A real estimation problem in the steel industry

Seminar at MSI class - November 3rd, 2021

Andrea Ghezzi, PhD Candidate at SYSCOP Lab

Andrea, who?

- Born in Bergamo, Italy, in 1996
- Graduated in October 2020 in Automation and Control Engineering from Politecnico di Milano
- From Dec 2020 to Sept 2021, researcher at Data Science R&D Department of Tenaris S.p.A. (in Dalmine, Italy)
- From October 2021, PhD Candidate at SYSCOP Lab

This is Tenaris

Global leader in pipes and related services for the world's energy industry

Serving the world's energy industry and other industrial applications.

5.1
US$ billion
Annual net sales (2020)

16
Countries
Manufacturing facilities

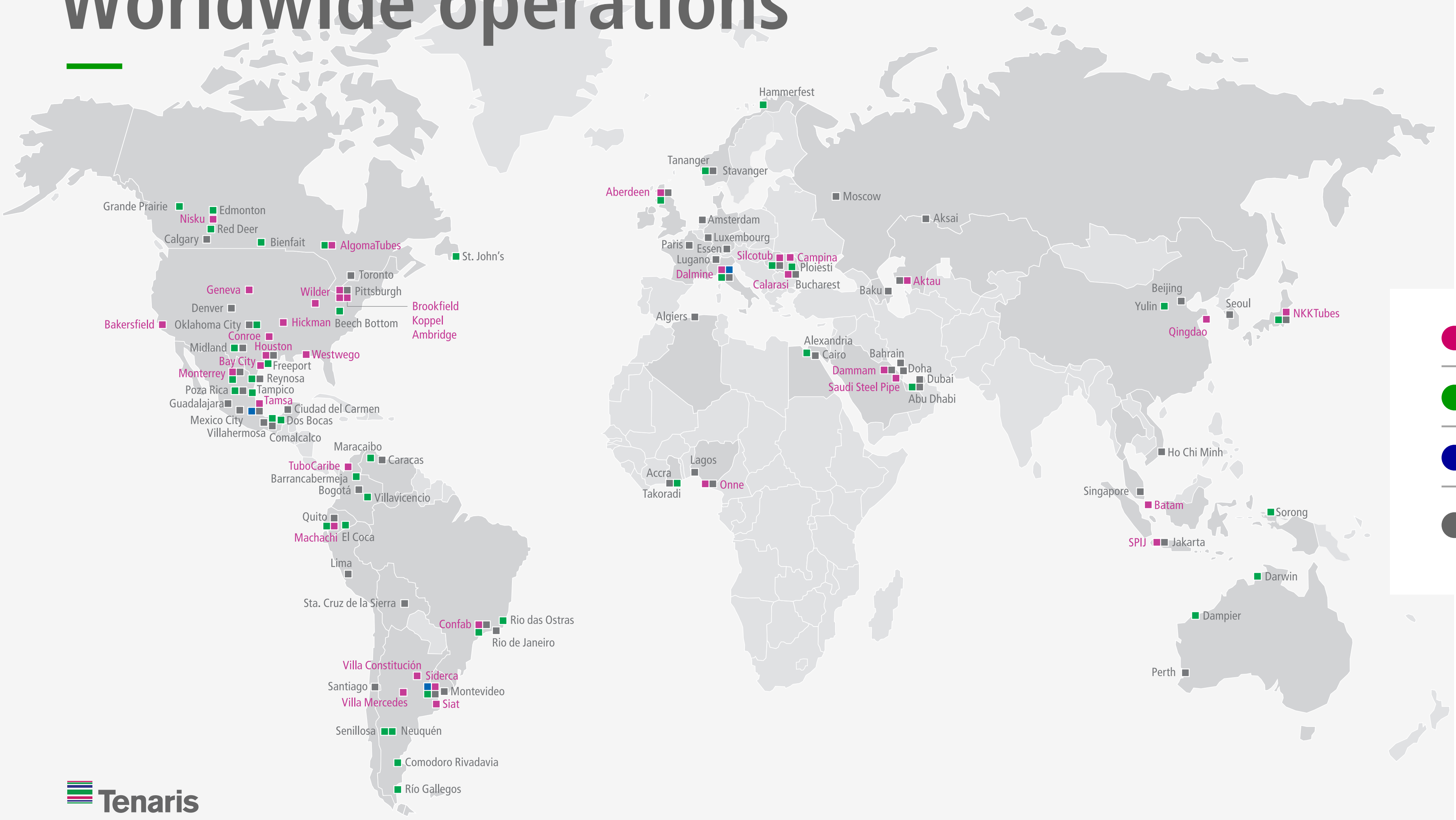
3
R&D Centers
Worldwide

3
Stock exchanges
New York, Italy, Mexico

19,000
Employees (approx.)
(2020)

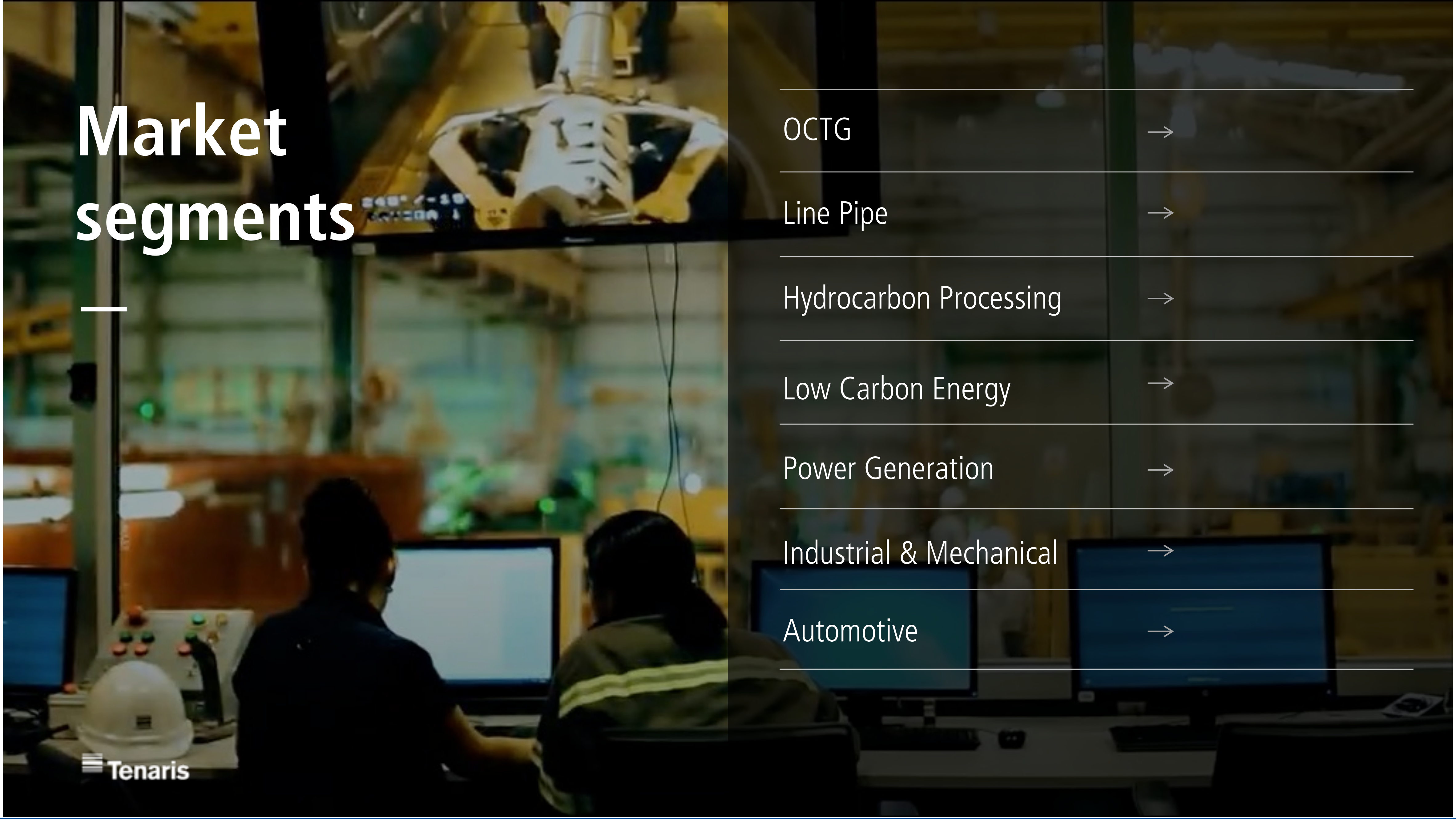
24
Countries
Services and distribution network

Worldwide operations



- Manufacturing Centers
- Service Centers
- R&D Centers
- Commercial/ Administrative Offices

Tenaris



Market segments

- OCTG →
- Line Pipe →
- Hydrocarbon Processing →
- Low Carbon Energy →
- Power Generation →
- Industrial & Mechanical →
- Automotive →

Outline

- Problem description
- Motivation
- Example

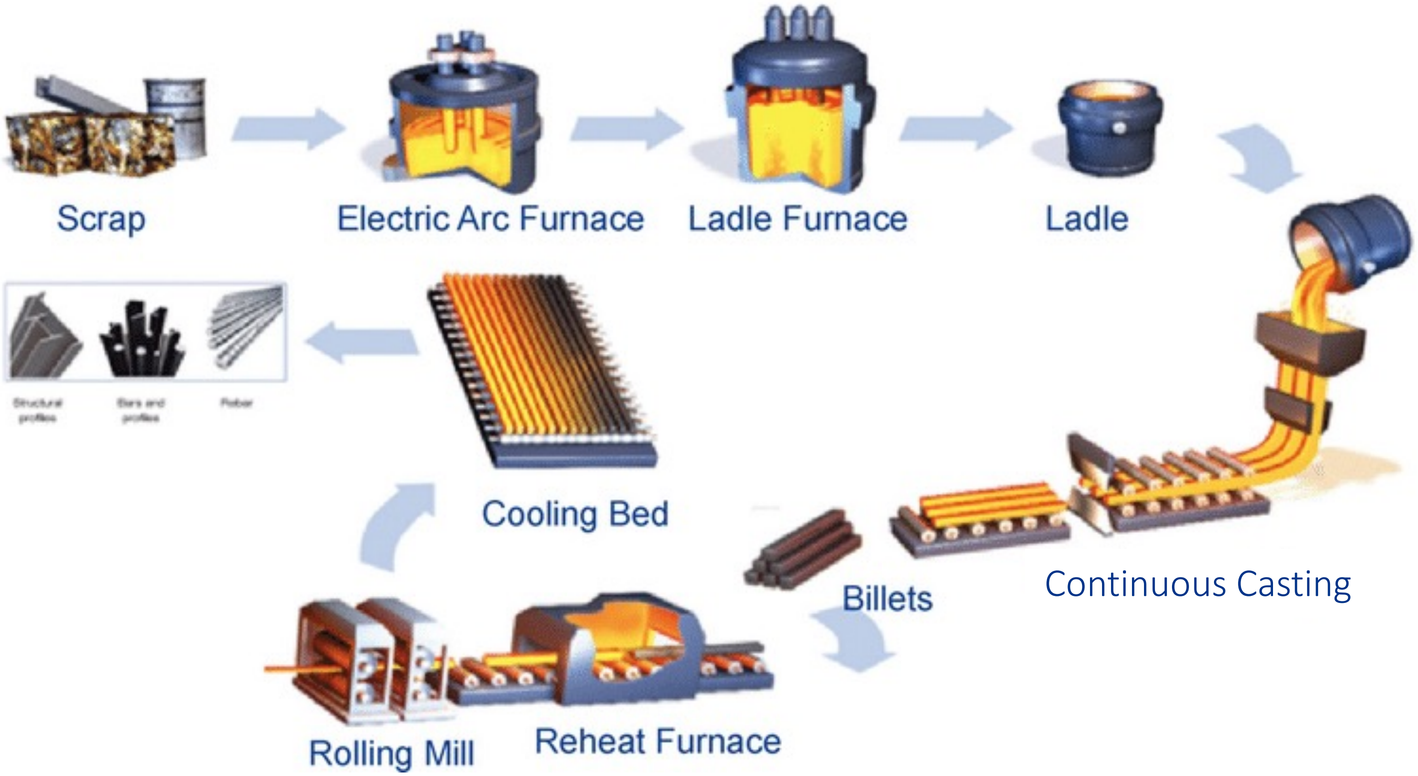
Steel recycling problem

- Steel manufacturing is one of the most energy demanding and polluting industries
- But steel itself is almost infinitely recyclable
- Indeed, most of the steel produced in developed countries is manufactured by melting steel scraps

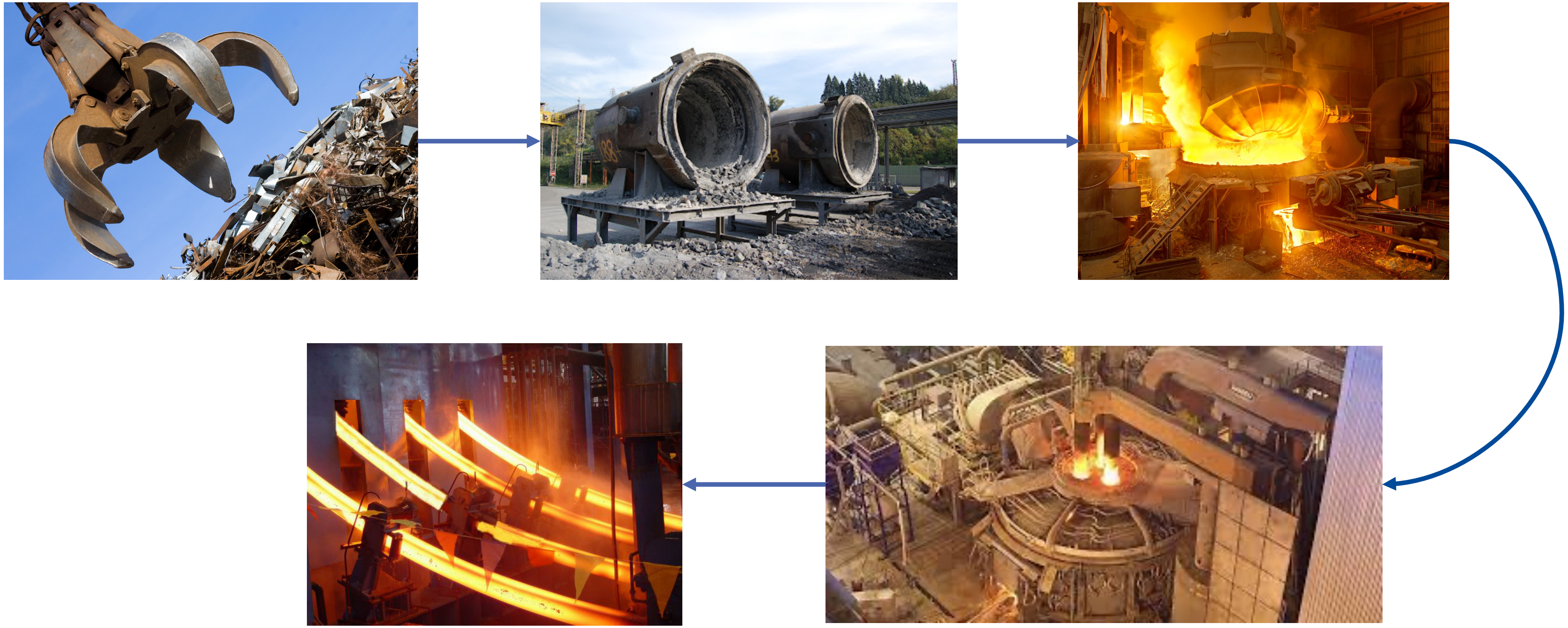
The raw materials – steel scraps

- The steel manufacturing companies buy steel scraps
- They come from:
  - Automotive industry
  - Building industry
  - Railway industry
  - …
- Each group of scraps has its own characteristics

The steel manufacturing process
In a nutshell



The steel manufacturing process



The reality

Motivation

- Steel is used in several applications, in each one specific properties are required (i.e. mechanical, chemical, …)
- The characteristics of the steel depend on the chemical elements contained

- Each type of new steel must satisfy constraints on the its chemical composition
- This set of constraints is called **prescription**

Motivation

- The two families of raw materials used to produce steel are:
  - **Steel scraps**: typically from 95% to 99% of the final mass of the steel
  - **Ferroalloys**: specific alloys that allows the final steel to reach high concentration of a specific element (i.e. Molybdenum, Nickel, …)

↓

Knowing perfectly the scraps allows us to always satisfy the prescription

↓

Quality guaranteed without waste

Motivation

Unfortunately knowing perfectly the scraps is not straightforward:

- The scrap suppliers declared the chemical composition only with large ranges or just on one element (i.e. Lead free)
- It's not possible to take direct measurements of each type of scraps:
  - Expensive
  - Not representative (the sample can be 1 kg taken from a pile of 5000 tons)

↓

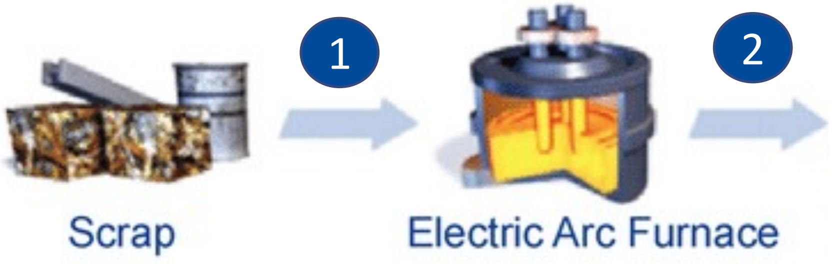
A statistical methodology is required!

Problem statement

- Regard a scrapyard with 5 type of scraps
- Regard the Copper (Cu) which is a pure pollutant:
  - It always worsens the performance of the final steel
  - It only comes from the scraps (not from the ferroalloys)
- We measure/know:
  - The masses of each scrap picked for each heat
  - The Copper contained in the final steel
  - The Copper is not subject to chemical reaction in the furnace
    - The Cu inside the scraps will be inside the final steel
- We want to characterize the Copper concentration in each type of scrap

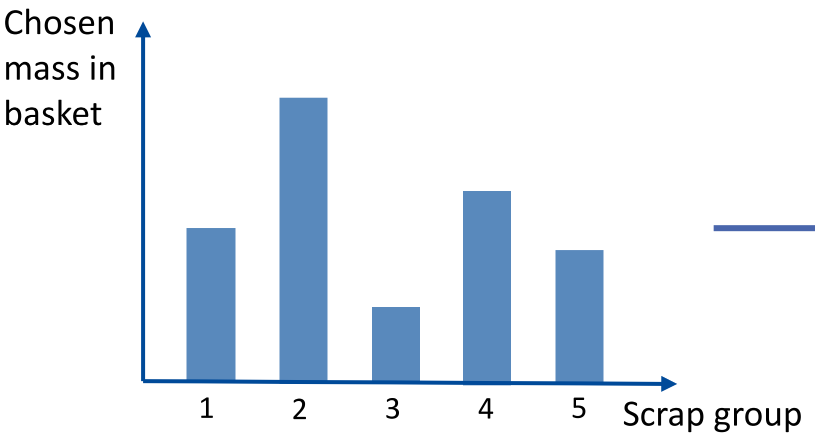
Modelling

Measurements



For each heat $k = 1, \dots, \mathrm{N}$ we measure:

1



$$\begin{bmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \\ \phi_4 \\ \phi_5 \end{bmatrix} \longrightarrow \phi(k)$$

2

A measurement of the chemical composition of the liquid steel before adding the felloalloys $\longrightarrow$ $y(k)$

Modelling

Sources of uncertainty

The **measurements** taken from the EAF **are subject to uncertainty**:

- Measurements are taken on a small sample
- Measurements can be taken some minutes before the steel bath is completely molten
- Sometimes to speed up the melting time for consecutive heats some liquid steel is kept in the furnace and it mixes with the next heat

Sequence of measurement noise
$\epsilon(k), k = 1, \dots, N$

Problem formulation

The relationship which binds the Copper inside the scraps to the Copper in the final steel is linear:

$$y(k) = \phi(k)^\top \theta + \epsilon(k), \qquad k = 1, \dots, N$$

where the unknown $\theta \in R^5$ is the concentration of Cu in the 5 scraps

→ Apply Linear Least Squares (LS) to find an estimate $\hat{\theta}_{LS}$ of θ_0 the real but unknown value

Linear Least Squares (LS)

→ **LS cost function:**

$$f(\theta) = \sum_{k=1}^{N} (y(k) - \phi(k)^\top \theta)^2 = \|y_N - \Phi_N \theta\|_2^2$$

$$\text{where } y_N = \begin{bmatrix} y(1) \\ \vdots \\ y(N) \end{bmatrix} \text{ and } \Phi_N = \begin{bmatrix} \phi(1)^\top \\ \vdots \\ \phi(N)^\top \end{bmatrix}$$

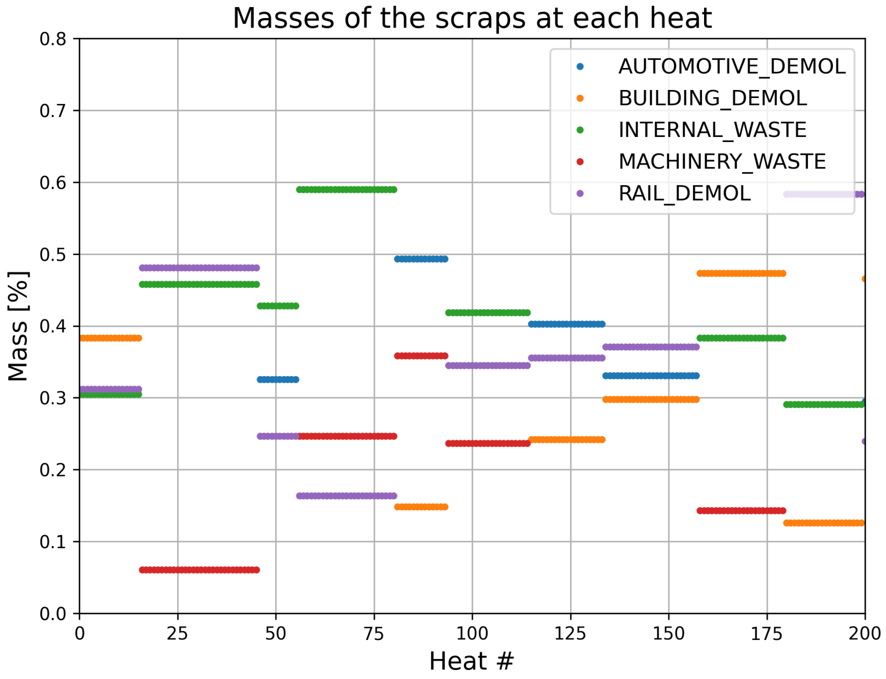
→ **LS optimization problem:**

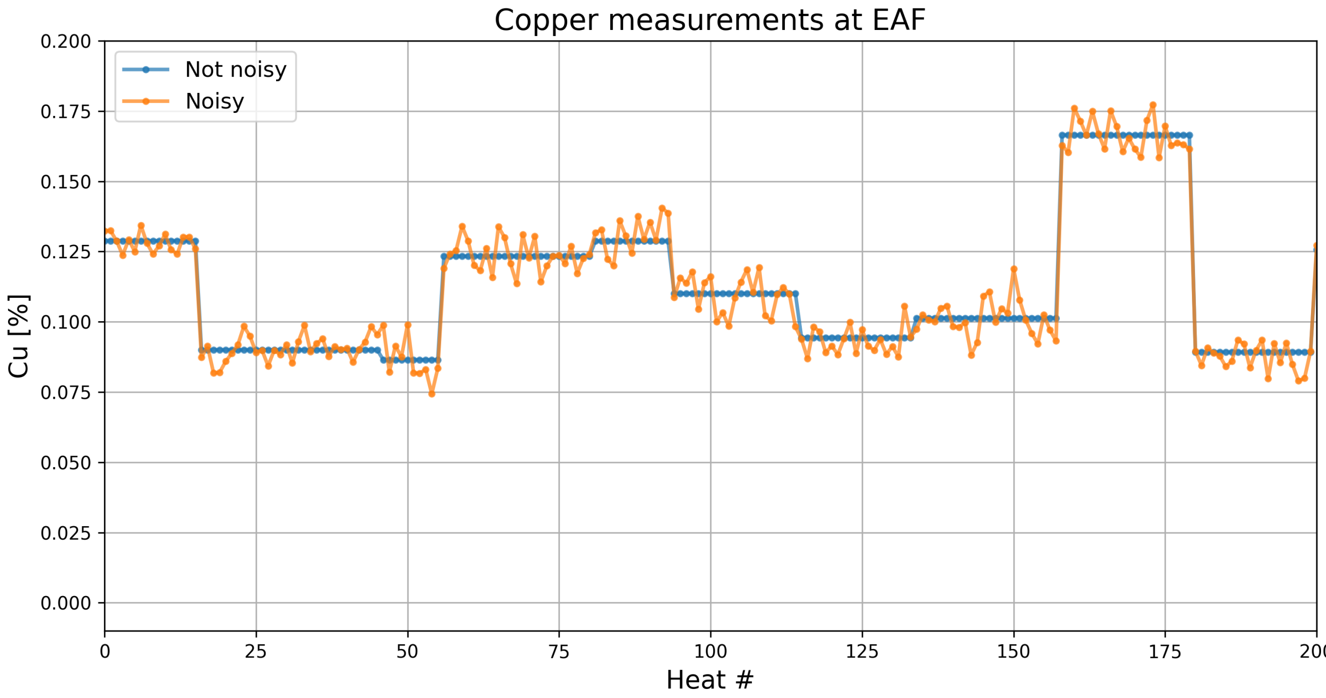
$$\hat{\theta}_{LS} = \arg\min_{\theta \in R^5} f(\theta)$$

Jupyter demo

Jupyter demo

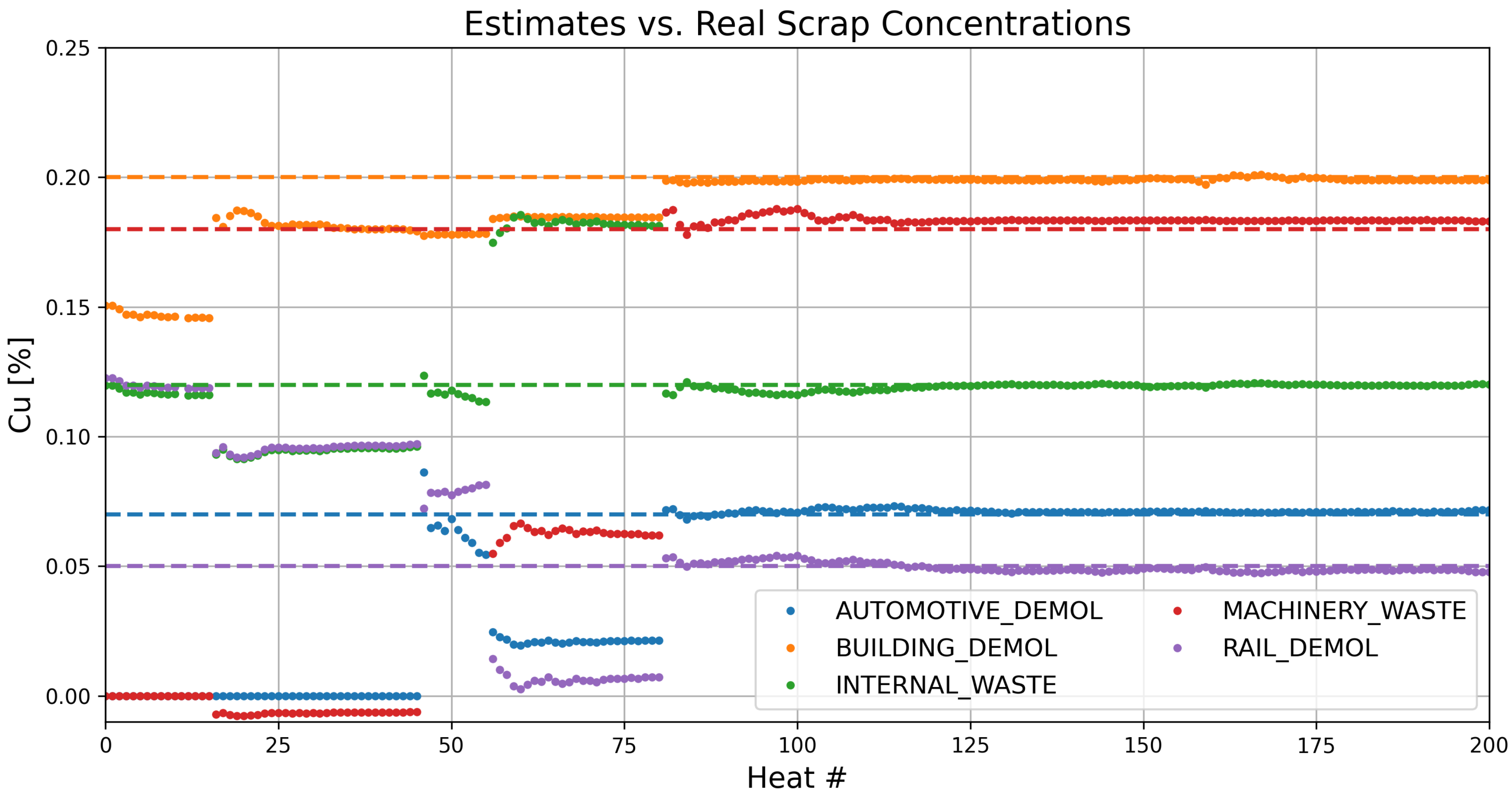
Measurements





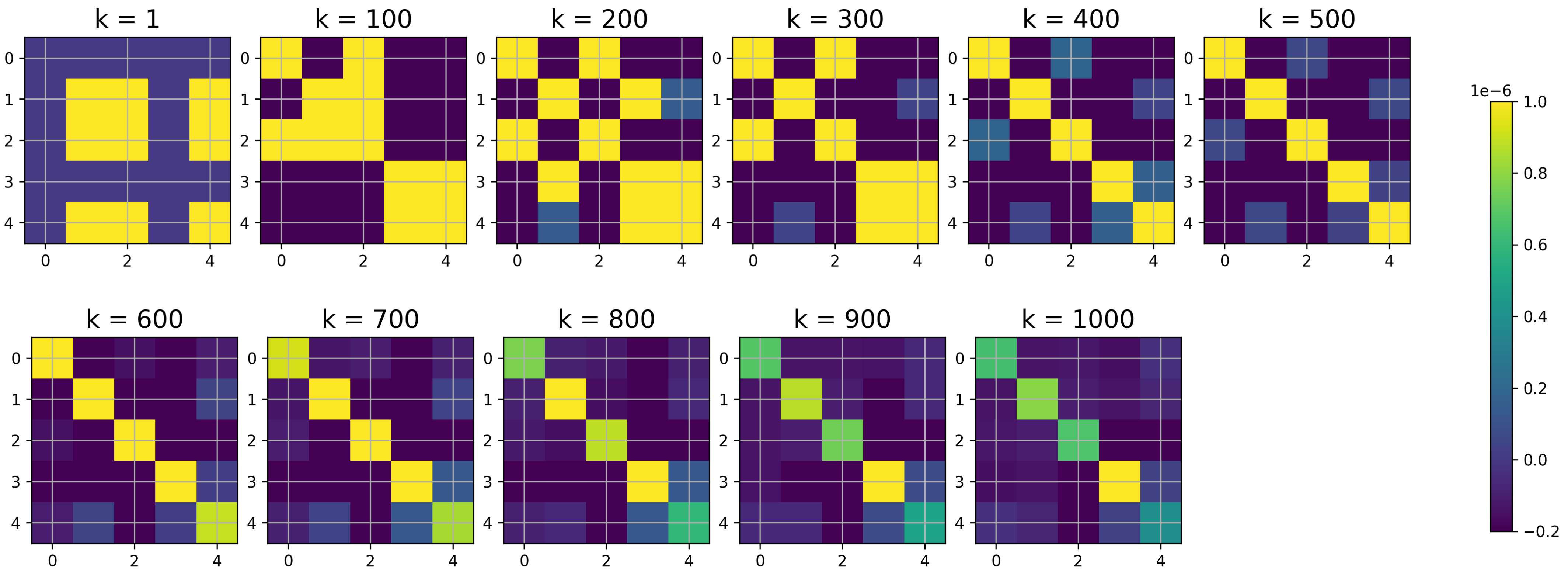
Jupyter demo

Estimates



Jupyter demo

Covariances of the least square estimator $\hat{\theta}_{LS}$ at different timesteps k



Thank you for your attention!