

# Six unsolved issues on the current state of MRM

June 2022

# Agenda

---

Data and documentation (still an issue!)

---

MRM culture: a capillarity problem

---

Automation of the model lifecycle: a pipe-dream?

---

Model risk of machine learning models

---

Model risk beyond models: non-models

---

MRM as a source of efficiency and cost reduction

---

Final remarks

---



---

## Data and documentation (still an issue!)

---

# Data and Documentation

Even if not the most innovative, data and doc are the greatest needs and what will contribute most to efficiency. Moreover, most of the supervisory findings and those that are most difficult to remedy have to do with data and documentation

## Fixing the Data

A major challenge is to ensure data quality, traceability, integrity and consistency. This implies identifying the data golden sources as a single informational system per model family.

### Data initiatives

- Historical data remediation
- Data dictionary and traceability
- Data quality
- Architecture harmonization
- Go to cloud

## Fixing the Documentation

A structural revision oriented to efficiency through the completeness, quality and accessibility of documents. For example: model documentation, frameworks, policies and other standards in the model life cycle.

### Documentation initiatives

- Assessment and fill in the gaps for taxonomy, the documentary tree and related policies.
- Historical remediation for all model documents

## In practice: tactical approaches

### Technical writers

A team of technical writers whose expertise is to:

- Assess current available model data and documentation.
- Re-structure and complete all model risk documentation, including model development, monitoring, implementation, validation, approval and use.

### Automate doc update

- Creating templates and self-updating documents to automate the generation of model-related documentation.
- Having guidelines with minimum core requirements aligned with regulation and MRM policies, quality checks and best practice.

*The highest efficiency return comes from data and doc initiatives*





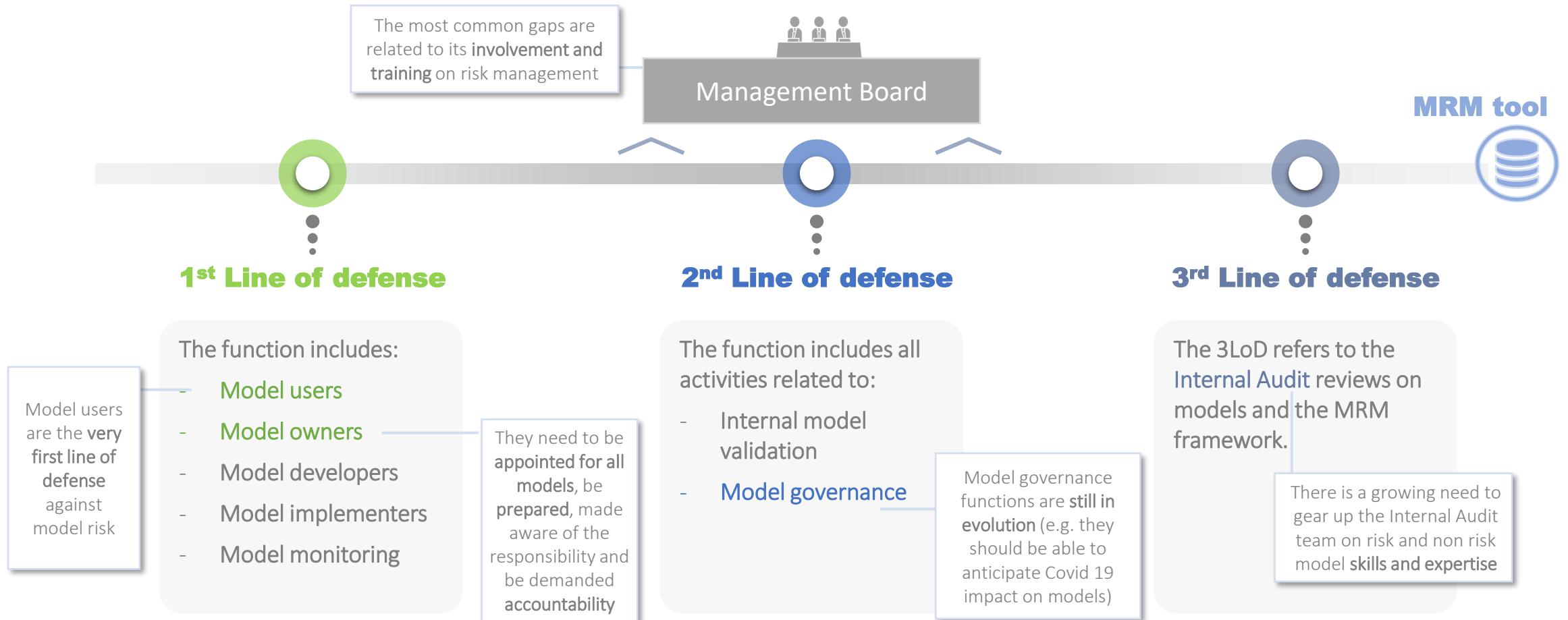


---

## MRM culture: a capillarity problem

---

A mature MRM culture means the involvement of the designated roles throughout the three lines of defense in a collaborative way of working, and the awareness of the Management Board





---

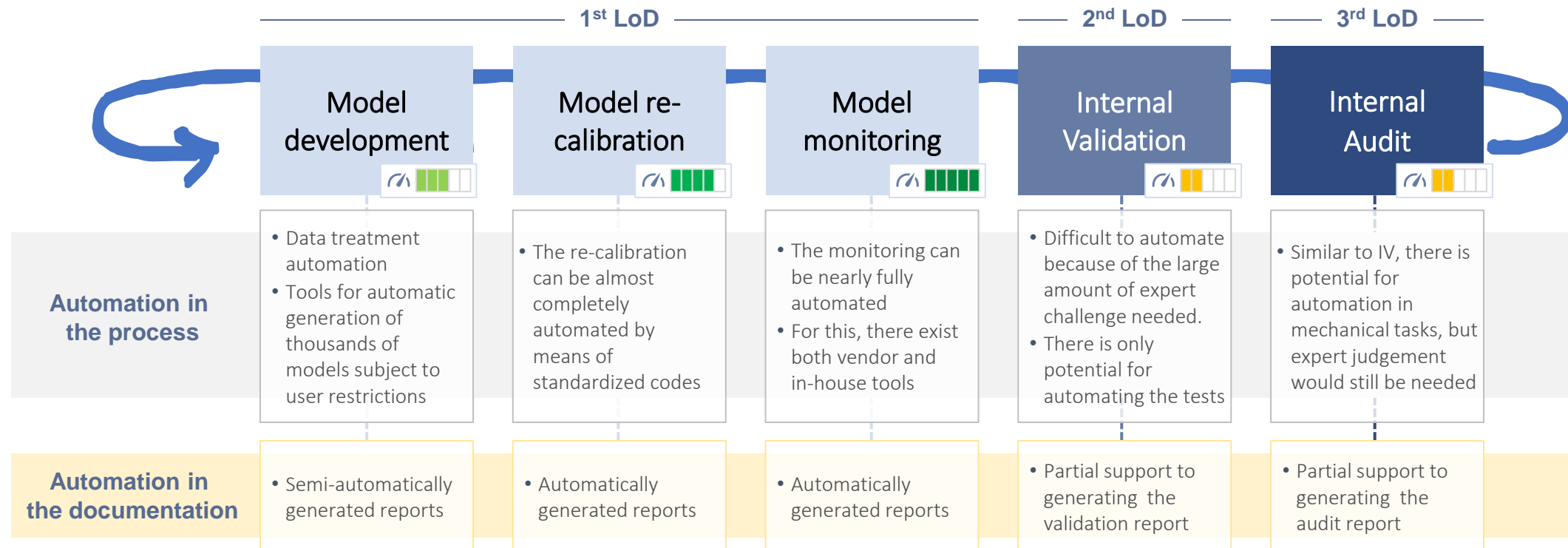
## Automation of the model lifecycle: a pipe-dream?

---

The potential for automation varies depending on the phase of the model life cycle; mechanical tasks can easily be automated, but human judgment is still irreplaceable in some phases

## >> Potential for automation through the model life cycle

 Automation potential







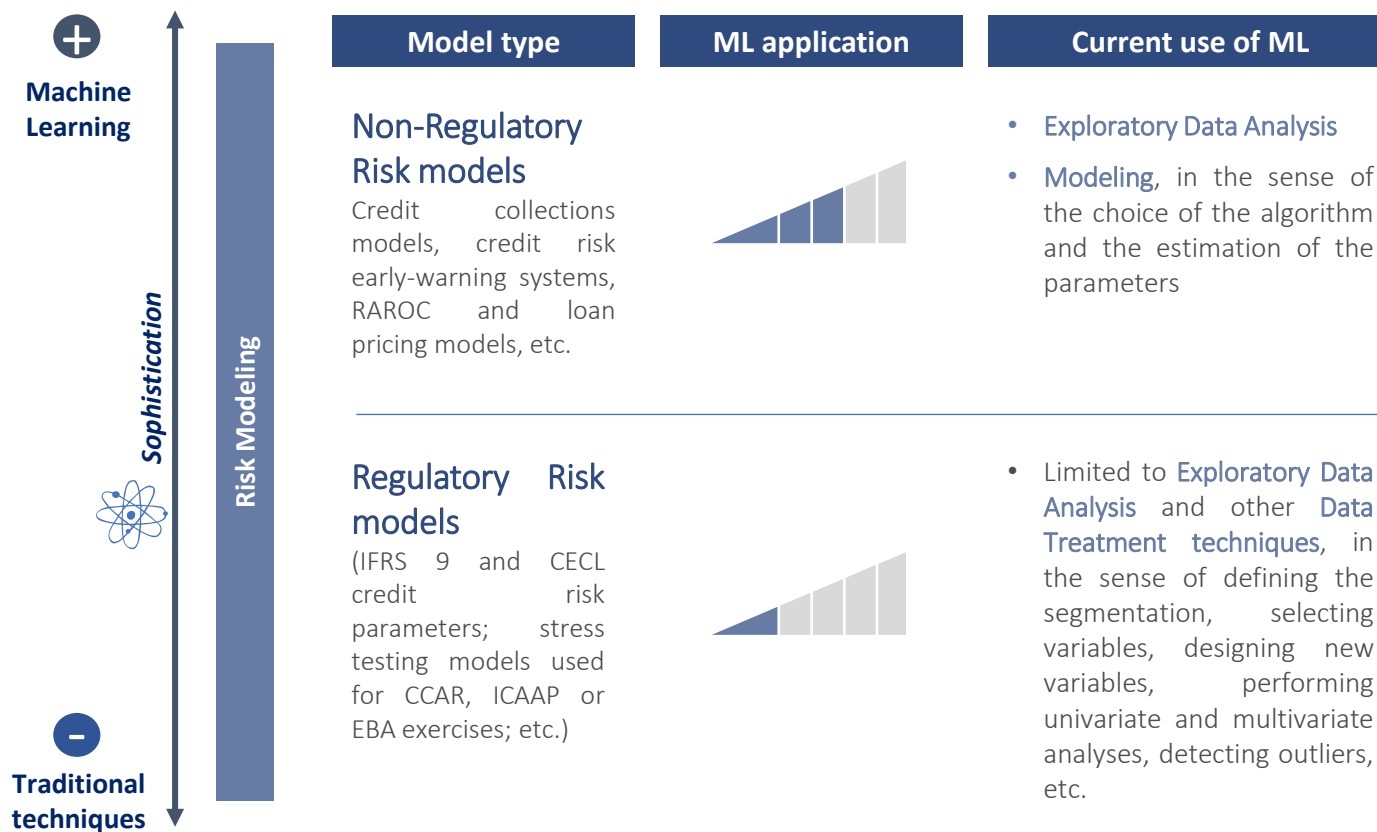
---

# Model risk of machine learning models

---

# Model Risk of Machine Learning Models

Regulatory risk models incorporate machine-learning techniques mostly for exploratory data analysis, while non-regulatory risk models include these techniques also for the actual modeling



## Machine Learning in Retail









### Case study

Use of a gradient boosting model for retail consumer loans admission:

- Increase in the accuracy of the model:** both type I and type II errors were reduced, and one year later the predictive power was sustained over new loans (no evidence of overfitting).
- Presenting the supervisor with this model for **AIRB approval would be problematic** (extension of response times, increase of the risk of rejection).
- Adopted **strategy:**
  - Limit the usage of the gradient boosting to the scoring of consumer loans strictly for admission
  - After two months, a traditional behavioral scoring model (logistic regression) was used
  - The sub-portfolio of retail mortgages with less than two months was rather immaterial in terms of exposure, so a PPU may be requested in the future.
- Through this regulatory workaround, the bank currently **benefits from the predictive power of an advanced machine-learning** technique for an essential process (credit admission).

## ML&AI models pose challenges throughout the entire model lifecycle: data treatment, model design, implementation, validation and documentation

	ITEM 	ISSUE 	SOLUTION 
Data	I Representativeness	<ul style="list-style-type: none"> <li>To ensure that the resulting dataset is still <b>representative of the population</b> it intends to model, even if data are gathered from a variety of sources, merged, transformed using automated ML techniques.</li> </ul>	<ul style="list-style-type: none"> <li>✓ <b>Traditional representativeness analyses</b> are usually valid.</li> <li>✓ Classical hypothesis tests like Z-test, Kolmogorov-Smirnov, Chi-Square, Kruskal-Wallis, etc.</li> <li>✓ Any deviations from representativeness to be justified by developers.</li> </ul>
	II Traceability and DQ	<ul style="list-style-type: none"> <li>To ensure that each data point used in the model estimation <b>traces back to the source system with no errors</b> in the extract, transform and load (ETL).</li> </ul>	<ul style="list-style-type: none"> <li>✓ Framework based on <b>checkpoints</b>, for e.g. comparison indicators in each intermediate dataset</li> <li>✓ Model developers should provide a why and how records are transformed or filtered out.</li> </ul>
	III Feature engineering	<ul style="list-style-type: none"> <li>To assess whether these synthetic variables are <b>correctly built, pertinent and fit-for-purpose</b>, especially when they have been created using ML techniques.</li> </ul>	<ul style="list-style-type: none"> <li>✓ Theoretical description of the algorithm for the feature engineering (when it is automated).</li> <li>✓ Deep analysis of the <b>actual variables in the resulting dataset</b>, qualitative and quantitative.</li> </ul>
Conceptual soundness	IV Model design	<ul style="list-style-type: none"> <li>To effectively challenge the <b>choice of the machine-learning algorithm</b>, including the hyperparameters.</li> </ul>	<ul style="list-style-type: none"> <li>✓ Model performance (predictive power) although important, <b>should not be the only criterion</b></li> <li>✓ <b>Additional drivers</b> should weigh in, such as interpretability, cost of implementing, maintaining and monitoring the model.</li> </ul>
	V Model assumptions and limitations 	<ul style="list-style-type: none"> <li>To assess that they have all been <b>identified and properly documented</b>.</li> <li>To effectively challenge a model that is <b>permanently changing</b>, even sometimes literally with every new observation? (dynamic learning)</li> </ul>	<ul style="list-style-type: none"> <li>✓ A deep assessment of the dynamic recalibration technique, including a <b>sensitivity analysis</b>.</li> <li>✓ <b>Periodic checkpoints</b> where the latest recalibrated version of the model is backtested and assessed for consistency and business reasonableness.</li> </ul>
	VI Explainability and interpretability 	<ul style="list-style-type: none"> <li>To make models <b>understandable and interpretable</b>.</li> </ul>	<ul style="list-style-type: none"> <li>✓ <b>Tentative industry standards</b> start to emerge, including LIME, SHAP and surrogate models.</li> <li>✓ Complemented with <b>large-scale sensitivity</b> and <b>selected case-by-case</b> scenario analyses.</li> </ul>
	VII Overfitting and bias	<ul style="list-style-type: none"> <li>To assess whether a complex ML model is <b>overfitted</b>, or if it may lead to <b>discriminating behaviour</b> due to bias in the input data or in the algorithm.</li> </ul>	<ul style="list-style-type: none"> <li>✓ Regularization techniques, k-fold cross-validation and a <b>strict control of the learning curve</b>.</li> <li>✓ <b>Individual case-by-case analysis</b> on a sample of representative data.</li> </ul>
Imp & Doc	VIII Implementation	<ul style="list-style-type: none"> <li>In the case of ML algorithms (complex structure, large sets of parameters, a dedicated platform with proprietary code libraries, and pose computational challenges) – How to validate that the model is <b>properly implemented</b>?</li> </ul>	<ul style="list-style-type: none"> <li>✓ Construct a <b>large dataset (some millions of observations)</b> real or synthetic), and ask both the developers and the implementers to score them; and check that the outcomes exactly match</li> <li>✓ <b>Periodically monitoring</b> if they continue to be fit-for-purpose (not only performance).</li> </ul>
	IX Documentation 	<ul style="list-style-type: none"> <li>Is there is a particular increase of attention to model documentation – <b>what level of depth</b> should this documentation reach for ML models?</li> <li>Example: consider an external code library (with the code for a random forest as sufficient doc), or <b>detail every step</b> in the random forest training algorithm?</li> </ul>	<ul style="list-style-type: none"> <li>✓ The model documentation should allow the <b>replication of the model by a third party</b>.</li> <li>✓ In general terms, the model documentation should ensure that <b>clarity, interpretability and traceability</b> of the models are up to the highest standards.</li> </ul>





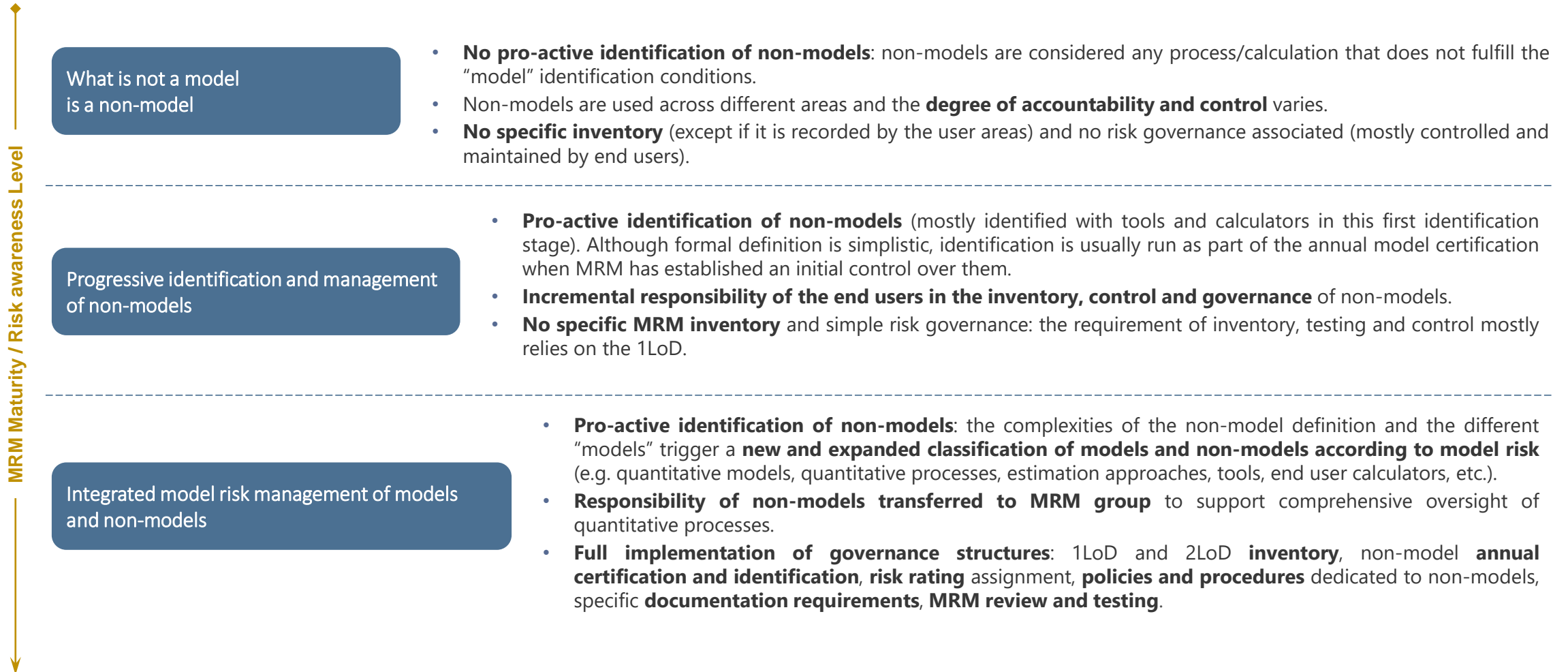
---

# Model risk beyond models: non-models

---



The identification and governance of non-models within the model risk realm substantially varies across entities, highly dependent on the degree of maturity of the function and the risk awareness created



# The complexity of (non-)model definition

The cornerstone of this evolution towards a comprehensive and integrated oversight is the definition of what a non-model is, based on their independent nature and characteristics

## A Model Is ...

Quantitative process that produces an uncertain rating / prediction / valuation / levels or segments

### Examples

- Any type of regression: multivariate / logistic / non-linear
- Machine Learning algorithms
- Quantitative calculations involving complex statistical, economic or financial theories
- Rules / detection algorithms



## A Non-Model Is

Quantitative process that produces a certain / factual / deterministic output

### Examples

- Expert judgment based on deterministic calculations (aggregation of information)
- Transformed financial or economic ratios



## (Conflicting) Examples

- Summary statistics or simple statistical concepts (interpolation)
- Look-up tables based on ratios resulting from basic statistical/empirical analysis or aggregation calculations
- Business / regulatory rules and calculation ratios



## In practice: what to do



### Be precise in the identification

Differentiation of models and non-model requires precision in the identification:

- Examples: LGD models based on look-up tables with two variables: leverage and asset valuation, or, Capital regulatory ratios and calculations. These models could be re-classified as non-models/tools or as non-complex models.

Critical elements to identify non-models extensively used:

- Certainty of the output.
- End user responsibility of the design and implementation (e.g. desktop).
- Complexity of assumptions involved.

### Be commensurate to risk

- Define criticality levels of non-models or tools to reinforce governance and review (and not jeopardize MRM function).
- Documentation template adapted to non-models critical elements: design, control, implementation, testing.





---

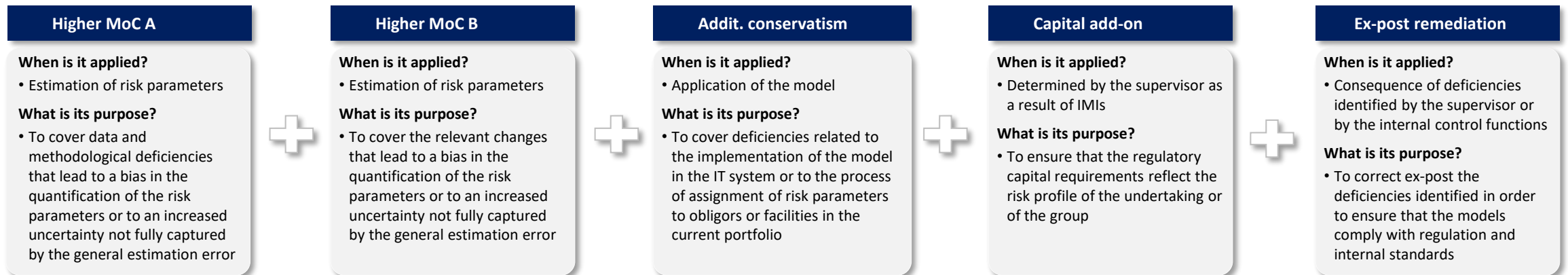
# MRM as a source of efficiency and cost reduction

---

# Cost of model-related weaknesses

**Focusing on the weaknesses affecting IRB models, these ones result in cost of conservatism, capital add-ons and ex-post remediation measures: this cost could be cut with an ex-ante program that leverages on the MRM framework and reinforces its effectiveness**

## Weaknesses affecting the IRB models: cost of conservatism and capital add-ons



### What is the cost for institutions? Case study

According to this case study:

- The common range of the %RWA (= RWA/EAD) increase due to the factors mentioned above is **(3.5; 6)** percentage points. Depending on the maturity / ambition level of the MRM framework implemented by the institution, the increase will be closer to the lower or upper limit of the range
- Implementing a comprehensive ex-ante program (as opposed to ex-post remediation measures) that leverages on the MRM framework and reinforces its effectiveness addressing the model-related weaknesses in aspects covered by overarching regulations would reduce the range of the RW increase to **(1; 3.5)** percentage points



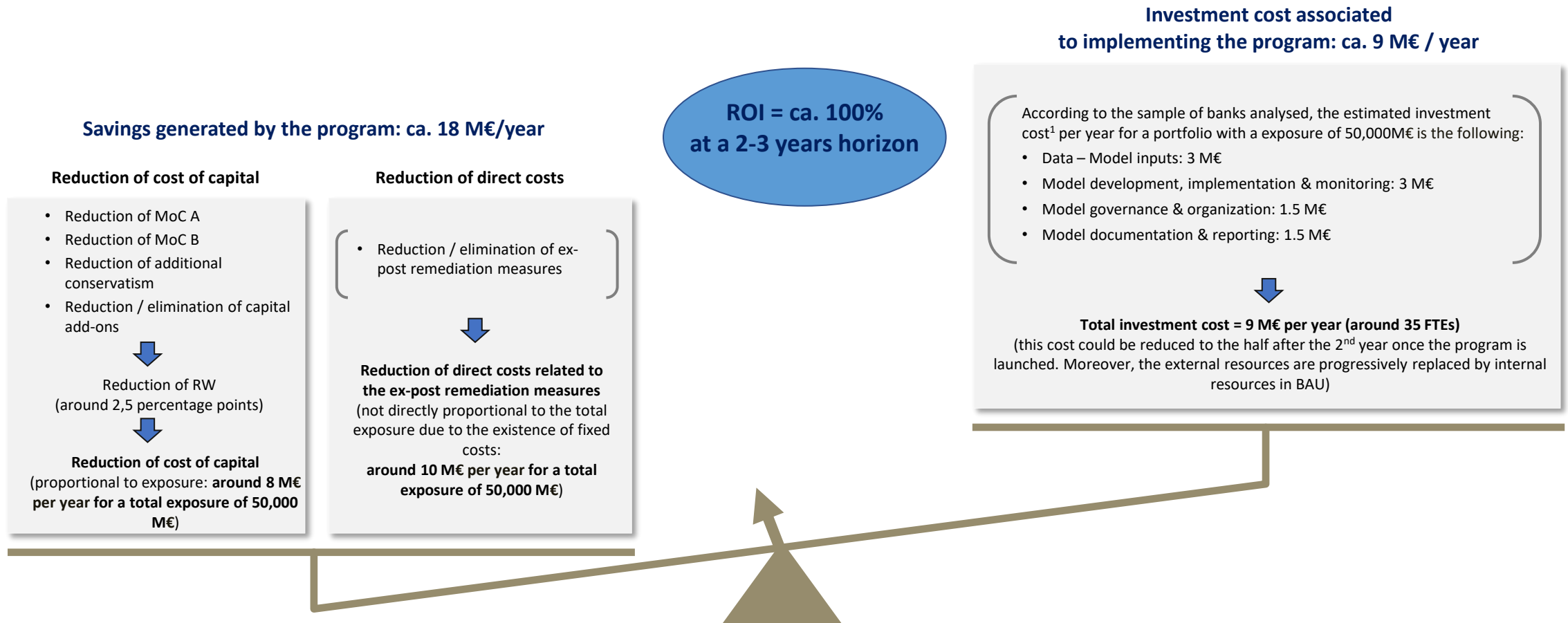
The %RWA is reduced on average 2,5 percentage points. Considering a total portfolio of 50,000M€ and a cost of capital equal to 8%, this extra 2,5% %RWA has a yearly cost for the bank equal to:  $\text{Cost} = 50.000 \times 2.5\% \times 8\% \times 8\% = \mathbf{8 \text{ M€}}$

- On the other hand, implementing this program would reduce / eliminate the ex-post remediation measures, which according to the sample analyzed would lower the direct costs in **ca. 10 M€ / year** for a total portfolio of 50.000 M€.



# Cost-benefit analysis

The implementation of a comprehensive ex-ante program that leverages on the MRM framework and reinforces its effectiveness is clearly profitable: the savings in costs of capital and ex-post remediation measures outweigh by far the investment cost



(1) Investment needed to apply the overarching measures that reinforce the MRM framework and mitigate findings and weaknesses such as the ones presented



---

## Final remarks

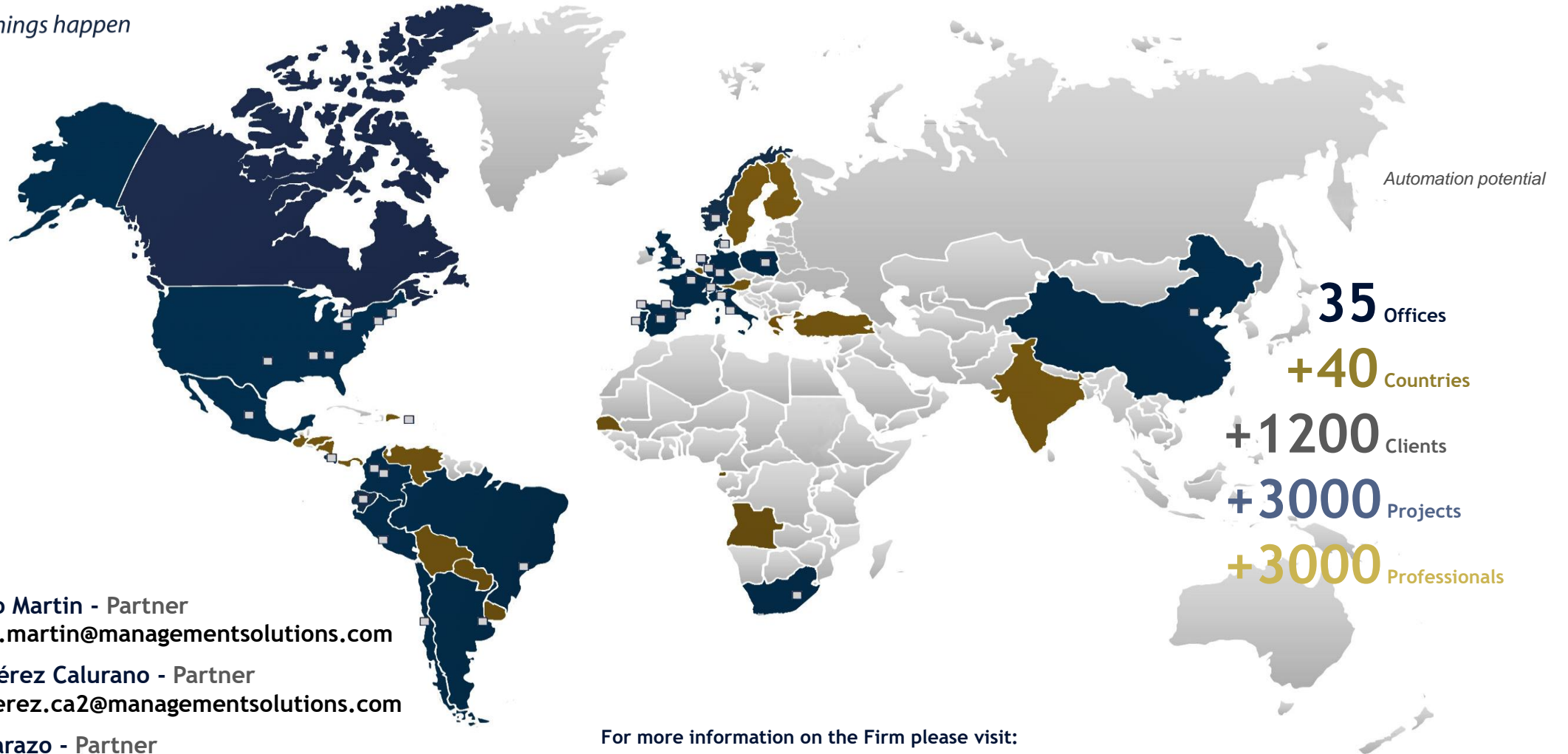
---

**Banks with an effective MRM practice rely on a number of key aspects to be at the forefront and ensure internal compliance with regulation, supervision and internal policies**

## Some tips

- 1** **Maturity:** there is still a long way to go, even for European front-runners and for major US G-SIBs.
- 2** **Stay close to supervisors and regulation updates:** there is still scarce regulation on MRM in Europe although the ECB commonly raises MRM-related findings covered by overarching (not MRM-specific) regulations.
- 3** **Efficiency as the next goal for advanced institutions:** only very mature MRM approaches are envisaging MRM as a source of synergies and effective reduction of costs, capital and provisions.
- 4** **Data and documentation as the first step towards efficiency:** this is the greatest need and what will contribute most to efficiency. Major institutions are already investing heavily on this, especially on the data side.
- 5** **Need to spread a model risk culture across the 3LoDs:** this implies raising awareness in senior management, gearing up with the right skillset, increasing commitment, involving model users and empowering model owners.
- 6** **Automation and machine learning have high potential:** advanced entities are incorporating these techniques, though not free from model risk-related challenges, to foster efficiency.

*We create value propositions for our clients and commit to their effective implementation*



#### CONTACT

Javier Calvo Martin - Partner  
[javier.calvo.martin@managementsolutions.com](mailto:javier.calvo.martin@managementsolutions.com)

Francisco Pérez Calurano - Partner  
[francisco.perez.ca2@managementsolutions.com](mailto:francisco.perez.ca2@managementsolutions.com)

José Luis Carazo - Partner  
[jose.luis.carazo@managementsolutions.com](mailto:jose.luis.carazo@managementsolutions.com)

For more information on the Firm please visit:

[www.managementsolutions.com](http://www.managementsolutions.com)

Or at our social network channels





# Annex: Component Modelling and AutoML

## A new modelling strategy

Both in academia and research, as well as in the financial industry but in a more emerging manner, the same trend is underlying: a new modelling strategy

### Context

- **Increased modelling needs** (more volume and complexity of models) with **unorganized and inefficient growth**.
- **Increased costs** (resources, software, IT infrastructure) and **shortage of qualified resources**.
- In this context, a **new trend** is emerging: **component modelling** and **AutoML** on new IT and SW infrastructures.

### Modelling Components

#### Level 0: Components (workflow fully editable)

- They must be applied regardless of the use case (atoms): missings, outliers, feature selection, sampling, ...

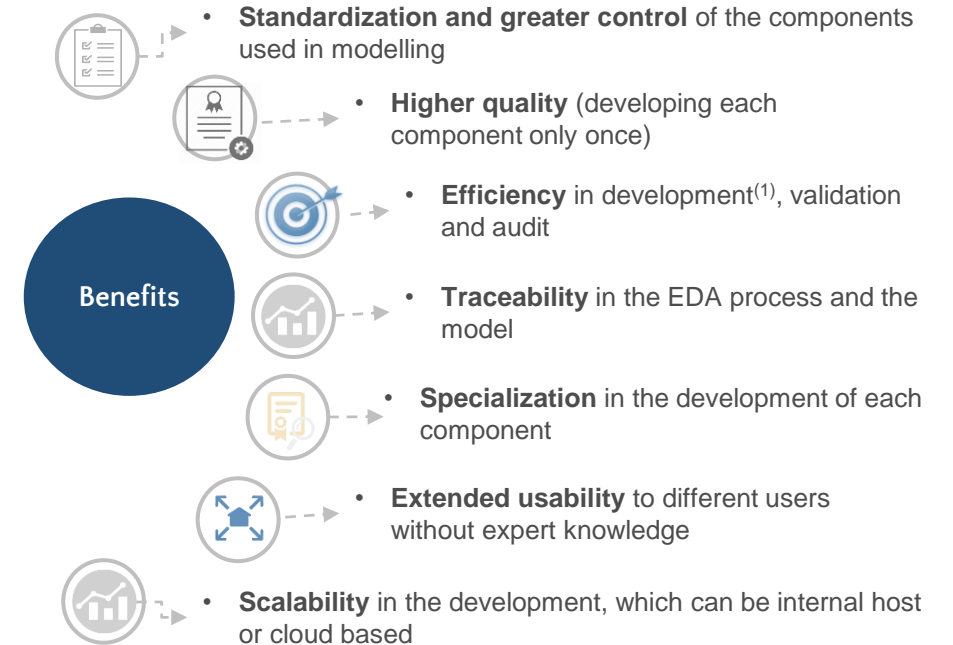
#### Level 1: Methodologies (workflow partially designed)

- Parameterizable combination of level 0 components: data processing, optimization of hyperparameters, ...

#### Level 2: AutoML (E2E automation)

- End-to-end modelling process configured with atoms and methodologies built with the previous levels.

### Benefits



### Use cases

- **Semi-automatic** model development
- **Challenger models** development (IV)
- Full **automation** of model **monitoring**
- **Automatic** generation of **reports**,...

**ModelCraft™ is a component modelling solution developed in a cloud environment whose main elements are data management, algorithm generation, reporting and interpretability, all in a customizable, parameterizable and scalable approach**

The screenshot shows the ModelCraft console with a SQL query being executed. The query is a CREATE TABLE statement followed by a SELECT statement that joins several tables: mexico\_financial, mexico\_behaviour, mexico\_personal, and mexico\_others. The results are displayed in a table with columns: new\_key, ISP\_D6INVN, ITMAXAUT, ITMAXNOM, ITRECNO, IMORAMAXNOM, ISALDOACTNOM, EPOR\_ENDEUDAMIENTO, INGRESO\_INFERIDO, and CUENTAS\_ABIERTAS\_OTRO. The first five rows of data are visible.

	new_key	ISP_D6INVN	ITMAXAUT	ITMAXNOM	ITRECNO	IMORAMAXNOM	ISALDOACTNOM	EPOR_ENDEUDAMIENTO	INGRESO_INFERIDO	CUENTAS_ABIERTAS_OTRO
1	1	162	3231	11	11	0	18148	100		1.0
2	2	12471	4220	-1	-1	0	0	0		0.0
3	3	0	0	-1	-1	0	0	0		0.0
4	4	0	0	-1	-1	0	0	0		0.0
5	5	4483	4260	35	35	0	0	0		2.0

### Components' Customization

- ✓ Inclusion of pre-defined standard level 0 components and also customized on-demand components
- ✓ Definition of level 1 and 2 components by users

### Reporting & Interpretability

- ✓ Automatic report generation
- ✓ Generation of functional and technical model interpretability reports

### Data Management

- ✓ DB Management
- ✓ Simultaneous connection with different data repositories in different languages (PostgreSQL, MariaDB, MySQL, Oracle, Terminal, ...)

### EDA & Algorithms

- ✓ Data processing and EDA
- ✓ Traditional techniques (e.g. logistic reg., trees, ARIMAX...)
- ✓ Advanced ML algorithms (random forest, boosting, SVM...)

### Scalability (Cloud)

- ✓ Its cloud architecture allows for auto-scaling in computing
- ✓ All logic is stored in Kubernetes containers allowing an easy and fast deployment

The screenshot shows the ModelCraft component modelling interface. It features a central canvas with a workflow diagram. The workflow starts with a 'Table Reader' component, followed by a 'Train Test' split, then 'Missing' and 'Drop Correlated' components. The data then flows into a 'Train' component, which branches into 'Cross Validation', 'Metrics', 'Predict', and 'Model Download'. The 'Predict' component leads to a 'CSV Download' component. The interface includes a sidebar with component categories like Input, AutoML, and Sampling, and a right-hand panel for component properties and logs.

Three case studies to illustrate how machine learning models and automation can be very successful...,  
... or not

### » Exploratory data analysis using ML



#### Case study

Usage of a XG Boost and an F-Race algorithm for a credit scoring model in one European bank:

- In a behavioural credit scoring model used for regulatory (IRB) purposes, the bank used a combination of an **XG Boost and an F-Race algorithm** for the initial variable selection process and the variable binning process.
- The algorithms were reviewed and approved by IV, though after many iterations due to the difficulty in replicating them. Nevertheless, in the words of the IV team, the usage of these machine-learning algorithms contributed to **accelerate and optimize these processes**.



### » Automated model generation



#### Case study

Usage of a stress testing time-series based (ARIMAX) model for validation in one European bank:

- The model used the expected macroeconomic regressors, had reasonable predictive power (R2), and complied with all the expected tests.
- IV used an **automated Python-based time-series development tool** to produce 2,000 challenger models based on the same data.
- The tool used **evolutionary (genetic) algorithms** to converge and produce only the models with the highest chances of suitability.
- As a result, IV found that **there was room for improvement in the predictive power** and there were combinations of regressors with business sense that would help to produce a significantly better model.



### » More complex is not always better



#### Case study

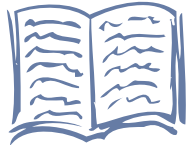
Usage of a complex perceptron neural network for a shadow-rating model in one European bank:

- The bank developed and submitted to IV a rather **complex perceptron neural network** with one hidden layer for a shadow-rating model.
- The IV team in turn tried a challenger model which was simply a multinomial logistic regression based on a **monotonic Box-Cox transformation** that linearized the variables.
- This model **roughly matched the predictive power** of the neural network, so the logistic regression was finally preferred.



### Three case studies to illustrate the degree of implementation of non-models' governance and risk management across entities

#### » Enhanced risk management



##### Case study

Full transference to MRM of non-models oversight and strong 1LoD governance and control

- **Full identification of models vs non-models:** different types of models identified depending on the complexity of methodologies and assumptions, and **non-models identified as end user tools and calculators.**
- **Transference of oversight to MRM** (independent identification questionnaires for models and non-models, integrated inventory, dedicated policies and templates) and reinforced governance by 1LoD (ownership/accountability identification, inventory, testing, self-identification and report).
- **Dedicated resources within the MRM team** with focus on non-models review and testing.

#### » Inception of active risk mgmt.



##### Case study

Initial oversight by MRM of non-models (vs models)

- **Full identification of models vs non-models with a simple approach** (focus on the nature of the output, quantitative estimates / forecast vs certain / factual / deterministic output).
- **Partial transference of oversight to MRM** (independent identification questionnaires for models /non-models, certification campaign run with the model certification stream, preliminary inventory, integration in MRM policy).
- **Resources in MRM are not specialized.** Monitoring and review is focused on **non-models that were reclassified from models.**

#### » Ad-hoc risk management



##### Case study

Re-classification of certain "models" due to specific program configurations

- **No active identification or management of non-models** from model risk management perspective. Control remains within 1LoD and mainly, within IT.
- **Specific identification of quantitative processes** (they may be considered models or non-models if the policy evolves) in the **suite of specific model families** (e.g. for stress test exercises) as per the requirements of the program: identification exclusive within the group of those processes and models, **dedicated template for model documentation**, included in annual reviews.