ESGI 188

European Study Group with Industry

26-30 May 2025

BCAM, Bilbao

basque center for applied mathematics

EXCELENCIA SEVERO OCHOA
Jul. 2014 - Dec. 2027

BAT
B Accelerator Tower

Bizkaia
foru aldundia
diputación foral

Lookiero

QUANTUM MADS

DANOBATGROUP

bcamath.org

Index

bcamath.org
Mazarredo 14, 48009 Bilbao | Basque Country, Spain | (+34) 946 567 842

Preface

The 188th European Study Group with Industry (ESGI 188), held from 26 to 30 May at the B Accelerator Tower (BAT) in Bilbao, brought together industrial challenges and mathematical expertise in a dynamic week of problem-solving and innovation. Organised by the Basque Center for Applied Mathematics (BCAM), in collaboration with Bizkaia Provincial Council and B Accelerator Tower, the event showcased the power of collaboration between academia and industry to address real-world problems.

Four companies presented specific challenges, seeking mathematical insights to enhance their operations:

- **Lookiero:** This online personal shopping service aimed to optimise its product recommendation algorithms to improve customer satisfaction and operational efficiency.
- **BAT (B Accelerator Tower):** As a hub for innovation, BAT sought strategies to foster collaboration among its diverse ecosystem of startups, corporations, and researchers.
- **Quantum Mads:** Specialising in quantum computing solutions, Quantum Mads presented a challenge related to optimising complex computational processes for industrial applications.
- **Danobatgroup:** A leader in industrial manufacturing, Danobat focused on enhancing predictive maintenance models to reduce downtime and improve machine reliability.

Throughout the week, interdisciplinary teams of mathematicians and scientists worked closely with company representatives to develop models and propose solutions. The collaborative environment fostered innovative approaches, demonstrating the practical impact of mathematical applications in industry. The event closed with presentations

ESGI 188 not only addressed immediate industrial challenges but also strengthened the ties between academia and the business sector, highlighting the mutual benefits of such partnerships. The success of this event underscores the value of continued collaboration to drive innovation and solve complex problems across various industries.

bcamath.org
Mazarredo 14, 48009 Bilbao | Basque Country, Spain | (+34) 946 567 842

The Scientific Committee of ESGI 188 was formed by the following members:

- Fernando García, BCAM
- Rubén Peña, BCAM
- Aritz Pérez, BCAM
- Felipe Ponce, BCAM

The Academic Coordinators Committee was formed by the following members:

- Javier González, Quantum Mads
- Jokin Muñoa, Danobatgroup
- Alex Sánchez, BAT Accelerator Tower
- Miguel Ángel Venganzones, Lookiero

bcamath.org
Mazarredo 14, 48009 Bilbao | Basque Country, Spain | (+34) 946 567 842

Challenge description

Challenge 1 – Predict the 5 years Life-Time Value (LTV) of a cohort of Lookiero's customers – Lookiero

PROBLEM DESCRIPTION

Life-Time Value (LTV) In marketing, customer life-time value (LTV) is a prognostication of the net profit contributed to the whole future relationship with a customer. The purpose of the customer lifetime value metric is to assess the financial value of each customer. Therefore, LTV is a key metric in every Business To Customer (B2C) business model.

There are different ways to formulate the LTV of a customer. Here, we propose a simple one:

Life-Time Value (time) = Repetition Rate (time) * Average Order Value (time)

The LTV is a time dependent metric, usually described as a time series. The Repetition Rate (RR) is the amount of purchases of a customer in a given time period, and the Average Order Value (AOV) is the average monetary spent of a customer over the purchases up to the given time period.

Usually, the LTV is calculated over a cohort, a group of customers sharing some characteristics, i.e., they ordered for the first time at the same year-month date, they belong to the same market, they belong to the same age group, etc.

OBJECTIVES / EXPECTED OUTCOMES

Lookiero Box

The main Lookiero business channel is Lookiero Box. Given some information from a customer (style, needs, body measurements, etc.), an expert (Personal Shopper) selects five garments that are sent to the customer' home in a Lookiero Box. The customer tries the garments and then pays for the one that wants to buy and returns the remainers for free.

This is a recurrent, mostly subscription-based business, with an approximately 2-months median frequency and a limited number of items to buy (up to 5).

bcamath.org

Mazarredo 14, 48009 Bilbao | Basque Country, Spain | (+34) 946 567 842

Lookiero Lookin

Lookiero Lookin is the ultra-personalized Lookiero e-commerce. Our fashion personalization algorithms provide an ultra-personalized e-commerce experience, where each customer can access a selection of garments and looks particularly suited for her style and preferences.

This is an on-demand business model with no limitations to the number of items a customer can purchase at the same time.

Challenge 2 – Analysis of the performance of startups in the Basque Country's technological entrepreneurship ecosystem compared to other ecosystems - B Accelerator Tower

PROBLEM DESCRIPTION

The Basque Country's technological entrepreneurship ecosystem is a dynamic and innovative ecosystem that has been growing in recent years, generating collaborative dynamics that involve universities, research centres, large corporations, Startups and the Public Administration itself, in addition to BAT B Accelerator Tower as the main unifying project of the ecosystem.

As far as emerging companies are concerned, and based on those certified by ENISA in February 2025, the Basque Country has 45 certified emerging companies out of the 1,370 existing at a national level.

In this context, BAT wants to analyse the performance in terms of economic performance and growth of emerging companies in the Basque ecosystem compared to emerging companies in other reference ecosystems at a national level.

OBJECTIVES / EXPECTED OUTCOMES

The expected outcomes of this challenge are:

- To extract the main metrics of volume and performance of the emerging companies in the Basque Country ecosystem, such as turnover, employment, share capital or the result of the year.
- To compare the obtained metrics with similar metrics from other comparable ecosystems and obtain conclusions.
- To cross-reference the obtained performance data with other economic variables such as the evolution of GDP, employment, etc.
- To obtain a methodology to measure the impact generated by the BAT activity in achieving the analysed performance.

bcamath.org
Mazarredo 14, 48009 Bilbao | Basque Country, Spain | (+34) 946 567 842

Challenge 3 – Quantum-Enhanced Optimization for Renewable Energy Systems - Quantum Mads

PROBLEM DESCRIPTION

The growing integration of renewable energies, particularly wind power combined with ultracapacitor-based energy storage, presents significant control and optimization challenges to ensure both stability and efficiency. Although the Linear Quadratic Regulator(LQR) approach is widely used for optimal control, solving it at large scale can lead to high computational complexity. To explore new paradigms, we propose mapping the LQR problem to a Quadratic Unconstrained Binary Optimization (QUBO) formulation for solution on digital annealers (e.g., Fujitsu Digital Annealer). This challenge leverages direct methods of optimal control, inspired by the reference article [1]. The goal isto demonstrate how quantum computing and annealing-basedmethods can enhance the speed and robustness of control decisionsin hybrid energy systems.

OBJECTIVES / EXPECTED OUTCOMES

The expected outcomes of this challenge are:

1. Quantum LQR Formulation

Define the LQR problem (using indirect methods) in a manner suitable for translation into a QUBO model.

Incorporate stability and feasibility constraints particular to a wind generation system with ultracapacitors.

2. Implementation and Validation on Digital Annealers

Encode the QUBO representation in the Fujitsu Digital Annealer (or another annealer) and compare the results with open-license classical solvers.

Assess computational time, solution accuracy, and sensitivity to variations in system dynamics (wind, load, etc.).

3. Demonstrations and Numerical Results

Present implementation examples in different scenarios, highlighting benefits (or limitations) of the quantum-oriented approach.

- Offer recommendations for scalability and possible extensions to other hybrid energy systems.

Challenge 4 – Robot dynamics for metal cutting processes - Danobatgroup

PROBLEM DESCRIPTION

In recent years, there has been renewed interest in using robots for metal cutting processes. In high-precision manufacturing, Computerized Numerical Control (CNC) machine tools are the standard for metal cutting processes because of their high stiffness and limited degrees of freedom, which allow for precise motion control. However, CNC machine tools are very expensive, difficult to relocate, and lack the flexibility to machine complex parts. On the other hand, robots are more affordable, easier to move, and offer greater flexibility.

Despite these advantages, robots still have several shortcomings: propagation of dimensional error between links; low-stiffness at the joints; joint backlash; and angular positioning errors at the joints due to the absence of a second encoder. These challenges have motivated various lines of research, in particular, to better understand the robot dynamics during metal cutting operations.

Modeling robot dynamics and vibrations is a difficult task due to the complex structure of links and joints in robots, as well as the constant variation of modal parameters. Engineers seek to understand how robotic arms respond to external and cutting forces, as these forces can generate unwanted vibrations that deteriorate the quality of the surface finish. By gaining deeper insights into these dynamics, researchers aim to develop strategies to suppress such vibrations and enhance robotic machining performance.

OBJECTIVES / EXPECTED OUTCOMES

To model the main factors affecting robot dynamics during metal cutting and provide recommendations to prevent instabilities.

Participants are expected to propose models for the metal cutting dynamics, develop theoretical approaches, perform numerical simulations, explore different stability criteria, discuss strategies to mitigate unwanted vibrations, and summarize their findings.

bcamath.org Mazarredo 14, 48009 Bilbao | Basque Country, Spain | (+34) 946 567 842

Final reports

Lookiero

1. Introduction

Lookiero is an online fashion retailer, whose business proposition is providing a personal shopper to their customers. Lookiero's main business channel is the Lookiero box, a subscription service in which Lookiero sends a 5-garment box to the customer, with items selected by a personal shopper to suit the customer based on style, needs and body measurements.

2. Objectives and Scope of the Challenge

The goal of this challenge is to predict the the 5-year Life-Time Value (LTV) of a cohort of Lookiero's customers. LTV is a predictor for the money the customer will expend in the company over the entire relationship period between the company and the customer. It is a time dependent metric and can be represented as a time series. We will be working with cohorts of customers, grouped by the date (year-month) they first ordered. Other cohorts: age group, market, ect. can also be considered.

2.1 Datasets Description

- `boxes_data`: Contains information regarding the main subscription service orders (lookiero box), including amount spent on each order and which customer spent the money on it. It has a customer id as it's primary key. This table contains 13 columns and 5184 entries for the orders from 02/2020 to 03/2025.

- `customer_data`: Contains data for all customers in who signed in on the same month and year to Lookiero. This includes demographics, size, preferences and other static variables. It has order id as primary key and customer id as a secondary key. This table contains 27 columns and 1914 entries for the 02/2020 customers.

- `Looking_data`: This file contains information regarding on-demand ecommerce purchases done by customers, outside the Lookiero box subscription services. It has

order id as primary key and customer id as a secondary key. This table contains 7 columns and 34 entries for the orders from 02/2020 to 03/2025.

3. Scientific Summary

Initially, we decided to just focus on the box data, as the looking data contained very few entries. We joined both the box and customer data tables into a single table: creating, for every month from 02/2020 to 03/2025, a column with total money spent by the custom in the company since they first register up to that particular month. Hence this new columns represented the life-time value for each customer for every month of the five year period. With this new dataset, we could proceed solve the problem. We will denote this dataset by $D = \{X_{static}, LTV\}$, where X_{static} corresponds the existing columns from the original `customer_data` table, and LTV denotes the new columns containing the LTV accumulated up to each month. We also denote by LTV_n the first n columns of LTV .

3.1 An initial approach using XGBoost

Our first approach to the problem was to consider a simple supervised learning model for each customer, and then averaging the LTV predictions across all members of the cohort to produce a cohort level prediction. The choice of classifier was motivated by the need to use a simple, but well performing method such as XGBoost as classifier and focus on solving the rest of the problem: Data processing, error metrics and problem modelling.

For this problem, we consider the set of labels $X = (X_{static}, LTV_0)$. While this contains the case $n = 0$, according to Lookiero's experience, having data from at least a few purchases is necessary to make accurate predictions.

Data Processing

In order to prepare the data for the classifiers we processed it by:

- Leave missing values as they are, since XGboost is designed to work with missing values (represented as np.nan for the Python implementation).
- Categorical data one-hot enconded, adding new binary columns represent- ing that data
- A few columns that had no relevant data are deleted: date of birth (as it was redacted), as well as ids, registration dates (only differ on the day).
- New variables are added. These are statistical variables summarizing the data from the first n months of LVT data. The goal is to aid the classifier in using that valuable information. They

bcamath.org Mazarredo 14, 48009 Bilbao | Basque Country, Spain | (+34) 946 567 842

include: average value, max value, slope of the regression and standard deviation.

Experimental Data

The following results are the result of training the XGBoost model on the previously described Datasets (corresponding to different values of *n*). The result come from using 10-fold cross-validation.

Table 3 presents the data for models including data from *n* = 0, 1 and 6

Month (n)	Model	MAE	RMSE
0	Mean	176.01 ± 15.33€	€317.83 ± 51.33
	Median	152.49 ± 18.13€	€332.20 ± 53.92
	XGBoost	181.52 ± 15.38€	€324.98 ± 46.02
1	Mean	178.10 ± 15.59€	€320.34 ± 51.76
	Median	154.34 ± 18.45€	€334.79 ± 54.38
	XGBoost	167.18 ± 13.99€	€316.66 ± 46.61
6	Mean	186.86 ± 16.72€	€332.78 ± 54.04
	Median	161.42 ± 19.83€	€347.68 ± 56.78
	XGBoost	$\mathbf{106.33 \pm 11.95}$€	€$\mathbf{235.91 \pm 36.66}$

Table 1: Model comparison

In Figure 1 we can see how both RSME and MAE decrease as we include more information in the predictor variables. We can also see in Figure 2 how the models fit the data, showing the predicted LTV, with the 6-month model fitting it better than the 1-month model.

4. Comprehensive approach using multiple year cohorts

At a later date in the challenge, more data was made available to us. Including 5 new triplets of tables. Each triplet corresponds to the year the customer of the cohort signed in 2021, 2022, 2023, 2024, 2025, and follows it's structure as described in Section 2.1, with `box_data`, `customer_data` and `looking_data` tables corresponding to the new years, adding up to a total of 6 triples of tables including the original 2020 cohort. Since orders only go up to march 2025, this new cohorts have less purchase data available, the only cohort with 5 years of purchase data is the one who registered in 2020. This presents a challenge in how to use this data to improve the model.

As stated, the objective is to predict 5 year LTV, which makes a supervised machine learning approach challenging, as it requires incorporating data that is not fully labeled, but still very

useful. What we propose is using

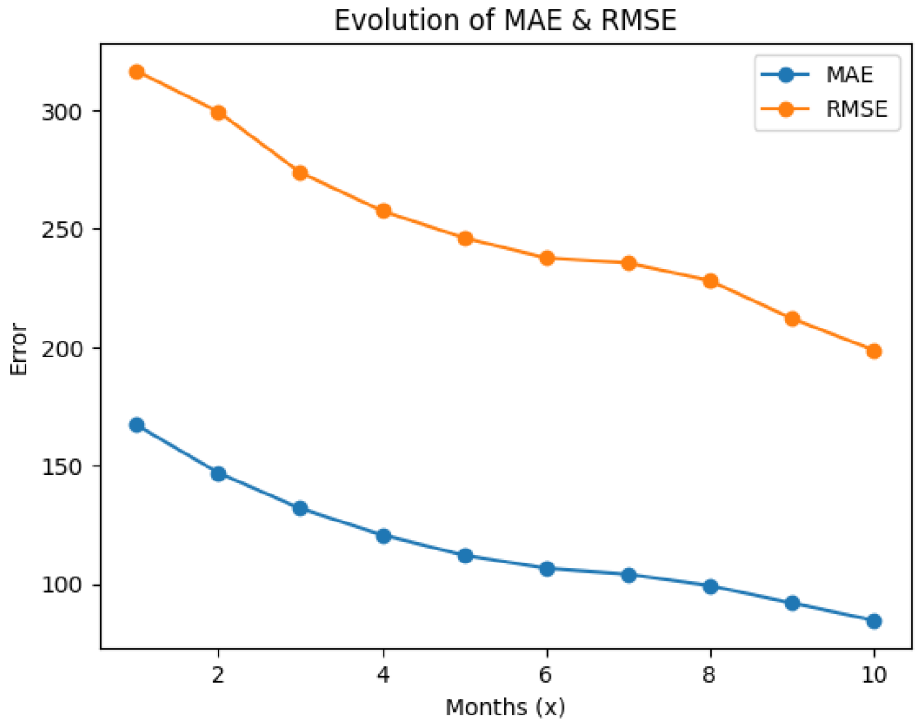


Figure 1: Evolution of the RMSE and MAE as more customer expense data is used as predictor data. Cohort 2020.

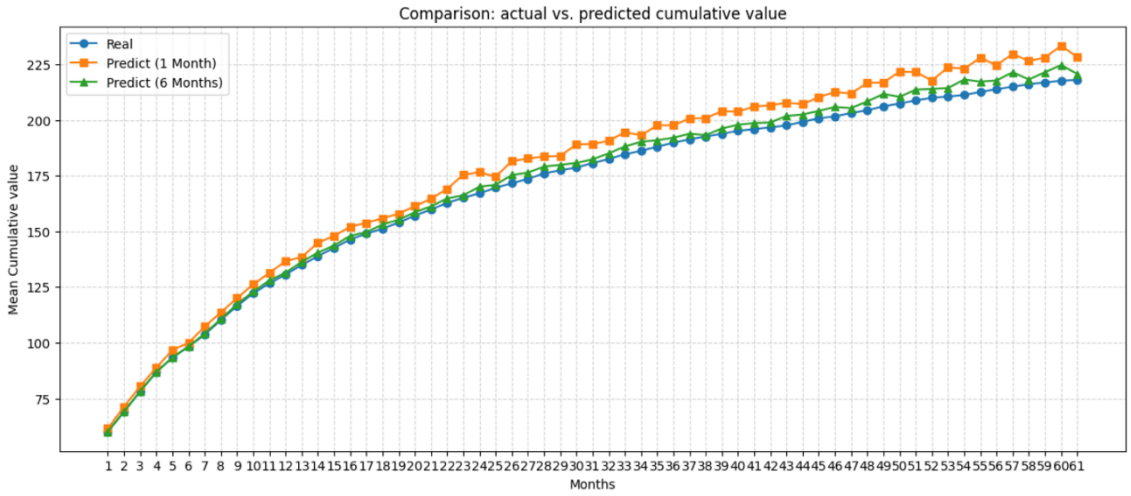
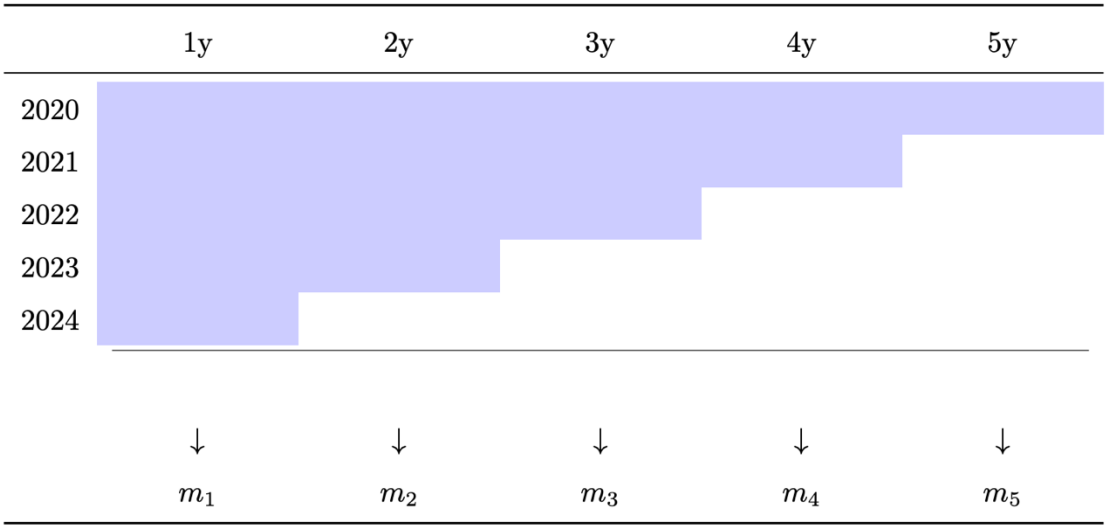


Figure 2: Monthly Life-Time value for 1 and 6 month models compared to real data. Cohort 2020.

Mazarredo 14, 48009 Bilbao | Basque Country, Spain | (+34) 946 567 842



Cohort	MAE	RMSE	R2
2021	109.49	203.34	0.60
2022	82.23	154.25	0.67
2023	55.31	103.70	0.71
2024	34.85	65.04	0.87

Table 2: Cohort result comparison

Segment	MAE	RMSE	R2
Year 1	20.75	46.15	0.92
Year 2	53.02	109.96	0.70
Year 3	103.62	200.29	0.57
Year 4	159.32	295.57	0.40
Year 5	164.16	326.66	0.31

Table 3: Model comparison by cohort

BAT

1. Introduction

This report presents an analytical study aimed at evaluating the impact of the BAT (Business Acceleration Team) ecosystem on the growth and performance of startups. The BAT initiative, launched in 2022, seeks to support early-stage companies through mentoring, networking, and access to resources. Our objective is to investigate whether participation in this ecosystem is associated with improved financial outcomes, particularly in terms of revenue growth. To do so, we have worked with a dataset of startups that includes financial indicators from 2021 to 2023, information about their sector (via CNAE classification), geographic location (Euskadi vs. other regions), and whether or not they are part of the BAT ecosystem. Using both exploratory analysis and predictive modeling techniques, we aim to identify patterns, test hypotheses, and simulate future scenarios that shed light on the influence of BAT participation.

2. Objectives and scope of the challenge

This challenge has different goals and questions to answer, including:

- To extract the main metrics of volume and performance of the emerging companies in the Basque Country ecosystem, such as turnover, employment, share capital, or the result of the year.
- Basque startups' evolution compared to the rest of the Spanish local ecosystems. Do Basque startups have stronger evolutions between 2021 and 2023 than the other Spanish startups?
- BAT Startups' evolution compared to Bizkaia and Basque Country. Did BAT have an influence on the progress made by the startups that are part of it compared to the ones that are not part of BAT?
- Based on the last three years of activity, can we predict/simulate how a startup will evolve within the next three years? For example, if a <30k revenue startup joins BAT, where will they be in 2028 based on the performance of BAT companies in the past?

bcamath.org
Mazarredo 14, 48009 Bilbao | Basque Country, Spain | (+34) 946 567 842

- Regarding sectors, is there any pattern related to startups in the same segment/market niche? Are there sectors that demand, on average, more employees than others, generate more revenue per employee, etc?

3. Scientific summary

Data Sources and Scope

The data we had for this challenge came from public sources on more than 1,000 startups in Spain both BAT-supported and non-BAT startups distributed in different geographical regions. The main information available about these startups included their revenue, number of employees, grants received, region, and CNAE code (Clasificación Nacional de Actividades Económicas). We also had financial data such as ROE, ROA, EBITDA, share capital, and the self-financing ratio generated by assets. We had some qualitative data that we could use if they were complete.

Data Quality

To extract valuable insights, our first task was to assess the quality and completeness of the data. A significant portion of the dataset had missing or non-public entries, particularly in variables like grant receipts, where over 80% of values were unavailable. This led us to focucompareily on revenue and number of employees rather than financial data.

Small startups are not required to share the data we need to properly analyze the situation, and they often choose not to do so in order to prevent rival companies from gaining insights into their financial position. This makes it challenging to access complete and reliable information. However, going forward, it is important for BAT to find ways to encourage startups to share this data, as it is essential for BAT's strategic analysis and decision-making.

Cleaning & Preprocessing Strategy

To ensure analytical robustness:

- We restricted our main focus of the analysis to variables with at least 40% availability. Thus, we focused on revenue and number of employees, (while we kept financials to see what we could for them) as these were available for the vast majority of startups, which were consistently reported across both BAT-supported and non-BAT startups.

- Growth metrics were computed:

bcamath.org Mazarredo 14, 48009 Bilbao | Basque Country, Spain | (+34) 946 567 842

- Revenue Growth: Absolute difference between years (i.e., revenue in euros in 2023 minus revenue in euros in 2022)
- Employment Growth: Percentage change year-over-year.

- Variables with low completeness or insufficient comparability were excluded
- All numerical columns used in the analysis were converted to proper numeric types and cleaned of infinite or invalid values.

Exploratory Data Analysis (EDA)

Once the data was cleaned and reorganized, we could begin our analysis. We started with an exploratory data analysis (EDA). Exploratory data analysis is a crucial step in most data science projects. It helps us understand the structure of the dataset, identify general trends, detect anomalies or outliers, and explore relationships between variables. EDA also allows us to assess the distribution of key metrics and supports the formulation of hypotheses that guide future modeling or more in-depth analysis.

In this step we focused on the revenue and the number of employees. As earlier said, we used absolute values rather than percentage change for revenue, because startups that had just launched and went from €0 to €10,000 in revenue would otherwise skew the results significantly if we used percentage growth. Using absolute values allowed us to better capture meaningful business growth across more established companies without the distortion caused by small base effects. For the number of employees, we used the percentage of growth because an evolution for a startup from 1 employee to 2 employees is more impactful than if a startup with already 20 employees hires another employee. Basically, this part of our analysis (EDA), could include different bar charts and histograms, trend plots, and also correlation analysis among different variables, but due to lack of data, the meaningfulness of such analysis would be under question.

We could also do some descriptive analysis of some statistical measures like average etc, or also doing some statistical tests for distributions, patterns and differences. We could also do some feature analysis to find the most important variables deriving different KPIs (in this study, we restricted us to three, as we mentioned earlier revenue, number of employees and grant taking), but it could be applied. If we had enough data, we could also do simulations for handling missing data through simulations or bootstrap methods, but we couldn't, as the available data was too limited.

Random Forest Regression

Random Forest is an ensemble learning method that builds many decision trees and averages their predictions to reduce variance and overfitting. It's particularly useful when:

- We have many non-linear relationships,
- We want to rank the importance of features,
- We want robustness to outliers and missing data.

4. Analysis and results

Exploratory analysis results

1. Revenue

The first approach we chose is to compare the median revenue growth for companies that are part of the BAT ecosystem, ones that are from Euskadi but not part of BAT and the other startups that are in Spain.

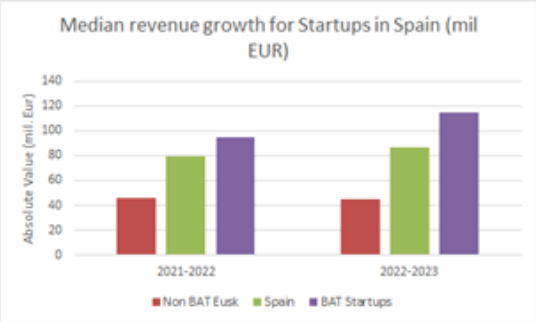
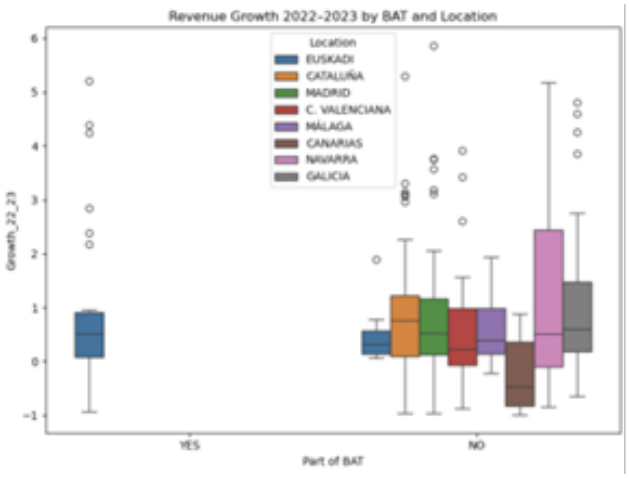


Image 1: Difference between BAT and not BAT startups in Spain.

To calculate the revenue growth, we just subtracted the revenue of the year by the one of the past years (for example for the revenue growth between 2021 and 2022: Revenue in 2022 - Revenue in 2021).

bcamath.org
Mazarredo 14, 48009 Bilbao | Basque Country, Spain | (+34) 946 567 842



As we see in the illustration above, BAT-supported Startups seem to have a higher revenue growth both times (~ +100k €/year). Moreover, the startups from Euskadi that are not part of BAT have the lowest revenue growth (~ +40k €/year).

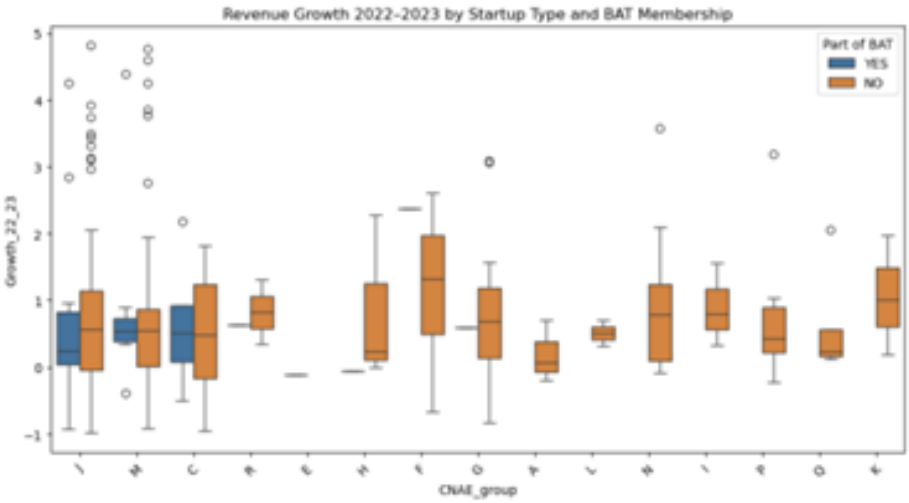


Image 2: Revenue growth (in percentages, eliminating growths higher than 600% to avoid the situations discussed above) compared to the BAT ecosystem (left) and BAT ecosystem according to the CNAE code (right).

According to the figures, it is shown that the revenue growth for BAT-supported startups, at least, are not lower than others in the whole country and for some regions are higher than others. It means they do not lose anything with likely getting higher growth than others in the same region.

2. Employment

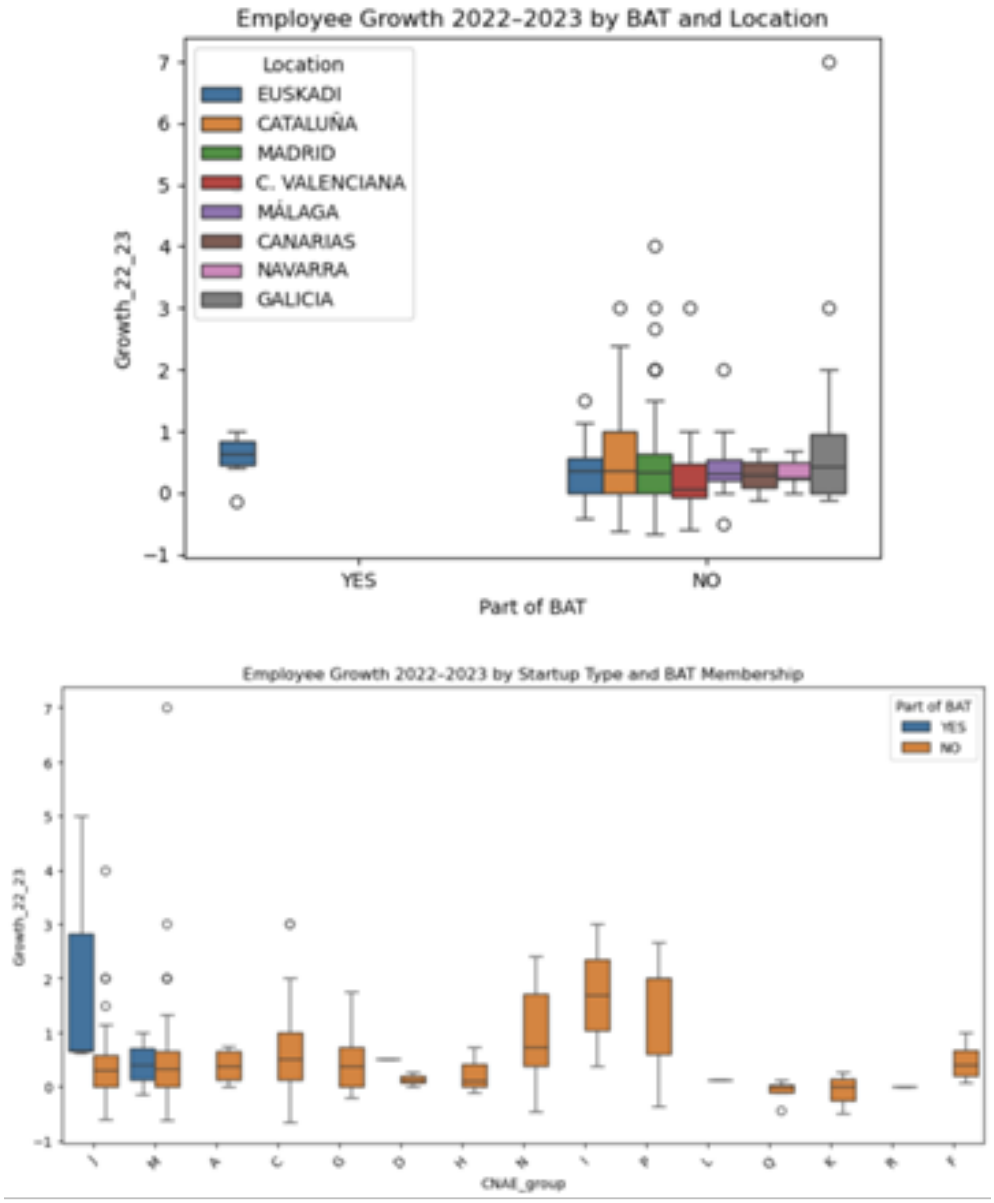


Image 3: Employee growth (in percentage) compared to the BAT ecosystem (left) and BAT ecosystem according to the CNAE code (right).

The above boxplots compares employee growth between BAT-supported and non-BAT startups within different regions. In the left fig, BAT-supported startups show higher median growth and similar variability in employee growth compared to non-BAT startups across all

regions. The BAT-supported group seems to outperform the regional average (in Basque Country and Bizkaia), which may suggest a positive effect of BAT support. As we can imply, even within the same regional context, BAT startups exhibit stronger employment growth, supporting the hypothesis that BAT plays an active role in scaling startup teams.

In the right fig, BAT's positive impact on employment growth is not uniform across sectors. Some industries seem to benefit more from BAT support — a signal that could guide targeted support strategies or sectoral prioritization.

For further explain the RF, we did an analysis of the effect of BAT by fixing other variables in RF. Holding all other features constant, being part of BAT increases the model's prediction for 2023 revenue by ~1.75 units (likely in €1,000s → ~€1,750). This can be interpreted as the marginal value of being part of BAT in the ecosystem which justifies the value of BAT in the system.

Model

To estimate the revenue of startups in the next year based on their characteristics and financial performance up to the previous years, we developed a regression model using a Random Forest Regressor. We selected this algorithm due to its ability to capture complex, non-linear relationships as we did not see any linear relationship between variables.

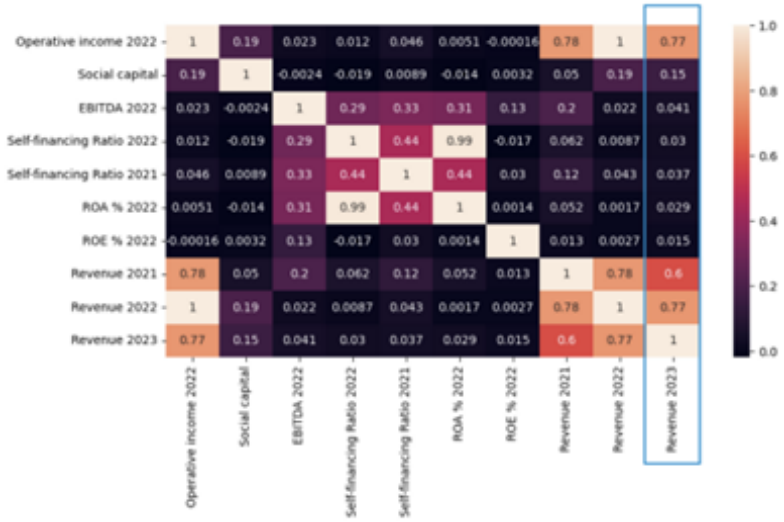


Image 4: Correlation matrix showing a strong relationship between revenue across years, particularly between Revenue 2022 and Revenue 2023 (ρ = 0.77), while other financial indicators show weak or no linear correlation with revenue.

The model was trained with:

- **Input features:** Revenue in 2021 and 2022, profitability ratios (ROA, ROE), operative income in 2022, social capital, EBITDA in 2022 and categorical features such as startup sector (CNAE), location, and whether the startup belongs to the BAT ecosystem.

- **Target variable:** Revenue in 2023.

We built the model using a scikit-learn function for random forest. First of all, the model preprocesses categorical data and then fits the random forest model. With this model it is possible to predict the revenue in the following year. If you use data from 2022-2023 the prediction will be for 2024.

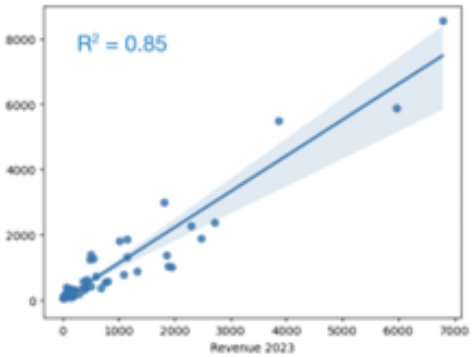


Image 5: Result of the regression model. The r- squared of this model using the test set is 0.85.

Mazarredo 14, 48009 Bilbao | Basque Country, Spain | (+34) 946 567 842

SHAP Analysis

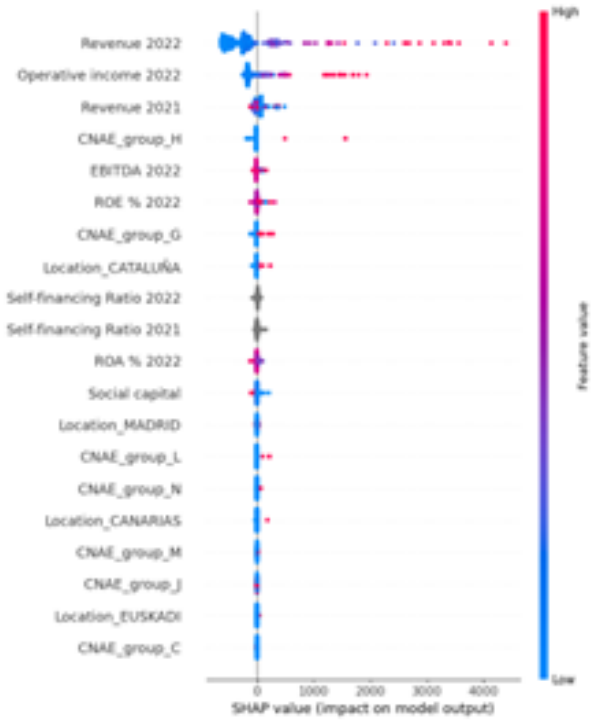


Image 6: Result of SHAP analysis.

To understand the behavior of the model, we applied SHAP (SHapley Additive exPlanations). This technique allowed us to identify the most influential features in the revenue prediction, helping us to see which economic variables have the most implications in the prediction.

As a conclusion, in the image it is shown that the most important variables are the revenue of the previous years. Also, high 2022 revenue value has a big impact on the value of the revenue of the next year.

Future improvements

Although we used a machine learning model (Random Forest) for its predictive power and flexibility, the analysis could be complemented or improved by incorporating segmented regression techniques. Segmented regression (or interrupted time series analysis) is particularly well suited to evaluate the effect of an intervention or structural change over time, such as the introduction of the BAT ecosystem in 2022. We tried testing it with existing data on this occasion, but the results were not significant so assuming that the database contains

few null values and that the annual monitoring information covers more years, this method could be successfully applied.

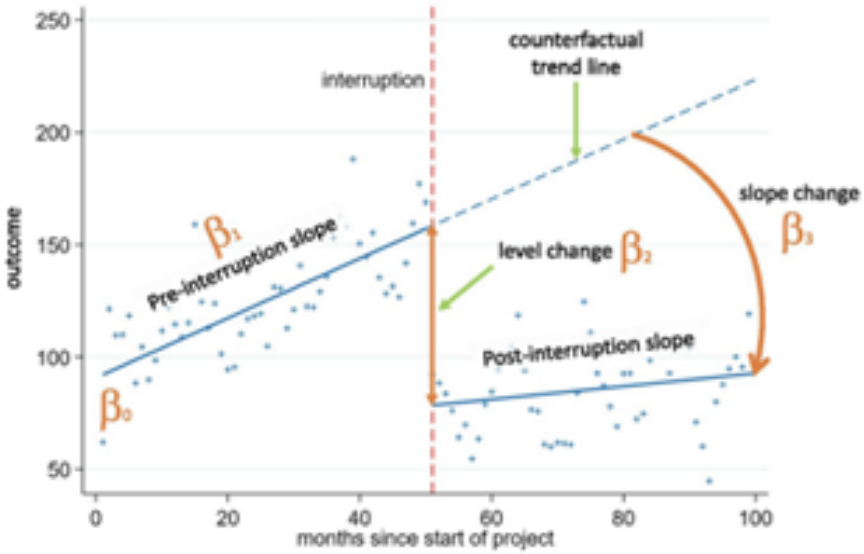


Image 7: Concept of segmented regression [1].

Value of Information

- model overfits due to tiny sample size,
- statistical tests lack power,
- comparison between BAT and non-BAT becomes fragile,
- generalization to future startups is weak.

We can use simulated BAT-supported startups to benchmark the marginal value of having more complete data. This is a way to test future data collection strategies. It is de-risking decisions before investing in more data and helping BAT justify the budget/cost to gather more complete startup data.

Simulation Procedure

- Extracting real distributions from complete BAT and non-BAT data (e.g., revenue, EBITDA, ROA, sector).

bcamath.org
Mazarredo 14, 48009 Bilbao | Basque Country, Spain | (+34) 946 567 842

- Fitting statistical distributions (normal, log-normal, or empirical) to each variable.
- Sampling from these distributions while preserving relationships (e.g., high EBITDA usually implies higher revenue).
- Assigning realistic categories like CNAE_group and Location based on observed frequencies.
  Setting "Part of BAT" = True to simulate them as BAT-supported startups.
- Then we came up with 2 scenarios
- Scenario A (realistic missing data): 7 real BAT companies + others with missing → limited insights.
- Scenario B (with simulation): 7 real BAT companies + 50 synthetic → test robustness of model, gain

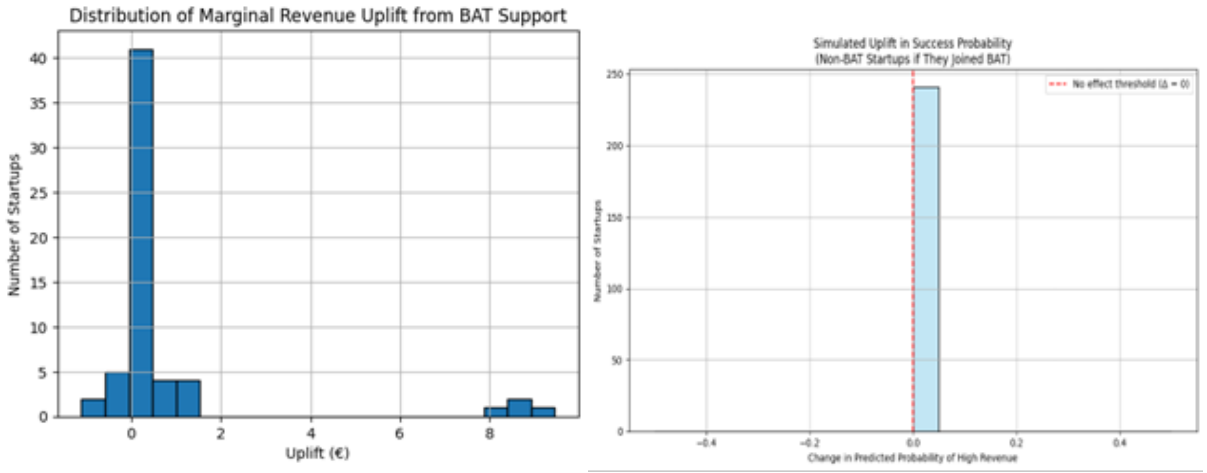


Image 10: These graphs show that getting support from BAT may have benefits for startups, in the simulated world.

Conclusions

To conclude, one of the main limitations of this analysis is the limited amount of available data, which constrains the statistical power and robustness of the findings. However, the exploratory analysis suggests that startups within the BAT ecosystem tend to show better performance indicators compared to those outside it, particularly in terms of revenue growth.

bcamath.org
Mazarredo 14, 48009 Bilbao | Basque Country, Spain | (+34) 946 567 842

While this trend is promising, a more definitive conclusion would require a larger dataset to validate these patterns with stronger evidence.

Despite these constraints, the predictive model allows us to estimate future performance based on historical data and key variables such as BAT participation, financial ratios, and sector classification. This enables scenario simulation for future years. Overall, the analysis has provided meaningful insights and has allowed us to address several of the core questions posed by the challenge.

Bibliography

[1] Turner, Simon & Karahalios, Amalia & Forbes, Andrew & Taljaard, Monica & Grimshaw, Jeremy & Mckenzie, Joanne. (2021). Comparison of six statistical methods for interrupted time series studies: empirical evaluation of 190 published series. BMC Medical Research Methodology. 21. 10.1186/s12874-021-01306-w.

Quantum mads

1. Introduction

1.1. *LQR Control for Wind Turbine Systems*

In modern wind turbine systems, achieving efficient energy conversion and maintaining system stability requires advanced control strategies. Among the various methods available, the Linear Quadratic Regulator (LQR) offers a robust approach for optimal control of dynamic systems. In our framework, the wind turbine control system is composed of three interconnected LQRs, each responsible for a different subsystem.

To initiate our analysis, we focus on the LQR designed for managing the ultracapacitor. This controller is responsible for regulating the DC Bus voltage by controlling the charge and discharge cycles of the capacitor through a component known as the bidirectional converter. Among the three LQRs considered, this one is the most straightforward, serving as a starting point for our exploration.

Figure 1.1 presents a block diagram illustrating the LQR structure associated with the ultracapacitor subsystem.

One of the initial steps in designing the LQR was to mathematically formulate the control problem. This involves defining a performance index (or cost function) that needs to be minimized, subject to a set of system constraints. The general form of such an optimization problem is given by (1.1):

$$\begin{aligned} &\text{minimize} \quad F(X), \quad \text{with } X \in \mathbb{R}^n \\ &\text{subject to} \quad g(X) \geq 0, \quad h(X) = 0 \end{aligned}$$

where:

- $X \in \mathrm{R}^n$ is the vector of decision variables,
- $F(X)$ is the objective (cost) function to be minimized,

- $g(X)$ represents inequality constraints,

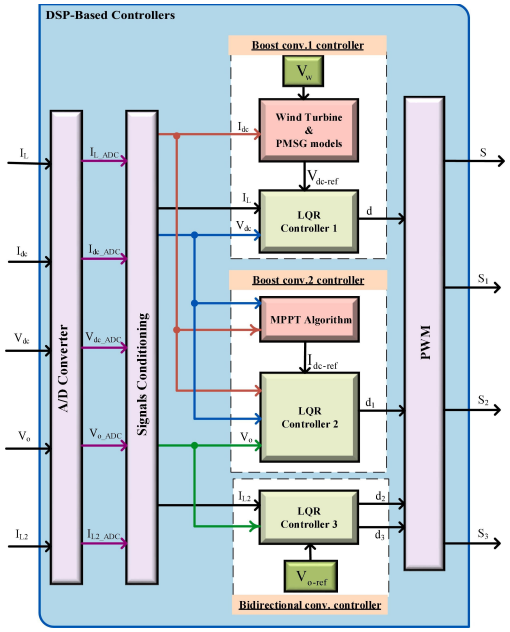


Figure 1.1: The DSP-based controllers block diagram.

- $h(X)$ denotes equality constraints.

To enable an efficient solution of the problem, the functional is typically reformulated using a Hamiltonian or equivalent representation, facilitating the derivation of optimal control laws.

1.2 Formulation of the optimization problem

The so-called LQR problem that is associated to finding the parameter values that minimize the fluctuations of the inputs and outputs of the ultracapacitor can be described as follows:

Definition 1. Let F: $R^n \times R^m \rightarrow R$ be a function of the type (1.2)

$$F(x,u) = x^T Q x + u^T R u, \quad \text{where } Q \in M_{n\times n} \text{ and } R \in M_{n\times m}.$$

Let S be a subset of curves on R^n that are sufficiently regular (i.e., S is composed of mappings

$t' \rightarrow x(t) \in R^n$ where $x(t)$ has derivatives up to a certain order and possibly satisfies some

constraints), and P and analogous subset of curves on R^m. We define the following cost function on $S \times P$: *(1.3)*

$$J[x,u] = \int_0^\infty F(x(t),u(t))dt = \int_0^\infty (x^T(t)Qx(t) + u(t)^T Ru(t))dt.$$

The goal is to, given a initial state $x_0 \in R^n$, find a curve $t \to u^*(t) \in R^m$ that minimizes the cost function J under the constraint that, for a given $t \to u(t)$, the curve $t \to x(t)$ must satisfy $x(0) = x_0$ and (1.4)

$$\begin{cases} \dot{x}(t) = Ax(t) + Bu(t), & \text{where } A \in M_{n\times n} \text{ and } B \in M_{n\times m}, \\ x(0) = x_0. \end{cases}$$

This problem can be seen as a problem of unconstrained minimization through the use of Lagrange multipliers.

Proposition 1. Let F and J be as stated above. The minimizer $t \to u*(t)$ and the corresponding trajectory $t \to x*(t)$ satisfying (1.4) for $t \to u*(t)$ are also the minimizers of the following unconstrained minimization problem with the cost function given by (1.5) and (1.6)

$$\begin{aligned} \tilde{J}[x,u,\lambda] &= \int_0^\infty (F(x(t),u(t)) + \lambda^T(t)(Ax(t) + Bu(t) - \dot{x}(t)))dt \\ &= \int_0^\infty \tilde{F}(x(t),u(t),\lambda(t),\dot{x}(t),\dot{u}(t),\dot{\lambda}(t))dt, \end{aligned}$$

Where (1.7)

$$\tilde{F}(x,u,\lambda,y,z,k) = F(x,u) + \lambda^T(Ax + Bu - y).$$

Given a Lagrangian $\tilde{F}(x, u, \lambda, y, z, k)$ as implemented in equation (1.5), its Hamiltonian is given by (1.8)

$$H(x,u,\lambda) = F(x,u) + \lambda^T(Ax + Bu).$$

Proposition 2. *Let t '→ x*(t), t '→ u*(t) and t '→ λ*(t) be the minimizer of the cost function* (1.9)

$$\tilde{J}[x,u,\lambda] = \int_0^\infty \tilde{F}(x(t),u(t),\lambda(t),\dot{x}(t),\dot{u}(t),\dot{\lambda}(t))dt.$$

Then, they satisfy the following equation (1.10)

$$H(x^*(t),u^*(t),\lambda^*(t)) = \min_{u\in\mathbb{R}^m} H(x^*(t),u,\lambda^*(t))$$

for all times t ∈ [0, ∞)

2. Direct Methods

In this chapter, we introduce several direct approaches for solving the finite-horizon Linear Quadratic Regulator (LQR) problem by mapping it to a Quadratic Unconstrained Binary Optimization (QUBO) formulation. We then leverage Fujitsu's quantum-inspired annealer to find high-quality solutions. Next, we extend the basic model to handle multiple control inputs and incorporate practical constraints. Finally, we explore how problem scaling can render classical optimization methods intractable—and how quantum-inspired techniques can overcome those challenges.

2.1. Penalty–Based Discretization

We approximate the infinite–horizon cost by a finite–horizon sum over *N* intervals of length Δ*t* = *T*/*N*. Defining

$$t_k = k\,\Delta t, \quad k = 0,\ldots,N, \qquad \Delta t = \frac{T}{N},$$

and using the explicit Euler scheme yields (2.1)

$$x_{k+1} = x_k + \Delta t\,(A\,x_k + B\,u_k) \iff x_{k+1} - \big(I - \Delta t\,A\big)x_k - \Delta t\,B\,u_k = 0.$$

We then introduce a penalty parameter $\lambda > 0$ to enforce the dynamics softly, obtaining the unconstrained cost (2.2)

$$H(x,u) = \Delta t \sum_{k=0}^{N-1} \big(x_k^\top Q\,x_k + u_k^\top R\,u_k\big) + \lambda \sum_{k=0}^{N-1} \big\| x_{k+1} - x_k - \Delta t\,(A\,x_k + B\,u_k)\big\|^2.$$

Equivalently, writing $\hat{A} = I - \Delta t\,A$, $\hat{B} = \Delta t\,B$, one has

$$H(\mathbf{x},\mathbf{u}) = \Delta t \sum_{k=0}^{N-1} \big(x_k^\top Q\,x_k + u_k^\top R\,u_k\big) + \lambda \sum_{k=0}^{N-1} \big\| x_{k+1} - \hat{A}\,x_k - \hat{B}\,u_k\big\|^2.$$

2.2. Exact Exponential Discretization

An alternative (and exact) discretization over each interval $[t_k, t_{k+1}]$ is (2.3)

$$x_{k+1} = e^{A\Delta t}x_k + \Big(\int_0^{\Delta t} e^{A\tau}\,d\tau\Big)B\,u_k = A_d\,x_k + B_d\,u_k,$$

Where

$$A_d = e^{A\Delta t}, \qquad B_d = \Big(\int_0^{\Delta t} e^{A\tau}\,d\tau\Big)B.$$

By introduction, (2.4)

$$x_k = A_d^k\,x_0 + \sum_{j=0}^{k-1} A_d^{k-j-1} B_d\,u_j.$$

2.3. QUBO Formulation

Substituting (2.4) into the finite-horizon cost $J = \Delta t \sum_{k=0}^{N-1} (x_k^\top Q x_k + u_k^\top R u_k)$, one obtains three contributions for each stage *k:*

- Constant: $(A_d^k x_0)^\top Q (A_d^k x_0)$.
- Linear in **u:** $2 \sum_{j=0}^{k-1} (A_d^k x_0)^\top Q (A_d^{k-j-1} B_d)\, u_j$.
- Quadratic in **u:** $\sum_{i,j=0}^{k-1} u_i \, (A_d^{k-i-1} B_d)^\top Q (A_d^{k-j-1} B_d)\, u_j$ plus the control penalty $u_k^\top R\, u_k$.

Collecting terms yields

$$J(\mathbf{u}) = \mathbf{u}^\top H\, \mathbf{u} \; + \; \mathbf{h}^\top \mathbf{u} \; + \; C,$$

Where

$$H_{ij} = \Delta t \sum_{k=\max(i,j)+1}^{N} (A_d^{k-i-1} B_d)^\top Q\, (A_d^{k-j-1} B_d) \; + \; \delta_{ij}\, R\, \Delta t, \tag{2.5}$$

$$h_i = 2\, \Delta t \sum_{k=i+1}^{N} (A_d^{k-i-1} B_d)^\top Q\, (A_d^k x_0), \tag{2.6}$$

$$C = \Delta t \sum_{k=0}^{N-1} (A_d^k x_0)^\top Q\, (A_d^k x_0). \tag{2.7}$$

Thus, the pure QUBO problem is

$$\min_{\mathbf{u} \in \{0,1\}^N} \; \mathbf{u}^\top H\, \mathbf{u} + \mathbf{h}^\top \mathbf{u} + C,$$

To compress it as a full QUBO problem, i.e, a problem with only quadratic operators,

we can transform our linear term setting that: (2.8)

$$H_{ij}^h = \begin{cases} h_i, & \text{if } j = i, \\ 0, & \text{else.} \end{cases}$$

then we define our quadratic operator as: (2.9)

$$F = H + H^h$$

and we will have our new problem (up to the constant): (2.10)

$$\min_{u \in \{0,1\}^N} u^\top F u,$$

2.4. Setting Smoothness to the Solution

Now we will improve our model adding the constraint that *x* must evolve smoothly. First let have the following formulation of the smooth LQR:

$$\min_{\{u_k\}_{k=0}^{\infty}} \quad \sum_{k=0}^{\infty} \left(x_k^\top Q x_k + u_k^\top R u_k \right) + \mu((\Delta x_k)^\top I(\Delta x_k)) \tag{2.11}$$

$$\text{subject to} \quad x_{k+1} = A x_k + B u_k, \tag{2.12}$$

$$x_0 \text{ given} \tag{2.13}$$

Let assume that we already discretised our model like in the previous section. Then let define: (2.14)

$$\Delta x_k = x_{k+1} - x_k = (A_d - I)x_k + B_d u_k,$$

Then we can express it as: (2.15)

$$\Delta x_k = (A_d - I)A_d^k x_0 + \sum_{j=0}^{k-1}(A_d - I)A_d^{k-j-1}B_d u_j + B_d u_k,$$

Now since we have our new model: (2.16)

$$\min \sum_{k=0}^{k-1}(x_k^\top Q x_x + u^\top R u)\Delta t + \mu((\Delta x_k)^\top I(\Delta x_k))$$

Following the same processes of the non-smooth model, we will get the following Quadratic, lineal and constant matrices:

$$H_{ij}^* = H_{ij} + \mu \sum_{k=\max(i,j)}^{N-2} M_{ki}^\top M_{kj}, \tag{2.17}$$

$$h_i^* = h_i + 2\mu \sum_{k=j}^{N-2} M_{ij}^T (A_d - I)A_d^k x_0, \tag{2.18}$$

$$C^* = C + \mu \sum_{k=0}^{N-2} ||(A_d - I)A_d^k x_0||^2, \tag{2.19}$$

Now we have the new QUBO problem with a new constraint, but we maintain the same dimensions of the new main F^* matrix.

We can keep adding some restrictions, however the dimension of the QUBO problem will not increase if we apply this restriction since the beginning of the original problem formulation. Therefore, constraints will not affect the complexity of the QUBO problem.

2.5. Fujitsu Digital Annealer

Every physical system tends to its state of minimum energy. This fact it's used by the

Quantum Annealers by codifying an Ising system; a physical system based on the interactions of spins, using the QUBO model to configure the value of interactions between spins. By slowly introducing the Hamiltonian of the Ising system, you let the system evolve naturally to its minimum energy state; where the solution to our problem is codified. This combined with quantum properties like (ii) superposition; that allows encountering the system in multiple configurations at the same time, and (ii) the tunnelling effect; useful to avoid high energy barriers, these kinds of processes guarantee that we reach the optimal solution given enough time; theoretically faster than the classical solvers.

Nowadays, there exists this kind of machine; D'Wave system it's a successful example of this; but, due to the technical issues with maintaining the superposition property of the system and the multiple sources of error that this machine has due to the actual NISQ era (Noisy intermediate-scale quantum era), these machines could not be used to resolve real problems yet.

In the meantime, Fujitsu Ltd. has created the Digital Annealer, DA for short, a specialized computer to run a Simulated Annealing algorithm (SA). The differential factor of DA with their competitors consists in a massive parallelized computer that can evaluate the QUBO model in multiple positions at the same time; and, by allowing a controlled noise in the movement, the system can surpass high energy barriers of the target function; emulating the tunnelling property of the quantum systems.

2.6. Conclusions

Quantum annealing via a QUBO formulation presents several compelling advantages over classical control approaches such as the Linear Quadratic Regulator (LQR):

1. **Handling Discrete and Combinatorial Decisions.** QUBO naturally encodes binary variables, making it ideal for problems with on/off actions, scheduling, or mode switches that classical LQR cannot handle directly.

2. **Escaping Local Minima.** Quantum tunneling in annealers helps traverse rugged energy landscapes and avoid entrapment in suboptimal local minima, a frequent pitfall for gradient-based or local-search classical methods.

3. **Parallel Exploration via Superposition.** Quantum annealers can explore many candidate solutions simultaneously, potentially accelerating convergence compared to the inherently sequential updates of classical optimizers.

4. **Flexible Constraint Modeling.** Complex requirements—nonlinear dynamics, stochastic disturbances, multi-objective trade-offs, and smoothness or rate-of-change penal- ties—can all be incorporated directly into the QUBO cost function by adding appropriate quadratic and linear terms.

Moreover, the QUBO+annealer paradigm scales beyond wind-turbine control into a variety of engineering and industrial domains:

- **Energy Systems:** Unit commitment, load balancing, and grid-wide scheduling.
- **Transportation:** Traffic signal optimization, vehicle routing, and logistics planning.
- **Manufacturing:** Job-shop scheduling, resource allocation, and supply-chain optimization.
- **Healthcare:** Staff rostering, operating-room scheduling, and patient-flow management.
- **Finance:** Portfolio optimization under risk and return constraints.

Finally, the comparative visualizations of state trajectories and control inputs demonstrate that:

$$\min_{\mathbf{u}\in\{0,1\}^N} \left(\mathbf{u}^\top H\,\mathbf{u} + \mathbf{h}^\top \mathbf{u}\right) \quad \text{(QUBO + annealer)} \quad \text{vs.} \quad \min_K \int \left(x^\top Q x + u^\top R u\right) dt \quad \text{(classical LQR)}$$

highlight both the competitive performance and the enhanced flexibility of the QUBO approach—especially when augmented with penalties for smoothness, stochastic robustness, or nonlinear effects—that make it a powerful tool for next-generation control and optimization tasks. (2.20)

$$A^T P + PA - PBR^{-1}B^T P + Q = 0$$

bcamath.org Mazarredo 14, 48009 Bilbao | Basque Country, Spain | (+34) 946 567 842

2.7. Classical LQR vs QUBO-LQR: Comprehensive Comparison for 100 Wind Turbines

2.7.1. Problem Definition Comparison

Aspect	Classical LQR	QUBO-LQR
Problem Type	Continuous optimization problem	Discrete binary optimization problem
Objective Function	$\min_{u[k]} J = \sum_{k=0}^{\infty} \left[x[k]^T Q x[k] + u[k]^T R u[k]\right]$	$\min_b b^T Q_{\text{QUBO}} b$
Decision Variables	$u[k] \in \mathbb{R}^{100}$	$b \in \{0,1\}^{8000}$
State Vector	$x[k] \in \mathbb{R}^{200}$	Same: $x[k] \in \mathbb{R}^{200}$
System Dynamics	$x[k+1] = Ax[k] + Bu[k]$	$x[k+1] = Ax[k] + Bu[k], \quad u_i[k] = \sum_{j=0}^{7} 2^j b_{i,j}[k] - 128$
Control Range	$u_i[k] \in [-5, 5]$	$u_i[k] \in \{-128, \ldots, 127\}$
Constraints	Penalty methods or constrained optimization	Penalty terms in QUBO matrix
Coupling	$\sum_{i=1}^{100} P_i u_i[k] = P_{\text{total}}[k]$	$\lambda \left(\sum_{i=1}^{100} P_i u_i[k] - P_{\text{total}}[k]\right)^2$

2.7.2. Mathematical Formulation Comparison

Component	Classical LQR	QUBO-LQR
System Matrix	$A = \text{diag}(A_1, \ldots, A_{100}),\ A_i = \begin{bmatrix} 1 & -0.01 \\ 0.01 & 0.999 \end{bmatrix}$	Same structure
Input Matrix	$B = \text{diag}(B_1, \ldots, B_{100}),\ B_i = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$	Same structure
Weight Matrices	$Q = \text{diag}(Q_1, \ldots, Q_{100}),\ Q_i = \begin{bmatrix} 1e^{-3} & 0 \\ 0 & 10 \end{bmatrix},\ R = \text{diag}(10, \ldots, 10)$	Embedded in QUBO
Solution Method	Riccati equation	Quantum annealing or QUBO solvers
Control Law	$u[k] = -Kx[k], \quad K = (R + B^T P B)^{-1} B^T P A$	$u[k] = f(b^*)$
Binary Encoding	Not applicable	$u_i[k] = \sum_{j=0}^{7} 2^j b_{i,j}[k] - 27$
QUBO Matrix Size	Not applicable	8000×8000

bcamath.org Mazarredo 14, 48009 Bilbao | Basque Country, Spain | (+34) 946 567 842

2.7.3. Computational Complexity Comparison

Metric	Classical LQR	QUBO-LQR
Problem Size	$n = 200$, $m = 100$	8000 binary variables
Time Complexity	$O(n^3) = O(8 \times 10^6)$	Quantum: $O(1)$, Classical: $O(2^{8000})$
Space Complexity	$O(n^2) = 40{,}000$	$O(m^2) = 64 \times 10^6$
Solution Time	Milliseconds	Quantum: Seconds, Classical: Hours/Days
Scalability	Poor	Excellent
Memory Usage	≈ 1.6 MB	≈ 512 MB
Hardware Requirements	CPU/GPU	Quantum annealer or HPC

2.7.4. Performance and Capabilities Comparison

Capability	Classical LQR	QUBO-LQR
Optimality	Global for continuous	Global for discretized
Real-time Performance	Excellent	Good
Constraint Handling	Limited	Excellent
Discrete Constraints	Difficult	Natural
Nonlinear Constraints	Challenging	Moderate
Multi-objective Optimization	Limited	Good
Robustness	High	Moderate
Implementation Complexity	Low	High

2.7.5. Practical Implementation Comparison

Aspect	Classical LQR	QUBO-LQR
Development Time	Days to weeks	Months
Software Tools	MATLAB, Python, C++	D-Wave Ocean, Qiskit
Hardware Cost	$1,000 - $10,000	$15M+ or $100K+
Maintenance	Low	High
Expertise Required	Control engineering	Quantum computing + optimization
Debugging	Straightforward	Complex
Validation	Established methods	Emerging methods
Industry Readiness	Mature	Experimental

2.7.6. Specific Example: Wind Turbine System

Parameter	Classical LQR	QUBO-LQR
LQR Gains	$K = [1.599, 1.270]$	Preserved per turbine
Control Formula	$u[k] = -1.599e_v[k] - 1.270e_i[k]$	$u_i[k] = \sum_{j=0}^{7} 2^j b_{i,j}[k] - 128$
Duty Cycle	$d[k] = 0.5 + 0.1u[k]$	Same formula per turbine
System Size	1 turbine, 2 states, 1 control	100 turbines, 200 states, 100 controls
Simulation Time	¡1 sec (20 sec sim)	10-60 sec (incl. solving)
Accuracy	Continuous precision	8-bit (256 levels)
DC Bus Coupling	None	$\sum_{i=1}^{100} P_i u_i = P_{\text{total}}$
Wind Speed Handling	Step change	Individual per turbine

2.7.7. Summary and Recommendations

Scenario	Recommended Approach
Small Scale (1-10 turbines)	Classical LQR
Medium Scale (10-50 turbines)	Classical LQR with approximations
Large Scale (50-100+ turbines)	QUBO-LQR
Complex Constraints	QUBO-LQR
Real-time Critical	Classical LQR
Research/Development	QUBO-LQR
Industrial Deployment	Classical LQR

Danobatgroup

1. Introduction

In high-precision manufacturing, Computerized Numerical Control (CNC) machine tools remain the standard for metal cutting processes, due to their high stiffness and limited degrees of freedom. These properties allow for highly accurate motion control and surface finish quality. However, CNC machines are expensive, difficult to relocate, and limited in their ability to adapt to complex geometries. In contrast, industrial robots offer greater affordability, flexibility, and ease of reconfiguration. Their ability to reach complex tool paths makes them an attractive alternative for advanced manufacturing. Despite these advantages, robots suffer from several downsides in the context of metal cutting: low joint stiffness, structural vibrations, positioning errors, and joint backlash. These limitations give rise to dynamic instabilities, including unwanted vibrations that degrade machining performance. Understanding and controlling these instabilities is an active area of research. A particularly critical phenomenon is regenerative chatter, where periodic variations in chip thickness caused by tool vibrations lead to self-excited oscillations. To tackle this, researchers must develop physics-based models that accurately capture the interaction between the robot dynamics and the cutting process.

2. Objectives and scope of the challenge

The goal of this challenge was to:

- Model the dynamics of a robotic cutting process, including chip regeneration effects.
- Formulate the governing equations as a delay differential system that reflects the time-dependent interaction between the cutting tool and the workpiece.
- Analyze the stability of this mathematical model using spectral methods.

While the problem is analytically complex and explicit conditions for stability could not be derived in general, we succeeded in:

- Deriving a delay-based model from first physical principles and geometrical assumptions.
- Identifying the structure of the system, including the influence of delayed vibration feedback on cutting forces.

- Expressing the model as a linear delay differential equation with periodic coefficients.
- Discussing qualitative stability conditions and outlining a numerical framework for future root-counting and simulation-based investigation.

This report documents the modeling process, the derivation of the dynamic system, and the analytical tools used to investigate its stability. Our work provides a solid foundation for future efforts in simulating, controlling, and stabilizing robotic metal cutting processes.

3. Modeling the Cutting Process

We assume that the tool moves in the plane, so we can describe its motion as $r(t) = r_0(t) + r_1(t)$, where t is the time, r_0 is the nominal motion, and r_1 is the departure from equilibrium; we can also see r as a complex-valued function. In DANOBAT's experiments, r_0 traces a circle.

The position of k-th edge is (1)

$$\theta_k(t) = \Omega t + 2\pi k/Z$$

where Ω[rad/s] > 0 is the speed of rotation, and Z is the number of teeth. The position of edge $k + 1$ at some time t_- is (2)

$$r(t_-) + Re^{i\theta_{k+1}(t_-)}$$

where R is the radius of the tool. Then, at time $t = t_- + \frac{2\pi}{\Omega Z} + \Delta t$ edge k starts engaging with the material, and it intersects the profile at the point (3)

$$r(t) + (R - h)e^{i\theta_k(t)}$$

where $h \geq 0$ is the chip thickness. Therefore, (4)

$$r(t_-) + Re^{i\theta_{k+1}(t_-)} = r(t) + (R - h)e^{i\theta_k(t)}$$

where we have to unknowns h and Δt. We reorganize the equation so that (5)

$$r_0(t) - r_0(t_-) + r_1(t) - r_1(t_-) = e^{i\theta_k(t)}[R(e^{i(\theta_{k+1}(t_-) - \theta_k(t)} - 1) + h]$$

where $\theta_{k+1}(t_-) - \theta_k(t) = -\Omega\Delta t$. The system we have to solve is nonlinear, so we exploit some simplifications.

3.1 Chip Thickness Derivation

The nominal motion r_0 is usually smooth and has negligible variations, so we approximate it as $r_0(t_-) - r_0(t) = r_0'(t)(t_- - t) + \mathcal{O}(t_- - t)^2$; actually, we will also assume that r''_0 is negligible. Assuming Δt is infinitesimally small, to first order (5) gets into(6)

$$\frac{2\pi r_0'(t)}{\Omega Z} + r_0'(t)\Delta t + r_1(t) - r_1(t - 2\pi/(\Omega Z)) + r_1'(t - 2\pi/(\Omega Z))\Delta t = e^{i\theta_k(t)}(-iR\Omega\Delta t + h)$$

(a) (b)

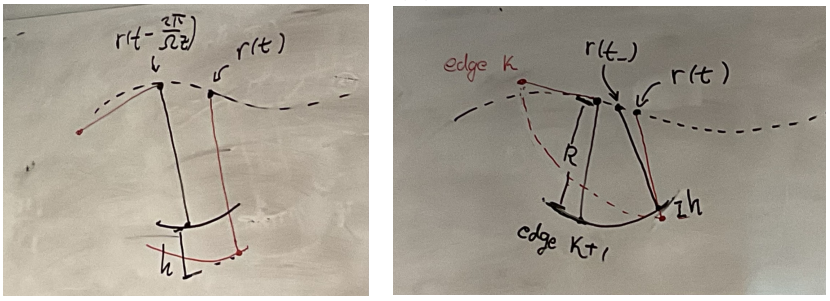


Figure 1: Chip thickness as computed (a) in Chapter 4.6 of [1], and (b) in this report. The chip thickness in (a) can be seen as an approximation of that in (b) when vibrations are moderated.

we will assume that vibrations are not very strong so that we can neglect the term $r_1'(t - 2\pi/(\Omega Z))\Delta t$. In this way, we have reached (7) (8)

$$a\Delta t + bh = c \tag{7}$$

$$a = -r_0'(t) - iR\Omega e^{i\theta_k(t)}, \quad b = e^{i\theta_k(t)}, \quad \text{and} \quad c = \frac{2\pi r_0'(t)}{\Omega Z} + \Delta r_1(t). \tag{8}$$

If we multiply (7) by $\overline{a}$ and take imaginary part we have that (9)

$$h_k = \frac{\mathrm{Im}(\bar{a}c)}{\mathrm{Im}(\bar{a}b)} = \frac{(2\pi/(\Omega Z))\,\mathrm{Re}(-i\bar{a}r_0'(t)) + \mathrm{Re}(-i\bar{a}\Delta r_1)}{R\Omega + \mathrm{Re}(\overline{r_0}'(t) i e^{i\theta_k(t)})}$$

3.2 Cutting Force Formulation

The force has two components: tangential and normal. The magnitudes of the forces are given by $F_t = K_t a h$ and $F_n = K_n a h$. Since the direction of the edge is $e^{i\theta_k}$, the total force on edge *k* is: (11)

$$F_k = \left(-ie^{i\theta_k(t)} K_t a h_k - e^{i\theta_k(t)} K_n a h_k\right) g(\theta_k)$$

Where *g* is the indicator function: (12)

$$g(\theta) = \begin{cases} 1, & \theta_{\text{in}} \le \theta \le \theta_{\text{in}} + \varphi, \\ 0, & \text{otherwise}, \end{cases}$$

With θ_{in} and $\theta_{\text{in}} + \varphi$ being the initial and exit angles, respectively. The value of φ depends on the setup of the cutting process, while θ_{in} depends on *r* and corresponds to the angle where *h* = 0, To avoid complexities, we assume $r_1 = 0$ so that:

$$v_{f1}\cos\theta_k + v_{f2}\sin\theta_k = 0 \quad \Rightarrow \quad \theta_{\text{in}} = -\arctan\left(\frac{v_{f2}}{v_{f1}}\right)$$

The total *force F = F0 + F1 + · · ·* is:

$$F(t) = -a\sum_{k=0}^{Z-1} h_k e^{i\theta_k(t)}(K_n + iK_r)g(\theta_k) \tag{13}$$

$$= -a\sum_{k=0}^{Z-1} e^{i\theta_k(t)}(K_n + iK_r)g(\theta_k)(\cos\theta_k(t), \sin\theta_k(t)) \cdot (v_f + \Delta r_1(t)) \tag{14}$$

$$= F_0(t) + F_1(t), \tag{15}$$

Where

$$F_0(t) = -a(K_n + iK_r) \sum_{k=0}^{Z-1} e^{i\theta_k(t)} g(\theta_k)(\cos\theta_k(t), \sin\theta_k(t)) \cdot v_f, \tag{16}$$

$$F_1(t) = -a(K_n + iK_r) \sum_{k=0}^{Z-1} e^{i\theta_k(t)} g(\theta_k)(\cos\theta_k(t), \sin\theta_k(t)) \cdot \Delta r_1(t). \tag{17}$$

If we change variables to $\theta = \Omega t$ then $t = \theta/\Omega$. That means $\Delta r_1(t)$ becomes $\Delta r_1(\theta)$, and the expressions become:

$$F_0(\theta) = -a(K_n + iK_r) \sum_{k=0}^{Z-1} e^{i\theta_k(\theta)} g(\theta_k) \left[\cos\theta_k(\theta) \cdot v_{f1} + \sin\theta_k(\theta) \cdot v_{f2}\right]$$

$$F_1(\theta) = -a(K_n + iK_r) \sum_{k=0}^{Z-1} e^{i\theta_k(\theta)} g(\theta_k) \left[\cos\theta_k(\theta) \cdot \Delta x(\theta) + \sin\theta_k(\theta) \cdot \Delta y(\theta)\right]$$

Hence, the vibrations r_1 follow the equation (18):

$$M\ddot{r}_1 + C\dot{r}_1 + Kr_1 = F_0(t) + F_1(t), \quad \text{where } \frac{dr_1}{dt} = \dot{r}_1$$

4. Derivation of the Delay Differential System

After th

e change of variables θ = Ωt, the position of the *k*-th edge becomes (19):

$$\theta_k(t) = \theta + \frac{2\pi k}{Z}$$

and the system (18) becomes (20):

$$r_1'' + \frac{M^{-1}C}{\Omega} r_1' + \frac{M^{-1}K}{\Omega^2} r_1 = \frac{M^{-1}}{\Omega^2} \left(F_0(\theta) + F_1(\theta)\right), \quad \text{where } \frac{dr_1}{d\theta} = r_1'.$$

To convert this into a first-order system, define the state vector (21):

$$q(\theta) = \begin{bmatrix} r_1(\theta) \\ r_1'(\theta) \end{bmatrix}$$

$$q' = \underbrace{f}_{\text{from } F_0(\theta)} + \underbrace{A(\theta)q(\theta) + B(\theta)q(\theta - \tau)}_{\text{from delayed part of } F_1(\theta)}$$

Here *f* is a small term corresponding to the contribution from F_0, while the stability of the system is primarily determined by the force component F_1. The state vector is $q = (r_1, r_1')$ and $A(\theta)$ and $B(\theta)$ are τ - periodic matrices, with $\tau = \frac{2\pi}{Z}$.

It is recommended to study the system's behavior when is r_0' *not constant, as it may significantly influence the dynamics.*

Assuming *f* = 0 for simplicity, the system reduces to (22):

$$q' = A(\theta)q + B(\theta)q(\theta - \tau)$$

Assuming A(θ) and B(θ) are τ -periodic, we try a solution of the form (23):

$$q(\theta) = ve^{\lambda\theta}$$

$\bar{A} = \frac{1}{\tau}\int_0^\tau A(\theta)\,d\theta.$ Substituting into the equation gives (24):

$$\lambda v = \left(\bar{A} + e^{-\lambda\tau}\bar{B}\right) v$$

This leads to the characteristic equation (25):

$$\det\left(\bar{A} + e^{-\lambda\tau}\bar{B} - \lambda I\right) = 0$$

Stability is guaranteed if all solutions λ satisfy Re(λ) < 0. We are particularly interested in the real part of the spectrum λ to analyze the system's stability.

4.1 Explicit Formulation of the Delay System

We denote the vibration displacement vector *r1* as:

$$r_1(\theta) = \begin{pmatrix} x(\theta) \\ y(\theta) \end{pmatrix}$$

To **compute the matrices** A(θ) **and** B(θ), we start by defining the delayed terms:

$$\Delta x(\theta) = x(\theta) - x(\theta - \tau), \quad \Delta y(\theta) = y(\theta) - y(\theta - \tau)$$

The full system is expressed as (26):

$$q'(\theta) = \begin{bmatrix} r_1'(\theta) \\ r_1''(\theta) \end{bmatrix} = \underbrace{\begin{pmatrix} 0 & I \\ -\frac{M^{-1}K}{\Omega^2} & -\frac{M^{-1}C}{\Omega} \end{pmatrix}}_{\text{System dynamics}} \begin{pmatrix} r_1(\theta) \\ r_1'(\theta) \end{pmatrix} + \underbrace{\begin{pmatrix} 0 \\ \frac{M^{-1}F_0(\theta)}{\Omega^2} \end{pmatrix}}_{f(\theta)} + \underbrace{V(\theta)}_{\text{Delay influence}}$$

To extract the S_x and S_y matrices from the expression for $F_1(\theta)$, you need to identify the coefficients that multiply $\Delta x(\theta)$ and $\Delta y(\theta)$ in the summation. These coefficients are essentially what define S*x* and S*y* in the system formulation.

— In the system dynamics, $F_1(\theta)$ contributes to the delayed feedback through terms

of the form:

$$V(\theta) = \frac{M^{-1}}{\Omega^2}\left[S_x \Delta x(\theta) + S_y \Delta y(\theta)\right]$$

So, from:

$$F_1(\theta) = -a(K_n + iK_r)\sum_{k=0}^{Z-1} e^{i\theta_k(\theta)} g(\theta_k)\left[\cos\theta_k(\theta)\cdot\Delta x(\theta) + \sin\theta_k(\theta)\cdot\Delta y(\theta)\right]$$

we can factor out Δ*x*(θ) and Δ*y*(θ) to isolate their coefficients:

$$F_1(\theta) = -a(K_n + iK_r)\begin{bmatrix}\sum_{k=0}^{Z-1} e^{i\theta_k(\theta)} g(\theta_k)\cos\theta_k(\theta) \\ \sum_{k=0}^{Z-1} e^{i\theta_k(\theta)} g(\theta_k)\sin\theta_k(\theta)\end{bmatrix}\cdot\begin{bmatrix}\Delta x(\theta) \\ \Delta y(\theta)\end{bmatrix},$$

$$F_1(\theta) = -a(K_n + iK_r)\begin{bmatrix}\sum_{k=0}^{Z-1} e^{i\theta_k(\theta)} g(\theta_k)\cos\theta_k(\theta) \\ \sum_{k=0}^{Z-1} e^{i\theta_k(\theta)} g(\theta_k)\sin\theta_k(\theta)\end{bmatrix}\cdot\begin{bmatrix}(x(\theta) - x(\theta-\tau)) \\ (y(\theta) - y(\theta-\tau))\end{bmatrix},$$

We define:

$$S_x = -a(K_n + iK_r)\sum_{k=0}^{Z-1} e^{i\theta_k(\theta)} g(\theta_k)\cos\theta_k(\theta),$$

$$S_y = -a(K_n + iK_r)\sum_{k=0}^{Z-1} e^{i\theta_k(\theta)} g(\theta_k)\sin\theta_k(\theta).$$

Delay Term Expansion:

$$V(\theta) = \begin{pmatrix} 0 \\ \frac{M^{-1}}{\Omega^2}\left[S_x(x(\theta) - x(\theta-\tau)) + S_y(y(\theta) - y(\theta-\tau))\right]. \end{pmatrix}$$

We now express this in matrix-vector form using the structure:

$$\begin{pmatrix} 0 \\ \frac{M^{-1}}{\Omega^2} S_x x(\theta) \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} x(\theta) \\ y(\theta) \\ x'(\theta) \\ y'(\theta) \end{pmatrix}$$

Where

$$A = B = D = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \quad C = \frac{M^{-1}}{\Omega^2} \begin{pmatrix} S_{x_0} & S_{y_0} \\ S_{x_1} & S_{y_1} \end{pmatrix}$$

Similarly, combining all delay contributions, we have

$$V(\theta) = \underbrace{\begin{pmatrix} 0_{2\times 2} & 0_{2\times 2} \\ \frac{M^{-1}}{\Omega^2}\begin{pmatrix} S_{x_0} & S_{y_0} \\ S_{x_1} & S_{y_1} \end{pmatrix} & 0_{2\times 2} \end{pmatrix}}\begin{pmatrix} x(\theta) \\ y(\theta) \\ x'(\theta) \\ y'(\theta) \end{pmatrix} + \underbrace{\begin{pmatrix} 0_{2\times 2} & 0_{2\times 2} \\ \frac{M^{-1}}{\Omega^2}\begin{pmatrix} S_{x_0} & S_{y_0} \\ S_{x_1} & S_{y_1} \end{pmatrix} & 0_{2\times 2} \end{pmatrix}}\begin{pmatrix} x(\theta-\tau) \\ y(\theta-\tau) \\ x'(\theta-\tau) \\ y'(\theta-\tau) \end{pmatrix}, \tag{27}$$

$$V(\theta) = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} x(\theta) \\ y(\theta) \\ x'(\theta) \\ y'(\theta) \end{pmatrix} + \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} x(\theta-\tau) \\ y(\theta-\tau) \\ x'(\theta-\tau) \\ y'(\theta-\tau) \end{pmatrix}. \tag{28}$$

Final system equation (29)**:**

$$q'(\theta) = A(\theta)q(\theta) + B(\theta)q(\theta-\tau) + f(\theta)$$

Where

$$A(\theta) = \begin{pmatrix} 0_{2\times 2} & I_{2\times 2} \\ -\frac{M^{-1}K}{\Omega^2} + \frac{M^{-1}}{\Omega^2}\begin{pmatrix} S_{x_0} & S_{y_0} \\ S_{x_1} & S_{y_1} \end{pmatrix} & -\frac{M^{-1}C}{\Omega} \end{pmatrix} = \begin{pmatrix} 0_{2\times 2} & I_{2\times 2} \\ -\frac{M^{-1}K}{\Omega^2} + \frac{M^{-1}}{\Omega^2}(S_x \quad S_y) & -\frac{M^{-1}C}{\Omega} \end{pmatrix},$$

$$B(\theta) = \begin{pmatrix} 0_{2\times 2} & 0_{2\times 2} \\ \frac{M^{-1}}{\Omega^2}(S_x \quad S_y) & 0_{2\times 2} \end{pmatrix},$$

$$f(\theta) = \begin{pmatrix} 0 \\ \frac{M^{-1}F_0(\theta)}{\Omega^2} \end{pmatrix}.$$

4.1.1 Special Case: stiffness in one direction: (y(ϑ) = 0)

Assume the arm is very stiff in the transverse direction and does not move:

$$y(\theta) = 0, \quad y'(\theta) = 0, \quad y(\theta - \tau) = 0, \quad y'(\theta - \tau) = 0$$

We define the reduced state vector:

$$q_x(\theta) = \begin{pmatrix} x(\theta) \\ x'(\theta) \end{pmatrix}$$

Then

$$A(\theta) = \begin{pmatrix} 0_{2\times 2} & I_{2\times 2} \\ -\frac{M^{-1}K}{\Omega^2} + \frac{M^{-1}}{\Omega^2}\begin{pmatrix} S_{x_0} & 0 \\ S_{x_1} & 0 \end{pmatrix} & -\frac{M^{-1}C}{\Omega} \end{pmatrix}$$

$$B(\theta) = \begin{pmatrix} 0_{2\times 2} & 0_{2\times 2} \\ \frac{M^{-1}}{\Omega^2}\begin{pmatrix} S_{x_0} & 0 \\ S_{x_1} & 0 \end{pmatrix} & 0_{2\times 2} \end{pmatrix},$$

$$f(\theta) = \begin{pmatrix} 0 \\ \frac{M^{-1}F_0(\theta)}{\Omega^2} \end{pmatrix}.$$

5. Stability Analysis of the Model

5.1 Counting roots

We are interested in the stability of (30)

$$x'' + ax' + b + cx(\cdot - \tau) = 0,$$

so solutions are similar to $x = e^{\lambda t}$ then $f(\lambda) = \lambda^2 + a\lambda + b + ce^{-\lambda\tau}$

The number of roots of a function *f* inside a curve Γ is (31)

$$N = -\frac{1}{2\pi i}\int_{\Gamma}\frac{f'(\lambda)}{f(\lambda)}\,d\lambda.$$

To test the stability, we will apply this to the function $f(\lambda) = \lambda^2 + a\lambda + b + ce^{-\tau\lambda}$, taking as Γ the union of the right half of the semi-circle with radius R → ∞, and the imaginary axis.

We parametrize the semi-circle Γ_1 as $Re^{-i\phi}$, for $\phi \in [-\pi/2, \pi/2]$, and the imaginary axis Γ_2 as *is*, for $s \in$ [−R, R]. Since $p(\lambda) = \lambda^2 + a\lambda + b$ has no root in the right-half plane when a > 0, then (32)

$$N = -\frac{1}{2\pi i}\int_{\Gamma}\left(\frac{f'(\lambda)}{f(\lambda)} - \frac{p'(\lambda)}{p(\lambda)}\right)d\lambda.$$

The term *p′/p* is introduced to make the integral more stable for numerical computations.

In the semi-circle Γ_1 we have (33)

$$\frac{iR}{2\pi i}\int_{-\pi/2}^{\pi/2}\left(\frac{f'(Re^{-i\phi})}{f(Re^{-i\phi})}-\frac{p'(Re^{-i\phi})}{p(Re^{-i\phi})}\right)e^{-i\phi}d\phi.$$

Since $|e^{-\tau Re^{-i\phi}}| = e^{-R\cos\phi} < 1 \text{ in } \Gamma_1$, and the integrand is bounded by $1/R^2$, then the integral vanishes in the limit when $R \to \infty$. It remains to compute the integral in the imaginary axis

$$-\frac{1}{2\pi}\int_{\mathbb{R}}\left(\frac{f'(is)}{f(is)}-\frac{p'(is)}{p(is)}\right)ds = -\frac{1}{2\pi}\int_{\mathbb{R}}\left(\frac{1}{f(is)}-\frac{1}{p(is)}\right)f'(is)\,ds+\frac{\tau c}{2\pi}\int_{\mathbb{R}}\frac{e^{-i\tau s}}{p(is)}ds$$
$$=\frac{c}{2\pi}\int_{\mathbb{R}}\frac{e^{-i\tau s}f'(is)}{p(is)f(is)}\,ds+\frac{\tau c}{2\pi}\int_{\mathbb{R}}\frac{e^{-i\tau s}}{p(is)}\,ds. \qquad (34)$$

We can compute the second integral explicitly. For that, notice that p(*is*) = $-(s-\mu_1)(s-\mu_2)$ where the roots lie in the upper half-plane and $\mu_2-\mu_1 = \sqrt{4b-a^2} > 0$, so (35)

$$\int\frac{e^{-i\tau s}}{p(is)}\,ds=\frac{1}{\mu_2-\mu_1}\int\frac{e^{-i\tau s}}{s-\mu_1}-\frac{e^{-i\tau s}}{s-\mu_2}\,ds.$$

Since τ > 0, the integrals vanish, and we get that the number of roots in the right-half plane is (36)

$$\frac{c}{2\pi}\int_{\mathbb{R}}\frac{e^{-i\tau s}f'(is)}{p(is)f(is)}\,ds.$$

Since the integrand decays as $1/s^3$, the integral is finite and well-defined

References

[1] Altintas Yusuf. Manufacturing automation. Cambridge University Press, 2nd edition, 2012.

List of participants

- Jorge Angulo – Basque Center for Applied Mathematics (BCAM)
- Garikoitz Artola – University of the Basque Country (EHU)
- Hossein Babazadeh – Norwegian University of Science and Technology (NTNU)
- Unax Barrainkua – Basque Center for Applied Mathematics (BCAM)
- Kartheek Reddy Bondugula – Basque Center for Applied Mathematics (BCAM)
- Sandra Burgos – University of the Basque Country (EHU)
- Ethève Cédric – Basque Center for Applied Mathematics (BCAM)
- Meriem Essadik – Basque Center for Applied Mathematics (BCAM)
- Mikel Pérez Arrieta – Basque Center for Applied Mathematics (BCAM)
- Yolan Perón – Basque Center for Applied Mathematics (BCAM)
- Kanika Rajain – Basque Center for Applied Mathematics (BCAM)
- Aingeru Ramos – University of the Basque Country (EHU)
- Kauê Rodrigues Alves – Basque Center for Applied Mathematics (BCAM)
- Pranjal Tamuly – Basque Center for Applied Mathematics (BCAM)
- Leo Vattoly – Basque Center for Applied Mathematics (BCAM)

Acknowledgements

BCAM as a BERC and Severo Ochoa Excellence research centre, wishes to thank to the company speakers, the academic coordinators and the researchers of each working team for their invaluable contributions to the scientific success of the 188 European Study Group with Industry. We also want to express our gratitude to the Provincial Council of Bizkaia (DFB), B Accelerator Tower (BAT), and BCAM stakeholders; Basque Government, Ikerbasque, Innobasque, the University of the Basque Country (EHU), Petronor Innovación and Bilbao City Hall as well as the institutions that financially supported the event:

Provincial Council of Bizkaia (DFB).


bcam
basque center for applied mathematics


EXCELENCIA
SEVERO
OCHOA
Jul. 2014 - Dec. 2027


BAT
B Accelerator Tower


Bizkaia
foru aldundia
diputación foral


Lookiero


QUANTUM
MADS


DANOBATGROUP