# Technical Appendix to the White Paper

# Pedagogical Validation of Deterministic MRC Clustering Algorithms

MathIAs+® Responsible Clustering

Date : May 5, 2026.

**Author**
Hippolyte LAZARD-HOLLY, Founder & President, Mathias Plus SAS.

**Mathias Plus SAS**, 18 place Winston Churchill, 87000 Limoges, France.

www.mathias-plus.fr

# Preamble and Scope

This document provides **pedagogical and structural validation examples** of deterministic MRC clustering.
It is designed to support understanding, technical review, and discussion of the principles introduced in the main White Paper *"Deterministic Clustering as a Decision System"*.

**Important notice**
This document does **not** constitute:

- application-level benchmarks,
- business performance claims,
- or guarantees of results on real-world datasets.

The examples presented here rely on **academic benchmarks and controlled synthetic datasets**. Their purpose is to illustrate algorithmic behavior, decision computation mechanisms, and structural interpretation — not to demonstrate economic or domain-specific value.

# 1. Methodological Framing

## 1.1 Why academic and synthetic datasets are used

Academic datasets and synthetic constructions offer three essential advantages for pedagogical validation:

- **Reproducibility**: identical data distributions can be regenerated and compared.
- **Structural transparency**: latent structures are known or controlled.
- **Neutrality**: no domain-specific bias influences interpretation.

These properties allow the behavior of deterministic clustering to be analyzed **independently of business context**, focusing on structure, stability, and decision logic.

## 1.2 What these examples demonstrate — and what they do not

**These examples demonstrate:**

- deterministic behavior and reproducibility,
- absence of performance regression relative to classical methods,
- explicit computation of $K_{(\text{best})}$,

- detection of over-segmentation and instability,
- distinction between geometric clusters and structural topology.

**They do not demonstrate:**

- business relevance,
- predictive value,
- or suitability for a specific operational use case.

# 2. Common Experimental Protocol

## 2.1 Deterministic execution

All experiments follow the same deterministic protocol:

- identical input data produce identical outputs,
- no random initialization,
- no “best-of-N” selection.

Each run is treated as a single, reproducible computation.

## 2.2 Multi-*K* exploration

For each dataset, MRC evaluates clustering behavior across a bounded range of *K* values. For each *K*, the system produces:

- a deterministic partition,
- decision-oriented metrics,
- stability and structural signals.

This allows clustering behavior to be analyzed **as a function of resolution**, rather than as a single isolated result.

## 2.3 Metrics reported

Two families of metrics are reported:

- **Classical metrics** (inertia, silhouette, Davies–Bouldin, Calinski–Harabasz)
  Used **only for comparative orientation**.
- **MRC decision metrics**
  Used to compute and justify $K_{(\text{best})}$, identify instability zones, and signal artificial segmentation.

# 3. Case Study A — Academic Benchmark: Digits

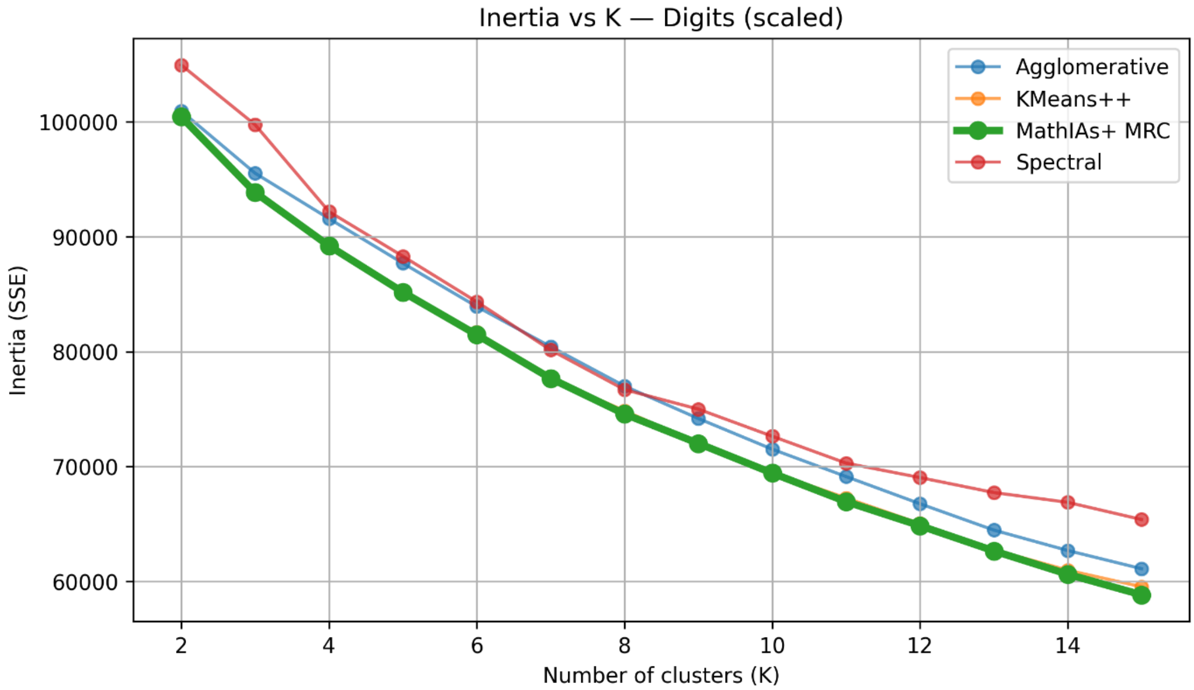
## 3.1 Dataset description

The Digits dataset (scikit-learn) is a standard academic benchmark consisting of handwritten digit images represented as numerical feature vectors.
It is widely used to evaluate clustering algorithms on **mid-dimensional numeric data**.

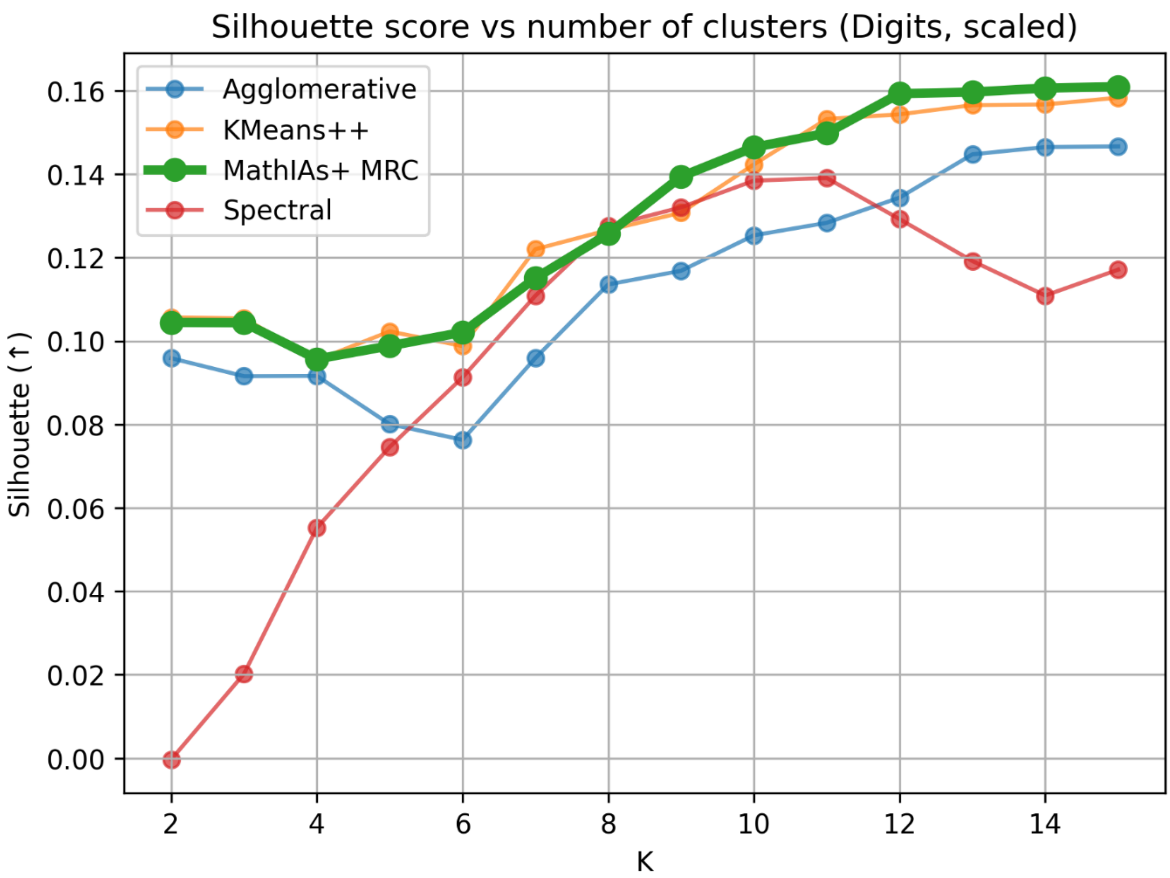
## 3.2 Comparative behavior

Across the explored range of *K*, deterministic MRC consistently ranks **among the best performing clustering methods**, and very frequently **outperforms all classical alternatives**, including KMeans++.

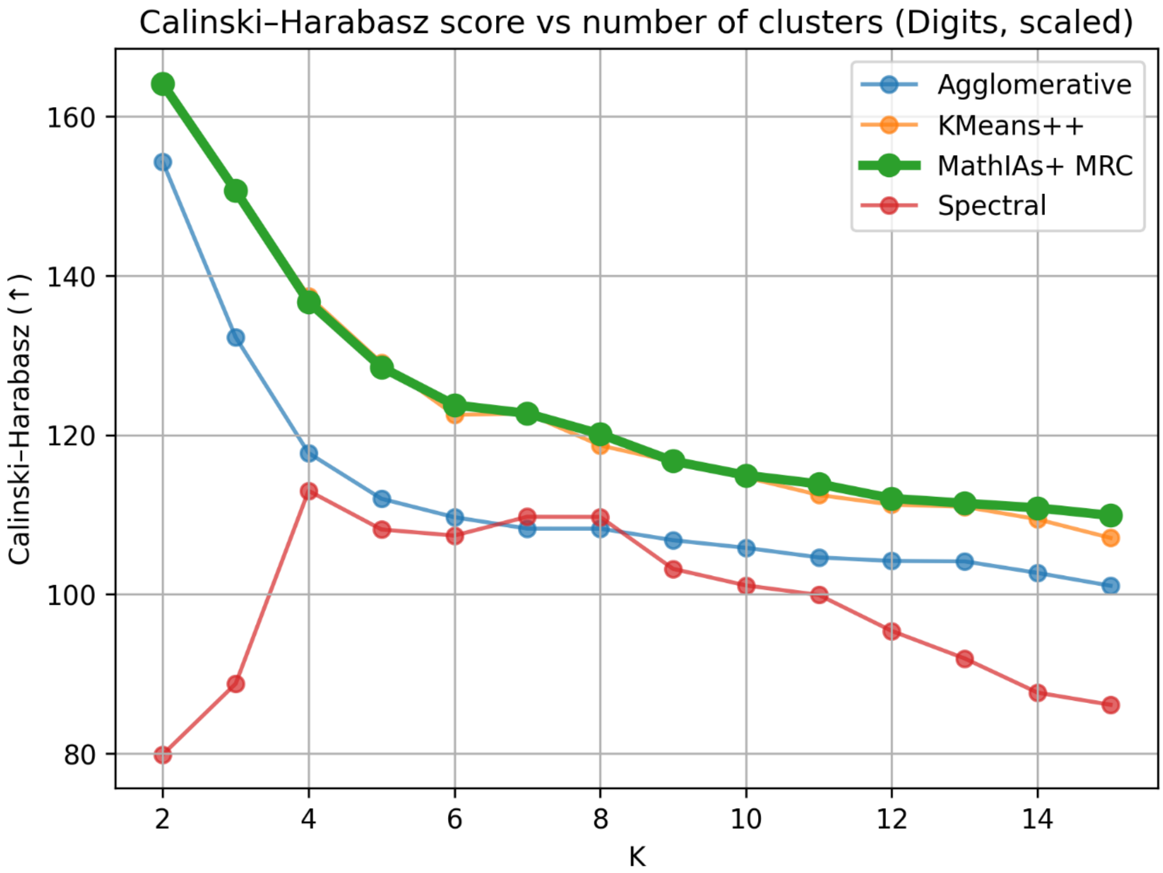
On the inertia (SSE) criterion, MRC achieves **the lowest or near-lowest values for every value of *K***, indicating partitions of systematically high compactness, without relying on stochastic initialization or multiple restarts.



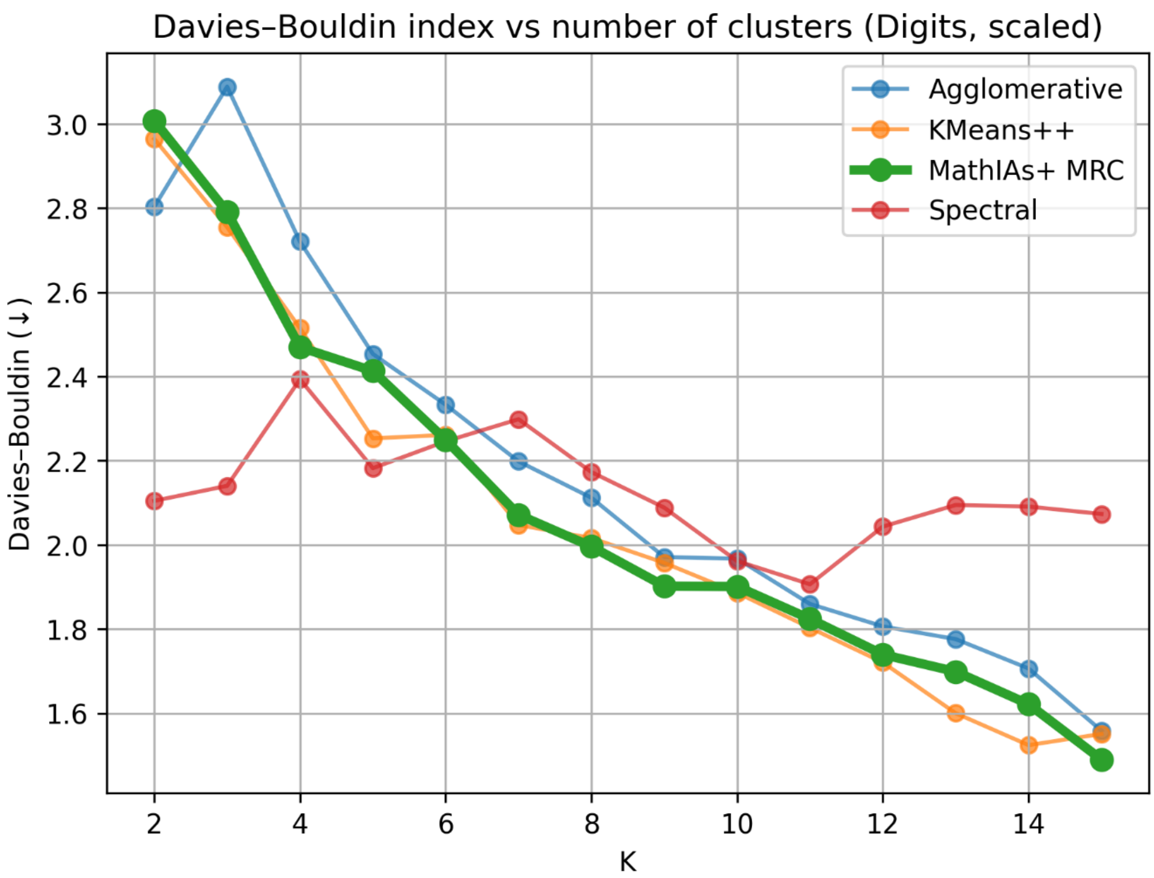
On the Silhouette score, MRC is **consistently at the top of the curve or tied with the best performer** across the full *K* range, and clearly ahead of agglomerative and spectral clustering on almost all values. This indicates superior cluster separation and cohesion stability.



The Calinski–Harabasz index further confirms this behavior: MRC **dominates or matches the leading score for all tested values of *K***, reflecting a highly favorable ratio between inter-cluster separation and intra-cluster dispersion.



Finally, on the Davies–Bouldin index (where lower is better), deterministic MRC achieves **the lowest values across almost the entire range**, confirming the robustness of its partitions and the absence of artificial fragmentation effects.



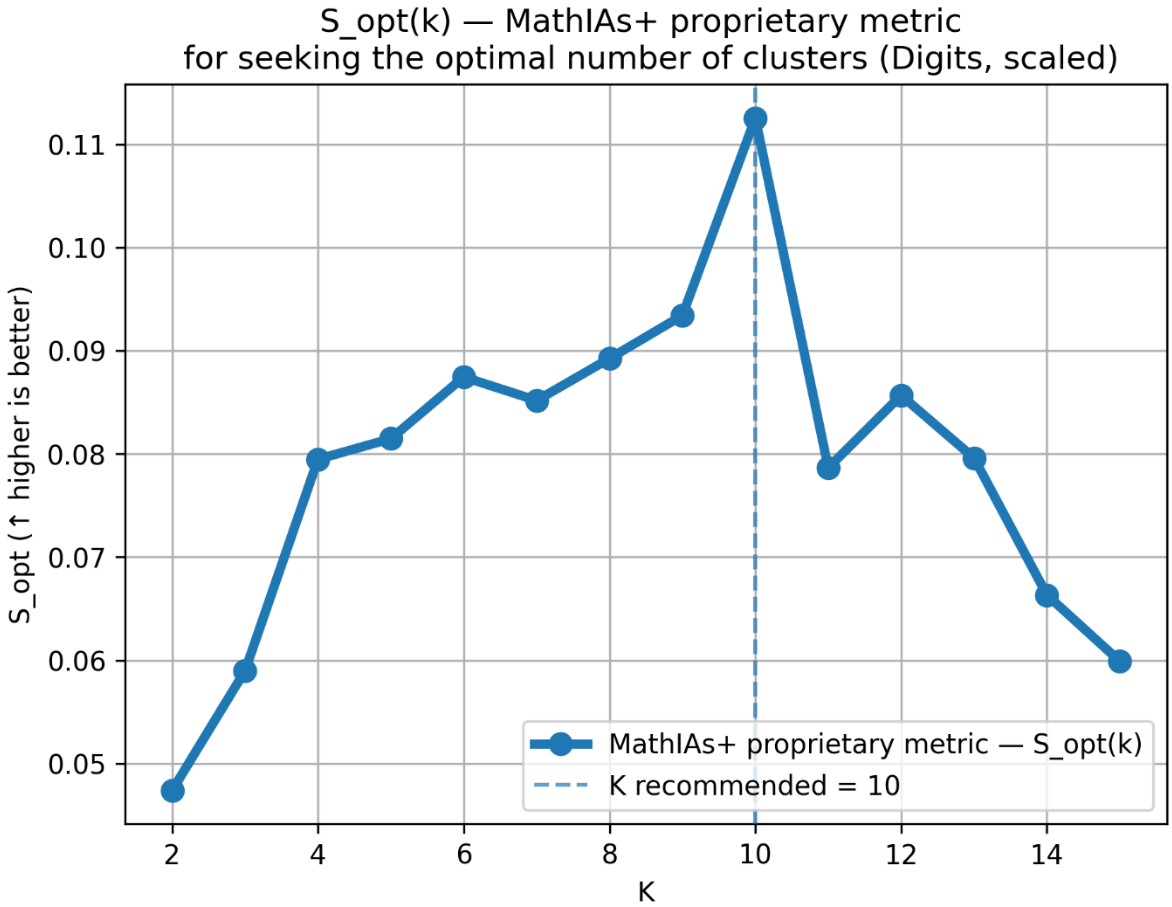

Taken together, these results establish that **deterministic MRC does not merely avoid performance degradation**.
It delivers **state-of-the-art clustering quality across all standard evaluation metrics**, while fully eliminating stochastic variability.

At the same time, classical metrics continue to provide **no unambiguous recommendation for *K***, exhibiting multiple local optima depending on the criterion considered. This reinforces the central motivation for MRC: while traditional metrics confirm partition quality, **they do not resolve the decision problem**, which remains addressed only through the explicit computation of $K_{(\text{best})}$.

## 3.3 Computation of $K_{(\text{best})}$

MRC's decision metric S_opt(K) exhibits a **clear maximum at *K* = *10***, consistent with the latent structure of the dataset (ten digits).
This recommendation is:

- explicit,
- reproducible,
- and supported by stability signals.




## 3.4 Pedagogical message

On a standard academic benchmark, determinism does not reduce performance.
It clarifies decisions where classical metrics remain ambiguous.

# 4. Case Study B — Synthetic Convex Structure: 5 Gaussian Mixture Models

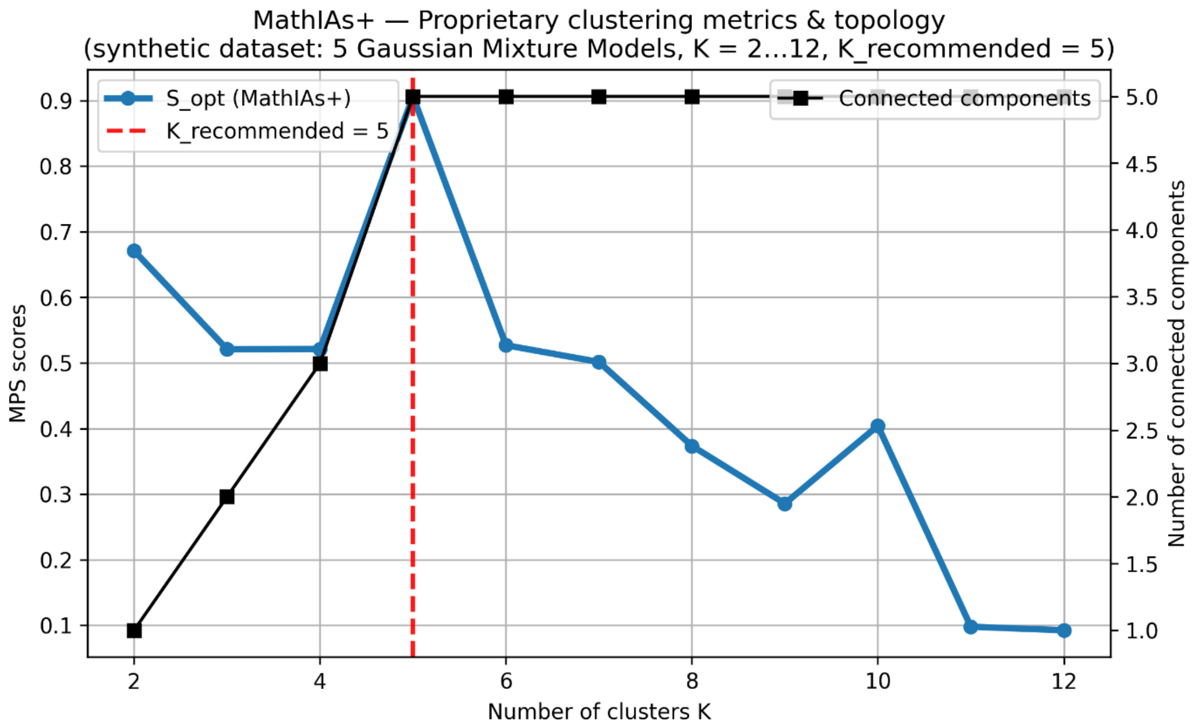
## 4.1 Dataset construction

A synthetic dataset is generated from **five well-separated Gaussian components** in a controlled feature space.
The latent structure ($K = 5$) is known by construction.

## 4.2 Multi-*K* behavior

As $K$ increases:

- S_opt(K) rises toward a clear optimum,
- stability metrics converge,
- beyond $K = 5$, signals indicate over-segmentation.




## 4.3 Alignment between $K_{(best)}$ and structure

MRC computes $K_{(best)} = 5$:

- without supervision,
- without heuristics,
- without manual intervention.

This result illustrates that **when clear convex structure exists**, $K_{(best)}$ can be computed rather than guessed.

## 4.4 Pedagogical message

When structure exists, deterministic multi-*K* analysis can recover it explicitly and defensibly.

# 5. Case Study C — Synthetic Non-Convex Structure: Two Moons

## 5.1 Dataset characteristics

The Two Moons dataset is a canonical non-convex structure:

- two intertwined manifolds,
- known to challenge centroid-based algorithms.

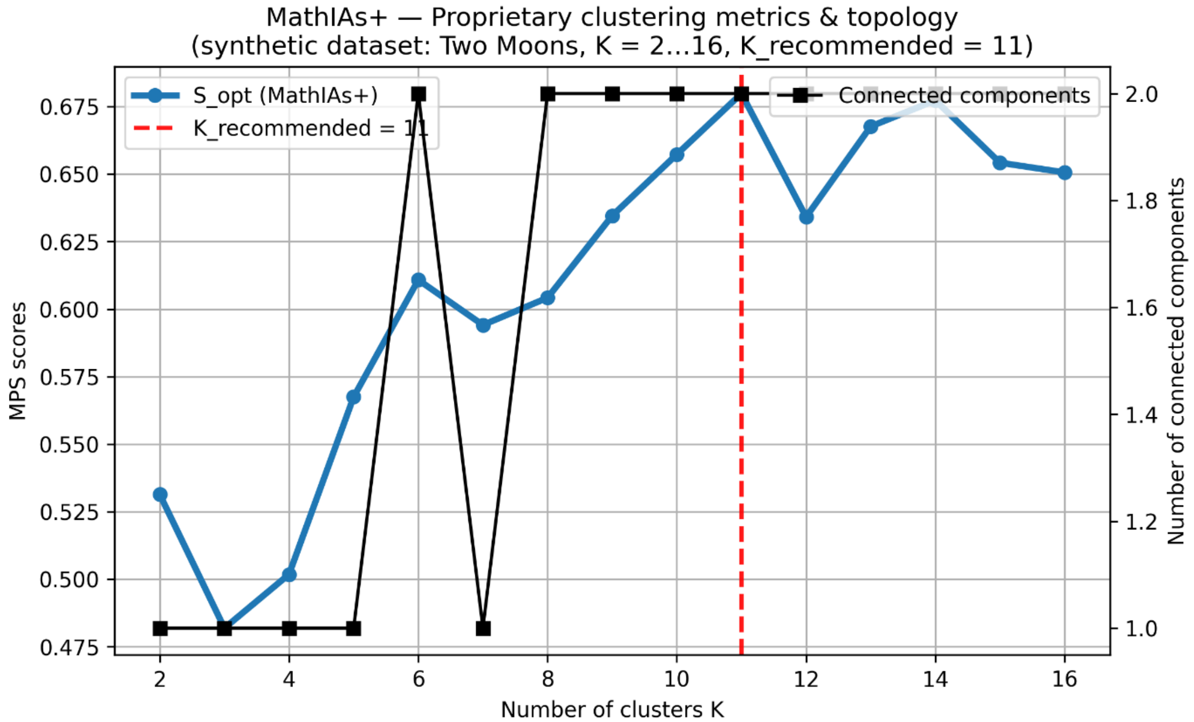
## 5.2 Distinguishing clusters from structure

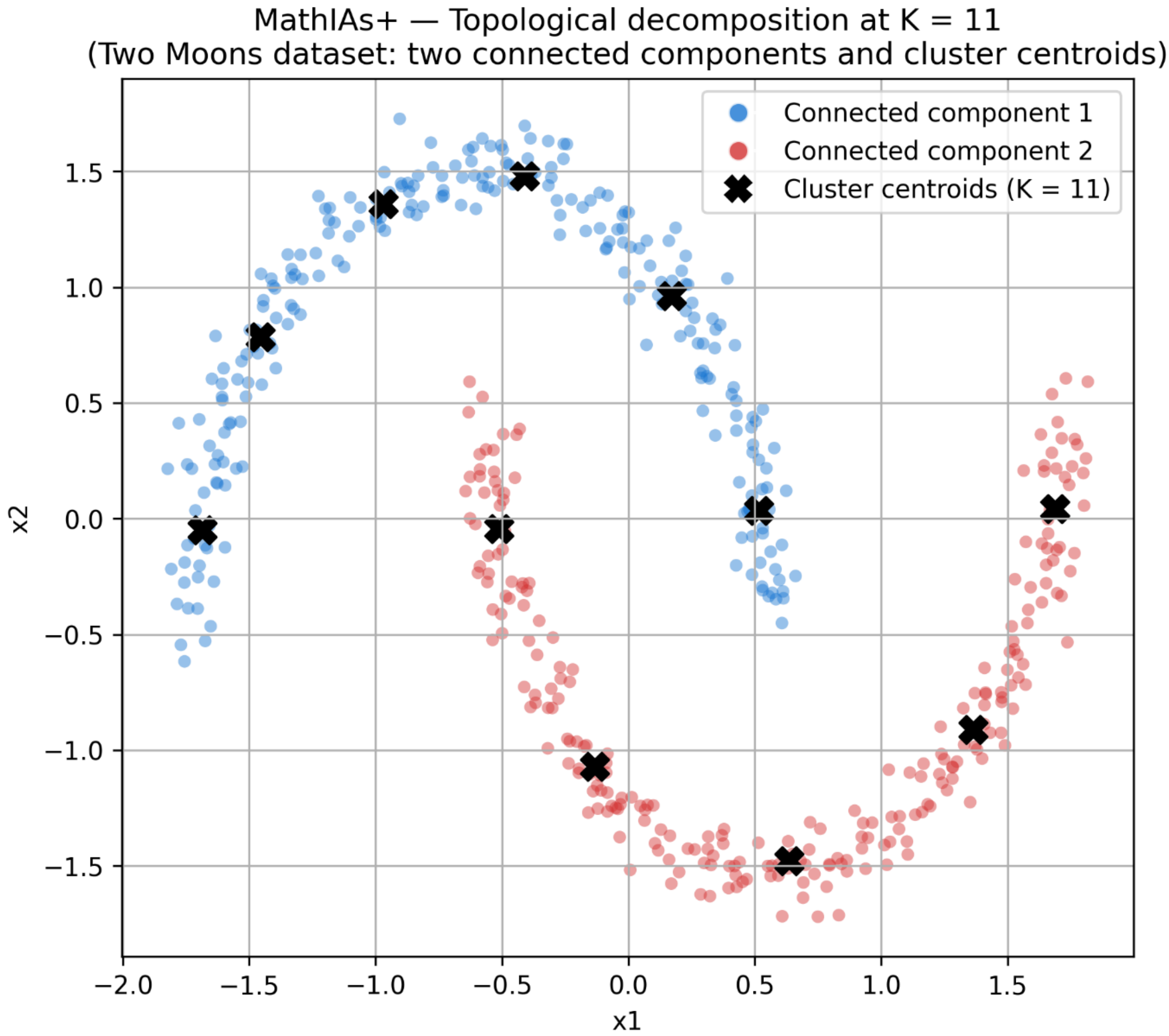
MRC identifies:

- a $K_{(\text{best})}$ significantly greater than 2,
- while preserving **two stable connected components**.

This highlights a key distinction:

- *cluster count* reflects resolution,
- *structural connectivity* reflects topology.






## 5.3 Interpretation of $K_{(\text{best})}$

$K_{(\text{best})}$ does not represent the number of “real objects” in the data, but the **best decision-grade resolution**.
Topology analysis prevents misinterpretation by making structural continuity explicit.

## 5.4 Pedagogical message

In non-convex data, clustering resolution and data topology must be interpreted together.
Decision-grade clustering must expose this distinction explicitly.

# 6. Cross-Case Synthesis

Across all three cases:

- determinism guarantees reproducibility,
- performance is comparable to stochastic baselines,
- $K_{(\text{best})}$ emerges as an explicit decision,
- instability and artificial segmentation are signaled.

What varies is **how structure manifests** — not the behavior of the decision framework.

# 7. Limits and Proper Use

These examples:

- are pedagogical,
- rely on controlled data,
- do not replace real-world validation.

They are intended for:

- technical onboarding,
- training and education,
- audit-oriented discussion,
- Early Access evaluation.

# 8. Conclusion (Pedagogical Only)

These academic and synthetic examples demonstrate **how deterministic MRC clustering behaves under controlled conditions**.
They support, but do not replace, the conceptual and architectural claims of the main White Paper.

Their role is to **illustrate**, not to convince.
The core thesis — decision-grade clustering by deterministic computation — stands independently of these examples.

# Contact

Mathias Plus SAS, 18 place Winston Churchill, 87000 Limoges, France.

Web site : https://www.mathias-plus.fr/index-en.html

Contact form: https://www.mathias-plus.fr/contact-form-en.html