Dimensionality Reduction Applied to Mechanical Design and Optimization

LMA2S Seminar, ONERA – 14/03/2025

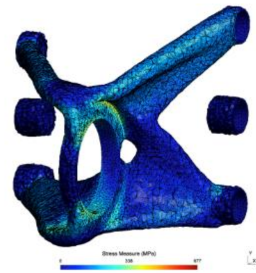
Rajan Filomeno Coelho
Contact: rajan.coelho@cenaero.be

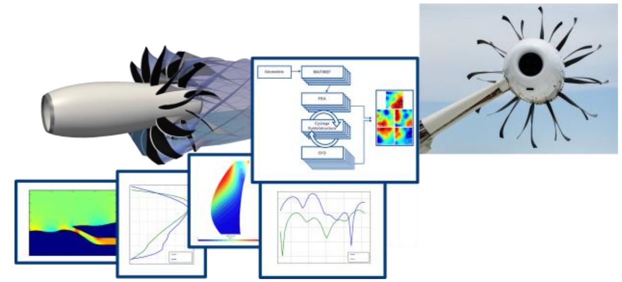
Outline

- Introduction
- Dimensionality Reduction
- Application to design optimization
- Conclusions and future prospects
- References

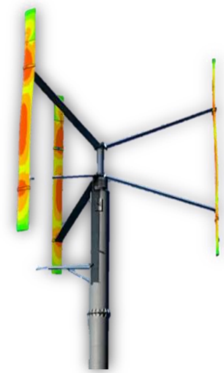
Introduction

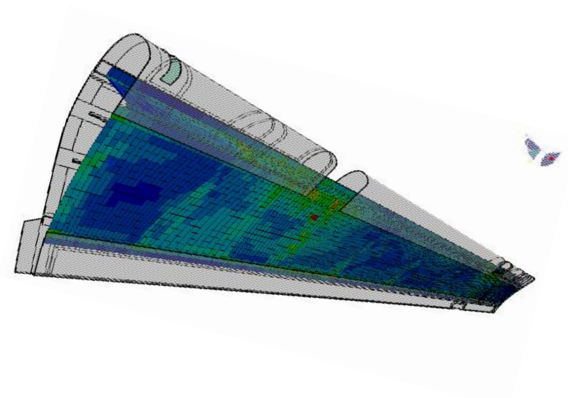
Cenaero





- Private research center, specialized in numerical simulations
- Two locations
  - Non-profit organization in Charleroi (BE)
  - Subsidiary in Moissy-Cramayel (FR)
- 80+ researchers
- Started in 2002 with the support of the European Regional Development Funds and the impulse of the Walloon aeronautical sector and universities





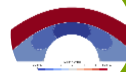
- Operational missions: collaborative research in close collaboration with industry and academia + services for companies (HPC supercomputing facilities, consultancy)

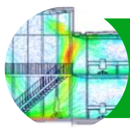
Research Themes at Cenaero

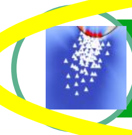

Aerostructures Design & Development

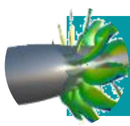

High Performance Computing

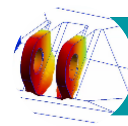

Aerostructures Engineering Environment

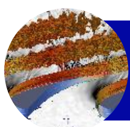

Hypersonic Flows & Phase-changing Materials

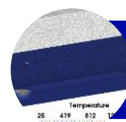

Buildings and Smart Cities

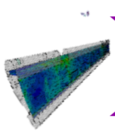

Machine Learning for Modeling, Optimization & Data Mining

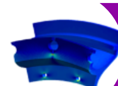

Turbomachinery Design


Manufacturability & Multidisciplinary Design for Manufacturing


High-Fidelity CFD & CAA for Aeronautical Applications

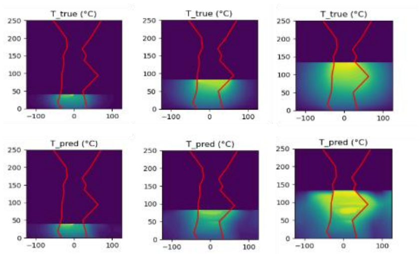

Metallic Manufacturing Processes Modeling


High Performance Composites

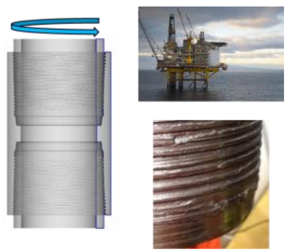

Multiscale Mechanics through Lifetime

Machine Learning & Optimization

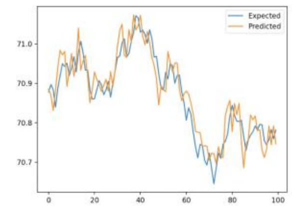
- **Machine Learning (ML) activities in various scientific and industrial projects**



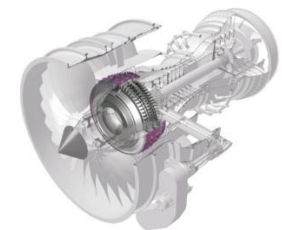
Prediction of physical fields



Detection of defects

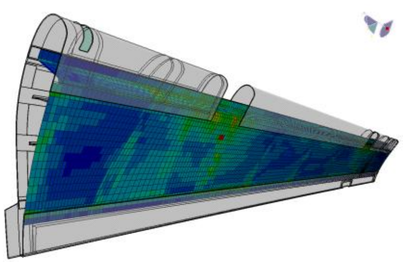


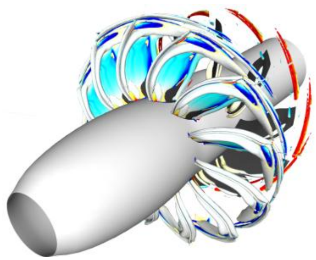
Time series prediction

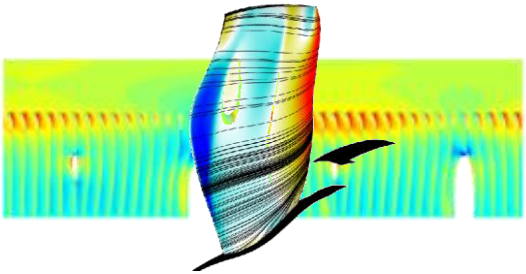


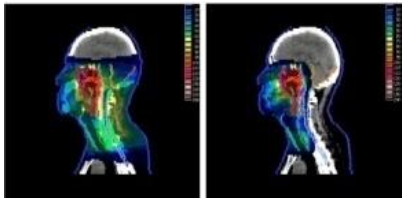
Dimensionality Reduction

- **In addition to ML activities, we also develop Minamo software for optimization and parametric studies**





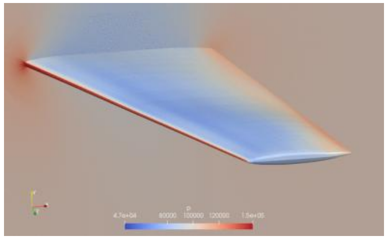


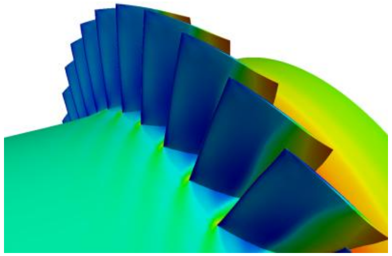


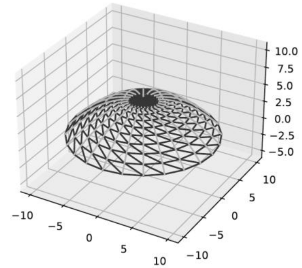
Cenaero

Dimensionality reduction in simulation-based problems

- **Dimensionality reduction in the case of Cenaero's activities:**
  - **Need to cope with high-dimensonal design spaces**
  - **Data coming from physics-based simulations**
  - **Application: engineering design and optimization (fields: aeronautics, biomedical, buildings, …)**



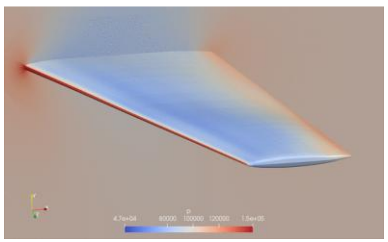


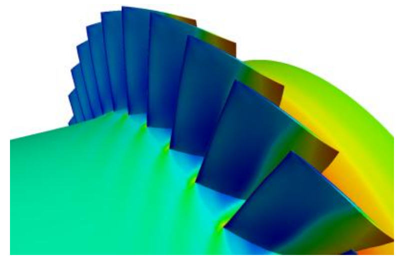


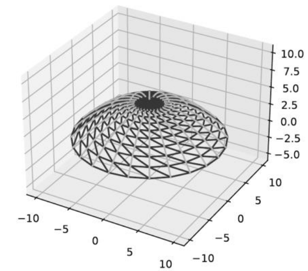
Cenaero

Dimensionality reduction in simulation-based problems

- **Dimensionality reduction in the case of Cenaero's activities:**
  - **Need to cope with high-dimensonal design spaces**
  - **Data coming from physics-based simulations**
  - **Application: engineering design and optimization (fields: aeronautics, biomedical, buildings, …)**





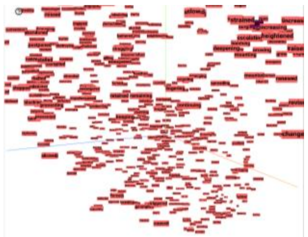


- **Context of this research:**
  - **WINGS project funded by the Walloon Region, with HPC resources from Lucia Tier 1 supercomputer (hosted by Cenaero and also funded by the Walloon Region)**
  - **Goal: develop dimensionality reduction methods for mechanical design examples characterized by highly flexible parametrizations**

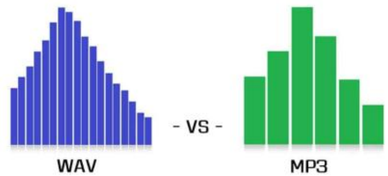
Dimensionality Reduction

Motivation for dimensionality reduction

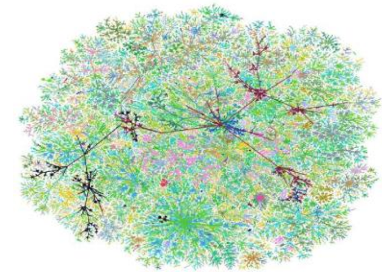
Text categorization of massive document resources



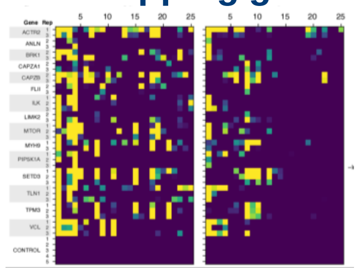
Acoustic signal compression



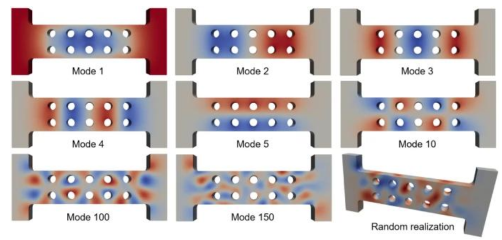
Data visualization of the internet network map



Identification of overlapping genes



Model reduction in physics-based simulations



- **Dimensionality reduction: popular in various application fields**
- **Black-box simulations ⇒ no analytical properties can be *a priori* assumed**
- **Restricted to continuous features (no integer, discrete, or categorical var.)**

Cenaero

Dimensionality reduction in design optimization

- **Three main paradigms:**

$\mathbf{x} = [\, x_1 \ldots x_N \,]^T$

Input space (e.g. CAD or FFD parameters)

Simulation

(CFD, CAA, FEA, ...)

$\mathbf{y} = [\, y_1 \ldots y_M \,]^T$

Output space (e.g. scalar responses extracted from velocity field, stresses, …)

Dimensionality reduction in design optimization

- **Three main paradigms:**

$\mathbf{x} = [\, x_1 \ldots x_N \,]^T$

Input space (e.g. CAD or FFD parameters)

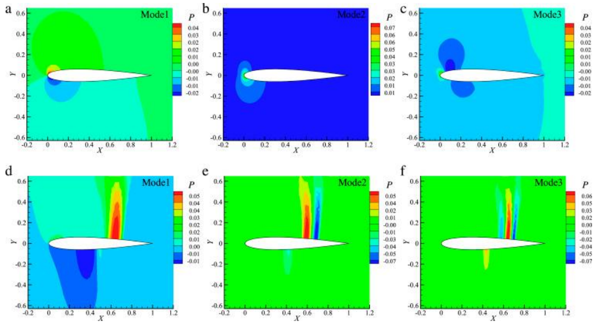
Simulation

(CFD, CAA, FEA, ...)

$\mathbf{y} = [\, y_1 \ldots y_M \,]^T$

Output space (e.g. scalar responses extracted from velocity field, stresses, …)

Model reduction on the **output space**
Ex.: POD on the CFD flow field



[Li & Zhang, 2016; Benamara *et al.*, 2017; etc.]

Dimensionality reduction in design optimization

- **Three main paradigms:**

$\mathbf{x} = [\, x_1 \,\dots\, x_N \,]^T$

Input space (e.g. CAD or FFD parameters)

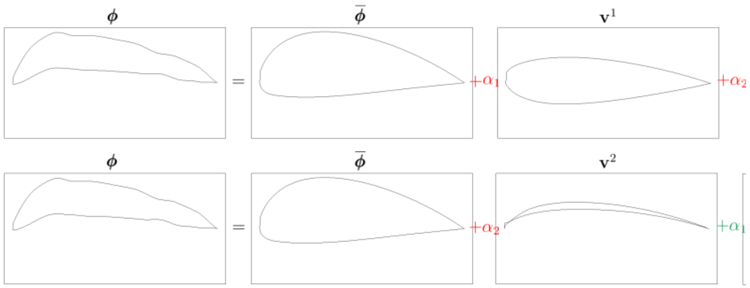
Simulation

(CFD, CAA, FEA, ...)

$\mathbf{y} = [\, y_1 \,\dots\, y_M \,]^T$

Output space (e.g. scalar responses extracted from velocity field, stresses, …)

Model reduction on the **input space**
Ex.: PCA on the airfoil geometries



[Gaudrie, Le Riche *et al.*, 2019]

Dimensionality reduction in design optimization

- ## Three main paradigms:

$\mathbf{x} = [\, x_1 \ldots x_N \,]^T$

Simulation

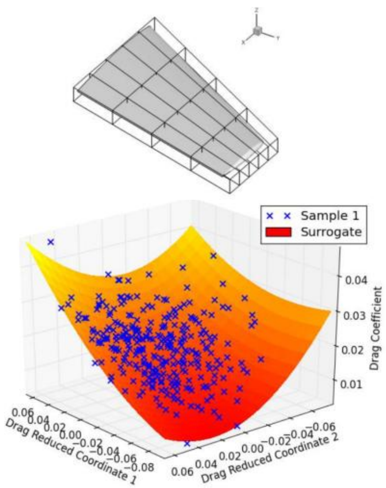
$\mathbf{y} = [\, y_1 \ldots y_M \,]^T$

Input space (e.g. CAD or FFD parameters)

(CFD, CAA, FEA, ...)

Output space (e.g. scalar responses extracted from velocity field, stresses, …)

Model reduction on the **input space**, taking the **ouputs** into account
Ex.: Active subspaces [Lucaczyk *et al.*, 2014]
⇒ Drag-reduced coordinates
⇒ Surrogate on a limited set of variables

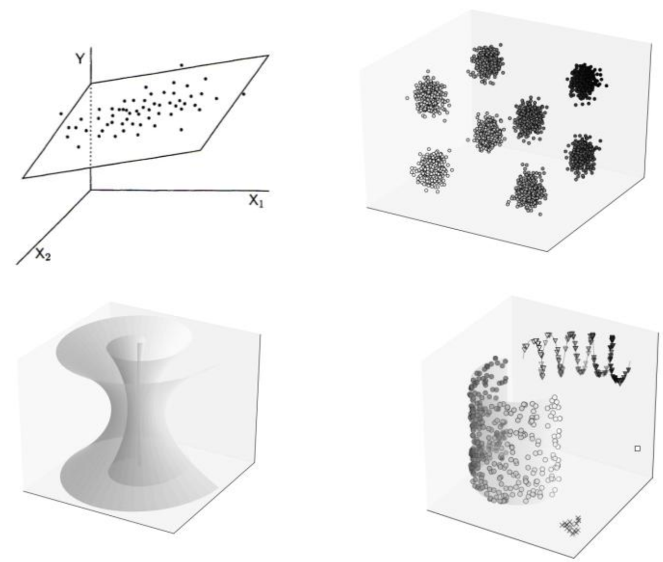


- Efficient, but often CPU expensive and not always flexible enough (need for adjoints, ...)
- **Our goal: develop an alternative *generic* and *affordable* dimensionality reduction methodology (on the input space)**

Dimensionality reduction

- **Generally speaking, data management issues involve:**
  - **Size (= number of samples)**
  - **Complexity (e.g., non-linearities, singularities, ...)**
  - **Dimensionality (= number of features/parameters/variables/inputs)**

- **To address these issues, several dimensionality reduction techniques have been proposed to:**
  - **Transform the original dataset into a new dataset representing low dimensionality...**
  - **… while maintaining as much as possible the original meanings of the data [Zebari *et al.*, 2020]**
  - **Their performances vary in the way they can capture or (not) the inner complexity of the data**

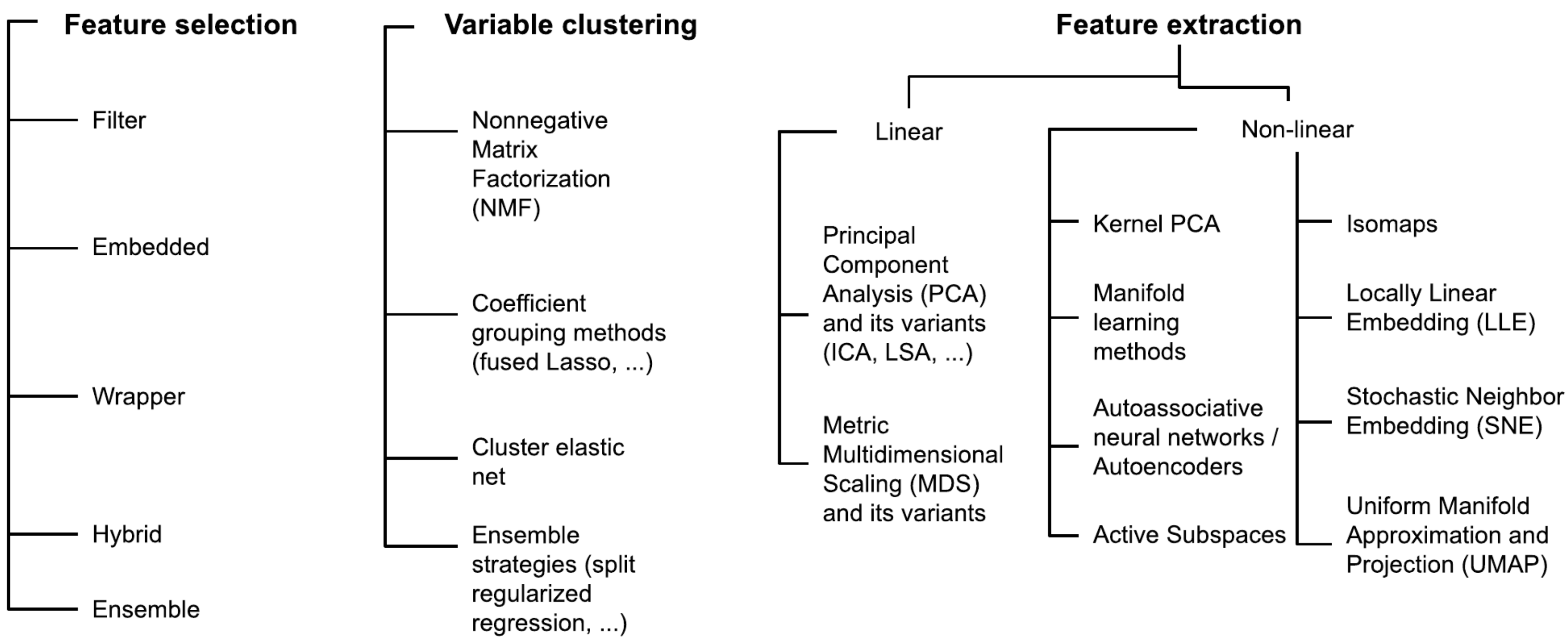


Dimensionality reduction: benefits expected

- **Benefits expected:**
  - **Reduction of the computational time and complexity of the subsequent approximation and optimization process**
  - **Avoids overfitting and noise**
  - **Better data visualization and interpretation**

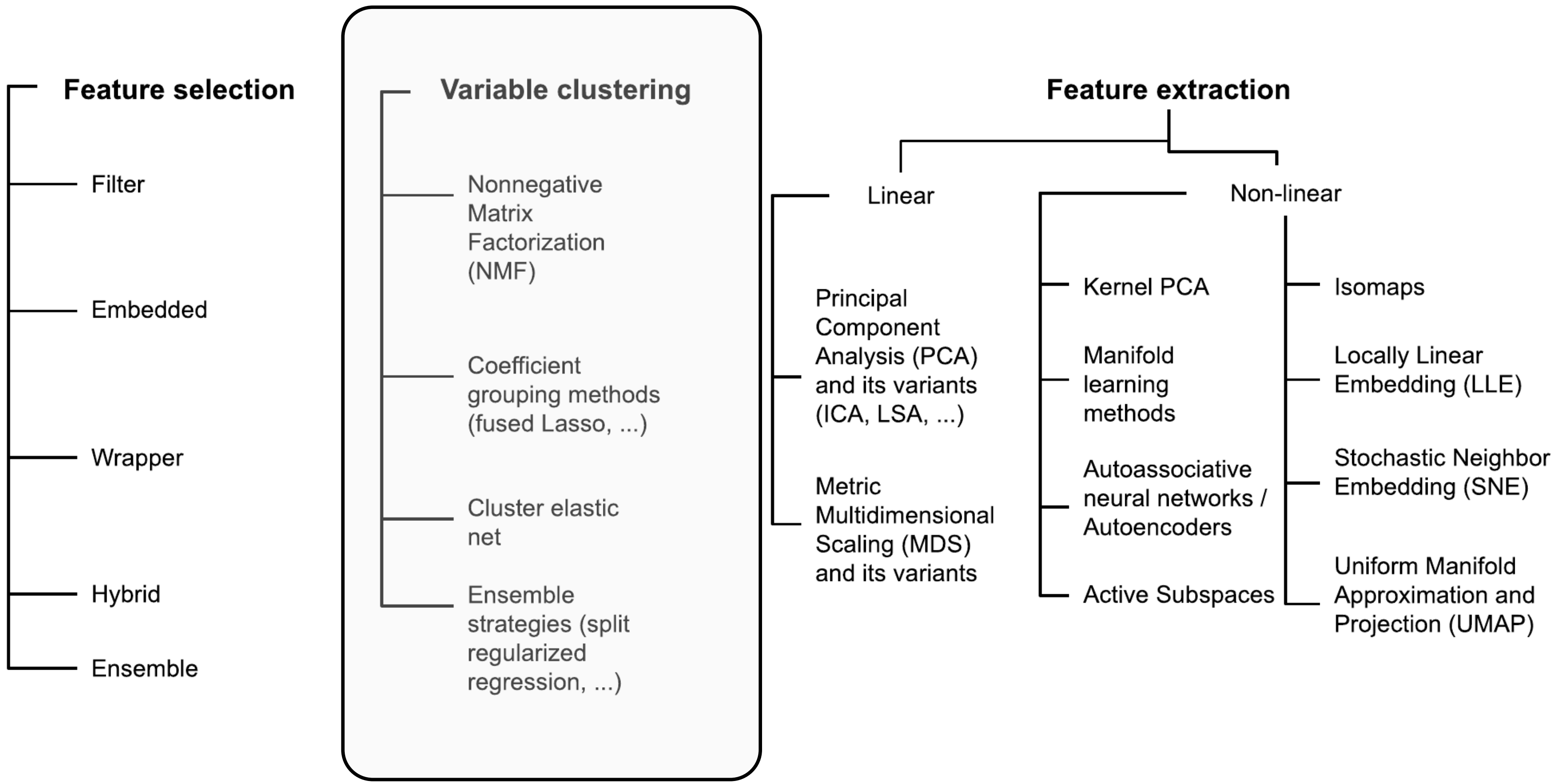
- **Active area of research**
  - **Lots of papers recently published**
    - In various fields (engineering, biotech, finance, etc.)
    - For several applications (image and text processing, data visualization, etc.)
  - **Historical trend of research**
    - First methods focus on global projections (typically: preserve covariance between dimensions, like Principal Component Analysis or PCA)
    - Now: more localized approaches (i.e. preserve pairwise distances or proximities as much as possible)

Taxonomy of dimensionality reduction techniques



Hou, C. K. J. and Behdinan, K. Dimensionality reduction in surrogate modeling: A review of combined methods. Data Science and Engineering, 7(4):402–427, 2022.

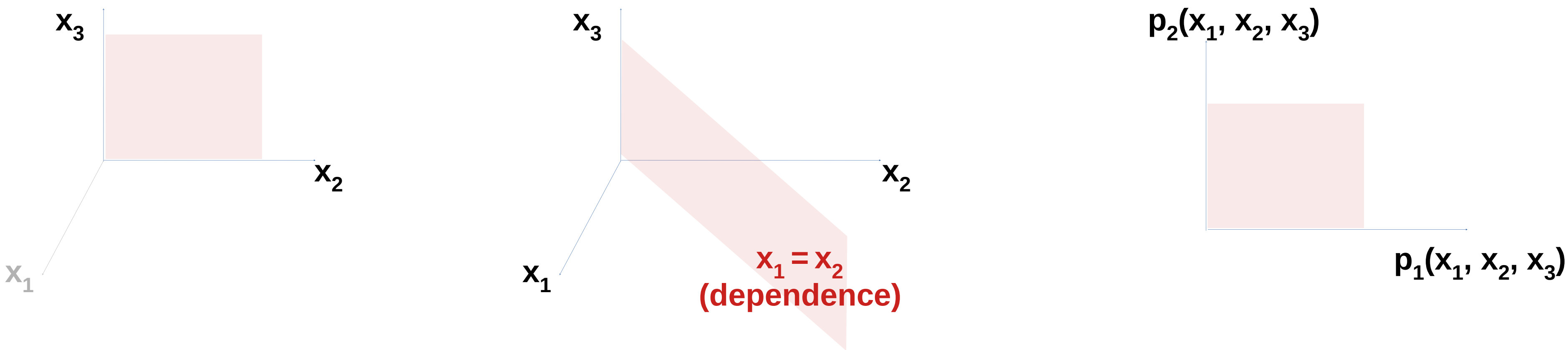
Taxonomy of dimensionality reduction techniques



Hou, C. K. J. and Behdinan, K. *Dimensionality reduction in surrogate modeling: A review of combined methods*. Data Science and Engineering, 7(4):402–427, 2022.

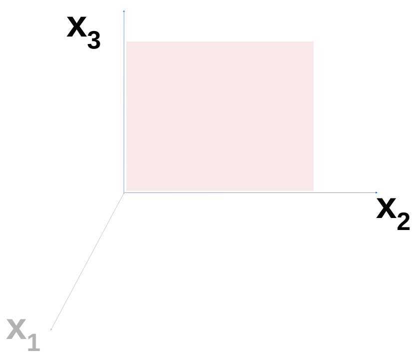
- **Visual comparison of dimensionality reduction paradigms:**

Feature selection **Variable clustering** **Feature extraction**

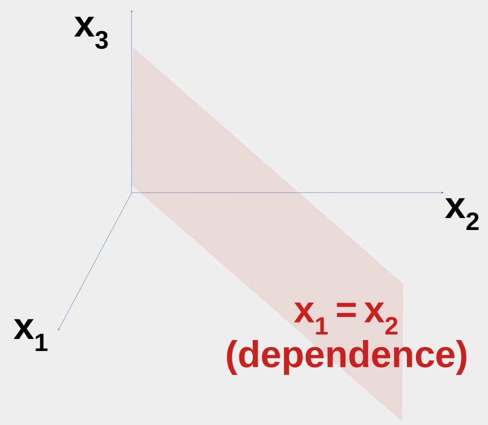


- **Visual comparison of dimensionality reduction paradigms:**

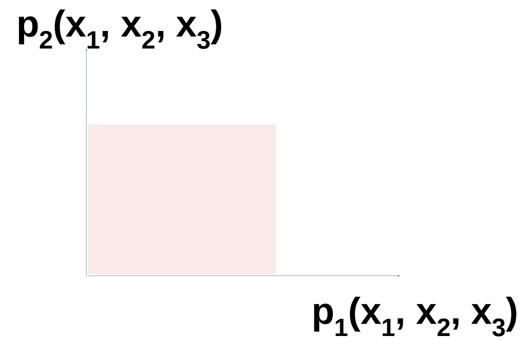
Feature selection



Variable clustering



Feature extraction



Focus on variable clustering:
⇒ **The original parameter space is preserved**
⇒ **Parameters are grouped according to a given criterion, but no one is discarded**

- **Analysis on the state-of-the-art in variable clustering:**
  - **Most methods imply the resolution of a minimization problem expressed as a linear regression or matrix factorization process, coupled with additional penalty functions to deal with grouping**
  - **Techniques directly or remotely inspired by data clustering (k-means, ...) are frequently integrated within the search for variable groups**
  - **Most methods work like in a black-box mode: no visualization tool is proposed to help the user apprehending the high-dimensional space**
  - **In structural and multidisciplinary optimization, the methods proposed usually involve, in one way or another, a significant level of user's expertise, or imply strong physical assumptions**

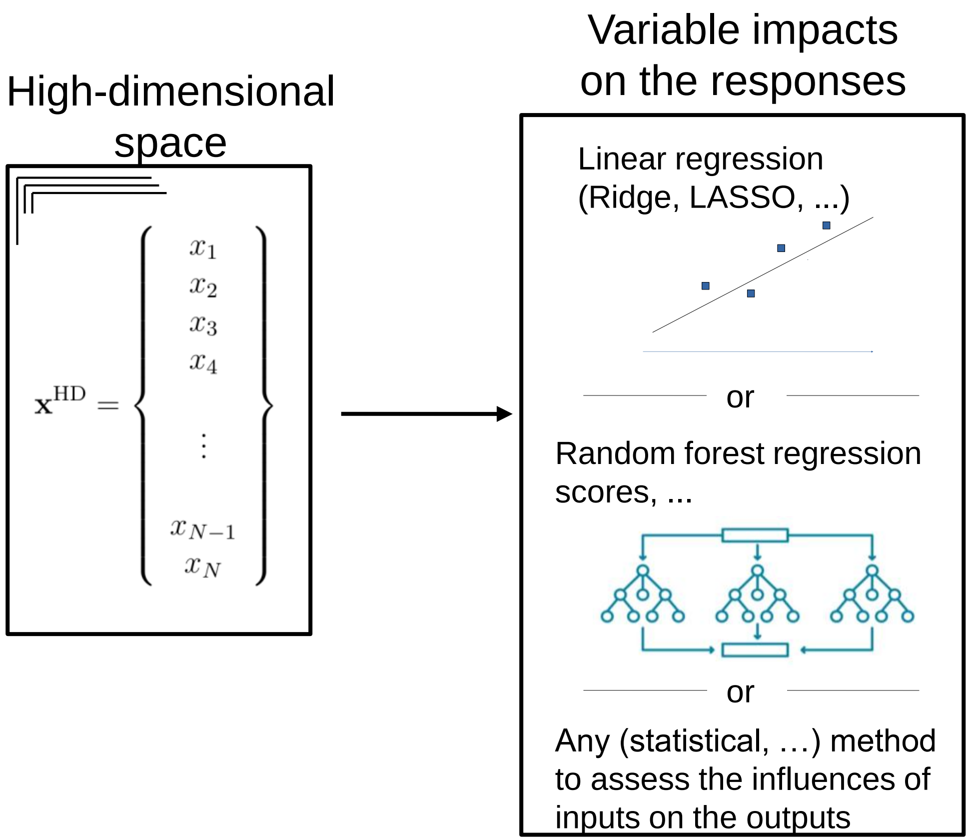
- **Let's remind our objective: devise a generic and flexible variable clustering strategy, with visualization functionalities**

- **Hypothesis: variables with similar impact on the responses are grouped together**

High-dimensional space

$$\mathbf{x}^{\mathrm{HD}} = \left\{ \begin{array}{c} x_1 \\ x_2 \\ x_3 \\ x_4 \\ \vdots \\ x_{N-1} \\ x_N \end{array} \right\}$$

- **Hypothesis: variables with similar impact on the responses are grouped together**



Variable clustering methodology

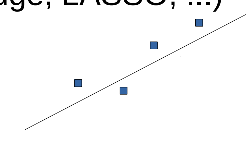
- **Hypothesis: variables with similar impact on the responses are grouped together**

High-dimensional space

$$\mathbf{x}^{\mathrm{HD}} = \begin{Bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ \\ \vdots \\ \\ x_{N-1} \\ x_N \end{Bmatrix}$$

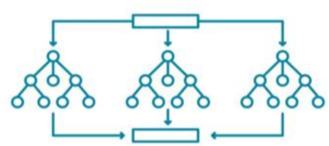
Variable impacts on the responses

Linear regression (Ridge, LASSO, ...)



or

Random forest regression scores, ...

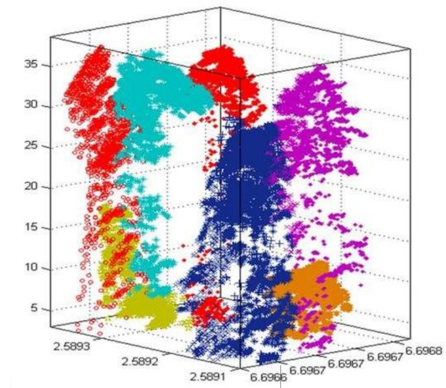


or

Any (statistical, …) method to assess the influences of inputs on the outputs

Variable clustering

DBSCAN, K-MEANS, ...



Variables grouped by importance

$$\mathbf{x}^{\mathrm{grouped}} = \begin{Bmatrix} x_1 \equiv p_1 \\ x_2 \equiv p_2 \\ x_3 \equiv p_2 \\ x_4 \equiv p_1 \\ \\ \vdots \\ \\ x_{N-1} \equiv p_n \\ x_N \equiv p_1 \end{Bmatrix}$$

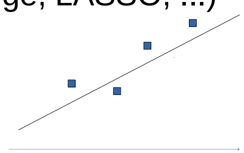
- **Hypothesis: variables with similar impact on the responses are grouped together**

High-dimensional space

$$\mathbf{x}^{\mathrm{HD}} = \left\{ \begin{array}{c} x_1 \\ x_2 \\ x_3 \\ x_4 \\ \vdots \\ x_{N-1} \\ x_N \end{array} \right\}$$

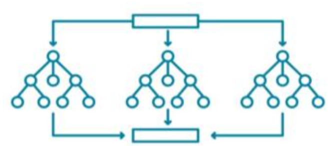
Variable impacts on the responses

Linear regression (Ridge, LASSO, ...)



or

Random forest regression scores, ...



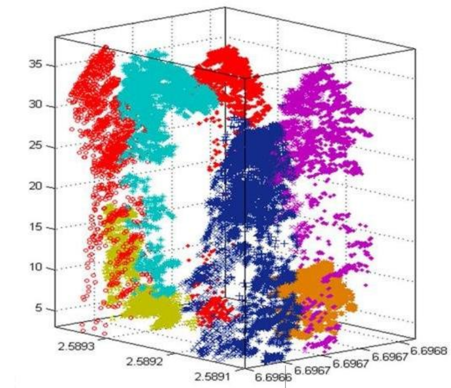
or

Any (statistical, …) method to assess the influences of inputs on the outputs

Issue: poor results observed with (even moderately) high number of variables and responses

Variable clustering

DBSCAN, K-MEANS, ...



Variables grouped by importance

$$\mathbf{x}^{\mathrm{grouped}} = \left\{ \begin{array}{c} x_1 \equiv p_1 \\ x_2 \equiv p_2 \\ x_3 \equiv p_2 \\ x_4 \equiv p_1 \\ \vdots \\ x_{N-1} \equiv p_n \\ x_N \equiv p_1 \end{array} \right\}$$

- **Hypothesis: variables with similar impact on the responses are grouped together**

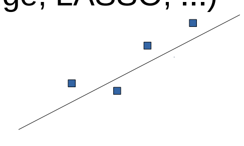
High-dimensional space

$$\mathbf{x}^{\text{HD}} = \left\{ \begin{array}{c} x_1 \\ x_2 \\ x_3 \\ x_4 \\ \vdots \\ x_{N-1} \\ x_N \end{array} \right\}$$

1

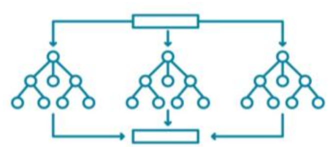
Variable impacts on the responses

Linear regression (Ridge, LASSO, ...)



or

Random forest regression scores, ...

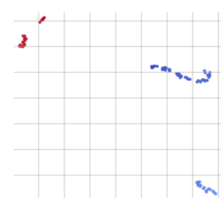


or

Any (statistical, …) method to assess the influences of inputs on the outputs

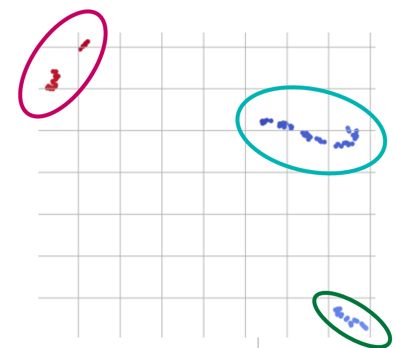
2

Solution: projection on an intermediate latent space, as a pre-conditioner of the clustering



3 Variable clustering

DBSCAN, K-MEANS, ...

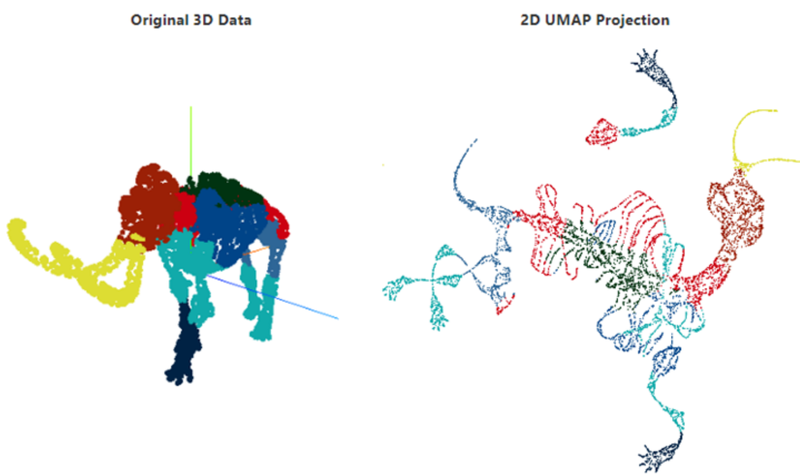


Variables grouped by importance

$$\mathbf{x}^{\text{grouped}} = \left\{ \begin{array}{c} x_1 \equiv p_1 \\ x_2 \equiv p_2 \\ x_3 \equiv p_2 \\ x_4 \equiv p_1 \\ \vdots \\ x_{N-1} \equiv p_n \\ x_N \equiv p_1 \end{array} \right\}$$

Three-step methodology

- **Idea: decomposition in three distinct steps 1, 2, and 3, using the UMAP dimensionality reduction technique at step 2:**
  - **UMAP (Uniform Manifold Approximation and Projection, by McInnes et al., 2018) has gained popularity in recent years in various DR contexts**
  - **Based on fuzzy algebraic topology and Riemannian geometry, UMAP constructs a graph in high dimensions to connect the data, and then performs an optimization to find the most similar graph in lower dimensions**



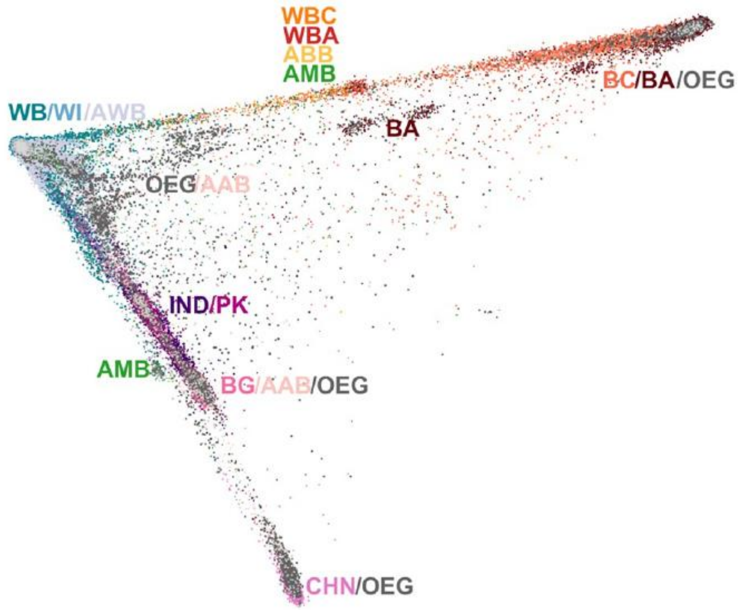
McInnes, L., Healy, J., Saul, N., and Großberger, L. *UMAP: Uniform Manifold Approximation and Projection*. Journal of Open Source Software, 3(29):861, 2018.

UMAP-based variable clustering

Cenaero

Applications of UMAP

- **Various fields of applications (genetics, image processing, ...)**
- **Example: analyzing genotype data among 488,377 individuals in the UK, according to ethnicity [Diaz-Papkovitch *et al.*, 2018]**
  - **Method: dimension reduction of genomic data combining Principal Component Analysis (PCA) with UMAP to illustrate population structure in large cohorts, and capture their relationships on local and global scales**
  - **Aims:**
    - Detect correlations between genetic diseases and ethnic origins
    - Genetic anthropology

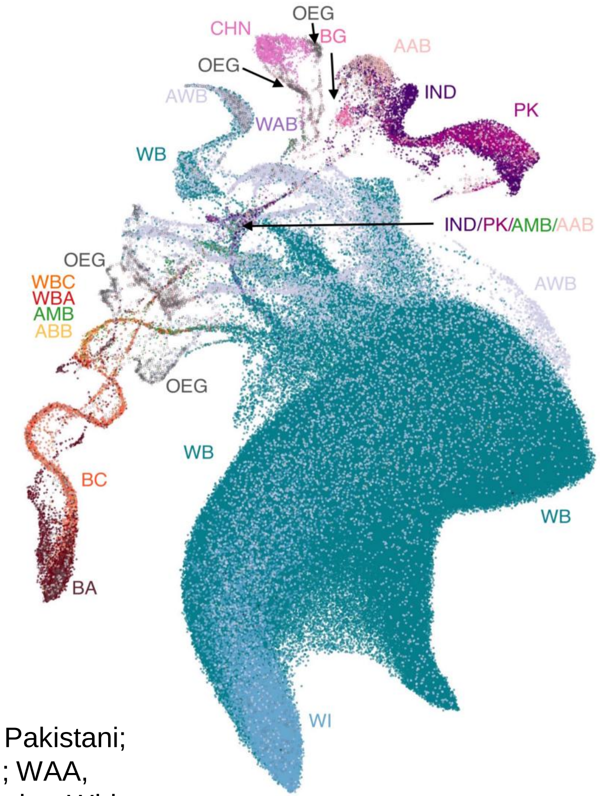


PCA (first 2 modes)

Legend: BA, Black African; BC, Black Caribbean; BG, Bangladeshi; CHN, Chinese; IND, Indian; PK, Pakistani; WB, White British; WI, White Irish; WBC, White and Black Caribbean; WBA, White and Black African; WAA, White and Asian; AAB, Any other Asian Background; ABB, Any other Black Background; AWB, Any other White Background; AMB, Any other Mixed Background; OEG, Other ethnic group.

Applications of UMAP

- **Various fields of applications (genetics, image processing, ...)**
- **Example: analyzing genotype data among 488,377 individuals in the UK, according to ethnicity [Diaz-Papkovitch *et al.*, 2018]**
  - **Method: dimension reduction of genomic data combining Principal Component Analysis (PCA) with UMAP to illustrate population structure in large cohorts, and capture their relationships on local and global scales**
  - **Aims:**
    - Detect correlations between genetic diseases and ethnic origins
    - Genetic anthropology



UMAP on first 10 PCA modes

Legend: BA, Black African; BC, Black Caribbean; BG, Bangladeshi; CHN, Chinese; IND, Indian; PK, Pakistani; WB, White British; WI, White Irish; WBC, White and Black Caribbean; WBA, White and Black African; WAA, White and Asian; AAB, Any other Asian Background; ABB, Any other Black Background; AWB, Any other White Background; AMB, Any other Mixed Background; OEG, Other ethnic group.

- **Step 0: design of experiments (dataset)**
- **Step 1: compute a k-neighbor connected graph in the high-dimensional space**

- **Step 2: define a symmetric *measure of similarity* for each pair of data points *i* and *j*:**

$$p_{ij} := p_{j|i} + p_{i|j} - p_{j|i} \cdot p_{i|j}$$

$$p_{j|i} := \begin{cases} \exp\left(-\frac{\left\|\mathbf{x}^{(i)}-\mathbf{x}^{(j)}\right\|_2 - \rho_i}{\sigma_i}\right) & \text{if } \mathbf{x}^{(j)} \in \mathcal{N}_i \\ 0 & \text{otherwise} \end{cases}$$

- **Step 3: compute a k-neighbor connected graph for the embedded points y*(i)* with a similarity measure in the latent space**

$$q_{ij} := \frac{1}{\left(1 + a\,\|\mathbf{y}^{(i)} - \mathbf{y}^{(j)}\|_2^{2b}\right)}$$

- **Step 4: the discrepancy between the similarity measures in the high- and low-dimensional spaces is computed by the fuzzy cross-entropy, to be minimized:**

$$c_1 := \sum_{i=1}^{N} \sum_{j=1, j\neq i}^{N} \left(p_{ij} \ln\left(\frac{p_{ij}}{q_{ij}}\right) + (1 - p_{ij}) \ln\left(\frac{1 - p_{ij}}{1 - q_{ij}}\right)\right)$$

- **Additional mathematical developments lead to a more tractable function to be minimized:**

$$\min_{\{\mathbf{y}_i\}_{i=1}^{n}} c_2 \quad := \quad \sum_{i=1}^{N} \sum_{j=1, j\neq i}^{N} \left[p_{ij} \ln\left(q_{ij}\right) + (1 - p_{ij}) \ln\left(1 - q_{ij}\right)\right]$$

UMAP algorithm: remarks

- UMAP has demonstrated its effectiveness when compared to other techniques like t-SNE or LargeVis in terms of CPU time and resources

- Further modifications have been proposed in the literature (DenseMap algorithm to regularize the cost function, variant for time series, etc.)

- UMAP can be used in high dimensions (not limited to 2D/3D visualization)

- UMAP is a stochastic algorithm: randomness used both when speeding up approximation steps, and when solving hard optimization problems

- UMAP has been promoted as an efficient pre-conditioner for clustering in a few studies [Allaoui et al., 2020; Hozumi et al., 2021]

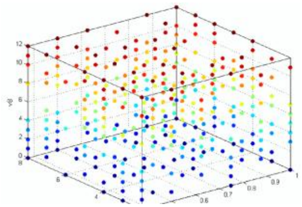
Application to design optimization

Let's go back to our original problem

- **Goal: optimize highly parametrized structural and mechanical systems**
- **Algorithm: surrogate-based optimizer (SBO) from Minamo**
- **Sequential approach followed:**

Design of Experiments

to evaluate the responses on a representative subset of the design space

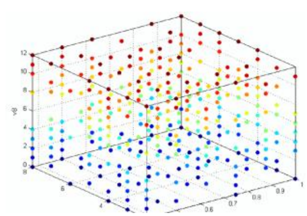


Let’s go back to our original problem

- Goal: optimize highly parametrized structural and mechanical systems
- Algorithm: surrogate-based optimizer (SBO) from Minamo
- Sequential approach followed:

Design of Experiments

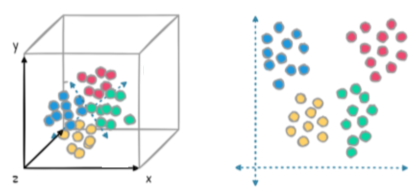
to evaluate the responses on a representative subset of the design space



→

Dimensionality reduction to group variables by importance

(using UMAP)



Cenaero

Let’s go back to our original problem

- Goal: optimize highly parametrized structural and mechanical systems
- Algorithm: surrogate-based optimizer (SBO) from Minamo
- Sequential approach followed:

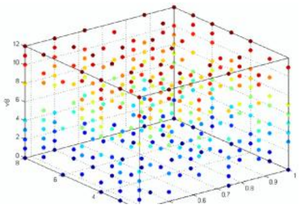
Design of Experiments

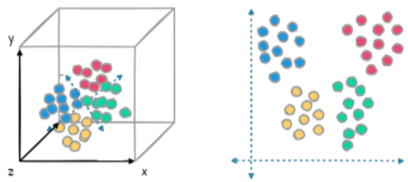
to evaluate the responses on a representative subset of the design space

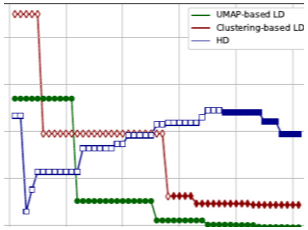
Dimensionality reduction to group variables by importance

(using UMAP)

Surrogate-based optimization (SBO) using Minamo



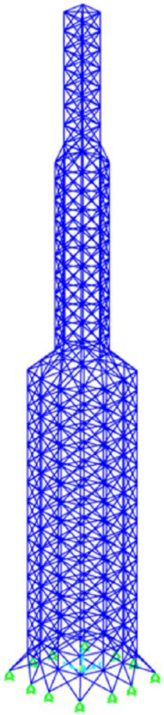




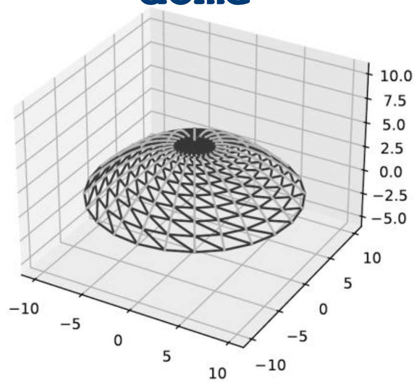
Cenaero

First validation: on structural test cases

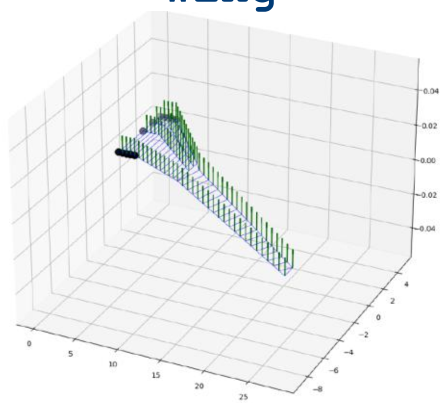
bldg



dome



wing



- **Variables: cross-sectional diameters of beams**
- **Dimensionality:**

Example	Number of variables
`dome`	$np = 696$
`bldg`	$np = 942$
`wing`	$np = 162$

- **Responses: outputs from a finite element analysis [FEAP program, by Taylor *et al.*, 2008]**

$$\left\{\begin{array}{l} \mathrm{mass}(\mathbf{x}) \\ \sigma^{\text{max, external load}}(\mathbf{x}) \\ d^{\text{max, external load}}(\mathbf{x}) \\ s^{\text{max, external load}}(\mathbf{x}) \\ \sigma^{\text{max, dead load}}(\mathbf{x}) \\ d^{\text{max, dead load}}(\mathbf{x}) \\ s^{\text{max, dead load}}(\mathbf{x}) \end{array}\right.$$

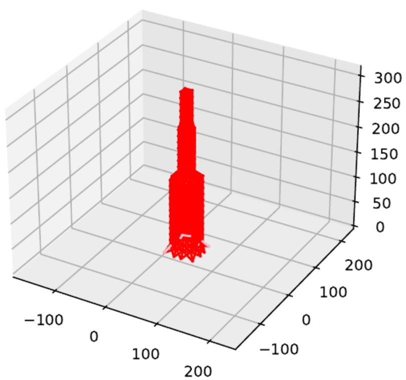
Cenaero

Is the UMAP step really useful?

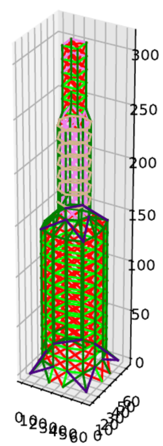
- **Direct clustering vs. UMAP-preconditioned clustering**
  - **Variable impacts on responses estimated by Ridge linear regression coefficients, clustering performed by DBSCAN**

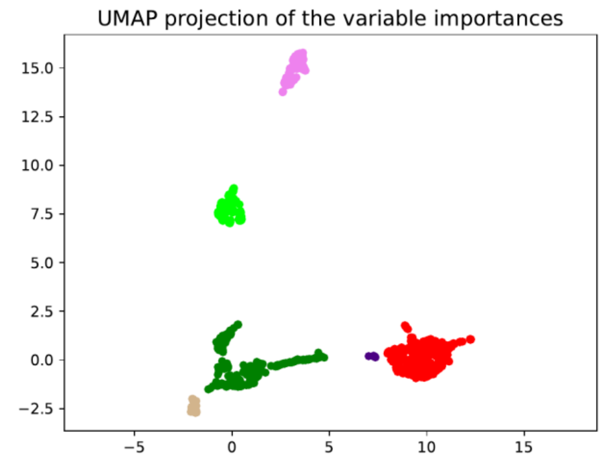
Average number of clusters found	Direct clustering	UMAP-based clustering
bldg	1	7
dome	2	4
wing	1	3

  - **Illustration on the skyscraper test case (`bldg`):**



Direct: 1 cluster found





UMAP-based: 6 clusters found

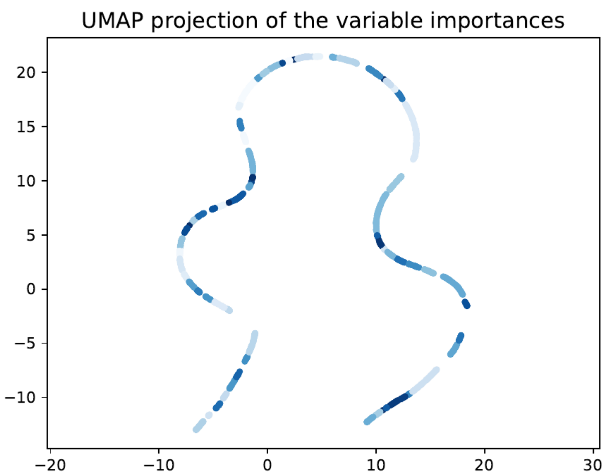
  - **DBSCAN without UMAP-preconditioning fails to identify meaningful clusters of variables (caveat: no hyperparameter tuning)**

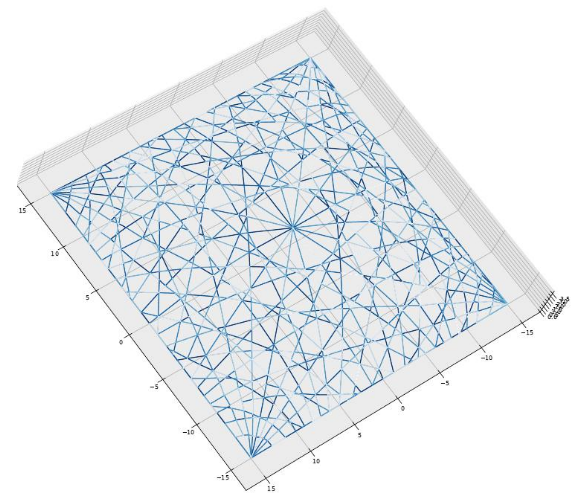
Complex structural application: Dutch Maritime Museum

- **Industrial case: structural optimization of the Dutch Maritime Museum [Descamps *et al.*, 2014 – ULB / Princeton University / Ney & Partners]**
  - **1954 design variables (cross-section areas)**
  - **Objective: mass – Constraints: on stresses, buckling, and displacements**

Complex structural application: Dutch Maritime Museum

- **Industrial case: structural optimization of the Dutch Maritime Museum [Descamps *et al.*, 2014 – ULB / Princeton University / Ney & Partners]**
  - **1954 design variables (cross-section areas)**
  - **Objective: mass – Constraints: on stresses, buckling, and displacements**

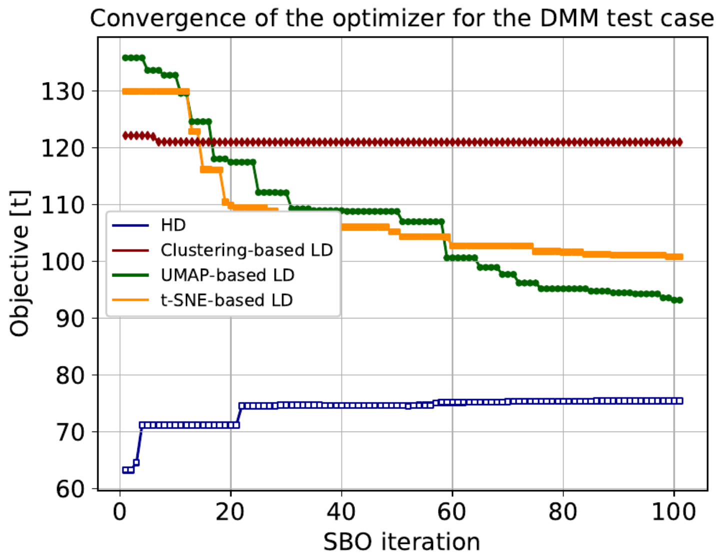


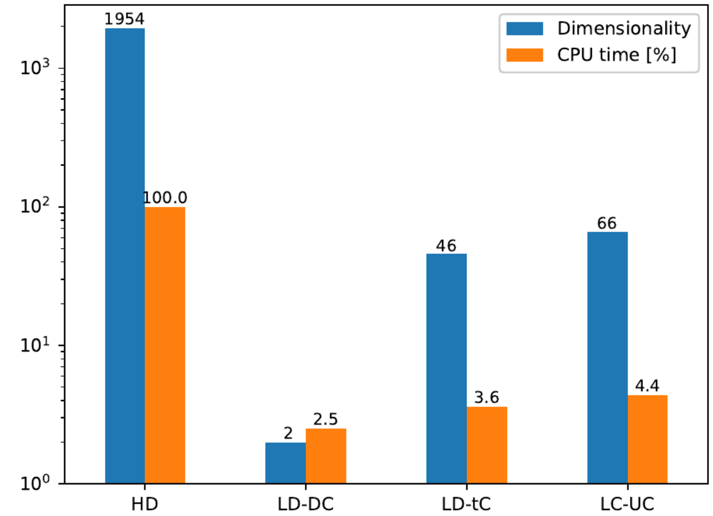


66 clusters detected by UMAP + DBSCAN

Complex structural application: Dutch Maritime Museum

- ## SBO results

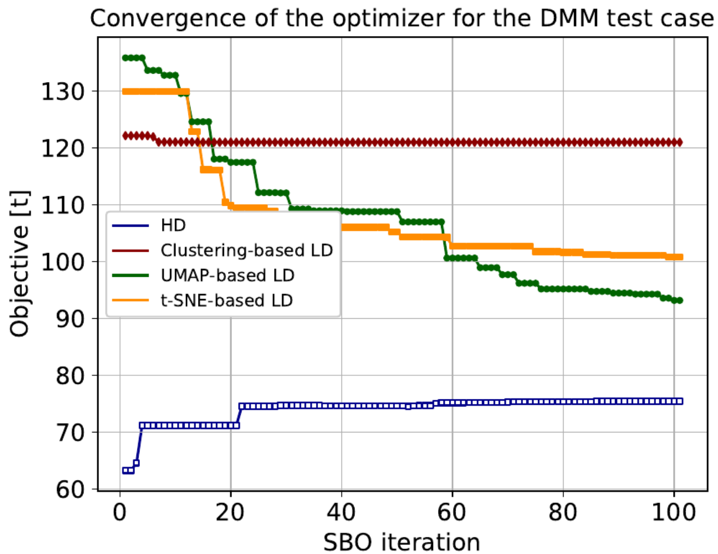


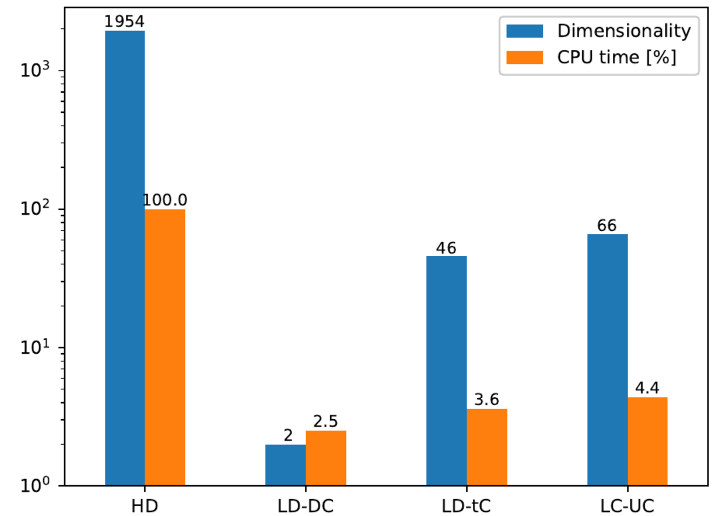


Filomeno Coelho, R., Sainvitu, C., and Benamara, T. *UMAP-based Dimensionality Reduction for Variable Grouping: Application to the Design Optimization of Truss Structures*. In 13th ASMO UK / 2nd ASMO-EUROPE / ISSMO Conference on Engineering Design Optimization Product and Process Improvement, Cenaero, Belgium, July 8-9, 2024.

Complex structural application: Dutch Maritime Museum

- **SBO results**





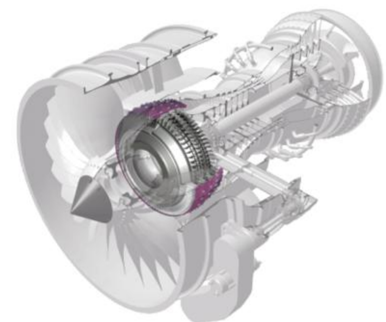
Filomeno Coelho, R., Sainvitu, C., and Benamara, T. *UMAP-based Dimensionality Reduction for Variable Grouping: Application to the Design Optimization of Truss Structures*. In 13th ASMO UK / 2nd ASMO-EUROPE / ISSMO Conference on Engineering Design Optimization Product and Process Improvement, Cenaero, Belgium, July 8-9, 2024.

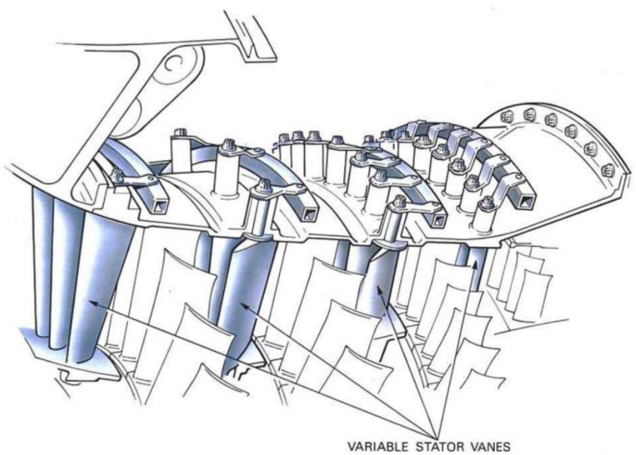
- **Take aways from these first results:**
  - **Strategy successfully tested on cases up to 1954 design variables, but on structural design cases only with symmetry / repeatibility patterns**
  - **What about different mechanical systems like compressors?**
  - **Should the same variable grouping be used during the whole SBO?**

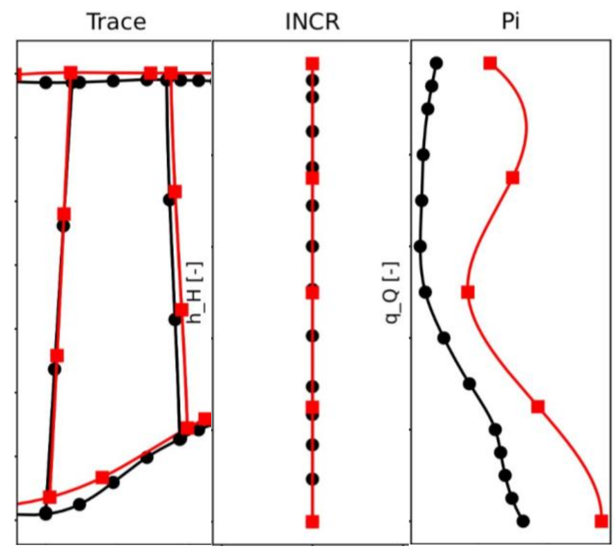
Cenaero

Multi-stage compressor (Safran Aero Boosters)

- **Number of input parameters:**
  - **N = 144 design variables**
  - **Simulation: axisymmetric throughflow CFD solver, widely used at the preliminary design stage in turbomachinery [Simon, 2007]**
  - **Rate of successful simulations: ~55%**



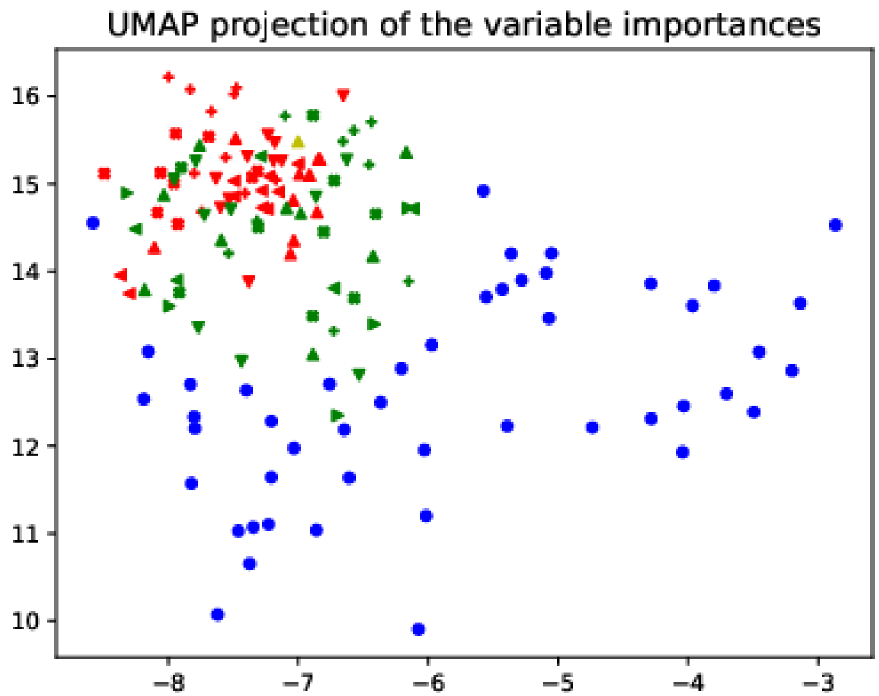




Nigro, R., Baert, L., Nyssen, F., de Cazenove, J., Dominique, J., Lepot, I., Veglio, M., and Princivalle, R. *Multi-fidelity aeromechanical design framework for high flow speed multistage axial compressors*. In: Proceedings of the ASME TurboExpo 2024 conference, London, UK, June 23-28, 2024.

Multi-stage compressor (Safran Aero Boosters)

- **Preliminary results on a multi-stage compressor (SAB)**
  - **Design of Experiments of 5 N = 720 geometries (LHS – Latin Hypercube Sampling)**
  - **Dimensionality reduction performed on 1175 responses**
  - **Agglomerative clustering better adapted to this configuration**



Color:

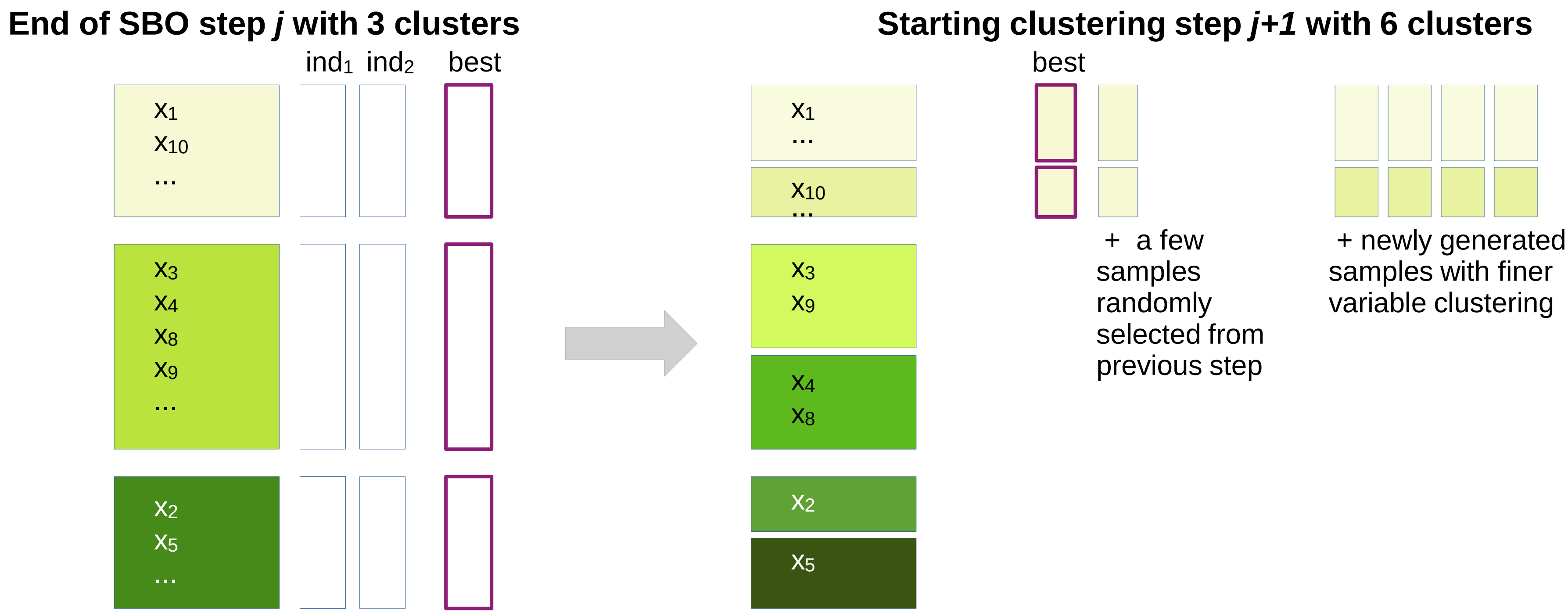
- If "flow_path": blue
- If "aero": red
- If "geom": green
- Otherwise: yellow

Marker:

- S1: v
- S2: ^
- S3: <
- ST: >
- R2: +
- R3: x

Multi-stage compressor (Safran Aero Boosters)

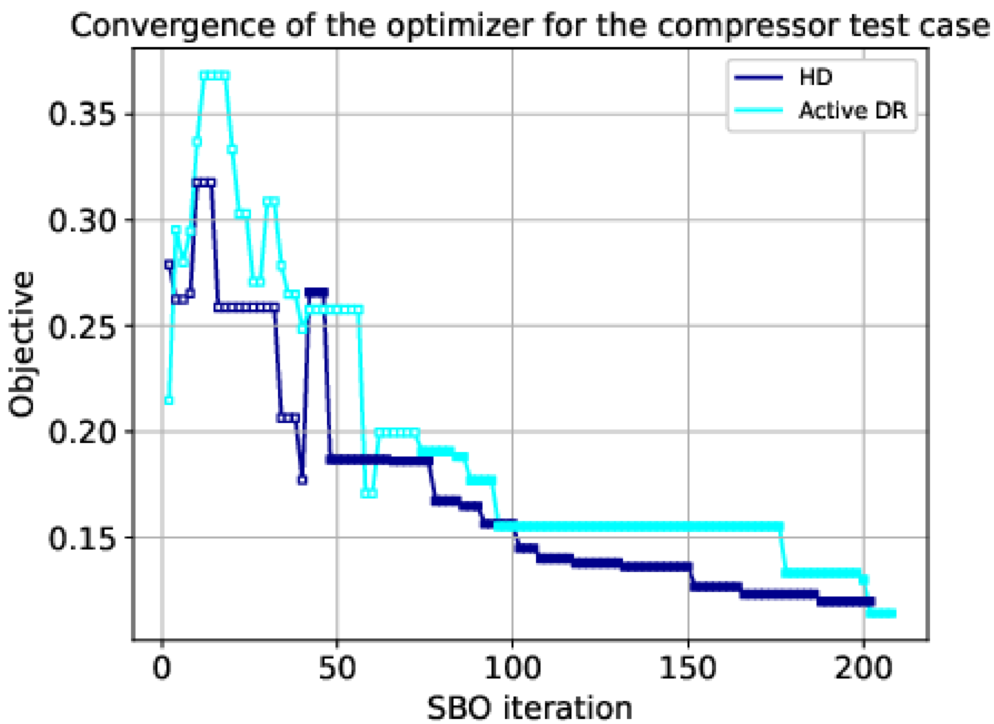
- **Idea: update the clustering of variables during the SBO**
  - **Caution: when switching from a “coarse-grained” to a “fine-grained” variable clustering, there must be a mapping between levels, in order to re-use points from previous SBO iterations:**

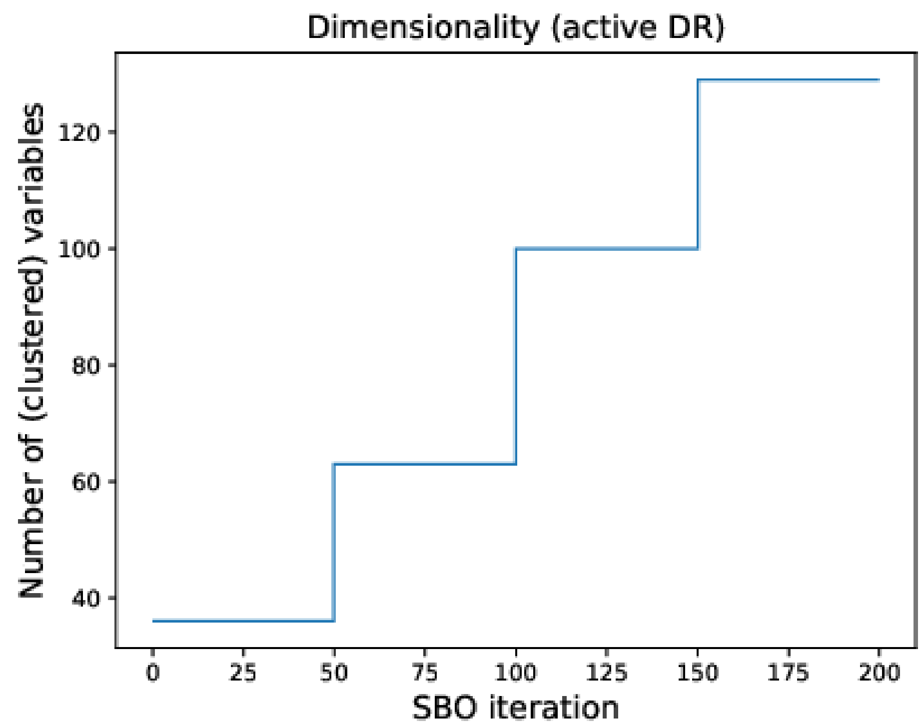


Cenaero

Multi-stage compressor (Safran Aero Boosters)

- **Preliminary results on a multi-stage compressor (SAB)**
  - **SBO: 1 objective (minimize fuel burn) and 122 aerodynamical constraints**
  - **Similar performances obtained, but with a decrease of ~50% CPU time required (faster training of the surrogates due to reduced dimensionalities)**





Conclusions and future prospects

Discussion

- **Design of Experiments**
  - Size of the design of experiments: compromise between accuracy of the results and cost of the simulations

- **Variable importances:**
  - Ridge regression: computationally inexpensive and often acceptable [Grömping, 2009], but might be insufficient in some cases
  - Currently investigated: Random forests (RF), Mutual information (MI) scores [Université de Namur]
  - Grouping dependent on active constraints only?

- **UMAP for variable clustering**
  - UMAP useful to pre-condition data for clustering
  - Strategy efficiently combined to surrogate-based optimization
  - Ongoing developments on active learning strategies

References

References

A Abbas, A Rafiee, and M Haase. DeepMorpher: Deep learning-based design space dimensionality reduction for shape optimisation. Journal of Engineering Design, pages 1–17, April 2023.

N Ahmad and A Nassif. Dimensionality Reduction: Challenges and Solutions. ITM Web of Conferences, 43:01017, March 2022.

M. Allaoui, M. L. Kherfi, and A. Cheriet. Considerably Improving Clustering Algorithms Using UMAP Dimensionality Reduction Technique: A Comparative Study,Image and Signal Processing, vol. 5, pp. 317-325, June 2020.

M Alswaitti, K Siddique, S Jiang, W Alomoush, and A Alrosan. Dimensionality Reduction, Modelling, and Optimization of Multivariate Problems Based on Machine Learning. Symmetry, 14(7), 2022.

C Audouze, F De Vuyst, and P B Nair. Reduced-order modeling of parameterized PDEs using time-space-parameter principal component analysis. International Journal for Numerical Methods in Engineering, 80:1025–1057, 2009.

C M Bishop. Pattern Recognition and Machine Learning. Springer-Verlag, Heidelberg, Germany, 2006.

D Brockhoff and E Zitzler. Dimensionality Reduction in Multiobjective Optimization: The Minimum Objective Subset Problem. In K-H Waldmann and U M Stocker, editors, Operations Research Proceedings 2006, pages 423–429. Springer Berlin/Heidelberg, 2007.

V Bruni, M L Cardinali, and D Vitulano. A Short Review on Minimum Description Length: An Application to Dimension Reduction in PCA. Entropy, 24(2), 2022.

References

M Á Carreira-Perpiñán. A Review of Dimension Reduction Techniques. Technical report, University of Sheffield.

C Cartis and Ad Otemissov. A dimensionality reduction technique for unconstrained global optimization of functions with low effective dimensionality. Information and Inference: A Journal of the IMA, 11(1):167–201, May 2021.

A Chakraborti, H P N Nagarajan, S Panicker, H Mokhtarian, E Coatanéa, and K T Koskinen. A Dimension Reduction Method for Efficient Optimization of Manufacturing Performance. Procedia Manufacturing, 38:556–563, 2019.

R Chauhan, D. Mathew. Review on Dimensionality Reduction Techniques. In International Conference on Computer Networks, Big data and IoT, Madurai, India, December 19-20 (2019), pages 356–362, March 2020.

A-A Christidis, L Laks, E Smucler, and R Zamar. Split Regularized Regression. Technometrics, 62(3):330–338, 2020.

R Connor and L Vadicamo. nSimplex Zen: A Novel Dimensionality Reduction for Euclidean and Hilbert Spaces. Technical report, University of Saint-Andrews, February 2023.

P G Constantine. Active Subspaces: Emerging Ideas for Dimension Reduction in Parameter Studies. SIAM Spotlights, 2015.

P G Constantine, E Dow, and Q Wang. Active Subspace Methods in Theory and Practice: Applications to Kriging Surfaces. SIAM J. on Scientific Computing, 36(4):1500–1524, 2014.

C de Bodt, J Lee, L Journaux, and L Sautot. Handbook of Proximity Relations, chapter Proximities in dimensionality reduction, pages 165–203. Edward Elgar Publishing, 2022.

References

A De Mauro, M Greco, and M Grimaldi. A formal definition of Big Data based on its essential features. Library Review, 65:122–135, March 2016.

A Diaz-Papkovich, L Anderson-Trocmé, and S Gravel. Revealing multi-scale population structure in large cohorts. PLOS Genetics, doi: 10.1371/journal.pgen.1008432, 2018.

A Franc. Linear Dimensionality Reduction. Technical report, French National Institute for Agriculture, Food, and Environment (INRAE), 2022.

A H Gandomi, K Deb, R C Averill, S Rahnamayan, and M Nabi Omidvar. Variable functioning and its application to large scale steel frame design optimization. Structural and Multidisciplinary Optimization, 66:13, 2023.

F J García-Fernández, M Verleysen, J A Lee, and I Díaz Blanco. Stability comparison of dimensionality reduction techniques attending to data and parameter variations. In Eurographics Conference on Visualization (EuroVis), 2013.

M Garcia Torres, F Gómes-Vila, B Melian, and J Marcos Moreno-Vega. High-dimensional feature selection via feature grouping: a Variable Neighborhood Search Approach. Information Sciences, 326:102–118, 2016.

A Gracia, S González, V Robles, and E Menasalvas. A methodology to compare Dimensionality Reduction algorithms in terms of loss of quality. Information Sciences, 270:1–27, 2014.

U Grömping. Variable Importance Assessment in Regression: Linear Regression versus Random Forest. The American Statistician, 63(4):308–319, November 2009.

References

T Hastie, R Tibshirani, and J Friedman. The Elements of Statistical Learning: Data Mining, Inference and Prediction. Springer Series in Statistics. Springer, New York, 2 edition, 2009.

J He, Y Luo, and K Wang, R Cuo. A dimensionality reduction and layout optimization scheme for grid-stiffened structures. Engineering Optimization, pages 1–17, 2022.

G Hinton and S Roweis. Stochastic Neighbor Embedding. Advances in neural information processing systems, 15.

C K J Hou and K Behdinan. Dimensionality reduction in surrogate modeling: A review of combined methods. Data Science and Engineering.

Y. Hozumi, R. Wang, C. Yin, and G.-W. Wei. UMAP-assisted K-means clustering of large-scale SARS-CoV-2 mutation datasets, Computers in Biology and Medicine, vol. 131, p. 104264, 2021.

IDC. The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things (#IDC_1672), IDC Whitepaper sponsored by EMC – April, 2014.

G James, D Witten, T Hastie, and R Tibshirani. An Introduction to Statistical Learning: with Applications in R. Springer, New York, 2014.

W Jia, M Sun, J Lian, and S Hou. Feature dimensionality reduction: a review. Complex & Intelligent Systems, 8:2663–2693, January 2022.

A C C Lemonge, H J C Barbosa, A L G A Coutinho, and C C H Borges. Multiple cardinality constraints and automatic member grouping in the optimal design of steel framed structures. Engineering Structures, 33:433–444, 2011.

References

D J C MacKay. Information Theory, Inference, and Learning Algorithms. Copyright Cambridge University Press, 2003.
O Maimon and L Rokach. Data Mining and Knowledge Discovery Handbook. Springer, 2010.
M A Mendez. Linear and Nonlinear Dimensionality Reduction from Fluid Mechanics to Machine Learning. Technical report, von Karman Institute, 2022.
R Marion. Statistical and machine learning methods for identifying clusters of variables: with applications in omics, ecology and psychology. PhD thesis, Université catholique de Louvain, Belgium, 2021.
L McInnes, J Healy, N Saul, and L Großberger. UMAP: Uniform Manifold Approximation and Projection. The Journal of Open Source Software, 3(29):861, 2018.
S Nanga, A Bawah, B Acquaye, M Billa, F Baeta, N Odai, S Obeng, and A Nsiah. Review of Dimension Reduction Methods. Journal of Data Analysis and Information Processing.
N Navaneeth and S Chakraborty. Surrogate assisted active subspace and active subspace assisted surrogate: A new paradigm for high dimensional structural reliability analysis. Com-
puter Methods in Applied Mechanics and Engineering, 24(2), 2022.
S Patil, S G Joshi, and S V Patil. Grouping & Optimization of Structural Design Using FCM Algorithm. Technical report, KLE Instititute of Technology, India, 2022.
K B Petersen and M S Pedersen. The Matrix Cookbook. Technical report, Technical University of Denmark, November 2012.

References

Y Qiu, J Bai, N Lui, and C Wang. Global aerodynamic design optimization based on data dimensionality reduction. Chinese Journal of Aeronautics, 31(4):643–659, 2018.

R Ran, T Gao, and Y Zhang. Transformer-based dimensionality reduction. Technical report, Chongqing Normal University, Bhina, 2022.

L Saul, K Weinberger, F Sha, J-H Ham, and D Lee. Semisupervised Learning, chapter Spectral Methods for Dimensionality Reduction. MIT Press, 2006.

B Schölkopf, A Smola, and K-R Müller. Kernel Principal Component Analysis. In W Gerstner, A Germond, M Hasler, and J-D Nicoud, editors, Artificial Neural Networks – ICANN'97, pages 583–588. Springer Berlin Heidelberg, 1997.

J-F Simon. Contribution to Throughflow Modelling for Axial Flow Turbomachines. PhD thesis, Université de Liège, Belgium, 2007.

P Sobczyk, S Wilczynski, M Bogdan, P Graczyk, J Josse, F Panloup, V Seegers, and M Staniak. VARCLUST: clustering variables using dimensionality reduction, 2020.

R Tibshirani, M Saunders, S Rosset, J Zhu, and K Knight. Sparsity and smoothness via the fused lasso. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67(1):91–108, 2005.

A Tugilimana, R Filomeno Coelho, and A P Thrall. An integrated design methodology for modular trusses including dynamic grouping, module spatial orientation, and topology optimization. Structural and Multidisciplinary Optimization, 60(2):613–638, April 2019.

References

S Ullah, D A Nguyen, H Wang, S Menzel, B Sendhoff, and T Bäck. Exploring Dimensionality Reduction Techniques for Efficient Surrogate-Assisted Optimization. In 2020 IEEE Symposium Series on Computational Intelligence (SSCI), pages 2965–2974, 2020.

L van der Maaten, E Postma, and J Van Den Herik. Dimensionality Reduction: A Comparative Review. Journal of Machine Learning Research, 10(1):66–71, 2007.

T R van Woudenberg and F P van der Meer. A grouping method for optimization of steel skeletal structures by applying a combinatorial search algorithm based on a fully stressed design. Engineering Structures, 249:13299, 2021.

S Velliangiri, S Alagumuthukrishnan, and S Iwin Thankumar Joseph. A review of Dimensionality Reduction Techniques for Efficient Computation. Procedia Computer Science, 165: 104–111, 2019.

E Vigneau and M Chen. Dimensionality reduction by clustering of variables while setting aside atypical variables. Electronic Journal of Applied Statistical Analysis, 9(1):134–153, 2016.

A Viswanath, A I J Forrester, and A J Keane. Dimension Reduction for Aerodynamic Design Optimization. AIAA Journal, 49(6):1256–1266, 2011.

R Walls and A Elvin. An Algorithm for Grouping Members in a Structure. Engineering Structures, 32(6):1760–1768, 2010.

Y Wang, H Yao, and S Zhao. Auto-encoder based dimensionality reduction. Neurocomputing, 184:232–242, 2016.

References

M Xiao, P Breitkopf, R Filomeno Coelho, C Knopf-Lenoir, M Sidorkiewicz, and P Villon. Model reduction by CPOD and Kriging–Application to the shape optimization of an intake port. Structural and Multidisciplinary Optimization, 41(4):555–574, 2010.

M Zandehshahvar, O Hemmatyar, Y Kiarashinejad, S Abdollahramezani, and A Adibi. Dimensionality Reduction Based Method for Design and Optimization of Optical Nanostructures Using Neural Network. In Frontiers in Optics + Laser Science APS/DLS, page FM5C.2. Optica Publishing Group, 2019.

R Zebari, A Mohsin Abdulazeez, D Zeebaree, D Zebari, and J Saeed. A Comprehensive Review of Dimensionality Reduction Techniques for Feature Selection and Feature Extraction. Journal of Applied Science and Technology Trends, 1:56–70, May 2020.

J Zhang, W Wang, Q Wu, and L Hu. A GAN-based dimensionality reduction technique for aerodynamic shape optimization. Technical report, University of Electronic Science and Technology of China, July 2021.