



cfms

COMITÉ FRANÇAIS DE MÉCANIQUE
DES SOLS ET DE GÉOTECHNIQUE

Gestion des Données et Nouvel Environnement Numérique en Géotechnique

LABORATOIRE SOUTERRAIN DE L'ANDRA : DONNEES BRUTES,
STRUCTURATION ET CAS D'USAGES DU MACHINE LEARNING

COTTON, Julien

15 NOVEMBRE 2022



L'Andra

Cadre général

Sa mission : L'Andra remplit avec engagement et responsabilité une mission d'intérêt général, confiée par l'État, au service des Français : **prendre en charge les déchets radioactifs produits par les générations passées et présentes, et les mettre en sécurité pour les générations futures**

- Établissement public à caractère industriel et commercial (Epic)
- Missions d'intérêt général Indépendante des producteurs de déchets radioactifs
- Placée sous la tutelle des ministères en charge de l'énergie, de la recherche et de l'environnement
- ≈ 650 salariés sur 4 sites



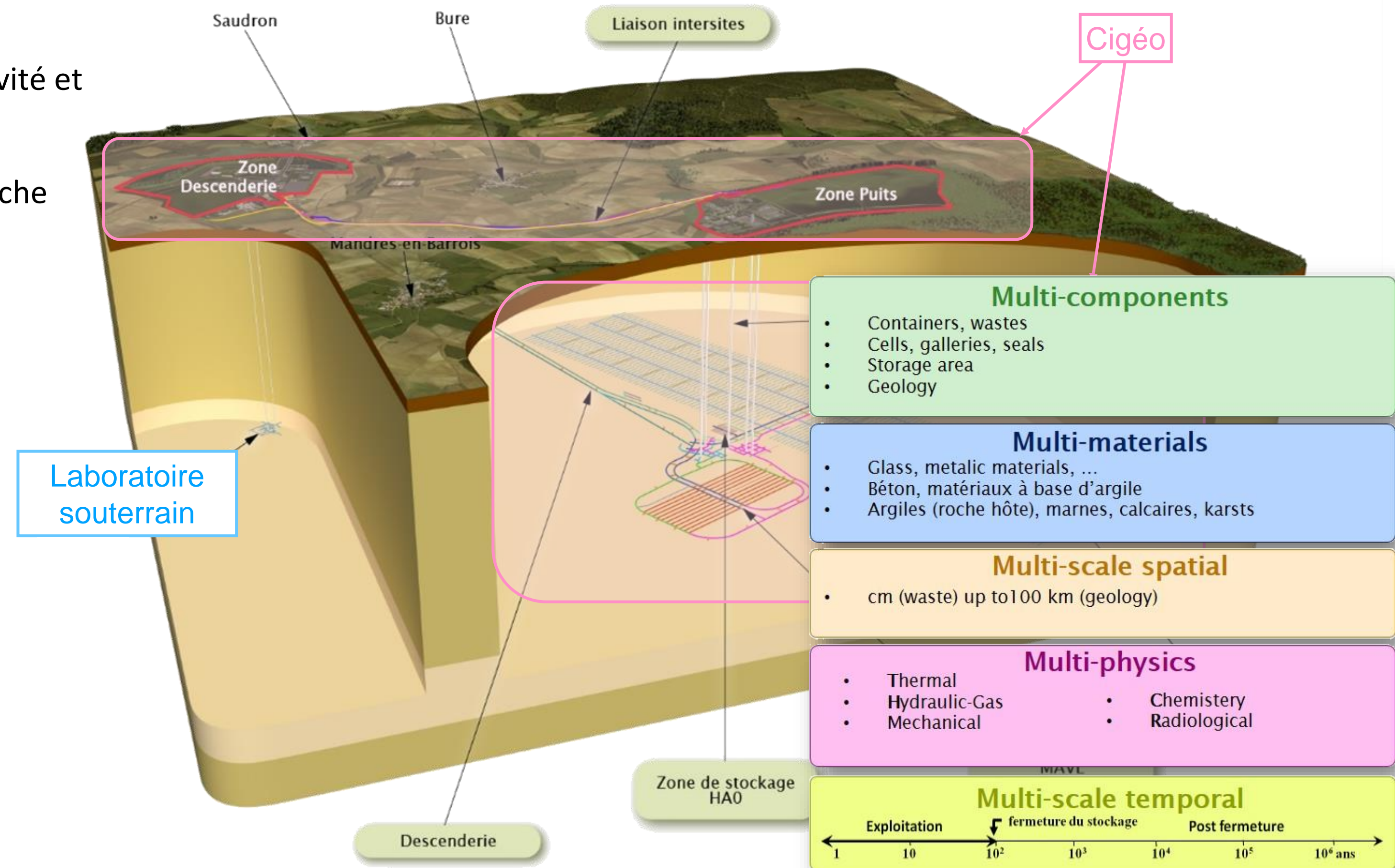
Cigéo

le projet de centre industriel de stockage en couche géologique profonde

- Un projet de stockage réversible pour les déchets de haute activité et de moyenne activité à vie longue
- Concept de stockage à 500 mètres de profondeur dans une couche argileuse (le Cox : Callovo-Oxfordien)
- Mise en exploitation prévue à l'horizon 2035-2040
- Un ouvrage hors normes : **multi-échelles**
- Un centre d'étude spécifique ouvert depuis le début des années 2000 en Meuse/Haute-marne **pour préparer et accompagner le développement progressif de Cigéo**



Le Laboratoire Souterrain de Meuse/Haute-Marne



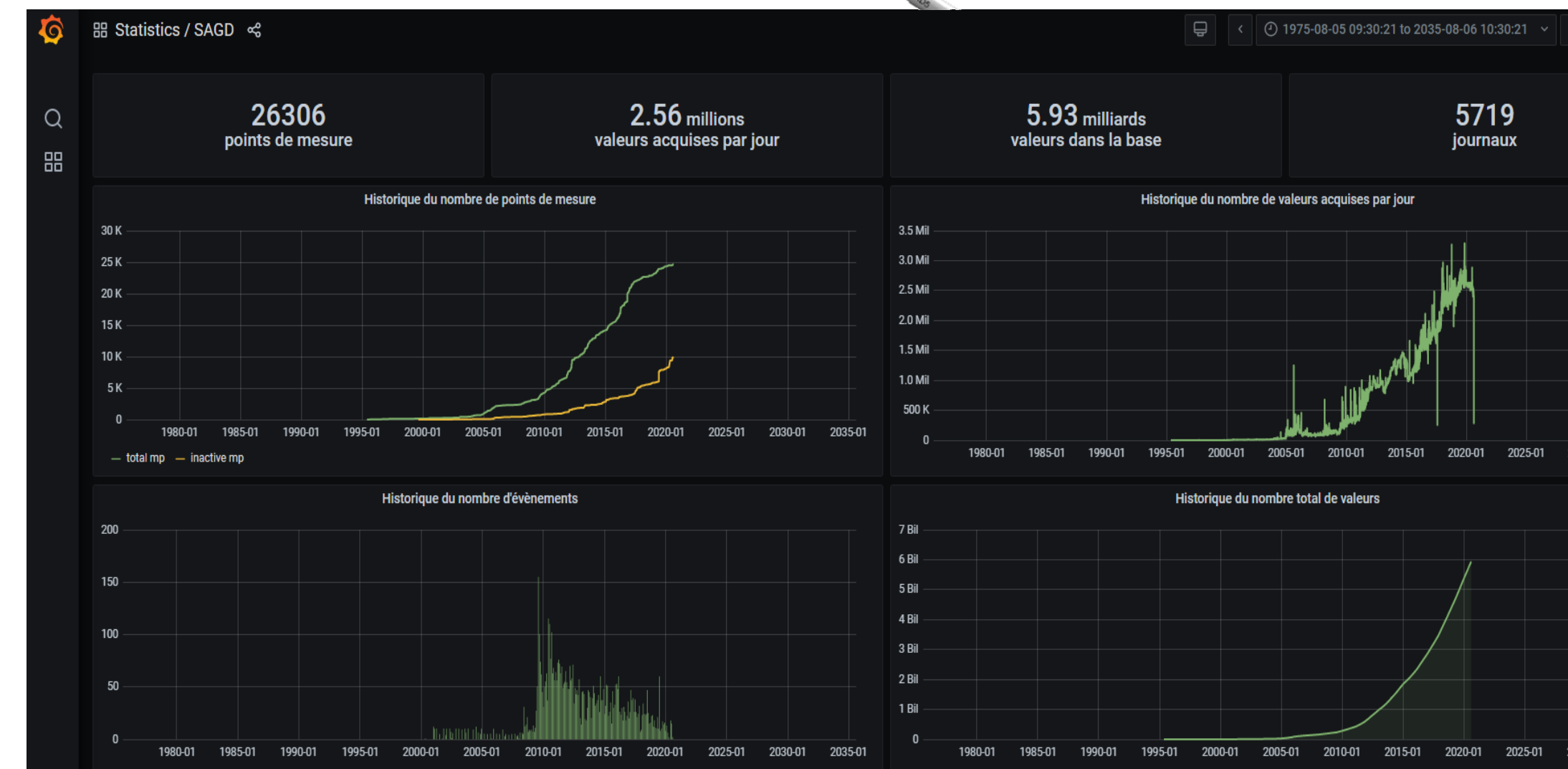
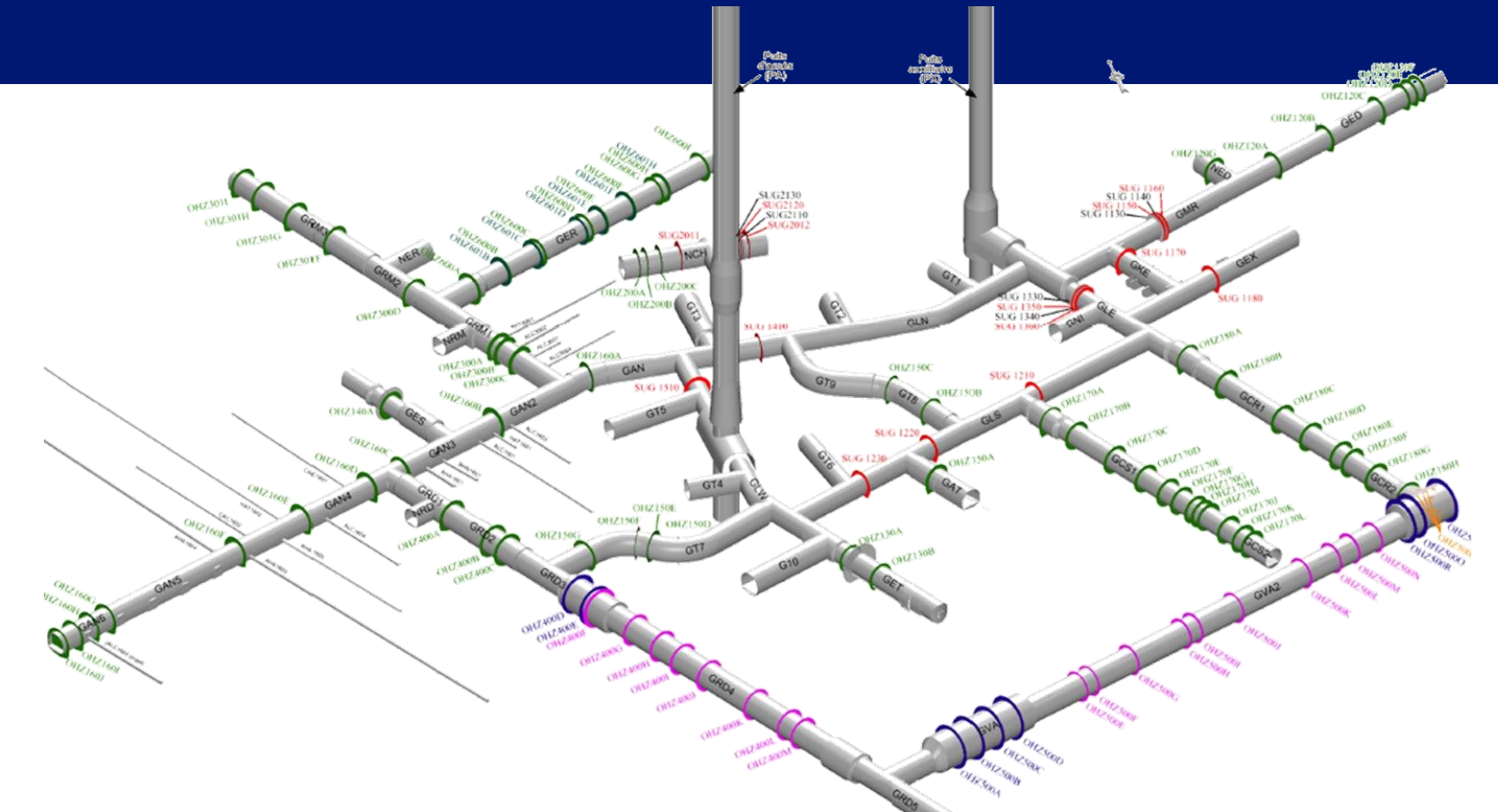
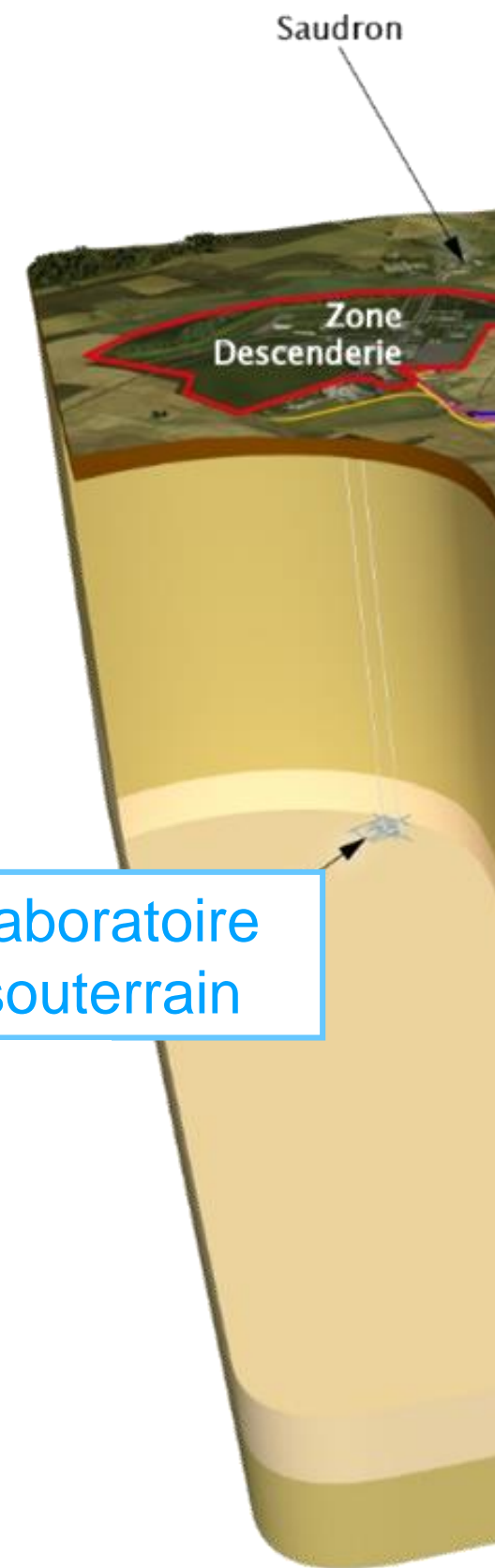
Le Laboratoire souterrain de recherche

Un des outils dont dispose l'Andra pour l'acquisition de connaissances

L'acquisition de connaissances est au cœur des enjeux de sûreté, de conception et de surveillance. Une connaissance « temps réel » et prédictive, incluant les incertitudes de toutes natures, supportée par d'importants moyens de caractérisation (données et processus), de mesures in situ (surface et fond), et de simulations numériques à différentes échelles de temps/espace.

Le Laboratoire souterrain, c'est :

- Plus de 2 Km de galeries expérimentales ;
- Plus de 2000 forages ;
- 2,5 millions de valeurs par jours pour un cumul de plus de 6 milliards de valeurs depuis les années 2000.



Le Laboratoire souterrain de recherche

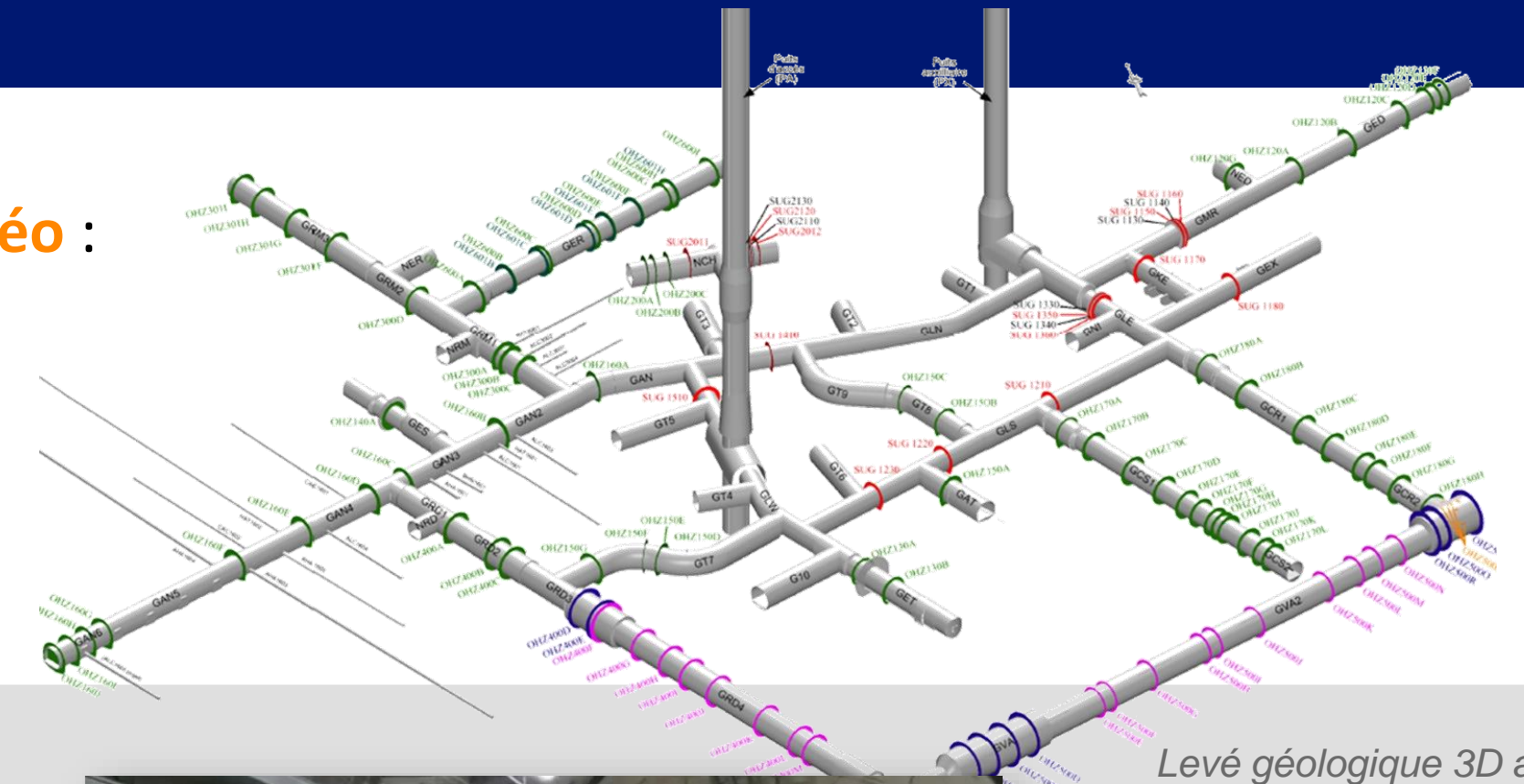
Un des outils dont dispose l'Andra pour l'acquisition de connaissances

Le Laboratoire souterrain est utilisé pour **préparer et accompagner le développement progressif de Cigéo** :

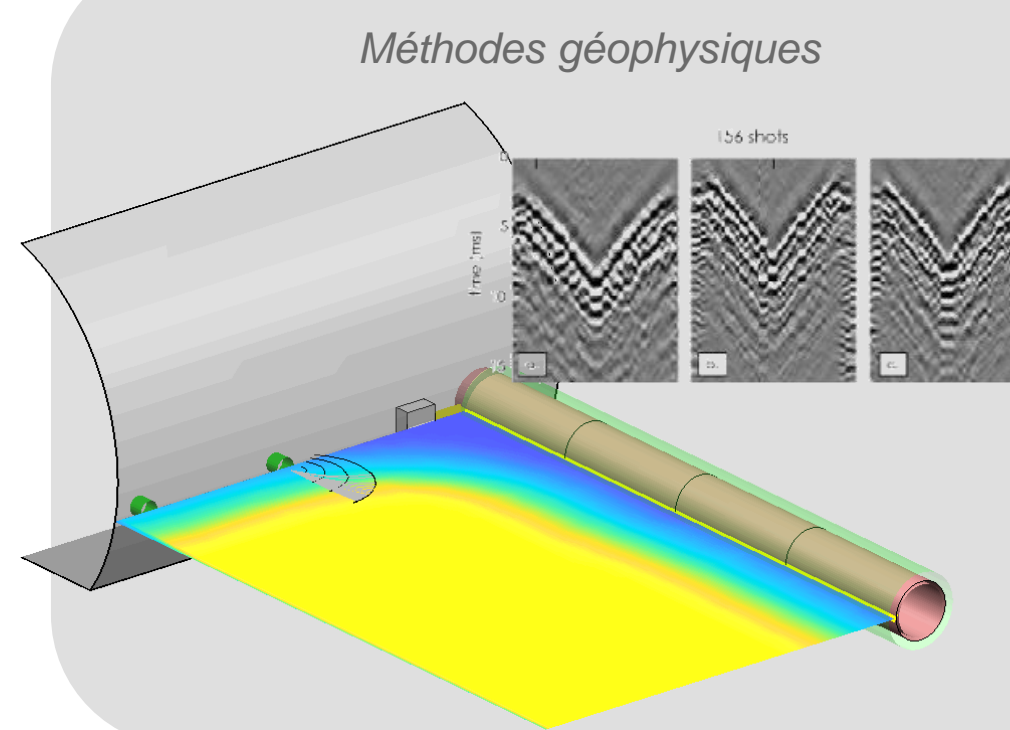
- étudier la roche et ses propriétés de confinement en conditions réelles ;
- mettre en œuvre les concepts de l'Andra sous forme de démonstrateurs à l'échelle (alévoles HA, MAVL, ouvrages de fermeture, remblais...);

- tester et évaluer des techniques innovantes ou plus classiques pour :

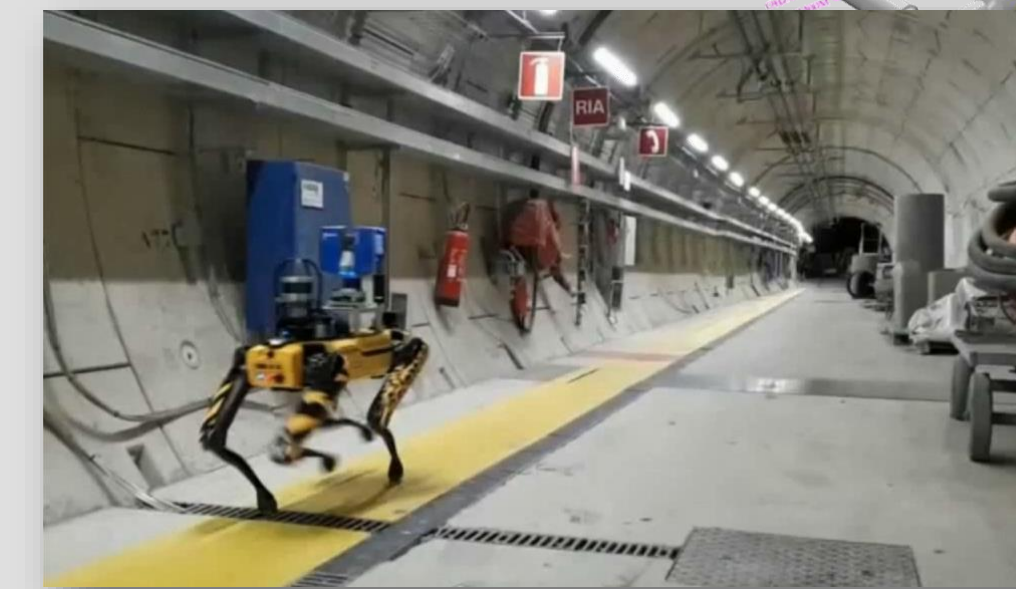
- Le suivi de réalisation des ouvrages **avant et pendant leurs constructions** : plans 2D et 3D, forages, levés géologiques, inspection (recette d'ouvrage);
- La surveillance de l'évolution des ouvrages **après réalisation** grâce à des dispositifs de surveillances à demeure (capteurs ponctuels, mesures réparties par Fibres Optiques FO) ;



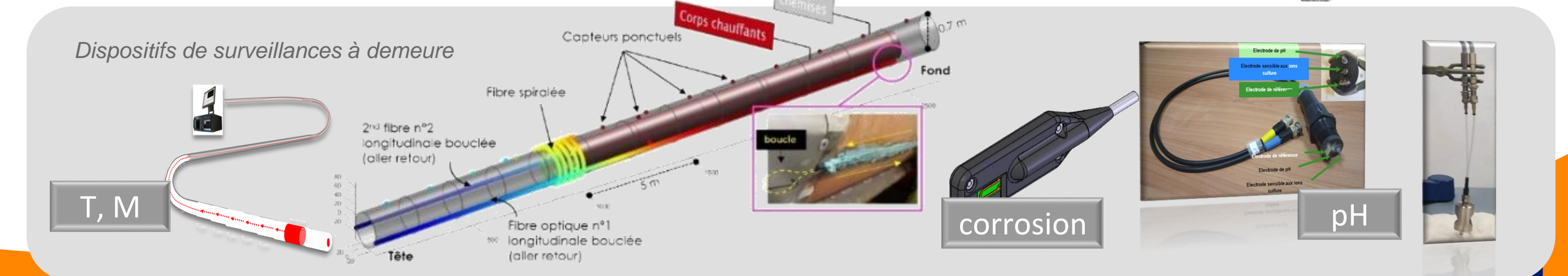
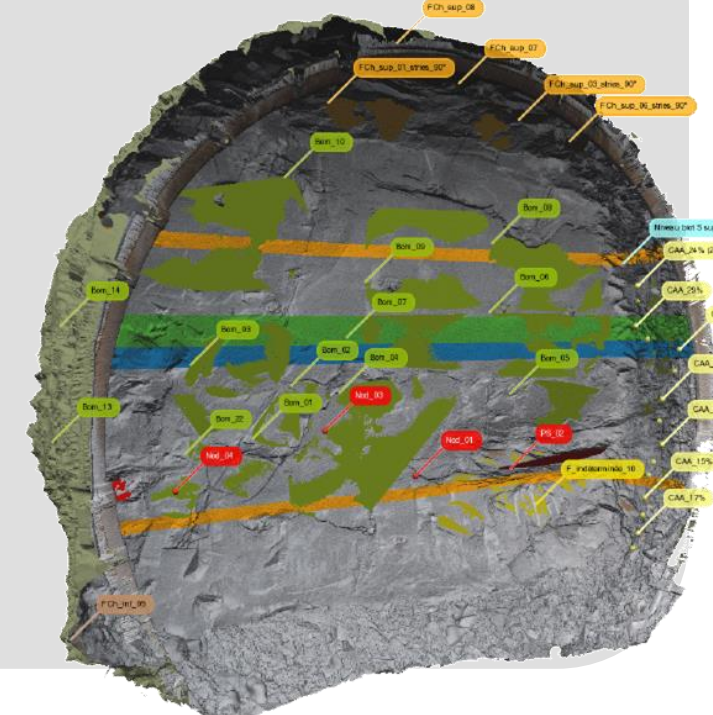
Levé géologique 3D au front



Méthodes géophysiques



Inspection robotisée au LIDAR 3D en galerie



Dispositifs de surveillances à demeure

Démonstrateur d'alvéole HA

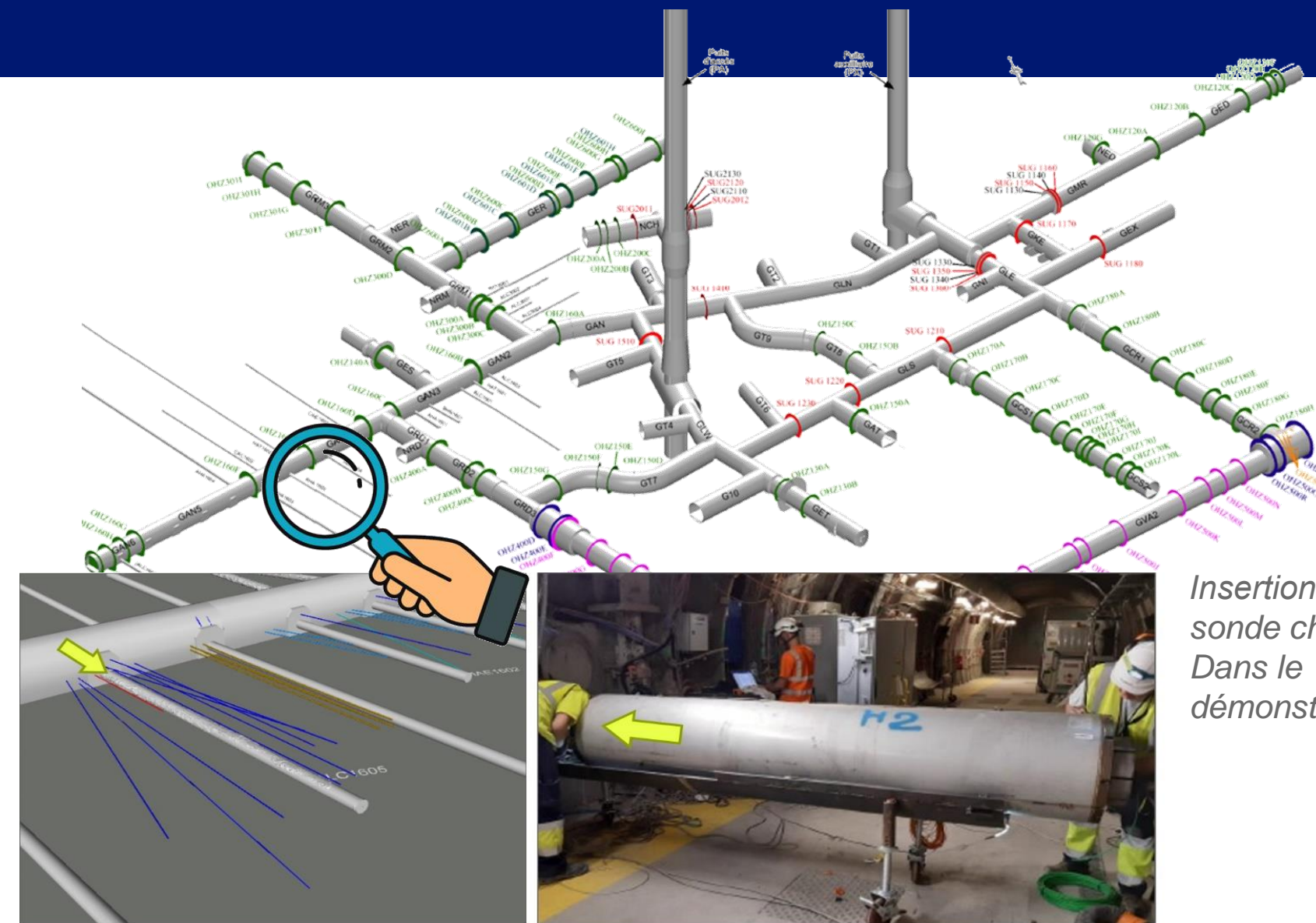
Impact d'un chargement thermique

Un démonstrateur d'alvéole HA : ALC1605

Ce démonstrateur permet de reproduire de manière artificielle **la montée en température** (initiée à l'été 2020) des déchets de haute activité. L'objectif général de l'expérimentation ALC1605 est **d'étudier l'impact d'un chargement thermique sur l'alvéole (sur le chemisage, sur la roche et sur le matériau de remplissage).**

Des données de toutes natures :

- Conception : 5 types de chemises utilisées (et instrumentées) ; 8 forages périphériques instrumentés, des sondes chauffantes (également instrumentées), etc...
- Inspection (recette d'ouvrage) : scan 3DHR avant/après tubage
- Surveillance (capteurs ponctuels et FO) : **6095** séries temporelles de mesures thermiques ; **3541** séries temporelles de mesures mécaniques ;



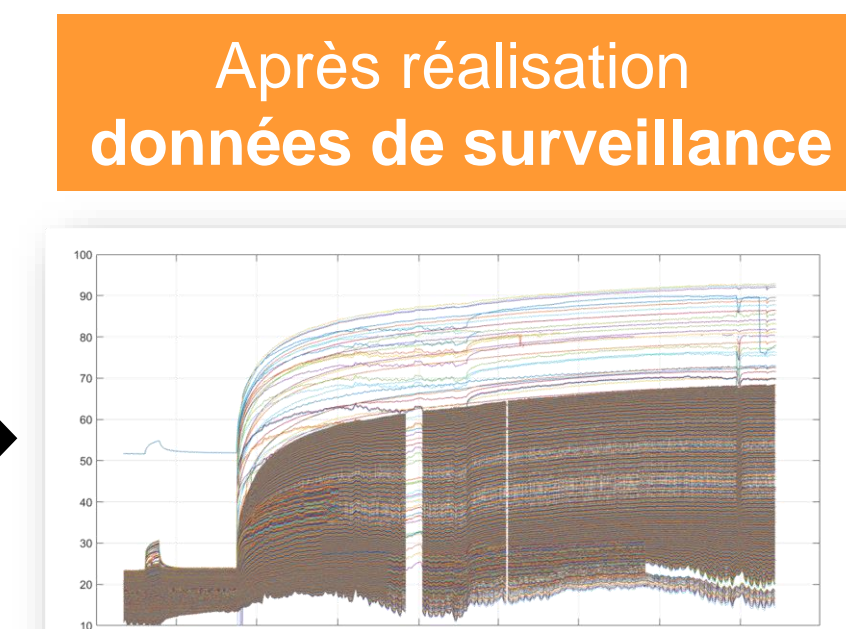
Insertion d'une sonde chauffante Dans le démonstrateur



Ex : Chemise utilisée pour le tubage du démonstrateur



Ex : Scan 3D trou nu et tubé



Ex : Mesures in situ

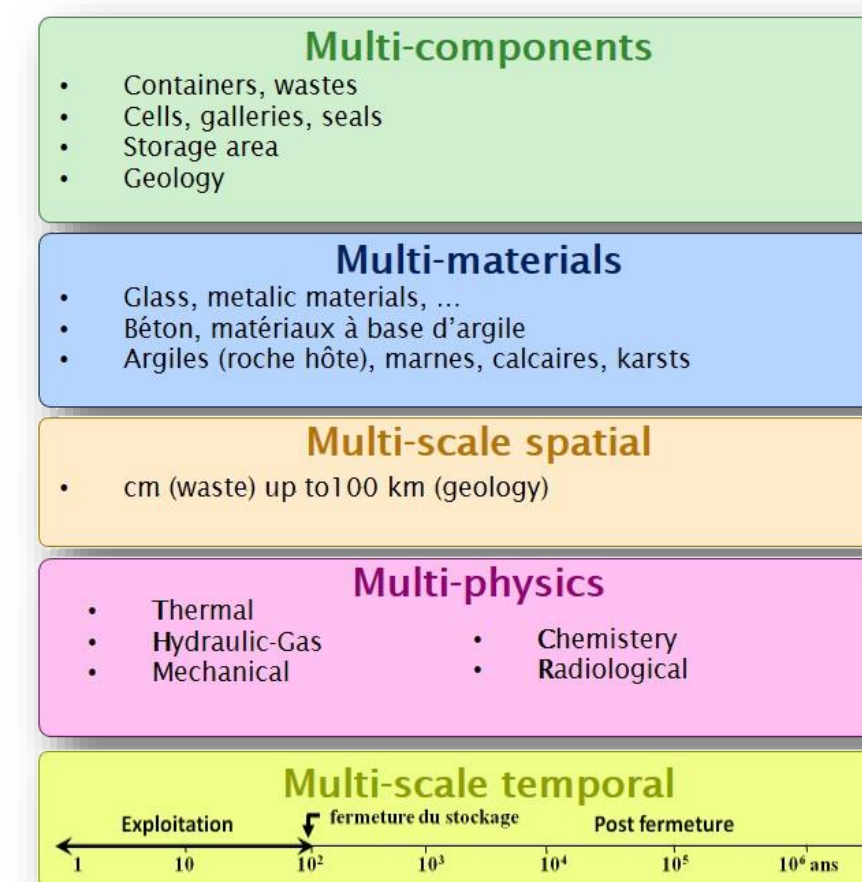
Structurer les données

gagner en efficacité sur toute la chaîne de données

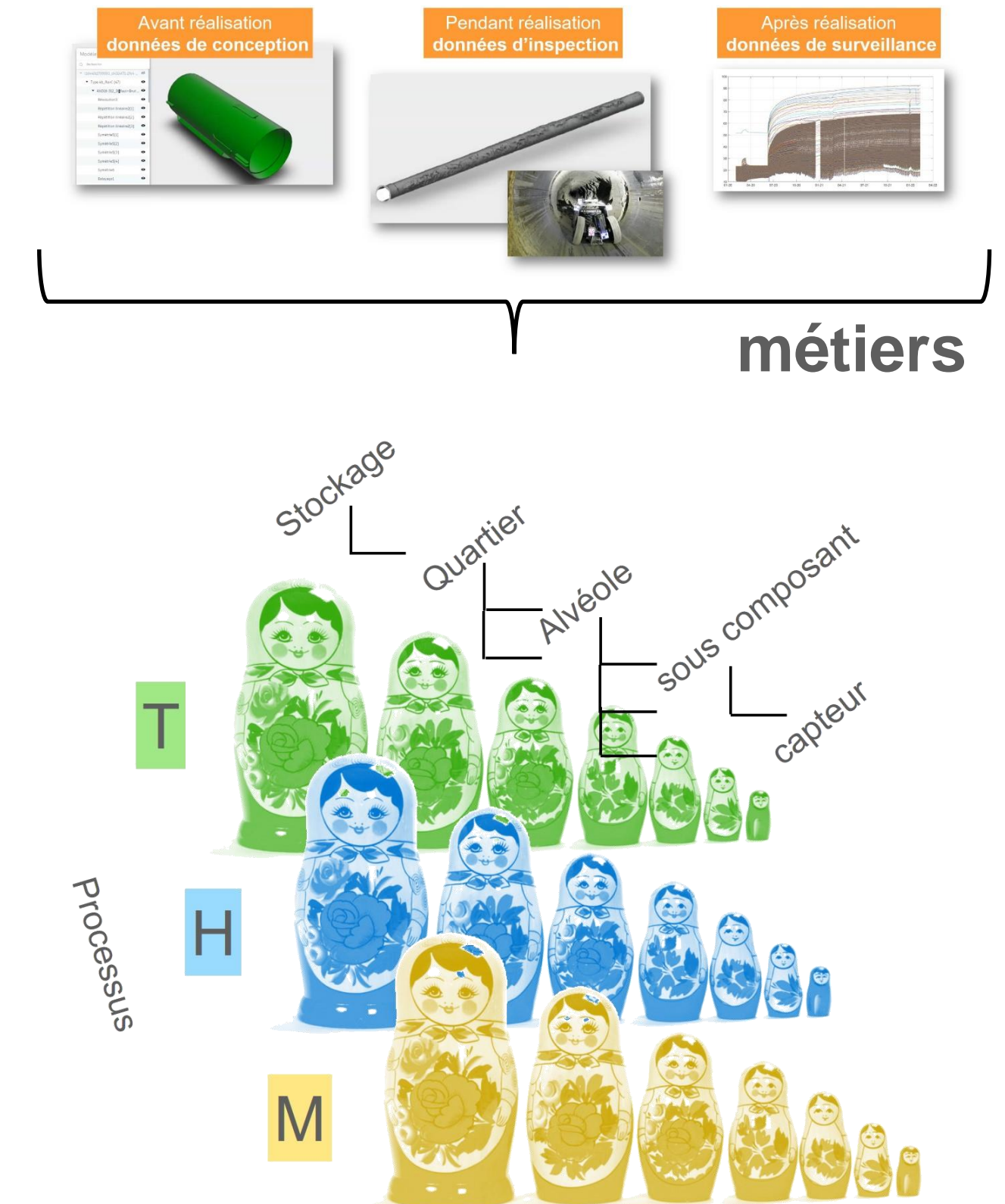
Une structuration des données en lien avec les spécificités du stockage (« gestion multi ») et en lien avec les spécificités métiers :

Quelques points de vues :

- **Technique (« gestion multi »)** : des capteurs peuvent être proches mais liés à des composants différents avec des natures et des comportements différents (roche, revêtement en acier, intérieur ou extérieur ...).
- **Métier de l'instrumentation** : famille de capteur, capteurs en particulier...
- **Projet** : Alvéole, sous-composant, processus...



projet



Structurer les données

gagner en efficacité sur toute la chaîne de données

Structure hiérarchique :

- Toutes les relations entre capteurs (et leurs attributs), composants et processus sont décrites numériquement par des liens « parent/enfant » dans la structure et des hyperliens extérieurs vers les autres bases de données

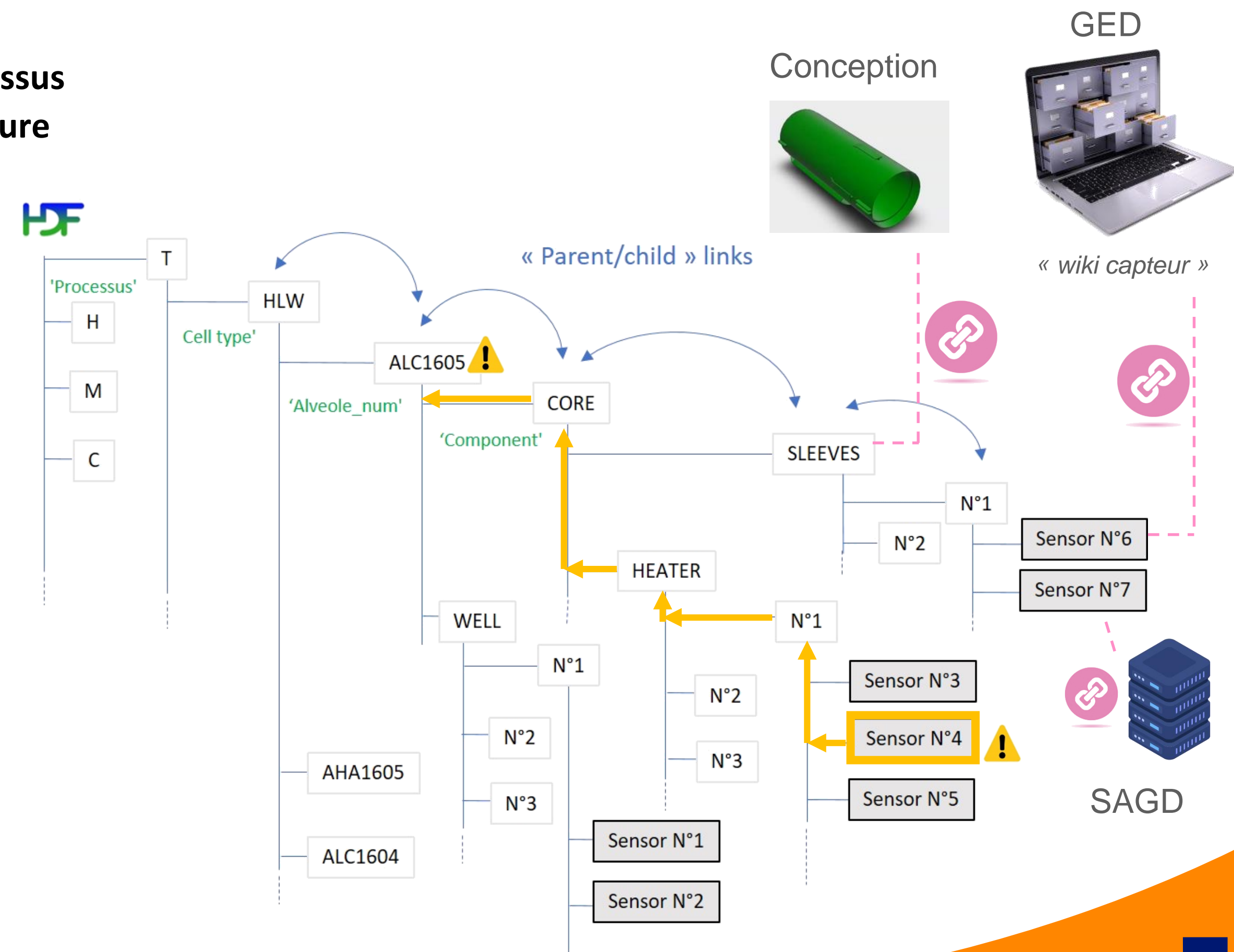
➔ collecte/sélection efficace des données en filtrant simplement la structure sur des champs choisis : processus et/ou composants/sous-composants.

➔ Lien et balises vers documents (GED) ; ex : fiches capteurs

➔ Lien et balises vers les éléments de conception

- La structure permet le transfert d'informations « multi-échelles »

➔ Ex : un indicateur de performance au niveau capteur peut être remonté jusqu'au niveau parent supérieur (« chemisage acier » au niveau intermédiaire ou « cellule HA » juste au-dessus)

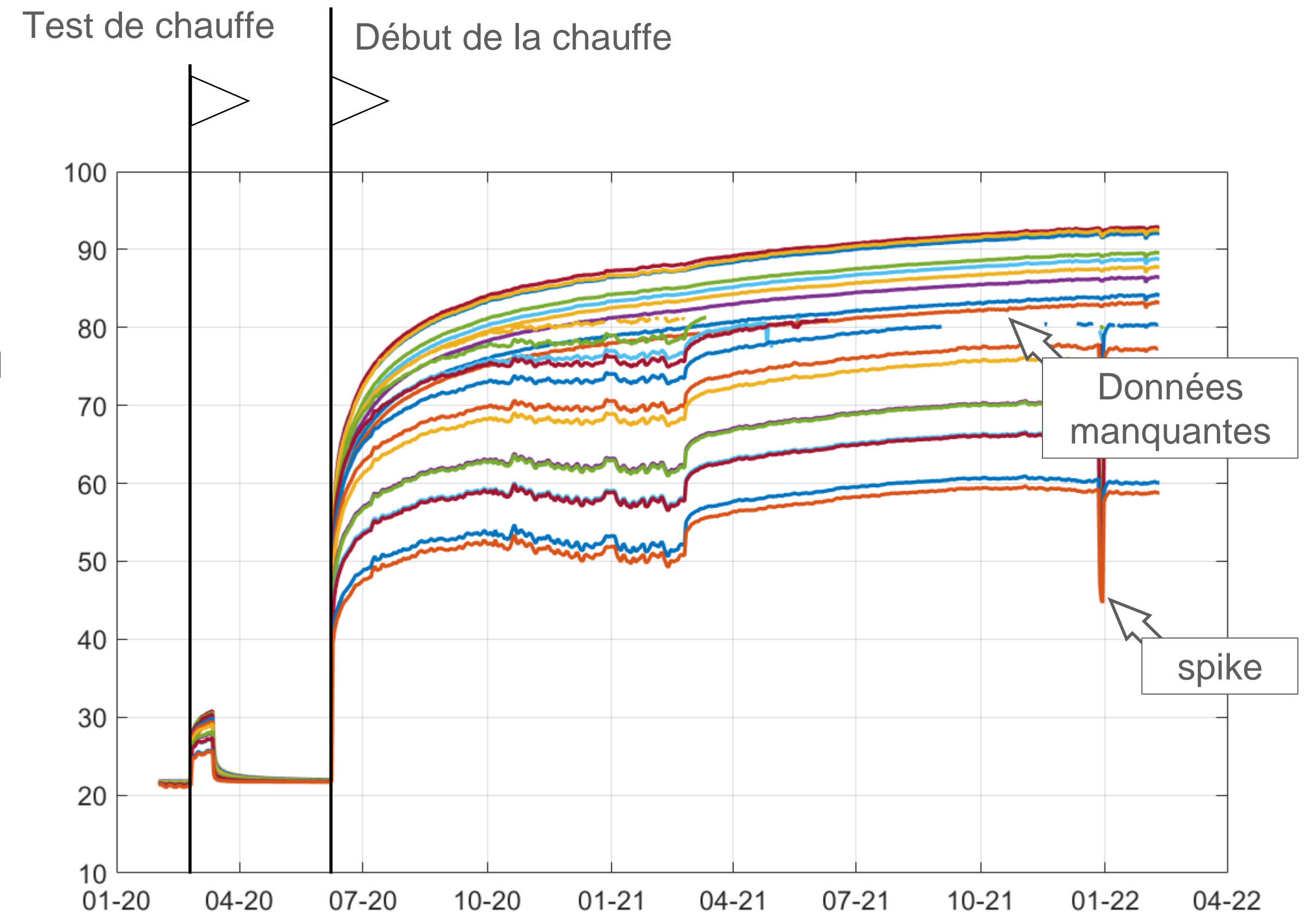
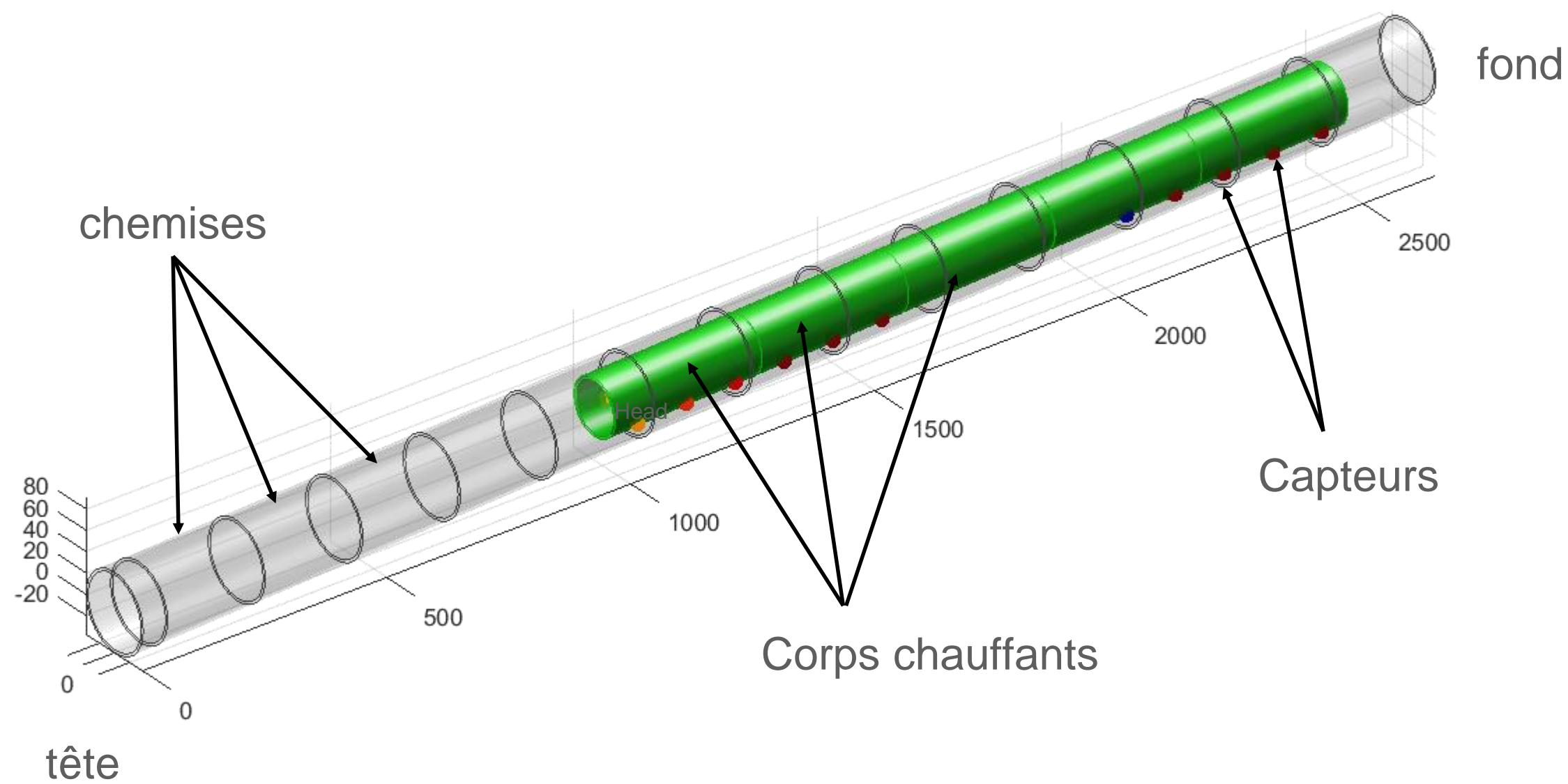


Structurer les données

gagner en efficacité sur toute la chaîne de données

HF Request (= filter)

```
{'Alveole_num' } == ALC1605  
{'Processus' } == T  
{'Ss_composant_type'} == corps chauffant
```

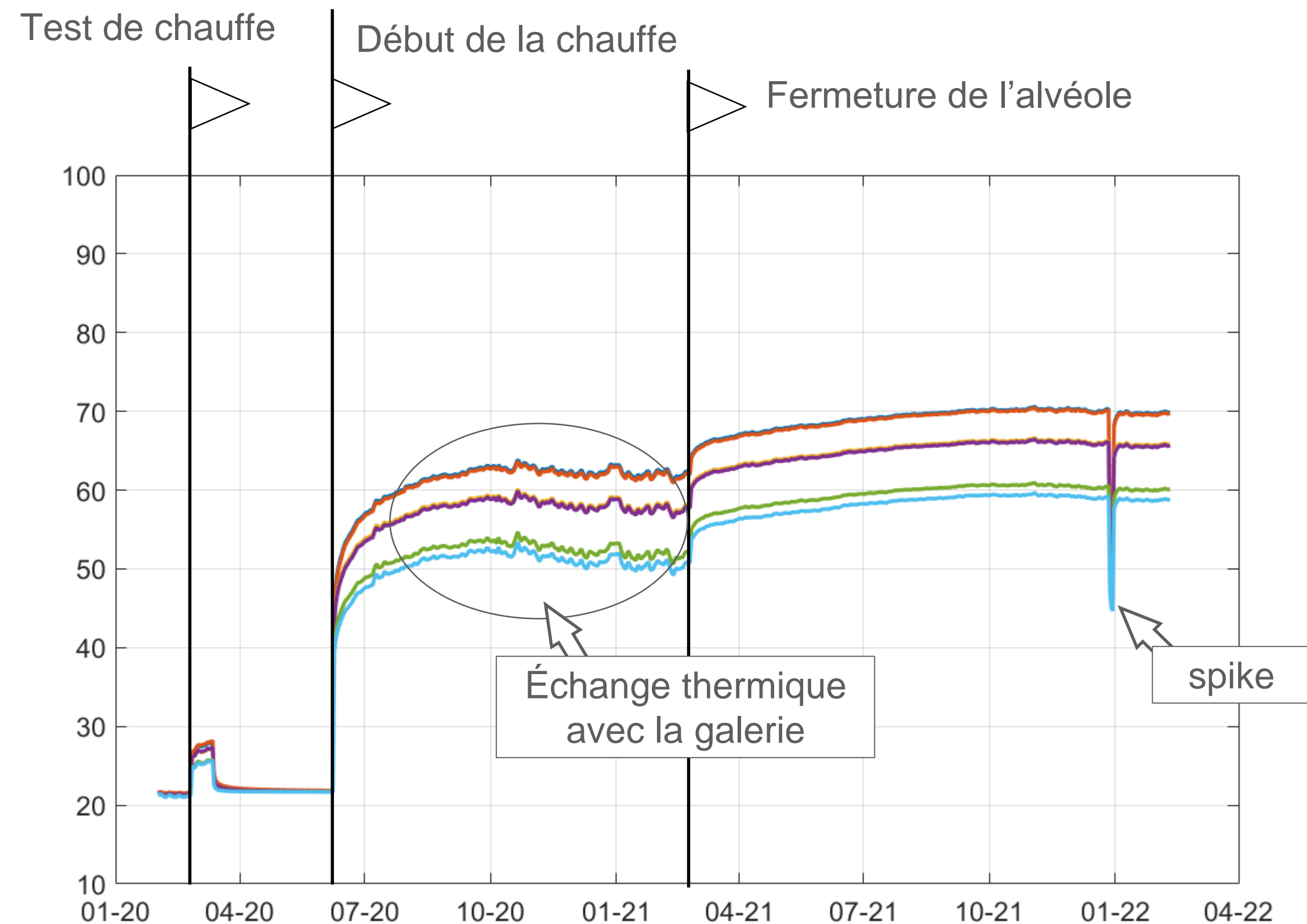
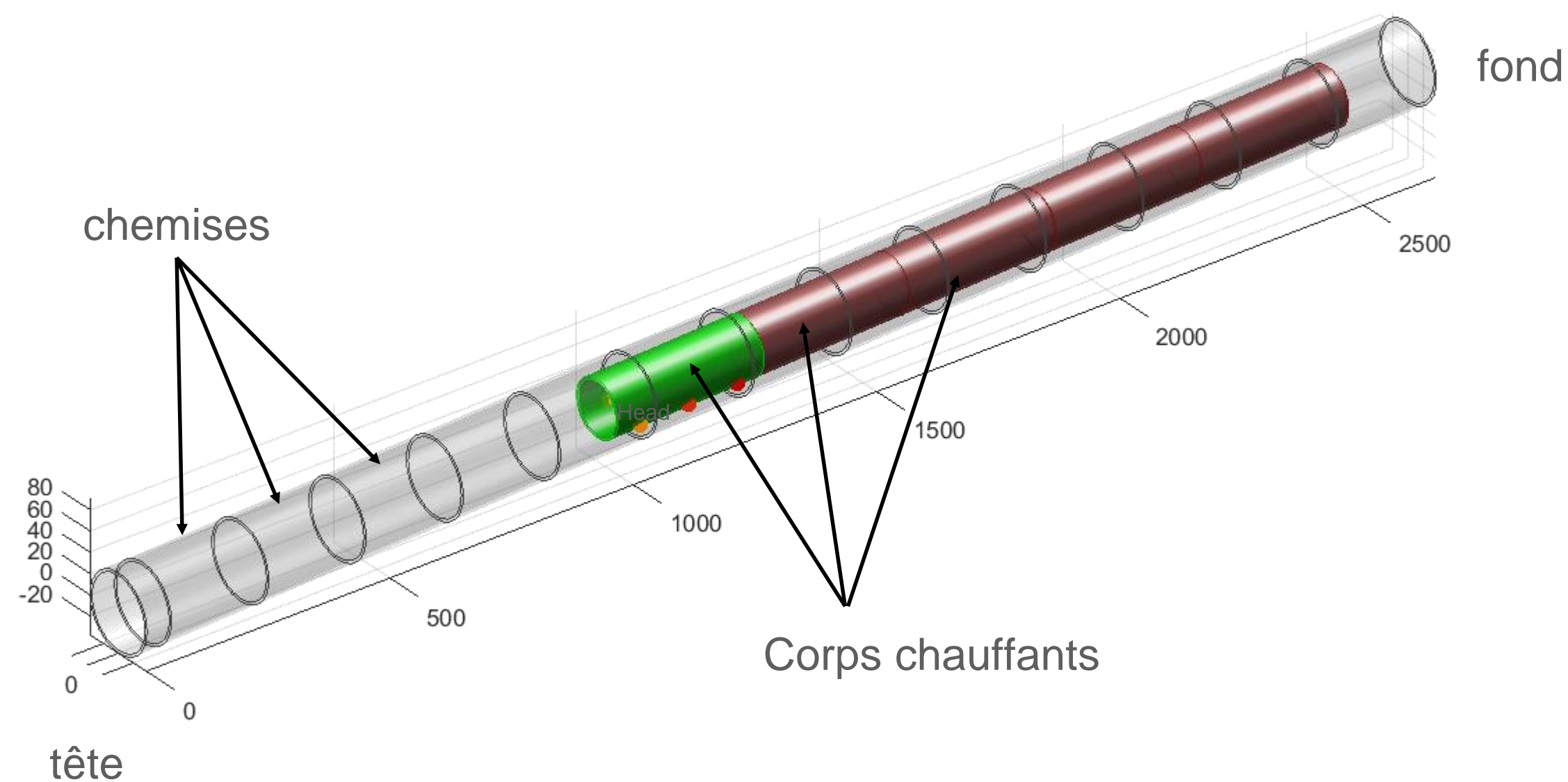


Structurer les données

gagner en efficacité sur toute la chaîne de données

HF Request (= filter)

```
{'Alveole_num' } == ALC1605
{'Processus' } == T
{'Ss_composant_type'} == corps chauffant
{'Ss_composant_num' } == 1005
```

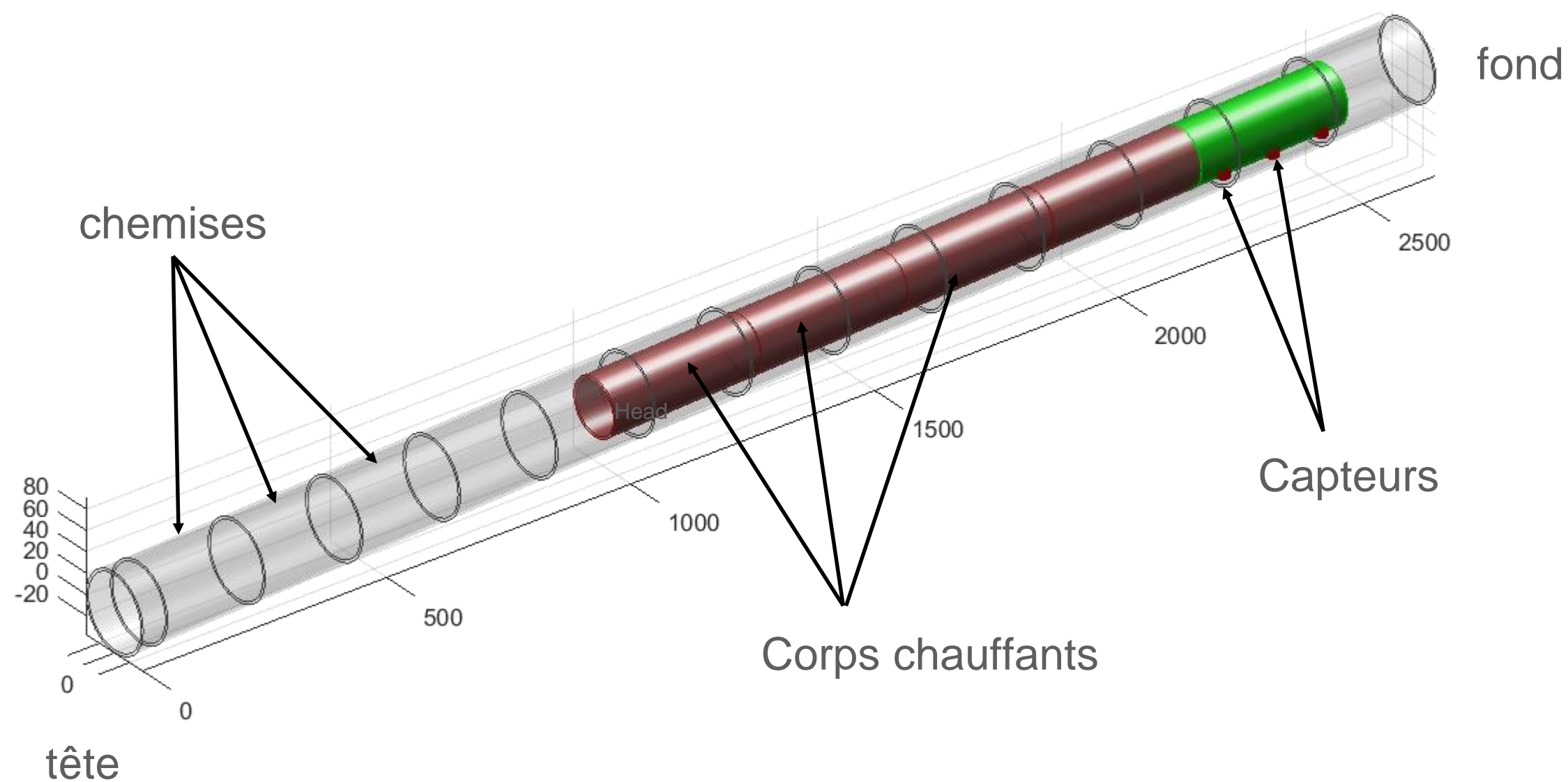


Structurer les données

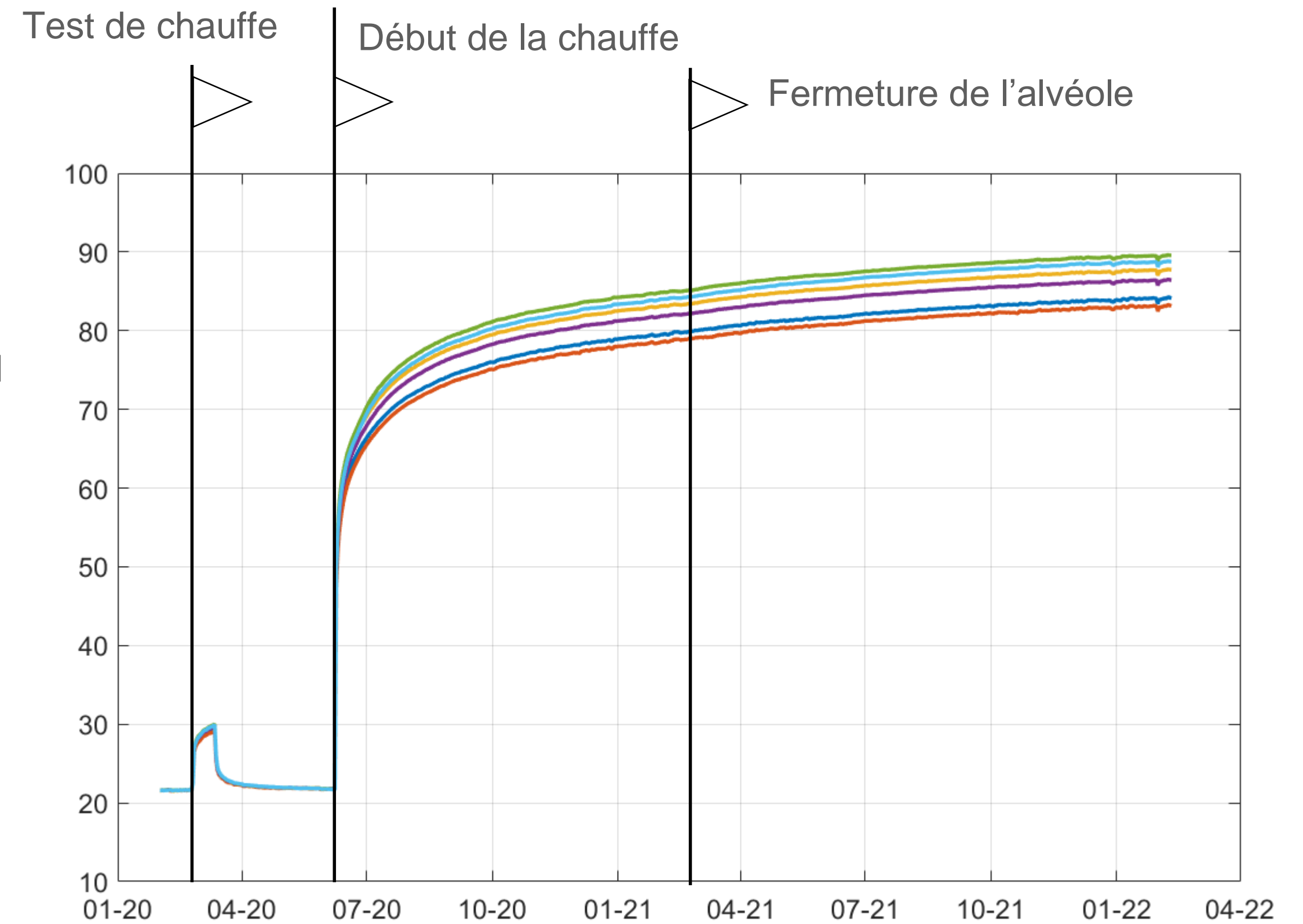
gagner en efficacité sur toute la chaîne de données

HF Request (= filter)

```
{'Alveole_num' } == ALC1605
{'Processus' } == T
{'Ss_composant_type'} == corps chauffant
{'Ss_composant_num' } == 1001
```



■ Composants sélectionnés

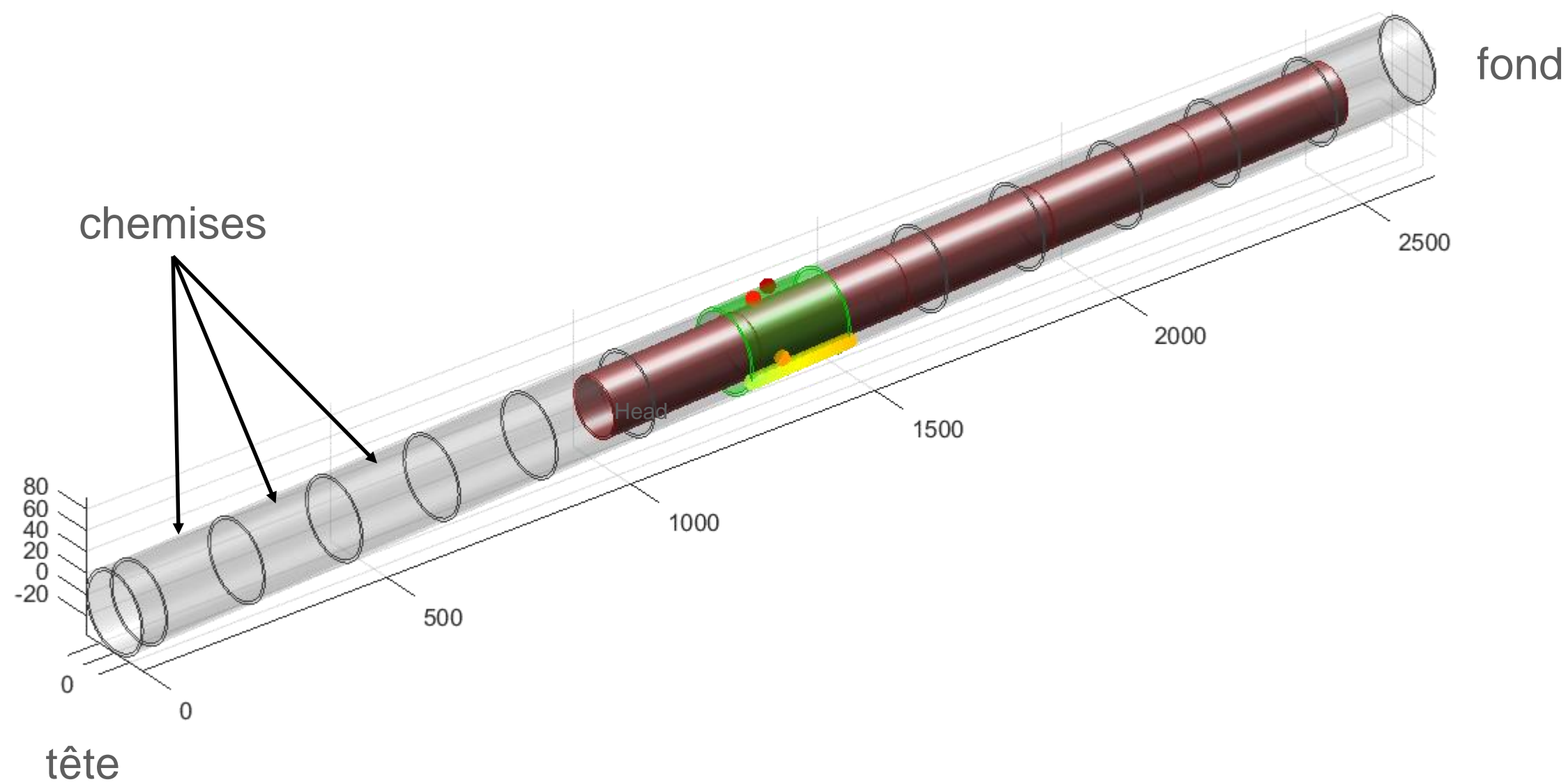


Structurer les données

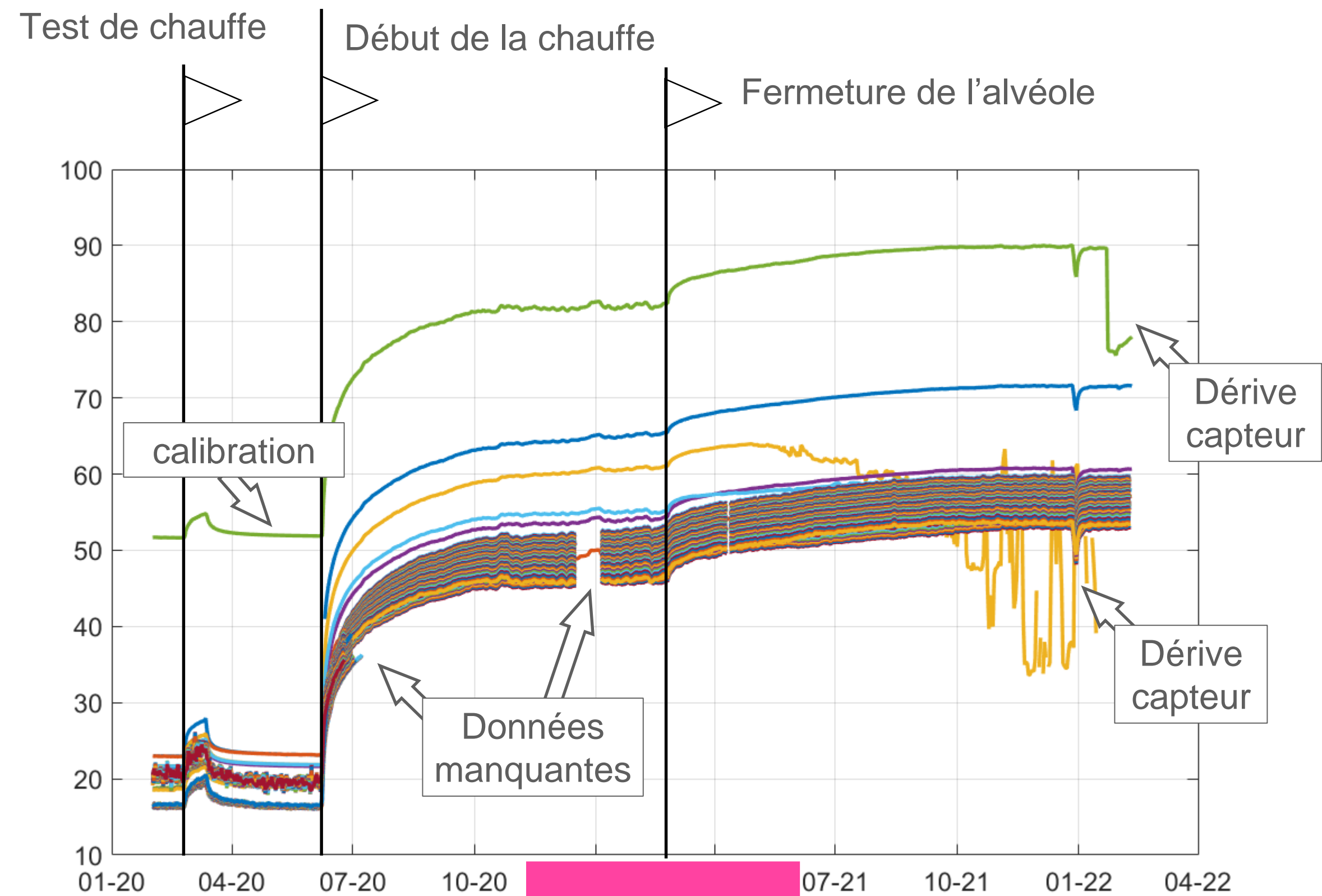
gagner en efficacité sur toute la chaîne de données

HF Request (= filter)

```
{'Alveole_num' } == ALC1605
{'Processus' } == T
{'Ss_composant_type'} == chemise
{'Ss_composant_num' } == 7
```



■ Composants sélectionnés



assurer la cohérence d'ensemble des données

Machine Learning

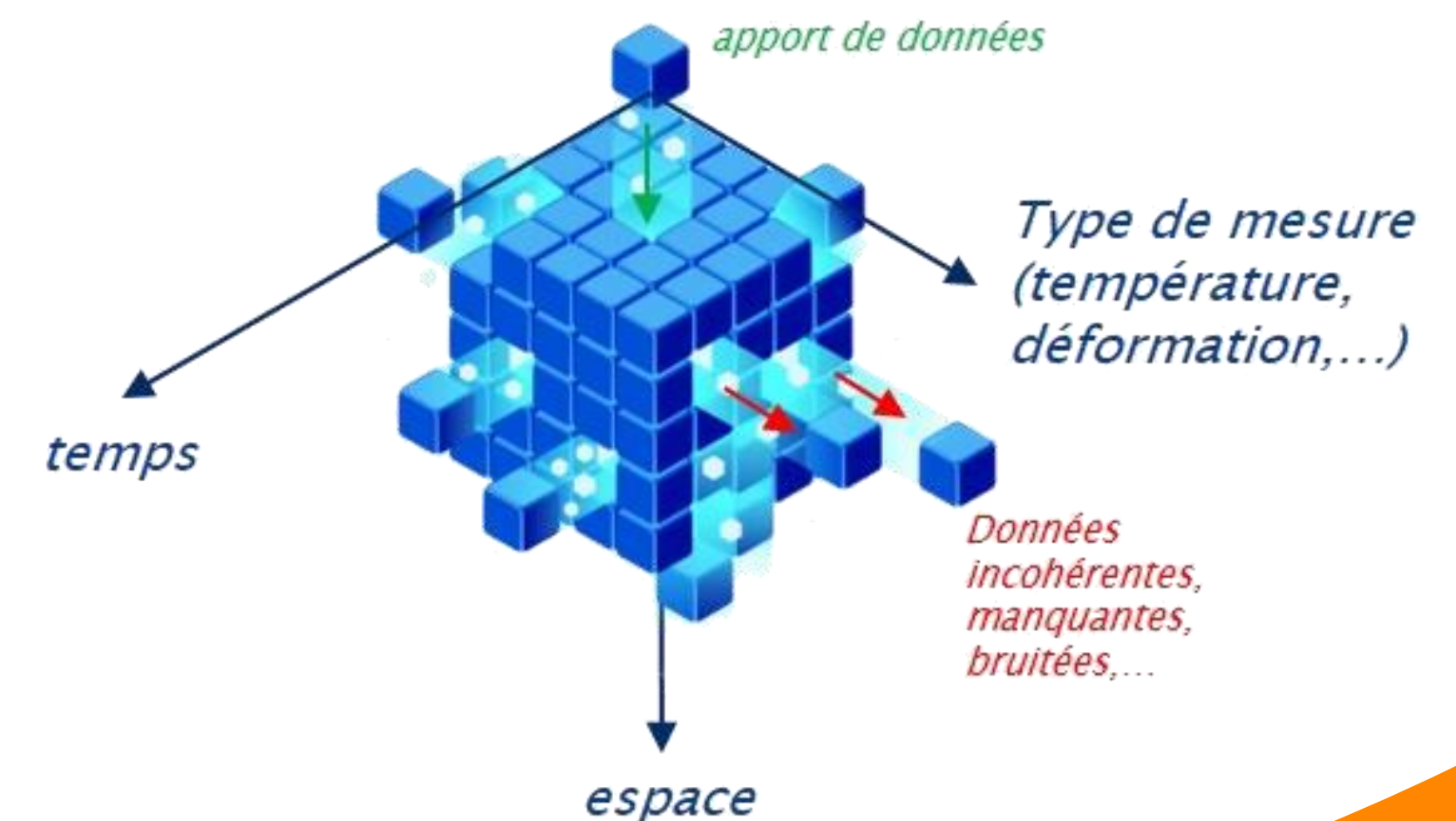
un outil pour assurer la cohérence d'ensemble des données

Il s'agit de pouvoir gérer la combinaison de différentes sources de données afin :

- d'améliorer la compréhension d'un système et de son évolution ;
- de fournir une interprétation plus précise et plus fiable des données en s'appuyant sur leur complémentarité ;
- de fournir des indicateurs de performance ou des outils d'aide à la décision durant la phase de fonctionnement de Cigéo.

La cohérence d'ensemble des données peut être affectée notamment par :

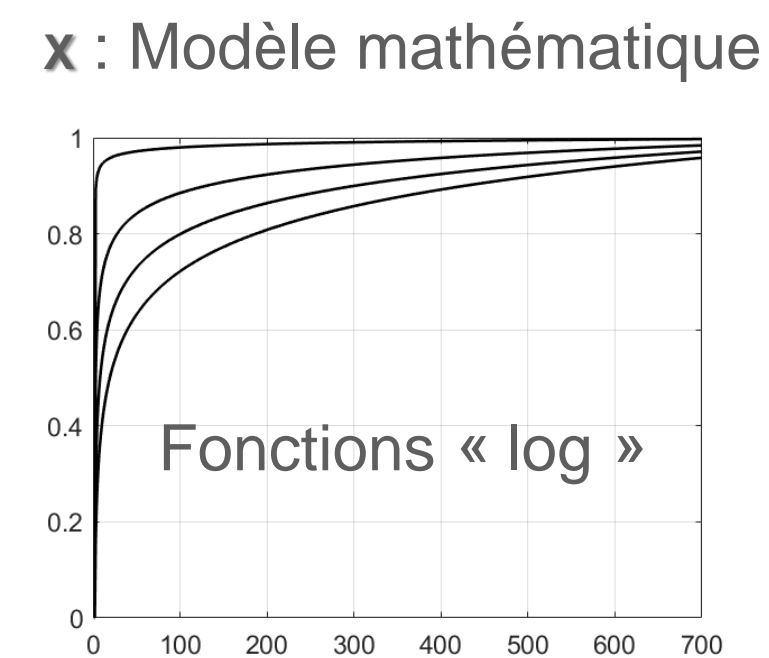
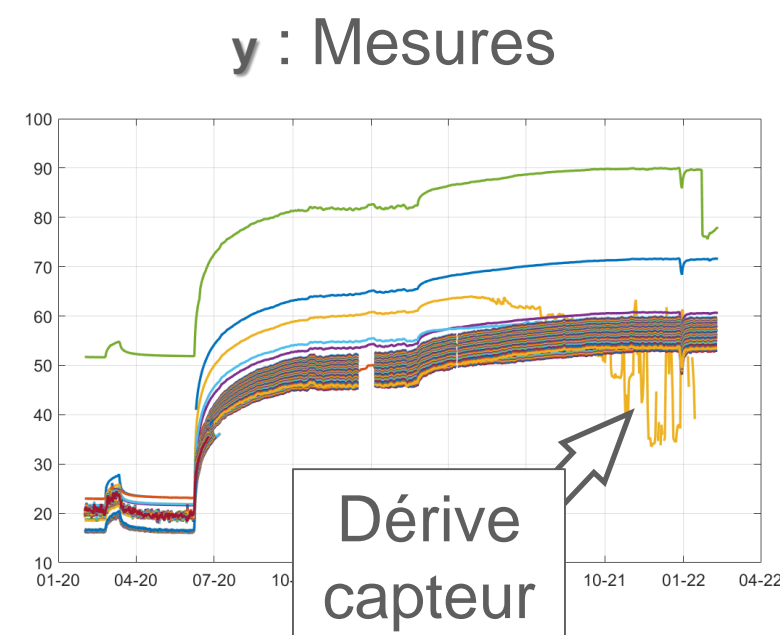
- La dérive de mesure et/ou la défaillance : la donnée disparaît subitement du réseau ;
- le remplacement : une donnée disparaît du réseau et une donnée de remplacement est introduite (au même endroit ou de manière déportée) ;
- l'ajout de nouveaux capteurs : des données additionnelles sont introduites dans le réseau ;
- une absence de mesures (capteurs non présents).



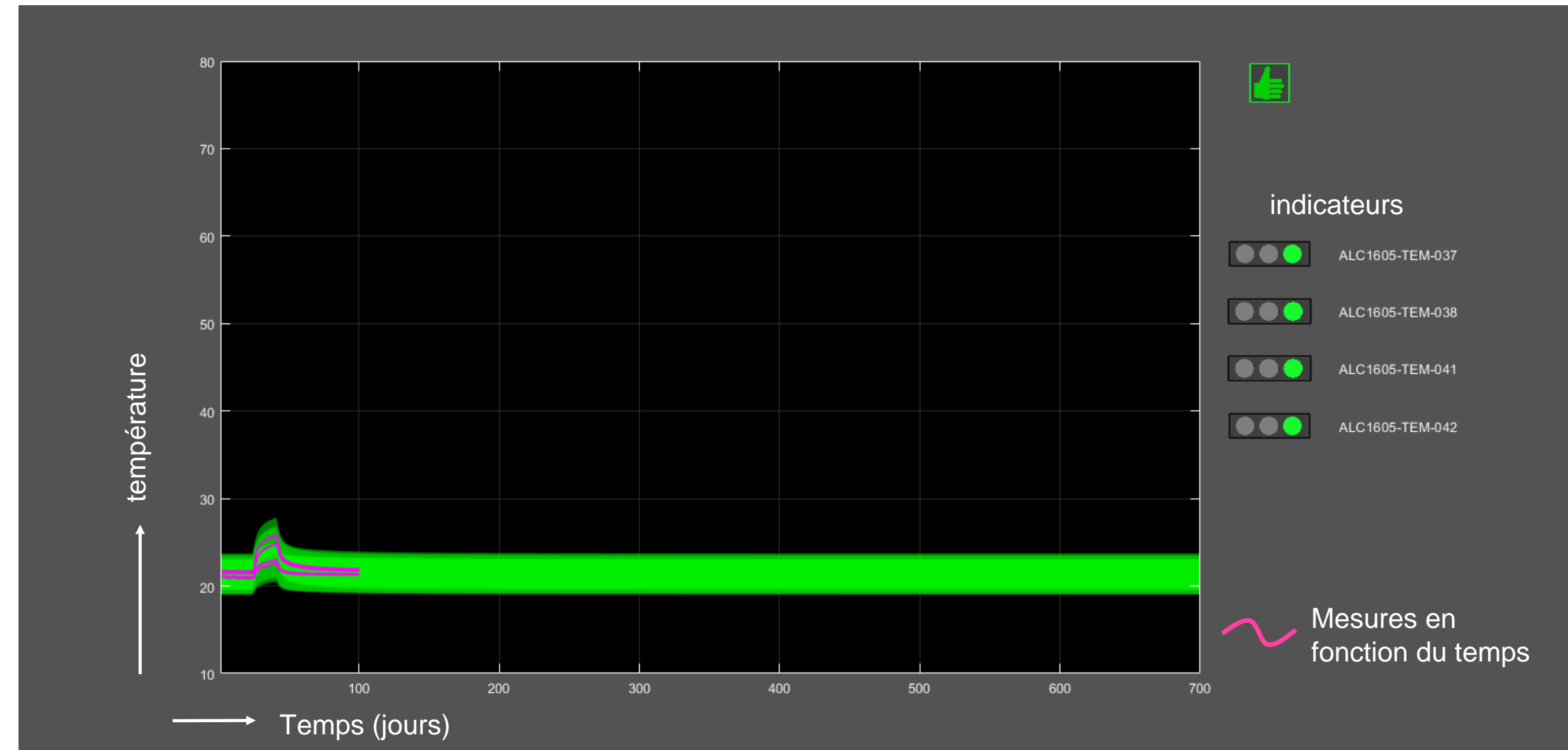
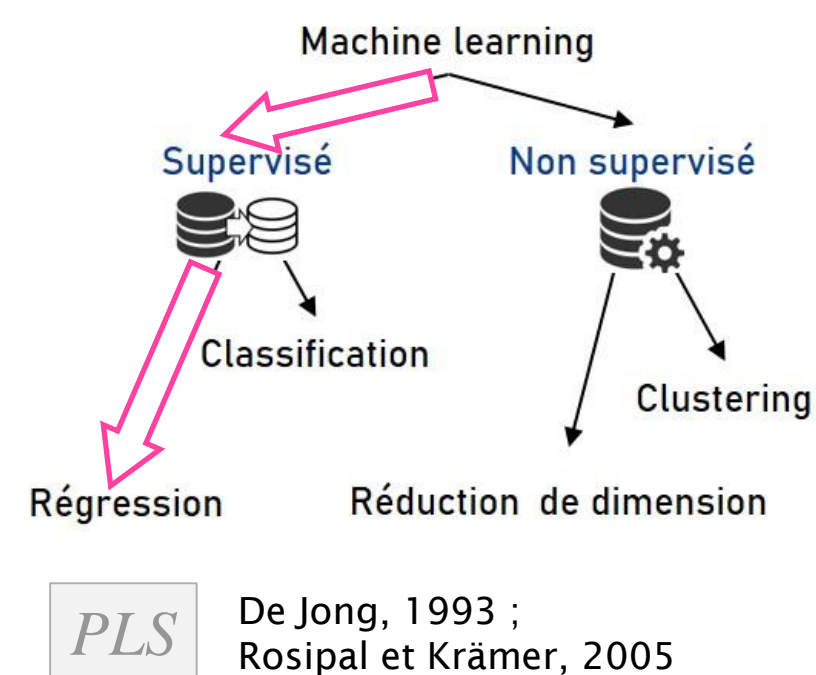
Machine Learning

cas d'usage : calage dynamique de modèles (ici mathématiques) pour la détection de dérives capteurs par apprentissage supervisé

- **L'apprentissage supervisé vise** à déterminer une fonction f permettant d'approximer les relations entre un **premier jeu de données d'entrée x** (des mesures, des images, des modèles...) et un **second jeu de données en sortie y** (des mesures, des paramètres, des noms d'objets...).



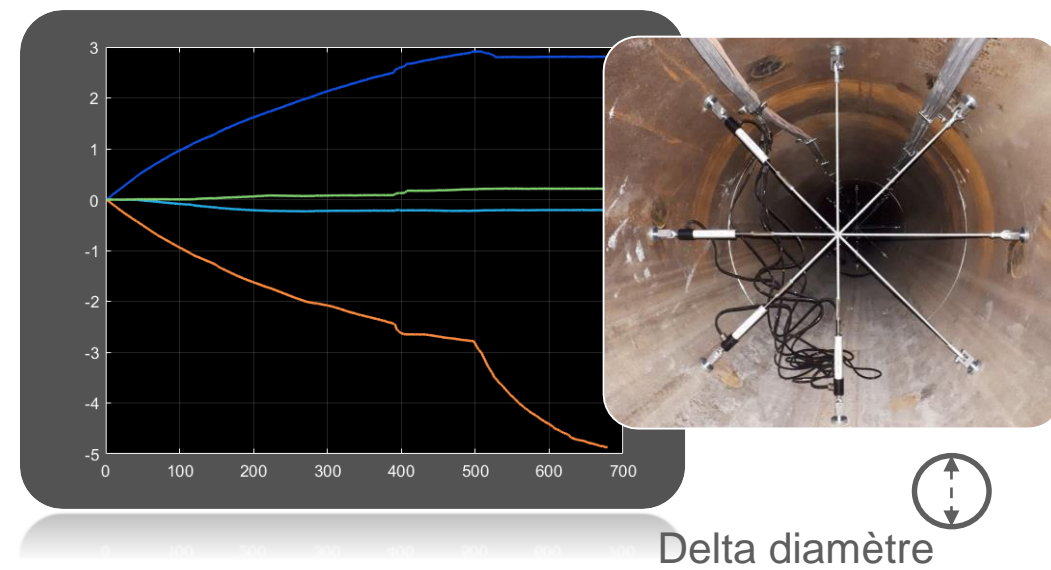
$$y = f(x)$$



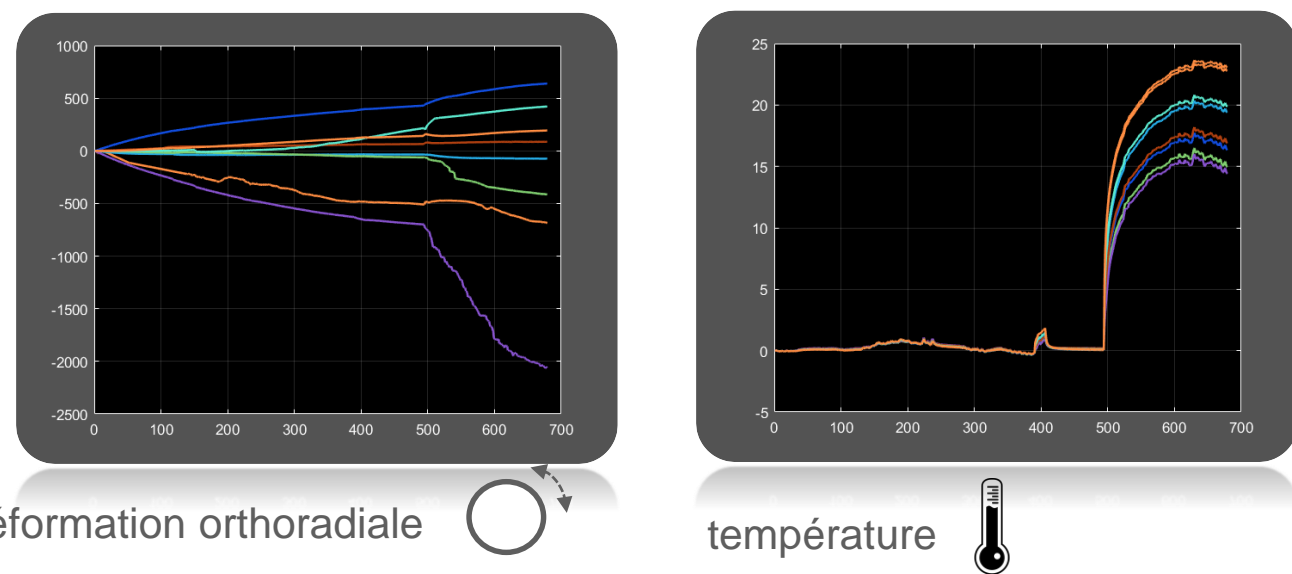
Machine Learning

cas d'usage : prédiction de l'ovalisation à partir de mesures indirectes relation cannes/extensomètres par apprentissage supervisé

y : Mesures des cannes à l'intrados (directes)



x : Mesures extensométriques Et thermiques à l'extrados (indirectes)



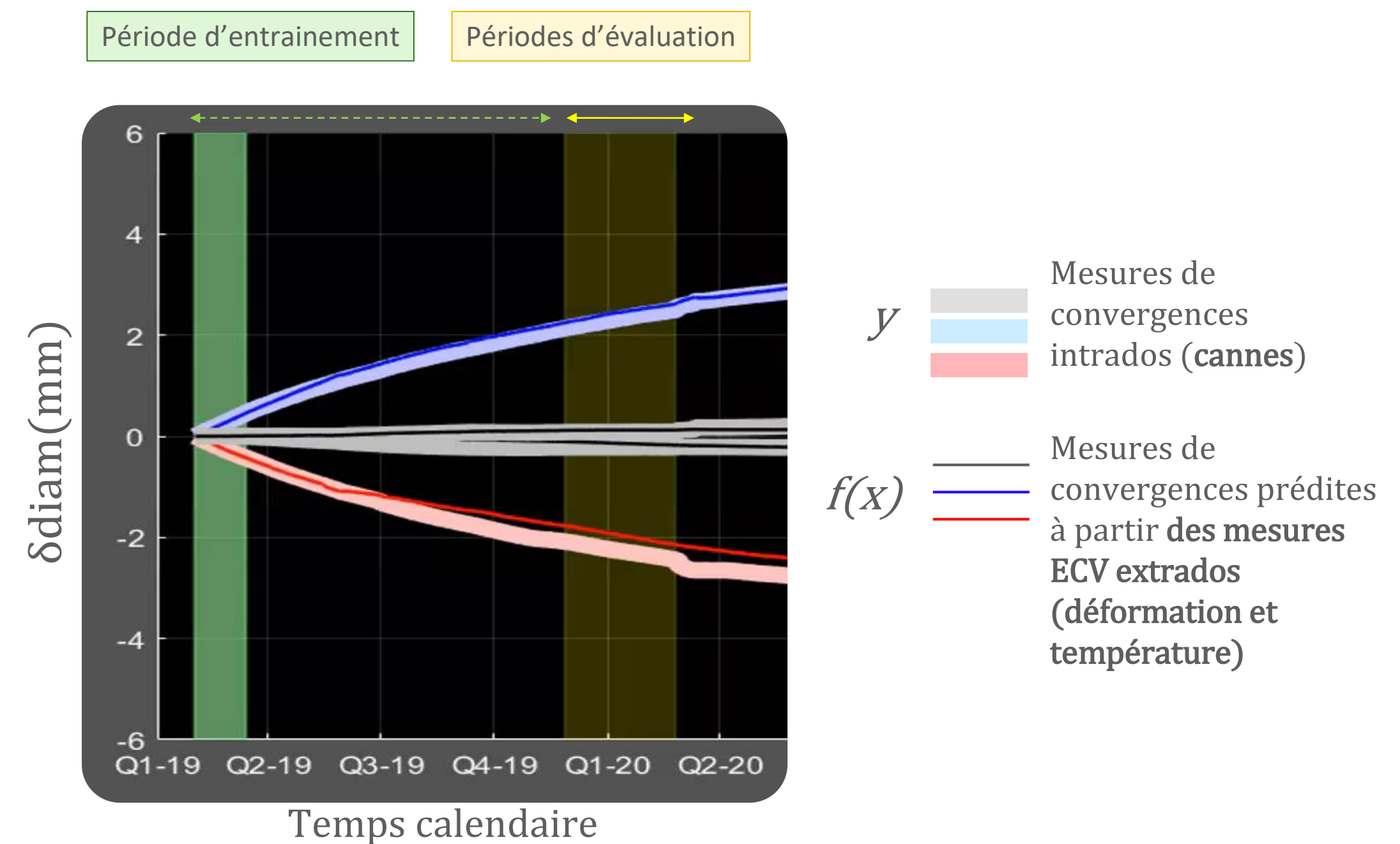
Sur une des sections de l'alvéole, des cannes mesurent directement son ovalisation mais constituent **un obstacle** dans l'alvéole, ce qui n'est pas le cas des ECV.

Après l'apprentissage, les cannes peuvent être temporairement ou définitivement retirées

$$y = f(x)$$

Si une donnée disparaît dans x , le machine learning ne peut plus prédire y : il faudrait réentraîner

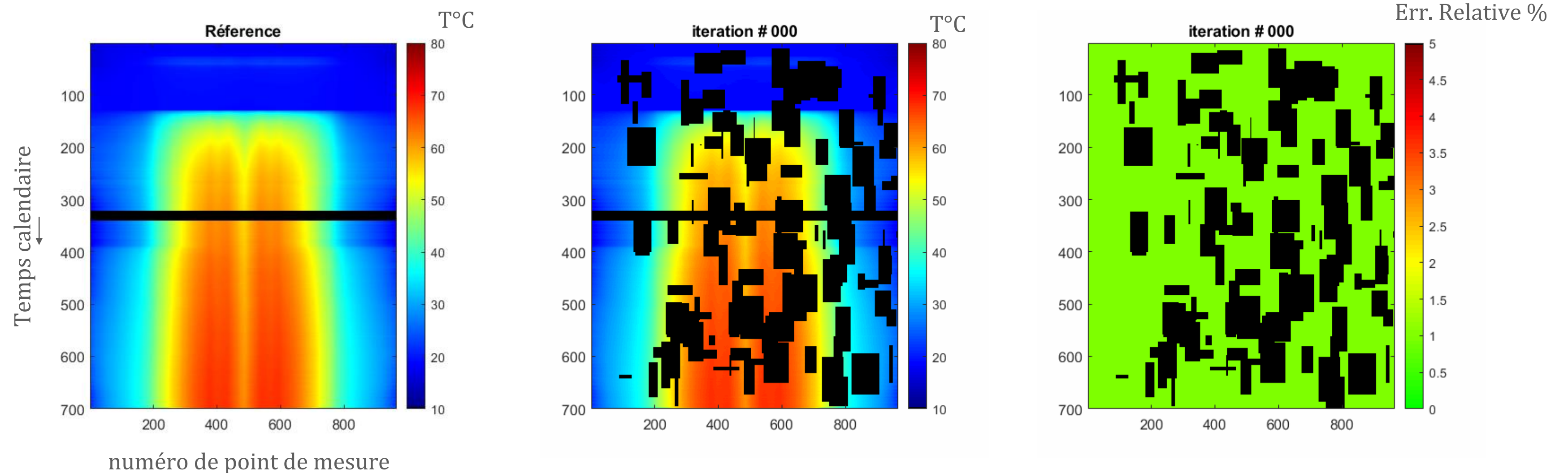
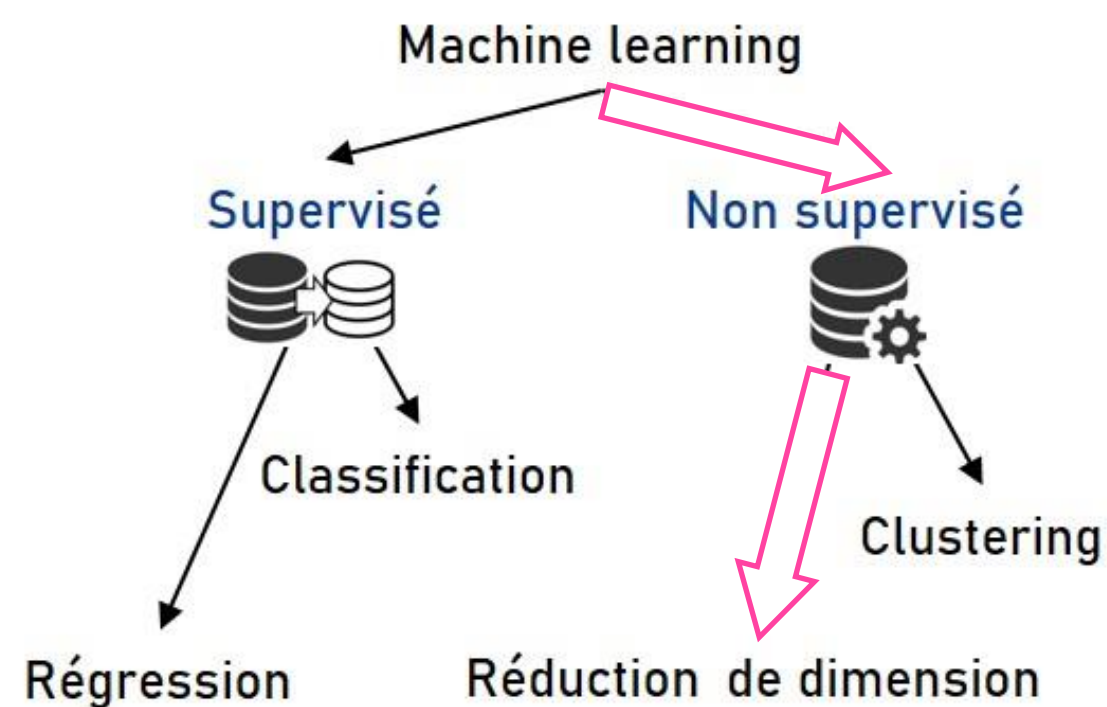
Apprentissage non-supervisé



Machine Learning

Reconstruction de données manquantes par apprentissage non-supervisé : R-PCA

- **L'apprentissage non supervisé** : les algorithmes sont laissés à leurs propres mécanismes (essentiellement des projections associées à des réductions de dimension) pour déterminer des structures intéressantes dans **le jeu de données d'entrée**.



SVT Cai et al. 2010

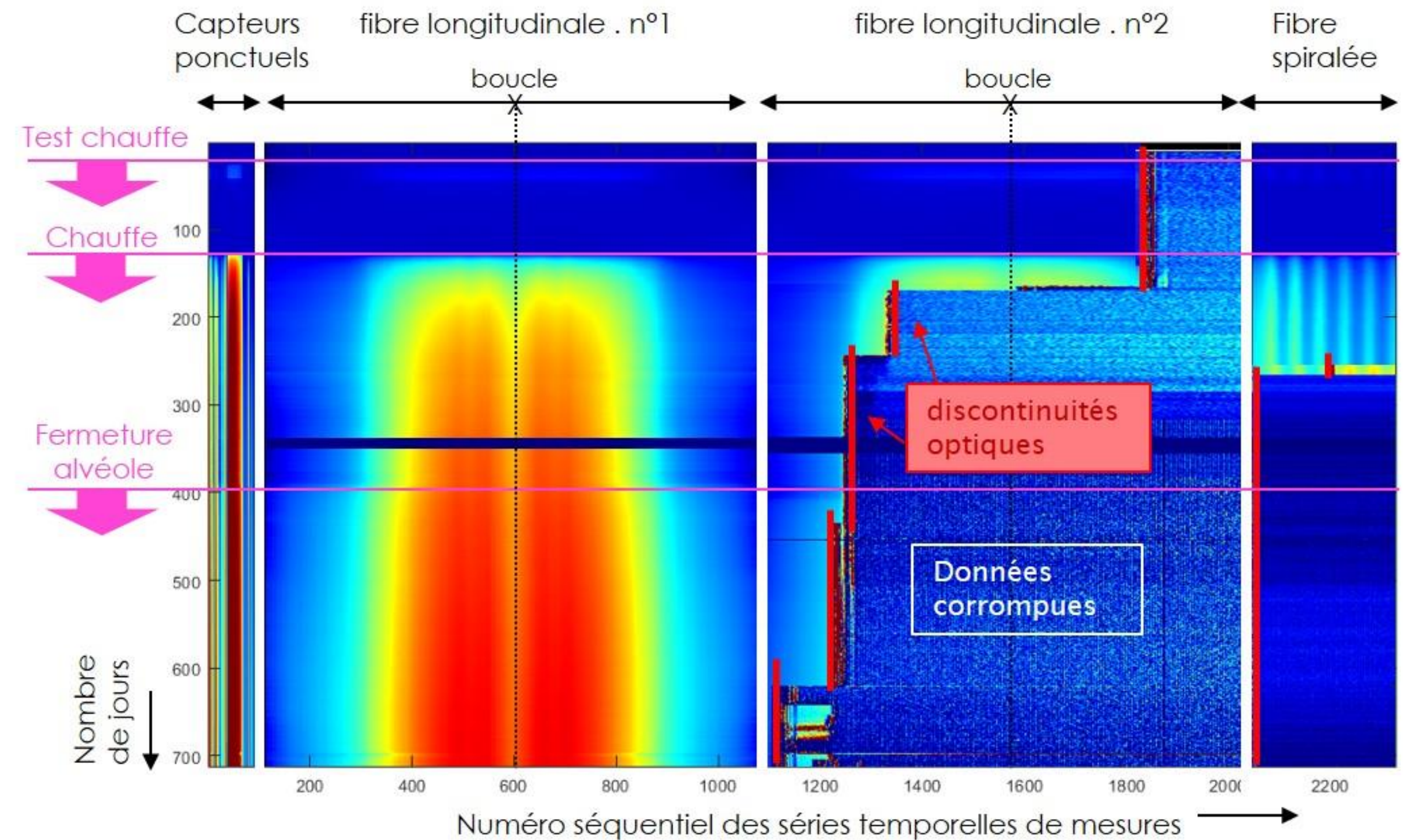
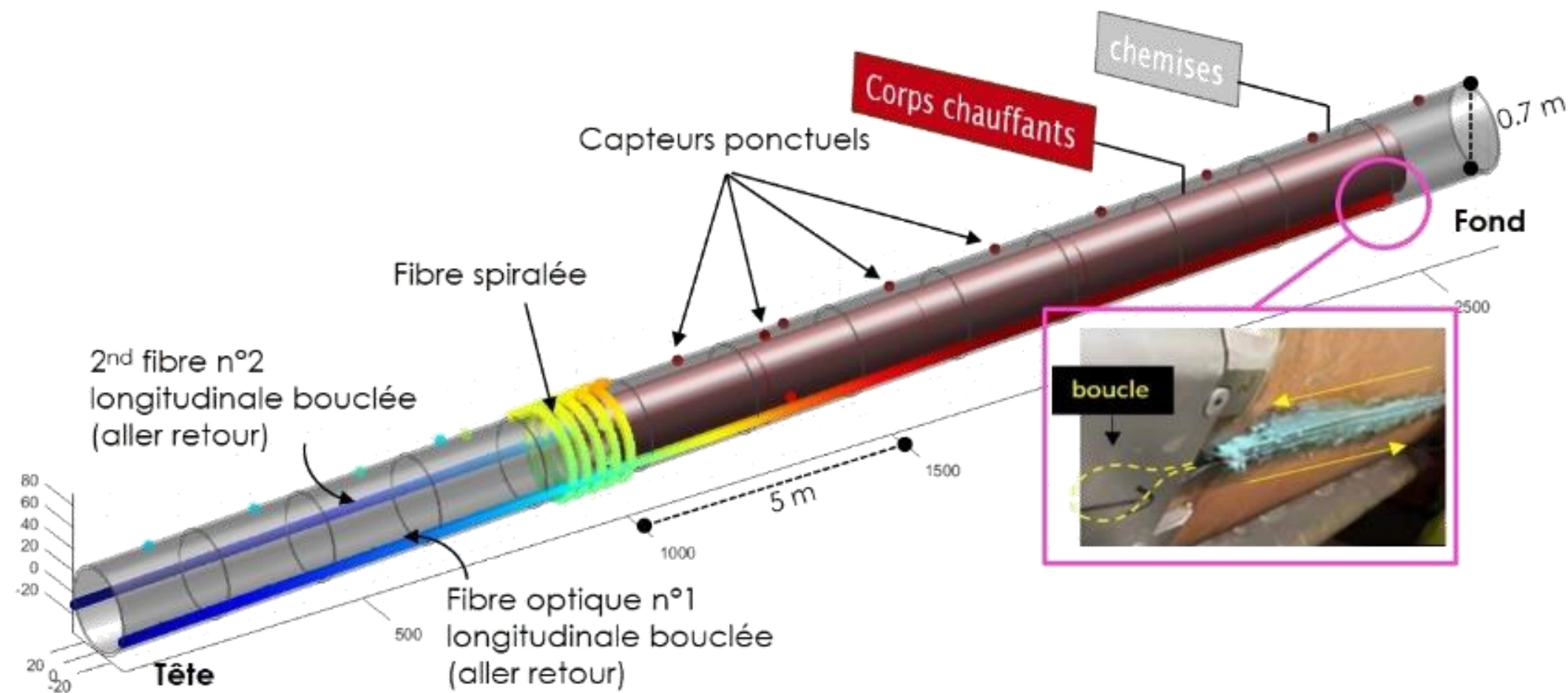
R-PCA Candès et al. 2009

- La **Robust-PCA** décompose itérativement des données d'entrée comme la somme d'une matrice de rang faible et d'une matrice sparse

➔ Sur ces données « trouées artificiellement », elle permet de bien reconstruire les données manquantes en se basant sur la cohérence d'ensemble

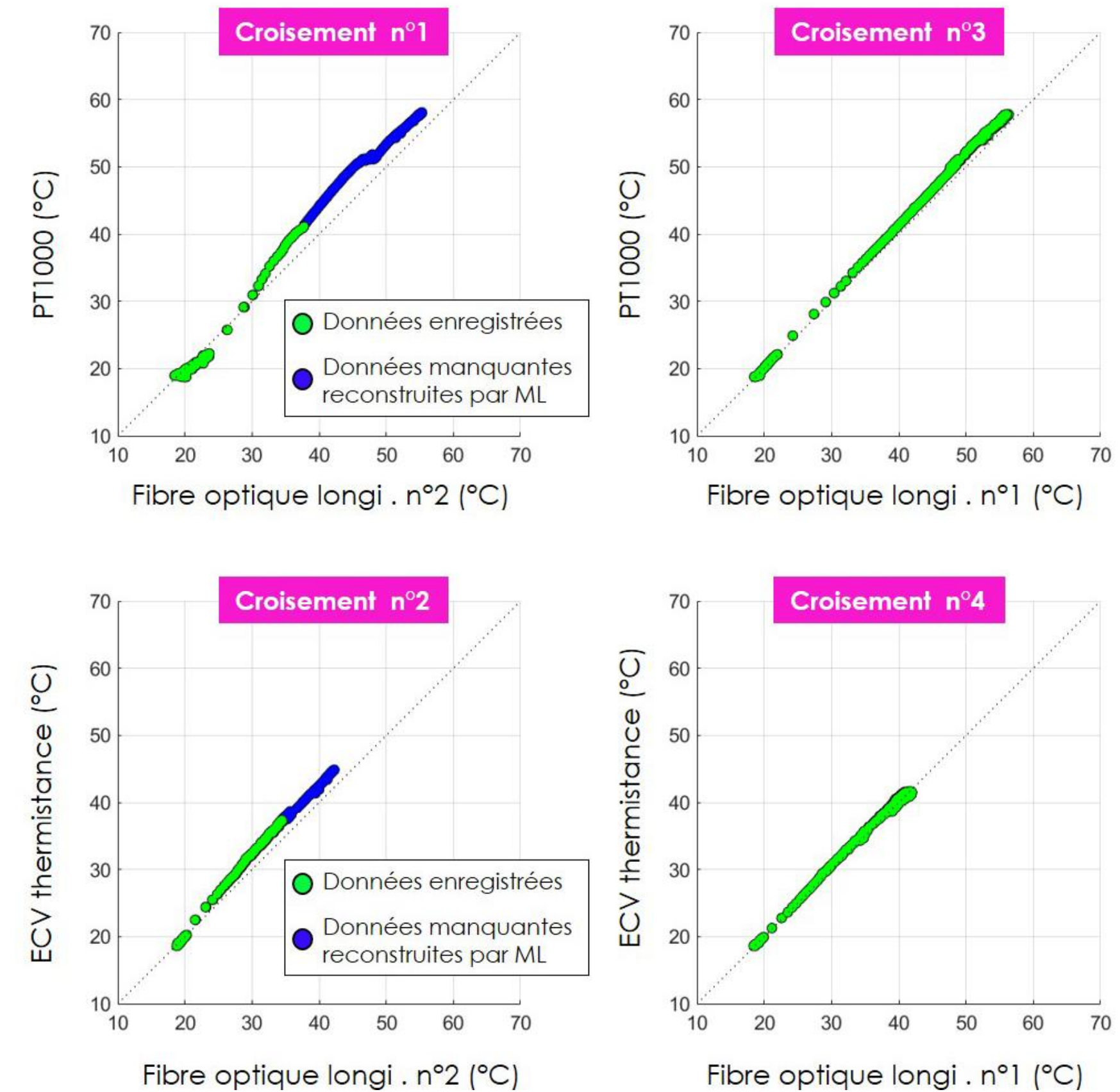
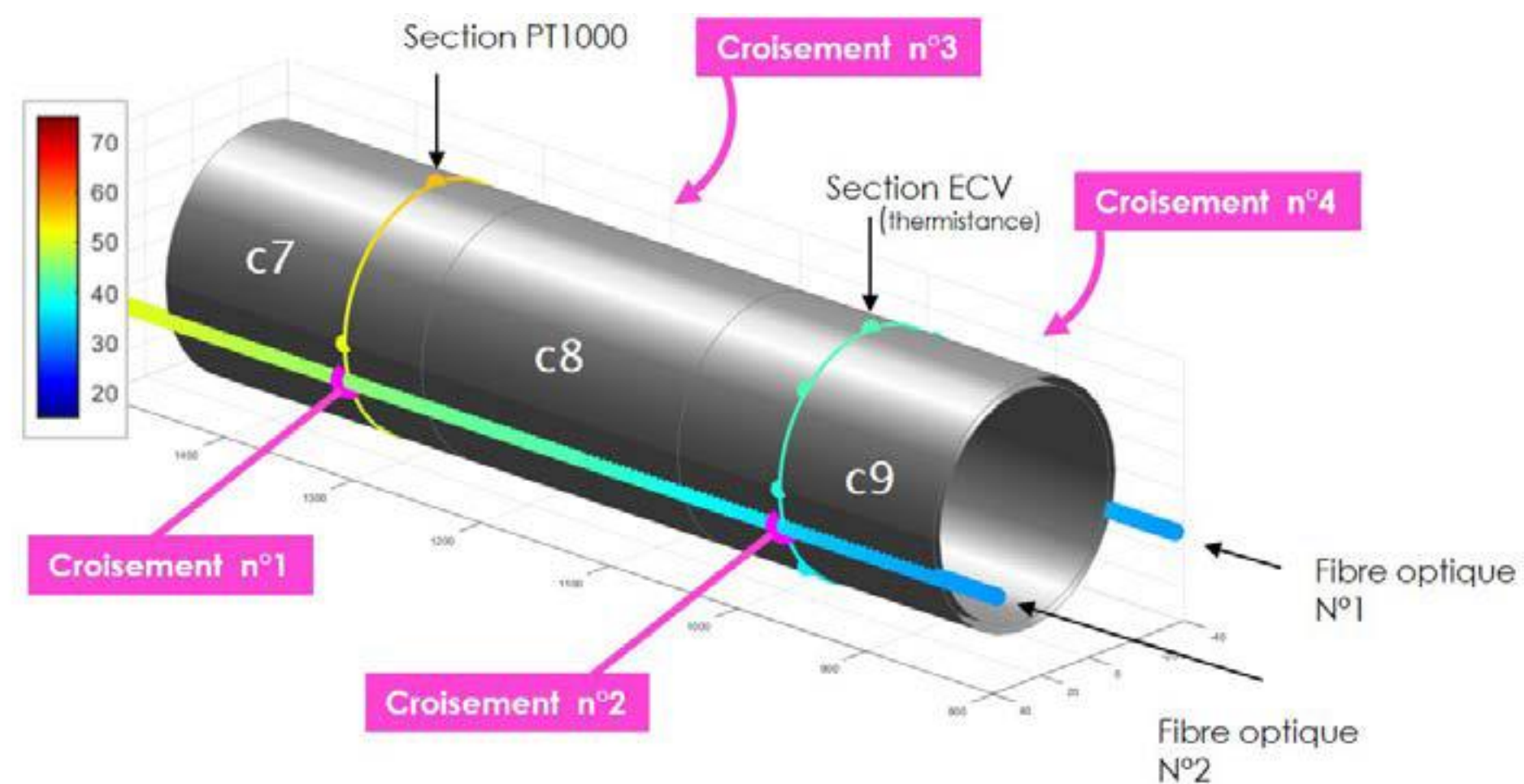
Machine Learning

Reconstruction de données manquantes par apprentissage non-supervisé : R-PCA



Machine Learning

Contrôle des données reconstruites



➔ Les données défailtantes de la FO et reconstruites par la R-PCA sont en cohérence avec les données des capteurs ponctuels présents au niveau des croisements

Conclusion

Structuration des données et Machine Learning – références biblio

➤ **La structuration des données** à l'Andra est en adéquation avec les spécificités du stockage, elle permet de gagner en efficacité sur toute la chaîne de donnée.

➔ Ouvrages hors normes : multi-échelle ;

➔ Besoins utilisateurs au niveau des métiers « sur le terrain » ;

➔ Au niveau projet

➤ **Le Machine Learning** propose un panel d'outils intéressants pour assurer et maintenir la cohérence d'ensemble des données vis-à-vis des échelles spécifiques du stockage, des phénomènes mesurés et sur des temporalités pluriséculaires.

➔ Un champ de nouvelles pratiques mais une utilisation progressive, maîtrisée et raisonnée

➤ Références

De Jong, S. (1993). SIMPLS: an alternative approach to partial least squares regression. *Chemometrics and intelligent laboratory systems*, 18(3), 251-263.

Rosipal, R., & Krämer, N. (2005, February). Overview and recent advances in partial least squares. In *International Statistical and Optimization Perspectives Workshop "Subspace, Latent Structure and Feature Selection"* (pp. 34-51). Springer, Berlin, Heidelberg.

J.-F. Cai, E.-J. Candès, Z. Shen (2010). A Singular Value Thresholding Algorithm for Matrix Completion. *SIAM Journal on Optimization*. 20. 1956-1982. 10.1137/080738970

E. J. Candès, X. Li Ma, J. Wright (2009). Robust Principal Component Analysis.