation in the CWE area since the introduction of flow-based)

Example of Antares output for a specific hour

area	ENS (MWh)
be	0
de	0
fr	1297
nl	0

**France in shortage  
And yet is exporting!**

Matching flow-based domain projected on BE-FR



In this example, France is experiencing a loss of load; however, as can be seen in the flow-based domain, France is exporting in the CWE area. Moreover, the energy exported by France is higher than its ENS. This means that the French loss of load is here due only due to its exports.

This is the result of an optimal solution (minimisation of ENS). As flow-based domains are not symmetric for all countries, counterintuitive phenomena can results from the optimisation of the overall system costs. For instance, without adequacy patch:

- Countries with ENS can export<sup>4</sup>
- Countries short in generation can import to avoid ENS, thus increasing or even creating ENS in other countries

**The adequacy patch has been added to the market coupling algorithm in order to bypass these counterintuitive results. It is applied only when at least one country has ENS. It redistributes ENS among the contributing countries.**

<sup>4</sup> In that case, the export of (e.g.) 1 MW from this country relax a critical branch which allow another country to import strictly more than 1 MW.

## 5.2 The adequacy patch in Antares

### 5.2.1 Principle

To integrate this patch to Antares' simulations, the chosen solution is to apply it as a **module ex-post**. When the simulations have ended, all situations where at least one country experiences ENS are selected and an adequacy patch is modelled to correct them.

The principle of this model is to re-compute the Net Positions of CWE countries so that the amount of ENS in all the area is minimized and the following constraints are respected:

- The flow-based limits remain respected
- Each country import is proportional to its contribution to ENS in the CWE area (i.e. ENS without exchanges with CWE countries)

### 5.2.2 Example 1: Creation of ENS

Going back to the example presented in Figure 19, the situation before adequacy patch is:

Figure 20: Antares situation, before adequacy patch

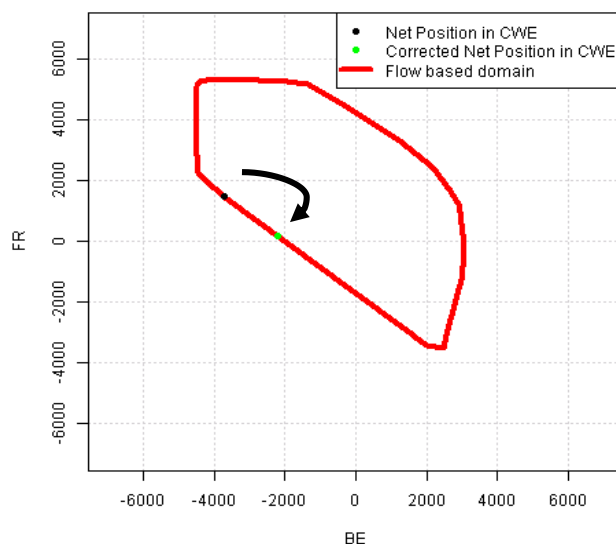
area	Initial Net position CWE (MWh)	ENS (MWh)	ENS without exchanges CWE (MWh)	
be	-3717	0	3717	Lack of generation in Belgium hidden by substantial imports
de	2103	0	0	
fr	1471	1297	0	ENS in France caused by high exports
nl	143	0	0	

The situation in output shows ENS in France: however, within the CWE area, France is exporting a value higher than its ENS. This proves that the ENS in France is caused by exchanges that the adequacy patch would prevent.

Moreover, the only country importing in this case is Belgium and Figure 20 illustrates that without exchanges, this country would be in ENS situation: the imports from France enable to fill this ENS.

When applying the adequacy patch on this case, the Net Position of each country are recomputed: the French exports and consequently the Belgian imports are reduced (Figure 21).

Figure 21: Flow-based domain projected on BE-FR, before and after adequacy patch



The consequences of these modifications of exchanges are displayed in Figure 22.

Figure 22: Antares situation, after adequacy patch

area	Corrected Net Position CWE (MWh)	Corrected ENS (MWh)
be	-2200	1517
de	1883	0
fr	174	0
nl	143	0

ENS:

- Compatible with the effective contribution of each country
- Solution provided by Adequacy Patch more expensive than "optimal solution"

In this final situation, France is no longer in ENS and keeps exporting but the export is less important and does not endanger the country's balance. As a consequence, Belgium imports less and experiences ENS, because there are not enough generation and imports to cover its entire needs. German and Netherlands' positions move a little to adapt to the new flow-based constraints.

This solution is not optimal: Belgian ENS is finally higher than the initial French ENS.

### 5.2.3 Example 2: Redistribution of ENS

The first example represented the attribution of all ENS of CWE to one country. Another case which requires the intervention of the adequacy patch is a wrong distribution among countries in loss of load situations.

Figure 23: Antares situation, before adequacy patch

area	Initial Net position CWE (MWh)	ENS (MWh)	ENS without exchanges CWE (MWh)
be	129	515	386
de	395	0	0
fr	4001	0	0
nl	-4525	14	4539

*Total ENS in CWE = 529 MWh*

ENS in Belgium increased by its exports

Netherlands : single importing country in CWE; its ENS is reduced

In this example (Figure 23), Belgium and the Netherlands are simultaneously experiencing ENS, but the Netherlands are drawing all exports from the area, Belgium is even exporting in addition.

This is another case which need the intervention of the adequacy patch as the initial distribution of ENS among the countries is not fair. The correction is then to rebalance the imports between the two countries, while respecting the physical limits of the flow-based domain.

Figure 24: Flow-based domain projected on BE-NL, before and after adequacy patch

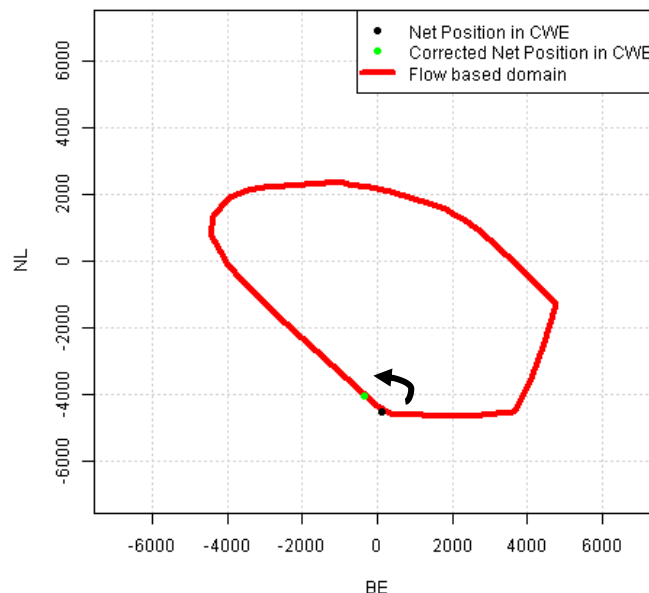


Figure 25: Antares situation, after adequacy patch

area	Corrected Net Position CWE (MWh)	UNSP. ENRG (MWh)
be	-344	42
de	389	0
fr	4001	0
nl	-4046	493

ENS :

- split of ENS pro-rata of situation without exchanges
- More expensive solution

*Total ENS in CWE = 535 MWh*

After the adequacy patch, both countries receive a part of the available exports, the final ENS is proportional to what they originally needed without exchanges.

This solution is not the optimal result, again, the total ENS over the area is higher after adequacy patch.

R package: `antaresFlowBased`

To apply the adequacy patch on a simulation, the function `adqPatch` is used at the end of an Antares simulation (calculation ex-post). It takes in input the last simulation results. It is possible to select the Monte Carlo years which are of interest for the user and a filter also enables to reduce computation time by loading the data only if ENS has happened in the year.

### 5.3 The case of existing strategic reserves

The calculation of ENS for each country takes into account all remaining production margin (if it exists). However, the modelling of strategic reserves has sometimes for consequence that, after the application of the adequacy patch, some countries have ENS while still having some available strategic reserves.

Strategic reserves are production units which can be activated by the grid operator during exceptional situations, and only for the use of the country.

**If a country possesses strategic reserves, a supplementary step checks whether there is or not available reserves if the country has ENS after the application of the adequacy patch.** If the reserve is not empty, the ENS is recomputed with it.

In the end, if a country has ENS after adequacy patch, it implies that there is no strategic reserves, or that they have been emptied.

R package: `antaresFlowBased`

The same function `adqPatch` enables to take strategic reserves into account: two parameters (`strategic_reserve_be` and `strategic_reserve_de`) lists the names of the possible strategic reserves in CWE.

## 6. Current users and future prospects

### 6.1 Users

The Antares modelling of flow-based domains is used by RTE for:

- The **French adequacy report of 2017**, based on the typical days presented in 2.3 (clustering with weighted hours), to simulate medium term (up to 5 years) generation adequacy.
- The **French Winter Adequacy Outlook 2017-2018**, which are based on only 4 typical days (matching the studied season)

It is also used by the Belgian grid operator, Elia, in the **Belgian 2017's adequacy study** which assesses the need for strategic reserve in winter 2018-19.

Finally, on the European scope, this is also used for the adequacy study of the **Pentalateral Energy Forum (PLEF)**, by the participating TSOs which use Antares. The other TSOs are however also basing their simulations on the same typical days.

### 6.2 Additional specification to typical days

For the PLEF study, another factor has been added to the model, a 37<sup>th</sup> binding constraint enabling to link the maximal capacity of the exchange between Germany and Switzerland to the typical day used. It is then possible to adapt other borders NTC to the chosen typical day, considering that they could be explained by a climatic situation. Although the creation of additional binding constraints is not handle by the R package, such an adaptation is currently possible and easily achievable.

### 6.3 Adaptability to the input flow-based domain

The main objective of the modelling - based on reducing the number of sides of the domains and standardizing them, is to shorten as much as possible the simulation time on Antares by ensuring the number of binding constraints is limited and no second member is set to infinity. Thanks to this method, **any domain can be represented with reasonable accuracy in Antares**, only by changing the second members. Second members being represented by time series, the simulation time would not be increased.

Using typical days is a way to extrapolate defined historical situations to a diversity of future horizons, but the number of domains is not restrictive: any flow-based domain can actually be modelled and used in the simulation. It is then possible to replay the past by modelling entire sections of the history.



Flow-based simulations are currently only used for short or medium term studies: because the typical days are based on the history, they can hardly take into account evolutions of the grid topology and generation location, especially over ten years' time. However, should long term domains be built, they can replace the history, be clustered to extract typical days and be projected thanks to the same method to be used on Antares long term simulations.

#### **6.4 Comparison between two domains**

Flow-based simulations can be used to price the difference between two flow-based domains: it is possible to compare their impact by running two Antares studies and comparing the results (costs, volume of exchanged energy etc.).

This could be used to compare the benefits of grid infrastructure evolutions for example, whereas it is very hard to compare two flow-based domains by only looking at their volumetric shape.

#### **6.5 Additional countries**

If new countries were to be added to the flow-based area, the shapes of the domains will evolve, increasing their number of dimensions. Fundamentally, the method will remain the same, the number of "sides" for example can be raised by only changing a parameter. However the R packages would need to be adapted for more dimensions in the geometrical shapes, especially for the algorithm distances calculation and the final weights on the binding constraints.

## 7. Annexes

### 7.1.1 Annex 1: Description of the modelling optimisation problem

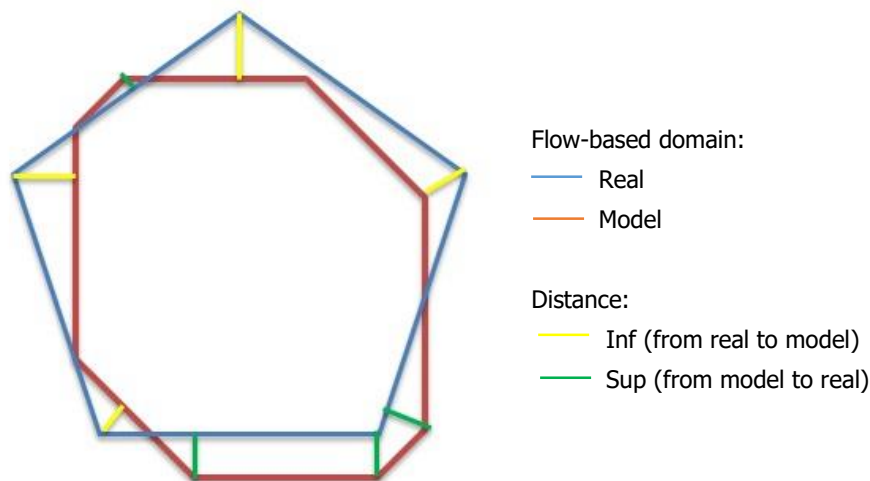
#### 7.1.1.1 Presentation

The algorithm aims to optimise the setting of the limits of the domain, by choosing them together as a set. Its principle is then to calculate the distance between the model and the real domain, and then adjust the limits of the domains to reduce this distance. The same problem faced for the clustering of the history (to choose the typical days, report to RTE's note *Typical Flow-based Days Selection*), is met: to calibrate the model, one needs a way to calculate the distance between two polyhedron.

Two distances are calculated, the distance from the real domain to the model, which is similar to the inf error (although the error is volumetric and it is a Cartesian distance), and the distance from the model to the real domain, the equivalent of the sup error. To do so, the vertices of each domain are identified and projected on the other domain's surface: the distances between each vertex and its projected point are summed up to get the distance from the first domain to the other one.

The chosen method follows then the same principle detailed for the typical days selection (RTE's note), except that the comparison is being made between a real domain and its model, and not two real domains. The main obstacle implied is that the model is not yet a defined shape, its sides are being moved, meaning that the vertices are also moving, always different. They must themselves be considered as variables, defined by an annex function.

Figure 26 : Distances between a real domain and its model



The algorithm calculates the two distances and position the limits of the model in order to minimize them. However, like the volumetric errors, the two distances are inversely linked, so a weighting must be applied to indicate which distance to favor. The results returned by this algorithm will then be the Pareto frontier for these two distances, each point being the optimal solution for a precise combination between inf and sup.

### 7.1.1.2 Algorithms

#### 7.1.1.2.1 Inputs

Two inputs are needed :

- The PTDF matrix describing the real domains, which is, by the process explained in RTE's note *Typical Flow-based Days Selection*, transformed into a matrix of vertices coordinates of the real domain
- A matrix B of vectors describing the chosen directions for the model (the sides of the model). The matrix has 36 lines, for the 36 directions extracted from the clustering algorithm presented in 4.2.3.1

Table 5: Example of coordinates of vertices

BE	DE	FR
-1096.21	4821.499	-3809
-1258.306	4322.37	-3336.598
-1301.912	6385.21	-3809
-519.2495	3338.604	-3809
...	...	...

BE	DE	FR
1	0	0
0	1	0
0	0	1
1	1	1
-1	1	1
1	-1	1
-1	-1	1

Table 6 : Example of matrix B (non clustered directions)

These data enable to identify equations defining the real domain X and the model Y:

$$X = \left\{ x = \sum_s \lambda_s x_s, \text{ avec } \sum_s \lambda_s = 1 \right\}$$

Where s is the set of vertices of X and  $x_s$  is a vertex of X.

$$Y = \{y \text{ so that } B \cdot y \leq b\}$$

Where B is the matrix of the projection axis and b the vector of the domain's limits on each axis.

b is the problem's variable.

#### 7.1.1.2.2 Algorithm 1: Coordinates of the model's vertices

Calculating the modelled domain vertices' coordinates is one of the key issues of the algorithm: the Y's vertices are at the same time inputs to calculate the distance from Y to X and outputs to the variations of b, the limits of the domain.

Alongside the optimization algorithm, a secondary algorithm is designed to express Y's vertices, based on the matrix B's coefficients and the second member b. To identify the vertices, each side of Y, meaning each line of the matrix B, is attributed an id. Let  $y_c$  be the intersection point between three plans (i, j, k) of Y.

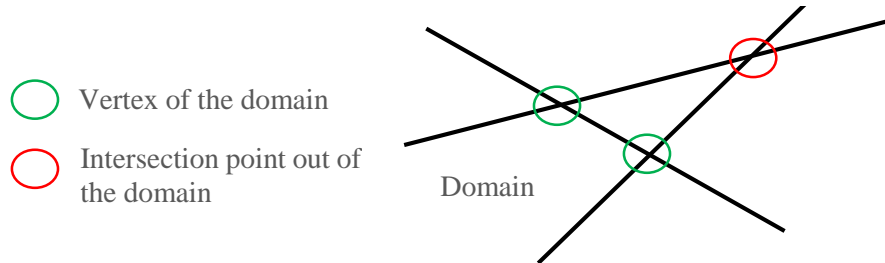
$$\begin{cases} B_i \cdot y_c = b_i \\ B_j \cdot y_c = b_j \\ B_k \cdot y_c = b_k \end{cases}$$

$y_c$  is defined by the triplet :

To ensure  $y_c$  belongs to the domain, it must also verify:

$$B \cdot y_c \leq b \quad \text{for all vectors of } B$$

Figure 27: Two kinds of intersection points in two dimensions



The output of this algorithm is a file containing all the vertices' coordinates, but under a specific format: as they are always evolving, the chosen way to identify them is  $(B_i \ B_j \ B_k)$ , meaning the coordinates of the three sides they belong to.

Figure 28: Example of Y's vertices coordinates

	Side i			Side j			Side k		
Vertices of Y	1	0	0	1	1	1	1	0	1
	1	0	0	1	1	1	1	1	0
	1	0	0	1	-1	1	1	0	1

### 7.1.1.2.3 Algorithm 2 : main optimisation problem

The optimization problem is divided in several parts (optimized at the same time).

One of them calculates the "inf" distance, from the real domain to the modelled one.

Let  $x_s$  be a vertex (three coordinates BE, DE, FR) of X. Each vertex  $x_s$  is projected and its projection  $y_s$  is calculated by minimizing the distance between the two points. An error, divided in two positive terms is defined via:

$$y_s = x_s + y_s^+ - y_s^-$$

Knowing that the projection  $y_s$  belongs to the model domain, the optimization must verify:

$$B \cdot y_s \leq b$$

$$B \cdot (x_s + y_s^+ - y_s^-) \leq b$$

The second part calculates similarly the "sup" error, from the model Y to the real domain X.

Let  $y_e$  be a vertex of Y, defined in output of the first algorithm under the format  $(B_i \ B_j \ B_k)$ . The three actual coordinates of the vertex are calculated by redefining  $y_e$  as the intersection point of the three planes and solving the system:

$$\begin{cases} B_i \cdot y_e = b_i \\ B_j \cdot y_e = b_j \\ B_k \cdot y_e = b_k \end{cases}$$

A projection of  $y_e$  is then calculated:  $x_e$  is obtained by minimizing the distance between the two points, two positive terms are defined in the error:

$$y_e = x_e + x_e^+ - x_e^-$$

$x_e$  belongs to the real domain, the algorithm must verify, with  $x_s$  being a vertex of X:

Under these constraints, the algorithm solves:

$$\min \left[ \frac{1-\alpha}{\text{Card}(X^s)} \cdot \sum_s |y_s^+ + y_s^-|^2 + \frac{\alpha}{\text{Card}(Y^e)} \cdot \sum_e |x_e^+ + x_e^-|^2 \right]$$

$$x_e = \sum_s \lambda_{es} x_s, \quad \text{with} \quad \sum_s \lambda_{es} = 1$$

$$y_e = \sum_s \lambda_{es} x_s + x_e^+ - x_e^-$$

With:

- $\alpha$  being the weighting coefficient enabling to favor one error despite the other
- $X^s, Y^e$  are the sets of vertices of X and Y. The error is divided by the number of vertices to make sure one domain is not too heavily considered

$\alpha$  is set depending on the targeted « inf » and « sup » errors. This method enables to reach very low levels of errors (depending on the domains), so  $\alpha$  is chosen to ensure similar levels for both errors. To achieve it, the algorithm is first run with  $\alpha$  at 0.5 and the volumetric errors are calculated on the obtained model. The final  $\alpha$  is reached by dichotomy by repeating these steps until the volumetric errors are similar.

#### 7.1.1.2.4 Fixed limits

To ensure the characteristics needed for adequacy studies are represented, 4 of the 36 constraints are manually fixed before the optimization algorithm, which do not change them afterwards.

These are the ones representing the maximal import for France, Belgium and Germany and the maximum export for Germany as well. They are set at the maximum value reached by the real domain.