Water distribution networks leaks estimation in faulty sensor context: a graph approach

K. Srinivasarengan, S. Aberkane, V. Laurain

25 Nov 2021

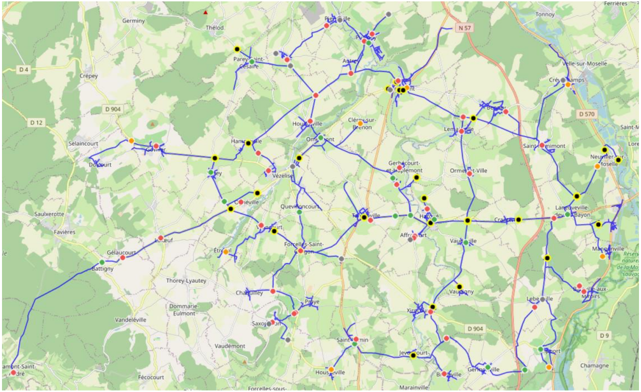
vincent.laurain@univ-lorraine.fr

Context

Status of Water Distribution Network (WDN)

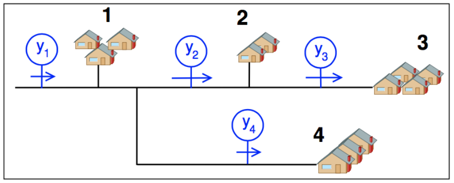
- Currently, at best 20%, of drinking water is lost
- How to best avoid leaks ? → **Detecting them and prioritise interventions**
- SPHEREAU FUI Project : Towards a more efficient water management

- Rural site in France
- 82 flowmeters
- Water direction is known
- 4 years data each 15 minutes



A case study

Just focus on one part



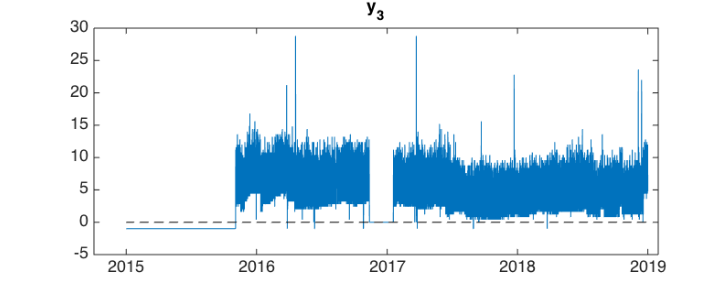
- **1,2,3,4** are District Metered Areas (DMA)
- y_1, y_2, y_3, y_4 are flowmeters

Is it possible to really estimate leaks?

Let us have a look at y_3

A case study

y_3 represents the consumption over a whole DMA



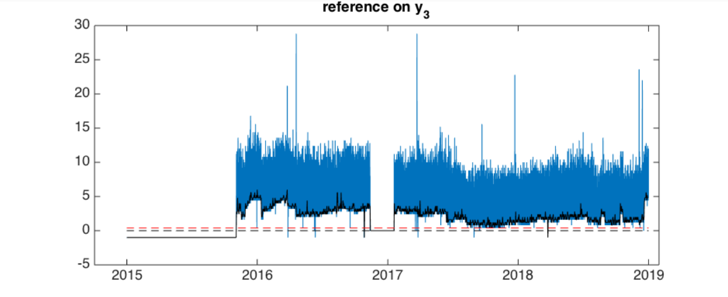
- -1 indicates a missing data
- 0 are real 0 (the sensor is dead)
- **Regular water consumption never reaches 0**

Is it possible to really estimate leaks?

Inseparability of night consumption and leaks $\rightarrow$ it is only possible to estimate leaks with respect to a **certain reference**

A case study

Possible references : Minimum night flow



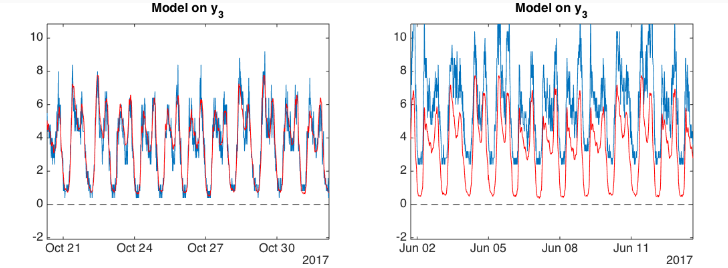
- Night flow has less variablility
- The reference is the minimum night flow
- **Leaks are the difference between measurements and reference**

drawback

What if leaks appear during the day?

A case study

More Robust approach : Modelling



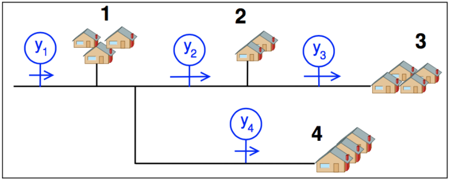
- **Contribution 1** : Kernel modelling of water demand
- Average over the day can be taken
- The reference is the difference between measurements and the model

In any case

The data must contain an unleaky time region
Leaks can be detected up to some reference

Water management traditional approach : the water balance

What about DMA 1?

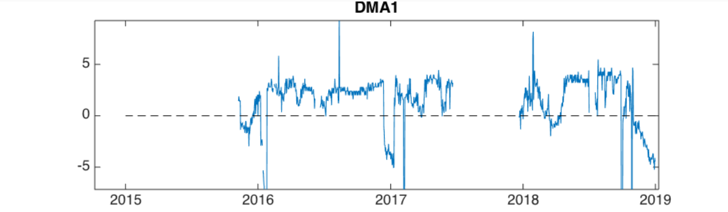


- Total Consumption : $\mathcal{T}_{\mathcal{C}_1}(k) = y_1(k) - y_2(k) - y_4(k)$
- Leaks : $\mathcal{L}_1(k) = y_1(k) - y_2(k) - y_4(k) - \tilde{r}_1(k)$
- Equivalent Leaks : $\mathcal{L}_1(k) = \varepsilon_1(k) - \varepsilon_2(k) - \varepsilon_4(k)$

$$\begin{cases} \mathcal{T}_{\mathcal{C}1}(k) = y_1(k) - y_2(k) - y_4(k) \\ \mathcal{T}_{\mathcal{C}2}(k) = y_2(k) - y_3(k) \\ \mathcal{T}_{\mathcal{C}3}(k) = y_3(k) \\ \mathcal{T}_{\mathcal{C}4}(k) = y_4(k) \end{cases} \qquad \begin{cases} \mathcal{L}_1(k) = \varepsilon_1(k) - \varepsilon_2(k) - \varepsilon_4(k) \\ \mathcal{L}_2(k) = \varepsilon_2(k) - \varepsilon_3(k) \\ \mathcal{L}_3(k) = \varepsilon_3(k) \\ \mathcal{L}_4(k) = \varepsilon_4(k) \end{cases}$$

When problems arise

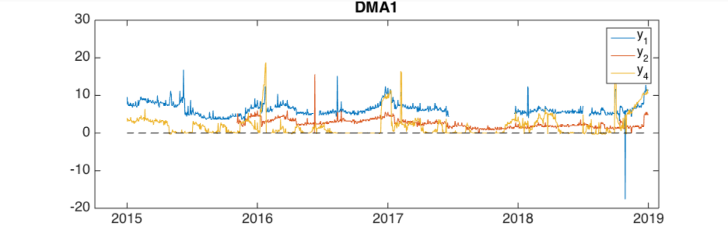
What about DMA 1?



- how to handle missing data?
- Sometimes Consumption becomes strongly negative?

When problems arise

Details on the sensors



- Strong correlation between y_1 and y_2
- Low correlation between y_4 and y_1
- Sometimes y_4 exceeds y_1
- $\rightarrow$ Sensors can be faulty

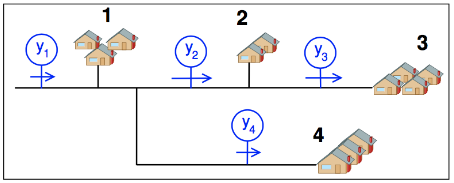
Problematic

Is it possible and how to identify leaks in a sensor faulty context?

Outline

Assumptions

Sensors may contain additive default



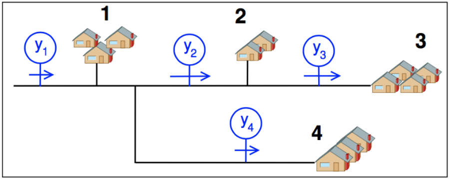
- $y_i(k) = y_i^o(k) + \mathcal{D}_i(k)$
- $\varepsilon_i(k) = y_i^o(k) + \mathcal{D}_i(k) - r_i(k)$

$$\begin{cases} \mathcal{L}_1(k)+\mathcal{D}_1(k)-\mathcal{D}_2(k)-\mathcal{D}_4(k)=\varepsilon_1(k)-\varepsilon_2(k)-\varepsilon_4(k) \\ \mathcal{L}_2(k)+\mathcal{D}_2(k)-\mathcal{D}_3(k)=\varepsilon_2(k)-\varepsilon_3(k) \\ \mathcal{L}_3(k)+\mathcal{D}_3(k)=\varepsilon_3(k) \\ \mathcal{L}_4(k)+\mathcal{D}_4(k)=\varepsilon_4(k) \end{cases}$$

Problematic

How to identify $\mathcal{L}_i(k)$ and $\mathcal{D}_i(k)$?

Assumptions



- Potentially, each DMA contains leaks
- Potentially, each sensor is faulty

$$\underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 & 1 & -1 & 0 & -1 \\ 0 & 1 & 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} \mathcal{L}_1(k) \\ \mathcal{L}_2(k) \\ \mathcal{L}_3(k) \\ \mathcal{L}_4(k) \\ \mathcal{D}_1(k) \\ \mathcal{D}_2(k) \\ \mathcal{D}_3(k) \\ \mathcal{D}_4(k) \end{bmatrix}}_{\mathbf{X}} = \underbrace{\begin{bmatrix} 1 & -1 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{\mathbf{B}} \underbrace{\begin{bmatrix} \varepsilon_1(k) \\ \varepsilon_2(k) \\ \varepsilon_3(k) \\ \varepsilon_4(k) \end{bmatrix}}_{\mathbf{Y}}$$

General Problematic

$$\underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 & 1 & -1 & 0 & -1 \\ 0 & 1 & 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} \mathcal{L}_1(k) \\ \mathcal{L}_2(k) \\ \mathcal{L}_3(k) \\ \mathcal{L}_4(k) \\ \mathcal{D}_1(k) \\ \mathcal{D}_2(k) \\ \mathcal{D}_3(k) \\ \mathcal{D}_4(k) \end{bmatrix}}_{\mathbf{X}} = \underbrace{\begin{bmatrix} 1 & -1 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{\mathbf{B}} \underbrace{\begin{bmatrix} \varepsilon_1(k) \\ \varepsilon_2(k) \\ \varepsilon_3(k) \\ \varepsilon_4(k) \end{bmatrix}}_{\mathbf{Y}}$$

Problematic 1

Is it possible to uniquely estimate $\mathbf{X}$? $\rightarrow$

- Number of equations is the number of DMA
- Number of unknowns $n_{DMA}+n_y$
- Underdetermined problem

Can we propose a regularized identification scheme?

Preliminary Problematic

$$\underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 & 1 & -1 & 0 & -1 \\ 0 & 1 & 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} 0 \\ 0 \\ \mathcal{L}_3(k) \\ \mathcal{L}_4(k) \\ 0 \\ 0 \\ \mathcal{D}_3(k) \\ 0 \end{bmatrix}}_{\mathbf{X}} = \underbrace{\begin{bmatrix} 1 & -1 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{\mathbf{B}} \underbrace{\begin{bmatrix} \varepsilon_1(k) \\ \varepsilon_2(k) \\ \varepsilon_3(k) \\ \varepsilon_4(k) \end{bmatrix}}_{\mathbf{Y}}$$

Preliminary Problematic

What about overdetermined cases

- We know which defaults are present
- Their value is still unknown

Preliminary Problematic

$$\underbrace{\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}}_{\mathbf{A_s}} \underbrace{\begin{bmatrix} \mathcal{L}_3(k) \\ \mathcal{L}_4(k) \\ \mathcal{D}_3(k) \end{bmatrix}}_{\mathbf{X_s}} = \underbrace{\begin{bmatrix} 1 & -1 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{\mathbf{B}} \underbrace{\begin{bmatrix} \varepsilon_1(k) \\ \varepsilon_2(k) \\ \varepsilon_3(k) \\ \varepsilon_4(k) \end{bmatrix}}_{\mathbf{Y}}$$

Preliminary Problematic

Is it possible to uniquely estimate $\mathbf{X}_s$? $\rightarrow$

- Answer is straight forward : Yes iff $rank(A_s) = n_X$
- But...Very uninteresting in practice
- **Water managers are not mathematicians!**

Can we give a comprehensive topological answer to this identifiability problem?
$\rightarrow$ **Use of graphs**

Outline

Strategy

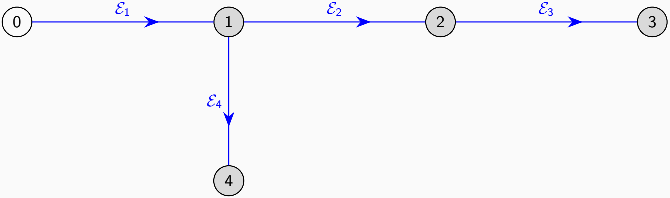
$$A = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & -1 & 0 & -1 \\ 0 & 1 & 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Train of thoughts

- Matrices A and B look like indidence matrices of graphs Let $\mathbf{I}_G$ be the incidence matrix of the directed graph G, then:
  - $\mathbf{I}_G(i,j) = 1$ if edge j has its arrow leaving node i
  - $\mathbf{I}_G(i,j) = -1$ if edge j has its arrow entering node i
  - $\mathbf{I}_G(i,j) = 0$ if edge j is not incident on node i,
- Note that by construction, matrices A and B embed directly the topology of the network.
- **What are the corresponding graphs?**

Graph/Matrices analogy

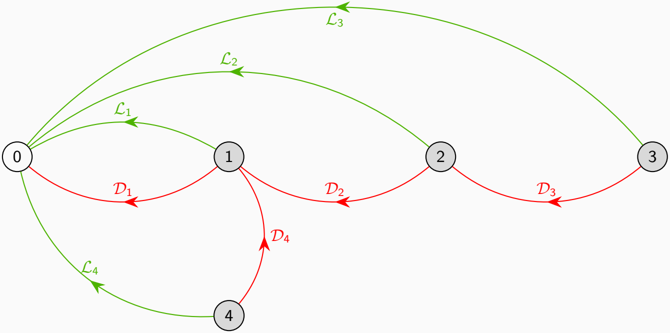
Residual Graph $G_{\mathcal{R}}$



$$\mathbf{I}_{G_{\mathcal{R}}} = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 1 & -1 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} -1 \quad 0 \quad 0 \quad 0 \\ \\ \mathbf{B} \\ \\ \end{bmatrix},$$

Graph/Matrices analogy

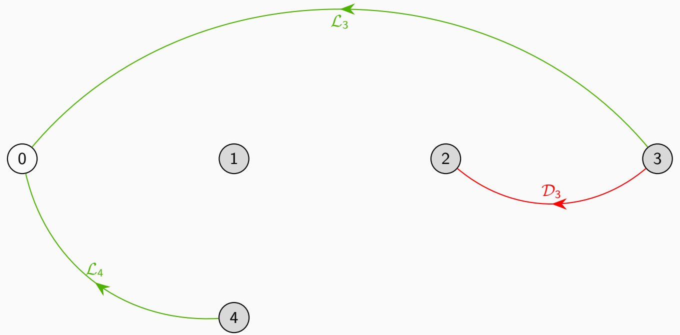
Fault Graph $G_{\mathcal{F}}$



$$\mathbf{I}_{G_{\mathcal{F}}} = \begin{bmatrix} -1 & -1 & -1 & -1 & -1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & -1 & 1 & 0 & 1 \\ 0 & -1 & 0 & 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 & -1 \end{bmatrix} = \begin{bmatrix} -1 \; -1 \; -1 \; -1 \; -1 \; 0 \; 0 \; 0 \\ -\mathbf{A} \end{bmatrix}$$

Graph/Matrices analogy

Fault Graph $G_{\mathcal{F}}$



$$\mathbf{I}_{G_{\mathcal{F}}} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \\ -1 & 0 & -1 \\ 0 & -1 & 0 \end{bmatrix} = \begin{bmatrix} 1 \quad 1 \quad 0 \\ \\ -\mathbf{A}_s \\ \\ \end{bmatrix},$$

Preliminary conclusions

Remarks

- **Contribution 2 : It is possible to derive an algorithmic construction of graphs for the given problem**
  - The residual graph has exactly the same topology as the physical network
  - In the proposed representation, sensors come as edges, which is the dual version of usual water network graph representations
  - A node 0 is needed : It adds a redundant equation
- Solving $\mathbf{A}X = \mathbf{B}Y \Leftrightarrow$ solving $-\mathbf{I}_{\mathcal{F}}X = \mathbf{I}_{\mathcal{R}}Y$ (and $rank(\mathbf{A}) = rank(\mathbf{I}_{\mathcal{F}})$)
- **What are the sufficient and necessary conditions on $G_{\mathcal{F}}$ to guarantee identifiability?**

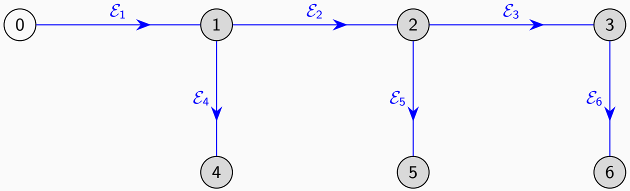
Contribution 3

Theorem 1 : The defaults are identifiable iff $G_{\mathcal{F}}$ is a directed tree

Example

Example

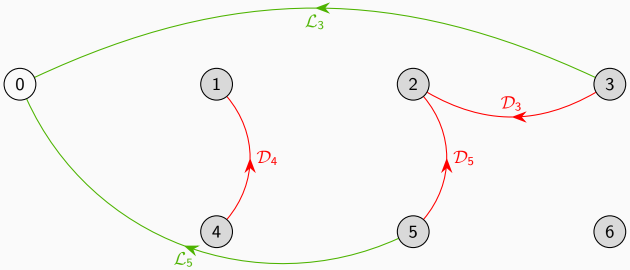
Let the following WDN



Is the default combination $\mathcal{L}_3, \mathcal{L}_5, \mathcal{D}_3, \mathcal{D}_4, \mathcal{D}_5$ identifiable?

Example

Answer : No



$G_{\mathcal{F}}$ is not a directed tree (it contains loops)

Other advantages

- This result can also indicate the resilience against appearance of new defaults
- Complexity for computing the rank of **A** by gaussian elimination: $\mathcal{O}(n^3)$
- Checking if $G_{\mathcal{F}}$ is a directed tree: $\mathcal{O}(n)$

Outline

Original Problem

What to do if the presence of default is unknown?

- The problem is underdetermined $\rightarrow$ Regularisation
- Physical assumption : Leaks must be positive

Contribution 4

$$\hat{X} = \min_{\mathcal{X}} \|\mathbf{A}\mathcal{X} - \mathbf{B}Y\|^2 + \mathcal{R}$$

$$s.t \quad \mathcal{L}_i > 0 \quad \forall i \in [1, \ldots, |E_{G_{\mathcal{R}}}| - 1]$$

$\mathcal{R}$ is the regularization term

Propositions for $\mathcal{R}$

- The solution with the least number of defaults is the most acceptable
- $\mathcal{R} = |X|_0 \rightarrow \ell_0$ norm relaxed as ℓ_1 norm = LASSO
- The solution closest to yesterday solution is the most acceptable (requires a ground truth on day 1)
- $\mathcal{R} = |\Delta X|_0 \rightarrow \ell_0$ norm relaxed as ℓ_1 norm = LASSO
- **Problem : possibly multiple solutions**

Proposition using previous results

Statement

- Theorem 1 $\rightarrow$ it is possible to determine all identifiable default structures
- Strategy : Replace one quadratic optimization scheme by multiple quadratic analytic solutions

Algorithm

- Step 1 (offline): determine all identifiable structures
- Step 2 (online) : identify all identifiable structures using LS
- Step 3 (online) : retain the physical plausible solutions
- Step 4 (online) : retain the solutions fulfilling the regularisation

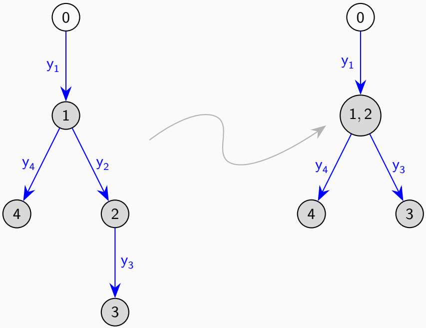
Advantages

- It is possible to display a range of solutions
- It is possible to enforce *a priori* by acting on step 3 (If we know some defaults are present)

Handling missing data

Merging nodes

What if sensor 2 is down?

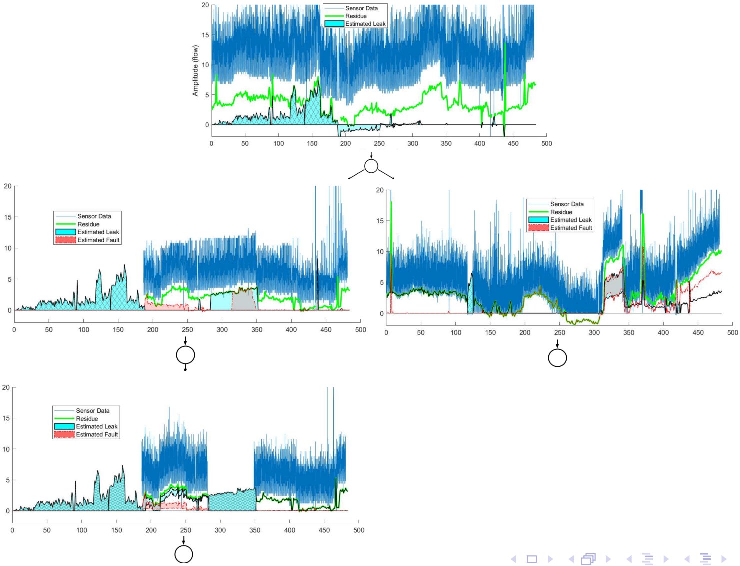


Consequences

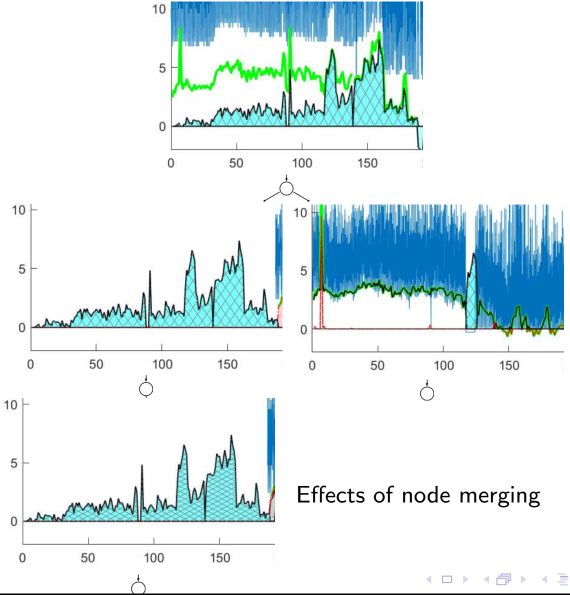
- Any leak $\mathcal{L}$ estimated on the merged DMA $1, 2$ corresponds to a leak positioned either on DMA 1 or on DMA 2

Outline

Results on a true data set

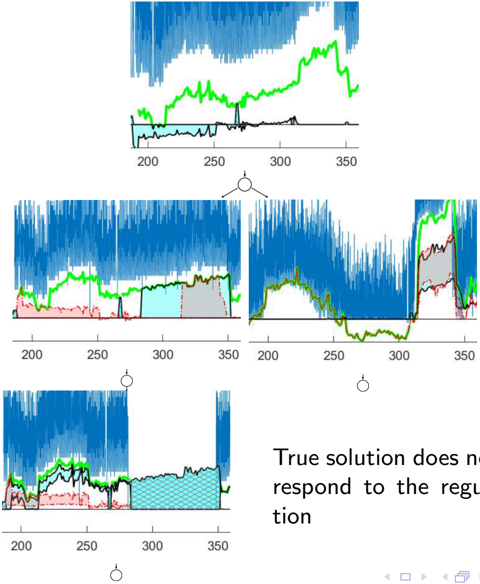


Results on a true data set (zoom 1)



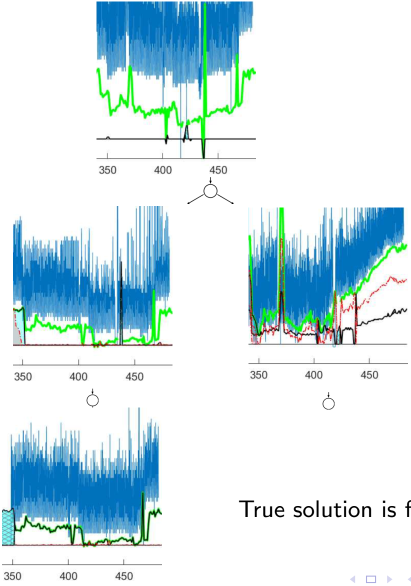
Effects of node merging

Results on a true data set (zoom 2)



True solution does not correspond to the regularisation

Results on a true data set (zoom 3)



True solution is found

Outline

1. Noisy sensors modelling
2. Graphs paralellism
3. Identification of leaks
4. Real dataset results
5. Conclusions

Conclusions

Context

- Leak identification in water dynamical networks
- Realistic assumptions : Sensors are faulty

Contributions

- Graphs transposition for the identifiability problem
- Sufficient and necessary identifiability conditions fully expressed in graph topology
- Proposition of a regularised identification scheme algorithm
- Test on real data

Water distribution networks leaks estimation in faulty sensor context: a graph approach

K. Srinivasarengan, S. Aberkane, V. Laurain

25 Nov 2021
vincent.laurain@univ-lorraine.fr