Important notice!

The students of the course of *Machine Learning for Process Engineering* are warned that the video, the slides, the software and the data shared by Prof. Facco are **strictly confidential** and can be used only to the purpose of preparing this specific exam.
Any other use is strictly forbidden.
Publication, divulgation or disclosure of these materials (or part of it) is legally punishable and will be legally pursued.

Università degli Studi di Padova

Department of Industrial Engineering

Machine Learning for Process Engineering

Academic year 2025-2026

Prof. Pierantonio Facco

CAPE-Lab, Computer Aided Process Engineering Laboratory

Email: pierantonio.facco@unipd.it

URL: https://research.dii.unipd.it/capelab

Prof. Facco's introduction

- **Associate professor**, Department of Industrial Engineering, University of Padova
  - **ICHI-01/C** – Theory of the development of chemical processes

- Education
  - Master Science in Chemical Engineering (2005, University of Padova)
  - **Ph.D. in Industrial Engineering – Chemical Engineering** (2009, University of Padova)

- Expertise and research activity:
  - **machine and deep learning**
    - **quality and process monitoring and soft-sensing**
    - **autonomous adaptive systems for model updating**
    - **artificial vision systems**
    - **data and sensor fusion**
    - **anti-fraud and anti-sophistication technologies** for pharmaceutical, manufacturing and food industries
    - **product, process and technology development, transfer and scale-up**
  - **process modelling**
    - **digital twins**
    - **hybrid models**
  - Statistical, dynamic and model-based **Design-of-Experiments**

Teaching activity

- **Courses**
  - **Machine learning for process engineering**
    - **Chemical & Process Engineering**
  - **Process optimization and scheduling**
    - **Chemical & Process Engineering** (from II semester 2026/27)
  - Machine learning for quality control in food production
    - Food & Health
  - Agri-food products market economy and Statistical methodologies for risk analysis
    - Food sanitary hygiene and safety

- **University of Padova – Texas State International Educational Hub** University of Padova (Italy)
  - organization, coordination and lecturing
  - **Industrial engineering statistics**

MOOC on digitalization

FutureLearn

Subjects Courses FutureLearn for business

Online Courses / Science, Engineering & Maths

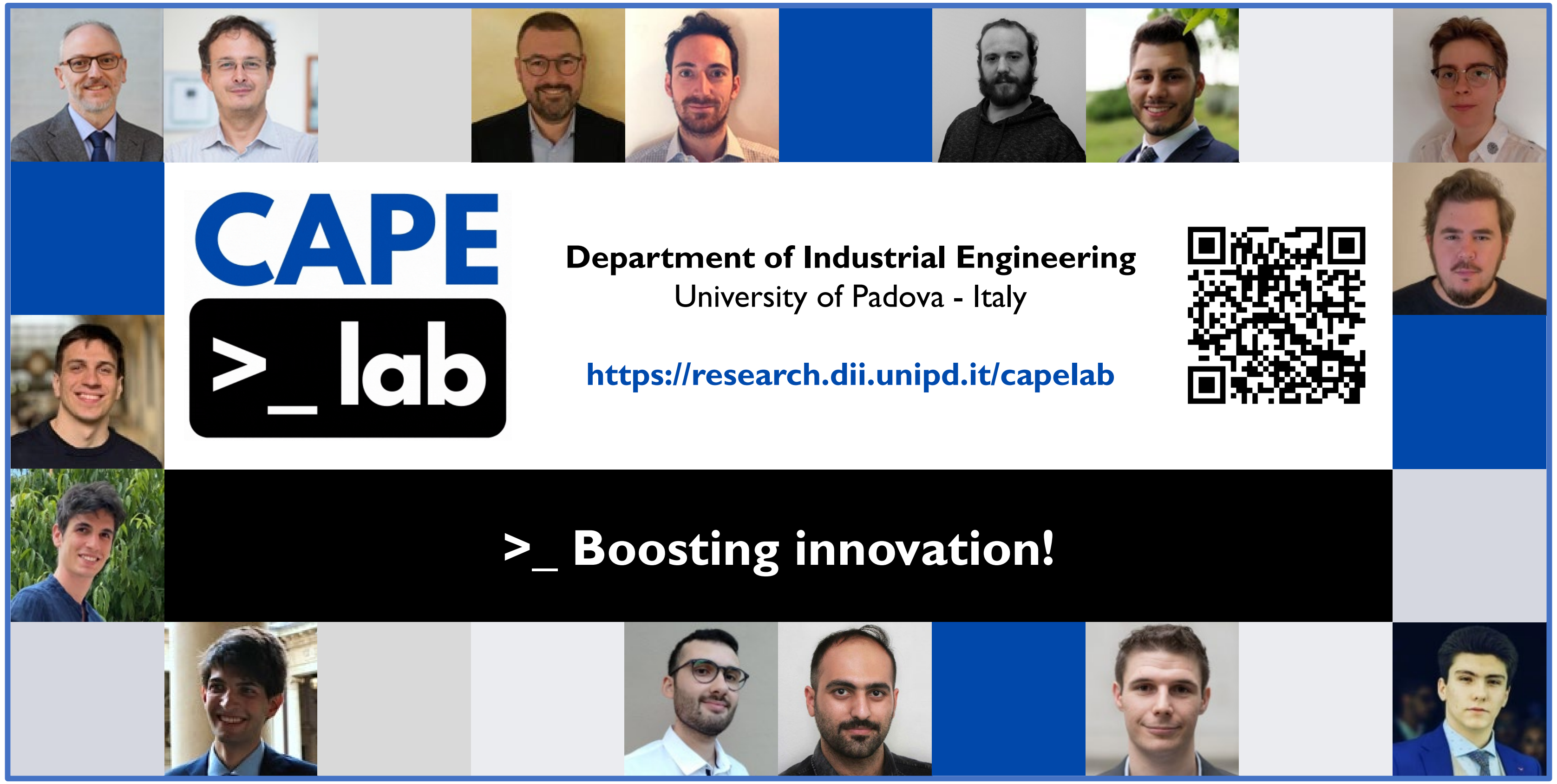
Università degli Studi di Padova

4.0 Shades of Digitalisation for the Chemical and Process Industries

Make the transition to Industry 4.0 as you learn how to streamline chemical and process engineer jobs with digital technologies.

Research group: CAPE-Lab

- Computer-Aided Process Engineering Laboratory
  - Research and technological development in **Process Systems Engineering**



Research achievements

- **Publications**
  - >150 publications: **66 papers on peer reviewed journals** + 2 book chapters + >80 conference proceedings…
  - >40 presentations **to international congresses and schools**, also as an invited and keynote speaker
- **Research and commercial funding attraction**
  - **>1.9 M€** in the last 8 years

Academic collaborations

- **University College London** (UK)
  - innovative methodologies for model-based design of experiments
- **Universitat Politecnica de Valencia** (Spain)
  - novel latent variable model inversion techniques under uncertainty
- **Imperial College**, London (UK)
  - biopharma cell culture modelling development
- **Rutgers University** (USA)
  - design space identification in the pharmaceutical industry
- **Louisiana State University** (USA)
  - monitoring of integrated circuits manufacturing
- At the **University of Padova**
  - **BioEra**
    - time expression of DNA/RNA sequencing for circadian rhythms identification
  - **Biamet**
    - discovering the role of exosomes in neuroblastoma
  - **PARLab**, University of Padova
    - artificial vision systems for microalgae cultivation
    - experimental protocols in microbioreactors
  - **Department of Animal Science and Public Health**, University of Padova
    - anti-sophistication technologies in food industry
  - **Department of Automation Engineering**, University of Padova
    - robust methods for image analysis

Industrial collaborations

- **Pharma and biopharm**
  - **Glaxo Smith Kline**
  - **Newchem**
  - Eli Lilly
  - Merial/Sanofi
  - Pfizer
- **Biomedical**: Fresenius Kabi
- **Biorefinery**: Novamont
- **Catalysis**: **Casale**
- **Cement: FLSmidth**
- **Food**: Buhler
- **Fine chemicals**
  - **Sirca**
  - BASF
  - Versalis-Eni
- **Eyewear**: Safilo
- **Manufacturing**
  - Unox
  - Carel

Tech transfer

- **Patent**
  - **Glaxo Smith Kline** (2024): Method for predicting production stability of clonal cell lines
    - digital tool to support the selection of clonal cell line for producing therapeutic proteins

- **Software**
  - **eFiber** (2012), Fresenius Kabi (Cavezzo, MO, Italy)
    - artificial vision system for microfiber diameter measurement from SEM images
  - **ADAM** (2022), Glaxo Smith Kline (Ware, U.K.)
    - metabolomics dynamics interpretation
  - **eKinetics** (2022). Glaxo Smith Kline (Ware, U.K.)
    - reaction kinetics characterization in drug substance development
  - **bGen** (2022), freeware
    - in-silico batch generator for process monitoring with small data
  - **COSMO** (2024), Glaxo Smith Kline (Ware, U.K.)
    - cell-line selection and process data analysis
  - **CURES Analyzer** (2025), Glaxo Smith Kline (Ware, U.K.)
    - culture replicate similarity analyzer in AMBR®15 experiments

Entrepreneurship

ProDig
Digital solutions for the real world

- I am a **founder**, and the **Chief Technology Officer** of **ProDig**
  - spin-off of the University of Padova
  - **hybrid modelling in the digitalization of (bio)pharm, chemical, manufacturing and food industries 4.0 and 5.0**

Introduce yourself, please!

- Could you, please, introduce yourself?
  - what is your background?
- What are your expectations from the course?
- Could you, please, use a paper sheet to do a place-card with your first name and family name?

Main objective for this course

- To teach part of my **industrial (and research) experience on machine learning and design of experiments**
- To transfer fundamental skills:
  - Machine Learning in process engineering
  - **data "translators"**
- To introduce all of you to (hopefully) new and useful disciplines:
  - **univariate statistics**
  - **multivariate statistical techniques**
    - exploratory analysis and data mining
    - process understanding and troubleshooting
    - quality improvement
    - classification
    - estimation/prediction (regression)
  - **multivariate statistical process control**
  - **design of experiments** and **response surface modelling**

Information on the course

- Detailed information on the course can be found in the **syllabus**
  - please, use your mobile phone to read this **QR code**:

… you will find the course syllabus here!

Università degli Studi di Padova | Course Catalogue | Go to the university portal | Search by | IT | EN

Home / Courses / Course [IN2925] / Teaching [INQ3103840]

[INQ3103840] - MACHINE LEARNING FOR PROCESS ENGINEERING

Year of enrolment 2025/2026

Save PDF

General information

Course: CHEMICAL AND PROCESS ENGINEERING
Course type: Master's Degree
Academic year: 2025/2026
Year of course: 1
Training activity type: Affine/Integrative
Language: English
ECTS credits: 6 ECTS (6 ECTS ING-IND/26)
Didactic Activity Type: Lesson
Evaluation: Final grade
Teaching period: Second semester (from 01/03/2026 to 15/06/2026)
Teaching type: Teaching with lectures
Holders: FACCO PIERANTONIO
Length: 48 hours (48 hours Lesson)
Attendance: Not mandatory
Didactic method: In presence
Subject area: ING-IND/26
Location: PADOVA
Corso a libera scelta: This teaching can be selected as a free-choice course

Course agenda

- **Introduction**
  - probability theory and univariate statistics

10% of the course

- **Machine Learning**
  - mining information from an overload of data
  - predicting useful information from available data
  - data classification and clustering

50% of the course

- **Design of experiments**
  - plan an experimental campaign
  - analyzing data variability in a critical manner
  - extract useful information from experiments

40% of the course

Lessons plan & topics

lesson	month	day	part	lesson	topic
1	February	24	intro	0	Introduction + Machine learning in process industry 4.0 and 5.0
2		25	ML	1	Introduction to machine learning
3	March	3	ML	2	Data challenges and probability theory
4		4	ML	3	How to describe data: random variables and descriptive statistics
5		10	ML	4	Probability distributions and hypothesis testing
6		11	ML	5	Dealing with large amounts of data: introduction to multivariate statistics and latent variables methods
			ML	6 Flipped	Principal component analysis, PCA
7		17	ML	7	Practical example of PCA
8		18	ML	8 + Laboratory 1	PCA in practice + Laboratory 1
9		24	ML	9	Statistical process control and process capability
10		25	ML	10	Process monitoring and monitoring charts
11		31	ML	11	Industrial process understanding, troubleshooting and process monitoring: production of copper and continuous processes
			ML	Laboratory 2	Industrial applications of fault detection and diagnosis: slurry-fed glass melter for nuclear waste stabilization
12	April	1	ML	12	Monitoring of batch processes
13		14	ML	13	Unsupervised identification of different groups of products/processes: clustering: k-means and hierarchical clustering
			ML	Laboratory 3	Batch process monitoring application: styrene-butadiene polymerization for the production of rubber
14		15	ML	14	Supervised models for product and process classification: regression models, KNN, linear and quadratic discriminant analysis
15		21	ML	15	Product quality prediction through multiple linear regression
			ML	16 Flipped	Partial least squares, PLS; applications to the plastic and biomedical industries
			ML	Laboratory 4	PLS applications: productivity prediction in the batch production of penicillin; NIR spectroscopy in food and gasoline industry
16		22	DoE	1	Introduction to design of experiments DoE
17		28	DoE	2	Procedure to apply DoE
18		29	DoE	3	Comparative experiments; example on the cement production
			DoE	4 Flipped	Understanding the qualitative effect of factors on processes and products: Analysis of Variance ANOVA
19	May	5	DoE	5	Improved process and product understanding through factorial designs
20		6	DoE	6	Factors effects on product quality and examples
21		12	DoE	7	DoE in practice: examples in the manufacturing and beverage industries
22		13	DoE	8	Unreplicated full factorial designs: applications to catalytic reactors and pressure vessels
23		19	DoE	9	Process/product initial exploration through fractional factorial designs and examples to chemical and mechanical industry
24		20	DoE	10	Process and product optimization through central composite designs: applications to the food and coke industries

Types of lessons

- Lessons in the **class**
- 3 **flipped lessons**
- Asynchronous telematic lessons of **computational laboratory**

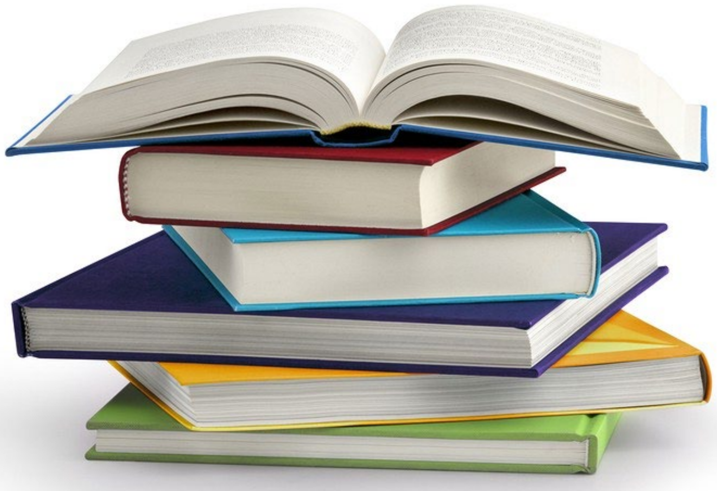
Teaching materials: lessons slides and videos

- The `.pdf` of the **slides** will be provided
- **Videos** of the lessons will be available in Moodle
  - be present in the class as much as possible :)
  - from a statistical analysis of the past years, I understood that the students who attend classes typically get better grades
- A couple of **books**
- Some useful **readings** will be suggested
  - important specific references are already available in the course Moodle

How to read the slides?

- The **important concepts** will be highlighted in bold blue
  - sometimes stressed with a blue box
- The **criticalities** and the **challenging issues** will be highlighted in bold red
- The **`suggested exercises hands-on PC`** will be proposed with green bold Courier New
  - stressed with a green box

Teaching materials: textbooks



- **Textbooks**:
  - Eriksson, Kettaneh-Wold, Trygg, Wikström, Wold (2006)
    Multi- and Megavariate Data Analysis : Part I - Basic Principles and Applications
    Umetrics Inc.
  - Montgomery (2012)
    Design and analysis of experiments
    Wiley
  - Montgomery (2004)
    Introduction to Statistical Quality Control
    Wiley
  - Eriksson, Johansson, Kettaneh-Wold, Wikström, Wold, S. (2008)
    Design of Experiments - Principles and Applications
    Umetrics Inc.
  - Hastie, Tibshirani, Friedman (2008).
    The Elements of Statistical Learning - Data Mining, Inference, and Prediction, 2nd Ed.
    Springer
  - Johnson, Wichern (2007)
    Applied Multivariate Statistical Analysis
    Pearson Prentice Hall
  - Chiang, Russell, Braatz (2001)
    Fault detection and diagnosis in Industrial systems
    Springer
  - Ogunnaike (2009)
    Random phenomena – Fundamentals of probability and statistics for engineers
    CRC Press

Hands-on-pc course

- You will be challenged to demonstrate your ability in dealing with **real industrial data**
- The course is conceived to make you **work on computers**
  - you will use of **your laptop**
- **No hard programming skills are required**

Teaching material: software

- **Proprietary software** will be provided
  - the use of the provided software is allowed **only to prepare the final exam of the course**
  - **virtual machines** are also available as an alternative, if needed

- **Software**
  - **Matlab®**, Matworks **- 2024 release**
    - UniPD campus license
  - **PLS_Toolbox®**, Eigenvector Research inc.
    - personal license of the software will be provided
  - **Minitab®**
    - personal license of the software will be provided

the fact that the software is **properly installed and works** in the laptop is a **student responsibility!**

Learning assessment

- **2 individual homework** (35% of the final grade)
  - homework #1 on Machine Learning (ML)
  - homework #2 on Design of Experiments (DoE)

- **Final exam** will be held **hands-on PC** (65% of the final grade):
  - 4 calls: June, July, September, January
  - how:
    - 1 numerical exercise on ML (10 points)
    - 1 numerical exercise on DoE (10 points)
    - 2 open questions (4 points each)
    - 5 multiple choice question (1 point each)

2h/2.5 h: pen, paper, laptop and all the printed support the student needs

30 min: pen and paper (oral)

- Final grade: $0.35\left(\frac{H1}{2}+\frac{H2}{2}\right)+0.65(WE)$
  - $H1$ = homework #1 grade
  - $H2$ = homework #2 grade
  - WE = written exam grade

where:
- $H1, H2, WE \in [0/30, 33/30]$
- $\geq 30.5/30$ means *cum laude*

- **The final exam will be held with your laptop**
  - you are responsible of how it works
- **No extra calls will be available**
  - no shortcuts

Homework

- They are typical problems you find in the final exam

- Deadlines:
  - 1st homework late April
    - deadline: 10 days/a couple of weeks later
  - 2nd homework 10 days before the final exam
    - deadline: 1 week before the final exam
  - …however, **we will fix the deadlines together**, based on students' needs

- Few time required to complete the homework (if you have already attended the lessons and the labs, and studied them):
  - They are the best way to prepare the final exam!

Everyday study, homework and self-assessment

- Everyday study is important!
  - remind that for every h of lesson in the class you are expected to study 2 h at home
- Every lecture I will challenge you with homework
  - **not mandatory**
  - however:
    - they take few minutes
    - they help you to get a **deeper practical understanding** of the proposed topic
    - you can **self-evaluate your preparation**
    - you can **pave a path for success** for the final exam
    - you will be ready to take the exam at the end of the course!

… you will be involved in active learning

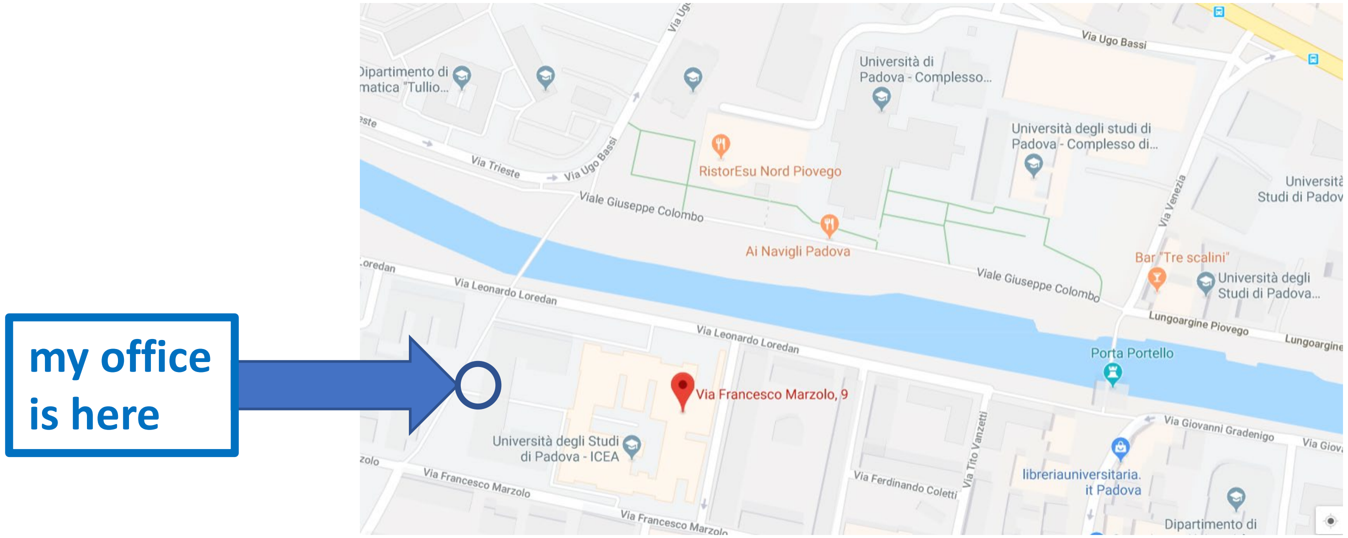
- Teamwork during the lessons
  - make your mates an extra-value
- Brainstorming and discussion
  - feel free to speak about your ideas
- Use (your) computer
  - hands-on-PC sessions are useful to practice!
- Use your mobile phone
  - I need your feedback

Subscribe to the course

- If you decide that this course fits to your expectations (and your needs, as an engineer) **subscribe to the course via Moodle**
  - no password required

Contacts

- Contacts:
  - email: pierantonio.facco@unipd.it
  - web page of Dr. Facco: https://www.unipd.it/en/contatti/rubrica/?detail=Y&ruolo=1&checkout=cerca&persona=FACCO&key=40D2DF502A70C84E3897E0375D3AE30B
  - web page of the CAPE-Lab research team: https://research.dii.unipd.it/capelab/
  - you can find me also in **LinkedIn**, **Instagram**, etc.
- To schedule a meeting, please contact me by email
  - consider that my agenda is typically packed for the next 2 weeks
- My office address at DII is: via Marzolo 9, 35131 Padova



Questions?

- Curiosities?
- Missing information?

Università degli Studi di Padova

Department of Industrial Engineering

Machine learning for process industry 4.0 and 5.0

Successful industrial applications

Academic year 2025-2026

Prof. Pierantonio Facco

CAPE-Lab, Computer Aided Process Engineering Laboratory

Email: pierantonio.facco@unipd.it

URL: https://research.dii.unipd.it/capelab

Why Machine Learning for process engineers?

Designing information, not just experiments

- You work for a **food industry**
  - you have to formulate a new product and launch it into the market as soon as possible to gain advantage on competitors
  - **an extended experimental campaign** must be performed to optimize the product formulation and guarantee its safety
  - the cost and duration of the experimentation should be minimized
- **….but:**
  - **experimental protocols for product formulation are based on the formulators experience**
  - **experiments are costly, time-consuming and require trained personnel**

- Questions:
  - which is the best recipe to guarantee the best taste of the new product?
  - how would you perform the experimental campaign?
- A **science-based experimental protocol** guarantee to obtain the highest informative content with the most parsimonious experimental campaign!
  - that is why we need appropriate methodologies for designing, performing and analyzing experiments

From bench to bioreactor: scaling up a biopharm product

- You work for a **multinational biopharmaceutical industry** which is developing a new monoclonal antibody for the treatment of cancer
  - you have to
    - improve process understanding
    - scale-up the cell culture from the lab to the industrial-scale plant
  - **hundreds, if not thousands of experiments** are carried out
  - a big amount of data (temperature, pH, dissolved oxygen, but also biological information) are collected from every scale
- **….but:**
  - **process understanding is left to the experience of scientists**
  - **scale-up is performed in a very "slow" manner**

- Questions:
  - which are the most productive and stable cell lines?
  - how would you scale-up the process?
  - how would you deal with uncertainty and large variability?
- Common sense and experience is not the best way to develop a new process and a new product!
  - that is why we need appropriate methodologies to deal with the wealth of information stored into the **available data and their variability**

Everything is under control… until it isn't

- You work for a **company producing plastic materials** mainly for construction engineering and automotive
  - you are in charge of controlling a polymerization reactor to produce an important polymer grade
  - **120 process variables** (temperatures, flowrates, pressures) are collected from the plant and recorded every second by process computers
  - the automatic control system is working properly
  - no alarms were warned in the last 3 days
- **….but:**
  - **the last batch product quality is out of specification**
  - **panic! Your company lost 400.000 €!!!**
- Questions:
  - which variable has changed?
  - how would you have identified it? By looking at 120 trends one by one?
- Nothing changed individually: what changed was the relation among variables!!!
  - that is why we need appropriate methodologies to deal **with large amounts of data**

Everything is under control… until it isn't

- You work for a **small-medium enterprise which manufactures resins** mainly for coatings
  - you are in charge of the quality control system
  - a product can be manufactured in different reactors, but you want to ensure a repeatably high quality
  - **dozens of process variables** are collected online from the plant
  - the automatic control system work properly
  - 2 hours ago the operators took a sample of resin from the reactor, and the product quality was OK
- **….but:**
  - **few minutes ago the product quality drifted out of specification**
  - **you want to propose the correction of the recipe to restore quality to specification**

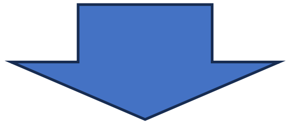
- Questions:
  - how could you anticipate the detection of out-of-specs?
  - how could you propose the appropriate corrective action?
- Predicting process anomalies and anticipating the corrective actions and maintenance is of paramount importance!
  - that is why we need appropriate methodologies to **predict process behavior**

Typical applications in process engineering

- Product formulation and design
- Process development and design
- Process understanding and troubleshooting
- Product and quality monitoring
  - fault detection and diagnosis
- Predictive maintenance
- Soft sensing
  - virtual sensors for hard-to-measure variables
- Process optimization
- Process control
- Efficiency and sustainability

Why now? The data revolution in process industry

- Industries 4.0 and 5.0 and digital transformation: increased availability of plant data
  - historians
  - sensors
  - lab measurements
  - etc.
- Computational power
  - (open-source) ML tools



- Competitive advantage through **data-driven decisions**
- Shift from purely mechanistic to **hybrid modelling**

Machine Learning vs. first-principles modeling

- **First-principles models**
  - (simplified) understanding of chemical, physical and biological phenomena
    - conservation, kinetics, or thermodynamics laws
  - interpretable and physically consistent
  - differential and algebraic equations modelled in programming languages

+

- **Data-based models**
  - utilize “black boxes” that depend heavily on the data quality
  - capture relationships among data
  - deal with non-linearities
  - handle uncertainty

- **Hybrid models**
  - combine the strengths of first-principles and data-driven approaches
  - address plant-model mismatches and nonlinear process complexities
  - tackle complex applications

This is not a machine learning course…

- This is a course about developing new products and processes, understanding and troubleshooting processes, monitoring processes and products, classifying products, extracting valuable information from systems under study
  - data are just a tool
  - black-box models are not enough
    - interpretability over hype
- Engineers must interpret and decide!

Here we are proposing the professional figure of the **process engineer of the near future**!

Machine learning for process industry 4.0 and 5.0

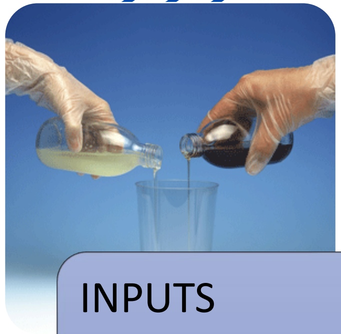
Successful industrial applications

Status-quo in process industry

data

process measurements

product analysis



INPUTS

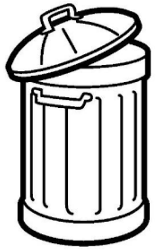
- checking raw materials
- fixed process parameters

PROCESS

- fixed process
- predetermined recipe
- low knowledge
- rigid regulation

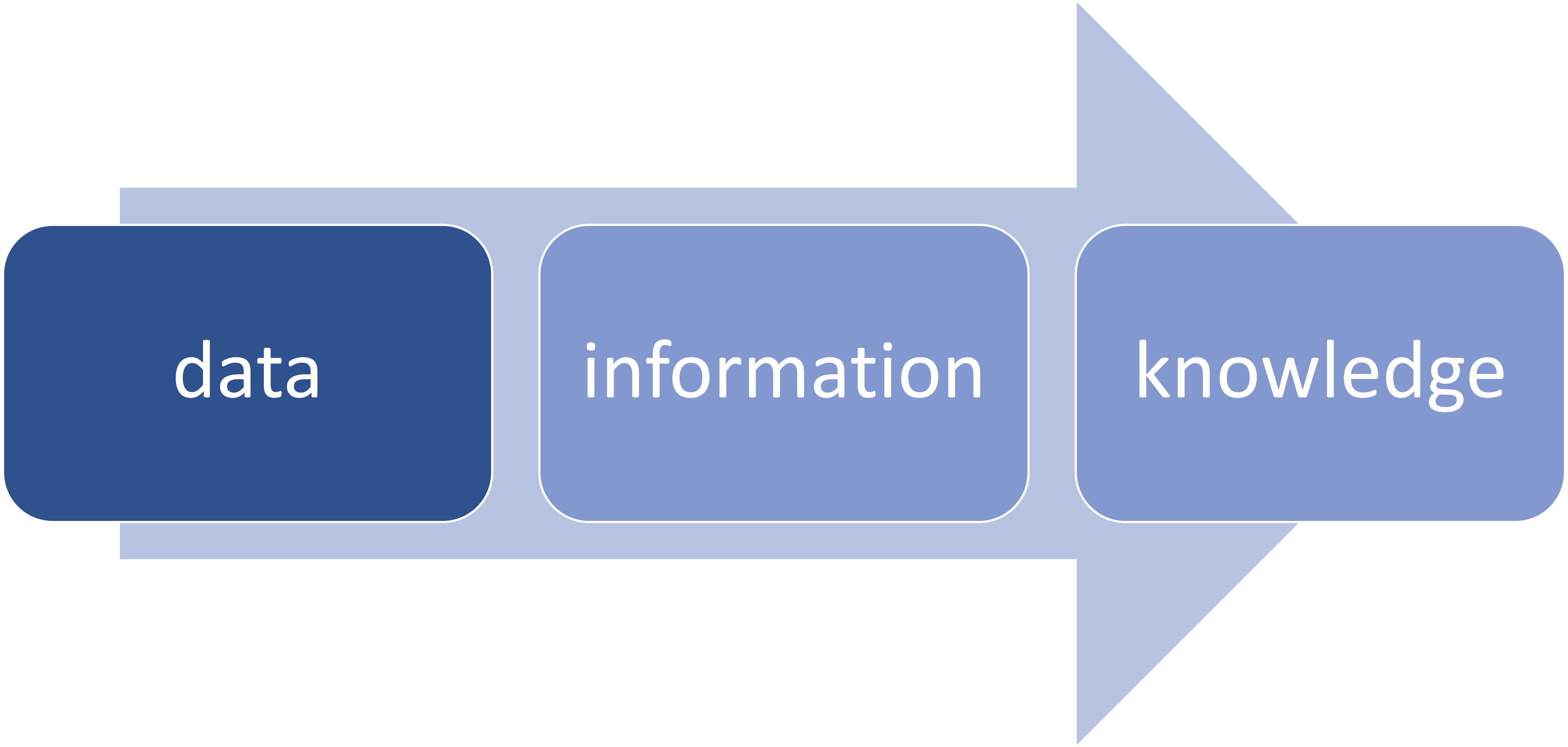
PRODUCT

- a-posteriori quality control
- expensive and time-consuming lab assays



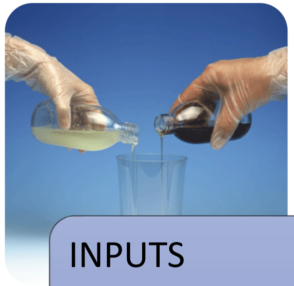
time loss
& waste

Data, information, knowledge



Benefits of data analytics

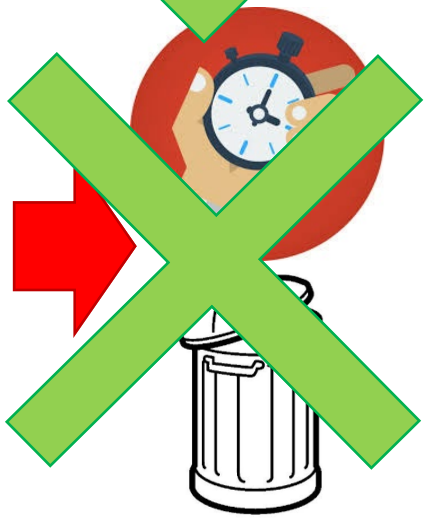
DATA → INFORMATION → KNOWLEDGE

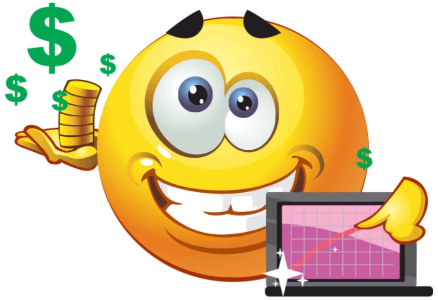


INPUTS
- checking raw materials
- fixed process parameters

PROCESS
- fixed process
- predetermined recipe
- low knowledge
- rigid regulation

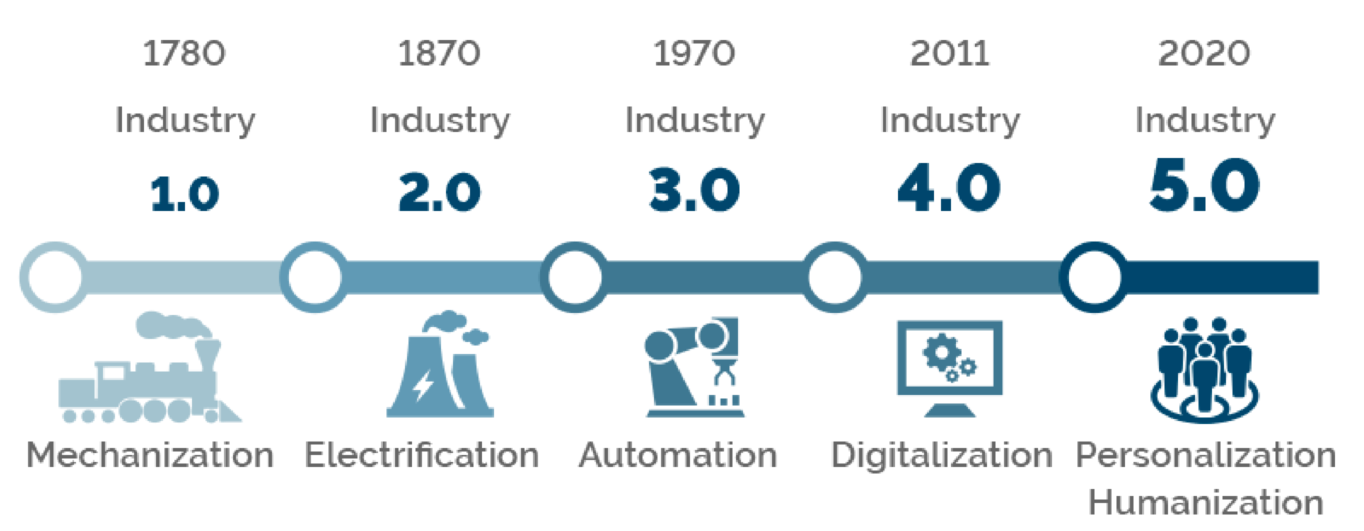
PRODUCT
- a-posteriori quality control
- expensive and time-consuming lab assays





4th and 5th industrial revolutions

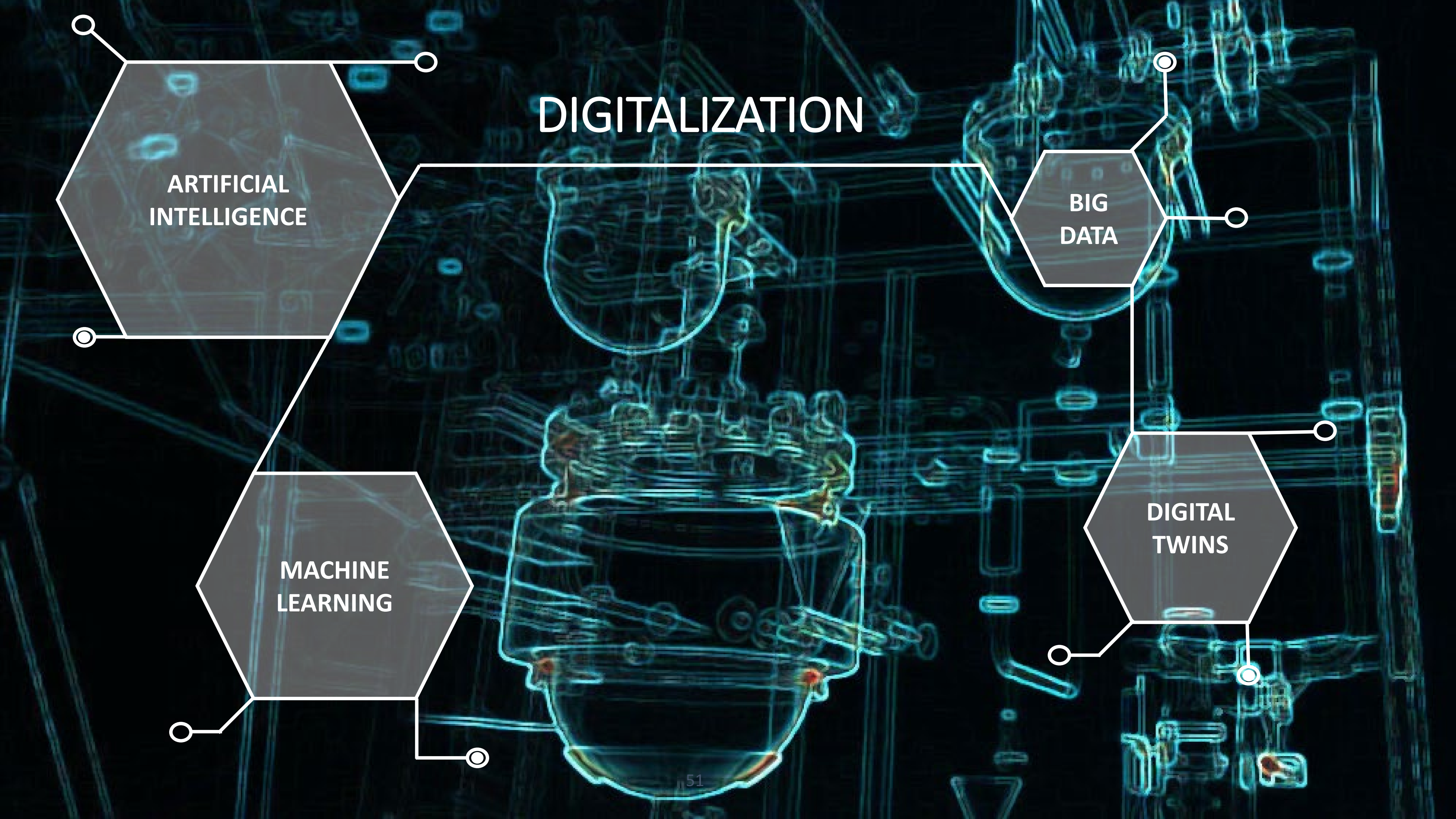
- **Industry 4.0**
  - strong industrial digitalization in process development, manufacturing, commercialization, …
  - smart integrated production:
    - cyber-physical systems
    - connected factories
    - cloud technology
    - IoT
    - **big data**
- **Industry 5.0**
  - human – technology combination
  - collaborative robotics
  - pillars
    - **people centricity**
    - sustainability
    - resilience



How can we help product development/manufacturing?

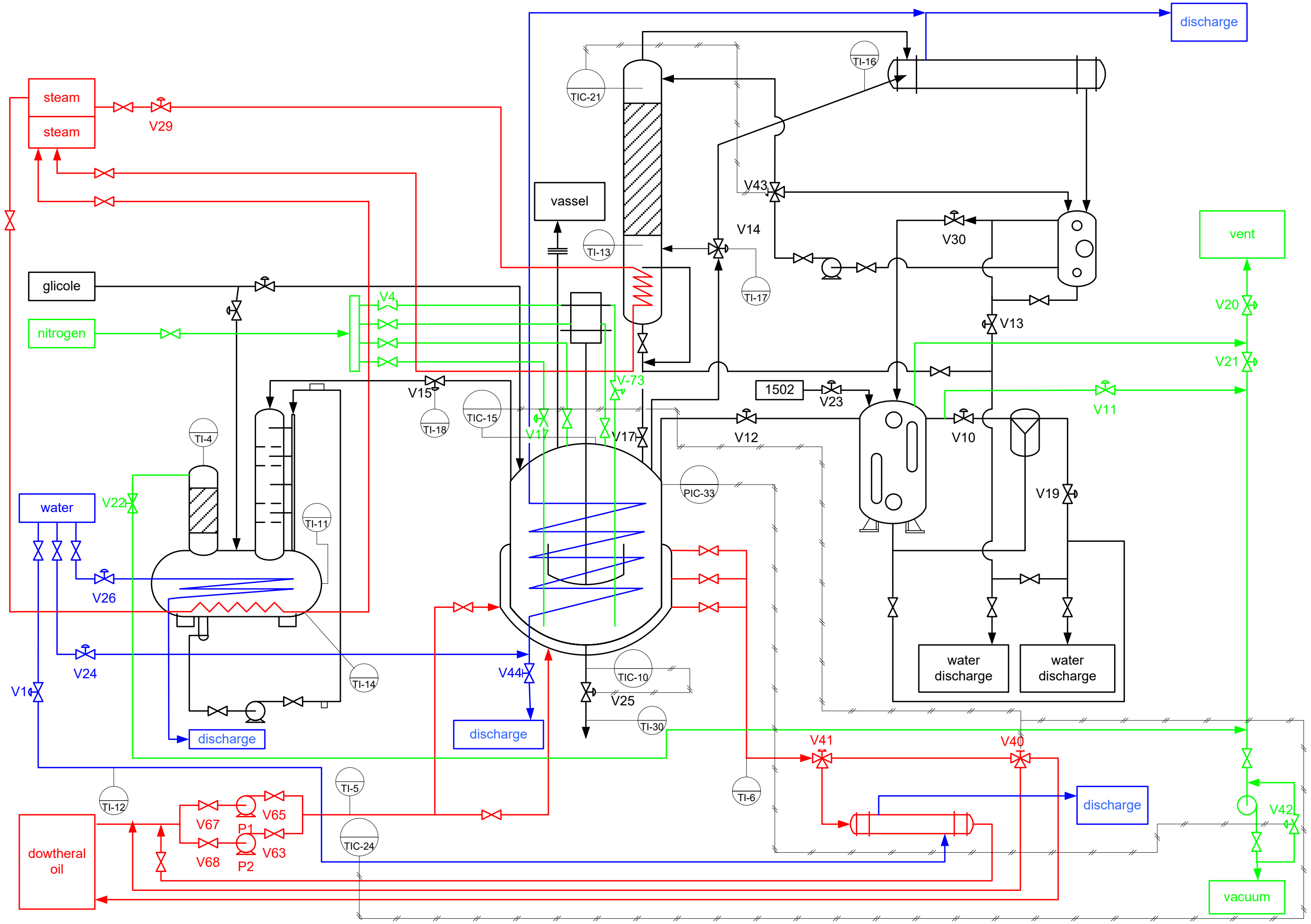
- Is it OK to develop a product and a process or to run a process based on the experience of experts and practitioners?
- Do **science-based methods** exist to accelerate the product/process development and to support the production of products of the desired quality?

- **DIGITALIZATION**
  - sensors and process analytical technologies (PAT) available at relatively low cost
  - dozens, hundreds, even millions of **data** collected and recorded every few seconds
    - online sensors
      - temperature
      - pressure
      - flow rates
      - pH, …
    - advanced technologies
      - spectra, …
      - images …


DIGITALIZATION
ARTIFICIAL INTELLIGENCE
BIG DATA
MACHINE LEARNING
DIGITAL TWINS

Let’s start from a chemical plant…

Typical plant layout



Online plant instrumentation data

process variables

batches

time

time	reactor temperature 2 [K]			pressure [bar]			flow rate [kg/min]			exchanger temperature 1 [K]		
	OP	PV	SP	OP	PV	SP	OP	PV	SP	OP	PV	SP
21/06/2022 00:00	0	415.9489	416	0	1.283862	1	0	0	0	0	349.2377	350
21/06/2022 00:01	0	415.8793	416	0	1.058317	1	0	0	0	0	350.0228	350
21/06/2022 00:01	0	416.1596	416	0	1.039562	1	0	0	0	20	350.3868	350
21/06/2022 00:02	0	416.1564	416	0	1.31754	1	0	0	0	20	350.7704	350
21/06/2022 00:02	0	415.5676	416	0	0.839107	1	0	0	0	20	351.0809	350
21/06/2022 00:03	0	415.985	416	0	1.139325	1	0	0.1	0	20	350.0602	350
21/06/2022 00:03	0	415.9176	416	0	1.167018	1	0	0.1	0	30	348.9559	350
21/06/2022 00:04	0	416.3139	416	0	0.951257	1	0	0	0	30	349.4804	350
21/06/2022 00:04	0	416.5466	416	0	1.043134	1	0	0	0	33	349.2569	350
21/06/2022 00:05	0	416.5546	416	0	0.766831	1	0	0.2	0	33	351.6453	350
21/06/2022 00:05	0	415.5682	416	0	0.770409	1	0	0.1	0	30	349.5691	350
21/06/2022 00:06	0	416.0387	416	0	1.020975	1	0	0	0	30	350.5237	350
21/06/2022 00:06	0	415.3929	416	0	1.144451	1	0	0	0	30	349.8653	350
21/06/2022 00:07	0	415.4432	416	0	1.517098	1	0	0	0	30	350.622	350
21/06/2022 00:07	0	415.9966	416	0	0.866622	1	0	0	0	20	349.4646	350
21/06/2022 00:08	0	416.7663	416	0	1.037466	1	0	0	0	20	349.0184	350
21/06/2022 00:08	0	415.6152	416	0	0.983501	1	0	0	0	2	349.0043	350
21/06/2022 00:09	0	416.1857	416	0	0.613395	1	0	0	0	0	350.3417	350
21/06/2022 00:09	0	415.8872	416	0	0.912207	1	0	0	0	0	349.8758	350
21/06/2022 00:10	0	416.5587	416	0	0.641064	1	0	0	0	0	349.8628	350

Example of laboratory data

quality variables

time

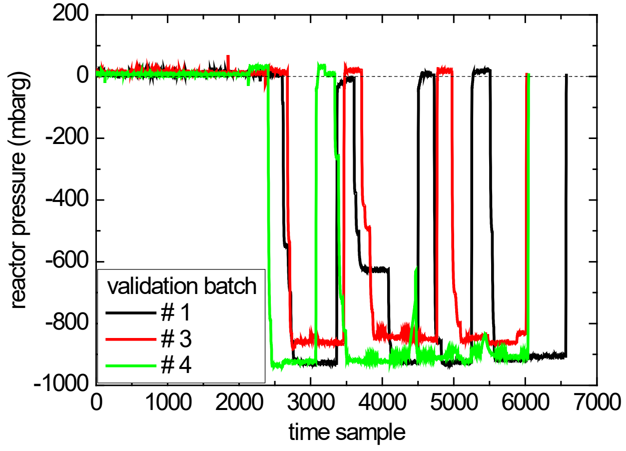
time	manual annotation	quality index 1	quality index 2
21/06/2022 00:00	start of new batch	13.68	31.36
21/06/2022 00:01		10.22	32.02
21/06/2022 00:01		12.21	29.73
21/06/2022 00:02		10.91	25.03
21/06/2022 00:02		12.61	23.38
21/06/2022 00:03		10.79	27.58
21/06/2022 00:03		12.97	27.33
21/06/2022 00:04		13.47	22.19
21/06/2022 00:04	correction	18.42	15.25
21/06/2022 00:05	after correction	11.61	20.62
21/06/2022 00:05		7.72	33.49
21/06/2022 00:06		10.32	26.88
21/06/2022 00:06		14.74	15.85
21/06/2022 00:07		9.85	31.81
21/06/2022 00:07		13.92	23.01
21/06/2022 00:08		12.24	25.65
21/06/2022 00:08		14.87	16.25
21/06/2022 00:09		8.07	37.22
21/06/2022 00:09		11.6	27.84
21/06/2022 00:10		9.58	29.93

Data types

- **Process data**
  - *hardware sensors*
  - *lab analysis*
  - *images*
  - *process analytical technologies*
    - *NIR spectra*
    - *etc…*

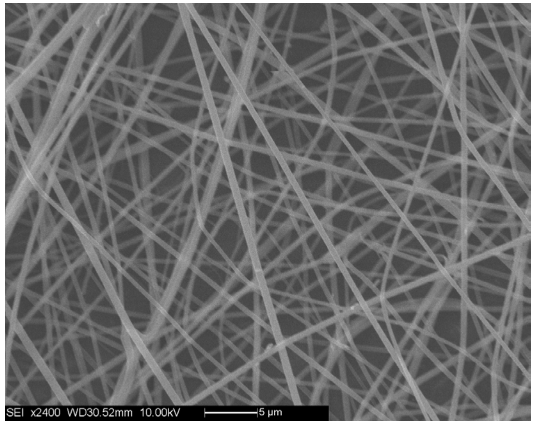
- **Data features**
  - *high quantity (giga/tera/zetta-byte)*
  - *correlation (also in space and time)*
  - *noise*
  - *missing data, etc…*

pressures, temperatures, etc…

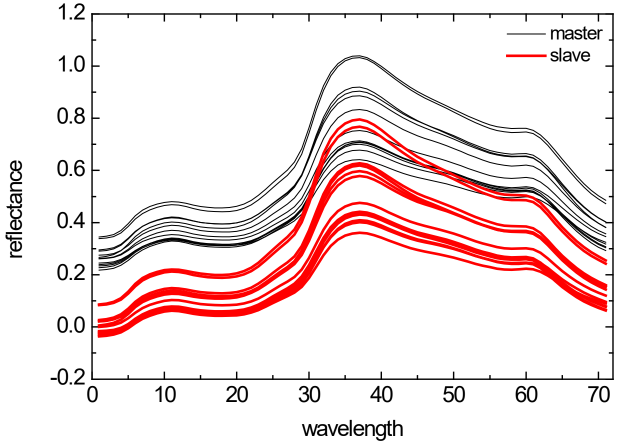


RGB images

SEM images



NIR spectra



"Big data" challenges

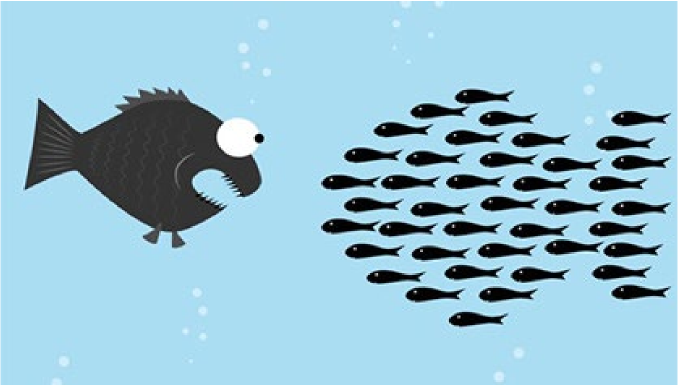
- The big data Vs:
  - volume
  - velocity
  - variety
  - veracity, etc…

Enormous quantity of data (zettabyte = 10^{21} byte) of different types (sensors measurements, images, spectra, etc…) available every second from different sources

Challenges and opportunities:
- *data transmission*
- *data integrity and storage*
- *data analytics*
  - *treatment and mining*
  - *integration*
  - *revision, etc…*

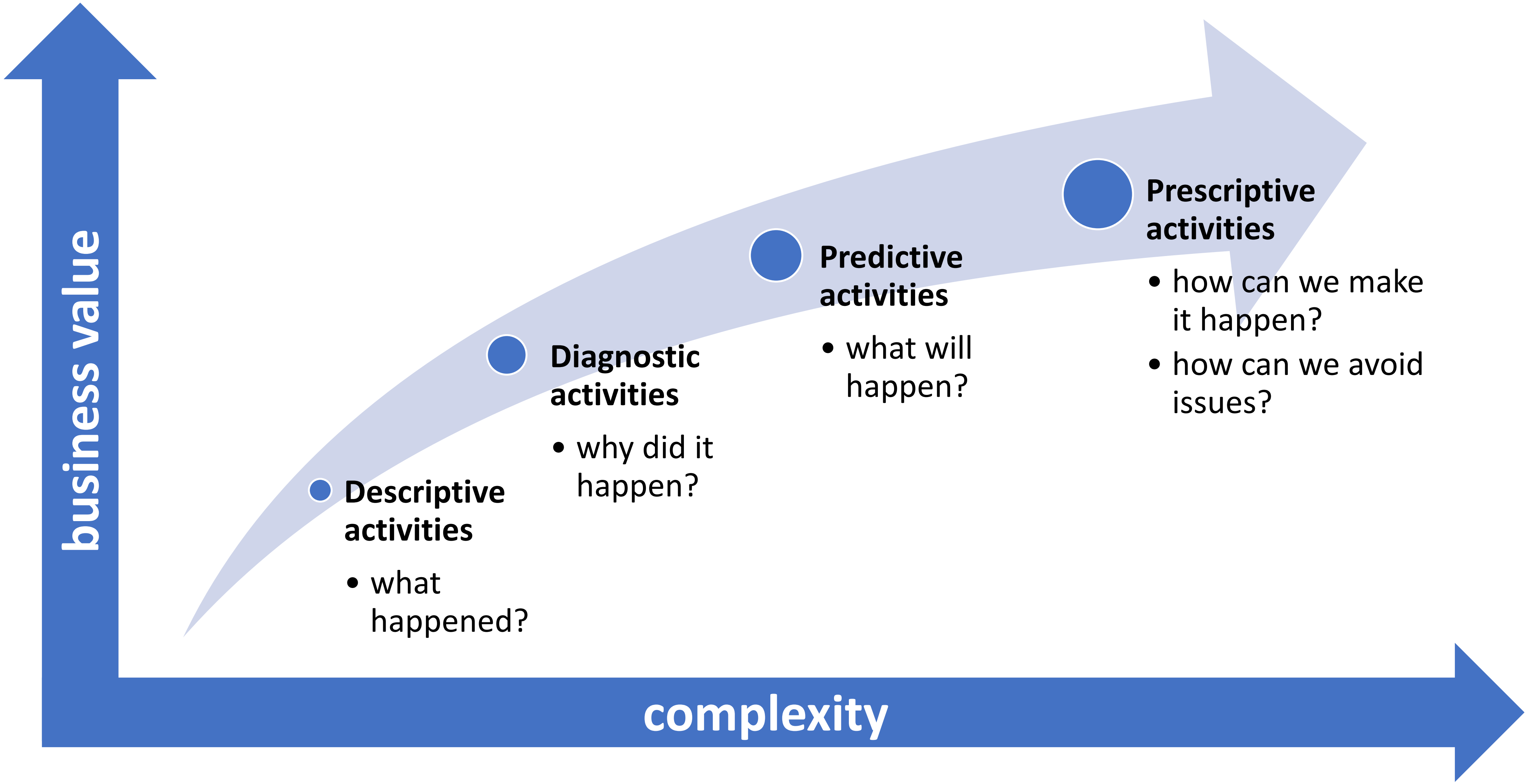
… and the “small data” challenges!

- The **small data** Vs:
  - scarce volumes
  - «batch» velocity
  - non-structured variety
  - veracity to be tested, etc…

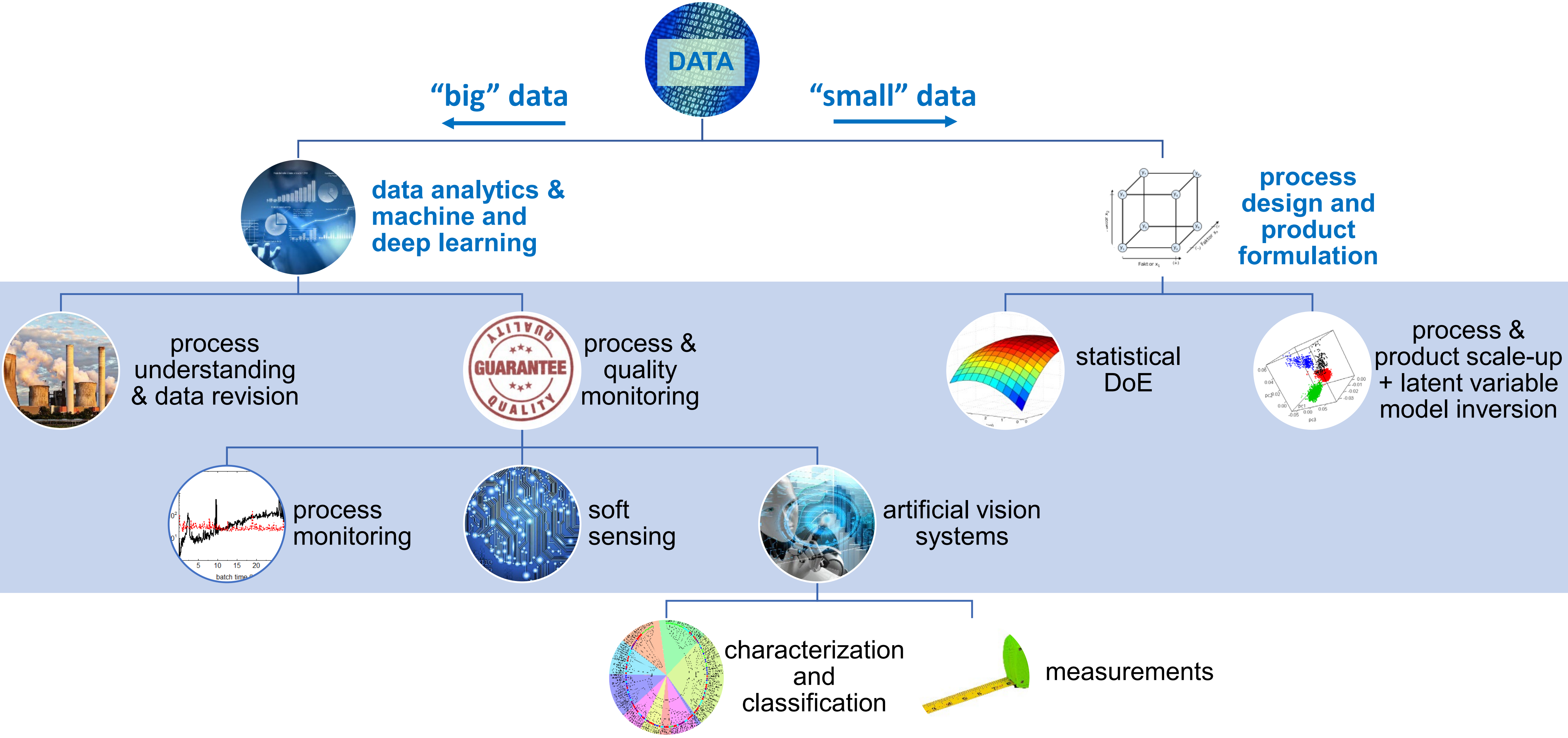


- Initial stages of **product and process development**:
  - few data
  - need of scaling up product and process from the lab scale to the commercial one
  - extended experimental campaigns at laboratory scale
  - few experiments at the pilot scale
  - fast launch of the product in the commercial-scale plant production
    - product and processes often not optimal
    - **experiments not designed at all**

Improving the business value through digitalization



Outline



DATA
"big" data
"small" data
data analytics & machine and deep learning
process design and product formulation
process understanding & data revision
process & quality monitoring
GUARANTEE
QUALITY
statistical DoE
process & product scale-up + latent variable model inversion
process monitoring
soft sensing
artificial vision systems
characterization and classification
measurements

Summary

- Successful industrial applications of:
  - process understanding and troubleshooting
  - process monitoring and product quality monitoring
  - artificial vision systems, spectroscopy and data fusion
  - product formulation, process design and process/product/technology transfer and scale-up

- Sectors:
  - pharmaceutical and bio-pharmaceutical industry
  - manufacturing industry
  - plastics manufacturing
  - semiconductor manufacturing

Process understanding & troubleshooting

Periodic data revision

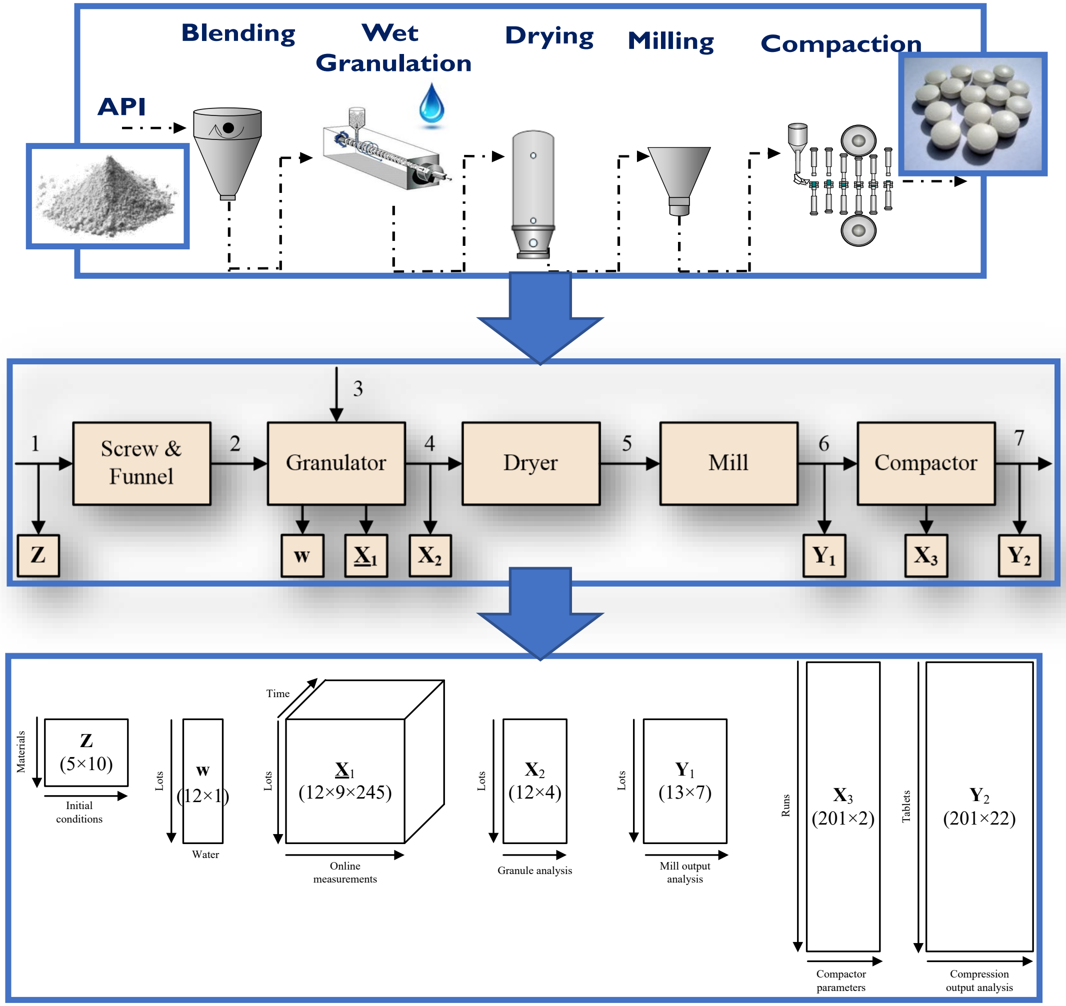
Cloud technologies

PROCESS UNDERSTANDING AND TROUBLESHOOTING

IN-DEPTH PROCESS **UNDERSTANDING**, IDENTIFICATION OF THE PROCESS PARAMETERS MOST AFFECTING PROCESS PERFORMANCE, STUDY OF THE QUALITY VARIABILITY, RAW MATERIAL CHARACTERIZATION, PERIODIC DATA REVISION

Process understanding/troubleshooting

- Initial data management and available datasets organization
- Process understanding
  - granulation for tableting
  - production of plastic materials
  - production of toothpaste
- Process troubleshooting
  - production of plastic material
  - production of ovens
- Periodic historical data revision
  - granulation and drying

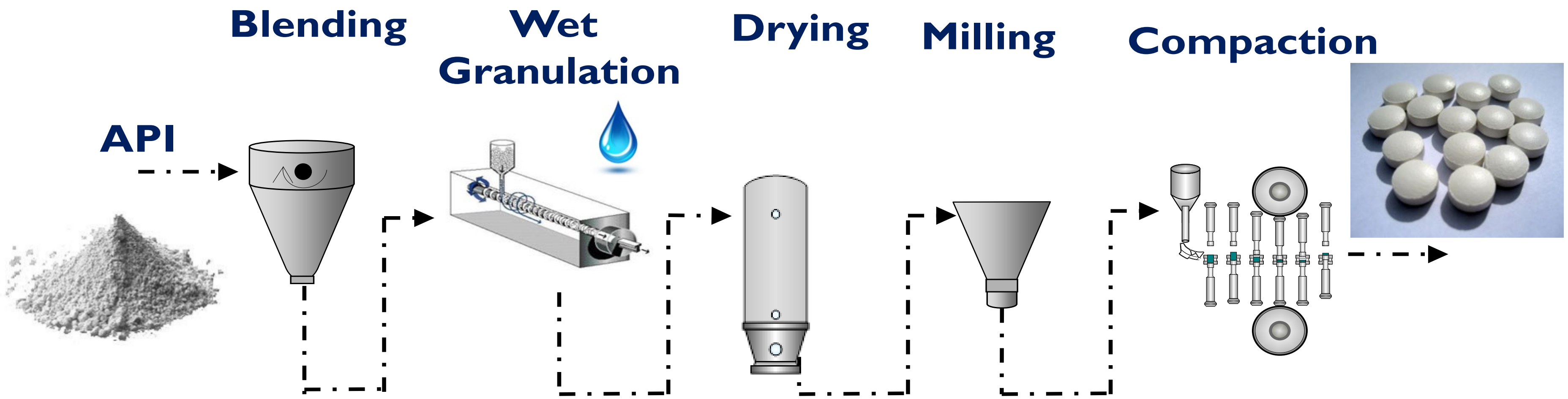


Process understanding in a continuous tableting process

in collaboration with: GlaxoSmithKline
Product Development R&D, Harlow (U.K.)

Tomba et al. (2013), *Int. J. Pharmaceutics*, 444, 25-39

Paracetamol tablets production



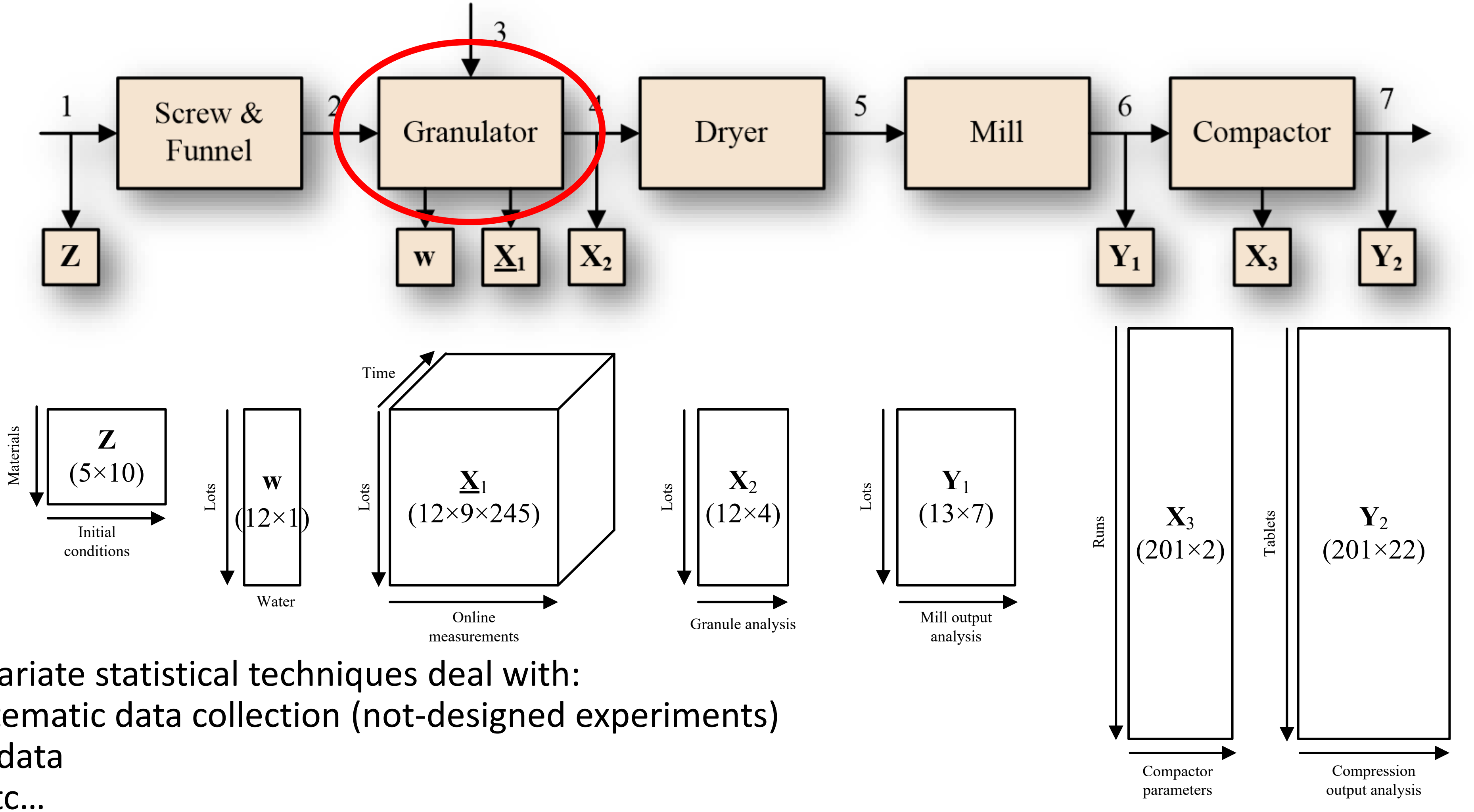
Issues

- experiments from different:
  - raw materials
  - process parameters
- high process and product variability
- high complexity of:
  - plant layout
  - interaction between consecutive unit operations and materials

Objectives:

- **data management**
- in-depth **process understanding**
  - what are the major sources of variability?
  - how do process parameters influence the product quality?

Data management



The multivariate statistical techniques deal with:

- non-systematic data collection (not-designed experiments)
- missing data
- noise, etc…

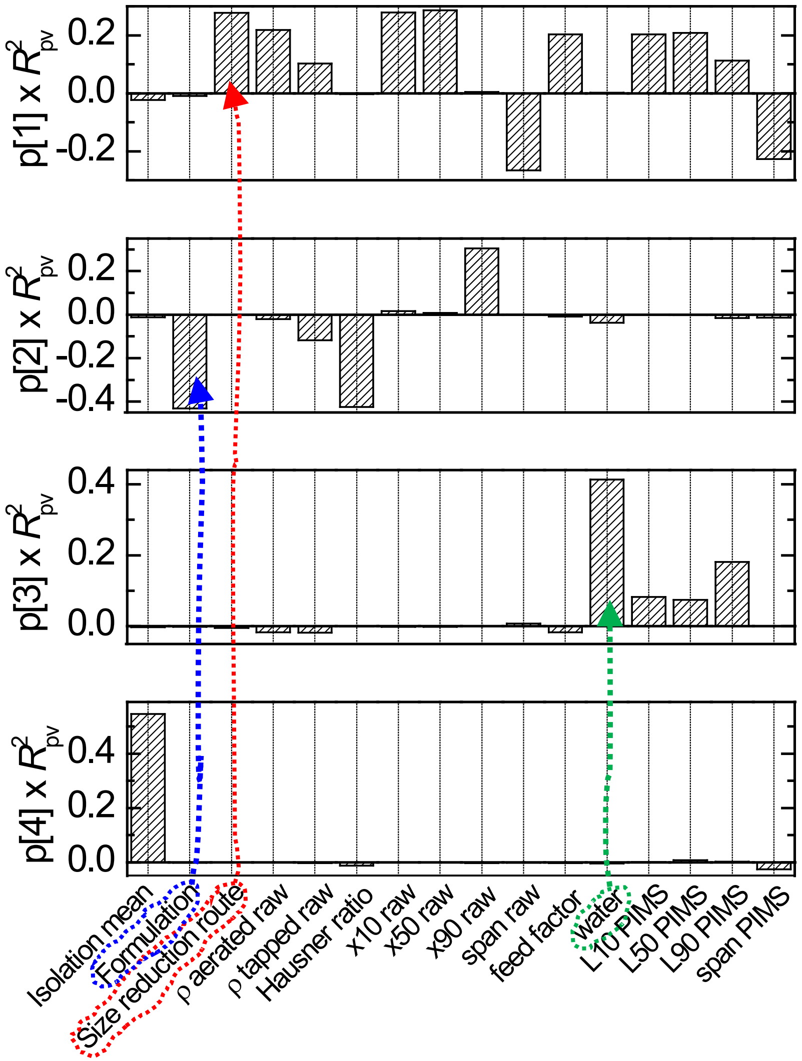
Process understanding

1st PC (52%)

2nd PC (24%)

3rd PC (12%)

4th PC (7%)



- Granulator
  - the major source of variability (52%) is the **way in which API is prepared**
    - correlated to granulate distribution
  - the second driving force is the **formulation** (24%)
    - correlated to powder flowability
  - the third driving force is related to **addition of H_2O** (12%)
    - correlated to the powder particle size distribution PSD

PROCESS AND PRODUCT QUALITY MONITORING

Process monitoring

Product quality forecasting

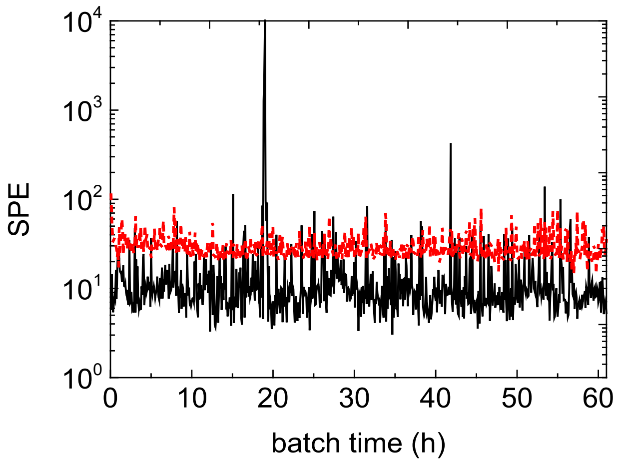
Artificial vision systems

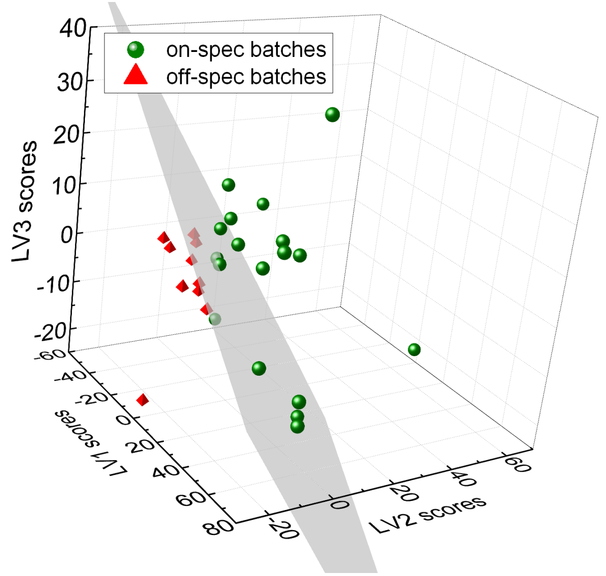
Process analytical techniques

MONITORING THE PROCESS AND THE PRODUCT QUALITY, TIMELY DIAGNOSING THE CAUSES OF **ANOMALIES AND MALFUNCTIONS**, PRODUCT QUALITY **PREDICTION AND CLASSIFICATION**

Process and quality monitoring

- **Process monitoring**
  - production of resins
  - production of avian vaccines
  - granulation and drying
- **Diagnosis** of the anomalies and malfunctions
- **Product quality monitoring, classification, estimation and prediction**
  - soft sensors for manufacturing of resins for coatings
  - production of avian vaccines
- **Adaptive updating** to new process conditions
  - manufacturing of resins for coatings
  - granulation for tablets manufacturing





Batch process monitoring for avian vaccines production

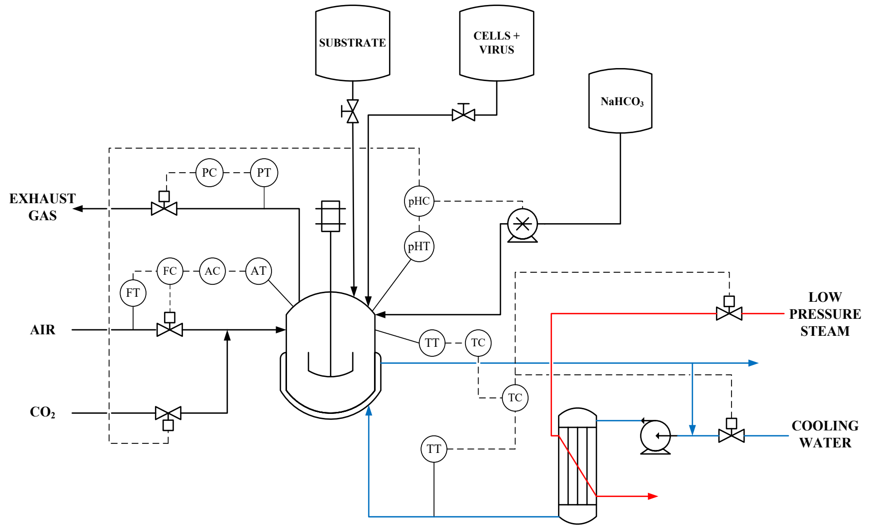
in collaboration with: Merial – a Sanofi company

Noventa Padovana (PD, Italy)

Largoni et al. (2015), *Journal of Biotechnology*, 211, 87-96

Avian vaccines production

- Avian reovirus: severe poultry disease
- Production process: dozens of operating stages
- **Infection** for reovirus multiplication:
  - 300/600 L bioreactors
  - suspension of H_2O, chicken embryo fibroblasts, substrate, additives, etc…
  - pH, temperature, pressure and dissolved O_2 are kept under automatic control



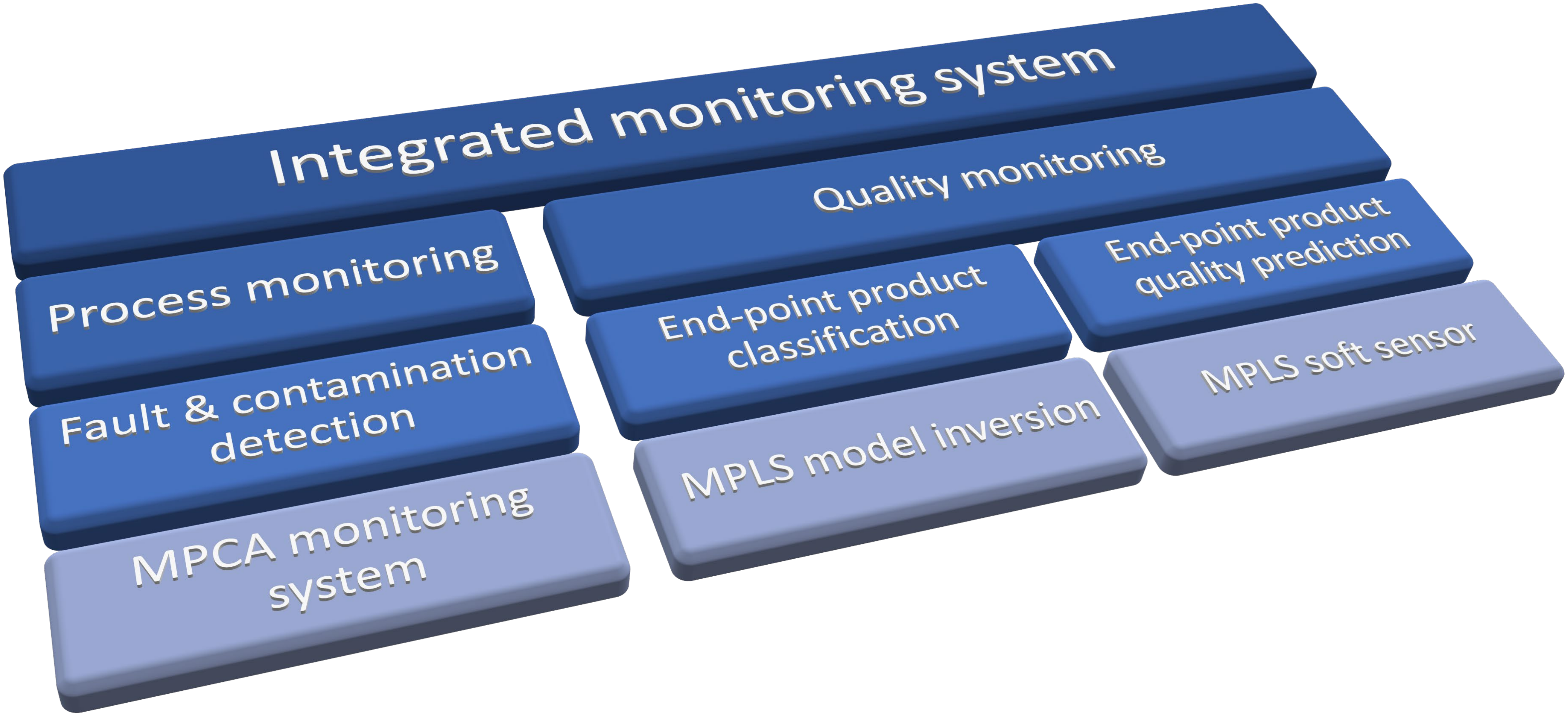
Issues:

- **difficult process control**
- few variables are surveilled by operating personnel
- **fixed recipe and batch length** (61 h)
- **final viral titer** is influenced by raw materials and process parameters:
  - lab analysis available **15 days after batch completion**
  - 30% of the batches are out-of-specification

Objectives:

- **real time monitoring** of the batch runs:
  - diagnose malfunctions and anomalies causes
- predict **final product quality**
- reduce scraps and reworks

Architecture of the monitoring system

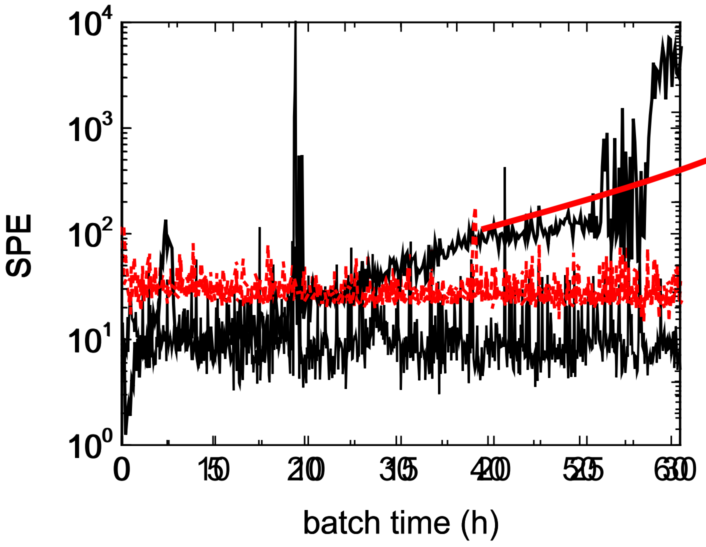


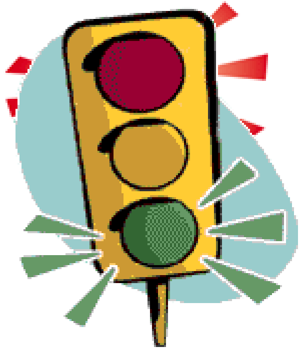
Process monitoring

- Surveille if process is running in standard operating conditions:
  - **detect anomalies, malfunctions and contaminations**
  - **diagnose** the causes of anomalies

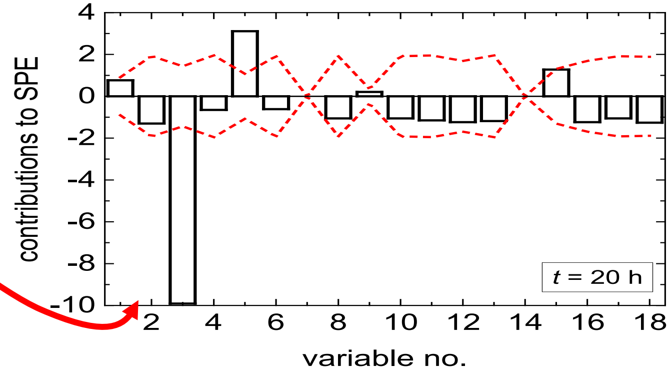
standard operating conditions

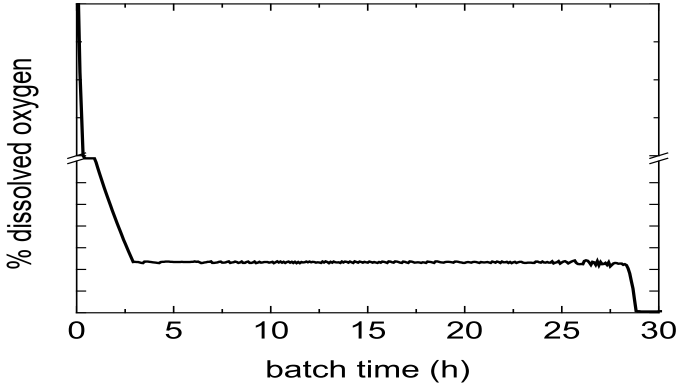
contamination





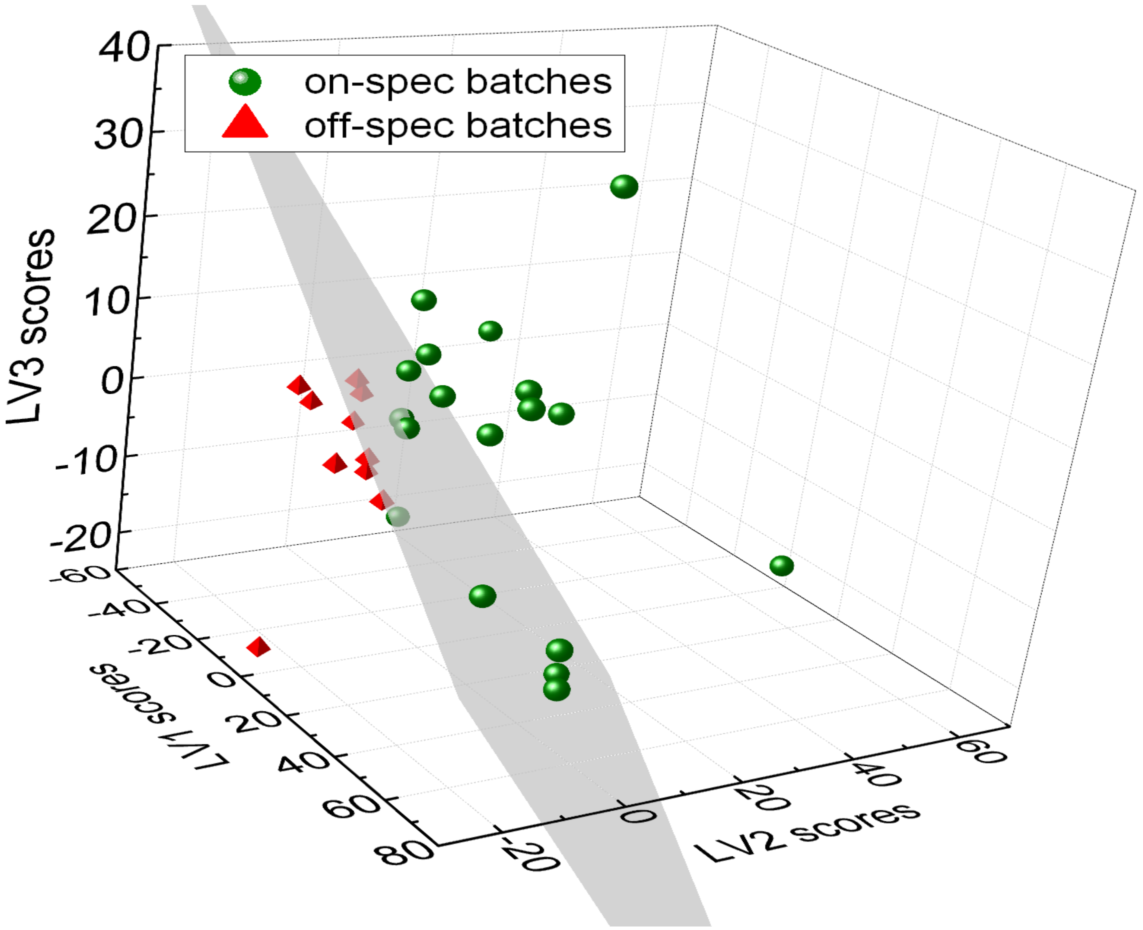
diagnosis





Quality monitoring and classification

- Anticipated out-of-spec detection:
  - optimal product classification
  - accurate classification already **30-40 h after batch start**



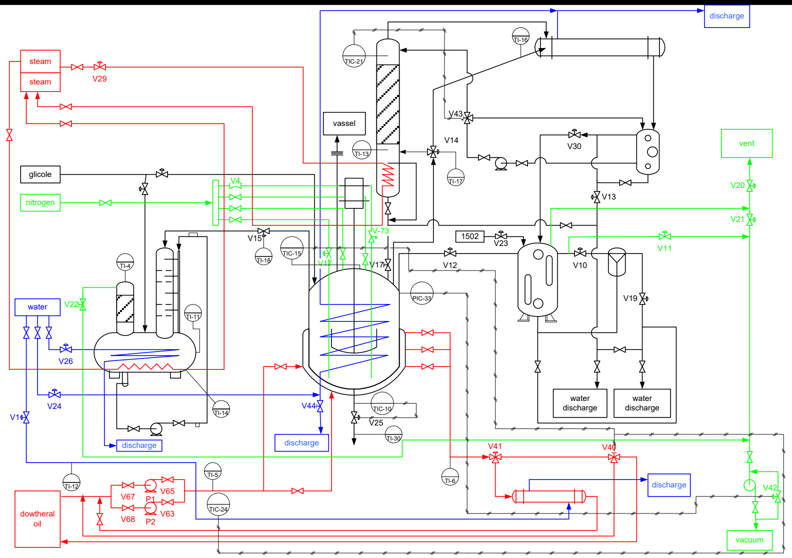
Online product quality monitoring

in collaboration with: SIRCA – Resins and coatings
San Dono di Massanzago (PD, Italy)

Facco et al. (2009), *J. Process Control*, 19, 520-529

Batch production of resins for coatings

- Production of **resins for coatings:**
  - semi-batch polymerization
  - 12 m^3 reactors
- Resins' quality indices:
  - **acidity**
  - **viscosity**



Issues:

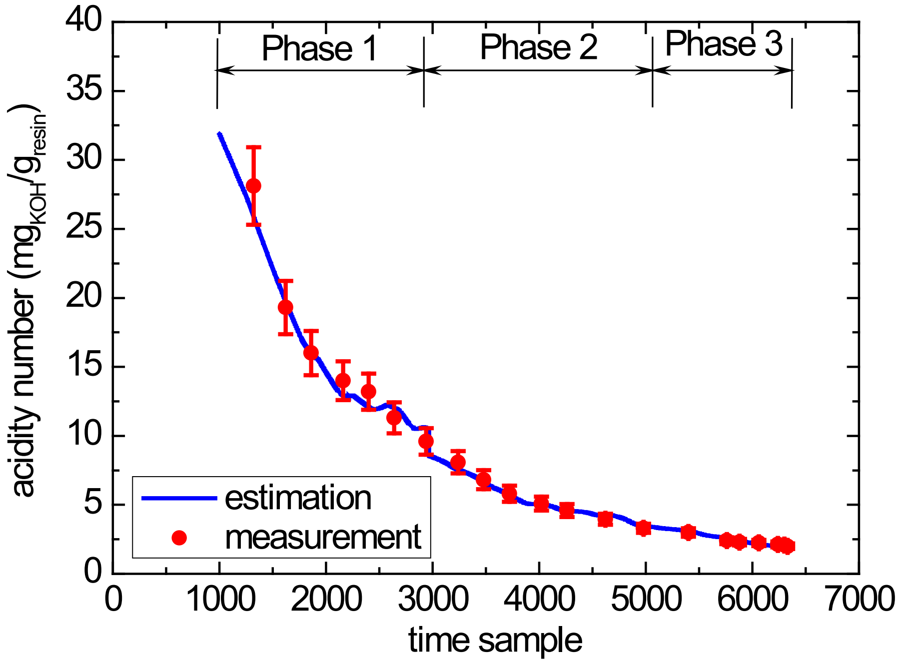
- raw materials variability
- plant operations are carried out manually from operating personnel
- manual measurements of quality indices
  - every 2 h
- manual corrections of the recipe
  - based on operators' experience
- high variability of the batch duration: 40-70 h

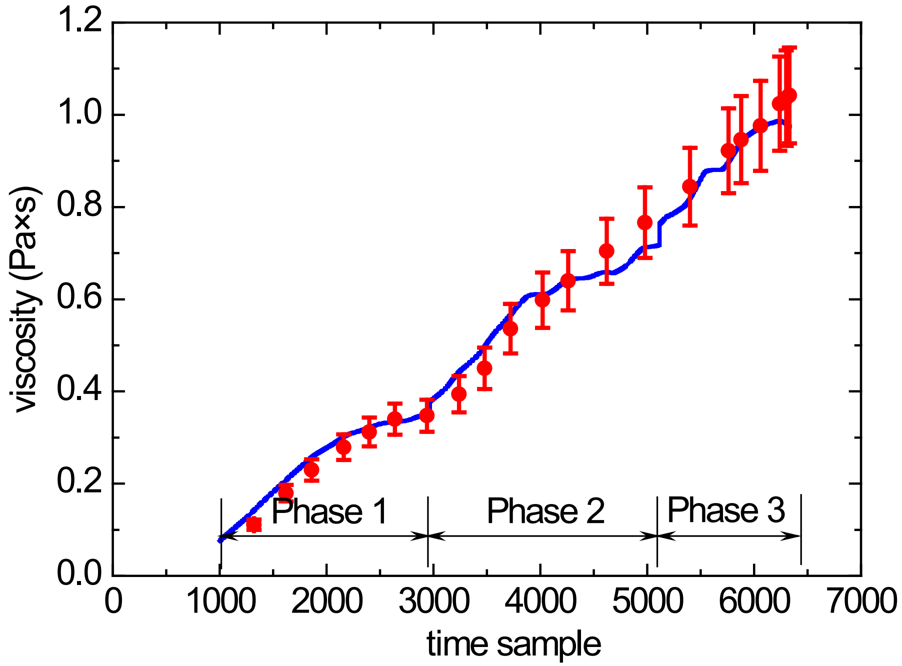
Objectives:

- **real time estimation of the product quality**
  - prompt corrections of the recipe
  - avoid out-of-spec
- **batch duration prediction** from the initial part of the batch
  - production organization
  - labor resources management

Virtual sensor

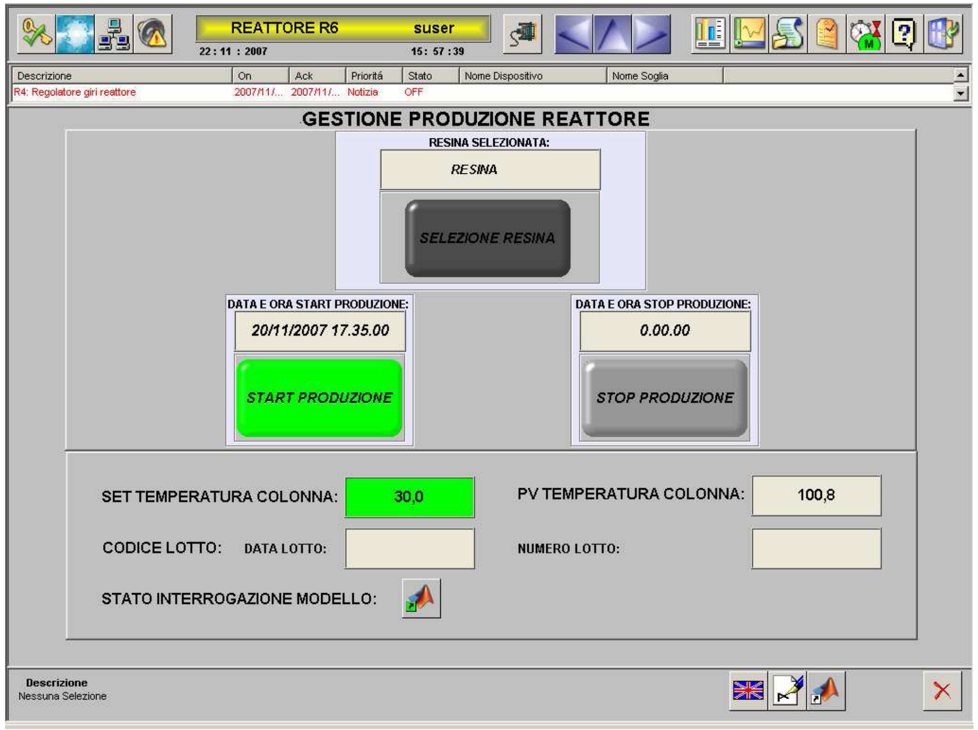
- **Product quality online estimation** (acidity e viscosity)
  - **frequency** every 30 s >> frequency of lab assays
  - **accuracy** = accuracy of lab measurements
- **Batch duration prediction**: accuracy < 4 h already 10 h from the batch start
  - correlated to initial temperature ramps management

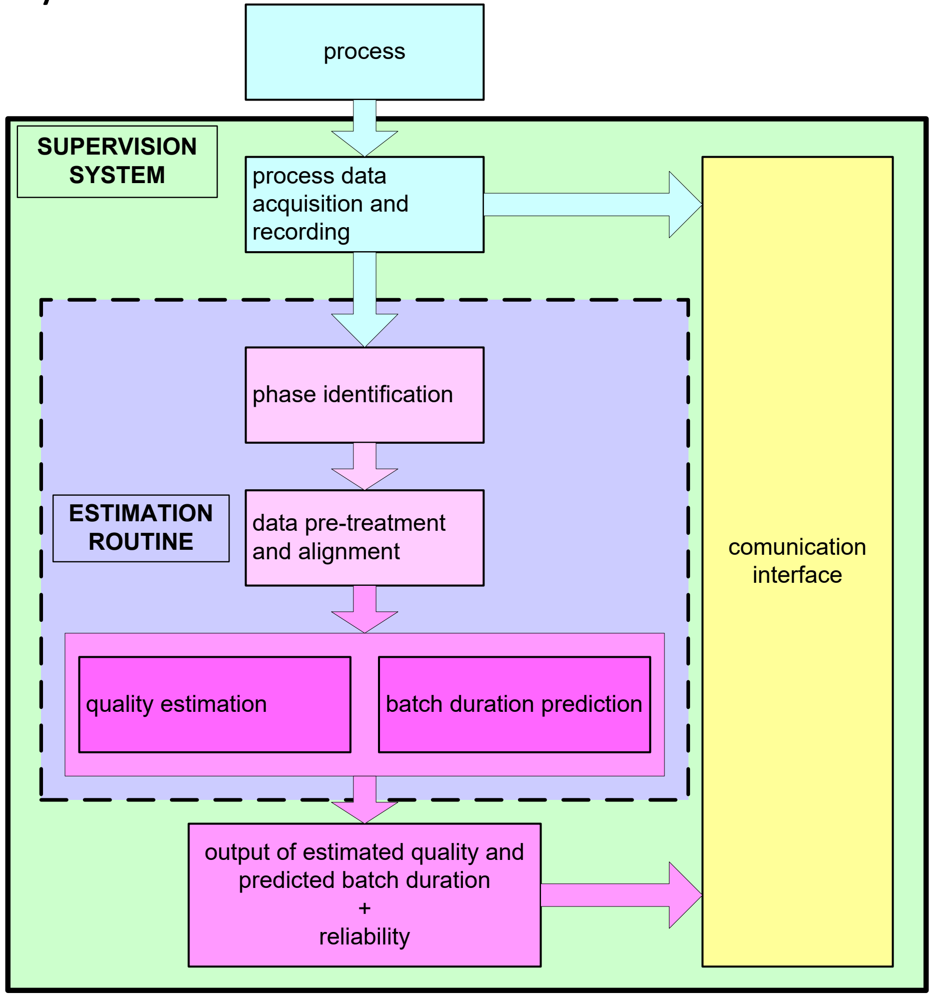




Industrial implementation

- **Physical implementation** integrated to SCADA Movicon® 9.1 supervision system :
  - deal with input data and pretreat them
  - estimates the quality indices and predicts batch duration
  - shows the outputs of the estimations/predictions and their reliability

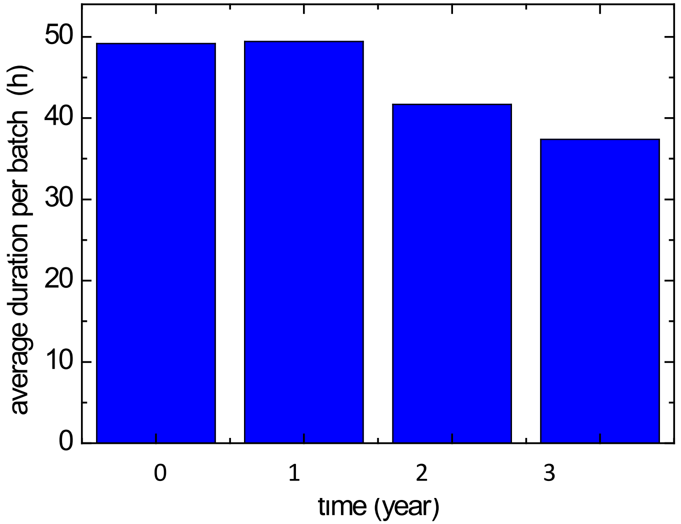


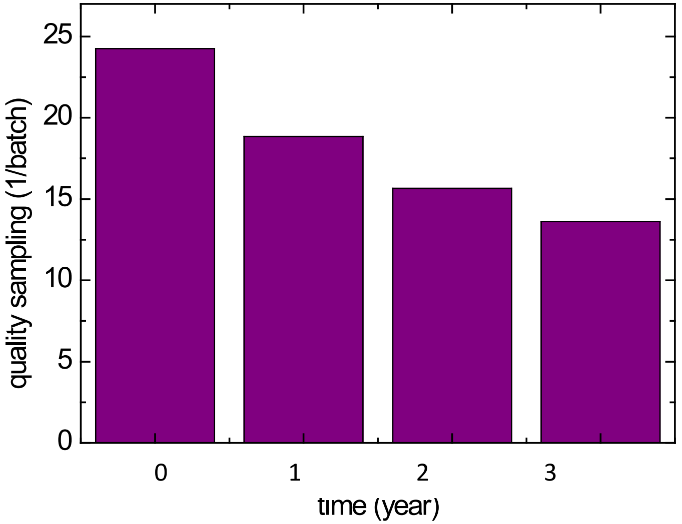


Operating and economic benefits

- "Faster" batches: -10 h duration
- Lower number of lab assays: -10 samples/batch
- Total savings:
  - *1000+ lab measurements*
  - *100+ h process*

- **Increased production: 250 000 kg/year**
- **Saving: 340 h/operators**
- + materials, instruments, etc.





Predicting drug solubility in organic solvents through high-throughput experimentation

in collaboration with: GlaxoSmithKline
Stevenage (U.K.)

Cenci et al. (2024). *Int. J. Pharma.*, 124233


DRUG SOLUBILITY IS A
CRITICAL QUALITY
PARAMETER TO BE
EVALUATED TO
UNDERSTAND DRUG
PERFORMANCE IN THE
HUMAN BODY

Evaluation of API solubility in organic solvents

Product

- Active Pharmaceutical Ingredient API → more than 90% as crystal product

Crucial property: **solubility in organic solvents**

- solvent system (solvent types and temperature)
- crystallization configuration
- amount of cooling required

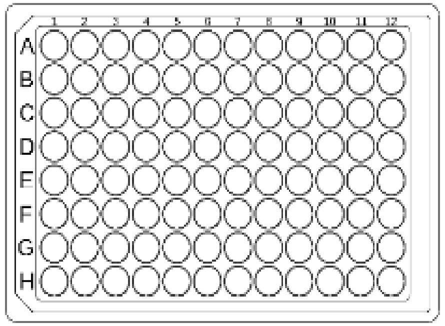
Process

- Crystallisation: **separate and purify API**

Experimental strategy to evaluate solubility

- high-throughput technology: 96-vials plate

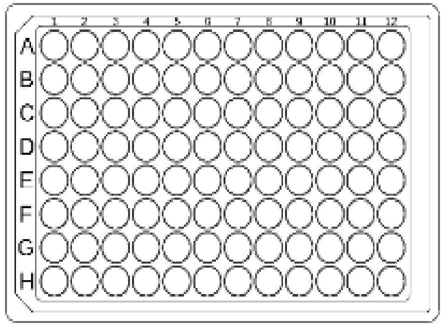
single solvents



- at $T_{\min}$ and $T_{\max}$

based on the **experts' knowledge**

binary mixtures



- at $T_{\min}$ and $T_{\max}$

Issues and objectives

- **Solubility experiments**
  - **unfeasible** to measure **all** possible mixtures of organic solvents
  - new and complex API molecules → data often **not** available **in literature**

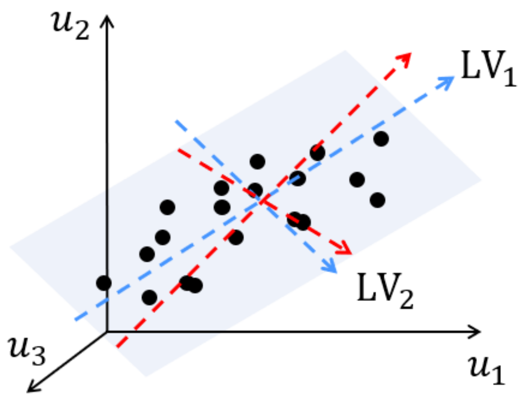
- **Solubility models**
  - a few solubility data for calibration, model predictions for the complete screening
  - **accurate prediction** of solid-liquid equilibria (SLE) is still challenging

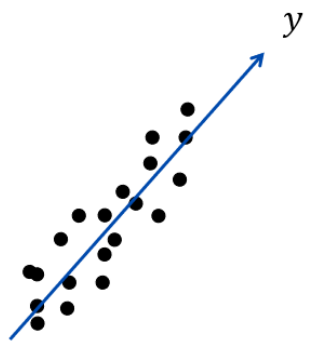
- **Thermodynamic models**
  - a few predict the correct order of magnitude
  - not enough **accuracy** for design purposes

- **Data-driven models**
  - QSPR (quantitative structure property relationship): relate solubility to molecular descriptors
    - mainly in **water**; a few organic solvents and even less binary or ternary **mixtures**
    - **>100 molecular descriptors**, issues: selection, decorrelation
    - **temperature** often not explicit
    - complex nonlinear models → a lot of **data** needed

Objective: accurately predicting **API solubility** in **mixtures of organic solvents** commonly employed for crystallisation

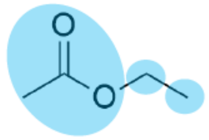
Solubility prediction through PLS





- **Input matrix U**
  - **temperature** explicitly included
  - molecular descriptors (**UNIFAC subgroups**) to identify solvent types:
    - weighted on mixture composition
- **Output vector y**
  - API solubility as $\log(S)$

Ethyl Acetate



n-Heptane

$$g_k = \sum_{i=1}^{N_o} \upsilon_{ik} x_i$$

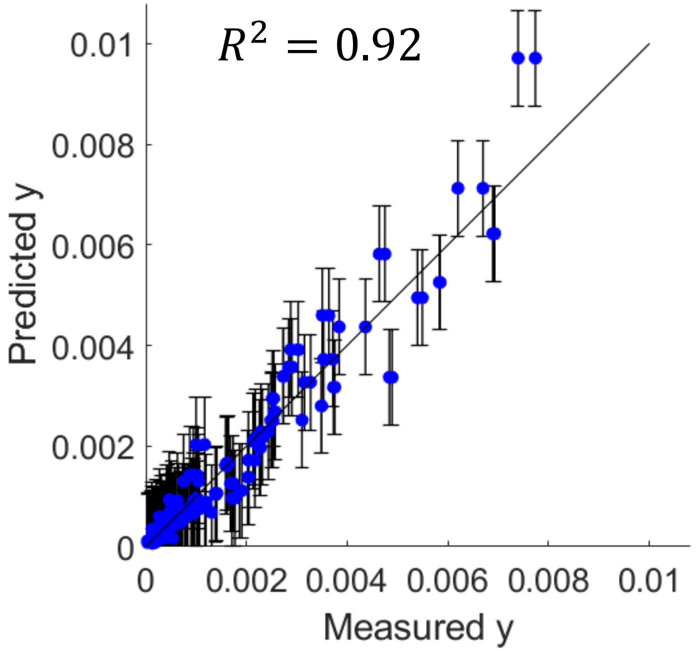
molar fraction of solvent i in the mixture

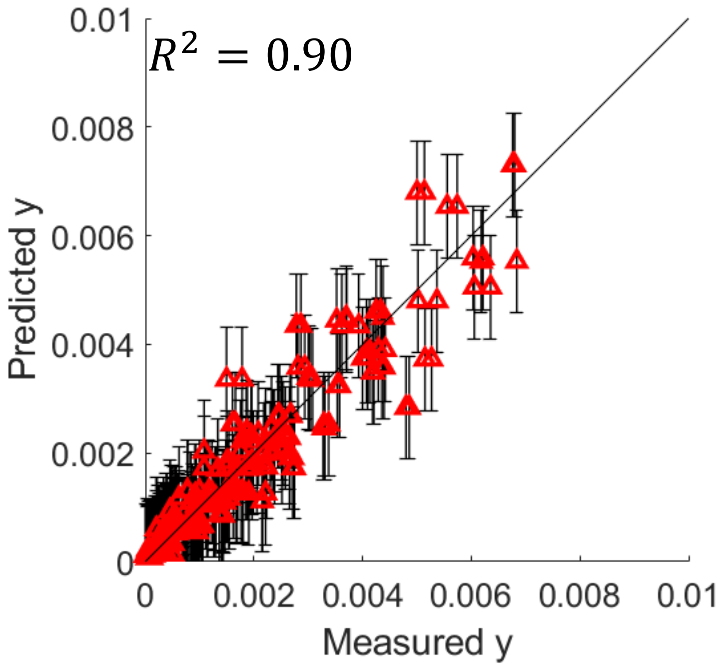
no. occurrences of subgroup k in solvent i

regressors:

$$g_1, \dots, g_k, \dots, g_{N_k}, g_1 g_2, \dots, g_{N_{k-1}} g_{N_k}, T$$

Experimental results: 1 API with 14 organic solvents





- **Satisfactory prediction accuracy**
  - similar performance in calibration and validation
  - **reliable in extrapolation**:
    - 50°C
    - different binary mixtures
    - ternary mixtures

- **178 calibration observations**
  - single solvents
  - few binary mixtures

one plate
at 20, 40°C

- **288 validation observations**
  - single solvents
  - binary mixtures **with different composition** and/or solvent types
  - **ternary mixtures**

at 20, 40, **50°C**

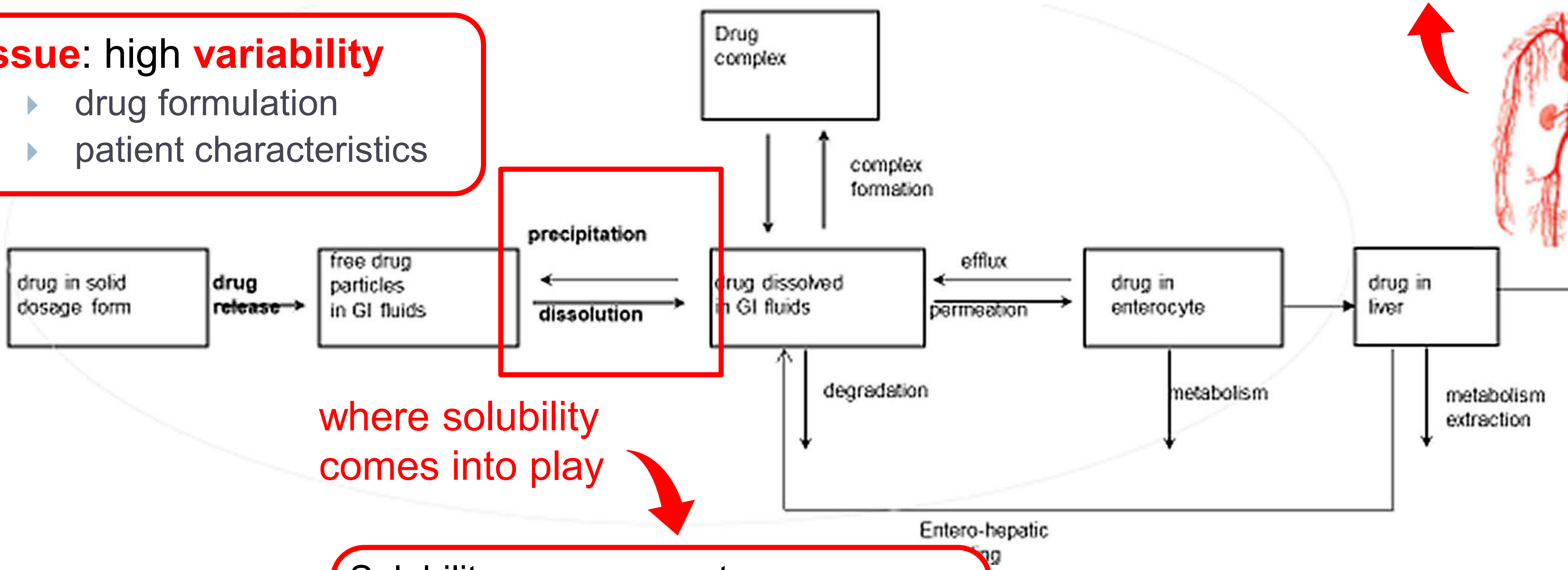
VALIDATED WITH 8 DIFFERENT TYPES OF DRUGS MINIMIZING THE **COST AND DURATION OF THE EXPERIMENTAL CAMPAIGNS**

Pharma industry 5.0

- Evaluating drug movement from administration to **bloodstream** to understand the sorption **rate and extent**

Drug concentration measurements
- often as plasma-concentration profiles

Issue: high **variability**
- drug formulation
- patient characteristics



where solubility comes into play

Solubility measurements
- seldom in Human Intestinal Fluids
- often in Simulated Intestinal Fluids

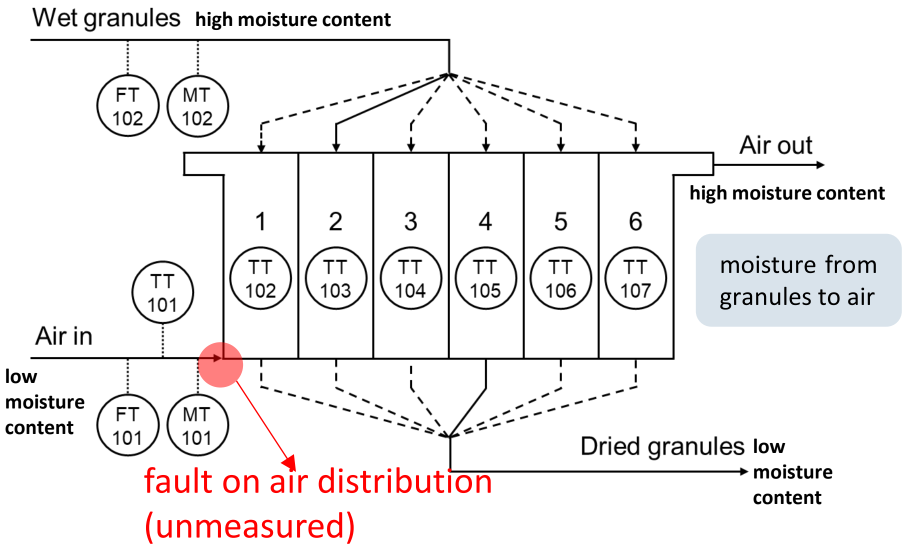
Development of hybrid monitoring systems

in collaboration with: Eli Lilly &C.
Indianapolis (USA)

Destro et al. (2020), *J. Process Control.*, 92, 333-351

Sequential drying of drug products

- Drying of drugs powders in the pharmaceutical industry
  - **production of tablets**



Issues:

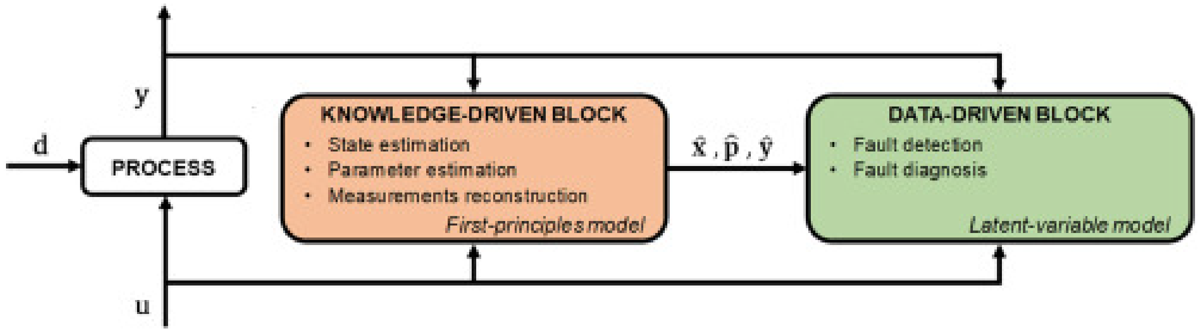
- raw materials variability
- high complexity of process layout
- few process measurements
  - malfunctions on unmeasured variables
- limited control of the process

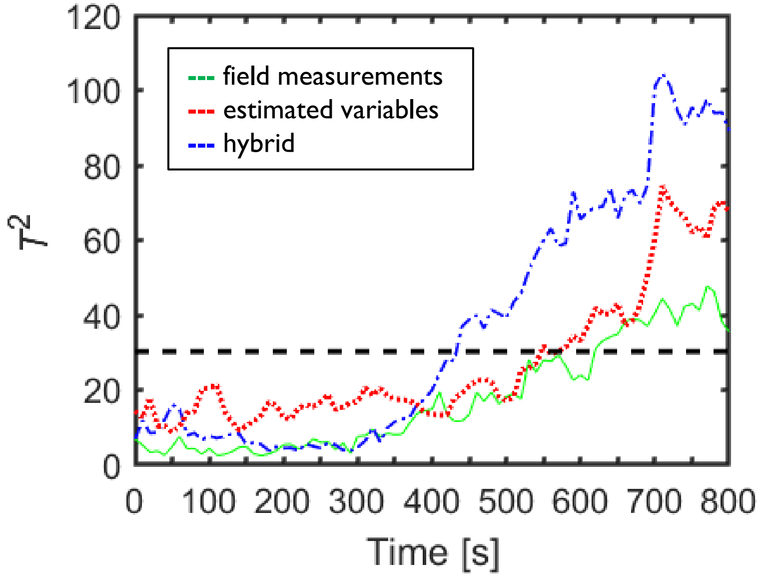
Objectives:

- **real time monitoring of the process**
- **product quality insurance**
- **prompt detection and diagnosis of faults, malfunctions and anomalies** despite few variables are measured online

Towards process metaverse through digital twins

- **Digital twins** are used to improve the monitoring performance
  - measurements are augmented in silico to improve the monitoring capability
- **Outcomes**:
  - prompter detection of anomalies and malfunctions
  - more accurate diagnosis of the faulty causes





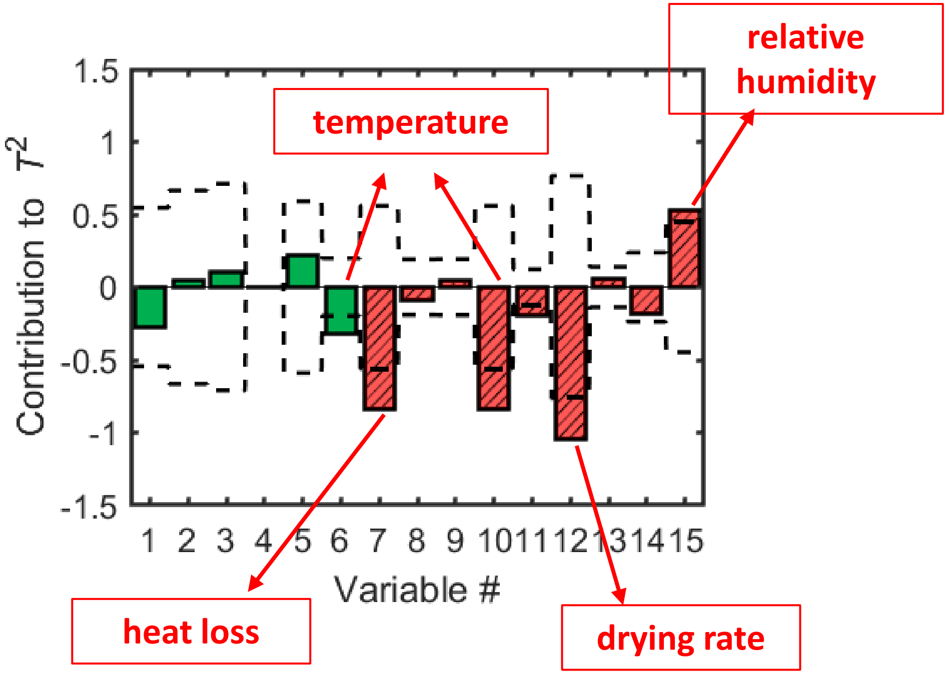
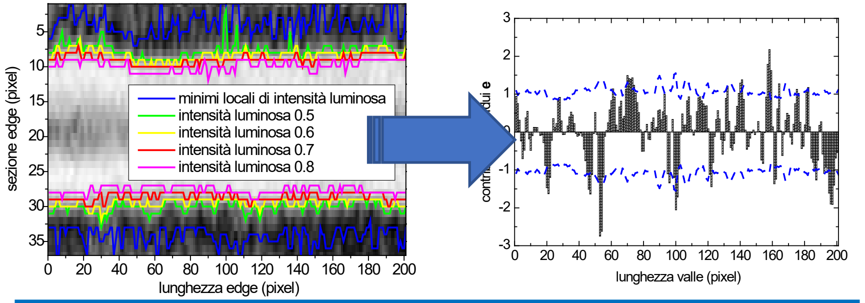
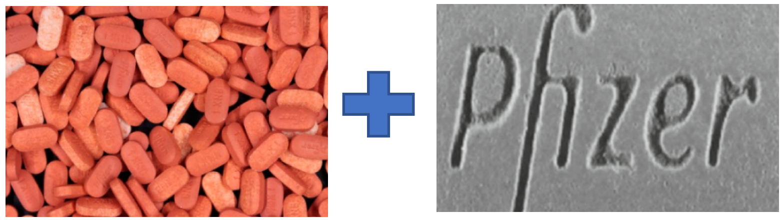


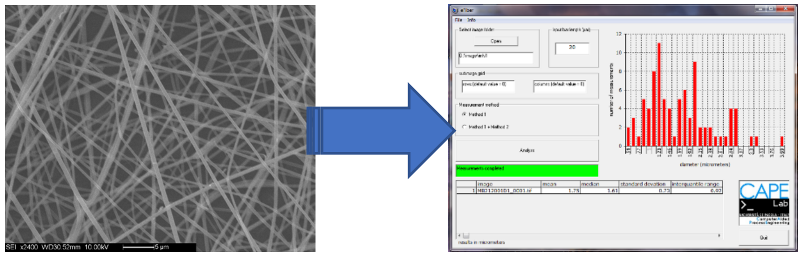
Image analysis, spectroscopy and data fusion

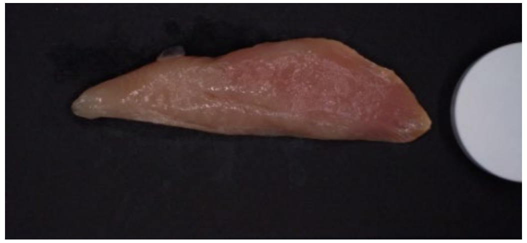
Visual product characterization

- Characterization, classification and measurement of important product features:
  - **defect detection and localization**
  - **roughness**
    - integrated circuits
  - **color & integrity**
    - tablet coating & logo
  - **texture** and chemical-physical characteristics (e.g.: permeability)
    - nanofiber membranes
  - data fusion for **anti-sophistication and anti-adulteration** systems
    - food industry
  - **measurement**
    - particle size distribution PSD of granulate material







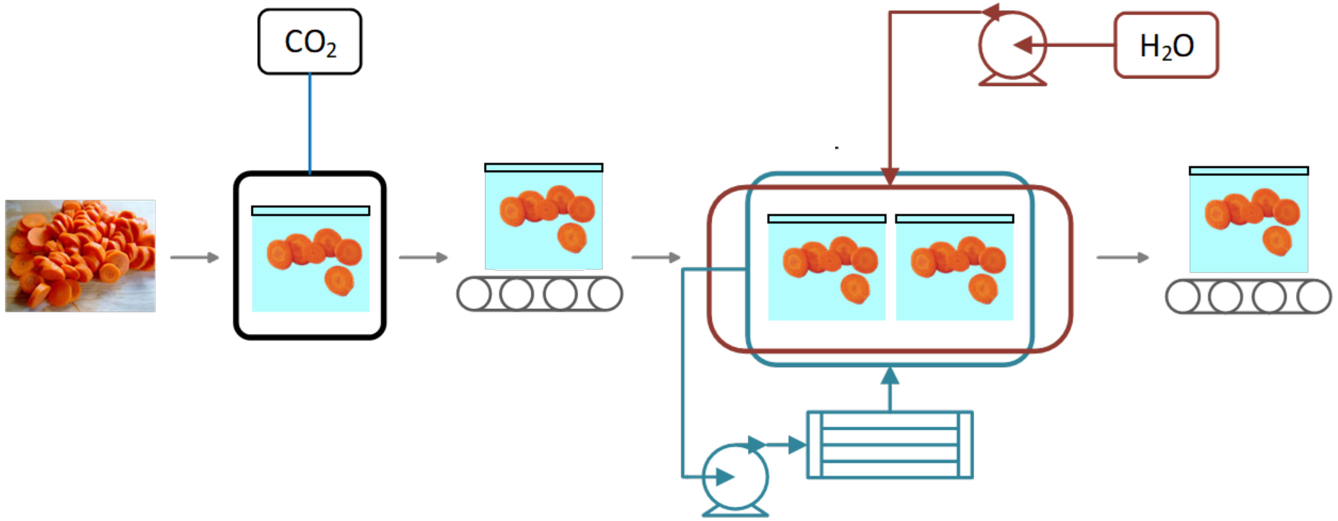


Optimization of the appearance in processed ready-to-eat carrots

Barberi et al. (2021). Foods, **10** (12), 2999

Food pasteurization

- **Food pasteurization**:
  - ready-to-eat food
  - packaging
  - food microbiological safety



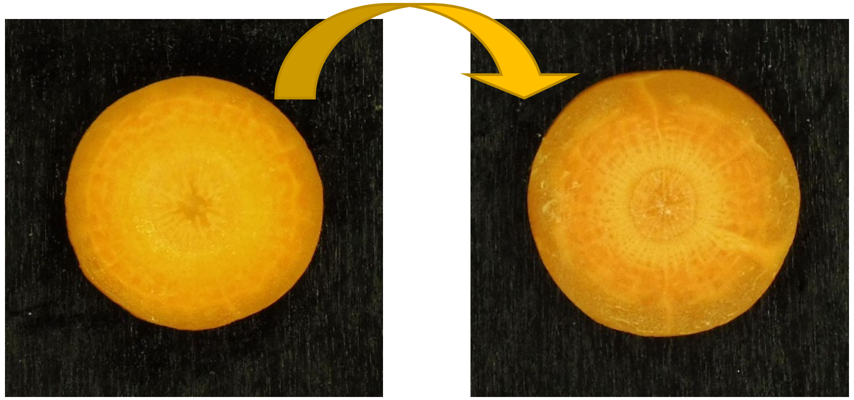
- **Issues**:
  - product appearance
    - product acceptability
  - costly microbiological tests
    - product healthiness and safety

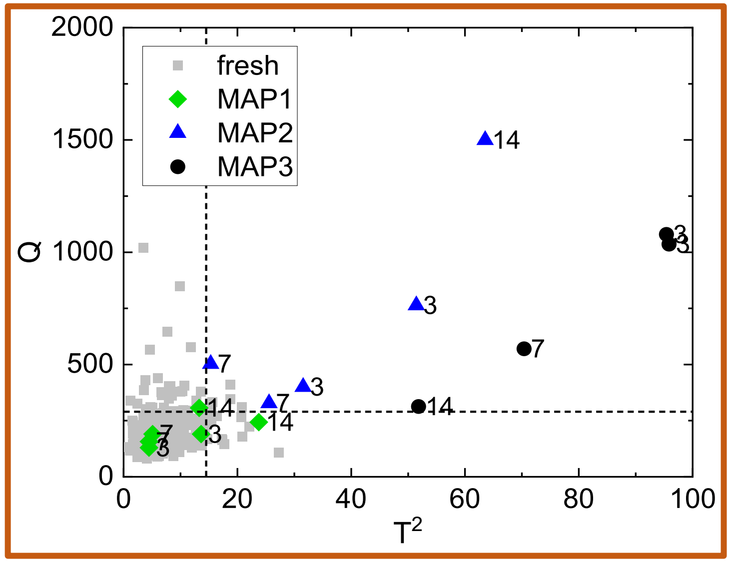
Objective: to develop an artificial vision system:

- to judge if the product appearance is fresh-like
- to minimize the microbial tests in the pasteurization process development

Carrots texture and color for a fresh-like appearance

- Artificial vision system for the product quality characterization
  - time and money are saved in the process development
  - fresh-like, fresh and healthy products are provided
  - microbiological safety is guaranteed





Characterization of granulate materials

Facco at al. (2017). Chem. Eng. Sci., 164, 246-257

Granulate materials

- Industrial fields of interest:
  - pharmaceutical industry
  - plastic materials
  - mining
  - food industry, etc…
- Morphological characterization:
  - particle size distribution PSD
  - granules size, shape, surface features, etc…

Issues:

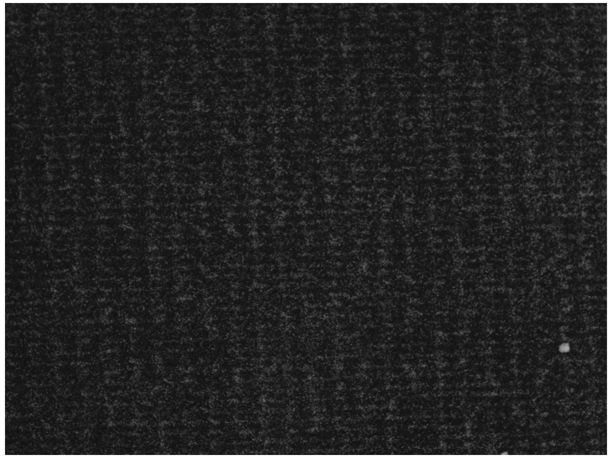
- lack of sensors
- lack of methodologies to analyze the **bulk** materials
- local inhomogeneity
- shuffling effect

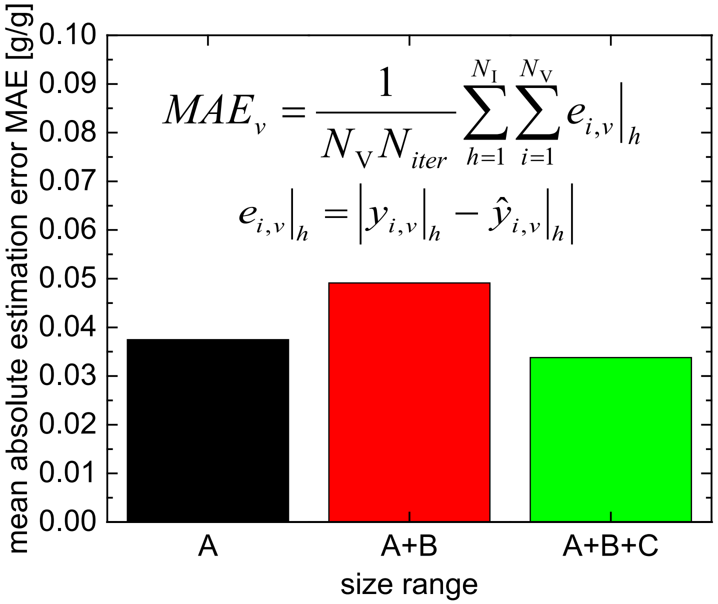
Objectives:

- **automatic measurement system** to characterize the granulate materials in **bulk**
  - both in static and dynamic applications
- **PSD estimation** from images and videos

PSD granulate estimation

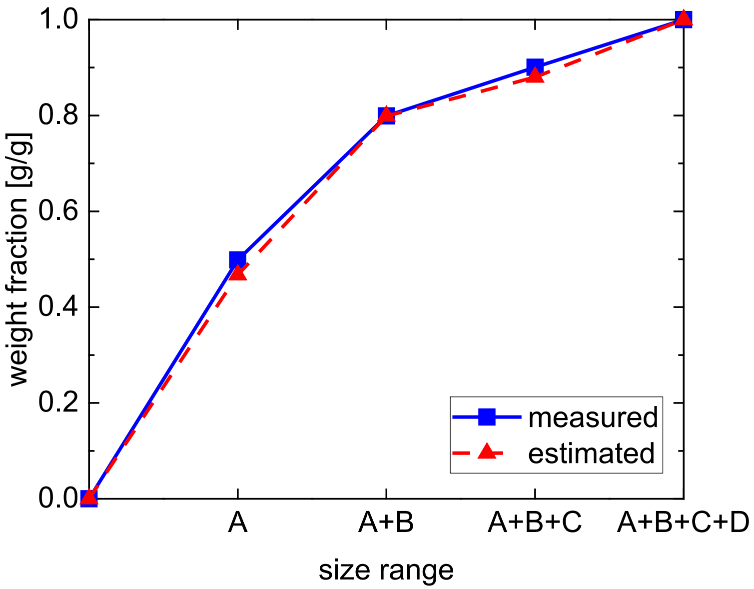
- **Fast and accurate** estimation system
- Works in both **static** and **dynamic** applications
- No significant performance loss:
  - at **different velocity** of the conveyor belt
  - with **segregated materials**





$$MAE_v = \frac{1}{N_V N_{iter}} \sum_{h=1}^{N_I} \sum_{i=1}^{N_V} e_{i,v}\Big|_h$$

$$e_{i,v}\Big|_h = \left| y_{i,v}\Big|_h - \hat{y}_{i,v}\Big|_h \right|$$

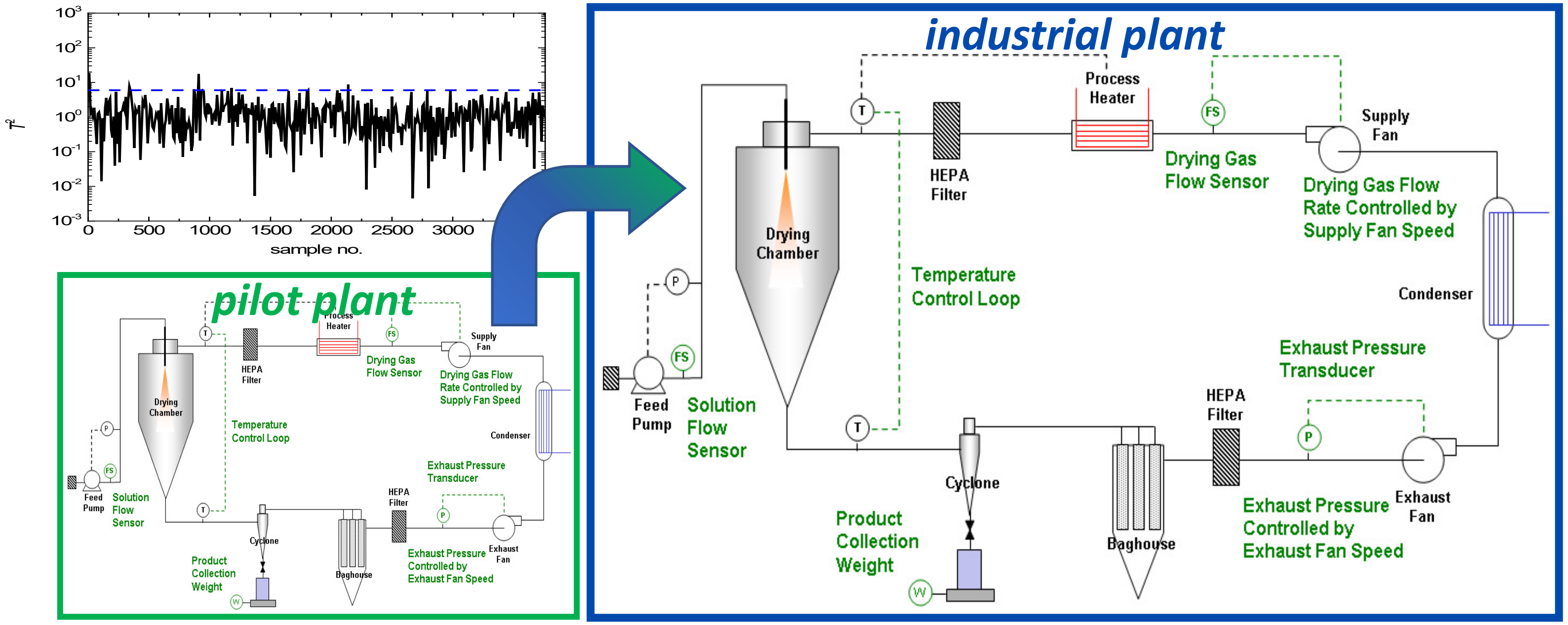


Product and process design, transfer and scale-up/down

Process and product scale-up

- **Product and process scale-up/down:**
  - product scale-up
    - drug carrier for inhalation
  - process scale-up/down
    - food industry: grain milling
    - **biopharmaceutical production of IgG1**

- **Technology transfer** between scales:
  - standardization of NIR spectrometers
  - **monitoring model transfer**
    - spray-drying process



Biopharmaceutical process scale-up/down

in collaboration with: GlaxoSmithKline

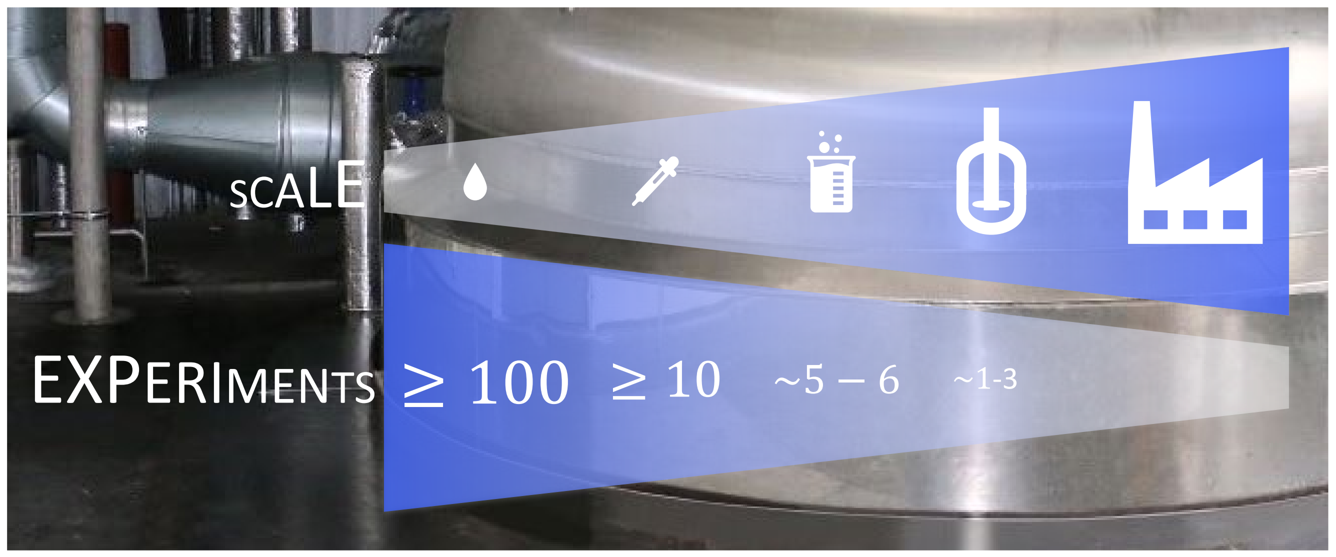
Biopharm Product Development R&D, Harlow (U.K.)

Facco et al. (2020), *Biochem. Eng. J.*, 164, 107791

MONOCLONAL ANTIBODIES ARE AN IMPORTANT CLASS OF DRUGS UTILIZED TO TREAT DIFFERENT TYPES OF DESEASES: INFECTIONS, IMMUNOLOGICAL DESEASES, CANCER, ETC.

MONOCLONAL ANTIBODIES DEVELOPMENT REQUIRES **LONG-LASTING AND EXPENSIVE EXPERIMENTAL CAMPAIGNS**

THE SCALE-UP **FROM LAB-SCALE TO COMMERCIAL-SCALE PLANT** REQUIRES DOZENS OF STAGES TO IDENTIFY THE **MOST PRODUCTIVE AND STABLE CELL LINES**



Biopharmaceutical process

- Biopharmaceutical process for drugs manufacturing

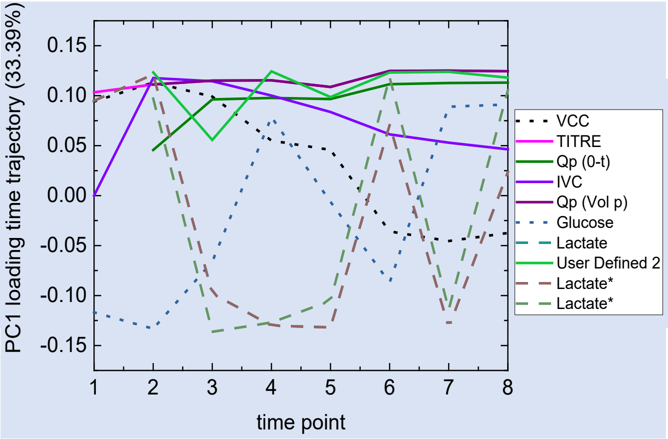
Issues:

- very wide biological variability
- experimental campaigns on different cell lines
- batch productions
- very “slow” process scale-up
  - very demanding experimentation on smaller scales
  - few experiments on large scales

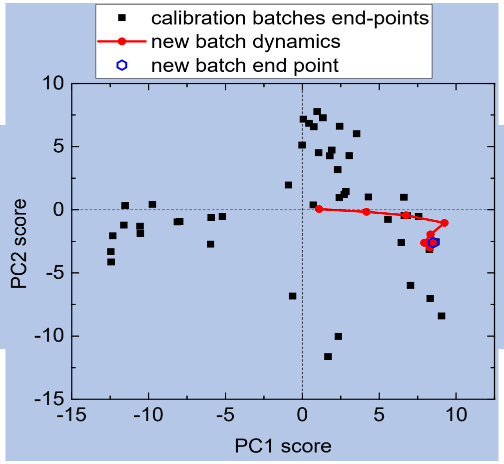
Objectives:

- higher **process understanding**
  - identifications of the major driving forces that act on the process variability
  - characterization of the differences throughout the scales
- identification of the **most promising cell lines** in terms of productivity
- study of the main **similarities and differences among scales**

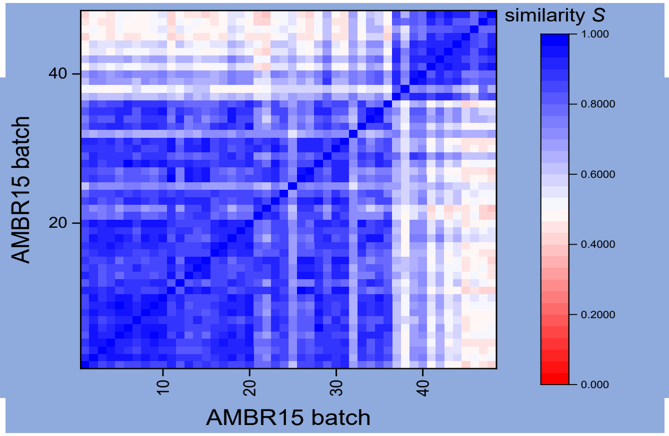
Main results



- Understanding on how the **driving forces** determine the **process variability**



- Identification of the most promising **cell lines** for what concerns productivity

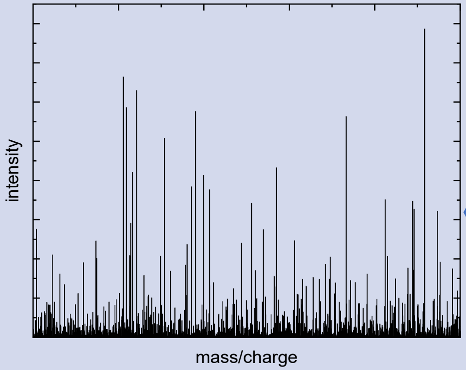


- Identification of:
  - similar batches within scale
  - between scales:
    - **similar physical phenomena**
    - **similar behavior of the clones**

Integration of biological and process information

Biological information on the CULTURE

- Biological understanding
- Culture optimization
- Host engineering



LINK

Engineering information on the PROCESS

- Process development
- Scale-up
- Plant management

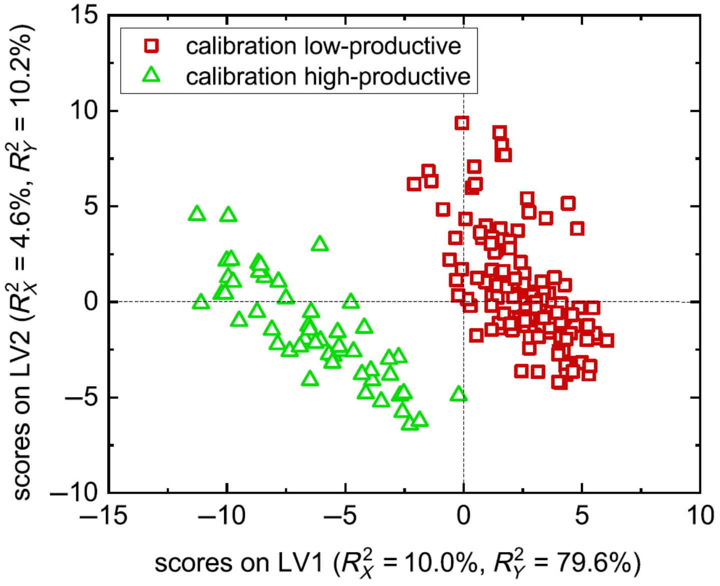
OUTCOMES

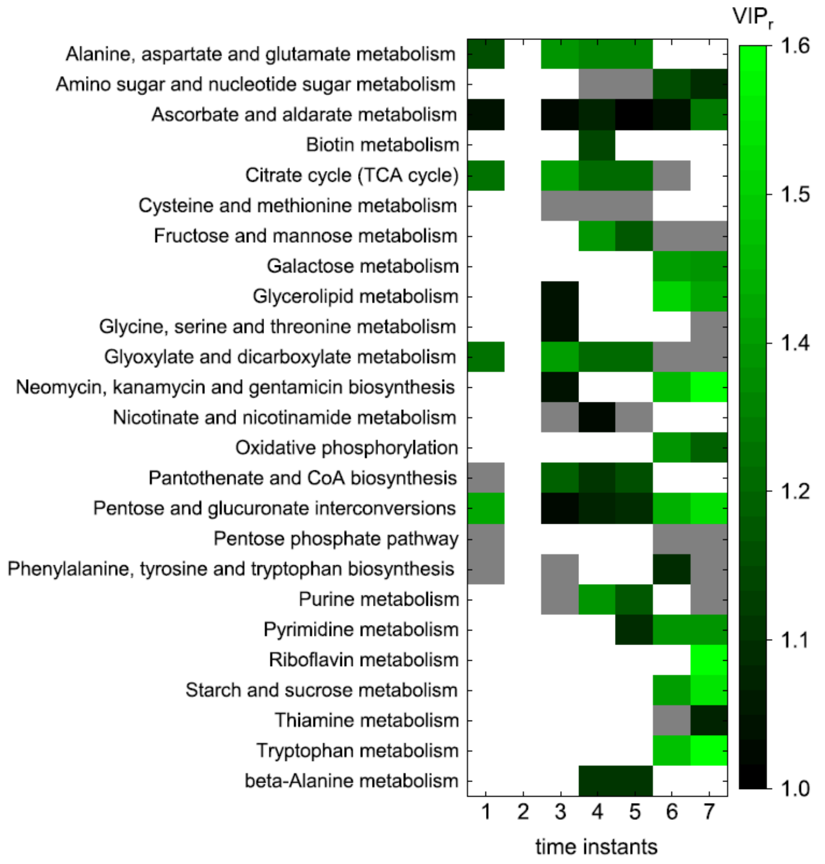
- Selection of **promising cell lines** through the integration of process and biological information (-omics data)
- Process optimization:
  - *development of **optimal feeding strategies***
  - *optimal management of the process based on cell lines metabolic characteristics*
- **Genetic engineering**: management/modification of the cell lines to be more robust to process variability

From cell metabolism to drug productivity

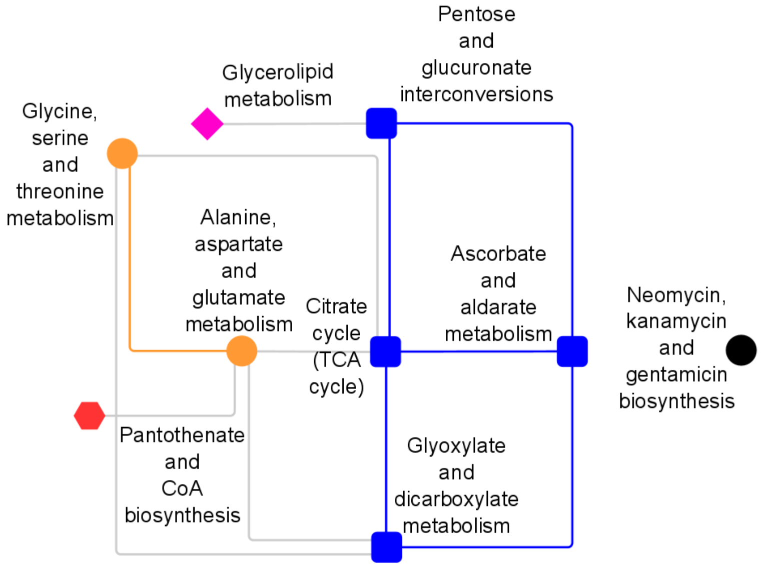
identification of the metabolites related to productivity

early identification of productive cell lines





identification of the metabolic pathways related to productivity



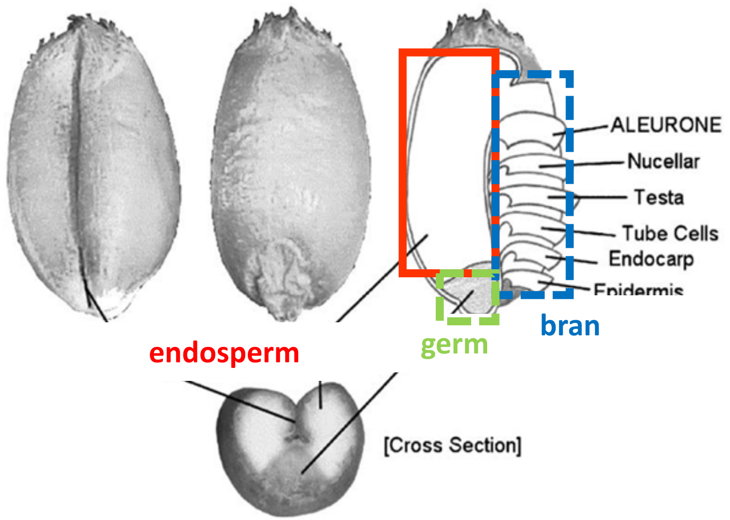
Product and process design in the food industry

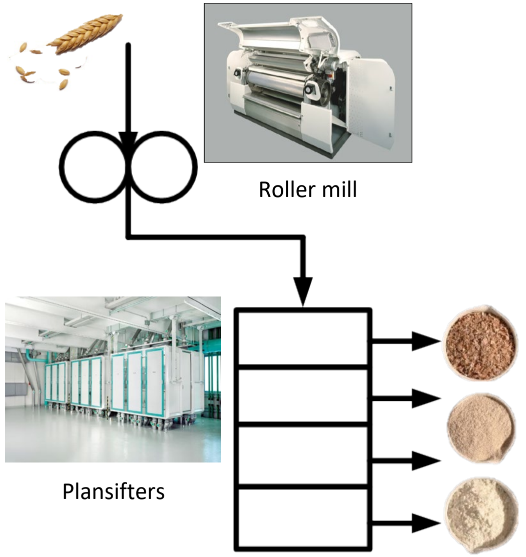
in collaboration with: Buhler
Grain processing division, Utzwill (Switzerland)

Dal Pastro et al. (2017), *Food and Bioprod. Proc.*, 106, 127-136

Grain milling

- **<u>Grain milling objective</u>**: collect as much and as pure endosperm as possible with an appropriate granulate size and at the minimum cost

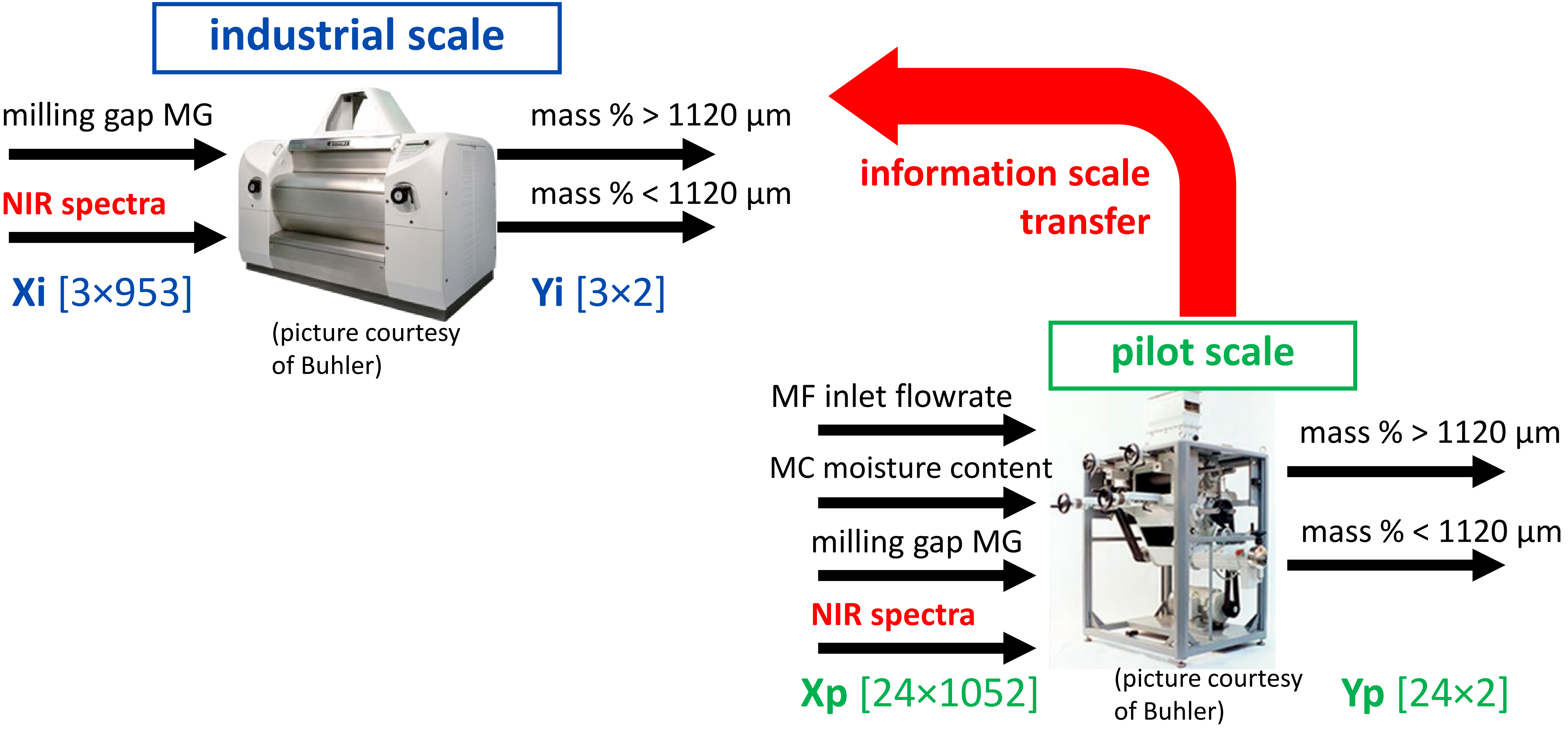




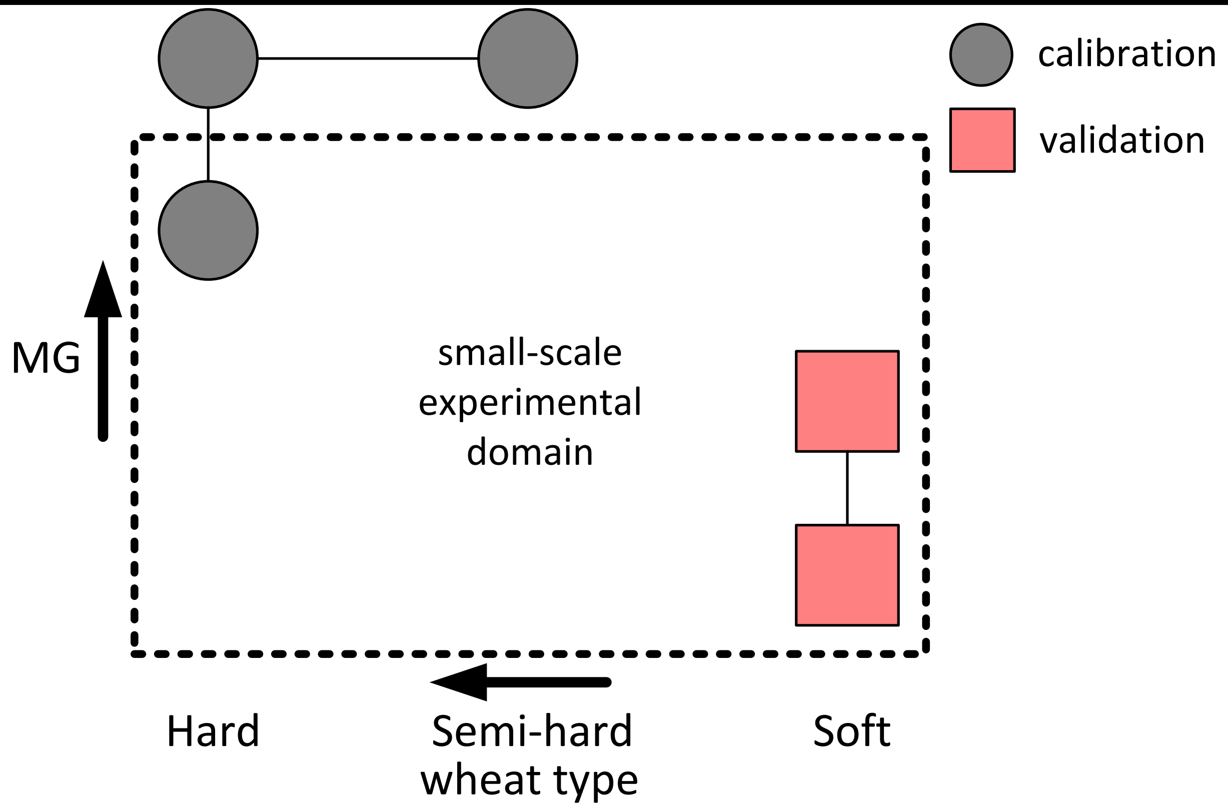
<u>Issues</u>:
- a single passage of milling + sieving is not sufficient
- complex plant layout
  - complicated interconnections among passages
- limited automation
  - plant control based on operators' experience

<u>Objectives</u>:
- to determine **the optimal process conditions of the industrial plant** to obtain the **desired flour quality** (i.e., particle size distribution)
  - depending on the **inlet raw material**
    - grain has a high variability depending on the geographical origin, the grain type, etc…
  - **through data collected in a pilot scale plant**

Transfer learning for scale-up



Industrial-scale dataset



- The industrial scale dataset does not cover the experimental domain of the desired product
  - the pilot scale experimentation does

Scale-up results

- Exploiting the **pilot scale data** the MG can be estimated in an accurate manner to run the milling in such a way as to produce the **desired flour**

desired product [mass % >1120 μm / total mass, mass % <1120 μm / total mass]	**Real MG**	**Estimated MG**
[62.7,37.3]	8	7.83
[72.4,27.6]	7.17	6.65

PLANNING EXPERIMENTAL CAMPAIGNS

In-silico formulation of new products

In-silico experimental campaigns

Design of experiments

Digital twins

Strawberries pasteurization and drying

Bertolini et al. (2020), *Journal of Supercritical Fluids.*, 164, 104914

Problem

- High-pressure CO_2 pasteurization and drying process:
  - weight loss
  - water activity
  - POD
  - PPO

Issues:

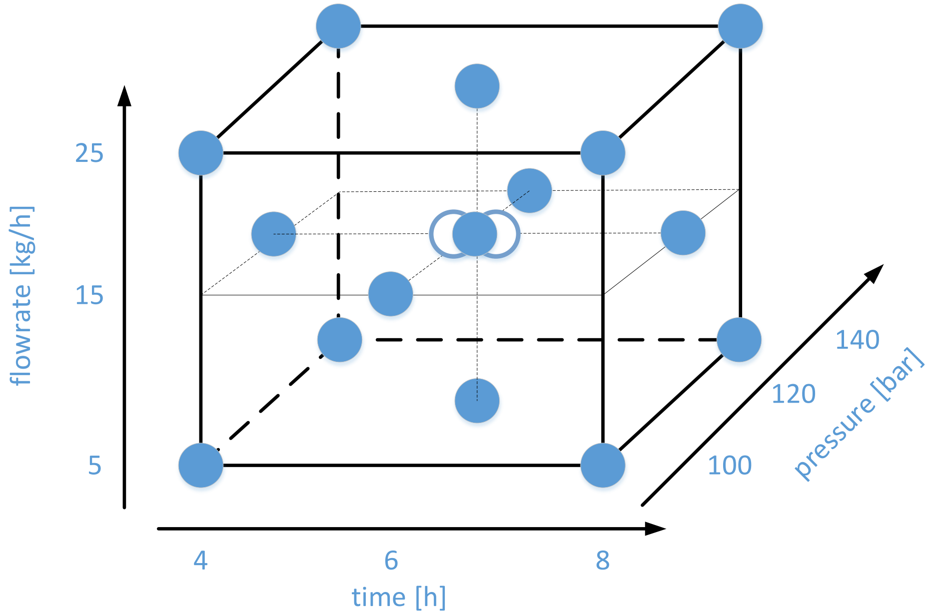
- complex task for the variability of the strawberries
- novel unknown process
- novel plant technologies

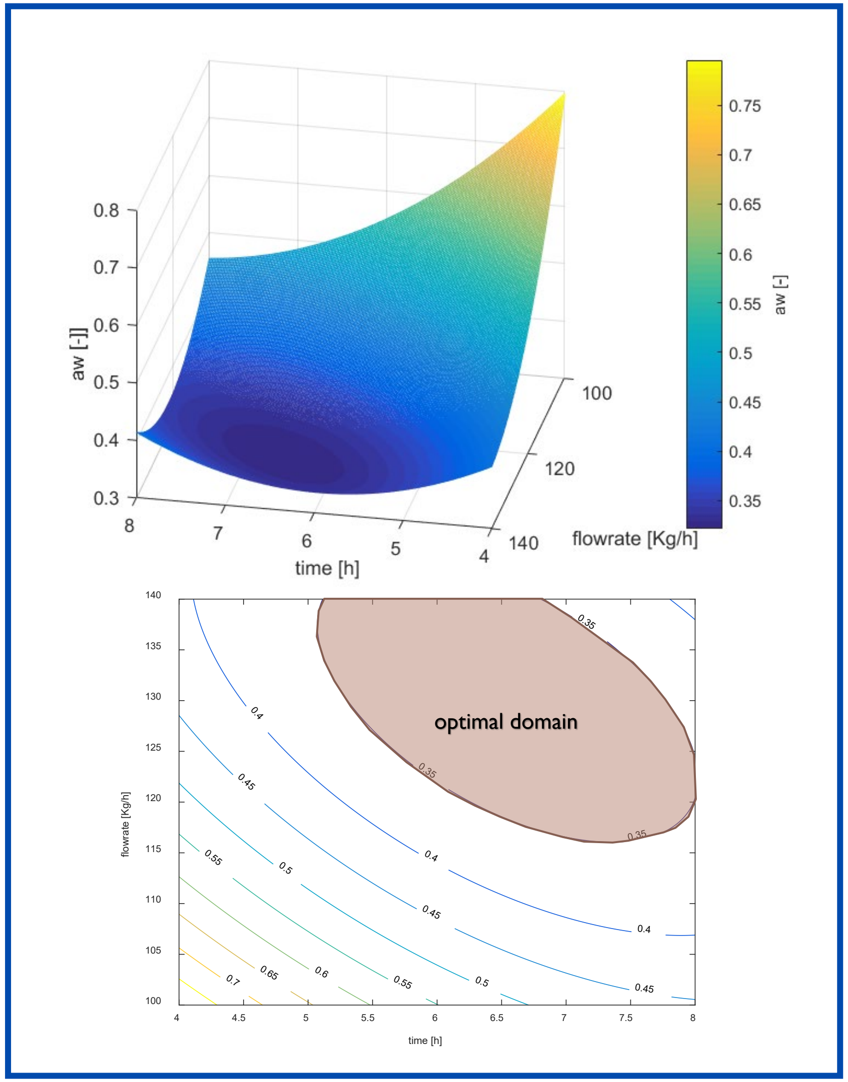
Objectives:

- to determine **the optimal process conditions** with the minimum experimental effort to obtain a **pasteurized and dried product** preserving
  - visual features
  - taste

Optimum process domain

- Suggested the most parsimonious strategy to optimize the process and identified the optimal process conditions for a fully pasteurized and dried product





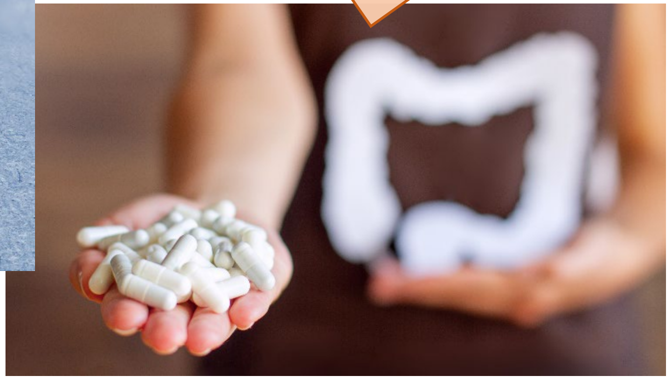
Model-based design of experiments in the production of tablets

in collaboration with: GlaxoSmithKline
Product Development R&D, Harlow (U.K.)

Cenci et al. (2022), Int. *J. Pharmaceutics*, 164, 104914

Tablets lubrication

- Production of tablets:
  - tablet press
  - tablet formulation
    - API
    - excipients
    - fillers
    - **lubricants**
    - etc…



Issues:

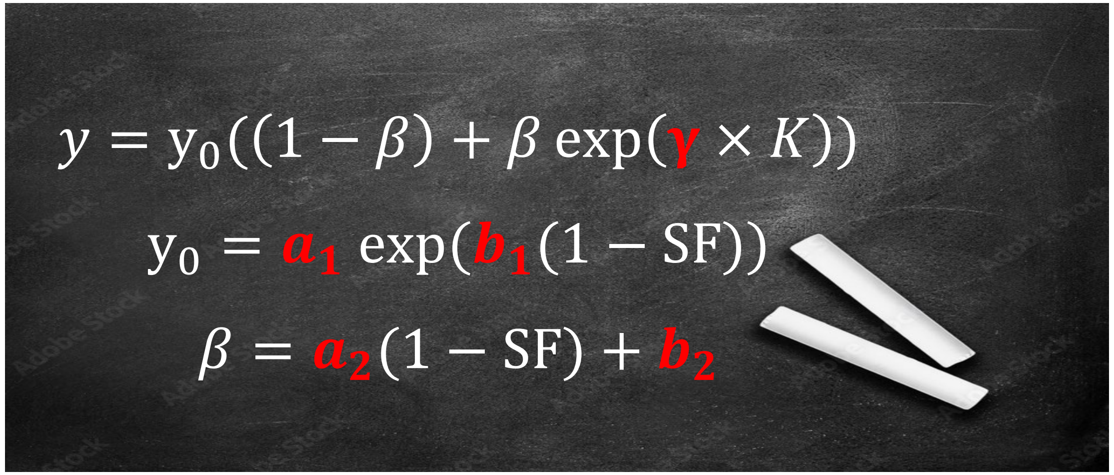
- lubricants are necessary for tablet processing
  - tablets detachment from the die
- lubricants impact negatively on:
  - tablets dissolution
  - patients correct drug adsorptions

Objectives:

- estimate the parameters of a model to **predict the correct amount of lubricant to be utilized**
- **minimum experimental effort** to obtain the parameters
- optimize the experiments in such a way they are maximally informative

Optimal model parameters estimation

- Suggested the optimal strategy to minimize the experimentation for the model parameters estimation
  - **60-70% reduction of the number of experiments** to be carried out
  - reduced experimental expenses (consider that 1 kg of API may cost up to 1 000 000 €)



IN-DEPTH PROCESS **UNDERSTANDING**, PROCESS AND QUALITY **MONITORING**, TIMELY SOLUTION OF MALFUNCTIONS, **PREDICTIVE MAINTENANCE**, PRODUCT **FORMULATION**, PROCESS **TRANSFER AND SCALE-UP**

METHODOLOGIES WERE DEVELOPED FOR THE IN-SILICO FORMULATION OF TABLETS IN SUCH A WAY AS TO **ACCELERATE THE EXPERIMENTATION:** NET REDUCTION OF THE NUMBER OF EXPERIMENTS


SOCIAL IMPACT: NEW DRUGS AVAILABLE FOR
PATIENTS IN SHORT TIMES
AND AT A
LOW COST

Conclusions

- **Exploiting historical data** from processes and products to
  - process **monitoring**
  - detect **anomalies** of a process
  - analyze the **defectiveness** of a product
  - **diagnose** the causes of anomalies and defects
  - characterize **product quality**
  - **design** new processes and products
  - process **scale-up/down**
  - **technology transfer** between scales
  - **shelf-life** determination
  - …
- Mature technologies for implementation and real-time use
- Data analysis + **engineering process knowledge!!!**

Data analytics:
turning data into business!

… per sempre a fianco a me!