

Complément IA

Quelques méthodes de détection d'anomalies dans des séries temporelles

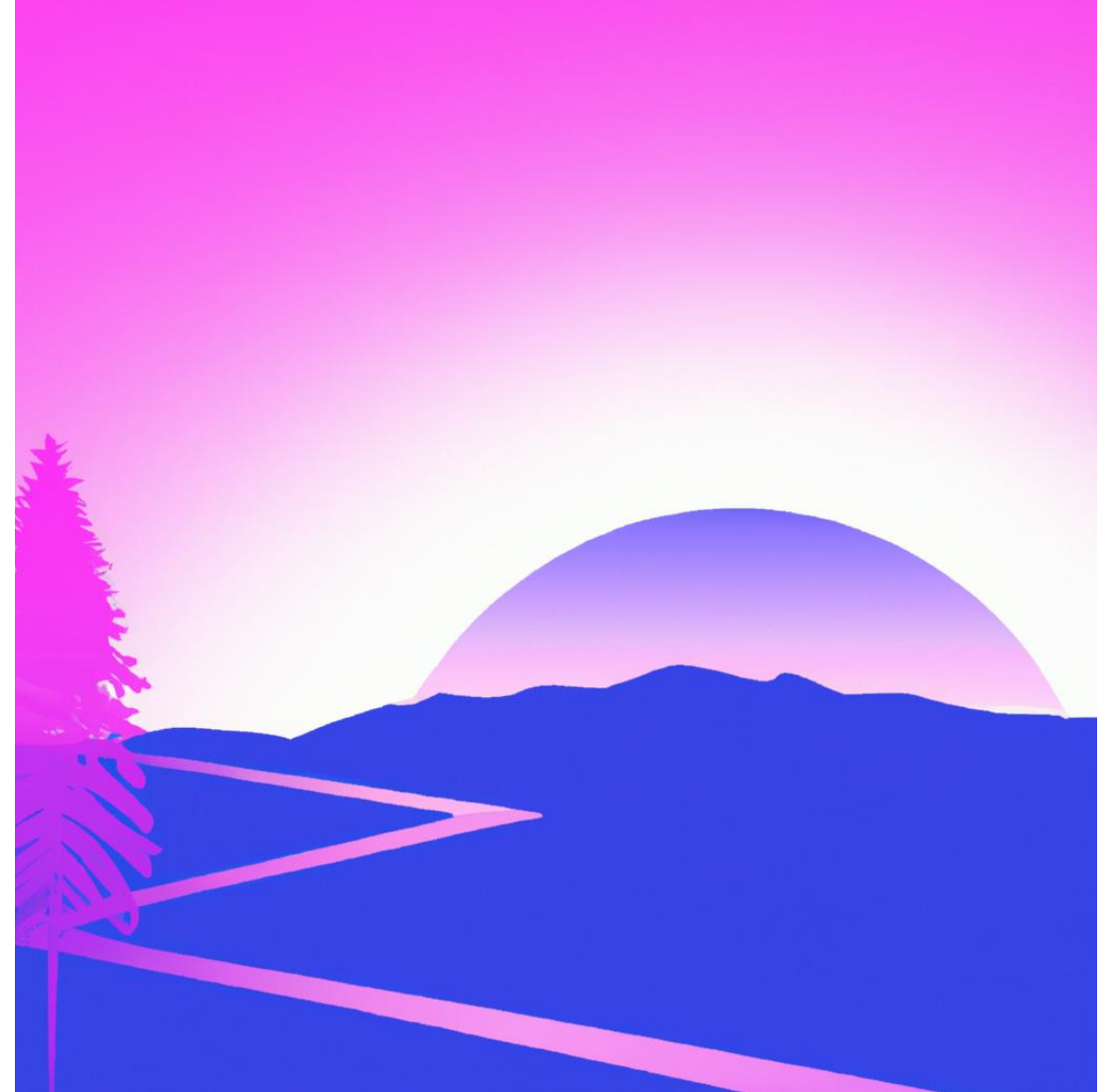
Isolation forests

Sparsity-based AD

1D-CNN data reconstruction

NAVAL GROUP

SEPTEMBRE 2025



Programme



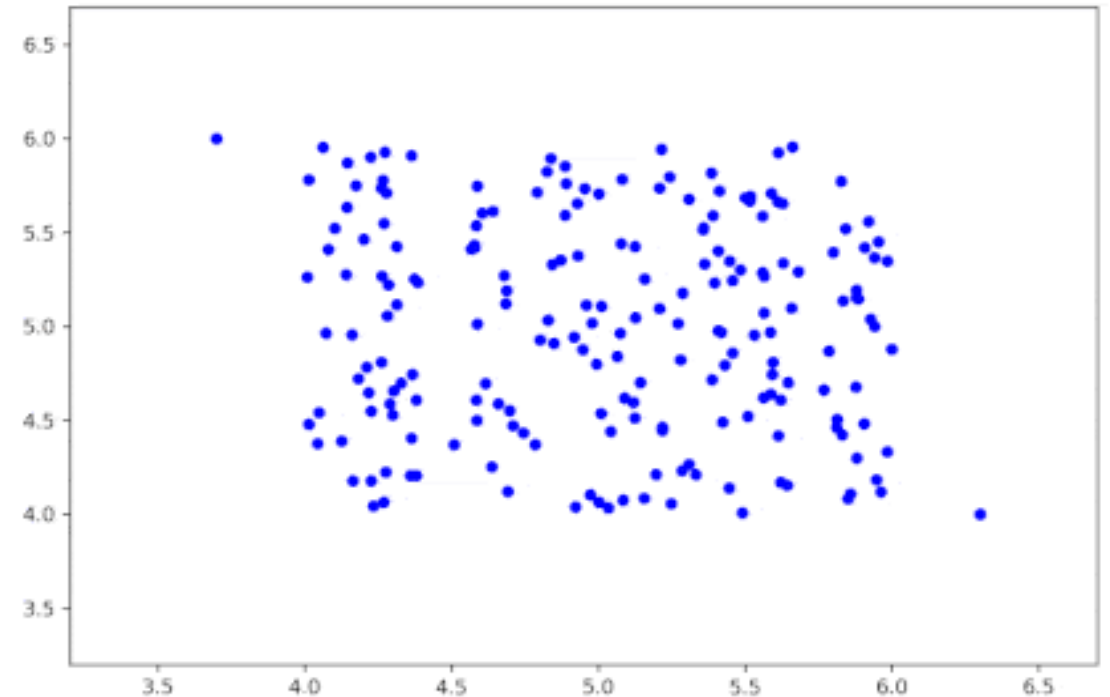
- 01 **Isolation forests**
- 02 **Sparsity-based AD**
- 03 **1D-CNN data reconstruction**

01

Isolation forests

Principe

- On sélectionne une **variable** (feature) de façon **aléatoire**.
- Ensuite on réalise un partitionnement aléatoire du jeu de données selon cette variable de telle sorte à obtenir **deux** sous-ensembles de données.
- On répète les deux étapes précédentes jusqu'à ce qu'une donnée soit **isolée**.
- De manière **récursive**, on répète les étapes précédentes.



Adaptation à des séries temporelles

- Isolation Forest n'est pas intrinsèquement prêt à gérer l'aspect séquentiel : il "voit" juste des vecteurs, pas une structure temporelle. Si on lui donne une série brute (un seul signal 1D), il va seulement chercher des valeurs numériques atypiques → il détecte des points aberrants, mais pas forcément des anomalies de type "changement de tendance" ou "rupture de régime".
- Cherche-t-on à identifier des anomalies dans un signal, ou un signal anormal parmi un ensemble de signaux ?
- Pour capturer l'aspect dynamique, on procède en général par une **transformation de la série avant de l'envoyer à notre Isolation Forest**.

Adaptation à des séries temporelles

- Pour capturer l'aspect dynamique, on procède en général par une **transformation de la série avant de l'envoyer à notre Isolation Forest** :

1. Fenêtrage

- On découpe la série en **vecteurs de longueur fixe** (par ex. 20 points consécutifs).
- Chaque fenêtre devient une observation pour Isolation Forest.
- Cela permet de détecter des motifs temporels inhabituels à l'intérieur d'un signal.

2. Caractéristiques dérivées

- On enrichit les données avec des *lag features* :
dérivées, variations locales, moyenne/écart-type dans une fenêtre, autocorrélations, FFT...
- Isolation Forest travaille alors sur ces représentations plus informatives.

3. Détection par "reconstruction" locale

- On calcule par exemple les **résidus** d'un modèle ARIMA, Prophet ou RNN → puis on applique Isolation Forest sur ces résidus pour isoler les séquences atypiques.

02

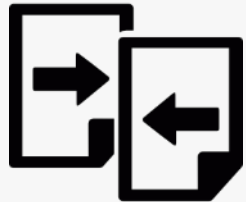
Sparsity-based AD

Version vulgarisée

- Un modèle « sparse » (parcimonieux) est utilisé pour représenter les séries temporelles de manière compacte tout en préservant les schémas rares.
- Cela implique l'utilisation d'une transformée en ondelettes.

Version vulgarisée

- Un modèle « sparse » (parcimonieux) est utilisé pour représenter les séries temporelles de manière compacte tout en préservant les schémas rares.
- Cela implique l'utilisation d'une transformée en ondelettes.

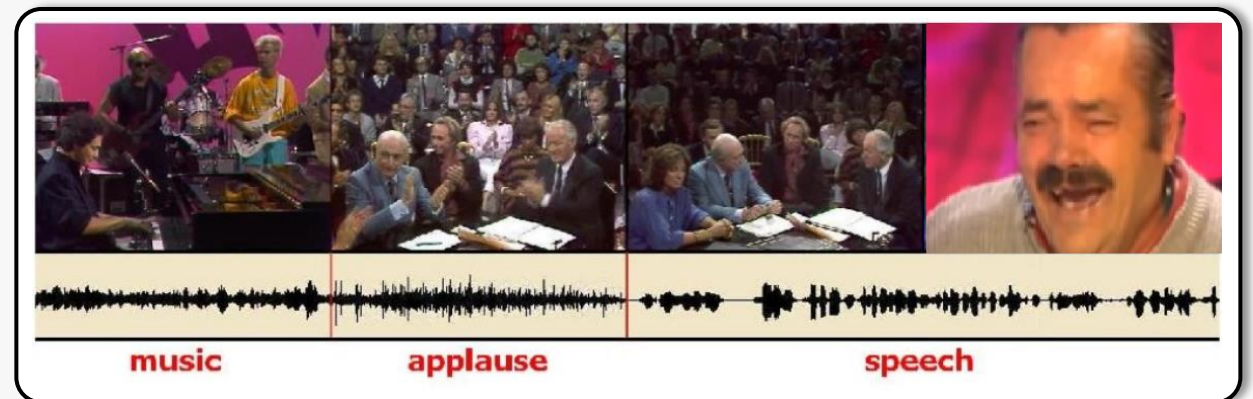
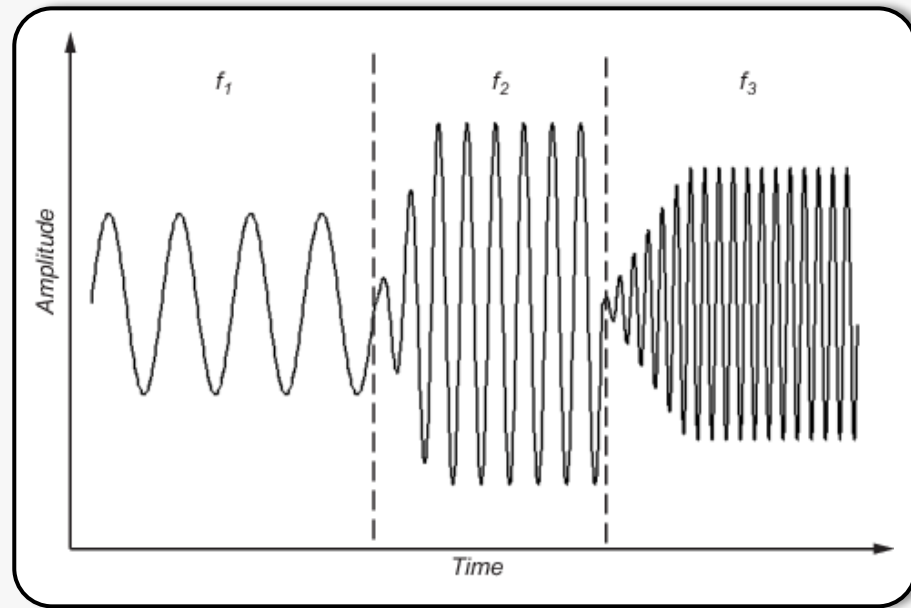


Analogie de la transformée en ondelettes avec un procédé classique ?

Qu'apporte la transformée en ondelettes ?

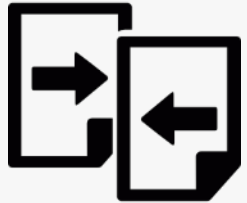
Version vulgarisée

- Un modèle « sparse » (parcimonieux) est utilisé pour représenter les séries temporelles de manière compacte tout en préservant les schémas rares.
- Cela implique l'utilisation d'une transformée en ondelettes.



Version vulgarisée

- Un modèle « sparse » (parcimonieux) est utilisé pour représenter les séries temporelles de manière compacte tout en préservant les schémas rares.
- Cela implique l'utilisation d'une transformée en ondelettes.
- On construit un ensemble d'objets (signaux temps/fréquence) élémentaires.
- Lors de la détection, on essaie en quelque sorte de reconstruire le signal de départ à partir de cet ensemble d'objets.

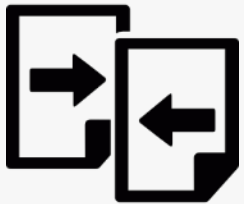


Analogie de cet ensemble avec ce qui est fait dans d'autres domaines de la data science ?

Est-ce une démarche supervisée ? Y a-t-il un fit ?

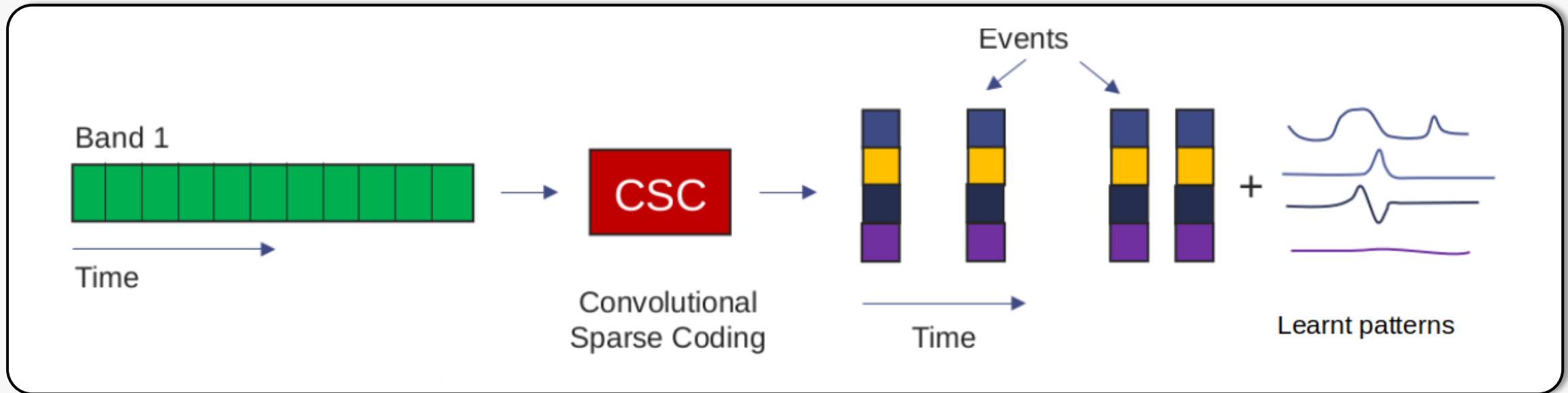
Version vulgarisée

- Un modèle « sparse » (parcimonieux) est utilisé pour représenter les séries temporelles de manière compacte tout en préservant les schémas rares.
- Cela implique l'utilisation d'une transformée en ondelettes.
- On construit un ensemble d'objets (signaux temps/fréquence) élémentaires.
- Lors de la détection, on essaie en quelque sorte de reconstruire le signal de départ à partir de cet ensemble d'objets.
- Les anomalies sont enfin scorées.

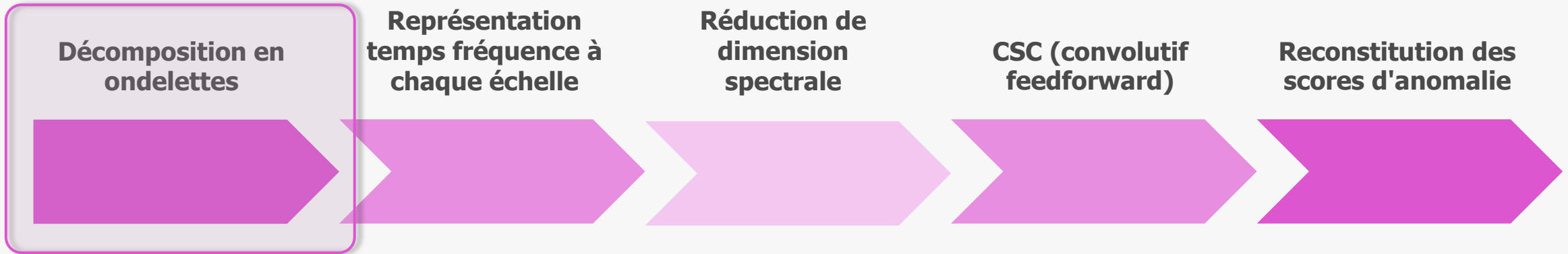


Comment ?

Version détaillée



Version détaillée



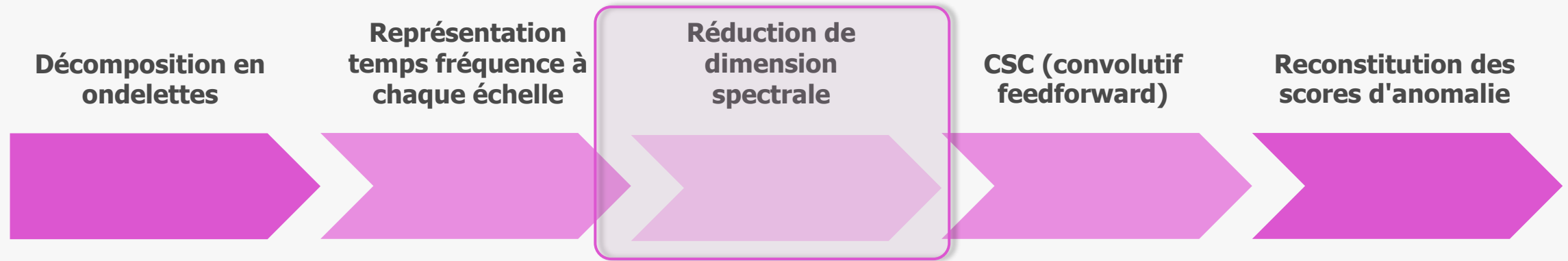
- ➡ Une séparation des échelles est effectuée à l'aide d'une **transformée en ondelettes**.
- ➡ Des **ondelettes de Daubechies** d'ordre 2 sont utilisées par défaut, ce qui signifie que les tendances quadratiques sont réduites à zéro dans les coefficients des ondelettes.
- ➡ La décomposition en ondelettes permet donc d'**éliminer les tendances polynomiales lisses** d'un degré donné et de **suivre les événements temporels à différentes échelles de temps**.

Version détaillée



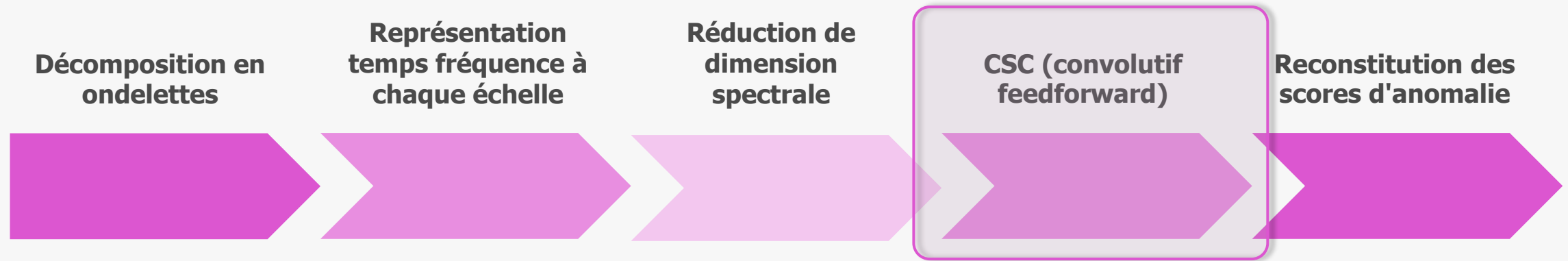
- ➡ Les coefficients d'ondelettes localisés dans le temps sont traités comme de nouvelles séries temporelles, qui sont ensuite décomposées en différentes **bandes de fréquences**, grâce à une représentation temps-fréquence.
- ➡ Une **transformée en cosinus discrète** est effectuée dans des fenêtres coulissantes avec un certain pourcentage de chevauchement. Il en résulte une nouvelle représentation du signal qui fournit l'évolution temporelle de son énergie dans différentes bandes de fréquence.
- ➡ **La taille de la fenêtre coulissante est un paramètre de lissage** qui détermine la finesse des variations spectrales potentiellement intéressantes. Le choix de ce paramètre doit donc être guidé par l'expertise du domaine. Le pourcentage de chevauchement est un paramètre de compression. Moins de chevauchement signifie moins d'échantillons temporels dans la représentation temps-fréquence, et donc une charge de calcul ultérieure potentiellement plus faible. Un chevauchement plus important signifie plus de redondance entre les échantillons voisins et moins de risque de ruptures artificielles qui seraient simplement des artefacts du fenêtrage. Le choix de ce paramètre doit donc reposer sur un compromis entre le temps de calcul et le risque de faux positifs.

Version détaillée



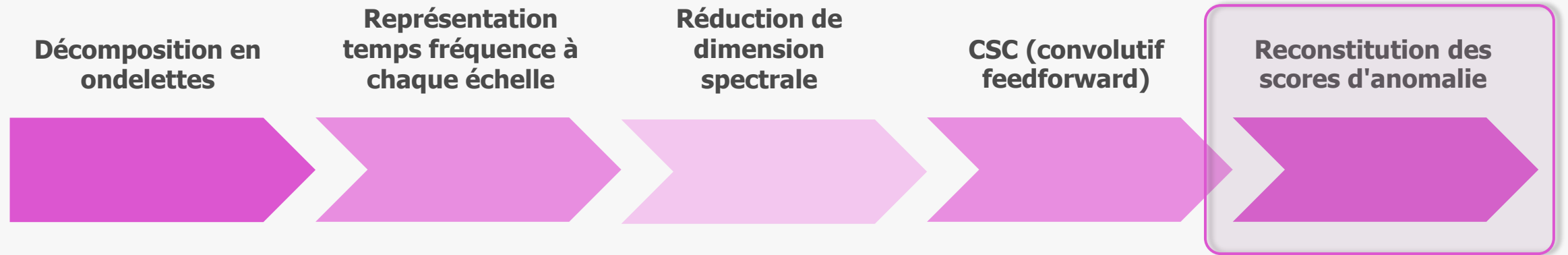
- ➡ Les représentations temps-fréquence obtenues à différentes échelles peuvent être comprimées davantage, étant donné la redondance qui existe entre les canaux de fréquence voisins.
- ➡ Dans cette étape, la **réduction de la dimension spectrale** se fait à l'aide d'un **auto-encodeur**.
- ➡ A noter que lorsque l'encodeur ne comporte qu'une seule couche, cela revient à apprendre une banque de filtres dans le domaine spectral, ce qui permet de résumer l'information dans un plus petit nombre de canaux de fréquence.

Version détaillée



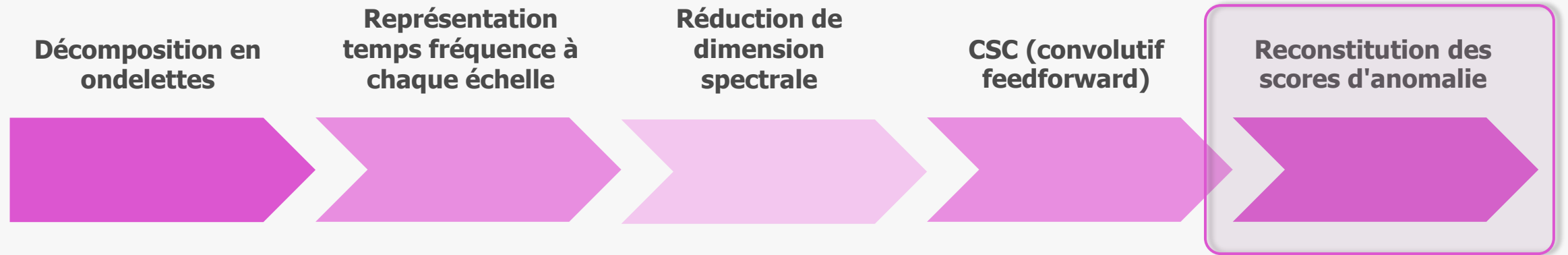
- ➡ À ce stade, on dispose de signaux très simplifiés à différentes échelles et bandes de fréquence pour détecter les événements temporels marquants.
- ➡ Afin de les délimiter, chacun des signaux spectro-temporels est décomposé en une combinaison linéaire parcimonieuse de motifs de glissement avec la contrainte importante que les activations du motif sont simultanées et à un seul instant dans la plage de temps couverte par le motif. Le vecteur des coefficients d'activation représente l'événement temporel. Les contraintes de simultanéité et d'exclusion imposées garantissent que ces vecteurs résument effectivement la dynamique locale du signal, sans aucune fuite d'information dans les vecteurs voisins, et qu'ils peuvent être comparés les uns aux autres en tant qu'événements temporels distincts.
- ➡ La taille des motifs appris implique donc un choix quant à l'extension temporelle des phénomènes capturés.
- ➡ Les événements temporels dont les vecteurs d'activation ont des normes non négligeables sont sélectionnés pour le calcul des scores d'anomalie.

Version détaillée



- ➡ Après tout ce travail, on dispose de plusieurs ensembles de vecteurs représentant des événements temporels à différentes échelles et dans différents canaux de fréquences. Toute méthode de détection d'échantillons hors distribution peut être utilisée pour terminer la détection d'anomalies.
- ➡ Si des labels supervisés sont disponibles, il est également possible d'utiliser ces vecteurs pour apprendre un classificateur d'anomalies de manière supervisée.
- ➡ Étant donné que dans un framework générique la distribution sous-jacente aux codes éparés est inconnue et potentiellement très irrégulière, on peut utiliser une **méthode non paramétrique** de détection d'anomalies, par exemple basée sur le **transport optimal** (à voir comme un outil géométrique pratique pour manipuler des distributions empiriques avec peu d'hypothèses sur leur structure globale).

Version détaillée



- ➡ Une première méthode consiste à résoudre une variable du problème de transport dans laquelle on minimise le coût du transfert de masse d'un point à un autre, au sein de la même distribution empirique, en pénalisant le non-transfert de masse.
- ➡ Dans une seconde méthode, chaque canal génère une distribution d'événements temporels proches de ceux des autres, et on calcule le barycentre de ces distributions dans le sens d'un transport optimal.
- ➡ Dans les deux approches, un coût moyen d'envoi et de réception de masse élevé pour un échantillon indique qu'il est isolé des autres.
- ➡ Une dernière étape consiste à passer des scores d'anomalies sur les événements temporels extraits aux scores d'anomalies sur les séries temporelles initiales à chaque instant.

03

1D-CNN data reconstruction

Version vulgarisée

- ➡ Dans cette méthode, l'architecture du 1D-CNN est similaire à celle utilisée dans l'apprentissage supervisé. Cependant, dans la détection d'anomalies non supervisée, le réseau est entraîné sans utiliser d'anomalies explicitement labellisées, et (si possible) uniquement sur des données normales.
- ➡ Au cours de l'entraînement, le réseau apprend à encoder les patterns et structures présents dans les données normales. L'objectif de l'entraînement est de minimiser l'erreur de reconstruction entre les données d'entrée et leur version reconstruite par le réseau. Dans cette phase, le réseau n'est (si possible) exposé qu'à des données normales.
- ➡ Une fois que le modèle est entraîné sur des données normales, il peut être utilisé pour reconstruire les séquences d'entrée pendant l'inférence. Les anomalies sont identifiées sur la base des divergences entre les séquences d'entrée originales et leurs reconstructions. En d'autres termes, les séquences présentant des erreurs de reconstruction ou des dissimilitudes importantes sont détectées comme anomalies.

Version vulgarisée

