

Manuel d'Utilisation et Rapport Technique

Système d'Analyse de Données Énergétiques Hétérogènes

Version 1.2

6 février 2026

Environnement : MATLAB R2020a ou supérieur

Table des matières

1	Introduction	3
1.1	Objectif du Logiciel	3
1.2	Périmètre Fonctionnel	3
2	Architecture Technique	3
2.1	Vue d'Ensemble du Système	3
2.2	Structure de la Base de Données	4
2.3	Le Système de Cache	4
3	Guide d'Utilisation	4
3.1	Installation et Préparation	4
3.1.1	Prérequis Système	4
3.1.2	Installation des Fichiers	5
3.1.3	Conventions de Nommage des Fichiers Excel	5
3.2	Étape 1 : Importation des Données	5
3.3	Étape 2 : Exploration Interactive	6
3.3.1	Navigation par Menu	6
3.4	Étape 3 : Génération de Graphiques	6
3.4.1	Sélection des Entités	6
3.4.2	Sélection de la Période	6
3.4.3	Affichage Automatique	7
4	Fonctionnalités d'Analyse Avancées	7
4.1	Vue d'Ensemble des Commandes	7
4.2	Analyses Graphiques	7
4.2.1	Régression Linéaire (<code>reg</code>)	7
4.2.2	Prédiction (<code>pred</code>)	8
4.2.3	Lissage (<code>smooth</code>)	8
4.2.4	Ajustement Polynomiale (<code>poly2</code>)	8
4.2.5	Normalisation Base 100 (<code>base100</code>)	8
4.3	Analyses Calculatoires	8
4.3.1	Dérivation (<code>derviv</code>)	8
4.3.2	Intégration Cumulée (<code>cumul</code>)	9
4.3.3	Taux de Croissance Annuel Moyen (<code>cagr</code>)	9
4.4	Analyses Statistiques Avancées	9
4.4.1	Volatilité / Risque (<code>vol</code>)	9
4.4.2	Détection d'Anomalies (<code>anom</code>)	10
4.5	Commandes Système	10
4.5.1	Export Excel (<code>export</code>)	10
4.5.2	Listage (<code>list</code>)	10

5	Gestion des Erreurs et Dépannage	10
5.1	Problèmes Courants	10
5.1.1	Graphique Vide ou Axes Incorrects	10
5.1.2	Détection Incorrecte de Structure (Wide/Long)	11
5.1.3	Unité Manquante ou Incorrecte	11
5.2	Messages d'Erreur Spécifiques	11
5.3	Validation des Résultats	11
6	Exemples d'Utilisation Pratiques	12
6.1	Cas d'Usage 1 : Comparaison Multi-Pays	12
6.2	Cas d'Usage 2 : Analyse de Volatilité	12
6.3	Cas d'Usage 3 : Prédiction Énergétique	13
7	Annexes	13
7.1	Annexe A : Métriques Statistiques Affichées	13
7.2	Annexe B : Commandes MATLAB Utiles	13
8	Glossaire	14

1 Introduction

1.1 Objectif du Logiciel

Ce logiciel, développé en MATLAB, sert à résoudre un problème critique rencontré par les analystes de données énergétiques : l'exploitation de rapports Excel (type *BP Statistical Review of World Energy*) dont la structure est **hétérogène, imprévisible et non standardisée**.

Ces fichiers présentent typiquement :

- Des entêtes décalés verticalement (ligne 2, 3, voire 6)
- Des unités de mesure dispersées (cellule A1, A3, ou ligne précédente)
- Des formats mixtes : données *Wide* (pays en lignes, années en colonnes) ou *Long* (années en lignes, séries en colonnes)
- Des cellules fusionnées, textes parasites, et incohérences typographiques

1.2 Périmètre Fonctionnel

Le logiciel offre trois capacités principales :

- 1. ETL Robuste** : Extraction, transformation et chargement automatisés avec détection intelligente de structure
- 2. Interface Interactive** : Navigation par menu avec recherche contextuelle
- 3. Analyse Visuelle** : Génération de graphiques avec outils statistiques (régression, prédiction, dérivation)

Public Cible

Ce logiciel s'adresse aux chercheurs, ingénieurs et analystes manipulant des séries temporelles énergétiques issues de sources documentaires (rapports Excel, bases de données statistiques).

2 Architecture Technique

2.1 Vue d'Ensemble du Système

Le logiciel repose sur une architecture modulaire en trois couches :

Couche	Description
Backend	Pipeline d'importation avec parsing brut et normalisation
Storage	Base de données persistante structurée en <code>containers.Map</code>
Frontend	Interface utilisateur interactive avec moteur de visualisation

2.2 Structure de la Base de Données

Les données sont stockées dans un fichier `Base_Complete_Energy.mat` selon une hiérarchie à trois niveaux :

```

1 BaseDeDonnees : Map
2   "2023" : Map
3     "Oil Production" : Table (100 45 )
4     "Coal Consumption" : Table (75 50 )
5     ...
6   "2024" : Map
7     ...

```

Listing 1 – Structure logique de la base de données

Avantages de cette architecture :

- Accès en $O(1)$ par année et feuille (pas de parsing répétitif)
- Support natif de structures hétérogènes (chaque table garde ses métadonnées)
- Persistance immédiate via `save` MATLAB

2.3 Le Système de Cache

Optimisation Critique

Chaque fichier traité est sauvegardé en `Cache_2023.mat`. Au prochain lancement, si le cache existe, le chargement prend **0.3 secondes au lieu de 45 secondes**.

```

1 fichierCache = fullfile(dossierExcel, ['Cache_', anneeReport '.mat',
2   ]);
3
4 if exist(fichierCache, 'file')
5   fprintf('[CACHE] Chargement rapide...\n');
6   charge = load(fichierCache);
7   DonneesAnnee = charge.DonneesAnnee;
8 else
9   DonneesAnnee = TraiterUnFichier(cheminFichier, fid);
10  save(fichierCache, 'DonneesAnnee');
end

```

Listing 2 – Logique de cache dans Importation.m

Pour regénérer les données : Supprimez manuellement les fichiers `Cache_*.mat`.

3 Guide d’Utilisation

3.1 Installation et Préparation

3.1.1 Prérequis Système

- MATLAB R2020a ou supérieur

- Toolboxes : *Statistics and Machine Learning* (optionnel pour `movmean`)
- Système d'exploitation : Windows, macOS ou Linux

3.1.2 Installation des Fichiers

1. Créez un dossier de travail (ex : C:\EnergyProject)
2. Placez tous les fichiers .m dans ce dossier :
 - Importation.m
 - Explorateur.m
 - boucle_analyse_pays.m
 - tracer_tableau.m
 - analyse_statistique.m
 - extraire_colonnes_temps.m
 - demander_choix.m
3. Créez un sous-dossier pour vos fichiers Excel (ou placez-les directement)

3.1.3 Conventions de Nommage des Fichiers Excel

RÈGLE CRITIQUE

Les fichiers Excel **doivent** contenir une année à 4 chiffres dans leur nom :

- BP_Statistical_Review_2023.xlsx
- Energy_Data_2024.xls
- Rapport_Annuel.xlsx (pas d'année détectable)

Cette année sert de clé dans la base de données. Si aucune année n'est détectée, le fichier sera indexé comme **Inconnu_1**.

3.2 Étape 1 : Importation des Données

1. Ouvrez MATLAB et naviguez vers votre dossier de travail :

```
1 cd 'C:\EnergyProject'
```

2. Lancez l'importation :

```
1 Importation()
```

3. Le processus affiche :

```
Début du traitement de 3 fichiers...
[TRAITEMENT] Lecture de BP_2023.xlsx...
[CACHE] Chargement rapide de BP_2024.xlsx...
Traitement terminé. Consultez Rapport_Importation.txt
```

Fichiers générés :

- Base_Complete_Energy.mat : Base de données complète
- Cache_*.mat : Fichiers de cache par année
- Rapport_Importation.txt : Journal d'erreurs

3.3 Étape 2 : Exploration Interactive

Lancez l'interface :

```
1 Explorateur()
```

3.3.1 Navigation par Menu

Menu Principal : Sélection de l'Année

```
--- MENU PRINCIPAL ---
>> Choix Année du rapport (ou 'q'=quitter, 'r'=retour): 2023
```

Menu Secondaire : Sélection de la Feuille Le système propose une recherche textuelle intelligente :

```
--- RAPPORT 2023 ---
>> Recherche Feuille de données (ex: "Solar") ou 'r'=retour: oil
```

Si plusieurs correspondances existent :

Résultats multiples :

1. Oil Production
2. Oil Consumption
3. Oil Prices

```
>> Numéro: 1
```

3.4 Étape 3 : Génération de Graphiques

3.4.1 Sélection des Entités

Une fois la feuille chargée, entrez les pays/entités à tracer (séparés par des virgules) :

```
[Analyse: Oil Production | Unité: Million tonnes]
>> Votre choix : France, Germany, United Kingdom
```

Correspondance partielle activée : Taper Unit trouvera United States et United Kingdom.

3.4.2 Sélection de la Période

```
Plage: 1965 - 2023
>> Année Début (Entrée=Min): 2000
>> Année Fin (Entrée=Max): 2023
```

Appuyez simplement sur Entrée pour utiliser toute la plage disponible.

3.4.3 Affichage Automatique

Le logiciel génère automatiquement :

1. Un tableau de statistiques descriptives (moyenne, médiane, écart-type)
2. Un graphique temporel (ligne) ou catégoriel (barre) selon le contexte
3. Une légende automatique (si moins de 20 séries)

4 Fonctionnalités d'Analyse Avancées

4.1 Vue d'Ensemble des Commandes

Après avoir généré un graphique, vous pouvez appliquer des analyses directement en tapant une commande :

Commande	Catégorie	Description
<code>reg</code>	Graphique	Affiche la tendance linéaire (régression)
<code>pred</code>	Graphique	Prédiction à +5 ans (extrapolation)
<code>smooth</code>	Graphique	Lissage par moyenne mobile (5 ans)
<code>poly2</code>	Graphique	Ajustement polynomial quadratique
<code>dériv</code>	Calcul	Calcule la dérivée (vitesse de variation)
<code>cumul</code>	Calcul	Intégrale cumulée (total sur la période)
<code>export</code>	Système	Exporte les données vers Excel
<code>list</code>	Système	Liste toutes les entités disponibles
<code>r</code>	Système	Retour au menu précédent

4.2 Analyses Graphiques

4.2.1 Régression Linéaire (`reg`)

Objectif : Identifier la tendance de croissance/décroissance moyenne.

Formule mathématique :

$$y = ax + b \quad \text{où } a = \text{pente (croissance annuelle)} \quad (1)$$

Interprétation Console :

> France : Croissance moyenne = 2.34 unités/an

Une pente positive ($a > 0$) indique une augmentation dans le temps.

4.2.2 Prédition (pred)

Extrapole la tendance linéaire sur les 5 années suivantes.

Limitation

Cette prédition suppose que la tendance historique se maintient. Elle est peu fiable pour des séries volatiles ou avec des ruptures structurelles.

Sortie Console :

```
> Germany : Prédiction à +5 ans = 145.67
```

4.2.3 Lissage (smooth)

Applique une moyenne mobile centrée sur 5 ans pour réduire le bruit.

Formule :

$$y_{\text{lissé}}(t) = \frac{1}{5} \sum_{i=-2}^{+2} y(t+i) \quad (2)$$

Utile pour identifier les tendances à long terme en supprimant les fluctuations annuelles.

4.2.4 Ajustement Polynomial (poly2)

Ajuste un polynôme de degré 2 : $y = ax^2 + bx + c$

Permet de capturer des dynamiques d'accélération ou de décélération (courbes paraboliques).

4.2.5 Normalisation Base 100 (base100)

Objectif : Comparer la **dynamique relative** de plusieurs pays ayant des ordres de grandeur très différents (ex : Chine vs Belgique).

Formule : Ramène la valeur de la première année (t_0) à 100 pour chaque série :

$$y_{\text{indexé}}(t) = \frac{y(t)}{y(t_0)} \times 100 \quad (3)$$

Interprétation Visuelle : Toutes les courbes partent du même point ($y = 100$).

- Courbe > 100 : Croissance par rapport à l'année de départ.
- Courbe < 100 : Déclin.

4.3 Analyses Calculatoires

4.3.1 Dérivation (deriv)

Calcule la **vitesse de variation** par différences finies :

$$\frac{dy}{dx} \approx \frac{y_{t+1} - y_t}{x_{t+1} - x_t} \quad (4)$$

Cas d'usage :

- Identifier les années de croissance maximale
- Déetecter les ruptures (changements brusques de politique)

Sortie Console :

> China : Pic de variation = 12.5 en 2007

4.3.2 Intégration Cumulée (cumul)

Calcule le **total cumulé** via la méthode des trapèzes :

$$\text{Cumul} = \sum_{i=1}^{n-1} \frac{y_i + y_{i+1}}{2} \times (x_{i+1} - x_i) \quad (5)$$

Cas d'usage :

- Production totale d'énergie sur une décennie
- Émissions cumulées de CO₂

4.3.3 Taux de Croissance Annuel Moyen (cagr)

Calcule le taux de croissance constant théorique qui permettrait de passer de la valeur initiale à la valeur finale (Lissage géométrique).

Formule :

$$\text{CAGR} = \left(\frac{y_{\text{fin}}}{y_{\text{début}}} \right)^{\frac{1}{n}} - 1 \quad (6)$$

Contrairement à la régression linéaire (**reg**) qui est additive, le CAGR est multiplicatif. Il est standard pour les analyses financières et énergétiques.

Sortie Console :

> India : CAGR = 5.42%

4.4 Analyses Statistiques Avancées

4.4.1 Volatilité / Risque (vol)

Mesure l'**instabilité** de la série temporelle en calculant l'écart-type mobile sur une fenêtre glissante (5 ans).

Formule :

$$\sigma_t = \sqrt{\frac{1}{N-1} \sum_{i=0}^{N-1} |y_{t-i} - \mu|^2} \quad (7)$$

Cas d'usage :

- Évaluer la sécurité énergétique (une production stable est moins risquée).
- Analyser la volatilité des prix (Spot vs Long Terme).

Sortie Console :

> Crude Oil : Volatilité Moyenne = 14.20

4.4.2 Détection d'Anomalies (anom)

Utilise un algorithme statistique pour identifier les points aberrants (*outliers*) qui s'écartent significativement de la médiane locale.

Méthode : Détection basée sur l'écart absolu médian (MAD) :

$$|y_t - \text{MoyenneMobile}| > 3 \times \sigma_{\text{local}} \quad (8)$$

Utilité

Permet de repérer automatiquement les **chocs exogènes** (guerres, crises économiques, erreurs de données) sans inspection manuelle.

Sortie Console :

> Venezuela : 3 anomalies détectées

4.5 Commandes Système

4.5.1 Export Excel (export)

Exporte le dernier sous-tableau tracé vers `Resultats_Export.xlsx`.

Format de Sortie

Le fichier Excel contient :

- Colonne 1 : Noms des entités
- Colonnes 2..N : Années sélectionnées
- Métadonnées (unité) dans les propriétés du fichier

4.5.2 Listage (list)

Affiche toutes les entités disponibles dans la feuille actuelle. Utile pour connaître l'orthographe exacte des pays.

5 Gestion des Erreurs et Dépannage

5.1 Problèmes Courants

5.1.1 Graphique Vide ou Axes Incorrects

Symptôme : Le graphique s'affiche mais sans données, ou les années sur l'axe X sont erronées.

Causes Possibles :

1. **Cache Corrompu** : Un fichier `Cache_*.mat` contient une version obsolète.
2. **Regex Années** : Les années ne sont pas détectées (ex : format "Year 2023" au lieu de "2023").

Solutions :

1. Supprimez tous les fichiers Cache_*.mat
2. Relancez Importation()
3. Vérifiez dans Rapport_Importation.txt les messages [SKIP]

5.1.2 Détection Incorrecte de Structure (Wide/Long)

Symptôme : Le logiciel effectue un "Auto-Pivot" alors que les données sont déjà au bon format.

Diagnostic : Ouvrez le fichier Rapport_Importation.txt et cherchez la ligne :

Structure détectée : Long

Solution : Modifiez le seuil de détection dans Importation.m (ligne 138) :

```
1 if countYearsHeader >= 3 % Au lieu de >= 2
2     structureType = 'Wide';
```

5.1.3 Unité Manquante ou Incorrecte

Symptôme : Le graphique affiche "Unité : Unknown" ou une valeur erronée.

Cause : L'algorithme de scan ascendant n'a pas trouvé de texte au-dessus de l'entête.

Solution Temporaire : Ajoutez manuellement l'unité après chargement :

```
1 Tableau.Properties(userData.Unit = 'Million_barrels_per_day');
```

5.2 Messages d'Erreur Spécifiques

Message	Signification
Aucun fichier Excel trouvé	Le dossier spécifié ne contient pas de fichiers .xls ou .xlsx
Structure illisible	La feuille ne contient aucune ligne dense détectable (probablement vide ou mal formattée)
Pas de données temporelles	Aucune colonne d'année n'a été détectée (vérifiez le format des entêtes)
Hors plage	Les années saisies sont en dehors de la période disponible

5.3 Validation des Résultats

Pour vérifier l'intégrité des données importées :

1. Chargez manuellement la base :

```
1 load('Base_Complete_Energy.mat');
2 keys(BaseDeDonnees) % Liste les années
```

2. Inspectez une table spécifique :

```

1 T = BaseDeDonnees('2023')('OilProduction');
2 head(T) % Affiche les 8 premières lignes
3 T.Properties.UserData % Affiche les métadonnées

```

3. Vérifiez les colonnes d'années :

```

1 [annees, ~] = extraire_colonnes_temps(T);
2 disp(annees) % Doit afficher [1965, 1966, ..., 2023]

```

6 Exemples d'Utilisation Pratiques

6.1 Cas d'Usage 1 : Comparaison Multi-Pays

Objectif : Comparer la production pétrolière de 5 pays sur 20 ans avec tendances.

```

1 >> Explorateur()
2 >> Choix Année : 2023
3 >> Recherche Feuille : oil production
4 >> Votre choix : USA, Russia, Saudi Arabia, China, Canada
5 >> Année But : 2000
6 >> Année Fin : 2023
7
8 % Graphique s'affiche avec 5 courbes
9
10 >> reg % Ajoute les tendances linaires
11 > USA : Croissance moyenne = -0.15 Mt/an
12 > Russia : Croissance moyenne = 0.45 Mt/an
13 > Saudi Arabia : Croissance moyenne = 0.12 Mt/an
14
15 >> export % Sauvegarde dans Excel

```

Listing 3 – Session complète

6.2 Cas d'Usage 2 : Analyse de Volatilité

Objectif : Identifier les périodes de forte variation dans les prix du pétrole.

```

1 >> Recherche Feuille : oil prices
2 >> Votre choix : Brent, WTI
3 >> Année But : 1980
4 >> Année Fin : 2023
5
6 >> deriv % Calcule la dérivée
7 > Brent : Pic de variation = 45.2 $/an en 2008
8 > WTI : Pic de variation = 42.8 $/an en 2008
9
10 % Interprétation : Crise financière 2008

```

6.3 Cas d'Usage 3 : Prédiction Énergétique

Objectif : Estimer la consommation d'électricité en 2030.

```

1 >> Recherche Feuille : electricity consumption
2 >> Votre choix : World
3 >> Ann e D but : 1990
4 >> Ann e Fin : 2023
5
6 >> smooth % Lisse les donn es
7 >> pred % Extrapolation +5 ans
8 > World : Pr diction +5 ans = 28,450 TWh (2028)
9
10 % Note : Pour 2030, relancer avec pr diction +7 ans
11 % ou ajuster manuellement dans analyse_statistique.m

```

7 Annexes

7.1 Annexe A : Métriques Statistiques Affichées

Pour chaque graphique, le logiciel calcule automatiquement :

Métrique	Formule MATLAB
Moyenne	mean(data, 'omitnan')
Médiane	median(data, 'omitnan')
Écart-type	std(data, 'omitnan')
Minimum	min(data)
Maximum	max(data)

7.2 Annexe B : Commandes MATLAB Utiles

```

1 % Lister toutes les ann es disponibles
2 load('Base_Complete_Energy.mat');
3 disp(keys(BaseDeDonnees))

4
5 % Inspecter une feuille sp cifique
6 T = BaseDeDonnees('2023')('Oil_Production');
7 summary(T)

8
9 % V rifier les m tadonn es
10 disp(T.Properties.UserData)

11
12 % Forcer la reg n ration du cache
13 delete('Cache_*.mat')
14 Importation()

15
16 % Exporter une table manuellement
17 writetable(T, 'export_manuel.xlsx')

```

Listing 4 – Commandes de débogage avancé

8 Glossaire

ETL Extract, Transform, Load - Processus de traitement de données

Wide Format Structure où les années sont en colonnes (lignes = entités)

Long Format Structure où les années sont en lignes (colonnes = séries)

Auto-Pivot Transformation automatique Long → Wide

Cache Fichier .mat contenant des données prétraitées

Metadata Informations annexes (unité, nom de feuille, structure)

Heuristique Algorithme basé sur des règles pratiques (non mathématiquement optimal)