



# Tetrachore User Guide

Laurent Oliversen · Copyright 2026, Laurent Oliversen. All rights reserved. · Build 2026-09-11

---

## Sommaire

### 1. Prise en main

- 1.1. Organisation de l'interface
- 1.2. Le panneau flottant du Workspace
- 1.3. Modes utilisateur
- 1.4. Forfaits et verrouillage

### 2. Noyau Géométrique

#### 2.1. Hypersolid Gen

- 2.1.1. Polygons (polygones, 2D)
- 2.1.2. Polyhedra (polyèdres, 3D)
- 2.1.3. Polychora (polychores, 4D)
- 2.1.4. N-Polytopes ( $N > 4$ )
- 2.1.5. Special
- 2.1.6. Réglages des quadriques (sphères, ellipsoïdes, tores)
- 2.1.7. Opérateurs de graphe
- 2.1.8. Tessellation (Pavage)

#### 2.2. Hypersolid Style

- 2.2.1. Vertices (sommets)
- 2.2.2. Edges (arêtes)
- 2.2.3. Faces
- 2.2.4. Arrière-plan

#### 2.3. Projection

- 2.3.1. Views
- 2.3.2. ND vers 4D
- 2.3.3. 4D vers 3D
- 2.3.4. Camera

#### 2.4. Motion

- 2.4.1. Dynamic Effects

- 2.4.2. Orbit (déplacements automatiques de caméra)
- 2.4.3. Auto-Rotate (auto-rotation)
- 2.4.4. Auto-Shear (auto-cisaillement)
- 2.4.5. Interpolation
- 2.4.6. Oscillation

## 2.5. Quelques exemples de projection

- 2.5.1. Tesseract

## 3. Ingestion de données

### 3.1. Dataset

- 3.1.1. Synthetic Data
- 3.1.2. Reference Data
- 3.1.3. Custom Points

### 3.2. Points & Labels

- 3.2.1. Familles de points
- 3.2.2. Affichage
- 3.2.3. Style des points
- 3.2.4. Encodage par variable
- 3.2.5. Dégradé de couleur
- 3.2.6. Couleurs fixes
- 3.2.7. Style des étiquettes

### 3.3. Data Import

- 3.3.1. Import CSV
- 3.3.2. Role Mapping
- 3.3.3. Axes Mapping
- 3.3.4. Sampling (échantillonnage)
- 3.3.5. Test Split
- 3.3.6. Prévisualisation

### 3.4. Preprocessing (pré-traitement)

- 3.4.1. Typage des colonnes
- 3.4.2. Imputation
- 3.4.3. Winsorization
- 3.4.4. Discretization
- 3.4.5. Category Binning (regroupement de modalités)
- 3.4.6. One-Hot Encoding (encodage disjonctif complet)

## 4. Exploration de données

### 4.1. Stats

- 4.1.1. Statistiques descriptives
- 4.1.2. Statistiques inférentielles

### 4.2. Data Exp (Exploration)

- 4.2.1. Selection Mode (orbite et sélection)
- 4.2.2. Data Scaling (mise à l'échelle)
- 4.2.3. Filtering (filtrage)
- 4.2.4. Orbital
- 4.2.5. Decision Boundary (frontière de décision)
- 4.2.6. Hy-Perf Cube
- 4.2.7. Valence
- 4.2.8. Style des valences et des orbitales
- 4.2.9. Trajectory (trajectoires)
- 4.2.10. Dimension Fading (estompage dimensionnel)
- 4.2.11. Marqueurs de scène
- 4.2.12. Surcouches des modules d'analyse

### 4.3. Data Style

- 4.3.1. Class Style (coloration par classe)
- 4.3.2. Couleurs de Valence
- 4.3.3. Layouts

## 5. Analyse

### 5.1. Lancer un module

### 5.2. Advanced Stats

- 5.2.1. Seuils pour les tests statistiques
- 5.2.2. Corrélations
- 5.2.3. Profiling (profilage)
- 5.2.4. Association

### 5.3. Hy-Personality

- 5.3.1. Sources de données
- 5.3.2. Techniques de projection
- 5.3.3. Filtrage des points de référence
- 5.3.4. Statistiques associées
- 5.3.5. Exemple d'utilisation : zodiaque inversé

### 5.4. Factor

- 5.4.1. Axes factoriels
- 5.4.2. Techniques factorielles
- 5.4.3. Statistiques associées

### 5.5. Clustering

- 5.5.1. Sources de données
- 5.5.2. Méthodes
- 5.5.3. Pre-KMeans
- 5.5.4. HCA (CAH)
- 5.5.5. Final Cleaning (nettoyage final)

5.5.6. Ellipsoïdes de cluster

5.5.7. Statistiques associées

## 5.6. Sup ML (machine learning supervisé)

5.6.1. Paramètres

5.6.2. Statistiques associées

## 5.7. Optimization

5.7.1. Paramètres

5.7.2. Statistiques associées

## 5.8. Forecast

5.8.1. Paramètres

5.8.2. Statistiques associées

## 5.9. Hy-Performance

5.9.1. Sources de données

5.9.2. Paramètres

5.9.3. Statistiques associées

5.9.4. Guide méthodologique

## 5.10. MMM

5.10.1. Paramètres

5.10.2. Statistiques associées

5.10.3. Guide méthodologique

## 5.11. Hyperion

# 6. Import / Export

6.1. Hypersolid

6.2. Config

6.3. Scene

6.4. Analytics

6.5. PDF Recap

6.6. Info Points

Cette application web est un outil d'exploration de polytopes et d'analyse de jeux de données multidimensionnelles, le tout propulsé par une interface 3D intuitive.  
Ce guide explique les concepts mathématiques et les fonctionnalités de l'application Tetrachore.

Vous pouvez télécharger cette documentation au format PDF sur :  
<https://tetrachore.com/docs/tetrachore-user-guide-fr-light.pdf>

# 1. Prise en main

---

## 1.1. Organisation de l'interface

L'application s'organise en onglets, listés dans la barre supérieure.

À gauche, les onglets de travail :

- **Workspace** : la scène 3D et ses panneaux de réglages. C'est le cœur de l'application (partie 2 et sections 3.1, 4.2, 4.3 de ce guide).
- **Data** : import et pré-traitement des données (partie 3).
- **Stats** : statistiques descriptives et inférentielles (section 4.1).
- **Analysis** : un menu déroulant réunissant tous les modules d'analyse — **Advanced Stats**, **Hy-Personality**, **Factor**, **Cluster**, **Sup ML**, **Forecast**, **Hy-Performance**, **MMM**, **Optimization** et **Hyperion** (partie 5).
- **Export** : toutes les extractions (partie 6).

À droite, les onglets de service : **User Guide** (ce document), **Upgrade**, **Contact**.

## 1.2. Le panneau flottant du Workspace

Dans l'onglet **Workspace**, tous les réglages sont regroupés dans un panneau flottant que l'on peut déplacer en le saisissant par son en-tête, et replier avec le bouton de réduction. Il se découpe en cinq sections, sélectionnables par les onglets en haut du panneau :

- **Hypersolid** : génération et style du polytope (sections 2.1 et 2.2).
- **Dataset** : jeux de données de référence, points personnalisés, style des points et étiquettes (section 3.1).
- **Proj** : rotations, cisaillements, projections et caméra (section 2.3).
- **Exp** : exploration des données projetées — filtres, orbitales, valences, marqueurs de scène (section 4.2).
- **Motion** : animation et interpolation (section 2.4).

Chaque section conserve l'état plié ou déplié de ses dossiers lorsque l'on navigue entre elles.

Certains onglets d'analyse proposent en outre une bascule vers la vue scène, qui affiche la scène 3D en gardant le panneau du module à portée de main.

## 1.3. Modes utilisateur

Un sélecteur, en haut à droite, choisit le **mode utilisateur**. Le mode ne change rien aux calculs : il masque simplement les onglets et les formes hors de son périmètre, pour alléger l'interface.

-  **Explorer** : exploration géométrique et personnalité. Donne accès à tout le catalogue de formes, ainsi qu'aux onglets **Hy-Personality** et **Factor**.
-  **Data Scientist** : analyse de données. Ouvre **Hyperion**, **Factor**, **Cluster**, **Sup ML**, **Optimization** et **Forecast**. Le catalogue de formes est restreint aux formes essentielles, qui servent de repère visuel.
-  **Business Analyst** : pilotage de la performance. Ouvre **Hyperion**, **Optimization**, **Forecast**, **Hy-Performance** et **MMM**. Catalogue de formes également restreint.
-  **Hypervision** : réunit les trois précédents, sans restriction.

Les onglets **Workspace**, **Data**, **Stats**, **Advanced Stats** et **Export** sont visibles dans tous les modes.

## 1.4. Forfaits et verrouillage

Chaque fonctionnalité est rattachée à un forfait : **Guest**, **Free**, **Hyper**, **Pro**. Chaque forfait contient strictement le précédent.

Une fonctionnalité qui n'est pas couverte par votre forfait **n'est pas masquée** : elle apparaît grisée, suffixée d'un cadenas , et le survol indique le forfait requis. Vous voyez donc toujours ce que l'application sait faire, même sans y avoir accès.

Le détail des forfaits — quotas de lignes et de colonnes compris — se trouve dans l'onglet **Upgrade**.

## 2. Noyau Géométrique

---

### 2.1. Hypersolid Gen

L'application génère des formes géométriques définies par leurs sommets et leurs arêtes dans un espace à N dimensions. L'application prend en charge les formes suivantes. Notez que les sphères sont traitées comme des polytopes pour les besoins de la visualisation, bien qu'elles soient techniquement des variétés lisses.

Le pipeline de génération et de projection comporte les étapes suivantes :

- Génération du Polytope (Maths 100% CPU)
- ND-4D : normalisation, rotation, cisaillement, projection (Maths CPU)
- 4D-3D : rotation, cisaillement, projection, normalisation, mise à l'échelle (Maths CPU)
- 3D-2D/1D : écrasement sur un plan ou une droite (Filtre Maths CPU)
- Assemblage de la Scène (Construction des sphères, lignes et faces)
- Caméra / Rendu GPU (Calcul de l'angle final et affichage à l'écran)

L'ordre des transformations se fait bien dans cette ordre : 1) rotation, 2) cisaillement, 3) projection.

Les axes dimensionnelles sont notées ainsi :

- X, Y, Z pour les axes usuels de la 3D (orientés à droite, en haut et vers l'avant)
- W pour la quatrième dimension
- V5, V6, V7, V8, ... pour toutes les dimensions supérieures

#### 2.1.1. Polygons (polygones, 2D)

- **Triangle** : simplexe 2D (3 sommets). Le polygone régulier le plus simple.
- **Square** : hypercube 2D (4 sommets). Polygone régulier.
- **Rectangle** : polygone dérivé du carré, aux côtés de longueur irrégulière.
- **Pentagon** : polygone régulier à 5 sommets.
- **Hexagon** : polygone régulier à 6 sommets.
- **G2 Roots** : structure basée sur les vecteurs du système de racines G2.
- **Circle** : le cercle 2D.
- **Ellipse** : l'ellipse 2D.

Note : **Circle** et **Ellipse** sont des quadriques plates. Un cercle unique n'ayant ni parallèles ni méridiens à tracer, ils n'exposent que le tracé **Mesh** et le réglage **Resolution** .

### 2.1.2. Polyhedra (polyèdres, 3D)

- **Tetrahedron** (Tétraèdre) : simplexe 3D (4 sommets). Le polyèdre régulier le plus simple.
- **Cube** (Hexaèdre) : hypercube 3D (8 sommets). Hexaèdre régulier.
- **Cuboid** (Parallélépipède) : hypercube aux arêtes de longueur irrégulière.
- **Octahedron** (Octaèdre) : orthoplexe 3D (6 sommets). Dual du Cube.
- **Dodecahedron** (Dodécaèdre) : polyèdre régulier avec 12 faces pentagonales.
- **Icosahedron** (Icosaèdre) : polyèdre régulier avec 20 faces triangulaires.
- **Cuboctahedron** (Cuboctaèdre) : solide d'Archimède avec 8 faces triangulaires et 6 faces carrées.
- **Rhomb. Dodeca.** (Dodécaèdre Rhombique ou Granatoèdre) : solide de Catalan (dual du Cuboctaèdre) avec 12 faces en losange.
- **Icosidodecahedron** (Icosidodécaèdre) : solide d'Archimède avec 20 faces triangulaires et 12 faces pentagonales.
- **Rhomb. Triaconta.** (Triacontaèdre Rhombique) : solide de Catalan (dual de l'Icosidodécaèdre) avec 30 faces en losange.
- **Sphere** : sphère usuelle en 3D. Traitée comme un polytope pour la visualisation.
- **Ellipsoid** : l'ellipsoïde 3D, une sphère étirée selon les axes.
- **Torus** (Tore) : tore de révolution. Ce n'est pas une sphère : il possède des parallèles, des méridiens et des cercles de **Villarceau** , mais ni hyperméridiens ni coques.

Note : **Sphere** , **Ellipsoid** et **Torus** sont des quadriques et disposent à ce titre des réglages de tracé décrits en 2.1.6.

### 2.1.3. Polychora (polychores, 4D)

- **Pentachoron** (Pentachore, 5-Cellules) : simplexe 4D (5 sommets). Le polychore 4D le plus simple.
- **Tesseract** (8-Cellules) : hypercube 4D (16 sommets). L'analogue 4D du cube.
- **4-Box** : hypercuboïde 4D (16 sommets). L'analogue 4D du cube avec des arêtes de longueur irrégulière.
- **Hexadecachoron** (Hexadécachore, 16-Cellules) : orthoplexe 4D (8 sommets). Dual du Tesseract.
- **Icositetrachoron** (Icositétrachore, 24-Cellules) : polychore 4D sans analogue 3D. Auto-dual.
- **Hecatonicosachoron** (Hécatonicosachore, 120-Cellules) : polychore 4D avec 120 cellules dodécaédriques.
- **Hexacosichoron** (Hexacosichore, 600-Cellules) : polychore 4D avec 600 cellules tétraédriques.
- **Glome** (3-Sphère) : analogue 4D d'une sphère.
- **3-Ellipsoid** (Ellipsoïde 4D) : analogue 4D d'un ellipsoïde.

- **Toratope** : la famille des tores en dimension 4. C'est ici que la famille est la plus riche : outre le tore de Clifford ( $T^2$ , plat), elle compte quatre membres qui sont des 3-variétés dans  $\mathbb{R}^4$  sans analogue dans aucune autre dimension. Le membre affiché se choisit avec **Torus Type** (voir 2.1.6).

### 2.1.4. N-Polytopes ( $N > 4$ )

Pour tous les polytopes généralisés, un slider **Dimension** permet de choisir le nombre d'axes spatiaux.

- **N-Simplex** : simplexe généralisé en N dimensions ( $N+1$  sommets).
- **N-Cube** : hypercube généralisé en N dimensions ( $2^N$  sommets). Les coordonnées sont des permutations de  $(\pm 1, \dots, \pm 1)$ .
- **N-Box** : hypercuboïde généralisé en N dimensions ( $2^N$  sommets).
- **N-Orthoplex** : polytope croisé généralisé en N dimensions ( $2N$  sommets). Les coordonnées sont des permutations de  $(\pm 1, 0, \dots, 0)$ .
- **N-Demicube** : polytope obtenu à partir d'un hypercube, en supprimant un sommet sur deux.
- **E•n Roots** (racines de  $E_n$ ) : structures basées sur les vecteurs du système de racines  $E_n$  (base pour la génération des polytopes de Gosset). Limité à la dimension 8, les systèmes de racines exceptionnels s'arrêtant à  $E_8$ .
- **Gosset K•21** : polytopes exceptionnels de Gosset jusqu'à 8 dimensions.
- **N-Sphere** : sphères en dimension N.
- **N-Ellipsoid** : ellipsoïde en dimension N.
- **N-Torus** :  $T^n \subset \mathbb{R}^{2n}$ , une famille de cercles par facteur. Le tore plat n'a pas de stratification radiale : il n'expose donc pas de coques.

### 2.1.5. Special

Cette catégorie réunit des structures qui ne relèvent d'aucune famille régulière.

- **Elysion Vault (3D)** (Coffre-fort d'Élysion) : forme 3D pointue obtenue à partir de 4 pyramides, utilisée comme logo par Laurent Oliversen, concepteur de Tetrachore. Deux réglages spécifiques, **Vault Gap X** et **Vault Gap Y**, modifient l'espacement entre les pyramides.
- **Elysion Cell (3D)** (Cellule d'Élysion) : prisme pentagonal droit — le motif du pavage personnalisé Elysion 120-Cell. Deux réglages spécifiques : **Cell Height** (hauteur de la cellule) et **Cell Gap** (espacement entre cellules).
- **Complex Hopf (4D)** : la fibration de Hopf sur  $\mathbb{C}$ , c'est-à-dire  $S^3 \rightarrow S^2$ . Construite en dimension 4, quelle que soit la position du slider **Dimension**.
- **Quatern. Hopf (8D)** : la fibration de Hopf sur les quaternions  $\mathbb{H}$ ,  $S^7 \rightarrow S^4$ . Construite en dimension 8.
- **Octon. Hopf (16D)** : la fibration de Hopf sur les octonions  $\mathbb{O}$ ,  $S^{15} \rightarrow S^8$ . Construite en dimension 16.

Note : contrairement aux sphères que l'on peut *tracer* comme une fibration, ces trois dernières formes **sont** la fibration. Elles sont toujours construites à la dimension de leur algèbre ( $\mathbb{C} \rightarrow 4$ ,  $\mathbb{H} \rightarrow 8$ ,  $\mathbb{O} \rightarrow 16$ ).

### 2.1.6. Réglages des quadriques (sphères, ellipsoïdes, tores)

Les formes quadriques — sphères, ellipsoïdes, tores et fibrations — ne sont pas décrites par une liste de sommets mais par des familles de courbes. Elles partagent le groupe de réglages suivant.

#### Mode de tracé

- **Tracing Mode** : mode de représentation. Les bases disponibles dépendent de la forme :
  - **Mesh** (maillage) : représentation à base de parallèles, méridiens et hyperméridiens.
  - **Stratified** (stratification) : imbrication de sphères de rang inférieur ; l'enchevêtrement est une autre manière de représenter la dimension supplémentaire.
  - **Hybrid** (hybride, dimension  $> 4$ ) : combine maillage et stratification.
  - **Hopf Fibration** : partition du globe en grands cercles entrelacés qui ne se touchent pas. L'entrée est renommée selon la fibration réellement tracée : **Hopf Fibration** en dimension 4, **Quaternionic Hopf** en dimension 8, **Octonionic Hopf** en dimension 16, et **Pseudo-Hopf** pour les ellipsoïdes — car il ne s'agit alors pas d'une vraie fibration, mais d'une déformation a posteriori de celle-ci.
- **Use Edges** : bascule la représentation de courbes vers des arêtes. Chaque famille est alors construite avec des arêtes et des sommets, ce qui permet la localisation des sommets et la gestion des couleurs, au prix de la courbure des cercles.
- **Resolution** : nombre de sommets par cercle. Contrôle la finesse du tracé.

#### Familles de cercles

Trois sliders comptent les cercles tracés. **Leur nom change selon la forme sélectionnée**, car ce qu'ils comptent n'est pas le même objet :

| Forme                                   | Slider 1        | Slider 2        | Slider 3        |
|-----------------------------------------|-----------------|-----------------|-----------------|
| Sphères, ellipsoïdes, glomes            | Parallels       | Meridians       | Hypermeridians  |
| Torus (tore de révolution 3D)           | Parallels       | Meridians       | Villarceau      |
| Toratope / N-Torus — Clifford           | Ring 1          | Ring 2          | Extra Rings     |
| Toratope — Tiger                        | Ring 1          | Ring 2          | Tube Circles    |
| Toratope — Spheritorus                  | Parallels       | Meridians       | Hypermeridians  |
| Toratope — Torisphere                   | Parallels       | Meridians       | Tube Circles    |
| Toratope — Ditorus                      | Outer Ring      | Inner Ring      | Tube Circles    |
| Fibrations $\mathbb{H}$ et $\mathbb{O}$ | Fiber Density 1 | Fiber Density 2 | Fiber Density 3 |

Sur les fibrations quaternionique et octonionique, les trois sliders ne comptent pas des familles distinctes : chacun pilote un sous-ensemble arbitraire des directions de fibres, et le découpage change avec le **Fiber Mode**. Les libellés numérotés sont donc les seuls honnêtes.

## Stratification

- **Shells** : nombre de coques (sphères de rang inférieur) constituant les sections, pour les modes stratifiés.
- **Shell Distance** : distribution des coques au sein de l'enveloppe. Plus la valeur est grande, plus les coques intérieures se tassent contre la coque extérieure ; plus elle est petite, plus les coques s'éloignent les unes des autres en se contractant vers le centre.

## Mode hybride (dimension > 4)

- **Suspension Depth** (profondeur de suspension) : étirement dimensionnel puis contraction polaire. Ce facteur modélise le nombre de dimensions représentées par une interconnexion (hyperméridienne) de glomes.
- **Glomes** : nombre de glomes par suspension dimensionnelle.
- **Susp. Hyper Mode** : mode de représentation des interconnexions entre les glomes d'une suspension — **Radial** (radiale), **Peripheral** (périphérique) ou **Helical** (hélicoïdale).

## Hyperméridiens et fibres

- **Hypermeridian Mode** : représentation des hyperméridiens en mode **Mesh** — **Magnetic** (magnétique), **Interlaced** (entrelacé), **Tilted** (incliné), **Equatorial** (équatorial) ou **Polar** (polaire).
- **Alternate Hyper-Modes** : disponible pour le tracé **Mesh** en dimension > 4. Alterne le mode de représentation des hyperméridiens selon la dimension.

- **Fiber Mode** : sélection du motif d'échantillonnage des fibres, en mode Hopf. Les modes offerts dépendent de la forme et de la dimension.

## Tores

- **Torus Type** : membre de la famille des tores à construire. **Clifford** est plat et existe en toute dimension ; les quatre autres — **Tiger**, **Spheritorus**, **Torisphere**, **Ditorus** — sont des 3-variétés dans  $\mathbb{R}^4$  et ne sont donc proposés qu'en dimension 4. Ramener **Tube Thickness** à 0 sur un Tiger redonne exactement le tore de Clifford : c'est bien la même famille.
- **Tube Thickness** : épaisseur du tube, exprimée comme le rapport  $r/(R+r)$  et normalisée pour que la figure conserve sa taille globale. Le slider s'arrête à la *corne* de la forme, c'est-à-dire là où le trou se referme.

## Étirements

Les formes étirables (ellipse, ellipsoïdes, boîtes, tores) disposent des réglages **Scale X**, **Scale Y**, **Scale Z** et **Scale W**, qui étirent la figure dans les dimensions souhaitées. Au-delà de la dimension 4 s'y ajoutent :

- **Scale V Base** : étirement dans la 5ème dimension. Sert de valeur de référence pour toutes les dimensions  $> 4$ .
- **Scale V Factor** : facteur d'étirement exponentiel dans les dimensions  $> 5$ .

Note : seules les quadriques tracées en mode maillage (**Mesh**) sont éligibles à la tessellation.

### 2.1.7. Opérateurs de graphe

Quatre opérateurs transforment le graphe sommets/arêtes du solide sélectionné. Leur éligibilité dépend de la forme exacte : la troncature exige une figure de sommet régulière, l'alternation un graphe biparti, le composé un solide non centralement symétrique.

- **Truncation** (troncature) : coupe chaque sommet au ratio  $t$  le long de ses arêtes. 0 = solide d'origine ; 0.5 = rectifié, c'est-à-dire aux milieux d'arêtes (par exemple cube  $\rightarrow$  cuboctaèdre). Réservé aux formes régulières exactes.
- **Alternate** (alternation de Coxeter) : supprime un sommet sur deux — cube  $\rightarrow$  tétraèdre, tesseract  $\rightarrow$  16-cellules,  $n$ -cube  $\rightarrow$   $n$ -demicube. Prend le pas sur **Truncation**.
- **Compound** (composé) : superpose au solide son inversion centrale ( $v \rightarrow -v$ ) — tétraèdre  $\rightarrow$  stella octangula, 5-cellules  $\rightarrow$  composé de deux 5-cellules. Prend le pas sur **Truncation**.
- **Positive Coords** : translate chaque axe pour que son minimum se place à 0. L'origine de la scène se retrouve dans un coin du solide au lieu de son centre (par exemple tesseract  $\pm 1 \rightarrow 0/2$ ).

À cela s'ajoutent deux réglages généraux du groupe **Generator** :

- **Type** : catégorie de formes proposée dans le dropdown **Shape** — **All** , **2D** , **3D** , **4D** , **ND (5+)** et **Special** . Les deux dernières ne sont offertes que si le plan souscrit donne accès aux dimensions supérieures.
- **Hypersolid Scale** : échelle uniforme du solide.

## 2.1.8. Tessellation (Pavage)

### 2.1.8.1. Tiling Types (types de pavages)

**Enable Tessellation** active le pavage. Pour remplir l'espace en répliquant les polytopes, vous pouvez ensuite choisir entre différents modes d'agencement via **Tiling Type** :

- **Prism** (prismes)
- **Honeycomb** (nids d'abeille) — proposé uniquement pour les formes éligibles
- **Swarm** (essaims)
- Les réseaux personnalisés propres à la forme sélectionnée (voir 2.1.8.5)

### 2.1.8.2. Prism (prisme)

Les prismes sont définis par un polytope de base répliqué dans une seule direction. Les sommets correspondants de chaque réplique sont reliés entre eux.

- **Nb Slices** (nombre de tranches) : nombre de répliques
- **Flat Base** (base plane) : le polytope qui sert de base au prisme est aplani dans le plan sélectionné — **None** , **XY** , **XZ** ou **YZ**
- **Tile Rotation** (rotation des tuiles) : chaque réplique tourne dans un plan orthogonal à la direction de cisaillement ; rotation proportionnelle à la distance à l'origine
- **Rotation Factor** : amplitude de la rotation, fonction de la distance à l'origine
- **Slice Shear X** , **Slice Shear Y** , **Slice Shear Z** : espacement entre les répliques, selon les axes X, Y et Z

### 2.1.8.3. Honeycomb (nid d'abeille)

Seules certaines formes sont éligibles au pavage de type « nid d'abeilles », comme les cubes ou les hexagones. Contrairement aux autres assemblages (prismes et essaims), ces pavages sont compacts et remplissent parfaitement l'espace sans vide.

- **Nb Slices X** , **Nb Slices Y** , **Nb Slices Z** : nombre de tranches dans les directions X, Y et Z

Les pavages 2D (du plan XY) disposent de deux réglages supplémentaires pour entrer dans la troisième dimension (ensembles disjoints de tranches dans la direction Z) :

- **Gap Z** : espace les tranches de pavage selon l'axe Z

- **Nexus Z** : connecte les sommets correspondants des répliques appartenant aux différentes tranches de Z

#### 2.1.8.4. Swarm (essaim)

Les essaims sont une généralisation des prismes et nids d'abeille. Ils sont les plus flexibles et personnalisables. Ce sont des tranches de répliques disjointes dans toutes les directions, pouvant être connectées par des nexus (arêtes qui relient les sommets correspondants de chaque réplique).

En termes de réglages, l'essaim réunit ceux des prismes et des nids d'abeille :

- **Nb Slices X**, **Nb Slices Y**, **Nb Slices Z** : nombre de tranches dans toutes les directions (idem nids d'abeille)
- **Nexus X**, **Nexus Y**, **Nexus Z** : un système de connexion dans toutes les directions
- **Gap X**, **Gap Y**, **Gap Z** : l'espacement entre les tranches se gère dans toutes les directions

À cela s'ajoutent les réglages propres aux prismes : **Slice Shear X**, **Slice Shear Y**, **Slice Shear Z**, **Flat Base**, **Tile Rotation** et **Rotation Factor**.

#### 2.1.8.5. Réseaux personnalisés

Certains polytopes ont des réseaux prédéfinis, proposés dans **Tiling Type** lorsque la forme correspondante est sélectionnée. Par exemple :

Pour les cubes :

- **Tesseract Net** (Patron de tesseract) : croix 3D constituée de N-cubes

Pour l'ElySION Vault :

- **ElySION Seal** (Sceau d'ÉlySION) : assemblage circulaire de Vaults
- **ElySION 2-Seal** : double assemblage circulaire de Vaults
- **ElySION 3-Seal** : triple assemblage circulaire de Vaults

## 2.2. Hypersolid Style

L'application utilise un système de coloration automatique pour aider à identifier les dimensions.

Les arêtes sont colorées en fonction de leur alignement sur les axes principaux dans l'espace d'origine en haute dimension.

Par défaut :

- X : Vert
- Y : Bleu
- Z : Orange
- W : Violet

- V5, V6, V7, V8, ... : Gris

À noter que ces couleurs sont déterminées de manière systématique pour les structures ayant des arêtes parallèles aux axes du repère. Cela s'applique donc aux hypercubes. Pour les figures ayant des arêtes inclinées dans plusieurs dimensions, la gestion des couleurs est plus subtile.

Les réglages stylistiques sont rangés dans quatre dossiers : **Vertices** (sommets), **Edges** (arêtes), **Faces** et **Background** (arrière-plan). Les effets de profondeur ne forment pas un dossier à part : chaque famille d'objets porte les siens, à la suite de ses réglages de style.

## 2.2.1. Vertices (sommets)

### 2.2.1.1. Style de base

- **Show Vertices** : affiche ou masque les sommets.
- **Vertex Info** : rend les sommets cliquables. Un clic ouvre le panneau d'information du point (voir 6.5).
- **Vertex Size** : taille des sommets.
- **Vertex Shape** : forme du marqueur — **Dot** (point), **Square** (carré), **Diamond** (losange), **Cross (+)** ou **Cross (x)**.
- **Vertex Color** : couleur des sommets.

### 2.2.1.2. Profondeur des sommets

L'activation des effets de profondeur se fait en cochant la case **Vertex Depth**. On peut alors contrôler l'opacité ("Alpha") des éléments géométriques en fonction de leur profondeur dans l'espace 4D/3D. Ces commandes se trouvent juste sous la case **Vertex Depth**. Selon l'objet dessiné, la profondeur est évaluée différemment :

- **Depth Quantiles** (quantiles de profondeur) : Discrétise les distances par rapport à l'origine en un ensemble de compartiments (ex : 10 quantiles), en les rendant dans des couches alpha distinctes via une approche X-Ray pour éviter les collisions visuelles (Z-Fighting).
- **Outer Alpha** (alpha extérieur) : Opacité de base pour les éléments les plus externes des objets.
- **Inner Alpha** (alpha intérieur) : Opacité de base pour les éléments les plus internes (le cœur) des objets.

Note : Le système crée automatiquement un dégradé linéaire d'opacité entre les éléments extérieurs et intérieurs. Si Outer Alpha < Inner Alpha, le dégradé est désactivé et la valeur Outer Alpha est appliquée uniformément à toutes les profondeurs.

- **Invert Alpha** (inverser l'alpha) : une bascule mathématique qui inverse l'opacité. Utile pour masquer l'extérieur et n'observer que le cœur solide.

- **Alpha Cutoff** (seuil alpha) : un seuil de rendu strict appliqué après le calcul final de l'alpha (et l'inversion). Tout élément dont l'opacité finale tombe en dessous de cette valeur est entièrement retiré du pipeline de rendu (masqué), ce qui améliore la clarté visuelle et les performances de rendu.

### 2.2.2. Edges (arêtes)

Les arêtes disposent des réglages suivants :

- **Show Edges** : affiche ou masque les arêtes.
- **Edge Info** : rend les arêtes cliquables, sur le même principe que **Vertex Info**.
- **Edge Width** : épaisseur du trait.
- **Edge Default Color** : couleur unique appliquée quand la coloration par axe est désactivée.

La coche **Edge Colors** permet de définir une couleur par type d'arête. Si elle est décochée, alors toutes les arêtes portent la couleur définie par **Edge Default Color**.

Si elle est cochée, alors le réglage suivant permet de gérer les couleurs pour des arêtes dont l'orientation n'est pas parallèle aux axes.

#### 2.2.2.1. Color Tie Break (gestion des égalités)

Vous pouvez définir les 5 couleurs associés à chaque axe du repère :

- **X Axis**, **Y Axis**, **Z Axis**, **W Axis**, **Vi Axes (i > 4)**.

Ces couleurs sont, selon certains objets, utilisés différemment (typiquement pour les sphères).

La liste **Color Tie Break** (gestion des égalités) :

Lorsqu'une arête n'est pas parfaitement alignée avec un seul axe (ex: arêtes diagonales d'un triangle), la couleur est déterminée par une règle de départage (choix entre 3 règles) :

- **Vertex** : Couleur basée sur l'index du sommet de début/fin.
- **Edge** : Couleur basée sur l'index de l'arête dans la liste.
- **Hash** : Couleur aléatoire cohérente basée sur les propriétés de l'arête.

Cette règle s'applique à la plupart des polytopes. Les rendus sont difficiles à anticiper, mais ils permettent généralement de distribuer les couleurs de manière différente.

Pour les sphères, d'autres techniques sont proposées :

Pour les sphères/ellipsoïdes en mode Mesh/Stratified Edges :

- **Circle** : seulement ; reproduit le même système de couleur par type de cercle que celui utilisé par les tracés de type Mesh/Stratified Circles.

Pour les sphères/ellipsoïdes en mode Hopf Fibration (complex, quaternionic, octonionic) :

- **Fiber** : attribue une couleur par typologie de cercles (associés à des nombres imaginaires différents).
- **Fano** : coloration dérivée du plan de Fano, qui encode la table de multiplication des octonions. Réservé à la fibration octonionique.
- **Polar** : attribue un dégradé de 3 couleurs (parmi les couleurs Axis X/Y/Z) pour distinguer les cercles proches du pôle nord, pôle sud et équateur.
- **Bipolar** : attribue deux dégradés de 3 couleurs (parmi les couleurs Axis X/Y/Z/W/V) pour distinguer les cercles proches du pôle nord, pôle sud et équateur.
- **Tripolar** : idem avec trois dégradés.
- **Quadripolar** : idem avec quatre dégradés

### 2.2.2.2. Contours des sphères

Comme les sphères sont générées différemment, elles utilisent une logique de colorisation spécifique qui dépend de son mode de tracé ( **Tracing Mode** ) :

Cas des Mesh Edges/Circles (maillages) : Les couleurs du maillage sont définies par les couleurs de la section Edges :

- Parallèles : couleur définie par **X Axis** ,
- Méridiens : couleur définie par **Y Axis** ,
- Hyperméridiens : couleur définie par **Z Axis** .

#### Cas de Hopf Fibration (fibration de Hopf)

Cette visualisation avancée utilise l'application de Hopf pour projeter la 3-Sphère ( $S^3$ ) comme un faisceau de cercles (dits "fibres") au-dessus d'une 2-Sphère ( $S^2$ ).

- Base : Une 2-Sphère où chaque point correspond à un cercle complet (fibre) dans la 3-Sphère.
- Fibres : Cercles disjoints qui remplissent l'espace 4D. L'échelle de 12 fibres permet de visualiser la structure sans encombrement.
- Coloration : les fibres de Hopf complexe (4D) sont colorées en fonction de leur position sur la 2-Sphère de base (longitudinale/latitudinale), révélant la structure de torsion. Elles utilisent uniquement les couleurs définies par **X Axis** et **Y Axis** . Les fibres de Hopf quaternionique et octonionique utilisent des systèmes de couleurs par typologie de cercles (soit en lien avec la structure mathématique, soit en lien avec leur proximité aux pôles).

#### Stratified Mode (équateur & pôles)

Dans ce mode, la 3-Sphère est décomposée en une série de coques 3D parallèles (collection de sphères classiques), de la même manière qu'un globe est tranché par des lignes de latitude. Les couleurs dépendent de la position de la coque dans l'imbrication.

- *Equator (Outer) Shell* (coque équatoriale extérieure = géométrie la plus large) : La "surface" de la sphère 4D. Représentée par X/Y. Géométriquement, c'est la coque la plus grande. Les parallèles et méridiens de cette sphère utilisent les couleurs définies par **X Axis** et **Y Axis** .

- *Pole (Inner) Shell (coque polaire intérieure = géométrie la plus petite)* : Le "cœur" de la sphère 4D. Représenté par Z/W. Ce sont géométriquement des points ou de très petits cercles. Les parallèles et méridiens de cette sphère utilisent les couleurs définies par **Z Axis** et **W Axis**.
- *Coques intermédiaires* : les couleurs des coquilles non extrêmes sont interpolées entre les deux précédentes : il s'agit donc d'un dégradé entre les coques intérieures et extérieures.

### Note : Artéfact de stratification en mode perspective

Pourquoi les coques intérieures semblent-elles parfois plus grandes que les coques extérieures ?

Lors de l'utilisation de la Projection Perspective, les objets plus proches de la caméra 4D apparaissent significativement plus grands. Si la **4D Distance** est petite (caméra proche de l'objet) :

- Une coque *Intermédiaire* (qui est géométriquement plus petite que l'équateur) se trouvait plus près de la caméra.
- Le grossissement perspectif peut faire apparaître cette coque intérieure *visuellement plus grande* que la coque équatoriale extérieure, allant potentiellement jusqu'à l'"engloutir".

Les couleurs restent correctes (car elles sont basées sur le rang) :

- La coque colorée en X/Y est toujours l'équateur géométrique (Rang 1).
- La coque colorée en Z/W est toujours le pôle géométrique (Rang N).

Si vous voyez une coque Z/W apparaître plus grande qu'une coque X/Y, il s'agit d'une pure illusion d'optique causée par la perspective 4D. Solution : modifier la *4D Distance* ou passer en *Projection Orthographique* pour restaurer la hiérarchie visuelle intuitive.

### 2.2.2.3. Profondeur des arêtes

L'activation des effets de profondeur se fait en cochant la case **Edge Depth**.

Les arêtes disposent des mêmes réglages que les sommets pour gérer la profondeur : **Depth Quantiles**, **Outer Alpha**, **Inner Alpha**, **Invert Alpha**, **Alpha Cutoff**. Voir les réglages de profondeur des sommets pour plus de détails.

### 2.2.2.4. Profondeur dans des sphères

On opère une distinction entre les arêtes (droites) et les coques (courbes).

#### Objets avec arêtes (polytopes et sphères utilisant le mode de tracé de type Mesh Edges)

L'opacité est calculée par distance sommet/arête par rapport à l'origine *dans l'espace 3D projeté* (pas l'espace 4D d'origine). Les distances sont triées et réparties dans *N* compartiments de quantiles.

Chaque compartiment reçoit une valeur alpha interpolée linéairement entre Outer Alpha et Inner Alpha :

Note sur la profondeur de la 2/3-Sphère :

Même si tous les points d'une 3-sphère sont équidistants de l'origine 4D par définition, la projection 4D→3D brise cette isotropie. En projection *perspective*, les cercles proches du pôle W (grand W) sont compressés vers l'origine 3D, tandis que les cercles équatoriaux (petit W) sont projetés plus loin. En projection *orthographique*, un parallèle proche du pôle nord se situe près de l'origine 3D, tandis qu'un parallèle équatorial se trouve à une distance unitaire (la plus grande). Le dégradé de profondeur sur un Glome en mode *Mesh Edges* est donc réel et physiquement significatif : il reflète la structure de profondeur 4D telle qu'elle est vue à travers la projection.

### Sphères avec coques (Stratified Edges et Stratified Circles)

La profondeur est évaluée structurellement en fonction de l'index mathématique de la coque (de 1 pour l'équateur géométrique à N pour le pôle). Comme les coques sont naturellement des topologies imbriquées, le dégradé alpha est continu et exact. Note : le réglage **Depth Quantiles** ne s'applique *pas* aux coques.

### Fibration de Hopf : profondeur basée sur les coques (sans arêtes)

La Fibration de Hopf est rendue comme un ensemble de \*cercles de fibres\* imbriqués topologiquement (boucles fermées), et non comme des arêtes discrètes entre sommets. Chaque cercle appartient à une coque mathématique (une couche de tore), numérotée du tore le plus externe à la fibre centrale la plus interne. L'alpha de profondeur est calculé directement à partir de l'index de la coque ; aucun tri de distance Z en 3D n'est nécessaire. La formule est la même que celle utilisée pour les objets Stratifiés/Coques. Comme les fibres de Hopf sont générées dans un ordre topologique strict (extérieur → intérieur), la structure mathématique encode intrinsèquement la profondeur. Cela rend la gestion de la profondeur exacte, continue et parfaitement cohérente avec les commandes Inner/Outer Alpha, sans aucune heuristique géométrique.

### Limitations - Mesh Circles uniquement

Dans le mode de tracé *Mesh Circles* pour la 2/3-Sphère, toute la géométrie des cercles est fusionnée en un seul maillage combiné sans shader de profondeur par cercle. L'alpha de profondeur ne s'applique *pas* aux cercles, seulement aux sommets. En revanche, le mode *Mesh Edges* pour les Sphères *prend bien en charge* la profondeur car un shader GPU évalue profondeur = distance par sommet au moment du rendu.

## 2.2.3. Faces

La bascule **Show Faces** superpose des polygones remplis sur le fil de fer du polytope. Le moteur prend en charge deux modes fondamentaux pour la détection et la génération des faces :

### 2.2.3.1. Méthodes de détection des faces

L'application utilise des générateurs algorithmiques spécialisés pour identifier les véritables faces géométriques du polytope. La méthode utilisée dépend de la famille du polytope. Les techniques utilisées sont :

- **Détection de cycles par graphe (DFS)** : utilisée pour les solides réguliers 3D/4D standard, le *N-Simplexe* et le *N-Cube* (jusqu'à  $N=5$ ). Une recherche en profondeur (Depth-First Search) explore le graphe des arêtes pour identifier les boucles fermées de 3, 4 ou 5 arêtes.
- **Recherche combinatoire par axes** : spécifique à la famille des *N-Orthoplexes* (N-Cross). Il identifie les triplets de sommets situés sur des axes orthogonaux distincts pour former des faces triangulaires parfaitement régulières. Ce mode est actif jusqu'à  $N=6$ .
- **Logique de sous-espace de parité** : spécifique à la famille des *N-Demicubes* (Groupe  $D_n$ ). Il identifie les sous-espaces cubiques 3D du hypercube parent et extrait les faces tétraédriques qui préservent la parité alternée des sommets. Ce mode est actif jusqu'à  $N=5$ .
- **Échantillonnage stochastique (dans les autres cas)** : pour les polytopes exceptionnels complexes (Gosset *E6*, *E7*, *E8*) ou tout objet dépassant les seuils de dimension mentionnés ci-dessus, la détection topologique est automatiquement contournée au profit d'un échantillonnage aléatoire afin de maintenir des performances interactives.

### 2.2.3.2. Rendu des faces

- **Show Faces** : affiche ou masque les faces.
- **Face Type** : mode de rendu.
  - **Solid** (mode plein) : les faces sont triangulées et rendues comme des surfaces remplies.
  - **Wireframe** (fil de fer) : seules les arêtes de bordure et les diagonales internes des faces identifiées sont dessinées. Cela révèle le « pavage » du polytope.
- **Wire Width** : épaisseur des traits, en mode **Wireframe**.
- **Face Color** : couleur des faces.
- **Face Alpha** : transparence des faces, quand les effets de profondeur sont désactivés.

### 2.2.3.3. Échantillonnage des faces

Pour les objets complexes ou de haute dimension où la détection topologique est trop coûteuse, le moteur peut échantillonner de manière stochastique des paires de sommets pour former des triangles (ou des quadrilatères/pentagones).

Dans le paragraphe suivant : une face est dite *petite* (respectivement *grande*) lorsque les sommets qui la délimitent sont, en moyenne, à une distance *inférieure* (respectivement *supérieure*) à une valeur donnée. La notion de distance se réfère à la distance euclidienne entre un sommet et l'origine du polytope (le barycentre dans la plupart des cas).

L'échantillonnage s'active avec **Face Sampling**. Les réglages associés sont les suivants :

- **Large/Small Cutoff** : ce paramètre est une distance normalisée qui permet de classifier les faces en deux catégories : petites ou grandes. Ce seuil est compris entre 0 et 1 car il représente le pourcentage de la plus grande distance observée parmi toutes les faces. Par exemple, un seuil à 0.5 signifie que les sommets d'une face doivent se trouver, en moyenne, à une distance (à l'origine) *supérieure*(respectivement *inférieure*) à 0.5 de la distance maximale (la face considérée comme "la plus éloignée") pour que la face soit considérée comme *grande*(respectivement *petite*).
- **Large Faces** (grandes faces) : nombre de faces classifiée *grandes* qui seront retenues (échantillonnées) pour la colorisation.
- **Small Faces** (petites faces) : nombre de faces classifiée *petites* qui seront retenues (échantillonnées) pour la colorisation.

Note pour les Polytopes Réguliers : Pour les polytopes parfaitement réguliers (Cubes, Tesseracts) où tous les sommets sont équidistants du barycentre, la population "Small" est mathématiquement vide car le seuil de distance normalisé est constant. Dans ce cas, seul l'échantillonnage des faces "Large" produira des résultats.

#### 2.2.3.4. Profondeur des faces

L'activation des effets de profondeur se fait en cochant la case **Face Depth** .

##### Z-Sorting & Alpha (Algorithme du Peintre)

Pour appliquer la profondeur aux faces, le moteur réorganise dynamiquement les faces de l'arrière vers l'avant à chaque image. Cela permet l'utilisation de dégradés de transparence basés sur la profondeur :

- **Front Face Alpha** : opacité des faces les plus proches de la caméra (la « peau »).
- **Back Face Alpha** : opacité des faces les plus éloignées de la caméra (le « cœur »).
- **Invert Face Alpha** : inverse le dégradé, sur le même principe que **Invert Alpha** pour les sommets et les arêtes.

Exemple d'usage : régler **Back Face Alpha** = 1.0 et **Front Face Alpha** = 0.2 vous permet de voir l'intérieur solide de la structure à travers une couche externe translucide.

### Couverture par type de polytope :

- ✓ Solides Platoniciens (3D) : triangles, carrés, pentagones
- ✓ Archimédiens / Catalan (cuboctaèdre, dodécaèdre rhombique) : types de polygones mixtes
- ✓ Polychores 4D : via détection de cycles par graphe
- ✓ N-Simplexe / N-Cube / N-Orthoplexe : jusqu'à la résolution disponible
- ✗ Sphère / Glome : Non applicable, rendu sous forme de cercles
- ✓ Gosset K.21 dim=5 (D5) : triangles + carrés, nombre gérable
- ✗ Gosset K.21 dim≥6 (E6, E7, E8) : les polytopes de Gosset E6, E7, E8 possèdent un nombre de 2-faces (faces planes) qui croît de manière exponentielle. Pour E6, le calcul des faces est rapide, mais projeter 2160 faces qui se chevauchent depuis la 6D crée un maillage opaque illisible sans gain visuel. Pour E7 et E8, le calcul prendrait plusieurs secondes et le résultat est tout aussi illisible. Seul le D5 (dim=5, ~40 sommets) expose ses faces, comme un compromis pratique entre l'information visuelle et la performance.

## 2.2.4. Arrière-plan

Cette section permet d'afficher des éléments de fond ou en surimpression.

### 2.2.4.1. Couleur et dégradé du fond

- **Transparent** : retire le fond. La scène est alors rendue sur un fond transparent, ce qui est utile pour les exports PNG destinés à être superposés.
- **Color** : couleur unie de l'arrière-plan, utilisée quand **Use Gradient** est décoché.
- **Use Gradient** : remplace la couleur unie par un dégradé à trois arrêts.

Le dégradé se règle ensuite avec :

- **Gradient Type** : **Linear** (linéaire) ou **Radial** (radial).
- **Color 1 (Top)**, **Color 2 (Mid)**, **Color 3 (Bottom)** : les trois couleurs du dégradé, du haut vers le bas.
- **Gradient Angle** : orientation du dégradé linéaire.
- **Mid Scale** : étendue de la couleur intermédiaire. Plus la valeur est grande, plus la bande centrale occupe de place au détriment des deux extrêmes.
- **Mid H Position** / **Mid V Position** : position horizontale et verticale du centre de la couleur intermédiaire.
- **Smoothing** : adoucit les transitions entre les trois arrêts.

### 2.2.4.2. Halo

Le halo est une source lumineuse diffuse peinte dans le fond, indépendante du dégradé. Il sert à détacher le solide de l'arrière-plan sans éclairer la scène 3D.

- **Halo** : active le halo.

- **Halo Color** : couleur du halo.
- **Halo Intensity** : intensité lumineuse.
- **Halo Scale** : rayon du halo.
- **Halo H Position** / **Halo V Position** : position horizontale et verticale du centre du halo.

### 2.2.4.3. Grille d'alignement

L'option **Alignment Grid** affiche une grille tangente aux sommets périphériques, afin d'ajuster les alignements lors des rotations par exemple.

### 2.2.4.4. Axes et plans de repère

Localisation : onglet **Workspace** , section **Exp** , dossier **Scene Markers** . Ces marqueurs décrivent la scène et non le solide, mais ils accompagnent aussi bien la géométrie que les données projetées.

#### Axes

Les axes permettent de se repérer dans les dimensions spatiales.

- **Show Scene Axes** : affiche les axes de l'espace tridimensionnel absolu (vu par la caméra), |X|, |Y| et |Z|. Ils sont encadrés de barres verticales pour ne pas les confondre avec les axes relatifs ci-après.
- **Show Relative Axes** : affiche les axes relatifs, solidaires du polytope et des éventuels points de données projetés. Quand on effectue un cisaillement ou une rotation, ces axes se décalent et tournent en même temps que le polytope et les données. Bien noter qu'ils ne représentent plus les coordonnées spatiales absolues, mais les coordonnées après projection/transformation : X, Y, Z, W, V5, V6, ..., V10.
- **Negative Extension** : prolonge les axes également du côté des valeurs négatives.
- **Negative Arrow** : ajoute une pointe aux axes du côté des valeurs négatives.
- **Axis Length** : longueur des axes (des flèches).
- **Axis Label Size** : taille des étiquettes (X, Y, Z...).
- **Axis Arrow Size** : taille de la pointe des axes.
- **Axis Ticks** : affichage des graduations unitaires (coordonnées 1, 2, 3...) pour les axes absolus de la scène et les axes relatifs des données/polytope. Les graduations sont différentes selon les axes, car les coordonnées relatives sont transformées par la projection spatiale.
- **Axis Grid** : ajoute un quadrillage dans les plans portés par les axes, pour lire les coordonnées à vue.
- **Axis Projection** : trace la projection orthogonale de chaque point sur les axes, sous forme de traits de rappel.

#### Plans

- **Plane |XY|** , **Plane |XZ|** , **Plane |YZ|** : affichage des plans du repère absolu (axes de la scène).
- **Plane Alpha** : opacité des plans affichés.

- **Plane Occlusion** : les plans masquent ce qui se trouve derrière eux, au lieu de rester translucides.
- **Plane Back Fade** : atténue la face arrière d'un plan quand la caméra passe de l'autre côté, pour que le plan ne dissimule pas la scène.
- **Back Alpha** : opacité résiduelle de cette face arrière atténuée.

## 2.3. Projection

La projection s'effectue dans cet ordre :

- Projection des dimensions supérieures à 4 vers l'espace 4D (X, Y, Z, W)
- Projection de l'espace 4D vers la 3D (les coordonnées X, Y, Z intègrent les vestiges de la coordonnée W par des effets de perspective)
- Éventuellement, un effondrement (projection orthographique) de la 3D vers la 2D
- Le tout s'affiche dans une scène 3D, au travers de la caméra

Dans l'interface, les réglages sont répartis par typologie d'opérations mathématiques : rotation, cisaillement, projection (avec distance focale). Mais conceptuellement, les dimensions sont traitées en cascade.

### 2.3.1. Views

Cette section permet d'appliquer des visualisations mathématiques spécifiques. Ces valeurs de rotation révèlent des projections caractéristiques pour chaque polytope.

Les vues ci-dessous permettent de capturer les éléments suivants :

- Topologique : Graphe de connectivité (sommets, arêtes)
- Combinatoire : Incidences entre k-faces
- Métrique : Angles dièdres, longueurs relatives
- Global : Périodicité, enveloppe, stratification

#### 2.3.1.1. Mode de projection unifiée

Les trois boutons ci-dessous permettent d'affecter tous les modes de projection (ND-4D, 4D-3D et Caméra) d'un seul coup :

- **All Perspective** : unifie l'ensemble du pipeline (Projection ND, Projection 4D, et Caméra) en mode *Perspective*. Cela garantit une profondeur focale cohérente tout au long de la réduction dimensionnelle.
- **All Ortho** : unifie l'ensemble du pipeline en mode *Orthographic*, révélant les proportions géométriques réelles sans distorsion.

- **Ortho-Perspective** : mode mixte. Les réductions dimensionnelles (ND→4D et 4D→3D) restent orthographiques — les proportions du polytope sont donc préservées — tandis que seule la caméra 3D passe en perspective. On obtient une scène lisible en profondeur sans que la structure du polytope soit déformée par deux étages de fuite successifs.

### 2.3.1.2. Projection Front et Petrie

Ces deux projections sont respectivement la plus simple et la plus complexe à mettre en œuvre. Elles disposent d'un bouton pour un accès immédiat.

- **Front** : réinitialise tous les angles de rotation et cisaillements (ND et 4D) à 0, et restaure la caméra à son orientation frontale par défaut. C'est le point de départ de toute analyse visuelle.
- **Petrie** : Configure les rotations et la projection pour montrer le *Polygone de Petrie* du polytope. Il s'agit d'une projection orthographique spécifique où le polygone projeté du polytope forme un périmètre régulier à N côtés. L'axe est aligné sur le polygone qui forme le plus long cycle hamiltonien zigzaguant entre toutes les facettes. C'est la seule vue qui dévoile la structure globale : elle révèle la périodicité de grande échelle, invisible dans les vues locales. Elle demeure unique en ce qu'elle capture la périodicité hamiltonienne, sans équivalent direct ailleurs.

### 2.3.1.3. Plans canoniques

Naviguer entre les différents plans canoniques en cliquant sur les boutons **◀ Prev** (précédent) et **Next ▶** (suivant). Ils incluent :

- **Vertex-first** (par le sommet) : l'axe passe par un sommet. Révèle la valence locale et le groupe de symétrie ponctuel : combien d'arêtes rayonnent, avec quelle symétrie angulaire.
- **Edge-first** (par l'arête) : l'axe est aligné sur une arête. Révèle les angles dièdres entre les deux (N-1)-facettes qui partagent cette arête - la "courbure locale" de l'objet. Souvent négligée, irremplaçable pour distinguer des polytopes à vues vertex-first similaires. Note : une arête délimite exactement 2 facettes de codimension 1 dans un polytope régulier convexe ; en général ce nombre dépend de la structure locale.
- **Cell-first** / **k-Facet-first** (par la cellule puis par la facette de dimension  $k < N$ ) : l'axe passe par le centre d'une cellule ou d'une facette de dimension 3, 4, ..., N-1. Révèle la stratification concentrique : comment les cellules/facettes s'organisent en couches successives autour du centre ("en oignon").
- **Schlegel** : correspond à une vue (N-1)-facet-first en mode perspective. Elle permet d'apercevoir l'ensemble des cellules d'un seul coup.
- **Equatorial** : correspond à une projection de Petrie, suivie d'une orbite de 90° dans le plan vertical, comme si on regardait le Petrie par la tranche. Elle permet de voir la stratification des facettes autour du barycentre.

Note : à l'exception de la vue Schlegel, toutes les vues des plans canoniques ne sont possibles qu'en mode de projection orthographique.

#### 2.3.1.4. Permutations de la base

Permuter les axes de la base trouvent son sens dans les analyses factorielles, pour visualiser la distribution des points selon 3 axes simultanés, parmi les N axes disponibles (X, Y, Z, W, V5, V6...). Voir la section des analyses factorielles pour plus de détails.

Chaque vue pré-enregistrée se traduit par une rotation de 90° selon un, deux ou trois axes. Cela permet, visuellement, de remplacer un axe de base (X, Y, Z) par un autre axe disponible.

Les 20 combinaisons les plus usuelles (jusqu'en dimension 6) sont proposées. La première correspond à la vue de l'espace XYZ. La dernière traduit la vue de l'espace W.V5.V6.

Naviguer entre les différents espaces factoriels en cliquant sur les boutons ◀ **Prev** (précédent) et **Next** ▶ (suivant).

Pour mieux appréhender les rotations, il conviendra d'afficher les axes relatifs ( **Show Relative Axes** ) qui traduisent l'orientation du polytope et des données en surimpression.

#### 2.3.1.5. Collapse (effondrement)

##### Étape optionnelle d'écrasement 2D/1D.

Dédiée à l'onglet de visualisation Factor, cette étape mathématique finale permet de compresser la scène 3D entièrement projetée (Polytope, données Zodiac/Astro/MBTI/Social, et données personnalisées) vers un plan 2D ou un axe 1D. Ceci est accompli en forçant à zéro une ou deux dimensions de la sortie 3D immédiatement avant le rendu.

##### Projections sur Plan (2D) :

- **XY** : Émule une vue aplatie en écrasant la profondeur ( $z = 0$ ).
- **XZ** : Aplatit l'axe vertical ( $y = 0$ ).
- **YZ** : Aplatit l'axe horizontal ( $x = 0$ ).

##### Projections sur Axe (1D) :

- **X**, **Y**, **Z** : Écrase deux dimensions simultanément pour projeter l'ensemble du cluster structurel sur une seule ligne. Ceci est particulièrement utile pour identifier les corrélations, la variance ou les écarts-types le long d'une composante principale spécifique.

Note : Comme cette projection est appliquée comme un filtre final sur la scène 3D plutôt que dans l'espace mathématique d'origine, les dimensions supérieures comme W (profondeur 4D) ou V (5D+) ne peuvent pas être isolées ici - elles sont déjà repliées dans X, Y et Z par les étapes précédentes. De plus, pour maintenir la lisibilité lorsque les points sont mathématiquement écrasés sur des coordonnées spatiales identiques, les étiquettes textuelles (Zodiaque/Astro/MBTI/Social et personnalisées) appliquent un décalage spatial intelligent (dérivé de leur coordonnée perdue d'origine) pour éviter les chevauchements visuels.

### 2.3.1.6. Vues personnalisées

Certains solides disposent de vues personnalisées (que l'on peut modifier). Naviguer entre les différentes vues en cliquant sur les boutons ◀ **Prev** (précédent) et ▶ **Next** (suivant).

### 2.3.2. ND vers 4D

Cette étape réduit les dimensions  $N > 4$  vers la 4D (X, Y, Z, W).

Le pipeline cascade de toutes les dimensions supérieures s'effectuent ainsi :

- Rotation initiale dans tous les plans ND sélectionnés
- Enchaînement de cisaillements (Shear) et de projections (à chaque étage de la cascade dimensionnelle)

Note : par sa logique additive non destructive, le cisaillement est beaucoup plus efficace que la rotation pour déplier les dimensions supérieures dans l'espace 3D visible (projeté sur les trois premiers axes X, Y, Z). Car les rotations dans des hautes dimensions peuvent s'effectuer dans des plans de rotation orthogonaux au plan de projection de chaque dimension vers les suivantes, compressant toute tentative de déploiement visuel.

La coche **Lock Hypersolid Rotation** permet de verrouiller le polytope et de le rendre insensible aux rotations (attention, le solide reste sensible aux cisaillements et à la distance focale des projections). Utile pour analyser des données, en faisant apparaître, par rotation, les axes relatifs des dimensions invisibles (W, V5, V6...) sans toucher au solide en tant que tel. Si un cube ou une sphère est affichée pour servir de repère, ces derniers restent immobiles durant les rotations, tandis que les données projetées (nuage de points), elles, continuent de pivoter avec les axes relatifs.

#### 2.3.2.1. Rotation ND

- **Circular Rotation** : effectue une rotation simultanée dans XV5, YV6, ZV7, WV8, V5V9, V6V10.
- **Rotation Factor** : Contrôle la décroissance exponentielle des rotations pour les dimensions supérieures.
- **Rotation Plane** qui contient une liste avec tous les plans de rotation supplémentaires. Il suffit d'en sélectionner un, puis d'ajuster le curseur qui apparaît juste en dessous.

#### 2.3.2.2. Cisaillement ND

- **Circular Shear** (cisaillement circulaire) : Ajoute une composante géométrique rotationnelle (alternance de directions selon X, Y, Z) au cisaillement linéaire de la 4D, créant des motifs en spirale pour les dimensions supérieures à 4.
- **Shear Factor** : Contrôle la décroissance exponentielle des décalages de cisaillement pour les dimensions supérieures.
- **Shear Plane** : sélection d'un plan spécifique pour effectuer un cisaillement précis selon ce plan uniquement.

### 2.3.2.3. Projection ND

Le mode se choisit avec le dropdown **ND -> 4D** .

#### Perspective

Les dimensions supérieures ( $D > 4$ ) sont mises à l'échelle par un facteur de perspective. Les objets ayant des coordonnées plus élevées dans ces dimensions sont projetés plus près du centre 4D, tout en appliquant un cisaillement cumulatif.

- **Base ND Distance** : détermine la distance de référence (de la caméra) pour la "perspective" des dimensions supérieures. Plus la valeur est élevée, plus les couches externes se rapprochent de l'origine (éloignement de la caméra, diminution de la perspective).
- **ND Distance Factor** : ajuste l'échelle exponentielle de la distance perspective à travers les dimensions. Une valeur élevée engendre une uniformisation des espacements entre les dimensions. Une valeur faible provoque un plus grand espacement des dimensions élevées.

En manipulant conjointement **Base ND Distance** et **ND Distance Factor** , il est ainsi possible de personnaliser la perspective global, ainsi que l'espacement ou le rapprochement entre les dimensions (ressenti compression/dilatation).

#### Orthographic

Réduction parallèle vers la 4D. La distance perspective et l'accumulation de profondeur sont strictement ignorées (W reste inchangé). La structure est révélée purement par le cisaillement linéaire. Les notions de distance (focale) sont ignorées. Il ne reste que les paramètres de cisaillement pour déplier la structure.

#### Oblique

La projection oblique n'est pas proposée dans les dimensions supérieures à 4, car elle devient très lourde à configurer. Une projection orthographique accompagnée d'une rotation ou d'un cisaillement (circulaire ici) est équivalent à une projection oblique.

### 2.3.3. 4D vers 3D

L'étape de projection vise à projette les points 4D ( $x, y, z, w$ ) dans l'espace 3D ( $x', y', z'$ ). La coordonnée W représente la "profondeur 4D".

Le traitement commence par un cisaillement linéaire 4D optionnel (dossier **Shear** , curseurs **XW** , **YW** , **ZW** ) vers les véritables coordonnées 3D, basé sur la profondeur W. Cela permet d'incliner l'objet 4D dans notre dimension physique avant le passage final de la caméra perspective ou orthographique.

#### 2.3.3.1. Rotation 4D

La rotation du polytope peut s'effectuer dans l'ensemble des plans 3D et 4D. Le dossier **Rotation** expose un curseur par plan : **XY** , **XZ** , **YZ** , **XW** , **YW** , **ZW** . Les trois derniers n'apparaissent qu'à partir de la dimension 4.

Note : les noms de ces curseurs suivent le layout d'axes actif (voir 4.3.3). Sous un layout renommant les axes, **XY** devient le couple correspondant.

### 2.3.3.2. Cisaillement 4D

Le cisaillement du polytope peut s'effectuer dans l'ensemble des plans 4D. Le dossier **Shear** expose un curseur par plan : **XW** , **YW** , **ZW** . Chacun déplace la position 3D en fonction de la profondeur 4D (W).

### 2.3.3.3. Projection 4D

#### Perspective

Perspective 4D standard. Les objets avec un w plus élevé (plus éloignés en 4D) apparaissent plus petits et convergent vers le centre.

Le mode se choisit avec le dropdown **4D -> 3D** . La **4D Distance** règle la distance focale de la caméra 4D. Des valeurs plus petites créent une distorsion de perspective plus forte.

#### Orthographic

Projection parallèle à W. La coordonnée W est complètement ignorée pour la position des points dans la 3D. Utile pour préserver les tailles géométriques réelles.

Important : en l'absence de rotation ou de cisaillement, il devient donc impossible de visualiser les composantes 4D. Il convient d'utiliser les curseurs de rotation et de cisaillement pour déployer la structure.

#### Oblique

Déplie la structure 4D en déplaçant les points en fonction de leur coordonnée W le long d'un angle 2D spécifique sur l'écran. Géométriquement, la projection oblique équivaut à une projection orthographique combinée à un cisaillement, mais elle est proposée comme un mode simplifié distinct, typiquement pour reproduire les perspectives Cavalier/Cabinet (cavalière).

- **4D Angle** : contrôle la direction du déplacement 4D dans le plan 3D XY.
- **4D Scale** : contrôle l'amplitude du déplacement (jusqu'où la profondeur 4D est « extraite »).

Note : le cisaillement manuel **XW** / **YW** / **ZW** est ignoré en mode Oblique 4D puisque l'angle et l'échelle obliques définissent entièrement le cisaillement 4D. Donc les cisaillements sont redondants.

## 2.3.4. Camera

La phase finale rend le fil de fer 3D sur votre écran 2D. Notez que la projection 2D/1D dans l'onglet Factor est effectuée avant le rendu de la caméra. C'est pourquoi nous pouvons toujours orbiter autour du polytope et des points écrasés.

### 2.3.4.1. Camera Type

Le type se choisit avec le dropdown **Type** du dossier **Camera** .

- **Perspective** : caméra 3D standard avec points de fuite. Les objets éloignés de la caméra apparaissent plus petits.
- **Orthographic** : projection parallèle. Les objets conservent leur taille quelle que soit la distance. Idéal pour les vues techniques.
- **Oblique** (cavalier/cabinet) : Un mode spécial qui combine la projection orthographique avec un cisaillement de l'axe Z (profondeur). L'oblique fournit une perspective cavalière où la face avant conserve sa forme réelle, tandis que la profondeur est représentée par des lignes diagonales. Deux paramètres supplémentaires apparaissent :
  - **Angle** : Direction de l'axe simulant la profondeur.
  - **Scale** : Amplitude du déplacement simulant la profondeur.

#### 2.3.4.2. Viewport

##### Fenêtre de vue (Viewport)

- **Roll** : Fait pivoter l'ensemble de la vue (roulis de la caméra) autour de la ligne de visée.

Note : il n'y a pas de cisaillement pour la caméra. Cela apporterait des inclinaisons indésirables aux polytopes.

#### 2.3.4.3. Camera Monitoring

Ces moniteurs se trouvent en bas du dossier **Camera** et affichent les coordonnées courantes de la caméra :

- **Orbit Phi** : orbite dans le plan horizontal.
- **Orbit Theta** : orbite dans le plan vertical.
- **Zoom** : niveau de zoom courant.

Données purement informatives (non modifiables). Pour modifier la caméra, il suffit d'utiliser la souris ou le trackpad.

Le dossier expose enfin une bascule de diagnostic :

- **Polytope Logs** : écrit dans la console du navigateur le détail de la génération et de la projection du polytope courant (nombre de sommets, d'arêtes, de faces, angles appliqués). Utile pour signaler un comportement inattendu.

## 2.4. Motion

### Uniquement pour le forfait Hyper.

L'application fournit des outils pour visualiser la géométrie en haute dimension par le mouvement.

### 2.4.1. Dynamic Effects

Ce dossier regroupe les effets « vivants » de la scène : leur pulsée sur le point sélectionné, leur de valence, et déplacements animés de la caméra.

- **Dynamic** : interrupteur général de ces effets. Décoché, la scène reste statique et les déplacements de caméra sont instantanés.
- **Pulse Speed** : vitesse de la pulsation lumineuse sur le point sélectionné.
- **Pulse Radius** : rayon maximal de la pulsation, exprimé en multiple de la taille du point sélectionné. Le rayon minimal est la taille du point elle-même.

### 2.4.2. Orbit (déplacements automatiques de caméra)

Certaines actions déplacent la caméra toutes seules : un double-clic sur un point, l'arrivée sur un plan canonique, un changement de permutation de base. Ce dossier décide de ce qui déclenche ces déplacements et de la manière dont ils se déroulent.

#### Déclencheurs

- **Best Plan Selection** : un double-clic sur un point (et sur les étapes intermédiaires de **Mid Point States**) fait pivoter les axes vers son *Best Cos<sup>2</sup> Plan*. Combinable avec **Magnet Selection** ; les deux décochés, le double-clic ne fait plus rien.
- **Magnet Selection** : oriente la caméra vers le point sélectionné, en le centrant au plus près. Lorsqu'une arrivée sur un *Best Plan* a lieu, ce réglage remplace la vue de face ; seul, le double-clic amène simplement la caméra sur le point.

#### Trajectoire des déplacements

Par défaut un déplacement automatique va droit au but. Les réglages ci-dessous lui ajoutent du mouvement, pour rendre la transition lisible plutôt qu'abrupte.

- **Orbit Adjustment** : interrupteur général des six curseurs ci-dessous. Décoché, aucun trajet supplémentaire n'est ajouté.
- **Pre-Motion Phi** : trajet de phi supplémentaire (en degrés) parcouru *avant* que le déplacement ne se stabilise. 360 correspond à un tour complet. Le signe est tiré au hasard ( $\pm$ ) à chaque geste.
- **Pre-Motion Theta** : idem pour theta.
- **Pre-Motion Zoom** : coup de zoom appliqué en cours de route, avant stabilisation. Positif = plus près, négatif = plus loin.
- **Post-Motion Phi** : décalage d'orbite (en degrés) ajouté *après* le déplacement, pour que l'arrivée ne soit pas parfaitement centrée. Signe tiré au hasard.
- **Post-Motion Theta** : idem pour theta.
- **Post-Motion Zoom** : ajustement du zoom d'arrivée.

## Scope (portée)

Le sous-dossier **Scope** désigne les gestes auxquels ces trajets supplémentaires s'appliquent :

- **Best Projection Plan** : les boutons *Best Contrib* / *Best Cos<sup>2</sup> Plan*.
- **Canonical Plan** : les boutons ◀ **Prev** / ▶ **Next** des plans canoniques.
- **Base Permutation** : les boutons ◀ **Prev** / ▶ **Next** des permutations de base.
- **Double-Click** : le double-clic sur un point, et les étapes intermédiaires.

Note : ces trajets supplémentaires ne sont visibles qu'en mode **Dynamic**.

### 2.4.3. Auto-Rotate (auto-rotation)

Fait pivoter automatiquement le polytope dans jusqu'à trois plans simultanés.

- **Auto-Rotate** : active la rotation automatique.
- **Rotation Speed** : contrôle la vitesse angulaire. Prend en charge les valeurs négatives pour inverser la direction.
- **Rotation Plane 1**, **Rotation Plane 2**, **Rotation Plane 3** : les plans de rotation à activer (XY, XZ, XW, YZ, YW, ZW, ou les plans ND pour  $N > 4$ ). La liste proposée dépend de la dimension du solide courant.

### 2.4.4. Auto-Shear (auto-cisaillement)

Fait osciller harmoniquement les facteurs de cisaillement pour faire « pulser » les dimensions supérieures dans la vue.

- **Auto-Shear** : active le cisaillement automatique.
- **Shear Amplitude** : amplitude de l'oscillation.
- **Shear Speed** : fréquence de l'oscillation.
- **Shear Plane 1**, **Shear Plane 2**, **Shear Plane 3** : les plans de cisaillement à activer (XW, YW, ZW, ou les plans ND pour  $N > 4$ ).

### 2.4.5. Interpolation

Effectue une transition fluide entre les préréglages d'état globaux.

- **Interpolate** : active la chaîne d'interpolation.
- **Start State** : l'état initial de la transition — **Front** ou **Petrie**.
- **Mid State** : un état intermédiaire optionnel — **None**, **Front** ou **Petrie**. Réglé sur **None**, la transition va directement de Start à End.
- **End State** : l'état cible final — **Front** ou **Petrie**.

- **Mid Point States** : liste d'étiquettes de points, séparées par des virgules (correspondance par inclusion). Chaque point devient une étape intermédiaire : la caméra fait pivoter les axes pour le centrer, exactement comme un double-clic. Ces étapes s'intercalent dans la chaîne.

### Chaîne d'interpolation

Un clic sur **Interpolation** déplace d'abord la vue actuelle vers **Start State**, puis vers **Mid State** (si activé) et les éventuelles étapes de **Mid Point States**, avant de se stabiliser sur **End State**.

### Rythme de la transition

- **Ease In/Out** : active une transition douce, en courbe « en S », plutôt qu'une vitesse constante.
- **Duration (s)** : durée de chaque segment de la transition.
- **Start Pause (s)** : temps d'arrêt sur l'état de départ avant de repartir.
- **Mid Pause (s)** : temps d'arrêt sur chaque étape intermédiaire.
- **End Pause (s)** : temps d'arrêt sur l'état d'arrivée.

### 2.4.6. Oscillation

**Oscillate** : si l'oscillation est activée, l'interpolation ne s'arrête pas à **End State**. Elle inverse la séquence et boucle indéfiniment entre Start, Mid et End. Les pauses de début, de milieu et de fin s'appliquent à chaque passage.

## 2.5. Quelques exemples de projection

### 2.5.1. Tesseract

#### Tesseract

Le tesseract dispose d'un grand nombre de projections célèbres.

La vue des *deux hexagones centrés* (disponible dans les **Custom Views** du panneau Projection) :

- Rotation : (XY, XZ, YZ) = (30°, 35.26°, 135°)
- Projection 4D perspective avec une projection caméra orthographique
- Utiliser la distance 4D pour jouer sur l'espacement entre les deux hexagones

La vue en *perspective cavalière double* (disponible dans les **Custom Views** du panneau Projection) :

- Projection 4D oblique avec une projection caméra oblique

## 3. Ingestion de données

---

### 3.1. Dataset

Localisation : onglet **Workspace** , section **Dataset** .

#### 3.1.1. Synthetic Data

Ces jeux de données sont aléatoires ou open source. Ils permettent de tester l'application sans avoir à importer de données.

- **5D Data** : données à 5 variables numériques pseudo-aléatoires
- **12D Data** : données à 12 variables numériques pseudo-aléatoires
- **Titanic Data** : données open source
- **Titanic Categ. Data** : idem que précédemment, mais avec des données discrétisées ; permet de tester les ACM
- **Titanic Mixed Data** : idem que précédemment, mais avec des données mixtes ; permet de tester les AFDM
- **MMM Data** : données pour tester le forecast et le MMM
- **Hy-Perf Single Data** : données pour tester le module Hy-Performance, avec effet simple
- **Hy-Perf Multi Data** : données pour tester le module Hy-Performance, avec effets multiples
- **Clear Imported Data** : vide le jeu de données courant et remet la scène à son état initial

#### 3.1.2. Reference Data

Ces jeux de données de référence sont purement cosmétiques et servent de repère visuel. Chacun est un menu déroulant dont la première entrée, **None** , désactive l'affichage.

- **Zodiac Data** : **Astro** (les 12 signes), **Malgovert** (les 13 clusters psychiques du roman éponyme), **Elysion 120-C Loop** et **Elysion 120-C H3**
- **Semiology Data** : **12 Circuits** et **6 Circuits** , les circuits neurologiques qui conditionnent le comportement, ainsi que **HBDI** (les quatre quadrants cérébraux de Herrmann)
- **MBTI-DISC-SS Data** : les référentiels comportementaux utilisés en entreprise — **MBTI (3D)** , **MBTI (4D)** , **Social Styles (3D)** , **DISC (3D)** et **DISC x SS (4D)** , ce dernier croisant DISC et Social Styles sur quatre axes
- **Task Management Data** : **Management Styles** et **Managerial Grid** (la grille de Blake et Mouton)

- **Clear Selection** : remet les quatre menus sur **None** d'un seul coup. Le bouton n'apparaît que si au moins un jeu est sélectionné.

Instant fiction : dans la catégorie Zodiac se trouve un jeu de données dit « Malgovert » qui correspond à la classification psychique des treize clusters du Zodiaque dans la série Malgovert de Laurent Oliversen. C'est un peu comme une classification en groupe sanguin, mais avec trois variables. Les points sont définis selon 3 coordonnées qui représentent la sensibilité des profils zodiacaux aux 3 sérums élyséens. La position spatiale de ces clusters est véritablement utilisée dans la fiction.

### 3.1.3. Custom Points

Cette section permet de saisir manuellement jusqu'à 10 coordonnées spatiales (X, Y, Z, W, V5, ..., V10). Le point résultant est affiché en surimpression sur le polytope. Vous pouvez lui associer une étiquette, le dupliquer ou le supprimer.

## 3.2. Points & Labels

Localisation : onglet **Workspace** , section **Dataset** , dossiers **Points** et **Labels** .

Ces réglages gouvernent l'apparence de tous les points de données projetés dans la scène, quelle que soit leur provenance, ainsi que celle de leurs étiquettes.

### 3.2.1. Familles de points

L'application distingue six familles de points, qui s'affichent et se colorent indépendamment :

| Famille                     | Origine                                       |
|-----------------------------|-----------------------------------------------|
| Imported                    | les lignes du jeu de données importé          |
| Reference                   | les jeux de données de référence (3.1.2)      |
| Factor Individuals          | les individus projetés dans le plan factoriel |
| Factor Numeric Features     | les variables numériques projetées            |
| Factor Categorical Features | les modalités qualitatives projetées          |
| Cluster                     | les centroïdes issus du clustering            |

### 3.2.2. Affichage

Le dossier `Points / Display` expose une case par famille :

`Show Imported Pts` , `Show Reference Pts` , `Show Fact Indiv Pts` , `Show Fact Num Pts` , `Show Fact Categ Pts` , `Show Cluster Pts` .

Le dossier `Labels / Display` fait de même pour les étiquettes :

`Show Imported Lbls` , `Show Reference Lbls` , `Show Fact Indiv Lbls` , `Show Fact Num Lbls` , `Show Fact Categ Lbls` , `Show Cluster Lbls` .

Afficher les points sans les étiquettes est le réglage courant sur les gros volumes ; l'inverse — étiquettes seules — sert à lire un plan factoriel dense.

### 3.2.3. Style des points

- `Point Shape` : forme du marqueur — `Dot` (point), `Square` (carré), `Diamond` (losange), `Cross (+)` ou `Cross (x)` .
- `Point Size` : taille de base des points.

### 3.2.4. Encodage par variable

Trois canaux visuels peuvent être pilotés par une colonne du jeu de données, plutôt que d'être fixes. C'est le moyen de faire porter à la scène une information supplémentaire sans ajouter d'axe.

- `Variable Point Size` : la colonne qui pilote la taille des points. `Variable Max Size` fixe la taille atteinte par la valeur maximale.
- `Variable Point Color` : la colonne qui pilote la couleur des points.
- `Variable Point Opacity` : la colonne qui pilote l'opacité. L'opacité encode naturellement une notion de certitude ou d'inclusion : privilégiez donc des métriques d'incertitude (erreur standard relative, valeur absolue de la marge, silhouette).
  - `Opacity Invert` : inverse la correspondance, pour qu'une valeur faible donne un point opaque — par exemple pour qu'une faible incertitude reste pleine.
  - `Min Opacity` : plancher d'opacité appliqué à la valeur minimale. Il n'est jamais nul, afin qu'un point ne disparaisse jamais silencieusement.
- `Hide Var From Space` : retire du calcul des coordonnées spatiales la variable utilisée pour l'encodage. Sans cette option, la variable serait comptée deux fois — une fois comme axe, une fois comme canal visuel.

### 3.2.5. Dégradé de couleur

Le dossier `Points / Gradient` définit la rampe de couleurs utilisée par `Variable Point Color` . Un moniteur `i` rappelle en tête de dossier la variable en cours d'encodage.

- **Diverging Colors** : utilise une rampe divergente (négatif → milieu → positif) ancrée sur 0, au lieu d'une rampe séquentielle min → max. C'est le bon choix dès que le zéro a un sens — un écart, une marge, un uplift.
- **Seq Min** / **Seq Max** : les deux couleurs de la rampe séquentielle.
- **Div Neg** / **Div Mid** / **Div Pos** : les trois couleurs de la rampe divergente.

### 3.2.6. Couleurs fixes

Quand aucune variable ne pilote la couleur, chaque famille porte sa couleur propre.

- **Points / Color** : **Reference Point** , **Data Point** , **Factor Indiv Point** , **Factor Feature Point** .
- **Labels / Color** : **Reference Label** , **Data Label** , **Factor Indiv Label** , **Factor Feature Label** .

### 3.2.7. Style des étiquettes

- **Label Size** : taille du texte.
- **Label Angle** : orientation de l'étiquette par rapport au point.
- **Label Offset** : distance entre l'étiquette et son point. Augmenter cet écart est le moyen le plus simple de désempiler des étiquettes qui se chevauchent.

## 3.3. Data Import

Localisation : onglet **Data** .

### 3.3.1. Import CSV

Vous pouvez charger des données en CSV avec le bouton **Upload CSV** . En fonction du plan souscrit, les données sont limitées à un certain nombre de lignes et de colonnes (indiquées dans l'interface). Le bouton **Clear Data** vide le jeu importé.

Les structures attendues sont :

- Format CSV UTF-8
- La première ligne correspond à des libellés (header)
- Séparateur : virgule, point-virgule, tabulation ou barre verticale "|"
- Les libellés de colonnes suivants sont réservés à des usages spécifiques : "label" (nom du point), "segment" (pour les statistiques inférentielles) et "target" (pour le machine learning supervisé)

Par défaut, si des colonnes x, y, z, w, v5, v6, v7... sont présentes dans le fichier, elles sont automatiquement assignées aux coordonnées spatiales correspondantes.

Si aucun libellé ne correspond à la convention précédente, alors les variables numériques sont assignées dans l'ordre de présence dans le fichier : la première colonne numérique est assignée à X, la seconde à Y, la troisième à Z, etc.

En cas de mélange (exemple : a, x, y, z, b, c), alors les libellés reconnus (x, y, z) sont assignés aux 3 coordonnées spatiales, et les libellés inconnus (a, b, c) sont assignés aux coordonnées spatiales suivantes, dans cet ordre : a -> w, b -> v5, c -> v6.

### 3.3.2. Role Mapping

Cet encart permet de définir les colonnes qui joueront un rôle spécifique pour l'analytique. Certains rôles sont interchangeable dans certains cas de figure.

Les colonnes pouvant jouer des rôles sont :

- **label** : nom qui sera affiché au-dessus du point sur la scène projetée
- **color** : couleur du point et du label
- **segment** : colonne de groupement utilisée pour les statistiques descriptives ou exploratoires (profiling)
- **target** : colonne de variable dépendante à expliquer, utilisée pour la sélection de variables (corrélation) et pour les algorithmes supervisés (LDA, Sup ML)
- **treatment** : colonne binaire pour le calcul des effets incrémentaux et uplifts entre deux groupes (profiling, Hy-Performance) ; exemples : essais cliniques, campagnes marketing...
- **flag\_test** : colonne binaire (0, 1) pour définir les lignes qui serviront à entraîner les modèles supervisés et celles pour les phases de test

Une fois les rôles assignés, cliquez sur **Apply Roles**.

### 3.3.3. Axes Mapping

Vous pouvez assigner les variables numériques sur les axes spatiaux de votre choix.

Il vous suffit de renseigner une combinaison Variable / Axe et de cliquer sur **Add**.

L'option **Merge T|C columns on same axes** n'est disponible que pour les données Hy-Performance. Elle permet de mutualiser les variables suffixées par \_T et celles par \_C. Par exemple, plutôt que d'avoir deux axes pour afficher NB\_T et NB\_C, l'option n'affiche qu'un seul axe NB qui fusionne les deux populations (T et C). Pour distinguer les deux populations, il faut soit afficher les labels, soit se rendre dans **Data Style**, et cocher la case **Colorize Treatment** (section **Class Style**).

### 3.3.4. Sampling (échantillonnage)

Sur un gros fichier, il est souvent préférable de travailler d'abord sur un échantillon : la scène reste fluide et les modèles s'ajustent en quelques secondes.

- **Size Mode** : mode d'expression de la taille — **Number** (un nombre de lignes) ou **Proportion** (une fraction du fichier).
- **Size Value** : la taille voulue, exprimée selon le mode choisi.
- **Sample Mode** : mode de prélèvement — **Random** (tirage aléatoire), **Head** (les premières lignes) ou **Tail** (les dernières). **Head** et **Tail** sont utiles sur des données chronologiques, où l'ordre du fichier porte du sens.
- **Sample** : applique l'échantillonnage.
- **Clear** : rétablit le fichier complet.

### 3.3.5. Test Split

Si la colonne **flag\_test** n'est pas définie, il est possible de la créer et de la remplir automatiquement de manière aléatoire en définissant un pourcentage de test ( **Test Size** ) et en cliquant sur **Random Split** . Le moniteur **Counts** affiche la répartition obtenue entre apprentissage et test. **Clear** supprime le découpage.

### 3.3.6. Prévisualisation

Une prévisualisation des données est disponible dans l'onglet.  
Dans l'en-tête se trouvent deux sections.

#### Comptage

- **Total rows** : nombre de lignes dans le jeu importé
- **Total columns** : nombre de colonnes dans le jeu importé
- **Numeric columns** : nombre de colonnes numériques
- **Categorical columns** : nombre de colonnes catégorielles
- **Rows retained** : nombre de lignes retenues selon le plan de l'utilisateur (pricing)
- **Spatial dim retained** : nombre de colonnes spatiales retenues (selon le plan de l'utilisateur) et donc visualisées dans la scène

#### Colonnes réservées

Cette liste de tags renvoie aux rôles réservés. Elle permet d'identifier les colonnes qui jouent un rôle particulier : uid, label, color, target, segment, treatment, flag\_test

#### Typage des colonnes

Pour chaque colonne, un tag indique le type de la colonne : numérique, catégorielle ou booléenne.

## 3.4. Preprocessing (pré-traitement)

Localisation : onglet **Data** .

Cette section permet d'effectuer un certain nombre de transformations parfois nécessaires pour exécuter des algorithmes (projections factorielles, clustering, Sup ML).

La coche **Keep Original Column** permet de conserver la colonne d'origine en plus de la colonne transformée : la source est gardée et la transformation est écrite dans une nouvelle colonne **<col>\_t**, que **Undo** supprime. Décochée, la transformation écrase la colonne sur place. Note : conserver l'original n'est pas toujours souhaitable, car cela induit une redondance d'informations.

Un champ **Select Column** permet de choisir la colonne à transformer ; chaque entrée y est suffixée de son type et de son état d'activation. Chaque transformation s'effectue et s'annule au moyen de boutons **Apply** et **Undo**.

Enfin, une coche **Re-Apply in Inference** permet de conserver le paramétrage de ces étapes pour les réutiliser plus tard sur de nouvelles données ; étape nécessaire pour l'utilisation de modèles entraînés (Factor, Clustering, Sup ML) sur un jeu transformé.

Le bouton **Clear Preprocessing** annule d'un coup l'ensemble des transformations appliquées au jeu de données.

### Ordre des étapes

Les transformations s'appliquent dans un ordre fixe, qui est celui des sections ci-dessous : typage → désactivation → imputation → winsorization → discrétisation → regroupement de modalités → encodage disjonctif. Les paramètres sont toujours estimés sur les seules lignes d'apprentissage ( **flag\_test** différent de 1), puis appliqués à toutes les lignes.

### 3.4.1. Typage des colonnes

Trois boutons sont disponibles pour changer le format des colonnes.

- **Make Numeric** : transforme la colonne en donnée numérique ; se projette spatialement dans la scène
- **Make String** : transforme la colonne en donnée qualitative ; ne se projette pas
- **Make Bool** : transforme la colonne en donnée binaire (true/false, yes/no, 0/1) ; ne se projette pas dans la scène ; correspond à des indicatrices ou des « dummies »

Les données string et bool sont très proches dans leur implication au sein des différentes fonctionnalités. La seule distinction vient du fait qu'une variable bool sera bien utilisée dans les modèles de régression linéaire et logistique (Sup ML), alors que les données string seront ignorées.

Utilisation ou non de la colonne :

- **Activate** : autorise la variable dans la scène et les différentes modélisations
- **Deactivate** : exclut la colonne

### 3.4.2. Imputation

Cette section permet de remplir les données vides.

- **Num Imputation** : valeur utilisée pour remplir les données numériques vides — `-1`, `0`, `mean` (moyenne des données non vides) ou `median` (médiane des données non vides)
- **Categ Imputation** : chaîne de caractères utilisée pour remplir les données catégorielles vides

Chacune des deux dispose de ses propres boutons **Apply** et **Undo**.

### 3.4.3. Winsorization

Cette section permet de caper les valeurs numériques extrêmes.

- **Tail Cut %** : pourcentage capé sur les valeurs hautes ; une valeur de 2% signifie que toutes les valeurs supérieures au 98e centile seront capées.
- **Two-Sided** : cape également les valeurs extrêmes basses ou négatives

### 3.4.4. Discretization

Cette section permet de discrétiser les données numériques en intervalles. Elles deviennent donc catégorielles (qualitatives). La discrétisation s'effectue en quantiles.

- **Nb Bins** : nombre d'intervalles (quantiles)

### 3.4.5. Category Binning (regroupement de modalités)

Cette section permet de regrouper les catégories qui ont des effectifs faibles et induisent de la rareté, donc de l'instabilité. C'est l'équivalent des valeurs extrêmes basses pour des données qualitatives.

- **Rare Category Label** : libellé de la catégorie fusionnée
- **Use Frequency** : utilise les proportions d'effectif plutôt que le comptage absolu
- **Count Threshold** : seuil en nombre d'occurrences en dessous duquel une catégorie est jugée rare. Utilisé quand **Use Frequency** est décoché.
- **Freq Threshold** : le même seuil, exprimé en fréquence. Utilisé quand **Use Frequency** est coché.

### 3.4.6. One-Hot Encoding (encodage disjonctif complet)

Cette section permet de transformer une variable catégorielle en indicatrices. Chaque catégorie de la variable devient une colonne binaire (bool, 0/1).

- **Keep all categories** : conserve toutes les catégories et les transforme en binaires ; si l'option est décochée, la dernière catégorie est ignorée pour éviter la multicollinéarité.

## 4. Exploration de données

---

### 4.1. Stats

Localisation : onglet **Stats** .

#### 4.1.1. Statistiques descriptives

Contient des analyses descriptives des données importées dans l'application :

- Tableau récapitulatif des métriques de dispersion : moyenne, écart-type, valeurs min et max, quartiles (Q1, Q2, Q3), coefficient de variation (CV), écart semi-interquartile (SIQR =  $(Q3-Q1)/2$ ), midhinge (MH =  $(Q1+Q3)/2$ ), trimoyenne (TM =  $(Q2 + MH)/2$ ), coefficient de dispersion interquartile (QCD = SIQR/MH)
- Boxplots (boîtes à moustache) : pour chacune des variables numériques, avec possibilité d'afficher une échelle distincte par variable ( **Different Scales** ), ou bien une échelle commune pour toutes les variables ( **Same Scale** ).
- Category Distribution : comptage des modalités présentes dans chacune des variable catégorielle, avec une liste pour choisir la variable

#### 4.1.2. Statistiques inférentielles

Vous pouvez effectuer des tests statistiques rapides : comparaisons de moyennes ou proportions, ainsi que des estimations de taille d'échantillon.

##### 4.1.2.1. Paramètres

- **Use Case** : comparaison de moyennes, comparaison de proportions, analyse de la puissance statistique
- **Value (numeric)** : pour les comparaisons de moyennes ; variable jouant le rôle de mesure numérique
- **Outcome (0/1)** : pour les comparaisons de proportions ; colonne binaire de succès (bool, 0/1, yes/no...) qui alimente le numérateur de la proportion
- **Group (category)** : variable jouant le rôle de groupe catégoriel
- **Variable A** et **Variable B** : les deux variables à comparer, lorsque le cas d'usage oppose deux colonnes plutôt que deux groupes d'une même colonne

Réglages spécifiques à l'analyse de puissance :

- **Design** : type de données pour l'analyse : moyennes ou proportions
- **Solve for** : valeur cible de l'analyse : taille de l'échantillon ou puissance ciblée
- **Cohen's d** : valeur de la taille d'effet désirée (pour deux moyennes)
- **Proportion 1** : valeur de la première proportion
- **Proportion 2** : valeur de la seconde proportion
- **N per Group** : taille de chaque groupe pour déterminer la puissance
- **Target Power (1-β)** : seuil de puissance jugé suffisant. Il matérialise la zone verte de la jauge, et c'est la cible retenue lorsque l'on résout en taille d'échantillon.

Réglages communs:

- **Method** : paramétrique (basé sur des distributions statistiques connues) ou non paramétrique (robuste aux outliers)
- **Alpha** : niveau de significativité (erreur de première espèce) pour rejeter l'hypothèse nulle
- **Alternative** : unilatéral ( **Greater (A > B)** = à droite ; **Less (A < B)** = à gauche) ou bilatéral ( **Two-Sided** )

Le bouton **Run Testing** lance le test ; **Clear** réinitialise les résultats.

#### 4.1.2.2. Statistiques associées

Les résultats se présentent avec deux zones :

- À gauche, les métriques du test et quelques statistiques descriptives
- À droite, des jauges pour la puissance, significativité et taille d'effet ; ainsi qu'une représentation de l'intervalle de confiance

## 4.2. Data Exp (Exploration)

Localisation : onglet **Workspace** , panneau **Data Exp** .

### 4.2.1. Selection Mode (orbite et sélection)

Ce dossier décide de ce que fait la souris (ou le trackpad) dans la scène, et sert de point de départ à la constitution de groupes manuels.

- **Orbit** : mode par défaut. Le geste fait tourner et déplacer la scène.
- **Select** : le geste dessine un rectangle qui capture les points projetés qu'il recouvre.
- **Tag Selection** : désigne le groupe natif dont on veut capturer les points — les mêmes groupes que ceux du système de valences. Déplacer ce menu ne sélectionne rien par lui-même.

- **Select Tag Points** : capture d'un coup tous les points portant le tag choisi ci-dessus, en remplacement de la sélection courante. Le mode d'interaction n'est pas modifié : on reste en orbite ou en sélection. Le bouton peut être recliqué autant de fois que voulu, ce qui en fait le chemin de retour après avoir affiné une sélection à la main.

Trois moniteurs en lecture seule rendent compte de l'état courant :

- **Current Mode** : **Orbit** ou **Select** . Le compte de la sélection reste visible même après un retour en orbite.
- **Selection** : nombre de points actuellement sélectionnés.
- **Selection as Group** : indique **Active** dès que la sélection a été matérialisée en variable de groupe.

L'enregistrement d'une sélection en groupe se fait dans le bloc **Manual Groups** , juste en dessous du dossier.

### 4.2.2. Data Scaling (mise à l'échelle)

La mise à l'échelle se lit comme un pipeline en deux étages successifs : d'abord les colonnes, puis les lignes. Chaque étage est un menu déroulant, car ses options sont des réponses concurrentes à une même question — on n'en applique qu'une.

#### Étage 1 — **Column Scaling** (par variable)

- **None** : aucune transformation.
- **Standard** : centrage-réduction. Retrait de la moyenne, puis division par l'écart-type. C'est le choix classique dès que les variables n'ont pas la même unité.
- **Robust** : variante résistante aux valeurs extrêmes. Le centrage se fait sur la médiane et la dispersion sur l'écart interquartile, si bien qu'une poignée d'outliers ne comprime plus tout le reste du nuage.

#### Étage 2 — **Row Scaling** (par point)

- **None** : aucune transformation.
- **Unit Norm** : chaque point est ramené à une distance unitaire de l'origine. Seule sa *direction* subsiste ; la magnitude est effacée.
- **Global Quantile Norm** : même principe, mais la distance de référence est un quantile de l'ensemble des distances plutôt que la norme de chaque point. Les points conservent alors leurs écarts relatifs, tout en étant ramenés à une échelle commune.
  - **Quantile** : le quantile utilisé comme référence. N'apparaît que pour ce mode.

#### **After Projection**

Ce réglage n'apparaît que si un **Row Scaling** est actif. Il définit à quel moment la mise à l'échelle des lignes s'opère :

- Case décochée : dans l'espace initial N-dimensionnel, avant toute rotation, cisaillement ou projection.
- Case cochée : après la cascade de projections, sur l'espace 3D résultant.

Il vaut mieux mettre à l'échelle *avant* la projection pour conserver un maximum d'information durant les projections. Néanmoins, la normalisation a posteriori permet d'obtenir un étalement sur la sphère unité de tous les points projetés, ce qui simplifie grandement l'exploration visuelle — même si beaucoup d'informations ont été perdues au passage.

#### Data Zoom

Ce curseur espace ou contracte tous les points projetés. Il agit comme un multiplicateur de toutes les coordonnées de l'espace 3D final, après mise à l'échelle. Ce n'est donc pas une véritable normalisation, mais un zoom visuel : les coordonnées ND sous-jacentes ne changent pas, et il ne fait par conséquent ni entrer ni sortir de points d'une orbitale (voir 4.2.4).

### 4.2.3. Filtering (filtrage)

La section de filtrage permet d'afficher les points projetés vérifiant certaines conditions.

- **Filter** : active le filtrage. Décoché, tous les points restent visibles quels que soient les critères saisis.
- **Inter-Conditions** : nature de la combinaison entre les trois volets de critères ci-dessous — **All (AND)** (intersection) ou **Any (OR)** (union). AND signifie que tous les volets renseignés doivent être satisfaits simultanément pour afficher un point ; OR qu'un seul suffit.

Chaque volet dispose en outre de sa propre condition interne, qui combine ses critères entre eux :

**Label Condition**, **Group Condition** et **Axis Condition**, chacune réglable sur **AND** ou **OR**.

Les critères ne sont pas appliqués en continu : il faut cliquer sur **Apply** pour les valider, et **Clear** pour tout remettre à zéro. Le moniteur **Status** rend compte du filtre courant et du nombre de points retenus.

#### 4.2.3.1. Libellés de recherche

Le premier volet concerne les libellés attachés aux points :

- **Case Sensitive** : effectuer des recherches sans considérer la casse. "LABEL" est perçu comme "label" ou "Label".
- **Whole Word** : effectuer des recherches en considérant l'entièreté du label. S'il est coché, alors la chaîne "Lab" ne sera pas retenue pour rechercher la chaîne "Label".
- **Show** : saisie manuelle d'une chaîne de caractère incluse dans le libellé pour déterminer le points à afficher ; vous pouvez mettre plusieurs mots recherché différents en les séparant par une virgule
- **Hide** : saisie manuelle d'une chaîne de caractère incluse dans le libellé pour déterminer le points à masquer ; vous pouvez mettre plusieurs mots recherché différents en les séparant par une virgule

#### 4.2.3.2. Group Display (affichage par groupe)

Ce volet filtre sur les rôles de colonnes et les groupes calculés. Chaque menu liste les modalités disponibles ; laisser un menu sur **None** revient à ne pas filtrer dessus.

- **Target** : modalités de la colonne cible.
- **Treatment** : modalités de la colonne de traitement.
- **Segment** : modalités de la colonne de segmentation.
- **Cluster** : les clusters, si un clustering a eu lieu.
- **Personality** : les familles de points de personnalité — features micro et macro de sémiologie, HBDI, profils et features DISC, Social Styles, Management Styles. Voir 5.3.

#### 4.2.3.3. Filtrage des valeurs d'axes

Ce volet concerne les propriétés des points projetés. Il nécessite la sélection d'un à trois axes : **Axis 1**, **Axis 2**, **Axis 3**. Si vous ne souhaitez pas tous les renseigner, vous pouvez les laisser sur **None**.

- **Pipeline Stage** : l'étage du pipeline sur lequel les coordonnées sont lues — coordonnées ND d'origine, ND standardisées, ND normalisées ou 3D projetées. C'est un point important : filtrer sur les coordonnées 3D revient à filtrer sur ce que l'on voit, filtrer sur les coordonnées ND revient à filtrer sur la donnée réelle.

Vous pouvez ensuite choisir quelles valeurs filtrer : par signe, par valeur absolue, ou par métrique factorielle.

- **Axis 1 Sign**, **Axis 2 Sign**, **Axis 3 Sign** : filtrer les points dont les coordonnées sont, au choix, **Positive** ou **Negative** sur l'axe correspondant. **None** désactive le critère pour cet axe.
- **Use Coords** : autoriser le filtrage sur les coordonnées
- **Coords Ineq** : le sens de la comparaison —  $\geq$ ,  $>$ ,  $\leq$ ,  $<$  ou  $=$ .
- **Absolute Coords** : liée à la coche précédente ; valeur absolue que doivent vérifier les coordonnées d'un point sur l'ensemble des axes précédemment définis (si condition AND) ou sur au moins un de ces axes (condition OR)
- **Use Contrib** : autoriser le filtrage sur les contributions factorielles
- **Contrib Ineq** : le sens de la comparaison pour la contribution.
- **Contribution** : idem que pour les coords, mais s'applique à la valeur de la contribution factorielle du point sur les axes définis dans les listes précédentes ; ne s'applique que sur des points de type "Factor", c'est-à-dire projeté via une PCA ou une MCA, qui ont participé à la formation des axes factoriels.
- **Use Cos2** : autoriser le filtrage sur les cosinus carrés factoriels
- **Equality Tolerance** : marge de tolérance appliquée à la comparaison  $=$ . Sans elle, une égalité stricte sur des flottants ne retiendrait pratiquement jamais aucun point.
- **Cos<sup>2</sup> Ineq** : le sens de la comparaison pour le cosinus carré.

- **Cos<sup>2</sup>** : idem que pour les coords, mais s'applique à la valeur du cosinus carré factoriel du point sur les axes définis dans les listes précédentes ; ne s'applique que sur des points de type "Factor", c'est-à-dire projeté via une PCA ou une MCA, qu'ils aient participé ou non à la formation des axes factoriels.

#### 4.2.3.4. Filtrage combiné

##### Combinaison des deux règles précédentes

Vous pouvez définir des conditions sur les chaînes de caractères et sur les axes. Auquel cas, la condition AND/OR s'applique à l'ensemble.

Par exemple, avec une condition AND, un point doit respecter les chaînes de caractères saisie et respecter les valeurs sur l'ensemble des axes définis. Avec une condition OR, un point doit respecter les chaînes de caractères saisie ou respecter les valeurs sur l'un des axes définis.

##### Influence des plans de projections automatiques (canoniques et factoriels)

Lorsque les filtres sont actifs, si vous cliquez sur un plan de projection automatique, à savoir, les plans canoniques (ex : vertex-first, edge-first, etc. Cf. Views) ou les plans factoriels ( **Best Contrib Plan** et **Best Cos<sup>2</sup> Plan** ), alors les axes résultants sont envoyés dans les axes de filtrage.

Cela permet de balayer différents espaces canoniques et factoriels tout en mettant à jour les filtres basés sur les nouveaux axes affichés à l'écran.

#### 4.2.4. Orbital

L'orbitale est un système de filtrage multidimensionnelle basé sur la distance. Son nom n'est pas un terme mathématique officiel, mais une référence à la physique-chimie atomistique qui définit l'orbitale comme une zone de l'espace à l'intérieure de laquelle on peut observer un électron avec un certain niveau de probabilité. Ici, il s'agit de mettre en évidence la proportion de la population qui se trouve dans la zone délimitée par l'orbitale.

Dans l'application, par défaut, une orbitale va masquer les points proches de l'origine et ne laisser que les points en périphérie (les outliers). On peut inverser la logique de masquage ( **Outlier Mode** ) et n'affiche que les points centraux (inliers).

Remarque importante :

l'orbitale agit dans l'espace ND, avant la projection et le **Data Scaling** . Ainsi, le data scaling ne permet pas de sortir ou d'entrer dans l'orbitale. Un outlier n'a aucune raison de devenir inlier parce qu'on zoome sur l'échelle des données.

Il existe néanmoins un moyen d'ajuster la taille du solide au jeu de données : utiliser **Hypersolid Scale** . Ce paramètre intervient au moment de la création du solide et donc de l'orbitale qui en découle.

##### 4.2.4.1. Types d'orbitale

L'orbitale s'affiche avec **Show Orbital** , et son type se choisit dans le menu **Shape** . Trois orbitales sont proposées :

- **Gaussian Ellipsoid** : un ellipsoïde multidimensionnel basé sur les distributions normales. Deux façons de le dimensionner : **Confidence (%)** , un niveau de confiance gaussien fondé sur la distance de Mahalanobis, ou **k ( $\sigma$ )** , un nombre d'écarts-types. La coche **Stat Fit** bascule entre les deux lectures.
- **Box** : un n-rectangle dont les longueurs sont définies par des centiles dans chacune des dimensions ; les curseurs **Quantile X** , **Quantile Y** , **Quantile Z** , **Quantile W** et **Quantile V** permettent d'ajuster ces longueurs.
- **Quantile Ellipsoid** : l'ellipsoïde inscrit dans le n-rectangle précédent ; l'option **Outer Ellipsoid** permet d'utiliser l'ellipsoïde circonscrit au n-rectangle.

Le moniteur **Inside | Outside** affiche en permanence le nombre de points situés à l'intérieur et à l'extérieur de l'orbitale — c'est la lecture directe de la proportion masquée.

En termes d'usage, l'ellipsoïde gaussien est simple à manipuler (un seul curseur **Confidence (%)** ), mais n'est approprié que pour les nuages de points à tendance gaullienne (distribution normale dans chacune des dimensions).

La "box" est une variante robuste de l'ellipsoïde précédent. Mais elle a tendance à devenir très stricte (beaucoup de points masqués) pour les hautes valeurs de quantiles. L'ellipsoïde quantile est une variante beaucoup plus douce de la boîte ; elle ne masque que les points distants dans plusieurs directions simultanées. Mais sa variante circonscrite (outer orbital), elle, est très restrictive (encore plus que la boîte). Il faut abaisser les seuils pour éviter de masquer tout le nuage.

D'un point de vue proportion de masquage, on a approximativement les valeurs suivantes :

- Dans un espace à n dimensions, le n-rectangle de quantile 90 dans toutes les dimensions englobe environ  $(0,9^n)$  % de la population (72% en 3D, 65% en 4D, 60% en 5D..., 35% en 10D).
- L'ellipsoïde inscrit dans le n-rectangle précédent ne couvre qu'une fraction rapidement décroissante : en 3D, l'orbitale ellipsoïde occupe 52% du rectangle, en 4D 30%... et en 10D tout juste 0,25% !
- L'ellipsoïde circonscrit au n-rectangle inverse la tendance. En 3D, l'orbitale occupe près de 2,7 fois plus d'espace, en 4D 1,2 fois plus, en 5D 8,3 fois plus... et ça croît de plus en plus vite !

Du plus tolérant au plus strict : ellipsoïde inscrit < n-rectangle < ellipsoïde circonscrit.

#### 4.2.4.2. Comparaison entre orbitale et filtrage

Quelle différence entre l'orbitale et le filtrage basé sur les coordonnées absolues ?

Les deux objets servent à filtrer la visualisation, mais possèdent chacun leurs nuances.

*Nombre de dimensions filtrées*

Le filtrage basé sur **Absolute Coords** ne s'applique que sur (au plus) trois axes définis dans les axes de filtrage. L'orbitale, quant à elle, s'applique sur toutes les dimensions possibles en simultané.

*Condition booléenne*

Le filtrage par coordonnées peut utiliser la condition AND ou OR, tandis que l'orbitale n'utilise qu'un principe AND par construction.

*Définition des seuils de filtrage*

Le filtrage par coordonnées nécessite d'explicitement la valeur seuil de la coordonnée qui sera vérifiée sur les axes de filtrage. L'orbitale fonctionne plutôt par une logique de quantile (proportion de points qui possèdent une coordonnée inférieure ou égale au seuil) ou probabiliste (intervalle de confiance gaussien basé sur la distance de Mahalanobis).

#### 4.2.5. Decision Boundary (frontière de décision)

Cette option permet d'afficher la frontière de décision de la régression logistique binaire ou l'hyperplan prédictif de la régression linéaire.

- **Show Boundary** : affiche l'hyperplan ou la frontière.
- **Boundary Alpha** : ajuste la transparence de l'hyperplan / frontière de décision.
- **Resolution** : ajuste la finesse du maillage de l'hyperplan / frontière de décision.
- **Iso Levels** : en classification, superpose trois coques de probabilité emboîtées ( $p = 0,25 / 0,5 / 0,75$ ) au lieu de la seule frontière à 0,5. On lit ainsi la pente de la décision, et pas seulement son emplacement.
- **PD Feature A** / **PD Feature B** : les deux variables portées par les axes de la surface de dépendance partielle. Toutes les autres variables sont maintenues à leur médiane d'apprentissage.
- **Residual Sticks** : trace, pour chaque point, un segment vertical entre sa valeur observée et la valeur prédite par le modèle. La longueur du segment est le résidu.
- **Residual Alpha** : transparence de ces segments.

Note : la frontière de décision n'est affichée que si la variable-cible est binaire ; l'hyperplan de prédiction n'est affiché que si le nombre de prédicteurs (variables explicatives) est supérieur ou égal à 2. Le cas dégénéré de la droite (régression simple) est ignoré.

#### 4.2.6. Hy-Perf Cube

Permet de modifier rapidement les axes affichés sur la scène pour analyser les jeux de données de Hy-Performance.

**Cube Preset** : ces presets permettent de mapper rapidement certains axes spatiaux aux variables du jeu de données transformés (budget, uplift, ROI, etc.).

Pour personnaliser davantage, il faut mapper les axes manuellement dans Dataset / Axes Mapping.

### 4.2.7. Valence

La logique de Valence (les liens de connectivité) repose sur une analyse de proximité dans l'espace : chaque point est relié à ses plus proches voisins, et la couleur du trait dit à quel rang de voisinage il correspond.

- **Show Lines** : affiche les lignes de valence.
- **Selection Only** : ne trace les lignes que depuis les points ou sommets sélectionnés, qui jouent alors seuls le rôle d'ancres. C'est le réglage à utiliser dès que le nuage est dense.

#### Classification des points

Le système regroupe tous les points affichés en trois grandes familles pour filtrer les types de liens :

- Reference : points issus des presets (Zodiaque, Malgovert, DISC, MBTI, etc.).
- Data : points importés via CSV ou ajoutés manuellement.
- Factor : points calculés par l'Analyse Factorielle (Individus ou Modalités).

Logique globale : les points Data sont les éléments mobiles et analysés, tandis que les points Reference, Factor et le Polytope servent de repères fixes (des ancres). On regarde donc comment la donnée utilisateur s'oriente par rapport à la structure interne de l'application

#### Valence Info

Si **Valence Info** est coché, alors des informations supplémentaires sont ajoutées dans le panneau latéral quand on clique sur un point : on peut consulter les différents voisins d'un point ainsi que leur distance respective, cela pour les différentes hypothèses de projection (ND, 3D, avec ou sans normalisation).

La coche **Use Inverse Dist** permet d'afficher, non pas les distances euclidiennes, mais les coefficients de pondérations par l'inverse des distances. Exemple : avec 3 voisins, la valeur affichée pour un voisin donné est l'inverse de la distance divisée par la somme (sur les 3 voisins) des inverses des distances.

#### Rang du voisinage

Le paramètre **K Neighbors** définit le nombre de rangs (de distance les plus proches) à afficher par point ; ce n'est pas le nombre exact de liaisons affichées (voir "Égalités" ci-dessous).

#### 4.2.7.1. Pipeline Stage

La notion de voisinage dépend de l'étape de projection. Rappel, la projection intègre toutes ces étapes :

- ND-4D : ND-standardisation, ND-normalisation (option), rotation, cisaillement, projection
- 4D-3D : rotation, cisaillement, projection, 3D-Normalisation (option), mise à l'échelle (option)
- 3D-2D/1D : écrasement

On peut effectuer une étude de voisinage dans la scène ND initiale avant la projection spatiale. Celle-ci peut inclure ou non une étape de normalisation. De même, le voisinage peut être effectué après la projection, dans la scène 3D résultante et visible à l'écran. Cette scène peut elle aussi inclure une normalisation.

En synthèse, les voisinages peuvent être considérés :

Le menu **Pipeline Stage** désigne l'étage retenu :

- **ND** : la valence est calculée sur les données initiales, espace ND sans normalisation
- **ND Stand** : la valence est calculée sur les données ND centrées réduites ("standardisées"), avant projection
- **ND Norm** : la valence est calculée sur les données ND normalisées, avant projection
- **3D** : la valence est calculée sur les données 3D projetées, avant normalisation 3D et avant mise à l'échelle
- **3D Scaled** : la valence est calculée sur les données projetées, après les éventuelles normalisations 3D et mises à l'échelle

#### 4.2.7.2. Directions des valences

Une connexion s'établit entre des points d'ancrage et des points de voisinage. Pour chaque point d'ancrage, on cherche les points de voisinages. Les voisins les plus proches du point d'ancrage sont dits de "rang 1". Les voisins suivants sont de rang 2, etc.

Pour chaque point P, tous les autres points sont triés par distance euclidienne. Les points à la même distance (avec une tolérance de  $10^{-5}$ ) sont regroupés en un seul rang. Avec  $K = 2$ , les deux rangs de distance les plus proches sont conservés - cela peut générer plus de 2 lignes si plusieurs points partagent la même distance.

Il existe deux algorithmes différents selon la nature de la connexion :

*Liens Mutuels (Internes à un groupe) : les ancrs et les voisins appartiennent au même groupe de points*

Exemples : Reference->Reference, Data->Data, Polytope->Polytope

- Calcul : pour deux points I et J, on calcule la position de J dans le

classement des voisins de I ( $\text{rank}_{\{IJ\}}$ ) et inversement ( $\text{rank}_{\{JI\}}$ ).

- Condition : le lien est tracé si le rang moyen  $(\text{rank}_{\{IJ\}} + \text{rank}_{\{JI\}}) / 2$  est

inférieur au paramètre K Neighbors.

- Logique : permet de ne montrer que les relations fortes et réciproques au sein d'une même population.

*Liens Unidirectionnels : ancrs et voisins sont des groupes différents*

Exemples pour : Data->Reference, Data->Factor, Zodiaque->Polytope.

- Calcul : On prend chaque point "Ancre" (ex: un point Data) et on cherche ses

plus proches voisins dans le groupe "Voisin" (ex: les sommets du Polytope).

- Condition : Le lien est tracé si le rang du voisin est inférieur à K Neighbors.
- Logique : C'est une logique d'attraction. On regarde vers quels points de référence ou vers quelles facettes du solide les données "tendent".

Dans le bloc de règles de valence, situé sous le dossier, vous pouvez ajouter toutes les liaisons désirées entre les groupes jouant le rôle d'ancrage et les groupes jouant le rôle de voisinage.

### Gestion des Égalités (Ties)

Lorsqu'un rang contient plus d'un voisin (une égalité), les lignes correspondantes peuvent être tracées en pointillés (cliquer sur **Ties Dash**) au lieu d'être pleines. Cela signale visuellement les voisinages ambigus. Une ligne est en pointillés si l'une ou l'autre de ses extrémités présente une égalité à son niveau de rang respectif.

- Tolérance : Si deux points sont à une distance quasi identique ( $10^{-5}$ ), ils

reçoivent le même rang.

- Visuel : Si un rang contient plusieurs points (ex-æquo), les lignes sont affichées

en pointillés (si l'option Dashed Ties est activée dans le panel Style).

### Code Couleur et Distances

- Dimensions : Les distances sont calculées sur les coordonnées finales (x, y, z)

après toutes les transformations (ND-4D-3D). Cela signifie que le zoom et les rotations 4D impactent directement les liens de valence.

- Couleurs : chaque niveau de proximité (rang du voisin) peut être identifié par une

couleur personnalisable

## 4.2.8. Style des valences et des orbitales

**Style des valences** (dossier **Valence Metrics / Style**)

- **Line Width** : épaisseur des lignes de valence.
- **Ties Dash** : trace en pointillés les lignes correspondant à un rang comportant plusieurs voisins ex æquo (voir « Gestion des égalités » ci-dessus).
- **Rank 1**, **Rank 2**, **Rank 3**, **Rank 4+** : les quatre couleurs associées aux rangs de voisinage.

**Style des orbitales** (dossier **Orbital / Style**)

- **Orbital Width** : épaisseur du tracé de l'orbitale.
- **Center Origin** : centre l'orbitale sur l'origine du repère plutôt que sur le barycentre du nuage.
- **Size Constraints** : active les trois bornes ci-dessous. Elles sont purement esthétiques et n'ont aucun effet mathématique : elles empêchent seulement une orbitale de devenir illisible parce que trop petite ou trop grande à l'écran.
  - **Absolute Size Min** : taille minimale absolue.
  - **Size Ratio Min** et **Size Ratio Max** : bornes de la taille rapportée à celle du solide.
- **3D Color**, **4D Color**, **ND Color** : couleurs de l'orbitale selon l'espace dans lequel elle est calculée.
- **Alpha** : transparence. Pour masquer une orbitale sans la désactiver, il suffit de régler son alpha à zéro.

#### 4.2.9. Trajectory (trajectoires)

Relie chronologiquement les points d'un même groupe, pour lire un déplacement plutôt qu'un nuage figé.

- **Show Trajectories** : trace une polyligne chronologique par groupe, à travers les points projetés.
- **Time Column** : la colonne qui donne l'ordre chronologique.
- **Group Column** : la colonne qui définit les groupes ; chaque modalité donne une trajectoire.
- **Trajectory Opacity** : transparence des polylignes.
- **Time Gradient** : colore chaque trajet du plus ancien au plus récent (sombre → clair), afin que le sens de parcours soit lisible.

#### 4.2.10. Dimension Fading (estompage dimensionnel)

Sur un nuage projeté depuis un espace de grande dimension, beaucoup de points se retrouvent superposés à l'écran alors qu'ils sont éloignés dans les dimensions non représentées. L'estompage rend cette masse cachée visible en jouant sur l'opacité.

- **Enable Dim Fading** : active l'estompage.
- **Opaque threshold** : seuil en dessous duquel un point reste pleinement opaque.
- **Min alpha** : opacité plancher des points les plus éloignés dans les dimensions masquées.
- **Show hidden-mass** : affiche l'indicateur de masse cachée, c'est-à-dire la part du nuage que la projection courante ne restitue pas.

#### 4.2.11. Marqueurs de scène

Le dossier **Scene Markers** (axes et plans de repère) est décrit en 2.2.4.4, avec les autres éléments de décor de la scène.

## 4.2.12. Surcouches des modules d'analyse

Trois dossiers n'apparaissent qu'une fois le module correspondant exécuté. Ils projettent son résultat dans la scène 3D.

**Association Rule Edge** — voir 5.2.4

- **Show Rule Edges** : trace une arête entre les éléments liés par une règle d'association.
- **Edge Min Lift** : masque les arêtes dont le lift est inférieur à ce seuil. Ce filtre est purement visuel et reste indépendant du **Min Lift** utilisé lors de l'extraction des règles.

**MMM Surface & Map** — voir 5.10

- **Fit Landscape** : affiche la grille de recherche des hyperparamètres ( $\theta \times ec$ ) d'un canal, colorée par score d'ajustement, à la place de la série temporelle. Nécessite d'avoir lancé un MMM local au préalable.
- **Landscape Channel** : le canal dont on affiche le paysage d'ajustement.
- **Collinearity Map** : affiche les canaux comme des points — la proximité traduit la co-évolution des dépenses transformées, la couleur le VIF. Un VIF élevé signale une collinéarité, donc une attribution peu fiable.

**Optimization 3D View** — voir 5.7

- **3D View** : ce que la scène représente — **None**, **Selection Plane** (le plan de sélection) ou **Feasibility Polytope** (le polytope des solutions admissibles).

## 4.3. Data Style

Localisation : onglet **Workspace**, section **Exp**, dossiers **Class Style** et **Layouts**.

Le style des points et des étiquettes eux-mêmes (taille, forme, familles affichées, encodage par variable) est décrit en 3.2. Cette section traite de ce qui vient par-dessus : la coloration par classe, la forme par classe, et la façon dont les axes et les nombres sont nommés et affichés.

### 4.3.1. Class Style (coloration par classe)

Cinq rôles peuvent servir à colorer les points. Ils ne s'excluent pas : ils sont empilés par ordre de priorité, et le premier qui s'applique à un point l'emporte.

| Priorité | Réglage              | Source de la couleur                        |
|----------|----------------------|---------------------------------------------|
| 1        | Colorize User Groups | les groupes tracés à la main au lasso       |
| 2        | Colorize Cluster     | les modalités de cluster                    |
| 3        | Colorize Target      | les modalités de la colonne cible           |
| 4        | Colorize Treatment   | les modalités de la colonne de traitement   |
| 5        | Colorize Segment     | les modalités de la colonne de segmentation |

Un point non couvert par un niveau retombe sur le suivant : un point sans groupe manuel prend la couleur de son cluster, et ainsi de suite.

Deux réglages réutilisent des palettes déjà définies ailleurs, plutôt que d'en créer une sixième :

- **Use Valence Colors** : affecte aux emplacements de couleur 1 à 4 la palette des valences (rangs 1 à 4+).
- **Use Orbital Colors** : affecte aux emplacements 5 à 7 la palette des orbitales (3D / 4D / ND) — ou aux emplacements 1 à 3 si les couleurs de valence sont désactivées.

### Forme par classe

- **Shape By** : encode un rôle de classe par la *forme* du point, ce qui est orthogonal à la couleur — on peut donc lire deux variables qualitatives à la fois. Au choix : **None** , **Manual Group** , **Cluster** , **Target** , **Treatment** ou **Segment** . La première modalité conserve la forme définie par **Point Shape** ; les suivantes reçoivent des glyphes distincts.
- **Category Shapes** : moniteur rappelant la correspondance modalité → forme effectivement retenue.

### 4.3.2. Couleurs de Valence

Pour la valence, chaque ligne relie deux points, A et B, et sa couleur traduit le rang du voisinage. Le point A peut classer B comme son voisin le plus proche (Rang 1), tandis que B peut classer A différemment (ex : Rang 2). Pour refléter cette relation bidirectionnelle, la couleur affichée est déterminée par la partie entière de la moyenne des deux rangs : Rang affiché =  $\lfloor (\text{rank}_{A \rightarrow B} + \text{rank}_{B \rightarrow A}) / 2 \rfloor$

Ceci garantit que la couleur reflète un accord mutuel : deux points qui se considèrent tous deux comme proches obtiennent une couleur forte (Rang 1), tandis que les relations asymétriques sont naturellement dégradées.

### 4.3.3. Layouts

Un *layout* renomme les axes spatiaux et adapte leurs couleurs, pour que la scène parle le vocabulaire du référentiel étudié plutôt que celui des mathématiques. Le calcul n'est en rien modifié : seul l'habillage change.

- **Layout** : le layout actif.
  - **Default** : affichage standard, axes X, Y, Z, W, V5...
  - **Imported Data** : les axes reprennent les noms des colonnes importées qui leur sont assignées.
  - **Malgover** : affichage alternatif utilisé par l'auteur dans le cadre du roman Malgover. Les coordonnées X, Y et Z sont renommées K, L et S. Elles désignent les sérums Kappa (Cocyte), Lambda (Léthé) et Styx (Sigma) dans l'univers fictif de Malgover. Les couleurs des axes correspondants sont modifiées pour refléter la couleur des sérums dans le roman : bleu pour le Cocyte (K), argent pour le Léthé (L) et vert pour le Styx (S).
  - **MBTI 4D**, **DISC 3D**, **SS 3D**, **DISC-SS 4D**, **Task Management** : affichages alternatifs qui reprennent les libellés des frameworks comportementaux utilisés en entreprise.
- **Label Chars** : nombre de caractères conservés pour les noms d'axes d'un layout. Par exemple, dans le layout **SS 3D**, les axes X et Z deviennent « Assertiveness » et « Responsiveness » ; en réglant le curseur sur 2, les libellés deviennent « As » et « Re ».
- **Number Decimals** : nombre de décimales affichées pour les nombres, dans toutes les tables et tous les panneaux de l'application.
- **Local Settings Reset** : remet à zéro les réglages d'affichage conservés localement par le navigateur, et rétablit les valeurs par défaut.

## 5. Analyse

---

### 5.1. Lancer un module

Aucun module d'analyse ne se recalcule tout seul : les réglages se posent d'abord, puis un bouton lance le calcul. Chaque module suit la même convention.

- **Run <Module>** : lance le calcul **dans le navigateur**. C'est la voie par défaut : le résultat est immédiat et la scène 3D se met à jour dans la foulée.
- **Run Large <Module>** : délègue le même calcul au **moteur Python distant**. Cette variante n'apparaît que pour les modules concernés et sur les forfaits qui y donnent droit. Elle sert dès que le volume de données rend le calcul local trop lent, ou que la méthode demandée dépasse ce que le navigateur sait faire tenir en mémoire.
- **Clear** : efface les résultats du module et retire ses points de la scène.

Les boutons portent le nom du module : **Run Correlation**, **Run Profiling**, **Run Association** et **Run Large Association**, **Run Factor**, **Run Clustering**, **Run Estimator** et **Run Large Estimator**, **Run Forecast** et **Run Large Forecast**, **Run Performance**, **Run MMM** et **Run Large MMM**, **Map Personality**, **Run Testing**, **Run** pour l'optimisation.

Un module lancé sur des données qui ont changé depuis reste affiché tel quel : c'est à vous de le relancer. Le rappel vaut particulièrement après un pré-traitement (partie 3.4) ou un ré-échantillonnage.

### 5.2. Advanced Stats

Ce volet complète l'onglet Stats. Ici, les analyses portent sur plusieurs paires de colonnes en parallèle. L'onglet se subdivise en trois parties : les corrélations (bivariées), le profiling (sur ou sous-représentation de certains attributs au sein de clusters) et les règles d'association.

#### 5.2.1. Seuils pour les tests statistiques

Il est possible de définir les tests statistiques souhaitées ainsi que les valeurs seuil permettant de coloriser les métriques. Ces seuils s'appliquent aux graphiques de corrélation et de profiling.

- **Num Testing** : test statistique souhaité pour les corrélation numériques (Pearson, Spearman, Kendall)
- **Cat Testing** : test(s) statistique(s) souhaité(s) pour les dépendances catégorielles (chi2 seulement, ou un mélange chi2 et test exact de Fisher pour les tableaux de contingence 2x2)

- **Mixed Testing** : test statistique utilisé pour l'analyse des écarts de valeurs centrales entre deux groupes (ANOVA F qui généralise le test T de Student, et Kruskal-Wallis qui généralise le test U de Mann-Whitney)
- **Alpha** : valeur seuil en deçà de laquelle les p-values deviennent vertes dans les tableaux.
- **Num Effect Threshold** : valeur seuil au-delà de laquelle les tailles d'effet numériques (en valeur absolue) deviennent vertes ; s'applique au R<sup>2</sup>, rho carré de Spearman, et tau de Kendall
- **Cat Effect Threshold** : valeur seuil au-delà de laquelle les tailles d'effet catégorielles (en valeur absolue) deviennent vertes ; s'applique au V de Cramer, h de Cohen et r de Wilcoxon
- **OR Effect Threshold** : valeur seuil au-delà de laquelle les métriques de type ratio et log-ratios (en valeur absolue) deviennent vertes ; s'applique aux (log)-odds ratios, (log-)lifts et Weights of Evidence
- **Mix Effect Threshold** : valeur seuil au-delà de laquelle les tailles d'effet catégorielles (en valeur absolue) deviennent vertes ; s'applique au d de Cohen, g de Hedges, Delta de Glass, delta de Cliff, Êta carré (test H) et Oméga carré (ANOVA)

## 5.2.2. Corrélations

### 5.2.2.1. Paramètres

Données à analyser (optionnel et cumulatif) :

- **Target** : variable-cible utilisée pour déterminer les variables les plus discriminantes
- **Imported Data** : données importées en CSV
- **Factor Data** : données issues d'une projection factorielle
- **Cluster Data** : label de s-cluster issus d'un clustering
- **Score Data** : scores calculés par un estimateur de Sup ML

### 5.2.2.2. Statistiques associées

#### Tableau des métriques de corrélation avec la cible

Tests statistiques de dépendance bivariable (Pearson et équivalents, khi-deux, ANOVA, etc.) avec la p-value et les tailles d'effet associée (R<sup>2</sup>, oméga carré, V de Cramer, etc.)

#### Tableau des métriques de corrélation entre variables explicatives

Idem que précédemment, mais entre les variables explicatives

#### Matrices des tailles d'effet (corrélation intra-variables)

Matrices de corrélation (taille d'effet) entre variables.

#### Barres des tailles d'effet vs Target

Intensité de la corrélation entre les variables explicatives et la variable-cible. Traduit le pouvoir discriminant.

## Matrices de corrélation de la scène

Matrices de corrélation entre données spatiales ou les données projetées par une méthode factorielle.

Pour ces matrices de corrélation, tous métriques sont proposées :

- **Dot Product** (produit scalaire normalisé) : somme des produits par composante, divisé par le produit des normes ;  $\text{dot}(u, v) = (u \cdot v) / (\|u\| \times \|v\|)$  (similarité cosinus) ; plage :  $[-1, 1]$ . 1 = aligné, 0 = orthogonal, -1 = opposé
- **Pearson** : coefficient de corrélation linéaire de Pearson, de -1 à 1 ; le coefficient est proche de zéro en cas de non-corrélation ;  $r(u, v) = \Sigma(u_i - \bar{u})(v_i - \bar{v}) / (\sigma_u \times \sigma_v)$  où  $\bar{u}$ ,  $\bar{v}$  sont les moyennes et  $\sigma$  sont les écarts-types des composantes des coordonnées.
- **Similarity** (indice de similarité) : rapport de distances entre une paire et la distance maximale du jeu de données ;  $\text{sim}(u, v) = 1 - d(u, v) / \text{maxDist}$  où  $d$  est la distance euclidienne et  $\text{maxDist}$  est la plus grande distance dans l'ensemble des points actifs ; vaut 0 quand deux points sont superposés, vaut 1 quand les deux points sont les plus distants parmi toutes les paires

Note : les auto-comparaisons diagonales de la matrice sont laissées vides.

### 5.2.3. Profiling (profilage)

Cette section est une description approfondie des valeurs caractéristiques que l'on retrouve dans des groupes d'intérêt (segment, target, cluster, score quantile), avec ou sans treatment.

#### 5.2.3.1. Paramètres

Données à analyser (optionnel et cumulatif) :

- **Imported Data** : données importées en CSV
- **Factor Data** : données issues d'une projection factorielle
- **Cluster Data** : label de s-cluster issus d'un clustering
- **Score Data** : scores calculés par un estimateur de Sup ML

Les paramètres spécifiques au profilage sont :

- **Group Variable** : variable utilisée pour définir les groupes à profiler ; correspond à l'une des variables d'intérêt (segment, target, cluster, score quantile)
- **Use Complement** : pour calculer les dénominateurs des ratios, soit on utilise la population totale, soit on utilise la population complémentaire au groupe étudié (= le "reste" de la population = les autres groupes). Si on coche cette option, c'est le complémentaire qui est utilisé ; permet de mettre en valeur les métriques des petits groupes.
- **Moment Metric** : variable mise en avant pour l'affichage des métriques de profilage numérique
- **Lift Metric** : variable mise en avant pour l'affichage des métriques de profilage catégoriel

### 5.2.3.2. Statistiques associées

#### Tableau de profilage

Cinq onglets centralisent toutes les métriques :

- **Profile** : synthèse générale ; combine une partie des vues lift et moment
- **Moment** : focus sur les variables descriptives numériques
- **Lift** : focus sur les variables descriptives catégorielles
- **Lift Net** : focus sur les variables descriptives catégorielles (1 ligne par catégorie) avec présence d'un treatment
- **Info Value** : focus sur les variables descriptives catégorielles (1 ligne par variable) avec ou sans présence de treatment ; elle complète la vue Lift / Lift Net avec des indicateurs agrégés

#### Barres de profilage

Une version graphique des résultats précédents.

Les graphiques sont regroupées par métrique de profilage. Une ligne par variable avec possibilité de trier par groupe.

- Les barres positives (une couleur par groupe) traduisent des catégories ou des valeurs hautes caractéristiques (typiques, surreprésentées) des groupes concernées.
- Les barres négatives (une couleur par groupe) traduisent des catégories ou des valeurs basses caractéristiques (sous-représentées) des groupes concernées.

### 5.2.4. Association

Les règles d'association déterminent des ensembles de catégories (variables qualitatives) dites "antécédentes" qui apparaissent en même temps qu'une catégorie d'intérêt dite "conséquente". Les règles ne fonctionnent qu'entre variables catégorielles.

#### 5.2.4.1. Paramètres

Les paramètres spécifiques aux règles d'association sont :

- **Group Variable** : variable facultative utilisée pour partitionner les règles : soit les règles sont globales, soit elles sont subdivisées selon toutes les modalités de la variable de groupe
- **Group as Consequent** : plutôt que de partitionner les règles, la variable groupe est cette fois-ci utilisée pour définir directement les valeurs de Conséquent.
- **Min Support** : support minimal pour retenir une règle ; c'est la proportion de lignes qui contiennent l'ensemble des catégories antécédentes ; recommandé 1-5%
- **Min Confidence** : confiance minimale pour retenir une règle ; c'est la proportion de lignes (parmi les lignes ayant tous les antécédents), pour lesquelles le conséquent apparaît ; permet de filtrer les règles peu fiables, même si elle n'est pas suffisante (le lift ci-après est nécessaire pour écarter les règles fiables mais sans intérêt) ; recommandé 60-80%

- **Min Lift** : surreprésentation minimale pour retenir une règle ; une règle ayant un lift de 1 est une sélection aléatoire ; une règle à fort lift indique que le conséquent a effectivement plus de chances de sortir lorsque les antécédents apparaissent (règle utile) ; un lift négatif permet justement de trouver les antécédents qui écartent l'apparition du conséquent (objectif opposé) ; recommandé 1,2
- **Max Antecedent Size** : nombre d'antécédents utilisés pour définir une règle ; plus ce nombre est grand, plus la combinatoire explose
- **Max Consequent Size** : nombre de conséquents utilisés pour définir une règle ; plus ce nombre est grand, plus la combinatoire explose
- **Top N** : filtrage visuel pour ne garder que les N premières lignes (mais on peut ordonner les lignes ou les filtrer, indépendamment de cette valeur)

Le bouton **Run Association** lance l'extraction dans le navigateur, **Run Large Association** la délègue au moteur Python (voir 5.1), et **Clear** réinitialise les résultats.

#### 5.2.4.2. Statistiques associées

Le tableau de résultat liste les règles ainsi que les métriques associées (support, confiance, lift, etc.). Le symbole  $\wedge$  signifie intersection (ET logique).

### 5.3. Hy-Personality

Voici la méthode "rigoureuse" pour déterminer les signes de Zodiaque qui vous ressemblent le plus !

Un questionnaire de personnalité intégré à l'application peut être saisi en ligne. Ils combinent des thématiques présentes dans plusieurs frameworks utilisés en entreprise (Social Style, DISC, MBTI) et aussi les descriptions lexicales du zodiaque astrologique.

Attention, ceci n'est pas un protocole d'analyse psychologique/psychométrique. Il s'agit d'un exemple illustratif montrant l'intérêt des projections factorielles appliquées à des questionnaires de toute sorte. L'objectif ici est montrer visuellement les ressemblances lexicales entre différents frameworks célèbres.

Remplir le questionnaire et projeter les résultats permet ainsi de visualiser les concepts ou attributs similaires entre vos réponses et les profils présents dans les différents frameworks. Cette approche repose sur une PCA ou une MCA des réponses des différents profils aux 18 questions du formulaire. C'est un jeu de données fourni par l'application. Il est libre, mais ne peut pas être modifié dans cette interface. En revanche, les réponses que vous ajoutez, elles, sont personnalisables.

Procédure : répondre aux 18 questions en bougeant le slider à gauche ou à droite (échelle à 5 degrés) en fonction de la réponse qui vous correspond le plus. Si vous hésitez, vous pouvez laisser le slider au centre (réponse équilibrée ou neutre).

Une fois terminé, saisissez un libellé et cliquez sur **Submit** pour projeter les résultats dans l'hyperespace factoriel !

### 5.3.1. Sources de données

Le questionnaire peut provenir de différentes sources :

- **Reference Points** : données déjà intégrées dans l'application (frameworks comportementaux et zodiaque)
- **User Form** : questionnaire en ligne affiché à l'écran
- **CSV** : questionnaire importé au format CSV (via la section **Form Import** de l'onglet actuel, ou la section d'import de l'onglet **Data** )

La section **Form Import** propose **Upload CSV** pour charger un fichier de réponses, et **Clear All Respondents** pour effacer d'un coup tous les répondants importés.

### 5.3.2. Techniques de projection

Les réponses au questionnaire (toutes sources confondues) sont projetées en utilisant une PCA ou une MCA déjà entraînée avec les données de référence de l'application.

- **Method** : **PCA** , **MCA** ou **MFA** (analyse factorielle multiple, quand les questions sont regroupées en blocs thématiques)
- **Keep All Factors** : conserve tous les facteurs (cela inclut également le bruit statistique dans l'explication des variances)
- **Nb Factors** : si l'option précédente est décochée, on peut définir ici le nombre de facteurs à conserver ; recommandé entre 3 et 8

Le bouton **Map Personality** lance la projection ; **Clear** réinitialise les résultats.

### 5.3.3. Filtrage des points de référence

La section **Reference Filtering** permet de choisir quels points de référence sont affichés. Chaque catégorie est subdivisée entre *Features*, qui correspondent aux critères de description, et *Profiles*, qui en sont la résultante — une combinaison de plusieurs critères.

Les cases disponibles sont :

- **Semio Micro Features** et **Semio Macro Features** : les circuits de sémiologie, au niveau fin puis agrégé
- **HBDI Features** : les quadrants cérébraux de Herrmann
- **MBTI Profiles** et **MBTI Features**
- **DISC Profiles** et **DISC Features**
- **SS Profiles** et **SS Features** : Social Styles
- **MS Profiles** et **MS Features** : Management Styles
- **MG Profiles** : Managerial Grid (grille de Blake et Mouton)

- **Zodiac Profiles** et **Zodiac Features**

Exemple : les points *Extraversion* et *Introversion*, qui définissent l'orientation de l'énergie, sont des *MBTI Features*, tandis qu'*INTP* est un *MBTI Profile* défini par la présence de la modalité « I » (pour Introvert). De même, le *Feu* est une *Zodiac Feature*, tandis qu'*Aries* (Bélier) est un *Zodiac Profile*.

### 5.3.4. Statistiques associées

#### Radar des réponses par profiles

Chaque sommet du radar représente une question (valeur entières de -2 à 2). Deux profils sont superposés dans ce radar pour visualiser la ressemblance globale de leurs réponses au questionnaire.

#### Réponses de tous les répondants par question

Le graphique en barres donne la réponse de tous les répondants à la question sélectionnée.

#### Matrices de corrélation de la scène

Matrices de corrélation entre réponses brutes au questionnaire de personnalité (valeurs entières de -2 à 2), et aussi entre réponses projetées (coordonnées factorielles).

### 5.3.5. Exemple d'utilisation : zodiaque inversé

Aller dans *Modeling / Personality*

Dans le panneau *Personality Map*, laisser les réglages par défaut :

- *Méthode* : PCA
- *Reference Points* : coché
- *User Form* : coché
- *CSV* : décoché

Remplir le questionnaire, donner un nom et cliquer sur *Soumettre*. On bascule alors sur *Workspace*.

Aller dans la section *Dataset* du panneau *Workspace*, dossier *Points / Display*, et cocher uniquement :

- *Show Reference Pts* : ce sont les points déjà dans l'application
- *Show Fact Indiv Pts* : ce sont les réponses au formulaire en ligne

Aller dans la section *Exp* du panneau *Workspace*.

- Dans *Data Scaling*, mettre *Row Scaling* sur *Unit Norm* et le *Data Zoom* aux alentours de 2 ou 3 selon le feeling.
- Dans *Valence*, cocher *Show Lines*, puis *Valence Info* et *Use Inverse Dist*, et mettre le *Pipeline Stage* sur *ND*.
- Toujours dans *Valence*, en dessous du slider *K Neighbors*, cocher uniquement *Factor-Reference*

Juste en dessous, dans la section *Filtering*, cocher *Filter* puis dans la liste *Personality*, choisir *Zodiac Profiles*.

Désormais, en cliquant sur le point correspondant à votre réponse, un panneau latéral apparaît avec des statistiques. Vous pouvez consulter le graphiques *ND Valence* et *ND-Norm Valence* qui indique votre "proportion" de chaque signe du Zodiaque (vus comme vos plus proches voisins en termes de réponse).

Pour mieux se rendre compte des distances, tout en bas du panneau, après les valences, cliquer sur le bouton *Best Cos<sup>2</sup> Plan*. Cela affichera l'espace dans lequel votre réponse est spatialement la mieux représentée. Les plus proches voisins ont alors tendance à se rapprocher visuellement, ce qui est plus cohérent.

## 5.4. Factor

Cet onglet permet d'appliquer des méthodes de réduction dimensionnelle aux points importés en CSV, ou au jeu de données déjà intégrés (données synthétiques, réponses au questionnaire de personnalité). Il suffit de choisir le nombre d'axes factoriels retenus et la technique utilisée.

### 5.4.1. Axes factoriels

- **Keep All Factors** : garde tous les axes factoriels possibles ; selon la méthode, cela correspond soit au nombre de variables numériques (PCA), aux nombres de modalités de variables catégoriels (MCA), à la somme des deux précédents (FAMD), ou au nombre de classes - 1 (LDA).
- **Nb Factors** : permet de choisir le nombre d'axes à retenir
- **Inference** : utilise, sur de nouvelles données, les matrices de projection déjà calibrées d'une précédente analyse factorielle

### 5.4.2. Techniques factorielles

- **PCA** : analyse en composantes principales (ACP) ; adaptée aux données numériques continues
- **MCA** : analyse des correspondances multiples (ACM) ; adaptée aux données catégorielles
- **FAMD** : analyse factorielles des données mixtes (AFDM) ; adaptée aux données mélangeant des numériques et catégorielles
- **MFA** : analyse factorielle multiple (AFM) ; adaptée aux données organisées en blocs thématiques, chaque bloc étant pondéré pour qu'aucun ne domine la projection du seul fait de son nombre de variables
- **LDA** : analyse factorielle discriminante linéaire de Fisher (AFD) ; adaptée aux données numériques incluant une variable-cible pour séparer des groupes
- **MCA+LDA** : analyse discriminante sur données qualitatives (méthode de Saporta) ; permet d'effectuer une LDA sur des données catégorielles

- **FAMD+LDA** : analyse discriminante sur données mixtes (déclinaison de la méthode de Saporta) ; permet d'effectuer une LDA sur des données mixtes

### Portée du calcul et surcouches

- **Process Individuals** : projette les individus (les lignes) dans l'espace factoriel.
- **Process Features** : projette les variables (les colonnes). Les deux cases sont indépendantes : sur un gros jeu de données, ne calculer que l'un des deux allège sensiblement la scène.
- **Correlation Sphere** : affiche un maillage sphérique unitaire dans l'espace des variables. Une variable numérique proche de la surface ( $\sum \cos^2 \approx 1$ ) est bien restituée par les trois axes affichés ; une variable proche du centre ne l'est pas. Nécessite que les points de variables numériques soient visibles (PCA, FAMD).
- **Partial Stars** : relie une modalité à ses composantes partielles, en analyse factorielle multiple — **selected** (pour le seul point sélectionné), **all** (pour tous) ou **off**.
- **Target** : la variable-cible, requise par les méthodes discriminantes (LDA et ses déclinaisons).
- **Inference** : applique un modèle factoriel déjà entraîné à de nouvelles lignes, au lieu de le réajuster.

Le bouton **Run Factor** lance la projection.

Une fois les données projetées, celles-ci sont visualisables dans l'onglet Workspace, et s'affichent par-dessus le polytope qui sert d'ancrage visuel. Il est recommandé d'utiliser un simple cube (voire un tesseract pour l'analyse 4D) ou une sphère pour faciliter l'exploration.

### 5.4.3. Statistiques associées

Dans l'onglet Factor, vous retrouverez un certain nombre de statistiques relatives à la méthode factorielle :

- Contribution des individus (des lignes) aux axes : un point éloigné sur un axe va fortement "contribuer" cet axe
- Contribution des features (des colonnes) aux axes : idem, mais du point de vue des variables
- Cosinus carré des individus : un point avec un cosinus carré élevé est bien représenté sur cet axe ; sa coordonnée projetée demeure fidèle
- Cosinus carré des features : idem, mais du point de vue des variables

À noter que les axes factoriels ont une correspondance sur les axes spatiaux :

- 1er axe factoriel (PCA\_1, MCA\_1) correspond à l'axe X
- 2e axe factoriel (PCA\_2, MCA\_2) correspond à l'axe Y

Et ainsi de suite avec les axes Z, W, V5, V6, etc.

Ainsi, lorsqu'on analyse les contributions ou cosinus des carrés, la valeur de la liste déroulante correspond à l'axe factoriel. C'est pour cela qu'ils sont affichés sous la forme "X (PCA\_1), Y (PCA\_2)" ou "X (MCA\_1), Y (MCA\_2)".

Mais si le jeu de données contient des coordonnées spatiales (nommées, elles aussi, X, Y, Z), alors on se retrouve à des situations où les axes X, Y, Z (du jeu de données avant projection) contribuent à l'axe X (qui est en fait le premier axe factoriel, après projection). Il faut donc garder en tête ce qui relève du jeu d'entrée et du résultat de la projection factorielle dans l'espace d'arrivée (et qui peut partager les mêmes lettres).

## 5.5. Clustering

Le clustering permet de regrouper les points par proximité spatiale.

### 5.5.1. Sources de données

La source de données à clusteriser comporte des nuances qui activent ou désactivent certaines options.

Les données numériques :

- **Imported Data** : données brutes importées (CSV) ; seules les variables numériques du CSV sont retenues
- **Spatial Data** : données projetées visuellement dans la scène : cela inclut les données numériques précédentes ainsi que les points de référence et les points ajoutés manuellement (onglet Workspace / Dataset) ; c'est donc une version enrichie de la précédente, mais elle n'intègre pas de possibilité de mode *Inference* (cf. ci-dessous)

Les données factorielles :

- **Factor Individuals** : données projetées des individus (lignes)
- **Factor Features** : données projetées des features (colonnes)

Les données de personnalité (Hy-Personality) :

- **Personality Form** : données projetées des répondants au questionnaire de personnalité

### 5.5.2. Méthodes

#### Méthodes de clustering

- **K-Means** : la méthode simple pour aller vite
- **Hybrid** : la méthode complète qui combine K-Means et HCA (classification ascendante hiérarchique) et potentiellement un second K-Means de "nettoyage périphérique"

- **Partition** : un cas particulier qui consiste à assigner, à chaque point spatial, le point de référence le plus proche ; cela nécessite donc de faire apparaître des points de référence (soit les points définis dans { **Dataset** }, soit les points projetés du questionnaire de { **Hy-Personality** })

#### Options spécifiques aux sources **Spatial Data** et **Personality Form**

- **Add Reference Points** : si la source est *Spatial Data*, alors cette option permet d'ajouter les points de référence ajoutés manuellement (onglet Workspace / Dataset) ; si la source est *Personality Form*, alors cette option permet d'ajouter les points projetés de type MBTI, DISC, SS, Zodiac.
- **Reference as Center** : utilise les points de référence comme seeds (points initiaux) pour le premier K-Means ; il désactive nécessairement le choix du nombre de *Centers*

#### Options spécifiques aux sources **Imported Data** et **Factor Data (Individuals ou Features)**

- **Inference** : application d'un modèle de clustering (déjà entraîné) sur de nouvelles données

#### Options communes

- **Standardize** : mise à l'échelle unitaire des données numériques

### 5.5.3. Pre-KMeans

Le premier volet consiste à appliquer une méthode de K-Moyennes sur le jeu initial.

- **N Init** : nombre de tentatives différentes pour déterminer le meilleur partitionnement
- **Use Sqrt(N) Centers** : utilise la racine carré du nombre de points comme nombre de centroïdes pour effectuer le partitionnement
- **N Centers** : si la règle de la racine carré est décochée, alors ce slider règle le nombre de centroïdes désiré. En général, un nombre situé entre la racine cubique et la racine carrée de la taille initiale (nombre de lignes) est un bon point de départ
- **Filter Outliers** : permet d'assigner les points isolés dans une classe à part ; se base sur la distance de Mahalanobis
- **Outlier Confidence %** : niveau de confiance utilisée pour définir les outliers

### 5.5.4. HCA (CAH)

La deuxième étape est une classification ascendante hiérarchique (méthode de Ward) appliquée aux centroïdes du Pre-KMeans.

- **HCA Auto-Cutoff** : détermine automatiquement le nombre de clusters finaux
- **Cutoff Metric** : métrique utilisée pour déterminer automatiquement le nombre de clusters ; peut se baser sur le coefficient de silhouette ou la courbe d'accélération du dendrogramme (le coude sur la courbe d'inertie inter-classes)

- **Min Nb Clusters** : nombre minimum de clusters souhaités (permet de chercher d'autres optima dans la méthode précédente)
- **Max Nb Clusters** : nombre maximum de clusters souhaités (permet de restreindre la recherches des optima dans la méthode précédente)
- **Nb Clusters** : si le mode Auto-Cutoff n'est pas coché, alors ce slider détermine le nombre de clusters finaux

### 5.5.5. Final Cleaning (nettoyage final)

Selon la distribution spatiale des données, les méthodes HCA peuvent souffrir d'allocations contestables pour les points situés en périphérie des clusters.

- **Post-KMeans** : réaffectation de Voronoï après la CAH. Une K-moyennes effectuée sur les barycentres de clusters de la CAH réaffecte les points périphériques au cluster le plus intuitif géométriquement, ce qui corrige les cellules de frontière.
- **Merging Threshold %** : à la toute fin, ce pourcentage permet de fusionner les clusters qui n'ont pas assez d'individus (proportion fixée par ce curseur)

Le bouton **Run Clustering** lance l'algorithme ; le bouton **Clear** réinitialise les résultats.

### 5.5.6. Ellipsoïdes de cluster

Une enveloppe de covariance peut être dessinée autour de chaque cluster, dans la scène 3D. C'est la lecture visuelle la plus directe de la qualité d'une segmentation : deux enveloppes qui se chevauchent signalent des clusters ambigus.

- **Show Ellipsoids** : affiche une coque de covariance translucide par cluster.
- **$k \cdot \sigma$**  : demi-étendue de la coque, exprimée en écarts-types.
- **Opacity** : transparence des coques.

### 5.5.7. Statistiques associées

#### Métriques de qualité

Les métriques de qualités affichées sont : coefficient de silhouette, Davies-Bouldin Index (DBI), Calinski-Harabasz Index, Dunn Index

#### Graphiques

La section suivante affiche le dendrogramme de la CAH, et les courbes d'accélération de l'inertie inter-classes. Puis vient la distribution de la population par cluster, les boîtes à moustache par variable spatiale, et les matrices de contingence avec les variables d'intérêt (Segment, Target, Treatment).

## 5.6. Sup ML (machine learning supervisé)

Sup ML permet d'entraîner des modèles de régression (linéaire) et de classification.

### 5.6.1. Paramètres

- **Source** : sélection des données à modéliser : soit les données brutes importées, soit les données projetées par méthode factorielle (sur les individus seulement)
- **Target** : variable-cible à expliquer
- **Estimator** : choix du modèle : régression linéaire ou régression logistique
- **Score Bins** : nombre de quantiles utilisés pour discrétiser le score prédit ; utile pour effectuer des analyses de profilage sur les scores (onglet Stats, section Profiling)
- **Inference** : application d'un modèle de clustering (déjà entraîné) sur de nouvelles données
- **Cluster Predictor** : utilise les classes de clustering (si elles existent) comme variables explicatives supplémentaires

Le bouton **Run Estimator** lance l'algorithme ; le bouton **Clear** réinitialise les résultats.

### 5.6.2. Statistiques associées

Les résultats sont disponibles dans deux onglets, un par échantillon (Train, Test).

#### Régression linéaire

Les métriques de qualités affichées sont : R2, MAE, RMSE.

Un panneau affiche les coefficients de la régression.

#### Régression logistique

Les métriques de qualités affichées sont : Balanced Accuracy, Macro Precision, Macro F1, ROC AUC, Gini.

Un panneau affiche la matrice de confusion, puis les coefficients de la régression (et les odds ratios associés).

Enfin, deux courbes sont proposées : ROC et lift (gains cumulés).

#### Scores et scores quantiles

Les scores du modèle sont discrétisés en score quantiles. Ces quantile peuvent être utilisés comme variable de groupes dans le profilage.

## 5.7. Optimization

Ce module cherche, parmi les individus du jeu de données, ceux qui maximisent ou minimisent une variable cible, sous les contraintes définies par les filtres actifs. Le résultat peut être projeté dans la scène (voir 4.2.12).

### 5.7.1. Paramètres

- **Target** : la variable à optimiser.
- **Direction** : le sens de l'optimisation — **maximize** ou **minimize**.
- **Individual ID** : la colonne identifiant chaque individu dans les résultats.
- **Error Margin** : marge de tolérance admise autour de l'optimum. Élargir cette marge fait remonter un ensemble de solutions quasi optimales plutôt qu'un point unique, ce qui est souvent plus exploitable.
- **Seed** : graine du générateur aléatoire. À graine fixée, deux exécutions donnent le même résultat — indispensable pour comparer deux paramétrages.

Le bouton **Run** lance la résolution ; **Clear** réinitialise les résultats.

### 5.7.2. Statistiques associées

Les résultats listent les individus retenus, la valeur atteinte par la cible, et l'écart à l'optimum théorique. La vue 3D associée ( **Optimization 3D View** , voir 4.2.12) permet d'afficher soit le plan de sélection, soit le polytope des solutions admissibles.

## 5.8. Forecast

Le modèle de prévision temporelle s'appuie sur un exponentiel de Holt-Winters.

### 5.8.1. Paramètres

- **Time Column** : variable temporelle (peut être un simple index)
- **Target** : variable-cible à prédire dans le temps
- **Seasonal Period** : nombre de périodes qui définissent une saison
- **Seasonality** : type de saisonnalité : additive ou multiplicative
- **Auto-fit  $\alpha/\beta/\gamma$**  : détermine automatiquement les coefficients du lissage exponentiel, par recherche sur grille minimisant l'erreur dans l'échantillon. Décoché, les trois coefficients se règlent à la main :
  - **Alpha (level)** : lissage du niveau (la hauteur de la courbe)
  - **Beta (trend)** : lissage de la tendance (l'inclinaison)
  - **Gamma (season)** : lissage de la composante saisonnière (la périodicité)
- **Back-Test Holdout** : nombre de points retirés du jeu d'entraînement pour tester le modèle
- **Inference Horizon** : nombre de points supplémentaires à prédire (données postérieures aux périodes importées)

Le bouton **Run Forecast** lance le calcul dans le navigateur. **Run Large Forecast** délègue le calcul au moteur Python distant : c'est la voie à prendre sur les gros volumes, où le calcul local devient trop lent.

## 5.8.2. Statistiques associées

### Métriques

Les résultats sont composés de courbes de visualisation ainsi que de métriques de qualité (dans les échantillons d'apprentissage et de back-test) : MAE (écart absolu moyen), RMSE (écart quadratique moyen), MAPE (pourcentage d'erreur absolu moyen).

### Graphiques

Les graphiques affichent les courbes réelles, calibrées ("fittées") et prédites.

On dispose également d'une décomposition des 3 piliers de la fonction construite : le niveau (la hauteur), la tendance (l'inclinaison), la saison (la périodicité).

## 5.9. Hy-Performance

Fonctionnalité bientôt disponible.

Objectif : étudier l'efficacité d'un traitement (essai clinique ou action marketing) sur une population, au moyen d'un A/B test intégrant des évaluations statistiques.

### 5.9.1. Sources de données

Données granulaires ou données agrégées

- **Keep All Factors** : garde t
- **Nb Factors** : perm
- **Inference** : utilise, se

### 5.9.2. Paramètres

Type d'analyse : absolue ou incrémentale

- **Control Group** : coché, chaque action est évaluée par un test cible-contre-contrôle. Décoché, l'estimation se fait en bras unique (préfixe **SOLO\_**), par intervalle de confiance — c'est le mode à utiliser quand le design n'a pas prévu de groupe témoin.
- **Alpha** : niveau de significativité (erreur de première espèce) pour rejeter l'hypothèse nulle
- **Alternative** : unilatéral (**Greater** = à droite ; **Less** = à gauche) ou bilatéral (**Two-sided**)
- **Multi-Test Correction** : correction pour tests multiples — **None (per action)** teste chaque action isolément ; **BH-FDR** re-arbitre l'ensemble des indicateurs de significativité sur le seuil de Benjamini-Hochberg, calculé sur toute la famille testée. Cette seconde option contrôle le taux de fausses découvertes au niveau alpha, et sa décision se propage aux montants incrémentaux.

- **Expected Ratio (T/C)** : garde-fou contre le déséquilibre d'échantillon. On y déclare le ratio cible/contrôle prévu par le design (1 = 50/50, 9 = 90/10). Un test du  $\chi^2$  par action alerte quand le partage réellement observé s'en écarte. Voir 5.9.4 pour le détail. La valeur 0 désactive le contrôle.

### Seuils d'analysabilité (fiabilité des mesures et instabilité liée au volume)

- **Analyz Min Conv** : seuil minimal pour considérer une action analysable
- **Max Relative Prec (%)** : valeur maximale tolérée du coefficient de variation, pour considérer une action analysable
- **Vol Increase Method** : méthode utilisée pour définir la stabilité des tests statistiques quand le volume augmente : on peut choisir d'augmenter la taille des témoins (Controls Only) ou bien les témoins et les traités (All Groups)
- **Vol Increase Factor** : facteur d'agrandissement des groupes pour tester la stabilité au volume
- **Amount Formula Switch** : valeur de convertis (dans chaque groupe) qui permet de choisir le type de formules utilisées pour les calculs d'uplifts ; quand un groupe a moins de convertis que le chiffre indiqué ici, alors on utilise les formules basées sur les volumes, sinon on utilise les formules basées sur les montants

### Calculs de montants incrémentaux

- **Inlier Salvage** : méthode de récupération/préservation des actions à faible volume dont la précision des moyennes est suffisante
- **Inlier Min Conv** : valeur minimale de convertis pour qu'une action soit "récupérable"
- **Inlier Max CV** : valeur maximale tolérée du coefficient de variation, pour considérer une action à faible volume "récupérable"

### Gestion des ROI

- **Breakeven K0 → 0** : les actions qui n'atteignent pas le break-even (les montants des convertis traités ne couvrent pas le budget) sont remplacés par zéro (plutôt que par des valeurs négatives)
- **Isolate Non-Zero IA** : si l'option est cochée, alors les actions significatives sont isolées et ne sont pas temporellement regroupées ; dans le cas contraire, seules les actions ROI+ (ROI positifs) sont isolées, mais les autres actions sont bien regroupées dans le temps.

## 5.9.3. Statistiques associées

#### PERIOD

Period of aggregation for the groups of actions and the financial KPIs: Q1, Q2, Q3, Q4, H1, H2, FY

#### AGG\_TYPE

Is the aggregate evaluated on actions (before grouping, coded "BG") or on groups of actions (after grouping, coded "AG")

YEAR

Year of the action

HALF\_YEAR

Semester of the action

QUARTER

Quarter of the action

MONTH

Month of the action

WEEK

Week of the action

DAY

Day of the action

STRATEGY

Strategy followed by the action

FLAG\_OUT

Remove from analysis

LEVEL1

Typology of the action (1st level category)

LEVEL2

Typology of the action (2nd level category)

LEVEL3

Typology of the action (3rd level category)

LEVEL4

Typology of the action (4th level category)

SEGMENT

LEVEL1 - LEVEL2 - LEVEL3 - LEVEL4

ACTION\_GROUP

(Month) - Strategy - Segment - ROI Flag - (ROI Trend)

Month and Trend are only appearing for certain ROI flags and to\_group flag

MARGIN

Margin Profit (in % of outstanding)

#### AMOUNT\_FACTOR

Factor used to assess the profit from margin and amount:  $\text{PROFIT} = \text{AMOUNT} \times \text{AMOUNT\_FACTOR}$   
MARGIN

Useful when margins depend on two multiplied variables (eg. monthly subscription price and number of months subscribed)

#### UNIT\_COST

Unit communication cost

#### BUDGET

Budget of the action (unit cost \* nb targets)

#### ANALYZABLE\_BUDGET

Part of the budget (in €) that is analyzable (ROI\_FLAG is not "NA")

#### ANALYZABLE\_BUDGET\_RATIO

Part of the budget (in %) that is analyzable (ROI\_FLAG is not "NA")

#### POSITIVE\_ROI\_BUDGET

Part of the budget (in €) that is profitable (ROI\_FLAG is "ROI+")

#### POSITIVE\_ROI\_BUDGET\_RATIO

Part of the budget (in %) that is profitable (ROI\_FLAG is "ROI+")

#### POS\_ROI\_ANA\_BUDGET\_RATIO

Part of the analyzable budget (in %) that is profitable (ROI\_FLAG is "ROI+")

#### NB\_T

Nb of Targets

#### NB\_C

Nb of Controls

#### CG\_SIZE

Control Group Size (%)

#### NB\_CONV\_T

Nb of Converted Targets

#### NB\_CONV\_C

Nb of Converted Controls

#### CONV\_RATE\_T

Conversion Rate Target = Conv Rate Target

#### CONV\_RATE\_C

Conversion Rate Control = Conv Rate Control

UPLIFT\_NB

Uplift in volume (delta rates in percentage points)

UPLIFT\_NB\_RATIO

Uplift in Volume Ratio (%)

AMOUNT\_T

Amount (Total Converted Amount) Target

AMOUNT\_C

Amount (Total Converted Amount) Control

SQUARED\_AMOUNT\_T

Squared Amount (Total Squared Converted Amount) Target

SQUARED\_AMOUNT\_C

Squared Amount (Total Squared Converted Amount) Control

WEIGHTED\_SQUARED\_AMOUNT\_C

Scaled Squared Amount Control

PROFIT\_A

Profit All Customers (Targets + Controls)

PROFIT\_T

Profit Target

PROFIT\_C

Profit Control

AVG\_AMOUNT\_A

Average Amount per Customer

AVG\_AMOUNT\_T

Average Amount per Target

AVG\_AMOUNT\_C

Average Amount per Control

AVG\_AMOUNT\_CONV\_A

Average Amount per Converted Customer

AVG\_AMOUNT\_CONV\_T

Average Amount per Converted Target

AVG\_AMOUNT\_CONV\_C

Average Amount per Converted Control

VAR\_AMOUNT\_T

Variance of the Amount Target

VAR\_AMOUNT\_C

Variance of the Amount Control

VAR\_AMOUNT\_CONV\_T

Variance of the Converted Amount Target

VAR\_AMOUNT\_CONV\_C

Variance of the Converted Amount Control

UPLIFT\_AMOUNT

Uplift in amount per customer (delta of the average amounts)

UPLIFT\_AMOUNT\_RATIO

Uplift in Amount Ratio (%)

UPLIFT\_AMOUNT\_CONV

Uplift in amount per converted customer (delta of the average tickets)

UPLIFT\_AMOUNT\_CONV\_RATIO

Uplift in Amount per Converted Customer Ratio (%)

UPLIFT\_TOTAL\_RATIO

Total Uplift Ratio (%)

The six columns below are the additive carriers the seven UPLIFT columns above are rebuilt from on an aggregated row (see 5.9.4.5). They are summed like any other total, then the uplifts are re-derived from them, for instance  $UPLIFT\_NB = (NB\_CONV\_T\_CG - CF\_NB\_CONV) / NB\_T\_CG$ . The `_CG` suffix marks the standardisation perimeter: cells with no control group are left out of both the numerator and the denominator.

NB\_T\_CG

Nb of Targets, restricted to the cells having a control group

NB\_CONV\_T\_CG

Nb of Converted Targets, restricted to the cells having a control group

AMOUNT\_T\_CG

Amount Target, restricted to the cells having a control group

CF\_NB\_CONV

Counterfactual converts: converted targets expected WITHOUT treatment, at target scale (per cell,  $NB\_T \times CONV\_RATE\_C$ )

CF\_AMOUNT

Counterfactual amount: amount expected WITHOUT treatment, at target scale (per cell,  $NB\_T \times AVG\_AMOUNT\_C$ )

CF\_AMOUNT\_CONV

Counterfactual converted amount: amount per converted customer expected WITHOUT treatment, at target scale (per cell,  $NB\_CONV\_T \times AVG\_AMOUNT\_CONV\_C$ )

CI\_VALID\_T

Is the Confidence Interval of RR valid?

CI\_VALID\_C

CI\_SE\_NB\_T

Standard Error of the Resp Rate Confidence Interval

CI\_SE\_NB\_C

CI\_LOWER\_NB\_T

Lower bound of the Resp Rate Confidence Interval

CI\_LOWER\_NB\_C

CI\_UPPER\_NB\_T

Upper bound of the Resp Rate Confidence Interval

CI\_UPPER\_NB\_C

CI\_SE\_AMOUNT\_(CONV)\_T

Standard Error of the Average Amount Confidence Interval \*[1]

CI\_SE\_AMOUNT\_(CONV)\_C

CI\_LOWER\_AMOUNT\_(CONV)\_T

Lower bound of the Average Amount Confidence Interval \*[1]

CI\_LOWER\_AMOUNT\_(CONV)\_C

CI\_UPPER\_AMOUNT\_(CONV)\_T

Upper bound of the Average Amount Confidence Interval \*[1]

CI\_UPPER\_AMOUNT\_(CONV)\_C

VOLUME\_INCREASE\_FACTOR

("VOL\_INC", see \*[2])

Strategy to increase volumes: it impacts NB, NB\_CONV, CONV\_RATE.

Note: this metric is not visible in the output. It's only an intermediate within the code.

Z\_STAT

Z Statistic

Z\_NORM\_PROBA

Cumulative probability of the Z-Stat regarding the Normal distrib

Z\_NORM\_P\_VALUE

P-Value of the Z-Stat regarding the Normal distrib

ODDS\_RATIO

Odds Ratio of the CONVancing contingency table

Z\_HYPERGEOM\_PROBA

Cumulative probability of the Z-Stat regarding the Hypergeometric distrib

Z\_HYPERGEOM\_P\_VALUE

P-Value of the Z-Stat regarding the Hypergeometric distrib

P\_VALUE\_NB

Retained P-value for volume significance

SS\_NB

Is the uplift in volume significant?

T\_STAT\_(CONV)

T Statistic \*[1]

T\_NORM\_PROBA\_(CONV)

Cumulative probability of the T-Stat regarding the Normal distrib \*[1]

T\_NORM\_P\_VALUE\_(CONV)

P-Value of the T-Stat regarding the Normal distrib \*[1]

T\_STUD\_PROBA\_(CONV)

Cumulative probability of the T-Stat regarding the Student distrib \*[1]

T\_STUD\_P\_VALUE\_(CONV)

P-Value of the T-Stat regarding the Student distrib \*[1]

P\_VALUE\_AMOUNT\_(CONV)

Retained P-value for amount significance \*[1]

SS\_AMOUNT\_(CONV)

Is the uplift in amount significant? \*[1]

SS\_DOUBLE\_(CONV)

Means "Significant in both amount and nb" \*[1]

SS\_ZERO\_(CONV)

Means "Not Significant in neither amount nor nb" \*[1]

SS\_DIV\_(CONV)

Means "Significant in amount (or nb) but not in nb (or amount)" \*[1]

RAW\_ANALYZABILITY

Is the action / action-group roughly analyzable regarding the CG existence and the Nb of Conversions?

CI\_LOWER\_NB\_C\_WITH\_CONV\_RATE\_T

Theoretical value of the Control's lower Conf Interval bound if we assume that  $CONV\_RATE\_C = CONV\_RATE\_T$

TRUE\_ZERO\_CONTROL

Is the zero number of Converted Control a reality or an issue due to low volume?

NET\_ANALYZABILITY

Is the action / action-group deCONVitably analyzable regarding the Conversions and the statistical significance stability?

NOTE FOR MULTI-EFFECT

- If one effect is analyzable then the global effect is analyzable

IA\_PC\_AMOUNT

Credit: additional Amount per customer based on amounts

Insurance: additional probability per customer based on amounts

IA\_PC\_NB

Credit: additional Amount per customer based on volumes

Insurance: additional probability per customer based on volumes

IA\_COHERENCE\_FLAG

Credit: retained formula for additional Amount

Insurance: retained formula for additional probability

"2" means amount-based formula

"1" means volume-based formula

"0" means insignificant ( $IA = 0$ )

The "New" method is also named "coherence rule": a action is coherent if, when increasing UPLIFT\_NB, we do not decrease AVG\_AMOUNT\_CONV\_T.

The "Old" method is also named "retake rule": if we are not significant in amount, then we might "retake" the IA from the significance in nb.

IA\_PC

Credit: retained additional amount per customer

Insurance: retained additional probability per customer

IA

Credit: retained additional Amount

Insurance: retained additional probability

ROI\_PC

ROI per customer

Note for multi-effect

> The global-effect-ROI-PC is calculated by computing the underlying effect-IA-PC (multiplied by their respective margin/lifetime) and by subtracting the unit cost once! → it's not the simple sum of the underlying effect-ROI-PC!

- If the global effect is NOT analyzable, then the ROI is cut to zero

ADD\_PROFIT\_PC

Additional Profit per customer

ROI

Total ROI

ADD\_PROFIT

Total additional Profit

ROI\_PU

ROI per unit invested

ROI\_PU\_ANA

ROI per unit invested and analyzable

ROI\_MAX\_PC

Theoretical maximum ROI per customer (if control never get Converted)

ROI\_MAX

Theoretical maximum ROI (if control never get Converted)

ROI\_FLAG

Typology of ROI/IA (situations of profitability)

TO\_GROUP

Is the action groupable?

We want to group the actions that are not analyzable or not significant

ROI\_AMOUNT\_WITHOUT\_SIGNIF

Theoretical ROI (based on amounts) without significance

ROI\_NB\_WITHOUT\_SIGNIF

Theoretical ROI (based on volumes) without significance

ROI\_TREND

Has the action (groupable) a positive or negative trend of ROI?

## 5.9.4. Guide méthodologique

Hy-Perf est conçu pour *refuser de conclure* quand les données ne permettent pas de conclure. Cette section explique les garde-fous statistiques, pour lire correctement les résultats - et leurs silences.

### 5.9.4.1. Analysabilité : quand l'outil refuse de conclure

Chaque action doit passer des contrôles découplés avant qu'un uplift ou un ROI ne soit retenu :

- **Estimabilité** ( `Analyz Min Conv` , défaut 10) : sous ~10 conversions, taux et moyennes ne sont tout simplement pas estimables - l'action est marquée NA.
- **Précision** ( `Max Relative Prec (%)` , défaut 20 %) : l'erreur-type relative des moyennes mesurées doit rester sous le seuil ; un effet « significatif » mais très imprécis n'est pas actionnable.
- **Routage des formules** ( `Amount Formula Switch` , défaut 50) : sous 50 conversions par groupe, les formules en montants sont trop instables - le moteur bascule sur les formules en volumes.

**Un NA n'est pas un bug : c'est l'outil qui vous dit que l'échantillon ne peut pas encore répondre à la question.** Les actions NA sont re-testées sous une augmentation simulée de volume ( `Vol Increase Method` / `Vol Increase Factor` ) : certaines sont « récupérables » - à relancer plus tard avec plus de volume - plutôt que des impasses. Les budgets des actions non analysables sont exclus des KPI analysables au lieu d'être comptés silencieusement.

### 5.9.4.2. Multiplicité et le défaut alpha = 0.1

Chaque action x effet x segment est testée à `Alpha` = 0.1 par défaut, sans correction entre tests. Ce défaut est délibéré : la philosophie du produit est une décision par action, pas de la découverte de masse. Mais il est permissif pour des décisions budgétaires : testez 100 actions sans effet réel et environ 10 sortiront quand même « significatives » par hasard. Le dashboard affiche ce nombre attendu de faux positifs à côté des agrégats.

Quand vous criblez un large catalogue d'actions et que l'image agrégée prime, passez `Multiplicity` sur BH-FDR (panneau Statistical Inference) : chaque flag de significativité est re-gaté sur le cutoff de Benjamini-Hochberg, contrôlant la part attendue de fausses découvertes à alpha, et se propage jusqu'aux IA et ROI.

### 5.9.4.3. Winner's curse sur les agrégats

Les KPI agrégés (budgets analysables, montants incrémentaux totaux, ROI de bucket) sont construits à partir des actions ayant passé le filtre de significativité. Cette sélection est optimiste par construction : parmi des mesures bruitées, celles qui passent sont plus souvent celles qui ont eu de la chance vers le haut. Le dashboard encadre donc l'agrégat : ROI total **avec** le filtre de significativité vs **sans** (colonnes ROI\_AMOUNT\_WITHOUT\_SIGNIF / ROI\_NB\_WITHOUT\_SIGNIF) - la vérité se situe généralement entre les deux.

### 5.9.4.4. SRM - déclarer le ratio cible/contrôle attendu

Si votre design assigne cible et contrôle selon un ratio connu (par ex. 90/10), déclarez-le via le contrôle `Expected Ratio (T/C)` ( `srm_expected_ratio` ; 0 = désactivé). Le moteur exécute alors un test du  $\chi^2$  de sample-ratio-mismatch par action (colonnes SRM\_CHI2 / SRM\_P\_VALUE / SRM\_WARNING) : un split réel qui dévie significativement du ratio déclaré signale le plus souvent un groupe de contrôle contaminé, le tueur silencieux le plus fréquent des analyses A/B en CRM. Sans ratio déclaré, le contrôle reste désactivé et ne lève jamais de fausse alerte.

### 5.9.4.5. Uplifts agrégés - standardisés sur la population ciblée

Un uplift est un écart entre deux taux, un par groupe. Le recalculer à partir des totaux d'une maille agrégée donne un chiffre faux : le bras cible s'agrége pondéré par ses effectifs cible, le bras contrôle par ses effectifs contrôle, alors que le résultat doit être pondéré par la cible - c'est là que le budget a été dépensé. Dès que la taille du groupe de contrôle varie d'une cellule à l'autre **et** qu'elle est corrélée au taux de base, l'écart se déplace ; il peut dépasser l'uplift lui-même et en inverser le signe (c'est le paradoxe de Simpson). Sur le jeu de test de référence, cette lecture surestimait l'uplift en volume de 54 % au total portefeuille.

Les colonnes UPLIFT d'une ligne agrégée sont donc **standardisées sur la population ciblée** : le moteur transporte depuis la maille fine six colonnes additives (voir 5.9.3) et re-dérive les sept uplifts à partir de ces sommes, à n'importe quelle granularité. Au niveau d'une cellule unique la valeur est inchangée ; seules les lignes agrégées bougent, et c'est exactement la correction voulue.

Trois conséquences à connaître pour les lire :

- **Ce sont des indicateurs descriptifs, pas des chiffres de décision.** Contrairement à IA, ils ne sont **pas** filtrés : les actions non analysables ou non significatives y comptent leur écart brut. IA reste le chiffre gaté sur lequel arbitrer, et IA / NB\_T sa forme standardisée par client. Ne reconstruisez jamais IA à partir des uplifts - le produit UPLIFT\_NB × AVG\_AMOUNT\_CONV\_A est le pire chemin, une moyenne de produits n'étant pas le produit des moyennes.
- **Les cellules sans groupe de contrôle sortent du périmètre.** Sans contrôle il n'y a pas de contrefactuel : les compter reviendrait à leur créditer la totalité de leurs conversions comme incrément. Elles quittent donc le dénominateur aussi bien que le numérateur. L'uplift publié est la moyenne sur les cellules **qui ont** un contrôle, pas sur toutes.

- **Sans groupe de contrôle du tout, la lecture est mono-bras.** UPLIFT\_NB vaut alors le taux de conversion cible, à toutes les mailles, et la nomenclature bascule sur les libellés mono-bras (voir 5.9.2).

## 5.10. MMM

Le MMM (Marketing Mix Modeling) est un puissant outil qui permet d'évaluer la contribution de leviers d'investissement à une variable-cible temporelle.

### 5.10.1. Paramètres

- **Time Column** : variable temporelle (peut être un simple index)
- **Target** : variable-cible à prédire dans le temps

#### Les canaux d'investissements

Cette section permet d'indiquer au modèle quelles sont les colonnes qui représentent les leviers marketing. Cette assignation est indispensable pour que le modèle fonctionne correctement.

#### Baseline

La baseline est la contribution "naturelle" à la variable-cible quand tous les autres investissements s'arrêtent.

- **Trend** : ajoute une inclinaison à la baseline
- **Seasonal Period** : pour spécifier la saisonnalité

#### Adstock

- **Carryover (max lag)** : fenêtre d'adstock géométrique — nombre de périodes passées sur lesquelles une dépense continue d'agir ; effet de rémanence (loi géométrique) d'un investissement sur plusieurs périodes après qu'il a eu lieu
- **Max Delay** : temps d'attente / retard (en périodes) maximal entre le moment de l'investissement et l'apparition de son effet sur la cible

#### Ridge

- **L2 Penalty** : force de la régularisation ridge pour gérer la multi-collinéarité ; elle tire la contribution des canaux vers zéro pour supprimer le bruit inutile
- **Non-Negative media** : contrainte qui interdit aux médias d'avoir une contribution négative

#### ROI Prior

- **Prior Weight** : force globale pour les priors de ROI ; cette valeur est appliquée par défaut sur chaque canal (lorsque la valeur du poids du canal est  $< 0$ )
- **ROI - <canal>** : valeur du ROI théorique (contribution) pour chaque canal (optionnel)

- **W - <canal>** : force du prior ROI du canal ; cette valeur remplace la valeur de Prior Weight lorsque celle-ci est  $\geq 0$  ; lorsque la valeur est  $< 0$ , alors c'est Prior Weight qui sert de valeur de remplacement

## Back-Test

- **Holdout** : nombre de points retirés du jeu d'entraînement pour tester le modèle
- **Reliability Bootstrap** : nombre d'échantillons utilisés pour établir l'intervalle de confiance des contributions à 90%

## 5.10.2. Statistiques associées

### Métriques

Les résultats sont composés de courbes de visualisation ainsi que de métriques de qualité (dans les échantillons d'apprentissage et de back-test) : R2, MAE (écart absolu moyen), RMSE (écart quadratique moyen), MAPE (pourcentage d'erreur absolu moyen).

### Graphiques

Le premier graphique représente les tranches de contributions.

Le second tableau expose les contributions de chaque canaux, ainsi que de quelques métriques associées (adstock, retard, investissement total, facteur d'inflation de variance, etc.).

Le troisième graphique affiche les courbes de réponse (courbes de saturation) qui représentent l'évolution de la contribution en fonction du budget. On note la présence d'un point particulier sur chaque courbe : le coude, c'est-à-dire le moment où la courbe commence rapidement à plafonner.

## 5.10.3. Guide méthodologique

Le MMM de Tetrachore tourne *entièrement dans votre navigateur* : pas de cloud, pas de transfert de données, pas de logiciel MMM à installer - un Marketing-Mix Model complet sur données locales. Cette autonomie s'accompagne d'un devoir d'honnêteté : cette section dit ce dont le modèle a besoin, ce que ses chiffres signifient, et quand il ne faut PAS les croire.

### 5.10.3.1. Prérequis données

- **Au moins 52 périodes régulières** - l'hebdomadaire est l'idéal, 104 semaines est confortable. Le moteur avertit sous 52 points et refuse les timestamps dupliqués.
- **Cadence régulière** : les lignes sont traitées comme équidistantes ; trous ou fréquences mélangées déforment le carryover - un avertissement pré-flight se déclenche quand la cadence semble irrégulière.
- **Dépense variable par canal** : un canal qui dépense le même montant à chaque période est inidentifiable (erreur bloquante). Le flighting - alternance on/off des dépenses - est ce qui rend l'attribution possible.

- **Une cible propre** : un KPI cohérent (ventes, conversions, revenu) sans changement de régime inexplicé ; ajoutez des colonnes de contrôle pour les chocs connus (prix, distribution, promos).
- **Saisonnalité couverte au moins deux fois** : une saisonnalité de période 52 exige 2+ ans de données, sinon les dummies saisonnières mangent le signal média (un avertissement se déclenche, avec celui du ratio paramètres/observations).

### 5.10.3.2. Lire les résultats

Pour les profils métier :

- **ROI** = contribution modélisée / dépense sur la fenêtre d'analyse. C'est une attribution par modèle, pas une expérimentation : la meilleure explication cohérente de votre historique, pas un fait causal mesuré.
- **robuste ✓** signifie que la contribution du canal est restée au-dessus de zéro à travers les rééchantillonnages bootstrap. C'est un indicateur de stabilité, PAS un test de significativité statistique.
- **Les notes de run comptent autant que les chiffres** : elles signalent non-convergence, saturation non identifiée (courbes en pointillés), collinéarité et problèmes de qualité de données. Un run couvert de notes est un run à ne pas croire.
- Pour les profils data :
- Menez avec le **MAPE holdout** : le backtest est chronologique et sans fuite (transforms, standardisation et priors sont recalibrés sur la tête d'apprentissage seule). Le R2 in-sample est la métrique la plus flatteuse - un simple sanity check.
- **Les intervalles de confiance** sont des intervalles bootstrap par blocs conditionnels aux transforms calibrés : l'incertitude de calibration n'y est pas incluse, l'incertitude réelle est donc plus large que l'intervalle affiché.
- **VIF > 10** sur un canal signifie que sa part individuelle ne peut pas être séparée de façon fiable des canaux corrélés - la contribution combinée peut être juste, la répartition entre eux ne l'est pas.

### 5.10.3.3. Trois limites structurelles

- **Attribution sans expérimentation** : un MMM par régression lit des motifs de corrélation dans le temps. Quand vous disposez de mesures expérimentales (geo-tests, lift studies), injectez-les comme priors de ROI ( **Prior Weight** , contrôles **ROI** / **W** par canal) - c'est exactement leur rôle, et c'est l'équivalent fréquentiste des priors bayésiens de Meridian.
- **Biais de bord** : l'adstock monte en charge au début de la série, et le carryover généré après la dernière période n'est jamais crédité - le ROI est légèrement sous-estimé pour les canaux à forte rémanence.

- **Forme de la saturation** : la pente de Hill est fixée à 1, les courbes de réponse sont donc concaves et saturantes, jamais en S. Quand la calibration pousse la demi-saturation d'un canal en haut de sa grille de recherche, la courbe est tracée **en pointillés** avec un coude évidé : **la saturation n'est pas identifiée pour ce canal - ne lisez pas un coude d'allocation budgétaire sur une courbe indicative.**

#### 5.10.3.4. Quand ne PAS croire le modèle

- Moins de ~52 points, cadence irrégulière, ou un run couvert d'avertissements.
- Le canal qui vous intéresse a une courbe de réponse en pointillés (saturation non identifiée) ou un VIF > 10.
- Le MAPE holdout est très au-dessus du MAPE train (surapprentissage), ou le solveur signale une non-convergence.
- Dépense concentrée sur un seul burst, ou deux canaux toujours actifs ensemble - aucun modèle ne peut les séparer.
- La cible a traversé un choc structurel (pandémie, refonte tarifaire) sans colonne de contrôle pour le couvrir.

#### 5.10.3.5. Validation et reproductibilité

Le moteur est benchmarké contre le dataset simulé public de Meridian : les classements de canaux sont compatibles au sens des intervalles de confiance, et les niveaux de contribution rentrent dans la bande bayésienne dès que des priors de ROI faibles sont utilisés (l'écart sans prior est documenté, avec ses causes). Une analyse de sensibilité seedée vérifie que le classement des ROI est stable sous bruit rééchantillonné. Chaque résultat est estampillé d'un `engineVersion` et d'un `configHash` : mêmes données, même config, même version ⇒ même sortie, vérifiable.

## 5.11. Hyperion

## 6. Import / Export

---

Localisation : onglet **Export** . Les extractions analytiques sont également accessibles depuis chaque onglet d'analyse.

### 6.1. Hypersolid

- **View (PNG)** : exporte l'image de visualisation en cours, au format PNG. La capture reprend exactement la scène affichée, arrière-plan compris — si **Transparent** est coché dans les réglages d'arrière-plan (voir 2.2.4.1), le PNG est produit sur fond transparent.
- **Copy Rotation** : copie dans le presse-papiers l'état de rotation courant du solide (tous les plans, ND compris), sous une forme réutilisable comme vue personnalisée.

### 6.2. Config

Vous pouvez exporter puis réimporter les configurations du workspace, c'est-à-dire l'ensemble des réglages de ses panneaux.

#### Export

- **Export All Config** : exporte tous les réglages au format JSON.

Vous pouvez aussi n'exporter qu'une sous-partie :

- **Export Hypersolid Config** : réglages de génération et de style appliqués au solide.
- **Export Projection Config** : réglages de rotation, de cisaillement et de projection.
- **Export Data Config** : réglages de style appliqués aux données projetées.

#### Import

- **Import Config** : réimporte un fichier de configuration précédemment exporté.

### 6.3. Scene

- **Projected Data (CSV)** : exporte les données spatiales projetées au format CSV.

## 6.4. Analytics

Localisation : onglet **Export** , ou bien en haut à droite de chaque onglet analytique (Stats, Factor, Clustering...).

On retrouve toutes les extractions possibles, rangées par module :

- Stats : statistiques descriptives, corrélation, profilage, association
- Hy-Personality : les métriques liées aux réponses projetées du formulaire de personnalité
- Factor : les contributions, cosinus et variance liées à la projection factorielle
- Cluster : les métriques, courbes et tableaux de contingence du clustering
- Sup ML : les métriques, coefficients et courbes des modèles de régression/classification
- Forecast / Time Series : les métriques du lissage exponentiel (Holt-Winters)
- Hy-Performance : les tableaux d'actions, de stratégies et de duels
- MMM : les contributions, courbes en tranches et courbes de saturation du MMM

## 6.5. PDF Recap

- **Charts Recap (PDF)** : rassemble en un seul PDF les graphiques produits par les modules d'analyse au cours de la session. Sert de compte rendu à partager, sans avoir à exporter chaque graphique séparément.

## 6.6. Info Points

Lorsque vous cliquez sur un point dans la scène, un certain nombre de données s'affichent dans le panneau latéral droit. Tout en bas de ce panneau se trouve un bouton pour télécharger ces données.