Off-line Data Validation for Water Network Modeling Studies

M. Quiñones* G, C. Verde and L. Torres****

***Universidad Tecnológica de La Habana J.A. Echeverría**

****II-Universidad Nacional Autónoma de México**

Nov 2019

Content

1. **Motivation & Objective**
2. **Case Study: *El Charro, Guanajuato, Mexico***
3. **Off-line Semi Automatic Data Validation Scheme**
4. **Density-Based Spatial Clustering, DBSCAN**
5. **Results & Conclusions**

1.- Motivation & Objective

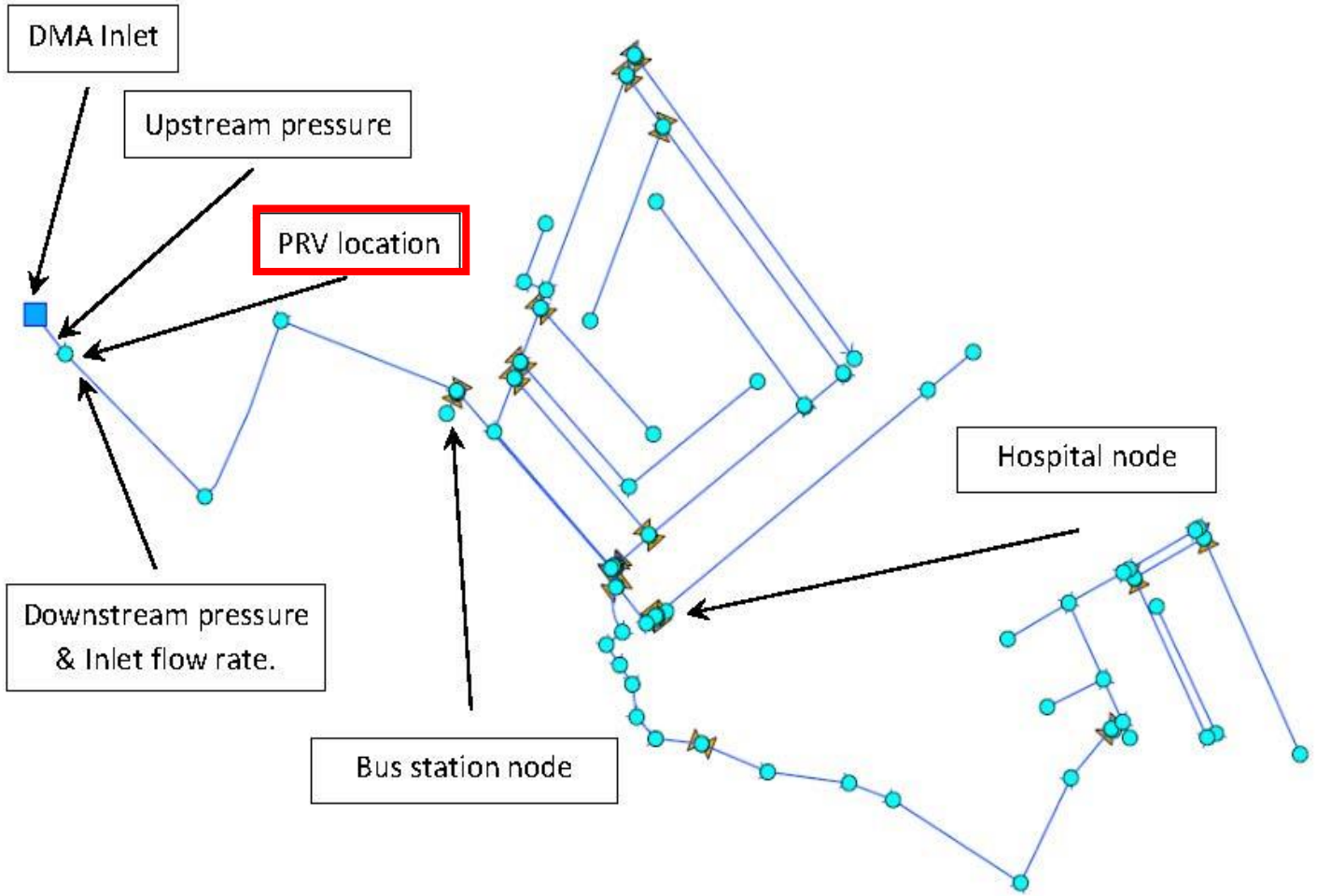
Motivation

- Water network (WN) operating studies are significantly affected by the real data quality.
- If raw data are not validated before they are used, the resulting studies and models could not be representative of the real behavior of the WN.

Objective

- The application of an off-line semi- automatic classifier that separates data of nominal & abnormal events into WNs.
- A simplified procedure to validate raw data of WNs by using machine learning techniques.

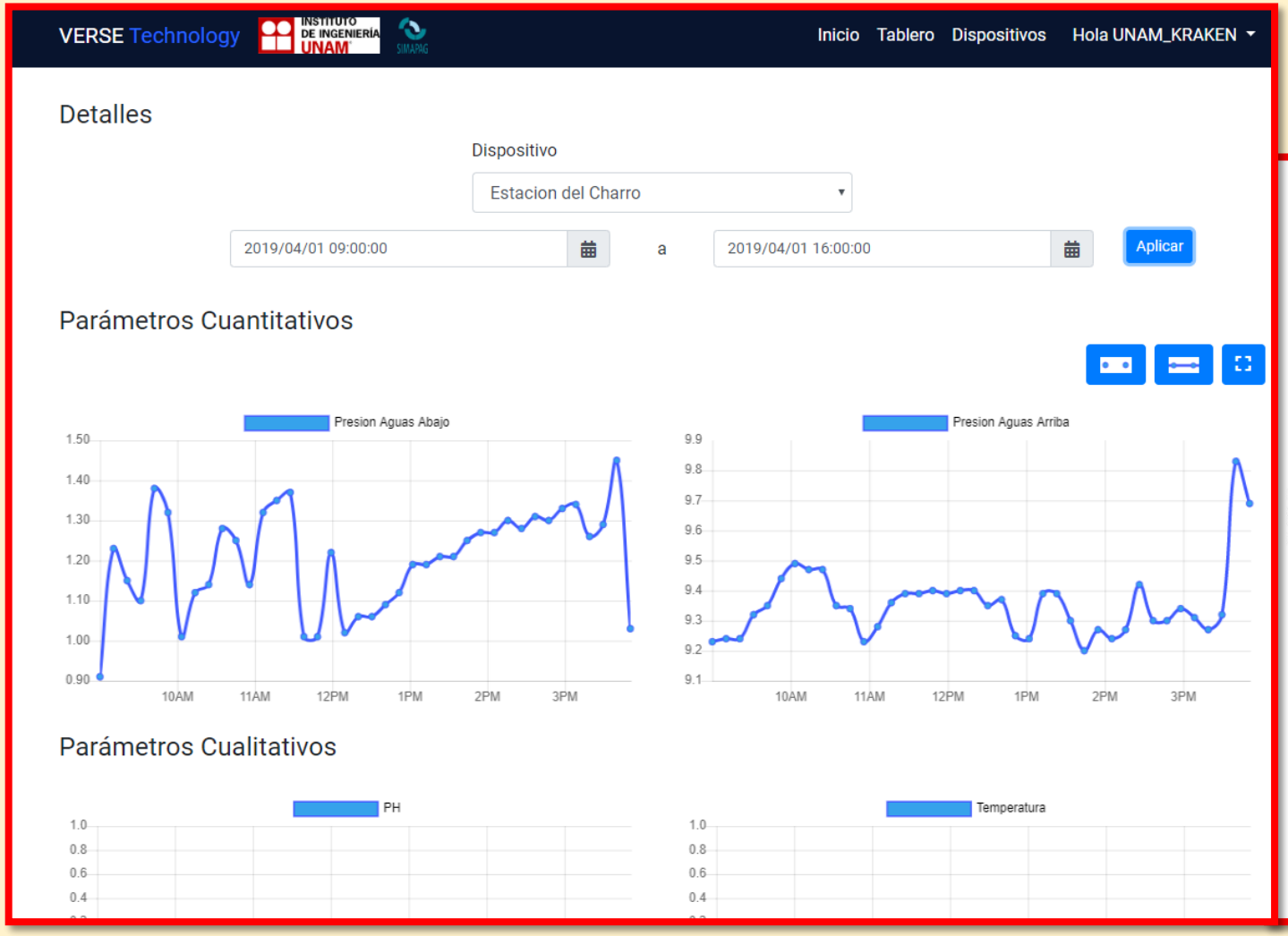
2.- Case study: DMA El Charro

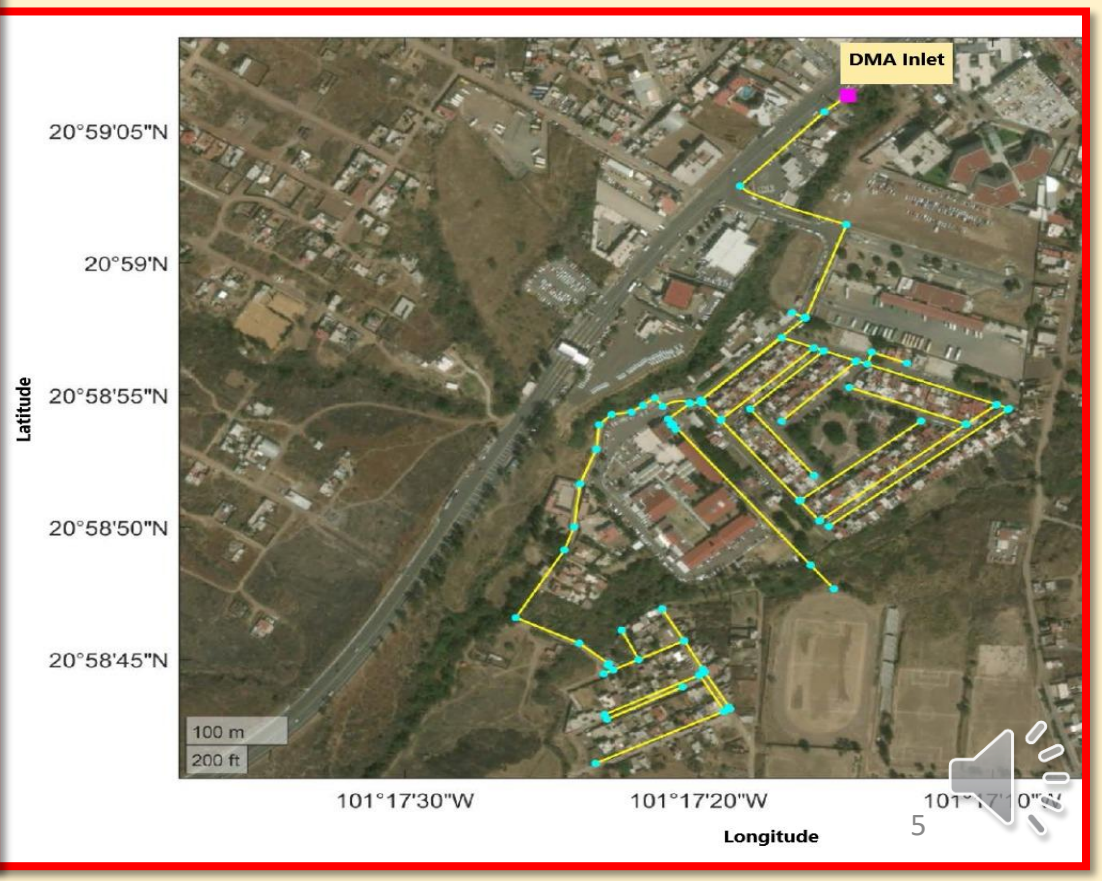


Characteristics	Quantity
Pipelines:	75
Nodes	90
Supply reservoir capacity:	1000 (m^3)
Customers:	2000
Average consumption per year:	3 (lps)
Installed Sensors:	1 upstream pressure transducer (kg/cm^2) 1 downstream pressure transducer (kg/cm^2) 1 inlet flowmeter (lps)
Valves:	1 pressure reducing valve [PRV]

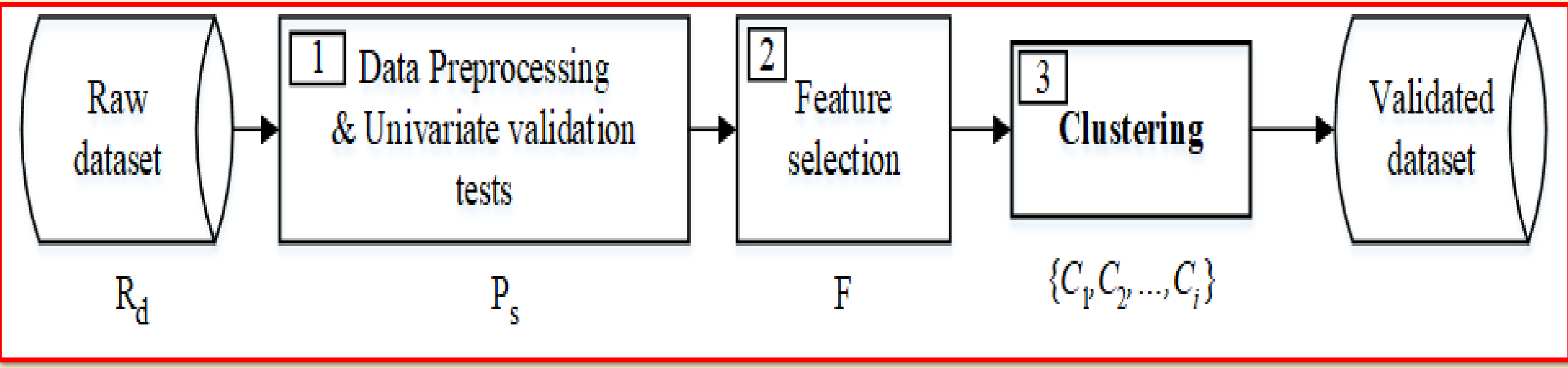
2.- Case study: DMA El Charro

- **Web platform & monitoring station**



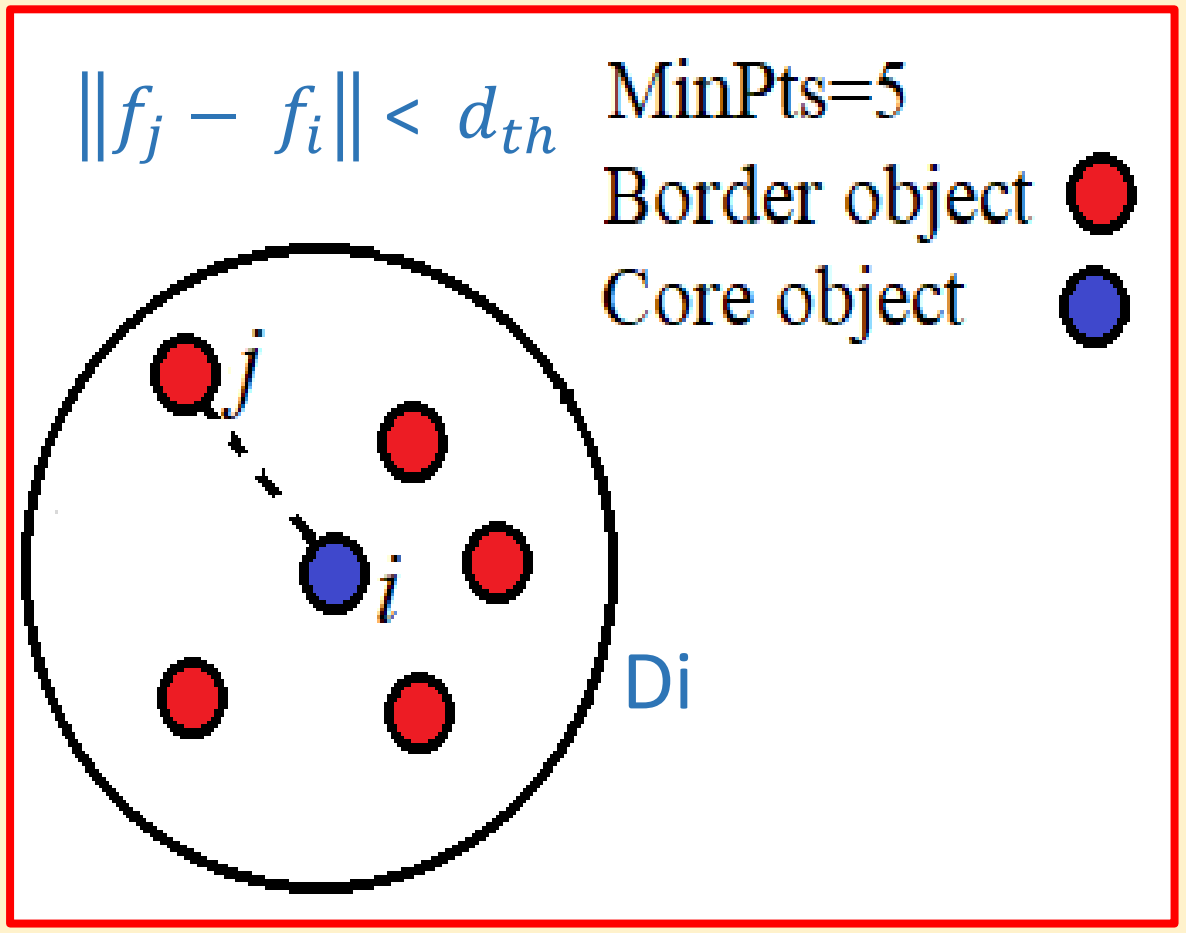


3.- Off-line Semi Automatic Validation Scheme

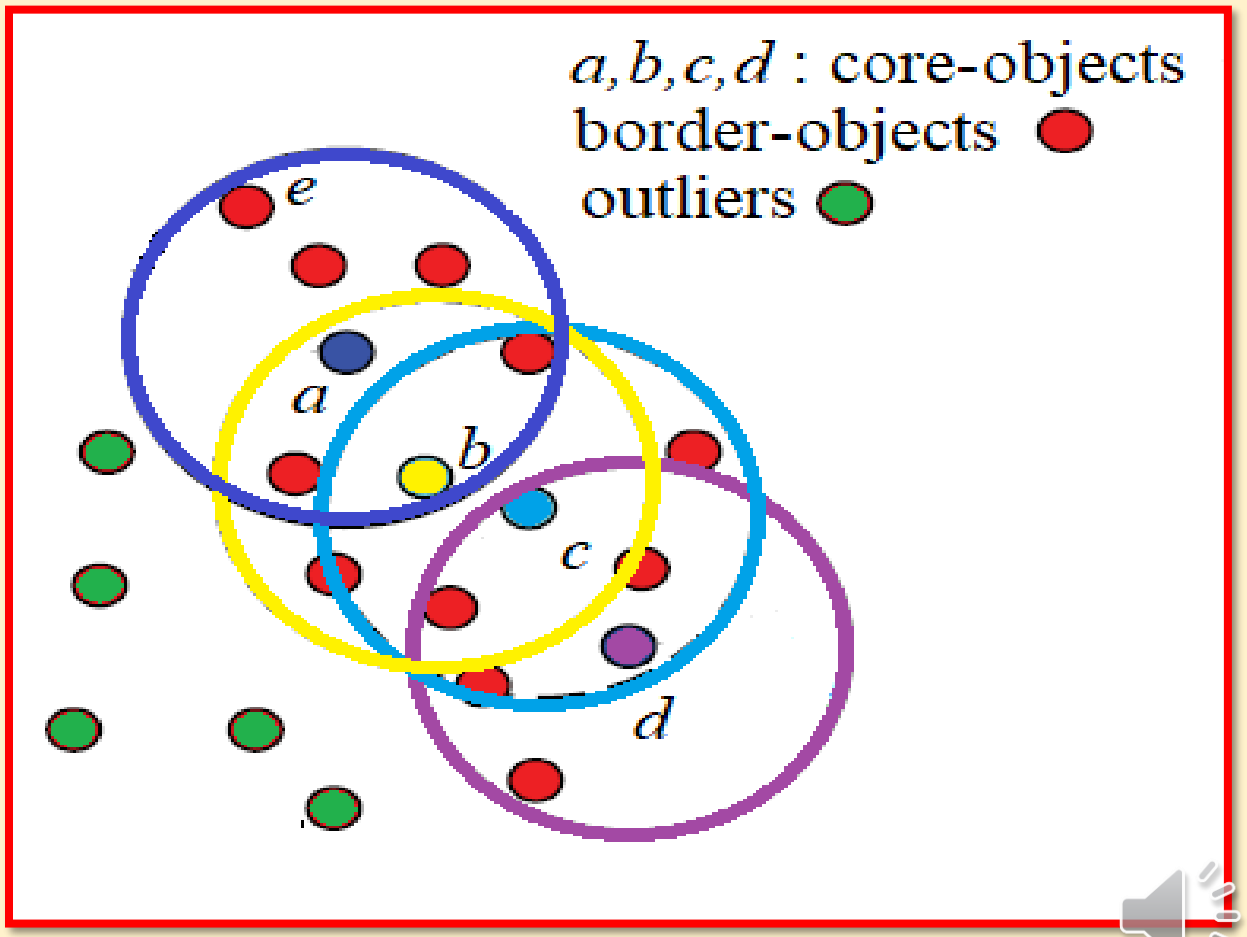


4.-Density-Based Spatial Clustering (DBSCAN)

- Object with its Neighborhood



- Density-based Cluster and outliers



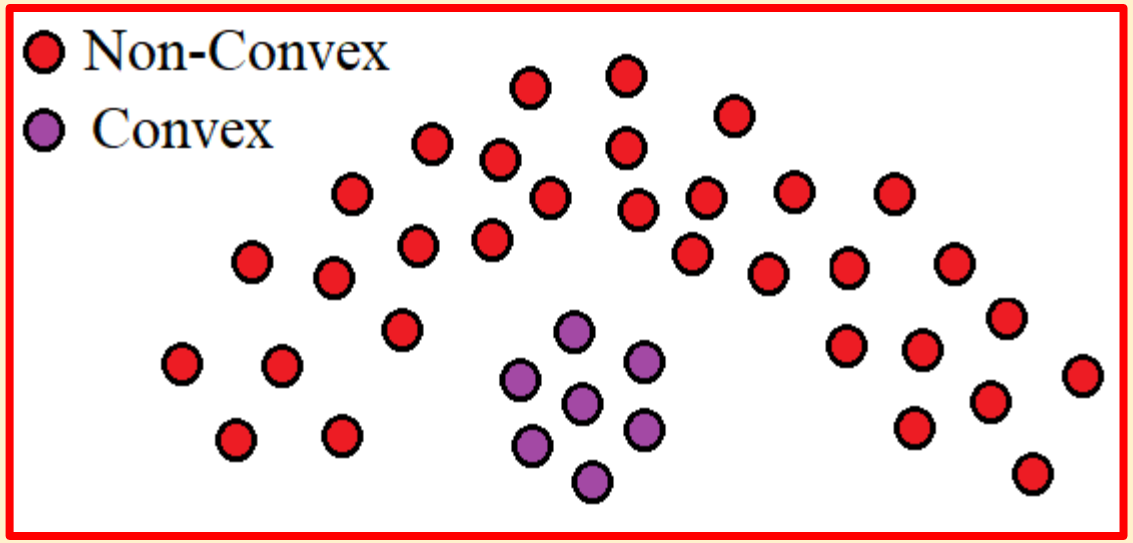
4. DBSCAN: Algorithm and properties

Algorithm

Data: F, $MinPts$, d_{th}, C: set of clusters, No: set of noise objects, i: number of clusters
Label all objects as not classified, $C = \emptyset, No = \emptyset, i = 0$;
for $\mathbf{f}_j \in F$ **do**
  if $\mathbf{f}_j$ *is not classified* **then**
    $DR_j = DensReach(\mathbf{f}_j)$
    if $|DR_j| > 1$ **then**
      Form a new cluster with all density-reachable objects
      Label cluster' objects as classified
      $C_i = DR_j, C = \{C, C_i\}, i = i + 1$
    if $\mathbf{f}_j$ *is not a border-object* **then**
      $No = No \cup \mathbf{f}_j$
  Label $\mathbf{f}_j$ as classified
end

Properties

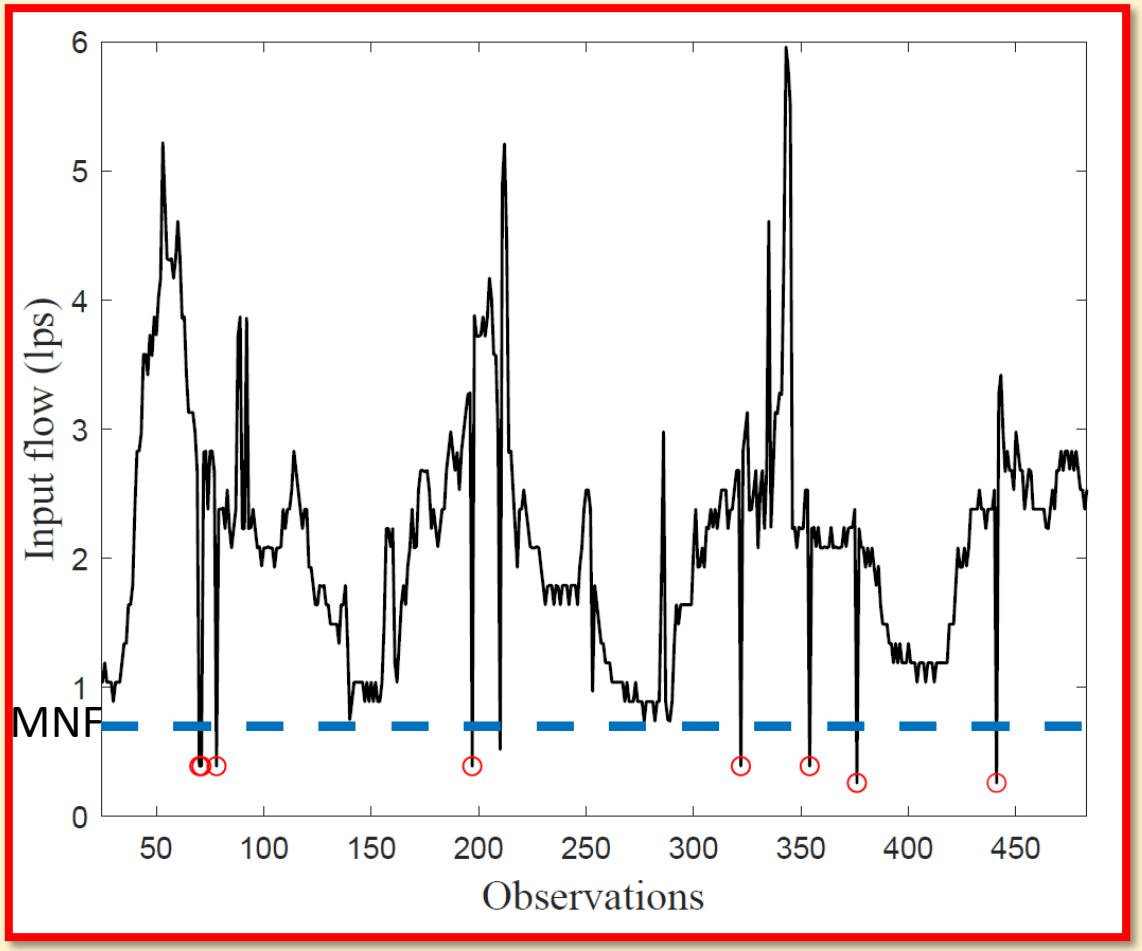
- Clustering of objects with non-convex shapes



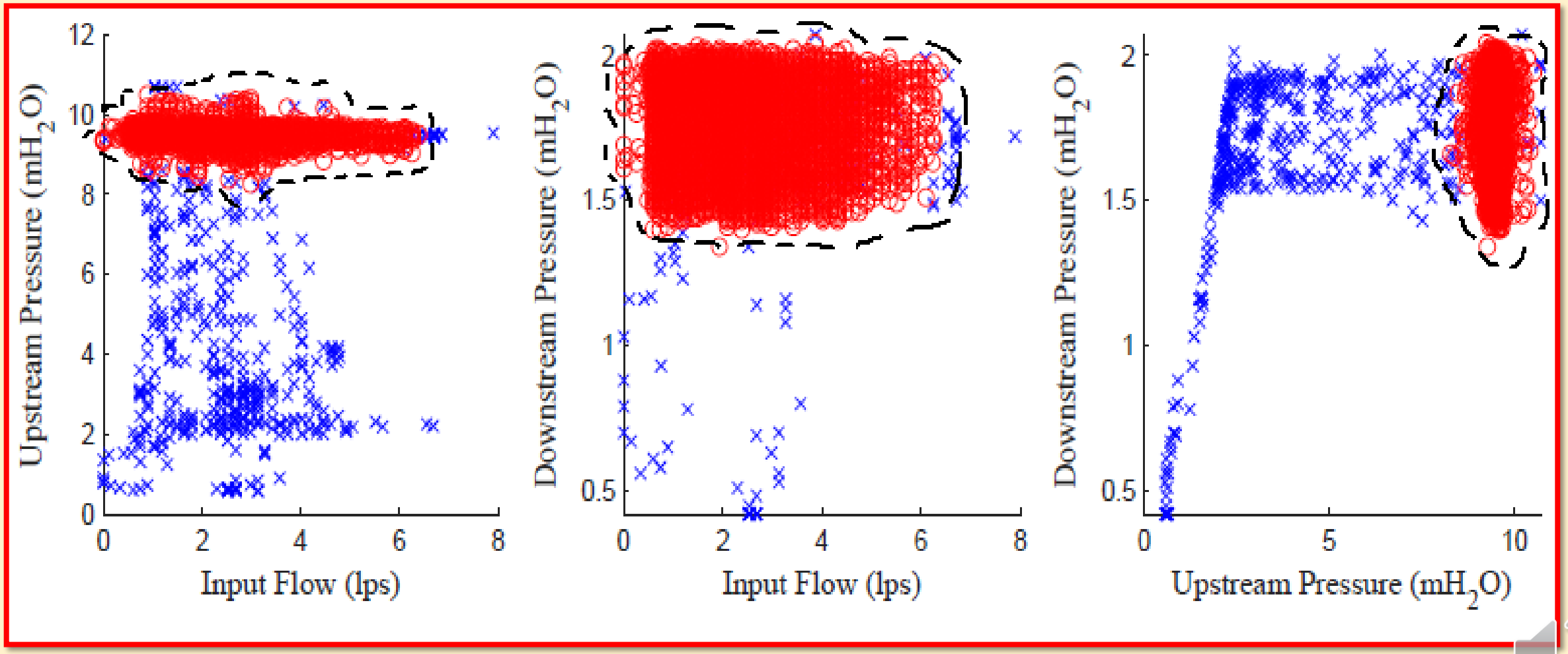
- Isolation of outliers from clusters

5.- Results & Discussion

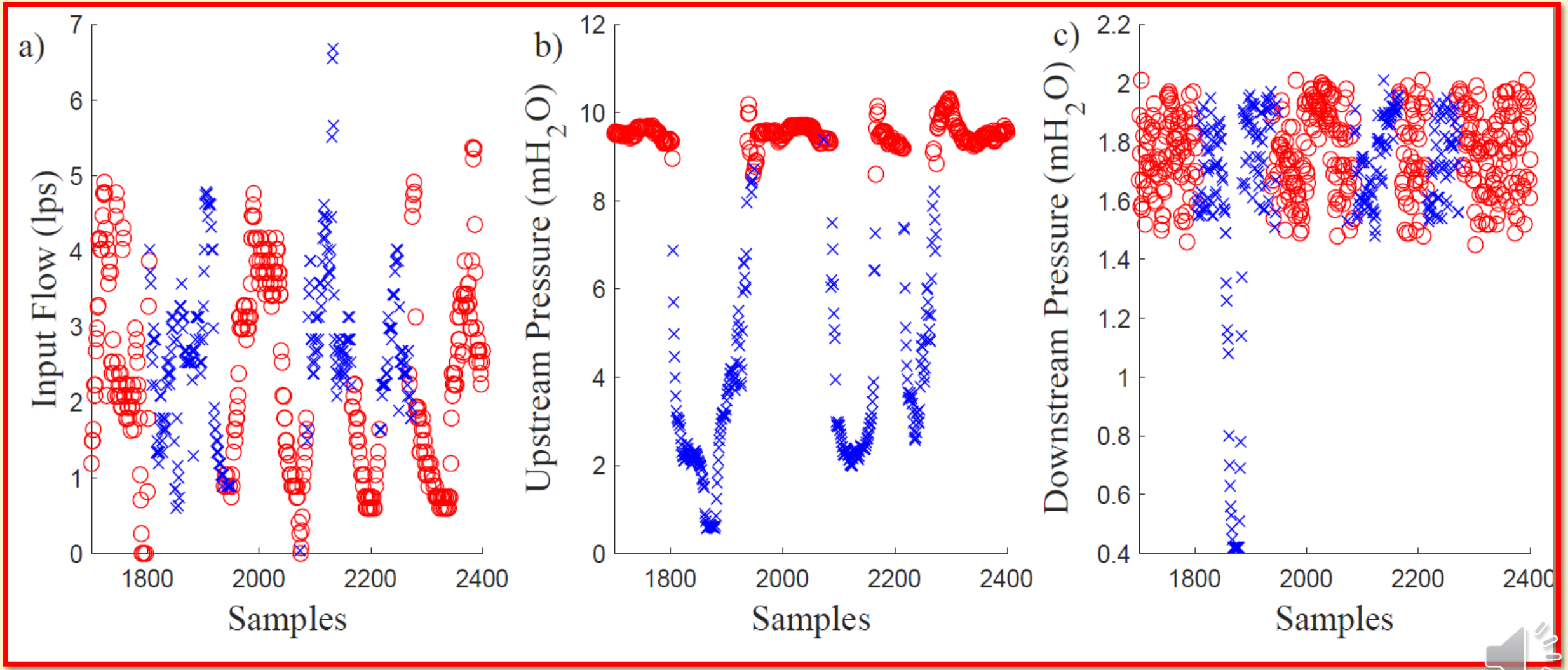
- **Preprocessing Tasks**



5. Clusters of Normal & Abnormal Data



5.- Draining of the reservoir



5.- Conclusions

- An off-line approach to data validation in WN is introduced.
- The core of the proposal is the application of an unsupervised clustering method without feature definition for the diverse data sets to be identified.
- The application of the cluster algorithm to the DMA El Charro allowed the identification of a systematic anomaly: the reservoir draining.
- Given the results, the network operators concluded the convenience of the pressure reducing valve for the DMA.

Thanks to you for the attention!
&
we are open to questions by email

- Marcos Quiñones-Grueiro: marcosqg88@gmail.com
- Cristina Verde: verde@unam.mx
- Lizeth Torres: ftorreso@ii.unam.mx