Monitoring cyberthreats in railway systems: A hybrid framework for detecting stealthy data tampering attacks

Société d'Automatique, de Génie Industriel & de Productique

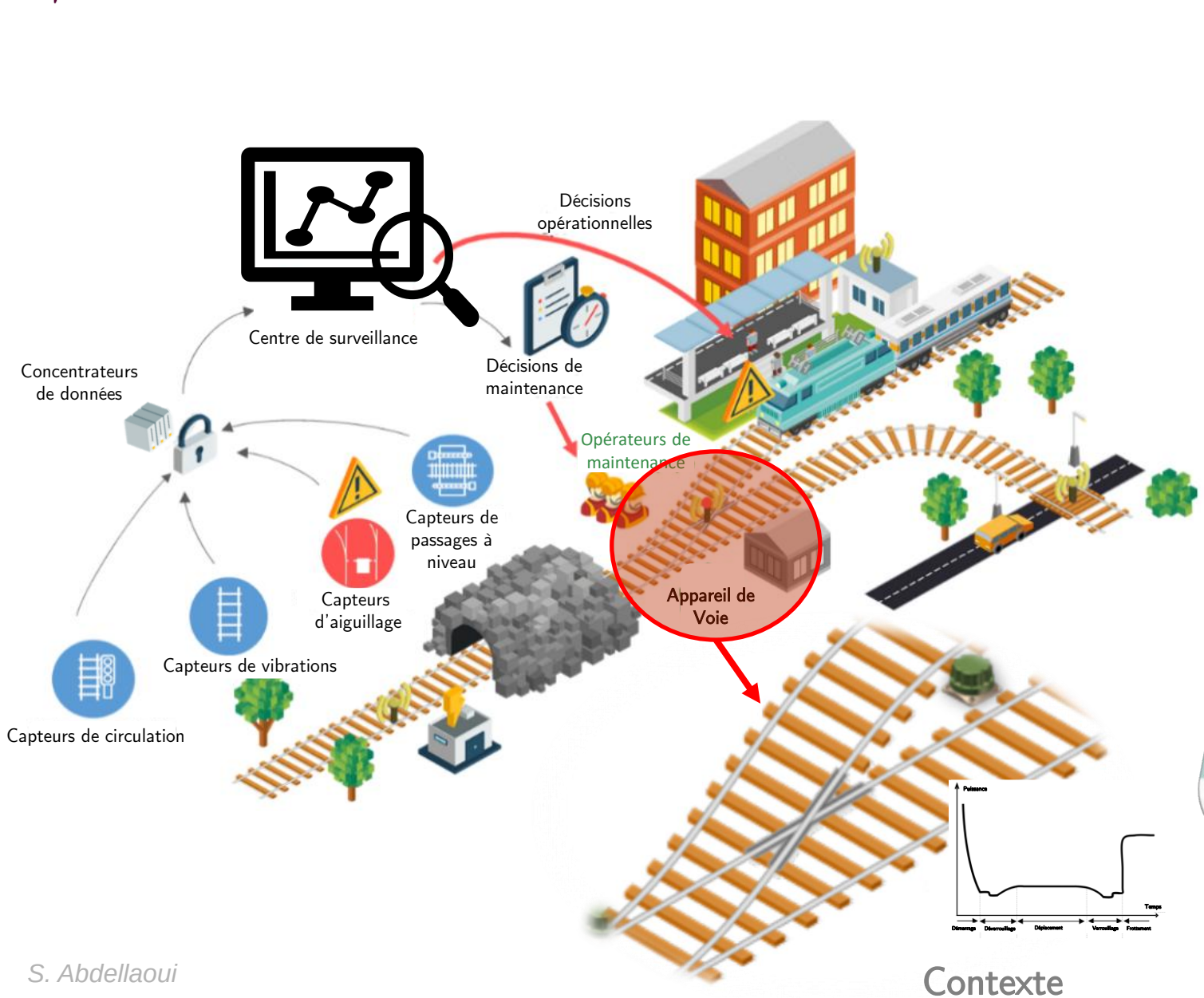
Sara ABDELLAOUI, Emil DUMITRESCU, Cédric ESCUDERO, Eric ZAMAÏ

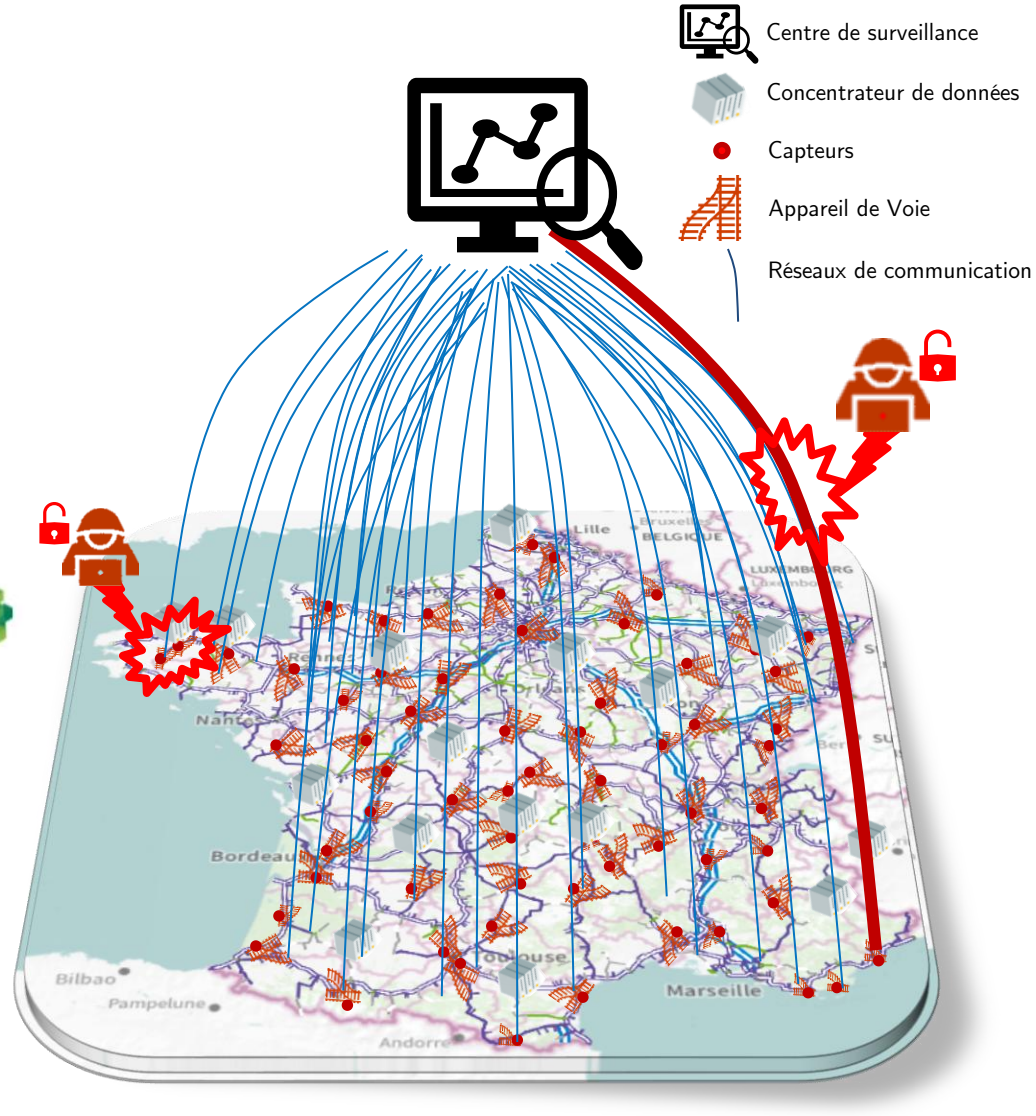
Département Automatique pour l'Ingénierie des Systèmes (**AIS**), **AMPERE, INSA Lyon**

Équipe *Resilient Industrial Cyber Systems* (**RICS**)

saraabdellaoui1805@gmail.com

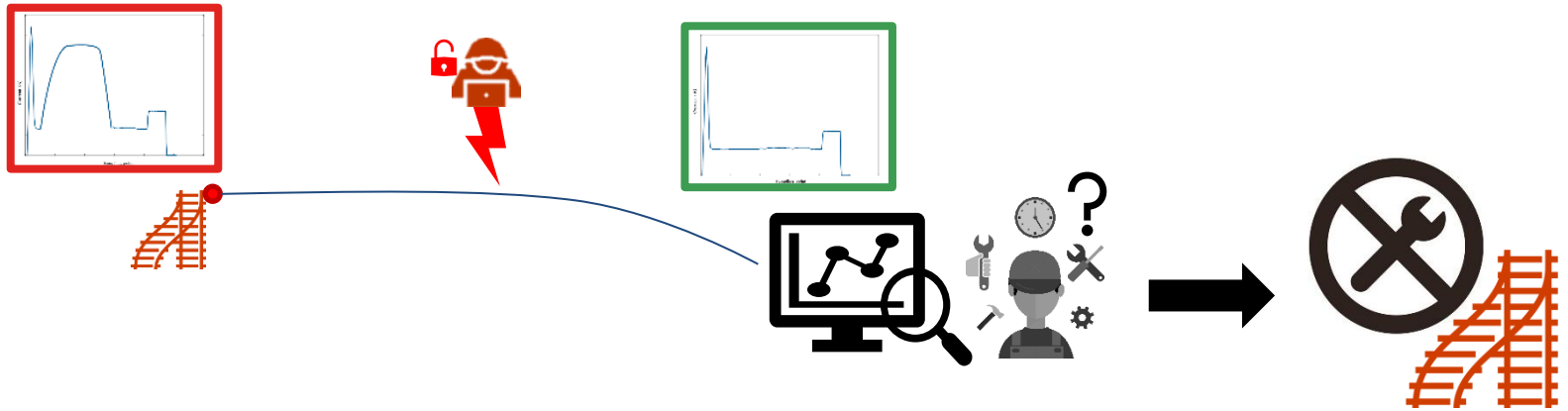
Centre de surveillance ferroviaire



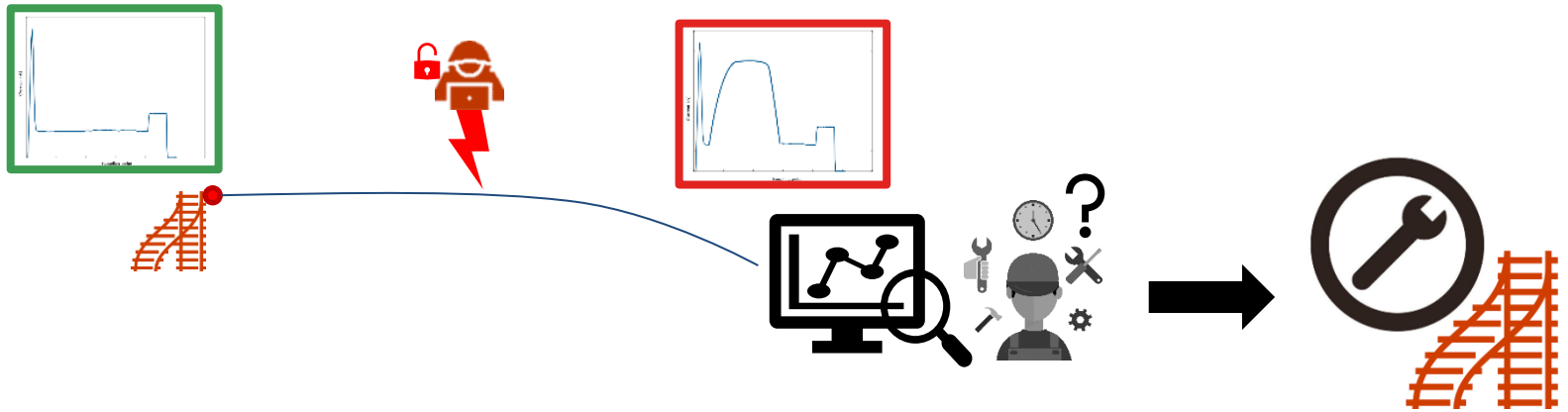


Corruption et manipulation des données de surveillance

- Attaque par masquage de défauts



- Attaque par injection de défauts



Hypothèses

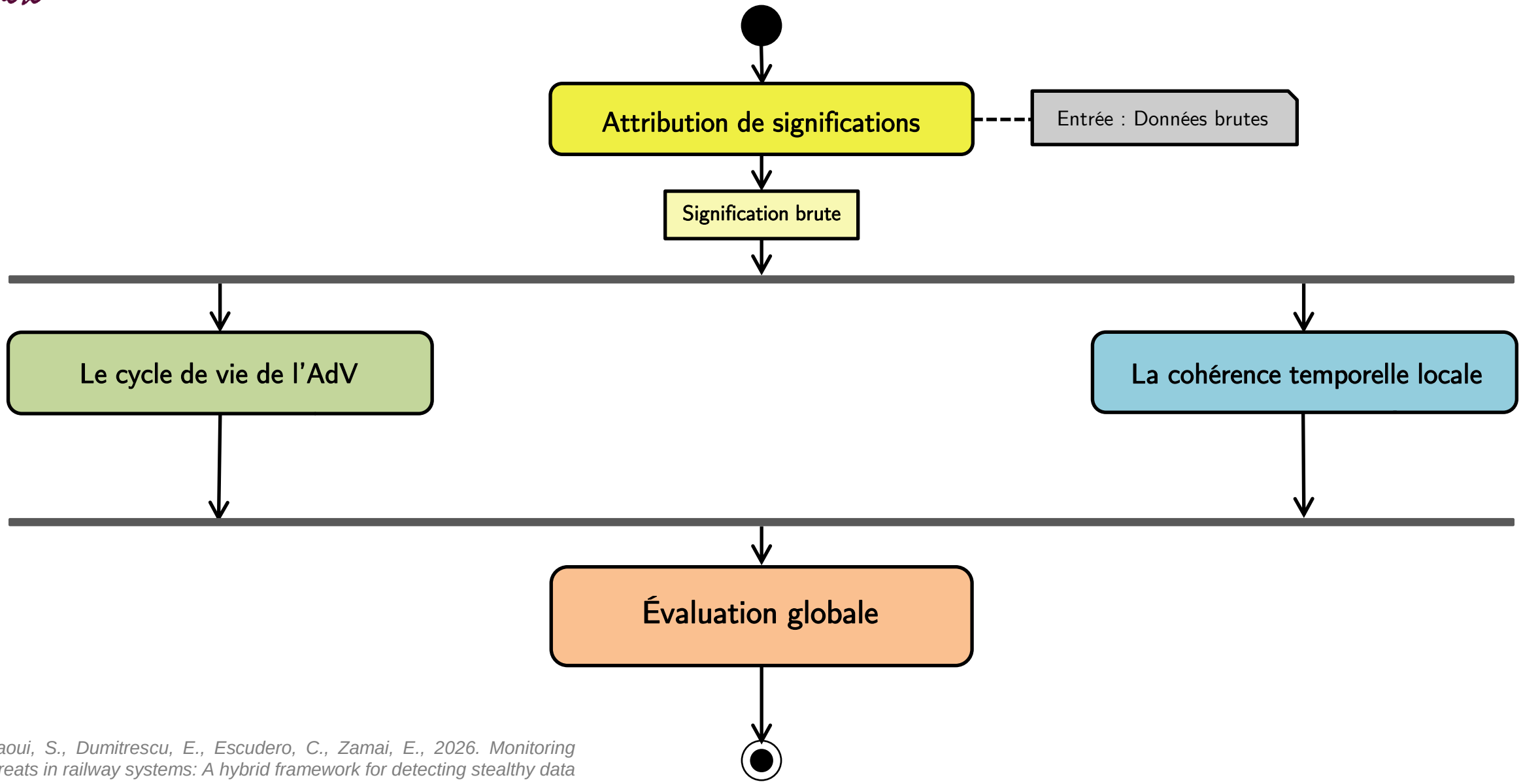
Connaissance et capacités des attaquants :

- Ressources importantes
- Bonne connaissance du système ferroviaire
- Connaissance limitée du cycle de vie de l'AdV

Cible et impact attendu :

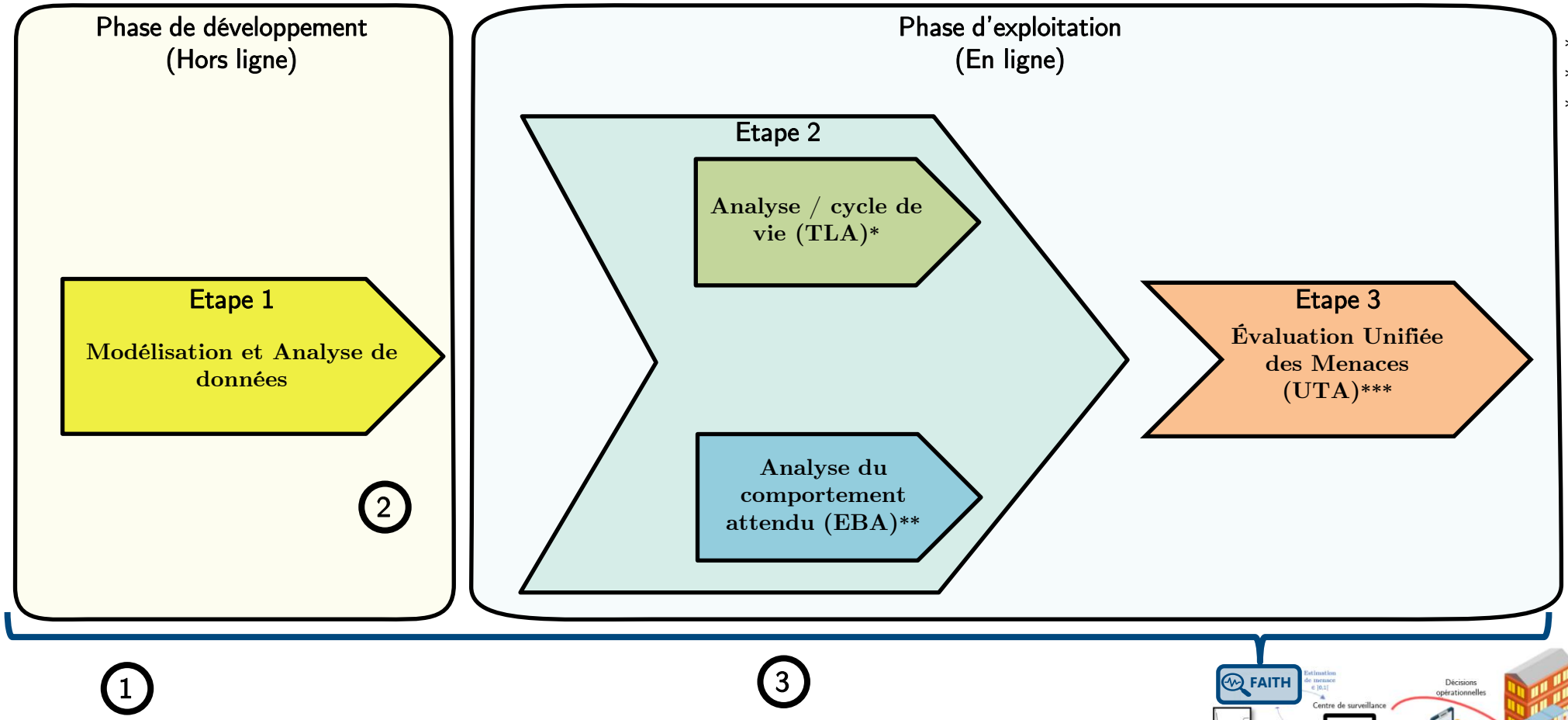
- Manipulations des données de manœuvre d'aiguillage
- Manipulations ciblées : courbes de puissance
- Injection de scenarios normaux ou de défauts

Principe de détection : contextualisation*

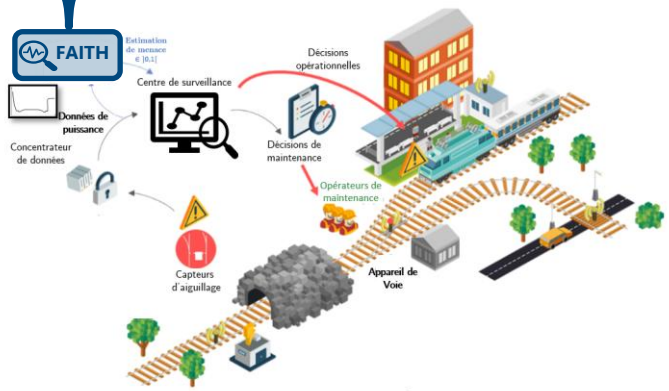


*Abdellaoui, S., Dumitrescu, E., Escudero, C., Zamai, E., 2026. Monitoring cyberthreats in railway systems: A hybrid framework for detecting stealthy data tampering attacks. Reliability Engineering & System Safety 266, 111747. https://doi.org/10.1016/j.ress.2025.111747

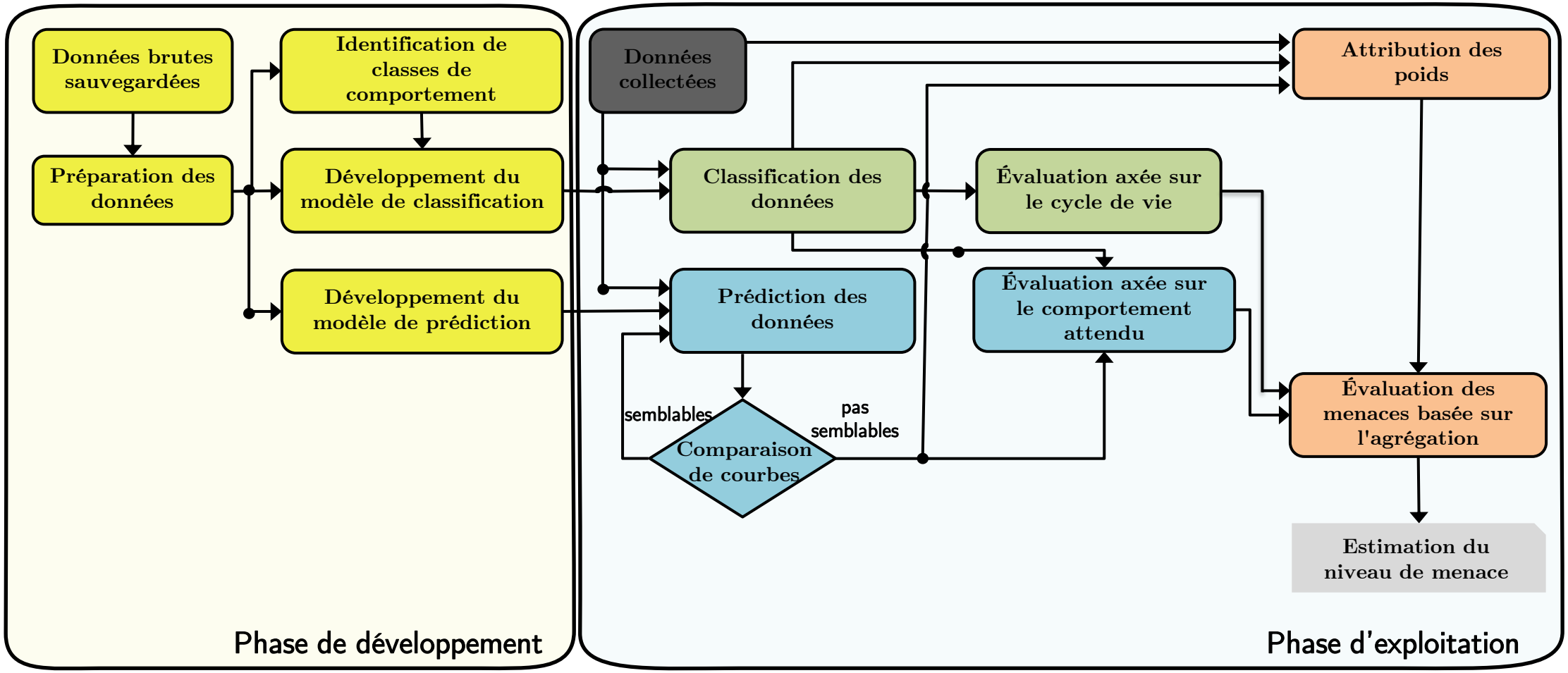
Méthodologie : de la modélisation de données à la détection d'attaques



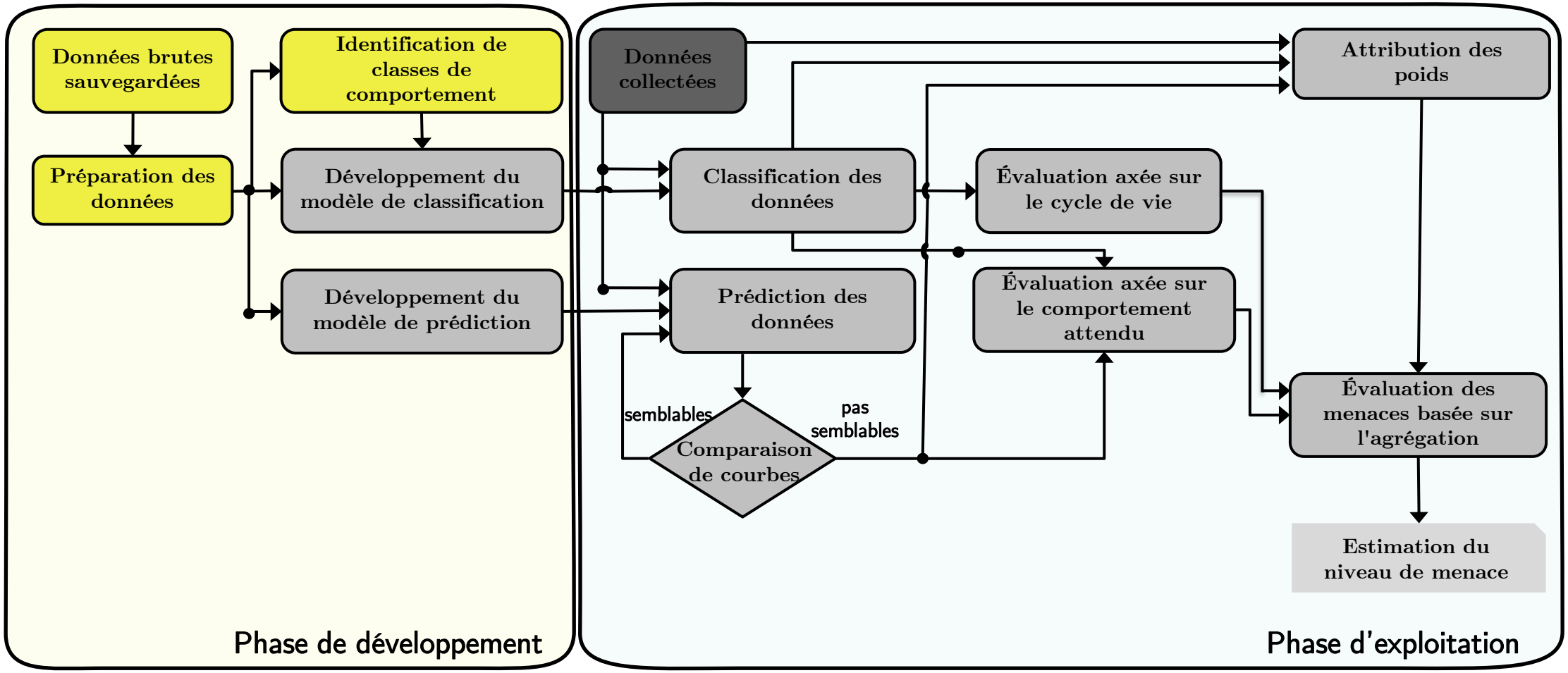
* Turnout Lifecycle Assessment
** Expected Behavior Assessment
*** Unified Behavior Assessment



Méthodologie : de la modélisation de données à la détection d'attaques



Méthodologie : de la modélisation de données à la détection d'attaques

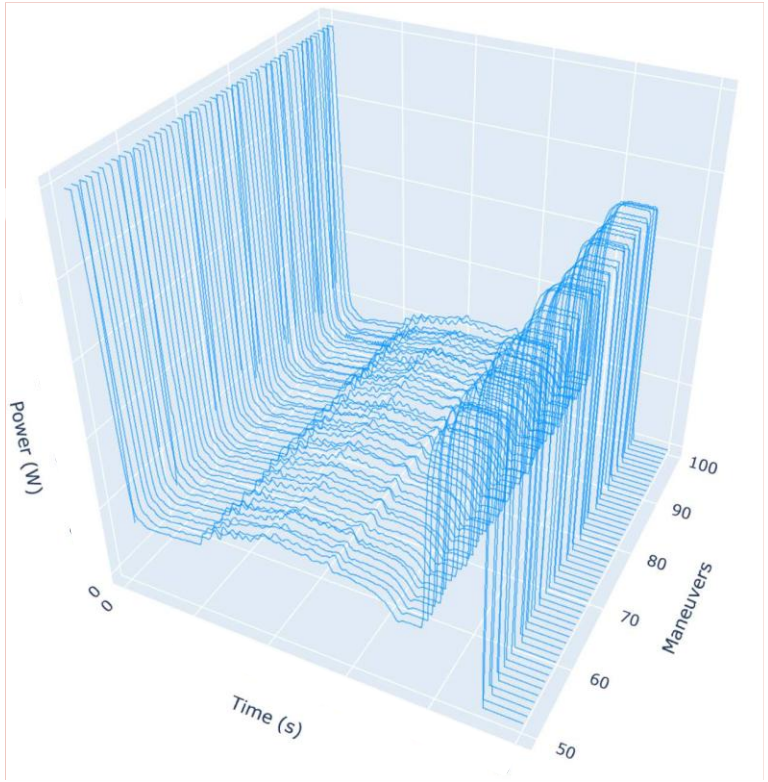


Attribution de significations : classes de comportements

Comportements normaux

$$E_n = \{e_{n_e}, e_{n_w}\}$$

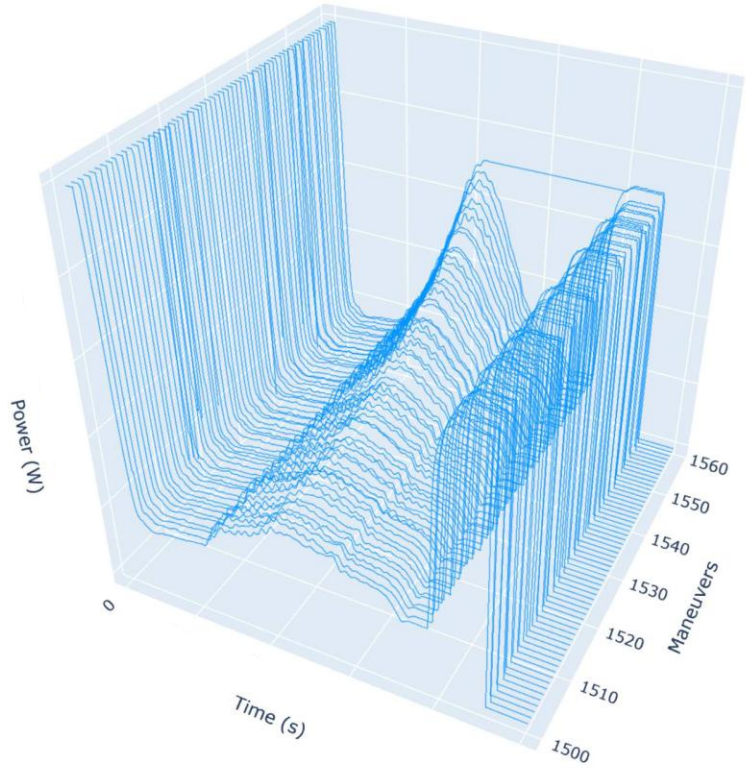
- **Fonctionnement normal**
- Vieillissement



Comportements anormaux

$$E_a = \{e_{a_p}, e_{a_m}, e_{a_s}\}$$

- **Détériorations progressives**
- Anomalies mineures
- Défaillances soudaines



Données et attaques injectées

- **Quantité** : 3879 courbes. - **Période** : 14 mois. - **Ensemble d'entraînement** : 2480 courbes. - **Ensemble de test** : 1399 courbes

Attaques injectées

A_1 : Défaillance soudaine

A_2 : Comportement de vieillissement

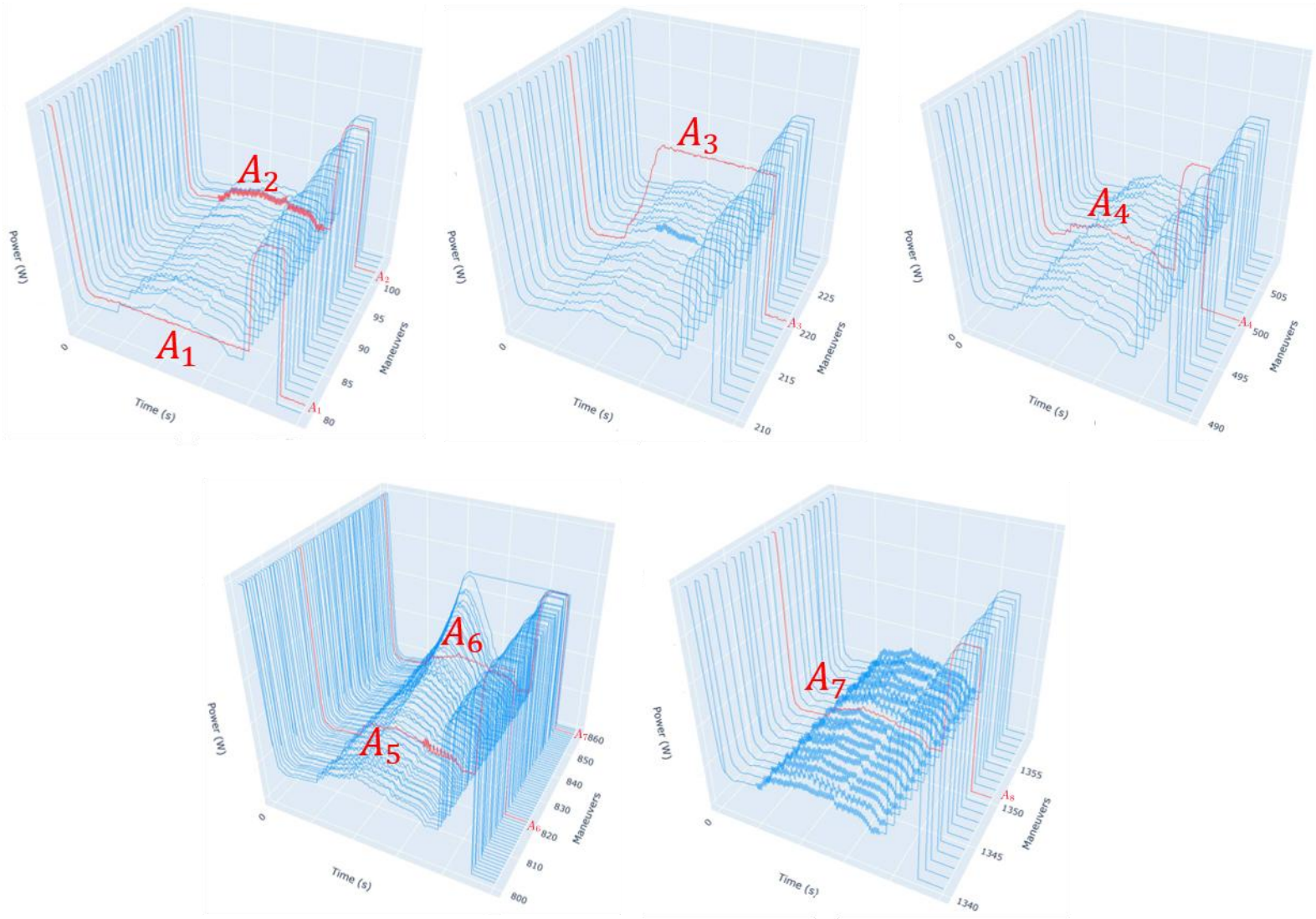
A_3 : Détériorations progressives

A_4 : Défaillance soudaine

A_5 : Anomalie mineure

A_6 : Comportement sans défaut

A_7 : Comportement sans défaut

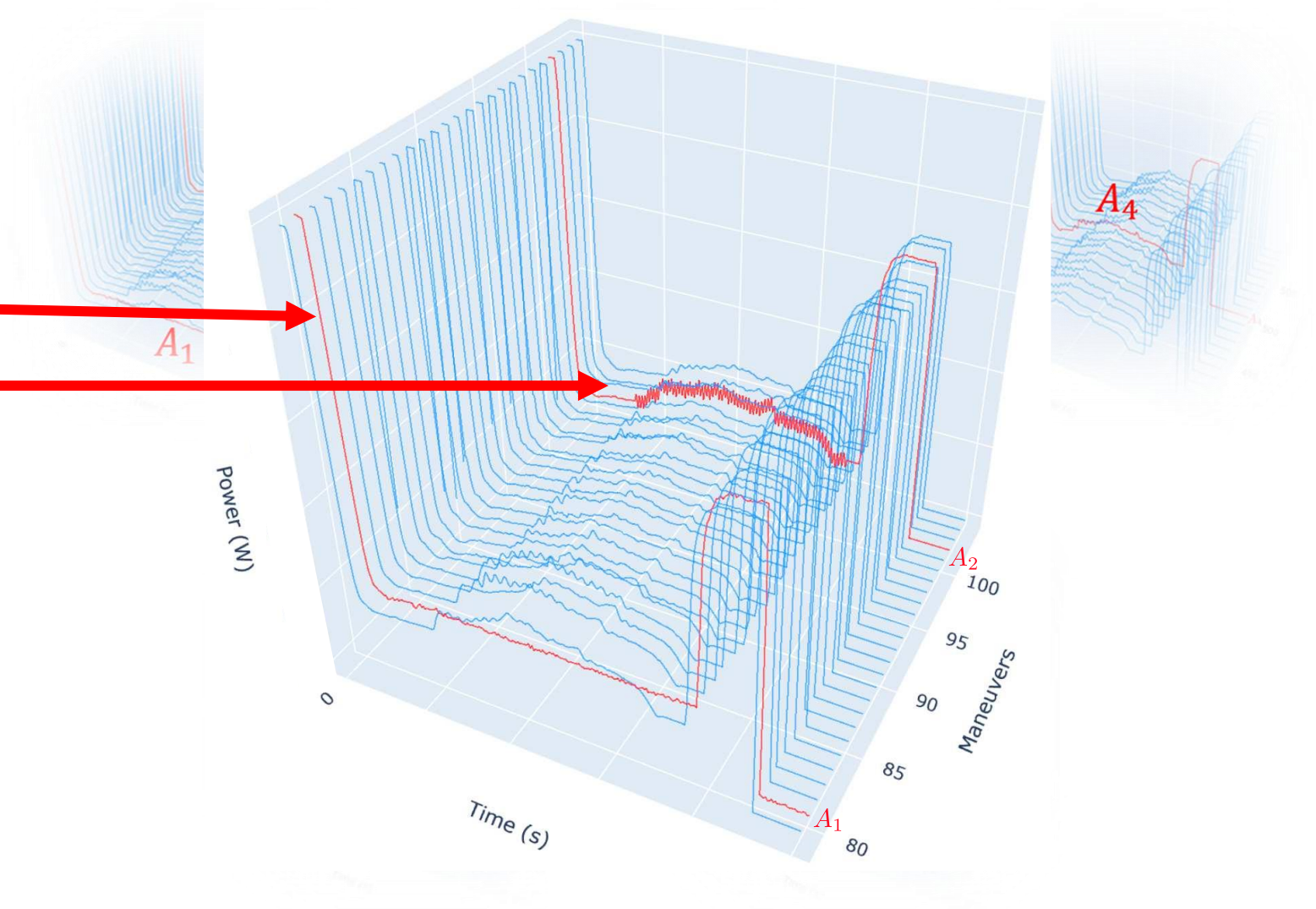


Données et attaques injectées

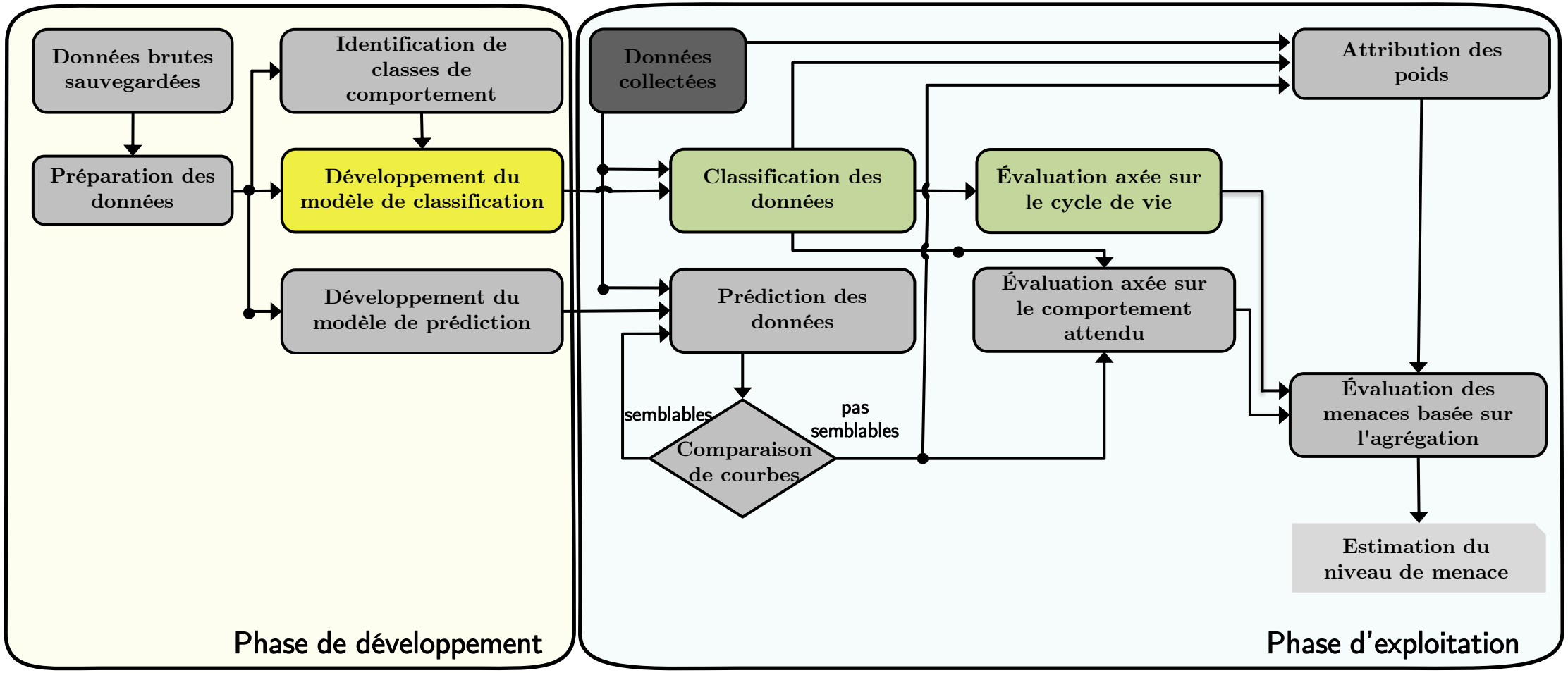
- **Quantité** : 3879 courbes. - **Période** : 14 mois. - **Ensemble d'entraînement** : 2480 courbes. - **Ensemble de test** : 1399 courbes

Attaques injectées

- A_1 : **Défaillance soudaine**
- A_2 : **Comportement de vieillissement**
- A_3 : Détériorations progressives
- A_4 : Défaillance soudaine
- A_5 : Anomalie mineure
- A_6 : Comportement sans défaut
- A_7 : Comportement sans défaut

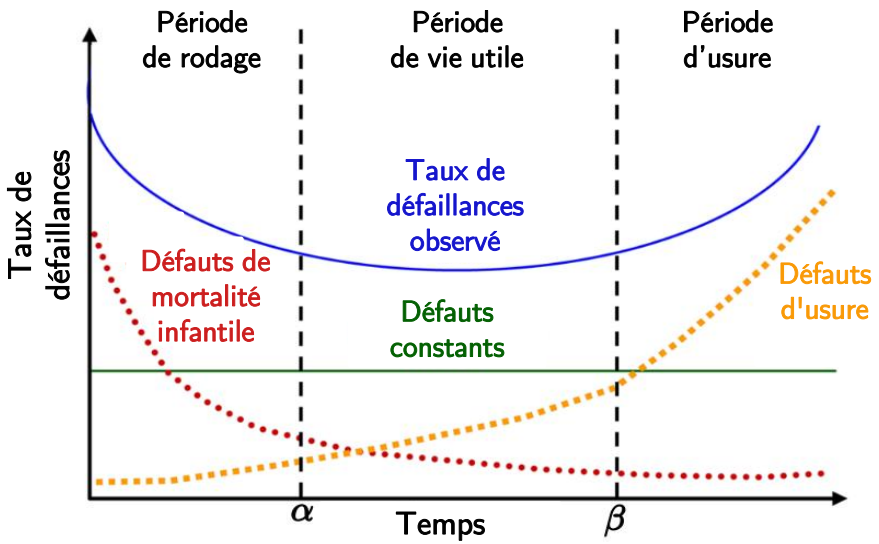


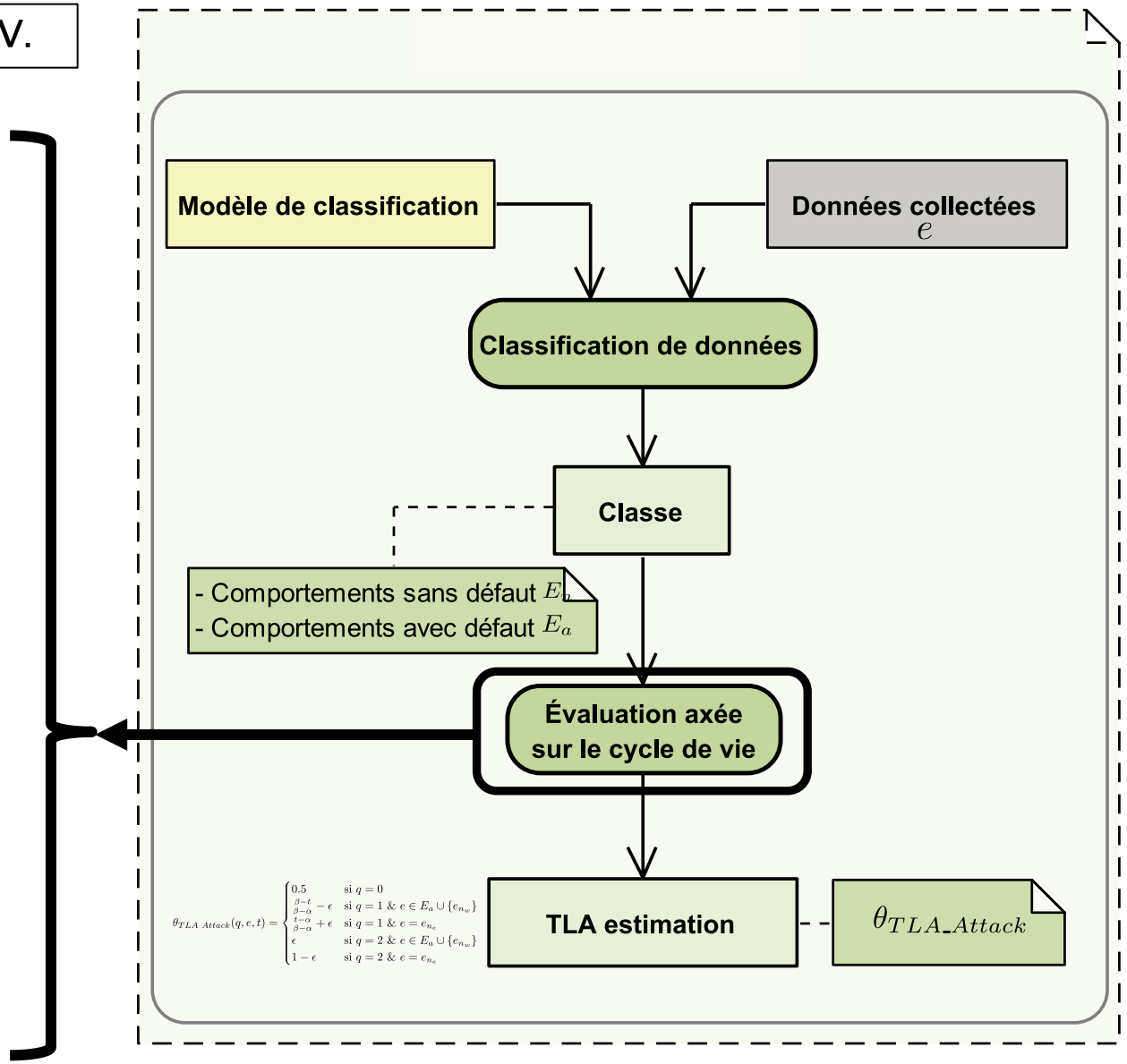
Analyse par rapport au Cycle de Vie (TLA)



Analyse par rapport au Cycle de Vie (TLA)

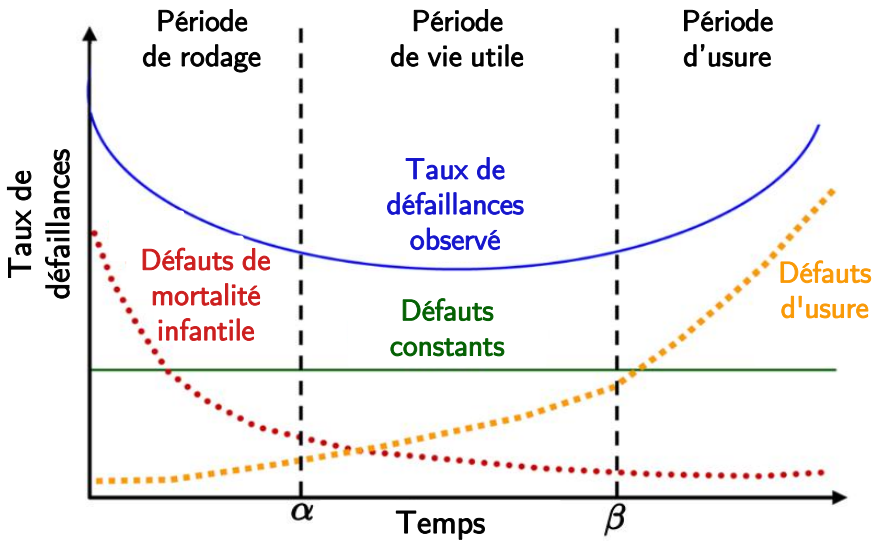
La signification d'une courbe dépend de l'état réel de l'AdV.





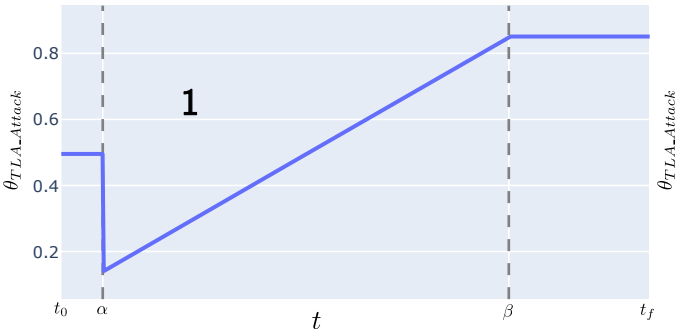
Analyse par rapport au Cycle de Vie (TLA)

La signification d'une courbe dépend de l'état réel de l'AdV.



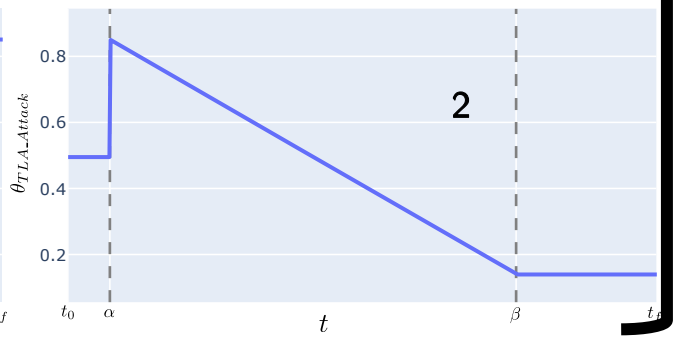
$e = e_{n_e}$

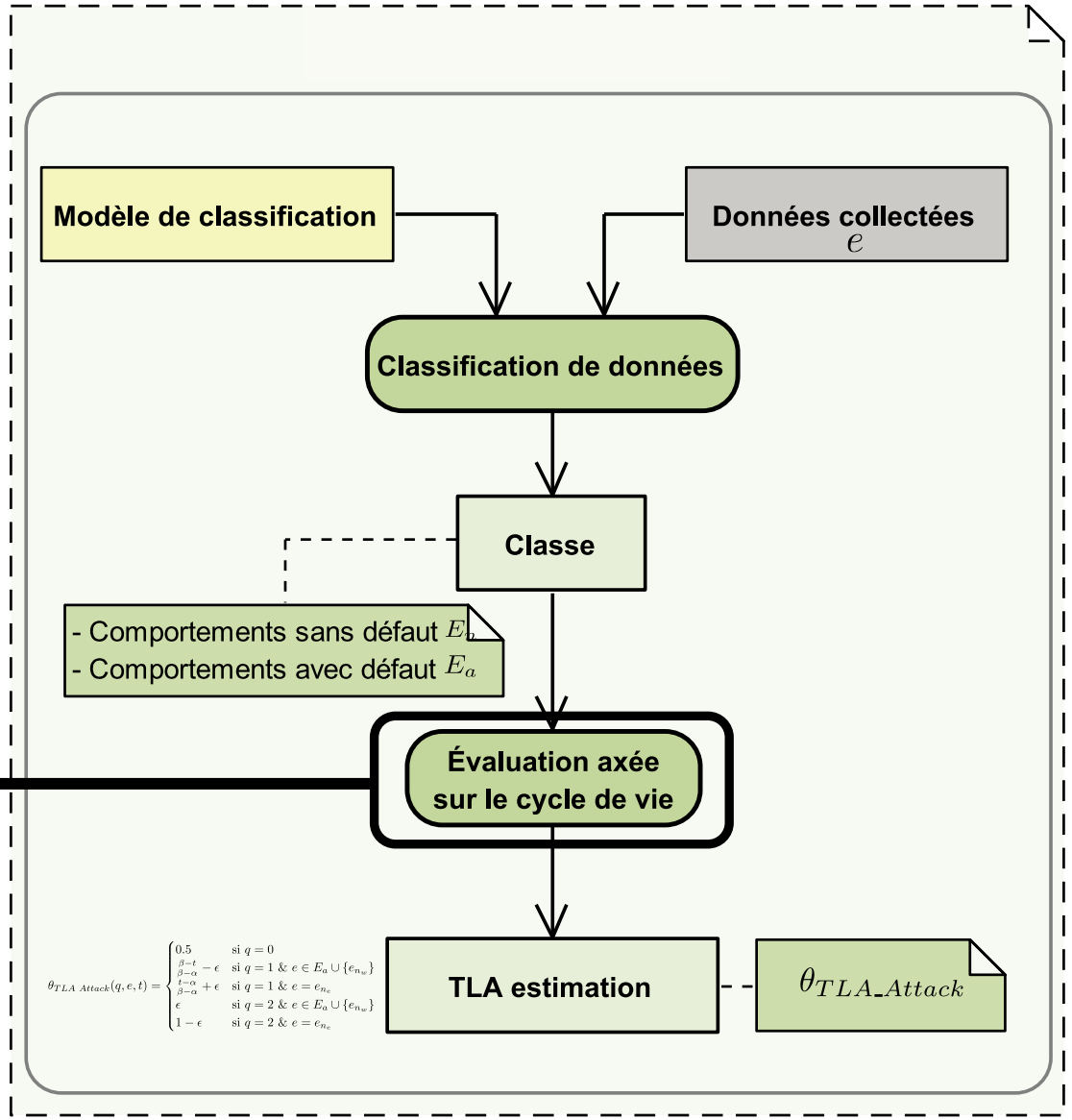
Niveau de suspicion



$e \in E_a \cup \{e_{n_w}\}$

Niveau de suspicion





Résultats de la méthode TLA

Caractéristiques du cycle de vie

- Durée de vie de l'AdV : $[t_0, t_f]$.
- $\boldsymbol{\alpha}$: $t_0 + 6d$.
- $\boldsymbol{\beta}$: $t_0 + 123d$.

$$\theta_{TLA\ Attack}(q, e, t) = \begin{cases} 0.5 & \text{si } q = 0 \\ \frac{\beta - t}{\beta - \alpha} - \epsilon & \text{si } q = 1 \ \& \ e \in E_a \cup \{e_{n_w}\} \\ \frac{t - \alpha}{\beta - \alpha} + \epsilon & \text{si } q = 1 \ \& \ e = e_{n_e} \\ \epsilon & \text{si } q = 2 \ \& \ e \in E_a \cup \{e_{n_w}\} \\ 1 - \epsilon & \text{si } q = 2 \ \& \ e = e_{n_e} \end{cases}$$

Résultats d'estimation du niveau de menace avec la méthode TLA

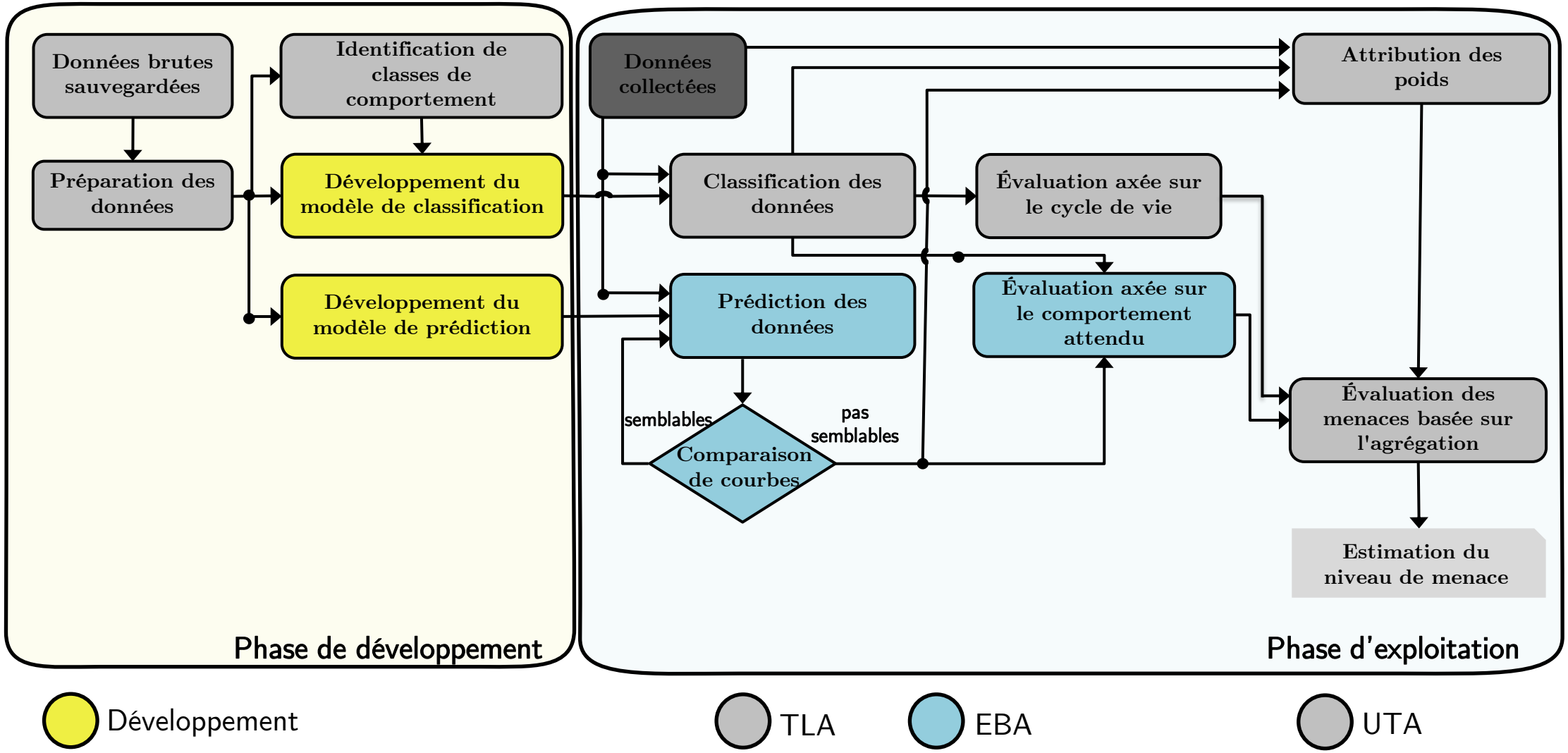
T	t	e	θ_{TLA_Attack}	
$[t_0, t_0 + 6d[$	$t_0 + 4d$	e_{a_m}	0.5	😐
	$t_0 + 5d$	e_{n_e}	0.5	
$[t_0 + 6d, t_0 + 123d[$	$\boldsymbol{t_0 + 8d}$	e_{a_s}		☹
	$t_0 + 10d$	e_{n_e}	0.1	☺
	$\boldsymbol{t_0 + 18d}$	e_{n_w}	**0.8**	☹
	$\boldsymbol{t_0 + 24d}$	e_{a_p}	**0.7**	☹
	$\boldsymbol{t_0 + 44d}$	e_{a_m}	**0.6**	☹
	$t_0 + 56d$	e_{a_s}	0.5	😐
	$t_0 + 90d$	e_{a_m}	0.2	☺
	$t_0 + 92d$	e_{a_p}	0.2	☺
	$\boldsymbol{t_0 + 94d}$	e_{n_e}	**0.8**	☹
	$\boldsymbol{t_0 + 114d}$	e_{n_e}	**0.9**	☹
$[t_0 + 123, t_0 + 155d[$	$t_0 + 125d$	e_{a_m}	0.1	☺
	$t_0 + 126d$	e_{a_m}	0.1	☺
	$t_0 + 143d$	e_{a_m}	0.1	☺
	$\boldsymbol{t_0 + 149d}$	e_{n_e}	**0.9**	☹
	$t_0 + 151d$	e_{a_s}	0.1	☺
	$t_0 + 154d$	e_{n_w}	0.1	☺

☺ Conforme

😐 Incertitude

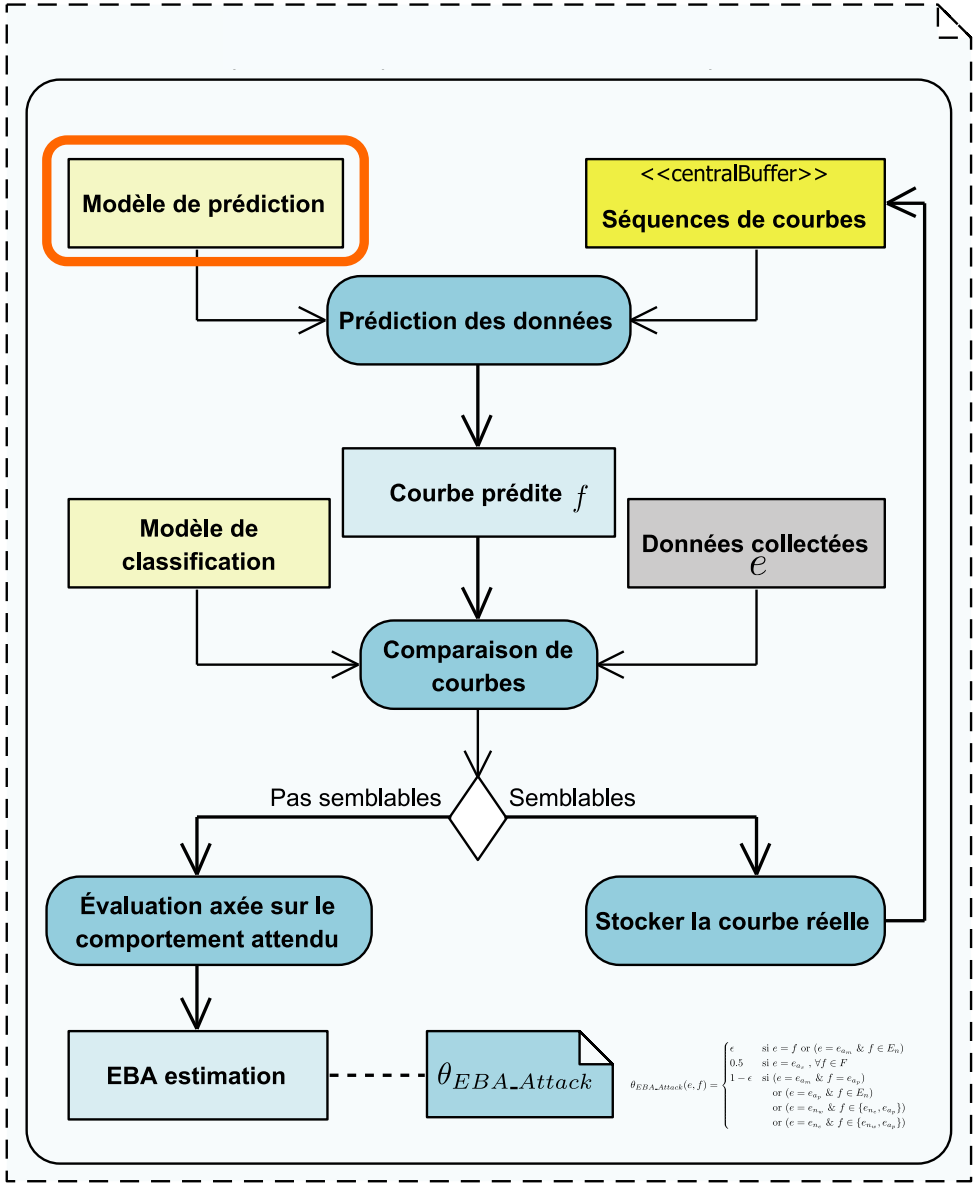
☹ Suspicion

Analyse du Comportement Attendu (EBA)



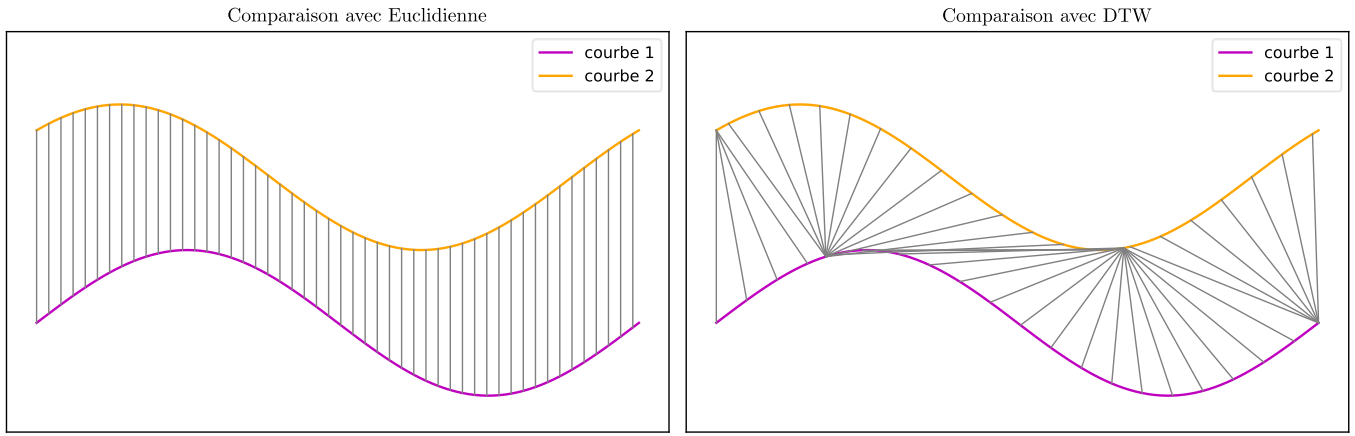
Analyse du Comportement Attendu (EBA)

Chaque courbe a du sens au sein de la séquence de courbes qui la précède.

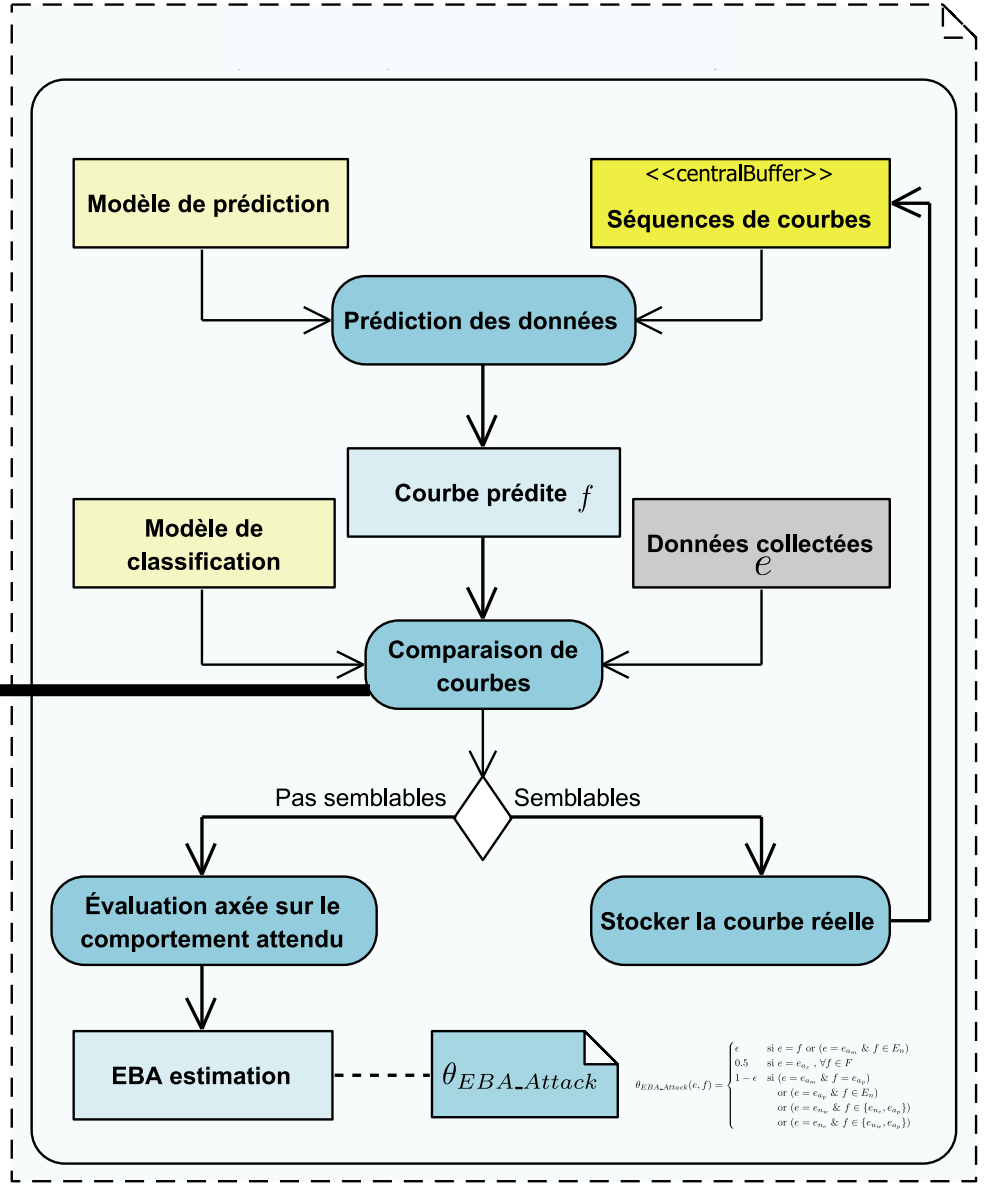


Analyse du Comportement Attendu (EBA)

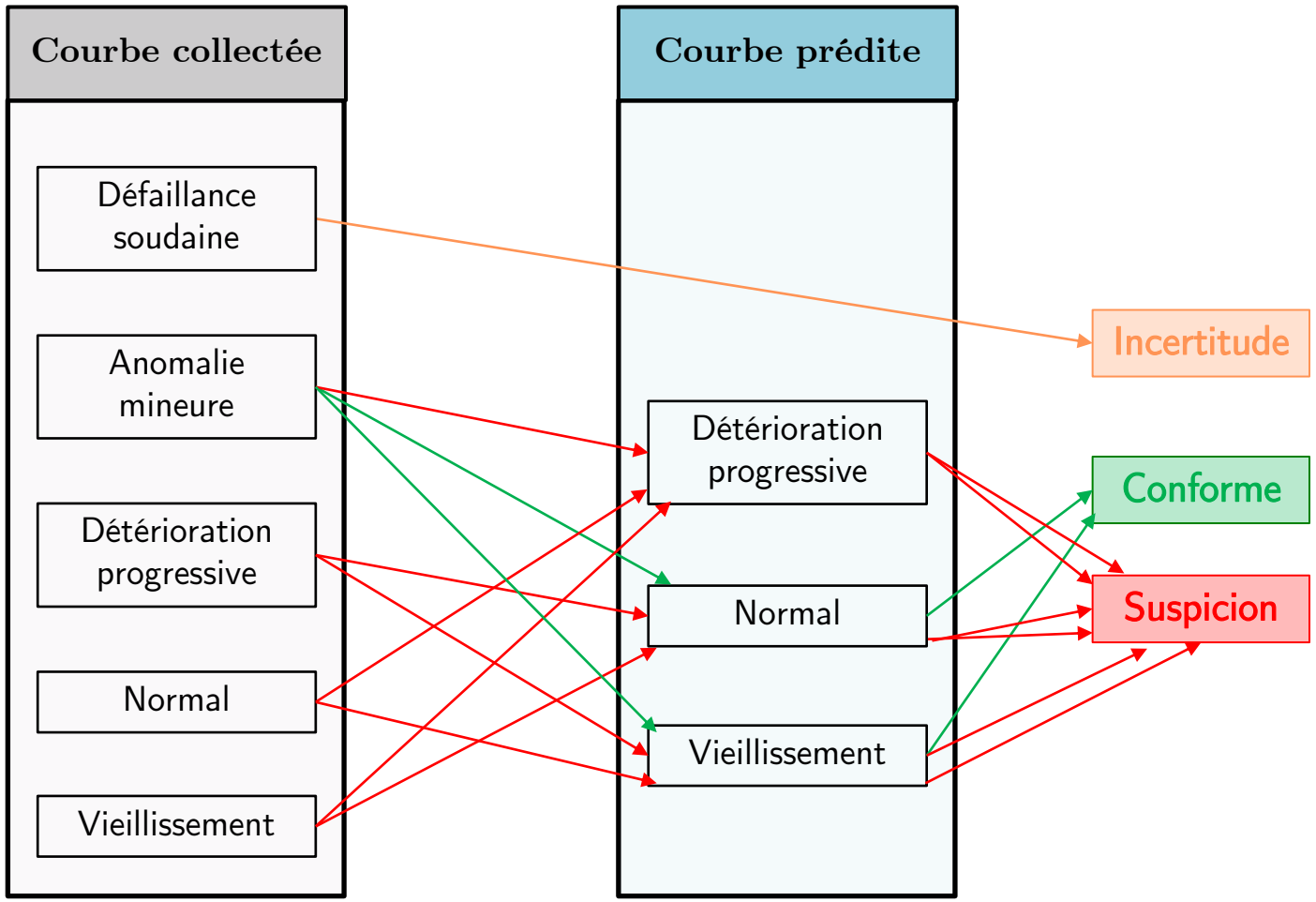
Chaque courbe a du sens au sein de la séquence de courbes qui la précède.

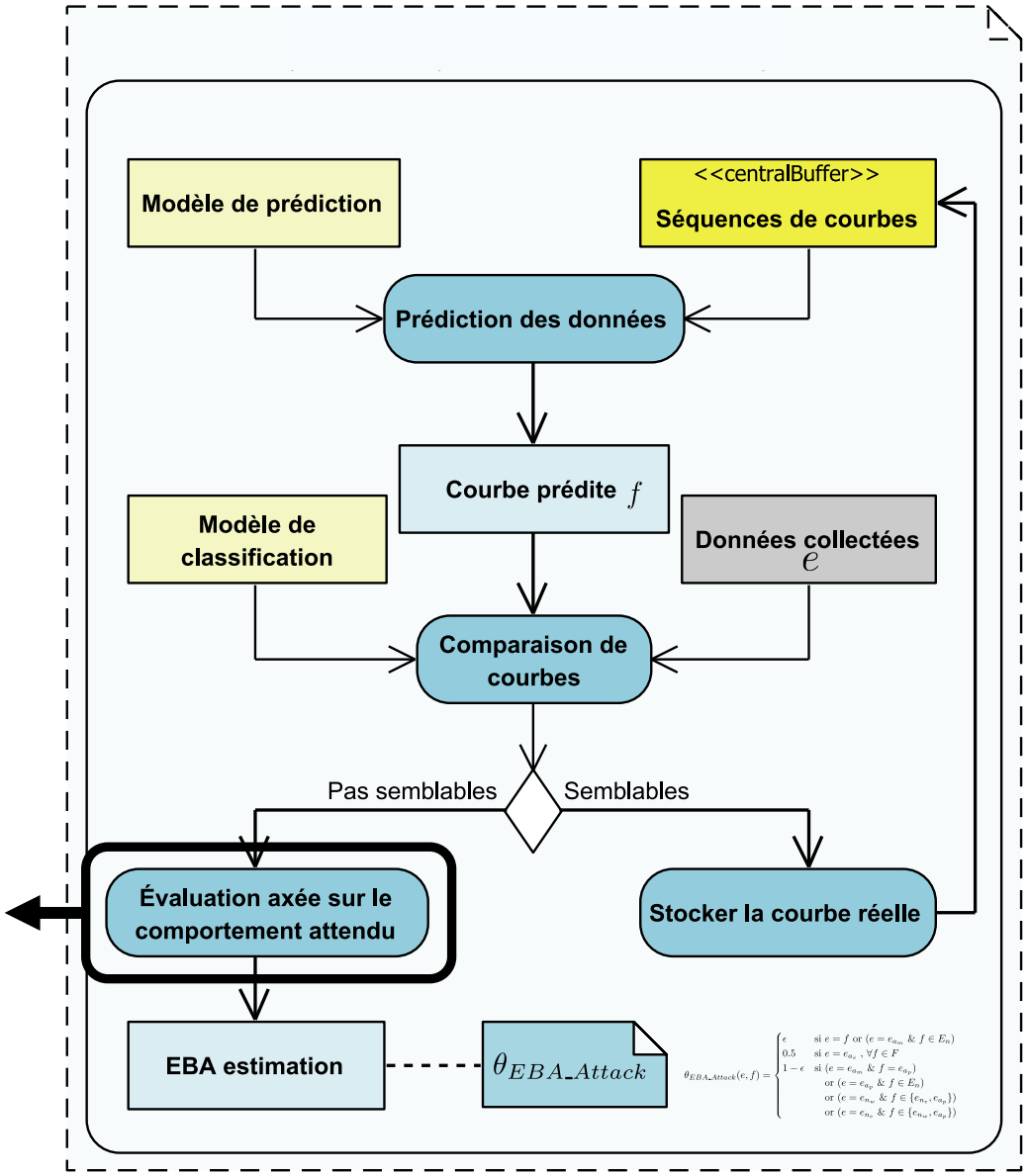


* Dynamic Time Warping



Analyse du Comportement Attendu (EBA)





Résultats de la méthode EBA

$$\theta_{EBA\_Attack}(e,f) = \begin{cases} \epsilon & \text{si } e = f \text{ or } (e = e_{a_m} \ \& \ f \in E_n) \\ 0.5 & \text{si } e = e_{a_s} \ , \forall f \in F \\ 1-\epsilon & \text{si } (e = e_{a_m} \ \& \ f = e_{a_p}) \\ & \quad \text{or } (e = e_{a_p} \ \& \ f \in E_n) \\ & \quad \text{or } (e = e_{n_w} \ \& \ f \in \{e_{n_e}, e_{a_p}\}) \\ & \quad \text{or } (e = e_{n_e} \ \& \ f \in \{e_{n_w}, e_{a_p}\}) \end{cases}$$

Résultats d'estimation du niveau de menace avec la méthode EBA

Seuil DTW = 15, Seuil ED = 4

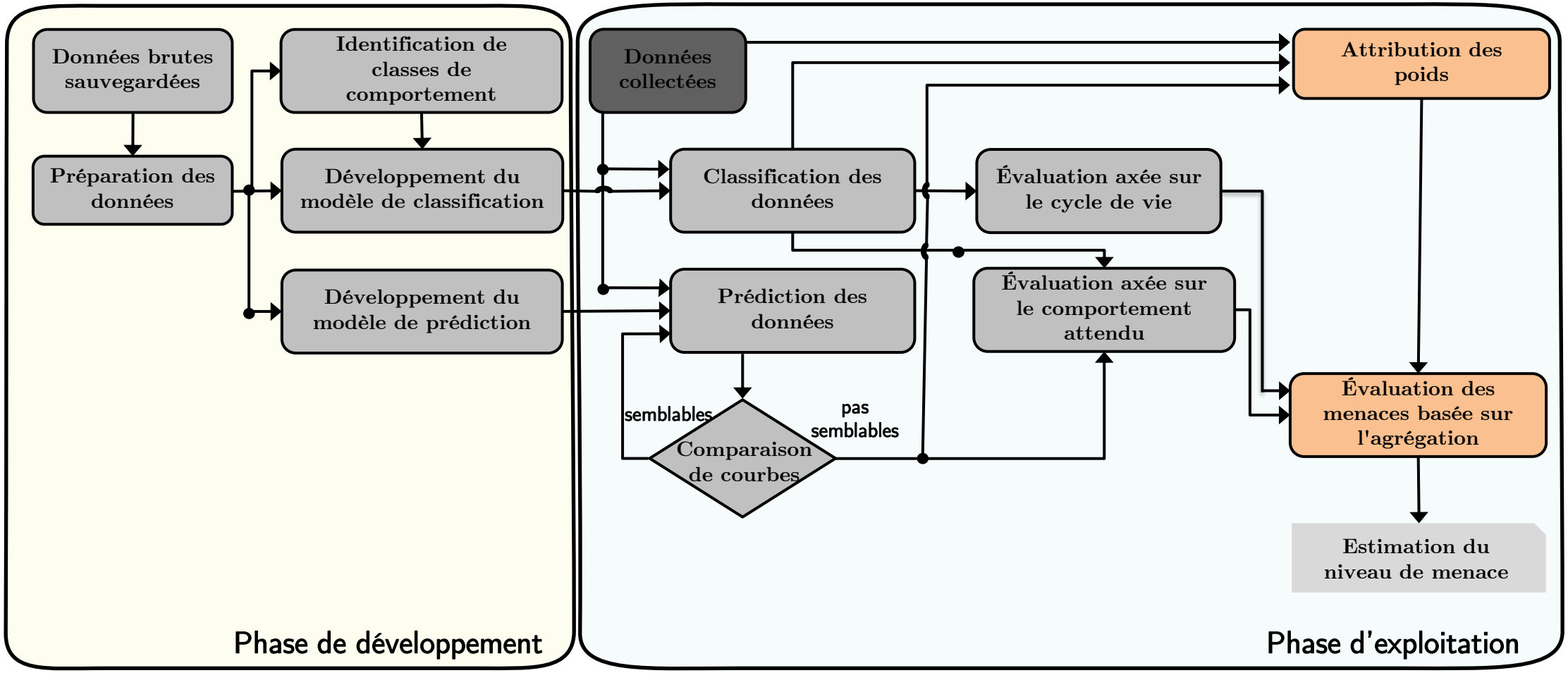
t	e	f	θ_{EBA_Attack}
$t_0 + 8d$	e_{a_s}	e_{n_e}	0.5 😐
$\boldsymbol{t_0 + 18d}$	e_{n_w}	e_{n_e}	**0.9** ☹
$\boldsymbol{t_0 + 24d}$	e_{a_p}	e_{n_e}	**0.9** ☹
$t_0 + 56d$	e_{a_s}	e_{n_e}	0.5 😐
$\boldsymbol{t_0 + 90d}$	e_{a_m}	e_{a_p}	**0.9** ☹
$\boldsymbol{t_0 + 94d}$	e_{n_e}	e_{a_p}	**0.9** ☹
$t_0 + 94d$	e_{a_p}	e_{a_p}	0.1 ☺
$t_0 + 102d$	e_{a_m}	e_{n_e}	0.1 ☺
$t_0 + 126d$	e_{a_m}	e_{n_w}	0.1 ☺
$t_0 + 127d$	e_{a_s}	e_{n_w}	0.5 😐
$t_0 + 151d$	e_{a_s}	e_{n_w}	0.5 😐

☺ Conforme

😐 Incertitude

☹ Suspicion

Évaluation Unifiée des Menaces (UTA)



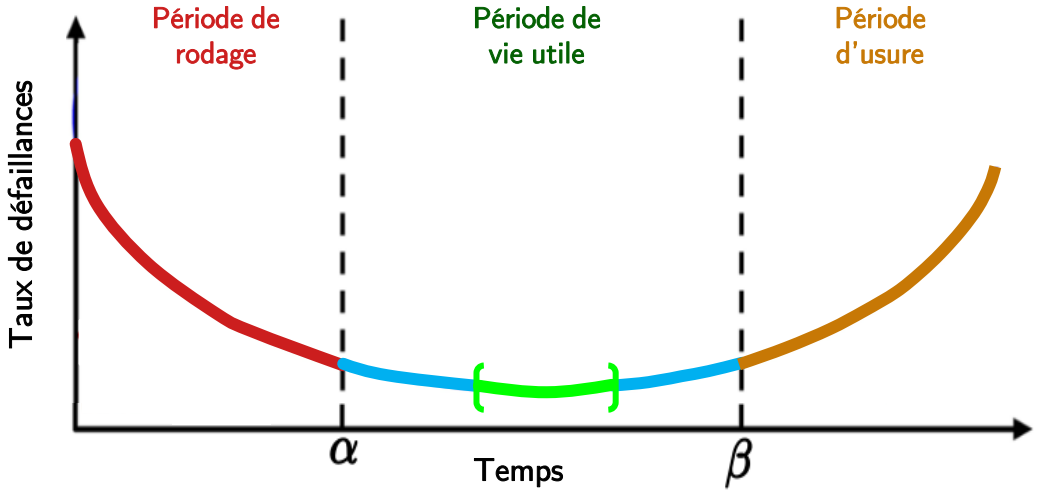
Évaluation des méthodes TLA et EBA

Méthode	Avantages	Limites
Analyse de Cycle de Vie (TLA)	Contextualisation de chaque courbe par rapport au cycle de vie de l'AdV.	- Incertitudes durant la période de rodage de l'AdV. - Incertitudes au milieu de la vie-utile de l'AdV.
Analyse du Comportement Attendu (EBA)	Contextualisation de chaque courbe par rapport à celles précédemment observées.	- Incertitudes liées aux défaillances soudaines.

Attribution des poids de confiance

t	e	f	ω_{TLA}	ω_{EBA}
$[t_0, \alpha[$	E_n	$E_n \cup \{e_{a_p}\}$	0	1
	e_{a_p}	E_n	0	1
	e_{a_m}	F	0	1
	e_{a_s}	F	0.5	0.5
$[\alpha, 0.3\beta + 0.7\alpha]$ *or* $[0.7\beta + 0.3\alpha, \beta[$	e_{a_s}	F	1	0
	e_{a_m}	F	0.1	0.9
	else	F	0.4	0.6
$]0.3\beta + 0.7\alpha, 0.7\beta + 0.3\alpha[$	e_{a_s}	F	1	0
	else	F	0.1	0.9
$[\beta, t_f[$	e_{n_c}	e_{n_w}	0.9	0.1
	e_{a_s}	F	1	0
	else	F	0.1	0.9

$$\omega_{TLA}, \omega_{EBA} : \mathbb{R}^+ \times E \times F \rightarrow [0, 1]$$



*Abdellaoui, S., Dumitrescu, E., Escudero, C., Zamai, E., 2026. Monitoring cyberthreats in railway systems: A hybrid framework for detecting stealthy data tampering attacks. Reliability Engineering & System Safety 266, 111747. https://doi.org/10.1016/j.ress.2025.111747

Évaluation des résultats

t	e	Comportement réel	TLA	EBA	UTA
$t_0 + 4d$	e_{a_m}	Légitime			
$t_0 + 5d$	e_{n_e}	Légitime			
$t_0 + 8d$	e_{a_s}	Attaque A_1			
$t_0 + 10d$	e_{n_e}	Légitime			
$t_0 + 18d$	e_{n_w}	Attaque A_2			
$t_0 + 24d$	e_{a_p}	Attaque A_3			
$t_0 + 44d$	e_{a_m}	Légitime			
$t_0 + 56d$	e_{a_s}	Attaque A_4			
$t_0 + 90d$	e_{a_m}	Attaque A_5			
$t_0 + 92d$	e_{a_p}	Légitime			
$t_0 + 94d$	e_{n_e}	Attaque A_6			
	e_{a_p}	Légitime			
$t_0 + 102d$	e_{a_m}	Légitime			
$t_0 + 114d$	e_{n_e}	Légitime			
$t_0 + 125d$	e_{a_m}	Légitime			
$t_0 + 126d$	e_{a_m}	Légitime			
$t_0 + 127d$	e_{a_s}	Légitime			
$t_0 + 143d$	e_{a_m}	Légitime			
$t_0 + 149d$	e_{n_e}	Attaque A_7			
$t_0 + 151d$	e_{a_s}	Légitime			
$t_0 + 154d$	e_{n_w}	Légitime			

Incertitude

Vrai Positif : attaque détectée

Vrai Négatif : comportement légitime détecté

Faux Négatif : attaque classée comme comportement légitime

Faux Positif : comportement légitime classé comme attaque

Conclusions

- Contextualisation basée sur le cycle de vie
- Analyse de cohérence comportementale temporelle
- Agrégation pour une interprétation globale
- Validation sur scénarios d'attaques simulés

UNIVERSITÉ DE LYON

INSA INSTITUT NATIONAL DES SCIENCES APPLIQUÉES LYON

Ampère

SAGIP
Société d'Automatique, de Génie Industriel & de Productique

Merci pour votre attention !

Sara ABDELLAOUI, Emil DUMITRESCU, Cédric ESCUDERO, Eric ZAMAÏ

Département Automatique pour l'Ingénierie des Systèmes (**AIS**), **AMPERE, INSA Lyon**

Équipe *Resilient Industrial Cyber Systems* (**RICS**)

saraabdellaoui1805@gmail.com