Formule D.D Cohérence IA

Cadre méthodologique de diagnostic des transitions de phase

Version consolidée

A) Objet

La **Formule D.D Cohérence** est un opérateur algorithmique **strictement descriptif** destiné à produire des diagnostics de **changement de régime durable** dans des systèmes complexes observables, à partir de données publiques.

La formule :

- ne produit aucune décision,
- ne fournit aucune recommandation,
- n’implique aucun jugement normatif,
- ne possède aucune capacité prédictive,
- n’établit aucune causalité.

Elle fournit exclusivement un **diagnostic descriptif de compatibilité avec une transition de phase durable**, au sens défini dans le présent cadre.

B) Unité d’analyse, granularité et période

- **Unité d’analyse** : système SSS observé (par exemple un projet open-source, un système financier, sanitaire ou biophysique).
- **Granularité temporelle** Δt\Delta tΔt : fixée ex ante et constante (jour, semaine ou autre).
- **Période d’analyse** [tstart,tend][t_{start}, t_{end}][tstart,tend] : fixée ex ante.

- **Fenêtre de calcul** m : fixée ex ante.

Aucune modification post hoc de Δt, m, de la période d’analyse ou de l’unité d’observation n’est autorisée sans déviation explicitement déclarée.

C) Sources de données admissibles

Les données utilisées doivent être :

- publiques,
- accessibles antérieurement à l’analyse,
- documentées et auditables.

Sources admissibles (exemples non exhaustifs) :

- systèmes de contrôle de version (commits, auteurs, tags, reverts),
- plateformes collaboratives (issues, PR/MR, reviews, temps de cycle),
- métriques de releases,
- données de sécurité (advisories, CVE),
- dépendances logicielles,
- réseaux de contributeurs ou de dépendances,
- séries financières, sanitaires ou environnementales publiques.

La formule est **indépendante du domaine d’application**.

D) Liste des indicateurs observables

$\mathbf{x}_t \in \mathbb{R}^d$

Avant toute extraction, la liste des d séries temporelles multivariées est **figée ex ante**.

Pour chaque composante, doivent être documentés :

- le nom,
- la définition,
- l'unité,
- les transformations éventuelles,
- les bornes plausibles,
- la règle de gestion des valeurs aberrantes.

Toute modification après pré-enregistrement constitue une **déviation déclarée**.

E) Prétraitement des données (règles figées)

Les règles suivantes sont fixées ex ante :

- granularité temporelle uniforme et alignement explicite des timestamps,
- absence d'imputation libre non documentée,
- absence d'événement ⇒ valeur 0 pour les compteurs,
- métriques non définies : règle fixée ex ante (valeur 0 ou exclusion),
- gestion des valeurs aberrantes par règle explicite (winsorisation, clip, exclusion ou conservation).

Toute modification après visualisation constitue une déviation déclarée.

F) Mesures critiques (fenêtre glissante)

Pour chaque instant $t \ge m$, on définit la fenêtre :

$W_t = \{t-m+1, \dots, t\}$

Pour chaque composante $j \in \{1,\dots,d\}$:

F.1 Variance

$V_{t,j} = \operatorname{Var}(x_{\tau,j} : \tau \in W_t)$

F.2 Autocorrélation (lag-1)

At,j=Corr(xτ,j,xτ−1,j:τ∈Wt∖{t−m+1})A_{t,j} = \operatorname{Corr}(x_{\tau,j}, x_{\tau-1,j} : \tau \in W_t \setminus \{t-m+1\})At,j=Corr(xτ,j,xτ−1,j:τ∈Wt∖{t−m+1})

F.3 Susceptibilité empirique

Soit utu_tut un signal exogène défini ex ante.

Si aucun signal n'est défini :

χt,j=0\chi_{t,j} = 0χt,j=0

Sinon :

χt,j=∣Cov(xτ,j,uτ)Var(uτ)+ε∣\chi_{t,j} = \left| \frac{\operatorname{Cov}(x_{\tau,j},u_\tau)}{\operatorname{Var}(u_\tau)+\varepsilon} \right|χt,j=Var(uτ)+εCov(xτ,j,uτ)

F.4 Entropie discrète

Discrétisation en KKK classes fixées ex ante :

Ht,j=−∑k=1Kp^t,j(k)log(p^t,j(k)+ε)H_{t,j} = -\sum_{k=1}^{K} \hat p_{t,j}(k)\log(\hat p_{t,j}(k)+\varepsilon)Ht,j=−k=1∑Kp^t,j(k)log(p^t,j(k)+ε)

G) Normalisation robuste

Référence initiale

T0(0)=[m−1, N0]T_0^{(0)} = [m-1,\, N_0]T0(0)=[m−1,N0]

(premières périodes disponibles).

Cette référence est **la seule utilisée pour la détection du premier instant diagnostiqué**.

Référence roulante post-transition

Après une transition diagnostiquée tk∗t_k^*tk∗, pour t≥tk∗+at \ge t_k^* + at≥tk∗+a :

T0(t)=[t−Lref+1, t]T_0(t) = [t-L_{ref}+1,\, t]T0(t)=[t−Lref+1,t]

La référence roulante **n'intervient jamais dans la détection initiale**, uniquement dans l'analyse post-transition ou les transitions ultérieures.

Z-scores robustes

Pour toute mesure M∈{V,A,χ,H}M \in \{V,A,\chi,H\}M∈{V,A,χ,H} :

Zt,j(M)=Mt,j−MedT0(t)(M·,j)MADT0(t)(M·,j)+εZ^{(M)}_{t,j} = \frac{M_{t,j}-\operatorname{Med}_{T_0(t)}(M_{\cdot,j})}{\operatorname{MAD}_{T_0(t)}(M_{\cdot,j})+\varepsilon}Zt,j(M)=MADT0(t)(M⋅,j)+εMt,j−MedT0(t)(M⋅,j)

Stabilisation optionnelle pré-enregistrée :

Zt,j(M)←clip(Zt,j(M),−zmax,zmax)Z^{(M)}_{t,j} \leftarrow \operatorname{clip}(Z^{(M)}_{t,j},-z_{max},z_{max})Zt,j(M)←clip(Zt,j(M),−zmax,zmax)

H) Score de criticité par composante

Ct,j=σ (αVZt,j(V)+αAZt,j(A)+αχZt,j(χ)+αHZt,j(H))C_{t,j} = \sigma\!\left(\alpha_V Z^{(V)}_{t,j} + \alpha_A Z^{(A)}_{t,j} + \alpha_\chi Z^{(\chi)}_{t,j} + \alpha_H Z^{(H)}_{t,j} \right)Ct,j=σ(αVZt,j(V)+αAZt,j(A)+αχZt,j(χ)+αHZt,j(H))

avec :

σ(z)=11+e−z\sigma(z)=\frac{1}{1+e^{-z}}σ(z)=1+e−z1

I) Agrégation globale et verrous

Score global (quantile bas)

Φt=Qq(Ct,1,…,Ct,d)\Phi_t = Q_q(C_{t,1},\dots,C_{t,d})Φt=Qq(Ct,1,…,Ct,d)

Verrou de proportion

Ft=1d∑j=1d1[Ct,j≥θ]F_t = \frac{1}{d}\sum_{j=1}^{d}\mathbf{1}[C_{t,j}\ge\theta]Ft=d1j=1∑d1[Ct,j≥θ]

J) Rupture globale avant / après

μtpre=1k∑s=t−2k+1t−kΦs\mu_t^{pre} = \frac{1}{k}\sum_{s=t-2k+1}^{t-k}\Phi_sμtpre=k1s=t−2k+1∑t−kΦs
μtpost=1k∑s=t−k+1tΦs\mu_t^{post} = \frac{1}{k}\sum_{s=t-k+1}^{t}\Phi_sμtpost=k1s=t−k+1∑tΦs

---

K) Rupture multivariée par composante

Δt,j=(1k∑s=t−k+1tCs,j)−(1k∑s=t−2k+1t−kCs,j)\Delta_{t,j} = \left(\frac{1}{k}\sum_{s=t-k+1}^{t}C_{s,j}\right) - \left(\frac{1}{k}\sum_{s=t-2k+1}^{t-k}C_{s,j}\right)Δt,j=(k1s=t−k+1∑tCs,j)−(k1s=t−2k+1∑t−kCs,j)

L) Règle de diagnostic candidate

$$\mathsf{Cand}(t)= (\Phi_t\ge\theta) \land (\forall s\in[t-r+1,t],\Phi_s\ge\theta) \land (\mu_t^{post}-\mu_t^{pre}\ge\kappa) \land (F_t\ge p) \land \left(\frac{1}{d}\sum_j\mathbf{1}[\Delta_{t,j}\ge\kappa_C]\ge p_C \right)$$

M) Transitions multiples (cooldown)

$$t_1^*=\min\{t:\mathsf{Cand}(t)\}$$

$$t_{k+1}^*=\min\{t\ge t_k^*+c:\mathsf{Cand}(t)\}$$

L'absence de t^* constitue un **résultat valide**.

N) Paramètres figés (exemple conservateur)

(inchangés par rapport à v1.0)

O) Tests de robustesse pré-enregistrés

- permutations par blocs,
- holdout temporel,
- concordance événementielle descriptive.

P) Sorties obligatoires

- instants $\{t_k^*\}$,
- séries Φ_t, F_t,
- métriques locales associées,
- paramètres exacts,

- hash du code et des données.

Q) Déviations au protocole

Toute modification post hoc doit être :

- explicitement déclarée,
- justifiée,
- accompagnée des résultats avec et sans modification.

Statut final

La **Formule D.D Cohérence IA** est un opérateur générique, non compensable et non normatif de **diagnostic descriptif de changements de régime durables** dans des systèmes complexes observables.