



EXCELENCIA
MARÍA
DE MAEZTU



 IFISC*

 **Universitat**
de les Illes Balears

 **CSIC**

**PROCEEDINGS OF THE 2022
SURF@IFISC FELLOWSHIPS**

Proceedings of the 2022 SURF@IFISC Fellowships

The SURF (Summer Undergraduate Research Fellowships) Program is offered by IFISC since 2013 (only interrupted in 2020 due to the Covid-19 pandemic) with the goal of attracting excellent undergraduate students with majors compatible with IFISC research lines and offers Summer internships at IFISC, typically during the month of July, advised by members of IFISC Claustro.

SURF attracted for the Summer of 2022 a total of 54 candidates (38 male and 16 female), and 6 candidates were selected. The candidates came from 32 different universities belonging to 13 different countries, while the candidates were citizens of 21 different countries. The universities with more candidates were Manchester U. (8), Autonomous University of Barcelona, UAB (6) and Complutense University of Madrid (5), while exactly half of the candidates (27) were Spanish citizens and no other country was represented by more than 3 candidates.

These are the proceedings:

1. Sebastian Høltedahl Castedo and Silvia Ortín González, *Prediction and assessment of childhood obesity using machine learning algorithms*
2. Adrián Nadal Rosa (advisor Gonzalo Manzano), *Stochastic thermodynamics and information*
3. Alberto Gallego Pozo, Pere Colet, and Damià Gomila, *Power grid stability in scenarios of large renewable penetration*
4. Pablo Cabrales, Alex Giménez, and Manuel A. Matías, *Analysis of biodiversity in Posidonia meadows with satellite images through machine learning*
5. Antonio A. Carpes and David Sánchez, *An ordinal pattern approach to word correlations*
6. Carlota Prieto Jiménez and Llorenç Serra, *Interferometry with topological insulator nanowires*

Prevention and Detection of Childhood Obesity using Machine Learning

Sebastian Holtedahl Castedo, Silvia Ortín González
 Instituto de Física Interdisciplinar y Sistemas Complejos, IFISC (CSIC-UIB)
 Campus Universitat de les Illes Balears, E-07122 Palma de Mallorca, Spain

Abstract

Over the course of this fellowship, I investigated different machine learning methods and tools in order to detect childhood obesity. It has been suggested that gut bacteria or microbiota aids the detection of obesity and we used species data from a few hundred children in order to improve and better train machine learning models. We struggled with the size of our samples but still managed to develop a few techniques that produced an Area under Curve (AuC) of $72 \pm 6\%$. This was done by reducing the feature size of the microbiota down to 10 features using XGBoost and then using Catboost as the classifier. Microbiota was found to aid the performance of the models but not to a very significant extent.

1 Introduction

Obesity is a health condition caused by the excessive accumulation of body fat which results in negative effects on health including cardiovascular diseases and diabetes. Heart conditions form the largest cause of death in our society and they are mainly the result of obesity. At this moment in history, due to the steady accumulation of capital, we live not in an age of scarcity but of abundance. The power of the culture industry, the marketing of unhealthy goods and the relatively low price of sustenance have resulted in an epidemic of overconsumption and obesity. In Europe, approximately 59% of adults and a third of children are considered obese [1]. Solutions to tackle this epidemic are thus most effective when targeted at children.

Certain people are more at risk of developing obesity and identifying these people is a way of implementing a form of preventative care, warning them that they are predisposed and thus encouraging them to change their habits. It has recently been suggested that certain micro-organisms found in areas in the body including the gut called microbiota can help in the identifying process [2].

Microbiota are micro-organisms such as bacteria that are found in the digestive tracts of the human body. In the gut alone there are between 300-1000 different species [3]. It is the composition and amount of each species that can be used in order to determine the predisposition to obesity.

It is relatively difficult to gather microbiota samples as they cannot be cultured due to oxygen sensitivity and growth factors [4]. Instead, DNA based methods are used including pyro sequencing and MPSS. There are a few meta-genomic methods that can be used for the sequencing. The Shotgun method involves including all the data and Amplicon 16s sequencing uses a PCR fragment and thus does not include all the data [4]. Instead, it uses Operational Taxonomic Units as a proxy for the detection of the dominant bacteria and the approximate relative abundances. The benefit of 16s sampling is that it is much cheaper and easier to administer.

The microbiotic data that we record for each of our samples contains hundreds of features which are difficult to interpret. To be able to derive meaningful predictions of obesity prone people, machine learning techniques can be used. In addition to using a supervised classification model to determine which children are most likely to become obese, it will be imperative to use feature selection models to reduce the number of features used in the classification process. To sum up, the task of this project is to investigate a number of feature selection and classification algorithms to find an optimal way of detecting and thus preventing obesity.

In general there are two major categories of machine learning strategies in artificial intelligence

(AI): supervised and unsupervised learning. In this project, we deal with supervised learning techniques. This means that for each data point $x(i)$ there exists a label $y(i)$ and the objective is to predict these labels. If the labels are a continuous variable this is called a regression problem. If, however, the labels are discrete as is the case in our dataset, one deals with a classification problem. By contrast, in unsupervised learning, there are no labels and the objective is then to characterize the structure and probability density of the data points in their respective hyperspace [5].

2 Data and Algorithms

The first dataset that we used for our project was a sample of 70 children with 75 microbiotic features and 90 metadata features. The microbiotic data included the relative abundances of each species from 16S sequencing. The metadata features contained contextual information about the person including exercise routine and diet. Metadata and microbiota data were taken at the start of the study and after 2 years, the children were checked to see if they were obese or not. A second dataset consisted of a sample of 216 children, 63 of which already were obese at the beginning of the study, with 187 microbiotic features and 49 metadata features. In order to evaluate the effectiveness of the microbiota in improving the machine learning models, we tested out our algorithms on the data with and without the metadata.

Although these datasets have over 5000 points of data each, it is a very difficult data set to use machine learning algorithms for. There are several reasons for this:

The first is what is often called the problem of ' $p \gg n$ '. This is where there are more features (p) than there are samples (n). High dimensional data can cause overfitting very easily and it is more difficult to find a unique solution.

Next, the sparsity of the data leads to near zero variance. There is likely to be many zero variance predictors - those with only one unique value and many near variance predictors - those with some unique values but of low frequency. This results in the model becoming unstable and cross validation is likely to fail thus techniques to handle these values have to be implemented [5].

Additionally, the dependency structure of the data leads to co-linear predictors. Especially when handling microbiotic data there are many interlinked features which result in only models that can handle correlated predictors to be used.

On a smaller note, a few of the features are categorical which generally are harder to deal with and require specific techniques or models to handle such as one-hot encoding.

As a result of the above challenges, in our project, we used feature selection methods before using classification algorithms. This is in order to reduce the dimensionality of the data. In fact, due to the extremely 'wide' nature of our data set it is likely that feature selection will prove a more important role than classification in the prediction of obesity.

Unlike most machine learning algorithms, we cannot take an entirely black-box approach to our algorithms as there is a genuine need for interpretability due to the medicinal nature of the project. This rules out powerful deep neural network methods.

For the feature selection methods, we used SKB, IG, mRMR and XGBoost. We explored reducing the microbiota features to the top 10 and 20 features. For each feature selection algorithm, we performed hyperparameter tuning by the way of a grid search.

SKB stands for select k-best and is an algorithm that scores each feature using a specific function and then removes all but k highest scoring features. In our case, we used the chi-squared function [6].

IG stands for information gain and is similar to SKB in that it removes all but the k highest scoring features but unlike SKB it uses a different scoring algorithm. It uses the gain in information or the Kullback-Leibler divergence which is a measure of the amount of information learned about a process from observing a certain random variable. If one can learn a lot about a process from a certain feature, then it has a high information gain and is more useful [6].

mRMR stands for Maximum Relevance-Minimum Redundancy and is a feature selection algorithm developed by Uber [7] to find the minimal optimal subset of features. This attempts to find

a minimum subset of features that together have the maximum possible predictive power. For our case, this is useful as we want to reduce the dimensionality of our feature space while maintaining the predictive power of the data.

Finally, XGBoost is a gradient boosting algorithm, a form of ensemble learning that uses gradient-boosted decision trees [8]. It fits a sequence of weak classifiers to data and then adjusts the weights on the correctly and falsely predicted samples. This method allows you to obtain feature importance and partial dependence which is important for our task.

For the classification machine learning algorithms, we used Random Forest, Catboost and LGBM. The reason for the choice of these algorithms was that they are interpretable and effective at dealing with higher dimensional data samples. For each classification algorithm, we also performed hyperparameter tuning by the way of a grid search.

Random Forests are one of the more elementary classification algorithms [6]. It uses an ensemble of decision trees that predict the classification of a sample point, majority voting is then used in order to decide what the overall model predicts. The power of Random Forests comes from the predictive power of a large ensemble of uncorrelated predictors. For our case it should work relatively well with our data, however, some of our microbiotic features have correlations between them and this will diminish the effectiveness of the solution.

CatBoost is a gradient boosting algorithm, similar to XGboost but utilises categorical features directly, which is extremely useful for our case [9]. Other machine learning algorithms have to convert alphanumeric or categorical data into numerical features before running.

Lastly, LGBM is another gradient boosting algorithm developed by Microsoft [10]. The major difference compared to other gradient boosting algorithms is that it grows leaf-wise as opposed to level-wise which results in faster computation and 'lighter' computation. Although it is usually used to deal with very large datasets, it will still be useful as a comparison.

For this project, we predominantly used Jupyter Notebook and Python 3. The machine learning libraries used for this project included: Sci-kit Learn, CatBoost, SK-Feature and LightGBM.

In order to evaluate the effectiveness of our models we used the area under the curve measure (AuC) and performed a stratified 7-fold cross validation. Feature importance was later evaluated using SHAP (Shapley Additive explanations) values [11].

The full data pipeline went as follows:

- Pre-process and import our data
- Implement a train test split (90,10) using sk-learn
- Feature Selection using the above methods
- Classification using above models. Hyper parameter tuning also conducted
- Evaluation using AuC and SHAP values

3 Results and discussion

First, we will investigate the dataset with 70 children where we used four different feature selection algorithms (SKB, IG, mRMR, XGboost) to reduce the feature space down to 20 features and 3 different classification algorithms (Random forests, CatBoost, LGBM).

For Random forests we can see from figure 1. that all feature selection methods failed to achieve a useful score with XGBoost achieving the highest average AuC of 0.65. However, the range of values across the board was quite high at approximately 20%. The cause for this is most likely the small number of samples present in our dataset. The low score can also be attributed to the fact that these features were highly inter-dependant which causes Random Forests to lose their advantage as mentioned previously.

For CatBoost we see an increase in the predictive power of our models as seen in figure 2. Here we see that Information Gain was the most successful feature selection algorithm achieving an AuC of

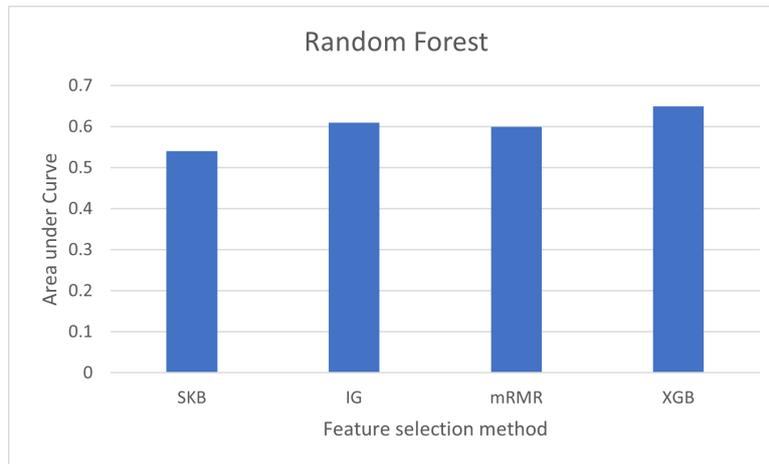


Figure 1: Random forest classification algorithm used with 4 different feature selection methods on the first dataset

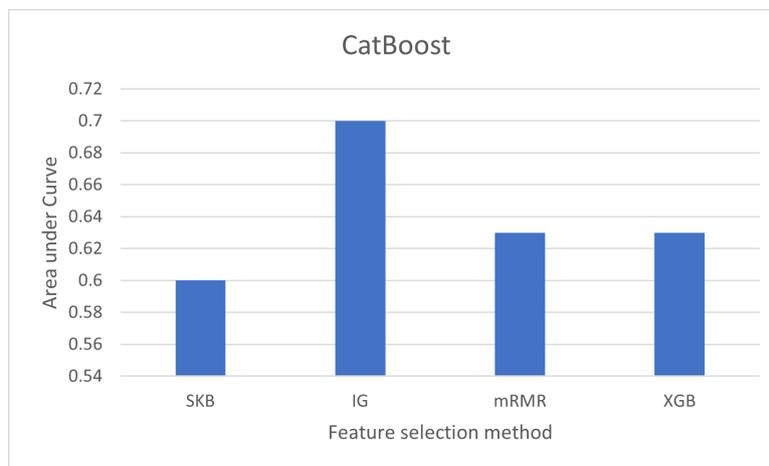


Figure 2: CatBoost classification algorithm used with 4 different feature selection methods on the first dataset

0.7. Again, the ranges on these values are quite high (around 15%) which does make these results less reliable.

Finally for LGBM (figure 3.) we see a similar score to that of Random Forests with a maximum score of 0.64 being achieved by mRMR. The reason for this low score is most likely due to the fact that it is a model designed to be used with very large datasets and thus is not very adapted to our use case. It had similar ranges to the Random Forest case.

For the second data set of 216 children we continued investigating 4 feature selection algorithms, this time varying the number of features reduced down to, but only used one classification algorithm - CatBoost. This was because in this second dataset there were many more categorical features which the other two models struggled to adapt to. We first repeated the training two times with two different random seeds in order to investigate the stability of our model. Next, we fixed the random seed and investigated the effect of only including the meta-data and the effect of removing any children who were noted obese at the start of the investigation.

Investigating the stability of the models (fig 4. and fig 5.) we found that for the most part, the variation between the different runs was less than the standard deviation. The only feature selection method that was unstable was mRMR whose difference in values was larger than 2 standard deviations. We found that using XGB with 10 microbiota features produced the best score of $72 \pm 6\%$ on average. With more time we would have run more than 2 random seeds.

Next, we investigated the effect of excluding the microbiota and of removing children who had

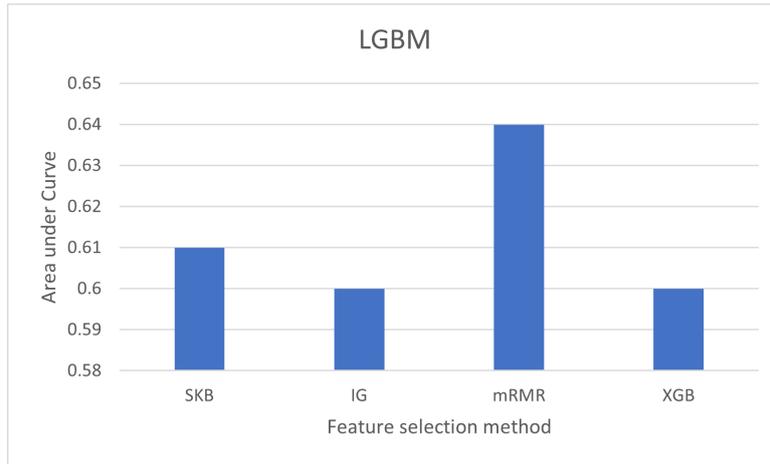


Figure 3: LGBM classification algorithm used with 4 different feature selection methods on the first dataset

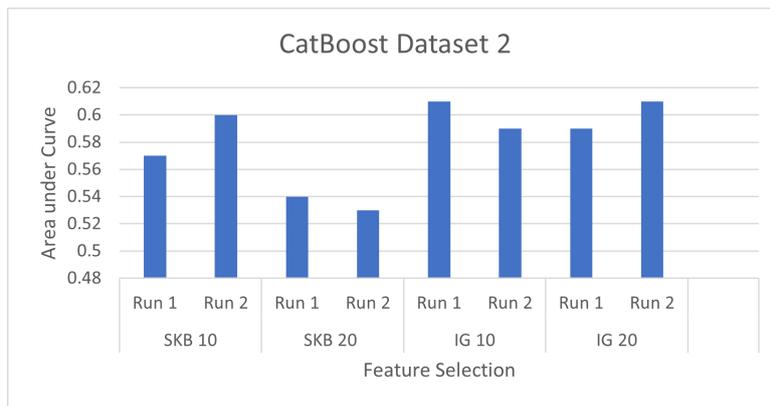


Figure 4: Catboost classification algorithm comparing SKB and IG feature selection which reduced the microbiota feature size to 10 and 20 respectively.

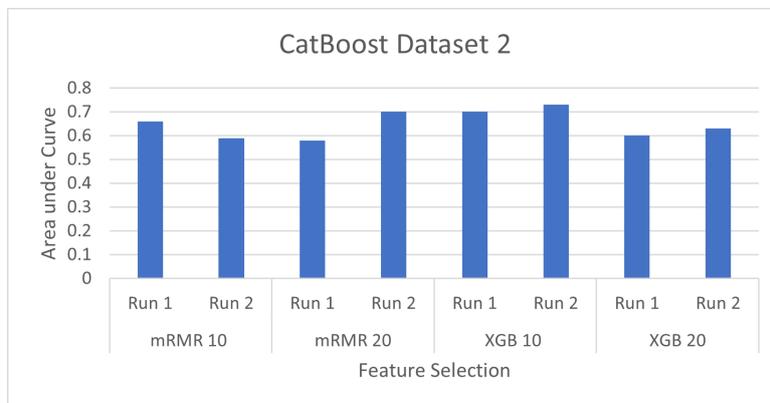


Figure 5: Catboost classification algorithm comparing mRMR and XGB feature selection which reduced the microbiota feature size to 10 and 20 respectively.

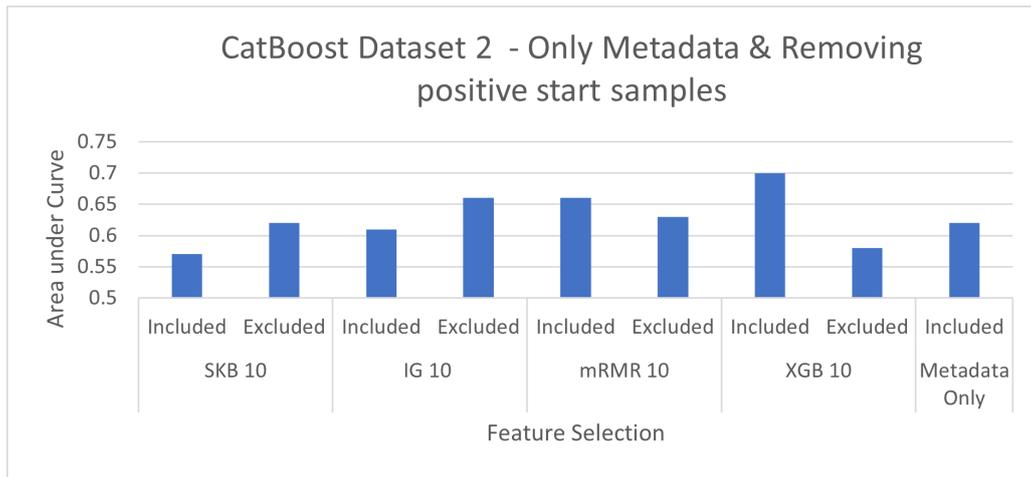


Figure 6: CatBoost classification algorithm comparing the effect of only using metadata features and removing samples that were noted to be obese at the start of the trial

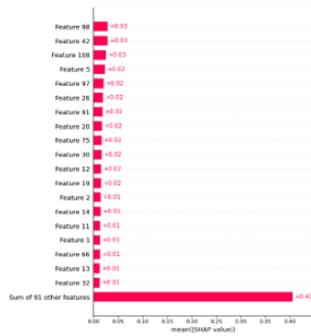


Figure 7: SHAP Values for SKB

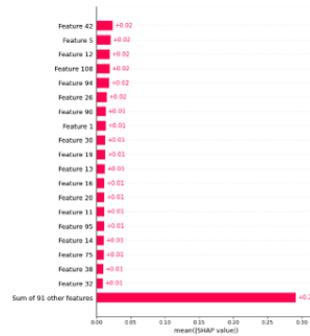


Figure 8: SHAP Values for IG

obesity at the start of the trial to see what produced the best results. From figure 6. we can see that removing the children who had obesity at the start of the trial had no significant improvement in the results. Furthermore, we can see that only including the metadata mostly scored worse results than those with feature selected microbiota features. From this, we can conclude that including the microbiota in helping to predict childhood obesity has a generally positive effect on the usefulness of our models, however, the benefits are not extreme and more investigations are needed in order to better understand the relationships present.

For the above feature selection algorithms, we used SHAP values in order to see if there was any correlation between the features that the algorithms selected. This is useful as it could tell us if certain species of microbiota can be used as markers for obesity. Figures (7-10) show the top 20 features found for the first data set using SKB, IG, mRMR and XGB. Although we could not find a substantial correlation between the different feature selection methods, feature numbers 42, 5, 26 and 12 appeared highly in all methods suggesting that they have a high predictive power when it comes to obesity. These corresponded to Anaerostipes, Coprococcus, Erysipelotrichaceae and Erysipelotrichaceae respectively.

4 Conclusions

In our investigations, we looked at trying to improve machine learning solutions to predict childhood obesity by using microbiota. Over the course of a few weeks, I looked at several methods currently employed and tested out several feature selection and classification algorithms.

We achieved the best prediction of $72 \pm 6\%$ AuC on the second data set with XGBoost reducing the feature size to 10 features and CatBoost as the classifier. The standard deviation of most of the results was around 6% and the range around 20%. Although we used a stratified 7-fold cross validation procedure the variation in the results was quite high which is most likely due to the very

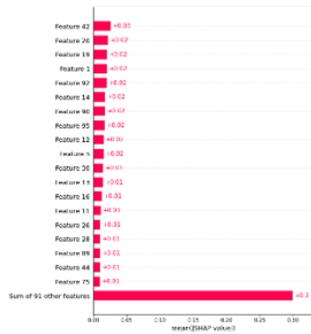


Figure 9: SHAP Values for mRMR

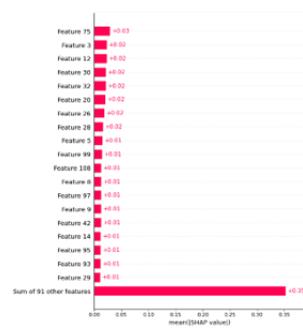


Figure 10: SHAP Values for XGB

small sample size. Including the microbiota data did increase the score of the models but not to a very large extent: some models seem to do better and worse with microbiota data. Finally, SHAP values were used to investigate any correlation between the most important features and found a few features that were present in all the feature selection methods.

The success of our investigation was hampered mainly by the lack of data. In most machine learning applications you have millions of samples to train your model with, having around 200 samples makes it very difficult to come up with any reproducible and stable model that can be used across the board. Hopefully in the future when more data is acquired further investigations can be made in order to find a stronger link between microbiota and obesity.

Acknowledgements

This work was supported by the SURF@IFISC fellowship. Francesc Serra Bisquerra aided with the pre-processing of the data and answering several questions.

References

- [1] World Health Organization. *New WHO report: Europe can reverse its obesity “epidemic”* (2022) <https://www.who.int/europe/news/item/03-05-2022-new-who-report--europe-can-reverse-its-obesity--epidemic>
- [2] Duncan, S. H., Lobley, G. E., Holtrop, G., Ince, J., Johnstone, A. M., Louis, P., & Flint, H. J. *Human colonic microbiota associated with diet, obesity and weight loss*. International journal of obesity, 32(11), 1720-1724 (2008)
- [3] Guarner, F; Malagelada, J *Gut flora in health and disease*. The Lancet. 361 (9356): 512–19. (2003)
- [4] Tang, Q., Jin, G., Wang, G., Liu, T., Liu, X., Wang, B., & Cao, H. *Current sampling methods for gut microbiota: a call for more precise devices*. Frontiers in cellular and infection microbiology, 151. (2020)
- [5] Mehta, P., Bukov, M., Wang, C. H., Day, A. G., Richardson, C., Fisher, C. K., & Schwab, D. *J.A high-bias, low-variance introduction to machine learning for physicists*. Physics reports, 810, 1-124. (2019)
- [6] <https://scikit-learn.org/> Accessed: July 2022
- [7] Zhao, Z., Anand, R., & Wang, M. *Maximum relevance and minimum redundancy feature selection methods for a marketing machine learning platform* In 2019 IEEE international conference on data science and advanced analytics (DSAA) (pp. 442-452). IEEE. (2019)
- [8] <https://xgboost.readthedocs.io/en/stable/> Accessed July 2022.
- [9] Yandex <https://catboost.ai/> Accessed July 2022
- [10] Microsoft <https://lightgbm.readthedocs.io/en/v3.3.2/> Accessed July 2022.
- [11] <https://shap.readthedocs.io/en/latest/index.html> Accessed July 2022

Stochastic thermodynamics and information

Student Adrián Nadal Rosa,

Supervisor Gonzalo Manzano,

Instituto de Física Interdisciplinar y Sistemas Complejos, IFISC (CSIC-UIB)

Campus Universitat de les Illes Balears, E-07122 Palma de Mallorca, Spain

Abstract

The main aim of this project was to study the basis of Stochastic Thermodynamics, able to describe small scale systems, unlike Classical Thermodynamics, linked only to the understanding of the macroscopic phenomena, usually at equilibrium or very close to it. In this comparison will be, hence, of crucial relevance to identify the equivalents of the principles of Thermodynamics into this small scale regime, where the fluctuations play a prominent role. The work was focused on testing the theory in few basic, but important examples with the aid of analytical calculations and numerical simulations. We simulated the stochastic dynamics of small systems in contact with one or several thermal reservoirs and checked the conservation of energy through the work and heat fluctuations. We also checked cornerstone nonequilibrium relations such as the Integral Fluctuation Theorem (FT) in several systems. Besides, a model of a thermal machine was examined, finding its different operating modes and under which conditions they appear. In this work we made use of simulations based on the Gillespie's algorithm and the first-reaction method.

1 Introduction

First of all, it should be depicted that this will be an introduction emphasised on concepts and results, rather than on mathematical derivations. For a further justification of the relations showed below one can read Reafs. [1] and [2].

1.1 Ensemble thermodynamics

We start by assuming a non-degenerate few-levels system in contact with only one heat reservoir, a thermal bath characterized by its temperature, T . Moreover, there will be an external driving, what is to say, an external agent that can control the energy of each level over time. We denote each state or level with m , and its associated energy with $\epsilon_m(\lambda)$, where λ is the representation of the external driving.

An ideal Markovian process is assumed, that is, the evolution of the system only depends on its current state and, hence has no memory, on the states visited before. In that conditions, the temporal evolution of the system can be determined by a Markovian master equation:

$$\dot{\mathbf{P}} = K\mathbf{P}, \quad \dot{P}_m = \sum_{m'} K_{m,m'} P_{m'}. \quad (1)$$

Here above P_m is the probability density function (PDF) of the state m and K the rate-matrix, with elements $K_{m,m'}$, containing the probability per unit of time that a transition from the state m' to m occurs. This rate-matrix must fulfill two basic conditions. First, during their evolution, the PDFs must remain as such and therefore conserve their normalization. This leads to the property showed below in two equivalent formulations:

$$K_{m,m} = - \sum_{m' \neq m} K_{m',m}, \quad \sum_m K_{m,m'} = 0. \quad (2)$$

Second, at equilibrium, a jump from state m to m' must be equally likely than a jump from m' to m . This is known as detailed balance¹, $K_{m,m'} P_m^{eq} = K_{m',m} P_m^{eq}$, that leads to the following

¹During this report the Boltzmann's constant is equal to 1, dimensionless.

relation between the elements of the rate-matrix², known as the Local Detailed Balance (LDB).

$$K_{m,m'} P_{m'}^{eq} = K_{m',m} P_m^{eq}, \quad \frac{K_{m',m}}{K_{m,m'}} = e^{-(\epsilon_{m'} - \epsilon_m)/T} \quad (3)$$

In this ensemble description, the energy and entropy of the system can be found by taking averages with respect to P_m . Asking for the temporal evolution of this quantities, one first obtains, for the energy, the expression:

$$E = \sum_m \epsilon_m P_m = \langle \epsilon_m \rangle, \quad \dot{E} = \sum_m \epsilon_m \dot{P}_m + \dot{\epsilon}_m P_m = \dot{Q} + \dot{W}. \quad (4)$$

Here above, heat is defined as the energy change due to the jumps between states and the work as the energy change due to variations of the energy of each state (due to the external driving).

For obtaining the entropy of the system, we use the surprise (from information theory) which we associate to every system microstate (i.e. to each level):

$$S = - \sum_m P_m \ln P_m = \langle - \ln P_m \rangle. \quad (5)$$

This corresponds to the Gibbs-Shannon entropy of the distribution P_m .

The temporal evolution of the entropy can be written in two terms: the entropy flow of the system, \dot{S}_e , directly related to the heat flow, and the entropy production, \dot{S}_{tot} :

$$\begin{aligned} \dot{S} &= \dot{S}_e + \dot{S}_{tot}, \\ \dot{S}_e &= \sum_{m,m'} K_{m,m'} P_{m'} \ln \frac{K_{m',m}}{K_{m,m'}} = \frac{1}{2} \sum_{m,m'} (K_{m,m'} P_{m'} - K_{m',m} P_m) \ln \frac{K_{m',m}}{K_{m,m'}} = \frac{\dot{Q}}{T}, \\ \dot{S}_{tot} &= \frac{1}{2} \sum_{m,m'} (K_{m,m'} P_{m'} - K_{m',m} P_m) \ln \frac{K_{m,m'} P_{m'}}{K_{m',m} P_m} \geq 0. \end{aligned} \quad (6)$$

1.1.1 Multiple reservoirs

In case that the transitions between states are coupled with different baths, characterized by their respective temperatures, T^ν , the system will not be able to reach, in general, an equilibrium state, but relaxes to a non-equilibrium steady state (NESS). This means that in the long-time limit, there will not be an equilibrium distribution for the system density function. Even though each bath can be at equilibrium with its corresponding transitions, the contact between all states turns out into a NESS. In this case, we just have to add to the above formulas a sum over all baths with which the system has transitions coupled.

We assume that the couplings with the baths do not interfere with each other. Hence, the rates of transition (i.e. the elements of the rate-matrix) will be a sum of the rates that there would be in case of having the system coupled just to one single bath. This means:

$$K_{m,m'} = \sum_\nu K_{m,m'}^\nu, \quad \dot{P}_m = \sum_\nu \dot{P}_m^\nu, \quad \dot{P}_m^\nu = \sum_{m'} K_{m,m'}^\nu P_{m'}. \quad (7)$$

The LDB is reached for a concrete temperature, as we saw in Eq. (3). Therefore, in a system in contact with several baths a general LDB cannot be found. Instead, one can find a LDB for each transition due to the possible coupling with each bath (i.e. with each temperature, T^ν). The resultant equations remain as follows:

$$\frac{K_{m',m}^\nu}{K_{m,m'}^\nu} = e^{-(\epsilon_{m'} - \epsilon_m)/T^\nu}. \quad (8)$$

Finally, following the same arguments used for the transition rates, the entropy and heat flows will be written as a sum of the flows generated by each bath:

$$\dot{Q} = \sum_\nu \dot{Q}^\nu, \quad \dot{S}_e = \sum_\nu \dot{S}_e^\nu, \quad \dot{S}_{tot} = \sum_\nu \dot{S}_{tot}^\nu. \quad (9)$$

²Note that the equilibrium distribution is given by the Boltzmann PDF, since we are working with the canonic ensemble.

1.2 Trajectory thermodynamics

Together with the ensemble description of the system in terms of probability densities, we can define a trajectory description, based on the dynamics that the system follows in every single experimental realization of the stochastic process.

The trajectory of the system is defined as the state in which the system is at each time, $X([0, t]) := \{x(0), x(\Delta t), \dots, x(t - \Delta t), x(t)\}$, following a concrete driving protocol, described by the values that a parameter λ , that controls the energy of the levels, takes along time, $\Lambda := \{\lambda(s) : 0 \leq s \leq t\}$. In the stochastic description, energy and entropy are described by the equations:

$$e = \epsilon_{m(t)}(t), \quad \dot{s} = -\ln P_{m(t)}(t). \quad (10)$$

The energy levels changes along time due to the external driving. Hence, the stochastic energy, e , is the energy of the state in which one finds the system, $\epsilon_{m(t)}$, at a concrete time $\epsilon_{m(t)}(t)$. The same goes for the stochastic entropy. Due to the external change of the energy levels, the associated PDFs changes in time, and therefore the surprise changes at each time.

To calculate the temporal evolution of these quantities will be easier making use of the equations:

$$\dot{f}_{m(t)}(t) = \sum_m \delta_{m,m(t)}^{Kr} \dot{f}_{m(t)}(t), \quad \dot{f}_{m(t)}(t) = \sum_m \dot{\delta}_{m,m(t)}^{Kr} f_{m(t)} + \delta_{m,m(t)}^{Kr} \dot{f}_{m(t)}. \quad (11)$$

Here above, $\dot{\delta}_{m,m'}^{Kr}$ is related to a jump from the state m' to the state m .

Using Eq. (11) one finds the description of stochastic work and heat along single trajectories. Again, the first is related to changes in the energy levels and the second is related to the jumps between states. Combining these two definitions, one finds the equivalent of the first law at the trajectory level, $\dot{e} = \dot{q} + \dot{w}$, with the heat and work flows described by the following expressions:

$$\dot{q} = \sum_m \epsilon_m(t) \dot{\delta}_{m,m(t)}^{Kr}, \quad \dot{w} = \sum_m \delta_{m,m(t)}^{Kr} \dot{\epsilon}_{m(t)}. \quad (12)$$

Focusing now on the entropy, one can find the same two contributions presented in the subsection of ensemble thermodynamics:

$$\dot{s} = \dot{s}_e + \dot{s}_{tot}, \quad \dot{s}_e = \frac{\dot{q}}{T}, \quad \dot{s}_{tot} = -\sum_m \left\{ \dot{\delta}_{m,m(t)}^{Kr} \ln \frac{P_m(t)}{P_m^{eq}(t)} + \delta_{m,m(t)}^{Kr} \frac{\dot{P}_m(t)}{P_m(t)} \right\}. \quad (13)$$

If averaging with respect to $K_{m,m'} P_{m'}$:

$$\begin{aligned} \dot{S}_{tot} &= \langle \dot{s}_{tot} \rangle = \sum_{m,m'} K_{m,m'} P_{m'} \ln \frac{P_{m'} P_m^{eq}}{P_{m'}^{eq} P_m} = \\ &= \frac{1}{2} \sum_{m,m'} (K_{m,m'} P_{m'} - K_{m',m} P_m) \ln \frac{K_{m,m'} P_{m'}}{K_{m',m} P_m} \geq 0. \end{aligned} \quad (14)$$

As said before, the surprise is used to calculate the entropy. One can find that the entropy variation at each time in a trajectory and the average over all the trajectories are given by the Shannon's entropy:

$$\begin{aligned} \Delta S(X([0, t])) &= -\ln P_t(x(t)) + \ln P_0(x(0)), \\ \langle \Delta S(X([0, t])) \rangle &= -\sum_{x(t)} P_t(x(t)) \ln P_t(x(t)) + \sum_{x(0)} P_0(x(0)) \ln P_0(x(0)). \end{aligned} \quad (15)$$

1.3 Path description

As explained before, the aim is finding the equivalents of the first and second principle of thermodynamics. As the reader can see, the above expressions do not lead to any direct physical conclusion. It will be more clarifying to see the cumulative entropy production, Δs_{tot} , along a trajectory, \mathbf{m} , with probability $\mathbf{P}_\Lambda(\mathbf{m})$, defined as:

$$\Delta s_{tot}(\mathbf{m}) := \ln \frac{\mathbf{P}_\Lambda(\mathbf{m})}{\bar{\mathbf{P}}_\Lambda(\mathbf{m})}, \quad \frac{P(\Delta s_{tot})}{P(-\Delta s_{tot})} = e^{\Delta s_{tot}}. \quad (16)$$

The quantities with the upper cross in Eq. (16) are referred to the backward process, this is, the inverted trajectory. Now there is just one more step left to get the equivalent of the second law, the integral FT.

$$\langle e^{-S_{tot}} \rangle = \sum_{X([0,t])} e^{-S_{tot}} P(X([0,t])) = \sum_{X([0,t])} \tilde{P}(\tilde{X}([0,t])) = 1. \quad (17)$$

One can deduce two corollaries from the integral FT. On the one hand, we have the first corollary, written as:

$$P(\Delta S_{tot} \leq \chi) \leq e^{-\chi}, \quad \chi \geq 0. \quad (18)$$

This means that even though the second law can be violated at the trajectory level, the likelihood of this event decays exponentially with respect to the deviation from zero. The second corollary is given by:

$$\langle S_{tot} \rangle \geq 0. \quad (19)$$

This latter one, is, in essence, the classical formulation of the second law. It can be understood as that if averaging with respect to all the trajectories (i.e., if calculating the macroscopic value of the total entropy production from the point of view of statistical mechanics), there must be a positive total entropy production.

A similar result can be obtained if the initial states of the forward and time-reversal process are assumed to be equilibrium distributions. This is called the Jarzinski's relation:

$$\langle e^{-w/T} \rangle = e^{-\Delta F^{eq}/T}. \quad (20)$$

The mathematical derivation of the Jarzinski's relations can be found in Reafs. [3].

As a final remark, the equivalent of the first law can be obtained directly, by just summing up all the contributions to the heat and work along the trajectory. Hence, the equivalents of the first and second principle at the trajectory level are presented below:

$$\begin{aligned} \Delta E(X([0,t])) &= Q(X[0,t]) + W(X[0,t]), \\ \langle e^{-S_{tot}} \rangle &= 1. \end{aligned} \quad (21)$$

2 Theoretical model

The project focused on three concrete systems: two levels coupled with one bath with a constant rate-matrix (constant energy levels), two levels coupled with one bath with a time dependent rate-matrix (time-dependent energy levels, this is, with an external driving) and three levels coupled with three different baths in the stationary state (a model of a thermal machine). A scheme of this systems is pictured in figure 1. The initial idea was to solve all of them analytically, and then simulate a large number of trajectories, in order to see if the quantities averaged respect to the trajectories recreate the obtained solutions from the master equation.

In the simulations based on the Gillespie's algorithm the elements of the rate matrix were written as follows:

$$K_{m'm} = \gamma e^{-(\epsilon_{m'} - \epsilon_m)/2T^\nu}. \quad (22)$$

The prefactor γ represents the number of transitions per unit time when the levels have equal energies and sets the time-scale of the problem. The factor 2 in the exponential is to guarantee that the LDB showed in Eq. (3) is fulfilled. The superscript ν in the temperature is just to signify that each pair of transitions (from m' to m and vice versa) can be coupled with a different bath.

2.1 Two levels with a constant rate-matrix

First the master equation has to be solved for this system. Using the property Eq. (2):

$$\begin{pmatrix} \dot{p}_0 \\ \dot{p}_1 \end{pmatrix} = \begin{pmatrix} K_{00} & K_{01} \\ K_{10} & K_{11} \end{pmatrix} \begin{pmatrix} p_0 \\ p_1 \end{pmatrix} = \begin{pmatrix} -K_{10} & K_{01} \\ K_{10} & -K_{01} \end{pmatrix} \begin{pmatrix} p_0 \\ p_1 \end{pmatrix}. \quad (23)$$

Where $p_0(t) + p_1(t) = 1$, since the sum of the probabilities of all states must be equal to one (the system must be in one state).

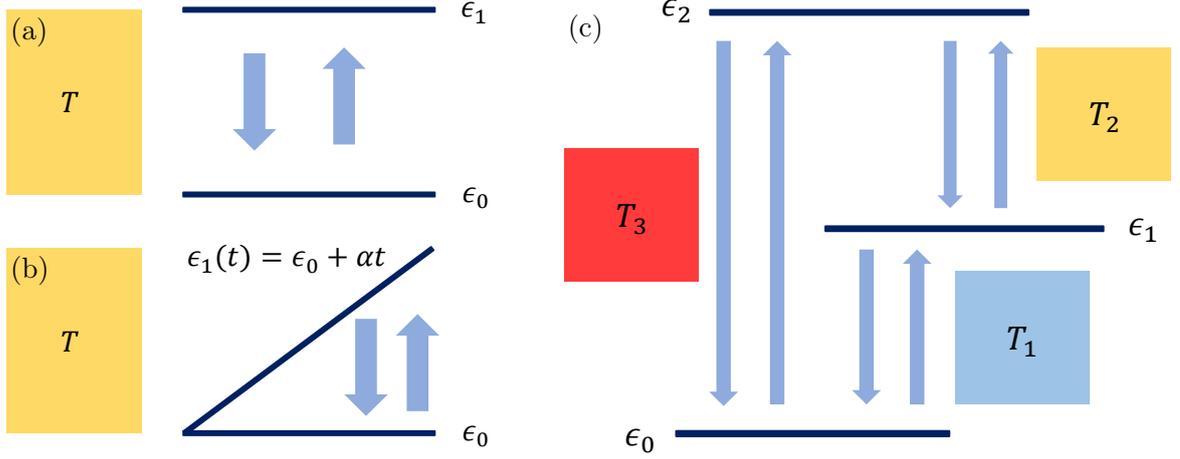


Figure 1: Schemes of the different studied systems. (a) Two levels with constant energies coupled with only one bath. (b) Two levels with time-dependent energies coupled with only one bath (external driving). (c) Three levels with constant energies coupled with three different baths (thermal machine).

The system of differential equations can be decoupled into two independent differential equations (ED):

$$\begin{cases} \dot{p}_0 + (K_{01} + K_{10})p_0 = K_{01} \\ \dot{p}_1 + (K_{01} + K_{10})p_1 = K_{10} \end{cases} \Rightarrow \begin{cases} \dot{p}_0 + 2\gamma \cosh\left(\frac{\epsilon_1 - \epsilon_0}{2T}\right) p_0 = \gamma e^{(\epsilon_1 - \epsilon_0)/2T} \\ \dot{p}_1 + 2\gamma \cosh\left(\frac{\epsilon_1 - \epsilon_0}{2T}\right) p_1 = \gamma e^{-(\epsilon_1 - \epsilon_0)/2T} \end{cases} \quad (24)$$

Notice that in this case, all the coefficients are constant. One can easily solve the equations with an homogeneous solution (making use of the characteristic equation) and a particular one, with a constant ansatz. The solutions are given by:

$$\begin{cases} p_0(t) = \left(p_{00} - \frac{K_{01}}{K_{01} + K_{10}}\right) e^{-(K_{01} + K_{10})t} + \frac{K_{01}}{K_{01} + K_{10}} \\ p_1(t) = \left(p_{10} - \frac{K_{10}}{K_{01} + K_{10}}\right) e^{-(K_{01} + K_{10})t} + \frac{K_{10}}{K_{01} + K_{10}} \end{cases} \quad (25)$$

This system was just a sort of "zero case", studied in order to understand the tools that were going to be used in the following cases. For that reason, only the approximation to the likelihood of occupation for the fundamental level was calculated. Two methods will be compared: the Gillespie's algorithm and the first-reaction.

As mentioned in the introduction, since the transitions are just coupled with one bath, in the long time limit the equilibrium distribution should be recreated, the Boltzmann's PDF. This can be easily seen in Eq. (25). The homogeneous solution decays exponentially with time. In the long time limit one gets, precisely, the Boltzmann's factor above the partition function, this is, the Boltzmann's PDF.

2.2 Two levels with external driving

The EDs to solve for this system are almost the same than those written in Eq. (24). The only difference is that now, $\epsilon_1 = \epsilon_1(t)$. The concrete new EDs are:

$$\begin{cases} \dot{p}_0 + 2\gamma \cosh\left(\frac{\alpha t}{2T}\right) p_0 = \gamma e^{\alpha t/2T} \\ \dot{p}_1 + 2\gamma \cosh\left(\frac{\alpha t}{2T}\right) p_1 = \gamma e^{-\alpha t/2T} \end{cases} \quad (26)$$

The analytical solution for the PDF of the fundamental state, that was gotten with the software *Mathematica*, is:

$$p_0(t) = e^{-\frac{4T\gamma}{\alpha} \sinh\left(\frac{\alpha t}{2T}\right)} \left\{ p_0(0) + \int_0^t e^{\frac{\alpha t}{2T}} e^{\frac{4T\gamma}{\alpha} \sinh\left(\frac{\alpha t}{2T}\right)} \gamma dt \right\}. \quad (27)$$

One can get the PDF for the upper level recalling $p_0(t) + p_1(t) = 1$. It is clear that the equation must be solved numerically for the simulations. The results showed in section 3 were obtained using the function `scipy.integrate.odeint()`.

The average respect to the simulations of thermodynamic quantities will be compared with the results predicted by the master equation. The expression for the average heat current in this case is:

$$\dot{Q}(t) = \sum_{m,m'} K_{m,m'} P_{m'} \ln \frac{K_{m',m}}{K_{m,m'}} = (\epsilon_1 - \epsilon_0)(K_{1,0}P_0 - K_{0,1}P_1) = \gamma\alpha t(e^{-\frac{\alpha t}{2T}}P_0(t) - e^{\frac{\alpha t}{2T}}P_1(t)). \quad (28)$$

First, for the heat flow, as shown in Eq. (6), the time-dependent rates of transition and the numerical solutions we got for the PDFs should be integrated. Nonetheless, this is not needed if ensuring an appropriate length for each step, Δt and making use of the `numpy.cumsum()` function with the elements, Q_i , written as:

$$Q(t) = \text{numpy.cumsum}(Q_i), \quad Q_i = \gamma\alpha t(e^{-\frac{\alpha t}{2T}}P_0(t) - e^{\frac{\alpha t}{2T}}P_1(t))(\Delta t). \quad (29)$$

To specify the ratio between the elements of the rate-matrix in Eq. (29) the LDB condition, Eq. (3), was used. In order to verify the FT, the total entropy production of each trajectory should be calculated using $s_{tot}(t) = -\ln P_{m(t)}(t) + \ln P_{m(0)}(0)$, which is the difference between the surprise assigned to the state in which the system is at a concrete time and the surprise assigned to the initial state of the trajectory.

Besides, one can compare whether the results are consistent. As explained several times before, one can identify two terms for the entropy of the system, one related to the interaction with the baths and another that will be eventually linked to the entropy production. Hence, S_{tot} can be calculated by subtracting S_e to the variation of entropy shown in Eq. (15). The final expression is given by:

$$S_{tot} = \langle \Delta S(X([0, t])) \rangle - \frac{Q}{T} = -\sum_{x(t)} P_t(x(t)) \ln P_t(x(t)) + \sum_{x(0)} P_0(x(0)) \ln P_0(x(0)) - \frac{Q}{T}. \quad (30)$$

2.3 Three levels in the stationary state

The trajectories in this model will be studied in the stationary state of the system, this is, in the limit in which the likelihood of occupation of each state remains constant at all times. That will make the system of equations easier to solve, since now it will be an homogeneous one, as presented below:

$$\dot{\mathbf{P}}^{st} = \mathbf{K}\mathbf{P}^{st} = 0 \Rightarrow \begin{pmatrix} -(K_{10} + K_{20}) & K_{01} & K_{02} \\ K_{10} & -(K_{01} + K_{21}) & K_{12} \\ K_{20} & K_{21} & -(K_{02} + K_{12}) \end{pmatrix} \begin{pmatrix} p_0 \\ p_1 \\ p_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \quad (31)$$

This system will be completed with the condition that the sum of the PDFs of all states must be equal to one. From now on, the PDFs in the stationary state will be denoted by $\mathbf{\Pi} := \mathbf{P}^{st}$, in order to simplify the notation. The concrete expression is not presented due to its length, but the solution for the system is given by:

$$\begin{cases} \Pi_1/\Pi_0 = (K_{02}K_{10} + K_{10}K_{12} + K_{12}K_{20})/(K_{01}K_{02} + K_{01}K_{12} + K_{02}K_{21}) \\ \Pi_2/\Pi_0 = (K_{01}K_{20} + K_{10}K_{21} + K_{21}K_{20})/(K_{01}K_{02} + K_{01}K_{12} + K_{02}K_{21}) \\ \Pi_0 + \Pi_1 + \Pi_2 = 1 \end{cases} \quad (32)$$

As explained in the introduction, this system, due to the coupling with different baths, will never reach an equilibrium state. In the long time limit a NESS will be observed. This is the studied model of a thermal machine, for which the functioning modes will be found, along with the conditions to observe them. But, first, the heat flow and the entropy production must be calculated. Again, the aim is finding the prediction of the master equation. Using Eq. (6), one can get, on the one hand, the following expression for the heat flows:

$$\begin{aligned} \dot{Q}_1 &= (\epsilon_1 - \epsilon_0)(K_{10}\Pi_0 - K_{01}\Pi_1) \\ \dot{Q}_2 &= (\epsilon_2 - \epsilon_1)(K_{21}\Pi_1 - K_{12}\Pi_2) \\ \dot{Q}_3 &= (\epsilon_2 - \epsilon_0)(K_{20}\Pi_0 - K_{02}\Pi_2) \end{aligned} \quad (33)$$

In these equations a positive contribution to the heat comes from a jump to an upper level and a negative contribution from a jump to a lower one. On the other hand, one can get for the total entropy flow the following expression:

$$\dot{S}_{tot} = - \sum_{x(t)} P_t(x(t)) \ln P_t(x(t)) + \sum_{x(0)} P_0(x(0)) \ln P_0(x(0)) - \sum_{\nu} \frac{Q_{\nu}}{T} = - \sum_{\nu} \frac{Q_{\nu}}{T}. \quad (34)$$

Since the system operates between three different thermal reservoirs, it can work as an autonomous absorption thermal machine. One can find two interesting operating modes: refrigerator and heat pump. The first one implies that heat is extracted from the bath at the lowest temperature, e.g. absorbing heat from a freezer chamber. In the latter, heat is instead pumped into the bath at the highest temperature, like a heater warming a house in the winter. The intermediate case represents a dissipator, in which energy is being dissipated due to a process of heat flowing from the bath with the highest temperature to the lowest ones. The conditions to have each operating mode are shown in the table (1).

| | $\langle \dot{Q}_1 \rangle$ | $\langle \dot{Q}_2 \rangle$ | $\langle \dot{Q}_3 \rangle$ |
|--------------|-----------------------------|-----------------------------|-----------------------------|
| Refrigerator | + | - | + |
| Heat Pump | - | + | - |

Table 1: Requirements that the heat flows of the baths must fulfill in order to get the different operating modes of the thermal machine.

In this simple system, once enforced the first condition, the others lead to the same inequality. Therefore, one can just focus on the sign of the heat flow of the first bath, obtaining the threshold given below:

$$(\epsilon_1 - \epsilon_0)_{thr} = \frac{(T_2 - T_3)T_1}{(T_2 - T_1)T_3} (\epsilon_2 - \epsilon_0). \quad (35)$$

The equality implies that all heats are settled to zero and hence the thermal machine is stopped. If the first energy gap, $\epsilon_1 - \epsilon_0$, is lower than the threshold value, one gets the refrigerator mode, otherwise, the heat pump one.

3 Results and discussion

In agreement with the initial aim of the project, the following results try to show the equivalency between the ensemble and the trajectory thermodynamic description. After simulating a large number of trajectories with the Gillespie's algorithm, and averaging respect to them, the results predicted by the master equation were recreated. The concrete expressions of heat and entropy for each system are presented where appropriate.

Once ensured that the values obtained for the different thermodynamic quantities are correct, the first and second principles, in their stochastic description, have been checked.

3.1 Two levels with a constant rate-matrix

As explained in the section 2, for this system the likelihood of occupation for the level with energy ϵ_0 , the fundamental level, was calculated. Both, the Gillespie's algorithm and the first-reaction method, recreate the Boltzmann's PDF in the long time limit, as showed in figure 2.

Even though both methods can lead to the correct solution for the PDFs, the Gillespie's algorithm makes the calculus of the thermodynamic quantities easier. For that reason, the rest of results will only be obtained with this algorithm.

The energy of the upper level is $\epsilon_1 = 1$. In this theoretical unit system, Boltzmann's constant equal to 1 dimensionless, energy and temperature have the same dimensions. Here ϵ_1 is defined as the basic unit of measurement of energy/temperature. The energy of the fundamental level is $\epsilon_0/\epsilon_1 = 0$ and the temperature of the bath $T/\epsilon_1 = 3$. Our measurement of time is defined by the frequency $\gamma = 0.1$.

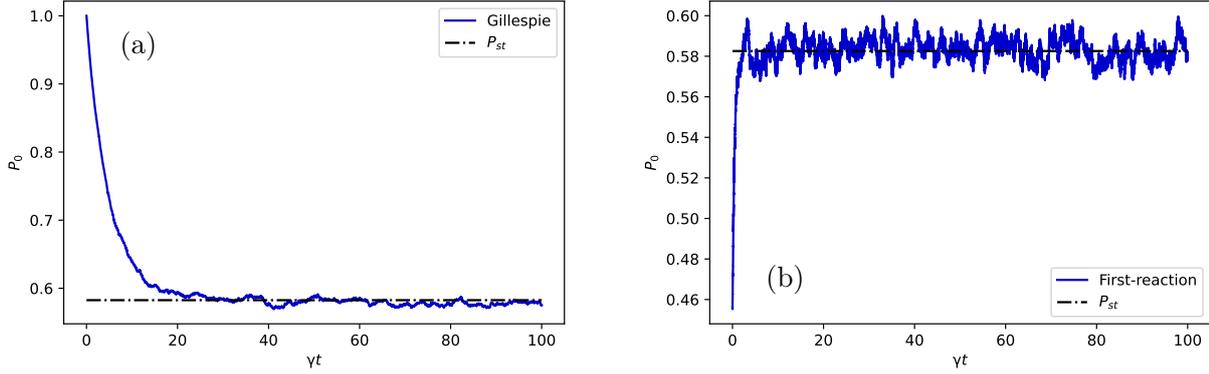


Figure 2: Approximation to de PDF of the fundamental level in the system with two levels with constant energies using: (a) Gillespie's algorithm with 10^4 simulations and $p_0(0) = 1$, $p_1(0) = 0$. (b) First-reaction method with a system of 10^4 elements.

3.2 Two levels with external driving

In the figure 3 the total entropy production and the heat are shown. The average respect to the simulations for the latter follows appropriately the prediction of the master equation. In the long time limit it seems that this is not the case for the entropy production. Nevertheless, the deviations are no larger than 1%, therefore the tendency is considered correct.

The first law is clearly fulfilled in figure 4. The energy of the system and the sum of work and heat are overlaped at each time. The FT shows, again, deviations no larger than 1%, so, it is assumed fulfilled.

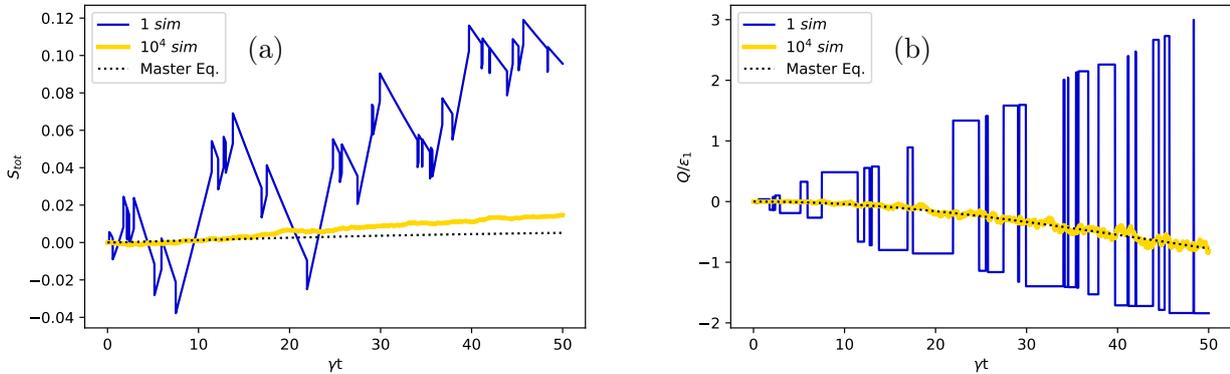


Figure 3: Total entropy change (a) and heat (b) for the system with two levels, coupled with one bath and with external driving.

3.3 Three levels in the stationary state

As for the system with two levels and external driving, in the figure 5 the equivalency between the master equation and the simulations is improved. The plots on the right side of the figure shows a visual representation of the dispersion due to the Gillespie's simulations.

In the figure 6 the FT is verified in two ways. On the one hand, verifying that the mean respect of the simulations recreates the expected value and, on the other, verifying the corollary presented in Eq. (18), with the exponential PDF as an upper bound for the likelihood of violating the classical second law.

Last but not least, in the figure 7 one can see the pair of values of the energy gaps for which the system is in a concrete operating mode. It can be understood as a phase diagram for the thermal

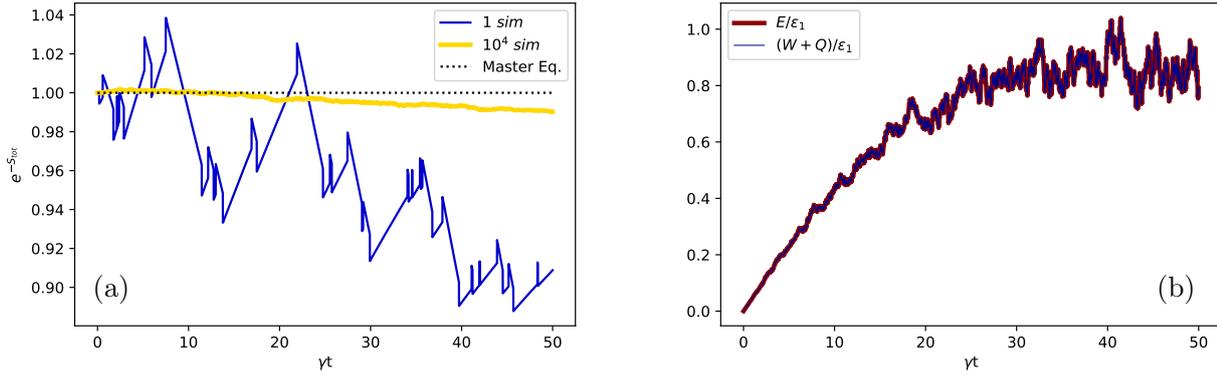


Figure 4: Verification of both, the Integral Fluctuation Theorem as the equivalent of the second law (a), and the first law in the stochastic description (b).

machine. In the z axes it is plotted \dot{Q}_1 . Negative values imply the heat pump mode (red), positive values imply refrigerator mode (blue).

In this case, the values of energy are: $\epsilon_1 = 10$, $\epsilon_0/\epsilon_1 = 0$ and $\epsilon_2/\epsilon_1 = 5$. And with settled temperatures: $T_1/\epsilon_1 = 1$, $T_2/\epsilon_1 = 7$ and $T_3/\epsilon_1 = 2$.

4 Conclusions

We have studied the main basic features of stochastic thermodynamics in different systems from both a theoretical perspective and from the point of view of their simulation in the computer. In particular we checked the first and second principles of thermodynamics at the stochastic and average levels of description, and tested the verification of the fluctuation theorem for entropy production, one of the cornerstone results in the field.

In this project, some of the prototypical systems have been considered, not only due to their low number of energy states, but also due to the number of interactions allowed. Nonetheless, one can easily add both more energetic levels or the chemical interaction to these systems by assuming that the heat reservoirs are also particle reservoirs, characterized by its temperature, T^ν , and chemical potential, μ^ν . Each state of the system will then have assigned a concrete number of particles in a concrete particle energy state. For instance, the number of proteins with which an enzyme is in contact in each step of the process it is in charge of. Each step would be a system state and being in contact with the enzyme or not a particle energy state.

Having had more time, concepts such as efficiency at maximum power and information to work conversion could have been studied, along with several systems like Brownian motors, biomolecules or quantum dots.

Acknowledgments

This work was supported by the SURF@IFISC fellowship, with great help and dedication of its supervisor, Gonzalo Manzano.

References

- [1] C. Van den Broeck, M. Esposito, *Ensemble and trajectory thermodynamics: A brief introduction* (2015).
- [2] Udo Seifert, *Stochastic thermodynamics: From principles to the cost of precision* (2018).
- [3] C. Jarzynski, *Nonequilibrium Equality for Free Energy Differences* (1997)

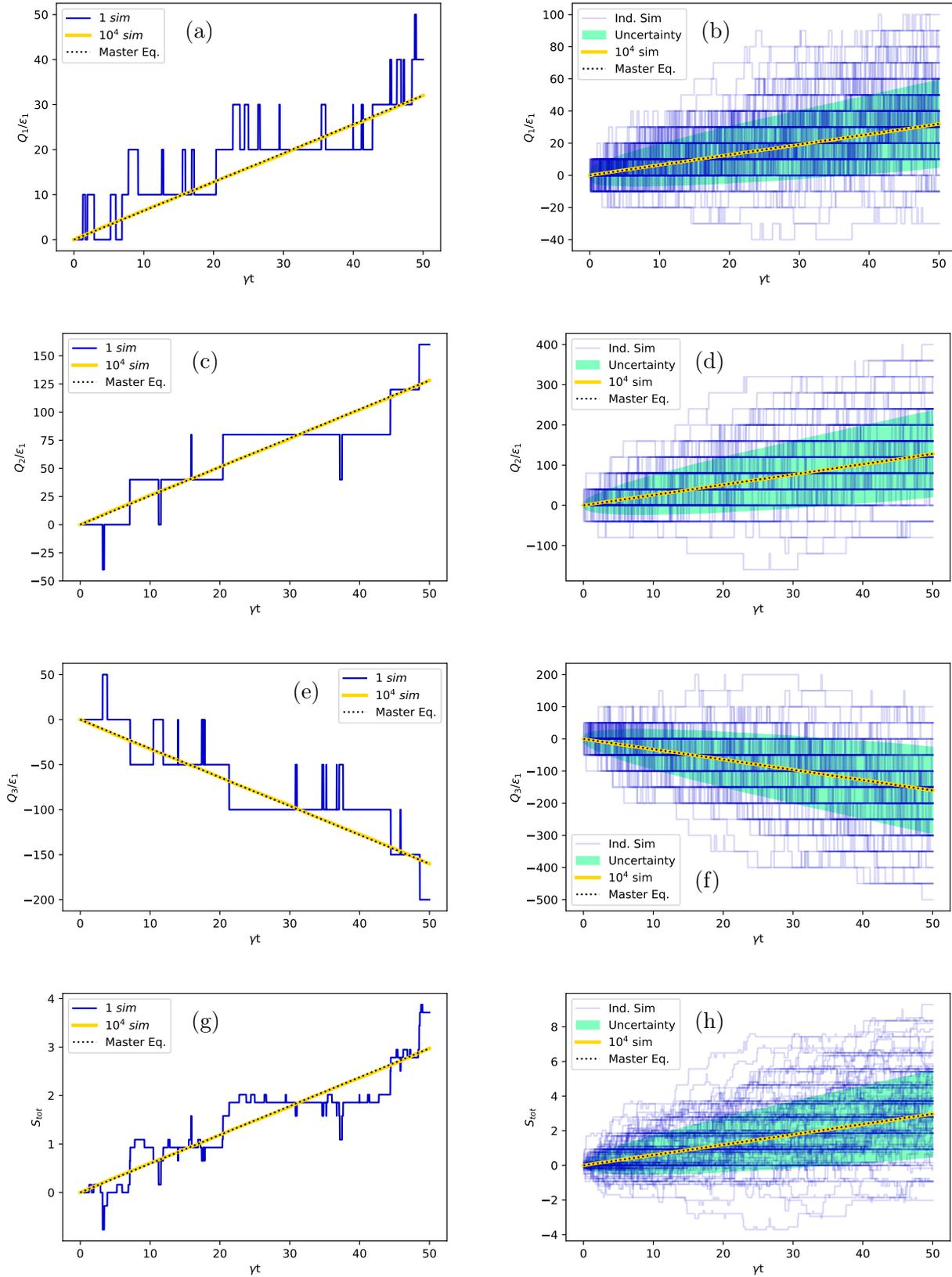


Figure 5: Heat exchanged with each bath and total entropy production for the system with three constant energy levels in the stationary state. On the right side of the figure, the the equivalency between the average respect to simulations and the master equation is being verified. On the left side there is a visual representation of the dispersion of the simulations generated. The "uncertainty" written in the legends refers to the standar deviation of the average values of the simulations.

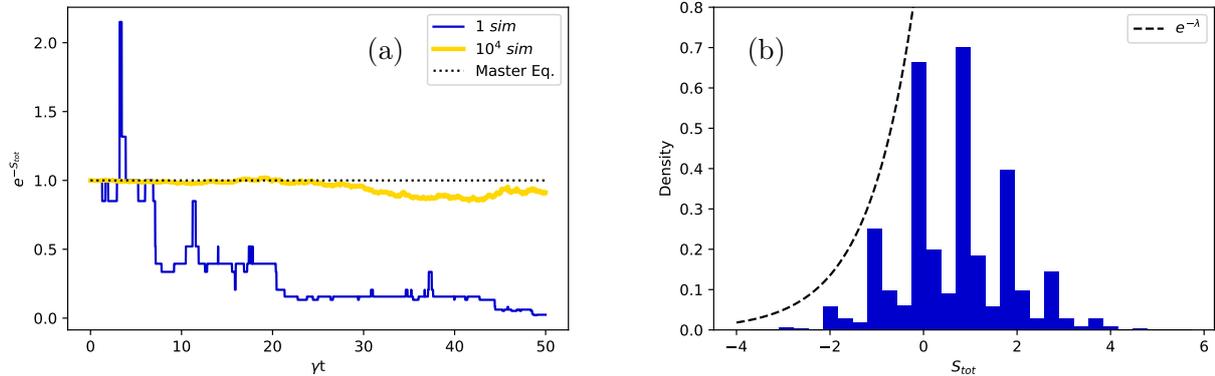


Figure 6: Verification of the Integral Fluctuation Theorem for the system with three constant energy levels. (a) The expected value of the FT compared with the approximation of the Gillespie's simulations. (b) Histogram of the values of entropy production and the exponential PDF that plays the role of an upper bound the deviations from the classical second law.

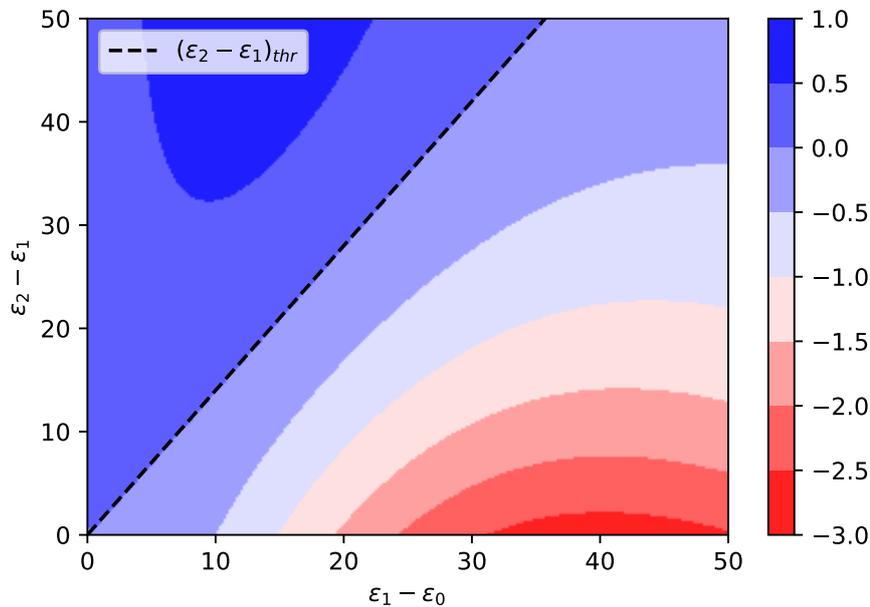


Figure 7: Operating modes for the thermal machine as a function of the energy gaps of the system. In the abscissa axis we have the basic unit of measurement of the energy. In the ordinate axis the second energy gap above the first. It is not written in the axes label for visual purposes.

Power grid stability in scenarios of large renewable penetration

Alberto Gallego Pozo, Pere Colet, Damià Gomila
 Instituto de Física Interdisciplinar y Sistemas Complejos, IFISC (CSIC-UIB)
 Campus Universitat de les Illes Balears, E-07122 Palma de Mallorca, Spain

Abstract

The electrical power grid is a complex interconnected network aimed to deliver electricity from producers to consumers. It is designed so that the power production and the power consumption are continuously balanced. Failure in a power transmission line can jeopardize this balance, possibly causing current to re-route itself to flow over transmission lines of insufficient capacity, which, in turn, results in further failures and extensive power outage. This report performs the stability analysis of the Balearic Islands' power grid, addressing the issue of cascading failures after a transmission line loss. In this scenario, the redundancy with which the Balearic Islands' power grid is built is verified, but without direct connection to mainland it is not possible to assure stability under failure of any individual transmission line.

1 Introduction

The power grid is the infrastructure that allows production and consumption of electricity. It consists of generation stations, electrical substations and transformers to step voltage up or down and electrical power transmission lines.

The generation stations convert some source of primary energy, such as nuclear, coal, thermal, hydroelectric, wind or solar energy, into electricity. The power grid is nearly always synchronous, operating with three phase alternating current (AC) frequencies synchronized. All the generators run at the same nominal frequency and they have to stay approximately in phase with each other and the grid. Production and consumption are balanced across the entire grid since, normally, energy is consumed as it is produced. Rotating generators have mechanisms to slightly adjust their rotational speed and hence, the generated power, to balance loading changes. When demand decreases, the grid is lightly loaded and the grid frequency runs above the nominal frequency, resulting in the generators reducing their output. Contrarily, when demand increases, the frequency slows, and generators increase their rotational speed to match the demand.

Electrical transmission lines allow the bulk movement of electricity from the generating site to electrical substations, which are connected to the distribution system and thus, to the end users. Power grid failures are normally associated with transmission lines failures. The loss of one region of the transmission network may cause current to re-route itself through lines of insufficient capacity, leading to failure in these lines, which will further exacerbate the problem. Ultimately, a cascading failure in the transmission lines can occur, leading to a power outage in a particular area. A power crisis can trigger enormous human and monetary losses, such as the one Texas suffered in 2021 [1] or the Northeast blackout of 2003 [2]. Thus, transmission networks have redundant pathways which allow line failures to occur while current is momentarily rerouted, until repairs are finished.

This report firstly presents a simple and general model for the dynamics of the high voltage power grid up to the transformer level. Then, it proposes a method to study the possibility of cascading failures in the transmission network after the loss of some transmission lines. This approach is then applied to analyze the stability of the Balearic Islands' power grid, which is used as a test case to examine the viability of the model to manage sudden finite perturbations as lines losses. The initial failures considered are losses of the main transmission lines of the islands' power grid.

2 Theoretical model

This section presents the model introduced in [3] for the dynamics of the high voltage power grid, based on the equivalent single line diagram of the actual power grid. Thus, an equivalent AC grid with a single phase is considered, instead of the real three phases of the current. The model consists of a set of differential equations that model the minimal elements of a power grid [4]: power plants, transmission lines and consumption nodes.

The power grid is modelled as a network of nodes, representing the production and consumption points, and links, through which current can be transferred from one node to another. Several transmission lines may contribute to each link, thus having the possibility to assess the failure of just one line in the case of double and triple lines. Transmission lines are undirected and power can flow in either sense. Nodes are described as non-linear oscillators, having a phase θ_i and a frequency ω_i , which are coupled via links and must be synchronized to the same reference frequency, which in the case of the Balearic Islands' power grid is $\omega_r = 50$ Hz. There are three types of nodes: conventional power plants, renewable energy sources and consumption points.

Nodes with conventional power plants are simply described by the swing equation [4]:

$$\frac{d\theta_i}{dt} = \omega_i \quad (1)$$

$$\frac{d\omega_i}{dt} = \frac{\omega_r^2}{2H_i P_i^G (\omega_i + \omega_r)} (P_i^m - P_i^e) \quad (2)$$

which models the power plant as a simple generator with a rotor attached on a turbine shaft rotating at the nominal frequency ω_r plus a small deviation ω_i . H_i is the inertia constant, P_i^G is the nominal capacity of the generator and P_i^m is the supplied mechanical power. P_i^e is the sum of the electric consumption at that node plus the net flow of power entering or leaving the node as a result of the power transmission from or to other nodes through each of the lines:

$$P_i^e = (1 + D_i \frac{\omega_i}{\omega_r}) P_i^l + \sum_j B_{ij} \sin(\theta_i - \theta_j) \quad (3)$$

The factor D_i is included to account for some fraction of the electric load being frequency-sensitive. B_{ij} is a parameter dependent on the line voltage V_{ij} and the impedance Z_{ij} :

$$B_{ij} = \frac{V_{ij}^2}{Z_{ij}} \quad (4)$$

In the development of the previous equations line losses have been neglected. Voltage fluctuations are also disregarded.

Conventional power plants include frequency control loops to guarantee operation close to the nominal grid frequency. Two different control mechanisms act in different time scales. Firstly, the load frequency control, which in the event that the mechanical supplied power P_m is lower (greater) than the electric demand P_e , rapidly stabilizes the frequency drop (increase) to a steady value by adjusting the mechanical power to the demand load.

$$\frac{dP_i^m}{dt} = \frac{1}{\tau_i} (P_i^s - P_i^m - \frac{P_i^c}{R_i} \frac{\omega_i}{\omega_r}) \quad (5)$$

where τ_i is the response time of the primary control, P_i^s is the spinning reserve power, P_i^c is the primary control capacity and R_i is the governor speed regulation parameter.

Later, the secondary control reduces the frequency deviation ω_i to zero by activating the spinning reserve power.

$$\frac{dP_i^s}{dt} = -k_i \frac{\omega_i}{\omega_r} - \lambda_i (P_i^s - P_i^{ref}) \quad (6)$$

where k_i is the gain of the secondary controller, λ_i is a parameter controlling the damping of the spinning reserve power, P_i^s is the spinning reserve power and P_i^{ref} is a set point value for the power plant.

Nodes without conventional power plants have low inertia and no mechanisms of frequency control. Their dynamics can be described by equations [1]-[3], imposing no supplied mechanical power $P_i^m = 0$ and very small effective inertia $H_i P_i^G$ as compared to conventional power plants.

Renewable energy sources are modelled as an additional supplied power on the particular nodes they are located.

The connection to mainland is not directly considered, but the input power this line would bring to the grid is considered by uniformly increasing the power produced by each power plant.

Each conventional power plant adds four equations ([1], [2], [5], [6]) to the global system of differential equations. Thus, if the number of nodes is n and the number of conventional power plants is n_p , a system of $2n + 2n_p$ differential equations has to be solved.

The failure in one transmission line can be included by setting the transmission parameter B_{ij} of line ij equal to zero at the breakdown instant.

The test case selected for the stability analysis is the Balearic Island's power grid, shown in the next Figures:



Figure 1: Mallorca's power grid. Green lines are 220 kV transmission lines, blue lines are 132 kV transmission lines and red lines are 66 kV transmission lines.

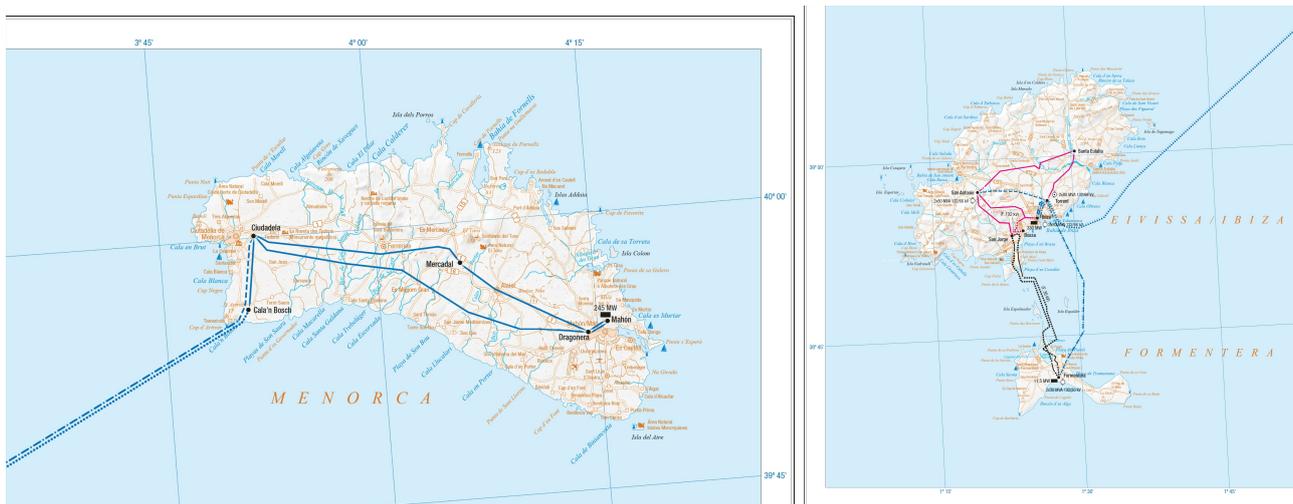


Figure 2: Menorca, Ibiza and Formentera's power grid.

The parameters appearing in equations [2]-[6] and the demand input data at each time of the day were provided by Pere Colet, Damià Gomila and María Marínez.

In order to assess the possibility of the cascading failure in the transmission lines and study its evolution, different initial failure conditions are considered. Firstly, each of the individual main lines, which correspond to those carrying a voltage of 220 kV, are initially set to be broken and the time evolution of the rest of the power grid is studied. In the case of double and triple lines, only failure in one line is imposed. Then, failures in more than one main transmission line are initially imposed. These two first cases could represent system failures caused by weather, transmission lines making contact with tree branches [2] or a malicious intrusion. The critical condition to determine the consequent failure of a transmission line is that its stress, defined as the ratio of the intensity carried out by the line to its nominal intensity, is larger than one:

$$stress_{ij} = \sigma_{ij} = \frac{B_{ij} \sin(\theta_i - \theta_j)}{B_{ijnominal}} > 1 \longrightarrow B_{ij} = 0 \quad (7)$$

A time interval of ten minutes is studied. In the first two minutes, the nominal system operation is considered. Then, the initial malfunction is imposed as a sudden perturbation and the system time evolution is analyzed. The time interval studied was selected to be 20:00-20:10 – 2020-07-14, since it corresponds to the night electricity peak and summer holidays, when demand is larger.

3 Results and discussion

This section presents the results for the three different failure cases studied, along with a discussion of the resultant power grid evolution.

3.1 Failure in one transmission line

Firstly, the initial failure of each of the high-voltage transmission lines is considered. One should note that double and triple lines count as two and three lines, respectively, and failure of only one of the lines is being considered in this section. It is observed that none of the failures, except for the one affecting the line connecting Valldurgent and Santa Ponsa, compromise the stability of the system.

| Initial failed line | Type | Voltage [220 kV] | Number of resultant failed lines |
|---------------------------|--------|------------------|----------------------------------|
| Son Reus-Llubi | Double | 220 | 0 |
| Son Reus-Son Orlandis | Double | 220 | 0 |
| Son Reus-Vallldurgent | Double | 220 | 0 |
| Cas Tresorer-Son Orlandis | Double | 220 | 0 |
| Son Orlandis-LLubi | Double | 220 | 0 |
| Llubi-Murterar | Double | 220 | 0 |
| Llubi-Es Bessons | Double | 220 | 0 |
| Vallldurgent-Santa Ponsa | Simple | 220 | 6 |

Table 1: Number of resultant failed lines after individual failure of each main transmission line

As an example of how the power grid reacts to a sudden downed line, the system evolution is studied for the case of failure of the line Son Reus-Llubi. The stress suffered by the neighbouring transmission lines, the frequency of the surrounding nodes and the neighbouring power plants frequency and generated power are monitored.

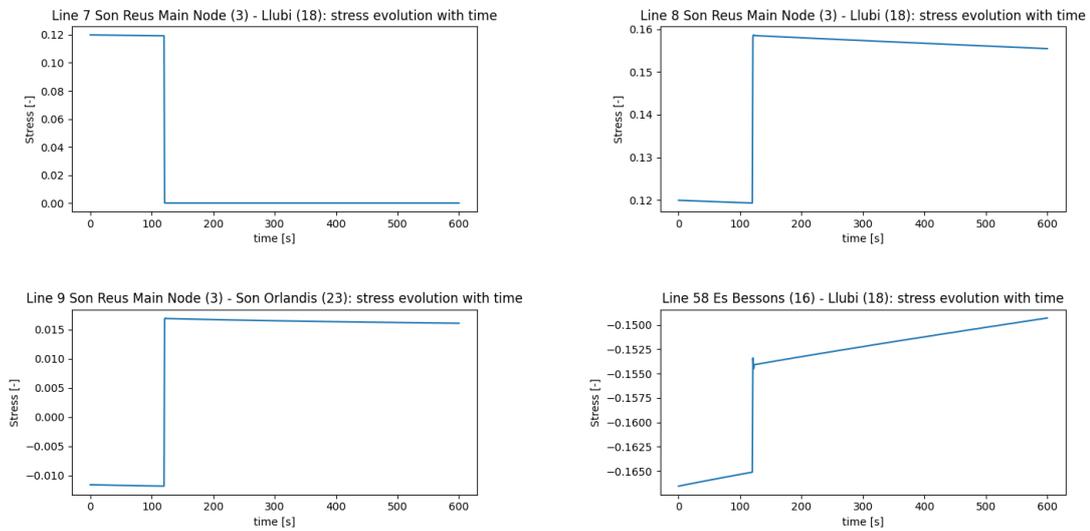


Figure 3: Son Reus-Llubi neighbouring lines stress evolution after initial failure of one of the lines connecting Son Reus and Llubi. The positive stress indicates that the current flows from the lower node number to the larger one.

After one Son Reus-Llubi line fails and its stress suddenly goes to zero, since Son Reus-Llubi is directly connected by a double line, the other line absorbs part of the stress. Part of the current flowing from Son Reus to Llubi is also diverted to the double neighbouring line Son Reus-Son Orlandis, as can be inferred from the increase in stress shown in Figure 3. In addition, the current flow leaving Llubi towards Es Bessons is suddenly decreased, due to the lower current inflow entering Llubi from Son Reus. Furthermore, the stress has been shown to practically immediately adapt to a new state after the sudden perturbation, since in just one integration time step (1 s), the stress is readjusted. The fact that the stress slightly varies during the time interval, specially in the case of line Es Bessons-Llubi, is due to the demand conditions varying every second.

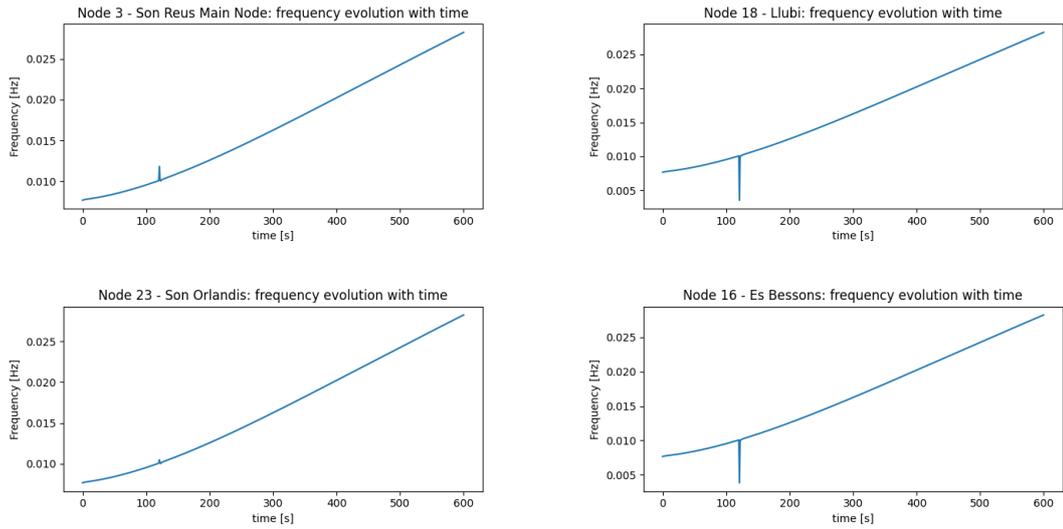


Figure 4: Son Reus-Llubi neighbouring nodes frequency evolution after initial failure of one of the lines connecting Son Reus and Llubi.

As for the nodes frequency, after the failure, a power imbalance in each of the neighbouring nodes is observed, resulting in a frequency variation, which is very quickly damped (in the order of the integration time step, 1 s). The system is still synchronized, with all nodes having almost the same frequency. Son Reus node experiences a first sudden frequency increase due to the net production and inflow of electrical power being instantaneously larger than the demand, as less current flows outwards just after failure of the line. Son Orlandis also experiences a slight positive frequency peak due to the larger inflow current diverted from Son Reus. On the other hand, Llubi and Es Bessons instantly receive less electrical power, leading to the demand being larger than the power inflow, which finally yields a sudden frequency decrease.

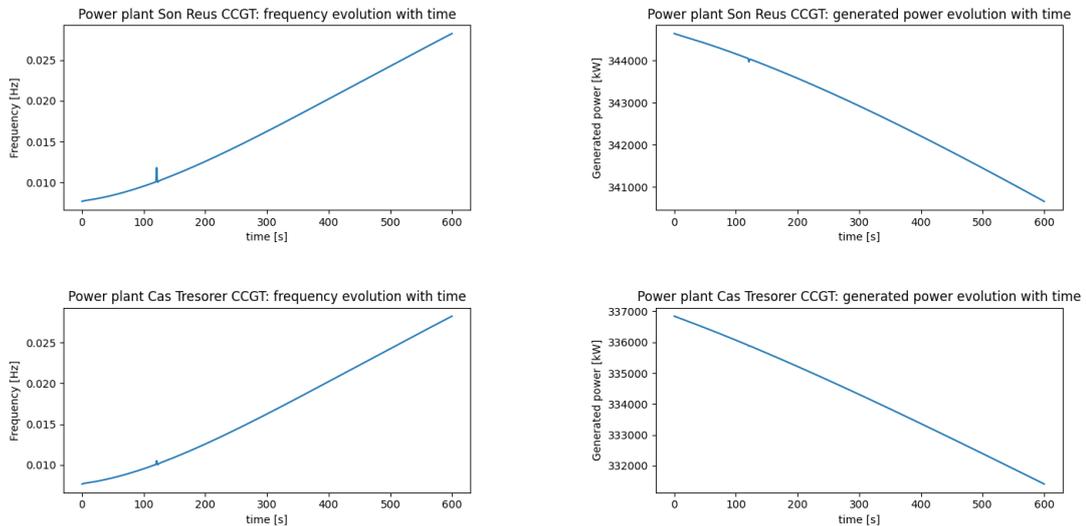


Figure 5: Son Reus-Llubi neighbouring power plants frequency and generated power evolution after initial failure of one of the lines connecting Son Reus and Llubi.

Son Reus power plant, just after the failure, needs to instantly decrease the mechanical power being generated, since it is not possible to divert all the previous current flowing through line Son Reus-Llubi to the neighbouring line in a very short time span (around 1 s). Conversely, Cas Tresorer power plant slightly increases the mechanical production to allow for a larger current flow from Son Orlandis to Llubi and hence, provide some power to reduce the negative power imbalance

in Llubi.

Once the overall behaviour of the system after a failure of a non-critical line is understood, interest is drawn to the power outage condition resulting from failure of the transmission line Valldurgent-Santa Ponsa. In a time span of three seconds a total of six lines go down (double lines count as two). The line connecting Valldurgent and Santa Ponsa is simple, and therefore, removing this simple high-capacity line, the current must be diverted to neighbouring lines of lower voltage and much lower capacity, which are unable to withstand the new stress (see Table 2).

| Failed line | Type | Time after initial failure [s] |
|------------------------|--------|--------------------------------|
| Valldurgent-San Agusti | Simple | 1 |
| Valldurgent-Andratx | Double | 2 |
| Santa Ponsa-Andratx | Double | 2 |
| Ibiza-Torrent | Simple | 3 |

Table 2: Transmission lines cascading failure after malfunction of Valldurgent-Santa Ponsa line.

As could be inferred from Figure 1 and being corroborated by Table 2 and Figures 6 and 7, after Valldurgent-Santa Ponsa failure, the electrical power required by Santa Ponsa will have to come from Andratx, which will then act as an intermediary between Valldurgent and Santa Ponsa. However, lines Valldurgent-Andratx and Santa Ponsa-Andratx have a lower capacity than Valldurgent-Santa Ponsa, and will not be capable of carrying the current without failure. Equivalently, the power required by Calviá, instead of directly flowing from Santa Ponsa, has to indirectly go from Valldurgent to San Agusti and from there to Calviá, leading to failure of line Valldurgent-San Agusti.

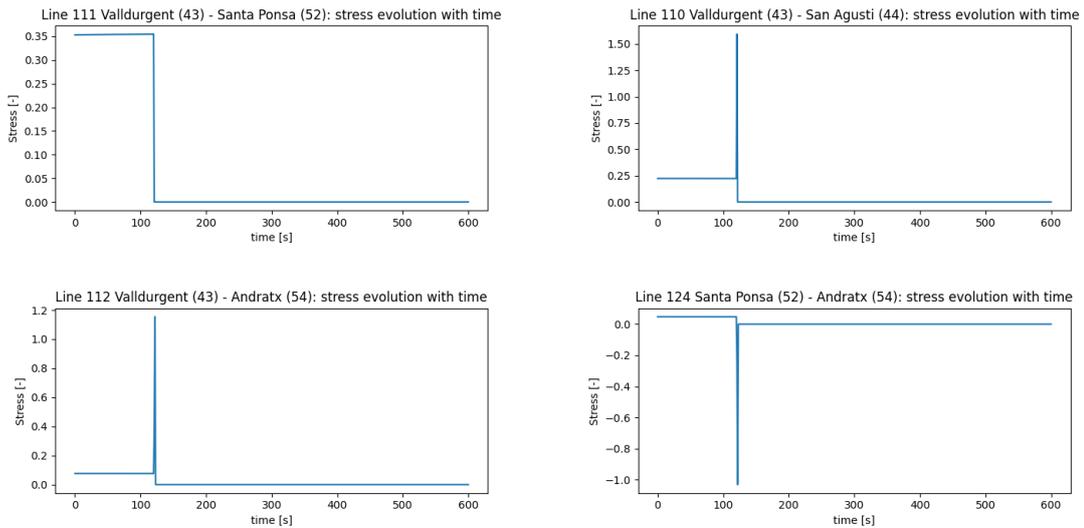


Figure 6: Valldurgent-Santa Ponsa neighbouring lines stress evolution after initial failure of line Valldurgent-Santa Ponsa. The positive stress indicates that the current flows from the lower node number to the larger one.

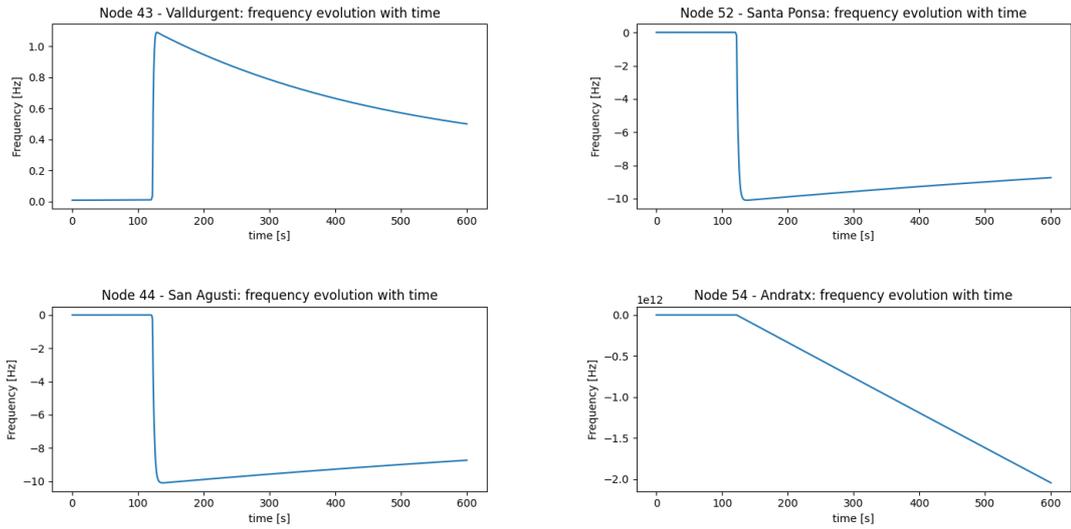


Figure 7: Vallldurgent-Santa Ponsa neighbouring nodes frequency evolution after initial failure of line Vallldurgent-Santa Ponsa.

Apart from the cascading failure, the nodes frequency evolution also needs to be monitored, since a large deviation from the nominal operating frequency (greater than around 2%), could compromise the correct operation of the power grid. Table 3 shows the frequency deviation of different nodes of the system after failure of line Vallldurgent-Santa Ponsa. Two different behaviours are observed. In the part corresponding to Ibiza, Formentera and the region of Mallorca which is disconnected from the rest of Mallorca, 3 seconds after the initial malfunction, several nodes reach frequency deviations around 3.3 Hz (6.6% of the nominal frequency), indicating a total power outage in these areas. On the other hand, the main bulk of Mallorca and Menorca, 6 seconds after the line failure, reach medium frequency deviations ($1.0505 \text{ Hz} = 2.1\%$ of the nominal frequency), which do not clearly indicate failure of the entire system. It may be possible to recover this part of the grid. It must be remarked that the negative frequencies found in the part corresponding to Ibiza, Formentera and a small part of Palma are the result of electrical demand being larger than the inflow power, since the main power plants, like Cas Tresorer or Son Reus, cannot send enough power through alternative lines and the power plants at Ibiza and Formentera cannot by themselves produce the necessary power. Contrarily, the positive frequencies found in the rest of Mallorca and Menorca come from the excess of power which still produce the main power plants in this area, due to the time delay in the decrease of production resulting from the plants inertia. This is the reason why this grid region may still correctly operate, since these power plants only have to reduce the produced power for the system to operate again in nominal conditions. It is also of interest to look at Andratx frequency, which blows up to a very large number, since this node is completely disconnected from the rest of the power grid.

| Node failed | time after initial failure [s] | Angular frequency after failure [rad/s] |
|----------------|--------------------------------|---|
| Ibiza | 3 | -3.2867 |
| Formentera | 3 | -3.2833 |
| San Agusti | 3 | -3.4951 |
| Calvia | 3 | -3.4957 |
| Santa Ponsa | 3 | -3.4960 |
| Palma Nova | 3 | -3.4957 |
| Andratx | 3 | -4.3076·10 ⁶ |
| Torrent | 3 | -3.2902 |
| Santa Eulalia | 3 | -3.2884 |
| San Antonio | 3 | -3.2892 |
| Bossa | 3 | -3.2867 |
| San Jorge | 3 | -3.2869 |
| EC Santa Ponsa | 3 | -3.4954 |
| Mahon | 6 | 1.0542 |
| Murterar | 6 | 1.0507 |
| Son Reus | 6 | 1.0505 |
| Cas Tresorer | 6 | 1.0505 |

Table 3: Nodes failure after malfunction in Valldurgent-Santa Ponsa transmission line. Data was shown until failure of the main power plants, when the power grid would collapse.

It is concluded that in order to assure the power grid stability after failure in one transmission line, the connection between Santa Ponsa and mainland needs to be considered, and that the Balearic Island’s power grid is not stable under one transmission line malfunction on its own.

3.2 Failure in two transmission lines

In this subsection, the failure of two simple lines or one complete double transmission line are considered.

| Initial failed line | Type | Voltage [220 kV] | Number of resultant failed lines |
|---------------------------|--------|------------------|----------------------------------|
| Son Reus-Llubi | Double | 220 | 0 |
| Son Reus-Son Orlandis | Double | 220 | 0 |
| Llubi-Son Orlandis | Double | 220 | 1 |
| Cas Tresorer-Son Orlandis | Double | 220 | 15 |
| Son Reus-Valldurgent | Double | 220 | 38 |
| Llubi-Murterar | Double | 220 | 0 |
| Llubi-Es Bessons | Double | 220 | 0 |
| Es-Bessons-Cala Mesquida | Triple | 132 | |

Table 4: Number of resultant failed lines after failure of two individual lines or one double line

Failure of Llubi-Son Orlandis only results in the consequent malfunction of one line, whereas Cas Tresorer-Son Orlandis and Son Reus-Valldurgent do lead to a cascading failure.

| Failed line | Type | Time after initial failure [s] |
|-------------------------------|--------|--------------------------------|
| Cas Tresorer-Son Molines | Double | 1 |
| Coliseo-Son Molines | Double | 1 |
| Cas Tresorer-San Juan de Dios | Double | 2 |
| Es Bessons-Llucmajor | Simple | 2 |
| Llucmajor-Son Orlandis | Simple | 2 |
| Llucmajor-Arenal | Simple | 2 |
| Arenal-San Juan de Dios | Simple | 2 |
| Arenal-Son Oms | Simple | 2 |
| San Juan de Dios-Son Oms | Simple | 2 |
| Son Molines-San Juan de Dios | Simple | 3 |
| Son Molines-Desaladora | Simple | 3 |
| San Juan de Dios-Aeropuerto | Simple | 3 |

Table 5: Transmission lines cascading failure after malfunction of Cas tresorer and Son Orlandis line.

As for the nodes frequency, it is observed that after cutting Llubi-Son Orlandis line, the overall grid remains within allowable limits (deviation frequency smaller than 0.1% of the nominal frequency) . On the other hand, removal of Cas Tresorer-Son Orlandis results in deviation frequencies larger than 2% of the nominal frequency for all nodes of the system, which will cause failure of the entire system. The reader may note that failure of this double line is equivalent to consider a massive failure of Cas Tresorer power plant, without which the system cannot correctly operate. Finally, the effect of removing of Son Reus-Vallldurgent double line is similar to the removal of Vallldurgent-Santa Ponsa, resulting in a power outage at Ibiza, Formentera and a small region in Mallorca.

4 Conclusions

A simple model to study the stability of a power grid with large renewable penetration under a sudden failure of a transmission line has been presented. The Balearic Islands' power grid has been used as a test case to show that the obtained results are coherent with reality. It has been shown that this model can deal with sudden perturbations without numerical errors building-up or spurious phenomena appearing. The redundancy with which the Balearic Island's power grid is designed was exposed by imposition of individual failures of each of the high-capacity lines. The only instability condition found after failure of one line was concluded to be due to the disregard of the connection between the islands' power grid and mainland, which is left for future work. Finally, malfunction of two high-capacity transmission lines was considered and it was found that this failure condition may result in a large power outage if it affects the lines connecting Mallorca and Ibiza or the double line which connects Cas Tresorer power plant to the power grid.

Overall, a simple and comprehensible model to help in the first design stages to comply with safety considerations and simulate transmission lines cascading failures has been presented. Special attention should be placed on the individual nodes frequency deviation with respect to the nominal one, as this can also be a critical condition.

Acknowledgments

I would like to thank Pere Colet and Damià Gomila. Without their support this work could not have been possible.

This work was supported by the SURF@IFISC fellowship.

References

- [1] Harvard Kennedy School. *Understanding Texas' energy grid failure*. 2021. URL: <https://www.hks.harvard.edu/faculty-research/policy-topics/environment-energy/understanding-texas-energy-grid-failure> (visited on 07/18/2022).

- [2] Wikipedia. The Free Encyclopedia. *Northeast blackout of 2003*. 2022. URL: https://en.wikipedia.org/wiki/Northeast_blackout_of_2003#Immediate_impact (visited on 07/18/2022).
- [3] Damia Gomila Pere Colet. “Modelling generation resources and consumption devices in VPP4Nodes”. In: (2022).
- [4] H. Saadat. *Power system analysis. The second edition*. McGraw Hill, 2003.

Analysis of biodiversity in Posidonia meadows with satellite images through machine learning

Pablo Cabrales, Alex Giménez, Manuel Matías

Instituto de Física Interdisciplinar y Sistemas Complejos, IFISC (CSIC-UIB)
 Campus Universitat de les Illes Balears, E-07122 Palma de Mallorca, Spain

Abstract

Accurately tracking the evolution of *Posidonia Oceanica* meadows is essential for the preservation of coastal Mediterranean ecosystems. In our study, we build and optimise an artificial neural network model (ANN) to identify seagrass species from satellite images of Mallorca’s Pollença bay. These images, which were taken by PlanetLabs and DigitalGlobe’s WorldView satellite systems, have a higher resolution than those used in similar studies and are available for the study of far larger areas. Accordingly, when we train and test our model on the same Planet image, and if we consider depths of up to 20m, we obtain accuracy scores around 92%. If we train and test our model on images from different days, the accuracy scores decrease but remain above 80%. We also apply a water column correction, investigate whether radiance ratios improve the accuracy of our model, and examine which frequency bands are the strongest predictors of the presence of Posidonia. Lastly, we compare our network with other ML models, such as SGD, Logistic Regression, LinearSVC, and Gradient Boosting. Although the accuracy scores are noticeably lower for the first three methods, Gradient boost performs almost as well as the ANN.

1 Introduction

Posidonia Oceanica is a species of seagrass which is only found in Mediterranean seabeds and which functions as a bioindicator. Moreover, Posidonia plays a fundamental role in carbon dioxide sequestration, oxygen production, harbouring of biodiversity, prevention of coastal erosion, or biomass production and transport, among other things. Posidonia meadows also form *climax communities*, which means that these ecosystems attain the highest possible level of ecological development and stability.

There are, however, several external threats facing these ecosystems. These include water contamination, which blocks light from reaching seabeds, fishing, boats anchoring, coastal constructions, eutrophication, and the presence of invasive algae species, such as the highly toxic, human-introduced, Caulerpa algae. Furthermore, because the rise of water temperatures leads to the acidification of sea waters and the proliferation of harmful algae species, climate change also places Posidonia in serious danger [1]. In this regard, scientists concur that temperatures above 28.5° could lead to massive Posidonia mortality events [2]. As an example, from August 7th to August 18th 2022, maximum water surface temperatures in Mallorca’s Pollensa bay have not dropped below the value of 28.5° [3].

Accordingly, being able to accurately track the evolution of Posidonia is of paramount importance. This concern, however, is far from a recent one, and several strategies have already been implemented. These strategies include the use of sound waves, underwater images, and the involvement of professional divers. These methods, however, are extremely time-consuming and expensive, and are not practical ways of tracking vast areas of Posidonia. In recent years, new approaches using satellite images and machine learning techniques have gained traction. Since these technologies are showing fast improvements, this could prove a much faster and more accurate way of detecting Posidonia.

In our study, we aim to build a neural network with which to identify Posidonia meadows and which we will train with satellite images of the coast of Mallorca. More specifically, we will use images from the bay of Pollença. Ideally, our model could help governments and coastal protection agencies to analyse the development of these ecosystems. We will also examine and test other machine learning methods. Our final goal is to obtain models which can consistently predict the presence of Posidonia on images taken on different days and with potentially different meteorological conditions.

2 Data

To train our model, we use data from the Atlas Posidonia Project [4], which uses side-scan sonar data and orthoimages to classify each region of the seabed as either Posidonia, sand, or other types of seagrass. This initiative was created to highlight areas where boat anchoring should not be allowed, as well as with general dissemination purposes. For the satellite images, we start by using a single image of the coast of Pollença from the highly potent DigitalGlobe’s WorldView-2 satellite [5]. This image has a multispectral spatial resolution of 1.8m GSD at nadir. This resolution yields around 17 million points on the whole Pollença bay (see Figure 1), roughly 40 % of which correspond to Posidonia meadows. The datasets are balanced prior to training, though.



Figure 1: Satellite image of the Pollença Bay.

The satellite data which we use for our model is the reflected radiance of light, comprising several frequencies. More specifically, the six frequency bands which we use are *Red Edge (RE)*, *Red(R)*, *Yellow(Y)*, *Green(G)*, *Blue(B)* and *Coastal Blue(CB)*. Although the satellite includes two additional infrared bands, we do not consider them in our analysis because of their insufficient penetrative power.

However, since the WorldView images are very costly, we also use images from Planet Labs’ flock of satellites [6]. These images have a lower multispectral spatial resolution of 3m GSD at nadir which is still above the resolution of images employed in similar studies and allows us to use an additional radiance band (*Green 2*). This resolution yields nearly 5 million points on the Pollença bay. The Planet image also includes two additional infrared bands, which again we will not consider.

These radiance bands are used as features for our models, together with an extra bathymetry (or depth) layer. The adjustments necessary for the spatial correspondence between the satellite, depth, and sonar data was done by IFISC’s Alex Giménez.

Additionally, we split our data into two groups for analysis. In the first of the two, we consider depths of up to 20m, in line with prior works on the subject [7]), allowing for an easier result comparison. In the second, we consider all depths in the studied region, reaching up to 34m, where the classification task will be more challenging.

3 Method

3.1 Neural network optimisation

As previously mentioned, we will use artificial neural networks (ANNs, see Figure 2) for the classification task. The structure of a neural network consists of multiple layers, each of them with a set of neurons that activate or fire with a given output. This output is the result of applying an activation function (see Figure 3) to a linear combination of the outputs from the previous layer plus a bias value. The network supplies the input layer, which is the first layer, with the variables or dataset features. In our case, these variables are the radiance bands of the satellite image and the depth variable. In the middle, hidden layers, several linear transformations are applied to the

data. The output layer, or final layer, produces a single value. In our case, if this value is greater than 0.5, the data point is predicted to be Posidonia, whereas if the value is lower than 0.5, the data point is predicted to be something else. These values are compared to the actual classification data values, and the neural network adjusts the neuron weights and biases accordingly.

The first stage of our study involves optimising the network hyperparameters. Hyperparameters are parameters that control the learning process but that are not derived via training. These hyperparameters include the number of layers, the number of neurons in each layer, the type of activation function in each layer, the learning rate, or the penalties given to features making the network inefficient. To find the optimal combination of these hyperparameters, we find the accuracy score of the network for each of these combinations. In our project, we use *Python's Optuna* library for this task.

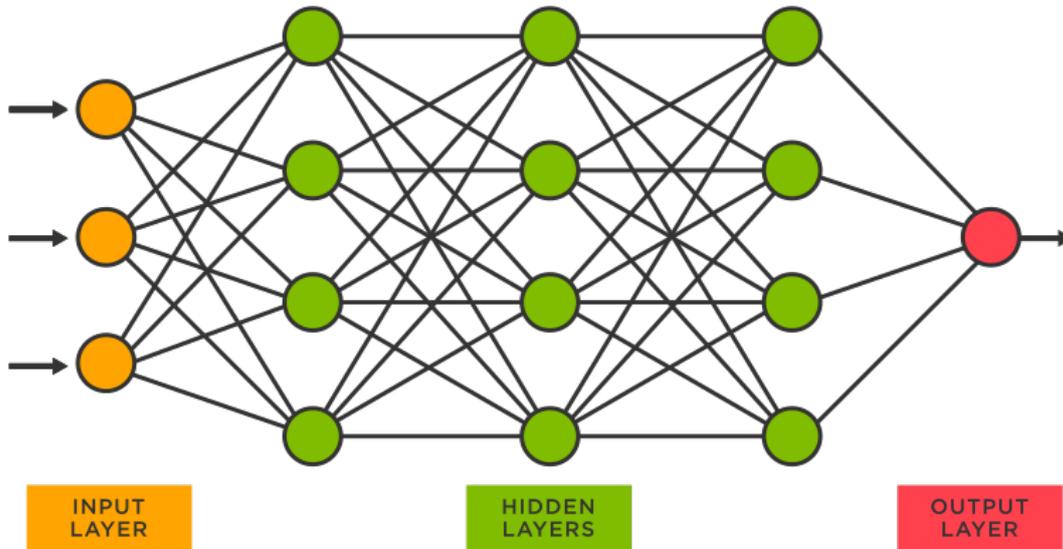


Figure 2: Basic structure of a sequential neural network.

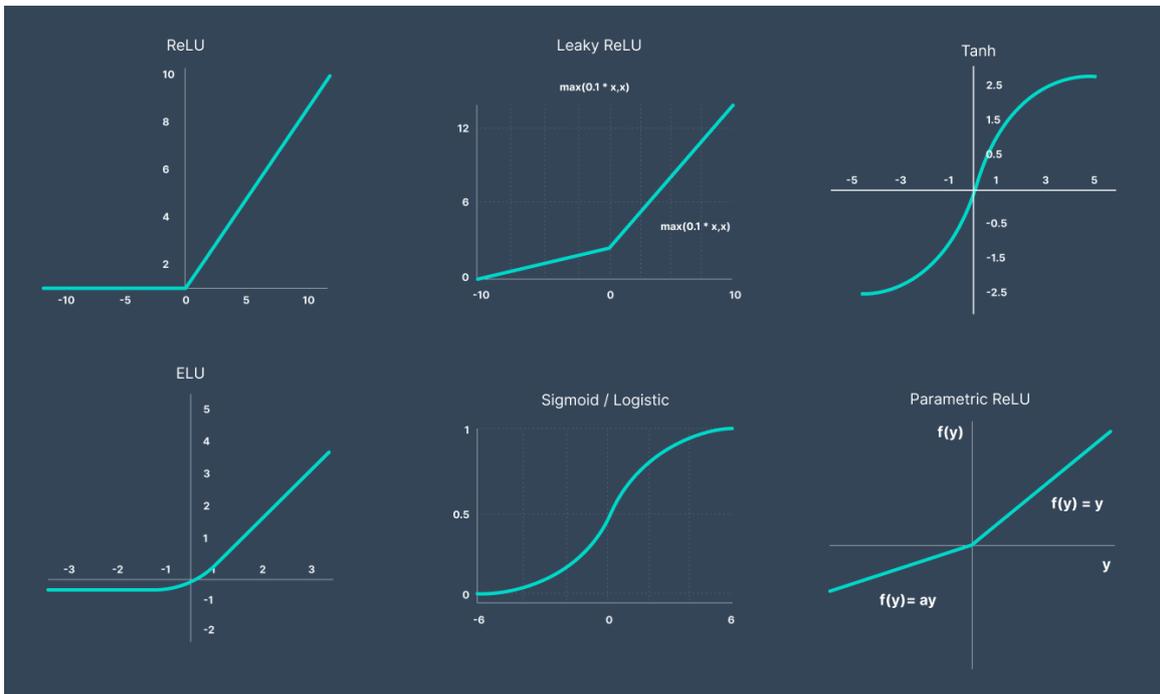


Figure 3: Six widely used activation functions.

3.2 Cross-validation accuracy

After finding our optimal network, we aim to obtain an accuracy score which makes the best possible use of the whole data set. To do this, we employ the *k-fold cross-validation* technique. In this way, we split the dataset into k folds, each of them with a different train-validation-test split (see Figure 4). Our ANN uses the training part of the dataset to generate different possible classification algorithms, while the validation set selects the algorithm which performs best and avoids overfitting to the training set. Finally, the test data is used to find the performance of the model on new, unseen data. In our case, as is standard practice, we split the dataset into $k = 5$ folds.

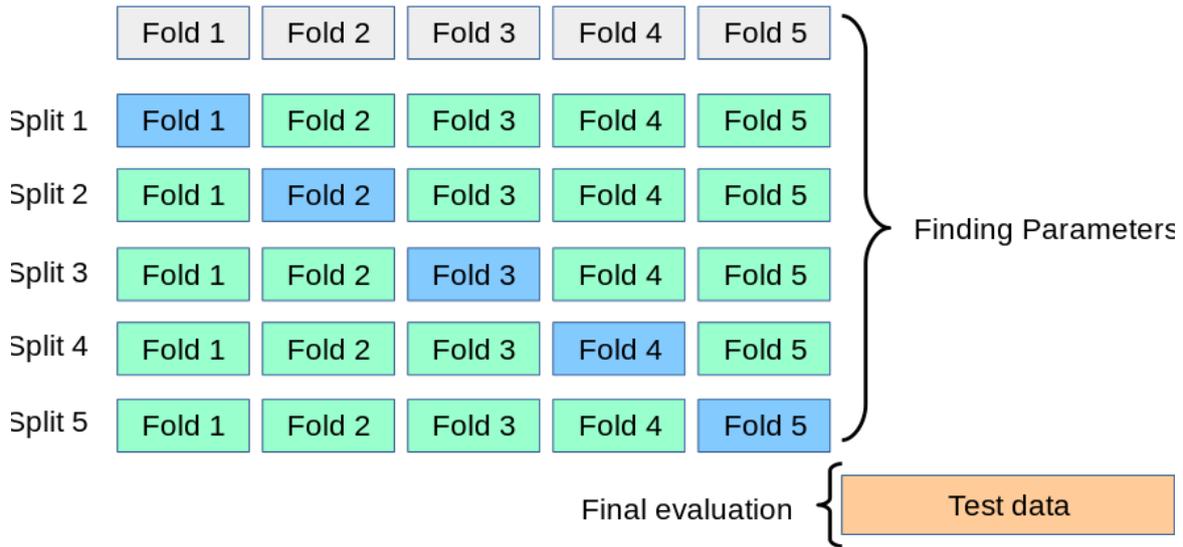


Figure 4: Schematic representation of a k -fold split. The training data is shown in green, the validation data is shown in blue, and the test data is shown in orange.

3.3 Feature analysis

Next, we wanted to understand how each model feature helped our network complete the classification task. To do so, we used SHAP (SHapley Additive exPlanations) values. The SHAP value for each feature is obtained by randomly rearranging its values within the dataset and measuring how much the network’s output changes as a result. In essence, we lose the insight which that feature provides for our network: the more the output changes, the more relevant the feature is and the higher the absolute SHAP value is. The calculations are readily performed via *Python*’s *shap* library.

Here, we expect the relevance to be determined, at least in part, by the “absorption” of different light frequencies in marine waters, as shown in Figure 5. The factors causing this “absorption” include light attenuation, scattering, refraction, and turbidity, among others [8].

3.4 Testing other machine learning methods

We expected the ANN to be the most effective model, but we also tested other machine learning models to see if they were any better at detecting the nuances of the data. These models are Gradient boosting, which combines several forest classifiers as weak learners; SGDCClassifier, which implements stochastic gradient descent; Logistic regressions, which fit a logistic function to the data to produce an output; and LinearSVC, or linear support-vector classification, which finds the best hyperplane to separate and classify the data points. We will rank their accuracy scores and compare them to our ANN.

3.5 Band ratios as features

Another relevant consideration is that two substrates might appear to be at different depths when they are not, since substrates affect how light is reflected. For instance, *Posidonia Oceanica* meadows appear deeper than sand at the same true depth, which could mislead our ANN. To prevent

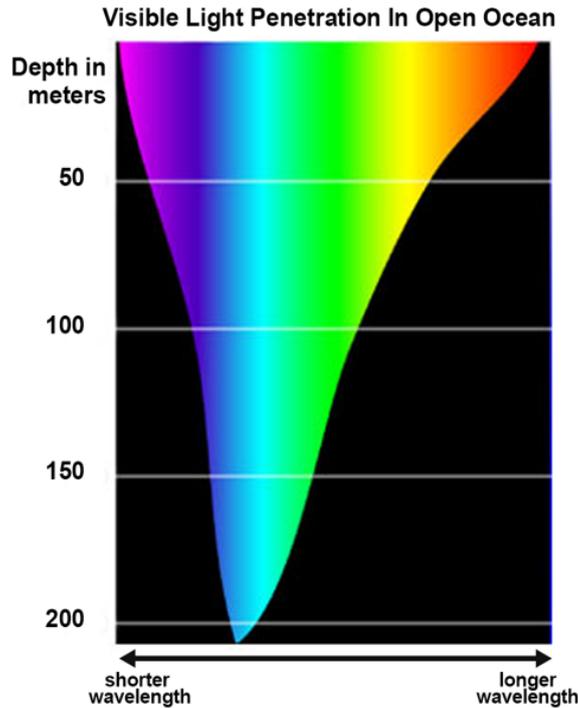


Figure 5: Spectrum of solar light penetration in marine waters.

this, researchers proposed implementing band ratios and normalised bands, based on the assumption that this ratio remains constant regardless of water depth.[9]

Accordingly, and similarly to what was done by Borfecchia et al., 2013 [10], we study how implementing the following normalised bands affects accuracy:

$$G/B, G/R, R/B, ND_1, ND_2 \quad (1)$$

Where

$$ND_1 = \frac{G - B}{G + B} \quad (2)$$

$$ND_2 = \frac{G - R}{G + R} \quad (3)$$

3.6 Water correction

Furthermore, the water column attenuates light and lowers signal-to-noise ratios, which affects reported reflectance, as stated in prior publications [11]. To get around this issue, we employ *Maritorena et al.'s* [4] model for regions with unchanging substrates and water quality, but with variable depth.

$$R_s = R_\infty + (R_B - R_\infty)e^{-2K_d Z} \quad (4)$$

Where R_s is the measured radiance, which is equal to the radiance at the water's surface after applying atmospheric corrections -already applied in the images we use-, R_∞ is the radiance from an infinitely deep water column, R_B is the radiance at the seabed, which we are ultimately trying to find, K_d is the attenuation coefficient and Z is the depth.

The depth/bathymetry data were readily available for our use. Knowing R_s , which are the measured values, and R_∞ , which are the radiance values at the deepest point in our image (34m for WorldView, but deeper for Planet), we can rewrite equation 4 and write equation 5. From this equation, we can perform a linear fit to find the values for K_d from the slope. Once we have obtained the values for the attenuation coefficients K_d , we can determine the corrected radiance bands R_B .

$$\log(R_s - R_\infty) = \log(R_B - R_\infty) - 2K_d Z \quad (5)$$

3.7 Comparing model performance in different days

The last part of the project aimed to test the validity of our trained model on images taken on different days under different atmospheric conditions. To do this, we use four images from four

different days: 3rd of April 2021 (“Day zero”), 7th of June 2021 (“Day one”, Figure 6a), 17th of April 2022 (“Day two”) and 26th of August 2020 (“Day three”, Figure 6b).

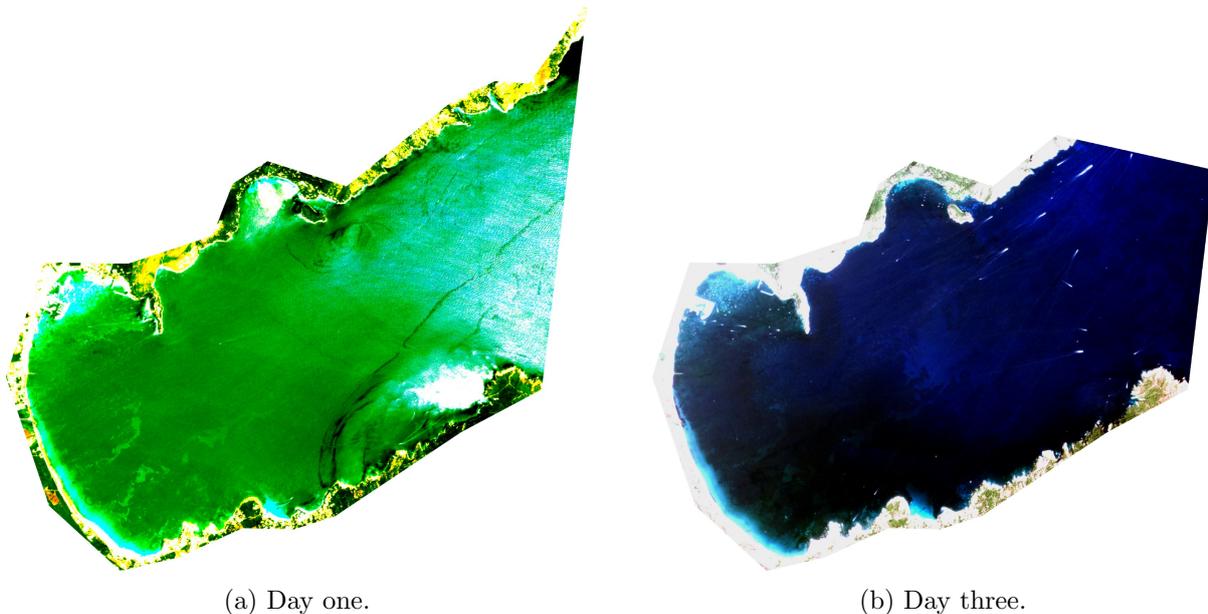


Figure 6: Comparison of two satellite images from different days.

In these images, we can directly see how both seabeds are not equally clear -the second day’s image is clearer- and how several boats are interfering with the day three image. Note that the difference in colour schemes between both images is only due to how they were plotted, not to the actual atmospheric conditions.

Now, we want to determine the accuracy of our model when it is trained on a given day and tested on a different one, shuffling through all possible combinations of these four days. Additionally, to analyse why these accuracy scores might differ, we look at how the radiance distributions differ for different images. We do this through several methods, such as the two-sample Kolmogorov-Smirnov (dis)similarity test [12], which measures the maximum difference between two cumulative distribution functions; principal component analysis, or PCA, which involves changing the basis in our feature hyperspace to find the direction that maximises the available information; and the cosine similarity test, which examines how the angles of these principal component vectors differ from each other.

4 Results and discussion

4.1 Neural network optimisation

For our initial testing, we built a simple, three-layer sequential neural network, as shown in Figure 7. The question mark symbol in the input and output cells represents the batch size, which is constantly adjusted by the network itself. The batch size is the number of samples processed before the model is updated, while the number of epochs is the number of full iterations across the dataset. Under the “input:” cell, the integer on the right indicates the number of input features, which is equal to the number of neurons in the preceding layer. Under the “output:” cell, the integer now references the neurons in the layer itself. In the top cell of each layer, the activation function and the type of layer are shown. Note that we do not consider the first layer as an actual neural network layer, as its sole function is to integrate the input data into the network and send it to the subsequent neurons.

After optimising our network, we obtain the model shown in Figure 8. This model has one fewer layer and more neurons in the single hidden layer. The activation functions also change, along with other parameters not included in the figure i.e. an L2 penalty or an optimised learning rate of 0.005.

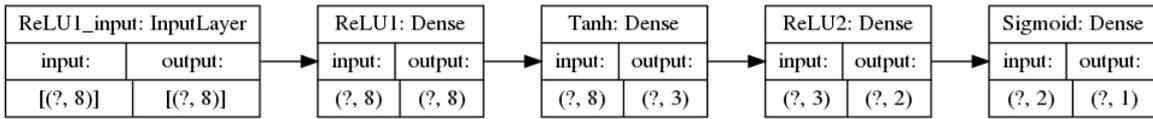


Figure 7: Original neural network. The question mark symbol refers to the batch size, which the ANN adjusts in every iteration.

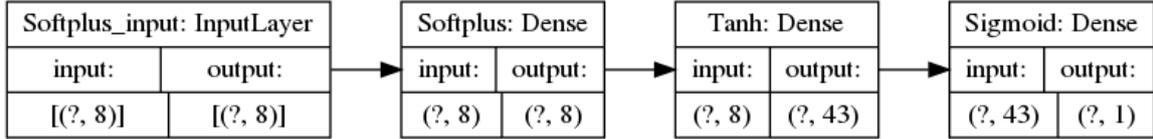


Figure 8: Optimised neural network. The question mark symbol refers to the batch size, which the ANN adjusts in every iteration.

4.2 Cross-validation accuracy

First, we look at the validation and training accuracy scores to see if our model is overfitting to the training data. As shown in Figure 9, the validation accuracy rises with the number of epochs, but remains slightly below the training accuracy, as expected.

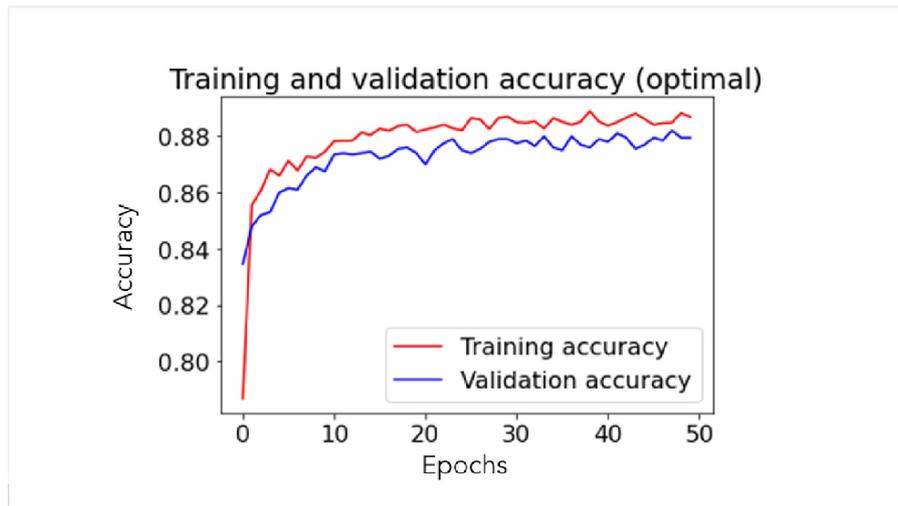


Figure 9: Learning curves for the optimal neural network.

We now examine whether, and if so, by how much, our optimised model outperforms the original one. From a 5-fold cross-validation test and starting with the WorldView image, we find the following accuracy values. For waters shallower than 20m, the accuracy of the original neural network is $91.5 \pm 0.1\%$, while the accuracy of the optimal neural network is $91.8 \pm 0.1\%$. Meanwhile, for all possible water depths, the accuracy of the original neural network is $88.9 \pm 0.1\%$, compared to $89.4 \pm 0.1\%$ for the optimal neural network.

If we now move to the Planet image, we obtain the following accuracy scores. For waters shallower than 20m, the accuracy of the original neural network is $91.9 \pm 0.1\%$, while the accuracy for the optimal neural network is $92.1 \pm 0.1\%$. Meanwhile, for all possible water depths, the accuracy of the original neural network is $87.0 \pm 0.1\%$, while the accuracy for the optimal neural network is $87.7 \pm 0.1\%$.

Comparing the results from both images, Planet performs better than WorldView data when examining shallower waters, probably because of the presence of the additional green band, whereas the opposite is true for deeper waters, where WorldView's higher spatial resolution gives it an edge. Furthermore, despite their simplicity, both networks exhibit significant accuracy scores, with the optimal neural network slightly outperforming the original. By optimising the model, we ensured that no model hyperparameter was impairing the predictive power of the network.

In Figure 10, we compare the ANN's prediction to the actual spatial distribution of the *Posidonia Oceanica* meadows. The green colour denotes the presence of Posidonia, while the red colour denotes

the presence of anything other than Posidonia (mostly sand). We can see how the network misses some of the smaller holes in the central Posidonia meadows, as well as some of the Posidonia patches on the right of the image, where the water is deeper. The bottom portion of the image corresponds to the northernmost part of Alcúdia bay, while the major central area corresponds to Pollença bay, as discussed previously.

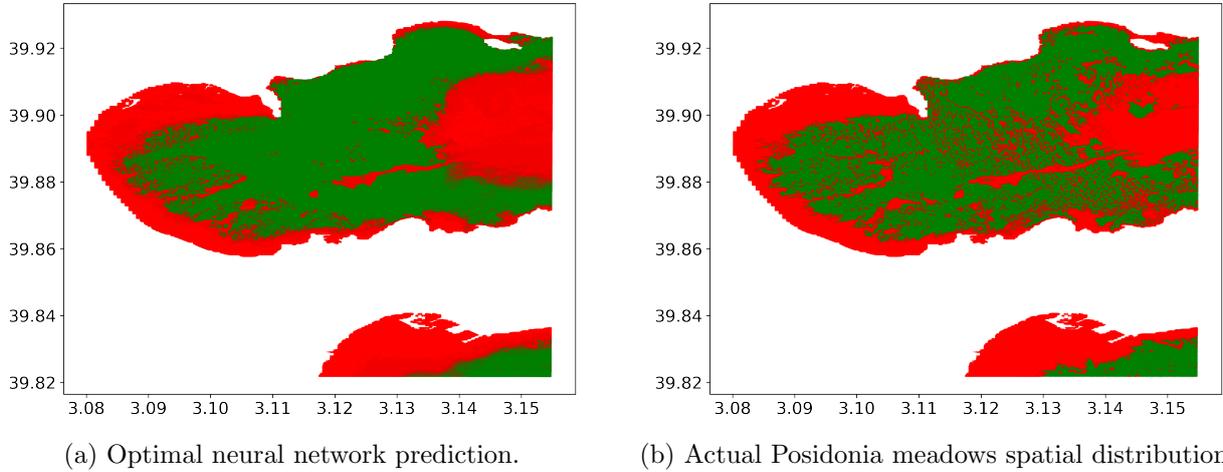


Figure 10: Comparison of predicted (left) versus actual (right) Posidonia meadows. In green, the posidonia meadows. In red, everything else.

4.3 Feature analysis

Regarding the SHaP values, the results are summarised in Figure 11 for the WorldView image, and in Figure 12 for the Planet image. In the WorldView image, when considering all depths, most of the predictive power comes from the green, blue, and depth features. The blue band, however, is less important when considering shallower waters. Moreover, since red light penetrates best in shallower waters (see Figure 5), it is more effective when considering depths of up to 20m. Finally, the red edge band appears to lack sufficient penetrative power in both cases and as a result has a very low SHaP value.

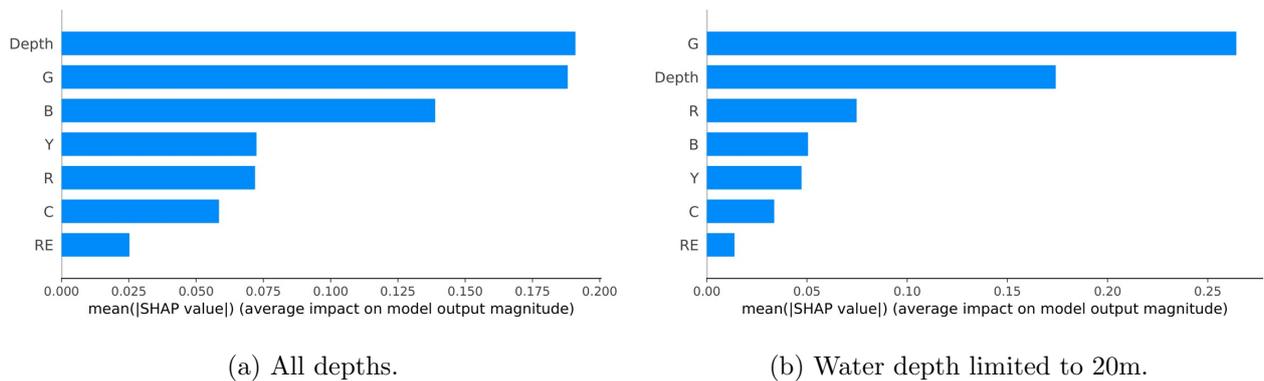


Figure 11: SHAP values for the features in the WorldView image.

For the Planet image, as a result of the added green band $G2$, we noticed a few differences. $G2$ was the second and third most important feature for both shallow and deep waters, respectively. Furthermore, the depth and the red band features no longer appeared to be as relevant. Everything else, however, remained similar to the prior case.

4.4 Testing other machine learning methods

After examining other machine learning methods (Tables 1 and 2), we find that our optimised artificial neural network (ANN) outperforms all of them, with the exception of the Gradient Boosting method for depths of up to 20m in the Planet image. Gradient boosting performs better than any

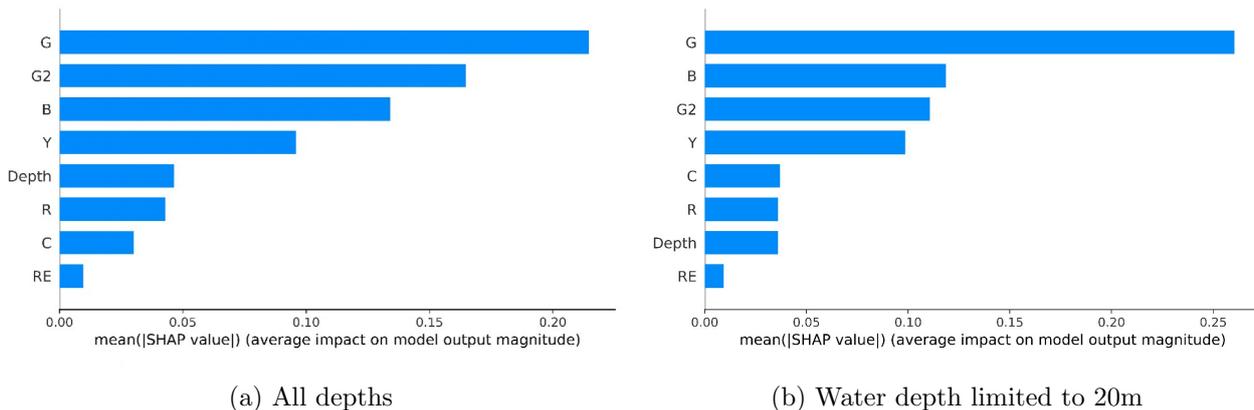


Figure 12: SHAP values for the features in the Planet image.

other method beside the ANN, and even more notably when considering deeper waters. In these regions, where visibility is worse, these other methods are not able to pick out the differences in transmitted radiance between the various seabed elements. The Gradient Boosting model and the ANN are approximately 2% more accurate than the other methods when looking at shallow waters, and this difference rises from 6% to 10% when considering deeper waters.

| | ANN (Original) | ANN (Optimal) | Grad. Boost | SGD | Log. Reg. | SVC |
|------------|----------------|---------------|-------------|-------|-----------|-------|
| All depths | 88.9% | 89.4% | 88.5% | 82.2% | 81.7% | 81.3% |
| Depth <20m | 91.5% | 91.8% | 91.5% | 89.0% | 89.1% | 89.0% |

Table 1: Precision of different ML models tested on the WorldView image

4.5 Band ratios as features

As previously described, we investigated several normalised indexes and band ratios as features. The accuracy scores obtained when testing on the WorldView image with the original neural network were 88.9 % and 91.4% for shallow waters and all depths, respectively. Without these new features, the cross-validation values were 88.9% and 91.5%, respectively. Evidently, these new features did little to improve our model. It is possible that our neural network already identified these reflectance-related effects through the training data. Nevertheless, training with these features on different ML models or other images could provide new information and prove insightful.

4.6 Water correction

Although applying a water correction had a positive effect on the results of similar studies [7], the accuracy of our model decreased by about 1% when using the corrected radiance values.

One difference with previous research of a similar nature is that we have many more data points, and our data 13, corresponding to Equation 5, deviates further from a clear trend. Another difference is that we use a neural network, which, without the need for additional adjustments, might be more effective at mitigating the effect of water absorption than other machine learning models. Nonetheless, more work in this area could potentially boost the network’s accuracy.

4.7 Comparing model performance in different days

The results after training and testing on different Planet images are shown in Figure 14. It is worth noting that, when training on the two clearest images, which are the day one and day two

| | ANN (Original) | ANN (Optimal) | Grad. Boost | SGD | Log. Reg. | SVC |
|------------|----------------|---------------|--------------|-------|-----------|-------|
| All depths | 87.0% | 87.7% | 86.5% | 77.4% | 77.4% | 77.4% |
| Depth <20m | 91.9% | 92.1% | 92.1% | 90.7% | 90.8% | 90.5% |

Table 2: Precision of different ML models tested on the Planet image

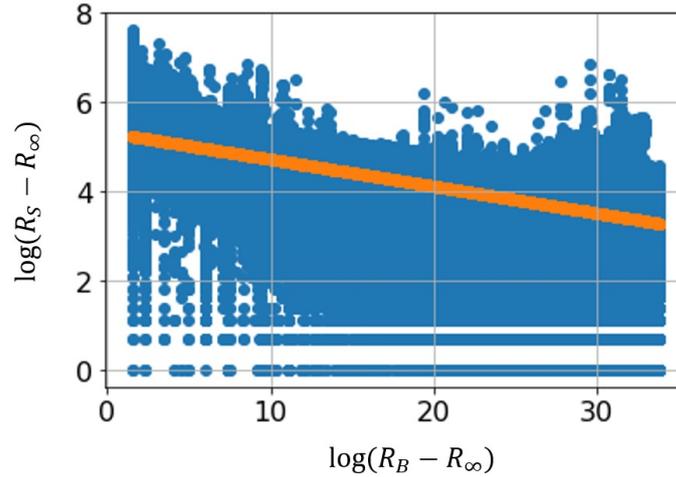


Figure 13: Data and fit from the red radiance band to find the appropriate attenuation coefficient, as described by Maritorena et al.’s model.

images, accuracy scores did not drop below 80 %, even when testing on the other two, less clear images. When training on the worse images, however, the accuracy scores are not as high. This might imply that a model built to study the evolution of *Posidonia* should only be trained with the best available images. However, we should be aware that these accuracy scores have standard deviations of about 2%.

Moving on to the principal component analysis, we found that the histograms of two images showed larger differences if the perceived quality gap between them was also large. As an example, Figure 15 shows that the day zero first principal component histogram, which originates from an image with lower quality and slightly less data (maybe due to the presence of clouds or reflections), lacks a *Posidonia* peak around the 4000 value point when comparing it to the day one histogram. Here, the red colour denotes the presence of *Posidonia*, while the green colour denotes the presence of anything other than *Posidonia*. Note that we only considered the first principal component because it explained 98% of the variation in our data.

To compare these distributions, we first use the Kolmogorov-Smirnov test or KS test. When applied to days 1 and 2, which appear to be the two best images, both subjectively and in terms of training accuracy scores, we obtain a KS value of 6.17. However, when looking at days 2 and 3, where 2 appears to be of much higher quality, the KS test results in a score of 10.81. This might offer a useful way to determine whether two images are compatible for training and testing. After considering the KS test, we employ the cosine similarity test. For days one and two, the test has a value of 0.999999840, while for days two and three, it has a value of 0.999999838. This test, however, did not prove consistent across different days and remains unreliable.

Furthermore, we briefly looked at how the radiance distributions differ for various bands and at different depths. However, due to time restrictions, we were unable to produce any significant findings. This area of the study needs more attention since it might offer a means of developing a reliable model which can consistently and accurately track the development of *Posidonia* meadows.

5 Conclusions and future work

In brief, we have successfully examined the effectiveness of artificial neural networks (ANN) for tracking *Posidonia* in Mediterranean seabeds through satellite images. More specifically, we have optimised our model to reach accuracy scores of 92% when looking at depths of up to 20m in the Planet image. We have also analysed other ML methods, among which Gradient boosting appeared to be the best, rivalling the ANN in terms of accuracy. Furthermore, we have studied which radiance bands provide the most useful information to our network and attempted to improve our model by introducing a water column correction and normalised band ratios. These attempts have not been successful, but merit further investigation. Finally, we have analysed how our network performs when trained and tested on different images, and what conditions affect the accuracy. When the satellite images are taken on clear days without the presence of boats, clouds, or reflections, accuracy scores remain consistently above 80 %. These values are comparable,

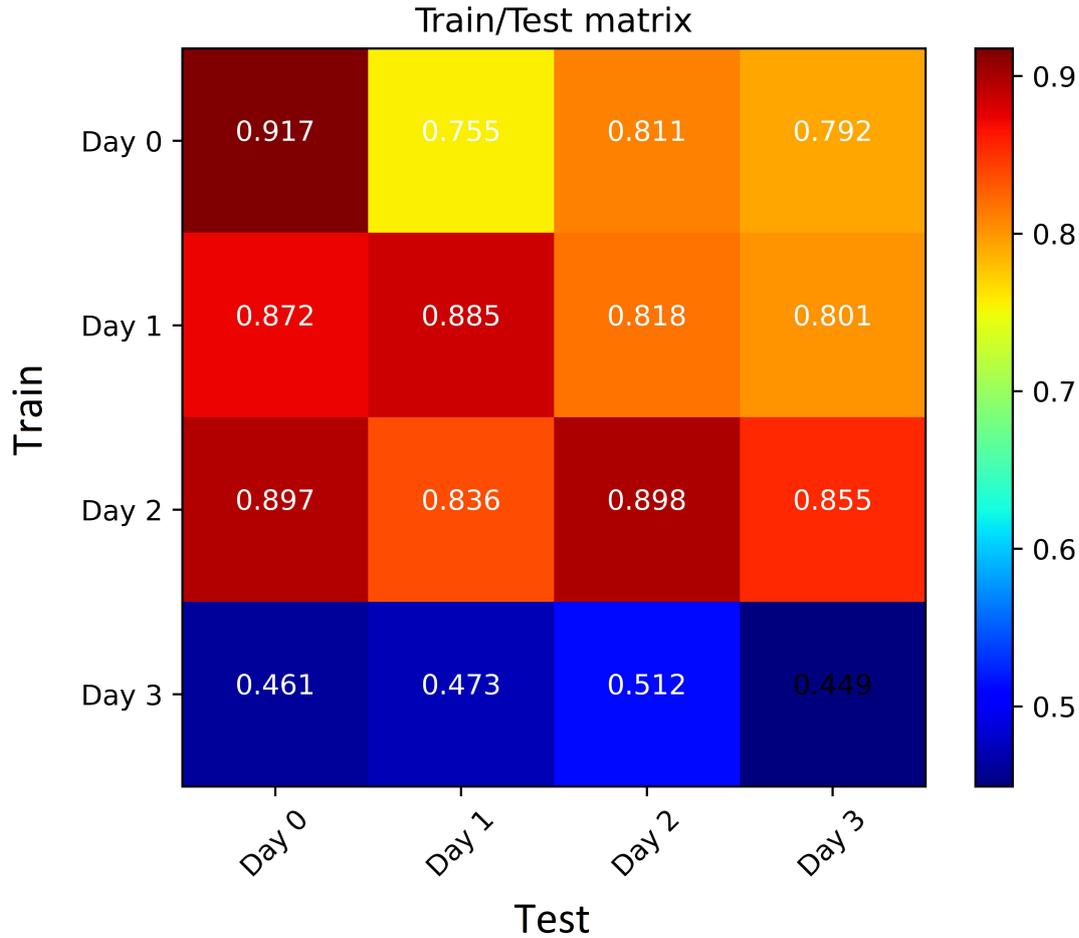


Figure 14: Test accuracy scores when training on one day and testing on another.

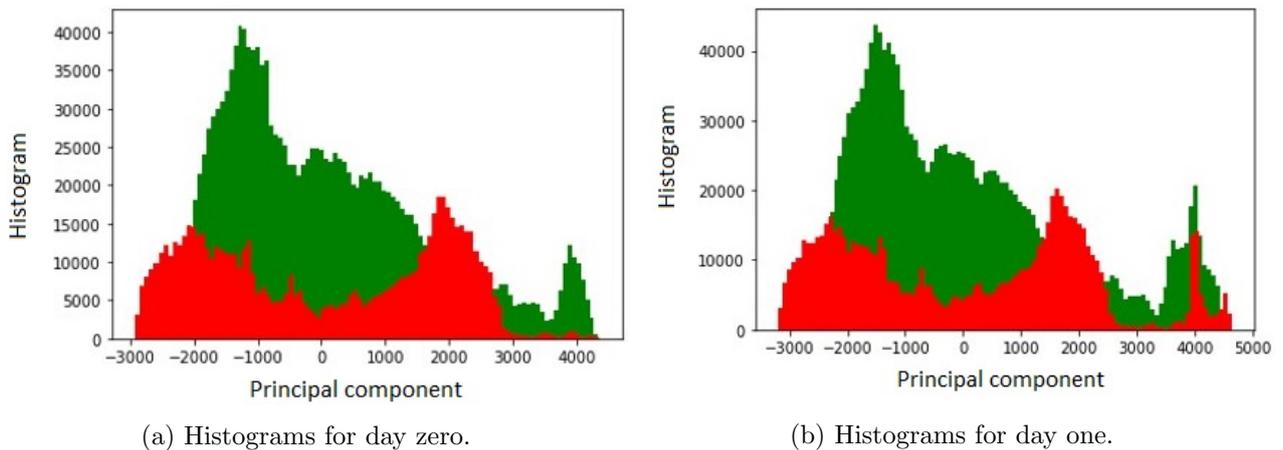


Figure 15: Comparison of our model's first principal components. In red, the posidonia meadows. In green, everything else.

and often better than those of similar studies, which often focus on shallower waters and smaller geographical areas [7] [10] [13] [14] [15].

This research could be advanced in a number of ways. First, we could also take the other seagrass species (e.g. *Cymodocea Nodosa*) and seabed elements into account. This might help our network discern what is Posidonia from what is not. Second, further analyses should be conducted to determine the conditions that make a satellite image suitable for model training. Third, more

work could be done to prevent problems affecting our images, such as the interactions at the air-water interface or light attenuation in water. Finally, we should consider training on the image of one geographical area and testing on the image of a different one. This way, our network could be used to study *Posidonia Oceanica* populations in different Mediterranean regions. Using an Alcudia bay photograph, we conducted a preliminary test, but the accuracy results were poor (60–70%) and the methodology needed to be reexamined.

In conclusion, since *Posidonia Oceanica* is crucial for the well-being of Mediterranean ecosystems and coastlines, scientific and political efforts should be made to ensure its prosperity. This line of research has the potential of improving the conservation and management of this species, and future developments on this subject remain highly exciting.

Acknowledgments

This work was supported by the SURF@IFISC fellowship. Furthermore, I am very grateful for Manuel Matías’ constant guidance and attention, and for the invaluable help from PhD student Àlex Giménez, who is incredibly original and productive, and who constantly pushed me forward in the best possible way. In brief, I am deeply thankful for the time spent in the program.

References

- [1] NÚRIA MARBÀ et al. “Regresión de praderas de *Posidonia oceanica* y calidad ambiental en el Parque Nacional del Archipiélago de Cabrera: Causas y Magnitud”. In: *Proyectos de investigación en parques nacionales. MAGRAMA* (2006), pp. 149–159.
- [2] Lucía Juliá Eivissa. *La regresión de la posidonia*. URL: <https://www.lavanguardia.com/vida/20191028/471230939177/la-regresion-de-la-posidonia.html>.
- [3] *Puerto Pollensa Sea Temperature*. URL: <https://seatemperatures.net>.
- [4] Agricultura i Pesca Govern de les Illes Balears Conselleria de Medi Ambient. *Atlas of Posidonia*. 2021. URL: <https://atlasposidonia.com/en/atlas-of-posidonia/>.
- [5] European Space Agency. *WorldView-3*. URL: <https://earth.esa.int/eogateway/missions/worldview-3>.
- [6] G2022 Planet Labs PBC. *Real-Time Satellite Monitoring with Planet*. URL: <https://www.planet.com/products/monitoring/>.
- [7] Dimos Traganos and Peter Reinartz. “Mapping Mediterranean seagrasses with Sentinel-2 imagery”. In: *Marine Pollution Bulletin* 134 (July 2017). DOI: 10.1016/j.marpolbul.2017.06.075.
- [8] Varsha Prasad, Rehna, and Rehna J. “Enhancing Underwater Gray Scale Images Using a Hybrid Approach of Filtering and Stretching Technique”. In: 1 (June 2015), pp. 33–36.
- [9] Heidi M Dierssen et al. “Ocean color remote sensing of seagrass and bathymetry in the Bahamas Banks by high-resolution airborne imagery”. In: *Limnology and oceanography* 48.1part2 (2003), pp. 444–455.
- [10] Flavio Borfecchia et al. “Mapping spatial patterns of *Posidonia oceanica* meadows by means of daedalus ATM airborne sensor in the coastal area of Civitavecchia (Central Tyrrhenian Sea, Italy)”. In: *Remote Sensing* 5.10 (2013), pp. 4877–4899.
- [11] Flavio Borfecchia et al. “*Posidonia oceanica* genetic and biometry mapping through high-resolution satellite spectral vegetation indices and sea-truth calibration”. In: *International Journal of Remote Sensing* 34.13 (2013), pp. 4680–4701.
- [12] Govindarajulu Z. “A brief survey of nonparametric statistics”. In: *Communications in Statistics - Theory and Methods* 5.5 (1976), pp. 429–453. DOI: 10.1080/03610927608827365. eprint: <https://doi.org/10.1080/03610927608827365>. URL: <https://doi.org/10.1080/03610927608827365>.

- [13] A Fornes et al. “Mapping Posidonia oceanica from IKONOS”. In: *ISPRS Journal of Photogrammetry and Remote Sensing* 60.5 (2006), pp. 315–322.
- [14] Erica Matta et al. “Mapping Posidonia meadow from high spatial resolution images in the Gulf of Oristano (Italy)”. In: *2014 IEEE Geoscience and Remote Sensing Symposium*. IEEE. 2014, pp. 5152–5155.
- [15] Vanina Pasqualini et al. “Use of SPOT 5 for mapping seagrasses: An application to Posidonia oceanica”. In: *Remote Sensing of Environment* 94.1 (2005), pp. 39–45.

AN ORDINAL PATTERN APPROACH TO WORD CORRELATIONS

Antonio A. Carpes, David Sánchez

Instituto de Física Interdisciplinar y Sistemas Complejos, IFISC (CSIC-UIB)

Campus Universitat de les Illes Balears, E-07122 Palma de Mallorca, Spain

Abstract

Words are the basic elements of a language. They are combined with each other to form up ideas and create more complicate structures. The syntactic rules are the framework in which words can be combined, as not every combination is valid. Sentences and phrases are the result of the existence of those rules and they can be totally different depending on the language or even within a certain language itself. In fact, languages show variations due to location, gender, age, social class, etc. In this work, we are concerned with the geographical variation and the study of dialects. The study of the geographical variations is performed by means of an ordinal pattern approach dealing with texts both in Mallorca and mainland Spain. With this analysis in hand, we are able to spot characteristic trigrams and tetragrams for each location. This opens a door to an analytical approach to the study of dialects in a language and shows the remarkable interest of this method.

1 Introduction

Languages are complex structures based on the linguistic unit which is the word. Words can be combined in order to form sentences and sentences can eventually form up texts. In spite of their complexity, words cannot be combined arbitrarily, but instead, the way words dovetail is constrained by the so called syntactic rules. These rules impose restrictions in a way that some linguistic structures are reinforced, while some others can be forbidden.

In this work, we did not take into account specific syntactic rules but instead the method employed is a statistical analysis whose fundamental concept is the occurrence frequency of words and groups of words in a given language. Because of this, the approach is valid for every language and does not require a previous knowledge of the language under study. This advantage of the method makes it so interesting because it provides a tool to study the differences between different languages in an analytical way.

This approach based on the occurrence frequency of words is not an arbitrary decision but is substantiated on the **Zipf's Law**.

Zipf's Law is an empirical law which states that if we were to have a ranking of the words of a language according to the rate of occurrence in the language, the frequency of each word of the ranking would be inversely proportional to its ranking position. This law is mathematically formulated as:

$$f(r) = \frac{C}{r^\alpha} \tag{1}$$

where $f(r)$ denotes the frequency of the word taking position r in the ranking. C stands for the frequency of the first word of the ranking and α is an exponent typically chosen equal to 1.

For instance, a set of data distributed according to the Zipf's Law would look as:

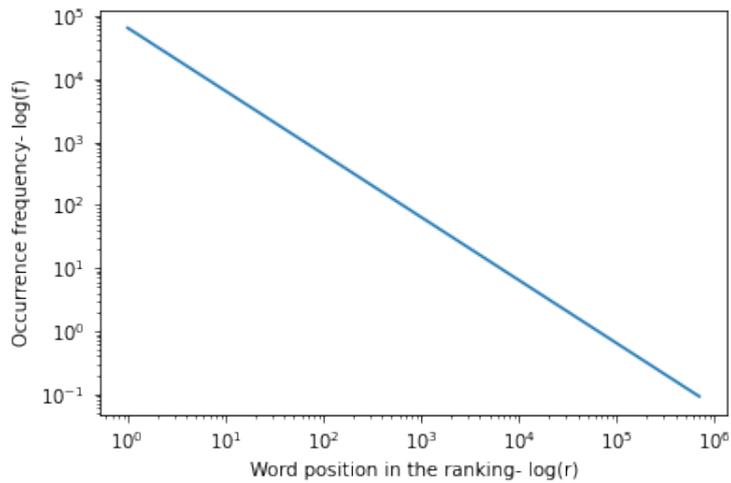


Figure 1: Line for the perfectly Zipf's Law correlated text

As an illustrative example, if we take the English Bible as a representative text to study the Zipf's Law in English we get the following results:

| | Word | Count |
|--------------|----------|-------|
| 0 | the | 63925 |
| 1 | and | 51700 |
| 2 | of | 34615 |
| 3 | to | 13558 |
| 4 | that | 12914 |
| ... | ... | ... |
| 12850 | tabering | 1 |
| 12851 | folks | 1 |
| 12852 | madian | 1 |
| 12853 | rhoda | 1 |
| 12854 | samos | 1 |

Table 1: English Bible's words ranked by their frequency of occurrence

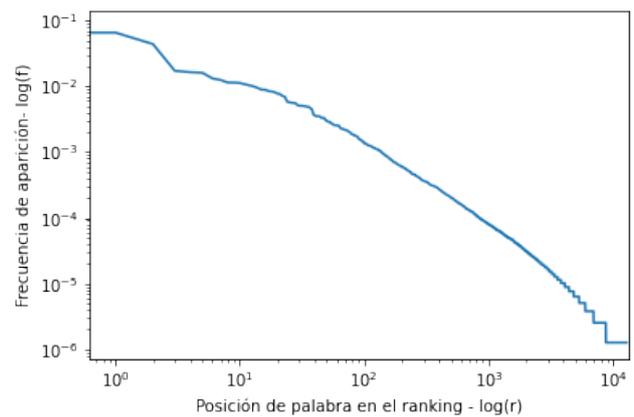


Figure 2: Zipf's curve associated to the English Bible

If we do the same analysis for the Spanish Bible we get the following results:

| | Word | Count |
|--------------|-------------|--------------|
| 0 | de | 44132 |
| 1 | y | 31705 |
| 2 | a | 20888 |
| 3 | que | 20347 |
| 4 | la | 19493 |
| 5 | el | 18365 |
| 6 | los | 17450 |
| ... | ... | ... |
| 27293 | convicta | 1 |
| 27294 | capturadlo | 1 |
| 27295 | vacilación | 1 |

Table 2: Spanish Bible’s words ranked by their frequency of occurrence

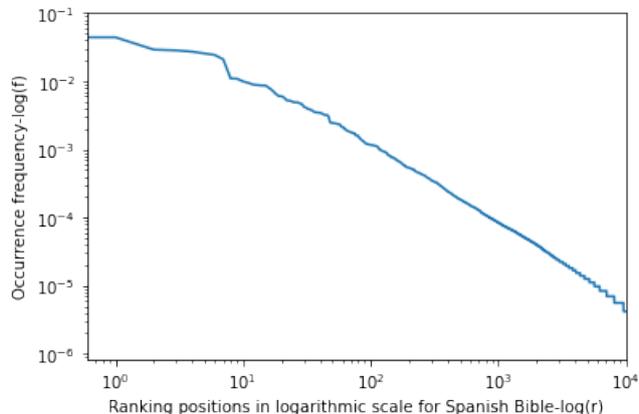


Figure 3: Zipf’s curve associated to the Spanish Bible

In Figs. 2 and 3 we observe that a real text shows deviations from a perfect Zipf law. In fact, this rule holds only for a finite interval of rankings, as demonstrated in Ref. [1]

2 Theoretical model

We consider a text long enough to be representative of a given language (hereafter, the corpus). Then, let P be the total number of words. We rank the words of the text according to their absolute frequency of occurrence in the text in such a way that we end up having a list from the most repeated word to the least. However, this process is not necessarily a bijection, since there can be words with the same absolute frequency and thus, the same ranking. The most appropriate way that has been designed to break down this equalities consists on adding small noise to the ranking numbers which are repeated in order to make them slightly different as shown in Ref. [2]

When we obtain the ranking positions in bijection with the words of the text, we can write the text as a list of numbers, where each number in the list is the position in the ranking of the corresponding word.

Once, the text has been transformed into a list of numbers, the next step is to extract a list of ordinal patterns as shown in Ref. [3]

For this aim, let us define the embedding dimension, D , as the number of words considered to extract a pattern. In this research we employ $D = 2, 3$ and 4 . For a given embedding dimension D , there exist $D!$ patterns, which is precisely the order of the finite symmetric group of permutations S_D .

We now label the $D!$ permutations of S_D with names from 1 to $D!$

Then we consider D consecutive words and their associate ranking positions, and arrange them from lowest to highest. After that, we calculate the permutation associated to the rearrangement that has been performed. Therefore, we have transformed the set of D words into a number corresponding to the permutation.

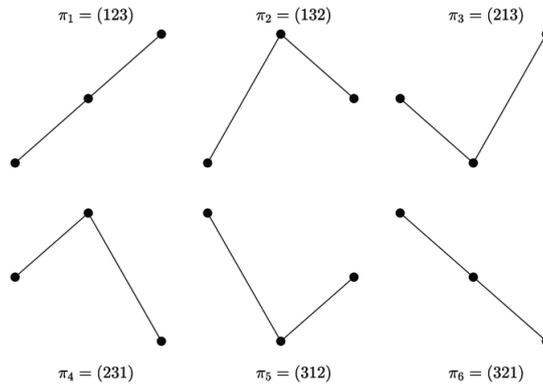


Figure 4: 6 patterns for $D = 3$

For instance, suppose we have a set of three words with associated ranking numbers $(34, 98, 4)$. Then, we shall order them from lowest to highest, thus getting $(4, 34, 98)$. The associated permutation is $(2, 3, 1)$ with number 4 (see Fig. 4).

The process previously shown is repeated along the whole text, taking a shift from a set of words to the consecutive one of one word.

A list of ordinal patterns is the result of this process. The following idea is to compute the probabilistic density of each pattern in the text, by simply taking the number of patterns of each type over the total number of patterns.

When computing the probabilistic density there are two options:

One is to simply compute the relative frequency of each pattern along the whole text and plot the different probabilities.

The second option is more intricate and not so relevant for our research. It consists of taking **temporal windows**, which are subtexts of 10^4 words and compute the relative frequency in that regions. Then, shift 10^3 words and repeat the same process with the next 10^4 words. This way, if we plot the relative frequency for a given pattern as a function of each temporal window, we arrive to a dynamic graphic which is characteristic for every given language.

In both cases, the relative frequency of the patterns is computed as:

$$\rho_i = \frac{f_i}{L} \quad (2)$$

where f_i stands for the absolute frequency of the i -th pattern and L stands for the total number of patterns that there are in a temporal window.

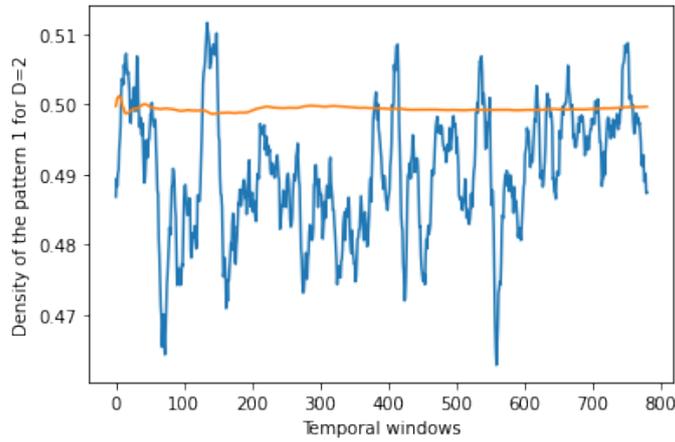
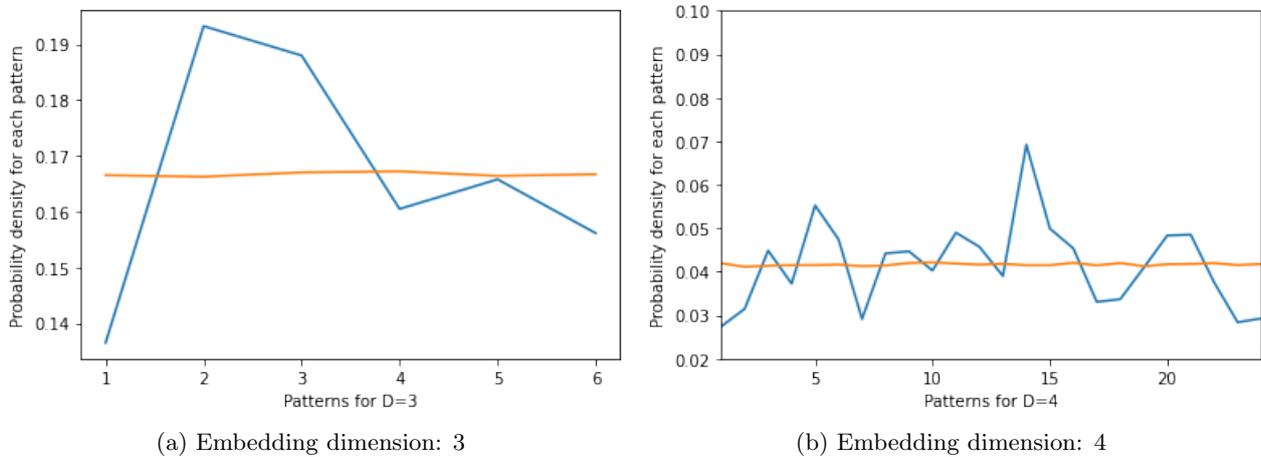


Figure 5: In blue, the density of the pattern 1 in each window for the original English Bible. In orange, the density of the pattern 1 in each window for the shuffled English Bible.

As expected, when the Bible is shuffled, the syntactic rules no longer apply, then the probability of finding pattern 1 is equal to that of finding pattern 2.

More interesting for our research is the stationary pattern probability distribution, which is the first case of the two cases discussed above. The idea is that only a window is considered, the whole text, and the density is calculated as in eqn. (2).

When applied to the English Bible, the following patterns are obtained:



(a) Embedding dimension: 3

(b) Embedding dimension: 4

Figure 6: Graphs for both original (blue) and shuffled English Bible (orange) ordinal patterns distribution

When applied to the Spanish Bible, these pattern graphs are obtained:

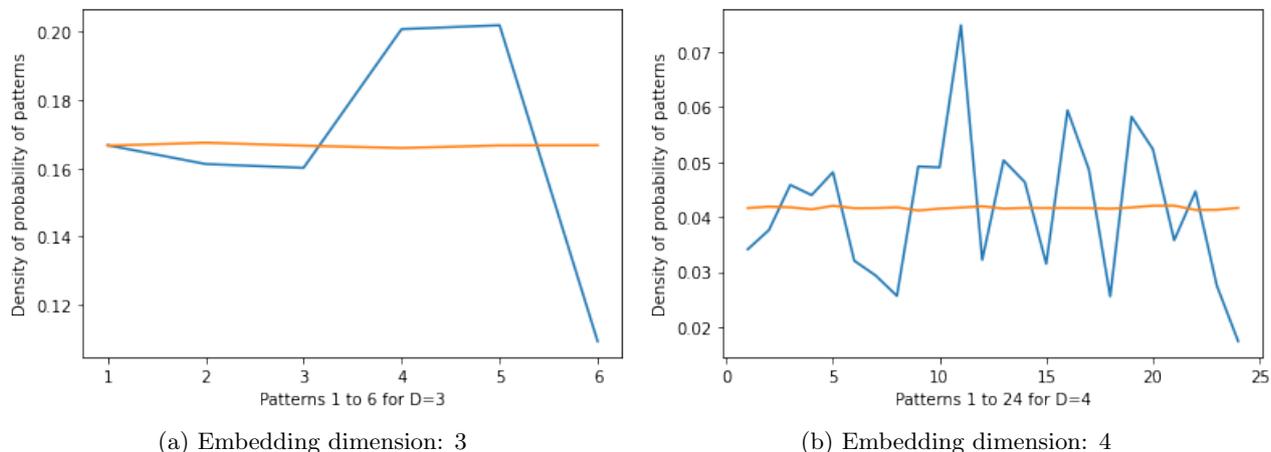


Figure 7: Graphs for both original (blue) and shuffled Spanish Bible (orange) ordinal patterns distribution

As we can easily check from the graphs above, the pattern for each embedding dimension is characteristic for a given language, as we know from Ref. [2]. For embedding dimension $D = 3$, in English, the second and third pattern are predominant, while in Spanish the fourth and fifth are the most repeated ones. This shows the power of our method. Its interest lies on the fact that a text can be studied and classified in different aspects according only to the density of pattern distribution, without a need to have fully understanding of the language.

Once this Bible analysis is performed we are in conditions to study two historical linguistic corpora. One of which is formed by texts written in Mallorca [4] in the XVII century and another formed by texts written in mainland Spain in that same century [5].

Details of these texts are given in the next table:

| | No. documents | No. tokens | No. types |
|----------|----------------------|-------------------|------------------|
| Charta | 94 | 76283 | 10362 |
| Mallorca | 514 | 293906 | 19550 |

The first idea is trying to obtain information from the texts based on the pattern extracted that can help us determine the linguistic differences between the different territories.

The first step is to plot the Zipf's curve for both texts in order to check that our texts are Zipf distributed which is the basis of our study.

For this, we first need to rank by frequency of appearance the words in each text, for which we obtain the following tables:

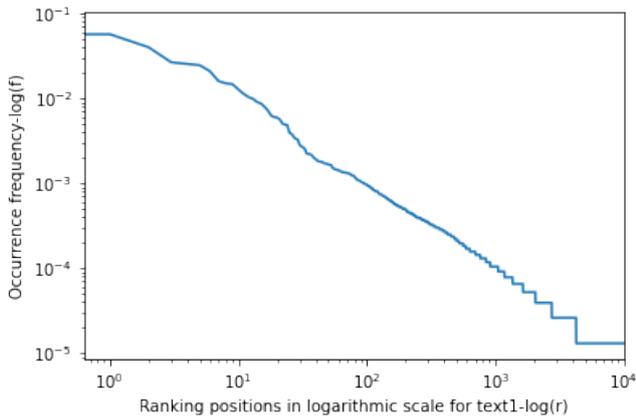
| | Word | Count |
|--------------|-----------|-------|
| 0 | de | 5287 |
| 1 | y | 4336 |
| 2 | que | 3046 |
| 3 | en | 2028 |
| 4 | la | 1943 |
| 5 | el | 1872 |
| ... | ... | ... |
| 10359 | seescuso | 1 |
| 10360 | reberente | 1 |
| 10361 | conforma | 1 |

Table 3: Text 1: corpus Charta ranked by frequency of occurrence

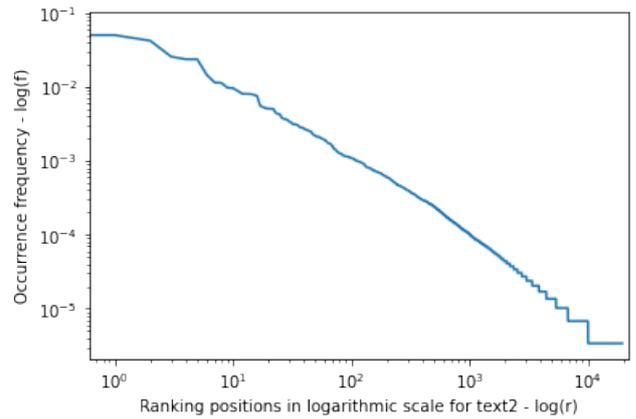
| | Word | Count |
|--------------|-----------|-------|
| 0 | de | 18967 |
| 1 | y | 14685 |
| 2 | que | 12343 |
| 3 | la | 7520 |
| 4 | en | 6930 |
| 5 | el | 6923 |
| ... | ... | ... |
| 19557 | probehido | 1 |
| 19558 | creio | 1 |
| 19559 | vestirse | 1 |

Table 4: Text 2: corpus Mallorca ranked by frequency of occurrence

This data leads us to the following Zipf's diagrams:



(a) Text 1: corpus charta



(b) Text 2: corpus Mallorca

Figure 8

As it is easily noticed, both texts roughly follow Zipf's distribution so we are now ready to employ our theoretical method on both texts.

3 Results and discussion

When the density pattern distribution is calculated for embedding dimension $D = 2, 3$ and 4 for both texts we arrive to the following graphs.

From now on blue stands for text 1, i.e. corpus Charta and orange stands for text 2, i.e., corpus Mallorca.

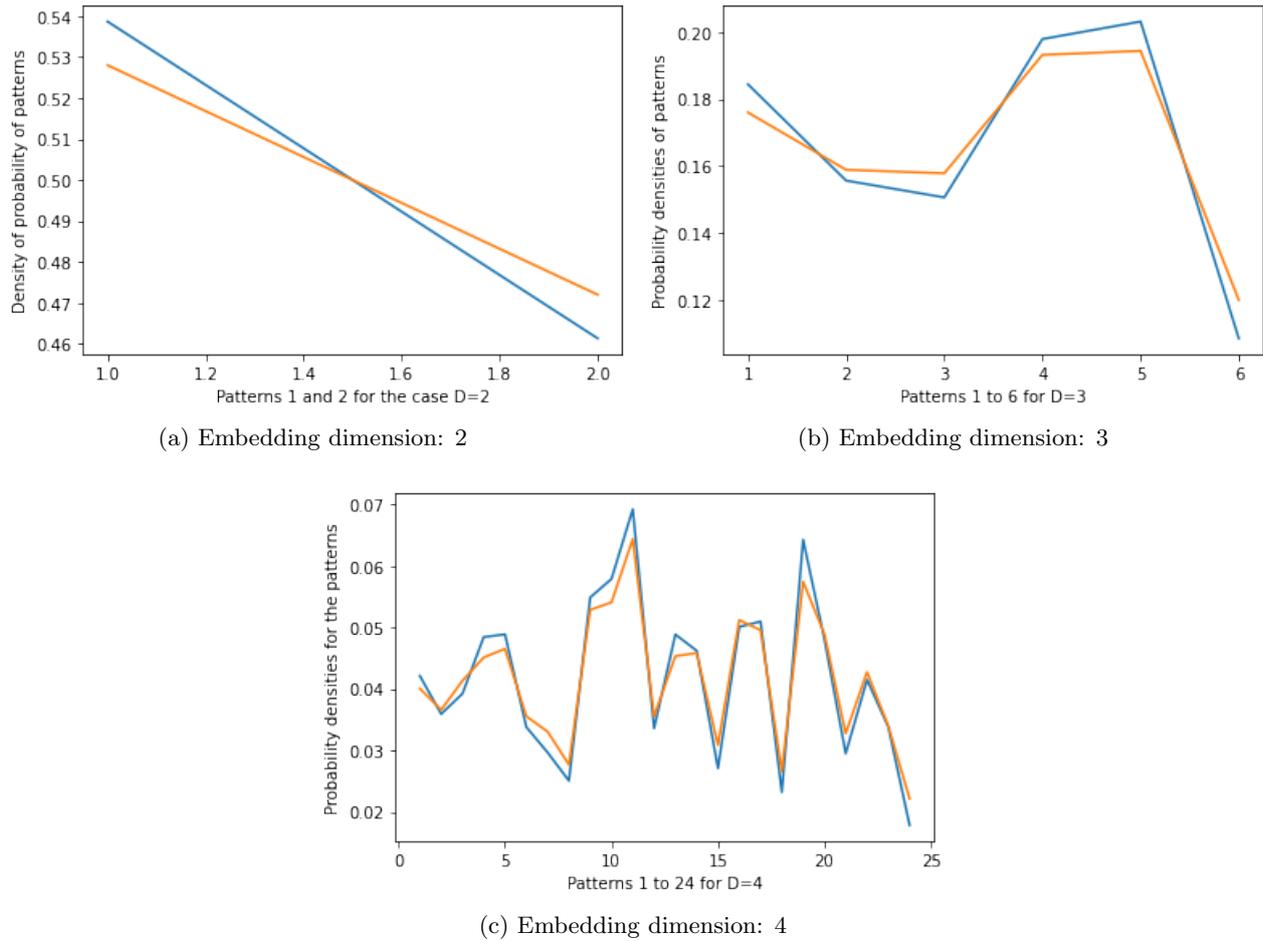


Figure 9

When we first look at these graphs, we notice the slight differences in pattern distribution densities. Our first task is to check that these fluctuations are not a consequence of statistical deviations. For this, we take the $D = 3$ pattern distribution and do as follow:

1. Divide both texts in 10 parts.
2. In each division, we compute the relative frequency of every pattern from 1 to 6.
3. We then compute the mean of the relative frequency for each pattern in both texts along the 10 divisions and the typical deviation for each pattern relative frequency mean.
4. Finally, we plot in the same graph both pattern distribution means with their respective typical deviation values as error bars.

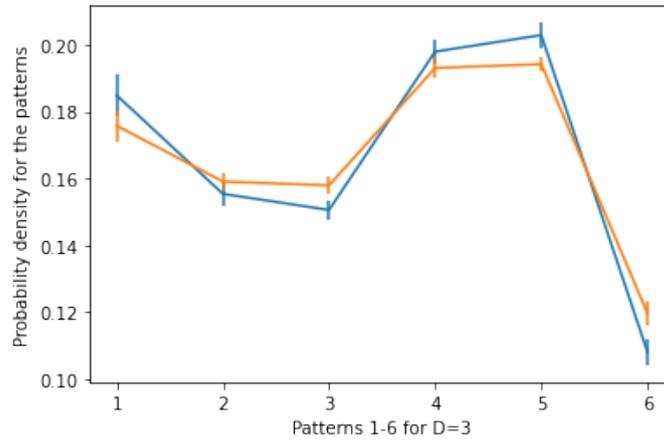


Figure 10

With this, we can be sure that our fluctuations are not due to statistical deviances, since the obtained values are incompatible as none of the points in the graph for each text happens to lie inside the error interval of the same given point for the other text.

Therefore, we understand that these fluctuations are the result of a different pattern distribution in each text according to linguistic differences. This is the key of our analysis and these linguistic differences is what we intend to unveil.

As an approach, the idea is to examine the most frequent bygrams and trigrams in both texts (since they are the most determining for the pattern densities for $D = 2$ and 3) and check whether there exist linguistic differences responsible for the ordinal pattern disagreements between both texts.

In the next tables, we can see the 15 more repeated bygrams and trigrams for each text with their associated ordinal patterns.

| | Bygram | Count | Pattern |
|-----------|------------|-------|---------|
| 0 | de la | 471 | 1 |
| 1 | que se | 309 | 1 |
| 2 | la ciudad | 301 | 1 |
| 3 | en el | 267 | 1 |
| 4 | a la | 248 | 2 |
| 5 | para que | 243 | 2 |
| 6 | señor don | 236 | 2 |
| 7 | el señor | 231 | 1 |
| 8 | de esta | 224 | 1 |
| 9 | de su | 209 | 1 |
| 10 | en la | 200 | 1 |
| 11 | de los | 186 | 1 |
| 12 | en este | 176 | 1 |
| 13 | lo que | 172 | 2 |
| 14 | don joseph | 172 | 1 |

Table 5: Text 1: corpus Charta. 15 most repeated bygrams

| | Bygram | Count | Pattern |
|-----------|------------|-------|---------|
| 0 | de la | 1567 | 1 |
| 1 | en la | 1248 | 2 |
| 2 | de que | 1045 | 1 |
| 3 | que se | 966 | 1 |
| 4 | lo que | 898 | 2 |
| 5 | y que | 882 | 1 |
| 6 | en el | 738 | 1 |
| 7 | bayle real | 686 | 2 |
| 8 | de su | 586 | 1 |
| 9 | que doy | 555 | 1 |
| 10 | de esta | 522 | 1 |
| 11 | que no | 513 | 1 |
| 12 | y de | 499 | 2 |
| 13 | doy fee | 493 | 1 |
| 14 | villa de | 488 | 2 |

Table 6: Text 2: corpus Mallorca. 15 most repeated bygrams

From these tables we can extract some conclusions:

We notice from the graph that pattern 1 is a little bit more common than pattern 2 in both texts. This can be probably associated to the fact that pattern 1 stands for the majority of the structures of the type:

1. Article + Noun
2. Preposition + Adverb/Pronouns
3. Preposition + Article

These are obviously very common structures in Spanish, since they suppose the natural way in which nouns are introduced in a text or an idea is connected to another.

On the other side, pattern 2 is a little bit less common and it is a pattern more associated to:

1. Pronoun + “que”.
2. Noun + Generalist Adjective (such as “bayle real”, in which the adjective “real” is much more frequent than the word “bayle”).
3. Conjunction + Preposition.

| | Trigram | Count | Pattern |
|-----------|--------------------------|--------------|----------------|
| 0 | el señor don | 155 | 2 |
| 1 | en este cauildo | 128 | 1 |
| 2 | de esta ciudad | 79 | 2 |
| 3 | los caualleros rexidores | 79 | 1 |
| 4 | la ciudad de | 76 | 4 |
| 5 | los señores don | 70 | 4 |
| 6 | de esta villa | 63 | 1 |
| 7 | del mes de | 63 | 4 |
| 8 | de su magestad | 59 | 1 |
| 9 | a la ciudad | 59 | 3 |
| 10 | don luis de | 56 | 4 |
| 11 | don juan de | 50 | 4 |
| 12 | dias del mes | 44 | 5 |
| 13 | la ciudad acuerdo | 44 | 1 |
| 14 | cauildo el señor | 43 | 5 |
| 15 | ciudad de malaga | 42 | 3 |
| 16 | de leon don | 42 | 2 |
| 17 | señor don joseph | 41 | 3 |
| 18 | la villa de | 41 | 4 |
| 19 | de la ciudad | 40 | 1 |

Table 7: Text 1: corpus Charta. 20 most repeated trigrams

| | Trigram | Count | Pattern |
|-----------|--------------------------|--------------|----------------|
| 0 | de que doy | 525 | 1 |
| 1 | que doy fee | 385 | 1 |
| 2 | la villa de | 281 | 4 |
| 3 | de edad de | 239 | 2 |
| 4 | de esta villa | 220 | 1 |
| 5 | del mes de | 217 | 4 |
| 6 | dios nuestro señor | 215 | 4 |
| 7 | reyno de mallorca | 206 | 5 |
| 8 | de esscribano real | 205 | 2 |
| 9 | en la villa | 197 | 3 |
| 10 | una señal de | 196 | 4 |
| 11 | assistente de esscribano | 194 | 5 |
| 12 | señal de cruz | 191 | 5 |
| 13 | mes y año | 185 | 3 |
| 14 | de lo que | 182 | 2 |
| 15 | presente curia real | 175 | 4 |
| 16 | en forma de | 175 | 4 |
| 17 | bennasser assistente de | 169 | 4 |
| 18 | del juramento que | 168 | 4 |
| 19 | del reyno de | 168 | 4 |

Table 8: Text 2: corpus Mallorca. 20 most repeated trigrams

As we have seen in the graph distribution for $D = 3$, the most frequent ordinal pattern is 5. This is again a logical result since as shown in these previous tables, the pattern 5 is associated with expressions such as “dias del mes”, “cauildo el señor”, “reyno de mallorca”, “assistente de esscribano” or “señal de cruz”. All of these expressions have in common the existence of two words with semantic charge connected by a preposition. More precisely, the last word is the word with the biggest semantic charge and the first one is a more general noun. This structure is very typical in Spanish and it operates connecting two words in which the first word alludes to the last one.

Moreover, we can use these tables to already try to distinguish the two texts. For instance, in the first text the most frequent trigram is “el señor don”, which does not appear in the most frequent

trigrams for the second text, so the presence of this trigram in a given text is a good indicative that the given text might belong to corpus Charta. The same thing occurs for trigrams such as “en este cauildo”, “los caualleros reidores”, “de su magestad” or “ciudad de malaga”, along with many others. These trigrams are found amongst the 20 most frequent trigrams, but not among the 20 most frequent in text 2. Therefore, they are a good factor to be taken into account when trying to decide whether a text belongs to corpus Charta or Mallorca.

If we now study the table associated to text 2, again we find characteristic trigrams which do not even appear on the first table, such as “de que doy”, “que doy fee”, “dios nuestro señor”, “reyno de mallorca”, “de esscribano real”, “assistente de esscribano” or “del reyno de”. These are some of the most repeated trigrams in the texts of corpus Mallorca and the presence of these structures in a given text is a good indicate of its procedence. Structures such as “reyno de mallorca” or “del reyno de” have a huge semantic meaning associated to the historical location, but there are some others such as “dios nuestro señor” or “que doy fee” which might reflect a given protocol when it comes to dealing with legal documents which is clearly different that the conventions existing at the time in mainland Spain. This is something we infer from the fact that trigrams such as “que doy fee” or “dios nuestro señor” do not appear among the most frequent for corpus Charta texts.

Now, we provide some examples of the tetragrams with ordinal pattern 11 which is the most repeated in both text 1 and text 2:

| Tetragram | Count | Pattern |
|---------------------------------|-------|---------|
| a cauildo en su | 36 | 11 |
| caualleros reidores los señores | 35 | 11 |
| se junto a cauildo | 24 | 11 |
| lo presidio el señor | 22 | 11 |
| joseph ponze de leon | 19 | 11 |
| diego ponze de leon | 18 | 11 |
| y reximiento de ella | 16 | 11 |
| su magestad y señores | 16 | 11 |
| manuel licardo de riuera | 15 | 11 |
| se juntto a cauildo | 13 | 11 |

Table 9: Text 1. Most frequent tetragrams with ordinal pattern 11

| Tetragram | Count | Pattern |
|-----------------------|-------|---------|
| una señal de cruz | 187 | 11 |
| del reyno de mallorca | 166 | 11 |
| que supiere y fuere | 118 | 11 |
| se afirma y ratifica | 96 | 11 |
| la tierra de dicho | 68 | 11 |
| años poco mas ô | 57 | 11 |
| â ella y dixo | 57 | 11 |
| la ciudad de palma | 48 | 11 |
| los pies de vs | 47 | 11 |
| le hace el testigo | 43 | 11 |

Table 10: Text 2. Most frequent tetragrams with ordinal pattern 11

As we can conclude from these tables, the pattern 11 is associated with structures of the type: preposition/article + word with large semantic charge + preposition + word with semantic charge. It is logical to obtain this pattern as the most frequent one, since this type of linguistic structure is very common in the language as it connects larger structures or ideas.

4 Conclusions

In this work we have analysed word correlations using an ordinal pattern approach. We have been able to conclude that this approach is very powerful since it systematizes the study of the differences between languages without the need of having a deep understanding of them. Besides, we have seen that the stationary probability distribution for the patterns is characteristic for a given language.

Our analysis has also allowed us to study different texts written in Spanish on various geographical locations. We have checked that the pattern analysis is sensitive to these geographical variations and thus, paves the way to a characterization of dialects for a given language.

Even though in our project we have not reached a fully decisive conclusion on how dialects can be characterized using this method, we have glimpsed a way that can lead us to eventually arriving to our objectives.

Acknowledgments

I would like to thank my advisor David Sánchez for his continuous guidance and help along the project, as well as, Manuel Matías and all of the IFISC team for this great opportunity to take my first steps towards research in an institution as reputed as IFISC.

This work was supported by a SURF@IFISC fellowship.

References

- [1] M. A. Montemurro, *Physica A* 300, 567 (2001).
- [2] David Sánchez, Luciano Zunino, Juan de Gregorio, Raúl Toral and Claudio Mirasso, *Universality and diversity in word patterns*, arXiv:2208.11175, (2022).
- [3] M. Zanin, F. Olivares, *Ordinal patterns-based methodologies for distinguishing chaos from noise in discrete time series*, *Communications Physics* 4, 190, 2021.
- [4] <http://http://www.corpusmallorca.es/>
- [5] <https://www.corpuscharta.es/>

Interferometry with topological insulator nanowires

Carlota Prieto Jiménez, Llorenç Serra

Instituto de Física Interdisciplinar y Sistemas Complejos, IFISC (CSIC-UIB)
 Campus Universitat de les Illes Balears, E-07122 Palma de Mallorca, Spain

Abstract

In this report, we have studied topological insulators (TI), materials that have the properties of a conventional insulator in the bulk but host topologically protected metallic states on the surface. We have studied the effect that underlies the TI, the quantum anomalous Hall effect. We have focused on the study of TI theoretically and we have modeled different situations, changing the magnetic field and the dimensions of the wire. We have also discussed the Dirac paradox in 1D and 3D from a theoretical point of view.

1 Introduction

A topological insulator (TI) is an electronic material that has a bulk band gap like an ordinary insulator but has protected conducting states on its edge or surface, therefore, electrons can only move along the surface of the material [1]. These states are possible due to the combination of a strong spin-orbit coupling (SOC) and a band inversion that leads to a time-reversal symmetry [2]. The surface states behave as massless Dirac fermions (a spin-1/2 particle that is different from its antiparticle) with a linear dispersion relation. An important characteristic of topological insulators is that the length has to be an integer number of k , the wavevector, or wavenumber depending on the dimension we are in. Otherwise, we would have backscattering.

The first models of 3D topological insulators were proposed by Volkov and Pankratov in 1985, and in 1987 by Pakhomov. Gapless 2D Dirac states were shown to exist at the band inversion contact in PbTe/SnTe and HgTe/CdTe heterostructures. In 2007, the existence of interface Dirac states in HgTe/CdTe was experimentally verified by Molenkamp's group in 2D topological insulators.

Due to unintentional doping because of anti-site defects, it is very challenging to fabricate three-dimensional topological insulator (3D TI) samples with an intrinsic Fermi level that lies in the bulk gap and near the Dirac point. This makes it difficult to resolve the transport properties of 3D TI surface states, with the majority of the charge carriers originating from the bulk. To get into a regime where the 3D TI surface-state (magneto) transport signatures are more pronounced, 3D TI nano samples with electrostatic gating are commonly considered.

In the bulk of a non-interacting topological insulator, the electronic band structure resembles an ordinary band insulator, with the Fermi level falling between the conduction and valence bands. On the surface of a topological insulator, there are special states that fall within the bulk energy gap and allow surface metallic conduction. Carriers in these surface states have their spin locked at a right angle to their momentum, which is called spin-momentum locking.

Therefore, due to this spin-momentum locking a paradox arises when we are in a 1+1 dimension (one spatial dimension and the time dimension), the Dirac paradox. If we only have massless Dirac models, we do not have solutions. However, if we have additional massless Dirac models we then have nontrivial scattering solutions [3].

Thus, we divided this report into two parts. In the first one, we studied the properties of the solutions of the energy as a function of the wavenumber. In the second part, we studied the Dirac paradox in the junction between the Dirac materials with opposite helicity.

2 Theoretical model

For 3D TI nanowires, we consider the 3D continuum model Hamiltonian [4]. We want to solve the Schrödinger equation:

$$H\psi = E(k)\psi. \quad (1)$$

with ψ the wave function that depends on the spatial components (x, y, z) and also depends on two parameters η_σ and η_τ which are the spin and the pseudo spin.

We consider the 3D continuum model Hamiltonian:

$$H(\mathbf{k}) = \epsilon(\mathbf{k}) + M(\mathbf{k})\tau_z + A_\perp\tau_x(k_x\sigma_x + k_y\sigma_y) + A_\parallel\tau_xk_z\sigma_z \quad (2)$$

where k_x, k_y, k_z are the three components of the wavevector and $\sigma_x, \sigma_y, \sigma_z$ and τ_x, τ_y, τ_z are the Pauli matrices for atomic-orbital and spin degrees of freedom, respectively. We are going to consider that the Dirac point energy corresponds to $E=0$.

To solve the equation we take into account the boundary conditions and the finite differences method. The boundary condition for the wave function satisfies $\phi_k(y, z, \eta_\sigma, \eta_\tau) = 0$. And, we use a mesh of points and calculate the finite differences between neighbor points.

The space is given by the position, the spin, and the pseudo-spin. The latter originated from the different atomic orbitals.

2.1 Hall effect

The Hall effect, the production of a voltage difference across an electrical conductor that is transverse to an electric current in the conductor and an applied magnetic field perpendicular to the current, has a quantized version. Quantum Hall Effect, QHE, is a consequence of the formation of well-defined Landau levels and is thus only possible in high-mobility samples and two-dimensional electron systems with strong external magnetic fields [5] with low temperatures. In QHE, the Hall conductance (the inverse of the Hall resistance, is extremely precise. It is a fractional multiple of e^2/h , with e being the electron's charge and h Planck's constant [6].

The anomalous Hall effect, or extraordinary Hall effect, occurs in ferromagnetic materials (and paramagnetic materials in a magnetic field) when the Hall resistivity includes an additional contribution that depends on the magnetization of the material and it is larger than the ordinary Hall effect. Thus it requires a combination of magnetic polarization and spin-orbit coupling to generate a finite Hall voltage. Thus, the quantized version is the quantum anomalous Hall effect (QAH), which arises when long-range ferromagnetic order is introduced into a topological insulator two-dimensional film, which opens an exchange gap on the surface and chiral edge states are formed due to broken time-reversal symmetry. Also, as the quantum Hall effect, the Hall conductivity acquires quantized values proportional to integer multiples of e^2/h .

2.2 Zeeman effect

The Normal Zeeman Effect is the splitting of spectral lines of an atomic spectrum caused by the interaction between the external magnetic field and the orbital magnetic moment. However, the Anomalous Zeeman Effect is due to the interaction between the magnetic field and the combined orbital and intrinsic magnetic momentum. Thus, in our topological insulator, we will have this effect. The effect mentioned is observed in the presence of an electron spin.

3 Results and discussion

With the experimental data we had and using a code in Fortran, we analyzed different situations in the topological insulator.

We started analyzing the dimensions of the sample, to have the effects that characterize a topological insulator and to have the energy band spectra. We got that we need a rectangular insulator with the dimensions of the y-axis greater than the z-axis, thus we need a flat conductor. Also, to avoid backscattering, we take an infinite insulator on the x-axis. Thus, we had a fixed length of the nanowire on the y and z axes and infinite along the x-axis. The dimensions we have are 200 Å on the y-axis and 50 Å on the z-axis. The scheme of the topological insulator is sketched in Fig. 1

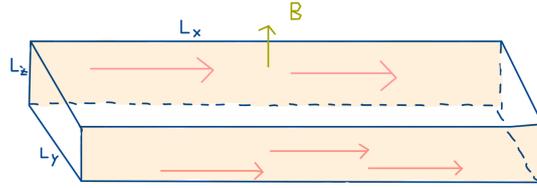


Figure 1: Topological insulator with a magnetic field on the z-axis

First, we analyzed the regime in which we did not have the Zeeman effect, thus the magnetic field is 0 T (Fig. 2). We solve the Schrödinger equation and we obtained the eigenvalues, that is the energies. Then, we represent the energies as a function of the wavenumber, the energy spectra. We have a gap between the upper band (conduction band) and lower band (valence band). This gap is typical of semiconductors.

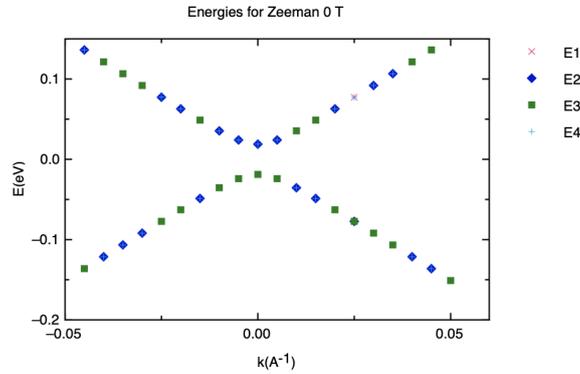


Figure 2: Energies as a function of the wavenumber with 0 Zeeman

Then, we add a magnetic field, we chose 0,2 T thus we get the anomalous Zeeman effect in which the spin blocks the direction of transport. It occurs due to the presence of both orbital and intrinsic magnetic moments. The gap among the bands that we had without the magnetic field disappears and in this case, we have surface states which guarantee electronic conduction. (Fig. 3).

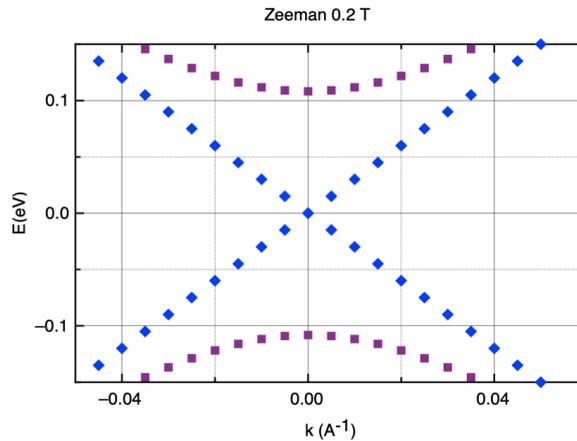


Figure 3: Energies as a function of the wavenumber with Zeeman of 0.2 T

We obtain two bands that cross the 0-point-energy with a 0 value of the wavenumber. The band that has a negative slope means that the group velocity of the electrons is negative, and the band that has a positive slope has electrons with positive velocity, as the group velocity is proportional to the derivative of the energy with respect to the wavenumber. The Fermi level (the thermodynamic work required to add one electron to the body) falls within the bulk band gap which is traversed by topologically-protected spin-textured Dirac surface states.

We now represent the energy as a function of the magnetic field for the different states (Fig. 4), when the wavenumber is 0 and the magnetic field is 0,2 T.

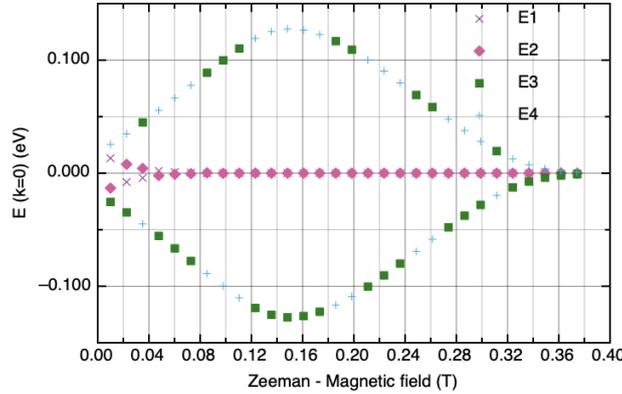


Figure 4: Energies as a function of the magnetic field

The gap closes for low values of the magnetic field, as shown in Fig. 5.

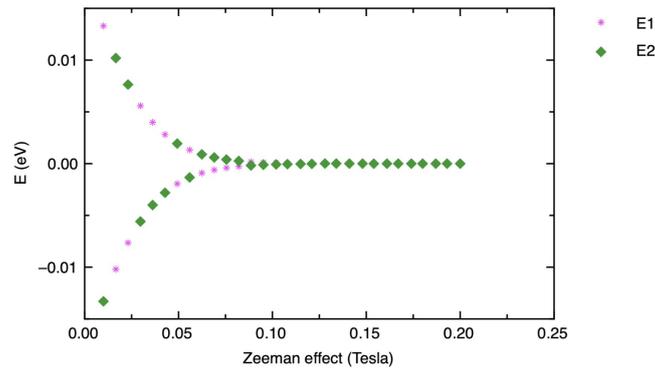


Figure 5: Energies as a function of low values of the magnetic field

Taking this into account we study four different states (states 1,2,3 and 4 as shown in Fig. 6) with a value of the wavenumber of -0.01 \AA^{-1} , and a magnetic field of 0.2 T to ensure the low-energy regime. The analysis will be analogous if we took a positive value of the wavenumber. State 1 has spin-up and state 2 has spin-down, both of them located in the bands that cross the 0 energy and wavenumber.

We first represent the density as a function of the length along the y and z axis for the four types of states (Fig. 7).

We get that in the case of type 1 and type 2 states, the density is located completely on one side of the conductor, on the left and right side of the y-axis respectively. This is due to the anomalous Hall effect. If we go to higher states, that is state 3 and 4 the density is more spread along the system. If we were to higher states, that is states with higher energy for the same wavenumber, we would have all the density along the system.

Then we represented the magnetization as a function of y and z. We obtain that the magnetization along the x-axis is irrelevant compared with the magnetization along the y-axis. This magnetization vanishes when doing the calculus due to the precision. The magnetization along the z-axis is also irrelevant. Thus, we only represented the magnetization along the y-axis (Fig. 8).

When representing the magnetization, we obtain, as we expected and again due to the manifestation of the anomalous Hall effect that for states 1 and 2, we have positive magnetization in one

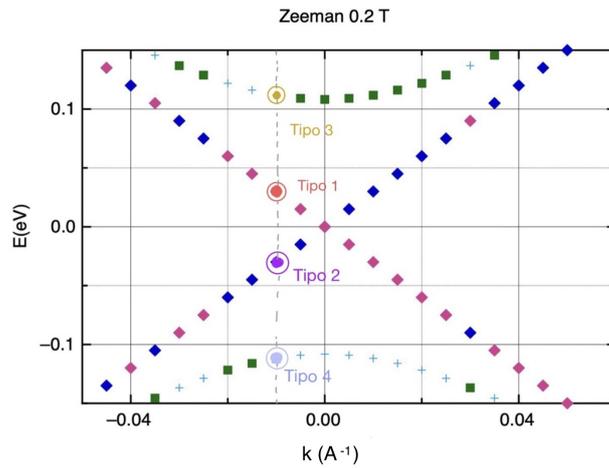


Figure 6: Energies as a function of the wavenumber with Zeeman of 0.2 T

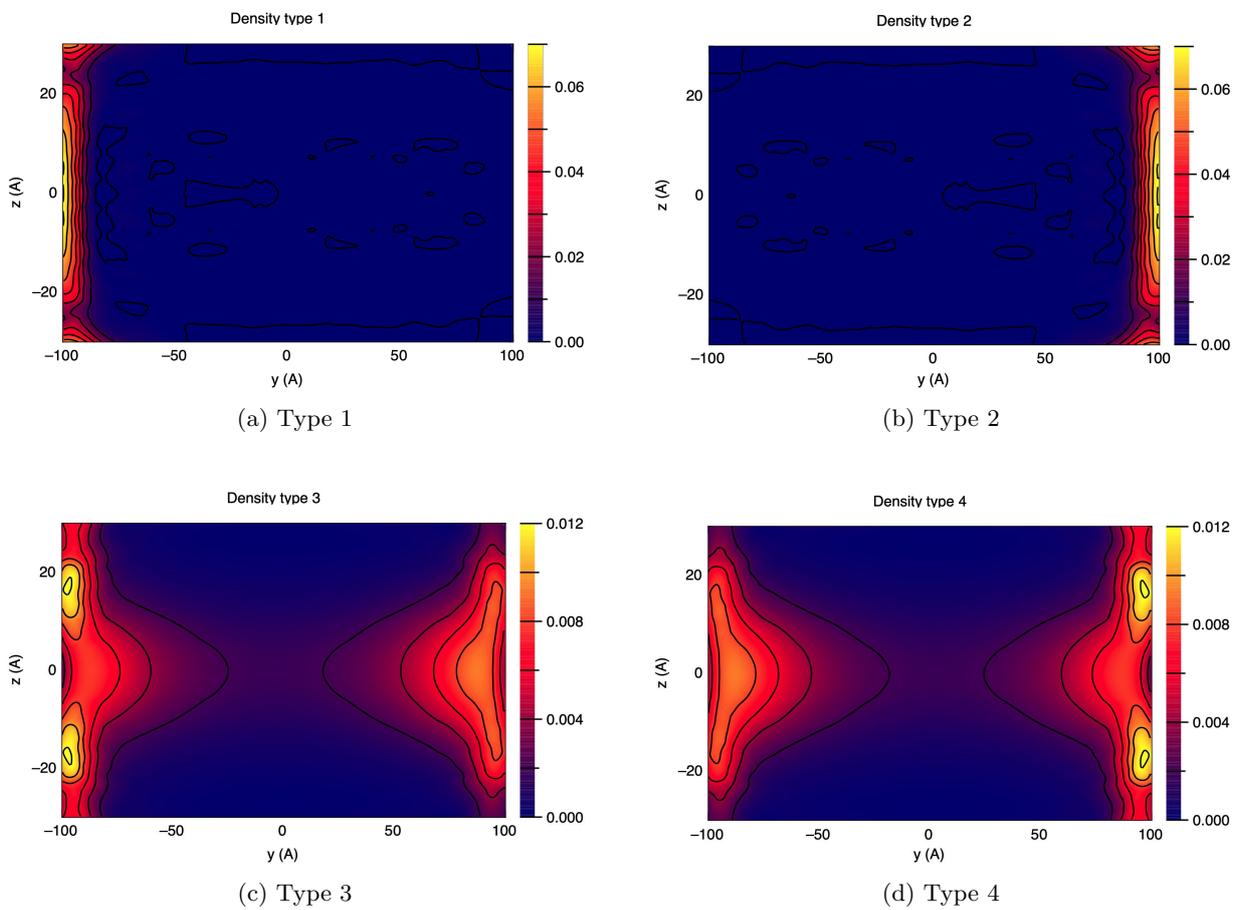


Figure 7: Density of states along the y-axis depending on the type of state

corner of one axis and negative magnetization in the positive or negative part, respectively of that axis. As it happened with the density, if we go to higher states, the magnetization spreads along the system.

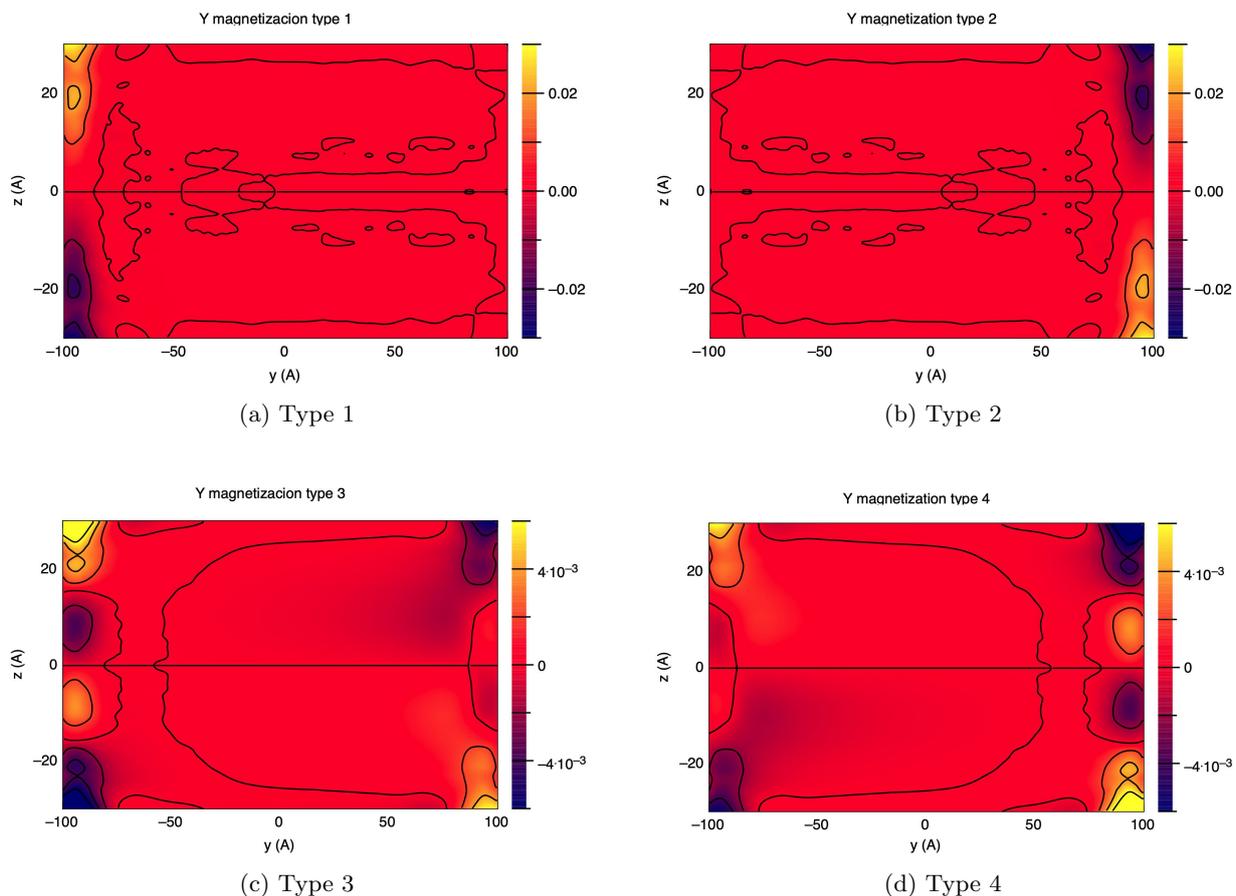


Figure 8: Magnetization along the y-axis depending on the type of state

3.1 Dirac paradox

At the interface between two massless Dirac models with opposite helicity, the electron can neither be transmitted nor reflected due to the spin-momentum locking. The helicity of the electron eigenstates of the massless Dirac cones is a sign encoding the locking between the propagation direction and the orientation of the spin degree of freedom. In 1 dimension (plus one temporal dimension) (Fig. 9), electrons cannot escape along the interface. In (Fig. 10) we obtain that if on the left-hand side of the junction, right-moving electrons are characterized by spin-up and left-moving electrons by spin-down and opposite on the right-hand side, a spin-up electron from the left going onto the interface can neither be transmitted nor reflected due to spin conservation. [2].



Figure 9: Dirac paradox in a system in 1 dimension

If in our model we only have massless Dirac modes we do not have a solution or we have trivial solutions. If we have additional massive Dirac modes, we can have non-trivial scattering solutions. If we have an inhomogeneous profile and therefore the helicity changes, the continuity equation forbids the existence of scattering states.

In 3D, we can have an evolution of the spin. The electron can jump from one side to another through the heterojunction and depending on the properties of this heterojunction.

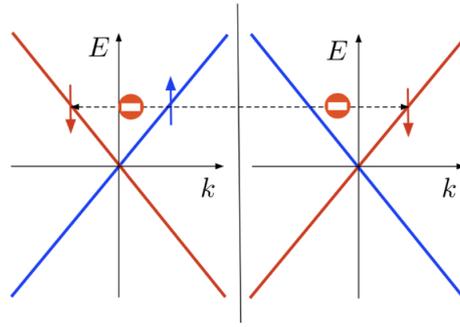


Figure 10: Dirac paradox in 1 dimension at a heterojunction between two massless Dirac models with opposite helicity. Figure from Gogin et al.

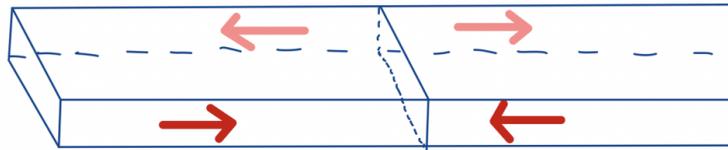


Figure 11: Dirac paradox in a system of 3 dimensions.

4 Conclusions

In this research, we have studied the properties of topological insulators when varying dimensions and magnetic fields. The results of the Dirac paradox are a conceptual advance in understanding heterojunctions and the limits we have when we do not have a topological insulator in three dimensions.

The importance of studying these types of systems and the advantages they have are due to their complexity. The complexity of topological insulators is since there are emerging properties, interactions between the components of the topological insulator, and thus between the properties

As an application, wires made of a topological insulator could enable highly stable qubits, the building blocks of future quantum computers. For instance, spintronic devices and dissipationless transistors. That is why it is important to deep into the research as they are going to be an important part of the second quantum revolution. Not only are they going to be important in quantum computers, but also in magnetoelectronic and optoelectronic.

Topological insulators need to be more studied. A possible extension of this research project could be studying the characteristics of the high-symmetry of the electronic bands in different materials.

Acknowledgments

I would like to thank the IFISC, Institute for Cross-Disciplinary Physics and Complex Systems, and the SURF (Summer Undergraduate Research Fellowship) program to allow me to work on this. I also would like to thank my advisor, Llorenç Serra, for all his work and for the time he spent with me working on this project.

This work was supported by the SURF@IFISC fellowship.

References

- [1] M.Z.Hasan and C.L.Kane, *Topological insulators* (2010).
- [2] Javier Osca, Kristof Moors, Bart Sorée and Llorenç Serra, *Fabry–Pérot interferometry with gate-tunable 3D topological insulator nanowires* (2021).
- [3] Leonid Gogin, Lorenzo Rossi, Fausto Rossi and Fabrizio Dolcini, *The Dirac paradox in 1 + 1 dimensions and its realization with spin–orbit coupled nanowires* (2022).
- [4] Chao-Xing Liu, Xiao-Liang Qi, HaiJun Zhang, Xi Dai, Zhong Fang, and Shou-Cheng Zhang, *Model Hamiltonian for topological insulators* (2010).
- [5] Gang Qiu, Peng Zhang, Peng Deng, Su Kong Chong, Lixuan Tai, Christopher Eckberg, and Kang L. Wang, *Experimental Observation of the Quantum Anomalous Hall Effect in a Magnetic Topological Insulator* (2013).
- [6] Gang Qiu, Peng Zhang, Peng Deng, Su Kong Chong, Lixuan Tai, Christopher Eckberg, and Kang L. Wang, *Mesoscopic Transport of Quantum Anomalous Hall Effect in the Submicron Size Regime* (2022).