Characterization and modelling data workflows for low carbon cement optimization

EMMC 2025, satellite workshop, April 7th, TU-Wien Vienna

Geoffrey Daniel (CEA Paris-Saclay) , Alexandre Ouzia (HeidelbergMaterials) and Sophie Schmid (TU-WIEN-IMWS)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement Nº 101091687.

Outline

MatCHMaker
Materials Characterisation & Modelling

- The EU-Project MatCHMaker
- The Use Case: Low carbon cement optimization
- Overall workflow
- Workflow on strength prediction (Sophie)
- Workflow on image analysis (Geoffrey)
- Data Lifecycle reflections

Outline

MatCHMaker
Materials Characterisation & Modelling

- The EU-Project MatCHMaker
- The Use Case: Low carbon cement optimization
- Overall workflow
- Workflow on strength prediction (Sophie)
- Workflow on image analysis (Geoffrey)
- Data Lifecycle reflections

MatCHMaker: Goals & Deliverables

MatCHMaker
Materials Characterisation & Modelling

Goals:

- to support the European industrial manufacturing with advanced characterisation methods and computational modelling, in order
- to lead the way to the reliable design of new and sustainable materials and processes, rapid upscaling, and effective quality control (QC).

	1. Develop a model-based innovation process (workflow) to accelerate advanced (multiscale and multiphases) materials design and validation including novel characterization (material properties / performance), modelling (data and physics based, engineering and numerical modelling) approaches and combination of theory with large-scale computational screening
	2. Reinforce traceability, integrity and interoperability of C&M data and workflows through provenance facilitation and a semantic approach using Elementary Multiperspective Material Ontology (EMMO) connected with CHADA and MODA methods
	3. From the ontological approach designed in the objective 2, propose an OPEN data REPOSITORY integrating intercorrelated concepts and data coming from the Workflow designed in the objective 1.

WP 2 and 3 overview

MatCHMaker
Materials Characterisation & Modelling

Goals: Enhance, automate and speed-up image analysis of electron microscopy images

Motivations:

- The Process-Microstructure-Properties paradigm & the importance of microscopy
- Many tasks are simple but lengthy and require expert knowledge in image analysis
- The potential of Machine learning for microscopy image an

Materials & Methods:

- Cement and concrete microstructures (2D BSE/EDX/Fib-SE
- SOEC / SOFC électrodes (3D – 2D transition)
- C/Pt PEMFC électrodes (tomography, PSSD)

Stakeholders:

- Academic: CEA (Paris, Grenoble), SINTEF (Norway), ASRO &
- Industrial: Heidelberg Materials, Toyota, Genvia

Deliverables: papers, a website with codes & cases studies

Timeline: 2022-2026

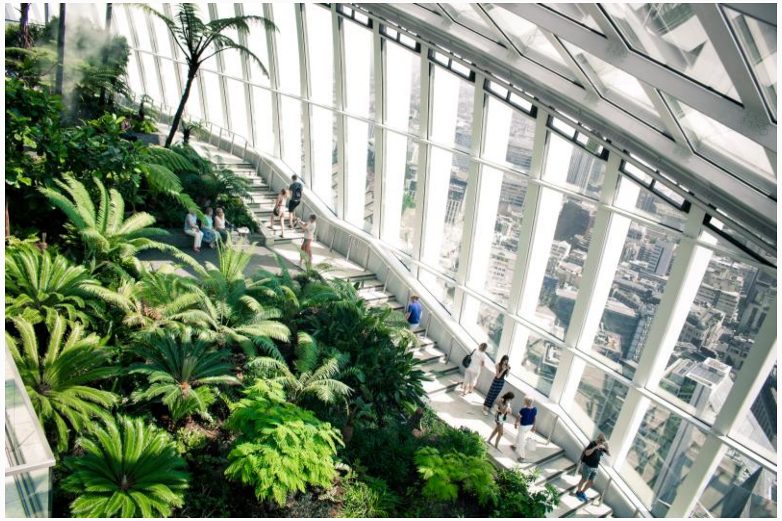
Outline

- The EU-Project MatCHMaker
- The Use Case: Low carbon cement optimization
- Overall workflow
- Workflow on strength prediction (Sophie)
- Workflow on image analysis (Geoffrey)
- Data Lifecycle reflections

Outline

- The EU-Project MatCHMaker
- The Use Case: Low carbon cement optimization
- Overall workflow
- Workflow on strength prediction (Sophie)
- Workflow on image analysis (Geoffrey)
- Data Lifecycle reflections

The global building materials industry faces major challenges



Climate change

In order to achieve the goals of the Paris Agreement for climate protection, the building materials industry needs to focus on decarbonisation.

Limited natural resources

For a more sustainable use of natural resources, industry must use fewer primary raw materials and rely more on recycling, for example.

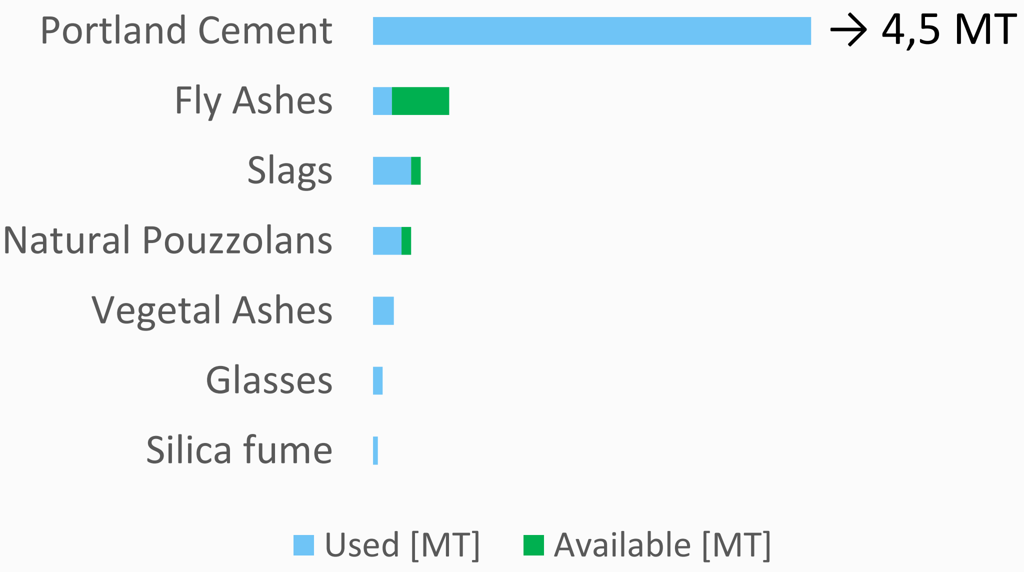
Digitalisation

In production and on construction sites, digital solutions are needed that make processes simpler, faster, safer, more sustainable and efficient.

Context: Rising CO2 prices and decreasing SCM

-

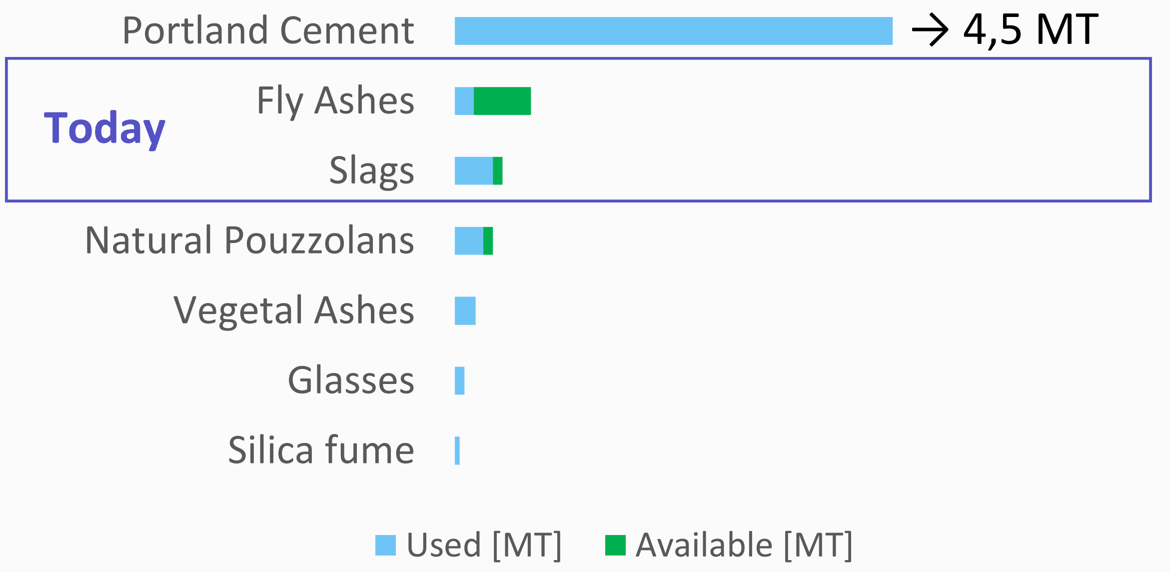
EU Carbon Permit price (06/12/2022: 87 €)



Context: Rising CO2 prices and decreasing SCM

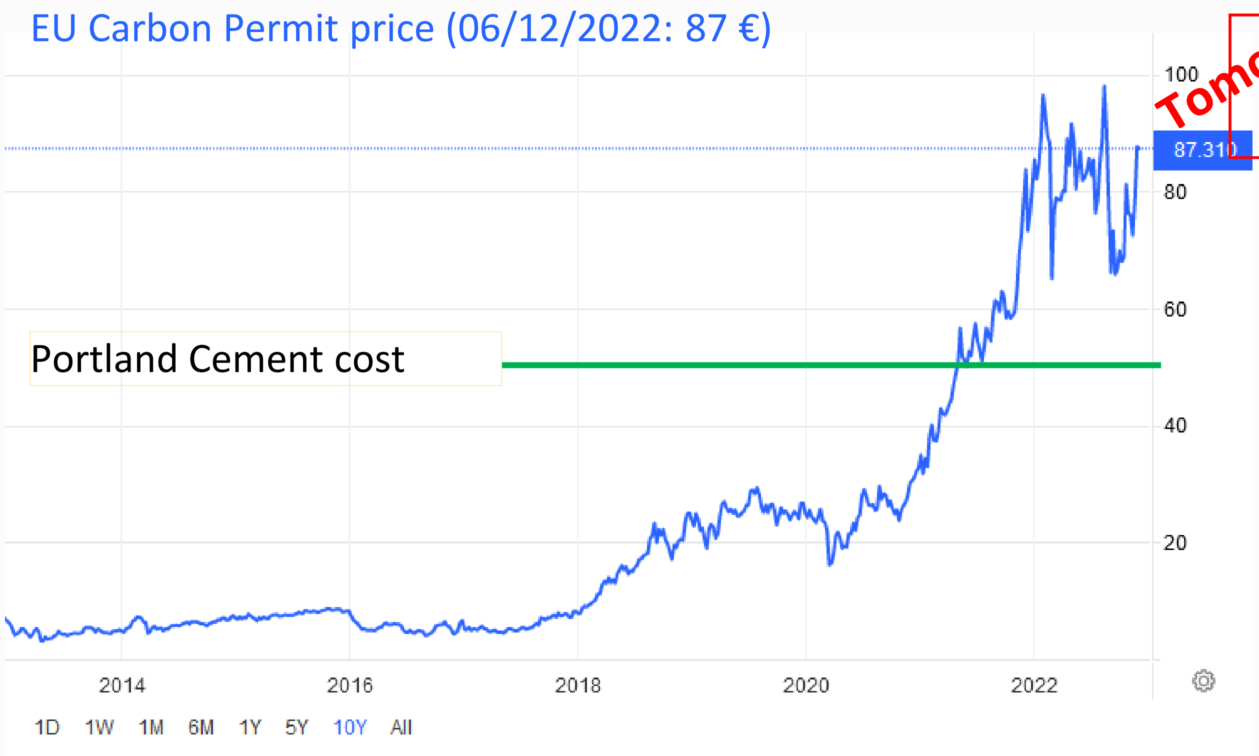
•

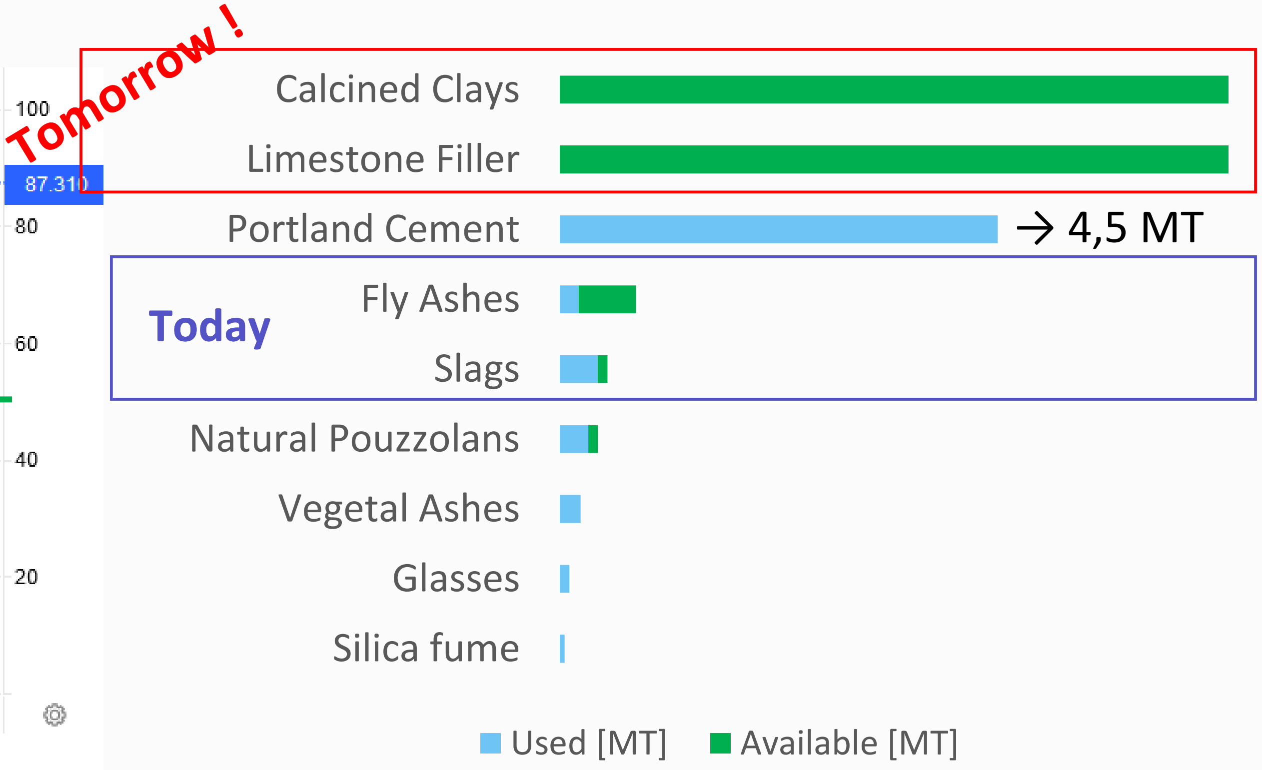
EU Carbon Permit price (06/12/2022: 87 €)



Context: Rising CO2 prices and decreasing SCM

•



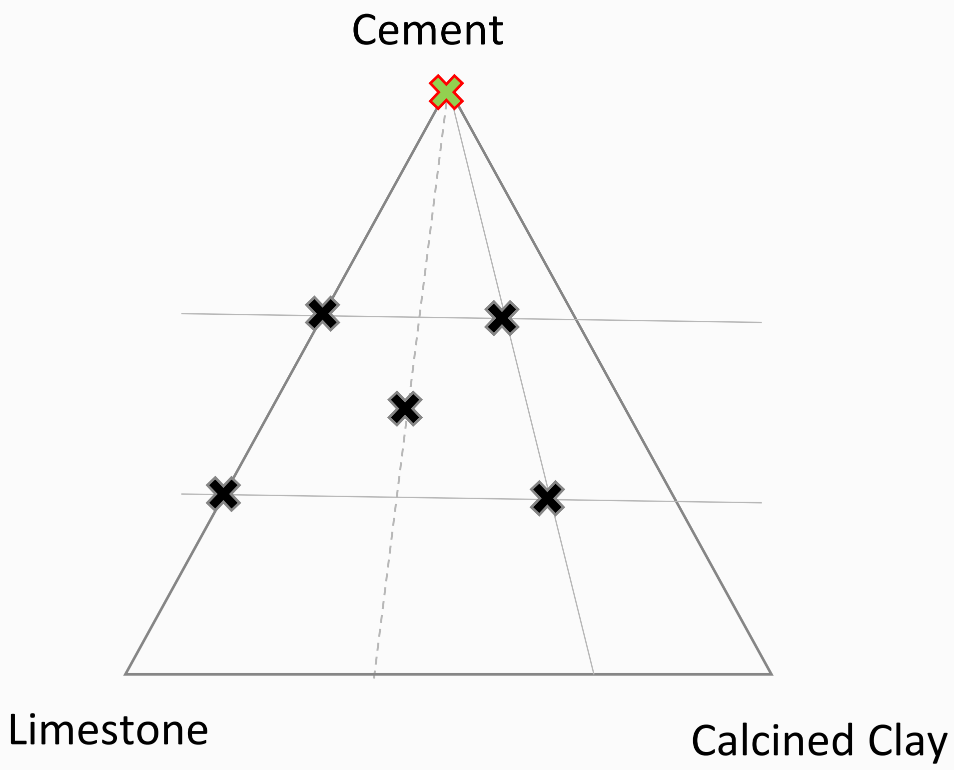


Goals and scope of the UC cement

- Goals:

- The key *applied* goals are the optimization of:
  - The mechanical (2,7 and 28 days) strength
- The key *fundamental* goals are to understand:
  - What limits the reaction rate
  - What limits the pore refinement

- Scope: Limestone – Calcined Clay - Cement



Outline

- The EU-Project MatCHMaker
- The Use Case: Low carbon cement optimization
- Overall workflow
- Workflow on strength prediction (Sophie)
- Workflow on image analysis (Geoffrey)
- Data Lifecycle reflections

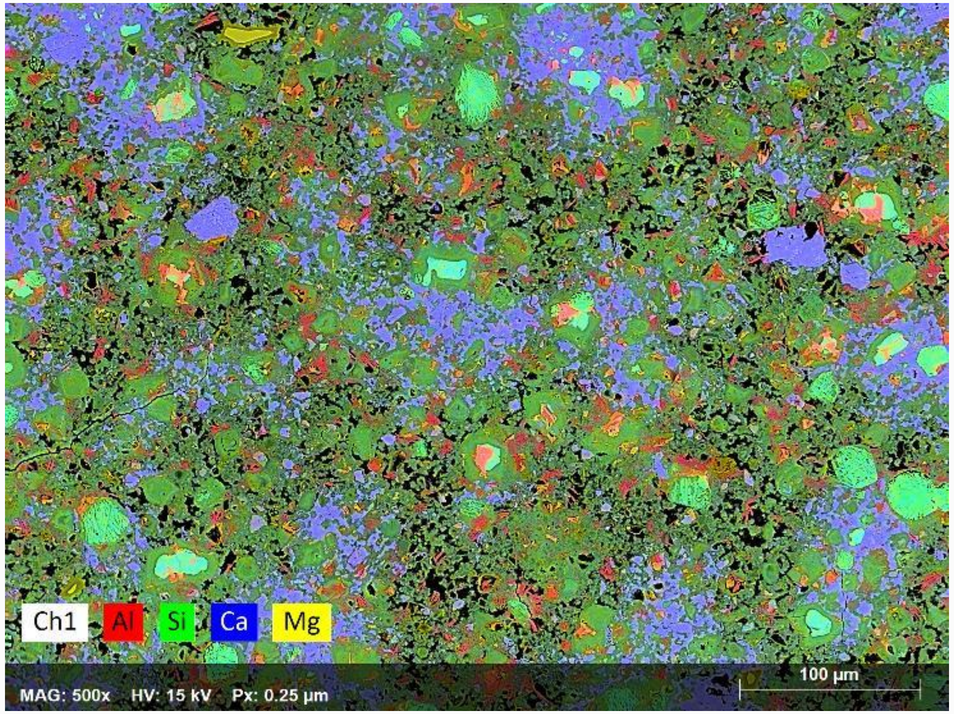
Outline

- The EU-Project MatCHMaker
- The Use Case: Low carbon cement optimization
- Overall workflow
- Workflow on strength prediction (Sophie)
- Workflow on image analysis (Geoffrey)
- Data Lifecycle reflections

The microstructure is key to understand the properties

Motivations: all the macroscopic properties derive from the microstructure. Hence the need to characterize it.

Input image



EU-Project MatCHMaker consortium

Overall Workflow

MatCHMaker
Materials Characterisation & Modelling

Heidelberg Materials

- Define design of experiments
- Prepare samples

SINTEF

- Acquire images (SEM / TEM)

cea PARIS-SACLAY

- From image compute phase assemblages

TU WIEN

- From phase assemblages compute strength

Outline

- The EU-Project MatCHMaker
- The Use Case: Low carbon cement optimization
- Overall workflow
- Workflow on image analysis (Geoffrey)
- Workflow on strength prediction (Sophie)
- Workflow on Data Lifecycle reflections

Outline

MatCHMaker
Materials Characterisation & Modelling

- The EU-Project MatCHMaker
- The Use Case: Low carbon cement optimization
- Overall workflow
- Workflow on image analysis (Geoffrey)
- Workflow on strength prediction (Sophie)
- Workflow on Data Lifecycle reflections

Image analysis workflow

Phases identification on SEM/TEM images

MatCHMaker
Materials Characterisation & Modelling

SEM acquisition

SINTEF HM

EDX: CSV, TXT

BSE: JPEG, TIFF, PNG

Data management

INPUT DATA FORMATTING

Stack of images $X \in \mathbb{R}^{w \times h \times c}$
$w \times h$: image dimensions (number of pixels)
c: number of channels (EDX + BSE)

Dimension processing

NPY

PIXEL MONO-PHASE SEPARATION

Clustering algorithm: Gaussian mixture models

Expert knowledge

HM

NPY

HYDRATES UNMIXING

Lasso regression, MLEM

Expert knowledge

HM

OUTPUT RESULTS

Phase assemblage (XLSX)

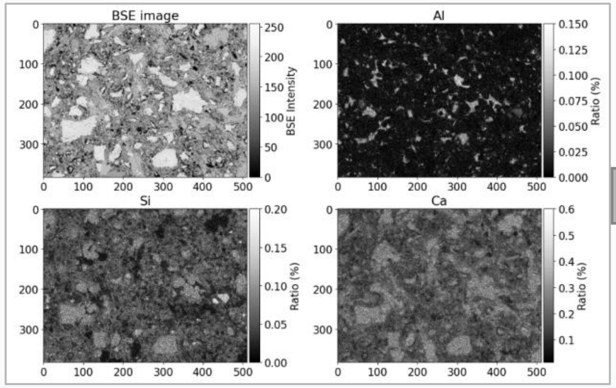
Images of the phases (JPEG)

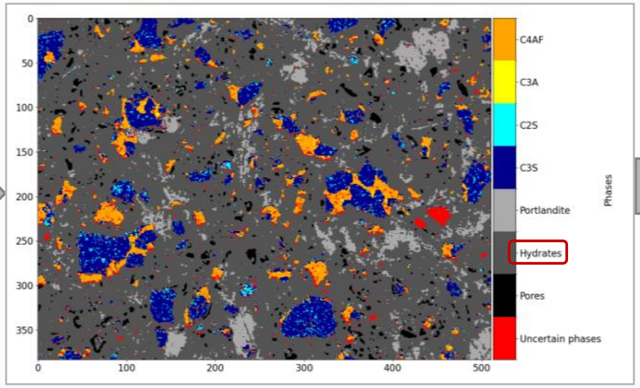
TU-WIEN HM

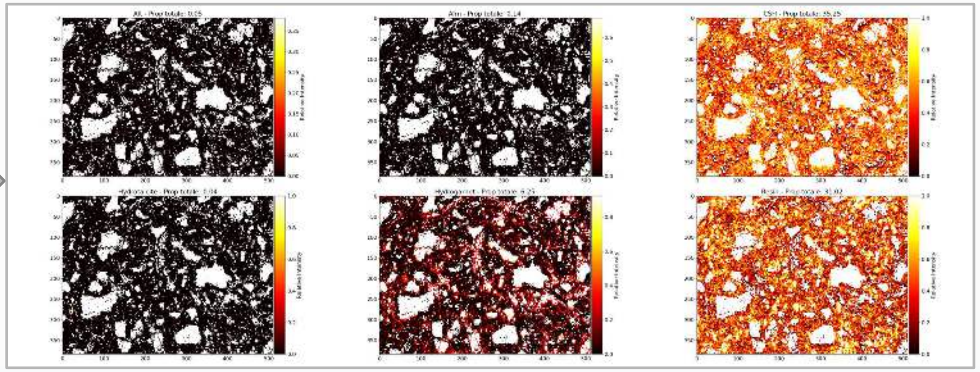
Exploitation for modelling

No ground truth: **Unsupervised ML**

Identification of pixels with similar composition







Outline

- The EU-Project MatCHMaker
- The Use Case: Low carbon cement optimization
- Overall workflow
- Workflow on image analysis (Geoffrey)
- Workflow on strength prediction (Sophie)
- Data Lifecycle reflections

Outline

- The EU-Project MatCHMaker
- The Use Case: Low carbon cement optimization
- Overall workflow
- Workflow on image analysis (Geoffrey)
- Workflow on strength prediction (Sophie)
- Data Lifecycle reflections

Physics-based modelling workflow

MatCHMaker
Materials Characterisation & Modelling

Multiscale Strength Modelling based on scale-transition methods of continuum micromechanics

Raw materials (HM)

Sample production, strength acquisition (TU Wien-IMWS)

samples

XRD, SEM acquisition (HM)

Image Analysis (CEA)

phase assemblage (SEM)

for TU Wien-IMWS

phase assemblage (XRD), SEM images

Numerical data — Input

- Phase assemblage (scalar*)
- Phase stiffness (scalar*)
- Ultimate strength of hydrate gel needle

Expert knowledge — Input

- Characteristic shape and inter-action of material phases
- Hierarchical organization

quantitative

Continuum micromechanics (scalar-tensor-scalar)

- Eshelby problems

$$\boldsymbol{\varepsilon}_i = \left[\mathbb{I} + \mathbb{P}_i^{\infty} : (\mathbb{C}_i - \mathbb{C}_\infty)\right]^{-1} : \mathbf{E}_\infty$$

- Scale transition

$$\mathbb{A}_i = \left[\mathbb{I} + \mathbb{P}_i^{\infty} : (\mathbb{C}_i - \mathbb{C}_\infty)\right]^{-1}$$

$$\left[\sum_{j=1}^{n} f_j \left[\mathbb{I} + \mathbb{P}_j^{\infty} : (\mathbb{C}_j - \mathbb{C}_\infty)\right]^{-1}\right]^{-1}$$

- Multiscale Analysis: from tensor to scalar*

stiffness: $\mathbb{C}_{cp} = \sum_{i=1}^{n} f_i \, \mathbb{C}_i : \mathbb{A}_i \longrightarrow E$

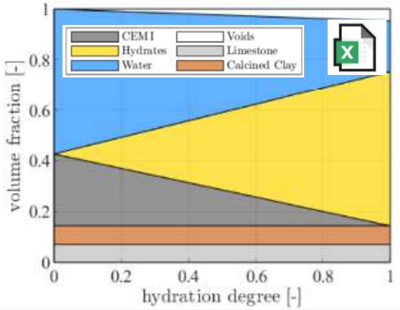
uniaxial compressive strength:

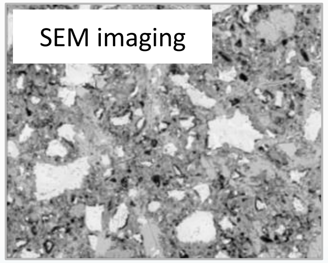
$$\bar{\sigma}^{vol}_{hyd;\varphi,\vartheta} = \sqrt{\lim_{f^{cp}_{hyd;\varphi,\vartheta}\to 0} \left(\frac{3\,k^2_{hyd}}{f^{cp}_{hyd;\varphi,\vartheta}}\, \mathbf{E}_{cp} : \frac{\partial \mathbb{C}_{cp}}{\partial k_{hyd;\varphi,\vartheta}} : \mathbf{E}_{cp}\right)}$$

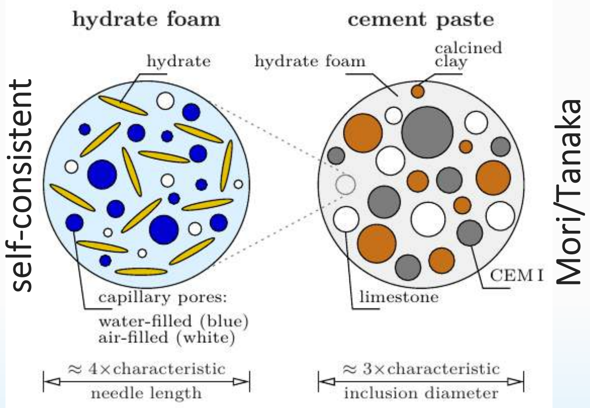
$$\bar{\sigma}^{dev}_{hyd;\varphi,\vartheta} = \sqrt{\lim_{f^{cp}_{hyd;\varphi,\vartheta}\to 0} \left(\frac{2\,\mu^2_{hyd}}{f^{cp}_{hyd;\varphi,\vartheta}}\, \mathbf{E}_{cp} : \frac{\partial \mathbb{C}_{cp}}{\partial \mu_{hyd;\varphi,\vartheta}} : \mathbf{E}_{cp}\right)}$$

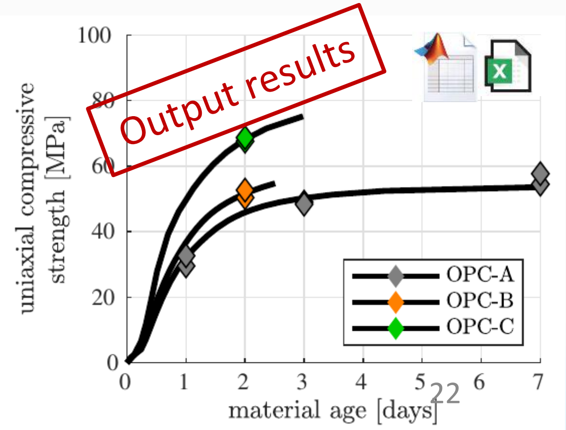
qualitative

Material organograms









* for different material ages

Outline

MatCHMaker
Materials Characterisation & Modelling

- The EU-Project MatCHMaker
- The Use Case: Low carbon cement optimization
- Overall workflow
- Workflow on image analysis (Geoffrey)
- Workflow on strength prediction (Sophie)
- Data Lifecycle reflections

Outline

MatCHMaker
Materials Characterisation & Modelling

- The EU-Project MatCHMaker
- The Use Case: Low carbon cement optimization
- Overall workflow
- Workflow on image analysis (Geoffrey)
- Workflow on strength prediction (Sophie)
- Data Lifecycle reflections

Overall Workflow

MatCHMaker
Materials Characterisation & Modelling

- Define design of experiments
- Prepare samples

- Acquire images (SEM / TEM)

Unpredictability of the problem solving process

Different experts need to sit every week at the same table to solve the problem together

Flows: of data, materials, and knowledge !

Data

Materials

Knowledge

- From image compute phase assemblages

- From phase assemblages compute strength

Data lifecycle: theory VS reality

- Our experience
  - Solving the scientific problem comes first
  - As the formulation of the problem evolve:
    - so does the vocabulary to describe it
    - the exact experimental data required also (iteration of acquisitions, most previous versions being discarded)
  - To formulate the problem clearly, a core team of experts in the different fields have to sit down for long hours every week to learn from one another and build the tools
  - Data life cycles tools only makes sense once there is a clear understanding and stable vocabulary

This widespread illustration is a fiction

Conclusions

- Challenges
  - Data diversity: different formats, instruments (brand and protocols), materials
  - Data dust: most data is actually discarded
  - Problem come first: experts needs to sit weekly at the same table and iterate on solving the problem
  - Expert knowledge is unavoidable:
    - No ground truth
    - Data scarcity
  - Data life cycle is by far not linear
- Best practices
  - Data management checkpoints once the problem is well understood
  - Scaling up the data management tools progressively
  - Team coordination

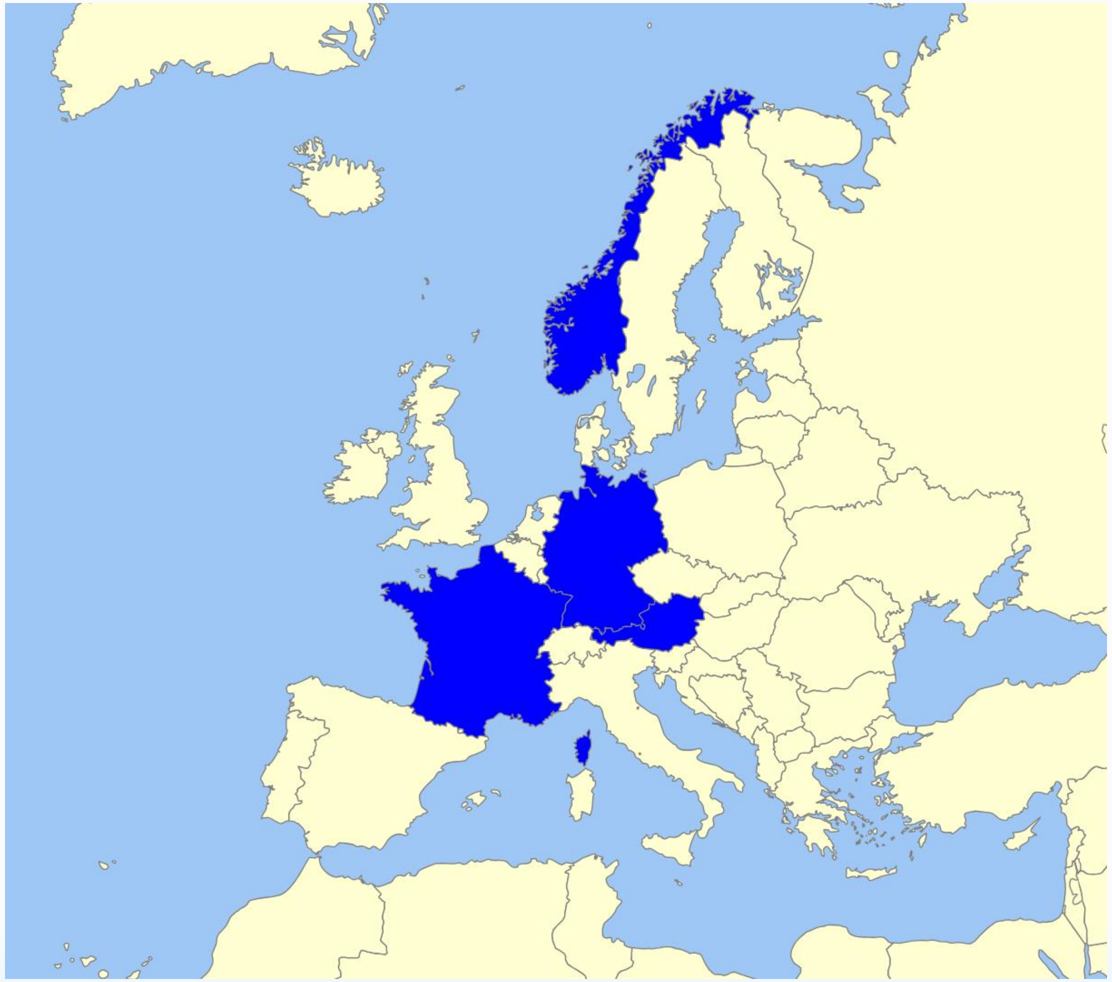
Thank you!

MatCHMaker

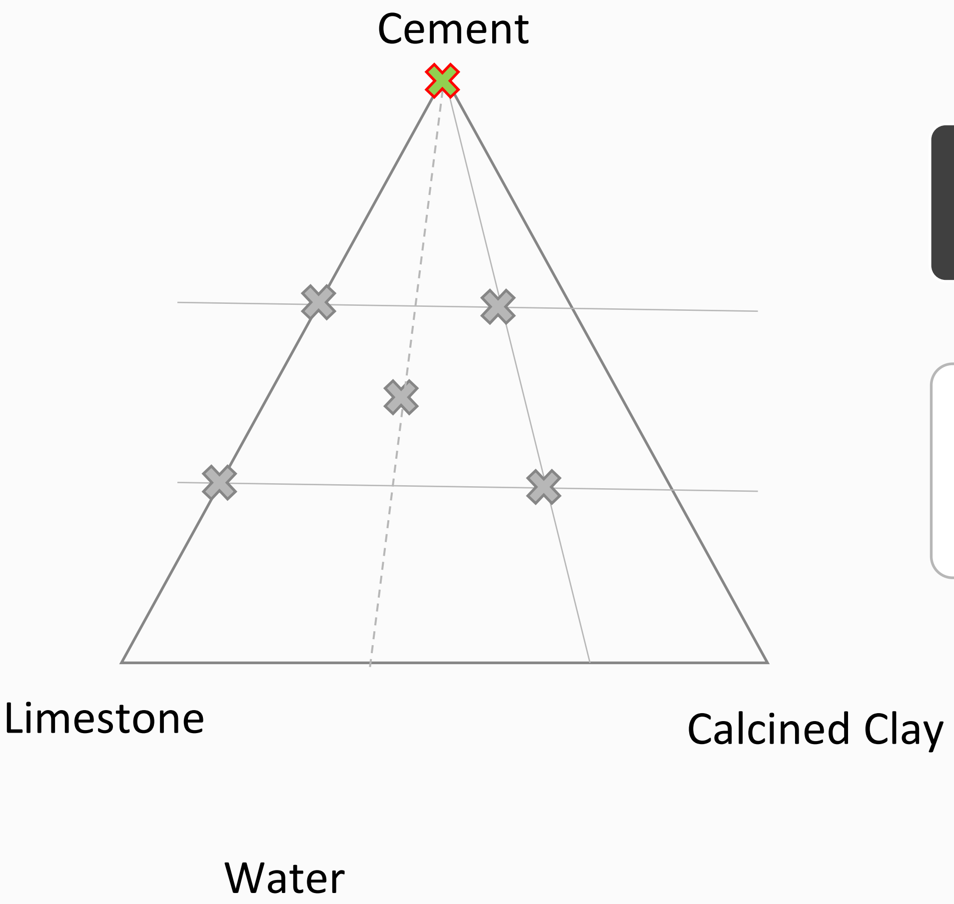
Materials Characterisation & Modelling

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement № 101091687.

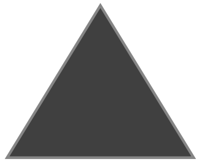
Overall Workflow



Approach: the Material Science paradigm and Design of experiments



Mix design → Microstructural features → Macroscopic properties



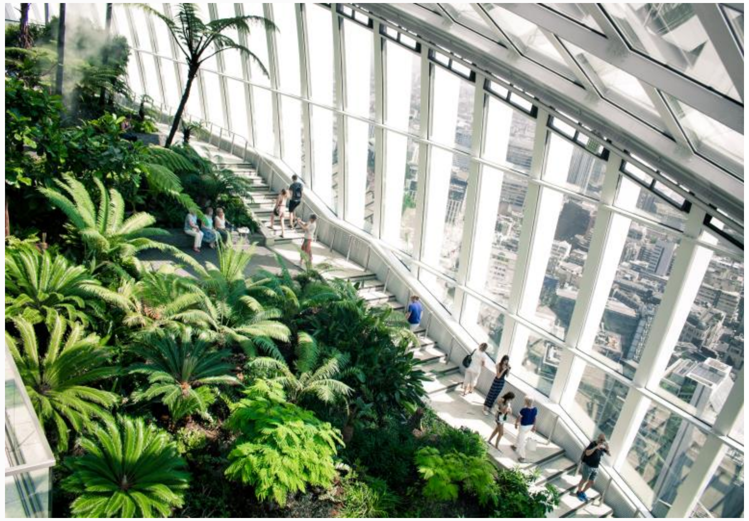
→

Microstructure:
- Phase assemblage
- Pore network

→

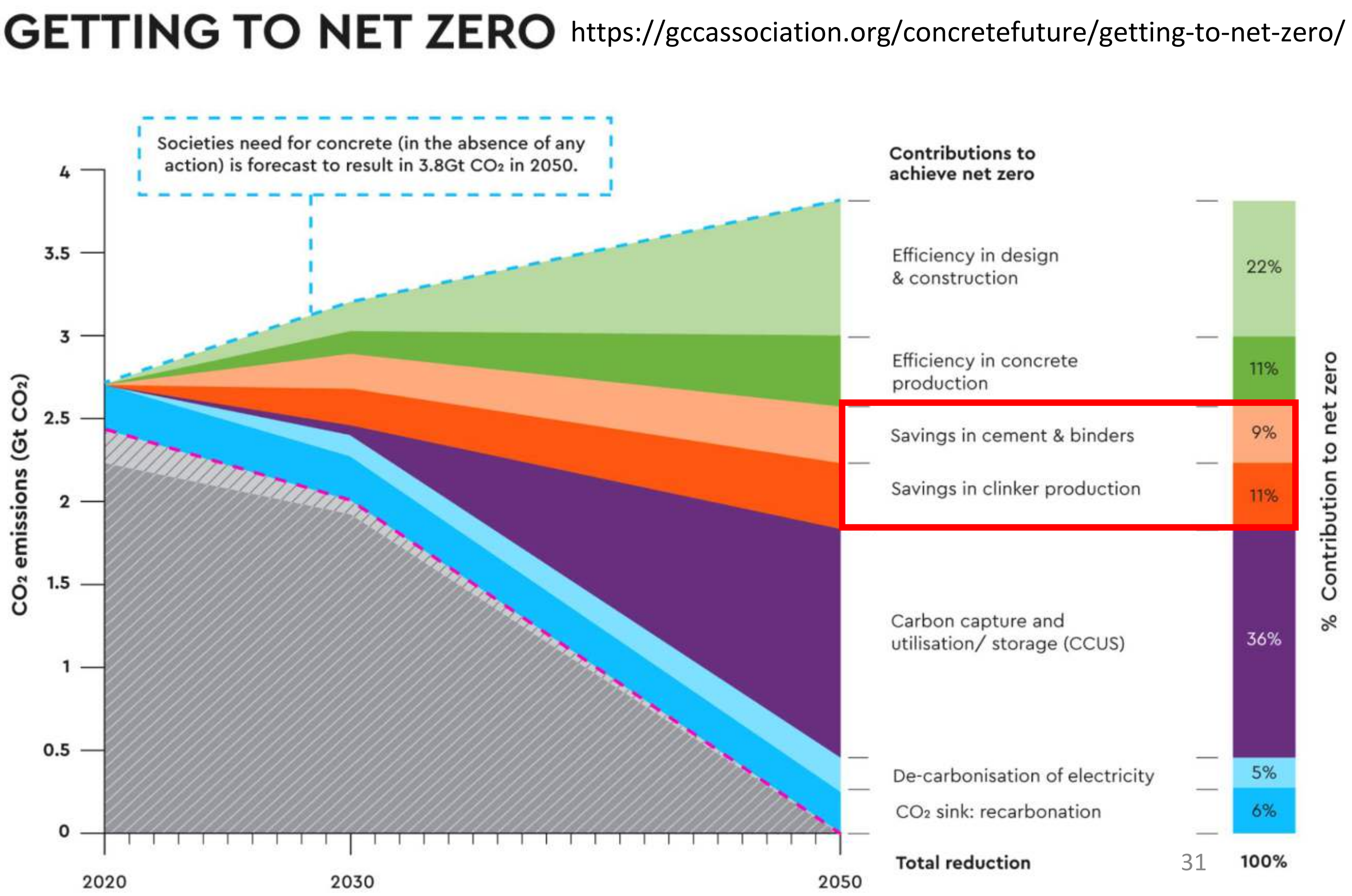
Macroscopic properties:
- Strength (2,7 & 28d)
- Durability

The global building materials industry faces major challenges



Climate change

In order to achieve the goals of the Paris Agreement for climate protection, the building materials industry needs to focus on decarbonisation.



Characterization and modelling practices

- Characterization techniques
  - Electron microscopies (SEM-BSE/EDX, TEM, EBSD)
  - Mechanical testing (strength, fatigue and creep)
  - And less common: tribology, electrochemistry, and in-situ techniques
- Modelling techniques
  - Finite-Element techniques
  - ML: for image analysis and prediction models
- Key remarks:
  - Integration is mostly manual, inconsistent data formats hamper efficient merging of experimental and simulation outputs
  - Minimal automation through the workflow: data exchanged mostly on clouds or by mails

Current Data Management Practices

- Data planning and documentation
  - Data management plan (DMP) are mostly not or partially implemented
  - Documentation tends to be custom and informal
- Data analysis and collaboration
  - Excel remains pervasing for initial logging and analysis
  - Custom scripts (python / Matlab) come second
- Sharing and storing
  - Most use cloud services
  - Some mention institutional repo: Zenodo, Sharepoint, LIMS systems, sftp servers
- Long-term preservation
  - Archiving is mostly on local / institutional servers
  - Concerns: cost, unclear ownership (& IP), minimal formal metadata
- Metadata and file formats
  - A nearly universal lack of standardized metadata
  - Multiple propriety instruments complicate file exchange

Key challenges identified

MatCHMaker
Materials Characterisation & Modelling

- Time, resources and organization priority
  - « Publish or Perish »: Many say that data management is lower priority than research task. *As a matter of fact, it is not needed for publication nor academic ranking of universities.*
  - Insufficient staff and training available: researchers often learn and do data management “on the fly”
- Data interoperability and standardization
  - Merging different data sources (different microscope, simulation inputs/outputs, etc) into cohesive workflows
  - Minimal cross-tool or cross-project standards hamper synergy
- Confidentiality and IP concerns
  - Industrial partners restrict certain data or procedures
- Scalability
  - Growing volume of large files overwhelm storage or HPC setups

A spectrum of use cases: exploratory project

- « Do we need data management at all? »
  - Data management tools are yet another bureaucratic hurdle. They are disrupting creativity and motivation
  - Science has progressed with minimal data management so far: do we actually even need more advanced tools?
- The Use Case: Exploratory / theoretical projects
  - The problem comes first: “If I had one hour to save the world, I would spend 55 minutes on defining the problem, and 5 minutes on solving it” (Einstein?)
  - Does data management helps in formulating the problem ?
  - Wittgenstein and language games: against a grandiose theoretical physicist ontology subsuming all other fields of science
  - If the problem is already clearly defined, is it still research or is it engineering?
  - As the formulation of the problem evolve:
    - so does the vocabulary to describe it
    - the exact experimental data required also (iteration of acquisitions, most previous versions being discarded)
  - To formulate the problem clearly, a core team of experts in the different fields have to sit down for long hours every week to learn from one another and build the
  - Data life cycles tools only makes sense once there is a clear understanding and stable vocabulary

This widespread illustration is a fiction: in reality the arrows should go in all directions

A spectrum of use cases: high-throughput projects

- Characteristics
  - Clear problem statement
  - Repetitive testing / simulations and high-data volume
  - Consistent data
  - Large team
- Clear pay-off of data lifecycles approaches

For high-throughput project this illustration is relevant

A spectrum of use cases: conclusions

- Characteristics

For high-throughput project this illustration is relevant

Recommendations

- The bottleneck to user engagmenet is not technical: everybody recognize the gain of a clean data management
- No silver bullet: adapt data management tools to the project type
  - Exploratory projects:
    - Do start with a DMD but aim at clear demonstrated results on a small dataset early in the project then set up place data management sprint / checkpoints.
    - Start with common sense standardization: file naming conventions, incremental metadata protocols
    - Division of labor: problem solving team doing the deep scientific problem formulation and developping tools, and acquisition teams supporting
    - Agile project management
  - High-throughput projects:
    - clearly a required
    - Waterfall project management
- Provide resources
  - Dedicated time
  - Include the role of data stewards
  - Provide documentation, tutorials and training (either at the beginning in high-througput projects, or at the intermediary stage in exploratory ones)
- Organizational aspects
  - Address the « Publish or Perish » issue: push institutions and editors to not accept publications without a minimal documented data
  - Build standard tools for the most used techniques (electron microscopies and mechanical testing).

Conclusion

MatCHMaker
Materials Characterisation & Modelling

- Data life cycle tools adoption is mostly low
  - Bottleneck to engagement are not technical
- Recommendations
- Data life cycle tools is context dependent
- Alignement with EMMC 2025 themes
  - Accelerated innovation and sustainability
  - Digitalization and interoperability
  - Industrial uptake
- Next steps:
  - Refining the survey and enlarging the pool (more projects, institutions, industrial partners)
  - Contact us if you want to take part of it!

Thank you!

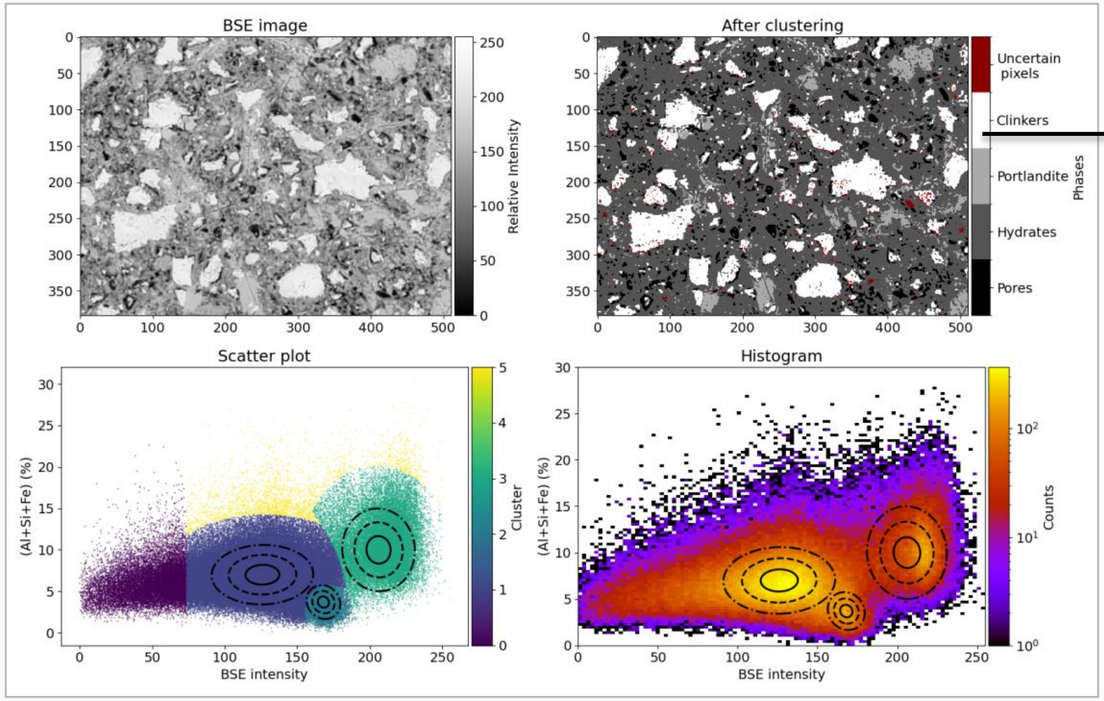
MatCHMaker
Materials Characterisation & Modelling

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement № 101091687.

Clustering processing

MatCHMaker
Materials Characterisation & Modelling

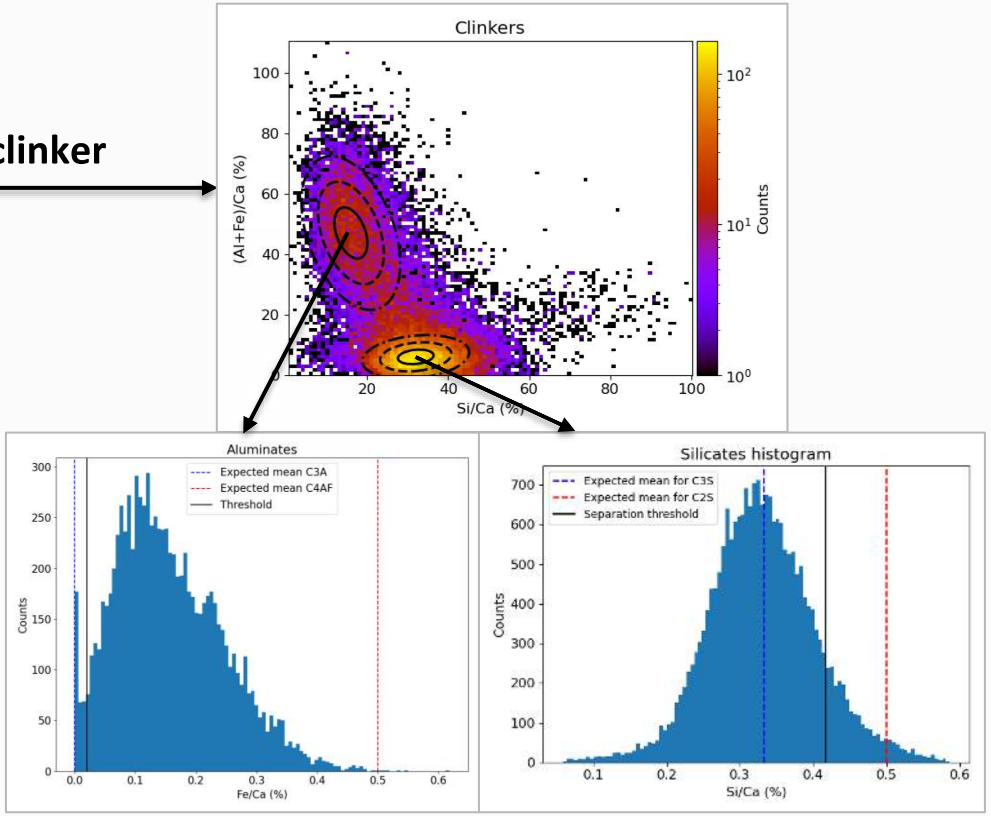
High-level clustering



Expert knowledge:

- Definition of the relevant input variables
- Initialisation of the means

Focus on clinker



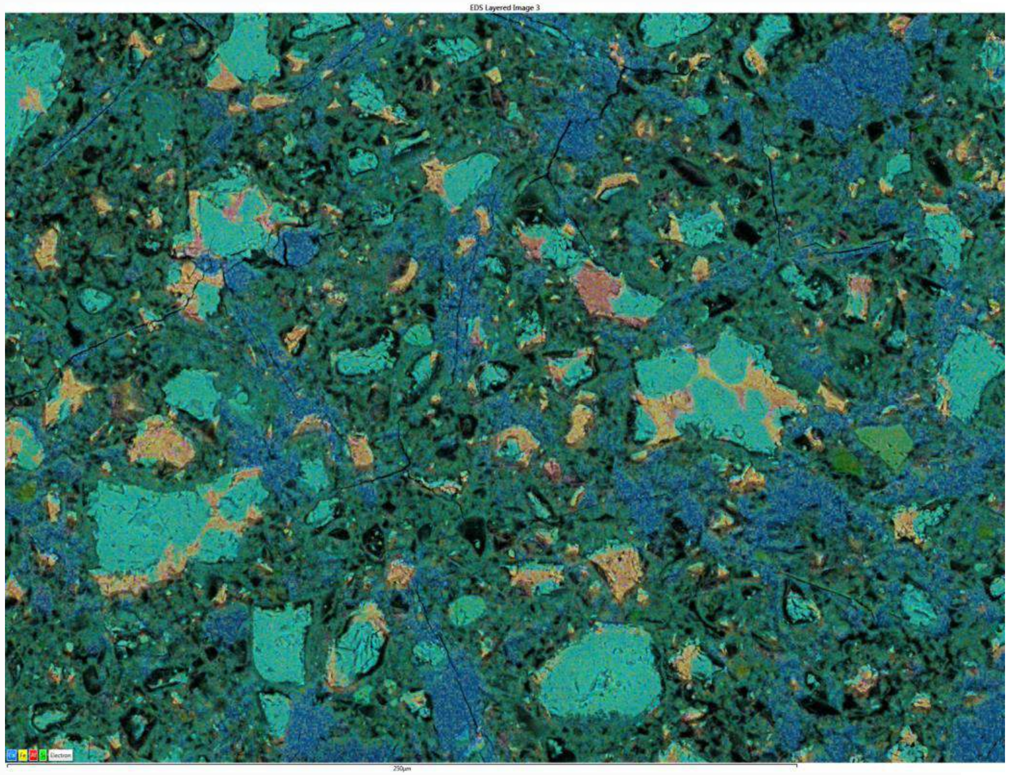
Expert knowledge:

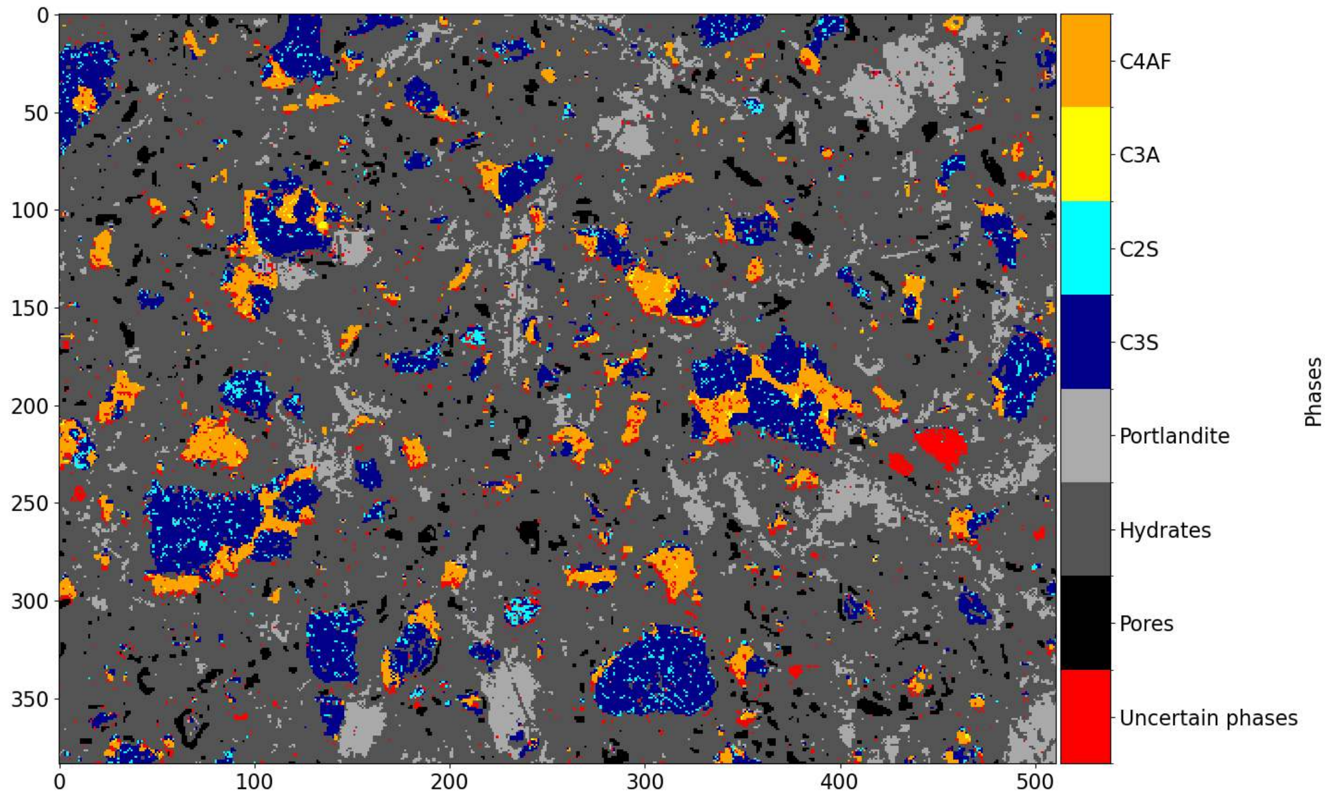
- Definition of the relevant input variables
- Thresholding values

Results

Portland cement (OPC), 2 days

Phase	Uncertain phases	Pores	Hydrates	Portlandite	C3S	C2S	C3A	C4AF
Quantity (%)	2,17	3,18	72,76	6,91	9,20	1,11	0,15	4,51





Refinement on hydrates

Portland cement (OPC), 2 days

Phase	Uncertain phases	Pores	Hydrates	Portlandite	C3S	C2S	C3A	C4AF
Quantity (%)	2,17	3,18	72,76	6,91	9,20	1,11	0,15	4,51

Problem

Phase unmixing at sub-pixel level

Proposed method

Lasso regression, given M matrix of phase composition, Y elemental pixel measurement

$$\hat{\boldsymbol{\lambda}} = \underset{\boldsymbol{\lambda} \geq \mathbf{0}}{\mathbf{argmin}} \|\boldsymbol{M\lambda} - \boldsymbol{Y}\|_2^2 + \boldsymbol{\alpha}\|\boldsymbol{\lambda}\|_1$$

Lasso ⍰ Sparsity in the number of coefficients

Limitations :
- M is not perfectly known
- Noise in measurement Y